

Information Navigation System with Discovering User Interests

Koichiro Yoshino, Yu Suzuki and Satoshi Nakamura

Graduate School of Information Science

Nara Institute of Science and Technology

8916-5, Takayama-cho, Ikoma, Nara, 6300192, Japan

{koichiro,ysuzuki,s-nakamura}@is.naist.jp

Abstract

We demonstrate an information navigation system for sightseeing domains that has a dialogue interface for discovering user interests for tourist activities. The system discovers interests of a user with focus detection on user utterances, and proactively presents related information to the discovered user interest. A partially observable Markov decision process (POMDP)-based dialogue manager, which is extended with user focus states, controls the behavior of the system to provide information with several dialogue acts for providing information. We transferred the belief-update function and the policy of the manager from other system trained on a different domain to show the generality of defined dialogue acts for our information navigation system.

1 Introduction

A large number of dialogue systems to assist daily life of users have been deployed in the real world, however, most are handled on goal-oriented architecture (Young et al., 2010), on which clear goals of users are assumed. In contrast, an information navigation system (Yoshino and Kawahara, 2015) to clarify ambiguous user goals through a dialogue was proposed. The system provides information about current talking topics with several dialogue acts of providing information according to the ky strength of the user's interest to the current topic. If the user's intention is ambiguous and the user does not have any strong interest or focus, the system provides general information about the current discussion topic to help the user decide. However, if the user has specified interests or focus for some contents, the system answers the

user's question and presents additional information proactively according to the detected user's interests, even if the user can not find the exact words to express his or her interests.

The partially observable Markov decision process (POMDP)-based management architecture of information navigation was proposed (Yoshino and Kawahara, 2014) with several dialogue acts of providing information, however, this study ran the system only on one limited domain. In this demonstration, we used the belief-update function and the policy function trained on the different domain (news navigation for baseball news) for the proposed system (information navigation for tourist) to investigate the robustness of the defined system architecture. We also introduce a new mechanism to select more related topics when the system selects a topic to be presented in the next sub-dialogue by introducing semantic similarity of dialogue topics (content).

The proposed system has speech interfaces with open-source speech recognition system Julius (Lee and Kawahara, 2009)¹ and text to speech system OpenJtalk². This system runs on a standalone machine without connecting other servers, however, any modules can be replaced by other modules on other servers because modules connect each other on TCP/IP protocol.

2 System Architecture

Information providing is a sub-function of the general purpose function defined in ISO-24617-2 standard dialogue act set (Bunt et al., 2012), which includes dialogue acts of providing new or requested information from the dialogue partner. We defined seven dialogue modules to fulfill the information demand of a user though interactions

¹<https://github.com/julius-speech/julius>

²<http://open-jtalk.sourceforge.net/>

Table 1: Relations between defined dialogue modules in our system and dialogue acts defined in ISO-24617-2 (Bunt et al., 2012)

Category in ISO-24617-2		DAs in the proposed system	
General purpose function	Information Providing	Inform	Topic presentation (TP)
			Storytelling (ST)
	Answer	Question Answering (QA)	
	Information seeking	Propositional question	Confirmation (CO)
Dimension specific function	Social Obligation	*	Greeting (GR)
	Turn Management	Turn release	Keep Silent (KS)

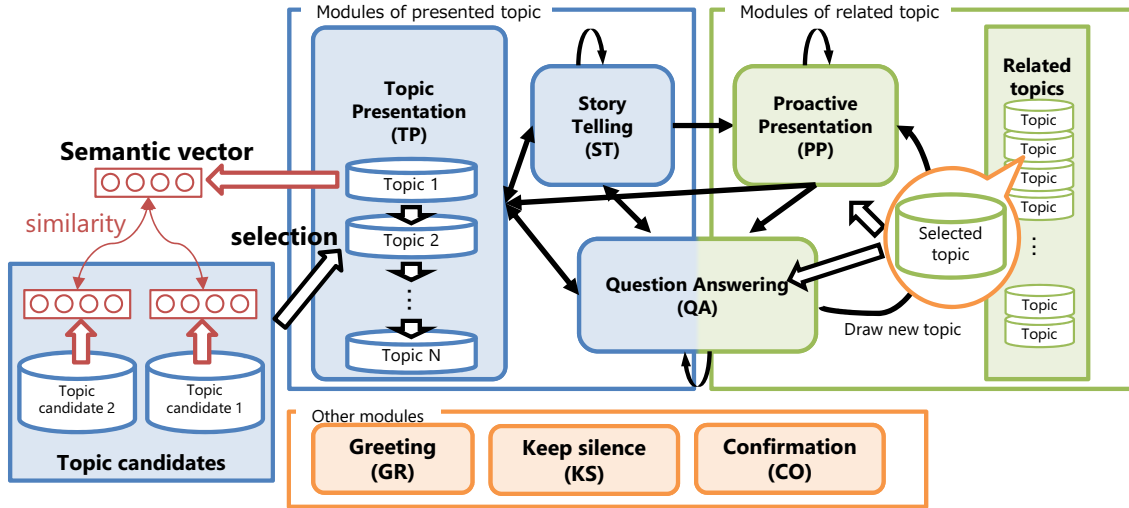


Figure 1: Overall architecture of proposed system.

by following this standard. The relations of defined modules and ISO-24617-2 dialogue acts are summarized in Table 1. We defined three sub-categories of dialogue act of “Inform” in the information providing function i.e., topic presentation (TP), storytelling (ST), and proactive presentation (PP), to enable smart information provision. The dialogue act “Answer” is also implemented as the question answering (QA) module in the same function. Our information navigation system basically provides information, but only one case in which the system uses the information-seeking function is confirmation (CO) to clarify the previous user intent. Dimension-specific functions support the general-purpose function; thus, minimum social obligation and turn-management functions are implemented as Greeting (GR) and Keep silent (KS).

Dialogue modules for these functions are defined for the task of news navigation (Yoshino and Kawahara, 2015), and the call of these modules at each turn is managed by the POMDP-based man-

ager. The call of system action is managed by the policy of POMDP, which takes into account the interests of the user (=user focus). The belief update is defined as,

$$b'_{s',f'} = P(s', f' | o'_{s'}, o'_{f'}) \sum_s P(s', f' | s, f, a) b_{s,f}, \quad (1)$$

where s denote the dialogue act of the user, f is the focus of the user on the current topic, and a is the selected dialogue module (dash means the next turn). The policy is trained in Q-learning as

$$\pi(b_{s,f}) = a \quad (2)$$

by using defined rewards and a user simulator constructed from dialogue data with users. The details are described in a previous study (Yoshino and Kawahara, 2015). The dialogue data of a different domain (news navigation for baseball news) is used to train the belief-update function and policy function for the current domain (sightseeing domain).

The order of topics (=documents of the information source) to be presented at each sub-dialogue is pre-defined in the current system. However, topics to be introduced by the system should be selected according to user interests in the dialogue history, which can be captured in the past dialogue. Thus, we introduce a mechanism to introduce new topics that will be more attractive for the user by using past topics that the user and system already discussed. We defined a semantic similarity measure to define the similarities of each pair of topics (=news). The system determines the next topic of the dialogue by using the similarity from topics that the user is interested in, to topic candidates to be presented as the next topic. A binary flag of user interest (=focus) can be predicted for each presented topic by using the existence of user focus in user utterances. Only vectors of flagged documents are used to select the next presenting document. If TP module is called by the dialogue manager, the TP module selects the next dialogue topic according to the similarity. The system calculates a vector of a document with word2vec (Mikolov et al., 2013) and the long-short term memory neural network (LSTM)-based encoder-decoder model (Cho et al., 2014) to define the similarity. Each word of new text is embedded into a fixed-length distributed representation with word2vec, and inputted into the LSTM-based encoder model. The LSTM-based encoder-decoder model is trained to restore the original sentences of news texts. The final vector of the encoding phase in this architecture is used as the vector of the document. News texts collected from a Kyoto prefecture sightseeing Web site³ was used as the training data of word2vec, LSTM-based encoder-decoder model, and content of the dialogue system. We used cosine similarity between the vectors of documents to calculate the similarities of each topic (news texts). The overall architecture is illustrated in Figure 1.

3 Dialogue example (translation from Japanese)

In the following example, the system starts a dialogue from the presentation of the system about some interesting places in Kyoto. In **User 2**, the user requires an overall description of the presented topic, however, the user does not have any specified focus word that indicates his inter-

³<http://www.kyototourism.org/en/>

est. Thus, the system drastically changes the topic when the system is requested to change the topic (**System 3**). In **System 4**, the user shows an interest with a focus on the word “matcha” and the focus detection module of the system detects that the user is interested in this current topic. The vector of the second topic (document) is used to select the third topic in **System 5**, and the selected topic is similar to the second topic (about “matcha”).

Dialogue example

System 0 (GR): Hello, I would like to introduce you to some interesting places in Kyoto. Let me know if you are interested.

System 1 (TP): There is a cherry blossom illumination event at the Kamo-gawa river.

User 2 (ST): Could you give me the details?

System 2 (ST): We have an illumination event at the banks of the Kamo-gawa river from Oike to Gojo, which is called the “flower corridor”.

User 3 (TP): Thanks, do you have any other topics?

System 3 (TP): There is a tea ceremony in Kyoto.

User 4 (QA): Does it involve matcha (traditional style green tea)?

System 4 (QA): Shoka-do added a new green-tea course in a quality tea room.

User 5 (TP): Interesting. Any other news?

System 5 (TP): There is a long line to wait for a new matcha parfait.

...

Acknowledgement

Part of this work is supported by Japan Science and Technology Agency PRESTO (JPMJPR165B). Part of this work have been achieved by “Research and Development on Fundamental and Utilization Technologies for Social Big Data”, the Commissioned Research of National Institute of Information and Communications Technology (NICT), JAPAN.

References

Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David R Traum. 2012. Iso 24617-2: A semantically-based standard for dialogue annotation. In *In Proc. 8th edition of the Lan-*

- guage Resources and Evaluation Conference*. Cite-seer, pages 430–437.
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *In Proc. The 2014 Conference on Empirical Methods on Natural Language Processing*. pages 1724–1734.
- Akinobu Lee and Tatsuya Kawahara. 2009. Recent development of open-source speech recognition engine julius. In *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*. Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee, pages 131–137.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. pages 3111–3119.
- Koichiro Yoshino and Tatsuya Kawahara. 2014. Information navigation system based on pomdp that tracks user focus. In *In Proc. Annual SIGdial Meeting on Discourse and Dialogue*. pages 32–40.
- Koichiro Yoshino and Tatsuya Kawahara. 2015. Conversational system for information navigation based on POMDP with user focus tracking. *Computer Speech & Language* 34(1):275 – 291.
- Steve Young, Milica Gašić, Simon Keizer, François Mairesse, Jost Schatzmann, Blaise Thomson, and Kai Yu. 2010. The hidden information state model: A practical framework for pomdp-based spoken dialogue management. *Computer Speech & Language* 24(2):150–174.