Natural Language Processing Technology for Agent Services

Hisako Asano, Katsuji Bessho, Kugatsu Sadamitsu, Kyosuke Nishida, and Kuniko Saito

Abstract

Artificial intelligence agents are beginning to take on the work traditionally performed by people at contact centers and information desks. This article introduces natural language processing technology for achieving agents that can absorb business knowledge, understand customer speech and respond to their requests, engage in casual conversation, and behave like a human being overall.

Keywords: agent, natural language processing, artificial intelligence

1. Introduction

An agent service is a type of service in which an artificial intelligence (AI) agent takes the place of a human being. Businesses such as contact centers and information desks that respond to inquiries based on specialized knowledge have growing needs. The user (customer) of an agent service makes inquiries or requests by talking or chatting by text with an agent. The agent, meanwhile, makes an attempt to understand what the user is saying, and based on a previously learned business flow and business knowledge, provides the information the user desires, or executes a procedure requested by the user after confirming with the user the information needed for that procedure. At this time, a natural response from the agent might include casual conversation in addition to dealing with the current task. Such an exchange between a user and agent is carried out in everyday language used by people, that is, in *natural language*.

We note here that the huge amount of specialized knowledge in business documents such as manuals that must be acquired by business personnel is also written in natural language. An AI agent, however, performs processing using computer-oriented (machine) languages (database query languages, web application programming interfaces (APIs), etc.) and structured data (relational databases, etc.). Therefore, information in user utterances and business documents expressed in natural language must be converted to a machine-readable form.

Agent-AI powered by the NTT Group AI technology called corevo® is technology that supports humans by interpreting the information they generate. Natural language processing technology is one important technology making up Agent-AI [1]. This article introduces natural language processing technology for achieving agents that can respond accurately to specialized inquiries from users while being equipped with a variety of knowledge groups enabling intelligent and friendly conversation on a wide range of topics with users (**Fig. 1**).

2. Natural-language equivalence judgement technology for understanding diverse expressions

An agent possesses knowledge expressed in natural language, and its function is to compare natural-language utterances made by the user with that natural-language knowledge. For example, frequently asked questions (FAQ) is a document or website link consisting of knowledge written in natural language in the form of question (Q) and answer (A) pairs. As such, it can be used to respond to a query from the user by comparing user utterances with FAQ questions and presenting to the user the answer (A) corresponding to the question (Q) matching those utterances.
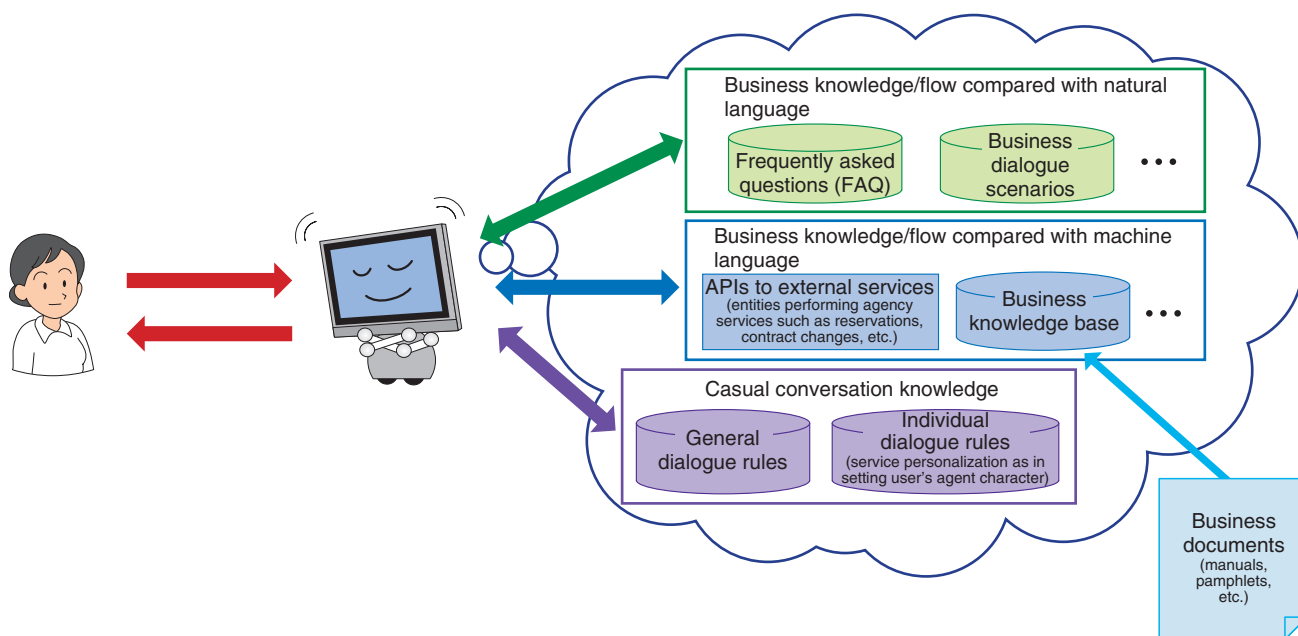
Fig. 1.   Agent equipped with diverse knowledge groups.

In addition, each dialogue rule in business dialogue scenarios (a set of dialogue rules based on a business flow) consists of a user utterance (U) and system utterance (S) pair. These rules can be used to compare a user's utterance with stored user utterances (U) and return the system utterance (S) corresponding to the user utterance (U) matching the actual utterance made by the user. However, agent knowledge targeted for comparison rarely matches a user utterance at the character-string level given the diversity of user expressions, so there is a need for technology that can determine equivalence between expressions in terms of intention.

To meet this need, NTT Media Intelligence Laboratories developed natural-language equivalence judgement technology (**Fig. 2**). This technology dynamically constructs quantified semantic concepts of words based on a large-scale Japanese language semantic dictionary and on word occurrence distribution in large volumes of text. It then judges semantic similarity in natural-language statements targeted for comparison. This approach makes it possible to understand diverse expressions characteristic of the Japanese language and to identify with high accuracy agent knowledge that semantically agrees with user utterances.

3.   Information extraction technology for providing information and executing external services from user utterances

NTT Media Intelligence Laboratories is also researching and developing technology for extracting information from user utterances for two main purposes. The first is to provide pinpoint information using business knowledge such as fees for specific product options. The second is to perform procedures or obtain information using systems external to the agent as in making/checking reservations or checking inventory.

First, we describe the flow of providing pinpoint information based on business knowledge, referring to **Fig. 3**. In response to a user request for information, the agent provides appropriate information using a knowledge base. For example, given the question "Until when will contract X provide compensation?" from the user, the system interprets this natural-language statement and returns "insurance period" for "contract X" as a reply.

Of course, there are many and varied products other than insurance policies in the world, and with this in mind, NTT Media Intelligence Laboratories has developed technology that enables robust information extraction even for user utterances related to unknown products through the use of general
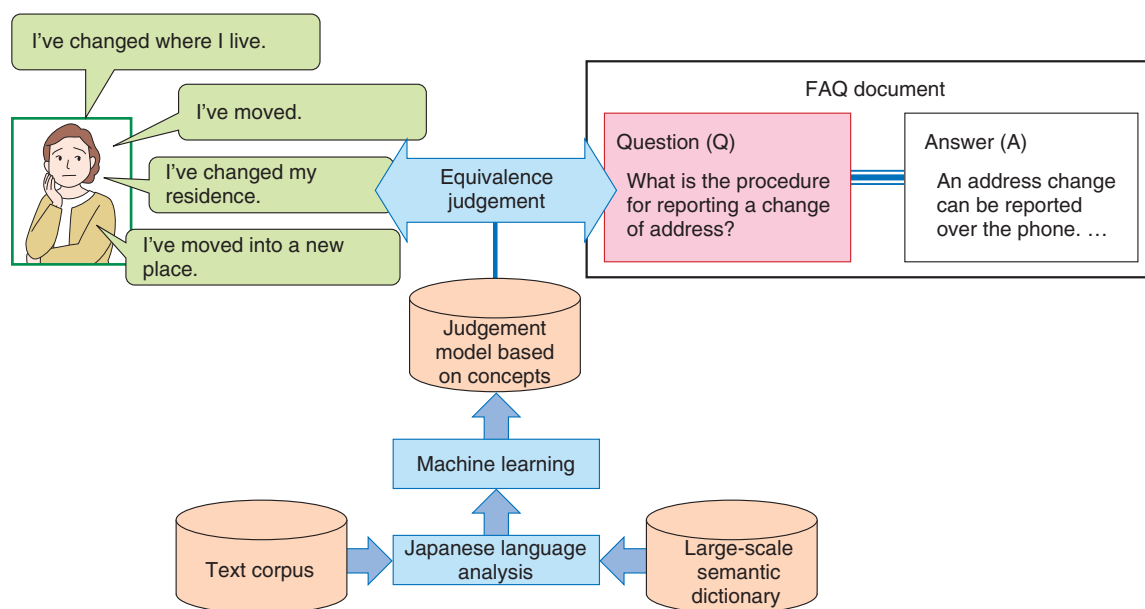
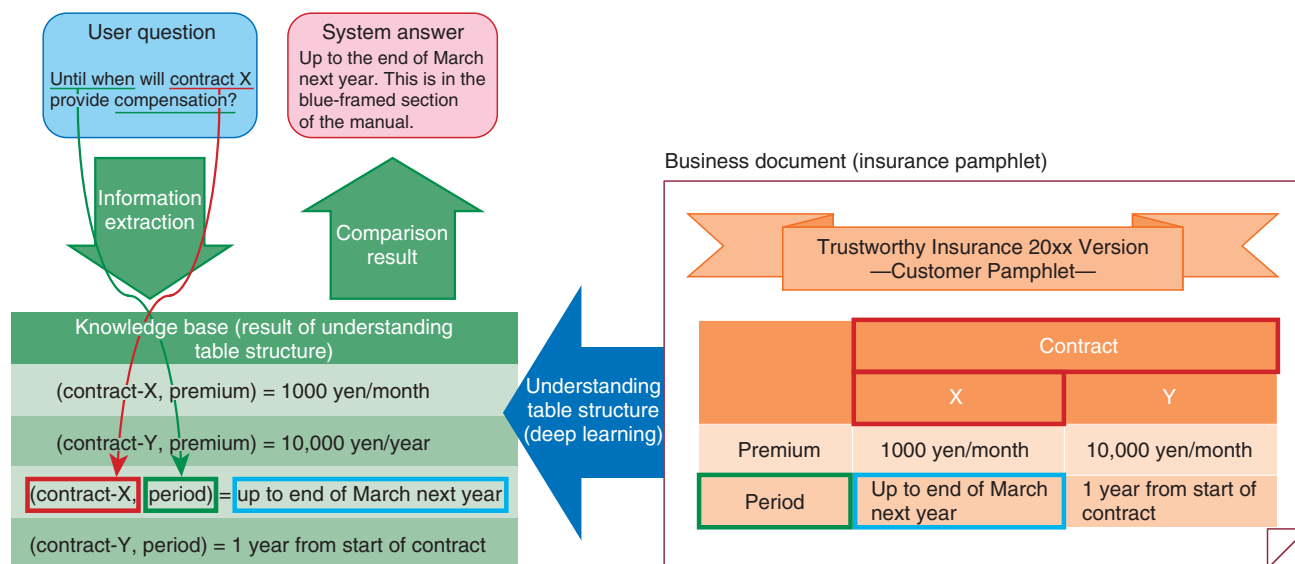Fig. 2.   Natural-language equivalence judgement technology.

Fig. 3.   Provision of pinpoint information based on business knowledge.

language features. This is achieved by automatically inferring the type of question posed. In this example, the technology would infer that the user is asking about "period" when speaking the words "until when" and would use this inference as a general language feature [2].

Next, in the case of external services, the system automatically performs processing to connect to an external API and satisfy the user's request. For example, in response to a user who simply says, "I would like to receive an insurance consultation on Saturday this week," it would be possible to automatically execute an appointment procedure by understanding the intention of specific keywords in the user's

3

utterance. Here, the system would infer the desired date to be "July 24" (the actual date indicated by "Saturday this week"), the user's objective to be "insurance consultation," and the desired processing to be "appointment."

4. Table classification technology for converting knowledge from human use to AI use

Discovering intelligent information within documents and putting it into an AI-readable form is an important role of natural language processing technology. Knowledge acquired from documents is useful for improving information retrieval and Q&A services, which search for information from a massive amount of data, and for supporting human activities in contact centers and elsewhere.

Documents are filled with information in various formats, and tabular data, in particular, group together important information. However, AI has not been able to make maximum use of such data as knowledge. Tables that we usually unconsciously read fall into a number of types, and we cannot acquire knowledge correctly from the table unless we can distinguish between the tables (Fig. 3). Research to identify different types of tables has been carried out since the first half of the 2000s, but the ability to accurately and exhaustively acquire knowledge from documents has not reached human level.

What kind of ability is necessary to identify table types that AI has not been able to achieve? It is the ability to understand both the meaning of the text written in each cell in the table and the semantically characteristic blocks of the cells. To address these two requirements, NTT Media Intelligence Laboratories proposed hybrid deep learning technology called TabNet that combines a neural network specialized in understanding sequence data as in text and a neural network specialized in discovering features from matrix data as in an image [3]. This technology makes it possible to read tables as humans do and to acquire knowledge from documents with high accuracy.

5. Dialogue rule database enabling replies in casual conversations with humans

Is it enough for agents to play a role only in replying to inquiries or supporting information desks? It has been reported that we humans have a tendency to chat with inanimate objects that have the look and feel of a human being [4]. To achieve agents that are more human in nature, more research is being carried out to find the means of achieving natural replies in casual conversations with humans.

To deal with casual conversation, an agent requires dialogue knowledge for responding appropriately to a wide range of user utterances. Although techniques exist for automatically creating dialogue knowledge from large volumes of text data, manually describing dialogue rules is an effective technique for achieving more accurate and effective replies. NTT Media Intelligence Laboratories has constructed a large-scale dialogue rule database. This database deals with a wide variety of expressions in user utterances by taking into account synonyms like *liquor* and *alcohol* and *drink* and *enjoy* and expressions like *what is* and *I want to know* that express the same intent. Furthermore, by repeatedly improving rules through dialogue experiments, we have succeeded in constructing a database comprising approximately 300,000 rules, which enables natural dialogues with humans with high accuracy.

In actual services, there are ways of creating more attractive agents with high-quality replies. For example, agent characters can be configured as needed, and a separate set of dialogue rules customized to the main task of a service can be prepared and used together with the dialogue rule database.

6. Future development

The technology presented here is gradually being implemented in the COTOHA^{TM} communication engine introduced in the feature article in this issue entitled, "COTOHA^{TM}: Artificial Intelligence that Creates the Future by Actualizing Natural Japanese Conversation" [5]. The plan going forward is to expand the use of this technology in NTT Group agent services.

NTT is committed to researching and developing technologies for enhancing functions that enable AI to learn on its own and for achieving more intelligent and personalized agents.

References

[1] Y. Matsuo, R. Higashinaka, H. Asano, and T. Makino, "Natural Language Processing Supporting Artificial Intelligence," NTT Technical Review, Vol. 14, No. 5, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201605fa2.html
[2] K. Sadamitsu, Y. Homma, R. Higashinaka, and Y. Matsuo, "Zero-shot Learning of Natural Language Understanding for Information Provision in Open Domains," The 30th Annual Conference of the Japanese Society for Artificial Intelligence, 2016 (in Japanese).
[3] K. Nishida, K. Sadamitsu, R. Higashinaka, and Y. Matsuo, "Understanding the Semantic Structures of Tables with a Hybrid Deep Neural

Network Architecture," Proc. of AAAI 2017 (Thirty-First AAAI Conference on Artificial Intelligence 2017), pp. 168–174, San Francisco, CA, USA, Feb. 2017.

[4] B. Reeves and C. Nass, "The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places," CSLI Publications and Cambridge University Press, 1996.

[5] N. Furutaka and M. Hamada, "COTOHA™: Artificial Intelligence that Creates the Future by Actualizing Natural Japanese Conversation," NTT Technical Review, Vol. 15, No. 8, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201708fa4.html

Hisako Asano
Senior Research Engineer, Supervisor, Audio, Speech, and Language Media Laboratory, NTT Media Intelligence Laboratories.
She received a B.E. in information engineering from Yokohama National University, Kanagawa, in 1991. She joined NTT Information Processing Laboratories in 1991. Her research interests include natural language processing, especially morphological analysis, and information extraction. She is a member of the Information Processing Society of Japan (IPSJ) and the Association for Natural Language Processing (NLP).

Kyosuke Nishida
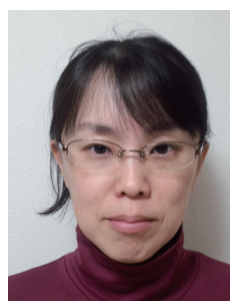Senior Research Engineer, NTT Media Intelligence Laboratories.
He received his B.E., MIS, and Ph.D. in information science and technology from Hokkaido University in 2004, 2006, and 2008. He joined NTT in 2009. He was a chief engineer at NTT Resonant from 2013 to 2015. Since 2015, he has been a senior research engineer at NTT Media Intelligence Laboratories. His current interests include natural language processing, web mining, and ubiquitous computing. He is a member of IPSJ, IEICE, and the Database Society of Japan.

Katsuji Bessho
Senior Research Engineer, Advanced Speech and Language Technology Group, Audio, Speech, and Language Media Laboratory, NTT Media Intelligence Laboratories.
He received a B.S. and M.S. in mathematics from Osaka University in 1992 and 1994 and a Ph.D. in engineering from Niigata University in 2009. He joined NTT Information and Communication Systems Laboratories in 1994. He is currently engaged in the research and development of natural language processing. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), IPSJ, and NLP.

Kuniko Saito
Senior Research Engineer, NTT Media Intelligence Laboratories.
She received a B.E. and M.E. in chemistry from the University of Tokyo in 1996 and 1998. She joined NTT Information and Communication Systems Laboratories in 1998. Her research focuses on natural language processing and spoken dialogue systems. She is a member of NLP and IPSJ.
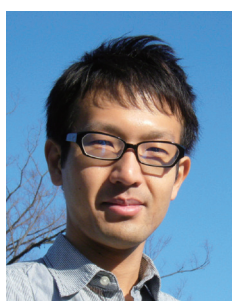
Kugatsu Sadamitsu
Researcher, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.
He received a B.E., M.E., and Ph.D. in engineering from the University of Tsukuba, Ibaraki, in 2004, 2006, and 2009. He joined NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories) in 2009. His current research interests include natural language processing and machine learning. He is a member of NLP.