Cross-Lingual Alignment of Contextual Word Embeddings in Natural Language Processing

Researchers have significantly improved the performance of unsupervised machine translation. A critical step of unsupervised machine translation is word embedding alignment, which aims to map word embeddings of the source language to the target language. Previous research of word embedding alignment focuses on traditional, context-independent word embeddings, such as word2vec, GloVe, etc. However, recent advances in contextual word embeddings raise new questions: How to build proper alignments between cross-lingual contextual word embeddings?

In this project, as a first step, we will look into the details of previously proposed methods for cross-lingual alignment of context-independent word embeddings. Based on previous work, we will then devise new methods for aligning cross-lingual contextual word embeddings, and apply such methods to downstream NLP applications, including machine translation, word similarity evaluation, etc. You will have access to powerful GPUs, and weekly discussions with two experienced PhD students in deep learning.

Requirements: Strong motivation, proficiency in Python, ability to read papers and work independently. Prior knowledge in deep learning is preferred.

Interested? Please contact us for more details!

Contact

- Zhao Meng: zhmeng@ethz.ch, ETZ G61.3
- Damian Pascual Ortiz: dpascual@ethz.ch, ETZ G93
- Yunpu Ma: cognitive.yunpu@gmail.com