Hitachi Materials Informatics Analytics Platform Assisting Rapid Development

Yoshihiro Osakabe, Akinori Asahara, Hidekazu Morita
Hitachi Ltd.
1-6-1, Marunouchi, Chiyoda-ku, Tokyo 100-8220, Japan
{yoshihiro.osakabe.fj@hitachi.com}

CEUR-WS.org/Vol-2600/short5.pdf

Abstract

The data science platform for materials developments is demonstrated. Due to the recent great advances in artificial intelligence, it becomes more realistic that the industrial application of materials informatics (MI) which is the data-driven approach to discover and investigate materials characteristics. However, it is not quite easy for materials manufacturers to set up MI analytics environments without any help. Therefore, we provide the user-friendly cloud-based IT platform for non-experts of IT enabling materials scientists in R&D departments to analyze their experimental data effectively for rapid developments.

Motivation

Product developments require significant time and costs to find the optimal combination of ingredients and parameters. Materials Informatics (MI) is an emerging study field based on the both informatics and materials science, with the goal of greatly reducing the resources and risks required to discover, invest, and deploy new materials (Curtarolo et al. 2013). Recently, artificial intelligence (AI) has improved the MI performance, thus the experimental candidates can be narrowed down without unnecessary trials and errors before its actual experiments to discover or create new materials with yet-to-be realized properties. In fact, US government has invested over $250 million to assist MI projects (Materials Genome Initiative 2011). The Novel Materials Discovery Laboratory in EU also opens new oppotunities to investigating MI by delivering analytics tools and open access repository of materials data (NOMAD Laboratory 2015). According to such outreach activities, there has been heavy demands of materials manufacturers for introducing MI-powered methodology into their R&D processes to increase their industrial competitiveness, and the number of startups in MI analytics services is increasing.

In figure 1, the concept of this demonstration is illustrated. In many cases, it is difficult for materials scientists to select suitable preprocessing method and effective algorithm to solve their problems because they do not have enough informatics knowledge, which means that they need the supports of informatics experts (data scientists). Their relation can be understood as that between a runner and his escort, thus this phase can be regarded as an "accompanying phase." Though this service style is common, it may remain the possibility that the informatics experts can not exactly understand the characteristics of target materials and obtain the know-hows materials scientists have. This problem will be solved if materials scientists can reach analysis results by themselves without the excessive IT and analytics knowledge. That phase can be understood as a "self-managing phase," and the MI analytics services should be shifted to that phase from accompanying phase for scaling up and rapid prototyping. It suggests the need of the informatics expert alternative and one-stop platform for storing, analyzing data and visualizing analysis results.

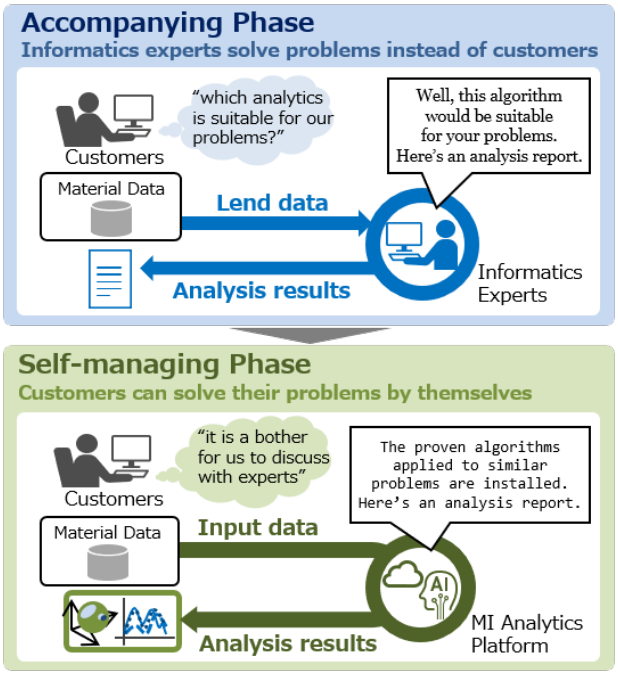

Figure 1: The concept of MIAP (MI Analytics Platform)

Copyright © 2020 held by the author(s). In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, F. van Harmelen (Eds.), Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice (AAAI-MAKE 2020). Stanford University, Palo Alto, California, USA, March 23-25, 2020. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Materials Informatics Analytics Platform

We have developed an IT platform of MI, called **Materials Informatics Analytics Platform (MIAP)**, for R&D teams of various manufacturing companies. This platform brings together all data into one place to make it easier for researchers to access and custom machine learning algorithms by themselves without any additional help of informatics experts. In fact, it includes the functions that support almost every step required for MI analytics.

Functions

MIAP is a cloud service thus the user interface is accessible via common web browsers. It mainly includes three functionalities; storing, analyzing and visualizing. In the following, their details are explained.

1. **Storing**
   In this platform, all input and output data is stored in PostgreSQL database servers. Various file types are acceptable; CSV, Microsoft Excel, NetCDF and so on. Graphical user interface (GUI) is utilized to upload and import data into databases. At the same time, it also receives SQL queries to manage data tables directly with the implemented query editor for complecated operations. With GUI for example, users can define the data type of each column without typing any complicated SQL queries.
2. **Analyzing**
   In general, MI problems are interpreted as regression and classification tasks. Thus, it supports the various well-known machine learning algorithms such as Random Forest (Breiman 2001), Gaussian Process (Rasmussen and Williams 2006), Support Vector Machine (Burges 1998) and Gradient Boosting (Friedman 2001). It also makes predictions and optimizations possible. In addition, one of the MIAP unique features is the implementation of the AI-based best practices of efficient methodologies for individual customers, which contributes to reduce their experiment iterations. In most cases, once users have developed their best practices, they can easily and repeatedly apply the same method to new data by themselves.
3. **Visualizing**
   It provides basic visualization tool to plot data in database by selecting target column and graph types (bar, line and pie graph). To check the learning performance, users only have to click on automatically generated truth-prediction scattering graphs. In addition, it provides the original UI tool derived from a Geospatial Information System (GIS) tool that draws animation along with time in 2D and 3D graphs. It means that users can see the time evolution of materials properties.

Demonstration

The usage of MIAP is demonstrated by taking an example of the search for an optimal recipe that improves the material properties of a ready-made product.

First, collect data accumulated in the process of making target product. Second, upload them to the MIAP database. MIAP automatically converts files with different formats into a predetermined format using KNIME, the open source software (KNIME 2019). Next, user specify the target material property expected to be improved as an objective variable, and other properties are set to explainable variables. After selecting algorithm for modeling and entering the output table name for the current attempt, the learning is started by pressing the execute button. These operations are very simple because almost all users have to do is just clicking on corresponding tabs. As shown in Figure 2, users can recognize the results are listed on results view window when the calculation is finished. Because the automatically generated truth-prediction scattering is shown with common indicators to score learning performace such as Root Mean Squared Error (RMSE) and correlation coefficient, it is possible for users to judge whether the learning is succeeded or not. After users can obtain well-trained model via iterational attempts, they can predict the target material property with candidate recipes to narrow down before the actual experiments for new products. In this way, MIAP assists to find the optimal recipes of ingredients or parameters, which contributes to reduce materials development resources.

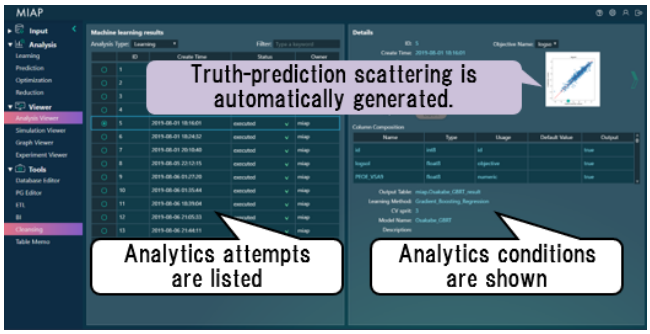

Figure 2: Screen capture of checking learning results

References

Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.

Burges, C. J. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2(2):121–167.

Curtarolo, S.; Hart, G. L.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; and Levy, O. 2013. The high-throughput highway to computational materials design. *Nature materials* 12(3):191.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 1189–1232.

KNIME. 2019. KNIME, https://www.knime.com/.

Materials Genome Initiative. 2011. In *Website of Materials Genome Initiative (MGI)*, https://obamawhitehouse.archives.gov/mgi.

NOMAD Laboratory. 2015. In *Website of Novel Materials Discovery (NOMAD) Laboratory*, https://nomad-coe.eu/industry/interaction-with-industry.

Rasmussen, C., and Williams, C. 2006. Gaussian processes for machine learning, model selection and adaptation of hyperparameters, chapter 5.