

Natural Language Processing Supporting Artificial Intelligence

Yoshihiro Matsuo, Ryuichiro Higashinaka, Hisako Asano, and Toshiro Makino

Abstract

Natural language processing technology is considered to be essential to achieving artificial intelligence (AI) that can substitute for humans in certain roles. Most robots and AI systems that appear in fiction are equipped with a natural language interface, and it is expected that *clever* AI will learn on its own from encyclopedias, news reports, and other textual sources. Exactly what kinds of AI functions will be achieved through natural language processing technology? This article introduces the role that natural language processing technology can be expected to play in achieving AI.

Keywords: artificial intelligence, natural language processing, knowledge processing

1. Introduction

In artificial intelligence (AI), a technical field that is difficult to define, natural language processing technology appears to be a major element. This is clearly demonstrated by the use of natural language conversations to measure computer intelligence in the Turing test, which is a well-known technique for determining whether a certain machine is intelligent. In measuring the intelligence of something in which intelligence or intellectual level cannot be directly observed, it is natural to adopt an approach in which intelligence is assessed based on responses to queries, since linguistic expression is surely a typical means of exhibiting intelligence.

What kind of relationship does AI then have with natural language processing? A schematic representation of the relationship between AI and natural language processing is shown in **Fig. 1**. Agent-AI, which is introduced in the feature article “Artificial Intelligence Research Activities and Directions in the NTT Group” [1] in this issue, is a form of AI that possesses abundant knowledge and a capacity for making judgments so that it can replace humans in certain tasks and assist humans in their daily lives. Scenarios in which Agent-AI would have a connection to natural language can be broadly divided into reading of

documents and interaction with the user. Thus, Agent-AI achieved with natural language processing would read documents extensively and accumulate knowledge, recognize user utterances, and make an appropriate response based on that accumulated knowledge. We consider that the Turing test could be a means of measuring in some way whether Agent-AI is producing effective responses.

In this article, we introduce how natural language processing technology has so far been used in AI industrial applications and the role it is expected to play in the future.

2. Text-to-knowledge technology for understanding the meaning of text

If we assume that natural language expressions demonstrate human intelligence, documents written in natural language should then be filled with intelligent information. The primary role of natural language processing technology is to find information that is needed and beneficial from this text filled with intelligent information. Yet, text itself is no more than a collection of character strings, so evaluating data in this form to find what is needed and beneficial cannot be done. However, if text can be arranged in a machine-readable form through natural language

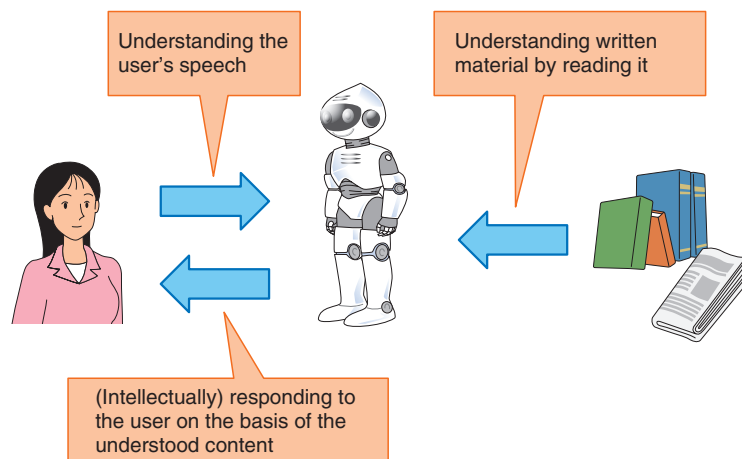


Fig. 1. Relationship between AI and natural language processing technology.

processing, information processing with a computational ability surpassing that of humans will become possible. To this end, we can consider applications of natural language processing to an information-searching technique that looks for desired information from within a huge amount of information, or to a text mining technique that consolidates, arranges, and presents information completely unreadable to humans in its original form in a relatively short period of time.

An example of applying the former technique is a question answering system [2]. A typical approach that most people would take to learn about something would be to search out a related section or entry from an encyclopedia, a textbook, or the web and retrieve an answer from that source. This is the so-called information-searching procedure, but if the sequence of steps in this procedure could be automated, it would be of great assistance in human activities.

Next, examples of applying the latter technique would be the mining of a contact-center log to analyze customers' voices and sentiment analysis of users' comments on the Internet [3]. While it may be possible to uncover valuable information straight from the voices and comments of consumers, differentiating what is good and bad from such a huge amount of information is not an easy task. Furthermore, such analysis would take time, so if this information was to be applied to some sort of decision-making, it could hardly be used for making rapid decisions. Thus, the ability to analyze a large volume of documents in a short period of time would be beneficial to corporations and organizations that need to

conduct such tasks.

What kind of language processing technology is therefore needed to meet the above requirements? That would be technology capable of extracting information from documents written in natural language and organizing that information as knowledge. We have undertaken the research and development of *rich indexing technology* for extracting semantic information from text and have been successful in extracting named entity expressions and opinion information and in identifying the meaning of named entities [3]. The automatic extraction of such information from text will enable the detection of answers in retrieved documents and the tabulation of information from large volumes of text.

3. Utterance understanding technology for a natural language ICT interface

Another function expected of natural language processing technology is a computer interface based on natural language.

Exchanges between humans and computers are normally carried out according to a specialized communication format consisting of menus, icons, and commands manipulated by a keyboard, mouse, or other device. This type of interface has been designed with efficient computer operation in mind, and a user who becomes proficient in its use can indeed perform efficient and effective operations. However, communication with a computer can be quite difficult for those uncomfortable with such an interface, and this problem is thought to be one of the factors giving rise

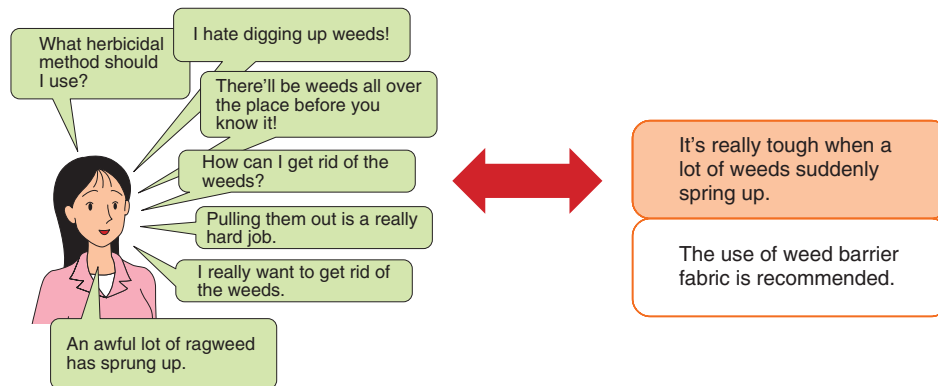


Fig. 2. Determining sameness between user questions and FAQ on paper.

to the so-called digital divide. In contrast, natural language can be viewed as an extremely basic human communication medium. Of course, a certain level of proficiency is also necessary to communicate with natural language, but this is a basic ability that most people come to possess through daily life and education.

The development of an information and communication technology (ICT) interface using natural language will create value in two ways. The first is that people will be able to use ICT and enjoy its benefits without having to develop special skills, and the second is that a standard means of communication between human beings and computers in the form of natural language will enable people and AI to coexist in society.

One example of the former is a voice agent. The operation of a smartphone is still rather complicated, and learning how to use its functions well is not that easy. Furthermore, while much knowledge can be obtained by querying a search engine, a certain amount of information technology literacy is needed in order to use a search engine effectively. Consequently, if questions could be posed to AI by voice using natural language, a further expansion of ICT users and usage scenarios could be expected.

An example of the latter type of value creation is a customer-service support terminal for use at a contact center or retail store. At contact centers, customers talk with operators by telephone. We can envision how a computer could listen in on these conversations and provide just the right information at just the right time. This kind of capability could make solving customers' problems more efficient. Similarly, at a retail store, a small robot could be positioned next to a

salesperson at a cash register or information counter. When the salesperson was discussing products with customers, the robot could promptly offer relevant brochures or documents when needed, thereby serving as a very capable assistant.

Common to the above examples is the need for utterance recognition technology. Several ways of recognizing an utterance can be considered. One would be to translate natural language into a computer language such as Structured Query Language (SQL), which is commonly used for database searching [4]. If a natural language query could be translated and entered into the SQL search conditions field, it would then be possible to search databases by using natural language. Another way of recognizing an utterance would be to judge the equivalence of an utterance in terms of its intention (**Fig. 2**). For example, let's consider a frequently asked questions (FAQ) search. An FAQ document consists of frequently asked questions, where knowledge is represented by pairs of questions and answers (Q&A). In this search, a user poses some kind of question (query). However, the query rarely agrees with a question at the character string level, but if it can be determined that the query has the same intention as a certain question, the correct answer can be provided.

4. Casual conversation technology for achieving human-like dialogue

The technologies described so far serve to make human life much more efficient, but they are not sufficient in themselves to achieve Agent-AI. If a human being and Agent-AI are to become true partners, they must be able to trust and understand each other and

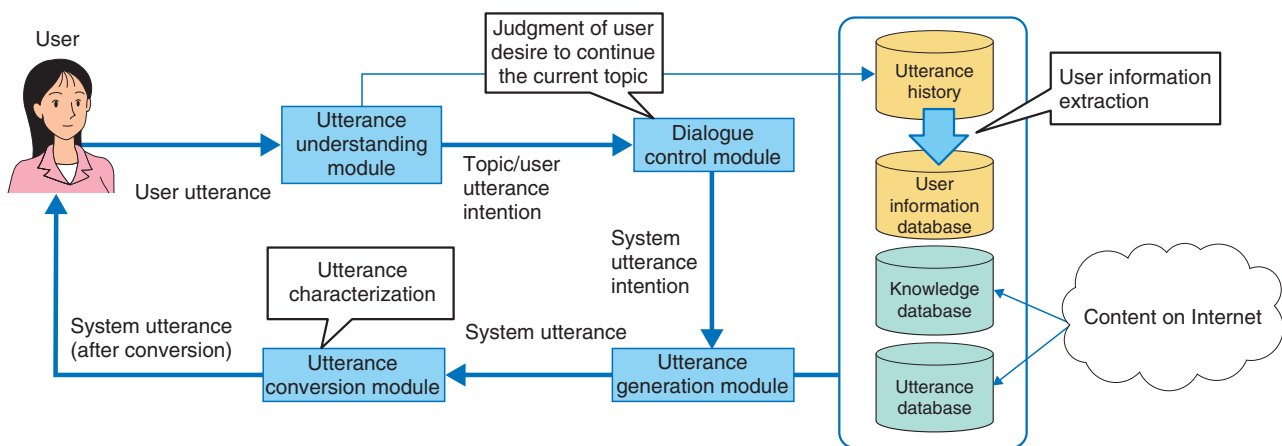


Fig. 3. Configuration of casual conversation technology.

even share emotions. When two people meet each other for the first time, they would hardly start talking about work from the start. It is only after they have become more familiar with each other that they would begin to talk freely. Furthermore, in times of need, partners should help, comfort, and emotionally support each other. Agent-AI must also be equipped with such functions. As is known from the Media Equation [5] theory, a human being interacting with a computer tends to behave as if the computer were human. Thus, the more intelligent a computer becomes, the more human in character it should be. In human society, a *casual conversation* is often the occasion for two people to get to know each other or to become emotionally closer.

With this in mind, NTT Media Intelligence Laboratories has been researching casual conversation technology [6]. This technology will enable a user to converse with a computer in everyday language on any topic. Casual conversation is, of course, easy for human beings but extremely difficult for computers. This is because current technology cannot easily handle the many and varied topics that a user may want to talk about, and it cannot develop a deep understanding in the way that a human can of complex subjects raised by the user. In the AI field, the former problem is known as the frame problem and the latter as the symbol grounding problem. Against this background, we have undertaken the development of casual conversation technology and have succeeded in achieving a certain level of casual conversation with a computer by combining large-scale text data and language processing technology. This technology is being used in casual-conversation

application programming interfaces for developer use and in commercially available communication toys.

A system based on casual conversation technology structures large volumes of content on the Internet (blogs, microblogs) using language processing technology and constructs an utterance database and knowledge database (Figs. 3 and 4). When a user utterance is input, the system recognizes the topic and intention of that utterance based on the context. It then produces a response in line with the user's intention by searching for an appropriate system utterance from an utterance database or by generating a system utterance using a knowledge database. If the user utterance is a question, the system uses question answering technology to obtain a word or phrase from the Internet as a response. This mechanism enables the system to produce responses on a variety of topics that may appear in casual conversation.

However, it is not sufficient in itself to achieve human-like behavior. Achieving Agent-AI that can form an attachment as a partner and carry on extended conversations will require mechanisms for remembering things about the conversation partner, expressing individuality, and understanding emotions. For this reason, we are equipping Agent-AI with techniques for user information extraction, utterance characterization, and judgment of user desire to continue the current topic. These techniques will enable the construction of user information databases based on user utterances, conversion of system utterances to ones expressing a personality, and a change of topic if the system detects that the user appears to have lost interest in the current topic.

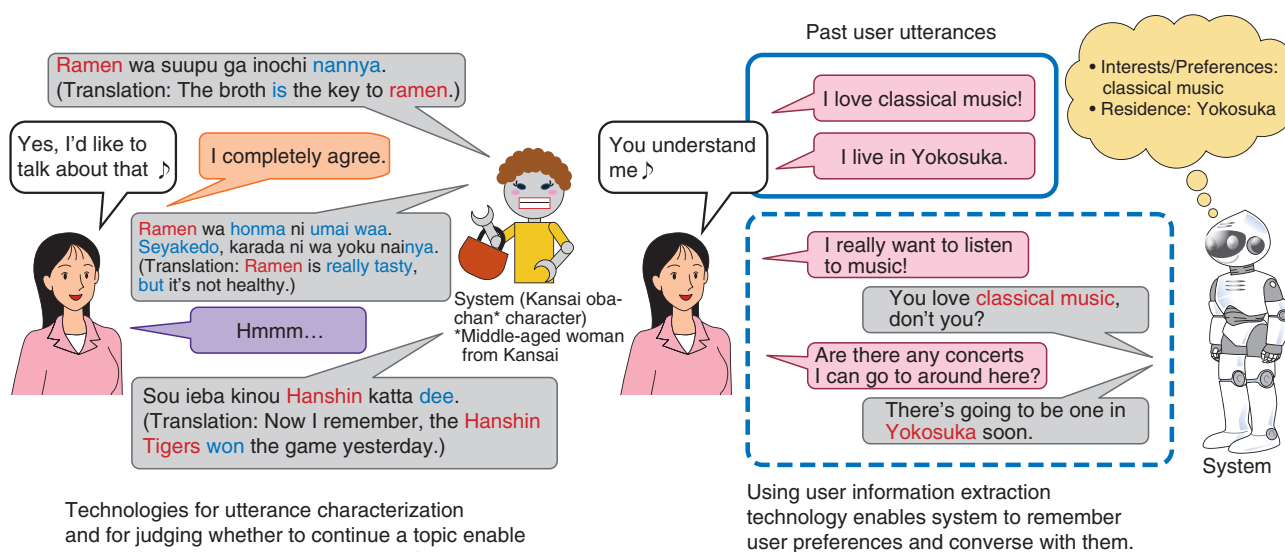


Fig. 4. Example of conversation enabled by casual conversation technology.

5. Future outlook

Finally, we would like to touch upon the Todai Robot Project (Can a Robot Get into the University of Tokyo?), a major AI project that has been attracting attention. This project is a grand challenge led by Japan's National Institute of Informatics to develop AI that can pass the entrance exam of the University of Tokyo. Specifically, it aims to achieve a high score in the National Center Test for University Admissions by 2016 and to pass the University of Tokyo exam by 2021. Passing the latter will require the development of AI that can take and pass tests in a variety of academic disciplines.

NTT has been participating in this project since 2014 and has been placed in charge of English as a target discipline [7]. In addition to learning vocabulary and grammar, English presents a variety of problems, including the reading and comprehension of long sentences and conversational text, the matching up of text with tables and figures, and the use of diverse language processing skills. Solving problems that human beings actually solve is what achieving Agent-AI is all about. The program that we have constructed for solving these English problems can presently achieve a level of performance equivalent to that of an average human examinee. Our aim, however, is to achieve Agent-AI that is useful in problem

solving and supportive to people through the experiences gained in this project and the results of researching and developing casual conversation technology.

References

- [1] T. Yamada, S. Takahashi, F. Naya, T. Ikebe, and S. Furukawa, "Artificial Intelligence Research Activities and Directions in the NTT Group," NTT Technical Review, Vol. 14, No. 5, 2016. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201605fa1.html>
- [2] R. Higashinaka, K. Sadamitsu, K. Saito, and N. Kobayashi, "Question Answering Technology for Pinpointing Answers to a Wide Range of Questions," NTT Technical Review, Vol. 11, No. 7, 2013. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201307fa4.html>
- [3] Special Feature: "Natural Language Processing Technologies for Portal Services," NTT Technical Review, Vol. 6, No. 9, 2008. <https://www.ntt-review.jp/archive/2008/200809.html>
- [4] K. Sadamitsu, J. Shimamura, G. Irie, S. Tarashima, T. Yoshida, R. Higashinaka, H. Nishikawa, N. Miyazaki, Y. Ijima, and Y. Nakamura, "Media Processing Technology for Achieving Hospitality in Information Search," NTT Technical Review, Vol. 13, No. 4, 2015. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa5.html>
- [5] 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.
- [6] 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 2014 (the 25th International Conference on Computational Linguistics), pp. 928–939, Dublin, Ireland, Aug. 2014.

- [7] R. Higashinaka, H. Sugiyama, H. Isozaki, G. Kikui, K. Dohsaka, H. Taira, and Y. Minami, "Taking the English Exam for the 'Can a Robot Get into the University of Tokyo?' Project," NTT Technical Review,

Vol. 13, No. 7, 2015.

<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507ra2.html>



Yoshihiro Matsuo

Group Leader, Senior Research Engineer, Supervisor, Audio, Speech, and Language Media Laboratory, NTT Media Intelligence Laboratories.

He received a B.S. and M.S. in physics from Osaka University in 1988 and 1990. He joined NTT Communications and Information Processing Laboratories in 1990. He moved to NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories) in 2001. His research interests include multimedia indexing, information extraction, and opinion analysis. He is a member of the Information Processing Society of Japan (IPSI) and the Association for Natural Language Processing (NLP).



Hisako Asano

Senior Research Engineer, 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 IPSJ and NLP.



Ryuichiro Higashinaka

Senior Research Scientist, Audio, Speech, and Language Media Laboratory, NTT Media Intelligence Laboratories.

He received a B.A. in environmental information, a Master of Media and Governance, and a Ph.D. from Keio University, Kanagawa, in 1999, 2001, and 2008. He joined NTT in 2001. His research interests include building question answering systems and spoken dialogue systems. From Nov. 2004 to Mar. 2006, he was a visiting researcher at the University of Sheffield in the UK. He received the Maejima Hisoka Award from the Tsushinbunka Association in 2014. He is a member of the Institute of Electronics, Information and Communication Engineers, the Japanese Society for Artificial Intelligence, IPSJ, and NLP.



Toshiro Makino

Senior Research Engineer, Supervisor, Audio, Speech, and Language Media Laboratory, NTT Media Intelligence Laboratories.

He received his B.E., M.E., and Dr. Eng. in electrical engineering from the University of Tokyo, in 1987, 1989, and 1992. He joined NTT Network Information Systems Laboratories in 1992. He is currently a senior research engineer, supervisor at NTT Media Intelligence Laboratories and is engaged in the research and development of natural language processing.