Selection method of an appropriate response in chat-oriented dialogue systems

Hideaki Mori
Kyoto Institute of Technology
/ Matsugasaki Sakyo-ku
Kyoto 6068585 Japan
`mori@ii.is.kit.ac.jp`

Masahiro Araki
Kyoto Institute of Technology
/ Matsugasaki Sakyo-ku
Kyoto 6068585 Japan
`araki@kit.ac.jp`

Abstract

Chat functionality is currently considered an important factor in spoken dialogue systems. In this paper, we explore the architecture of a chat-oriented dialogue system that can continue a long conversation with users and can be used for a long time. To achieve this goal, we propose a method combining various types of response generation modules, such as a statistical model-based module, a rule-based module, and a topic transition-oriented module. The core of this architecture is a method for selecting the most appropriate response based on a breakdown index and a willingness index.

1 Introduction

In recent years, there have been some research and development case studies on open-domain chat dialogue systems. The merit of chat functionality in a dialogue system is to encourage the daily use of the system so as to accustom the user to the speech interface. Moreover, chat dialogue functionality can give a user a sense of closeness to the system, especially for novice users of the speech interface. Considering this situation, the requirements of a chat dialogue system are (1) to maintain a longer dialogue without a breakdown of the conversation and (2) to maintain the long duration of use. We call the property of the first requirement as "continuous" and that of the second as "long-term." The aim of this paper is to propose a framework for realizing a continuous and long-term chat-oriented dialogue system.

In previous research studies on chat dialogue systems, the central theme of these studies is how to generate an appropriate and natural response to the user's utterance (Higashinaka et al., 2014),

(Xiang et al., 2014). There was little effort to realize both continuous and long-term features in chat-oriented dialogue systems.

The chat dialogue system's robustness to respond to any user utterance is a key functionality that must be implemented to make it continuous. Therefore, a statistical response generation method is used in recent chat-oriented dialogue systems. Moreover, appropriateness and naturalness of the response are required. To realize these functionalities, Higashinaka et al. proposed a method for evaluating the coherence of the system utterance to judge the latter's appropriateness(Higashinaka et al., 2014).

On the other hand, a chat system with a long-term feature should have the ability to keep the user interested and not bored. For example, it should be able to provide a new topic in a chat based on the recent news or seasonal event. It should also be able to develop a current topic for the dialogue by bringing up related topics. In general, it is difficult to realize such a topic shift in a statistical method. The rule-based method or a hybrid of rule and statistics is appropriate for implementing such functionalities.

Because of this difference in methods in implementing a suitable functionality for a continuous and long-term chat dialogue system, it is difficult to realize the aforementioned functionalities in one response generation module. Such module could be complex and difficult to maintain. Therefore, it is reasonable to implement the elemental functionalities in separate modules and combine them to generate one plausible response for the purpose of the continuous and long-term chat dialogue.

In this paper, we propose a framework for chat-oriented dialogue systems that can continue a long conversation with users and that can be used for a long-term. To achieve this goal, we pro-

pose a combination method of various types of response generation modules, such as a statistical model-based module, rule-based module, and topic transition-oriented module. The core of this architecture is a selection method of the most appropriate response based on the breakdown index and willingness index.

The rest of the paper is organized as follows. In Section 2, we explain the architecture of combining multiple response generation modules. In Section 3, we describe a selection method of the most appropriate response from several hypotheses. In Section 4, the demo description shows the details of the demonstration system. Finally, we conclude the paper in Section 5.

2 Response generation method

To realize a continuous chat dialogue, the system needs to be robust to various user utterances. Statistical methods (Sugiyama et al., 2013) (Banchs and Li, 2012) are popular in realizing the robust response generation. These methods can also generate a high-quality response in terms of appropriateness and naturalness. On the flip side of this strength, the system response tends to be confined to the expectations and, sometimes, the user considers it boring. As a result, the appropriateness and naturalness are not necessarily connected with the long-term use of the system.

Occasional and sometimes unexpected topic shift could make the chat interesting, but it requires a different response generation algorithm aiming for an appropriate and natural response.

Keeping the interest of the user in the chat system for a long-term requires changing the behavior of the system. If the system's utterance is gradually matching the user's preference, the user can feel a sense of closeness to the system. Such behavior is difficult to implement using the statistical method only. Some type of control by handwritten rule is required to begin the conversation with a new topic from the system side. In addition, the functionality of delivering the news filtered by the user's preference can encourage the daily use of the system. Such dialogue does not require robust dialogue management. The simple pattern is beneficial for both the user and the system.

As a result, the requirement of a continuous and long-term chat dialogue system is "to generate an appropriate and natural response as a majority behavior, but sometimes the system may generate an unexpected but interesting response and, sometimes, may start the chat by following the user's preference and recent news/topics." It is natural to divide the aforementioned, sometimes conflicting, functionality into individual specific modules and select the most plausible response among the candidates. Figure 1 shows our proposed architecture for realizing multiple response generations and the selection method. In the architecture, we used the following three chat dialogue systems:

- Rule-based system: This chat system is based on the ELIZA type system (Weizenbaum, 1966).

- Statistical model-based system: This one uses the NTT chat dialogue API (Yoshimura, 2014).

- Topic transition-oriented system: This one is implemented with a sequence-to-sequence model (Sutskever et al., 2014). First, the system extracts topics from user utterances and generates the nearest topic utterance in the word embedded space made by Word2Vec (Mikolov et al., 2013). By doing this, the chat system aims to generate a response that has related but unexpected contents.

The rule-based system can reply naturally when the rules match the user utterance appropriately, but it does not have a wide coverage. The statistical model-based system can respond to various topics, but sometimes it replies inappropriately. The topic transition-oriented system tends to generate unnatural responses, but sometimes it can generate appropriate ones and stimulate the user's willingness to chat effectively. We try to realize a continuous and long-term chat-oriented dialogue system by using the good aspects of these modules.

3 Evaluation method of the system response

As a result of the requirements discussed in Section 2, we created the following two evaluation indices:

Breakdown Index (BI):
    This index determines how natural the system utterance is.

Willingness Index (WI):
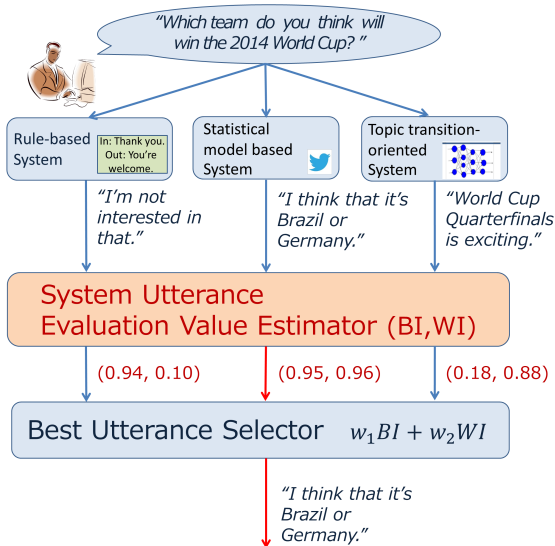   This one determines how the user's willingness is stimulated.

Figure 1: Proposed architecture for realizing multiple response generation modules, and the selection method.

Figure 2: Architecture of the demonstration system.

To create an estimator for the BI, we used a chat-oriented dialogue corpus collected by the dialogue task group of Project Next NLP, Japan[1]. We collected training data from this corpus based on bag-of-words (unigram) from 1000 utterances (10 * 100 dialogues), which have breakdown annotations by 24 participants for each utterance, and used a linear-kernel support vector machine (SVM) as the regressor for the target value.

To create an estimator for the WI, we calculated the similarity between user-system utterance and tweet-reply pair, and use the similarity as the WI. According to the online research[2] to Japanese user (1,496 people) who use Twitter one day a week or more, the top three purposes of using Twitter are "collecting infomation about their own hobbies", "as a pleasure", and "communicating with their friends and family". Thus, Japanese users mainly use Twitter for pleasure and communicating with familier person. Therefore We calculating similarity by using twitter copus as WI.

The method of calculating WI shows as follows. First, we applied NFKC (Normalization Form Compatibility Composition) to the sentences and removed inappropriate tweets such as tweets from bots. We collected about 205k tweet pairs and built the model based on the Paragraph Vec-
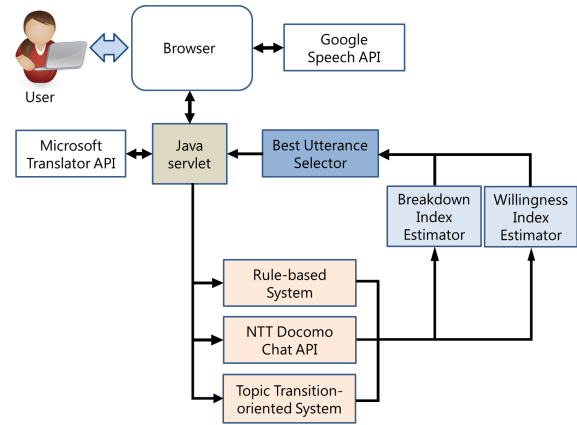
tor model (Le and Mikolov, 2014). Paragraph Vector is an unsupervised algorithm that learns fixed-length feature representations from variable-length pieces of texts, such as sentences, paragraphs, and documents. We used the Paragraph Vector model for vectorizing sentences and estimated the semantic similarity by calculating the cosine similarity. We get 10-best tweets which similar to user utterance, calculate similarity between the reply and system utterance, and use the maximum value as WI.

Finally, the proposed system calculates the weighted sum of BI and WI, and selects the utterance that has the highest weighted sum as a final output. The weight is set to optimize the system output by using development test set.

4 Demo description

Our chat-oriented dialogue system was implemented based on the proposed method described in Sections 2 and 3. Figure 2 shows the architecture of the demonstration system. This system aims to select the most appropriate response by considering its naturalness and willingness. The proposed chat-oriented dialogue system works on a Japanese sentence only. Therefore, the demonstration system translates Japanese sentences to English ones using the Microsoft Translator API and shows the dialogue in both Japanese and English.

5 Conclusion

In this work, we propose a selection method of the most appropriate response by considering its naturalness and willingness. Both a breakdown index

---

[1] https://sites.google.com/site/projectnextnlp/english-page
[2] http://www.opt.ne.jp/news/pr/detail/id=2341

230

and a willingness index, which are related to continuous and long-term functionality, respectively, contribute to deciding what a good utterance is in a chat dialogue. In future work, we plan to conduct an experimental evaluation on the continuous and long-term use of the chat dialogue system.

References

Rafael E Banchs and Haizhou Li. 2012. IRIS: a chat-oriented dialogue system based on the vector space model. In *Proceedings of the ACL 2012 System Demonstrations*, pages 37–42. Association for Computational Linguistics.

Ryuichiro Higashinaka, Toyomi Meguro, Kenji Imamura, Hiroaki Sugiyama, Toshiro Makino, and Yoshihiro Matsuo. 2014. Evaluating coherence in open domain conversational systems. In *INTERSPEECH*, pages 130–134.

Quoc V. Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. http://arxiv.org/abs/1405.4053. Mon, 02 Jun 2014 08:30:36 +0200.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *ICLR Workshop*.

Hiroaki Sugiyama, Toyomi Meguro, Ryuichiro Higashinaka, and Yasuhiro Minami. 2013. Open-domain utterance generation for conversational dialogue systems using web-scale dependency structures. In *The 14th annual SIGdial Meeting on Discourse and Dialogue*, pages 334–338.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Joseph Weizenbaum. 1966. ELIZA a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.

Yang Xiang, Yaoyun Zhang, Xiaoqiang Zhou, Xiaolong Wang, and Yang Qin. 2014. Problematic Situation Analysis and Automatic Recognition for Chinese Online Conversational System. *CLP 2014*, page 43.

Takeshi Yoshimura. 2014. Casual Conversation Technology Achieving Natural Dialog with Computers.