Information Navigation System Based on POMDP that Tracks User Focus

Koichiro Yoshino   **Tatsuya Kawahara**

School of Informatics, Kyoto University

Sakyo-ku, Kyoto, 606-8501, Japan

`yoshino@ar.media.kyoto-u.ac.jp`

Abstract

We present a spoken dialogue system for navigating information (such as news articles), and which can engage in small talk. At the core is a partially observable Markov decision process (POMDP), which tracks user's state and focus of attention. The input to the POMDP is provided by a spoken language understanding (SLU) component implemented with logistic regression (LR) and conditional random fields (CRFs). The POMDP selects one of six action classes; each action class is implemented with its own module.

1 Introduction

A large number of spoken dialogue systems have been investigated and many systems are deployed in the real world. Spoken dialogue applications that interact with a diversity of users are available on smart-phones. However, current applications are based on simple question answering and the system requires a clear query or a definite task goal. Therefore, next-generation dialogue systems should engage in casual interactions with users who do not have a clear intention or a task goal. Such systems include a sightseeing navigation system that uses tour guide books or documents in Wikipedia (Misu and Kawahara, 2010), and a news navigation system that introduces news articles updated day-by-day (Yoshino et al., 2011; Pan et al., 2012). In this paper, we develop an information navigation system that provides information even if the user request is not necessarily clear and there is not a matching document in the knowledge base. The user and the system converse on the current topic and the system provides potentially useful information for the user.

Dialogue management of this kind of systems was usually made in a heuristic manner and based on simple rules (Dahl et al., 1994; Bohus and Rudnicky, 2003). There is not a clear principle nor established methodology to design and implement casual conversation systems. In the past years, machine learning, particularly reinforcement learning, have been investigated for dialogue management. MDPs and POMDPs are now widely used to model and train dialogue managers (Levin et al., 2000; Williams and Young, 2007; Young et al., 2010; Yoshino et al., 2013b). However, the conventional scheme assumes that the task and dialogue goal can be clearly stated and readily encoded in the RL reward function. This is not true in casual conversation or when browsing information.

Some previous work has tackled with this problem. In a conversational chatting system (Shibata et al., 2014), users were asked to make evaluation at the end of each dialogue session, to define rewards for reinforcement learning. In a listening dialogue system (Meguro et al., 2010), levels of satisfaction were annotated in logs of dialogue sessions to train a discriminative model. These approaches require costly input from users or developers, who provide labels and evaluative judgments.

In this work, we present a framework in which reward is defined for the quality of system actions and also for encouraging long interactions, in contrast to the conventional framework. Moreover, user focus is tracked to make appropriate actions, which are more rewarded.

2 Conversational Information Navigation System

In natural human-human conversation, participants have topics they plan to talk about, and they progress through the dialogue in accordance with the topics (Schegloff and Sacks, 1973). We call this dialogue style "information navigation." An example is shown in **Figure 1**. First, the speaker
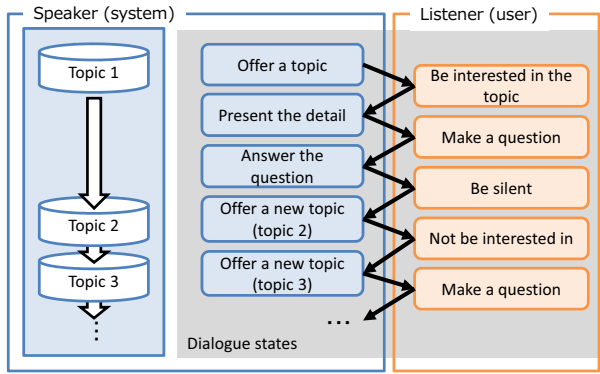
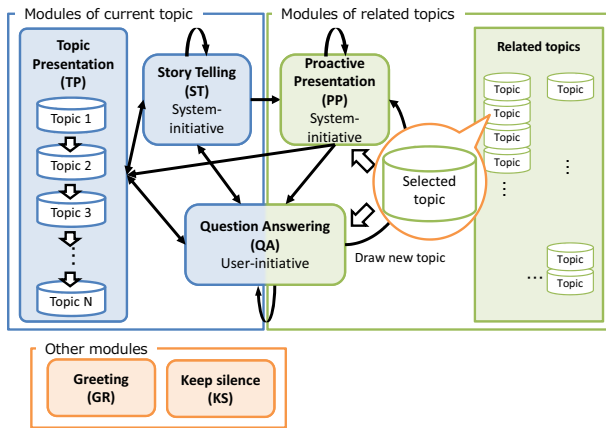Figure 1: An example of information navigation.

Figure 2: Overview of the information navigation system.

offers a new topic and probes the interest of the listener. If the listener shows interest, the speaker describes details of the topic. If the listener asks a specific question, the speaker answers the question. On the other hand, if the listener is not interested in the topic, the speaker avoids the details of that topic, and changes the topic. Topics are often taken from current news.

In our past work, we have developed a news navigation system (Yoshino et al., 2011) based on this dialogue structure. The system provides topics collected from Web news texts, and the user gets information according to his interests and queries.

2.1 System overview

An overview of the proposed system is depicted in **Figure 2**. The system has six modules, each of which implements a class of actions. Each module takes as input a recognized user utterance, an analyzed predicate-argument (P-A) structure and the detected user focus.

The system begins dialogues by selecting the "topic presentation (TP)" module, which presents a new topic selected from a news article. The system chooses the next module based on the user's response. In our task, the system assumes that each news article corresponds to a single topic, and the system presents a headline of news in the TP module. If the user shows interest (positive response) in the topic without any specific questions, the system selects the "story telling (ST)" module to give details of the news. In the ST module, the system provides a summary of the news article by using lead sentences. The system can also provide related topics with the "proactive presentation (PP)" module. This module is invoked by system initiative; this module is not invoked by any user request. If the user makes a specific question regarding the topic, the system switches to the "question answering (QA)" module to answer the question. This module answers questions on the presented topic and related topics.

The modules of PP and QA are based on a dialogue framework which uses the similarity of P-A structures (Yoshino et al., 2011). This framework defines the similarity of P-A structures between user queries and news articles, and retrieves or recommends the appropriate sentence from the news articles. This method searches for appropriate information from automatically parsed documents by referring to domain knowledge that is automatically extracted from domain corpus.

Transitions between the modules are allowed as shown in Figure 2. The modules "greeting (GR)" and "keep silence (KS)" are also implemented. GR module generates fixed greeting patterns by using regular expression matching. In terms of dialogue flow, these modules can be used at any time.

2.2 User focus in information navigation

"Focus" in discourse is "attentional state (that) contains information about the objects, properties, relations, and discourse intentions that are most salient at any given point." (Grosz and Sidner, 1986). The user has specific attention to an object if the user utterance contains the focus. In this work, we define the user focus as "the main piece of information of interest to the user." It makes a central component when making a reply or selecting relevant topics at the current dialogue state. For example, given "Did Ichiro perform bril-

```
SYS, tp, "Matsui, a part-time coach for the Giants, demonstrated 5 homeruns. The Giants president
          told him "Come back as a player!""
USR, st, 0, " Give me the detail?", "Give:me/0 the:detail/0"
SYS, st, "Matsui joined the extra batting training in the Giants' camp, and demonstrated his batting
          to young players."
USR, qa, 1, "Will Matsui become the Giants' manager in the future?", "Matsui/0 will:become/0
          the Giants' manager/1 in the future/0"
SYS, qa, "The director of Giants told him "Come back as a player!""
USR, tp, 0, "Next news, please.", "Next:news/0 please/0"
SYS, tp, " To be a ace pitcher, has Fujinami improved from the rookie year?
...
```

Figure 3: An example of annotation for collected dialogue. System utterances have a tuple of three elements separated by a comma: speaker, called module and utterance. User utterances have a tuple of four elements: speaker, the module the user request falls in, binary information of user focus, utterance and user focus annotation on each phrase or P-A element. (This example is translated from Japanese)

liantly?," user focus is "Ichiro" because the system reply should include information on Ichiro. This information is annotated on content words or named entities in a user utterance. In the POMDP, decisions are made based on whether any user focus was detected in the user's utterance.

3 Spoken Language Understanding (SLU)

In this section, we present the spoken language understanding components of our system. It detects the user's focus and intention and provides these to the dialogue manager. These spoken language understanding modules are formulated with a statistical model to give likelihoods which are used in POMDP.

3.1 Dialogue data

We collected 606 utterances (from 10 users) with a rule-based dialogue system (Yoshino et al., 2011). We annotated two kinds of tags: user intention (6 tags defined in Section 3.3), and focus information defined in Section 2.2. An example of annotation is shown in **Figure 3**. We highlighted annotation points in the bold font.

To prepare the training data, each utterance was labeled with one of the six modules, indicating the best module to respond. In addition, each phrase or P-A elements is labeled to indicated whether it is the user's focus or not. The user focus is determined by the attributes (=specifications of words in the domain) and preference order of phrases to identify the most appropriate information that the user wants to know. For example, in the second user utterance in Figure 3, the user's focus is the phrase "the Giants' manager". These tags are annotated by one person.

3.2 User focus detection based on CRF

To detect the user focus, we use a conditional random field (CRF) [1]. The problem is defined as a sequential labeling of the focus labels to a sequence of the phrases of the user utterance. Features used are shown in the **Table 1**. ORDER features are the order of the phrase in the sequence and in the P-A structure. We incorporate these features because the user focus often appears in the first phrase of the user utterance. POS features are part-of-speech (POS) tags and their pairs in the phrase. P-A features are semantic role of the P-A structure. We also incorporate the domain-dependent predicate-argument (P-A) scores that are defined with an unsupervised method (Yoshino et al., 2011). The score is discretized to 0.01, 0.02, 0.05, 0.1, 0.2, 0.5.

Table 2 shows the accuracy of user focus detection, which was conducted via five-fold cross-validation. "Phrase" is phrase-base accuracy and "sentence" indicates whether the presence of any user focus phrase was correctly detected (or not), regardless of whether the correct phrase was identified. This table indicates that WORD features are effective for detecting the user focus, but they are not essential for in the sentence-level accuracy. In this paper, we aim for portability across domains; therefore the dialogue manager only uses the sentence-level feature, so in our system we do not user the WORD features.

3.3 User intention analysis based on LR

The module classifies the user intention from the user utterance. We define six intentions as below.

- *TP*: request to the TP module.

---
[1] CRFsuite (Okazaki, 2007).

34

Table 1: Features of user focus detection.

feature type	feature
ORDER	Rank in a sequence of phrases
	Rank in a sequence of elements of P-A
POS	POS tags in the phrase
	POS tag sequence
POSORDER	Pair of POS tag and its order in the phrase
P-A	Which semantic role the phrase has
	Which semantic roles exist on the utterance
P-AORDER	Pair of semantic role and its order in the utterance
P-A score	P-A templates score
WORD	Words in the phrase
	Pair of words in the phrase
	Pair of word and its order in the phrase

Table 2: Accuracy of user focus detection.

	Accuracy
phrase	86.7%
phrase + (WORD)	90.3%
sentence (focus exist or not)	99.8%
sentence (focus exist or not) + (WORD)	99.8%

- *ST*: request to the ST module.

- *QA*: request to the QA module.

- *GR*: greeting to the GR module.

- *NR*: silence longer than a threshold.

- *II*: irrelevant input due to ASR errors or noise.

We adopt logistic regression (LR)-based dialogue act tagging approach (Tur et al., 2006). The probability of user intention o given an ASR result of the user utterance h is defined as,

$$P(o|h) = \frac{\exp(\omega \cdot \phi(h,o))}{\Sigma_o \exp(\omega \cdot \phi(h,o))}. \quad (1)$$

Here, ω is a vector of feature weights and $\phi(h,o)$ is a feature vector. We use POS, P-A and P-A templates score as a feature set. In addition, we add a typical expression feature (TYPICAL) to classify *TP*, *ST* or *GR* tags. For example, typical expressions in conversation are "Hello" or "Go on," and those in information navigation are "News of the day" or "Tell me in detail." Features for the classifier are shown in the **Table 3**.

The accuracy of the classification in five-fold cross-validation is shown in **Table 4**. The TYP-

Table 3: Features of user intention analysis.

feature type	feature
POS	Bag of POS tags
	Bag of POS bi-gram
P-A	Bag of semantic role labels
	Bag of semantic role labels bi-gram
	Pair of semantic role label and its rank
P-A score	P-A templates score
TYPICAL	Occurrence of typical expressions

Table 4: Accuracy of user intention analysis.

	All features	without TYPICAL
TP	100%	100%
ST	75.3%	64.2%
QA	94.1%	93.5%
GR	100%	100%
II	16.7%	16.7%
All	92.1%	90.2%

ICAL feature improves the classification accuracy while keeping the domain portability.

3.4 SLU for ASR output

ASR and intention analysis involves errors. Here, s is a true user intention and o is an observed intention. The observation model $P(o|s)$ is given by the likelihood of ASR result $P(h|u)$ (Komatani and Kawahara, 2000) and the likelihood of the intention analysis $P(o|h)$,

$$P(o|s) = \sum_h P(o,h|s) \quad (2)$$

$$\approx \sum_h P(o|h)P(h|u). \quad (3)$$

Here, u is an utterance of the user. We combine the N-best ($N = 5$) hypotheses of the ASR result h.

4 Dialogue Management for Information Navigation

The conventional dialogue management for task-oriented dialogue systems is designed to reach a task goal as soon as possible (Williams and Young, 2007). In contrast, information navigation does not always have a clear goal, and the aim of information navigation is to provide as much relevant information as the user is interested in. Therefore, our dialogue manager refers user involvement or engagement (=level of interest) and the user focus

(=object of interest). This section describes the general dialogue management based on POMDP, and then gives an explanation of the proposed dialogue management using the user focus.

4.1 Dialogue management based on POMDP

The POMDP-based statistical dialogue management is formulated as below. The random variables involved at a dialogue turn t are as follows:

- $s \in I_s$: user state
  User intention.

- $a \in K$: system action
  Module that the system selects.

- $o \in I_s$: observation
  Observed user state, including ASR and intention analysis errors.

- $b_{s_i} = P(s_i|o^{1:t})$: belief
  Stochastic variable of the user state.

- π: policy function
  This function determines a system action a given a belief of user b. π^* is the optimal policy function that is acquired by the training.

- r: reward function
  This function gives a reward to a pair of the user state s and the system action a.

The aim of the statistical dialogue management is to output an optimal system action $\hat{a}^t$ given a sequence of observation $o^{1:t}$ from 1 to t time-steps.

Next, we give the belief update that includes the observation and state transition function. The belief update of user state s_i in time-step t is defined as,

$$b_{s'_j}^{t+1} \propto \underbrace{P(o^{t+1}|s'_j)}_{\textbf{Obs.}} \sum_{s_i} \underbrace{P(s'_j|s_i, \hat{a}_k)}_{\textbf{Trans.}} b_{s_i}^t. \quad (4)$$

Obs. is an observation function which is defined in Equation (3) and **Trans.** is a state transition probability of the user state. Once the system estimates the belief $b_{s_i}^t$, the policy function outputs the optimal action $\hat{a}$ as follows:

$$\hat{a} = \pi^*(b^t). \quad (5)$$

4.2 Training of POMDP

We applied Q-learning (Monahan, 1982; Watkins and Dayan, 1992) to acquire the optimal policy π^*. Q-learning relies on the estimation of a Q-function, which maximizes the discounted sum of future rewards of the system action a^t at a dialogue

turn t given the current belief b^t. Q-learning is performed by iterative updates on the training dialogue data:

$$Q(b^t, a^t) \Leftarrow (1 - \varepsilon)Q(b^t, a^t)$$
$$+ \varepsilon[R(s^t, a^t) + \gamma \max_{a^{t+1}} Q(b^{t+1}, a^{t+1})], \quad (6)$$

where ε is a learning rate, γ is a discount factor of a future reward. We experimentally decided $\varepsilon = 0.01$ and $\gamma = 0.9$. The optimal policy given by the Q-function is determined as,

$$\pi^*(b^t) = \underset{a^t}{\operatorname{argmax}} Q(b^t, a^t). \quad (7)$$

However, it is impossible to calculate the Q-function for all possible real values of belief b. Thus, we train a limited Q-function given by a Grid-based Value Iteration (Bonet, 2002). The belief is given by a function,

$$b_{s_i} = \begin{cases} \eta & \textbf{if } s = i \\ \frac{1-\eta}{|I_s|} & \textbf{if } s \neq i \end{cases}. \quad (8)$$

Here, η is a likelihood of $s = i$ that is output of the intention analyzer, and we selected 11 discrete points from 0.0 to 1.0 by 0.1. We also added the case of uniform distribution. The observation function of the belief update is also given in a similar manner.

4.3 Dialogue management using user focus

Our POMDP-based dialogue management chooses actions based on its belief in: the user intention s and the user focus f (0 or $1 \in J_f$). The observation o is controlled by hidden states f and s that are decided by the state transition probabilities,

$$P(f^{t+1}|f^t, s^t, a^t), \quad (9)$$
$$P(s^{t+1}|f^{t+1}, f^t, s^t, a^t). \quad (10)$$

We constructed a user simulator by using the annotated data described in Section 3.1.

Equation (10) is also used for the state transition probability of the belief update. The equation of the belief update (4) is extended by introducing the previous user focus f_l and current user focus f'_m information,

$$b_{s'_j}^{t+1} = \underbrace{P(o^{t+1}|s'_j)}_{\textbf{Obs.}}$$
$$\times \sum_i \underbrace{P(s'_j|f'_m, f_l, s_i, \hat{a}_k)}_{\textbf{Trans.}} b_{s_i, f_l}^t. \quad (11)$$

Table 5: Rewards in each turn.

state s	focus f	action a					
		TP	ST	QA	PP	GR	KS
TP	0	+10	-10	-10	-10	-10	-10
	1						
ST	0	-10	+10	-10	0	-10	-10
	1						
QA	0	-10	+10	+10	-10	-10	-10
	1		-10	+30	+10		
GR	0	-10	-10	-10	-10	+10	-10
	1						
NR	0	+10	-10	-10	-10	-10	0
	1	-10			+10		
II	0	-10	-10	-10	-10	-10	+10
	1						

Figure 4: Effect of introduction of the user focus in simulation.

The resultant optimal policy is,

$$\hat{a} = \pi^*(b^t, f_l). \tag{12}$$

4.4 Definition of rewards

Table 5 defines a reward list at the end of a each turn. The reward of +10 is given to appropriate actions, 0 to acceptable actions, and -10 to inappropriate actions.

In Table 5, pairs of a state and its apparently corresponding action, *TP* and TP, *ST* and ST, *QA* and QA, *GR* and GR, and *II* and KS, have positive rewards. Rewards in bold fonts (**+10**) are defined for the following reasons. If the user asks a question (*QA*) without a focus (e.g. "What happened on the game?"), the system can continue by story telling (ST). But when the question has a focus, the system should answer the question (QA), which is highly rewarded (**+30**). If the system cannot find an answer, it can present relevant information (PP). When the user says nothing (*NR*), the system action should be decided by considering the user focus; present a new topic if the user is not interested in the current topic (*f*=0) or present an article related to the dialogue history (*f*=1).

Reward of +200 is given if 20 turns are passed, to reward a long continued dialogue. The user simulator terminates the dialogue if the system selects an inappropriate action (action of $r = -10$) five times, and a large penalty -200 is given to the system.

5 Evaluations of Dialogue

We evaluated the proposed system with two experiments; dialogue state tracking with real users and average reward with a user simulator. For the evaluation, we collected an additional 312 utterances (8 users, 24 dialogues) with the proposed dialogue system.
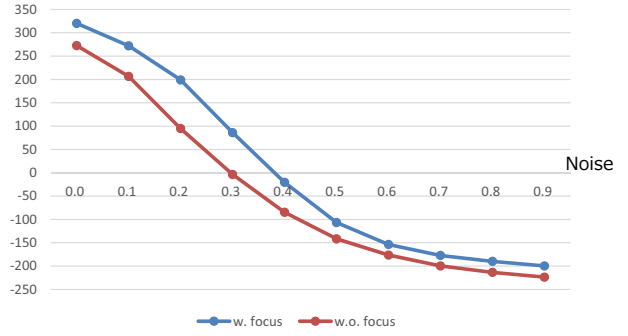
5.1 Evaluation of dialogue manager with user simulator

First, we evaluated the dialogue manager with user simulation that is constructed from the training corpus (Section 3.1). In this evaluation, the system calculated average reward of 100,000 dialogue sessions between the system and the user simulator given a fixed noise rate. **Figure 4** shows the effect of the user focus. By introducing the user focus, the system receives higher rewards than the model without the user focus. Especially, the proposed model is more robust with a noise level of 10–30% that spoken dialogue systems often encounter (Yoshino et al., 2013a).

5.2 Evaluation of dialogue state tracking with real users

Dialogue state tracking (DST) is a task of tracking the correct user state with a noisy input (e.g. ASR and NLU errors) (Williams et al., 2013). It tries to maximize the probability of the belief of the correct states, but we evaluated the accuracy of the 1-best result of the belief update. We also evaluated the average reward of each dialogue. The baseline system is a rule-based dialogue manager which we previously implemented. The baseline system is operated by a score of the question-answering module using predicate-argument structures (Yoshino et al., 2011) and regular expressions for TP and GR modules.

The DST accuracy and average reward are shown in **Table 6**. This result shows that the proposed method tracks the dialogue state of the user with a high accuracy. Our statistical framework

Dialogue 1

Dialogue	State info
Sys: Hello. Usr: Hello. Tell me today's news. Sys: Tanaka showed a crazy split. His colleague Sabathia was stunned.	$s=TP$, $P(o\|s)$=0.8, $b(s=TP)$=0.800, f=0 a=TP, $b(s=TP\ f=0)$=0.8
Usr: What is the crazy split? Sys: His split comes just like a four-seam, but it drops quickly in front of the home plate.	$s=QA$, $P(o\|s)$=0.4, $b(s=QA)$=0.934, f=1 a=QA, $b(s=QA,\ f=1)$=0.9
Usr: Where was he training? Sys: Tanaka practiced at the bullpen in the New York Yankees' camp on Feb. 20th.	$s=QA$, $P(o\|s)$=0.8, $b(s=QA)$=0.973, f=1 a=QA, $b(s=QA,\ f=1)$=1.0
Usr: (pause) Sys: By the way, he also practiced a drop two-seam on Feb. 18.	$s=NR$, $P(o\|s)$=1.0, $b(s=NR)$=1.00, f=1 a=PP, $b(s=NR,\ f=1)$=1.0
Usr: Does his two-seam drop? Sys: The catcher said "Tanaka's two-seam not only slides, but also drops."	$s=QA$, $P(o\|s)$=0.6, $b(s=QA)$=0.867, f=1 a=QA, $b(s=QA,\ f=1)$=0.9
Usr: OK, any other news? Sys: Yankees' Ichiro is training strenuously!	$s=TP$, $P(o\|s)$=0.8, $b(s=TP)$=0.984, f=0 a=TP, $b(s=TP,\ f=0)$=1.0
Usr: Tell me more. Sys: Ichiro came to the field an hour earlier than usual and trained strenuously. …	$s=ST$, $P(o\|s)$=0.9, $b(s=ST)$=0.871, f=0 a=ST, $b(s=ST,\ f=0)$=0.9

Dialogue 2

Dialogue	State info
Sys: Hello. Usr: Hello. Tell me today's news? Sys: Matsui, a part-time coach for the Giants, demonstrated 5 homeruns. The Giants president told him "Come back as a player!"	$s=TP$, $P(o\|s)$=0.8, $b(s=TP)$=0.800, f=0 a=TP, $b(s=TP\ f=0)$=0.8
Usr: What happened? Sys: The godzilla cannon came back to the Giants' spring camp in Miyazaki.	$s=QA$, $P(o\|s)$=0.8, $b(s=QA)$=0.532, f=0 a=ST, $b(s=QA,\ f=0)$=0.5
Usr: What is the Godzilla cannon? Sys: The Godzilla coach demonstrated his home-run batting to young players.	$s=QA$, $P(o\|s)$=0.8, $b(s=QA)$=0.806, f=1 a=QA, $b(s=QA,\ f=1)$=0.8
Usr: Next please. Sys: To be a ace pitcher, has Fujinami improved from the rookie year? …	$s=TP$, $P(o\|s)$=0.8, $b(s=TP)$=0.986, f=0 a=TP, $b(s=TP,\ f=0)$=1.0

Figure 5: A dialogue example. (This example is translated from Japanese)

Table 6: Accuracy of dialogue state tracking.

	rule	focus POMDP
Accuracy of tracking (1-best)	0.561 (=175/312)	0.869 (=271/312)
Average reward	-22.9	188.6

improved SLU accuracy and robustness against ASR errors, especially reducing confusions between question answering (*QA*) and topic presentation (*TP*). Moreover, belief update can detect the *TP* state even if the SLU incorrectly predicts *QA* or *ST*.

5.3 Discussion of trained policy

An example dialogue is shown in **Figure 5**. In the example, the system selects appropriate ac-tions even if the observation likelihood is low. At the 4th turn of Dialogue 1 in this example, the system with the user focus responds with an action of proactive presentation a=PP, but the system without the user focus responds with an action of topic presentation a=TP. At the 2nd turn of Dialogue 2, the user asks a question without a focus. The confidence of $s=QA$ is lowered by the belief update, and the system selects the story telling module a=ST. These examples show that the training result (=learned policy) reflects our design described in Section 4.4: It is better to make a proactive presentation when the user is interested in the topic.

6 Conclusions

We constructed a spoken dialogue system for information navigation of Web news articles updated day-by-day. The system presents relevant infor-

mation according to the user's interest, by tracking the user focus. We introduce the user focus detection model, and developed a POMDP framework which tracks user focus to select the appropriate action class (module) of the dialogue system. In experimental evaluations, the proposed dialogue management approach determines the state of the user more accurately than the existing system based on rules. An evaluation with a user simulator shows that including user focus in the dialogue manager's belief state improves robustness to ASR/SLU errors.

In future work, we plan to evaluate the system with a large number of real users on a variety of domains, and optimize the reward function for the information navigation task.

Acknowledgments

References

Dan Bohus and Alexander I. Rudnicky. 2003. Ravenclaw: Dialog management using hierarchical task decomposition and an expectation agenda. In *Proceedings of the 8th European Conference on Speech Communication and Technology*, pages 597–600.

Blai Bonet. 2002. An e-optimal grid-based algorithm for partially observable Markov decision processes. In *Proceedings of International Conference on Machine Learning*, pages 51–58.

Deborah A. Dahl, Madeleine Bates, Michael Brown, William Fisher, Kate Hunicke-Smith, David Pallett, Christine Pao, Alexander Rudnicky, and Elizabeth Shriberg. 1994. Expanding the scope of the ATIS task: the ATIS-3 corpus. In *Proceedings of the workshop on Human Language Technology*, pages 43–48.

Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Ryuichiro Higashinaka, Katsuhito Sudoh, and Mikio Nakano. 2006. Incorporating discourse features into confidence scoring of intention recognition results in spoken dialogue systems. *Speech Communication*, 48(3):417–436.

Tatsuya Kawahara. 2009. New perspectives on spoken language understanding: Does machine need to fully understand speech? In *Proceedings of IEEE workshop on Automatic Speech Recognition and Understanding*, pages 46–50.

Kazunori Komatani and Tatsuya Kawahara. 2000. Flexible mixed-initiative dialogue management using concept-level confidence measures of speech recognizer output. In *Proceedings of the 18th conference on Computational linguistics*, pages 467–473.

Esther Levin, Roberto Pieraccini, and Wieland Eckert. 2000. A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Transactions on Speech and Audio Processing*, 8(1):11–23.

Toyomi Meguro, Ryuichiro Higashinaka, Yasuhiro Minami, and Kohji Dohsaka. 2010. Controlling listening-oriented dialogue using partially observable markov decision processes. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 761–769.

Teruhisa Misu and Tatsuya Kawahara. 2010. Bayes risk-based dialogue management for document retrieval system with speech interface. *Speech Communication*, 52(1):61–71.

George E. Monahan. 1982. State of the art? a survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1):1–16.

Naoaki Okazaki. 2007. CRFsuite: a fast implementation of Conditional Random Fields (CRFs).

Yi-Cheng Pan, Hung yi Lee, and Lin shan Lee. 2012. Interactive spoken document retrieval with suggested key terms ranked by a markov decision process. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(2):632–645.

Emanuel A. Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica*, 8(4):289–327.

Tomohide Shibata, Yusuke Egashira, and Sadao Kurohashi. 2014. Chat-like conversational system based on selection of reply generating module with reinforcement learning. In *Proceedings of the 5th International Workshop Series on Spoken Dialog Systems*.

Gokhan Tur, Umit Guz, and Dilek Hakkani-Tur. 2006. Model adaptation for dialog act tagging. In *Proceedings of IEEE workshop on Spoken Language Technology*, pages 94–97. IEEE.

Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning*, 8(3):279–292.

Jason D. Williams and Steve Young. 2007. Partially observable Markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422.

Jason D. Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. 2013. The dialog state tracking challenge. In *Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 404–413.

Koichiro Yoshino, Shinsuke Mori, and Tatsuya Kawahara. 2011. Spoken dialogue system based on information extraction using similarity of predicate argument structures. In *Proceedings of the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 59–66.

Koichiro Yoshino, Shinsuke Mori, and Tatsuya Kawahara. 2013a. Incorporating semantic information to selection of web texts for language model of spoken dialogue system. In *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, pages 8252–8256.

Koichiro Yoshino, Shinji Watanabe, Jonathan Le Roux, and John R. Hershey. 2013b. Statistical dialogue management using intention dependency graph. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pages 962–966.

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.