

В.С. Тормозов

аспирант

(Тихоокеанский государственный университет)

г. Хабаровск, Российская Федерация

А.Л. Золкин

канд. техн. наук

(Самарский филиал Волжского государственного университета водного транспорта)

г. Самара, Российская Федерация

К.А. Василенко

(Колледж сервиса и дизайна при Владивостокском университете экономики и сервиса)

г. Владивосток, Российская Федерация

Настройка, обучение и тестирование нейронной сети долгой краткосрочной памяти для задачи распознавания образов

Статья посвящена искусственным нейронным сетям прямого распространения с блоками долгой краткосрочной памяти (ДКП). Выполнено исследование способностей этой нейронной сети по распознаванию разрозненных образов. В статье предлагается линейное переключение блоков ДКП в зависимости от значения проходящего сигнала. Выполнено сравнение полученных результатов с результатами работы нейронной сети прямого распространения аналогичной структуры.

Ключевые слова: блок долгой краткосрочной памяти; искусственный интеллект; искусственная нейронная сеть; машинное обучение; распознавание образов.

V.S. Tormozov

Postgraduate Student

(Pacific State University)

Khabarovsk, Russian Federation

A.L. Zolkin

Cand. of Techn. Sciences

(Samara branch of Volga State University of Water Transport)

Samara, Russian Federation

K.A. Vasilenko

(College of Service and Design at the Vladivostok University of Economics and Service)

Vladivostok, Russian Federation

Configuring, Training and Testing a Long-term Short-term Memory Neural Network for Pattern Recognition

The article is devoted to artificial neural networks of direct distribution with blocks of long short-term memory (LSTM). A study of the abilities of this neural network to recognize disparate images has been performed. The article suggests linear switching of LSTM blocks depending on the value of the transmitted signal. The results obtained are compared with the results of the operation of a neural network of direct distribution of a similar structure.

Keywords: LSTM block; artificial intelligence; artificial neural network; machine learning; pattern recognition.

DOI: 10.25791/asu.3.2020.1171

При разработке автоматизированных интеллектуальных систем классификаций, в условиях неопределенности, немаловажную роль играет смысловой контекст решаемой задачи. В том числе к такому контексту относится информация о прошлых состояниях системы. Сегодня высокую популярность для задач

классификации входных образов приобретают нейросетевые математические модели.

Для искусственных нейронных сетей (ИНС) контекст может быть извлечен из вектора поступающих на вход сигналов и из значений весовых коэффициентов обученной сети [1]. Немаловажным преимуществом

нейросетевого подхода является его способность распараллеливания процедуры обучения и тестирования на современных микропроцессорах и GPU [2]. Однако при учете контекста значительного объема растет количество нейронных связей и сложность процедур обучения и работы сети. Вместо поступления контекста от входных значений контекст может также временно храниться в специальном буфере памяти, откуда в дальнейшем он может быть извлечен и использован в качестве сигнала при работе ИНС. Такой вид памяти носит название *долгой краткосрочной памяти* (ДКП), в которой сохраняются предыдущие значения входного вектора.

Информационно-технические методы решения прикладных задач приобретают все более широкую область применения [3, 4]. Для предварительной обработки и фильтрации зашумленных аналоговых сигналов может быть использована ИНС, схема которой представлена на рисунке 1.

Нейроны первого слоя принимают сигналы от блока ДКП и обрабатывают также как и нейроны скрытого слоя. В качестве нелинейной дифференцируемой функцией активизации была выбрано нормальное распределение для нейронов всей сети.

Такая ИНС может быть применена для предсказания последующей последовательности значений синусоидального сигнала с зашумлением. В блоке памяти сохраняются значения десяти предыдущих входных сигналов. Учебная выборка состоит из значений одной синусоиды $\frac{1}{2} + \frac{1}{2} \sin(x_i)$ в диапазоне $[0; 2\pi]$.

Тестовая выборка состоит из значений суммы трех синусоид переменных частот в том же диапазоне (рис. 2). Погрешность обучения достигла 8,5 % после $7 \cdot 10^5$ итераций обучения, в то время как погрешность тестирования составила 16,7 %.

Распознавание образов – это область науки, посвященная разработке принципов, построению программных и (или) аппаратных систем, предназначением которых является определение того, какому из заранее заданных классов принадлежит рассматриваемый объект. Объектом может быть любой сигнал, явление, реально существующий предмет или абстракция и другое.

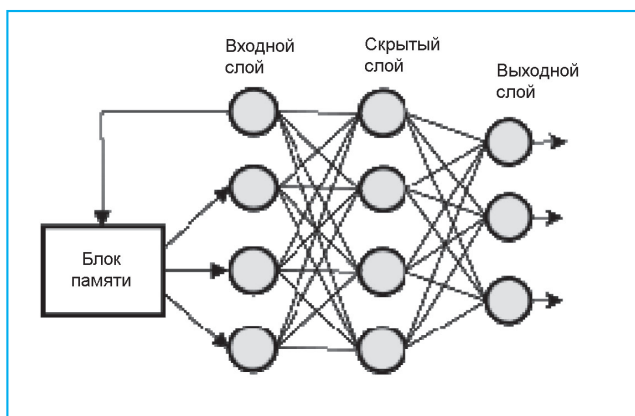


Рис. 1. ИНС прямого распространения с блоком ДКП

Объект определяется конечной последовательностью признаков или свойств. Признаками называются *данные, полученные в результате измерения объекта*, для того чтобы использовать их в процессе классификации. Множество признаков одного объекта образует так называемый *вектор признаков*. Этот вектор признаков может быть отмечен как точка в гиперпространстве признаков. Каждому образу сопоставлен только один вектор признаков, также как и вектору признаков сопоставлен только один образ. Соотнесение объекта одному из классов называется *классификацией*. Классификация выполняется с применением эталонных объектов – объектов, класс которых уже известен [5].

Образ – описание объекта, используемое для решения задачи распознавания образов, соотнесения объекта одному из классов. Для того чтобы было возможно решить задачу распознавания образов как задачу дискриминантного анализа [6], необходимо чтобы все возможные образы объектов разбивались на конечное множество классов. Также должно быть задано множество эталонных объектов. Каждый эталонный объект принадлежит одному из классов. Существует также объект, принадлежность которого какому-либо классу неизвестна. Задача распознавания образов заключается в том, чтобы установить какому классу принадлежит этот объект [0]. Либо разработать алгоритм, выполняющий это для любого рассматриваемого объекта. Чем чаще метод или алгоритм верно определяет принадлежность объекта классу, тем точнее система распознавания образов.

Качество работы системы распознавания может быть оценено с точки зрения точности назначения класса предъявляемым объектам. Различают распознавание с обучением и без обучения. Если распознавание выполняется с применением имеющегося множества эталонных объектов, то задачу распознавания называют *классификацией с обучением*.

Нейросетевой подход с сохранением предыдущих значений в блоках ДКП – очень эффективный инструмент как для задач распознавания, так и для задач прогнозирования временных рядов. Тем не менее, у такого подхода присутствуют ограничения по функциональности, связанные с предопределенностью блока ДКП. Развитием нейросетевой архитектуры с ДКП является ИНС, сохраняющая промежуточные значения проходящих сигналов во внутренних слоях. Такое усовершенствование позволяет сети производить самонастройку с помощью блоков памяти, производящих чтение и запись в долговременную память значений сигнала.

Обзор существующих исследований

Sainath и др. [0] получили хорошие результаты по метрике коэффициента словесных ошибок (КСО), обучая глубокую РНС с ДКП с помощью многомасштабного ввода. Weston [0] реализовал РНС с ДКП для обработки и формирования текстового сценария, в ходе которого множество действующих лиц могут перемещаться между помещениями, забирая и передвигая

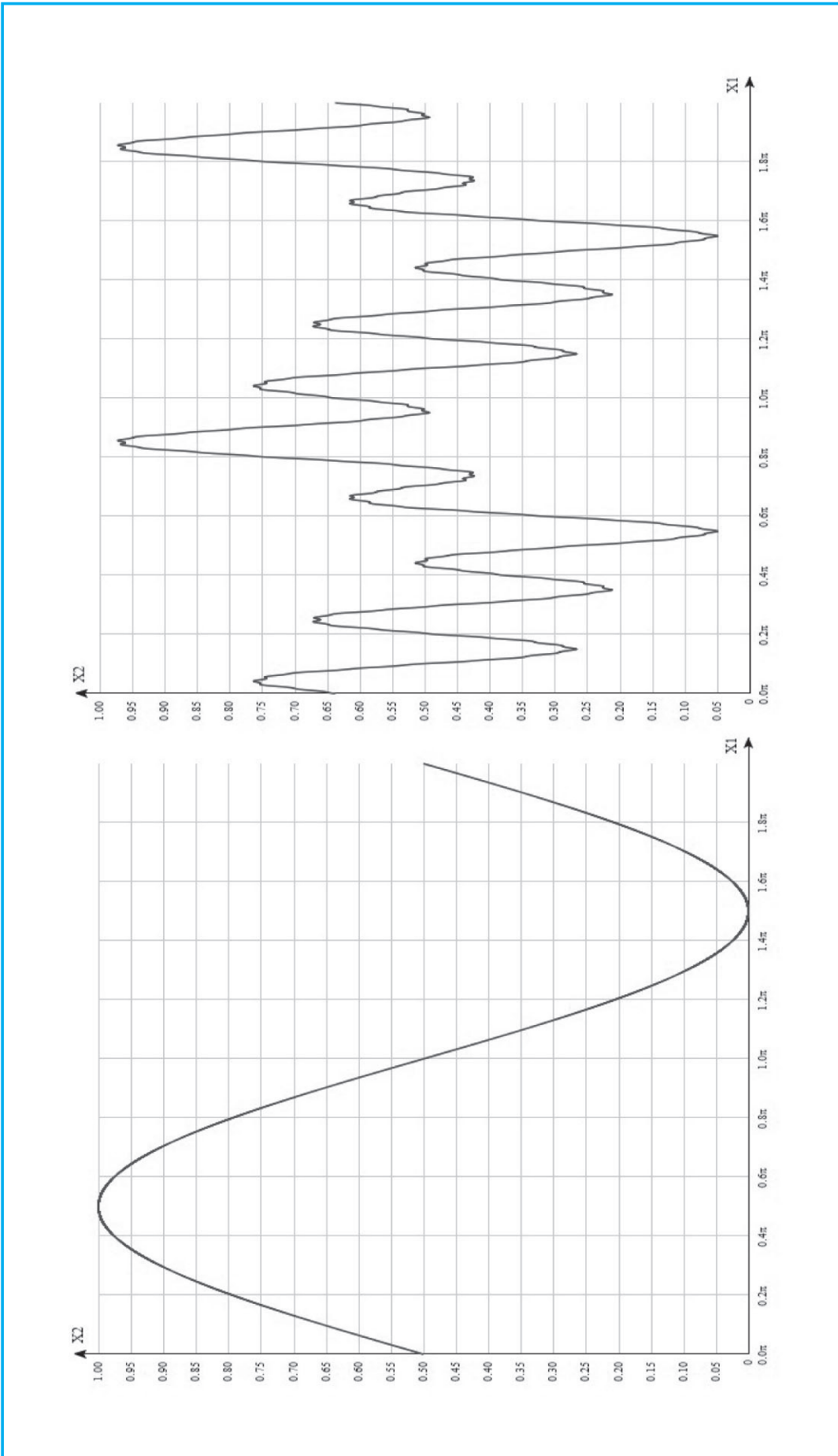


Рис. 2. Обучающая и тестируемая функции ИНС

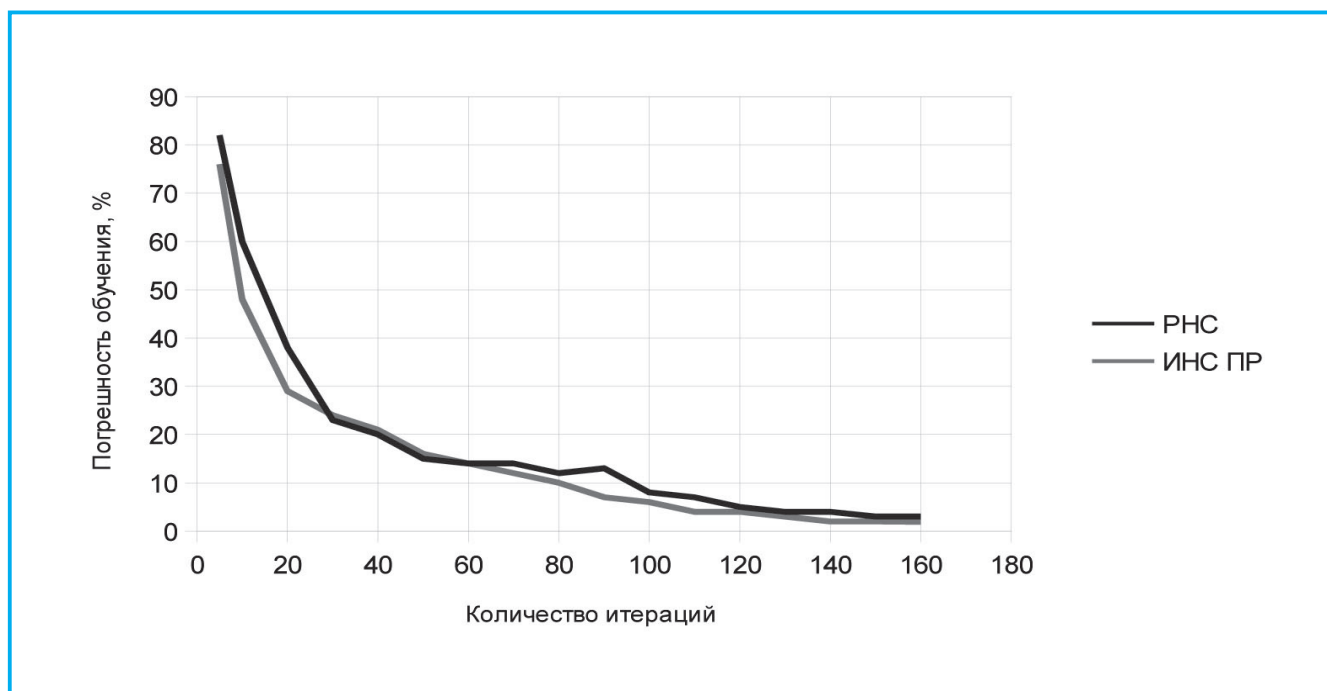


Рис. 3. Графики погрешностей обучения сетей ИНС ПР с блоками ДКП и РНС для задачи классификации

объекты. С внедренными блоками долгосрочной памяти построенная и обученная сеть показала лучшие экспериментальные результаты, по сравнению с обычной РНС при запросах о прошлых отдаленных данных.

Постановка задачи

Архитектура исследуемой нейронной сети включает в себя ИНС прямого распространения с блоком ДКП, связанным с каждым нейроном в каждом из скрытых слоев. Входной и выходной слои составлены только из стандартных нейронов, в то время как скрытые слои составлены из нейронов с подключенными ДКП. Выход каждого нейрона формируется сигмоидальной функцией комбинации его входных сигналов. Выход нейрона работает как линейный переключатель активации подключенного к нему блока ДКП. Нейронная сеть такой архитектуры в процессе обучения приобретает способность извлекать записанный в блоки памяти контекст по мере необходимости.

ИНС с блоками ДКП имеет входной и выходной слои, состоящие из стандартных нейронов и скрытых слоев, которые в свою очередь состоят из стандартных нейронов с подключенными к ним блоками ДКП. Нейроны скрытых слоев имеют сигмоидальную функцию активизации. Каждый нейрон связан с блоком памяти, и каждый блок памяти состоит из множества буферов памяти. Выходы нейронов используются в качестве линейных переключателей, чтобы при необходимости активизировать определенный буфер памяти.

Процедуры обучения и тестирования сети состоят из трех базовых этапов:

- 1) прямое распространение сигналов по сети через каждый слой;
- 2) обучение сети с помощью алгоритма обратного распространения ошибки;
- 3) сохранение значений сигнала памяти в каждом скрытом слое как ортогональный набор в многомерном пространстве.

Вектор состояния вывода для стандартных нейронов каждого слоя увеличивается с каждым проходом сигнала через скрытые слои, содержащие ДКП. Выходной сигнал каждого блока памяти – это содержание того буфера памяти, к которому он соотнесен выходной сигнал подключенного к нему нейрона. Выполняется линейное переключение блока памяти, в зависимости от того, какому отрезку допустимых значений сигнала относится проходящий через нейрон сигнал. При таком подходе нейрон выступает в роли линейного переключателя, который включает конкретный буфер памяти в зависимости от значения выхода данного нейрона. Нейроны скрытого слоя получают входной сигнал от предыдущего слоя, в то время как блоки памяти воспринимают входной сигнал только от нейрона, к которому они подключены [10]. Обучение с использованием алгоритма обратного распространения ошибки для ИНС с блоками памяти аналогично обучению ИНС прямого распространения с использованием этого алгоритма [11].

Каждый нейрон скрытого слоя содержит блок ДКП. Таким образом, в сети сохраняются воспоминания о предыдущих прохождениях сигнала. Каждое уникальное воспоминание в нейроне ортогонально всем другим воспоминаниям, представленным в этом блоке. Последовательность воспоминаний представлена в виде

вектора в многомерном пространстве и называется *составной памятью*. Направление вектора воспоминаний указывает на то, какие воспоминания присутствуют, а значения величины вектора при проекции на оси каждого воспоминания указывают на порядок, в котором они произошли [12]. Вновь занесенные значения в блок памяти задаются ортогональным отображением выходного сигнала подключенного нейрона и соседнего нейрона. Ближайшие соседние нейроны могут быть сформулированы в пространственном смысле: как нейроны того же скрытого слоя с близкими значениями индексов. Ближайшие соседние нейроны несут информацию о состоянии предыдущего слоя.

Результаты исследования

Набор данных состоит из выборки 8-ми случайно сгенерированных многомерных нормальных распределений. Целью работы сети является классификация, из какого распределения была взята заданная точка. Помимо этого были протестированы: ИНС прямого распространения (ИНС ПР) и РНС аналогичной архитектуры, с таким же количеством нейронов на каждом слое. Классификационная погрешность исследуемой РНС составила 1,9 %. Результаты проведенного исследования показаны на рисунке 3.

Выводы

В статье рассмотрена и продемонстрирована работа ИНС ПР, которая сохраняет предыдущие значения сигнала в блоках памяти ДКП, связанных с нейронами скрытых слоев. Такая ИНС показывает хорошие результаты для задач классификации и распознавания образов в сравнении с рекуррентными нейронными сетями (РНС). Преимущество сетей такого типа в том, что они используют блоки памяти, ассоциированные с каждым нейроном скрытого слоя, что позволяет сохранять контекст при формировании шаблонов распознавания.

Список литературы

1. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // *Proceedings of the national academy of sciences*. 1982. Vol. 79, no. 8, pp. 2554–2558.
2. Василенко К.А., Мишин В.В. Архитектура процессоров и микропроцессоров // *V (62) ежегодная научно-практическая конференция преподавателей, студентов и молодых ученых СКФУ «Молодая наука – 2017»*. Пятигорск: Изд-во ПФ СКФУ. 2017, pp. 48–49.
3. Резак Е.В., Барсукова Н.В. Преимущества использования интерактивных электронных учебников в учебном процессе студентов вуза // *IV Всероссийская научно-практическая конференция на английском языке с международным участием «Professional*

english in use». Хабаровск: Изд-во ТОГУ, 2018. С. 195–196.

4. Золкин А.Л. Разработка информационно-управляющей системы для контроля износа коллекторов тяговых электродвигателей // *Вестник Донецкой академии автомобильного транспорта*. 2019, № 2. С. 65–74.
5. Тормозов В.С. Метод детектирования и классификации транспортных средств на спутниковых снимках сверхвысокого разрешения // *Промышленные АСУ и контроллеры*. 2019, № 6. С. 18–24.
6. Федосеев А.А., Фрышкина Е.А. Распознавание образов // *Ученые заметки ТОГУ*. 2018. Т. 9, № 2. С. 475–479.
7. Тормозов В.С. Автоматическое детектирование дорожного покрытия на космических снимках сверхвысокого разрешения // *Молодые ученые – Хабаровскому краю. Материалы XIX краевого конкурса молодых ученых и аспирантов*. Хабаровск: Издательство Тихоокеанского государственного университета. 2017. С. 131–135.
8. Sainath T.N., Vinyals O., Senior A., Sak H. Convolutional, long short-term memory, fully connected deep neural networks // *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2015, pp. 4580–4584.
9. Weston, Jason, Antoine Bordes, Sumit Chopra. Memory networks // *arXiv conference – 2014*.
10. Hochreiter S., Bengio Y., Frasconi P. and Schmidhuber J. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies // *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press. 2001, pp. 114–132.
11. Graves Alex, Ivo Danihelka, Greg Wayne. Neural Turing Machines // *arXiv conference – 2014*.
12. Estebon M.D. Perceptrons: An Associative Learning Network // *Virginia Tech*. 1997.

References

1. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities // *Proceedings of the national academy of sciences*. 1982. Vol. 79, no. 8, pp. 2554–2558.
2. Vasilenko K.A., Mishin V.V. Arkhitektura protsessorov i mikroprotsessorov. *V (62) ezhegodnaya nauchno-prakticheskaya konferentsiya prepodavateley, studentov i molodykh uchenykh SKFU «Molodaya nauka – 2017»* [Architecture of processors and microprocessors. V (62) annual scientific and practical conference of teachers, students and young scientists of SKFU «Young Science – 2017»]. Pyatigorsk: Izd-vo PF SKFU [Pyatigorsk: Publishing house of the Pyatigorsk branch of the North Caucasian Federal University]. 2017, pp. 48–49.
3. Rezak Ye.V., Barsukova N.V. Preimushchestva ispolzovaniya interaktivnykh elektronnykh uchebnikov v uchebном protsesse studentov vuza. *IV Vserossiyskaya nauchno-prakticheskaya konferentsiya na angliyskom*

- yazyke s mezhduнародnym uchastiem «Professional English in Use» [The benefits of using interactive electronic textbooks in the learning process of university students. IV All-Russian Scientific and Practical Conference in English with international participation «Professional English in Use»]. Khabarovsk: Izd-vo TOGU [Khabarovsk: Pacific State University Publishing House]. 2018, pp. 195–196.*
4. Zolkin A.L. Razrabotka informatsionno-upravlyayushchey sistemy dlya kontrolya iznosa kollektorov tyagovykh elektrodvigateley [Development of an information management system for controlling the wear of traction electric motor collectors]. *Vestnik Donetskoy akademii avtomobilnogo transporta* [Bulletin of the Donetsk Academy of Road Transport]. 2019, no. 2, pp. 65–74.
 5. Tormozov V.S. Metod detektirovaniya i klassifikatsii transportnykh sredstv na sputnikovykh snimkakh sverkhvysokogo razresheniya [Method of Detecting and Classifying Vehicles on Ultra-high Resolution Satellite Images]. *Promyshlennyye ASU i kontrolyery* [Industrial Automatic Control Systems and Controllers]. 2019, no. 6, pp. 18–24.
 6. Fedoseev A.A., Fryshkina Ye.A. Raspoznavanie obrazov [Pattern recognition]. *Uchenyye zametki TOGU* [Study Notes Pacific State University]. 2018. Vol. 9, no. 2, pp. 475–479.
 7. Tormozov V.S. Avtomaticheskoe detektirovanie dorozhnogo pokrytiya na kosmicheskikh snimkakh sverkhvysokogo razresheniya. *Molodye uchenyye – Khabarovskomu krayu. Materialy XIX kraevogo konkursa molodykh uchenykh i aspirantov* [Automatic detection of pavement in satellite images of ultra-high resolution. Young Scientists – Khabarovsk Territory. Materials of the XIX regional competition of young scientists and graduate students.]. Khabarovsk: Izdatelstvo Tikhoookeanskogo gosudarstvennogo universiteta [Khabarovsk: Pacific State University Publishing House]. 2017, pp. 131–135.
 8. Sainath T.N., Vinyals O., Senior A., Sak H. Convolutional, long short-term memory, fully connected deep neural networks // *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2015, pp. 4580–4584.
 9. Weston, Jason, Antoine Bordes, Sumit Chopra. Memory networks // *arXiv conference – 2014*.
 10. Hochreiter S., Bengio Y., Frasconi P. and Schmidhuber J. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies // *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press. 2001, pp. 114–132.
 11. Graves Alex, Ivo Danihelka, Greg Wayne. Neural Turing Machines // *arXiv conference – 2014*.
 12. Esteban M.D. Perceptrons: An Associative Learning Network // *Virginia Tech*. 1997.

Информация об авторах

Тормозов Владимир Сергеевич, аспирант кафедры «Вычислительная техника», соискатель ученой степени кандидата технических наук, старший преподаватель кафедры «Программное обеспечение вычислительной техники и автоматизированных систем»

Федеральное государственное бюджетное образовательное учреждение высшего образования «Тихоокеанский государственный университет»

680054, Российская Федерация, г. Хабаровск, ул. Тихоокеанская, д. 136

Золкин Александр Леонидович, кандидат технических наук

Самарский филиал Федерального государственного бюджетного образовательного учреждения высшего образования «Волжский государственный университет водного транспорта»

443099, Российская Федерация, г. Самара, Молодогвардейская ул., д. 62

Василенко Константин Александрович, преподаватель высшей категории

Колледж сервиса и дизайна при Владивостокском университете экономики и сервиса (ВГУЭС)

690092, Российская Федерация, г. Владивосток, ул. Добровольского, 20

Information about the authors

Tormozov Vladimir Sergeevich, Postgraduate Student of the Department «Computer Engineering», Applicant for Degree of the Scientific Degree of Candidate of Technical Sciences, Senior Lecturer of the Department of Computer Software and Automated Systems

Federal State Budgetary Institution of Higher Education «Pacific State University»

680054, Russian Federation, Khabarovsk, Tikhoookeanskaya Str., 136

Zolkin Aleksandr Leonidovich, Candidate of Technical Sciences

Samara branch of the Volga State University of Water Transport of the Federal State Budgetary Educational Institution of Higher Education

443099, Russian Federation, Samara, Molodogvardeiskaya Str., 62

Vasilenko Konstantin Aleksandrovich, Senior Lecturer of the Highest Category

College of Service and Design at the Vladivostok University of Economics and Service

690092, Russian Federation, Vladivostok, Dobrovolskaya Str., 20