NEC's Approach to Big Data

KAGEYASU Taichi, MIKAMI Sawako, KAWATSU Masato, NAKAO Toshiyasu, OKADA Yoshiaki

Abstract

NEC believes that the process of creating value from big data in order to lead to the solution of various issues should be regarded as consisting of three steps. These are sensing, (digitalization of the real world), analytics (analyses & inferences) and actuation (control & guidance). The application of big data to the innovative value chain that connects the three "MONOZUKURI" features of; "MAKE," "CARRY" and "SELL" will contribute to the creation of social values that will support a comfortable and secure social life. This paper reports on NEC's approach to the utilization of big data and discusses the measures actually taken to deal with it.

Keywords

Big data, three steps, sensing, analytics, actuation, invariant analysis, heterogeneous learning, recognizing textual entailment, rapid machine learning

1. Introduction

Increases in the speed/capacity of network technologies such as the Internet, the progress of various computing environments such as cloud computing and the dissemination of devices featuring various sensors and permanent network connections such as smartphones have recently brought about an explosive growth in the amount of distributed and stored data.

These systems include not only structured data like those handled by traditional databases but also non-structured data such as videos and texts, the generation and distribution of which are growing rapidly. Such large volume data of a wide variety is called "big data." This is expected to offer commercially useful knowledge to promote policy making. How to create new values from it is one of our current challenges.

In this paper, we report on NEC's proposed methodology for applying big data together with actual cases and measures taken to support it.

2. Market Trend

Domestic big data usage in Japan is expected to expand rapidly with an average annual growth rate of 27.0%. A 2014-2015 comparative survey of the approach taken by Japanese enterprises to big data indicates that those enterprises adopting measures are changing from advanced ones to general ones, with enterprises "taking certain measures" increasing from 28.6% to 32.0% and those "having no knowledge" decreasing significantly from 31.6% to 12.1%. These figures suggest that the interest of our customers is shifting from "becoming aware of the concept of big data" to those "utilizing big data".

Changes are also being observed in the volume of inquiries made by our customers. Since NEC started a project aiming at encouraging enterprises to utilize big data in 2012, we have been receiving thousands of inquiries every year. Interest in 2012 was oriented toward the concept of big data, such as "what is big data?" as well as to the technologies for handling it. Recent inquiries however indicate more interested in methods of utilizing big data in order to create value and in its practical application, such as in "how to utilize big data," "where to start using big data" and so on.

Below, we explain the approach taken by NEC in response to the needs of customers who want to create value from big data.

3. NEC's Approach

At NEC we consider it important to understand the process of creating value from big data by enabling the solution of various issues by taking three steps. These are "sensing (digitalization of real world," "analytics (analyses & inferences) and "actuation (control & guidance) (**Fig. 1**).

The "sensing" step generates data that expresses the information and events in the real world. "Analytics" extracts the properties and changes from the data and turns them into significant information. Finally, based on the information, the "actuation" step actuates the control of machines and the actions of persons in the real world in order to create value for customers.

For example, in the case of traffic jam prediction for automobiles, sensing corresponds to the collection of data such as the vehicle location and passage information. Analytics relates to the prediction of production, time and the length of congestions, and actuation to the control of signals for reducing congestion and the provision of congestion avoiding route information to vehicles. Data on vehicles passing and other information is collected from thousands of optical beacons installed nationwide. This allows the prediction of detailed traffic jam information, providing it to vehicles, and enabling the car navigation systems to provide routing information for avoiding traffic jams. Values such as reductions in travel time and fuel consumption may thereby be provided for drivers. At the same time, we can also create value for the whole of society, such as via efficient energy usage and pollution prevention. This approach characterizes NEC's concept of "utilization of big data in three steps."

One frequent cause of failure in creating expected value from the use of big data is in considering only one of the three elements of sensing, analytics and actuation. In the case of traffic jams mentioned above, if analytics alone is considered, the data would be insufficient and of low-quality, so it is impossible to predict traffic jam locations accurately. Also, if the prediction result does not lead to the driver's judgment being applied to avoid a traffic jam, it would be impossible to provide the value of decreasing the travel time. Instead of enhancing specific processes via ICT, the classification of the target into the three steps of "sensing," "analytics" and "actuation" and analysis of the process from data collection to control value creation as a unity will provide enhanced value. This approach characterizes NECs big data utilization platform.

It is naturally not so easy to plan, design and build these processes of data utilization. Consequently, we are preparing a consultation service called the "Big Data Discovery Program" aiming at big data utilization based on collaboration between the customer and NEC.

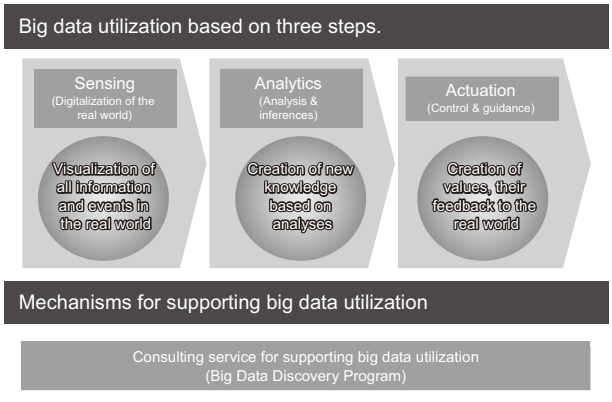

Fig. 1 Big data utilization based on three steps.

3.1 NEC's Sensing

NEC possesses a wide range of original sensing technologies. These include component parts such as; the piezoelectric vibration sensor, terminals such as the smart power distribution panel and media analysis technologies such as image/voice recognition. Solutions supporting social infrastructures such as satellite systems and ocean-bottom earthquake observation systems are also available. We introduce two examples of state-of-the-art sensing in the following.

NEC's was ranked the No. 1 vendor for three consecutive years in the testing of the U.S. NIST (National Institute of Standards and Technology) for its face recognition technology. This was achieved by overwhelming other competitors with its high search accuracy with the low recognition error rate of 3.1% and a search time of more than 3 million items per second in satisfying the national security requirement. Our technology has already been installed in more than 20 countries worldwide. For example, the Hong Kong Immigration Department introduced an immigration gate management system that can perform face recognition while each person is sitting in a car. The system succeeded in providing both the benefits of increasing the speed/efficiency of immigration control and in enhancing the rigorousness of personal authentication.

NEC is also acting as the main contractor in charge of the design, manufacturing and testing of the satellite system and major components of the GCOM-W (Global Change Observing Mission - Water "SHIZUKU") of the Japan Aerospace Exploration Agency (JAXA). NEC's support for the GCOM-W is almost total because we are also in charge of the initial post launch functionality checks and of the development of the earth based system for supporting the satellite operations from the ground. The GCOM-W continually acquires data on the precipitation, atmospheric water vapor, sea surface temperatures and sea surface wind velocities, at the global scale by installing the microwave radiometers that are used in the more accurate fields of climate forecasting and water resource management.

3.2 NEC's Analytics

NEC possesses world leading analytics engines. The present subsection deals with heterogeneous mixture learning, invariant analysis, recognizing textual entailment, and rapid machine learning technologies.

The heterogeneous mixture learning technology automatically discovers specific regularities of relationships among various kinds of data, and it switches the referenced rule according to the analyzed data. This enables highly accurate prediction and abnormality detection even with "data with variable regularities" that has been hard to analyze with the "traditional machine learning that discovers and references only one regularity" (**Fig. 2**).

The Invariant analysis technology automatically collects a large amount of time-series data from multiple sensors and automatically models the invariant relationships between sensors that work in normal conditions in the form of relation expressions. Comparison of the values predicted from the models with the real-time data makes it possible to detect "unusual" behavior. Invariant analysis technology can discover relationships between sensors that are often overlooked by experts because it is based on automatic extraction via machine learning.

Recognizing textual content is a technology for determining if two texts have the same meaning with high accuracy by considering the importance of each word in a text and the structure of each text, such as its subject and predicate. The analysis achieved by previous technologies was focused on matching and mismatching of the words contained in the two texts. Such a method was accompanied by incorrect analysis result issues such as expression of a same meaning with different words or expression of different meanings with a single word. Recognizing textual content contributes to improvements in the CS and the creation of new services. It enables the extraction of the documents to be managed from a large volume of documents such as mails and financial reports with high accuracy and without omission and to automatically categorize customer opinions.

rapid machine learning is a deep learning technology that learns and recognizes big data, including video, audio and text data without complicated processing involving human labor. It supports operations that necessitate human judgments and inferences. As this technology is capable of fast, highly accurate matching without the need to set rules, it is useful for example when selecting job applicants that match the employment information.

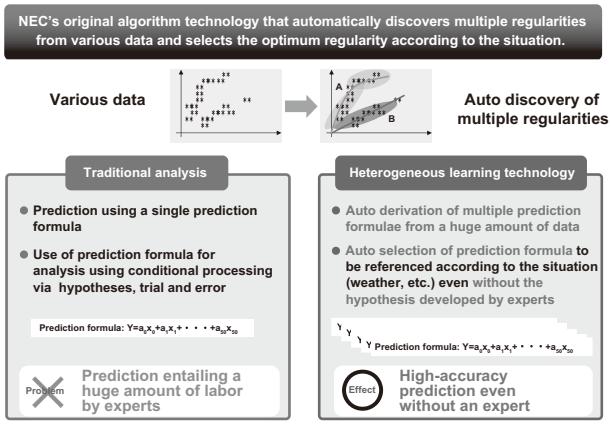

Fig. 2 Features of heterogeneous learning technology.

3.3 NEC's Actuation

NEC is conducting research into actuation that quantitatively identifies status and situations of humans and things in order to control them to be optimum status by adopting the future predictions obtained via analytics.

An example of the application of this technology is the demonstrative survey of a sewer culvert using a robot, created jointly with the City of Funabashi. The survey employs a piping management system, which uses a special survey robot and aims at the efficient reconstruction and extension of the life expectancy of sewer pipes. The robot automatically discovers the locations to be treated based on a model obtained via machine learning from image data so that the system can automate the survey work that has previously been done using human labor. This brings about a significant improvement in the efficiency of pipe status identification that usually imposes important constraints on the control of sewage facilities. It thereby enables efficient control and guidance by preventing incidents such as road damage due to pipe aging and it strategically narrows down the issues to be subjected to more detailed surveys.

3.4 Big Data Discovery Program

The Big Data Discovery Program is a consulting service that clarifies the purpose of big data utilization and the analysis method that supports the planning of specific remedial actions. Experienced consultants and data scientists at NEC support anxious customers that make comments such as "We don't know where to begin" or "How do we utilize the data in our business" so that they can quickly begin to utilize big data.

4. Approach and Actual Cases of Utilization in the Enterprise Domain

This section introduces actual cases of big data utilization in accordance with the conceptual features that characterize the issue, "Value chain innovation" connecting the "MAKE," "CARRY" and "SELL" (**Table**). With regard to the "resupply part demand prediction," "merchandise demand prediction," "plant failure prediction/monitoring" and "food/beverage de-

Table Examples of big data utilization in the enterprise domain.

Examples of big data utilization		Provided values	Sensing	Analytics	Actuation (Examples)
MAKE	Manufacturing plant failure monitoring*	Advancement/ optimization of operations	Sensor data	Invariant analysis	Repair/improvement of abnormal parts
	Prediction of demand for new products*	Improvement of product/service values	Purchase data	Heterogeneous mixture learning	Optimization of inventory holdings
	Equipment monitoring	Improvement of product/service values	Camera video	rapid machine learning	Elimination of abnormal parts
"CARRY"	Optimization of service parts*	Advancement/ optimization of operations	Shipment achievements, failure data	Heterogeneous mixture learning	Reduction of inventory holdings
	Prediction of degradation	Advancement/ optimization of operations	Sensors	Heterogeneous mixture learning	Repair/ improvement of abnormal parts
"SELL"	Prediction of demand for daily delivery of goods	Advancement/ optimization of operations	POS data	Heterogeneous mixture learning	Optimization of order quantity
	Analyses of customer opinions	Customer acquisition/ retention, sales promotion		Textual entailment	Improvement of manual countermeasures
	Compliance monitoring	Enhancement of information management, information crime/illegality	Mails, daily reports	Textual entailment	Warning to the relevant person

* Topics discussed in other papers in the feature pages of this issue.

mand prediction," these topics are detailed in other papers in the feature pages of this issue.

4.1 "MAKE"

In the case of petrochemical and steel manufacturing plants, the safe operation of the production facilities is an important issue. NEC's invariant analysis technology automatically visualizes the "typical" performance from a large volume of sensor data of the plant and thereby enables them to avoid incidents and, if one should occur, to reduce the lead time for identifying the cause.

The profits of drink and food sales companies are affected greatly by the settings of the initial production quantities of new products, but such predictions are very difficult because little data is available at the time of the release of a new product. NEC has solved this problem and predicted the sale quantity for a month after release with high accuracy. This was achieved using the heterogeneous learning technology in order to discover multiple regularities automatically from the shipment and actual sales data of new and past products.

One of the problems with equipment monitoring using camera video is the necessity of checking all of the video visually in order to survey suspicious individuals and articles exhaustively. When the rapid machine learning technology is applied, efficient equipment monitoring is enabled because any suspicious individuals or articles are identified automatically and these are then reported to the monitoring staff.

4.2 "CARRY"

One of the common issues experienced by companies in the manufacturing industry is the reservation of an optimum inventory of service parts. When the heterogeneous learning technology is applied, a highly accurate inventory prediction becomes possible. This is achieved by learning details of the latest service parts shipments, as well as the relationship between any associated information, such as the quantity and failure rate of the equipment and the predicted shipment demand for the future. This strategy has actually succeeded in reducing inventory sizes by about 20%.

The transport and physical distribution industries have difficulties in predicting the replacement timings of the consumable parts used in transportation equipment. The heterogeneous learning technology enables high-accuracy prediction of the degree of degradation of consumable parts based on data obtained by attaching sensors to them.

4.3 "SELL"

Issues of the retail industry such as in the convenience stores and food supermarkets include the disposal of daily delivered goods with short expiration dates, such as boxed lunches and prepared foods, and the loss of sales opportunities due to inventory stock depletion. When the heterogeneous learning technology is applied, it becomes possible for the sales volume of all daily delivered goods to be predicted with high accuracy by analyzing the relationships of a wide variety of data. This includes the past sales of the store, the amount of disposed goods, weather forecasting, calendar information and event/campaign information.

The main issues experienced in customer service offices and at the quality promotion HQs are how to reflect the opinions of the many customers that are optimally collected by their

contacts and call centers for the management of businesses including effective product development. We applied the RTE (recognizing textual entailment) technology to categorize texts automatically at the equivalent accuracy level to that of human capability so that we have succeeded in supporting our customers with the timely action planning.

It is vital for any enterprise to discover any information that would put compliance at risk (bid rigging, cartels, insider trade, etc.) from the huge volume of mails and business reports that are increasing day by day. However, this is difficult work if it must be done visually. RTE recognizes the value of textual entailment exhaustively and successfully identifies expressions containing suspicious features.

5. Conclusion

In the above, we introduced NEC's approach to the utilization of big data and discussed actual cases of utilization in the enterprise domain. Many of the introduced cases are those in which values are implemented by analyzing data that customers have already obtained. In the future it can be expected that greater value will be created by collecting field information that has not yet been compiled into data. This will be done by optimally sensing and combining the obtained data via actuation, automated control, etc. The recently announced "NEC Industrial IoT" solution also indicates that big data utilization based on three steps discussed above is capable of providing customers with greater value.

Under the "Mid-term Management Plan 2015," the NEC Group is globally promoting the "social solution businesses" that provide social values such as safety, security, efficiency and equality. NEC are working to create an efficient, sophisticated society in which people may lead brighter, affluent lives based on the fusion of advanced ICT and other innovative solutions.

Authors' Profiles

KAGEYASU Taichi
Assistant Manager
Big Data Strategy Division

MIKAMI Sawako
Assistant Manager
Big Data Strategy Division

KAWATSU Masato
Manager
Big Data Strategy Division

NAKAO Toshiyasu
Senior Expert
Big Data Strategy Division

OKADA Yoshiaki
Executive Specialist
Big Data Strategy Division

Information about the NEC Technical Journal

Thank you for reading the paper.
If you are interested in the NEC Technical Journal, you can also read other papers on our website.

Link to NEC Technical Journal website

Japanese

English

Vol.10 No.1 Special Issue on Enterprise Solutions to Support a Safe, Secure and Comfortable Life
- Value Chain Innovation Linking "MAKE," "CARRY" and "SELL" -

Vol.10 No.1
December, 2015

Special Issue TOP