Collection and Analysis of Meaningful Dialogue by Constructing a Movie Recommendation Dialogue System

Takashi Kodama
Kyoto University

Ribeka Tanaka
Kyoto University

Sadao Kurohashi
Kyoto University
NII CRIS

{kodama, tanaka, kuro}@nlp.ist.i.kyoto-u.ac.jp

1 Introduction

Intelligent dialogue systems must be able to produce the utterance that match to the previous utterance of the conversational partner, as well as the *dialogue context* more broadly. This context can be influenced by former utterances, dialogue participants intentions, focus of interests, and so many other factors.

Especially, in human-human dialogues, dialogue participants have some purposes with the dialogue even in chat dialogues, e.g., telling some information and influencing the partner. Based on the purposes, we humans maintain a coherent dialogue for multiple turns, take care of what the partner thinks, and behave appropriately based on the former conversation. Handling such factors that govern the multiple dialogue turns is one of the biggest issues towards a better dialogue system.

In order to address this issue, it is essential to analyze multi-turn dialogue data. However, because the recent studies on chat dialogues have paid most attention to dealing with the variety of user utterances in open-domain chit-chat, the resultant dialogues are locally appropriate but not suitable for context-level analysis.

In this study, we propose a new large human-machine dialogue dataset for Japanese. We constructed a scenario-based movie recommendation dialogue system and collected 1060 dialogues by crowdsourcing. Figure 1 shows an example of collected dialogues[1]. The system starts the dialogue with a movie-related topic such as news topic, introduces a related movie, and presents some recommendation points. The system produced mostly natural dialogues, while it also raised some context-level errors that are worth considering toward a better dialogue system.

S1: It is a hot topic that it was revealed by the 11th that the second child was born to actor Shota Someya and actress Rinko Kikuchi.
U1: That's a good news.
S2: Shota Someya appeared in the movie called "The Lion in March: Part 1."
U2: Right.
S3: There is a way of attacking in *Shogi* for each players, and their face express the conflict and distress in the game. You will be moved and nailed to the picture.
U3: I want to see it once.
S4: The performance of the casts is the best, so the video and music are at a satisfactory level.
U4: I'm interested in it.
S5: I think you can enjoy it.

Figure 1: Example of collected dialogue (S is the system and U is the user).

2 System Configuration

Figure 2 shows an overview of our proposed system. There are two main components in our system: a movie database and a dialogue manager. A movie database stores information of 213 movies acquired by web crawling. For each movie, we prepare the system's utterances beforehand as a *scenario* using the information of the movie database.

To produce natural dialogue, we created scenarios based on the three strategies: (1) the system starts the dialogue by introducing the surrounding topic and later presents the title of the recommended movie; (2) sentences that describe recommended points of the movie are uttered multiple times; (3) in principle, the system utters according to the scenario, as if it takes the dialogue initiative. Since the last point can be a cause of disagreement, we designed the dialogue manager to answer easy questions and to react to some typical user utterances.

[1] Examples of dialogues presented in this paper are originally in Japanese and were translated by the authors.

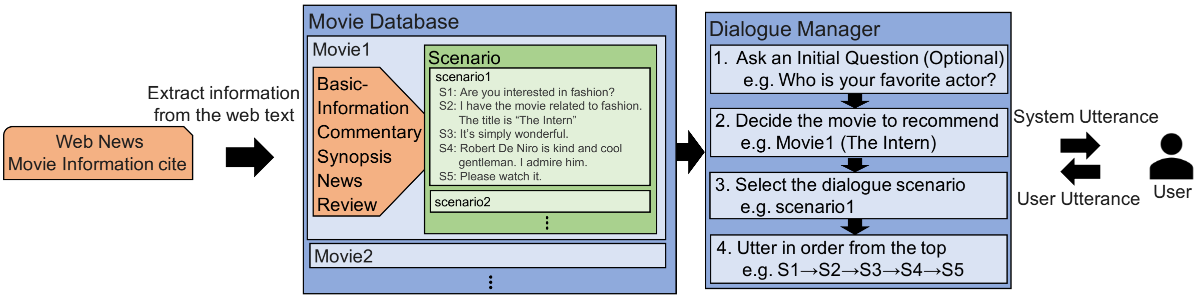
Web News
Movie Information cite
Extract information from the web text
Movie Database
Movie1
Basic-Information
Commentary
Synopsis
News
Review
Scenario
scenario1
S1: Are you interested in fashion?
S2: I have the movie related to fashion. The title is "The Intern"
S3: It's simply wonderful.
S4: Robert De Niro is kind and cool gentleman. I admire him.
S5: Please watch it.
scenario2
Movie2
Dialogue Manager
1. Ask an Initial Question (Optional) e.g. Who is your favorite actor?
2. Decide the movie to recommend e.g. Movie1 (The Intern)
3. Select the dialogue scenario e.g. scenario1
4. Utter in order from the top e.g. S1→S2→S3→S4→S5
System Utterance
User Utterance
User

Figure 2: The overview of movie recommendation dialogue system

Based on the three strategies, we created one or more scenarios for each movie. In the first part of the scenario, the system started with movie-related topic and presented recommended movies. In the remaining part, the system made two utterances that describe the recommendation points and finally prompted the user to watch the movie. We extracted recommendation points from the user reviews of a film review website and converted them to appropriate syntactic form by some rules.

Which movie to recommend as well as which scenario to use are decided by the dialog manager. It drives the dialogue based on each scenario by referring to the movie information if necessary.

3 Dialogue Collection and Analysis

We collected dialogue by our movie recommendation system on crowdsourcing. After each dialogue, we asked the worker to answer some questionnaire. We selected the following four referring to the evaluation metrics adopted by Hiraoka et al. (2013); Li et al. (2018); Yoshino et al. (2018).

(1) **Potential Interest**: Do you like movies?

(2) **Watching Experience**: Have you seen this recommended movie?

(3) **Persuasiveness**: Do you want to see this recommended movie?

(4) **Naturalness of Flow**: Was the flow of dialogue natural?

The result we got suggested that our scenario-based system was able to produce natural utterances that keep the dialogue purpose.

We labeled each of the system utterances in the collected dialogues whether it is natural in the concerned dialogue context. We conducted this annotation task also on crowdsourcing. For each system utterance except the first utterance, we asked three workers to annotate it with one of the three labels, *Natural*, *Possibly Unnatural*, and *Unnatural*. When an utterance is judged as unnatural by multiple annotators, we further classified its error type. We adopted the error taxonomy provided by Higashinaka et al. (2015a,b) and we further classified the error types for some subcategories based on the observed cases. As they proposed, we also distinguished the four hierarchical levels, namely, *utterance-level*, *response-level*, *context-level*, and *environment-level* (see Higashinaka et al. (2015a) for details). Table 1 shows the distribution of error categories in dialogue data proposed by Higashinaka et al. (2015a) as well as in dialogue data we collected. Although the direct comparison is difficult, our data contains less utterance-level errors and more context-level errors compared to dialogue collected by Higashinaka et al. (2015a). This suggests dialogues we collected contain more errors concerned with dialogue context, which are worth analyzing in more depth towards a better dialogue system.

Main category	H+	Ours
Utterance-level	12.7%	4.1%
Response-level	51.1%	41.6%
Context-level	29.9%	54.3%
Environment-level	6.3%	0.0%

Table 1: Distribution of error categories (H+ is numbers obtained by Higashinaka et al. (2015a)).

4 Conclusion

We proposed a human-machine dialogue collection in Japanese. The dialogue data will be made available for research use on a request basis. Future research will have to investigate more on the cause of contextual errors and the way to avoid the unnatural utterances.

Acknowledgments

This research was supported by NII CRIS Contract Research 2019.

References

Ryuichiro Higashinaka, Kotaro Funakoshi, Masahiro Araki, Hiroshi Tsukahara, Yuka Kobayashi, and Masahiro Mizukami. 2015a. Towards taxonomy of errors in chat-oriented dialogue systems. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 87–95, Prague, Czech Republic. Association for Computational Linguistics.

Ryuichiro Higashinaka, Masahiro Mizukami, Kotaro Funakoshi, Masahiro Araki, Hiroshi Tsukahara, and Yuka Kobayashi. 2015b. Fatal or not? finding errors that lead to dialogue breakdowns in chat-oriented dialogue systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2243–2248, Lisbon, Portugal. Association for Computational Linguistics.

Takuya Hiraoka, Yuki Yamauchi, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. 2013. Dialogue management for leading the conversation in persuasive dialogue systems. In *Proceedings of IEEE automatic speech recognition and understanding workshop (ASRU)*, pages 114–119. IEEE.

Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 9725–9735. Curran Associates, Inc.

Koichiro Yoshino, Yoko Ishikawa, Masahiro Mizukami, Yu Suzuki, Sakriani Sakti, and Satoshi Nakamura. 2018. Dialogue scenario collection of persuasive dialogue with emotional expressions via crowdsourcing. In *Proceedings of the 11th Language Resources and Evaluation Conference*, Miyazaki, Japan. European Language Resource Association.