

End-to-end approaches to speech recognition and language processing

Jan Chorowski *

Abstract

End-to-end techniques solve complex machine learning tasks by building models that can be trained by optimizing a single joint loss criterion. Therefore all of a model's components collaborate to solve the task at hand. In this talk I will present attention-based recurrent neural networks, which are well suited for speech and language processing. I will show how one can build translation [1], speech recognition [2] and speech translation [3] systems using those networks. I will also demonstrate how similar attention-based recurrent networks can be applied to NLP tasks, such as parsing [4].

References

- [1] Bahdanau, D., Kyunghyun, C., and Yoshua, B., *Neural machine translation by jointly learning to align and translate*. ICLR 2015.
- [2] Chorowski, J., Bahdanau, D., Serdyuk, D., Cho, K., and Bengio, Y. *Attention-based models for speech recognition*. NIPS 2015.
- [3] Weiss, R., Chorowski, J., Jaitly, N., Wu, Y., and Chen, Z. *Sequence-to-Sequence Models Can Directly Transcribe Foreign Speech*. INTERSPEECH 2017
- [4] Chorowski, J., Zapotoczny, M., and Rychlikowski, P. *Read, Tag, and Parse All at Once, or Fully-neural Dependency Parsing*. arXiv preprint arXiv:1609.03441.

*University of Wroclaw