

В.Ю. ОСИПОВ, С.В. КУЛЕШОВ, Д.И. МИЛОСЕРДОВ, А.А. ЗАЙЦЕВА,  
А.Ю. АКСЕНОВ

## РЕКУРРЕНТНЫЕ НЕЙРОННЫЕ СЕТИ С НЕПРЕРЫВНЫМ ОБУЧЕНИЕМ В ЗАДАЧАХ МНОГОФУНКЦИОНАЛЬНОЙ ОБРАБОТКИ НОВОСТНЫХ ПОТОКОВ

*Осипов В.Ю., Кулешов С.В., Милосердов Д.И., Зайцева А.А., Аксенов А.Ю.*  
**Рекуррентные нейронные сети с непрерывным обучением в задачах многофункциональной обработки новостных потоков.**

**Аннотация.** Главной задачей использования нейронных сетей является оперативное и точное решение различных творческих задач, таких как анализ и синтез новостных потоков при сохранении непрерывности обучения. Результатом такой обработки могут быть дайджесты, новостные потоки, прошедшие фильтрацию, а также прогнозы событий, позволяющих обеспечивать проактивность в управленческих решениях. Известные методы обработки новостей нейронными сетями и реализующие их технические решения не в полной мере обеспечивают решение возникающих в этой области задач. Необходимо расширить их функциональные возможности, совершенствовать пространственно-временное связывание сигналов в рекуррентных нейронных сетях. При обработке новостных потоков одновременно с непрерывным обучением рекуррентных нейронных сетей следует осуществлять селекцию, распознавание, восстановление, прогнозирование и синтез новостей. Для снижения остроты проблемы предлагается перспективный метод многофункциональной обработки новостных потоков с применением рекуррентных нейронных сетей с логической организацией слоев и непрерывным обучением. Метод основан на развитии ассоциативной обработки текстовой информации в потоковых рекуррентных нейронных сетях с управляемыми элементами. Ключевыми особенностями этого метода являются многофункциональная обработка информационных потоков с изменяющимися законами появления новостей. Метод предусматривает оперативный отбор, распознавание, восстановление, прогнозирование и синтез новостей на основе глубокой ассоциативной непрерывной обработки связей между текстовыми элементами. Реализующая предлагаемый метод нейросетевая система отличается от известных решений новыми элементами, связями между ними, а также выполняемыми функциями. По результатам экспериментов подтверждена расширенная функциональность метода. Выявлены новые особенности обработки новостных текстов потоковыми РНС. Предлагаемые решения могут найти применение при создании интеллектуальных систем нового поколения не только для обработки текстов, но и других видов информации.

**Ключевые слова:** рекуррентные нейронные сети, интеллектуальная обработка новостей, многофункциональность, непрерывность обучения.

**1. Введение.** Качественная и своевременная обработка новостных потоков, связанная с интерпретацией событий, может стать основой для построения прогнозов, построения автоматических систем поддержки принятия решений с проактивным управлением, и соответственно принятия важных управляющих решений, в том числе с серьезными экономическими последствиями. Такая обработка предполагает анализ больших потоков информации от различных

источников, которые могут содержать как истинные, так и ложные сообщения. Вручную выполнить такой анализ в сжатые сроки крайне затруднительно, что приводит к необходимости использования для обработки новостных потоков методов машинной интеллектуальной обработки. Выделяемые новости должны соответствовать следующим критериям качества: быть востребованы целевыми потребителями, обладать новизной (то есть не иметь дублирований с ранее обработанными) и релевантностью на текущий момент времени (соответствие новостей порождающим их информационным событиям), быть достоверными, полными и непротиворечивыми. Входной поток новостей должен подвергаться интеллектуальной обработке, обеспечивающей соответствие выделенным критериям. Результатом обработки могут быть дайджесты, новостные потоки, прошедшие фильтрацию, а также прогнозы событий, позволяющих обеспечивать проактивность в управленческих решениях. Для обработки больших потоков данных в целях получения качественных новостей применим ряд известных методов [1–5]. Для приведения различного вида данных к текстовой цифровой форме на выбранном языке применяют методы словесного описания изображений [6], системы машинного перевода [7]. При машинной обработке решаются задачи распознавания и классификации текстов и их фрагментов, определения тональности, тематической направленности [5, 8], при этом формируются тематические модели, отражающие структуру тематических коллекций. В основном эта обработка сводится к распознаванию и аннотированию текстов, формированию рефератов [9, 10]. При этом исходят из наиболее часто встречающихся слов и их сочетаний. Такой подход трудно применим для определения новизны сообщений, которые на начальном этапе жизни слабо выделяются по частоте встречаемости слов. Однако на его основе можно осуществлять прогнозы будущих тем [11, 12], а также различных общественных и экономических событий [13–16], связанных с ними. Прогнозирование содержания текстов при использовании тематических моделей сводится лишь к определению наборов слов без тесной связи между ними. Наиболее отработанными решениями выступают методы предсказания слов и окончаний в предложениях [17]. Среди общих недостатков решений на основе ассоциативно-вероятностного подхода к обработке текстов в части прогнозирования, селекции и анализа новостей можно выделить:

- необходимость учета широкого контекста;

- существенно ограниченная точность по обнаружению новых по содержанию сообщений, их классификации и прогнозированию при анализе малых объемов текста;
- слабый учет содержания обрабатываемых слов и связей между ними в текстах;
- невысокая интеллектуальность выделения и анализа новостей в потоках, по сравнению с ручной обработкой;
- значительная вычислительная сложность решаемых задач.

Наиболее перспективным направлением для совершенствования машинной обработки новостных потоков выступает разработка и применение новых интеллектуальных нейросетевых систем. Чем больше творческих задач могут решать такие системы, тем выше их уровень интеллектуальности.

Нейронные сети (НС) позволяют успешно связывать и запоминать элементы текстов на различных уровнях иерархии. Посредством этого в нейронных сетях формируются постоянно обновляемые пространственно-временные модели воспринимаемых сигналов, несущих информацию о текстах. Потенциально эти модели применимы для решения различных задач, связанных с обнаружением, распознаванием, прогнозированием и генерацией новостей. Целесообразно, чтобы эти задачи решались на одной и той же нейросетевой системе без прерывания ее обучения. Известен ряд работ по традиционной нейросетевой обработке новостных потоков [18–22] и другие.

В подобных задачах применяют как НС сети прямого распространения, так и рекуррентные нейронные сети (РНС). Успешность обработки нейронными сетями новостных потоков зависит не только от характеристик применяемых сетей, но и от совершенства кодирования текстов и декодирования результатов на выходе НС. В частности, для обработки текстов в импульсных НС они должны быть закодированы в виде последовательностей совокупностей единичных импульсов. Вид такого кодирования во многом может влиять на уровень обработки текстов, оперативность и точность получаемых результатов. Несомненно, показатели обработки текстов в НС сетях зависят от накопленного опыта, их структуры, размера, логической организации, реализуемых правил, обеспечивающих связывание, ассоциативное запоминание и извлечение сигналов из памяти сетей. Применение традиционных НС позволило совершенствовать интеллектуальную обработку новостных потоков. Однако это только в пределах решения задач классификации и распознавания текстов. Расширенными возможностями по

интеллектуальной обработке новостных потоков обладают потоковые рекуррентные нейронные сети (РНС). К их представителям следует отнести прежде всего GRU (Gated Recurrent Units) и LSTM (Long short-term memory) сети [23], а также потоковые РНС с управляемыми элементами [5, 24, 25]. Сети GRU и LSTM, хотя и обладают значительными возможностями по управлению своими параметрами, однако требуют обучения с учителем. При изменении законов поведения обрабатываемых новостных потоков эти сети необходимо переобучать. Этого недостатка лишены самообучающиеся потоковые РНС с управляемыми элементами [5, 24–27]. В отличие от широко известных РНС они могут быть наделены различными логическими структурами, и обладают возможностями по управлению ассоциативными взаимодействиями обрабатываемых сигналов в зависимости от текущих слоев. Однако свойства потоковых РНС, включая управление ассоциативным взаимодействием сигналов, во многом не исследованы. Открытыми остаются вопросы по решению на одних и тех же структурах потоковых РНС множества различных задач анализа и синтеза.

В настоящей статье развиваются взгляды на процессы управляемой обработки текстов в этих РНС применительно к селекции, распознаванию, восстановлению, прогнозированию и синтезу новостей при непрерывном обучении для различных классов информационных потоков.

Предлагается новый метод многофункциональной обработки новостей потоковыми РНС и реализующая его программная нейросетевая система, отличающаяся от известных решений функциями, составом элементов и связей между ними.

**2. Метод.** Известные функции обработки новостей потоковыми РНС [5] включают в себя: извлечение данных из потоков, фильтрацию данных, кодирование, нейросетевое выделение текущих и прогнозирование содержания возможных информационных событий. Представление входных данных основано на предобработке (нормализации) новостных текстов с использованием NLP-инструментария [28] с последующим формированием связей между основными лексическими единицами на основе синтаксических конструкций [29]. При этом идентификаторы распределяются на основе словарей, как предопределенных заранее, так и формируемых в процессе обработки. Сами слова при вводе в РНС представляются соотношенными в пространстве и времени наборами единичных импульсов. В состав нейросетевой системы, реализующей прототип предлагаемого метода, входят две одинаковые потоковые РНС, а

также блок управления сетями [5, 25]. Схема потоковой РНС включает в себя два слоя, связанные через единичные задержки управляемыми синапсами, а также собственный блок управления [24]. Эти РНС за счет реализуемых пространственных сдвигов сигналов при передаче от слоя к слою наделяются прозрачными логическими структурами. Сигналы при обработке за счет этих сдвигов продвигаются вдоль слоев. Предусмотрена возможность управления направленностью и параметрами расходимости единичных сигналов при передаче от слоя к слою. Однако не проработаны вопросы, как управлять ассоциативными взаимодействиями сигналов внутри сетей при селекции новостей, их распознавании, восстановлении, прогнозировании и генерации при сохранении непрерывного обучения. Кроме этого не совершенны правила управления самими РНС со стороны внешнего блока управления. Предлагается метод многофункциональной обработки новостей потоковыми РНС и реализующая его система, обеспечивающая преодоление этих недостатков.

Поясним предлагаемый метод на примере нейросетевой системы, реализующей его. Структура этой системы приведена на рисунке 1.

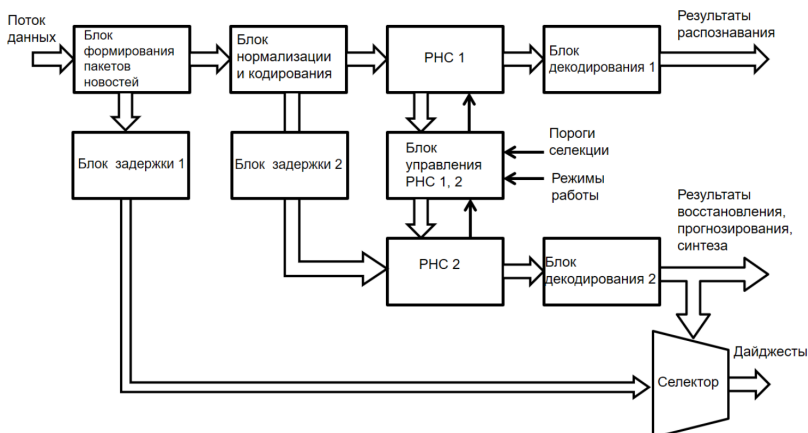


Рис. 1. Структура нейросетевой системы, реализующей метод

В этой системе используются две одинаковые импульсные потоковые РНС с блок-схемой, отраженной на рисунке 2.

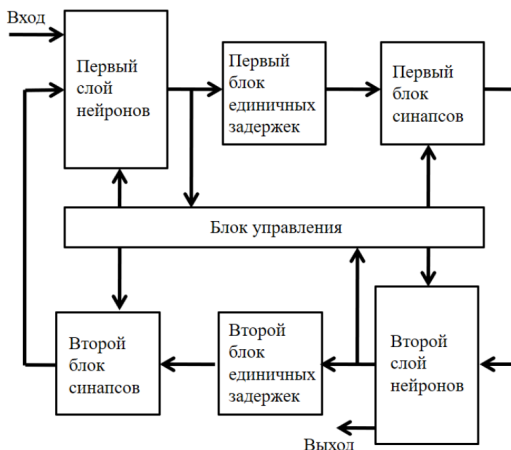


Рис. 2. Блок-схема потоковой РНС с управляемыми элементами

Эти РНС наделяются одной и той же логической структурой. Пример одной из возможных логических структур этих РНС приведен на рисунке 3, где 1, 3 – направления продвижения совокупностей единичных импульсов (SSP) вдоль и между слоями; 2 – линии разделения слоев сети на логические поля за счет реализуемых пространственных сдвигов SSP при передаче от слоя к слою.

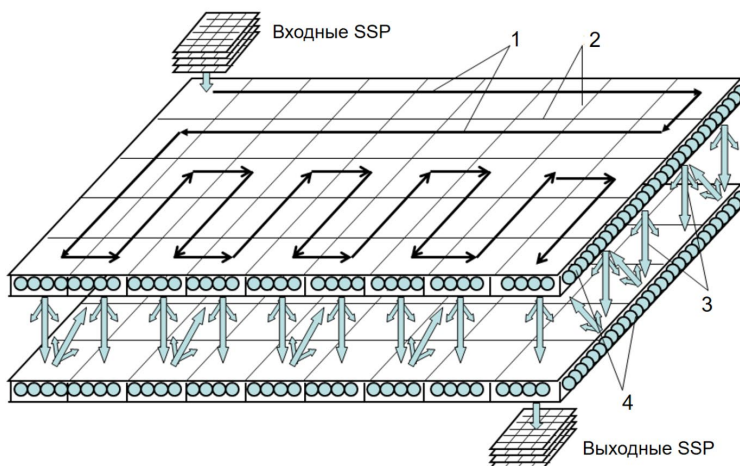


Рис. 3. Пример логической структуры потоковой РНС

Значение потенциала на выходе каждого  $i$ -го импульсного нейрона принимающего слоя РНС определяется как:

$$x_i(t) = \begin{cases} 1, \text{ если } \sum_{j=1}^N x_j(t) \cdot w_{ij}(t) \geq I_0; t_{i0} \geq T_R \\ 0, \text{ в других случаях} \end{cases}.$$

В этом выражении приняты обозначения:

$x_j(t)$  – значения выходных потенциалов нейронов передающего слоя РНС;

$w_{ij}(t)$  – веса синапсов, связывающие  $i$ -й принимающий нейрон с  $j$ -ми передающими нейронами;

$N$  – число нейронов в каждом слое РНС;

$I_0$  – порог возбуждения нейрона;

$t_{i0}$  – время, прошедшее после предыдущего возбуждения  $i$ -го нейрона;

$T_R$  – время невосприимчивости нейронов после возбуждения.

Веса  $w_{ij}(t)$  синапсов представляются в виде:

$$w_{ij}(t) = k_{ij}(t) \cdot \beta_{ij}(t) \cdot \eta_{ij}(t),$$

где  $k_{ij}(t)$  – весовой коэффициент,  $\eta_{ij}(t)$  – функция ослабления сходящихся единичных образов,  $\beta_{ij}(t)$  – функция ослабления расходящихся единичных импульсов, передаваемых от  $j$ -х нейронов к  $i$ -м нейронам,

$$k_{ij}(t) = th(\gamma \cdot g_{ij}(t));$$

$$th(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}};$$

$\gamma$  – коэффициент обучаемости;

$g_{ij}(t)$  – условное число импульсов, прошедших через  $ij$ -й синапс,

$$g_{ij}(t) = g_{ij}(t - \Delta t) \pm \Delta g_{ij}(t);$$

$\Delta g_{ij}(t)$  – приращение  $g_{ij}(t)$ . Это приращение положительное, когда возбуждающий импульс приходит на принимающий нейрон в состоянии ожидания;

$$\beta_{ij}(t) = 1 / (1 + \alpha_{ij} \cdot r_{ij}(t)),$$

$$r_{ij}(t) = ((\Delta x_{ij}(t) + n_{ij}(t)d)^2 + (\Delta y_{ij}(t) + m_{ij}(t)q)^2)^{1/2};$$

$$n_{ij}(t) = \pm 0, 1, \dots, D-1; m_{ij}(t) = \pm 0, 1, \dots, B-1;$$

где  $r_{ij}$  – удаленность связываемых через синапсы нейронов (расстояний между ними на плоскости  $X, Y$ ). Полагается, что расстояние между взаимодействующими слоями нейронной сети стремится к нулю;  $\Delta x_{ij}, \Delta y_{ij}$  – проекции связи  $j$ -го нейрона с  $i$ -м на оси  $X, Y$  без учета пространственных сдвигов;  $d, q$  – величины единичных сдвигов, соответственно, по координатам  $X, Y$ ;  $D, B$  – число, соответственно, столбцов и строк, на которые разбивается каждый слой нейронной сети за счет сдвигов;  $n_{ij}, m_{ij}$  – кратность единичных сдвигов при передаче импульсов от  $j$ -х нейронов к  $i$ -м нейронам. Произведение  $d \times q$  определяет площадь рабочего поля каждого слоя сети. Эта площадь равна числу входящих в поле нейронов.

Стирание информации о единичных импульсах с синапсов осуществимо за счет частичного отражения единичных импульсов от слоев сети.

Варьируя значениями параметров  $\alpha_{ij}$  (коэффициентами ослабления) можно существенно изменять направленность ассоциативного взаимодействия единичных импульсов в сети. В рассматриваемом случае значения  $\alpha_{ij}$  изменяются так, чтобы усиливался вызов сигналов из памяти сети в интересующем направлении.

Принципиальное отличие структуры нейросетевой системы на рисунке 1 от известных решений [5, 25] состоит в том, что дополнительно используются временные задержки, селектор и вводятся новые связи между блоками, а также она наделяется новыми функциями, реализуемыми блоком управления сетями. Селектор предназначен для выделения из обрабатываемых пакетов новостей тех, которые обладают требуемой информационной новизной.



**3.1. Реализуемые функции.** В соответствии с рисунком 1 предлагаемый метод предусматривает формирование пакета новостей, требующих обработки. Каждый новостной текст перед вводом в первую потоковую рекуррентную нейронную сеть (РНС 1) кодируется последовательностью идентификаторов в соответствии с описанным выше методом, и преобразуется в последовательность совокупностей единичных импульсов (СЕИ). После подачи последовательности СЕИ в РНС 1 в ней формируется пространственно-временная модель входных сигналов с учетом ранее запомненной информации. Эта модель постоянно обновляется при обработке СЕИ. Одновременно с непрерывным обучением РНС 1 решается задача классификации текстов, за счет ассоциативного вызова из памяти сети известных атрибутов. Блок управления периодически считывает и анализирует информацию с РНС 1. При таком анализе он распознает типы состояний РНС 1. Выделяются пять состояний в зависимости от уровня и пространственных характеристик загрузки сети СЕИ: 1 – начальное состояние; 2 – равномерной редкой загрузки сети; 3 – полной загрузки; 4, 5 – при загрузке сети в области ее входа и выхода, соответственно. В зависимости от этих состояний или по внешним командам блок управления сетями запускает один из основных режимов работы РНС 2 и самой нейросетевой системы. Это следующие режимы: селекция, восстановление текста, прогнозирование содержания, синтез новостей. После обработки СЕИ в РНС 1, 2 результаты декодируются и преобразуются в текстовую форму. Рассмотрим особенности работы нейросетевой системы в этих режимах.

**3.2. Непрерывное обучение и распознавание.** При поступлении СЕИ в РНС 1 они продвигаются вдоль слоев сети с входа на выход. При таком продвижении между СЕИ в РНС 1 устанавливаются и запоминаются пространственно-временные связи через изменение весов синапсов. Также изменяются состояния импульсных нейронов РНС 1. В обобщенном виде в векторной форме и дискретном времени описание работы РНС 1 можно свести к выражениям (1) – (3):

$$\mathbf{W}_t = \varphi(\mathbf{W}_{t-1} \circ \mathbf{U}_{t-1}, \mathbf{X}_{t-1}, \mathbf{S}_{t-1}^a), \quad (1)$$

$$\mathbf{X}_t = \psi(\mathbf{W}_{t-1} \circ \mathbf{U}_{t-1}, \mathbf{X}_{t-1}, \mathbf{S}_{t-1}^a), \quad (2)$$

$$\mathbf{S}_t^b \subset \mathbf{X}_t, \quad (3)$$

где  $\mathbf{W}_t, \mathbf{W}_{t-1}$  – вектора синапсов на моменты времени  $t$  и  $t-1$ ;  $\mathbf{X}_t, \mathbf{X}_{t-1}$  – вектора состояний нейронов;  $\mathbf{U}_{t-1}$  – управляющие воздействия;  $\circ$  – символ поэлементного умножения значений векторов;  $\mathbf{S}_{t-1}^a, \mathbf{S}_t^b$  – состояния входа и выхода РНС 1.

Заметим, что каждый импульсный нейрон может находиться в одном из следующих состояний: ожидание, возбуждение, временная невосприимчивость. Под распознаванием РНС 1 новостных текстов в нашем случае понимается отнесение их известному классу. В непрерывно обучаемую РНС 1 наряду с текстом  $\mathbf{S}_{t-1}^{a1}$  можно подавать характеризующий его класс  $\mathbf{S}_{t-1}^{a2}, \mathbf{S}_{t-1}^a = \mathbf{S}_{t-1}^{a1} \cup \mathbf{S}_{t-1}^{a2}$ . При обработке текста с учетом накопленного опыта при отсутствии на входе информации о классе РНС 1 может ассоциативно вызывать его из памяти при соответствующем управляющем воздействии.

**3.3. Нейросетевая селекция.** Для перехода в этот режим работы нейросетевой системы помимо нахождения РНС 1 в состоянии полной загрузки должны выполняться условия (4) – (6):

$$W_{\Sigma}(t) - W_{\Sigma}(t - \Delta t) > \Delta W_0, \quad (4)$$

$$\Delta W_{\Sigma}(t) = \sum_{m=1}^2 \sum_{i=1}^N \sum_{j=1}^N \Delta w_{mij}(t), \quad (5)$$

$$\Delta w_{mij}(t) = w_{mij}(t) - w_{mij}(t - \Delta t). \quad (6)$$

Суммарный вес  $W_{\Sigma}(t)$  синапсов РНС 1 на текущий момент времени  $t$  должен превысить аналогичный их вес  $W_{\Sigma}(t-\Delta t)$  на момент  $t-\Delta t$  на заданную величину  $\Delta W_0$ . Этот прирост определяется согласно (5), где  $m$  – номер слоя сети;  $N$  – число нейронов в каждом слое;  $w_{mij}(t), w_{mij}(t-\Delta t)$  – значения весов отдельных синапсов на моменты времени  $t$  и  $t-\Delta t$ . При выполнении условия (4), соответствующего наличию новых элементов в обрабатываемом тексте, блок управления считывает приращения  $\Delta w_{mij}(t)$  весов синапсов из РНС 1 в РНС 2 с умножением их на заданный коэффициент  $\zeta$ . Если пренебречь задержкой на считывание, то весам  $w_{mij}^*(t)$  синапсов РНС 2 присваиваются значения  $\Delta w_{mij}(t) \cdot \zeta$ . Затем через РНС 2 с заданным порогом возбуждения нейронов пропускается блок закодированного текста, вызвавшего отмеченный прирост суммарных весов синапсов. В данном случае РНС 2 реализует интеллектуальную фильтрацию сигналов. В результате на выходе РНС 2 формируются усеченные

СЕИ, несущие только информацию о новых элементах текста. Затем в селекторе выделяется конкретный блок в исходном тексте, как весомое новостное событие.

**3.4. Нейросетевое восстановление.** Восстановление новостных текстов реализуется в РНС 2 только после установления факта, что РНС 1 находится в состоянии загрузки сети в области ее входа. В этом случае блоком управления считывается в РНС 2 информация с РНС 1 о значениях весов синапсов и состояниях нейронов. Затем РНС 2 начинает функционировать в ускоренном времени (режим, при котором за один такт работы РНС 1 в РНС 2 реализуется  $K$  тактов,  $K \gg 1$ ) с усилением ассоциативного вызова сигналов из памяти сети в направлении ее выхода.

С формальной точки зрения этот процесс можно представить в виде (7) – (8):

$$\{\mathbf{W}_t, \mathbf{X}_t\} \rightarrow \{\mathbf{W}_t^*, \mathbf{X}_t^*\} \rightarrow \{\mathbf{W}_t^{*U} = \mathbf{W}_t^* \circ \mathbf{U}_t^*, \mathbf{X}_t^*\}, \quad (7)$$

$$\mathbf{X}_t^* = \psi(\mathbf{W}_t^{*U}, \mathbf{X}_{t-1}^*), \quad \mathbf{S}_t^{*b} \subset \mathbf{X}_t^*. \quad (8)$$

В (7) – (8)  $\mathbf{U}_t^*$  – это вектор управления направленностью ассоциативного взаимодействия сигналов в РНС 2 при их восстановлении;  $\circ$  – символ поэлементного умножения значений векторов; звездочка при векторах синапсов и состояний нейронов означает, что они относятся к РНС 2;  $\psi(\cdot)$  – функция переходов, звездочка при обозначении времени отражает его ускоренное течение. Усиление ассоциативного взаимодействия сигналов в этом режиме можно реализовывать как путем непосредственного изменения весов синапсов, так и за счет изменения функций ослабления расходящихся единичных импульсов, в зависимости от которых рассчитываются эти веса.

Заметим, что восстановление текстовых новостей в рассматриваемом случае сводится к решению обратной задачи анализа.

**3.5. Нейросетевое прогнозирование.** Получение содержания будущих информационных событий на основе анализа текущих событий осуществимо, прежде всего, когда РНС 1 находится в состоянии полной загрузки или частичной загрузки в области входа сети. Для прогнозирования при полной загрузке РНС 1 блоком управления, как в предыдущем режиме, считывается информация с этой сети в РНС 2 о значениях весов синапсов и состоянии нейронов.

Затем РНС 2 запускается в работу в ускоренном времени с усилением ассоциативного вызова сигналов, в отличие от предыдущего режима, в направлении входа сети. В результате сеть вызовет из своей памяти текущими СЕИ связанные сигналы, отражающие содержание будущих новостей. В этом случае прогнозирование сводится только к решению прямой задачи анализа текста. Когда предусматривается прогнозирование будущих новостей при нахождении РНС 1 в состоянии частичной загрузки в области входа сети, предлагается сначала восстанавливать последовательность СЕИ (решать задачу обратного анализа) в РНС 2, а затем реализовывать в ней задачу прямого анализа. Точность прогнозирования будущих новостей может оцениваться стандартными показателями: MAE (mean absolute error), MAPE (mean absolute percentage error), RMSE (root-mean square error).

**3.6. Нейросетевой синтез.** Синтез новостей в РНС 2 осуществим, исходя из следующих условий. Первая РНС должна находиться в состоянии равномерной, но редкой загрузки сети совокупностями единичных импульсов, несущими информацию об обрабатываемых текстах. Также из РНС 1 в РНС 2 должна быть считана информация о значениях весов синапсов и состояниях нейронов. В этом случае РНС 2 должна быть запущена на исполнение в ускоренном режиме с равномерным ассоциативным вызовом сигналов из ассоциативной памяти. При таком вызове информации из памяти сети редкие СЕИ дополняются связанными с ними по содержанию сигналами и сформируется корректная последовательность слов, составляющая новость. В итоге посредством анализа состояний РНС 1 и управления ассоциативным вызовом сигналов из памяти РНС 2 реализуемы, как задачи нейросетевого синтеза, так прямая и обратная задачи анализа новостных текстов.

**4. Результаты и обсуждение.** Оценим возможности предложенного метода по решению на одной и той же структуре нейросетевой системы задач селекции, распознавания, восстановления, прогнозирования и синтеза новостей в виде текстов.

**4.1. Исходные данные.** Для проведения экспериментов был сформирован тестовый набор частично-упорядоченных по времени опубликования текстов новостей из русскоязычных новостных агрегаторов (с сайтов <http://finam.ru>, <http://rbc.ru>, <http://mail.ru> и статьи с сервиса <http://zen.yandex.ru>) без предварительной селекции по тематическим доменам. Они составили 300000 текстов без дублирования. Для подготовки датасетов осуществлялся полный сбор всех доступных страниц с ресурса с последующим анализом, фильтрацией и индексацией. Все собранные из открытых источников

документы преобразовывались из исходного формата (HTML, XML, DOC, PDF, ODT) в формат простого текста без разметки, служебных областей и рекламных материалов. Они нормализовались по форматированию (удалялись лишние служебные символы) и приводились к единой кодовой таблице UTF-8.

**4.2. Особенности потоковых РНС.** В первом варианте метод реализовывался программной нейросетевой системой, включающей две одинаковые потоковые двухслойные РНС со следующими характеристиками. Число нейронов в каждом слое РНС составляло 9600 единиц. Слои разбивались на логические поля с размером  $60 \times 20 = 1200$  нейронов. Реализовывалась линейная схема продвижение СЕИ вдоль слоев сети. Во втором варианте применялись потоковые РНС с размерами слоев по 28800 нейронов. Каждый слой содержал восемь логических полей по 3600 нейронов. Сигналы в виде СЕИ продвигались вдоль слоев по спиральной схеме. В третьем варианте число нейронов в каждом слое РНС равнялось 1890 единиц. Слои разбивались на логические поля с размером  $6 \times 7$  нейронов. Продвижение СЕИ вдоль слоев осуществлялось по петлевой схеме. Рассматривалось непосредственное кодирование слов перед вводом их в РНС и кодирование связей между ними.

**4.3. Результаты селекции новостей.** С использованием первого варианта нейросетевой системы, реализующей метод, проводилась серия из десяти экспериментов по обнаружению новизны обрабатываемых текстовых фрагментов по сравнению с уже обработанными новостями. В каждом эксперименте новостной поток состоял из двух тематических групп по 75 текстов. Общая длина последовательности составляла 150 текстов. Оценивались приросты суммарных весов синапсов при обработке таких последовательностей и ошибки обнаружения факта новизны обрабатываемого текста. Использовался пороговый обнаружитель. Результаты приведены на рисунках 4, 5 и в таблице 1.

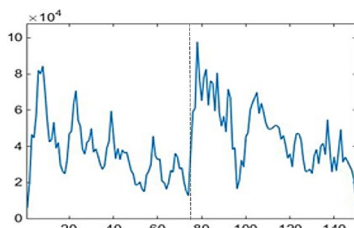


Рис. 4. Изменение суммарных весов синапсов во времени при смене тематики обрабатываемых текстов

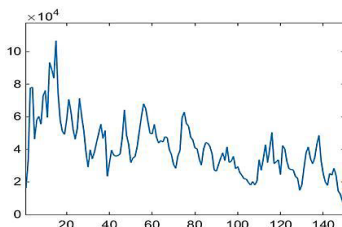


Рис. 5. Изменение суммарных весов синапсов при обработке текстов со смешанной тематикой

Таблица 1. Временные задержки ( $\Delta T$ ) и ошибки определения новизны при различных порогах обнаружения (Thr)

№ эксперимента	Thr $\times 1.8$		Thr $\times 2.0$		Thr $\times 2.2$	
	$\Delta T$	Ошибка	$\Delta T$	Ошибка	$\Delta T$	Ошибка
1	5	0	7	0	-	1
2	-	1	-	1	-	1
3	2	0	3	0	3	0
4	0	0	1	0	1	0
5	5	0	5	0	-	1
6	2	0	2	0	2	0
7	5	0	8	0	9	0
8	0	0	0	0	0	0
9	2	0	2	0	2	0
10	2	0	2	0	2	0
Среднее	2.55	0.1	3.33	0.1	2.71	0.3

Установлено, что при смене тематической группы обрабатываемых текстов резко скачкообразно изменяется суммарный вес синапсов РНС (рисунок 4). Согласно рисунку 4 смена тематики текста произошла на 75 такте работы РНС, а всплеск суммарного веса синапсов проявился на 78 такте. В случаях перемешивания тематик в текстах таких резких скачков не наблюдается (рисунок 5). Он проявляется лишь при начальном вводе смешанного текста в РНС. Установлено, что возможности выявления факта новизны обрабатываемых текстов значительно зависят от принятого абсолютного или относительного порога ее обнаружения (таблица 1). В таблице 1 приведены ошибки второго рода – пропуска событий. С увеличением порога они растут. При понижении порога растут

ошибки первого рода – ложные срабатывания. Порог может задаваться на основе ранее накопленного опыта и требований к безошибочности выявления фактов новизны. За счет фильтрации во второй РНС текстов по новым связям и последующей селекции новых блоков текста в исходном новостном потоке удавалось формировать новостные дайджесты.

**4.4. Результаты прогнозирования.** С использованием второго варианта реализации нейросетевой системы, реализующей метод, проводились эксперименты по управляемому прогнозированию содержания, восстановлению и синтезу новостей. При этом оценивались возможности такого прогнозирования содержания новостей в зависимости от ширины кодирования связей между словами обрабатываемого текста (Таблица 2). Установлено, чем шире учет связей между словами, тем точнее результаты прогнозирования. Однако горизонт такого прогнозирования с приемлемой точностью из-за большого числа анализируемых событий и быстро меняющейся обстановки не велик. При горизонте прогнозирования равном 2 часам его точность в эксперименте 1 составляет 67.4 %, а при 4 часах - не превышает 53.8 %. В этих экспериментах период прогнозирования равнялся 2 часам. При периоде прогнозирования 30 мин. с горизонтом 2 часа ошибки могут не превышать 23 %. Для такой оценки использовались стандартные показатели MAE, MAPE, RMSE с учетом правил, отраженных в [5].

Таблица 2. Результаты прогнозирования содержания новостей

Индикатор	Эксперимент 1 с учетом широких связей		Эксперимент 2 с учетом только смежных связей	
	2 часа	4 часа	2 часа	4 часа
Среднее число связей	908.0	881.5	329.8	317.3
MAE	295.7	407.5	137.9	223.2
MAPE	32.6	46.2	41.8	70.4
RMSE	342.3	456.3	175.2	263.3

Также оценивались возможности прогнозирования новостных событий предлагаемым методом при высокоуровневом кодировании входных потоков. Для этого использовался третий вариант реализации нейросетевой системы с петлевой схемой продвижения СЕИ вдоль слоев. При этом каждое слово в новостном потоке кодировалось своим

единичным импульсом. В РНС 1 вводились последовательности непересекающихся СЕИ.

В частности, были обработаны последовательности закодированных слов “*Two roads diverged in a yellow wood, and sorry I could not travel both and be one traveler, long I stood and looked down one as far as I could to where it bent in the undergrowth*” из произведения “*The Road Not Taken*” Роберта Ли Фроста. Эти последовательности слов, закодированные соответствующими СЕИ, отражались на первом слое РНС 1, как показано на рисунке 6. Стрелками на рисунке 6 обозначены направления продвижения СЕИ вдоль слоев.

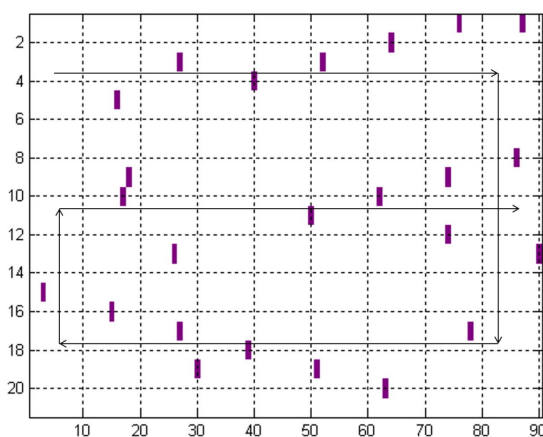


Рис. 6. Состояние первого слоя РНС 1 после введения в нее анализируемой закодированной последовательности слов

После однократного восприятия РНС 1 данных в нее вводился ограниченный набор “*Two roads diverged in a yellow wood, and sorry I could not travel both*” этих же закодированных слов. Затем информация о состоянии РНС 1 считывалась в РНС 2, и на ней путем ускоренного ассоциативного вызова сигналов из памяти осуществлялось прогнозирование последующих слов.

Пример состояния первого слоя RNN 2 перед таким вызовом показан на рисунке 7а. Результат этого вызова сигналов отражен на рисунке 7б. Он выделен штрихпунктирными линиями.



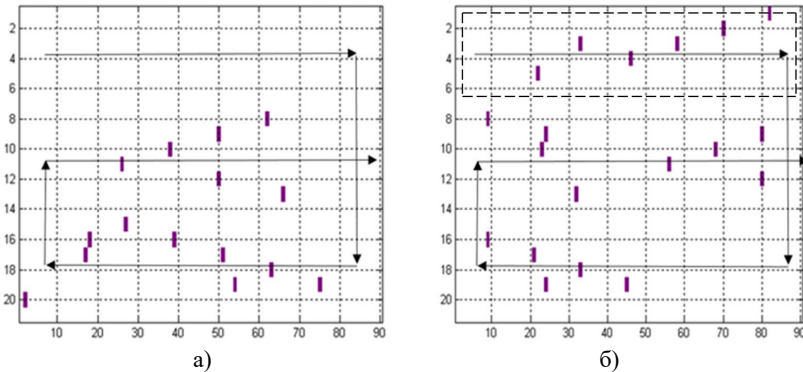


Рис. 7. Результаты прогнозирования закодированной последовательности слов: а) состояние первого слоя РНС 2 до ассоциативного вызова СЕИ из памяти; б) состояние этого слоя после такого вызова

По этому результату после расшифровки из памяти РНС 2 вызывалась полная последовательность. Состояние первого слоя РНС 2 на рисунке 7б отражает лишь часть такого вызова. Выделенный штрихпунктирными линиями прогноз на рисунке 7б соответствует словам *“be one traveler, long I stood”*. Установлено, что отсутствие не только пересечения между обрабатываемыми СЕИ, но и их повторений позволяет исключать ложные выбросы. Данные результаты соответствуют не ожидаемому прогнозированию информационных событий, а соответствуют предиктивному формированию последующих языковых единиц на основе выявленных РНС языковых особенностей.

**4.5. Результаты синтеза и восстановления.** В рамках исследования проводились эксперименты по управляемому восстановлению и синтезу новостей в рамках предложенного метода при втором варианте его реализации. В интересах оценки возможностей восстановления пропущенных элементов текста нейросетевой системой с накопленным опытом в нее вводились усеченные совокупности связей между словами. Путем наблюдения за состояниями слоев РНС 2 оценивались устраненные пропуски этих связей. Напомним, каждая связь между словами, кодируется соотношенным в пространстве и времени единичным импульсом. Восстановление каждой такой связи дает эффект, равный извлечению из ассоциативной памяти сети сразу двух связанных слов. Результаты исследования показали, что уровень восстановления закодированных связей сильно зависит от числа активированных нейронов РНС и

значений весов синапсов с учетом реализуемого их усиления. В ограниченных размерах слоев РНС отношение числа восстанавливаемых связей по отношению к числу связей в неискаженном варианте в экспериментах составило 0.4. Заметим, что слишком сильное усиление ассоциативного вызова сигналов из памяти в направлении входа РНС при большом числе уже активированных нейронов может приводить к ложным вызовам. С учетом этого уровень усиления ассоциативного вызова сигналов из памяти должен согласовываться с уровнем активизации РНС. Примерно такие же оценки возможностей РНС по синтезу новостей были получены, как и при восстановлении. Однако наблюдалась одна особенность. Если при нейросетевом восстановлении новостей энергия активных нейронов направлялась в локализованное пространство, то при синтезе она распределялась во всем пространстве принимающих слоев РНС. В этом случае к усилению ассоциативного вызова сигналов из памяти РНС должны предъявляться другие требования.

**5. Заключение.** В результате выполненного исследования разработан метод многофункциональной обработки новостей потоковыми рекуррентными нейронными сетями с непрерывным обучением. Предложена реализующая его нейросетевая система.

Установлены новые закономерности в поведении этих сетей при ассоциативной обработке новостей, в том числе проявление фактов новизны в сообщениях через изменения суммарных весов синапсов в РНС. Разработанный метод в отличие от известных подходов обладает расширенными функциональными возможностями по решению различных задач анализа и синтеза новостных текстов при непрерывном обучении. Творческие задачи селекции, распознавания, восстановления, прогнозирования и синтеза новостей могут решаться на РНС с одной и той же структурой. Это достигается за счет развития правил ассоциативного взаимодействия сигналов в РНС и управления самими нейронными сетями. Для решения каждой из этих задач не требуется разрабатывать свою нейросетевую систему. Предложен критерий новизны текстов, обрабатываемых в РНС. Реализующая предлагаемый метод нейросетевая система отличается от известных решений новыми элементами, связями между ними, а также выполняемыми функциями. Используемые в ее составе потоковые рекуррентные нейронные сети с управляемыми элементами в отличие от сети GRU и LSTM не требуют обучения с учителем. По результатам экспериментов подтверждена расширенная функциональность метода. Выявлены новые особенности обработки новостных текстов потоковыми РНС. Отражено влияние на

показатели такой обработки как текущих параметров сетей, так и различных вариантов кодирования элементов текста при вводе в РНС. Предлагаемые решения могут найти применение при создании интеллектуальных систем нового поколения не только для обработки текстов, но и других видов информации.

### Литература

1. Wu. Y. Language independent web news extraction system based on text detection framework. *Information Sciences*. 2016. vol. 342. pp.132-149.
2. Kaur G., Bajaj K. News Classification and Its Techniques: A Review. *IOSR Journal of Computer Engineering*. 2016. vol. 18. no. 1. pp. 22-26.
3. Zhang H., Boons F., Riza B.-N. Whose story is it anyway? Automatic extraction of accounts from news articles. *Information Processing & Management*. 2019. vol. 56 . no. 5. pp. 1837-1848.
4. Papagiannopoulou E., Tsoumakas G. Local word vectors guiding keyphrase extraction. *Information Processing & Management*. 2018. vol. 54. no. 6. pp. 888-902.
5. Osipov V., Kuleshov S., Zaytseva A., Levonevskiy D., Miloserdov D. Neural network forecasting of news feeds. *Expert systems with applications*. 2021. vol. 169. p. 114521.
6. Ji Z., Chen K., Wang H. Step-Wise Hierarchical Alignment Network for Image-Text Matching. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAAI-2021)*. pp. 765- 771.
7. Rivera-Trigueros I. Machine translation systems and quality assessment: a systematic review. *Language Resources & Evaluation* 10 April, 2021.
8. Chaudhary and Bali. EASTER: Simplifying Text Recognition using only 1D Convolutions. *The 34th Canadian Conference on Artificial Intelligence, Vancouver, 2021*.
9. Grosman J., Furtado P., Rodrigues A., Schardong G., Barbosa S., Lopes H. Eras: Improving the quality control in the annotation process for natural language processing tasks. *Information Systems*. 2020. vol. 93. p. 101553.
10. Ashari A., Riasetiawan M. Document summarization using TextRank and semantic network. *International journal intelligent systems and applications*. 2017. vol. 1. pp. 26-33.
11. Mele I., Bahrainian S., Crestani F. Event mining and timeliness analysis from heterogeneous news streams. *Information Processing and Management*. 2019. vol. 56. pp. 969–993.
12. Curiskis S., Drake B., Osborn T., Kennedy P. An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management*. 2019. p. 102034.
13. Rezaeinia S., Rahmani R., Ghodsi A., Veisi H. Sentiment Analysis Based on Improved Pre-Trained Word Embeddings. *Expert Systems with Applications*. 2019. vol. 117, pp. 139-147.
14. Hemmatian F., Sohrabi M. A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*. 2019. vol. 52. pp. 1495–1545.
15. Li J., Chen W., Gu B., Fang J., Li Z., Zhao L. Measuring semantic relatedness with knowledge association network. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2019. vol. LNCS 11446. pp. 676-691.

16. Ardia D., Bluteau K., Boudt K. Questioning the news about economic growth: Sparse forecasting using thousands of newsbased sentiment values. *International Journal of Forecasting*. 2019. vol. 35 (4). pp. 1370 – 1386.
17. Sutskever I., Martens J., Hinton G. Generating Text with Recurrent Neural Networks. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, 2011*. pp. 1017-1024.
18. Widodo A., Naomi N., Purnomo F. Prediction of Research Topics Using Combination of Machine Learning and Logistic Curve. *Journal of Theoretical and Applied Information Technology*. 2013. vol. 49 (3). pp. 725 – 732.
19. Cabana A., Mizraji E., Valle-Lisboa J. A neural model that implements probabilistic topics. *Neurocomputing*. 2016. vol. 171. pp. 1099-1107.
20. Wang R., Zhou D., He Y. ATM: Adversarial-neural Topic Model. *Information Processing & Management*. 2019. vol. 56(6). p. 102098.
21. Wei Wei, Guo C. A text semantic topic discovery method based on the conditional co-occurrence degree. *Neurocomputing*. 2019. vol. 368. pp. 11 – 24.
22. Mukhina K., Visheratin A., Nasonov D. Urban events prediction via convolutional neural networks and Instagram data. *Procedia Computer Science*. 2019. vol. 156. pp. 176-184.
23. Recurrent Neural Network Tutorial (2015), Part 4 – Implementing a GRU/LSTM RNN with Python and Theano – WildML. <https://dennybritz.com/posts/wildml/recurrent-neural-networks-tutorial-part-4/>.
24. Osipov V., Osipova M. Space-time signal binding in recurrent neural networks with controlled elements. *Neurocomputing*. 2018. vol. 308. pp. 194–204.
25. Osipov V., Nikiforov V., Zhukova N., Miloserdov D. Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers. *Neural Computing and Applications*. 2020. vol. 32. pp. 14885-14897.
26. Осипов В.Ю. Ассоциативная интеллектуальная машина. Информационные технологии и вычислительные системы. 2010. № 2. С. 59 – 67.
27. Osipov V., Zhukova N., Subbotin A., Glebovskiy P., Evnevich E. Intelligent escalator passenger safety management. *Scientific reports*. 2022. vol. 12. pp. 5506.
28. Zhang J. El-Gohary N.M. Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking. *Journal of Computing in Civil Engineering*. 2016. vol. 30(2). pp. 1–14. doi: 10.1061/(ASCE)CP.1943-5487.0000346.
29. Kuleshov S., Zaytseva A., Aksenov A. Natural Language Search and Associative-Ontology Matching Algorithms Based on Graph Representation of Texts. In: Silhavy, R., Silhavy, P., Prokopova, Z. (eds) *Intelligent Systems Applications in Software Engineering*. *Advances in Intelligent Systems and Computing*. Springer, Cham, 2019. vol 1046. pp. 285–294. doi 10.1007/978-3-030-30329-7\_26.

**Осипов Василий Юрьевич** — д-р техн. наук, профессор, директор, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук. Область научных интересов: математическое моделирование, интеллектуальные системы, нейронные сети, информационная безопасность. Число научных публикаций — 170. osipov\_vasilyu@mail.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812) 328-0887.

**Кулешов Сергей Викторович** — д-р техн. наук, профессор РАН, главный научный сотрудник, лаборатория автоматизации научных исследований, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук. Область научных интересов: ассоциативно-онтологический подход к анализу интернет-контента,

цифровые программно-определяемые инфокоммуникационные системы, обработка изображений и видеоданных, сжатие данных, обработка текстов, поисковые системы. Число научных публикаций — 140. kuleshov@iiias.spb.su; 14 линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812) 323-5139.

**Милосердов Дмитрий Игоревич** — руководитель группы, центр системного анализа и моделирования АО НТЦ РЭБ. Область научных интересов: нейронные сети, моделирование, обработка потоков данных, прогнозирование. Число научных публикаций — 12. dmmil94@yandex.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0887.

**Зайцева Александра Алексеевна** — канд. техн. наук, старший научный сотрудник, лаборатория автоматизации научных исследований, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук. Область научных интересов: методы и технологии обработки больших данных, обработка текстов, поисковые системы. Число научных публикаций — 98. cher@iiias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)323-5139.

**Аксенов Алексей Юрьевич** — канд. техн. наук, старший научный сотрудник, лаборатория автоматизации научных исследований, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук. Область научных интересов: цифровая обработка сигналов, методы обработки и компрессии 3D-данных, в том числе полученных с помощью 3D-сканеров, поисковые системы. Число научных публикаций — 51. a\_aksenov@iiias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)323-5139.

**Поддержка исследований.** Работа выполнена при финансовой поддержке АНО «Аналитический центр при Правительстве Российской Федерации» (ИГК 000000D730321P5Q0002), соглашение № 70-2021-00141, а также бюджетной темы № FFZF-2022-0005.

V. OSIPOV, S. KULESHOV, D. MILOSERDOV, A. ZAYTSEVA, A. AKSENOV  
**RECURRENT NEURAL NETWORKS WITH CONTINUOUS  
LEARNING IN PROBLEMS OF NEWS STREAMS  
MULTIFUNCTIONAL PROCESSING**

*Osipov V., Kuleshov S., Miloserdov D., Zaytseva A., Aksenov A.* **Recurrent Neural Networks with Continuous Learning in Problems of News Streams Multifunctional Processing.**

**Abstract.** The main task of using neural networks is the prompt and accurate solution of various creative tasks, including the analysis and synthesis of news flows, while maintaining the continuity of learning. The result of such processing can be digests, filtered news streams, as well as event forecasts that allow for proactivity in management decisions. Known methods of news processing by neural networks and technical solutions that implement them do not fully provide a solution to the problems that arise in this area. It is necessary to expand their functionality, and improve the space-time signal binding in recurrent neural networks. When processing news flows, simultaneously with continuous training of recurrent neural networks, selection, recognition, restoration, prediction and synthesis of news should be carried out. To reduce the severity of the problem, a promising method of multifunctional processing of news flows is proposed using recurrent neural networks with a logical organization of layers and continuous learning. The method is based on the development of associative processing of textual information in streaming recurrent neural networks with controlled elements. The key features of this method are the multifunctional processing of information flows with changing laws of news appearance. The method provides for operational selection, recognition, restoration, forecasting and synthesis of news based on deep associative continuous processing of links between text elements. The neural network system that implements the proposed method differs from the known solutions by new elements, connections between them, as well as by the functions performed. The results of the experiments confirmed the extended functionality of the method. New features of processing news texts by streaming RNNs are revealed. The proposed solutions can be used to create a new generation of intelligent systems not only for word processing but also for other types of information.

**Keywords:** recurrent neural networks, intelligent news processing, multifunctionality, continuity of learning.

## References

1. Wu. Y. Language independent web news extraction system based on text detection framework. *Information Sciences*. 2016. vol. 342. pp. 132-149.
2. Kaur G., Bajaj K. News Classification and Its Techniques: A Review. *IOSR Journal of Computer Engineering*. 2016. vol. 18. no. 1. pp. 22-26.
3. Zhang H., Boons F., Riza Batis-ta-Navarro. Whose story is it anyway? Automatic extraction of accounts from news articles. *Information Processing & Management*. 2019. vol. 56. no. 5. pp. 1837-1848.
4. Papagiannopoulou E., Tsoumakas G. Local word vectors guiding keyphrase extraction. *Information Processing & Management*. 2018. vol. 54. no. 6. pp. 888-902.
5. Osipov V., Kuleshov S., Zaytseva A., Levonevskiy D., Miloserdov D. Neural network forecasting of news feeds. *Expert systems with applications*. 2021. vol. 169. p. 114521.
6. Ji Z., Chen K., Wang H. Step-Wise Hierarchical Alignment Network for Image-Text Matching. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-2021)*. pp. 765- 771.

7. Rivera-Trigueros I. Machine translation systems and quality assessment: a systematic review. *Language Resources & Evaluation*, 10 April, 2021.
8. Chaudhary and Bali. EASTER: Simplifying Text Recognition using only 1D Convolutions. *The 34th Canadian Conference on Artificial Intelligence*, Vancouver, 2021.
9. Grosman J., Furtado P., Rodrigues A., Schardong G., Barbosa S., Lopes H. Eras: Improving the quality control in the annotation process for natural language processing tasks. *Information Systems*. 2020. vol. 93. p. 101553.
10. Ashari A., Riaseitawan M. Document summarization using TextRank and semantic network. *International journal intelligent systems and applications*. 2017. vol. 1. pp. 26-33.
11. Mele I., Bahrainian S., Crestani F. Event mining and timeliness analysis from heterogeneous news streams. *Information Processing and Management*. 2019. vol. 56. pp. 969–993.
12. Curiskis S., Drake B., Osborn T., Kennedy P. An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management*. 2019. pp. 102034.
13. Rezaeinia S., Rahmani R., Ghodsi A., Veisi H. Sentiment Analysis Based on Improved Pre-Trained Word Embeddings. *Expert Systems with Applications*. 2019. vol. 117. pp. 139-147.
14. Hemmatian F., Sohrabi M. A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*. 2019. vol. 52. pp. 1495–1545.
15. Li J., Chen W., Gu B., Fang J., Li Z., Zhao L. Measuring semantic relatedness with knowledge association network. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2019. vol. LNCS 11446. pp. 676-691.
16. Ardia D., Bluteau K., Boudt K. Questioning the news about economic growth: Sparse forecasting using thousands of newsbased sentiment values. *International Journal of Forecasting*. 2019. vol. 35(4). pp.1370 – 1386.
17. Sutskever I., Martens J., Hinton G. Generating Text with Recurrent Neural Networks. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, 2011*. pp. 1017-1024.
18. Widodo A., Naomi N., Purnomo F. Prediction of Research Topics Using Combination of Machine Learning and Logistic Curve. *Journal of Theoretical and Applied Information Technology*. 2013. vol. 49. no. 3. pp. 725 – 732.
19. Cabana A., Mizraji E., Valle-Lisboa J. A neural model that implements probabilistic topics. *Neurocomputing*. 2016. vol. 171. pp. 1099-1107.
20. Wang R., Zhou D., He Y. ATM: Adversarial-neural Topic Model. *Information Processing & Management*. 2019. vol. 56(6). p. 102098.
21. Wei Wei, Guo C. A text semantic topic discovery method based on the conditional co-occurrence degree. *Neurocomputing*. 2019. vol. 368. pp. 11 – 24.
22. Mukhina K., Visheratin A., Nasonov D. Urban events prediction via convolutional neural networks and Instagram data. *Procedia Computer Science*. 2019. vol. 156. pp. 176-184.
23. *Recurrent Neural Network Tutorial (2015), Part 4 – Implementing a GRU/LSTM RNN with Python and Theano – WildML*. Available at: <https://dennybritz.com/posts/wildml/recurrent-neural-networks-tutorial-part-4/>. (accessed 27.10.2022).
24. Osipov V., Osipova M. Space-time signal binding in recurrent neural networks with controlled elements. *Neurocomputing*. 2018. vol. 308. pp. 194–204.

25. Osipov V., Nikiforov V., Zhukova N., Miloserdov D. Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers. *Neural Computing and Applications*. 2020. vol. 32. pp. 14885-14897.
26. Osipov V.Yu. *Associativnaya intellektual'naya mashina. Informacionnye tekhnologii i vychislitel'nye sistemy*. 2010. no. 2. pp. 59 – 67.
27. Osipov V., Zhukova N., Subbotin A., Glebovskiy P., Evnevich E. Intelligent escalator passenger safety management. *Scientific reports*. 2022. vol. 12. p. 5506.
28. Zhang J. El-Gohary N.M. Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking. *Journal of Computing in Civil Engineering*. 2016. vol. 30. no. 2. pp. 1–14. doi: 10.1061/(ASCE)CP.1943-5487.0000346.
29. Kuleshov S., Zaytseva A., Aksenov A. Natural Language Search and Associative-Ontology Matching Algorithms Based on Graph Representation of Texts. In: Silhavy, R., Silhavy, P., Prokopova, Z. (eds) *Intelligent Systems Applications in Software Engineering. Advances in Intelligent Systems and Computing*. Springer, Cham, 2019. vol. 1046. pp. 285–294. doi 10.1007/978-3-030-30329-7\_26.

**Osipov Vasily** — Ph.D., Dr.Sci., Professor, Director, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: mathematical modeling, intelligent systems, neural networks, information security. The number of publications — 170. osipov\_vasily@mail.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812) 328-0887.

**Kuleshov Sergey** — Ph.D., Dr.Sci., Professor of the RAS, Chief researcher, Laboratory of research automation, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: associative-ontological approach to the analysis of Internet content, digital software-defined infocommunication systems, image and video data processing, data compression, text processing, search engines. The number of publications — 140. kuleshov@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812) 323-5139.

**Miloserdov Dmitry** — Head of the group, Center for system analysis and modeling JSC STC EW. Research interests: neural networks, modeling, data flow processing, forecasting. The number of publications — 12. dmmil94@yandex.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0887.

**Zaytseva Alexandra** — Ph.D., Senior researcher, Laboratory of research automation, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: methods and technologies for processing big data, word processing, search engines. The number of publications — 98. cher@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(921)318-0136.

**Aksenov Alexey** — Ph.D., Senior researcher, Laboratory of research automation, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: digital signal processing, methods of processing and compression of 3D data, including, search engines. The number of publications — 51. a\_aksenov@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)323-5139.

**Acknowledgements.** This work was supported by the Analytical Center for the Government of the Russian Federation (IGK 000000D730321P5Q0002), agreement No. 70-2021-00141, and by the Budget № FFZF-2022-0005.