

Abstract

Semantic role labeling (SRL) is an essential task for understanding natural language, which allows for identifying events and their participants in a text by means of semantic roles (eg. agent, theme, goal). As such, it has been proven useful for a wide range of tasks in natural language processing, such as information extraction, question answering, machine translation, etc. However, the amount of manually labelled data necessary for building SRL systems is unfortunately not available for most languages given the time and resources required for their creation. Various cross-lingual methods have been suggested in order to create such models for resource-poor languages by means of model transfer or annotation projection while often making use of existent monolingual SRL systems and parallel language corpora.

The proposed thesis builds upon the work of Kozhevnikov and Titov (2013) who proposed a model transfer method for SRL of English, French, Czech and Chinese languages by making use of shared feature representations, such as cross-lingual clustering or cross-lingual distributed word representations, and machine learning. Even though they have demonstrated competitive results without using word-aligned parallel corpora, they still make use of syntactic information for the transfer. In this project, we show that by using a state-of-the-art neural SRL system (He, Lee, Lewis, et al., 2017) and pretrained cross-lingual word embeddings (Smith et al., 2017) we can achieve competitive results even without using any syntactic information. On the CoNLL-2009 data, our best models achieve weighted F1 score of 74.44 and 79.86 for French and Chinese language, respectively.

Key words: semantic role labelling, model transfer, cross-lingual word embeddings, deep learning, bi-LSTM