Available online at www.sciencedirect.com
ScienceDirect
Procedia Technology 17 (2014) 122 – 129
Procedia Technology

Conference on Electronics, Telecommunications and Computers – CETC 2013

Proposal of open-ended dialog system based on topic maps

Ryosuke Ota, Masaomi Kimura*

Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan
Department of Information Science and Engineering, Shibaura Institute of Technology, Tokyo, Japan

Abstract

In this study, we proposed an open-ended dialog system based on topic maps. It is based on the fact that keywords and dialog answers are mapped to topics and their relationships are considered as associations. Based on this structure, we proposed a method to determine adequate answers that are related to questions asked by the users. We also employed the history of input questions to determine answers based on the context intended by the user. We conducted an experiment to evaluate our system and found that in most cases, the system generated relatively good answers. In the future, we will improve the topic map to contain a significant number of topics and answers, and to improve responses when topics change.

Peer-review under responsibility of ISEL – Instituto Superior de Engenharia de Lisboa, Lisbon, PORTUGAL.
Keywords: Topic Maps; Dialog System; Semantic Web

1. Introduction

Over the past years, there have been many studies on human-like chatting systems. Eliza [1] is a pioneering work on such systems. However, Eliza often repeats a user's input. This is because it does not understand the meaning of sentences [2]. It simply inserts keywords from the user's input into templates and returns them to the user [1]. To solve this problem, certain systems gather past dialogs and utilize them [3]. However, these systems do not consider topic changes that might occur.

In this study, we focus on the fact that topics often change during a dialog. Fig. 1 illustrates an example of such topic changes in daily conversation. In this example, the topic varies from a meal to a trip. Such time-sequential topic variations are referred to as topic changes.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: ma12019@shibaura-it.ac.jp

Peer-review under responsibility of ISEL – Instituto Superior de Engenharia de Lisboa, Lisbon, PORTUGAL.
doi:10.1016/j.protcy.2014.10.219

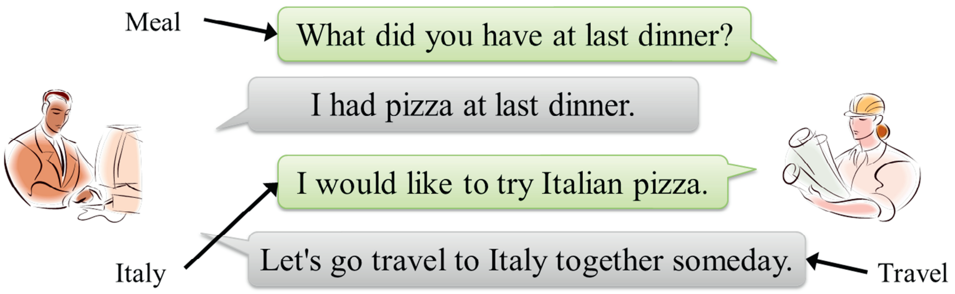

Fig. 1. Example of topic changes in daily conversation.

There are many studies that classify topic changes [4]. These studies classify topic changes as follows:

- A derivation-type change is a change to the topic that is directly connected to the last topic.
- A reappearance-type change is a change to the topic that formerly appeared in the dialog.
- A new-topic-type change is a change to a new topic that has not previously appeared in the dialog.

In addition, these studies reported that meaningful chats usually contain frequent derivation-type topic changes in human communication. Considering this, we propose a method that realizes derivation-type topic changes.

To realize this type of topic change, we utilize a data structure defined in the ISO 13250 topic maps standard [5]. This standard represents the relations of knowledge as a graph structure. By using topic maps, topic changes in a dialog are realized as the transition of the corresponding portion in the topic map's graph structure.

The remainder of this paper is organized as follows. In Section 2, we introduce topic maps. In Section 3, we propose the application of topic maps to realize a dialog system and introduce the utilization of dialog history. Section 4 includes our experiment, results, and a discussion of the experimental results. Finally, Section 5 presents our conclusions.

2. Topic Maps

Topic maps [5] is an international standard defining a data structure to represent relationships between subjects. A topic map consists of various components, such as topics, associations, and occurrences. A *topic* represents a concept (e.g., a person, an event) and an *association* represents a relationship between topics. An *occurrence* represents an information resource for topics (e.g., a photograph of the person, a web page of the event). Fig. 2 illustrates the structure of relationships between topics, associations, and occurrences.

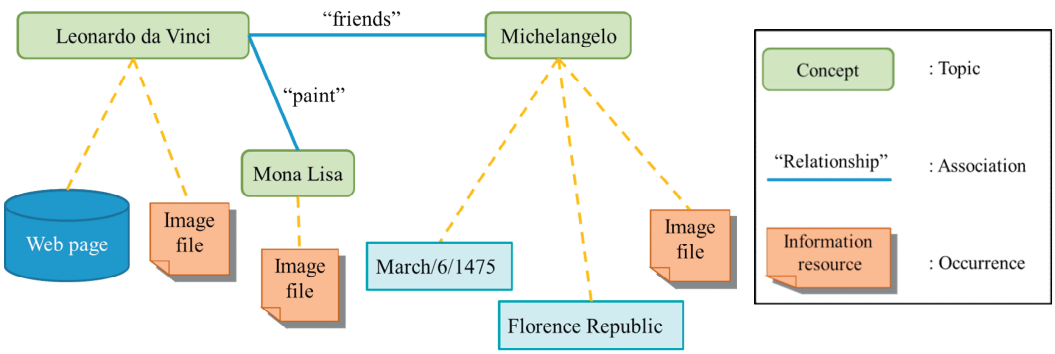

Fig. 2. Relationships between topics, associations, and occurrences.

In this figure, "Leonardo da Vinci" and "Mona Lisa" are topics. The association "friends" represents the friend relationship between Leonardo da Vinci and Michelangelo. The topic "Leonardo da Vinci" has its occurrences, a web page and an image file.

Note that Fig. 2 does not clarify whether "Leonardo da Vinci" painted "Mona Lisa" or vice versa. In fact, we need information pertaining to the roles played by the topics in the association. Topic maps use an *association role* to define such roles. We should also note that the original structure of a topic map is an undirected graph. However, because of association roles, we can introduce a direction to each association. In this study, this fact can help navigate through topics on the topic map structure.

3. Method

3.1. Application of topic maps to realize the dialog system

In this study, we regard a topic map as a knowledge base that stores the relationships between dialog topics and answers. In a dialog, a human expresses his knowledge as sentences. In sentences, nouns express concepts and verbs represent their relationships. Table 1 shows the relationships between parts of speech and their corresponding components of the topic map. In our dialog system, we utilize this correspondence to deduce an answer from the sentence input by the user. Occurrences do not appear in Table 1 since they do not have a corresponding part of speech. Later, we utilize them to represent answers output by the dialog system.

Table 1. Relationships between parts of speech and components of a topic map.

Parts of speech in Japanese language	Functions	Corresponding components of a topic map
Nouns, Pronouns	Representation of concepts	Topics
Verbs, Adjectives	Representation of relationships between concepts	Associations
Auxiliary verbs, Particles	Representation of an objective of an action indicated by a verb	Association roles

Fig. 3 illustrates an example of the topic maps proposed in this study. This topic map corresponds to the sentence "A dog dreads fireworks." The nouns "dog" and "fireworks" are topics, and the verb "dread" connects these nouns as an association. They have their own roles (e.g., subject, object), and both are connected by the answer, Answer001, that has an occurrence representing "Do not force him to get used to you." We should remind that a topic map essentially has the structure of an undirected graph. However, the roles, "subject" and "object", virtually define the directions of edges in the topic map. This enables to distinguish the input sentences, "dog dreads firework" and "firework dreads dog".

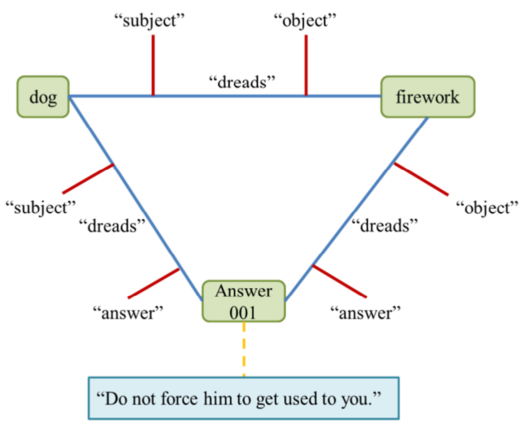

Fig. 3. Example of the proposed topic map

3.2. Derivation-type topic changes realized by a topic map

To realize derivation-type topic changes, we propose a neighborhood search in a topic map.

In our setting, neighbor topics in the topic map have relationships corresponding to their connecting verb. Let us identify a dialog topic to the corresponding topic in the topic map. Then, a topic connected to a previously appearing topic in the dialog should be valid as a potential topic used by the system to lead the dialog. Assuming a high clustering property discussed in the context of complex networks, next-nearest neighbor topics in our topic map are also expected to share some relationships. Therefore, in our method, we focus on the topics of not only the nearest but also of the next-nearest neighbor to the one appearing in the sentence input by the user. The procedure for this search is as follows:

1 The dialog system obtains topics (nouns) in an input sentence.

2 It searches nearest and next-nearest neighbor topics in the topic map.

3 The system returns an answer whose topic is connected to the pair of topics obtained in Step 2.

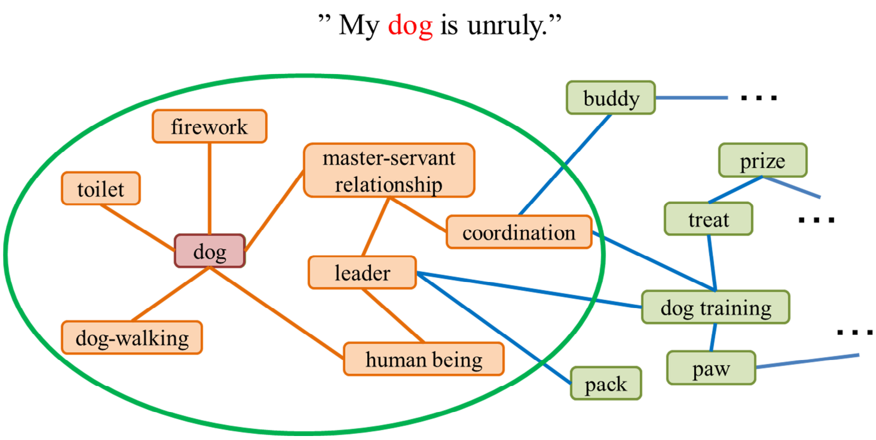

Fig. 4. Example of our neighborhood topic search. The sentence “My dog is unruly” is assumed to be input. The topic in this sentence is a noun “dog.” Its nearest/next-nearest neighbor topics are in the green circle.
Answer topics and their associations are ommited in this figure.

Fig. 4 illustrates our neighborhood search. Here the sentence “My dog is unruly” is assumed to be input by the user. The neighborhood search starts from the topic “dog.” The topics surrounded by the green circle are the nearest/next-nearest neighbor to the “dog.” The system randomly selects an answer topic connected to the obtained topics.

3.3. Utilization of dialog history

Not only the dialog system but also the user can make a derivation-type topic change. Therefore, we need to consider the relationship between the sentences that the user continually inputs. We extend the method in Section 3.2 to extract the topics commonly related to the latest dialog and its previous topics stored in the dialog history.

The procedure for the extended neighborhood search is as follows:

1 The system obtains topics in thelatest sentence input by the user.

2 It searches all topics that are nearest/next-nearest neighbors to topics obtained in Step 1.

3 It iterates Step 1 to Step 2 for each sentence previously input by the user in the session.

4 It returns an answer that is randomly selected from a common set of results obtained after Step 3.

Note that utiliztion of all input sentences causes extraction of unnecessary topics. This is because they may have little relationship with the current context of conversation. To prevent this, we limit the target dialog sentences to the latest sentence and its previous.

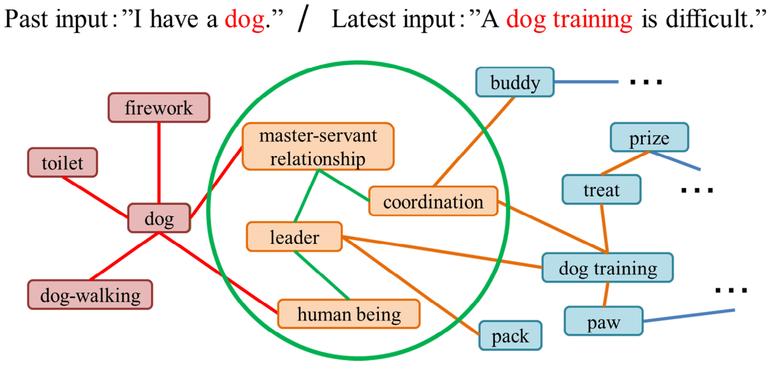

Fig. 5. Extended search taking account of dialog history.

Fig. 5 illustrates the above procedure. Here we assume that two sentences "I have a dog" and "Dog training is difficult" are input in this order. The search starts from the topics "dog" and "dog training." The green circle shows the resulting common set of topics. The system randomly selects an answer connected to a topic in the circle.

3.4. Implementation

We implemented the proposed topic change method as a search subsystem in our dialog system. Fig. 6 illustrates the system configuration diagram of the dialog system. The neighborhood search process searches answers by using the method in Section 3.3. "Save the dialog history process" saves the user's input and the answer given by the dialog system. In our system, we utilized CaboCha [6] as a dependency parser of sentences. We assumed that the sentences input by users are in Japanese. As a topic map database management system, we used TOME, which was previously developed by our project team [7].

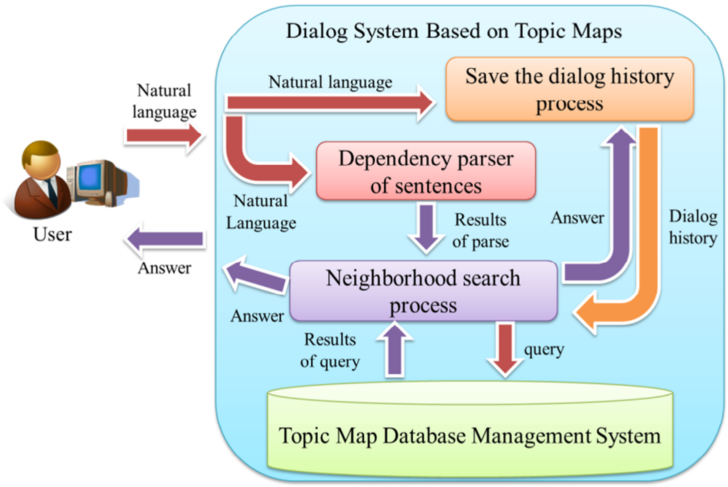

Fig. 6. System configuration diagram of the dialog system

4. Experiments

We conducted an experiment to evaluate the extent to which our system can generate dialog with derivation-type topic changes. Our topic map consisted of 311 topics (including 85 answers) and 265 associations related to dog training.

The subjects were eight university students, all between the ages of 21 and 25 years. We asked them to consult the system regarding dog training. We instructed the subjects that they could stop the dialog at any time. They produced a total of 172 statements.

User1 > My dog hates the walk.
System > Why do not you change the course of the walk?
User1 > My dog does not listen to me.
System > You should teach the master-servant relationship to your dog.
User1 > My dog does not give his paw to me.
System > Dog training needs the patience.

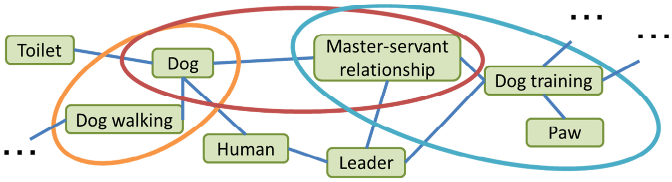

Fig. 7. Part of a derivation-type topic change in the resultant dialog and its corresponding part of the topic map. Original sentences in the dialogs are in Japanese.

Fig. 7 shows part of the resultant dialog with derivation-type topic changes. Note that the statements were translated from Japanese to English and may not encompass grammatical accuracy in some cases. The system sequentially answered comments about dog walking, master–servant relationships, and dog training. In 172 comments answered by our system, we found three types of answers related to the three subjects. Though small in number, these results suggest that our system can potentially realize derivation-type topic changes in a realistic setting.

We looked into the dialogs where topic changes did not occur and evaluated their causes. In most cases at the beginning of a dialog, the system could not find answer topics and answered with its default answer, “I don’t know.” This is because initial topics did not exist in either the input sentence or the dialog history in the topic map.

In addition, there were cases when the system replied with the same answer to different inputs. This occurred because the system randomly chose an answer if it obtained multiple answer candidates. Based on this, it is considered that there are few answer candidates available based on the contents of the input sentence and dialog history; hence, the nearest neighbor search results in selection of the previous answer. Fig. 8 shows an example when the system replied with the same answer to different inputs.

User2 > I want relieve my dog’s stress.
System > It is known that a dog is insensible to the chemical substance that does not exist in nature.
User2 > My dog litters my room with garbage.
System > It is known that a dog is insensible to the chemical substance that does not exist in nature.

Fig. 8. Example when the system replied with the same answer to different inputs.

Further, there were cases where differences of opinion existed between the system and the subject. Fig. 9 shows an example of this. Here the subject inputted the speech “his dog is not attached to him,” and the system replied with an answer. As the conversation continued, the user input a topic completely different and the system responded with topics related to “his dog is not attached to him” again. This is because the system is significantly dependent on dialog histories.

User3 > My dog is not attached to me.
System > It is said that a dog wags its tail to a leader.
......
User3 > My dog dislikes the air conditioner.
System > Did you scold your dog severely?

Fig. 9. Example of differences in opinions between the system and the subject.

In summary of the experimental results, sequential variation covered topics ranging from dog walking to dog training in 172 statements. This result suggests that our system can realize derivation-type topic changes in a realistic setting. We looked into the dialogs where topic changes did not occur and discussed their causes and solutions. First, we evaluated the case when the system could not find answer topics at the start of the dialogs. This is because topics in the input sentence or dialog history did not have associations in the topic map. Therefore, we should improve the topic map to have a significant number of topics and associations. Second, we evaluated the case when the system replied with the same answer to different inputs. This is because there are few answer candidates that are obtained from the input sentence or the dialog history. Therefore, we need to consider an improved method to realize topic changes since this problem is unsolvable with only derivation-type topic changes. Finally, we evaluated the case when differences of opinion occurred between the system and the subject. This occurs because the system is significantly dependent on dialog histories. Our conclusion is that it is necessary to focus on the latest input more tightly than on the dialog history. Therefore, we need to consider a weighting system for the dialog history.

In the topic map used in this experiment, we gave the answer sentences by hand. We conducted the same experiment but the used topic map is generated from the data disclosed in the Internet. Topics and associations were extracted from the sentences in articles in Yahoo! Japan News, and answers were extracted from tweets in Twitter. First, the sentences including nouns in the tweets were extracted from the articles. Then, we created associations representing the dependency relations between a subject (=the noun in the tweet) and a verb in the sentences. We created topics that represent the subjects and the verbs. We collected data in the term from December 1st to December 13th in 2013. The resultant topic map had 8,840 topics, 86,074 associations and 4092 answers. The subjects were seven university students and generated 124 statements. We found that 10 answers generated by our system made derivation-type topic changes. This indicates that the result that our system can give us answers with derivation-type topic change does not depend on the topic map used in the experiment.

5. Conclusion

In recent years, there have been many studies on systems that "chat" with the user. However, these systems do not consider topic changes. Therefore, in this study, we focus on the fact that topics often change during a dialog.

To realize derivation-type topic changes, we proposed a dialog system utilizing topic maps. We regarded the topic map as a knowledge base that stores relationships between dialog topics and answers, and proposed the application of topic maps to realize the dialog system. In addition, we proposed a neighborhood search in the topic map to realize derivation-type topic changes. In addition, we proposed the utilization of dialog history since not only the dialog system but also the user may make a derivation-type topic change.

The experimental results suggest that our system can realize derivation-type topic changes in a realistic setting. However, we need to consider solving the cases in which topic changes were not followed properly by the dialog system.

In the future, we will improve the topic map by increasing the number of topics and answers. In addition, we need to discuss and implement improved topic changing methods. Finally, we need to consider a weighting system for the dialog history to reduce the system's dependency on previous topics.

References

[1] J. Weizenbaum, ELIZA-A Computer Program for the Study of Natural Language Communication between Man and Machine, Communications ACM, 9(1), 36-45, 1966.
[2] D. Hofstadter, Preface 4: The Ineradicable Eliza Effect and Its Dangers, from Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought, Basic Books: New York, 1995.
[3] Y. Yamamoto, Extraction of Sense of Incompatibility in Computer Conversation Paying Attention to Correspondence of Past Conversation, Forum on Information Thechnology, 7(2), 293-294, 2008.
[4] Y. Hong, An overview of topic shifting research: Focusing on the types and strategies, Japanese language education. Special issue, The state of the art in second language acquisition and instruction research: a guidepost to Japanese language education for the new century, 2005, pp.159-185, 2005.
[5] ISO/IEC 13250-2: Information Technology–Topic Maps–Part 2: Data Model. International Organization for Standardization, 2006.
[6] T. Kudo, Y. Matsumoto, Fast Methods for Kernel-Based Text Analysis, ACL 2003.
[7] Y. Kuribara, T. Hosoya, M. Kimura, Tome: Topic maps database extended. Proceedings of SEATUC, 2010.