Nizhny Novgorod State University
Institute of Information Technologies, Mathematics and Mechanics
Department of Computer Software and Supercomputer Technologies

Educational course «Introduction to deep learning using the Intel® neon™ Framework»

Lecture №7
Recurrent neural networks

Supported by Intel

Kustikova Valentina

Nizhny Novgorod
2018

Content

1 Abstract

The goal of this lecture is to study networks with feedback between different neuron layers – ***recurrent neural networks*** [2]. Initially, recurrent networks are constructed for processing sequences of the same data type, i.e. the order of representing objects to the network input is important. Typical examples of tasks in which this feature is clearly found are the tasks of speech recognition (processing of a sequence of sounds, processing of natural language texts). Nevertheless, a number of computer vision problems are highlighted, which are also successfully solved using recurrent networks (tracking objects on video).

A recurrent network approximates the behavior of any dynamic system. The introduction of recurrent networks requires a generalization of the concept of ***a computational graph***. The computational graph of recurrent neural networks can contain loops that reflect the dependence of the variable value at the next time from its current value. First, we describe the idea of deploying recurrent networks in a computational graph that has a repeating structure, which usually reflects a sequence of events [3]. As the classical examples of recurrent networks, Elman's and Jordan's networks are represented. We consider the procedure of training recurrent networks on the example of the Elman's network. A general scheme of the backpropagation through time [3] method for training network parameters is introduced. Possible ways of constructing deep recurrent networks are given [3, 4]: a conventional recurrent network, a deep transition recurrent network (DT-RNN), a deep output recurrent network (DO-RNN).

If there are enough long input sequences, during the training process the network "forgets" the information about previous objects. In some cases, it becomes necessary for the network to "remember" the information about the objects at the beginning of the sequence. Thereby, networks of long short-term memory (LSTM) are introduced [5 – 7]. A general scheme for constructing a long short-term memory unit is described, and the classical structure of the LSTM-unit is cited [7]. The practical application of recurrent networks is demonstrated based on the example of the task of classifying a person's sex by photo using the Intel® neon™ Framework. The recurrent network being developed consists of a set of convolutional layers and two recurrent blocks constructed according to the method proposed in [8].

2 Literature

2.1 Books

1. Haykin S. Neural Networks: A Comprehensive Foundation. – Prentice Hall PTR Upper Saddle River, NJ, USA. – 1998.
2. Osovsky S. Neural networks for information processing. – 2002.
3. Goodfellow I., Bengio Y., Courville A. Deep Learning. – MIT Press. – 2016. – [http://www.deeplearningbook.org].

2.2 Further reading

4. Pascanu R., Gulcehre C., Cho K., Bengio Y. How to Construct Deep Recurrent Neural Networks [https://arxiv.org/pdf/1312.6026.pdf].
5. Hochreiter S., Schmidhuber J. Long short-term memory // Neural Computation. – 1997. – P.1735-1780. – [http://www.bioinf.jku.at/publications/older/2604.pdf].
6. Greff K., Srivastava R.K., Koutnık J., Steunebrink B.R., Schmidhuber J. LSTM: A Search Space Odyssey // Transactions on Neural Networks and Learning Systems. – 2017. – [https://arxiv.org/pdf/1503.04069.pdf].

2.3 References

7. Understanding LSTM Networks [http://colah.github.io/posts/2015-08-Understanding-LSTMs].
8. Visin F., Ciccone M., Romero A., Kastner K., Cho K., Bengio Y., Matteucci M., Courville A. ReSeg: A Recurrent Neural Network-based Model for Semantic Segmentation // In CVPR Deep Vision Workshop, 2016. – 2016. – [https://arxiv.org/abs/1511.07053].