OPEN-DOMAIN PERSONALIZED DIALOG SYSTEM USING USER-INTERESTED TOPICS IN SYSTEM RESPONSES

Jeesoo Bang, Sangdo Han, Kyusong Lee and Gary Geunbae Lee

Pohang University of Science and Technology
Department of Computer Science and Engineering
Pohang, South Korea

ABSTRACT

We built a personalized example-based dialog system that constructs its responses by considering entities that the user has uttered, and topics in which the user has expressed interest. The system analyzes user input utterances, then uses DBpedia and Freebase to extract relevant entities and topics. The extracted entities and topics are stored in personal knowledge memory and are used when the system selects responses from the example database and generates responses. We conducted a human experiment in which evaluators rated dialog systems based on subjective metrics. The proposed dialog system that uses topics that are of interest to the user achieved higher evaluation scores for both personalization and satisfaction than the baseline systems. These results demonstrate that the use of topics in the system response provides a sense that the system pays attention to the user's utterances; as a consequence the user has a satisfactory dialog experience.

Index Terms— Chat system, chatbot, open-domain dialog, knowledge base, topic-based dialog

1. INTRODUCTION

With the rapid proliferation of smart phones, the number of people who use spoken dialog systems has also increased. People use spoken dialog systems to make the phone complete some tasks, such as checking weather, setting alarms, and play music. This kind of dialog system that assists in some tasks is called a task-oriented dialog system. The user utterances of the task-oriented dialog system are often standardized and have clear objectives. However, people sometimes try to interact with a dialog system by uttering utterances that do not have objectives. Open-domain dialog systems can respond to these types of user utterance; here we address an open-domain dialog system as a chat-oriented dialog system.

An open-domain dialog system communicates with users for the purpose of entertainment by having conversation with users like humans do [1, 2, 3]. We aim to build a dialog system that talks like a human, specifically, one that talks like a friend of the user, and that can build a rapport with each user. Heintzman et al. [4] define rapport as a "communication characterized by warmth, enthusiasm, and interest". Research on rapport-building has determined that effective rapport building entails referring to the user by name, remembering the user's interests, and constantly updating the interests [5, 6, 7].

Personalization has studied for various chat-oriented dialog systems. ALICE [2] uses pattern rules or instructions to extract user information (i.e. name, hobby) from the user input, and utilizes this information to generate responses. Kim et al.'s personalized dialog system uses user-uttered noun phrases and triple pattern match for personalization [8]. However, extracting personal knowledge with rules, noun phrases or triple patterns is too restricted for use in generating user-related system responses, because the system response can use user-uttered entities that have been stored in the system, but cannot use topics in which the user has expressed interest, hereafter user-interested topics, even if they have been mentioned several times. Existing personalized dialog systems do not use topics in the system responses, so the responses are restricted to user-uttered entities.

We propose a system that extracts both an entity and its corresponding topics from the user input, and generates responses that refer to those topics. We used DBpedia[1] and Freebase[2] to extract entities and topics from the user input and from sentences in an example database. Our proposed system is different from previous personalized dialog systems in: 1) remembering and updating user-interested topics, 2) using user-interested topics when selecting the system response and in entity replacement, and 3) uses these topics to generate the system response when no dialog examples exist in the example database.

2. RELATED WORK

ALICE [2] is a rule-based chatbot based on Artificial Intelligence Markup Language (AIML). It has more than 40,000 categories, where each category combines stimulus and re-

[1] http://wiki.dbpedia.org/
[2] http://www.freebase.com/

978-1-4799-7291-3/15/$31.00 ©2015 IEEE

sponse, called the "pattern" and "template" respectively. ALICE uses pattern rules or instructions to extract user information (i.e. name, hobby) from the user input, and utilizes it when generating responses. However, developing a rule-based system entails tremendous human labor and expert knowledge to make rules that can cover all kinds of user utterance. Therefore, we used an example-based dialog model (EBDM) [9] to build and maintain the system without much human labor.

Kim et al.[8] proposed a personalized dialog system that uses triple patterns and noun phrases. The system extracts triples and noun phrases from the user input utterance, then stores the information. The noun phrases are used to select a response from response candidates; to rate the candidates, the system give a higher score to a candidate that includes a noun phrase that the user has uttered than to other candidates. However, this use of only noun phrases has the limitation that it cannot capture a topic that different entities share. An entity that has been uttered many times is considered more relevant to the user than is a coherent topic, even if the user uttered different entities for a single topic. For example, if a user uttered different entities such as football clubs, football teams, and football players, the system should consider that the user is interested in football. Our approach uses topics to enlarge the scope of system responses to consider user interest.

In Kim et al.'s work, a user-related triple is specially stored in the memory by matching manually-generated triple patterns. The stored user-related triple is used in the system response if a triple extracted from a system response candidate matches the user-related triple. For example, the system stores the triple [I][like][cheesecake] from the user utterance "I like cheesecake", and uses this triple for replacing the entity in the triple [You][like][carbonara] extracted from a system response "You like carbonara, yummy!" using person-change rule (i.e. I-You). The replaced system response will be "You like cheesecake, yummy!" and the user gets the feeling that the system knows about him/her. However, a triple [You][like][tennis] extracted from a system response "You like tennis, let's play" is also replaced with the triple [I][like][cheesecake], and the entity replacement will result in the response "You like cheesecake, let's play", which is awkward. This problem can be solved by using named entity types rather than instances in entity replacement; we used this strategy in this work.

Previous example-based dialog systems [10, 11] search similar example pairs by relaxing the condition of searching; this relaxation can result in selection of inappropriate system responses if a similar example is not included in the example database. In this case to give an adequate system response to the user, and also a response that matches user interests, we generated responses by considering user-interested topics.

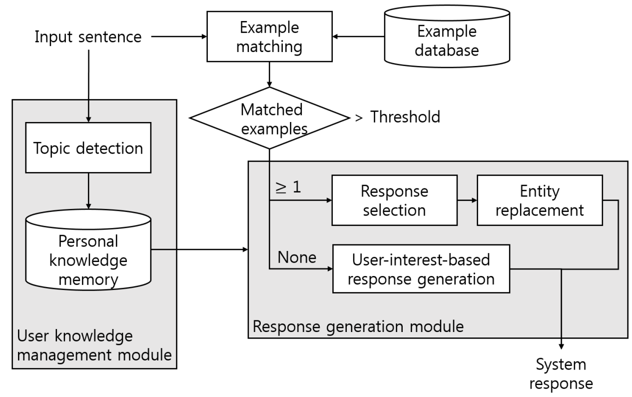

Fig. 1. Architecture of the proposed example-based personalized dialog system which uses user-interested topics when it generates system responses

3. EXAMPLE-BASED PERSONALIZED DIALOG SYSTEM

We propose an example-based personalized dialog system which uses user-interested topics in the system response. (Fig. 1)

3.1. User knowledge management module

The user knowledge management module consists of a topic-detection module that extracts an entity and its corresponding topic from the user input utterance; and a personal knowledge memory that stores user-uttered entities and topics.

The topic-detection module extracts an entity and its corresponding topic from the user input utterance. We used a knowledge base for entity and topic detection: we used DBpedia Spotlight [12] to extract named entities, and Freebase to detect entity types. Freebase may have a lot of types for an entity, among them we used the notable type that Freebase offers for each entity. The notable type has a hierarchical feature: general type and specific type. We used both types as a topic and entity type. For example, consider a user-uttered sentence "I want to see Messi in person": first, the entity "Lionel Messi" would be extracted by DBpedia Spotlight, and then "Lionel Messi" is searched in Freebase which has a notable type "soccer/football_player". We used the general type "soccer" as a topic, and the specific type "football_player" as a type.

The detected entities and topics are stored in the personal knowledge memory. Personal knowledge memory has two different memories: topic memory and entity memory. The topic memory stores topics and their frequencies which show how many times the entities that have the topics were uttered by a user. The entity memory stores user-uttered entities with their corresponding topics and types, and the frequencies of these utternces.

3.2. Example matching

Our personalized dialog system uses example-based dialog model (EBDM) [9] to generate system responses. EBDM uses data consists of query-response pairs, where the query is representative of the user input U_{DB} to the system, and the response is representative of the system response S_{DB}. Response retrieval for EBDM works by matching the user input utterance U_{Input} with U_{DB} in the example database, then returning the response that corresponds with the most-closely-matching query. The example matching process can be performed in various ways: dialog act and part-of-speech (POS)-tagged token based similarity [10], TF-IDF weighted cosine similarity [3], or recent sentence embedding [13].

Use of sentence embedding in an example-based dialog system achieved competitive evaluation result with respect to TF-IDF cosine similarity model and a superior result when the queries had out-of-vocabulary words [13]. We also used sentence embedding during the example-matching phase. We used deep structured semantic model [3] (DSSM) [14] for sentence embedding to match examples. We calculated cosine similarity

$$cos_{sim}(U_{Input}, U_{DB}) = \frac{U_{Input} \cdot U_{DB}}{\|U_{Input}\|\|U_{DB}\|} \quad (1)$$

between vectors U_{Input} and U_{DB}, and if the similarity exceeded a threshold we added the corresponding S_{DB} to the set of response candidates. We heuristically set the threshold to 0.6 to determine whether a user input utterance matched the example pair in the database. When we constructed example database, we pre-calculated the sentence embedding vector of each user-input sentence U_{DB} and the pre-detected topics of each system response S_{DB}.

3.3. Response generation module

3.3.1. Response selection and entity replacement

Multiple example pairs can be selected for a user input utterance; in that situation, one of them must be chosen. Previous work selected the response by calculating the similarity between U_{Input} and U_{DB}, then choosing S_{DB} for U_{DB} that had the highest similarity. The utterance similarity is calculated using lexico-semantic similarity and the discourse history similarity [9]. We modified the example score calculation module to select from the multiple system response candidates an example that is most relevant to the user. Each system response in the database has topics extracted beforehand, and each topic has a score $p(topic)$ calculated by user-uttered frequencies.

[3]Available at http://research.microsoft.com/en-us/downloads/731572aa-98e4-4c50-b99d-ae3f0c9562b9/

U	I like football.
S	Football game!
U	Yes, especially Primera Liga.
S*	Tell me more about [sports league].
U	I like Real Madrid most in Primera Liga.
S*	You are talking about [football team], right?
U	Yes.
U	I am bored. Let's talk about something.
S	What would you like to talk about.
U	Artists?
S*	Let's talk about [Real Madrid].
U	Well, that's also good topic to talk about.

Fig. 2. A realistic dialog example using user-interested topics in system responses. U: user input utterance, S: system response, S*: system response that uses user-interested topics

$$p(topic) = \frac{count(topic)}{\sum_{t \in TopicMemory} count(t)} \quad (2)$$

The system uses topic score to calculate the example score for each response candidate

$$ExampleScore = \alpha[cos_{sim}(U_{Input}, U_{DB})] + (1-\alpha)p(topic_{S_{DB}}) \quad (3)$$

where $topic_{S_{DB}}$ denotes the topic that is considered by the corresponding S_{DB} for U_{DB}.

The system selects the system response that has the highest example score among the candidates. If several responses have the same example score, one of them is selected randomly. We arbitrarily set α = 0.5.

With the selected system response, we conducted entity replacement: (1) check whether the selected system response has an entity (system response entity), then (2) check whether the entity memory includes an entity that is of the same type as the system response entity. If several entities in the entity memory have same type, the most frequently-uttered entity by the user is selected to replace the system response entity.

3.3.2. User-interest-based response generation

The example-based dialog system may not find similar example pairs in the database; if this happens, the system tries to find one by relaxing the search conditions. The relaxation can be done using correlation-based relaxation [11] or POS-weight-based relaxation [10]. In our system, we did not try to relax the condition of example matching by lowering the threshold; instead, we used the personal knowledge memory to guide generation of responses in which the user may be interested. (Fig. 2)

We generate the response based on the user input utterance, if the user uttered an entity. We used user uttered topic

Table 1. Evaluation result of the non-personalized system, triple-pattern-matching system and the proposed topic-using system

Statement	Dialog System					
	Non-person		Triple-match		User-topic	
	mean	s.d.	mean	s.d.	mean	s.d.
1. The system seems to know me well	2.18	0.57	2.36	0.48	2.82	0.96
2. The system knows me well over time	2.63	0.88	3.00	0.95	3.27	0.86
3. The conversation with the system was interesting	3.00	0.85	2.73	0.96	3.36	0.64
4. The conversation with the system was satisfactory	2.64	0.64	2.36	0.64	3.27	0.45
5. The system talks about my interest	2.73	0.96	2.82	0.94	3.73	0.86
6. The system seems to be interested in me	3.09	0.79	2.82	1.03	4.18	0.57
Overall	2.71	0.85	2.68	0.89	3.44	0.86

entity and its type to generate a system response that continues the conversation and does not degrade the user dialog experience by introducing a new dialog topic. The response generation process uses templates to effectively exploit user-uttered entities and topics. We adopted the rapport-building phrases of the interview from [15] to conduct the rapport-building templates. For example, if the user input is "I like Messi" and the database contains no similar example pair, then the response is generated using the user-uttered entity "Messi" with its type "soccer/football_player" like "Tell me more about football player", to encourage the user to talk more and not to feel that the dialog has been disrupted.

However, if the user input utterance does not have an entity, the system suggests a new topic from its topic memory; the purpose is to interest the user, and to encourage him or her to continue talking. The system chooses the new topic by selecting one randomly from the top N topics in which the user has expressed interest. For example, if the user says "I feel bored" and there is no similar example pairs in the database, the system generates responses using the user-interested topics, that user already has uttered several times. The system uses both types and entities, and therefore can generate "Let's talk about Messi" or "Well, do you want to talk about football player?" if the user has uttered several times about "Messi".

4. EXPERIMENTAL SETUP

We used MovieDic [16] and the scripts from the TV shows (Friends, Modern Family) to generate an example database for the dialog system. We refined the data [17], and got 8,447 example pairs for this purpose. We compared our dialog system that uses user-interested topics (User-topic) to two existing systems to evaluate the effect of the proposed methods on the user's dialog satisfaction. The other systems were one that does not use personalization methods (Non-person), and Kim et al.'s personalized dialog system that uses triple-match methods [8] (Triple-match).

To evaluate the systems, we designed an experiment for 11 evaluators to utilize the systems. The evaluators used each system for 15 min, then completed a questionnaire that evaluated each dialog system on a scale of 1 - 5, with positive adjectives anchoring the high end and negative adjectives anchoring the low end. We requested the evaluators mainly to talk about themselves, but the form of dialog was free.

5. RESULTS AND DISCUSSION

The users evaluated the proposed personalized system as more personalized and also more satisfactory than the two other baseline systems (Table 1).

Questions 1 and 2 evaluated how the personal knowledge is managed by each system. The Non-person system has no personal knowledge management methods, and therefore had the lowest evaluation scores. The Triple-match system selected a system response that had a noun phrase that the user has uttered, so this response is more relevant to the user than is a random response. Therefore, this system was given a higher score than Non-person system in personal knowledge management evaluation. The proposed system that remembers and updates user-interested topics and that uses the information when selecting the system response, was given the highest evaluation score. The score of the question 2 clearly shows that the proposed system can update the list of user-interested topics, and that its response change over time. The evaluation scores of questions 1 and 2 show that the users considered that User-topic system knows them well, and knows them increasingly well as time elapsed.

The questions 3 and 4 evaluated the use of user-interested topics in response selection and entity replacement of system responses. The Triple-match system was given lower evaluation scores than Non-person system. This low score shows that the entity replacement of Triple-match system sometimes harmed the system response. Entity replacement by Triple-match system does not consider the entity type. In contrast, User-topic system was given higher scores than Non-person system, even though it also uses entity replacement. This result shows that the using types in entity replacement does not harm the dialog experience, and that it generates system responses that the user prefers.

The questions 5 and 6 evaluated quality of responses generation based on user-interested topics. Non-person and Triple-match systems had similar evaluation scores around 2.8–3.0. Both systems relax the search condition, when no matched examples are found, whereas User-topic system generates a system response by considering user interests in the same situation. These evaluation scores were much higher for User-interest system than for other systems, especially the score of question 6. We consider that this high score is caused by the response generation, because most response generation templates were requests to continue talking about the user's interests.

From these evaluations, we conclude that the proposed system (User-topic) gives more satisfaction to users than did the other systems; i.e., that considering user-interests when generating system responses can increse a user's satisfaction with an open-domain dialog system.

6. CONCLUSIONS

This study reports a personalized example-based dialog system that builds a rapport with users by considering their interests when it chooses a response. The system stores user-uttered entities and their topics and types, and user-interested topics, then uses this information to select a response among the response candidates with entity replacement by entity type, and to generate a topic-based response. The proposed system gave more dialog satisfaction to users than did two baseline systems. These results will be useful in development of personalized dialog systems and multi-turn open dialog systems.

7. ACKNOWLEDGE

This work was supported by the ICT R&D program of MSIP/IITP. [R0126-15-1117, Core technology development of the spontaneous speech dialogue processing for the language learning] and by the ICT R&D program of MSIP/IITP [R0101-15-0176, Development of Core Technology for Human-like Selftaught Learning based on a Symbolic Approach]

8. REFERENCES

[1] Joseph Weizenbaum, "Eliza–a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[2] Richard S Wallace, *The anatomy of ALICE*, Springer, 2009.

[3] Rafael E Banchs and Haizhou Li, "Iris: a chat-oriented dialogue system based on the vector space model," in *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 2012, pp. 37–42.

[4] Mark Heintzman, Dale G Leathers, Roxanne L Parrott, and Adrian Bennett Cairns, "Nonverbal rapport-building behaviors' effects on perceptions of a supervisor," *Management Communication Quarterly*, vol. 7, no. 2, pp. 181–208, 1993.

[5] Linda L Wolcott, "Distant, but not distanced," *TechTrends*, vol. 41, no. 5, pp. 23–27, 1996.

[6] William Buskist and Bryan K Saville, "Creating positive emotional contexts for enhancing teaching and learning," *APS Observer*, vol. 19, pp. 12–13, 2001.

[7] Ernest Mendes, "What empathy can do," *Educational Leadership*, vol. 61, no. 1, pp. 56–59, 2003.

[8] Yonghee Kim, Jeesoo Bang, Junhwi Choi, Seonghan Ryu, Sangjun Koo, and Gary Geunbae Lee, "Acquisition and use of long-term memory for personalized dialog systems," in *Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*, pp. 78–87. Springer, 2014.

[9] Cheongjae Lee, Sangkeun Jung, Seokhwan Kim, and Gary Geunbae Lee, "Example-based dialog modeling for practical multi-domain dialog system," *Speech Communication*, vol. 51, no. 5, pp. 466–484, 2009.

[10] Jeesoo Bang, Hyungjong Noh, Yonghee Kim, and Gary Geunbae Lee, "Example-based chat-oriented dialogue system with personalized long-term memory," in *Big Data and Smart Computing (BigComp), 2015 International Conference on*. IEEE, 2015, pp. 238–243.

[11] Cheongjae Lee, Sungjin Lee, Sangkeun Jung, Kyungduk Kim, Donghyeon Lee, and Gary Geunbae Lee, "Correlation-based query relaxation for example-based dialog modeling," in *Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on*. IEEE, 2009, pp. 474–478.

[12] Pablo N Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer, "Dbpedia spotlight: shedding light on the web of documents," in *Proceedings of the 7th International Conference on Semantic Systems*. ACM, 2011, pp. 1–8.

[13] Lasguido Nio, Sakriani Sakti, Graham Neubig, Tomoki Toda, and Satoshi Nakamura, "Improving the robustness of example-based dialog retrieval using recursive neural network paraphrase identification," in *Spoken Language Technology Workshop (SLT), 2014 IEEE*. IEEE, 2014, pp. 306–311.

[14] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck, "Learning deep structured semantic models for web search using clickthrough data," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 2013, pp. 2333–2338.

[15] Kim P Roberts, Michael E Lamb, and Kathleen J Sternberg, "The effects of rapport-building style on children's reports of a staged event," *Applied Cognitive Psychology*, vol. 18, no. 2, pp. 189–202, 2004.

[16] Rafael E Banchs, "Movie-dic: a movie dialogue corpus for research and development," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*. Association for Computational Linguistics, 2012, pp. 203–207.

[17] Nio Lasguido, Sakriani Sakti, Graham Neubig, Toda Tomoki, and Satoshi Nakamura, "Utilizing human-to-human conversation examples for a multi domain chat-oriented dialog system," *IEICE TRANSACTIONS on Information and Systems*, vol. 97, no. 6, pp. 1497–1505, 2014.