

ПРЕДВАРИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ ОБРАБОТКИ РАЗНОТИПНЫХ БИОМЕТРИЧЕСКИХ ДАННЫХ МЕТОДАМИ DATA MINING

Дюк В.А., Жвалевский О.В., Рудницкий С.Б., Толстоногов Д.А.

УДК 519.25

Дюк В.А., Жвалевский О.В., Рудницкий С.Б., Толстоногов Д.А. **Предварительные результаты обработки разнотипных биометрических данных методами data mining.**

Аннотация. Рассматривается задача совместной обработки разнотипных биометрических данных, полученных в результате измерений пульса, микровибраций лица, и оценки уровня тревожности на основе психологического тестирования. В частности, находятся скрытые признаки, различающие две группы практически здоровых испытуемых: 1) людей различного пола и возраста; 2) беременных женщин. Приводятся предварительные результаты анализа экспериментальных данных методами data mining.

Ключевые слова: Data Mining, временные ряды, обработка данных, классификация.

Duke V.A., Zhvalevsky O.V., Rudnitsky S.B., Tolstonogov D.A. **The Preliminary Results of Mixed Biometrical Data Processing by Data Mining Methods.**

Abstract. The problem of joint processing of polytypic biometric data from puls and microvibration measuring and uneasiness level estimation by psychology testing is considered. In particular, Hidden discriminate features of two groups of healthy testees: 1) people of different gender and age; 2) pregnant women. The preliminary results of experimental data processing by Data Mining methods are given.

Keywords: Data Mining, time series, preprocessing, classification.

1. Введение. В статье рассматривается задача совместной обработки разнотипных биометрических данных, полученных в результате измерений пульса, микровибраций лица, и оценки уровня тревожности испытуемого. Пульс измерялся простым датчиком, основанным на фотозффекте; сама пульсограмма получалась при выставлении некоторого заданного уровня (порога) и регистрации т.н. «М-М интервалов». Микровибрации лица измерялись при помощи приложения AuraVibraImage компании «Элсис» посредством вычисления межкадровой разности с последующим вычислением интегральных показателей по каждому кадру. Уровень тревожности определялся в ходе психологического тестирования при помощи программы «Интегративный тест тревожности», разработанной авторским коллективом института им. Бехтерева под руководством Л. И. Вассермана. Три указанные измерительные системы собраны в виде программно-измерительного комплекса, предназначенного для определения психосоматического статуса испытуемого, и проходящего апробацию в Российском научном центре радиологии и

хирургических технологий (ФГУ «РНЦРХТ Росмедтехнологий», бывший ЦНИИРИ).

Онкологами уже давно высказывается гипотеза о том, что развитие онкологических заболеваний в определенном смысле может иметь общие проявления с развитием плода у беременных женщин. Представляет интерес анализ биометрических данных на предмет значимых различий между практически здоровыми женщинами детородного возраста и беременными с последующей проверкой найденных различий (в случае их обнаружения) на онкологических пациентах.

В статье решается задача классификации для объектов двух классов: практически здоровые испытуемые, и беременные женщины. Используются следующие методы Data Mining: нейронные сети, метод ближайшего соседа с метрикой, основанной на сжатии, и метод информационного резонанса.

2. Постановка задачи. Каждый метод измерений, включая и психологическое тестирование, предоставляет исследователю разнородный набор числовых данных. Это и исходные временные ряды значений измеряемых физиологических показателей, и результаты обработки временных рядов в виде некоторой совокупности признаков, и величины, принимающие значения в шкалах наименований и порядковых шкалах, и результаты вторичной обработки данных в виде оценок и интегральных показателей. При этом предполагается, что все необходимые измерения производятся практически одновременно в ходе непрерывного сеанса за одно посещение испытуемого. Таким образом, ставится задача совместной обработки разнотипных биометрических данных. Сюда относится поиск и исследование зависимостей между различными признаками и группами признаков, снижение размерности признакового пространства с целью определения наиболее информативных для решаемой задачи признаков, и собственно решение задачи классификации испытуемых.

Для проверки гипотезы о сходстве проявлений при онкологических заболеваниях и беременности необходимо обследовать, по крайней мере, три группы испытуемых:

- 1) практически здоровые женщины детородного возраста;
- 2) практически здоровые беременные женщины;
- 3) онкологические больные.

Первые две группы можно использовать в качестве обучающей выборки и выяснять различие между объектами двух данных групп, а

третью использовать в качестве контрольной выборки для подтверждения полученного ранее результата (если искомое различие будет найдено).

Фактически, при апробации программно-инструментального комплекса в ФГУ «РНЦРХТ Росмедтехнологий» группой «Хрономедицина» под руководством д.м.н М.А. Бланка, был собраны экспериментальные данные для двух групп практически здоровых испытуемых:

- I. 97 людей различного возраста, из которых 45 мужчин и 55 женщин;
- II. 190 беременных (и практически здоровых) женщин, находящихся на различных сроках беременности.

В силу того, что указанные выборки нельзя признать репрезентативными, предполагаемые результаты следует оценивать как предварительные, необходимые для того, чтобы оценить саму возможность совместной обработки разнотипных биометрических данных.

3. Исходные данные. В результате измерений была собрана база экспериментальных данных (см. табл. 1).

Табл. 1. **Классы испытуемых**

№	Класс испытуемых	кол-во	расшифровка
0	Класс неизвестен	28	
1	Проба	7	пробные измерения
2	Контроль	85	основная группа
3	Первичный пациент	11	
4	Процтер	18	в процессе терапии
5	Контррем	37	в ремиссии
6	Обследование	4	дополнительные
7	Контроль грав	194	беременные
	ВСЕГО	408	

Для каждого испытуемого из экспериментальных данных были получены признаки, приведенные в табл. 2 (звёздочка «*» означает, что название признака начинается с указанных букв):

Табл. 2. **Общие признаки**

Num	Уникальный идентификатор испытуемого
Age	Возраст испытуемого
Sex	Пол испытуемого
Vibro*	28 признаков, полученных при анализе виброизображения
Puls*	22 признака, полученных при анализе пульса
ИТТ*	12 признаков, отвечающих ИТТ
ИТТAnswer*	Ответы на 30 вопросов, полученные при прохождении ИТТ

Виброизображение лица получалось при помощи приложения AuraVibraImage, основанного на накоплении межкадровой разности. В результате обработки каждого кадра в файле протокола фиксировались значения четырёх показателей, табл. 3:

Табл. 3. **Данные, получаемые для виброизображения**

Показатель	Описание
Integrated All (IA)	Интегральная интенсивность пульсации изображения за интервал времени по всему кадру
$f(IA)$	Частота измерения IA (основной метод расчёта частоты)
$(IA^2)/f(IA)$	Расчётная величина
FPS	Текущая частота кадров

Результаты измерений представляют собой временные ряды, регистрируемые в продолжение времени длительностью около трёх минут. Чтобы получить набор признаков, брались последние 100 отсчётов каждого временного ряда. Это делалось для того, чтобы обеспечить одни и те же условия измерений, а, значит, однородность анализируемых данных. Такой подход можно использовать при разведочном анализе данных, когда необходимо гарантировать отсутствие в данных привнесённых особенностей, вызванных начальной реакцией испытуемого на саму процедуру измерений, и использовать данные, полученные для различных испытуемых при прочих равных условиях. Затем каждый из полученных временных рядов обрабатывался по одной и той же схеме, в результате которой для каждого ряда было вычислено по семь признаков следующего вида, всего 28 признаков, см. табл. 4:

Табл. 4. Признаки, вычисляемые из виброизображений

Признак	Способ вычисления	Название заканчивается на...
Отношение СКО/Ср. Знач	Среднеквадратичное отклонение (СКО) нормируется средним значением (Ср. Знач)	*Std
Энтропия	Интервал значений разбивается на 14 отрезков, строится гистограмма и подсчитывается энтропия распределения по известной формуле	*Entropy
Фрактальная размерность Хигучи	Для системы подпоследовательностей анализируемого ряда вычисляется взвешенное среднее их энтропий	*Fractality
Первые четыре сингулярных собственных числа	Исходный ряд преобразуется в матрицу ряда методом «Гусеница»	*EigenValue1 *EigenValue2 *EigenValue3 *EigenValue4

При более глубоком анализе необходимо брать исходные временные ряды, поскольку при усечении могла быть потеряна важная информация о протекании физиологических процессов. В частности, существующие в рядах циклы могут быть смещены по фазе у разных испытуемых, ряды могут иметь различную длину, и т.п. Визуальное рассмотрение анализируемых временных рядов обнаруживает в данных импульсные всплески (см., например, рис. 1), при этом непонятно, являются ли они особенностями процесса накопления межкадровой разности, следствием каких-либо иных технических свойств измерительной подсистемы, либо движениями пациента, т.е. артефактами.

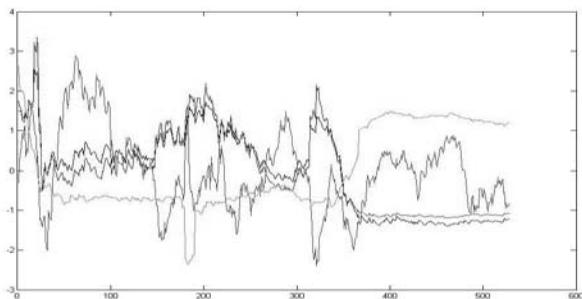


Рис. 1. Пример исходных данных для виброизображения.

В конечном итоге, информативность параметров, вычисляемых в программе AuraVibraImage, не очевидна. Это означает, что необходим дополнительный анализ исходных данных, а именно — непосредственно самого изображения.

Пульсометрия заключается в получении последовательности М-М интервалов, т.е. промежутков между последовательными моментами времени, когда сигнал преодолевает заданный пороговый уровень, являющийся частью реализации датчика. Программа-клиент получает только сигналы о срабатывании датчика. В программе «Пульсомер» производится автоматическая обработка каждой пульсограммы, в результате которой вычисляются следующие признаки, см. табл. 5:

Табл. 5. Признаки, полученные из пульсограмм

Имя	Описание
kol	Количество М-М интервалов после корректировки и отбрасывания выпадающих значений
Nmax	Полное количество измеренных М-М интервалов
time	Среднее значение М-М интервала
freg	Средняя частота пульса, 1/мин
Mmin	Наименьшее значение М-М интервала
Mmax	Наибольшее значение М-М интервала
dx	Наибольшая по абсолютной величине разность между средним значением R-R интервала и его граничными значениями (вариационный размах)
AM	Амплитуда моды М-М интервалов;
m	Число сдвигов до полной декорреляции
M	Мода значений М-М интервалов;
sigma	Среднее квадратическое отклонение ряда М-М интервалов
var	Коэффициент вариации ряда М-М интервалов
per	Коэффициент асимметрии, оценивающий наличие переходных процессов
wper	Показатель, характеризующий выраженность переходных процессов
stab	Показатель, характеризующий стабильность М-М интервалов
rr	Показатель, характеризующий интенсивность медленных изменений М-М интервалов
idw	Показатель, характеризующий интенсивность дыхательных волн
k	Коэффициент автокорреляции при сдвиге на одно измерение
Tmax	Общее время измерения М-М интервалов
nW	Размер массива результатов спектрального анализа

Имеется также массив диагностических оценок (см. Табл. 6).

Табл. 6. Массив диагностических оценок

Имя	Описание
diagnos1	Суммарный эффект регуляции
diagnos2	Функция автоматизма
diagnos3	Вегетативный гомеостаз
diagnos4	Устойчивость регуляции
diagnos5	Активность подкорковых нервных центров

Результаты «интегративного теста тревожности» непосредственно задействованы не были. Поэтому описание этих признаков опущено.

4. Методика предварительной обработки данных. Из потенциально информативных признаков, полученных при обработке результатов измерений микродвижений лица, необходимо выделить те, которые с достаточной точностью разделяли бы классы I и II. Задача осложняется наличием пропусков и аномальных значений признаков.

Общеизвестно, что одна из наиболее важных частей исследования и наиболее трудозатратная и сложная — это предварительная подготовка данных. При правильной постановке задачи подготовка данных часто является ключом к решению проблемы [1]. Исходя из этого, к исходной матрице экспериментальных данных применяется следующая последовательность преобразований.

Прежде всего, из анализа исключаются признаки «возраст» и «пол», поскольку очевидными правилами принадлежности к классу II являются женский пол и детородный возраст.

Далее, обрезаются существующие аномальные выбросы и значения всех признаков преобразуются к неотрицательным путем их сдвига на соответствующие положительные величины.

Затем, диапазон значений категориальных признаков сдвигается так, чтобы уменьшить размерность матрицы входов при их бинарном кодировании. Например, диапазон значений $1, \dots, N$ преобразуется в диапазон $0, \dots, N - 1$.

Ко всем признакам, имеющим ординальные непрерывные значения, применяется преобразование Бокса-Кокса [2] следующего вида:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda \neq 0; \\ \log \lambda, & \lambda = 0. \end{cases}$$

Это преобразование приближает исходные функции распределения признаков к нормальному распределению. Параметр λ вычисляется процедурой оптимизации, примененной к соответствующему временному ряду.

Далее, признаки нормализуются так, чтобы они имели нулевое среднее и единичное стандартное отклонение:

$$y := \frac{y - \text{mean}(y)}{\text{std}(y)}.$$

Для того, чтобы перейти к дискретным значениям признаков, область допустимых значений каждого из ординальных признаков разбивается на 10 ячеек, в которые значения признаков попадает с приблизительно одинаковой вероятностью, исходя из известных квантилей для нормального распределения. Значению признака ставится в соответствие номер его ячейки.

Полученная дискретная матрица преобразуется к бинарному виду. Для этого ординальные переменные кодируются по принципу «термометра» бинарной строкой длиной в заданное число ячеек. Например, значение 3 ординального признака кодируется строкой $[1,1,1,0, \dots, 0]$. Категориальные признаки кодируются бинарной строкой следующим образом. Первая категория кодируется строкой $[1,0,0, \dots, 0]$, вторая — $[0,1,0, \dots, 0]$, и т.п. Число строк матрицы, соответственно, увеличивается.

Полученная бинарная матрица, состоящая из строк-примеров и столбцов-признаков, разбивается случайным образом на обучающее и тестовое множества.

Для непосредственного сравнения исходных временных рядов, полученных при измерениях пульса и микровибраций лица используется другая методика предварительной обработки данных, которая заключается в следующем.

К каждому из исходных рядов применяется преобразование Гильберта-Хуанга [3], являющееся более современной альтернативой преобразованию Фурье и вейвлет-преобразованию, поскольку оно полностью адаптивно, применимо к нестационарным рядам и не налагает никаких априорных предположений об исследуемом ряде. Данное преобразование разлагает исходный ряд на сумму приблизительно ортогональных, модулированных по амплитуде и частоте компонент с нулевыми средними — так называемых «эмпирических мод». Ортогональные разложения обычно используют для лучшего сжатия данных, т.к. декомпозированные данные имеют

более простую структуру.

Каждая компонента преобразования для каждого ряда нормализуется независимо к нулевым средним и единичным стандартным отклонениям, и они объединяются в общий ряд-строку.

Полученный длинный временной ряд аппроксимируется символьной строкой, с использованием тех же приемов, что описывались ранее.

Отдельно с исходными рядами проделываются те же манипуляции, но без первоначального применения преобразования Гильберта-Хуанга.

5. Решение задачи классификации. Сначала для нахождения специфических признаков, разделяющих примеры на классы I и II, к полученной обучающей бинарной матрице был применен классификатор на основе нейронной сети с методом обучения, аналогичным алгоритму TurboProp от Ward Systems. Как и предполагалось, нейронная сеть не смогла найти различия в классах даже на обучающем множестве, поскольку нейросети малоприменимы к задачам с относительно большим числом признаков и относительно малым числом примеров.

Затем, к полученной обучающей бинарной матрице был применен оригинальный алгоритм В.А. Дюка [4], основанный на эффекте информационного структурного резонанса. При этом информативными оказались лишь два признака, полученные из данных пульсометрии.

Признаки, кодирующие виброизображения, в данной задаче оказались *неинформативными*. Поэтому, вместо того, чтобы использовать 28 признаков, искусственным образом полученных для микродвижений лица, сравнивались непосредственно исходные временные ряды.

Общеизвестно, что простейший метод ближайшего соседа при классификации временных рядов способен дать во многих ситуациях результаты, лучшие или не худшие, чем другие более сложные методы. Метод ближайшего соседа заключается в том, что класс данного объекта определяется классом «ближайшего» к нему в определенном смысле объекта, чей класс заранее известен.

К сложностям задачи относятся те факты, что ряды измерений параметров имеют разную длину и не синхронизированы по фазе, поэтому традиционные метрики, типа евклидовой, для вычисления «расстояния» между рядами оказываются неприменимы.

В таких случаях обычно применяют меру расстояния, основанную

на «динамическом искривлении времени» (Dynamic Time Warping, DTW), где расстояние между двумя строками Q и C вычисляется по формуле:

$$DTW(Q, C) = \min \left(\sqrt{\sum_{k=1}^K w_k} / K \right),$$

Где «искривленный путь» w вычисляется при помощи методов динамического программирования [5].

В данной работе использовался альтернативный подход, основанный на теории сложности по Колмогорову. Поскольку сложность по Колмогорову в общем случае не может быть вычислена явно, для вычисления соответствующей метрики вместо теоретической сложности использовалась степень сжатия данных при архивировании текстовых строк с данными задачи общеизвестным алгоритмом ZIP. В работе [6] показано, что этот подход для достаточно длинных рядов предпочтительнее, чем DTW.

Применялась следующая мера расстояния между текстовыми строками:

$$dist(x, y) = \frac{size([x, y])}{size(x) + size(y)},$$

где $size(x)$ - размер файла при сжатии строки x алгоритмом ZIP, $[x, y]$ - операция объединения строк. Следует заметить, что эта мера расстояния не является метрикой в обычном смысле, так как не удовлетворяет всем аксиомам метрики: очевидно, что расстояние от объекта до самого себя не равно нулю.

Были вычислены попарные расстояния между всеми объектами, принадлежащими двум рассматриваемым классам. Объекты относились к тому или иному классу исходя из их близости к объектам известного класса.

Вместо проверки алгоритма на тестовом множестве использовался метод Leave-One-Out-Cross-Validation, когда каждый объект поочередно становился тестовым, а остальные объекты – обучающими. Необходимо принять во внимание, что этот метод имеет тенденцию недооценивать ошибки прогноза при малом числе примеров [7].

7. Результаты. При помощи информационного структурного резонанса удалось найти высокоточные правила разделения классов по следующим двум признакам: `diagnos5` и `Mmin`. Точность разделения классов по этим признакам на тестовом множестве составила 89,34%.

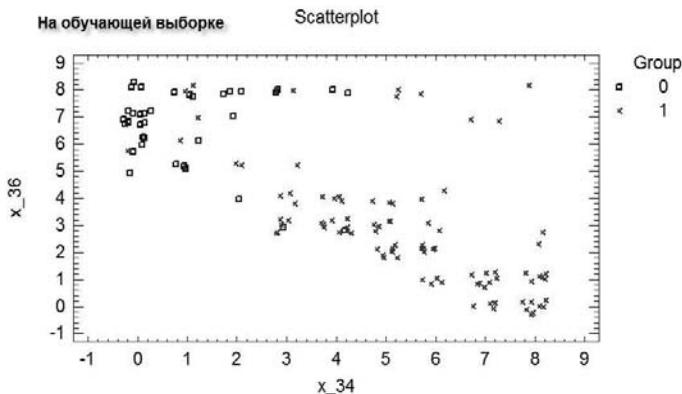


Рис. 2. График разброса между классами в координатах информативных признаков на обучающей выборке.

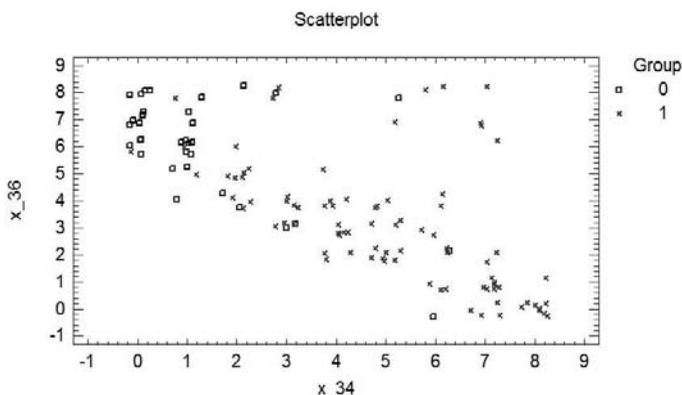


Рис. 3. График разброса между классами в координатах информативных признаков на тестовой выборке.

Остальные признаки оказались неинформативными.

Решая задачу классификации непосредственно для исходных временных рядов виброизображений, значимого различия между искомыми классами не удалось найти, как с использованием преобразования Гильберта-Хуанга, так и без него. Точность классификации непосредственно пульсограмм на тестовом множестве оказалась чуть больше 70%.

8. Обсуждение. Актуальная проблема клинической медицины — это определение характерных проявлений онкологических заболеваний, что может иметь первостепенное значение для их ранней диагностики.

Разумеется, полученные признаки могут свидетельствовать лишь о беременности и не иметь отношения к раковой опухоли. Если говорить о факте беременности, правильнее искать различия между беременными и небеременными только среди женщин детородного возраста. Однако, для этого недостаточно экспериментального материала, чтобы можно было говорить о статистической достоверности результатов.

Отсутствие найденных информативных признаков в виброизображениях может быть результатом не только отсутствия информации в виброизображениях как таковой, но и в возможных некорректно проведенных экспериментах и полученных в результате этого артефактах в данных, и т.п.

9. Заключение. В работе идёт речь исключительно о методах совместной обработки данных. Однако, в действительности, следует гарантировать одновременное измерение пульсограммы и получение виброизображения, чтобы была возможность применить методы комплексной обработки данных. Поэтому дальнейшие исследования будут направлены на решение задачи одновременной регистрации разнородных физиологических параметров в рамках единого программно-инструментального комплекса с последующей обработкой разнотипных биометрических данных методами Data Mining.

Литература

1. *Dorian Pyle*, Data Preparation For Data Mining, Morgan Kaufmann, 1999.
2. *Box, George E.P.; Cox, D.R.* An analysis of transformations // Journal of the Royal Statistical Society, 1964, Series B 26: 211–246.
3. Hilbert-Huang Transform and Its Applications / Editors Norden E Huang, Samuel S P Shen. 2005, World Scientific Publishing.
4. Дюк В.А., Эмануэль В.Л. Информационные технологии в медико-биологических исследованиях. СПб.: Питер. 2003. 525 с.
5. *Keogh, E.* Exact indexing of dynamic time warping. In 28th International Conference on Very Large Data Bases. Hong Kong. pp 406-417.
6. *Keogh, E., Lonardi, S. and Ratanamahatana, C.* Towards Parameter-Free Data Mining. In proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, WA, Aug 22-25, 2004.
7. *Martens, H.A.; Dardenne, P.* Validation and verification of regression in small data sets 1998, 44, 99-121.

Дюк Вячеслав Анатольевич — д.т.н., ведущий научный сотрудник лаборатории биомедицинской информатики Санкт-Петербургского института информатики и автоматизации РАН. Область научных интересов: информационные технологии в медико-биологических исследованиях; методы и алгоритмы классификации, диагностики, распознавания образов и прогнозирования; поиск неперiodических шаблонов с джокерами в последовательностях чисел и символов; искусственный интеллект, обнаружение знаний в базах данных (Data Mining). Число научных публикаций — 80. duke@spiiras.nw.ru; СПИИРАН, 14 линия, дом 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-5411, факс +7(812) 328-4450.

Dyuk Viacheslav Anatolievich — Dr. Sci. (Tech.); leading researcher, Laboratory of Biomedical Informatics, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: Data Mining in medical and biologic researches; methods and algorithms of classification, diagnostics, pattern recognition and forecasting; search of acyclic patterns with jokers in sequences of numbers and symbols. The number of publications — 80. duke@spiiras.nw.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-5411, fax +7(812)328-4450.

Жвалевский Олег Валерьевич — научный сотрудник лаборатории биомедицинской информатики Учреждения Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН). Область научных интересов: методы обработки временных рядов, методы комплексной обработки данных, базы биометрических данных, интеграция приложений. Число научных публикаций — 5. ozh@spiias.ru, СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-5411, факс +7(812)328-4450.

Zhvaleyevsky Oleg Valerievich — researcher, Biomedical Information Laboratory, , St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: time series processing methods, data complex processing methods, biometrical data bases, applications integration. The number of publications — 5. roudnitsky@spiias.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-5411, fax +7(812)328-4450.

Рудницкий Сергей Борисович — д.т.н., заведующий лабораторией биомедицинской информатики Учреждения Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН). Область научных интересов: дистанционная биометрия, хронобиология, комплексная обработка сигналов, радионавигация. Число научных публикаций — 80. roudnitsky@spiias.ru, СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-5411, факс +7(812)328-4450.

Roudnitsky Sergey Borisovich — Dr. Sci. (Tech.); leading researcher, Biomedical Information Laboratory, , St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: distance biometry, chronobiology, cybernetics, complex signal processing, and radionavigation. The number of publications — 80. roudnitsky@spiias.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-5411, fax +7(812)328-4450.

Толстоногов Дмитрий Александрович — к.ф.-м.н.; старший научный сотрудник лаборатории биомедицинской информатики Учреждения Российской академии наук

Санкт-Петербургский институт информатики и автоматизации РАН (СПИИРАН). Область научных интересов: информационные технологии в медико-биологических исследованиях; методы и алгоритмы классификации, диагностики, распознавания образов и прогнозирования; искусственный интеллект, обнаружение знаний в базах данных (Data Mining). Число научных публикаций — 12. dat@spiias.ru, СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-5411, факс +7(812)328-4450.

Toltsonogov Dmitry Aleksandrovich — PhD, Sci. (Tech.); senior researcher, Biomedical Information Laboratory, , St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: Data Mining in medical and biologic researches; methods and algorithms of classification, diagnostics, pattern recognition and forecasting. The number of publications — 12. dat@spiias.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-5411, fax +7(812)328-4450.