Researching AI Technologies to Support and Substitute Human Activities

Hidenori Tanaka, Masaki Kitahara, and Yoshinori Kusachi

Abstract

The coming of digital transformation has been gaining interest, and its development seems to be accelerating during the COVID-19 pandemic. Competition in artificial intelligence (AI) technologies is also intensifying, and platformers are accumulating massive amounts of data as learning data. Amid these trends, NTT Media Intelligence Laboratories is promoting the research and development (R&D) of AI technologies to support and substitute human activities by taking advantage of the technologies and expertise that it has cultivated. This article introduces this R&D initiative.

Keywords: support of human activities, substitute for human activities, AI technology

1. Introduction

NTT Media Intelligence Laboratories has been involved in the research and development (R&D) of technologies for processing various forms of media, such as voice, audio, language, images, and video, and has put several of these technologies into practice. More recently, it has made a number of contributions to business such as operator support at contact centers [1], provision of artificial intelligence (AI) agents [2, 3], sound-collection technology for emergency call systems [4], and video-compression devices for 4K/8K broadcasting [5].

However, today's market environment is going through major changes and competition is intensifying. Every industry is experiencing digital transformation (DX), which is revolutionizing business models through digitalization, and the development of DX is expected to accelerate to deal with the COVID-19 pandemic. The advent of deep learning has ignited a third AI boom, and the basic algorithms of AI technologies are now available to anyone. Moreover, learning data that contributes to the improved performance of AI systems is now being collected on a large

scale by platformers, such as Google, Apple, Facebook, and Amazon, so a world in which AI performance increases daily is becoming a reality.

In response to this external environment, the NTT Group is promoting initiatives such as B2B2X (business-to-business-to-X) and DX toward the creation of a *smart world* as part of its Medium-Term Management Strategy. It is also promoting the concept of the Innovative Optical and Wireless Network (IOWN) to build a smart world through innovative technologies.

Against this background, NTT Media Intelligence Laboratories is leveraging the technologies and expertise cultivated through R&D activities in media processing to pursue the R&D of AI technologies for supporting and substituting human activities—the source of all value—and the R&D of Digital Twin Computing (DTC) to create new value from a medium- to long-term perspective [6]. In this article, we introduce the R&D of AI technologies toward an application domain in which AI supports and substitutes human activities.

2. Overview of AI technologies to support and substitute human activities

Several possible scenarios can be considered when talking about an application domain in which AI supports and substitutes human activities. For example, in terms of increasing efficiency, AI can not only improve the productivity of operators at contact centers and AI agents but also improve business processes and productivity in offices and improve the quality of life as new value. We expect the COVID-19 pandemic to drive the penetration of teleworking and online meetings and transform the way in which they should be carried out.

However, applying current AI technologies to such scenarios is difficult. For example, when attempting to support and substitute human activities on a more personal level based on a person's environment, there would be a need to obtain data on that person and environment, but there are situations in which a large amount of data cannot be obtained. Under such conditions, it would be difficult to achieve the full potential of AI. Taking speech-recognition technology as an example, differences in its application would arise in telephone calls, meetings, etc. depending on whether only speech-to-text processing is sufficient or whether processing as far as speaker identification is necessary. In the case of online meetings, there may even be more differences in expected performance and other requirements.

NTT Media Intelligence Laboratories has begun to address these issues by focusing its efforts on achieving efficient learning with a relatively small amount of data, producing new AI technologies, and achieving breakthroughs in the performance of current technologies.

3. Current initiatives in AI technologies to support and substitute human activities

The Feature Articles in this issue introduce a group of technologies that we are currently developing. First, in the article "Media-processing Technologies for Ultimate Private Sound Space" [7], the focus is on sound as a major element of technologies that can produce new AI technologies, such as for understanding surrounding conditions from sound, listening to only the person you want to hear, and eliminating sound you do not want to hear. Our aim with these technologies is to achieve an ultimate private space that can be expected in applications such as teleworking.

Next is "Saxe: Text-to-Speech Synthesis Engine Applicable to Diverse Use Cases" [8]. Regarding speech-synthesis technology for generating speech for virtual announcers, AI agents, etc., this article introduces technology for high-accuracy reading of heteronyms according to context and deep-neural-network speech-synthesis technology for reproducing a variety of speaker characteristics at low cost. These technologies are aimed at achieving breakthroughs in the performance of current technologies and efficient learning with a relatively small amount of data.

In "Speech Recognition Technologies for Creating Knowledge Resources from Communications" [9], we envision speech recognition in scenarios such as meetings and face-to-face customer service. In addition to techniques for improving current speech-to-text technology, this article introduces technologies for producing new AI technologies such as for extracting the gender, emotions, and other characteristics of the speaker from sound.

In "Knowledge and Language-processing Technology that Supports and Substitutes Customer-contact Work" [10], we introduce technologies for achieving breakthroughs in the performance of current technologies and producing new AI technologies. These technologies include document summarizing with a certain specified length according to the application scenario and response analysis for improving the productivity of an operator involved in inside (remote) sales.

Finally, to achieve a fourth-dimensional (4D) digital platform [11] that integrates various types of sensing data in real time and enables a variety of future predictions, the article "Spatial-information Processing Technology for Establishing a 4D Digital Platform" [12] introduces technology for digitalizing real-world space and point-cloud-coding technology for efficiently compressing 3D data that include temporal changes. These are technologies for efficient learning with a small amount of data and achieving breakthroughs in the performance of current technologies.

4. Anticipated use cases

We now introduce use cases of AI technologies for supporting and substituting human activities that take into account the recent COVID-19 pandemic (**Fig. 1**). In creating personal space, the aim is to construct a pseudo personal space for teleworking and a space that ensures privacy within one's home. In online
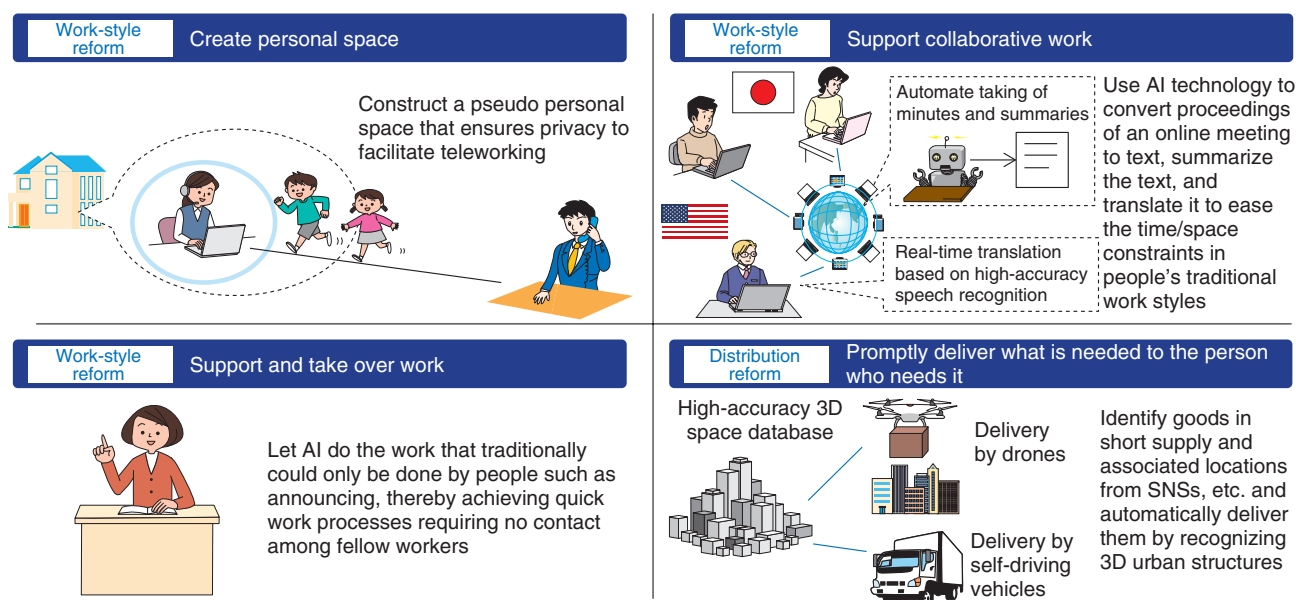
Fig. 1.   Use cases of AI technologies to support and substitute human activities.

meetings, AI technologies will be used to support innovative collaborative work by converting the proceedings of a meeting to text, summarizing and translating the meeting, and easing the time/space constraints in people's traditional work styles. Additionally, by letting AI do the work traditionally done by people, such as announcing, we anticipate quick work processes that require no contact among fellow workers. Finally, from the viewpoint of distribution reform, we envision the possibility of identifying goods in short supply and associated locations from social networking services (SNSs), etc. and automatically delivering those goods by recognizing 3D urban structures.

5.   Future outlook

The surrounding environment is changing rapidly. To provide various means of supporting and substituting human activities, technology must deal accordingly with these changes. To this end, we plan to promote R&D with a flexible frame of mind while understanding both the macro and micro aspects of these changes.

References

[1]   ForeSight Voice Mining, https://www.ntt-tx.com/products/foresight_vm/
[2]   Website of NTT Communications' AI services, https://www.ntt.com/en/services/ai.html
[3]   Website of NTT DOCOMO's my daiz AI agent service (in Japanese), https://www.nttdocomo.co.jp/service/mydaiz/
[4]   Press release issued by NTT on Feb. 19, 2018 (in Japanese). https://www.ntt.co.jp/news2018/1802/180219c.html
[5]   Press release issued by NTT, "NTT Develops the Worlds' Best Performance 8K HEVC Real-time Encoder through Dedicated LSI Use," Feb. 15, 2016.
https://www.ntt.co.jp/news2016/1602e/160215a.html
[6]   White paper on Digital Twin Computing, https://www.ntt.co.jp/svlab/e/DTC/DTC_Whitepaper_en_2_0_0.pdf
[7]   M. Fukui, S. Saito, and K. Kobayashi, "Media-processing Technologies for Ultimate Private Sound Space," NTT Technical Review, Vol. 18, No. 12, pp. 43–47, 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202012fa6.html
[8]   Y. Ijima, N. Kobayashi, H. Yabushita, and T. Nakamura, "Saxe: Text-to-Speech Synthesis Engine Applicable to Diverse Use Cases," NTT Technical Review, Vol. 18, No. 12, pp. 48–52, 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202012fa7.html
[9]   Y. Nakazawa, T. Mori, Y. Yamaguchi, Y. Shinohara, and N. Miyazaki, "Speech Recognition Technologies for Creating Knowledge Resources from Communications," NTT Technical Review, Vol. 18, No. 12, pp. 53–58, 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202012fa8.html
[10]   K. Nishida, K. Saito, T. Amakasu, K. Iso, and S. Nishioka, "Knowledge and Language-processing Technology that Supports and Substitutes Customer-contact Work," NTT Technical Review, Vol. 18, No. 12, pp. 59–63, 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202012fa9.html
[11]   Press release issued by NTT, "4D Digital Platform™"; Integrates Various Sensing Data in Real Time and Enables Future Predictions," Mar. 26, 2020.
https://www.ntt.co.jp/news2020/2003e/200326c.html
[12]   Y. Yao, K. Kurata, N. Ito, S. Ando, J. Shimamura, M. Watanabe, R.

Tanida, and H. Kimata, "Spatial-information Processing Technology for Establishing a 4D Digital Platform," NTT Technical Review, Vol. 18, No. 12, pp. 64–70, 2020.

Hidenori Tanaka
Senior Research Engineer/Producer, Planning Section, NTT Media Intelligence Laboratories.
He received a B.E., M.E., and Ph.D. in science and technology from Keio University, Kanaga-wa, in 2004, 2006, and 2011. After joining NTT in 2006, he engaged in research on computer vision and pattern recognition. He has been involved in technology marketing and R&D management.

Yoshinori Kusachi
Chief Producer, Planning Section, NTT Media Intelligence Laboratories.
He received a B.E. in computer science from Kyoto University in 1995 and an M.E. in information science and Ph.D. in information engineering from Nara Institute of Science and Technology in 1997 and 2007. He joined NTT in 1997 and engaged in research and practical application development in the fields of image processing, computer vision, and pattern recognition. He has also been involved in R&D management.

Masaki Kitahara
Senior Research Engineer, Scene Analysis Technology Group, Universe Data Handling Laboratory, NTT Media Intelligence Laboratories.
He received a B.E. and M.E. in industrial and management engineering from Waseda University, Tokyo, in 1999 and 2001. After joining NTT in 2001, he engaged in the development of compression and rendering algorithms for 3D video. He transferred to NTT Advanced Technology (NTT-AT) and managed various software development projects. In 2009, he returned to NTT and has been engaged in the development of video encoding and AI technologies and R&D planning. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).