Response Generation Based on Hierarchical Semantic Structure with POMDP Re-ranking for Conversational Dialogue Systems

Jui-Feng Yeh　　　　　　**Yuan-Cheng Chu**

Department of Computer Science and Information Engineering,
National Chiayi University,
No.300 Syuefu Rd., Chiayi City 60004, Taiwan (R.O.C.).

Ralph@mail.ncyu.edu.tw　s1000444@mail.ncyu.edu.tw

Abstract

Conversational spoken dialogue systems can assist individuals to communicate with machine to obtain relevant information to their problems efficiently and effectively. By referring to relevant response, individuals can understand how to interact with an intelligent system according to recommendations of dialogue systems. This work presents a response generation based on hierarchical semantic structure with POMDP Re-ranking for conversational dialogue systems to achieve this aim. The hierarchical semantic structure incorporates the historical information according to dialogue discourse to keep more than one possible values for each slot. According to the status of concept graph, the candidate sentences are generated. The near optimal response selected by POMDP Re-ranking strategy to achieve human-like communication. The MOS and recall/precision rates are considered as the criterion for evaluations. Finally, the proposed method is adopted for dialogue system in travel domain, and indicates its superiority in information retrieval over traditional approaches.

1 Introduction

Intelligent space is one of the new trends about computing environment construction. From providing the natural intelligent human machine interaction, conversational dialogue systems play an essential role in iterative communication. Let us now attempt to extend the observation into the frameworks of spoken dialogue systems, in viewpoints of input and output aspects, speech recognition and speech synthesis provide the main acoustic interfaces between users and dialogue management. However, the semantic extraction and generating of natural language processing plays more essential roles for human machine interactions. As shown in Figure 1, a spoken dialogue system is composed of three components: speech recognition and natural language processing, dialogue management, and response generating and text to speech. Actually, we should now look more carefully into the results obtained in speech recognition and natural language processing. Since the accuracy of speech recognition is not near to perfect, it will cause the natural language misunderstanding.
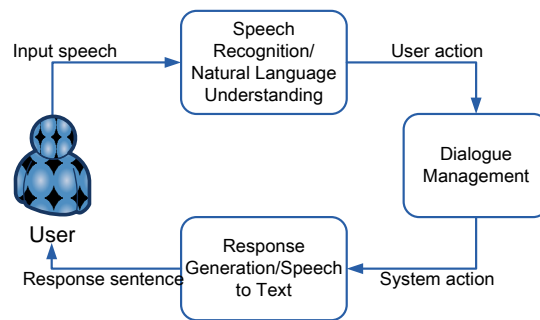

Figure 1. Overview of spoken dialogue systems.

That is to say, it is hard for conversational dialogue systems to fill the values in the semantic slots perfectly. Dialogue management with only limited information about semantic slots is unable to decide the correct system actions. Based on the incorrect system actions, the indisposed response generated by the system will make the system unfriendly. In the latest decades, some research efforts on response generating are invested for improving the quality of conversational dialogue systems.

The goal of natural language generating is aimed at obtaining the sentence that is suitable to understand for users. Herein, there are three categories of sentence generating: template-based, rule-based and statistics-based approaches. Template-based approaches were first developed for generating the sentence in natural language processing (Lemon 2011; Bauer 2009; Zhan 2010). Fang et

al. used the mixed template for constructing the declarative sentences. The declarative sentences were further converted into interrogative sentences by changing the word order and verb forms (Fang et al. 2006). Compared to template-based approaches, rule-based approaches were designed to provide more flexibility and desired sentences (Reiter and Dale 2000; Reiter 1996). Three main modules are included here in rule-based approaches: content determination and text planning, sentence planning and surface realization. Content determination and text planning are designed to decide the on the information communicated in a generated text. Dialogue management plays essential roles in content determination in conversational dialogue systems. According to the results of content determination, sentence planning selects and organizes the propositions, events and states to generating the sentence which usually contain one issue. Its main function is to select and adapt linguistic forms so that the generated sentences are suitable for the local context in which they are to appear. Surface realization is designed to create a syntactic representation in the form of a generated sentence given the semantic concepts. The overall flow of the rule-based approach is illustrated in Figure 2.

The third category about sentence generation is statistics-based approaches. The statistics-based approaches are also called as trainable generation. Instead of the predefined templates and rules, trainable approaches build the models and the corresponding parameters using the gathered corpus. Branavan et al. used the reinforcement learning to predict causal relationship between content and event, the causal relationship was further adapted to derive the higher level content determination (Branavan et al. 2012). Walker et al. proposed trainable sentence planner, DSyntS, to enhance the variousness of generated sentences (Walker et al. 2001; Walker et al. 2002; Melcuk 1988). Stent et al. added the rhetorical knowledge into the sentence planner to form the system, SPaRKy (Stent et al. 2004).
Since the response generating plays an essential role in conversational dialogue systems, the excellent response generating will cause the system more practical. For avoiding the limitation of human labeling, this paper invests a statistical approach based on hierarchical semantic structure with partially observable Markov decision process (POMDP) re-ranking strategy to produce the more spontaneous speaking style output.
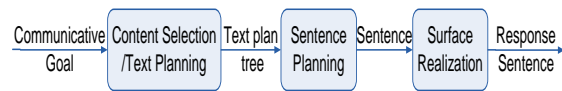

Figure 2. The flow chart of rule-based sentence generation that composed three modules.

The rest of this paper is organized as follows. Section 2 describes the proposed method and the related important modules in conversational dialogue system. Next, Section 3 presents the detail description about the proposed method especially in hierarchical semantic structure and partial observable Markov decision process (POMDP) re-ranking strategy. Experiments to evaluate the proposed approach and the related discussion are presented in Section 4. Concluding remarks are finally made in Section 5.

2 The proposed system framework

With this investment, we want to generate the human-like response of conversational dialogue systems by statistical approach. Herein, the system framework is divided into two phases: training and generation as described in Section 2.1 and 2.2 respectively.

2.1 Training phase

Human to human conversations are gathered as the training corpus. The utterance obtained from the corpus is first as the seed sentence for equivalent utterance expansion. The keywords and corresponding sentence pattern are extracted from the utterance and are fed into a web search engine by API to gather the expanding utterances. Basically, the utterance with the near-meaning will be recalled as the candidates for further processing. For each candidate utterance, the semantic objects embedded here will be extracted as the values and filled into semantic slots defined in conceptual graph. Due to the hypernym (superordinate) plays an essential role in information retrieval and dialogue management, eHownet, developed by academia Sinica, Taiwan, is used as the knowledge based to provide relative information. The expanded utterances and the original one are all ranked. The ranking can be composed of system pre-ranking and human adjustment. According the results of ranking and the status of concept graph, the POMDP parameters are estimated for generation phase. One of the most important issues is the reward function training.
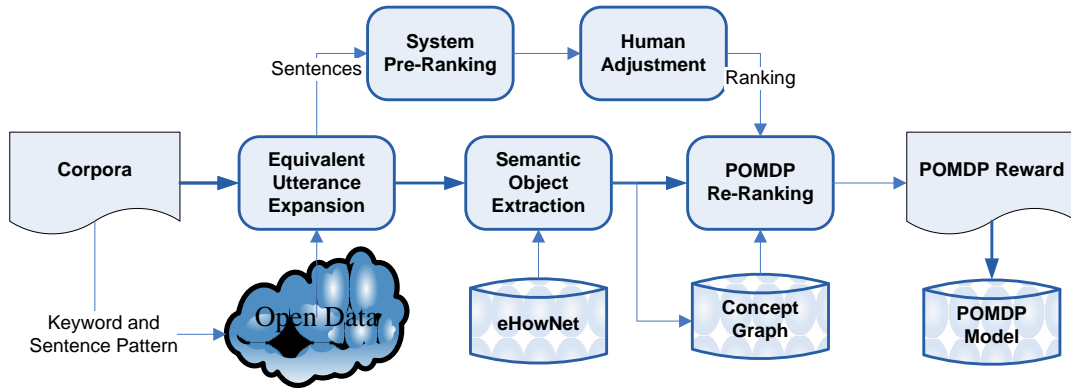
Figure 3. The system flow of training phase in proposed approach

2.2 Generation phase

As described in previous sections, response generation can be divided into content determination, sentence planning and surface realization. Herein, we combined the sentence planning and surface realization. That is to say, the proposed generation phase consists of two models: content selection and sentence planning and surface realization.

Semantic objects are first extracted from user's input utterance and fed into corresponding semantic slots defined in concept graph. According to the absent information in concept graph, the proposed method will decide the response contents. The results of the response content decision will be further fed into sentence planning and equivalent utterance expansion. Sentence planning will form the basic word set and sentence patterns for surface realization. The equivalent utterance expansion will gather the relevant sentence from the internet and the select candidate sentence with potential for surface realization. Open data contain many sentences is fit for the response generation. After the surface realization, the acceptable spoken utterances are obtained. A POMDP reward function is used to rank the spoken utterances according the POMDP model obtained in the training phase. According to the result of POMDP re-ranking, the generated response is obtained.
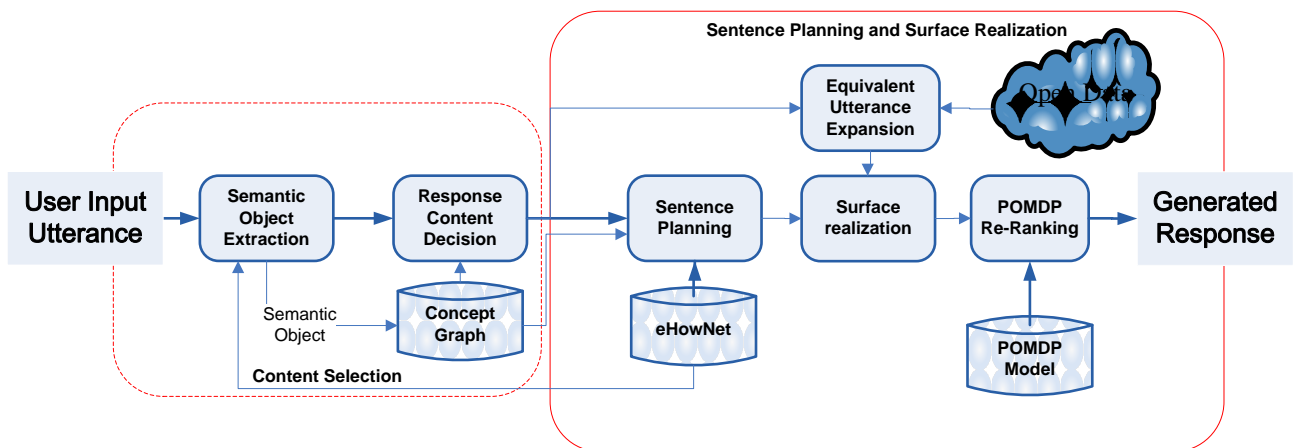

Figure 4. The system flow of generation phase in proposed approach

3 The proposed hierarchical semantic structure with POMDP re-ranking

This investment proposed a various response generation method for conversational dialogue systems. Actually, it is important for the practicability of spoken conversation interface how to increase adoption in response generation. Context determination decides which meaning would be carried in response. Realization strategies for dialogue responses depend on communicative confidence levels and interaction management goals. However, the only one value kept for each semantic slot in traditional dialogue management makes some information lost resulting in the persecution of users. Enrich the number of the response utterances and their sentence will increase users' delight. Some sentence patterns and linguistics material will enrich the natural language generation significantly. In fact, these issues connect to the conversational dialogue system practice or not. In this section, we may consider the subject under the following heads: conceptual graph and hierarchical semantic structure and POMDP re-ranking strategy. It seems reasonable to consider response generation through two types of organization.

3.1 Hierarchical semantic structure

Thinking ways about speech is very essential because it provides insight into the utility of human communication. In other words, that human uses communication as a tool to further their own ends not merely in human to human communication but also in human machine interactions. Ignoring semantic relations among semantic objects causes the exactly extracting the values from spoken utterance hard in traditional spoken dialogue systems. Conceptual graph is adopted as the knowledge representation for describing the semantic relations in this paper. Compared to the semantic slot with only one value, this investment proposed a hierarchical semantic structure to store the potential values for corresponding semantic slot by a linked list.

Concept graph is one of formalisms for knowledge representation. Herein, we used them to represent the conceptual schemas used in conversational dialogue systems. An example of conceptual graph for speech act, 訂票 (booking ticket for the train), is illustrated in Figure 5. Speech act "訂票 (booking ticket for the train)" is the centre in the conceptual graph. Some non-

terminal nodes denote the concepts. Here, the item concept refers to the relationship between certain symbols and signifiers such as semantic objects. Semantic objects are regarded as the possible values for some semantic slots. For example, the non-terminal concept "旅程 (journey)", in the top left of Figure 5, is composed of two semantic slots "起點(departure)" and "終點(destination)" and their relation. When talking about the "終點(destination)," users say "台北 (Taipei)" and the dialogue management imagines "台北(Taipei)" is the potential value of the semantic slot "終點(destination)". Then you have just used a signifier (the word "台北 (Taipei)") to indicate a semantic slot "終點 (destination). A large part of semantics is language, which uses words to symbolize things. Signifiers may be ideas, nouns, places, objects or feelings corresponding to semantic slots defined in conversational dialogue system.
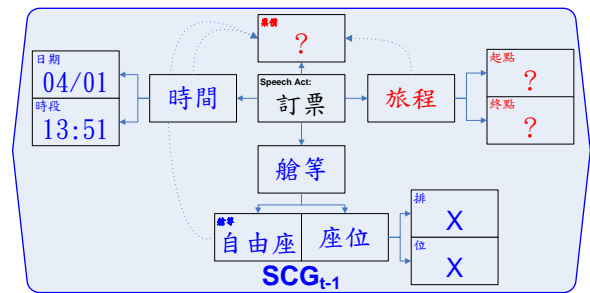

Figure 5. An example of conceptual graph for the speech act "booking train tickets"

This paper proposed a linked list based semantic slot for keeping more than one possible values for each semantic slot. In other word, the proposed system will extract all possible semantic objects from discourse. This structure provides more flexible combinations for the response generation. According to these combinations, the response utterances will be re-ranked by POMDP strategy.

3.2 POMDP re-ranking strategy

Response utterances can be further divided into two source categories. The first category comes from sentence planning and surface realization by the system according to the concept graph and eHownet. The second category comes from internet data from the equivalent utterance expansion. Finally, POMDP is adopted as the re-ranking process to select the near optimal utterance to be the generated response.

Due to the conversational dialogue is an interactive process. Considering the current user utterance and predicting the next user utterance, generated responses are re-ranked by POMDP. The POMDP adopted as the response generation operates as follows. At each time-step, that is to say, one turn in dialogue the state on the discourse record is in some unobserved state s_t. Due to the values in semantic slot is not exactly sure, the concept graph is partially observable. Since s_t is not known exactly, a distribution over possible states called a belief state b_t is maintained where $b_t(s_t)$ indicates the probability of being in a particular state s_t. Based on *both*, the dialogue management selects an action a_t, generating the response to user, receives a reward Rt, and transitions to next unobserved state, the corresponding concept graph at $t+1$. Here, syntactic and semantic scores are used to calculate the reward. We call it as s_{t+1}, where s_{t+1} depends only on s_t and a_t. The dialogue system then receives an observation o_{t+1}, which is dependent on s_{t+1} and a_t. Herein, o_{t+1} means the speech act and the semantic object carried in user utterance at turn $t+1$.This process is represented graphically as an influence diagram in Figure 6.
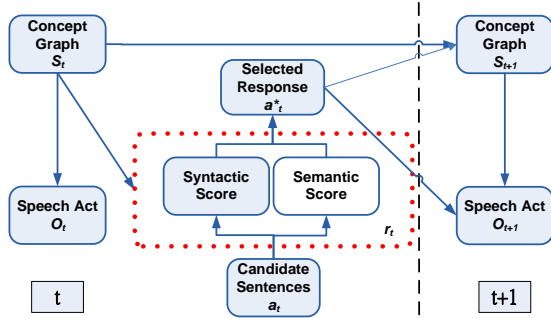

Figure 6. Illustration about the proposed POMDP re-ranking strategy

Given an existing belief state b_t, the last system action a_t, and a new observation o_{t+1}, the new updated belief state b_{t+1} is given by

$$b_{t+1}(s_{t+1}) = \eta P(o_{t+1}|b_{t+1}, a_t) \sum_{s_t} P(s_{t+1}|s_t, a_t) b(s_t) \quad (1)$$

Where η denotes the normalization factor. It can be calculated as equation (2).

$$\eta = P(o_{t+1}|b_t, a_t) \quad (2)$$

The standard optimizing process of POMDP is used for estimating of the action policy.

4   Experimental results

For evaluating the performance of the proposed method, a corpus contains 243 dialogues with 7,445 sentences are used for training. A conversational dialogue system using mandarin in travel domain is developed for assessing dynamically. Ignore of the error resulted from speech recognition engine, five dialogues for each individual to obtain the statistics.

To evaluate the performance of the proposed method, the subjective evaluation, the mean opinion score (MOS), is used to measure the qualities of the voice transformation approaches. The opinion score was r is expressed in one number, from 1 to 5 (1 means bad and 5 denotes excellent). MOS is quite subjective, as it is based figures that results from what is perceived by people during tests. Twenty two individuals are asked to be the users using the conversation dialogue system developed in travel domain in this paper. Five dialogues with MOS scores for each individual during two weeks are recorded for further evaluation. Another system based on template response generation is also developed for comparison (Lee et al. 2009). Four aspects, variety, naturalness, suitability, intelligibility, are used to appraise the response systems. The experimental results are shown in Figure 7.
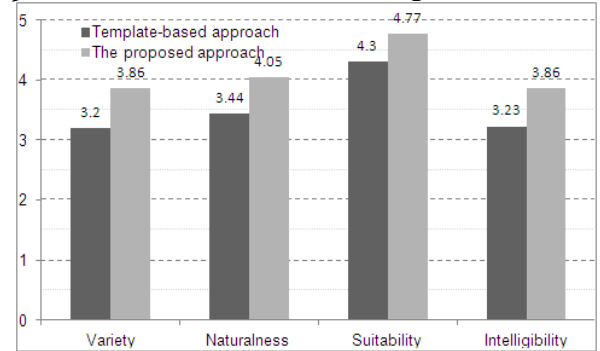

Figure 7. Evaluation results about template-based and the proposed approaches

According to the results, the suitability of these two approaches is high enough. Due to either template approach or the proposed approach are both able to provide the right information for users. The proposed approach outperforms the template approach significantly in variety and naturalness. These results show the concept graph and POMDP re-ranking ability to obtain improvement.

5   Conclusions

A new approach to generate responses for conversational dialogue systems has been presented in this study. The algorithm is based on the idea of hierarchical semantic structure of concept graph and POMDP re-ranking strategy. Linked list based semantic slot is applied to extract the values of semantic objects from input utterance. The two sentence generation sources: natural language generation and gathering open data from the internet are used to keep the variety of generated responses. POMDP re-ranking further selects near optimal utterance considering of the status of concept graph. The experimental results verified that the proposed approach results in keeping more information in concept graph and various responses generated especially in variation and naturalness. The future works include applying more precise estimation for POMDP.

Acknowledgments

References

Lemon, O. 2011. Learning What to Say and How to Say it: Joint Optimization of Spoken Dialogue Management and Natural Language Generation. *Journal Computer Speech and Language*. 25(2): 210-221.

Bauer, D. 2009. Statistical Natural Language Generation as Planning. *Master's thesis of Department of Computational Linguistics, Saarland University*, Saarbrucken, Germany.

Zhan, W.D. 2010. A Brief Introduction to Natural Language Understanding and Generation. *Terminology Standardization & Information Technology* 4.

Fang, Z.-W., Du, L.-M., Yu, S.-Y. 2006. A Chinese Sentence Generator Based on Hybrid-Template for Spoken Dialogue System. *Journal of the Graduate School of the Chinese Academy of Sciences*, 23(1): 23-30.

Reiter, E., and Dale, R. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

Reiter, E. 1996. *Building Natural-Language Generation Systems*. In Alison Cawsey, ed., Proceedings of the AI and Patient Education Workshop, Glasgow, GIST Technical Report G95.3, Department of Computing Science, University of Glasgow.

Branavan, S.R.K., Kushman, N., Lei, T., Barzilay, R. 2012. Learning High-Level Planning from Text. *Proceeding ACL '12 Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*: Vol. 1: 126-135.

Walker, M. A. Rambow, O., Rogati, M. 2001. SPoT: A Trainable Sentence Planner. *Proceedings of the 2nd Annual Meeting of the North American Chapter of the Associataion for Computational Linguistics.*

Walker, M. A. O. Rambow, O., Rogati, M. 2002. Training a Sentence Planner for Spoken Dialogue using Boosting. *Computer Speech and Language*: Special Issue on Spoken Language Generation, 2002.

Melcuk, I. A. 1988. *Dependency Syntax: Theory and Practice*, SUNY, Albany, New York.

Stent, A., Prasad, R., and Walker, M. 2004. Trainable Sentence Planning for Complex Information Presentation in Spoken Dialog Systems. *Proceedings of ACL*.

Lee, C., Jung, S., Kim, S., Lee, G. G. 2009. Example-based Dialog Modeling for Practical Multi-domain Dialog System. *Speech Communication* 51(5): 466−484.