EARTH MATERIALS SCIENCE IN A DATA-DRIVEN PARADIGM

The 21st century is often called the era of the fourth scientific paradigm, which is characterized by data-driven science and that follows on from the first three paradigms of the experimental, theoretical, and computational sciences. Data-science methodologies that utilize sophisticated mathematical techniques to process available datasets are often called machine learning (ML) and artificial intelligence (AI). Such strategies have already been utilized as practical tools for applications in industry, society, daily life and, increasingly in recent years, in the natural sciences and in engineering.

Our research group at the Japan Agency for Marine-Earth Science Technology (JAMSTEC) has introduced these data-driven approaches to Earth materials science by forming collaborations not only with information scientists but also with a wide range of natural scientists, including geophysicists, astronomers, and medical scientists. This article briefly introduces our recent research in three areas (Fig. 1).

The first area of research is the super-resolution measurement. In April 2019, the first direct image of a black hole was released by the Event Horizon Telescope project. Our collaborators contributed to this historical accomplishment by developing the mathematical super-resolution technique. We used a similar method, using GPS data, to image in detail the spatial distribution of slip on a subduction boundary. Recently, we proposed a novel super-resolution technique for processing images derived from laser ablation inductively coupled plasma mass spectrometry (LA–ICP–MS) (Aonishi et al. 2018). In this approach, super-resolution is accomplished by forward modeling the diffusion of ablated aerosol particles and non-negative signal deconvolution.

The second area of research is the treatment of high-dimensional data. Geochemists have compositional data from several to several tens of dimensions (e.g., elements or isotopes). However, it is difficult to understand the overall structure of high-dimensional data by conventional graphical approaches using several elements that have been selected based on the scientists’ intuition and experience. In data science, several powerful methodologies that utilize high-dimensional data have been developed for discrimination and feature extraction and these have been applied to various problems, such as medical diagnostics. We have introduced such methods to geophysical and geochemical data analyses, such as the geochemical discrimination of global magmatic tectonic settings, to find the optimal combination of important elements for discrimination and to find the most representative magma composition of each setting that can be objectively obtained (Ueki et al. 2018).

The third area of research is the integration of numerical simulation with data observation. In geoscience, numerical simulation is becoming increasingly important for understanding the complex structures and behaviors of natural systems. Data assimilation techniques, which combine large time-series observation data sets with fluid-dynamics simulations, have been developed for the atmospheric sciences and utilized in weather forecasting, estimating past climates, and predicting future climates. We extended a data assimilation technique to recover
past dynamics from a spatial data set, which is a common data type in Earth materials science, rather than from the time-series data. Further, we applied this method to estimate the pressure–temperature–time (P–T–t) path from the compositional zoning of a garnet grain (Kuwatani et al. 2018).

Other examples of our research include a study by Yasumoto et al. (2018) involving the use of probabilistic clustering and regression techniques to achieve the precise quantitative compositional mapping of a sample by an electron microprobe analyzer (EPMA). In another study, Yoshida et al. (2018) proposed a new type of material-transfer process during subduction metamorphism using a machine learning technique that extracts common topics from a collection of documents.

As seen above, data-science methodologies can be applied to a wide variety of different targets because they share a common mathematical structure. This demonstrates the importance of collaboration between different scientific fields. In Earth materials science, we expect the use of advanced data-science methods to become commonplace in the future. However, at present, we are at the stage of trial and error and do not have a systematic way of determining the most appropriate method for a particular dataset and objective. Thus, the challenge has just begun. In the coming years, we will continue to create innovations via interdisciplinary collaboration.

Tatsu KUWATANI (JAMSTEC)

REFERENCES
Yasumoto A and 5 coauthors (2018) A rapid and precise quantitative electron probe chemical mapping technique and its application to ultrahigh-pressure eclogite from the Moldanubian Zone of the Bohemian Massif (Nové Dvory, Czech Republic). American Mineralogist 103: 1690-1698