Evaluation of Listening-oriented Dialogue Control Rules based on the Analysis of HMMs

Toyomi Meguro†*, Yasuhiro Minami*†*, Ryuichiro Higashinaka*‡*, Kohji Dohsaka*†

† NTT Communication Science Laboratories, NTT Corporation
‡ NTT Cyber Space Laboratories, NTT Corporation
{meguro.toyomi, minami.yasuhiro, higashinaka.ryuichiro, dohsaka.kohji}@lab.ntt.co.jp

10.21437/Interspeech.2011-295

Abstract

We have been working on listening-oriented dialogues for the purpose of building listening agents. In our previous work [1], we trained hidden Markov models (HMMs) from listening-oriented dialogues (LoDs) between humans, and by analyzing them, discovered a distinguishing dialogue flow of LoD. For example, listeners suppress their information giving and self-disclosure, and instead, increase acknowledgments and questions to elicit speakers' utterances. As an initial step for building listening agents, we decided to create dialogue control rules based on our analysis of the HMMs. We built our rule-based system and compared it with three other systems by a Wizard of Oz (WoZ) experiment. As a result, we found that our rule-based system achieved as much user satisfaction as human listeners.
Index Terms: Listening-oriented dialogue, Dialogue system, Wizard of Oz

1. Introduction

We have been working on listening-oriented dialogue (LoD) in which one conversational participant listens attentively to the other (see Fig. 5 for a typical LoD). Our aim is to build listening agents that can implement a listening process in which users can satisfy their desire to speak and have themselves heard.

For this purpose, we have been analyzing LoDs conducted between humans by comparing them with casual conversation in which conversational participants have no predefined roles. For this analysis, we used hidden Markov models (HMMs) to model LoDs and casual conversation, and compared state transitions that represent dialogue flows. Figure 1 shows an example of a trained HMM. By examining such transitions, we found that LoD and casual conversation have significantly different dialogue flows. For example, in LoD:

- Listeners ask questions actively with a frequent insertion of self-disclosure, which seems in concordance with social penetration theory [2], which states that one needs to self-disclose to prompt others to speak about themselves.
- Listeners suppress information giving and self-disclosure, and instead, increase acknowledgments and questions to elicit speakers' utterances.

Since such insights would help us build automated listening agents, we decided to encode them as dialogue control rules and build a dialogue control module. Here, we create rules by hand because this is a typical and basic way of building less task-oriented dialogue systems [3, 4, 5], and handcrafted rules can be flexible at the initial stage of research. In addition, handcrafted rules have the following advantages:

- It is not necessary to collect a large number of dialogues generally needed for training stochastic models. Note that the state space can be very large for less task-oriented dialogues.

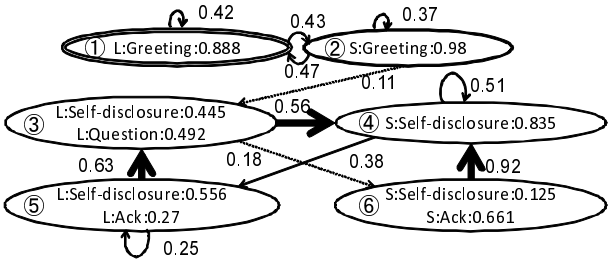


Figure 1: HMM trained from LoDs between humans [1].

- It is safe in terms of system development because a completely wrong action cannot be chosen.
- It is easy to make a system perform certain actions, to modify system actions, and to maintain the system.

The disadvantage may be that it is often difficult to write down rules for all possible dialogue states, which may make system actions seem rigid and inflexible, although this can be overcome by adopting stochastic models with large dialogue data.

Acknowledging that both handcrafted rules and stochastic models have advantages and disadvantages, we weigh highly the advantages of handcrafted rules and build our prototype system using handcrafted rules, although we have also begun working on using stochastic models [6]. To verify the effectiveness of our rules, we create three other systems and compare them with our rule-based prototype by a Wizard of Oz (WoZ) experiment.

Note that this paper deals with the dialogue control module of a listening agent, which receives the user's dialogue act (i.e., meaning representation of an utterance) as input, and outputs the system's next dialogue act. The module does not deal with surface utterances. Although other modules such as language understanding and language generation are important, we consider dialogue control to be the most important module for listening agents because it decides the flow of dialogue and that, according to our analysis, is what characterizes LoDs.

2. Related work

There is emerging work on building listening agents. Shitaoka et al. [3] investigated the functions of listening agents, focusing especially on their response generation component. Their system takes the confidence score of speech recognition into account and changes the system response accordingly; that is, it repeats the user utterance or utters an empathic utterance for high-confidence user utterances, and it makes a back-channel when the confidence is low. Here, the system's emphatic utterances can be "I'm happy" or "It's a pity", depending on whether a positive or negative expression is included in user utterances.

Copyright © 2011 ISCA

Their system's response generation only uses the speech recognition confidence and the polarity of user utterances as cues to choose its actions. Currently, their system does not take into account the content of an utterance, such as dialogue acts.

To create listening agents that achieve high smoothness, a switching mechanism between "active listening mode", in which a system is a listener and actively utters back-channeling utterances, and "topic presenting mode", in which the system is a speaker, has been proposed [4, 5]. Here, the system uses a heuristic function to maintain a high user interest level and to keep the system in the active listening mode. Dialogue control is done by handcrafted rules. Our motivation is similar to theirs, but differs in that we want to build a listening agent that gives the user a sense of being heard not only by maintaining the user's interest by back-channeling but also by various linguistic actions.

Maatman et al. [7] have also been working on building listening agents, but they focus only on non-linguistic actions, such as gestures. We consider linguistic actions to be equally important in LoD because they concern the content of a dialogue.

3. Rule-based LoD control module

We made our dialogue control rules based on the analysis of our HMMs [1]. The rules were made to output dialogue acts in response to user dialogue acts. Table 1 lists the dialogue acts that can be exchanged between a listening agent and a user. The table also shows generation templates corresponding to the dialogue acts that were used by WoZ experimenters to create utterances. Note that, to enable more fine-grained exchanges, the dialogue act set here has been extended from when we performed our analysis [1].

Figure 2 shows all of the rules that we created. One dialogue act corresponds to one sentence and a user can utter multiple sentences in one utterance. The rules respond to the last dialogue act of the user utterance and generate the next system dialogue act. The rules basically encode the results of our analysis although we needed to insert some rules using insight from previous studies. In the rules, after a user's self-disclosure, the system self-discloses and then asks a question. This is because, in Fig. 1, the typical flow of LoD is that when the speaker self-discloses, the listener first self-discloses and then asks a question. Also, the rules sparingly output INFORMATION, which is only returned when requested by the user or in accordance with the user's information giving.

On the basis of previous studies, the rules return the self-disclosure with the same sub-tag for a question, which is reasonable in cooperative dialogue. The rules also have some mirroring actions [8]; for example, for a user's sympathy, the system also sympathizes with the user. Also, in the rules, SYMPATHY is output or inserted frequently. For example, after two or more user's actions of REPEAT, PARAPHRASE, etc., the system sympathizes with the user. This is because, in previous research [9], sympathy has been found effective for keeping the user engaged in conversation. For active listening practice, paraphrasing and repeating have been found effective [10] because they show a high level of understanding by the listener. Therefore, after two or more user's actions of INFORMATION, APOLOGY, CONFIRMATION, etc., we made the rules return REPEAT or PARAPHRASE.

4. Experiment

4.1. Experimental setup

To verify the effectiveness of our rules, we performed a WoZ experiment. In the experiment, the wizards manipulated seven systems in all. Four of the systems were those that we wanted to compare: one with our rules and three other systems for comparison; and the other three systems were the experimental systems that we wanted to test for future development. We only deal with the former four systems and discuss their results in this report, but note that we made sure that each of the four systems was evaluated under the same condition in terms of experimental order. We describe our four systems in detail in the next section. The WoZ experimenters did not have any knowledge about the systems except that they needed to convert user utterances into dialogue acts, input them to the system, receive system dialogue acts, and finally generate utterances from the system dialogue acts to communicate with the user. Here, generation is done using the generation templates (see Fig. 2). The topic of all dialogues was "food".

We recruited 18 participants (nine males; nine females) as speakers and three participants (two females; one male) as wizards (i.e. listeners). The participants did not include the authors. The WoZ experimenters and the participants were kept in different rooms and they only used text to communicate. Each participant talked with two WoZ experimenters. Each partici-

Table 1: Definitions of dialogue acts with sample generation templates. The templates were used by a WoZ experimenter to generate utterances for the dialogue acts. The templates were adapted for the topic of "food" for our experiment. Except for REPEAT, PARAPHRASE and OTHER, each dialogue act has several generation templates. Templates are shown in square brackets. (time) can be replaced with lunch, breakfast or dinner and (day) with tomorrow, today etc. (food) is filled with arbitrary food by a WoZ experimenter according to a dialogue context.

Dialogue act	Definition and sample generation templates
GREETING	Greeting and confirmation of a dialogue theme. e.g.[Hello.],[The theme of dialog is XX.]
INFORMATION	Delivery of objective information. [(place-name) is famous for (food).]
SELF-DISCLOSURE	Disclosure of one's preferences and feelings.
sub: fact	[I ate (food) for (time) (day).]
sub: experience	[I have eaten (food).]
sub: habit	[I always go out to eat.]
sub: preference (positive)	[I like (food).] [(food) is delicious.]
sub: preference (negative)	[I don't like (food).] [(food) is not delicious.]
sub: preference (neutral)	[(food) is (adjective: either positive or negative).]
sub: desire	[I want to eat (food) for (time) (day).]
sub: plan	[I will eat (food) for (time) (day).]
ACKNOWLEDGMENT	Encourages the conversational partner to speak by a backchannel. [Well.] [Aha.]
QUESTION	Utterances that expect answers.
sub: information	[Please tell me about (food).]
sub: fact	[What did you eat for (time) (day)?]
sub: experience	[Have you eaten (food) before?]
sub: habit	[Do you usually go out to eat?]
sub: preference	[Do you like (food)?]
sub: desire	[What do you want to eat for (time) (day)?]
sub: plan	[What will you eat for (time) (day)?]
SYMPATHY	Sympathetic utterances and praises. [Me, too]
NON-SYMPATHY	Negative utterances. [I don't think so.]
CONFIRMATION	Confirm what the conversation partner said. [Really?]
PROPOSAL	Encourage the partner to act. [Please eat (food).]
REPEAT	Repeat the partner's most recent utterance.
PARAPHRASE	Paraphrase the partner's most recent utterance.
APPROVAL	Bring up or show goodwill toward the partner. [Absolutely!]
THANKS	Express one's thanks [Thank you.]
APOLOGY	Express one's regret [I'm sorry.]
FILLER	Filler between utterances. [Uh.],[Let me see.]
ADMIRATION	Express one's affection. [A-ha-ha.]
OTHER	Other utterances.

```
function generate_action(user_action) returns system_action
  num ← random(1..3)
  system_action ← {}
  if user_action = GREETING then
    return GREETING × num
  else if user_action = INFORMATION then
    system_action ← SYMPATHY
    system_action ← INFORMATION × num
    return system_action
  else if user_action = S-DISC then
    system_action ← user_action × num
    system_action ← QUESTION(sub:(random.sub))
    return system_action
  else if user_action = ACKNOWLEDGMENT then
    return S-DISC(sub: fact) × num
  else if user_action = QUESTION then
    if user_action.sub = information then
      return INFORMATION × num
    else if user_action.sub = preference then
      return S-DISC(sub:pref+) × num
    else
      return S-DISC (sub: user_action.sub) × num
  else if user_action = SYMPATHY or NON-SYMPATHY then
    system_action ← REPEAT
    system_action ← user_action × num
    return system_action
  else
    if user_action = REPEAT or PARAPHRASE
       or THANKS or FILLER then
      system_action ← SYMPATHY
    else if user_action = PROPOSAL or APPROVAL then
      system_action ← THANKS
    else if user_action = APOLOGY or OTHER then
      system_action ← REPEAT
    else if user_action = CONFIRMATION then
      system_action ← PARAPHRASE
    else if user_action = ADMIRATION then
      system_action ← ADMIRATION
    system_action ← S-DISC (sub: fact) × num
    return system_action
end
```

Figure 2: Our dialogue control rules written as a function. S-DISC is shorthand for SELF-DISCLOSURE and PREF+ for preference positive. "←" means adding a dialogue act to a list, "× num" a repetition of a preceding dialogue act "num" times, and "random.sub" means a randomly chosen sub tag.

pant talked with a WoZ using the four different systems, which yielded eight dialogues per participant. Each dialogue lasted approximately 15 minutes. After each dialogue, the participants filled out questionnaires to rate their user satisfaction level by answering the question, "Did you feel that you were listened to?" on a 7-point Likert scale.

4.2. Systems

4.2.1. Rule-based dialogue control

This system is our proposed system and uses the rules in Fig. 2. We call this system the **Rule-based system**.

4.2.2. HMM dialogue control

This system uses for dialogue control an HMM trained from dialogue-act annotated LoD data between humans that we collected [6]. The annotation was done using the tag set in Table 1. The data include 1260 LoDs. We have this system to compare our rules with a stochastic model. The system (hereafter, **HMM system**) chooses the most probable dialogue act at each step. The system makes a prediction as to whether it should continue its utterance or pass the turn to the user. If the most probable action at the next step is that of the user, the system stops generating dialogue acts. If the most probable dialogue act is that of the system, the system continues to generate dialogue acts by greedily choosing the most probable dialogue act at each step. The number of states for the HMM was 16. Figure 3 shows the graphical representation of the HMM we used. The figure shows only six states to make it easy to understand the dialogue flow.

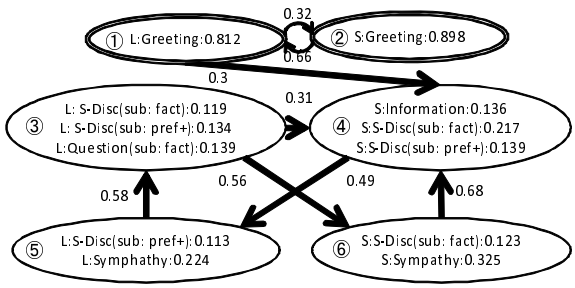


Figure 3: HMM used for HMM System. For brevity, only output probabilities over 0.1 and transition probabilities over 0.3 are shown.

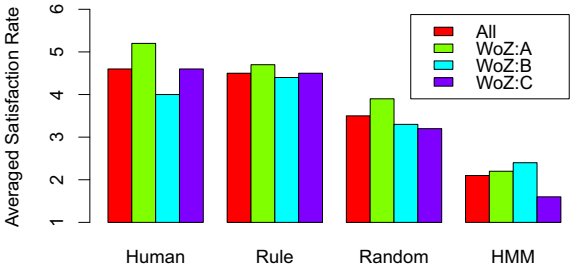


Figure 4: Averaged user satisfaction ratings.

4.2.3. Human dialogue control

In this system, the WoZ experimenters choose the system dialogue acts (both the dialogue acts and the number of dialogue acts) at their discretion. We regard this system as an upper bound. We call this system the **Human system**.

4.2.4. Random dialogue control

This system chooses the system dialogue acts randomly. The number of dialogue acts to be output is also random from one to three. We call this system the **Random system**.

4.3. Experimental results

Figure 4 shows the averaged user satisfaction ratings for the four systems. Except between the Human system and Rule-based system, there are significant differences between all systems ($p<0.01$) by a statistical test (t-test). Surprisingly, the user satisfaction rating of the Rule-based system almost equaled that of the Human System, and it greatly outperformed the HMM system, the stochastic model. In addition, the Rule-based system worked preferably for a person who is not a good listener; WoZ:B's score of the Human system was not that high compared to other WoZ experimenters, but WoZ:B's score of the Rule-based system was higher than the Human system and comparable to other WoZ experimenters, suggesting that our rules could be used as possible aids for listeners.

Figures 5 and 6 show the example dialogues conducted using the Rule-based system and HMM system. In the dialogue of the Rule-based system, WoZ experimenters asked questions after their self-disclosures, which closely follows our analysis, and thus obtained high user satisfaction ratings. On the other hand, the HMM system always chose the same action, SELF-

Utterance	Dialogue act
L: Nice to meet you.	GREETING
S: Nice to meet you, too.	GREETING
L: The theme of dialogue is "Most impressive food".	GREETING
S: OK.	SYMPATHY
L: OK.	REPEAT
S: My most impressive food is the spit-roasted silkworms at a Chinese restaurant in Ikebukuro.	S-DISC(sub: fact)
L: I ate Chinese food last night.	S-DISC(sub: fact)
I ate it in Ikebukuro.	S-DISC(sub: fact)

Figure 5: Excerpt of LoD using the Human system. S stands for speaker, and L for listener. The dialogue was originally in Japanese and was translated by the authors.

Utterance	Dialogue act
L: Nice to meet you.	GREETING
S: Nice to meet you, too.	GREETING
L: The theme of dialogue is "Today's food".	GREETING
S: Yes, I ate spaghetti.	S-DISC(sub: fact)
L: The theme of dialogue is "Today's food".	GREETING
S: Yes, I understand.	SYMPATHY
L: I ate grilled fish for dinner yesterday.	S-DISC(sub: fact)
S: It sounds delicious.	S-DISC(sub: pref+)
L: I can cook grilled fish.	S-DISC(sub: fact)

Figure 6: Excerpt of LoD using the HMM system.

DISCLOSURE (sub: fact), repeatedly because it is the most probable. Even if the users showed a variety of dialogue acts, the HMM system chose the same dialogue act and that repetition made dialogues monotonous, which led to its poor performance. An appropriate amount of self-disclosure would make social relationships between systems and users stronger. However, if the system always self-discloses, this would make users feel that they are not being listened to attentively.

The Random system generated all kinds of dialogue acts. Although there was no tangible dialogue flow, because of the randomness, the dialogue was not monotonous, which probably led to higher user satisfaction ratings than the HMM system.

To further investigate the quality of dialogues by each system, we calculated the Kullback-Leibler (KL) divergence between (a) an HMM trained from human-human LoDs and (b) HMMs trained from the data of each system; that is, the more similar an HMM is to that of an HMM trained from human-human dialogue or from the data of the Human system, the better the system should be.

Table 2 shows the KL divergence of HMMs for each system. Low scores mean that the distributions of the HMMs (i.e., output distribution of dialogue acts) are similar. The first row shows the divergence between (a) and (b). The second row shows the KL divergence between an HMM trained from the data of the Human system and (b). From the first row, we can see that the distribution of human-human data and that of the Human system are very similar. This suggests that there were no unrealistic restrictions in the experiment and the participants could talk as usual, confirming that the dialogues in the WoZ experiment were realistic. The dialogue distribution of the Random system is closer to that of human-human dialogues than the Rule-based system, meaning that the dialogue of the Rule-based system could still be monotonous and the number of our rules could still be small. We may need more randomness in the Rule-based system. However, notice that the user satisfaction ratings of the Random system are low. Therefore, too much randomness would not be effective. Focusing on the second row, the dialogue act distribution of the Rule-based system is more similar to that of the Human system than the Random system. This

Table 2: Kullback-Leibler divergence between HMMs.

	Human	HMM	Rule	Random
Human-human LoD [6]	0.9	4.4	2.8	2.4
Human System	N/A	4.7	1.4	1.7

may be attributable to the templates used by the WoZ experimenters; the templates could have restricted the experimenters' freedom slightly in the Human system.

5. Conclusions and Future work

In this paper, we built a dialogue control module for a listening agent using handcrafted rules based on our analysis of the HMMs trained from human-human LoDs. Then, we evaluated the rules by comparing them with three other systems by a WoZ experiment. As a result, we found that our Rule-based system performed as well as the Human system. We also found that the HMM-based system is worse than other systems because it repeatedly generated the same dialogue acts that appeared frequently in the training data, showing some limitation of stochastic models.

As future work, we plan to continue improving and augmenting our rules. However, it would be difficult to write down rules for all possible dialogue states. For this reason, we have already started examining a method to apply partially observable Markov decision processes to automate dialogue control from data [6]. Whether rule-based or statistic-based, we would like to investigate ways to create listening agents that maximize user satisfaction.

6. Acknowledgements

This work was partially supported by a Grant-in-Aid for Scientific Research on Innovative Areas, "Founding a creative society via collaboration between humans and robots" (21118004), from MEXT, Japan.

7. References

[1] T. Meguro, R. Higashinaka, K. Dohsaka, Y. Minami, and H. Isozaki, "Analysis of listening-oriented dialogue for building listening agents," in *Proc. SIGDIAL*, 2009, pp. 124–127.

[2] I. Altman and D. Taylor, *Social Penetration: The Development of Interpersonal Relationships*. Holt Rinehart and Winston, 1973.

[3] K. Shitaoka, R. Tokuhisa, T. Yoshimura, H. Hoshino, and N. Watanabe, "Active listening system for dialogue robot," in *Proc. SIG-SLUD, JSAI*, vol. 58, 2010, pp. 61–66, (in Japanese).

[4] S. Yokoyama, D. Yamamoto, Y. Kobayashi, and M. Doi, "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*, vol. 2010-SLP-80, no. 4, 2010, pp. 1–6, (in Japanese).

[5] Y. Kobayashi, D. Yamamoto, T. Koga, S. Yokoyama, and M. Doi, "Design targeting voice interface robot capable of active listening," in *Proc. HRI*, 2010, pp. 161–162.

[6] T. Meguro, R. Higashinaka, Y. Minami, and K. Dohsaka, "Controlling listening-oriented dialogue using partially observable Markov decision processes," in *Proc. COLING*, 2010, pp. 761–769.

[7] R. M. Maatman, J. Gratch, and S. Marsella, "Natural behavior of a listening agent," *Lecture Notes in Computer Science*, vol. 3661, pp. 25–36, 2005.

[8] R. B. Cialdini, *Influence: Science and Practice*. Allyn & Bacon, 2000.

[9] R. Higashinaka, K. Dohsaka, and H. Isozaki, "Effects of self-disclosure and empathy in human-computer dialogue," in *Proc. SLT*, 2008, pp. 108–112.

[10] G. Buck, *Assessing listening*. Cambridge University Press, 2001.