

ISSN 2713-3192  
DOI 10.15622/ia.19.6  
<http://ia.spcras.ru>

ТОМ 19 № 6

# ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

# INFORMATICS AND AUTOMATION



СПб ФИЦ РАН

Санкт-Петербург  
2020

# INFORMATICS AND AUTOMATION

Volume 19 № 6, 2020

Scientific and educational journal primarily specialized in computer science, automation, robotics, applied mathematics, interdisciplinary research

Founded in 2002

---

## Founder and Publisher

St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)

---

## Editor-in-Chief

**R. M. Yusupov**, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia

---

## Editorial Council

<b>A. A. Ashimov</b>	Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan
<b>N. P. Veselkin</b>	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
<b>I. A. Kalyaev</b>	Prof., Dr. Sci., Academician of RAS, Taganrog, Russia
<b>Yu. A. Merkur'yev</b>	Prof., Dr. Sci., Academician of the Latvian Academy of Sciences, Riga, Latvia
<b>A. I. Rudskoi</b>	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
<b>V. Sgurev</b>	Prof., Dr. Sci., Academician of the Bulgarian Academy of Sciences, Sofia, Bulgaria
<b>B. Ya. Sovetov</b>	Prof., Dr. Sci., Academician of RAE, St. Petersburg, Russia
<b>V. A. Soyfer</b>	Prof., Dr. Sci., Academician of RAS, Samara, Russia

---

## Editorial Board

<b>O. Yu. Gusikhin</b>	Ph. D., Dearborn, USA
<b>V. Delic</b>	Prof., Dr. Sci., Novi Sad, Serbia
<b>A. Dolgui</b>	Prof., Dr. Sci., St. Etienne, France
<b>M. N. Favorskaya</b>	Prof., Dr. Sci., Krasnoyarsk, Russia
<b>M. Zelezny</b>	Assoc. Prof., Ph.D., Plzen, Czech Republic
<b>H. Kaya</b>	Assoc. Prof., Ph.D., Utrecht, Netherlands
<b>A. A. Karpov</b>	Assoc. Prof., Dr. Sci., St. Petersburg, Russia
<b>S. V. Kuleshov</b>	Dr. Sci., St. Petersburg, Russia
<b>A. D. Khomonenko</b>	Dr. Sci., St. Petersburg, Russia
<b>D. A. Ivanov</b>	Prof., Dr. Habil., Berlin, Germany
<b>K. P. Markov</b>	Assoc. Prof., Ph.D., Aizu, Japan
<b>R. V. Meshcheryakov</b>	Prof., Dr. Sci., Moscow, Russia
<b>N. A. Moldovian</b>	Prof., Dr. Sci., St. Petersburg, Russia
<b>V. Yu. Osipov</b>	Dr. Sci., St. Petersburg, Russia
<b>V. K. Pshikhopov</b>	Prof., Dr. Sci., Taganrog, Russia
<b>A. L. Ronzhin</b>	Prof., Dr. Sci., Deputy Editor-in-Chief, St. Petersburg, Russia
<b>H. Samani</b>	Assoc. Prof., Ph.D., New Taipei City, Taiwan, Province of China
<b>V. Skormin</b>	Prof., Ph.D., Binghamton, USA
<b>A. V. Smirnov</b>	Prof., Dr. Sci., St. Petersburg, Russia
<b>B. V. Sokolov</b>	Prof., Dr. Sci., St. Petersburg, Russia

---

**Editor:** E. P. Miroshnikova

**Technical editor:** M. S. Avstriyskaya

**Interpreter:** E.N. Mesheryakova

---

## Editorial office address

14-th line V.O., 39, SPIIRAS, St. Petersburg, 199178, Russia,

e-mail: ia@spcras.ru, web: <http://ia.spcras.ru>

**The journal is indexed in Scopus**

The journal is published under the scientific-methodological supervision of Department for Nanotechnology and Information Technology of the Russian Academy of Sciences

© St. Petersburg Federal Research Center of the Russian Academy of Sciences, 2020

# ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 19 № 6, 2020

Научный, научно-образовательный журнал с базовой специализацией  
в области информатики, автоматизации, робототехники, прикладной математики  
и междисциплинарных исследований.

Журнал основан в 2002 году

---

## Учредитель и издатель

Федеральное государственное бюджетное учреждение науки  
«Санкт-Петербургский Федеральный исследовательский центр Российской академии наук»  
(СПб ФИЦ РАН)

---

## Главный редактор

**Р. М. Юсупов**, чл.-корр. РАН, д-р техн. наук, проф., Санкт-Петербург, РФ

---

## Редакционный совет

<b>А. А. Ашимов</b>	академик Национальной академии наук Республики Казахстан, д-р техн. наук, проф., Алматы, Казахстан
<b>Н. П. Веселкин</b>	академик РАН, д-р мед. наук, проф., Санкт-Петербург, РФ
<b>И. А. Каляев</b>	академик РАН, д-р техн. наук, проф., Таганрог, РФ
<b>Ю. А. Меркурьев</b>	академик Латвийской академии наук, д-р, проф., Рига, Латвия
<b>А. И. Рудской</b>	академик РАН, д-р техн. наук, проф., Санкт-Петербург, РФ
<b>В. Сгурев</b>	академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария
<b>Б. Я. Советов</b>	академик РАН, д-р техн. наук, проф., Санкт-Петербург, РФ
<b>В. А. Соيفер</b>	академик РАН, д-р техн. наук, проф., Самара, РФ

---

## Редакционная коллегия

<b>О. Ю. Гусихин</b>	д-р наук, Диаборн, США
<b>В. Делич</b>	д-р техн. наук, проф., Нови-Сад, Сербия
<b>А. Б. Долгий</b>	д-р наук, проф. Сент-Этьен, Франция
<b>М. Железны</b>	д-р наук, доцент, Пльзень, Чешская республика
<b>Д. А. Иванов</b>	д-р экон. наук, проф., Берлин, Германия
<b>Х. Кайя</b>	д-р наук, доцент, Утрехт, Нидерланды
<b>А. А. Карпов</b>	д-р техн. наук, доцент, Санкт-Петербург, РФ
<b>С. В. Кулешов</b>	д-р техн. наук, Санкт-Петербург, РФ
<b>К. П. Марков</b>	д-р наук, доцент, Аизу, Япония
<b>Р. В. Мещеряков</b>	д-р техн. наук, проф., Москва, РФ
<b>Н. А. Молдовян</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>В.Ю. Осипов</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>В. Х. Пшихопов</b>	д-р техн. наук, проф., Таганрог, РФ
<b>А. Л. Ронжин</b>	д-р техн. наук, проф., зам. главного редактора, Санкт-Петербург, РФ
<b>Х. Самани</b>	д-р наук, доцент, Синьбэй, Тайвань, КНР
<b>В. А. Скормин</b>	д-р наук, проф., Бингемптон, США
<b>А. В. Смирнов</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>Б. В. Соколов</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>Л. В. Уткин</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>М. Н. Фаворская</b>	д-р техн. наук, проф., Красноярск, РФ
<b>А. Д. Хомоненко</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>Л. Б. Шереметов</b>	д-р техн. наук, Мехико, Мексика

---

**Выпускающий редактор:** Е. П. Мирошникова

**Технический редактор:** М. С. Австрийская

**Переводчик:** Е. Н. Мещерякова

---

## Адрес редакции

199178, г. Санкт-Петербург, 14-я линия В.О., д. 39

e-mail: [ia@spcras.ru](mailto:ia@spcras.ru), сайт: <http://ia.spcras.ru>

**Журнал индексируется в международной базе данных Scopus**

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий,  
в которых должны быть опубликованы основные научные результаты диссертации  
на соискание ученой степени доктора и кандидата наук»

Журнал выпускается при научно-методическом руководстве Отделения нанотехнологий  
и информационных технологий Российской академии наук

© Федеральное государственное бюджетное учреждение науки

«Санкт-Петербургский Федеральный исследовательский центр Российской академии наук», 2020  
Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных  
в составе печатного периодического издания - журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ»  
статей по текущим экономическим, политическим, социальным и религиозным вопросам  
с обязательным указанием имени автора статьи и печатного периодического издания

## CONTENTS

### ***Mathematical Modeling, Numerical Methods***

S. Mikoni, D. Burakov  
JUSTIFICATION AND CLASSIFICATION OF EVALUATION FUNCTIONS USED IN RATING METHODS OF MULTI-CRITERIA CHOICE 1131

A. Batenkov, K. Batenkov, A. Bogachyov, V. Mishin  
MATHEMATICAL MODEL OF OBJECT CLASSIFIER BASED ON BAYESIAN APPROACH 1166

O. German, S. Nasr  
NEW METHOD FOR OPTIMAL FEATURE SET REDUCTION 1198

### ***Artificial Intelligence, Knowledge and Data Engineering***

C. Sarmiento, J. Savage  
COMPARISON OF TWO OBJECTS CLASSIFICATION TECHNIQUES USING HIDDEN MARKOV MODELS AND CONVOLUTIONAL NEURAL NETWORKS 1222

H.N. Phat, N.T.M. Anh  
VIETNAMESE TEXT CLASSIFICATION ALGORITHM USING LONG SHORT TERM MEMORY AND WORD2VEC 1255

### ***Digital Information Telecommunication Technologies***

S. Dvornikov, S. Dvornikov  
EMPIRICAL APPROACH TO ESTIMATING THE IMMUNITY OF PHASE MODULATION SIGNALS WITH A CONTINUOUS PHASE 1280

I. Ereemeev, M. Tatarka, F. Shuvaev, A. Cyganov  
COMPARATIVE ANALYSIS OF CENTRALITY MEASURES OF NETWORK NODES BASED ON PRINCIPAL COMPONENT ANALYSIS 1307

M. Peregudov, A. Steshkovoy  
CENTRALLY RESERVED ACCESS MODEL TO THE MEDIUM IN DIGITAL RADIO COMMUNICATION NETWORKS 1332

## СОДЕРЖАНИЕ

### **Математическое моделирование и прикладная математика**

С.В. Микони, Д.П. Бураков  
ОБОСНОВАНИЕ И КЛАССИФИКАЦИЯ ОЦЕНОЧНЫХ ФУНКЦИЙ, ПРИМЕНЯЕМЫХ В РЕЙТИНГОВЫХ МЕТОДАХ МНОГОКРИТЕРИАЛЬНОГО ВЫБОРА 1131

А.А. Батенков, К.А. Батенков, А.Г. Богачев, В.В. Мишин  
МАТЕМАТИЧЕСКАЯ МОДЕЛЬ КЛАССИФИКАТОРА ОБЪЕКТОВ НА ОСНОВЕ БАЙЕСОВСКОГО ПОДХОДА 1166

О.В. Герман, С.Н. Наср  
НОВЫЙ МЕТОД ОПТИМАЛЬНОГО СОКРАЩЕНИЯ МНОЖЕСТВА ПРИЗНАКОВ 1198

### **Искусственный интеллект, инженерия данных и знаний**

К. Сармьенто, Х. Саваж  
СРАВНЕНИЕ ДВУХ МЕТОДОВ КЛАССИФИКАЦИИ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕМ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ И СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ 1222

Х.Н. Фат, Н.Т.М. Ань  
АЛГОРИТМ КЛАССИФИКАЦИИ ВЬЕТНАМСКОГО ТЕКСТА С ИСПОЛЬЗОВАНИЕМ ДОЛГОЙ КРАТКОСРОЧНОЙ ПАМЯТИ И WORD2VEC 1255

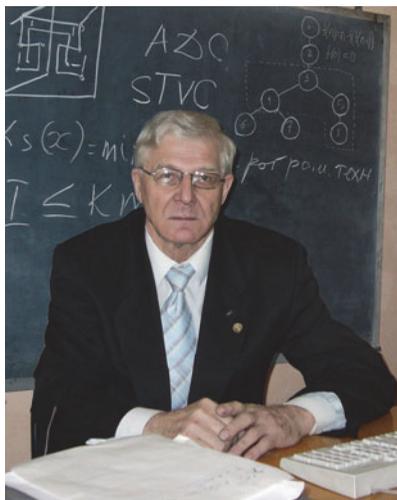
### **Цифровые информационно-телекоммуникационные технологии**

С.В. Дворников, С.С. Дворников  
ЭМПИРИЧЕСКИЙ ПОДХОД К ОЦЕНКЕ ПОМЕХОУСТОЙЧИВОСТИ СИГНАЛОВ ФАЗОВОЙ МОДУЛЯЦИИ 1280

И.Ю. Еремеев, М.В. Татарка, Ф.Л. Шуваев, А.С. Цыганов  
АНАЛИЗ МЕР ЦЕНТРАЛЬНОСТИ УЗЛОВ СЕТЕЙ НА ОСНОВЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ 1307

М.А. Перегудов, А.С. Стешковой  
МОДЕЛЬ ЦЕНТРАЛИЗОВАННО-ЗАРЕЗЕРВИРОВАННОГО ДОСТУПА К СРЕДЕ В СЕТЯХ ЦИФРОВОЙ РАДИОСВЯЗИ 1332

## ПАМЯТИ УЧЕНОГО



2 ноября 2020 года в возрасте 81 года ушел из жизни замечательный ученый, заслуженный деятель науки РФ, доктор технических наук, профессор Виктор Васильевич Александров.

Виктор Васильевич стоял у истоков создания Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН), а затем многие годы возглавлял лабораторию автоматизации научных исследований СПИИРАН.

После окончания Ленинградского физико-механического техникума в 1958 году начал работать в Ленинградском научно-исследовательском радиотехническом институте (ЛНИРТИ). Участвовал в летных испытаниях радиоэлектронной аппаратуры – доплеровских и астроинерциальных систем. Без отрыва от производства учился и в 1964 году получил диплом с отличием Ленинградского политехнического института по специальности «Автоматика и телемеханика». Работая в 1965–1967 годах в Лаборатории бионики ЛНИРТИ, разработал цифровые методы классификации и распознавания на основе алгоритмов быстрых преобразований Адамара – Уолша.

Большая часть жизни Виктора Васильевича тесно связана с наукой и работой в научных организациях. В 1972-1975 годах был доцентом кафедры биомедицины ЛЭТИ; с 1975 года – стал заведующим лабораторией вновь созданной организации Академии наук: с 1975 года – это ЛВЦ (отдел ФТИ им. Иоффе), с 1978 года – ЛНИИВЦ АН СССР, с 1985 года – ЛИИАН АН СССР, с 1991 года – СПИИРАН.

За выдающиеся успехи в науке в 1990 году Виктор Васильевич был удостоен премии Центрального Комитета КПСС, а в 1993 году – почетного звания «Заслуженный деятель науки РФ» и Премии Джона фон Неймана по информатике. Награжден медалью «В память 300-летия Санкт-Петербурга» и медалью ордена «За заслуги перед Отечеством» II степени.

Научные исследования Виктора Васильевича были связаны с новым подходом к пространственным преобразованиям структур

данных. Этот подход основан на предложенном и развитом им аппарате самоподобного иерархического развертывания  $n$ -мерных пространств, использующим принцип заполнения пространства кривых (ЗПК), что послужило основой для разработки адаптивно-динамической структуры данных как компьютерного аналога декартовой системы координат, организации ассоциативной адресации и поиска информации в ЭВМ и построения специализированных ЭВМ пирамидально-рекурсивной архитектуры.

Выдающийся ученый, В.В. Александров создал научную школу в области обработки данных, изображений и сигналов. Им предложено и развито новое научное направление имитационного моделирования развивающихся систем и методы рекурсивно-фрактального синтеза. Разработаны теоретические основы программируемой технологии обработки данных и семантического анализа аудио- и видеопотоков данных.

Результаты исследований Виктора Васильевича Александра опубликованы в 25 монографиях и более чем 300 научных статьях, часть которых опубликована и в журнале «Труды СПИИРАН».

Коллеги помнят Виктора Васильевича как выдающегося ученого, блестящего лектора, футуролога, предвидевшего научные и социальные изменения, и в то же время как доброго и отзывчивого человека. Его жизненная и профессиональная позиция снискала заслуженное уважение коллег и учеников. Нам всем будет очень не хватать его! Но нам остались его книги, статьи и память о нем...

С.В. МИКОНИ, Д.П. БУРАКОВ

**ОБОСНОВАНИЕ И КЛАССИФИКАЦИЯ ОЦЕНОЧНЫХ  
ФУНКЦИЙ, ПРИМЕНЯЕМЫХ В РЕЙТИНГОВЫХ МЕТОДАХ  
МНОГОКРИТЕРИАЛЬНОГО ВЫБОРА**

*Микони С.В., Бураков Д.П. Обоснование и классификация оценочных функций, применяемых в рейтинговых методах многокритериального выбора.*

**Аннотация.** Проанализированы предложенные ранее исследователями рекомендации по применению методов многомерного оценивания объектов. Отмечена слабая обоснованность этих рекомендаций, следующая из поверхностной систематизации методов многомерного оценивания. Рекомендации ориентированы не на классы задач многомерного оценивания объектов, а на различные области человеческой деятельности. Однако в каждой сфере человеческой деятельности имеет место широкий спектр задач оценивания объектов различной природы. В связи с этим признана актуальность более тщательной систематизации методов многомерного оценивания.

Учитывая разноплановость методов многомерного оценивания, решено ограничиться систематизацией методов, применяющих оценочные функции, и на этой основе предложить общие рекомендации по их применению.

Обзор методов многомерного оценивания с единой позиции потребовал уточнения применяемой в них терминологии. На основе формальной модели установлены отношения между понятиями «предпочтение», «критерий» и «показатель». Для выделения методов, применяющих оценочные функции, введено понятие целевого значения показателя. Относительно его расположения на шкале показателя введены понятия идеальной и реальной целей. Соответствующие этим целям критерии разделены на целевые и ограничительные. С применением предложенной терминологии проанализированы наиболее известные методы многомерного оценивания. Из них выделена группа методов, применяющих оценочные функции.

Рассмотрены варианты оценочных функций, создаваемых на основе критерия и постулатов теории ценности и полезности. На основе сходства областей определения и значений различных оценочных функций установлена взаимосвязь между ними. Относительно целевого значения показателя они разделены на функции достижения цели и функции отклонения от цели. Показана взаимная дополнительность этих функций. Выделена группа функций отклонения от цели, которая позволяет упорядочивать объекты раздельно по штрафам и поощрениям относительно достижения реальной цели. Для отношения соответствия введено понятие нормы. На примере медицинских анализов показано практическое применение функций отклонения от нормы с применением как минимаксной, так и средневзвешенной обобщающей функции для установления рейтинга на множестве объектов.

Выявленное в процессе исследования сходство и различие оценочных функций положено в основу классификации использующих их методов многомерного оценивания. Различие оценочных функций по трудоемкости их создания отражено в предложенной методике их применения.

**Ключевые слова:** предпочтение, показатель, критерий, целевое значение, оценочная функция, функция ценности, функция полезности, достижение цели, отклонение от цели, функциональный выбор, многомерное оценивание объектов, рейтинговый метод

**1. Введение.** Извечная потребность человека в оценивании объектов любого вида и происхождения воплотилась в теоретические изыскания, начатые трудами [1-5]. Именно в послевоенный период с

развитием информатики как средства реализации методов принятия решений было разработано их теоретическое обоснование. В последующие годы методы теории принятия решений (ТПР) разделились на отдельные группы и конкретизировались применительно к особенностям оцениваемых объектов. Ежегодно публикуются сотни работ, в которых методы ТПР используются для решения конкретных практических задач. Методы оценивания объектов по многим показателям (многомерного оценивания – ММО) нашли широкое применение на практике, охватывая все виды человеческой деятельности и объекты любой природы [6-12].

Ввиду многочисленности прикладных публикаций их общий обзор не представляется возможным. Он может быть сужен конкретной сферой приложения методов ТПР. Другой путь – сопоставление наиболее популярных методов многомерного оценивания с целью выявления их особенностей. Остановимся для примера на двух работах по этой теме, отечественной и зарубежной, опубликованных в 2013 году. Перечень рассматриваемых в них методов многокритериального выбора включает как методы оптимизации, изучаемые в рамках исследования операций, теории управления и искусственного интеллекта, так и подходы, основанные на моделировании предпочтений лица, принимающего решение (ЛПР). Методы первой группы, такие как линейное программирование и эволюционное моделирование, применяются для решения задач многокритериального выбора. В частности, целевое программирование рассматривается как обобщение линейного программирования в случае наличия нескольких целей оптимизации.

В [13] достаточно полно охвачены классические методы оптимизации. В ней приводится обширный перечень как отечественных, так и зарубежных средств решения задач многокритериального выбора. Но данное исследование носит информационный характер, поскольку в нем отсутствует сопоставление рассматриваемых методов и средств их реализации. В работе [14] представлен обзор англоязычных работ в области многокритериальной оптимизации. Как отмечают авторы, они рассмотрели наиболее популярные методы многомерного оценивания (в их терминах – Multi-Criteria Decision Making Methods, сокращенно – MCDM), к которым относятся:

1. Многомерная теория полезности (Multi-Attribute Utility Theory – MAUT).
2. Метод анализа иерархий (Analytic Hierarchy Process – AHP).
3. Теория нечётких множеств (Fuzzy Set Theory – FST).
4. Рассуждения по прецедентам (Case-Based Reasoning – CBR).
5. Анализ данных (Data Envelopment Analysis – DEA).

6. Простой метод многомерного упорядочения (Simple Multi-Attribute Rating Technique – SMART).
7. Целевое программирование (Goal Programming – GP).
8. Метод «исключение и выбор, отражающие реальность» (ELECTRE).
9. Метод организации ранжирования предпочтений для оценки обогащенного оценивания (PROMETHEE).
10. Простое аддитивное взвешивание (Simple Additive Weithing – SAW).
11. Метод упорядоченного предпочтения через сходство с идеальным решением (Technique for Order of Preference by Similarity Ideal Solution – TOPSIS).

Здесь следует отметить некорректность перечисления авторами в одном ряду как методов MCDM, так и используемых для их разработки теорий (например, FST само по себе не является методом MCDM, хотя элементы теории нечётких множеств активно используются, в том числе и в методах многомерного оценивания). В отличие от [13] в данной работе на основе анализа особенностей рассматриваемых методов предлагаются рекомендации по областям их применения. В основу рекомендаций положена трудоемкость построения модели и размерность задачи оценивания.

К недостаткам обзора следует отнести поверхностность рекомендаций по применению рассматриваемых методов. Рекомендации ориентированы не на классы задач ММО объектов, а на различные области деятельности. Однако, несмотря на большое различие областей деятельности, их оценивание не ограничивается решением частных задач. В каждой области деятельности имеет место широкий спектр задач оценивания объектов различной природы. Отмеченный недостаток обусловлен, с одной стороны, чрезмерным охватом методов ММО, а с другой стороны, отсутствием теоретического обоснования рекомендаций. Решению этой проблемы должна способствовать разработка научно-обоснованных классификаций методов ММО. Такие классификации, упорядочивая текущее знание, обладают также прогностическим свойством, что позволяет применять системный подход не только для анализа существующих моделей и средств их реализации, но и для создания новых вариантов моделей. Отсутствие таких классификаций в [14] послужило побудительным мотивом для исследований в этом направлении, результаты которых излагаются в настоящей работе.

На понимание задач оценивания объектов и классифицирование применяемых для их решения методов существенно влияет терминология предметной области (Про). Эта про-

блема актуальна для отечественной терминологии, в частности, по причине неточного перевода широко заимствуемых англоязычных терминов. В качестве примера приведем термин «метод анализа иерархий» (МАИ), который можно трактовать как анализ существующей иерархии показателей. В оригинале метод называется «Analytic Hierarchy Process» (АНР). Буквально оно переводится как «Метод аналитической иерархии». А под *иерархией* автор термина Т. Саати понимает структуру «цель → критерии → альтернативы».

Проблема уточнения смысла термина упрощается применением системного подхода к анализу моделей, изложенного в монографии [15]. В ней предложено рассматривать модель в трех аспектах: функциональном (Ф-модель), операционном (О-модель) и структурном (С-модель) и в их сочетаниях. Модель задачи описывается структурно-функциональной моделью (СФ-моделью), а процессы ее *построения* и *решения* описываются операционными, а точнее структурно-операционными моделями (СО-моделями). Примером СО-модели является алгоритм. В этих терминах любой метод относится к классу СО-моделей. Таким образом, термин АНР характеризует именно СО-, а не СФ-модель задачи, а его смысл заключается в *иерархическом процессе оценивания альтернатив*.

Для ясности дальнейшего изложения уточним еще ряд применяемых терминов. Введенный ранее термин «методы многомерного оценивания» (ММО) обобщает методы оценивания объектов по многим показателям: Multi-Attribute Rating Technique, Multi-Attribute Optimization, Multi-Objective Optimization, Multi-Criteria Optimization. Под словом «оптимизация» в них подразумевается цель упорядочения объектов для выбора из них того, который в наибольшей степени удовлетворяет заданным требованиям.

В отечественной литературе часто не делается различий между терминами «показатель» и «критерий». Показателем фиксируется количественное или качественное значение некоторого свойства объекта, а критерием задается требование к значениям этого показателя. Следует упомянуть также о соотношении терминов «атрибут» и «показатель». Слово *attribute* имеет смысл *неотъемлемого свойства* объекта, а показатель используется для отражения результата *измерения* этого свойства в количественной или качественной шкале. В английском языке ему соответствует слово *indicator*.

Поскольку классификация должна охватывать методы ММО, изложенные в разных терминах, исследование предваряется объяснением применяемых терминов. На этой основе осуществлен обзор наиболее известных методов ММО. Из них выделена группа методов,

применяющих оценочные функции. Сходство и различие этих функций положено в основу предлагаемой классификации группы методов ММО. Оно используется для выработки общих рекомендаций по применению рассматриваемых методов ММО. Заключительный пример демонстрирует расширенные возможности одного из методов ММО как следствие предложенной классификации.

**2. Краткий обзор методов ММО.** Как было отмечено ранее, в обзоре [14] отсутствует разделение моделей и методов оценивания и отнесение их к определенным классам. Для их первичного различия примем за основу деления принцип *создания* моделей оценивания. Модель ММО создается либо на основе модели оптимизации, либо на основе предпочтений ЛППР. Модели первого типа назовем *оптимизационными*, а второго типа – *экспертными* моделями. Их родоначальниками являются научные направления «Исследование операций», «Теория управления» для моделей первого типа и «Системный анализ» – для моделей второго типа. Несмотря на присутствие оптимизационного и экспертного факторов в моделях обоих типов, предлагаемое разделение удобно в смысле указания на первичность каждого из этих факторов.

Модель предпочтений ЛППР реализует отношения предпочтения:

- на множестве значений  $j$ -го показателя  $R_Y \subset Y_j \times Y_j$ ;
- на множестве показателей  $R_F \subset F \times F$ ;
- на множестве объектов  $R_X \subset X \times X$ .

Модель ПрО инвариантна по отношению к различным моделям предпочтений. Именно они и определяют различие методов ММО, которые и являются предметом дальнейшего рассмотрения.

По способу задания предпочтения разделим модели ММО на *связанные* и *несвязанные* по предпочтениям с альтернативами. Связанная модель предпочтений ЛППР отношения совмещает все три типа отношений предпочтения. В несвязанной модели предпочтений ЛППР отношение предпочтения на множестве объектов реализуется на основе первых двух типов предпочтений.

В связанной модели отношение предпочтения  $R_{пр,j} \subseteq X \times X$  по каждому показателю представляется матрицей парных сравнений (МПС). Значения предпочтений на множестве объектов задаются экспертом. В наиболее концентрированной форме эта модель реализуется в методе АНР Т. Саати [3]. В различных формах этот тип модели реализуется другими методами, названными в англоязычной литературе *нерейтинговыми* методами оценивания (*outranking methods*) [14].

Такое название «от противного» отражает назначение этих моделей. Их основная цель – выбор наилучшей альтернативы (объекта), хотя следствием процесса выбора является упорядочение альтернатив.

К этому классу принадлежат все методы, основанные на применении отношения предпочтения на множестве альтернатив. К ним относятся метод ANP (Analytic Network Process) [16] как обобщение метода АНР, метод ELECTRE (Elimination Et Choix Traduisant la Realite – исключение и выбор, отражающее реальность) [17], PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation) [18] и отечественный метод вербального анализа альтернатив (ВАР), разработанный коллективом академика О. И. Ларичева [19].

Трудоемкость построения связанной модели методом АНР Т. Саати оценивается количеством операций, требуемых для сопоставления  $N$  альтернатив по  $n$  показателям. Она определяется по формуле:

$$Q_{\text{АНР}} = \frac{N(N-1)}{2} \cdot n. \quad (1)$$

Согласно формуле (1) трудоемкость построения сравнительной модели быстро увеличивается с ростом размерности альтернатив, что ограничивает применение этих моделей для решения задач оценивания большой размерности. На начальном этапе развития ТПР это ограничение не являлось критическим. Основное внимание уделялось формализации предпочтений ЛПП и проблеме независимости критериев для предпочтению. Ограничение на размерность задачи оказалось критическим для оценивания большого количества объектов и сложных объектов, характеризующихся десятками показателей. Примером задач первого типа является определение рейтинга 85 регионов РФ по показателям социально-экономического развития [10]. Примером задач второго типа является оценивание качества и технического уровня сложных систем [6]. Для решения таких задач востребованы «рейтинговые» модели и методы оценивания.

В «рейтинговых» методах оценивания не лучшие, то есть неоптимальные объекты не исключаются из рассмотрения, а посему не называются альтернативами. В этих моделях первично упорядочение всех объектов, на основе которого выбираются наилучшие варианты.

Трудоемкость построения иерархической модели ММО на основе несвязанной модели с общими предпочтениями для всех объектов оценивается формулой:

$$Q = n + \frac{1}{2} \sum_{i=1}^k n_i \cdot (n_i - 1). \quad (2)$$

Второе слагаемое в формуле (2) отражает использование сравнительных моделей для вычисления весовых коэффициентов, приме-

няемых для получения средневзвешенных оценок объектов. В отличие от формулы (1), сравнению подвергаются не оцениваемые объекты, а характеризующие их показатели. В иерархической модели ММО показатели разбиваются по их назначению на группы. В  $i$ -ю группу,  $i = \overline{1, k}$ , включается  $n_i$  показателей. Таким образом, большие (многомерные) и сложные (неоднородные) задачи решаются гибридными методами. В данном случае рейтинговый метод (ranking method) оценивания дополняется нерейтинговым методом (outranking method). Так, например, в [20] рассматривается методика многокритериального двухуровневого анализа пунктов размещения гидроэлектростанций, сочетающая в себе этапы применения методов МАУТ и АНР для выбора наилучшего решения.

Формула (2) отражает независимость модели предпочтений ЛПП от модели ПрО или, иными словами, несвязность этих моделей. Она проявляется в том, что объем предпочтений  $Q$  не зависит от числа оцениваемых объектов  $N$ . Это и является основанием для именования методов упорядочения объектов на основе несвязанной модели рейтинговыми методами, поскольку отсутствует ограничение на число оцениваемых объектов.

Для сравнения несопоставимых по ресурсам объектов применяются два способа: вычисление отношения значений показателя к имеющимся ресурсам для обеспечения сопоставимости объектов относительно общего целевого значения  $c_j$  и индивидуальное планирование целевого значения  $j$ -го показателя  $c_{ji}$  для каждого  $i$ -го объекта,  $i = \overline{1, N}$ . Трудоемкость построения иерархической модели ММО на основе нормативной модели с индивидуальными требованиями для  $N$  объектов оценивается формулой:

$$Q = N \cdot n + \frac{1}{2} \sum_{i=1}^k n_i \cdot (n_i - 1). \quad (3)$$

Для вычисления обобщенных оценок объектов по всем показателям применяются различные обобщающие функции (ОФ). Наибольшее распространение на практике получили аддитивная (АОФ) и мультипликативная (МОФ) функции, дающие средневзвешенные оценки объектов [21].

В работе [22] был предложен подход, основанный на замене функций полезности целями (targets). Однако в действительности задание целевых значений не противоречит заданию полезности на шкале показателя, что будет показано ниже.

В работе [4] была поставлена под сомнение парадигма многокритериального оценивания, основанная на получении компромисса в достижении частных целей с применением средневзвешенных оценок объектов. Достижению максимальной полезности (ценности) была противопоставлена парадигма достижения поставленной цели (Reference Objective), трактуемой точкой в  $n$ -мерном пространстве. Утверждалось, что «любая точка в целевом пространстве, независимо от того, достижима она или нет, идеальна или нет, может использоваться вместо весовых коэффициентов для получения функций, которые имеют минимумы только в точках Парето». В качестве таковой функции был предложен минимаксный критерий частных отклонений от поставленной цели (точки в пространстве).

В [23] в качестве источников частных целей предложено использовать ограничительные критерии с реальными целями. Метод условной оптимизации [24] предполагает исключение из упорядочения объектов, не удовлетворяющих ограничениям. В работе [23], как и в [4], все объекты подлежали упорядочению по общим оценкам независимо от достижения ими частных целей. На этом основании он был назван *методом мягких притязаний* или *приближением к образцу*. Однако в отличие от работы [4] не отвергались ОФ, дающие средневзвешенные оценки объектов. Таким образом, независимо от выбранной ОФ методы упорядочения объектов разделим на методы *достижения* и *отклонения* от цели.

**3. Постановка задачи.** Требования к объектам задаются на основе бинарного отношения предпочтения  $R_{пр}$  [21], обладающего свойствами нестрогого порядка, и потому образованного объединением отношений *превосходства*  $R_{>}$  и *соответствия*  $R_{\equiv}$ , представляющих отношения строгого порядка и эквивалентности, то есть  $R_{пр} = R_{>} \cup R_{\equiv}$ .

В функциональной форме отношение предпочтения  $R_{пр} \subseteq X \times X$ , связывающее пару объектов  $x_i, x_k \in X$ , то есть  $x_i Pr_{пр} x_k$ , описывается двухместным предикатом  $Pr_{пр}(x_i, x_k)$ . *Превосходство*  $x_i > x_k$  объекта  $x_i$  над объектом  $x_k$  представляется предикатом  $\succ(x_i, x_k)$ , а *соответствие* объектов  $x_i$  и  $x_k$  – предикатом  $\equiv(x_i, x_k)$ . Отношения превосходства и соответствия объектов  $x_i$  и  $x_k$  по  $j$ -му показателю представляются предикатами  $\succ(f(x_i), f(x_k))$  и  $\equiv(f(x_i), f(x_k))$ , задающими составное отношение предпочтения  $R_{пр/j} = R_{>,j} \cup R_{\equiv,j} \subseteq X \times X$ . Здесь и далее используются следующие обозначения:

1)  $X = \{x_1, \dots, x_i, \dots, x_N\}$  – множество оцениваемых объектов (альтернатив),  $|X| = N$ .

2)  $F = \{f_1, \dots, f_j, \dots, f_n\}$  – множество показателей, отражающих различные свойства объектов,  $|F| = n$ .

3)  $Y_j$  – область значений (домен)  $j$ -го показателя, также называемая шкалой этого показателя. Она определяется как некоторый отрезок  $[y_{j,\min}, y_{j,\max}]$ ,  $j = \overline{1, n}$ .

4)  $y_j = f_j(x)$  – значение  $j$ -го показателя для объекта  $x$ .

При количественных оценках  $j$ -го показателя в качестве предикатов превосходства применяются отношения «>» («быть больше») и «<» («быть меньше»), а в качестве предикатов соответствия – отношения «=» («быть равным») и « $\in [ ]$ » («принадлежать отрезку»).

Предпочтение  $R_{pr,j}$  преобразуется в *критерий* заданием целевого значения  $c_j$  в роли базы сравнения с любой величиной  $y_j = f_j(x)$  на шкале  $j$ -го показателя, задаваемой отрезком  $[y_{j,\min}, y_{j,\max}]$ . Целевое значение  $c_j$  задается либо на границе шкалы показателя ( $c_j = y_{j,\max}$ , или  $c_j = y_{j,\min}$ ) или на точных гранях множества  $Y_j$ , либо в промежуточной точке шкалы ( $y_{j,\min} < c_j < y_{j,\max}$ ). Граничным целям соответствуют критерии  $y_j \rightarrow \max$  и  $y_j \rightarrow \min$ . В [21] они названы *целевыми* критериями, а преследуемые ими цели – *идеальными* целями. Отметим, что здесь и далее символическая запись « $y \rightarrow \text{extr}$ » означает требование поиска экстремального (наибольшего либо наименьшего) значения величины  $y$  на некотором множестве.

Целевой критерий характеризуется однородностью предпочтения на *всей шкале*  $j$ -го показателя. Оно реализуется заданием в качестве базы сравнения в предикате  $Pr(y_j, c_j)$  целевого значения  $c_j$  на одной из границ шкалы  $j$ -го показателя. Базе сравнения  $c_j = y_{j,\min}$  соответствует *максимизация* значений  $j$ -го показателя, то есть  $y_j \rightarrow \max \Leftrightarrow Pr_>(y_j, y_{j,\min})$ , а базе сравнения  $c_j = y_{j,\max}$  – *минимизация* его значений, то есть  $y_j \rightarrow \min \Leftrightarrow Pr_<(y_j, y_{j,\max})$ , поскольку оценка выполнения требования, например максимизации, реализуется путем сравнения значения  $y_j$  с наименьшим возможным значением  $y_{j,\min}$ .

Цель, назначенная в промежутке между граничными значениями шкалы показателя, названа *реальной*, а содержащий ее критерий – *ограничительным* критерием по аналогии с оптимизационной моделью математического программирования. При реализации предпочтения превосходства с включением отношения равенства  $y_j = c_j$  формируются ограничительные критерии «больше или равно» (ограничение «снизу»)  $Pr_{\geq}(y_j, c_j)$  и «меньше или равно» (ограничение «сверху»)  $Pr_{\leq}(y_j, c_j)$ . В инфиксной форме они записываются как  $y_j \geq c_j$  и  $y_j \leq c_j$  соответственно.

При реализации отношения соответствия с базой сравнения  $c_j$ ,  $y_{j,\min} < c_j < y_{j,\max}$  формируется *точечный* ограничительный критерий

рий  $Pr_{=} (y_j, c_j)$ , а с базой сравнения  $[c_{j,н}, c_{j,в}]$  – *интервальный* ограничительный критерий  $Pr_{\lceil} (y_j, [c_{j,н}, c_{j,в}])$ . В этом критерии  $c_{j,н}$  – нижняя ( $y_{j,\min} < c_{j,н}$ ), а  $c_{j,в}$  – верхняя ( $c_{j,в} < y_{j,\max}$ ) граница интервала требуемых значений  $j$ -го показателя, определяемого соответствующим отрезком.

Конечно, называя любой показатель критерием, автор держит требования к нему «в уме». Однако формально значение  $j$ -го показателя  $y_j = f_j(x)$  и оценка его соответствия требуемому значению  $c_j$  двухместным предикатом  $Pr(y_j, c_j) \in \{1, 0\}$  – не одно и то же.

Модель задачи ММО состоит из двух частей [21]: модели ПрО и модели предпочтений ЛПР. Модель ПрО содержит информацию о сопоставляемых объектах и их свойствах. Ее формальной моделью является  $n$ -мерное отношение мощности  $N R_{SD} \subseteq Y_1 \times \dots \times Y_n$ ,  $n = |F|$ ,  $N = |X|$ . В силу конечности множеств  $X$  и  $F$  эта модель представляется прямоугольной таблицей «Объекты/Показатели (Признаки)». В строках таблицы, помеченных именами объектов из  $X$ , перечисляются значения всех показателей для этих объектов, а столбцы озаглавлены именами показателей из множества  $F$ . Произведение  $N \cdot n$  характеризует размерность задачи оценивания. На размерность задачи влияет также неоднородность показателей, проявляющаяся в различии их доменов. Различие заключается не только в принадлежности доменов показателей множествам действительных  $\mathbf{R}$  или целых  $\mathbf{Z}$  (при балльных оценках) чисел, но и в выделенных из них подмножествах для каждого показателя. Иными словами, каждый показатель  $f_j(x)$ ,  $j = \overline{1, n}$ , измеряется в собственной шкале  $Y_j = [y_{j,\min}, y_{j,\max}]$ .

Модель предпочтений ЛПР выражается бинарными отношениями предпочтения  $R_{пр,j}$ , а они, в свою очередь, задают соответствующие им критерии. Как было отмечено выше, в функциональной записи любой критерий, как соответствия, так и превосходства, представляет собой некоторый предикат, принимающий значение «истина» (1), если заданная цель выполняется, и «ложь» (0) — в противном случае, играя роль *оценочной функции*  $f_o(y_j)$   $j$ -го показателя (ОцФ), то есть функции, оценивающей, удовлетворяет ли значение  $j$ -го показателя  $y_j$  заданному требованию. При расширении области значений ОцФ  $f_o(y_j)$  с множества  $\{0, 1\}$  на множество  $[0, 1]$  переходом от двузначной (булевой) логики к бесконечнозначной логике, ее значение характеризует *степень удовлетворения* заданного требования.

Заметим, что в смысле меры ОцФ является показателем, но измеряющим не свойство объекта, а *меру соответствия* этого свой-

ства заданному требованию. Обобщенный показатель, вычисляемый обобщением значений частных ОцФ по некоторому правилу, сам по себе тоже не является критерием. Он становится таковым при задании на его шкале некоторого отношения предпочтения. Для вычисления значений обобщенного показателя для объектов, характеризующихся разнородными показателями, должна использоваться единая шкала. В качестве таковой наиболее востребована абсолютная шкала  $[0, 1]$ , универсальная по отношению к аксиомам теории измерений. Эта шкала является областью значений ОцФ  $f_o(y_j)$ . В случае использования ОцФ с единой шкалой значений  $[0, 1]$  предпочтения на шкале обобщенного показателя естественным образом задаются отношениями  $\{“\geq”, “>”, “\rightarrow \max”\}$  или  $\{“\leq”, “<”, “\rightarrow \min”\}$ .

Разумеется, каждая ОцФ по-своему отображает отношение предпочтения на шкале  $[y_{j,\min}, y_{j,\max}]$   $j$ -го показателя. Однако общность областей определения и значений позволяет свести созданные разными способами ОцФ в единую систему. В основу этих способов положены аксиомы критерия и теории ценности/полезности [25].

Согласно [21] к ОцФ  $f_o(y_j)$ , областью определения которой является шкала  $j$ -го показателя  $Y_j = [y_{j,\min}, y_{j,\max}]$ ,  $j = \overline{1, n}$ , относятся следующие функции:

- ценности  $v$ :  $Y_j \rightarrow [0, 1]$ ;
- полезности  $u$ :  $Y_j \rightarrow [-1, 1]$ ;
- плановая  $s$ :  $Y_j \rightarrow [0, 200\%]$ ;
- принадлежности объекта  $l$ -му классу по  $j$ -му показателю

$\mu_l$ :  $Y_j \rightarrow [0, 1]$ ,  $l = \overline{1, k}$  ;

- отклонения от цели  $d$ :  $Y_j \rightarrow [-1, 1]$ .

В [21] методы, реализующие сопоставление объектов на шкалах ОцФ, были названы методами *функционального выбора* (ФВ) в отличие от методов *критериального выбора* (КВ), реализующих сопоставление объектов на исходных шкалах показателей. Англоязычные имена этих методов (ranking/outranking methods), отражающие их назначение, более ситуативны, чем именование по «принципу действия», так как в процессе применения назначение обозначаемого может расширяться.

Учитывая разноплановость рассмотренных методов ММО, на наш взгляд, затруднительно конкретизировать их применение по примеру работы [14]. Обоснованием рекомендаций по применению методов ММО должна служить систематизация, отражающая сходство и различие методов. В основу такой систематизации положим виды ОцФ, используемые рейтинговыми методами.

Их сходство и различие применим для выработки общих рекомендаций по применению этих методов.

Поэтому далее подробно рассмотрим ОцФ, их свойства и способы применения в рейтинговых задачах, на основании чего построим классификацию методов ФВ и определим возможные варианты их использования.

**4. Оценочные функции, реализующие отношение превосходства.** Согласно самому названию этого отношения и соответствующим ему предикатам реализующие его ОцФ являются монотонными в области определения  $Y_j = [y_{j,\min}, y_{j,\max}]$ ,  $j = \overline{1, n}$ .

ОцФ, создаваемая на основе критерия превосходства, отображает значения из шкалы  $j$ -го показателя  $[y_{j,\min}, y_{j,\max}]$  в шкалу  $[0, 1]$ :  $f_n: Y_j \rightarrow [0, 1]$ . Для предпочтения  $y_j \geq c_j$  функция  $f_{n,\max}(y_j)$  на отрезке  $[y_{j,\min}, c_j]$  вычисляется по формуле:

$$f_{n,\max}(y_j) = \frac{y_j - y_{j,\min}}{c_j - y_{j,\min}} + b. \quad (4)$$

В формуле (4) величина  $b$  – нижняя граница значений функции  $f_{n,\max}$ . Для области значений  $[0, 1]$   $b = 0$ . Линейное отображение  $f_n: Y_j \rightarrow [0, 1]$  называют *нормирующей функцией*. Таким образом, мера достижения объектом  $x$  цели, заданной на границе шкалы, оценивается линейной нормирующей функцией  $f_n(y_j)$   $j$ -го показателя.

В [21] нормирующая функция показателя трактуется как линейный вариант функции ценности  $v(y_j)$  или полезности  $u(y_j)$ . Нормирующая функция  $f_{n,\max}$  единственна в случае идеальной цели, совпадающей с верхней границей шкалы:  $c_j = y_{j,\max}$  (выбор по максимуму) [5]. В [21] она называется функцией *достижения идеальной цели* (ДИЦ). Приближение к цели сопровождается увеличением значения функции  $f_n(y_j)$   $j$ -го показателя и при достижении идеальной цели  $f_n(c_j) = 1$ . Пример функции ДИЦ показан на рисунке 1а линией 1.

В отличие от идеальной цели реальная цель  $c_j$ , задаваемая ограничительными критериями  $\geq(y_j, c_j)$  и  $\leq(y_j, c_j)$ , отвечает условию  $y_{j,\min} < c_j < y_{j,\max}$  на шкале  $[y_{j,\min}, y_{j,\max}]$ . В [21] такая точка  $c_j$  рассматривается как *реальная цель* (target). Реальной цели ставится в соответствие приемлемое значение полезности (ценности)  $f_n(c_j) = u_{jn}$  в диапазоне  $0 < u_{jn} \leq 1$ .

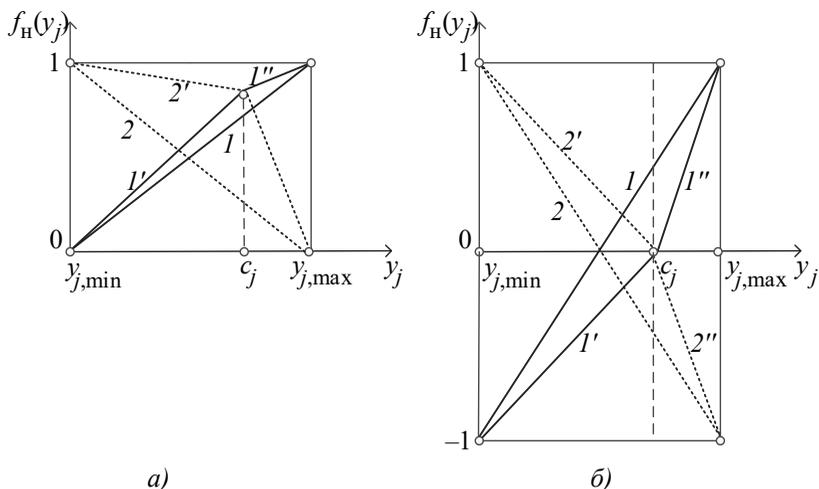


Рис. 1. Линейные и кусочно-линейные варианты оценочных функций:  
 а) – функции ценности; б) – функции полезности

В общем случае, когда  $c_j \neq (y_{j,\max} + y_{j,\min})/2$ , целевое значение  $c_j$  становится точкой излома ОцФ  $j$ -го показателя, а сама функция становится составной. Первая ее часть  $f'_{\text{н},\max}(y_j)$ , отражающая достижение цели для предпочтения  $y_j \geq c_j$  на отрезке шкалы показателя  $[y_{j,\min}, c_j]$ , вычисляется по формуле (1). Вторая часть ОцФ  $j$ -го показателя  $\geq(y_j, c_j)$  – функция  $f''_{\text{н},\max}(y_j)$ , отражающая превышение цели  $c_j$ , вычисляется по формуле:

$$f''_{\text{н},\max}(y_j) = \frac{y_j - c_j}{y_{j,\max} - c_j} + b. \quad (5)$$

В целом ОцФ  $j$ -го показателя, моделирующая предпочтение критерия  $\geq(y_j, c_j)$ , представляет собой кусочно-линейную функцию  $f_{\text{н},\max}$ . В [21] она названа *функцией достижения реальной цели* (ДРЦ). На рисунке 1а первая ее часть  $f'_{\text{н},\max}(y_j)$  представлена линией 1', а вторая часть  $f''_{\text{н},\max}(y_j)$  – линией 1''. Аналогичным образом создаются функции ДРЦ  $f_{\text{н},\min}(y_j)$ ,  $f'_{\text{н},\min}(y_j)$  и  $f''_{\text{н},\min}(y_j)$  для критерия  $\leq(y_j, c_j)$ , представленные на рисунке 1а линиями 2, 2' и 2''.

Разновидностью функций достижения цели является *плановая функция*  $s(y_j)$ . Область ее значений ограничена сверху двукратным выполнением плана. Относительно невыполнения и перевыполнения

плана область значений плановой функции также может быть сведена к биполярной абсолютной шкале  $[-1, 1]$ .

Значение планового показателя  $s(y_j) \in [0, 100)$  означает *невыполнение* плана, именуемое на языке полезности потерями, а значение планового показателя  $s(y_j) \in (100, 200]$  означает *перевыполнение* плана. Выполнению плана соответствует значение  $s(y_j) = 100\%$ . Особенность плановой шкалы дает возможность индивидуального планирования показателей для каждого оцениваемого объекта  $x \in X$ . Это позволяет сравнивать несравнимые по ресурсам объекты разной природы по плановой функции  $s(y_j)$  в одном  $n$ -мерном пространстве показателей.

Плановое значение  $s(y_j)$   $j$ -го показателя задается на шкале  $[y_{j,\min}, y_{j,\max}]$  (см. рис. 1б). Подобно ОцФ  $f_{n,\max}$ , соответствующей критерию  $\geq(y_j, c_j)$ , плановая функция  $s(y_j)$  представляет собой также кусочно-линейную функцию. Ее составляющие  $s'(y_j)$  и  $s''(y_j)$  вычисляются по формулам (4) и (5). На рисунке 1б они представлены линиями 1' и 1". При трактовке невыполнения плана потерями, которые измеряются на отрицательной полуоси, в формуле (4) нижняя граница  $b$  функции  $s(y_j)$   $b = -1$ . В частном случае при  $c_j = (y_{j,\max} + y_{j,\min})/2$  плановая функция проходит через среднюю точку шкалы  $[y_{j,\min}, y_{j,\max}]$  (см. линию 1 на рис. 1б).

Случай  $c_j \neq (y_{j,\max} + y_{j,\min})/2$  означает изменение предпочтения превосходства в точке  $c_j$  на втором участке шкалы показателя. При увеличении числа участков с разными предпочтениями растет и число кусков кусочно-линейной ОцФ [2]. При устремлении числа участков с разными предпочтениями к бесконечности получаем гладкую нелинейную ОцФ.

Исследуя лотерею как модель предпочтений ЛПР, фон Нейман и Моргенштерн [1] выявили зависимость типа нелинейности ОцФ от склонности/несклонности ЛПР к риску. Стремление ЛПР к максимальному выигрышу при достижении цели моделируется ОцФ, выпуклой вниз (см. линию 1' на рис. 2а). Чем ближе значение показателя к целевому, тем быстрее возрастает значение ОцФ. Стратегия несклонности к риску, то есть «довольствования малым» моделируется ОцФ, выпуклой вверх (см. линию 1" на рис. 2а). Она быстро возрастает в начале шкалы, но по мере приближения к цели скорость возрастания замедляется. В простейшем случае нелинейность таких функций моделируется возведением в соответствующую степень  $k$  линейной нормирующей функции (при  $k > 1$  функция выпукла вниз, а при  $0 < k < 1$  – выпукла вверх).

Отказ от лотереи трактуется авторами [1] как безразличие к риску, то есть отсутствие интереса ЛПР к прибыли и к потерям. Это-

му условию соответствует полезность  $u(c_j) = 0$  в точке  $c_j$  на рисунке 2б. Область значений функции полезности с учетом потерь является bipolarной, то есть  $U_j \rightarrow [-1, 1]$ .

Для уменьшения потерь ЛППР склонен к риску на участке шкалы  $[y_{j,\min}, c_j]$ , устремляясь к точке безразличия  $c_j$  (см. кривую 1' на рис. 2б). Достигнув ее, он может предпочесть несклонность к риску на участке  $(c_j, y_{j,\max}]$  (см. кривую 1'' на рис. 2б).

Учет склонности/несклонности ЛППР к риску можно использовать для перехода от кусочно-линейной функции полезности в кусочную нелинейную функцию (см. рис. 2б), возведением ее в степень  $k > 0$ . Дополнительно для представления нелинейной функции полезности с перегибом в целевой точке  $c_j$  и областью значений  $[-1, 1]$  можно использовать логистическую функцию следующего вида:

$$f_{\text{онл}}(y_j) = m \cdot \left(1 + \exp(-t \cdot (y_j - c_j))\right)^{-1} + b. \quad (6)$$

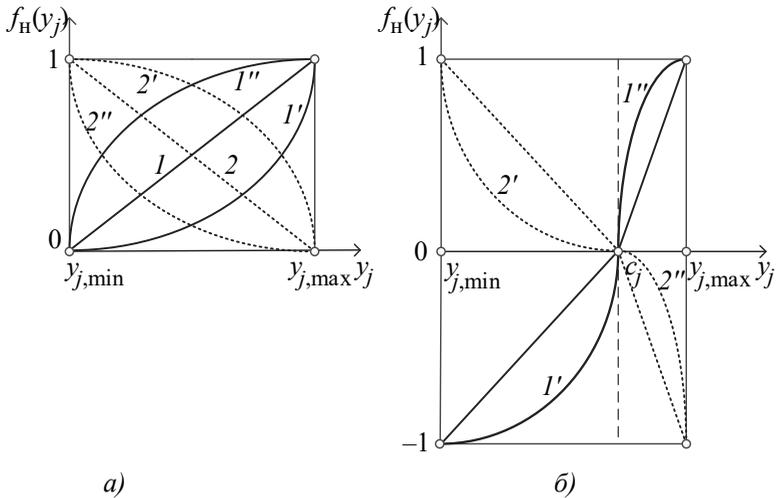


Рис. 2. Нелинейные оценочные функции:  
 а) – функции ценности; б) – функции полезности

Здесь  $c_j$  соответствует целевой точке, параметр  $t$  определяет степень нелинейности функции (6) и ее направление: при  $t > 0$  функция возрастает (соответствует кривой 1'–1'' на рис. 2б), а при  $t < 0$  – убывает (соответствует кривой 2'–2'' на рис. 2б). Масштабирующие параметры  $m$  и  $b$  определяют разброс области значений функции: при  $m = 1$  и

$b=0$  функция имеет своей областью значений отрезок  $[0, 1]$ , стандартный для логистических функций, а при  $m=2$ ,  $b=-1$  отрезок  $[-1, 1]$ , характерный для функций полезности.

Заметим, что переход от задания предпочтений от двух к бесконечному числу интервалов увеличивает информативность ФП при условии соответствия параметров ее нелинейности предпочтениям ЛПП. Один из способов решения проблемы параметризации типовых ФП в многомерной модели предпочтений предложен в [26].

**5. Оценочные функции, реализующие отношение соответствия.** В отношении соответствия за достижение цели следует принять истинность предикатов  $Pr_{=} (y_j, c_j) = 1$  при точечном соответствии и  $Pr_{\subseteq} (y_j, [c_{j,n}, c_{j,v}]) = 1$  – при интервальном соответствии. Применительно к отношению соответствия больше подходит понятие *нормы*. Истинность указанных предикатов означает соответствие значения  $y_j$  норме  $c_j$  ( $y_j = c_j$ ) или  $[c_{j,n}, c_{j,v}]$  ( $y_j \in [c_{j,n}, c_{j,v}]$ ).

Задание нормы через точку  $c_j$  и отрезок  $[c_{j,n}, c_{j,v}]$  на шкале  $j$ -го показателя можно рассматривать как *класс* его допустимых значений, а выявление принадлежности классу – как задачу классификации по принципу «годен/не годен». В этом смысле ОцФ соответствия значения  $y_j$  норме  $c_j$  или отрезку  $[c_{j,n}, c_{j,v}]$  примем за функцию принадлежности  $\mu_n(y_j)$ .

Стопроцентная принадлежность объекта  $x$  норме по  $j$ -му показателю имеет место в точке  $c_j$  для  $Pr_{=} (y_j, c_j) = 1$  и на отрезке  $[c_{j,n}, c_{j,v}]$  для  $Pr_{\subseteq} (y_j, [c_{j,n}, c_{j,v}]) = 1$ . На рисунке 3а точечная принадлежность представлена точкой  $\mu_n(c_j) = 1$ . На рисунке 3б интервальная принадлежность представлена горизонтальной линией  $1^0$ .

Вертикальные штриховые линии выделяют подобласти преобразованного отображения  $\mu_n$ , принадлежащие области его определения  $[y_{j,\min}, y_{j,\max}]$ . Для практических задач представляет интерес не только принадлежность значения показателя норме, но и степень приближения к ней. В простейшем случае примем линейный закон приближения к норме от границ шкалы  $[y_{j,\min}, y_{j,\max}]$ .

Степень приближения к норме от левой границы шкалы до  $\mu_n(c_j) = 1$  вычисляется так же, как и для левой нормирующей функции ограничительного критерия  $f'_{n,\max}(y_j)$  по формуле (1) с  $b=0$ . Степень приближения к норме от правой границы шкалы до  $\mu_n(c_j) = 1$  вычисляется по формуле (7):

$$f''_{n,\max}(y_j) = \frac{y_{j,\max} - y_j}{y_{j,\max} - c_j}. \quad (7)$$

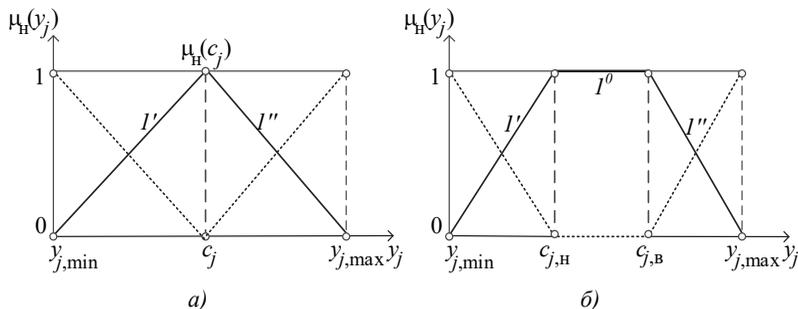


Рис. 3. Немонотонные оценочные функции:  
 а) – треугольная; б) – трапецидальная

На рисунке 3а левый фронт функции принадлежности  $\mu_H(y_j)$  точечного соответствия норме представлен линией  $I'$ , а правый ее фронт – линией  $I''$ . Для вычисления фронтов функции принадлежности  $\mu_H(y_j)$  интервального соответствия в формуле (4)  $c_j$  заменяется на  $c_{j,н}$ , а в формуле (6) – на  $c_{j,в}$ . Полученные функции соответствуют треугольной и трапецидальной функциям принадлежности, применяемым в нечёткой классификации.

В работах [27, 28] утверждается, что «на практике гораздо лучшие результаты получаются, когда используются простейшие функции принадлежности: треугольные и трапециевидные». Тем не менее по аналогии с линейными функциями, реализующими отношение превосходства, для линейных функций соответствия норме также могут быть введены нелинейные аналоги, как показано на рисунке 4.

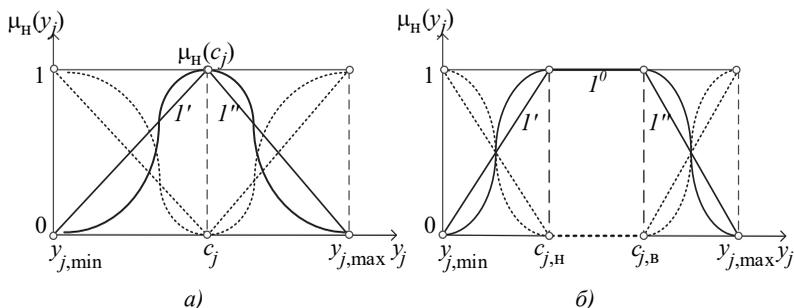


Рис. 4. Немонотонные нелинейные оценочные функции: а) – точечная норма; б) – интервальная норма

Нелинейная ОцФ соответствия точечной норме  $c_j$  задается следующим образом:

$$\mu_n^p(y_j; c_j, t) = \exp\left(-t \cdot (c_j - y_j)^2\right). \quad (8)$$

Здесь параметр  $c_j$ , собственно, значение нормы, а параметр  $t_j > 0$  регулирует степень крутизны функции, то есть скорость ее убывания по мере удаления значения показателя  $y_j$  от нормы  $c_j$ . Это выражение определяет колоколообразную кривую, подобную функции плотности нормального распределения, при этом норма  $c_j$  аналогична математическому ожиданию, а  $t$  – дисперсии случайной величины.

ОцФ соответствия интервальной норме  $[c_{j,n}, c_{j,v}]$ , как и соответствующая функция на рисунке 3, является составной: поскольку интервальная норма представляет собой комбинацию двух ограничений  $y_j \geq c_{j,n}$ , и  $y_j \leq c_{j,v}$  соответственно. Соответствие цели, определяемой ограничением «снизу» или «сверху» может быть описано функцией, которой соответствует логистическая кривая (6) с параметрами при  $m = 1$  и  $b = 0$ . Таким образом, ОцФ соответствия интервальной норме  $[c_{j,n}, c_{j,v}]$ , изображенная кривой на рисунке 4б, представляется следующим выражением:

$$\mu_n^h(y_j; c_{n,j}, t_1, c_{v,j}, t_2) = f_{\text{онл}}^h(y_j; c_{n,j}, t_1) \cdot f_{\text{онл}}^v(y_j; c_{v,j}, t_2). \quad (9)$$

Здесь функции  $f_{\text{онл}}^h$  и  $f_{\text{онл}}^v$  соответствуют левому и правому фронтам, а параметры  $t_1$  и  $t_2$  регулируют отдельно их нелинейность, причем  $t_1 > 0$ , а  $t_2 < 0$ .

По сравнению с линейными и кусочно-линейными ОцФ нелинейные функции обладают такими преимуществами, как большая информативность за счет нелинейности и простота аналитического описания, позволяющая учесть склонность ЛППР к риску на разных участках шкалы относительно заданной цели.

**6. Функция отклонения от цели.** В [4] функция отклонения от цели трактовалась как мера недостижимости цели. В настоящей работе понятие отклонения от цели расширено в направлении превышения цели и применено к функциям ценности/полезности. Согласно принципу «от противного» функции отклонения дополняют функции достижения и соответствия цели до 1. На графиках ОцФ этот принцип выражается через симметрию соответствующих функций. На рисунках 1-4 ОцФ отклонения от цели представлены пунктирными линиями. Отсюда следует, что принцип дополнительности позволяет применять функции отклонения от цели для решения задач упорядочения и классификации объектов наряду с функциями достижения и соответствия цели.

Мера несоответствия цели (норме) отображается функцией отклонения  $d(y_j)$ . Чем дальше в любую сторону от цели  $c_j$  или от границ

нормы  $[c_{j,н}, c_{j,в}]$  отстоит значение  $y_j$   $j$ -го показателя, тем больше значение функции  $d(y_j)$ . Функция  $d(y_j) = 0$  в том случае, когда имеет место либо точное соответствие цели ( $y_j = c_j$ ), либо принадлежность  $y_j$  отрезку  $[c_{j,мин}, c_{j,макс}]$ . Таким образом, объекты, оцениваемые по степени недостижения цели или несоответствия норме, упорядочиваются относительно критерия  $d(y_j) \rightarrow \min$ .

Особый интерес представляет раздельное рассмотрение вариантов недостижимости и превышения цели, свойственного отношению превосходства с реальными целями. В [23] предложено степень недостижимости цели оценивать штрафами, измеряемыми на положительной полуоси графика функции отклонения от цели. Степень превышения цели измеряется областью поощрений на отрицательной полуоси графика. На рисунке 5а изображена кусочно-линейная функция отклонения от целевого значения  $c_j$  для ограничения « $\geq$ » («не менее» или «снизу»), а на рисунке 5б – для ограничения « $\leq$ » («не более» или «сверху»).

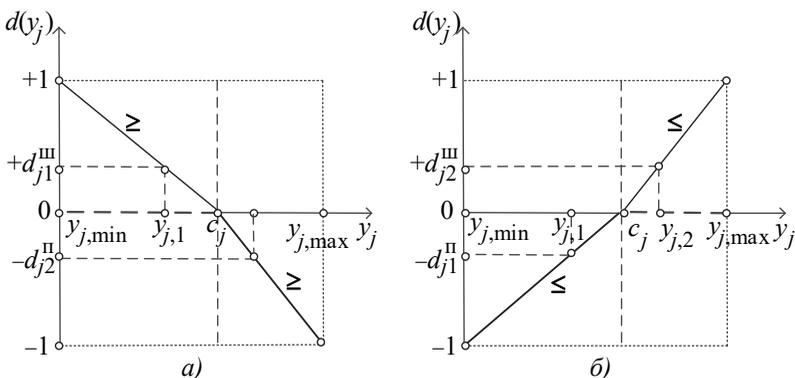


Рис. 5. Функции отклонения: а) – снизу (не менее); б) – сверху (не более)

На рисунке 5а значению  $y_{j1} < c_j$  сопоставляется штраф  $d_{j1}^{III}$ , а на рисунке 5б – поощрение  $-d_{j1}^{III}$ . Аналогичным образом значению  $y_{j2} > c_j$  на рисунке 5а сопоставляется поощрение  $-d_{j2}^{II}$ , а на рисунке 5б – штраф  $d_{j2}^{III}$ .

Относительные отклонения от цели для ограничения «снизу» вычисляется по формуле:

$$d^n(y_j) = \begin{cases} d_{II}^n(y_j) = \frac{c_j - y_j}{y_{j,max} - c_j}, & \text{если } y_j > c_j, \\ d_{III}^n(y_j) = \frac{c_j - y_j}{c_j - y_{j,min}}, & \text{если } y_j < c_j. \end{cases} \quad (10)$$

Верхняя формула в выражении (10), отражающая превышение цели, представляет функцию поощрений  $d_n$  на отрицательной полуоси графика, а нижняя формула – функцию штрафов  $d_m$  на положительной полуоси графика (см. рис. 5а).

Функции поощрений и штрафов для ограничения «сверху» (см. рис. 5б) представлены соответственно в выражении:

$$d^B(y_j) = \begin{cases} d_n^B(y_j) = \frac{y_j - c_j}{c_j - y_{j,\min}}, & \text{если } y_j < c_j, \\ d_m^B(y_j) = \frac{y_j - c_j}{y_{j,\max} - c_j}, & \text{если } y_j > c_j. \end{cases} \quad (11)$$

Как следует из изложенного, функция отклонения  $d(y_j)$  обобщает две функции – штрафов  $d_m(y_j)$  и поощрений  $d_n(y_j)$ , определенных на противоположных полуосях функции  $d(y_j)$ :  $d(y_j) = d_m(y_j) + d_n(y_j)$ , при этом на тех участках шкалы, где  $d_m(y_j) \neq 0$ ,  $d_n(y_j) = 0$  и наоборот. Эта особенность предоставляет дополнительные возможности в оценивании объектов.

В отличие от функций отклонения  $d(y_j)$ , образуемых на основе предикатов превосходства, функции, образуемые на основе предикатов соответствия, симметричны относительно цели (нормы), следовательно, отклонения от нормы в любую сторону оцениваются только функцией штрафов  $d_m(y_j)$ . При линейной трактовке отклонений от нормы на рисунке ба,б изображены треугольная и трапециевидальная функции отклонения от нормы, подобные функциям принадлежности норме (см. рис. 3а,б).

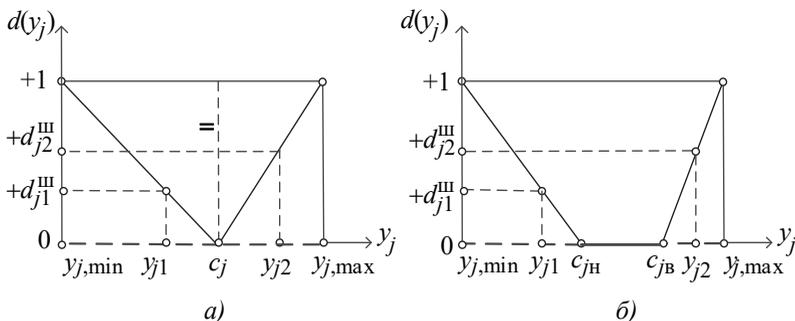


Рис. 6. Функции отклонения: а) – для ограничения «равно»; б) – для ограничения «интервал»

Пунктиром на оси абсцисс показаны участки шкалы, не удовлетворяющие соответствующему ограничению. Значению показателя на этом участке начисляется штраф, пропорциональный величине отклонения.

Для точечного (равно) и интервального ограничений относительное отклонение значения  $y_j = f_j(x)$  объекта  $x \in X$  по  $j$ -му показателю от цели  $c_j$  или отрезка  $[c_{н,j}, c_{в,j}]$  всегда положительно (см. рис. 6а,б).

Относительное отклонение значения  $j$ -го показателя для точечного ограничения определяется выражением (12), состоящим из двух функций штрафов – нижней  $d_{ш,н}$  и верхней  $d_{ш,в}$ :

$$d^p(y_j) = \begin{cases} d_{ш,н}^p(y_j) = \frac{|y_j - c_j|}{c_j - y_{j,\min}}, & \text{если } y_j < c_j, \\ d_{ш,в}^p(y_j) = \frac{|y_j - c_j|}{y_{j,\max} - c_j}, & \text{если } y_j > c_j. \end{cases} \quad (12)$$

При использовании интервального ограничения  $y_j \in [c_{н,j}, c_{в,j}]$  относительное отклонение внутри отрезка равно 0 ( $d^h(y_j) = 0$ ), если все точки внутри него равноценны. При выходе значения  $y_j$  за любую границу отрезка  $[c_{н,j}, c_{в,j}]$  относительное отклонение определяется по формуле (12), где в верхнюю формулу вместо  $c_j$  подставляется нижняя  $c_{н,j}$ , а нижнюю формулу – верхняя граница  $c_{в,j}$  отрезка  $[c_{н,j}, c_{в,j}]$ .

Следует отметить, что функции отклонения от нормы также могут быть преобразованы в нелинейные варианты подобно функциям достижения цели, изображенным на рисунках 2 и 4, дополнением до единицы соответствующих нелинейных функций соответствия.

В [4] требованию максимизации полезности показателя противопоставлялась минимизация расстояния до заданной точки в многомерном пространстве, а аддитивной обобщающей функции — минимаксный критерий для выбора объекта с наименьшим из максимальных отклонений от цели. Отсюда делался вывод о ненужности весовых коэффициентов, используемых в АОФ. Однако АОФ может использоваться и для нахождения обобщенного отклонения от вектора целей в тех случаях, когда необходимо учитывать важность показателей. Результирующее отклонение  $d^*(x)$  объекта  $x \in X$  от цели по всем видам ограничений вычисляется по формуле:

$$d^*(x) = \sum_{j=1}^l w_j \cdot d_j^h(y_j) + \sum_{j=1}^u w_j \cdot d_j^b(y_j) + \sum_{j=1}^s w_j \cdot d_j^p(y_j) + \sum_{j=1}^e w_j \cdot d_j^h(y_j). \quad (13)$$

Здесь  $d_j^\Theta(y_j)$  — функция отклонения, заданная для  $j$ -го показателя в соответствии с критерием, сформулированным для этого показателя,  $\Theta \in \{н, в, р, и\}$ , а  $y_j = f_j(x)$  — значение объекта  $x \in X$  по  $j$ -му показателю. Верхние индексы сумм соответствуют количеству показателей с заданным типом ограничения, причем  $l + u + s + e = n$ , а  $n$  — общее количество показателей. Веса показателей удовлетворяют условию нормировки:  $\sum_{j=1}^k w_j = 1$ .

Для оценивания суммарного штрафа  $d_{ш}^*(x)$  для объекта  $x \in X$  в сомножителях первых двух сумм обобщающей функции отклонения (10) используются только соответствующие функции штрафов из формул (10-12), а для оценивания суммарной премии  $d_{п}^*(x)$  — соответствующие функции штрафов из формул (7, 8). При этом используются только 2 первых члена выражения (13), так как только отклонения для ограничений «снизу» и «сверху» могут иметь ненулевые значения функции поощрения  $d_{п}(y_j)$ :

$$d_{п}^*(x_i) = \sum_{j=1}^l w_j \cdot d_{п,j}^H(x_i) + \sum_{j=1}^u w_j \cdot d_{п,j}^B(x_i). \quad (14)$$

Требованию минимизации отклонения от цели соответствует целевая функция:  $d^*(x) \rightarrow \min$ .

Рассмотрим типовые задачи упорядочения по функциям отклонения.

1. *Соответствие норме.* Объект  $x$  удовлетворяет всем ограничениям, если  $d_{ш}^*(x) = 0$ , что соответствует решению задачи *отбора* или положительному результату *контроля*.

2. *Отклонение от цели.*

а. *Средневзвешенное отклонение.* Вычисляется по формуле (13) с сомножителями в первых двух суммах, вычисляемых по формулам (10) и (11) — и по штрафам, и по премиям (алгебраические суммы первых двух членов формулы).

б. *Лучший из худших.* Отклонение вычисляется по формуле:

$$d_{\text{ММХ}}^*(x_i) = \min_{i=1, N} \max_{j=1, n} (d_j^\Theta(y_j)). \quad (15)$$

3. *Упорядочение*

а. *По штрафам.* *Общий штраф* вычисляется по формуле (13) с сомножителями в первых двух суммах, в которых используются только штрафы из формул (10-12) — отклонение *только по штра-*

фам,  $d^*_{ин}(x) \rightarrow \min$ .

б. По премиям. Общая премия вычисляется по формуле (14) с сомножителями в первых двух суммах, вычисляемых по премиям из формул (10, 11) – отклонение *только по премиям*,  $d^*_{ин}(x) \rightarrow \min$ .

Упорядочение по штрафам и премиям можно использовать для реализации метода ДИЦ, если цель  $c_j$  ставить на границах шкал ( $c_j = y_{j,\min}$  для  $y_j \geq c_j$ ) и ( $c_j = y_{j,\max}$  для  $y_j \leq c_j$ ) для штрафов и на противоположных границах – для премий.

**7. Методы функционального выбора.** Применим рассмотренные выше свойства ОцФ для систематизации использующих их методов. По отношению к форме задания предпочтения ЛПР разделим ОцФ на две группы: по заданию критерия или функции полезности. Первая группа характеризуется заданием целевого значения показателя – на границах или внутри его шкалы. По отношению к цели разделим первую группу на две подгруппы: по функции достижения цели (ДЦ) и функции отклонения от цели (ОЦ). В первую подгруппу входят методы (см. рис. 7):

1. Метод достижения идеальной цели (ДИЦ).
2. Метод достижения реальной цели (ДРЦ).
3. Плановый метод.

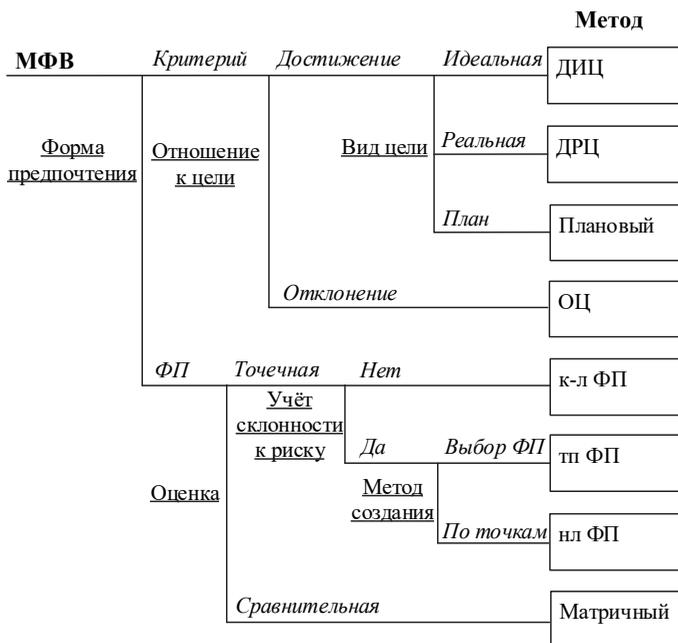


Рис. 7. Классификация методов ФВ

Модели предпочтений первых двух методов формируются путем задания целевых и ограничительных критериев соответственно. Плановый метод отличается от метода ДРЦ отображением значений показателей на шкалу  $[0\%, 200\%]$  с  $s_j(c_{ji}) = 100\%$  и заданием индивидуального планового значения  $c_{ji}$  по  $j$ -му показателю для каждого  $i$ -го оцениваемого объекта,  $i = \overline{1, N}$ .

ОцФ, применяемые в методах достижения цели, подлежат максимизации. Значения обратных по форме функций отклонения от цели, наоборот, подлежат минимизации. По форме и знаку они дополнителины по отношению к функциям ДЦ и образуют *семейство методов*, использующих различные варианты функции отклонения от цели. Целевые значения показателей задаются в методах ДИЦ, ДРЦ и плановом методе. В методе ДИЦ целевое значение совпадает с одной из границ шкалы показателя. По отношению к цели методы достижения цели и отклонения от цели решают противоположные задачи и, следовательно, дополнителины друг к другу.

Вторую группу методов образуют методы теории многомерной полезности:

1. Метод с кусочно-линейной функцией полезности (к-л НП).
2. Метод с типовыми функциями полезности (тп ФП).
3. Метод с нелинейной функцией полезности (нл ФП).

Они различаются способами создания функций полезности. Функции полезности в методах 2 и 3 строятся с учетом склонности/несклонности ЛПР к риску. Модель предпочтений метода 2 проще модели метода 3 по построению, поскольку создание ФП показателя заменяется выбором типовой функции полезности. Однако усредненные значения нелинейности типовых функций требуют уточнений, выполняемых в процессе отладки [9].

Характеристики моделей предпочтений, используемых перечисленными методами, представлены в таблице 1.

Принцип дополнителиности отклонения от полезности имеет место и для методов многомерной теории полезности по отношению ко всем точкам функций полезности. В расширенной трактовке полезности к методам, использующим функцию полезности, относится матричный метод [3]. Он использует для вычисления дискретной функции приоритетов матрицу парных сравнений. К значениям, вычисляемым на основе МПС, по умолчанию предъявляется требование максимизации. Эти значения используются и как весовые коэффициенты обобщающей функции, и как полезность сопоставляемых объектов.

Таблица 1. Характеристики моделей предпочтений рейтинговых методов

№ п/п	Число интервалов	Цель	Склонность к риску	Функция полезности	Метод оценивания
1	1	ИЦ	–	линейная	ДИЦ
2	2	РЦ	–	к-л с целью	ДРЦ
3	2	план	–	плановая	Плановый
4	>2	max	–	к-л ФП	к-л ФП
5	≥1	max	+	типовая	тп ФП
6	>2	max	+	нелинейная	нл ФП

Все методы, представленные на рисунке 7, были реализованы в системе выбора и ранжирования СВБРЬ, разработанной в Петербургском государственном университете путей сообщения [29]. Эта система допускает применение совместного задания предпочтений как в форме критериев, так и в форме функций полезности, получаемых различными способами. Пример комплексного применения методов при решении задачи выбора жилья был продемонстрирован в [30].

**8. Методика обоснования вариантов использования методов ФВ.** На выбор метода оценивания объектов влияет размерность решаемой задачи, пропорциональная числу объектов  $N$  и числу характеризующих их показателей  $n$ . Для определения границы ее сложности примем  $N=n=7$ . Это число характеризует усредненную способность человека к умозрительному решению задач. Согласно формуле (1) трудоемкость решения такой задачи составит  $Q=147$  ( $\approx 150$ ). Таким образом, при превышении этой величины целесообразно для оценивания объектов применять рейтинговые методы. Из рейтинговых методов в настоящей работе выделены методы функционального выбора, ориентированные на упорядочение объектов, характеризующихся неоднородными показателями.

Помимо размерности задачи на выбор метода ФВ влияет компетенция ЛПР в оцениваемой предметной области. Она проявляется в его способности задать целевое значение показателя и определить полезность на шкалах показателей на основе склонности/несклонности к риску [1]. Из этих соображений может быть предложен следующий подход к выбору метода ФВ для решения задачи упорядочения объектов высокой размерности.

1. Если ЛПР не обладает информацией о целевых значениях показателей или они не имеют значения в решаемой задаче, выбирается метод ДИЦ. Этот метод отражает «жадный» подход, заключающийся в желании достичь граничных значений всех показателей.

2. Если ЛПР может задать целевые значения показателей и их приемлемую полезность (от  $>0,5$  до  $1,0$ ), выбирается метод ДРЦ. Этот

метод отражает рациональный подход к оцениванию объектов, заключающийся в формулировании разумных требований к показателям.

3. В случае несравнимости оцениваемых объектов по ресурсным показателям применяется плановый метод, как отображение метода ДРЦ на процентную шкалу  $[0\%, 100\%]$  с  $s_j(c_{ji}) = 100\%$ . Поскольку необходимо задать целевое значение  $c_{ji}$  для  $i$ -го оцениваемого объекта  $i = \overline{1, N}$ , трудоёмкость построения иерархической модели этого метода оценивается по формуле (3).

4. Если ЛПР способен задать приемлемую полезность для нескольких промежуточных точек на шкале показателя, выбирается метод с кусочно-линейной функцией полезности (к-л ФП).

5. Если ЛПР способен определить свою склонность/несклонность к риску на двух интервалах шкалы  $j$ -го показателя по обе стороны от целевого значения  $c_j$ , он выбирает одну из типовых функций метода тп ФП.

6. Если ЛПР способен определить свою склонность/несклонность к риску на нескольких интервалах шкалы показателя, выбирается метод с нелинейными ФП (нл ФП).

Из семейства методов отклонения от цели практический интерес представляют методы с явно выраженной целью, дополнительные по отношению к методам ДИЦ, ДРЦ и плановому методу.

При наличии интервальных целей  $[c_{j,н}, c_{j,в}]$ ,  $j = \overline{1, n}$ , принадлежность фактических значений показателей заданным интервалам позволяет выполнить параметрический контроль оцениваемых объектов по принципу «годен/не годен». Практический интерес представляет также упорядочение объектов по невыполнению и перевыполнению заданного плана (цели). Для этой цели используются полуоси функции отклонений по штрафам и отклонениям. Эта задача имеет многочисленные применения, такие, как оценка результатов клинических анализов (например – крови) в медицине, параметрический контроль в технике, непрерывное наблюдение (мониторинг) за объектами любой природы и пр.

Подытоживая рекомендации, следует подчеркнуть, что в основу рекомендаций по применению рейтинговых методов положены компетенции ЛПР в оцениваемой предметной области. Более того, компетенции ЛПР могут быть различными в отношении разных свойств оцениваемого объекта. А это означает, что в рамках одной модели ММО ЛПР может применять различные оценочные функции для разных показателей в зависимости от степени исследования оцениваемого свойства объекта. В целом, чем более глубоким является знание предметной области и чем больше времени имеется для создания модели оценивания, тем большее доверие должны вызывать результаты оценивания. Другим

способом повышения доверия к результатам является комплексное применение рейтинговых методов для решения поставленной задачи, что позволяет сравнивать результаты, полученные разными методами.

**9. Пример.** Рассмотрим следующий пример использования функций отклонения для задач параметрического контроля, решаемых в медицине. Одной из задач такого типа является задача проверки состояния здоровья пациента на основе результатов различных лабораторных исследований, например, клинического или биохимического анализа крови. Соответствие состояния пациента клинической норме задается набором референсных значений измеряемых показателей.

Под референсным значением показателя понимается медицинский термин, употребляемый при проведении и оценке лабораторных исследований, который определяется на основе *среднего* значения определенного лабораторного показателя, полученного в результате массовых обследований здорового населения.

Клинический анализ крови позволяет оценить содержание гемоглобина в системе красной крови, количество эритроцитов, цветовой показатель, количество лейкоцитов и тромбоцитов, рассмотреть лейкограмму и измерить скорость оседания эритроцитов (СОЭ). С помощью данного анализа можно выявить анемию, воспалительные процессы и так далее.

В упрощенном примере оценим отклонение состояния здоровья от клинической нормы для пяти условных пациентов по таким основным показателям, как количественное содержание кровяных телец в венозной крови, а также содержанию гемоглобина и СОЭ. Исходные данные приведены в таблице 2.

Таблица 2. Результаты клинического анализа крови пяти пациентов

Пациенты	Лейкоциты, $10^9$ кл/л	Эритроциты, $10^{12}$ кл/л	Гемоглобин, г/л	Гематокрит, %	Тромбоциты, $10^9$ кл/л	Лимфоциты, %	СОЭ, мм/ч
Пациент 1	3,9	4,4	100	40	160	35	9,0
Пациент 2	4,2	5,0	36	42	152	26	7,0
Пациент 3	7,7	2,7	18	50	390	41	6,5
Пациент 4	9,4	5,3	120	27	100	30	8,0
Пациент 5	6,6	7,2	143	41	400	25	8,3
Референсные значения показателей крови							
	[4, 9]	[4,3, 5,5]	[20, 140]	[39, 49]	[150, 400]	[25, 40]	[0, 8]
Шкалы показателей							
	[0, 10]	[0, 10]	[0, 200]	[0, 100]	[0, 500]	[0, 100]	[0, 50]

В нижней части таблицы приведены референсные значения, соответствующие нормальному (здоровому) состоянию пациента, и шкалы показателей, указывающие на теоретически возможный разброс значений каждого из показателей.

Учитывая разные шкалы и единицы измерения показателей крови, вычислим относительные отклонения от нормы. С учетом того, что все требования заданы только интервальными нормами, для вычисления отклонений будем использовать функцию, представленную на рисунке 6б, и формулу (12) для нижней и верхней границ нормы. Отклонение от нормы  $d(y_j) = 0$  отсутствует при соответствии норме  $[c_{j,н}, c_{j,в}]$ , заданной соответствующими референсными значениями.

*Эксперимент 1.* Оценим состояние здоровья пациентов по всем показателям, применив к частным отклонениям от нормы обобщающую функцию (15). Максимальное отклонение от нормы у каждого пациента помечено жирным шрифтом в таблице 3. Они дублированы в столбце  $d^*_{mm}(x)$ .

Таблица 3. Ранжирование пациентов по минимаксному отклонению от нормы

	$d(y_1)$	$d(y_2)$	$d(y_3)$	$d(y_4)$	$d(y_5)$	$d(y_6)$	$d(y_7)$	$d^*_{mm}(x)$	Ранг
Пациент 1	0,03	0,00	0,00	0,00	0,00	0,00	0,02	<b>0,03</b>	2
Пациент 2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	<b>0,00</b>	1
Пациент 3	0,00	0,37	0,10	0,02	0,00	0,02	0,00	<b>0,37</b>	3
Пациент 4	0,40	0,00	0,00	0,31	0,33	0,00	0,00	<b>0,40</b>	5
Пациент 5	0,00	0,38	0,05	0,00	0,00	0,00	0,01	<b>0,38</b>	4

В последнем столбце таблицы 3 пациенты упорядочены в направлении увеличения максимальных отклонений от нормы. Соответствие норме по всем показателям обнаружено только у пациента 2. Достоинством применения обобщающей функции (15) является акцентирование внимания ЛПП на наихудшем показателе у каждого объекта.

*Эксперимент 2.* Оценим состояние здоровья пациентов по всем показателям, применив к частным отклонениям от нормы обобщающую функцию (13). С учетом того, что все требования заданы только интервальными нормами, в этой формуле используется только последняя сумма частных отклонений от нормы. В отсутствие экспертной информации о важности показателей для оценивания состояния здоровья пациентов, весовые коэффициенты всех показателей в формуле (13) приняты равными:  $w_j = 1/7, j = 1, \dots, 7$ . Результаты оценивания приведены в таблице 4. Результаты упорядочения пациентов по минимаксной и усредняющей обобщающей функциям различаются в местах пациентов 3 и 5.

Таблица 4. Ранжирование пациентов по усредненному отклонению от нормы

	$d(y_1)$	$d(y_2)$	$d(y_3)$	$d(y_4)$	$d(y_5)$	$d(y_6)$	$d(y_7)$	$d^*(x)$	Ранг
Пациент 1	0,03	0,00	0,00	0,00	0,00	0,00	0,02	0,007	2
Пациент 2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,000	1
Пациент 3	0,00	0,37	0,10	0,02	0,00	0,02	0,00	0,073	4
Пациент 4	0,40	0,00	0,00	0,31	0,33	0,00	0,00	0,149	5
Пациент 5	0,00	0,38	0,05	0,00	0,00	0,00	0,01	0,062	3

Небольшое различие в полученных рейтингах пациентов позволяет в принципе согласиться с утверждением автора [4] о приемлемости замены обобщающих функций с весовыми коэффициентами минимаксной обобщающей функцией. Однако исключение из рассмотрения показателей, у которых отклонения от нормы не максимальны, влияет на обобщенную оценку их здоровья. Это тем более следует учитывать при диагностике болезни, зависящей от разной важности показателей.

*Эксперимент 3.* Упорядочим пациентов по степени сопротивляемости организма инфекциям. Из принятых для экспериментов показателей для определения иммунитета организма к инфекциям врач обращает внимание прежде всего на соответствие нормам лейкоцитов и лимфоцитов. Назначим весовые коэффициенты для них соответственно  $w_1=0,5$  и  $w_6=0,4$ , оставив весовые коэффициенты прочих факторов равными,  $w_j=0,02$ .

Результаты упорядочения пациентов по состоянию их иммунитета с применением средневзвешенных оценок, вычисленных по формуле (13), приведено в таблице 5.

При указанных условиях пациент 5 переместился с третьего на второе место, а пациент 1 – со второго на третье в силу того, что у пациента 1 большее отклонение от нормы по наиболее важному фактору «уровень лейкоцитов», а у пациента 5, несмотря на наличие отклонений от нормы по менее важным факторам, наблюдается нулевое отклонение от нормы по обоим наиболее важным факторам.

Таблица 5. Ранжирование пациентов по средневзвешенному отклонению от нормы

	$d(y_1)$	$d(y_2)$	$d(y_3)$	$d(y_4)$	$d(y_5)$	$d(y_6)$	$d(y_7)$	$d^*(x)$	Ранг
Пациент 1	0,03	0,00	0,00	0,00	0,00	0,00	0,02	0,013	3
Пациент 2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,000	1
Пациент 3	0,00	0,37	0,10	0,02	0,00	0,02	0,00	0,017	4
Пациент 4	0,40	0,00	0,00	0,31	0,33	0,00	0,00	0,213	5
Пациент 5	0,00	0,38	0,05	0,00	0,00	0,00	0,01	0,009	2

Рассмотрение этого примера показывает широкие возможности практического применения методов отклонения от цели в отличие от утверждений автора работы [4]. Во-первых, методы отклонения от цели не ограничиваются применением отношения превосходства. Во-вторых, минимаксная обобщенная оценка не исключает, а дополняет усредняющую и средневзвешенную обобщающие функции.

Таким образом, применение методов отклонения от цели расширяет диапазон задач, решаемых с применением методов многомерного оценивания объектов любой природы и назначения. При трактовке объектов временными интервалами с применением методов отклонения решается, в частности, задача идентификации аритмии.

**10. Заключение.** В основу систематизации методов многомерного оценивания положено понятие *цели*, присущее как формальной модели критерия, так и оценивающим функциям показателей. Из многообразия методов ММО для систематизации выделены рейтинговые методы, использующие для упорядочения объектов частные и обобщенные ОцФ показателей.

Показано соотношение методов достижения цели и отклонения от цели для отношений превосходства и соответствия. Отмечена особенность функций отклонения от реальной цели с областью значений  $[-1, 1]$ , которая позволяет выполнять как совместное, так и раздельное упорядочение объектов по штрафам и поощрениям.

Сходство ОцФ по областям определения и значений дало основание для систематизации рейтинговых методов ММО относительно разновидностей и способов задания целевых значений показателей. Линейная нормирующая функция критерия признана за частный случай нелинейной функции ценности, а функция ценности — за частный случай функции полезности. Показано влияние степени склонности/не склонности ЛПР к риску на нелинейность ОцФ. Предложены общие рекомендации по выбору рейтинговых методов относительно трудоемкости создания модели ММО. На конкретном примере показана правомерность применения наряду с минимаксным способом обобщения частных функций отклонения от цели усредняющей и средневзвешенной обобщающих функций. Показано, что варьирование весовыми коэффициентами средневзвешенной обобщающей функции расширяет спектр задач, решаемых с применением функций отклонения от цели.

Авторы выражают благодарность д.т.н., профессору Б. В. Соколову за поддержку исследований и полезные советы, данные при подготовке работы к публикации.

### Литература

1. *Neumann J.V., Morgenstern O.* Theory of Games and Economic Behavior // Princeton University Press. 1953. 586 p.
2. *Keeney R.L., Raiffa H.* Decisions with Multiple Objectives: Preferences and Value Tradeoffs // Wiley. 1976. 452 p.
3. *Saaty T.L.* The Analytic Hierarchy Process: Planning, Priority Setting, Resources Allocation // McGraw-Hill. 1980. 586 p.

4. *Wierzbicki A.P.* The Use of Reference Objectives in Multiobjective Optimization // Multiple Criteria Decision Making Theory and Application. Lecture Notes in Economics and Mathematical Systems. 1980. vol. 177. pp. 468–486.
5. *Hwang S.L., Yoon K.* Multiple Attributes Decision Making Methods and Applications // Berlin Heidelberg. 1981. 269 p.
6. *Семенов С.С.* Оценка качества и технического уровня сложных систем: практика применения метода экспертных оценок // М.: Ленанд. 2015. 350 с.
7. *Abastante F. et al.* Choice architecture for architecture choices: evaluating social housing initiatives putting together a parsimonious AHP methodology and the Choquet integral // Land Use Policy. 2018. vol. 78. pp. 748–762.
8. *De Boni A., Roma R., Ottomano Palmisano G.* Fishery Policy in the European Union: A Multiple Criteria approach for assessing sustainable management of Coastal Development Plans in Southern Italy // Ocean and Coastal Management. 2018. vol. 163. pp. 11–21.
9. *Greco S., Ishizaka A., Matarazzo B., Torrisi G.* Stochastic multiattribute acceptability analysis: an application to the ranking of Italian regions // Regional Studies. 2018. vol. 52.n. 4. pp. 585–600.
10. *Бураков Н.А., Бухвальд Е.М., Кольчугина А.В.* Ранжирование субъектов российской федерации на основе регионального индекса экономического развития // Федерализм. 2019. № 3. С. 149–171.
11. *Ogryszak W. et al.* Large-scale periodic routing problems for supporting planning of mobile personnel tasks // Advances in Intelligent Systems and Computing. 2018. vol. 559. pp. 205–216.
12. *Vergara-Solana F., Araneda M., Ponce-Díaz G.* Opportunities for strengthening aquaculture industry through multicriteria decision-making // Reviews in Aquaculture. 2019. vol. 11. no. 1. pp. 105–118.
13. *Хабарова Д.С.* Обзор программных комплексов многокритериальной оптимизации // Прикладная информатика. 2013. № 2(44). С. 102–112.
14. *Velasquez M., Hester P.T.* An Analysis of Multi-Criteria Decision Making Methods // International Journal of Operations Research. 2013. vol. 10. no. 2. pp. 56–66.
15. *Микони С.В., Соколов Б.В., Юсутов Р.М.* Квалиметрия моделей и полимодельных комплексов // М.: РАН. 2018. 314 с.
16. *Saaty T.L.* The analytic hierarchy and analytic network measurement processes: Applications to decisions under Risk // European Journal of Pure and Applied Mathematics. 2008. vol. 1. no. 1. pp. 122–196.
17. *Roy B.* Classement et choix en présence de points de vue multiples (la méthode ELECTRE) // La Revue d'Informatique et de Recherche Opérationnelle (RIRO). 1968. vol. 8. pp. 57–75.
18. *Brans J.P., Vincke P.* A preference ranking organisation method: The PROMETHEE method for MCDM // Management Science. 1985. vol. 31. no. 6. pp. 647–656.
19. *Ларичев О.И.* Вербальный анализ решений // М.: Наука. 2006. 181 с.
20. *Шакиров В.А., Панкратьев П.С.* Методика многокритериального двухуровневого анализа пунктов размещения электростанций // Искусственный интеллект и принятие решений. 2017. № 1. С. 69–83.
21. *Микони С.В.* Теория принятия управленческих решений // СПб.: Лань. 2015. 448 с.
22. *Bordley R., LiCalzi M.* Decision Analysis with Targets instead of Utilities // Decisions in Economics and Finance. 2000. vol. 23. no. 1. pp. 53–74.
23. *Mikoni S.V.* Method of choice by approximation to a pattern // Proceedings of the 4th International Conference NITE'2000. 2000. vol. 1. pp. 156–159.
24. *Перегудов Ф.И., Тарасенко Ф.П.* Основы системного анализа: Учеб. 2-е изд., доп. // НТЛ. 1997. 396 с.

25. *Микони С. В.* Аксиоматика методов многокритериальной оптимизации на конечном множестве альтернатив. // Труды СПИИРАН. 2016. Вып. 44. С. 198–214.
26. *Микони С. В., Бураков Д. П.* Отладка типовых одномерных функций полезности в модели многомерной полезности // Известия Петербургского университета путей сообщения. 2019. Т. 16(2). С. 131–144.
27. *Kosheleva O., Kreinovich V., Shahbazova S.* Type-2 Fuzzy Analysis Explains Ubiquity of Triangular and Trapezoid Membership Functions // Recent Developments and the New Direction in Soft-Computing Foundations and Applications. Studies in Fuzziness and Soft Computing. 2018. vol. 393. pp. 63–75.
28. *Gholamy A., Kosheleva O., Kreinovich V.* How to explain the efficiency of triangular and trapezoid membership functions in applications to design // Онтология проектирования. 2019. Т. 9. № 2(32). С. 253–260.
29. Сайт научной школы «Многокритериальный выбор на конечном множестве альтернатив». URL: <http://mcd-svir.ru/> (дата обращения: 26.08.2020).
30. *Mikoni S. V.* Application of the Universal Decision Support System SVIR to Solving Urban Problems // Digital Transformation and Global Society. DTGS 2016. Communications in Computer and Information Science. 2016. vol. 674. 016. pp. 1–14.

**Микони Станислав Витальевич** – д-р техн. наук, профессор, ведущий научный сотрудник, лаборатория информационных технологий в системном анализе и моделировании, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: системный анализ, принятие решений, интеллектуальные технологии. Число научных публикаций – 320. [smikoni@mail.ru](mailto:smikoni@mail.ru); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7 (812) 328-01-03.

**Бураков Дмитрий Петрович** – канд. техн. наук, старший научный сотрудник, лаборатория информационных технологий в системном анализе и моделировании, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: системный анализ, теория принятия решений. Число научных публикаций – 35. [burakovdmityr8@gmail.com](mailto:burakovdmityr8@gmail.com); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7 (812) 328-0103.

**Поддержка исследований.** Работа выполнена при финансовой поддержке РФФИ (проекты № 19-08-00989-а, № 20-08-01046) в рамках бюджетной темы № 0073–2019–0004.

S. MIKONI, D. BURAKOV  
**JUSTIFICATION AND CLASSIFICATION OF EVALUATION  
FUNCTIONS USED IN RATING METHODS OF MULTI-CRITERIA  
CHOICE**

---

*Mikoni S., Burakov D. Justification and Classification of Evaluation Functions used in Rating Methods of Multi-criteria Choice.*

**Abstract.** The recommendations on the application of methods of multidimensional estimation (MDE) of objects, proposed in the paper Velasquez M., Hester P.T. «An Analysis of Multi-Criteria Decision Making Methods», are analyzed. The weak substantiation of these recommendations, resulting from the superficial systematization of MDE methods, is noted. The recommendations are focused not on the classes of MDE methods, but on various areas of activity. However, in each area of activity there is a wide range of tasks for evaluating objects of various nature. In this regard, the urgency of a more thorough systematization of MDE methods is recognized.

Taking into account the diversity of MDE methods, it was decided to limit ourselves to the systematization of methods that use evaluation functions (EF), and on this basis to offer general recommendations for their application.

The review of MDE methods from a unified position required clarification of the terminology used in them. On the basis of the formal model of the criterion, the relationship between the concepts of "preference", "criterion" and "indicator" is established. To highlight the methods that use evaluation functions, the concept of the target value of the indicator is introduced. Regarding its location on the indicator scale, the concepts of ideal and real goals are introduced. The criteria corresponding to these goals are divided into target and restrictive ones. Using the proposed terminology, a review of the most well-known MDE methods was carried out. Of these, a group of methods using evaluation functions is distinguished.

Variants of evaluation functions created on the basis of the criterion and postulates of the theory of value and utility are considered. On the basis of the similarity of the domains of definition and the meanings of EFs, the relationship between them is established. Regarding the target value of the indicator, they are divided into the functions of achieving the goal and functions of deviation from the goal. The mutual complementarity of these functions is shown. A group of functions of deviation from the goal is highlighted, which allows us to order objects separately according to penalties and rewards in relation to achieving a real goal. The concept of norm is introduced for the correspondence relation. On the example of medical analyzes, the practical application of deviation functions from the norm is shown using both the minimax and the weighted average generalizing function to establish a rating on a set of objects.

The similarities and differences of the EFs revealed in the course of the study form the basis for the classification of the MDE methods that use them. The difference in EFs in terms of the complexity of creation is reflected in the proposed methodology for their application.

**Keywords:** Preference, Indicator, Criterion, Target Value, Evaluation Function, Value Function, Utility Function, Goal Achievement, Deviation from the Goal, Functional Choice, Multidimensional Evaluation of Objects, Rating Method

---

**Mikoni Stanislav** – Ph.D., Dr.Sci., Professor, Leading Researcher, Laboratory of Information Technologies in the System Analysis and Modeling, St.-Petersburg Federal Research Center of the Russian Academy of Sciences (SPb FRC RAS). Research interests: system analyses, decision making, intellect technologies. The number of publications – 320. smikoni@mail.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7 (812) 328-01-03.

**Burakov Dmitry** – Ph.D., Senior Researcher, Laboratory of Information Technologies in System Analysis and Modeling, St.-Petersburg Federal Research Center of the Russian Academy of Sciences (SPb FRC RAS). Research interests: system analysis, decision-making theory. The number of publications – 35. burakovdmitry8@gmail.com; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7 (812) 328-0103.

**Acknowledgements.** This research is supported by RFBR (grants No 19–08–00989–a, No 20–08–01046) within the budgetary theme No 0073–2019–0004.

## References

1. Neumann J.V., Morgenstern O. *Theory of Games and Economic Behavior*. Princeton University Press. 1953. 586 p.
2. Keeney R.L., Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Wiley. 1976. 452 p.
3. Saaty T.L. *The Analytic Hierarchy Process: Planning, Priority Setting, Re-sources Allocation*. McGraw-Hill. 1980. 586 p.
4. Wierzbicki A.P. *The Use of Reference Objectives in Multiobjective Optimization. Multiple Criteria Decision Making Theory and Application. Lecture Notes in Economics and Mathematical Systems*. 1980. vol. 177. pp. 468–486.
5. Hwang S.L., Yoon K. *Multiple Attributes Decision Making Methods and Applications*. Berlin Heidelberg. 1981. 269 p.
6. Semenov S.S. *Oценка kachestva i tekhnicheskogo urovnya slozhnykh sistem: praktika primeneniya metoda ekspertnykh ocnok* [Assessment of the quality and technical level of complex systems: the practice of applying the method of expert assessments]. M.: Lenand. 2015. 350 p. (In Russ.).
7. Abastante F. et al. Choice architecture for architecture choices: evaluating social housing initiatives putting together a parsimonious AHP methodology and the Choquet integral. *Land Use Policy*. 2018. vol. 78. pp. 748–762.
8. De Boni A., Roma R., Ottomano Palmisano G. Fishery Policy in the European Union: A Multiple Criteria approach for assessing sustainable management of Coastal Development Plans in Southern Italy. *Ocean and Coastal Management*. 2018. vol. 163. pp. 11–21.
9. Greco S., Ishizaka A., Matarazzo B., Torrisi G. Stochastic multiattribute accept-ability analysis: an application to the ranking of Italian regions. *Regional Studies*. 2018. vol. 52. no. 4. pp. 585–600.
10. Burakov N.A., Buhval'd E.M., Kol'chugina A.V. [Ranking of the subjects of the Russian Federation based on the regional index of economic development]. *Federalizm – Federalism*. 2019. vol. 3. pp. 149–171. (In Russ.).
11. Ogryszak W. et al. Large-scale periodic routing problems for supporting planning of mobile personnel tasks. *Advances in Intelligent Systems and Computing*. 2018. vol. 559. pp. 205–216.
12. Vergara-Solana F., Araneda M., Ponce-Diaz G. Opportunities for strengthening aquaculture industry through multicriteria decision-making. *Reviews in Aquaculture*. 2019. vol. 11. no. 1. pp. 105–118.
13. Habarova D.S. [Review of software packages for multicriteria optimization]. *Prikladnaya informatika – Applied Informatics*. 2013. vol. 2(44). pp. 102–112. (In Russ.).
14. Velasquez M., Hester P.T. An Analysis of Multi-Criteria Decision Making Methods. *International Journal of Operations Research*. 2013. vol. 10. no. 2. pp. 56–66.
15. Mikoni S.V., Sokolov B.V., Yusupov R.M. *Kvalimetriya modelej i polimodel'nykh kompleksov* [Qualimetry of models and polymodel complexes]. M.: RAN. 2018. 314 p. (In Russ.).
16. Saaty T.L. The analytic hierarchy and analytic network measurement processes: Ap-

- plications to decisions under Risk. *European Journal of Pure and Applied Mathematics*. 2008. vol. 1. no. 1. pp. 122–196.
17. Roy B. Classement et choix en présence de points de vue multiples (la méthode ELECTRE). *La Revue d'Informatique et de Recherche Opérationnelle (RIRO)*. 1968. vol. 8. pp. 57–75.
  18. Brans J.P., Vincke P. A preference ranking organisation method: The PROMETHEE method for MCDM. *Management Science*. 1985. vol. 31. no. 6. pp. 647–656.
  19. Larichev O.I. *Verbal'nyj analiz reshenij* [Verbal decision analysis]. M.: Nauka. 2006. 181 p. (In Russ.).
  20. Shakirov V.A., Pankrat'ev P.S. [Technique of multi-attribute two-level analysis of power plant sites]. *Iskusstvennyj intellekt i prinyatie reshenij – Artificial intelligence and decision making*. 2017. vol. 1. pp. 69–83. (In Russ.).
  21. Mikoni S.V. *Teoriya prinyatiya upravlencheskih reshenij* [Management decision making theory]. SPb.: Lan'. 2015. 448 p. (In Russ.).
  22. Bordley R., LiCalzi M. Decision Analysis with Targets instead of Utilities. *Decisions in Economics and Finance*. 2000. vol. 23. no. 1. pp. 53–74.
  23. Mikoni S.V. Method of choice by approximation to a pattern. Proceedings of the 4th International Conference NITE'2000. 2000. vol. 1. pp. 156–159.
  24. Peregodov F.I., Tarasenko F.P. *Osnovy sistemnogo analiza: Ucheb. 2-e izd., dop.* [Fundamentals of Systems Analysis: A Textbook. 2nd edition, revised.]. NTL. 1997. 396 p. (In Russ.).
  25. Mikoni S.V. [Axioms of multicriteria optimization methods on a finite set of alternatives]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2016. vol. 1(44). pp. 198–214.
  26. Mikoni S.V., Burakov D.P. [Setting up the typical one-dimensional utility functions in a multi-dimensional utility model]. *Izvestiya Peterburgskogo universiteta putej soobshcheniya – Proceedings of Petersburg Transport University*. 2019. Issue 16(2). pp. 131–144. (In Russ.).
  27. Kosheleva O., Kreinovich V., Shahbazova S. Type-2 Fuzzy Analysis Explains Ubiquity of Triangular and Trapezoid Membership Functions. Recent Developments and the New Direction in Soft-Computing Foundations and Applications. *Studies in Fuzziness and Soft Computing*. 2018. vol. 393. pp. 63–75.
  28. Gholamy A., Kosheleva O., Kreinovich V. How to explain the efficiency of triangular and trapezoid membership functions in applications to design. *Ontologiya proektirovaniya – Ontology of designing*. 2019. vol. 9. no. 2(32). pp. 253–260.
  29. Sajt nauchnoj shkoly “Mnogokriterial'nyj vybor na konechnom mnozhestve al'ternativ” [Website of the scientific school “Multi-criteria choice on a finite set of alternatives”]. Available at: <http://mcd-svir.ru/> (accessed: 26.08.2020). (In Russ.).
  30. Mikoni S.V. Application of the Universal Decision Support System SVIR to Solving Urban Problems. *Digital Transformation and Global Society. DTGS 2016. Communications in Computer and Information Science*. 2016. vol. 674. pp. 1–14.

А.А. БАТЕНКОВ, К.А. БАТЕНКОВ, А.Г. БОГАЧЕВ, В.В. МИШИН  
**МАТЕМАТИЧЕСКАЯ МОДЕЛЬ КЛАССИФИКАТОРА  
ОБЪЕКТОВ НА ОСНОВЕ БАЙЕСОВСКОГО ПОДХОДА**

*Батенков А.А., Батенков К.А., Богачев А.Г., Мишин В.В.* **Математическая модель классификатора объектов на основе байесовского подхода.**

**Аннотация.** Утверждается, что первостепенное значение в решении задачи классификации занимают: нахождение условий разбиения генеральной совокупности на классы, определение качества такого расслоения и верификация модели классификатора. Рассмотрена математическая модель нерандомизированного классификатора признаков, полученных без учителя, когда априори не задается число классов, а лишь устанавливается его верхняя граница. Математическая модель приведена в виде постановки минимаксной условной экстремальной задачи и представляет собой задачу поиска матрицы принадлежности объектов к какому-либо классу. В основе разработки классификатора признаков находится синтез двумерной плотности вероятностей в пространстве координат: классы – объекты. С помощью обобщенных функций вероятностная задача поиска минимума Байесовского риска сведена к детерминированной задаче на множестве нерандомизированных классификаторов. Вместе с тем использование специально введенных ограничений фиксирует нерандомизированные правила принятия решений и погружает целочисленную задачу нелинейного программирования в общую непрерывную нелинейную задачу. Для корректного синтеза классификатора необходимы дисперсионная кривая изотропной выборки и характеристики качества классификации в зависимости от суммарной внутриклассовой и межклассовой дисперсии. Задача классификации может быть интерпретирована как частная задача теории катастроф. В условиях ограниченных исходных данных найден минимаксный функционал, отражающий качество классификации при квадратичной функции потерь. Математическая модель представлена в виде задачи целочисленного нелинейного программирования и приведена с помощью полиномиальных ограничений к виду общей задачи нелинейного непрерывного программирования. Найденные необходимые условия расслоения на классы. Эти условия могут быть использованы как достаточные при проверке гипотезы о существовании классов.

**Ключевые слова:** нерандомизированный классификатор признаков, верхняя граница числа классов, минимакс, условная экстремальная задача, целочисленная задача нелинейного программирования

**1. Введение.** Выясним цели классификации, которые могут различаться в зависимости от особенностей конкретной прикладной области знаний, и место задачи классификации в общей теории системного анализа [1-3]. Известно, что под моделью понимается некоторый материальный или мысленно представляемый объект (образ объекта), который в процессе изучения замещает объект-оригинал, сохраняя некоторые важные для данного исследования типичные его свойства. Модель – это всегда упрощенное изображение реального объекта, процесса или явления. Тогда модель можно определить как некоторое сжатое изображение исходных данных. В случае их дискретного множества

такую модель можно интерпретировать как замену более мощного множества менее мощным множеством, что в общем случае соответствует задаче классификации. В связи с этим будем рассматривать классификатор как совокупность моделей (по числу классов) представления выхода некоторого источника исходных данных, где каждый класс представлен своей эталонной моделью [3, 4].

В рамках такой концепции системного анализа (синтеза) решение задачи классификации занимает центральное место и позволяет достичь следующих результатов:

1. Разбить (расслоить) множество объектов на подмножества (классы) схожих объектов для дальнейшей раздельной обработки данных.

2. Установить наличие однородности (изотропности) исходных данных, если расслоение на классы невозможно (например, в случае выборки из непрерывного многомерного равномерного распределения).

3. Сократить объем хранимых данных, оставив по одному представителю от каждого класса (кластера), то есть решить задачу сжатия данных (построить векторный квантователь).

4. Выделить нетипичные объекты, которые не подходят ни к одному из кластеров, а именно выявить уникальные (новые по отношению к существующим) модели.

5. Построить иерархию множества объектов.

В первом и втором случаях особое внимание обращают на идентификацию числа кластеров. В третьем случае важнее обеспечить высокую степень сходства объектов внутри каждого кластера, а число классов выбирают исходя из условий функционирования системы. В четвертом случае наибольший интерес представляют отдельные объекты, не вписывающиеся ни в один из кластеров, а потому они могут описывать новые знания об интересующей проблеме. В пятом случае создают вложенные модели классификаторов, где значимость приобретает степень согласованности критериев качества классификации на различных уровнях древообразной иерархической структуры.

Анализ предполагаемых результатов и приоритетных направлений их достижения позволяет сделать вывод, что первостепенное значение в решении задачи классификации занимают: нахождение условий разбиения генеральной совокупности на классы, определение качества такого расслоения и верификация модели классификатора, что является целью исследования.

**2. Условия правильной классификации.** Известно [5-7], что разбиение множества на классы – это действие распределения объектов по классам на основе сходства объектов внутри класса и их отличия от объектов других классов.

Условия правильной классификации [8].

1. Подмножества (классы) попарно не пересекаются.

2. Объединение всех подмножеств (классов) совпадает с исходным множеством.

Необходимо отметить, что условия правильной классификации могут характеризовать построение классификатора по обучающей выборке (на этапе обучения), но не соответствуют условиям правильной классификации по контролирующей выборке (на этапе контроля), так как могут быть применены рандомизированные правила принятия решений, и обучающее множество не тождественно контролирующему множеству. Кроме того, неизвестно, как разделить предоставленный статистический материал на обучающую и контролирующую выборки. Обозначим еще ряд нерешенных вопросов и проблем в автоматической классификации без учителя:

1. Как учесть при классификации неточность измерений признаков?

2. Насколько формализовано понятие класса?

3. Почему классы не должны пересекаться?

4. Допускает ли исходное множество расслоение на классы и сколько их?

5. При каких условиях классификатор становится векторным квантователем? [9-11].

Чтобы ответить хотя бы на некоторые из поставленных вопросов, разрабатываемая математическая модель должна быть достаточно общей, а значит, иметь как можно больший набор стратегий игры с природой. В этой связи рассмотрена задача поиска количественного функционала (критерия) и его параметров, отражающих качество классификации объектов без учителя.

**3. Байесовский подход при построении математической модели классификатора.** Для построения математической модели классификатора применим Байесовский подход [12-15], который используется во многих теориях (принятия решений, оценивания параметров, распознавания образов и т.д.).

Пусть  $\mathbf{Y}_n = (y_{1n}, y_{2n}, \dots, y_{Ln})^T$ ,  $\forall n = \overline{1, N}$  – дискретное множество из  $N$  векторов в пространстве размерности  $L$ . На физическом уровне это вводит  $N$  векторов признаков, подлежащих классификации, которые характеризуются дескриптором из  $L$  компонент. Далее будем ассоциировать  $N$  с числом объектов, подлежащих классификации. Использование Байесовского подхода предполагает задание вероятностной меры в пространстве признаков. Будем считать, что известен вероятностный ряд появления  $n$ -го объекта для классификации:

$P(n), \forall n = \overline{1, N}$ . Если признаки всех объектов различны, то можно считать ряд равновероятным:

$$P(n) = \frac{1}{N}, \forall n = \overline{1, N}.$$

Далее, пусть  $\mathbf{Z}_k = (z_{1k}, z_{2k}, \dots, z_{LK})^T, \forall k = \overline{1, K}$  – дискретное множество из  $K$  эталонных векторов, каждый из которых представляет центроид одного из  $K$  классов, и  $\mathbf{Z}'_k = (z'_{1k}, z'_{2k}, \dots, z'_{LK})^T, \forall k = \overline{1, K}$  – дискретное множество, тождественное эталонному множеству. Эти множества не известны в задаче классификации и могут быть определены после анализа решения исходной задачи.

На рисунке 1 представлена модель процедуры оценки [12], которую в дальнейшем будем использовать для формирования математической модели классификатора объектов.

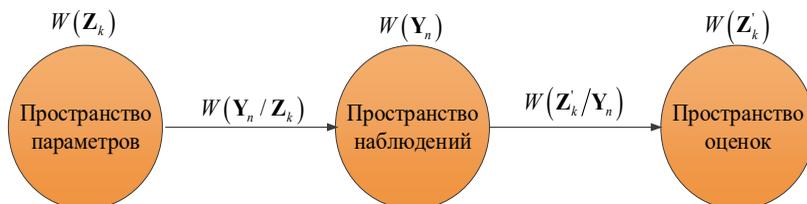


Рис. 1. Модель процедуры оценки

Применим формулу Байеса [12] для задачи классификации:

$$W(\mathbf{Z}'_k / \mathbf{Y}_n) = \frac{W(\mathbf{Z}_k)W(\mathbf{Y}_n / \mathbf{Z}_k)}{W(\mathbf{Y}_n)}, k = 1, \dots, K; n = 1, \dots, N;$$

где  $W(\mathbf{Z}'_k / \mathbf{Y}_n)$  – апостериорная плотность вероятности эталонной модели  $\mathbf{Z}_k$  при наблюдении объекта  $\mathbf{Y}_n$ ;  $W(\mathbf{Z}_k)$  – априорная плотность вероятности эталонной модели  $\mathbf{Z}_k$ ;  $W(\mathbf{Y}_n / \mathbf{Z}_k)$  – функция правдоподобия наблюдения  $\mathbf{Y}_n$  при реализации эталонной модели  $\mathbf{Z}_k$ ;  $W(\mathbf{Y}_n)$  – плотность вероятности наблюдений  $\mathbf{Y}_n$ .

В основе байесовского подхода к решению задачи классификации лежит разложение дискретной двумерной плотности вероятности

сти (в пространстве эталоны – признаки) на безусловные и условные дискретные плотности:

$$W(\mathbf{Z}_k, \mathbf{Y}_n) = W(\mathbf{Z}_k / \mathbf{Y}_n)W(\mathbf{Y}_n) = W(\mathbf{Z}_k)W(\mathbf{Y}_n / \mathbf{Z}_k), \\ k = 1, \dots, K, n = 1, \dots, N.$$

В наиболее общем виде представим ее с помощью смеси плотностей вероятности непрерывных случайных величин на основе обобщенных функций:

$$W(\mathbf{Z}, \mathbf{Y}) = \sum_{k=1}^K \sum_{n=1}^N x_{k,n} \delta(\mathbf{Z} - \mathbf{Z}_k) \delta(\mathbf{Y} - \mathbf{Y}_n),$$

где  $x_{k,n}$  – вероятность назначения в  $k$ -й кластер  $n$ -го объекта;  $\delta(\mathbf{Z})$  – обобщенная  $\delta$ -функция [13].

Далее определим маргинальные плотности вероятностей следующим образом:

$$W(\mathbf{Z} / \mathbf{Y}) = \sum_{k=1}^K \frac{x_{k,n}}{\sum_{i=1}^K x_{i,n}} \delta(\mathbf{Z} - \mathbf{Z}_k) 1(\mathbf{Y} - \mathbf{Y}_n), \quad \forall n = 1, \dots, N.$$

Вывод выражения для апостериорной плотности вероятности принадлежности  $n$ -ого объекта  $k$ -ому классу приведен в приложении 1.

Аналогично апостериорной плотности получим формулировку функции правдоподобия:

$$W(\mathbf{Y} / \mathbf{Z}) = \sum_{n=1}^N \frac{x_{k,n}}{\sum_{s=1}^N x_{k,s}} 1(\mathbf{Z} - \mathbf{Z}_k) \delta(\mathbf{Y} - \mathbf{Y}_n), \quad \forall k = 1, \dots, K, \\ W(\mathbf{Y}) = \sum_{n=1}^N \delta(\mathbf{Y} - \mathbf{Y}_n) \sum_{k=1}^K x_{k,n}.$$

Здесь введена специальная разрывная  $1(\mathbf{Z})$ -функция:

$$1(\mathbf{Z}) = \begin{cases} 1, & \mathbf{Z} = \mathbf{0}, \\ 0, & \mathbf{Z} \neq \mathbf{0}. \end{cases}$$

Ее основные свойства приведены в приложении 2 [16, 17]. Условия нормировки для введенных плотностей показаны в приложении 3.

При таком подходе решение задачи классификации сводится к формированию двумерной плотности вероятностей  $W(\mathbf{Z}, \mathbf{Y})$ , а именно к нахождению матрицы  $\mathbf{X}$  вероятностей назначений, размера  $K \times N$ , объектов в кластеры при ограничениях, представляющих собой условия нормировки. Однако обязательными дополнительными ограничениями должно быть то, что все представленные объекты необходимо классифицировать, то есть каждый объект должен принадлежать хотя бы одному классу, и то, что класс может не содержать объектов (быть пустым), так как априори неизвестно число классов. Если число классов известно, то это порождает задачу векторного квантования.

**4. Матричная модель классификатора.** Заметим, что искомая двумерная дискретная плотность распределения вероятностей состоит из трех компонент: матрицы  $\mathbf{X}$ , множества признаков и множества эталонных векторов  $\mathbf{Z}'_k = (z'_{1k}, z'_{2k}, \dots, z'_{LK})^T$ ,  $\forall k = \overline{1, K}$ . Элементы множеств признаков и эталонных векторов связаны между собой с помощью матрицы  $\mathbf{X}$  – вероятностей назначений в  $k$ -й кластер  $n$ -го объекта. Тогда  $\mathbf{X}$  может рассматриваться как искомый двумерный ряд распределения дискретных случайных величин  $\{k = \overline{1, K}\}$  и  $\{n = \overline{1, N}\}$ , и следовательно, допустимо получение любой маргинальной плотности вероятности, применяемой при байесовском подходе. Таким образом, если имеется матрица  $\mathbf{X}$  с элементами  $x_{k,n}$ , удовлетворяющих условиям:

$$x_{k,n} \geq 0, \sum_{k=1}^K \sum_{n=1}^N x_{k,n} = 1, \forall k = \overline{1, K}, n = \overline{1, N},$$

то можно найти маргинальные плотности, опуская обозначения соответствующих  $\delta$ ,  $1$ ,  $\varepsilon$ -функций, то есть определить маргинальные ряды распределения:

$$P(k, n) = x_{k,n}, P(k/n) = \frac{x_{k,n}}{\sum_{s=1}^N x_{s,n}}, P(k) = \sum_{n=1}^N x_{k,n},$$

$$P(n/k) = \frac{x_{k,n}}{\sum_{s=1}^K x_{k,s}}, P(n) = \sum_{k=1}^K x_{k,n}, \forall k = \overline{1, K}, n = \overline{1, N}.$$

Для упрощения записи в матричном виде маргинальные ряды представим в виде:

$$\begin{aligned} P(k, n) &= \mathbf{X}, P(k / n) = \mathbf{X} \left[ \text{diag}(\mathbf{X}^T \mathbf{I}_k) \right]^{-1}; \\ P(k) &= \mathbf{X} \mathbf{I}_n, P(n / k) = \mathbf{X}^T \left[ \text{diag}(\mathbf{X} \mathbf{I}_n) \right]^{-1}, \\ P(n) &= \mathbf{X}^T \mathbf{I}_k, \forall k = \overline{1, K}, n = \overline{1, N}. \end{aligned}$$

где  $\text{diag}(\mathbf{x})$  – оператор диагонализации вектора  $\mathbf{x}$ .

Однако поиск неизвестной матрицы вероятностей назначений  $\mathbf{X}$  значительно усложняет задачу классификации вследствие большого количества ограничений на ее элементы [18]. Поэтому будем искать решение в виде матрицы назначений  $\mathbf{X}'$  с булевыми элементами. При этом сама матрица назначений  $\mathbf{X}'$  естественно взаимосвязана с матрицей вероятностей назначений  $\mathbf{X}$  [19, 20]. Введение булевой матрицы назначений возможно ввиду того, что классификатор нерандомизированный.

Для этого определим связь условного маргинального ряда вероятностей  $P(k / n) = \mathbf{X}'$  и двумерного ряда вероятностей  $P(k, n) = \mathbf{X}$  в матричном виде:

$$\mathbf{X}' \mathbf{G} = \mathbf{X}, \mathbf{G} = \text{diag}(\mathbf{X}^T \mathbf{I}_k), \mathbf{X}' = \mathbf{X} \left[ \text{diag}(\mathbf{X}^T \mathbf{I}_k) \right]^{-1}, \mathbf{X}' \text{diag}(\mathbf{X}^T \mathbf{I}_k) = \mathbf{X},$$

в скалярном виде:

$$x_{k,n} = x'_{k,n} b_n, k = \overline{1, K}, n = \overline{1, N}, b_n = \sum_{m=1}^K x_{m,n}, x_{k,n} = x'_{k,n} \sum_{m=1}^K x_{m,n},$$

где  $\mathbf{I}_k = \underbrace{[1 \ \dots \ 1 \ \dots \ 1]}_k^T$  – единичный вектор размера  $K \times 1$ .

В задаче классификации будем использовать следующие условия нормировки, которые порождают ограничения в создаваемой математической модели:

$$\mathbf{I}_k^T \mathbf{X} \mathbf{I}_n = 1; \tag{1}$$

$$\mathbf{I}_k^T \mathbf{X} = \mathbf{I}_n^T; \tag{2}$$

$$x'_{k,n} = \{0; 1\}, \forall k = \overline{1, K}, n = \overline{1, N}, \mathbf{X}' \bullet \mathbf{X}' - \mathbf{X}' = \mathbf{0}; \quad (3)$$

$$\begin{aligned} x_{k,n} &\geq 0, \forall k = \overline{1, K}, n = \overline{1, N}; \\ \mathbf{X} - \mathbf{Q} \bullet \mathbf{Q} &= \mathbf{0}, \end{aligned} \quad (4)$$

где  $\mathbf{A} \bullet \mathbf{B}$  – произведение Адамара матриц  $\mathbf{A}$  и  $\mathbf{B}$ ;  $\mathbf{Q}$  – матрица дополнительных переменных размера  $K \times N$ , необходимых для перехода от ограничений неравенств (на неотрицательность переменных) к ограничениям типа равенств;  $\mathbf{0}$  – нулевая матрица размера  $K \times N$ .

Ограничение (1) означает, что матрица  $\mathbf{X}$  – представляет собой дискретный двумерный ряд распределения вероятностей.

Ограничения (2), наложенные на апостериорную плотность вероятности, обеспечивают классификацию каждого объекта, то есть их принадлежность одному или нескольким классам с определенной вероятностью. Тогда составляющие:

$$\mathbf{X}' = \mathbf{X} \left[ \text{diag}(\mathbf{X}^T \mathbf{I}_k) \right]^{-1}$$

интерпретируются как условные вероятности принадлежности  $n$ -го объекта  $k$ -му классу. Если эти компоненты определяются как элементы булевого множества  $\{0; 1\}$ , то классификатор оказывается нерандомизированным.

Ограничения (3) задают детерминированное множество решающих правил, то есть разрабатываемая математическая модель является нерандомизированным классификатором, и погружают дискретную булеву задачу в непрерывную.

Ограничения (4) позволяют интерпретировать составляющие матрицы  $\mathbf{X}$  как вероятности.

Совокупность ограничений (2) и (3) обуславливает разбиение [21, 22] множества из « $N$ » объектов на « $K$ » классов, то есть вероятность принадлежности к  $k$ -му классу при условии наблюдения объекта с номером  $n$ . Однако условная плотность вероятности не определяет полностью основу математической модели классификатора – двумерную плотность вероятности, которая имеет большее число степеней свободы [23, 24]. В условиях равновероятного ряда появления объектов для классификации нетрудно увидеть, что ограничения (1) и (4) являются избыточными и  $\mathbf{X}'$  однозначно определяет  $\mathbf{X}$ .

Очень важно отметить, что ограничения на функцию правдоподобия не вводятся, что обеспечивает возможное существование пустых

кластеров. Поэтому при автоматической классификации можно задавать лишь верхнюю границу числа классов либо осуществлять полное разбиение объектов на неупорядоченные классы, так как возможное число классов всегда определяется числом объектов:  $1 \leq K \leq N$ .

Тогда рассматриваемая математическая модель должна содержать постановку задач определения:

- сходства объектов внутри класса;
- отличия объектов класса от объектов других классов;
- числа классов;
- оптимального разбиения множества объектов на классы;
- гипотетического (представительного) объекта внутри каждого класса.

**5. Межклассовая метрика.** Для того чтобы сформулировать такую задачу, необходимо ввести метрику (расстояние) между и внутри множеств  $\mathbf{Y}_n, \forall n = \overline{1, N}$ , и  $\mathbf{Z}_k, \forall k = \overline{1, K}$ .

Эта метрика должна удовлетворять следующим требованиям:

- обладать известными свойствами метрики в смысле теории метрических пространств;
- отражать физическую сущность функции потерь при решении задачи минимизации Байесовского риска;
- иметь не высокую вычислительную сложность;
- учитывать статистические свойства результатов измерений признаков;
- использовать полученный классификатор для решения последующих задач (например, применение нелинейных функционалов для синтеза разделяющих многомерных поверхностей за счет увеличения размерности пространства [25- 27]).

Конструирование такой метрики представляет собой самостоятельную научную задачу, которой будут посвящены дальнейшие исследования. Поэтому здесь мы ограничимся математической моделью классификатора признаков в смысле экстремальной задачи с ограничениями при использовании обобщенной метрики евклидова пространства [28-30]:

$$\rho(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{W} (\mathbf{x} - \mathbf{y})},$$

где  $\mathbf{W}$  – положительно определенная квадратная матрица размера  $(L \times L)$ .

**6. Условия построения классификатора.** Разрабатываемая математическая модель классификатора признаков основана на следую-

щих утверждениях, рассматриваемых сначала как некоторые гипотезы, которые в последующем могут быть доказаны.

Так при разбиении множества на классы цели являются противоречивыми. С одной стороны, при увеличении числа классов сходство объектов внутри класса увеличивается (расстояние между объектами одного класса уменьшается), а с другой стороны, уменьшается межклассовое различие (расстояние между представителями разных классов уменьшается). Поэтому обоснование числа классов проводится эвристически, без достаточного научного обоснования. Кроме того, множество может вообще не расслаиваться на классы, вследствие недостаточного статистического материала и (или) неправильного отбора признаков, и такие ситуации должны быть обнаружены. Поэтому концептуально выдвигается лишь гипотеза о существовании расслоения исходного множества на классы с помощью разрабатываемой математической модели. Эта гипотеза на заключительном этапе подлежит проверке (верификации), а именно: допускает ли исследуемое пространство признаков решение задачи классификации? В связи с этим рассмотрение достаточных условий правильной классификации является актуальной задачей.

Кроме того, конструирование функционала качества классификации предполагает рассмотрение двух граничных условий: 1) кластер является единственным; 2) число классов равно объему выборки. В первом случае межклассовая дисперсия минимальная (равна нулю), а внутриклассовая дисперсия максимальная. Во втором случае наоборот: межклассовая дисперсия максимальная, а суммарная внутриклассовая дисперсия минимальная (равна нулю). Кроме того, если суммарная внутриклассовая дисперсия может быть выбрана для количественного описания сходства объектов внутри классов, то межклассовая дисперсия недостаточно описывает отличие объектов разных классов, так как она определяется лишь координатами эталонов каждого класса и не учитывает «тонкого» распределения объектов внутри класса.

В противоположность задаче классификации в задаче векторного квантования основное внимание уделяется минимизации суммарной внутриклассовой дисперсии, а значит, все классы должны быть не пустыми, и конструирование функционала качества векторного квантователя не вызывает затруднений. Этой задаче будут посвящены последующие работы.

Интересно рассмотрение задачи классификации с позиций теории информации [9]. Известно [9], что дифференциальная энтропия монотонно возрастает с возрастанием дисперсии генеральной совокупности и достигает своего максимального значения для установленных плотностей вероятностей случайной величины в зависимости от ее области определения. Поэтому можно высказать следующие гипотезы, что:

1. При расслоении выборки на классы ее суммарная внутриклассовая дисперсия должна быть меньше максимально возможной, то есть, когда нет расслоения при заданном числе классов.

2. Чтобы идентифицировать расслоения, должна быть некоторая гипотетическая выборка (образец), в которой потенциально невозможно выделить кластеры (например, выборка из многомерного равномерного распределения [31, 32]);

3. Проверка предположения о наличии расслоения на классы может быть проведена путем сравнения рабочего и гипотетического множеств (например, с помощью расстояния Хаусдорфа [33], использования информационной меры Кюльбака – Лейблера [34] и т. д.).

4. Система замкнута [1], а значит ее энтропия и, следовательно, дисперсия не зависят от способа разбиения выборки на классы, то есть:

$$D_m(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) + D_v(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \text{const}(\mathbf{Y}),$$

где  $\mathbf{X}$  – матрица размера  $K \times N$  назначений объектов в кластеры;  $K$  – число классов;  $N$  – число классифицируемых объектов;  $D_m$  – межклассовая дисперсия (дисперсия эталонов);  $D_v$  – суммарная по всем классам внутриклассовая дисперсия;  $\mathbf{Y}$  – матрица координат размера  $L \times N$  распознаваемого множества объектов;  $L$  – размерность пространства признаков;  $\mathbf{Z}$  – матрица координат размера  $L \times K$  представителей классов.

5. Существует канонический вид гипотетической и рабочей выборок, основанный на свойстве замкнутости классификатора.

Эти гипотезы наталкивают на комплексное решение проблемы классификации:

– определение дисперсионных последовательностей (кривых), образуемых различными разбиениями гипотетической (образцовой) изотропной выборки;

– приведение рабочей выборки к каноническому виду;

– проверка гипотезы о расслоении рабочей выборки на кластеры;

– синтез классификатора;

– верификация полученной математической модели.

В условиях, когда гипотеза о расслоении рабочей выборки на классы подтверждается (проверке этой гипотезы будут посвящены последующие исследования), основной задачей можно считать задачу синтеза классификатора по некоторому функционалу качества.

Функционал качества сформируем в общем виде (по принципу гарантированного результата) без априорного задания числа классов (оно естественно ограничено сверху объемом выборки) и эталонов на множестве нерандомизированных классификаторов, задаваемых как

множество всех неупорядоченных разбиений исходного множества объектов [35, 36]. При этом каждое разбиение (классификатор) будем характеризовать двумя параметрами: межклассовой дисперсией и суммарной внутриклассовой дисперсией.

**7. Дисперсии разбиений.** Известен теоретико-множественный подход к постановке задачи классификации [5]. Однако его применение совместно с математическим программированием [37] невозможно. Поэтому математическую модель представим как некоторую условную экстремальную задачу в формальном матричном виде.

Для этого конкретизируем межклассовую дисперсию.

Следуя Байесовскому подходу, будем рассматривать оценки  $\{\mathbf{Z}'\}$  множества эталонов  $\{\mathbf{Z}\}$  при заданной функции стоимости (потерь), наблюдая множество признаков  $\{\mathbf{Y}\}$ . Так как, в ходе решения задачи классификации необходимо определить: 1) априорное множество эталонов  $\{\mathbf{Z}\}$ ; 2) правило принятия решения о принадлежности объекта к классу; 3) апостериорное множество оценок  $\{\mathbf{Z}'\}$ , то естественно положить, что оценки должны совпадать с эталонами  $\{\mathbf{Z}'\} \equiv \{\mathbf{Z}\}$  и также должны быть найдены в результате решения задачи классификации.

Известно [9], что в случае квадратичной функции потерь, а также в достаточно широком классе других функций стоимости, оптимальная оценка есть среднее значение апостериорной плотности вероятности признаков класса. Тогда применительно к задаче детерминированной классификации с необходимыми плотностями вероятностей:

$$\begin{aligned} W(\mathbf{Z}, \mathbf{Y}) &= W(\mathbf{Z} / \mathbf{Y}) W(\mathbf{Y}) = \\ &= \sum_{n=1}^N \sum_{k=1}^K x_{k,n} \frac{\delta(\mathbf{Z} - \mathbf{Z}_k) 1(\mathbf{Z} - \mathbf{Y}_n)}{\sum_{s=1}^K x_{s,n}} \delta(\mathbf{Y} - \mathbf{Y}_n) \sum_{s=1}^K x_{s,n} = \\ &= \sum_{n=1}^N \sum_{k=1}^K x_{k,n} \delta(\mathbf{Z} - \mathbf{Z}_k) 1(\mathbf{Y} - \mathbf{Y}_n) \delta(\mathbf{Y} - \mathbf{Y}_n); \\ W(\mathbf{Y}) &= \sum_{n=1}^N \delta(\mathbf{Y} - \mathbf{Y}_n) \sum_{k=1}^K x_{k,n}; \\ W(\mathbf{Z} / \mathbf{Y}) &= \sum_{k=1}^K \frac{x_{k,n}}{\sum_{s=1}^K x_{s,n}} \delta(\mathbf{Z} - \mathbf{Z}_k) 1(\mathbf{Y} - \mathbf{Y}_n), \forall n = 1, \dots, N, \end{aligned}$$

составим матрицу оптимальных оценок  $\mathbf{Z}'$  размера  $(L \times N)$ , которая может быть выражена как:

$$\mathbf{Z}'_n = \int_{\Omega_{\mathbf{Z}}} \mathbf{Z} W(\mathbf{Z} / \mathbf{Y}) d\mathbf{Z} = \mathbf{Z} \mathbf{X}'_n \mathbf{1}(\mathbf{Y} - \mathbf{Y}_n), \forall n = \overline{1, N};$$

$$\mathbf{Z}' = [\mathbf{Z}'_1, \mathbf{Z}'_2, \dots, \mathbf{Z}'_N] = \mathbf{Z} \mathbf{X}',$$

где  $\mathbf{Z}'_n$  – оценка  $\mathbf{Z}_n$  при наблюдении вектора признаков  $\mathbf{Y}_n$ ;  $\mathbf{Z}'_n \in \{\mathbf{Z}'_k = (z_{1k}, z_{2k}, \dots, z_{LK})^T\}, \forall k = \overline{1, K}$ ;  $\mathbf{Z}$  – формируемое эталонное множество – матрица размера  $(L \times K)$ ;  $\mathbf{X} = \mathbf{X}' \text{diag}(\mathbf{X}'^T \mathbf{I}_k)$  – матрица размера  $(K \times N)$  апостериорных вероятностей эталонов при наблюдении вектора признаков  $\mathbf{Y}_n$ ;  $\mathbf{X}'_n$  –  $n$ -й столбец матрицы  $\mathbf{X}'$ .

Тогда можно найти среднее значение на множестве оценок:

$$\mathbf{Z}'_s = \mathbf{Z} \mathbf{X}' \mathbf{X}'^T \mathbf{I}_k,$$

и определить дисперсию оценок как

$$D_m(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \sum_{k=1}^K \sum_{n=1}^N x_{k,n} \int_{\Omega_{\mathbf{Y}}} \int_{\Omega_{\mathbf{Z}}} \left[ (\mathbf{Z}' - \mathbf{Z}'_s)^T \times \right. \\ \left. \times \mathbf{W}(\mathbf{Z}' - \mathbf{Z}'_s) \right] \delta(\mathbf{Z}' - \mathbf{Z}'_n) \delta(\mathbf{Y} - \mathbf{Y}_n) d\mathbf{Z} d\mathbf{Y} = \\ = \sum_{k=1}^K \sum_{n=1}^N x_{k,n} (\mathbf{Z}'_n - \mathbf{Z}'_s)^T \mathbf{W}(\mathbf{Z}'_n - \mathbf{Z}'_s).$$

Эталоны  $\mathbf{Z}_k$  будем искать как матрицу координат эталонов  $\mathbf{Z}$  размера  $L \times K$ . Тогда после подстановки оценок и их среднего, используя фильтрующее свойство  $\delta$ -функции, получим:

$$D_m(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \text{tr} \left\{ \left( \mathbf{Z} \mathbf{X}' - \mathbf{I}_n^T \otimes \{ \mathbf{Z} \mathbf{X}' \mathbf{X}'^T \mathbf{I}_k \} \right)^T \mathbf{W} \times \right. \\ \left. \times \left( \mathbf{Z} \mathbf{X}' - \mathbf{I}_n^T \otimes \{ \mathbf{Z} \mathbf{X}' \mathbf{X}'^T \mathbf{I}_k \} \right) \text{diag}(\mathbf{X}'^T \mathbf{I}_k) \right\},$$

где  $\text{tr}(\mathbf{A})$  – след квадратной матрицы  $\mathbf{A}$ ;  $\mathbf{A} \otimes \mathbf{B}$  – произведение Кронекера матриц  $\mathbf{A}$  и  $\mathbf{B}$ .

Отметим, что здесь под межклассовой дисперсией понимается не дисперсия эталонного множества, мощность которого равна  $K$ , а дисперсия рабочей выборки после замены каждого объекта всех клас-

сов его эталонным значением для класса, которому он принадлежит, а значит, мощность преобразованного множества равна  $N$ .

Теперь введем и рассмотрим суммарную внутриклассовую (усредненную на множестве признаков класса) дисперсию, которая в общем случае задается выражением:

$$D_v(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \sum_{k=1}^K \int_{\Omega_{\mathbf{Y}_k}} \int_{\Omega_{\mathbf{Z}_k}} (\mathbf{Y} - \mathbf{Z})^T \mathbf{W}(\mathbf{Y} - \mathbf{Z}) \mathcal{W}(\mathbf{Z}, \mathbf{Y}) d\mathbf{Z} d\mathbf{Y};$$

$$\Omega_{\mathbf{Y}} = \bigcup_k \Omega_{\mathbf{Y}_k}, \bigcap_k \Omega_{\mathbf{Y}_k} = \emptyset, \Omega_{\mathbf{Z}} = \bigcup_k \Omega_{\mathbf{Z}_k}, \bigcap_k \Omega_{\mathbf{Z}_k} = \emptyset, \text{card } \Omega_{\mathbf{Z}_k} = 1, \forall k = \overline{1, K};$$

$$D_v(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \sum_{k=1}^K \sum_{n=1}^N x_{k,n} \int_{\Omega_{\mathbf{Y}}} \int_{\Omega_{\mathbf{Z}}} (\mathbf{Y} - \mathbf{Z}')^T \mathbf{W} \times$$

$$\times (\mathbf{Y} - \mathbf{Z}') \delta(\mathbf{Z}' - \mathbf{Z}'_n) 1(\mathbf{Y} - \mathbf{Y}_n) \delta(\mathbf{Y} - \mathbf{Y}_n) d\mathbf{Y} d\mathbf{Z}'.$$

Используя фильтрующее свойство  $\delta$ -функции окончательно получим:

$$D_v(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \sum_{k=1}^K \sum_{n=1}^N x_{k,n} (\mathbf{Y}_n - \mathbf{Z}'_n)^T \mathbf{W}(\mathbf{Y}_n - \mathbf{Z}'_n),$$

где  $\mathbf{Z}'_n$  –  $n$ -й столбец матрицы  $\mathbf{Z}'$ .

Суммарную внутриклассовую дисперсию определим с помощью следующей конструкции:

$$D_v(K, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \mathbf{I}_k^T \mathbf{X} \text{diag} \left[ (\mathbf{Y} - \mathbf{Z} \mathbf{X}')^T \mathbf{W}(\mathbf{Y} - \mathbf{Z} \mathbf{X}') \right] \mathbf{I}_n.$$

Для сокращения размерности пространства решений учтем наличие связи между искомыми переменными и ограничениями задачи. Так:

$$\text{diag}(\mathbf{X}^T \mathbf{I}_k) = \frac{1}{N} \mathbf{I}, \mathbf{X} = \frac{1}{N} \mathbf{X}', \mathbf{X}'^T \mathbf{I}_k = \mathbf{I}_n,$$

где  $\mathbf{I}$  – единичная матрица.

С этой же целью будем искать координаты центроидов как взвешенное среднее векторов, находящихся в кластере:

$$\mathbf{Z} = \mathbf{W} \mathbf{Y} \mathbf{X}'^T \left( \text{diag}(\mathbf{X}' \mathbf{I}_n) \right)^{-1}.$$

Полученные выражения для дисперсий содержат явную зависимость от предполагаемого числа классов  $K$ , которая выражается числом строк в матрице назначений  $\mathbf{X}'$ . Для исключения этой зависимости определим эту матрицу для максимально возможного числа классов размером  $(N \times N)$ .

Тогда выражения для дисперсий записываются в виде:

$$D_m(\mathbf{X}') = \frac{1}{N} \operatorname{tr} \left\{ \left( \mathbf{W} \mathbf{Y} \mathbf{X}'^T [\operatorname{diag}(\mathbf{X}' \mathbf{I}_n)]^{-1} \mathbf{X}' - \frac{1}{N} \mathbf{I}_n^T \otimes \left\{ \mathbf{W} \mathbf{Y} \mathbf{X}'^T [\operatorname{diag}(\mathbf{X}' \mathbf{I}_n)]^{-1} \mathbf{X}' \mathbf{I}_n \right\} \right)^T \times \right. \\ \left. \times \mathbf{W} \left( \mathbf{W} \mathbf{Y} \mathbf{X}'^T [\operatorname{diag}(\mathbf{X}' \mathbf{I}_n)]^{-1} \mathbf{X}' - \frac{1}{N} \mathbf{I}_n^T \otimes \left\{ \mathbf{W} \mathbf{Y} \mathbf{X}'^T [\operatorname{diag}(\mathbf{X}' \mathbf{I}_n)]^{-1} \mathbf{X}' \mathbf{I}_n \right\} \right) \right\}, \quad (5)$$

$$D_v(\mathbf{X}') = \frac{1}{N} \mathbf{I}_k^T \mathbf{X}' \times \operatorname{diag} \left\{ \left[ \mathbf{Y} - \mathbf{W} \mathbf{Y} \mathbf{X}'^T (\operatorname{diag}(\mathbf{X}' \mathbf{I}_n))^{-1} \mathbf{X}' \right]^T \mathbf{W} \times \right. \\ \left. \times \left[ \mathbf{Y} - \mathbf{W} \mathbf{Y} \mathbf{X}'^T (\operatorname{diag}(\mathbf{X}' \mathbf{I}_n))^{-1} \mathbf{X}' \right] \right\} \mathbf{I}_n. \quad (6)$$

Максимальные значения (5) и (6) используем для определения канонического вида рабочей выборки, для чего произведем нормировку матрицы признаков  $\mathbf{Y}$  по правилу:

$$\mathbf{Y}_s = \frac{\mathbf{Y}}{\sqrt{s}},$$

где  $s = D_m(\mathbf{I}) = D_v(\mathbf{I}_n)$ , что означает, что максимальное значение межклассовой дисперсии наблюдается, когда число классов равно числу объектов классификации, а максимальное значение суммарной внутриклассовой дисперсии – если класс один. Эти значения одинаковы и равны выборочной дисперсии исходной выборки, что несложно показать (для упрощения  $\mathbf{W} = \mathbf{I}$ ):

$$s = \max \{ D_m(\mathbf{X}') \} = \frac{1}{N} \operatorname{tr} \left\{ \left( \mathbf{Y} - \frac{1}{N} \mathbf{I}_n^T \otimes \{ \mathbf{Y} \mathbf{I}_n \} \right)^T \left( \mathbf{Y} - \frac{1}{N} \mathbf{I}_n^T \otimes \{ \mathbf{Y} \mathbf{I}_n \} \right) \right\},$$

которое получено из (5), и

$$s = \max \{Dv(\mathbf{X}')\} = \frac{1}{N} \mathbf{I}_n^T \text{diag} \left[ \left( \mathbf{Y} - \frac{1}{N} \mathbf{Y} \mathbf{I}_n \mathbf{I}_n^T \right)^T \left( \mathbf{Y} - \frac{1}{N} \mathbf{Y} \mathbf{I}_n \mathbf{I}_n^T \right) \right] \mathbf{I}_n,$$

определяемое из (6). Нетрудно заметить, что эти выражения для масштабирующей константы эквивалентны.

**8. Функционал качества классификатора.** Теперь введем некоторый функционал качества классификатора:

$$\Phi(\mathbf{Y}_r, \mathbf{H}_s, \mathbf{X}') \xrightarrow{\mathbf{X}'} \text{extr}, \quad (7)$$

где  $\mathbf{Y}_r$  – рабочая выборка канонического вида;  $\mathbf{H}_s$  – гипотетическая (образцовая) выборка канонического вида;  $\mathbf{X}'$  – искомое разбиение рабочей выборки на классы.

Для конструирования функционала будем использовать зависимости дисперсий канонических рабочей и гипотетической выборок от разбиений на классы:  $D_{v_s}(\mathbf{X}'_i, \mathbf{H}_s)$  – суммарная внутриклассовая дисперсия изотропной канонической выборки;  $D_{v_s} = D_{v_s}(\mathbf{Y}_r, \mathbf{X}')$  – суммарная внутриклассовая дисперсия канонической рабочей выборки.

Чтобы получить формализованное описание функционала (7), были проведены исследования зависимости межклассовой и суммарной внутриклассовой дисперсий от неупорядоченных разбиений путем их полного перебора на различных рабочих выборках. По сути, функционал (7) должен ввести отношение порядка [1] на множестве всех возможных неупорядоченных разбиений, определяемых числом Белла [38] и обладающих следующими свойствами:

- число классов в разбиении не фиксируется и ограничено сверху числом объектов  $N$ , что позволяет избавиться от явного задания числа кластеров в задаче классификации;

- условия (2) и (3) формально задают искомые разбиения (область определения функционала (7)), хотя могут описывать и упорядоченные разбиения при дальнейшем развитии задачи классификации;

- полный перебор всех разбиений осуществляется формированием матриц назначений  $\mathbf{X}'$  путем использования известного рекуррентного свойства [35] (необходимые свойства разбиений приведены в приложении 4);

- для идентификации оптимального расслоения должна быть некоторая гипотетическая изотропная выборка (образец), в которой потенциально невозможно выделить кластеры;

– значения признаков объектов изотропной выборки ограничены сверху и снизу и образуют компактное [33] множество, чему удовлетворяет выборка из непрерывного многомерного равномерного распределения [31] (его определение приведено в приложении 5).

Для того чтобы здесь не конкретизировать область  $S$ , приведем изотропную выборку к каноническому виду с помощью вышеприведенных масштабирующих констант  $S$ . Это должно обеспечить некоторый стандарт дисперсионной последовательности (кривой), определяемой различными разбиениями гипотетической (образцовой) изотропной выборки (рис. 2). Эта последовательность не должна зависеть от размерности признакового пространства и числа объектов классификации, если оно велико. Конечно, этот факт нуждается в строгом математическом доказательстве. Приведем лишь экспериментальное подтверждение существования такой последовательности при малом числе объектов и для двумерного признакового пространства, так как вычислительная сложность экспериментов экспоненциально возрастает с ростом числа объектов и увеличением размерности пространства признаков.

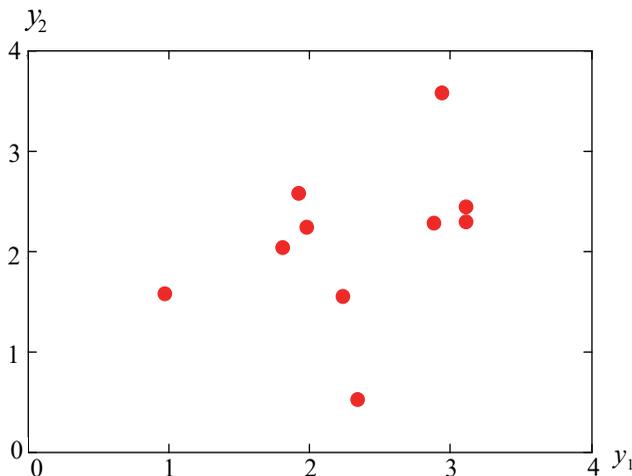


Рис. 2. Корреляционное поле изотропной выборки

На рисунке 3 изображены зависимости суммарной внутриклассовой дисперсии  $D_{v_s}$  от разбиений изотропных канонических выборок различных объемов. При этом разбиения упорядочены по возрастанию дисперсии, а значит, образуют монотонную неубывающую последовательность. Значения по оси абсцисс не имеют физического смысла, а лишь образуют порядковую шкалу  $\nu$  на множестве разбиений.

Здесь использовались выборки объемом:  $N = 7, 8, 9, 10$ , для которых числа Белла равны:  $B = 877, 4140, 21147, 115975$  соответственно.

Представленные графики позволяют заметить особенности последовательности:

- дисперсия изменяется дискретно плавно без резких скачков с ростом объема выборки;
- при увеличении числа классифицируемых объектов кривая, отображающая последовательность, стремится к некоторому предельному положению;
- при  $N \rightarrow \infty$  дисперсионная кривая изотропной канонической выборки может носить характер устойчивой закономерности.

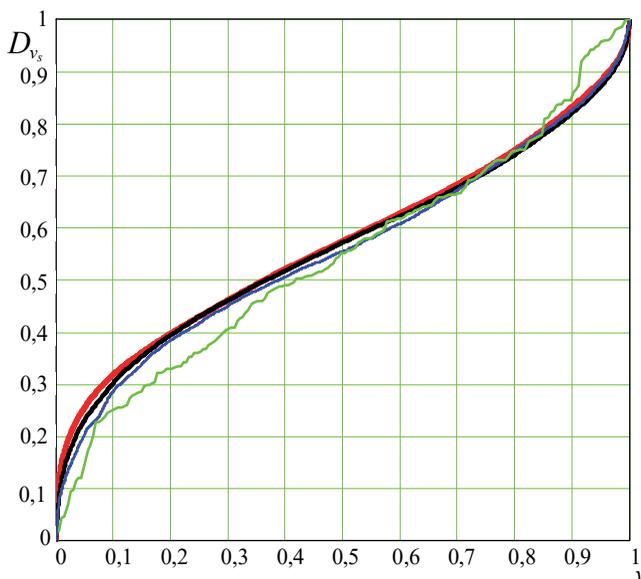


Рис. 3. Дисперсионные кривые изотропной выборки

**9. Дисперсионные кривые.** Теперь рассмотрим поведение дисперсионной кривой анизотропной выборки, то есть выборки с возможным расслоением на классы (рис. 4). Из большого количества экспериментов по исследованию анизотропных выборок в работе обсуждается один из них (наиболее представительный). Выявленные свойства дисперсионной кривой устойчиво сохраняются и в других экспериментах.

На рисунке 5 изображены зависимости суммарной внутриклассовой дисперсии  $D_{v_s}$  и межклассовой дисперсии  $D_{m_s}$  от разбиений

изотропной (рис. 2) и анизотропной (рис. 4) канонических выборок объема  $N=10$ . При этом разбиения упорядочены по уменьшению межклассовой дисперсии, а значит, образуют монотонную невозрастающую последовательность. Здесь же на графике изображена сумма внутриклассовой и межклассовой дисперсий выборок канонического вида. Эта сумма не зависит от разбиения исходной рабочей совокупности и равна ее общей дисперсии, а после приведения к каноническому виду соответствует единичному значению.

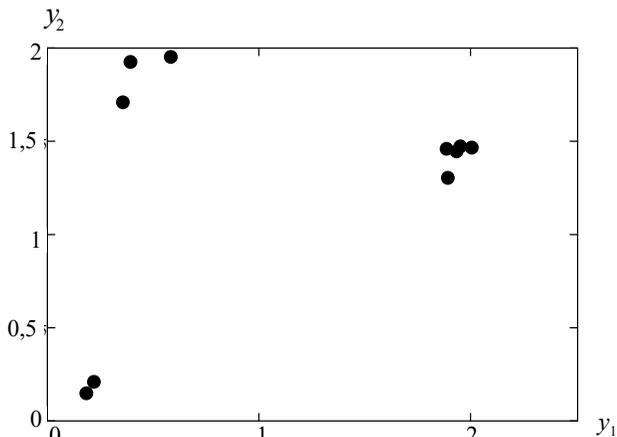


Рис. 4. Корреляционное поле анизотропной выборки

Значения по оси абсцисс на рисунке 5, как и ранее на рисунке 3, не имеют физического смысла, а лишь образуют порядковую шкалу  $\nu$  на множестве разбиений.

Установлено, что дисперсионные кривые изотропной и анизотропной выборок отличаются друг от друга тем больше, чем больше неоднородность анизотропной выборки. При дальнейших исследованиях это свойство может быть использовано для проверки гипотезы о расслоении статистического материала на классы.

Дискретные зависимости на рисунке 5 аппроксимированы кусочно-линейными непрерывными кривыми. Поэтому детализируем их с точностью до дискретных отсчетов. Особый интерес представляет условие, когда число классов значительно меньше, чем объем выборки:  $K \ll N$  при малых (больших) значениях суммарной внутриклассовой дисперсии (межклассовой дисперсии). Для этих условий представлены зависимости на рисунках 6 и 7 суммарной внутриклассовой  $D_{v_s}$  и межклассовой  $D_{m_s}$  дисперсий соответственно.

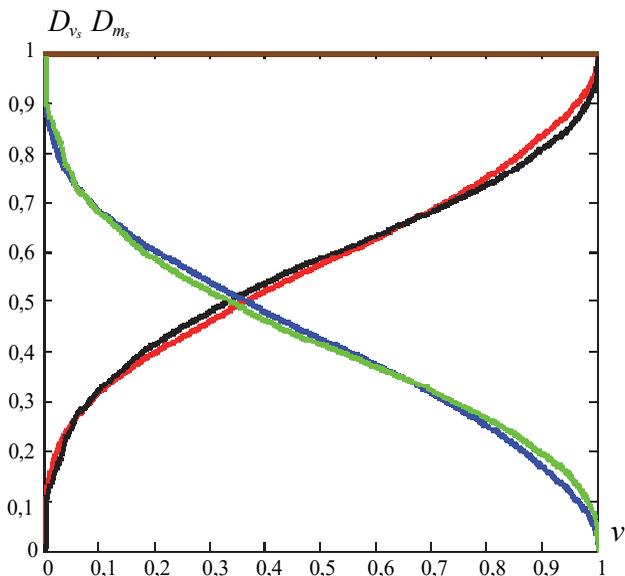


Рис. 5. Дисперсионные кривые изотропной и анизотропной выборок

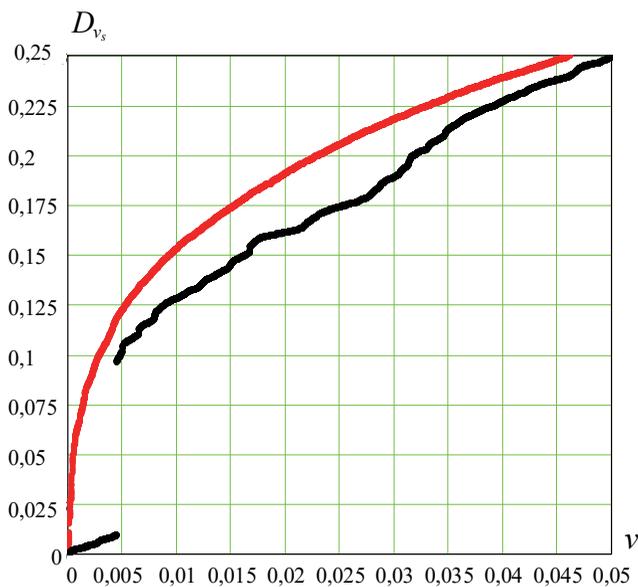


Рис. 6. Внутриклассовые дисперсионные последовательности изотропной и анизотропной выборок

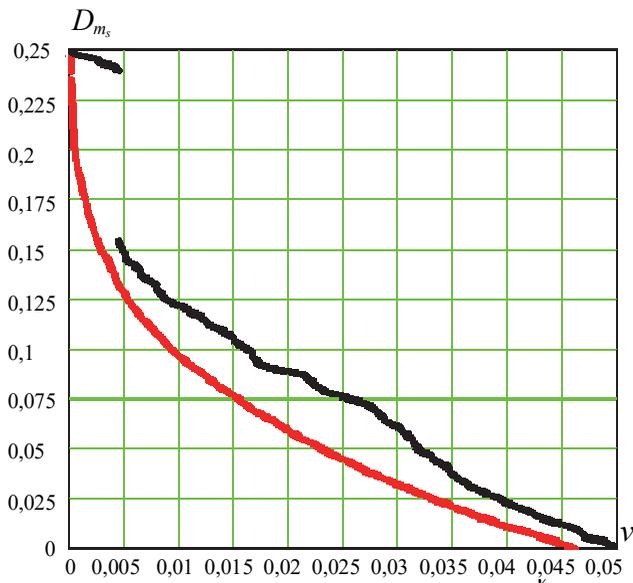


Рис. 7. Межклассовые дисперсионные последовательности изотропной и анизотропной выборок

В обоих случаях расчета дисперсий по (5) и (6) на графиках наблюдаются разрывы области допустимых значений функционалов для анизотропной выборки. Используем это свойство для формирования функционала качества классификации (7).

**10. Задача классификации.** Так многочисленные исследования (эксперименты) показали, что при наличии классов наблюдается неоднородность выборки. Эта неоднородность проявляется в виде существенных скачков межклассовой дисперсии (суммарной внутриклассовой дисперсии) разбиений, упорядоченных по их убыванию (возрастанию). Наибольший выброс идентифицирует наилучшее разбиение на классы. Значит надо искать минимум (максимум) производной какой-либо из дисперсий на множестве всех возможных разбиений. Однако производной в классическом смысле здесь не существует, так как производная слева не равна производной справа и, кроме того, функционал определен на дискретном множестве [39, 40]. По сути, это означает, что происходит качественное изменение количественной характеристики разбиения при небольшом изменении его параметров. Тогда задачу классификации можно отнести к задачам теории катастроф.

Важно отметить, что в случае изотропных выборок катастрофы нет, и она наблюдается, когда существует некоторая организация

в системе, генерирующей анизотропную выборку. Поэтому конструкция функционала (7) должна учитывать, в какой мере решение задачи классификации отлично от наиболее близкого решения при замене анизотропной выборки изотропной. Это можно сформулировать в виде следующей минимаксной задачи:

$$\min_{\mathbf{X}'_a} \left| D_{v_s}(\mathbf{X}'_i, \mathbf{H}_s) - D_{v_s}(\mathbf{X}'_a, \mathbf{Y}_r) \right| \rightarrow \max_{\mathbf{X}'_i} \quad (8)$$

где  $D_{v_s}(\mathbf{X}'_i, \mathbf{H}_s)$  – суммарная внутриклассовая дисперсия изотропной канонической выборки;  $D_{v_s}(\mathbf{X}'_a, \mathbf{Y}_r)$  – суммарная внутриклассовая дисперсия анизотропной канонической выборки;  $\mathbf{X}'_i$  – матрица назначений объектов в классы изотропной канонической выборки;  $\mathbf{X}'_a$  – матрица назначений объектов в классы анизотропной канонической выборки;  $\mathbf{H}_s$  – матрица признаков объектов изотропной канонической выборки;  $\mathbf{Y}_r$  – матрица признаков объектов анизотропной канонической выборки.

Аналогичный функционал может быть записан и для межклассовой дисперсии:

$$\min_{\mathbf{X}'_a} \left| D_{m_s}(\mathbf{X}'_i, \mathbf{H}_s) - D_{m_s}(\mathbf{X}'_a, \mathbf{Y}_r) \right| \rightarrow \max_{\mathbf{X}'_i}$$

Поскольку дисперсионная кривая изотропной выборки уникальна, она может быть табулирована или аппроксимирована с заданной точностью с помощью некоторой функции  $f_i(x)$ . Теперь задачу (8) можно записать как:

$$\min_{\mathbf{X}'_a} \left| f_i(x) - D_{v_s}(\mathbf{X}'_a, \mathbf{Y}_r) \right| \rightarrow \max_x \quad (9)$$

Тогда, математическую модель можно представить как задачу нахождения матрицы назначений  $\mathbf{X}'_a$  и вспомогательной переменной  $x$ , удовлетворяющих (10) при известных матрицах признаков объектов  $\mathbf{Y}_r$ , обобщенной метрики эвклидова пространства  $\mathbf{W}$  и эталонной дисперсионной кривой  $f_i(x)$ , при ограничениях на компоненты матрицы назначений (2), (3). Вычисление (9) производится с помощью выражения (6).

Задачи (3), (2), (6), (9) можно классифицировать как задачи целочисленного нелинейного программирования на булевом множестве решений, которое задается с помощью нелинейных уравнений. В этой задаче неявно осуществляется оптимизация по числу классов, так как на строки матрицы  $X'_a$  не накладывается никаких дополнительных ограничений, кроме булевости переменных. Это может приводить к наличию нулевых (пустых) кластеров вследствие ограничения снизу и сверху числа классов ( $1 \leq K \leq N$ ), а также вследствие предлагаемой минимаксной математической модели задачи классификации. Разработке алгоритма решения поставленной задачи будут посвящены последующие работы.

Решение задач (9), (6), (2), (3) определяет наилучшее разбиение множества объектов на классы, а по достигнутому значению функционала (9) можно судить о качестве классификации.

В целом можно выделить следующие основные этапы решения задачи классификации:

1. Определение эталонной дисперсионной кривой изотропной выборки.

2. Нормирование рабочей выборки относительно ее общей дисперсии, то есть приведение рабочей выборки к каноническому виду.

3. Проверка гипотезы о расслоении на классы, а именно: можно выделить классы или нет?

4. Решение задачи синтеза классификатора, то есть поиск оптимального разбиения в случае истинности гипотезы о расслоении на классы.

5. Верификация полученного решения и определение качества классификации.

**11. Заключение.** В рамках общей задачи теории принятия решений при байесовском подходе задача классификации признаков объектов с точки зрения разбиения множества на подмножества в условиях правильной классификации представляет собой модель системы распознавания, где функция правдоподобия определяется в ходе решения задачи, а класс правил распознавания образован на множестве известных признаков объектов, что определяет теоретическую значимость результатов работы. Тогда задача классификации сводится к задаче определения вероятностей принадлежности исследуемых объектов к некоторым классам. При этом постулируется тождественность эталонного множества и множества оценок классов.

В результате разработана математическая модель классификатора признаков, которая предполагает синтез двумерной плотности вероятностей в пространстве координат: классы – объекты. Использование обобщенных функций позволило свести вероятностную задачу поиска минимума байесовского риска к детерминированной задаче на

множестве нерандомизированных правил принятия решений при помощи специально введенных ограничений.

В подобных условиях для корректного синтеза классификатора необходима дисперсионная кривая изотропной выборки, а также использование суммарной внутриклассовой и межклассовых дисперсий для характеристики качества классификации, что предполагает интерпретацию задачи классификации как частной задачи теории катастроф. Ограниченность исходных данных предопределила поиск минимаксного функционала, отражающего качество классификации при квадратичной функции потерь. В итоге разработанную математическую модель оказалось возможным рассматривать в форме задачи целочисленного нелинейного программирования, а последующее использование полиномиальных ограничений позволило ее преобразовать к виду общей задачи нелинейного непрерывного программирования, что обуславливает практическую значимость работы, как способа решения NP – полных комбинаторных задач. В дополнение к этому найдены необходимые условия расслоения на классы, которые могут быть использованы как достаточные при проверке гипотезы о существовании классов.

Дальнейшие исследования в области разработки моделей классификаторов должны быть направлены на синтез динамических классификаторов [41], для чего необходимо:

- разделить представленный статистический материал на обучающую и контролирующую выборки;
- найти эффективный алгоритм получения решения задачи классификации в рамках представленной математической модели;
- выполнить верификацию разработанной модели классификатора и по возможности проверку его адекватности;
- предусмотреть процедуру обучения классификатора, как в признаковом пространстве объектов для снижения его размерности (отбора информативных признаков), так и в пространстве самих объектов для расслоения их на классы;
- определить, при каких условиях задачу синтеза классификатора нужно решать заново в динамике изменения признаков объектов.

### Литература

1. *Блауберг И.В.* Проблемы методологии системного исследования // М.: Мысль. 1970. 454 с.
2. *He H.* A Deep Research in Classic Classification Network // IOP Conf. Series: Materials Science and Engineering. 2020. 740 p.
3. *Wehrmann J., Cerri R., Barros R.C.* Hierarchical Multi-Label Classification Networks // Proceedings of the 35th International Conference on Machine Learning, Stockholm (PMLR 80). 2018. pp. 5075–5084.
4. *Li P., Wang D., Wang L., Lu H.* Visual Tracking by Dynamic Matching-Classification Network Switching // Pattern Recognition. 2020. pp. 107419.

5. *Айвазян С.А.* Прикладная статистика. Классификация и снижение размерности // М.: Финансы и статистика. 1989. 606 с.
6. *Pontone S., Grimaldi G.* What is the Technique without a Proper Classification? // Int J Gastroenterol Disord Ther. 2015. pp.113.
7. *Nwafor G., Onwukwe C.* On Proper Classification and Placement of Students in Nigerian University Systems Using Discriminant Analysis // American Journal of Applied Mathematics and Statistics. 2014. pp. 394–397.
8. *Мандель И.Д.* Кластерный анализ // М.: Финансы и статистика. 1988. 176 с.
9. *Галлагер Р.Дж.* Теория информации и надежная связь // М.: Сов. Радио. 1974. 719 с.
10. *Xiaotong L. et al.* Learning a Deep Vector Quantization Network for Image Compression // IEEE WESCANEX 93 Communications, Computers and Power in the Modern Environment – Conference Proceedings. 1993. pp. 299–312.
11. *Janabi Samaher A.-J., Abed S.M.* Digital Video Scenes Recognition using  $M_{jib}$ -EA and Learning Vector Quantization Network // Journal of Babylon University. vol. 9. no. 24. 2016. pp. 2362–2373.
12. *Ван Трус Г.Л.* Теория обнаружения, оценок и модуляции // М.: Сов. радио. 1972. 744 с.
13. *Батенков К.А.* Синтез детерминированных нелинейных дискретных отображений непрерывных каналов связи // Труды СПИИРАН. 2016. № 2(45). С. 75–101.
14. *Kipping D.* A Bayesian Approach to the Simulation Argument // Universe. 2020. vol. 6. no. 8. pp. 109.
15. *Mihnea A., John H.* A Bayesian Approach for Asset Allocation // International Journal of Statistics and Probability. 2020. vol. 4. pp. 1–14.
16. *Тихонов В.И.* Марковские процессы // М.: Сов. радио. 1977. 488 с.
17. *Schlegel M., White A., Patterson A., White M.* General Value Function Networks. 2018. URL: [arxiv.org/pdf/1807.06763v1.pdf](https://arxiv.org/pdf/1807.06763v1.pdf) (дата обращения: 21.10.2020).
18. *Батенков К.А.* Точные и граничные оценки вероятностей связности сетей связи на основе метода полного перебора типовых состояний // Труды СПИИРАН. 2019. Т. 18. № 5. С. 1093–1118.
19. *Taketoshi Y., Masahiro I.* Extended responsibility assignment matrix (ERAM) suitable for a cross functional project // EDULEARN17 Proceedings. 2017. pp. 4825–4834.
20. *Qingge J., Haoqiang Y.* Online Multiple Object Tracking with Reid Feature Extraction Network and Similarity Matrix Network // Journal of Physics: Conference Series, Volume 1544, 2020 5th International Conference on Intelligent Computing and Signal Processing (ICSP). 2020. pp. 20–22.
21. *Эндрюс Г.* Теория разбиений // М.: Наука. 1982. 255 с.
22. *Menchaca B. et al.* Technique for setting network communication parameters : publ. no. WO/2011/008515, publ. date 20.01.2011, int. appl. No. PCT/US2010/040298, int. fil. date 29.06.2010.
23. *Rahman, H., Sheikh, N.U., Saleh Al-Qahtani, H., Hazra, T.K.* Partitioned network with Adaptive Mobile Sinks // 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). 2019. pp. 1098–1103.
24. *Chalupa D.* Partitioning Networks into Cliques: A Randomized Heuristic Approach // Information Sciences and Technologies Bulletin of ACM Slovakia. 2014. vol. 6. pp. 1–8.
25. *Ланнэ А.А.* Нелинейные динамические системы: синтез, оптимизация, идентификация // Л.: ВАС. 1985. 240 с.
26. *Kostrikov I., Bruna J., Panozzo D., Zorin D.* Surface Networks. 2017. URL: [arxiv.org/pdf/1705.10819.pdf](https://arxiv.org/pdf/1705.10819.pdf) (дата обращения: 21.10.2020).
27. *Pinghua G., Jieping Y., Changshui Z.* Robust multi-task feature learning // KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. August 2012. pp. 895–903.

28. *Гантмахер Ф.П.* Теория матриц // М.: Физматлит. 2010. 559 с.
29. *Lee S., Song B.* Transformation of Non-Euclidean Space to Euclidean Space for Efficient Learning of Singular Vectors // IEEE Access. 2020. vol. 8. pp. 127074–127083.
30. *Vandyapadhyay S., Fomin F., Simonov K.* On Coresets for Fair Clustering in Metric and Euclidean Spaces and Their Applications. 2020. URL: [arxiv.org/pdf/2007.10137.pdf](https://arxiv.org/pdf/2007.10137.pdf) (дата обращения: 21.10.2020).
31. *Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф.* Справочник по теории вероятностей и математической статистике // М.: Наука. 1985. 640 с.
32. *Borзов А.В., Labunets L.V., Steshenko V.B.* Noncanonical Spectral Model of Multidimensional Uniform Random Fields // J. Comput. Syst. Sci. Int. 2017. vol. 57. pp. 874–889.
33. *Федоров В.В.* Численные методы максимина // М.: Наука. 1979. 278 с.
34. *Кловский Д.Д., Конторович В.Я., Широков С.М.* Статистическая теория связи // М.: Связь. 1974. 247 с.
35. *Kreher D.L.* Combinatorial algorithms : Generation, enumeration, a. search // CRC press, Cop. 1999. 329 p.
36. *Батенков А.А., Батенков К.А., Фокин А.Б.* Методы формирования множеств состояний телекоммуникационных сетей для различных мер связности // Труды СПИИРАН. 2020. № 3 (19). С. 644–673.
37. *Химмельблау Д.* Прикладное нелинейное программирование // М.: Мир. 1975. 534 с.
38. *Яблонский С.В.* Введение в дискретную математику : учебное пособие для студентов вузов, обучающихся по специальности «Прикладная математика» // М.: Высш. шк. 2006. 384 с.
39. *Батенков К.А.* Числовые характеристики структур сетей связи // Труды СПИИРАН. 2017. № 4(53). С. 5–28.
40. *Батенков А.А., Батенков К.А.* Анализ и синтез структур сетей связи по детерминированным показателям устойчивости // Труды СПИИРАН. 2018. № 3(58). С. 128–159.
41. *Nooka S.P. et al.* Adaptive hierarchical classification networks // 2016 23rd International Conference on Pattern Recognition (ICPR). 2016. pp. 3578–3583.

**Батенков Александр Александрович** – д-р техн. наук, профессор, профессор кафедры, кафедра электроники, радиотехники и систем связи, Орловский государственный университет имени И.С. Тургенева (ОГУ). Область научных интересов: статистическая теория связи, модели и методы обработки сигналов, вычислительные алгоритмы. Число научных публикаций – 110. [pustur@yandex.ru](mailto:pustur@yandex.ru); ул. Комсомольская, 95, 302026, Орел, Россия; р.т.: +7-906-570-16-66.

**Батенков Кирилл Александрович** – д-р техн. наук, доцент, сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: статистическая теория связи, модели и методы обработки сигналов, анализ качества сетей связи. Число научных публикаций – 152. [pustur@yandex.ru](mailto:pustur@yandex.ru); ул. Приборостроительная, 35, 302015, Орел, Россия; р.т.: 4862-54-97-63.

**Богачёв Андрей Геннадьевич** – канд. техн. наук, сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: модели и методы обработки сигналов. Число научных публикаций – 26. [pustur@yandex.ru](mailto:pustur@yandex.ru); ул. Приборостроительная, 35, 302015, Орел, Россия; р.т.: 4862-54-97-63.

**Мишин Владислав Владимирович** – канд. техн. наук, доцент, заведующий кафедрой, кафедра электроники, радиотехники и систем связи, Орловский государственный университет имени И.С. Тургенева (ОГУ). Область научных интересов: вычислительные алгоритмы. Число научных публикаций – 106. [pustur@yandex.ru](mailto:pustur@yandex.ru); ул. Комсомольская, 95, 302026, Орел, Россия; р.т.: +7-906- 570-16-66.

## ПРИЛОЖЕНИЕ 1.

Вывод выражения для апостериорной плотности вероятности принадлежности  $n$ -ого объекта  $k$ -ому классу:

$$\begin{aligned}
 W(\mathbf{Z} / \mathbf{Y}) &= \frac{\sum_{k=1}^K \sum_{n=1}^N x_{k,n} \delta(\mathbf{Z} - \mathbf{Z}_k) \delta(\mathbf{Y} - \mathbf{Y}_n)}{\sum_{n=1}^N \sum_{k=1}^K x_{k,n} \delta(\mathbf{Y} - \mathbf{Y}_n)} = \\
 &= \frac{\sum_{k=1}^K \delta(\mathbf{Z} - \mathbf{Z}_k) \sum_{n=1}^N x_{k,n} \delta(\mathbf{Y} - \mathbf{Y}_n)}{\sum_{n=1}^N \delta(\mathbf{Y} - \mathbf{Y}_n) \sum_{k=1}^K x_{k,n}} = \\
 &= \sum_{k=1}^K \delta(\mathbf{Z} - \mathbf{Z}_k) \sum_{n=1}^N x_{k,n} \frac{\delta(\mathbf{Y} - \mathbf{Y}_n)}{\sum_{n=1}^N \delta(\mathbf{Y} - \mathbf{Y}_n) \sum_{k=1}^K x_{k,n}} = \\
 &= \sum_{k=1}^K \delta(\mathbf{Z} - \mathbf{Z}_k) \sum_{n=1}^N x_{k,n} \frac{\delta(\mathbf{Y} - \mathbf{Y}_n)}{\sum_{n=1}^N \delta(\mathbf{Y} - \mathbf{Y}_n) \sum_{k=1}^K x_{k,n}} = \\
 &= \sum_{k=1}^K x_{k,n} \frac{\delta(\mathbf{Z} - \mathbf{Z}_k) 1(\mathbf{Y} - \mathbf{Y}_n)}{\sum_{k=1}^K x_{k,n}}, \\
 W(\mathbf{Z}) &= \sum_{k=1}^K \delta(\mathbf{Z} - \mathbf{Z}_k) \sum_{n=1}^N x_{k,n}.
 \end{aligned}$$

## ПРИЛОЖЕНИЕ 2.

Свойства специальной разрывной функции  $1(\mathbf{Z})$ . По сути, эта функция определяет отношение

$$1(\mathbf{Z}) = \frac{\delta(\mathbf{Z} - \mathbf{Z}_i)}{\delta(\mathbf{Z})} \text{ при } \mathbf{Z}_i \rightarrow \mathbf{Z};$$

и поэтому так же является обобщенной [16, 17] функцией.

В качестве приближения  $1(\alpha, \mathbf{Z})$ -функции к идеальной предлагается использовать экспоненциальную функцию вида:

$$\varepsilon(\alpha, \mathbf{Z}) = e^{-\alpha(\mathbf{z}^T \mathbf{Z})}, \alpha \rightarrow \infty,$$

где  $\alpha$  – параметр настройки аппроксимации разрывной функции непрерывной дифференцируемой функцией.

На рисунке П2 изображена  $\varepsilon(\alpha, x)$  при различных значениях  $\alpha$  :

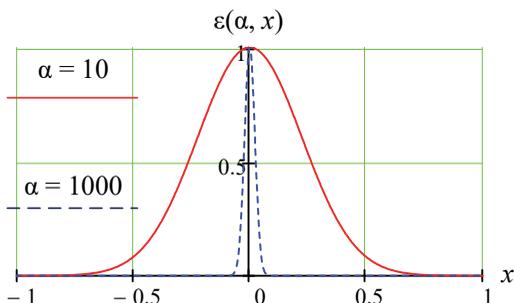


Рис. П2. График функции  $\varepsilon(\alpha, x)$  при различных значениях  $\alpha$

### ПРИЛОЖЕНИЕ 3.

**Условия нормировки для введенных плотностей.** Целесообразно их записать в виде:

$$\begin{aligned} \int_{\Omega_{\mathbf{Z}}} \int_{\Omega_{\mathbf{Y}}} W(\mathbf{Z}, \mathbf{Y}) d\mathbf{Z} d\mathbf{Y} &= \int_{\Omega_{\mathbf{Z}}} \int_{\Omega_{\mathbf{Y}}} \sum_{k=1}^K \sum_{n=1}^N x_{k,n} \delta(\mathbf{Z} - \mathbf{Z}_k) \delta(\mathbf{Y} - \mathbf{Y}_n) d\mathbf{Z} d\mathbf{Y} = \\ &= \sum_{k=1}^K \sum_{n=1}^N x_{k,n} = 1; \\ \int_{\Omega_{\mathbf{Z}}} W(\mathbf{Z} / \mathbf{Y}) d\mathbf{Z} &= \int_{\Omega_{\mathbf{Z}}} \sum_{k=1}^K \frac{x_{k,n}}{\sum_{s=1}^K x_{s,n}} \delta(\mathbf{Z} - \mathbf{Z}_k) 1(\mathbf{Y} - \mathbf{Y}_n) d\mathbf{Z} = \\ &= \sum_{k=1}^K \frac{x_{k,n}}{\sum_{s=1}^K x_{s,n}} 1(\mathbf{Y} - \mathbf{Y}_n) = 1(\mathbf{Y} - \mathbf{Y}_n), \forall n = \overline{1, N}; \end{aligned}$$

$$\int_{\Omega_Z} W(\mathbf{Z}) d\mathbf{Z} = \int_{\Omega_Z} \sum_{k=1}^K \sum_{n=1}^N x_{k,n} \delta(\mathbf{Z} - \mathbf{Z}_k) d\mathbf{Z} = 1;$$

$$\int_{\Omega_Y} W(\mathbf{Y} / \mathbf{Z}) d\mathbf{Y} = \int_{\Omega_Y} \sum_{n=1}^N \frac{x_{k,n}}{\sum_{s=1}^N x_{k,s}} 1(\mathbf{Z} - \mathbf{Z}_k) \delta(\mathbf{Y} - \mathbf{Y}_n) d\mathbf{Y} =$$

$$= \sum_{n=1}^N \frac{x_{k,n}}{\sum_{s=1}^N x_{k,s}} 1(\mathbf{Z} - \mathbf{Z}_k) = 1(\mathbf{Z} - \mathbf{Z}_k), \forall k = \overline{1, K};$$

$$\int_{\Omega_Y} W(\mathbf{Y}) d\mathbf{Y} = \int_{\Omega_Y} \sum_{n=1}^N \sum_{k=1}^K x_{k,n} \delta(\mathbf{Y} - \mathbf{Y}_n) d\mathbf{Y} = 1,$$

где  $\Omega_X$  – область определения переменных  $\mathbf{X}$  (скаляра, вектора матрицы и т. п.)

#### ПРИЛОЖЕНИЕ 4.

**Необходимые свойства разбиений.** Разбиение  $Q_n$  множества  $\{1, \dots, n\}$  однозначно определяет разбиение  $Q_{n-1}$  множества  $\{1, \dots, n-1\}$ , которое получается из  $Q_n$  после удаления элемента  $n$  (или пустого блока) из соответствующего блока. Также, если имеется разбиение  $W = \{B_1, \dots, B_k\}$  множества  $\{1, \dots, n-1\}$ , то можно отыскать все разбиения  $Q_n$  множества  $\{1, \dots, n\}$ , для которых  $Q_{n-1} = W$ , то есть такие разбиения:

$$B_1 \cup \{n\}, \dots, B_k; \dots, B_1, \dots, B_k \cup \{n\}; B_1, \dots, B_k, \{n\}.$$

#### ПРИЛОЖЕНИЕ 5.

**Определение непрерывного многомерного равномерного распределения [31].** Пусть  $S \subset R^n$  – борелевское множество с конечной лебеговой мерой  $\lambda(S)$ ; говорят, что вектор  $(\xi_1, \dots, \xi_n)$  имеет равномерное распределение в области  $S$ , если плотность совместного распределения  $f_{(\xi_1, \dots, \xi_n)}(x_1, \dots, x_n)$  постоянна в области  $S$  и равна нулю вне этой области:

$$f_{(\xi_1, \dots, \xi_n)}(x_1, \dots, x_n) = \begin{cases} \frac{1}{\lambda(S)}, & (x_1, \dots, x_n) \in S, \\ 0, & (x_1, \dots, x_n) \notin S. \end{cases}$$

A. BATENKOV, K. BATENKOV, A. BOGACHYOV, V. MISHIN  
**MATHEMATICAL MODEL OF OBJECT CLASSIFIER BASED ON  
BAYESIAN APPROACH**

*Batenkov A., Batenkov K., Bogachyov A., Mishin V. Mathematical Model of Object Classifier based on Bayesian Approach.*

**Abstract.** The paper claims that the primary importance in solving the classification problem is to find the conditions for dividing the General complexity into classes, determine the quality of such a bundle, and verify the classifier model. We consider a mathematical model of a non-randomized classifier of features obtained without a teacher, when the number of classes is not set a priori, but only its upper bound is set. The mathematical model is presented in the form of a statement of a minimax conditional extreme task, and it is a problem of searching for the matrix of belonging of objects to a class, and representative (reference) elements within each class. The development of the feature classifier is based on the synthesis of two-dimensional probability density in the coordinate space: classes-objects. Using generalized functions, the probabilistic problem of finding the minimum Bayesian risk is reduced to a deterministic problem on a set of non-randomized classifiers. At the same time, the use of specially introduced constraints fixes non-randomized decision rules and plunges the integer problem of nonlinear programming into a General continuous nonlinear problem. For correct synthesis of the classifier, the dispersion curve of the isotropic sample is necessary. It is necessary to use the total intra-class and inter-class variance to characterize the quality of classification. The classification problem can be interpreted as a particular problem of the theory of catastrophes. Under the conditions of limited initial data, a minimax functional was found that reflects the quality of classification for a quadratic loss function. The developed mathematical model is classified as an integer nonlinear programming problem. The model is given using polynomial constraints to the form of a General problem of nonlinear continuous programming. The necessary conditions for the bundle into classes are found. These conditions can be used as sufficient when testing the hypothesis about the existence of classes.

**Keywords:** Non-randomized Feature Classifier, Upper Bound of the Number of Classes, Minimax, Conditional Extreme Problem, Integer Problem of Nonlinear Programming

**Batenkov Aleksandr** – Ph.D., Dr.Sci., Professor, Professor, Department of Electronics, Radio Engineering and Communication Systems, Orel State University named after I.S. Turgenov (OSU). Research interests: statistical communication theory, models and methods of signal processing, computational algorithms. The number of publications – 110. [pustur@yandex.ru](mailto:pustur@yandex.ru); 95, Komsomolskaya str., 302026, Orel, Russia; office phone: +7-906-570-16-66.

**Batenkov Kirill** – Ph.D., Dr.Sci., Associate Professor, Employee, Academy of Federal Guard Service. Research interests: statistical communication theory, models and methods of signal processing, communications network quality analysis. The number of publications – 152. [pustur@yandex.ru](mailto:pustur@yandex.ru); 35, Priborostroitelnaya str., 302015, Orel, Russia; office phone: 4862-54-7-63.

**Bogachev Andrey** – Ph.D., Employee, Academy of Federal Guard Service. Research interests: models and methods of signal processing. The number of publications – 26. [pustur@yandex.ru](mailto:pustur@yandex.ru); 35, Priborostroitelnaya str., 302015, Orel, Russia; office phone: 4862-54-7-63.

**Mishin Vladislav** – Ph.D., Associate Professor, Head of a Department, Department of Electronics, Radio Engineering and Communication Systems, Orel State University named after I.S. Turgenov (OSU). Research interests: computational algorithms. The number of publications – 106. [pustur@yandex.ru](mailto:pustur@yandex.ru); 95, Komsomolskaya str., 302026, Orel, Russia; office phone: +7-906- 570-16-66.

**References**

1. Blauberg I.V. *Problemy metodologii sistemnogo issledovaniya* [Problems of system research methodology]. M.: Mysl. 1970. 454 p. (In Russ.).
2. He H. A Deep Research in Classic Classification Network. IOP Conf. Series: Materials Science and Engineering. 2020. 740 p.
3. Wehrmann J., Cerri R., Barros R.C. Hierarchical Multi-Label Classification Networks. Proceedings of the 35th International Conference on Machine Learning (PMLR 80). 2018. pp. 5075–5084.
4. Li P., Wang D., Wang L., Lu H. Visual Tracking by Dynamic Matching-Classification Network Switching. Pattern Recognition. 2020. pp. 107419.
5. Ayvazyan S.A., Bukhshtaber V.M., Enyukov I.S., Meshalkin L.D. *Prikladnaya statistika. Klassifikaciya isnizhenie razmernosti :Sprav. izd.* [Applied statistics. Classification and reduction of dimension: Reference ed.]. M.: Finansy i statistika. 1989. 606 p. (In Russ.).
6. Pontone S., Grimaldi G. What is the Technique without a Proper Classification?. *Int J Gastroenterol Disord Ther.* 2015. pp. 113.
7. Nwafor G., Onwukwe C. On Proper Classification and Placement of Students in Nigerian University Systems Using Discriminant Analysis. *American Journal of Applied Mathematics and Statistics.* 2014. pp. 394–397.
8. Mandel I.D. *Klasternyy analiz* [Cluster analysis]. M.: Finansy i statistika. 1988. 176 p. (In Russ.).
9. Gallager R.G. Information Theory and Reliable Communication. Springer. 1968. (Russ. ed.: Gallager R.G. Teoriya informacii i nadezhnaya svyaz. M.: Sov. Radio. 1974. 719 p.).
10. Xiaotong L. et al. Learning a Deep Vector Quantization Network for Image Compression. IEEE WESCANEX 93 Communications, Computers and Power in the Modern Environment – Conference Proceedings. 1993. pp. 299–312.
11. Janabi Samaher A.-J., Abed S.M. Digital Video Scenes Recognition using  $M_{jiv}$ -EA and Learning Vector Quantization Network. *Journal of Babylon University.* vol. 9. no. 24. 2016. pp. 2362–2373.
12. Van Trees H.L. Detection, estimation, and linear modulation theory, Hoboken, New Jersey: John Wiley & Sons, Inc. 1968. (Russ. ed.: Van Trees H.L. Teoriya obnaruzheniya, ocenok i modulyacii. M.: Sov. radio. 1972. 744 p.).
13. Batenkov K.A. [Synthesis of deterministic nonlinear discrete mappings of continuous communication channels]. *Trudy SPIIRAN – Proceedings of SPIIRAS.* 2016. vol. 2(45). pp. 75–101 (In Russ.).
14. Kipping D. A Bayesian Approach to the Simulation Argument. *Universe.* 2020. vol. 6. no. 8. pp. 109.
15. Mihnea A., John H. A Bayesian Approach for Asset Allocation. *International Journal of Statistics and Probability.* 2020. vol. 4. pp. 1–14.
16. Tihonov V.I., *Markovskie process* [Markov processes]. M.: Sov. radio. 1977. 488 p. (In Russ.).
17. Schlegel M., White A., Patterson A., White M. General Value Function Networks. 2018. Available at: [arxiv.org/pdf/1807.06763v1.pdf](https://arxiv.org/pdf/1807.06763v1.pdf) (accessed: 21.10.2020).
18. Batenkov K. A. [Accurate and boundary estimates of the connectivity probabilities of communication networks based on the method of full iteration of typical States]. *Trudy SPIIRAN – Proceedings of SPIIRAS.* 2019. Issue 18. vol. 5. pp. 1093–1118 (In Russ.).
19. Taketoshi Y., Masahiro I. Extended responsibility assignment matrix (ERAM) suitable for a cross functional project. EDULEARN17 Proceedings. 2017. pp. 4825–4834.
20. Qingge J., Haoqiang Y. Online Multiple Object Tracking with Reid Feature Extraction Network and Similarity Matrix Network. *Journal of Physics: Conference Series, Volume 1544, 2020 5th International Conference on Intelligent Computing and Signal Processing (ICSP).* 2020. pp. 20–22.

21. Endryus G. *Teoriya razbieniij* [Theory of partitions]. M.: Nauka. 1982. 255 p.
22. Menchaca B. et al. Technique for setting network communication parameters : publ. no. WO/2011/008515, publ. date 20.01.2011, int. appl. No. PCT/US2010/040298, int. fil. date 29.06.2010.
23. Rahman, H., Sheikh, N.U., Saleh Al-Qahtani, H., Hazra, T.K. Partitioned network with Adaptive Mobile Sinks. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). 2019. pp. 1098–1103.
24. Chalupa D. Partitioning Networks into Cliques: A Randomized Heuristic Approach. *Information Sciences and Technologies Bulletin of ACM Slovakia*. 2014. vol. 6. pp. 1–8.
25. Lanne A.A. *Nelinejnye dinamesicheskie sistemy: sintez, optimizaciya, identifikaciya* [Nonlinear dynamical systems: synthesis, optimization, identification]. L.: VAS. 1985. 240 p. (In Russ.).
26. Kostrikov I., Bruna J., Panozzo D., Zorin D. Surface Networks. 2017. Available at: [arxiv.org/pdf/1705.10819.pdf](https://arxiv.org/pdf/1705.10819.pdf) (accessed: 21.10.2020).
27. Pinghua G., Jieping Y., Changshui Z. Robust multi-task feature learning. KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. August 2012. pp. 895–903.
28. Gantmaher F.R. *Teoriya matric* [Matrix Theory]. M.: Fizmatlit. 2010. 559 p. (In Russ.).
29. Lee S., Song B. Transformation of Non-Euclidean Space to Euclidean Space for Efficient Learning of Singular Vectors. *IEEE Access*. 2020. vol. 8. pp. 127074–127083.
30. Bandyapadhyay S., Fomin F., Simonov K. On Coresets for Fair Clustering in Metric and Euclidean Spaces and Their Applications. 2020. Available at: [arxiv.org/pdf/2007.10137.pdf](https://arxiv.org/pdf/2007.10137.pdf) (accessed: 21.10.2020).
31. Korolyuk V.S., Portenko N.I., Skorohod A.V., Turbin A.F. *Spravochnik po teorii veroyatnostej i matematicheskoj statistike* [Handbook of probability theory and mathematical statistics]. M.: Nauka. 1985. 640 p. (In Russ.).
32. Borzov A.B., Labunets L.V., Steshenko V.B. Noncanonical Spectral Model of Multidimensional Uniform Random Fields. *J. Comput. Syst. Sci. Int.* 2017. vol. 57. pp. 874–889.
33. Fedorov V.V. *Chislennyye metody maksimina* [Numerical methods of Maximin]. M.: Nauka. 1979. 278 p. (In Russ.).
34. Klovsikij D.D., Kontorovich V.Ya., Shirokov S.M. *Modeli nepreryvnyh kanalov svyazi na osnove stohasticheskikh differencial'nyh uravnenij* [Models of continuous communication channels based on stochastic differential equations]. M.: Svyaz. 1984. 247 p. (In Russ.).
35. Kreher D.L. Combinatorial algorithms: Generation, enumeration, a. search. CRC press, Cop. 1999. 329 p.
36. Batenkov K.A., Batenkov A.A., Fokin A.B. [Methods for generating sets of telecommunication network States for various connectivity measures]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2020. Issue 3(19). pp. 644–673 (In Russ.).
37. Himmelblau D. *Prikladnoe nelinejnoe programmirovaniye* [Applied nonlinear programming]. M.: Mir. 1975. 534 p. (In Russ.).
38. Yablonskij S.V. *Vvedenie v diskretnuy umatematiku :uchebnoe posobie dlyi astudentov vuzov, obuchayushchihsiya po special'nosti "Prikladnaya matematika"* [Introduction to discrete mathematics: a textbook for University students studying in the specialty "Applied mathematics"]. M.: Vyssh. shk. 2006. 384 p. (In Russ.).
39. Batenkov K.A. [Numerical characteristics of the structures of communication networks]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2017. vol. 4(53). pp. 5–28 (In Russ.).
40. Batenkov K.A., Batenkov A.A. [Analysis and synthesis of communication network structures based on deterministic stability indicators]. *Trudy SPIIRAN – Proceedings of SPIIRAS*. 2018. vol. 3(58). pp. 128–159 (In Russ.).
41. Nooka S.P. et al. Adaptive hierarchical classification networks. 2016 23rd International Conference on Pattern Recognition (ICPR). 2016. pp. 3578–3583.

O. GERMAN, S. NASR

**NEW METHOD FOR OPTIMAL FEATURE SET REDUCTION***German O., Nasr S. New Method for Optimal Feature Set Reduction.*

**Abstract** A problem of searching a minimum-size feature set to use in distribution of multidimensional objects in classes, for instance with the help of classifying trees, is considered. It has an important value in developing high speed and accuracy classifying systems. A short comparative review of existing approaches is given. Formally, the problem is formulated as finding a minimum-size (minimum weighted sum) covering set of discriminating 0,1-matrix, which is used to represent capabilities of the features to distinguish between each pair of objects belonging to different classes. There is given a way to build a discriminating 0,1-matrix. On the basis of the common solving principle, called the group resolution principle, the following problems are formulated and solved: finding an exact minimum-size feature set; finding a feature set with minimum total weight among all the minimum-size feature sets (the feature weights may be defined by the known methods, e.g. the RELIEF method and its modifications); finding an optimal feature set with respect to fuzzy data and discriminating matrix elements belonging to diapason [0,1]; finding statistically optimal solution especially in the case of big data. Statistically optimal algorithm makes it possible to restrict computational time by a polynomial of the problem sizes and density of units in discriminating matrix and provides a probability of finding an exact solution close to 1.

Thus, the paper suggests a common approach to finding a minimum-size feature set with peculiarities in problem formulation, which differs it from the known approaches. The paper contains a lot of illustrations for clarification aims. Some theoretical statements given in the paper are based on the previously published works.

In the concluding part, the results of the experiments are presented, as well as the information on dimensionality reduction for the coverage problem for big datasets. Some promising directions of the outlined approach are noted, including working with incomplete and categorical data, integrating the control model into the data classification system.

**Keywords:** Multidimensional Data, Classification, Feature Selection, Minimum-size Covering Problem, Group Resolution Principle

**1. Introduction.** One of important applied problems in data mining, control and system analysis is reduction of the feature set used in a model (e.g. classification or recognition ones). This problem attracts serious attention [1-5]. There are three common groups (and their combinations) of methods to realize feature set reduction including filtering, wrapper, and embedded methods. They give different results from the viewpoint of accuracy and computational complexity.

Filtering methods [6, 7] are computationally effective but do not provide (in general) classification and prognostic accuracy of the model, because, for instance, they do not take into account (in general) the internal links between features, e.g. multigroup co-relation coefficients and dependencies.

The main idea of the filtering methods is to estimate feature ratings (weights)  $W$  and use some threshold to remove features with small

ratings. A good and widely known filtering method is RELIEF [8]. In this method, for each sample object  $Ob_x$  and each feature  $f_y$  one defines the nearest object  $Ob_1$  from the same class (say,  $A$ ) and the nearest object  $Ob_2$  from the opposite class (say,  $B$ ) for two-classes classification problem. Obviously, feature  $f_y$  differs between  $A$  and  $B$  quite well if its value for class  $A$  is clearly greater than its value for class  $B$ . This observation constitutes the idea of RELIEF method which uses iterative process to re-evaluate feature  $f_y$  weight  $W[f_y]$  by adding to it the value of  $diff(f_y, Ob_x, Ob_2)/m$  and subtracting the value of  $diff(f_y, Ob_x, Ob_1)/m$ , where  $diff(\dots)$  stands for the distance between two objects with respect to feature  $f_y$  and  $m$  denotes the number of pairs  $(Ob_x, Ob_1), ((Ob_x, Ob_2))$ .

Besides RELIEF and its modifications, one can mention principal component analysis [9], supporting vector machine and other feature scoring methods [6, 10].

The other group of methods is united under the title *wrapper* methods [11, 12]. They estimate quality of the feature set  $\{f_1, f_2, \dots, f_z\}$  for instance by learning neural network with the inputs  $f_1, f_2, \dots, f_z$  and providing the following assesment of the resulting classification accuracy. This technique is extensively consuming computational resources and cannot be recommended for big feature sets as it requires to consider  $O(2^d)$  feature subsets to reveal the smallest one with satisfying classification capabilities (where  $d$  stands for the total number of the features).

To smooth drawbacks of the above two groups of methods the *embedded* methods were suggested [13]. A good example is C4.5/CART [14, 15] methods based on information gain calculations in clasification trees. Suppose that each of the objects belongs either to class  $A$  or to class  $B$  (but only to one of them). Next suppose that one selects attribute (feature)  $f_y$  to split all the objects accordingly to value of  $f_y$ . To simplify our considerations, admit then that  $f_y \in \{0, 1\}$ . Now divide all the objects into two subsets:  $f_{y0}$  where  $f_y = 0$ , and  $f_{y1}$  where  $f_y = 1$ . In general, the representatives of initial classes  $A$  and  $B$  may be among samples in  $f_{y0}$  and  $f_{y1}$ . In GINI-index based method [16] (which lies in the basis of CART – classification and recognition tree) a quality of the splitting is associated with the score value

$$h_i = 1 - \sum_{j \in A, B} \left( \frac{n_j}{|f_{yi}|} \right)^2$$

computed for each subset  $f_{yi}$  where  $n_A(n_B)$  stands

for the number of objects of class  $A$  (class  $B$ ) in  $f_{y_i}$  and  $|f_{y_i}|$  denotes cardinality of  $f_{y_i}$ . It is accepted that the best selection corresponds to attribute  $f_y$  minimizing average value of  $h_0(h_1)$ .

One can then split  $f_{y_0}$  and  $f_{y_1}$  in the similar way with the help of another attribute, say  $f_z$ , and so on in order to build a classifying tree with the attributes located in its nodes. Clearly, this heuristical approach delivers in general not optimal solution, that is, the set of attributes defined accordingly to the method may not be a minimum-size one. The main advantage of this embedded method is that it has good computational characteristics and results in the ready-to-use classifying tree.

However, there is no common platform in the above groups of methods regarding possible peculiarities in problem specification. These peculiarities are linked, for instance, to data fuzziness, weightedness, possible redundancy, incompleteness, quantitative nature, big sizes, etc.

The goal of the paper is to propose such a platform. It uses a technique to solve a 0,1–matrix covering problem on the basis of some common principle (called group resolution principle – *GRP*) applicable to the different problem specifications including (among the others):

- finding an exact minimum-size feature set;
- finding a feature set with minimum total weight among all the minimum-size feature sets;
- finding an optimal feature set with respect to fuzzy data;
- finding statistically optimal solution especially in the case of big data.

Among the advantages of this platform are also eliminating feature redundancy problem and possibility to deal with qualitative data. This may serve a good starting position for future investigations in the marked areas.

**2. Discrimination Matrix.** Suppose, a normalized data set is given (Table 1). There are 2 classes ( $A$  and  $B$ ), 6 features  $\{f_1, f_2, \dots, f_6\}$  and 8 objects  $\{i_1, i_2, \dots, i_8\}$ . Our task is to define a minimum-size feature set and build a classifying model, for instance, in the form of a classifying tree or a neural network.

Let us give a general formal statement of the problem. Denote by  $D_i = \langle f_{i1}, f_{i2}, \dots, f_{iK} \rangle$ ,  $C_i$  the  $i$ th row ( $i = 1, N$ ) of the data set with the corresponding feature values  $f_{ip}$  (quantitative data,  $p = 1, K$ ), and  $C_i$  standing for the class label. Without loss of generality we shall use two classes. Denote by  $Ob_i = \langle f_{i1}, f_{i2}, \dots, f_{iK} \rangle$  the  $i$ th object (in vector form) of the data set. Let us require the following conditions to be true:

$$\exists r \exists s (C_r \neq C_s), \quad 1 \leq r, s \leq N, \quad (1)$$

$$\forall r \forall s (C_r \neq C_s) \rightarrow Ob_r \neq Ob_s. \quad (2)$$

Table 1. The normalized data set

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	Class
$i_1$	0.8	0.5	0.0	0.77	0.33	0.33	$A$
$i_2$	1.0	0.5	0.5	1.0	0.0	0.0	$A$
$i_3$	0.4	0.25	0.5	1.0	0.0	0.0	$B$
$i_4$	0.2	0.0	0.75	0.0	0.0	0.66	$B$
$i_5$	0.7	1.0	0.75	0.44	1.0	1.0	$A$
$i_6$	0.0	1.0	1.0	0.44	1.0	0.83	$B$
$i_7$	0.0	0.5	1.0	0.33	0.66	0.83	$A$
$i_8$	0.4	1.0	0.75	0.44	0.33	0.66	$B$

Let  $\pi \subseteq \{1, \dots, K\}$  be some set of the unique integer indices and  $f(\pi)$  stand for the (sub)set of the features with indices from  $\pi$ . Let  $Ob_r(\pi)$  denote new vector obtained from  $Ob_r$  on the features  $f(\pi)$ . The feature set minimization problem is stated as to find  $\pi$  with minimum power  $|\pi|$  providing:

$$\forall r \forall s (C_r \neq C_s) \rightarrow Ob_r(\pi) \neq Ob_s(\pi). \quad (3)$$

Clearly, condition (3) warrants that one can build a classifying tree ( $CT$ ) for the data set using the features from the covering set  $\pi$ . We, however, omit the question about the sizes of  $CT$  leaving it for experiments.

Consider an arbitrary column in the normalized data Table (e.g. column  $f_1$ ).

*Definition 1.* Feature  $f_t$  discriminates between two objects  $x \in A$  and  $y \in B$  if and only if  $f_{xt} \neq f_{yt}$ .

We shall also use another notation  $(i_x, f_t)$  instead of  $f_{xt}$ . For example,  $f_1$  discriminates between  $(i_2, f_1)$  and  $(i_3, f_1)$  and does not discriminate between  $(i_6, f_1)$  and  $(i_7, f_1)$ .

*Notice 1.* Discrimination between the same classes is not considered as it has no sense.

*Notice 2.* As features may be incorrectly defined due to different reasons, one may use level  $\Delta > 0$  of discrimination ( $\Delta$  has rather small positive value) and consider that feature  $f_t$  discriminates between two samples  $x \in A$  and  $y \in B$  if and only if  $f_{xt} \geq (f_{yt} + \Delta)$  or  $f_{xt} \leq (f_{yt} - \Delta)$ .

Now one is in position to build a discrimination matrix  $DM$  with elements  $dm_{kij} = 1$  if and only if feature  $f_k$  discriminates between samples  $i$  and  $j$ ; otherwise  $dm_{kij} = 0$  (take into account *Notice 1*)

This matrix is given by Table 2. The columns containing only «1s» or «0s» are deleted as redundant.

Table 2. Reduced matrix  $DM$

	1,8	2,3	2,4	4,5	5,6	5,8	6,7
$f_1$	1	1	1	1	1	1	
$f_2$	1	1	1	1			1
$f_3$	1		1		1		
$f_4$	1		1	1			1
$f_5$				1		1	1
$f_6$	1		1	1	1	1	

The rows correspond to the features. Each column is represented by pair  $(i, j)$  with  $i$  and  $j$  specifying rows in Table 1. For instance, consider row  $f_2$  and column (5,6) with «0» (empty value) at the intersection. This situation means that feature  $f_2$  does not discriminate between rows 5 and 6 in Table 1.

*Definition 2.* Row  $a$  covers column  $b$  if  $a$  contains «1» in column  $b$ .

*Definition 3.* A set of rows  $\pi = \{a_1, a_2, \dots, a_r\}$  is a covering one for  $DM$  if for each column  $b$  from  $DM$  there is at least one row from  $\pi$  that covers  $b$ .

We are interested in a minimum-size covering set  $\pi$ . In our example, one of the minimum-size covering sets is  $\pi = \{f_1, f_2\}$ . So, instead of 6 features it is sufficient to use only 2. Basing on the found feature set, one can build a  $CT$  with the Python script given in Appendix.

One can prove then the next

*Proposition 1.* Any minimum-size cover  $\pi$  of the matrix  $DM$  defines the corresponding minimum-size feature set.

*Proof.* Any proper subset  $\pi^*$  of  $\pi$  ( $\pi^* \subset \pi$ ) does not cover some column in  $DM$ . Let this column be  $(i, j)$  and let row  $i$  belong to class  $A$  and row  $j$  belong to class  $B$ . The values of the features from  $\pi^*$  are the same both in objects  $i$  and  $j$ . By this, it is impossible to uniquely define by means of features  $\pi^*$  to which classes  $i$  and  $j$  belong.

It is a well-known fact that finding a minimum-size covering set of 0,1-matrix is NP-complete problem. We shall consider some exact method for its solution with polynomial efficiency in average. This method uses a group resolution principle suggested and substantiated in [17, 18].

**3. Group Resolution Principle (Method).** Group resolution method enables one to find a minimum-size covering set of a 0,1-matrix  $B$ . It represents an iterative process, with unique covering sets  $\pi_i$  found at each iteration by means of some heuristic technique. The following heuristic method may be used: at each iteration  $q$  find a column (amidst those remained undeleted) with minimum number of 1s. Let this column be  $r_q$ . Call  $r_q$  a *syndromic* column. Then find a row  $f_q$  (amidst undeleted rows), covering  $r_q$ , with maximum number of units. Call the unit element («1») at the intersection of row  $f_q$  and column  $r_q$  a *syndromic* element. Include  $f_q$  into a covering set formed at the iteration  $q$ . Delete then all the rows containing 1s in the column  $r_q$ . Also delete all the columns covered by row  $f_q$ . The  $q$ th iteration continues till there remains at least one undeleted column.

For each  $\pi_i$ , a special (*syndromic*) matrix is being built. From that matrix, one finds a new column-resolvent and adds it to  $B$  to contract the search area. The process repeats till an empty resolvent is generated. It is warranted that sooner or later a totally zero resolvent will be generated what indicates to finishing of the *GRP*. The best solution found to this moment represents a minimum-size cover.

To explain the details of the group resolution method, let us consider an example of some *DM* (Table 3a, excluding columns  $w$  and  $res_1$ ).

Table 3. Example of 0,1-matrix (a) and syndromic submatrix (b)

a)													b)			
	w	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	res1	c1	c2	c4	c10
$f_1$	4	1		1					1				1			
$f_2$	2				1	1	1			1					1	
$f_3$	5			1		1	1				1					1
$f_4$	6		1				1		1					1		
$f_5$	2	1		1				1		1			1			
$f_6$	3		1		1			1		1	1	1		1	1	1

Thus, select column  $c_1$  and covering it row  $f_5$  include into  $\pi_1$ . Then delete the rows and columns as explained above and get the next matrix (Table 4).

Table 4. *GRP* in action

	$c_2$	$c_4$	$c_5$	$c_6$	$c_8$	$c_{10}$
$f_2$		1	1	1		
$f_3$			1	1		1
$f_4$	1			1	1	
$f_6$	1	1				1

Now select column  $c_2$  and row  $f_4$ . Extend current cover to  $\pi_1 = \{f_5, f_4\}$ . Delete the rows and columns with respect to this new selection: namely, delete columns  $c_2, c_6, c_8$  and rows  $f_4, f_6$ .

Acting by analogy (select column  $c_4$  first and  $c_{10}$  next), form the resulting covering set  $\pi_1 = \{f_5, f_4, f_2, f_3\}$ . This solution was delivered by heuristic method without warrants of optimality. The essence of this heuristic method (introduced by *A.D. Zakrewsky* and here slightly modified) is to select the columns with minimum number of units first and in those columns find a covering row with maximum number of units. It is time now to generate a logical consequence of the columns  $c_1, c_2, c_4, c_{10}$  – their *group resolvent*. To do this, let us select submatrix of the initial matrix with columns  $c_1, c_2, c_4, c_{10}$  (Table 3b). This submatrix is called *syndromic*. A group resolvent is a new column defined according to the following *rule (RSA)*: it contains 1s only in the rows of syndromic matrix with two or more 1s. Add this resolvent ( $res_1$ ) to initial matrix in Table 3a.

Now perform the second iteration ( $q = 2$ ). Omitting the details, find  $\pi_2 = \{f_6, f_1, f_3\}$  with syndromic columns  $res_1, c_1, c_5$  correspondingly. To produce a new group resolvent, form syndromic matrix for this cover with columns  $res_1, c_1, c_5$  and rows  $f_1, \dots, f_6$ . According to *RSA*, the group resolvent is empty. The entire process terminates with the best solution found –  $\pi_2 = \{f_6, f_1, f_3\}$ . This solution stands for a minimum size cover, we have been looking for. Theoretical backgrounds of *GRP* can be found in [17, 18].

The computational complexity estimation of *GRP* is given in the final part of the paper and testifies to its polynomial properties in average with the

required number of iterations submitted to expression [17]  $O\left(\frac{n \cdot m \cdot p}{\sqrt{1-p}}\right)$ ,

where  $n$  ( $m$ ) stands for the number of rows (columns) in original matrix  $DM$  and  $p$  is the density of units (i.e. the number of units divided by  $n \cdot m$ ) ( $p$  is supposed to be not close to 0 or 1).

The evident drawback of the described method is continuous growth of the matrix sizes due to adding of new group resolvents. Now we introduce a new technique to eliminate this drawback within the frame of enhanced version of *GRP*.

**4. An Enhanced Version of *GRP*.** The following material is based on [19]. In the method realizing enhanced version of *GRP* new group resolvents (starting from some time point) are not added to matrix  $B$  as additional columns but overlap some previously generated resolvents. The total number of the added group resolvents does not exceed the number of rows in  $B$ .

Provide the following reasoning. Let a cover  $\pi_i$  was found at iteration  $i$  by sequential including rows  $\alpha_1, \alpha_2, \dots, \alpha_k$ . Suppose that new iteration  $i+1$  entirely repeats the previous iteration  $i$ . This means that the same syndromic columns and the same rows  $\alpha_1, \alpha_2, \dots, \alpha_k$  are selected in the same order including some additional new row(s). At the moment of including row  $\alpha_k$  into  $\pi_{i+1}$  ( $\alpha_k$  is the last one in  $\pi_i$ ) matrix  $B$  cannot be entirely destroyed, otherwise one gets  $\pi_i = \pi_{i+1}$ , which is impossible according to *GRP* theoretical properties [17, 18]. This means that at least one column  $\beta$  should remain undeleted and  $\beta$  is not covered by any one of rows  $\alpha_1, \alpha_2, \dots, \alpha_k$ . But column  $\beta$  must be at this moment totally zero as all the rows having «1s» in  $\beta$  will be deleted (because the same syndromic columns are selected for rows  $\alpha_1, \alpha_2, \dots, \alpha_k$  at the iterations  $i$  and  $i+1$ ). Evidently, this is impossible and enables one to come to one of the next conclusions:

- either  $\pi_{i+1}$  has less than  $k$  rows

or

- at one of the steps  $1, 2, \dots, k$  when forming cover  $\pi_{i+1}$  in the selected syndromic column there would be smaller amount of units in comparison to the syndromic column selected at the same step while forming cover  $\pi_i$ .

This decisive reasoning enables one to restrict the number of added group resolvents by only those ones which were used as syndromic columns

at the current iteration of *GRP*. Obviously, the total number of these syndromic columns cannot exceed the number of rows in  $B$ . The rest group resolvents which were not used as syndromic columns at the current iteration of *GRP* can be excluded without loss of solution. One can use any one of the excluded columns to replace it with a new group resolvent (this means that new group resolvent simply overlaps the old column without extending current matrix  $B$  sizes).

Thus, the enhanced method enables one not to exceed the memory region restricted by 2-dimension array with  $m$  rows and  $N+m$  columns (where  $N$  stands for initial number of columns in  $B$ ).

We now consider the weighted case of *GRP* [18]. One may be interested in finding among all minimum-size covering sets that one with maximum total weight (representing sum of the weights of rows (features) from this covering set). It is important in the case when new samples not included in original data set should be classified later. A feature weight represents in general a score evaluating its classifying capabilities. The general idea is that the greater the total weight of the selected features, the less likely it is to re-train the classification model. We introduce a specific formulation of the covering problem and solve it by means of the modified *GRP* version.

**5. Weighted Case of *GRP*.** Suppose, each row  $i$  of 0,1-matrix  $B$  is assigned an integer weight  $w_i$ . Formulate a problem as to find from all minimum-size covers of  $B$  a cover  $\pi^*$  such that:

$$\begin{aligned} \forall r \forall s (C_r \neq C_s) \rightarrow Ob_r(\pi^*) \neq Ob_s(\pi^*); \\ \exists \pi ((|\pi| < |\pi^*|) \& \forall r \forall s (C_r \neq C_s) \rightarrow Ob_r(\pi) \neq Ob_s(\pi)); \quad (4) \\ \sum_{i \in \pi^*} w_i \rightarrow \min, \end{aligned}$$

where  $C_r, C_s$  belong to class labels in the data set  $D = \{Ob_i\}, i = 1, N$ .

This formulation is different from a classical one which requires to find a cover  $\pi_k$  (not obligatory a minimum-size one) with a minimum sum of the weights of rows in  $\pi_k$ . Clearly, the weights  $w_i$  may initially be defined as negated RELIEF-weights. To solve a problem, we address to weighted case of *GRP* and modify it to meet our goals. The idea behind the method is to use *GRP* as in section 3 to find a minimum-size cover  $\pi_k$  at the iteration  $k$  and then on the syndromic matrix, corresponding to  $\pi_k$ , build a new group resolvent *res* (in a new fashion described in [18]) such that if an optimal solution has not been found yet, it would cover *res*. For the details, let us

consider an example from Table 3a again with additional column  $w$  representing the row weights.

The procedure finds a minimum-size cover first. It is the same as in unweighted case of *GRP*. We previously found a minimum-size cover  $\pi_2 = \{f_6, f_1, f_3\}$  with syndromic columns  $res_1, c_1, c_5$  correspondingly (Table 5).

Table 5. Weighted syndromic matrix

	$c_1$	$c_5$	$res_1$	$w$
$f_1$	1			4
$f_2$		1		2
$f_3$		1		5
$f_4$				6
$f_5$	1			2
$f_6$			1	3

The weight of the cover  $\pi_2$  is denoted as  $w(\pi_2) = 4 + 5 + 3 = 12$ . To form a group resolvent, we use rule *RSB* suggested in [18] with peculiarity of the syndromic matrix having no rows with two or more 1s. Namely, divide the rows into two subsets  $SR_{\pi}^I$  and  $SR_{\pi}^{II}$ . Subset  $SR_{\pi}^I$  contains those rows with no more than one unit in each of them. In our case due to observed specificity of syndromic matrix,  $SR_{\pi}^I$  coincides with the set of all rows, i.e  $SR_{\pi}^I = \{f_1, f_2, \dots, f_6\}$ . Subset  $SR_{\pi}^{II}$  contains the rest rows of matrix  $B$ , not belonging to  $SR_{\pi}^I$ . In our case  $SR_{\pi}^{II} = \emptyset$ .

Now define the low boundary  $LB(\pi_2)$  by means of the cover  $\pi_2 = \{f_6, f_1, f_3\}$  and its syndromic matrix. Denote set of columns of syndromic matrix by  $\Omega(\pi_2)$ . Let us for each column  $z$  from  $\Omega(\pi_2)$  denote by  $\rho(z)$  the row in  $SR_{\pi}^I$  with minimum weight which covers column  $z$ .

Then

$$LB(\pi_2) = \sum_{z \in \Omega(\pi_2)} w(\rho_z). \tag{5}$$

In the example,  $LB(\pi_2) = 2 + 2 + 3 = 7$ . Now let us compare  $LB(\pi_2) = 7$  and the weight of the cover  $\pi_2$   $w(\pi_2) = 12$ .

*Proposition 2.*

1. If with respect to the current cover  $\pi_i$   $w(\pi_i) \leq LB(\pi_i)$  then in supposition that  $\pi_i$  is not optimal solution, each optimal solution contains at least one row from  $SR_{\pi}^I$ . Consequently, for this case, if  $SR_{\pi}^I = \emptyset$  then  $\pi_i$  is an optimal solution.

2. If  $w(\pi_i) > LB(\pi_i)$  then one needs to move minimum number of rows from  $SR_{\pi}^I$  (no one should belong to current cover  $\pi_i$ ) into  $SR_{\pi}^II$  to provide.

$$- w(\pi_i) \leq LB(\pi_i);$$

- conditions for *RSB* to generate a group resolvent as a column with the units standing only in the rows belonging to  $SR_{\pi}^II$ .

The easiest way to provide 2b is as follows: for each column  $c_x$  in  $\Omega$  find the corresponding value  $\Delta_x = w_{\max}(x) - w_{\min}(x)$ , where  $w_{\max}(x)$  ( $w_{\min}(x)$ ) is the maximum (minimum) value of the weight of some row covering column  $c_x$ . If there is only one row  $\alpha$  covering  $c_x$  or  $\Delta_x = 0$  then this column is ignored. In the example (Table 5), one has  $\Delta_{c_1} = 4 - 2 = 2$ ,  $\Delta_{c_5} = 5 - 2 = 3$ . Find maximum value among  $\Delta_x$ . In the example, this is  $\Delta_{c_5}$ . Consequently, one needs to transfer row  $f_2$  (covering

$c_5$  and having minimum weight value) from  $SR_{\pi}^I$  to  $SR_{\pi}^II$ :  $SR_{\pi}^I = \{f_1, f_3, \dots, f_6\}$ ,  $SR_{\pi}^II = \{f_2\}$ . This transfer leads to  $LB(\pi_2) = 2 + 3 + 5 = 10 < w(\pi_2) = 12$ . It is necessary to make one more transfer. There remains only one possibility: to transfer row  $f_5$  (covering  $c_1$  and having minimum weight value) to  $SR_{\pi}^II$ :  $SR_{\pi}^I = \{f_1, f_3, f_4, f_6\}$ ,  $SR_{\pi}^II = \{f_2, f_5\}$ .

Now  $LB(\pi_2) = 4 + 3 + 5 = 12 = w(\pi_2) = 12$  and one can build a group resolvent  $res_2$  with the units only in the rows  $f_2, f_5$ . Add this resolvent to original matrix in Table 3a and resume searching a minimum-size cover.

Point 2 of the *Proposition 2* can be reformulated in a stronger form. Namely, let  $w^*$  denote the minimum weight of the best minimum-size covering set  $\pi^*$  found at the previous iterations and  $w(\pi_i)$  as earlier stands for the weight of the current minimum-size cover. Then the strengthened form of proposition 2 is as follows:

*Proposition 2 (strengthened form)* [18].

1. If with respect to the best cover  $\pi^*$  and the current cover  $\pi_i$   $w^* \leq LB(\pi_i)$  then in supposition that  $\pi^*$  is not optimal solution, each optimal solution contains at least one row belonging to  $SR_{\pi}^{II}$ . Consequently, for this case, if  $SR_{\pi}^{II} = \emptyset$  then  $\pi^*$  is an optimal solution.

2. If  $w^* > LB(\pi_i)$  then one needs to move minimum number of rows from  $SR_{\pi}^I$  (no one should belong to current cover  $\pi_i$ ) into  $SR_{\pi}^{II}$  to provide.

–  $w^* \leq LB(\pi_i)$ ;

– conditions for *RSB* to generate a group resolvent as a column with the units standing only in the rows contained in  $SR_{\pi}^{II}$ .

*Proposition 3.* New resolvents generated by the rules of proposition 2 may be overlapped in the iterations of *GRP* as in the unweighted case of *GRP* (that is, in the case they are not used as syndromic columns at some future iteration(s)).

*Proof.* From syndromic matrix one has got a new column – group resolvent  $res_x$ . It excludes current minimum size cover  $\pi_x$ . Consider next iteration  $x+1$ . Again, as previously,

– either at one of the steps  $1, 2, \dots, k$  when forming next cover  $\pi_{x+1}$  in the selected syndromic column there would be smaller amount of units in comparison to the syndromic column selected at the same step while forming cover  $\pi_x$

or

–  $\pi_{x+1}$  has less than  $k$  rows.

The last is impossible since  $\pi_x$  is a minimum size cover. Therefore, there remains the first possibility directing the computations in a new way previously not passed. This notice remains valid with respect to any new group resolvent with no relation to the previously added group resolvents provided that they were not used as syndromic columns at the current iteration of *GRP*. This observation also provides finiteness of the searching process.

Now we have all necessary rules to lead the searching process to its finish. Omitting the details, the next cover is  $\{f_6, f_5, f_4, f_2\}$  with syndromic matrix shown in Table 6a.

Add new resolvent  $res_3$  (Table 6a) to original matrix (instead of  $res_2$ , since  $res_2$  was not used as syndromic column) and resume a search.

Perform new iteration of *GRP*. Now one finds  $\pi_{fin} = \{f_1, f_6, f_2\}$  with total weight  $w(\pi_{fin}) = 9$  and syndromic matrix given in Table 6b. For this syndromic matrix one comes to a conclusion about optimality of  $\pi_{fin}$  (as its group resolvent is empty and  $\pi_{fin}$  has minimum total weight among the previously found minimum-size covering sets).

Table 6. Syndromic matrix for  $\{f_6, f_5, f_4, f_2\}$  with resolvent  $res_3$  (a) and syndromic matrix for  $\pi_{fin} = \{f_1, f_6, f_2\}$  with empty resolvent (b)

a)						b)			
	$c_1$	$c_5$	$c_8$	$res_1$	$res_3$	$c_2$	$c_5$	$res_3$	$w$
$f_1$	1		1		1			1	4
$f_2$		1					1		2
$f_3$		1					1		5
$f_4$			1			1			6
$f_5$	1								2
$f_6$				1		1			3

**6. Fuzzy Case of *GRP*.** Now suppose that some samples in original data set are characterized with fuzzy measure  $\mu_A(\mu_B)$  of belonging to class *A* (*B*). This supposition leads to a discrimination matrix with elements from diapason  $[0, 1]$ . Indeed, if some feature  $f_x$  discriminates between row  $r$  and row  $q$  (standing for the objects  $i_r, i_q$  from different classes, e.g. *A* and *B* respectively) then at the intersection of row  $f_x$  and column  $(r, q)$  of *DM* one places the value of  $\mu(f_x, r, q) = \mu_A(i_r) \cdot \mu_B(i_q)$ . The covering problem now should be reformulated as described below.

*Definition 4.* Let row  $a$  contain  $\mu(a, b) > 0$  in column  $b$ . Then they say that row  $a$  covers column  $b$  in a fuzzy mode with fuzzy measure  $\mu_A(a, b)$ .

*Definition 5.* A set of rows  $\pi = \{a_1, a_2, \dots, a_r\}$  is a covering fuzzy set for *DM* if for each column  $b$  from *DM* there is at least one row from  $\pi$  that covers  $b$  in a fuzzy mode.

Let  $\pi = \{a_1, a_2, \dots, a_r\}$  be a fuzzy covering set. Define for each column  $j$  of *DM* the value  $v_j(\pi) = \max(\mu(f_1, j), \mu(f_2, j), \dots, \mu(f_W, j))$  where  $W$  stands for the number of columns in original *DM*.

*Definition 6.* An optimal minimum-size fuzzy covering set  $\pi^*$  is defined as that one having 1) minimum number of rows among all fuzzy covering sets for  $DM$  and 2) providing maximum value of

$$\Psi_{\pi^*} = \sum_{j=1, W} v_j(\pi^*).$$

Formally,

$$\begin{aligned} & \forall r \forall s (C_r \neq C_s) \rightarrow Ob_r(\pi^*) \neq Ob_s(\pi^*); \\ & \exists \pi (|\pi| < |\pi^*|) \ \& \ \forall r \forall s (C_r \neq C_s) \rightarrow Ob_r(\pi) \neq Ob_s(\pi); \quad (6) \\ & \Psi_{\pi^*} = \sum_{j=1, W} v_j(\pi^*) \rightarrow \max, \end{aligned}$$

where  $C_r, C_s$  belong to class labels in the data set  $D = \{Ob_i\}, i = 1, N$ .

*Notice 3.* It is important to emphasize that group resolvents added at the iterations of algorithm with fuzzy matrix  $DM$  do not influence upon the value of  $\Psi_{\pi}$ .

To find an optimal minimum-size fuzzy covering set, one can use a slight modification of *GRP* for the weighted case considered above practically basing on the same ideas. Again, let us take a matrix (Table 7) as an example to illustrate the details.

Table 7. Fuzzy matrix  $DM$

	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$c_8$	$c_9$	$c_{10}$
$f_1$	0.6		1					0.8		
$f_2$				0.9	0.9	1			1	
$f_3$			0.8		0.7	0.5				0.7
$f_4$		1				0.5		1		
$f_5$	0.8		0.8				1		1	
$f_6$		1		1			1		1	0.9

First, one is looking for some minimum-size cover (with no regards to its weight  $\Psi_{\pi}$ ). This process has a peculiarity when making group resolvents only. One is acting as though fuzzy elements are crisp-valued (as in *GRP*). Namely, find column with minimum number of non-zero elements (it is a syndromic column). Then define a row with maximum number of non-zero elements which covers this column in a fuzzy mode. Then reduce the matrix according to the *GRP* rules and continue till the new covering set is defined.

The first covering set  $\pi_1 = \{f_5, f_4, f_2, f_3\}$  with the syndromic columns  $c_1, c_2, c_4, c_{10}$ . Find their group resolvent (Table 8a).

*Definition 7.* A fuzzy group resolvent is defined as that one, containing units in the rows with two or more non-zero elements in syndromic matrix rows.

Thus, in Table 8a the fuzzy resolvent contains the only unit in row  $f_6$ .

This phase finishes with an empty resolvent found on the last syndromic matrix. Now, a new group resolvent should be generated to provide condition 2) in *Definition 6*. For clearness, consider the covering set  $\pi_2 = \{f_6, f_1, f_3\}$  with fuzzy syndromic columns  $res_1, c_1, c_5$  (Table 8b).

Let us formulate the rule *RSC* for the fuzzy case of *GRP*.

1. If in the syndromic matrix for the last minimum-size cover there are no fuzzy elements (different from 0 and 1) then optimal solution is found corresponding to the best solution found in the previous iterations including the last one. Otherwise.

2. In each column of the syndromic matrix mark (with «\*») all non-zero elements with the values higher than the value of the syndromic element in this column (see Table 9 as an example). If there are no marked elements then algorithm finishes with optimal solution, corresponding to the best solution found in the previous iterations including the last one. Otherwise.

3. Form a new group resolvent as a column containing units in the rows with marked elements and zeroes in the rest rows.

Table 8. Fuzzy syndromic submatrix *DM* with fuzzy group resolvent  $res_1$  (a); Fuzzy syndromic matrix for  $\pi_2 = \{f_6, f_1, f_3\}$  (b)

a)						b)		
	$c_1$	$c_2$	$c_4$	$c_{10}$	$res_1$	$c_1$	$c_5$	$res_1$
$f_1$	0,6					0,6		
$f_2$			0,9				0,9	
$f_3$				0,7			0,7	
$f_4$		1						
$f_5$	0,8					0,8		
$f_6$		1	1	0,9	1			1

After forming a fuzzy group resolvent add it to the current matrix *DM* and resume iterations.

Again, the column-resolvents may overlap the previously built group resolvents provided they were not used in syndromic matrix. Therefore, the finiteness of entire process is based on the following fact.

Each time when iterations resume after adding new group resolvent, the computations are being performed in a new direction (see proof of Proposition 3) due to the following decisive point: at one of the steps  $1, 2, \dots, k$  when forming next cover  $\pi_{x+1}$  in the selected syndromic column, there would be smaller amount of units in comparison to the syndromic column selected at the same step while forming cover  $\pi_x$ .

Table 9. Fuzzy resolvent (rule RSC)

	$c_1$	$c_5$	$res_1$	<i>fuzzy resolvent</i>
$f_1$	0,6			
$f_2$		0,9*		1
$f_3$		0,7		
$f_4$				
$f_5$	0,8*			1
$f_6$			1	

**7. Approximate Covering Procedure.** One can restrict the *GRP* iterations before getting empty group resolvent as was shown in [17, 18]. Let us reproduce some estimations of complexity of approximate covering procedure. Denote by  $n$  ( $m$ ) the number of rows (columns) in *DM*; let  $p$  stand for the density of units, that is,  $p$  is equal to the total number of units in *DM* divided by  $(n \cdot m)$ . Let  $k$  stand for the covering set size. Accordingly to [17, 18], the number of iterations required in order that mathematical expectation ( $M_k$ ) of the number of covers with  $k$  rows becomes  $M_k < 1$  can be defined from the condition:

$$(n + 0.5) \cdot \ln n - (k + 0.5) \cdot \ln k - (n - k + 0.5) \cdot \ln(n - k) + \frac{1}{12} \cdot n - \frac{1}{12k + 1} - \left( -\frac{1}{12(n - k) + 1} - m \cdot \ln(1 - \varepsilon_k) \right) \leq -1.5, \tag{7}$$

where

$$\varepsilon_k = \left(1 - \frac{p \cdot n}{n}\right) \times \left(1 - \frac{p \cdot n}{n-1}\right) \times \dots \times \left(1 - \frac{p \cdot n}{n-k+1}\right). \tag{8}$$

These relations are obtained in supposition that in «average case» the density of units in the group resolvents is approximately the same as  $p$  what is confirmed for quite a big amount of experimental data (excluding extreme cases with very low density  $p$  or its closeness to 1).

Formulas (7), (8) enable one to stop iterations before getting an empty resolvent. The estimation of the number  $I$  of iterations is of the form:

$$I = O \left( \frac{m \cdot p \cdot \left( n - \ln \left( \frac{z^2 + 2}{2} - \sqrt{\left( \frac{z^2 + 2}{2} \right)^2 - 1} \right) \right)}{\sqrt{1 - p}} \right), \quad (9)$$

which can be simplified to

$$I = O \left( \frac{m \cdot p \cdot (n + 2.41)}{\sqrt{1 - p}} \right). \quad (10)$$

Here,  $z$  can be selected from the well-known rule of « $z\sigma$ » (e.g.  $z = 3$  or higher). The rule of « $3\sigma$ » means that a value of a normally distributed random variable  $x$  with mean  $x_{mean}$  falls in the diapason  $[x_{mean} - 3 \cdot \sigma, x_{mean} + 3 \cdot \sigma]$  with a probability close to 0.997.

So, in average the approximate method behaves itself like a polynomial computational method for a given density  $p$  of units (see the concomitant considerations in [17, 18]).

Now consider the last question: how to restrict the original number of samples in the learning data set.

**8. Restriction of Data Set Sizes.** For multi-dimension data one can use  $\varepsilon$ -nets (see for instance [20]) with nodes covering data samples in the following way: for each multi-dimension data object there exists one and only one node in  $\varepsilon$ -net the (Euclidean or other) distance to it does not exceed  $\varepsilon$ . Building  $\varepsilon$ -net is again a minimum-size covering problem. So, in order to simplify computational expences one can use  $K$ -nearest neighbors method (see papers review [21]) to build  $K > 0$  clusters such that each data object gets directly to one and only one cluster. We do not restrict this formulation by the condition that  $K$  should have a minimum value. Then one can use the cluster centroids instead of data objects from original data set [22]. This gives us a solution to reduction of the sizes of discrimination matrix.

Another possibility is connected to use random sampling technique developed in applied statistics. This approach, and the previously mentioned, need special attention.

**9. Experiments.** For estimation of the described approach on the basis of *GRP* and minimum-size matrix covering technique we used *DecisionTreeClassifier* (*DTC*) and *RFECV* (recursive feature elimination with cross-validation) methods provided by Python programming language. Comparative results with *DecisionTreeClassifier* are placed in Table 10, with *RFECV* – in Table 11. The original data sets contained two classes with randomly generated binary vectors and unit density randomly chosen from diapason [0.2, 0.5]. The first column in Tables 10, 11 indicates to original amounts of features and samples. The second and the third columns define the resulting amounts of features found by the corresponding method. The order of the numbers (experimental results) in the second column corresponds to the comparative results of the same experiments in the third column for each row of the Table.

The calculation time of each experiment with *GRP* was in the worst case three times longer compared to the *DTC* method but did not exceed 10 seconds on IBM Pentium 2.1GHz.

Table 10. Comparative results with *DecisionTreeClassifier*

Features, samples (original Data set)	Feature set sizes found with <i>GRP</i>	Feature set sizes found by <i>DTC</i>
15, 100	7, 9, 14	14, 14, 15
20, 100	9, 8, 9, 7, 12, 7	16, 18, 18, 14, 17, 15
30, 100	9, 9	20, 23
40,100	8, 9, 12	17, 19, 20
50, 100	9, 10, 8	21, 20, 19

Table 11. Comparative results with *RFECV*

Features, samples (original Data set)	Feature set sizes found with <i>GRP</i>	Feature set sizes found by <i>RFECV</i>
20, 100	17, 14, 10, 9	18, 16, 12, 13
30, 100	10, 11, 13	13, 14, 14
40,100	12, 12, 9, 9	14, 15, 10, 12
50, 100	12, 7, 16	14, 8, 19

One can see that *GRP* provides stable superiority over Python techniques with practically acceptable computation time. Developing the ideas of the section 8 of this paper, we also performed experiments with big data sets containing 300 multidimensional binary objects (vectors) what exceeds the predefined limitations on our program for *GRP*-based solution technique. In experiments, two classes of objects were generated

with different mathematical expectation and standard deviation. The best results were obtained for 20-30 clusters with classification accuracy near 90%. However, increasing the number of clusters did not improve accuracy of classification and even worsened it. This problem remains open for further investigations.

A series of 30 experiments was conducted to find out relations between original features amount ( $OFA\{20, 30, 40, 50, 60\}$ ), number of objects ( $N\{80, 100, 120, 130, 140, 150, 160\}$ ), minimum feature amounts ( $FAGRP$ ) found by  $GRP$ -based covering technique and number of nodes in the classifying tree ( $Nnod$ ) created for  $FAGRP$  and  $N$ . There were generated two classes of objects with predefined probabilities of units for 0,1-valued features. The following conclusions can be made:

1.  $FAGRP = O(k \cdot OFA^{0.5})$  with a constant  $k$  in  $[0.8-3.0]$  (in majority of cases  $k$  is near to 1.5).
2.  $Nnod$  depends on  $FAGRP$  ( $OFA$ ) in unstable mode within given 0,1-distribution of the feature values and fixed  $N$  (Table 12a).
3.  $Nnod$  has no clear tendency to growth with increasing  $N$  and fixed  $OFA$  as is illustrated by Table 12b.

Table 12. Results of experiments ( $N = 100$ ) (a); Results of experiments for different  $N \in [100, 160]$ ,  $OFA = 40$  (b)

a)			b)		
Features, samples (original Data set)	$FAGRP$	$Nnod$	Features, samples (original Data set)	$FAGRP$	$Nnod$
20, 100	7	41	40, 100	6	41
30, 100	7	55	40, 120	9	115
40, 100	6	35	40, 140	8	83
50, 100	9	39	40, 150	9	63
60, 100	23	87	40, 160	8	67

As the last example, consider feature set reduction in image recognition. An original image represents digit «4» placed within a square divided by cells (Fig. 1). Each cell in Figure 1 stands for the feature  $f_i$ . Initially, there are 64 features. Each feature is either zero (no part of the digit «4» is within the corresponding cell) or unit (the cell contains some part of the digit «4»).

One can randomly distort the image by clearing some cells with parts of the digit or painting empty cells like in Figure 2. In experiments, we

randomly obtained 50 distorted images of the digit «4» with the slight modifications (cluster *A*). Also there were generated 50 random samples with chaotic distribution of the empty and colored cells (cluster *B*). The *GRP*-based method left 23 features accordingly to minimum-size cover of the discriminating matrice. This result was obtained practically instantly for a single *GRP*-iteration despite the big sizes of the discriminating matrix (64 rows and 2500 columns).

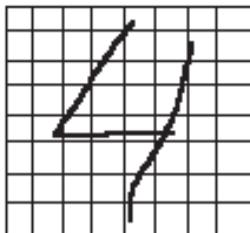


Fig. 1. Recognition of the digit «4»

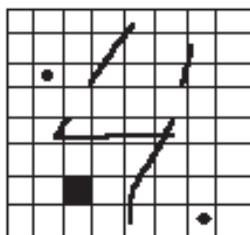


Fig. 2. Distorted image

**10. Conclusion.** The total approach outlined here competes well with the known methods and gives better solutions in majority of cases especially with big initial amounts of the features. It makes possible to operate with discrimination matrices with some hundreds of features (this amounts to 300 in our program). To extend the practical boundaries of the realized technique it was suggested to use clusterization of the input data sets which showed promising results, although they are needed in future investigations. The *GRP*-based approach may serve a common platform for different feature selection models and can be extended in the following directions: processing incomplete (impure) data, processing qualitative data, integrating control models in classification process, modeling practical systems in different areas *etc.*

**Appendix.** By means of Python programming language let us build a classifying regression tree (CRT) on the reduced feature set with two features  $f_1, f_2$ . One can use the next Python script (List. 1) which can be applied in general case for  $n > 2$  features.

```
import numpy as np
from sklearn.tree import DecisionTreeRegressor
import matplotlib.pyplot as plt
# Create datasets
X = np.array([[0.8, 0.5], [1.0, 0.5], [0.4, 0.25], [0.2,
0.0],[0.7,1.0],[0.0,1.0],[0.0,0.5],[0.4,1.0]])
X_test = np.array([[0.94, 0.5],[0.85,0.3],[0.3,0.3]])
Y = np.array([1.0,1.0,0.0,0.0, 1.0, 0.0, 1.0, 0.0])
# Fit model
regr_1 = DecisionTreeRegressor(max_depth=2)
#regr_2 = DecisionTreeRegressor(max_depth=5)
regr_1.fit(X, Y)
# Predict
y_1 = regr_1.predict(X_test)
print y_1
```

Listing. 1. Python code to build classification tree

The columns  $f_1, f_2$  in Table 1 are represented as array X (features) and Y (classes). The regression tree is created in operator:

$$\text{regr\_1} = \text{DecisionTreeRegressor}(\text{max\_depth}=2).$$

In  $y_1 = \text{regr\_1.predict}(X\_test)$  one verifies how the model predicts test values (defined in array X\_test).

This script provides the output in the form of array [1, 1, 0] with the first two 1s defining class A and last 0 defining class B for the two-features inputs [0.94, 0.5], [0.85, 0.3], [0.3, 0.3] respectively.

## References

1. Shah S.A., Shabbir H.M., Rehman S., Waqas M. A comparative study of feature selection approaches: 2016–2020. *International journal of scientific and engineering research*. 2020. vol. 11. no. 2. pp. 469–478.
2. Khun K., Johnson K. Feature engineering and selection. A practical approach for predictive models. CRC Press. 2019. 310 p.
3. Bachu V., Anuradha J. A review of feature selection and its methods. *Cybernetics and information technologies (Bulgary)*. 2019. vol. 19. no. 1. pp. 3–22.
4. Hameed S., Petinrin O., Hashi A., Saeed F. Filter-wrapper combination and embedded feature selection for gene expression data. *International journal of advances in soft computing and its applications*. 2018. vol. 10. no. 1. pp. 91–105.
5. Sanchez-Pinto L.N., Venable L.R., Fahrenbach J., Churpek M. Comparison of variable selection methods for clinical predictive modeling. *International journal of medical informatics*. 2018. vol. 116. pp. 10–17.
6. Li J. et al. Feature selection: A data perspective. *ACM Computer surveys*. 2017. vol. 50. no. 6. pp. 94:2–94:46.

7. Guyon I. et al. Feature Extraction. Foundations and Applications. Springer. 2006. 762 p.
8. Urbanowicz R.J. et al. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*. 2018. vol. 8. no. 5. pp. 189–203.
9. Liu Y., Singleton A., Arribas-Bel D. A principal component analysis (PCA)-based framework for automated variable selection in geodemographic classification. *GEO-Spatial Information Science*. 2019. vol. 22. no. 4. pp. 251–264.
10. Khanna R., Awad M. Efficient learning machines: Theories, Concepts, and Applications for engineers and system designers. Apress. 2015. 247p.
11. Mao Y., Yang Y. A wrapper feature subset selection based on randomized search and multilayer structure. *BioMed Research International*. 2019. vol. 2019. pp. 1–9.
12. Hui K.H. et al. An improved wrapper-based feature selection method for machinery fault diagnosis. *PLoS ONE*. 2017. vol. 12. no. 12. pp. 1–10.
13. Lal T.N., Chapelle O., Weston J., Elisseeff A. Embedded methods. Series in Fuzzy and soft computing. 2006. vol. 207. pp. 137–165.
14. Sudrajat R., Irianingsih I., Krisnawan D. Analysis of data mining classification by comparison of C4.5 and ID algorithms. IOP Conference Series: Materials and Engineering. 2017. vol. 166. pp.012031.
15. Krishna M. et al. Predicting student performance using classification and regression trees. *International Journal of Innovative Technology and Exploring Engineering*. 2020. vol. 9. no. 3. pp. 3349–3356.
16. Suneetha N., Hari Ch., Sunilkumar V. Modified Gini index classification: a case study on hart disease dataset. *International journal on computer science and engineering*. 2010. vol. 2. no. 6. pp. 1959–1965.
17. German O.V., Naidenko V.G. [Statistically optimal algorithm for the minimum-size covering problem] *Jekonomika i matematicheskie metody – Economics and mathematical methods*. Moscow. 1993. Issue 29. vol. 4. pp. 662–667. (In Russ.).
18. German O.V. [The generalized statistically optimal method to find minimum weighted covering set for 0,1-matrix] *Jekonomika i matematicheskie metody – Economics and mathematical methods*. oscar. 1994. Issue 30. vol. 4. pp. 139–150. (In Russ.).
19. German O.V. *Jekspertnye sistemy* [Expert systems]. Minsk. Belorusskij gos. universitet informatiki i radioelektroniki. 2008. 91 p. (In Russ.).
20. Kamenev G.K., Kamenev I.G. Primenenie metodov mnogomernogo analiza dlja izuchenija sociologicheskikh sovokupnostej [Applications of the methods of multidimension analysis for learning social aggregates]. Proc. of the department of mathematical modeling of economic systems. Computer Center «Informatics and Control» of the Russian Academy of Sciences. 2017. 91p. (In Russ.).
21. Bhatia N. et al. Survey of Nearest Neighbor Techniques. *International Journal of Computer Science and Information Security*. 2010. vol. 8. no. 2. pp. 302–304.
22. Sun L., Chen G., Xiong H., Guo C. Cluster analysis in data-driven management and decisions. *Journal of Management Science and Engineering*. 2017. vol. 2. no. 4. pp. 227–251.

**German Oleg** – Ph.D., Associate Professor, Department of Information Technologies in Automated Systems, Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: applied logic, informatics, cybernetics. The number of publications – 140. ovgerman@tut.by; 6, Petrusya Brovki str., 220600, Minsk, Belarus; office phone: +375 17 2938823; fax: +375 17 2702033.

**Nasrh Sara** — Postgraduate Student, Department of Information Technologies in Automated Systems, Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: informatics, decision making. The number of publications – 10. sara.nasrh@gmail.com; 6, Petrusya Brovki str., 220600, Minsk, Belarus; office phone: +961 3997163; fax: +375 17 2702033.

О.В. ГЕРМАН, С.Н. НАСР  
**НОВЫЙ МЕТОД ОПТИМАЛЬНОГО СОКРАЩЕНИЯ  
МНОЖЕСТВА ПРИЗНАКОВ**

*Герман О.В., Наср С.Н. Новый метод оптимального сокращения множества признаков.*

**Аннотация.** Рассматривается задача нахождения минимального по размеру множества атрибутов, используемых для распределения многомерных объектов по классам, например на основе деревьев решений. Задача имеет важное значение при разработке высокопроизводительных и точных классифицирующих систем. Приведен краткий сравнительный обзор известных методов. Задача сформулирована как отыскание минимального (взвешенного) покрытия на различающей  $\{0,1\}$ -матрице, которая служит для описания возможности атрибутов разделять пары объектов из разных классов. Приведено описание способа построения различающей матрицы. Сформулированы и решены на основе общего разрешающего принципа групповых резолюций следующие варианты задачи: отыскание минимального по размеру множества атрибутов на заданном входном наборе данных; отыскание минимального по размеру множества атрибутов с минимальным суммарным весом атрибутов (в качестве весов атрибутов можно использовать величины, определяемые на основе известных алгоритмов, например на основе метода RELIEF); нахождение оптимального взвешенного нечеткого покрытия для случая, когда элементы различающей матрицы принимают значения в диапазоне  $[0,1]$ ; определение статистически оптимального покрытия различающей матрицы (например, для входных наборов данных больших размеров). Статистически оптимальный алгоритм позволяет ограничить время решения полиномом от размеров задачи и плотности единичных элементов в различающей матрице и при этом обеспечить близкую к единице вероятность отыскания точного решения.

Таким образом, предлагается общий подход к определению минимального по размеру множества атрибутов, учитывающий различные особенности в постановке задачи, что отличает данный подход от известных. Изложение содержит многочисленные иллюстрации с целью придать ему максимальную ясность. Ряд теоретических положений, приводимых в статье, основывается на ранее опубликованных результатах. В заключительной части представлены результаты экспериментов, а также сведения о сокращении размерности задачи о покрытии для больших массивов данных. Отмечаются некоторые перспективные направления изложенного подхода, включая работу с неполными и качественными данными, интегрировании управляющей модели в систему классификации данных.

**Ключевые слова:** многомерные данные, классификация, минимизация размера множества атрибутов, задача о минимальном покрытии, принцип групповых резолюций

### Литература

1. *Shah S.A., Shabbir H.M., Rehman S., Waqas M.* A comparative study of feature selection approaches: 2016–2020 // International journal of scientific and engineering research. 2020. vol. 11. no. 2. pp. 469–478.
2. *Khun K, Johnson K.* Feature engineering and selection. A practical approach for predictive models // CRC Press. 2019. 310 p.
3. *Bachu V., Anuradha J.* A review of feature selection and its methods // Cybernetics and information technologies (Bulgary). 2019. vol. 19. no. 1. pp. 3–22.
4. *Hameed S., Petinrin O., Hashi A., Saeed F.* Filter-wrapper combination and embedded feature selection for gene expression data // International journal of advances in soft computing and its applications. 2018. vol. 10. no. 1. pp. 91–105.
5. *Sanchez-Pinto L.N., Venable L.R., Fahrenbach J., Churpek M.* Comparison of variable selection methods for clinical predictive modeling // International journal of medical informatics. 2018. vol. 116. pp. 10–17.

6. *Li J. et al.* Feature selection: A data perspective // ACM Computer surveys. 2017. vol. 50. no. 6. pp. 1–45.
7. *Guyon I. et al.* Feature Extraction. Foundations and Applications // Springer. 2006. 762 p.
8. *Urbanowicz R.J. et al.* Relief-based feature selection: Introduction and review // Journal of biomedical informatics. 2018. vol. 8. no. 5. pp. 189–203.
9. *Liu Y., Singleton A., Arribas-Bel D.* A principal component analysis (PCA)-based framework for automated variable selection in geodemographic classification // GEO-Spatial Information Science. 2019. vol. 22. no. 4. pp. 251–264.
10. *Khanna R., Awad M.* Efficient learning machines: Theories, Concepts, and Applications for engineers and system designers // Apress. 2015. 247p.
11. *Mao Y., Yang Y.* A wrapper feature subset selection based on randomized search and multilayer structure // BioMed Research International. 2019. vol. 2019. pp. 1–9.
12. *Hui K.H. et al.* An improved wrapper-based feature selection method for machinery fault diagnosis // PloS ONE. 2017. vol. 12. no. 12. pp. 1–10.
13. *Lal T.N., Chapelle O., Weston J., Eliseeff A.* Embedded methods // Series in Fuzzy and soft computing. 2006. vol. 207. pp. 137–165.
14. *Sudrajat R., Irianingsih I., Krisnawan D.* Analysis of data mining classification by comparison of C4.5 and ID algorithms // IOP Conference Series: Materials and Engineering. 2017. vol. 166. pp.012031.
15. *Krishna M. et al.* Predicting student performance using classification and regression trees // International Journal of Innovative Technology and Exploring Engineering. 2020. vol. 9. no. 3. pp. 3349–3356.
16. *Suneetha N., Hari Ch., Sunilkumar V.* Modified Gini index classification: a case study on hart disease dataset // International journal on computer science and engineering. 2010. vol. 2. no. 6. pp. 1959–1965.
17. *Герман О.В., Найденко В. Г.* Статистически оптимальный алгоритм для задачи о минимальном покрытии // Экономика и математические методы. 1993. Т. 29. № 4. С. 662–667.
18. *Герман О.В.* Обобщенный статистически оптимальный метод решения задачи о минимальном взвешенном покрытии 0,1-матрицы // Экономика и математические методы. 1994. Т. 30. № 4. С. 139–150.
19. *Герман О.В.* Экспертные системы // Минск. Белорусский гос. университет информатики и радиоэлектроники. 2008. 91с.
20. *Каменев Г.К., Каменев И.Г.* Применение методов многомерного анализа для изучения социологических совокупностей // М. Труды отдела математического моделирования экономических систем ВЦ ФИЦ ИУ РАН. 2017. 91р.
21. *Bhatia N. et al.* Survey of Nearest Neighbor Techniques // International Journal of Computer Science and Information Security. 2010. vol. 8. no. 2. pp. 302–304.
22. *Sun L., Chen G., Xiong H., Guo C.* Cluster analysis in data-driven management and decisions // Journal of Management Science and Engineering. 2017. vol. 2. no. 4. pp. 227–251.

**Герман Олег Витольдович** – канд. техн. наук, доцент, кафедра информационных технологий автоматизированных систем, Белорусский государственный университет информатики и радиоэлектроники (БГУИР). Область научных интересов: прикладная логика, информатика и кибернетика. Число научных публикаций – 140. ovgerman@tut.by; ул. Петруся Бровки, 6, 220600, Минск, Беларусь; р.т.: +375 17 2938823; факс: +375 17 2702033.

**Наср Сара Набих** – аспирантка, кафедра информационных технологий автоматизированных систем, Белорусский государственный университет информатики и радиоэлектроники (БГУИР). Область научных интересов: информатика, принятие решений. Число научных публикаций – 10. sara.nasrh@gmail.com; ул. Петруся Бровки, 6, 220600, Минск, Беларусь; р.т.: +961 3997163; факс: +375 17 2702033.

C. SARMIENTO , J. SAVAGE  
**COMPARISON OF TWO OBJECTS CLASSIFICATION  
TECHNIQUES USING HIDDEN MARKOV MODELS AND  
CONVOLUTIONAL NEURAL NETWORKS**

---

*Sarmiento C., Savage J. Comparison of Two Objects Classification Techniques using Hidden Markov Models and Convolutional Neural Networks.*

**Abstract.** This paper presents a comparison between discrete Hidden Markov Models and Convolutional Neural Networks for the image classification task. By fragmenting an image into sections, it is feasible to obtain vectors that represent visual features locally, but if a spatial sequence is established in a fixed way, it is possible to represent an image as a sequence of vectors. Using clustering techniques, we obtain an alphabet from said vectors and then symbol sequences are constructed to obtain a statistical model that represents a class of images. Hidden Markov Models, combined with quantization methods, can treat noise and distortions in observations for computer vision problems such as the classification of images with lighting and perspective changes.

We have tested architectures based on three, six and nine hidden states favoring the detection speed and low memory usage. Also, two types of ensemble models were tested. We evaluated the precision of the proposed methods using a public domain data set, obtaining competitive results with respect to fine-tuned Convolutional Neural Networks, but using significantly less computing resources. This is of interest in the development of mobile robots with computers with limited battery life, but requiring the ability to detect and add new objects to their classification systems.

**Keywords:** Hidden Markov Models, Image Classification, Computer Vision, Pattern Recognition

---

**1. Introduction.** In the field of computer vision, it has always been required to interpret visual content captured in sensors, providing information to the systems to carry out tasks that are useful. One of these tasks is to recognize what types of objects are in a work environment.

By knowing the meaning of the data that comes from cameras, it's possible to determine the current state of the environment such as knowing where an object is or not in a certain region of space. With this, it's possible to carry out planning, navigation, and manipulation tasks in robotic systems.

Automatic planning or briefly called planning can be described as a set of techniques to represent knowledge, calculate a temporal sequence of actions, and obtain a final configuration to complete one or more tasks. It's common to establish a representation of the internal state of a robotic system and a representation for the environment, by calculating a sequence of actions, a transition between said states is achieved until reaching a terminal state, this last configuration represents the execution of one of several goals.

Autonomous navigation is the ability of a robotic system to estimate its position, calculate a trajectory to a given target in the environment, and

follow a dynamically generated path in order to avoid moving obstacles. To achieve this, it's necessary to build a representation of the environment such as a map, recognize distinctive markings to self-locate through the use of sensors or vision systems and finally estimate a trajectory to the place that has been selected as a destination.

Robotic manipulation, in the context of robots operating in service tasks, refers to the design of control systems that operate mobile manipulators in spaces shared with humans. These places have a certain degree of structure such as a kitchen, a living room, or a bedroom but to carry out manipulation of objects in these environments it's necessary to have computer vision systems. Computer vision can facilitate the control of actions on the manipulator, obtaining descriptions of the shape or texture that would be difficult to do using other types of perception.

Systems based on computer vision can perform tasks such as creating a representation of the environment, self-locate and navigate autonomously, estimate the best way to take an object, all this through the use of digital cameras. This has helped the integration of planning, navigation and manipulation systems in mobile robots to service to people.

The first computer vision systems were based on the extraction of visual features, these had to be robust to changes in lighting, perspective, and scale which were designed manually. These representations were later classified using supervised or unsupervised machine learning techniques, or in variants and compositions of these two options [1].

At present, the most accurate image classification systems are based on what is called Artificial Neural Networks (ANNs). The artificial neurons are used as fundamental building blocks to create new structures, hierarchically and with locality properties in their arrangement, in what is called Convolutional Neural Networks (CNNs) [2].

CNNs are the image detection systems with highest performance [3], with varied and specialized architectures depending on the problem where they are applied. A disadvantage of these detection systems is that the mathematical theory on which they are based is the gradient descent optimization theory [4,5], which requires highly dimensional differentiable functions. This is a problem in terms of computing requirements since it's necessary to store millions of parameters, perform calculations on each of them, and update their values. Due to a large number of parameters very small learning rates are used to maintain convergence and stability, this is a disadvantage since many examples are required to achieve a balance between the generalization of the system and it's precision.

Service robots have gained relevance in recent years [6, 7]. This is because with current technology it's already possible to assemble robots for common tasks such as surveillance, customer service, disinfection of spaces or even home delivery [8]. As a result of research in this branch of robotics, competitive communities have emerged such as RoboCup, which is an international scientific initiative that encourages research and communication of discoveries and its applications.

Mobile service robots must be autonomous, the use of energy and computing must be more efficient to extend the service time that is provided. Currently, these systems use specialized computers with low power consumption and with architectures that allow CNNs to be executed in short intervals of time, some systems, even use laptops as a common rule [9]. In this context, this research shows that it's possible to use discrete Hidden Markov Models (HMMs) as an inference method for classifying images using a data set with a reduced number of classes, achieving similar results to CNNs but using generic computers such as a laptop, with the advantage of executing the classification faster and therefore consuming fewer resources.

The remaining of the paper is divided as follows. In Section 2 we present a summary of HMMs applications for computer vision. Then, in Section 3 we introduce our probabilistic approach to solve the image recognition task and in Section 4 we describe experimental results. Finally, the main findings are discussed in Section 5.

**2. Related Work.** Hidden Markov Models (HMMs) are described in the writings of L. E. Baum [10], in them the theory is described to use Markov chains as a tool to analyze time series. Later its use becomes popular due to the applications found by L.R. Rabiner [11] for speech recognition. In his work, numerical stability techniques for algorithms, preprocessing techniques, and details about the implementation are documented. In [12], a great variety of applications and variations to the original method is compiled. However, HMMs have mainly been used for recognition of symbol sequences in the area of bioinformatics, handwriting recognition, and as a method of pattern recognition for electrical signals in biomedicine.

In the case of recognition from digital images, there are applications such as face recognition [13], in which a number of different architectures are described, such as using one-dimensional HMMs but allowing transitions between various stacked models. This represents different types of sequences as they are in order from top to bottom and from left to right. The symbols are obtained by applying a fixed-size sliding window and calculating the discrete cosine transform over pixel values. Clustering techniques were used to obtain a

discrete alphabet, which is a common practice to discretize observation vectors, in this case, obtained by the coefficients.

The application of HMMs for thermal image classification is shown in [14], which describes its application to classify images of breast cancer. This work merges grayscale images and thermal images represented in the RGB color space. Through binarization of constant size windows on the image, a unique symbol of each section is obtained. Finally, an HMM with two hidden states classifies a sequence as diseased or healthy tissue.

Another application in the field of medical imaging is written in [15]. This work describes the application of HMMs for the detection of blood vessels, where the HMMs have the function of estimating the next most probable state in the images to remove pixel occlusions in binarized retinal images.

In a previous work [16], we expanded the application of HMMs for point clouds detection. To carry out the classification of objects and places, 3D keypoints are detected, and using 3D descriptors it's possible to obtain observation vectors. One HMM was trained with these sequences for each class of location or object, and a pattern was detected. A disadvantage is that 3D descriptors are usually vectors with many dimensions, slowing down the quantization process to obtain a discrete alphabet. This can be solved by using a more compact representation.

The year 2010, an image classification challenge arises as a proposal to improve the performance of classification systems. Today this is known as the ImageNet challenge [17]. This is a benchmark to compare detection systems on millions of images with hundreds of different categories.

In the case of CNNs, they are based on the paradigm of having groups of neurons emulating the convolution operation by applying filter banks. The year 2012 training process using graphics processors was implemented, resulting in an inference design highly dependent on parallelization with excellent results such as AlexNet [18], which was the architecture with the best recognition rate in the ImageNet data set. For this reason, most of the image classifiers are of the CNN type.

Over time, variants with more parameters emerged, such as the so-called VGGNet in 2014 [19], an architecture that manages to reduce the percentage of detection error to less than 8%, obtaining a first place that year in the ImageNet challenge. With the increase in the number of parameters, new problems appear such as the vanishing of gradient values due to the number of stacked layers. To solve this, a new architecture is proposed, such as residual blocks [20] and therefore, the authors manage to add a greater number of layers without degrading the accuracy of the classification system, obtaining an error

close to 3 % in the ImageNet challenge of the year 2015. This architecture was called ResNet.

Today, there are many variations to the architecture. Some have improvements such as being optimized to require fewer parameters and computing time thinking of mobile devices [21]. In this work, we will use VGGNet and ResNet as a reference to compare HMMs in the same task, classification of images, in a domestic service context.

**3. HMM for Image Recognition.** A Hidden Markov Model (HMM) is a set of discrete or continuous variables, representing a time-series observed  $\{o_1, o_2, \dots, o_T\} = o_{1:T}$ . These variables  $o_{1:T}$  are assumed to be generated by an internal configuration that changes successively over time  $\{s_1, s_2, \dots, s_t\} = s_{1:T}$  but which we cannot observe, so they are assumed to be hidden variables. The probability that an observation  $o_t$  is presented at a certain time  $t$  only depends on the current state, that is

$$p(o_t | o_{1:T}, s_{1:T}) = p(o_t | s_t). \quad (1)$$

The stochastic process associated with the states is causal and stationary with a finite number of states, so it's represented as a finite automaton with edges labeled with the probability of transiting from one vertex to another. This process only depends on state  $s_t$  at time  $t$ , and the previous state  $s_{t-1}$ , expressed as follows:

$$p(s_t | s_{1:T}) = p(s_t | s_{t-1}). \quad (2)$$

The probability of observing a sequence of variables  $o_{1:T}$  in a fixed period  $T$  is given as

$$p(o_{1:T}, s_{1:T}) = p(o_1 | s_1) \prod_{t=2}^T p(o_t | s_t) p(s_t | s_{t-1}). \quad (3)$$

Where the probability of transit between internal states  $s_{t-1}$  to  $s_t$  is defined as

$$p(s_{t=j} | s_{t-1=i}) = a_{ij}. \quad (4)$$

The probability of observing a variable  $o_t$  depends on hidden state  $s_t$ , and is defined as:

$$p(o_{t=j} | s_{t=i}) = b_{ij}. \quad (5)$$

The parameters of the model are listed below:

- a finite set of  $N$  states  $S = \{s_1, s_2, \dots, s_N\}$ ;
  - a finite alphabet  $M$ , with symbols  $V = \{v_1, v_2, \dots, v_M\}$ ;
  - an initial probability distribution for each state  $\pi = [\pi_1, \pi_2, \dots, \pi_N]$ ;
  - a transition matrix  $\mathbf{A} = \{a_{ij}\}$ , with  $i, j \in [1, N]$ ;
  - and an emission matrix  $\mathbf{B} = \{b_{nm}\}$  with  $m \in V$ , in some state  $n \in S$ .
- Briefly the parameter set of an HMM is denoted as  $\lambda = (S, V, \mathbf{A}, \mathbf{B}, \pi)$ .

In order to be able to use an HMM for image classification, we can establish a temporal order that corresponds uniquely with visual features located in the image. This is possible if we fragment each image into regions of constant size and always generate observations with the same trajectory. Since each image will be different in brightness and perspective, elevation angle, and rotation angle (using as reference the object captured in each image), an HMM is ideal for finding a statistical description of the image.

**3.1. Image Feature Extraction.** We propose to find a representation automatically using clustering techniques. For this, fixed-size windows  $w_{n,m}^t$  are established on the image, and from these, vectors are obtained by concatenating the three channels of the image, forming an observation vector  $o_t$  of size  $1 \times 3 \times n \times m$ . Figure 1 shows these sections.

A visual alphabet is built in an unsupervised way using the mini-batch k-means clustering method [22] for all vectors  $o_t$  that are selected as training samples. The clusters obtained assign an index to each vector  $o_t$  according to a similarity metric (euclidean distance is chosen). The quantization process has the objective of eliminating noise, obtaining centroids representing the average of the content in each window, which eliminates disturbances.

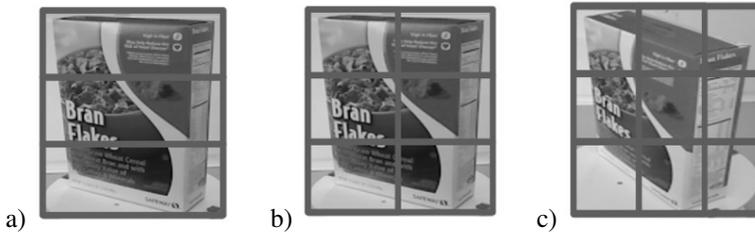


Fig. 1. Example of the sections created to obtain visual features (images from [23]). a) Three windows with size  $n = 74$ ,  $m = 224$ ; b) Six windows with size  $n = 74$ ,  $m = 112$ ; c) Nine windows with size  $n = 74$ ,  $m = 74$

This method enables an automatic but distinctive visual content representation between different objects, quickly and with compact representation because only clusters centroid list is stored. Then, each vector can be statistically treated using an HMM to generate a probabilistic model for each image

class. The trained classes generate a higher probability, and observations of other classes or untrained examples generate a lower probability or zero tendencies. The next section explains this procedure.

**3.2. HMMs for Feature Detection.** In computer vision, Hidden Markov Models (HMMs) are a useful tool because they model the appearance of symbols  $o_{1:T}$  that we want to use as comparison patterns. These patterns can be built by associating a sequence of windows  $w_{n,m}^{t:T}$  on a image at each instant  $t$ , so we can say that the sequence  $\{o_1, o_2, \dots, o_T\} = \{w_{n,m}^{t=1}, w_{n,m}^{t=2}, \dots, w_{n,m}^{t=T}\}$  represents a complete unit. Where  $m, n$  are the window dimensions, and the variable  $t$  assign a fixed path on the image as time increases.

The architecture selected to model the sequences is the so-called ergodic HMM. This structure is used since it allows to obtain the probability of transition between all states, avoiding constraints on model dynamics. The Figure 2 shows this configuration between states. The number of states  $T$  is associated with the number of symbols that describe each image, and this number can be arbitrarily long, but to reduce processor usage time in an autonomous robotic system, it's preferable to use a small number of symbols in each sequence. This also makes it easy to scale a detection system to new classes without degrading response time.

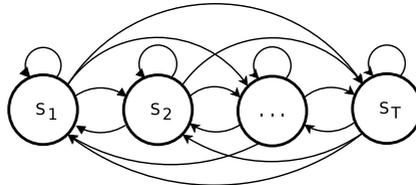


Fig. 2. An example of ergodic HMM

The detection system used is based on a set of HMMs, where each one model  $\lambda_k$  has been trained with examples of sequences that correspond to images of a single class  $k$ . This allows learning a distribution of these sequences and assigning a single model. In the end, given a test sequence  $o_{1:T}$ , each model  $\lambda_k$  is evaluated, and the one with the highest probability is the one that best explains the dynamics of the sequence. This is defined as

$$k \{P(o_{1:T} | \lambda_k)\}. \tag{6}$$

Figure 3 illustrates this procedure. This method is efficient since inference can be made in parallel as well as training.

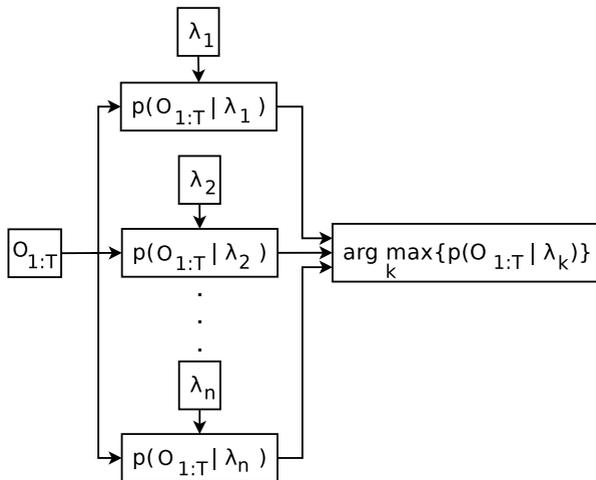


Fig. 3. Block diagram of the inference system using HMMs

The following section shows how it's possible to obtain a classification with enough precision to compete with pre-trained and fine-tuned CNNs, maintaining equally competitive training and inference times, using only a laptop processor.

**4. Experiments and Results.** For experiments, the Hidden Markov Model (HMM) shown in Figure 2 has been used. This architecture has been chosen since there is no predefined model that describes transitions between each window  $w_{n,m}^i$  created in each image, so a transition between all the states is allowed. This architecture is used with  $T = 3$ ,  $T = 6$  and  $T = 9$ .

The data set published in [23] is used since the objective is to test our method in a context of a mobile service robot, where objects to train and detect are those that could be found inside a house, distributed in different classes.

This data set is made up of 51 classes of objects and each class has a variable number of instances, but only one instance of each class is used in this work.

If necessary, a detection system based on HMMs can be scaled to new classes without the need for retraining but simply by adding new models  $\lambda_k$  to the structure shown in the Figure 3. This applies equally to instances of the same objects.

The data set images were captured with a fixed RGB-D camera, a rotating base and each class present an unbalanced number of images. Images are captured with an object rotation between 0 and up to 720 degrees, so

they are sub-sampled in half at regular intervals to obtain an object coverage of approximately 0 to 360 degrees. This subset also contains images with an elevation angle of 30 and 60 degrees with respect to the object, different lighting and perspective. Figure 4 shows an example of the training images used.

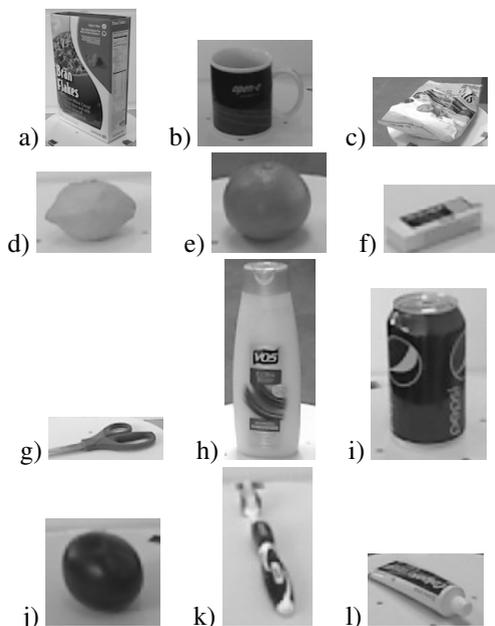


Fig. 4. Twelve examples of objects contained in the data set for testing [23]: (a cereal\_box; b) coffe\_mug; c) food\_bag; d) lemon; f) orange; g) rubber\_eraser; h) scissors; i) shampoo; j) soda\_can; k) tomato; k) toothbrush; l) toothpaste. All classes are listed in the first column of Table 9 to Table 13

For the tests three recognition system architectures have been used. The first architecture is based on work of Rabiner [11] and is shown in Figure 5. The second and third architectures are shown respectively in the Figure 14 and Figure 15, the latter are a composition of the first architecture in what is known as an ensemble method. This is described in Section 4.3.

For the first architecture, a rescaling is applied to each image in the training set, and this new scale is 224 by 224 pixels. Then each image is transformed into the CIE Lab color space [24] and three, six, and nine sections of the image were created.

For the case where there are three windows in the image, vectors of size  $1 \times (74 \times 224 \times 3)$  are obtained, where  $w_{74,224}^t$  are the size of the window. The factor  $\times 3$  appears since there are three channels in the image corresponding to channels  $L$ ,  $a$  y  $b$ . In the case of extracting six windows with size  $w_{74,112}^t$ , there are vectors of size  $1 \times (74 \times 112 \times 3)$ . For nine windows with size  $w_{74,74}^t$  the vectors are of dimension  $1 \times (74 \times 74 \times 3)$ .

To quantize all vectors scikit-learn library was used [25], it incorporates optimizations based on openMP, the use of AVX2 instructions (Advanced Vector Extensions 2) and LaPACK (Linear Algebra PACKage) library support. Mini batch k-means algorithm [22] is used, and it's initialized with k-means++ for faster convergence [26]. The number of centroids was predefined with values 64, 128, 256 and 512. Furthermore, mini-batch k-means is implemented using the Elkan algorithm [27], this algorithm uses triangular inequalities, so it's not necessary to search for a new vector through all centroids and mini-batch k-means avoids doing searches using all data (it samples mini-batches speeding up calculations). The quantization process is performed offline.

The symbol sequences are obtained by associating an identifier index of the centroids with a vector to be compared, which generates a sequence  $o_{1:T}$  for each image. These sequences are generated to train each model  $\lambda_k$  or during the inference process. Each model is trained by each category of objects in the data set and subsequently stored as a later use file. In the end, the model  $\lambda_k$  that has the highest probability calculated given the sequence  $o_{1:T}$  is the one that assigns the detected class. The architecture of Figure 5 shows the proposed detection system.

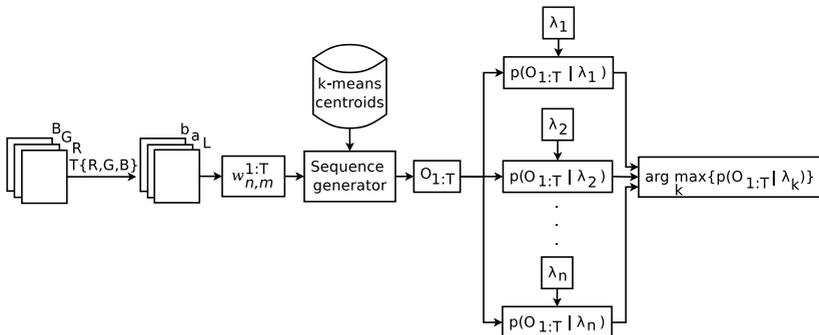


Fig. 5. Block diagram of the recognition system. A transformation from RGB to CIE Lab color space, construction of sections in the image, and construction of symbol sequences to be subsequently detected by a set of previously trained HMMs

The experiments were organized as follows:

- an HMM with three states for each class of images. A quantizer for 64, 128, 256, and 512 centroids;
- an HMM with six states for each class of images. A quantizer for 64, 128, 256, and 512 centroids;
- an HMM with nine states for each class of images. A quantizer for 64, 128, 256, and 512 centroids.

The experiments were performed five times, using the subset of images that cover the object from approximately 0 to 360 degrees, partitioned into 75% for training and 25% for testing. In the case of CNNs, 10 % of the data set is taken for training validation but from the section for testing.

For the training set, all images are randomly preprocessed with one of the following transformations: rotations of up to 25 degrees, horizontal and vertical displacement of 10% with respect to the center of the original image, skew transformation of 20% on the horizontal and vertical axis, reflections in the horizontal and vertical axis and zoom of up to 10%. The validation set (only for CNNs) and testing set (HMMs and CNNs) does not include transformations. All this in order to compare performance against CNNs typically trained with data augmentation techniques.

A laptop with 32 GB of RAM, a model GeForce 1070 with 8 GB GPU, and a 4-core seventh-generation CPU was used for testing. The results are reported as weighted Average Precision (wAP) and weighted Average Recall (wAR) due to imbalance in the number of images between classes. Accuracy, precision, recall, and F1-score are also reported [28].

Weighted Average Precision is defined as follows:

$$\frac{1}{\sum_{i=1}^K \hat{y}_i} \sum_{i=1}^K \hat{y}_i P_i. \quad (7)$$

Weighted Average Recall is defined as:

$$\frac{1}{\sum_{i=1}^K \hat{y}_i} \sum_{i=1}^K \hat{y}_i R_i. \quad (8)$$

The index  $K$  is the total number of classes and  $\hat{y}_i$  is the total number of images labeled as class  $i$ .

Precision  $P_i$  and Recall  $R_i$  are defined for each class  $i$  as:

$$P_i = \frac{t_p}{t_p + f_p}. \quad (9)$$

$$R_i = \frac{t_p}{t_p + f_n}. \quad (10)$$

Where  $t_p$  means true positives,  $f_p$  false positives and  $f_n$  false negatives.

**4.1. HMMs Results.** Figure 6 shows the qualification result for 51 object classes. Of 5 experiments carried out, weighted Average Precision (wAP) is reported, this since each class of objects contains a different number of images, so imbalance must be taken into account, weighting the precision for each class with respect to the number of classified images. In the same way, Figure 7 shows the detection average for weighted Average Recall (wAR).

The highest percentage of recognition is carried out by HMMs using 512 symbols as an alphabet, as is shown in Figure 6. The use of three states to form models  $\lambda_k$  results in a wAP value equal to 0,97, but wAR does not increase above one hundredth from models with six and nine states. Therefore, using three states increases the detection of positive cases but does not increase for negative cases, as shown in Figure 7.

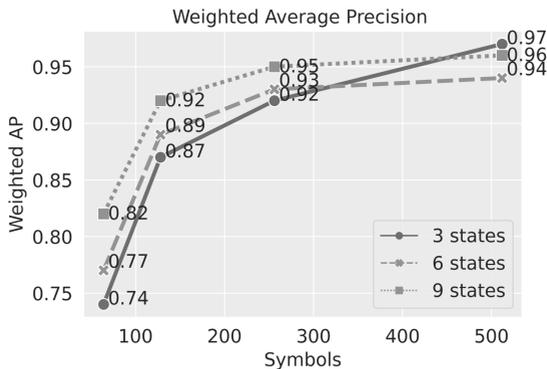


Fig. 6. Weighted Average Precision for all HMMs

Figure 8 shows the processing time used to obtain an alphabet using the mini-batch k-means method. For the case of 512 symbols and three states, the peak processing time is shown close to 220 seconds, this is due to the fact that the chains with three states use the largest vector sizes, so the clustering process is slower.

Table 1 in row one shows each vector's average query time obtained from each window. The average time used to obtain the sequence of symbols of each image is included in row two. Only the case for 512 symbols is shown, since it's the case with the slowest access of all the symbol variants. The query

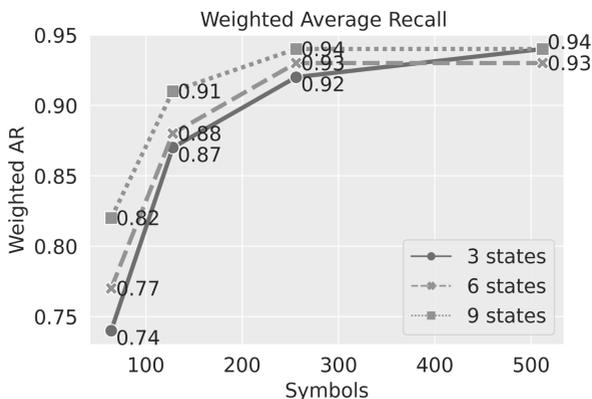


Fig. 7. Weighted Average Recall of all HMMs

process reuses the Elkan method, so on average, the time of 2.7 milliseconds in the worst case is never exceeded.

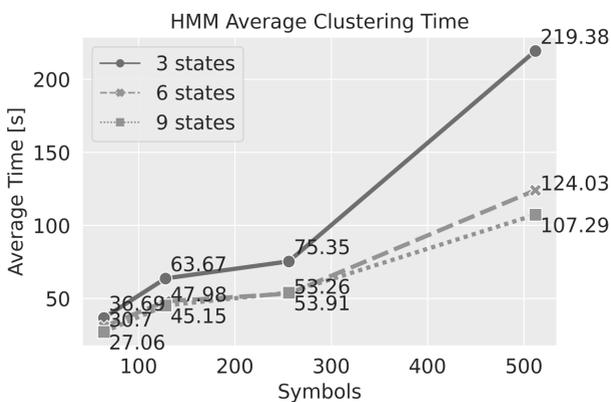


Fig. 8. Average clustering time for HMMs

Table 1. Average Query Time for 512 symbols

	3 windows	6 windows	9 windows
single vector	0,00088 s	0,000425 s	0,000288 s
single image	0,002639 s	0,002551 s	0,00259 s

In the case of training time, HMMs with the highest number of states, such as those with six and nine states, are the ones that take the longest, this is shown in Figure 9. Even so, training time does not exceed 0,200 seconds in the worst case, on average for each HMM.

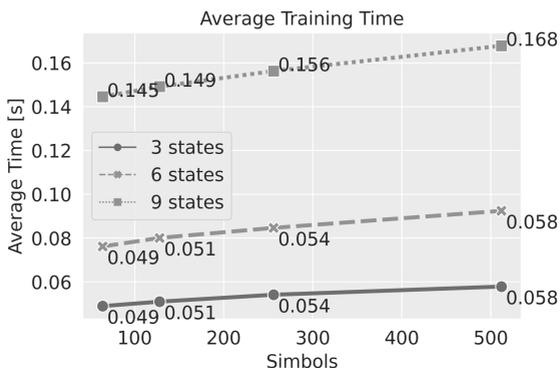


Fig. 9. Average training time for HMMs

Inference time depends on the number of states and the number of transitions between them. Times range from approximately  $2.54 \times 10^{-6}$  seconds to  $16.4 \times 10^{-6}$  seconds as shown in Figure 10.

Less number of symbols generates similarity between learned distributions, so it's more likely to go through many states until the end of a complete sequence, this increases inference time for 64 and 128 symbols. On the contrary, more symbols generate more distinctive observations, with more differentiated statistics, so that when evaluating an input sequence, fewer states are passed, resulting in a shorter inference time for 256 and 512 symbols. It's useful to see how fast an image can be classified using HMMs. The inference time do not exceed  $20 \times 10^{-6}$  seconds on average in any HMM variant and number of symbols.

The amount of memory used in critical parts of the system was monitored. Creating an alphabet using larger image sections (as was done in the case of three states in each HMM) has as a consequence a high consumption of system memory because all the data must be stored for the queries reaching a peak of approximately eight giga bytes of occupation as shown in Figure 11.

In comparison, HMMs with six and nine states use smaller vectors that do not exceed the size of two point one gigabytes of memory.

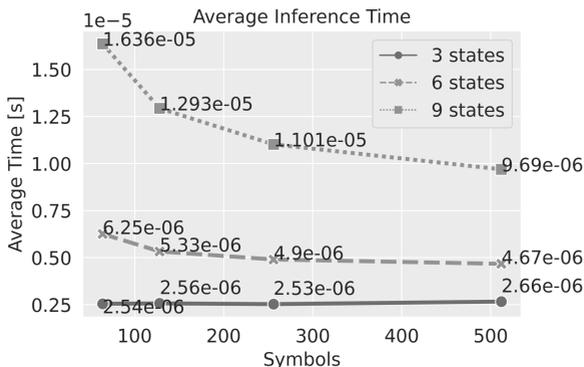


Fig. 10. Average inference time for HMMs

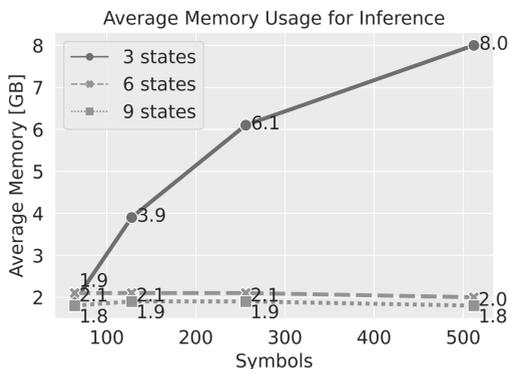


Fig. 11. Average memory usage for inference with HMMs

For the memory used in the clustering process, again the HMMs with three states demand storage to build the corresponding alphabet as seen in Figure 12. It reaches the memory occupancy peak of up to twenty-two point eight gigabytes, which is a disadvantage. The Figure 6 shows that this three-state model with 512 symbols is the one with the highest wAP but requires excess memory for both inference and clustering.

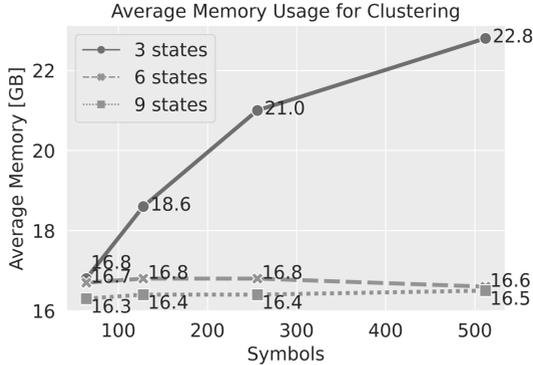


Fig. 12. Average memory usage for clustering

In Section 4.3, we will explain how to solve the memory consumption problem by discarding models with high memory occupancy and high computation time to form an HMMs ensemble.

**4.2. CNNs Results.** In the case of Convolutional Neural Networks (CNNs), two variants were used. The first called VGG-19 [19] and the second ResNet-50 [20]. These are chosen due to their high level of accuracy in tests carried out in the ImageNet challenge, in addition to the fact that previously trained coefficients are in the public domain included in the latest version of the library named TensorFlow [29].

VGG-19 and ResNet-50 were tested by freezing pre-trained coefficients obtained from ImageNet data set. Only classification layers were removed, and two hidden and fully connected layers were added with 256 neurons. A softmax classification layer for 51 classes is also added and data augmentation is used as described in Section 4.

These architectures were designed to classify millions of images, but they perform well in this compact data set. Table 2 shows the weighted Average Precision (wAP) in first row, weighted Average Recall (wAR) is reported in second row. VGG-19 achieves a wAP of 1,0 and a wAR of 1,0. On the other hand, Resnet-50 achieves 0,96 in wAP and 0,94 in wAR through all five experiments.

Table 2. CNNs Average Results

	VGG-19	ResNet-50
Weighted Average Precision	1,0	0,96
Weighted Average Recall	1,0	0,94

For the case of training time and inference time, due to the massive amount of parameters, they turn out to be slower than any HMM. Table 3 shows in row one that VGG-19 has an average inference time per image close to 0,35 seconds, and an average training time of 1882.6 seconds is reported in row two. For ResNet-50 inference, time is 0,183 seconds, and an average training time of 2846.8 seconds is shown in the second column.

Table 3. Time for Training and Inference for CNNs

	VGG-19	ResNet-50
Average Inference Time	0,3498 s	0,183 s
Average Training Time	1882.6 s	2846.8 s

The memory consumption of these classifiers is composed of occupation in GPU and system memory. The two architectures use the same amount of system and GPU memory during training and inference, as reported in Table 4.

Table 4. Memory usage for CNNs

	VGG-19	ResNet-50
GPU	7.8 GB	7.8 GB
System	2.8 GB	2.8 GB

The Figure 13 shows a wAP and wAR comparison of the best HMMs with respect to VGG-19 and ResNet-50 at the end of the experiments. In the case of wAP, VGG-19 gets the best test performance, scores 1,0 in 5 experiments. Behind this classifier is the HMM with three states and 512 symbols. The HMM with nine states and 512 symbols and ResNet-50, the latter with similar capacity to recognize positive and negative images of each class. The six-state, 512-symbol HMM lags behind all with a wAP equal to 0,94.

In the case of wAR, Figure 13 shows each value with a triangular marker. ResNet-50 and HMMs with three and nine states obtain a value equal to 0,94, and the HMM with six states obtain 0,93. From these results, it can be concluded that CNNs perform well in detecting positive and negative examples. However, HMMs are more likely to make mistakes in positive examples classified as negative, which decreases the recall value.

**4.3. HMMs Ensemble.** During the experiments, 51 HMMs were trained, these also in four varieties of symbols and three types of architectures, giving a total of 612 trained HMMs (51x4x3 models). So the question arises: Is it possible to reuse some of these HMMs to create a more robust classification system? The answer is developed below.

First, we must establish that not all HMMs are useful for this purpose since models with three states make extensive use of memory and processing

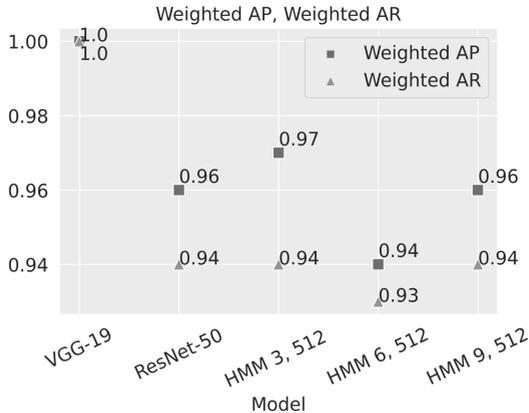


Fig. 13. Weighted Average Precision and Weighted Average Recall for best models. HMM 3, 512 for three states and 512 symbols. HMM 6, 512 for six states and 512 symbols. HMM 9, 512 for nine states and 512 symbols

time in their variants of 512 symbols. However, models with nine states and 512 symbols are the most accurate. Thus, HMMs with less memory use and, if possible, with greater precision and recall were selected. The only HMMs that simultaneously fulfill these two characteristics are the chains of short length or small observation vectors, the set of models with three states and 64 symbols, and the models with nine states and 512 symbols.

This results in using three types of HMMs, two types of symbols (64 and 512), and 51 models for classification, so it's proposed to make an ensemble of 306 HMMs. This might seem implausible, but because HMMs are compact in memory during the inference process in addition to a speed in the order of microseconds, 306 sequentially evaluated models must have a lower execution time than a CNN like VGG-19 or ResNet-50. The number of parameters for VGG-19 is approximately 143.6 million and for ResNet-50 it's 25.6 million for the ImageNet data set. For the case of our HMMs ensemble, it can be calculated taking into account the following: the size of each vector (having three, six, and nine windows) and the number of symbols to store (64 and 512 symbols). This is  $[74 \times 224 \times 3 \times (64 + 512)] + [74 \times 112 \times 3 \times (64 + 512)] + [74 \times 74 \times 3 \times (64 + 512)]$  which results in approximately 52.4 million parameters to store the centroids. In the case of 306 HMMs, we have less than 0,55 million parameters counting each matrix  $\mathbf{A}_k$ ,  $\mathbf{B}_k$  and  $\pi_k$  in each model  $\lambda_k$ .

It's important to note that the data set to obtain the extraction of visual features in CNNs is enormous, close to 1,3 million images (ImageNet data set), which makes its direct application difficult in systems that do not have such a large amount of data. The HMMs allow training with a compact data set, and in a modular way since the performance of a classification system is increased by adding or removing models  $\lambda_k$ , adding or removing lists of centroids obtained in the clustering process, and there is no dependence between models during their training. On the contrary, in the case of a CNN, retraining is necessary to add or remove classes.

To construct the ensembles, the approach of a soft and hard classifier was used. The soft classifier calculates a joint probability on all models already trained and selects a class with the highest joint probability. This system is shown in Figure 14. The hard classifier works as HMMs do, selecting the model with the highest probability but with the difference that now one vote per class is accumulated. The most voted class is the selected class, Figure 15 shows this system.

Table 5 shows in the first row the wAP result of the ensembles. These two new classifiers achieve a wAP value of 0,98. For the case of wAR which is shown in the second row, the classifier with soft-decision lags behind with a wAR of 0,95.

Thus, it's confirmed that it's possible to ensemble HMMs with a different number of hidden states to increase the overall performance. Ensemble 2 obtains a wAP equal to 0,98 and a wAR equal to 0,98, lagging behind VGG-19 with scores equal to 1,0.

Table 5. Results for HMMs Ensemble

	HMM Ensemble 1	HMM Ensemble 2
Weighted Average Precision	0.98	0.98
Weighted Average Recall	0.95	0.98

Inference time when assembling the models is triggered, reaching values of up to 0,0022 seconds for the best ensemble case. This is reported in Table 6. The HMM ensemble, even evaluated sequentially, is faster than an average of 0,183 seconds of inference per image in ResNet-50.

Table 6. Average Inference Time

	HMM Ensemble 1	HMM Ensemble 2
Average Inference Time	$4.410^{-5}$ s	0,002237 s

The memory usage increases because it's necessary to load all the centroids used by each HMM, but even with this, it only increases the memory

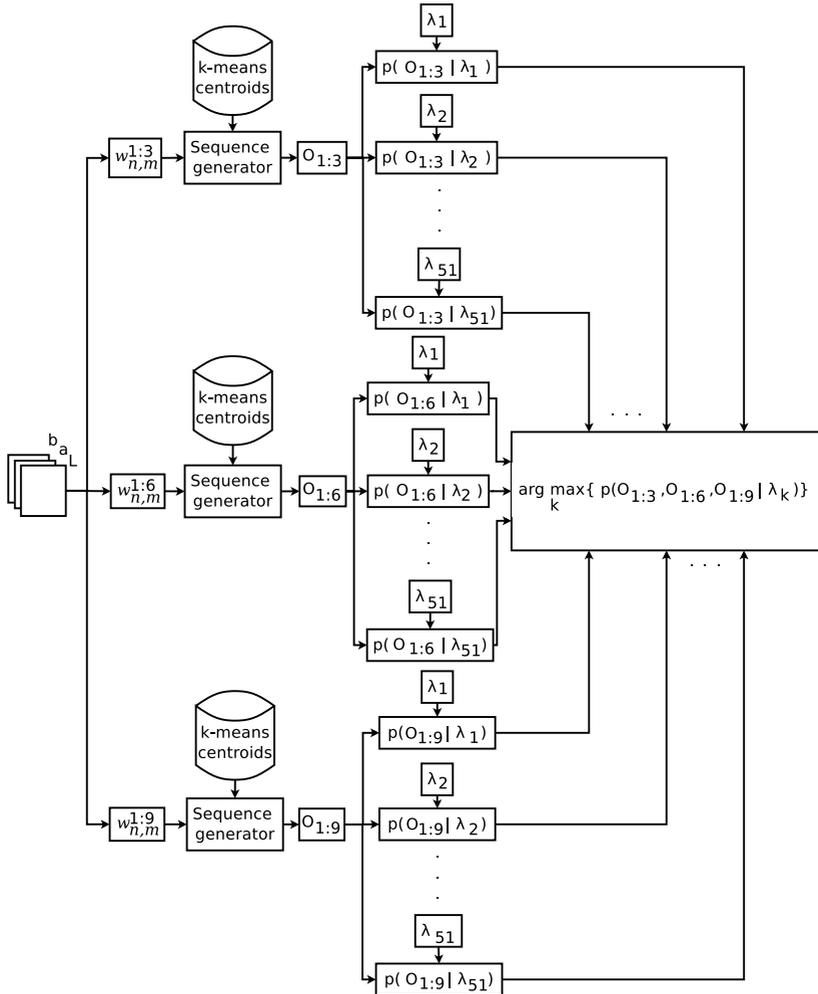


Fig. 14. Ensemble 1. A soft classifier based on HMMs

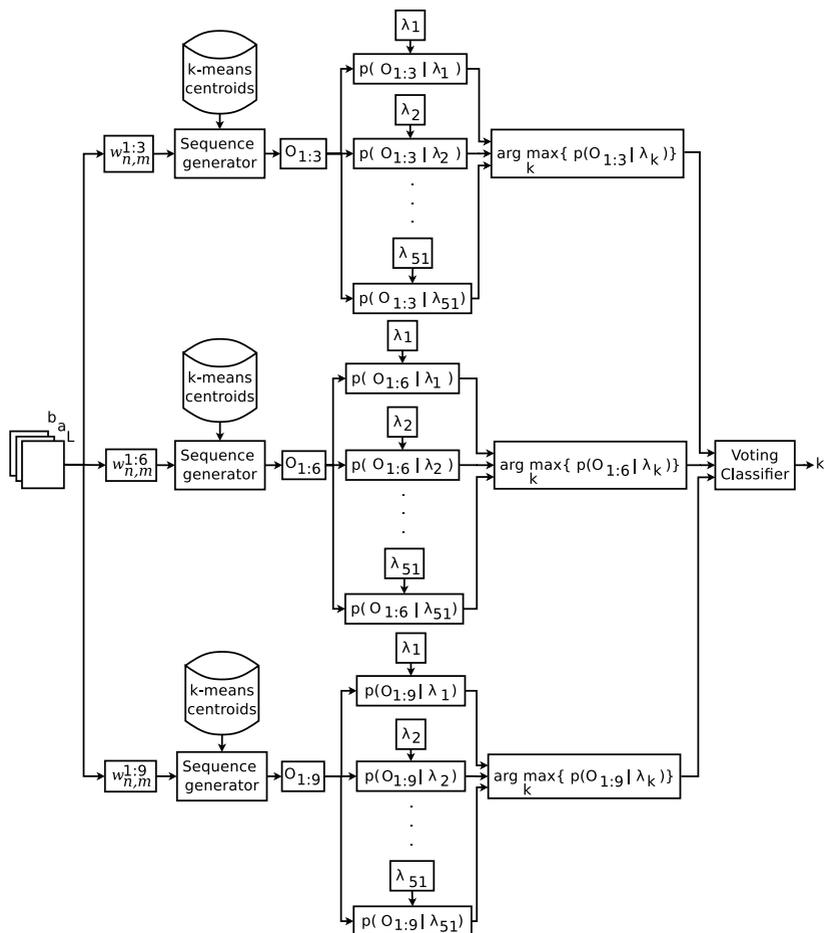


Fig. 15. Ensemble 2. A hard classifier based on HMMs

usage up to a limit of two point eight gigabytes on average, as reported in the Table 7 (a 32-bit representation is used as well as CNNs parameters).

Table 7. Average Memory Usage

	HMM Ensemble 1	HMM Ensemble 2
Average Memory Usage	2.8 GB	2.8 GB

Finally, a summary of wAP and wAR is shown in Figure 16 for the best models evaluated in this work. VGG-19 is the system with the most accuracy, precision and best recall, scoring 1,0 in all metrics. Table 9 presents the full report.

HMMs Ensemble 2 ranks second among the best classifiers, but this by far uses generic resources such as memory and CPU and turns out to be faster than VGG-19 and ResNet-50, verifying that HMMs are an effective tool for image classification (see Table 13 for full report).

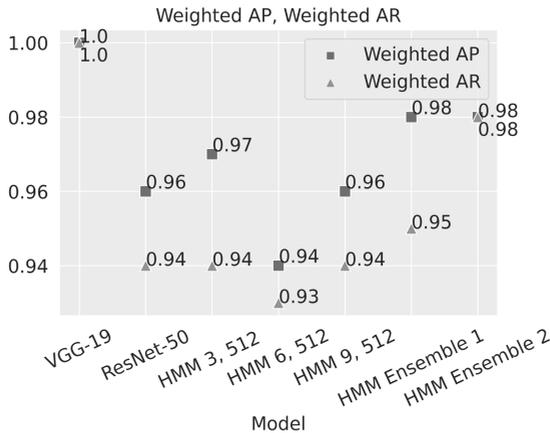


Fig. 16. Final metrics of best classifiers. HMM 3, 512 for three states and 512 symbols. HMM 6, 512 for six states and 512 symbols. HMM 9, 512 for nine states and 512 symbols

The other image classifiers developed have similar performance in the task. ResNet-50 has a wAP with 0,96 value, Table 10 shows the complete classification report.

The HMM of nine states and 512 symbols is reported with detail in Table 11. This is the model with less resource consumption and better detection metrics simultaneously. Also, the Table 12 shows the performance for ensemble 1. The HMM report with three states and 512 symbols is skipped

due to its high memory consumption, and HMMs with six states and 512 symbols due to its low precision.

**5. Conclusions.** In this paper, a practical comparison was presented using publicly available data. It has been demonstrated experimentally that the application of the theory of HMMs is useful to create a classification system for images containing objects. This task has been widely developed with the use of deep convolutional neural networks but these models, even with great precision, are complex to integrate into systems that require high energy efficiency due to the millions of parameters that must be adjusted and then transported to a final program.

Table 8 summarizes the three best models tested, their values of wAP, wAR, inference time for an image, and the number of parameters. Row three for the case of HMM Ensemble 2 is the time required to obtain the complete inference on a 224x224 size image. This time is the sum of 6 average queries with 512 symbols, which correspond to three different architectures of HMMs and two different types of symbols.

Table 8. Results comparison

	HMM Ensemble 2	VGG-19	ResNet-50
wAP	0,98	1,0	0,96
wAR	0,98	1,0	0,94
Avg. Time per Image	0,018071 s <sup>1</sup>	0,3498 s	0,183 s
Number of Parameters	52.95 millions <sup>2</sup>	143.6 millions	25.6 millions

Weighted Average Precision (wAP). Weighted Average Recall (wAR). <sup>1</sup> Query time for an image in 6 centroid lists with 512 symbols, plus run all 306 models. <sup>2</sup> All centroids for 64 and 512 symbols and 306 HMMs

The HMMs ensemble 2 had the best value of wAP and wAR of all proven HMMs, but without reaching the performance of VGG-19. Its main advantage is that although with the increase of parameters, it still achieves an execution speed ten times higher than ResNet-50 using a moderate performance CPU that only has four cores.

All the models achieve a good recognition capacity, equal to or greater than 0,94 in wAP also for wAR, but it's possible to build an ensemble with half of these models to reach up to 0,98 for both wAP and wAR. This result means a reasonable ability to distinguish between positive and negative examples. Note that tests demonstrate the power of the method using data augmentation and an unbalanced data set per class. Of course, a disadvantage of the method is that it requires the construction of a visual vocabulary using unsupervised methods, so it's not possible to modify the representation online. Using a more compact and distinctive visual representation, in addition to a segmentation subsystem, will be the reason for future work.

Table 9. Classification report for VGG-19

Class	Precision	Recall	F1-score	Support
apple	1,00	1,00	1,00	380
ball	1,00	1,00	1,00	515
banana	1,00	1,00	1,00	460
bell_pepper	1,00	1,00	1,00	395
binder	1,00	1,00	1,00	445
bowl	1,00	1,00	1,00	365
calculator	1,00	1,00	1,00	365
camera	1,00	1,00	1,00	395
cap	1,00	1,00	1,00	395
cell_phone	1,00	1,00	1,00	345
cereal_box	1,00	1,00	1,00	360
coffee_mug	1,00	1,00	1,00	345
comb	1,00	1,00	1,00	355
dry_battery	1,00	1,00	1,00	355
flashlight	1,00	1,00	1,00	380
food_bag	1,00	1,00	1,00	495
food_box	1,00	1,00	1,00	495
food_can	1,00	1,00	1,00	515
food_cup	1,00	1,00	1,00	490
food_jar	1,00	1,00	1,00	485
garlic	1,00	1,00	1,00	490
gluestick	1,00	1,00	1,00	505
greens	1,00	1,00	1,00	455
hand_towel	1,00	1,00	1,00	490
instant_noodles	1,00	1,00	1,00	485
keyboard	1,00	1,00	1,00	445
kleenex	1,00	1,00	1,00	490
lemon	1,00	1,00	1,00	370
lightbulb	1,00	1,00	1,00	390
lime	1,00	1,00	1,00	400
marker	1,00	1,00	1,00	505
mushroom	1,00	1,00	1,00	490
notebook	1,00	1,00	1,00	500
onion	1,00	1,00	1,00	485
orange	1,00	1,00	1,00	450
peach	1,00	1,00	1,00	425
pear	1,00	1,00	1,00	425
pitcher	1,00	1,00	1,00	355
plate	1,00	1,00	1,00	450
pliers	1,00	1,00	1,00	370
potato	1,00	1,00	1,00	380
rubber_eraser	1,00	1,00	1,00	495
scissors	1,00	1,00	1,00	390
shampoo	1,00	1,00	1,00	495
soda_can	1,00	1,00	1,00	395
sponge	1,00	1,00	1,00	400
stapler	1,00	1,00	1,00	370
tomato	1,00	1,00	1,00	425
toothbrush	1,00	1,00	1,00	375
toothpaste	1,00	1,00	1,00	515
water_bottle	1,00	1,00	1,00	350
Average	1,00	1,00	1,00	21900
Weighted Avg	1,00	1,00	1,00	21900
Accuracy			1,00	21900

Table 10. Classification report for ResNet-50

Class	Precision	Recall	F1-score	Support
apple	1,00	0,93	0,97	380
ball	1,00	0,63	0,77	515
banana	1,00	0,96	0,98	460
bell_pepper	1,00	1,00	1,00	395
binder	1,00	0,80	0,89	445
bowl	1,00	0,79	0,89	365
calculator	0,97	1,00	0,99	365
camera	1,00	1,00	1,00	395
cap	1,00	1,00	1,00	395
cell_phone	0,95	1,00	0,97	345
cereal_box	1,00	0,93	0,96	360
coffee_mug	1,00	1,00	1,00	345
comb	1,00	0,93	0,96	355
dry_battery	1,00	1,00	1,00	355
flashlight	1,00	1,00	1,00	380
food_bag	1,00	1,00	1,00	495
food_box	1,00	0,99	0,99	495
food_can	0,95	1,00	0,98	515
food_cup	0,96	0,89	0,92	490
food_jar	1,00	1,00	1,00	485
garlic	1,00	0,89	0,94	490
glue_stick	0,99	1,00	1,00	505
greens	1,00	1,00	1,00	455
hand_towel	0,99	0,99	0,99	490
instant_noodles	1,00	1,00	1,00	485
keyboard	1,00	0,98	0,99	445
kleenex	0,99	1,00	0,99	490
lemon	0,96	0,96	0,96	370
lightbulb	0,91	1,00	0,95	390
lime	0,96	0,95	0,96	400
marker	1,00	0,86	0,93	505
mushroom	0,80	1,00	0,89	490
notebook	1,00	1,00	1,00	500
onion	1,00	0,46	0,63	485
orange	0,66	1,00	0,80	450
peach	1,00	0,72	0,84	425
pear	1,00	0,79	0,88	425
pitcher	1,00	0,94	0,97	355
plate	1,00	1,00	1,00	450
pliers	1,00	1,00	1,00	370
potato	1,00	0,58	0,73	380
rubber_eraser	1,00	1,00	1,00	495
scissors	0,90	1,00	0,95	390
shampoo	1,00	0,98	0,99	495
soda_can	0,83	1,00	0,91	395
sponge	0,81	1,00	0,89	400
stapler	0,97	1,00	0,99	370
tomato	0,44	1,00	0,61	425
toothbrush	0,85	1,00	0,92	375
toothpaste	1,00	0,99	1,00	515
water_bottle	1,00	1,00	1,00	350
Average	0,96	0,94	0,94	21900
Weighted Avg	0,96	0,94	0,94	21900
Accuracy			0,94	21900

Table 11. Classification report for HMM with 9 states and 512 symbols

Class	Precision	Recall	F1-score	Support
apple	0,33	0,99	0,50	380
ball	0,99	0,95	0,97	515
banana	0,98	0,91	0,95	460
bell_pepper	0,99	0,97	0,98	395
binder	0,99	0,97	0,98	445
bowl	1,00	0,99	1,00	365
calculator	0,85	0,81	0,83	365
camera	0,87	0,89	0,88	395
cap	0,82	0,91	0,86	395
cell_phone	0,96	0,94	0,95	345
cereal_box	0,95	0,88	0,91	360
coffee_mug	0,99	0,90	0,94	345
comb	0,94	0,86	0,90	355
dry_battery	0,99	0,86	0,92	355
flashlight	0,88	0,82	0,85	380
food_bag	0,97	0,94	0,96	495
food_box	1,00	0,93	0,96	495
food_can	1,00	0,97	0,98	515
food_cup	0,99	0,98	0,99	490
food_jar	1,00	0,96	0,98	485
garlic	0,96	0,98	0,97	490
glue_stick	1,00	0,98	0,99	505
greens	0,99	0,92	0,96	455
hand_towel	0,99	0,94	0,96	490
instant_noodles	0,95	0,94	0,94	485
keyboard	0,99	0,84	0,91	445
kleenex	1,00	0,92	0,96	490
lemon	1,00	0,98	0,99	370
lightbulb	0,98	0,91	0,94	390
lime	1,00	0,99	1,00	400
marker	0,95	0,89	0,92	505
mushroom	0,99	0,92	0,95	490
notebook	0,98	0,93	0,95	500
onion	0,99	1,00	0,99	485
orange	1,00	0,99	1,00	450
peach	1,00	0,98	0,99	425
pear	1,00	0,99	1,00	425
pitcher	0,97	0,94	0,96	355
plate	1,00	1,00	1,00	450
pliers	0,96	0,85	0,90	370
potato	1,00	0,98	0,99	380
rubber_eraser	0,96	0,96	0,96	495
scissors	0,90	0,83	0,86	390
shampoo	0,99	0,96	0,98	495
soda_can	1,00	0,95	0,98	395
sponge	1,00	0,99	1,00	400
stapler	0,85	0,84	0,84	370
tomato	1,00	0,98	0,99	425
toothbrush	0,95	0,93	0,94	375
toothpaste	0,92	0,89	0,90	515
water_bottle	0,98	0,99	0,99	350
Average	0,96	0,93	0,94	21900
Weighted Avg	0,96	0,94	0,94	21900
Accuracy			0,94	21900

Table 12. Classification report for HMMs Ensemble 1

Class	Precision	Recall	F1-score	Support
apple	0,27	1,00	0,43	380
ball	1,00	0,97	0,98	515
banana	1,00	0,92	0,96	460
bell_pepper	1,00	0,98	0,99	395
binder	1,00	0,98	0,99	445
bowl	1,00	0,99	1,00	365
calculator	1,00	0,80	0,89	365
camera	0,97	0,90	0,93	395
cap	0,87	0,93	0,90	395
cell_phone	1,00	0,93	0,96	345
cereal_box	0,99	0,91	0,95	360
coffee_mug	1,00	0,91	0,95	345
comb	0,99	0,86	0,92	355
dry_battery	1,00	0,86	0,93	355
flashlight	0,97	0,84	0,90	380
food_bag	1,00	0,96	0,98	495
food_box	1,00	0,94	0,97	495
food_can	1,00	0,97	0,99	515
food_cup	1,00	0,99	0,99	490
food_jar	1,00	0,96	0,98	485
garlic	1,00	0,97	0,99	490
glue_stick	1,00	0,99	1,00	505
greens	1,00	0,92	0,96	455
hand_towel	1,00	0,97	0,98	490
instant_noodles	1,00	0,93	0,96	485
keyboard	1,00	0,88	0,94	445
kleenex	1,00	0,96	0,98	490
lemon	1,00	0,99	1,00	370
lightbulb	0,99	0,95	0,97	390
lime	1,00	0,99	1,00	400
marker	1,00	0,92	0,96	505
mushroom	1,00	0,97	0,98	490
notebook	0,99	0,93	0,96	500
onion	1,00	0,99	1,00	485
orange	1,00	0,99	1,00	450
peach	1,00	0,98	0,99	425
pear	1,00	0,99	0,99	425
pitcher	1,00	0,99	0,99	355
plate	1,00	1,00	1,00	450
pliers	1,00	0,84	0,91	370
potato	1,00	0,98	0,99	380
rubber_eraser	1,00	0,96	0,98	495
scissors	0,98	0,83	0,90	390
shampoo	1,00	0,97	0,98	495
soda_can	1,00	0,96	0,98	395
sponge	1,00	1,00	1,00	400
stapler	0,94	0,90	0,92	370
tomato	1,00	0,99	1,00	425
toothbrush	0,98	0,94	0,96	375
toothpaste	0,99	0,93	0,96	515
water_bottle	1,00	0,99	0,99	350
Average	0,98	0,95	0,96	21900
Weighted Avg	0,98	0,95	0,96	21900
Accuracy			0,95	21900

Table 13. Classification report for HMMs Ensemble 2

Class	Precision	Recall	F1-score	Support
apple	0,97	1,00	0,98	380
ball	1,00	1,00	1,00	515
banana	0,99	1,00	0,99	460
bell_pepper	1,00	1,00	1,00	395
binder	1,00	1,00	1,00	445
bowl	1,00	1,00	1,00	365
calculator	0,98	0,98	0,98	365
camera	0,88	0,93	0,90	395
cap	0,73	0,86	0,79	395
cell_phone	0,97	1,00	0,98	345
cereal_box	0,97	0,97	0,97	360
coffee_mug	1,00	0,97	0,98	345
comb	0,99	0,96	0,97	355
dry_battery	1,00	0,95	0,97	355
flashlight	0,94	0,88	0,91	380
food_bag	0,99	1,00	0,99	495
food_box	1,00	0,99	1,00	495
food_can	1,00	1,00	1,00	515
food_cup	0,99	1,00	1,00	490
food_jar	1,00	1,00	1,00	485
garlic	0,91	1,00	0,95	490
glue_stick	1,00	1,00	1,00	505
greens	1,00	1,00	1,00	455
hand_towel	1,00	1,00	1,00	490
instant_noodles	0,99	1,00	1,00	485
keyboard	1,00	0,75	0,85	445
kleenex	1,00	0,98	0,99	490
lemon	1,00	1,00	1,00	370
lightbulb	0,90	0,95	0,92	390
lime	0,99	1,00	0,99	400
marker	0,99	0,99	0,99	505
mushroom	0,99	0,91	0,95	490
notebook	1,00	1,00	1,00	500
onion	1,00	1,00	1,00	485
orange	1,00	1,00	1,00	450
peach	1,00	1,00	1,00	425
pear	0,98	1,00	0,99	425
pitcher	0,99	0,98	0,99	355
plate	1,00	1,00	1,00	450
pliers	0,99	0,98	0,99	370
potato	1,00	0,99	0,99	380
rubber_eraser	0,98	1,00	0,99	495
scissors	0,97	0,98	0,98	390
shampoo	1,00	1,00	1,00	495
soda_can	1,00	1,00	1,00	395
sponge	1,00	1,00	1,00	400
stapler	0,88	0,88	0,88	370
tomato	1,00	1,00	1,00	425
toothbrush	0,94	0,98	0,96	375
toothpaste	0,98	0,98	0,98	515
water_bottle	0,99	1,00	1,00	350
Average	0,98	0,98	0,98	21900
Weighted Avg	0,98	0,98	0,98	21900
Accuracy			0,98	21900

## References

1. Dhall D., Kaur R., Juneja M. Machine Learning: A Review of the Algorithms and Its Applications. Proceedings of International Conference on Recent Innovations in Computing (ICRIC'2019). 2019. pp. 47–63.
2. Dhruv P., Naskar S. Image Classification Using Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN): A Review. International Conference on Machine Learning and Information Processing (ICMLIP'2019). 2020. pp. 367–381,
3. Wang W et al. Development of convolutional neural network and its application in image classification: a survey. *Optical Engineering*. 2019. vol. 58. no. 4. pp. 1–19.
4. Gambella C., Ghaddar B., Naoum-Sawaya J. Optimization problems for machine learning: A survey. *European Journal of Operational Research*. Available at: <https://doi.org/10.1016/j.ejor.2020.08.045> (accessed: 20.09.2020).
5. Dogo E.M. et al. A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks. International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS'2018). 2018. pp. 92–99.
6. Belanche D., Casal L.V., Flavián C., Schepers J. Service robot implementation: a theoretical framework and research agenda. *The Service Industries Journal*. 2020. vol. 40 no. 3-4. pp. 203–225.
7. Torras C. Service Robots for Citizens of the Future. *European Review*. 2016. vol. 24. no. 1. pp. 17–30.
8. Matamoros M. et al. Robocup at home 2019: Rules and regulations. Available at: <http://www.robocupathome.org/rules/2019rulebook.pdf> (accessed: 20.09.2020).
9. Matamoros M., Rascon C., Wachsmuth S., Moriarty A.W., Kummert J., Hart J, Pfeiffer S., van der Brugh M., St-Pierre M. Robocup at home 2019: Rules and regulations. Available at: [http://www.robocupathome.org/rules/2019\\_rulebook.pdf](http://www.robocupathome.org/rules/2019_rulebook.pdf). (accessed 20.09.2020).
10. Baum L.E., Petrie T., Soules G., Weiss N. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics*. 1970. vol. 41. no.1. pp. 164–171.
11. Rabiner L.R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Readings in Speech Recognition. Morgan Kaufmann Publishers Inc. 1990. pp. 267–296.
12. Mor B., Garhwal S., Kumar A. A Systematic Review of Hidden Markov Models and Their Applications. Archives of Computational Methods in Engineering. Available at: <https://doi.org/10.1007/s11831-020-09422-4>. (accessed: 20.09.2020).
13. Corcoran P., Iancu C. Hidden Markov Models in Automatic Face Recognition – A Review. *Reviews, Refinements and New Ideas in Face Recognition*. Available at: <https://doi.org/10.5772/17664> (accessed: 20.09.2020).
14. Rastghalam R., Pourghassem H. Breast cancer detection using MRF-based probable texture feature and decision-level fusion-based classification using HMM on thermography images. *Pattern Recognition*. 2016. vol. 51. pp. 176–186.
15. Hassan M. et al. Robust Hidden Markov Model based intelligent blood vessel detection of fundus images. *Computer Methods and Programs in Biomedicine*. 2017. vol. 151. pp. 193–201.
16. Sarmiento C. et al. Feature detection using Hidden Markov Models for 3D-visual recognition. IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC'2019). 2019. pp. 1–6.
17. Russakovsky O. et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*. 2015. vol. 115. no. 3. pp. 211–252.

18. Krizhevsky A., Sutskever I., Hinton G. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems (NIPS'2012)*. 2012. pp. 1097–1105.
19. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations (ICLR'2015)*. 2015
20. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2016)*. 2016. pp. 770–778.
21. Howard A.G. et al. MobileNets:Efficient Convolutional Neural Networks for Mobile Vision Applications. *CoRR*. Available at: <http://arxiv.org/abs/1704.04861> (accessed: 20.09.2020).
22. Sculley D. Web-scale k-means clustering. *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. 2010. pp. 1177–1178.
23. Lai K., Bo L., Ren X., Fox D. A large-scale hierarchical multi-view RGB-D object dataset. *IEEE International Conference on Robotics and Automation (ICRA'2011)*. 2011. pp. 1817–1824.
24. Sharma G., Wu W., Dalal E.N. The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Research and Application*. 2005. vol. 30. no. 1. pp. 21–30.
25. Pedregosa F. et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*. 2011. vol. 12. no. 85. pp. 2825–2830.
26. Arthur D., Vassilvitskii S. K-means++: The advantages of careful seeding. *ACM-SIAM Symposium on Discrete Algorithms (SODA'07)*. 2007. pp. 1027–1035.
27. Elkan C. Using the triangle inequality to accelerate k-means. *Proceedings of the 20th International Conference on Machine Learning (ICML'03)*. 2003. pp. 147–153.
28. Marina S.M., Lalpalme G. A systematic analysis of performance measures for classification tasks. *Information Processing Management*. 2009. vol. 45. no. 4 pp. 427–437.
29. Abadi M. et al. TensorFlow: Large-scale machine learning on heterogeneous systems. Available at: [www.tensorflow.org](http://www.tensorflow.org) (accessed: 20.09.2020).

**Sarmiento Carlos** — PhD student, Laboratory of Biorobotics, National Autonomous University of Mexico (UNAM). Research interests: computer vision, reinforcement learning, autonomous navigation. The number of publications — 2. [ing.adriansarmiento@comunidad.unam.mx](mailto:ing.adriansarmiento@comunidad.unam.mx); Circuito Exterior S/N, Ciudad Universitaria, 04510, Mexico City, Mexico; office phone: +52(55)56223041.

**Savage Jesus** — Ph.D., Professor, Professor, Electrical Engineering Department, National Autonomous University of Mexico (UNAM). Research interests: rch interests: autonomous mobile robots, digital signal processing, computer architectures. The number of publications — 100. [robotssavage@gmail.com](mailto:robotssavage@gmail.com); Circuito Exterior S/N, Ciudad Universitaria, 04510, Mexico City, Mexico; office phone: +52(55)56223041.

**Acknowledgements.** This research is supported by UNAM-CONACYT.

К. САРМЬЕНТО, Х. САВАЖ  
**СРАВНЕНИЕ ДВУХ МЕТОДОВ КЛАССИФИКАЦИИ ОБЪЕКТОВ  
С ИСПОЛЬЗОВАНИЕМ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ  
И СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ**

*Сармьенто К., Саваж Х. Сравнение двух методов классификации объектов с использованием скрытых марковских моделей и сверточных нейронных сетей.*

**Аннотация.** Представлено сравнение дискретных скрытых марковских моделей и сверточных нейронных сетей для классификации изображений. После разбивки изображений на части целесообразно получить векторы, которые представляют локальные визуальные структуры, одновременно определяющие изображения глобально через пространственную последовательность. С использованием методов кластеризации создается алфавит из указанных векторов, а затем конструируются последовательности символов, которые описывают статистические модели, соответствующие классам изображений. Скрытые марковские модели в сочетании с методами квантования могут обрабатывать шум и искажения в наблюдениях для решения проблем компьютерного зрения, таких как классификация изображений с изменением освещения и перспективы.

Протестированы архитектуры, основанные на трех, шести и девяти скрытых состояниях, в пользу скорости обнаружения и низкого использования памяти. Также были протестированы два типа ансамблевых моделей. Точность предлагаемого метода была оценена с помощью общедоступных данных; полученные результаты оказались сравнимы с известными оценками при использовании тонко настроенных сверточных нейронных сетей, но требовали значительно меньших вычислительных ресурсов. Результат представляет интерес при разработке мобильных роботов с вычислительными устройствами, имеющими ограниченное время автономной работы, но требующими способности обнаруживать и добавлять новые объекты в свои системы классификации.

**Ключевые слова:** скрытые марковские модели, классификация изображений, компьютерное зрение, распознавание образов

**Сармьенто Карлос** — аспирант, лаборатория биороботики, Национальный автономный университет Мексики (НАУМ). Область научных интересов: компьютерное зрение, машинное обучение с подкреплением, автономная навигация. Число научных публикаций — 2. [ing.adriansarmiento@comunidad.unam.mx](mailto:ing.adriansarmiento@comunidad.unam.mx); Циркуито Экстериор б/н, Университетский город, 04510, Мехико, Мексика; р.т.: +52(55)56223041.

**Саваж Хесус** — канд. техн. наук, профессор, профессор, электротехнический отдел, Национальный автономный университет Мексики (НАУМ). Область научных интересов: автономные мобильные роботы, цифровая обработка сигналов, архитектуры компьютеров. Число научных публикаций — 100. [robotssavage@gmail.com](mailto:robotssavage@gmail.com); Циркуито Экстериор б/н, Университетский город, 04510, Мехико, Мексика; р.т.: +52(55)56223041.

**Поддержка исследований.** Данное исследование поддержано UNAM-CONACYT.

### Литература

1. *Dhall D., Kaur R., Juneja M. Machine Learning: A Review of the Algorithms and Its Applications // Proceedings of International Conference on Recent Innovations in Computing (ICRIC'2019). 2019. pp. 47–63.*

2. *Dhruv P., Naskar S.* Image Classification Using Convolutional Neural Net-work (CNN) and Recurrent Neural Network (RNN): A Review // International Conference on Machine Learning and Information Processing (ICMLIP'2019). 2020. pp. 367–381.
3. *Wang W et al.* Development of convolutional neural network and its application in image classification: a survey // Optical Engineering. 2019. vol. 58. no. 4. pp. 1–19..
4. *Gambella C., Ghaddar B., Naoum-Sawaya J.* Optimization problems for ma-chine learning: A survey // European Journal of Operational Research. Availa-ble at: <https://doi.org/10.1016/j.ejor.2020.08.045> (accessed: 20.09.2020).
5. *Dogo E.M. et al.* A Comparative Analysis of Gradient Descent-Based Optimization Algorithms on Convolutional Neural Networks // International Confer-ence on Computational Techniques, Electronics and Mechanical Systems (CTEMS'2018). 2018. pp. 92–99.
6. *Belanche D., Casal ´o L.V., Flavi ´an C., Schepers J.* Service robot implementation: a theoretical framework and research agenda // The Service Industries Journal. 2020. vol. 40 no. 3-4. pp. 203–225.
7. *Torras C.* Service Robots for Citizens of the Future // European Review. 2016. vol. 24. no. 1. pp. 17–30.
8. *Zachiotis G.A. et al.* A Survey on the Application Trends of Home Service Robotics // IEEE International Conference on Robotics and Biomimetics (ROBIO'2018). 2018. pp. 1999–2006
9. *Matamoros M. et al.* Robocup at home 2019: Rules and regulations. Available at: <http://www.robocupathome.org/rules/2019rulebook.pdf> (accessed: 20.09.2020).
10. *Baum L.E., Petrie T., Soules G., Weiss N.* A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains // The Annals of Mathematical Statistics. 1970. vol. 41. no.1. pp. 164–171.
11. *Rabiner L.R.* A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Readings in Speech Recognition // Morgan Kaufmann Publishers Inc. 1990. pp. 267–296.
12. *Mor B., Garhwal S., Kumar A.* A Systematic Review of Hidden Markov Models and Their Applications. Archives of Computational Methods in Engineering. Available at: <https://doi.org/10.1007/s11831-020-09422-4>. (accessed: 20.09.2020).
13. *Corcoran P., Iancu C.* Hidden Markov Models in Automatic Face Recognition – A Review. Reviews, Refinements and New Ideas in Face Recognition. Avail-able at: <https://doi.org/10.5772/17664> (accessed: 20.09.2020).
14. *Rastghalam R., Pourghassem H.* Breast cancer detection using MRF-based probable texture feature and decision-level fusion-based classification using HMM on thermography images // Pattern Recognition. 2016. vol. 51. pp. 176–186.
15. *Hassan M. et al.* Robust Hidden Markov Model based intelligent blood vessel detection of fundus images // Computer Methods and Programs in Biomedicine. 2017. vol. 151. pp. 193–201.
16. *Sarmiento C. et al.* Feature detection using Hidden Markov Models for 3D-visual recognition // IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC'2019). 2019. pp. 1–6.
17. *Russakovsky O. et al.* ImageNet Large Scale Visual Recognition Challenge // International Journal of Computer Vision. 2015. vol. 115. no. 3. pp. 211–252.
18. *Krizhevsky A., Sutskever I., Hinton G.* Imagenet classification with deep convolutional neural networks // Neural Information Processing Systems (NIPS'2012). 2012. pp. 1097–1105.
19. *Simonyan K., Zisserman A.* Very Deep Convolutional Networks for Large-Scale Image Recognition // International Conference on Learning Representations (ICLR'2015). 2015.

20. *He K., Zhang X., Ren S., Sun J.* Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2016). 2016. pp. 770–778.
21. *Howard A.G. et al.* MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. CoRR. Available at: <http://arxiv.org/abs/1704.04861> (accessed: 20.09.2020).
22. *Sculley D.* Web-scale k-means clustering. Proceedings of the 19th International Conference on World Wide Web (WWW '10). 2010. pp. 1177–1178.
23. *Lai K., Bo L., Ren X., Fox D.* A large-scale hierarchical multi-view RGB-D object dataset. IEEE International Conference on Robotics and Automation (ICRA'2011). 2011. pp. 1817–1824.
24. *Sharma G., Wu W., Dalal E.N.* The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. Color Research and Application. 2005. vol. 30. no. 1. pp. 21–30.
25. *Pedregosa F. et al.* Scikit-learn: Machine learning in python. Journal of Machine Learning Research. 2011. vol. 12. no. 85. pp. 2825–2830.
26. *Arthur D., Vassilvitskii S.* K-means++: The advantages of careful seeding. ACM-SIAM Symposium on Discrete Algorithms (SODA'07). 2007. pp. 1027–1035.
27. *Elkan C.* Using the triangle inequality to accelerate k-means. Proceedings of the 20th International Conference on Machine Learning (ICML'03). 2003. pp. 147–153.
28. *Marina S.M., Lapalme G.* A systematic analysis of performance measures for classification tasks. Information Processing Management. 2009. vol. 45. no. 4 pp. 427–437
29. *Abadi M. et al.* TensorFlow: Large-scale machine learning on heterogeneous systems. Available at: [www.tensorflow.org](http://www.tensorflow.org) (accessed: 20.09.2020).

N.H. PHAT, N.T.M. ANH  
**VIETNAMESE TEXT CLASSIFICATION ALGORITHM USING  
LONG SHORT TERM MEMORY AND WORD2VEC**

*Phat N.H., Anh N.T.M. Vietnamese Text Classification Algorithm using Long Short Term Memory and Word2Vec.*

**Abstract.** In the context of the ongoing forth industrial revolution and fast computer science development the amount of textual information becomes huge. So, prior to applying the seemingly appropriate methodologies and techniques to the above data processing their nature and characteristics should be thoroughly analyzed and understood. At that, automatic text processing incorporated in the existing systems may facilitate many procedures. So far, text classification is one of the basic applications to natural language processing accounting for such factors as emotions' analysis, subject labeling etc. In particular, the existing advancements in deep learning networks demonstrate that the proposed methods may fit the documents' classifying, since they possess certain extra efficiency; for instance, they appeared to be effective for classifying texts in English. The thorough study revealed that practically no research effort was put into an expertise of the documents in Vietnamese language. In the scope of our study, there is not much research for documents in Vietnamese. The development of deep learning models for document classification has demonstrated certain improvements for texts in Vietnamese. Therefore, the use of long short term memory network with Word2vec is proposed to classify text that improves both performance and accuracy. The here developed approach when compared with other traditional methods demonstrated somewhat better results at classifying texts in Vietnamese language. The evaluation made over datasets in Vietnamese shows an accuracy of over 90%; also the proposed approach looks quite promising for real applications.

**Keywords:** Text Classification, Natural Language Processing, Data Processing, Long short term memory, Word2Vec

**1. Introduction.** Automatic text classification is intended for processing new documents based on their similarity with the other ones in the training model. In this paper the text classification algorithm is proposed aimed at solving certain issues (classifying topics and positive-negative comments) based on titles. The following example is given to explain the text classification. At that, the news dataset is selected and represented as:

$$N = (n_1, n_2, \dots, n_n). \quad (1)$$

The news corresponding labels are:

$$C = (c_1, c_2, \dots, c_m). \quad (2)$$

The articles labeling will be performed automatically according to label (C).

The traditional methods often classify documents based on dictionaries. However, the development of deep learning (DL) models has been much more effective and used widely for classifying documents [1, 2].

Besides, the methods for Vietnamese language are quite limited. The most difficult point for processing documents in Vietnamese is to determine word boundaries. In English language words are groups of meaningful characters separated by spaces in sentences. Therefore, it is not difficult to select in a sentence the words in English language. However, word boundaries are not defined for Vietnamese and depend on context of sentences. Several key properties are:

- independent syntax;
- single, complex, repeated, and compound words;
- phrase clustering.

Figure 1 gives an example of Vietnamese phrase clustering [3]; and shows that there exist two ways to understand this sentence. If the second way of words clustering is accepted, the sentence will have not the same meaning as the method.

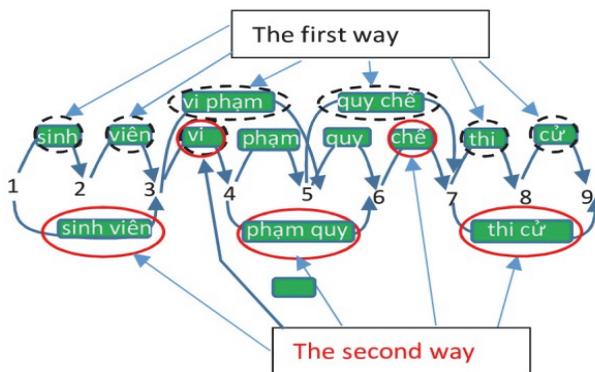


Fig. 1. An example of clustering words in Vietnamese language

Words' clustering in sentences is an important step in preprocessing documents in Vietnamese. If the sentence is identified by the first method, it would be classified as a law. Otherwise, certain misunderstanding would lead to another label. Therefore, the accuracy of phrase clustering is very important. If the clustering of words is not reliable, the label classification of text may be wrong.

In the paper, the long short-term memory (LSTM) model is used at first to classify text since gates of LSTM are able to filter information similar to input ports. Therefore, the data from the past will be adjusted.

Besides, here also proposed to combine LSTM and Word2vec methods to improve the method accuracy.

The rest of the paper is organized as follows. Section 2 describes the related works; Section 3 presents the theoretical basis; Section 4 shows the results of the model use; Section 5 contains conclusions and some new directions proposed for further development.

**2. Related Work.** The amount of text data must be processed prior to their use. The problem is how to extract information from the data source. The data nature and their characteristics should be clearly understood in order to apply the necessary methodologies. Therefore, time and effort for their classifying will be saved.

DL models have appeared to be quite successful at natural language processing (NLP) [3-12]. Applications of DL to NLP are mentioned as phrase classification algorithm [13] and the main content of Vietnamese text [14].

Currently, there exist many methods to improve English texts classification based on artificial neural (AI) networks [15-25]. In [15] the authors use new LSTM model to classify text. The model with prior training is able to solve the multi-dimensional data processing problem by traditional methods. In [20] the authors use convolutional neural network (CNN) and recurrent neural network (RNN) for classification. In [16] the authors combine two models CNN and bidirectional recurrent neural network (BRNN). In the model the authors use the bidirectional class to replace the pooling class in CNN and help to store the long-term dependencies of input chains.

In addition, the support vector machine (SVM) algorithm is used to reduce required memory [17]. SVM algorithm is well applicable to the text classification problem. It saves memory since only a subset of points is used for training and forming process for new data. Therefore, the necessary points are stored while making decisions in a result of / for decision-making. SVM possesses certain flexibility due to change between linear and non-linear methods. However, it does not clearly indicate the calculus of probability yet, since its classification only focuses on dividing objects into two layers by super-flat.

Many studies have been performed [18, 26-28] in regard to Vietnamese texts classification. Classification of texts in Vietnamese via traditional methods and topics is often based on [18]. Topic model uncovers-abstract documents. This method guarantees stability and provides a relative accuracy. However, it is quite difficult and time-consuming as well as costly. Therefore, SVM algorithm is applied to classifying texts in Vietnamese [26]. This classification is able to adjust the parameters automatically. However, it displays worse results comparing with an

application to English language (the accuracy for Vietnamese language is 80.72% comparing with 89% for English [17]).

In [27] the authors used traditional methods like Naïve Bayes (NB) and maximum entropy (ME). NB is a simple method to solve problems regarding data classification based on statistic. However, its disadvantage is that there is no link between the characteristics. Therefore, ME is used to estimate label probability based on sentence characteristics. Besides, in [27] the authors also use DL method as LSTM and Bi-LSTM for Vietnamese language classification. In [28] the authors employ a combination of LSTM and CNN for separating words in a sentence. This approach helps the model to learn both adjacent and distant data. To the best of our knowledge, the classification of words by subject using LSTM and Word2vec has not been published before.

Based on the analyzed above, the topic based DL model for text classification could be proposed since it has demonstrated certain improvements for classifying documents in Vietnamese. In the model the LSTM and Word2vec methods are combined. As a result, the model improved both accuracy and speed processing for input of data in Vietnamese language.

### 3. System

**3.1. The System Overview.** An overview of the system is shown in Figure 2. The system consists of two main blocks, namely training and predicting parts. In two blocks, a combination of two algorithms Word2vec and LSTM is proposed to increase accuracy for classification of texts in Vietnamese. Besides, the preprocessing step suitable for data in Vietnamese is used. The steps details are given in the following sections.

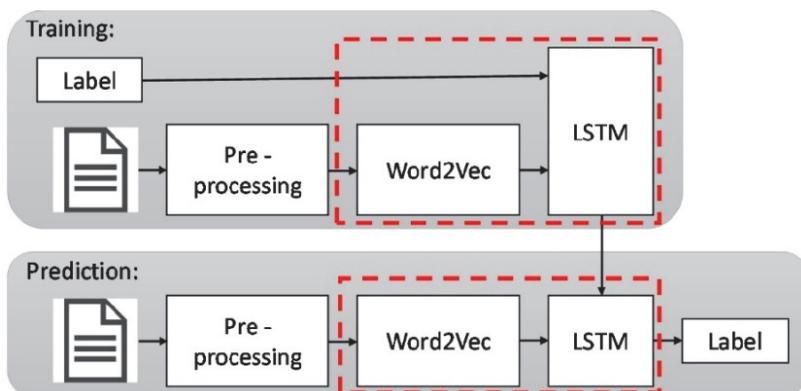


Fig. 2. Diagram of system structure

**3.2. Data Collection.** The difficulty for classifying is that the dataset of topics in Vietnamese is limited. Datasets usually consist of 50 to 100 raw

texts. In this paper the VNTC dataset [4] suitable for the done research is used. The dataset is updated and information from official electronic newspapers like [29-32] is chosen by users according to each topic. It contains 10 topics of 33756 and 50373 articles for training and testing respectively. Data are selected and aggregated as shown in Figures 3 and 4.

### DATA FOR TRAINING ON 10 TOPICS

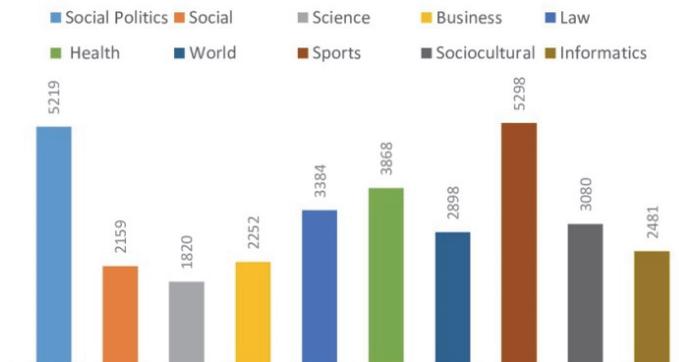


Fig. 3. Result of dataset training [4]

### DATA FOR TESTING ON 10 TOPICS

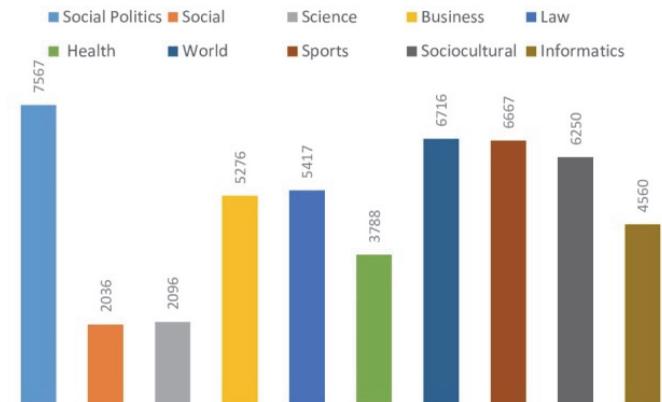


Fig. 4. Result of dataset testing [4]

**3.3. Data Preprocessing.** At the preprocessing step three small steps namely Vietnamese words separation, data cleaning, stop-word separation are executed as follows.

*Vietnamese words separation.* As was mentioned in Part 1, the separation of Vietnamese words greatly affects the output results. In recent years the Vietnamese language processing community has grown stronger. There appeared many libraries related to NLP for Vietnamese language that have produced good results [18, 26-28].

Table 1 shows the accuracy of methods for separating Vietnamese words. In the current paper the *vnTokenizer* and *Python* language [33] are used. To separate words the Vietnamese *Pyvi* library for VNTC dataset [33] is used. This library is a combination of maximum matching (MM) algorithm and SVM model to solve both ambiguity and unknown words recognition to improve the shortcomings for other methods. It proposes a language classifier for Vietnamese as follows: O – single word for one language, B – first language for one multi-language word, I – intermediate language for one multi-language word, and E – final language for one multilingual word. MM is considered to be the simplest dictionary based on splitting word; it attempts to match the longest word dictionary. However, this method is not able to solve the problem of ambiguity since it only recognizes words in dictionary. Therefore, the combined system of MM + SVM has reduced the ambiguity by contextual words. In [33] the library has an accuracy of 97,86% for Vietnamese words.

Table 1. Accuracy of Vietnamese words separation methods

Tokenizer	F1-Score (%)
RDRsegmente [34]	97,90
iPTDP-v2 [35]	97,90
UETsegmenter [10]	97,87
vnTokenizer [3]	98,5
JvnSegmenter [36]	97,06

*Data cleaning.* After separating words the text reveals many special characters and punctuation that simplify the system. To solve the problem, all uppercase letters were converted into lowercase and punctuation was removed.

*Stop-word separation.* Then the stop-words are to be removed. Stop-words are understood as words not important for classification. Besides, a number of concatenation or quantitative words is not discriminatory at their classifying. In addition, stop-words have no taxonomic values that appear in most documents. Therefore, these words are eliminated to reduce computation time and memory during the processing. At this step the stop-words Vietnamese dictionary [37] are used. Upon these words removing the text will be considered a set of the representation important words.

**3.4. Extracting Feature by Word2vec.** After removing stop-words the text with the important words will be received. However, there are still many documents' characteristics, and they have to be shorten.

Then words will be vectorized. The vector representation of words plays an important role. Word embedding is responsible for mapping from word or phrase to a real number vector.

In this section the Word2vec model is used. The model represents words as real vectors with specified dimensions. It is an unsupervised learning model, and is one of the first models of word embedding that uses vector each word based on contexts; it maps a set of words to a vector space where each of them is represented by  $n$  real numbers. Word2vec is the neural network with only one hidden layer. Input is a large set of words and output are vector spaces. Each single word is assigned a corresponding vector as shown in Figure 5 [38, 39].

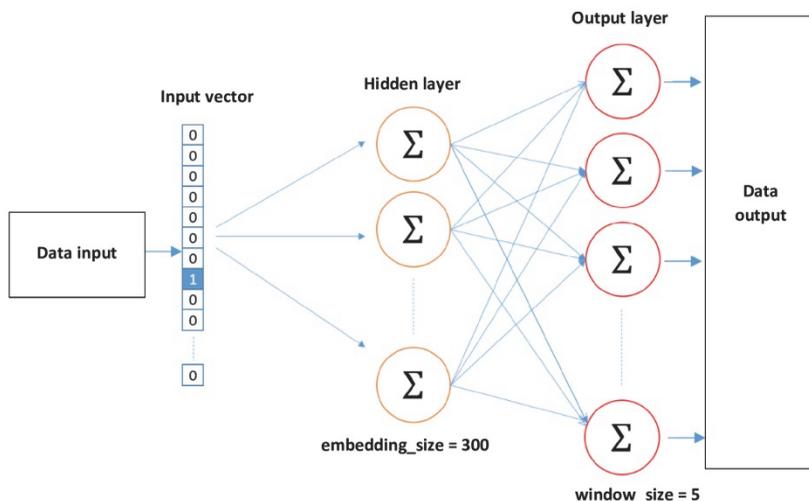


Fig. 5. Proposal of Word2vec model with  $embedding\_size = 300$  and  $window\_size = 5$

**The model of input is one-hot-vector:** Each word will be in a form  $x_1, x_2, \dots, x_v$ , where  $v$  is the number of vocabulary. Each word is a vector with a value of “1” equivalent to the order of words in vocabulary and the rest will be of “0” value.

The matrix between the input and hidden layer is  $W$  (its dimension is  $V \times N$ ) whose activation function is linear. The matrix between the hidden layer and the output is  $W'$  (its dimension is  $V \times N$ ) whose activation function is *softmax*.

Each row of  $W$  is  $N$ -dimensional vector representing  $V_w$ . Each row of  $W$  is  $v_w^T$ .

Output matrix of the hidden layer is  $W' = w'_{i,j}$ . The score for each word is calculated as follows:

$$u_j = v'_{wj} * h, \tag{3}$$

where  $v'_{wj}$  is  $j$  column of  $W'$  matrix. Then the *softmax* trigger function is used as follows:

$$P(w_j / w_I) = y_i = \frac{\exp(u_j)}{\sum_{j'=1}^v \exp(u_{j'})} = \frac{\exp(v'_{wj} v_{WI})}{\sum_{j'=1}^v \exp(v'^T W_{j'} v_{WI})}, \tag{4}$$

where  $v_w$  and  $v_{w'}$  are two vectors.

After training model, the weight of each word is updated continuously.

Therefore, the calculations can be performed by distances. The words appear together in context or synonyms belonging to the same vocabulary.

Word2vec algorithm has two approaches (as shown in Fig. 6) as follows:

- continuous bag of words (CBOW) model gives context (surrounding words) and guesses the appearing probability from destination;
- skip-gram model gives the current word and guesses the probability of context words.

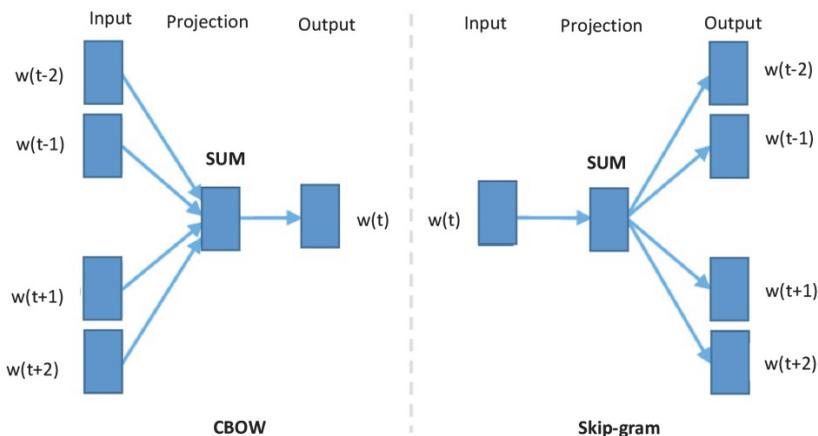


Fig. 6. Two approaches of word2vec algorithm [40]

The neural network architecture of skip-gram and CBOW consists of three layers:

1. *Input* is the input of network and context words around a target or a present word.

2. *Projection* contains the hidden layers of network for parameters calculating.

3. *Output* is a softmax function that calculates the probability for target words or distributes into vocabulary. Based on the feed and back propagation models, the methods would be found to optimize the parameters to predict the results accuracy.

CBOW advantage is that it requires less memory to store large matrices. It is essentially probabilistic since implementation is better than Skip-gram. However, the disadvantage is that words with different meanings are still represented by one vector of words.

**3.5. Labeling by LSTM.** RNN for NLP has been of interest lately [15-25]. The main idea of RNN is to use sequences of information. In traditional neural networks all inputs and outputs are independent of each other. If inputs go through the hidden layer and the outputs connect among the classes, they will be combined via a function to calculate the current and output layers.

The RNN calculation is performed as follows:

–  $x_t$  is the input at t that is one-hot vector corresponding to size of  $n \times 1$ ;

–  $s_t$  is the hidden state at t that is calculated based on both the pre-hiding state and input. The front words will affect the output as follows:

$$s_t = f(Ux_t + Ws_{t-1}). \quad (5)$$

The function  $f$  is usually a linear function as tang hyperbolic (tanh). The function is used to adjust the information passing through the system. All values are assigned to the range (-1, 1). When vectors go through neural network, they undergo many calculations. In the process, several components become too large. The function will help to refine the difference. To make calculations for the first element,  $s_{t-1}$  to 0 or random values should be initialized. The memory is empty without data;

–  $o_t$  is the output at t that is a probability vector of predicting words by learning information from all previous inputs as follows:

$$o_t = g(Vs_t), \quad (6)$$

where  $g$  is the activation function.

The RNN is described as shown in Figure 7.

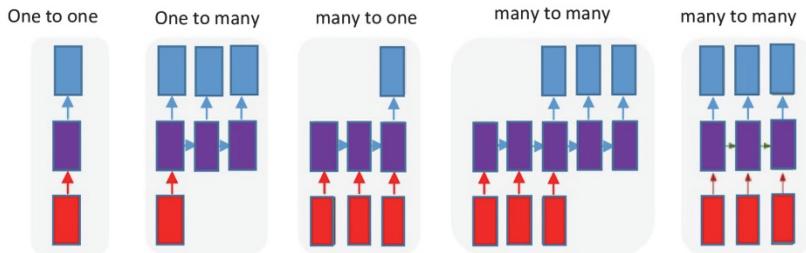


Fig. 7. Types of problems of RNN [41]

– *one to one* is for neural network (NN) and CNN with one input and output. For example, the input is the image and the output identifies whether it is a motorcycle;

– *one to many* is a problem with one input and many outputs. For example, the input is one image and the output is many descriptive words.

– *many to one* has many inputs and one output. For example, the input is multiple images from video and the output is one action.

– *many to many* has many inputs and outputs, for example translating from English into Vietnamese.

LSTM is an extending version of RNN designed to solve long-term dependencies. RNN is a neural network containing a loop. The network is capable of storing information. Information is passed from layer to layer. The output of hidden layers depends on the information. RNN has been commonly used for NLP or sequential data problems. However, its architecture is simple since the ability to link long-distance layers is poor. It is incapable of remembering information from long distance data and, thus, the first elements of input sequence usually has no great effect on the following steps. As a result, RNN is influenced by the derivatives during learning and vanishing gradient. LSTM network is designed to solve the problem; it only remembers relevant and other information to discard them.

The LSTM network consists of many interconnected cells as shown in Figure 8 [21]. The idea of LSTM is to add cell internal state ( $s_t$ ) and three ports of input and output ( $f_t$  forget gate,  $i_t$  input, and  $\theta_t$  output). At each time step, the ports receive the input  $x_t$  (representing an element of input sequence) and  $h_{t-1}$  that obtains from output of the memory cells from previous time step ( $t-1$ ). The port has the function of selecting information for each different purpose. They are defined as follows:

- *Forgotten gate* removes unnecessary information from  $s_t$ ;
- *Input port* filters the necessary information to add  $s_t$ .
- *Output port* determines information from  $s_t$  that is used as output.

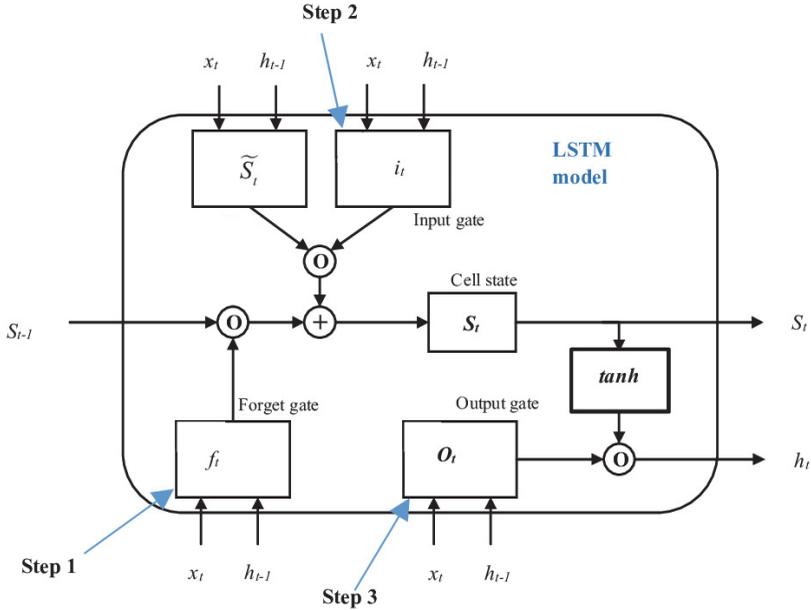


Fig. 8. An example of LSTM model

During the implementation,  $s_t$  and  $h_t$  are calculated as follows:

At the first step, LSTM cell determines the information that needs to be removed from  $s_t$  at  $t-1$ . The value of  $f_t$  is calculated based on  $x_t$ ,  $h_{t-1}$  and  $b_f$ . The sigmoid function converts all activation values between “0” and “1” as follows:

$$f_t = \sigma(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f). \quad (7)$$

At the second step, LSTM cell determines the information that needs to be added to  $s_t$ . The step involves two calculations for  $\tilde{s}_t$  and  $i_t$ .  $\tilde{s}_t$  represents information that can be added to  $s_t$ . as:

$$\tilde{s}_t = \tanh(W_{\tilde{s},x}x_t + W_{\tilde{s},h}h_{t-1} + b_{\tilde{s}}). \quad (8)$$

$i_t$  of input port at time  $t$  is calculated:

$$i_t = \sigma(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i). \quad (9)$$

In the next step,  $s_t$  is calculated as follows:

$$s_t = f_t * s_{t-1} + i_t * \tilde{s}_t. \quad (10)$$

Finally,  $h_t$  is:

$$o_t = \sigma(W_{o,h}h_{t-1} + b_o); \quad (11)$$

$$h_t = o_t * \tanh(s_t), \quad (12)$$

where  $W_{\tilde{s},x}, W_{\tilde{s},h}, W_{f,x}, W_{f,h}, W_{i,x}$  and  $W_{i,h}$  are weight matrices in each LSTM cell and  $b_f, b_{\tilde{s}}, b_i$ , and  $b_o$  are bias vectors.

**3.6. Convolution Neural Network (CNN).** CNN has a number of advantages for image and text processing. Therefore, it is studied in regard to NLP application. However, the biggest disadvantage of RNN is that it takes a too much time to train a model. Therefore, the researchers hope to use CNN to reduce training time while still achieving the same results as those of RNN.

CNN is a type of artificial neural network (ANN) with multiplayer perceptron's, namely convolution and pooling. It is understood as a convolutional class that transforms an input into a different output. CNN simply includes a few layers of convolution combining nonlinear activation functions as *ReLU* or *tanh* to create abstract of information for the next layer.

In the feed-forward neural network (FNN) model, the layers are directly connected to each other through weight ( $^w$ ). These layers are called the fully connected or affine layers. In CNN model layers are linked through convolution. Specifically, the next layer is a convolution result from the previous layer since local connections exist.

CNN is a collection of convolution classes that overlaps and used for nonlinear activation functions similar to ReLU and then for activating weights in nodes. Each class after the activation function will generate abstract of information for the next one. In the FNN model each input is used for subsequent layers.

In training process CNN automatically learns values through filter classes. There are two aspects to consider; they are location invariance and compositionality. If this object is projected under different angles, the accuracy of the algorithm will be significantly affected with the same object. Pooling layer is the immutability for translation, rotation, and scaling. Local aggregation gives levels of information representation from lower to higher and abstract through the convolution. The composite layer is inserted periodically after each pair of convolutions and the nonlinear activation layer. It reduces the spatial dimensions of a sample and the total parameters of network. Therefore, the

aggregate layer makes sense to control the overload in network. There are two types of algorithms common in aggregation classes as follows:

- *MaxPooling* extracts the maximum value of child function;
- *AveragePooling* extracts the average value of sub-functions.

RNN is a specialized neural network that processes a series of values. At the network application, the text to be a two-dimension matrix is considered where rows are the tokens of sentence, and their values are the vector values of each token. The size of input matrix is as follows:  $height = len(tokens)$  and  $width = dimension(wordembedding)$ . At convolution implementation, the filter (also two-dimension matrix) is used with  $widthvalue = Input\ Matrix$ , and height value is usually 3, 4 or 5. The application of CNN to NLP can be described as shown in Figure 9.

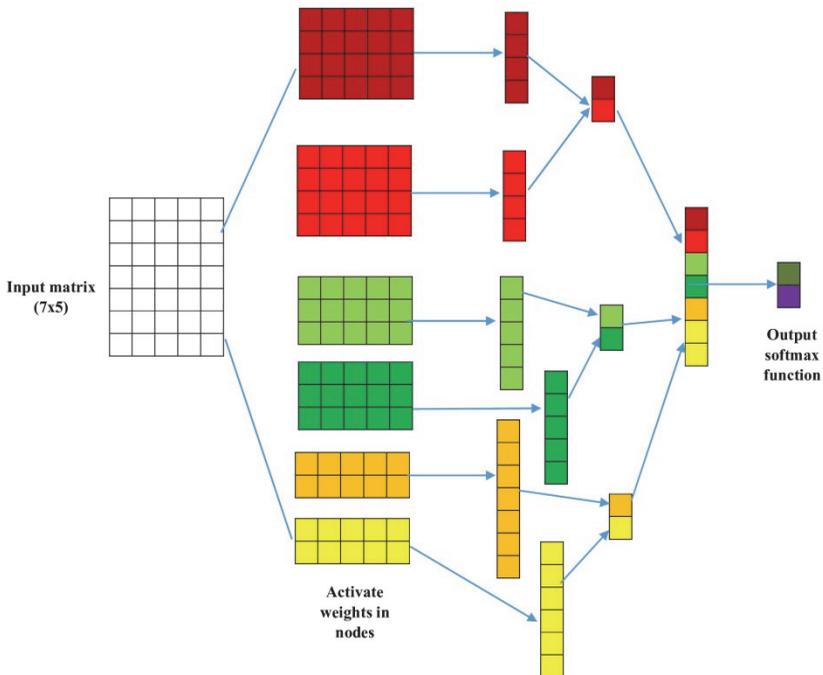


Fig. 9. Diagram of applying CNN to NLP model [42]

## 4. Simulation and Results

### 4.1. Setup

**4.1.1. Dataset.** VNTC dataset is used [4] including 33756 and 50373 articles for training and testing sets, respectively. The input will be separated from the words via *pyvi* library [43] and cleaned stop-words (special characters and punctuation).

**4.1.2. Pre-processing Data.** All documents after separating from via *pyvi* library are processed into words for the next processing step. The function of auxiliary words will be omitted to increase the performance as well as to reduce the number of characteristics for classification model.

**4.1.3. Extracting Feature.** The text is converted into-vector to create a directory for using document representations. Each word will be assigned a natural value. They are placed with the corresponding numbers. Each text is converted into an array of natural numbers that are assigned fixed words. It will be converted into a vector form consisting of one number and “0” elements.

Word2vec will be used to improve text features. It was found out that the pre-train of the *gensim* library is not suitable for the considered dataset. Therefore, the dataset was built by the authors. The Word2vec model was subjected to retraining in order to improve the problem with  $embedding_{size} = 300$ ,  $window_{size} = 5$  for 80000 documents. Figure 10 is the result of training model using  $t - SNE$ .

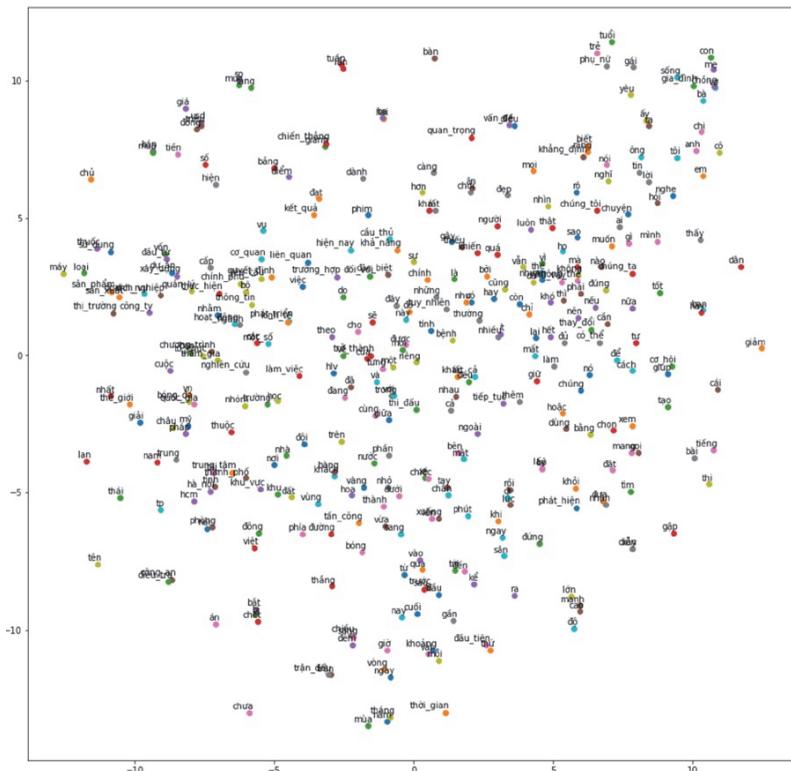


Fig. 10. Selection words of vector space from Word2vec

**4.1.4. Text Classification.** Since a newsletter usually ranges from 500 to 1000 words, the following was chosen: the maximum length of each document is 400 and the size of word embedding is 300 in order to reduce the input of network. The choice of text size and vector space to solve the multi-dimensional data problem of traditional methods is very important. Therefore, was selected the training of the model with a hidden class of 128 units and use of *RMSProp* with a learning rate of 0.001 for the optimal function and a dropout of 0.4.

The embedding layer with a 300-dimensional vector is used to represent a word. Then, was used spatial dropout 1D to remove 1D feature of maps and increase independence between them. The LSTM class is used with 128 memory units. The model of output layer consists of 10 topics. Activation function is used as *softmax* to classify multiple layers and loss function is categorized as *cross\_entropy*. In addition, *batchsize* is set to 64 and *epochs* is 35. Detailed flowchart of LSTM algorithm is shown in Figure 11.

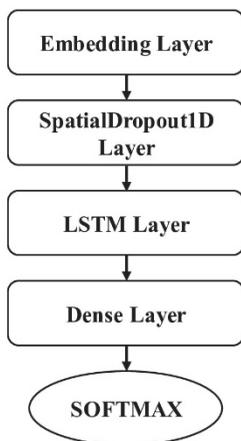


Fig. 11. Algorithm flowchart for LSTM model

**4.1.5. Evaluation Criteria.** The precision, recall, and F1-score criteria were used to evaluate the model performance. The parameters are shown in Table 2 [44] where:

- TN is the result where model accurately predicts the negative class;
- TP is the result where model accurately predicts the positive class;
- FN is the result where model incorrectly predicts the negative class;
- FP is the result where model incorrectly predicts the positive class.

Table 2. Confution matrix

Labels	Negative prediction	Positive prediction
Negative act	True negative(TN)	False positive(FP)
Positive act	False negative(FN)	True positive(TP)

*Precision* is the ratio of positive points determining by *TP* and *FP* as:

$$Precision = \frac{TP}{TP + FP}. \quad (13)$$

*Recall* is the ratio of positive points determining by *TP* and *FN* as:

$$Recall = \frac{TP}{TP + FN}. \quad (14)$$

*F1-score* is the harmonic mean of precision and recall (assuming these two quantities are non-zero) determining by the expression:

$$\frac{2}{F1} = \frac{1}{Precision} + \frac{1}{Recall}. \quad (15)$$

**4.2. Results.** The results of the model are shown in Tables 3 and 4.

Table 3. Comparing the results of LSTM model and the proposal

Model	Precision (%)	Recall (%)	F1-score (%)
LSTM	92.36	91.81	92.09
The proposal (LSTM + Word2Vec)	95.55	95.93	95.74

Table 4. Comparing the results of CNN model and the proposal

Model	Precision (%)	Recall (%)	F1-score (%)
CNN	84.48	83.02	83.25
The proposal (CNN+ Word2Vec )	84.13	84.89	84.01

Table 3 shows the results of classification for the text in Vietnamese. The accuracy of model using LSTM with Word2vec is 4% higher than that

using LSTM. The result is achieved by pre-training of the Word2vec model. Besides, Word2vec is able of preventing the overfitting that reduces the number of training parameters and improves accuracy.

Also was executed the model based on CNN and CNN + Word2vec as shown in Table 4. Two important features of CNN are local sensing and weight sharing. When CNN algorithm is applied to the problem, the input is considered as a two-dimensional matrix of size  $400 \times 300$ . In the case under consideration the fixed length of each text is 400 and size of each word is 300. In the CNN model is used the Conv1D class with a filter of 128 and kernel-size of 5. For the Pooling class, we use *MaxPool1D* with *pool - size* = 2. In the next class, we continue to use *Conv1D* with a filter of 256. Finally, we use *GlobalAvgPool1D* with the *softmax* activation function.

The results given in Tables 3 and 4 show that the accuracy of model based on LSTM + Word2vec is much better than that of CNN. CNN demonstrates great advantages in image processing. However, the model has low results in text classification. It can be seen that LSTM gives good results in text data processing.

Besides, the received results are also compared with other methods results. The results are shown in Table 5, Figures 12 and 13.

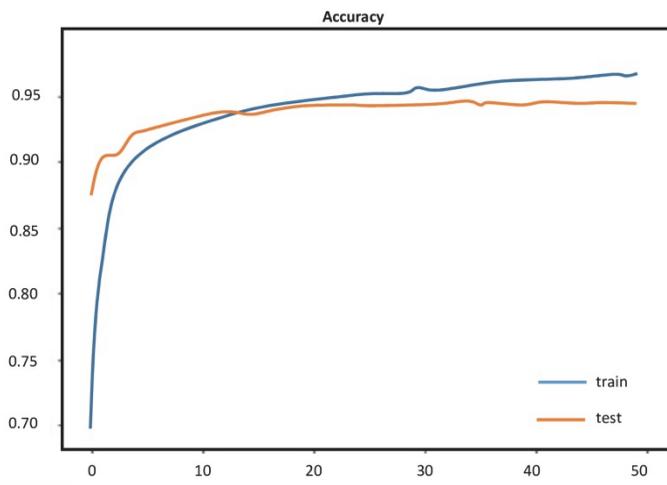


Fig. 12. Loss value of proposal's model

It could be seen in Table 5, that while using the same number of extracting features the proposed approach gains a better accuracy than the other methods.

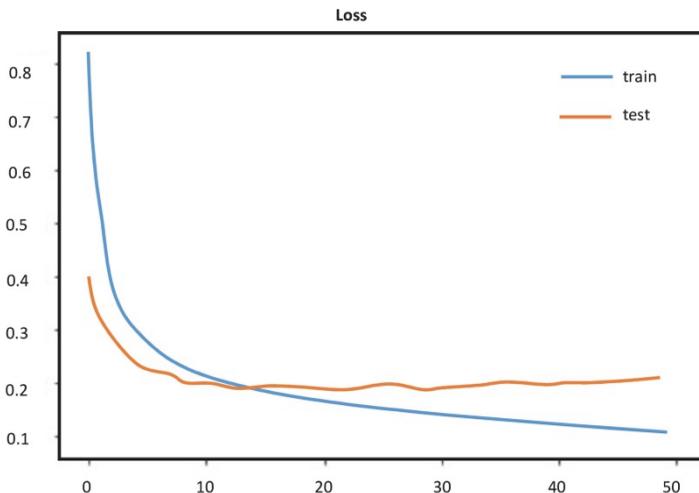


Fig. 13. Accuracy value of proposed model

Table 5. Comparing the results of proposed method with some other methods

Method	Technology	Data	Number of extracting features	Accuracy (%)
Proposed	LSTM, Wor2vec, and CNN	Vietnam data (VNTC) [4]	300	93.8
[18]	Topic modeling (Naïve Bayes theory)	VLSP (20000 sentences)	829 and 339	from 83.00 to 94.07
[19]	Support vector machine (SVM)	4162 documents [29]	7721	80.72
[45]	Naïve Bayes, Maxent, LSTM, and Bi-LSTM	Vietnamese students of feed- back corpus for sentiment analysis (16000 feed-back) [47]	300	from 81.2 to 89.6
[4]	SVM,kNN, and NGram	VNTC [4]	N/A	93.4; 84.67; and 97.1
[47]	SVM, Random forest, SVC, and neural network	VNTC [4]	N/A	96.52; 99.21; 99.22; and 99.75

**5. Conclusion.** The documents classification plays an important role in exploiting big data. The paper presented the method of text classification based on LSTM and CNN models. LSTM algorithm overcomes the vanishing gradient problem that helps in identifying the associations between sentence characteristics, key words, and contextual words. The proposed approach contribution is the use of the Word2vec model for classifying text based on title. Besides, the features learning was executed by Word2vec method that helps to link them together to improve the model efficiency. In the Word2vec method, was used a 300-dimension to reduce the number of characteristics in comparison with other methods. It helps to increase processing speed and avoid a curse of dimensionality.

Also, was performed the model without using Word2vec. The simulation (model) shows that the results are better with Word2vec use. The selection of characteristics from Word2vec helps the model to select appropriate features and to improve accuracy. Also were compared the models using LSTM and CNN. The results show that LSTM is better than CNN.

However, some disadvantages were identified as follows:

- training time of model is longer than that of other methods (SVM and CNN);

- the method will not work correctly if the sentences are too long.

In the future, therefore, the plan is to improve the classification model by combining other DL models to change derivative disappearance.

## References

1. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural computation*. 1997. vol. 9. pp. 1735–1780.
2. Sak H., Senior A., Beaufays F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv preprint arXiv:1402.1128.2014.
3. Phuong L.-H., Nguyen H., Roussanaly A., Ho T. A hybrid approach to word segmentation of Vietnamese texts. *Lecture Notes in Computer Science*. 2013. vol. 5196. pp. 240–249.
4. Hoang V.C.D., Dinh D., Nguyen N. le, Ngo H.Q. A comparative study on Vietnamese text classification methods. 2007 IEEE International Conference on Research, Innovation and Vision for the Future. 2007. pp. 267–273.
5. Ngo Q.H., Dien D., Winiwarer W. A hybrid method for word segmentation with English-Vietnamese bilingual text. 2013 International Conference on Control, Automation and Information Sciences (ICCAIS). 2013. pp. 48–52.
6. Jindal P., Jindal B. Line and word segmentation of handwritten text documents written in Gurmukhi script using mid point detection technique. 2015 2nd International Conference on Recent Advances in Engineering Computational Sciences (RAECS). 2015. pp. 1–6.
7. Gao Y. et al. Wacnet: Word segmentation guided characters aggregation net for scene text spotting with arbitrary shapes. 2019 IEEE International Conference on Image Processing (ICIP). 2019. pp. 3382–3386.
8. Charoenpornswat P., Schultz T. Improving word segmentation for Thai speech translation. 2008 IEEE Spoken Language Technology Workshop. 2008. pp. 241–244.
9. Yu C. et al. Term extraction from Chinese texts without word segmentation. 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT). 2017. pp. 1–4.

10. Nguyen T., Le A. A hybrid approach to Vietnamese word segmentation. 2016 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF). 2016. pp. 114–119.
11. Zhang Z. et al. Effective subword segmentation for text comprehension. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2019. vol. 27. no. 11. pp. 1664–1674.
12. Bal A., Saha R. An improved method for handwritten document analysis using segmentation, baseline recognition and writing pressure detection. *Procedia Computer Science*. 2016. vol. 93. pp. 403–415.
13. Nguyen T.V., Tran H.K., Nguyen T.T.T., Nguyen H. Word segmentation for Vietnamese text categorization: An online corpus approach. RIVF06. 2005. vol. 172. pp. 1–6.
14. Nguyen T., Lung V.D. Extracting the main content of Vietnamese scientific documents based on the structure. *Vietnam Journal of Science and Technology (VJST)*. 2014. vol. 52. no. 3. pp. 269–280.
15. Xiao L., Wang G., Zuo Y. Research on patent text classification based on word2vec and LSTM. 2018 11th International Symposium on Computational Intelligence and Design (ISCID). 2018. vol. 01. pp. 71–74.
16. Hassan A., Mahmood A. Efficient deep learning model for text classification based on recurrent and convolutional layers. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). 2017. pp. 1108–1113.
17. Sarkar A., Chatterjee S., Das W., Datta D. Text classification using support vector machine. *International Journal of Engineering Science Invention*. 2015. vol. 4. no. 11. pp. 33–37.
18. Linh B.K. et al. Vietnamese text classification based on topic modeling. 9th Fundamental and Applied IT Research (FAIR). 2016. vol. 01. pp. 532–537.
19. De T.C., Khang P.N. Classify text with supported vector learning machine and decision tree. *Can Tho University Journal of Science*. 2012. vol. 21. pp. 269–280.
20. Radhika K., Bindu K.R. A text classification model using convolution neural network and recurrent neural network. *International Journal of Pure and Applied Mathematics*. 2018. vol. 119. pp. 1549–1554.
21. Fischer T., Krauss C. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*. 2018. vol. 270. no. 2. pp. 654–669.
22. Sebastiani F. Machine learning in automated text categorization. *ACM Computing Surveys*. 2001. vol. 34. pp. 1–47.
23. Yasotha R., Charles E.Y.A. Automated text document categorization. 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS). 2015. pp. 522–528.
24. Farhoodi M., Yari A. Applying machine learning algorithms for automatic Persian text classification. 2010 6th International Conference on Advanced Information Management and Service (IMS). 2010. pp. 318–323.
25. Krendzelak M., Jakab F. Text categorization with machine learning and hierarchical structures. 2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA). 2015. pp. 1–5.
26. Giang N.L., Hien N.M. Classification of Vietnamese documents using support vector machine. *NU Journal of Science: Computer Science and Communication Engineering*. 2005. pp. 1–10.
27. Nguyen P., Hong T., Nguyen K., Nguyen N. Deep learning versus traditional classifiers on Vietnamese students' feedback corpus. 2018 5th NAFOSTED Conference on Information and Computer Science (NICS). 2018. pp. 75–80.
28. Vo Q., Nguyen H., Le B., Nguyen M. Multi-channel LSTM-CNN model for Vietnamese sentiment analysis. 9th International Conference on Knowledge and Systems Engineering (KSE). 2017. pp. 24–29.
29. Vnexpress, The most read Vietnamese newspaper. 2020. Available at: <https://e.vnexpress.net/> (accessed: 05.12.2019).
30. Tuoitre, Tuoitre news. 2020 Available at: <https://tuoitre.vn/> (accessed: 05.12.2019).

31. Thanhnien, Thanhnien online newspaper. 2020. Available at: <https://thanhnien.vn/a> (accessed: 05.12.2019).
32. NLD, Nguoilaocong online newspaper. 2020. Available at: <https://nld.com.vn/> (accessed: 05.12.2019).
33. Trung T.V. Python Vietnamese Core NLP Toolkit. 2019. Available at: <https://github.com/trungtv/pyvi> (accessed: 05.12.2019).
34. Nguyen D.Q. et al. A fast and accurate Vietnamese word segmenter. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). 2018. pp. 2582–2587.
35. Nguyen D.Q., Verspoor K. An improved neural network model for joint pos tagging and dependency parsing. Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. 2018. pp. 1–11.
36. Nguyen C.-T. et al. Vietnamese word segmentation with CRFs and SVMs: An investigation. Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation 2006. pp. 215–222.
37. Le V.-D. Detailed explanation of Word2Vector Skip-gram. 2015. Available at: <http://www.programmingsought.com/article/8383114826/> (accessed: 05.12.2019).
38. Ma L., Zhang Y. Using word2vec to process big text data. 2015 IEEE International Conference on Big Data (Big Data). 2015. pp. 2895–2897.
39. Barazza L. How does Word2Vec’s Skip-Gram work? 2017. Available at: <https://becominghuman.ai> (accessed: 19.02.2017).
40. Landthaler J. et al. Extending thesauri using word embeddings and the intersection method. ASAIL@ ICAIL. 2017. vol. 8. no. 1. pp. 112–119.
41. An S. Recurrent Neural Networks. 2017. Available at: <https://www.cc.gatech.edu/san37/post/dlhc-rnn/> (accessed: 10.10.2019).
42. Zhang Y., Wallace B. A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification // arXiv preprint arXiv:1510.03820. 2015.
43. Le V.-D. Vietnamese stopwords, 2015. Available at: <https://github.com/stopwords/vietnamese-stopwords> (accessed: 05.12.2019).
44. Ting K.M. Confusion Matrix. Boston. MA: Springer US. 2010. pp. 209–209.
45. Nguyen P., Hong T., Nguyen K., Nguyen N. Deep learning versus traditional classifiers on Vietnamese students’ feedback corpus. 2018 5th NAFOSTED Conference on Information and Computer Science (NICS). 2018. pp. 75–80.
46. Nguyen K.V. et al. UIT-VSFC: Vietnamese students’ feedback corpus for sentiment analysis. 2018 10th International Conference on Knowledge and Systems Engineering (KSE). 2018. pp. 19–24.
47. Van T.P., Thanh T.M. Vietnamese news classification based on bow with key-words extraction and neural network. 2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES). 2017. pp. 43–48.

**Phat Huu Nguyen** – Ph.D., Dr.Sci., Lecturer, School of Electronics and Telecommunications, Hanoi University of Science and Technology (HUST). Research interests: digital image and video processing, wireless networks, ad hoc and sensor network, and intelligent traffic system (ITS) and internet of things (IoT). The number of publications – 55. [phat.nguyenhuu@hust.edu.vn](mailto:phat.nguyenhuu@hust.edu.vn); 1, Dai Co Viet str., Hanoi, Viet Nam; office phone: +84(243)869-2242; fax: +84(243)869-2242.

**Anh Nguyen Thi Minh** – Bachelor, School of Electronics and Telecommunications, Hanoi University of Science and Technology (HUST). Research interests: natural language processing, artificial intelligence applications. The number of publications – 1. [anh.ntm165774@sis.hust.edu.vn](mailto:anh.ntm165774@sis.hust.edu.vn); 1, Dai Co Viet str., Hanoi, Viet Nam; office phone: +84(243)869-2242; fax: +84(243)869-2242.

**Acknowledgements.** This research is carried out in the framework of the project funded by the Ministry of Education and Training (MOET), Vietnam under the grant B2020-BKA-06. The authors would like to thank the MOET for their financial support.

Х.Н. ФАТ, Н.Т.М. АНЬ  
**АЛГОРИТМ КЛАССИФИКАЦИИ ВЬЕТНАМСКОГО ТЕКСТА С  
ИСПОЛЬЗОВАНИЕМ ДОЛГОЙ КРАТКОСРОЧНОЙ ПАМЯТИ И  
WORD2VEC**

*Фат Х.Н., Ань Н.Т.М.* Алгоритм классификации вьетнамского текста с использованием долгой краткосрочной памяти и Word2Vec.

**Аннотация.** В условиях текущей четвертой промышленной революции вместе с развитием компьютерных технологий увеличивается и количество текстовых данных. Следует понимать природу и характеристики этих данных, чтобы применять необходимые методологии. Автоматическая обработка текста экономит время и ресурсы существующих систем. Классификация текста является одним из основных приложений обработки естественного языка с использованием таких методов, как анализ тональности текста, разметка данных и так далее. В частности, недавние достижения в области глубокого обучения показывают, что эти методы хорошо подходят для классификации документов. Они продемонстрировали свою эффективность в классификации англоязычных текстов. Однако по проблеме классификации вьетнамских текстов существует не так много исследований. Последние созданные модели глубокого обучения для классификации вьетнамского текста показали заметные улучшения, но тем не менее этого недостаточно. Предлагается автоматическая система на основе длинной краткосрочной памяти и Word2Vec моделей, которая повышает точность классификации текстов. Предлагаемая модель продемонстрировала более высокие результаты классификации вьетнамских текстов по сравнению с другими традиционными методами. При оценке данных вьетнамского текста предлагаемая модель показывает точность классификации более 90%, поэтому может быть использована в реальном приложении.

**Ключевые слова:** классификация текста, естественная языковая обработка, обработка данных, длинная краткосрочная память, Word2Vec

**Фат Хуу Нгуен** – д-р техн. наук, преподаватель, факультет электроники и телекоммуникаций, Ханойский научно-технический университет. Область научных интересов: цифровая обработка изображений и видео, беспроводные сети, одноранговые и сенсорные сети, интеллектуальная транспортная сеть (ITS) и Интернет вещей (IoT). Число научных публикаций – 55. phat.nguyenhuu@hust.edu.vn; Дай Ко Вьет, 1, Ханой, Вьетнам; р.т.: +84(243)869-2242; факс: +84(243)869-2242.

**Ань Нгуен Тхи Минь** – бакалавр, факультет электроники и телекоммуникаций, Ханойский научно-технический университет. Область научных интересов: обработка естественного языка, приложения искусственного интеллекта. Число научных публикаций – 1. anh.ntm165774@sis.hust.edu.vn; ул. Дай Ко Вьет, 1, Ханой, Вьетнам; р.т.: +84(243)869-2242; факс: +84(243)869-2242.

**Поддержка исследований.** Данное исследование проводится в рамках проекта, финансируемого Министерством образования и науки Вьетнама в рамках (грант B2020-VKA-06).

### Литература

1. Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. 1997. vol. 9. pp. 1735–1780.

2. *Sak H., Senior A., Beaufays F.* Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition // arXiv preprint arXiv:1402.1128.2014.
3. *Phuong L.-H., Nguyen H., Roussanaly A., Ho T.* A hybrid approach to word segmentation of vietnamese texts // Lecture Notes in Computer Science. 2013. vol. 5196. pp. 240–249.
4. *Hoang V.C.D., Dinh D., Nguyen N. le, Ngo H.Q.* A comparative study on Vietnamese text classification methods // 2007 IEEE International Conference on Research, Innovation and Vision for the Future. 2007. pp. 267–273.
5. *Ngo Q.H., Dien D., Winiwarter W.* A hybrid method for word segmentation with english- vietnamese bilingual text // 2013 International Conference on Control, Automation and Information Sciences (ICCAIS). 2013. pp. 48–52.
6. *Jindal P., Jindal B.* Line and word segmentation of handwritten text documents written in Gurmukhi script using mid point detection technique // 2015 2nd International Conference on Recent Advances in Engineering Computational Sciences (RAECS). 2015. pp. 1–6.
7. *Gao Y. et al.* Wacnet: Word segmentation guided characters aggregation net for scene text spotting with arbitrary shapes // 2019 IEEE International Conference on Image Processing (ICIP). 2019. pp. 3382–3386.
8. *Charoenpornasawat P., Schultz T.* Improving word segmentation for Thai speech translation // 2008 IEEE Spoken Language Technology Workshop. 2008. pp. 241–244.
9. *Yu C. et al.* Term extraction from Chinese texts without word segmentation // 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT). 2017. pp. 1–4.
10. *Nguyen T., Le A.* A hybrid approach to Vietnamese word segmentation // 2016 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF). 2016. pp. 114–119.
11. *Zhang Z. et al.* Effective subword segmentation for text comprehension // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2019. vol. 27. no. 11. pp. 1664–1674.
12. *Bal A., Saha R.* An improved method for handwritten document analysis using segmentation, baseline recognition and writing pressure detection // Procedia Computer Science. 2016. vol. 93. pp. 403–415.
13. *Nguyen T.V., Tran H.K., Nguyen T.T.T., Nguyen H.* Word segmentation for Vietnamese text categorization: An online corpus approach // RIVF06. 2005. vol. 172. pp. 1–6.
14. *Nguyen T., Lung V.D.* Extracting the main content of Vietnamese scientific documents based on the structure // Vietnam Journal of Science and Technology (VJST). 2014. vol. 52. no. 3. pp. 269–280.
15. *Xiao L., Wang G., Zuo Y.* Research on patent text classification based on Word2Vec and LSTM // 2018 11th International Symposium on Computational Intelligence and Design (ISCID). 2018. vol. 01. pp. 71–74.
16. *Hassan A., Mahmood A.* Efficient deep learning model for text classification based on recurrent and convolutional layers // 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). 2017. pp. 1108–1113.
17. *Sarkar A., Chatterjee S., Das W., Datta D.* Text classification using support vector machine // International Journal of Engineering Science Invention. 2015. vol. 4. no. 11. pp. 33–37.
18. *Linh B.K. et al.* Vietnamese text classification based on topic modeling // 9th Fundamental and Applied IT Research (FAIR). 2016. vol. 01. pp. 532–537.
19. *De T.C., Khang P.N.* Classify text with supported vector learning machine and decision tree // Can Tho University Journal of Science. 2012. vol. 21. no. a. pp. 269–280.

20. *Radhika K., Bindu K.R.* A text classification model using convolution neural network and recurrent neural network // International Journal of Pure and Applied Mathematics. 2018. vol. 119. pp. 1549–1554.
21. *Fischer T., Krauss C.* Deep learning with long short-term memory networks for financial market predictions // European Journal of Operational Research. 2018. vol. 270. no. 2. pp. 654–669.
22. *Sebastiani F.* Machine learning in automated text categorization // ACM Computing Surveys. 2001. vol. 34. pp. 1–47.
23. *Yasotha R., Charles E.Y.A.* Automated text document categorization // 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICI-CIS). 2015. pp. 522–528.
24. *Farhoodi M., Yari A.* Applying machine learning algorithms for automatic Persian text classification // 2010 6th International Conference on Advanced Information Management and Service (IMS). 2010. pp. 318–323.
25. *Krendzelak M., Jakab F.* Text categorization with machine learning and hierarchical structures // 2015 13th International Conference on Emerging eLearning Technologies and Applications (ICETA). 2015. pp. 1–5.
26. *Giang N.L., Hien N.M.* Classification of Vietnamese documents using support vector machine // VNU Journal of Science: Computer Science and Communication Engineering. 2005. pp. 1–10.
27. *Nguyen P., Hong T., Nguyen K., Nguyen N.* Deep learning versus traditional classifiers on Vietnamese students' feedback corpus // 2018 5th NAFOSTED Conference on Information and Computer Science (NICS). 2018. pp. 75–80.
28. *Vo Q., Nguyen H., Le B., Nguyen M.* Multi-channel LSTM-CNN model for Vietnamese sentiment analysis // 9th International Conference on Knowledge and Systems Engineering (KSE). 2017. pp. 24–29.
29. *Vnexpress.* The most read Vietnamese newspaper. 2020. URL: <https://e.vnexpress.net/> (дата обращения: 05.12.2019).
30. *Tuoi tre, Tuoi tre news.* 2020. URL: <https://tuoi tre.vn/> (дата обращения: 05.12.2019).
31. *Thanh nien, Thanh nien online newspaper.* 2020. URL: <https://thanh nien.vn/a> (дата обращения: 05.12.2019).
32. *NLD, Nguoi lao dong online newspaper.* 2020. URL: <https://nld.com.vn/> (дата обращения: 05.12.2019).
33. *Trung T.V.* Python Vietnamese Core NLP Toolkit. 2019. URL: <https://github.com/trungtv/pyvi> (дата обращения: 05.12.2019).
34. *Nguyen D.Q. et al.* A fast and accurate Vietnamese word segmenter // Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). 2018. pp. 2582–2587.
35. *Nguyen D.Q., Verspoor K.* An improved neural network model for joint post tagging and dependency parsing // Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. 2018. pp. 1–11.
36. *Nguyen C.-T. et al.* Vietnamese word segmentation with CRFs and SVMs: An investigation // Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation 2006. pp. 215–222.
37. *Le V.-D.* Detailed explanation of Word2Vector Skip-gram. 2015. URL: <http://www.programmingsought.com/article/8383114826/> (дата обращения: 05.12.2019).
38. *Ma L., Zhang Y.* Using word2vec to process big text data // 2015 IEEE International Conference on Big Data (Big Data). 2015. pp. 2895–2897.
39. *Barazza L.* How does Word2Vec's Skip-Gram work? 2017. URL: <https://becominghuman.ai> (дата обращения: 19.02.2017).
40. *Landthaler J. et al.* Extending thesauri using word embedding's and the inter-section method // ASAIL@ ICAIL. 2017. vol. 8. no. 1. pp. 112–119.

41. *An S.* Recurrent Neural Networks. 2017. URL: <https://www.cc.gatech.edu/san37/post/dlhc-rnn/> (дата обращения: 10.10.2019).
42. *Zhang Y., Wallace B.* A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification // arXiv preprint arXiv:1510.03820. 2015.
43. *Le V.-D.* Vietnamese stopwords, 2015. URL: <https://github.com/stopwords/vietnamese-stopwords> (дата обращения: 05.12.2019).
44. *Ting K.M.* Confusion Matrix. Boston // MA: Springer US. 2010. pp. 209–209.
45. *Nguyen P., Hong T., Nguyen K., Nguyen N.* Deep learning versus traditional classifiers on Vietnamese students' feedback corpus // 2018 5th NAFOSTED Conference on Information and Computer Science (NICS). 2018. pp. 75–80.
46. *Nguyen K.V. et al.* UIT-VSFC: Vietnamese students' feedback corpus for sentiment analysis // 2018 10th International Conference on Knowledge and Systems Engineering (KSE). 2018. pp. 19–24.
47. *Van T.P., Thanh T.M.* Vietnamese news classification based on bow with key-words extraction and neural network // 2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES). 2017. pp. 43–48.

С.В. Дворников, С.С. Дворников  
**ЭМПИРИЧЕСКИЙ ПОДХОД К ОЦЕНКЕ  
ПОМЕХОУСТОЙЧИВОСТИ СИГНАЛОВ ФАЗОВОЙ  
МОДУЛЯЦИИ**

*Дворников С.В., Дворников С.С. Эмпирический подход к оценке помехоустойчивости сигналов фазовой модуляции.*

**Аннотация.** Высокая спектральная эффективность сигналов с непрерывной фазовой модуляцией определила их известность и активное применение в различных радиотехнических проектах. Уникальность свойств таких сигналов связана с сохранением непрерывности их фазы при смене информационных посылок на длительности символа. Вместе с тем до недавнего времени из всего широкого класса сигналов с непрерывной фазовой модуляцией наибольшее распространение получили различные вариации так называемых сигналов частотной модуляцией с минимальным сдвигом. Однако это далеко не единственные представители класса сигналов с непрерывной фазовой модуляцией, обладающие свойством высокой спектральной компактности.

Исследованы не менее интересные сигналы этого класса, формируемые посредством двойной фазовой модуляции. Представлены аналитические выражения их синтеза, обоснована их принадлежность к классу сигналов с непрерывной фазовой модуляцией. Также исследованы временные свойства фазовой функции, рекомендованной МСЭ-R SM.328-11 для синтеза сигналов с непрерывной фазовой модуляцией, приведены временные и частотные фрагменты сигналов с минимальным сдвигом в сравнении сигналами с двоичной фазовой манипуляцией. Представлены этапы аналитического вывода модели помехоустойчивости сигналов с непрерывной фазовой модуляцией по показателю вероятности битовой ошибки на основе эмпирического подхода. Показана общность полученной модели с известным выражением для сигналов с минимальным сдвигом путем исследования разностной функции ошибки аппроксимации (ошибка порядка  $10^{-3}$ ), что позволило получить более компактное представление разработанной модели применительно к сигналам с двойной фазовой модуляцией. Доказано, что такие сигналы обладают более высокими свойствами помехоустойчивости по отношению к сигналам с минимальным сдвигом (порядка 0,5 дБ по уровню ошибки  $10^{-5}$ ). Указанный результат получен на основе исследования функций различия, определяемых разностью между сигнальными символами соответствующих информационным значениям «1» и «0». Определены направления дальнейшего исследования.

**Ключевые слова:** сигналы с непрерывной фазовой модуляцией, помехоустойчивость сигналов, вероятность битовой ошибки, спектральные свойства сигналов

**1. Введение.** Ограниченность частотного ресурса стимулирует поиск спектрально-эффективных сигналов, которые, с одной стороны, обладали бы приемлемой помехоустойчивостью, а с другой – отвечали требованиям компактности их спектрального представления. Поиск компромиссных решений способствовал разработке так называемых сигналов с непрерывной фазовой модуляцией (*Continuous Phase Modulation – CPM*).

Так в [1] в ходе исследования методов модуляции, обеспечивающих эффективное использование занимаемой полосы частот, авторы пришли к необходимости сохранить непрерывность фазы при смене

сигнальных символов. В результате была разработана частотная манипуляция с непрерывной фазой, получившая название *Continuous Phase Frequency Shift Keying (CPFSK)*. Действительно, сигналы двойной *CPFSK (Binary Continuous Phase Frequency Shift Keying – BCPFSK)* с индексом модуляции, равным  $m_c = 0,715$ , показали более высокую спектральную эффективность по отношению к сигналам двоичной фазовой манипуляции (*Binary Phase Shift Keying – BPSK*).

Широкое признание новый вид модуляции получил после публикации [2], где модуляция *CPFSK* на *m-ичный* случай была обобщена. В том числе получены оптимальные параметры модуляционного формата (здесь и далее понятие модуляционного формата рассмотрено с позиций [3]) по показателю минимума вероятности битовой ошибки (*Bit-Error-Rate – BER*) в соответствии с канальным уровнем отношения сигнал/шум (*Signal-to-Noise Ratio – SNR*).

Поскольку не всегда удобно использовать рассчитанные значения оптимальных параметров модуляционных форматов сигналов, то интересными представляются результаты оценки потерь при отходе в практических решениях от оптимальных значений. Именно поэтому работы [4-7] определили практическую направленность сигналов *CPM*.

Другим преимуществом сигналов *CPM* является то, что непрерывность фазы позволяет в радиотехнических системах использовать достаточно надежные и недорогие усилители мощности с нелинейными характеристиками [6].

Активное практическое применение получила Гауссова частотная модуляция с минимальным сдвигом (*Gaussian Minimum Shift Keying – GMSK*) как разновидность *CPFSK* с индексом модуляции  $m_c = 0,5$  [8], в которой последовательность прямоугольных информационных импульсов сглажена посредством фильтра Гаусса [9, 10].

Несмотря на то, что с момента разработки сигналов *CPM* прошло более 40 лет, исследования в этой области по-прежнему активно продолжаются, о чем свидетельствуют многочисленные публикации.

Так в [7] разработана оптимальная модель, позволяющая осуществлять синтез сигналов *CPM* с плавным изменением фазы на длительности символа. В [11] разработан механизм совместной оценки времени символа, сдвига частоты и сдвига фазы для сигналов с непрерывной фазовой модуляцией. Обоснована гипотеза, что дисперсия фазового сдвига при наличии частотной ошибки должна совпадать с условной границей Крамера – Рао для любых сигналов цифровой фазовой модуляции. В [12] предложен модулятор слабого рассеяния для формирования изображений с когерентной модуляцией вместо двоичного случайного фазового модулятора с сильной рассеивающей спо-

способностью, что позволило улучшить конечное разрешение. В [13] методология синтеза сигналов с непрерывной фазовой модуляцией применена к акустическим каналам передачи информации, в частности обоснована оптимальная скорость передачи. В [15] рассмотрены вопросы построения модели радиосигналов с модуляцией с непрерывным изменением фазы зарубежных спутниковых систем связи в пространстве параметров фазовых диаграмм. Исследованы условия, при которых достигается требуемая точность оценок параметров фазовых диаграмм, и как итог – обеспечивается достоверность распознавания классов сигналов. В [17] исследованы вопросы влияния каналов с многолучевым распространением радиоволн на битовую вероятность ошибки сигнальных форматов с непрерывной модуляцией. В [14, 16, 18, 19] рассмотрены сигналы двойной модуляции, также относящиеся к классу сигналов *CPM*. Исследованы их свойства, оценена спектральная эффективность.

Вместе с тем непрерывность фазовой функции, обеспечивающая спектральную компактность сигналов *CPM*, имеет и негативные моменты: именно отсутствие разрыва затрудняет принятие решения о границах сигнальных символов, особенно в условиях шумов высокой интенсивности, что в конечном итоге ведет к ухудшению их помехоустойчивости.

Поскольку физические свойства сигналов *CPM* определяются законом изменения фазовой функции, то достаточно сложно получить универсальную аналитическую модель для расчета вероятности битовой ошибки. Поэтому в [8] предложен обобщенный подход, основанный на результатах моделирования, что не всегда удобно в теоритических исследованиях. С учетом указанных обстоятельств предложен эмпирический подход к выводу выражения вероятностной оценки помехоустойчивости сигналов *CPM*, основанный на принципе аналогий.

**2. Общие свойства сигналов *CPM*.** В [9] предложена обобщенная аналитическая модель сигналов *CPM*. Эту модель с учетом формирующей фазовой функции  $\phi(t, \mathbf{a})$  (где  $\mathbf{a}$  – вектор, определяющий информационную  $M$ -арную битовую последовательность независимых символов, имеющих значения  $\pm 1, \pm 3, \dots, (M-1)$ , с равной вероятностью их проявления равной  $1/M$ ) можно представить в следующем виде:

$$s(t, \mathbf{a}) = \sqrt{\frac{2E_b}{T}} \cos(2\pi f_c t + \phi(t, \mathbf{a}) + \phi_0), 0 \leq t \leq NT, \quad (1)$$

где  $E_b$  – энергия бита (сигнального символа при бинарной модуляции);  $T$  – длительность бита (сигнального символа при бинарной модуляции);  $f_c$  –

частота несущего колебания,  $\phi_0$  – начальная фаза (в общем случае при когерентной обработке может не учитываться);  $N$  – количество элементов в битовой последовательности, определяющих размерность вектора  $\mathbf{a}$ .

Заметим, что формирующая фазовая функция  $\phi(t)$  при синтезе *CPM* определяется модулирующей функцией мгновенной частоты первичного сигнала, поэтому представим ее следующим образом:

$$\phi(t) = \sum_{k=1}^N a_k g(t - (k-1)T), (k-1)T \leq t \leq kT, \quad (2)$$

где  $g(*)$  – функция мгновенной частоты.

Заметим, что на длительности символа  $T$  форма функции мгновенной частоты  $g(*)$  изменяется в соответствии с текущим значением информационного символа  $a_k$ , являющегося элементом вектора  $\mathbf{a}$ .

$$\mathbf{a} = (a_1, a_2, \dots, a_N). \quad (3)$$

При этом

$$a_k \in \{\pm 1, \pm 3, \dots, \pm(M-1)\}, 1 \leq k \leq N. \quad (4)$$

Согласно (4) именно значение параметра  $M$  в конечном счете определяет число возможных вариантов форм функций мгновенной частоты на длительности символа. Далее в работе будут рассматриваться только те сигналы, которые удовлетворяют условию  $M = 2$ .

В рамках анализируемой ситуации будем полагать, что  $T$  представляет интервал, в пределах которого формирующая фазовая функция  $\phi(t, \mathbf{a})$  на длительности символа будет определяться только изменением функции мгновенной частоты сигнала:

$$g(T) = 2\pi \int_0^T b(\tau) d\tau,$$

где  $b(\tau)$  для сигналов *CPFSK* определяется согласно [2] как:

$$b(\tau) = \begin{cases} \frac{m_c \tau}{2T} & \text{при } 0 \leq \tau \leq T; \\ 0 & \text{при } \tau \geq T, \end{cases} \quad (5)$$

где  $m_c$  – индекс модуляции.

Вместе с тем в рекомендации МСЭ-R SM.328-11 [9] предложено аналитическое описание функции мгновенной частоты  $g(t, n, L)$ , определяющей формирующую фазовую функцию, как функции от трех параметров в виде многочлена:

$$g(t, n, L) = \frac{1}{4} + n \left( \frac{1}{T} - \frac{L}{2} \right) + \frac{5-8nL}{L^3} \left( \frac{1}{T} - \frac{L}{2} \right)^3 + \frac{16nL-12}{L^5} \left( \frac{1}{T} - \frac{L}{2} \right)^5. \quad (6)$$

Здесь  $0 \leq t \leq T$ .

В выражении (6) значения  $n$  и  $L$  представляют собой параметры формирующего импульса сигнальной конструкции СРМ [9].

Очевидно, что результирующие свойства сигналов СРМ как раз и определяются поведением функции  $g(t, n, L)$  на длительности элемента сигнала (рис. 1) при фиксированных значениях параметров  $n$  и  $L$ .

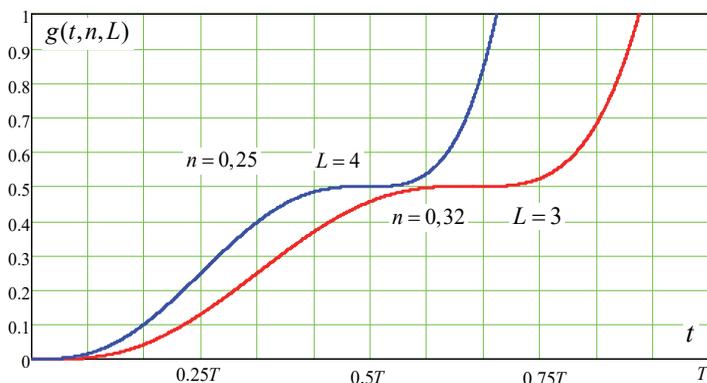


Рис. 1. Фронт возрастания нормализованного импульса, формирующего фазу

На рисунке 1 показаны графики  $g(t, n, L)$  для рекомендованных МСЭ-R SM.328-11 параметров  $n = 0,25$  и  $L = 4$ , а также  $n = 0,32$  и  $L = 3$ .

Выбор указанных параметров обусловлен тем, что они обеспечивают локализацию 95% всей энергии в пределах 62% и 69% занимаемой полосы частот.

Так на рисунке 2 показаны зависимости импульсов  $g(t, n, L)$  единичной длины, то есть при  $t = T$ , от параметра  $n$ , для различных целых значений  $L = 3, 4$  и 5.

Анализ результатов, представленных на рисунках 1 и 2, показывает, что даже значения параметров  $n$  и  $L$ , рекомендованные в МСЭ-R

SM.328-11, носят приближенный характер, что подтверждает сложность применения универсальных решений для практических приложений.

С теоретических позиций рациональным подходом при формировании сигналов *CPM* возможно применение формирующих импульсов прямоугольной формы (*REC*) в виде приподнятого косинуса (*RC*) (полуцикловой синусоиды (*HCS*)) и линейной частотной модуляции (ЛЧМ) [20].

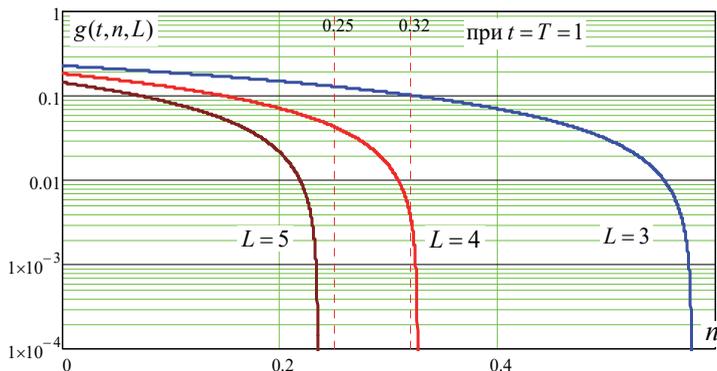


Рис. 2. Зависимость формирующего фазу импульса от параметра  $n$

Отметим, что представленные в настоящей работе результаты моделирования поведения функции  $g(t, n, L)$  при рекомендованных значениях параметров  $n$  и  $L$  являются уникальными.

При всем разнообразии в классе сигналов *CPM* наибольшее распространение получили сигналы частотной модуляции с минимальным сдвигом (*Minimum Shift Keying – MSK*) и с индексом модуляции  $m_c = 0,5$ , у которых изменение фазы носит линейный характер на длительности формирующего импульса.

Учитывая, что для демодуляции сигналов *CPM*, как правило, применяют алгоритм Витерби [17], то для обеспечения наилучшего значения достоверности по критерию максимального правдоподобия целесообразен выбор индекса модуляции  $m_c$ , определяемого соотношением двух простых целых чисел  $k$  и  $n$  в следующей пропорции:

$$m = v / u. \quad (7)$$

Тогда в соответствии с (7) значение индекса модуляции будет определять число фазовых состояний  $\theta_k$  сигнала *CPM*, значения которых в моменты времени  $t = (N - 1)T$  можно представить как:

$$\theta_k = \left\{ 0, \frac{\pi\nu}{u}, \frac{2\pi\nu}{u}, \dots, \frac{(u-1)\pi\nu}{u} \right\}, \quad (8)$$

когда  $\nu$  четное, и:

$$\theta_k = \left\{ 0, \frac{\pi\nu}{u}, \frac{2\pi\nu}{u}, \dots, \frac{(2u-1)\pi\nu}{u} \right\}, \quad (9)$$

когда  $\nu$  нечетное.

Поскольку для сигналов *MSK* индекс модуляции равен  $m_c = 0,5$ , то значение изменения их фазовой функции в соответствии с выражением (5) будет удовлетворять условию:

$$b(\tau) = \begin{cases} 0, & \tau < 0 \\ \frac{\tau}{4T}, & 0 \leq \tau \leq T \\ \frac{1}{4}, & \tau > T \end{cases} \quad (10)$$

Согласно выражению (10) результирующее значение фазы будет возрастать или убывать на величину  $\pi/2$  в зависимости от значения информационного импульса.

На рисунке 3 представлены изменения сигнала *MSK*  $s_{MSK}(t)$  в соответствии с изменениями функции формирующих импульсов  $a(t)$ .

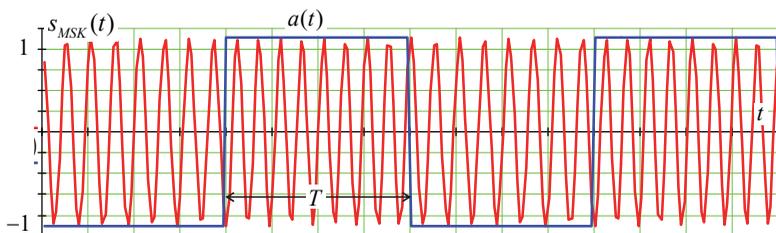


Рис. 3. Фрагмент сигнала *MSK*, изменяющегося в соответствии с формирующими импульсами меандра

На рисунке 3  $a(t)$  представляет собой импульсную последовательность на основе вектора  $\mathbf{a} = (a_1, a_2, \dots, a_N)$  определяемого формулой (3). Здесь  $a(t)$  – меандр.

На рисунке 4 представлены фрагменты изменения функции мгновенной фазы, в соответствии с изменениями функции формирующих импульсов  $a(t)$ .

Понятие функции мгновенной фазы, определяемой (8) и (9), в [8] предложено рассматривать как разность значений фазовой функции на длительности элемента сигнала:

$$\theta(t, a_k) = \phi(t, a_k) - \phi(t, a_{k-1}) = \sum_{k=1}^N a_k g(t - (k-1)T). \quad (11)$$

В выражении (1)  $(k-1)T \leq t \leq kT$ .

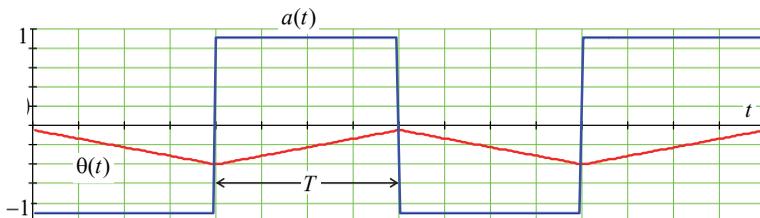


Рис. 4. Фрагмент изменения формирующей фазовой функции в соответствии с импульсной последовательностью в виде меандра для сигнала *MSK*

Оправданность такого подхода определяется особенностями обработки сигналов *CPM* на приеме, когда текущее значение символа можно рассматривать как отношение мгновенной фазы на его временных границах, взвешенное функцией формирующих импульсов:

$$a_k = \frac{\theta_k}{\theta_{k-1}} g(t = T). \quad (12)$$

Отсутствие разрывов мгновенной фазы как определяющего параметра фазовой функции как раз и обеспечивает компактность спектрального представления сигналов *MSK*, в частности по отношению к сигналам двойной фазовой манипуляции (*Binary Phase Shift Key – BPSK*) [21].

В качестве примера на рисунках 5 и 6 показаны фрагменты сигналов *MSK*  $s_{MSK}(t)$  и *BPSK*  $s_{BPSK}(t)$  совместно с их спектрами для информационной импульсной последовательности  $a(t)$  в виде меандра.

Даже визуальный анализ результатов, представленных на рисунках 5 и 6, позволяет заключить, что сигналы *MSK* имеют более компактное спектральное представление по отношению к сигналам

*BPSK*. Однако свойство спектральной компактности без высокой помехоустойчивости еще не гарантирует практическую привлекательность модуляционных форматов.

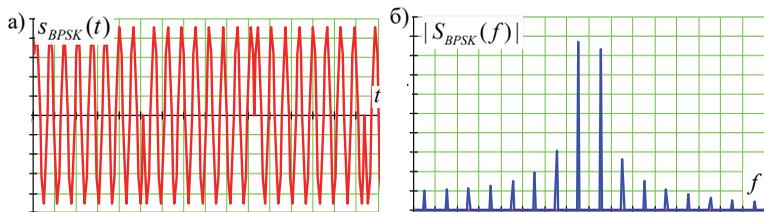


Рис. 5. Временной фрагмент сигнала *BPSK* – а) и его спектр – б)

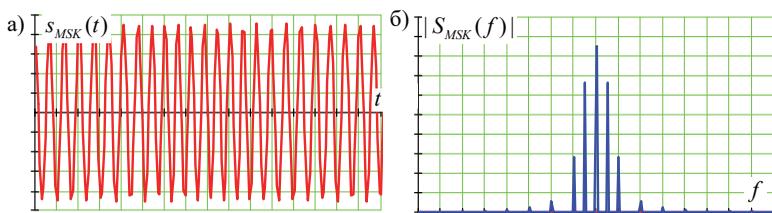


Рис. 6. Временной фрагмент сигнала *MSK* – а) и его спектр – б)

**3. Обоснование эмпирической модели оценки помехоустойчивости сигналов *PCM*.** Классические подходы к оценке помехоустойчивости сигналов базируются на поиске различий принимаемой реализации и эталонного описания [22]. В их основе лежат вероятностные методы, которые предполагают использование отношения правдоподобия  $\Lambda$ , рассчитываемое как отношение функций распределения следующих вероятностных гипотез.

1. Гипотеза  $H_1 | \mathbf{a} = H_1 : z(t) = s(t, a_k = 1 | \mathbf{a}) + x(t)$ , согласно которой в принятой реализации  $z(t)$ , содержащей полезный сигнал  $s(t)$  и мешающие шумовые излучения  $x(t)$  (далее – шум), переданный бит  $a_k$  информационной последовательности  $\mathbf{a}$  соответствует информационному символу «1».

2. Гипотезы  $H_0 | \mathbf{a} = H_0 : z(t) = s(t, a_k = 0 | \mathbf{a}) + x(t)$ , согласно которой в принятой реализации  $z(t)$ , содержащей полезный сигнал  $s(t)$  и шум  $x(t)$ , переданный бит  $a_i$  информационной последовательности  $\mathbf{a}$  соответствует информационному символу «0».

В рассмотренных гипотезах будем полагать, что полезный сигнал  $s(t)$  для гипотезы  $H_1$  соответствует сигналу, переносящему ин-

формационный символ «1», а для гипотезы  $H_0$  соответствует сигналу, переносящему информационный символ «0».

Тогда для гауссова канала (*Additive White Gaussian Noise – AWGN*) функции распределения вероятностей будут иметь следующий вид [23, 24]:

$$\begin{aligned} p(z(t) | H_1, \mathbf{a}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(z(t) - M[z(t)])^2}{2\sigma^2}\right), \\ p(z(t) | H_0, \mathbf{a}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(z(t) - M[z(t)])^2}{2\sigma^2}\right). \end{aligned} \quad (13)$$

В выражениях (13)  $M[z(t)]$  – математическое ожидание входной реализации  $z(t)$  на длительности элемента сигнала;  $\sigma^2 = N_0 / 2$  – дисперсия шума, определяемая значением его спектральной плотностью мощности  $N_0$ .

Заметим, что различия распределений  $p(z(t) | H_1, \mathbf{a})$  и  $p(z(t) | H_0, \mathbf{a})$  определяются в соответствии с различиями сигналов  $s(t, a_k = 1)$  и  $s(t, a_k = 0)$  между собой.

Поскольку  $\Lambda$  рассчитывается как отношение функций распределения вероятностей [27], определяемые формулой (13), можно записать:

$$\Lambda = \frac{p(z(t) | H_1, \mathbf{a})}{p(z(t) | H_0, \mathbf{a})} \underset{0}{\overset{1}{>}} 1. \quad (14)$$

Учитывая, что для каждой выборки, содержащей информационный символ «1», значение  $M[z(t)]$  будет определяться полезным сигналом вида  $s_1(t) = s(t, a_i = 1)$ , а для выборки, в которой содержится информационный символ «0» – полезным сигналом вида  $s_0(t) = s(t, a_i = 0)$ , то отношение правдоподобия (14) можно привести к следующему виду:

$$\Lambda = \frac{\exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^N (z_k^2(t) - 2z_k(t)\bar{s}_1(t) + \bar{s}_1^2(t))\right)}{\exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^N (z_k^2(t) - 2z_k(t)\bar{s}_0(t) + \bar{s}_0^2(t))\right)} \underset{0}{\overset{1}{>}} 1, \quad (15)$$

где  $\bar{s}_1(t)$ ,  $\bar{s}_0(t)$  представляют собой усредненные значения принимаемой выборки на длительности элемента (символа) обрабатываемой реализации для соответствующих гипотез  $H_1$  и  $H_0$  в канале *AWGN* с дисперсией шума  $\sigma^2$ .

Таким образом, в (15)  $z_k(t)$  представляет собой  $k$ -тую входную реализацию на длительности элемента сигнала из выборки, включающей  $N$  реализаций. В то же время отношение правдоподобия (14) для условий когерентной обработки при приеме сигналов *CPM* может быть рассчитано как:

$$\Lambda = M[\Lambda | s(t, a = \{0, 1\}, \mathbf{a})] + \sqrt{\rho(N_0)}. \quad (16)$$

Здесь  $M[\Lambda | s(t, a = \{0, 1\}, \mathbf{a})]$  – усредненное значение отношения правдоподобия по результатам обработки  $N$  реализаций;  $\rho(N_0)$  – дисперсия функции различий, обусловленная спектральной плотностью мощности шума  $N_0$  в канале с *AWGN*.

Для байесовского критерия принятия решения расчет  $\Lambda | s(t, a = \{0, 1\})$  может быть осуществлен на основе усреднении результатов корреляции между функцией различий  $r(t) = \bar{s}_1(t, a = 1) - \bar{s}_0(t, a = 0)$  и усредненными значениями  $\bar{s}_1(t)$  и  $\bar{s}_0(t)$  с каждой  $k$ -ой входной реализацией  $z_k(t)$ .

Тогда в соответствии с [14] имеем:

$$\Lambda_k = \int_0^{kT} z_k(t)r(t) dt \begin{matrix} > \\ < \end{matrix} 0, \text{ где } k = 1, 2, \dots, N. \quad (17)$$

Далее:

$$\begin{aligned} M[\Lambda_k | s(t, a = \{0, 1\}, \mathbf{a})] &= \int_0^{kT} z_k(t)r(t) dt = \\ &= \int_0^{kT} s_1(t, a = 1, \mathbf{a})[\bar{s}_1(t, a = 1) - \bar{s}_0(t, a = 0)] dt. \end{aligned} \quad (18)$$

В выражении (18) в качестве входной реализацией  $z_k(t)$  представлен сигнал, соответствующий информационной «1», определяемой текущим значением элемента  $a_k$  вектора  $\mathbf{a}$ .

$$\bar{s}(t, a = \{1, 0\}) = \sum_{k=1}^n s(t, a = \{1, 0\}, a_k). \quad (19)$$

Формула (18) представляет собой усредненные значения полезного сигнала, полученные на усреднении  $n$  символов, в идеале на  $N$ .

Заметим, что допустимо условие, когда  $n = N$ , однако при достаточно большом размере вектора  $\mathbf{a}$   $n < N$ .

Важным моментом в определении значения  $\Lambda$  является расчет величины  $\rho(N_0)$ , которая является функцией параметра спектральной плотности мощности шума в канале с *AWGN*.

Значение  $\rho(N_0)$  обосновано в [4]:

$$\rho(N_0) = \frac{N_0}{2} \int_0^{nT} [\bar{s}_1(t, a=1) - \bar{s}_0(t, a=0)]^2 dt. \quad (20)$$

Тогда вероятность ошибки, возникающей в результате обработки принимаемой реализации сигнала *CPM*, можно рассчитать в соответствии со следующей формулой:

$$p^{CPM} = Q \left[ \frac{M[\Lambda_k | s(t, a = \{0, 1\}, a_k)]}{\sqrt{\rho(N_0)}} \right], \quad (21)$$

где  $Q(x) = 1 - \Phi(x)$ , а  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{\tau^2}{2}\right) d\tau = \frac{1}{2} \left( 1 + \operatorname{erf} \frac{x}{\sqrt{2}} \right)$  –

функция нормального интегрального распределения;

$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_0^x \exp\left(-\frac{\tau^2}{2}\right) d\tau$  – функция ошибок [23].

Анализ выражения (21) показывает, что значение вероятности битовой (в рассматриваемой ситуации – символьной) ошибки определяется структурным различием сигналов, соответствующих информационной «1»  $\bar{s}_1(t, a=1)$  и информационному «0»  $\bar{s}_0(t, a=0)$  на физическом уровне.

На рисунке 7 показаны функции различий сигналов  $\bar{s}_1(t, a=1)$  и  $\bar{s}_0(t, a=0)$ , накапливаемых на длительности символа для *MSK*  $\Sigma_r^{MSK}(T)$  и *BPSK*  $\Sigma_r^{BPSK}(T)$ , рассчитываемых по формуле:

$$\Sigma_r(T) = \sum_{t=0}^T |\bar{s}_1(t, a=1) - \bar{s}_0(t, a=0)|. \quad (22)$$

При моделировании полагалось  $T = 64$  дискретных отсчета.

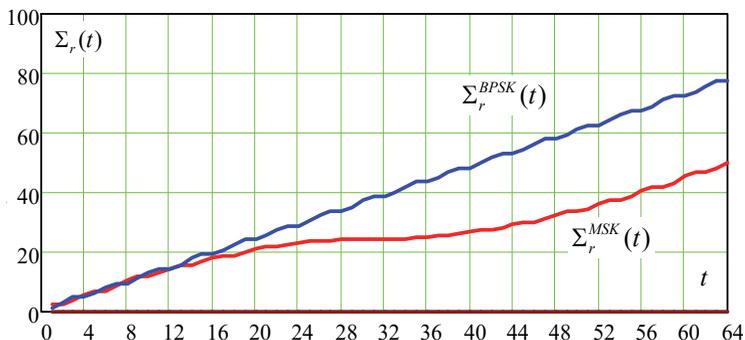


Рис. 7. Функции различных сигналов *MSK* и *BPSK* на длительности элемента сигнала

Поскольку теоретически сложно в условиях шумов получить аналитические выражения для усредненных значений  $\bar{s}_1(t, a=1)$  и  $\bar{s}_0(t, a=0)$ , то результаты, представленные на рисунке 7, получены путем моделирования сигналов *MSK* и *BPSK* в среде *MathCAD* при отношении сигнал/шум 18 дБ.

Для идеальных условий можно воспользоваться аналитическими выражениями, например [28].

Для сигналов *MSK*:

$$s_{MSK}(t, a_k) = \sqrt{\frac{2E_b}{T}} \cos \left[ 2\pi \left( f_0 + \frac{a_k}{4T} \right) t + \phi_k \right], \quad (k-1)T \leq t \leq kT. \quad (23)$$

Здесь  $f_0$  – несущая частота;  $a_k = \pm 1$  – информационные символы, являющиеся элементами вектора  $\mathbf{a}$  и передающиеся со скоростью  $\upsilon = 1/T$ ;  $\theta_k$  представляет собой фазовую функцию на длительности  $k$ -го элемента (символа) сигнала.

В соответствии с (23) получаем при  $a_k = +1$  положительные набег частоты, равный  $\frac{1}{4T}$ , а при  $a_k = -1$  – отрицательный набег

$-\frac{1}{4T}$ . При этом на длительности текущего символа значение полной фазы  $\theta_k$  сохраняется постоянным, равным  $\theta_k = 0$  или  $\theta_k = \pi$ . Указанное постоянство достигается строгим выполнением следующего рекурсивного соотношения:

$$\phi_k = \left[ \phi_{k-1} + \frac{\pi k}{2} (a_{k-1} - a_k) \right] \text{ по модулю } 2\pi. \quad (24)$$

При квадратурной модуляции аналитическое выражение для сигнала *MSK* имеет вид:

$$s_{MSK}(t, a_k) = \cos(\theta_k) \cos\left(\frac{\pi t}{2T}\right) \cos(2\pi f_0 t) - a_k \cos(\theta_k) \sin\left(\frac{\pi t}{2T}\right) \sin(2\pi f_0 t), (k-1)T \leq t \leq kT, \quad (25)$$

где  $\theta_k = 0$  или  $\theta_k = \pi$ .

Для сигналов *BPSK*:

$$s_{BPSK}(t, a_k) = \sqrt{\frac{2E_b}{T}} \cos[2\pi f_0 t + \pi a_k + \phi_0], (k-1)T \leq t \leq kT. \quad (26)$$

Здесь  $\phi_0$  – начальная фаза.

Согласно результатам, представленным на рисунке 7, сигналы *BPSK* обладают лучшей контрастностью элементов, определяющих информационные значения «1» и «0», по сравнению с сигналами *MSK*, что согласуется с теорией потенциальной помехоустойчивости [22]. Правомерность данного подхода позволяет использовать его при анализе новых сигналов, в том числе в классе сигналов *PCM*. Например, в [19] рассмотрены сигналы, формируемые посредством двойной фазовой модуляции (*Dual Phase Modulation – DPM*), которые также являются представителями класса сигналов *PCM*.

Аналитический синтез сигналов *DPM* состоит в том, что первоначально информационную импульсную последовательность  $g(t)$  используют в качестве манипуляционной для модуляции низкочастотного колебания таким образом, чтобы на длительности информационной

посылки укладывался строго один период. В результате получают низкочастотный фазоманипулированный сигнал  $u_{BPSK}(t)$  (рис. 8) [16].

Затем низкочастотным фазоманипулированным сигналом  $u_{BPSK}(t)$  модулируют фазы высокочастотного колебания. Аналитически реализацию рассмотренных процедур можно описать следующим образом:

$$\begin{aligned} s_{DPM}(t) &= \sqrt{\frac{2E_b}{T}} \cos(2\pi f_c t + u_{BPSK}(t) + \phi_0) = \\ &= \sqrt{\frac{2E_b}{T}} \cos(2\pi f_c t + m \cos(2\pi \Omega t) + \phi_0). \end{aligned} \quad (27)$$

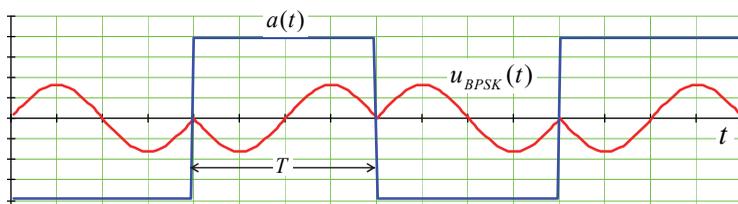


Рис. 8. Фрагменты импульсной последовательности в виде меандра и соответствующей ей формирующей фазовой функции, представленной в виде низкочастотного фазоманипулированного сигнала с для сигнала  $DPM$

В выражении (27)  $\Omega = 1/T$  – частота следования информационной импульсной последовательности. Поскольку  $u_{BPSK}(t)$  является непрерывной функцией, то результирующий сигнал  $DPM$   $s_{DPM}(t)$  удовлетворяет требованиям соответствия сигналов  $PCM$ .

На рисунке 9 представлен сигнал  $DPM$  и его спектр по аналогии с демонстрируемыми фрагментами на рисунках 5 и 6.

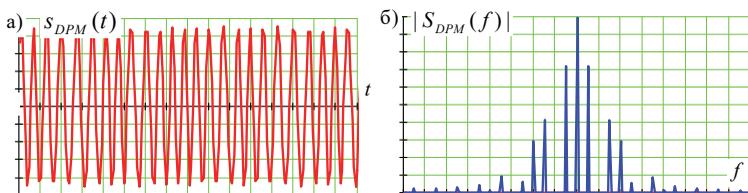


Рис. 9. Временной фрагмент сигнала  $DPM$  – а) и его спектр – б)

Согласно [14, 19] сигналы  $DPM$  обладают достаточно высокой спектральной эффективностью. Однако их помехоустойчивость была оценена только по результатам практического эксперимента.

Вместе с тем, используя функцию различий  $r(t)$ , которая представлена в выражении (18), и модель (21), можно получить аналитическое выражение для оценки помехоустойчивости сигналов *DPM*.

Так, на рисунке 10 показана  $\Sigma_r^{DPM}(T)$  – функция различий сигналов  $\bar{s}_1(t, a=1)$  и  $\bar{s}_0(t, a=0)$ , накапливаемых на длительности символа *DPM* в сравнении с аналогичной функцией  $\Sigma_r^{MSK}(T)$ .

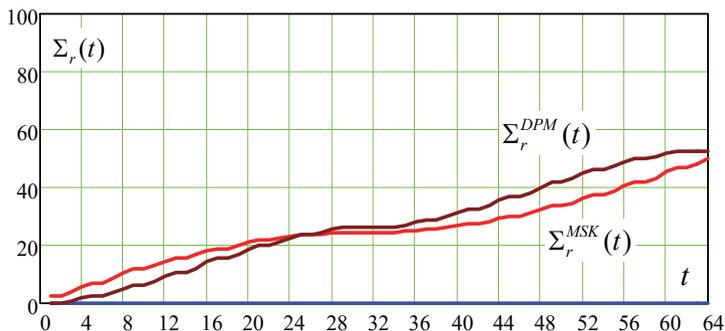


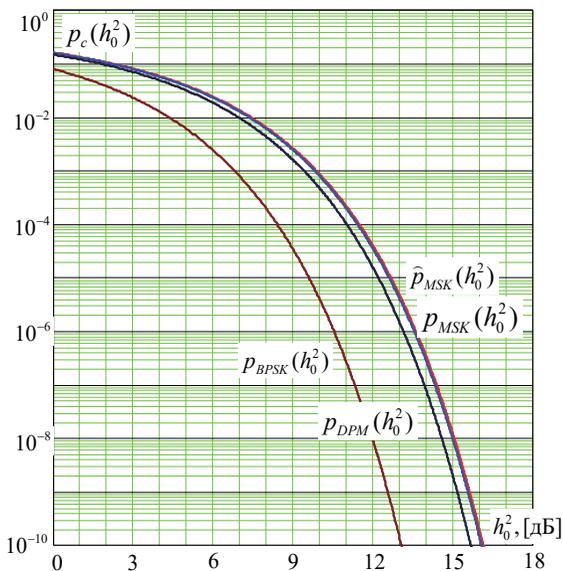
Рис. 10. Функции различий сигналов *MSK* и *DPM* на длительности элемента сигнала

Согласно полученным результатам контрастность сигналов  $\bar{s}_1(t, a=1)$  и  $\bar{s}_0(t, a=0)$  у *MSK* ниже, чем у *DPM*, что свидетельствует о более высокой помехоустойчивости сигналов *DPM*. Данный факт подтверждается при моделировании с использованием выражения (21).

На рисунке 11 представлены графики зависимости вероятности битовой ошибки от  $SNR$   $h_0^2$ , характеризующих помехоустойчивость сигналов *MSK* и *DPM*, в сравнении с сигналами *BPSK*.

Следует заметить, что график вероятности битовой ошибки, полученный для сигнала *MSK* на основе модели (21)  $p_{MSK}(h_0^2)$ , фактически совпадает с величиной  $\hat{p}_{MSK}(h_0^2)$ , рассчитанной в соответствии с общепринятыми выражениями:

$$\hat{p}_i^{MSK} = Q\left[\sqrt{h_0^2}\right] = \frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{h_0^2}{2}}\right). \quad (28)$$

Рис. 11. Вероятность символьной ошибки для сигналов *CPM*

Так, согласно [28], если сигнал, модулированный *MSK*, когерентно обнаруживается в интервале  $T$  секунд как сигнал, модулированный *FSK*, то эта вероятность будет ниже, чем у *BPSK* на 3дБ, что наглядно подтверждает рисунок 11.

Для оценки близости значений  $p_{MSK}(h_0^2)$  и  $\hat{p}_{MSK}(h_0^2)$  на рисунке 12 показана функция ошибки, рассчитанная в соответствии с выражением:

$$\Delta^{MSK}(h_0^2) = |p_{MSK}(h_0^2) - \hat{p}_{MSK}(h_0^2)|. \quad (29)$$

Анализ графика функции  $\Delta^{MSK}(h_0^2)$  показывает, что на уровне  $h_0^2$  порядка 9-10 дБ и выше различия между значениями значений  $p_{MSK}(h_0^2)$  и  $\hat{p}_{MSK}(h_0^2)$  не превышают 0,01%.

Полученный результат соответствия между  $p_{MSK}(h_0^2)$  и  $\hat{p}_{MSK}(h_0^2)$  позволяет путем дальнейшей аппроксимации выражения (21) получить аналитическую модель уже сигнала *DPM*, которую можно представить как:

$$\hat{p}_i^{DPM} = Q\left[\sqrt{1, (1)h_0^2}\right] = \frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{h_0^2}{1,82}}\right). \quad (30)$$

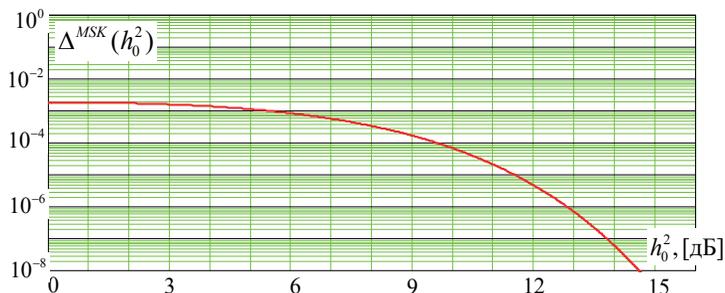


Рис. 12. Зависимость ошибки аппроксимации графика вероятности битовой ошибки для сигнала *MSK* от *SNR*

На рисунке 13 показана зависимость ошибки аппроксимации  $\Delta^{DPM}(h_0^2)$ , определяемой в соответствии с выражением (29), но уже для сигнала *DPM*.

Так, функция  $\Delta^{DPM}(h_0^2)$  характеризует различия между значением  $p_{DPM}(h_0^2)$ , рассчитанным в соответствии с обобщенной моделью для сигналов класса *CPM* (19) и значением  $\hat{p}_{DPM}(h_0^2)$ , вычисление которого осуществлялось согласно аналитической модели (24). Графическая зависимость  $\Delta^{DPM}(h_0^2)$  от *SNR* показана на рисунке 13

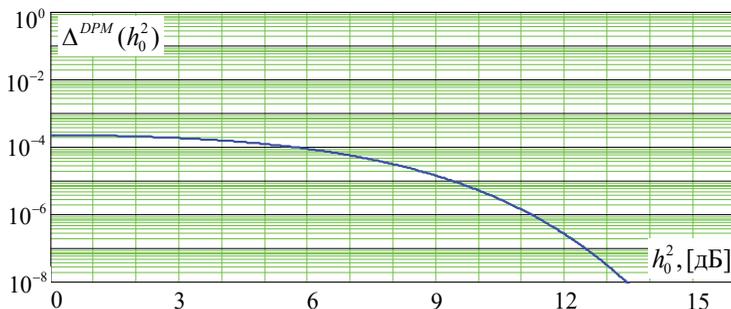


Рис. 13. Зависимость ошибки аппроксимации графика вероятности битовой ошибки для сигнала *DPM* от *SNR*

Анализ изменения функции  $\Delta^{DPM}(h_0^2)$ , представленной на рисунке 13, показывает, что в диапазоне изменения уровня  $h_0^2$  от 0 до 7 дБ величина ошибки составляет порядка  $10^{-4}$  и стремительно снижается к нулю с возрастанием величины ОСШ.

Различия абсолютных значений функций ошибок  $\Delta^{DPM}(h_0^2)$  и  $\Delta^{MSK}(h_0^2)$  связаны с погрешностью расчетов.

Следует отметить уникальность выражения (30), поскольку аналитические выражения для оценки помехоустойчивости сигналов DPM ранее получены не были.

**4. Анализ спектральных свойств сигналов DPM.** Полученный результат позволяет заключить, что сигналы DPM обладают более высокими свойствами помехоустойчивости по отношению к сигналам MSK, что в целом согласуется с выводами, сделанными в [16]. Но при этом сигналам MSK имеют лучшую спектральную компактность.

В частности, на рисунках 14-16 показаны спектры сигналов BPSK, MSK и DPM, модулированных одинаковой случайной последовательностью при равной символьной скорости.

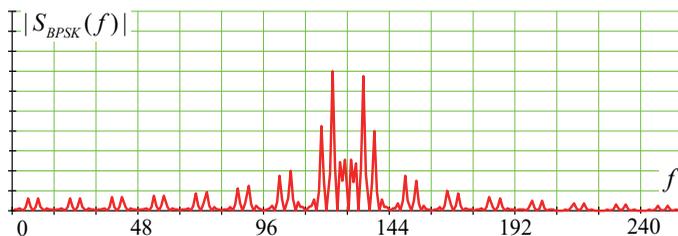


Рис. 14. Спектральное представление сигнала BPSK

Графики спектров построены в результате моделирования в среде MathCAD. В качестве вектора информационных символов была определена последовательность вида:

$$\mathbf{a} = \{0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1\}. \quad (31)$$

Здесь и далее по оси абсцисс представлены дискретные отсчеты частоты, а для построения спектров использовались стандартные процедуры расчета быстрого преобразования Фурье среды MathCAD

Даже визуальный анализ спектральных представлений тестируемых сигналов позволяет судить о достаточно высокой спектральной

эффективности сигналов класса  $CPM$ , к которым относятся  $s_{MSK}(t)$  и  $s_{DPM}(t)$ , по отношению к сигналу  $BPSK$ .

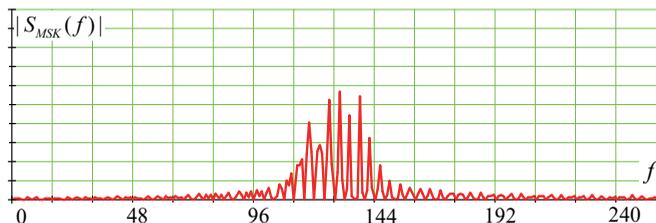


Рис. 15. Спектральное представление сигнала  $MSK$

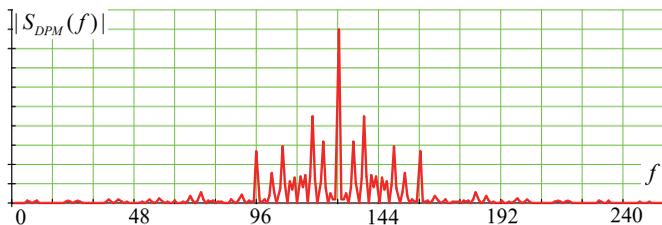


Рис. 16. Спектральное представление сигнала  $DPM$

Вместе с тем спектры сигналов  $s_{MSK}(t)$  и  $s_{DPM}(t)$  имеют существенные различия.

Так спектральные компоненты  $|S_{MSK}(f)|$ , сгруппированные в пределах значимой полосы частот, имеют относительно близкие значения амплитуд, что делает спектр «равномерно плотным». В то же время в структуре спектра сигнала  $s_{DPM}(t)$  наблюдаются мощные спектральные выбросы в пределах значимой полосы частот, характерные для сигналов фазовой манипуляции.

Другой особенностью функции  $|S_{DPM}(f)|$  является явно выраженная ее симметрия относительно центральной компоненты, характерная для сигналов амплитудной манипуляции или угловой модуляции с малым индексом.

В интересах более детального анализа была построена разностная функция  $\Delta_R(f) = |S_{DPM}(f)| - |S_{MSK}(f)|$  (рис. 17). Особенность функции  $\Delta_R(f)$  в том, что ее положительные значения свидетельствуют о превалировании спектральных компонент  $s_{DPM}(t)$ , а отрицательные – указывают на доминирование спектральных компонентов сигнала  $s_{MSK}(t)$ .



деляемую информационной составляющей; фазовращатель (ФВ); коммутирующее устройство (КУ) и фазовый модулятор (ФМ).

Формирование сигнала  $DPM$  в соответствии с представленным устройством осуществляется следующим образом.

Генератор НЧ формирует низкочастотное колебание  $z(t)$ , которое поступает на прямой и инверсный вход коммутирующего устройства. При этом после фазовращателя на инверсный вход КУ поступает инвертируемое колебание  $\bar{z}(t)$ , а на управляющий вход коммутирующего устройства подается информационная импульсная последовательность  $a(t)$ . При этом длительность информационного импульса последовательности  $a(t)$  соответствует периоду колебания  $z(t)$ . В результате на выходе КУ формируется низкочастотный фазоманипулированный сигнал  $u_{BPSK}(t)$ , который поступает на фазовый модулятор, где выступает в качестве модулирующего сигнала для высокочастотного колебания  $w(t)$ . А уже на выходе ФМ формируется результирующий сигнал  $DPM$   $s_{DPM}(t)$ .

**6. Заключение.** Представленный эмпирический подход позволил получить вероятностную модель оценки помехоустойчивости сигналов  $CPM$ , основанную на принципе аналогии по отношению к известным выражениям. При этом ошибка аппроксимации полученных на его основе графиков по отношению к известным моделям, в частности сигнала  $MSK$ , составила десятые доли процента, что позволяет судить о правомерности и состоятельности такого подхода.

Кроме того, принцип аналогий дал возможность впервые получить аналитическое выражение расчета вероятности битовой ошибки в зависимости от ОСШ для сигналов  $DPM$ , также являющихся представителями  $CPM$ .

Справедливости ради стоит отметить, что результат получен для наиболее простого случая – канала  $AWGN$ , поэтому дальнейшие исследования авторы связывают с исследованием канала с райсовскими замираниями [29].

Эмпирический характер обобщенной модели также требует дальнейшего ее изучения и уточнения. В его основу, аналогично минимальному евклидовому расстоянию, положены структурные различия между сигналами, которые соответствуют разным информационным символам. Но если евклидово расстояние – величина детерминированная и в большей степени применима к сигналам, у которых параметры на длительности символа стационарны, то предложенная в работе функция различий применима и к излучениям с вариативными параметрами, к которым относятся сигналы  $CPM$ .

Чтобы повысить точность оценки с использованием разработанного подхода, в дальнейшем планируется использовать методы совместной частотно-временной обработки, предложенные в [30], а также негармонические базисы анализа [31].

### Литературы

1. *Osborne W., Luntz M.* Coherent and Noncoherent Detection CPFSK // IEEE Transactions on Communications. 1974. vol. 22(8). pp. 1023–1036.
2. *Schonhoff T.* Symbol Error Probabilities for m-ary CPFSK: Coherent and Noncoherent Detection // IEEE Transactions on Communications. 1976. vol. 24(6). pp. 644–652.
3. *Аладинский В.А. и др.* Патент на изобретение RU 2261476 C1, 27.09.2005. Способ распознавания радиосигналов.
4. *Aulin T., Rydbeck N., Sundberg C.E.* Continuous Phase Modulation--Part II: Partial Response Signaling // IEEE Transactions on Communications. 1981. vol. 29(3). pp. 210–225.
5. *Miyakawa H., Harashima H., Tatsui N., Tanaka Y.* Digital Phase-Modulation Scheme Using Phase-Continuous Waveform // Electronics and Communications in Japan (English translation of Denshi Tsushin Gakkai Zasshi). 1975. vol. 58(12). pp. 35–42.
6. *Zhou L., Martin P.A., Taylor D.P., Horn C.* MLSE diversity receiver for partial response CPM // 12th IEEE International Conference on Communication Systems 2010. 2010. pp. 501–505.
7. *Qi J. et al.* Research on an optimization method for a partially responsive continuous phase modulated (CPM) signal based on an optimal generic function // Symmetry. 2019. vol. 11. no. 9. pp. 1114.
8. *Zourob M.* Detection, Receivers, and Performance of CPFSK and CPCK" // Electronic Thesis and Dissertation Repository. 2013. 1179 p.
9. Рекомендация МСЭ-R SM.328-11. Спектры и ширина полосы излучений (Вопрос МСЭ-R 222/1).
10. *Кузьмин Е.В., Зограф Ф.Г.* Параметризованная модель квадратурного модулятора MSK-сигнала в OrCAD // Электронный научный журнал «Инженерный вестник Дона». 2016. № 1. С. 2007–2016.
11. *Zhai W., Li Z., Si J., Bai J.* Performance analysis of a joint estimator for timing, frequency, and phase with continuous-phase modulation // IET Communications. 2016. vol. 10. no. 3. pp. 263–271.
12. *Pan X., Liu C., Zhu J.* Phase retrieval with extended field of view based on continuous phase modulation // Ultramicroscopy. 2019. vol. 204. pp. 10–17.
13. *Rhyou C., Park S., Lee H.* Optimal rate for continuous phase modulation in standing surface acoustic waves // Journal of Mechanical Science and Technology. 2019. vol. 33. no. 8. pp. 3819–3829.
14. *Дворников С.С., Дворников С.В.* Формирование сигналов с непрерывной фазой для передачи дискретной информации // Информационные технологии. 2016. Т. 22. № 6. С. 435–440.
15. *Абакумов А.Н., Еремеев И.Ю., Кадуков Е.П.* Модель радиосигналов с модуляцией с непрерывным изменением фазы зарубежных спутниковых систем связи в пространстве параметров фазовых диаграмм // Труды Военно-космической академии имени А.Ф.Можайского. 2019. № 671. С. 71–79.
16. *Дворников С.В., Дворников С.С., Манаенко С.С., Пишеничников А.В.* Спектрально-эффективные сигналы с непрерывной фазой // Вестник Воронежского государственного технического университета. 2016. Т. 12. № 2. С. 87–93.
17. *Тамбовский С.С.* Применение сигналов семейства CPM для организации связи с БПЛА // DSPA: Вопросы применения цифровой обработки сигналов. 2016. Т. 6. № 1. С. 82–84.

18. *Дворников С.С.* Обоснование параметров фазомодулированных сигналов для высокоскоростных систем передачи информации // Вопросы радиоэлектроники. Серия: Техника телевидения. 2016. № 3. С. 43–48.
19. *Дворников С.С.* Спектрально-эффективные формы сигналов с непрерывной фазой для передачи дискретной информации // Вопросы радиоэлектроники. Серия: Техника телевидения. 2016. № 1. С. 86–93.
20. *Султанов А.Х. и др.* Методика определения оптимального коэффициента скругления спектра в беспроводных цифровых системах связи для устранения межсимвольной интерференции // Электротехнические и информационные комплексы и системы. 2013. Т. 9. № 2. С. 96–101.
21. *Yuan R., Cheng J.* Free-Space Optical Quantum BPSK Communications in Turbulent Channels // 2018 IEEE Globecom Workshops (GC Wkshps). 2018. pp. 1–6.
22. *Котельников В.А.* Теория потенциальной помехоустойчивости // М.: Государственное энергетическое изд-во. 1956. 152 с.
23. *Вентцель Е.С.* Теория вероятностей // М.: Академия. 2005. 576 с.
24. *Дворников С.В. и др.* Теоретическое обоснование синтеза ансамбля биортогональных сигналов с повышенной помехоустойчивостью // Вопросы радиоэлектроники. Серия: Техника телевидения. 2015. № 5. С. 16–20.
25. *Yang G. et al.* CPFSK signal detection in white and bursty impulsive noises // IEEE Access. 2019. vol. 7. pp. 74029–74039.
26. *Нахмансон Г.С., Маснев И.Н.* Прием модифицированного фазоманипулированного широкополосного сигнала корреляционным приемником с входным полосовым фильтром // Телекоммуникации. 2020. № 7. С. 17–23.
27. *Левин Б.Р.* Теоретические основы статистической радиотехники // М.: Радио и связь. 1989. 653 с.
28. *Скляр Б.* Цифровая связь. Теоретические основы и практическое применение // М.: Вильямс. 2003. 1104 с.
29. *Osyphchuk S., Moshynska A., Pieshkin A., Shmihel B.* The effect of the features of signal-code constructions forming on indicators of functionality and reliability of communication systems based on the 802.11 N/AC standards // Sciences of Europe. 2018. vol. 26-2(26). pp. 38–47.
30. *Дворников С.В., Сауков А.М.* Модификация частотно-временных описаний нестационарных процессов на основе показательных и степенных функций // Научное приборостроение. 2004. Т. 14. № 3. С. 76–85.
31. *Афанасьев В.В., Логинов С.С.* Негармонический спектральный анализ MULTISCROLL СИСТЕМ с динамическим хаосом // Системы синхронизации, формирования и обработки сигналов. 2019. Т. 10. № 3. С. 4–7.

**Дворников Сергей Викторович** – д-р техн. наук, профессор, профессор кафедры, кафедра радиотехнических и оптоэлектронных комплексов (кафедра 21), Санкт-Петербургский государственный университет аэрокосмического приборостроения (ГУАП). Область научных интересов: общая теория связи, помехозащищенность линий радиосвязи, методы совместной частотно-временной обработки сигналов, вейвлет-анализ. Число научных публикаций – 298. [practicdsv@yandex.ru](mailto:practicdsv@yandex.ru); ул. Большая Морская, 67, 194000, Санкт-Петербург, Россия; р.т.: +7-905-275-50-02.

**Дворников Сергей Сергеевич** – канд. техн. наук, начальник лаборатории, лаборатория боевого применения войск связи, Военная академия связи им. Маршала Советского Союза С.М. Буденного. Область научных интересов: теория передачи сигналов, спектральная эффективность сигналов, помехозащищенность каналов управления и связи радиотехнических систем. Число научных публикаций – 82. [dvornik.92@mail.ru](mailto:dvornik.92@mail.ru); Тихорецкий пр., 3, 194064, Санкт-Петербург, Россия; р.т.: +7-812-247-98-20.

S. DVORNIKOV, S. DVORNIKOV

**EMPIRICAL APPROACH TO ESTIMATING THE IMMUNITY OF PHASE MODULATION SIGNALS WITH A CONTINUOUS PHASE**

*Dvornikov S., Dvornikov S. Empirical Approach to Estimating the Immunity of Phase Modulation Signals with a Continuous Phase.*

**Abstract.** The high spectral efficiency of signals with continuous phase modulation (CPM) has determined their popularity and active use in various radio engineering projects. The uniqueness of the properties of CPM signals is associated with the preservation of the continuity of their phase when changing information messages for the duration of a symbol. At the same time, until recently, of the entire wide class of signals with continuous phase modulation, the most widespread were various variations, the so-called Minimum Shift Keying (MSK) signals. However, these are far from the only representatives of the class of CPM signals with the property of high spectral compactness. This article examines no less interesting signals of this class, formed by means of Dual Phase Modulation (DPM). In particular, analytical expressions of their synthesis are presented, their belonging to the class of CPM signals is substantiated. In addition, the article investigates the temporal properties of the phase function recommended by ITU-R SM.328-11 for the synthesis of signals with continuous phase modulation, presents the time and frequency fragments of MSK signals in comparison with signals with Binary Phase Shift Keying (BPSK). The stages of the analytical derivation of the model of noise immunity of PCM signals in terms of the probability of a bit error based on an empirical approach are presented. The generality of the obtained model with the known expression for MSK signals is shown by studying the difference function of the approximation error (error of the order of  $10^{-3}$ ), which made it possible to obtain a more compact representation of the developed model in relation to DPM signals. It has been proven that DPM signals have higher noise immunity properties in relation to MSK signals (about 0.5 dB at an error level of  $10^{-5}$ ), using the results of studying the difference functions determined by the difference between the signal symbols corresponding to the information values "1" and "0". The directions of further research are determined.

**Keywords:** Signals with Continuous Phase Modulation, Signal Noise Immunity, Bit Error Probability, Spectral Properties of Signals

**Dvornikov Sergey** – Ph.D., Dr.Sci., Professor, Professor, Professor of a Department of Radio Engineering and Optoelectronic Complexes (Department 21), State University of Aerospace Instrumentation (SUAI). Research interests: general theory of communication, noise immunity of radio communication lines, methods of joint time-frequency signal processing, wavelet analysis. The number of publications – 298. [practicdsv@yandex.ru](mailto:practicdsv@yandex.ru); 67, Bolshaya Morskaya str., 194000, St. Petersburg, Russia; office phone: +7-905-275-50-02.

**Dvornikov Sergey** – Ph.D., Head of the Laboratory, Laboratory of Combat use of Forces, Military Academy of Communications named after Marshal of the Soviet Union SM Budenogo. Research interests: theory of signal transmission, spectral efficiency of signals, noise immunity of control and communication channels of radio engineering systems. The number of publications – 82. [dvornik.92@mail.ru](mailto:dvornik.92@mail.ru); 3, Tikhoretskiy pr., 194064, St. Petersburg, Russia; office phone: +7-812-247-98-20.

**References**

1. Osborne W., Luntz M. Coherent and Noncoherent Detection CPFS. *IEEE Transactions on Communications*. 1974. vol. 22(8). pp. 1023–1036.

2. Schonhoff T. Symbol Error Probabilities for m-ary CPFSK: Coherent and Noncoherent Detection. *IEEE Transactions on Communications*. 1976. vol. 24(6). pp. 644–652.
3. Aladinsky V.A. et al. *Patent na izobreteniyu RU 2261476 C1, 27.09.2005*. [Method for recognizing radio signals]. (In Russ.).
4. Aulin T., Rydbeck N., Sundberg C.E. Continuous Phase Modulation--Part II: Partial Response Signaling. *IEEE Transactions on Communications*. 1981. vol. 29(3). pp. 210–225.
5. Miyakawa H., Harashima H., Tatsui N., Tanaka Y. Digital Phase-Modulation Scheme Using Phase-Continuous Waveform. *Electronics and Communications in Japan (English translation of Denshi Tsushin Gakkai Zasshi)*. 1975. vol. 58(12). pp. 35–42.
6. Zhou L., Martin P.A., Taylor D.P., Horn C. MLSE diversity receiver for partial response CPM. 12th IEEE International Conference on Communication Systems 2010. 2010. pp. 501–505.
7. Qi J. et al. Research on an optimization method for a partially responsive continuous phase modulated (CPM) signal based on an optimal generic function. *Symmetry*. 2019. vol. 11. no. 9. pp. 1114.
8. Zourob M. "Detection, Receivers, and Performance of CPFSK and CPCK". Electronic Thesis and Dissertation Repository. 2013. 1179 p.
9. Recommendation ITU-R SM.328-11. [Emission spectra and bandwidth] (Question ITU-R 222/1).
10. Kuzmin E.V., Zograf F.G. [Parameterized model of the quadrature modulator of the MSK signal in OrCAD]. *Elektronnyy nauchnyy zhurnal "Inzhenernyy vestnik Dona" – Electronic scientific journal "Engineering Bulletin of the Don"*. 2016. vol. 1. pp. 2007–2016. (In Russ.).
11. Zhai W., Li Z., Si J., Bai J. Performance analysis of a joint estimator for timing, frequency, and phase with continuous-phase modulation. *IET Communications*. 2016. vol. 10. no. 3. pp. 263–271.
12. Pan X., Liu C., Zhu J. Phase retrieval with extended field of view based on continuous phase modulation. *Ultramicroscopy*. 2019. vol. 204. pp. 10–17.
13. Rhyou C., Park S., Lee H. Optimal rate for continuous phase modulation in standing surface acoustic waves. *Journal of Mechanical Science and Technology*. 2019. vol. 33. no. 8. pp. 3819–3829.
14. Dvornikov S.S., Dvornikov S.V. [Formation of signals with a continuous phase for the transmission of discrete information]. *Informatsionnyye tekhnologii – Information technologies*. 2016. Issue 22. vol. 6. pp. 435–440. (In Russ.).
15. Abakumov A.N., Eremeev I.Yu., Kadukov E.P. [Model of radio signals with modulation with continuous phase change of foreign satellite communication systems in the space of parameters of phase diagrams]. *Trudy Voyenno-kosmicheskoy akademii imeni A.F.Mozhayskogo – Proceedings of the A.F. Mozhaisky Military Space Academy*. 2019. vol. 671. pp. 71–79. (In Russ.).
16. Dvornikov S.V., Dvornikov S.S., Manaenko S.S., Pshenichnikov A.V. [Spectrally effective signals with continuous phase]. *Bulletin of the Voronezh State Technical University – Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta*. 2016. Issue 12. vol. 2. pp. 87–93. (In Russ.).
17. Tambovskiy S.S. [Application of signals of the CPM family for the organization of communication with the UAV]. *DSPA: Digital Signal Processing Application Issues – DSPA: Voprosy primeneniya tsifrovoy obrabotki signalov*. 2016. Issue 6. vol. 1. pp. 82–84. (In Russ.).
18. Dvornikov S.S. [Substantiation of parameters of phase-modulated signals for high-speed data transmission systems]. *Voprosy radioelektroniki. Seriya: Tekhnika televiziyeniya – Radio electronics issues. Series: Television technology*. 2016. vol. 3. pp. 43–48. (In Russ.).

19. Dvornikov S.S. [Spectrally efficient continuous phase waveforms for transferring discrete information]. *Voprosy radioelektroniki. Seriya: Tekhnika televideniya – Radio electronics issues. Series: Television technology*. 2016. vol. 1. pp. 86–93. (In Russ.).
20. Sultanov A.Kh. et al. [Methodology for determining the optimal spectrum rounding factor in wireless digital communication systems to eliminate intersymbol interference]. *Elektrotekhnicheskiye i informatsionnyye komplekxy i sistemy – Electrical and information complexes and systems*. 2013. Issue 9. vol. 2. pp. 96–101. (In Russ.).
21. Yuan R., Cheng J. Free-Space Optical Quantum BPSK Communications in Turbulent Channels. 2018 IEEE Globecom Workshops (GC Wkshps). 2018. pp. 1–6.
22. Kotelnikov V.A. *Teoriya potentsial'noy pomexoustoychivosti* [The theory of potential noise immunity]. M.: Gosudarstvennoye energeticheskoye izd-vo. 1956. 152 p. (In Russ.).
23. Wentzel E.S. *Teoriya veroyatnostey* [Probability theory]. M.: Academia. 2005. 576 p. (In Russ.).
24. Dvornikov S.V. et al. [Theoretical substantiation of the synthesis of an ensemble of biorthogonal signals with increased noise immunity]. *Questions of radioelectronics. Series: Television technology – Voprosy radioelektroniki. Seriya: Tekhnika televideniya*. 2015. vol. 5. pp. 16–20. (In Russ.).
25. Yang G. et al. CPFSSK signal detection in white and bursty impulsive noises. *IEEE Access*. 2019. vol. 7. pp. 74029–74039.
26. Nakhmansov G.S., Masnev I.N. [Reception of a modified phase-shift keyed broadband signal by a correlation receiver with an input bandpass filter]. *Telekommunikatsii – Telecommunications*. 2020. vol. 7. pp. 17–23. (In Russ.).
27. Levin B.R. *Teoreticheskiye osnovy statisticheskoy radiotekhniki* [Theoretical foundations of statistical radio engineering]. M.: Radio i svyaz. 1989. 653 p. (In Russ.).
28. Sklar B. *Tsifrovaya svyaz. Teoreticheskiye osnovy i prakticheskoye primeneniye* [Digital communication. Theoretical foundations and practical application]. M.: Vilyams. 2003. 1104 p. (In Russ.).
29. Osypchuk S., Moshynska A., Pieshkin A., Shmihel B. The effect of the features of signal-code constructions forming on indicators of functionality and reliability of communication systems based on the 802.11 N/AC standards. *Sciences of Europe*. 2018. vol. 26-2 (26). pp. 38–47.
30. Dvornikov S.V., Saukov A.M. [Modification of time-frequency descriptions of non-stationary processes based on exponential and power-law functions]. *Nauchnoye priborostroyeniye – Scientific instrument-making*. 2004. Issue 14. vol. 3. pp. 76–85. (In Russ.).
31. Afanasyev V.V., Loginov S.S. [Nonharmonic spectral analysis of MULTISCROLL SYSTEMS with dynamic chaos]. *Sistemy sinkhronizatsii, formirovaniya i obrabotki signalov – Systems of synchronization, formation and processing of signals*. 2019. Issue 10. vol. 3. pp. 4–7. (In Russ.).

И.Ю. ЕРЕМЕЕВ, М.В. ТАТАРКА, Ф.Л. ШУВАЕВ, А.С. ЦЫГАНОВ  
**АНАЛИЗ МЕР ЦЕНТРАЛЬНОСТИ УЗЛОВ СЕТЕЙ НА ОСНОВЕ  
МЕТОДА ГЛАВНЫХ КОМПОНЕНТ**

*Еремеев И.Ю., Татарка М.В., Шуваев Ф.Л., Цыганов А.С. Анализ мер центральности узлов сетей на основе метода главных компонент.*

**Аннотация.** Анализ сетей разнообразной природы, которыми являются сети цитирования, а также социальные или информационно-коммуникационные сети, включает изучение топологических свойств, позволяющих оценивать взаимосвязи между узлами сети и различные характеристики, такие как плотность и диаметр сети, связанные подгруппы узлов и тому подобное. Для этого сеть представляется в виде графа – совокупности вершин и ребер между ними. Одной из важнейших задач анализа сетей является оценивание значимости узла (или в терминах теории графов – вершины). Для этого разработаны различные меры центральности, позволяющие оценить степень значимости вершин сетевого графа в структуре рассматриваемой сети.

Существующее многообразие мер центральности порождает проблему выбора той, которая наиболее полно описывает значимость центральности узла.

Актуальность работы обусловлена необходимостью анализа мер центральности для определения значимости вершин, что является одной из основных задач изучения сетей (графов) в практических приложениях.

Проведенное исследование позволило с использованием метода главных компонент среди известных мер центральности выявить коллинеарные меры, которые в дальнейшем можно исключить из рассмотрения. Это позволяет уменьшить вычислительную сложность расчетов, что особенно важно для сетей с большим числом узлов, и повысить достоверность интерпретации получаемых результатов при оценивании значимости узла в рамках анализируемой сети при решении практических задач.

Выявлены закономерности представления различных мер центральности в пространстве главных компонент, что позволяет классифицировать их с точки зрения близости образов узлов сети, формируемых в определяемом применяемыми мерами центральности пространстве.

**Ключевые слова:** метод главных компонент, мера центральности, граф, кластеризация, мера сходства

**1. Введение.** При анализе социальных сетей обеспечении информационной безопасности информационно-коммуникационных сетей (ИКС) и других приложений теории графов важную роль играет исследование мер центральности как отдельных узлов, так и сети в целом. На сегодняшний день известно около тридцати различных мер центральности, применяемых при анализе структуры и топологических свойств сетей [1-6], которые базируются на различных подходах к оценке степени важности узла сети. На рисунке 1 в качестве примера представлены три наиболее известные меры центральности с показателями значимости узлов.

Рассмотрим граф:

$$G = (V, E),$$

где  $V$  – вершины графа, а  $E$  – ребра (рис. 1а).

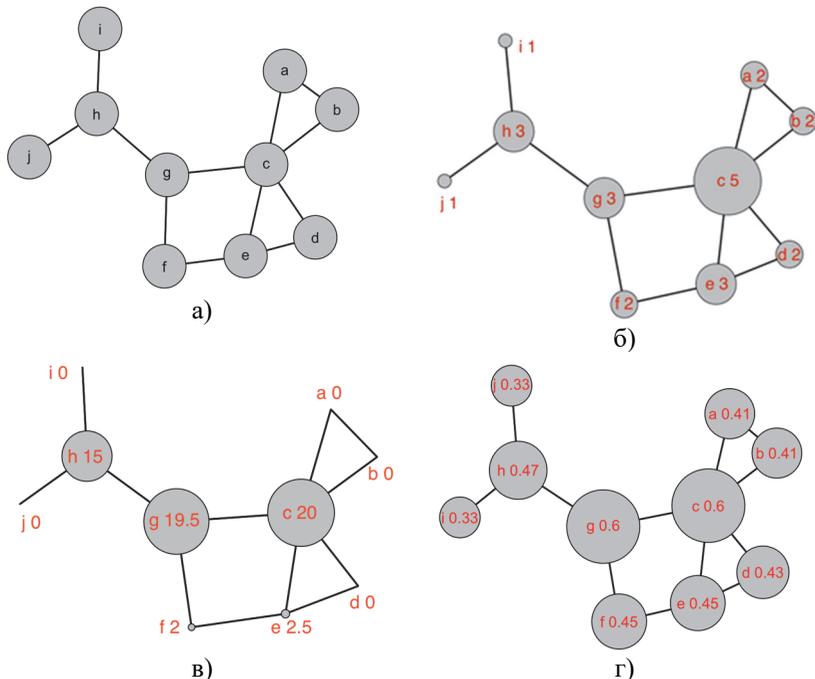


Рис. 1. Пример графов с различными мерами центральности вершин информационно-телекоммуникационных сетей: а) структура рассматриваемой ИКС; б) центральность по степени; в) центральность по посредничеству; г) центральность по близости

Будем полагать, что в рассматриваемом графе ребра являются ненаправленными и имеют одинаковый вес.

Центральность по степени определяется следующим образом (рис. 1б):

$$C_i = \deg(V_i),$$

где  $\deg$  – степень  $i$ -го узла  $V$  графа  $G$  (число ребер узла).

Центральность по посредничеству (рис. 1в) характеризует, насколько важную роль вершина играет на пути «между» парами других вершин графа, в том смысле, что пути между другими вершинами должны проходить через рассматриваемую вершину. Вершина с наибольшим значением центральности по посредничеству является важной, так как позволяет от-

слеживать или контролировать поток информации в сети и может быть рассчитана в соответствии с выражением:

$$P_i = \sum_{v_j < v_k} g_{v_j v_k}^* (V_i) / g_{v_j v_k},$$

где  $g_{v_j v_k}^*$  – общее число кратчайших путей из узла  $V_i$  и  $V_k$ ;  $g_{v_j v_k} (V_i)$  – число кратчайших путей, включающих вершину  $V_i$ .

Центральность по близости (рис. 1г) вершины графа определяется как величина, обратная сумме расстояний от узла  $V_i$  до всех остальных узлов:

$$B_i = (\sum_{V_j=1}^V L(V_i, V_j))^{-1},$$

где  $L(V_i, V_j)$  – расстояние между вершинами графа  $V_i$  и  $V_k$ .

Многообразие способов определения мер центральности позволяет получить огромное количество признаков, характеризующих анализируемые сети, что несомненно повышает требования к вычислительному ресурсу. В таких условиях возникает проблема выбора наиболее информативной меры или некоторого множества мер, позволяющих анализировать топологию исследуемых сетей.

Цель исследования – определить возможность сокращения признаков пространства, формируемого различными мерами центральности узлов анализируемых сетей. Данная тема исследована слабо, так как только в последние 2-3 года появились вычислительные мощности, достаточные для имитационного моделирования и исследования свойств сетей, которые имеют в своем составе тысячи вершин и связывающих их ребер [7-9].

В качестве информативного признака при исследовании выбран вклад мер центральности в формирование главных компонент при проведении компонентного анализа. Решается дополнительная задача, которая заключается в выявлении схожих мер центральности на основе метода главных компонент (МГК) и кластерного анализа. Выявленные группы мер центральности для разных типов сетей сравниваются между собой, что позволяет утверждать о взаимосвязи и однородности мер центральности.

Исследование основывается на базовых положениях теории графов и математической статистики. Совокупность узлов ИКС и связей между ними представлена в виде графа. При этом меры централь-

ности описывают потенциальную значимость вершин графа. Сравнительный анализ мер центральности проводится с использованием ряда реальных сетей, а также модельных сетей: на основе предпочтительно-го присоединения Барабаши – Альберта и модели малого мира Уоттса – Строгатца. Именно эти модели случайных графов наиболее полно описывают топологические свойства реальных сетей [10-13].

**2. Модели сетей и меры центральности узлов.** Для исследования использовались 8 моделируемых сетей: 1-4 – Уоттса – Строгатца, 5-8 – Барабаши – Альберта, и 8 реальных сетей.

**2.1. Модель предпочтительного присоединения Барабаши – Альберта.** Как показывает практика, новые вершины с большей вероятностью соединяются с вершинами, которые занимают выдающееся положение в сети, то есть имеют наивысшие показатели центральности по степени [11, 14]. Для анализа реальных ИКС разработаны свободно масштабируемые модели, самой распространенной среди которых является модель предпочтительного присоединения Барабаши – Альберта [14]. Формируется модель сети в виде графа по следующему принципу:

1. В начальный момент времени  $t = 0$  есть  $V_t$  несвязных вершин.

2. На каждом шаге ( $t = 1, 2, 3, \dots$ ) будем добавлять новую вершину с  $E_t$  ребрами.

3. Количество ребер, с которыми приходит в граф новая вершина, фиксировано, но соединяется она с уже существующей вершиной сети с вероятностью, пропорциональной степени этой вершины.

Максимально приближена к реальным сетям модификация модели Барабаши – Альберта с фиксированным параметром распределения вероятностей соединения вершин [10, 11, 14]. При таком варианте построения модели вводится кортеж распределения вероятностей образования вершин, в котором  $p_1$  – вероятность изолированности вершины;  $p_2$  – вероятность соединения вершины с одной вершиной;  $p_3$  – вероятность соединения вершины с двумя вершинами. Рассмотренный кортеж заменяет этапы 2-3 построения модели Барабаши – Альберта в классическом варианте и позволяет регулировать соотношения между изолированными вершинами и вершинами с необходимым для моделирования количеством связей. На рисунке 2а изображен граф, построенный по модели Барабаши – Альберта без учета распределения вероятностей. Число вершин графа фиксировано и равно 500, а вновь появляющиеся вершины соединяются с вершинами с большей степенью. На рисунке 2б представлен граф, построенный по модели Барабаши – Альберта с уче-

том кортежа распределения вероятностей образования вершин, который имеет следующие значения:  $p_1 = 0.25$ ,  $p_2 = 0.5$ ,  $p_3 = 0.25$ .

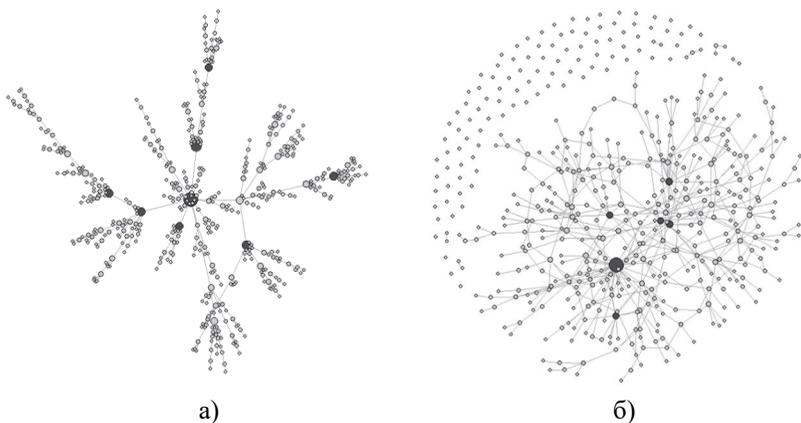


Рис. 2. Граф на основе модели Барабаши – Альберта: а) с равновероятно присоединенными вершинами; б) с заданными вероятностями присоединения вершин

На рисунке 2б видно, что граф содержит изолированные вершины и вершины, обладающие высокой степенью (увеличены в размерах). Таким образом, модель Барабаши – Альберта возможна в двух вариантах: одно- и двухпараметрическом. Параметрами являются число вершин  $V$  и вероятность присоединения фиксированного числа вершин  $p$  на каждом шаге моделирования [10, 11].

**2.2. Модель малого мира Уоттса – Строгатца.** Предложена в 1998 году американскими учеными Д. Уоттсом и С. Строгатцем [15-18]. Модель представляет собой одномерную регулярную решетку, состоящую из  $V$  вершин, причем каждая из них соединена только с  $k$ -ближайшими соседями, и на нее наложены периодические граничные условия (вероятность  $p$  соединения с любыми другими вершинами, кроме соседних, равна нулю), то есть при  $p = 0$  решетка свернута в кольцо. После этого каждая связь с заданной вероятностью  $p$  перебрасывается на другую случайно выбранную вершину. При  $p = 0$  получаем граф с исходной решеткой. На рисунке 3 представлены графы на основе модели малого мира Уоттса – Строгатца. Эти графы имеют четыре различные вероятности переключения ребер, что приводит к различным топологиям графа.

Таким образом, модель малого мира является трехпараметрической, управляемой тремя параметрами: количеством вершин  $V$ , количеством соседних узлов для каждой вершины  $k$ , вероятностью соединения с другими вершинами  $p$ .

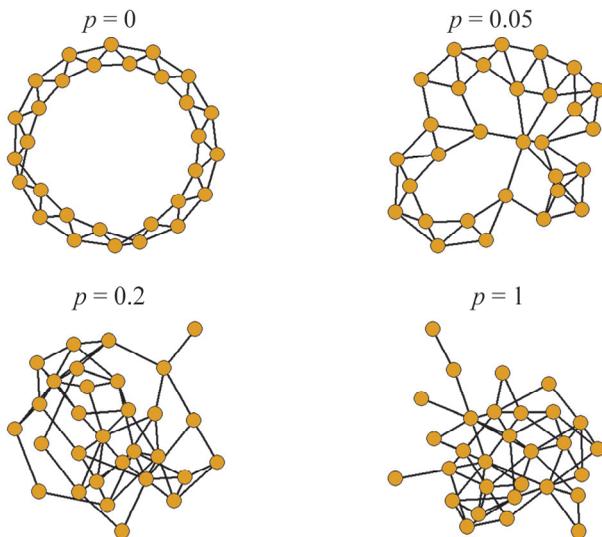


Рис. 3. Граф на основе модели малого мира Уоттса – Стругатца с изменяемой вероятностью переключения ребер

**2.3. Используемые реальные сети.** Реальные сети выбраны из пакета «нетворкдата» (англ. «networkdata») – самого крупного набора сетевых данных для языка статистического моделирования «R». Характеристики сетей представлены в таблице 1. Сети преобразованы к ненаправленным.

Выбранные реальные данные отражают различные стороны человеческой деятельности, а именно: сети 9-10 ИКС в датацентрах одной из коммерческих организаций; 11 – сеть межбелковых взаимодействий в человеческом организме; 12 – сеть взаимодействий между участниками выставки в Дублине в 2009 году; 13 – сеть дружеских отношений между студентами Австралийского национального университета; 14 – данные фиксируют случаи с местных пресс-релизов, выпущенных сенаторами США; сети 15-16 – взаимодействие между персонажами в произведениях Уильяма Шекспира.

Таблица 1. Характеристики исследуемых сетей

Номер сети\ Характеристика	Число вершин	Число ребер	Дополнительные параметры
1	100	100	$p = 0,01; k = 2$
2	100	100	$p = 0,05; k = 2$
3	100	300	$p = 0,1; k = 3$
4	100	400	$p = 0,2; k = 4$
5	100	143	$p = \{0; 0,5; 0,5\}$
6	100	424	$p = \{0; 0,25; 0,15; 0,1; 0,15; 0,1; 0,2; 0,5\}$
7	100	265	$p = \{0; 0,25; 0,25; 0,25; 0,25\}$
8	100	99	Без дополнительных параметров
9	27	69	Без дополнительных параметров
10	26	137	Без дополнительных параметров
11	212	244	Без дополнительных параметров
12	410	17298	Без дополнительных параметров
13	216	2672	Без дополнительных параметров
14	92	477	Без дополнительных параметров
15	47	228	Без дополнительных параметров
16	52	314	Без дополнительных параметров

**2.4. Исследуемые меры центральности.** Наиболее широко распространены около 30 мер центральности [19-21], которые активно продвигаются своими разработчиками. Однако в настоящее время отсутствует четкое понимание алгоритмов выбора и применения мер при различных условиях для определения величины центральности узла графа. Для более наглядного графического представления проводимых в исследовании экспериментов мерам центральности присвоены сокращения (табл. 2) [20, 21].

Таблица 2. Исследуемые меры центральности

№ п/п	Мера центральности	Сокращение	№ п/п	Мера центральности	Сокращение
1	По степени	С	13	Центральность подграфа	ЦП
2	Взвешенная степень вершины	ВСВ	14	Среднее расстояние узла	СР
3	По посредничеству	П	15	Барицентр	БЦ
4	По близости	Б	16	Близость по Фриману	БФ
5	Собственный вектор	СВ	17	Близость по Латору	БЛ
6	Ранг страницы	РС	18	Близость остатков	БО
7	Эксцентриситет	ЭКС	19	Межкликсовая связь	МКС
8	Авторитетность Кляйнберга	АК	20	Плотность максимальной компоненты окрестности узла	ПМКО
9	Концентрация Кляйнберга	КК	21	Линейная центральность	ЛЦ
10	Геодезическая центральность	ГЦ	22	Марковская центральность	МЦ
11	Центральность по вектору Лапласа	ЦЛ	23	Радиальная центральность	РЦ
12	Центральность «рычага»	ЦР	–	–	–

В ходе проведенного исследования были выбраны 23 меры центральности, которые представлены в библиотеке «centiserve» языка программирования «R» [19-21]. Предпочтение отдавалось мерам центральности, не требующим дополнительного параметра для настройки.

**3. Основные этапы сравнительного анализа.** На рисунке 4 представлена структурная схема сравнительного анализа мер центральности узлов графа на основе метода главных компонент. Он состоит из трех основных этапов, каждый из которых разделяется на подэтапы.

**3.1. Этап подготовки данных.** При решении задач машинного обучения и анализа данных около 30-40 % времени уходит на подготовку данных. Некорректно подготовленные данные могут существенно исказить результаты эксперимента.

*1.1. Предобработка данных.* Включает ряд преобразований.

- извлекается гигантская компонента графа, то есть извлекается максимальный связный подграф. Данное действие необходимо для избавления от висячих вершин и преобразования графа к связному;
- удаляются петли графа, то есть ребра, инцидентные одной и той же вершине;
- полученный граф преобразуется к неориентированному.

Рассмотренные преобразования не снижают достоверности дальнейшего анализа в силу того, что большинство мер центральности рассчитываются для связных и неориентированных графов.

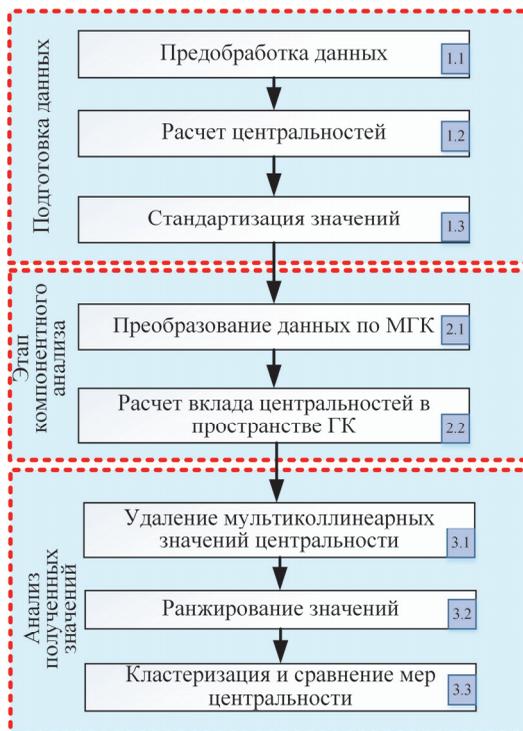


Рис. 4. Структурная схема сравнительного анализа меры центральности узлов сетей на основе метода главных компонент

1.2. Расчет мер центральности осуществляется для каждой вершины каждого из рассматриваемых графов (табл. 1). Полученные значения  $y_{ij}$  помещаются в список данных. При этом каждая ячейка списка соответствует одной исследуемой сети и представляет собой матрицу рассчитанных значений  $N$  мер центральности для  $n$  вершин сетевого графа:

$$Y_{[N,n]} = \|y_{ij}\|_N^n.$$

1.3. Стандартизация значений. Результаты вычислений центральности имеют различную размерность, поэтому выполняется их

нормировка по величине среднеквадратического отклонения значений центральности узлов для каждой меры.

**3.2. Этап компонентного анализа.** Компонентный анализ в рамках метода главных компонент (МГК) представляет собой совокупность статистических приемов обработки данных, которые позволяют сконцентрировать содержащуюся в исходном массиве данных информацию за счет перехода к меньшему числу наиболее информативных факторов – главных компонент (ГК).

2.1. Преобразование данных по МГК. Исходными данными являются стандартизированные значения  $y_{ij}$ . Каждую строку матрицы можно представить как реализацию  $n$ -мерного случайного вектора:

$$\hat{Y}_{\langle n \rangle} = \langle \hat{y}_1, \hat{y}_2, \dots, \hat{y}_n \rangle.$$

Свойства данного случайного вектора с достаточной для практики точностью описываются вектором математических ожиданий:

$$\bar{Y}_{\langle n \rangle} = \langle \bar{y}_1, \bar{y}_2, \dots, \bar{y}_n \rangle$$

и корреляционной матрицей, содержащей линейные коэффициенты корреляции Пирсона:

$$K_{[n]} = \|K_{j_1, j_2}\|_n, \text{ где } K_{j_1, j_2} = \frac{\sum_{i=1}^N (\hat{y}_{j_1, i} - \bar{y}_{j_1})(\hat{y}_{j_2, i} - \bar{y}_{j_2})}{\sqrt{\sum_{i=1}^N (\hat{y}_{j_1, i} - \bar{y}_{j_1})^2 \sum_{i=1}^N (\hat{y}_{j_2, i} - \bar{y}_{j_2})^2}}.$$

Метод ГК основывается на предположении, что любой  $j$ -й признак может быть представлен в виде линейной комбинации ГК  $f_i$ :

$$y_j = a_{1,j}f_1 + a_{2,j}f_2 + \dots + a_{n,j}f_n, \quad [j = 1(1)n],$$

где  $f_1, f_2, \dots, f_n$  – главные компоненты;  $a_{m,j}$  – вес  $m$ -й ГК в  $j$ -м признаке.

Главные компоненты рассчитываются таким образом, чтобы первая из них давала максимально возможный вклад в суммарную дисперсию наблюдений, вторая – максимальный вклад в дисперсию, оставшуюся в суммарной дисперсии за вычетом первой главной компоненты и так далее. Таким образом, задача анализа главных

компонент сводится к тому, чтобы найти такое линейное ортогональное преобразование  $n$  наблюдаемых признаков, которое позволит получить совокупность  $n$  некоррелированных нормированных переменных  $f_i, [i=1(1)n]$ , дисперсии  $\sigma^2$  которых обладают следующим свойством:

$$\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_n^2.$$

Такое преобразование эквивалентно преобразованию исходной корреляционной матрицы  $K_{[n]}$  к матрице вида:

$$K_{[n]}^T K_{[n]} = \begin{vmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{vmatrix}.$$

По дисперсии, соответствующей каждой ГК, можно оценить вклад этой компоненты в формирование общей дисперсии совокупности результатов расчета мер центральности и тем самым выделить существенные компоненты. При этом возможны следующие варианты использования результатов вычисления ГК:

- расчет вклада каждого наблюдаемого признака в ГК и их ранжирование по степени вклада;
- выявление мультиколлинеарных признаков, то есть имеющих тесную корреляционную взаимосвязь;
- разделение наблюдаемых признаков по группам (кластеризация).

2.2. Для каждой из исследуемых сетей осуществляется расчет вклада различных мер центральности в пространстве ГК. Для расчета удельного вклада каждой меры центральности используется следующее выражение [22]:

$$Con_i = \frac{f_1(y_i)f_2(y_i)}{\sum_{k=1}^I f_1(y_k)f_2(y_k)}, [i=1(1)N],$$

где  $f_1(y_i)$  и  $f_2(y_i)$  – значение вклада  $i$ -й меры в ГК1 и ГК2.

Таким образом, значение удельного вклада  $Con$  (от англ. «contribution») является произведением вклада в ГК1 и ГК2  $i$ -й меры центральности, нормированной на сумму вкладов остальных мер в ГК1 и ГК2.

**3.3. Этап анализа полученных значений.** Заключительный этап анализа. На данном этапе полученные результаты интерпретируются.

3.1. Выделение коллинеарных значений центральности. Рассчитанные значения вклада центральностей записываются в матрицу. Чтобы минимизировать ошибку на последующих этапах метода, необходимо сократить признаковое пространство, то есть выявить и удалить дублирующие друг друга значения  $Con$  различных мер центральности.

3.2. Ранжирование полученных значений. Массив со значениями вкладов мер центральности (с учетом предыдущего преобразования) ранжируется для каждой сети.

3.3. Кластеризация и сравнение мер центральности. На данном этапе производится кластеризация рассчитанных значений мер центральности в пространстве ГК  $f_1$  и  $f_2$ , в ходе которой схожие меры разбиваются на группы методом иерархической кластеризации. Далее сравнивается состав кластеров мер центральности для всех исследуемых сетей. В качестве меры схожести результатов кластеризации используется коэффициент Фулкса – Мэллова [23].

**4. Апробация метода.** Полученные в результате этапа 2 значения вклада ГК представлены в таблице 3. Результаты анализа полученных значений на подэтапе 3.1 – 3.2 представлены в таблице 4.

С данными, полученными в результате имитационного моделирования, и с реальными данными производились преобразования в соответствии с этапами сравнительного анализа, представленными на рисунке 4. На подэтапе 1.1 в результате предобработки каждая из 16 сетей (табл. 1) приведена к виду связного неориентированного графа без петель. На подэтапе 1.2 для каждого графа рассчитаны 23 меры центральности вершин (табл. 2). Полученные данные стандартизировались, после чего подавались на второй этап анализа.

На втором этапе выполнен расчет вклада каждой меры центральности в первую и вторую главные компоненты (табл. 3).

Таблица 3. Расчет вклада мер центральностей в первую и вторую главные компоненты

	БЛ	ЛЦ	БЦ	БФ	Б	МЦ	РЦ	СР	БО	ЖС	СВ	АК	КК	П	Ш	КВБ	С	ГМК	РС	ЦР	МКС
1	7,95	7,67	7,67	7,67	7,67	7,61	7,13	7,13	6,83	6,78	5,49	5,49	5,49	4,08	1,47	0,99	0,99	0,86	0,67	0,18	0,08
2	7,32	6,48	6,48	6,48	6,48	7,39	6,25	6,25	7,33	3,85	3,29	3,29	3,29	4,71	4,75	4,01	4,01	0,27	3,15	1,80	1,07
3	6,14	5,54	5,54	5,54	5,54	6,48	5,42	5,42	5,96	1,33	4,26	4,26	4,26	4,94	5,95	5,53	5,53	0,02	5,11	3,53	0,95
4	5,53	5,17	5,17	5,17	5,17	5,61	5,14	5,14	5,40	0,67	5,12	5,12	5,12	4,55	5,51	5,34	5,34	0,13	5,18	4,18	1,75
5	5,30	5,03	5,03	5,03	5,03	5,12	4,26	4,26	5,26	3,22	5,33	5,33	5,33	4,40	5,05	4,86	4,86	0,99	4,52	2,10	4,90
6	5,12	5,06	5,06	5,06	5,06	5,14	4,59	4,59	5,00	2,85	5,10	5,10	5,10	3,90	4,98	5,13	5,13	0,01	5,04	4,29	4,28
7	5,24	5,14	5,14	5,14	5,14	5,13	4,40	4,40	5,12	2,72	5,29	5,29	5,29	4,11	4,94	5,07	5,07	0,05	4,92	3,21	4,80
8	5,75	4,95	4,95	4,95	4,95	5,73	3,98	3,98	5,74	3,54	4,63	3,09	3,09	4,62	5,64	4,83	4,83	4,83	4,62	1,21	4,83
9	5,41	4,82	4,82	4,68	4,68	5,78	3,55	3,55	4,89	1,24	4,12	4,12	4,12	5,54	5,85	5,14	5,82	0,90	5,23	4,30	5,51
10	5,41	5,36	5,36	5,47	5,47	6,08	4,98	4,98	5,30	1,73	2,18	2,18	2,18	3,28	6,07	4,20	6,05	4,86	4,12	4,99	4,24
11	5,76	5,12	5,12	5,12	5,12	5,44	4,42	4,42	5,77	3,74	4,65	4,65	4,65	4,40	5,10	4,27	4,27	3,61	3,36	1,97	4,26
12	6,83	6,18	6,18	6,18	6,18	6,48	5,89	5,89	6,70	3,91	2,88	2,88	2,88	1,55	6,01	6,01	6,01	0,02	4,57	3,28	1,27
13	4,77	5,34	5,34	5,32	5,32	6,56	5,02	5,02	4,05	1,63	4,18	4,18	4,18	2,95	6,46	5,07	6,53	0,98	4,87	4,81	2,23
14	5,04	4,99	4,99	4,99	4,99	5,03	4,42	4,42	4,91	2,66	5,03	5,03	5,03	3,26	4,85	5,04	5,04	3,64	4,92	3,89	3,47
15	5,56	4,80	4,80	4,39	4,39	6,14	4,08	4,08	5,15	2,91	4,47	4,47	4,47	4,44	6,22	4,37	5,83	2,99	4,28	3,16	3,18
16	5,42	5,21	5,21	5,13	5,13	5,79	4,63	4,63	5,27	1,05	3,39	3,39	3,39	3,27	5,61	4,95	5,84	3,74	4,99	5,08	3,70

Таблица 4. Ранжированные значения вклада мер центральности для анализируемых сетей

Ц/№ сети	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
МЦ	1	1	1	1	2	3	1	9	4	1	2	1	2	3	2	2
БЛ	2	4	2	5	7	6	7	8	9	6	4	2	12	6	5	4
БО	3	2	6	7	8	10	11	7	10	8	3	5	13	10	9	7
ЛЦ	4	8	9	9	10	9	10	13	11	7	11	8	8	8	11	8
Б	5	7	8	8	11	8	9	14	13	5	12	7	9	9	12	10
РЦ	6	12	10	11	14	11	14	15	15	10	15	9	11	12	15	12
КК	7	10	13	10	3	5	2	11	14	16	14	13	10	5	10	16
П	8	9	11	14	13	14	13	12	6	15	9	15	14	16	8	15
ЭКС	9	13	15	16	15	16	16	16	16	17	16	12	16	17	17	17
ЦЛ	10	3	3	2	1	7	5	1	1	2	1	6	3	7	1	3
ВСВ	11	6	5	3	5	2	3	4	8	13	6	3	4	2	6	11
С	12	5	4	4	4	1	4	3	2	3	5	4	1	1	3	1
РС	13	11	7	6	9	4	6	10	7	14	13	10	5	4	7	9
ЦР	14	15	12	12	16	13	15	17	12	9	17	11	6	13	13	6
ЦП	15	14	14	13	12	12	12	2	3	4	8	14	7	11	4	5
МКС	16	16	16	15	6	15	8	5	5	12	7	16	15	15	14	13
ПМКО	17	17	17	17	17	17	17	6	17	11	10	17	17	14	16	14

На подэтапе 3.1 для поиска коллинеарных значений проанализирован расчет вклада центральностей (табл. 3). Выявлено 9 коллинеарных мер центральности, заштрихованных в таблице 3. Коллинеарные меры центральности распределены по следующим группам:

- первая группа: ЛЦ || БЦ || БФ || Б;
- вторая группа: РЦ || СР;
- третья группа: СВ || АК || КК.

Меры ВСВ и С коллинеарны частично, поэтому исключать их из дальнейших расчетов не будем.

После удаления коллинеарных мер БЦ, БФ, Б, СР, СВ, АК, оставшиеся меры центральности узлов ранжируются в зависимости от вклада *Con* для каждой сети. Результаты такого преобразования представлены в таблице 4. Среди мер центральности видны явные аутсайдеры – это меры центральности ПМКО, МКС, ЭКС. Первую позицию, в большинстве сетей занимает мера марковской центральности (МЦ).

МЦ – это мера, основанная на концепции случайного обхода графа. В ней используется среднее время первого прохождения от каждой вершины до каждой другой вершины. Оно показывает насколько тесно каждая вершина связана с другой. Среднее время первого прохода от вершины *a* к вершине *b* – это среднее число шагов,

получаемые в процессе случайного блуждания. Случайные блуждания с большей вероятностью быстрее достигают вершин, занимающих главенствующее положение в сети.

В результате изучения публикаций по данной предметной области установлено, что МЦ незаслуженно обделена вниманием исследователей. Так, в русскоязычном сегменте сети Интернет не найдено ни одного научного труда, посвященного этой мере центральности. В англоязычном – статьи, датированные 2003 годом [24].

На рисунках 5 и 6 представлены корреляционные окружности для некоторых из исследуемых сетей. Корреляционная окружность – это способ визуализации результатов анализа данных методом главных компонент. Окружность имеет единичный радиус, что соответствует суммарному вкладу первой и второй компонент. Чем ближе к ней значение координат вектора, характеризующего ту или иную меру центральности, тем выше его вклад в указанные компоненты. Схожие по степени вклада в ГК1 и ГК2 меры центральности расположены рядом.

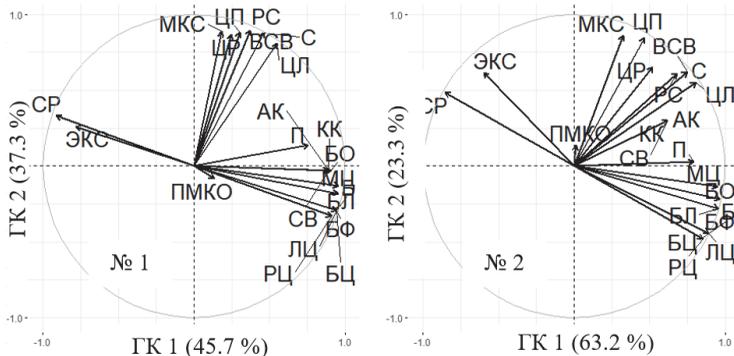


Рис. 5. Корреляционные окружности для моделей сетей № 1-2

На рисунке 5 корреляционные окружности сверху: для моделей сетей 1 и 2, построенных на базе модели Уоттса – Строгатца (слева) и на базе модели Барабаша – Альберта (справа). На рисунке 5 видно, что для каждой сети обособлена мера СР и ЭКС. Остальные меры имеют склонность к группированию и сильно коррелированы между собой.

Представленные на рисунке 6 корреляционные окружности относятся реальным сетям 9 (слева) и 10 (справа). Обособленность мер ЭКС и ПМКО сохраняется и для реальных сетей. Остальные меры так же как в смоделированных сетях склонны образовывать группы и имеют положительную высокую корреляцию между собой.

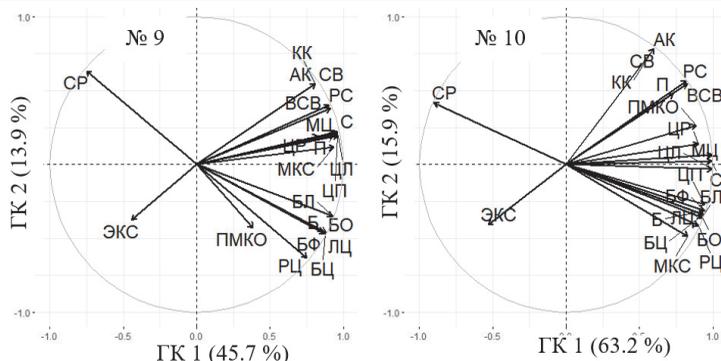


Рис. 6. Корреляционные окружности для реальных сетей № 9-10

Закономерности группировки данных позволяют применить методы кластерного анализа для оценивания близости мер центральности. Результатом иерархической кластеризации мер центральности в пространстве ГК является дендрограмма или дерево, построенное по матрице сходства, которая определяет расстояние между парами кластеров. Самыми распространенными методами расчета расстояния являются метод одиночной связи, метод полной связи, метод средней связи, центроидный метод и метод Уорда. Для выбора метода расчета расстояния между кластерами воспользуемся коэффициентом попарной корреляции Пирсона строк матрицы сходства и матрицы кофенетического расстояния результатов кластеризации, выполненной с использованием различных методов расчета расстояния. Результаты расчетов приведены в таблице 5.

Таблица 5. Значения коэффициента корреляции для различных методов расчета расстояния

Вид расстояния	Уорда	Одиночной связи	Полной связи	Средней связи	Центроидный
Корреляция	0,95	0,867	0,789	0,829	0,851

Из таблицы 5 следует, что наилучшим является расстояние Уорда, так как корреляция между матрицей расстояний Уорда и матрицей кофенетического расстояния наибольшая. Следовательно, расстояние Уорда и будем применять при построении дендрограмм. Результаты иерархической кластеризации представлены в виде дендрограмм на рисунках 7 и 8.

На рисунке 7 видно, что группы мер центральности, полученные в результате кластеризации, в основном схожи. Из общей картины выделяются кластеры ЭКС и ПМКО, число кластеров для всех сетей

равно четырем. Таким образом, меры центральности делятся на 4 кластера, из которых можно выделить два крупных и два малых. Малые кластеры имеют различный состав, что позволяет их отнести к случайным выбросам. Для выявления различий между кластерами, образованными мерами центральности, рассчитаем их попарное сходство. Для этого будем использовать коэффициент Фолкса – Мэллова [23].

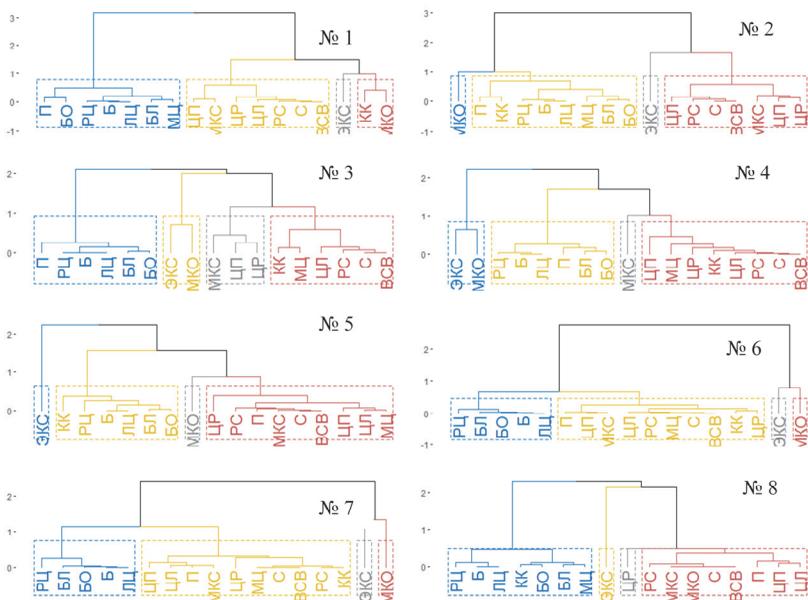


Рис. 7. Результаты кластеризации мер центральности для исследуемых сетей, построенных на базе моделей (№ 1-8)

Коэффициент Фолкса – Мэллова – это мера сходства между двумя результатами кластеризации. Рассмотрим, как рассчитывается коэффициент Фолкса – Мэллова. При сравнении результатов кластеризации при условии одинакового количества кластеров возможны 4 варианта положения объектов кластеризации (точек):

1.  $YY$  – количество объединенных в одинаковые кластера точек при кластеризации  $A_1$  и  $A_2$ .
2.  $YN$  – количество точек присутствующих в одном кластере при кластеризации  $A_1$ , но не  $A_2$ .
3.  $NY$  – количество точек присутствующих в одном кластере при кластеризации  $A_2$ , но не  $A_1$ .
4.  $NN$  – количество точек не совпадающих ни в  $A_1$  ни в  $A_2$ .

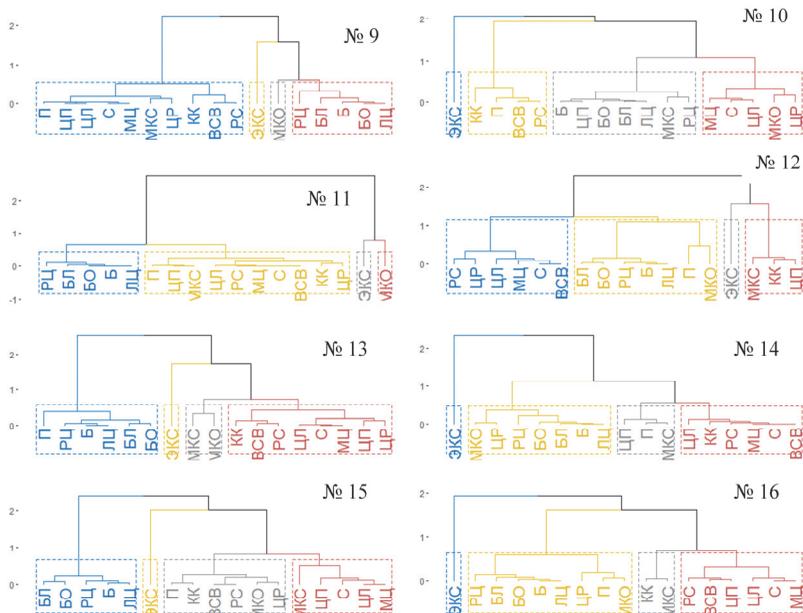


Рис. 8. Результаты кластеризации мер центральности для исследуемых реальных сетей (№ 9-16)

Коэффициент Фолкса – Мэллова рассчитывается согласно выражению:

$$C = \frac{YU}{\sqrt{(YU + YN) \times (YU + NU)}}$$

Коэффициент изменяется в пределах  $[-1; 1]$ . Чем его значение выше, тем более похожи друг на друга результаты кластеризации.

Результаты расчета коэффициента Фолкса – Мэллова представлены в виде матрицы попарных сравнений на рисунке 9.

Таким образом, можно утверждать, что рассматриваемые меры центральности узлов сгруппированы примерно одинаково для всех исследуемых сетей, что говорит об устойчивости полученных результатов. По матрице попарных сравнений можно сделать вывод, что группировка мер центральности имеет схожесть для всех исследуемых 16 сетей. Значения коэффициента Фолкса – Мэллова не опускаются ниже 0,41, а максимальные значения в ряде случаев достигают 1.

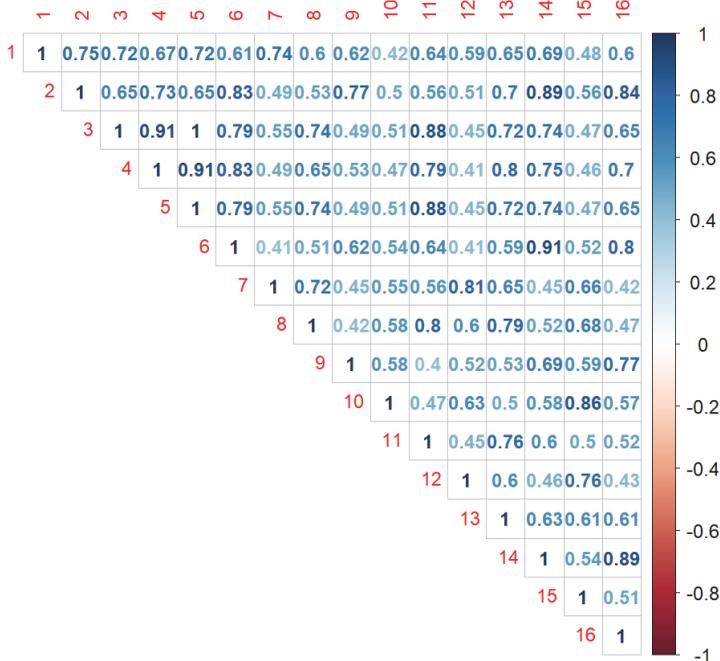


Рис. 9. Матрица попарных сравнений меры сходства кластеров на основе коэффициента Фулкса – Мэллова

Рассчитаем статистическую значимость положения меры в ранжированном списке ее вклада в ГК. На рисунке 10 в виде столбчатой диаграммы показана зависимость статистической значимости для мер центральности от значения вклада в первые три ГК. Статистическая значимость рассчитывалась в соответствии с критерием согласия Пирсона. Выбор только трех ГК обусловлен тем, что в конкретном исследовании они определяют 95% обобщенной дисперсии данных.

Как видно на рисунке 10, наибольший вклад при оценке важности узла сети вносят меры ЦП, С, МЦ. Значения по вертикальной оси пропорциональны наблюдаемым частотам встречаемости меры центральности на соответствующей позиции значимости по всем исследуемым графам.

Цветом обозначена значимость отклонения ожидаемых и наблюдаемых частот в этой ячейке, если значения нормализованных стандартизированных остатков больше 0.025, можно считать, что в этой ячейке зафиксированы статистически значимые отклонения, что говорит о степени важности меры центральности в контексте всех исследованных сетей.

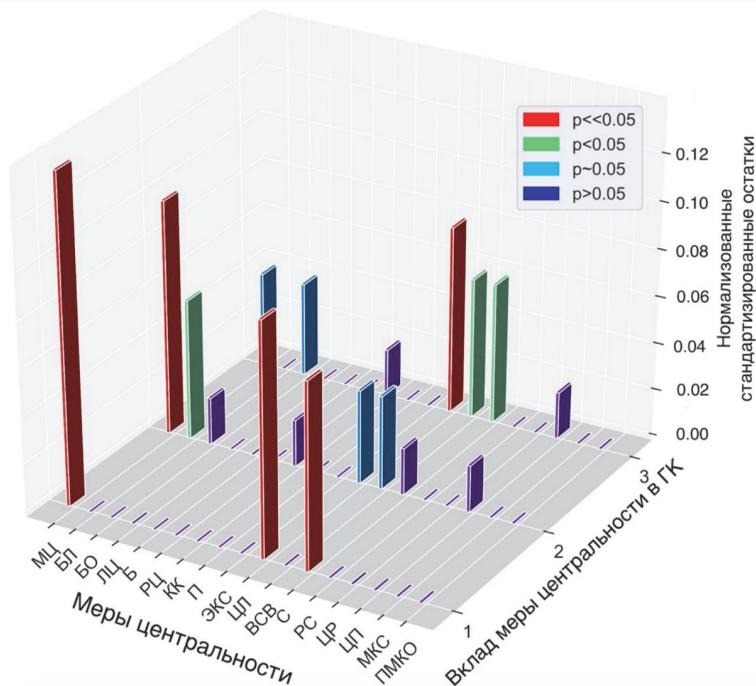


Рис. 10. Значения вклада мер центральности в первые три компоненты по всем исследуемым графам сетей

При этом под ожидаемыми частотами принимаются вычисленные по критерию Хи-квадрат равновероятные частоты, иначе говоря, отклонение расстояния Пирсона в значимых пределах позволяет сказать о статистической значимости положения меры центральности в ранжированном списке ее вклада в ГК в сравнении с равновероятным распределением.

Таким образом, на основании проведенного сравнительного анализа статистически обоснована значимость мер центральности, что позволяет в дальнейшем начинать анализ структуры ИКС с наиболее значимых мер. Кроме того, обоснована возможность сокращения признакового пространства, формируемого различными мерами центральности узлов при анализе ИКС различной природы.

**5. Заключение.** Полученные результаты составляют определенный вклад в развитие приложений теории графов в задачах исследования сетей разнообразной природы и обосновывают возможность сокращения признакового пространства, формируемого различными мерами центральности узлов анализируемых сетей.

Использован набор данных о модельных и реальных информационно-коммуникационных сетях различной природы, для которых меры центральности узлов имеют достаточно высокий уровень статистической значимости, что соответствует коэффициенту кластеризации на уровне не менее 0,7.

Анализ 23 мер центральности узлов сети методом главных компонент показал, что 9 исследуемых мер центральности являются коллинеарными. Коллинеарные меры центральности распределены по трем группам. Первая группа включает линейную центральность, барицентральность, центральность по Фриману и по близости. Вторая группа – центральность по среднему расстоянию и центральность «рычага». Третья группа – центральность по собственному вектору, центральность по авторитетности Кляйнберга и по концентрации Кляйнберга.

В результате ранжирования вклада мер центральности в значения главных компонент для каждой сети выявлено, что меры центральности по плотности максимальной компоненты окрестности узла, межкликковой связи и эксцентриситету являются малозначимыми, а первую позицию в большинстве сетей занимает марковская центральность.

Обоснованное применение метода Уорда для расчета расстояния между кластерами, формируемыми различными мерами центральности в пространстве главных компонент позволило, осуществив кластеризацию, сделать вывод о группировании мер центральности по четырем кластерам. По матрице попарных сравнений результатов кластеризации, полученной с использованием коэффициента Фоулкса – Мэллова, можно сделать вывод о том, что группировка мер центральности имеет схожесть для всех исследуемых 16 сетей, что говорит об устойчивости полученных результатов.

Наибольший вклад в результат оценивания важности узла сети вносят меры центральности по степени, марковской центральности и центральности подграфа.

Для дальнейших исследований представляют интерес меры центральности по среднему расстоянию узла и по эксцентриситету, так как они расположены обособленно в пространстве главных компонент относительно остальных мер центральности, а также меры центральности по степени, марковской центральности и центральности подграфа, как наиболее значимые при оценке важности узла сети.

### **Литература**

1. *Bonchi F., De Francisci G., Riondato M.* Centrality Measures on Big Graphs: Exact, Approximated, and Distributed Algorithms // Proceedings of the 25th International Conference Companion on World Wide Web. 2016. pp. 1017–1020.

2. *Щербакова Н.Г.* Меры центральности в сетях // Проблемы информатики. 2015. № 1. С. 18–30.
3. *Берейхин С.В., Ляпунов В.М., Щербакова Н.Г.* Мера важности научной периодики – «Центральность по посредничеству» // Проблемы информатики. 2014. № 3. С. 53–63.
4. *Юдина М.Н.* Узлы в социальных сетях: меры центральности и роль в сетевых процессах // Омский научный вестник. 2016. № 4. С. 161–165.
5. *Brandes U., Borgatti S., Freeman L.* Maintaining the duality of closeness and betweenness centrality // Social Networks. 2016. vol. 44. pp. 153–159.
6. *Minoo A. et al.* A Systematic Survey of Centrality Measures for Protein-Protein Interaction Networks // BMC Systems Biology. 2018. vol. 12. no. 1. pp. 80.
7. *Chen P-Y., Choudhury S., Hero A.,* Multi-centrality graph spectral decompositions and their application to cyber intrusion detection // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2016. pp. 4553–4557.
8. *Lu B., Sun H., Harris P., Xu M.* Shp2graph: Tools to Convert a Spatial Network into an Igraph Graph in R // ISPRS Int. J. Geo-Inf. 2018. vol. 7. pp. 293.
9. *Csardi G., Nepusz T.* The IGRAPH software package for complex network research // InterJournal, Complex Systems. 1695. 2006. vol. 1695. no. 5. pp. 1–9.
10. *Шуваев Ф.Л., Татарка М.В.* Анализ динамики мер центральности математических моделей случайных графов // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 2. С. 249–256.
11. *Шуваев Ф.Л., Татарка М.В.* Анализ математических моделей случайных графов, применяемых в имитационном моделировании информационно-коммуникационных сетей // Вестник Санкт-Петербургского университета ГПС МЧС России. 2020. № 2. С. 67–77.
12. *Van Mieghem P., Ge X., Schumm P., Trajanovski S., Wang H.* Spectral graph analysis of modularity and assortativity // Phys. Rev. 2010. vol. 82. no. 5. P. 056113.
13. *Barzel B., Biham O.* Quantifying the connectivity of a network: the network correlation function method // Phys. Rev. 2009. vol. 80. pp. 046104.
14. *Barabasi A.* Network Science // Cambridge university press. 2016. 453 p.
15. *Watts D., Strogatz H.* Collective dynamics of «Small-world» networks // Nature. 1998. vol. 393. pp. 440–442.
16. *Hartmann A., Mézard M.* Distribution of diameters for Erdős-Rényi random graphs // Phys. Rev. 2018. vol. 97. no. 3. pp. 032128.
17. *Le C., Levina E., Vershynin R.* Concentration and regularization of random graphs // Random Structures & Algorithms. 2017. vol. 51. no. 3. pp. 538–561.
18. *Gibson H., Vickers P.* Using adjacency matrices to lay out larger small-world networks // Applied soft computing. 2016. vol. 42. pp. 80–92.
19. *Jalili M. et al.* CentiServer: A Comprehensive Resource, Web-Based Application and R Package for Centrality Analysis // PLoS ONE. 2015. vol. 10. no. 11. pp. 0143111.
20. *Oldham, S. et al.* Consistency and differences between centrality measures across distinct classes of networks // PLoS ONE. 2019. vol. 14. no. 7. pp. 0220061.
21. *Bloch F., Jackson M., Tebaldi P.* Centrality measures in networks // SSRN. 2016. 42 p.
22. *Lê S., Josse J., Husson F.* FactoMineR: A Package for Multivariate Analysis // Journal of Statistical Software. 2008. vol. 25. no. 1. pp. 1–18.
23. *Depaolini M., Ciucci D., Calegari S., Dominoni M.* External Indices for Rough Clustering // Lecture Notes in Computer Science. 2018. vol. 11103. pp. 378–391.
24. *White S., Smyth P.* Algorithms for estimating relative importance in networks // Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003. pp. 266–275.

**Еремеев Игорь Юрьевич** – д-р техн. наук, профессор, профессор, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: обработка сигналов со сложной частотно-временной структурой в условиях априорной неопределенности относительно параметров сигналообразования. Число научных публикаций – 100. [eremeeviu@yandex.ru](mailto:eremeeviu@yandex.ru); ул. Ждановская, 13, 197198, Санкт-Петербург, Россия; р.т.: +7(812)347-97-70; факс: +7(812)347-95-57.

**Татарка Максим Васильевич** – канд. техн. наук, докторант, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: теория вероятностей и математическая статистика, распознавание образов, анализ больших объемов данных. Число научных публикаций – 22. [maksimtbv@gmail.com](mailto:maksimtbv@gmail.com); ул. Ждановская, 13, 197198, Санкт-Петербург, Россия; р.т.: +7(812)347-97-70; факс: +7(812)237-12-49.

**Шуваев Федор Леонидович** – канд. техн. наук, научный сотрудник, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: теория вероятностей и математическая статистика, распознавание образов. Число научных публикаций – 8. [cadetfed@mail.ru](mailto:cadetfed@mail.ru); ул. Ждановская, 13, 197198, Санкт-Петербург, Россия; р.т.: +7(812)347-97-70; факс: +7(812)237-12-49.

**Цыганов Андрей Сергеевич** – канд. техн. наук, старший преподаватель, Военно-космическая академия имени А.Ф. Можайского (ВКА им. А.Ф. Можайского). Область научных интересов: теория вероятностей и математическая статистика, распознавание образов, цифровая обработка сигналов. Число научных публикаций – 18. [rogudchik@mail.ru](mailto:rogudchik@mail.ru); ул. Ждановская, 13, 197198, Санкт-Петербург, Россия; р.т.: +7(812)347-97-70; факс: +7(812)237-12-49.

I. EREMEEV, M. TATARKA, F. SHUVAEV, A. CYGANOV  
**COMPARATIVE ANALYSIS OF CENTRALITY MEASURES OF  
NETWORK NODES BASED ON PRINCIPAL COMPONENT  
ANALYSIS**

*Eremeev I., Tatarka M., Shuvaev F., Cyganov A. Comparative Analysis of Centrality Measures of Network Nodes based on Principal Component Analysis.*

**Abstract.** The analysis of networks of a diverse nature, which are citation networks, social networks or information and communication networks, includes the study of topological properties that allow one to assess the relationships between network nodes and evaluate various characteristics, such as the density and diameter of the network, related subgroups of nodes, etc. For this, the network is represented as a graph – a set of vertices and edges between them. One of the most important tasks of network analysis is to estimate the significance of a node (or in terms of graph theory – a vertex). For this, various measures of centrality have been developed, which make it possible to assess the degree of significance of the nodes of the network graph in the structure of the network under consideration.

The existing variety of measures of centrality gives rise to the problem of choosing the one that most fully describes the significance and centrality of the node.

The relevance of the work is due to the need to analyze the centrality measures to determine the significance of vertices, which is one of the main tasks of studying networks (graphs) in practical applications.

The study made it possible, using the principal component method, to identify collinear measures of centrality, which can be further excluded both to reduce the computational complexity of calculations, which is especially important for networks that include a large number of nodes, and to increase the reliability of the interpretation of the results obtained when evaluating the significance node within the analyzed network in solving practical problems.

In the course of the study, the patterns of representation of various measures of centrality in the space of principal components were revealed, which allow them to be classified in terms of the proximity of the images of network nodes formed in the space determined by the measures of centrality used.

**Keywords:** Principal Component Analysis, Measure of Centrality, Graph, Clustering, Measure of Similarity

**Eremeev Igor** – Ph.D., Dr.Sci., Professor, Professor, Mozhaisky Military Space Academy. Research interests: difficult time-and-frequency structure signal processing under the conditions generation signal parameters prior indetermination. The number of publications – 100. eremeeviu@yandex.ru; 13, Zdanovskaya str., 197198, St. Petersburg, Russia; office phone: +7(812)347-97-70; fax: +7(812)347-95-57.

**Tatarka Maxim** – Ph.D., Doctoral Student, Mozhaisky Military Space Academy. Research interests: probability theory and mathematical statistics, pattern recognition, analysis of data. The number of publications – 22. maksimtbv@gmail.com; 13, Zdanovskaya str., 197198, St. Petersburg, Russia; office phone: +7(812)347-97-70; fax: +7(812)237-12-49.

**Shuvaev Fedor** – Ph.D., Researcher, Mozhaisky Military Space Academy. Research interests: probability theory and mathematical statistics, pattern recognition, analysis of data. The number of publications – 8. cadetfed@mail.ru; 13, Zdanovskaya str., 197198, St. Petersburg, Russia; office phone: +7(812)347-97-70; fax: +7(812)237-12-49.

**Tsyganov Andrey** — Ph.D., Senior Lecturer, Mozhaisky Military Space Academy. Research interests: probability theory and mathematical statistics, pattern recognition. The number of publications — 18. porudchik@mail.ru; 13, Zdanovskaya str., 197198, St. Petersburg, Russia; office phone: +7(812)347-97-70; fax: +7(812)237-12-49.

## References

1. Bonchi F., De Francisci G., Riondato M. Centrality Measures on Big Graphs: Exact, Approximated, and Distributed Algorithms. Proceedings of the 25th International Conference Companion on World Wide Web. 2016. pp. 1017–1020.
2. Shherbakova N.G. [Centrality measures in networks]. *Problemy informatiki – Informatics problems*. 2015. vol. 1. pp. 18–30. (In Russ.).
3. Beredihin S.V., Ljapunov V.M., Shherbakova N.G. [Measure of importance of scientific periodicals – «Centrality by mediation»]. *Problemy informatiki – Informatics problems*. 2014. vol. 3. pp. 53–63. (In Russ.).
4. Judina M.N. [Nodes in social networks: measures of centrality and role in network processes]. *Omskij nauchnyj vestnik – Omsk Scientific Bulletin*. 2016. № 4. pp. 161–165. (In Russ.).
5. Brandes U., Borgatti S., Freeman L. Maintaining the duality of closeness and betweenness centrality. *Social Networks*. 2016. vol. 44. pp. 153–159.
6. Minoo A. et al. A Systematic Survey of Centrality Measures for Protein-Protein Interaction Networks. *BMC Systems Biology*. 2018. vol. 12. no. 1. pp. 80.
7. Chen P-Y., Choudhury S., Hero A., Multi-centrality graph spectral decompositions and their application to cyber intrusion detection. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2016. pp. 4553–4557.
8. Lu B., Sun H., Harris P., Xu M. Shp2graph: Tools to Convert a Spatial Network into an Igraph Graph in R. *ISPRS Int. J. Geo-Inf*. 2018. vol. 7. pp. 293.
9. Csardi G., Nepusz T. The IGRAPH software package for complex network research // *InterJournal, Complex Systems*. 1695. 2006. vol. 1695. no. 5. pp. 1–9.
10. Shuvaev F.L., Tatarka M.V. [Analysis of the dynamics of measures of centrality of mathematical models of random graphs]. *Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki – Scientific and technical bulletin of information technologies, mechanics and optics*. 2020. Issue 20. vol. 2. pp. 249–256. (In Russ.).
11. Shuvaev F.L., Tatarka M.V. [Analysis of mathematical models of random graphs used in the simulation of information and communication networks]. *Vestnik Sankt-Peterburgskogo universiteta GPS MChS Rossii – Bulletin of St. Petersburg University State Fire Service of the Ministry of Emergencies of Russia*. 2020. vol. 2. pp. 67–77. (In Russ.).
12. Van Mieghem P. et al. Spectral graph analysis of modularity and assortativity. *Phys. Rev*. 2010. vol. 82. no. 5. P. 056113.
13. Barzel B., Biham O. Quantifying the connectivity of a network: the network correlation function method. *Phys. Rev*. 2009. vol. 80. pp. 046104.
14. Barabasi A. Network Science. Cambridge university press. 2016. 453 p.
15. Watts D., Strogatz H. Collective dynamics of «Small-world» networks. *Nature*. 1998. vol. 393. pp. 440–442.
16. Hartmann A., Mézard M. Distribution of diameters for Erdős-Rényi random graphs. *Phys. Rev*. 2018. vol. 97. no. 3. pp. 032128.
17. Le C., Levina E., Vershynin R. Concentration and regularization of random graphs. *Random Structures & Algorithms*. 2017. vol. 51. no. 3. pp. 538–561.
18. Gibson H., Vickers P. Using adjacency matrices to lay out larger small-world networks. *Applied soft computing*. 2016. vol. 42. pp. 80–92.
19. Jalili M. et al. CentiServer: A Comprehensive Resource, Web-Based Application and R Package for Centrality Analysis. *PLoS ONE*. 2015. vol. 10. no. 11. pp. 0143111.
20. Oldham, S. et al. Consistency and differences between centrality measures across distinct classes of networks. *PLoS ONE*. 2019. vol. 14. no. 7. pp. 0220061.
21. Bloch F., Jackson M., Tebaldi P. Centrality measures in networks. SSRN. 2016. 42 p.
22. Lê S., Josse J., Husson F. FactoMineR: A Package for Multivariate Analysis. *Journal of Statistical Software*. 2008. vol. 25. no. 1. pp. 1–18.
23. Depaolini M., Ciucci D., Calegari S., Dominoni M. External Indices for Rough Clustering. *Lecture Notes in Computer Science*. 2018. vol. 11103. pp. 378–391.
24. White S., Smyth P. Algorithms for estimating relative importance in networks. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003. pp. 266–275.

М.А. ПЕРЕГУДОВ, А.С. СТЕШКОВОЙ  
**МОДЕЛЬ ЦЕНТРАЛИЗОВАННО-ЗАРЕЗЕРВИРОВАННОГО  
ДОСТУПА К СРЕДЕ В СЕТЯХ ЦИФРОВОЙ РАДИОСВЯЗИ**

*Перегудов М.А., Стешковой А.С.* **Модель централизованно-зарезервированного доступа к среде в сетях цифровой радиосвязи.**

**Аннотация.** Централизованно-зарезервированный доступ к среде в сетях цифровой радиосвязи семейства стандартов IEEE 802.11 является альтернативой случайному множественному доступу к среде типа CSMA/CA и в основном используется при передаче голосовых и видеосообщений в режиме реального времени. Область применения централизованно-зарезервированного доступа к среде определяет интерес к нему со стороны злоумышленников. Однако оценка эффективности централизованно-зарезервированного доступа к среде в условиях потенциально возможных деструктивных воздействий не проводилась, а потому сложно определить вклад этих воздействий в снижение эффективности такого доступа. Представлена аналитическая модель централизованно-зарезервированного доступа к среде, учитывающая не только этап его функционирования, но и этап установления в условиях деструктивных воздействий со стороны злоумышленника. Причем в модели этап установления централизованно-зарезервированного доступа к среде отображает последовательную взаимосвязь такого доступа, синхронизации элементов сетей цифровой радиосвязи и случайного множественного доступа к среде типа CSMA/CA. Установлено, что коллизии в канале передачи данных, вызванные деструктивными воздействиями, способны исключить централизованно-зарезервированный доступ к среде еще на этапе его установления. Модель применима при проектировании сетей цифровой радиосвязи семейства стандартов IEEE 802.11, оптимизации работы таких сетей и обнаружении потенциально возможных деструктивных воздействий со стороны злоумышленника.

**Ключевые слова:** централизованно-зарезервированный доступ к среде, синхронизация, случайный множественный доступ к среде, средство коммутации и управления, абонентский терминал, IEEE 802.11

**1. Введение.** Для оптимального функционирования сетей цифровой радиосвязи (СЦР) требуется оценка их эффективности не только на этапе проектирования, но и на этапе эксплуатации, особенно в условиях деструктивных воздействий (ДВ). В [1, 10-16] приведены аналитические модели функционирования процедур канального уровня СЦР, а именно синхронизации, управления мощностью, зарезервированного доступа к среде, случайного множественного доступа к среде (СМДС) типа ALOHA, S-ALOHA, CSMA/CA и CSMA/CD. В [2-6] рассмотрены модели физического уровня СЦР, а в [7-9] представлены алгоритмические модели взаимодействия ее элементов.

Наиболее доступными для злоумышленника являются СЦР семейства стандартов IEEE 802.11, на уровне доступа к среде которых сосредоточены самые опасные уязвимости [13, 16-22]. В работах [13, 16] представлены ДВ на уровне синхронизации и случайного множественного доступа к среде. В [17-19] рассмотрены деструктив-

ные воздействия на уровне аутентификации и ассоциации, а в [20, 21] – методы реализации таких воздействий. В [22] предложен алгоритм обнаружения ДВ.

К основным процедурам уровня доступа к среде сетей цифровой радиосвязи семейства стандартов IEEE 802.11 относятся аутентификация, ассоциация, синхронизация, СМДС типа CSMA/CA и централизованно-зарезервированный доступ к среде (ЦЗДС) [23-26]. При этом ЦЗДС является альтернативой СМДС типа CSMA/CA и применяется при передаче голосовых и видеосообщений в режиме реального времени.

Сегодня в СЦР семейства стандартов IEEE 802.11 в условиях ДВ эффективность оценивается только для синхронизации [16] и СМДС типа CSMA/CA [13]. Известны математические модели централизованно-зарезервированного доступа к среде [14, 27-39]. Однако модель [14] не учитывает этап установления ЦЗДС, а модели [27-39] – помимо этапа установления ЦЗДС не учитывают деструктивные воздействия.

Таким образом, разработка математической модели оценки эффективности централизованно-зарезервированного доступа к среде в сетях цифровой радиосвязи семейства стандартов IEEE 802.11 с учетом этапов его установления и функционирования в условиях деструктивных воздействий, является актуальной задачей.

**2. Анализ существующих работ.** Существует ряд исследований [14, 27-39], посвященных централизованно-зарезервированному доступу к среде в сетях цифровой радиосвязи. В [14] приведена обобщенная модель ЦЗДС в условиях ДВ без учета этапа его установления. В данной модели в качестве показателя эффективности выступает вероятность обслуживания пакета данных. В [27] с использованием методов теории вероятности представлена зависимость избыточности переданных байт от размеров служебных и пользовательских пакетов данных, а также предложена модификация ЦЗДС с целью уменьшения накладных расходов на служебные данные. В [28] эффективность ЦЗДС представлена в виде задержки передачи данных, показана зависимость задержки передачи пакетов данных от количества абонентских терминалов (АТ) в сети. Зависимость мощности передатчиков АТ и средств коммутации и управления (СКУ) от ослабления сигнала связи на трассе при различных скоростях передачи данных описана в [29], а зависимость энергопотребления АТ от скорости передачи данных и их количества в СЦР – в [30, 31]. В [33] эффективность ЦЗДС представлена в виде зависимости пропускной способности канала передачи данных от количества АТ в сети и размера пакетов данных без учета вероятностных характеристик при их передаче. Зависимость времени резервирования канала передачи данных от момента наступления ЦЗДС, а также зависимость количества первично переданных паке-

тов данных, вступивших в коллизию, и повторно переданных пакетов данных от количества АТ в сети представлены в работах [34-36]. В [37, 38] рассмотрены модели модифицированного ЦЗДС без показателя эффективности существующего ЦЗДС, а приоритетное обслуживание пакетов данных при таком доступе рассмотрено в [39]. При этом в [27-39] результирующие зависимости не учитывают потенциально возможные ДВ, а также успешность передачи синхронизирующего пакета и количество доступных АТ, которые определяют процесс установления ЦЗДС.

В связи с этим возникает потребность в создании модели, которая бы учитывала и этап установления централизованно-зарезервированного доступа к среде, и этап его функционирования в условиях деструктивных воздействий.

**3. Описательная модель централизованно-зарезервированного доступа к среде в сетях цифровой радиосвязи семейства стандартов IEEE 802.11.** В результате анализа описательной модели централизованной синхронизации элементов СЦР [16], потенциально возможных ДВ со стороны злоумышленника [17-22], спецификаций семейства стандартов IEEE 802.11 [24-26] и алгоритмической модели ЦЗДС [23] разработана структурная схема такого доступа (рис. 1). Она включает SKU, АТ и злоумышленника. В качестве SKU в СЦР семейства стандартов IEEE 802.11 может выступать ведущее устройство радиомоста или точка доступа, а в качестве АТ – ведомое устройство радиомоста.

В результате анализа спецификаций семейства стандартов IEEE 802.11 [24-26] выявлено, что ЦЗДС включает в себя этапы его установления и функционирования.

При установлении централизованно-зарезервированного доступа к среде средство коммутации и управления сравнивает количество абонентских терминалов в сетях цифровой радиосвязи семейства стандартов IEEE 802.11 с максимальным количеством таких терминалов, которые способен обслужить такой доступ. Затем SKU включает информацию о начале и длительности интервала ЦЗДС в содержание синхронизирующего пакета *Beacon*, который широковещательно рассылается каждые 100 мс. Получив такой пакет *Beacon*, АТ резервирует среду в соответствии с картой ЦЗДС, представленной на рисунке 2. Однако до ЦЗДС пакет *Beacon*, как и все пакеты данных в СЦР семейства стандартов IEEE 802.11, является участником случайного множественного доступа к среде типа CSMA/CA и, следовательно, может оказаться в коллизии (наложении пакетов). При этом АТ не получают пакет *Beacon*, и ЦЗДС не установится. В ходе СМДС типа CSMA/CA все  $N$  устройств СЦР семейства стандартов IEEE 802.11 осуществляют передачи пакетов данных с вероятностью  $p$ .

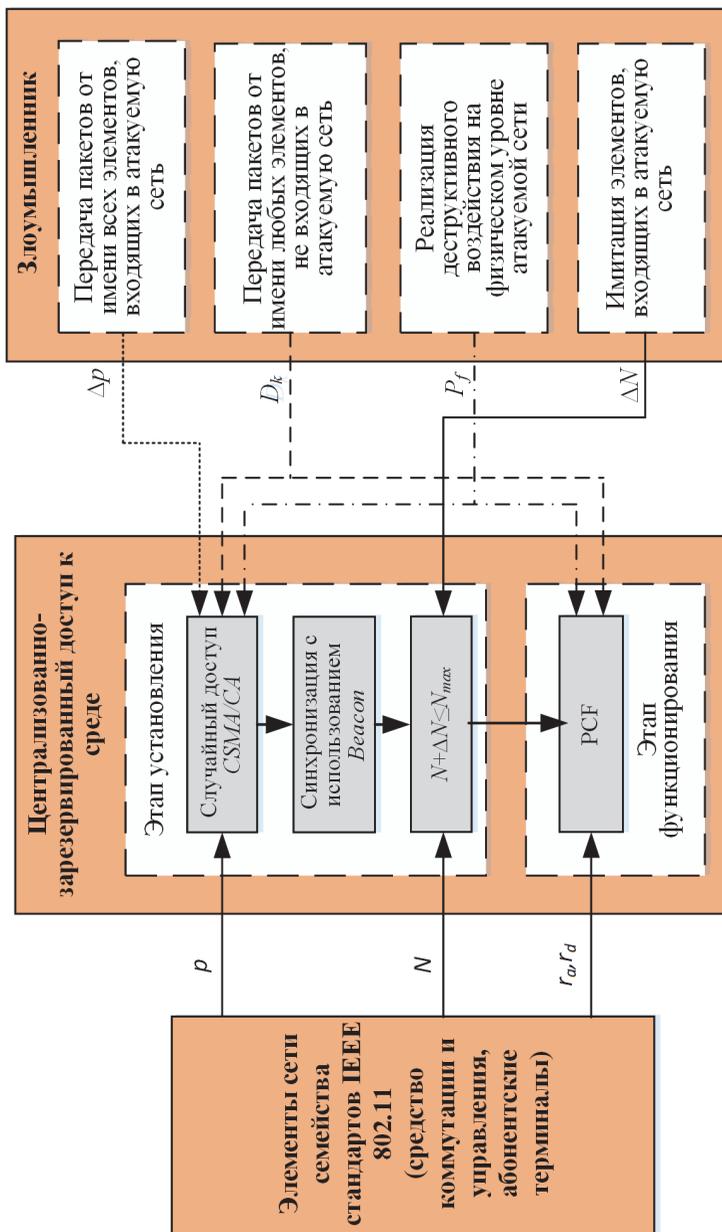


Рис. 1. Структурная схема ЦЗДС в СДР семейства стандартов IEEE 802

При функционировании ЦЗДС средство коммутации и управления в соответствии с точечной координационной функцией PCF [24-26] осуществляет последовательный опрос всех АТ сети на предмет определения наличия данных для передачи, причем для уменьшения избыточности данное средство передает объединенный пакет пользовательских и служебных данных с вероятностью  $r_a$ . При этом в качестве служебных данных выступают данные опроса и подтверждения получения пользовательских данных. При получении пакета с опросом АТ отвечает пакетом пользовательских данных с подтверждением с вероятностью  $r_d$  спустя уменьшенный межпакетный интервал *SIFS*. Если для передачи отсутствуют пользовательские данные, то АТ игнорирует пакет с опросом. Затем спустя межпакетный интервал *PIFS* СКУ продолжает опрос остальных АТ.



Рис. 2. Карта ЦЗДС в СЦР семейства стандартов IEEE 802.11

На рисунке 3 представлено информационное взаимодействие абонентских терминалов и средства коммутации и управления при функционировании централизованно-зарезервированного доступа к среде в сетях цифровой радиосвязи семейства стандартов IEEE 802.11.

Пакет пользовательских данных *Data* может содержать в своих полях информацию о подтверждении АСК (пакет подтверждения принятых данных) и опросе СКУ абонентских терминалов. Также пакет опроса АТ *Poll* может содержать в своих полях информацию о подтверждении АСК.

Из анализа работ [1-6, 10-16] следует, что злоумышленник в целях блокирования ЦЗДС на этапах его установления и функционирования может осуществлять следующие деструктивные воздействия:

- передачу пакетов данных от имени всех  $N$  легитимных АТ, которые входят в атакуемую сеть, с вероятностью  $\Delta p$ ;
- передачу пакетов данных от имени любых  $K$  элементов, не входящих в атакуемую сеть, с вероятностью  $D_k$ ;
- имитацию  $\Delta N$  легитимных АТ, входящих в атакуемую сеть;

– формирование помехи на физическом уровне атакуемой сети с вероятностью  $P_f$ .

В настоящей работе рассматриваются только деструктивные воздействия на физическом и канальном уровнях эталонной модели взаимодействия открытых систем и не рассматриваются такие воздействия на сетевом и вышестоящих уровнях.

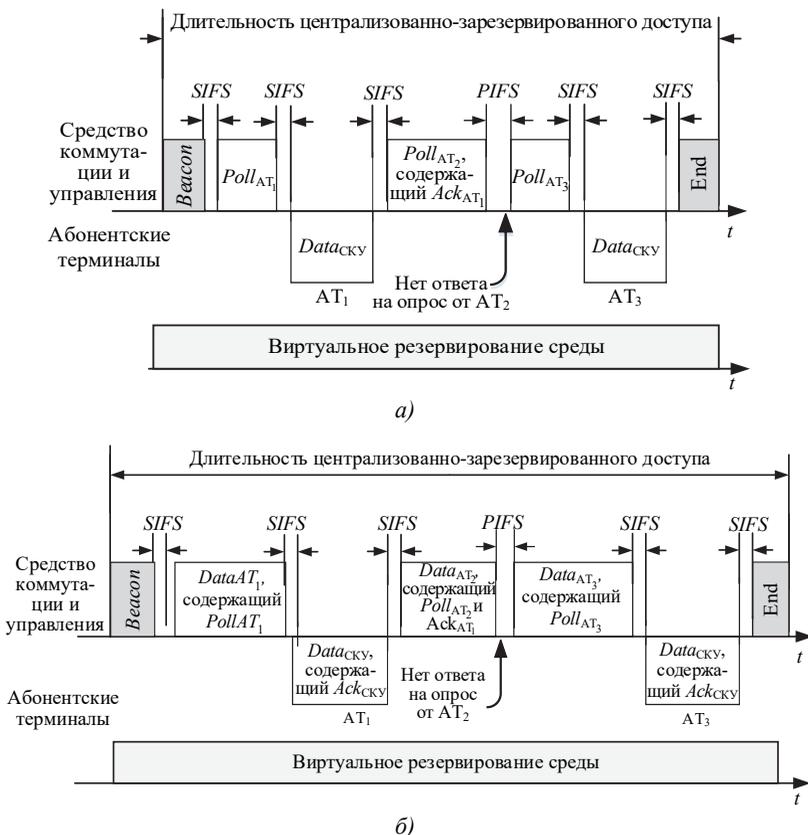


Рис. 3. Информационное взаимодействие AT и SKU: а) при отсутствии данных для передачи у SKU; б) при наличии данных для передачи

**4. Аналитическая модель централизованно-зарезервированного доступа к среде в сетях цифровой радиосвязи семейства стандартов IEEE 802.11.** Из анализа описательной модели ЦЗДС в СЦР семейства стандартов IEEE 802.11 следует, что события успешной передачи синхронизирующего пакета *Beacon*, проверки

ограничения по количеству АТ при установлении такого доступа и доставки пакетов пользовательских данных при его функционировании совместны и независимы друг от друга. Поэтому под эффективностью ЦЗДС в СЦР семейства стандартов IEEE 802.11 будем понимать вероятность успешной передачи пакетов пользовательских данных в течение средней длительности интервала такого доступа при его успешном установлении. Данную вероятность предлагается определять следующим выражением:

$$\Omega_{cam} = \Omega_{con} \cdot \Omega_{syn} \cdot \Omega_{sea}, \quad (1)$$

где  $\Omega_{con}$  – вероятность выполнения ограничения по количеству АТ, участвующих в ЦЗДС;  $\Omega_{syn}$  – вероятность успешной передачи пакета *Beacon*;  $\Omega_{sea}$  – вероятность успешной передачи пакетов пользовательских данных при функционировании такого доступа.

#### 4.1. Ограничение по количеству абонентских терминалов сетей цифровой радиосвязи семейства стандартов IEEE 802.11, участвующих в централизованно-зарезервированном доступе к среде.

Коэффициент выполнения ограничения по количеству АТ, участвующих в ЦЗДС, будем определять по формуле:

$$\Omega_{con} = \begin{cases} 1, & \text{если } N + \Delta N \leq N_{max}; \\ 0, & \text{если } N + \Delta N > N_{max}, \end{cases} \quad (2)$$

где  $N$  – количество АТ в СЦР семейства стандартов IEEE 802.11;  $\Delta N$  – количество симитированных злоумышленником легитимных АТ, входящих в атакуемую сеть;  $N_{max}$  – максимально возможное количество АТ, которое способно опросить СКУ за один интервал ЦЗДС.

Коэффициент выполнения ограничения по количеству АТ, участвующих в ЦЗДС, учитывает ДВ, направленное на имитацию легитимных АТ, которые входят в атакуемую сеть. При реализации такого воздействия еще на этапе установления ЦЗДС до передачи синхронизирующего пакета такой доступ будет прекращен.

Максимальное возможное количество АТ, способных участвовать в ЦЗДС, определяется в соответствии с [27] выражением:

$$N_{max} = \frac{T_{cam} - (T_{Beacon} + T_{SIFS}) - T_{CFend}}{T_{vr}}, \quad (3)$$

где  $T_{cam}$  – длительность интервала ЦЗДС;  $T_{Beacon}$  – длительность передачи пакета *Beacon*;  $T_{SIFS}$  – длительность межпакетного интервала *SIFS*;

$T_{CFend}$  – длительность передачи пакета, сигнализирующего об окончании ЦЗДС  $CFend$ ;  $\overline{T}_{vr}$  – среднее время опроса абонентского терминала СКУ.

**4.2. Синхронизация абонентских терминалов сетей цифровой радиосвязи семейства стандартов IEEE 802.11, участвующих в централизованно-зарезервированном доступе к среде.** Согласно [16] вероятность успешной передачи пакета *Beacon* имеет следующий вид:

$$\Omega_{syn} = \frac{\left(1 - k_a \frac{\tau}{2(\overline{T}_m + T_{DIFS})}\right) \left(T_{PIFS} + k_b \frac{(\overline{T}_m + T_{PIFS} - \tau)^2}{2(\overline{T}_m + T_{DIFS})}\right)}{T_{PIFS} + k_a \frac{(T_{TBTT} + \overline{T}_m - \tau)\tau}{2(\overline{T}_m + T_{DIFS})} + k_b \frac{(\overline{T}_m + T_{PIFS} - \tau)^2}{2(\overline{T}_m + T_{DIFS})}}, \quad (4)$$

где  $\overline{T}_m$  – средняя длительность передачи пакета данных АТ;  $T_{DIFS}$  – длительность межпакетного интервала *DIFS*;  $T_{PIFS}$  – длительность межпакетного интервала *PIFS*;  $T_{TBTT}$  – длительность повторяющегося интервала синхронизации *TBTT*;  $k_a$  – коэффициент создания коллизии синхронизирующего пакета *Beacon*;  $k_b$  – коэффициент занятости канала передачи данных;  $\tau$  – длительность минимального временного интервала, из которого состоят межпакетные интервалы и пакеты данных.

Коэффициент создания коллизии пакета *Beacon* определяется по формуле [16]:

$$k_a = \frac{P_{vrN-1} \tau}{P_{frN-1} \tau + P_{vrN-1} (\overline{T}_m + T_{DIFS} - \tau)}, \quad (5)$$

а коэффициент занятости канала передачи данных – выражением [16]:

$$k_b = 1 - \frac{(1 - P_{vrN-1}) \tau}{P_{scN-1} T_{sc} + P_{clN-1} T_{cl} + P_{frN-1} \tau}, \quad (6)$$

где  $P_{vrN-1}$  – вероятность занятости канала передачи данных одним из  $N-1$  элементов СЦР семейства стандартов IEEE 802.11;  $P_{frN-1}$  – вероятность свободного канала передачи данных при  $N-1$  элементах сети;  $P_{scN-1}$  – вероятность успешной передачи пакета данных одним из  $N-1$  элементов сети;  $P_{clN-1}$  – вероятность создания коллизии  $N-1$  эле-

ментом сети;  $T_{sc}$  – средняя длительность успешной передачи пакета данных;  $T_{cl}$  – средняя длительность коллизии.

Вероятности  $P_{fr_{N-1}}$ ,  $P_{sc_{N-1}}$ ,  $P_{cl_{N-1}}$ ,  $P_{r_{N-1}}$  [16] имеют вид:

$$\begin{aligned}
 P_{fr_{N-1}} &= (1 - (p + \Delta p))^{N-1} (1 - P_f) \prod_{k=1}^K (1 - D_k); \\
 P_{sc_{N-1}} &= (N-1)p(1 - (p + \Delta p))^{N-2} (1 - P_f) \prod_{k=0}^K (1 - D_k); \\
 P_{cl_{N-1}} &= 1 - P_{fr_{N-1}} - P_{sc_{N-1}}; \\
 P_{r_{N-1}} &= 1 - P_{fr_{N-1}}; \\
 0 &< p + \Delta p \leq 1,
 \end{aligned}
 \tag{7}$$

где  $p$  – вероятность передачи АТ или СКУ пакета данных при СМДС типа CSMA/CA;  $\Delta p$  – вероятность передачи злоумышленником пакетов данных от имени  $N$  легитимных элементов, входящих в атакуемую сеть;  $D_k$  – вероятность передачи злоумышленником пакетов данных от имени  $K$  элементов, не входящих в атакуемую сеть;  $P_f$  – вероятность формирования помехи на физическом уровне атакуемой сети.

Вероятность  $p$  определяется путем решения системы уравнений, отражающей особенности СМДС типа CSMA/CA [13, 16]:

$$\left\{ \begin{aligned}
 p &= \frac{2(1 - P_{r_N})}{W_0(1 - P_{r_{N-1}}) \sum_{i=0}^{m-1} (2P_{r_{N-1}})^i + W_0(2P_{r_{N-1}})^m + 1}; \\
 P_{r_{N-1}} &= 1 - \left[ (1 - (p + \Delta p))^{N-1} (1 - P_f) \prod_{k=1}^K (1 - D_k) \right]; \\
 P_{r_N} &= 1 - \left[ (1 - (p + \Delta p))^N (1 - P_f) \prod_{k=1}^K (1 - D_k) \right],
 \end{aligned} \right.
 \tag{8}$$

где  $W_0$  – значение счетчика отсрочки повторной передачи;  $m$  – количество повторных попыток передач.

Средние длительности передачи пакета данных  $\overline{T_m}$ , создания коллизии  $T_{cl}$  и успешной передачи такого пакета  $T_{sc}$  в соответствии с [13, 16] для основного механизма СМДС типа CSMA/CA определяются следующим образом:

$$\begin{cases}
 \left\{ \begin{array}{l}
 \overline{T}_m = \overline{T}_{data} + \sigma + T_{SIFS} + T_{Ack} + \sigma; \\
 T_{cl} = \overline{T}_{data} + T_{DIFS} + \sigma;
 \end{array} \right. , \\
 \text{если } (E[P_z] \leq \overline{T}_{data}); \\
 \left\{ \begin{array}{l}
 \overline{T}_m = E[P_z] + \sigma + T_{SIFS} + T_{Ack} + \sigma; \\
 T_{cl} = E[P_z] + T_{DIFS} + \sigma;
 \end{array} \right. , \\
 \text{если } (E[P_z] > \overline{T}_{data}); \\
 T_{sc} = \overline{T}_{data} + \sigma + T_{SIFS} + T_{Ack} + \sigma + T_{DIFS},
 \end{cases} \quad (9)$$

для дополнительного механизма CSMA/CA:

$$\begin{cases}
 \left\{ \begin{array}{l}
 \overline{T}_m = T_{Rts} + \sigma + T_{SIFS} + T_{Cts} + \sigma + T_{SIFS} + \\
 + \overline{T}_{data} + \sigma + T_{SIFS} + T_{Ack} + \sigma; \\
 T_{cl} = T_{Rts} + T_{DIFS} + \sigma;
 \end{array} \right. , \\
 \text{если } (E[P_z] \leq \overline{T}_{data}); \\
 \left\{ \begin{array}{l}
 \overline{T}_m = T_{Rts} + \sigma + T_{SIFS} + T_{Cts} + \sigma + T_{SIFS} + \\
 + E[P_z] + \sigma + T_{SIFS} + T_{Ack} + \sigma; \\
 T_{cl} = E[P_z] + T_{DIFS} + \sigma;
 \end{array} \right. , \\
 \text{если } (E[P_z] > \overline{T}_{data}); \\
 T_{sc} = T_{Rts} + \sigma + T_{SIFS} + T_{Cts} + \sigma + T_{SIFS} + \\
 + \overline{T}_{data} + \sigma + T_{SIFS} + T_{Ack} + \sigma + T_{DIFS},
 \end{cases} \quad (10)$$

где  $\overline{T}_{data}$  – средняя длительность передачи пакета пользовательских данных *Data*;  $T_{Ack}$  – длительность пакета *Ack* подтверждения успешной доставки пакета пользовательских данных *Data*;  $T_{Rts}$  – длительность пакета *Rts* запроса на получение доступа к среде;  $T_{Cts}$  – длительность пакета *Cts* подтверждения успешной доставки пакета *Rts*;  $\sigma$  – задержка распространения сигнала;  $E[P_z]$  – средняя длительность передачи средством злоумышленника.

**4.3. Передача пакетов пользовательских данных при функционировании централизованно-зарезервированного доступа к среде в сетях цифровой радиосвязи семейства стандартов IEEE 802.11.** На основании описания этапа функционирования ЦЗДС

и учета ДВ представим вероятность успешной передачи пакетов пользовательских данных при функционировании такого доступа как отношение среднего времени передачи пакетов пользовательских данных в условиях ДВ к максимальному среднему времени такой передачи в рамках одного цикла опроса СКУ абонентского терминала:

$$\Omega_{sea} = \frac{\overline{T_{S_{vr}}}(1 - P_{z_{tr}})}{\overline{T_{max_{vr}}}(1 - P_{z_{tr}}) + P_{z_{tr}}E[P_z]}, \quad (11)$$

где  $\overline{T_{S_{vr}}}$  – среднее время передачи пакетов пользовательских данных в рамках одного цикла опроса СКУ абонентского терминала;  $\overline{T_{max_{vr}}}$  – максимальное среднее время передачи;  $P_{z_{tr}}$  – вероятность передачи средством злоумышленника;  $E[P_z]$  – средняя длительность передачи средством злоумышленника.

Среднее время передачи пакетов пользовательских данных  $\overline{T_{S_{vr}}}$  составляет полезную часть среднего времени передачи всех пакетов данных  $\overline{T_{vr}}$  в рамках одного цикла опроса СКУ абонентского терминала. При этом с учетом вероятностей передач пакетов данных АТ и СКУ  $\overline{T_{vr}}$  в соответствии с [27] примет вид:

$$\begin{aligned} \overline{T_{vr}} &= T_{SIFS} + T_{poll}(1 - r_a) + \overline{T_{data}}r_a + \overline{T_{data}}r_d + T_{PIFS}(1 - r_d) + T_{SIFS}r_d = \\ &= T_{SIFS} + T_{SIFS}r_d + T_{poll}(1 - r_a) + \overline{T_{data}}(r_a + r_d) + T_{PIFS}(1 - r_d), \end{aligned} \quad (12)$$

где  $T_{poll}$  – длительность передачи пакета опроса СКУ;  $r_d$  – вероятность передачи пакета пользовательских данных АТ при ЦЗДС;  $r_a$  – вероятность передачи пакета данных с опросом СКУ;  $\overline{T_{data}}$  – средняя длительность передачи пакета пользовательских данных  $Data$ .

В (12) одним из слагаемых является среднее время передачи пакетов пользовательских данных в рамках одного цикла опроса СКУ абонентского терминала:

$$\overline{T_{S_{vr}}} = \overline{T_{data}}(r_a + r_d). \quad (13)$$

Вероятности  $r_a$  и  $r_d$  представляют собой показатели загруженности сети и варьируются в пределах от 0 до 1. Значение 1 вероятности  $r_a$  достигается только при передаче СКУ потоковых мультимедийных данных, а значение 0 – только при приеме данных. Аналогично и для  $r_d$ .

Для  $r_a = 1$  и  $r_d = 1$  из (12) получим максимальное среднее время передачи в рамках одного цикла опроса СКУ абонентского терминала:

$$\overline{Tmax_{vr}} = 2\overline{T_{data}} + 2T_{SIFS}. \quad (14)$$

Средняя длительность передачи пакета пользовательских данных  $Data$  определяется исходя из скорости передачи данных в канале связи и размера полезной нагрузки [40]:

$$\overline{T_{data}} = T_{preamble} + T_{signalExtension} + \left( \frac{22 + (L_{header} + \overline{L_{data}})8}{R} \right), \quad (15)$$

где  $T_{preamble}$  – длительность преамбулы;  $T_{signalExtension}$  – длительность поля расширения сигнала;  $L_{header}$  – объем заголовка полезной нагрузки;  $\overline{L_{data}}$  – средний размер полезной нагрузки;  $R$  – скорость передачи данных.

Средний размер полезной нагрузки пакета пользовательских данных  $Data$  [40] имеет вид:

$$\overline{L_{data}} = \frac{\sum_{i=1}^N L_{data_i}}{N}, \quad (16)$$

где  $L_{data_i}$  – объем полезной нагрузки пакета пользовательских данных  $Data$ , передаваемой  $i$ -ым элементом СЦР семейства стандартов IEEE 802.11.

Вероятность передачи средством злоумышленника, учитывающая потенциально возможные ДВ, определяется:

$$Pz_{tr} = 1 - (1 - P_f) \prod_{k=1}^K (1 - D_k). \quad (17)$$

На основании рассмотренных показателей в аналитической модели предложена система показателей эффективности ЦЗДС в СЦР семейства стандартов IEEE 802.11, которая представлена на рисунке 4.

Приведенная аналитическая модель с ее функциями, параметрами и характеристиками соответствует реальному процессу ЦЗДС в СЦР семейства стандартов IEEE 802.11, описанному в спецификациях [24-26].



Рис. 4. Система показателей эффективности ЦЗДС в СЦР семейства стандартов IEEE 802.11

Из приведенной на рисунке 4 системы показателей видно, что известный показатель – вероятность успешной передачи пакета *Beacon*, оценивающий эффективность синхронизации элементов СЦР,

является новым частным показателем эффективности ЦЗДС, а также в качестве исходных данных для оценки эффективности такого доступа применяются новые показатели – количество симитированных злоумышленником легитимных АТ, вероятности передачи пакетов пользовательских данных АТ и СКУ. Также из системы показателей видно, что ЦЗДС зависит от централизованной синхронизации элементов СЦР, которая, в свою очередь, зависит от СМДС типа CSMA/CA.

**5. Результаты исследования.** Рассмотрена СЦР семейства стандартов IEEE 802.11n, основные характеристики которой приведены в таблице 1.

Таблица 1. Основные характеристики СЦР семейства стандартов IEEE 802.11n

Параметр	Значение
Длительность интервала ЦЗДС $T_{cam}$ , мкс	32000
Длительность межпакетного интервала $DIFS$ , мкс	28
Длительность межпакетного интервала $PIFS$ $T_{PIFS}$ , мкс	19
Длительность межпакетного интервала $SIFS$ $T_{SIFS}$ , мкс	10
Длительность повторяющегося интервала синхронизации $T_{BTT}$ $T_{BTT}$ , мкс	$10^5$
Длительность минимального временного интервала $\tau$ , мкс	9
Задержка распространения сигнала $\sigma$ , мкс	1

Данные основные характеристики справедливы для СЦР стандартов IEEE 802.11b/g/n [41]. Сети цифровой радиосвязи стандартов IEEE 802.11a/ac отличаются длительностью межпакетных интервалов и минимальным временным интервалом. Однако при теоретических исследованиях вне зависимости от значений основных характеристик СЦР семейства стандартов IEEE 802.11 поведение рассматриваемых зависимостей не меняется.

Результаты теоретических и экспериментальных исследований эффективности ЦЗДС представлены на рисунках 5-8.

Из анализа результатов теоретических исследований, приведенных на рисунках 5-8, следует:

1. С увеличением вероятностей передачи пользовательских данных АТ и СКУ возрастает эффективность ЦЗДС в СЦР семейства стандартов IEEE 802.11 (рис. 5 и 6). Она стремится к эффективности централизованной синхронизации элементов аналогичных сетей, которая отражает качество установления ЦЗДС. При этом эффективность централизованной синхронизации элементов СЦР семейства стандартов IEEE 802.11 не зависит от вероятностей передачи пользовательских данных АТ и СКУ.

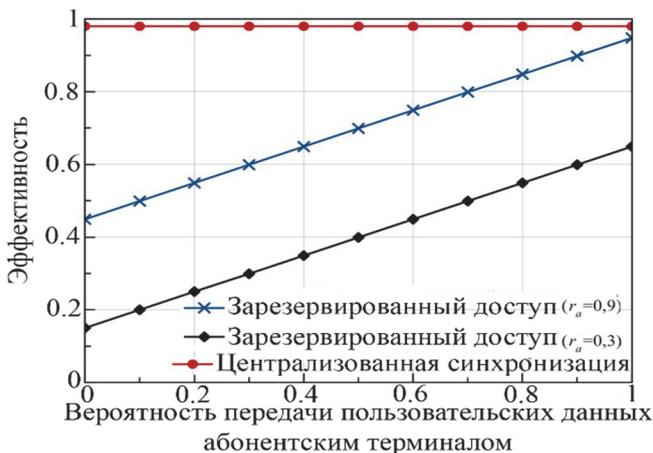


Рис. 5. Зависимости эффективностей ЦЗД и централизованной синхронизации от вероятности передачи пользовательских данных

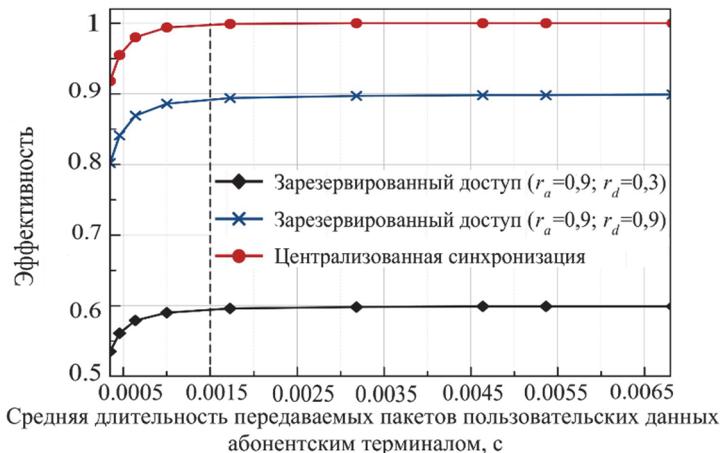


Рис. 6. Зависимости эффективностей ЦЗД и централизованной синхронизации от средней длительности передаваемых пакетов пользовательских данных

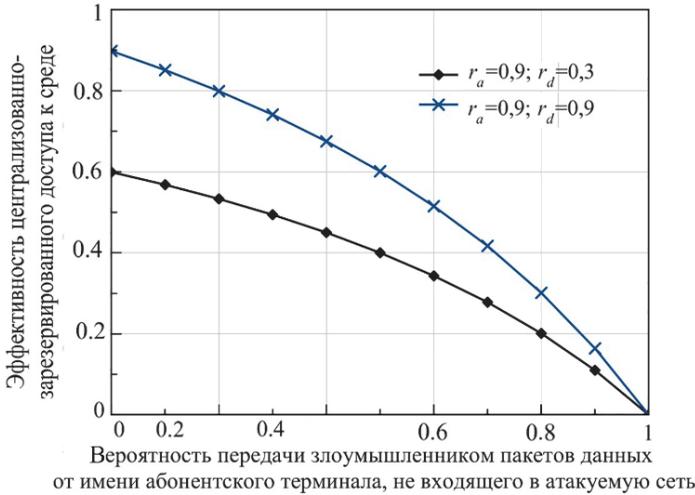


Рис. 7. Зависимости эффективности ЦЗДС от вероятности передачи злоумышленником пакетов данных от имени АТ, не входящего в атакуемую сеть

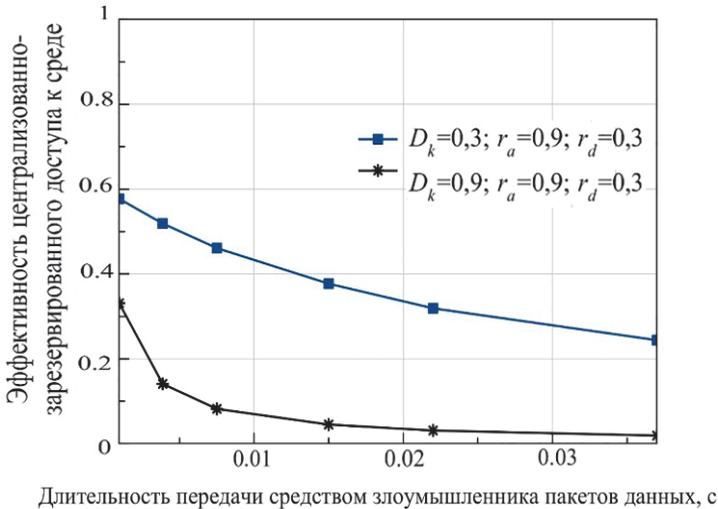


Рис. 8. Зависимости эффективности ЦЗДС от длительности передачи средством злоумышленника пакетов данных

2. При длительностях передаваемых пакетов пользовательских данных более 1,5 мс эффективность ЦЗДС в СЦР семейства стандартов IEEE 802.11 – константа (рис. 6). С уменьшением длительности передаваемых пакетов пользовательских данных от отметки в 1,5 мс эффективность ЦЗДС также уменьшается. Это обстоятельство связано с ростом коллизий (столкновений пакетов) в канале передачи данных, которые влияют на установление ЦЗДС.

3. Централизованно-зарезервированный доступ к среде в СЦР семейства стандартов IEEE 802.11 не устойчив к ДВ со стороны злоумышленника (рис. 7 и 8). Однако только при непрерывной передаче средством злоумышленника пакетов данных эффективность ЦЗДС стремится к нулю. Такое поведение злоумышленника является его демаскирующим признаком.

4. При увеличении длительности передачи средством злоумышленника пакетов данных эффективность ЦЗДС в СЦР семейства стандартов IEEE 802.11 уменьшается (рис. 8). Причем эффективность уменьшится в 10 раз с изменением длительности с 1 на 20 мс при вероятности передачи средством злоумышленника, равным 0,9.

Полученные теоретические результаты подтверждают правильность описания реального процесса ЦЗДС в СЦР семейства стандартов IEEE 802.11, что свидетельствует об адекватности разработанной модели. Причем в [16] с использованием аналитических выражений (4) – (10), реализованных в программном комплексе оптимизации работы СЦР [42], проведена экспериментальная оценка эффективности централизованной синхронизации элементов СЦР семейства стандартов IEEE 802.11, которая отвечает за установление ЦЗДС в таких сетях. При этом результаты экспериментального исследования эффективности централизованной синхронизации элементов СЦР семейства стандартов IEEE 802.11 практически совпадают с результатами теоретического исследования и отличаются на сотые доли числа.

**6. Заключение.** Разработана аналитическая модель централизованно-зарезервированного доступа к среде в сетях цифровой радиосвязи семейства стандартов IEEE 802.11, основанная на применении теорий вероятностей и массового обслуживания. Новизна предложенной модели состоит в учете этапа установления централизованно-зарезервированного доступа к среде, потенциально возможных деструктивных воздействий со стороны злоумышленника и комплексной оценке эффективности такого доступа. В модели на этапе установления отражена связь случайного множественного доступа к среде типа CSMA/CA с централизованной синхронизацией элементов сетей цифровой радиосвязи, а также связь такой синхронизации с централизо-

ванно-зарезервированным доступом к среде. При этом случайный множественный доступ к среде типа CSMA/CA не зависит от централизованно-зарезервированного доступа к среде. Комплексная оценка эффективности централизованно-зарезервированного доступа к среде учитывает как этап его функционирования, так и этап установления. Определено, что коллизии в канале передачи данных способны исключить централизованно-зарезервированный доступ к среде еще на этапе его установления. Модель применима при проектировании сетей цифровой радиосвязи семейства стандартов IEEE 802.11, оптимизации работы таких сетей и обнаружении потенциально возможных деструктивных воздействий со стороны злоумышленника.

Использование разработанной модели централизованно-зарезервированного доступа к среде в сетях цифровой радиосвязи семейства стандартов IEEE 802.11 в сравнении с известными моделями позволяет комплексно количественно оценить эффективность как на этапе его функционирования, так и на этапе установления. Также ее применение позволяет оценивать вклад деструктивных воздействий в снижение эффективности ЦЗДС. Централизованно-зарезервированный доступ к среде является одной из основных процедур канального уровня сетей цифровой радиосвязи семейства стандартов IEEE 802.11, поэтому полученный результат приближает к количественной оценке эффективности функционирования канального уровня СЦР семейства стандартов IEEE 802.11 в целом.

Представленная модель может быть использована при разработке методики оценки эффективности ЦЗДС в СЦР семейства стандартов IEEE 802.11, программную реализацию которой целесообразно включить в состав программных комплексов оптимизации [42] и диагностирования [43] СЦР. При этом оптимизация таких сетей направлена на поддержание в режиме реального времени требуемой эффективности функционирования СЦР путем реализации в такой сети для ее абонентов управляющих воздействий, изменяющих значения характеристик сети. К таким характеристикам СЦР также относятся исходные данные системы показателей эффективности ЦЗДС, приведенные на рисунке 4.

Модель применима в системах проектирования сетей цифровой радиосвязи, в которых для требуемой эффективности функционирования таких сетей рассчитываются значения их характеристик. Также модель может использоваться в системах обнаружения потенциально возможных деструктивных воздействий со стороны злоумышленника за счет определения резкого снижения эффективности ЦЗДС при известных характеристиках СЦР.

В рамках рассматриваемого направления научных исследований дальнейшей проработки требует модель канального уровня сетей цифровой радиосвязи семейства стандартов IEEE 802.11.

### Литература

1. *Макаренко С.И.* Подавление пакетных радиосетей со случайным множественным доступом за счет дестабилизации их состояния // Журнал радиоэлектроники. 2011. № 9. С. 2. URL: [www.jre.cplire.ru/jre/sep11/4/text.pdf](http://www.jre.cplire.ru/jre/sep11/4/text.pdf) (дата обращения: 03.06.2020).
2. *Титов К.Д., Завалишина О.Н.* Оценка помехоустойчивости системы связи стандарта IEEE 802.11ac при воздействии помех // Успехи современной радиоэлектроники. 2019. № 12. С. 191–196.
3. *Deniau V et al.* IEEE 802.11n Communications in the Presence of FrequencySweeping Interference Signals // IEEE Transactions on Electromagnetic Compatibility. 2017. vol. 59. no. 5. pp. 1625–1633.
4. *Scalia L., Tinnirello I., Giustiniano D.* Side effects of ambient noise immunity techniques on outdoor IEEE 802.11 deployments // GLOBECOM. Proceedings of the Global Telecommunications Conference. 2008. pp. 1–6.
5. *Титов К.Д., Липатов А.О., Завалишина О.Н.* Оценка помехоустойчивости системы связи стандарта IEEE 802.11n при воздействии помех с учётом структуры пакета передаваемых данных // Теория и техника радиосвязи. 2019. № 4. С. 95–107.
6. *Макаренко С.И.* Динамическая модель системы связи в условиях функционально-разноразового информационного конфликта наблюдения и подавления // Системы управления, связи и безопасность. 2015. № 3. С. 122–185.
7. *Аганесов А.В., Макаренко С.И.* Модель воздушно-космической сети связи с иерархическим принципом ретрансляции информационных потоков // Радиотехнические и телекоммуникационные системы. 2015. № 4. С. 43–51.
8. *Бойко А.А.* Способ аналитического моделирования процесса распространения вирусов в компьютерных сетях различной структуры // Труды СПИИРАН. 2015. Вып. 5. С. 196–211.
9. *Бойко А.А., Обущенко Е.Ю., Щеглов А.В.* Особенности синтеза полного множества тестовых способов удаленного информационно-технического воздействия на пространственно распределенные системы информационно-технических средств // Вестник Воронежского государственного университета. 2017. № 2. С. 33–45.
10. *Перегудов М.А., Бойко А.А.* Модель процедуры случайного множественного доступа к среде типа S-ALOHA // Информационно-управляющие системы. 2014. № 6. С. 75–81.
11. *Перегудов М.А., Бойко А.А.* Оценка защищенности сети пакетной радиосвязи от имитации абонентских терминалов на уровне процедуры случайного множественного доступа к среде типа S-ALOHA // Информационные технологии. 2015. № 7. С. 527–534.
12. *Перегудов М.А., Семченко И.А.* Оценка эффективности случайного множественного доступа к среде типа ALOHA при голосовых соединениях, передаче служебных команд, текстовых сообщений и мультимедийных файлов в условиях деструктивных воздействий // Труды СПИИРАН. 2019. Т. 18. № 4. С. 887–911.
13. *Перегудов М.А., Стешковой А.С., Бойко А.А.* Вероятностная модель процедуры случайного множественного доступа к среде типа CSMA/CA // Труды СПИИРАН. 2018. Вып. 4(59). С. 92–114.
14. *Перегудов М.А., Бойко А.А.* Модель процедуры зарезервированного доступа к среде сети пакетной радиосвязи // Телекоммуникации. 2015. № 6. С. 7–15.

15. *Пережудов М.А., Бойко А.А.* Модель процедуры управления питанием сети пакетной радиосвязи // Телекоммуникации. 2015. № 9. С. 13–18.
16. *Пережудов М.А., Степиковой А.С.* Модель централизованной синхронизации элементов сетей цифровой радиосвязи со случайным множественным доступом к среде типа CSMA/CA // Труды СПИИРАН. 2020. Т. 19. № 1. С. 128–154.
17. *Liu C., Qiu J.* Performance study of 802.11w for preventing DoS attacks on wireless local area networks // Wireless personal communications. 2017. no. 95. pp. 1031–1053.
18. *Kaur J.* Mac Layer Management Frame Denial of Service Attacks // International Conference on Micro-Electronics and Telecommunication Engineering. 2016. pp. 155–160.
19. *Filipek J., Hudec L.* Securing mobile ad hoc networks using distributed firewall with PKI // IEEE 14th International Symposium on Applied Machine Intelligence and Informatics. 2016. pp. 321–325.
20. *Yacchirena A. et al.* Analysis of attack and protection systems in Wi-Fi wireless networks under the Linux operating system // IEEE International Conference on Automatica. 2016. pp. 1–7.
21. *Liu C., Qiu J.* Performance study of 802.11w for preventing DoS attacks on wireless local area networks // Wireless personal communications. 2017. no. 95. pp. 1031–1053.
22. *Noman H.A., Abdullah S.M., Mohammed H.I.* An Automated Approach to Detect Deauthentication and Disassociation Dos Attacks on Wireless 802.11 Networks // International Journal of Computer Science Issues. 2015. vol. 12. pp. 1694–1784.
23. *Пережудов М. А., Степиковой А. С., Щеглов А. В.* Описательная модель канального уровня сетей цифровой радиосвязи семейства стандартов IEEE 802.11 // Системы управления, связи и безопасности. 2020. № 3. С. 203–221.
24. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications // IEEE Computer Society LAN MAN Standards Committee. 1997.
25. IEEE standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements PART 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications // IEEE Std. 802.11–2012. 2012. pp. 1–2793.
26. IEEE Standards Association/IEEE Computer Society. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; Amendment 4: Enhancements for Very High Throughput for Operation in Bands Below 6 GHz // IEEE Std. 802.11–2013. pp. 1–425.
27. *Kanjanavapastit A., Landfeldt B.* An Analysis of a Modified Point Coordination Function in IEEE 802.11 // Proceedings of IEEE 14th Personal, Indoor and Mobile Radio Communications. 2003. vol. 2. pp. 1732–1736.
28. *Sikdar B.* An analytic model for the delay in IEEE 802.11 PCF MAC based wireless networks // IEEE Transactions on Wireless Communications. 2007. vol. 4. no. 6. pp. 1542–1560.
29. *Qiao D., Choi S., Soomro A., Shin G.* Energy-Efficient PCF Operation of IEEE 802.11a Wireless LAN // Proc. IEEE INFOCOM. 2002. vol. 2. pp. 580–589.
30. *Guan Z., Yang Z. J., He M.* Energy-efficient analysis of an IEEE 802.11 PCF MAC protocol based on WLAN // Journal of Ambient Intelligence & Humanized Computing. 2018. pp. 1–11.
31. *Zheng G., Zhi-Jun Y., Min H.* Energy-efficient analysis of an IEEE 802.11 PCF MAC protocol based on WLAN // Journal of Ambient Intelligence and Humanized Computing. 2018. pp. 1–11.
32. *Eyadeh, A., Jarrah, M., Aljumaili, A.* Modeling and simulation of performance limits in IEEE 802.11 point-coordination function // International Journal of Recent Technology and Engineering. 2019. vol. 8(4). pp 5575–5580.

33. *Noman H.M.* PCF and DCF Performances Evaluation for a Non Transition 802.11 Wireless Network using OPNET Modular // International Journal of Soft Computing and Engineering. 2017. vol. 7. pp. 2231–2307.
34. *Sarmah S., Sharma S.K.* Performance Analysis of IEEE 802.11 WLANs by varying PCF, DCF and EDCF to Enhance Quality of service // International Journal of Computer Applications. 2016. pp. 138.
35. *Dhaliwal A.S.* Analyzing the Impact of DCF and PCF on WLAN Network Standards 802.11a, 802.11b and 802.11g // Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. 2013. no. 7. pp. 1594–1598.
36. *Chen D., Garg S., Kappes M., Trivedi K.* Supporting VBR VoIP traffic in IEEE 802.11 WLAN in PCF mode // Avaya Laboratories. 2002. vol. 26. 538 p.
37. *Vishnevsky V., Lyakhov A.* Analytical Study of IEEE 802.11 PCF for regional and Metropolitan Area Networks // Cybernetics and Information Technologies. 2005. vol. 5. no. 2. pp. 117–136.
38. *Liu Q., Zhao D., Zhou D.* An analytic model for enhancing IEEE 802.11 point coordination function media access control protocol // European transactions on telecommunications. 2011. vol. 22. pp. 332–338.
39. *Kaur I., Bala M., Bajaj H.* Performance evaluation of wlan by varying PCF, DCF and enhanced DCF slots to improve quality of service // IOSR Journal of Computer Engineering. 2012. vol. 2. pp. 29–33.
40. *Shigeo S., Daiki T.* Bistable Behavior of IEEE 802.11 Distributed Coordination Function // 22nd International Symposium on Wireless Personal Multimedia Communications. 2019. pp. 1–6.
41. *Burton M.* 802.11 Arbitration // Certified Wireless Network Professional Inc. Durham. 2009. 24 p.
42. Свидетельство о государственной регистрации программы для ЭВМ 2018614894 Российская Федерация. Программный комплекс оптимизации работы сетей радиосвязи; правообладатели и авторы А.А. Бойко, М.А. Перегудов, И.А. Семченко, А.С. Стешковой. – № 2018612052; заявл. 05.03.2018; опубл. 19.04.2018.
43. Свидетельство о государственной регистрации программы для ЭВМ 2019665751 Российская Федерация. Программный комплекс диагностирования сетей цифровой радиосвязи; правообладатели и авторы М.А. Перегудов, И.С. Дегтярев, А.Я. Уманский, И.А. Семченко, А.С. Стешковой, А.В. Щеглов. – № 2019664891; заявл. 21.11.2019; опубл. 28.11.2019.

**Перегудов Максим Анатольевич** — канд. техн. наук, начальник, заместитель начальника, научно-исследовательская лаборатория, научно-исследовательский отдел, Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (ВУНЦ ВВС «ВВА»). Область научных интересов: методы и системы защиты информации. Число научных публикаций — 15. [maharegudov@mail.ru](mailto:maharegudov@mail.ru); ул. Ст. Большевиков, д. 54А, 394064, Воронеж, Россия; р.т.: +7(473)236-5228; факс: +7(473)244-7860.

**Стешковой Анатолий Сергеевич** — научный сотрудник, научно-исследовательский отдел, Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (ВУНЦ ВВС «ВВА»). Область научных интересов: методы и системы защиты информации. Число научных публикаций — 6. [9515431635@mail.ru](mailto:9515431635@mail.ru); ул. Ст. Большевиков, 54А, 394064, Воронеж, Россия; р.т.: +7(473)236-5228; факс: +7(473)244-7860.

M. PEREGUDOV, A. STESHKOVY  
**CENTRALLY RESERVED ACCESS MODEL TO THE MEDIUM IN  
DIGITAL RADIO COMMUNICATION NETWORKS**

*Peregudov M., Steshkovy A. Centrally Reserved Access Model to the Medium in Digital Radio Communication Networks.*

**Abstract.** Currently, centrally reserved access to the medium in the digital radio communication networks of the IEEE 802.11 family standards is an alternative to random multiple access to the environment such as CSMA/CA and is mainly used in the transmission voice and video messages in real time. Centrally reserved access to the environment determines the scope of interest in it from attackers. However, the assessment of effectiveness of centrally reserved access to the environment under the conditions of potentially possible destructive impacts was not carried out and therefore it is impossible to assess the contribution of such impacts to the decrease in the effectiveness of such access. Also, the stage establishing of centrally reserved access to the environment was not previously taken into account. Analytical model development of centrally reserved access to the environment under the conditions of destructive influences in digital radio communication networks of the IEEE 802.11 family standards. A mathematical model of centrally reserved access to the environment has been developed, taking into account not only the stage of its functioning, but also the stage of formation under the conditions of destructive influences by the attacker. Moreover, in the model the stage of establishing centrally reserved access to the medium displays a sequential relationship of such access, synchronization elements in digital radio communication networks and random multiple access to the medium of the CSMA/CA type. It was established that collisions in the data transmission channel caused by destructive influences can eliminate centrally reserved access to the medium even at the stage of its establishment. The model is applicable in the design of digital radio communication networks of the IEEE 802.11 family of standards, the optimization of such networks of the operation, and the detection of potential destructive effects by an attacker.

**Keywords:** Centrally Reserved Access to the Medium, Synchronization, Random Multiple Access to the Medium, Switching and Control Mean, Subscriber Terminal, IEEE 802.11

**Peregudov Maksim** — Ph.D., Head of Laboratory, Assistant Head of Department, Research Laboratory, Research Department, Military Education-science Center of Military Aviation Forces "Military Aviation Academy Named for Prof. N.E. Zhukovsky and J.A. Gagarin" (MESC MAF "MAA"). Research interests: methods and systems of information protection. The number of publications — 15. maxaperegudov@mail.ru; д. 54А, St. Bolshevikov str., 394064, Voronezh, Russia; office phone: +7(473)236-5228; fax: +7(473)244-7860.

**Steshkovy Anatoliy** — Researcher, Research Department, Military Education-science Center of Military Aviation Forces "Military Aviation Academy Named for Prof. N.E. Zhukovsky and J.A. Gagarin" (MESC MAF "MAA"). Research interests: methods and systems of information protection. The number of publications — 6. 9515431635@mail.ru; 54А, St. Bolshevikov str., 394064, Voronezh, Russia; office phone: +7(473)236-5228; fax: +7(473)244-7860.

## References

1. Makerenko S.I. [Suppression of packet radio networks with random multiple access due to destabilization of their state]. *ZHurnal radioelektroniki – Journal of Electronics*. 2011. no. 9. p. 2. Available at: [www. http://jre.cplire.ru/jre/sep11/4/text.pdf](http://jre.cplire.ru/jre/sep11/4/text.pdf) (accessed 10.12.2019). (In Russ.).

2. Titov K.D., Zavalishina O.N. [Assesment of noise immunity of standart data transmisions IEEE 802.11ac under the influence of interference]. *Uspekhi sovremennoi radioelektroniki – Advances in modern radio electronics*. 2019. no. 12. pp. 191–196 (In Russ.).
3. Deniau V. et al. IEEE 802.11n Communications in the Presence of FrequencySweeping Interference Signals. *IEEE Transactions on Electromagnetic Compatibility*. 2017. vol. 59. no. 5. pp. 1625–1633.
4. Scalia L., Tinnirello I., Giustiniano D. Side effects of ambient noise immunity techniques on outdoor IEEE 802.11 deployments. *GLOBECOM. Proceedings of the Global Telecommunications Conference*. 2008. pp. 1–6.
5. Titov K.D., Lipatov A.O., Zavalishina O.N. [Assessment of noise immunity of IEEE 802.11n communication system in case of intentional interference taking into account the structure of the transmitted data packet]. *Teoriya i tekhnika radiosvyazi – Theory and technique of radio communication*. 2019. no. 9. pp. 95–107. (In Russ.).
6. Makarenko S.I. [Dynamic Model of Communication System in Conditions the Functional Multilevel Information Conflict of Monitoring and Suppression]. *Sistemy upravleniya, svyazi i bezopasnost' – Systems of Control, Communication and Security*. 2015. no. 3. pp. 122–185. (In Russ.).
7. Aganesov A.V., Makarenko S.I. [Aerospace communications network model with traffic routing hierarchical principle]. *Radiotekhnicheskie i telekommunikacionnye sistemy – Radio engineering and telecommunication systems*. 2015, no. 4, pp. 43–51. (In Russ.).
8. Bojko A.A. [Method of analytical modeling of viruses propagation process in computer networks with different topology]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 5. pp. 196–211. (In Russ.).
9. Boyko A.A., Obushenko E.Y., Shcheglov A.V. [About synthesis of a full set of test methods of remote information-technical impacts on spatially distributed systems of information-technical tools]. *Vestnik Voronezhskogo gosudarstvennogo universiteta – Bulletin of Voronezh State University*. 2017. vol. 2. pp. 33–45. (In Russ.).
10. Peregudov M.A., Boyko A.A. [Model procedure of random multiple access to the environment type S-ALOHA]. *Informacionno-upravljajushhie sistemy – Information-control systems*. 2014. vol. 6. pp. 75–81. (In Russ.).
11. Peregudov M.A., Boyko A.A. [Estimation of security of a network packet radio from imitation of user's terminals at level of the procedure of random multiple access to the environment type S-ALOHA]. *Informacionnye tekhnologii – Information Technology*. 2015. vol. 7. pp. 527–534. (In Russ.).
12. Peregudov M.A., Semchenko I.A. [Evaluation of efficiency of random multiple access to ALOHA type environment with voice connections, transfer of service commands, text messages and multimedia files in destructive impact conditions]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2019. vol. 18. no. 4. pp. 887–911. (In Russ.).
13. Peregudov M.A., Steshkovoy A.S., Boyko A.A. [Probabilistic random multiple access procedure model to the CSMA/CA type medium]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2018. vol. 4. pp. 92–114. (In Russ.).
14. Peregudov M.A., Boyko A.A. [Model of reserved access procedure to environment of packet radio network]. *Telekommunikacii – Telecommunications*. 2015. vol. 6. pp. 7–15. (In Russ.).
15. Peregudov M.A., Boyko A.A. [Model of power supply control procedure of packet radio network]. *Telekommunikacii – Telecommunications*. 2015. vol. 9. pp. 13–18. (In Russ.).
16. Peregudov M.A., Steshkovoy A.S. [Digital radio networks centralized elements synchronization model with random multiple access to the CSMA/CA type medium]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2020. vol. 19. no. 1. pp. 128 – 154. (In Russ.).

17. Liu C., Qiu J. Performance study of 802.11w for preventing DoS attacks on wireless local area networks. *Wireless personal communications*. 2017. no. 95. pp. 1031–1053.
18. Kaur J. Mac Layer Management Frame Denial of Service Attacks. *International Conference on Micro-Electronics and Telecommunication Engineering*. 2016. pp. 155–160.
19. Filipek J., Hudec L. Securing mobile ad hoc networks using distributed firewall with PKI. *IEEE 14th International Symposium on Applied Machine Intelligence and Informatics*. 2016. pp. 321–325.
20. Yacchirena A. et al. Analysis of attack and protection systems in Wi-Fi wireless networks under the Linux operating system. *IEEE International Conference on Automatica*. 2016. pp. 1–7.
21. Liu C., Qiu J. Performance study of 802.11w for preventing DoS attacks on wireless local area networks. *Wireless personal communications*. 2017. vol. 95. pp. 1031–1053.
22. Noman H.A., Abdullah S. M., Mohammed H.I. An Automated Approach to Detect Deauthentication and Disassociation Dos Attacks on Wireless 802.11 Networks. *International Journal of Computer Science Issues*. 2015. vol. 12. 1694–1784 p.
23. Peregudov M.A., Steshkovoy A.S., Shcheglov A.V. Descriptive Model of Networks Broadband Access Link Layer for the IEEE 802.11 Standards Family. *Systems of Control, Communication and Security*. 2020. vol. 3. pp. 203–221. (In Russ.).
24. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Computer Society LAN MAN Standards Committee*. 1997.
25. IEEE standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements PART 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std. 802.11–2012*. 2012. pp. 1–2793.
26. IEEE Standards Association/IEEE Computer Society. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; Amendment 4: Enhancements for Very High Throughput for Operation in Bands Below 6 GHz. *IEEE Std. 802.11–2013*. 2013. pp. 1–425.
27. Kanjanavastit A., Landfeldt B. An Analysis of a Modified Point Coordination Function in IEEE 802.11. *Proceedings of IEEE 14th Personal, Indoor and Mobile Radio Communications*. 2003. vol. 2. pp. 1732–1736.
28. Sikdar B. An analytic model for the delay in IEEE 802.11 PCF MAC based wireless networks. *IEEE Transactions on Wireless Communications*. 2007. vol. 4. no. 6. pp. 1542–1560.
29. Qiao D., Choi S., Soomro A., Shin G. Energy-Efficient PCF Operation of IEEE 802.11a Wireless LAN. *In Proc. IEEE INFOCOM*. 2002. vol. 2. pp. 580–589.
30. Guan Z., Yang Z. J., He M. Energy-efficient analysis of an IEEE 802.11 PCF MAC protocol based on WLAN. *Journal of Ambient Intelligence & Humanized Computing*. 2018. pp. 1–11.
31. Zheng G., Zhi-Jun Y., Min H. Energy-efficient analysis of an IEEE 802.11 PCF MAC protocol based on WLAN. *Journal of Ambient Intelligence and Humanized Computing*. 2018. pp. 1–11.
32. Eyadeh A., Jarrah M., Aljumaili A. Modeling and simulation of performance limits in IEEE 802.11 point-coordination function. *International Journal of Recent Technology and Engineering*. 2019. vol. 8(4). pp. 5575–5580.
33. Noman H.M. PCF and DCF Performances Evaluation for a Non Transition 802.11 Wireless Network using OPNET Modular. *International Journal of Soft Computing and Engineering*. 2017. vol. 7. pp. 2231–2307.
34. Sarmah S., Sharma S. K. Performance Analysis of IEEE 802.11 WLANs by varying PCF, DCF and EDCF to Enhance Quality of service. *International Journal of Computer Applications*. 2016. pp. 138.

35. Dhaliwal A.S. Analyzing the Impact of DCF and PCF on WLAN Network Standards 802.11a, 802.11b and 802.11g. *Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*. 2013. vol. 7. pp. 1594–1598.
36. Chen D., Garg S., Kappes M., Trivedi K. Supporting VBR VoIP traffic in IEEE 802.11 WLAN in PCF mode. *Avaya Laboratories*. 2002. vol. 26.
37. Vishnevsky V., Lyakhov A. Analytical Study of IEEE 802.11 PCF for regional and Metropolitan Area Networks. *Cybernetics and Information Technologies*. 2005. vol. 5. no. 2. pp. 117–136.
38. Liu Q., Zhao D., Zhou D. An analytic model for enhancing IEEE 802.11 point coordination function media access control protocol. *European transactions on telecommunications*. 2011. vol. 22. pp. 332–338.
39. Kaur I., Bala M., Bajaj H. Performance evaluation of wlan by varying PCF, DCF and enhanced DCF slots to improve quality of service. *IOSR Journal of Computer Engineering*. 2012. vol. 2. pp. 29–33.
40. Shigeo S., Daiki T. Bistable Behavior of IEEE 802.11 Distributed Coordination Function. 22nd International Symposium on Wireless Personal Multimedia Communications. 2019.
41. Burton M. 802.11 Arbitration. Certified Wireless Network Professional Inc. Durham. 2009. 24 p.
42. Boyko A.A., Peregudov M.A., Semchenko I.A., Steshkovej A.S. [Optimizing the operation of radio communication networks software package]. *Svidetel'stvo ob ofitsial'noi registratsii programm dlya EVM – The Certificate on Official Registration of the Computer Program*. 2018. vol. 2018614894. (In Russ.).
43. Peregudov M.A., Degtyarev I.S., Umanskij A.YA., Semchenko I.A., Steshkovej A.S., SHCHeglov A.V. [Software package for diagnosing digital radio networks]. *Svidetel'stvo ob ofitsial'noi registratsii programm dlya EVM – The Certificate on Official Registration of the Computer Program*. 2019. vol. 2019665751. (In Russ.).

---

Signed to print 03.12.2020

Printed in Publishing center GUAP, 67, B. Morskaya, St. Petersburg, 190000, Russia

---

The journal is registered in the Russian Federal Agency for Communications and Mass-Media Supervision, certificate ПИ № ФС77-79228 dated September 25, 2020  
Subscription Index П5513, Russian Post Catalog

---

Подписано к печати 03.12.2020. Формат 60×90 1/16. Усл. печ. л. 13,37. Заказ № 293.

Тираж 300 экз., цена свободная.

Отпечатано в Редакционно-издательском центре ГУАП, 190000, Санкт-Петербург, Б. Морская, д. 67

---

Журнал зарегистрирован Федеральной службой по надзору в сфере связи и массовых коммуникаций, свидетельство ПИ № ФС77-79228 от 25 сентября 2020 г.

Подписной индекс П5513 по каталогу «Почта России»

ISSN 2713-3192



9 772713 319007 >