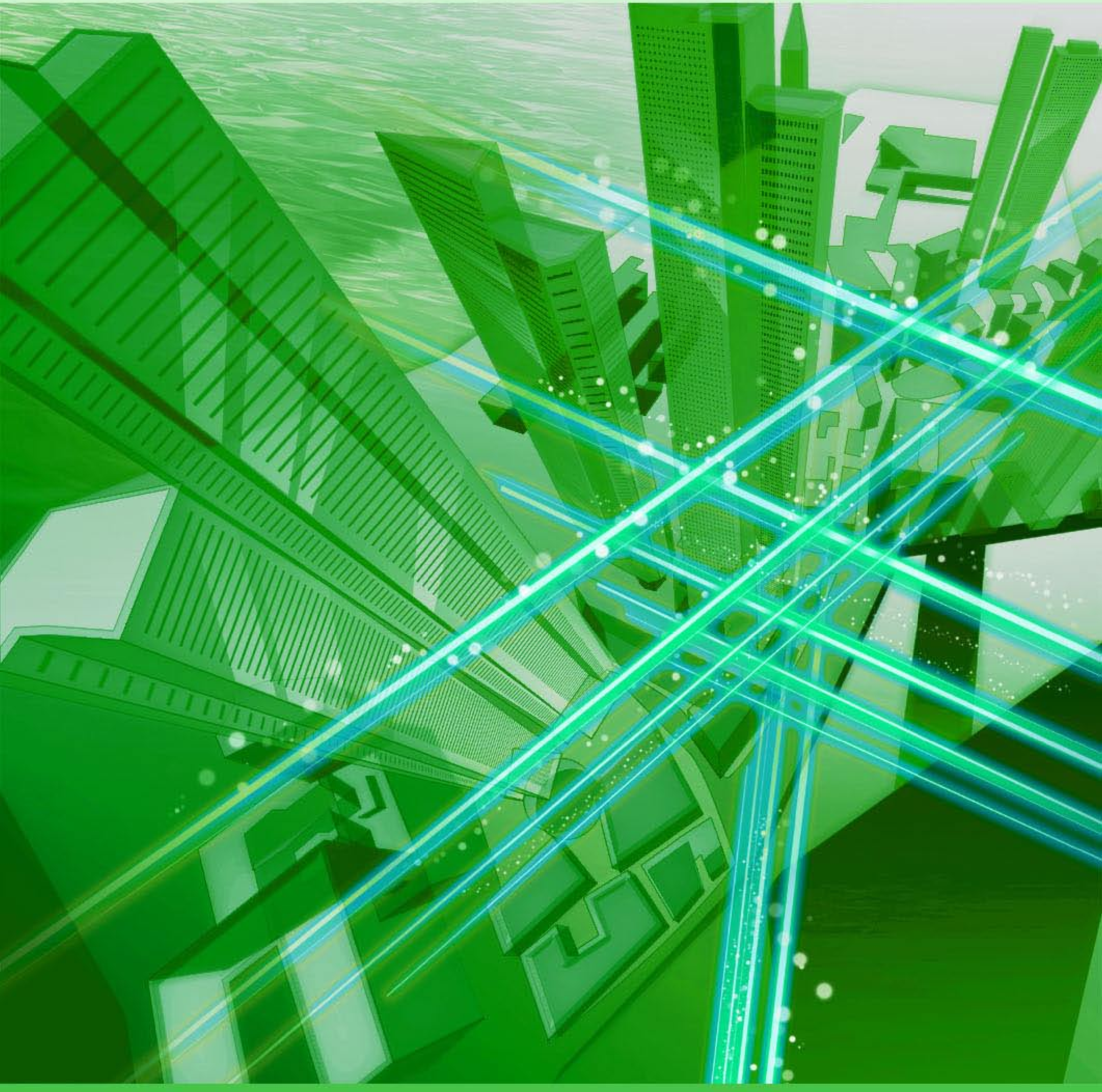
2
2016

NTT Technical Review



February 2016 Vol. 14 No. 2

NTT Technical Review

February 2016 Vol. 14 No. 2

Battle of Equals between NTT Laboratories and NTT Business Companies via OpenStack—A New Challenge toward the Next-generation Cloud Infrastructure

(From left) Susumu Okuhira, Hiroshi Sakai, and Masayuki Hayashi

Hiroshi Sakai, Senior Research Engineer, NTT Software Innovation Center

Susumu Okuhira, Director, Cloud Services, NTT Communications

Overview

OpenStack has been attracting attention as open source software for integrated management of IaaS (infrastructure as a service) consisting of servers, networks, storage, and other resources. Masayuki Hayashi, an NTT Communications manager and an active "cloud evangelist," spoke with Hiroshi Sakai, Senior Research Engineer at NTT Software Innovation Center, which has been involved with the OpenStack community since its early days, and Susumu Okuhira, Director of Cloud Services at NTT Communications, which manages the operation of OpenStack as actual services. He asked them about the background to adopting OpenStack for cloud services and about new services designed to achieve an Enterprise Cloud (next-generation cloud infrastructure).

Keywords: OpenStack, cloud, open source software

Transition to open source software (OSS) for economical and rapid development

Hayashi: As a cloud evangelist this past year, I have taken on the work of steering the cloud toward business with an objective point of view. Although any company is apt to promote its own products, I have tried to offer advice from a neutral standpoint, but at the same, I believe I can say without bias that the cloud services offered by NTT Communications (NTT Com) are superior. One reason for this is the use of an OpenStack system, which NTT Com and NTT laboratories have been collaborating on. Today, I would like to ask both of you about NTT's involvement with OpenStack. To begin with, I would like to ask Mr. Sakai about OpenStack activities at NTT Software Innovation Center (SIC).

Sakai: Despite being part of NTT laboratories, SIC is performing research and development with a focus on

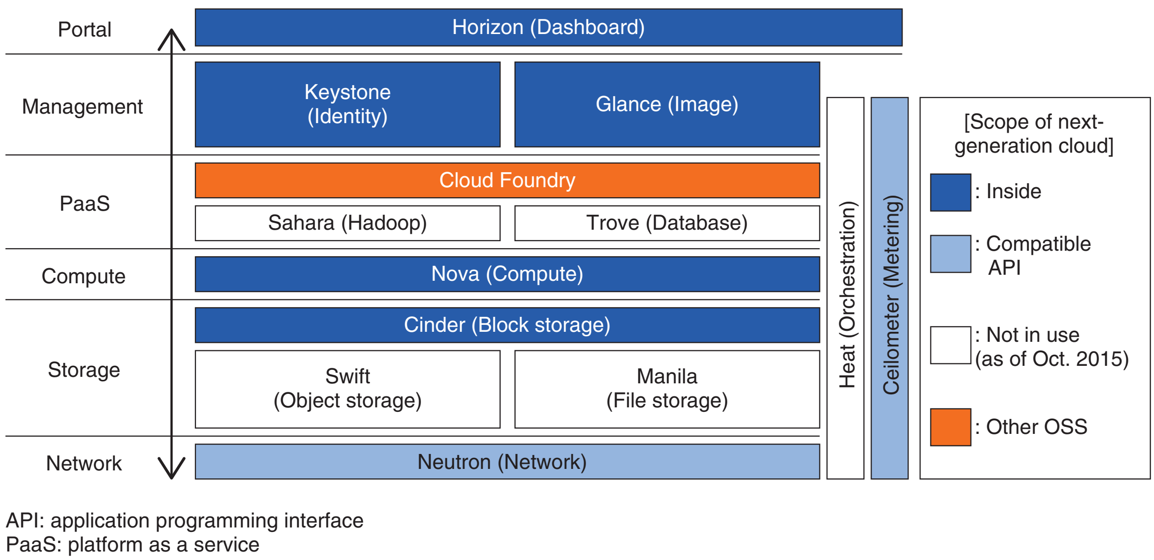


Fig. 1. OpenStack components in next-generation cloud.

practical use. At SIC, about 40 individuals are presently involved in OpenStack development in cooperation with NTT business companies such as NTT Communications and NTT DOCOMO with the aim of releasing a commercial system. OpenStack is a system that implements resources such as servers, networks, and storage on the cloud using virtualization technology (**Fig. 1**). Until recently, the approach taken by NTT laboratories was to execute projects on their own and develop services and products superior to those of other companies as proprietary technologies. OpenStack, however, represents a complete reversal to that approach as it aims to create a better system through the joint efforts of many people around the world.

Hayashi: SIC's participation in OpenStack began quite early. Can you provide some details on this involvement?

Sakai: There had been an increasing number of cases in which general enterprises throughout the world were adopting Amazon Web Services[*1] to run their internal business applications. Although NTT laboratories are generally quick to pick up on new technologies, there have been not a few times in which global competitors have invested a large amount of resources and taken the lead while we were still in the development stage. In technology fields in which the commoditization of services is accelerating, as in cloud-based IaaS, it is believed that the development of services by a large community of software developers in an open-source manner can result in good products inexpensively and quickly compared with the development of services as proprietary technologies by a single company. It is for this reason that we chose to make a transition to commercial development using OSS five years ago.

We have been participating in the OpenStack community nearly from the beginning of its development in 2010. At that time, a new version of the software was released every three months (every six months at present), and beginning with the initial release named OpenStack Austin, every subsequent release has been given a name in alphabetical order (**Table 1**). SIC has been involved with releases since the second release named OpenStack Bexar, so we have been participating in the OpenStack community practically from the very start. An OpenStack version named OpenStack Mitaka is scheduled for release in the spring of 2016. Here, the name "Mitaka" refers to the train station nearest to NTT Musashino R&D Center in which SIC is located, so the selection of this name for this release is truly gratifying for us.

*1 Amazon Web Services is a trademark of Amazon.com, Inc. and its affiliates in the United States and other countries.

Table 1. List of OpenStack release names.

Release name	Release date	Code names of included components
Austin	October 21, 2010	Nova, Swift
Bexar	February 3, 2011	Nova, Glance, Swift
Cactus	April 15, 2011	Nova, Glance, Swift
Diablo	September 22, 2011	Nova, Glance, Swift
Essex	April 5, 2012	Nova, Glance, Swift, Horizon, Keystone
Folsom	September 27, 2012	Nova, Glance, Swift, Horizon, Keystone, Quantum, Cinder
Grizzly	April 4, 2013	Nova, Glance, Swift, Horizon, Keystone, Quantum, Cinder
Havana	October 17, 2013	Nova, Glance, Swift, Horizon, Keystone, Neutron, Cinder, Heat, Ceilometer
Icehouse	April 17, 2014	Nova, Glance, Swift, Horizon, Keystone, Neutron, Cinder, Heat, Ceilometer, Trove
Juno	October 16, 2014	Nova, Glance, Swift, Horizon, Keystone, Neutron, Cinder, Heat, Ceilometer, Trove, Sahara
Kilo	April 30, 2015	Nova, Glance, Swift, Horizon, Keystone, Neutron, Cinder, Heat, Ceilometer, Trove, Sahara, Ironic
Liberty	October 16, 2015	Nova, Glance, Swift, Horizon, Keystone, Neutron, Cinder, Heat, Ceilometer, Trove, Sahara, Ironic, Zaqar, Manila, Designate, Barbican

Hayashi: Mr. Okuhira, please tell us about the work of Cloud Services at NTT COM and your duties as Director.

Okuhira: I began my cloud-related work in 2011, the year that the Cloud Services department was established. Prior to that, my work focused on the development of large-scale financial systems. In that work, I frequently interfaced with researchers in NTT laboratories in such fields as e-money, smart cards, and encryption, and I received much help from these individuals in various ways from the time that I was a young employee. The system that Cloud Services initially worked on was Cloud^n, which was launched in the United States in March 2012. Cloud^n is a public cloud featuring rich application programming interfaces (APIs) and self-management functions. Although NTT Com has traditionally focused on the global datacenter business, cloud-related orders are on the increase, and while revenues are still low, the cloud business is seen as one pillar of future growth.

Hayashi: Can you tell us how OpenStack came to be adopted in Cloud^n?

Okuhira: OpenStack was one of several options as the initial platform for Cloud^n, but since we sought stability in the early stages of Cloud^n as a commercial system, we decided to go with Apache-based CloudStack originally developed by Cloud.com but now provided as OSS (**Fig. 2**). OpenStack later came to be used by SIC and elsewhere within the NTT Group as its stability improved, and it was decided in 2013 to adopt it in the Cloud^n platform. Actually, it was decided to use OpenStack for a new service three months before the launch of that service. A number of problems arose before and after this service launch, but SIC staff responded to them rapidly, which enabled early implementation. At present, we are providing a stable service without any major problems.

Cloud Services is also developing a next-generation cloud infrastructure (**Fig. 3**) that will adopt OpenStack Juno as one of its platforms. In the year leading up to the Juno release, I met with Mr. Sakai every week and had many discussions with him on how to combine Juno with software-defined networking (SDN). We also incorporated the opinions of the business companies in a process that promoted community activities.

A relationship on equal terms between NTT laboratories and business companies bears fruit

Hayashi: It appears that the collaboration between SIC and NTT Com is producing results.

Sakai: In the development of a commercial system, it has often been the case that NTT laboratories would listen to the needs of a business company and then go on to improve product quality to a level that could withstand the demands of commercial operation. However, if such a relationship were to continue, it would not be unusual for the relationship between the business company and NTT laboratories to unconsciously fall into a consignment type of relationship

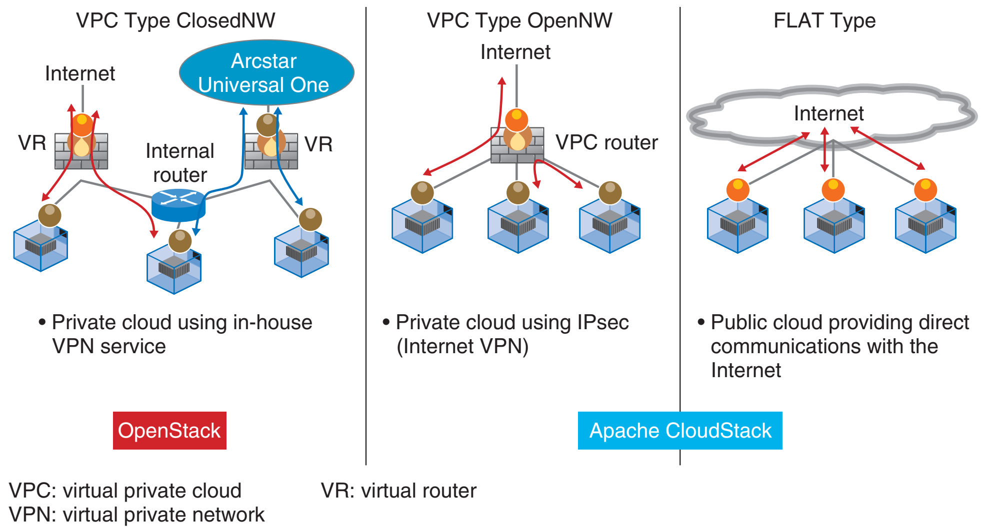


Fig. 2. Different configurations of Cloud^n Compute.

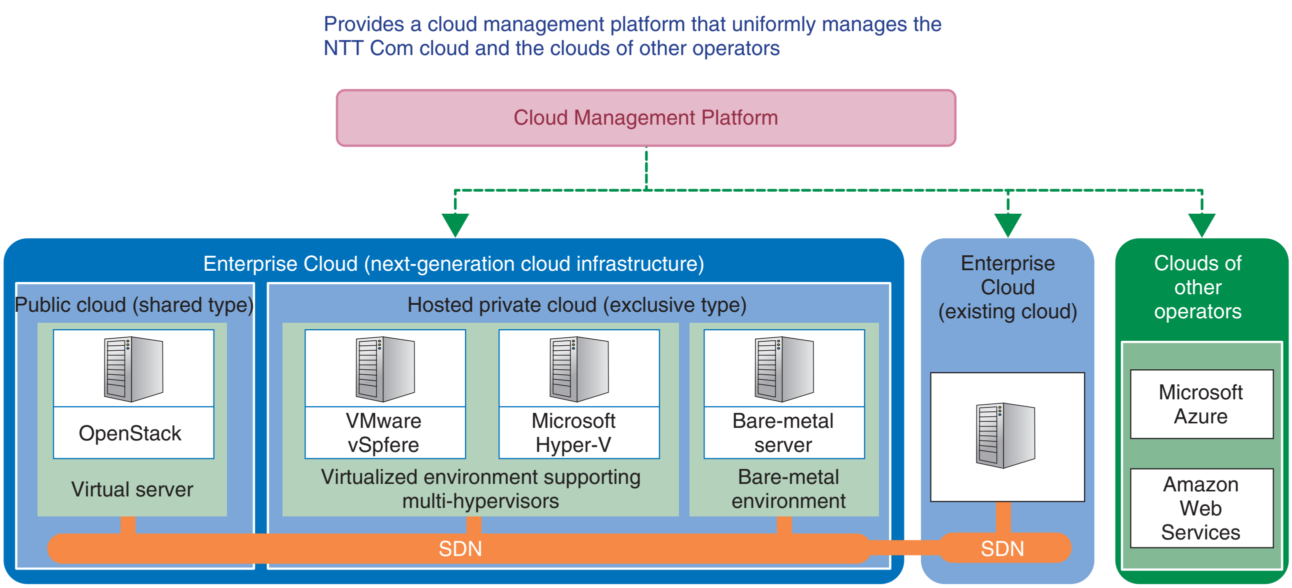


Fig. 3. Enterprise Cloud: features of next-generation cloud infrastructure.

between an order placer and an order receiver. However, that kind of relationship has not occurred between Mr. Okuhira and me. I have been able to voice the opinions of NTT laboratories without restraint, and we have been able to talk until reaching an agreement on equal terms. Forming such a relationship has been difficult in the past, but it evolved quickly in this project.

Okuhira: A multi-tenant scheme and scalability are essential to the provision of cloud services, and for a service provider, high availability (HA) is an important requirement. These functions, however, are not incorporated in OpenStack. I have had many discussions with staff at SIC as to how we can develop these functions in a sound way that befits NTT services.

Sakai: OpenStack is an accumulation of various types of OSS, so its development and troubleshooting in the event of a problem requires the use of all

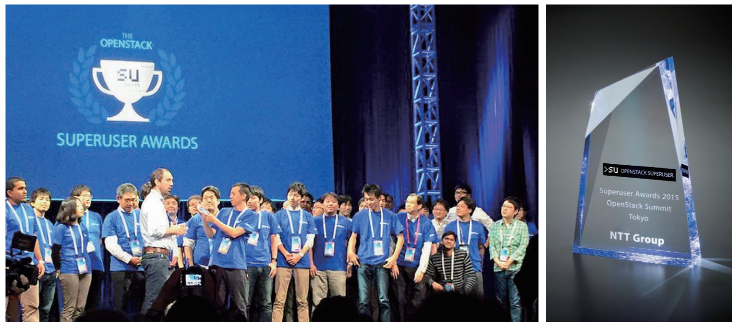


Photo 1. Receiving the Superuser Award and view of trophy.

knowledge pertaining to such a variety of OSS. Indeed, when a problem occurs just before service launch, an immediate investigation is necessary to determine what exactly is happening. However, since OpenStack operates on the basis of many distributed modules, logs for analysis purposes will be output in a separate and scattered manner at such a time. Clarifying the cause of the problem under such conditions would be time consuming, and for this reason, we have attached common identifications (IDs) to logs across modules. The specifications for this common ID and the HA function mentioned earlier by Mr. Okuhira are currently being proposed by SIC to the OpenStack community, which has praised these proposals.

Hayashi: The NTT Group received the Superuser Award at the recent OpenStack Summit Tokyo. Can you explain the reasons for receiving this award?

Sakai: More than 5000 people representing 56 countries attended the OpenStack Summit Tokyo held in Shinagawa from October 27–30, 2015. The OpenStack Summit welcomes OpenStack developers from around the world twice a year, and this meeting was held in Japan for the first time. The Superuser Award goes to one company for its achievements in implementing OpenStack in a way that creates value and for its contributions to the OpenStack community. The NTT Group was one of four finalists and was the first company from Asia to be chosen for this award (**Photo 1**). Through this award, NTT Com was of course commended for its application of OpenStack to Cloud^{n}, while NTT Resonant was commended for deploying OpenStack in the infrastructure of the "goo" web portal. It was also noted that NTT DOCOMO and NTT DATA use OpenStack, the former in its mail services and the latter in SI (system integration) matters. In short, the NTT Group received the Superuser Award because of the way in which its business companies were using OpenStack in actual business applications and for its contributions to the OpenStack community.

Okuhira: I believe that this award is the result of five years of contributions led by Mr. Sakai. The services provided by the business companies have become more stable as a result. We have also been recognized for our decision to continue using OpenStack in the next-generation cloud infrastructure. When the press release announcing this award [1] was issued, we received many inquiries from overseas affiliated companies, and I sensed their great expectations for us. I also think that this news served to publicize our cloud services among our customers and to spread the word that the NTT Group is deeply involved in OpenStack activities.

Hayashi: Which OpenStack components are the focus of attention at SIC?

Sakai: One is the Neutron module for network virtualization. In the beginning, the implementation of network-related functions was lacking compared to functions for servers and storage. For example, a single virtual machine was unable to support multiple network interfaces, so we focused our energy on upgrading such functions that were insufficient for enterprise use and proposed those changes to the OpenStack community. By the way, NTT Com adopted network virtualization by SDN in commercial services early on by global standards, and NTT laboratories made contributions here as well, such as by modifying plug-ins in OpenStack to incorporate a variety of SDN products. More recently, Neutron adopted the "Ryu" Border Gateway Protocol function developed and proposed by SIC as an open source SDN infrastructure. In this way, I am proud of the contributions that we at SIC are continuing to make to the OpenStack community in this field. At the same time, we are actively contributing to the Swift module that provides object storage services. In OpenStack, engineers actively involved in development work are chosen to be "core developers" in specific modules via a recommendation system. SIC has produced several core developers in the Neutron and Swift modules. Recent efforts in this regard include our proposal for a function that spans multiple OpenStack modules, by which I mean the application of common IDs to logs as I described earlier, and studies on new technologies such as NFV (network functions virtualization) and containers. A project for providing an HA function that we have named "Masakari" after my nickname also attracted interest at the recent OpenStack Summit.

Hayashi: There is also the idea of performing independent development outside of contributions to the OpenStack community. Where do you draw the line between those two pursuits?

Sakai: This is a difficult problem. Certainly, in the past, there was a strategy of holding on to independently developed functions to differentiate our services, but recently, we are apt, in principle, to submit our developments to the community. This is because failing to submit such a function early on may result in the appearance of that function in the community through other code. That function may then become mainstream, which would make us mere imitators. Of course, for functions supporting NTT carrier-grade operations or NTT proprietary advanced security functions, for example, we have decided that it is best that we keep these closed and continue to develop them independently. In such cases, however, we would try as much as possible to leave the main body of OpenStack untouched by implementing these original functions in plug-in form as external functions that would interface with OpenStack. In a May 2015 press release, it was announced that NTT laboratories had incorporated an NTT original high-speed secret sharing engine called "Super High-speed Secret Sharing" in OpenStack Swift in plug-in form thereby enhancing storage confidentiality [2]. This is a good example of such independent development.

Personnel training toward global rollout of the next-generation cloud infrastructure

Hayashi: What kinds of contribution activities take place within the OpenStack community?

Sakai: OpenStack is truly an open community. In the OSS community, although completed source code is open, development tends to be closed to specific companies. In contrast, the development process in OpenStack is very democratic; someone who would like to add something to OpenStack will make a proposal, which can be incorporated in OpenStack only if everyone else approves. This process begins by submitting one's desired function in the form of specifications, writing the code for that function, repeatedly making corrections based on reviews received from core developers, and finally obtaining consensus. It is only in this way that a proposal can come to be incorporated within the body of OpenStack. What source code will come to be incorporated in releases that are issued once every six months depends on the judgment of people in charge of each module, but some problems can arise in this process. For example, an abundance of proposed functions may result in some being overlooked, or the addition of new functions A and B may produce a conflict. Such problems have to be solved. A person who can consistently solve software bugs is very useful and highly appreciated in the OpenStack community, and his or her opinions may have great influence. In such ways, SIC will continue to be active in the OpenStack community in cooperation with the entire NTT Group.
Okuhira: At NTT Com too, we are adding members involved in OpenStack and seek to educate and train a variety of personnel in this field. In the end, however, there is a limit to the scale of human resources that we can contribute, but we seek to be active in the OpenStack community in cooperation with the entire NTT Group centered around SIC.

Hayashi: In the next-generation cloud infrastructure, what kind of new services using OpenStack will be offered?

Okuhira: In the next-generation cloud infrastructure, OpenStack will be used in virtual servers. We also plan to develop OpenStack into a form in which it can be used in financial systems and private networks while we fortify network communications between datacenters. Up until recently, OpenStack had been used only in services within Japan, but the Enterprise

Cloud using the next-generation cloud infrastructure is being rolled out globally. In the past, the sales system was based on paper-based applications, but from here on, it will be a paperless system and all services will be usable through APIs. In addition, use of the Keystone module will make it unnecessary to independently install user authentication/authorization functions for each component.

Hayashi: I see that there are plans to develop services using OpenStack even for industries that demand high reliability such as the financial sector. What is the outlook for research and development that includes OpenStack?

Sakai: We are looking to create added value by combining a customer's business data with information from the Internet of Things (IoT). One example is predictive maintenance, which is now a popular topic. As a typical scenario in this field, we can envision how the analysis of IoT data can indicate that a particular piece of equipment is about to fail. Then, in addition to advising the customer of this situation, we could also provide information on where an order for new equipment at minimal cost could be placed, or in the case of repairs, when the engineer in charge could visit the customer's premises to replace components, etc. Such processing will require more than just collecting IoT data in a simple and safe manner. It will involve a number of issues that have to be resolved, such as processing collected data and matching it up with cloud data as well as ensuring the secure flow of data between enterprises. I would like to see SIC make an all-out effort to make NTT's network and cloud a main player in achieving such a world.

Hayashi: Please tell us about future developments centered on the next-generation cloud infrastructure.

Okuhira: Our plan is to extend OpenStack to include a menu oriented to enterprise applications such as SAP*2 and Oracle*3. Furthermore, in addition to providing APIs and automation and self-management functions, we are also thinking about a service implementation system that enables a customer with insufficient facilities to add new facilities in a short period of time. Moreover, in addition to developing services, we are also thinking about ways of optimizing maintenance operations after service launch, an optimal way of writing API reference codes, and a mechanism that would enable customers to link up with each other so that necessary services could be provided when needed. It is not our aim to just keep up with the times but to be ahead of the times.

Experiencing the joy of people around the world using the results of our research

Hayashi: What would you say to young researchers?

Sakai: I believe that the joy of putting something that one has developed out into the world and seeing customers use it for the first time is a sensation that no one who has never felt it can understand. To achieve widespread use of a development as opposed to simply introducing it into business, many difficult and undesirable things need to overcome, but for people who would like to experience such emotion and joy together, I heartily recommend that they enter the field of OSS. The code that one writes becomes a foundation to be used by other people in the world, and there is much joy in seeing that happen. To begin with, please use yourself what you have developed!
Okuhira: I would like to ask young researchers in NTT laboratories to become involved early with NTT business companies. In particular, I think that accumulating a variety of experiences is good for researchers in OSS development. If you visit a business company, you will have to handle complaints from customers, so that will also help to improve your communication skills. Having the feeling that your group's technology in itself is not suitable for business is also a valuable experience.

Hayashi: Finally, can you tell us how you spend your free time?

Sakai: Well, there is a racetrack nearby, so I often go there with my children. Watching horses in a green, natural environment different from my everyday environment is emotionally soothing. A horse's eyes are truly beautiful. I also have fun playing mahjong with my family at home.
Okuhira: I enjoy working up a sweat playing softball with others in my neighborhood, and I also participate in PTA (parents and teachers association) activities. When at work, it is easy for me to motivate people as needed because of my position, but with people in my neighborhood, I have a horizontal relationship, so if I cannot skillfully convey what I would

*2 SAP is a registered trademark of SAP AG in Germany and in several other countries.
*3 Oracle is a registered trademark of Oracle Corporation.

like others to do, they will not do it. I therefore think about the best way of communicating with people to make things move along. I believe that it's important to have points of contact with all kinds of people and to have conversations with people outside the company. We should not become obsessed with our work; private activities can also be useful in various aspects of our lives.

Hayashi: Thank you very much!

References

[1] NTT Com press release issued in October 2015.
http://www.ntt.com/aboutus_e/news/data/20151027.html

[2] NTT press release issued in May 2015.
http://www.ntt.co.jp/news2015/1505e/150518a.html

■ Interviewees profiles

Hiroshi Sakai

Senior Research Engineer, Supervisor, Cloud System SE Project, NTT Software Innovation Center.

He received a B.E. and M.E. in electronic engineering from the University of Tokyo in 1991 and 1993. Since joining NTT in 1993, he has contributed to the development of network service systems including intelligent networks and the Next Generation Network. His current research interests include an OSS-based cloud management system and software-defined networking.

Susumu Okuhira

Director, Cloud Services, NTT Communications Corporation.

He joined NTT in 1987 and engaged in constructing various service systems involving electronic money, IC (integrated circuit) cash cards, and the certificate authority system for the Japanese Bankers Association. He was also involved in a large-scale financial systems project. He has been developing cloud services since 2011. He is currently responsible for managing a next-generation cloud infrastructure project.

■ Interviewer profile

Masayuki Hayashi

Cloud Evangelist, NTT Communications Corporation.

He joined NTT in 1995. He was engaged in sales to small- and mid-size businesses, after which he was involved in corporate sales in Malaysia as well as planning and operating international events. He joined NTT Communications after a reorganization of NTT and worked in the areas of business planning, corporate sales, and market development. He was also responsible for projects related to the Japanese government's cloud service and information technology policies. His involvement in marketing of cloud services began in June 2011. He took up his current position in October 2014.

R&D Activities at Machine Learning and Data Science Center

Naonori Ueda

Abstract

The Machine Learning and Data Science Center (MLC) was established in April 2013 as a research and development hub of big data analysis technologies at NTT laboratories with the aim of creating innovative services from a wide variety of big data. MLC uses machine learning and data mining technologies cultivated by NTT laboratories and a parallel-distributed processing platform (Jubatus) for high-efficiency and real-time data analysis to develop diverse big data analysis technologies and support big data services. This article introduces these big data activities at MLC.

Keywords: big data, machine learning, data science

1. Introduction

Machine learning and data mining are considered to be promising technologies for analyzing diverse types of big data. However, a single core technology is insufficient for achieving innovative services that make use of big data, so a composite technology that combines key technologies is essential. A parallel-distributed computation platform is also needed to efficiently process huge volumes of data. Furthermore, to conduct technology trials in the field and to uncover business needs, close discussions with specialists in applied fields are needed. Against this background, the Machine Learning and Data Science Center (MLC) was established in April 2013 at NTT laboratories as an inter-laboratory collaborative organization focused on big data analysis [1]. At MLC, however, we do not limit ourselves to collaboration within NTT laboratories. We also pursue open innovation in collaboration with NTT business companies, NTT Group companies, and even other firms and research institutions. Our aim here is to perform research and development (R&D) and provide technology support for various types of big data analysis both inside and outside the NTT Group. At present, our R&D activities at MLC are focused on the following technologies:

- Machine learning, data mining engine
- High-efficiency computation engine based on Jubatus
- Spatio-temporal multidimensional collective data analysis
- Detection of network-fault premonitory signs

We are supporting such big data analysis efforts at NTT laboratories through the brand name *himico*. Other Feature Articles in this issue provide detailed descriptions of the above, but this article provides an overview of these R&D areas and explains their importance.

2. Internet of Things

It has been more than five years since the concept of *big data* was first proposed. At first, however, it was more of a concept than a reality, and there were questions as to how big data differed from business intelligence, and how big data analysis differed from existing analysis technologies. In short, definitions of big data and big data analysis were unclear from the start. In time, though, advances in sensor technologies would enable the use of sensors in all sorts of fields including social infrastructures, medicine and healthcare, transportation, and agriculture, and at present, an environment is evolving in which massive amounts of data can be collected and analyzed in real time. This is none other than the birth of a concept

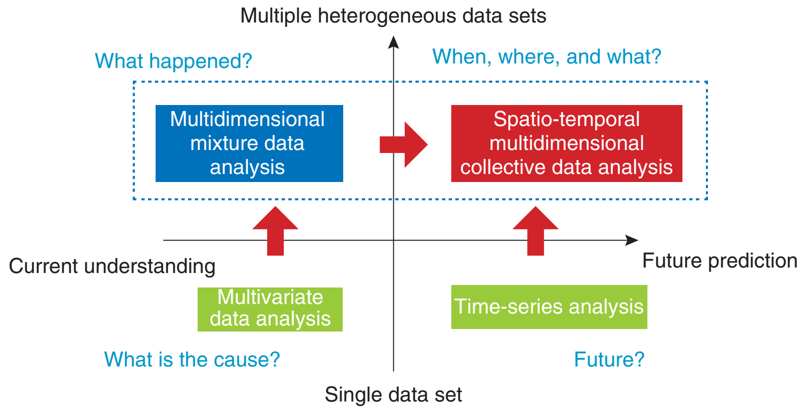


Fig. 1. Core technologies for the era of big data and IoT.

called the Internet of Things (IoT), and it is through IoT that the true nature of big data is finally revealing itself.

The R&D vision for big data analysis technologies in the IoT era as envisioned by MLC is shown in **Fig. 1**. In this vision, regression analysis (also called multivariate data analysis), which is representative of conventional data analysis, aims to describe an objective variable using multiple explanatory variables. That is to say, it is an analysis technique for determining whether an objective variable (for example, sales) can be expressed as a function of explanatory variables (factors). A famous case study of regression analysis is the estimation of wine prices by Professor Orley Ashenfelter. Using three explanatory variables (winter rainfall, average temperature during growing season, rainfall during harvesting season), he claimed on the basis of data only that the price of wine (objective variable: y) could be expressed by the following formula:

y = 12.145 + 0.00117 × (winter rainfall) + 0.0614 × (average temperature during growing season) – 0.00386 × (rainfall during harvesting season)

As it turns out, this formula correctly predicted the eventual price of 1989 Bordeaux wine before harvesting it, thereby demonstrating the usefulness of data analysis. Although not called *big data analysis* at the time, Ashenfelter's formula is now regarded as a classic example of such.

Also important in big data analysis is clustering technology that extracts clusters (groups) of data with similar features from big data to uncover latent value hidden in data. In the IoT era, however, there is a need for technology that can extract latent information spanning heterogeneous (multidimensional mixture) data sets, which cannot be discovered by individually analyzing different sets of data. To this end, NTT laboratories have developed a technique called multidimensional mixture data analysis that enables multiple heterogeneous sets of data to be simultaneously analyzed. The usefulness of this technique has been demonstrated through actual field trials.

3. Spatio-temporal multidimensional collective data analysis

The IoT era will also feature the collection of real-world information linked to time and space (location) from all sorts of sensors. This type of data is called spatio-temporal data. With reference to Fig. 1, time-series analysis attempts to model the temporal interaction or cause-and-effect relationship between data, but spatio-temporal analysis attempts to construct models that also consider the spatial dynamics among data. To analyze the spatio-temporal behavior of people and things and predict when, where, and what in real time, NTT laboratories established a research theme called spatio-temporal multidimensional collective data analysis. This form of analysis considers time and space along multidimensional axes and uses past data from a certain period of time to learn about the mutual relationships between time and space with respect to the *flow* of people, things, information, and other factors. Additionally, as an application example, R&D is underway on anticipatory people-flow induction based on spatio-temporal prediction. The importance of anticipatory induction and the usefulness of the proposed technique were demonstrated through computer simulation of a major event on a scale of

5000 people. Information on the above topics can be found in the feature article "From Multidimensional Mixture Data Analysis to Spatio-temporal Multidimensional Collective Data Analysis" [2] included in this issue.

4. Parallel-distributed processing platform (Jubatus)

In the IoT era, we can expect massive amounts of diverse types of data to be generated from moment to moment. Under these conditions, there is a need for a high-efficiency processing platform that can analyze such streams of data in real time. The feature article "The Latest Developments in Jubatus, an Online Machine-learning Distributed Processing Framework" [3] included in this issue provides an overview of Jubatus, presents examples of recent developments, and discusses future developments.

Jubatus is open source software developed jointly by Preferred Infrastructure, Inc. and NTT laboratories prior to the establishment of MLC in 2011. In contrast to the Hadoop parallel distributed processing platform that performs batch processing on stored data, Jubatus performs parallel distributed processing oriented to online data processing, which means it is a processing platform specially suited to the IoT era. The usefulness of Jubatus has already been demonstrated in the high-speed analysis of data generated by social networking services (SNS), but more recently, it has come to be applied in a variety of areas including real-time recognition of images from surveillance cameras (dress, behavior, etc.).

5. Ultrahigh-speed graph mining engine

In comparison to conventional tabular data, graph structured data that represent the relations among diverse items of information such as people, things, and places are referred to as *unstructured data*. As the name implies, graph structured data are represented by a graph in which the edges of the graph constitute connections between nodes or items of data. Graph structured data can represent a variety of phenomena such as links between web pages, friend relationships in SNS, and road networks. The process of extracting useful latent information hidden in graph structures such as the extraction of a strongly connected group of nodes is called *graph mining*, which has recently been a focus of research in the field of data mining.

NTT laboratories have been researching and developing a graph-mining engine named Grapon for ultrahigh-speed analysis of large-scale graphs. One important application of Grapon is balanced granularity partitioning. In a verification experiment performed by NTT DATA in China on traffic-signal control for eliminating road congestion based on a multi-agent simulation, the use of balanced granularity partitioning proved to be effective for accelerating simulation processing. Specifically, NTT DATA achieved high-speed parallel simulation here by treating the road network as graph data and partitioning the resulting large-scale graph into smaller well-balanced graphs to level the parallel processing load of congestion prediction. Information on this topic can be found in the feature article "Advanced Processing and Analytics for Large-scale Graphs" [4] included in this issue.

6. Network fault detection, premonitory signs and prediction, optimization

The increasing scale and complexity of networks in recent years has been accompanied by an increasing number of cases in which network faults and their causes are difficult to detect and deduce. At NTT laboratories, R&D has begun on a new type of big data analysis in the network field that uses machine learning and data mining technologies to estimate network conditions. This form of analysis inputs network operation data and traffic data such as system log (syslog) data that has not been used effectively in the past. The aim here is to develop technologies that use operation/traffic data to predict and detect network operation problems (faults, congestion, etc.), determine and visualize network conditions, and isolate problem factors.

It is common in conventional fault detection using syslog data to apply methods based on previously established rules (empirical knowledge). However, the conversion to software and the virtualization of network functions in software-defined networking means that network devices and their configuration and the syslogs and patterns generated will change dynamically, so formulating rules beforehand is difficult. In response to this problem, NTT laboratories have developed a fault detection method based on patterns generated in syslogs. This method is described in the feature article "Improving Network Management and Operation with Machine Learning and Data Analytics" [5] included in this issue.

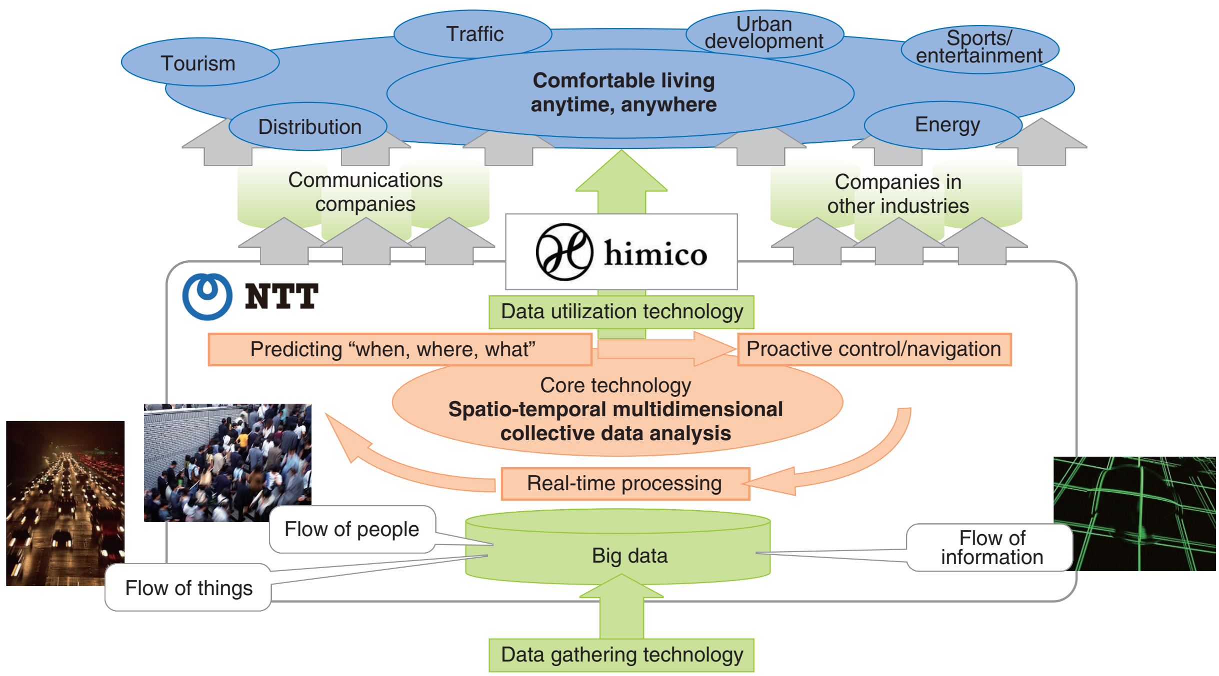


Fig. 2. Realization of comfortable living by analyzing traffic data in a broad sense.

7. Trial based on merging of big data technologies

In addition to the development of elemental technologies for big data analysis, MLC also conducts technology trials in collaboration with partners both inside and outside NTT. One example is the Fukuoka tourism trial targeting foreign visitors to Fukuoka City in Japan. For this trial, we merged big data analysis technologies developed by NTT laboratories to establish advanced, high-performance big data usage technology and tested its usefulness through the course of this trial. Specifically, we distributed a smartphone app to foreign visitors in Fukuoka and analyzed their movements based on the location and attribute information obtained from the app. This analysis helped us to discover interesting movement patterns and improve the opening rate of messages pushed to the app based on the analysis of tourist behavior. Information on this trial is provided in the feature article "Advanced, High-performance Big Data Technology and Trial Validation" [6] included in this issue.

8. Future developments

At MLC, we plan to collaborate with NTT laboratories and external parties to find problem solutions that are difficult to develop by a single organization. Our aim is to achieve a comfortable level of traffic in a broad sense that includes network traffic, people flow, and vehicle flow. We intend to accomplish this by using spatio-temporal multidimensional data analysis to develop technology for making real-time, highly accurate predictions of where and when an event will occur and to establish proactive control and induction technology (**Fig. 2**). Additionally, we aim to establish globally unprecedented and business-creating technologies for isolating the causes of silent failures and congestion and to achieve proactive network operation.

Today, at the dawn of the IoT and big data era, MLC in unison with the NTT Group plans to form collaborative relationships in rolling out next-generation big data business as the value partner that customers continue to select.

References

[1] Website of MLC (in Japanese), http://www.kecl.ntt.co.jp/rps/lab/mlc.html

[2] F. Naya and H. Sawada, "From Multidimensional Mixture Data Analysis to Spatio-temporal Multidimensional Collective Data Analysis," NTT Technical Review, Vol. 14, No. 2, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201602fa2.html

[3] T. Hayashi, M. Umeda, M. Sawada, K. Isagai, A. Yamanaka, and M. Tsunakawa, “The Latest Developments in Jubatus, an Online Machine-learning Distributed Processing Framework,” NTT Technical Review, Vol. 14, No. 2, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201602fa3.html
[4] Y. Iida, Y. Kishimoto, Y. Fujiwara, H. Shiokawa, J. Arai, and S. Iwamura, “Advanced Processing and Analytics for Large-scale Graphs,” NTT Technical Review, Vol. 14, No. 2, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201602fa4.html
[5] K. Ishibashi, T. Hayashi, and K. Shiomoto, “Improving Network Management and Operation with Machine Learning and Data Analytics,” NTT Technical Review, Vol. 14, No. 2, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201602fa5.html
[6] K. Noguchi, Y. Sato, and H. Shiohara, “Advanced, High-performance Big Data Technology and Trial Validation,” NTT Technical Review, Vol. 14, No. 2, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201602fa6.html

Naonori Ueda

Director of Machine Learning and Data Science Center, Senior Distinguished Scientist, NTT Communication Science Laboratories.

He received his B.S., M.S., and Ph.D. in communication engineering from Osaka University in 1982, 1984, and 1992. He joined NTT Yokosuka Electrical Communication Laboratories in 1984. In 1994, he moved to NTT Communication Science Laboratories, where he has been researching statistical machine learning, Bayesian statistics, and their applications to web data mining. From 1993 to 1994, he was a visiting scholar at Purdue University, Indiana, USA. He is a guest professor at the National Institute of Informatics and a visiting professor at Kyoto University. He is a Fellow of the Institute of Electronics, Information and Communication Engineers and a member of the Information Processing Society of Japan and the Institute of Electrical and Electronics Engineers. He became Senior Distinguished Scientist in April 2013 and was appointed Director of Machine Learning and Data Science Center in July the same year, after serving as Director of NTT Communication Science Laboratories for three years.

From Multidimensional Mixture Data Analysis to Spatio-temporal Multidimensional Collective Data Analysis

Futoshi Naya and Hiroshi Sawada

Abstract

This article introduces a multidimensional mixture data analysis technique that can efficiently extract significant features that transect different types of data with multiple attributes such as application logs available on the Web and sensor data collected from IoT (Internet of Things) sensors. The basic algorithm and an example of an application to review site data analysis are described. Spatio-temporal data modeling and an extension of a spatio-temporal multidimensional collective data analysis technique for predicting the time and place of near-future events are also explained.

Keywords: data mining, machine learning, spatio-temporal analysis

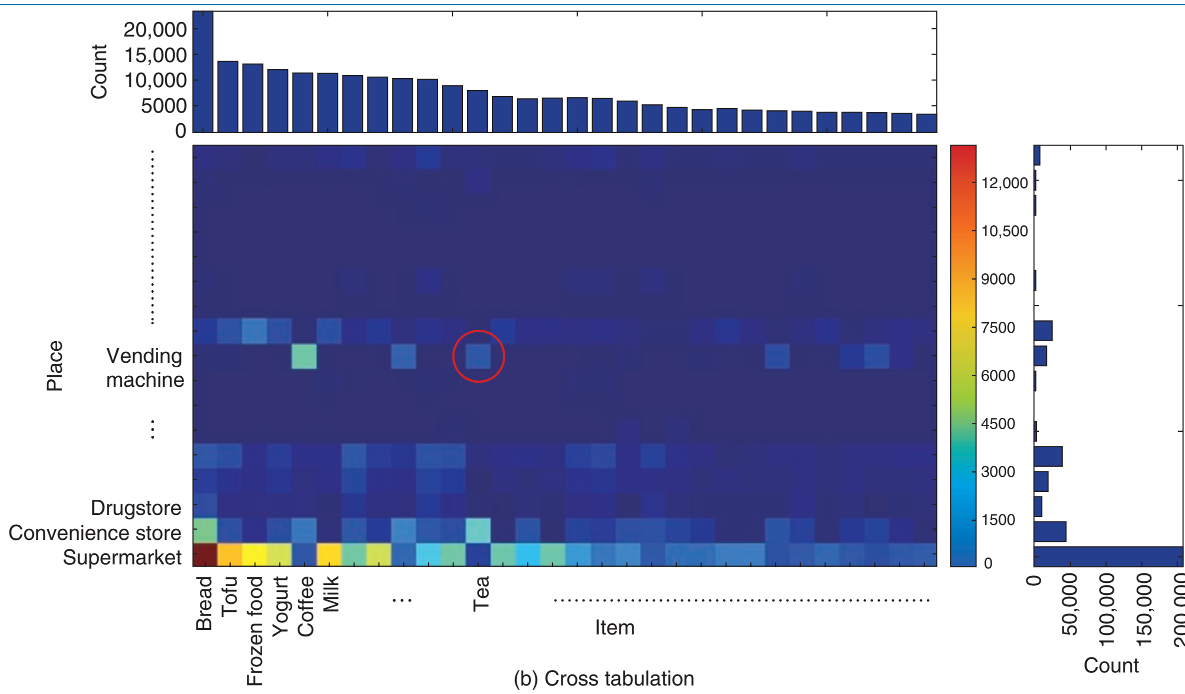
1. Introduction

As the use of e-commerce sites, smartphone apps, the Internet of Things (IoT) devices, and other such services and devices increases, various types of data related to the behavior of people and the movement of things are being generated. For example, the purchase log of an e-commerce site includes data such as the product item purchased, the time and place, and the customer age group and gender. Characteristic customer class and product groups are extracted by analyzing the huge amount of accumulated data (**Fig. 1(a)**). Another recent trend is application programs, mainly for smartphones, that obtain user location data from GPS (Global Positioning System), Wi-Fi, and beacons with user consent. Combining such location data with the purchase log information described above enables finer analysis that includes more details concerning the purchase circumstances, such as whether the purchase was made while the purchaser was out or at home.

Previous methods for extracting such customer classes and product groups have used cross tabulation, which, for example, aggregates the number of items sold for each combination of place and product, and represents the results in table format or as a matrix (**Fig. 1(b)**). The type of shop and the product item are referred to as attributes or axes. In the example shown in Fig. 1, there are two attributes (place and item), so the data are said to have two axes. The attributes can have various values because the place may be a supermarket, convenience store, or other shop, and the items might be coffee, tea, or another product. From the aggregation results, it is possible to determine the trend in a single attribute value, such as what items have high sales numbers for a particular place, or the places where a particular item is sold. However, the results become more complex as the number of possible attribute values increases. Because the data focus on combinations that have high aggregate numbers, it is difficult to extract features such as “Sales of black tea are relatively high, even for vending machines,” as indicated by the red circle in Fig. 1(b).

(a) Purchase log

Date	Time	Place	Customer	Age/Sex	Item
5/2/2015	07:31	Convenience store	A	30s/m	Coffee
5/2/2015	18:03	Supermarket	B	40s/f	Milk
8/2/2015	17:51	Supermarket	C	20s/m	Bacon
11/2/2015	12:18	Drugstore	A	30s/m	Shampoo
11/2/2015	13:03	Supermarket	D	30s/f	Chocolate
⋮	⋮	⋮	⋮	⋮	⋮



(b) Cross tabulation

Fig. 1. Example of cross tabulation of purchase log.

2. Multidimensional data analysis

One way to overcome the problem described above is multidimensional data analysis, in which a matrix that represents two-dimensional aggregate data is decomposed to the sum of K matrices (**Fig. 2**). The data represented by the K individual matrices T_1 to T_K are called clusters. The aggregate values of individual clusters obtained by multidimensional data analysis are biased toward a particular place or item, so it is easier to identify trends in individual clusters. For example, the items that are often sold in vending machines can be extracted in cluster 1; cluster 2 reveals that most bread is sold in supermarkets, and only about half that amount is sold in convenience stores; and cluster K indicates that the top item in supermarket sales is milk.

Our concern here is the method used in matrix decomposition. We take the original aggregate data as an $I \times J$ matrix T and decompose it such that it can be approximated as the sum of K matrices T_1 to T_K ($T \cong T_1 + T_2 + \ldots + T_K$). In the example described above, I is the number of places and J is the number of items. In that case, the individual matrices that result from the decomposition, T_i ($i = 1 \ldots K$), are also $I \times J$ matrices. Although the way the matrix is decomposed is arbitrary, multidimensional data analysis decomposes the original matrix so that each resulting matrix T_i ($i = 1 \ldots K$) is the product of $I \times 1$ column vectors a_i and $1 \times J$ row vectors b_i ($T_i = a_i \times b_i$), where we impose the constraint that the elements of each vector take non-negative values (0 or higher). The above decomposition can then be written as $T_1 + T_2 + \ldots + T_K = a_1 \times b_1 + a_2 \times b_2 + \ldots + a_K \times b_K = (a_1\ a_2 \ldots a_K) \times (b_1\ b_2 \ldots b_K) = A \times B$. This is nothing other than an approximation of the original $I \times J$ matrix T as the product of $I \times K$ matrix A and $K \times J$ matrix B ($T \cong A \times B$), where K is much smaller than I and J ($K << I, J$).

Matrices A and B, which are called factor matrices, are obtained by initially setting their elements to

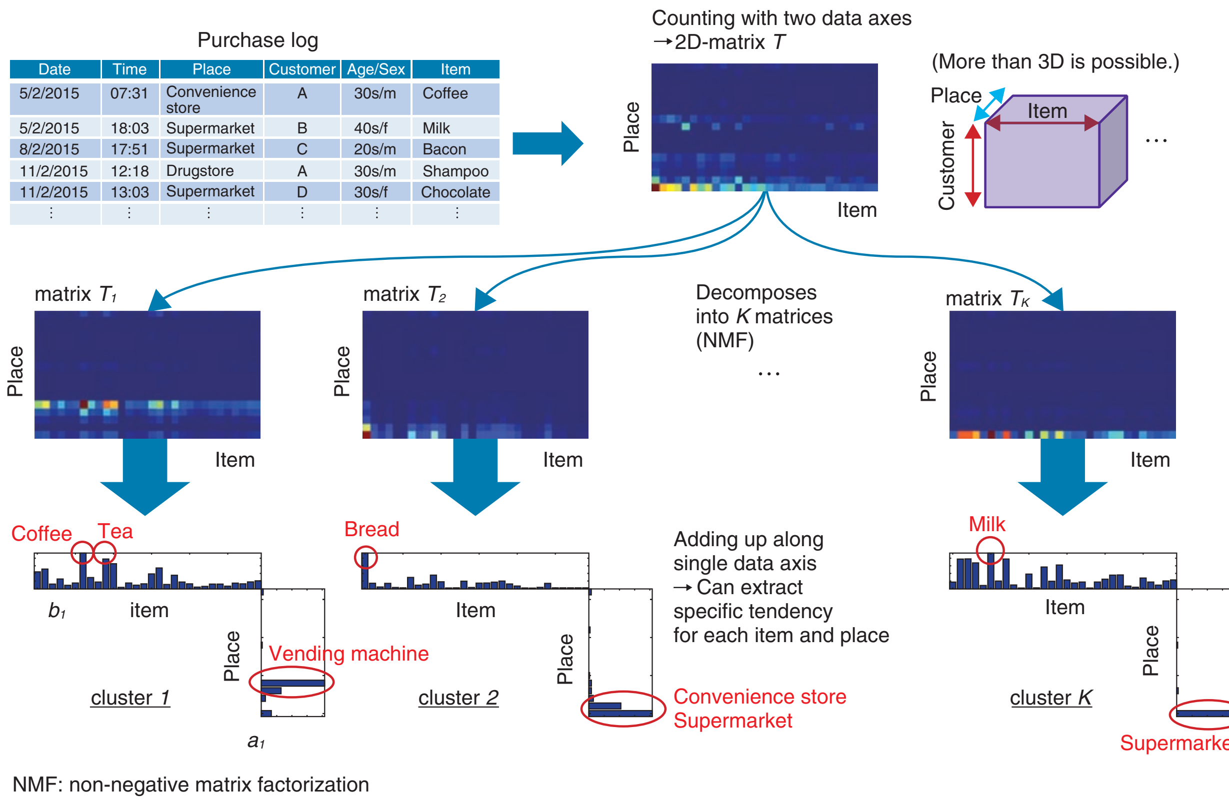


Fig. 2. Concept of multidimensional data analysis.

random values and then reconstructing them such that $A \times B = T'$, where the factor matrices A and B are updated so as to minimize the error between T' and the original matrix T. This process is repeated until the error value converges and the final factor matrices A and B are obtained. In the analysis example described above, A is the factor that corresponds to place, and B is the factor that corresponds to item. This method is a machine learning technique referred to as non-negative matrix factorization (NMF) and is commonly applied in data mining [1]. This approach can also be applied to high-dimensional data (tensors) that involve three or more attribute values, in which case it is referred to as non-negative tensor factorization (NTF).

3. Extension to multidimensional ‘mixture’ data analysis

NTF can be used to extract latent features from data by decomposing a matrix (tensor) of data that aggregates various attribute combinations into the product of factor matrices. However, as variation in the attribute values increases, bias in the aggregate data is more likely to develop. Analyzing aggregate data to which a new attribute (user age group) is incorporated in addition to place and item attributes would make it possible to extract customer clusters that indicate trends based on user class, place, and purchased item. Nevertheless, the number of attribute combinations increases explosively as the number of attributes increases and the data become sparser, meaning that the aggregate values for most of the combinations are zero.

The data aggregation for the case in which the purchase log shown in Fig. 1 involves four attributes (axes) comprises 240 million value combinations (e.g., 24 time values, one for each hour of the day × 100 place values × 1000 users × 100 products). Although it depends on the type of data, 99.9% of the elements for combinations of four or more attributes actually become zero. That is to say, less than 0.1% of all of the elements in the aggregate data have non-zero values, and it is difficult to obtain satisfactory results from factor analysis of such extremely sparse aggregate data.

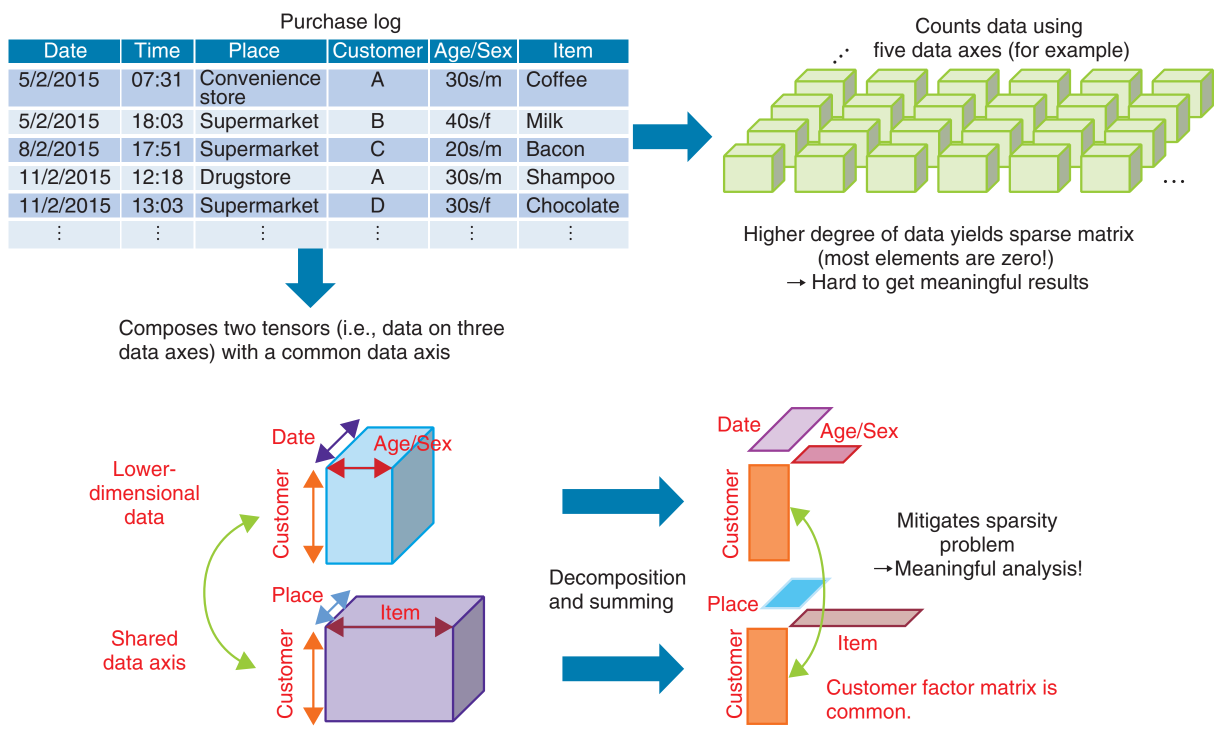


Fig. 3. Concept of multidimensional complex data analysis.

NTT laboratories have developed non-negative multiple tensor factorization (NMTF) as a solution to mitigate the sparseness problem. Instead of aggregating the original data as a high-dimensional tensor (five dimensions in this case), this approach constructs multiple tensors or matrices that represent non-sparse aggregate data (having mostly non-zero elements). The attributes (axes) are reduced to three or fewer, and the analysis is performed assuming that factor matrices share common attributes among multiple tensors or matrices (e.g., the customer axis in **Fig. 3**). Because multiple items of aggregate data are combined for the multidimensional analysis, this method is referred to as a multidimensional mixture data analysis technique.

4. Example using a Yelp data set

To evaluate the effectiveness of multidimensional mixture data analysis with NMTF, we applied it to an open data set collected by the Yelp review site in the United States (**Fig. 4**). Yelp data are available for research purposes [2] and contain a variety of review information on commercial facilities in the form of attributes, including the shop name, category, location, customers, rating, terms used in reviews, and day and time of *check-ins*, which provide customers' self-reported location via social networking service applications. From such data, it is possible to aggregate the following three types of tensor data, for example.

(1) Review tensor: user × shop × day of the week. The value is the number of reviews.

(2) Check-in tensor: shop × time × location (longitude and latitude). The value is the number of check-ins.

(3) Term frequency tensor: user × category × term. The value is the term-frequency of each word that appears in a review.

The tensors have the user and shop axes in common. By applying NMTF for the three tensors described above, we can obtain significant clusters that capture features with many attribute combinations such as "Weekend morning leisure activities" and "Meals at Japanese restaurants" (**Fig. 5**). From these clusters, it is possible to obtain results for areas where there are many facilities or restaurants, the day and time period when facilities are most often used, the geographic distribution, and the terms often used in reviews. The details of the analysis algorithm and

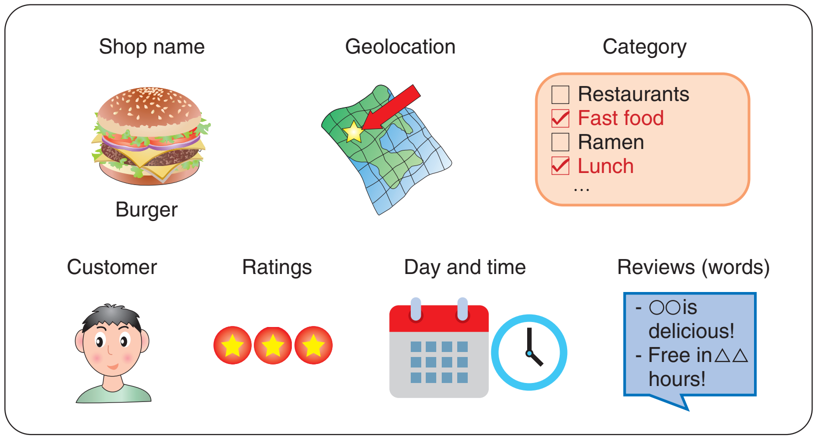


Fig. 4. Yelp data set.

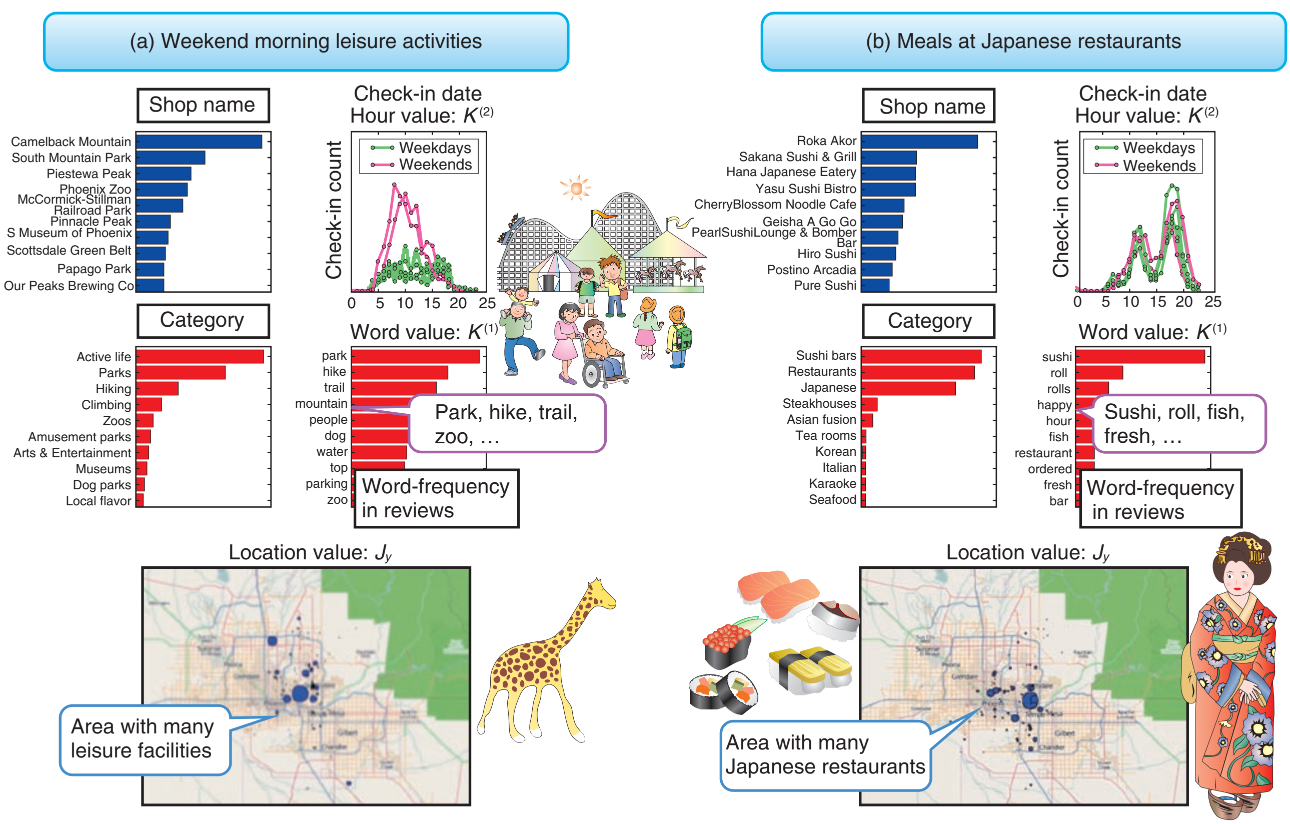


Fig. 5. Cluster examples extracted by multidimensional mixture data analysis.

results are presented in a published study [3]. In this way, NMTF can efficiently extract significant clusters from data with multidimensional attributes while mitigating the sparseness problem.

5. Future development

Although multidimensional mixture data analysis by NMTF can efficiently extract latent feature clusters from data, other factors such as the temporal

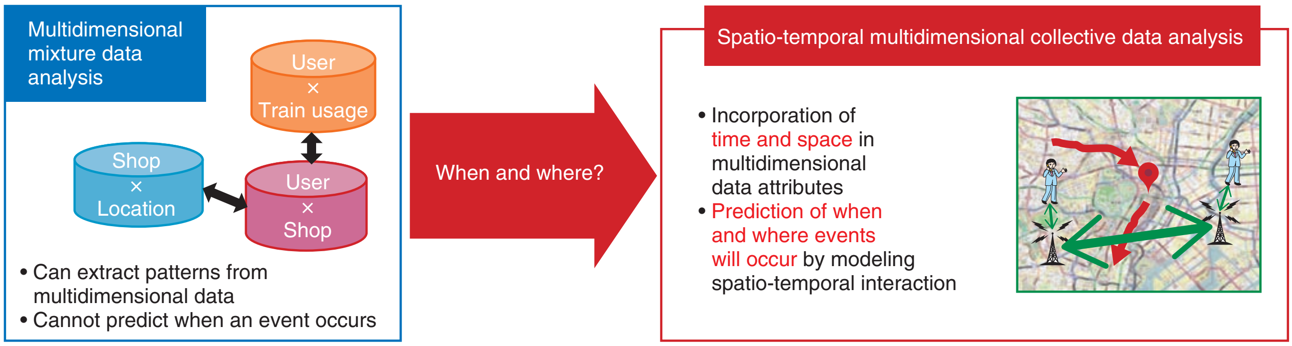


Fig. 6. Extension to spatio-temporal multidimensional collective data analysis.

cause-and-effect relationships among the elements of aggregated data or the spatial adjacency are not taken into account. In particular, the analysis of location data for users and vehicles can be based on an area divided into a spatial mesh and the number of persons and vehicles observed in each time period. However, even when NMTF is applied to such data, it is not possible to obtain analysis results that predict the times at which events such as congestion will occur in the future. For that reason, we plan to push forward with the development of innovative spatio-temporal multidimensional collective data analysis techniques that enable modeling of the temporal relationships of multidimensional data and prediction of future events. Spatio-temporal multidimensional collective data analysis considers four data aspects (time, space, multidimensional, and collective) for the purpose of gaining insight on near-future events (**Fig. 6**). The collective aspect is for estimating the spatio-temporal flow of people or traffic in cases where individuals cannot be recognized, such as when counting the number of persons or vehicles in a spatial mesh and only aggregate statistical data are available.

Looking toward 2020, we will continue with research and development on the use of spatio-temporal multidimensional collective data analysis techniques and real-time observation data to predict near-future events such as congestion in order to implement proactive navigation to relieve congestion at large-scale event venues [4]. We will also investigate how this research can be applied to stabilize the communication infrastructure.

References

[1] H. Sawada, "Nonnegative Matrix Factorization and Its Applications to Data/Signal Analysis," J. IEICE, Vol. 95, No. 9, pp. 829–833, 2012.

[2] Yelp Academic Dataset, https://www.yelp.com/academic_dataset

[3] K. Takeuchi, R. Tomioka, K. Ishiguro, A. Kimura, and H. Sawada, "Non-negative Multiple Tensor Factorization," Proc. of ICDM 2013 (IEEE International Conference on Data Mining), pp. 1199–1204, Dallas, USA, 2013.

[4] N. Ueda, F. Naya, H. Shimizu, T. Iwata, M. Okawa, and H. Sawada, "Real-time and Proactive Navigation via Spatio-temporal Prediction," Proc. of the First International Workshop on Smart Cities: People, Technology and Data, in conjunction with Ubicomp 2015, pp. 1559–1566, Osaka, Japan, 2015.

Futoshi Naya
Senior Research Scientist, Supervisor, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received a B.E. in electrical engineering, an M.S. in computer science, and a Ph.D. in engineering from Keio University, Kanagawa, in 1992, 1994, and 2010. He joined NTT Communication Science Laboratories in 1994. From 2003 to 2009, he was with Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International (ATR). His research interests include communication robots, sensor networks, pattern recognition, and data mining in cyber physical systems. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Robotics Society of Japan, the Society of Instrument and Control Engineers, and the Institute of Electronics, Information and Communication Engineers (IEICE).

Hiroshi Sawada
Senior Research Engineer, Supervisor, Proactive Navigation Project, NTT Service Evolution Laboratories.

He received a B.E., M.E., and Ph.D. in information science from Kyoto University in 1991, 1993, and 2001. He joined NTT in 1993. His research interests include statistical signal processing, audio source separation, array signal processing, machine learning, latent variable models, graph-based data structures, and computer architecture. From 2006 to 2009, he served as an associate editor of the IEEE Transactions on Audio, Speech & Language Processing. He received the Best Paper Award of the IEEE Circuit and System Society in 2000, the SPIE ICA Unsupervised Learning Pioneer Award in 2013, and the Best Paper Award of the IEEE Signal Processing Society in 2015. He is an associate member of the Audio and Acoustic Signal Processing Technical Committee of the IEEE Signal Processing Society and a member of IEEE, IEICE, and the Acoustical Society of Japan.

The Latest Developments in Jubatus, an Online Machine-learning Distributed Processing Framework

Takashi Hayashi, Masayoshi Umeda, Masato Sawada, Kaori Isagai, Akihiro Yamanaka, and Mitsuaki Tsunakawa

Abstract

Mobile terminals and other devices are now adopting multiple sensors, which generates voluminous data sets (big data). This necessitates rapid analysis of big data to understand the latest trends and current events. In this article, we introduce the latest developments in the open source community and commercial support activities for the distributed processing framework called *Jubatus*. Case studies are introduced to show that it offers deep analysis of big data in real time.

Keywords: real-time analysis, parallel and distributed architecture, online machine learning

1. Introduction

The popularity of the Internet and the rapid adoption of information and communication technology mean that large volumes of a wide variety of data sets are being generated. Examples include user data from various social networking services (SNSs) such as Twitter*1 and Facebook*2, log files from network equipment and servers, and data sent from vehicle-mounted sensors and home appliances.

We introduce here the online machine-learning distributed processing framework Jubatus and some examples of its application. It moves beyond simple statistical aggregation and keyword searches of big data and offers real-time and deep analysis of big data.

2. Jubatus

NTT's laboratories and Preferred Infrastructure, Inc. (PFI) commenced work on the Jubatus project in 2011 and released it as open source software (OSS) in October of that year [1–3]. Jubatus was developed with two goals in mind, and three features were added in order to achieve these goals, as shown in the green circles in **Fig. 1**.

The first goal is highly scalable processing performance for real-time analysis, which is achieved through the parallel and distributed architecture of Jubatus. The second goal, deep analysis, is realized by the adoption of machine learning techniques.

Real-time analysis enables quick response through sequential processing of continuously generated streams of data without temporary storage. The parallel and distributed architecture enables scale-out by adding servers in the same way as Hadoop and other large-scale data processing platforms. Given that it is often difficult to set rules to govern system processing in advance (since the rules are unknown, it is difficult to express constraints as rules, or the rules change over time), Jubatus adopts the machine learning approach to learn rules from data examples.

*1 Twitter is a registered trademark of Twitter, Inc. in the United States and other countries.

*2 Facebook is a registered trademark of Facebook, Inc. in the United States and other countries.

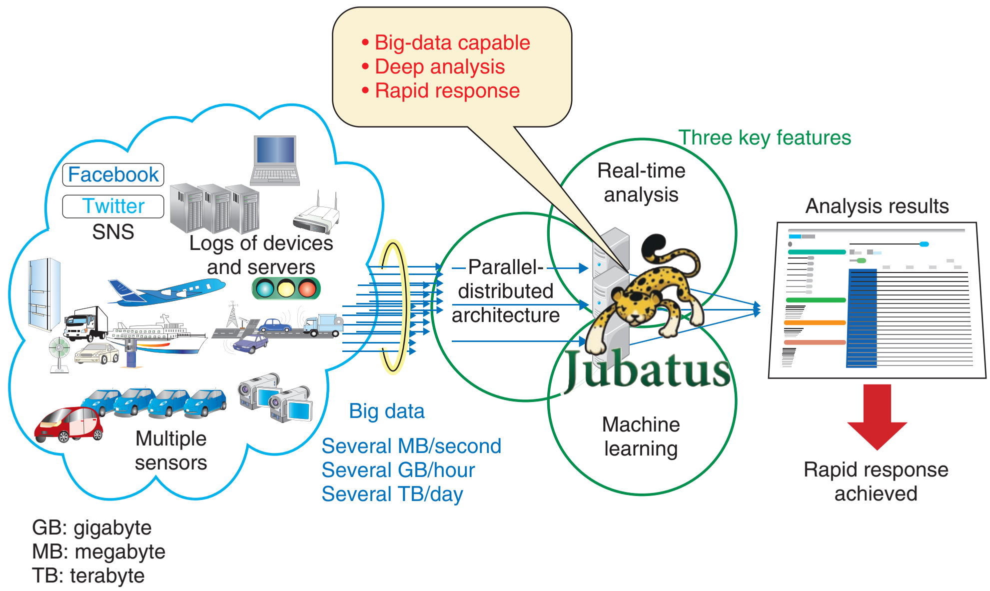


Fig. 1. Jubatus.

Jubatus has offered for the first time in the world the combination of real-time analysis, parallel and distributed processing, and machine learning [4].

3. Jubatus application examples

The three features of Jubatus are useful in detecting the onset of abnormalities or failures. For example, in the case of a factory, rapid response (real-time analysis) can be achieved by sequentially processing the data and logs output by a variety of sensors. The parallel and distributed architecture of Jubatus makes it easy to add extra servers whenever needed, so real-time response is possible even when the service area is extremely large and the number of sensors is enormous. Rather than adding dedicated systems to catch faults, the anomaly detection algorithm of Jubatus and its machine learning approach enable anomalies to be identified from the data of many sensors even if the detection pattern is unknown to the operator.

An example of the machine learning approach being applied to server log data is shown in **Fig. 2**. Principal components are extracted and plotted in three dimensions; data points that are unusual are shown in red. This representation yields regular and irregular patterns and enables unknown anomalies to be predicted.

To expand the application area of Jubatus, researchers at NTT's Machine Learning and Data Science Center are working in tandem with researchers involved in the *himico* initiative [5] on a proof of concept (PoC)[*3] project targeting areas such as network fault detection.

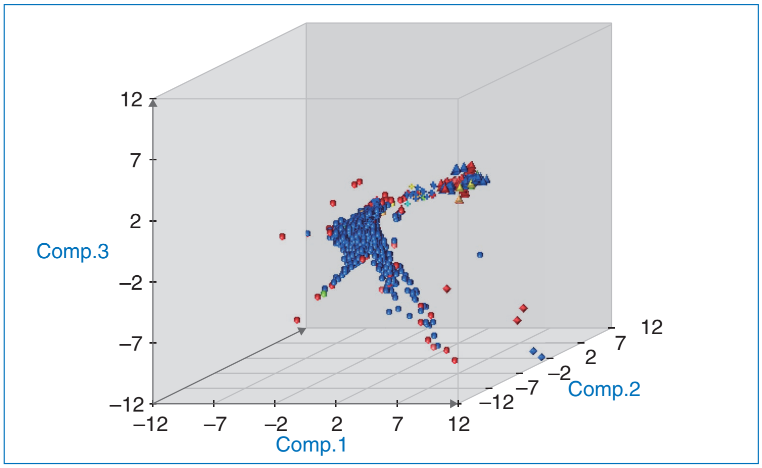
4. Latest development trends

Since its release as OSS in 2011, Jubatus has been repeatedly upgraded (nine times in 2014 alone), with emphasis placed on improving productivity. Consequently, it now offers an extremely efficient programming environment. In addition, data scientists and professionals in different fields are continually enhancing its various analysis algorithms. The algorithms currently available in Jubatus are listed in **Table 1**. As one example, the Bandit algorithm allows multiple solutions to a problem to be assessed in parallel, even while the service is in operation, and it determines which solution best suits the user's current situation. This algorithm is extremely useful for product recommendations and advertisement placements.

The analysis algorithms listed in the table take the form of plug-in modules, and user-written analysis modules can be loaded as plug-ins that can be freely selected and applied as needed.

JubaQL is the latest of our research results to be published. JubaQL offers an interactive interface with

*3 PoC: Minimal realization of a proposal sufficient to confirm its viability.



• Detecting unusual new patterns (shown in red) in log data; each data point has 41 dimensions.
• All data must be sampled to provide total inspection.
• This anomaly detection algorithm is based on unsupervised learning.

Fig. 2. Predicting failure from irregularities.

Table 1. Analysis algorithms implemented in Jubatus.

Analysis algorithm	Function	Use case example
Classification	Classifying data into categories	• Spam mail blocking • Categorizing tweets
Recommendation	Recommending items similar to given item	• Recommending items for EC sites • Linking search requests to ads
Regression	Estimating value from given data	• Predicting power consumption • Predicting share prices
Statistics	Aggregating frequency, standard deviation, maximum value, and minimum value statistics	• Sensor monitoring • Detecting data anomalies
Graph mining	Finding centroid and shortest route, etc., for given graph	• Social community analysis • Network structure analysis
Anomaly detection (Outlier detection)	Finding abnormal values (outliers) within a given data set	• Fraud detection • Fault detection
Clustering	Assigning given data points to a specified number of groups without using training data	• Segment analysis • Topic classification
Burst detection	Detecting bursts in traffic without using thresholds	• Detecting traffic congestion • Detecting SNS storms
Bandit	During service operation, tailoring service choices to suit the user's circumstances	• Personalized item recommendations • Linking users to ad sites

EC: electronic commerce

SQL (Structured Query Language)-like syntax to access the rich online machine learning functions available. Some of the benefits provided by JubaQL are shown in **Fig. 3**. Prior to its release, client programs that were needed for repeatedly evaluating different combinations of analysis processes had to be written in Ruby, Python, or another programming language. JubaQL offers highly efficient analysis processing and enhanced productivity.

5. OSS community activities and commercial support system

As mentioned above, Jubatus has been released as

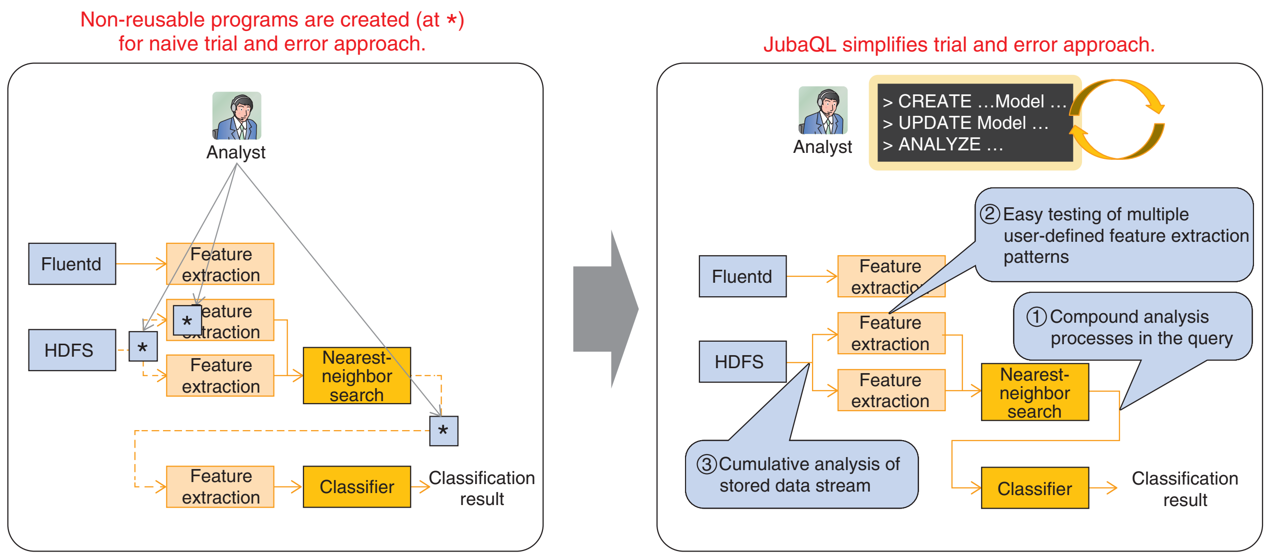


Fig. 3. Benefits of JubaQL.

an OSS project. Many of the distributed processing tools needed to undertake big data analysis such as Hadoop have been published as OSS projects and are attracting the interest of many developers and researchers. This has created an extensive ecosystem that has yielded new use cases and examples of data utilization. Our aim in releasing Jubatus as an OSS project is to connect with more users and thus discover new use cases and ideas for data utilization.

Our OSS project is extremely lively with over 2000 code commits, and we place emphasis on continually improving its usability. Lively community activities are also taking place, with some 590 participants attending two hackathons, four hands-on meetings, three casual talks, and one casual meetup event. The casual talks have been very interesting, as they introduced usage reports from organizations other than the original developers, NTT laboratories and PFI, which confirms that the use of the system is expanding.

Another significant indicator is the drive to introduce Jubatus to commercial products and thus expand its use by businesses. This will require careful utilization of commercial knowledge through selection of the most suitable analysis algorithm and performance tuning. To better support businesses, NTT Software Corporation established a commercial Jubatus support service in January 2014; it offers a wide variety of support functions to enhance the commercial application of Jubatus.

6. Case studies

6.1 Case 1: WatchBee

Jubatus is used in the summary display function of WatchBee [6], a reputation analysis service provided by NTT IT Corporation (NTT-IT). This function analyzes large-scale data such as SNS data, extracts the features of each kind of data, and groups similar data based on the correlation between features. Rapid data analysis is essential for this function because large quantities of new data are created every minute, and WatchBee must analyze that data every 30 minutes. The online processing of Jubatus has substantial advantages for this kind of deep and large-scale real-time analysis.

6.2 Case 2: hitoe

Jubatus is well-suited for analyzing sensor data. With conventional techniques, threshold values are necessary to classify human poses and motions when analyzing their three-axis accelerometer data*4 acquired from wearable devices such as *hitoe* [7]. However, it is difficult to determine the proper threshold values. Furthermore, optimum values may differ from person to person because of their different physiques and manners of poses and gestures.

Jubatus can automatically find the criteria for

*4 Three-axis accelerometer data: To determine sensor orientation, data captured on the sensor's X, Y, and Z axes are processed to identify the direction of gravity.

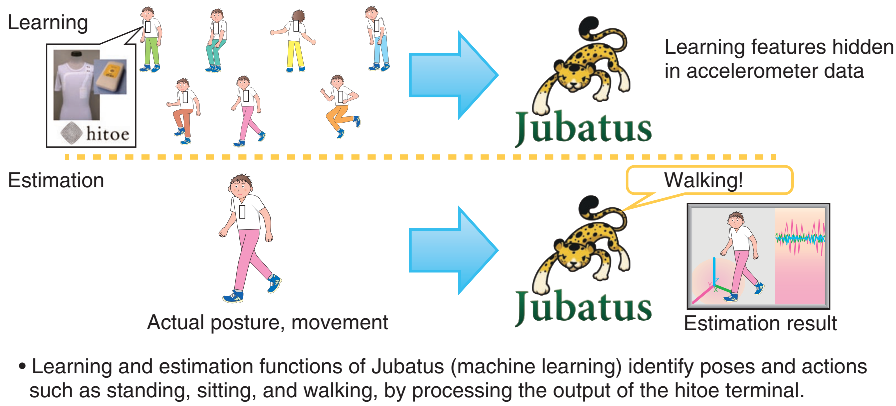


Fig. 4. Example of Jubatus used with hitoe.

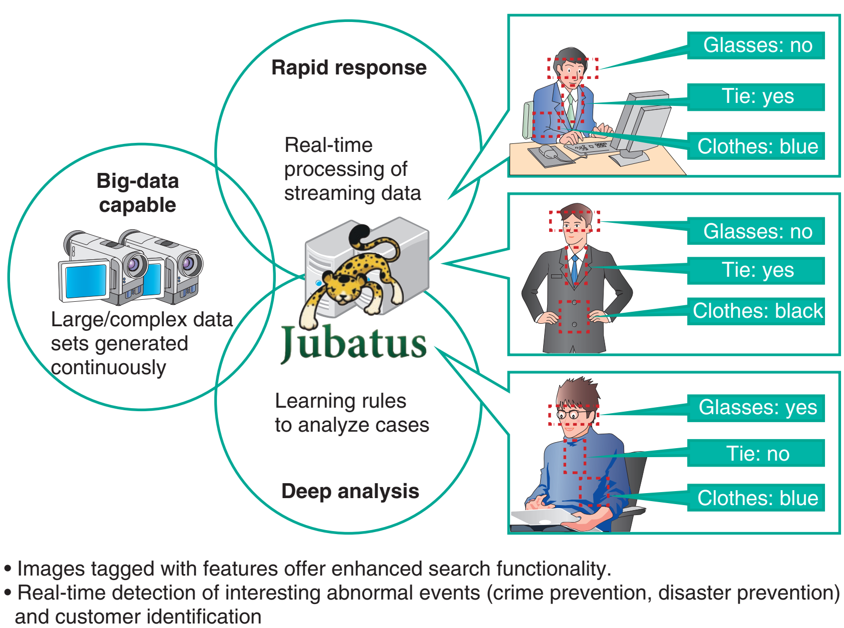


Fig. 5. Analyzing surveillance videos.

establishing thresholds by learning from the sensor data of each pose and motion (**Fig. 4**). Personalized criteria for every individual are generated by Jubatus since it is able to distinguish the differences in personal physiques and in the manner of poses and gestures. Furthermore, Jubatus can improve the accuracy of criteria by using its incremental learning function on misclassified data.

6.3 Case 3: Surveillance video analysis

Jubatus can also be used to analyze footage captured in surveillance videos (**Fig. 5**). Current security systems are inefficient, as the stored videos must be manually rewound to around the time of the incident and then manually reviewed. This can allow crime precursors to be overlooked. The classification power of Jubatus can be used to analyze the videos as they are captured and to extract notable features such as the color of clothing, the presence of a backpack, or actions such as using a smartphone or reading a book. The image data are tagged with the features, which greatly facilitates subsequent processing and search operations.

We intend to utilize the latest video recognition technology such as *deep learning* [8] to further enhance the accuracy of attribute classification.

7. Future developments

Current activities are mainly directed towards strengthening the analysis algorithms of Jubatus and enhancing the usability of JubaQL. A long-term goal remains raising user awareness of Jubatus. Future activities will focus on introducing Jubatus to business applications, improving service quality (higher operability and reliability), and adding peripheral functions. Finally, we intend to greatly expand the application area of Jubatus by introducing it in new fields such as AI (artificial intelligence) and IoT (Internet of Things).

References

[1] Website of Jubatus, http://jubat.us/
[2] Twitter account of Jubatus, https://twitter.com/jubatusofficial
[3] Website of GitHub, Jubatus, https://github.com/jubatus/jubatus
[4] K. Horikawa, Y. Kitayama, S. Oda, H. Kumazaki, J. Han, H. Makino, M. Ishii, K. Aoya, M. Luo, and S. Uchikawa, "Jubatus in Action: Report on Realtime Big Data Analysis by Jubatus," NTT Technical Review, Vol. 10, No. 12, 2012. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201212fa5.html
[5] NTT press release issued on February 18, 2015 (in Japanese). http://www.ntt.co.jp/news2015/1502/150218a.html
[6] Website of NTT-IT, WatchBee (in Japanese). http://www.ntt-it.co.jp/product/watchbee/
[7] K. Takagahara, K. Ono, N. Oda, and T. Teshigawara, ""hitoe"—A Wearable Sensor Developed through Cross-industrial Collaboration," NTT Technical Review, Vol. 12, No. 9, 2014. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201409ra1.html
[8] Website of Deep Learning, http://deeplearning.net/

Takashi Hayashi

Senior Research Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.

He received a B.E. and M.E. in electrical engineering from Keio University, Kanagawa, in 1997 and 1999. He joined NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories) in 1999, where he researched database management system (DBMS) technology. From 2003 to 2010, he developed video conference systems at NTT Bizlink. He is currently engaged in research and development (R&D) of a scalable distributed computing framework for big data. He is a member of the Information Processing Society of Japan.

Masayoshi Umeda

Senior Research Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.

He received a B.E. in electrical engineering from the University of Electro-Communications, Tokyo, in 1991. He joined NTT Information Telecommunication Networks Laboratory in 1991 and researched and developed DBMS technology. He is currently working on the development of Jubatus.

Masato Sawada

Senior Research Engineer, NTT Software Innovation Center.

He received an M.E. in information and computer sciences from Osaka University in 2000. He joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 2000. He has been researching and developing a large-scale full text search engine. He is also currently involved in the R&D of data processing software using machine learning.

Kaori Isagai

Senior Research Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.

She received a B.E. and M.S. in information engineering from Nagoya University, Aichi, in 1994 and 1996. She joined NTT Multimedia Network Laboratories in 1996 and engaged in R&D of large scale Internet protocol networks. From 2011 to 2015, she developed session control systems for NGN (Next Generation Network) at NTT EAST. Since moving to NTT Software Innovation Center in 2015, she has been incubating and developing the market for Jubatus.

Akihiro Yamanaka

Engineer, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.

He received an M.S. in mathematics from Waseda University, Tokyo, in 2010. He joined NTT in 2010 and is involved in data analysis.

Mitsuaki Tsunakawa

Senior Research Engineer, Supervisor, Distributed Data Processing Platform SE Project, NTT Software Innovation Center.

He received a B.S. in mathematics from Tsukuba University, Ibaraki, in 1990 and joined NTT Communications and Information Processing Laboratories the same year. He has been researching DBMS technology and the integration of heterogeneous information sources. He is currently engaged in R&D of a scalable distributed computing framework for real-time analysis of big data. He is a member of the Institute of Electronics, Information and Communication Engineers.

Advanced Processing and Analytics for Large-scale Graphs

Yasuhiro Iida, Yasunari Kishimoto, Yasuhiro Fujiwara, Hiroaki Shiokawa, Junya Arai, and Sotetsu Iwamura

Abstract

Expectations have been rising in recent years for capabilities to extract hidden knowledge in friendship relationships, product purchase relationships, and business transaction relationships for use in promoting sales, increasing work efficiency, and other such purposes. A graph is a data structure that represents data relationships. Data mining on graph data is called graph mining. Here, we describe advanced graph mining techniques that can be used to instantly discover hidden knowledge in large-scale graphs which lead to improved work efficiency through high-speed graph processing. We also present examples of graph mining applications.

Keywords: graph mining, clustering, label propagation

1. Introduction

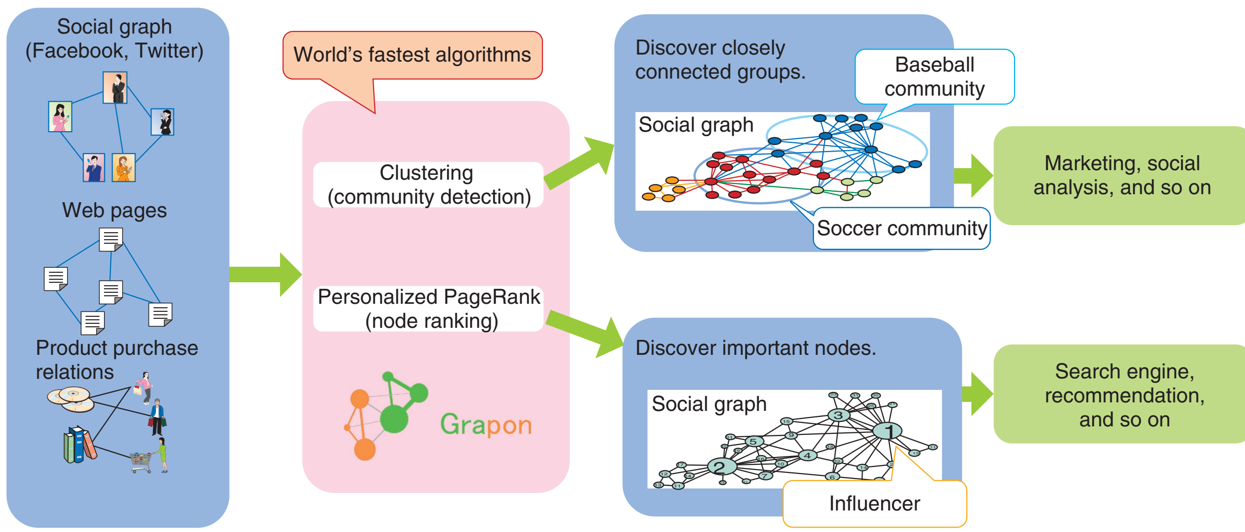
With the explosive growth of social media such as social networking services (SNSs) over the last few years, the data structures of big data have extended beyond the conventional simple table format. They now include graphs, which can represent the relationships among many kinds of information, including people, objects, and places. In a graph structure, data are represented as nodes, and the relationships between data are represented as edges. Graph structures can represent the relationships of many kinds of data that we see all around us, including the web pages and links on the Internet, the friendship relationships in SNSs, product and purchase relationships, and the relationships of roads and intersections.

The research and development (R&D) on graph technology being done at NTT laboratories includes the grouping of similar data (nodes) in large-scale graphs, the discovery of important nodes, and the suitable partitioning of large-scale graphs for handling on multiple machines. By utilizing these technologies in cooperation with other companies and academic society, we aim to produce new and more practical technology.

2. Graph mining technology for efficient analysis of large-scale graphs

Graph mining refers to the process of discovering hidden relationships in a graph such as a group of people who have strong friendship relationships or a group of people who have a high degree of influence in a certain area. Research on graph mining methods and applications has been increasing in recent years, and we are beginning to see uses in the business world as well. However, a huge amount of processing time is required when working with large graphs on the scale of the population of Japan. For such large-scale graphs, efficient graph mining that involves trial and error is not possible.

To address this problem, NTT laboratories have developed *Grapon*, a graph mining engine that is capable of performing clustering and ranking of large-scale graphs at high-speed (**Fig. 1**). Clustering (community detection) can be used to efficiently discover groups of people or things that have strong relationships in a graph. Such information can be used in various ways such as in developing marketing and product sales strategies. Personalized PageRank (node ranking) can be used to instantly rank people or things by importance according to how they are



*Facebook is a registered trademark of Facebook, Inc. in the United Sates and other countries.
Twitter is a registered trademark of Twitter, Inc. in the United Sates and other countries.

Fig. 1. Grapon, NTT's graph mining engine.

related. Such information is useful for making product recommendations.

Grapon implements an algorithm that we devised and achieves the world's highest-speed graph processing. For clustering analysis in a graph on a scale of about 100 million nodes, for example, our algorithm reduces the processing time from several hours to about three minutes [1, 2]. We are currently moving forward with R&D to enable more kinds of data analysis and to develop even more advanced graph mining technology. In this article, we describe our most recent work, including techniques for partitioning graphs into clusters of equal granularity at high speed, techniques for efficiently discovering nodes that serve as mediators by connecting multiple groups in a graph, and techniques for instantly inferring node category membership.

3. Partitioning graphs using equal granularity

Graph partitioning is an important technique with large-scale graphs since such graphs should be stored in multiple storage systems, and they are computed in a parallel manner by using multiple machines. We have developed a novel graph partitioning algorithm called *balanced granularity partitioning* that divides graphs into same-sized partitions without sacrificing clustering quality (clustering property) [3]. Maintaining the clustering property means that strongly connected nodes are not split apart but are put into the same cluster. Also, if the clusters are formed so they have about the same size, weights can be assigned to each edge, and the sum of the edge weights can be used as a basis for partitioning.

The balanced granularity partitioning first generates a lot of small-sized clusters as an intermediate step. Those small clusters are then merged into the same partition in order to build final partitions that have the same cluster size (**Fig. 2(a)**). For example, we can apply our balanced granularity partitioning for traffic congestion prediction by using road networks. In this case, roads and their intersections are respectively represented as nodes and edges in a graph. In addition, we can also represent the amount of traffic on a road by using a weighted edge in the graphs. For example, if there is a road network in an urban area, we can expect its traffic conditions to change dramatically from time to time. Obviously, if the traffic congestion prediction and simulation system requires a large runtime, the system will not make sense to users. Hence, to predict traffic conditions in near real time, it is important to reduce the prediction and simulation runtime by using parallel computing on multiple machines and to use graph partitioning techniques that balance the workload across the machines.

The equal-granularity clustering technique enables sufficiently fast parallel processing of simulations by

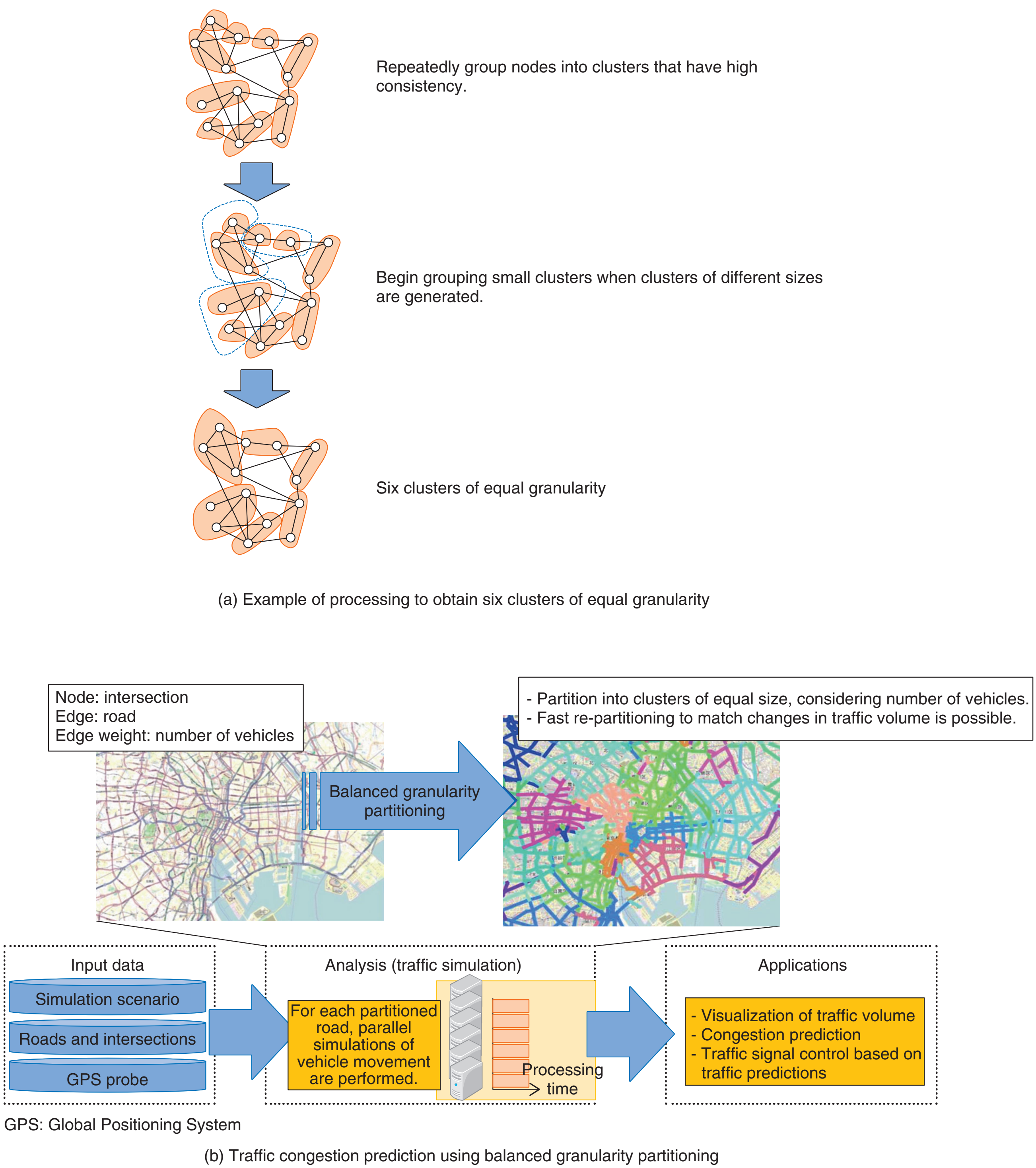


Fig. 2. Partitioning graphs using equal granularity.

fast partitioning of the graph to match the number of parallel simulations, and it evens out the processing load by producing partitions of equal granularity (**Fig. 2(b)**).

We evaluated our equal-granularity clustering technique in a traffic simulator done in collaboration with

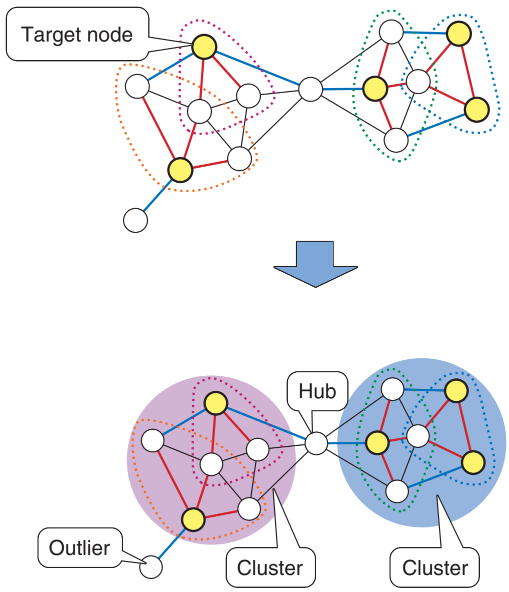


Fig. 3. Discovery of mediators.

NTT DATA Corporation, and we confirmed that our technique ran at least 10 times faster than METIS, the de-facto standard open-source graph partitioning software. Furthermore, we also experimentally verified that the equal-granularity clustering technique produced well-sized balanced partitions for large-scale graphs. NTT DATA Corporation is currently working on the application of this technique by conducting demonstration experiments for various use cases. Our technique is not limited to the above congestion prediction; we can also apply it to traffic volume visualization, traffic signal control, and other such purposes.

4. Efficient discovery of persons or things that serve as mediators

In recent years, techniques have been developed for finer graph analysis that goes beyond the discovery of groups of strongly connected nodes. One of the most important techniques is the Structural Clustering Algorithm for Networks (SCAN). SCAN is a graph clustering technique that tries to find clusters based on node similarities (i.e., structural similarities). Unlike traditional graph clustering techniques, SCAN finds not only clusters but also hubs, which are nodes that bridge multiple clusters, and outliers, which are noise components included in a graph. SCAN regards nodes that are not included in clusters as hubs or outliers. If a non-clustered node bridges multiple clusters, SCAN determines the node to be a hub; otherwise the node is an outlier. As mentioned previously, SCAN provides good clustering results; however, it requires substantial computation time to find all clusters, hubs, and outliers in large-scale graphs since SCAN has to iteratively compute all edges in the graph.

Hence, we developed a novel structural clustering algorithm that can analyze data at a practical high speed, even with large-scale graphs [4]. Specifically, our approach forms clusters of only some of the nodes rather than evaluating cluster membership for all nodes. If the clustering process produces clusters that intersect, the clusters are combined into a single cluster, and any nodes that do not belong to either cluster are regarded as either hubs or outliers (**Fig. 3**). By taking this approach, we can increase the processing speed by a factor of 20 compared to the conventional method without loss of accuracy. Discovery of clusters, hubs, and outliers at the same time takes more runtime than simple cluster analysis, but this large increase in speed makes it possible to perform the processing for a graph with hundreds of thousands of nodes in a matter of seconds.

If we apply this technique to friend relationships, for example, a person who does not belong to any group is classified as a hub. As a mediator, that person is considered to contribute to multiple groups. Hence,

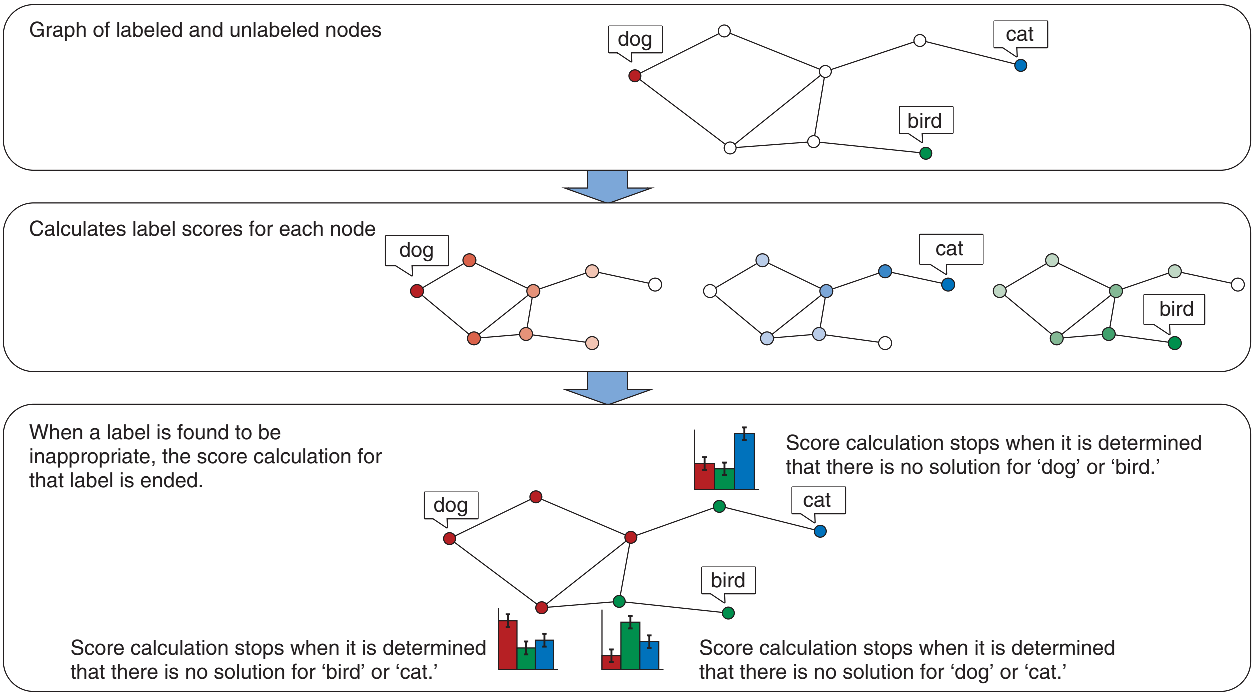


Fig. 4. Categorizing.

he/she might be expected to spread information about products to multiple groups and to serve as a route for conveying information from one group to another. In addition, since persons who are classified as hubs do not belong to any particular group, they can also be considered as candidates for work that requires a neutral position such as evaluation or investigation tasks. Another example is analyzing a graph of the transaction relationships among companies to identify companies that could serve as mediators. This would provide useful information that could be used in business decisions such as selecting companies that have influence in multiple industries for commercial cooperation. The existence of such possible applications has motivated us to continue evaluating the applicability of this technology.

5. Efficient categorizing of people and objects

One important task in data analysis is categorizing. Products can be classified according to customer interests and preferences such as health-consciousness or a preference for upscale products. News articles can be classified by subject such as politics or entertainment. One node classification technique for graph mining is label propagation. Each node in the graph is assigned label information, and nodes that have different labels are regarded as belonging to different categories.

Categorizing based on label propagation involves assigning labels to unlabeled nodes when given a small number of labeled nodes. Consider a graph that relates documents by similarity and has nodes that are labeled *dog*, *cat*, and *bird*, for example. The unlabeled nodes can be assigned labels according to their similarity to the labeled nodes. To do this, a score is calculated for each label, and the label with the highest score is applied to that node. A problem with this calculation is that the calculation may be recursive, depending on the structure of the graph, so the computational cost can be very high. We have developed a fast label propagation technique that can analyze large-scale graphs at a practical high speed [5].

Conventional label propagation involves evaluating each label to determine if it is the correct label, so the computational load is very high. Our approach is to halt the calculation of label scores when it is determined that a particular label is clearly not correct (**Fig. 4**). By doing so, we achieved an increase in speed by a factor of from 2 to 400 relative to the conventional technique. The processing for node categorization by label propagation usually takes longer

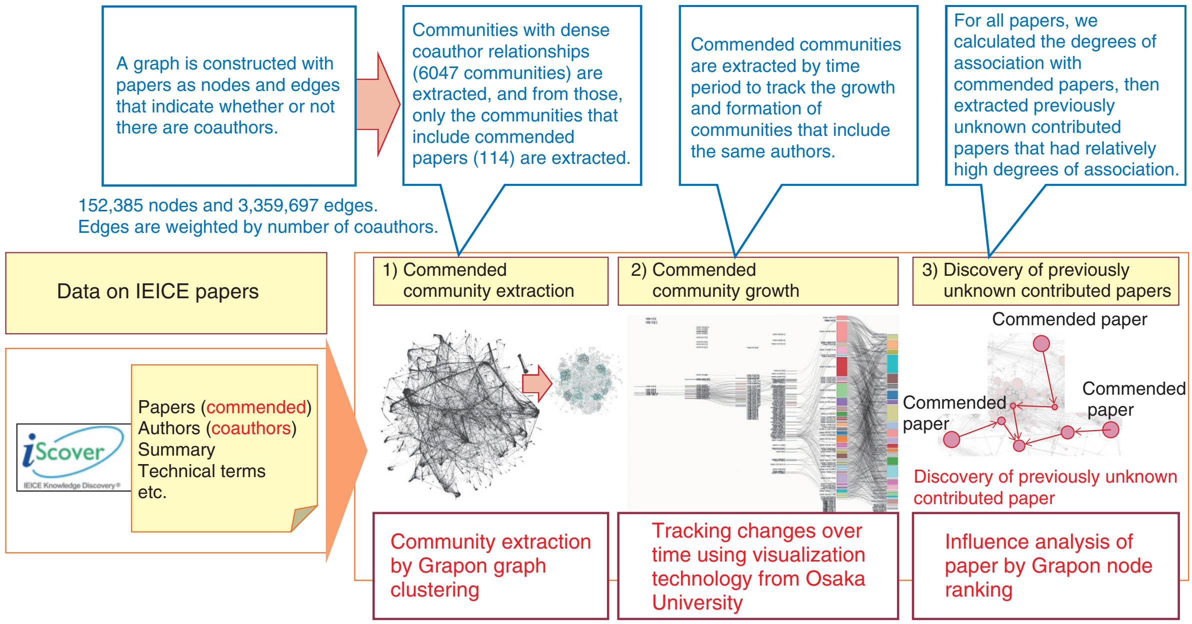


Fig. 5. Discovery of community trends and previously unknown contributed papers.

than cluster processing, but the increase in speed that we have achieved makes it possible to perform the processing for a graph with tens of thousands of nodes in just a few seconds.

Potential applications of this technique include classification of customers in a graph of purchase relationships and co-marketing of products that belong to the same category. Other possibilities include categorizing companies that are in the supply chain of a corporate group by capital grouping in order to make business decisions such as choosing companies for collaboration or accepting companies as clients or customers. We are currently evaluating the applicability of this technique.

6. Verification of the utility of the technology

We are moving forward with R&D to verify and refine the technology and to derive new research topics by applying Grapon in various fields. An important example is our participation in the data analysis competition sponsored by the Joint Association Study Group of Management Science (JASMAC), a gathering of over 600 data analysis experts from academic and corporate organizations, in which we conducted trials that involved applying Grapon in the field of marketing. Those trials confirmed the utility of Grapon for high-speed extraction of products for recommendation and discovery of product combinations that are suitable for cross-selling from large-scale POS (point of sales) data.

Another example is our participation together with Osaka University in the I-Scover Challenge 2014, which involved deriving useful knowledge from the textual data of over 160,000 research papers of the Institute of Electronics, Information and Communication Engineers (IEICE). Our results in that event confirmed that Grapon was effective for recognizing trends in communities that produce commended papers from the co-authorship relationships among several tens of thousands of authors and for discovering previously unknown contributed papers, and that it analyzed such large-scale data at high speed (**Fig. 5**).

7. Toward analysis of data on larger scale

We are witnessing the ever-growing scale of graph-structured data; for example, the total number of web pages has reached a scale of billions of nodes. To deal with this situation, we have begun studies on parallel processing techniques for efficient graph mining on multiple CPUs (central processing units) rather than on a single powerful CPU. For example, data layout optimization by reassigning node identification numbers promotes the use of data within the CPU cache memory. This optimization allows us to obtain maximum benefit from parallel processing by reducing the amount of communication between CPUs as well as achieving faster operation of single CPUs. Although there are open source tools for parallel graph mining, for example, GraphLab, Grapon achieves 60 times better performance than GraphLab by using the data layout optimization [6, 7]. We will continue our R&D on more scalable parallel processing on multiple CPUs to cope with the increasing graph scale.

8. Future developments

We have briefly described our R&D on graph mining technology to achieve more efficient and more complex analysis of data that continues to increase in scale. In the future, we will continue to refine the graph mining techniques and increase their ease-of-use by applying our technology in various fields with the objectives of creating new value and contributing to a smart society.

References

[1] NTT press release issued on February 13, 2013 (in Japanese). http://www.ntt.co.jp/news2013/1302/130213b.html

[2] Y. Iida, Y. Kishimoto, Y. Fujiwara, H. Shiokawa, and M. Onizuka, "Finding Communities and Ranking in Large-scale Graphs: Fast Algorithms and Applications," Journal of JSAI, Vol. 29, No. 5, pp. 472–479, 2014.

[3] T. Fujimori, H. Shiokawa, and M. Onizuka, "Optimization of Graph Partitioning for Distributed Graph Processing," Proc. of DEIM Forum 2015 (the 7th Forum on Data Engineering and Information Management), E5-2, Fukushima, Japan, 2015.

[4] H. Shiokawa, Y. Fujiwara, and M. Onizuka, "SCAN++: Efficient Algorithm for Finding Clusters, Hubs and Outliers on Large-scale Graphs," The Proceedings of the VLDB Endowment (PVLDB), Vol. 8, No. 11, pp. 1178–1189, 2015.

[5] Y. Fujiwara and G. Irie, "Efficient Label Propagation," Proc. of ICML 2014 (the 31st International Conference on Machine Learning), pp. 784–792, Beijing, China, 2014.

[6] J. Arai, H. Shiokawa, T. Yamamuro, and M. Onizuki, "Scalable Parallel Graph Processing by Optimized Vertex Order," DEIM Forum 2015, E5-3, Fukushima, Japan, 2015.

[7] J. Arai, H. Shiokawa, T. Yamamuro, M. Onizuki, and S. Iwamura, "Rabbit Order: Just-in-time Parallel Reordering for Fast Graph Analysis," Proc. of IPDPS 2016 (the 30th IEEE International Parallel and Distributed Processing Symposium), Chicago, USA, to appear.

Yasuhiro Iida

Senior Research Engineer, Distributed Computing Technology Project, NTT Software Innovation Center.

He received a B.S. and M.S. in applied physics engineering from the University of Tokyo in 1998 and 2000. In 2000, he joined NTT Information Sharing Platform Laboratories. His recent research area is data mining and data management. He is a member of the Association for Computing Machinery (ACM).

Hiroaki Shiokawa

Researcher, Distributed Computing Technology Project, NTT Software Innovation Center.

He received a B.S. in information science and an M.E. and Ph.D. in engineering from University of Tsukuba in 2009, 2011, and 2015. He joined NTT in 2011 and has been studying graph data management, graph mining algorithms, distributed computing, and databases. He is a professional member of ACM.

Yasunari Kishimoto

Research Engineer, Distributed Computing Technology Project, NTT Software Innovation Center.

He received a B.E. and M.E. from Kyushu University, Fukuoka, in 1989 and 1991. He joined NTT in 1991 and studied directory systems, billing systems, and data mining. He is a member of the Information Processing Society of Japan (IPSJ).

Junya Arai

Researcher, Distributed Computing Technology Project, NTT Software Innovation Center.

He received a B.S. and a Master of Information Science and Technology degree from the University of Tokyo in 2011 and 2013. He joined NTT in 2013 and has been studying parallel distributed processing and graph mining algorithms. He is a member of ACM and DBSJ.

Yasuhiro Fujiwara

Distinguished Technical Member, Distributed Computing Technology Project, NTT Software Innovation Center.

He received a B.E. and M.E. from Waseda University, Tokyo, in 2001 and 2003, and a Ph.D. from the University of Tokyo in 2012. He joined NTT in 2003. His research interests include data mining, databases, natural language processing, and artificial intelligence. He is a member of IPSJ, IEICE, and the Database Society of Japan (DBSJ).

Sotetsu Iwamura

Senior Research Engineer, Supervisor, Distributed Computing Technology Project, NTT Software Innovation Center.

He received a B.S., M.S., and Ph.D. in electronic engineering from the University of Tokyo in 1989, 1991, and 1994. He joined NTT Telecommunication Networks Laboratories in 1994. Since then, he has been researching mobile computing and high-speed computer networks based on asynchronous transfer mode (ATM), mobile computing, as well as an in-house cloud computing platform for NTT R&D. He is currently managing a research group on big data processing, such as fast data-mining algorithm, parallel-distributed processing architecture. He is a member of IPSJ.

Improving Network Management and Operation with Machine Learning and Data Analytics

Keisuke Ishibashi, Takanori Hayashi, and Kohei Shiomoto

Abstract

In this article, we describe the current and future direction of our research on the use of machine learning and data analytics to solve network management and operation problems. Specifically, we introduce ways to predict or detect network failures using social networking services or network logs, ways to extract workflows using operator logs, and methods to predict mobile traffic by investigating traffic generation factors.

Keywords: network failure, self-operation, traffic analysis

1. Introduction

We are conducting research on methods to solve network management and operation problems using machine learning and data analytics. Machine learning can extract latent rules and generation models of network behavior through big data analytics, and those rules and models can be used for predicting and optimizing network management and operation, which improves network service and reduces costs. In this article, we introduce methods to predict and detect network failures using a social networking service (SNS) or network logs, extract workflows from operator logs, and predict mobile traffic from traffic generation factors.

2. Predicting and detecting network failures

To minimize the impact on services caused by a network failure, we must detect a failure, or a fault that leads to a failure, before it occurs or as soon as possible after it occurs. Current failure detection methods are rule based, where rules between network logs and failure modes are set manually in advance. By trapping a log, we can detect a corresponding network failure. However, much progress has been made recently in software-based network function virtualization, which means that the network configuration is dynamically changing. Consequently, building those rules and updating them are difficult and time-consuming tasks. In addition, because of the increasing network complexity and the growing number of network roles, the monitoring information (network logs) obtained using existing methods is insufficient for monitoring the network status. To solve this problem, we are working on enhancing network monitoring and improving the accuracy of failure detection.

2.1 Enhancing monitoring objects

Current network monitoring techniques are based on the use of internal network data such as network logs and external network data such as service monitoring and user claims. To improve the coverage and agility of monitoring, we are implementing failure detection based on SNS data. However, because SNS data consist of free-format text messages and include a huge number of messages other than those related to network outages, we have developed a technique to accurately extract messages that correspond to

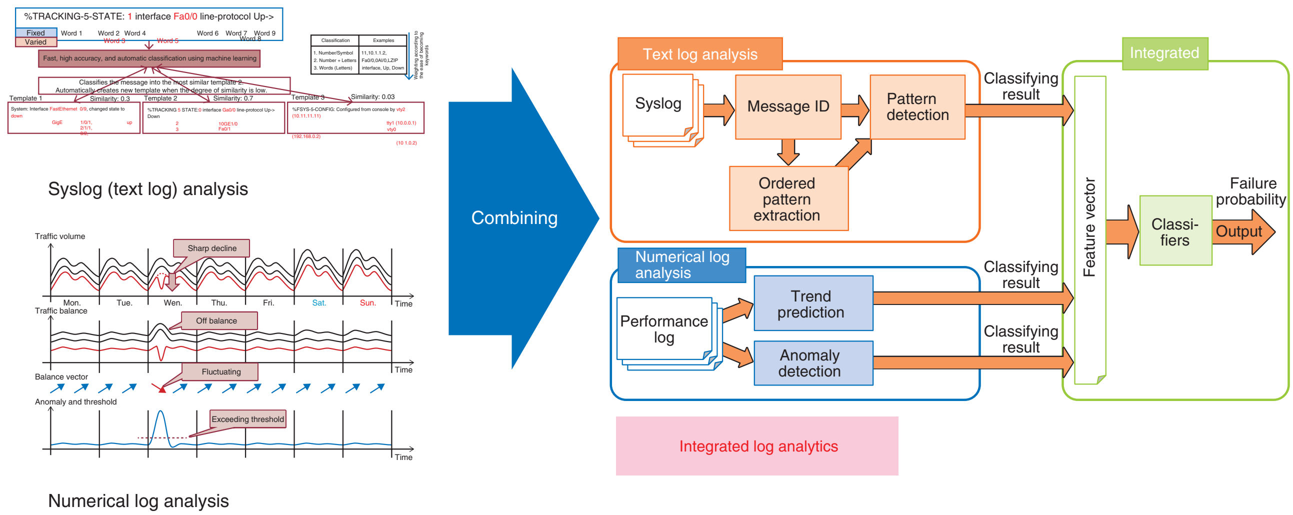


Fig. 1. Log analysis technique.

network failures. In addition, by estimating the locations of such messages, we can estimate the location and the impact of the network failure.

Network logs that are used for network monitoring are divided into numerical logs such as CPU (central processing unit) and interface loads, and messages in text format such as syslog messages. Numerical logs are used for detecting network failures by applying predefined thresholds. However, defining thresholds for a huge number of logs is difficult. We apply statistical outlier detection based on a non-supervised machine learning method. In addition, we apply this method to obtain the time series of each log as well as the correlations among logs.

Text logs are used for failure detection by monitoring keywords. However, this involves looking at the relations between keywords. The network status is not directly monitored, and therefore, the accuracy needs to be improved. We adopt a machine learning technique such as clustering to monitor text logs that are not only keyword based but also message based and that generate patterns. We also try to merge both numerical logs and text logs to detect network failures that are not detected for them individually. (**Fig. 1**).

2.2 Improving detection accuracy

Adopting a statistical method for detection as described in the previous section can potentially result in a *miss* detection (false positive) or detection error (false negative). In addition, the current supervised machine learning method requires both positive (failure mode) and negative (normal mode) samples, but network failures seldom occur and positive samples are not easily obtained. Here, we focus on the partial area under ROC (receiver operating characteristic) curve (pAUC), which is the index for a balance between false positives and negatives, and we try to adopt a technique that directly optimizes pAUC [1].

3. Extracting operation workflow

When network operators resolve network failures, if the resolution processes are not fixed and no manuals are available, operators must take action based on their knowledge, which of course depends on their experience. This increases the time needed to resolve the failure (the time-to-fix), especially for non-skilled operators. This in turn increases the need for Runbook Automation (RBA), which enables auto-operation in the event of network failures. However, building a workflow (scenario) to be used for RBA is a time-consuming task.

To solve this problem, we take two approaches to extract and visualize workflows in resolving network failures.

First, we develop a technique to extract workflows using a trouble ticket log where operators manually record the processes they carry out from the time a failure starts to when the problem has been resolved. Within a trouble ticket, operators record resolution processes, which provide useful information for building workflows for the processes. However, these records consist of free-format text data, which means

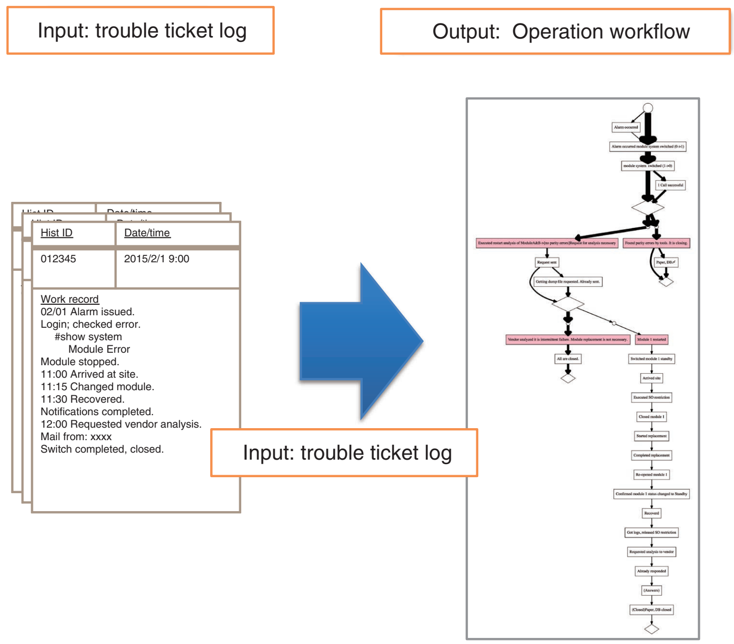


Fig. 2. Extracting operation workflow using trouble ticket log.

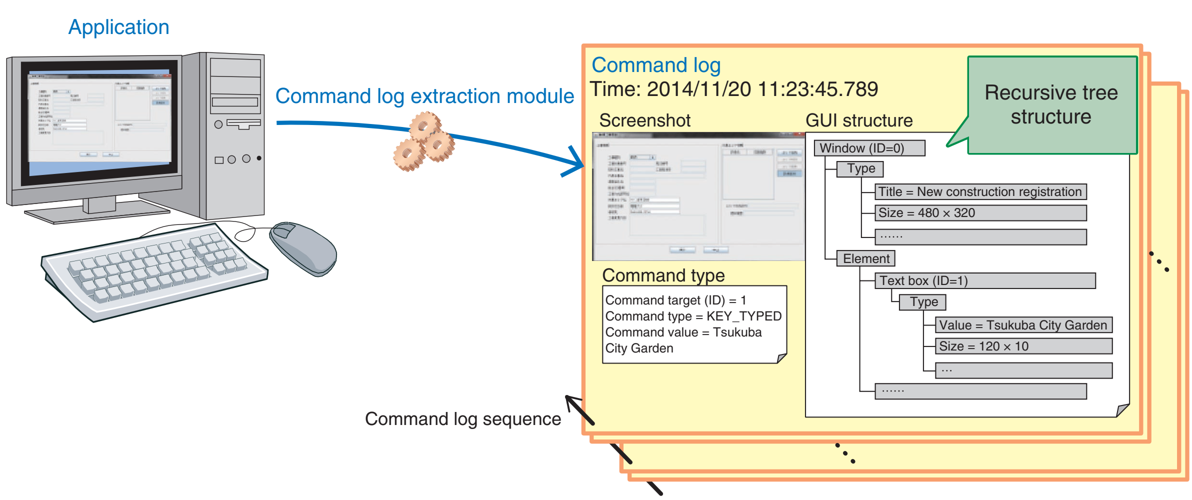


Fig. 3. Extracting operation workflow using GUI.

that the same process can be recorded using different words, and some processes may not be recorded. Therefore, we adopt a sequence alignment technique to adjust and complement these records (**Fig. 2**).

However, some processes, specifically the initial processes implemented for critical failures, tend not to be recorded because the actions are taken prior to recording them. To extract workflows for such processes, we should not rely on trouble ticket logs, but rather use command logs. To extract command logs for graphical user interface (GUI) applications, we have developed a technique to independently build GUI command log sequences on applications (**Fig. 3**).

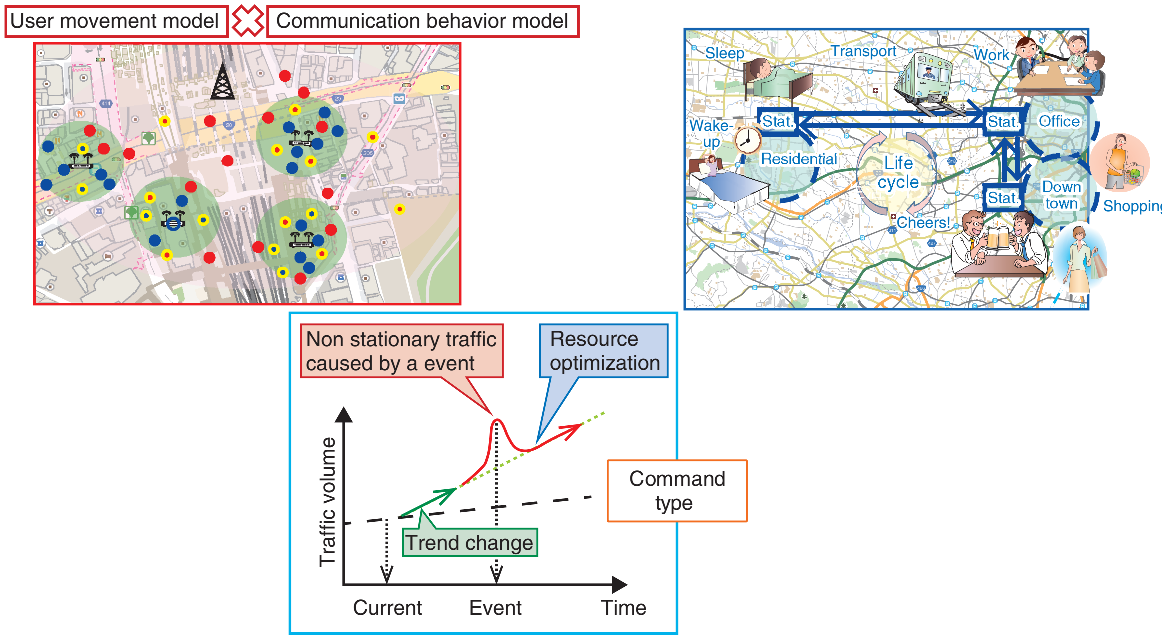


Fig. 4. High accuracy traffic prediction.

4. Traffic prediction through its generation process

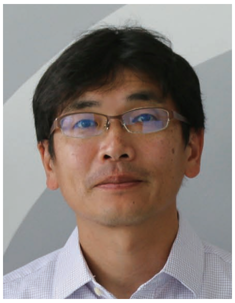
Traffic prediction is based on using past values to extrapolate future values. However, the drawback of this method is that it cannot adapt to changes in the traffic generation mechanism such as application usage or popular content. Specifically, mobile traffic is critically affected by changes in human movement such as those occurring during sports or music events. We are developing a technique to predict future traffic based on not only past traffic values but also traffic generation mechanisms such as human movement patterns and application usages (**Fig. 4**). We also adopt a method for long-term future traffic prediction by analyzing and feeding back prediction errors.

5. Future direction

In this article, we described our research that focuses on solving problems related to network management and operation by applying machine learning and data analytics. We plan to move forward with this research from analyzing and predicting the network status to optimizing it.

Reference

[1] O. Komori and S. Eguchi, "A Boosting Method for Maximizing the Partial Area under the ROC Curve," BMC Bioinformatics, Vol. 11, pp. 314–330, 2010.



Keisuke Ishibashi

Senior Research Engineer, Supervisor, NTT Network Technology Laboratories.

He received a B.S. and M.S. in mathematics from Tohoku University, Miyagi, in 1993 and 1995, and a Ph.D. in information science and technology from the University of Tokyo in 2005. Since joining NTT in 1995, he has been researching traffic issues in computer communication networks. He received the Young Researcher's Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2002, the Information Network Research Award in 2002 and 2010, and the Internet Architecture Research Award in 2009. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), IEICE, and the Operations Research Society of Japan.

Takanori Hayashi

Senior Research Engineer, Communication Traffic & Service Quality Project, NTT Service Integration Laboratories.

He received his B.E., M.E., and Ph.D. in engineering from the University of Tsukuba, Ibaraki, in 1988, 1990, and 2007. Since joining NTT in 1990, he has been conducting research on subjective quality assessment of multimedia telecommunications and network performance measurement methods. He is currently working on a multimodal quality assessment method over IP networks. He is a member of IEICE.

Kohei Shiomoto

Senior Manager of Communication & Traffic Service Quality Project, NTT Network Technology Laboratories.

He received his B.E., M.E., and Ph.D. in information and computer sciences from Osaka University in 1987, 1989, and 1998. He joined NTT in 1989 and began researching asynchronous transfer mode (ATM) traffic control and ATM switching system architecture design. During 1996–1997, he was a Visiting Scholar at Washington University in St. Louis, MO, USA. During 1997–2001, he directed architecture design for the high-speed IP/MPLS (multiprotocol label switching) label switching router research project at NTT Network Service Systems Laboratories. He also conducted research on photonic IP router design and routing algorithms, and Generalized MPLS (GMPLS) routing and signaling standardization at NTT Network Innovation Laboratories and then at NTT Network Service Systems Laboratories, and he was involved in GMPLS standardization in the Internet Engineering Task Force. He led the IP Optical Networking Research Group in NTT Network Service Systems Laboratories from April 2006 to June 2011 and the traffic engineering research group at NTT Service Integration Laboratories from July 2011 to June 2012. He has been in his present position since July 2012. He has chaired various committees of the Asia-Pacific Board of the IEEE Communications Society and served as Secretary for International Relations of the Communications Society of IEICE. He has also been involved in organizing several international conferences including MPLS, iPOP, and WTC. He received the Young Engineer Award from IEICE in 1995 and the Switching System Research Award from IEICE in 1995 and 2000. He is a fellow of IEICE and a member of IEEE and the Association for Computing Machinery.

Advanced, High-performance Big Data Technology and Trial Validation

Ken-ichi Noguchi, Yoshihide Sato, and Hisako Shiohara

Abstract

Tourism to Japan is expected to increase as the year 2020 approaches. In every possible scenario, people will need access to stable transportation and communication services as well as other social infrastructure. NTT is researching and developing solutions to meet these requirements by using big data technologies. In this article, we present one such advanced, high-performance big data technology, and we describe the results of a field test (the Fukuoka trial) of services for tourists visiting Japan.

Keywords: big data, clustering, prediction

1. Introduction

With Tokyo chosen as the host for the 2020 Summer Olympic and Paralympic Games, it will become important to stabilize communication networks while understanding and relieving congestion in the areas around such massive events. Similarly, a stable societal infrastructure is also needed in order to deal with any large-scale disasters that might occur. We can expect to be able to address these issues (i.e., understand and relieve congested conditions) through the application of big data technologies using location data from now-ubiquitous smartphones and car navigation systems.

However, NTT's big data methodology does not simply involve analyzing location data independently. Our research and development (R&D) efforts are instead aimed at discovering insights that could not be gleaned from a single category of data alone. To do this, we analyze several disparate categories of data together; the data include a variety of user attributes and behaviors as well as local geographical and weather data. We believe that this type of analysis will allow us to provide appropriate information based on various factors such as location data, user attributes, time of day, and weather during events in the near future. Furthermore, we expect to be able to apply this information to guidance services and to services that control said guidance by predicting changes in both human and network flows from a variety of factors in the event of a disaster.

In this article, we present an advanced, high-performance big data technology we have developed in order to analyze data across a wide variety of categories, particularly location data. We first explain the technology's overarching structure along with each analysis technique. Then we present our analysis of a sample trial we conducted with overseas tourists visiting Fukuoka City.

2. Advanced, high-performance big data technology

We developed our advanced, high-performance big data technology primarily to analyze the flow of people, a process that involves examining location data and other types of activity logs (**Fig. 1**).

In recent years, technologies such as GPS (Global Positioning System) have made it easy to obtain the geographic coordinates of smartphones and other portable computing devices. However, it has been difficult to analyze user activity such as that related to sightseeing and lodging via ordinary statistical techniques (e.g., sums and averages) on this kind of movement tracking data, which represent a massive time series. Our visit identification techniques can solve this problem by predicting where people have visited or stayed from movement tracking data that

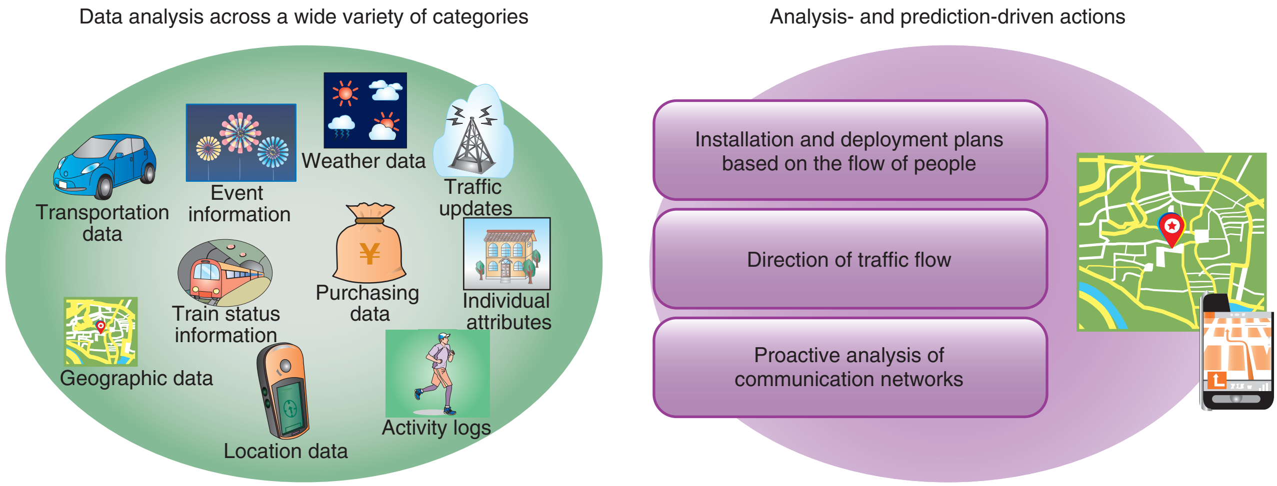


Fig. 1. Overview of advanced, high-performance big data technology.

contain many measurement errors, noise, and missing values.

At the core of our advanced, high-performance big data technology is our multidimensional composite data (cluster) analysis, which can collectively analyze data across a variety of categories, including user attributes (e.g., age and residence), weather data, and predicted visits.

Furthermore, we have prepared real-time predictive crowd counting techniques that can analyze users' location data in real time to predict the size and movements of crowds of people several hours in the future.

Our advanced, high-performance big data technology is thus characterized by their ability to use three core components together on a single platform (**Fig. 2**). By providing a single platform for integrated data analysis with a simple user interface, shared data preprocessing, visualization of prediction results, and other features, we can reduce the amount of work that must be done by data analysts.

2.1 Visit identification

When analyzing user movement tracking data, we first need to convert it into information that tells us how long users spent at any given location, taking into account the effects of noise caused by measurement errors in any location data obtained from smartphones. NTT Service Evolution Laboratories has researched and developed a visit identification technique that estimates both the locations people have visited as well as their actual destinations (e.g., parks and cafés) from movement tracking data with overlaid noise. We are attempting to improve the accuracy of our identifications by taking individual preferences and time spent at any given location into account rather than relying on physical proximity alone. Furthermore, our technique takes advantage of parallelization to allow it to quickly process the movement tracking data for many different people in a short amount of time.

2.2 Multidimensional composite data analysis

This clustering technique is at the heart of analyzing data together across many diverse categories; it is based on non-negative multiple tensor factorization (NMTF), which is explained in detail in this issue's feature article, "From Multidimensional Mixture Data Analysis to Spatio-temporal Multidimensional Collective Data Analysis" [1]. Although an analysis of many different categories of data at the same time will take place in an extremely large (sparse) data space, NMTF reduces computational costs and allows data to be analyzed more efficiently.

Clustering results are characterized by a combination of several factors when they are visualized or semantically interpreted. For example, one such result might be that during weekends, most café patrons in the western districts are women in their 30s from the eastern districts. This allows us to discover properties that we could not have found through simple combinations.

The actual data analysis work requires us to repeatedly change parameters (e.g., input data weights and

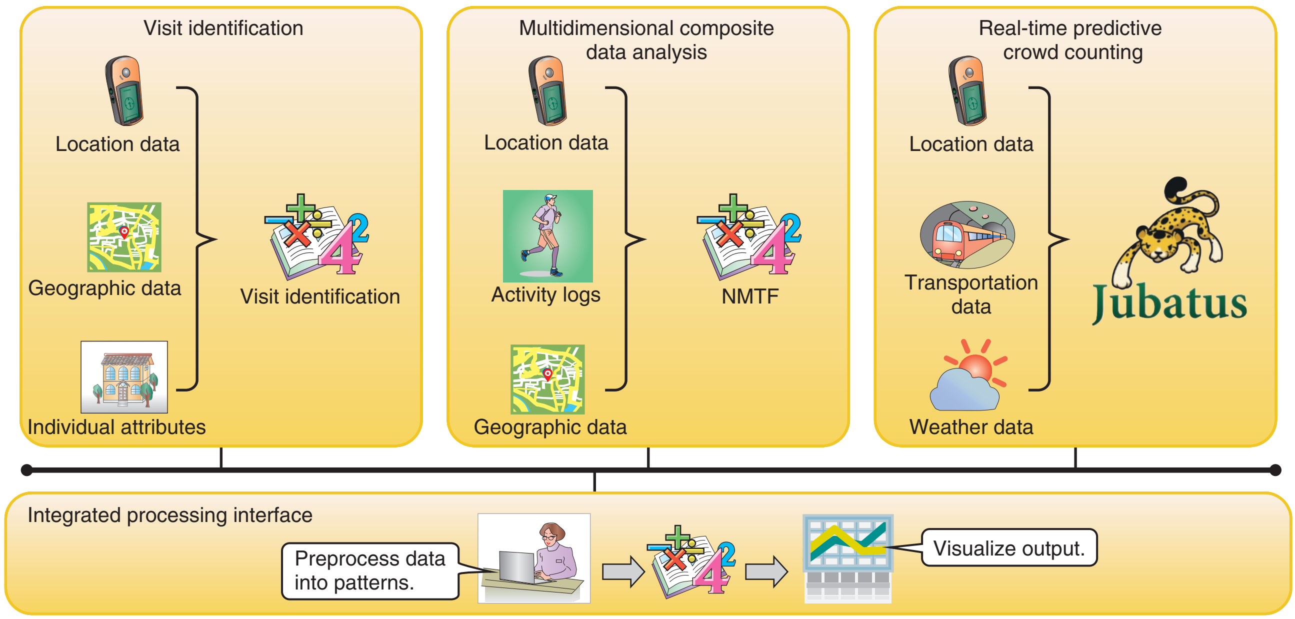


Fig. 2. System structure.

cluster counts) and re-analyze the data while interpreting the results. A graphical user interface (GUI) allows anyone to run our multidimensional composite data analysis even without a detailed understanding of the underlying algorithms.

2.3 Real-time predictive crowd counting

This technique analyzes location data in real time to estimate how many people are or will be in a given area. Activity logs and other location data are constantly being updated. To keep up with this flow of data, we need to add the latest updates to our accumulated time series and analyze them in real time. Our implementation of this real-time analysis uses Jubatus [2], a distributed processing framework for online machine learning that is also presented in one of this issue's feature articles [3].

To visualize the analyzed data, we use a GUI with a heat map overlaid on a geographic map. The actual process of analyzing the data involves tuning our machine learning models on the basis of a side-by-side comparison of the actual and expected results. We expect this to allow us to make more accurate predictions of the number of people in any given area. For example, we believe that we can apply this technique to the constantly shifting flows of people during an event to predict how many people with particular attributes will be in each area and thus anticipate where to post personnel.

3. Trial application with Fukuoka tourism

To validate the effectiveness of our techniques, we participated in a field trial intended for tourists visiting Japan from overseas. We present here the results of analyzing actual data from the trial.

During the trial, we distributed a smartphone application (or tourist app) to tourists visiting Fukuoka City and its surrounding tourist destinations. In addition to connecting to free Wi-Fi hotspots, the app provided useful services such as tourist information and coupons relevant to its user's activities and circumstances. With the tourists' advance consent, we collected the app's usage logs, location data, and individual characteristics such as gender, year of birth, and country of residence.

By analyzing such data with our advanced, high-performance big data technology, we were able to find patterns in tourist trips and activities, estimate how many tourists would visit individual areas, and explore other avenues of inquiry.

3.1 Analyzing patterns in trips and activities

We collected movement tracking data in the form of

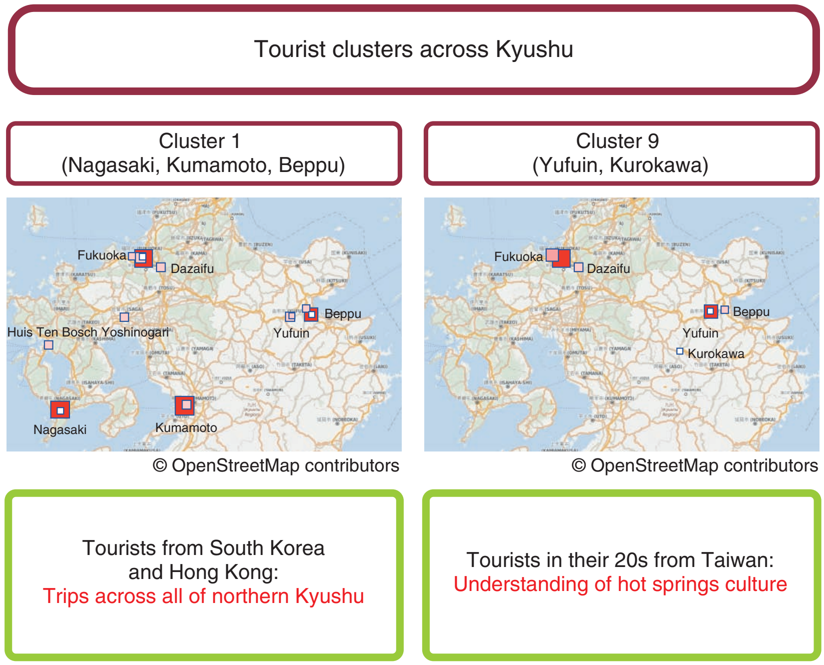


Fig. 3. Trip patterns over a large geographic area.

latitude and longitude coordinates at fixed time intervals. However, this is not appropriate for determining where a trip begins (e.g., at a store) and ends (e.g., at a tourist destination) because the data sets are large and include location data while people are in transit. Using our visit identification technique, we converted movement tracking data into *visits* (location data indicating where we determined that tourist app users stayed for some period of time). By further dividing locations into 500-m^2 areas and counting the number of times each user stayed in each area, we were able to analyze users' movement from one area to the next.

These techniques produced a multidimensional data set with information on each user, including areas visited, movement patterns between areas, times of day, gender, age (by decade), and country of residence. We then used our multidimensional data analysis techniques to cluster the data. We analyzed trips and activities for both the island of Kyushu and the city of Fukuoka to discover patterns that were characteristic of each (**Figs. 3** and **4**). We believe that taking these activity patterns into account could allow tourist campaigns to both attract and satisfy more tourists.

(1) Patterns in trips and activities over large areas

- Most of the tourists were from South Korea and Hong Kong; their travels took them to Nagasaki, Kumamoto, Beppu, and other major tourist destinations all across the northern region of Kyushu.
- There were many younger tourists from Taiwan who traveled to hot spring resorts, including the Yufuin hot springs in Oita Prefecture and the Kurokawa hot springs in Kumamoto Prefecture.

(2) Patterns in trips and activities over small areas

- A noticeable number of tourists in (or near) their 30s from South Korea and Hong Kong visited locations throughout downtown Fukuoka.
- Taiwanese tourists in their 30s traveled large distances east and west across the urban districts of Fukuoka City during the day but returned to the downtown area at night.

3.2 Improving push notification open rates by analyzing trips and activities

In conducting our multidimensional composite data analysis, we augmented users' location data and attributes with category data on points of interest (such as tourist destinations and stores) in each area; this allowed us to cluster the data according to each area's characteristics such as how many tourist attractions and restaurants it contains. Because we can estimate how tourists' individual attributes are related to the

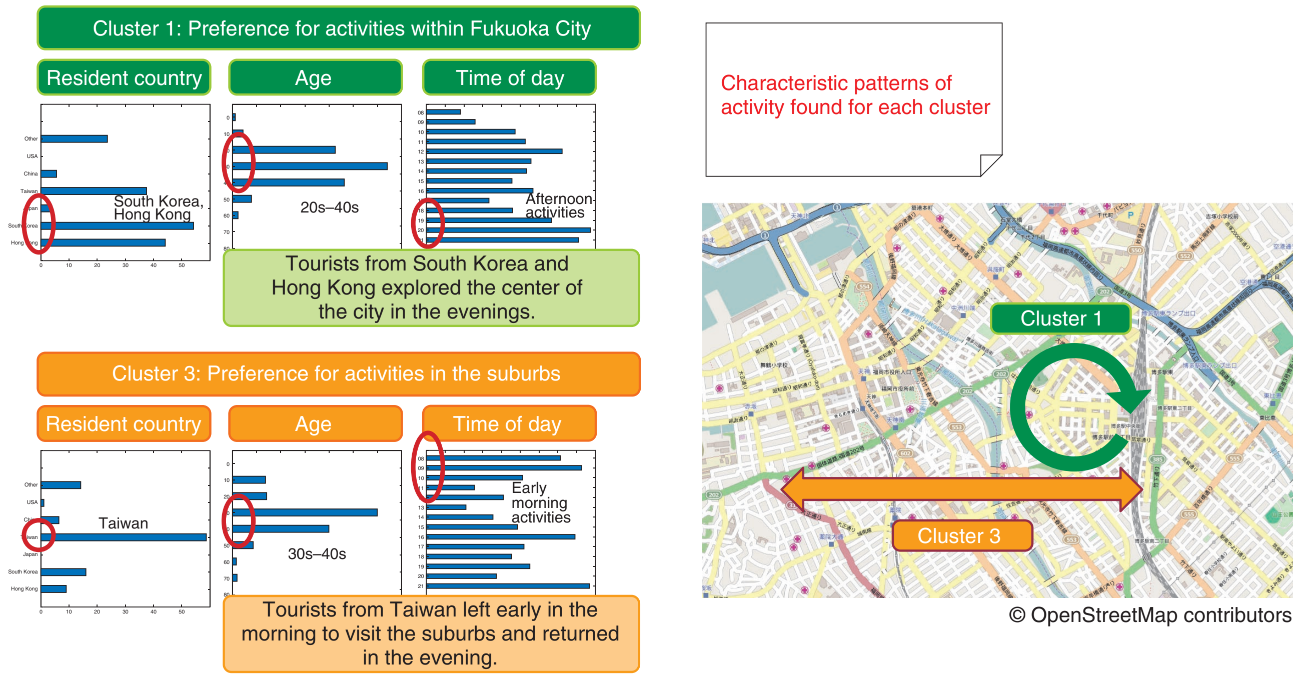


Fig. 4. Minute trip patterns.

areas and categories of places they are likely to visit, we can also make effective recommendations based on those attributes. On the basis of our actual clustering results from this trial, we sent push notifications to our tourist app with information on tourist spots and categories tailored to individual combinations of country of residence, age (by decade), and gender. We found that these notifications were more than three times as likely to be opened as notifications sent out according to manual rules (**Fig. 5**).

3.3 Verifying the accuracy of real-time predictive crowd counting

We analyzed location data to predict how many people would be downtown several hours in the future. To fill in the gaps presented by missing location data and improve the accuracy of our results, we interpolated the location data before making any predictions.

After analyzing input data between December 1, 2014 and March 22, 2015, we found that our predictions still remained accurate when the number of tourists doubled during the Chinese New Year (from late February until early March).

4. Future plans

Rather than limiting ourselves to studying the flow of people, we will continue to pursue R&D efforts targeted at making social infrastructure—such as transportation and communication services—more stable. In addition to analyzing data, we will gather proof of the effectiveness of big data technologies for forecasting, control, and other applications.

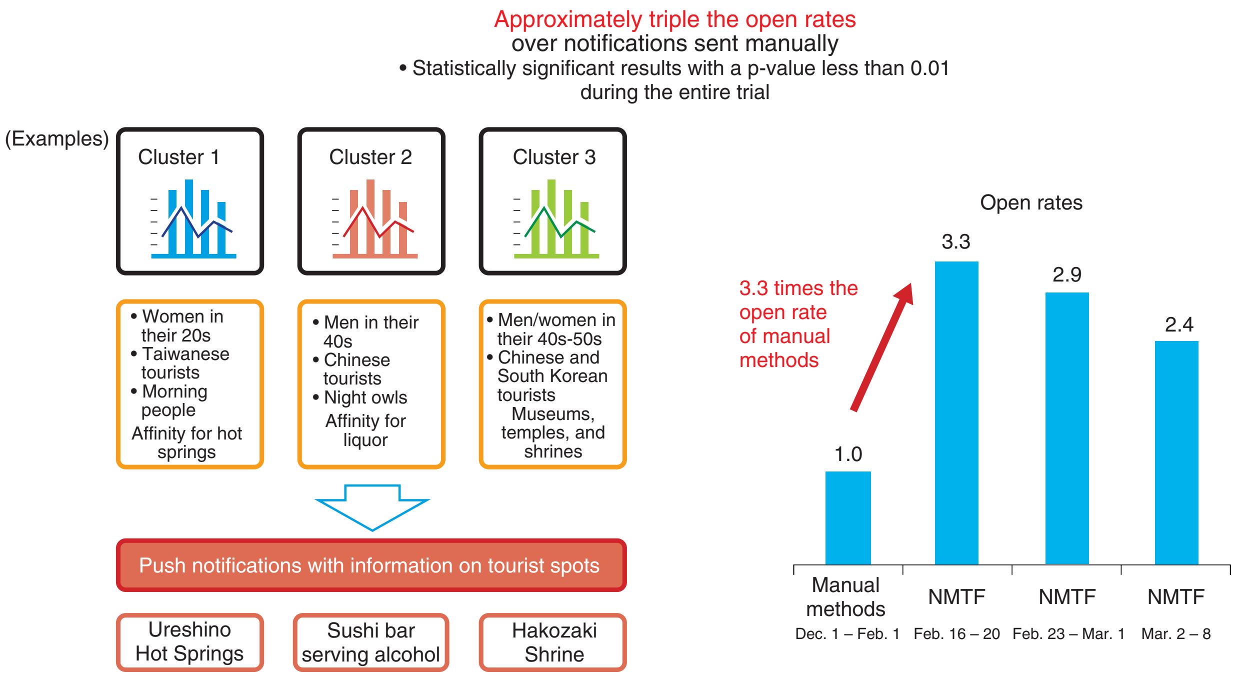


Fig. 5. Experimental results for push notifications.

References

[1] F. Naya and H. Sawada, "From Multidimensional Mixture Data Analysis to Spatio-temporal Multidimensional Collective Data Analysis," NTT Technical Review, Vol. 14, No. 2, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201602fa2.html

[2] Website of Jubatus, http://jubat.us/en/

[3] T. Hayashi, M. Umeda, M. Sawada, K. Isagai, A. Yamanaka, and M. Tsunakawa, "The Latest Developments in Jubatus, an Online Machine-learning Distributed Processing Framework," NTT Technical Review, Vol. 14, No. 2, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201602fa3.html

Ken-ichi Noguchi
Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. in electronic physics and an M.E. in human system science from Tokyo Institute of Technology in 2001 and 2003. He joined NTT in 2003. His current research interests include audio signal analysis and processing.

Hisako Shiohara
Senior Research Engineer, Proactive Navigation Project, NTT Service Evolution Laboratories.

She received a B.E. and M.E. in physics from Osaka University in 1990 and 1992. She joined NTT in 1992. She has been engaged in developing an advanced system for analyzing people flow data.

Yoshihide Sato
Senior Research Engineer, Proactive Navigation Project, NTT Service Evolution Laboratories.

He received a B.E. in electrical and electronics engineering and an M.E. in informatics from Kyoto University in 2000 and 2002. He joined NTT in 2002. His current research interests include systems for analyzing and optimizing the flow of people.

Overview of Network Synchronization Technology Standardization in ITU-T

Kaoru Arai and Makoto Murakami

Abstract

Network synchronization technologies are fundamental to telecommunications network services around the world. They are standardized internationally by the Telecommunication Standardization Sector of the International Telecommunication Union, Study Group 15. In this article, we present an overview of time and phase synchronization technologies that have been gaining attention in recent years and explain how these technologies can be achieved. We also introduce the initiatives being implemented for their international standardization and review the other topics being discussed for future standardization.

Keywords: network synchronization, PTP, ITU-T SG15

1. Introduction

With telephone or leased line network services using time division multiplexing (TDM) technologies such as synchronous optical network (SONET)/synchronous digital hierarchy (SDH) or asynchronous transfer mode (ATM), it has conventionally been necessary to match clock frequencies between network systems, or in other words, to synchronize networks, so that data can be transmitted and received. Synchronization technologies are fundamental network service technologies that are widely used by telecommunications carriers around the world. The standardization of these technologies has been ongoing for many years in international organizations including the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T). The NTT Group has made proactive contributions to international standardization efforts related to synchronization while developing equipment to generate and distribute high-quality clock frequencies.

Many telecommunications carriers are facing difficulties nowadays due to the aging of TDM network equipment and maintenance cost increases. Nevertheless, packet transport technologies have been developed that are cheaper than TDM technologies but provide a similar level of quality, which has accelerated the movement to replace TDM networks [1]. Synchronous Ethernet technologies have been developed to distribute clock frequencies on packet networks to establish network synchronization; these technologies were standardized internationally by ITU-T in 2008 [2]. Furthermore, in addition to requirements for frequency synchronization between equipment, the emergence in recent years of applications requiring time and phase synchronization with absolute time means that the areas in which synchronization technologies are applied are expanding. In addition to the mobile field, time and phase synchronization technologies are anticipated for uses such as matching timing with electricity storage and supply in the energy sector and high frequency trading in the financial and securities fields.

2. Technologies achieving network synchronization

There are mainly two classes of technology that enable synchronization of time and phase (**Fig. 1**). The first entails the acquisition of standardized time

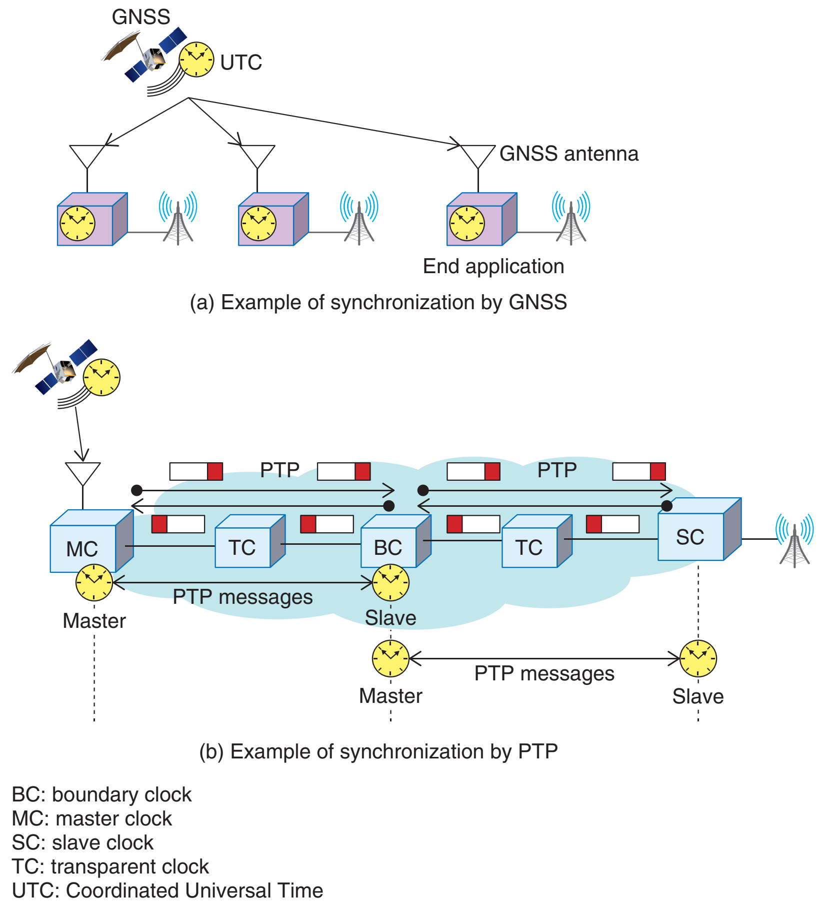


Fig. 1. Time and phase synchronization technologies.

data via radio signals received from global navigation satellite systems (GNSSs) such as the Global Positioning System (GPS). However, there are concerns about degradation in GNSS reception sensitivity due to constraints on antenna installation because of obstacles such as buildings in urban areas, the effects of radio interference, and solar winds or unfavorable weather conditions.

The other class consists of methods that entail synchronization via a network by using a protocol such as Network Time Protocol (NTP) or Precision Time Protocol (PTP). With NTP, clients request time synchronization from an NTP server that maintains a time standard. The clients perform time and phase synchronization in consideration of the time taken for a response to a request and the time between the server and the client.

With PTP, devices that perform synchronization take on the roles of master and slave. These devices mutually exchange PTP packets appended with time data (t_1 to t_4), as shown in **Fig. 2**. Based on the packet transmission and reception time, slave devices calculate the delay between devices and the time offset between the master and slave so that the slave can match its time with the master.

NTP uses a timestamp in the application layer and has synchronization accuracy to 1 ms. In contrast, PTP offers high-precision time and phase synchronization to 1 µs or lower with a hardware timestamp function and compensation for transmission path delay. In addition to the time and phase synchronization characteristics given in **Table 1**, PTP is also characterized by its ability to synchronize frequency by calculating the transmission interval from the PTP packet master as well as the reception interval at the slave.

Thus, PTP's superiority over NTP in synchronization precision and scope of applications means that PTP technologies are gaining attention in a wide range of business fields requiring high-precision time

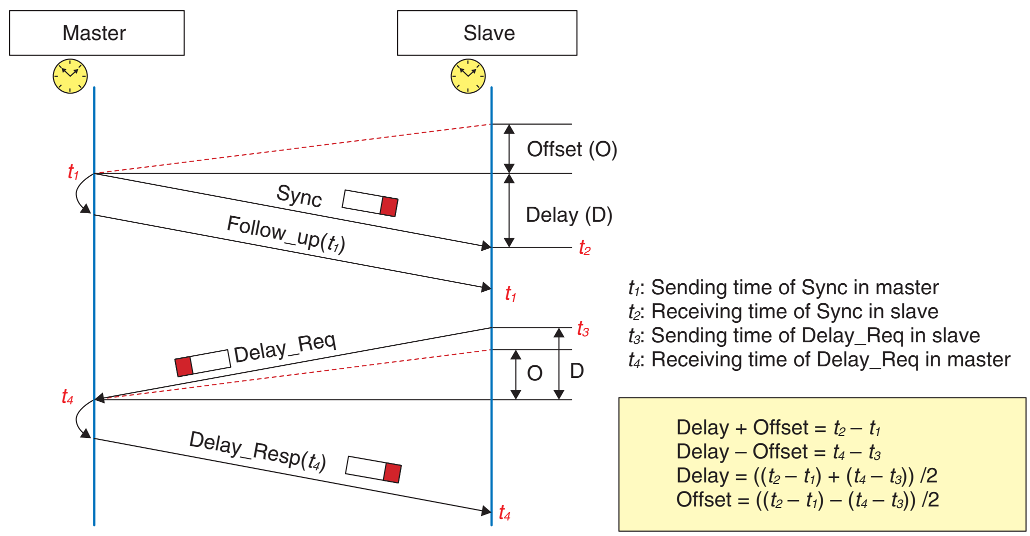


Fig. 2. Principle of PTP.

Table 1. Synchronization methods.

Synchronization methods		Function	
		Frequency	Time/Phase
Packet based synchronization	PTP	○	○
Physical based synchronization	Synchronous Ethernet	○	×
GNSS (GPS, Galileo, GLONASS, etc.)		○	○

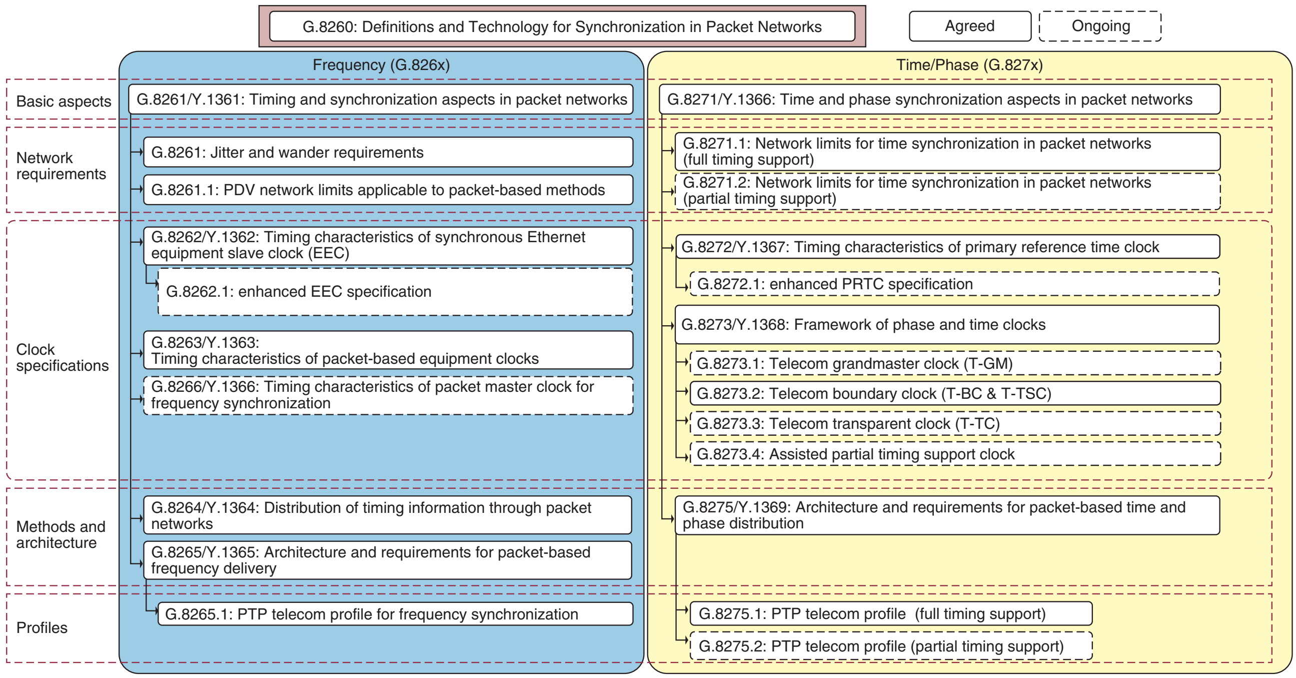
GLONASS: GLObal NAvigation Satellite System (Russian)

and phase synchronization, while PTP defined by IEEE (Institute of Electrical and Electronics Engineers) is being standardized by ITU-T for use in the telecommunications field.

However, a number of issues must be solved in order to achieve high-precision time and phase synchronization with PTP on a telecom network. For example, if wavelength division multiplexing (WDM) equipment or multiplex equipment that does not support PTP is used, uplink and downlink delay asymmetry must be corrected in the PTP offset calculation. In addition, network congestion can cause packet delay fluctuation or PTP packet loss, which can seriously affect synchronization precision. ITU-T is actively discussing problems that can have particularly serious impacts, and NTT has proposed delay asymmetry compensation methods to ITU-T, which were specified as recommendations in standardization documentation in 2014.

3. ITU-T recommendation systems and standardization trends regarding network synchronization

ITU-T discusses numerous issues including network synchronization precision, requirements for configuring networks, profiles, and testing and measurement methods for telecommunications. Through standardization, ITU-T prescribes the quality for synchronization applications to achieve interoperability between telecommunications providers and interconnectivity between equipment. The ITU-T standardization recommendation system related to synchronization is shown in **Fig. 3**. Synchronization standardization recommendations are mainly for synchronous Ethernet, or are roughly divided into PTP frequency



PDV: packet delay variation

Fig. 3. ITU-T recommendations of synchronization.

synchronization with the G.826x series recommendations, and time and phase synchronization with the G.827x series recommendations, where standardization of the former is almost complete. Standardization of the latter has been finalized and covers the requirements and limitations for configuring synchronized networks centered on time and phase synchronization using PTP and related equipment and network architecture. Currently, however, with new demands from end users for time and phase synchronization, there is a need to amend existing recommendations or create new ones, which has spurred discussions among carriers and system/device vendors around the world.

4. Diversification of synchronous network structures

ITU-T is targeting a number of profiles to achieve network synchronization. One of these, shown in **Fig. 4(a)**, is called full timing support (FTS) and has already been standardized as G.8275.1. FTS uses a master clock supplied with time data from GNSS, and it is a profile in which time is distributed to end applications via networks on which all devices have PTP functions.

In contrast to this, a profile called partial timing support (PTS), in which only some devices on the network support PTP, has also been proposed. PTS has been targeted for the G.8275.2 standards recommendations for 2016. Many issues remain such as defining switching actions that address degradation of time and phase synchronization quality on packet networks with PTS; hence, discussions are ongoing. In particular, assisted PTS used in conjunction with GNSS synchronization is attracting attention, as shown in **Fig. 4(b)**.

Behind this lie differences in the status of synchronization equipment implementation and in the thinking about the network structures of different carriers. With FTS, carriers building new synchronous packet networks can build them so they are compliant with standards, whereas carriers that have already deployed packet equipment that does not support PTP will need to set up new equipment, replace equipment, or add functions to achieve FTS. These carriers would like to be able to set up time and phase synchronization with assisted PTS used in conjunction with GNSS antennas installed only where high-precision synchronization is required.

Thus, to implement time and phase synchronization in the future, carriers will have to select the appropriate FTS or PTS profile to suit their network structures.

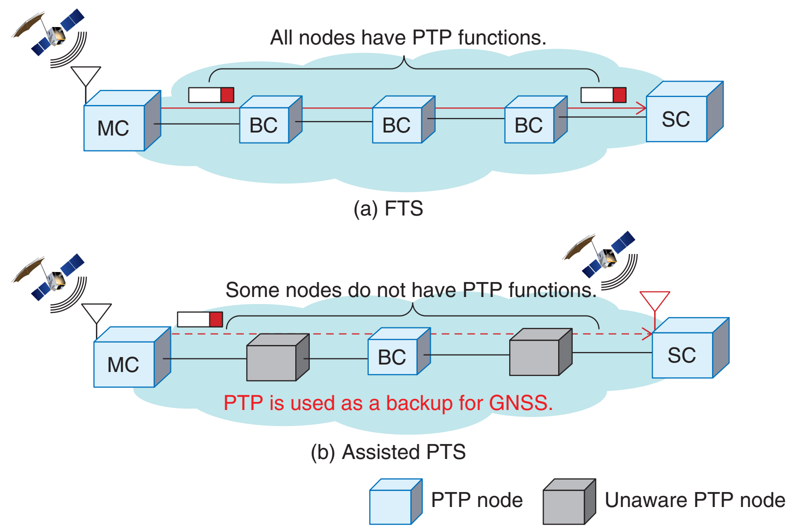


Fig. 4. Illustration of (a) full timing support (FTS) and (b) assisted partial timing support (PTS).

Table 2. Time and phase requirement classes.

Level of accuracy	Range of requirements	Typical applications
1	1 ms–500 ms	Billing, alarms
2	5 µs–100 µs	IP delay monitoring
3	1.5 µs–5 µs	LTE TDD (large cell)
4	1 µs–1.5 µs	UTRA-TDD, LTE-TDD (small cell) Wimax-TDD
5	x ns–1 µs	Wimax-TDD (some configurations)
6	< x ns	Some LTE-Advanced features

IP: Internet protocol
TDD: time division duplex

This will require detailed consideration of various factors such as reliability and operability, including issues such as new PTP equipment deployment and comparative costs of installing antennas for synchronization with GNSS.

5. High-precision time and phase synchronization

A future standardization issue for ITU-T is to make synchronization more precise. Therefore, ITU-T has defined six levels of precision for time and phase synchronization, as listed in **Table 2**. Current ITU-T recommendations are for standardization at precision level 4. However, mobile applications are being introduced that require time and phase synchronization precision at or below 1 µs. These include the LTE (Long Term Evolution)-Advanced inter-base station coordination technology called coordinated multipoint transmission (CoMP)[*1]. Additionally, there is demand for time and phase synchronization precision down to several hundreds of nanoseconds for technologies used to determine user terminal positions from base stations. For these reasons, ITU-T has begun discussions on new recommendations to expand the existing recommendations to the aforementioned six levels of precision.

*1 CoMP: Technology linking mobile coverage cells to suppress terminal throughput degradation due to interference from other cells at the cell’s edge. High-precision time and phase synchronization is required between mobile cells.

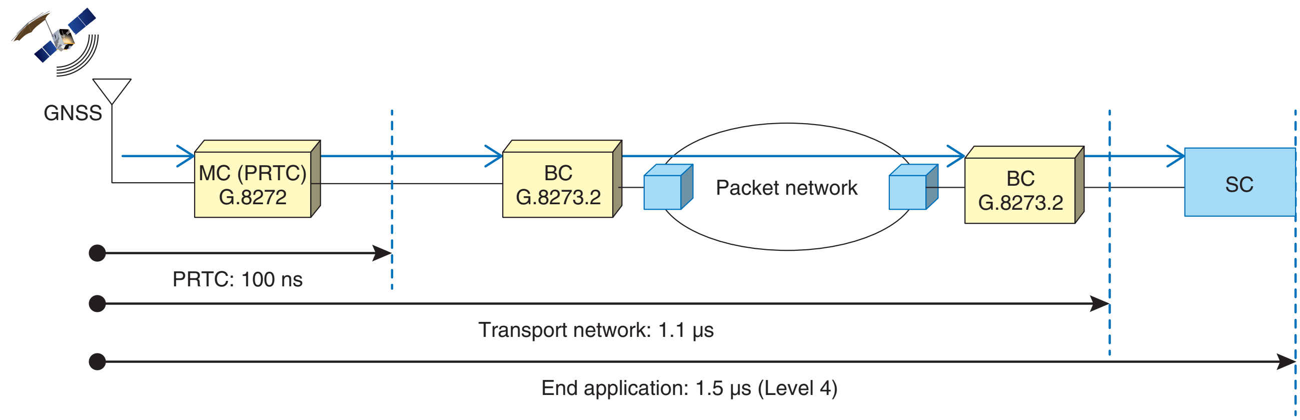


Fig. 5. Network limit of time/phase synchronization.

Precision of time and phase synchronization is defined as the absolute time offset to the end application from the time the standard equipment on the network known as a primary reference time clock (PRTC), which is supplied with time data by GNSS, starts the application. This is represented as the culmination of segment errors, which are typically time errors in PRTC, time errors in transmission segments, and time errors in end applications, as shown in **Fig. 5**. Current ITU-T recommendations describe designs and standards for segment time error tolerance that satisfy precision level 4 (1.5 µs), although to satisfy precision levels 5 and 6 (1 µs or lower), the permissible segment time error tolerance will have to be reduced. Thus, to reduce PRTC time errors, enhanced PRTC standards are under discussion that will reduce the 100-ns time error tolerance currently defined for PRTC to 30 ns in order to achieve higher precision and quality. The enhanced PRTC is due to be standardized as G.8272.1 in 2016. Because the scope of standards to be prescribed by G.8272.1 has only just been defined, detailed discussions on it will begin in earnest in the future.

There will also likely be similar discussions on network segment synchronization precision other than PRTC. For example, high precision boundary clock (BC) time data distribution devices are actively being discussed in response to demands from providers for high-precision time and phase synchronization.

For this reason, to provide high-quality network services and mobile applications, carriers all over the world including the NTT Group will have to consider their network synchronization equipment development and deployment while keeping pace with these demands and standardization trends in time and phase synchronization.

6. Future outlook

To build synchronous networks flexibly and efficiently in step with increases in network capacity and falling network costs, ITU-T plans to discuss systems such as time and phase synchronization methods on optical transport network and low-cost transparent clock (TC)*2 technologies that circumvent BC functions. Furthermore, since time and phase synchronization technologies are expected to provide fundamental support to the 4G (fourth generation)-LTE and 5G next-generation mobile systems, various requirements for network construction and operation from the carrier perspective are being sought. NTT also intends to stay informed about trends in services relating to time and phase synchronization both in Japan and internationally and will continue to proactively contribute to the international standardization of related technologies.

References

[1] M. Murakami and Y. Koike, "Highly Reliable and Large-capacity Packet Transport Networks: Technologies, Perspectives, and Standardization," J. Lightwave Technol., Vol. 32, No. 4, pp. 805–816, 2014.

*2 TC: A BC terminates received PTP packets as a slave and has functions to replay the PTP messages it receives as a master. In contrast, a TC compensates for the time the PTP message is retained in equipment, which enables transparent transfer of PTP packets to downstream BC. Thus, because the PTP packet termination and replay function is eliminated in a TC, a TC is potentially cheaper than a BC.

[2] Y. Koike and M. Murakami, "Status of Packet Transport Network Standardization in ITU-T," NTT Technical Review, Vol. 7, No. 7, 2009.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200907gls.html

Kaoru Arai
Engineer, NTT Network Service Systems Laboratories.
He received a B.S. and M.S. in applied physics from Tokyo University of Science in 2010 and 2012. Since joining NTT Network Service Systems Laboratories in 2012, he has been researching and developing packet transport systems such as clock supply systems and leased line systems. He has been participating in ITU-T SG15 activities concerning synchronization and packet transport technologies since 2014.

Makoto Murakami
Senior Research Engineer, NTT Network Service Systems Laboratories.
He received a Ph.D. in electrical engineering from the University of Tokyo in 2009. He initially engaged in research and development (R&D) of long haul transmission systems using optical amplifiers and coherent modulation/demodulation schemes at the emergence of those technologies. After completing development and deployment of a commercial optically amplified submarine system, he continued R&D of WDM systems to further increase the fiber transmission capacity. From 2001 to 2003, he worked for NTT Communications, where he was involved in the construction and operation of international communication networks mainly in the Asia-Pacific region. Since 2003 he has been an active participant in ITU-T SG15 as head of the Japanese delegation and has also been involved in R&D and standardization of large-capacity optical transport networks. He is currently the chairman of the transport networks and EMC (Electro-Magnetic Compatibility) Working Group in the Telecommunication Technology Committee (TTC) of Japan. He received the Accomplishment Award from the ITU Association of Japan in 2015 and the Distinguished Service Award from TTC in 2015.

Case Studies of Failures and Countermeasures in Access Network Facilities

Abstract

This article describes case studies of failures and countermeasures in access network facilities. This is the thirty-third article in a bimonthly series on practical field information on telecommunication technologies. This month's contribution is from the Access Engineering Group, Technical Assistance and Support Center, Maintenance and Service Operations Department, Network Business Headquarters, NTT EAST.

Keywords: access network facilities, failure case studies, damage caused by wildlife

1. Introduction

NTT's access network facilities are deployed throughout the country, with most installed in outdoor environments. As a result, failures occur every year because of weather factors such as rain, wind, snow, and lightning; the harmful effects of special environments containing salt, sulfidizing gas, or other elements; and damage caused by wildlife. In fiscal years 2013 and 2014, the Technical Assistance and Support Center received 97 inspection requests in relation to access network facilities (**Fig. 1**). We introduce three of these cases in this article.

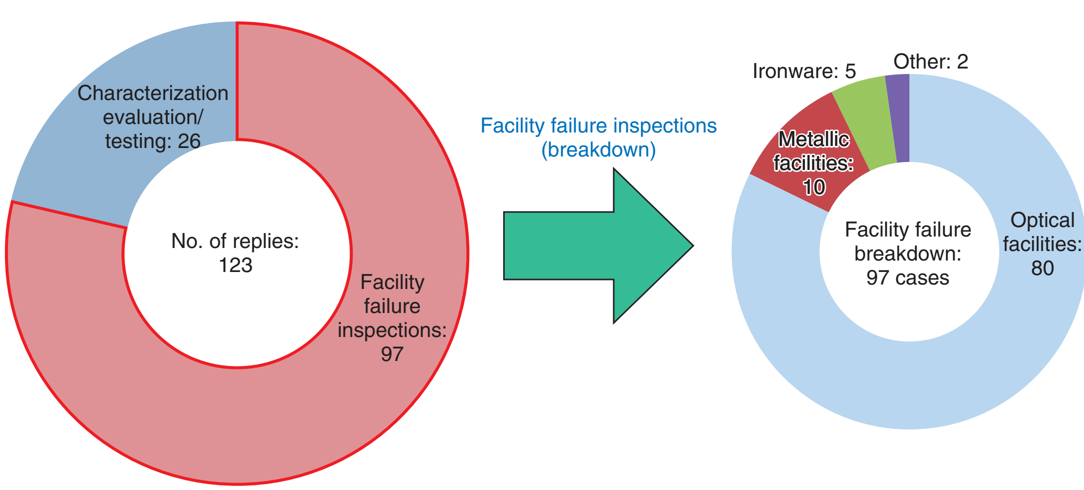


Fig. 1. Number of technical assistance and support requests.

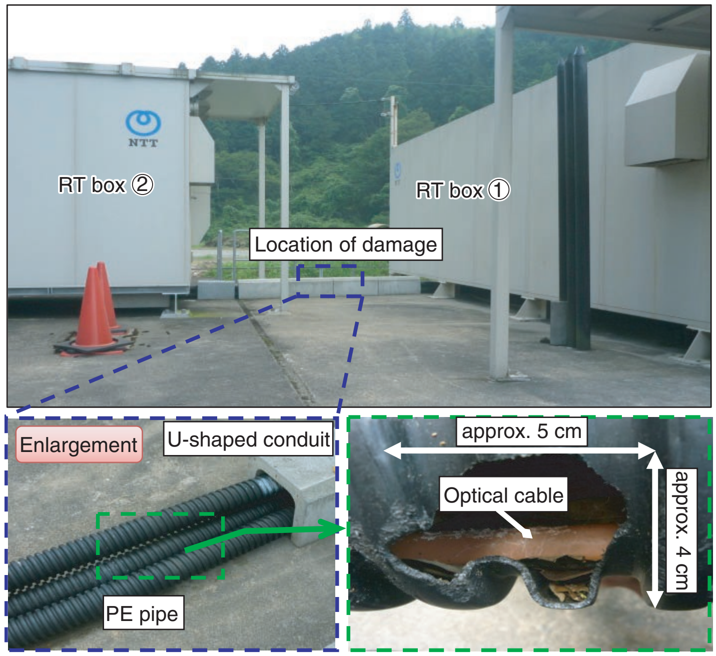


Fig. 2. Location of damage.

2. Introduction of failure case studies

2.1 Optical cable breakage

2.1.1 Overview

Disconnections in optical links between node equipment housed in remote terminal (RT) boxes occurred twice over a period of two days. An inspection performed by the local maintenance department revealed damage at a particular location in optical cable drawn between the RT boxes.

2.1.2 Inspection results and inference of cause

The optical cable between the RT boxes was drawn inside hard polyethylene (PE) pipe covered by a concrete U-shaped conduit. In this case, a 5 × 4 cm hole was discovered in one of three PE pipes within which tree leaves were scattered about (**Fig. 2**). A digital microscope was used to observe the optical cable and the PE pipe at the damaged location. This observation revealed incisor teeth marks (two upper teeth 0.7 mm in width and two lower teeth 0.6 mm in width) thought to be those of rodents. The widths of the teeth led us to infer that the damage-inflicting wildlife was a small rodent such as a house mouse (**Fig. 3**).

2.1.3 Countermeasure

As a countermeasure to damage caused by wildlife, it is important that facilities that require protection not be exposed. In this case study, it is thought that an animal penetrated the U-shaped conduit and gnawed on the cable in the process of building a nest inside the PE pipe. We therefore recommend that the entrance to a U-shaped conduit or any other location of penetration be sealed off and that any locations where wildlife-related damage has occurred in the past be periodically inspected.

2.2 Fiber breakage within optical closure

2.2.1 Overview

Breakage of an optical fiber occurred in two optical fiber cables within an optical drop closure at the bottom of an outdoor 8-branch splitter. An inspection performed by the local maintenance department revealed that the two failures occurred at the same time according to the monitoring-system log and that core fibers suffered breakages 8–10 cm from splice-on connectors within the closure (**Figs. 4** and **5**).

2.2.2 Inspection results

The splice-on connectors connected to the damaged fibers were visually examined, and no damage or other anomalies were found. The surfaces of the fractures in the fibers were then observed by scanning electron microscope, and it was found that the fibers fractured at an angle (**Table 1**).

2.2.3 Reproducibility experiment and inference of cause

In this case study, it was assumed that some type of external force during construction was applied to the

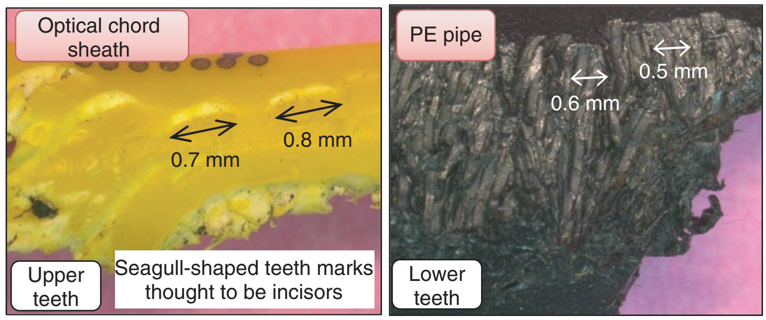


Fig. 3-1. Observations by digital microscope.

	Brown rat	Black rat	Field mouse	House mouse
Upper	2.0 mm	1.7 mm	1.0 mm	0.7 mm
Lower	2.0 mm	1.7 mm	1.5 mm	0.7 mm

Fig. 3-2. Incisor widths of various rodents.

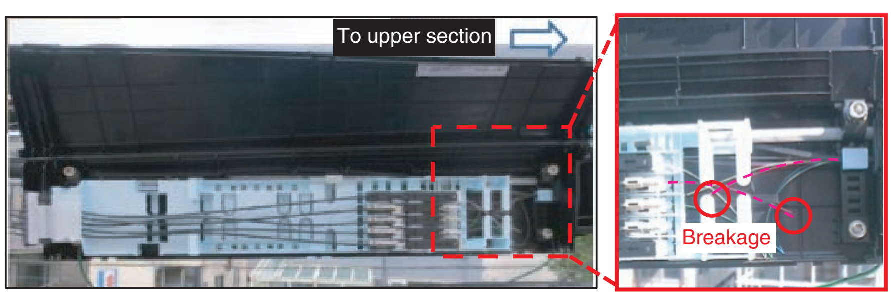


Fig. 4. Breakage locations.

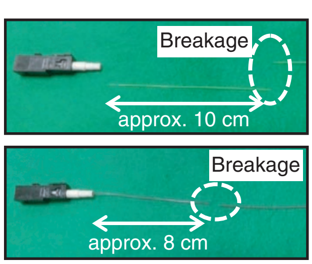


Fig. 5. Fiber breakage locations.

core fibers, which caused the fractures. We performed a reproducibility experiment for three types of external forces caused by (1) catching the optical fiber on the closure frame, (2) damaging the optical fiber with nippers, and (3) cramming the fiber into position with the closure lid (**Table 2**). The results revealed that the fractures could not be reproduced by external force (1) and that the fracture surface took on a mountain shape by external force (2). However, for external force (3), although a cramming effect could not be

Table 1. Fracture surface of failed fibers.

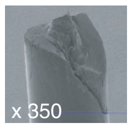
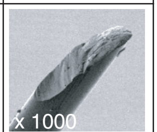
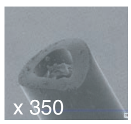
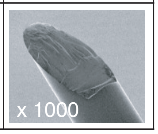
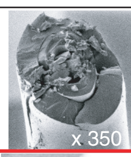
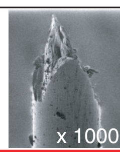
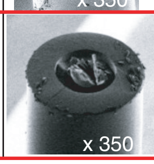
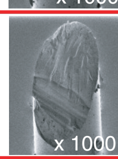
	With sheath	Without sheath
2nd fiber		
4th fiber		

Table 2. Results of reproducibility experiment (fracture surface).

	With sheath	Without sheath
(1) Catching	Fiber breakage not reproducible	
(2) Nipper damage		
(3) Cramming		

achieved using existing work procedures, it was possible to reproduce the shape of the fractured fibers in this case study when using procedures outside of installation standards. In an optical drop closure with splice-on connectors, the excess cutoff length of fiber cores (cutting position) is predetermined, but if the cutoff position is longer than that predetermined length, it is known that the surplus length of the fiber core may protrude from the closure when being stowed and end up being crammed in by the closure lid. The fracture of a crammed-in optical fiber is diagonally shaped, and since that was similar to the received fractured sample, it was concluded that the type of fracture in this case study was caused by an external force such as cramming applied to a core fiber that had been cut off at the wrong position.

2.2.4 Countermeasure

Optical fibers are vulnerable to scratches, cuts, and bends, so they should be handled with the utmost care. We recommend that stipulated surplus lengths be observed when assembling a connector and that sufficient caution be taken to prevent fibers from getting crammed in when opening or closing the optical closure.

2.3 Breakage in 4T indoor optical cable in building pipe

2.3.1 Overview

When new optical cable was drawn through a pipe already containing optical cables (indoor optical fibers) in a customer's main distribution frame (MDF) room, a breakage occurred in the existing optical cable.

2.3.2 Results of on-site inspection

The pipe opening through which optical cables pass in the MDF room was installed in the downward direction at a height about 60 cm from the floor, and it was found that a portion of the edge of the pipe opening had been chipped away (with a width of about 10 mm), as shown in **Fig. 6**. In addition, the damaged optical cable suffered a breakage near the pipe opening. This breakage took the form of groove-shaped damage with a maximum width of 3.7 mm that occurred at an angle to the longitudinal direction of the cable.

2.3.3 Reproducibility experiments

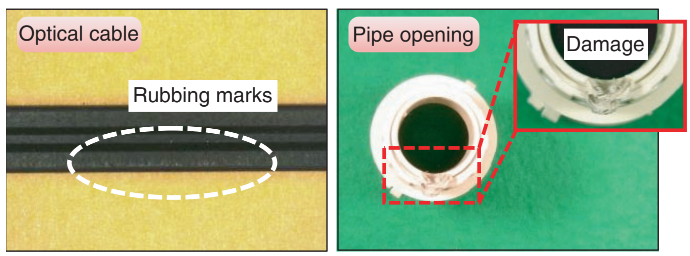
We conducted reproducibility experiments to isolate the cause of this breakage. First, we inserted 15 m of a threading wire (6.5 mm in diameter) into a pipe in which multiple optical cables already existed and pulled it out while in contact with the pipe opening, repeating this process three times. In each case, only rubbing marks appeared on the outer sheath of the existing optical cables, but it was found that the edge of the pipe opening was damaged in the form of groove-shaped chipping similar to the damage in the on-site pipe (**Fig. 7 (1)**). We then conducted another experiment, this time by fitting existing optical cable into the groove-shaped damaged section of the pipe opening and again inserting and pulling out a threading wire three times. For two of the three times, we found that we were able to reproduce the state in which the outer sheath of the existing optical cable was scraped away to expose the optical fiber (**Fig. 7 (2)**).

2.3.4 Inference of cause

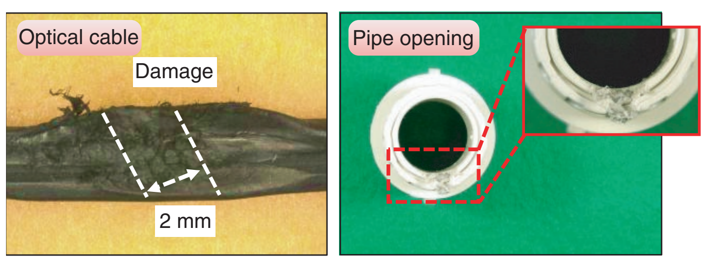
The results of the inspection and reproducibility experiments led us to infer that this type of failure occurred in the following process.

1) When performing threading operations through a pipe using a threading wire in the past, contact

Fig. 6. On-site conditions (MDF room).



Reproducibility experiment (1)



Reproducibility experiment (2)

Fig. 7. Results of reproducibility experiments.

between the edge of the pipe opening and the threading wire generated friction that eventually damaged the pipe opening.

2) Then, when performing threading again in a state in which an existing optical cable was stuck in place within the damaged edge of the pipe opening, the optical cable and the threading wire rubbed against each other, thereby damaging the optical cable.

2.3.5 Countermeasure

It is important that care be taken so as not to damage existing facilities during work operations. For the type of failure presented in this case study, we recommend that field personnel use threading wire carefully so as not to damage pipe openings and that they pay attention to the state of existing cable.

3. Conclusion

This report introduced case studies of failures in access network facilities and their countermeasures. By leveraging its knowledge and experience in the maintenance and operation of diverse facilities accumulated over many years, the Technical Assistance and Support Center will continue to handle requests from all concerned for technical consultation, assistance, and support and to provide information on ways to further improve the reliability of access network facilities.

NTT Communications India Pvt. Ltd. Netmagic Solutions Pvt. Ltd.

Takashi Nogami

Vice President, Global Product Development, Netmagic Solutions Pvt. Ltd.

Abstract

NTT Communications India (NTT Com India) is a managed MPLS (multiprotocol label switching), network integration, and system integration company in India. It has headquarters in Delhi/Gurgaon and five other branches across India. Netmagic Solutions is a sister company of NTT Com India that joined the NTT Communications Group in 2012. It offers cloud, managed hosting, and datacenter services in India. The sister companies are working closely together to deliver the benefits of total ICT (information and communication technology) solutions to our customers in India.

Keywords: business in India, datacenter, cloud

1. Introduction

In 2003, NTT Communications (NTT Com) decided to establish a representative office to deliver its international circuits to mainly Japanese customers who were using NTT Com's circuits outside of India, and to establish new locations in India. In 2005, in order to handle the increasing demand in India for its international circuits, NTT Com established a 100% subsidiary, NTT Communications India (NTT Com India). Thus, the year 2015 was the 10th anniversary of NTT Com India, and in April 2015, a large celebration was held with customers in Gurgaon, which is close to Delhi, the capital city of India. As of December 2015, NTT Com India has six offices, including the Delhi/Gurgaon headquarters and branches in Mumbai, Bangalore, Chennai, Neemrana, and the newest branch in Ahmedabad, which was established in April 2015. NTT Com India has about 100 employees across India who deliver network integration (NI) and system integration (SI) services mainly to Japanese/US/European multinational companies.

In 2012, to enhance the business in India, NTT Com acquired Netmagic Solutions (Netmagic), a cloud, managed hosting, and datacenter company based in Mumbai. Netmagic has a total of nine datacenters throughout India: two in Bangalore, one in Chennai, one in Noida (a satellite city of Delhi), and five in Mumbai, which together employ about 1000 staff. Netmagic has been recognized as an *operational excellence* managed service provider and recently built a new cutting-edge datacenter known as the Mumbai 5 Data Center (Mumbai DC5). This datacenter functions as Netmagic's headquarters and also hosts the NTT Com India Mumbai branch. NTT Com India and Netmagic provide an information and communication technology (ICT) single point of contact (SPOC) for customers through delivery of integrated NI, SI, cloud, managed hosting, and datacenter services (**Fig. 1**).

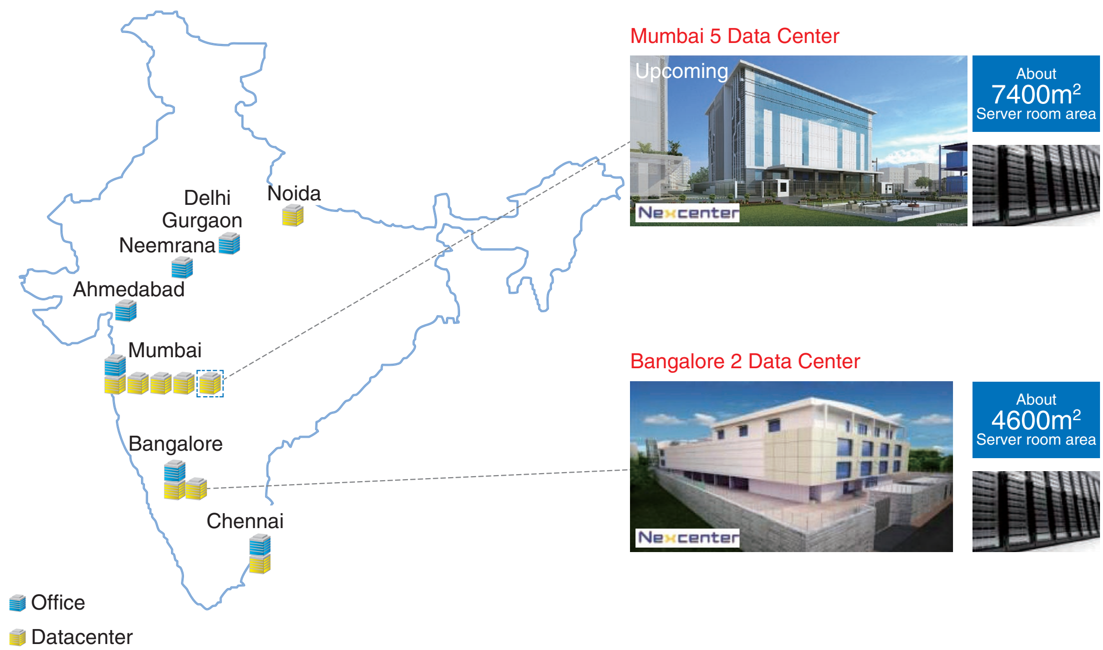


Fig. 1. NTT Com India offices and Netmagic datacenters.

2. Business overview

Many people recognize India as an information technology (IT)-business process outsourcing (BPO) destination country. There are several reasons for this, but it is mainly because there are many engineers who can speak English, and the time-zone difference is 12 hours from the US. Therefore, India can serve as a midnight IT factory for US companies to enhance round-the-clock agile software development or round-the-clock remote infrastructure management for critical ICT systems from India to the US. According to the CIA World Factbook, thanks to IT-BPO, the tertiary sector of industry in India was responsible for earning 57.6% of GDP (gross domestic product) even though population-wise it accounts for only 31% [1].

The NTT Group also recognizes India as a nation with huge ICT potential so has invested a lot in operations there. The NTT Group has seven major subsidiaries (major = owning more than a 51% share of the company), with more than 15,000 employees in India. NTT Com India/Netmagic promotes collaboration not only among five NTT Com subsidiaries (NTT Com India, Netmagic, Arkadin India, Emerio India, and Virtela India) but also with NTT DATA India and Dimension Data India, by enhancing cross-selling optimization. In July 2010, I was assigned as Mumbai branch manager of NTT Com India, and in October 2014, was newly assigned to Netmagic. Consequently, this is my sixth year in Mumbai. More details on these two NTT Com subsidiaries in India are provided below.

2.1 NTT Com India

The main business focus of NTT Com India is NI such as router/switch/firewall settings, and SI such as server/other ICT item settings, for multinational companies in India. One of the most important action items for those multinational companies at the initial stage in India is to find the best ICT partner in India. Actually, in the early stage, multinational companies need to focus on their own core business, and it is difficult to handle ICT setup—a key item to establish a business in India—by themselves. NTT Com India can serve as a kind of representative chief information officer (CIO) for customers by offering an ICT SPOC function.

In particular, the low quality of access line circuits is a huge headache for all those multinational companies that enjoy high quality services outside of India. To satisfy those customers' needs, NTT Com India decided in 2015 to add one more core product, national long distance services in India, by which NTT Com India can deliver its own managed multiprotocol label switching (MPLS) services in India.

Some people say that it is not difficult to provide MPLS in India, and this can be true if operational quality is not a big issue. NTT Com is well known throughout the world as an MPLS provider with operational excellence, and high quality is its core competency. By bringing operational excellence skills from its global experience, NTT Com India can deliver global standard quality MPLS in India.

Incidentally, in Japan, telecom providers can deliver carrier services with another carrier's telecom equipment, but in India, in order to deliver carrier services, the provider must own its equipment. This means that in India, no one can simply resell carrier services the way that we can in Japan. By delivering its own MPLS service, NTT Com India can be a total ICT solution partner providing services including NI, SI, and MPLS with operational excellence.

2.2 Netmagic

Netmagic is one of the top three datacenter providers in India and also offers cloud computing and managed hosting. Netmagic has nine datacenters across India, and the latest one is the Mumbai 5 Data Center that was launched on October 28, 2015 (**Photos 1** and **2**). It is one of the largest datacenters in India and provides cutting-edge, high quality functionality. It was designed by NTT Communications and follows the quality criteria of Nexcenter, which is a global seamless service brand for NTT Com's datacenters.

Netmagic was established by Mr. Sharad Sanghi, the current managing director and chief executive officer (CEO), in 1998. One of the missions of his business is to provide operational excellence, which is the same as NTT Com. In 2011, in order to deliver datacenter/cloud service solutions to a US-based banking customer, NTT Com India formed a strategic alignment with Netmagic. One year after that, NTT Com decided to acquire Netmagic to expand its business in India, so the relationship between NTT Com India and Netmagic became stronger and tighter. At that time, I was the Mumbai branch manager of NTT Com India and decided to migrate my office to Netmagic headquarters in order to maximize the synergy of the two organizations. Now these two organizations can be considered as one team, NTT Com/Netmagic (**Photo 3**).

Photo 1. Opening ceremony of new cutting-edge Mumbai 5 Data Center.

Photo 2. Mr. Sharad Sanghi (left), CEO of Netmagic, and Mr. Motoo Tanaka (right), SVP of NTT Com and head of Nexcenter, at Mumbai 5 Data Center opening ceremony.

3. Five core products in India

In India, NTT Com/Netmagic offers five main core product portfolios: colocation, hosting, cloud computing, remote infrastructure management, and a national long distance network (managed MPLS). (**Fig. 2**)

3.1 Colocation

Some people refer to colocation as *housing*, which means offering space, power, and air-conditioning for computers/servers with logical and physical security. To deliver high value for customers, Netmagic prepares redundant commercial power routes, redundant generators, and redundant UPS (uninterruptible power supplies), so that customers' servers never

Photo 3. Mr. Sharad Shangi (left), CEO of Netmagic, and Mr. Tetsuya Shoji (right), CEO of NTT Com, at Netmagic headquarters/NTT Com India Mumbai branch reception.

have power failures. As mentioned, NTT Com has the same datacenter quality criteria as Nexcenter, which defines those redundancy levels to deliver global seamless colocation services. All nine Netmagic datacenters follow these criteria, so customers can expect the same Nexcenter quality that is delivered outside of India from the NTT Com datacenter. Netmagic has five datacenters in Mumbai, two in Bangalore, and one each in Chennai and Noida. Therefore, our customers can choose one/some of them for their ICT engineers' convenience, or to maintain the fitness of their computer system architecture. Some customers choose a Mumbai datacenter as a primary site and a Bangalore datacenter as a secondary site. Other customers choose the Chennai datacenter as the primary site and the customers' captive server room in Delhi as a secondary site. Thus, NTT Com India/Netmagic offers flexible solutions.

3.2 Hosting

Hosting is a kind of service by which the service provider owns the ICT equipment such as servers and charges a monthly recurring fee to customers, in other words, an OPEX (operational expenditure) model (the opposite of capital expenditure (CAPEX) model). In this model, customers can reduce several problematic issues, for instance, expensive equipment costs, because NTT Com India/Netmagic, as a large ICT provider, has strong price negotiating power with manufacturers. The more "standard" a model is, the lower the rate NTT Com India/Netmagic can offer the service at. In addition to the inexpensive costs for equipment, NTT Com India/Netmagic can offer high quality operational excellence at a reasonable cost because the NTT Com India/Netmagic operation team works everyday on those standardized models. Therefore, the team has the skills and expertise to manage and trouble-shoot the models. The skilled team can deliver value to our customers in the form of operational excellence.

3.3 Cloud computing

Cloud computing is an on-demand and pay-per-use based remote and virtualized computer service. In hosting, customers can choose items from a list of manufacturers. In contrast, in our cloud computing services, the customer can choose from a list of items such as CPU (central processing unit) power, memory size, and storage size, and via a consolidated website called the customer portal site, the customer can easily establish virtualized computers that are physically located inside NTT Com India/Netmagic datacenters.

We call the physical setup a grid, and NTT Com India/Netmagic has four cloud grids in India, in Mumbai, Bangalore, Chennai, and Noida. Customers can choose one or a combination of them to meet their needs. Again, NTT Com India/Netmagic can deliver flexible solutions to our customers. Some customers use our Mumbai cloud grid as web servers for their end customers and use the Bangalore colocation space for their database servers. Cloud computing can reduce some customers' problems. For instance, some e-commerce customers had to pay a huge amount of money for seasonally fluctuating computer resources such as peak traffic during the Thanksgiving sales season in the US or during Diwali (or Dipawali: a Hindu festival) in India. To adjust the peak resources, those e-commerce companies need to prepare maximum resources that are excessive for non-peak seasons. By choosing NTT Com India/Netmagic cloud services that offer the pay-per-use model, those customers can reduce much of the cost of their ICT investment.

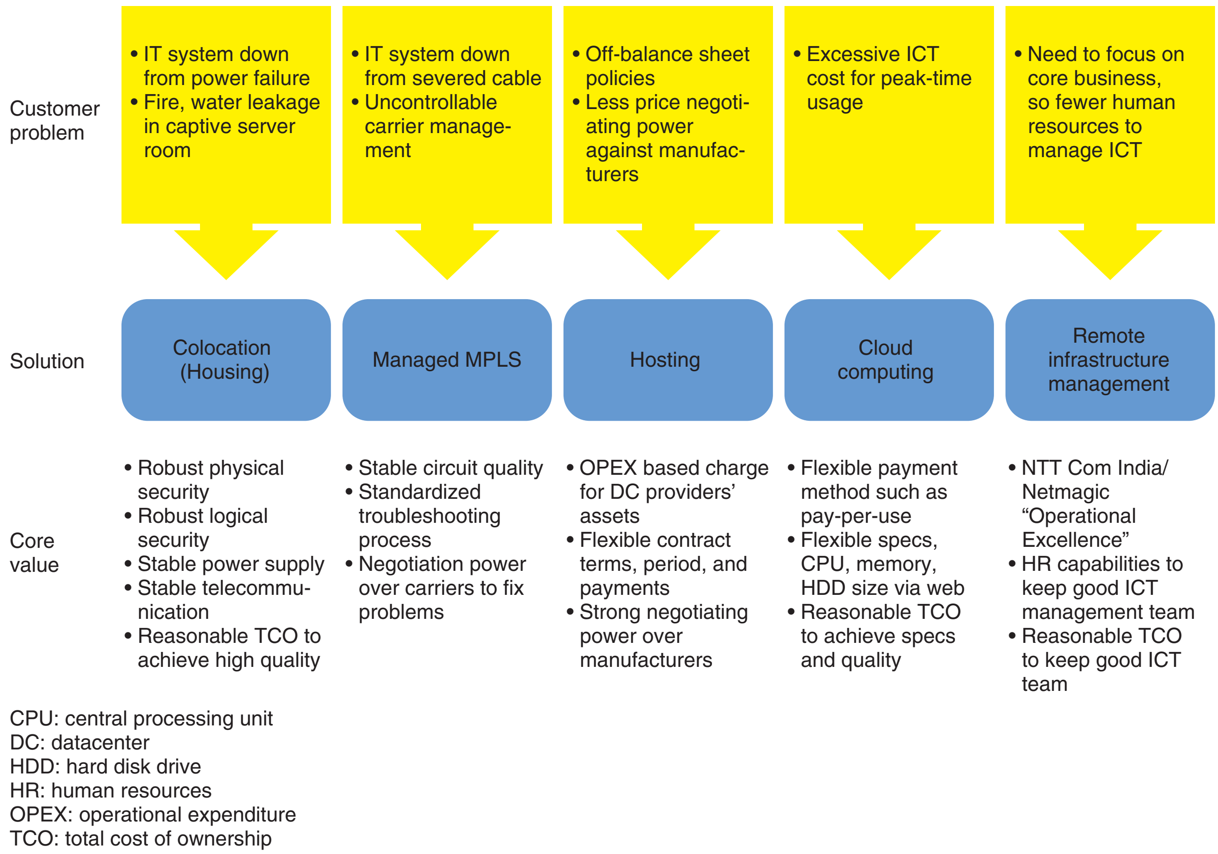


Fig. 2. Five core products in India.

3.4 Remote infrastructure management (RIM)

This service is a kind of BPO. NTT Com India/Netmagic offers *Remote monitoring for ICT equipment* + *Troubleshooting* + *Change management* as representative of IT management services for customers. RIM does not necessarily include colocation or cloud services, and NTT Com India/Netmagic is very flexible in offering the following combinations of services, or even other services, depending on customer requirements.

1) Colocation + hosting + RIM
2) Colocation + customer equipment + RIM
3) Customer server room + customer equipment + RIM

In India, NTT Com India/Netmagic is now focusing largely on RIM to eliminate customer problems. As I described, the ICT industry is one of the main business streams in India, so retaining good ICT engineers is a major headache for many non-ICT companies that are focusing on their core business and cannot invest a lot in ICT. As an ICT partner, NTT Com India/Netmagic can be a kind of CIO representative for those non-ICT companies by maintaining an ICT equipment monitoring/management team that can be shared with other customers or dedicated to a particular company. Customers can opt for a shared RIM solution if they require a standardized and inexpensive solution. If customers need complex operations, NTT Com India/Netmagic can deliver dedicated RIM solutions. Again, NTT Com India/Netmagic is flexible enough to propose both solutions depending on the customers' requirements and conditions.

3.5 National long distance network service (NLD)—managed MPLS

This is a new service for NTT Com India. In December 2015, NTT Com India obtained approval to get a network carrier NLD license from the Department of Telecommunications in the Indian Ministry of Communications & Information Technology. With this license, NTT Com India can deliver managed MPLS across India. Before this service was launched, NTT Com India/Netmagic was able to deliver datacenter related services but could not provide MPLS, meaning that customers had to go to other carriers for it even if they preferred hosting in Netmagic's datacenter. Some customers, in their RFP (request for proposal), state that "The provider needs to deliver both datacenter and MPLS services to eliminate demarcation issues when trouble occurs," and because of that, NTT Com India/Netmagic could not take part. With this managed MPLS, however, NTT Com India/Netmagic can offer its total ICT services to our customers, especially those who are expecting operational excellence for both datacenters and MPLS.

4. Next actions

NTT Com India/Netmagic will continue delivering one-stop ICT solutions by integrating five core products and providing them with operational excellence to customers. Operational excellence is not a one-time value but a primary long-term objective, so NTT Com India/Netmagic is working hard every day to increase value for its customers.

Reference

[1] CIA World Factbook, https://www.cia.gov/library/publications/the-world-factbook/geos/in.html

NTT Com India/Netmagic short column

Cricket, the most popular sport in India

In Japan, cricket is not a popular sport, but in India, it is by far the most popular one. We have a professional cricket league called the India Premier League (IPL), and professional players are stars with numerous fans all over the nation. On television there are several 24/7 dedicated cricket channels on which we can watch recent and past matches held in and outside India. On weekends and holidays, there are always a lot of young people playing cricket on fields or even on the road.

Of course, Netmagic/NTT Com India team members are also greatly interested not only in watching cricket games but also in playing them. We have our own Netmagic Premier League (NPL) that holds a tournament every year in February, with games on Saturday and Sunday. NPL is similar to the Sports Day events in Japan although not as casual. The team members are very serious, so two months prior to the NPL season, the team members start to practice with the goal of getting the winner's trophy. Because they take their playing very seriously, I, as a novice Japanese cricket player, cannot casually say, "Let me play" on NPL day. However, it is fun to watch the games, so every year I go to the event.

The origin of cricket is the same as that of baseball, so there are many similarities but also many differences. First of all, a team has 11 players, instead of 9 in baseball. There is a bowler instead of a pitcher, and a keeper instead of a catcher. The 9 fielders catch the cricket ball with their bare

The author (right) at an NPL game, wearing the same uniform as the players.

hands instead of with gloves.

The striker or batsman (instead of a batter), tries to hit the ball to get points. There are no fouls, so the striker can hit a ball 360 degrees around him/her. The cricket pitch (field) is oval, and there is a wicket, which is the set of three poles, or stumps, behind the batsman.

The striker is replaced when he is out. There are mainly three kinds of outs in cricket. The first one is *caught out*, in which the ball the batsman hit is caught directly by the fielder without bouncing on the ground. This seems to be the same rule as in baseball. The second one is *run out*, which is similar to a *first base force out* in baseball, although there are some minor differences (not enough space to write them in detail here, but they can be found on a wiki site or elsewhere). The last one is a *wicket out*, which is totally different from baseball. As can be seen in the photo, the wicket consisting of three stumps is located behind the batsman. The bowler can get a wicket out if the ball hits the wicket and knocks a stump down. The batsman tries to hit the ball so as to prevent the bowler from knocking the wicket down. In this sense, the batsman is carrying out defense like a goal keeper in soccer, and the bowler is carrying out offense like a kicker in soccer. This is really interesting and totally different from baseball.

There are various ways to get points, referred to as a *run*, but for simplicity here, the batsman tries to hit the ball to keep from being out. If the ball is hit and reaches the field boundary without being caught by the fielders, the batsman gets four runs, which is kind of like reaching second base in baseball, and if the hit ball goes past the boundary without bouncing on the ground, the batsman gets six runs, similar to a homerun in baseball.

Under the original rules, it can take as long as five days (!) to complete a game, which is tough for professional sports. Therefore, the IPL now has a *20 overs rule* (1 over = 6 balls), that is, 120 balls for each match. Logically, the batsman can get 720 runs (6 × 120) at maximum, and the bowler can get 11 outs for 11 balls at minimum. Thus, the team can get somewhere between 0–720 runs, although in the IPL, the average is 100–200 runs each, and it takes around four hours to complete the game (still very long!). In the NLP, the game duration is shorter, about 30 minutes, to give many members the chance to play.

In India, people all over the nation enjoy playing or watching cricket in all seasons. Come to India to play cricket! It is fun!

At NPL 2015. Bowler (left) is about to bowl.

The batsman hits the ball.

External Awards

2014 LOIS Young Researcher Award
Winner: Rika Mochizuki, NTT Service Evolution Laboratories
Date: May 14, 2015
Organization: Institute of Electronics, Information and Communication Engineers (IEICE) Information and Systems Society, Technical Committee on Life Intelligence and Office Information Systems (LOIS)

For "Impression Extraction for Comparison Services of the Restaurant."
Published as: R. Mochizuki, T. Watanabe, K. Tanaka, and T. Yamada, "Impression Extraction for Comparison Services of the Restaurant," IEICE Tech. Rep., Vol. 113, No. 381, LOIS2013-49, pp. 59–64, Jan. 2014.

2014 LOIS Research Award
Winner: Yoshiaki Seki, Naoko Chiba, NTT Secure Platform Laboratories; and Yoshiaki Hashimoto, The University of Tokyo
Date: May 14, 2015
Organization: IEICE Information and Systems Society, LOIS

For "Risk Assessment of Information Leak in Twitter Use of Employee."
Published as: Y. Seki, N. Chiba, and Y. Hashimoto, "Risk Assessment of Information Leak in Twitter Use of Employee," IEICE Tech. Rep., Vol. 113, No. 479, LOIS2013-75, pp. 125–130, Mar. 2014.

IPSJ Activity Contribution Award
Winner: Mitsuaki Akiyama, NTT Secure Platform Laboratories
Date: June 3, 2015
Organization: Information Processing Society of Japan (IPSJ)

For his contribution to the establishment and development of a community of researchers and implementers in the information security field.

2014 ICSS Research Award
Winner: Yuta Ishii, Takuya Watanabe, Waseda University; Mitsuaki Akiyama, NTT Secure Platform Laboratories; and Tatsuya Mori, Waseda University
Date: June 11, 2015
Organization: IEICE Technical Committee on Information and Communication System Security (ICSS)

For "Understanding Android Apps that Are Similar to Legitimate Ones."
Published as: Y. Ishii, T. Watanabe, M. Akiyama, and T. Mori, "Understanding Android Apps that Are Similar to Legitimate Ones," IEICE Tech. Rep., Vol. 114, No. 489, ICSS2014-94, pp. 187–192, Mar. 2015.

MVE Award
Winner: Keita Suzuki, The University of Tokyo; Masanori Yokoyama, NTT Service Evolution Laboratories; Yuki Kinoshita, The University of Tokyo; Takayoshi Mochizuki, Tomohiro Yamada, NTT Service Evolution Laboratories; Sho Sakurai, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose, The University of Tokyo
Date: July 2, 2015
Organization: IEICE Technical Committee on Multimedia and Virtual Environment (MVE)

For "Method of Enhancing Effect of Social Touch in Remote Communication between Males by Changing Impression of Gender."
Published as: K. Suzuki, M. Yokoyama, Y. Kinoshita, T. Mochizuki, T. Yamada, S. Sakurai, T. Narumi, T. Tanikawa, and M. Hirose, "Method of Enhancing Effect of Social Touch in Remote Communication between Males by Changing Impression of Gender," IEICE Tech. Rep., Vol. 115, No. 125, MVE2015-9, pp. 7–12, July 2015.

DICOMO 2015 Presentation Award
Winner: Keisuke Tsunoda, NTT Service Evolution Laboratories
Date: July 10, 2015
Organization: Executive Committee of IPSJ DICOMO (Multimedia, Distributed, Cooperative, and Mobile) 2015 Symposium

For "Estimation of Cognitive Performance Change Using Heart Rate Variability."
Published as: K. Tsunoda, A. Chiba, H. Chigira, T. Ura, O. Mizuno, and T. Tanaka, "Estimation of Cognitive Performance Change Using Heart Rate Variability," Proc. of DICOMO 2015, 8D-1, pp. 1694–1702, Tokyo, Japan, July 2015 (in Japanese).

DICOMO 2015 Best Paper Award
Winner: Keisuke Tsunoda, Akihiro Chiba, Hiroshi Chigira, Tetsuya Ura, Osamu Mizuno, and Tomohiro Tanaka, NTT Service Evolution Laboratories
Date: August 24, 2015
Organization: Executive Committee of IPSJ DICOMO (Multimedia, Distributed, Cooperative, and Mobile) 2015 Symposium

For "Estimation of Cognitive Performance Change Using Heart Rate Variability."
Published as: K. Tsunoda, A. Chiba, H. Chigira, T. Ura, O. Mizuno, and T. Tanaka, "Estimation of Cognitive Performance Change Using Heart Rate Variability," Proc. of DICOMO 2015, 8D-1, pp. 1694–1702, Tokyo, Japan, July 2015 (in Japanese).

CRYPTO2015 Best Paper Award/Best Young Researcher Award
Winner: Yousuke Todo, NTT Secure Platform Laboratories
Date: August 18, 2015
Organization: International Association for Cryptologic Research

For "Integral Cryptanalysis on Full MISTY1."
Published as: Y. Todo, "Integral Cryptanalysis on Full MISTY1," Proc. of CRYPTO 2015 (the 35th International Cryptology Conference), Santa Barbara, CA, USA, Aug. 2015.

Editor's Choice Blue Ribbon at World Maker Faire 2015
Winner: Motohiro Makiguchi, NTT Service Evolution Laboratories
Date: September 28, 2015
Organization: Maker Media, Inc.

For "Compact Virtual Theater—Kirari! For Mobile—."
I demonstrated a device that makes it possible to watch an aerial image that is available when using the Kirari! technique on smartphones.
Published as: M. Makiguchi, "Compact Virtual Theater—Kirari! For Mobile—," World Maker Faire New York 2015, New York, USA, Sept. 2015.

MWS 2015 Best Student Paper Award

Winner: Yuta Ishii, Takuya Watanabe, Waseda University; Mitsuaki Akiyama, NTT Secure Platform Laboratories; and Tatsuya Mori, Waseda University
Date: October 22, 2015
Organization: MWS (Anti Malware Engineering Workshop) 2015 Executive Committee, IPSJ Special Interest Group on Computer Security

For "A Large-scale Analysis of Cloned Android Apps."
Published as: Y. Ishii, T. Watanabe, M. Akiyama, and T. Mori, "A Large-scale Analysis of Cloned Android Apps," Proc. of MWS 2015, Nagasaki, Japan, Oct. 2015.

PWS 2015 Excellent Paper

Winner: Takuya Watanabe, Waseda University; Mitsuaki Akiyama, NTT Secure Platform Laboratories; and Tatsuya Mori, Waseda University
Date: October 22, 2015
Organization: PWS (Privacy Workshop) 2015 Executive Committee, IPSJ Special Interest Group on Computer Security

For "RouteDetector: Tracking Your Location with 9-Axis Sensors."
Published as: T. Watanabe, M. Akiyama, and T. Mori, "RouteDetector: Tracking Your Location with 9-Axis Sensors," Proc. of PWS 2015, Nagasaki, Japan, Oct. 2015.

2016 OSA Fellow Member

Winner: Masahito Tomizawa, NTT Network Innovation Laboratories
Date: October 20, 2015
Organization: Optical Society of America (OSA)

For his significant contribution to the advancement of optics and photonics.

69th Mainichi Publishing Culture Award in Natural Science Category

Winner: Junji Watanabe, NTT Communication Science Laboratories
Date: November 3, 2015
Organization: The Mainichi Newspapers

For "Somatic Intelligence for Generating Information."
Published as: J. Watanabe, "Somatic Intelligence for Generating Information," Kagaku-Dojin Publishing Co., Inc., Kyoto, 2014 (in Japanese).

TMF Most Significant Contribution to Frameworx

Winner: Shingo Horiuchi, NTT Access Service Systems Laboratories, and Takayuki Nakamura, NTT Comware
Date: November 4, 2015
Organization: TM Forum

For "Maximizing Profitability with NFV Orchestration."

IEEE SPS Japan Chapter Student Conference Paper Award

Winner: Daichi Kitamura, The Graduate University for Advanced Studies (SOKENDAI); Nobutaka Ono, National Institute of Informatics; Hiroshi Sawada, NTT Service Evolution Laboratories; Hirokazu Kameoka, NTT Communication Science Laboratories; and Hiroshi Saruwatari, The University of Tokyo
Date: November 5, 2015
Organization: IEEE (Institute of Electrical and Electronics Engineers) Signal Processing Society (SPS) Japan Chapter

For "Efficient Multichannel Nonnegative Matrix Factorization Exploiting Rank-1 Spatial Model."
Published as: D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Efficient Multichannel Nonnegative Matrix Factorization Exploiting Rank-1 Spatial Model," Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015, pp. 276–280, Brisbane, Australia, May 2015.

MNC 2014 Most Impressive Presentation

Winner: Daiki Hatanaka, Imran Mahboob, Koji Onomitsu, and Hiroshi Yamaguchi, NTT Basic Research Laboratories
Date: November 11, 2015
Organization: MNC 2014 (27th International Microprocesses and Nanotechnology Conference) Organizing Committee

For "All-Mechanical Bistable Memory in a Phonon Waveguide."
Published as: D. Hatanaka, I. Mahboob, K. Onomitsu, and H. Yamaguchi, "All-Mechanical Bistable Memory in a Phonon Waveguide," Proc. of MNC 2014, 5D-2-3, Fukuoka, Japan, Nov. 2014.

MNC 2014 Most Impressive Poster

Winner: Kenji Yamazaki and Hiroshi Yamaguchi, NTT Basic Research Laboratories
Date: November 11, 2015
Organization: MNC 2014 Organizing Committee

For "Renovation of Three-dimensional Electron Beam Lithography System for Improvement of Positioning Accuracy and Reduction of Turnaround Time."
Published as: K. Yamazaki and H. Yamaguchi, "Renovation of Three-dimensional Electron Beam Lithography System for Improvement of Positioning Accuracy and Reduction of Turnaround Time," Proc. of MNC 2014, 6P-7-6, Fukuoka, Japan, Nov. 2014.

Certificate of Recognition

Winner: Nei Kato, Hiroki Nishiyama, Katsuya Suto, Daiki Murayama, Hideki Kuribayashi, Graduate School of Information Sciences, Tohoku University; Yoshitaka Shimizu, Yasuo Suzuki, NTT Network Innovation Laboratories
Date: November 19, 2015
Organization: Municipality of San Remigio, Cebu, Philippines

For recognition of their tireless dedication and their willingness to share and support the disaster-resilient network experiments.

GN Workshop 2015 Best Paper Award

Winner: Keisuke Tsunoda, Akihiro Chiba, Hiroshi Chigira, Kazuhiro Yoshida, and Osamu Mizuno, NTT Service Evolution Laboratories
Date: November 28, 2015
Organization: IPSJ Special Interest Group on Groupware and Network Services (GN)

For "Online Estimation Method of a Cognitive Performance Using Heart Rate Variability."
Published as: K. Tsunoda, A. Chiba, H. Chigira, K. Yoshida, and O. Mizuno, "Online Estimation Method of a Cognitive Performance Using Heart Rate Variability," Proc. of GN Workshop 2015, pp. 1–10, Tokyo, Japan, July 2015 (in Japanese).

SSS'14 Young Author Prize

Winner: Masaya Murata, NTT Communication Science Laboratories
Date: December 5, 2015
Organization: The Institute of Systems, Control and Information Engineers (ISCIE)

For "Gaussian Unscented Filter."
Published as: M. Murata, H. Nagano, and K. Kashino, "Gaussian Unscented Filter," Proc. of SSS'14 (the 46th ISCIE International Symposium on Stochastic Systems Theory and Its Applications), Kyoto, Japan, Nov. 2014.

HCG Research Award

Winner: Reiko Aruga, Shin-ichiro Eitoku, Tae Sato, Toru Sadakata, and Tomohiro Tanaka, NTT Service Evolution Laboratories
Date: December 5, 2015
Organization: IEICE Human Communication Group (HCG) Special Interest Group on Human Communication Science (HCS)

For "Fundamental Study on Interactive Labeling Method Based on Retrieval Property."
Published as: R. Aruga, S. Eitoku, T. Sato, T. Sadakata, and T. Tanaka, "Fundamental Study on Interactive Labeling Method Based on Retrieval Property," IEICE Tech. Rep., Vol. 114, No. 440, HCS2014-99, pp. 149–154, Jan. 2015.

Papers Published in Technical Journals and Conference Proceedings

Magnetooptical and Crystalline Properties of Sputtered Garnet Ferrite Film Using a Stress Relaxation Buffer Layer

A. Furuya, A. Sasaki, H. Morimura, O. Kagami, and T. Tanabe

Japanese Journal of Applied Physics, Vol. 54, p. 09MF02, September 2015.

The purpose of this study was to improve the optical characteristics of garnet ferrite films sputter-deposited on a glass substrate. The magnetooptical properties of the garnet ferrite film are strongly influenced by the thermal stress imposed on the substrate during crystallization. The condition of the interface between the garnet film and the substrate during the initial film deposition affects the magnetooptical characteristics of the entire film. In particular, we revealed the effect of stress generated at the interface on the crystallinity of the deposited garnet films with a stress relaxation buffer layer by observing the film cross section and the film surface. In addition, we qualitatively estimated the effect of cracking in the garnet film on a glass substrate.

Development of High-efficiency Downsized Power Conversion (380 V to −48 V) Equipment for HVDC Power Supply System

N. Hanaoka, A. Takahashi, T. Tanaka, N. Yamashita, and M. Sugahara

Proc. of the 37th International Telecommunications Energy Conference (INTELEC 2015), pp. 12–17, Osaka, Japan, October 2015.

This paper proposes and describes the performance of high-efficiency downsized power conversion equipment for high-voltage direct-current (HVDC) use. The equipment is small enough to mount in the dead space of a rack. Configuring 380 V/–48 V conversion equipment with existing technology lowers the efficiency of the entire power supply system and increases the installation area. To solve these problems, it has become necessary to develop equipment with an efficiency of 97% or greater and with dimensions small enough for mounting in the dead space of a rack. To this end, we have achieved high efficiency and small size by adopting a circuit configuration that uses a compact, high-efficiency chip converter as used on motherboards in ICT equipment and that disconnects the voltage step-up circuit during normal operation.

Concept of New Power Supply System Topology Using 380 V and 48 V DC Bus for Future Datacenters and Telecommunication Buildings

T. Tanaka, N. Hanaoka, A. Takahashi, K. Asakimori, T. Iwato, A. Sakurai, and N. Yamashita

Proc. of INTELEC 2015, pp. 114–119, Osaka, Japan, October 2015.

A new power supply topology is proposed for using various types of power supply systems together in datacenters and telecom buildings. In this topology, the power supply chain from the building power receiving point to the IC chips in the ICT equipment is organized into blocks to simplify the number of DC bus conversion steps. Two types of voltage for the DC bus are proposed: 380 V and 48 V. This topology is expected to enable the design of power supply systems that are highly efficient, highly reliable, and less expensive and that can be flexibly applied in buildings that are large or small according to reliability requirements.

Subcarrier Restoration for Survivable Multi-flow Transponders in Elastic Optical Networks

T. Tanaka, T. Inui, W. Imajuku, and A. Hirano

Proc. of the 21st Asia-Pacific Conference on Communications (APCC 2015), Kyoto, Japan, October 2015.

We discuss survivability considerations for multicarrier-based elastic optical networks, focusing on the multicarrier transponder (MCT). We explain the necessity of subcarrier restoration that can recover multicarrier connections using backup sub-channels. An initial evaluation shows our restoration scheme improves transponder reliability.

Proposal of a Simple CD ROADM with Contention-free Add/Drop Function

A. Iwaki, A. Sahara, and M. Fukutoku

Proc. of APCC 2015, Kyoto, Japan, October 2015.

We propose a simple configuration for colorless and directionless (CD) reconfigurable optical add/drop multiplexers that enables contention-free add/drop operation to be achieved. In the configuration, we apply a combination of multiple small-port-count CD add/drop banks (CD banks) and round-robin CD bank assignment. Evaluation results show that the proposed configuration can substantially reduce intra-node contention.

Discovering and Describing Image Contexts from Socially Curated Contents

A. Kimura, K. Ishiguro, A. M. Alvarez, M. Yamada, K. Kataoka, and K. Murasaki

IPSJ Transactions on Mathematical Modeling and Its Applications (TOM), Vol. 8, No. 3, pp. 10–25, November 2015 (in Japanese).

This paper proposes a novel method for discovering a set of image *contents* sharing a specific *context* with the help of image collections obtained from content curation platforms. Socially curated contents are promising to analyze various kinds of multimedia information, since they are manually filtered and organized based on specific individual preferences, interests, or perspectives. Our method fully exploits the process of social curation: how images are manually grouped together by users, and how images are distributed in the platform. Our method reveals that images with a specific context are naturally grouped together and every image includes really various contexts that cannot necessarily be verbalized by texts. In addition, we show how the image contexts obtained from socially curated contents can be used for several tasks such as image classification, data visualization, and image retrieval. The key idea is to incorporate the contexts as side information to derive an embedding transformation so that images with similar contexts are close in the embedding space. Through experiments with images obtained from Pinterest, we show that our proposed method benefits several image-related tasks such as visualization, image classification, and image retrieval.

Ultralow Bias Power All-optical Photonic Crystal Memory Realized with Systematically Tuned L3 Nanocavity

E. Kuramochi, K. Nozaki, A. Shinya, H. Taniyama, K. Takeda, T. Sato, S. Matsuo, and M. Notomi

Applied Physics Letters, Vol. 107, No. 22, p. 221101, November 2015.

An InP photonic crystal nanocavity with an embedded InGaAsP active region is a unique technology that has realized an all-optical memory with a sub-micro-watt operating power and limitless storage time. In this study, we employed an L3 design with systematic multi-hole tuning, which realized a higher loaded *Q* factor (> 40,000) and a lower mode volume (0.9 μm^3) than a line-defect-based buried-heterostructure nanocavity (16,000 and 2.2 μm^3). Excluding the active region realized a record loaded *Q* factor (210,000) in all for InP-based nanocavities. The minimum bias power for bistable memory operation was reduced to 2.3 ± 0.3 nW, which is about 1/10 that of the previous record of 30 nW. This work further established the capability of a bistable nanocavity memory for use in future ultralow-power-consumption on-chip integrated photonics.

Emphasized Accent Phrase Prediction from Advertisement Text for Text-To-Expressive Speech Synthesis

H. Nakajima, H. Mizuno, and S. Sakauchi

IPSJ Journal, Vol. 56, No. 12, pp. 2384–2394, December 2015 (in Japanese).

Realizing expressive text-to-speech synthesis requires development of both text processing and rendering of natural expressive speech. This paper focuses on the former as a front-end task in the production of synthetic speech and investigates a novel method for predicting emphasized accent phrases from advertisement text information. For this purpose, we examine features that can be accurately extracted by text processing based on current text-to-speech synthesis technologies. Among features, the word surface string of the main content and function words and the part-of-speech of main function words in an accent phrase are found to have higher potential on predicting whether the accent phrase should be emphasized or not through the calculation of mutual information between emphasis label and features of Japanese advertisement sentences. Experiments confirm that emphasized accent phrase prediction using support vector machine (SVM) offers encouraging accuracies for the system which requires emphasized accent phrase locations as context information to improve speech synthesis qualities.

Coherent Raman Beat Analysis of the Hyperfine Sublevel Coherence Properties of $^{167}Er^{3+}$ Ions Doped in an Y_2SiO_5 Crystal

D. Hashimoto and K. Shimizu

Journal of Luminescence, Vol. 171, No. 3, pp. 183–190, December 2015.

In this paper, we describe the use of coherent Raman beat (CRB) spectroscopy designed to reveal the phase coherence time t_2 of the ground state hyperfine sublevels of $^{167}Er^{3+}$ ions doped in an Y_2SiO_5 crystal cooled to a cryogenic temperature. The short optical coherence time T_2 (~1 μs) and sublevel population lifetime t_1 (~0.1 s) of the dopant Er^{3+} ions make it difficult to implement Raman echoes and optically detected spin echoes for measuring t_2, respectively. In contrast, CRB is insensitive to the T_2 and t_1 values and worth employing if we can cope with inhomogeneous dephasing. By comparing carefully with the effects of sublevel inhomogeneous broadening Δ_{12} on the various Raman signals, we can estimate the t_2 value even when $t_2 > 1/\Delta_{12}$. The ground-state t_2 values of the Λ-type three-level system identified in the $^{167}Er^{3+}$ ions are approximately 12 and 50 μs when the dopant concentrations are 0.005 and 0.001 at%, respectively. These t_2 values are long enough to realize electromagnetically induced transparency and coherent population trapping with an Er^{3+}:Y_2SiO_5 crystal at a wavelength of 1.5 μm.

Large-scale Collection and Analysis of Personal Questions for Dialogue Agents

H. Sugiyama, T. Meguro, and R. Higashinaka

Transactions of the Japanese Society for Artificial Intelligence, Vol. 31, No. 1, pp. 1–9, January 2016.

In conversational dialogue, a talker sometimes asks questions that relate to the other talker's personality, such as his/her favorites and experiences. This behavior also appears in conversational dialogues with a dialogue system; therefore, the system should be developed so that it responds to this kind of question. Previous systems realized this function by creating question-answer pairs by hand. However, there is no work that examines the coverage of the created question-answer pairs over real conversations. This study analyzes a huge amount of question-answer pairs created by many question-generators, with one answer-generator for each character. Our analysis shows that 41% of personality questions that appeared in real conversations are covered by the created pairs. We also investigated the types of questions that are frequently asked.