Chapter 32
Developing Non-goal Dialog System Based on Examples of Drama Television

Lasguido Nio, Sakriani Sakti, Graham Neubig, Tomoki Toda, Mirna Adriani, and Satoshi Nakamura

Abstract This paper presents a design and experiments of developing a non-goal dialog system by utilizing human-to-human conversation examples from drama television. The aim is to build a conversational agent that can interact with users in as natural a fashion as possible, while reducing the time requirement for database design and collection. A number of the challenging design issues we faced are described, including (1) filtering and constructing a dialog example database from the drama conversations and (2) retrieving a proper system response by finding the best dialog example based on the current user query. Subjective evaluation from a small user study is also discussed.

32.1 Introduction

Natural language dialog systems have so far mostly focused on two main dialog genres: goal-oriented dialog (such as ATIS flight reservation [1], DARPA communicator dialog travel planning [2]) and non-goal-oriented dialog (such as chatterbot systems like Eliza [3] or Alice [4]). Though various techniques have been proposed, data-driven approaches to dialog have become the most common method used in

L. Nio (✉)
Nara Institute of Science and Technology, Ikoma, Japan

Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
e-mail: lasguido@is.naist.jp

S. Sakti • G. Neubig • T. Toda • S. Nakamura
Nara Institute of Science and Technology, Ikoma, Japan
e-mail: ssakti@is.naist.jp; neubig@is.naist.jp; tomoki@is.naist.jp; s-nakamura@is.naist.jp

M. Adriani
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
e-mail: mirna@cs.ui.ac.id

J. Mariani et al. (eds.), *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialog Systems into Practice*, DOI 10.1007/978-1-4614-8280-2_32,

dialog agent design. Example-based dialog modeling (EBDM) is one of several data-driven methods for deploying dialog systems. The basic idea of this approach is that a dialog manager (DM) uses dialog examples that are semantically indexed in a database, instead of domain-specific rules or probabilistic models [5]. With various sources of natural conversation examples, the usage of EBDM techniques has great potential to allow more efficient construction of natural language dialog systems.

Many studies have been conducted to develop technologies related to EBDM, such as a back-end workbench for implementing EBDM [6], query relaxation based on correlation for EBDM [7], and confirmation modeling for EBDM [8]. However, tedious and time-consuming design, collection, and labeling of a large set of user-system interactions are often required. Moreover, the scripted design scenarios in a lab typically result in unnatural conversations, with users responding differently from what is found in real situation. Consequently, many studies use EBDM to find the best responses or utilize template from available log databases [9]. To address this problem, some studies have proposed using Twitter data or crowdsourcing over large databases [10]. These techniques are also used by chat bots like Jabberwacky.[1] and Cleverbot[2] However, on the other hand, the issue of how to handle uncontrolled conversation content still remains.

One way to overcome these problems was proposed by [11] IRIS (Informal Response Interactive System), which uses a vector space model to implement a chat-oriented dialog system based on movie scripts [12]. Following their work, we further make improvements on the retrieval system by using a semantic similarity formula [13] with examples from drama television. The aim is to build a conversational agent that could interact with users as naturally as possible, while reducing the time requirement for database design and collection. One of the advantages of using examples from drama television is that the conversation content is more natural than scripted lab dialog design, since it contains some humorous dialog conversation. Yet, it is still within controlled drama scenes. More or less, drama television also affect the way people communicate. To build an example database, we propose a *tri-turn* unit for dialog extraction and semantic similarity analysis techniques to help ensure that the content extracted from raw movie/drama script files forms appropriate dialog examples.

32.2 System Overview

Figure 32.1 shows an overview of our system architecture. The system includes two components: (1) filtering and construction of a dialog example database from the drama conversations and (2) retrieval of a proper system response by finding the

[1] Jabberwacky—http://www.jabberwacky.com.

[2] Cleverbot—http://www.cleverbot.com.

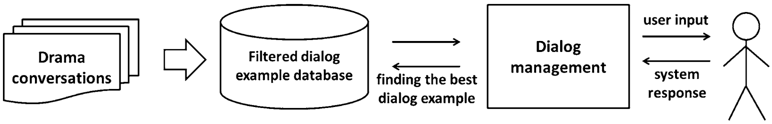

Fig. 32.1 System overview

best dialog example based on the current user query. Each of these components is described in the following sections.

32.3 Filtering Data

In EBDM, one of the important tasks is to filter and construct a dialog example database from the drama conversations. The challenge is that many drama dialog-turn conversations are not two-way "query-and-response" sentences. Even consecutive dialog turns may contain disjoint conversations from more than two persons/actors, which makes identifying the query and response difficult (see Table 32.1). In this study, to make sure the dialog examples are based on two-way "query-and-response" sentences, we select dialog data by proposing a concept called the trigram-turn sequence or *tri-turn*.

An example of a tri-turn dialog is shown in Table 32.2. The first and last utterances of the tri-turn are performed by the same person or actor (i.e., Joey), while the second turn is performed by another actor (i.e., Rachel). When a tri-turn pattern exists, we can generally assume that the two-actor conversation has a two-way "query-and-response" format.

After extracting the tri-turn from a dialog script, all words in all tri-turns were labeled by part of speech (POS) tagger and named entity (NE) recognizer. NE generalization was performed with normalizing all person or place name into general form such as "Joey" to "that man" or "Japan" to "that place."

Semantic similarity matching (similar to the approach introduced in [13]) is performed to ensure a high semantic relationship between each dialog turn in the dialog pair data. The formula requires two sentences (S_1 and S_2) and its synset (S_{syn1} and S_{syn2}) as an input. As shown in Eq. 32.1, the similarity is computed using WordNet[3] synsets in each dialog turn. Finally, the tri-turn dialogs exceeding a similarity threshold are extracted and included into the database

$$sem_{sim}(S_1, S_2) = \frac{2 \times |S_{syn1} \cap S_{syn2}|}{|S_{syn1}| + |S_{syn2}|}. \tag{32.1}$$

[3]Wordnet—http://wordnet.princeton.edu/.

Table 32.1 Example of dialog conversations in *Friends* drama television with multiple actors

Actor	Sentence
Rachel	Oh, he is precious! Where did you get him?
Ross	My friend Bethel rescued him from some lab.
Phoebe	That is so cruel! Why? Why would a parent name their child Bethel?
Chandler	Hey, that monkey's got a Ross on its ass!
Monica	Ross, is he gonna live with you, like, in your apartment?

Table 32.2 Example of a tri-turn with two actors from the *Friends* drama television

Actor	Sentence
Joey	I might know something.
Rachel	I might know something too.
Joey	What's the thing you know?

32.4 Dialog Management

The dialog management consists of two important elements, the dialog template and the response search. Both are described in the following.

32.4.1 Dialog Template

Figure 32.2 shows the overall dialog system template. It mainly consists of three conversations states: the *greeting* state, the *discussion* state, and the *farewell* state. The system responses for *greeting* and *farewell* states will be selected randomly from a handmade template combined with *greeting* and *discussion* examples in the database. For the *discussion* state, every time the system receives a user input, it generates the response with the highest similarity score from the example database. If no example is found, the system will respond "I don't understand what you mean" and send a new topic. To avoid repetitive responses, the system will search responses from dialog turns that have not been selected previously.

32.4.2 Retrieving Proper Response

A proper system response is retrieved by measuring both semantic and syntactic relations. These two measures are combined using linear interpolation as shown below:

$$sim(S_1,S_2) = \alpha \times sem_{sim}(S_1,S_2) + (1-\alpha) \times cos_{sim}(S_1,S_2). \tag{32.2}$$

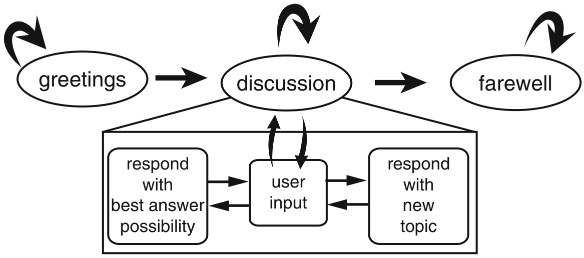

Fig. 32.2 Dialog system template

This value is calculated over the user input sentence (S_1) and every input examples on database (S_2). These values are calculated using semantic similarity in WordNet as a semantic factor and POS tag cosine similarity

$$cos_{sim}(S_1, S_2) = \frac{S_1 \cdot S_2}{\| S_1 \| \| S_2 \|} \tag{32.3}$$

as a syntactic factor. In this study, we assume the semantic factor is more important than the syntactic factor, so we set the interpolation coefficient α to be 0.7. Finally, if there is more than one retrieved example on database, the system will give more priority to the shortest response.

32.5 Experiments and Evaluation

We conduct our experiments using *Friends*[4] TV show scripts. Parsing the HTML data is done with the Perl CPAN HTML-Parser[5] and the system is built in the Python environment using the Python NLTK tools.[6]

The *Friends* TV show's scripts are written in English and contain 5 seasons, with a total of 112 episodes. Each episode contains several scenes and each scene contains several dialog turns. The total number of scenes and dialog turns in the corpus are 1,437 and 26,658, respectively. By applying the filter function, the total number of gathered tri-turns is 18,690. Based on semantic similarity, we finally extract and collect 1,901 tri-turn dialogs with high similarity.

The main objective of the evaluation in this preliminary study is to investigate the usefulness or effectiveness of different approaches in giving responses through the example database. Three different systems are evaluated: (1) only data filtering, (2) data filtering and NE generalization, and (3) data filtering, NE generalization, and priority selection on the shortest related answer.

[4]Friends Scripts—http://ufwebsite.tripod.com/scripts/scripts.htm.

[5]Perl CPAN—http://search.cpan.org/dist/HTML-Parser/Parser.pm.

[6]NLTK—http://nltk.org.

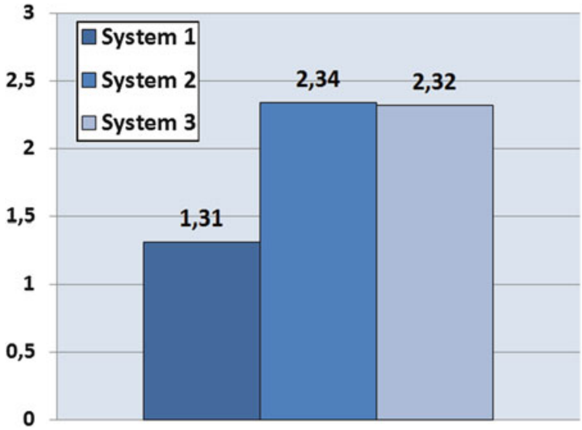

Fig. 32.3 Evaluation result

Table 32.3 Dialog conversations snippet between system and user

Actor	Sentence
System	Hi.
User	How are you?
System	What are you doing here?
User	I just wanted to see how it's going.
System	Well, it's going okay.
User	Where have you been?
System	Oh, just had a baby.
User	This is unbelievable.
System	So, what is this information worth to you, my friend?
User	It's not that bad.
System	Not that bad? Did you hear the hammer story?
User	Which one do you have?

All systems are evaluated by asking a human evaluator to give a score between 1 and 3 for every dialog-turn response. Score 1 means that the system is not giving a natural and relevant response to the user input and score 3 means that the system is giving a natural and relevant response to the user input. Each system is evaluated ten times by ten different evaluators and results are shown in Fig. 32.3.[7] It is expected that system 2 with full capabilities would outperform system 3 and system 1. However, it also reveals that the NE generalization gave a significant improvement in naturalness, while selecting a short controlled response had no significant difference. A conversation snippet between the user and system 2 can be seen in Table 32.3.

[7]The standard deviation from system 1 to 3 is 0.563, 0.768, and 0.764.

32.6 Conclusion

This paper addressed a non-goal EBDM dialog system based on examples from a TV show's dialog scripts. Filtering is performed to capture relevant dialog chat in the example corpus. We compared three different approaches to giving responses using an example database. The results reveal that the NE generalization from conversation in tri-turns gives a significant effect of naturalness, while selecting a short controlled response has no significant difference. However, much of the work shown in this paper is a preliminary work. Many improvements should be done to present a better non-goal dialog system. Future work could be done by adding a learning process to the system, so that it can remember the context of the conversation. Furthermore, compounding other examples from other data sources is also necessary to extend the system response.

References

1. Seneff, E., Hirschman, L., Zue, V.: Interactive problem solving and dialogue in the ATIS domain. In: Proceedings of the Fourth DARPA Speech and Natural Language Workshop, pp. 354–359 (1991)
2. Walker, M., Aberdeen, J., Boland, J., Bratt, E., Garofolo, J., Hirschman, L., Le, A., Lee, S., Narayanan, S., Papineni, K., Pellom, B., Polifroni, J., Potamianos, A., Prabhu, P., Rudnicky, A., Sanders, G., Seneff, S., Stallard, D., Whittaker, S.: DARPA communicator dialog travel planning systems: The June 2000 data collection. In: Proceedings of EUROSPEECH, pp. 1371–1374 (2000)
3. Weizenbaum, J.: Eliza - a computer program for the study of natural language communication between man and machine. Commun. ACM **9**(1), 36–45 (1966)
4. Wallace R.: Be Your Own Botmaster. A.L.I.C.E A.I. Foundation, California (2003)
5. Lee, C., Jung, S., Kim, S., Lee, G.: Example-based dialog modeling for practical multi-domain dialog system. Speech Commun. **51**(5), 466–484 (2009)
6. Jung, S., Lee, C., Lee, G.: Dialog studio: An example based spoken dialog system development workbench. In: Proceedings of the Dialogs on Dialog: Multidisciplinary Evaluation of Advanced Speech-Based Interactive Systems. Interspeech2006-ICSLP Satellite Workshop, Pittsburg (2006)
7. Lee, C., Lee, S., Jung, S., Kim, K., Lee, D., Lee, G.: Correlation-based query relaxation for example-based dialog modeling. In: ASRU, pp. 474–478. Merano (2009)
8. Kim, K., Lee, C., Lee, D., Choi, J., Jung, S., Lee, G.: Modeling confirmations for example-based dialog management. In: SLT, pp. 324–329. Berkeley, California (2010)
9. Murao, H., Kawaguchi., N., Matsubara, S., Yamaguchi, Y., Inagaki, Y.: Example-based spoken dialogue system using WOZ system log. In: SIGDIAL, pp. 140–148. Saporo (2003)
10. Bessho, F., Harada, T., Kuniyoshi, Y.: Dialog system using real-time crowdsourcing and Twitter large-scale corpus. In: SIGDIAL, pp. 227–231. Seoul (2012)
11. Banchs, R.E., Li, H.: IRIS: A chat-oriented dialogue system based on the vector space model. In: ACL (System Demonstrations), pp. 37–42 (2012)
12. Banchs, R.E.: Movie-dic: A movie dialogue corpus for research and development. In: ACL (2), pp. 203–207. The Association for Computer Linguistics (2012). URL http://dblp.uni-trier.de/db/conf/acl/acl2012-2.html
13. Liu, D., Liu, Z., Dong, Q.: A dependency grammar and wordnet based sentence similarity measure. J. Comput. Inform. Syst. **8**(3), 1027–1035 (2012)