

# **N-best Parse Rescoring Based on Dependency-Based Word Embeddings**

Yu-Ming Hsieh,                      Wei-Yun Ma

Institute of Information Science

Academia Sinica

E-mail: {morris, ma}@iis.sinica.edu.tw

## **Abstract**

Rescoring approaches for parsing aim to re-rank and change the order of parse trees produced by a general parser for a given sentence. The re-ranking quality depends on the precision of the rescoring function. However it is a challenge to design an appropriate function to determine the qualities of parse trees. No matter which method is used, Treebank is a widely used resource in parsing task. Most approaches utilize complex features to re-estimate the tree structures of a given sentence [1, 2, 3]. Unfortunately, sizes of treebanks are generally small and insufficient, which results in a common problem of data sparseness. Learning knowledge from analyzing large-scaled unlabeled data is compulsory and proved useful in the previous works [4, 5, 6]. How to extract useful information from unannotated large scale corpus has been a research issue. Word embeddings have become increasingly popular lately, proving to be valuable as a source of features in a broad range of NLP tasks [7, 8, 9]. The *word2vec* [10] is among the most widely used word embedding models today. Their success is largely due to an efficient and user-friendly implementation that learns high quality word embeddings from very large corpora. The *word2vec* learns low dimensional continuous vector representations for words by considering window-based contexts, i.e., context words within some fixed distance of each side of the target words. Another different context type is dependency-based word embedding [11, 12, 13], which considers syntactic contexts rather

than window contexts in *word2vec*. Bansal et al. [8] and Melamud et al. [11] show the benefits of such modified-context embeddings in dependency parsing task. The dependency-based word embedding can relieve the problem of data sparseness, since even without occurrence of dependency word pairs in a corpus, dependency scores can be still calculated by word embeddings [12]. In this paper, we proposed a rescoring approach for parsing, based on a combination of original parsing scores and dependency word embedding scores to assist the determination of the best parse tree among the  $n$ -best parse trees. There are three main steps in our rescoring approach. The first step is to have the parser to produce  $n$ -best parse trees with their structural scores. For each parsed tree including words, part-of-speech (PoS) and semantic role labels. Second, we extract word-to-word associations (or called word dependency, a dependency implies its close association with other words in either syntactic or semantic perspective) from large amounts of auto-parsed data and adopt *word2vecf* [13] to train dependency-based word embeddings. The last step is to build a structural rescoring method to find the best tree structure from the  $n$ -best candidates. We conduct experiments on the standard data sets of the Chinese Treebank. We also study how different types of embeddings influence on rescoring, including word, word with semantic role labels, and word senses (concepts). Experimental results show that using semantic role labels in dependency embeddings has best performance. And the final experiments results indicate that our proposed approach outperforms the best parser in Chinese. Furthermore we attempt to compare the performance of using the traditional conditional probability method with our approach. From the experimental results, the embedding scores can relax data sparseness problem and have better results than the traditional approach.

**Keywords:** Word Embeddings, Parsing, Word Dependency, Rescoring.

## References

- [1] K. Hayashi, S.Kondo, and Y. Matsumoto, "Efficient stacked dependency parsing by forest reranking," *Transactions of the ACL*, vol. 1, pp. 139-150, 2013.
- [2] L. Shen, A. Sarkar, and A. Toshi, "Using LTAG based features in parse reranking," in *Proceedings of EMNLP*, pp. 89-96, 2003.
- [3] E. Charniak and M. Johnson, "Coarse-to-fine  $n$ -best parsing and MaxEnt discriminative reranking," in *Proceedings of ACL 2005*, pp. 173-180, 2015.

- [4] A. Wu, “Learning verb-noun relations to improve parsing,” in *Proceedings of the Second SIGHAN workshop on Chinese Language Processing*, pages 119-124, 2003.
- [5] K. Yu, D. Kawahara, and S. Kurohashi, “Chinese Dependency Parsing with Large Scale Automatically Constructed Case Structures,” in *Proceedings of the 22nd International Conference on Computational Linguistics (COLING2008)*, pp. 1049-1056, 2008.
- [6] Y. Hsieh, J. S. Chang, and K. Chen, “Ambiguity Resolution for Vt-N Structures in Chinese,” in *Proceedings of EMNLP 2014*, pp. 928-937, 2014.
- [7] J. Turian, L. Ratinov, and Y. Bengio, “Word representations: A simple and general method for semi-supervised learning,” in *Proceedings of ACL*, pp.384–394, 2010.
- [8] M. Bansal, K. Gimpel, and K. Livescu, “Tailoring continuous word representations for dependency parsing,” in *Proceedings of ACL 2014*, pp. 809-815, 2014.
- [9] R. Socher, A. Perelygin, J. Wu, J. Chuang, C.D. Manning, A.Y. Ng, and C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank,” in *Proceedings of EMNLP*, pp. 1631-1642, 2013.
- [10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in *Proceedings of ICLR. 2013*.
- [11] O. Melamud and D. McClosky and S. Patwardhan and M. Bansal, “The Role of Context Types and Dimensionality in Learning Word Embeddings,” in *Proceedings of NAACL-HLT 2016*, pp. 1030-1040, 2016.
- [12] O. Melamud, O. Levy, and I. Dagan, “A Simple Word Embedding Model for Lexical Substitution,” in *Proceedings of NAACL-HLT 2015*, pp. 1-7, 2015.
- [13] O. Levy and Y. Goldberg, “Dependency-Based Word Embeddings,” in *Proceedings of ACL 2014*, pp. 302-308, 2014.