National Institute of Information and Communications Technology

NICT NEWS

2018 No.2 Vol. 468

FEATURE

Promoting Open Innovation in AI Science Field

NICT NEWS 2018 No.2 Vol.468
National Institute of Information and Communications Technology

CONTENTS

FEATURE

TOPICS

Cover photo: "Deep learning program," Upper left photo: "AI research computer"
The cover photo shows an example of a machine learning program for image recognition (i.e., examining an image to determine what it depicts). Since this was developed based on — and inspired by — the neural network of the human brain (especially the cerebral cortex), it is called a neural network. Since the layers that process information are deeply stacked, it is also referred to as deep learning (deep neural network). These have been attracting attention in recent years because their ability to recognize images and sounds has increased to the point where it is comparable to or even higher than that of human beings.
The picture on the upper left shows part of a computer for AI research. Recent breakthroughs in AI have often occurred due to advances in hardware such as high performance and high capacity data storage devices and computing devices that can analyze data at high speed.

INTERVIEW

Helping Society with NICT's Advanced AI Technology

Takaaki SAEKI
Associate Director General, AI Science Research and Development Promotion Center (AIS)

2016 Managing Director, Strategic Program Produce Office, Social Innovation Unit, Open Innovation Promotion Headquarters. Current position in April 2017.

The research and development of ICT is following many different directions, but for the support of people who send and receive information and the analysis of these function, artificial intelligence (AI) is playing an increasingly important role in diverse situations. The AI Science Research and Development Promotion Center (AIS), established on April 1, 2017, is an organization that promotes social implementations by gathering knowledge on AI-related R&D that has been carried out at NICT, such as natural language processing, knowledge processing and studies of information and neural networks. We spoke to the Center's Associate Director General Takaaki SAEKI about the founding of this organization.

A center for open innovation in the AI Science field

— What can you tell us about the circumstances behind the founding of AIS in the 2017 fiscal year?

Saeki: It took place as a response to two major trends. One was related to the strengthening of open innovation functions mentioned in NICT's 4th Medium- to Long-Term Plan that began in April 2016. This plan arranged NICT's research into five major categories: "Sensing fundamentals," "Integrated ICT," "Data utilization and analytics platform," "Cybersecurity," and "Frontier research." AI-related research is treated as part of the Data utilization and analytics platform category, and it was recognized that there was a need for a hub organization promoting open innovation in AI-related field.

The other was the growing recognition that R&D and industrialization are major issues for Japan in the field of AI. In April 2016, the Ministry of Internal Affairs and Communications (MIC), the Ministry of Education, Culture, Sports, Science and Technology (MEXT), and the Ministry of Economy, Trade and Industry (METI) were called upon by the Prime Minister to establish a Strategic Council for AI Technology. This Council discussed future AI technology R&D targets and an industrialization road map, with input provided by the NICT (supervised by the MIC), the Institute of Physical and Chemical Research (RIKEN, supervised by the MEXT), and the National Institute of Advanced Industrial Science and Technology (AIST, supervised by the METI). Then in March 2017, the council's findings were summarized in the document "Artificial Intelligence Technology Strategy." To promote this strategy, the NICT also needed a central organization to coordinate and collaborate with outside parties.

These two initiatives eventually converged with the establishment of AIS.

A two-wheeled system comprising the "Collaborative Research Laboratory of AI Science" and the "Collaboration Promotion Office"

— What sort of specific topics are you working on towards the social implementation of AI technology?

Saeki: I think the departments where research is actually being done differ slightly in their approach to social implementations according to the content of their research. For example, the group working on machine translation was originally involved with a "Global Communication Plan" aimed at the social implementation of a multilingual speech translation system together with the MIC. One of the main short-term goals of this project was achieving some sort of social implementation in time for the Tokyo 2020 Olympic and Paralympic Games. Consequently, this group is already working with many companies and developing various different tools.

However, in departments that are more concerned with basic research, there are still a lot of projects that have no specific activity aimed at social implementations. I think the core of our activities at AIS consists of measuring how much of an impact this research can have on society, and bridging across other research organizations and private companies that are likely to be interested in it.

I think each department needs to be supported in a different way, so we are currently working on an exchange of opinions between various research departments in NICT. Since AIS also includes people from private enterprise, we are also making use of their knowledge while pro-

INTERVIEW

Helping Society with NICT's Advanced AI Technology

posing new ideas of ourselves.

— Tell us about the current organization at AIS for promoting this sort of activity.

Saeki: The current director of the AIS is Dr. Tokuda, who is also NICT president. The AIS also consists of two associate director generals, including myself, and two sections — the Collaborative Research Laboratory of AI Science, and Collaboration Promotion Office.

The staff of the Collaborative Research Laboratory of AI Science include not only full-time researchers based at the headquarters in Koganei, but also researchers at each site/department involved with AI-related R&D within NICT who are working on data acquisition and the utilization of research results and are thereby also working with people who are more deeply involved with open innovation on joint projects with the headquarters.

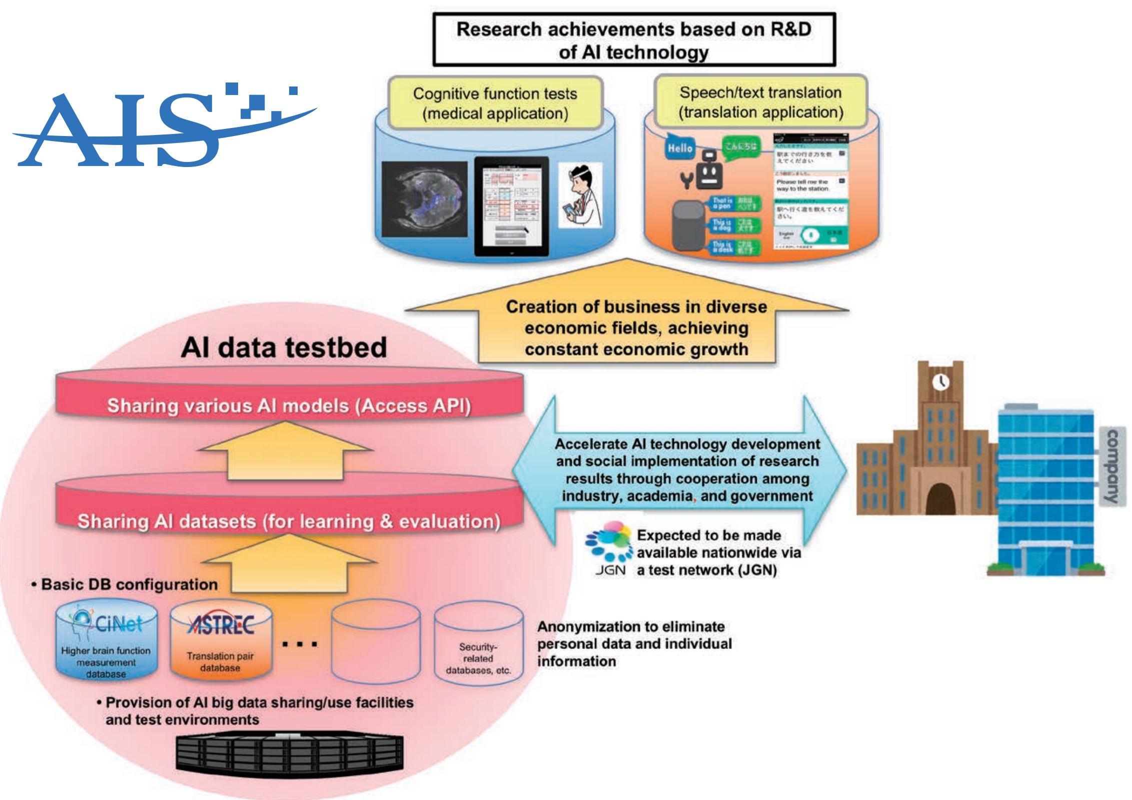
The Collaboration Promotion Office backs up the activities of the collaborative laboratories and engages in cooperative negotiations with departments acting as counterparts at the Strategic Council for AI Technology, especially AIST and RIKEN. For example, in May 2017, we concluded an agreement with AIST on the promotion of collaboration and cooperation in the information and communications field, and we began a joint study on dedicated neural machine translation for patent documents and system configurations that facilitate them. We also held the "Second Joint Symposium on Next-Generation Artificial Intelligence Technology" together with the MIC. The Promotion Office is therefore responsible for carrying out such tasks as concluding memoranda of understanding with external organizations and planning events relating to AI.

Collaborating with external organizations, and sharing for a win-win relationship

— I understand that one of the things being worked on at AIS is "Hon'yaku (translation) Bank," which began operations in September 2017. What is the current state of this project?

Saeki: Particularly in recent AI research and development, it is said that the accumulation of basic data is essential as well as the development of computer resources and algorithms. We therefore drew up plans to create an AI data testbed to obtain a firmer grasp of not only com-

Overview of the AI Science Research and Development Promotion Center (AIS)



At the Second Joint Symposium on Next-Generation Artificial Intelligence Technology hosted by AIS (May 22, 2017)

Closing address by Dr. Tokuda, the NICT President, who is also the Director of the AIS

Osaka University was the venue for this second symposium.

puter resources but also data that can be used for AI.

Another mission of the AIS is to share this sort of AI-related data with others and explore what can be done using this data in collaboration with outside organizations. The Hon'yaku Bank is a part of this activity.

Studies have shown that — particularly in machine translation — the quantity of translation data is closely related to the accuracy of the translation, but purchasing this sort of data can be very expensive. On the other hand, it is speculated that private companies and other organizations are hoarding huge quantities of multilingual sets of translation data such as technical documents that could be salvaged and recycled to improve the accuracy of translation engines. This is the purpose of our Hon'yaku Bank. The accumulation of data is useful for AI research, and could also benefit private companies that have provided data if progress is made in social implementations. We are therefore starting out with the idea that it should be possible to create a win-win relationship.

— It seems that there are going to be a lot of interesting developments like this in the future.

Saeki: It is not long since AIS was inaugurated, and we started with individual specific initiatives on a small scale including the Hon'yaku Bank. But eventually I hope we can instigate a data-centric ecosystem in diverse fields related to AI.

For example, with regard to brain information data handled by the Center for Information and Neural Networks (CiNet), we will publish models after performing modeling operations including anonymization, and we will explore its potential use for social implementation, such as whether it is possible to develop real stories together with private companies. In addition to machine translation and brain information, I hope to extend this movement to include all the AI-related data in the NICT's possession.

In the recent AI boom, various organizations and companies have been actively engaged in R&D, but there are a number of things for which the NICT is entitled to pride itself on as being among the world's best research in specific fields. I hope to raise awareness of this research both internally and externally, and among private enterprises and other research institutions as well. I want us to be able to say that new technologies can be implemented in society in various fields, and that by working together, it's possible to discover new ways forward. The mission of AIS is to realize an organization that plays such a role.

An Approach to R&D Collaborations on AI Technology for Solving Social Issues and Open Innovation

Michiaki IWAZUME
Director of Collaborative Research Laboratory, AI Science Research and Development Promotion Center (AIS)
After completing a doctoral course, Dr. Iwazume successively held the positions of JSPS post-doctoral researcher, assistant professor of Kinki University and research scientist at the RIKEN Brain Science Institute. He joined NICT in 2005, and was appointed to his current position in 2017.
He has been conducting studies of knowledge-based systems and web information processing. Ph.D. (Engineering).

At the Collaborative Research Laboratory of AI Science Research and Development Promotion Center (AIS), we are cooperating with domestic and overseas research institutes on the research and development of AI to solve social issues by exploiting NICT's strengths in machine learning and other aspects of brain communication technology and artificial intelligence (AI), as well as security technology, remote sensing technology, and big data utilization connected with these fields.

Background

Nowadays, barely a day goes by without a news item on some aspect of AI, such as AI speakers, self-driving cars, robots, and chess-playing machines. Against this background of heightened interest in AI, a machine learning technique called deep learning that uses multi-layer neural networks has started to show performance exceeding that of conventional techniques in fields such as image recognition and speech recognition. Although multi-layer neural networks are not a new concept (they were first studied in the 1960s), they were not practical at first because they required huge amounts of computer processing time and because it turned out to be difficult to achieve successful learning in multi-layer systems. However, with recent advances in the processing power and availability of ICT, vast quantities of data have been accumulated and there are now computing environments that are capable of high-speed analysis. Combined with improved learning methods, it is now possible to perform learning in multi-layer neural networks with dozens of layers.

AI technology R&D strategy of NICT

To develop a smart AI system with current machine learning technology, a large amount of high quality data is required. Here, high quality data means a learning data set that includes labels such as correct/incorrect that match the task to be solved (Figure 1). Image data sets for handwritten character recognition and general object recognition have been published and shared on the Internet and are available for use by researchers, but datasets that are useful for

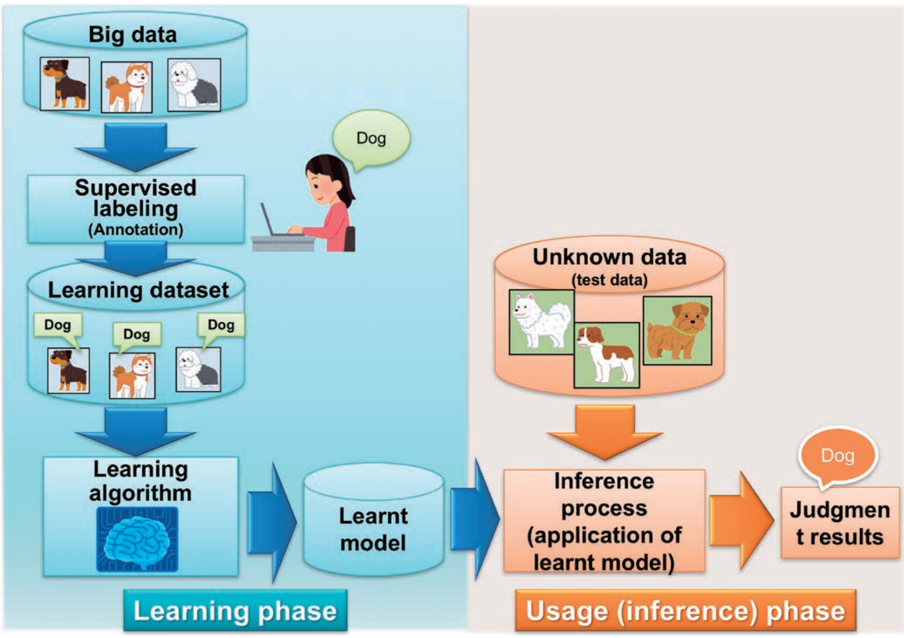


Figure 1 Flow of AI system development

social tasks are often left unpublished or have not been properly integrated. With current technology, maintaining this sort of data set requires that labels are created manually, and is therefore generally very expensive.

Meanwhile, NICT recognized the importance of data long before the arrival of the current AI boom, and built itself a multilingual corpus and a large-scale web archive that can be used for the research and development of advanced AI technologies such as speech recognition, machine translation, and social intelligence analysis. Moreover, in order to clarify the ability of the human brain to learn based on a small amount of experience and with low power consumption, we are working on the creation of advanced brain information databases using technologies such as fMRI and EEG to adapt to the next generation of information communications technology. Furthermore, a diverse variety of data have been measured and collected for the development of ICT-based technologies such as security systems, remote sensing, and wired/wireless communication (Figure 2). This includes valuable data that can only be measured by NICT. In collaboration with each of NICT's research centers, we will apply AI technologies such as machine learning to NICT's dataset assets, and in order to research and develop demonstration systems and accelerate social implementations, we are promoting the implementation and use of an advanced AI data testbed.

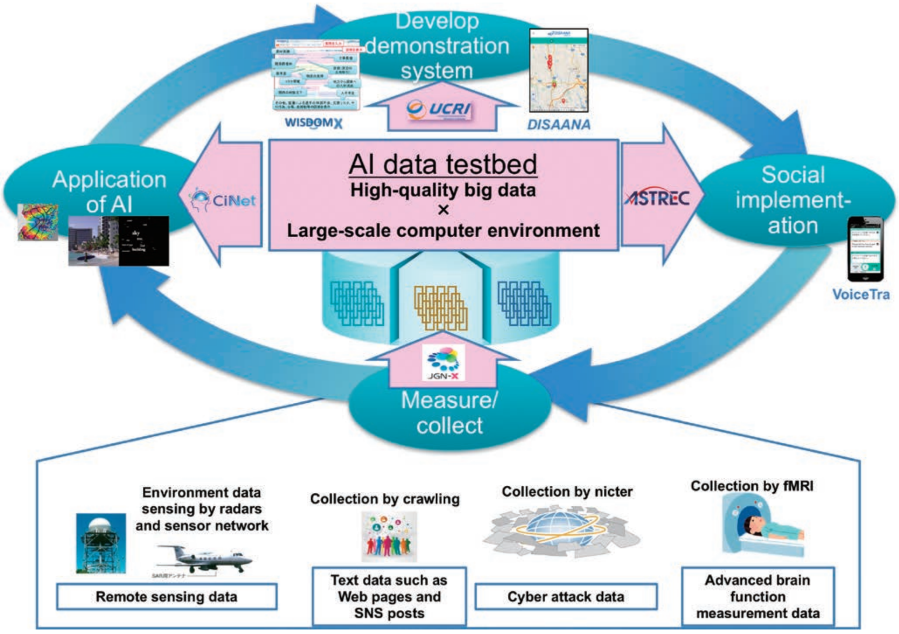


Figure 2　NICT's R&D strategy for AI technology

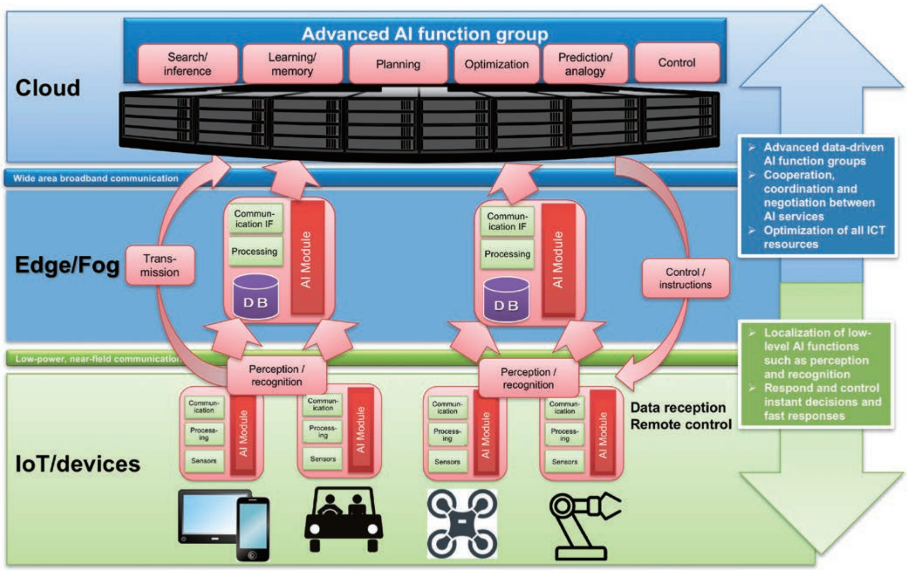


Figure 3　Near Future vision of AI and ICT Network

Future prospects

How will AI technology evolve in the future? Currently, it is at the stage where the accuracy of individual functions such as speech recognition and visual object recognition has dramatically improved. Taking the evolution of living organisms as an analogy, one might say that AI systems have acquired sensory organs like eyes and ears to perceive the outside world, and have started to develop brains with higher functions like recognition and identification. When a huge amount of information is transmitted from IoT devices every moment, it will probably be difficult to process it with a centralized cloud processor due to restrictions such as network communication costs and response times. It is expected that this problem will be addressed by incorporating perception and recognition functions into IoT devices so they can perform primary processing, following which the segmented information can be handed over to edge/fog servers (i.e., servers that are close to network terminals and sensors, or servers that are situated mid-way between terminals and cloud equipment) that perform higher-order processing functions such as inference, optimization and control (Figure 3). If this happens, the whole architecture of ICT will be transformed. I hope to continue exploring how humans and AI can coexist with ICT while tacking concrete social issues.

Toward Automation of Cybersecurity Operations Using Machine-Learning Technologies

Takeshi TAKAHASHI
Research Manager, Cybersecurity Laboratory, Cybersecurity Research Institute/Senior Researcher, Collaborative Research Laboratory, AI Science R&D Promotion Center (AIS)

Received a Ph.D. in telecommunications from Waseda University in 2005. He has worked as a researcher at Tampere University of Technology, and at Roland Berger Ltd. as a Business Consultant. He joined NICT in 2009. His research interests include Internet security and network protocols.

In recent years, the Internet has faced various cyber threats, including ransomware and the hijacking of IoT devices by botnets. To respond to these threats without sufficient numbers of security experts, it is necessary to promote the automation of security operations in addition to training additional human resources. We contribute to the advancement of this automation by making use of artificial intelligence (AI) technology centered on machine learning. For example, we are working on R&D aimed at bringing about various technical revolutions such as classifying and identifying malware (malicious software) and its functions, detecting changes in the trends of malware activity on the Internet, detecting abnormal communication between organizations, and predicting future attacks.

Since AI technology and cybersecurity technology originally developed from different communities, effectively integrating the technologies of both domains is laborious and time-consuming. However, if we can achieve a broad fusion between the two, then it would lead to great advances in the automation of security. The AI Science Research and Development Promotion Center (AIS) was founded within NICT in FY2017 to accelerate this integration process, discover and develop human resources familiar with both fields, and play a key role in the evolution of cybersecurity technology.

Table Examples of data owned by NICT

Category	Examples of datasets
Darknet-related data	Data on attack-related communication to unused IP addresses; pcap files, statistical information, malicious host information, etc.
Live net-related data	Data on communication within NICT; pcap files, flow data, alerts generated by security appliances, etc.
Malware-related data	Malware samples, results of static and dynamic analyses, etc.
Spam-related data	Spam (double bounce) email data, statistical information, etc.
Android application-related data	APK files, market app information, etc.
Blogs and articles	Twitter tweets, security vendor blogs, etc.

Datasets as a source of research competitiveness

In the cybersecurity field, datasets (accumulated data) provide a competitive edge to research and development. Machine learning is technology that learns and analyzes large amounts of data, so it is important to collect large quantities of high quality data for this purpose. As shown in the table, we have accumulated a lot of data. For example, we have

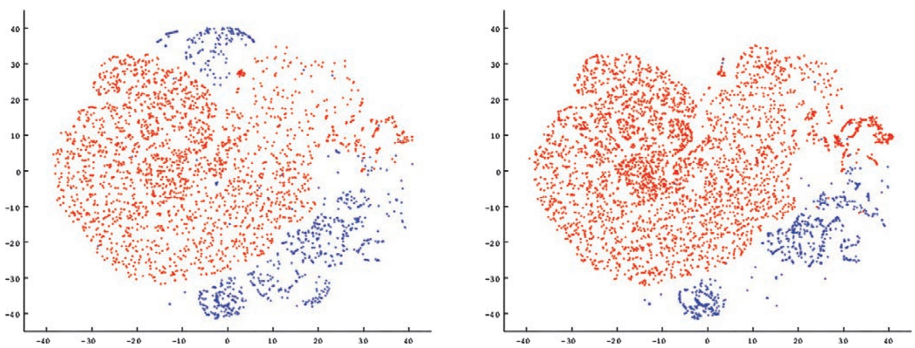


Figure 1 Feature vectors projected into two dimensions by t-SNE. The left figure shows the situation in the initial month of the experiment, and the right figure in the following month. Feature vectors associated with the DDoS attack are shown in red, and other feature vectors are shown in blue.

spent about ten years collecting, analyzing, and accumulating data from traffic on the "darknet" (unused IP addresses).

Examples of the use of machine learning

We are studying how this accumulated data can be used for analysis and automation using machine learning. Here are a few examples:

Packer identification: To make their creations harder to analyze, many malware authors use obfuscation tools called "packers," which are tools that encrypt executable files without compromising their functionality. When analyzing malware, it is important to identify which packer was used. Using a support vector machine (SVM), we can identify packers by classifying the binary sequences of packer decryption routines with an identification accuracy of 99.46% [1].

DDoS attack detection: We are using machine learning for the prompt detection of DDoS (distributed denial of service) attacks on the Internet. Backscatter* can be observed in the darknet, and this can be analyzed to enable the prompt detection of DDoS attacks. Darknet traffic is classified using 17 characteristics for each IP address, on the basis of which we implement classification processing using machine learning to determine whether or not the backscatter is related to a DDoS attack. When analyzing real traffic on the Internet, it is important to iterate learning considering that these trends can change any moment. Using a dimensional compression algorithm called t-SNE, we condense these feature vectors down to just two dimensions. Figure 1 shows a plot of the feature vectors obtained at the start of the experiment (left), and in the following month (right). In these figures, the red parts are feature vectors associated with DDoS attacks, and the other parts are shown in blue. You can see how these red and blue regions gradually change over time [2].

Android application analysis: Android apps are available from various online markets such as Google Play, but are sometimes mixed with malware. It is therefore important to be able to judge whether or not these apps are in fact malware. We have developed an Android

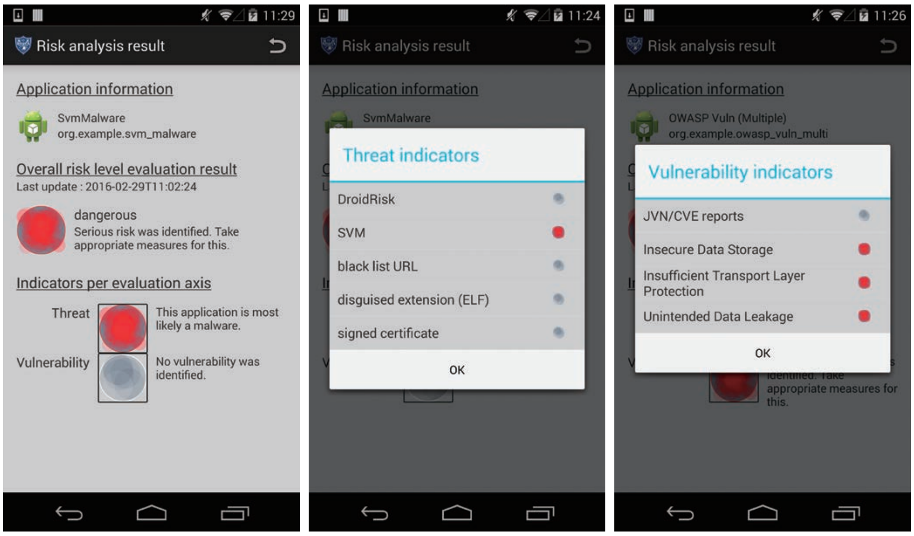


Figure 2 **This tool evaluates risk in terms of both threats and vulnerabilities. The left figure shows the overall evaluation, the middle figure shows the threat evaluation, and the right figure shows the vulnerability evaluation.**

app analysis technique that extracts features from Android apps, while at the same time collecting their descriptive text and category information from online markets. This information is fed into an SVM, which determines with 94.07% accuracy whether or not the app is malware [3]. As shown in Figure 2, we have also created a tool to visualize the analysis results. In this tool, the presence or absence of alerts is visualized (red lights on or off) from the viewpoints of both threats and vulnerabilities, and SVM is used to analyze the threat information.

Future issues and current direction

We expect that the scope of machine learning applications in the security field will expand rapidly in the future. We will continue to bolster our competitiveness by collecting and storing data. At the same time, we will redefine the issues to be addressed, and work on improving machine learning technology for this purpose. Furthermore, to apply machine learning technology to real security operations, we will need further technical innovations such as real-time security provision.

For the Tokyo 2020 Olympic and Paralympic Games, we are hoping to accelerate our research and development in order to promote the automation of security operations.

References

[1] R. Isawa et al., "An Accurate Packer Identification Method using Support Vector Machine," IEICE Trans. Fundamentals, Vol. E97-A, No. 1, pp. 253–263, 2014.

[2] I. Skrjanc et al., "Evolving Cauchy Possibilistic Clustering and Its Application to Large-Scale Cyberattack Monitoring," IEEE Symposium Series on Computational Intelligence, Hawaii, 2017.

[3] T. Takahashi et al., "The Usability of Metadata for Android Application Analysis," The 23rd International Conference on Neural Information Processing, Kyoto, 2016.

＊ Backscatter: Response (SYN-ACK) packets from a server that is being hit by a DoS (SYN-flood) attack with falsified IP addresses

A Research Platform Linking AI Technology to Brain Information Databases

Towards the Implementation of Publicly Available Cognitive-Function Evaluation Software

Tetsuya SHIMOKAWA
Senior Researcher, Collaborative Research Laboratory of AI Science, AI Science Research and Development Promotion Center / Neural Information Engineering Laboratory, Center for Information and Neural Networks

Joined NICT in 2010 after gaining a Ph.D. in Engineering and serving as Research Associate and Specially Appointed Associate Professor at Osaka University. Since then, he has been engaged in research on network analysis of the human brain.

Previous studies of AI focused on the development of expert systems that formulate the decision-making processes of human experts for applications in fields such as medicine. However, since expert decisions tend to be somewhat subjective, it turned out to be very hard to formulate the knowledge of experts, and this approach to decision making fell into decline. In this study, our aim is to extract not only subjective decisions from medical experts, but also test results and other objective medical data. By teaming up with medical and information science experts, we arrived at a database that can be processed with AI techniques. This article introduces our efforts in developing specific social implementations, which is the mission of the AI Science Research and Development Promotion Center (AIS) and presents a research platform that links AI technology to a database of brain information on psychiatric patients.

Background

At the Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT), we are conducting research aimed at fusing neuroscience with information and communications technology. In April 2017, the AIS was founded to bridge between AI research and applications.

Why should an information and communications research institute be studying the brain? The human brain contains about 10 billion nerve cells, and its complexity is far beyond any artificial neural network that has so far been developed. Moreover, its power consumption is said to be roughly 10 to 20 watts. Therefore, in order to develop new information and communications technology that is inspired by the brain, we have been studying the structure of the human brain, especially the networks formed by nerve cells in the brain.

The role of AIS in connecting the database core and analysis core

In addition to the brain information data and analytical techniques that have been accumulated by CiNet, we are also sharing huge quantities of schizophrenia-related brain information data owned by the COCORO*1 research group led by Associate Professor Ryota Hashimoto of Osaka University, Graduate School of Child Development (invited to AIS from October 2017), and we are working on the construction of a database for brain data from healthy volunteers and patients.

When disparate fields like medicine and information engineering are brought together in a joint study of this sort, there are certain issues that need to be addressed. This situation is illustrated in Figure 1. Medical institutions centered on the medical department form the "database core," and research institutions with knowledge

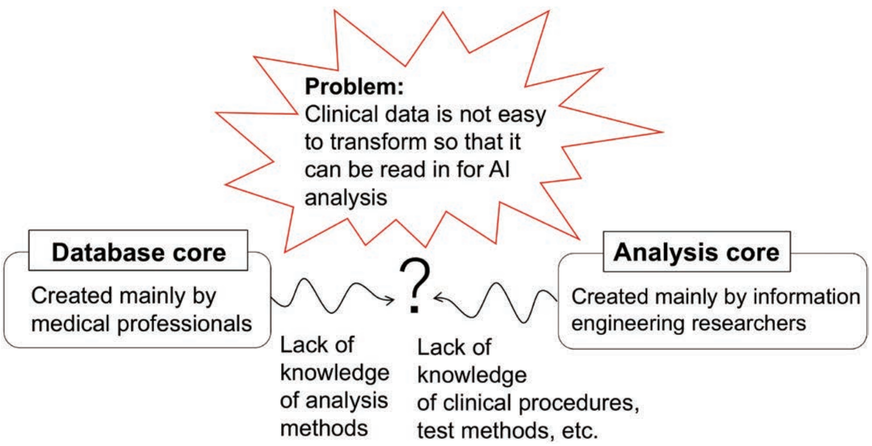


Figure 1 Problems that have arisen in research combining the medical field (database core) and information engineering (analysis core)

*1 COCORO (Cognitive Genetics Collaborative Research Organization), http://www.sp-web.sakura.ne.jp/lab/cocoro.html

*2 Sumiyoshi et al., Psychiatry Research, vol. 245, pp. 371–378, 2016.

*3 Hiroto Ito, Takashi Fukuda, Hideo Iwase, Atsushi Nishida, and Yasuyuki Okumura (2011): "Report on the FY2010 Social Welfare Promotion Project for Disabled Persons (Estimation of Social Costs of Psychiatric Disorders)," Juntendo University, pp.1–39.

*4 NICT commissioned study: Research and development of new network infrastructure technology for the future (FYs2016-2020) "Creation of evolutionary networked information processing infrastructure architecture based on extended bow-tie structures that can describe spatiotemporal dynamics" (Osaka University) of network analysis and AI analysis form the "analysis core." The database core lacks knowledge of analysis methods, while the analysis core lacks knowledge of clinical and examination methods. This gives rise to a problem in that the clinical data cannot be converted into a form that is easy to analyze.

Therefore, when we set up AIS, we created a joint research group comprising members from CiNet centered on AIS and three members from Osaka University as the "data decipherment core" to connect these two cores (Figure 2).

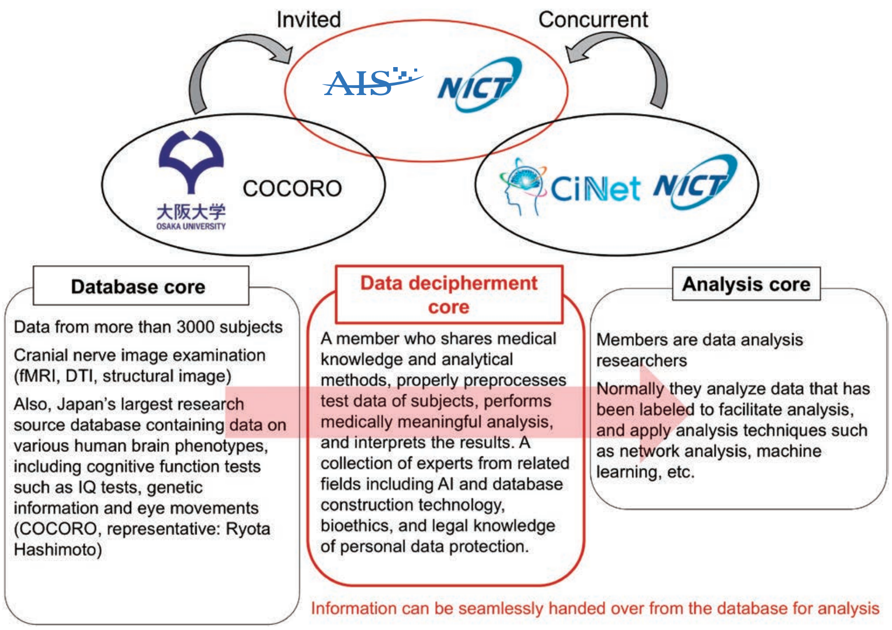


Figure 2 Role of the AIS data decipherment core connecting the database core and analysis core

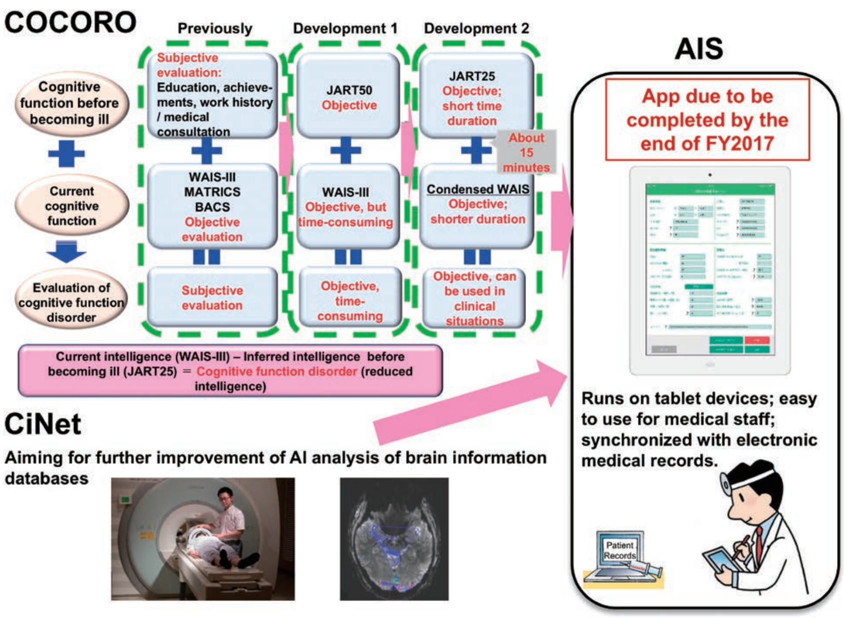


Figure 3 Development of software for the measurement of cognitive function with a shorter test

Development of software for the measurement of cognitive dysfunction

As a specific example of the sort of research that is possible in a project consisting of three cores, we will take a look at the development of software for the measurement of cognitive dysfunction. We mainly use fMRI scanners to measure and analyze brain function images, but when doing so, it is important to take the attributes of the subjects themselves into consideration. Of course, one needs to consider the subject's gender and age, but tests of cognitive function level such as IQ tests are of greater importance, even from a clinical viewpoint. If these tests of cognitive function level are performed using typical methods such as WAIS-III, then each test can take up to two hours to complete. However, a group headed by Associate Professor Hashimoto at Osaka University has succeeded in reducing the test duration to about 15 minutes by narrowing down the essential test items based on data from previous tests of nearly 400 subjects including healthy volunteers and people with psychiatric disorders.*[2] At AIS, we are currently conducting further analysis using AI technology such as machine learning to promote the development of a more precise and shortened test method for cognitive function level. Finally, as a social implementation, we are developing an application that can run on tablet devices (Figure 3).

Future prospects

According to a survey conducted by the Ministry of Health, Labor and Welfare, there are estimated to be at least two million patients with psychiatric disorders in Japan, with a social cost estimated to be more than 8 trillion yen.*[3] By combining medical knowledge with ICT, we expect that our efforts will support the discovery of brain biomarkers for psychiatric disorders whose prompt detection and diagnosis have so far been difficult, and contribute to the establishment of new diagnostic and therapeutic methods.

The importance of this research is not limited to the medical field. It is also strongly expected to have a major impact on the development of new information and communications technology. To understand what it is that makes the human brain function, we need to study not only healthy subjects but also subjects with different kinds of psychiatric disorders. Currently, in collaboration with a study commissioned by NICT,*[4] we are regularly exchanging information with a research group at the Graduate School of Information Science and Technology at Osaka University, while performing abstract modeling of analysis results of brain information database, and research aimed at applying our findings to information and communications technology. In the future, in order to apply this technology to the fields of medicine and info-communications and develop more social implementations, we will continue with the research and development for the construction of a brain information database containing information from healthy volunteers and psychiatric patients, and for the implementation of research projects that use this database.

Construction of an AI Data Testbed through "Hon'yaku Bank"

Keita HIKITA
Assistant Chief, Collaboration promotion office, AI Science Research and Development Promotion Center
Joined NICT in 2011 after graduating from law school with a Juris Doctor (JD). His various positions include working at the intellectual property office and legal and compliance office, and on secondment at the Ministry of Internal Affairs and Communications.

At AI Science Research and Development Promotion Center (AIS), as part of the construction of an AI data testbed, we are running a service called Hon'yaku Bank in partnership with Advanced Speech Translation Research and Development Promotion Center (ASTREC). This article introduces the history of the Hon'yaku Bank service, the details of this service, and its future prospects.

Efforts to construct an AI data testbed

At AIS, to promote the strategic R&D of AI technology in an open innovation style, we are working on the collection of data for an AI data testbed. As part of this initiative, we are building a "Hon'yaku (translation) Bank" by accumulating translation data that can be used for research and development of machine translation technology.

Current status of machine translation technology

At NICT, as part of our Global Communication Plan,*1 which aims to provide a social implementation of multilingual speech translation technology by 2020, we are researching and developing machine translation technology (called "neural translation") based on a deeply hierarchical neural network structure that is often seen in AI technology. To make neural translation more accurate, NICT added improvements to the neural network that it is working on in research and development, and since a greater quantity of translation data is known to result in better translation accuracy, it is essential to secure a suitable quantity of translation data (see Figure 1). For example, NICT is cooperating with and receiving translation data from the Japan Patent Office for the machine translation of patent documents.*2 In neural translation, since the translation accuracy is greater in fields where there is a lot of translation data available, a neural translation engine (available on the Min'na no Jido Hon'yaku @ TexTra® web site*3) that has been improved by this translation data has become particularly accurate at translating long texts such as patents and manuals.

Lack of translation data

Translation data is a collection of original sentences in various languages, paired with the results of translating these sentences. Currently, NICT receives translation data not only from the Patent Office but also from local governments, companies, various organizations, and so on, but the translation data NICT uses for research and development is mainly collected by crawling web pages.

However, the translation data published on web pages is only a tiny fraction of the translation data held by society. It's a fair assumption that many companies and other organizations have untapped resources of translation data in

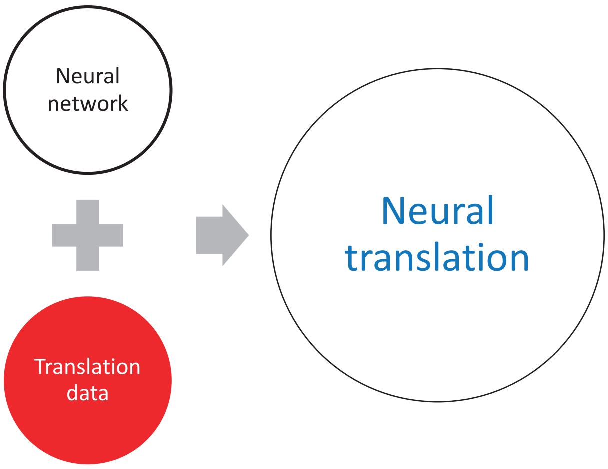


Figure 1 Neural translation

*1 http://www.soumu.go.jp/main_content/000285578.pdf
*2 https://www.nict.go.jp/info/topics/2016/04/160401-1.html
*3 https://mt-auto-minhon-mlt.ucri.jgn-x.jp
*4 https://www.nict.go.jp/press/2017/09/08-1.html

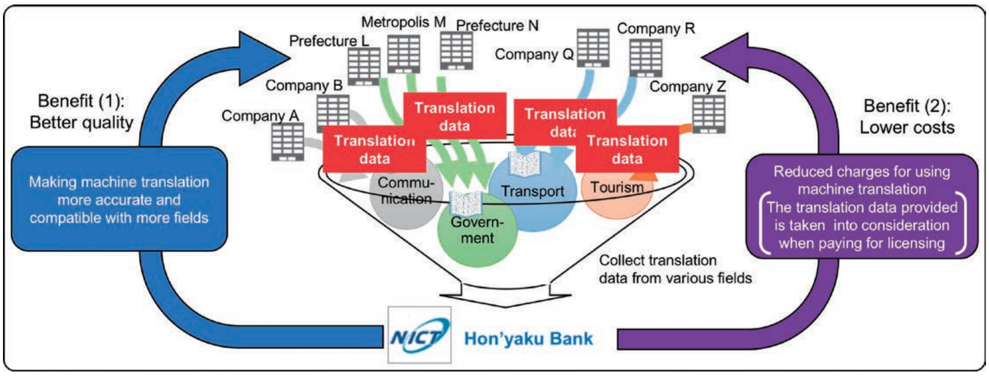


Figure 2 The Hon'yaku Bank concept

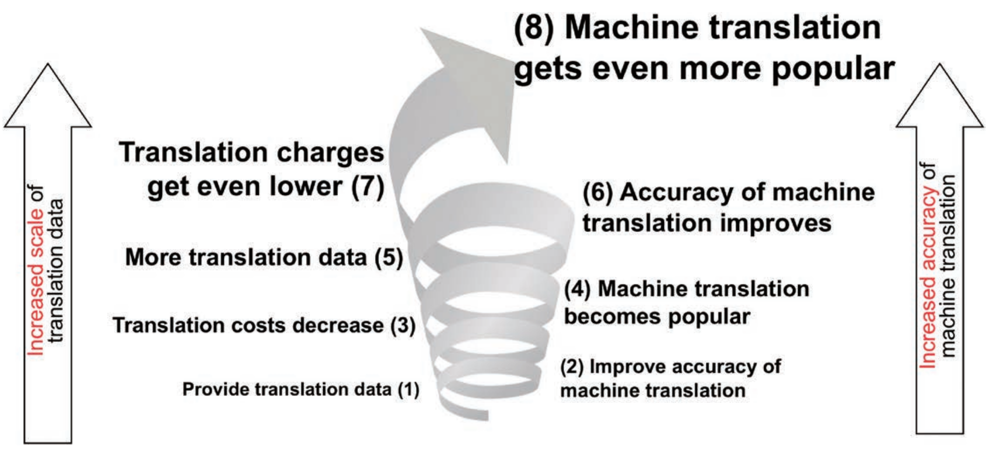


Figure 3 The ecosystem of machine translation

the form of technical documents and manuals. If Japan's machine translation technology is to be capable of competing with the rest of the world, the current quantity of translation data is insufficient. We therefore need to accumulate more translation data in various fields. To this end, the number of companies and other organizations that actively provide translation data will have to increase. It is also necessary to make it generally known that translation data is needed to improve the accuracy of machine translation.

The launch of Hon'yaku Bank

NICT has therefore launched a translation bank service called Hon'yaku Bank together with the Ministry of Internal Affairs and Communications.*4 In the future, we will work towards accumulating translation data from all over Japan at NICT to improve Japan's translation technology in terms of both its accuracy and the range of fields that it covers.

Anyone can submit translation data to NICT. There are three ways of doing so:

1. By registering bilingual translations at web sites that use the Min'na no Jido Hon'yaku @ TexTra® web site, whereupon they immediately contribute to the automatic translation R&D objectives of NICT.
2. By entering into an agreement to provide translation data to NICT. There are currently 29 organizations that provide translation data (as of the end of August 2017).
3. By providing translation data according to a license for the use of machine translation technology from NICT.

To clarify the benefits of providing translation data to NICT, we have prepared a new mechanism that accepts translation data when entering into a new license agreement for the use of NICT machine translation technology. This makes it possible to reduce the burden by considering the translation data that is expected to be provided at the time of license fee calculation (see Figure 2).

Translation data provided to us will be accumulated along with other translation data as part of the AI data testbed and will be used to improve the accuracy of machine translation. Then by allowing people to use our machine translation technology with improved accuracy, we will be able to gather more translation data, resulting in further improvements in translation accuracy. We believe this will create an ecosystem where people can use highly accurate automatic translation while nurturing automatic translation for the whole of society (see Figure 3).

From "All Japan" to "All Asia"

From now on, we will work on publicizing Hon'yaku Bank with the aim of immediately accumulating a million sentences from a hundred companies to make a hundred million sentences of translation data. In addition, since neural translation is language-independent, and Hon'yaku Bank is also language-independent, it will be possible to perform highly accurate translations in any language. So, in the future, if we can operate Hon'yaku Bank together with public research organizations throughout Asia to enable collection of translation data from all over the Asia region, then we should be able to develop translation technology that is even more accurate. This will contribute to economic and social revitalization not just in Japan but throughout Asia.

Through such efforts, we hope to eliminate language barriers by providing highly accurate translations in various fields, and we want to change Japan from a country where people have difficulty learning foreign languages into a country where people can communicate using the language of every other country in the world.

Encryption methods for servers and services that are compatible with long-term privacy protection and the use of big data

— The world's first homomorphic cryptosystem that allows privacy to be protected across generations —

Privacy protection technology is needed to facilitate the use of personal data in diverse fields such as medical care and finance. In the revised Personal Information Protection Act, measures were put in place to balance the use of big data and privacy protection. We have invented a cryptosystem that offers an improved level of security while allowing various operations to be performed directly on the encrypted data, thereby supporting not only the long-term security of information but also the use of this information in big data applications.

■Technical overview, applications, etc.

When data is encrypted, it cannot be directly subjected to complex processing. Instead, it must first be downloaded and decrypted to recover the plaintext before any calculations are performed. In this method, there is a risk of data leakage when the data is decrypted, which means it is necessary to introduce systems with a higher level of security. Also, in existing systems where the security level is updated while the data remains encrypted, it is impossible to process the data after the second encryption because the data format has to be converted and split apart.

Due to these technical issues, existing technology is regarded as being able to safeguard privacy information relating to individuals — such as information about diseases, treatments, and genetic information — for several years or decades.

With our invention, encryption and decryption are performed using a technique called "lattice-based cryptography" which is a cryptosystem that is difficult to break even with quantum computers, and in the updating process, we use a technique called "proxy re-encryption" whereby the destination of the new ciphertext can be altered without performing decryption. In a demonstration experiment, we found that encrypted linear regression calculations can be performed a hundred times faster than before, and that it is possible to increase the security level in a secure state where the data remains encrypted.

Based on these achievements, it has become possible to work safely with personal information such as information about diseases, treatments, and genetic information for long periods of time. As a result, it is possible to extract and calculate statistical data of diseases and treatments over a long period of time without seeing any personal information such as medical charts.

■ Searching for uses, applications, and collaboration partners

To promote the application of this technology in the real world, we have been performing verification trials of operations using anonymized public medical data. If this technique turns out to be useful, then we expect it will be introduced in medical information systems and gene preservation services. We are also studying applications in the fields of financial engineering and marketing research.

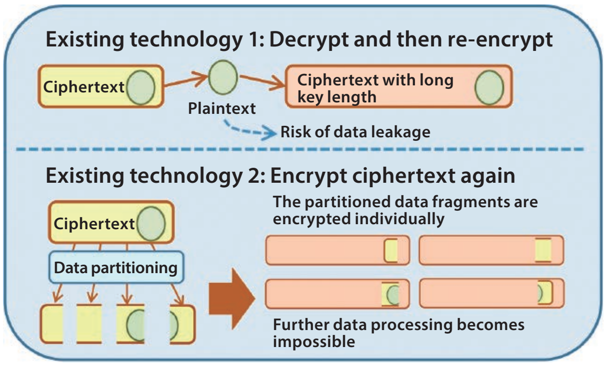


Figure 1 Conventional key length modification method

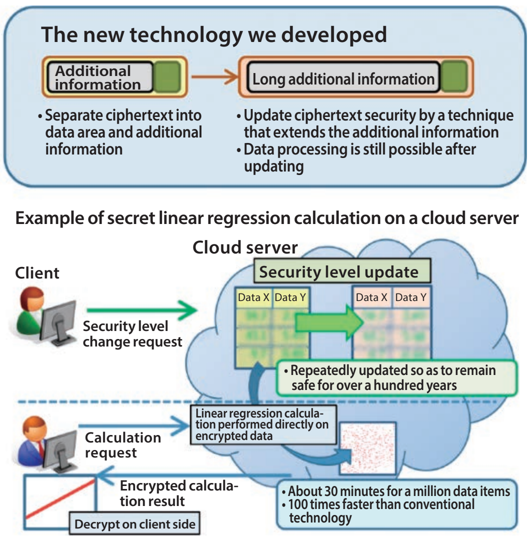


Figure 2 Our new homomorphic cryptosystem where the key length can be changed

<Patent information>

Publication No.: Patent Pub 2016-131271

Name of invention: Server, service method

<Contact (Inquiries, etc.)>

Intellectual Property Promotion Office, Innovation Promotion Department

E-mail: ippo@ml.nict.go.jp

TEL: +81-42-327-6950 FAX: +81-42-327-6659

Smartphone / tablet apps and translation website from NICT

At NICT, we are researching and developing multilingual speech translation technology. As part of our “Hon'yaku Bank” project, we are operating a translation website and providing diverse translation apps (introduced in this issue). The apps and website are free for anyone to use (except for Internet traffic costs), so you’re welcome to use them for various purposes.

Multilingual speech translation app

VoiceTra®

VoiceTra is a speech translation app that translates your spoken words into foreign languages.

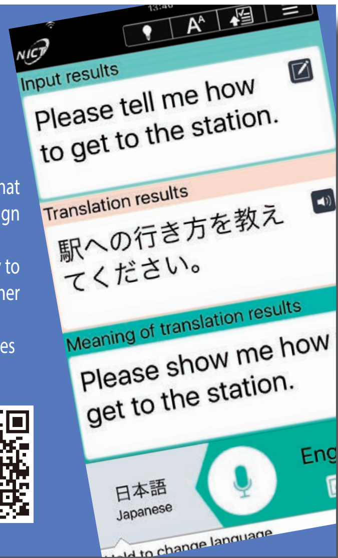
It has a simple user interface that is easy to operate and allows users to check whether the translation results are correct.

VoiceTra supports 31* different languages

* including dialects of Chinese and Portuguese

iOS 8.0 or later

Android 4.1 or later



Communication app for people with impaired hearing

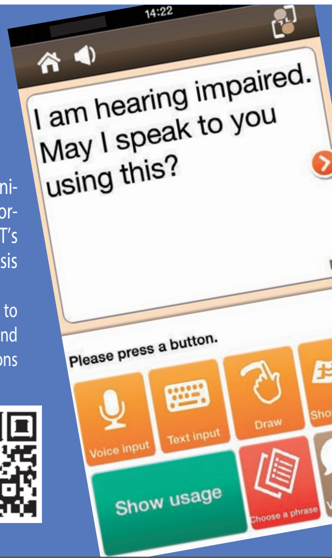
KoeTra

An app that supports smooth communication between hearing-impaired and normal-hearing people by making use of NICT’s speech recognition and speech synthesis technology.

Features: Speech input, speech output, easy to operate, allows fixed phrases to be registered and used, supports emoji, allows group conversations on multiple terminals

* Japanese language version also available http://www.koetra.jp/

iOS 8.0 or later



Service provided by: FEAT Limited

Speech converted directly to text, Supports conversations by finger writing

SpeechCanvas® Available only in Japanese

An app that uses NICT’s speech recognition technology to provide strong support for conversations with hearing-impaired people.

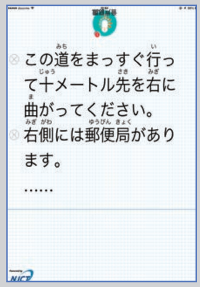
As words are spoken, they are displayed one by one together with furigana characters on the screen. It is also possible to draw pictures and characters on the screen with a finger. Since these operations are simple to perform and easy to understand, anyone can use this interface without any difficulty.

Furthermore, since the app can perform speech recognition even when disconnected from the Internet, it can be relied upon to work anywhere and at any time.

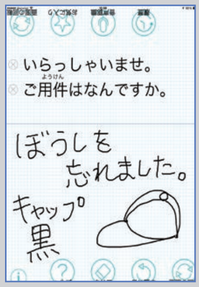
iOS 7.0 or later

Android 4.0 or later

Service provided by: FEAT Limited



To begin speech recognition, tap the speech recognition button. Words appear as text on the screen as they are spoken.



The handwriting area is below the bar in the middle of the screen. By touching the handwriting area directly with a finger, it is possible to draw letters and pictures.

Tapping the rotate button causes the screen display to rotate through 180° without having to turn the display device around. This is a very useful feature when holding a conversation across a shop counter, for example.

A machine translation site that everyone can use

Min'na no Jido Hon'yaku @ TexTra®

This is a site where users can create their own machine translation engines by adding translation examples and languages. After registering (for free), users can try out pre-registered machine translations, have files translated automatically, and translate words themselves on the web site.

https://mt-auto-minhon-mlt.ucri.jgn-x.jp/

A new type of site for supporting and sharing translations

Min'na no Hon'yaku

This is a site where documents from all over the world are translated with the cooperation of many people. Everyone can register and use the site for free, and various articles on subjects such as computers, current affairs, culture, cooking and encyclopedia entries have already been translated and published. The services of this site are also used by various NGOs and NPOs that conduct translation activities.

http://trans-aid.jp/

NICT NEWS 2018 No.2 Vol. 468
Published by **Public Relations Department, National Institute of Information and Communications Technology**
Issue date: Mar. 2018 (bimonthly)

4-2-1 Nukui-Kitamachi, Koganei, Tokyo
184-8795, Japan
TEL: +81-42-327-5392 FAX: +81-42-327-7587

E-mail: publicity@nict.go.jp
URL: http://www.nict.go.jp/en/
Twitter: @NICT_Publicity

Subscription applications accepted by E-mail or on the Web site.
ISSN 2187-4050 (Online)