International Journal of Computer Processing of Oriental Languages, Vol. 15, No. 3 (2002) 231–243
© Chinese Language Computer Society & World Scientific Publishing Company

Korean Speech Act Analysis System Using Hidden Markov Model with Decision Trees

SONGWOOK LEE* AND JUNGYUN SEO†

Department of Computer Science, Sogang University
1 Sinsu-dong Mapo-gu, Seoul 121-742, Republic of KOREA
*gospelo@nlprep.sogang.ac.kr
†seojy@ccs.sogang.ac.kr

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

Analyzing speech act is important for understanding the intention of a speaker and the flow of dialogue in discourse analysis. This study analyzes speech act by using syntactic pattern information which is tagged in dialogue corpus. We apply bigram Hidden Markov Model(HMM) to the task of computing speech act. We compute speech act probabilities for each utterance with forward algorithm. When computing the speech act probabilities to find the best path in HMM, there arises a sparse data problem. We smoothed probabilities through class probabilities of decision trees.

Keywords: Speech Act Analysis; HMM; Forward Probability; Decision Tree; Class Probability.

1. Introduction

In a natural language dialogue, speech act is an abstraction of a speaker's intention which the speaker wants to represent through utterance. The computer system must consider the semantic information and the flow of dialogue to interpret speech act. However, it is difficult to infer the speech act from a surface utterance because the same utterance may represent a different speech act according to the context.

Recently, there have been many works for the speech act analysis utilizing tagged dialogue corpus with machine learning methods [1]–[9]. [5] analyzes speech act and discourse structure by using Maximum Entropy Model. Most of the previous works are based on features such as cue phrases, change of speaker, short utterances, utterance length, speech acts n-grams, and word n-grams. We also use many syntactic features which are tagged in dialogue corpus. [6]

Correspondence should be sent to: Songwook Lee, gospelo@nlprep.sogang.ac.kr

investigates a method of automatically selecting cue phrases which have useful information for speech act tagging. [7] and [8] interprets speech act via decision trees and obtains linguistic rules which decide speech acts. The decision tree chooses useful features from all given features to construct a tree. However, cumulative error occurs in some previous works which use speech act n-grams as a context [2, 5, 7, 8]. For example, cumulative error occurs in the decision tree model which use the previous utterance's speech act as a context to classify current utterance's speech act. If an error occurs in the current utterance's speech act analysis then the error is propagated to the next utterance's speech act classification phase and the error is used by decision tree as the wrong context. Consequently the errors are cumulated until a dialogue ends [7, 8]. By the way, in HMM, this error propagation does not occur since n-gram contexts are used in the state transitions. The HMM chooses the best sequence of states after trying all possible state transitions. [9] uses HMM in the speech act classification and smoothes transition probability of speech act n-grams with backed off method. [9] obtains observation probability from the combination of speech recognized words, true words and prosodic information by parameter estimation.

[10] proposes decision tree to smooth trigram transition probability for the POS tagging. We also use it for solving sparse data problem which occurs in the HMM. In this paper, we reflect all possible previous utterances' speech acts as a context to analyze the current utterance's speech act by navigating all possible paths in the HMM. Because of the large number of parameters, this method has difficulties in accurately estimating small probabilities from limited amounts of training data. We acquire observation probabilities and transition probabilities by using two decision trees to get reliable estimates of probabilities. These probabilities are used in forward algorithm to analyze speech acts. Decision tree determines automatically the size of the context which should be used to estimate probabilities. Thus, we can avoid the sparse data problem. We use CART as a tool for constructing a decision tree [11, 12]. The main difference between our proposed model and [9] is that we use decision trees to solve sparse data problem in the speech act analysis.

2. Dialogue Corpus

We use [5]'s Korean dialogue corpus which was transcribed from recordings in domains such as hotel reservation, airline reservation, and tour reservation. This corpus consists of 528 dialogues with 10,285 utterances. Each utterance in a dialogue is manually annotated with discourse knowledge such as speaker (SP), syntactic pattern (ST), speech acts (SA) and discourse structure (DS) information.

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

SP has a value of either “User” or “Agent”. ST consists of the selected syntactic features of an utterance. Surface utterances can be different, but can convey the same meaning. ST is used to generalize them. It includes *Sentence Type*, *Main-Verb*, *Aux-Verb*, *Tense*, *Negativeness of Sentence* and *Clue-Word*. *Main-Verb* has lexical words which are performative [2, 5].

We will use SA, ST, and SP which are extracted from the annotated dialogue corpus for our experiment. We do not consider discourse structure. We use 17 types of speech acts that appear in the dialogue corpus. Figure 1 shows a part of the annotated dialogue corpus. Table 1 shows the syntactic features of a syntactic pattern.

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

/SP/User /KS/**미국 조지아대 어학연수에 참가 신청을 한 학생인데요**.(*Migug Georgia-dae eo-hag-yeon-su-e cham-ga sin-cheong-eul han hag-saeng-in-de-yo*.) (I’m a student and registered for a language course at University of Georgia in U.S.) /ST/[decl,be,present,no,none,none] /SA/introducing-oneself /DS/[2] /SP/ User /KS/**숙소에 관해서 문의할 사항이 있어서요**.*(Sug-so-e gwan-hae-seo mun-ui-hal sa-hang-i iss-eo-seo-yo*.) (I have some questions about lodgings.) /ST/[decl,paa,present,no,none,none] /SA/ask-ref /DS/[2] ->Continue	/SP/Agent /KS/**조지아대학의 어학연수 코스는 대학에 기숙사를 제공하고 있습니다**.(*Georgia-dae-hag-ui eo-hag-yeon-su course-neun dae-hag-e gi-sug-sa-reul je-kong-ha-go iss-seub-ni-da*.) (There is a dormitory in University of Georgia for language course students.) /ST/[decl,pvg,present,no,none,none] /SA/response /DS/[2] /SP/User /KS/**그럼 식비는 연수비에 포함이 되어 있는 건가요**?(*Geu-reom sig-bi-neun yeon-su-bi-e po-ham-i doe-eo iss-neun-geon-ga-yo?*) (Then, is meal included in tuition fee?) /ST/[yn_quest,pvg,present,no,none,then] /SA/ask-if /DS/[2.1]

Figure 1. A part of the anonotated dialogue corpus.

Table 1. Syntactic features used in the syntactic pattern.

Syntactic feature	Example	Total
Sentence Type	yn_quest, decl, wh_quest, imperative	4
Main-Verb	be, know, ask, promise, etc.	88
Tense	present, future, past	3
Negativeness	no, yes	2
Aux-Verb	want, will, possible, serve, seem, intend, etc.	31
Clue Word	yes, okay, and, then, hello, instead of, etc.	26

3. HMM for Speech Act Tagging

Let U_i denote i_{th} utterance in a dialogue. Each of the syntactic features, speakers, and speech acts of U_i are represented as follow:

SA_i	: Speech Act of U_i
SP_i	: Speaker of U_i
SenType	: Sentence type of U_i
MainV	: Main-Verb of U_i
Tense	: Tense of U_i
Neg	: Negativeness of U_i
AuxV	: Aux-Verb of U_i
Clue	: Clue-Word of U_i

3.1. HMM

We adopt the method which solved POS tagging problem in the analysis of speech act. Generally, POS tagging problem is to find the sequence of POS for a given sequence of words. Let $W_1,...,W_n$ be a sequence of words and $C_1,...,C_n$ be a possible sequence of POS. POS tagging problem is to find the sequence of $C_1,...,C_n$ which maximizes (1) [13]. ((1) is a case of first order Markov Model.)

$$\Pi_{i=1,...,n} P(C_i|C_{i-1})*P(W_i|C_i) \quad (1)$$

Let us change (1) to solve speech act problem. Let $U_1,...,U_n$ be a sequence of utterance of a dialogue. SA_i is a speech act of U_i, Analyzing speech act can be defined as finding $SA_1,...,SA_n$ which maximizes (2) for a $U_1,...,U_n$. Like POS tagging, we assume that the current speech act is affected only by the previous speech act, and assume that an utterance appears in a speech act independent of the utterances in the preceding or succeeding speech acts.

$$\Pi_{i=1,...,n} P(SA_i|SA_{i-1})*P(U_i|SA_i) \quad (2)$$

In order to use (2), the whole utterance of a certain dialogue should be given as an input, prior to the determination of speech act. Therefore, it is difficult to apply this method to the determination of speech act. On the face of it, we resort to forward algorithm which considers the input only up to a specific state. The forward algorithm needs current utterance and previous utterances as input. Figure 2 is an algorithm which is transformed to solve speech act problem from POS tagging problem [14]. The speech act probablity that utterance U

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

appears in a speech act category SC_t out of $SC_1,...,SC_T$ could be estimated by Step 3. We choose the speech act for an utterance which has a maximum speech act probability.

Step1: Initialization

For t = 1 to T do

$$SEQSUM(t,1) = P(P_1 \mid SC_t) * P(SC_t \mid \phi)$$

Step2: Computing the Forward Probabilities

For i = 2 to N do

For t = 1 to T do

$$SEQSUM(t,i) = \sum_{u=1,T} (P(SC_t \mid SC_u) * SEQSUM(u,i-1)) * P(U_i \mid SC_t)$$

Step3: Computing the Speech act Probabilities

For i = 1 to N do

For t = 1 to T do

$$P(SA_i = SC_t) = \frac{SEQSUM(t,i)}{\sum_{u=1,T} SEQSUM(u,i)}$$

Figure 2. The forward algorithm for computing the speech act probabilities.

$P(U_i|SC_t)$ at Step2 in Figure 2 is an observation probability for a given utterance. Since a kind of utterance is infinite in a real dialogue, we approximate the utterance as speakers and a syntactic pattern P_i.

$$P(U_i|SC_t) := P(P_i|SC_t) \tag{3}$$

$$P(P_i|SC_t) = P(SP_{i-2}, SP_{i-1}, SP_i, SenType, MainV, Tense, Neg, AuxV, Clue \mid SC_t) \tag{4}$$

(4) could be estimated by counting the observation which occurs at the same time. This estimation is problematic since many frequencies are so small that the corresponding probabilities cannot be estimated reliably. Some difficulties are caused by zero frequencies. It is difficult to decide whether the corresponding observation is incorrect or just rare. So (4) is often modified by replacing zero probabilities with a small value and renormalizing the probabilities, so that they sum up to 1 [15]. We will solve this sparse data problem by using decision tree.

3.2. Smoothing probability with a decision tree

Our proposed model is based on HMM with the use of decision tree. We acquire $P(P_i|SC_t)$ from class probabilities of a trained decision tree.

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

$P(P_i|SC_t)$ in (3) could be rewritten as (5) according to Bayes' rule.

$$P(P_i|SC_t) = P(SC_t|P_i)*P(P_i)/P(SC_t) \quad (5)$$

Since we are interested in finding the SA_i that has the maximum value of $P(SA_i = SC_t)$, $P(P_i)$ will not affect the answer. Thus (5) could be estimated as (6).

$$P(P_i|SC_t) : = P(SC_t|P_i)/P(SC_t) \quad (6)$$

By (4), (6) could be rewritten as (7).

$$P(P_i|SC_t) : = P(SC_t |SP_{i-2}, SP_{i-1}, SP_i, SenType, MainV, Tense, Neg, AuxV, Clue)/P(SC_t) \quad (7)$$

$P(SC_t |SP_{i-2}, SP_{i-1}, SP_i, SenType, MainV, Tense, Neg, AuxV, Clue)$ in (7) could be computed by class probabilities of decision tree. Class probabilities are computed by the distribution of training data which are distributed in terminal node after constructing a decision tree with training data. Thus we can avoid sparse data problem with class probabilities. In the same manner, we can acquire other probability $P(SC_t|SC_u)$ at Step 2 in Figure 2 from class probabilities of the other decision tree. The best sequence of speech acts for utterances could be easily computed by using forward algorithm.

3.3. Class probability of the decision tree

Our purpose of using decision tree is to get smoothed $P(SC_t |SP_{i-2}, SP_{i-1}, SP_i, SenType, MainV, Tense, Neg, AuxV, Clue)$ in (7). When decision tree is constructed, each terminal node has data distribution which has been classified by the non-terminal nodes of the tree. This data distribution represents class probabilities.

Figure 3 shows part of a decision tree. In Figure 3, we can easily acquire smoothed probability of $P(SC_t |SP_{i-2}, SP_{i-1}, SP_i, SenType, MainV, Tense, Neg, AuxV, Clue)$ as $P(SC_t | SP_i, SenType, MainV)$ and $P(SC_t | SenType, MainV)$ when a given utterance matches the condition of terminal nodes such as *[SP_i = Agent or null, MainV = 수고하다 (su-go-ha-da) or 계시다 (gyeo-si-da), SenType = wh_quest]* and *[MainV = 수고하다 (su-go-ha-da) or 계시다 (gyeo-si-da), SenType = yn_quest]* by considering not all the features but selected features such as SP_i, *SenType* and *MainV*. In this way, we can get $P(SC_t |SP_{i-2}, SP_{i-1}, SP_i, SenType, MainV, Tense, Neg, AuxV, Clue)$ and $P(SC_t|SC_u)$ from the terminal nodes of two decision trees which have been trained independently. We constructed two kinds of decision tree; one for observation probabilities, and the other for transition probabilities. For observation probabilities $P(SC_t|P_i)$, the

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

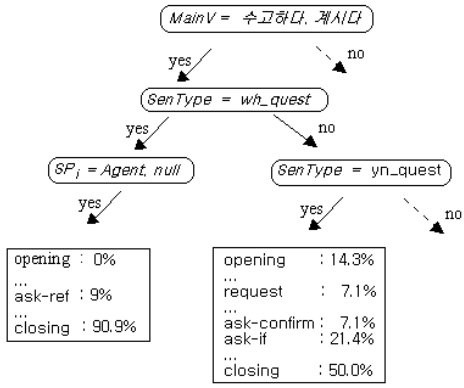

Figure 3. A part of a decision tree.

decision tree was trained with speakers and six syntactic features in order to decide the speech act, for transition probabilities $P(SC_t|SC_u)$, it was trained with only the previous speech act to decide the current speech act.

3.3.1. Construction of the decision tree

We used twoing criterion in order to grow tree [11, 12]. For any node t and class $c=1,...J$, suppose that there is a candidate split s of the node which divides it into t_L and t_R such that a proportion P_L of the cases in t goes into t_L and a proportion P_R goes into t_R. The twoing criterion to be maximized is:

$$P_L P_R /4 \ S_{j=1,...,J} \ (|p(c|t_L)-p(c|t_R)|^2) \tag{8}$$

The goal is to make the probability that a class c object goes to the left as different as possible from the probability that it goes to the right. Given contexts, this node is expanded recursively on each of the two proportions of training data, satisfying twoing criterion. The recursive expansion of the decision tree stops if the next split would generate at least one subset of contexts whose size is below some predefined threshold. In our experiment, we define it as 10.

3.3.2. Pruning the decision tree

Each of the candidate subtrees is used to classify the data in the test sample. One subtree with the lowest overall misclassification rate is selected. We used 10-fold cross validation method to test a subtree. For each n, n=1, ..., 10, the nth case is set aside and decision tree is constructed by using the other 9 cases. Then the nth case is used as a single-case test sample. One optimal tree whose misclassification rate is the lowest is selected.

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

4. Experiment and Result

We used the tagged corpus explained in Section 2. We divided the corpus into the training data with 428 dialogues, 8,349 utterances, and the test data with 100 dialogues, 1936 utterances. We used the same corpus set which was used in [5].

Table 2 shows the accuracy of two HMMs. One does not use decision trees to estimate observation probability and transition probability, the other does. We will call the later model as the *smoothed HMM* as a matter of convenience.

Table 2. Results of a speech act analysis.

Model	Test data (%)	Training data (%)
HMM	52.6	86.2
Smoothed HMM	81.5	85.7

As we expected, the accuracy of HMM with test data is much lower, compared with the accuracy with training data, because of sparse data problem. Table (2) shows that a decision tree is effective for the sparse data problem. Sparse data problem may occur when counting frequencies of features to compute observation probabilities.

Table 3 compares our smoothed HMM with related previous works [2, 3, 4, 5, 8, 9]. As shown in Table 3, the proposed smoothed HMM shows a better result than previous works such as [2] and [5].

Table 3. Results of a speech act analysis.

	Test data (%)	Training data (%)
(Lee et all. 1997)[2]		78.6
(Reithinger et all. 1997)[3]	74.7	
(Samuel et all. 1998)[4]	73.2	
(Tanaka et all. 1999)[8]	75.4	
(Choi et all. 1999)[5]	80.5	
(Stolcke et all. 2000)[9]	65.0	
Smoothed HMM	81.5	85.7

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

[5] analyzes speech act and discourse structure using maximum entropy model. Interestingly, our proposed model without considering discourse structure shows a little better result than [5]. We presume that such result would be caused by the following reasons. Firstly, most of the utterances are affected mainly by their previous utterance. Hence, it is enough to use the previous utterance's speech act as context. Secondly, it is very difficult to tag discourse structure into dialogue corpus consistently. It is also difficult to represent and analyze discourse structure well. Therefore the errors caused by discourse structure analysis would be propagated to the speech act analysis as in [5].

It is difficult to compare the proposed model directly with previous works like [3], [4], [8], and [9] because dialogue corpus and speech acts used in experiments are different. We will test our model using the same data with the same speech acts used in those works in future works.

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

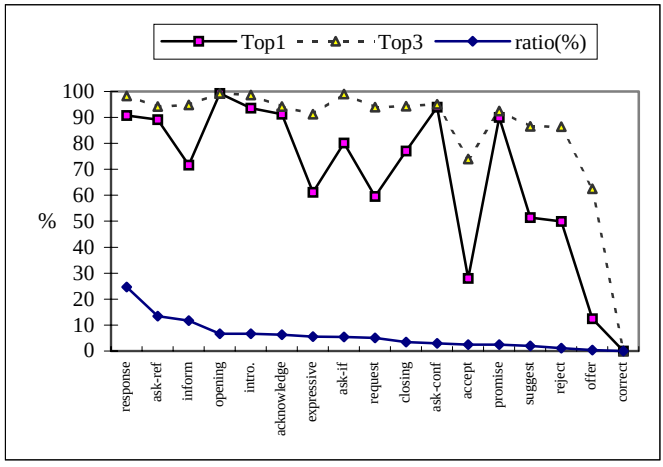

Figure 4. Distribution of speech acts in training data and each speech act's accuracy on the test data.

The errors are mainly caused by insufficient training data and syntactic features. Figure 4 shows the distribution of speech acts in descending order in the training data and each speech act's accuracy on the test data. As we can see in the figure, each speech act's accuracy is proportional to the frequency of each speech act in the training data. The accuracies of speech acts having low frequencies in the training data such as 'accept', 'suggest', 'reject', 'offer', and 'correct' are lower than other speech acts. As in Figure 4, utterances that have speech acts such as 'inform', 'expressive', 'request', 'accept', 'suggest', 'reject',

'offer', and 'correct' were not predicted as well as others. Especially, utterances like 'inform', 'request', 'suggest', and 'offer' are leading utterances that do not depend on previous utterances but initiate new sub-dialogue. Thus it is difficult to predict these speech acts by using contexts [1]. Low frequencies of these utterances in the training data make it more difficult than other utterances. That is to say, data is insufficient for guessing the speech acts of these utterances than other utterances. Interestingly, 'promise' is predicted well although this utterance is rare since it has syntactic features which other speech acts do not have. Actually, every syntactic pattern of 'promise' have 'serve' or 'will' as an *aux-verb* feature. So, it is important to select useful features which can represent both utterance and speech act.

Table 4 shows the pair of correct speech acts and their mostly wrong-predicted speech acts. Speech acts such as 'inform', 'expressive', and 'request' have comparably enough training data but resulted in low performance as shown in Figure 4.

Table 4. Examples of errors.

Correct speech act	Mostly wrong-predicted speech act and its error ratio
inform	response(66%)
expressive	closing(57%)
request	inform(62%)
closing	expressive(73%)

Table 4 shows an example of incorrectly predicted speech acts. It is very difficult to discriminate between some pairs of speech acts represented in Table 4. This result is mainly due to, we believe, the inconsistency of speech act tagging and the incompleteness of the syntactic features.

Most errors occurred with speech acts that are different but have the same syntactic pattern. Table 5 shows examples of syntactic patterns with various utterances which have different speech act. As Table 5 shows, many utterances tagged 'inform' have the same syntactic pattern and context with some utterances tagged 'response' in training corpus. This indicates that the syntactic pattern which represents a surface utterance needs more useful features. Especially, the features which play an important role in predicting speech act must be expanded,

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

and the features which do not have an effect on speech act must be reduced. Thus finding useful knowledge which can discriminate these speech acts is important. This could be another research area to investigate.

Table 5. Examples of syntactic patterns which represent various utterances.

Syntactic pattern [decl, pvg, present, no, none, none]	
expressive	저희 연구소를 방문하신다니 참 환영입니다. *(Jeo-hui yeon-gu-so-reul bang-mun-ha-sin-da-ni cham hwan-yeong-ib-ni-da.)* (You are welcome if you are planning to visit our research center.)
suggest	괜찮으시다면 구월 이십오일 목요일에 방문할까 생각 중인데요. *(Goen-chanh-eu-si-da-myeon gu-wol i-sib-o-il mog-yo-il-e bang-mun-hal-kka saeng-gag-jung-in-deo-yo.)* (If you don' t mind, I would like to visit at Thurs. 25, Sept.)
inform	저희들이 세 명이 가는데요. *(Jeo-hui-deul-i se myeong-i ga-neun-de-yo.)* (Three members will go to.)
response	그 안에 아침 저녁 식사가 포함되어 있습니다. *(Geu an-e a-chim jeo-neog sig-sa-ga po-ham-doe-eo iss-seub-ni-da.)* (It includes breakfast, lunch and dinner.)
Syntactic pattern [decl, pvg, present, no, want, none]	
ask-ref	한국에 차로 여행하는데 정보를 얻고 싶습니다. *(Han-gug-e cha-lo yeo-haeng-ha-neun-de jeong-bo-reul eod-go sip-seub-ni-da.)* (I want to get some information for car tour in Korea.)
response	그곳에 연구 활동을 좀 살펴보고 싶습니다. *(Geu-gos-e yeon-gu hwal-dong-eul jom sal-pyeo-bo-go sip-seub-ni-da.)* (I want to study a research activity in there.)
inform	그 방문 후에 관광을 좀 했으면 좋겠어요. *(Geu bang-mun hu-e gwan-gwang-eul jom haess-eu-myeon joh-gess-eo-yo.)* (After visiting, I would like to have a tour.)
request	지금 예약하고 싶은데요. *(Ji-geum yeo-yag-ha-go sip-eun-de-yo.)* (I want to reserve it now.)

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

5. Conclusions and Future Work

In this paper, we proposed an analysis of speech act by the smoothed HMM which combines HMM and decision trees. We computed the speech act probabilities for each utterance with forward algorithm. Two decision trees provided the observation probabilities and transition probabilities, respectively. Many features can be included in the statistical model avoiding sparse data

problem using class probabilities of decision tree. Class probabilities are computed by the distribution of training data in a terminal node of decision tree. Each terminal node has smoothed probabilities according to the conditions of its precedent nodes.

The proposed model resulted in better performance than other previous works by smoothing each observation probabilities and transition probabilities which are used in HMM. To improve our speech act analysis system, more well annotated dialogue corpus and more useful syntactic features will be needed. Based on our proposed model, we will pursue a discourse structure analysis. It would be necessary to experiment on different corpus to show that our proposed model is good for predicting speech act.

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.

Acknowledgements

This research was supported by Brain Korea 21 Project.

References

[1] M. Nagata and T. Morimoto, "First Steps towards Statistical Modeling of Dialogue to Predict the Speech Act Type of the Next Utterance", *Speech Communication*, 15, 1994, pp. 193–203.

[2] J.W. Lee, J.Y. Seo and G.C. Kim, "A Dialogue Analysis Model With Statistical Speech Act Processing For Dialogue Machine Translation", in *Proceedings of Spoken Language Translation (Workshop in conjunction with (E)ACL'97)*, 1997, pp. 10–15.

[3] N. Reithinger and M. Klesen, "Dialogue Act Classification Using Language Models", in *Proceedings of EuroSpeech-97*, 1997, pp. 2235–2238.

[4] K. Samuel, S. Carberry and K. Vijay-Shanker, "Dialogue Act Tagging with Transformation-Based Learning", in *Proceedings of COLING-ACL98*, Montr'eal, Canada, 1998, pp. 1150–1156.

[5] W.S. Choi, J.M. Cho and J.Y. Seo, "Analysis System of Speech Acts and Discourse Structures Using Maximum Entropy Model", in *Proceedings of COLING-ACL99*, 1999, pp. 230–237.

[6] K. Samuel, S. Carberry and K. Vijay-Shanker, "Automatically Selecting Useful Phrases for Dialogue Act Tagging", in *Proceedings of the Fourth Conference of the Pacific Association for Computational Linguistics*, Waterloo, Ontario, Canada, 1999.

[7] S.W. Lee and J.Y. Seo, "Korean Speech Act Analysis Using Decision Tree", in *Proceedings of the Conference on Hangul and Korean Language Information Processing*, 1999, pp. 377–381.

[8] H. Tanaka and A. Yokoo, "An Efficient Statistical Speech Act Type Tagging System for Speech Translation Systems", in *Proceedings of COLING-ACL99*, 1999, pp. 381–388.
[9] A. Stolcke and K. Ries *et al*., Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech, *Computational Linguistics*, 26, No. 3, 2000, pp. 339–373.
[10] H. Schmid, "Probabilistic Part-of-Speech Tagging Using Decision Trees", in *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, 1994, pp. 44–49.
[11] L. Breiman, J. Friedman and R. Olshen *et al*., *Classification and Regression Trees*, Wadsworth International Group, Belmont, California, 1984.
[12] D. Steinberg and P. Colla, *CART Interface and Documentation*, Salford Systems, 1997.
[13] E. Charniak, *Statistical Language Learning*, A Bradford Book, MIT Press, 1993.
[14] J. Allen, *Natural Language Understanding*, Benjamin/Cummings, 1994.
[15] M. Collins and J. Brooks, "Prepositional Phrase Attachment through a Backed-Off Model", in *Proceedings of the Third Workshop on Very Large Corpora*, Cambridge, Massachusetts, ACL, 1995, pp. 27–38.

Int. J. Comp. Proc. Lang. 2002.15:231-243. Downloaded from www.worldscientific.com
by UNIVERSITY OF GEORGIA on 04/28/15. For personal use only.