Frustratingly Easy Cross-Lingual Transfer for Transition-Based Dependency Parsing

Summarized By: Faizan E Mustafa

Introduction:

Most of the research work in NLP is focused on a few commercially important languages such as english. This makes cross-lingual techniques very useful for low-resourced languages. Coss-lingual techniques enable us to develop useful applications also for low-resource language. In this work, authors transfer the dependency parser between two languages using annotation transfer.

Annotation transfer requires parallel corpus as we need word alignments to infer syntactic structure from source dependencies. Deliberate effort is required to make a parallel corpus and the resulting word alignments are noisy which demands to use heuristics to deal with non-isomorphism. This study offers an easy baseline solution for annotation transfer which does not require previously used complex heuristics.

Methodology:

The approach is based on simple observation that it is possible to learn a good parser from reliable partially annotated data. A transition-based dependency parser based on arc-eager algorithm is used to incrementally build dependency trees. At each step of the process, the highest scoring action is selected. The training algorithm uses dynamic oracle to update the weights using the perceptron rule when the prediction action does not result in gold dependency. Dynamic oracle is an error-correction algorithm; it does not update the weights when the chosen action results in gold dependency. This property makes us able to use it as it is even if partially annotated data is available. That means when no reference dependency is known, all the actions are considered correct and no weights are updated.

To test our observation regarding the dynamic oracle, a controlled experiment is performed to compare the performance when using n% of the sentences versus using full train data but only n% of the dependency relations known. The plot in Figure 1 shows that it is indeed possible to learn a parser from partially annotation data. It can also be seen that the number of training samples can be reduced without hurting the performance significantly and for a similar number of dependencies known, better results are obtained when more sentences are annotated, even if the annotation is partial.

This paragraph shows how learning from partially annotated data can be used for cross-lingual transfer. If we have sentence-aligned parallel data where a well-resourced source language has been parsed and have corresponding sentences of desired low-resource language, then dependencies can be projected using alignment links. However, this results in noisy alignments due to many-to-many alignments and unaligned tokens. Heuristics has been proposed to deal with noisy data but they generally produce unreliable annotations. The approach in this paper is relatively simple. Firstly, Giza++ is used to generate alignments which uses the intersection heuristic to generate reliable 1:1 alignment. Secondly, different simple heuristics are used which includes removing non-projective dependencies from target sentences, ignoring sentences for which less than

80% of the words are aligned and dropping the alignments that links words with different POS tags. This approach reduces the noise and only gives the most reliable alignments.

This suggests that we can learn a transition-based parser with dynamic oracle and beam search using reliable partially annotated data obtained by removing noise from the alignments.

Data and Experiment:
Six languages namely German, English, Spanish, French, Italian and Swedish are used from the Universal Dependency Treebank(UDT). A subset of Europarl corpus is used to get the parallel sentences.

To build a parser for target language, MateParser trained with a beam size of 40 on the UDT is used to parse the source sentences. Afterwards, the method described in methodology is used to project the dependencies on the target side to obtain partially annotated data. Once we have partially annotated target data, an arc-eager parser with dynamic oracle can be used to train a target language parser. The resulting parser is evaluated using the UDT test data with Unlabeled Attachment Score as a metric. The overview of the method is illustrated in Figure 1X.
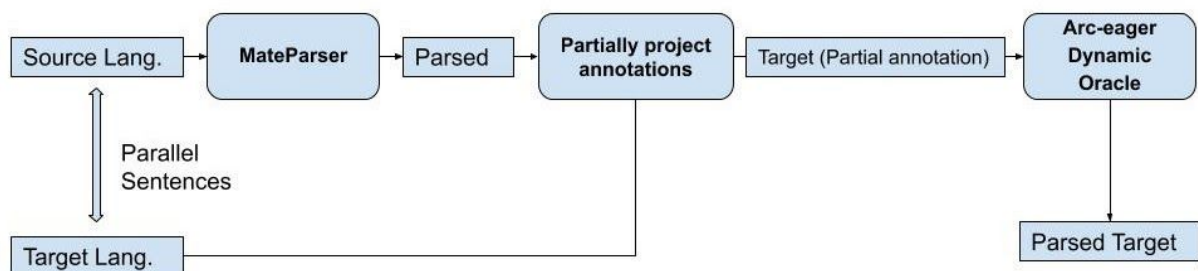

Figure 1X Overview of the approach.

Results:
The metric scores obtained from this approach is compared with three other annotation transfer approaches. The results are shown in Table 1 shows that our model scores much better than the McDonald et al. approach. It also outperforms or is at par for different languages when compared to the Ma & Xia ,and Rasooli & Collins approach. However, Rasooli & Collins outperforms our approach when trained on full projected trees. One more positive aspect of our approach is that it is computationally cost effective.

Conclusion and Future Work:
In my opinion, the potential drawback for this approach is that it still requires parallel data which is usually a result of deliberate effort and not easily available for some domains. Nevertheless, the experiments show the effectiveness of using reliable partially annotated data. However, the results with other models are not comparable due to differences in training conditions such as beam size. The approach provides a strong baseline approach for cross-lingual dependency parsers. The learning from this approach(usefulness of partially annotated data) could be used in some other approach. So, this model could also be combined with other approaches to get a strong ensemble approach.