A Multi-Strategic Concept-Spotting Approach for Robust Understanding of Spoken Korean

Changki Lee, Jihyun Eun, Minwoo Jeong, Gary Geunbae Lee, YiGyu Hwang, and Myung-Gil Jang

We propose a multi-strategic concept-spotting approach for robust spoken language understanding of conversational Korean in a hostile recognition environment such as in-car navigation and telebanking services. Our concept-spotting method adopts a partial semantic understanding strategy within a given specific domain since the method tries to directly extract pre-defined meaning representation slot values from spoken language inputs. In spite of partial understanding, we can efficiently acquire the necessary information to compose interesting applications because the meaning representation slots are properly designed for specific domain-oriented understanding tasks. We also propose a multi-strategic method based on this concept-spotting approach such as a voting method. We present experiments conducted to verify the feasibility of these methods using a variety of spoken Korean data.

Keywords: Robust spoken language understanding, information extraction, concept spotting, spontaneous speech.

Manuscript received Aug. 11, 2006; revised Jan. 02, 2007.

This research was supported by the Ministry of Information and Communication, Korea, under the Information Technology Research Center support program supervised by the Institute of Information Technology Assessment; IITA-2005-C1090-0501-0018 and was also funded by the second BK21 program of the Ministry of Education of Korea.

Changki Lee (phone: + 82 42 860 6879, email: leeck@etri.re.kr), YiGyu Hwang (email: yghwang@etri.re.kr), and Myung-Gil Jang (email: mgjang@etri.re.kr) are with Embedded S/W Research Division, ETRI, Daejeon, Korea.

Jihyun Eun (email: tigger@postech.ac.kr) is with Future Technology Laboratory, KT, Seoul, Korea.

Minwoo Jeong (email: stardust@postech.ac.kr) and Gary Geunbae Lee (gblee@postech.ac.kr) are with the Department of Computer Science and Engineering, POSTECH, Pohang, Korea.

I. Introduction

The understanding of spoken language has been studied extensively since the late 1980s. The basic roles of spoken language understanding (SLU) are to analyze the output of the speech recognizer and to assign a meaning representation that can be used by the dialog manager [1]. That is, SLU involves two primary component technologies: automatic speech recognition (ASR) and natural language understanding (NLU) [2]. However, the integration of speech and natural language is not a simple matter. In contrast to written language, spoken language has grammatically poor structure because people speak immediately without thinking in depth when they converse with other people. Furthermore, SLU is much more complicated due to abundant ambiguities in natural language and very noisy speaking environments. Moreover, the output from the speech recognizer is likely to contain misrecognized words as well as features of spontaneous speech such as filled pauses, restarts, repetitions and repairs (see Table 1).

To overcome these speech recognition limitations, we attempt to understand spoken languages by a concept-spotting approach which aims to extract only essential factors for

Table 1. Examples of spontaneous speech phenomena.

Disfluency	Example
Filled pauses	**Um** I do **uh** some **uh** paper writing.
Restarts	**The new train** it is so fast.
Repetitions	**The movie** the movie is really interesting.
Repairs	I took this book from **the library** the bookstore.

pre-defined meaning representations. This approach is quite different from conventional approaches because it bypasses the general semantic structure and tries to directly extract the domain dependent task structures. For SLU, this approach claims only partial understanding rather than a level of full understanding because it is only interested in pre-defined meaning representation slots. In spite of this partial understanding, we can acquire the necessary information to make many interesting applications possible from the slot values because the slots are properly designed for domain-oriented language understanding tasks. We also propose a multi-strategic method to map directly from the input sentences to the intended meaning structures. Specifically, we propose a voting-based selection from multiple classifiers for robust domain-dependent spoken language understanding.

The remainder of this paper is organized as follows. In section II, we review some previous SLU studies. In section III, we describe our concept-spotting approach for an SLU system. In section IV, we describe a voting-based multi-strategic approach, and in section V, we present our experimental results. Finally, section VI presents our conclusions.

II. Previous Studies

The two main types of handling methods for spoken language understanding are rule-based methods and statistical methods.

Language understanding systems that use a large set of rules to explain the syntactic and semantic possibilities for spoken utterances suffer from a lack of robustness when faced with the wide variety of ill-formed spoken sentences that people use. One reason for this weakness is that, for most limited domains, a traditional syntactic explanation of a sentence is often much less focused than a direct explanation of the meaning of the sentence in terms of the words spoken and the relations [3] between the words. In order to remedy this syntactic generality, these systems have typically been implemented via hand-crafted semantic level grammar rules and some form of robust parsers such MIT's TINA [4] and CMU's PHOENIX [5]. However, this semantic grammar approach carries a high development cost because it is necessary to compose different hand-crafted grammars according to several domain applications. It can also lead to fragile operations since users do not typically know which grammatical constructions are supported by the system.

An alternative approach is to map directly from word strings to the intended meaning structures. In this approach, hand-crafted grammars and rules are replaced by statistical models that are automatically learned from training data. Statistical methods are attractive because they can be easily adapted to new conditions (tasks or languages) if appropriate training data is available. That is, these methods are portable and cost-efficient but should not decrease understanding accuracy. Statistical methods for spoken language understanding have already been investigated and proposed in AT&T's CHRONUS [6], BBN's hidden understanding model (HUM) [7] and the hidden vector state (HVS) model [8]. These models were primarily motivated by the single technique which has been extremely successful in speech recognition and natural language processing—the hidden Markov model (HMM). However, these previous models try to learn the general semantic structures with multiple mapping using a single HMM-based learning method. Our concept-spotting approach is different in two respects. First, we try to extract direct task-level semantic structures. Second, we use multiple classification and named-entity tagging methods for diverse structure learning for practical dialog management.

More recently, Wang and others [9] proposed a composite model to integrate an HMM with context-free grammar to apply a semantic to the statistical model before the semantic structure. They also built a conditional-random-field (CRF)-based SLU system and reported that the conditional model reduced the SLU slot-error rate by 17% over the generative HMM/CFG composite model on ATIS data. In contrast to [9], we use a multi-strategic approach which addresses the SLU problem as three different statistical learning tasks.

The most relevant prior work is [10], in which a multi-stage Thai SLU system was developed. It consists of three stages: concept extraction, goal identification, and concept-value recognition. This system is similar to our method in terms of multi-level representation of a user's utterances, such as goal identification and concept-value recognition. In contrast to [10], we can use an advanced multi-classifier combination to identify the dialog act and the main goal for each utterance. Moreover, our concept-value (or component slot-value) extraction is based on the state-of-the-art machine learning approach, CRF, which has been previously applied with promising results to various tasks such as part-of-speech (POS) tagging [11], table extraction [12], shallow parsing [13], information extraction for research papers [14], and spoken language understanding [15].

III. Concept-Spotting Approach for Robust SLU

In this section, we first describe the detailed methodology of our concept-spotting approach and the definition of the meaning representation for this approach in the SLU system. Then, we describe the basic features and a method using this concept-spotting approach.

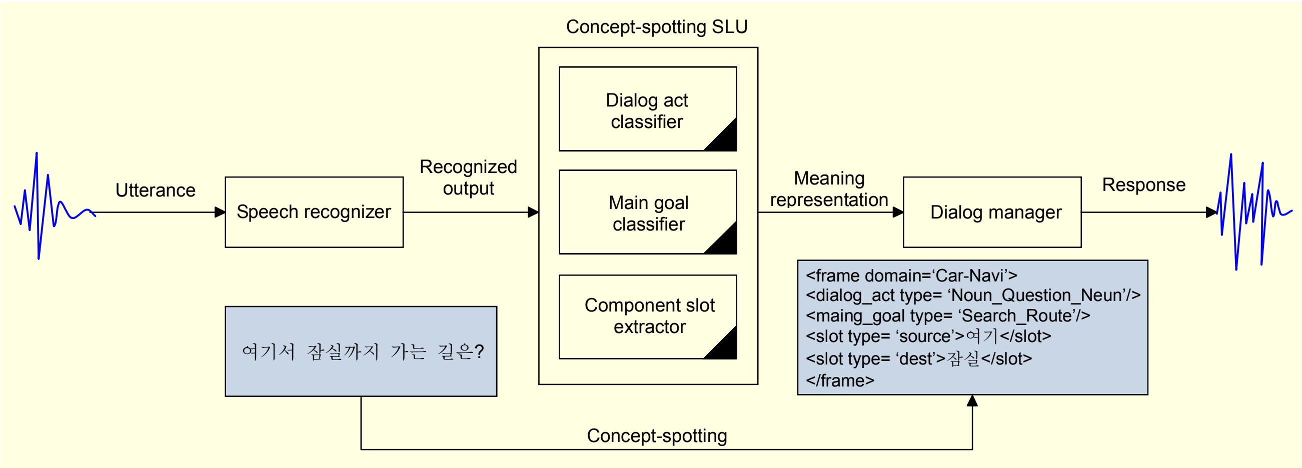


Fig. 1. Structure of the concept-spotting SLU system.

1. Concept-Spotting SLU System

Basically, our concept-spotting approach is based on the idea of information extraction (IE) in text processing. The goal of IE is to build systems which find and link relevant information while ignoring extraneous and irrelevant information. The desired knowledge in IE can be described by a relatively simple and fixed template with slots which need to be filled [16]. These characteristics of IE are well suited for natural language processing tasks, such as POS tagging and partial parsing.

In our concept-spotting approach, several essential factors for understanding spoken language and dialog management are extracted based on the pre-defined domain-dependent meaning representation. Thus, the meaning representation should be solidly defined to express the intentions of the users as fully as possible but should not be too complex to be managed in an SLU system. The meaning representation directly reflects a task structure for dialog management by including dialog acts, a main goal or theme, and possible slot structures for the main goal.

2. Meaning Representation Design

Our concept-spotting approach attempts to understand spoken language by extracting essential factors for pre-defined slots. Figure 1 illustrates the structure of the concept-spotting SLU in a spoken dialog system.

This approach addresses the problem of SLU as three levels of understanding problems: dialog act, main goal, and component slot. A dialog act and main goal present the meaning of an utterance at the discourse level, and it is approximately the equivalent of intent or subject slot in a practical dialog system. The dialog act is a domain-independent and surface-level concept, but the main goal is a domain-specific and functional-level concept. A component slot is a generalized identifier of a named entity such as a person, location, organization, or time. In the SLU problem, we itemize the component slots as the domain-specific semantic meanings of words. Table 2 shows an example of our representation scheme for the telebanking service domain. We solve the tasks on dialog act and main goal slots as classification tasks. The value of a dialog act slot is assigned from one of the classes which designate the surface-level speech acts, such as *yn_question*, *wh_question*, *request*, and so on for each sentence. Similar to the value assignment of the dialog act slot, the value of the main goal slot is assigned from one of the classes of the main application actions in a specific domain, for instance, in the telebanking service domain, such as *confirm_qualified*, *search_info*, *confirm_info*, and so on. The tasks for the component slots, such as, *info*, *period*, and *rate* in the telebanking service domain are solved as named entity recognition tasks.

Table 2. Example of meaning representation in the telebanking service domain.

요새 (nowadays)	인기 있는 예금이 (popular savings)	뭐가 (what)	있지 (are)
Dialog_act = wh_question			
Main_goal = search_info			
Slot.period = '요새' (nowadays) Slot.info = '인기 있는 예금' (popular savings)			

3. Utterance Classification

To assign classes for dialog act and main goal prediction, we use conditional maximum entropy (ME) classifiers and linguistically-motivated features to model the classifiers.

The criterion of ME is based on the idea that the most uniform distribution among the probability distributions that satisfy the given constraints is ideal. This classifier offers a clean way to combine diverse pieces of contextual evidence in order to estimate the probability of a certain class occurring in a certain linguistic context [17].

The objective of this modeling is to find the *y* that maximizes the conditional probability $p(y|x)$, which is expressed as *p*. In our model of IE-based SLU, *x* is the context of a sentence and *y* is the value of a pre-defined slot (dialog act or main goal) in the sentence. Given *k*-feature constraints, the conditional probability is

$$p(y|x) = \frac{1}{Z}\exp\left(\sum_{k=1}^{K}\lambda_k f_k(y,x)\right),$$

where *k* is the number of features, f_k denotes the features, λ_k denotes the weighted parameters for features in the ME model, and *Z* is a normalization factor to ensure that $\Sigma\, p(y|x) = 1$.

Maximum entropy has been applied to various NLP tasks such as sentence boundary detection, POS tagging, and parsing. Also, the ME classifier is good for integrating information from many heterogeneous information sources. Thus, for concept-spotting SLU tasks using the ME framework, we use the following basic diverse linguistic features for classification and extraction:

- Lexical word features: Lexical word features include current, previous, and next lexical words. They are clear and powerful features for classification, but the lexical words appearing in training data are limited and data sparseness can arise.
- POS tag features: POS tag features include current, previous and next lexical tags. Because POS tag features have the role of categorizing each lexical word, they are also important for classification.
- Last-N word features: In Korean, the classes of sentences are often influenced by the last words in sentences; hence, these last-N word features are also included for classification.

4. Component Slot Extraction

Next, the problem of component slot extraction for our concept-spotting SLU can be stated as a sequential supervised learning problem. Given word sequence x, the component slot extractor finds the best probable component slot **s***, which can be defined by the extracted frames with a slot/value structure as

$$s^* = \operatorname*{argmax}_{\mathbf{s}}\ p(\mathbf{s}|\mathbf{x}).$$

Now, we can formalize the SLU problem as a sequential labeling problem given **x**. In this case, input word sequences **x** are not only lexical strings, but also multiple linguistic features. To extract semantic frames from users' utterance inputs, we use a linear-chain CRF model, which is a model that assigns a joint probability distribution over labels that are conditional on the input sequences where the distribution respects the independent relations encoded in a graph [11].

A common special-case graph structure of CRF, which is suitable for our SLU problem, is a linear chain in which a first-order Markov assumption corresponds to a finite state machine. A linear-chain CRF is defined as follows. Let G be an undirected model over sets of random variables *y* and *x*. As in our SLU case, $\mathbf{s} = \{s_t\}$ and $\mathbf{x} = \{x_t\}$ for $t = 1,\cdots,T$, such that **s** is a semantic class labeling of an observed word sequence **x**. The graph G with parameter $M = \{\mu,\cdots\}$ defines a conditional probability for a state (or label) sequence $\mathbf{s} = s_1,\cdots,s_T$, given an in-put sequence $\mathbf{x} = x_1,\cdots,x_T$, to be

$$p(\mathbf{s}|\mathbf{x}) = \frac{1}{\mathbf{Z(x)}}\exp\left(\sum_{t=1}^{T}\sum_{k=1}^{K}\mu_k g_k(s_{t-1}, s_t, \mathbf{x}, t)\right),$$

where $Z(\mathbf{x})$ is the normalization factor that makes the probability of all state sequences sum to one, $g_k(s_{t-1},s_t,x,t)$ is an arbitrary linguistic feature function which is often binary-valued in NLP tasks, and μ_k is a trained parameter associated with feature g_k. The feature functions can encode any aspect of a state transition, $s_{t-1} \rightarrow s_t$, and the observation (a set of observable features) x centered at the current time *t*. Large positive values for μ_k indicate a preference for such an event, while large negative values make the event unlikely. Note that CRF is a structured version of ME, that is, ME is identical to a state-independent CRF model.

We use a limited memory version of the quasi-Newton method (L-BFGS) to optimize the objective function, that is, the 2-norm penalized log-likelihood. The L-BFGS method converges super-linearly to the solution; therefore it can be an efficient optimization technique on large-scale NLP problems [18]. To train the model, we use the following features for the component slot extraction task.

- We use the lexical word w_i and POS tags t_i with indicators for specific words at location *i*, or locations within five words of *i* (-2, -1, 0, +1, +2 words on current position *i*).

IV. Voting-Based Multi-Strategic Approach for Robust SLU

As previously stated, the tasks on dialog act and main goal slots are treated as classification tasks. We separately adjust several classifiers based on ME to these tasks, namely, a support vector machine (SVM), and artificial neural networks (NN). The performance of the classifiers drop sharply when erroneously recognized utterances are given as input for the classification. For this reason, we propose combining these classifiers using a simple voting method for better performance and more robust spoken language understanding.

The structure of a voting classifier is shown in Fig. 2. Three classifiers use the same basic features as in the previous basic spotting method using the ME classifier.

In addition to our ME framework, we use two more classifiers for the voting: SVM and NN. An SVM is a powerful machine learning model introduced by Vapnik [19], which offers the possibility to train generalized, nonlinear classifiers in high dimensional spaces. An SVM generalization error is not related to the input dimensionality of the problem but to the margin with which it separates the data [20]. Although SVM is generally a binary classifier, we used an implementation of a multi-class version (LIBSVM) [21] which is modified for multiple classes to apply to dialog act and main goal classification. On the other hand, NNs provide a general and practical method for learning real-valued, discrete-valued, and vector-valued functions in a way that is robust to noise in training data. In a rough analogy, an NN is built out of a densely interconnected set of simple units, where each unit takes a number of real-valued inputs and produces a single real-valued output. The Stuttgart Neural Network Simulator (SNNS) toolkit [22] was used to construct an NN, which was the radial-basis-function (RBF) type with 2 hidden layers. The NN architecture consists of 830 units comprising an input layer, 4 units comprising an output layer, and 100*100 units of hidden layers.

Many of the simple techniques that aim to combine much evidence into a single prediction are based on voting [23]. One of the several voting methods is a simple process. Based on the predictions of the different base classifiers, a final prediction is chosen as the classification with a plurality of votes [24], [25]. In our case, if the three classifiers disagree with each prediction, then the prediction of the classifier that has the best individual performance should receive the vote.

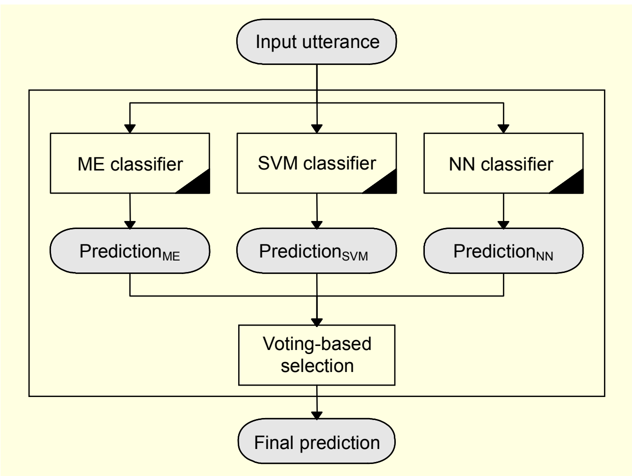


Fig. 2. Structure of the voting classifier.

V. Experiment

1. Data Sets

We evaluated our system on two sets of Korean dialog data; Car-Navi (car navigation service domain) and Telebank (tele-banking service domains). These two data sets were collected and annotated for developing spoken dialog systems. They consist of three classes: dialog act, main goal, and component slots. Some statistics of the two data sets are presented in Table 3. The average number of words per utterance is less than 7 to 14 words and the average numbers of classes per utterance is nearly 1 to 3 in these data sets. The pre-defined concepts are 5/4/6 class types of dialog act, main goal, and component slots for the Car-Navi domain and 5/25/17 classes for the Telebank domain. Figure 3 shows the statistic of the annotated labels of the two data sets.[1)]

All experiments were implemented in C++ and executed on Linux with XEON 2.8 GHz dual processors and 2.0 GB of main memory.

Table 3. Corpus statistics.

Data set	# sent	# word	# slot	DA	MG	CS
Car-Navi	462	6,552	1,335	5	4	6
Telebank	2,239	16,481	2,191	5	25	17

2. Concept-Spotting SLU Results

To evaluate our concept-spotting SLU system, we used 10-fold cross validation using the Car-Navi and Telebank data sets. We used two Korean speech recognizers: LG-Elite (LG Electronics Institute of Technology) for the Car-Navi domain and an HTK-based Korean speech recognizer (morpheme-based recognition system). Their performance indicates word error rates (WER) of 39.0% and 18.71% for the Car-Navi and Telebank data sets, respectively. Our concept-spotting SLU system was trained on transcripts and tested on both transcripts and spoken utterances.

The results in Table 4 show that the concept-spotting SLU

1) In Figure 3(d), *search_info and *confirm include (confirm_benefit, confirm_care, confirm_info, confirm_limit, confirm_qualified, and confirm_service) and (search_benefit, search_calculate, search_count, search_how_to, search_info, search_kind, search_limit, search_profit, search_qualified, search_rate, search_refund, search_service, and search_state) for visualizing the class distribution.

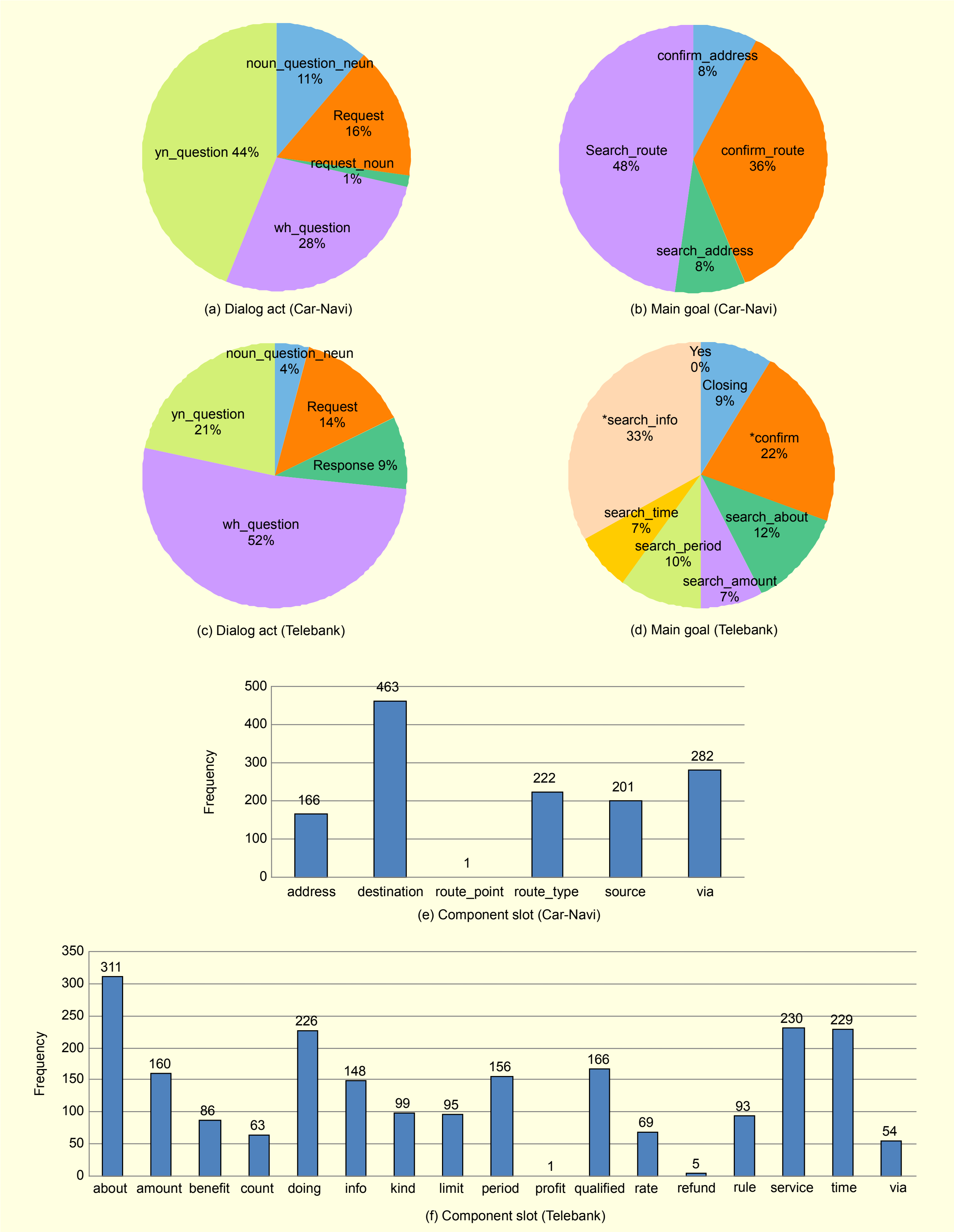


Fig. 3. Statistics for Car-Navi and Telebank data sets.

performance differences of transcript versus spoken input for the two data sets are only 11.9, 9.7, and 13.3 reduction in F-measure[2)] although the WER with spoken input is 39.0% higher than that with transcripts. The decreased performance rates for dialog act classification, main goal classification, and component slot extraction tasks with the spoken input are significantly less than the increased WER rate, demonstrating that our concept-spotting SLU method is very robust to highly erroneous speech inputs. Table 5 shows the results of our concept-spotting SLU for the Telebank data set. This result is more sensible to speech recognition noises, because users' utterances are more complex than simple query-oriented dialogs.

Table 4. SLU performance in the Car-Navi domain (F1).

	WER 0%	WER 39%	Reduction rate
Dialog act	98.9	87.0	11.9 ↓
Main goal	99.3	89.6	9.7 ↓
Component slot	91.2	77.9	13.3 ↓

Table 5. SLU performance in the Telebank domain (F1).

	WER 0%	WER 19%	Reduction rate
Dialog act	97.63	85.75	11.88 ↓
Main goal	95.80	82.76	13.04 ↓
Component slot	88.67	83.40	5.27 ↓

3. Results of Voting Methods

To evaluate the multiple-classifier method, we used the Telebank data. In the case of the dialog act slot, the respective performance figures of the classifiers were F-measure 97.63, 98.66, and 97.95 for the transcript, and 85.75, 85.41, and 86.61 for the spoken input in Table 6. The results in this table show that the performance of ME, SVM, and NN classifiers for dialog act classification decreases as F-measure by 11.88, 13.25, and 11.34 when tested with spoken input (WER 18.71%). The performance of the combined classifier shows more endurance, decreasing only 10.72 in F-measure.

A similar performance trend can be confirmed for the main goal recognition task. In this case, the number of classes for the main goal slot was 25 and the performance figures of the classifiers were F-measure 95.80, 96.38, and 94.16 for the transcript and 82.76, 81.24, and 79.33 for the spoken input in Table 7. The results in this table demonstrate that the performance of the classifiers for main goal classification decreases as F-measure by 13.04, 15.14, and 14.83 when tested with spoken inputs. As in the previous case, the performance of the combined classifier shows more endurance, decreasing only 11.43.

As the results of the dialog act and the main goal recognition tasks demonstrate, in general, the decreased understanding performance when tested with the several different error-level spoken inputs is significantly less than the increased rate of WER in speech recognition.[3)] Figures 4 and 5 illustrate these trends by comparing the slowly-decreasing performance of multiple classifiers (especially, voting classifiers) with the steeply-decreasing speech recognition performance (increases in WER). Therefore, the experimental results show that our multi-classifier based concept-spotting method is very robust for spoken language inputs that have large errors.

Table 6. Performance of the dialog act classifiers in the telebanking service domain.

	WER 0%	WER 19%	Reduction rate
ME classifier	97.63	85.75	11.88 ↓
SVM classifier	98.66	85.41	13.25 ↓
NN classifier	97.95	86.61	11.34 ↓
Voting classifier	98.30	87.58	10.72 ↓

Table 7. Performance of the main goal classifiers in the telebanking service domain.

	WER 0%	WER 19%	Reduction rate
ME classifier	95.80	82.76	13.04 ↓
SVM classifier	96.38	81.24	15.14 ↓
NN classifier	94.16	79.33	14.83 ↓
Voting classifier	96.56	85.13	11.43 ↓

VI. Conclusion

This paper proposed a new concept-spotting approach as a spoken language understanding method for highly erroneous Korean speech recognition environments. We demonstrated that the IE paradigm can be successfully used for spoken language understanding by finding the values of the pre-defined slots of a task-specific meaning representation using various observation features of a sentence. The decreased

2) Evaluation of SLU is measured in terms of F1-measure which is the harmonic mean between the precision (P) and recall (R) for concept type/value matching with the gold standard, and is defined as $F1 = (2 * P * R) / (P + R)$.

3) In this result, we used three levels of WERs to evaluate performance decreases. Using the error-corrective adaptation method proposed in [26], we obtained the WERs of 18.71, 14.10, and 12.21%.

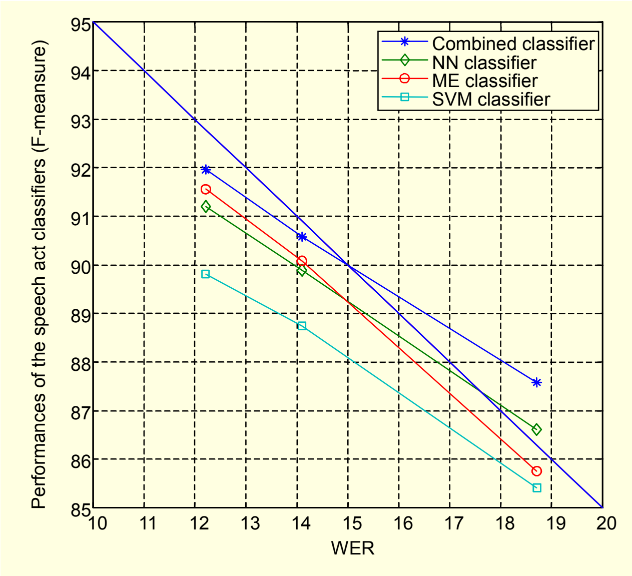


Fig. 4. Performance trend of the dialog act classifiers in the telebanking service domain.

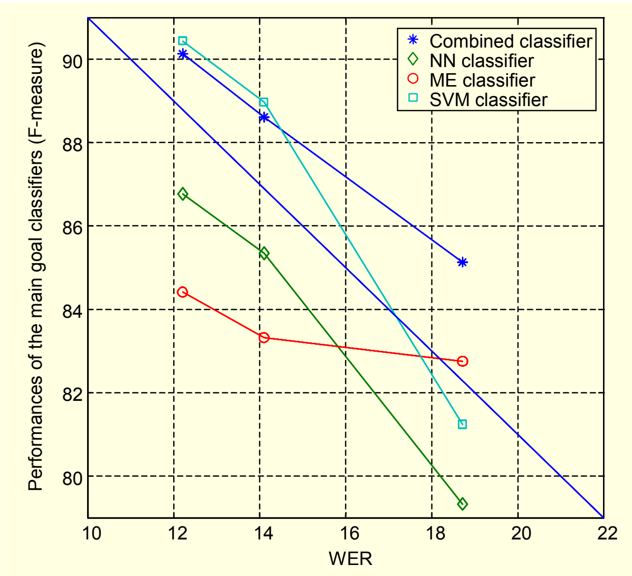


Fig. 5. Performance trend of the main goal classifiers in the telebanking service domain.

performance rates of dialog act classification, main goal classification, and component slot extraction tasks with spoken input are significantly less than the increased WER rate in speech recognition.

We also proposed a multi-strategic method to directly classify sentences into the intended meaning representations. In particular, we proposed a voting-based selection method from several classifiers (ME, SVM, and NN). According to the results of the dialog act and main goal recognition tasks, the voting-based method showed more robust performance when tested with spoken Korean inputs compared to the single classifier-based concept-spotting method.

References

[1] M.F. McTear, *Spoken Dialogue Technology*, Springer, 2004.

[2] R. Cole (ed.), *Survey of the State of the Art in Human Language Technology*, Cambridge University Press, New York, NY, USA, 1997.

[3] C.-K. Lee, G.-B. Lee, and M.-G. Kim, "Dependency Structure Applied to Language Modeling for Information Retrieval," *ETRI J.*, vol. 28, no. 3. 2006, pp. 337-346.

[4] S. Seneff, "TINA: A Natural Language System for Spoken Language Applications," *Computational Linguistics*, vol. 18, no. 1, 1992, pp. 61-86.

[5] W. Ward and B. Pellom, "THE CU-Communicator System," *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*, Keystone, Colorado, 1999.

[6] E. Levin and R. Pieraccini, "CHRONUS, the Next Generation," *Proc. 1995 ARPA Spoken Language Systems Technical Workshop*, 1995, pp. 269-271.

[7] R. Schwartz, S. Miller, D. Stallard, and J. Makhoul, "Hidden Understanding Models for Statistical Sentence Understanding," *Proc. ICASSP*, 2:1479, 1997.

[8] Y. He and S.J. Young, "Semantic Processing using the Hidden Vector State Model," *Computer Speech and Language*, vol. 19, no. 1, 2005, pp. 85-106.

[9] Y. Wang, A. Acero, M. Mahajan, and J. Lee, "Combining Statistical and Knowledge-Based Spoken Language Understanding in Conditional Models," *Proc. COLING/ACL*, July 2006.

[10] C. Wutiwiwatchai and S. Furui, "A Multi-Stage Approach for Thai Spoken Language Understanding," *Speech Communication*, In Press, Corrected Proof, Available online 21 April 2005.

[11] J. Lafferty, A. McCallum, and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," *Proc. Int'l Conf. Machine Learning*, 2001.

[12] D. Pinto, A. McCallum, X. Lee, and W.B. Croft, "Table Extraction Using Conditional Random Fields," *Proc. ACM SIGIR*, 2003.

[13] F. Sha and F. Pereira, Shallow Parsing with Conditional Random Fields, *Proc. HLT-NAACL*, 2003.

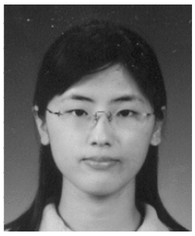
[14] F. Peng and A. McCallum, "Accurate Information Extraction from Research Papers using Conditional Random Fields," *Proc. HLT-NAACL*, 2004.

[15] M. Jeong and G.G. Lee, "Exploiting Non-Local Features for Spoken Language Understanding," *Proc. COLING/ACL*, July 2006.

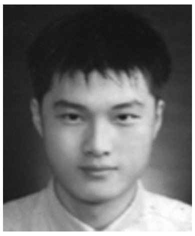
[16] J. Cowie and W. Lehnert, "Information Extraction," *Communications of the ACM*, vol. 39, no. 1, 1996, pp. 80-91.

[17] A. Ratnaparkhi, *Maximum Entropy Models for Natural Language Ambiguity Resolution*, PhD thesis, University of Pennsylvania, 1998.
[18] R. Malouf, "A Comparison of Algorithms for Maximum Entropy Parameter Estimation," *Proc. 6th CoNLL*, 2002.
[19] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, NY, USA, 1995.
[20] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, 1998, pp. 121-167.
[21] C. Chang and C. Lin, LIBSVM - *A Library for Support Vector Machines*, National Taiwan University, 2006.
[22] A. Zell, G. Mamier, M. Vogt, N. Mache, R. Hubner, S. Doring, K.U. Herrmann, T. Soyez, M. Schmalzl, T. Sommer, A. Hatzigeorgiou, D. Posselt, T. Schreiner, B. Kett, G. Clemente, and J. Wieland, *Stuttgart Neural Network Simulator, User Manual*, version 4.1, University of Stuttgart, 1994.
[23] T.G. Dietterich, "Ensemble Methods in Machine Learning," J. Kittler and F. Roli (ed.*) First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science*, 2002, pp. 1-15.
[24] E. Brill and J. Wu, Classifier Combination for Improved Lexical Disambiguation, *Proc. COLING-ACL*, 1998, pp. 191-195.
[25] R. Mihalcea, "Classifier Stacking and Voting for Text Filtering," *The Eleventh Text Retrieval Conference (TREC 2002)*, 2002.
[26] M. Jeong, J. Eun, S. Jung, and G.G. Lee, "An Error-Corrective Language-Model Adaptation for Automatic Speech Recognition," *Interspeech 2005-Eurospeech*, Lisbon, 2005.

Changki Lee received the BS degree in computer science from KAIST, Korea, in 1999. He received the MS and PhD degrees in computer engineering from POSTECH, Korea, in 2001 and 2004, respectively. Since 2004, he has been with Electronics and Telecommunications Research Institute (ETRI), Korea, as a senior member of research staff. He has served as a reviewer for some international journals such as *Information System and Information Processing & Management*. His research interests are natural language processing, information retrieval, information extraction, question answering, spoken language understanding, and semantic web.

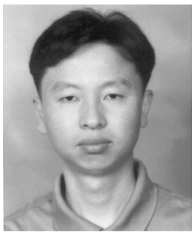


Jihyun Eun received the BS and MS degrees in computer science and engineering from POSTECH, Korea, in 2002 and 2006. Since 2006, she has been with Korea Telecom as a member of research staff. Her research interests currently include natural language processing and spoken language understanding.



Minwoo Jeong received the BS degree in computer engineering from Chonbuk National University in 2003. He received the MS degree in computer science and engineering from POSTECH, Korea, in 2005. He now is a PhD student of POSTECH. His primary research interests are language understanding and modeling, spoken dialog systems, text mining, and practical machine learning methods for NLP.

Gary Geunbae Lee received his BS and MS degrees in computer engineering from Seoul National University in 1984 and 1986, respectively. He received his PhD in computer science from UCLA in 1991 and was a Research Scientist in UCLA. He was an Assistant Professor from 1992 to 1996, and an Associate Professor since 1997 at POSTECH, Korea. In 2004, he was promoted to a Full Professor in POSTECH. He has authored more than 100 papers in international journals and conferences and has served as a technical committee member and reviewer for several international conferences such as ACL, COLING, ACM SIGIR, IRAL, EMNLP, and so on. His current research interests include natural language processing, biological text mining, spoken language understanding, and TTS systems.



YiGyu Hwang received the BS degree in computer science from Chonbuk National University, Korea, in 1993. He received the MS and PhD degrees in computer science & statistics from Chonbuk National University, Korea, in 1995 and 2001, respectively. Since 2001, he has been with Electronics and Telecommunications Research Institute (ETRI), Korea, as a senior member of research staff. His research interests are natural language processing, information retrieval, and text mining.

Myung-Gil Jang received the BS and MS degrees in computer science and statistics from Pusan National University, Korea, in 1988 and 1990. He received the PhD degree in information science from Chungnam National University, Korea, in 2002. He was with System Engineering Research Institute (SERI), Korea, from 1990 to 1997 as a researcher. Since 1998, he has been with Electronics and Telecommunications Research Institute (ETRI), Korea, as a Senior/Principle Member of Research Staff. His research interests are natural language processing, information retrieval, question answering, knowledge and dialogue processing, media retrieval/management, and semantic web.