Developing Non-Goal Dialog System based on Examples of Drama Television

Lasguido[†,‡], Sakriani Sakti[†], Graham Neubig[†], Tomoki Toda[†], Mirna Adriani[‡], 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.

1 Introduction

Natural language dialogue systems have so far mostly focused on two main dialogue 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 dialogue 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 modelling for EBDM [8]. However, tedious and time consuming design, collection, and labeling of a large set of user-system interactions is 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

---

Lasguido, S. Sakti, G. Neubig, T. Toda, S. Nakamura
[†]Nara Institute of Science and Technology (Japan)
e-mail: {lasguido,ssakti,neubig,tomoki,s-nakamura}@is.naist.jp

Lasguido, M. Adriani
[‡] Faculty of Computer Science, Universitas Indonesia (Indonesia)
e-mail: lasguido@ui.ac.id,mirna@cs.ui.ac.id

the best responses or utilize template from available log databases [9]. To address this problem, some studies have proposed of 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 advantages of using examples from drama television is that the conversation content is more natural than scripted lab dialog design, since contain some humorous dialog conversation. Yet, it is still within controlled drama scenes. To build a 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 an appropriate dialog examples.

2 System Overview

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

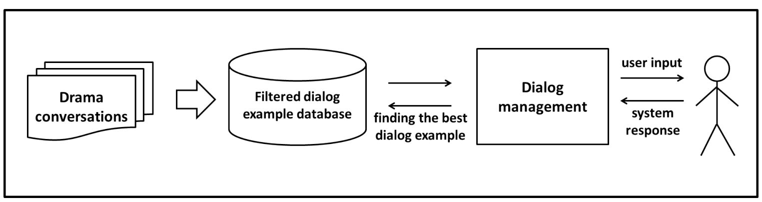

Fig. 1 System overview.

3 Filtering Data

In EBDM, one of 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 1). In this study, to make sure the dialog examples are based on two-way "query-and-response" sen-

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

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

tences, we select dialog data by proposing a concept called the trigram turn sequence or *tri-turn*.

Table 1 Example of dialog conversations in Friends drama television[3] 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 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?

An example of a tri-turn dialog is shown in Table 2. The first and last utterance 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 was labeled by part of speech (POS) tagger and named entity (NE) recognizer. NE generalization were performed with a 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. 1, the similarity is computed using WordNet[4] synsets in each dialog turn. Finally, the tri-turn dialogs exceeding a similarity threshold are extracted and included into the database

$$sem_{sim}(S_1,S_2) = \frac{2 \times |S_{syn1} \cap S_{syn2}|}{|S_{syn1}| + |S_{syn2}|}. \tag{1}$$

4 Dialog Management

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

4.1 Dialog Template

Figure 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 hand-made template combined with *greeting* and *discussion* examples in the

[2] http://ufwebsite.tripod.com/scripts/scripts.htm

[4] http://wordnet.princeton.edu/

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.

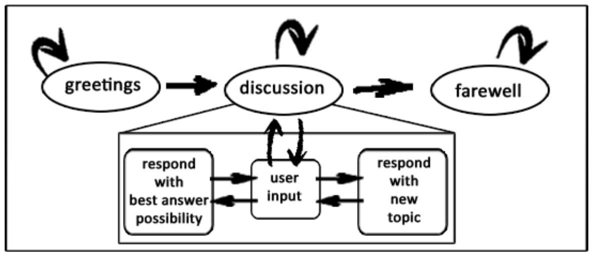

Fig. 2 Dialog System Template

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

$$sim(S_1,S_2) = \alpha \times sem_{sim}(S_1,S_2) + (1-\alpha) \times cos_{sim}(S_1,S_2). \quad (2)$$

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

$$cos_{sim}(S_1,S_2) = \frac{S_1 \cdot S_2}{\| S_1 \| \| S_2 \|} \quad (3)$$

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 examples on database, the system will give more priority to the most shortest response.

5 Experiments and Evaluation

We conduct our experiments using Friends 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

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

[6] http://nltk.org

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.

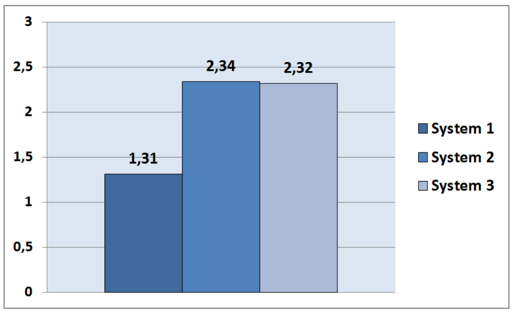

Fig. 3 Evaluation result

All systems are evaluated by asking a human evaluator to give a score between 1-3 for every dialog turn response. The 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 10 times by 10 different evaluators and results are shown in Fig. 3[7]. It is expected that system 2 with full capabilities would outperform the 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 the system 2 can be seen in the Table 3.

Table 3 Dialog conversations snippet between system and the 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?

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

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 give 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 improvement 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 the system 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. E. Seneff, L. Hirschman, and V. Zue, "Interactive problem solving and dialogue in the ATIS domain," in *Proc. of the Fourth DARPA Speech and Natural Language Workshop*, 1991, pp. 354–359.
2. M. Walker, J. Aberdeen, J. Boland, E. Bratt, J. Garofolo, L. Hirschman, A. Le, S. Lee, S. Narayanan, K. Papineni, B. Pellom, J. Polifroni, A. Potamianos, P. Prabhu, A. Rudnicky, G. Sanders, S. Seneff, D. Stallard, and S. Whittaker, "DARPA communicator dialog travel planning systems: the June 2000 data collection," in *Proc. of EUROSPEECH*, 2000, pp. 1371–1374.
3. J. Weizenbaum, "Eliza - a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966.
4. R.Wallace, *Be Your Own Botmaster*, A.L.I.C.E A.I. Foundation, 2003.
5. C. Lee, S. Jung, S. Kim, and G.G. Lee, "Example-based dialog modeling for practical multi-domain dialog system," *Speech Communication*, vol. 51, no. 5, pp. 466–484, 2009.
6. S. Jung, C. Lee, and G.G. Lee, "Dialog studio: An example based spoken dialog system development workbench," in *Proc. of the Dialogs on dialog: Multidisciplinary Evaluation of Advanced Speech-based Interactive Systems. Interspeech2006-ICSLP satellite workshop,*, Pittsburg, USA, 2006.
7. C. Lee, S. Lee, S. Jung, K. Kim, D. Lee, and G.G. Lee, "Correlation-based query relaxation for example-based dialog modeling," in *Prof. of ASRU*, Merano, Italy, 2009, pp. 474–478.
8. K. Kim, C. Lee, D. Lee, J. Choi, S. Jung, and G.G. Lee, "Modeling confirmations for example-based dialog management," in *Prof. of SLT*, Berkeley, California, USA, 2010, pp. 324–329.
9. H. Murao, N. Kawaguchi., S. Matsubara, Y. Yamaguchi, and Y. Inagaki, "Example-based spoken dialogue system using WOZ system log," in *Proc. of SIGDIAL*, Saporo, Japan, 2003, pp. 140–148.
10. F. Bessho, T. Harada, and Y. Kuniyoshi, "Dialog system using real-time crowdsourcing and Twitter large-scale corpus," in *Proc. of SIGDIAL*, Seoul, South Korea, 2012, pp. 227–231.
11. Rafael E. Banchs and Haizhou Li, "IRIS: a chat-oriented dialogue system based on the vector space model," in *ACL (System Demonstrations)*, 2012, pp. 37–42.
12. Rafael E. Banchs, "Movie-dic: a movie dialogue corpus for research and development.," in *ACL (2)*. 2012, pp. 203–207, The Association for Computer Linguistics.
13. D. Liu, Z. Liu, and Q. Dong, "A dependency grammar and wordnet based sentence similarity measure," *Journal of Computational Information Systems*, vol. 8, no. 3, pp. 1027–1035, 2012.