PAPER Special Section on Advances in Modeling for Real-world Speech Information Processing and its Application

Utilizing Human-to-Human Conversation Examples for a Multi Domain Chat-Oriented Dialog System

Lasguido NIO[†a)], ***Nonmember***, **Sakriani SAKTI**[†b)], ***Member***, **Graham NEUBIG**[†c)], ***Nonmember***, **Tomoki TODA**[†d)], ***and*** **Satoshi NAKAMURA**[†e)], ***Members***

SUMMARY This paper describes the design and evaluation of a method for developing a chat-oriented dialog system by utilizing real human-to-human conversation examples from movie scripts and Twitter conversations. The aim of the proposed method 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) constructing an appropriate dialog corpora from raw movie scripts and Twitter data, and (2) developing an multi domain chat-oriented dialog management system which can retrieve a proper system response based on the current user query. To build a dialog corpus, we propose a unit of conversation called a tri-turn (a trigram conversation turn), as well as extraction and semantic similarity analysis techniques to help ensure that the content extracted from raw movie/drama script files forms appropriate dialog-pair (query-response) examples. The constructed dialog corpora are then utilized in a data-driven dialog management system. Here, various approaches are investigated including example-based (EBDM) and response generation using phrase-based statistical machine translation (SMT). In particular, we use two EBDM: syntactic-semantic similarity retrieval and TF-IDF based cosine similarity retrieval. Experiments are conducted to compare and contrast EBDM and SMT approaches in building a chat-oriented dialog system, and we investigate a combined method that addresses the advantages and disadvantages of both approaches. System performance was evaluated based on objective metrics (semantic similarity and cosine similarity) and human subjective evaluation from a small user study. Experimental results show that the proposed filtering approach effectively improve the performance. Furthermore, the results also show that by combing both EBDM and SMT approaches, we could overcome the shortcomings of each.

key words: *dialog corpora, response generation, example-based dialog modeling, semantic similarity, cosine similarity, machine translation*

1. Introduction

The continuous growth of information technology is having an increasingly large impact on many aspects of our daily lives. The issue of communication via speech between human beings and information-processing machines is also becoming more important [1]. A common dream is to realize a technology that allows humans to communicate or have dialogs with machines through natural and spontaneous speech.

Manuscript received September 28, 2013.
Manuscript revised January 18, 2014.

[†]The authors are with the Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630–0192 Japan.
a) E-mail: lasguido.kp9@is.naist.jp
b) E-mail: ssakti@is.naist.jp
c) E-mail: neubig@is.naist.jp
d) E-mail: tomoki@is.naist.jp
e) E-mail: s-nakamura@is.naist.jp

DOI: 10.1587/transinf.E97.D.1497

Natural language dialogue systems have so far mostly focused on two main dialogue genres: goal-oriented dialog (such as ATIS flight reservation [2], DARPA Communicator dialog travel planning [3]) and non-goal-oriented dialog (such as chatterbot systems like Eliza [4] or Alice [5]). Dialog systems can also be described by the amount of human intervention used in their construction, ranging from entirely hand-made to completely data-driven. Seminal work often limited interactions to a specific scenario (e.g. a Rogerian psychotherapist [4]) or were based on complex, knowledge-rich rule-based systems for generating responses, which required large amounts of human effort to create or add new rules [5].

Example-based dialog modeling (EBDM) is data-driven approach for deploying dialog systems. It uses dialog examples that are semantically indexed to a database. Proper responses for user input are generated based on these dialog examples. Consequently, to achieve good coverage on various types of natural conversation, recording of a large data set of real human-to-human conversation is necessary, which is tedious and time consuming. Common solutions use handmade scripted dialog scenarios which may result in unnatural conversations. Some studies also propose constructing dialog examples from available log databases, such conversation between human subjects and the Wizard of OZ (WOZ) system [6], or human-to-human conversation in Twitter [7].

However, covering all possible patterns that may exist in real human-to-human conversation is still difficult. Currently, most EBDM systems rely on either canned responses by providing error messages [8] or templates for generation which may result a completely incomprehensible response [9]. On the other hand, SMT has been successfully used to address various NLP tasks [10]–[12]. The investigation of SMT as an approach for response generation has also been introduced by [13].

The goal of our work is to create a dialog agent that can interact with the user in as natural a fashion as possible. In this paper we focus on addressing the following two main challenging issues:

Corpus construction: We propose a method to utilizing human-to-human conversation examples from movies and and Twitter data. The aim is to gain insights in how to build a conversational agent that can interact with users in as natural a way as possible, while reduc-

Copyright © 2014 The Institute of Electronics, Information and Communication Engineers

ing the time requirement for database design and collection. Then, to help ensure that the content extracted from raw movie/drama script files consists of appropriate dialog-pair (query-response) examples, we propose using a unit called a tri-turn for extraction, as well as semantic similarity analysis techniques.

Dialog management: We investigate various data-driven approaches to dialog management, including two EBDM techniques (syntactic-semantic similarity retrieval and TF-IDF based cosine similarity retrieval) and using phrase-based SMT to learn a conversational mapping between user-input and system-output dialog-pairs. We also propose a simple, but effective way to perform system combination of example-based and SMT-based techniques into one dialog management framework. Experimental results demonstrate that our combined system shows promise for overcoming the shortcomings of each approach.

In the next session, we first describe how to construct dialog corpora from raw movie scripts and Twitter data. We then briefly describe the proposed dialog management systems based on EBDM and SMT technique in Sect. 3. Experimental set-up is briefly describe in Sects. 4. Then, a detailed evaluation of our proposed approaches is presented in Sect. 5. Some related research works are discussed in Sect. 6. Finally, conclusions are drawn in Sect. 7.

2. Construction of Dialog Corpora

The dialogue corpora that we constructed in this study are based on dialog-pair sentences. As the movie scripts and Twitter data used in this work contain very different types of text, we go through different processes to construct them.

A movie scripts is a conversational manuscript that portray the conversations between and actions of between actors in a movie. Figure 1 (a) illustrates an example of one movie scene with four actors talking to each other. The corresponding raw movie scripts that are available from the web are usually written in HTML files shown in Fig. 1 (b). The dialog between actors are arranged in chronological order. Consequently, the conversation dialog contained in movie scenes does not have a clear indication of which utterances are responses to a particular utterance. Therefore, it is important to find a solution that is able to construct appropriate dialog-pair examples from raw movie scripts files. As shown in Fig. 1 (c), we perform dialog tri-turn extraction to find the candidate of dialog-pairs.

In contrast to the movie scripts data, the text on Twitter often represents real conversations between two or more people. Therefore, we do not perform dialog tri-turn extraction in order to extract the related dialog-pair sentences. Instead, the challenge with handling Twitter data is how to ensure the integrity of the sentences. In this case, it is necessary to filter out sentence pairs that are not likely to be useful for training the system.

Unifying both data sources into one dialog corpus, we

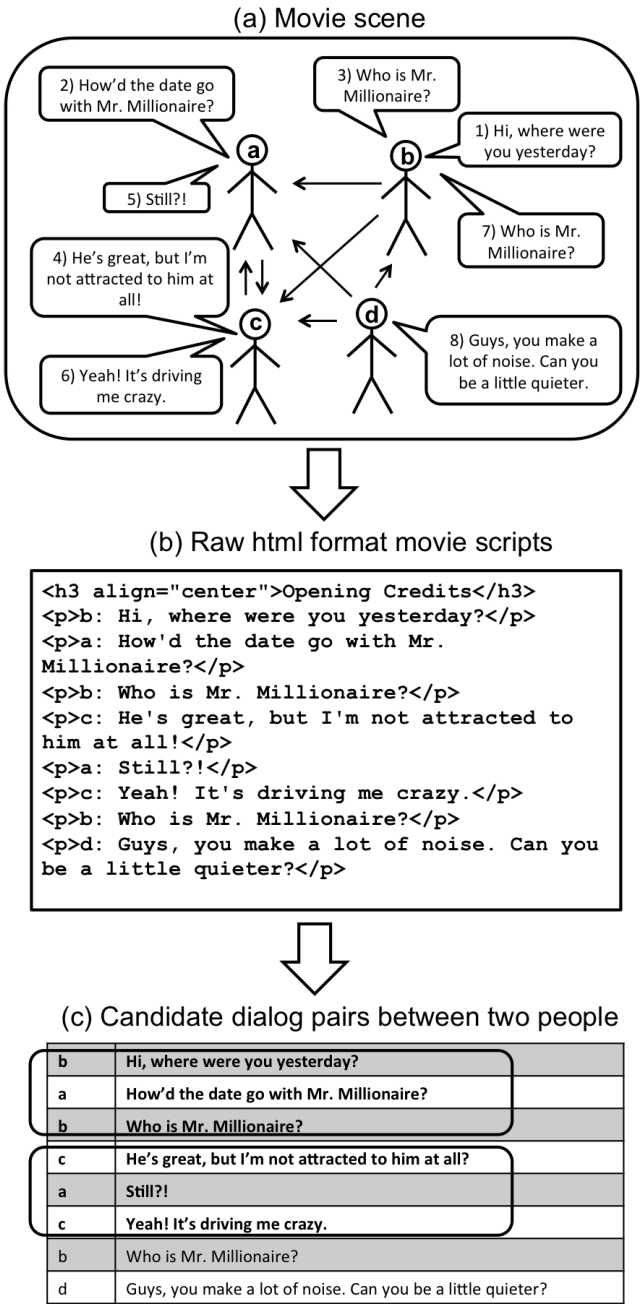


Fig. 1 Dialog corpora construction from movie script.

define two basic types of information about each dialog: actor and utterances. The utterances are the actual content of each dialog turn in the movie scripts or tweets. The actor refers to the character name in the movies, or the name of the Twitter user that posted each tweet. This actor and utterance information will be utilized to construct the dialog corpus.

The details of dialog corpus construction, as illustrated in Fig. 2, consists of three main steps: (1) preprocessing, which remove unnecessary information and normalize the text, (2) a dialog-pairs extraction, which ensures that the conversation is between two people talking each other, and

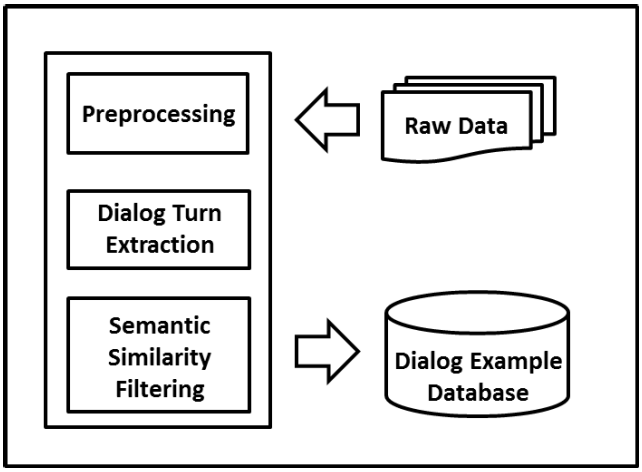


Fig. 2 Dialog corpora construction from movie script.

(3) a semantic similarity calculation, which ensure that the each query-response pair is semantically related. We describe the details of each step in the following sections.

2.1 Preprocessing

Preprocessing of the movie scripts is done by transforming raw HTML files into easily readable text format. Since we use a variety source of movie scripts that had various formats, we implemented several parsing algorithms to fetch the information from the raw movie conversation. Furthermore, unnecessary explanatory information about the movie scenes is also removed.

For the Twitter data, preprocessing removes information about person identity, hash tags, and URLs. Next, for both data sets all the words in the sentences are labeled with parts of speech (POS) and named entities (NE). Finally, to ensure the integrity of the Twitter data, English language filtering [†] and non-standard word (NSW) normalization [14] is also performed.

2.2 Dialog Turn Extraction

To ensure that the dialog example database contains only query-response pairs, we propose a simple and intuitive method for selection of the dialog data: trigram turn sequences, or *tri-turn*. A tri-turn is defined as three turns in a conversation between two actors X and Y that has the pattern X-Y-X. In other words, within a tri-turn the first and last dialog turn are performed by the same actor and the second dialog turn is performed by the other actor. Next the query-response pairs are made by separating the tri-turn pattern X-Y-X into two pairs, X-Y and Y-X.

We found that when we observed this pattern, in the great majority of the cases this indicated that the first and second utterances (X-Y pair), as well as the second and third utterances (Y-X pair), formed a proper input-response pair

Actor	Correlated tri-turn
c	He's great, but I'm not attracted to him at all!
a	Still?!
c	Yeah! It's driving me crazy.
Actor	**Un-correlated tri-turn**
b	Hi, where were you yesterday?
a	How'd the date go with Mr. Millionaire?
b	Who is Mr. Millionaire?

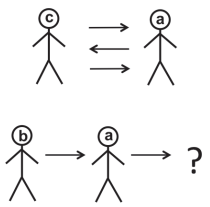


Fig. 3 Example of a tri-turn with two actors.

as shown in the c-a-c tri-turn in Fig. 3. However, noisy cases which contain uncorrelated turns still exist (see the b-a-b tri-turn in Fig. 3), this happens because the speakers are not actually speaking to each-other. To address this problem, we perform further filtering using the semantic similarity measure described in the following section.

2.3 Semantic Similarity

Semantic similarity (similar approach introduced in [15]) is used to ensure a strong semantic relationship between each dialog turn in the dialog-pair data. As shown in Eq. (1), it computes the similarity between WordNet [††] synsets in each dialog turn. The dialog-pairs with high similarity are then extracted and included into database. S_{syn1} and S_{syn2} respectively is a group of WordNet synsets for each word in the sentence S_1 and S_2 that are linked by a complex network of lexical relations. The similarity of sentence pair X-Y where S_1 = X and S_2 = Y can be obtained by calculating the relations between S_{syn1} and S_{syn2}. Where $|S_{syn1} \cap S_{syn2}|$ is a number of co-occurring WordNet synsets and $|S_{syn1}|+|S_{syn2}|$ is a total number of effective WordNet synsets.

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

3. Dialog Management Systems

The overview of our dialog management method system is shown in Fig. 4. It utilizes dialog templates within three states: the *greeting* state, *free-chat* state, and *farewell* state. For every user input query, dialog management will forward to the response generators systems. The response generators then provide various possible responses, which finally voted by the system for the best responses and forwards it as an system responses output. In this study, various approaches of response generators are investigated including syntactic-semantic similarity and TF-IDF based cosine similarity retrievals of EBDM techniques, as well as phrase-based SMT approaches which are discussed in the next section. Note that, as current focus here is to investigate the optimal technique for retrieving a proper system response based on the current user query, the utilizing of user history and dialog context will not be discussed in this paper.

[†]http://search.cpan.org/~ambs/Lingua-Identify-0.51/

[††]http://wordnet.princeton.edu/

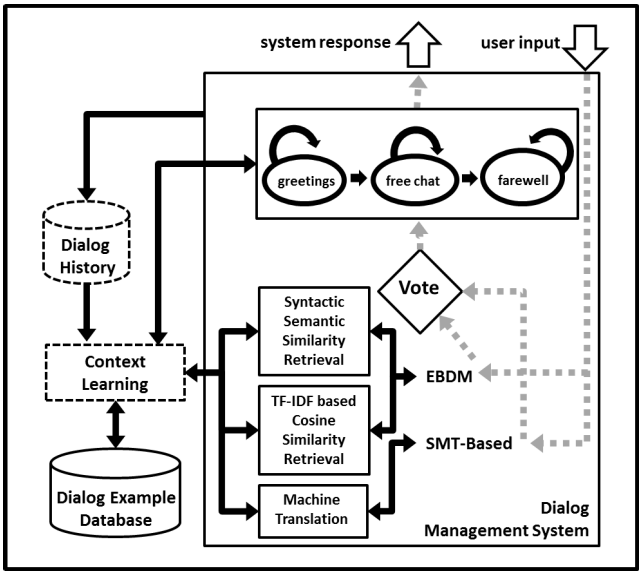


Fig. 4 Dialog management systems.

3.1 Example-Based Dialog Modeling

In EBDM, the system chooses a response from the examples stored in the dialog database. In order to do so, it computes a similarity measure between the user input and the query part of the query-response pairs, and returns the associated response for the query with the highest similarity. In this work, we examine syntactic-semantic similarity and TF-IDF based cosine similarity as two similarity measures for use in EBDM.

3.1.1 Syntactic-Semantic Similarity Retrieval

A proper system response is retrieved by measuring both semantic and syntactic relations. These two measures are combined using linear interpolation as shown in Eq. (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 cosine similarity (Eq. (3)) over part-of-speech (POS) tag vectors as a syntactic factor.

$$sim(S_1, S_2) = \alpha[sem_{sim}(S_1, S_2)] + (1 - \alpha)[cos_{sim}(S_1, S_2)] \tag{2}$$

where

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

In this work, we assumed that the semantic factor is more important than syntactic factor, so we set the interpolation coefficient α to be 0.7.

3.1.2 TF-IDF Based Cosine Similarity Retrieval

Cosine similarity over the term vector as described in Eq. (3) is used to retrieve a proper system response. To increase the emphasis on important words, an additional TF-IDF weighting (Eq. (4)) is performed to construct the term vector [16].

$$TFIDF(t, T) = F_{t,T} \log\left(\frac{|T|}{DF_t}\right) \tag{4}$$

We define $F_{t,T}$ as a term frequency 't' in a sentence 'T', and DF_t as a total number of sentence in the query-response pairs that contain term 't'.

Table 1 Total characters involved in one movie.

Characters	Percentage
1 – 10	4.40%
11 – 20	28.62%
21 – 30	16.94%
31 – 40	18.02%
41 – 50	14.71%
≥ 51	17.30%

3.2 SMT-Based

The second approach we tested was based on SMT [13]. With this approach, the dialog-pair data is treated as a parallel corpus for training an SMT system. Given the trained SMT system, the user dialog is treated as an input and "translated" into the system response. The system response is chosen to be system output S of maximal probability given the user input T

$$\hat{S} = \arg\max_{S} P(S \mid T). \tag{5}$$

4. Experimental Set-Up

In this paper, the movie script dialog is collected through Friends TV show scripts [†], The Internet Movie Script Database [††], and The Daily Script [†††]. This resulted in a total of 1,786 movie scripts with 1,042,288 dialog-pairs. After performing dialog turn extraction and semantic similarity filtering, the total number of dialog-pairs is 86,719. Table 1 shows the total different characters involved in one movie. Mostly around 28.62% collected movie scripts is played by 11 – 20 different characters. Only 4.40% collected movie scripts is played by 1 – 10 different characters. Therefore besides the main characters, the movie scripts usually composed by a cameo e.g. *"a man in the radio"*, *"man 1"*, *"radio"*, and so on. These cameo characters contribute to increase the characters variation on a single movie show.

For the Twitter data, Twitter *tweets* was collected through the Twitter API [††††], resulting in a total of 1,076,447 dialog-pairs. After performing language filtering and semantic similarity filtering, the total number of dialog-pairs

[†]http://ufwebsite.tripod.com/scripts/scripts.htm
[††]http://imsdb.com/
[†††]http://dailyscript.com/
[††††]http://dev.twitter.com

Table 2 Example of different system responses.

Input	Shall we eat at my house?
Response 1	Sorry, I ate already.
Response 2	Yes, sure.
Response 3	Of course, will you cook?
Response 4	Great! But, where is your house?

was reduced to 67,500 and 7,048 respectively.

After extracting all the dialog-pairs, we randomly selected 500 and 1000 query-response pairs from Twitter and movie conversation dialog, respectively, as a test set (the query-response pairs are denoted as $\langle Q_{test}, R_{test} \rangle$). Then, the remaining dialogs will be used as dialog examples for EBDM, and training data for SMT (the dialog-pairs here are denoted as $\langle Q_{train}, R_{train} \rangle$).

The natural language processing tools and Wordnet synsets used were provided by NLTK toolkit [†], and the example-based TF-IDF based cosine similarity retrieval was performed using Apache Lucene [††]. For the SMT approach, Moses [†††] was used to build the translation model and perform translation for the dialog system. Here, four-gram language models built with the Kneser-Ney smoothing and the lexicalized distortion model were used.

Given a query from the test set (Q_{test}), the EBDM will search the closest query examples using syntactic-semantic similarity retrieval: sim(Q_{test}, Q_{train}) or TF-IDF based cosine similarity retrieval: cos(Q_{test}, Q_{train}), and output a response of R_{output}. However, as there may be a number of valid system responses for a single user query (see the example in Table 2), it is not trivial to evaluate the system performance. Therefore, in this study, we attempt to assess each output response within semantic, syntactic, and human judge aspect. We investigate the performance of our system by utilizing TF-IDF based cosine similarity and syntactic-semantic based similarity to represent the semantic and syntactic aspect, and subjective evaluation to represent the human judge aspect. For objective evaluation, the R_{output} are evaluated by computing similarity with R_{test}: sim(R_{output}, R_{test}) and cos(R_{output}, R_{test}). Also when performing subjective evaluation in user-system interaction, Q_{test} are given, and the users are evaluate the naturalness of R_{output} in comparison with R_{test}.

5. Evaluation

5.1 Evaluation of Dialog Corpora Construction

To demonstrate the effect of semantic similarity and tri-turn filtering in our data, we compare our system performance with and without the tri-turn filtering. In this comparison, the TF-IDF based cosine similarity and syntactic-semantic similarity retrieval methods were used to retrieve responses

[†]http://nltk.org
[††]http://lucene.apache.org/
[†††]http://statmt.org/moses/

Table 3 Various approach on the system response generation.

Approach	Abbreviation
EBDM	
- Syntactic-Semantic Similarity Retrieval	sssr
- TF-IDF based Cosine Similarity Retrieval	csm
SMT	smt
Combination EBDM and SMT	comb

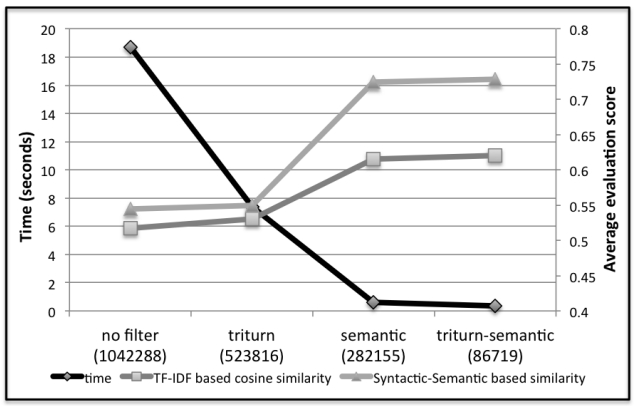


Fig. 5 Filtering effect on the movie data.

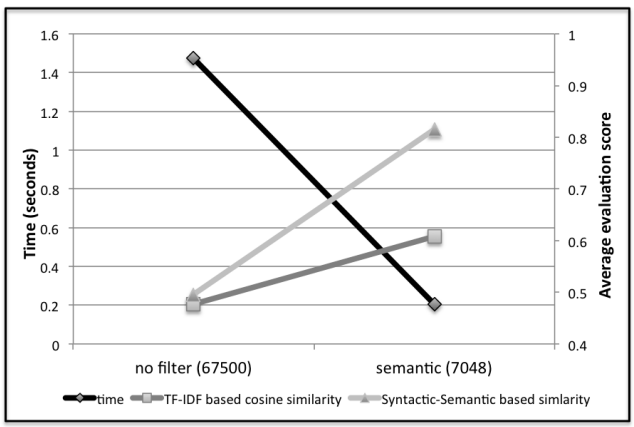


Fig. 6 Filtering effect on the Twitter data.

(Table 3 shows the various approaches to retrieve the system response). Semantic similarity filtering could improve the performance significantly over the tri-turn filtering. However, the application of the tri-turn filtering has a role in reducing the amount of training example while maintaining the evaluation score result. On the other hand, the difference in the amount of training examples resulting from the filtering process also affect the response retrieval time. Figure 5 and 6 depict average system evaluation score improvement and response time per input query for each applied filter. The number in the horizontal axis shows the amount of dialog-pairs in each filtering steps.

5.2 Evaluation of the Dialog Management System

Objective evaluation presented in Fig. 7 and Fig. 8 is per-

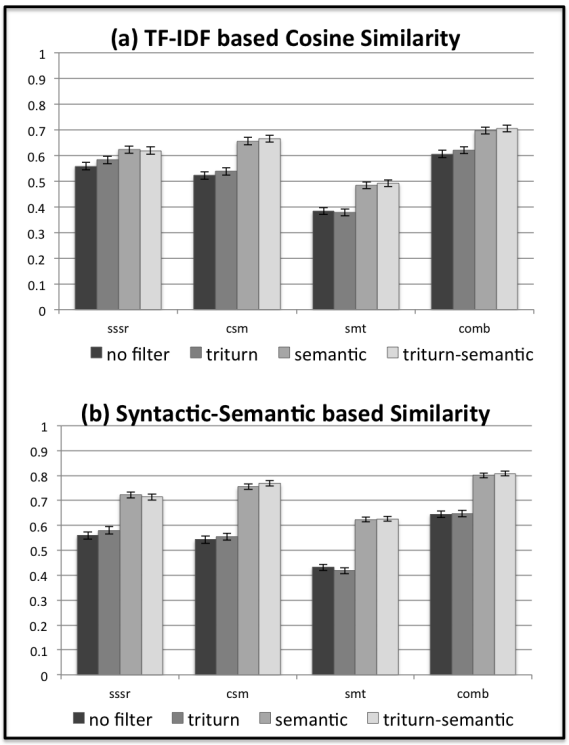


Fig. 7 Objective evaluation results on the movie data by various data-driven approaches.

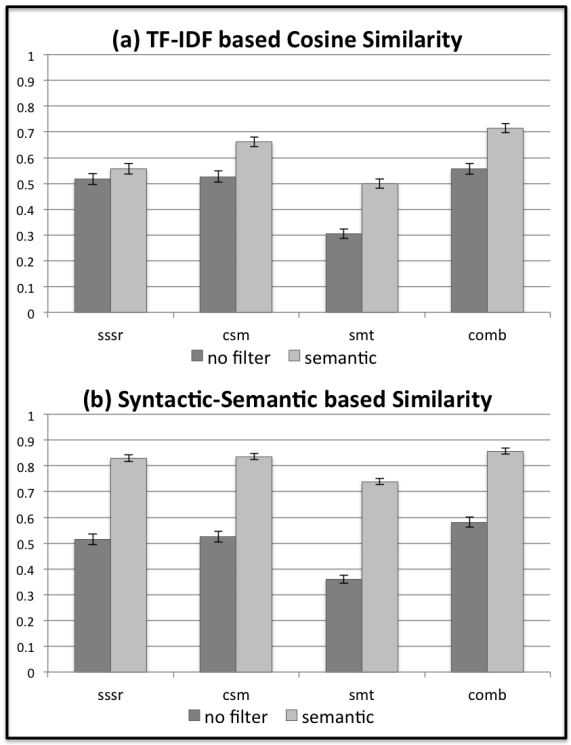


Fig. 8 Objective evaluation results on the Twitter data by various data-driven approaches.

formed using TF-IDF based cosine similarity and syntactic-semantic based similarity (with $p < 0.05$). The results reveal that, within EBDM approach, TF-IDF based cosine similarity retrieval (denoted as "csm") gives a better response score than syntactic-semantic similarity retrieval (denoted as "sssr"). The csm approach exceeds the sssr approach because this approach utilizes cosine similarity over the TF-IDF vector, while the sssr approaches compute cosine similarity over the POS tag vector. Furthermore, the tri-turn and semantic similarity filtering methods manage to increase the response score.

Comparing the best EBDM approach "csm" againsts the SMT approach "smt", csm always give a better performance than smt. Analyzing the data in more detail, we found that csm is better in handling when dialogs close to Q_{test} exists in Q_{train}, while smt can provide a better output when there is no dialogs in Q_{train} similar with Q_{test}. Combining both approaches (denoted as "comb") in which the system uses EBDM if the similarity between user input and dialog examples exceeds given threshold, and responds with SMT output otherwise, could overcome the shortcomings of each approach. The objective evaluation results on system combination given various thresholds are presented in Fig. 9. The results reveal that the best system is provided by the combined system. The optimum score shown here is achieved by 0.4 and 0.6 for movie and Twitter data respectively.

A cross-domain evaluation between movie and Twitter data is also performed. In this experiment we use

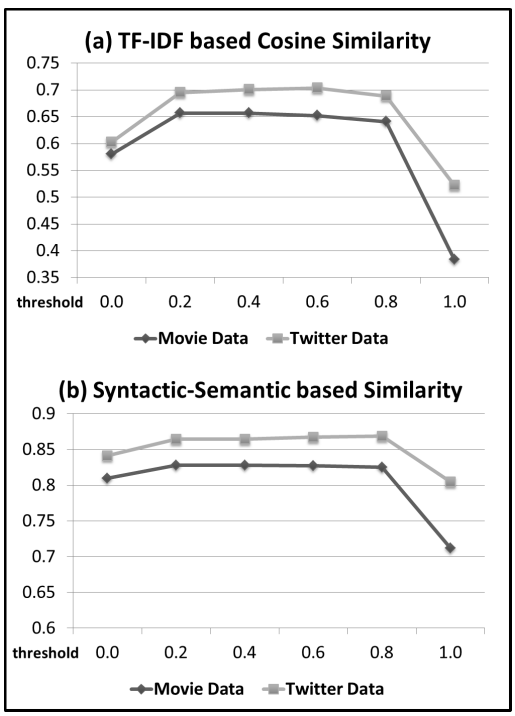


Fig. 9 Objective evaluation results of the combined system given various threshold.

the "comb" retrieval approach to retrieve responses within movie and Twitter filtered data. Both the Twitter and movie test data is tested in the movie, Twitter, and combined movie

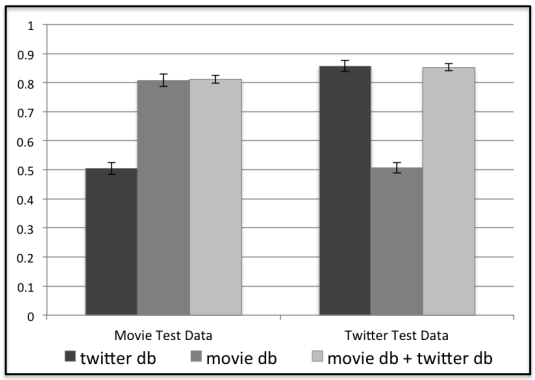


Fig. 10 Combination retrieval approach on cross-domain using syntactic-semantic similarity as an evaluation metric.

and Twitter database. The results of the cross-domain evaluation can be seen in the Fig. 10.

In the subjective evaluation, 40 human annotators were asked to give a naturalness score between 1-5 of the system output, with higher scores indicating that the system is giving a natural and relevant system response to the user input. Each person assesses 140 randomly selected query-response pairs that were evenly distributed over all systems. The results of this evaluation are shown in Fig. 11. We also prepare a dummy system as a baseline which outputs a response by simply repeating the user input, i.e. user-input: "How are you?", then the system's output is also: "How are you?". For greeting conversations, this simple approach may work. But, for the other cases, the system may result in a completely incomprehensible response.

Along with objective evaluation, the results show that the csm approach significantly outperform the smt approach. This may indicate that the smt responses consists of several matching phrases with the reference, but have not yet reached the naturalness of real human responses. For instance, for a query input "I'll call you back.", the smt system will responded "I call me back.". Because this sentence is incomprehensible, many people will prefer the dummy system response "I'll call you back." instead of the smt response. This factor seems to affect the system combination as well, where it reduced the score slightly compared with the csm approach. Furthermore, the results of the subjective evaluation also demonstrate slightly higher scores on filtered data. This shows that the tri-turn and semantic similarity filtering methods manage to increase the naturalness of the response.

6. Related Work

There have been a number of related works into EBDM and SMT-based approaches for response generation in data-driven chat [17]–[19]. Work by [8] proposes a generic dialog modeling framework for a multi-domain dialog systems to simultaneously manage goal-oriented and chat dialogs for information access and entertainment. However, the chat-oriented dialog only includes small talk limited to 10 top-

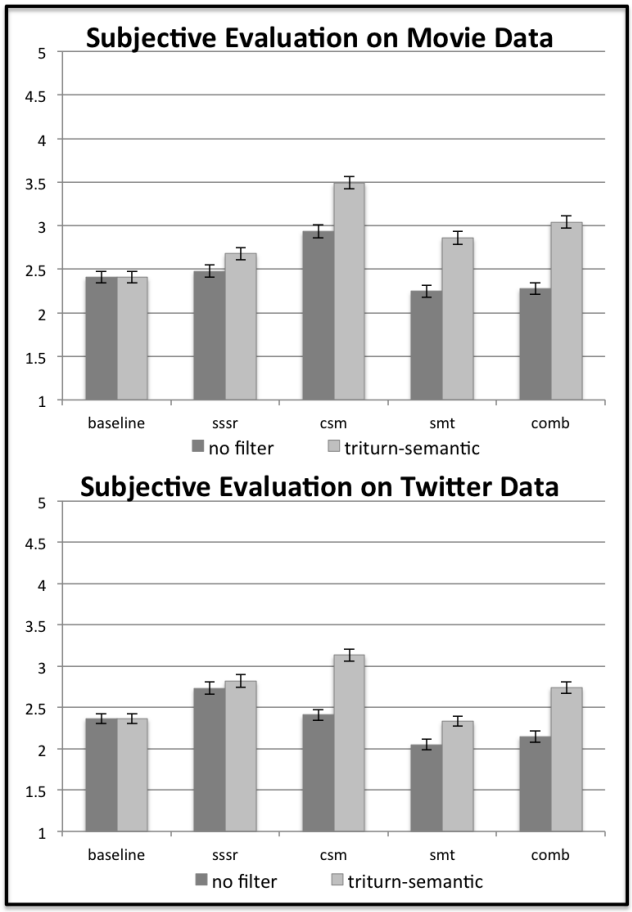


Fig. 11 Subjective evaluation result on the movie and Twitter data by various data-driven approaches.

ics of daily conversation. Furthermore, if the system cannot find similar examples to determine the next system action, it simply defines a "No Example" output error and provides an in-coverage example of what the user could say at the current dialog state. Finally, [20] introduce IRIS (Informal Response Interactive System), a chat oriented dialog system using movie scripts that is based on a similar cosine similarity in vector space model. However, the system did not filter any uncorrelated consecutive scripts in the movie data, and, as the authors state, this causes failures and diminishes the ability to maintain a consistent conversation.

Despite the relatively large interest in data-driven approaches for chat or response generation, there is surprisingly little work comparing and contrasting approaches or data sources. In this work, we attempt make an empirical evaluation that will contrast these approaches and provide a reference for future development in the area. In addition, to the best of our knowledge, our method to combine example-based and SMT-based response generation in dialog modeling, and utilizing tri-turn in dialog systems is also different from previous work.

7. Conclusion

In this work, we investigated several approaches to build a data-driven chat-oriented dialog systems. The proposed tri-turn extraction and semantic similarity filtering are able to extract dialog-pair examples from multi-speaker dialog of raw movie scripts and Twitter data. Experimental results also reveal that that the tri-turn and semantic filtering improve the objective evaluation metrics score (TF-IDF based cosine similarity and syntactic-semantic similarity evaluation metrics). Furthermore we also find out that this approach also helps to reduce the retrieval response time by reducing dialog examples in training set. It relies on no explicit domain knowledge, and should therefore be applicable to other dialog applications with little or no modification.

Furthermore, we compared three different approaches to giving responses using an example database. We found that the example-based approach is very good in handling the queries which are similar to the examples in the database, but achieves poor performance in handling the queries which are far different from existing examples. On the other hand, SMT-based systems showed the opposite tendency. We also introduced a system that combines example-based and SMT-based approaches to take advantage of the characteristics of both approaches.

As future work, investigating ways to improve the naturalness and cohesion of responses generated by the SMT approach may be necessary. Adding a learning process that considers the context of the conversation could also lead to further improvements. This would allow the system to both remember the context of the conversation and expand the example database. Therefore, including the user history and dialog context in the dialog management system is a promising future direction. Combining other approaches in the chat-oriented dialog framework could also demonstrate interesting results.

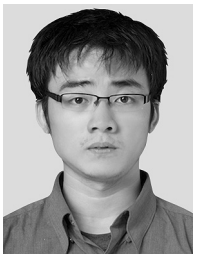
Acknowledgements

Part of this work was supported by the collaborative research with I2R [†].

References

[1] J. Holmes and W. Holmes, Speech Synthesis and Recognition, Taylor & Francis, London, UK, 2001.
[2] E. Seneff, L. Hirschman, and V. Zue, "Interactive problem solving and dialogue in the ATIS domain," Proc. Fourth DARPA Speech and Natural Language Workshop, 1991, pp.354–359, 1991.
[3] 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," Proc. EUROSPEECH, 2000, pp.1371–1374, 2000.
[4] J. Weizenbaum, "Eliza – computer program for the study of natural language communication between man and machine," Commun. ACM, vol.9, no.1, pp.36–45, Jan. 1966.
[5] R. Wallace, Be Your Own Botmaster, A.L.I.C.E A.I. Foundation, California, USA, 2003.
[6] H. Murao, N. Kawaguchi, S. Matsubara, Y. Yamaguchi, and Y. Inagaki, "Example-based spoken dialogue system using WOZ system log," Proc. SIGDIAL, pp.140–148, Saporo, Japan, 2003.
[7] F. Bessho, T. Harada, and Y. Kuniyoshi, "Dialog system using real-time crowdsourcing and twitter large-scale corpus," Proc. SIGDIAL, pp.227–231, Seoul, South Korea, 2012.
[8] C. Lee, S. Jung, S. Kim, and G.G. Lee, "Example-based dialog modeling for practical multi-domain dialog system," Speech Commun., vol.51, no.5, pp.466–484, May 2009.
[9] N. Chambers and J. Allen, "Stochastic language generation in a dialogue system: Toward a domain independent generator," Proc. SIGDIAL, pp.9–18, Cambridge, Massachusetts, USA, 2004.
[10] A. Echihabi and D. Marcu, "A noisy-channel approach to question answering," Proc. ACL, Morristown, pp.16–23, NJ, USA, 2003.
[11] Y.W. Wong and R. Mooney, "Learning for semantic parsing with statistical machine translation," Proc. HLT/NAACL, pp.439–446, New York City, NY, USA, 2006.
[12] Y.W. Wong and R. Mooney, "Generation by inverting a semantic parser that uses statistical machine translation," Proc. HLT/NAACL, pp.172–179, Rochester, NY, USA, 2007.
[13] A. Ritter, C. Cherry, and W.B. Dolan, "Data-driven response generation in social media," Proc. 2011 Conference on Empirical Methods in Natural Language Processing, pp.583–593, Edinburgh, Scotland, UK, July 2011.
[14] R. Sproat, A.W. Black, S.F. Chen, S. Kumar, M. Ostendorf, and C. Richards, "Normalization of non-standard words," Computer Speech and Language, vol.15, no.3, pp.287–333, 2001.
[15] D. Liu, Z. Liu, and Q. Dong, "A dependency grammar and wordnet based sentence similarity measure," J. Computational Information Systems, vol.8, no.3, pp.1027–1035, 2012.
[16] G. Salton, A. Wong, and C.S. Yang, "A vector space model for automatic indexing," Commun. ACM, vol.18, no.11, pp.613–620, Nov. 1975.
[17] S. Jung, C. Lee, and G. Lee, "Dialog studio: An example based spoken dialog system development workbench," Proc. Dialogs on dialog: Multidisciplinary Evaluation of Advanced Speech-based Interactive Systems, Interspeech2006-ICSLP satellite workshop, Pittsburg, USA, 2006.
[18] C. Lee, S. Lee, S. Jung, K. Kim, D. Lee, and G. Lee, "Correlation-based query relaxation for example-based dialog modeling," Proc. ASRU, pp.474–478, Merano, Italy, 2009.
[19] K. Kim, C. Lee, D. Lee, J. Choi, S. Jung, and G. Lee, "Modeling confirmations for example-based dialog management," Proc. SLT, pp.324–329, Berkeley, California, USA, 2010.
[20] R.E. Banchs and H. Li, "IRIS: A chat-oriented dialogue system based on the vector space model," ACL (System Demonstrations), pp.37–42, 2012.

[†]http://www.i2r.a-star.edu.sg/



Lasguido Nio received his B.C.S. and his M.C.S. degree from Universitas Indonesia, Indonesia in 2012 and 2013. He is currently a first year Doctor Student in Nara Institute of Science and Technology, Nara, Japan. His research interest include information retrieval and dialog systems.

Sakriani Sakti received her B.E. degree in Informatics (cum laude) from Bandung Institute of Technology, Indonesia, in 1999. In 2000, she received "DAAD-Siemens Program Asia 21st Century" Award to study in Communication Technology, University of Ulm, Germany, and received her MSc degree in 2002. During her thesis work, she worked with Speech Understanding Department, DaimlerChrysler Research Center, Ulm, Germany. Between 2003–2009, she worked as a researcher at ATR SLC Labs, Japan, and during 2006–2011, she worked as an expert researcher at NICT SLC Groups, Japan. While working with ATR-NICT, Japan, she continued her study (2005–2008) with Dialog Systems Group University of Ulm, Germany, and received her PhD degree in 2008. She actively involved in collaboration activities such as Asian Pacific Telecommunity Project (2003–2007), A-STAR and U-STAR (2006–2011). She also served as a visiting professor of Computer Science Department, University of Indonesia (UI) in 2009–2011. Currently, she is an assistant professor of the Augmented Human Communication Lab, NAIST, Japan. She is a member of JNS, SFN, ASJ, ISCA, IEICE amd IEEE. Her research interests include statistical pattern recognition, speech recognition, spoken language translation, cognitive communication, and graphical modeling framework.

Graham Neubig received his B.E. from University of Illinois, Urbana-Champaign, U.S.A, in 2005, and his M.E. and Ph.D. in informatics from Kyoto University, Kyoto, Japan in 2010 and 2012 respectively. He is currently an assistant professor at the Nara Institute of Science an Technology, Nara, Japan. His research interests include speech and natural language processing, with a focus on machine learning approaches for applications such as machine translation, speech recognition, and spoken dialog.

Tomoki Toda was born in Aichi, Japan on January 18, 1977. He earned his B.E. degree from Nagoya University, Aichi, Japan, in 1999 and his M.E. and D.E. degrees from the Graduate School of Information Science, NAIST, Nara, Japan, in 2001 and 2003, respectively. He was a Research Fellow of JSPS in the Graduate School of Engineering, Nagoya Institute of Technology, Aichi, Japan, from 2003 to 2005. He was an Assistant Professor of the Graduate School of Information Science, NAIST from 2005 to 2011, where he is currently an Associate Professor. He has also been a Visting Researcher at the NICT, Kyoto, Japan, since May 2006. From March 2001 to March 2003, he was an Intern Researcher at the ATR Spoken Language Communication Research Laboratories, Kyoto, Japan, and then he was a Visiting Researcher at the ATR until March 2006. He was also a Visiting Researcher at the Language Technologies Institute, CMU, Pittsburgh, USA, from October 2003 to September 2004 and at the Department of Engineering, University of Cambridge, Cambridge, UK, from March to August 2008. His research interests include statistical approaches to speech processing such as voice transformation, speech synthesis, speech analysis, speech production, and speech recognition. He received the 18th TELECOM System Technology Award for Students and the 23rd TELECOM System Technology Award from the TAF, the 2007 ISS Best Paper Award and the 2010 ISS Young Researcher's Award in Speech Field from the IEICE, the 10th Ericsson Young Scientist Award from Nippon Ericsson K.K., the 4th Itakura Prize Innovative Young Researcher Award and the 26th Awaya Prize Young Researcher Award from the ASJ, the 2009 Young Author Best Paper Award from the IEEE SPS, the Best Paper Award (Short Paper in Regular Session Category) from APSIPA ASC 2012, the 2012 Kiyasu Special Industrial Achievement Award from the IPSJ, and the 2013 Best Paper Award (Speech Communication Journal) from EURASIP-ISCA. He was a member of the Speech and Language Technical Committee of the IEEE SPS from 2007 to 2009. He is a member of IEEE, ISCA, IPSJ, and ASJ.

Satoshi Nakamura received his B.S. from Kyoto Institute of Technology in 1981 and Ph.D. from Kyoto University in 1992. He was a director of ATR Spoken Language Communication Research Laboratories in 2000–2008, and a vice president of ATR in 2007–2008. He was a director general of Keihanna Research Laboratories, National Institute of Information and Communications Technology, Japan in 2009–2010. He is currently a professor and a director of Augmented Human Communication laboratory, Graduate School of Information Science at Nara Institute of Science and Technology. He is interested in modeling and systems of spoken dialog system, speech-to-speech translation. He is one of the leaders of speech-to-speech translation research projects including C-STAR, IWSLT and A-STAR. He headed the world first network-based commercial speech-to-speech translation service for 3-G mobile phones in 2007 and VoiceTra project for iPhone in 2010. He received LREC Antonio Zampoli Award, the Commendation for Science and Technology by the Ministry of Science and Technology in Japan. He is an elected board member of ISCA, International Speech Communication Association, and an elected member of IEEE SPS, speech and language TC.