KIPS Tr. Software and Data Eng.
Vol.5, No.11 pp.555~562 pISSN: 2287-5905

http://dx.doi.org/10.3745/KTSDE.2016.5.11.555

Korean Semantic Role Labeling Based on Suffix Structure Analysis and Machine Learning

Miran Seok† · Yu-Seop Kim††

ABSTRACT

Semantic Role Labeling (SRL) is to determine the semantic relation of a predicate and its argu-ments in a sentence. But Korean semantic role labeling has faced on difficulty due to its different language structure compared to English, which makes it very hard to use appropriate approaches developed so far. That means that methods proposed so far could not show a satisfied perfor-mance, compared to English and Chinese. To complement these problems, we focus on suffix information analysis, such as josa (case suffix) and eomi (verbal ending) analysis. Korean lan-guage is one of the agglutinative languages, such as Japanese, which have well defined suffix structure in their words. The agglutinative languages could have free word order due to its de-veloped suffix structure. Also arguments with a single morpheme are then labeled with statistics. In addition, machine learning algorithms such as Support Vector Machine (SVM) and Condi-tional Random Fields (CRF) are used to model SRL problem on arguments that are not labeled at the suffix analysis phase. The proposed method is intended to reduce the range of argument instances to which machine learning approaches should be applied, resulting in uncertain and inaccurate role labeling. In experiments, we use 15,224 arguments and we are able to obtain approximately 83.24% f1-score, increased about 4.85% points compared to the state-of-the-art Korean SRL research.

Keywords : Semantic Role Labeling, Suffix Structure Analysis, Josa, Eomi, Machine Learning, Support Vector Machine, Conditional Random Fields

접사 구조 분석과 기계 학습에 기반한 한국어 의미 역 결정

석 미 란† · 김 유 섭††

요 약

의미 역 결정은 한 문장에서 술어와 그것의 논항간의 의미 관계를 결정해주는 것을 말한다. 한편 한국어 의미 역 결정은 영어와는 다른 한국어 고유의 특이한 언어 구조 때문에 많은 어려움을 가지고 있는데, 이러한 어려움 때문에 지금까지 제안된 다양한 방법들을 곧바로 적용하기에 어려움이 있었다. 다시 말하자면, 지금까지 제안된 방법들은 영어나 중국어에 적용했을 때에 비해서 한국어에 적용하면 낮은 성능을 보여주었던 것이다. 이러한 어려움을 해결하기 위하여 본 연구에서는 조사나 어미와 같은 접사구조를 분석하는 것에 초점을 맞추었다. 한국어는 일본어와 같은 교착어의 하나인데, 이들 교착어에서는 매우 잘 정리되어 있는 접사구조가 어휘에 반영되어 있다. 교착어는 바로 이들 잘 정의된 접사 구조 때문에 매우 자유로운 어순이 가능하다. 또한 본 연구에서는 단일 형태소로 이루어진 논항은 기초 통계량을 기준으로 의미 역 결정을 하였다. 또한 지지 벡터 기계(Support Vector Machine: SVM)과 조건부 무작위장(Conditional Random Fields: CRFs)와 같은 기계 학습 알고리즘을 사용하여 앞에서 결정되지 못한 논항들의 의미 역을 결정하였다. 본 논문에서 제시된 방법은 기계 학습 접근 방식이 처리해야 하는 논항의 범위를 줄여주는 역할을 하는데, 이는 기계 학습 접근은 상대적으로 불확실하고 부정확한 의미 역 결정을 하기 때문이다. 실험에서는 본 연구는 15,224 논항을 사용하였는데, 약 83.24%의 f1 점수를 얻을 수 있었는데, 이는 한국어 의미 역 결정 연구에 있어서 해외에서 발표된 연구 중 가장 높은 성능으로 알려진 것에 비해 약 4.85%의 향상을 보여준 것이다.

키워드 : 의미 역 결정, 접사 구조 분석, 조사, 어미, 기계 학습, 지지 벡터 기계, 조건부 무작위장

1. 서 론

Semantic parsing of sentences is believed to be an important task on the road to natural language un-derstanding, with immediate applications in tasks such as information extraction and question answer-ing [1]. The primary task of semantic role labeling (SRL) is to indicate exactly what semantic relations hold among a predicate and its associated participants and properties, with these relations drawn from a prespecified list of possible semantic roles for that predicate (or class of predi-cates) [2].

※ 이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2015R1A2A2A01007333).
† 비 회 원 : 필아이티(주) IT사업본부 주임연구원
†† 종신회원 : 한림대학교 융합소프트웨어학과 교수
Manuscript Received : October 4, 2016
Accepted : October 12, 2016
* Corresponding Author : Yu-Seop Kim(yskim01@hallym.ac.kr)

Previous research on SRL can be divided into the case frame-based method and the corpus-based method. In the corpus-based method, support vector machine (SVM) [3-6], maximum entropy [7-9], and conditional random fields (CRF) [10-15] are widely used. However, Korean doesn't have enough linguistic resources for SRL, such as Proposition Bank [16]. That means that the case frame-based method and machine learning based methods could not show a satisfied performance, compared to English and Chinese. Therefore, this research adds a methodology based on suffix information analysis of the Korean language, one of the agglutinative lan-guages. The agglutinative languages, such as Japanese and Korean, have free word order. This causes the fact that the word order could not play a role in analyzing the syntactic and semantic structure of a sentence, compared to western lan-guages like English.

However, the suffix information of the languages tells a lot about the syntax and semantics.

Agglutinative languages have forms of languages that a function of a word is determined by its affixes. Agglutina-tive languages are characterized by word configuration that multiple morphemes (including prefix and suffix) are attached to central morpheme (stem). In this case, stem and each affix is always keeping their morphological word form.

In Korean, suffixes, such as josa (case suffix) and eomi (verbal ending), play a very important role in syntactic parsing and SRL. Fig. 1 shows an example of Josa and Eomi.

Geuga nareul mideumeuro, nado geureul midneunda.
(he) (me) (believe)(because), (I) (him) (believe)
Because he believes me, I believe him too.

Fig. 1. Example of Josa (Blue) and Eomi (Red). The Green Texts Indicate the Predicates. 2

Eomies (red characters), placed at the end of predicates (verbs or adjectives), are often used as part of the words, whereas josas (blue characters) represent the grammatical relationship between words or add different meaning to the root word, which is the noun, pronoun, and rhetoric. Kim et al. (2014) [17] used 11 features of different types re-garding josa and eomi, and the performance increased greatly compared to when only general features, used in English SRL, were used.

We suggest a hybrid approach for SRL that uses Korean suffix information analysis and a machine learning method, CRF in this paper. We also test SVM to compare its performance to CRF's. We used a semantically anno-tated Korean corpus tagged on the syntactic corpus anno-tated by the Electronic and Telecommunications Research Institute of Korea[1] to train and test our model.

2. Suffix Structure in Korean

2.1 Josa (Case Suffix)

For English, a word 'I' has various noun forms like 'I, my, me' in accordance with its grammatical role, but in Korean (agglutinative language), josa such as '-ga, -eul' is added after the noun root 'na' to be applied its syntactic role.

(1) Cheol-su-ga chaeg-eul ilg-neun-da.
(Cheol-su reads a book)

In example (1), '-ga' is sticking behind the word 'Cheolsu', and 'Cheolsu-ga' is giving an indicate that the word has a subjective role. '-eul' is also sticking behind 'chaeg (book)', and indicates that 'chaeg (book)' is an ob-jective word.

Josa can be divided into 3 types: a case josa, a connection josa and an auxiliary josa. A case josa shows a grammar function of a root word to which the josa is attached; subjective, objective, adverbial, complemental, determi-native, and vocative josas are included in this category.

(2) a. Seon-Saeng-nim-kkeseo O-sin-da.
(A teacher comes here.)
b. Chug-gu hyeob-hoe-eseo dae-hoe-reul ju-choe-han-da.
(Football Association will host the tournament.)

In example (1), '-ga' is basically used as a subjective josa, but '-kkeseo' in exmaple (2a) is used when referring to honorific. Also, '-eseo' is used when indicating institution or organization as 2b.

(3) Cheol-su-wa Yeong-su-neun o-raen chin-gu-i-da.
(Cheol-su and Yeong-su are old friends.)

(4) Ppang-man meok-ji mal-go, u-yu-do ma-syeo-ra.
(Don't eat bread only, but drink milk.)

A connection josa plays the role of connecting two words into a constituency having a single gram-matical role. In example (3), '-wa' serves to connect 'Cheolsu' and

1) http://www.etri.re.kr.

'Youngsu' into a single constituency whereas an auxiliary josa refers to a josa with its own meaning. In (4), '-man' or '-do' has been used as objective case josas, but it represents another special meaning unlike '-eul' or '-reul'. '-man' adds a meaning of exclusiveness to a normal objective word and '-do' adds a meaning of concurrency to the objective word.

In Fig. 1, '-ga' is the subjective case josa while '-reul' is the objective case josa and '-do' is an auxiliary josa that has the meaning of 'too'. Some of these josas are mapped to specific semantic roles with 80% or higher accuracy. Table 1 shows some of them. (Table 1, 2 and 3 shows data statistics extracted from 10,000 Korean semantically annotated sentences, which are built manually). We used this data in Korean SRL.

Table 1. Josas in the First Column of the Table Shows 80% or Higher Accuracy When They are Mapped into Specific Semantic Roles. The Second Column Shows the Mapped Role and the Third Column Means its Accuracy. The Final Column Shows Their Frequency in Our Corpus.

Josa	Role	Accuracy (%)	Freq.
-Eul (Theme)	ARG1	97.1	4,195
-Man (Only)	ARG1	89.4	104
-Cheo-reom (Like)	ARGM-EXT	85.6	97
-E-seo (At)	ARGM-LOC	80.8	1069

2.2 Eomi (Verbal Ending)

In Korean, a verval ending (eomi) has very complicated rules for its transformation. Predicates could have diverse syntactic and semantic funcitons with changing their eomies and their combinations.

Fig. 2 shows the classification of eomies. The prefinal eomi refers to one that cannot locate at the end of a word and it exists in between the root word and the final eomi, representing honorific, modesty, and tense. The tense pre-final eomi shows future tense or past tense.

(5) a. Na-neun ye-jeon-e I chaeg-eul ilg-eot-da.
(I read this book before.)
b. Nae-il geu-gos-eu-ro ga-get-da.
(I will go there tomorrow.)

'-eot-, -get-' of (5) is to represent a pre-final tense eomis, and '-eot-' represents the past tense and '-get-' represents the future tense. Out of the final eomi, the connection eomi does not end the sentence but connects two sentences; whereas a closing eomi closes a sentence.

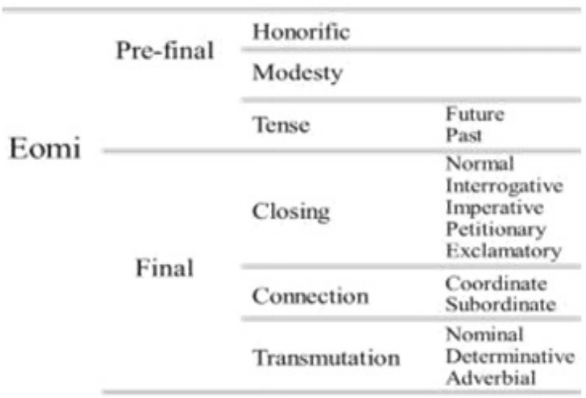

Fig. 2. Eomi is Composed of Two Classes, One for Pre-Final and the Other for Final. Pre-Final Eomi has Honorific, Modesty and Tense. Final Eomi also has Three Classes, Closing, Connection and Transmutation.

(6) a. In-saeng-eun jjal-go ye-su-reun gil-da.
(Life is short, and art is long.)
b. Ba-ram-i bul-myeon-seo bi-ga on-da.
(It rains with the wind blowing)

'-go' and '-myeon-seo' of (6) are coordinate connection eomis which connect two sentences in paral-lel. For example, "Ye-su-reun gil-go in-saeng-eun jjal-da. (Art is long, life is short.)"

A transmutation eomi can change the properties of a sentence to a adnomial or noun phrase.

(7) a. Jeo-gi-ga nae-ga sal-deon gos-i-da.
(That's where I lived.)
b. Mom-eul um-jig-i-gi him-deul-da.
(Moving the body is difficult.)

'-deon' changes a sentence ("I live") to a adnomial phrase ("where I live"), which is called 'adnomial eomi'. In contrast, '-gi' plays a role that changes a sentence ("move the body") as a noun phrase ("moving the body"), which is called 'nominal eomi'.

In Fig. 1, '-eumeuro' is a subordinate connection eomi that represents a causal relationship, and '-neunda' is a normal closing eomi that describes the current situation or fact.

Table 2 shows some eomies and their mapped semantic roles in PropBank. First column lists commonly used eomis and the second column shows their classes defined in Fig. 2. The third column and the fourth column shows the most mapped semantic roles and the mapping proportion, respectively. And the final colum shows frequencies occurred in our corpus.

Table 2. Eomis and Their Semantic Roles

Eomi	Class	Role	Acc. (%)	Freq.
-Go (And)	Coordinate Connection	ARGM-DIS	97	3,353
-Gi-ddaemin-e (Because)	Subordinate Connection	ARGM-CAU	96	174
-eot (Past)	Past Tense Prefinal	ARGM-DIS	75	1,335
-Da-go (Quotation)	Subordinate Connection	ARG1	58	337

3. Korean Semantic Role Labeling

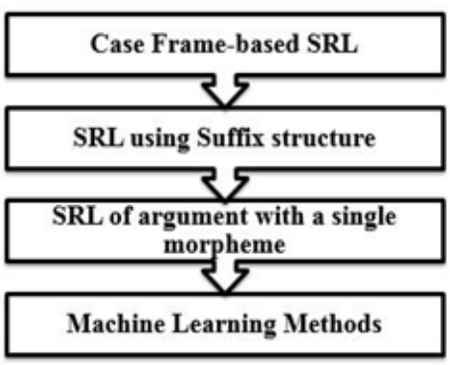

Fig. 3. The Whole Process of Korean Semantic Role Labeling

Fig. 3. is the whole process of the SRL methodology proposed in this paper. The role labeling is mainly composed of four stages. First, we label the roles of arguments by using the case frame files provided by Korean PropBank. Arguments which cannot be labeled in case frame based SRL stage should be labeled by using suffix analysis. Arguments with a single morpheme are then labeled with statistics. Finally, arguments having suffix and single morpheme arguments which could not show 80% or higher accuracy are then labeled by using machine learning frameworks, such as support vector machines or conditional random fields.

3.1 Case frame-based SRL

Case frame-based semantic role labeling has very simple method. When a predicate 'si-do-ha-da' is found in a sentence, we look up the frame files to get the correct file about 'si-do-ha-da'. In this example, we find two arguments syntactically tagged as sbj and obj and then label the sbj argument as arg0 and label the obj argument as arg1. However, other arguments than sbj and obj could not be labeled and the arguments should be labeled in later stages.

Because a predicate word commonly has multiple senses, the appropriate sense should be selected to find appropriate case structure in a given context. Suppose that a predicate has senses, S_i, $1 \leq i \leq N$ in its frame file and in the sentence including the predicate has noun words, C_j, $1 \leq j \leq M$. Each sense S_i has an example sentence and noun words E^i_ks, $1 \leq k \leq L$ are extracted from the sentence. We select the most appropriate sense S_i as follows:

$$\arg max_i \frac{1}{M \times L} \sum_{k=0}^{L} \sum_{j=0}^{M} X(C_j, E_k^i) \quad (1)$$

where $X(\alpha, \beta)$ is an estimate function to calculate semantic similarity between two words α and β. The similarity function could be one of those which have been widely used [18, 19]. In this paper, we use the method described by [20].

For example, if we meet a predicate 'si-do-ha-da' and find that the sense is 01, then we extract arguments whose syntactic feature is sbj and obj and give them arg0 and arg1 roles, respectively. If there exist arguments other than sbj or obj, the arguments are passed to next stage to be labeled.

3.2 SRL using Suffix Structure

If an argument could not be labeled in the above stage, suffixes, josa (case suffix) and eomi (verbal ending), were used.

Arguments with a josa (a case suffix) were labeled with the dominant roles listed in Table 1. For example, since 97% of arguments containing '-Eul' are labeled to ARG1 in semantically annotated data, we label ARG1 to arguments ended with the josa '-Eul'.

An eomi is handled at the same way as a josa but a more complecated way. In Korean, an argument could be transformed from a predicate word and this transformed predicate could have Eomis (Verbal Ending) like a general predicate. This kinds of arguments could be divided into four cases, as shown in Fig. 4.

- Case 1 : Radix + Final
- Case 2 : Radix + Prefinal + Final
- Case 3 : Radix + Auxiliary predicate + Final
- Case 4 : Radix + Auxiliary predicate + Prefinal + Final

Fig. 4. A Predicate Having Eomis could be Classified into 4 Cases in View of Its Internal Components Combination.

In case 1, an argument could be composed of only a stem word and its final eomi. A word 'doe-eo' is composed of a verbal predicate 'doe-da' and a subordinate connection final eomi 'eo'. In this case, the argument could be labeled by the role of the highest percentage of '-eo', if the percentage is higher than 80%.

With Case 2, an argument is composed of a stem word, a prefinal eomi and final eomi. For example, a word 'bal-jeon-ha-yeot-deon' is broken down into a stem 'bal-jeon-ha-da', a past tense prefinal eomi 'yeot' and a adnomial transformation final eomi 'deon'. When we decide the role of this type of an argument, we extract dominant role percentages of its prefinal and final eomis and select the role with higher percentage. Of course, the higher percentage should be higher than 80% to be decided as its role.

Case 3 and case 4 of predicates have a stem word, an auxiliary predicate, and then eomi(es). An auxiliary predicate connects to its main predicate and complement the meaning of the predicate. Removing the main predicate will make a sentence invalid; however, removing an auxiliary predicate will not affect the validity of a sentence. For example, an argument 'but-eo-it-go' is composed of a stem word 'but-da', an auxiliary predicate 'eo-it' and a coordinate connection final eomi 'go' in case 3. For the case 4, we can take an example 'dop-go-i-seot-da'. This example could be decomposed of a stem word 'dop-da', an auxiliary predicate 'go-i', a past tense prefinal eomi 'seot' and a normal closing final eomi 'da'. Be-cause the auxiliary predicate could not affect the main meaning of the argument, we don't consider the auxiliary predicate in the role selection phase. Case 3 has a same selection method as case 1, and case 4 as case 2.

3.3 SRL of a single morpheme argument

If an argument doesn't have even a josa (case suffix) or an eomi (verbal ending) and is written with a single morpheme, this argument is to be labeled with a role of 80% or higher percentage in PropBank corpus as shown in Table 3. For example, because an argument 'Geu-reo-na' (in English 'but') has been labeled as ARGM-DIS with 100% percentage, we give a role ARGM-DIS whenever we meet an argument 'geu-reo-na'.

Table 3. Roles of Arguments Consisting of one Morpheme (Mapped to a Specific Role Morea than 80% of Accuracy)

Arguments	Role	Acc. (%)	Freq.
Geu-reo-na (But)	ARGM-DIS	100	271
Da (All)	ARGM-EXT	100	98
An (Not)	ARGM-NEG	98	1,335
Jal (Well)	ARGM-MNR	90	156

3.4 SRL by Machine Learning

Arguments which have not been labeled by above processes are to be assigned their roles by machine learning methods, a 2-level Support Vector Machine (SVM) like Fig. 5 or Conditional Random Fields (CRF) model in 3.4.3.

1) Features

We used general features independent on a specific language and Korean specific features proposed by [17].

2) Support Vector Machines (SVM)

When using the multiclass SVM from the beginning, it showed a lower performance because SVM predicts one class out of 17 classes at once. Therefore, it is better to first classify whether the role is ARGN or ARGM and then to classify the argument into one of the classes included in ARGN or ARGM. ARGN includes ARG0, ARG1, ARG2 and ARG3. In Korean PropBank, ARG4 has not been used. ARGM means an argument simply modifying a predicate. We used two types of SVM for each level, one is SVM-light and the other is multiclass SVM. The optimization algorithms used in SVM-light are descrived in [21]. While the SVM-light was used to predict whether the argument is labeled as ARGN or ARGM, the multiclass SVM predicts the detailed semantic role on the basis of the predicted results of earlier SVM.

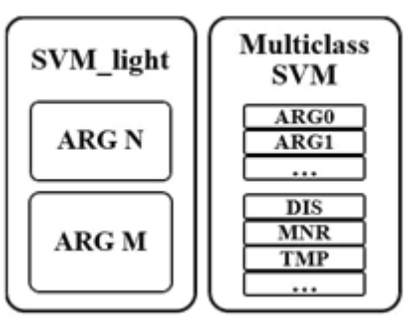

Fig. 5. 2-Level SVM

The multiclass SVM uses the multi-class formulation described in [22], but optimizes it with an algorithm that is very fast in the linear case. To solve this optimization problem, multiclass SVM uses an algorithm that is based on Structural SVMs [23]. For a training set (x_1,y_1) ... (x_n,y_n) with labels yi in [1..k], it finds the solution of the following optimization problem during training.

$$\min \frac{1}{2}\sum_{i=1}^{k} w_i * w_i + \frac{C}{n}\sum_{i=1}^{n}\epsilon_i$$
$$s.t.\ \text{for all}\ y \in [1..k]$$
$$[x_i \bullet w_{yi}] \geqq [x_i \bullet w_y] + 100 * \Delta(y_i, y) - \xi_i \ldots$$
$$[x_i \bullet w_{yn}] \geqq [x_n \bullet w_y] + 100 * \Delta(y_n, y) - \xi_n$$

w and ξ_i each denotes a parameter vector and slack variables. C is the usual regularization parameter that trades off margin size and training error. $(y_1; y)$ is the loss function that returns 0 if y_n equals y, and 1 otherwise.

3) Conditional Random Fields (CRFs)

A common classifier, like SVM, predicts the labels for a single argument without regard to the neigh-boring constituents. On the other hand, CRF model predicts its label considering the neighboring words. CRF is mainly used to label prediction, analysis of the text in the natural language. CRF can be described briefly as a conditional probability of the output sequence for the input sequence [24].

In a CRF, each feature function $f()$ is a function that takes in as input: a sentence s, the position i of a word in the sentence, the label l_i of the current word, and the label l_{i-1} of the previous word. Arguments labeled by CRF were predicted among 17 classes. Next, assign each feature function f_j a weight λ_j.

$$score(l|s) = \sum_{j=1}^{m}\sum_{i=1}^{n}\lambda_j f_j(s, i, l_i, l_{i-1}) \quad (2)$$

$$p(l|s) = \frac{\exp[score(l|s)]}{\sum_{l'}\exp[score(l'|s)]} = \frac{\exp[\sum_{j=1}^{m}\sum_{i=1}^{n}\lambda_j f_j(s, i, l_i, l_{i-1})]}{\sum_{l'}\exp[\sum_{j=1}^{m}\sum_{i=1}^{n}\lambda_j f_j(s, i, l'_i, l'_{i-1})]} \quad (3)$$

In Equation (2), we can now score a labeling l of given sentence s. The first sum runs over each feature function j, and the inner sum runs over each position i of the sentence. Finally, we can transform these scores into probabilities $p(l \mid s)$ between 0 and 1 by exponentiating and normalizing:

4. Experimental Results

This study used 10,000 sentences in the ETRI's syntactic annotated corpus and obtained 78,086 argu-ments. Approximately 80%, or 62,862 arguments, were used as the training data, and approximately 20%, or 15,224 arguments were used as the test data.

When predict semantic role with [17]'s method (only CRF) using the above data, we get 78.39% of F1 score. Whereas when predict semantic role with our method, we get 83.24% of F1 score (see Table 4).

First, we predict roles of 4,468 arguments, which have corresponding roles in frame files, with case frame-based method which yields 94.74% F1-score. About 10% of arguments appearing in the frame file fails to find correct semantic roles. A predicate could have multiple senses and each sense has its own case frame. An argument could select wrong predicate sense and wrong semantic role.

Among arguments on which SRL could not be per-formed with the case-frame based method, there were 1,721 arguments that had Josa (Case Suffix) which yielding F1-score of 90.24%. The greatest error is caused from the fact that a josa '-Eul' included in the objective argument is

Table 4. The Number of Labeling, Precision, Recall and F1-Score

Methods	Labeling	Precision (%)	Recall (%)	F1 score
State-of-the-art	15,224	78.08	79.04	78.39
Case-frame	4,468	90.00	100	94.74
Josa (Case Suffix)	1,721	82.22	100	90.24
Eomi (Verbal Ending)	3,415	78.65	100	88.05
One Morpheme	1,217	95.40	100	97.65
2-level SVM	4,403	55.63	62.56	58.89
CRF	4,403	59.92	61.40	60.67
Total-SVM	15,224	77.07	89.17	82.68
Total-CRF	15,224	78.31	88.84	83.24

mainly mapped to ARG1, but sometimes '-Eul' is mapped to ARG2. For example, when a josa '-Eul' is used in an argument of a predicate 'Dang-ha-da (suffer)' and 'Kyung-heom-ha-da (experience)', the argument should be annotated to ARG2, not ARG1.

Out of 3,415 arguments with one or more eomi(es) (Verbal ending), 2,686 arguments are correspond-ing to the manual annotation results, yielding 88.05% F1-score. 1,217 arguments with single morphemes yield 97.65% of F1. And finally, 4,403 arguments show 58.89% of F1-score and 60.67% of F1-score when they are input to machine learning algorithms, 2-level SVM and CRF respectively. In total, we acquire 83.24% of F1-score by using CRF algorithm, 0.56% points higher than SVM and 4.85% points higher than the state-of-art published by [17].

5. Conclusion

This paper is about Korean semantic role labeling by using a hybridized methodology, alleviating the dis-advantages of the simple case frame-based and CRF method. We utilize the suffix structure specialized in Korean language. F1-score has been increased compared to the case without suffix structure information analysis. Our method is intended to reduce cases to which machine learning methods should apply.

In the future, for each suffix structure analysis phase, we should find the ways to improve the per-formance in each stages. Also we should find the way to represent the suffix structure with a constant number of features. The features could be used when we apply a machine learning based method.

References

[1] V. Punyakanok, D. Roth, and W. Yih,. "The Importance of Syntactic Parsing and Inference in Semantic Role Labeling," *Computational Linguistics*, Vol.34, No.2, pp.257-287, 2008.

[2] L. Marquez, X. Carreras, K. C. Litkowski, and S. Stevenson, "Semantic Role Labeling: An Introduction to the Special Issue," *Computational Linguistics*, Vol.34, No.2, pp.145-159, 2008.

[3] S. Pradhan, W. Ward, K. Hacioglu, J. H. Martin, and D. Jurafsky, "Semantic Parsing using Support Vector Machines," *HLT-NAACL*, pp.233-240, 2004.

[4] H. A. Schwartz, F. Gomez, and C. Millward, "A Semantic Feature for Verbal Predicate and Semantic Role Labeling using SVMs," *FLAIRS Conference*, pp.213-218, 2008.

[5] T. Mitsumori, M. Murata, Y. Fukuda, K. Doi, and H. Doi, "Semantic Role Labeling using Support Vector Machines," Association for Computational Linguistics, pp.197-200, 2005.

[6] R. T. Tsai, W. Chou, Y. Su, Y. Lin, C. Sung, H. Dai, I. T. Yeh, W. Ku, T. Sung, and W. Hsu, "BIOSMILE: A Semantic Role Labeling System for Biomedical Verbs using a Maximum-Entropy Model with Automatically Generated Template Features," *BMC bioinformatics*, Vol.8, No.325, pp.1-15, 2007.

[7] N. Kwon, M. Fleischman, and E. Hovy, "Framenet-based Semantic Parsing using Maximum Entropy Models," *Proceedings of the 20th International Conference on Computational Linguistics*, 2004.

[8] T. Liu, W. Che, and S. Li, "Semantic Role Labeling with Maximum Entropy Classifier," *Journal of Software*, Vol.18, No.3, pp.565-573, 2007.

[9] Z. P. Jiang and H. T. Ng., "Semantic Role Labeling of NomBank: A Maximum Entropy Ap-proach," *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp.138-145, 2006.

[10] T. Cohn and P. Blunsom, "Semantic Role Labelling with Tree Conditional Random Fields," *Proceedings of the Ninth Conference on Computational Natural Language Learning*, pp.169-172, 2005.

[11] W. Aziz, M. Rios, and L. Specia, "Improving Chunk-based Semantic Role Labeling with Lexical Features," *Proceedings of Recent Advances in Natural Language Processing*, pp.226-232, 2011.

[12] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," *Proceedings of the 18th International Conference on Machine Learning*, pp.282-289, 2001.

[13] F. Sha and F. Pereira, "Shallow Parsing with Conditional Random Fields," *Proceedings of the Human Language Technology Conference and North American Chapter of the Association for Computational Linguistics*, pp.213-220, 2003.

[14] S. Arora, F. Lin, H. Shima, and M. Wang, "Tree Conditional Random Fields for Japanese Semantic Role Labeling," Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA., 2008.

[15] E. Moreau and I. Tellier, "The Crotal SRL System: A Generic Tool based on Tree Structured CRF," *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, pp.91-96, 2009.

[16] M. Palmer, D. Gildea, and P. Kingsbury, "The Proposition Bank: An Annotated Corpus of Seman-tic Roles," *Computational Linguistics*, Vol.31, No.1, pp.71-106, 2005.

[17] Y. Kim, H. Chae, B. Snyder, and Y. Kim. "Training a Korean SRL System with Rich Morphological Features," Association for Computational Linguistics (ACL). pp.637 - 642, 2014.

[18] P. Resnik, “Using Information Content to Evaluate Semantic Similarity in a Texonomy,” *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp.448-453, 1995.

[19] E. Terra and C. L A Clarke, “Frequency Estimates for Statistical Word Similarity Measures,” *Proceeding of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pp.165-172, 2003.

[20] M. Seok, C. Park, J. Kim, H. Song, and Y. Kim, “Korean Semantic Role Labeling using Korean PropBank Frame Files,” *Proceeding of the International Multi-Conference on Engineering and Technology Innovation*, 2015.

[21] T. Joachims, “Learning to Classify Text Using Support Vector Machines: The Springer International Series in Engineering and Computer Science,” New York, NY., 2002.

[22] K. Crammer and Y. Singer, “On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines,” *Journal of Machine Learning Research*, Vol.2, pp.265-292, 2001.

[23] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun, “Support Vector Ma-chine Learning for Interdependent and Structured Output Spaces,” *Proceedings of the Twenty-first International Conference on Machine Learning*, pp.104-111, 2004.

[24] C. Sutton and A. McCallum, “An Introduction to conditional random fields,” *Foundation and Trends in Machine Learning*, Vol.4, No.4, pp.267-373, 2006.

석 미 란

e-mail : smr4880@phill-it.com

2016년 한림대학교 융합소프트웨어학과 (공학석사, 학석사통합)

2016년~현 재 필아이티(주) IT사업본부 주임연구원

관심분야 : 자연언어처리, 가상입력시스템, 기계학습

김 유 섭

e-mail : yskim01@hallym.ac.kr

1992년 서강대학교 전자계산학과(공학사)

1994년 서울대학교 컴퓨터공학과 (공학석사)

2000년 서울대학교 컴퓨터공학과 (공학박사)

2002년~현 재 한림대학교 융합소프트웨어학과 교수

관심분야 : 자연언어처리, 기계학습, BioNLP, 의미 분석