A Neural Network Segmenter

Ashkan Alinejad

November 14, 2016

Contents

- Previous Works
- Basic Idea
- Future works

- **segmentation** for Machine Translation is the task of dividing up sentences into pieces that needs to be translated together.
- Ideally our segmentation needs to consider many parameters.
- Most of previous works are based on statistical and mathematical approaches [1].

- My works is based on this paper.
- They have considered Latency beside Quality in order to find better results.
- Their algorithm needs to have the whole sentence.

**Learning Segmentations that Balance Latency versus Quality
in Spoken Language Translation**

Hassan S. Shavarani   Maryam Siahbani   Ramtin M. Seraj   Anoop Sarkar

School of Computing Science
Simon Fraser University, Burnaby BC. Canada
{sshavara,msiahban,rmehdiza,anoop}@cs.sfu.ca

- Their algorithm is one of the best segmenters for machine translation.
- generate a great dataset for my work.
- **My Goal** : Designing and implementing a real-time MT segmenter based on DRNNs.

2 main ideas:

1. Output of my network is probability
2. My network consider the segmentation character as a separate word and tries to predict the next word.

Future Works

- Working on different methods of decoding my output.
- Use my system in a Neural Machine Translator.

References

📄 Yusuke Oda, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura.
Optimizing segmentation strategies for simultaneous speech translation.

📄 Hassan S. Shavarani, Maryam Siahbani, Ramtin Mehdizadeh Seraj, and Anoop Sarkar.
Learning segmentations that balance latency versus quality in spoken language translation.
In *International Workshop on Spoken Language Translation (IWSLT)*, Da Nang, Vietnam, 2015.

📄 Minh-Thang Luong, Hieu Pham, and Christopher D. Manning.
Effective approaches to attention-based neural machine translation.
CoRR, abs/1508.04025, 2015.

Thanks for your attention !