Open-technology oriented Japanese voice dialogue system

Kazuki SUZUKI（鈴木和樹）

In AI-based services, the technology is closed and the algorithms developed by companies tend to be kept secret. This is a barrier to social implementation even for services with a wide range of applications such as dialogue systems. However, there is a growing trend among some companies to open up their technologies to the public, for example, by releasing the language models they have learned. In this research, we define ‘‘Open- technology oriented’’ procedure as the combination of such open technologies and original algorithms to increase the openness.

The purpose of this study is to construct an open-technology oriented Japanese voice dialogue system and to clarify its usefulness. Specifically, we will verify the technology of dialogue management, which greatly affects the performance of a dialogue system, and develop a system that can be provided as a service. We have divided the dialogue management into two dialogue engines and verified their technologies. Figure 1 shows a schematic diagram of the Japanese voice dialogue system developed in this study. This system has been developed based on an open-technology oriented. While there are various possible domains to which the dialogue system can be applied, we chose a call center of a car rental service to be a specific domain. In addition, we aimed to make the system applicable to other topics as well. The dialogue system is divided into three main functions. (1) speech recognition, (2) dialogue management, and (3) speech synthesis. For (1) and (3), we used publicly available APIs (Microsoft Azure's Speech Services). We thought that a more flexible dialogue system could be realized by combining task-oriented dialogue and non-task-oriented dialogue engines. Specifically, we combined task-oriented dialogue using BERT[1], which has been attracting attention in the field of natural language processing, and non-task-oriented dialogue using Reformer[2], which is a relatively new machine learning technology capable of generating dialogue.

For the task-oriented dialogue engine, two techniques were tested: "Word2vec + Random Forest" and BERT. We compared the dialogue classification accuracy and found that BERT improved the accuracy by 25 points. We also developed a filtering function using NSP with BERT. The performance evaluation showed that BERT can filter correctly with about 75% accuracy.

In constructing the non-task oriented dialogue engine, we adopted a publicly available corpus of chatty dialogues as a dataset (Table 1) and Reformer as the machine learning algorithm. The dataset is a set of dialogue patterns between persons and the system with annotation. Since the accuracy of the initial data set is not high, we abstracted understandable dialogue patters from the dataset to improve the accuracy. As a result, we could improve slightly the accuracy of dialogue generation. Next, we compared the accuracy of dialogue generation by changing hyperparameters during reformer training. However, we were not able to achieve highly flexible dialogues using the Reformer. In the future, we will aim to further improve the accuracy of the dialogue generation model using Reformer.

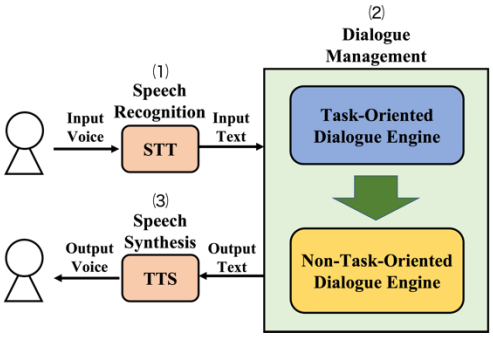

Figure 2 : System overview

Table 5 : Dataset (excerpt)

System	User
はいおはよー	音楽はお好きですか
心から音楽を楽しんでるんです	音楽はいいですよね
音楽はライブがいいですね	ライブにはあまり行かないです

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova : "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (2018)

[2] Nikita Kitaev, Łukasz Kaiser, Anselm Levskaya, Reformer: “The Efficient Transformer” (2020)