Technology Reports
Dialogue Systems | Chat-oriented Dialogue | Chatbots
Special Articles on AI Supporting a Prosperous and Diverse Society

Highly Customizable Chat-oriented Dialogue System

Service Innovation Department Yuiko Tsunomori Kanako Onishi

Most conventional casual chat-oriented dialogue systems have had limitations, such as only being able to respond within a specific domain, being difficult to customize, or being limited to certain use cases. NTT DOCOMO has developed a chat-oriented dialogue engine that is able to eliminate utterances that are not appropriate to the use case by assigning a particular linguistic style for a character to the system and enabling the priorities of the system utterance generator to be changed. This enables the chatbot to be customized to suit the use case.

1. Introduction

Conversation agents, such as smart speakers and the "my daiz*1" application, are becoming popular recently. Most of these agents respond to user input that has some kind of intent (task), such as "Please set an alarm," or "What is the weather like today?" The ability to request tasks through dialogue is extremely convenient for users. NTT DOCOMO released such a voice agent application in March 2012, called "Shabette Concier," and it has been very popular, attracting large numbers of users. The main purpose of this application is to respond to task-oriented user input, but not all input received has been of this type. A large amount of chat-oriented dialogue has actually been received. Unfortunately, Shabette Concier does not have functionality to respond sensibly to chat-oriented dialogue, so it is not very satisfying for users in such cases. As such, NTT DOCOMO developed a chat-oriented dialogue Application Programming Interface (API)*2 based on technology from the NTT

*1 my daiz: A speech dialogue agent that runs on smartphones and tablets, providing a wide range of information suited to the user.
*2 API: An interface that enables software functions to be used by another program.

Media Intelligence Laboratories to respond to the users' desire for chat-oriented dialogue. It has been available on the docomo Developers support site [1] since 2013.

Generally, toys and robots that incorporate chat-oriented dialogue systems are intended for use in various use cases, so there is good potential to realize ongoing, engaging conversation that will prompt users to continue using them for a long time. However, most earlier chat-oriented dialogue systems could only respond within a specific domain (range of conversation topics), were difficult to customize, or were limited to certain use cases. For example, it may be desirable to avoid difficult topics in toys for children, but a filter on system utterances appropriate to the use case is difficult to implement. The chat-oriented dialogue API also experienced similar issues.

As such, NTT DOCOMO developed a chat-oriented dialogue engine that can be customized according to the use case. The chat-oriented dialogue engine is part of a common platform called the natural dialogue platform. It realizes open-domain dialogue and is able to handle a wide range of topics, with utterances generated from large amounts of data on the Web. It is also highly customizable and can be used in all kinds of use cases. Specifically, it can be customized by assigning a linguistic style, by avoiding utterances not suitable to the use case, and by setting priorities for the system utterance generator.

This article describes technologies used to implement the highly customizable chat-oriented dialogue engine and introduces an application example called "kataraiⓇ*3", which is a chat-oriented dialogue service utilizing the agent.

2. Natural Dialogue Platform Overview

The architecture of the natural dialogue platform is shown in **Figure 1**. It consists of four basic engines: "Scenario dialogue," "Intention interpretation*4", "Knowledge Q&A," and "Chat-oriented dialogue," and can realize all kinds of dialogue by freely combining each of these engines. Parts of this platform are published as xAIML SUNABA [2], in the form of a descriptive language specification and development environment.

1) The scenario dialogue engine implements dialogue between the user and the system according to a scenario prepared beforehand. Dialogues with a story line are realized by preparing system utterances that match with user utterances beforehand. More complex dialogue scenarios can also be described by linking with external services using the external link functionality. Dialogue scenarios are described using xAIML, an NTT DOCOMO extension to Artificial Intelligence Markup Language (AIML)*5 [3], a language for describing software agents. xAIML is able to perform more flexible matching by describing conditional branches, and by normalizing and finding superordinate concepts for sentences.
2) The intention interpretation engine automatically classifies user utterances, including ambiguous expressions, into utterance intentions called "tasks" (e.g.: "weather" or "news"). It is also able to extract information needed for each task from the user's utterance (e.g.: location, time and date, etc.).
3) The knowledge Q&A engine [4] uses databases and other sources to respond to user utterances

*3 kataraiⓇ: A trademark or registered trademark of NTT DOCOMO Corp.

*4 Intention interpretation: Technology that uses machine learning and so forth to determine the user's intention from the user's utterances (natural language). User intentions are called "tasks." For example, all the utterances "What's tomorrow's weather?," "I wonder if tomorrow will be fine?," and "Is it going to rain tomorrow?" are judged as weather tasks.

*5 AIML: A description technique for constructing an interactive agent.

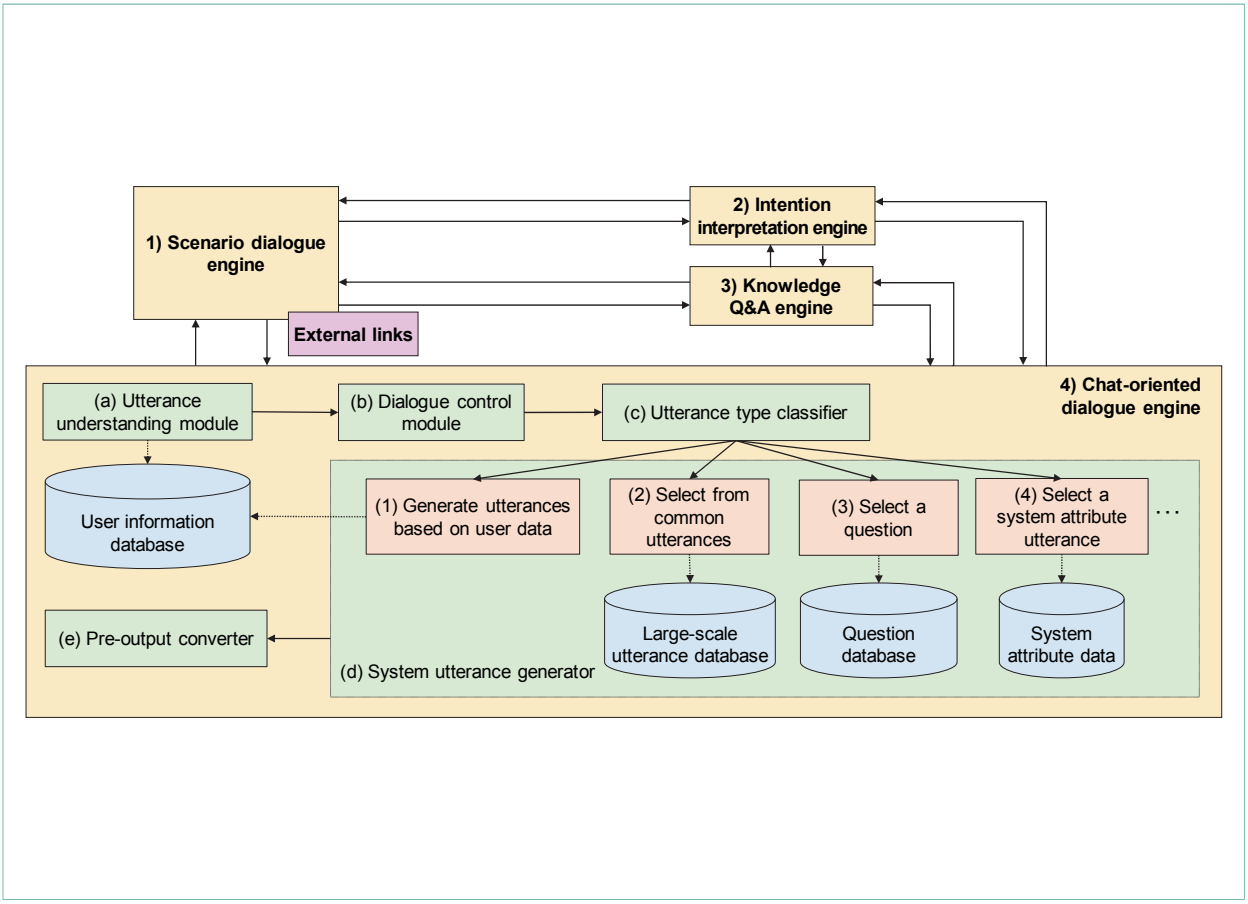

Figure 1 Natural dialogue platform architecture

asking general knowledge questions. For example, if a user inputs "What is the height of Mt. Fuji?", the system will respond with "3,776 m."

4) The chat-oriented dialogue engine is described below.

These engines are linked together to realize more-natural dialogue.

3. Chat-oriented Dialogue Engine

This section describes the processing sequence and customization features of the chat-oriented dialogue engine.

3.1 Processing Sequence

The chat-oriented dialogue engine processes input user utterances as follows.

(a) The utterance understanding module analyzes user input sentences, and infers focus points (words that express the topic) and dialogue acts*6 [5]. If the user's utterance also includes information about the user (interests, preferences, etc.), this information is also extracted and stored in the user information database [6].

*6 Dialogue acts: The type of utterance, according to the speaker's intention. E.g.: "Empathize," "Question," etc.

(b) The dialogue control module decides the dialogue act that the system will output next, based on the history of preceding utterances.

(c) The utterance type classifier decides which utterance module will be used to generate the next system utterance, using both rules and a machine learning[*7] model.

(d) The system utterance generator accommodates several modules and the module selected by the utterance type classifier decides what system utterances to output.

Here, we describe four of the many modules available.

(1) Generate utterances based on user data

Outputs an utterance using information stored in the user information database (e.g.: "Come to think of it, you like reading, don't you? How about reading a favorite book to relax?").

(2) Select from common utterances

Selects an utterance from a large utterance database by combining focus points and dialogue acts that the system needs to output next (e.g.: "Strawberries are delicious."). Note that the large database is composed of utterances linked with focus points, and contains approximately 40 million entries. A large volume of Web data was analyzed, selecting focus points and associated noun-predicate pairs that represent "who did what." Noise was eliminated by only using pairs that appear more frequently than a set threshold. These noun-predicate pairs were then converted to declarative[*8] sentences and to utterances that conform to each of the dialogue acts, and then stored in the utterance database [7].

(3) Select a question

Questions about the user are asked (e.g. "What are your hobbies?") to get user information or to change topics.

(4) Select a system attribute utterance

For user utterances asking for system information (e.g.: "What's your name?"), utterances are generated using system attribute data (e.g.: "My name is Mariko.").

(e) The pre-output converter modifies the selected system utterance in terms of inflection or special vocabulary of the specified linguistic style [8], and outputs the system utterance.

3.2 Customization Features

By customizing the engine, a chatbot[*9] capable of chat-oriented dialogue suitable for the use case can be developed. Customization features are described below.

1) Assigning Linguistic Styles (System Attribute Data, Pre-output Converter)

By providing pre-defined system attribute data, character profile data can be used in utterances. If text related to the character is available (e.g.: character blog articles, etc.), utterances can be generated from it automatically and added to the utterance database, increasing the possibility that topics linked to the character will be supported. The pre-output converter is also able to alter system utterances according to the speaking style of the specified character.

2) Eliminating Unsuitable Utterances (Large-scale Utterance Database)

The definitions of utterances that are unsuitable

*7 Machine learning: A framework that enables a computer to learn useful judgment standards through statistical processing from sample data.

*8 Declarative: A statement that includes a subject and verb. A declarative is not a question, command or an exclamation.

*9 Chatbot: A program that automatically conducts dialog with people with speech or text chat.

will differ depending on the use case. In technical collaboration with the NTT Media Intelligence Laboratories, NTT DOCOMO has developed technology to attach labels to sensitive information in focus points and system utterances associated with them. By "sensitive," we mean system utterances that may not be desirable, such as utterances that could violate ethical standards, taboos or other sensitive topics. The engine applies sensitive data labels to system utterances stored in the utterance database, decides which labels to remove based on the use case, and removes the utterances associated with those labels from the database.

3) Changing the Priorities of the System Utterance Generator (Utterance Type Classifier)

The system utterance generator module decides which system utterance to output next, and the priorities for utterance type class can be adjusted. It is also able to concatenate utterances selected from multiple modules and output them as one system utterance.

4. The "katarai" Chat-oriented Dialogue Service

The chat-oriented dialogue engine has been used for the "katarai" [9] commercial service. katarai is provided as an API that can engage in chat-oriented dialogue, and can produce more natural conversation by combining it with the scenario dialog engine and the knowledge Q&A engine. By using the customization functions of the chat-oriented dialogue engine, katarai can be used in a wide range of use cases. Some use cases for katarai are described below.

4.1 The ASTRO BOY Communication Robot

The ASTRO BOY communication robot (hereinafter referred to as "ASTRO BOY") sold by Kodansha Co. Inc. is a robot capable of speech conversation (**Photo 1**). ASTRO BOY is able to engage in conversation when not connected to the Internet, but can converse on a much broader range of topics using katarai when connected to the Internet. Conversation while connected to the Internet is implemented using a scenario dialogue engine and katarai. For example, if the user says, "Can you tell me some popular buzzwords?" ASTRO BOY will ask the user what year to find buzzwords for, and if the user says "2000," he can talk about buzzwords from that year. However, if the user then says "So, the Olympics are almost here!" the

Photo 1 ASTRO BOY communication robot (©TEZUKA PRO / KODANSHA)

scenario does not suggest what should come next, so ASTRO BOY cannot respond based only on the scenario. For input that is not prescribed by scenario, ASTRO BOY uses katarai to respond. The wide range of scenarios and katarai enable ASTRO BOY to respond to a wide range of utterances.

4.2 NTT DOCOMO "Onshoko-roid" Online Shop

The inquiries system installed in the NTT DOCOMO online shop has a character called "Onshoko-roid," who can answer questions related to the online shop, such as "Can I make a reservation?" or "Can you tell me about the student discount?" (**Figure 2**). The NTT DOCOMO FAQ chatbot is used to answer questions, but users often enter utterances not anticipated by the FAQ chatbot, such as "I'm hungry," or "I want a hamburger." By combining the FAQ chatbot with katarai, questions that the FAQ chatbot cannot answer can be answered by katarai. In this kind of application, it is common to associate character settings such as an icon with the FAQ chatbot, so often there are also questions about the character profile, such as its name or interests. By introducing katarai, these sorts of questions can also be answered appropriately.

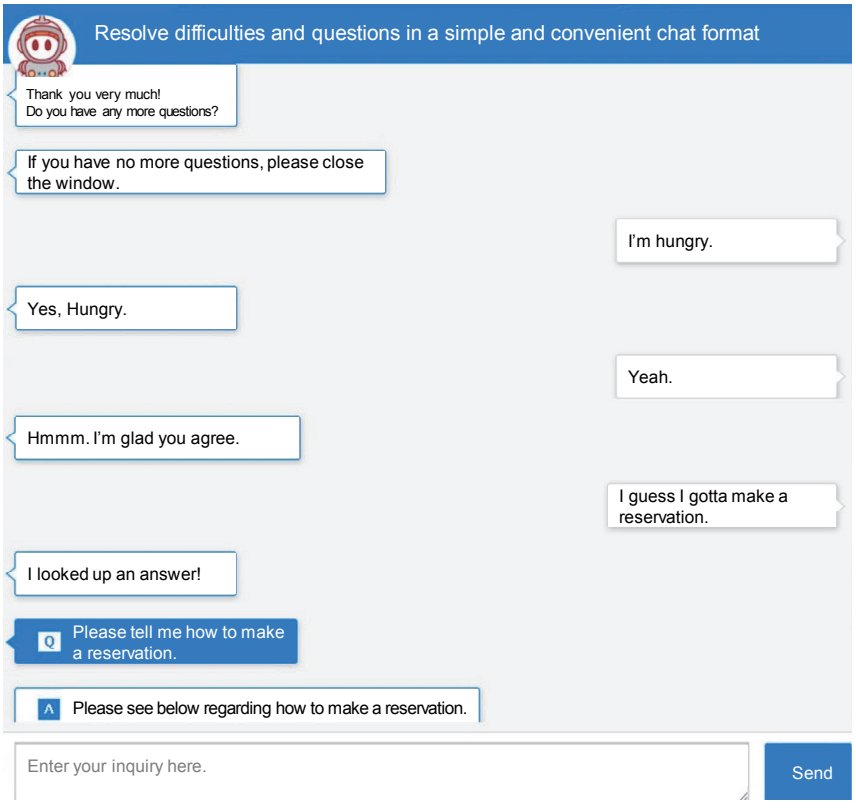

Figure 2 DOCOMO online shop, "Onshoko-roid"

5. Conclusion

This article has described a customizable chat-oriented dialogue engine technology. The engine is able to respond to all kinds of user input by building a system utterance database using large amounts of data from the Internet, and it can be customized to suit the use case. In particular, it can be customized by eliminating utterances that are not appropriate for the use case or for an assigned character, and by setting priorities in the system utterance generator module.

The "katarai" commercial service is already deployed and using the chat-oriented dialogue engine, and is able to respond to various use cases through use of the customization features. In the future, we intend to continue improving the chat-oriented dialogue engine to conduct even more natural dialogue, selecting issues from the real services where it is being used.

REFERENCES

[1] docomo Developers support Web site. https://dev.smt.docomo.ne.jp/
[2] xAIML SUNABA Web site. https://docs.xaiml.docomo-dialog.com
[3] R. S. Wallace: "The anatomy of A.L.I.C.E.," Parsing the Turing Test, Springer, 2009.
[4] W. Uchida et al.: "Knowledge Q&A: Direct Answers to Natural Questions," NTT DOCOMO Technical Journal, Vol.14, No.4, pp.4–9, Apr. 2013.
[5] T. Meguro, R. Higashinaka, K. Dohsaka and Y. Minami: "Creating a Dialogue Control Module for Listening Agents Based on the Analysis of Listening-oriented Dialogue," IPSJ Journal, Vol.53, No.12, pp.2787–2801, Dec. 2012 (In Japanese).
[6] T. Hirano, N. Kobayashi, R. Higashinaka, T. Makino and Y. Matsuo: "User Information Extraction for Personalized Dialogue Systems," Proc. of SemDial, 2015.
[7] C. Miyazaki, T. Hirano, R. Higashinaka and Y. Matsuo: "Towards an Entertaining Natural Language Generation System: Linguistic Peculiarities of Japanese Fictional Characters," Proc. of SIGDIAL, 2016.
[8] R. Higashinaka, K. Imamura, T. Meguro, C. Miyazaki, N. Kobayashi, H. Sugiyama, T. Hirano, T. Makino and Y. Matsuo: "Towards an open-domain conversational system fully based on natural language processing," Proc. of COLING, pp.928–939, Aug. 2014.
[9] katarai Web site. https://www.katar.ai/