Improving Long-length Dependency Parsing by Parser Ensemble

Kun Yu
Graduate School of Information Science and Technology
The University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo, Japan
kunyu@is.s.u-tokyo.ac.jp

Junichi Tsujii
Graduate School of Information Science and Technology
The University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo, Japan
tsujii@is.s.u-tokyo.ac.jp

Abstract

In this paper, we propose a parser ensemble method to combine a statistical parser with a deterministic parser, in order to improve the parsing accuracy of dependencies with long length. Experimental results based on Penn Chinese Treebank show that not only the accuracy of dependencies with long length was increased by the proposed method, but also the overall accuracy of the statistical parser was improved significantly.

1 Introduction

Dependency parsing has gained a wide interest recently and used successfully in many natural language processing applications. Among the various dependency parsing models, statistical dependency parsing (Collins, 1996) finds correct dependency structure in a global point of view; deterministic dependency parsing (Yamada and Matsumoto, 2003; Niver and Scholz, 2004; Hall et al., 2006; Nivre et al., 2007) uses a greedy algorithm that approximates a globally optimal solution by making a sequence of locally optimal choices (Hall et al., 2006). But, both the two approaches cannot deal with the dependency relations, in which the head and modifier are not near to each other, very well. For statistical parsing, noises are easy to be introduced during decoding. For deterministic parsing, the error of short length dependency is easy to be propagated to following choices (McDonald and Nivre, 2007).

In this paper, we call the dependency relations in which head and modifier are not near to each other as ‘long-length dependency’. To improve the accuracy of long-length dependency parsing, we propose a parser ensemble method to combine the two different parsing approaches together. Our purpose is to see if the combination of two ‘*weak*’ models could produce a ‘*strong*’ model for long length dependency parsing.

We did experiments on Penn Chinese Treebank (Xue et al., 2002) to validate the feasibility of the proposed method. Results prove that by using parser ensemble, the accuracy of the statistical parsing model on long-length dependency was greatly improved. Besides, the dependency accuracy of the statistical parser was increased significantly.

The following part is organized as follows. Section 2 introduces the proposed parser ensemble method in detail. Experimental results and discussion are listed in Section 3. Finally, Section 4 gives a brief conclusion and a direction of future work.

2 Parser Ensemble: Combining Statistical Parsing with Deterministic Parsing

In this work, we pay attention to combine a statistical parsing model with a deterministic parsing model by parser ensemble, in order to improve the accuracy of the statistical parser on long-length dependency. Currently, we focus on Chinese dependency parsing.

2.1 Statistical Dependency Parsing Model

We use a Chinese lexicalized statistical dependency parser proposed in (Yu et al., 2008). This parser gives a probability $P(T|S)$ to each possible dependency tree T of an input sentence $S=w_1,w_2,\ldots,w_n$ (w_i is a node representing a word with its pos-tag), and outputs the dependency tree T^* that maximizes $P(T|S)$ (see equation 1). CKY algorithm is used to decode the dependency tree from bottom to up.

$$T^* = \arg\max_T P(T \mid S) \tag{1}$$

When computing the probability, large scale lexicalized case structures are applied in this parsing model. Although the usage of lexical information help recognize correct dependency, the accuracy of long-length dependency is still not so good. Table 1 lists the dependency accuracy of this statistical parser on Penn Chinese Treebank[1]. It shows that with the increase of dependency length, the parsing accuracy dropped greatly.

Table 1 Result of the statistical dependency parser.

Dep. Length	*UAS*[2] (%)
1	93.01
2	85.87 (-7.14)
3	82.66 (-10.35)
4	80.55 (-12.46)
5	79.15 (-13.86)
5+	72.35 (-20.66)

2.2 Deterministic Dependency Parsing Model

For the deterministic parsing model, we use a Chinese deterministic dependency parser proposed in (Yu et al., 2007). This model is a shift-reduce model. It divides parsing into three-steps, which are *sentence segmentation*, *sub-sentence parsing* and *parsing combination*, to decrease the effect of error propagation.

Table 2 lists the dependency accuracy of this parser on the same training and testing data set as used in Section 2.1. It shows that because of not using any extra lexical information, the accuracy of this deterministic parser is much lower than that of the statistical parser in Table 1. But, the trend that parsing accuracy decreases with the increase of dependency length is almost the same as that of the statistical parsing model.

Table 2 Result of the deterministic dependency parser.

Dep. Length	*UAS* (%)
1	93.43
2	84.41 (-9.02)
3	82.51 (-10.92)
4	77.31 (-16.12)
5	77.65 (-15.78)
5+	65.01 (-28.42)

2.3 Parser Ensemble by Reparsing

Although neither the statistical parsing model nor the deterministic parsing model is good at recognizing long-length dependency, we hope the different mechanisms of the two parsing models could help each other. Therefore, we propose a parser ensemble method as following:

Step1: parsing the input sentence by the deterministic parser;
Step2: collecting the dependencies created by the deterministic parser as *constituent set*;
Step3: deleting the dependencies containing punctuations from the *constituent set*.
Step4: re-parsing the input sentence by the statistical parser, but assigning a penalty to the probability of dependency that does not exist in the *constituent set* during decoding.

In the current work, we assign a static penalty p for all the dependencies that do not exist in the *constituent set*. This penalty p is set as 0.5 in the following experiments.

3 Results and Discussion

3.1 Experimental Setting

We use Penn Chinese Treebank 5.1 as data set in this experiment. 9,684 sentences from Section 001-270 and 400-931 are used as training data. 346 sentences from Section 271-300 are used as testing data. Penn2Malt[3] is used to transfer the phrase structure of Penn Chinese Treebank to dependency structure. Gold-standard word segmentation and pos-tag are applied in all the experiments. Unlabeled attachment score (*UAS*) is used as evaluation metric. Because of the difficulty of assigning cor-

[1] We use the same training and testing data set that will be introduced in Section 3.
[2] *UAS* means unlabeled attachment score (Buchholz and Marsi, 2006).

[3] http://w3.msi.vxu.se/~nivre/research/Penn2Malt.html

rect head to Chinese punctuation, we calculate *UAS* only on the dependency relations in which the modifier is not punctuation.

Two parsing models are evaluated in the experiments:

- *w/o PE*: statistical parsing model without parser ensemble;
- *w/PE*: statistical parsing model with parser ensemble.

3.2 Results and Discussion

Experimental results are shown in Figure 1 and Figure 2. The two figures show that by using the proposed parser ensemble method, the accuracy of the statistical parser is improved on all the dependencies with different length. Besides, the accuracy improvement was more than 2% when the dependency length is larger than 2 (except for the two cases: *length* = 8 and *length* > 10). These results prove that although both the statistical parser and the deterministic parser are not good at long length dependency recognition, combining them by the proposed parser ensemble method helped improve the accuracy of long length dependencies greatly.

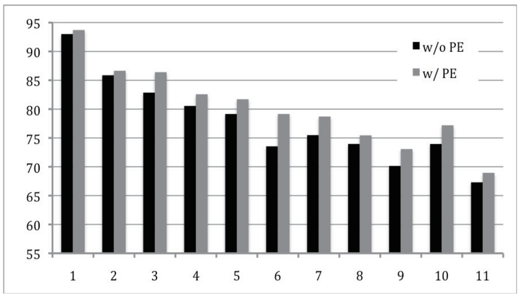

Figure 1. *UAS* of dependency with different length[4].

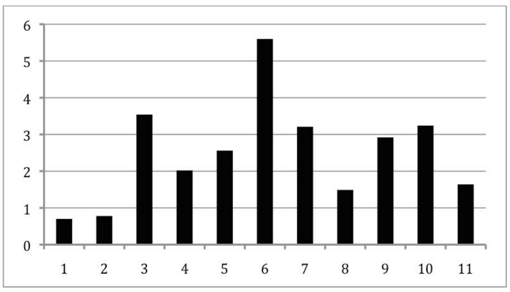

Figure 2. Improvement of *UAS* of dependency with different length[5].

[4] Horizontal axis is the length of dependency relation, vertical axis is *UAS* (%).
length = 11 means the length of dependency relation is more than 10.

Table 3 lists the overall accuracy of the statistical parser. It provides additional evidence that the proposed parser ensemble method is helpful to significantly (McNemar's test: $p<0.005$) improve the accuracy of the statistical parser.

Table 3 Results of statistical dependency parser.

	UAS (%)
w/o PE	87.26
w/ PE	88.17 (+0.91)

4 Conclusion and Future Work

In this paper, we propose a parser ensemble method to combine a statistical dependency parser with a deterministic dependency parser, to improve the accuracy of long-length dependency parsing. Experimental results based on Penn Chinese Treebank prove that although both the two parsing models are not good at long-length dependency parsing, using the proposed parser ensemble method increased the accuracy of long length dependency greatly.

The future work under consideration includes refining the selection of *constituent set* from the output of deterministic parsing; assigning dynamic penalty for different types of dependencies, adapting the proposed parser ensemble method to other languages, and so on.

References

S.Buchholz and E.Marsi. 2006. CoNLL-X Shared Task on Multilingual Dependency Parsing. In *Proceedings of the 10th Conference on Computational Natural Language Learning*.

M.Collins. 1996. A New Statistical Parser Based on Bigram Lexical Dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*. pp. 184-191.

J.Hall et al.. 2006. Discriminative Classifiers for Deterministic Dependency Parsing. In *Proceedings of Coling-ACL 2006*. pp. 316-323.

R.McDonald and J.Nivre. 2007. Characterizing the Errors of Data-driven Dependency Parsing Models. In *Proceedings of EMNLP-CoNLL 2007*.

J.Nivre and M.Scholz. 2004. Deterministic Dependency Parsing of English Text. In *Proceedings of the 20th*

[5] Horizontal axis is the length of dependency relation, vertical axis is the improvement of *UAS* (%).

International Conference on Computational Linguistics. pp. 64-70.

J.Nivre et al.. 2007. MaltParser: A Language-independent System for Data-driven Dependency Parsing. *Natural Language Engineering*. 13(2): 95-135.

N.Xue, F.Chiou and M.Palmer. 2002. Building a Large-Scale Annotated Chinese Corpus. In *Proceedings of the 18th International Conference on Computational Linguistics*.

H.Yamada and Y.Matsumoto. 2003. Statistical Dependency Analysis with Support Vector Machines. In *Proceedings of the 9th International Workshop on Parsing Technology*. 2003.

K.Yu et al.. 2007. A Three-step Deterministic Parser for Chinese Dependency Parsing. In *Proceedings of NAACL-HLT* 2007.

K.Yu et al.. 2008. Chinese Dependency Parsing with Large Scale Automatically Constructed Case Structures. In *Proceedings of the 22nd International Conference on Computational Linguistics*.