KSII The 11th Asia Pacific International Conference on Information Science and Technology(APIC-IST) 2016.

Improving Korean dependency parsing performance using predicate-argument features

Chang-Uk Shin and Jeong-Won Cha
Department of Computer Engineering, Changwon National University
Changwon, South Korea
[e-mail: papower1@changwon.ac.kr, jcha@changwon.ac.kr]
*Corresponding author: Jeong-Won Cha

Abstract

This paper proposes the use of predicate-argument feature to improve the performance of Korean dependency parsing. To better understand the errors our baseline system is still making, we examined five folds (44,610 errors in total in 100% of the evaluation data) and identified the major categories of errors: word with no particle, topic markers. To resolve such problems, this paper proposes a novel predicate-argument feature and inference model. Our experimental results on Korean dependency parsing task show that the proposed method leads to significant parsing performance improvement over a state-of-the-art baseline, and the system using the proposed method achieved the best parsing performance in Korean.

Keywords: CRFs, dependency parsing, cascaded chunking, predicate-argument feature

1. Introduction

A syntactic parsing is to find components of the sentence. We perform a Korean syntactic parsing using cascaded chunking[1]. We convert a syntactic parsing task to labeling task. Suppose that right next word is a head word of current word, we label a functional tag.
There are two types of errors in labeling methods we will do. One is a functional tag error and the other is a dependency error. The functional tag error is to take a wrong functional tag. That is, even though the word has a subject tag, it has an object tag. This is because of a topic marker or absence of particle. The dependency error is to take a wrong link between words.

Example is as follows:
(꽃이 (((핀 정원에) 소풍을) 갔다.))"
kko-chi pin jeong-won-e so-pung-eul gat-da.
(We took a stroll at flower garden.)

This case that even if '꽃이(kko-chi)' is the subject of '핀(pin)', '꽃이' is the subject of '갔다.'. In this paper, we focus on the first case.

2. Related Works

Dependency parsing methods are separated into two big parts. One is graph-based, and the other is transition-based method. Graph-based dependency parsing method find the highest scored tree for given sentence. So normally it takes over than $O(n^3)$.
Transition-based method set the sentence to buffer first, and check relation a top word of stack and first word of buffer with pushing each word one by one from buffer to stack.

3. Error classes of Cascaded Chunking and proposed feature

3.1 Baseline system

It is well-known that feature selection is very important for machine learning. And, recognition of good features is challenging due to the overfitting problem. So, developing good features is difficult.
We used first/last POS tags, phrase tag, verbal('1' when next word is verbal word, '-' otherwise), and suffixes(POS tag when that

This research was supported by the Ministry of Trade, Industry & Energy(MOTIE), Korea Institute for Advancement of Technology(KIAT) through the fostering project.

word has verbal-suffix morpheme, '-' otherwise) as our baseline features.

3.2 Error classes of baseline system

Table 1 shows the distribution of Sejong parsed corpus [2] basis of the functional tag of a modifier word and the phrase tag of a head word. Table 2 shows that the result of baseline system. In table 2, the number of words that have 'SBJ' as its functional tag is bigger than the number of table 1, large amount of this came from 'AJT' and 'OBJ'. And this difference is also shown in 'AJT' column and 'OBJ' column of table 2.
Errors among 'AJT', 'SBJ', 'OBJ' are occurring frequently when the word has no postpositional morpheme or specific one. In baseline system, it is hard to classify due to lack of features.

3.3 Predicate-argument feature

As we argued above, baseline system has a weakness of classifying 'AJT', 'SBJ', 'OBJ'. This weakness is common in cascaded chunking or unlexicalized model as subject/object/adjunct (SBJ/OBJ/AJT respectively) of phrase has almost same tag based features. Our goal for this paper is to handle this without lexicalized features. To archive this, we propose predicate-argument feature.
Predicate-argument feature is gathered from Sejong verbal semantic dictionary [2] and it has 'frame' information which means the postpositional morphemes those can be attached to the verbal word. Totally 22,732 words are listed in dictionary and coverage on Sejong parse corpus is 98.11%. Some examples are listed on table 3.

	-	AJT	MOD	SBJ	OBJ	etc
NP	82,096	3,919	109,462	1,288	953	15,375
VP	105,177	138,840	1,988	72,813	60,591	10,809
VNP	19,983	5,203	16,528	5,486	496	687
ETC	1,657	968	69	196	23	54

Table. 1. Distribution of Sejong parse corpus

	-	AJT	MOD	SBJ	OBJ	etc
NP	85,508	1,522	109,073	650	515	14,976
VP	106,502	136,041	1,598	82,308	58,996	8,971
VNP	20,040	3,731	16,325	5,573	441	417
ETC	957	329	120	70	6	22

Table. 2. Distribution of baseline system result

word	postpositional word lists
가깝다	가/에/와/에서
폭발하다	가
헤어지다	가/와

Table. 3. Examples of Predicate-argument features

4. Experiments and discussion

To show the effectiveness of the 'Predicate-argument' feature, we performed 5-fold cross-validation over all(include Base) experiences.
We used Sejong parsed corpus as we analyzed above and it has about 50k sentences.

Experiments	UAS	FUNC	LAS
Baseline	84.39	93.26	80.72
with Predicate-argument feature	85.14	93.7	81.91

Table. 4. Performance of Baseline system and proposed one.

After adding 'Predicate-argument' feature, LAS(labeled attachment score) is improved 1.19%. it seems that the feature helps to choose right arc and label as classifying verbal words with our novel features.

5. Conclusions

In this paper, we presented a novel feature which is gathered from Sejong verbal semantic dictionary and described why that is useful for Korean dependency parser.

References

1. Steven P. Abney, "Parsing By Chunks", In principle-Based Parsing, Kluwer Academic Publishers, 1991
2. Sejong parse corpus, semantic dictionary Available: http://www.sejong.or.kr