Contribution ID: **301**   Type: **Sectional reports**

Russian-language speech recognition system based on DeepSpeech

Thursday 13 September 2018 14:00 (15 minutes)

The paper examines the practical issues in developing a speech-to-text system using deep neural networks.
The development of a Russian-language speech recognition system based on DeepSpeech architecture is described. The Mozilla company's open source implementation of DeepSpeech for the English language was used as a starting point.

The system was trained in a containerized environment using the Docker technology. It allowed to describe the entire process of component assembly from the source code, including a number of optimization techniques for CPU and GPU. Docker also allows to easily reproduce computation optimization tests on alternative infrastructures.

We examined the use of TensorFlow XLA technology that optimizes linear algebra computations in the course of neural network training. The number of nodes in the internal layers of neural network was optimized based on the word error rate (WER) obtained on a test data set, having regard to GPU memory limitations. We studied the use of probabilistic language models with various maximum lengths of word sequences and selected the model that shows the best WER.

Our study resulted in a Russian-language acoustic model having been trained based on a data set comprising audio and subtitles from YouTube video clips. The language model was built based on the texts of subtitles and publicly available Russian-language corpus of Wikipedia's popular articles. The resulting system was tested on a data set consisting of audio recordings of Russian literature available on voxforge.com—the best WER demonstrated by the system was 18%.

Authors: Prof. DEGTYAREV, Alexander (Professor); SHALEVA, Anna (Saint-Petersburg State University); FE-DOSEEV, G. (Saint-Petersburg State University); SEDOVA, O. (Saint-Petersburg State University); IAKUSHKIN, Oleg (Saint-Petersburg State University)

Presenter: SHALEVA, Anna (Saint-Petersburg State University)

Session Classification: 11. Big data Analytics, Machine learning

Track Classification: 11. Big data Analytics, Machine learning