Abstract

Cross-lingual transfer is an approach for developing NLP systems for low-resource languages where NLP tools for the resource-constrained languages are developed using the resources of languages for which the labelled data are available. The transfer methods have in applied in several tasks with varying degree of success.

In this thesis, we focus on the development of dependency parsers for low-resourced languages. We particularly aim to address the syntactic gaps between the source and target languages under different settings to improve the quality of the transferred parsers. The cross-lingual transfer approaches are faced with several challenges arising from the differences in vocabulary, syntax and morphological structure of the source and target languages that significantly restricts the performance of cross-lingual systems. Research in cross-lingual transfer is primarily aimed towards addressing these challenges in order to improve quality of the transferred systems.

To this end, we propose a transformation method to reduce the syntactic differences between a given source-target language pair where the word order differ significantly from predominant ordering. We present an approach for cross-lingual transfer of dependency parser so that the parser trained on a single source language can more effectively cater to diverse target languages. We also propose a treebank selection strategy when multiple source languages are used to train the transfer parser model and also a technique to determine the number of languages whose combination gives the best results. We also explore a chunk-aware cross-lingual parsing strategy to address the syntactic difference between two languages where we train the parser using the chunks as units instead of the words. Finally we present a classification of the different cross-lingual model transfer approaches reported in the literature based techniques adopted by them to address the challenges related to transfer parsing. We also present a quantitative comparison of some of the cross-lingual transfer parsing methods based on these aspects.

Keywords: Dependency parsing, Cross-lingual dependency parsing, Cross-lingual model transfer, Delexicalization, Multi-source dependency parsing, Treebank transformation