

Master internship in Natural Language Processing and Machine Learning: Cross-lingual transfer with multi-lingual BERT via linguistically informed fine-tuning

Caio Corro

- **Duration:** 5-6 months, during the year 2020
- **Location:** LIMSI, Orsay (south of Paris)
- **Supervisor:** Caio Corro
- **Team:** Spoken Language Processing / Traitement Automatique de la Parole
- **Contact:** caio.corro@limsi.fr

1 Context

Recently, much attention has been paid to large scale pre-training of context-sensitive representations (or context-sensitive word embeddings), in particular ELMO [1] and BERT [2] models. The main idea is to pre-train the first layers of a neural network on a large amount of unlabeled data before fine-tuning the rest of the network on a downstream task. As such, context-sensitive representations allow to lower annotation cost and improve classification performance on a wide range of tasks.

The multilingual BERT model pre-trains context sensitive representations on a collection of texts in 104 languages instead of texts in a single language. One question that arises is whether we can use the multilingual BERT model for cross-lingual learning, that is training a model on a subset of these languages (source languages) and testing it on a different subset (target languages). This problem is both important under a research perspective (how can we learn multi-lingual representations of typologically diverse languages?) and under an applied industry perspective (i.e. increase language coverage of NLP-based products at low cost). Previous work observed that cross-lingual transfert based on multi-lingual BERT works best for typological similar languages (i.e. languages with similar word order), which is expected but disappointing [3].

This internship will focus on multilingual dependency parsing with the Universal Dependency treebank.¹ Previous work has considered re-ordering source language sentences with respect to word order in target languages [4]. However, re-ordering is not possible for unsupervised large scale pre-training where syntactic structures is not annotated. A different line of work proposed to force word order statistics at test time using constraints [5], but this method is based on a costly lagrangian optimization procedure and cannot be applied on a per sentence basis. Alternatively, we propose to explore fine-tuning methods for multi-lingual BERT model using a linguistically informed training algorithm, i.e. to use dominant word order information (is the object placed before or after the verb in a given language?) to ensure unsupervised transfer to target languages.

2 Missions

The successful candidate will develop neural network architectures and training algorithms for cross-lingual generalization of pre-trained context-sensitive representations. The main evaluation task will be cross-lingual dependency parsing. As there are many ways to tackle this problem, the specific approach will be determined by the intern aspiration, which could be for example posterior regularization or latent variable modeling. In a nutshell, the aim is to:

- propose a method for cross-lingual generalization of multi-lingual BERT using typological information;
- evaluate the proposed method on cross-lingual parsing;
- evaluate if results generalize to other tasks, for example cross-lingual named entity recognition.

¹<https://universaldependencies.org/>

References

- [1] Matthew Peters et al. *Deep Contextualized Word Representations*
- [2] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*
- [3] Telmo Pires et al. *How multilingual is Multilingual BERT?*
- [4] Lauriane Aufrant et al. *Zero-resource Dependency Parsing: Boosting Delexicalized Cross-lingual Transfer with Linguistic Knowledge*
- [5] Tao Meng et al. *Target Language-Aware Constrained Inference for Cross-lingual Dependency Parsing*