Materials Technology Development Using Materials Informatics

Masashi Hatada, Shinji Kawamoto, and Takeshi Kizaki

Materials informatics (MI) is a development approach that actively utilizes computer science, such as computational and data science, in addition to conventional experimental and theoretical science. We have positioned MI as one of the fundamental materials technologies and applied it to a wide range of products, which has led to various outcomes. This paper introduces the most frequently used MI methods and presents a case in which we succeeded in significantly shortening the development period of cable sheaths by utilizing data science.

1. Introduction

In recent years, materials informatics (MI) has attracted attention in the field of materials technology development. In October 2012, the Massachusetts Institute of Technology (MIT) and Samsung announced that they had developed a solid electrolyte with very high ionic conductivity[1]. However, a major Japanese automobile manufacturer had been working on almost the same substance for several years and had patented it in 2011. Furthermore, the development was done without any experiments, using only data and calculations, which had a huge impact on materials researchers around the world. This led to a global trend toward MI, and national projects were started one after another. The Japan Science and Technology Agency (JST) proposed the promotion of MI in 2013[2], and three ministries have since started national projects. MI is now recognized as a fundamental technology that supports the development of materials technologies, and its full-scale introduction is progressing at various companies.

Because of the wide range of our business domains and the many different types of materials that we handle, it is necessary to strengthen the fundamental materials technology that will lead to increased competitiveness of our products. Therefore, we have been working on MI since 2018. Although it is still in the process of development and no formal definition exists, we define MI as not only new materials development but also process optimization. It has also become clear that computer science is only a tool, and domain knowledge such as expertise, knowledge, experience, and know-how in the domain where MI is applied is important in order to use MI as a tool. For that reason, we have developed activities that lead to more practical results by providing opportunities for materials engineers themselves to learn data science and by constructing computational environments. By using data science to design the experimental conditions for compound optimization, the time taken to develop the cable sheaths was significantly shortened.

Material Technology R&D Department

2. Overview of main MI methods

Among the MI methods, "Bayesian optimization," "regression and classification modeling," "model interpretation," and "persistent homology" are frequently utilized in materials technology development. These four methods are outlined in this section. In the cable sheath development described later, the two main MI methods that we used were Bayesian optimization and model interpretation.

2.1 Bayesian optimization

This method optimizes the experimental conditions sequentially by using past experimental data to find useful conditions for the next experiment[3,4]. Figure 1 shows an illustration of Bayesian optimization. In Figs. 1(a)-1(c), the horizontal axis shows the parameters, the vertical axis shows the objective variables, and the blue-circled plot points show the experimental results. Bayesian optimization first creates a probability model from the experimental data and estimates the mean and variance over the search range. The results are shown in Fig. 1(a). The blue line shows the mean value and the gray area shows the variance. Based on these results, an index called the acquisition function is calculated to represent the value of the experiment. The red line in Fig. 1(b) shows the acquisition function. The next experiment is conducted under the conditions in which the acquisition function is the highest, and the results are added to update the probability model, as shown in Fig. 1(c). Bayesian optimization is the process of searching for the optimal conditions by repeating these steps.

Conventionally, the relationship between each parameter and the objective variable is determined by repeating a large number of experiments while using a technique called "one factor at a time (OFAT)," which determines the change in the objective variable when only one variable is changed. However, this method tends to result in a localized search. Bayesian optimization, on the other hand, can conduct a relatively global search by appropriately setting
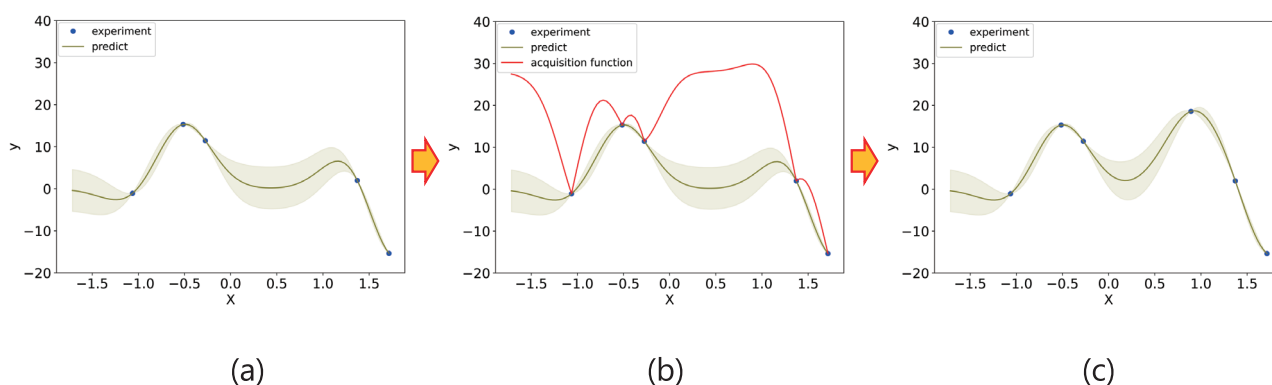
(a)　　　　　　　　　(b)　　　　　　　　　(c)

Fig. 1. Summary of Bayesian optimization.
　(a) Creation of probability model, (b) Calculation of acquisition functions, (c) Experiments and updating of probability model.

the acquisition function.

In Bayesian optimization, initial data are necessary to create a probability model. Several experimental design methods have been proposed for collecting initial data, such as the use of orthogonal tables and D-optimal design[5)6)]. The D-optimal design is expected to obtain more informative initial data than the orthogonal table because the number of levels is fixed in the orthogonal table, while the D-optimal design does not require a fixed number of levels. The D-optimal design also has the advantage that the number of data to be collected can be freely set, making it suitable for experiments that are often limited by time and cost.

2.2 Regression and classification models

If there is already a large number of data, optimization can be achieved by inverse analysis of the regression or classification models that have been created. Inverse analysis means that the optimal conditions are found by repeating the forward analysis in the search space. In the case of optimization by inverse analysis, it is important to consider whether the search conditions are within the adaptive range of the model. In addition, it is necessary to accumulate the know-how for each material (product) to be analyzed, because it is also important to know what kind of features should be used in creating the model.

2.3 Model Interpretation

In general, there is a tradeoff between model accuracy and interpretability. Models with usually high prediction accuracy, such as deep learning, tend to be blackbox models. In MI, interpretation is often more important than prediction. Therefore, models with low prediction accuracy but high interpretability, such as decision trees, are also used. Recently, XAI (explainable and interpretable AI) techniques such as LIME, SHAP, partial dependence plots, and permutation importance are being increasingly used[7)]. The important point is that the results obtained are only interpretations of the model, not causal relationships. The validity of the interpreted results should be confirmed by verification experiments.
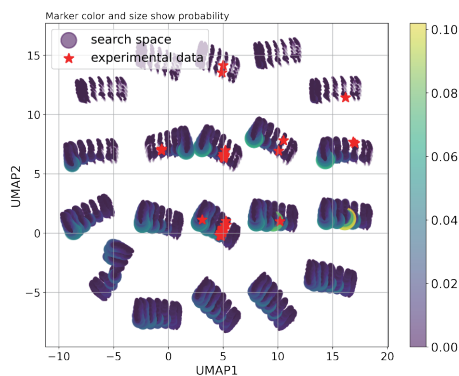
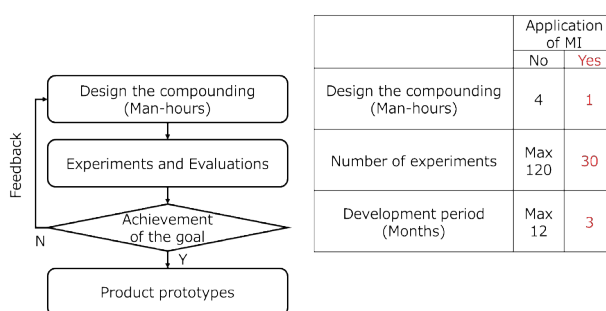**Fig. 2. Visualization of the acquisition functions in the search space.**

		Application of MI	
		No	Yes
Design the compounding (Man-hours)		4	1
Number of experiments		Max 120	30
Development period (Months)		Max 12	3

Fig. 3. Comparison of compound development with and without MI application.

2.4 Persistent homology

This is a mathematical analysis method for characterizing the topological structures such as connected components, holes, rings, and voids, using the mathematical concept of topology. For example, it was used to analyze the hierarchical relationship of ring structures in the atomic arrangement of glass, and revealed structures that could not be discovered in the past[8)9)]. Structural information obtained from various image data and molecular dynamics is often used in materials development, and expectations for persistent homology are increasing.

3. Optimization of sheath material composition

3.1 Design of experiments for compound optimization

One of the cable sheath materials that we were manufacturing posed the problem that its tensile strength and elongation did not have a margin to the standard. There were also concerns that the supply of base rubber might cease, so it was necessary to quickly establish an alternative compound that would improve the tensile strength and elongation. For example, the sheaths consisted of 14 different compounding agents, and testing each of them at three different levels would have resulted in approximately 4.78 million (3^{14}) combinations, which would have been a huge search space. In addition, the need to improve the tensile strength and elongation while meeting various specifications made the development of this compound a multi-objective optimization. Some of these properties were trade-offs, and in the conventional method, the experimental design had to be based on the experience of the engineer in charge, and a large number

of experiments had to be conducted. Therefore, we decided to utilize Bayesian optimization for the experimental design.

Figure 2 shows the acquisition function (the value of the next experimental condition) when analyzing experimental data at a certain time. The multidimensional experimental conditions are reduced to two dimensions using the UMAP method and then visualized. The location of the circle shows the experimental condition, and the size and color of the circle show the size of the acquisition function. Asterisks show the experimental conditions of the experimental data. It is possible to check the distribution of the experimental conditions over the search space. The condition with the highest acquisition function is the next experimental condition. The results are used for modeling again, and the acquisition function is updated. This process is repeated for optimization. As a result, the optimization was successfully conducted in approximately 30 experiments. Figure 3 shows the results of a comparison between the conventional method and the method using MI. It can be seen that the development period was significantly shortened by using MI. In addition, during the Bayesian optimization process, some conditions were presented that had not been considered by the engineers in charge, and when they tried the compound, it was found that relatively good characteristics could be obtained.

In this way, we reaffirmed that MI can be used as a tool to give engineers new insights and opportunities to consider.
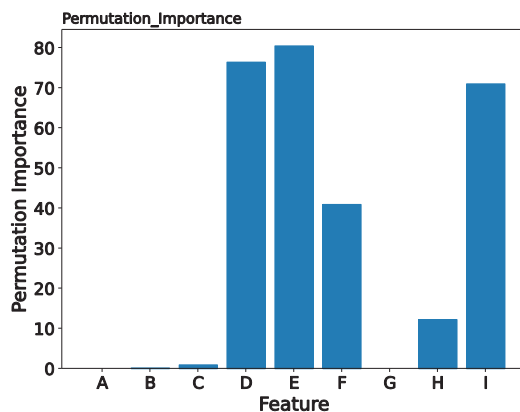
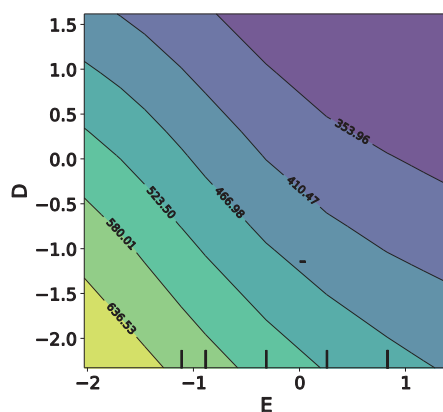Fig. 4. Identification of critical factors for compounding materials.

Fig. 5. Correlations between D and E compounding agents and
elongation considered by the model.

3.2 Model interpretation

In the compounding optimization described above, the interpretation of the probability model used for Bayesian optimization was also conducted in parallel. Figure 4 shows the results of interpreting the probability model with elongation as the objective variable in terms of permutation importance. This figure shows the magnitude of the impact of each compound on the predictions of the model, indicating the large importance of compounding agents D, E, and I. The relationship between compounding agents D and E on the elongation was checked using a partial dependence plot. The results are shown in Fig. 5. The correlation between the target features and the objective variable is obtained by eliminating the effects of features other than the one targeted by the marginalization. It can be seen that the elongation is negatively correlated with compounding agents D and E. These results agree with the domain knowledge and support the effectiveness of MI.

4. Conclusion

The MI method was used to develop sheath materials. Bayesian optimization was used to design the experimental conditions, which significantly shortened the development period. The analysis results were visualized using UMAP, a dimensionality reduction method. In addition, model interpretation was conducted using permutation importance and partial dependence plots. We will continue using MI to help develop more efficient material technologies.

Reference

1) Y. Mo, et al., First Principles Study of the Li10GeP2S12 Lithium Super Ionic Conductor Material, Chem. Mater. 24(1), pp. 15–17, 2011.

2) Center for Research and Development Strategy, Materials Design by Digital Data Driven Method, CRDS-FY2013-SP-01, 2013.

3) M. Nakayama, et al., Materials Optimization for All-solid-state Li Secondary Batteries by Materials Informatics, Surface Finishing Society of Japan, No. 2, pp. 84–90, 2021.

4) K. Matsui, et al., Introduction to Bayesian Optimization and Its Application to Material Engineering, Mater. Jpn., pp. 12–16, 2019.

5) S. Pruksawan, et al., Prediction and optimization of epoxy adhesive strength from a small dataset through active learning, Science and Technology of Advanced Materials, Vol. 20, 1010–1021, 2019.

6) T. Morishita, et al., Initial Sample Selection in Bayesian Optimization for Combinatorial Optimization of Chemical Compounds, ACS Omega, 8, pp. 2001–2009, 2023.

7) Y. Kameya, The Past and the Future of Explainable AI Techniques, IEICE Fundamentals Review, Vol. 16, No. 2, pp. 83–92, 2022.

8) Y. Hiraoka, et al., Hierarchical structures of amorphous solids characterized by persistent homology, Proceedings of the National Academy of Sciences, Vol. 113, No. 26, pp. 7035–7040, 2016.

9) Y. Onodera, et al., Structure and properties of densified silica glass: Characterizing the order within disorder, NPG Asia Mater., Vol. 12, pp. 85, 2020.