HIERARCHICAL SPEECH-ACT CLASSIFICATION FOR DISCOURSE ANALYSIS

SANGWOO KANG

Department of computer science, Sogang University,
Seoul, 121-742, Republic of Korea
swkang@soang.ac.kr

Corresponding author : YOUNGJOONG KO

Department of Computer Engineering, Dong-A University,
840, Hadan 2-dong, Saha-gu,
Busan, 604-714, Republic of Korea
Tel.: +82-051-200-7782; fax: +82-051-200-7783.
youngjoong.ko@gmail.com

JUNGYUN SEO

Department of Computer Science & Interdisciplinary Program of Integrated Biotechnology,
Sogang University
Seoul, 121-742, Republic of Korea
seojy@sogang.ac.kr

The analysis of a speech act is important for dialogue understanding systems because the speech act of an utterance is closely associated with the user's intention in the utterance. This paper proposes a speech act classification model that effectively uses a two-layer hierarchical structure generated from the adjacency pair information of speech acts. The proposed model has two advantages when adding hierarchical information to speech act classification; the improved accuracy of the speech act classification and the reduced running time in the testing phase. As a result, it achieves higher performance than other models that do not use the hierarchical structure and has faster running time because Support Vector Machine classifiers can efficiently be arranged on the two-layer hierarchical structure.

***Keywords*:** natural language processing, discourse analysis, speech act classification, hierarchical structure, dialogue system.

1. Introduction

A dialogue system is a software program that enables a user to interact with a system using natural language (Lee et al., 2010). An essential task of the dialogue system is to understand what the user says. Because a speech act is a linguistic action intended by a speaker, the dialogue system must first identify speech acts that imply the user's intentions.

Some initial approaches for speech act classification have been based on knowledge such as recipes for plan inference and domain specific knowledge (Litman and Allen, 1987; Carberry, 1989). Since these knowledge-based models depend on costly handcrafted knowledge, it is difficult to extend them to more complex domains. Various machine learning approaches have been utilized to identify speech acts in order to overcome this problem (Samuel et al., 1999; Reithinger and Klesen, 1997; Choi et al., 1999). Recently, in many applications that require front-end speech recognition, the prosodic information as well as the lexical information is considered as a significant feature because this prosodic information contained in the speech signal can provide another source of complementary information (Dielmann and Renals, 2008; Laskowski and Shriberg, 2009; Huda et al., 2009; Levinson, 1983; Rangarajan et al., 2009).

Currently, research on hierarchical classification is receiving considerable attention from researchers. It seems natural to derive some hierarchy from many different kinds of speech acts in order to effectively discriminate between them. In general, the speech acts can be divided into several categories in a hierarchical structure. Therefore, we propose an effective speech act classification model with a two-layer hierarchical classification method. In our model, the hierarchy of speech acts is built up by the principle of the adjacency pair (Grosz, 1995; Levinson, 1983). The adjacency pair is defined as a pair of utterances that are adjacent and ordered as first and second parts, where a particular type of the first part requires a particular type of the second part: "ask-if," "ask-ref" and "ask-confirm" for the first part and "response" for the second part. Since most general dialogues are constructed by this principle, we can easily divide all the speech acts into several categories generated from each part of the adjacency pair with similar characteristics. We finally build a two-layer hierarchical structure of the speech acts; the first layer is composed of the adjacency pair types and one *other type*, and the second layer is organized by individual speech acts. Since this layered hierarchical structure is based on the principle of the adjacency pair which many actions in conversation are

accomplished through, it can be easily applied to most speech act classification tasks with various speech act sets. To verify the generality of our model, we use two different types of dialogue corpora in our experiments. Finally, our model improved performance in both of two corpora that are composed of different speech act sets and are constructed in different domains.

From the viewpoint of running time, the classification complexity of our model can be reduced because a range of classifications in the second layer is limited to one category in the first layer.

This paper is organized as follows. In section 2, we explain the two-layer hierarchical structure of speech acts and the proposed classification method. Section 3 describes our experimental results. The final section states the conclusions.

2. Related Work

Initial state of speech act classification was based on a rule that is extracted from a tagged dialogue corpus, such as linguistic rules, or dialogue grammar (Grosz, 19 95; Lee, 1996; Lambert, 1993). Lee proposed a two-step speech act classification s ystem that uses linguistic rules and a dialogue flow diagram; the first step involve s surface speech act classification, and the second step performs a deeper level of speech act classification. The surface speech acts are selected using linguistic infor mation of the current utterance and a linguistic rule that is extracted from a tagge d dialogue corpus. All possible surface speech acts are selected in this step of sur face speech act classification. In the deep speech act classification step, the most suitable speech act is selected from among the surface speech acts using contextual information, such as dialogue flow diagrams.

Rule-based speech act classification depends on handcrafted knowledge that is costly to produce, so it is not easy to scale up and expand the acts into domains. Recently, statistical speech act classification using a tagged dialogue corpus has be en proposed in order to solve such problems (Kim et al., 2004; Lee and Seo, 2002; Choi et al., 2005). Most previous works on speech act classification have used two feature types: sentential features and contextual features. Sentential features reflect linguistic characteristics, and are extracted from the surface utterance by a linguistic analyzer, such as a morphological analyzer, syntactic parser or semantic analyzer.

Contextual features reflect the relationship between the current utterance and the previous utterance. A syntactic pattern consists of the selected syntactic features of an utterance, which then approximate the utterance (Lee et al., 1997). In an actual dialogue, a speaker can express an identical meaning using different surface utterances, based on the speaker's personal linguistic background. A syntactic pattern generalizes these surface utterances based on their syntactic features. In this regard, Lee and Seo used sentence type, main verbs, auxiliary verbs and clue words to determine syntactic patterns, and elaborate the values of their syntactic features.

Many statistical models have been applied to speech act classification. For Korean speech act classification, a Hidden Markov Model (HMM) and Maximum Entropy Model (MEM) have been used as statistical models. Lee and Seo applied a bigram HMM in order to classify speech acts (Lee and Seo, 2002). They computed speech act probabilities for each utterance using a forward algorithm. When computing the speech act probabilities in order to find the best path in HMM, the problem of sparse data arises. To solve the sparse data problem, they smoothed the probabilities based on the class probabilities of decision trees. Choi et al. proposed a statistical dialogue classification model that can perform both speech act classification and discourse structure analysis using MEM (Choi et al., 2005). Their model can acquire discourse knowledge from a discourse-tagged corpus in order to resolve ambiguities. In addition, they defined the discourse segment boundary in order to represent the structural relationship of the discourse based on two consecutive utterances in a dialogue, and used them to statistically analyze both the speech act of an utterance and the discourse structure of a dialogue.

3. Speech Act Classification by Using a Two-layer Hierarchical Structure of Speech Acts

3.1. *Two-layer hierarchical structure of speech acts*

An adjacency pair is an example of conversational turn-taking. An adjacency pair is composed of two utterances by two speakers, one following the other. The spe aking of the first utterance (the first part of the pair; the first turn) provokes a re

sponding utterance (the second part of the pair; the second turn).

In this study, speech acts in the first layer are divided into 3 categories: *Question, response* and *other type*. These categories are assigned according to the characteristics of each part of the adjacency pair. The *question* and *response types* are parts of the adjacency pair, and the *other type* is the category for speech acts that can be uttered alone. The second layer consists of speech acts that are involved in each category of the first layer. In the end, we grouped 16 speech acts into these 3 categories.

The *question type* corresponds to the first part of the adjacency pair and its utterances are active, like the demand for information. This type includes "question" ("ask-if," "ask-ref," "ask-confirm"), "suggest," "offer" and "request." The *response type* corresponds to the second part of the adjacency pair and its utterances are the responses to the first part. This type includes "accept," "reject," "response," "acknowledge," "express" and "promise." Finally, the *other type* consists of speech acts that can be independently used without being a pair. This type includes "opening," "closing" and "introducing-oneself."

Since adjacency pair is a basic principle to make conversational turn-taking, our two-layered hierarchical structure has an advantage that it can be adapted to the other dialogue analyses. Table 1 shows the two-layer hierarchical structure of the speech acts used in our corpus.

[Table 1. Two-layer hierarchical structure of speech acts.]

3.2. *Speech act classification*

For speech act classification, the speech act of current utterance can be expressed by Eq. (1) (W. Choi, 2005).

$$SA(U_i) \approx argmax_{s_{i,j}} P(S_{i,j}|F_i)P(S_{i,j}|SA(U_{i-1})) \quad (1)$$

$SA(U_i)$ denotes the speech act of the i^{th} utterance (U_i) and $S_{i,j}$ denotes j^{th} candidate speech act of the i^{th} utterance (U_i), given a dialogue including n utterances. Since we assume that the current speech act is dependent on the sentential features set (F_i) of current utterance (U_i) and the speech act ($SA(U_{i-1})$) of the previous utterance (U_{i-1}).

Sentential features contain lexical and morphological informative clues for determining the current speech act, and the previous speech act provides contextual information.

The feature extraction method proposed by Kim et al. is used in this model and it has exhibited the best performance in Korean speech act classification (Kim et al., 2004). This method assumes that the sentential features in an utterance are extracted from the lexical information of clue words and the sequence of Part-of-Speech (POS) tags, and these features provide very effective information for analyzing the speech acts of utterances. As a result, the sentential features are composed of words annotated with POS tags and POS bi-grams of all the words in an utterance; these features can be extracted by only using a morphological analyzer. Only a speech act tag of previous utterance by the Markov assumption is also used as a contextual feature.

Speech act classification is not a problem of finding an optimum path of speech acts throughout all the utterances of a dialogue, because dialogue analysis has already been carried out in real time in dialogue systems. Thus, in our model, we employ a Support Vector Machine (SVM) (Vapnik, 1995), which has been widely used and has demonstrated significant performance in various learning tasks (Kim et al., 2011), since HMM and Conditional Random Fields (CRFs) are not appropriate classification models for speech act classification. In our experimental settings, given the user's input in real time, the speech act of the input utterance is detected, rather than looking for a full sequence of speech acts that constitute a dialogue. This is the same environment as that used in a real dialogue system. It is impossible for a speech act system in a real dialogue system to analyze the sequence of all the speech acts of the entire conversation, because the system cannot foresee the whole conversation.

Therefore, in this study, it was not effective to use HMM or CRFs for optimized sequential labeling, because entire conversations were not obtained in our experimental environment. These models can also cause slowdowns due to the large amount of computation involved.

Equation (2) for SVM represents the equation of the hyper-plane in a high-dimensional space called the kernel space.

$$f_x = W^T \bullet X + b = 0 \tag{2}$$

If X is the vector of the features, then the discriminant function is given by f_x. • denotes the inner product, and b is a constant. The vector W is a normal vector that is perpendicular to the hyper-plane. The SVM is designed such that $0 < f_x$ for positive examples and $0 > f_x$ for negative examples as long as the data is separable.

In addition, we use the binary feature-weighting scheme, which is known to perform well in speech act classification, because each feature in an utterance rarely occurs more than once. Therefore, the vector X of each utterance in this model consists of the speech act ($SA(U_{i-1})$) of previous utterance and the sentential feature set ($F_i=\{f_{i,1}, ...f_{i,n}\}$) and each element of this vector is represented by a binary feature- weighting scheme as shown in the following Eq. (3).

$$W_{i,k} = 0 \text{ (if nonexistent) or } 1 \text{ (otherwise)} \quad (3)$$

In general, the SVM model requires as many classifiers as the number of speech acts to be classified, because SVM typically provides only a binary decision function. Figure 1 illustrates and compares a flat structure (a) and a hierarchical structure (b) of speech acts. In our research, a hierarchical classification method using an SVM is applied to the speech act classification. As can be seen Figure 1, our hierarchical classification model is constructed using 19 SVMs. Test examples (utterances) are passed through the classifiers of the first layer (3 SVM classifiers), and are then designated as one type among the 3 types of speech acts (*question, response* and *other type*) of the first layer.

[Fig. 1. Speech acts trees for flat and hierarchical classifications]

The test utterances are finally classified into one speech act among the speech acts included in the assigned type by the classifiers of the second layer*; Question, response* and *other type* consist of 6, 7 and 3 SVM classifier, respectively. Finally, we can observe that the proposed model is able to employ fewer classifiers than are used by flat classification models.

For example, the flat classification needs a total of 16 classification tasks for 16 speech acts, whereas the hierarchical classification needs at most 10 classification tasks, as shown in Figure 1. Therefore, we think that the proposed model can be more an efficient speech act classification model in real-time systems like dialogue systems in particular.

4. Experimental Evaluation

4.1. *Data set*

We use two different types of Korean dialogue corpus corpora for applying our hierarchal strategy to various dialogue environments. And both corpora are trained and tested separately. The first one is collected from real fields including hotel, airline and tour reservations. This corpus consists of 528 dialogues (19.5 utterances per dialogue and 16 speech act) and 10,281 utterances (training data (8,349) and test data (1,932)). Each utterance in the dialogues was manually annotated with a speaker (SP) and a speech act (SA). Table 2 shows a part of the annotated dialogue corpus used in the experiment. Especially, this corpus was used to compare the performance of previous models besides testing our model.

[Table 2. Part of the annotated dialogue corpus]

The second corpus is collected from different domain of schedule management and consists of different speech acts when comparing with the first corpus. This corpus consists of 954 dialogues (22.3 utterances per dialogue and 11 speech acts) and 21,310 utterances (train data (17,054) and test data (4,256)).

4.2. *Experimental results*

4.2.1. *Comparing the proposed model with the baseline model*

In order to evaluate the proposed model, we implemented a baseline model with a flat structure of speech acts. We can also calculate the F_1-score and the cost measure for each speech act in our experimental environment (Deisy, 2010). F_1-score is given in Eq. (4) and is the harmonic mean of the precision (Eq. (5)) and the recall (Eq. (6)).

$$F_1 - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

$$Precision = \frac{number\ of\ correctly\ classified\ speech\ acts}{number\ of\ speech\ acts\ classified} \quad (5)$$

$$Recall = \frac{number\ of\ correctly\ classified\ speech\ acts}{number\ of\ correct\ speech\ acts} \quad (6)$$

The cost measure is the rate of misclassification of speech acts. The cost measure is given in Eq. (7), and the classification cost considers both of the miss (Eq. (8)) and false_alarm (Eq. (9)) of the test in order to compute the score.

$$Cost\ measure = miss + false\_alarm \quad (7)$$

$$Miss = \frac{number\ of\ incorrect\ speech\ acts\ that\ are\ classified}{number\ of\ correct\ speech\ acts} \quad (8)$$

$$False\_alarm = \frac{number\ of\ correct\ speech\ acts\ that\ are\ misclassified}{number\ of\ incorrect\ speech\ acts} \quad (9)$$

The proposed model achieved higher performance, higher F_1-scores and lower costs, than the baseline model through almost all the speech acts in both corpora, as can be seen from Table 3.

[Table 3. Comparing F_1 and Cost measures for baseline and proposed models in individual speech act: B.model, P.model are the baseline and proposed model respectively]

In several speech acts, the performances of the proposed model are lower than those of the baseline models. We observed that most of uncorrected utterances in “accept” and “reject” speech acts (the first corpus) and “inform” speech act (the second corpus) were misclassified into “response” speech act. The distributions of speech acts in both corpora are biased toward the “response” speech act. In addition, the SVM classifier in hierarchical classification is certainly learned from more skewed distribution of speech acts in the case of the *response type* in the second layer because the portion of the “response” speech act in the *response type* is bigger than that of total corpus. We can

observe that the “confirm” and “opening” speech acts have zero performance. They occur only 5 and 6 times in the second corpus, and it made that kind of poor performance.

Table 4 shows the differences in the micro- and macro-average accuracy scores between the proposed and baseline models.

[Table 4. Performance differences between the proposed and baseline models]

Micro-average accuracy was calculated by dividing the number of correctly classified utterances by the total number of classified utterances, whereas macro-average accuracy was calculated as the average of the accuracy values of all the speech acts. The micro-average accuracy of the proposed model was 3% and 2% higher than that of the baseline model in the first and second corpora, respectively, and the macro-average accuracy of the proposed model was 5% and 9% higher than that of the baseline model in the first and second corpora, respectively. It means that we achieved more improvement in speech acts (e.g., “offer”) with a small number of utterances than in those (e.g., “response”) with a large number of utterances.

4.2.2. *Comparing the proposed model with other previous models*

This section explains the results obtained using the proposed model and other, previous speech act analysis models. Table 5 shows these other, previous models of different types, and their performance.

[Table 5. Performance of the proposed model and other previous models on the first corpus]

The first and second model used rule-based approaches. The first model defined rules such as dialogue transition networks in order to apply the structural information of a discourse (Lee et al., 1997). The second model used a fuzzy trigram model (Kim and Seo, 2003), which used a membership function in fuzzy set theory instead of conversational probability distributions. They were not, however, adequate to deal with a variety of dialogues, since they used a restricted rule-based model. Furthermore, these rule-based models were not better than statistical models such as HMM or MEM.

The third model used a smoothed HMM, which combined HMM and decision trees (Lee and Seo, 2002). This model computed the speech act probabilities for each utterance,

using a forward algorithm. Decision trees provided the observation probabilities and transition probabilities, and were constructed based on syntactic patterns. The fourth model exploited MEM (Choi et al., 2005). This model used discourse information drawn from a discourse-tagged corpus. The last model (the proposed model) used an SVM model with a two-layer hierarchical structure. The proposed model only used the sentential features, which were composed of words annotated with POS tags and POS bi-grams. As a contextual feature, the speech act tag of a previous utterance was also used.

We report the performance of each model based on the use of the same test data set (the first corpus) and an evaluation metric that are used in this paper. The proposed model applies a small feature set to be easier to extract than other statistical models, and shows significantly better performance than MEM and HMM.

5. Discussions

Most dialogue systems are designed to achieve the objectives of the user. Thus, the conversation involves repetitive questions (or requests) and responses. In our paper, the first layer of the speech act hierarchy consists of *question*, *response* and *other types* of speech acts, and actual dialogues are generated by a pair of a *question* and *response type* in a real dialogue system. A *question type* implies the intent to ask for information, while *response types* can assume the form of a variety of appropriate responses to a *question type*. Even if the classification result of a speech act in the second layer is incorrect, the user intent of the dialogue system (asking for information) is maintained (in Table 6, see the high degrees of accuracy in the first layer). As a result, the dialogue system can generate some kind of response, even if it is more difficult to generate an exact response. In Table 6, all the degrees of accuracy of the first-layer classification exceed 0.96 in both corpora.

Table 6 shows the differences between the performance of the proposed and baseline models in each layer of the hierarchical structure.

[Table 6. Performance differences in each layer of the hierarchical structure]

The proposed model can minimize the difficulty of dealing with errors in the second layer because the classification performance of the first layer is much higher. Although

an input utterance may ultimately be misclassified in the second layer, it is most likely to be correctly classified in the first layer. That is, a misclassified utterance probably contains one speech act that is semantically similar to a correct speech act. Thus, we expect that our dialogue manager can handle misclassified utterances more easily. In addition, when the proposed method is applied in real dialogue systems, the task success rate of the system's operation can be expected to be greatly improved.

We here verify that the running time of our model in the testing phase is much faster than that of the baseline model; our model needs only about 40% of running time of the baseline model. In the first corpus, the running time of the baseline model and proposed model are 1.57 and 0.63 second respectively and are 3.35 and 1.28 second in the second corpus: 1,932 test utterances in the first corpus and 4,256 test utterances in the second corpus. That is the reason why the proposed model can use fewer classifiers than the baseline model.

6. Conclusions

We proposed an effective speech act classification using two-layer hierarchical model. For constructing this model, we used the adjacency pair principle. The proposed model showed higher performance than the baseline model without the hierarchical structure and other previous models. In addition, the proposed model has a faster running time.

Acknowledgements

This work was supported by the IT R&D program of MKE/KEIT. [10041678, The Original Technology Development of Interactive Intelligent Personal Assistant Software for the Information Service on multiple domains] and this research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2009-0065895).

References

A. Dielmann and S. Renals, 2008, Recognition of Dialogue Acts in Multiparty Meetings Using a Switching DBN, *IEEE Trans. Audio, Speech and Language Processing*, 16(7), pp. 1303–1314.

B. Grosz, Discourse and Dialogue, 1995, In *Survey of the State of the Art in Human Language Technology*, Center for Spoken Language Understanding, pp.227- 254.

C. Deisy, 2010, A Novel Term Weighting Scheme MIDF For Text Categorization, *Journal of Engineering Science and Technology*, 5(1), 99. 94-107

C. Lee, S. Jung, K. Kim, D. Lee and G. Lee, 2010, Recent Approaches to Dialog Management for Spoken Dialog Systems, *Journal of Computing Science and Engineering*, 4(1), pp. 1-22.

D. Litman and J. Allen, 1987, A Plan Recognition Model for Subdialogues in Conversations, *Cognitive Science*, 11, pp. 163-200.

H. Kim and J. Seo, 2003, An Efficient Trigram Model for Speech Act Analysis in Small Training Corpus, *J. Cognitive Science*, 4(1), pp.107–120.

H. Kim, C. Seon, J. Seo, 2011, Review of Korean Speech Act Classification: Machine Learning Methods, *Journal of Computing Science and Engineering*, 5(4), pp. 288-293.

H. Lee, *Analysis of Speech Act for Korean Dialogue Sentences*, 1996, MS Thesis, Sogang University.

J. Lee, J. Seo and G.C. Kim, 1997, A Dialogue Analysis Model with Statistical Speech Act Processing for Dialogue Machine Translation, In *Proceeding Spoken Language Translation Workshop in conjunction with EACL*, pp.10–15.

K. Kim, H. Kim and J. Seo, 2004, A Neural Network Model with Feature Selection for Korean Speech Act Classification, *International Journal of Neural System*, 14(6), pp. 407-414.

K. Laskowski and E. Shriberg, 2009, Modeling Other Talkers for Improved Dialog Act Recognition in Meetings, In *Proceeding Interspeech*, pp. 2783–2786.

K. Samuel, S. Carberry and K. Vijay-Shanker, 1999, Automatically Selecting Useful Phrases for Dialogue Act Tagging, In *Proc. 4th Conference of the Pacific Association for Computational Linguistics*.

L. Lambert, 1993, *Recognizing Complex Discourse Acts: A Tripartite Plan-based Model of Dialogue*, Ph.D. thesis, The University of Delaware.

N. Reithinger and M. Klesen, 1997, Dialogue Act Classification Using Language Model, In *Proc. Of the Enrospeech*, pp.2235-2238.

S. Carberry, 1989, A Pragmatics-Based Approach to Ellipsis Resolution, *Computational Linguistics*, 15(2), pp. 75-96.

S. Huda, J. Yearwood and R. Togneri, 2009, A Stochastic Version of Expectation Maximization Algorithm for Better Estimation of Hidden Markov Model. *Pattern Recognition Letters*, 30(14), pp. 1301-1309.

S. Levinson, 1983, *Pragmatics*, Cambridge University Press, Cambridge, UK.

S. Lee and J. Seo, 2002, A Korean Speech Act Analysis System Using Hidden Markov Model with Decision Trees, *International Journal of Computer Processing of Oriental Languages*, 15(3), pp. 231-243.

S. Rangarajan, S. Narayanan and S. Bangalore, 2009, Modeling the Intonation of Discourse Segments for Improved Online Dialog Act Tagging, In *Proceeding ICASSP*, pp. 5033–5036.

V. Vapnik, 1995, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York.

W. Choi, J. Cho and J. Seo, 1999, Analysis System of Speech Act and Discourse Structures Using Maximum Entropy Model, In *Proc. of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pp.230-237.

W. Choi, H. Kim and J. Seo, 2005, An Integrated Dialogue Analysis Model for Determining Speech Acts and Discourse Structures. *IEICE Trans. on Information and Systems*, E88-D(1), pp.150-157.

Table 1. Two-layer hierarchical structure of speech acts.

First layer	Second layer
Question type	Ask-if *(e.g., Is the price of meals included in hotel charges?)* Ask-ref *(e.g., What kind of room do you want?)* Ask-confirm *(e.g., Is your signature right?)* Offer *(e.g., Would you like me to show you how?)* Suggest *(e.g., How about Hawaii?)* Request *(e.g., Please, reserve a seat on that flight.)*
Response type	Accept *(e.g., Yes, please.)* Reject *(e.g., I don't want.)* Response *(e.g., A single room, please)* Acknowledge *(e.g., Yes, all right.)* Inform *(e.g., I have some questions about lodgings.)* Express *(e.g., Sorry, we couldn't help you fast.)* Promise *(e.g., I'll make arrangements for you to be met at the airport.)*
Other type	Closing *(e.g., Thank you, see you again.)* Opening *(e.g., Hello?)* Introducing-oneself *(e.g., Good morning, this is travel agency.)*

Table 2. Part of the annotated dialogue corpus

(Speaker (SP), Korean (KS), English (EN), Speech Acts (SA))

Tag	Value
SP	Customer
KS	미국 조지아대 어학연수에 참가 신청을 한 학생인데요.
EN	I'm a student, and I'm registered for a language course at the University of Georgia in the U.S.
SA	Introducing-oneself
SP	Customer
KS	숙소에 관해서 문의할 사항이 있어서요.
EN	I have some questions about lodgings.
SA	Request
SP	Clerk
KS	조지아대학의 어학연수 코스는 대학에 기숙사를 제공하고 있습니다.
EN	There is a dormitory in the University of Georgia for language course students.
SA	Response
SP	Customer
KS	그럼 식비는 연수비에 포함이 되어 있는 건가요?
EN	Then, is a meal included in the tuition fee?
SA	Ask-if

Table 3. Comparing F_1 and Cost measures for baseline and proposed models in individual speech act: B.model, P.model are the baseline and proposed model respectively.

Speech act	1st corpus				Speech act	2nd corpus			
	F_1		Cost			F_1		Cost	
	B.model	**P.model**	B.model	**P.model**		B.model	**P.model**	B.model	**P.model**
ask-if	0.760	**0.861**	0.421	**0.149**	ask-if	0.745	**0.810**	0.956	**0.212**
ask-ref	0.849	**0.921**	0.215	**0.097**	ask-ref	0.940	**0.940**	0.814	**0.117**
ask-confirm	0.880	**0.939**	0.204	**0.061**	ask-confirm	0.000	**0.490**	0.632	**0.546**
offer	0.000	**0.167**	1.421	**0.875**	confirm	0.000	**0.000**	1.221	**1.223**
suggest	0.559	**0.500**	0.498	**0.622**	request	0.849	**0.910**	0.176	**0.144**
request	0.810	**0.748**	0.198	**0.422**	accept	0.424	**0.805**	0.282	**0.218**
accept	0.695	**0.492**	0.447	**0.680**	response	0.875	**0.935**	0.826	**0.112**
response	0.815	**0.971**	0.312	**0.035**	inform	0.530	**0.465**	0.542	**0.609**
reject	0.765	**0.691**	0.361	**0.409**	express	0.965	**0.944**	0.629	**0.107**
acknowledge	0.850	**0.934**	0.236	**0.078**	opening	0.000	**0.000**	1.311	**1.311**
inform	0.705	**0.756**	0.442	**0.188**	greeting	0.910	**0.930**	0.161	**0.112**
express	0.710	**0.773**	0.418	**0.230**					
promise	0.709	**0.949**	0.489	**0.051**					
closing	0.720	**0.697**	0.347	**0.357**					
opening	0.970	**0.975**	0.049	**0.048**					
introduce-oneself	0.965	**0.989**	0.161	**0.021**					
Macro-average	0.735	**0.773**	0.389	**0.270**		0.567	**0.657**	0.686	**0.428**

Table 4. Performance differences between the proposed and baseline models

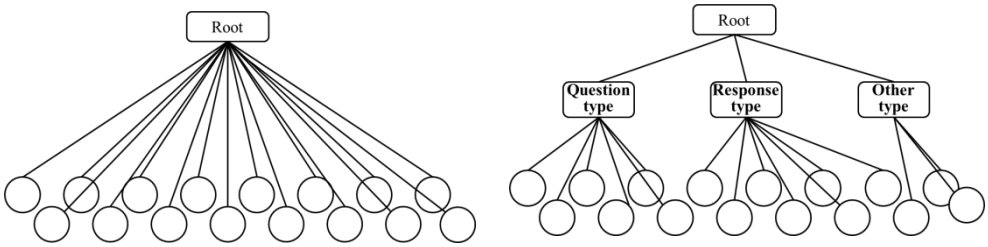
Model	1st corpus		2nd corpus	
	Micro-average	Macro-average	Micro-average	Macro-average
Baseline model	0.82(1579/1932)	0.74	0.89(3790/4256)	0.58
Proposed model	0.85(1633/1932)	0.79	0.91(3870/4256)	0.67
Improvement	**+3%**	**+5%**	**+2%**	**+9%**

Table 5. Performance of the proposed model and other previous models in the first corpus

Classification model	Feature set	Measurement	Score
Discourse analysis model with dialogue a transition network	Grammar by dialogue transition network	accuracy	0.75
Trigram model with a membership function	Word trigram	accuracy	0.77
HMM with decision trees	Syntactic pattern	accuracy	0.82
MEM with discourse information	Syntactic pattern, discourse structure tag	accuracy	0.83
SVM (Proposed model) with a hierarchical structure	Clue word and POS, POS bigram	accuracy	0.85

Table 6. Performance differences in each layer of the hierarchical structure

Model	1st corpus		2nd corpus	
	Micro-average in the 1st layer	Micro-average in the 2nd layer	Micro-average in the 1st layer	Micro-average in the 2nd layer
Baseline model	-	0.82(1579/1932)	-	0.89(3790/4256)
Proposed model	**0.96(1844/1932)**	**0.85(1633/1932)**	**0.97(4125/4256)**	**0.91(3870/4256)**



(a) (b)

Fig. 1. Speech acts trees for flat and hierarchical classifications