

Stochastic Discourse Modeling in Spoken Dialogue Systems Using Semantic Dependency Graphs

Jui-Feng Yeh, Chung-Hsien Wu and Mao-Zhu Yang

Department of Computer Science and Information Engineering

National Cheng Kung University

No. 1, Ta-Hsueh Road, Tainan, Taiwan, R.O.C.

{jfyeh, chwu, mzyang}@csie.ncku.edu.tw

Abstract

This investigation proposes an approach to modeling the discourse of spoken dialogue using semantic dependency graphs. By characterizing the discourse as a sequence of speech acts, discourse modeling becomes the identification of the speech act sequence. A statistical approach is adopted to model the relations between words in the user's utterance using the semantic dependency graphs. Dependency relation between the headword and other words in a sentence is detected using the semantic dependency grammar. In order to evaluate the proposed method, a dialogue system for medical service is developed. Experimental results show that the rates for speech act detection and task-completion are 95.6% and 85.24%, respectively, and the average number of turns of each dialogue is 8.3. Compared with the Bayes' classifier and the Partial-Pattern Tree based approaches, we obtain 14.9% and 12.47% improvements in accuracy for speech act identification, respectively.

1 Introduction

It is a very tremendous vision of the computer technology to communicate with the machine using spoken language (Huang et al., 2001; Allen et al., 2001). Understanding of spontaneous language

is arguably the core technology of the spoken dialogue systems, since the more accurate information obtained by the machine (Higashinaka et al., 2004), the more possibility to finish the dialogue task. Practical use of speech act theories in spoken language processing (Stolcke et al. 2000; Walker and Passonneau 2001; Wu et al., 2004) have given both insight and deeper understanding of verbal communication. Therefore, when considering the whole discourse, the relationship between the speech acts of the dialogue turns becomes extremely important. In the last decade, several practicable dialogue systems (McTEAR, 2002), such as air travel information service system, weather forecast system, automatic banking system, automatic train timetable information system, and the Circuit-Fix-it shop system, have been developed to extract the user's semantic entities using the semantic frames/slots and conceptual graphs. The dialogue management in these systems is able to handle the dialogue flow efficaciously. However, it is not applicable to the more complex applications such as "Type 5: the natural language conversational applications" defined by IBM (Rajesh and Linda, 2004). In Type 5 dialog systems, it is possible for the users to switch directly from one ongoing task to another. In the traditional approaches, the absence of precise speech act identification without discourse analysis will result in the failure in task switching. The capability for identifying the speech act and extracting the semantic objects by reasoning plays a more important role for the dialog systems. This research proposes a semantic dependency-based discourse model to capture and share the semantic objects among tasks that switch during a dialog for semantic resolution. Besides

acoustic speech recognition, natural language understanding is one of the most important research issues, since understanding and application restriction on the small scope is related to the data structures that are used to capture and store the meaningful items. Wang et al. (Wang et al., 2003) applied the object-oriented concept to provide a new semantic representation including semantic class and the learning algorithm for the combination of context free grammar and N-gram.

Among these approaches, there are two essential issues about dialogue management in natural language processing. The first one is how to obtain the semantic object from the user's utterances. The second is a more effective speech act identification approach for semantic understanding is needed. Since speech act plays an important role in the development of dialogue management for dealing with complex applications, speech act identification with semantic interpretation will be the most important topic with respect to the methods used to control the dialogue with the users. This paper proposes an approach integrating semantic dependency graph and history/discourse information to model the dialogue discourse (Kudo and Matsumoto, 2000; Hacioglu et al., 2003; Gao and Suzuki, 2003). Three major components, such as semantic relation, semantic class and semantic role are adopted in the semantic dependency graph (Gildea and Jurafsky, 2002; Hacioglu and Ward, 2003). The semantic relations constrain the word sense and provide the method for disambiguation. Semantic roles are assigned when the relation established among semantic objects. Both semantic relations and roles are defined in many knowledge resources or ontologies, such as FrameNet (Baker et al., 2004) and HowNet with 65,000 concepts in Chinese and close to 75,000 English equivalents, is a bilingual knowledge-base describing relations between concepts and relations between the attributes of concepts with ontological view (Dong and Dong 2006). Generally speaking, semantic class is defined as a set with the elements that are usually the words with the same semantic interpretation. Hypernyms that are superordinate concepts of the words are usually used as the semantic classes just like the Hypernyms of synsets in WordNet (<http://www.cogsci.princeton.edu/~wn/>) or definitions of words' primary features in HowNet. Besides, the approach for understanding tries to find the implicit semantic dependency between the con-

cepts and the dependency structure between concepts in the utterance are also taken into consideration. Instead of semantic frame/slot, semantic dependency graph can keep more information for dialogue understanding.

2 Semantic Dependency Graph

Since speech act theory is developed to extract the functional meaning of an utterance in the dialogue (Searle, 1979), discourse or history can be defined as a sequence of speech acts, $H^t = \{SA^1, SA^2, \dots, SA^{t-1}, SA^t\}$, and accordingly the speech act theory can be adopted for discourse modeling. Based on this definition, the discourse analysis in semantics using the dependency graphs tries to identify the speech act sequence of the discourse. Therefore, discourse modeling by means of speech act identification considering the history is shown in Equation (1). By introducing the hidden variable D_i , representing the i -th possible dependency graph derived from the word sequence W . The dependency relation, r_k , between word w_k and headword w_{kh} is extracted using HowNet and denoted as $DR(w_k, w_{kh}) \equiv r_k$. The dependency graph which is composed of a set of dependency relations in the word sequence W is defined as

$$D_i(W) = \{DR_1^i(w_1, w_{1h}), DR_2^i(w_2, w_{2h}), \dots, DR_{m-1}^i(w_{m-1}, w_{(m-1)h})\}.$$

The probability of hypothesis SA^t given word sequence W and history H^{t-1} can be described in Equation (1). According to the Bayes' rule, the speech act identification model can be decomposed into two components, $P(SA^t | D_i, W, H^{t-1})$ and $P(D_i | W, H^{t-1})$, described in the following.

$$\begin{aligned} SA^* &= \arg \max_{SA^t} P(SA^t | W, H^{t-1}) \\ &= \arg \max_{SA^t} \sum_{D_i} P(SA^t, D_i | W, H^{t-1}) \\ &= \arg \max_{SA^t} \sum_{D_i} P(SA^t | D_i, W, H^{t-1}) \times P(D_i | W, H^{t-1}) \end{aligned} \quad (1)$$

where SA^* and SA^t are the most probable speech act and the potential speech act at the t -th dialogue turn, respectively. $W = \{w_1, w_2, w_3, \dots, w_m\}$ denotes the word sequence extracted from the user's utterance without considering the stop words. H^{t-1} is the history representing the previous $t-1$ turns.

2.1 Speech act identification using semantic dependency with discourse analysis

In this analysis, we apply the semantic dependency, word sequence, and discourse analysis to the identification of speech act. Since D_i is the i -th possible dependency graph derived from word sequence W , speech act identification with semantic dependency can be simplified as Equation (2).

$$P(SA^t | D_i, W, H^{t-1}) \cong P(SA^t | D_i, H^{t-1}) \quad (2)$$

According to Bayes' rule, the probability $P(SA^t | D_i, H^{t-1})$ can be rewritten as:

$$P(SA^t | D_i, H^{t-1}) = \frac{P(D_i, H^{t-1} | SA^t) P(SA^t)}{\sum_{SA_t} P(D_i, H^{t-1} | SA_t) P(SA_t)} \quad (3)$$

As the history is defined as the speech act sequence, the joint probability of D_i and H^{t-1} given the speech act SA^t can be expressed as Equation (4). For the problem of data sparseness in the training corpus, the probability, $P(D_i, SA^1, SA^2, \dots, SA^{t-1} | SA^t)$, is hard to obtain and the speech act bi-gram model is adopted for approximation.

$$\begin{aligned} & P(D_i, H^{t-1} | SA^t) \\ &= P(D_i, SA^1, SA^2, \dots, SA^{t-1} | SA^t) \quad (4) \\ &\cong P(D_i, SA^{t-1} | SA^t) \end{aligned}$$

For the combination of the semantic and syntactic structures, the relations defined in HowNet are employed as the dependency relations, and the hypernym is adopted as the semantic concept according to the primary features of the words defined in HowNet. The headwords are decided by the algorithm based on the part of speech (POS) proposed by Academia Sinica in Taiwan. The probabilities of the headwords are estimated according to the probabilistic context free grammar (PCFG) trained on the Treebank developed by Sinica (Chen et al., 2001). That is to say, the headwords are extracted according to the syntactic structure and the dependency graphs are constructed by the semantic relations defined in HowNet. According to previous definition with independent assumption and

the bigram smoothing of the speech act model using the back-off procedure, we can rewrite Equation (4) into Equation (5).

$$\begin{aligned} & P(D_i, SA^{t-1} | SA^t) \\ &= \alpha \prod_{k=1}^{m-1} P(DR_k^i(w_k, w_{kh}), SA^{t-1} | SA^t) + \quad (5) \\ & \quad (1 - \alpha) \prod_{k=1}^{m-1} P(DR_k^i(w_k, w_{kh}) | SA^t) \end{aligned}$$

where α is the mixture factor for normalization.

According to the conceptual representation of the word, the transformation function, $f(\cdot)$, transforms the word into its hypernym defined as the semantic class using HowNet. The dependency relation between the semantic classes of two words will be mapped to the conceptual space. Also the semantic roles among the dependency relations are obtained. On condition that SA^t , SA^{t-1} and the relations are independent, the equation becomes

$$\begin{aligned} & P(DR_k^i(w_k, w_{kh}), SA^{t-1} | SA^t) \\ &\cong P(DR_k^i(f(w_k), f(w_{kh})), SA^{t-1} | SA^t) \quad (6) \\ &= P(DR_k^i(f(w_k), f(w_{kh})) | SA^t) P(SA^{t-1} | SA^t) \end{aligned}$$

The conditional probability, $P(DR_k^i(f(w_k), f(w_{kh})) | SA^t)$ and $P(SA^{t-1} | SA^t)$, are estimated according to Equations (7) and (8), respectively.

$$\begin{aligned} & P(DR_k^i(f(w_k), f(w_{kh})) | SA^t) \\ &= \frac{C(f(w_k), f(w_{kh}), r_k, SA^t)}{C(SA^t)} \quad (7) \end{aligned}$$

$$P(SA^{t-1} | SA^t) = \frac{C(SA^{t-1}, SA^t)}{C(SA^t)} \quad (8)$$

where $C(\cdot)$ represents the number of events in the training corpus. According to the definitions in Equations (7) and (8), Equation (6) becomes practicable.

2.2 Semantic dependency analysis using word sequence and discourse

Although the discourse can be expressed as the speech act sequence $H^t = \{SA^1, SA^2, \dots, SA^{t-1}, SA^t\}$, the dependency graph D_i is determined mainly by W , but not H^{t-1} . The probability that defines semantic dependency analysis using the words sequence and discourse can be rewritten in the following:

$$\begin{aligned} P(D_i | W, H^{t-1}) \\ = P(D_i | W, SA^{t-1}, SA^{t-2}, \dots, SA^1) \\ \cong P(D_i | W) \end{aligned} \quad (9)$$

and

$$P(D_i | W) = \frac{P(D_i, W)}{P(W)} \quad (10)$$

Seeing that several dependency graphs can be generated from the word sequence W , by introducing the hidden factor D_i , the probability $P(W)$ can be the sum of the probabilities $P(D_i, W)$ as Equation (11).

$$P(W) = \sum_{D_i : yield(D_i)=W} P(D_i, W) \quad (11)$$

Because D_i is generated from W , D_i is the sufficient to represent W in semantics. We can estimate the joint probability $P(D_i, W)$ only from the dependency relations D_i . Further, the dependency relations are assumed to be independent with each other and therefore simplified as

$$P(D_i, W) = \prod_{k=1}^{m-1} P(DR_k^i(w_k, w_{kh})) \quad (12)$$

The probability of the dependency relation between words is defined as that between the concepts defined as the hypernyms of the words, and then the dependency rules are introduced. The probability $P(r_k | f(w_k), f(w_{kh}))$ is estimated from Equation (13).

$$\begin{aligned} P(DR_k^i(w_k, w_{kh})) \\ \equiv P(DR_k^i(f(w_k), f(w_{kh}))) \\ = P(r_k | f(w_k), f(w_{kh})) \\ = \frac{C(r_k, f(w_k), f(w_{kh}))}{C(f(w_k), f(w_{kh}))} \end{aligned} \quad (13)$$

According to Equations (11), (12) and (13), Equation (10) is rewritten as the following equation.

$$\begin{aligned} P(D_i | W) &= \frac{\prod_{k=1}^{m-1} P(DR_k^i(w_k, w_{kh}))}{\sum_{D_i : yield(D_i)=W} \prod_{k=1}^{m-1} P(DR_k^i(w_k, w_{kh}))} \\ &= \frac{\prod_{k=1}^{m-1} \frac{C(r_k, f(w_k), f(w_{kh}))}{C(f(w_k), f(w_{kh}))}}{\sum_{D_i : yield(D_i)=W} \prod_{k=1}^{m-1} \frac{C(r_k, f(w_k), f(w_{kh}))}{C(f(w_k), f(w_{kh}))}} \end{aligned} \quad (14)$$

where function, $f(\cdot)$, denotes the transformation from the words to the corresponding semantic classes.

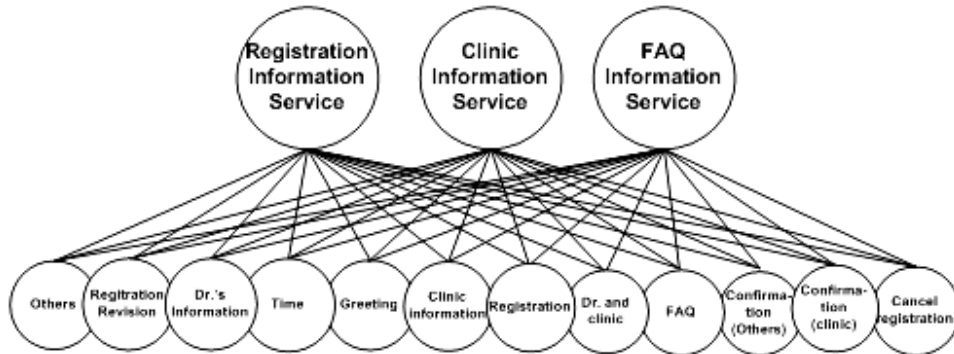


Figure 1. Speech acts corresponding to multiple services in the medical domain

3 Experiments

In order to evaluate the proposed method, a spoken dialogue system for medical domain with multiple services was investigated. Three main services: registration information service, clinic information service, and FAQ information service are used. This system mainly provides the function of on-line registration. For this goal, the health education documents are provided as the FAQ files. And the inference engine about the clinic information according to the patients' syndromes is constructed according to a medical encyclopedia. An example is illustrated as figure 2:

System: 午安。需要服務嗎? Good Afternoon. May I help you?
User: 是的，我要看醫生。 Yes, I should be diagnosed by doctor.
System: 請問您要掛哪一科? Which department would you like to make an appointment for?
User: 我不知道。 I do not know.
System: 請問你有那些症狀? What symptoms are with you?
User: 我喉嚨痛又發燒。而且全身酸痛。 I am with a sore throat and a fever. I ached all over, too.
System: 好的，這些症狀多久了? Well, how long have you had these symptoms?
User: 好幾天了。 For several days.
System: 建議您掛內科醫師。 My suggestion is that you may visit the Internal Medicine.
User: 請問內科何時有門診? What's the visiting schedule of the Internal Medicine?
System: 您可以參考醫院門診表，決定您的時間，我們的時間有... You can check the Services Schedule and decide a convenient time for you. The Available time for now is....

Figure 2 An example of dialog

12 Speech acts are defined and shown in Figure 1. Every service corresponds to the 12 speech acts with different probabilities.

The acoustic speech recognition engine embedded in dialog system based on Hidden Markov Models (HMMs) was constructed. The feature vector is parameterized on 26 MFCC coefficients. The decoding strategy is based on a classical Viterbi algorithm. The evaluation results by the character error rate (CER) for a Chinese speech recognition system is 18.3 percent and the vocabulary size of the language is 25,132.

3.1 Analysis of corpus

The training corpus was collected using the on-line recording from National Cheng Kung University Hospital in the first phase and the Wizard-of-Oz method in the second phase. Totally, there are 1,862 dialogues with 13,986 sentences in the corpus. The frequencies of the speech acts used in the system are shown in Figure 3.

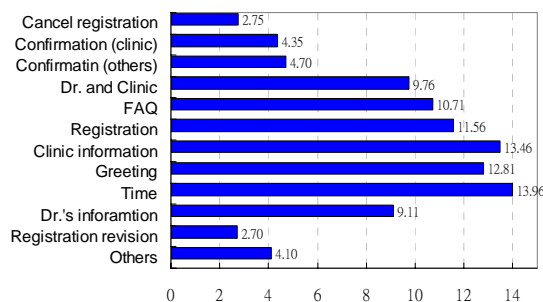


Figure 3 Frequencies for each speech act

The number of dialogue turns is also important to the success of the dialogue task. According to the observation of the corpus, we can find that the dialogues with more than 15 turns usually failed to complete the dialogue, that is to say, the common ground cannot be achieved. These failed dialogues were filtered out from the training corpus before conducting the following experiments. The distribution of the number of turns per dialogue is shown in Figure 4.

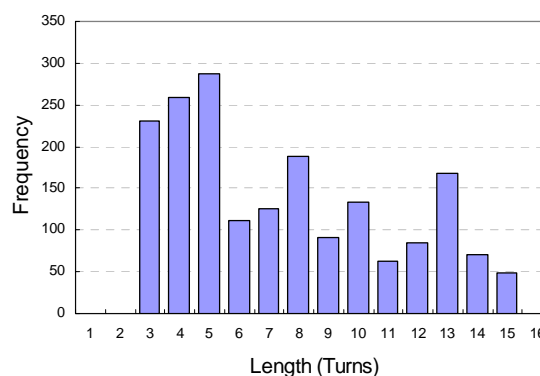


Figure 4. The distribution of the number of turns per dialogue

3.2 Precision of speech act identification related to the corpus size

The size of the training corpus is crucial to the practicability of the proposed method. In this experiment, we analyze the effect of the number of sentences according to the precision rate of the speech act using the semantic dependency graphs with and without the discourse information. From the results, the precision rates for speech act identification achieved 95.6 and 92.4 percentages for the training corpus containing 10,036 and 7,012 sentences using semantic dependency graphs with and without history, respectively. This means that semantic dependency graph with discourse outperforms that without discourse, but more training data are needed to include the discourse for speech act identification. Fig. 5 shows the relationship between the speech act identification rate and the size of the training corpus. From this figure, we can find that more training sentences for the semantic dependency graph with discourse analysis are needed than that without discourse. This implies discourse analysis plays an important role in the identification of the speech act.

3.3 Performance analysis of semantic dependency graph

To evaluate the performance, two systems were developed for comparison. One is based on the Bayes' classifier (Walker et al., 1997), and the other is the use of the partial pattern tree (Wu et al., 2004) to identify the speech act of the user's utterances. Since the dialogue discourse is defined as a sequence of speech acts. The prediction of speech

act of the new input utterance becomes the core issue for discourse modeling. The accuracy for speech act identification is shown in Table 1.

According to the observation of the results, semantic dependency graphs obtain obvious

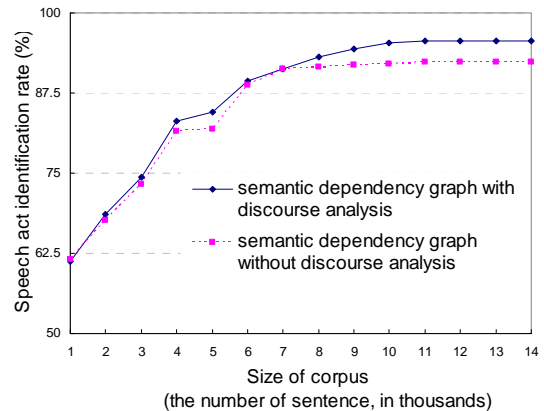


Figure 5. The relation between the speech act identification rate and the size of training corpus

improvement compared to other approaches. The reason is that not only the meanings of the words or concepts but also the structural information and the implicit semantic relation defined in the knowledge base are needed to identify the speech act. Besides, taking the discourse into consideration will improve the prediction about the speech act of the new or next utterance. This means the discourse model can improve the accuracy of the speech act identification, that is to say, discourse modeling can help understand the user's desired intension especially when the answer is very short.

Speech act	Semantic dependency graph		PPT	Bayes' Classifier
	With discourse analysis	Without discourse analysis		
Clinic information (26 sentences)	100 (26)	96.1 (25)	88 (23)	92 (24)
Dr.'s information (42 sentences)	97 (41)	92.8 (39)	66.6 (28)	92.8 (39)
Confirmation(others) (42 sentences)	95 (40)	95 (40)	95 (40)	95 (40)
Others (14 sentences)	57.1 (8)	50 (7)	43 (6)	38 (5)
FAQ (13 sentences)	70 (9)	53.8 (7)	61.5 (8)	46 (6)
Clinic information (135 sentences)	98.5 (133)	96.2 (130)	91.1 (123)	93.3 (126)
Time (38)	94.7 (36)	89.4 (34)	97.3 (37)	92.1 (35)
Registration (75)	100 (75)	100 (75)	86.6 (65)	86.6 (65)
Cancel registration (10)	90 (9)	80 (8)	60 (6)	80 (8)
Average Precision	95.6	92.4	85	88.1

Table 1 The accuracy for speech act identification

For example, the user may only say “yes” or “no” for confirmation. The misclassification in speech act will happen due to the limited information. However, it can obtain better interpretation by introducing the semantic dependency relations as well as the discourse information.

To obtain the single measurement, the average accuracy for speech act identification is shown in Table 1. The best approach is the semantic dependency graphs with the discourse. This means the information of the discourse can help speech act identification. And the semantic dependency graph outperforms the traditional approach due to the semantic analysis of words with their corresponding relations.

The success of the dialog lies on the achievement of the common ground between users and machine which is the most important issue in dialogue management. To compare the semantic dependency graph with previous approaches, 150 individuals who were not involved in the development of this project were asked to use the dialogue system to measure the task success rate. To filter out the incomplete tasks, 131 dialogs were employed as the analysis data in this experiment. The results are listed in Table 2.

	SDG¹	SDG²	PPT	Bayes'
Task completion rate	87.2	85.5	79.4	80.2
Number of turns on average	8.3	8.7	10.4	10.5

SDG¹:With discourse analysis, **SDG²**:Without discourse

Table 2 Comparisons on the Task completion rate and the number of dialogue turns between different approaches

We found that the dialogue completion rate and the average length of the dialogs using the dependency graph are better than those using the Bayes' classifier and partial pattern tree approach. Two main reasons are concluded: First, dependency graph can keep the most important information in the user's utterance, while in semantic slot/frame approach, the semantic objects not matching the semantic slot/frame are generally filtered out. This approach is able to skip the repe-

tion or similar utterances to fill the same information in different semantic slots. Second, the dependency graph-based approach can provide the inference to help the interpretation of the user's intension.

For semantic understanding, correct interpretation of the information from the user's utterances becomes inevitable. Correct speech act identification and correct extraction of the semantic objects are both important issues for semantic understanding in the spoken dialogue systems. Five main categories about medical application, clinic information, Dr.'s information, confirmation for the clinic information, registration time and clinic inference, are analyzed in this experiment.

	SDG	PPT	Bayes'
Clinic information	95.0	89.5	90.3
Dr.'s information	94.3	71.7	92.4
Confirmation (Clinic)	98.0	98.0	98.0
Clinic	97.3	74.6	78.6
Time	97.6	97.8	95.5

SDG:With discourse analysis

Table 3 Correction rates for semantic object extraction

According to the results shown in Table 3, the worst condition happened in the query for the Dr.'s information using the partial pattern tree. The mis-identification of speech act results in the un-matched semantic slots/frames. This condition will not happen in semantic dependency graph, since the semantic dependency graph always keeps the most important semantic objects according to the dependency relations in the semantic dependency graph instead of the semantic slots. Rather than filtering out the unmatched semantic objects, the semantic dependency graph is constructed to keep the semantic relations in the utterance. This means that the system can preserve most of the user's information via the semantic dependency graphs. We can observe the identification rate of the speech act is higher for the semantic dependency graph than that for the partial pattern tree and Bayes' classifier as shown in Table 3.

4 Conclusion

This paper has presented a semantic dependency graph that robustly and effectively deals with a variety of conversational discourse information in the spoken dialogue systems. By modeling the dialogue discourse as the speech act sequence, the predictive method for speech act identification is proposed based on discourse analysis instead of keywords only. According to the corpus analysis, we can find the model proposed in this paper is practicable and effective. The results of the experiments show the semantic dependency graph outperforms those based on the Bayes' rule and partial pattern trees. By integrating discourse analysis this result also shows the improvement obtained not only in the identification rate of speech act but also in the performance for semantic object extraction.

Acknowledgements

The authors would like to thank the National Science Council, Republic of China, for its financial support of this work, under Contract No. NSC 94-2213-E-006-018.

References

- J. F. Allen, D. K. Byron, D. M. Ferguson, L. Galescu, and A. Stent. 2001. Towards Conversational Human-Computer Interaction. *AI Magazine*.
- C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of COLING/ACL*. 86-90
- K. J. Chen, C. R. Huang, F.Y. Chen, C. C. Luo, M. C. Chang, and C.J. Chen. 2001. Sinica Treebank: Design Criteria, representational issues and implementation. In *Anne Abeille, editor, Building and Using Syntactically Annotated Corpora*. Kluwer. 29-37
- Z. Dong and Q. Dong. 2006. HowNet and the computation of meaning. *World Scientific Publishing Co Inc*.
- J. Gao, and H. Suzuki. 2003. Unsupervised learning of dependency structure for language modeling. In *Proceedings of ACL 2003*, 521-528.
- D. Gildea and D. Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3). 245-288.
- K. Hacioglu, S. Pradhan, W. Ward, J. Martin, and D. Jurafsky. 2003. Shallow semantic parsing using support vector machines. *Technical Report TR-CSLR-2003-1, Center for Spoken Language Research, Boulder, Colorado*.
- K. Hacioglu and W. Ward. 2003. Target word detection and semantic role chunking using support vector machines. In *HLT-03*.
- R. Higashinaka, N. Miyazaki, M. Nakano, and K. Aikawa. 2004. Evaluating Discourse Understanding in Spoken Dialogue Systems. *ACM Transactions on Speech and Language Processing (TSLP)*, Volume 1, 1-20.
- X. Huang, A. Acero, and H.-W. Hon. 2001. *Spoken Language Proceeding*. Prentice-Hall, Inc.
- T. Kudo and Y. Matsumoto. 2000. Japanese Dependency Structure Analysis Based on Support Vector Machines. In *Proceedings of the EMLNP*. 18-25
- M. F. McTEAR. 2002. Spoken Dialogue Technology: Enabling the Conversational User Interface. *ACM Computer Surveys*, Vol 34, No. 1, 90-169..
- B. Rajesh, and B. Linda. 2004. Taxonomy of speech-enabled applications (<http://www106.ibm.com/developerworks/wireless/library/wi-tax/>)
- J. Searle. 1979. *Expression and Meaning: Studies in the Theory of Speech Acts*. New York, Cambridge University Press.
- A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Van Ess-Dykema, and M. Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics* 26(3), 339--373.
- M. A. Walker, D. Litman, C. Kamm, and A. Abella, 1997. PARADISE: a general framework for evaluating spoken dialogue agents. In *Proceedings of the ACL*, 271-280
- M. Walker and R. Passonneau. 2001. DATE: a dialogue act tagging scheme for evaluation of spoken dialogue systems. In *Proceedings of the first international conference on Human language technology research*. 1-8.
- Y.-Y. Wang and A. Acero. 2003. Combination of CFG and N-gram Modeling in Semantic Grammar Learning, In *Proceedings of the Eurospeech Conference*. Geneva, Switzerland. September 2003.
- C.-H. Wu, J.-F. Yeh, and M.-J. Chen. 2004. Speech Act Identification using an Ontology-Based Partial Pattern Tree. in *Proceedings of ICSLP 2004*, Jeju, Korea, 2004.