NRI

Special Edition

Sharing insights without exposing data

- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

10.December.2018

Nomura Research Institute, Ltd.

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

Executive Summary

Data protection is a key issue amid mounting expectations vis-à-vis Big Data. NRI spoke to Dr. Alex Pentland, a global authority on data science and an international leader in the Big Data space, about how to best promote data utilization while protecting privacy and ensuring information security.

Alex 'Sandy' Pentland

Professor, Massachusetts Institute of Technology

Professor Alex 'Sandy' Pentland directs the MIT Connection Science and Human Dynamics labs and previously helped create and direct the MIT Media Lab and the Media Lab Asia in India. He is one of the most-cited scientists in the world, and Forbes recently declared him one of the "7 most powerful data scientists in the world" along with Google founders and the Chief Technical Officer of the United States. He has received numerous awards and prizes such as the McKinsey Award from Harvard Business Review, the 40th Anniversary of the Internet from DARPA, and the Brandeis Award for work in privacy.

Shigeki Hayashi

*Senior Corporate Managing Director
–Financial Technology Solutions, NRI*

Joined NRI in 1988. Initially assigned to PMS Development Department; later transferred to Insurance Systems Department. Promoted to General Manager of Project Development Department, Financial Solution Sector and then General Manager of Innovative Systems for Financial Industry Department. Seconded to Nomura Holdings in 2007. Returned to NRI in 2009 as General Manager of New Insurance Systems Marketing Department. Appointed Executive Officer and Deputy Division Manager–Insurance Solution in 2012. Promoted to Division Manager–Insurance Solution Division in 2014. Appointed Senior Corporate Managing Director in 2016. Division Manager–Financial Technology Solution since 2017.

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

Social Physics

Shigeki Hayashi: I'm honored to have the opportunity to speak with you.

I feel Japan's business environment is sort of deadlocked for several reasons. One is a fear that global platforms like Google and Amazon will put Japanese out of work. Another is strong social pressure stemming from rapid aging of the population, including challenges like rethinking estate planning and even how people work. Other issues include the distinctly Japanese tendency toward overkill in providing services and personal data protection laws that are so restrictive they're perversely anti-consumer.

I think the Social Physics and Network Intelligence that you propound may be the key to breaking through this deadlock.

To begin, could you talk about Social Physics as a driver of innovation?

Alex Pentland: The term "Social Physics" is very old. It dates back to around 1800, when the word "physics" was coined. But back then there were insufficient data and mathematics to be able to understand people's behavior. What I've done is to use new data resources and advanced AI and machine learning techniques to build better models of humans.

There are two main traditions for understanding people. One is the Adam Smith rational individual model, which assumes people individually behave rationally. The other is an Eastern tradition that sees people as the fabric of society. Interestingly, Adam Smith believed in the Eastern vision, not the rational individual. He said people were a mixture of both.

When we look at these huge Big Data sets about where people go, what they buy, how they behave, we find that people are indeed a mixture of the two. In an environment with good information, the best choices for individuals can be the best choices for society also. This is because of social learning. As an individual, my best strategy is to learn from others. It's the best way to see what works and what doesn't.

Hayashi: So individuals make up society and, through learning, behave like each

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

other.

Pentland: Yes, but the other side is that society needs innovation. Society needs individuals to try different things, not to all be the same. So human behavior combines individual searching with the leverage of social learning.

One key to making that combination really work is good information. People have to know what works for other people like them. Conversely, two things inimical to a properly functioning society are secrecy and advertising.

Hayashi: AI and machine learning work best when applied to voluminous data that don't change much over time. I assume they would have difficulty dealing with constantly changing data like the information flows and human decision-making you're talking about.

Pentland: Correct. Machine learning and AI, like economics, are based on the concept of equilibrium. They assume things are in a steady state. They also start out completely ignorant, without any knowledge of the subject matter. This is one reason AI and machine learning require so much data, which means they aren't very good at detecting trends and fads.

The new ways we're using to think about people and what they're doing can be much more efficient than conventional AI and machine learning, particularly for financial market trends, which rapidly change all the time.

Share insights not data

Hayashi: Another focus of your research is how to share data while ensuring data security and protecting privacy.

Pentland: Because AI and machine learning are all about data, data collection best practices are crucially important.

The natural way for humans to learn is by individually sharing outcomes with each other. This implies that machine learning requires data-sharing among people, but we

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

have to worry about privacy and security also. So I've developed an approach that preserves privacy and security while allowing insights gleaned from data, not the data themselves, to be shared.

First, the core idea is that instead of you giving me your data, I ask you questions. There are no copies of your data outside of your possession. You get to see not only the questions asked of your data but also who is asking and why. And you get to choose whether to answer.

Hayashi: Wouldn't it be a hassle for individuals to decide whether to answer every single question asked of them?

Pentland: Yes but you could join an institution that would make such decisions and act on your behalf.

Such an arrangement is analogous to having a financial institution manage your money. Just like you keep your money in a bank, you can have a data bank that holds your data.

Secondly, this whole process would be completely encrypted at all times. And, from a legal standpoint, it wouldn't constitute data sharing because your data would never be decrypted at any point.

Hayashi: Why is this completely new approach to data sharing now necessary?

Pentland: With cybercrime and cyberattacks set to inevitably keep increasing, everyone needs a system far more resilient than firewalls. There's a growing realization that fundamental security is at stake unless we adopt very different IT infrastructure. With the system we are proposing, you don't need a firewall or passwords. You can steal my encrypted data. It doesn't matter.

One critical point is that any proposal that requires changing everything is not feasible because it would impose too much of a burden on users. Under our proposal, all you'd have to do is answer questions if you want to. You don't have to. And you could continue using your outdated PC or whatever.

So our approach is easier from a hardware and software standpoint. It's easier from a legal standpoint also, because you'd know in advance exactly what questions will be

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

asked and that everything is legal.

Hayashi: Your proposal sounds good if we were creating a system from scratch, but with data already being shared, wouldn't it be difficult to transition to a new system?

Pentland: If data are being shared, there's nothing wrong with that. But you might decide to stop sharing and instead answer questions, which is what they're doing in Europe.

One of the major IT security risks today stems from data sharing. With so many players in the ecosystem, you really don't know what happens to data you've shared. If you give your data to a trusted entity, that entity may outsource data processing to a subcontractor that in turn subcontracts to another data processor. You don't know who these subcontractors are or what they're doing with your data.

The system we propose enables you to minimize such risks while also keeping precise track of your data.

How new data sharing method would be used

Hayashi: Specifically how would such a data sharing method be used?

Pentland: Say, for example, a bank wants to know if it's holding the same investments as other banks. If many banks own the same investments, that would be a risk for all of them. But banks cannot disclose such information to each other because it's their customers' data and their trade secret. With our method, banks could collectively find out the composition of their investments in aggregate without sharing customer data or divulging their secrets.

Hayashi: They could reduce everyone's risk while abiding by the law.

Pentland: That's right. By extension, you could ask these sorts of questions across the entire financial system to detect trends and potentially minimize systemic risk.

Every country today has financial and economic indicators calculated by the central bank and/or financial institutions but they're often not really critical indicators; nor are

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -
vol.293

they real-time. Governance and learning in the financial system are consequently not as good as they could be. Our approach is a way of building a much smarter, lower-risk financial system, enabling people to share insights without exposing data.

Hayashi: What kind of real-time financial data do you have in mind?

Pentland: One example of base data generally not used today is foot traffic in and out of stores.

Currently, store data are mostly monthly, weekly or daily sales data, which don't tell you things like how many people went into the store or where the customers who made purchases came from. We have spinoff companies that collect such data for real. They're setting up systems in shopping malls. Such data also tell you about the economic health of the city, which districts are prospering. And the data are collected daily, enabling you to see patterns of human behavior that drive the financial system.

Hayashi: And the data are aggregated, so there's no threat to privacy.

Pentland: Correct. This system is currently used for things like government planning, crime control and health planning. In Israel, it's used to promote greater innovation, particularly in transportation, by sharing insights to allow companies to do better. In China, the focus is on sharing data to achieve higher-quality, more inclusive growth.

The United Nations' Sustainable Development Goal (SDG) program is using these sorts of techniques to measure everything about society, even things like sustainability and justice. That's why I'm on the Board of Directors of the Global Partnership for Sustainable Development Data, which promotes innovative data utilization for the sake of sustainable development.

Hayashi: You said earlier that decisions that are good for individuals are beneficial for society also, provided they are based on good information. However, no matter how specifically survey questions are worded, some people may answer untruthfully or capriciously. How do you deal with that?

Pentland: That's a very hard question, but society has some answers for it. For instance, Japan, like every other country, has a national statistical office. The statistics it puts out are broadly trusted and generally accurate. Although some problems exist with statistics, they are known problems. I see overcoming such problems as

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

a long-term scientific project. Societies have put in place ways to know more about themselves. If those ways are flawed, citizens debug them and, if necessary, revamp them. People have been doing this now for two centuries in some places.

Hayashi: Social Physics had a similar genesis, didn't it?

Pentland: Yes, it did. I believe that the biggest barrier to a better society today is that data cannot be shared. Individuals' privacy and companies' proprietary ownership of data have made it difficult to ask certain questions. But these difficulties aren't insurmountable. With the methods we've developed, it's now possible to get insights and answers without risking personal or proprietary data.

International differences in data protection laws

Hayashi: When new data-sharing arrangements are used, a key issue is how to comply with different countries' laws. Isn't such compliance difficult when legal systems widely differ internationally?

Pentland: Yes. My research group's goal is to get similar compatible standards everywhere. Without this sort of compatibility, our data-sharing model's utility is limited.

This is why I put so much emphasis on working with not only the US but also the EU and China. If you can establish compatibility among a lot of the major players, things can happen. Japan also should pursue compatibility with the rest of the world.

Hayashi: The EU, China and the US seem to be moving in different directions in terms of data protection laws.

Pentland: I disagree. I was involved in helping to set up the EU's General Data Protection Regulation (GDPR) from the very beginning. It was realized that if Europe adopted certain standards, the US would have to follow suit because there's so much cross business.

The process is of course very different between Europe and the US. The European

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

legal system is based on Napoleonic law. The Europeans look at what could go wrong and take legal precautions against such adverse outcomes. In American law, you wait until there's actual harm and then try to remedy it.

Hayashi: How about China?

Pentland: The Chinese are of course very new to the whole idea of law but they actually have a law very similar to GDPR. My book on this subject was just translated into Chinese by the government and published in China. And we are receiving research funding from Chinese state-owned enterprises. I believe the Chinese have a great deal of interest in this area.

How to overcome psychological barriers to data sharing

Hayashi: In Japan, both companies and individuals seem to have a psychological barrier that needs to be surmounted with respect to data security. For example, we see strong resistance to moving data stored on a home PC or company computer to some unknown location in the cloud. This may be more of an emotional issue than a logical one.

Pentland: The problem is that people haven't taken to heart the idea that encrypted data can be safe and yet usable.

Hayashi: People have long been dependent on authorities like religion or the state. Japanese have a particularly strong tendency to defer to authority. I feel that what you're proposing is extreme democracy, where citizens, not the state, control their personal information. I doubt that Japanese would readily embrace such informational autonomy.

Pentland: In Japan, people are very willing to support society. They even take out their garbage at exactly the designated time.

If I were to say, "You should put your data into a data cooperative so everybody in society gets better healthcare," you'd do it. This is the idea of "data for good." To

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

gain people's understanding, you have to be very concrete. You can't just say, "Give me your data. We'll do something good." You have to say, "Give me the right to ask questions of your health data and we can make the hospitals this much better;" or "Give me the right to ask questions of your financial data so we can make the financial system much safer and more stable." Ask concrete questions for specific purposes.

In the US, this is known as informed consent. People have to know in advance what's being done and why. They have to have the right to withdraw. The EU's GDPR is basically the same.

Hayashi: What about China?

Pentland: In China, you have to contribute data to support the state whether you want to or not. But the same is true to some extent even in the US. If you don't contribute certain data, you forfeit important rights. So, for practical purposes, you have to provide certain data.

Companies' motivation for sharing data

Hayashi: Japanese companies also are starting to utilize Big Data, but they seem resistant to forming collaborations that involve data sharing.

For example, a railway company that issues its own brand of e-money has information on who gets on and off its trains at which stations, but there's a strong social taboo against using such information externally. If the railway and a major convenience store chain that also issues e-money were to share data with each other, they might be able to offer new services to the elderly. At present, however, the prospect of them doing so seems remote.

Pentland: This is the type of project we're doing in Senegal and Colombia in the aim of bringing together data from private companies and the government to be able to gain insights beneficial to society, most notably poor people and minorities.

Why do companies share their data with society as an open data resource? The

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

biggest reason is to be a good citizen that contributes to society by providing data as a public service.
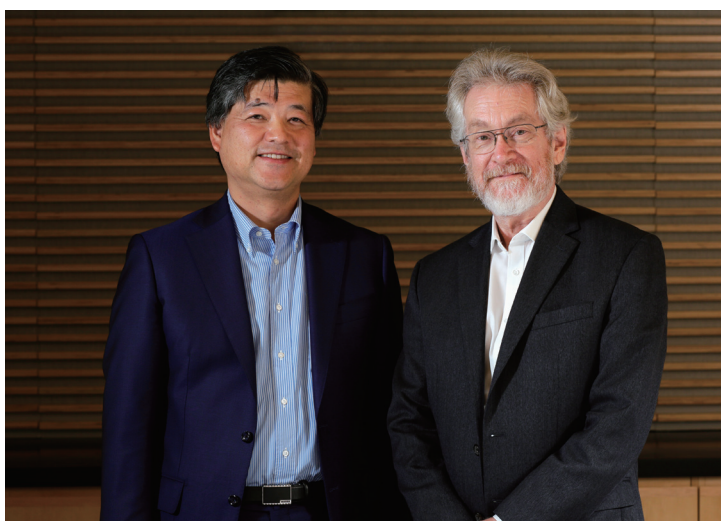
The data that the companies are sharing are aggregated geographically or otherwise. Such averaged data have little value to the companies. Citizens, however, can gain insights through fair and open access to such aggregated data. Meanwhile, the companies' data contributions familiarize people with the idea that sharing data and insights can benefit society. Contributing data is a way of both benefiting society and educating the public.

For instance, at a telecom carrier we've partnered with, the people who support our research are the marketing department because giving data to the country, to other companies, to the people, is a way of being seen as a good player in society.

Hayashi: Interesting! Do you usually interface with companies through their marketing departments?

Pentland: Sometimes. We're doing so with two other companies, a bank and transportation company, involved in the same initiative.

I believe this approach would be effective even in China. With the Chinese still in the process of deciding what their open data laws will be like, there's an opportunity for a company to say, "Here are some data. Use them for good." It's a way of making an argument to the authorities, to the government, that the law should be open to certain types of sharing.

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

Hayashi: It may influence how the Chinese government actually enforces the law.

Pentland: Exactly. You need to persuade regulators and legislators that data sharing can be societally beneficial, that they should not eradicate the good in an attempt to restrict the bad.

In Japan, the situation might be a little different than in Europe or the US but not so different. It helps to also have a famous university or other institution involved in supporting any future data-sharing initiatives together with MIT. We could promote discussion of how data could be used for good and what limits should be imposed.

Hayashi: While NRI is a for-profit company, I believe we earn the right to pursue profits by contributing to society. We intend to be at the forefront of the open-data discussion. Thank you for the pearls of wisdom you shared today.

Special Edition
Sharing insights without exposing data
- Interview with Alex 'Sandy' Pentland by Shigeki Hayashi -

vol.293

about NRI

Founded in 1965, Nomura Research Institute (NRI) is a leading global provider of system solutions and consulting services with annual sales above $4.4 billion. NRI offers clients holistic support of all aspects of operations from back- to front-office, with NRI's research expertise and innovative solutions as well as understanding of operational challenges faced by financial services firms. The clients include broker-dealers, asset managers, banks and insurance providers. NRI has its offices globally including New York, London, Tokyo, Hong Kong and Singapore, and over 13,000 employees.
 For more information, visit https://www.nri.com/en