Improving Korean Speech Acts Analysis by Using Shrinkage and Discourse Stack

Kyungsun Kim[1], Youngjoong Ko[2], and Jungyun Seo[3]

[1] Information Retrieval Division, Diquest.Inc, Seocho-dong,
Seocho-gu, Seoul, 137-070, Korea
kksun@diquest.com
[2] Dept. of Computer Engineering, Dong-A University, 840,
Hadan 2-dong, Saha-gu, Busan, 604-714, Korea
yjko@dau.ac.kr
[3] Dept. of Computer Science and Interdisciplinary Program of Integrated Biotechnology,
Sogang University, Seoul, 121-742, Korea
seojy@sogang.ac.kr

Abstract. A speech act is a linguistic action intended by a speaker. It is important to analyze the speech act for the dialogue understanding system because the speech act of an utterance is closely tied with the user's intention in the utterance. This paper proposes to use a speech acts hierarchy and a discourse stack for improving the accuracy of classifiers in speech acts analysis. We first adopt a hierarchical statistical technique called shrinkage to solve the data sparseness problem. In addition, we use a discourse stack in order to easily apply discourse structure information to the speech acts analysis. From the results of experiments, we observed that the proposed model made a significant improvement for Korean speech acts analysis. Moreover, we found that it can be more useful when training data is insufficient.

1   Introduction

To understand a natural language dialogue, a dialogue system must be able to make out the speaker's intentions indicated by utterances. Since the speech act of an utterance is very important in understanding a speaker's intentions, it is an essential part of a dialogue system. However, it is difficult to infer the speech act from a surface utterance because the utterance may represent more than one speech act according to the context [5][7].

Various machine learning models have been used to efficiently classify speech acts such as MEM (Maximum Entropy Model) [1], HMM (Hidden Markov Model) with Decision Tree [8][11], Neural Network Model [5]. And there are also studies on methods of automatically selecting efficient features with useful information for speech acts analysis [5][10]. Since the machine learning models can efficiently analyze a large quantity of data and consider many different feature interactions, they can provide a means of associating features of utterances with particular speech acts.

Generally, it is hard to create enough the number of examples for each speech act in the training examples. Thus this situation has been one of the main causes for errors occurred in speech acts analysis. That is, the sparse data problem from low

frequency of some speech acts has commonly occurred in the previous research [8]. Due to the problem, the accuracy of each speech act in previous research tends to be proportional to the frequency of each speech act in the training data. Therefore, we first focus on how to scale up statistical learning methods to solve the sparseness problem of training data in speech acts analysis. Then we propose to construct the commonly-available hierarchies of speech acts and apply a well-understood technique from Statistics called shrinkage to our speech acts analysis system. It provides improved estimates of parameters that would otherwise be uncertain due to limited amounts of training data [3]. The technique uses a hierarchy to shrink parameter estimates in data sparse children toward the estimates of the data-rich ancestors in ways that are probably optimal under the appropriate conditions [9]. We employ a simple form of shrinkage that creates new parameter estimates for a child by a linear interpolation of all hierarchy nodes from the child to the root.

In addition, discourse structure information can be used to identify the speech acts of utterances [1]. But most previous research has used only speech acts of previous utterances without considering discourse structure information to determine the speech act of current utterance. Therefore, in order to use discourse structure information for analyzing speech acts, we design a simple discourse stack. By using the discourse stack, the discourse structure information is easily applied to speech acts analysis.

In this paper, we propose a new speech acts analysis model to improve the performance by using shrinkage and discourse structure information. From the results of experiments, the proposed system showed significant improvement in comparison with previous research.

The rest of this paper is organized as follows. Section 2 explains the proposed speech acts analysis system in detail. In section 3, we discuss the empirical results in our experiments. The final section presents conclusions.

2   The Proposed Speech Acts Analysis System

The proposed system consists of two modules as shown in Fig. 1: one module to extract features from training data and the other module to build up a hierarchy of
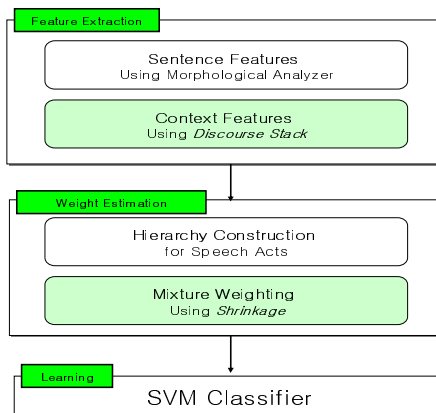
Fig. 1. The overview of the proposed system

speech acts and estimate weights of each feature on the hierarchy by shrinkage. Each process of Fig. 1 is explained in the following sections.

2.1 Feature Extraction

2.1.1 Sentence Features Extraction

We assume that clue words and a sequence of POS tags in an utterance provide very effective information for analyzing the speech act of the current utterance. We extract informative features for speech acts analysis using a Morphological analyzer; they are called the sentence features. The sentence features consist of content words annotated with POS tags and POS bi-grams of all words in an utterance. Fig. 2 shows an example of sentence feature extraction.
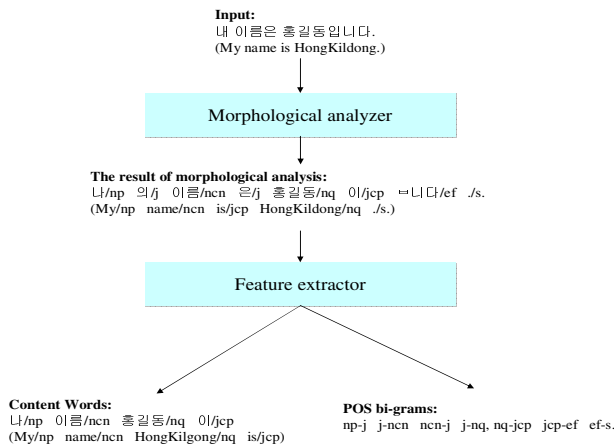
Input:
내 이름은 홍길동입니다.
(My name is HongKildong.)

Morphological analyzer

The result of morphological analysis:
나/np 의/j 이름/ncn 은/j 홍길동/nq 이/jcp ㅂ니다/ef ./s.
(My/np name/ncn is/jcp HongKildong/nq ./s.)

Feature extractor

Content Words:
나/np 이름/ncn 홍길동/nq 이/jcp
(My/np name/ncn HongKilgong/nq is/jcp)

POS bi-grams:
np-j j-ncn ncn-j j-nq, nq-jcp jcp-ef ef-s.

Fig. 2 An example of sentence feature extraction

```
For each utterance

Begin
    if(Move a sub-dialogue?)
        Use speech acts of previous utterance and Sub-dialogue Start (SS)
        Push speech acts of current utterance.
    else if(Return from a sub-dialogue?)
        Use speech acts that pop in discourse stack and Sub-dialogue End (SE)
    else
        Use speech acts of previous utterance and Dialogue Continue (DC)
End
```

2.1.2 Context Features Extraction

Most previous research uses the speech act of previous utterance as context feature (CF1 in Table 1) [5][8]. Since discourse structure information represents the relationship between two consecutive utterances, it is efficient to use discourse structure

information for speech acts analysis [1]. Especially, the speech act of seventh utterance in Table 1 (UID: 7) is tied with that of second utterance (UID: 2). In our system, we first design a discourse stack to easily detect discourse structure information and extract the discourse structure information from the discourse stack for context features. Context features of our system consist of speech acts of previous utterance and markers of discourse structure information (CF2 in Table 1). An algorithm for discourse stack is described as the following:

Table 1. An example of Context Feature

UID	DS	Utterance	Speech Acts	CF1	CF2
1	1	방을 하나 예약하고 싶은데요 (I would like to reserve a room)	Inform	Dialog-start	Dialog-start, NULL
2	1.1	어떤 방을 원하시죠? (What kind of room do you want?)	Ask-ref	Inform	Inform, SS
3	1.1.1	어떤 종류의 방이 있습니까? (What kind of room do you have?)	Ask-ref	Ask-ref	Ask-ref, SS
4	1.1.1	더블룸과 싱글룸이 있습니다. (We have single and double rooms)	Response	Ask-ref	Ask-ref, DC
5	1.1.2	방값이 얼마죠? (How much are those rooms?)	Ask-ref	Response	Response, DC
6	1.1.2	싱글은 삼만원이고 더블은 사만원 입니다. (Singles cost 30,000 won and doubles cost 40,000 won.)	Response	Ask-ref	Ask-ref, DC
7	1.1	싱글룸으로 해주세요. (A single room, please)	Response	Response	Ask-ref, SE

* UID: ID of utterances, DS: Discourse Structure, CF1: Using speech acts of previous utterances as features (Context Feature Type1), CF2: Using Discourse Structure Information by Discourse Stack as features (Context Feature Type2), Speech acts and discourse structure information were annotated by human.

2.2 The Feature Weight Calculation by Shrinkage in a Hierarchy of Speech Acts

Data sparseness is a common problem in mechanical learning fields. For speech acts analysis, the problem becomes more serious because it is a time-consuming and difficult task to collect dialogue examples and construct dialogue training data tagged with a lot of information for various application areas. Therefore, we apply the shrinkage technique to solve this data sparseness problem in speech acts analysis. The shrinkage technique was verified in its efficiency for text classification tasks learned with insufficient training data. Therefore, we first build up a hierarchy of speech acts to estimate the weight of features for each speech act by the shrinkage technique.

2.2.1 The Hierarchy Construction for Speech Acts

To model a dialogue system, the dialogue grammar has commonly used and it has observed that dialogues consist of adjacency pairs of the types of utterances such as
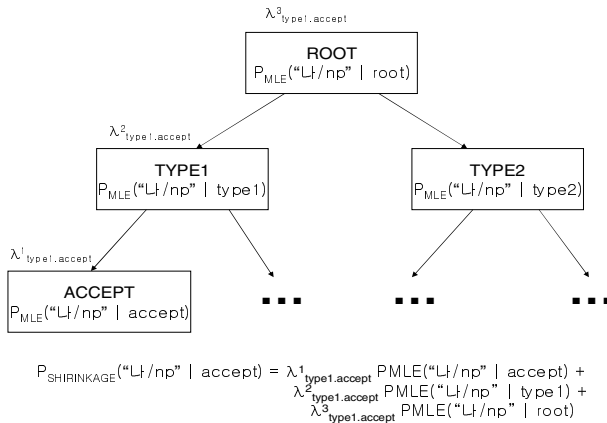
Table 2. The Hierarchy of Speech Acts

	Parent	Child
Root	Type1: Utterances of request type	Ask-if
		Ask-ref
		Ask-confirm
		Offer
		Suggest
		Request
	Type2: Utterances of response type	Accept
		Response
		Reject
		Acknowledge
	Type3: Utterances with a speaker emotion	Expressive
		Promise
		Closing
	Type4: Utterances of usually life	Opening
		Introducing-oneself
		Correct
		Inform

request-type and response-type [2][8]. Therefore, our speech acts hierarchy is built up according to this grammar. Table 2 shows the structure of our speech acts hierarchy.

2.2.2 Mixture Weighting Model by Shrinkage in a Hierarchy of Speech Acts

The shrinkage technique estimates the probability of a word as the weighted sum of the maximum-likelihood estimates from leaf to root in a hierarchy [9]. This estimate process can give us a possibility to resolve the data sparseness problem in some speech acts with insufficient examples. Fig.3 shows that the shrinkage-based estimate of the probability of a feature ("나/np") given a speech act class ("Accept") is calculated from a weighted sum of the maximum-likelihood estimates from leaf to root.

Fig. 3. An example of the shrinkage-based estimate of the probability of features

Let $\{\hat{\theta}_j^1, \hat{\theta}_j^2, ..., \hat{\theta}_j^k\}$ be k such estimates, where $\hat{\theta}_j^k = \theta_j$ is the estimate at the leaf, and *k-1* is the depth of speech acts s_t in a hierarchy of Speech Acts. The interpolation weights among the ancestors of speech acts s_t are written $\{\lambda_j^1, \lambda_j^2, ..., \lambda_j^k\}$, where $\sum_{i=1}^k \lambda_j^i = 1$. We write $\breve{\theta}_j$ for the new estimate of the speech act-conditioned feature probabilities based on shrinkage. The new estimate for the probability of feature f_t given speech act s_j is as follows:

$$\breve{\theta}_{jt} = P(f_t \mid s_j; \breve{\theta}_j) = \lambda_j^1 \hat{\theta}_{jt}^1 + \lambda_j^2 \hat{\theta}_{jt}^1 + ... + \lambda_j^k \hat{\theta}_{jt}^1 . \tag{1}$$

We derive empirically optimal weights using the following iterative procedure:

Initialize:

Set the λ_j 's to some initial values, say $\lambda_j^i = \dfrac{1}{k}$

Iterate:

1. Calculate the degree to which each estimate predicts the features f_t in the held-out feature set, H_j , from speech acts s_j :

$$\beta_j^i = \sum_{w_t \in H_j} P(\hat{\theta}_j^i \text{ was used to generate } f_t) = \sum_{w_t \in H_j} \frac{\lambda_j^i \hat{\theta}_{jt}^i}{\sum_m \lambda_j^m \hat{\theta}_{jt}^m} \tag{2}$$

2. Compensate the degree for loss that is caused by large variation of each degree :

$$\beta_i^j = \beta_i^j + \frac{\sum_m \beta_j^m}{m} \tag{3}$$

3. Derive new weights by normalizing the β's :

$$\lambda_j^i = \frac{\beta_j^i}{\sum_m \beta_j^m} \tag{4}$$

Terminate: Upon convergence of the likelihood function

2.3   The SVM Classifier

Support Vector Machines (SVM) is one of the state-of-the-art classifiers for classification tasks [6][12]. Since SVM has shown the high performance in various research areas, we also employ it in our method. In our method, we use the linear models offered by SVM[light] [4] and $\breve{\theta}_{jt}$, which are calculated by formula (1), are used as the feature weights of speech acts for the SVM classifier.

Table 3. The part of mixture weights learned by shrinkage-based estimation

# training documents	Speech Acts			Mixture Weights		
	Root	Parent	Child	Root	Parent	Child
250	Root	Type1	Ask-ref	0.289	0.32	0.39
			Suggest	0.257	0.275	0.467
		Type2	Expressive	0.263	0.335	0.4
		Type3	Reject	0.259	0.269	0.47
		Type4	Inform	0.297	0.336	0.366
8349	Root	Type1	Ask-ref	0.282	0.295	0.422
			Suggest	0.217	0.22	0.562
		Type2	Expressive	0.229	0.279	0.49
		Type3	Reject	0.212	0.215	0.571
		Type4	Inform	0.26	0.332	0.406

3   Empirical Evaluation

3.1   Experimental Data

We used the Korean dialogue corpus which has used in previous research [1][5][8]. This corpus was transcribed from recordings in real fields such as hotel reservation, airline reservation and tour reservation and consists of 528 dialogues, 10,285 utterances (19.48 utterances per dialogue). Each utterance in dialogues is manually annotated with a speaker (SP), a speech act (SA) and a discourse structure (DS). This annotated dialogue corpus has 17 types of speech acts. Table 4 shows a part of the annotated dialog corpus and Table 5 shows the distribution of speech acts in the annotated dialogue corpus.

Table 4. A part of the annotated dialogue corpus

Tag	Values
SP	Customer
KS	미국 조지아대 어학연수에 참가 신청을 한 학생인데요.
EN	I'm a student and registered for a language course at University of Georgia in U.S.
SA	Introducing-oneself
DS	[2]
SP	Customer
KS	숙소에 관해서 문의할 사항이 있어서요.
EN	I have some questions about lodgings.
SA	Request
DS	[2]

Table 5. The distribution of speech acts in corpus

Speech act type	Ratio (%)	Speech act type	Ratio (%)
Accept	2.49	Introducing-oneself	6.75
Acknowledge	5.75	Offer	0.4
Ask-confirm	3.16	Opening	6.58
Ask-if	5.36	Promise	2.42
Ask-ref	13.39	Reject	1.07
Closing	3.39	Request	4.96
Correct	0.03	Response	24.73
Expressive	5.64	Suggest	1.98
Inform	11.9	Total	100

We divided the annotated dialogue corpus into the training data with 428 dialogues, 8,349 utterances (19.51 utterances per dialogue), and the testing data with 100 dialogues, 1,936 utterances (19.36 utterances per dialogue).

3.2 Primary Experimental Results

3.2.1 The Performances of Speech Acts Analysis Model Using Shrinkage and Discourse Stack

In order to verify the proposed method, we made four kinds of speech acts analysis systems which use different kind of features. The Baseline System used default features such as sentence features and context features [5]. The Second system (Type 1) was built up to verify the shrinkage technique. Its features were the same as those of the first system but they were weighted by the shrinkage technique. The third System (Type 2) used the discourse structure information from the proposed discourse stack without shrinkage. Finally, the fourth system (Type 3) combined the discourse structure information and the shrinkage technique.

Table 6 shows the results of four speech acts analysis systems. As shown in Table 6, the performances of the proposed systems (Type 1,2,3) are better than the baseline system. The proposed system of Type 3 reported the best performance.

3.2.2 The Improvement of the Proposed System Using the Shrinkage Technique in Sparse Data

Here, we verify the facts that the shrinkage technique can improve the speech acts analysis when training data is sparse. We first compare the system with shrinkage (Type 3) and the system without shrinkage (Type 2). Fig. 4 shows the changes of performance in each number of training data from 250 to 8439. The proposed system with shrinkage obtains the better performance over all intervals in Fig. 4. Especially, the shrinkage technique provides more improvement when the amount of training data is small. This is a proof that the shrinkage technique can become an effective solution for sparse data problem from insufficient training data.

Table 6. The results of four speech acts analysis systems (precision %)

Speech acts	Baseline System	Proposed System (Type1)	Proposed System (Type2)	Proposed System (Type3)
Accept	36.00%	50.00%	38.00%	50.00%
Acknowledge	91.30%	91.30%	92.75%	95.65%
ask-confirm	92.68%	96.34%	93.90%	95.12%
ask-if	84.16%	86.14%	86.14%	89.11%
ask-ref	89.88%	91.05%	90.66%	91.44%
Closing	60.00%	61.43%	67.14%	71.43%
Correct	0.00%	0.00%	0.00%	0.00%
Expressive	85.84%	83.19%	87.61%	83.19%
Inform	70.00%	70.00%	76.00%	75.60%
Introducing-oneself	98.58%	98.58%	97.87%	98.58%
Offer	12.50%	12.50%	12.50%	12.50%
Opening	97.60%	96.80%	96.80%	96.80%
Promise	92.50%	92.50%	87.50%	90.00%
Reject	68.18%	72.73%	68.18%	68.18%
Request	71.43%	73.81%	70.24%	69.05%
Response	96.49%	96.07%	96.07%	96.07%
Suggest	56.76%	56.76%	56.76%	62.16%
TOTAL	85.18%	85.85%	86.31%	87.04%
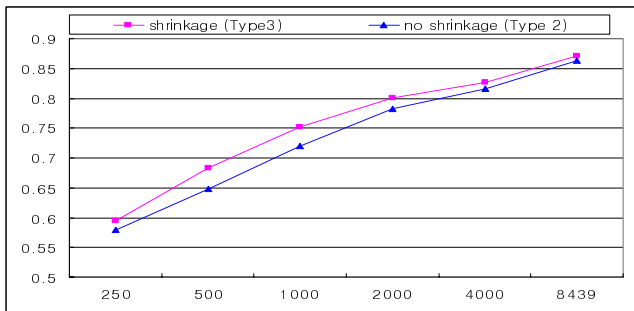
Fig. 4. The performance according to different number of training data

We then compare performances between the system of Type 2 and the system of Type 3 according to distribution of each speech act. As shown in Fig. 5, the proposed system (Type 3) with the shrinkage technique shows higher performance in speech acts with insufficient examples such as 'Accept', 'Closing', 'Promise' and 'Suggest'.
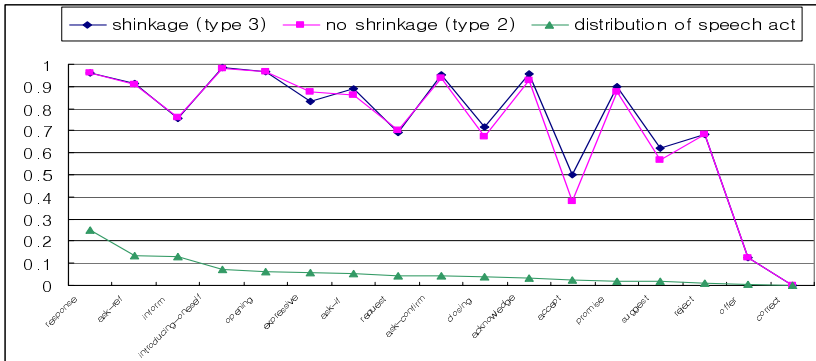
Fig. 5. The comparison of the performances for the shrinkage technique according to the distribution of speech acts

3.2.3   The Comparison of Performance with Other Speech Acts Analysis Models

Table 7 shows results from the proposed model and previous speech acts analysis models: the maximum entropy model (MEM) [1], the decision tree model (DTM) [8], and the neural network model (NNM) [5]. We report the performance of each system when using the same test data set as that of this paper. As a result, the proposed model achieved the highest performance.

Table 7. The experimental results of the proposed model and other previous models

Model	Precision (%)
MEM	83.4%
DTM	81.7%
NNM	85.2%
The propose model	87.0%

In the experiment, it is difficult to compare the proposed model directly with the other models because input features are different respectively. Even though direct comparisons are impossible, we think that the proposed model is more robust and efficient than MEM and DTM. In MEM and DTM, they used many kinds of high level linguistic knowledge than ours such as sentence type, tense, modality and so on. Nevertheless, the performances of them are lower than that of the proposed model. Moreover, the proposed model is more effective than NNM because the performance of the proposed model is better than that of NNM in spite of using same features.

4   Conclusions

In this paper, we proposed the new speech analysis model to improve speech acts analysis by using the shrinkage technique and the discourse stack. We first made a

hierarchy of speech acts by dialogue grammar for shrinkage and then estimate the probability of each feature on the hierarchy by the shrinkage technique. In experimental results, the proposed model is more effective for classifying speech acts. Especially, the shrinkage technique achieved more improvement when training data is sparse. Therefore, the shrinkage technique can be applied to the real applications that suffer from the data sparseness problem. We also proposed to use the discourse stack for easily extracting discourse structure information. As a result, the proposed model with shrinkage and the discourse stack showed the better performance than other speech acts analysis models.

Acknowledgement

References

1. Choi, W., Cho, J. and Seo, J.: Analysis System of speech acts and Discourse Structures Using Maximum Entropy Model, In Proceedings of COLING-ACL99, (1999), 230-237
2. Grosz, B.: Discourse and Dialogue, In Survey of the State of the Art in Human Language Technology, Center for Spoken Language Understanding, (1995), 227-254
3. James, W. and Stein, C.: Estimation with Quadratic Loss, In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1, University of California Press, 361-379
4. Joachims, T.: Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In European conference on machine learning (ECML), (1998), 137-142
5. Kim, K., Kim, H. and Seo, J.: A Neural Network Model with Feature Selection for Korean Speech Act Classification, International Journal of Neural System, VOL. 14 NO. 6, (2004), 407-414
6. Ko, Y., Park, J, Seo, J.: Improving Text Categorization Using the Importance of Sentences, Information Processing & Management, Vol. 40, No. 1, (2004), 65-79
7. Lee, J., Kim, G., and Seo, J.: A Dialogue Analysis Model with Statistical Speech Act Processing for Dialogue Machine Translation, In Proceedings of ACL Workshop on Spoken Language Translation, (1997), 10-15
8. Lee, S. and Seo, J.: A Korean Speech Act Analysis System Using Hidden Markov Model with Decision Trees, International Journal of Computer Processing of Oriental Languages. VOL. 15, NO. 3, (2002), 231-243
9. MacCallum, A., Rosenfeld, R., Mitchell, T. and Ng, A.Y.: Improving Text Classification by Shrinkge in a Hierarchy of Classes, In Proceedings of the International Conference on Machine Learning. (1998)
10. Samuel, K., Caberry, S., and Vijay-Shanker, K.: Automatically Selecting Useful Phrases for Dialogue Act Tagging, In Proceedings of the Fourth Conference of the Pacific Association for Computational Linguistics, (1999)
11. Tanaka, H. and Yokoo, A.: An Efficient Statistical Speech Act Type Tagging System for Speech Translation Systems, In Proceedings of COLING-ACL99, (1999), 381-388
12. Vapnik, V.: The Nature of Statistical Learning Theory, Springer Verlag, New York, (1995)