Open Innovation
Big Data Integration Research Center

Director General   Koji Zettsu

The Big Data Integration Research Center conducts research and development of data collection and analysis aimed at making effective use of real-space information obtained from the environment and human society. We are also developing data mining technology that will integrate advanced environmental data with social data to analyze their cross-domain associations. This will facilitate model case studies of environmental influences on social systems such as transportation. We are also conducting R&D on methods that feedback the analysis results to sensors and devices in real space, and on sensor technologies that provide efficient and effective feedback. This will allow us to create, develop, and verify platform technology for implementing mechanisms capable of advanced situation recognition and behavior support, with the goal of optimizing social systems.

R&D of fundamental technology for cross-domain data collaborations

We are developing a data analytics platform for cross-domain data collaborations in order to exploit various forms of social big data and data acquired by various forms of sensing technology. We are also developing smart services for a smart sustainable society and constructing a data utilization testbed to support this development, and we are engaged in open innovation promotion activities involving regional demonstrations.

So far, we have developed a method for discovering localized association patterns from heterogeneous urban sensing data through the mutual optimization of association rule extraction and spatiotemporal clustering, and we have applied this method to the prediction of traffic risks in the times of torrential rain based on rainfall data (XRAIN; phased array weather radar), traffic data (congestion, accidents), and social media data (Twitter). In an evaluation experiment, we achieved an improvement of about 60-80% in accuracy compared with a conventional method (Apriori) for association rule extraction alone, and we demonstrated the efficacy of our method for association analysis of spatiotempo-

rally-skewed sensing data such as data related to natural disasters. Also, to develop support for transportation and mobility, we prototyped an application system for risk-adaptive map-based navigation where routes are searched according to the user's risk tolerance level based on traffic risks estimated from rainfall data. The results of an evaluation performed using a drive simulator with 30 test users showed that approximately 86% of users select alternative routes to avoid risks when presented with information about risk severity and driving costs (distance and time) (Fig.1).

On the other hand, we are also conducting R&D on the use of atmospheric environment data. In collaborative research with Nagoya University, we have developed a portable sensor that acquires data on personal exposure to air pollution (PM2.5), and we have developed a method for predicting personal exposure based on lasso regression analysis. In preliminary tests conducted from March through May 2017, we confirmed that it is capable of predicting personal exposure to PM2.5 with 80% accuracy (with an error range of ±15 µg/m3). We have also built a field demonstration system on the NICT integrated
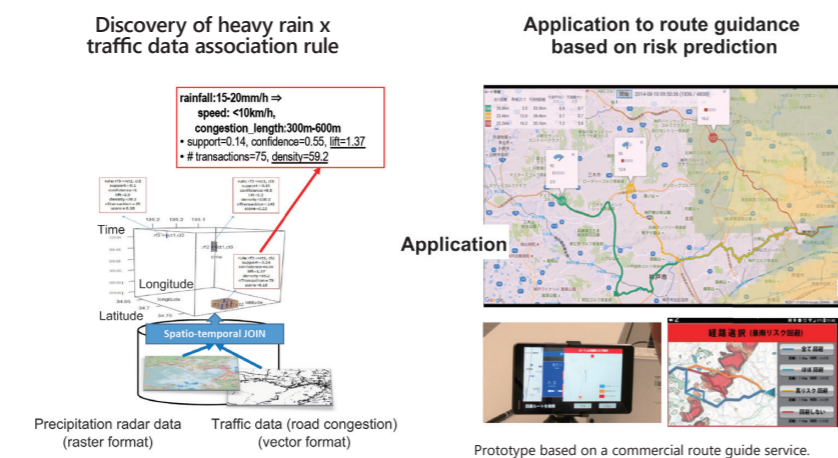
testbed that collects atmospheric environment data (PM2.5 levels, temperature and humidity, etc.) and health data (heart rate, autonomic nervous balance, etc.) and prepares digital maps including charts and comments showing how they are associated. A field demonstration trial with the participation of local residents aimed at finding healthy air was held in Fukuoka City between March 10 and April 8, 2018. A total of 69 people took part in this trial, including members of a running club, a regional open innovation group (One Japan in Kyushu), local students, and an IT volunteer group (Code for Fukuoka). The results of a questionnaire survey showed that most of them found the study very interesting, and included positive comments such as "It was a good opportunity to consider how my own vital data is affected by the environment," "It was interesting to see data about what I feel every day," and "It was fun using the latest sensor technology and learning about NICT's work."
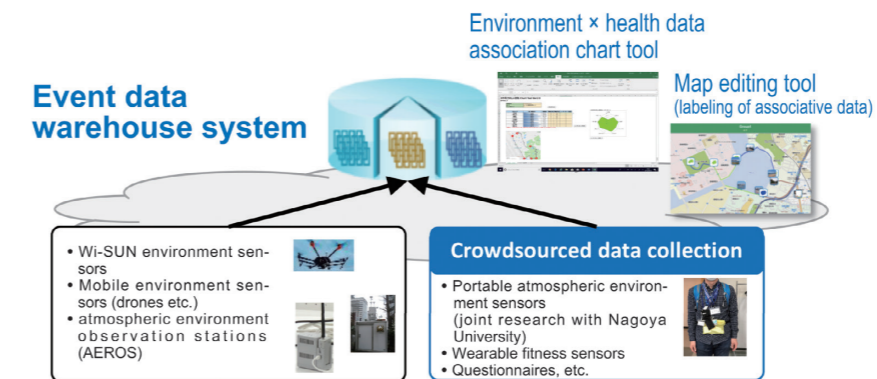
R&D of social big data analysis infrastructure

In the R&D of large-scale information integrated visualization technology, we are working on a 3D visualization method that facilitates understanding of the effects of day-to-day events in the real world, and how these effects spread across space and time. By focusing on position-related expressions in micro-blogging streams (of which Twitter is a prime example), our system associates tweets with location coordinates (either directly attached to the tweet or based on words such as the names of places or establishments). It then produces a multi-layered 3D visualization by checking for local events and wide-area events, using animation to show how this information changes over time. We are currently working on applying this system to the analysis and visualization of heavy rain risk by integrating it with rainfall data obtained by a weather radar (Fig.3).

Also, in the research and development of

data mining technology targeting non-textual data, we have developed a new algorithm that recursively searches tree-structured data converted from a time-series database based on periodic-frequency indicators that measure the temporal periodicity of item sets, and efficiently discovers all the item sets with partial periodicity. On the other hand, in large-scale graph data analysis technology, we have confirmed that the cost of communication can be reduced by 12% on average by running GraphSlice (an efficient distributed processing framework for social graphs) on Apache Spark in order to develop a scalable distributed graph database engine suitable for cloud environments.

Furthermore, in the R&D of human behavior analysis using social media data, we have succeeded in detecting tweets that strongly influence purchasing behavior with high accuracy (F-value of 0.53) based on multiple clues such as the relevance of the tweet contents, the proximity of tweet posting times, and the degree of closeness between users.

Fig.2 : Experimental demonstration of smart IoT for an environment and health monitoring system aimed at improving air quality

A scene at the Datathon held in Fukuoka in March to April 2018

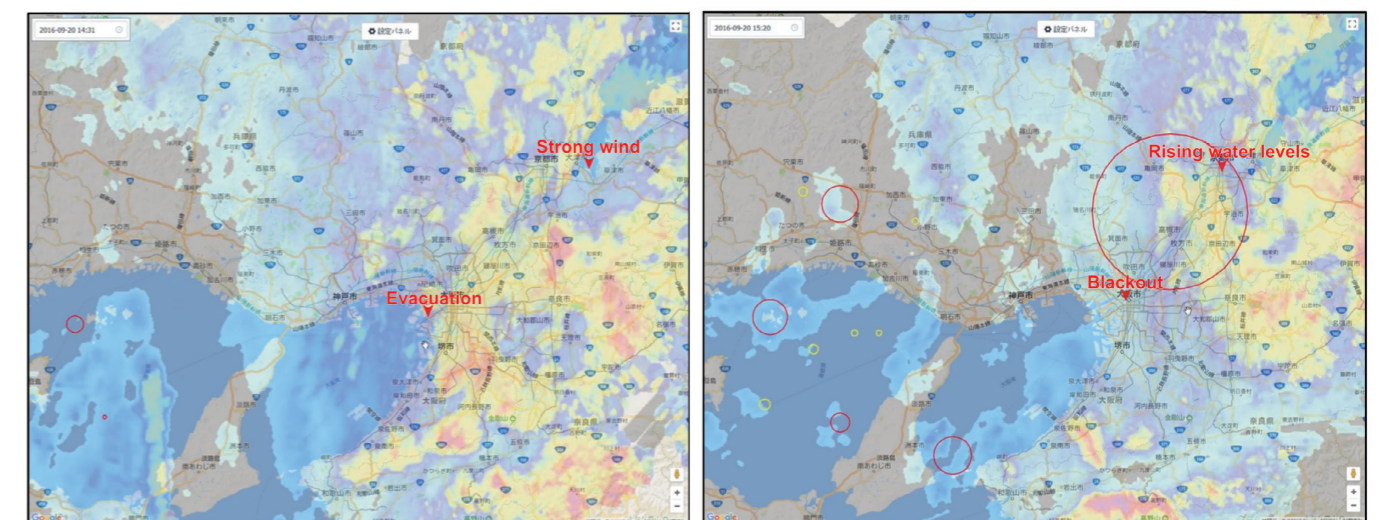Fig.1 : Mobility support based on a association analysis of heavy rain and traffic data

Fig.3 : Integrated visualization of heavy rain data (PANDA, XRAIN) and geospatial word-clouds based on social big data