National Institute of Informatics News ISSN 1883-1966 (Print) ISSN 1884-0817 (Online)

NII Today

64
Dec. 2017
National Institute of Informatics News

NII Interview

[Project research director Ken-ichi Kawarabayashi explains]

The algorithm that powers the networked society

[A research collaboration with NEC Data Science Research Laboratory]

Taking cutting-edge numerical algorithms from *basic science* to *business*

Discovering hidden regularity of web data using machine learning

Using information science to shed light on biological dynamics and networks

Feature

Decoding the world with network analysis

ERATO Kawarabayashi Large Graph Project

This English language edition
NII Today corresponds to
No. 78 of the Japanese edition

The algorithm that powers the networked society

Cultivating the brainpower that fuels the IT industry

Ken-ichi Kawarabayashi Principles of Information Research Division, National Institute of Informatics Professor / Advisor to the Director General / Global Research Center for Big Data Mathematics, Research Director, JST ERATO Kawarabayashi Large Graph Project Professor, School of Multidisciplinary Sciences, National University Sokendai, Graduate University for Advanced Studies

Interviewer: Jun-ichi Taki Senior Staff Writer, Nihon Keizai Shimbun

Graph theory is a branch of mathematics that uses graphs, which are collections of vertices and edges, to study the structure and properties of various types of networks. NII's Professor Kenichi Kawarabayashi leads the Kawarabayashi Large Graph Project, part of the Japan Science and Technology Agency (JST)'s ERATO (Exploratory Research for Advanced Technology) project running October 2012 to March 2018, to develop algorithms for rapid analysis of enormous, constantly changing networks, of which the Internet is the canonical example. We asked Professor Kawarabayashi to describe the objectives and achievements of his project.

Taki To start, perhaps you could tell us about some of the accomplishments of the Kawarabayashi Large Graph Project.

Kawarabayashi Well, I could describe any number of the many algorithms that we have developed for accelerating the analysis of large graphs, 69 research articles that we have published in top international journals, and 94 papers that we have had accepted at international conferences (as of October 2017).

However, what I would really like to emphasize here is the human talent that we have cultivated through this project. We have provided research funding to between 60 and 70 early-career researchers, primarily graduate students and postdocs, of which some 20 have gone on to academic positions at the University of Tokyo, NII, Kyoto University, Tokyo Institute of Technology, and other institutions. This will be a lasting legacy of our project.

Ken-ichi Kawarabayashi

Many people have pointed out that Japan is facing a shortage of researchers and engineers in IT (information and communication)-related fields, but not many people have a clear understanding of what needs to be done. The technologies in demand today are completely different from what could have been pursued 10 years within the confines of any one academic discipline. The fact that the ERATO program can transcend the boundaries between departments and disciplines in its provision of research funding makes it an extremely valuable framework.

Taki What sort of people are in highest demand?

Kawarabayashi Many of the problems visible in the world today can be reduced to graph or network models, including those involving transportation networks and supply chains, not to mention the obvious example of the Internet, which is growing by leaps and bounds every day. You can think of graphs as tools for visualizing the structure of networks to make them easier to understand. The world's population is less than 10 billion, and it is not that hard to handle networks in the range of 10-to-the-power-of-10 data nodes. However, if you start assigning attributes to each individual—age, address, and so on—then things start to become difficult when the number of attributes is around 100. The most serious issue facing companies like Google or Microsoft that use networks is how to keep up with networks that are constantly changing.

For example, the results of U.S. presidential elections affect stock prices, and there is a major need for the ability to extract information from networks and predict events that are going to take place in the world a little in advance. The question is *how far* in advance you can extract that information, and this is where the power of algorithms really becomes essential. No matter how beautiful a program you write, you are not going to get anywhere if the underlying algorithm is no good: the bigger the data set, the more crucial the algorithms. Developing fast algorithms is key. To give you a very simple example, suppose I asked you to sum the integers from 1 to 100. Many people know that there is a much faster way to do this than simply adding all the numbers


one by one.

The question of *which* algorithm is fastest depends on the task at hand. There is no such thing as a fully general algorithm capable of rapid analysis in all cases. You have to modify your approach as appropriate for the situation. It is the ability to produce answers even when the question changes that only humans have. It is thus crucial to have people equipped with enough general-purpose capability to be able to come up with ideas in any and all scenarios. For this purpose, programming skills alone are not sufficient. You also need mathematical ability and the capacity to think about things in profound theoretical ways and develop them longitudinally. People trained in the skill of thinking deeply and abstractly can transfer problems from one domain to another; they understand the connections between different fields of research and can think about problems in a way that blends approaches from multiple disciplines. This tends to generate insight; people suddenly realize, *hey, if I tackle the problem this way, then I'll be able to solve it.* This is something that computers cannot do. The question for human researchers is *how much depth does this person have?*

Taki I guess such people tend to aggregate in Silicon Valley in the United States, right?

Kawarabayashi Many of the top 10 companies by market capitalization are U.S. IT firms such as Apple. However, when you look at even the top 100 Japanese companies, no more than a few of them are on that list. The days in which companies could create value simply by building things are over. In today's world, value is added through services, and information is essential for services. I think that the leading manufacturing companies in Japan understand that such an evolution is taking place, but it seems to me that they are not keeping pace with the dynamic transformations happening in technology and the IT industry.

In order to stay abreast of the latest developments around the world, one really needs to be an honest-to-goodness researcher. The reason is that, if you are not one of the people recognized as a first-rate researcher within the community, then it is difficult to have that immediate grasp of the significance of cutting-edge research. You have to understand the academic significance of new trends. At a fundamental level, the fruits of research considered academic are directly tied to developing and improving new services in the IT world. For example, the head of Microsoft's R&D division is a former researcher. You need leaders with that sort of instantaneous, reflexive understanding of technology.

One of the things that I hear about Japanese IT companies is that, when a top-level executive is somebody from a liberal-arts background, then nobody knows how much of a detailed, nitty-gritty explanation of research results they should provide. They tell me that if they try to explain everything from the very beginning, then there will not be enough time to get to an explanation of the really innovative new developments. Japanese companies need leaders who understand IT research.

Taki The IT industry seems to have a different understanding of the meaning of basic research. The style of R&D in IT companies is not your typical stage-by-stage progression from basic research to applications and commercialization.

Kawarabayashi When I was living in the United States, I learned that some brilliant theoretical researchers in discrete mathematics and computational science were working at such places as AT&T Laboratories and Microsoft Research, and I would ask them point-blank what they were working on. Their role was not to contribute directly to developing products or services but rather to address issues that arose in the development process. They came up with solutions for various bottlenecks that arose at the development stage, and when problems could not be solved, they explained why the problems were impossible. This kind of thing cannot be done without mathematical ability.

Also, in order to be exposed to a variety of problems on a regular basis, one needs to make contact with the research going on in the world and understand the best work being done. For this purpose, researchers need to devote more than 50% of their time each day to their own theoretical research and produce first-rate results. Even at Google Research, I think that this basic structure is retained.

It is not necessary that every single researcher or engineer have a particularly deep background in mathematics. It is a relatively small core of elite researchers that forms the wellspring of Google's research strength. Training those elite researchers is an important imperative for Japan's competitiveness. There are talented young researchers in Japan, but if we are not careful, then they are all going to wind up working for Google. We need to cultivate a crop of elite researchers and create productive environments in which they can work. (Photography by Yusuke Sato)

A Word from the Interviewer

"If all you are doing is mathematics research, then you do not *need* tens of millions of yen." Kawarabayashi smiled as he said this. There is no doubt that training personnel is what gives meaning to Kawarabayashi's project, more so than for other ERATO projects. One also gets the sense that they have largely succeeded in their mission.

The information and communication networks that pervade our lives, which have grown at a breathtaking pace, amount to nothing less than game-changing technologies affecting every aspect of modern society, from the way we behave as individuals to the nature of business and government. One often hears commentary emphasizing the importance of *making things,* but in fact Japan needs to put more effort into understanding networks, which could be said without overstatement to *define* the space of our lives today, and focusing on the research needed to extract useful information from them.

Jun-ichi Taki

Senior Staff Writer Nihon Keizai Shimbun

Born in 1956, Jun-ichi Taki graduated from Waseda University with a degree in politics and economics and joined the Nihon Keizai Shimbun, where he worked in regional bureaus and as a corporate reporter before taking charge of the paper's science, technology, and environmental division in the mid-1980s. He is the author, in Japanese, of *Eko-uma ni nore! (Ride that eco-horse!)* (Shogakukan) and co-author of *Kansen-sho Retto (The Archipelago of Infectious Disease)* (Nihon Keizai Shimbun-sha), among other works.

Taking cutting-edge numerical analysis from *basic science* to *business*

"The best part about participating in the project is getting to work with such mathematically gifted people"

Junji Sakai
Senior Manager, Data Mining Technology Group
NEC Data Science Research Laboratory

NEC has maintained a collaborative relationship with the ERATO Kawarabayashi Large Graph Project ever since it began in 2012. Junji Sakai is a Senior Manager of the Data Mining Technology Group at NEC's Data Science Research Laboratory, from which two early-career researchers participated in the project; we asked him to discuss the significance of contributing as a private-sector corporation to a government-led project.

We want to establish data science as a tool for solving societal problems

—— What areas of research does NEC Data Science Research Laboratory pursue?

Sakai The laboratory was established in 2016 to bring various different types of research, from pattern recognition and image processing to AI-related problems such as machine learning and optimization, under one roof. We intentionally chose the name *data science*, and not *AI* or *media,* to destroy any remaining trace of boundaries between research fields.

For example, one of our researchers may develop a new algorithm as part of a basic research project and take on the challenge of presenting it at an international conference, while another researcher might analyze a customer's data and solve various problems. The idea is to create a good team and produce results without establishing any uniform across-the-board goals.

—— Tell us about the ERATO Kawarabayashi Large Graph Project.

Sakai At NEC, our *Network Graph Theories and Optimization Group* started participating in this project as research collaborators around 2012. Two NEC researchers, Shinji Ito and Akihiro Yabe, participated in the collaboration, and a paper summarizing its results was accepted to a top international conference in the world of information science. The two of them still participate in regular events associated with the project.

This is not research that we do on a contractual basis, nor is it a situation in which we have researchers permanently assigned to the project; we connect with the project on a slightly looser basis.

—— Why did you decide to get involved in the project?

Sakai The central pillar of NEC's business is the mission of "solving societal problems" by developing such things as computers and networks; that is, we are in the process of becoming a social solutions business.

We digitize real-world problems, understand them numerically, and apply proper algorithms to do optimization and control. Then the results of our work get fed back to society in the form of reduced waste in logistics and shipping, or production facilities, for example. For this to be possible, IT businesses such as ours need basic skills in various fields of mathematics.

NEC has also had a hand in data-science research for a very long time. It started with character recognition for postal codes and then expanded to such things as pattern recognition in images, with achievements including facial recognition technology, which is still ongoing today. Over the past few years, data science has replaced computers and networks as the new IT battleground, so we at NEC want to demonstrate our strength in this arena. We feel that research collaborations with academia are a valuable

Junji Sakai

After graduating from the Kyoto University Graduate School of Informatics in 1994, Junji Sakai joined NEC, where he has worked on the design and commercialization of parallel software technologies for use in cellular phones and other embedded devices. Since April 2016, he has been in charge of strategy planning and operation of the data mining R&D group.

NEC researcher Shinji Ito, who participates with Akihiro Yabe in the ERATO Kawarabayashi Large Graph Project

opportunity to sharpen our basic skills.

This project includes the participation of many researchers from the academic world, from whom we learn valuable points of view through our discussions. A major attraction is the ability to switch back and forth between basic research—such as graph-theoretic or other discrete-mathematics approaches, or robust optimization that gives insight into the variances of predicted values—and consideration of algorithms that contribute to society and business.

Instilling a feel for the landscape of advanced research and applying it as a research strategy

—— In what specific areas do you expect to see progress as a result of your participation in this project?

Sakai Recently, we have been doing research on predictive optimization. This is an approach in which we apply a visualization technique to the real world, predict the future, and construct an optimization plan for achieving desired outcomes.

Actually, when it comes to future prediction, NEC has already been developing predictive analysis technologies such as Hetero-Mixture Learning (HML) for about the last 4 years. However, the field of research on prediction-based optimization is only partially developed, and it is too early to tell whether this will be useful for business.

For example, if we are considering a retail store, we might want to predict such things as *how many customers will visit the store* and *which items will customers buy and in what quantities.* We could do this via such techniques as machine learning based on past data, and this allows computation with a certain level of accuracy.

In contrast, the optimization problem of designing an action plan to compute, based on a series of predictions, *which merchandise to stock in which quantities and how to arrange items on shelves to maximize profit* is not something that you can attack simply, right ? There are a large number of parameters that interact with each other in complicated ways—for example, if one product sells well, then another product will not sell as well—and it is difficult to solve the problem in a realistic amount of time.

By participating in this project and exchanging ideas with people active in the academic world, we develop what you might call a feel for the landscape of numerical algorithms and other advanced research. Then we can make more appropriate decisions about the directions in which to proceed with our research strategies for addressing optimization problems. This is one of the most significant benefits of our participation in the project as a corporation. We are also expecting to get a lot of help when it comes to developing our own algorithms.

In search of talented individuals who can bridge the gap between basic and applied research

—— As a private-sector corporation, what sorts of individuals do you hire as in-house researchers?

Sakai The critical shortage that we face in our corporate R&D environment is of people with the ability to read and understand a cutting-edge paper on numerical research, fill in gaps regarding any missing knowledge or skills, and then apply it to a real-world problem. People who can translate societal or business problems into numerical problems, exchange opinions with academics, and look for pathways to solutions. People with the ability to bridge the basic-applied gap in this way are extremely valuable.

In fact, the two early-career researchers that I mentioned earlier joined NEC after having been involved with the Kawarabayashi Large Graph Project while working on their masters' degrees. They both have a passionate interest in turning the numerical methods that they learned in academia into useful tools for the real world. It is hard to find first-rate people with such extensive mathematical background through the typical recruiting process. Participating in the project gives us access to people with a deep knowledge of the mathematics, which is another of the major advantages for us.

In some cases, an algorithm for solving problems in one type of business environment turns out to have applications in totally different industries as well. Thus, we are always reminding our research staff of the *one-to-many* principle. If they develop a method for solving one real-world problem, then we want them to expand it into a general-purpose solution with a broad range of applications. When we discover numerical methods with broad applicability and develop them into NEC solutions, we can serve more customers.

Of course, having said that, we do not actually think that any one corporation can develop cutting-edge algorithms in a closed environment. We develop tools in the open, making use of discussions with academia. Then we custom-develop the associated technologies needed to apply these algorithms to real-world problems, protect these with patents, and package them into solutions. We believe that this sort of approach is viable.

NEC researcher Akihiro Yabe

（Interview/Report by Naoki Asakawa, Nikkei BP . Photography by Yusuke Sato）

Photos (in this page)

ERATO Award Ceremony Season IV (2017.8.3 – 4, Hitotsubashi Hall, National Center of Sciences Building 2F)
Courtesy: JST ERATO Kawarabayashi Large Graph Project

Discovering hidden regularity of web data using machine learning

Finding needs and trends from user searches and posts

Takuya Konishi
Project Researcher, Graph mining & Web & AI groups
Global Research Center for Big Data Mathematics
JST ERATO Kawarabayashi Large Graph Project
National Institute for Informatics

This project has produced impressive results on algorithms for accelerating calculations and extracting new knowledge from the World Wide Web, the canonical example of a large graph. Examples include research on the connoisseurs of trending information on the web and the development of new methods for identifying patterns to deduce the intentions of users from their web-search queries. Applications of this work include improving the performance of web searches, detecting trends and targeting advertisements, and increasing the precision of marketing. We asked Takuya Konishi, a project researcher involved in this work, about new algorithms exploiting machine-learning techniques.

Inferring user intentions from search queries

The Kawarabayashi Large Graph Project involves many research collaborations with corporations, including project researcher Takuya Konishi's work on discerning user intentions from web search queries, which is a collaboration with Yahoo Japan Corporation.

Search queries are the inquiries that users enter into search engines; generally speaking, they correspond to search keywords. It is common for users to combine multiple words to form a search query, and this can furnish a clue to help understand the intent of the user's search. "We want to magnify the information contained in the search query," says Konishi, "to identify the user's background thought process."

Common patterns emerge for search queries used for similar purposes. For example, consider the queries *Chicago / hotel / luxury* and *NY / inn / cheap*. In both cases, the queries include words corresponding to *place names*, *travel*, and *criteria*, indicating shared, common topics. Combining these topics, we can describe both queries in terms of the same pattern: *place name / travel / criteria.* Discovering this sort of pattern, consisting of a combination of topics, is the goal of this research.

Learning patterns gives us clues to help understand what the user is thinking. In the example of the previous paragraph, we might conclude that the user is thinking, "I'd like to go on a trip." With this knowledge, we can propose appropriate search queries or recommend advertisements. We can sometimes display advertisements similar to previously entered search keywords, but to do this requires advance specifications of which advertisements to show in response to which specific input keywords. Understanding the patterns underlying user queries simplifies the task of matching the user with relevant advertising.

Automatically identifying topics via unsupervised learning

Discovering patterns requires both the correspondence between pairs of topics, which corresponds to keywords and patterns, and the correspondence between topics and words. For example, "Chicago" is a place name, but this is not clear to the search engine. Similarly, words such as "NY" or "Washington" belong to the same topic, namely, place names. Note that the term *topic* here simply refers to the distribution of words. The question is then how we go about achieving learning of this content.

In cases such as this, *probabilistic latent variable models* are frequently used. We assign a latent variable to each topic and envision a model in which queries are generated on the basis of these variables. What we actually know are the search queries, so we perform *unsupervised learning* to obtain the topics, which are represented as the discrete probability distributions of words, from the query log.

Accelerating the learning algorithm and discovering new knowledge

Konishi and his colleagues refer to their proposed method as the *pairwise coupled topic model (PCTM)*. Its characteristic feature is that it splits up search terms into all possible pairs of words and estimates their topic pairs. This allows PCTM to model the co-occurrence of two topics, i.e., the relationships between topics. By estimating the topic pairs appropriately, the original topic sequence information is recovered, and the learning algorithm of PCTM can be performed at moderate cost (Figure).

Konishi's team has also proposed a method for further accelerating the learning algorithm. This method focuses on the strength of relationships between topics. For example, travel-related topics are easily connected to topics related to place names but are difficult to connect to topics involving electrical products. The method reduces the computational time devoted to learning such weakly correlated pairs of topics, yielding both expressions for relationships between topics and a reduction in actual computation time. Indeed, Konishi says the method gives a

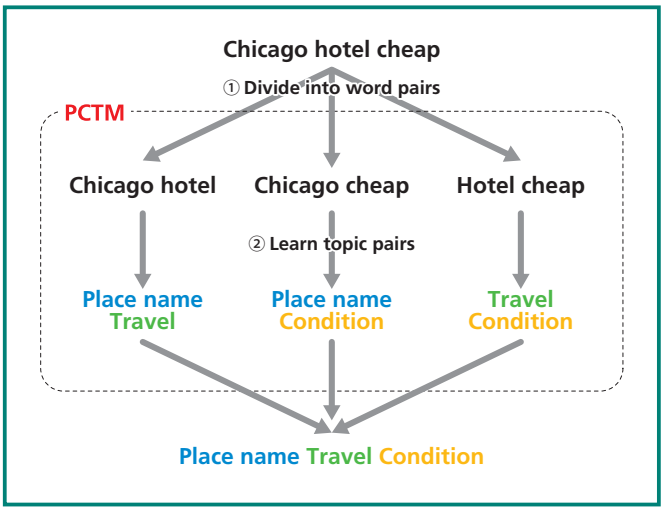

Figure Learning pairs of topics by decomposing search queries into all possible pairs of words. The original pattern (topic string) can be reconstructed via probabilistic inference from pairs of topics.

speedup of more than 10-fold without reducing accuracy.

Because conventional methods do not consider connections between topics, their representation mixes together tourism-related topics (to take one example) with topics related to place names. In contrast, PCTM allows each topic to be expressed separately.

Also, by subdividing categories in the manner of unsupervised learning, Konishi's team has succeeded in discovering connections that humans failed to notice. For example, in the category of topics related to place names, some items are easily connected to tourism thanks to region-specific characteristics; others have stronger associations with lifestyle-related matters such as rent, real estate, and hospitals; and still others have connections to words related to travel to distant destinations. Thus, one sees clearly that pairs of place names with strongly interconnections are distinct. Konishi notes that this sort of data-mining intelligence capable of discovering new rules in data sets is another outstanding feature of the methods that his team is developing.

The actual data set that the researchers used contains data for some 1 million search queries. This is just a small fraction of the queries entered by Yahoo! Japan users each day, but it suffices to yield significant expressions for topics.

Other research topics: rapid detection of trends and identification of user preferences

Another branch of Konishi's research involves discovering *trendspotters*, users with a knack for early identification of online trends. Specific trends frequently arise on SNS, and some individuals are particularly adept at sniffing them out before others; these are the trendspotters of the web, and Konishi's team set out to develop algorithms for identifying them automatically. Following trendspotters allows trends to be detected at earlier stages. Konishi and colleagues designed a method that uses machine learning to pinpoint such users, and can now do so with good accuracy. One interesting finding was that trendspotters do not necessarily have large numbers of followers.

Yet another branch of Konishi's work involves developing algorithms to generate music recommendations. These methods discern the musical preferences of users by analyzing their choice of songs, their reactions to songs, and other behavior, and then choose songs to recommend. In addition to faster selection of recommendations that accurately reflect users' tastes, the methods can both recommend matching songs and identify other possible taste preferences that users themselves may not be aware of, improving the accuracy of recommendations and accelerating the algorithm. The technique has already been adopted by a music-streaming service for smartphones. "The most interesting thing about this work," says Konishi, "was the experience of working with actual raw data not prepared for research purposes."

The increasing importance of linking different datasets

Going forward, Konishi's aspirations are not limited to making predictions and discovering patterns; rather, he explains, "I'd like to understand web events a little more deeply." This includes discovering the background knowledge or causal effects behind the events. For example, the growth of trends will be affected by the user interactions. Is it just a coincidence that trendspotters are able to sniff out new fads or are they truly experts? Getting to the bottom of this question, Konishi explains, will require one to not only analyze one dataset deeply but also link different datasets.

"I want to tackle some harder problems, not just evaluating whether the metrics are merely good or bad, but really understanding the relevant underlying factors."

(Interview/Report by Kazumichi Moriyama.
Photography by Yusuke Sato)

Takuya Konishi

Using information science to shed light on biological dynamics and networks

"We want to use the movements of ants to explore common dynamical patterns of living organisms."

Masato Abe

JST ERATO Kawarabayashi Large Graph Project
Project Researcher, Complex Network and Map Graph Group
Global Research Center for Big Data Mathematics* National Institute of Informatics
(*Title as of date of the interview)

The Kawarabayashi Large Graph Project also studies complex networks that arise in the actual world around us, including human behavior and societies of living organisms. The work of Masato Abe, Project Researcher in the Complex Network and Map Graph Group, focuses on the movement (dynamics) of ants. By studying the activity of worker ants using both movies and computer-based automated tracking systems, Abe not only established that the ants responsible for caring for eggs and larvae work nonstop but also searched for movement patterns allowing males and females to meet efficiently. Here he introduces his unique initiatives to explore the dynamics of living organisms and pinpoint structures and mechanics common to all living creatures.

Understanding relationships quantitatively

Masato Abe's research applies mathematical methods to understanding the networks and movements of groups of living organisms. Though fascinated by living creatures, Abe says he could never get the hang of molecular biology but felt immediately at home with tools such as mathematical biology, which applies mathematical models to ecological systems to quantify diversity and relationships, and evolutionary ecology, which simulates the dynamics of evolution.

The subject of his observations is the *Diacamma* ant, a social insect. Native to Okinawa, these large (around 1 cm) ants form colonies, each ranging from a few dozen to a few hundred individuals, making them easy to handle. Although the ants responsible for building the nest are all roughly identical in size and shape, the victors in battles for superiority earn the right to reproduce and become queen ants, with the others serving as workers. The task of younger workers is nursing, that is, caring for eggs and feeding larvae, but as workers advance in age, they travel outside the nest and bring back food.

Masato Abe

Abe and colleagues used methods of network science to analyze observational data and discovered hierarchical patterns in the behavior of worker ants. They found that some ants served in what might be termed a "mid-level management" position. This sort of relationship could never be discerned by studying the behavior of individual ants in isolation. The ability to identify patterns in the collective behavior of groups from the relationships between individuals is one of the most compelling advantages of network science.

Even one-on-one chance encounters are governed by rules for efficient movement

The characteristic feature of animals "is that they move," explains Abe. Where is the animal, and at what time? One of the challenges facing Abe and his colleagues is to analyze such questions quantitatively via methods of information science. Recent years have witnessed the development of *biologging* technology, in which animals are tagged with indicators to track their motion; this has enabled objective, quantitative studies. A central motivation for Abe's team is to discern from this sort of data the universal laws governing animal behavior.

For example, Abe explains that there are even rules for optimizing the efficiency of one-on-one chance encounters, such as between male and females. The problem of moving around to find a target object is known as a *random-search* problem. All living organisms—from bacteria, to animals, such as birds and even people (for example, those in hunter-gatherers tribes), all the way down to the immune cells in our bodies—solve random-search problems.

The typical behavior of living creatures is to travel a short distance

in a straight line, make a sudden change in direction, and repeat the process; however, from time to time, an animal will traverse a long distance in a straight line. Such as motion sequence is known as a *Lévy flight*, and it is thought to govern the motion patterns of living organisms when they do not know the location of their target object. Carnivores searching for prey and immune cells attacking pathogen entering the body both use Lévy flights to increase their chances of finding their target.

Now consider what happens if two individuals *want* to meet, as in the case of males and females seeking to mate. Abe and his colleagues designed a theoretical model premised on the assumption that two individuals wish to find each other and discovered that, to maximize the likelihood of chance encounters, the two individuals should move in different ways: one should execute a random walk, changing direction in irregular ways at a fixed time interval, and the other should execute a Lévy flight. Thus, the optimal search strategy for males and females seeking partners is to move in distinct patterns.

"Perhaps this is one of the reasons for the asymmetries between males and females at birth," Abe speculates. The explanation for the entirely distinct movement patterns of eggs and sperm, and thus the reason for the emergence of gender, may lie buried in these findings.

The activity times of ants vary with the need for childcare

One of the reasons that Abe chose to investigate the structure of networks of ants is their being extremely good examples of networks for which actual real-world data can be obtained. With colonies ranging from a few dozen to a few hundred individuals, they are easy to observe, and although *individual* ants are not particularly intelligent, their *collective* decision making leads to exquisitely honed behavior. But *why* do ants behave so skillfully? "There is a redundancy and robustness there lacking from man-made entities today," explains Abe, "that we want very badly to understand."

More specifically, Abe's team developed an automated tracking system that uses movies and computer analysis to obtain the locations of individual entities automatically and analyze them. Most recently, the team has found that, when worker ants are in the presence of eggs or larvae in need of care, their circadian rhythms (the biological processes, varying in time with a period of roughly 24 hours, that furnish internal body clocks) simply *turn off*, leaving the ants to work nonstop, without rest. Under ordinary conditions, ants, like other animals, have a day/night rhythm. However, when worker ants are added one at a time to a small vessel containing eggs or larvae at risk of dying if not fed or groomed, the ants proceed to work nonstop. On the other hand, when placed together with pupae in less dire need of care, the worker ants exhibit a restoration of circadian rhythms (Figure).

"The most interesting aspect of this finding," Abe notes, "is that the circadian rhythms vary flexibly depending on the nature of the work at hand." The numbers of eggs and larvae present in ant nests vary in time; the task of caring for them, it seems, does not fall on specific individual colony members, but rather the division of labor shifts dynamically depending on the total amount of work to be done.

Searching for the mechanics of life

Groups of living organisms can act in coherent organized ways with flexibility and robustness. For example, schools of mackerel or flocks of starlings can behave exactly as if they were single entities. After large numbers of physicists began to analyze collective behavior, it was suggested that the synchronized subsets of groups and the sizes of the regions over which correlations exist may exhibit *scale-free* (independent of group size) behavior, and that biological systems may have *critical points* (boundaries at which phase transitions, such as the transition from gas to liquid, can occur). The hypothesis that various biological systems, including collections of nerve cells in the brain and the state of expression of genes in cells, may exist at critical points on the cusp of phase transitions has also been proposed.

Abe thinks that the notion of biological critical points may be a common framework for understanding biological systems. In physics, phase transitions are induced by changes in temperature or pressure. What are the corresponding stimuli acting at biological critical points? In the case of ants, it is possible to investigate responses to a variety of external disturbances.

In the future, Abe will be moving from NII to RIKEN, and this upcoming transition has spurred him to investigate human behavior as well. His current project involves studying patterns in the meandering patterns of early-stage dementia patients with the hope of pinpointing how they differ from those of healthy subjects.

The fascinating thing about biological motion is its spontaneity. Robots produce outputs in response to inputs, but biological organisms simply move spontaneously of their own volition. "I think that sort of spontaneity may be crucial for living organisms," Abe suggests. "I'd like to discover the mechanics and the physical laws common to all forms of life, to identify the unique mechanics of biology. To figure out the mechanics governing life itself has been a lifelong goal." (Interview/Report by Kazumichi Moriyama. Photography by Yusuke Sato)

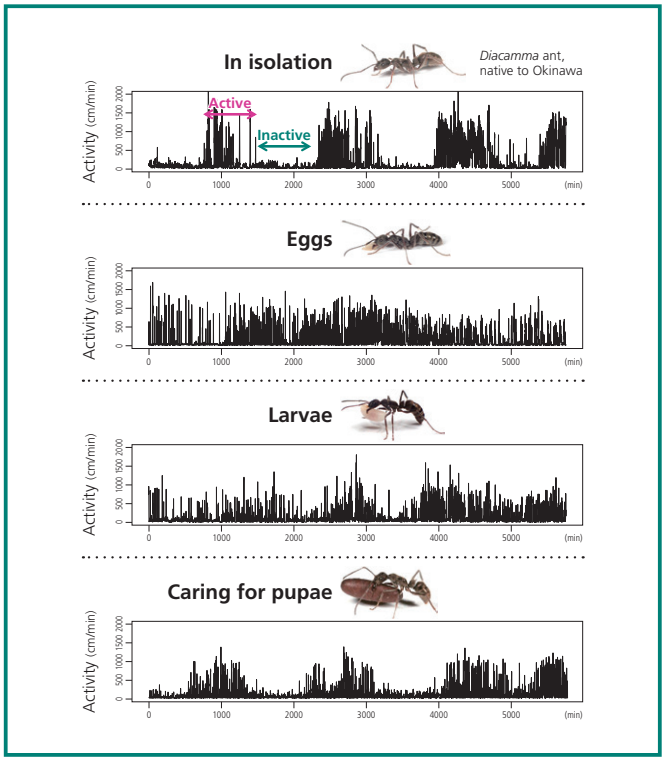

Figure Variation in ant activity. When attending to eggs or larvae, activities such as grooming, feeding, and other forms of selfless care are required; in this case, circadian rhythms cease. In contrast, pupae are enclosed in a cocoon, eliminating the need for care such as feeding and grooming; in this case the circadian rhythms of worker ants reappear.

NEWS 1

Using CPS/IoT to realize safe and secure systems for infrastructure upkeep and management

SIP research results submitted to CEATEC JAPAN 2017

At the CEATEC JAPAN 2017 information-technology exhibition, which ran October 3–6, 2017, at the Makuhari Messe convention center in Chiba, Japan, the National Institute of Informatics (NII) teamed with Hokkaido University, Tsukuba University of Technology, and Nagaoka University of Technology to present an exhibit titled "Protecting infrastructure through the IoT: Possibilities for monitoring technology and platforms for sensitive data management" (photo).

The exhibit presented results of the R&D of Integrated Data Management Platform for Civil Infrastructure Sensing project, one program area of the Strategic Innovation Promotion Program led by the Council for Science, Technology, and Innovation of Cabinet Office, Government of Japan. Today, the aging of infrastructure projects built during Japan's high economic growth period presents the nation with the challenges of increasing costs to maintain and improve infrastructure and growing risk of serious accidents due to the collapse of bridges, tunnels, roads, and other facilities. The project seeks to address this issue by designing new infrastructure-management protocols that use monitoring and prediction techniques based on state-of-the-art information and communication technologies; in addition to preventing serious accidents, a goal of the initiative is to minimize infrastructure life-cycle costs.

The CEATEC exhibit presented and demonstrated several component technologies developed in conjunction with this project: indexing technologies storage, compression, and cleansing technologies for large quantities of sensor data arising from the various types of sensors being developed; indexing techniques to reduce search times; an analytical environment for visualizing data and analytical results from infrastructure databases; and a monitoring system that uses a chip-scale atomic clock to guarantee time synchronization between multiple sensors even under conditions precluding the use of synchronization from GPS or wireless networks.

NEWS 2

Presenting voice-synthesis technology and SINET

Intra-University Research Institute Symposium 2017

The Intra-University Research Institute Symposium 2017, an event designed for researchers at NII and other intra-university research institutes across Japan to come together and describe their research to a broad swath of the general public, was held on October 8 at UDX in Akihabara, Tokyo. The slogan of the event was "Meet the researchers! An exhibition of intra-university research."

NII presented a poster and offered demonstrations highlighting the research of Junichi Yamagishi, Associate Professor in the Digital Content and Media Sciences Research Division, on voice recognition and voice synthesis technologies. Associate Professor Yamagishi has developed a technology known as *speaker adaptation*, in which recorded voices of multiple individuals are processed to create an *average voice*, which is then merged with the voice of a given speaker to produce a voice resembling that of the speaker. Whereas conventional methods require enormous volumes of voice data, recordings with total lengths of tens of hours or more for one individual, speaker adaptation can mimic a given speaker's voice given only 10 minutes' worth of voice data.

The demonstration featured Bit the Data Dog—NII's official mascot, familiar from NII PR material—on a large-screen display. Visitors used a microphone to pose questions to Bit, who used voice-recognition technology to interpret the questions and respond in appropriate ways as NII representatives provided explanations of the technology and introduced the researchers involved (top photo) In the "Researcher Talks" held in the main auditorium, researchers from various institutions discussed the compelling nature of research and introduced some of their latest

research projects. NII's contribution to this series featured Associate Professor Yamagishi delivered a talk titled "Digital voice cloning technology and its applications," which described research in which the voices of individuals who were losing or had lost their voices due to illness could be reconstructed by recording the voice of the individual and combining this with an average voice profile (bottom photo).

NEWS 3

NII exhibits "Platforms for Open Science Research Data" at the 19th Library Fair and Forum

19th Library Fair and Forum

At the 19th Library Fair and Forum, which ran November 7-9 at Pacifico Yokohama, NII presented a new academic information platform for open science in its exhibit on the topic of "Platforms for open science research data."

In recent years, the notion of *open science*, in which not only research articles but also data, software, and other resources are widely released and shared with the general public via platforms such as the Internet, has attracted attention as a new way of advancing research. The sharing of research data and other resources among separate institutions requires new open-science techniques for managing research data, including common naming conventions and storage formats. To this end, in April 2017, NII established the Research Center for Open Science and Data Platform (RCOS), under the leadership of Center Director Kazutsuna Yamaji of the Digital Content and Media Sciences Research Division. The center designs platforms for research data with a particular focus on three pillars of an IT infrastructure: *management*, *release*, and *search*.

For several days, the NII booth was staffed by RCOS researchers offering explanations and presentations on open-science research data platforms (photo).

NEWS 4

French Innovation Minister Vidal visits NII

Discussion of collaborations and future work in the field of information science

On October 2, Frédérique Vidal, France's Minister of Higher Education, Research, and Innovation, visited NII with a delegation of 14, including France's Ambassador to Japan, Laurent Pic.

NII Director General Masaru Kitsuregawa introduced Minister Vidal to a broad range of NII research initiatives in fields including big data, security, and artificial intelligence, as well as to enterprises such as SINET (Science Information NEtwork), an academic information network designed and operated by NII. Vidal also met with three NII researchers—Professor Shin Nakajima of the Information and Society Research Division, Professor Imari Sato of the Digital Content and Media Sciences Research Division, and Professor Katsumi Inoue of the Principles of Informatics Research Division—who described their latest research and discussed opportunities for collaboration with France.

Vidal also received a briefing on research initiatives at the Japan-France Laboratory for Informatics, an organization comprising six Japanese and French institutions, including NII and France's Centre National de la Recherche Scientifique (CNRS), from JFLI Japan Co-Director Professor Kae Nemoto of NII's Principles of Informatics Research Division, as well as presentations from researchers at the University of Tokyo and Keio University.

The visit was characterized by a spirited exchange of views, with Vidal asking how researchers tackled the challenge of research problems straddling multiple disciplines. Overall, it was an occasion to reaffirm the commitment on both sides to further collaboration in the field of information science (photo).

NEWS 5

Topics: Applications of Big Data, Control Theory, and Communications Infrastructure

FY2017 General public lecture series: "Frontiers of Information Science", Sessions 3, 4, and 5

On September 12, The National Institute of Informatics held Session 3 in its FY2017 "Frontiers of Information Science" lecture series for the general public, featuring Project Researcher Noriaki Koide of the Research Organization of Information and Systems (photo, left) delivering a lecture titled "Society as seen through big data: applications of web/WiFi big data".

Mr. Koide's research focuses on applying big data obtained from the Internet or Wi-Fi to tourism policy. The lecture discussed ongoing research and development efforts in several areas, including systems for visualizing the operating status of hotels and other accommodations and the tourism activities of visitors to Japan, techniques for counting tourists visiting Japan in real time from accumulations of Wi-Fi logging data, and methods for predicting future numbers of tourists visiting Japan.

"People are looking to use big data to make tourism institutions more effective," Koide explains. "Understanding big data obtained from the Internet and WiFi in real time and making good use of the data can help hotels or tourism businesses, as well as local governments, provide services and programs more efficiently."

Session 4 was held on October 18, with Associate Professor Masako Kishida of the Principles of Informatics Research Division (middle photo) lecturing on "The science of designing motion: how control theorists see and think about the world".

Control theory is used in almost every imaginable field, from automobiles and robots to medical care and transportation. Although the problems that arise in such broadly disparate systems might seem at first glance to have nothing in common, in fact they often exhibit common structures and properties. If one can understand their operating principles on fundamental levels and develop general methods and theories with broad applicability, then many problems can be solved simultaneously.

Research in control theory uses numerical models to identify and abstract the fundamental essence of various specific systems, analyze the movements of systems, and determine the techniques and conditions needed to make the system behave as desired. "When you describe things in terms of partial differential equations," Associate Professor Kishida explains, "even totally different systems can look the same. The world of control theory is all about manipulating the solutions of partial differential equations to design movements for actual objects in the real world."

Session 5 was held on November 16, with Associate Professor Takashi Kurimoto of the Information Systems Architecture Science Research Division (right photo) lecturing on "The infrastructure reflects the era: from the telephone to the SINET academic information network."



Associate Professor Kurimoto first traced the history of information communication technology and communication infrastructure as they have evolved through the ages: from land-line telephones, to pagers, to PHS devices and cellular phones, and eventually to the Internet. Kurimoto next discussed the SINET academic information network designed and operated by the National Institute of Informatics. SINET is an information and communication network that links over 850 universities and research institutions across Japan; since April 2016, the system has been operating on the SINET5 platform, which interconnected the nation via ultra-high-speed 100 Gbps channels and strengthened its overseas channels as well.

SINET is used to, among other purposes, share large-scale experimental facilities and supercomputers; it has become an important part of the infrastructure supporting academic research in Japan. Associate Professor Kurimoto, who is involved in SINET5 R&D and operation, described a series of initiatives to improve the capabilities of SINET through LAN virtualization technology and higher communication speeds.

SNS

"Hey, this is great!"

Hottest articles on Facebook and Twitter (September – November 2017)

National Institute of Informatics, NII (official) Facebook
www.facebook.com/jouhouken/

The 5th session of the public lecture series "Frontiers of Information Science" will be on November 16. This lecture will feature Associate Professor Takashi Kurimoto discussing how "The infrastructure reflects the era: from the telephone to the SINET academic information network". (October 31, 2017)

National Institute of Informatics, NII (official) Twitter
@jouhouken

News Release: Experiencing a quantum neural network in the cloud (November 20, 2017)

Bit on Twitter! Twitter
@NII_Bit

The day is almost here for "Managing research data in the era of open science"! Have you registered yet? Don't forget! (November 13, 2017)

* Some text edited/omitted.

Essay

The Essential Features of ERATO Projects in the Field of Informatics

Shin-ichi Minato
Professor, Hokkaido University Graduate School of Information Science and Technology [Title as of date of reporting]
Visiting Professor, National Institute of Informatics

The ERATO program of the Japan Science and Technology Agency (JST) ranks among the largest of all research funders in Japan (1–1.2 billion yen over 5.5 years) and boasts a history of more than 35 years; it is, without question, the flagship program of JST. Established for the purpose of training future Nobel laureates, it has chosen four or five projects each year from all branches of the natural sciences. In the 2000s, as it became clear that informatics was also an important discipline, the program also began selecting one informatics-related project every 1 to 2 years. I have been in charge of ERATO since 2009, or 3 years before the start of the Kawarabayashi ERATO.

The official names of ERATO projects have traditionally been taken from the names of the individual project leaders, a practice which reflects the extent to which the success of a project depends on the personal qualities and skills of its leader and the importance of new, independent thinking not available in existing organizations or disciplines. There is no fixed prototype for who an ERATO project leader should be. After all, we are looking to this person to do research that, 10 or 15 years down the road, people will recall and say, "Wow, that was really where everything got started."

ERATO projects in the field of informatics frequently involve specialized configurations that would be unthinkable within the traditional ERATO system. Conventional ERATO projects typically proceed by preparing a large-scale experimental apparatus or facility, perhaps one of only a few of its kind in the world, identifying a single star researcher to be the PI (principal investigator) and hiring large numbers of postdoctoral researchers on fixed-term contracts to follow the PI's instructions to conduct highly focused experiments as a team. However, in the field of informatics, and particularly for research on algorithms, there is no need for massive experimental equipment. Instead, all that is required is an expert team, consisting of some fixed number of brilliant researchers, and an environment in which the researchers can have deep discussions to ferment new ideas.

Because informatics projects are freed from the shackles of large-scale experimental facilities, we have outstanding young researchers all throughout Japan, many of whom we have already hired for regular employment. It would be difficult to gather a group of people like this on fixed-term employment contracts. Thus, in the Minato ERATO project that I led, we not only hired postdocs but also established loose collaborative relationships with talented researchers at universities and in companies across Japan whom we would gather from time to time for intense discussions—basically, we set out to create an open community of nomadic researchers. This type of structure would have been unimaginable within the existing ERATO system, but we were fortunate that JST quickly realized how important it was. In contrast, the Kawarabayashi ERATO project has restricted its sights to a younger generation of researchers (most in their 20s). Their goal is to build a community by attracting a large number of promising postdocs and graduate students who did well in programming contests and offering them specialized training before they pursue careers in academia, and they have been extremely successful.

In ERATO projects, for better or for worse, the leader has a large amount of dictatorial authority, and there are many cases in which a leader with strong convictions and a willingness to put them forward can succeed even with behavior that deviates from typical conventions. Thanks to our ERATO accomplishments, JST and Japanese Ministry are starting to understand that successful big-projects in the field of informatics can come in many different formats. I think ERATO has value as a practical proving ground for new proposed project structures appropriate for the era.

There is no fixed definition of success or failure in an ERATO project; the leader sets his or her own final goals. It is also the responsibility of leaders to think about how to wrap up the project and how to proceed with next steps. If the goal is for people in 10 or 15 years to look back and recognize the research as the birth of a new field, then we cannot have the team simply collapse and disappear the instant that they cross the finish line! It is important to proceed at an appropriate pace so that the baton can be passed to the next generation of leaders just joining the race.

My dearest hope is that, by passing the baton little by little, on many different occasions to a wide variety of people, we will be able to continue surging forward at a sustainable pace.

Future Schedule

January 30 | FY2017 General public lecture series: "Frontiers of Information Science", Session 7 The possibilities of online education: Analyzing learning logs for educational applications Speaker: Assistant Professor, Masako Furukawa, Information and Society Research Division For more information or to register, visit **http://www.nii.ac.jp/event/shimin/**

Notes on cover illustration

The cover image depicts steam jets suddenly erupting with great force from the head of a robot at the precise instant the robot, looking at a map, identifies the optimal route passing by a number of given points. Here the head of the robot plays the role of discovering algorithms for rapidly analyzing networks.

Weaving Information into Knowledge
NII

National Institute of Informatics News [NII Today]
No. 64 Dec 2017
[This English language edition NII Today corresponds to No. 78 of the Japanese edition.]
Published by National Institute of Informatics, Research Organization of Information and Systems
Address | National Center of Sciences 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
Publisher | Masaru Kitsuregawa **Editorial Supervisor** | Ichiro Satoh
Cover illustration | Toshiya Shirotani **Copy Editor** | Madoka Tainaka
Production | MATZDA OFFICE CO., LTD., Sci-Tech Communications Inc.
Contact | Publicity Team, Planning Division, General Affairs Department
TEL | +81-3-4212-2164 **FAX** | +81-3-4212-2150 **E-mail** | kouhou@nii.ac.jp

Search "NII Today"!

Bit (NII Character)

http://www.nii.ac.jp/en/about/publications/today/