Dependency Parsing for Weibo: An Efficient Probabilistic Logic Programming Approach

William Yang Wang, Lingpeng Kong, Kathryn Mazaitis, William W. Cohen
Carnegie Mellon University, Pittsburgh, PA, USA.

1. motivation: Weibo attracts 30% of Internet users, but NLP techniques for analyzing Weibo are not well studied.
2. question: Can we find efficient inference and learning methods for dependency arc prediction on Weibo?
3. goals:
   - Develop a new Chinese Weibo treebank
   - Make parser programmable via theory engineering
   - Efficient non-linear first-order probabilistic logic learning
   - Effective and efficient inference of the dependency structure

4. a new Chinese Weibo dependency treebank
   - Free: available at http://www.cs.cmu.edu/~yww/data/WeiboTreebank.zip
   - Annotation method: FUDG (Schneider et al., 2013) and GFL annotation tool (Mordowanec et al., 2014).
   - Training set #tokens: 14,774.
   - Development set #tokens: 3,223.
   - Test set #tokens: 1,700.

5. experimental results

6. conclusions
   - A new Chinese Weibo dependency treebank: we provide a freely available Chinese Weibo dependency treebank.
   - Programmable dependency parsing on Weibo: we show that it is easy to use language-specific parsing theory for Weibo parsing in ProPPR.
   - Promising results: we show that with language and genre specific first-order theory, our performance is better than an off-the-shelf Stanford parser and a state-of-the-art multilingual parser that is trained on the same in-domain data.