Controlling Listening-oriented Dialogue using Partially Observable Markov Decision Processes

Toyomi Meguro†, **Ryuichiro Higashinaka**‡, **Yasuhiro Minami**†, **Kohji Dohsaka**†
†NTT Communication Science Laboratories, NTT Corporation
‡NTT Cyber Space Laboratories, NTT Corporation
meguro@cslab.kecl.ntt.co.jp
higashinaka.ryuichiro@lab.ntt.co.jp
{minami,dohsaka}@cslab.kecl.ntt.co.jp

Abstract

This paper investigates how to automatically create a dialogue control component of a listening agent to reduce the current high cost of manually creating such components. We collected a large number of listening-oriented dialogues with their user satisfaction ratings and used them to create a dialogue control component using partially observable Markov decision processes (POMDPs), which can learn a policy to satisfy users by automatically finding a reasonable reward function. A comparison between our POMDP-based component and other similarly motivated systems using human subjects revealed that POMDPs can satisfactorily produce a dialogue control component that can achieve reasonable subjective assessment.

1 Introduction

Although task-oriented dialogue systems have been actively researched (Hirshman, 1989; Ferguson et al., 1996; Nakano et al., 1999; Walker et al., 2002), recently non-task-oriented functions are starting to attract attention, and systems without a specific task that deal with more casual dialogues, such as chats, are being actively investigated from their social and entertainment aspects (Bickmore and Cassell, 2001; Higashinaka et al., 2008; Higuchi et al., 2008).

In the same vein, we have been working on listening-oriented dialogues in which one conversational participant attentively listens to the other (hereafter, listening-oriented dialogue). Our aim is to build listening agents that can implement such a listening process so that users can satisfy their desire to speak and be heard. Figure 1 shows an excerpt from a typical listening-oriented dialogue. In the literature, dialogue control components for less (or non-) task-oriented dialogue systems, such as listening agents, have typically used hand-crafted rules for dialogue control, which can be problematic because completely covering all dialogue states by hand-crafted rules is difficult when the dialogue has fewer task restrictions (Wallace, 2004; Isomura et al., 2009).

To solve this problem, this paper aims to automatically build a dialogue control component of a listening agent using partially observable Markov decision processes (POMDPs). POMDPs, which make it possible to learn a policy that can maximize the averaged reward in partially observable environments (Pineau et al., 2003), have been successfully adopted in task-oriented dialogue systems for learning a dialogue control module from data (Williams and Young, 2007). However, no work has attempted to use POMDPs for less (or non-) task-oriented dialogue systems, such as listening agents, because user goals are not as well-defined as task-oriented ones, complicating the finding of a reasonable reward function.

We apply POMDPs to listening-oriented dialogues by having the system learn a policy that simultaneously maximizes how well users feel that they are being listened to (hereafter, user satisfaction) and how smoothly the system generates dialogues (hereafter, smoothness). This formulation is new; no work has considered both user satisfaction and smoothness using POMDPs. We collected a large amount of listening-oriented dialogues and annotated them with dialogue acts and also obtained subjective evaluation results for them. From them, we calculated the rewards and learned the POMDP policies. We evaluated the dialogue-act tag sequences of our POMDPs using human subjects.

Utterance	Dialogue act
S: Good evening.	Greeting
The topic is “food,” nice to meet you.	Greeting
L: Nice to meet you, too.	Greeting
S: I had curry for dinner.	S-Disc (sub: fact)
Do you like curry?	Question (sub: pref)
L: Yes, I do.	Sympathy
S: Really?	Repeat
Me, too.	Sympathy
L: Do you usually go out to eat?	Question (sub: habit)
S: No, I always cook at home.	S-Disc (sub: habit)
I don’t use any special spices, but I sometimes cook noodles using soup and curry.	S-Disc (sub: habit)
L: That sounds good!	S-Disc (sub: pref (positive))

Figure 1: Excerpt of a typical listening-oriented dialogue. Dialogue topic is “food.” Dialogue acts corresponding to utterances are shown in parentheses (See Table 1 for meanings): S-Disc stands for Self-disclosure; pref for preference; S for speaker; and L for listener. The dialogue was translated from Japanese by the authors.

The next section introduces related work. Section 3 describes our approach. Section 4 describes our collection of listening-oriented dialogues. This is followed in Section 5 by an evaluation experiment that compared our POMDP-based dialogue control with other similarly motivated systems. The last section summarizes the main points and mentions future work.

2 Related work

With increased attention on social dialogues and senior peer counseling, work continues to emerge on listening-oriented dialogues. One early work is (Maatman et al., 2005), which showed that virtual agents can give users the sense of being heard using such gestures as nodding and head shaking. Recently, Meguro et al. (2009a) analyzed the characteristics of listening-oriented dialogues. They compared listening-oriented dialogues and casual conversations between humans, revealing that the two types of dialogues have significantly different flows and that listeners actively question with frequently inserted self-disclosures; the speaker utterances were mostly concerned with self-disclosure.

Shitaoka et al. (2010) also investigated the functions of listening agents, focusing on their response generation components. Their system takes the confidence score of speech recognition into account and changes the system response accordingly; it repeats the user utterance or makes an empathic utterance for high-confidence user utterances and makes a backchannel when the confidence is low. The system’s empathic utterances can be “I’m happy” or “That’s too bad,” depending on whether a positive or negative expression is included in the user utterances. Their system’s response generation only uses the speech recognition confidence and the polarity of user utterances as cues to choose its actions. Currently, it does not consider the utterance content or the user intention.

In order for listening agents to achieve high smoothness, a switching mechanism between the “active listening mode,” in which the system is a listener, and the “topic presenting mode,” in which the system is a speaker, has been proposed (Yokoyama et al., 2010; Kobayashi et al., 2010). Here, the system uses a heuristic function to maintain a high user interest level and to keep the system in an active listening mode. Dialogue control is done by hand-crafted rules. Our motivation bears some similarity to theirs in that we want to build a listening agent that gives users a sense of being heard; however, we want to automatically make such an agent from dialogue data.

POMDPs have been introduced for robot action control (Pineau et al., 2003). Here, the system learns to make suitable movements for completing a certain task. Over the years, POMDPs have been actively studied for applications to spoken dialogue systems. Williams et al. (2007) successfully used a POMDP for dialogue control in a ticket-buying domain in which the objective was to fix the departure and arrival places for tickets. Recent work on POMDPs indicates that it is possible to train a dialogue control module in task-oriented dialogues when the user goal is obvious. In contrast, in this paper, we aim to verify whether POMDPs can be applied to less task-oriented dialogues (i.e., listening-oriented dialogues) where user goals are not as obvious.

In a recent study, Minami et al. (2009) applied POMDPs to non-task-oriented man-machine interaction. Their system learned suitable action control of agents that can act smoothly by obtaining rewards from the statistics of artificially generated data. Our work is different because we use real human-human dialogue data to train POMDPs for dialogue control in listening-oriented dialogues.

3 Approach

A typical dialogue system has utterance understanding, dialogue control, and utterance generation modules. The utterance understanding module comprehends user natural-language utterances, whose output (i.e., a user dialogue act) is passed to the dialogue control module. The dialogue control module chooses the best system dialogue act at every dialogue point using the user dialogue act as input. The utterance generation module generates natural-language utterances and says them to users by realizing the system dialogue acts as surface forms.

This paper focuses on the dialogue control module of a listening agent. Since a listening-oriented dialogue has a characteristic conversation flow (Meguro et al., 2009a), focusing on this module is crucial because it deals with the dialogue flow. Our objective is to train from data a dialogue control module that achieves a smooth dialogue flow that makes users feel that they are being listened to attentively.

3.1 Dialogue control using POMDPs

The purpose of our dialogue control is to simultaneously create situations in which users feel listened to (i.e., user satisfaction) and to generate smooth action sequences (i.e., smoothness). To do this, we automatically and statistically train the reward and the policy of the POMDP using a large amount of listening-oriented dialogue data. POMDP is a reinforcement learning framework that can learn a policy to select an action sequence that maximizes average future rewards. Setting a reward is crucial in POMDPs.

For our purpose, we introduce two different rewards: one for user satisfaction and the other for smoothness. Before creating a POMDP structure, we used the dynamic Bayesian network (DBN) structure (Fig. 2) to obtain the statistical structure of the data and the two rewards.

The random values in the DBN are as follows: s_o and s_a are the dialogue state and action state, o is a speaker observation, a is a listener action, and d is a random variable for an evaluation score that indicates the degree of the user being listened to. This evaluation score can be obtained by questionnaires, and the variable is used for calculating a user satisfaction reward for the POMDP. The DBN arcs in Fig. 2 define the emission and transition probabilities. $\Pr(o'|s'_o)$ is the emission probability of o' given s'_o. $\Pr(d|s_o)$ is the emission probability of d given s_o. $\Pr(s'_o|s_o, a)$ is a transition probability from s_o to s'_o given a. The DBN is trained using the EM algorithm. Using the obtained variables, we calculate the two reward functions as follows:

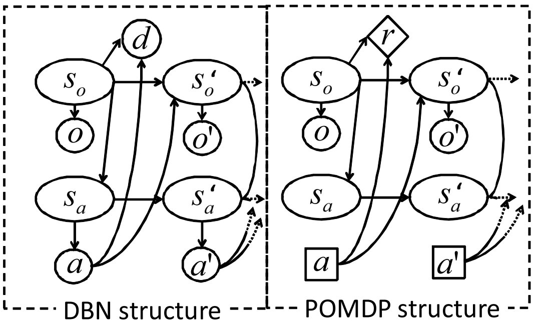

Figure 2: DBN and POMDP structures employed in this paper. Note that a in the POMDP is isolated from other states because it is decided by a learned policy.

(1) Reward for user satisfaction This reward is obtained from the d variable by

$$r_1((s_o, *), a) = \sum_{d=min}^{max} d \times \Pr(d|s_o, a),$$

where * is arbitrary s_a and min and max are minimum and maximum evaluation scores.

(2) Reward for smoothness For smoothness, we maximize the action predictive probability given the history of actions and observations. The probability is calculated from listening-oriented dialogue data. s_a is introduced for estimating the predictive probability of action a and for selecting a to maximize the predictive probability.

We set $\Pr(a|s_a) = 1$ when $a = s_a$ so that s_a corresponds one-on-one with a. Then, if $a_t = s_a$ at time t is given, we obtain

$$\begin{aligned}
&\Pr(a_t|o_1, a_1, \ldots, a_{t-1}, o_t) \\
&\quad = \sum_{s'_a} \Pr(a_t|s'_a) \Pr(s'_a|o_1, a_1, \ldots, a_{t-1}, o_t) \\
&\quad = \Pr(s_a|o_1, a_1, \ldots, o_{t-1}, a_{t-1}, o_t)
\end{aligned}$$

Consequently, maximizing the predictive probability of a equals maximizing that of s_a. If we set 1.0 to reward $r_2((*, s_a), a)$ when $s_a = a$, the POMDP will generate actions that maximize their predictive probabilities. We believe that this reward should increase the smoothness of a system action sequence since the sequence is generated according to the statistics of human-human dialogues.

Converting a DBN into a POMDP The DBN is converted into a POMDP (Fig. 2), while maintaining the transition and output probabilities. We convert d to r as described above.

The system is in a partially observed state. Since the state is not known exactly, we use a distribution called "belief state" b_t with which we obtain the average reward that will be gained in the future at time t by:

$$V_t = \sum_{\tau=0}^{\infty} \gamma^{\tau} \sum_{s} b_{\tau+t}((s_o, s_a)) r((s_o, s_a), a_{\tau+t}),$$

where τ is a discount factor; namely, the future reward is decreased by τ. A policy is learned by value iteration so that the action that maximizes V_t can be chosen. We define $r((s_o, s_a), a)$ as follows:

$$r((s_o, s_a), a) = r_1((s_o, *), a) + r_2((*, s_a), a).$$

By balancing these two rewards, we can choose an action that satisfies both user satisfaction and smoothness.

4 Data collection

We collected listening-oriented dialogues using human subjects who consisted of ten listeners (five males and five females) and 37 speakers (18 males and 19 females). The listeners and speakers ranged from 20 to 60 years old and were all native Japanese speakers. Listeners and speakers were matched to form a listener-speaker pair and communicated over the Internet with our chat interface. They used only text; they were not allowed to use voice, video, or facial expressions. The speakers chose their own listener and freely participated in dialogues from 7:00 pm to midnight for a period of 15 days. One conversation was restricted to about ten minutes. The subjects talked about a topic chosen by the speaker. There were 20 predefined topics: money, sports, TV and radio, news, fashion, pets, movies, music, housework and childcare, family, health, work, hobbies, food, human relationships, reading, shopping, beauty aids, travel, and miscellaneous. The listeners were instructed to make it easy for the speakers to say what the speakers wanted to say. We collected 1260 listening-oriented dialogues.

4.1 Dialogue-act annotation

We labeled the collected dialogues using the dialogue-act tag set shown in Table 1. We made these tags by selecting, extending, and modifying those from previous studies that concerned human listening behaviors in some way (Meguro et al., 2009a; Jurafsky et al., 1997; Ivey and Ivey, 2002). In our tag set, only question and self-disclosure tags have sub-category tags. Two annotators (not the authors) labeled each sentence of our collected dialogues using these 32 tags. In dialogue-act annotation, since there can be several sentences in one utterance, one annotator first split the utterances into sentences, and then both annotators labeled each sentence with a single dialogue act.

4.2 Obtaining evaluation scores

POMDPs need evaluation scores (i.e., d) for dialogue acts (i.e., a) for training a reward function. Therefore, we asked a third-party participant, who was neither a listener nor a speaker in our dialogue data collection, to evaluate the user satisfaction levels of the collected dialogues. She rated each dialogue in terms of how she would have felt "being heard" after the dialogue if she had been the speaker of the dialogue in question. She provided ratings on the 7-point Likert scale for each dialogue. Since she rated the whole dialogue with a single rating, we set the evaluation score of each action within a dialogue using the evaluation score for that dialogue.

We used a third-person's evaluation and not the original person's to avoid the fact that the evaluative criterion is too different between humans; identical evaluation scores from two people do not necessarily reflect identical user satisfaction levels. We highly valued the reliability and consistency of the third-person scores. This way, at least, we can train a policy that maximizes its average reward function for the rater, which we need to verify first before considering adaptation to two or more individuals.

5 Experiment

5.1 Experimental setup

The experiment followed three steps.

Greeting	Greeting and confirmation of dialogue theme. e.g., Hello. Let's talk about lunch.
Information	Delivery of objective information. e.g., My friend recommended a restaurant.
Self-disclosure	Disclosure of preferences and feelings.
sub: fact	e.g., I live in Tokyo.
sub: experience	e.g., I had a hamburger for lunch.
sub: habit	e.g., I always go out for dinner.
sub: preference (positive)	e.g., I like hamburgers.
sub: preference (negative)	e.g., I don't really like hamburgers.
sub: preference (neutral)	e.g., Its taste is near my homemade taste.
sub: desire	e.g., I want to try it.
sub: plan	e.g., I'm going there next week.
sub: other	
Acknowledgment	Encourage the conversational partner to speak. e.g., Well. Aha.
Question	Utterances that expect answers.
sub: information	e.g., Please tell me how to cook it.
sub: fact	e.g., What kind of curry?
sub: experience	e.g., What did you have for dinner?
sub: habit	e.g., Did you cook it yourself?
sub: preference	e.g., Do you like it?
sub: desire	e.g., Don't you want to eat rice?
sub: plan	e.g., What are you going to have for dinner?
sub: other	
Sympathy	Sympathetic utterances and praises. e.g., Me, too.
Non-sympathy	Negative utterances. e.g., Not really.
Confirmation	Confirm what the conversation partner said. e.g., Really?
Proposal	Encourage the partner to act. e.g., Try it.
Repeat	Repeat the partner's utterance.
Paraphrase	Paraphrase the partner's utterance.
Approval	Broach or show goodwill toward the partner. e.g., Absolutely!
Thanks	Express thanks e.g., Thank you.
Apology	Express regret e.g., I'm sorry.
Filler	Filler between utterances. e.g., Uh. Let me see.
Admiration	Express affection. e.g., Ha-ha.
Other	Other utterances.

Table 1: Definition and example of dialogue acts

In the first step, we created our POMDP system using our approach (See Section 3.1). We also made five other systems for comparison that we describe in Section 5.2. Each system outputs dialogue-act tag sequences for evaluation. The dialogue theme was "food" because it was the most frequent theme and accounted for 20% of our data (See Table 2 for the statistics); we trained our POMDP using the "food" dialogues. We restricted the dialogue topic to verify that our approach at least works with a small set. Since there is no established measure for automatically evaluating a dialogue-act tag sequence, we evaluated our dialogue control module using human subjective evaluations. However, this is very difficult to do because dialogue control modules only output dialogue acts, not natural language utterances.

	All	Food (subset of All)
# dialogues	1260	250
# words	479881	94867
# utterances per dialogue	28.2	29.1
# dialogues per listener	126	25
# dialogues per speaker	34	6.8
# dialogue acts	67801	13376
inter-annotator agreement	0.57	0.55

Table 2: Statistics of collected dialogues and dialogue-act annotation. Inter-annotator agreement means agreement of dialogue-act annotation using Cohen's κ.

In the second step, we recruited participants who created natural language utterances from dialogue-act tag sequences. In their creating dialogues, we provided them with situations to stimulate their imaginations. Table 3 shows the situations, which were deemed common in everyday Japanese life; we let the participants create utterances that fit the situations. These situations were necessary because, without restrictions, the evaluation scores could be influenced by dialogue content rather than by dialogue flow.

For this dialogue-imagining exercise, we recruited 16 participants (eight males and eight females) who ranged from 19 to 39 years old. Each participant made twelve dialogues using two situations. For assigning the situations, we first created four conditions: (1) a student and living with family, (2) working and living with family, (3) a student and living alone, and (4) working and living alone. Then the participants were categorized into one of these conditions on the basis of their actual lifestyle and assigned two of the situations matching the condition.

For each situation, each participant created six imaginary dialogues from the six dialogue-act sequences output by the six systems: our POMDP and the other five systems for comparison. This process produced such dialogues as shown in Figs. 5 and 6. The dialogue in Fig. 5 was made from a dialogue-act tag sequence of a human-human conversation using No. 1 of Table 3. The dialogue in Fig. 6 was made from the sequence of our POMDP using No. 2 of Table 3.

In the third step, we additionally recruited three judges (one male and two females) to evaluate the imagined 192 (16 × 2 × 6) dialogues. The judges were neither the participants who made dialogues nor those who rated the collected listening-oriented dialogues. Six dialogues made from one situation were randomly shown to the judges one-by-one, who then filled out questionnaires to indicate their user satisfaction levels by answering this question on a 7-point Likert scale: "If you had been the speaker, would you have felt that you were listened to?"

5.2 Systems for comparison

We created our POMDP-based dialogue control and five other systems for comparison.

POMDP We learned a policy based on our approach. We used "food" dialogues (See Section 4), and the evaluation scores were those described in Section 4.2. This system used the policy to generate sequences of dialogue-act tags by simulation; user observations were generated based on emission probability, and system actions were generated based on the policy.

In this paper, the total number of observations and actions was 33 because we have 32 dialogue-act tags (See Table 1) plus a "skip" tag. In learning the policy, an observation and an action must individually take turns, but our data can include multiple dialogue-act tags in one utterance. Therefore, if there is more than one dialogue-act tag in one utterance, a "skip" is inserted between the tags. The state numbers for S_o and S_a were 16 and 33, respectively. In this experiment, we set 10 to $r_2((*, s_a), a)$.

EvenPOMDP We arranged a POMDP using only the smoothness reward (hereafter, EvenPOMDP) by creating a POMDP system with a fixed evaluation score; hence user satisfaction is not incorporated in the reward. When using fixed (even) evaluation scores for all dialogues, the effect of the user satisfaction reward is denied, and the system only generates highly frequent sequences. We have EvenPOMDP to clarify whether user satisfaction is necessary. The other conditions are identical as in the POMDP system.

HMM We modeled our dialogue-act tag sequences using a Speaker HMM (SHMM) (Meguro et al., 2009a), which has been utilized to model two-party listening-oriented dialogues. In a SHMM, half the states emit listener dialogue acts, and the other half emit speaker dialogue acts. All states are connected to each other. We modeled the "food" dialogues using an SHMM, and made the model generate the most probable dialogue-act tag sequences. More specifically, first, a dialogue-act tag was generated randomly based on the initial state. If the state was that of a listener, we generated a maximum likelihood action and the state was randomly transited based on the transition probability. If the state was that of a speaker, we randomly generated an action based on the emission probability and the state was transited using the maximum likelihood transition probability.

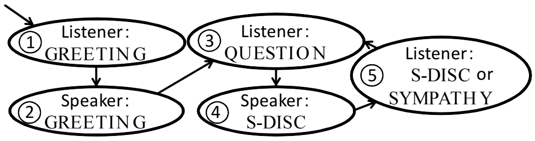

Figure 3: Structure of rule-based system

Rule-based system This system creates dialogue-act tag sequences using hand-crafted rules that are based on the findings in (Meguro et al., 2009a) and are realized as shown in Fig. 3. A sequence begins at state ① in Fig. 3, and one dialogue act is generated at each state. At state ③, a sub-category tag under QUESTION is chosen randomly, and at state ④, a matched sub-category tag under SELF-DISCLOSURE is chosen. At state ⑤, the listener's SELF-DISCLOSURE or SYMPATHY is generated randomly.

Human dialogue sequence This system created dialogue-act tag sequences by randomly choosing dialogues between humans from the collected data and used their annotated tag sequences.

Random This system simply created dialogue-act tag sequences at random.

5.3 Experimental results

Figure 4 shows the average subjective evaluation scores. Except between HMM and EvenPOMDP, there was a significant difference ($p<0.01$) between all systems in a non-parametric multiple comparison test (Steel-Dwass test). The dialogues shown in Figs. 5 and 6 were generated by the systems. The dialogue in Fig. 5 was made from human dialogue sequences, and the one in Fig. 6 was made from POMDP.

	With whom	What day	What time	What	Where	Who made
1	family	weekday	around 6:00 pm	grilled salmon	home	mother
2	family	weekend	around 7:00 pm	potato and meat	home	mother
3	co-workers	weekday	at noon	boiled seaweed	lunch box	myself
⋮	⋮	⋮	⋮	⋮	⋮	⋮
32	friend	weekday	at noon	hamburger	school cafeteria	N/A

Table 3: Dialogue situations relating to everyday Japanese life

We qualitatively analyzed the dialogues of each system and observed the following characteristics:

POMDP At a dialogue's beginning, the system greets several times and shifts to a different phase in which listeners ask questions and self-disclose to encourage speakers to reciprocate.

Rule-based The output of this system seems very natural and easy to read. The dialogue-act tags followed reasonable rules, making it easier for the participants to create natural utterances from them.

Human conversation The dialogues between humans were obviously natural before they were changed to tags from the natural-language utterances. However, human dialogues have randomness, which makes it difficult for the participants to create natural-language utterances from the tags. Hence, the evaluation score for this system was lower than the "Rule-based."

HMM, EvenPOMDP Since these systems continually output the same action tags, their output was very unnatural. For example, greetings never stopped because GREETING is most frequently followed by GREETING in the data. These systems have no mechanism to stop this loop.

POMDP successfully avoided such continuation because its actions have more varied rewards. For example, GREETING is repeated in EvenPOMDP because its smoothness reward is high; however, in POMDP, although the smoothness reward remains high, its user satisfaction reward is not that high. This is because greetings appear in all dialogues and their user satisfaction reward converges to the average. Therefore, such actions as greetings do not get repeated in POMDP. In POMDP, some states have high user satisfaction rewards, and the POMDP policy generated actions to move to such states.

Random Since this system has more variety of tags than HMM, its evaluation scores outperformed HMM, but were outperformed by POMDP, which learned statistically from the data.

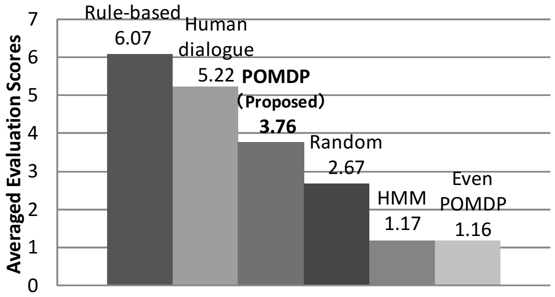

Figure 4: System scores. Except between POMDP and EvenPOMDP, significant differences exist among all systems (p<0.01).

From our qualitative analysis, we found that POMDP can generate more satisfying sequences than HMM/EvenPOMDP because it does not fall into the loop of frequent dialogue-act tag sequences. This suggests the usefulness of incorporating two kinds of rewards into the policy and that our approach for setting a reward is promising.

However, with the proposed POMDP, unnatural sequences remain; for example, the system suddenly output THANKS, as shown in Fig. 6. The number of states may have been too small. We plan to investigate what caused this in the future.

In our qualitative analysis, we observed that randomness in dialogues might hold a clue for improving evaluation scores. Therefore, we measured the perplexity of each system output using dialogue-act trigrams and obtained 72.8 for "Random," 27.4 for "Human dialogue," 7.4 for "POMDP," 3.2 for "HMM," 2.5 for "EvenPOMDP," and 1.7 for "Rule-based."

The perplexity of the human dialogues is less than that of the random system, but humans also exhibit a certain degree of freedom. On the other hand, POMDP's perplexity is less than the human dialogues; they still have some freedom, which probably led to their reasonable evaluation scores. Considering that HMM and EvenPOMDP, which continually output the same dialogue acts, had low

	Utterance	Dialogue act
S:	Hello.	GREETING
L:	Nice to meet you	GREETING
S:	I had dinner at home today.	S-DISC (sub: fact)
	Do you like grilled salmon?	QUESTION, PREF
L:	Yes, I think so.	SYMPATHY
	I sometimes want to have a fancy meal.	S-DISC (sub: desire)
S:	Deluxe.	REPEAT
	Me too.	SYMPATHY
L:	Do you usually do your own cooking?	QUESTION (sub: habit)
S:	No, I don't.	S-DISC, HABIT
	I always buy my meals at the convenience store.	S-DISC (sub: habit)
L:	I like the lunch boxes of convenience stores	S-DISC (sub: pref (positive))

Figure 5: Excerpt of listening-oriented dialogue that participant imagined from tag sequences of human conversations. Dialogue was translated from Japanese by the authors.

	Utterance	Dialogue act
L:	Nice to meet you.	GREETING
	Where and who did you have dinner with today?	QUESTION (sub: fact)
S:	I had "niku-jaga" (meat and beef) with my family at home.	S-DISC (sub: fact)
L:	Oh.	ADMIRATION
S:	I think it is normal to eat with your family at home.	S-DISC (sub: pref (neutral))
L:	Thanks.	THANKS
	Do you have any brothers or sisters?	QUESTION (sub: fact)
	Soon, my brother and his wife will visit my home.	S-DISC (sub: plan)
S:	I see.	SYMPATHY
L:	I want to use expensive meat in my "niku-jaga."	S-DISC (sub: desire)
	Oh.	ADMIRATION
	Please give me your recipe.	QUESTION (sub: information)
S:	My friends claim that my "niku-jaga" is as good as a restaurant's.	INFORMATION
L:	I'd love to try it	S-DISC (sub: desire)

Figure 6: Excerpt of a listening-oriented dialogue made from tag sequences of POMDP

evaluation scores, we conclude that randomness is necessary in non-task-oriented dialogues and that some randomness can be included with our approach. We do not discuss "Rule-based" here because its tag sequence was meant to have small perplexity.

6 Conclusion and Future work

This paper investigated the possibility of automatically building a dialogue control module from dialogue data to create automated listening agents.

With a POMDP as a learning framework, a dialogue control module was learned from the listening-oriented dialogues we collected and compared with five different systems. Our POMDP system showed higher performance in subjective evaluations than other statistically motivated systems, such as an HMM-based one, that work by selecting the most likely subsequent action in the dialogue data. When we investigated the output sequences of our POMDP system, the system frequently chose to self-disclose and question, which corresponds to human listener behavior, as revealed in the literature (Meguro et al., 2009a). This suggests that learning dialogue control by POMDPs is achievable for listening-oriented dialogues.

The main contribution of this paper is that we successfully showed that POMDPs can be used to train dialogue control policies for less task-oriented dialogue systems, such as listening agents, where the user goals are not as clear as task-oriented ones. We also revealed that the reward function can be learned effectively by our formulation that simultaneously maximizes user satisfaction and smoothness. Finding an appropriate reward function is a real challenge for less task-oriented dialogue systems; this work has presented the first workable solution.

Much work still remains. Even though we conducted an evaluation experiment by simulation (i.e, offline evaluation), human dialogues obviously do not necessarily proceed as in simulations. Therefore, we plan to evaluate our system using online evaluation, which also forces us to implement utterance understanding and generation modules. We also want to incorporate the idea of topic shift into our policy learning because we observed in our data that listeners frequently change topics to keep speakers motivated. We are also considering adapting the system behavior to users. Specifically, we want to investigate dialogue control that adapts to the personality traits of users because it has been found that the flow of listening-oriented dialogues differs depending on the personality traits of users (Meguro et al., 2009b). Finally, although we only dealt with text, we also want to extend our approach to speech and other modalities, such as gestures and facial expressions.

References

Bickmore, Timothy and Justine Cassell. 2001. Relational agents: a model and implementation of building user trust. In *Proc. SIGCHI conference on human factors in computing systems (CHI)*, pages 396–403.

Ferguson, George, James F. Allen, and Brad Miller. 1996. TRAINS-95: towards a mixed-initiative planning assistant. In *Proc. Third Artificial Intelligence Planning Systems Conference (AIPS)*, pages 70–77.

Higashinaka, Ryuichiro, Kohji Dohsaka, and Hideki Isozaki. 2008. Effects of self-disclosure and empathy in human-computer dialogue. In *Proc. IEEE Workshop on Spoken Language Technology (SLT)*, pages 108–112.

Higuchi, Shinsuke, Rafal Rzepka, and Kenji Araki. 2008. A casual conversation system using modality and word associations retrieved from the web. In *Proc. 2008 conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 382–390.

Hirshman, Lynette. 1989. Overview of the DARPA speech and natural language workshop. In *Proc. DARPA Speech and Natural Language Workshop 1989*, pages 1–2.

Isomura, Naoki, Fujio Toriumi, and Kenichiro Ishii. 2009. Evaluation method of non-task-oriented dialogue system using HMM. *IEICE Transactions on Information and Systems*, J92-D(4):542–551.

Ivey, Allen E. and Mary Bradford Ivey. 2002. *Intentional Interviewing and Counseling: Facilitating Client Development in a Multicultural Society*. Brooks/Cole Publishing Company.

Jurafsky, Dan, Elizabeth Shriberg, and Debra Biasca, 1997. *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual*.

Kobayashi, Yuka, Daisuke Yamamoto, Toshiyuki Koga, Sachie Yokoyama, and Miwako Doi. 2010. Design targeting voice interface robot capable of active listening. In *Proc. 5th ACM/IEEE international conference on Human-robot interaction (HRI)*, pages 161–162,

Maatman, R. M., Jonathan Gratch, and Stacy Marsella. 2005. Natural behavior of a listening agent. *Lecture Notes in Computer Science*, 3661:25–36.

Meguro, Toyomi, Ryuichiro Higashinaka, Kohji Dohsaka, Yasuhiro Minami, and Hideki Isozaki. 2009a. Analysis of listening-oriented dialogue for building listening agents. In *Proc. 10th Annual SIGDIAL Meeting on Discourse and Dialogue (SIGDIAL)*, pages 124–127.

Meguro, Toyomi, Ryuichiro Higashinaka, Kohji Dohsaka, Yasuhiro Minami, and Hideki Isozaki. 2009b. Effects of personality traits on listening-oriented dialogue. In *Proc. International Workshop on Spoken Dialogue Systems Technology (IWSDS)*, pages 104–107.

Minami, Yasuhiro, Akira Mori, Toyomi Meguro, Ryuichiro Higashinaka, Kohji Dohsaka, and Eisaku Maeda. 2009. Dialogue control algorithm for ambient intelligence based on partially observable markov decision processes. In *Proc. International Workshop on Spoken Dialogue Systems Technology (IWSDS)*, pages 254–263.

Nakano, Mikio, Noboru Miyazaki, Jun ichi Hirasawa, Kohji Dohsaka, and Takeshi Kawabata. 1999. Understanding unsegmented user utterances in real-time spoken dialogue systems. In *Proc. 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 200–207.

Pineau, Joelle., Geoff. Gordon, and Sebastian Thrun. 2003. Point-based value iteration: An anytime algorithm for POMDPs. In *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1025–1032.

Shitaoka, Kazuya, Ryoko Tokuhisa, Takayoshi Yoshimura, Hiroyuki Hoshino, and Narimasa Watanabe. 2010. Active listening system for dialogue robot. In *JSAI SIG-SLUD Technical Report*, volume 58, pages 61–66. (in Japanese).

Walker, Marilyn, Alex Rudnicky, John Aberdeen, Elizabeth Owen Bratt, Rashmi Prasad, Salim Roukos, Greg S, and Seneff Dave Stallard. 2002. DARPA communicator evaluation: progress from 2000 to 2001. In *Proc. International Conference on Spoken Language Processing (ICSLP)*, pages 273–276.

Wallace, Richard S. 2004. *The Anatomy of A.L.I.C.E.* A.L.I.C.E. Artificial Intelligence Foundation, Inc.

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

Yokoyama, Sachie, Daisuke Yamamoto, Yuka Kobayashi, and Miwako Doi. 2010. Development of dialogue interface for elderly people –switching the topic presenting mode and the attentive listening mode to keep chatting–. In *IPSJ SIG Technical Report*, volume 2010-SLP-80, pages 1–6. (in Japanese).