Technology Reports (Special Articles)

`Chatbot` `Character` `Word Play`

Special Articles on AI for Improving Services and Solving Social Problems

Technologies for Achieving Entertaining Dialogue and Friendly Chatbots

Service Innovation Department **Yuiko Tsunomori   Yutaro Shiramizu**

Recent years have seen increasing demand for chatbots in all sorts of applications. In dialogue with chatbots whose purpose is essentially communication, it is important that the user finds that dialogue to be friendly and entertaining. To this end, we developed technology for reproducing specific characters and technology for engaging in word play such as *nazokake* riddles. These technologies make it possible to give chatbots an entertaining nature. We also developed a dialogue application running on a chat tool as a means of providing entertaining dialogue.

1. Introduction

Chatbots such as smart speakers and dialogue agent applications are becoming increasingly popular. Most of these accept various types of task requests such as "Please set an alarm" or "What will the weather be like today?" These chatbots are designed with this in mind, that is, to carry out a requested task, but in the case of "Shabette Concier," a voice agent released by NTT DOCOMO in 2012, many user utterances came to be input not with the aim of requesting a task but simply to communicate, i.e., to carry on a conversation. To respond to such user input, NTT DOCOMO developed a chat-oriented dialogue Application Programming Interface (API)[*1] based on technology from NTT Media Intelligence Laboratories and released it for public use on the docomo Developers support (dDs) [1] website in 2013. Then, in 2018, NTT DOCOMO released the chat-oriented "katarai"

All company names or names of products, software, and services appearing in this journal are trademarks or registered trademarks of their respective owners.

*1 API: An interface that enables software functions to be used by another program.

[2] service developed on the basis of this chat-oriented dialogue API in collaboration with Intermedia Planning, Inc. Rather than carrying out tasks, the purpose of this API and service is to communicate with the user, so it is important that such dialogue be friendly and entertaining to the user.

To achieve dialogue of this type, we developed character chat technology for reproducing specific characters and word play technology that can achieve chatbot dialogue having game-like characteristics as in posing riddles. These technologies make it possible to give chatbots an entertaining nature that makes them friendly and approachable to users. Additionally, to make it easy for users to enjoy dialogue having an entertaining nature (hereinafter referred to as "entertaining dialogue"), we developed an entertaining dialogue application that runs on a chat tool.

In this article, we describe the above character chat technology, word play technology, and entertaining dialogue application and present application examples.

2. Character Chat Technology

2.1 Overview

There is a high possibility that giving chatbots personalities can achieve dialogue that is friendly to users. For example, creating a character that a user likes or an original character can arouse a user's interest.

Based on technology [3] from NTT Media Intelligence Laboratories, character chat technology reproduces the dialogue of arbitrary characters. This technology creates utterance-pair data (questions and answers) in the character style desired

and constructs a dialogue system based on that data. In addition to constructing a new dialogue system, this technology can give personality to an existing dialogue system by combining the two systems.

Character chat technology achieves high-accuracy dialogue that reproduces a character at low cost with real-time characteristics. In use, it first extracts candidates by searching an utterance-pair database and then decides on a final utterance based on a score calculated by a variety of techniques. In this way, it can narrow down the candidates for an answer obtained from the initial search and achieve dialogue with a real-time feel even without a high-performance server.

2.2 Process Flow

The system configuration of character chat technology is shown in **Figure 1**. The system performs the following processing on input of a user utterance.

First, the utterance-pair search block searches for and extracts multiple answer candidates with respect to the input user utterance from an utterance-pair database that holds utterance pairs (question-answer pairs) for an arbitrary character. Specifically, this block searches the database for questions similar to the input user utterance and extracts a set of question-and-answer pairs as candidates corresponding to the user's question.

Next, the utterance-score calculation block calculates multiple scores using a variety of techniques including a translation model and word2vec[*2] model. In general, a translation model learns using a bilingual corpus[*3] that assigns a correspondence between sentences in two languages (such as Japanese

---

*2 word2vec: A technique that analyzes text data and represents the meaning of each word in vector form.
*3 Corpus: A language resource consisting of a large volume of text and utterances, etc. collected and stored in a database.
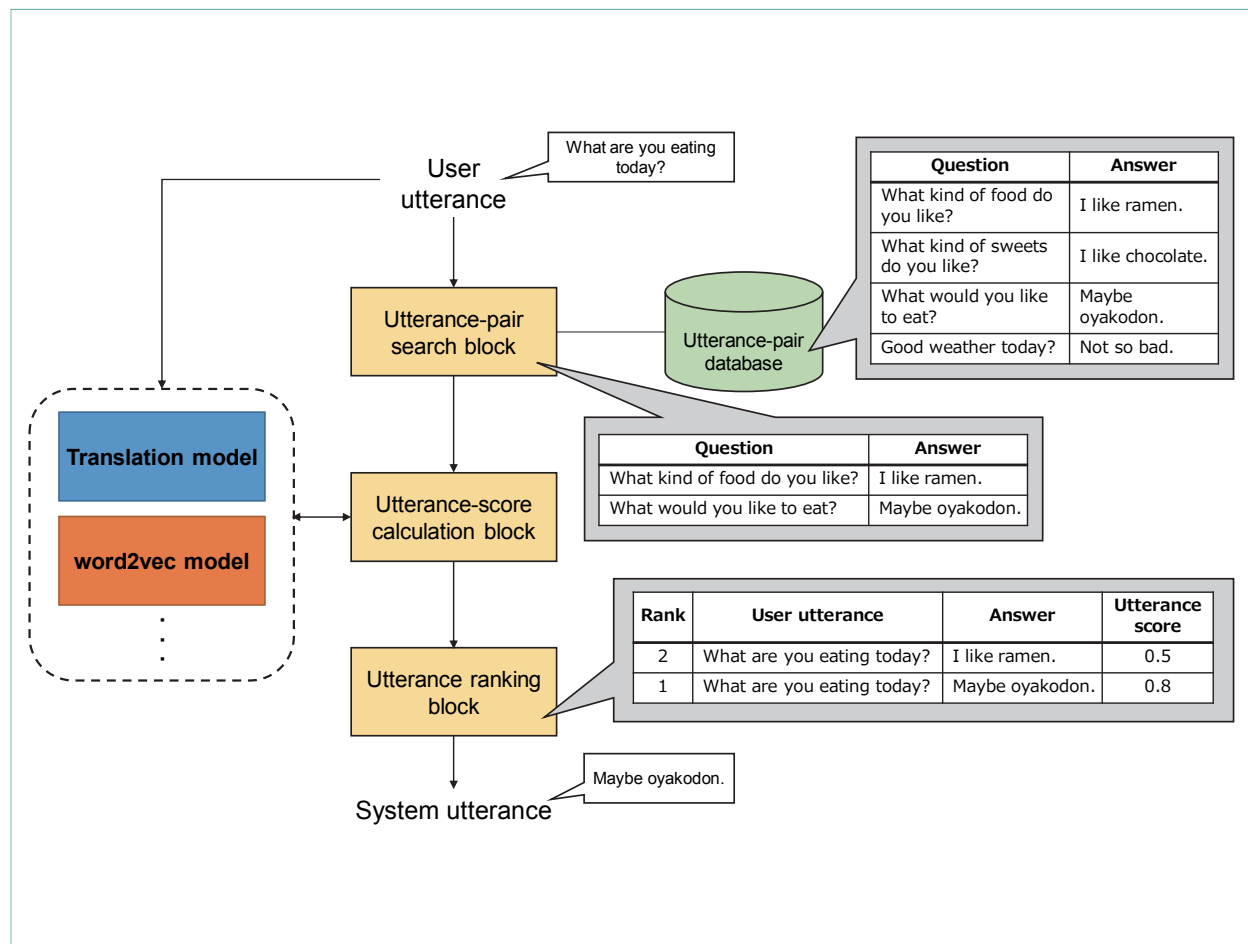
Figure 1 System configuration of character chat technology

and English). Our translation model, however, learns using a large volume of utterance pairs that assigns a correspondence between questions and answers. Specifically, for each question-answer pair consisting of the input user utterance and an answer candidate received as input, this translation model calculates a score on the naturalness of that user-utterance/answer-candidate pair. The word2vec model calculates a score on the similarity between the input user utterance and answer candidate.

Finally, the utterance ranking block integrates multiple scores and performs utterance ranking.

The utterance with the highest score is output as the final system utterance (answer).

2.3 Application Example

"AI Jimmy" is the first commercial service to make use of character chat technology. Scheduled for provision by Laugh & Peace Mother Co. Ltd., AI Jimmy is an application that enables voice interaction with Artificial Intelligence (AI) that reproduces the Japanese entertainer Mr. Jimmy Onishi (**Figure 2**). That is, it can achieve dialogue that reproduces Mr. Jimmy Onishi by creating

Figure 2   AI Jimmy

and learning question-answer pairs by actually having Mr. Jimmy Onishi himself give answers to a variety of questions. For example, if the user inputs the question "How are you?" the application would respond in Mr. Jimmy Onishi's way of speaking by outputting "Me? I'm doing great—I can't stop!" Character chat technology can make answers sound all the more like Mr. Jimmy Onishi.

Voice interaction in AI Jimmy is achieved by using a DOCOMO AI Agent API and a speech synthesis model developed by NTT Media Intelligence Laboratories in addition to character chat technology. This DOCOMO AI Agent API features speech-recognition and speech-synthesis functions that enable voice interaction with the user. Here, speech synthesis in the DOCOMO AI Agent API makes use of an original speech-synthesis model developed in collaboration with NTT Media Intelligence Laboratories and trained with the speech of Mr. Jimmy Onishi.

3. Word Play Technology

3.1   Overview

Giving chatbots dialogue functions with game-like characteristics has a high possibility of heightening user interest and giving the user a feeling of "I'd like to talk more." This technology is being used to provide the four games of Yamanote Line, *Hyakunin Isshu* (traditional Japanese card game), *nazokake* riddles, and quiz.

1) Yamanote Line

The Yamanote train line is the famous loop line of Tokyo. Here, Yamanote Line is an interactive game based on specific themes such as "names of stations of the Yamanote line" in which the user gives answers such as "Shinjuku" and "Shibuya." You lose by duplicating an answer or giving an incorrect answer.

One issue that arises here when the user inputs an answer is "orthographical variants" (different

notational forms). For example, if the user inputs the Japanese characters "新宿三丁目駅" corresponding to "Shinjuku-sanchome Station" in English (where the Chinese character 駅 = station), we can envision other forms of input with the same meaning such as "新宿三丁目" where the suffix駅/station is omitted or "新宿3丁目駅" where the Chinese numeral 三 meaning "three" is replaced by the Arabic numeral "3." We deal with this problem by automatically collecting information on orthographical variants from Wikipedia and regularly updating our dictionary. Additionally, as one means of enhancing game-like characteristics in dialogue, we establish an order for system answers according to the degree to which a name is known (name recognition) as calculated from the number of Wikipedia page views, link frequency, etc. For example, if the theme is "names of stations of the Yamanote line" as described above, the system would first answer with station names having relatively high name recognition such as "Shinjuku" and "Shibuya" and answer later with station names having relatively low name recognition such as "Tamachi."

2) *Hyakunin Isshu*

Hyakunin Isshu is a traditional Japanese card game based on a classical collection of one hundred short poems (each five lines in length) by one hundred poets called *Ogura Hyakunin Isshu*. In the basic format of this game, someone reads out the first three lines of one of those poems and the player looks for and picks up the card corresponding to the last two lines of that poem from among a set of cards each written with the last two lines of a different poem.

This word play function operates in a similar way. The system presents the user with a "reading card" showing the first half of a poem and multiple "pickup" cards showing the last half of different poems including the one corresponding to the first half of the poem shown on the reading card. The user selects one of those pickup cards presented by the system, which then informs the user whether the card selected showing the last half of a poem corresponds to the reading card showing the first half of a poem, i.e., whether the two cards correspond to the same poem (right/wrong answer).

In the "my daiz" member[*4] "Word Play" described later, the system displays a reading card while reading out the lines on that card by speech synthesis. Here, however, instead of directly uttering the text displayed on the reading card (written in historical kana orthography), a developer first makes some adjustments to that character string (such as replacing words with others having the same reading, inserting commas, etc. to enhance or diminish intonation) to achieve a natural card reading appropriate to *Hyakunin Isshu* using an existing speech synthesis function.

3) *Nazokake* Riddles

A *nazokake* riddle has the format "Given two words X and Y that are apparently different in meaning, what do they have in common? The answer is Z." In the "katarai" service, the user can choose one of four random words displayed by the system or input any word desired as theme X to enjoy a *nazokake* riddle. The construction of this game is based on text data collected from Social Networking Services (SNSs) using posts that have been converted to an "X, Y, Z" format consisting of three items of data.

---

*4 **my daiz member:** An extension function (service) that can be added to my daiz (see *5).

4) Quiz

We developed a quiz function as an interactive game that can be played as much as one likes. In this game, the system presents a random problem and four options as answers combining the correct answer and incorrect answers. The user may ask for a hint with respect to the given problem if needed. If the user answers correctly, the system assigns a score to that answer according to the difficulty of the problem.

In the "Word Play" member described below, total scores can be compared (ranked) among users by fixing the problems prepared in a certain period and tabulating the scores of each user.

3.2 Application Example

The "my daiz*5" speech dialogue agent offers "Word Play" as an interactive game using this technology (**Figure 3**). Although the main function of this agent is to provide information such as train service updates and weather, it includes an interactive game function with the aim of breaking down any psychological resistance that users may have about using agents and getting them to enjoy interaction with agents.

The "my daiz" member Word Play enables button-based operations by a Graphical User Interface (GUI)*6 as an original feature in addition to voice/text-based operations. This feature reduces input mistakes caused by speech-recognition errors,

Figure 3   Word Play

typing errors, etc. In addition, the quiz function adds mechanisms for improving the user retention rate such as applying a limit to the number of answers that can be given per day and assigning a score according to the level of difficulty of each question. The *Hyakunin Isshu* function, meanwhile, can read out the first three lines of a poem using "my daiz" speech synthesis to enable the user to practice for competitive play and to provide a learning effect.

4. Katarai Slack Application

4.1 Background

Chat tools like Slack®*7 are attracting attention due to the recent nationwide trend toward tele-working. However, issues particular to remote working are coming to light such as "a decline in communication" since there are less opportunities for interacting and conversing in a face-to-face manner in real space. Under these conditions, we can expect dialogue that is also entertaining to stimulate person-to-person and person-to-thing communication. With this in mind, we developed a dialogue application that runs on Slack—a leading chat tool—to make it easy for users to enjoy entertaining dialogue. As an application that can be installed in Slack, it is the first to have a function for enlivening dialogue within a channel*8.

4.2 Functions

The following functions become available simply by installing this Slack application.

(1) Replies to posts

This function posts a natural reply to the content of a message posted on that channel. The text making up the reply uses output from the katarai chat-oriented dialogue engine. To prevent the channel log (record of messages) from flowing in a disordered manner, this function issues replies with a fixed probability instead of replying to every user utterance.

(2) Word play that picks up keywords in posts

If words stored in the database of *nazokake* riddles described above should happen to be included in messages posted on the channel, this function will reply with word play with a fixed probability. For example, given the post "I would love to have some ramen," the application might reply with "What do the words 'ramen' and 'Gion Festival' (one of the most famous festivals in Japan) have in common? The answer is *dashi* (which, in Japanese, can mean either 'soup stock' or 'parade float')."

(3) Automatic posting of *nazokake* riddles and quizzes

This function automatically posts content such as *nazokake* riddles to liven up a channel. This can prevent chat-oriented channels from becoming buried or inactive and induce users to make posts on such channels.

We administered a questionnaire consisting of eight items to nine users who participated in a trial of this application (**Figure 4**). Then, when using "4. Undecided" on a seven-point Likert scale*9 as baseline, we conducted a one sample t-test*10 against the evaluation mean for each question and found that a significant difference on a 5% level existed for items like "I wanted to send messages" and "It lowered barriers to sending messages."

---

*7 Slack®: A registered trademark of Slack Technologies, Inc.
*8 Channel: In Slack, a place for organizing and sharing messages in units of projects, teams, etc.
*9 Likert scale: A type of answer scale used in questionnaires and other types of psychological examinations. It enables the responder to indicate the extent to which he/she agrees with

the statement presented. A Likert scale is generally a five-point scale but seven-point and nine-point scales are also used.
*10 One sample t-test: A statistical test that determines whether the mean of a population is equal to a specific value. In this article, we conducted this test under the hypothesis (null hypothesis) that the mean of questionnaire results is 4.0 (= specific value).
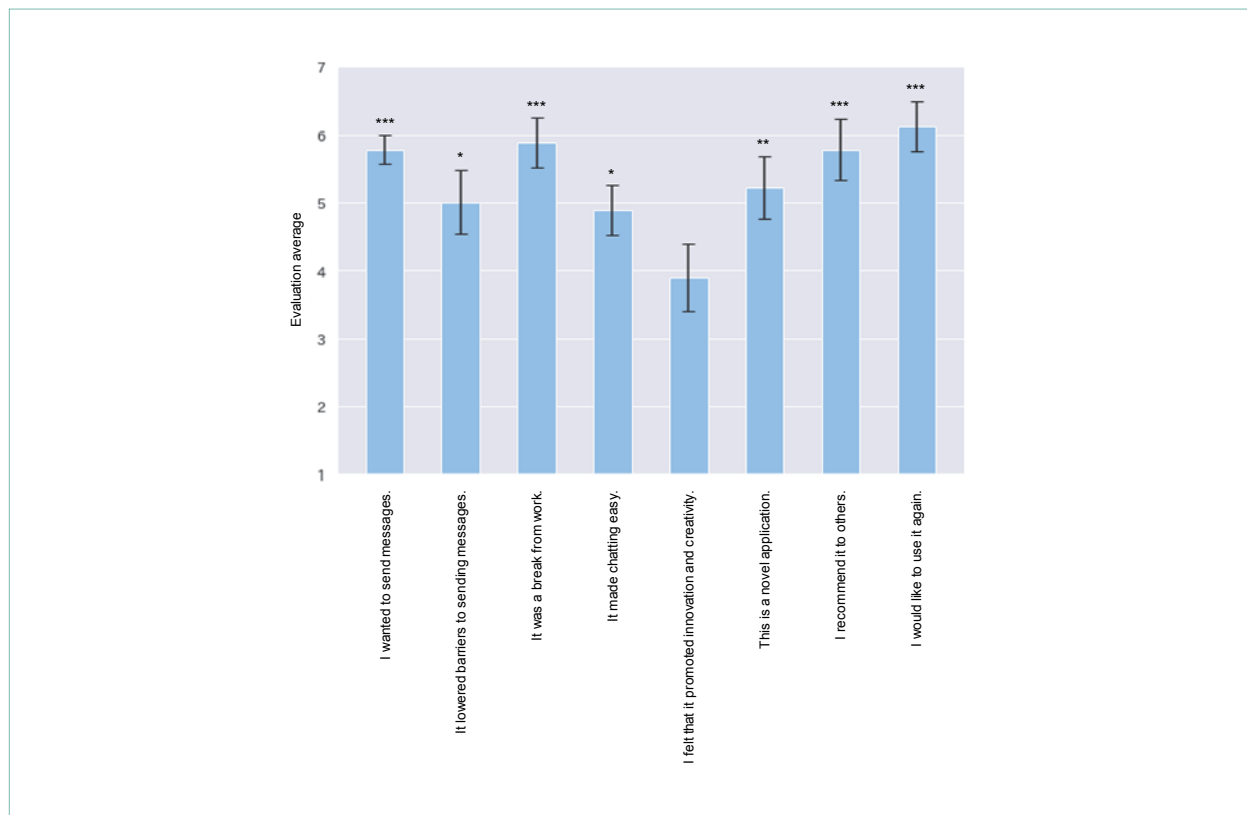
Figure 4   Result of questionnaire on katarai application

These results suggest that using this application should lower the psychological barriers to posting messages on a channel and eliminate a sense of isolation due to a drop in face-to-face interaction while also promoting communication within the organization. In the figure, the error bars indicate standard deviation (variation in collected data) and the number of asterisks (*) at the top of each bar graph indicate level of significance (*: $p < 0.05$; **: $p < 0.01$; *** $p < 0.005$).

5. Conclusion

In this article, we described character chat technology, word play technology, and an entertaining dialogue application. Character chat technology makes it possible to reproduce dialogue with any type of character by constructing a dialogue system based on utterance-pair data in the style of the desired character. Word play technology provides four types of games including "Yamanote Line" and "*nazokake* riddles" constructed on the basis of a knowledge base such as Wikipedia. These technologies make it possible to give chatbots an entertaining nature that makes them friendly and approachable to users. We also developed a Slack application that provides entertaining dialogue to make it easy for users to enjoy entertaining and friendly interaction.

Going forward, we plan to extract problems

from actual services using these technologies and to make improvements toward services that can provide dialogue that is even friendlier and more entertaining to the user.

REFERENCES

[1] docomo Developer support website.
https://dev.smt.docomo.ne.jp/

[2] katarai website.
https://www.katar.ai/

[3] R. Higashinaka, M. Mizukami, H. Kawabata, E. Yamaguchi, N. Adachi and J. Tomita: "Role play-based question answering by real users for building chatbots with consistent personalities," Proc. of SIGDIAL, Jul. 2018.