NTT DOCOMO

Technical Journal

Vol.21 No.2 | Oct. 2019

NTT docomo

Intellectual Property Strategy for the Reiwa Era

General Manager of
Intellectual Property Department
Tadanobu Ando

In Japan, the era name has changed from Heisei to Reiwa marking the dawn of a new era for the nation. Looking back at the history of mobile phones in Japan, handheld compact phones first appeared at the beginning of the Heisei era (early 1990s), so it is easy to see that the evolution of mobile phones during this period has truly been amazing. Of interest here is that it was NTT DOCOMO itself that drove this evolution forward. In this sense, I look forward to seeing what NTT DOCOMO will have created by the time that the Reiwa era comes to an end.

In addition to welcoming this new era in Japan, 2019 stands to be a year of major events such as the Rugby World Cup and the entry of Rakuten, Inc. into the Japanese mobile communications market. At the same time, pre-commercial service of 5G, or the "fifth generation mobile communications system," is scheduled to begin in 2019. The 5G system has the potential of transforming not only mobile carriers but everyone's lifestyle as well. In this article, I would like to take a look at this 5G megatrend from the viewpoint of intellectual property.

The various technologies and services of 5G have been discussed and international standards have been developed by the 3rd Generation Partnership Project (3GPP), an international standards organization. NTT DOCOMO has been actively participating in 3GPP since its founding in 1998 and has made many contributions toward advancing mobile communications, enhancing customer convenience, etc. Through these activities, NTT DOCOMO has acquired many patents in 3G and LTE that it licenses under fair, reasonable, and non-discriminatory terms. As of March 2019, NTT DOCOMO held about 14,000 patents, 40% of which fall under the category of essential patent*1 related to communications standards.

Discussions on 5G began at 3GPP in 2015, and since then, NTT DOCOMO has become even more involved in 3GPP standardization activities. According to a report [1] by an outside investigative agency, NTT DOCOMO ranks first in the world in the number of 5G patent applications among carriers (and sixth among all companies). NTT DOCOMO has figured prominently within this friendly competition with other companies.

In terms of ultra-high-speed communications, ultra-low-latency, and simultaneous connection of many terminals, 5G will achieve levels way beyond what can be presently imagined. However, what we feel to be the true significance of 5G will be the solutions and services that use 5G.

NTT DOCOMO has launched the "5G Open Partner Program" to promote co-creation with its corporate customers. A variety of solutions are now being studied with an eye toward pre-commercial service launch.

The birth of new technologies supporting 5G leads to new patents and intellectual property, and NTT DOCOMO's contribution to new technologies and related intellectual property is one proof of this process.

Making contributions to communications-related standardization activities is connected to the goal of embodying NTT DOCOMO ideas in the form of standardization and disseminating technology, but it also implies the generation of licensing revenues as a secondary effect. Taking technologies up to LTE, for example, NTT DOCOMO has obtained an appropriate amount of revenue by concluding licensing contracts with smartphone manufacturers and others. We can therefore expect a new source of licensing revenue with the future spread of 5G.

Furthermore, in addition to the introduction of 5G, NTT DOCOMO has made a turn toward business operations centered about the "membership base" of its "d POINT CLUB," a point program that anyone can join regardless of whether they have a line subscription or not. This is the beginning of a major transition for NTT DOCOMO. There is no change here in NTT DOCOMO's stance of deepening its relationship with customers—it is just the approach that will be changing greatly. This includes major reform through digital transformation*2 and the existence of new technologies.

Together with "5G rollout" and "transformation into business management," it is also important to "innovate and take action" in the field of intellectual property. The NTT DOCOMO Intellectual Property Department is itself evolving. By producing new NTT DOCOMO intellectual property in cooperation with all concerned, we wish to contribute not only to NTT DOCOMO business but also to the sustainable development of industry and society.

REFERENCE

[1] Cyber Creative Institute: "Cyber Creative Institute analyzes "Application trend of ETSI standard essential patent (5G-SEP) candidates contributing to realization of 5G and proposal trend of contributions for standards,"" Feb. 2019. https://www.cybersoken.com/file/press190206eng.pdf

*1 Essential patent: A patent for which it is necessary to obtain a license from its owner to avoid infringement when manufacturing or selling a product complying with a standard.

*2 Digital transformation: The changes that the digital technology causes or influences in all aspects of human life.

NTT DOCOMO
Technical Journal
Vol.21 No.2 Oct. 2019

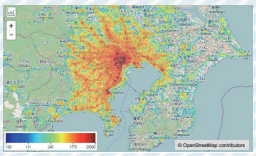
[Contents]

DOCOMO Today

Technology Reports & Topics (Special Articles)

Special Articles on AI Supporting a Prosperous and Diverse Society

Technology Reports



(P.27)

Topics

(P.36)

Technology Reports

(P.42)

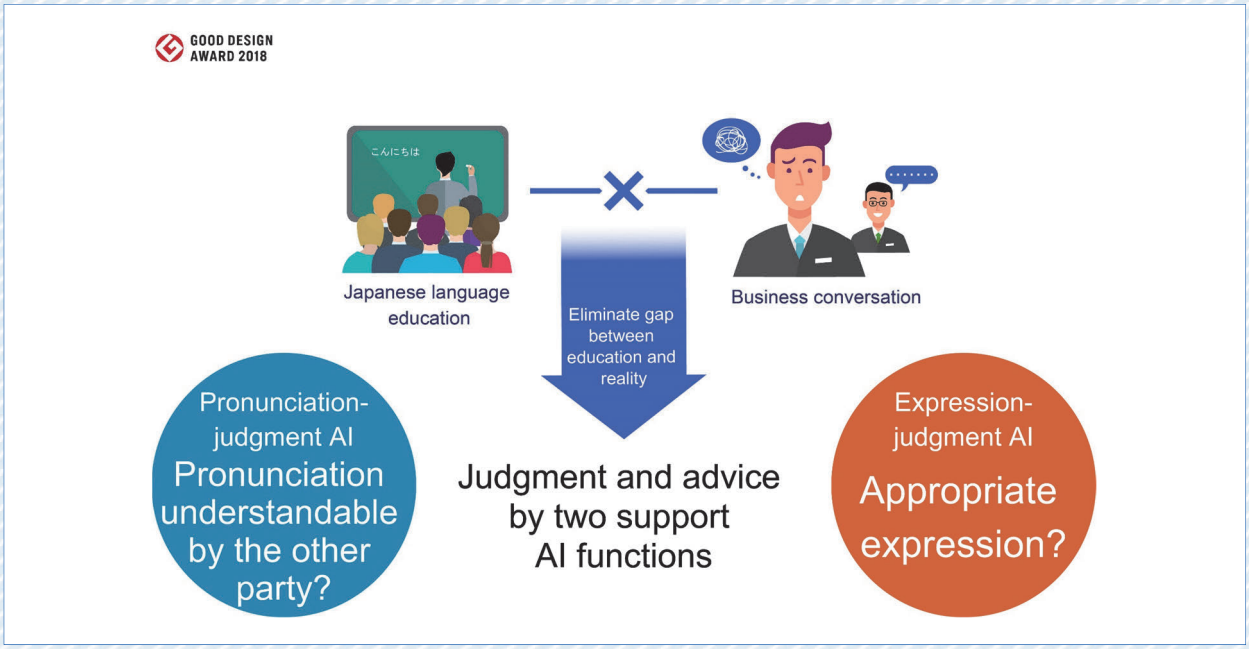
The “Office Link Voice Conferencing Service”
—A New Telephone Conferencing System Using the Office Link Platform— 42

Telephone Extension Solution | Telephone Conference Service | Office Link

News

(P.52)

NTT Group Receives the “Derwent Top 100 Global Innovators 2018-19” Award
—NTT DOCOMO Activities Contribute to Earning This Award— 52



Technology Reports (Special Articles) Special Articles on AI Supporting a Prosperous and Diverse Society (P.4)
JLT features

Technology Reports
Japanese Training AI
Special Articles on AI Supporting a Prosperous and Diverse Society

"Japanese Language Training AI" Supporting Japanese Conversation Training for Foreigners

Communication Device Development Department Shin Oguri Misa Tanaka†

The number of foreign workers in Japan has been increasing annually reaching 1.28 million in 2017. However, in conventional Japanese language education for foreigners, a gap has existed between the acquisition of "correct" reading/writing, pronunciation, and use of expressions on the one hand and practical Japanese conversation in real business situations on the other hand. Foreign workers are bewildered by this gap, which may hinder their work duties and even isolate them in the workplace. "Japanese Language Training AI" is a Japanese conversation training support service that was developed to solve these problems.

1. Introduction

In Japan, the number of foreign workers has been increasing annually reaching 1.28 million in 2017 [1]. Nevertheless, a shortage of human resources is still a problem in a variety of industries, so a bill was passed to revise the Immigration Control Act [2]. Enacted on April 1, 2019, this revision expands the range for which foreign workers having certain specialties and skills can be accepted, so the need for Japanese language education is expected to grow for both foreign workers and the companies accepting them.

As a result of conducting interviews with foreign workers, we found that there were some who had studied Japanese in their home countries before coming to Japan, passed the Japanese-Language Proficiency Test, and acquired a certain level of

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

† Currently Solution Service Department

Japanese. However, due to the gap between Japanese studied at a Japanese language school or in textbooks and Japanese used in the workplace, they could not communicate well, which hindered their work and left them feeling isolated in the workplace. This situation led some to even return to their home countries.

To eliminate this gap between conventional Japanese language education and conversation in real business situations, NTT DOCOMO developed Japanese conversation training support service "Japanese Language Training AI" (hereinafter referred to as "JLT") as a departure from conventional language teaching materials centered about the memorization of example sentences. This service features a function that enables the user to freely create conversation that he or she would actually like to speak and practice with. It also judges whether that conversation is made up of appropriate "words and expressions" and offers advice as well (**Figure 1**).

This service was achieved by forming a cross-organizational joint team composed of NTT DOCOMO R&D departments and corporate sales and marketing departments and developed as a "TOPGUN" project that aims to solve social and business issues. To solve the problems in conventional Japanese language education that cannot necessarily be said to be practical, this service has undergone hypothesis testing through verification experiments and its app has been improved. This article describes the JLT service and its development as a TOPGUN project.

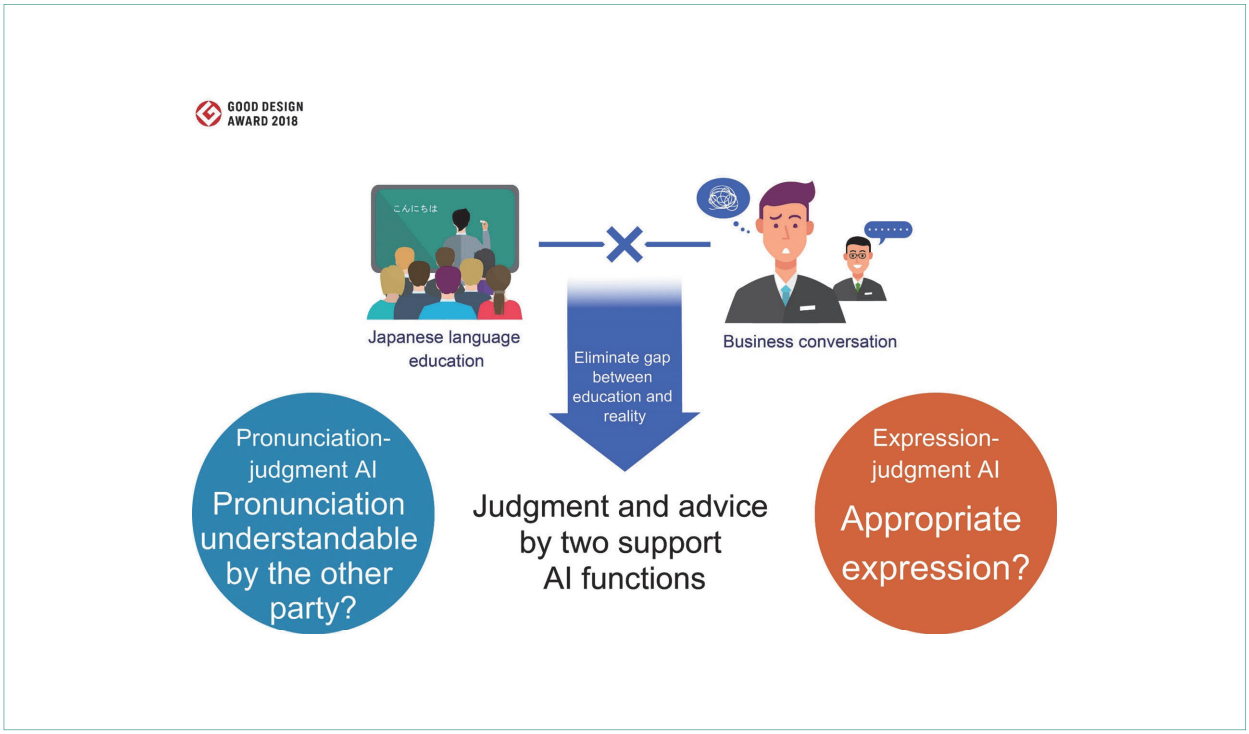


Figure 1 JLT features

2. JLT Overview

In contrast to conventional language teaching materials centered about the memorization of example sentences, JLT features NTT DOCOMO-developed AI functions (pronunciation judgment, expression judgment) that enable the user to freely create conversations that he or she would actually like to speak and to learn practical Japanese that can be understood by a Japanese native speaker.

The JLT service also provides training content useful in actual business situations for various fields and applications (dining, lodging, IT, retail sales, caregiving, and job-hunting activities) (**Figure 2**). Furthermore, to enable Japanese language training specific to the work of individual companies, NTT DOCOMO can provide a customer with customized training content.

The following describes the pronunciation-judgment function, expression-judgment function, and conversation-creation function.

2.1 Pronunciation-judgment AI

Pronunciation-judgment AI is a function that asks the user to read an example sentence in Japanese and judges not whether the result is correct Japanese pronunciation but whether it’s pronunciation that can be understood by a Japanese native speaker. It also offers advice on improving pronunciation. For example, if the user mistakenly says “nimotsu wo *omochi* itashimasu” (“please let me carry your baggage”) as “nimotsu wo *omachi* itashimasu,” a Japanese native speaker would still understand the meaning. While a conventional Japanese conversation training support service would treat this as a mistake, the JLT service would judge this to be “GOOD” since it’s a statement that could be understood while advising the user that “*omochi*” is correct (**Figure 3**).

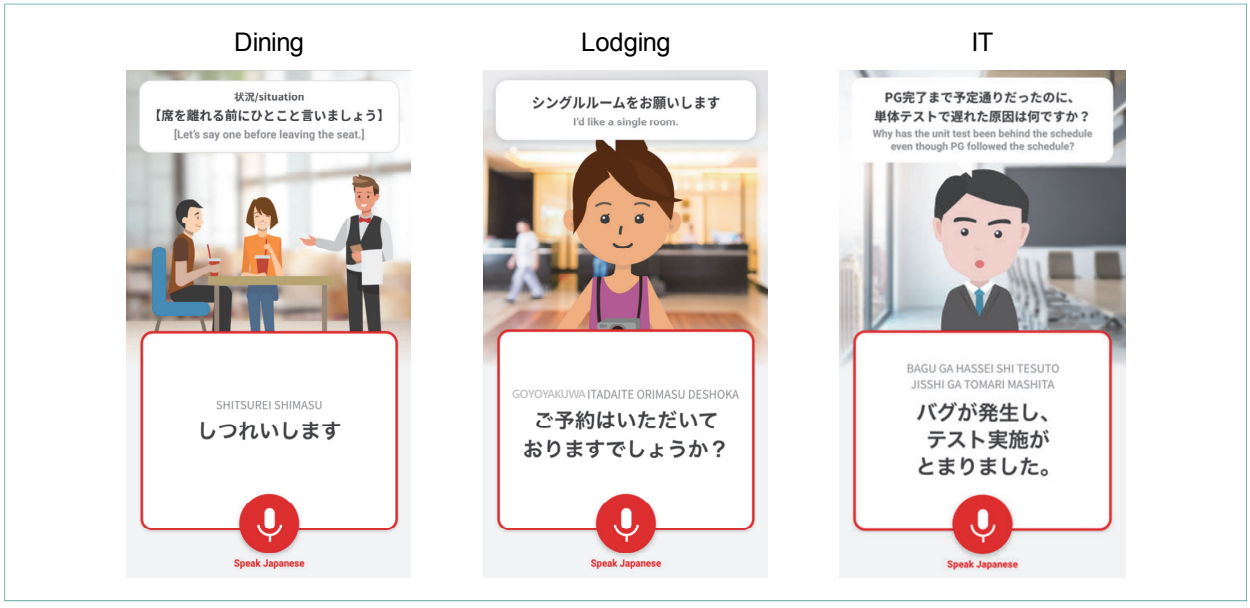


Figure 2 Training content by industry and application

2.2 Expression-judgment AI

Expression-judgment AI asks the trainee to speak in Japanese a phrase presented in the trainee’s native language (English or Vietnamese at present) and judges whether the expression or wording used would be understandable to a Japanese native speaker while offering advice if needed. For example, if the trainee renders the English sentence “If you take this bus, you can get to the station.” as “kono basu de eki ni ikemasu,” the function would judge it to be “GOOD” since the meaning is understandable but would advise the trainee that “kono basu ni noreba, eki ni tsukimasu” is a better choice of words (**Figure 4**). Another feature

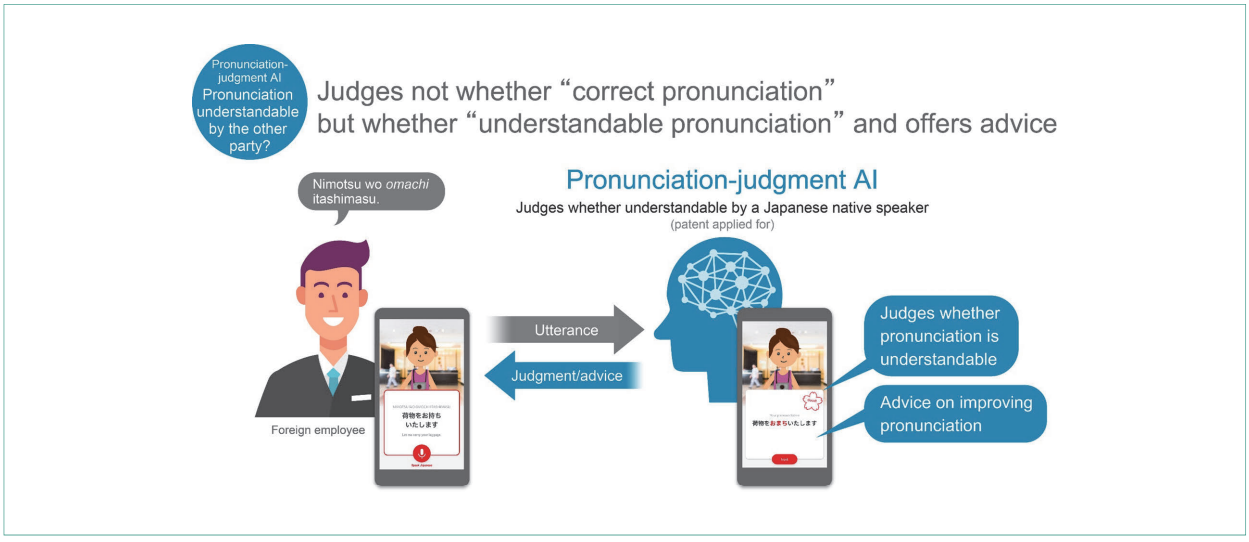


Figure 3 Pronunciation-judgment function

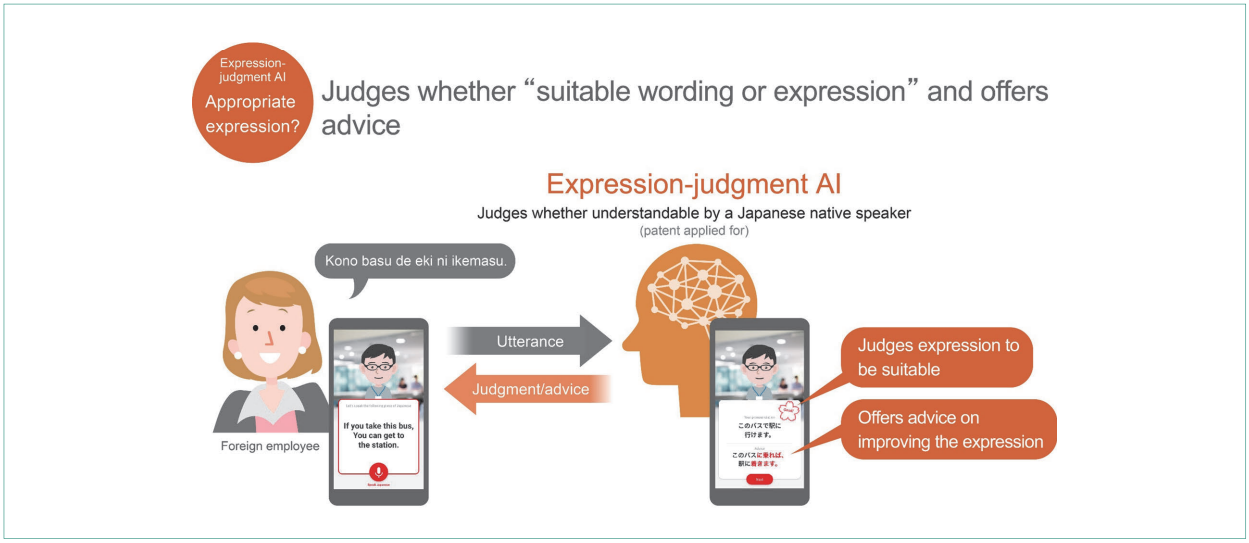


Figure 4 Expression-judgment function

of this function is that it gives a good evaluation to even a different expression such as "kono basu de eki ni ikemasu yo" as long as it expresses the correct meaning.

2.3 Conversation-creation Function

The conversation-creation function enables the user to create original training content by inputting a practice phrase in the user's native language (English or Vietnamese) into the user's smartphone by speech or text. The JLT service therefore supports not only training with preinstalled training content but also user needs in the manner of "I would like to try saying this too in such a scenario."

3. JLT Configuration and Technology

The configuration of JLT is shown in **Figure 5**. In JLT, pronunciation-judgment processing is performed on the terminal app while expression-judgment processing is performed on a server. We here describe the AI technology behind the pronunciation-judgment function.

The key feature of pronunciation-judgment AI is that it runs on the terminal app. Although there are systems that construct an acoustic model from the user's (foreigner's) spoken Japanese and perform associated processing on the server side, the cost of creating such an acoustic model and constructing a server for that purpose has been an issue. The pronunciation-judgment AI library*1 that we have developed uses the text of a practice sentence and the text of the speech recognition result to judge pronunciation. Specifically, the function partitions the speech recognition result into individual words, removes the vowels in those

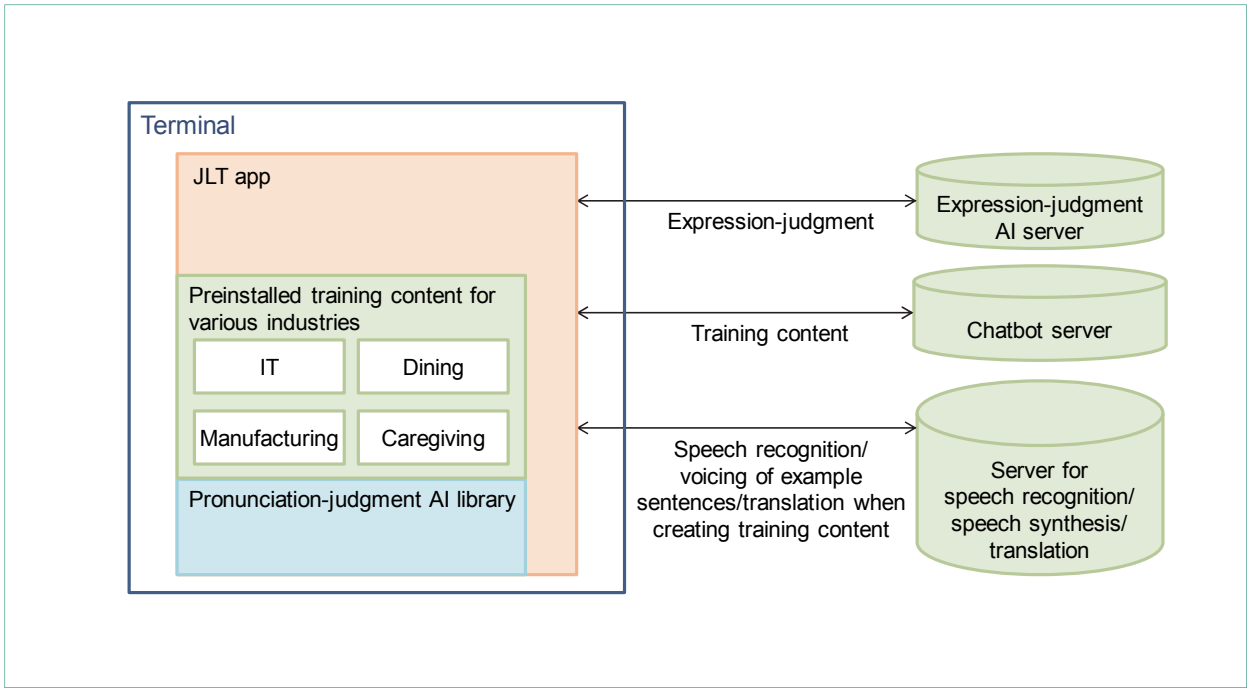


Figure 5 Function configuration

*1 Library: A collection of general-purpose software programs in a reusable form.

words and substitutes consonant groups taking pronunciation similarity into account, and calculates the degree of similarity with respect to the combinations of all the words included in the text of the correct sentence and all the words included in the text of the speech recognition result (**Figure 6**).

The JLT service also implements a function for removing fillers before performing judgment. This makes it possible to appropriately judge the pronunciation of an utterance mingled with fillers sounds that would intrinsically be understood by a Japanese native speaker. Japanese spoken by a foreigner may include fillers (such as "ah"), repetitions, etc. that act as noise. Consequently, the results of judgment may be low even if the utterance is understandable to a Japanese native speaker. For example, given "kono basu de eki ni ikemasu" as the correct sentence, the user may utter "kono basu de eki ni *ah* ikemasu" so that the speech recognition result would be exactly that. That is to say, if simply comparing the correct sentence with the speech recognition result of the user's utterance, the latter would turn out to be inappropriate with respect to the former due to the frequent use of "ah" in speaking resulting in a judgment of "error."

However, JLT performs judgment after removing fillers and repetitions so that the user's utterance in this case would be judged to be appropriate with respect to the correct sentence.

Pronunciation-judgment AI is not limited to judging the Japanese spoken by foreigners—it can also be applied to judging the English spoken by Japanese.

In this regard, a Japanese person skillful in English (corresponding to a TOEIC®*2 score of 800) and a Japanese person weak in English (corresponding to a TOEIC score of 400) were each asked to utter

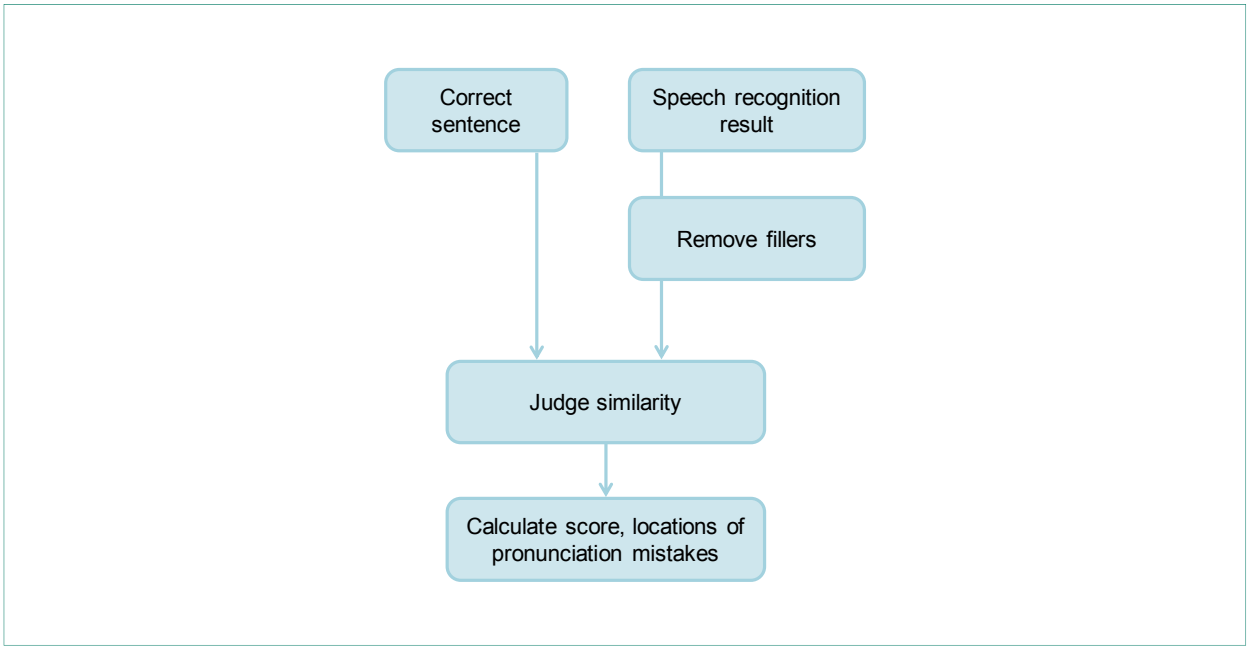
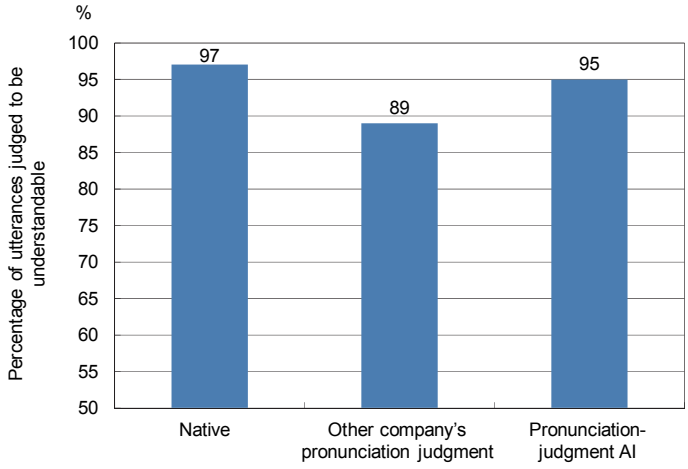


Figure 6 Pronunciation-judgment processing

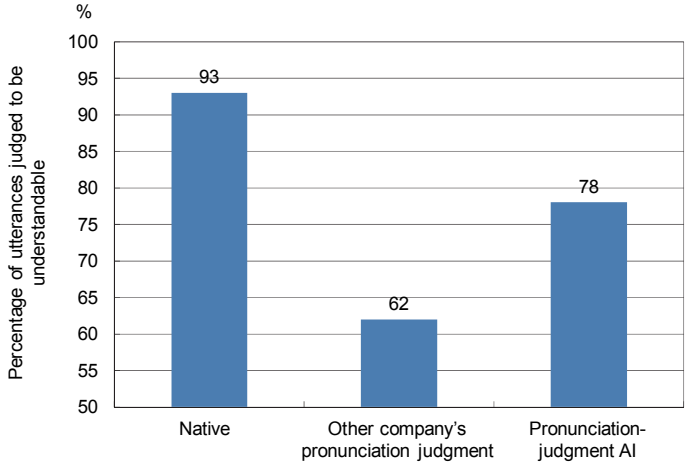
*2 TOEIC®: A registered trademark of Educational Testing Service (ETS). This product is not endorsed or approved by ETS.

150 example sentences in English. **Figure 7** shows the results of judging the pronunciation of those speakers by an English native speaker, another company’s pronunciation-judgment system, and our pronunciation-judgment AI.

In judging whether pronunciation was understandable, our pronunciation-judgment AI demonstrated a performance approximately 5% higher with respect to utterances by the Japanese person skillful in English and approximately 16% higher



(a) Utterances of Japanese skillful in English



(b) Utterances of Japanese weak in English

Figure 7 Performance evaluation of pronunciation-judgment AI

with respect to utterances by the Japanese person weak in English compared with the other company's pronunciation-judgment system. These results show that this technology is effective in judging whether pronunciation is understandable.

4. Verification Experiment with FPT Japan Holdings

To test the training support effect of the current version of JLT, we have been conducting verification experiments as a NTT DOCOMO TOPGUN project [3]. In this article, we introduce the verification experiment that we are conducting with FPT Japan Holdings Co., Ltd., which is the Japanese arm of FPT Software, the largest IT company in Vietnam.

While technical competence is, of course, essential, FPT Japan Holdings recognizes that communication in Japanese is also vitally important in getting customers in the Japanese market to entrust their work to another company with peace of mind. For this reason, the company is focusing its efforts on language acquisition by its Vietnamese employees by inviting a Japanese language lecturer every weekend and holding Japanese conversation classes free of charge for Vietnamese engineers living in Japan. There is also a plan to open an "FPT Japanese Language School" in Tokyo sometime in the future. Among these initiatives at FPT supporting Japanese language education, we have begun a verification experiment to explore the possibility of using JLT.

In the experiment, ten Vietnamese system engineers working at FPT Software Japan Co., Ltd., a subsidiary of FPT Japan Holdings, have been using JLT equipped with training content oriented to the IT industry.

Comments such as those below have been received on JLT.

- The many items of content having different degrees of difficulty and designed for various conditions enable training that can be tailored to individual employees. We expect communication between customers and employees to be vitalized as a result.
- I feel an improvement in my conversational ability since I have to think about expressions and words on my own.

5. Conclusion

This article presented an overview of "Japanese Language Training AI," explained pronunciation-judgment AI technology, and described a verification experiment conducted with FPT. After receiving a 2018 Good Design Award, JLT is expected to develop even further from here on [4] [5]. We are currently providing JLT to companies undertaking Japanese language education for foreign staff and technical interns, companies helping foreigners with living in Japan, organizations that support international students, etc. We are also promoting tests to assess the JLT training effect in foreign staff education, interview practice for international students, and other applications [6] [7]. Looking to the future, we plan to provide multilingual support so that foreigners studying Japanese overseas can make good use of the JLT service.

REFERENCES

[1] Cabinet Office: "On a Foreign Work Force," Feb. 2018

(In Japanese).
https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2018/0220/shiryo_04.pdf

[2] Ministry of Justice: “Bill to Revise Immigration Control and Refugee Recognition Act and Ministry of Justice Establishment Act,” (In Japanese).
http://www.moj.go.jp/nyuukokukanri/kouhou/nyuukokukanri05_00010.html

[3] NTT DOCOMO: “TOPGUN Corporate Site—Japanese Language Training AI,” (In Japanese).
https://www.nttdocomo.co.jp/biz/special/topgun/story06.html

[4] Japan Institute of Design Promotion: “List of Recipients of 2018 Good Design Award—Japanese Language Training AI.”
https://www.g-mark.org/award/describe/47048?token=st2v3GuNJW&locate=en

[5] “NTT DOCOMO Receives 2018 Good Design Award,” NTT DOCOMO Technical Journal, Vol.27, No.2, p.54, Jul. 2019 (In Japanese).

[6] NTT DOCOMO News Release: “(Notice) Trial Provision Begins of Japanese Conversation Training Service for Foreigners “Japanese Language Training AI”—AI judges and advises Japanese-specific “pronunciation” and “expressions”—,” Oct. 2018 (In Japanese).
https://www.nttdocomo.co.jp/info/news_release/2018/10/03_00.html

[7] NTT DOCOMO News Release: “Verification Experiment of Japanese Conversation Training Service for Foreigners “Japanese Language Training AI” Begins with Provision to Companies—Tests training effectiveness when used in education of foreign staff—,” Apr. 2019 (In Japanese).
https://www.nttdocomo.co.jp/binary/pdf/info/news_release/topics_190418_00.pdf#page=1

Technology Reports | Natural Language Processing | Machine Translation | Document Classification

Special Articles on AI Supporting a Prosperous and Diverse Society

Automatic Domain Prediction in Machine Translation

Service Innovation Department Soichiro Murakami Atsuki Sawayama Toshimitsu Nakamura Hosei Matsuoka Wataru Uchida

Machine translation can be applied to a variety of domains such as restaurants, lodging facilities, and transport agencies each of which differs in terms of conversation, vocabulary, phrasing, and their translation. It is therefore common to create a machine translation engine specialized for each domain using a corpus specific to that domain to improve translation performance. However, when faced with a translation task targeting multiple domains, the user must select multiple engines, which detracts from the convenience of machine translation. In response to this problem, NTT DOCOMO has developed technology for automatically predicting the domain of the machine translation engine from the text input by the user. This makes it possible to automatically select the optimal machine translation engine for the input text.

1. Introduction

In recent years, the number of foreign travelers visiting Japan has been increasing dramatically resulting in a sudden increase in "inbound demand." Against this background, voice translation services using speech recognition technology and machine translation technology are coming to be introduced for achieving smooth communication with foreign travelers in restaurants and other eating/drinking establishments, at lodging facilities, on public transportation, etc. Voice translation services

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

are also being introduced at medical institutions that will likely be used by ill or injured travelers to make the purpose of an examination or treatment understandable to the patient. In this way, voice translation services are coming to be used across a wide range of scenarios.

Amid this trend, Neural Machine Translation (NMT)*1 is attracting attention in the field of machine translation [1] [2]. "NMT" refers to the use of a bilingual corpus*2 to train a large-scale Neural Network (NN)*3, a scheme that has come to achieve more fluent and accurate translations than conventional statistical machine translation [3].

In NMT, using a large and high-quality bilingual corpus specific to a certain domain*4 can improve translation performance in that domain. It is therefore common to prepare a machine translation engine*5 specialized for each domain in voice translation services based on NMT. However, the sudden increase in inbound demand is being accompanied by an increase in domains that will likely require voice translation services. Furthermore, while a machine translation engine specific to each domain is needed, having to select which machine translation engine to use for each domain is troublesome for the user.

In response to these problems, NTT DOCOMO developed automatic domain prediction technology for automatically identifying the domain of input text. This technology predicts the domain of text input by the user by voice or keyboard so that a machine translation engine specific to that domain can be automatically selected for translation.

This article describes this domain prediction technology for automatically predicting usage scenarios in voice translation services.

2. Issues in Voice Translation Services

Voice translation services specific to overseas trips and customer service for foreign travelers include VoiceTra®*6, a voice translation app from the National Institute of Information and Communications Technology (NICT), ili®*7, an offline translation device for customer service from Logbar Inc., and POCKETALK®*8, a translation device from Sourcenext Corporation. NTT DOCOMO for its part provides "Jspeak" translation app for smartphones to facilitate face-to-face communication when making an overseas trip or when interacting with foreign travelers within Japan.

These examples show how voice translation services have been developed in diverse ways and how machine translation has come to be used in a wide range of domains. However, the content needing translation, the words and phrases used, and their translation depend on the domain such as restaurants, lodging facilities, public transportation, etc. For this reason, machine translation engines specialized for individual domains have been introduced to improve translation performance. Yet, for the user using a voice translation service, having to select a dedicated machine translation engine for each usage scenario takes time and effort. It is therefore considered that this troublesome task could be avoided if it were possible to predict the domain from the text input by the user and automatically select the optimal machine translation engine.

3. Automatic Domain Prediction

The automatic domain prediction technology that

*1 NMT: Machine translation technology using NNs (see *3), a machine-learning technique.
*2 Corpus: Language resource consisting of a large volume of text, utterances, etc. collected and stored in a database.
*3 NN: An entity that numerically models nerve cells within the human brain (neurons) and the connections between them. It is composed of an input layer, an output layer and hidden layers and is able to approximate complex functions by varying the number of neurons and layers and the strength of connections between layers.
*4 Domain: A usage scenario in machine translation.

we have developed extracts features from the text input by the user and performs document classification by machine learning*9 to predict the domain appropriate to the input text.

An overview of the system is shown in **Figure 1**. This figure shows the flow of classifying the text input by the user into one of several predetermined domains using a document classifier*10 and sending a translation request to the translation engine specialized for that domain. Here, "document classifier" refers to a device that classifies text into one of several predetermined classifications.

3.1 Automatic Domain Prediction as Document Classification

This technology performs document classification by predicting the domain of the input text. "Document classification" means the classification of text input to the voice translation service into one of several predefined labels. Here, "label" refers to a domain such as restaurants, lodging, or transportation. In the field of Natural Language Processing (NLP), it is common to construct a document classifier by training a machine-learning model using training data consisting of pairs of documents and labels.

An example of classification using a document classifier is shown in **Figure 2**. In this example, the document classifier extracts features from the text "Return visits to the clinic are received at counter 5" input into the voice translation service and predicts "medical care" from among the predefined domain labels using a machine-learning technique.

3.2 Training Data for Document Classifier

The training of a document classifier that uses a machine-learning technique requires the use of training data that pairs up input text of a voice translation service and domain labels.

In machine-learning techniques, model performance generally improves as the amount of training data increases. Furthermore, when constructing training data, care must be taken to prevent an imbalance in which data pairs in one domain

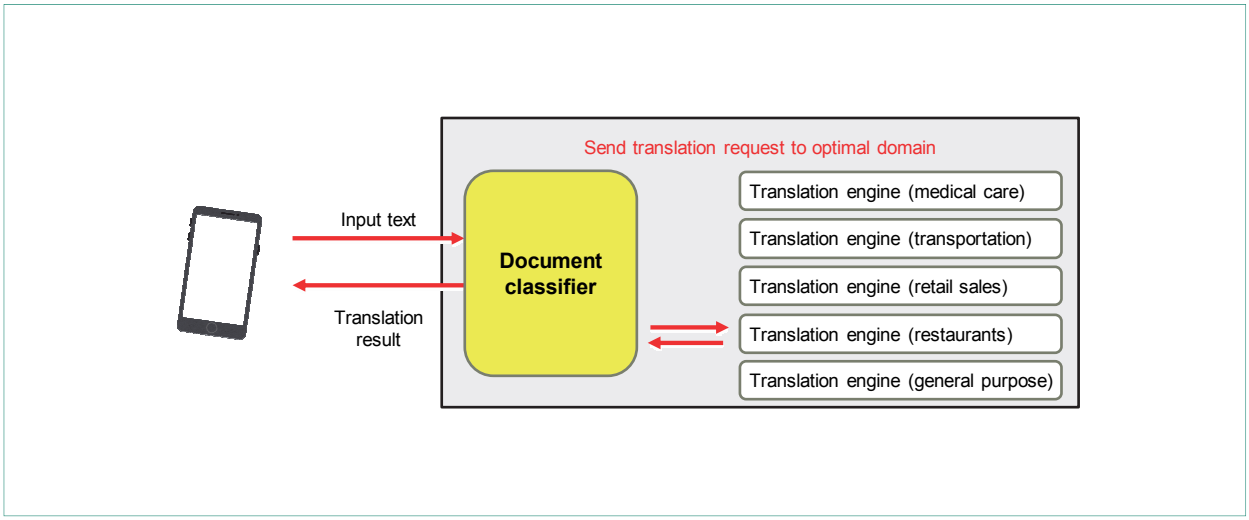


Figure 1 System overview

*5 Machine translation engine: Software for performing machine translation.
*6 VoiceTra®: A registered trademark of the National Institute of Information and Communications Technology (NICT).
*7 ili®: A registered trademark of Logbar Inc.
*8 POCKETALK®: A registered trademark of Sourcenext Corporation.
*9 Machine learning: Technology that enables computers to acquire knowledge, decision criteria, behavior, etc. from data, in ways similar to how humans acquire these things from perception and experience.

are many or few in number compared with that of another domain. This is to avoid the problem of over-fitting in which the accuracy of classification drops for text in a domain with a small amount of data.

Examples of document-classifier training data are listed in **Table 1**. Among these examples, the label "medical care" is attached to the text "When feeling dizzy, do you sweat or shiver with cold?" reflecting its domain.

3.3 Machine-learning Technique of Document Classifier

We here describe our system's document classifier that uses Long-Short Term Memory (LSTM) [4], which is a type of Recurrent NN (RNN) that introduces recurrent connections*11 in a NN. LSTM is widely used in the field of NMT that handles variable-length text. An overview of the feedforward NN*12 and RNN is shown in **Figure 3**.

In the hidden layer of an RNN such as LSTM,

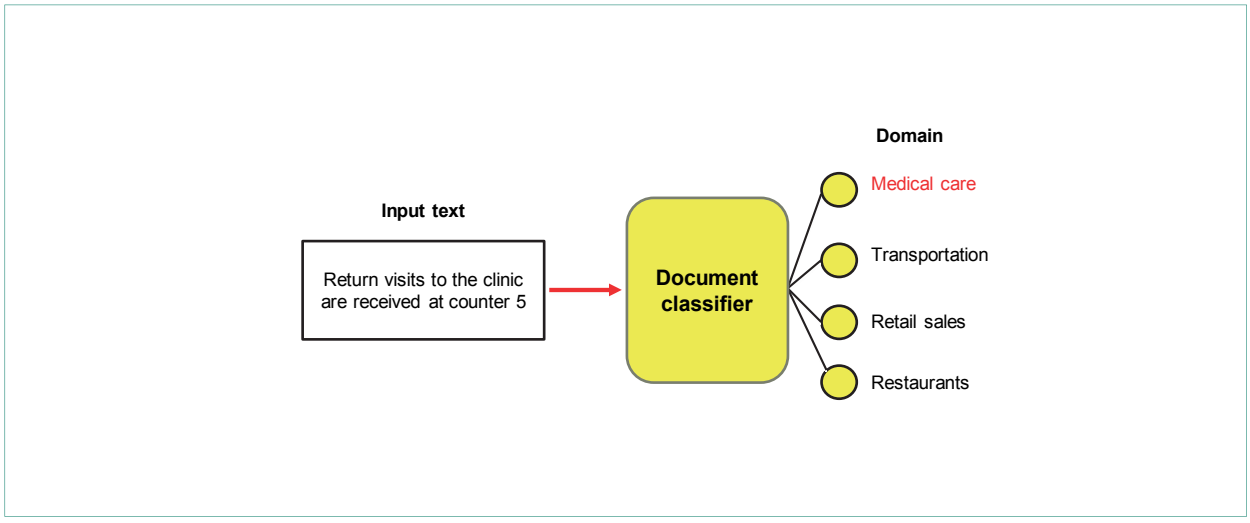


Figure 2 Overview of document classifier

Table 1 Examples of training data for a document classifier

Text	Domain
Can I see a doctor?	Medical care
When feeling dizzy, do you sweat or shiver with cold?	Medical care
This smart card cannot be charged here, so please do it beforehand.	Transportation
Arrival time may differ from the timetable depending on the weather or road conditions.	Transportation
Please bring your receipt to return or exchange any items.	Retail sales
Where is the toothpaste?	Retail sales
All items on the menu except for Japanese sake and shochu are all-you-can-drink.	Restaurants
All juices are 100% with no sugar added.	Restaurants

*10 Classifier: A device that classifies input into one of several predetermined classifications based on its feature values.

*11 Recurrent connections: Connections that are made in a recurrent manner.

*12 Feedforward NN: A NN that propagates signals only in a single direction in the order of input layer, hidden layers, and output layer without any recurrent connections in the network.

the inner state vector at the immediately previous time point $t-1$ can be taken over at the next time point t enabling flexible handling of variable-length input such as text. Furthermore, since text can be input in a time-series manner, the context information of that text can be expressed as a fixed-length vector called a context vector. In short, the use of an RNN enables the extraction of feature values[*13] that represent context from the input text.

An example of a decision made by a document classifier using LSTM is shown in **Figure 4**. In this example, the number of dimensions of the LSTM vector is 200. A document classifier using LSTM determines which domain the input text conforms to most based on the fixed-length context vector created from the input text using a NN. First, the

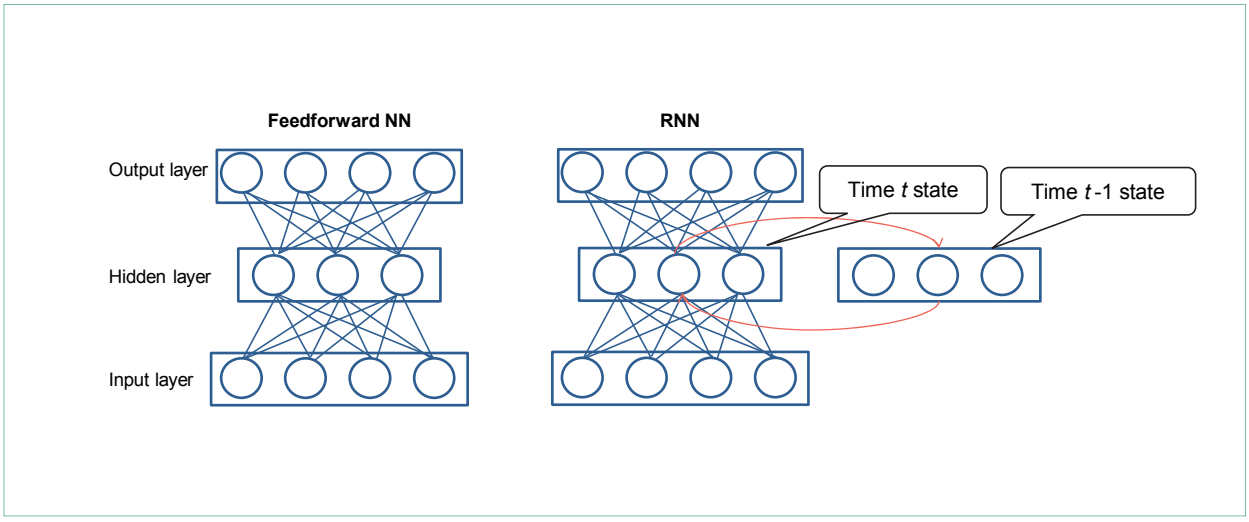


Figure 3 Feedforward NN and RNN

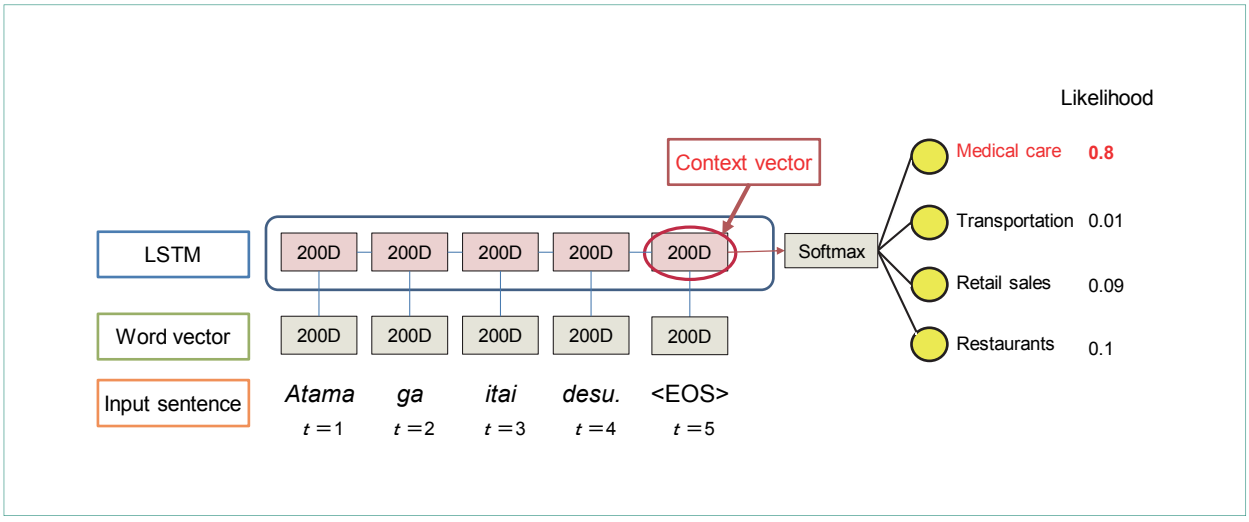


Figure 4 Document classifier using LSTM

*13 Feature values: Values extracted from data, and given to that data to give it features.

document classifier applies morphological analysis[*14] to the input Japanese sentence "頭が痛いです." ("*Atama ga itai desu.*" or "My head hurts.") to get the word-partitioned input word string "頭が 痛い です <EOS>" (*atama-ga itai desu* <EOS>). Here, "<EOS>" is a pseudo token[*15] that expresses the end of the sentence. Next, the classifier inputs the 200-dimension word vectors obtained by vectorizing each word of the input word string into the LSTM one-by-one and calculates the context vector expressing the context information of the input sentence. Finally, it uses the Softmax function[*16] based on this context vector to calculate the likelihood that the input sentence conforms to any one domain and uses those likelihood values to predict the domain most suitable for that input sentence.

3.4 Accuracy of Domain Prediction

We performed training of an LSTM-based document classifier using paired data consisting of text and labels and measured the accuracy of classification with respect to text data. In the experiment, we defined medical care, transportation, retail sales, and restaurants as the target domains and prepared 1,000 sentences of text data for each domain.

Domain prediction accuracy is summarized in Table 2. Examining the classification accuracy (*F* value[*17]) of each domain, it can be seen that the accuracy of this document classifier is generally high. In addition, the average processing time of domain prediction per sentence was approximately 12 ms, which indicates that domain prediction could be performed within a realistic processing time in actual use.

3.5 Application Example

Next, we describe an example of applying this system to machine translation using domain prediction technology (Fig. 1). The input text (in Japanese) was "このレストランではカリフォルニア産の高級ワインが召し上がれます." ("*Kono resutoran de wa kariforunia san no kokyu wain ga meshiagaremasu.*"). Using the document classifier, the system performed automatic domain prediction of this text and predicted the domain to be "restaurants." The system then translated the input text using the machine-translation engine for the restaurants domain resulting in the following translation:

"You can enjoy California high-quality wine at this restaurant."

However, on translating the input text using a

Table 2 Domain prediction accuracy

Domain	LSTM			No. of examples
	Precision	Recall	*F* value	
Medical care	0.99	0.92	0.95	1,000
Transportation	0.95	0.95	0.95	1,000
Retail sales	0.87	0.94	0.91	1,000
Restaurants	0.92	0.92	0.92	1,000

*Average processing time: 12 ms per sentence

*14 Morphological analysis: The task of dividing text written in natural language into morphemes—the smallest units of meaning in a language—and determining the part of speech of each.
*15 Token: A character or character string treated as the smallest unit of text.
*16 Softmax function: A function used to calculate probability values when normalizing the total output of a NN to 1.0.
*17 *F* value: A scale used for comprehensive evaluation of accuracy and exhaustiveness, and it is calculated as the harmonic mean of precision and recall.

general-purpose machine-translation engine, the following result was obtained:

"This restaurant has a high quality wine in California."

On comparing these translation results using a machine-translation engine dedicated to the restaurants domain and a general-purpose machine-translation engine, it can be seen that the dedicated translation engine can translate in a more fluent manner using phrases typical of that domain.

In this way, the use of automatic domain prediction technology enables higher quality translation by translating with a machine-translation engine optimal to the domain of the input text.

4. Conclusion

Given a voice translation service used in a variety of domains, this article described technology for automatically selecting the optimal machine translation engine using automatic domain prediction so that translation can be performed with an engine matching the user's domain.

Future plans include the development of domain prediction technology with even higher levels of accuracy and the development of domain prediction technology using information other than text.

REFERENCES

[1] I. Sutskever, O. Vinyals and Q. V. Le: "Sequence to Sequence Learning with Neural Networks," Advances in neural information processing systems, pp.3104–3112, 2014.

[2] D. Bahdanau, K. Cho and Y. Bengio: "Neural Machine Translation by Jointly Learning to Align and Translate," In Proc. of the 3rd International Conference on Learning Representations, 2014.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N Gomez, L. Kaiser and I. Polosukhin: "Attention is All You Need," In Proc. of Advances in Neural Information Processing Systems 30, pp.5998–6008, 2017.

[4] S. Hochreiter and J. Schmidhuber: "Long Short-Term Memory," Neural computation, Vol.9, No.8, pp.1735–1780, 1997.

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

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*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.

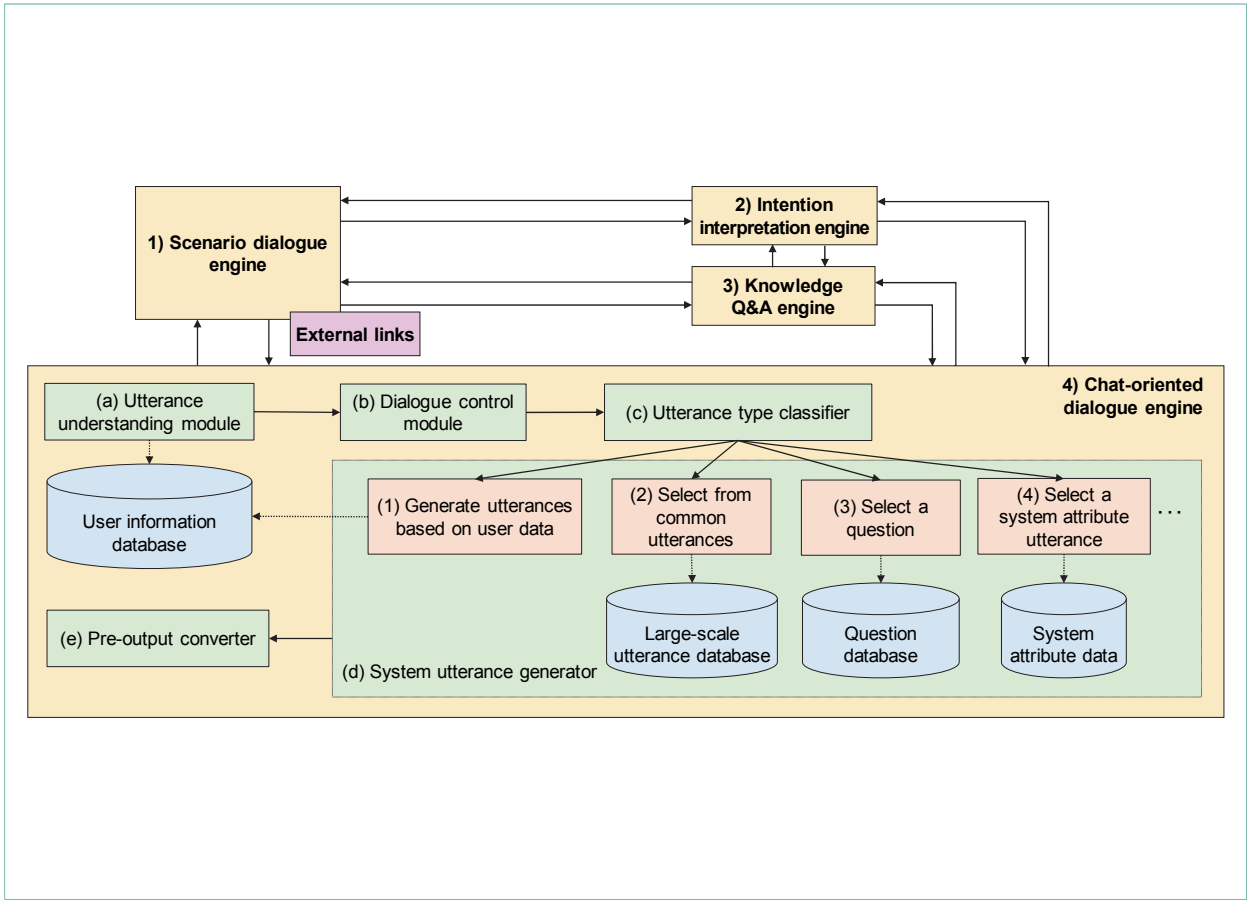


Figure 1 Natural dialogue platform architecture

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.

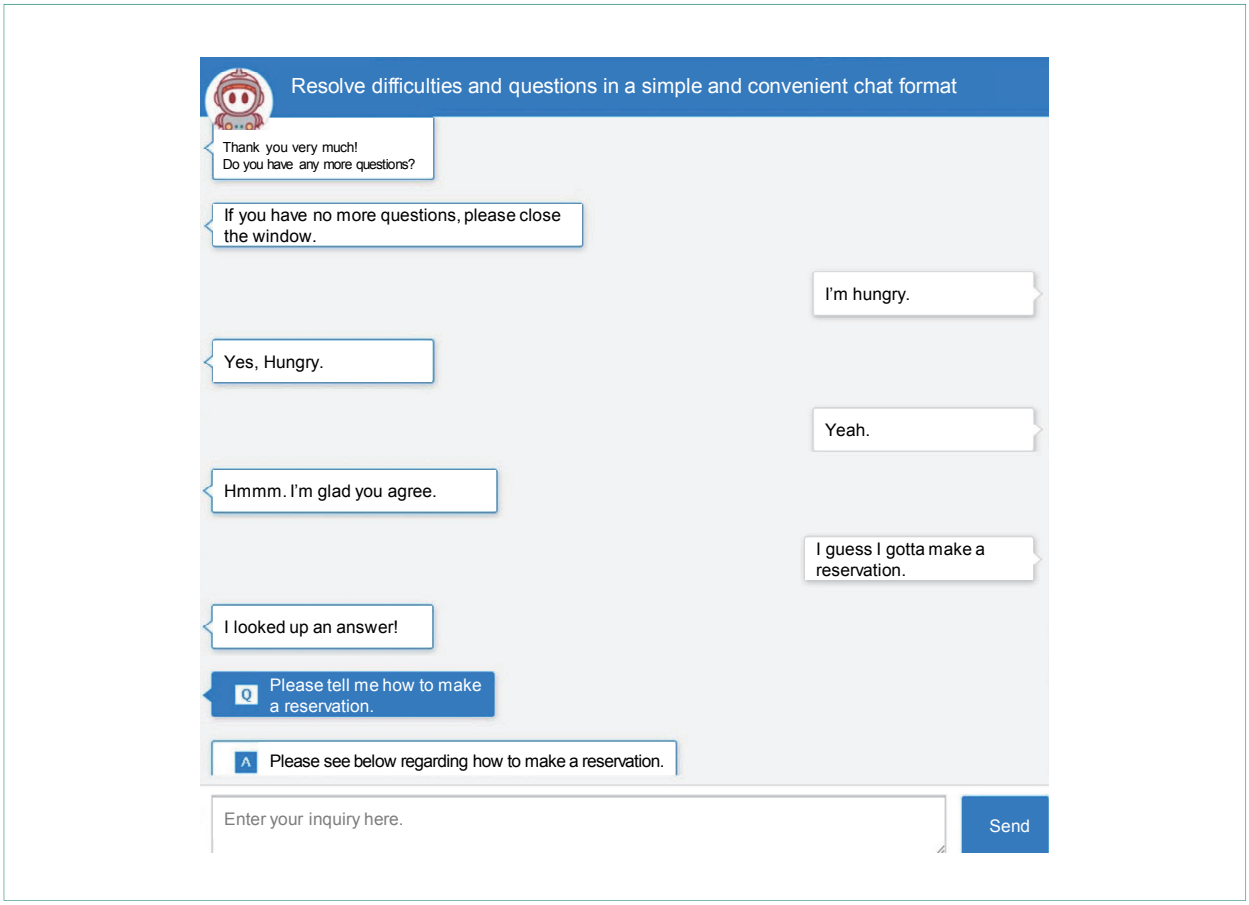


Figure 2 DOCOMO online shop, "Onshoko-roid"

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/

Technology Reports | Real-time Population Statistics | Congestion Prediction | Machine Learning

Special Articles on AI Supporting a Prosperous and Diverse Society

Avoiding Tokyo Bay Aqua Line Congestion Using Traffic Congestion Forecasting AI —Prediction Based on Statistical Processing of Mobile Phone Network Operations Data—

Research Laboratories Masayuki Terada Hiroto Akatsuka
Platform Business Department Tomohiro Nagata
Corporate Sales and Marketing Department I Satoshi Nakanishi

Traffic Congestion Forecasting AI is a technology that can predict the occurrence of traffic congestion, including size and times, by applying AI technology to real-time population statistics that are created in near-real-time from mobile telephone network operations data. Predictions are made based on the number of people out on a given day, making accurate predictions possible, accounting for effects such as weather or special events. This article gives an overview of Real-time Population Statistics and Traffic Congestion Forecasting AI, and introduces a trial performed in cooperation with NEXCO East on the Tokyo Bay Aqua Line. An overview of the trial, evaluation of prediction accuracy, and results of a survey of users participating in the trial are discussed.

1. Introduction

Frequent traffic congestion has been a major issue for many years in Japan. The resulting economic losses have been estimated at over 10 trillion yen per year [1], exerting a strong negative pressure on economic activity. Beyond effects on the economy, we are also all familiar with the related decrease in the quality of daily life. As an example, a common experience is encountering a traffic jam on the way home from an outing on the weekend and how this can diminish the enjoyable memories of the event.

While the roads that tend to get congested and

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

* Real-time Population Statistics: A service providing mobile-network spatial statistics in near real time. Displays aggregate population by area and attributes but does not include identifying information. These population statistics are created according to the mobile Spatial Statistics Guidelines.

the days and times are generally known, there can be large differences in day-to-day conditions: whether congestion will occur, how large it will be, when it will start, and when it will clear. For example, routes home from popular tourist destinations often get congested. One can even encounter an unexpected large traffic jam in an area that does not normally get congested if large numbers of people gather on that day for an event or other reason. On the other hand, if the weather is bad and people stay home, congestion could decrease or not occur at all.

As such, knowledge of approximately how many people are actually out on that day (turnout) is needed to predict what will actually happen in the during times when people are returning home, including the areas around their destinations. If we can know the turnout on a day quantitatively, we can expect to be able to predict congestion that will occur during a return-home period.

Traffic Congestion Forecasting AI is a new traffic prediction technology developed with a focus on this relationship between turnout and congestion. It is able to comprehend turnout on a given day using Real-time Population Statistics, a technology that estimates human populations throughout Japan using operations data from the NTT DOCOMO mobile telephone network, and based on this information, it is able to predict congestion and the scale and time-frame of the congestion.

This article gives an overview of Real-time Population Statistics and Traffic Congestion Forecasting AI and describes tests conducted in collaboration with East Nippon Expressway Co. Ltd. (NEXCO East) on the Tokyo Bay Aqua Line expressway starting in December 2017.

2. Real-time Population Statistics

Real-time Population Statistics is a new form of population statistics arising from R&D to make a real-time version of Mobile Spatial Statistics*1 [2], a commercial service that has been offered since 2013. It is able to estimate population distributions throughout Japan on a 500 m grid*2 (with some areas such as centers of designated cities on a 250 m grid) according to attributes such as age group (in 5-year increments) and place of residence (city, town, etc.). Data fluctuations can be provided at 10-minute intervals, approximately 20 minutes after the fact. In other words, population distributions at 12:00 are available by 12:20, those at 12:10 are available by 12:30, and so on.

An example visualization of Real-time Population Statistics is shown in **Figure 1**. This is an illustration of a population distribution at noon on a weekday. Each grid section is colored according to the population density in the cell, using blue, green, yellow, and red, in order of increasing population density.

Figure 2 shows the number of visitors to the Sumida River Fireworks Display on a given year at 8 pm, on a 500 m grid (defining the number of visitors as the increase in population compared with the usual population). Here, the number of visitors is represented by red, with darker shades indicating more visitors in that grid section.

The red areas are concentrated along the Sumida River, showing that many spectators were gathered there, but there are also two areas slightly east of the river with concentrations of people. Investigation on the following day revealed that, although they are somewhat far from the river, these

*1 Mobile Spatial Statistics: Population statistical data generated according to the "Mobile Spatial Statistics Guidelines," from NTT DOCOMO mobile network operations data. Population distributions on a grid (see *2) and by municipal boundaries are estimated such that individual users cannot be identified, using an estimation of the number of mobile phones currently in each base-station area and adjusting based on base-station area data, NTT DOCOMO phone usage rates and other information.

*2 Grid: Land divided into sections based on latitude and longitude.

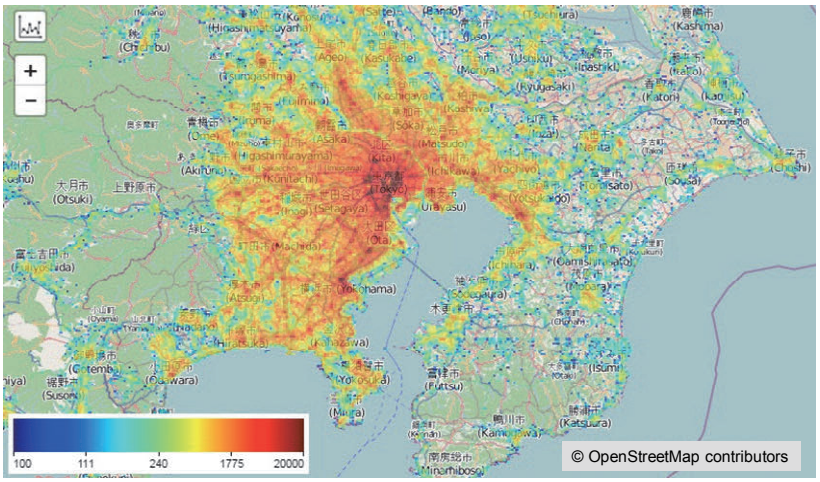


Figure 1 Real-time Population Statistics example

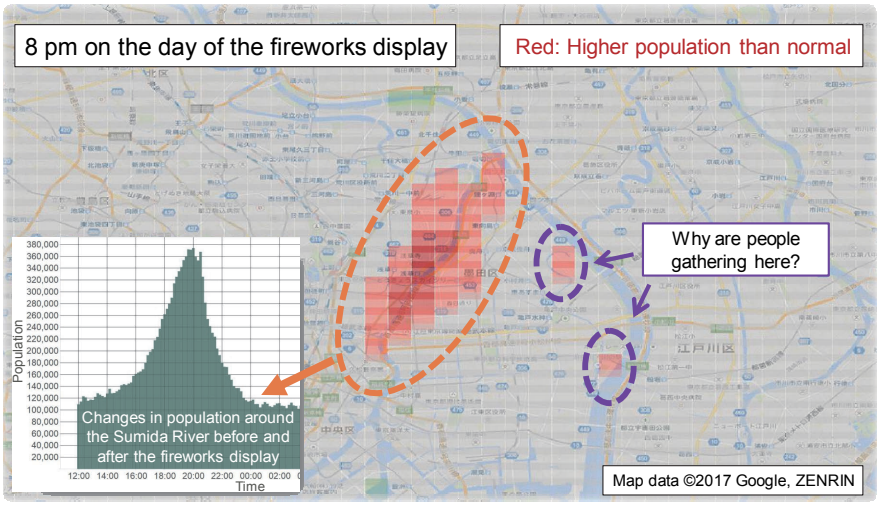


Figure 2 Results of estimating number of visitors to the Sumida River Fireworks Display in a given year

are little-known spots that are good for viewing the fireworks with few obstructions.

This illustrates the strength of Real-time Population Statistics, providing quantitative data on the fluctuations in population density throughout Japan according to attributes such as age and place of residence in near-real-time, including gatherings of people in areas, even without knowing why they may be gathering there.

3. Predicting the Future Based on Real-time Population Statistics

Real-time Population Statistics makes it possible to know the dynamics of population distributions or changes in how people are gathering. This suggests that dynamics of social phenomena and economic trends that are correlated to the movements of people can be estimated from Real-time Population Statistics.

A correlation is a relationship in which, when one value increases or decreases, another value also increases or decreases. For example, when the temperature increases in the summer, sales of ice cream also increase. Conversely, sales decrease when it is cool. We say there is a correlation between temperature and ice cream sales. By calculating this relationship based on past temperature fluctuations and ice cream sales records, one of the values can generally be estimated if the other value of is known.

In another example, water levels in rivers rise when it rains and drop after the rain stops. As with this example of precipitation and water levels, some correlations also involve a time difference. The value that changes first is called the leading indicator, and the value that changes later is called the lagging indicator. Since precipitation changes before the water levels, precipitation is a leading indicator for water levels. If the correlation can be computed based on past fluctuations in precipitation (considering time differences) and water levels in rivers, it will be possible to predict future fluctuations in water levels from precipitation leading up to the present. In other words, the leading index has the potential to predict future values of the lagging indicator.

Accordingly, changes in phenomena that are correlated to population fluctuations can be estimated based on Real-time Population Statistics, even if they cannot be observed directly. Human behavior involves a wide range of social and economic activities, and various social phenomena and economic trends are correlated to the movements of people. In particular, these correlations can also have a time difference, as with the correlation between precipitation and river water levels, so when population is the leading indicator, future changes in the social phenomena may be predictable.

4. Traffic Congestion Forecasting AI

Traffic conditions are an example where the future can be predicted. In implementing Traffic Congestion Forecasting AI, NTT DOCOMO has focused on increases and decreases in population in a given area as a leading indicator for increases and decreases in traffic demand on routes taken to return home from that area. The system is able to accurately predict traffic conditions several hours later, which has conventionally been difficult. This is done by making predictions based on observed population distributions, which are the basis of traffic demand. Thus, it can predict changes in traffic conditions from the afternoon until late at night based on populations observed at mid-day and earlier.

This gives users an opportunity to check traffic forecasts after lunch and revise plans for going home in the afternoon based on the information. Thus, it can help reduce the misfortune of encountering a traffic jam on the way home and having enjoyment ruined, as touched on earlier in this article. If an

increasing number of people act to avoid the congestion based on predictions, traffic demand will diffuse over time, relaxing or even eliminating the congestion itself. If people also avoid the congested times by deciding to have dinner before going home, for example, they will also spend more time or money in the area. In these ways, changes in user behavior based on prediction information can be expected to mitigate congestion by distributing traffic and stimulate economic activity in surrounding areas.

4.1 Technical Overview

The traffic predictions from Traffic Congestion Forecasting AI are implemented using population distributions obtained from Real-time Population Statistics and by applying a type of AI technology called machine learning*3. Specifically, a congestion prediction model that formulates the relationship between population and traffic conditions is created by training it using data from a set period in the past, consisting of population distributions together with the traffic history for the same day. When making predictions each day, population distributions from noon that day are presented to the congestion prediction model to obtain results predicting traffic conditions during a return-home period. This is presented schematically in **Figure 3**.

Here, we want to stress that by population, we do not mean simply the number of people in the given area, but rather, the population distribution on the grid, by attribute, as obtained from Real-time Population Statistics.

As an example, we consider congestion predictions on Tokyo Bay Aqua Line, an expressway crossing the Tokyo Bay between Kawasaki City in Kanagawa Prefecture and Kisarazu City in Chiba Prefecture. It is also the subject of a trial described below.

If the turnout on the Boso Peninsula is large

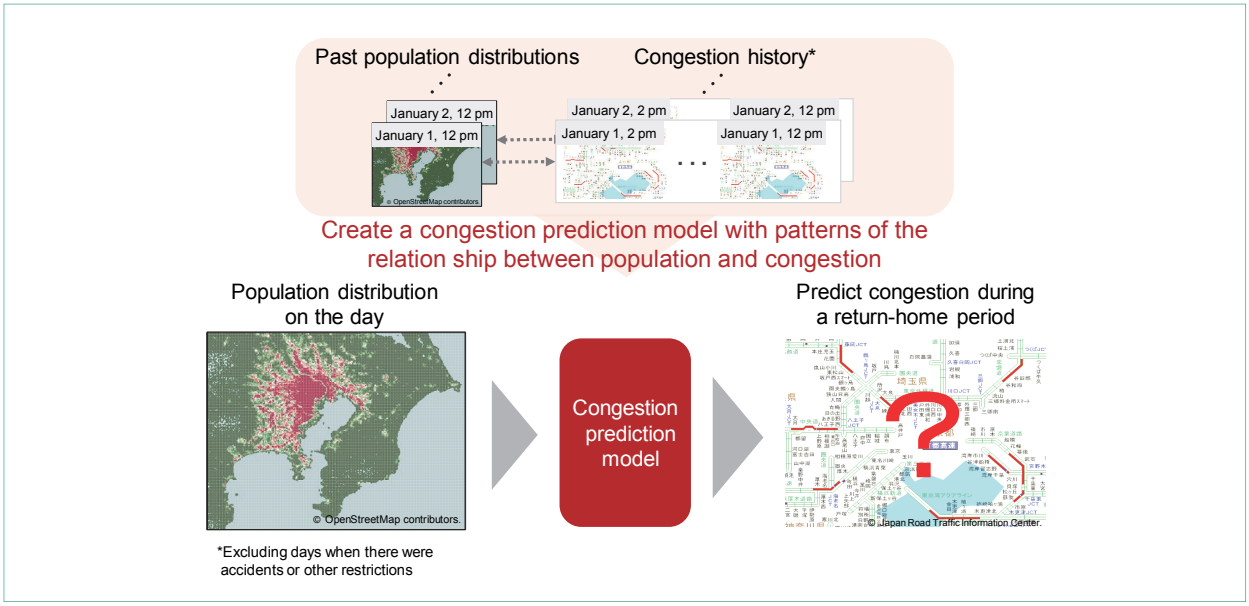


Figure 3 Traffic Congestion Forecasting AI organization

*3 Machine learning: A framework that enables a computer to learn the relationships between inputs and outputs by statistical processing of examples.

and mainly of people living in Chiba, there is almost no effect on congestion on the Aqua Line. On the other hand, if there are many people from Tokyo and Kanagawa, congestion on the Aqua Line during a return-home period can be severe. Even within Tokyo, whether people are coming from the east or the west of Tokyo can greatly affect traffic congestion.

Effects can also differ greatly depending on where visitors stay on the Boso Peninsula. Most people in the north of the peninsula will use the Keiyo Expressway or other routes, and people in the south end may use the Tokyo Bay Ferry. These proportions also differ depending on where they are going. For example, a larger proportion of people going to a golf course will be travelling by car, compared with other destinations, and they will tend to arrive earlier in the morning and go home earlier. As such, we can expect they will contribute to increasing traffic demand earlier in the day.

In this way, the effects of population distributions on congestion are not determined simply by the total number of people but can differ greatly according to location and other attributes. Traffic Congestion Forecasting AI is able to make predictions incorporating such differences by using AI to formulate such effects of population distributions by attribute on congestion, obtained from Real-time Population Statistics.

4.2 Tokyo Bay Aqua Line Trial

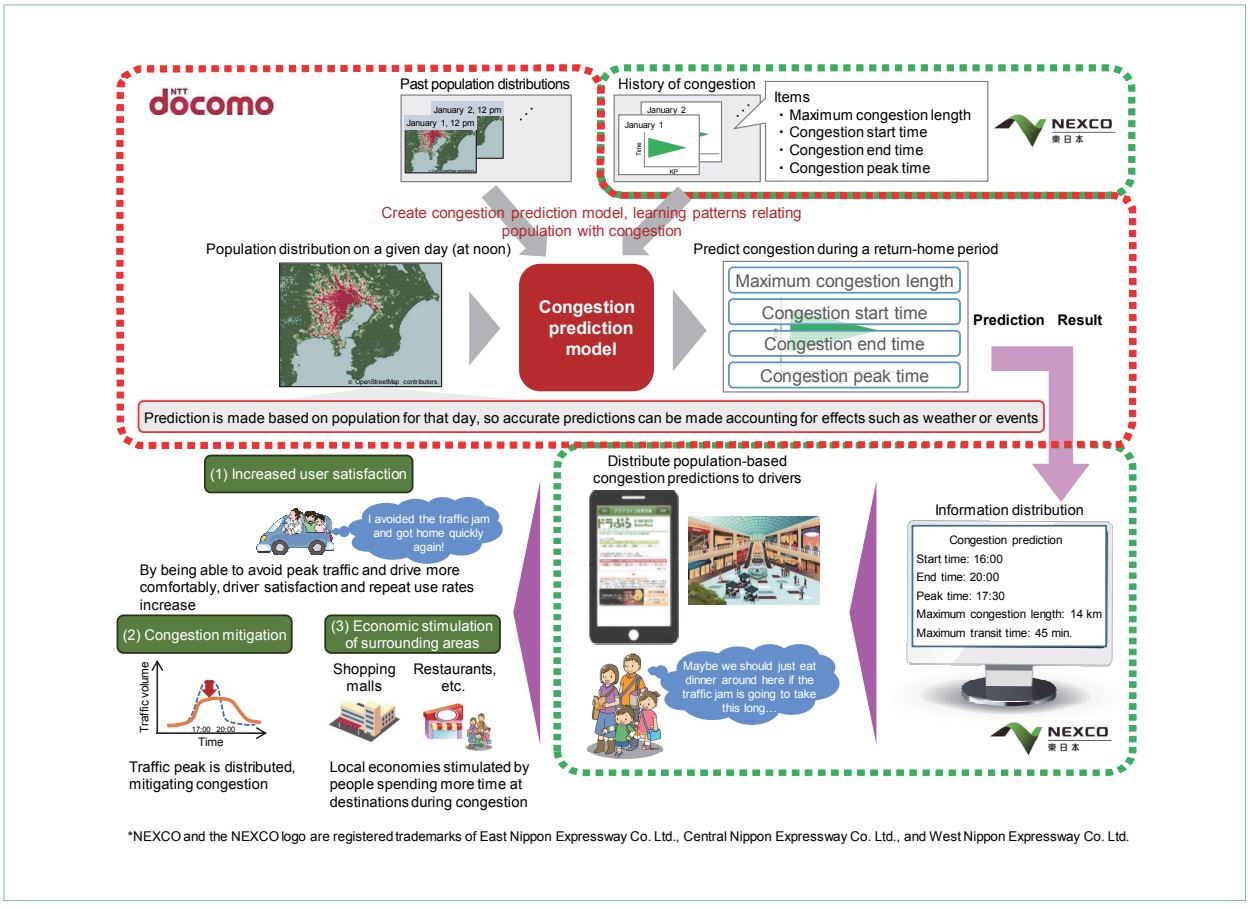
As part of implementation trials verifying the utility of Traffic Congestion Forecasting AI, we conducted a trial in collaboration with NEXCO East on the Tokyo Bay Aqua Line starting in December 2017 [3].

In this trial, we used population distributions at noon on the Boso Peninsula to predict congestion on the Kawasaki-bound lanes of the Tokyo Bay Aqua Line, which often occurs on weekend evenings and into the night. In particular, we predicted whether congestion would occur during the period from 14:00 to 24:00 based on population distributions by attribute at 12:00 in the Boso area, including residential areas. When congestion was predicted, we also predicted the start and end times, the peak time, and the physical length of the congestion at the peak time. In December 2018, we implemented new methods based on customer survey results for predicting the time required to travel the length of the congestion and the traffic demand every 30 minutes over the same time period. The survey results and new methods are described in detail below.

The results predicted by Traffic Congestion Forecasting AI are provided every day to the driving public through the Drive Plaza Web site operated by NEXCO East, which provides information on expressways in Japan. During the trial, in addition to the congestion prediction results, coupons were also issued, offering discount for meals and shopping at Kisarazu and other locations (called "Yorutoku coupons"). This was intended to mitigate congestion by spreading the return traffic over time and to stimulate local economies. An overview of the trial is shown in **Figure 4**.

4.3 Evaluation of Prediction Accuracy

Before providing prediction information to the public, we evaluated the accuracy of Traffic Congestion Forecasting AI. The evaluation was conducted over two years and four months, between



*NEXCO and the NEXCO logo are registered trademarks of East Nippon Expressway Co. Ltd., Central Nippon Expressway Co. Ltd., and West Nippon Expressway Co. Ltd.

Figure 4 Overview of collaboration with NEXCO East

January 2015 and April 2017 (excluding days with accidents and other traffic restrictions), and congestion was predicted on each day during the period using Leave-One-Out Cross Validation (LOOCV)*4. Ground truth data used for training and examination consisted of traffic history data maintained by NEXCO East for the time of the trial.

We used two indices for evaluation: the Missed-Alarm Rate (MAR) is the number of days congestion occurred even though it was predicted not to occur divided by the total number of days congestion did occur. The False-Alarm Rate (FAR) is the number of days when congestion did not occur even though it was predicted to occur divided by the total number of days congestion was predicted to occur. The accuracy of Traffic Congestion Forecasting AI is shown in **Table 1**, compared with results from "Congestion Forecast Calendar," which has been provided earlier by NEXCO East, as a benchmark.

As an example, compared with results from Congestion Forecast Calendar, the MAR went from 6% to 1% and FAR from 18% to 0% for congestion longer than 10 km. These can both be considered great improvements. For the FAR in particular, the results were improved overall. Note that only population data, and no other information (day of

*4 LOOCV: A method for evaluating the accuracy of a statistical predictor.

the week, weather, event information, etc.), was used for these predictions. These results were obtained by giving only the population distributions from Real-time Population Statistics for that day to the congestion prediction model trained with past data.

4.4 Survey Results and Introduction of a New Method

As part of the joint trial, a Web survey regarding the trial was conducted during the period from March 20 to July 9, 2018 [4]. The survey was completed by people who agreed to participate after learning of the survey through pamphlets placed in tourism facilities on the Tokyo Bay Aqua Line and in Chiba Prefecture, e-mails distributed to users of the Drive Plaza*5 and Drive Traffic*6 Web sites, and banner advertisements. Excerpts of the results are shown in **Table 2** and **Figure 5**.

Over 90% of the respondents to the survey indicated an intention to use the service in the future. In particular, approximately 95% of respondents presumed likely to use the Aqua Line frequently (those living outside of Chiba prefecture in the Kanto area, using it for leisure more than once every six months) indicated an intention to use it in the future.

The survey also confirmed a strong demand for information to be provided by time period as a desired feature of Traffic Congestion Forecasting AI in the future. Given this intent to use the service and requests for features, we developed new technology for predicting at 30-minute intervals the time needed to traverse the Aqua Line and traffic demand. We then updated the pages providing Traffic Congestion Forecasting AI information on the

Table 1 Evaluating accuracy of Traffic Congestion Forecasting AI

(a) Missed-alarm rates for actual congestion length

Congestion length	Missed-alarm rate	
	Congestion Forecast Calender	Traffic Congestion Forecasting AI
15 km and greater	2%	0%
10 km and greater	6%	1%
5 km and greater	7%	3%

(b) False-alarm rates for predicted congestion length

Congestion length	False-alarm rate	
	Congestion Forecast Calender	Traffic Congestion Forecasting AI
15 km and greater	6%	0%
10 km and greater	18%	0%
5 km and greater	22%	6%

Table 2 Intention to use Traffic Congestion Forecasting AI in the future

	Will use	Will not use
Total (n = 12,538)	90.1%	9.9%
Customers using the Aqua Line frequently (n = 1,784)	94.5%	5.5%

*5 Drive Plaza: A Web site that publishes information useful for driving holidays, mainly for expressways. Provides search for routes and tolls as well as information regarding tolls, discounts, service areas, and the areas under the jurisdiction of NEXCO East.

*6 Drive Traffic: A Web site publishing traffic information for expressways throughout Japan. Includes mainly real-time traffic restrictions and congestion, congestion forecasts, and scheduled restrictions.

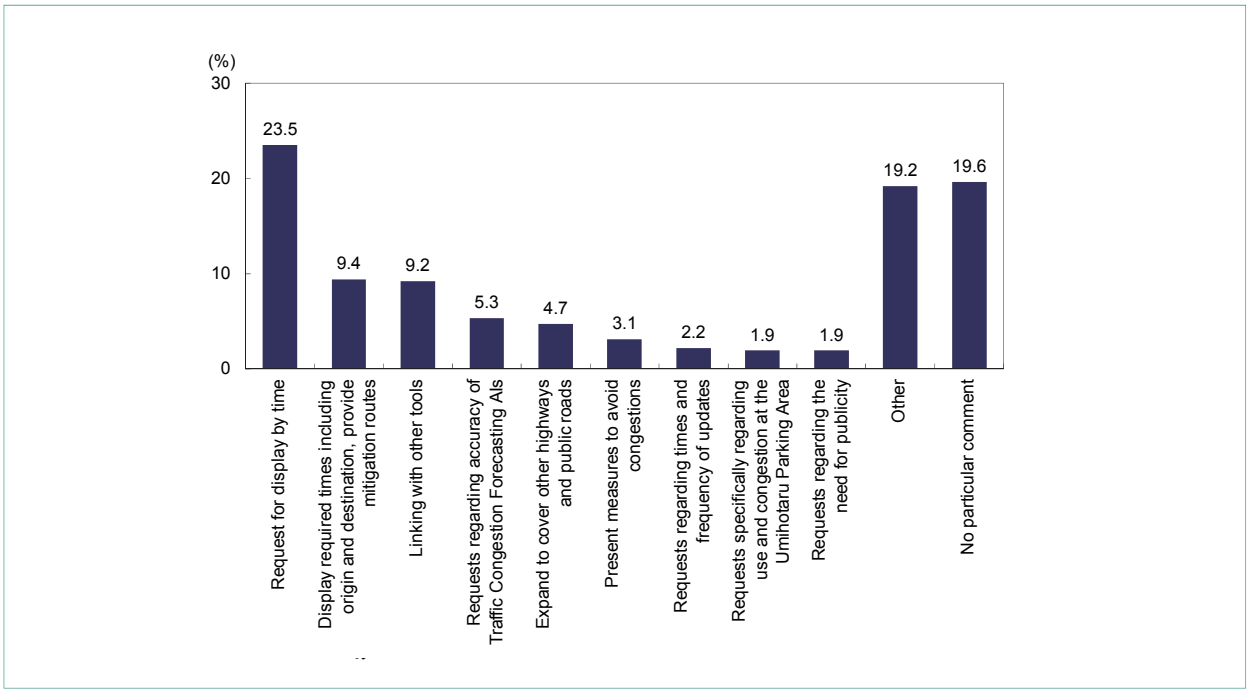


Figure 5 Opinions and requests regarding Traffic Congestion Forecasting AI (compiled from open comment field)

Drive Plaza Web site and began a new trial providing this information in December 2018 [4].

5. Conclusion

This article described Traffic Congestion Forecasting AI, a system that is able to predict traffic congestion in the *future* from Real-time Population Statistics, which estimates *current* population distributions for all of Japan based on mobile telephone network operations data. The article also described trials of Traffic Congestion Forecasting AI conducted in collaboration with NEXCO East on the Tokyo Bay Aqua Line expressway.

We are continuing trials of Traffic Congestion Forecasting AI to verify its effects and any issues and will improve and extend the system based on the results of the trials. We will continue technical development to realize more comfortable driving environments that will enable more drivers to avoid congestion.

REFERENCES

[1] H. Tanaka: "Goodbye Congestion," Nikkei Computer, pp.44–51, Aug. 2018 (In Japanese).
[2] Mobile Spatial Statistics Web page (In Japanese). https://mobaku.jp/
[3] NTT DOCOMO Press Release: "(Notice) NEXCO East and NTT DOCOMO begin trials of 'Traffic Congestion Forecasting AI' on the Tokyo Bay Aqua Line," Nov. 2017 (In Japanese).
[4] NTT DOCOMO Press Release: ""Traffic Congestion Forecasting AI' on the Tokyo Bay Aqua Line provides transit times every 30 minutes," Dec. 2018 (In Japanese).

Topics

Image Recognition | Specific Object Recognition | Food Product Judgment

Special Articles on AI Supporting a Prosperous and Diverse Society

A Food Product Judgment System Supporting Food Diversity —Enabling People Who Have Food and Drink Prohibitions to Select Foods Simply with an App—

Service Innovation Department Seiya Kojima †1 Fatina Putri †2

Certain food or drinks or ways of consuming foods are restricted by some cultures or religions. This is known as having food and drink prohibitions. There are many examples of foods whose consumption is prohibited by certain religions, for instance under the Islamic categorization of foods as Halal or Haram*1, there are many things that Muslims are not permitted to consume such as pork, pork-derived products or alcohol. In addition to religious or cultural reasons, there are also many vegetarians all over the world whose diets entail partial or full avoidance of animal-based foods for health reasons and so forth.

As more and more visitors from overseas are expected to visit Japan with the 2020 Olympics as a trigger, there will be many visitors with such food and drink prohibitions. This means it will be necessary to handle an unprecedented diversity of foods, for example handling greater numbers of food products labeled with Halal certification.

Conventionally, when Muslims or vegetarians who cannot read Japanese buy some food at a Japanese convenience store or a supermarket, they have to pick up each item and check the ingredients written in Japanese using a translation app, or check if the product is okay to eat by taking a photo of it and sending it to a friend for confirmation using a social networking service, etc. before purchasing the item. This inconvenience has even resulted in cases of travelers to Japan bringing foods from their own country to consume during their stay.

To address this issue, NTT DOCOMO has developed a "food product judgment system" for purchasing foods in convenience stores and supermarkets that enables people who have food and drink prohibitions to determine whether they can consume a product just by photographing shelved merchandise using their smartphone, etc. before they make purchase [1].

This system consists of two functions.

- The first is a food product recognition function that uses DOCOMO's "shelved merchandise

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

†1 Currently, Solution Service Department
†2 Served at the Service Innovation Department until the end of June 2019.

image recognition engine" [2]. This image recognition engine enables identification of various items on display from an image captured of the shelved merchandise.

- The second is a function to judge food products for people who have food and drink prohibitions. By combining data on the ingredients in a product with information about food and drink prohibitions (food product judgment logic), the system determines whether the product can be consumed by people who have food and drink prohibitions such as Muslims or vegetarians.

With these two functions, the system enables people who have food and drink prohibitions to determine whether they can consume a product just by photographing the shelved merchandise using their smartphone, etc. (**Figure 1**). Thus, the system reduces the bother of purchasing foods because it eliminates the need for the users to hold a product and translate what is written on its packaging to decrypt its ingredients.

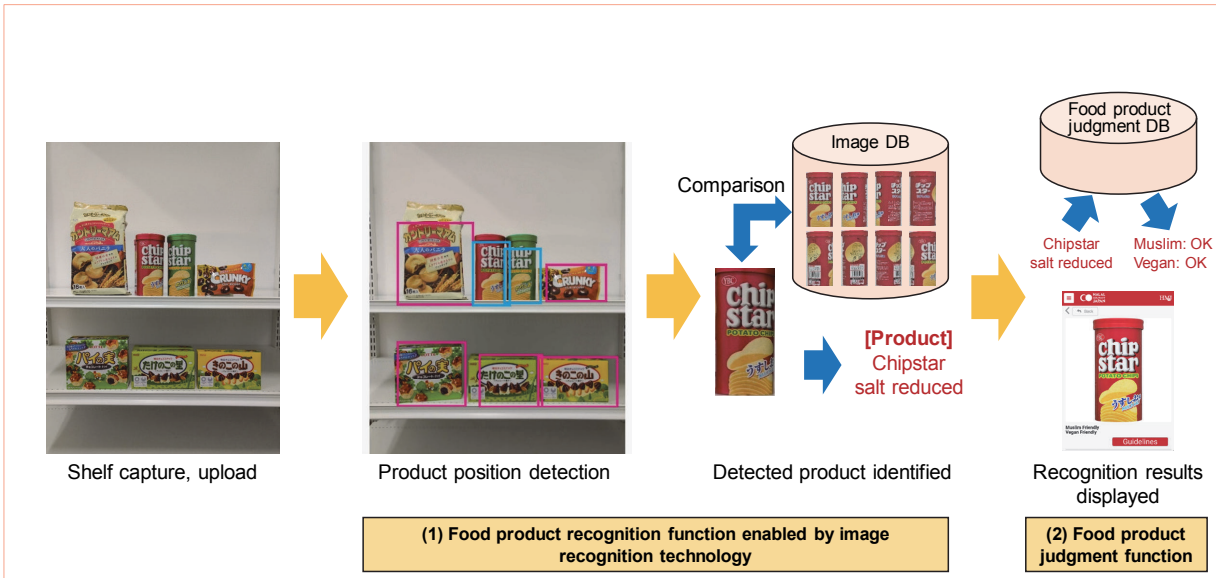
This article describes the details of the food product judgment system we developed.

1) Food Product Recognition Function Enabled by Image Recognition Technology

The food product judgment system is achieved with two image recognition technologies.

(1) Object detection technology using deep learning*2

The first technology is object detection technology that uses deep learning to analyze the position information of products from an image of shelved merchandise (**Figure 2**). An orange frame is displayed for the results of object detection, and the upper left and



Vegan: A person who practices veganism, a style of diet in which all animal-based foodstuffs are avoided, including dairy products.

Figure 1 Food product judgment system recognition flow

*1 Halal, Haram: Items that are allowed under Islam law are referred to as Halal, while items that are not allowed are referred to as Haram. Usually applied to determine whether dishes or ingredients can be consumed by Muslims.

*2 Deep learning: A method of machine learning (see *3) using a multilayered neural network.

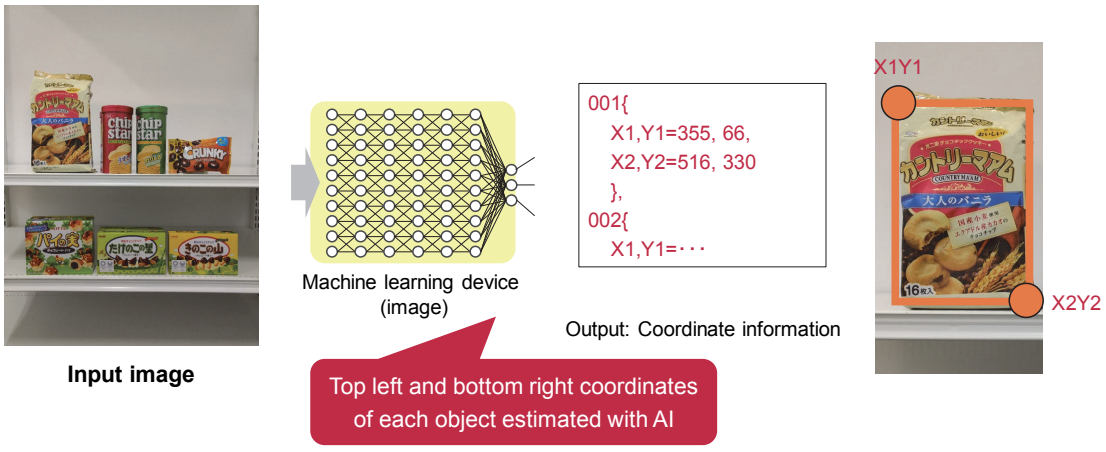


Figure 2 Object detection technology

bottom right coordinates of the frame surrounding the product are estimated. Because DOCOMO's object detection technology uses deep learning, the object detection engine must undergo machine learning*3 in advance to detect the desired objects. Hundreds and thousands of images of product displays in various actual stores and annotation data*4 have been prepared for this deep learning. Teaching this data to the system enables it to detect products with a high degree of accuracy even when products are crammed into tiny display spaces. Please refer to reference [3] for details of the object detection technology algorithm.

(2) Specific object recognition technology using local feature values*5

Second is specific object recognition technology that uses local feature values to identify products from partial images in the merchandise area detected as described above (**Figure 3**). As the target image, the merchandise area is input and compared to large amounts of product image data preregistered in an image database. This identifies products in the merchandise area by determining whether the input image is similar to any preregistered images. Images of products captured from various angles are preregistered in the database to make it possible to compute the similarity of the input image to the preregistered images. However, this could be impractical because the level of similarity is computed by comparing with all of the large number of images in the database, which could take several tens of seconds or more for one image of shelved merchandise. We addressed this issue with our specific object

*3 Machine learning: Technology that enables computers to acquire knowledge, decision criteria or behaviors, etc. from data in ways similar to how humans acquire these things from perception and experience.

*4 Annotation data: In this article, refers to metadata indicating what is in an image.

*5 Local feature values: Extracted from data, values (numbers) that characterize the data. In this article, "feature values" refers specifically to image feature values, which are characteristic points (corners) extracted from the image and the surrounding distribution of brightness.

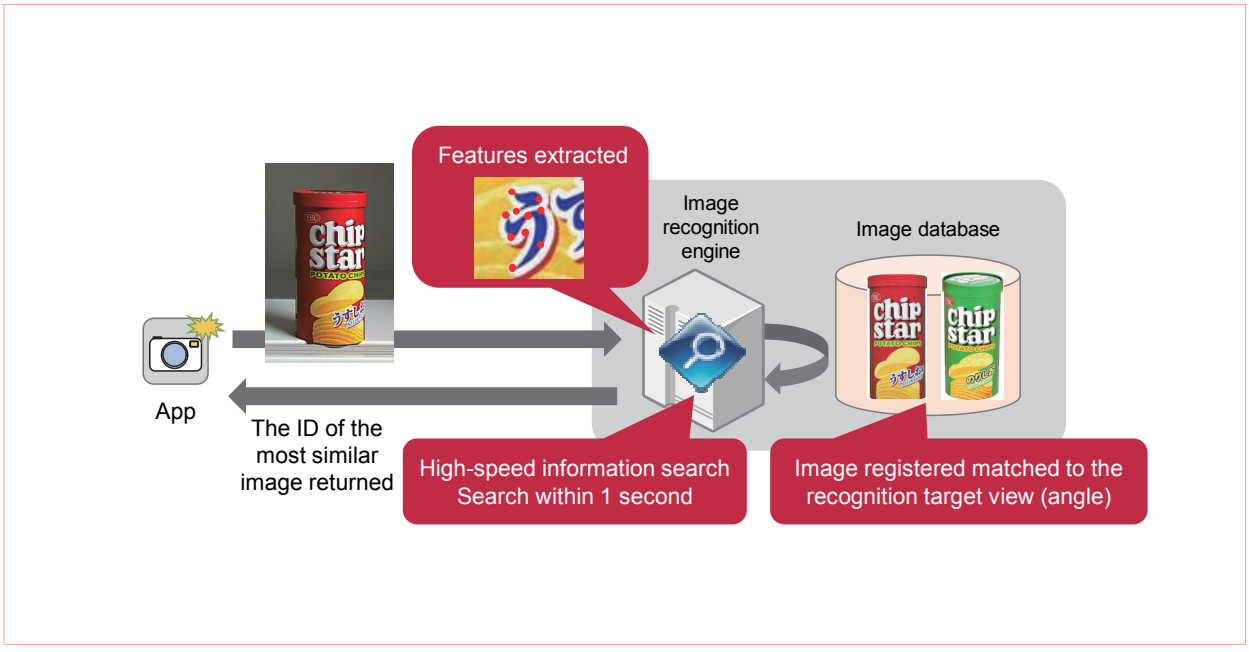


Figure 3 Specific object detection technology

recognition technology. With this technology, high-speed, high accuracy recognition of not only front view images but also images from different angles enabled by an algorithm we developed within one second from the several million preregistered images captured from various angles in the large database. Please refer to reference [4] for details of the specific object detection algorithm.

Using these two technologies, the position and the identity of the product are recognized.

2) Food Product Judgment Function

Food product judgment is enabled for products in the photograph by referencing recognition results from the image recognition engine with the database. Judgment information based on information about product ingredients is preregistered in the database to determine whether products are Muslim or vegetarian-friendly. The food product judgment logic has been achieved through collaboration with FOOD DIVERSITY Inc., a company in Japan taking initiatives regarding food and drink prohibitions particularly for Muslims and vegetarians. Products are judged to be Muslim or vegetarian-friendly based on information about the primary ingredients*6 of products.

3) Application Provision

A trial offering is underway of a food product judgment service incorporating the food product judgment system with the aforementioned two functions in the "Halal Gourmet Japan*7" restaurant search app provided by FOOD DIVERSITY Inc. for Muslims and vegetarians (**Figure 4**) [5]. This app displays food products that are Muslim or vegetarian-friendly in the image captured of shelved merchandise in different colors – Muslim-friendly products are displayed in a red frame, while Muslim and

*6 Primary ingredients: The ingredients that directly comprise a final product. With foods products in particular, these are the ingredients listed on the product label.

*7 Halal Gourmet Japan: A smartphone application designed for Muslims that enables search of restaurant information, etc., and is operated by FOOD DIVERSITY Inc.

vegetarian-friendly products are displayed in a blue frame. Products that are not Muslim and vegetarian-friendly or are unregistered are displayed with a white frame. Users can tap the product in the colored frame to display details about it. "Muslim-friendly" is displayed for Muslim-friendly products.

Figure 4 Food product judgment service trial offering

In addition to judgments about whether products are Muslim or vegetarian-friendly, the details screen also clarifies ingredient names and informs whether the products are Muslim or vegetarian-friendly under the different judgments of different people*8 in a way that is easy to understand (**Figure 5**).

This article has described the two functions of the food product judgment system we developed, and introduced a trial offering of a food product judgment service using those functions.

The NTT DOCOMO image recognition technology used with this system is capable of identifying food products on display just by capturing an image of the shelved merchandise, and thus in addition to Muslims and vegetarians, it could be used to provide services to people with various other food issues just by mapping information to various

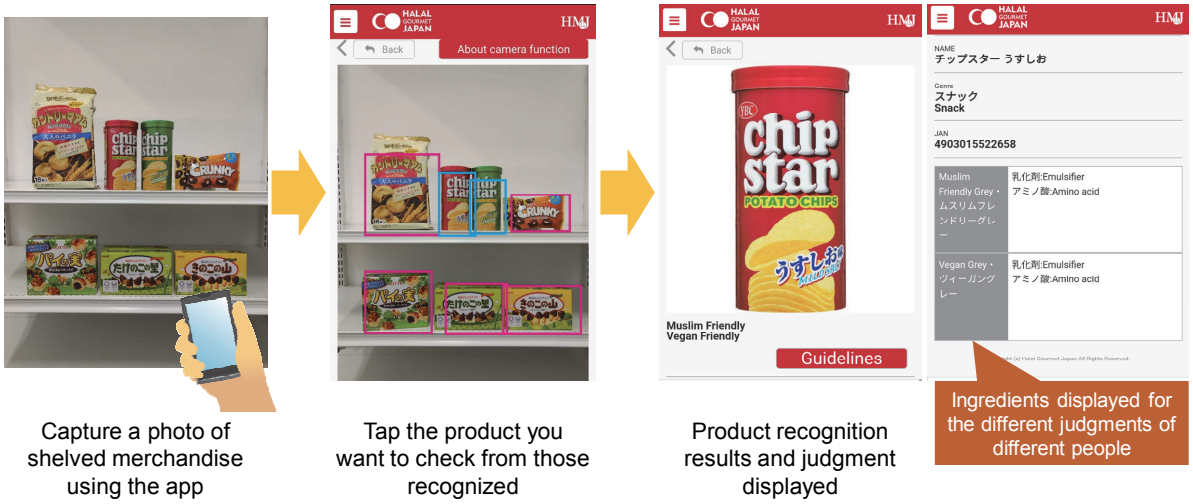


Figure 5 "Halal Gourmet Japan" food product judgment service usage image

*8 Different judgments of different people: In this system, this refers to such things as fresh cream. Because Muslim judgments are based on a promise between oneself and Allah, ultimately decisions about whether a food product can be consumed are up to the individual, and standards are not uniform.

products. For example, information about allergens*9 could be added so that people with various allergies could determine whether food products are consumable, or information about low-protein or low sugar foods or even information for picky eaters, etc. could be added to enable judgments. We plan to make judgment of these possible in the future.

Since only a limited number and types of products are currently registered, we will expand the numbers and types of food products that can be identified by this system by partnering with food providers and retailers.

This will enable the system to respond smoothly to various dietary restrictions. We greatly expect that this system will help to get the increasing numbers of visitors to Japan to recognize that when it comes to food, Japan is a safe place where people can feel confident about the things they eat.

REFERENCES

[1] NTT DOCOMO Press Release: "(Announcement) NTT DOCOMO Develops a "Food Product Judgment System" that Lets Muslims and Vegetarians Determine with an App what Food Products They Can Eat – Just by taking a picture of shelved merchandise with a smartphone –," Sep. 2018 (In Japanese). https://www.nttdocomo.co.jp/info/news_release/2018/09/26_01.html

[2] NTT DOCOMO Press Release: "DOCOMO Launches AI Engine for Fast, Accurate Shelf Analysis – Recognizes shelf allocation by analyzing photos of shelved merchandise –," Mar. 2018. https://www.nttdocomo.co.jp/english/info/media_center/pr/2018/0316_00.html

[3] H. Akatsuka et al.: "A Retail Shelving Analysis Solution Using Image Recognition – Recognizes Shelving Allocation and Quantifies Inventory by Analyzing Photos of Shelved Merchandize," NTT DOCOMO Technical Journal, Vol.20, No.2, pp.22–31, Nov. 2018.

[4] H. Akatsuka et al.: "High-speed, Large-scale Image Recognition and API," NTT DOCOMO Technical Journal, Vol.17, No.1, pp.10–17, Jul. 2015.

[5] HALAL MEDIA JAPAN: ""Just Take a Picture" and You Will Know Which Products are "Muslim-Friendly"." https://www.youtube.com/watch?v=78fsdhHb8CQ

*9 Allergen: A substance which can cause an allergic reaction. Among foodstuffs, allergens are defined as specific ingredients in food items such as shrimp, crab or wheat that can cause a high incidence or severity of symptoms of allergic reactions.

Technology Reports | Telephone Extension Solution | Telephone Conference Service | Office Link

The “Office Link Voice Conferencing Service” —A New Telephone Conferencing System Using the Office Link Platform—

Core Network Development Department Taku Nakamura Tomohiro Takami
Masahiro Hayakawa
DOCOMO Technology, Inc. Core Network Division Jyunpei Miyoshi

As one of its voice communications services for business-use, NTT DOCOMO provides “Office Link” for extension service between the fixed telephone lines and business-use mobile phones of corporate users. This service has been well received not only for the availability of PBX extensions within companies, but also its availability in FOMA/Xi (VoLTE) areas. As an addition to the service, NTT DOCOMO developed the “Office Link Voice Conferencing Service” as a new telephone conferencing system incorporated into the Office Link platform. This article describes an overview of the service and some of its technical aspects.

1. Introduction

From September 2010, NTT DOCOMO began providing the “Office Link” service to connect corporate Private Branch eXchange (PBX)*1 with its FOMA network and to make in-house telephone extensions accessible from DOCOMO mobile telephones [1]. Just as the name says, the service makes it possible to use a company’s telephone extensions in locations outside the office. In addition to one-to-one communications between telephone extensions, recent years have seen growing needs for systems to handle “many-to-many” conferencing. Thus, NTT DOCOMO developed new telephone conferencing service functions for its Office Link in-house telephone extension service system, which has already been established as a network service, and began providing its “Office Link Voice Conferencing Service” as a service for simultaneous broadcast on both external lines and in-house extensions.

©2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

*1 PBX: An enterprise private branch exchange. It has functions for both extension and external line connections.

To implement this system, NTT DOCOMO included rich functionality with the Web Customer Control[*2] to improve user convenience, and made efforts to reduce facility division loss when processing conferences with large numbers of people on one server to maximize facility efficiency. This new system holds promise for deployment as a telephone extension solution with high added value.

This article describes an overview of the Office Link Voice Conferencing Service, and how it is realized.

2. Details of the Office Link Voice Conferencing Service Provision

2.1 Service Overview

NTT DOCOMO began providing the Office Link Voice Conferencing Service as an additional Office Link service available with in-house extensions and with nationwide FOMA/Voice over LTE (VoLTE)[*3]. Conventionally, the Voice Meeting[*4] service provided by NTT DOCOMO only handled participation in conferences via external lines, whereas this service makes use of the Office Link platform to provide availability to in-house extensions as well. The service makes it possible for customers to engage in simultaneous broadcast with an extension number no matter where they are in Japan, and since communications fees are included in the flat Office Link rates, it also offers savings on the communication fees associated with conventional calling of external numbers. **Table 1** shows the characteristics of the Office Link Voice Conferencing Service and the Voice Meeting service.

2.2 Functions Provided with the Office Link Voice Conferencing Service

The Office Link Voice Conferencing Service can be used as a participant-calling-type or system-calling-type telephone conferencing system, and enables connection with telephone extensions used in offices as well as ordinary external mobile and fixed phones. We describe the participant-calling-type and system-calling-type conferencing offered by this service as follows.

1) Participant-calling-type Conferencing

Participant-calling-type conferencing is a form of conferencing in which members participate by calling a conference number. The conference host uses the Web Customer Control to book the conference in advance. When it is time to call the conference, the conference host performs operations to open the conference. Methods of opening the conference include the conference host calling the conference number from their terminal, or opening the conference from the Web Customer Control screen. Once the conference has been opened, participants can join by calling the conference number notified

Table 1 Office Link broadcast service, Voice Meeting service

Service name	Conference type	Terminal type	Max. number of participants
Office Link Voice Conferencing Service	Participant-calling-type/ system-calling-type	In-house telephone extensions/ external phones	200 persons
Voice Meeting service	System-calling-type	External phones	200 persons

*2 Web Customer Control: A Web site (Web Customer Control Site) on the Office Link platform accessible from a PC, smartphone or i-mode browser, which enables users to make and edit settings for conferences and holding conferences from a Web screen.

*3 VoLTE: A function to provide voice services over LTE using packet switching technologies.

*4 Voice Meeting: A service that enables simultaneous broadcasting with external line calling. The service has been providing participant-calling-type telephone conferencing from January 7, 2019.

beforehand from their terminals (their mobile phones or in-house telephone extensions, etc.). The flow for usage is described in **Figure 1**.

Participant-calling-type conferences do not require participants to preregister and participants can join a conference from any terminal. Thus, this type can be used for regular meetings or conferences where participants are not fixed in advance.

2) System-calling-type Conferencing

System-calling-type conferencing is a form of conferencing in which preregistered participants are called when the conference starts. The conference host uses the Web Customer Control to book the conference in advance and register conference participants. When it is time to call the conference, the Office Link Voice Conferencing Service system calls the terminals of all the participants on the participant list. Then, participants join the conference by responding to the call. In the same way as participant-calling-type conferencing, methods of opening the conference include the conference host calling the conference number from their terminal, or opening the conference from the Web Customer Control screen. The flow for usage is described in **Figure 2**.

Because preregistering participants is required with system-calling-type conferencing, it can be used for conferencing in which the participants to be called are clearly defined such as meetings for emergency information sharing.

3. Office Link Voice Conferencing Service System Overview

3.1 System Configuration

Figure 3 shows the structure of the Office Link Voice Conferencing Service system.

To realize this service, the Office Link platform system (1), which manages extension services, uses the conference server (2) and conference information management server (3) to provide the telephone conferencing service with accessibility from both in-

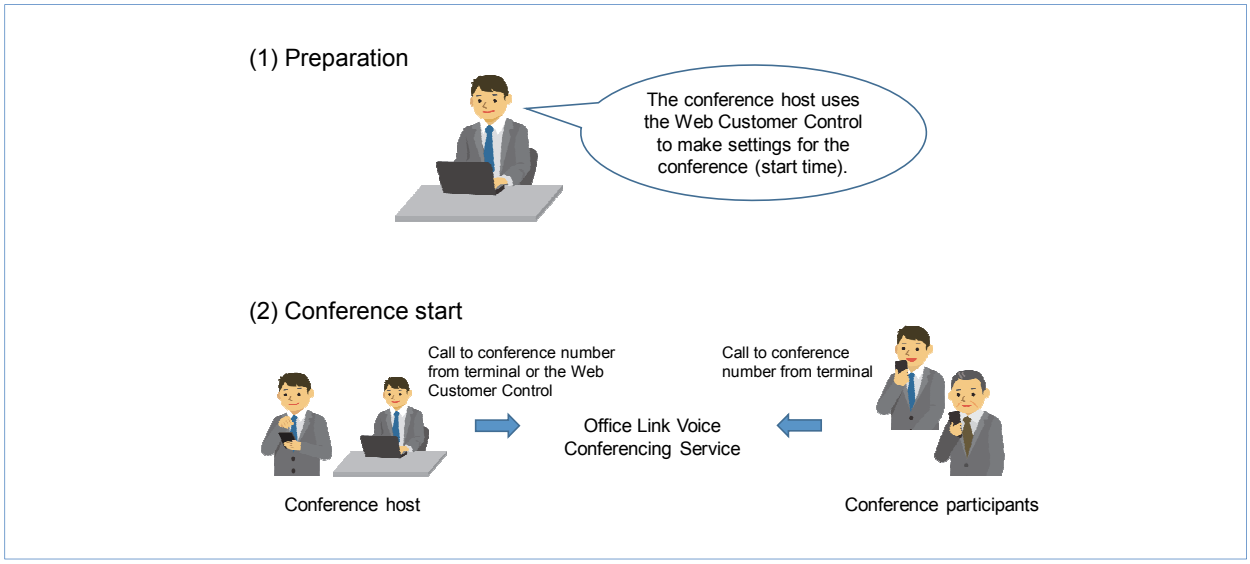


Figure 1 Flow of usage for participant-calling-type conferencing

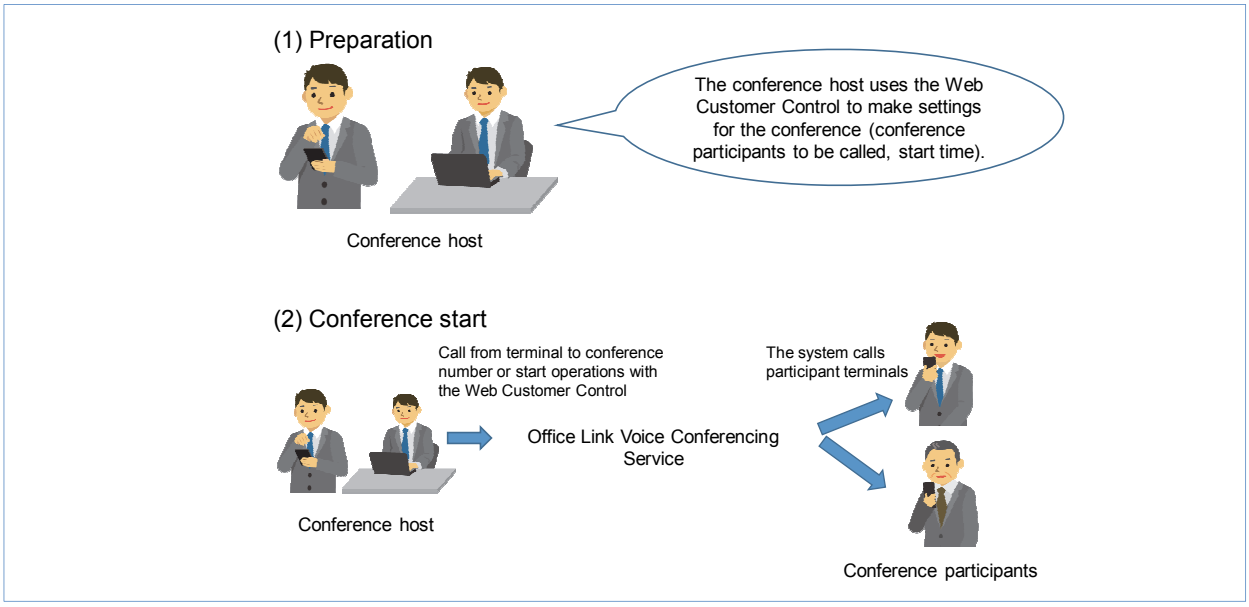


Figure 2 Flow of usage for system-calling-type conferencing

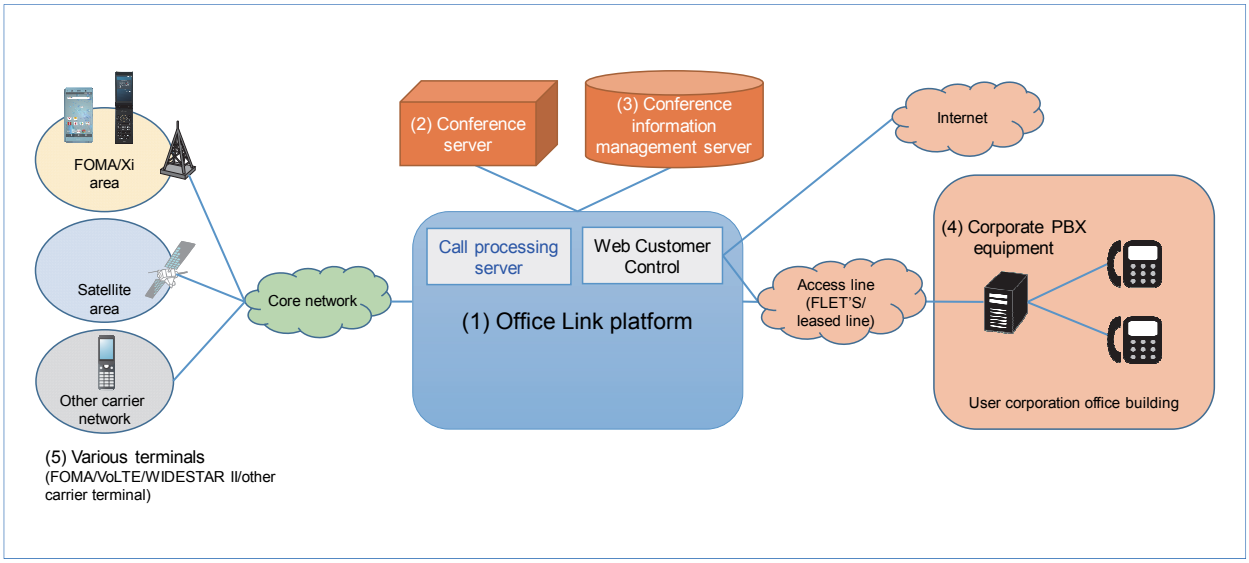


Figure 3 Overview of the Office Link Voice Conferencing Service system configuration

house extensions and external lines. The telephone conferencing service call control and voice transfer uses the core network*5 in the same way as the conventional Office Link, while the conference server performs voice mixing and distribution for the connected terminals.

3.2 Function Distribution

(1) Office Link platform

Incoming/outgoing call and voice transfer

*5 Core network: A network consisting of switching entities and subscriber information management equipment, etc. Mobile terminals communicate with the core network via the radio access network.

control functions provided with the call processing server of the Office Link platform are also provided with the Office Link Voice Conferencing Service. Also, in participant-calling-type conferencing, conference settings such as registration of participant lists are provided over the conventional Web Customer Control functions on the Office Link platform.

(2) Conference server

The conference server manages telephone conferencing, receives requests to reserve conferences, and commences conferences. After a conference starts, the conference server performs voice mixing and simultaneous distribution to participant terminals, and detects Dual-Tone Multi-Frequency (DTMF) tones*6 with the Office Link Voice Conferencing Service as telephone conference U-Plane*7 processing.

(3) Conference information management server

The conference information management server performs management and approval of conference information based on reservations made by the conference host with the Web Customer Control, and orders the call processing server to call participants based on participant lists.

(4) Corporate PBX equipment

The corporate PBX equipment connects to the Office Link platform via an access line and makes it possible to use the Office Link Voice Conferencing Service by using IP telephone extensions within customer premises.

(5) Various terminals (FOMA/VoLTE/WIDESTAR II*8/other carrier terminal)

When using the Office Link Voice Conferencing Service, connection to the conference server through incoming/outgoing calls to terminal numbers and conference operations (acquiring speaking rights, etc.) through DTMF tone transmissions are performed from various terminals. Both of these are available with existing telephone functions. FOMA/VoLTE terminals on the DOCOMO network can also originate and receive calls with Office Link extension numbers. All communications between various terminals and the Office Link platform are VoIP. Non-VoIP communications such as those using circuit switching are terminated in the core network and converted to VoIP communications.

3.3 Conference Server Cascade Connection Method

The voices of participants in a conference are mixed in the Office Link Voice Conferencing Service. However, since up to 200 people connect simultaneously with this service, if only one conference server is used and the number of people exceeds the number of spare channels, the conference cannot start. Furthermore, if a conference is held in which the number of participants exceeds the spare channels of a single conference server it is difficult to fully utilize the spare channels which degrades the facility usage rate.

To address this issue, we made it possible to hold conferences using cascade connections to multiple conference servers if the channels required for the number of participants in a conference cannot be

*6 DTMF tone: Also referred to as a push signal. The tones can be used to send a total of 16 different signals using four combinations of the numerals 0 through 9 and the asterisk(*), pound sign (#), and high and low tones from A to D.

*7 U-Plane: In contrast to the C-Plane, which carries signaling traffic and is responsible for routing, U-Plane is used for the transmission of user data. On the Office Link platform, user data refers to VoIP calling audio (RTP/RTCP).

*8 WIDESTAR II: The name of a satellite telephone service provided by NTT DOCOMO.

covered by a single conference server. **Figure 4** shows this connection method. In this method, a few spare channels in multiple conference servers are used with cascade connections to enable provision of large conferences, which enables the maximum usage rate of conference server connection channels and prevents the rejection of conferences due to a lack of connection channels.

4. Method of Achieving New Services

4.1 Mixing Extensions and External Phones in the Same Conference

It’s possible to set conference phone numbers used for telephone conferencing with the Office Link Voice Conferencing Service. The conference phone numbers are call destination numbers with participant-calling-type conferencing, and original caller numbers with system-calling-type conferencing.

The conference phone number is any extension number set by the user for each conference or a 050- number temporarily issued to users by NTT DOCOMO. Conference participants can set extension numbers from the conference settings screen of the Web Customer Control. This makes it possible for participation in a telephone conference from both extensions and external lines, to suit the various terminal types of the conference participants.

4.2 Participant-calling-type Conferencing System Operations

Figure 5 shows the sequence for holding a participant-calling-type conference. The sequence of holding a conference is the same for all types of terminals because all terminals that can connect

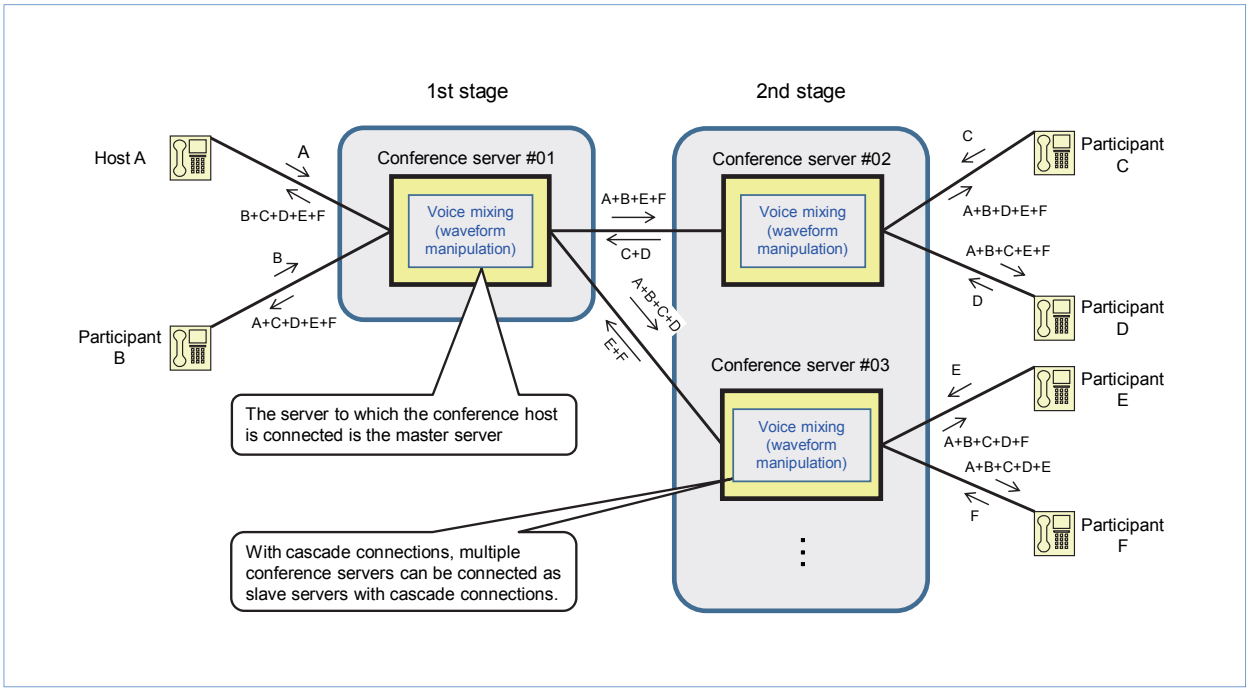


Figure 4 Cascade connection method

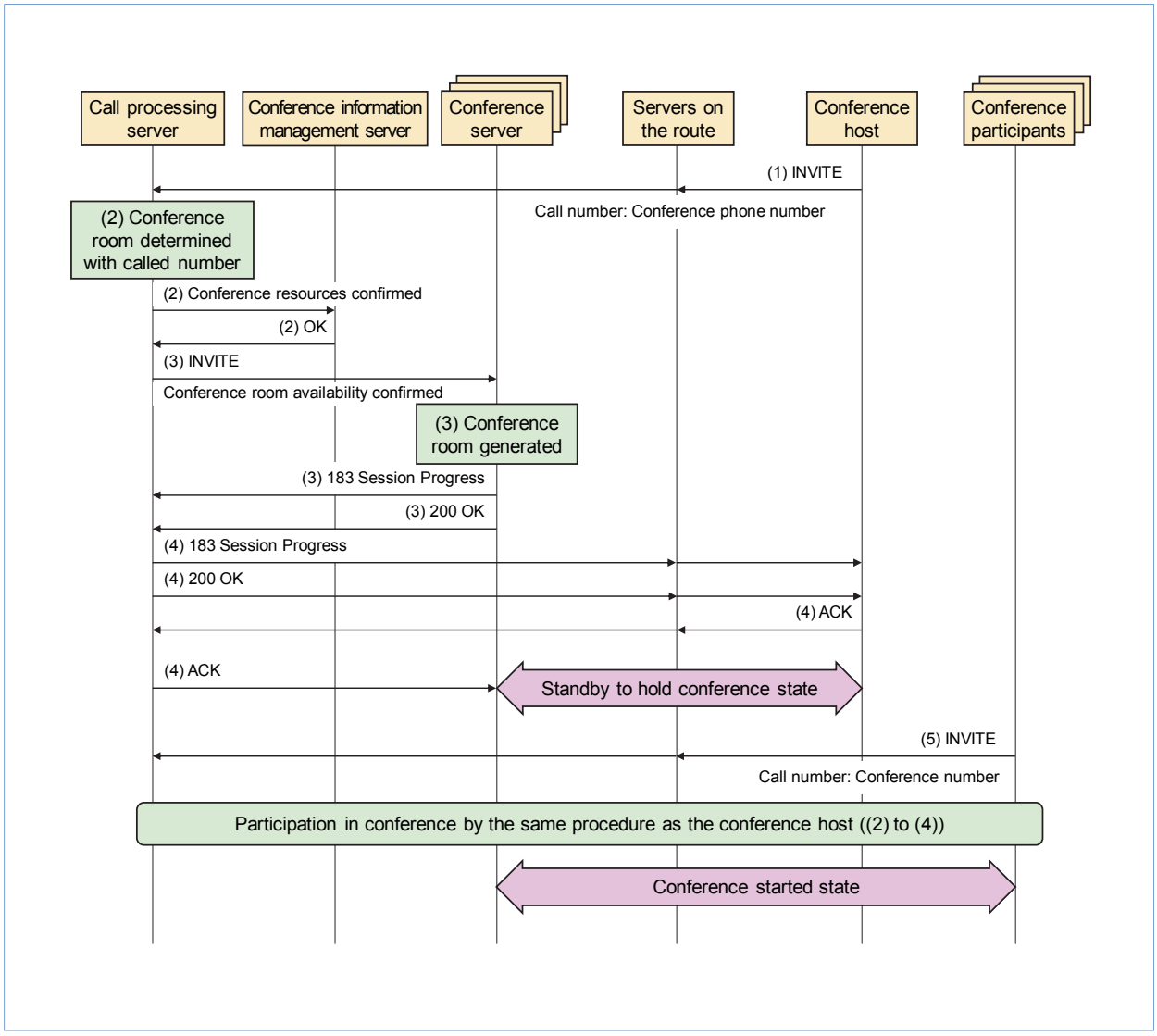


Figure 5 Sequence for starting a participant-calling-type conference

to the Office Link Voice Conferencing Service are processed with VoIP on the Office Link platform.

(1) When the conference host calls a conference number, a Session Initiation Protocol (SIP)*9 signal INVITE (start session*10 request) arrives at a call processing server in the system.

(2) The call processing server that received INVITE confirms whether the conference was booked in advance based on the conference information (conference phone number) notified in the INVITE. After confirmation, if the request can be processed, the conference information management server confirms the availability of conference resources.

(3) If resources are available, the call processing

*9 SIP: A call control protocol defined by the Internet Engineering Task Force (IETF) and used for IP telephony with VoIP, etc.

*10 Session: A virtual communication path for transmitting data or the transmission of data itself.

server sends a request to confirm the presence of a conference to the conference server. Since the conference was not yet generated when the conference host calls, a conference is generated and a SIP signal 183: Session Progress/200: OK is sent to the call processing server.

(4) The call processing server that receives the signal sends the same signal to the conference host, and a connection is established between the conference host and the system when the conference host responds to the signal.

(5) After the conference host establishes the connection, conference participants call the conference phone number and participate in the conference through the same procedure as the conference host.

4.3 System-calling-type Conferencing System Operations

With system-calling-type conferencing, conferences can start on a date specified when the conference host generates the conference or be held immediately with the Web Customer Control (hereinafter referred to as “Web Customer Control-generated”). With the former, the sequence of operations is the same as the participant-calling-type conference until the connection is established between the conference host and the system. After that, the call processing server calls the conference participants on the preregistered participant list, who can participate in the conference simply by responding.

Figure 6 shows the sequence for holding a Web Customer Control-generated conference. With this type of conference, the call processing server calls both the conference host and participants.

(1) The conference host presses the start conference button after inputting conference information with the Web Customer Control to initiate processing in the system.

(2) The server that provides the Web Customer Control functions (Web Customer Control server) checks conference information to confirm that conditions to hold the conference are satisfied. If conditions are satisfied, conference reservation information is written into the conference information management server by the Web Customer Control server, conference resources are confirmed, and if available the call processing server calls the conference host.

(3) Then, the call processing server asks the conference server if there is a conference, and because a conference has not been generated at this point, the conference server generates a conference.

(4) If a conference has been generated, the call processing server sends INVITE to the conference host, and a connection between the conference host and the system is established when the conference host responds.

(5) After that, the call processing server references the participant list stored in the conference information management server, and then calls participants and starts the conference when they respond.

5. Conclusion

With the growing diversity of DOCOMO’s in-house telephone extension services for business-use,

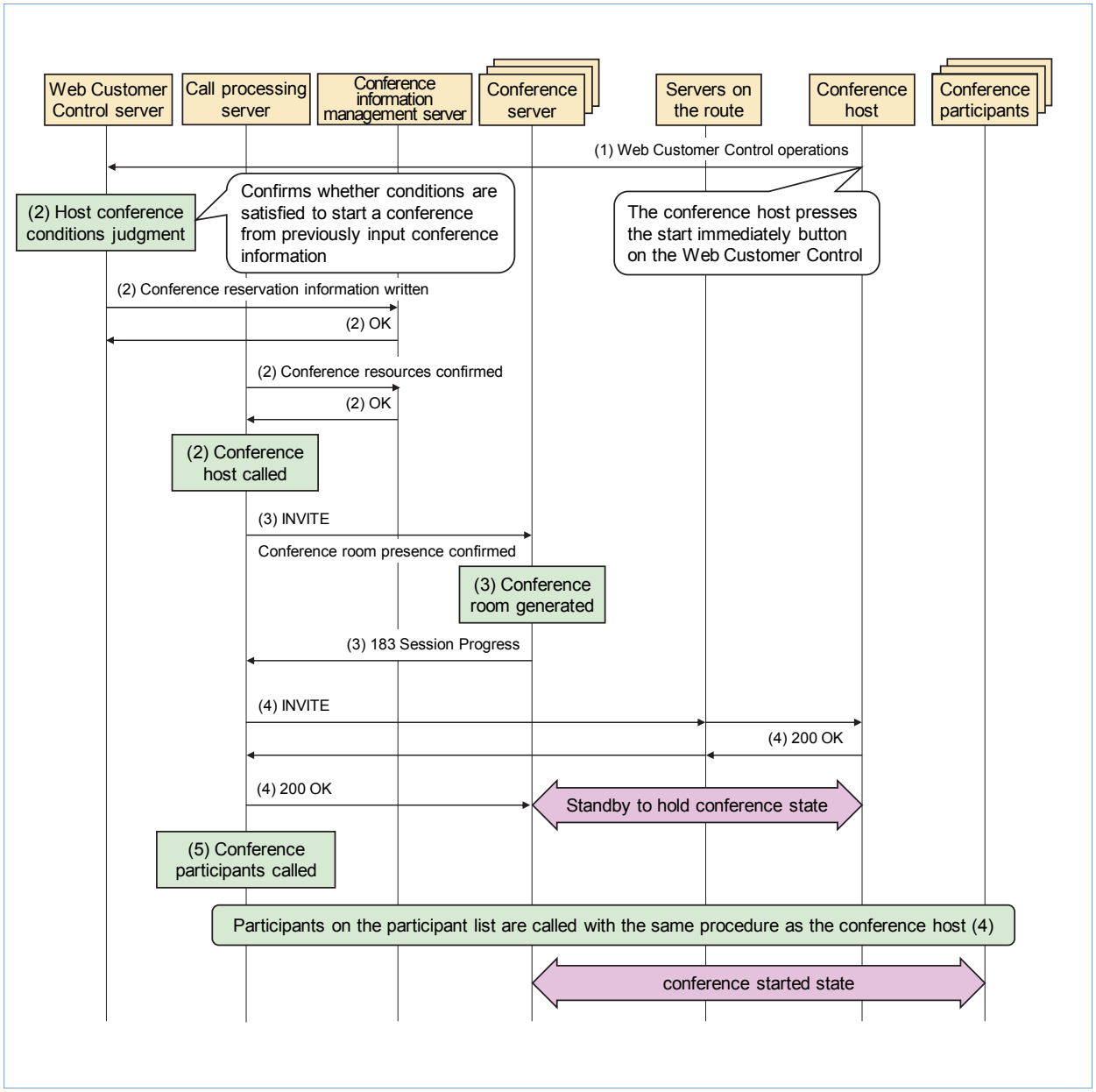


Figure 6 Sequence for holding a system-calling-type conference (Web Customer Control-generated)

NTT DOCOMO enhanced the Office Link platform and began providing the Office Link Voice Conferencing Service as a service enabling simultaneous broadcast even for in-house extensions. This has enabled new telephone conferencing using the in-house telephone extensions of users with Office Link subscriptions, and enables business support across a wide range of corporate user scenes.

Going forward, to promote DOCOMO’s “+d[*11]” midterm strategy of co-creating value with partners

*11 +d: The name of an NTT DOCOMO initiative for creating new value with partner companies.

such as linking the Office Link system with third parties, we will develop corporate services with even more value by advancing the system to enable linkage of services both in and out of NTT DOCOMO with the Office Link platform.

REFERENCE

[1] S. Koga et al.: “Office Link System for FOMA Internal Line Connection,” NTT DOCOMO Technical Journal, Vol.11, No.4, pp.33–38, Mar. 2010.

NTT Group Receives the “Derwent Top 100 Global Innovators 2018-19” Award
—NTT DOCOMO Activities Contribute to Earning This Award—

In January 2019, Nippon Telegraph and Telephone Corporation (hereinafter referred to as “NTT”) was recognized as one of the “Derwent Top 100 Global Innovators 2018-19” by Clarivate Analytics (Headquarters: Philadelphia, U.S.A.). This is the 8th consecutive year that NTT has received the award, following from recognition in the former “Thomson Reuters Top 100 Global Innovators.”

Clarivate Analytics presents awards to 100 global companies selected as “innovative organizations who are committed to respecting and protecting IPRs and who successfully develop valuable patented inventions with strong global impact.” The selection is made through a methodology developed by Clarivate Analytics themselves and based on four criteria: Volume (of filed patents), Success (the percentage of patents granted to patent applications filed with patent offices), Globalization and Influence. This is the eighth consecutive year that NTT has received this award.

NTT was awarded for global recognition of the advanced R&D undertaken by the entire NTT Group, and the value of the inventions and achievements brought about through these efforts.

As a member of the NTT Group, NTT DOCOMO has made many applications worldwide for patents for fundamental technologies in the mobile communications field such as 5G and LTE. DOCOMO’s approaches to R&D activities were also appraised as crucial elements leading to the winning of the award this time around.

NTT DOCOMO
Technical Journal Vol.21 No.2

Editorship and Publication
NTT DOCOMO Technical Journal is a quarterly journal edited by NTT DOCOMO, INC. and published by The Telecommunications Association.

Editorial Correspondence
NTT DOCOMO Technical Journal Editorial Office
R&D Strategy Department
NTT DOCOMO, INC.
Sanno Park Tower
2-11-1, Nagata-cho, Chiyoda-ku, Tokyo 100-6150, Japan
e-mail: dtj@nttdocomo.com

Copyright
© 2019 NTT DOCOMO, INC.
Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.