Spoken Dialogue Robot for Watching Daily Life of Elderly People

Koichiro Yoshino[1,2,3], Yukitoshi Murase[1], Nurul Lubis[1], Kyoshiro Sugiyama[1], Hiroki Tanaka[1], Sakti Sakriani[1,2], Shinnosuke Takamichi[4], Satoshi Nakamura[1,2]

Abstract The number of aged people is increasing. The influence of solitude on both physical and mental health of those seniors is a social problem that needs an urgent solution in advanced societies. We propose a spoken dialogue robot that looks over elderly people through conversations by using functions of life-support via information navigation, attentive listening, and anomaly detection. In this paper, we describe a demonstration system implemented in the conversational robot.

1 Introduction

Social relationships are essential for promoting mental health. People who perceived the absence of positive social relationship tend to have a higher risk on both the physical and the mental health than people who have a family living together. They are depressed or socially isolated since they tend to have less communication. In the aged society, the number of aged people in solitude increases, since friends and relatives often do not have much time to be a constant companion. On the other hand, the demand for professional caregivers for older adults already outstrips supply. Thus technologies that increase opportunities for conversation are highly expected [4, 2, 3, 1].

Communication robots are expected to solve these problems, as agents that use communicative functions to prevent isolating aged people. These systems reduce loneliness by listening to their self-disclosure attentively and increase the number of contacts to societies by providing information about daily events. These systems can also detect anomalies in aged people through daily conversations, for example, response delay caused by dementia [15].

[1]Nara Institute of Science and Technology, Japan
[2]RIKEN Center for Advanced Intelligence Project, Japan
[3]PRESTO, Japan Science and Technology Agency, Japan
[4]The University of Tokyo, Japan

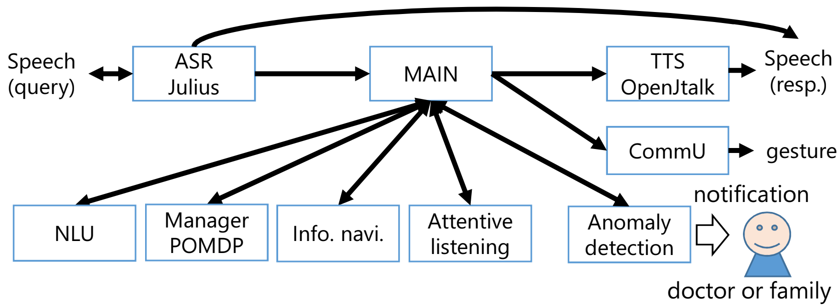

Fig. 1 System architecture.

In this work, we built a dialogue robot for watching aged people through daily conversations. The system has functions for preventing aged people from being socially isolated. The system is implemented in communicative robot called "CommU[1]" with a speech interface. The system has functions of information navigation, attentive listening, and anomaly detection, as shown in Fig. 1.

Some existing works built dialogue systems on robots for social conversations [8, 13]. These works tried to build systems in typical social roles to realize natural conversations. On the other hand, some works of dialogue systems tried to build dialogue functions for helping aged society [17, 5, 12]. Our work builds a dialogue robot in some social roles with aged people according to acquired knowledge of these existing works. Our system also tries the detection of anomaly caused by dementia.

2 System Architecture

2.1 Dialogue scenario

The system has two dialogue functions: "information navigation" and "attentive listening." The system starts a dialogue with the information navigation function; the function initiates dialogue and informs the user recent news written in web articles in the manner of an information-navigation dialogue [19]. The system moves to attentive listening mode that listening to the user opinion on the current topic, when the focus detection predicts higher user's interest. The focus detection module is implemented by conditional random fields (CRF). The system continues to predict the user's interest during the attentive listening mode by using focus detection and dialogue act classification modules based on logistic regression [18]. Once the system predicts that the user wants to know other news, the system proposes another piece of news of the day by using the information navigation module again. Before

[1] https://www.vstone.co.jp/products/commu/index.html

the topic transition, the system inserts a question generated by the anomaly detection module. The answer to the question is used to detect anomalies, particularly dementia in our case.

2.2 Dialogue data

We collected dialogue data of human-human dialogue following our dialogue scenario of information navigation and attentive listening [20]. A dialogue was conducted between an aged person and a person who listened the aged person, that is, a professional counselor, a professional care-taker, or a student. Sixty dialogues were collected in total from 24 aged people, who are more than 65 years old, and 15 listeners. Each utterance was transcribed and annotated with dialogue acts to use in training the dialogue modules as follows.

2.3 Speech interfaces

The dialogue system has speech interfaces including automatic speech recognition (ASR) and text to speech (TTS). We used Julius[2] [9] as an ASR decoder by adapting its language model with the collected dialogue data. Julius works as a server mode in deep neural network-hidden Markov model (DNN-HMM) mode and controls any speech activities of the user with its voice activity detection (VAD).

Open JTalk[3] [7] was used to generate the speech output of the system. A model of Open JTalk was first trained from the reading-style voice [14] and then adapted using the voices of a professional counselor, who is working in the area of care-taking for aged people, to build synthetic sounds that are easy for aged people to catch.

2.4 Natural language understanding

The user utterances transcribed by the ASR module are sent to the natural language understanding (NLU) module to extract information, which is necessary for deciding the next action of the system. Our NLU module consists of two functions: focus detection and dialogue act classification. The focus detection module detects whether an utterance made by the user contains any focus words, "an attentional state contains information about the objects, properties, relations, and discourse intentions that are most salient at any given point" [6, 18], or not, by using CRF. The

---

[2] http://julius.osdn.jp/

[3] http://open-jtalk.sourceforge.net/

dialogue act classifier selects a dialogue act for the user utterance, that is, what the intention is of the user in the dialogue level, based on logistic regression. Defined classes are based on our previous work [18].

2.5 Dialog manager

Once the NLU module predicts the focus and the dialogue act of the user utterance, these states are used to decide the next system actions. While the system is working on the "information navigation" function, the system action is decided by a policy function trained under a partially observable Markov decision process (POMDP). The policy decides the actions of the system from three actions: topic presentation (topic changing), storytelling and question answering. The policy is trained on annotated training data and it decides an action using information form NLU modules.

When the system changes the topic, the system randomly generates a question that is used to detect anomalies. Our system randomly selects a question from a question set for predicting dementia [16, 15]. The answer of the aged person is sent to the anomaly detection server, and the result is sent to people who monitor the elderly, e.g., medical doctor or family.

2.6 Information navigation module

Utterances generated in information navigation mode are based on our previous study [19]. The system can provide information written in web news text (news of the day) with the following functions.

- Topic presentation: Description of available news (topics) including topic changes.
- Storytelling: Description of a summary of the current topic. The description is created from headers of web news articles.
- Question answering: Answering user's question.
- Proactive presentation: Additional information provision to previous system utterance. The sentence is generated from related news articles to the current article (topic).

We expect this module to talk with the user on a variety of domains, to prevent the isolation from the society.

2.7 Attentive listening module

Attentive listening mode is implemented to elicit positive emotions from the user by following our previous work [10, 11]. The system selects a response from a response pool, which is constructed from the training dialogue data. Our dialogue data is annotated with dialogue acts; thus, utterances annotated as feedback functional utterances, for example, short feedback utterances such as “I see” or “Right”, are stored in the response pool. These utterances are also annotated with emotional impacts, that is, the degree of the effect on changing the emotion of the dialogue partner. The system select a response from the response pool by considering the connection to the contexts and the emotional impact.

2.8 Anomaly detection

For the anomaly detection function, we focused on dementia. Predicting dementia from old response of old people is highly demanded on the aged society. It is crucial to predict dementia in the early stages because dementia is progressive. Dementia patients show signs in their communication even if the problem is at an early stage; however, the discovery of such symptoms from aged people in solitude is often delayed because they have fewer interactions with others than aged people living with families or other people.

Recently, there have been pieces of research on predicting early-stage dementia through human-agent interactions by asking questions [16, 15]. These works use computer avatar with spoken dialog functionalities that produces spoken queries, and try to predict dementia tendency using language and speech features extracted from user utterances by using support vector machines or logistic regressions. We implemented the dementia detection module by inserting typical questions randomly during topic transition in dialogues because it is expected that dementia can be predicted through natural conversations. Responses to these questions from aged people are sent to the dementia detection module and used for detecting. If the system predicts any indications of dementia, it will notify medical doctors or families.

2.9 Communication robot

We implemented our system in a communication robot, CommU, since it has high expressiveness and familiarity. We set action patterns for the dialogue functions to encourage positive dialogue. These patterns were manually designed.

Acknowledgement

This work is supported by JST PRESTO (JPMJPR165B) and JST CREST (JP-MJCR1513).

References

1. Bemelmans, R., Gelderblom, G.J., Jonker, P., De Witte, L.: Socially assistive robots in elderly care: A systematic review into effects and effectiveness. Journal of the American Medical Directors Association **13**(2), 114–120 (2012)
2. Broekens, J., Heerink, M., Rosendal, H., et al.: Assistive social robots in elderly care: a review. Gerontechnology **8**(2), 94–103 (2009)
3. Callejas, Z., López-Cózar, R.: Designing smart home interfaces for the elderly. SIGACCESS Access. Comput. (95), 10–16 (2009)
4. Feil-Seifer, D., Mataric, M.J.: Defining socially assistive robotics. In: Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on, pp. 465–468. IEEE (2005)
5. Georgila, K., Wolters, M., Moore, J.D., Logie, R.H.: The match corpus: a corpus of older and younger users interactions with spoken dialogue systems. Language Resources and Evaluation **44**(3), 221–261 (2010)
6. Grosz, B.J., Sidner, C.L.: Attention, intentions, and the structure of discourse. Computational linguistics **12**(3), 175–204 (1986)
7. Group, H.W., et al.: The Japanese TTS system. Open JTalk (2015)
8. Jokinen, K.: Dialogue models for socially intelligent robots. In: International Conference on Social Robotics, pp. 127–138. Springer (2018)
9. Lee, A., Kawahara, T., Shikano, K.: Julius—an open source real-time large vocabulary recognition engine. In: Seventh European Conference on Speech Communication and Technology (2001)
10. Lubis, N., Sakti, S., Yoshino, K., Nakamura, S.: Eliciting positive emotion through affect-sensitive dialogue response generation: A neural network approach. In: AAAI (2018)
11. Lubis, N., Sakti, S., Yoshino, K., Nakamura, S.: Eliciting positive emotional impact in dialogue response selection. In: Advanced Social Interaction with Agents, pp. 135–148. Springer (2019)
12. Miehle, J., Bagci, I., Minker, W., Ultes, S.: A social companion and conversational partner for the elderly. In: Advanced Social Interaction with Agents, pp. 103–109. Springer (2019)
13. Milhorat, P., Lala, D., Inoue, K., Zhao, T., Ishida, M., Takanashi, K., Nakamura, S., Kawahara, T.: A conversational dialogue manager for the humanoid robot erica. In: Advanced Social Interaction with Agents, pp. 119–131. Springer (2019)
14. Sonobe, R., Takamichi, S., Saruwatari, H.: Jsut corpus: free large-scale japanese speech corpus for end-to-end speech synthesis. arXiv preprint arXiv:1711.00354 (2017)
15. Tanaka, H., Adachi, H., Ukita, N., Ikeda, M., Kazui, H., Kudo, T., Nakamura, S.: Detecting dementia through interactive computer avatars. IEEE journal of translational engineering in health and medicine **5**, 1–11 (2017)
16. Ujiro, T., Tanaka, H., Adachi, H., Kazui, H., Ikeda, M., Kudo, T., Nakamura, S.: Detection of dementia from responses to atypical questions asked by embodied conversational agents. In: Proc. Interspeech 2018, pp. 1691–1695 (2018)
17. Wolters, M., Georgila, K., Moore, J.D., MacPherson, S.E.: Being old doesn’t mean acting old: How older users interact with spoken dialog systems. ACM Trans. Access. Comput. **2**(1), 2:1–2:39 (2009)
18. Yoshino, K., Kawahara, T.: Information navigation system based on pomdp that tracks user focus. In: Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pp. 32–40 (2014)
19. Yoshino, K., Kawahara, T.: Conversational system for information navigation based on pomdp with user focus tracking. Computer Speech & Language **34**(1), 275–291 (2015)
20. Yoshino, K., Tanaka, H., Sugiyama, K., Kondo, M., Nakamura, S.: Japanese dialogue corpus of information navigation and attentive listening annotated with extended iso-24617-2 dialogue act tags. In: LREC (2018)