

DEEP LEARNING MODEL FOR BILINGUAL SENTIMENT CLASSIFICATION OF SHORT TEXTS

Y.B. Abdullin^{a,b}, V.V. Ivanov^c

^a «Hibrain» LTD, Astana, 010000, Kazakhstan

^b Kazan Federal University, Kazan, 420008, Russian Federation

^c Innopolis University, Innopolis, 420500, Russian Federation

Corresponding author: nomemm@gmail.com

Article info

Received 30.11.16, accepted 23.12.16

doi: 10.17586/2226-1494-2017-17-1-129-136

Article in English

For citation: Abdullin Y.B., Ivanov V.V. Deep learning model for bilingual sentiment classification of short texts. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2017, vol. 17, no. 1, pp. 129–136. doi: 10.17586/2226-1494-2017-17-1-129-136

Abstract

Sentiment analysis of short texts such as Twitter messages and comments in news portals is challenging due to the lack of contextual information. We propose a deep neural network model that uses bilingual word embeddings to effectively solve sentiment classification problem for a given pair of languages. We apply our approach to two corpora of two different language pairs: English-Russian and Russian-Kazakh. We show how to train a classifier in one language and predict in another. Our approach achieves 73% accuracy for English and 74% accuracy for Russian. For Kazakh sentiment analysis, we propose a baseline method, that achieves 60% accuracy; and a method to learn bilingual embeddings from a large unlabeled corpus using a bilingual word pairs.

Keywords

sentiment analysis, bilingual word embeddings, recurrent neural networks, deep learning, Kazakh language

Acknowledgements

This work is supported by the Russian Science Foundation (project 15-11-10019 "Text mining models and methods for analysis of the needs, preferences and consumer behaviour"). The authors thank the Everware team (<https://github.com/orgs/everware/people>) for access to their platform. The authors would like to thank Yerlan Seitkazinov, Zarina Sadykova and Aliya Sitdikova for manual annotation of the Kazakh sentiment corpus.

УДК 28.23.37

МОДЕЛЬ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ДВУЯЗЫЧНОЙ КЛАССИФИКАЦИИ ТОНАЛЬНОСТИ КОРОТКИХ ТЕКСТОВ

Е.Б. Абдуллин^{a,b}, В.В. Иванов^c

^a ТОО «Hibrain», Астана, 010000, Казахстан

^b Казанский федеральный университет, Казань, 420008, Российская Федерация

^c Университет Иннополис, Иннополис, 420500, Российская Федерация

Адрес для переписки: nomemm@gmail.com

Информация о статье

Поступила в редакцию 30.11.16, принята к печати 23.12.16

doi: 10.17586/2226-1494-2017-17-1-129-136

Язык статьи – английский

Ссылка для цитирования: Абдуллин Е.Б., Иванов В.В. Модель глубокого обучения для двуязычной классификации тональности коротких текстов // Научно-технический вестник информационных технологий, механики и оптики. 2017. Т. 17. № 1. С. 129–136. doi: 10.17586/2226-1494-2017-17-1-129-136

Аннотация

Исследованы проблемы классификации коротких текстов (сообщения в Twitter, комментарии из новостных порталов) при недостатке контекстной информации. Предложена модель глубокой нейронной сети, использующей двуязычные векторные представления слов для эффективного решения проблемы классификации тональности текста конкретной пары языков. Предложенный подход применен к двум корпусам двух различных языковых пар: английский-русский и русский-казахский. Показан способ обучения классификатора на одном языке и применения его для предсказания тональности на другом. Предлагаемый подход позволил достичь 73% точности для английского языка и 74% точности для русского языка. Впервые получены результаты анализа тональности на казахском языке с точностью до 60%. Предложен метод создания двуязычных векторных представлений слов из больших размеченных корпусов с использованием словаря переводов.

Ключевые слова

анализ тональности текста, двуязычные векторные представления слов, рекуррентные нейронные сети, глубокое обучение, казахский язык

Благодарности

Работа выполнена при финансовой поддержке Российского научного фонда (проект 15-11-10019 «Разработка моделей и методов text mining, семантической обработки текстов в задачах анализа потребностей, предпочтений и поведения потребителей»). Авторы выражают благодарность команде Everware за доступ к платформе (<https://github.com/orgs/everware>), а также Ералану Сейткаzinову, Зарине Садыковой и Альфии Ситдиковой за создание корпуса казахских текстов с разметкой.

Introduction

Sentiment analysis is an actively studied problem. Consumers use the web as an advisory body in influencing their view on matters. Knowing what is said on the web gives the possibility to react upon negative sentiment and to monitor positive sentiment. The social media connect the entire world and, thus, people can much more easily influence each other. Hundreds of millions of people around the world actively use websites such as Twitter and news portals to express their thoughts. That is why, there is a growing interest in sentiment analysis of texts where people express their thoughts or their opinion across a variety of domains such as commerce [1] and health [2, 3].

Sentiment analysis is the process of automatically determining sentiment expressed in natural language. As social media cover almost the entire world, a sentiment expressed by the users of social media is written in a multitude of languages. Here we face a new problem. For some languages, e.g. Kazakh language, there is no large enough labeled corpora to use them as a training data for sentiment analysis. The problem, we study in this paper, is to determine the general opinion expressed in texts written in one natural language, taking into account another language and how to apply such information in training a sentiment classifier.

Here we focus on short texts, in particular, on social media messages: news comments and micro-blogging posts. Sentiment analysis of such texts is challenging because of the limited amount of contextual data in this type of text. In this work, we propose a deep recurrent neural network that uses bilingual word embeddings to capture semantic features between words of two languages. We perform experiments on two language pairs: English-Russian and Russian-Kazakh. A sentiment has been one of the two classes: positive and negative.

In this paper, we describe an approach to building bilingual word embeddings and how to use it to create a deep neural network classifier, that achieves a competitive performance on sentiment analysis for the Russian language. We evaluate the model on a baseline in sentiment analysis for the Kazakh language.

Related Work

Distributed representations of words also known as word embeddings have been introduced as a part of neural network architectures for statistical language modeling ([4–7]). Generally, word embeddings is a very natural idea that treats words like math objects. Classical approach to building word embeddings constructs one-hot encoding, where each word corresponds with its own dimension. Obviously, there is a necessity to train representations for individual words, basically, as reduction of dimensionality. In particular, distributed word representations solve this problem. It maps each word occurring in the dictionary to a Euclidean space, attempting to capture semantic attitudes between words as geometric relationships. Thus, distributed word representations are very useful in different NLP tasks such as semantic similarity [8], information retrieval [9] and sentiment analysis [10].

There are few methods to build multilingual word embedding. In particular, Zou et al. [8] introduced bilingual word embeddings through utilizing Machine Translation word alignments to translational equivalence. Vulic and Moens [11] proposed a simple effective approach of learning bilingual word embeddings from non-parallel document-aligned data. Also Lu et. al [12] extend the idea of learning deep non-linear transformations of word embeddings for two languages, using the deep canonical correlation analysis.

Sentiment analysis task of short text is a very popular task in NLP. Mohammad et. al [13] described one of the state-of-the-art Twitter message-level sentiment classifying using SVM. Dos Santos and Gatti [10] proposed a deep convolutional neural network exploiting character-level and word-level embeddings to perform sentiment analysis of short texts, and achieved state-of-the-art results in binary classification, with 85.7% accuracy. Although, there are many works related to these models, little work has been done to use bilingual word embeddings to improve sentiment analysis, especially, for the Kazakh language.

Bilingual Word Embeddings

Assume, that we have two large not aligned corpora in the source language W^S and the target language W^T , respectively, and a set of bilingual word pairs (dictionaries) V^S and V^T for each language. Our goal is to generate vectors x and y in space R^{S+T} and retain semantic relationships between vectors from both source spaces and supplement them with new relationships between words of two languages. For example, in the joint

semantic space the Russian word 'школа' (school) is expected to be close to its Kazakh translation 'мектеп' (school). Besides, words 'школы' (plural form of the Russian word 'школа', schools) and 'мектептер' (plural form of the Kazakh word 'мектеп', schools) that are not contained in the dictionary, are also expected to be near.

There are several ways to solve this problem. We consider two methods in this paper. We propose a relatively straightforward method to creating multilingual word embeddings. The main idea of this method is generating a single "pseudo-bilingual" corpus through mixing source corpora with a second language. In the first step, we clean dictionaries V^S and V^T depending on the frequency of words in their corpora. We delete very common words using threshold. Due to the fact that words are commonly used in different contexts, they have different meanings. Following that, we have randomly splitted source language corpus to two parts and replace every n -th word in the first half with direct translation given in the dictionary V^S . Exactly the same step we apply to target language corpus. It has been done in order to extend the context of using particular word in two languages. Having bilingual contexts for each word in pseudo-bilingual corpus, we train the final model and construct a shared multilingual embedding space. The second method is to train word embeddings for each language and then applying linear regression transform word embeddings from the source to target language. This method was proposed by Mikolov et al. [14]. The objective function in regression task looks as follows:

$$\min_{\beta} \sum_i \|\beta x_i - y_i\|^2,$$

where β is a transformation matrix we have to calculate; x_i and y_i are word vectors of source and target language word spaces respectively.

Neural Network Architecture

As a basic structure of deep neural network we use the Long Short-Term Memory (LSTM) model proposed by Hochreiter and Schmidhuber [15]. LSTM model is a type of recurrent neural network (RNN). In a traditional recurrent neural network, during the gradient back-propagation phase, the gradient signal can end up being multiplied many times by the weight matrix associated with the connections between the neurons of the recurrent hidden layer (general LSTM architecture shown in Fig. 1).

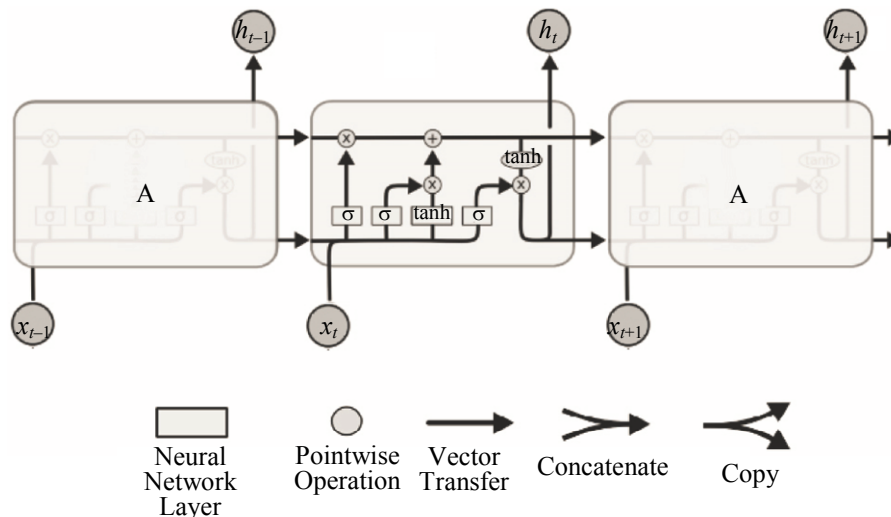


Figure 1. This diagram shows LSTM memory block architecture (taken from [17]). Here x_t – input values of neural network (NN) in the moment t , σ and \tanh – activation functions of hidden layers, A – simple LSTM unit

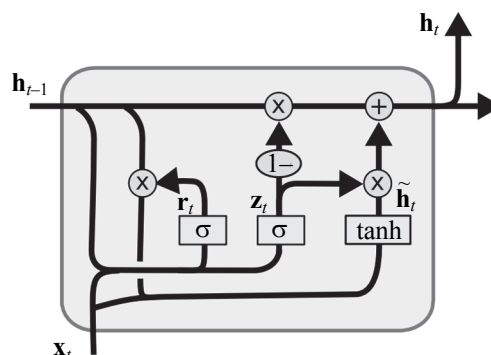


Figure 2. This diagram shows GRU memory block architecture (taken from [17]). Here x_t is input values of neural network (NN) in the moment t , h_t is output values of NN and r_t and z_t are hidden states of memory block

This means that, the magnitude of weights in the transition matrix can have a strong impact on the learning process. RNN makes all predictions sequentially, and the hidden layer from one prediction is fed to the hidden layer of the next prediction. This gives the network "memory", in the sense that the results from previous predictions can inform future predictions. LSTMs are explicitly designed to avoid the long-term dependency problem. Thus, LSTM networks are especially good in sequence labeling tasks [16].

Likewise, we are interested in evaluating the performance of one more recently proposed recurrent unit - GRU. A gated recurrent unit (GRU) was proposed by Cho et al. [18] to make each recurrent unit adaptively capture dependencies of different time scales. Similarly to the LSTM unit, the GRU has gating units that modulate the flow of information inside the unit, however, without having a separate memory cells. GRU memory block architecture is shown in Fig. 2.

In order to score a sentence, the network takes the sequence of words in the sentence as input, and passes it through a sequence of layers where features with increasing levels of complexity are extracted.

Scoring and Network Training

A sentence \mathbf{x} is given with n words $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$, which have been converted to joint word-level embeddings. We use a special padding token for sentences with small sizes. Then we get sentence-level representation passing word-level embeddings through two LSTM layers. Finally, the vector \mathbf{r}_x , the obtained feature vector of sentence \mathbf{x} , is processed by two fully connected (dense) neural network layers, which extract one more level of representation and compute a score for each sentiment label $c \in C$ as a logistic classifier.

The network was trained using RMSProp ([19, 20]), that worked better than using an annealed learning rate. We use dropout [21] as a powerful regularizer, even when the network was only two layers deep. Also we use dropout technique to regularize hidden states in LSTM layer. So that we have achieved good generalization capability getting opportunity to train a neural network in one language and predict in another.

Experimental Setup and Results

We apply our model for two different language pairs: English-Russian and Russian-Kazakh. As the English sentiment labeled dataset we use the Stanford Twitter Sentiment corpus introduced by Go et al. [22]. In our experiments, to speedup the training process we use only a sample of the training data consisting of 100 K randomly selected tweets. As the Russian dataset we use Russian Twitter corpus introduced by Rubtsova and Zagorulko [23]. For the Kazakh dataset we collect a corpus from news comments including about 1400 documents. Table 1 shows how we splitted the mentioned datasets. At the preprocessing step we have deleted sentence boundaries, non-letter characters (except apostrophe symbol) and have replaced all URLs to hashtag "#Replace- dUrl". Also we removed all emoticons, because training corpora was built using emoticon labeling and it has a huge impact to final results, whereas our goal is to achieve competitive results in bilingual text evaluations.

Datasets	Tweets	Documents	Classes
English	Train	80 000	2
	Test	20 000	
Russian	Train	80 000	2
	Test	20 000	
Kazakh	Train	1100	2
	Test	300	

Table 1. Sentiment analysis datasets

Unsupervised Learning of Bilingual Word Embeddings

Word embeddings play very important role in the model architecture. They meant to capture syntactic and semantic information that is very important to sentiment analysis. In our experiments, we perform unsupervised learning of word embeddings using the word2vec tool [24] that implements the continuous bag-of-words and skip-gram architectures for computing vector representations of words [6]. We use the English Wikipedia corpus, a collection of Russian news documents and a collection of Kazakh news documents [25] as a source of unlabeled data. We removed all documents that are less than 50 characters long. Also we lowered case all words and substituted each numerical digit by a "0" (e.g., 25 becomes 00). The resulting cleaned corpora contains about 280 million tokens for English, about 190 million tokens for Russian and about 20 million tokens for Kazakh.

After the preprocessing we start to "mix" mentioned corpora to each other in the following manner. We select replacing window size of 6 and further the same window size will be used for training skip-gram. Following that we get two corpora for English-Russian pair and Russian-Kazakh pair.

When running the word2vec tool, we set that a word must occur at least 4 times in order to be included in the vocabulary, and the resulted vocabulary is of about 900 K entries for English-Russian pair and about 600 K for Russian-Kazakh pair. The training time for the English-Russian pair corpus is around 4hrs and around 1h for

This section describes the network architectures and training details for the experimental results reported in this paper. The code for reproducing these results can be obtained from https://github.com/eabdullin/nlp_mthesis. The implementation is based on Keras library and Theano as backend using CPU. But also there is a possibility to use GPU. More detailed description of using GPU with Keras may be found in [30].

Training data	English		Russian	
	Accuracy	ROC AUC	Accuracy	ROC AUC
Our approach (to building a bilingual word embeddings)				
100% English	0.73	0.80	0.59	0.62
75% English and 25% Russian	0.73	0.81	0.67	0.76
50% English and 50% Russian	0.74	0.81	0.70	0.78
Linear transformation approach				
100% English	0.69	0.74	0.55	0.60
75% English and 25% Russian	0.70	0.77	0.59	0.60
50% English and 50% Russian	0.71	0.77	0.60	0.64
100% English, SVM (Go et al. [22])	0.82	–	–	–
100% English, NB (Go et al. [22])	0.83	–	–	–

Table 3. Evaluating English-Russian pair model

Training data	Russian		Kazakh	
	Accuracy	ROC AUC	Accuracy	ROC AUC
Our approach (to building a bilingual word embeddings)				
100% Russian	0.71	0.79	0.55	0.58
98% Russian and 2% Kazakh	0.72	0.80	0.56	0.64
95% Russian and 5% Kazakh	0.72	0.79	0.58	0.67

Table 4. Evaluating Russian-Kazakh pair model

Conclusion

In this work we present an approach to performing bilingual sentiment analysis. We propose a new relatively simple approach to building word embeddings. The main contributions of the paper are:

1. the new approach to building bilingual word embeddings;
2. the idea of using pre-trained bilingual word embeddings in neural network architecture;
3. experimental results for Kazakh sentiment analysis.

Proposed method may be used to perform a sentiment analysis in different language that has not enough labeled corpora. For this purpose, researches need to have only dictionaries to translate words. As a future work, we would like to build Kazakh sentiment labeled corpus using our classification model. Additionally, we would like to check the impact of performing the semi-supervised learning.

Литература

References

1. Jansen B.J., Zhang M., Sobel K., Chowdury A. Twitter power: tweets as electronic word of mouth // *Journal of the American Society for Information Science and Technology*. 2009. V. 60. N 11. P. 2169–2188. doi: 10.1002/asi.21149
2. Chew C., Eysenbach G. Pandemics in the age of twitter: content analysis of tweets during the 2009 H1N1 outbreak // *PLoS One*. 2010. V. 5. N 11. Art. e14118. doi: 10.1371/journal.pone.0014118
3. Paul M.J., Dredze M. You are what you tweet: analyzing twitter for public health // *ICWSM*. 2011. V. 20. P. 265–272.
4. Bengio Y., Schwenk H., Senecal J.-S., Morin F., Gauvain J.-L. Neural probabilistic language models // *Innovations in Machine Learning*. 2006. V. 194. P. 137–186. doi: 10.1007/3-540-33486-6_6
5. Collobert R., Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning // *Proc. 25th Int. Conf. on Machine Learning*. 2008. P. 160–167. doi: 10.1145/1390156.1390177
6. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space // *Proceedings of Workshop at ICLR*, 2013.
7. Pennington J., Socher R., Manning C.D. Glove: global vectors for word representation // *Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. V. 14. P. 1532–
1. Jansen B.J., Zhang M., Sobel K., Chowdury A. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 2009. vol. 60, no. 11, pp. 2169–2188. doi: 10.1002/asi.21149
2. Chew C., Eysenbach G. Pandemics in the age of twitter: content analysis of tweets during the 2009 H1N1 outbreak. *PLoS One*, 2010, vol. 5, no. 11, art. e14118. doi: 10.1371/journal.pone.0014118
3. Paul M.J., Dredze M. You are what you tweet: analyzing twitter for public health. *ICWSM*, 2011, vol. 20, pp. 265–272.
4. Bengio Y., Schwenk H., Senecal J.-S., Morin F., Gauvain J.-L. Neural probabilistic language models. *Innovations in Machine Learning*, 2006, vol. 194, pp. 137–186. doi: 10.1007/3-540-33486-6_6
5. Collobert R., Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning. *Proc. 25th Int. Conf. on Machine Learning*, 2008, pp. 160–167. doi: 10.1145/1390156.1390177
6. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space. *Proceedings of Workshop at ICLR*, 2013.
7. Pennington J., Socher R., Manning C.D. Glove: global vectors for word representation. *Proc. Conf. on Empirical Methods in Natural Language Processing EMNLP*, 2014, vol. 14, pp.

1543. doi: 10.3115/v1/d14-1162
8. Zou W.Y., Socher R., Cer D.M., Manning C.D. Bilingual word embeddings for phrase-based machine translation // *EMNLP*, 2013. P. 1393–1398.
 9. Manning C.D., Raghavan P., Schütze H. et al. *Introduction to Information Retrieval*. Cambridge University Press, 2008. V. 1. N 1.
 10. dos Santos C.N., Gatti M. Deep convolutional neural networks for sentiment analysis of short texts // *Proc. 25th Int. Conf. on Computational Linguistics*. Dublin, Ireland, 2014. P. 69–78.
 11. Vulic I., Moens M.-F. Bilingual word embeddings from non-parallel document-aligned data applied to bilingual lexicon induction // *Proc. 53rd Annual Meeting of the Association for Computational Linguistics (ACL 2015)*. 2015. doi: 10.3115/v1/p15-2118
 12. Lu A., Wang W., Bansal M., Gimpel K., Livescu K. Deep multilingual correlation for improved word embeddings // *Proc. Annual Conference of the North American Chapter of the ACL (NAACL)*. Denver, Colorado, 2015. P. 250–256. doi: 10.3115/v1/n15-1028
 13. Mohammad S.M., Kiritchenko S., Zhu X. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets // *arXiv preprint*. 2013. arXiv:1308.6242.
 14. Mikolov T., Le Q.V., Sutskever I. Exploiting similarities among languages for machine translation // *arXiv preprint*. 2013. arXiv:1309.4168.
 15. Hochreiter S., Schmidhuber J. Long short-term memory // *Neural Computation*. 1997. V. 9. N 8. P. 1735–1780. doi: 10.1162/neco.1997.9.8.1735
 16. Graves A. *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer, 2012. 146 p. doi: 10.1007/978-3-642-24797-2
 17. Olah C. *Understanding LSTM networks*. 2015. Available at: <http://colah.github.io/posts/2015-08-Understanding-LSTMs> (accessed: 30.11.16).
 18. Cho K., van Merriënboer B., Bahdanau D., Bengio Y. On the properties of neural machine translation: encoder-decoder approaches // *Proc. Workshop on Syntax Semantics and Structure in Statistical Translation*. 2014.
 19. Tieleman T., Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude // *COURSERA: Neural Networks for Machine Learning*. 2012. V. 4. P. 2.
 20. Dauphin Y.N., de Vries H., Chung J., Bengio Y. Rmsprop and equilibrated adaptive learning rates for non-convex optimization // *arXiv preprint*. 2015. arXiv:1502.04390.
 21. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting // *The Journal of Machine Learning Research*. 2014. V. 15. N 1. P. 1929–1958.
 22. Go A., Bhayani R., Huang L. *Twitter sentiment classification using distant supervision* // *Technical Report CS224N*. Stanford, 2009. V. 1. P. 12.
 23. Rubtsova Y.V., Zagorulko Y.A. An approach to construction and analysis of a corpus of short Russian texts intended to train a sentiment classifier // *The Bulletin of NCC*. 2014. V. 37. P. 107–116.
 24. Google. *Tool for computing continuous distributed representations of words*. Available at: <https://code.google.com/p/word2vec> (accessed: 30.11.16).
 25. Makhambetov O., Makazhanov A., Yessenbayev Z., Matkarimov B., Sabyrgaliyev I., Sharafudinov A. Assembling the kazakh language corpus // *Proc. Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, 2013. P. 1022–1031.
 26. Van der Maaten L., Hinton G. Visualizing data using t-sne // *Journal of Machine Learning Research*. 2008. V. 9. P. 2579–2605.
 27. Chollet F. *Keras: Theano-based deep learning library*. 2015. Available at: <https://github.com/fchollet> (accessed: 30.11.16).
 28. Bergstra J., Breuleux O., Bastien F., Lamblin P., Pascanu R., Desjardins G., Turian J., Warde-Farley D., Bengio Y. Theano: a cpu and gpu math expression compiler // *Proc. Python for Scientific Computing Conference (SciPy)*. Austin, 2010. V. 4. P. 3.
 29. *Machine Learning in Python*. Available: <http://scikit-learn.org> (accessed: 30.11.16).
 8. Zou W.Y., Socher R., Cer D.M., Manning C.D. Bilingual word embeddings for phrase-based machine translation. *EMNLP*, 2013, pp. 1393–1398.
 9. Manning C.D., Raghavan P., Schütze H. et al. *Introduction to Information Retrieval*. Cambridge University Press, 2008, vol. 1, no. 1.
 10. dos Santos C.N., Gatti M. Deep convolutional neural networks for sentiment analysis of short texts. *Proc. 25th Int. Conf. on Computational Linguistics*. Dublin, Ireland, 2014, pp. 69–78.
 11. Vulic I., Moens M.-F. Bilingual word embeddings from non-parallel document-aligned data applied to bilingual lexicon induction. *Proc. 53rd Annual Meeting of the Association for Computational Linguistics AC*, 2015. doi: 10.3115/v1/p15-2118
 12. Lu A., Wang W., Bansal M., Gimpel K., Livescu K. Deep multilingual correlation for improved word embeddings. *Proc. Annual Conference of the North American Chapter of the ACL, NAACL*. Denver, Colorado, 2015, pp. 250–256. doi: 10.3115/v1/n15-1028
 13. Mohammad S.M., Kiritchenko S., Zhu X. Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint*, 2013, arXiv:1308.6242.
 14. Mikolov T., Le Q.V., Sutskever I. Exploiting similarities among languages for machine translation. *arXiv preprint*, 2013. arXiv:1309.4168.
 15. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, vol. 9, no. 8, pp. 1735–1780. doi: 10.1162/neco.1997.9.8.1735
 16. Graves A. *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer, 2012, 146 p. doi: 10.1007/978-3-642-24797-2
 17. Olah C. *Understanding LSTM networks*. 2015. Available at: <http://colah.github.io/posts/2015-08-Understanding-LSTMs> (accessed: 30.11.16).
 18. Cho K., van Merriënboer B., Bahdanau D., Bengio Y. On the properties of neural machine translation: encoder-decoder approaches. *Proc. Workshop on Syntax Semantics and Structure in Statistical Translation*, 2014.
 19. Tieleman T., Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, 2012, vol. 4, p. 2.
 20. Dauphin Y.N., de Vries H., Chung J., Bengio Y. Rmsprop and equilibrated adaptive learning rates for non-convex optimization. *arXiv preprint*, 2015, arXiv:1502.04390.
 21. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 2014, vol. 15, no. 1, pp. 1929–1958.
 22. Go A., Bhayani R., Huang L. *Twitter sentiment classification using distant supervision*. *Technical Report CS224N*. Stanford, 2009, vol. 1, p. 12.
 23. Rubtsova Y.V., Zagorulko Y.A. An approach to construction and analysis of a corpus of short Russian texts intended to train a sentiment classifier. *The Bulletin of NCC*, 2014, vol. 37, pp. 107–116.
 24. Google. *Tool for computing continuous distributed representations of words*. Available at: <https://code.google.com/p/word2vec> (accessed: 30.11.16).
 25. Makhambetov O., Makazhanov A., Yessenbayev Z., Matkarimov B., Sabyrgaliyev I., Sharafudinov A. Assembling the kazakh language corpus. *Proc. Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, 2013, pp. 1022–1031.
 26. Van der Maaten L., Hinton G. Visualizing data using t-sne. *Journal of Machine Learning Research*, 2008, vol. 9, pp. 2579–2605.
 27. Chollet F. *Keras: Theano-based deep learning library*. 2015. Available at: <https://github.com/fchollet> (accessed: 30.11.16).
 28. Bergstra J., Breuleux O., Bastien F., Lamblin P., Pascanu R., Desjardins G., Turian J., Warde-Farley D., Bengio Y. Theano: a cpu and gpu math expression compiler. *Proc. Python for Scientific Computing Conference*. Austin, 2010, vol. 4, p. 3.
 29. *Machine Learning in Python*. Available: <http://scikit-learn.org> (accessed: 30.11.16).

- (accessed: 30.11.16).
30. Chollet F. *Keras: Deep learning library for Theano and tensorflow*. Available: <http://keras.io> (accessed: 30.11.16).

Авторы

Абдуллин Еламан Бердикулулы – разработчик программного обеспечения, ТОО «Hibrain», Астана, 010000, Казахстан; младший научный сотрудник, Казанский федеральный университет, Казань, 420008, Российская Федерация, a.elaman.b@gmail.com

Иванов Владимир Владимирович – кандидат физико-математических наук, научный сотрудник, Университет Иннополис, Иннополис, 420500, Российская Федерация, nomemm@gmail.com

Authors

Yelaman B. Abdullin – software developer, «Hibrain» LTD, Astana, 010000, Kazakhstan; junior scientific researcher, Kazan Federal University, Kazan, 420008, Russian Federation, a.elaman.b@gmail.com

Vladimir V. Ivanov – PhD, scientific researcher, Innopolis University, Innopolis, 420500, Russian Federation, nomemm@gmail.com