Unsupervised Neural Dependency Parsing

Yong Jiang, Wenjuan Han and Kewei Tu
School of Information Science and Technology, ShanghaiTech University

Unsupervised Dependency Parsing

Given a set of unannotated sentences, our goal is to learn a dependency model that can parse test data with high accuracy.

Dependency Model with Valence (DMV)

Dependency Model with Valence (DMV) is the first model to outperform the left-branching baseline. It generates both the parse tree and the sentence in a recursive manner.

Learning

- EM algorithm
- Hard EM algorithm

Results on Eight Other Languages

We also applied our approach on datasets of eight additional languages from the PASCAL Challenge on Grammar Induction (Gelling et al., 2012). We ran our approach using the hyper-parameters from the previous experiment on the new datasets without any further tuning.

Results on the WSJ Dataset

Results on the basic DMV model and compared the results against (Cohen and Smith, 2009) and (BergKirkpatrick et al., 2010), both of which have very similar motivation as ours in that they also utilize the correlation between POS tags to learn the basic DMV model.

Results on the extended DMV model (Headden III et al., 2009; Gillenwater et al., 2010) (with the maximum valence value set to 2 for both CHILD and DECISION rules). As shown in the table, we achieve comparable accuracy with recent state-of-the-art systems. If we initialize our model with the grammar learned by Tu and Honavar (2012), the accuracy of our approach can be further improved.

Reference

- Cohen and Smith, 2009
- BergKirkpatrick et al., 2010
- Headden III et al., 2009
- Gillenwater et al., 2010
- Tu and Honavar, 2012

Model Analysis

A visualization of the distances between embeddings of different POS tags. For example, VBP (Verb, non-3rd person singular present), VBD (Verb, past tense) and VBZ (Verb, 3rd person singular present) can be seen to be close to each other, and they indeed have very similar syntactic behavior.