Analyzing Disaster Risks from "Big Data"

Data Harmonization to Make a New Discovery

Various types of data – population in a certain area, number of offices, and disaster records of the tsunami that occurred in the past – are collected and accumulated by governments and international organizations. Such valuable data have been collected from the real world, based on a massive survey such as the population census. However, the accumulated data ("Big Data") are too enormous for us humans to easily understand the meaningful information contained inside. The computational ability of supercomputers may help us to make new findings from the enormous data. A researcher from Kyoto University has formed a group to derive disaster risks of a tsunami on the society by harmonizing data from completely different dimensions: environmental and socioeconomic perspectives.

Achievements

- The frequency of a tsunami reaching each area has been calculated based on records of tsunami (environmental data) for the past thousand years after dividing the entire country of Japan into approximately 1-kilometer grid squares.
- Disaster risks at each area have been clarified by multiplying the frequency of tsunami by the economic values (which have been calculated from socioeconomic data such as population, the number of firms, the number of workers and the number of travelers in each approximately 1-km grid square).

For the Future

- This result is a very interesting finding that clearly indicated how Japanese people have acquired land use over the years while avoiding natural disasters such as tsunamis.
- Future advancement of this research, in which data collected from different fields are integrated, will prompt us to acquire helpful knowledge not only in disaster countermeasures but also in various fields such as education, healthcare and welfare, import and export as well as economic reform.
Analyzing Socioeconomic Risks Hidden in Regions with Supercomputers

In order to clarify socioeconomic risks hidden in regions, the entire country of Japan was divided into areas of approximately 1-kilometer grid squares based on digital national land data provided by the Ministry of Land, Infrastructure, Transport and Tourism. Information on each area was combined with data on tsunami for the past thousand years, provided by the National Oceanic and Atmospheric Administration. This led to an estimate of the frequency of a hazardous event, namely how frequently a tsunami with a certain height will reach the specific area.

A Very Rare Occurrence of a Tsunami Once in a Few Centuries: Calculating Such Frequency

A huge tsunami is a very rare phenomenon, and a great disaster like the Great East Japan Earthquake rarely will occur every few hundred years. The probability of a huge tsunami reaching a specific area will be mistakenly estimated as zero, if there has been no occurrence of a huge tsunami during the observation period used for the calculation. To solve this problem, a mathematical method using the "likelihood" (maximum likelihood estimation) was utilized. Changing the shapes of probability distribution in terms of tsunami frequency, a wide range of calculations was performed by presuming the probability as high or low; and then the most accurate values were found. It would take approximately a week to perform these calculations once for each of over a hundred thousand areas in Japan by using parallel computers at a laboratory. However, the ability of supercomputers at the Institute of Statistical Mathematics has brought the calculation time down to just 85 minutes at minimum.

First, data for each of approximately 1-kilometer grid squares were obtained from the government statistics, such as the population and workforce, as well as the numbers of offices and tourists. The data numerically quantifies the economic value of the area. Then the value was further multiplied by the tsunami frequency to calculate socioeconomic damage (See Figure 1).

Analyzing Big Data to Learn the Socioeconomic Risks of a Tsunami

Figure 2 shows the visualization of computed socioeconomic loss by a tsunami which will be suffered if no countermeasures are put in place inside each area throughout Japan.

Thus the present statistical approach, which is different from conventional numerical simulations for tsunami forecast, was applied to the Big Data. This approach estimated risks of a very rare disaster that may occur every few hundreds of years. If this method had been applied before the Great East Japan Earthquake in 2011, risks hidden in the affected areas could have been revealed.

There is a history that socioeconomic and environmental data have been collected and accumulated to be studied in different fields of science. If we further promote the use of supercomputers continuously to analyze Big Data and harmonize various types of Big Data, we would probably find solutions for the future, such as "Precaution measures again tsunami are necessary in areas analyzed to have a high frequency of tsunami with great economic loss".

---

*1 National Oceanic and Atmospheric Administration (NOAA): http://www.noaa.gov
*2 Occurrence of tsunami has been calculated using the altitude of each area alone. Damage could be prevented by measures such as breakwaters.