Advanced Science and Technology Letters
Vol.120 (AIA 2015), pp.798-802
http://dx.doi.org/10.14257/astl.2015.120.157

Expansion of Feature Information for Korean Semantic Role Labeling

Byeong-Cheol Jo[1,2],Mi-ran Seok[1,2], Hye-Jeong Song[1,2], Chan-Young Park[1,2], Jong-Dae Kim[1,2], Yu-seop Kim[1,2]1

[1]Department of Convergence Software, Hallym University, Korea
[2]Bio-IT Research Center, Hallym University, Korea
max91128@naver.com,smr4880@hanmail.net, {hjsong, cypark, kimjd, yskim01}@hallym.ac.kr

Abstract. Semantic role labeling is defined as determination of the semantic relation between a predicate and various arguments that are dependent on the given predicate. In this study, automatic semantic role labeling using 10,000 sentences in a semantic role tagged corpus constructed from a Korean syntax tagged corpus was conducted. In the Korean language, the grammatical relation between particle and word ending as well as their semantic relation is very important. When features based on the affix information created in this study were added to the basic features used in previous studies on semantic role labeling of languages, an F1 score of approximately 80.83% was obtained.

Keywords: Feature, Korean affix, josa, Eomi.

1 Introduction

Semantic Role Labeling (SRL) is defined as determination of the semantic relation between the predicate of a sentence and arguments that are dependent on that predicate. SRL may be regarded as the process of mapping a grammatical relation comprising "subject" and "object" between the predicate of a sentence and arguments as a semantic relation comprising "agent," "experiencer," and "theme"—a process that is typically carried out after parsing [1, 2]. SRL can be used to improve performance in various areas of natural language processing, including machine translation, information extraction, and question and answer.

According to Bjorkelund et al. [3], the performance of the SRL of agglutinative Japanese is very low. In agglutinative languages such as Japanese and Korean, because of inflection a grammatical relation cannot be presented, and challenges such as severe word fluctuation and myriad morpheme features are used.

Using Korean PropBank, which complies with the Proposition Bank [4] system as a semantic role tagged corpus, in this study, a Korean SRL system that can automatically construct a semantically tagged corpus was implemented. In the Korean

[1] He is a corresponding author

ISSN: 2287-1233 ASTL

language, diverse grammatical and semantic variations are formed as various affixes are combined with one etymon. Accordingly, this study aimed to improve the performance of SRL by expressing the Korean language's affix information as features. Combination of the affix features presented in this study with the features presented in previous studies resulted in SRL performance improvement of approximately 0.6%.

2 Related Work

Three main methods are utilized for SRL: the corpus based method [5], the case frame based method [6].

The case frame based method labels semantic roles using a case frame and selectional restriction and finds a frame that is suitable for predicate-argument relations. However, it has the following shortcomings: (1) construction of case frames is difficult, and (2) it cannot be applied to sentence forms that are not described in the frame [1-2].

The corpus based method labels semantic roles through machine learning, particularly supervised learning, after tagging the semantic role to a corpus. Although this method has the advantage of being stable, it has a weakness in that the performance of the SRL system depends inordinately on the quality and quantity of the constructed learning corpus [1-2].

In this paper, an automatic SRL method that employs the corpus based method is proposed.

3 Semantic Tag

Table 1. Semantic tag

Semantic role of PropBank	
ARG0 (agent)	M-EXT (size)
ARG1 (patient)	M-INS (equipment)
ARG2 (start point)	M-LOC (place)
ARG3 (end point)	M-MNR (method)
M-ADV (adverbial phrase)	M-NEG (negation)
M-CAU (cause)	M-PRD (qualification of predicate)
M-CND (condition)	M-PRP (purpose)
M-DIR (direction)	M-TMP (time)
M-DIS (connection of sentence)	

For semantic tagging, the semantic role of arguments in parsed sentences was examined for dependent predicates and appropriate semantic tagging was performed

by analyzing examples of the frame files of PropBank. Table 1 illustrates the semantic role of PropBank.

4 Features

Table 2. General features and new features

General Features	Korean Features	New features
A_stem/P_stem	A-JosaExist	Josa_80
A_POS_LV1/P_POS_LV1	A-JosaClass	One-word exist
A_POS_LV2/P_POS_LV2	A-JosaLength	One-word stem
A_CASE/P_CASE	A-JosaMorphemes	
A-LeftSiblingStem	A-JosaIdentity	
A-LeftSiblingPOS_LV1	A-EomiExist	
A-LeftSiblingPOS Lv2	A-EomiClass Lv1	
A-RightSiblingPOS Lv1	A-EomiClass Lv2	
A-RightSiblingPOS Lv2	A-EomiLength	
P-ParentStem	A-EomiMorphemes	
P-ChildStemSet	A-EomiIdentity	
P-ChildPOSSet Lv1		
P-ChildCaseSet		

In the Korean language, an affix containing a particle or word ending plays a very important role in parsing and semantic analysis. Unlike the existing language in English-speaking countries, because the Korean language does not have word order restrictions, information on the location of a word cannot be used for SRL. On the other hand, there are many cases in which syntax and word ending are combined in various forms, resulting in the syntax and meaning of words being determined. Therefore, use of affix information should have a significant effect on SRL performance.

5 Conditional Random Fields (CRFs)

CRFs are a class of statistical modeling methods used for structural prediction such as pattern recognition and machine learning. In this study, CRF suite was used and the semantic role was predicted using the average perception generated by the CRF algorithm.

In this model, $x = \{x_1, x_2, x_3, \ldots, x_T\}$ is the input data in which components are connected in sequence, and $Y = \{y_1, y_2, y_3, \ldots, y_T\}$ is the label for each component of the input data. In other words, when a new is given, a y value is predicted using the model:

$$p(y|x) = \frac{1}{z(x)} \prod_{t=1}^{T} \exp\{\sum_{k=1}^{K} \omega_k f_k(y_t, y_{t-1}, x_t)\} \quad (1)$$

Where z(x) standardizes the probability value, f_k is feature function, and ω_k is the weight of the feature. In this study, CRF suite was used and the semantic role was predicted using the average perceptron generated by the CRF algorithm.

6 Results of Experiments

Of 10,000 semantic role tagged sentences in a corpus, 8,000 were used as training data and 2,000 as test data for evaluation. In the experiments conducted, basic features and new features were combined. When the SRL experiment was conducted on only basic features with equivalent data, an F1 score of approximately 68.5% was obtained. When Korean features were added to the basic features, a score of 80.26% was obtained, and when new features were added, a score of 80.83% was obtained. As this study used the features specialized in the Korean language, the result showed an approximate improvement of 0.6%.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (2015R1A2A2A01007333) and by the Ministry of Education, Science and Technology (2010-0010612).

References

1. Kim, B., Lee, Y., Na, S., Kim, J., Lee, J., "Bootstrapping for Semantic Role Assignment of Korean Case Marker", Korea Information Science Society (2006), vol.33, no.1,pp.4-6.
2. Lee, C., Lim, S., Kim, H., "Korean Semantic Role Labeling Using Structured SVM", J. of KIISE, 42(2), (2015)
3. Björkelund, A., Hafdell, L., and Nugues, P. Multilingual semantic role labeling. In Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, Association for Computational Linguistics, (2009)
4. Palmer, M., Gildea, D., and Kingsbury, P. The proposition bank: An annotated corpus of semantic roles, Computational linguistics, 31(1), 71-106, (2005)
5. Hacioglu, K., Pradhan, S., Ward, W., Martin, J. H., and Jurafsky, D. Semantic role labeling by tagging syntactic chunks. In CoNLL-2004 Shared Task, (2004)

6. Kurohashi, S., and Nagao, M. A method of case structure analysis for Japanese sentences based on examples in case frame dictionary, IEICE TRANSACTIONS on Information and Systems, 77(2), (1994)

