Optimal recurrent neural network model in paraphrase detection*

A. N. Smerdov, O. Yu. Bakhteev, V. V. Strijov

Abstract: The paper addresses the problem of choosing the optimal recurrent neural network. In this paper the optimal criteria is evidence lower bound. The paper investigates variational inference methods to approximate the posterior distribution of the model parameters. As a particular case the normal distribution of the parameters with different types of the covariance matrix is investigated. The authors propose a method of pruning parameters with the highest probability density in zero to increase the model marginal likelihood. As an illustrative example, a computational experiment of multiclass classification on SemEval 2015 dataset have been carried out.

Keywords: *deep learning; recurrent neural network; neural network pruning; variational approach.*

References

[1] Sutskever, I., O. Vinyals, and Q. V. Le. 2014. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*. 3104–3112. Available at: `https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf` (accessed December 29, 2017)

[2] Bishop, C. M. Pattern Recognition and Machine Learning. Springer. 2006.

[3] Kuznetsov, M. P., A. A. Tokmakova, and V. V Strijov. 2016. Analytic and stochastic methods of structure parameter estimation. *Informatica*. 607–624.

[4] Popova, M. S., and V. V Strizhov. 2015. Selection of optimal physical activity classification model using measurements of accelerometer. *Informatika i ejo primenenija*. 9(1):76–86.

*This research was supported by RFBR, project 16-07-01160, and by Government of the Russian Federation, agreement 05.Y09.21.0018.

[5] Sanborn, A., and J. Skryzalin. 2015. Deep Learning for Semantic Similarity. *CS224d: Deep Learning for Natural Language Processing — Stanford, CA, USA: Stanford University*. Available at: https://cs224d.stanford.edu/reports/SanbornAdrian.pdf (accessed December 29, 2017)

[6] Pennington, J., R. Socher, and C. D. Manning. 2014. Glove: Global vectors for word representation. *Proceedings of the Empiricial Methods in Natural Language Processing*. doi:10.3115/v1/D14-1162

[7] Rong, X. 2014. word2vec parameter learning explained. *Arxiv*. Available at: https://arxiv.org/abs/1411.2738 (accessed December 29, 2017)

[8] Shi, T., and Z. Liu. 2014. Linking GloVe with word2vec. *Arxiv*. Available at: http://arxiv.org/abs/1411.5595 (accessed December 29, 2017)

[9] Zolotov, V., and D. Kung. 2017. Analysis and optimization of fastText linear text classifier. *Arxiv*. Available at: https://arxiv.org/ftp/arxiv/papers/1702/1702.05531.pdf (accessed December 29, 2017)

[10] Graves, A. 2011. Practical variational inference for neural networks. *Advances in Neural Information Processing Systems 24 (NIPS 2011)*. 2348–2356. Available at: http://papers.nips.cc/paper/4329-practical-variational-inference-for-neural-networks.pdf (accessed December 29, 2017)

[11] Le Cun, Y., J. S Denker., and S. A Solla. 1989. Optimal Brain Damage. *Proceedings of NIPS-89*. 2. P. 598–605. Available at: https://papers.nips.cc/paper/250-optimal-brain-damage.pdf (accessed December 29, 2017)

[12] Hassibi, B., D. G. Stork, and G. J Wolff. 1993. Optimal brain surgeon and general network pruning. *Neural Networks., IEEE International Conference on. — IEEE*. 293-299.

[13] Smerdov, A. N. Computational experiment code. Available at: https://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group474/Smerdov2017Paraphrase/code/ (accessed December 29, 2017)

[14] Dataset of sentences with different types of similarity. Available at: http://alt.qcri.org/semeval2015/task2/index.php?id=data-and-tools (accessed December 29, 2017)

[15] Glove python library. Available at: https://github.com/stanfordnlp/GloVe (accessed December 29, 2017)