Tuning of Bayesian optimization for materials synthesis: simulation of the one-dimensional case

Ryo Nakayama a, b, Ryota Shimizu a,b, Taishi Haga c, Takefumi Kimura c, Yasunobu Ando b, c, Shigeru Kobayashi a, Nobuaki Yasuo a, c, Masakazu Sekijima a and Taro Hitsug i a,g

School of Materials and Chemical Technology, Tokyo Institute of Technology, Tokyo, Japan; bPRESTO, Japan Science and Technology Agency, Kawaguchi, Japan; cDepartment of Physics, Tokyo Institute of Technology, Tokyo, Japan; dResearch Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan; eTokyo Tech Academy for Convergence of Materials and Informatics (TAC-MI), Tokyo Institute of Technology, Tokyo, Japan; fDepartment of Computer Science, Tokyo Institute of Technology, Yokohama, Japan; gDepartment of Chemistry, The University of Tokyo, Tokyo, Japan

ABSTRACT

Materials exploration requires the optimization of a multidimensional space including the chemical composition and synthesis parameters such as temperature and pressure. Bayesian optimization has attracted attention as a method for efficient multidimensional optimization. Appropriate choices of the acquisition function and initial values of the hyperparameters of the kernel functions are essential for the Bayesian optimization of synthesis conditions in a small number of experiments. However, to date, there has been little discussion on how to tune Bayesian optimization for materials exploration, and no guidelines have been provided for materials scientists. In this study, we investigated the optimum initial values of the hyperparameters in Bayesian optimization using one-dimensional model functions that mimic actual materials syntheses. The optimal lengthscale and variance for different process windows of materials synthesis were investigated. It was shown that the use of an appropriate acquisition function and suitable initial values of the hyperparameters of the kernel functions enable the optimization of synthesis conditions in a small number of trials. These results provide insight for enabling fully automated and autonomous materials synthesis using Bayesian optimization and robotics for materials exploration.

ARTICLE HISTORY

Received 22 December 2021
Revised 5 April 2022
Accepted 11 April 2022

KEYWORDS

Bayesian optimization; machine learning; autonomous material synthesis; materials exploration

1. Introduction

Recent developments in science and technology have rapidly expanded the range of possible multicomponent materials [1,2] that can be explored to meet the increasing demands of industrial and economic development [3,4]. For example, the total number of small organic molecules has been estimated to be more than 10^60, forming a vast ‘chemical space’ [1]. In addition, the number of controllable synthesis parameters has increased due to the use of sophisticated experimental procedures and equipment. Accordingly, material scientists need to optimize the composition and synthesis parameters to obtain the desired material in a large search space. As it is difficult and, in some
cases, impossible to find the best chemical compositions and synthesis conditions in a multidimensional search space using an unguided trial-and-error approach, a more efficient and streamlined method is required to facilitate the exploration of multi-component complex materials.

To obtain the desired materials in a small number of experiments, we focus on Bayesian optimization [2,5–13]. This technique finds the global optimum of a black-box function and is used in many fields, such as computational model tuning in machine learning [14,15] and cognitive science [16]. Compared to grid searching, Bayesian optimization can find the global optimum in fewer trials [14]. Accordingly, the experimental conditions in material synthesis are expected to be optimized in a small number of trials using Bayesian optimization. The combination of Bayesian optimization and robotics has paved the way for fully automatic and autonomous material syntheses [10,17]. In a previous study, we reported an autonomous thin-film synthesis system for inorganic materials and minimized the electrical resistance of Nb-doped TiO₂ thin films by varying the oxygen partial pressure during deposition [2]. To further increase the efficiency of this powerful approach for obtaining novel functional materials from the large material search space, reducing the number of experiments for optimization (trials) is highly desirable.

When applying Bayesian optimization to minimize the number of trials, appropriately selecting the acquisition and kernel functions and their hyperparameters is crucial. Especially in autonomous material synthesis combining Bayesian optimization and robotics without human scientists, adjusting the hyperparameters in the middle of the loop is not desirable; therefore, it is important to choose appropriate hyperparameters at the beginning of the experiment. However, there has been inadequate discussion of guidelines to set suitable hyperparameters for material scientists, which hinders the actual use of Bayesian optimization in materials synthesis [11–13,18]. Furthermore, only a few researchers have considered issues that are important in actual materials synthesis, such as the half-width of the process window and multiple local minima and maxima. Moreover, the number of trials required to find the best synthesis conditions under the post-hyperparameter optimized condition has not been sufficiently explored.

In this study, we investigated the optimal initial values of the hyperparameters suitable for material synthesis. We applied Bayesian optimization to a one-dimensional (1D) model function that mimics an actual material synthesis. To simulate actual syntheses, we assumed a synthesis temperature range of 200–1200°C and determined the temperature at which optimal physical properties can be obtained. We emphasize the importance of process windows in the synthesis of functional materials and compare the number of syntheses required for the optimization of various process windows. Our findings enable global optimization with minimum number of trial syntheses by choosing appropriate hyperparameters, thereby accelerating materials exploration.

2. Methods

Bayesian optimization predicts the relationship between explanatory variables (synthetic conditions, e.g., temperature and pressure) and objective variables (e.g., transmittance and electrical resistance) from the acquired data and sequentially searches for the global optimum based on the prediction curve and confidence interval. Bayesian optimization is performed as follows. 1) Obtain the initial data points. 2) Calculate the prediction curve and confidence interval (standard deviation) for the objective variable from the obtained data. 3) From the prediction model, rank the synthetic conditions based on an index called the acquisition function to obtain a candidate global optimum. 4) Acquire the data again based on the candidate explanatory variables suggested in step 3, for which the global optimum can be obtained, and return to step 2. In the search for the best synthesis conditions, steps 2–4 are repeated. In Bayesian optimization, the prediction model for an objective variable (property value) is created using Gaussian process regression or other method [7,19]. The property value under a certain experimental condition x is expressed in the form of a probability distribution (usually a normal distribution), p(y|x). The prediction curve is obtained from the mean of the probability distribution, and the standard deviation (variance) of the probability distribution represents the uncertainty of the prediction.

In this study, we generated one-dimensional model functions that mimic the synthesis of LiCoO₂ [20] and the Nb-doped TiO₂ transparent conductor [21] (Figure 1). Considering the width of the process window, Pₜₘ, is essential when generating model functions that mimic material synthesis. Pₜₘ is the range of synthesis conditions in which the desired property values can be obtained. For example, a Pₜₘ of 100°C means that the properties do not change even if the temperature deviates above or below the optimum temperature by approximately 50°C. In many cases, Pₜₘ is bounded by 80–90% of the maximum (and minimum) of the physical property value. The smaller the Pₜₘ, the greater the number of experiments required to optimize the synthesis conditions. For instance, the Pₜₘ in the synthesis of LiCoO₂ and Nb-doped TiO₂ is around 100°C and this value is often observed in material synthesis. Furthermore, in actual material synthesis, physical properties often have local solutions [2,22]. In this study, we added a local solution to the model.
function in order to mimic a more complex material synthesis than in the case of LiCoO$_2$ and Nb-doped TiO$_2$.

To observe clearly the effects of the process window ($P_w$) and local solutions on the number of trials required to find the global maximum, we generated simple 1D model functions by summing the three Gaussian peaks corresponding to the global maximum, local maximum, and background. Gaussian peaks can be expressed as follows:

$$f(x) = H \exp \left( -\frac{(x - \mu)^2}{2\sigma_{\text{Gauss}}^2} \right)$$

where $H$ is the peak height, $\mu$ is the peak position, and $\sigma_{\text{Gauss}}$ corresponds to the standard deviation of the peak. The values of $H$ for the three Gaussian peaks corresponding to the maximum, local maximum, and background peaks were set to 1.2, 0.7, and 0.3, respectively. Note that we assumed that the physical property measurements were sufficiently accurate and that the noise in the experimental data was negligibly small. In this study, $P_w$ was defined as the full width at 90% of the peak maximum. In this case, $P_w = \sigma_{\text{Gauss}}$, and $P_w$ is related to the full width at half maximum (FWHM) by $\text{FWHM} = 2P_w\sqrt{2\ln2}$. Three types of 1D model functions with different $P_w$ values of 100, 60, and 40°C were generated. For the local maximum and background peaks, $P_w$ was set to 120°C and 400°C respectively. Five hundred model functions were generated, where the positions $\mu$ of the global maximum, local maximum, and background were randomly determined. To prevent the overlapping of global and local maximum peaks, the peaks were separated by at least twice the $P_w$ of the local maximum peak (240°C).

The most commonly used radial basis function (RBF) kernel was adopted in Gaussian process regression [23] to build the prediction model, and the initial values of the hyperparameter of kernel functions (lengthscale, and variance) were optimized. The setting of the kernel function plays an important role when developing a prediction model, because the shapes of the prediction curves and confidence intervals depend on the kernel function. For example, small and large values of the lengthscale lead to larger and smaller variations, respectively, for a given interval of an explanatory variable, such as temperature. On the other hand, small and large values of the variance lead to smaller and larger scale changes, respectively, in the objective variables (physical parameters) in the predicted curve.

This study also investigated the difference in the number of trials required for optimization between two acquisition functions, namely, the upper confidence bound (UCB) and expected improvement (EI) [24,25]. The acquisition function determines the next searching point in Bayesian optimization, which depends on the type of acquisition function and its hyperparameters. The UCB is a typical acquisition function and is defined as follows:

$$A_{\text{UCB}}(x) = \mu(x) + \kappa \sigma(x)$$

where $\kappa$ is the hyperparameter of the UCB. In this study, the effect of $\kappa$ was investigated. Adjusting $\kappa$ changes the strategy in finding the global maximum; hence, we can control the balance between ‘exploitation’ (searching for areas with high expected values) and ‘exploration’ (searching for areas with large standard deviations between the probability distribution and a few data points). On the other hand, EI is another acquisition function and is defined as follows:

$$A_{\text{EI}} = E[\left(f - \tau\right)I(f > \tau)] = (\mu - \tau)\Phi(t) + \sigma\varphi(t)$$

where $\tau$ is the largest $y$ among $N$ observations $y_N = f(x_N)$, $\mu$ is the mean value of the probability distribution, and $\sigma$ is the standard deviation [7].

![Figure 1](image1.png)
Here, \( t = (\mu - \tau)/\sigma \) is the difference between \( \mu \) and \( \tau \) normalized by the standard deviation \( \sigma \) at each \( x \). Moreover, \( I(\delta > \tau) \) is an indicator function that takes the value of 1 when \( \delta > \tau \) is true and 0 otherwise. Lastly, \( \Phi(t) \) and \( \phi(t) \) are the cumulative distribution function and density function of the standard normal distribution, respectively. For both EI and UCB, the prediction curve is updated by sequentially experimenting with the function value \( f(x) \) at \( x \), where the acquisition function is maximized.

The GP [26] Python package for Gaussian process regression was used. The initial values of the lengthscales and variance were set to 0.1–10 and 1–5, respectively. In this study, the grid for the 1D simulation was evenly divided into 51 points in the range of 200–1200°C where one grid spacing corresponds to 20°C. Thus, lengthscales = 1 (grid) corresponded to 20°C. Note that the lengthscales and variance were optimized using the truncated Newton method, which is one of the gradient methods that maximizes the marginal likelihood in each iteration. However, the hyperparameters obtained by the gradient method are generally local solutions rather than true global optimal solutions. Therefore, it is important to provide appropriate initial values for the hyperparameters of the kernel function to obtain an appropriate prediction model suitable for explaining the relationship between the experimental data. The objective variables were not standardized before the Gaussian process fitting. The hyperparameter \( \tau \) was set to 1–20 when UCB was used as the acquisition function. For the first five trials, the conditions were selected randomly or at equal intervals. As the grid was spaced at 20°C five points at 200, 440, 700, 940, and 1200°C were selected for the equal interval selection. The sixth and subsequent synthesis conditions were predicted by Bayesian optimization until the 50th trial, and previously used trials were not selected again. In this study, we judged that the maximum value was successfully explored when a value with a magnitude of 90% of the maximum value of the model function was found in the calculation.

To compare the number of trials for optimization quantitatively, the \( N_{90\%} \) index was introduced. This index is the number of trials required to find the global maximum value in 90 out of 100 model functions with the same value of \( P_w \) but different peak positions. Thus, \( N_{90\%} \) corresponds to the number of synthesis in which the global maximum can be reached with 90% probability. In this study, one set of calculations included 100 model functions, and the mean and standard deviation of \( N_{90\%} \) were evaluated for five sets (= 100 \times 5 model functions). The computational cost of the individual iterations of Bayesian optimization is low, and the calculation can be easily performed on any laptop. However, the supercomputer TSUBAME3.0 at the Tokyo Institute of Technology was employed to perform the parallel calculations and speed up the simulations.

3. Results and discussion

3.1. Case of EI

Figure 2 shows a typical example of the synthesis optimization process with \( P_w = 100, 60, \) and 40°C. Here, the acquisition function is EI. An animation of the entire search process is provided in the Supplemental Material. We explain the flow of optimization using the model function with \( P_w = 100°C \) as an example. The upper panels in Figure 2(a–c) show the prediction curve (blue line) and confidence interval (green region) based on the obtained data (white and red circles) that correspond to the measured data in actual experiments. The Bayesian optimization algorithm selects the next search point according to the temperature with the largest acquisition function (lower panels in Figure 2(a–c)). For example, the prediction curve, confidence interval, and acquisition function shown in Figure 2(a) were obtained based on trials up to the seventh trial. The acquisition function (Figure 2(a), bottom) has two peaks at \( 250°C \) near the global maximum and at \( -1000°C \) near the local maximum. The eighth trial was chosen from the maximum in the acquisition function, specifically, the temperature near 1000°C (light blue circle in Figure 2(a), bottom). The result of the eighth trial is represented by the red circle in Figure 2(b). The updated prediction curve, confidence interval, and acquisition function, according to the eighth trial, are depicted in Figure 2(b). The maximum of the acquisition function was \( 250°C \) and was chosen as the ninth trial condition. By repeating this process, the temperature that gives the maximum value of the physical property (best synthesis condition) was successfully found in the ninth search for this model function (Figure 2(c)). Similarly, in the cases of \( P_w = 60 \) and 40°C, the prediction curve and confidence interval were obtained based on the acquired data, and finally the best synthesis conditions were determined (Figure 2(d–i)). With decreasing value of \( P_w \), the number of experiments required to find the maximum value increases as the search around the local solution increases (Figure 2(d–i)).

3.2. \( P_w \) dependence of EI

To investigate the effect of \( P_w \) in more detail, we show histograms of the number of trials required to find the global maximum for 100 model functions with \( P_w = 100, 60, \) and 40°C (Figure 3). At \( P_w = 100°C \) 90% of the global maxima were found in 10 trials. With a smaller \( P_w \) value, the center of the histogram shifts to the right, indicating that the number of required trials tends to increase. When \( P_w \) decreases, the width of the histogram increases, because even when the value of \( P_w \) is small, if the initial point is near the peak of the global maximum value, the number of syntheses required to find the global maximum decreases. Notably, all model functions reach the global maximum within 50 searches.
Table 1 summarizes the obtained \( N_{90\%} \) values for five equally spaced and five randomly selected initial points. At \( P_w = 100, 60, \) and 40°C the \( N_{90\%} \) values calculated using 500 model functions are 9.6 ± 0.5, 13.4 ± 0.5, and 17.6 ± 0.9, respectively. We also compared the \( N_{90\%} \) values of the five randomly chosen initial points and five equally spaced initial points that include both ends (Table 1). No significant variation in the obtained \( N_{90\%} \) values with the initial value selection method was observed. Therefore, equally spaced initial points, including both ends, were used for the subsequent simulations.
Figure 3. Histogram of the number of syntheses required to find the maximum value of the model function with (a) $P_w = 100$, (b) 60, and (c) 40°C when the expected improvement (EI) is used as the acquisition function, lengthscale = 3, and variance = 1. Five initial points equally spaced between 200 and 1200°C including both edges, were used. Each histogram was obtained from 100 model functions with the same value of $P_w$ but different peak positions.

Table 1. Process window ($P_w$) dependence of $N_{90\%}$, the acquisition function was the expected improvement (EI). Lengthscale = 3, variance = 1. $N_{90\%}$ was calculated from five sets ($= 500$) of model functions with the same value of $P_w$ but different peak positions. The five initial points were chosen in two ways: either equally spaced, including both edges, or randomly.

| Initial points         | $P_w$/$^\circ$C | $N_{90\%}$     |
|------------------------|-----------------|----------------|
| Equally spaced         | 100             | 9.6 ± 5        |
|                        | 60              | 13.4 ± 5       |
|                        | 40              | 17.6 ± 9       |
| Randomly selected      | 100             | 9.2 ± 1.0      |
|                        | 60              | 13.2 ± 8       |
|                        | 40              | 19.2 ± 1.0     |

3.3. Case of UCB

Next, we examined the number of trials required to find the maximum using the UCB acquisition function. Table 2 shows the obtained $N_{90\%}$ values when $\kappa$ of the UCB was varied from 1 to 20. For $P_w = 100^\circ$C $N_{90\%}$ shows a small dependence on $\kappa$. In contrast, for small $P_w$, $N_{90\%}$ strongly depends on $\kappa$. In Bayesian optimization, it is important to use an acquisition function that provides an appropriate balance between ‘exploitation’ to search peaks with highly expected values and ‘exploration’ to search points with large variances. For small $\kappa$, only the areas around a local maximum are searched, resulting in an increase in the number of trials necessary to find the best conditions in the model function with $P_w = 60$ and 40°C. Considering the overall $N_{90\%}$ for the three $P_w$ values, $\kappa = 10$ is considered the best hyperparameter. Typical search processes for model functions using the UCB are shown in Figure S1, and the histograms are depicted in Figure S2.

Here, we compare the number of trials required to find the maximum when using the EI and UCB as the acquisition functions. For model functions with small $P_w$ values, the EI is superior to the UCB. For example, for $P_w = 40^\circ$C the $N_{90\%}$ value of the UCB is 24.6 ± 2.1, whereas the $N_{90\%}$ value of the EI is 17.6 ± 9, indicating that the number of trials required for the EI is approximately 70% of that for the UCB. Compared to the EI, the width of the histogram is wider for the UCB, indicating a tendency of the UCB to fall into a local maximum when the initial point is not near the global maximum value (Figures 3, S1 and S2).

3.4. Lengthscale and variance dependence

Furthermore, we investigated the effects of the lengthscale and variance on $N_{90\%}$ (Table 3, the acquisition function is the EI). For all three cases of $P_w$, the smallest $N_{90\%}$ value was obtained for lengthscale = 1 and variance = 1. We note that even after optimizing the lengthscale and variance, $N_{90\%}$ is smaller in the EI than in the UCB (Table S1). With increasing lengthscale and/or variance, $N_{90\%}$ increased, particularly for $P_w = 60$ and 40°C. For $P_w = 40^\circ$C when using lengthscale = 5 and variance = 10, $N_{90\%}$ increased by a factor of 2.4 relative to the value obtained using the best conditions (lengthscale = 1, variance = 1). This result highlights the importance of tuning the hyperparameters.

A typical example of the search process for the best conditions (lengthscale = 1, variance = 1) and the worst conditions (lengthscale = 5, variance = 10) for
Table 3. Dependence of $N_{\text{search}}$ on the initial values of the lengthscale and variance, where the expected improvement is the acquisition function. The calculation for $N_{\text{search}}$ was similar to those used for the data in Tables 1 and 2. Five initial points equally spaced between 200 and 1200°C and including both edges were selected.

| Lengthscale | Variance | 1 (± 20°C) | 3 (± 60°C) | 5 (± 100°C) |
|-------------|----------|------------|------------|------------|
| 100         | 0.1      | 8.0 ± 0.0  | 8.6 ± 0.5  | 10.4 ± 0.8 |
|             | 1        | 8.0 ± 0.0  | 9.6 ± 0.5  | 13.4 ± 0.9 |
|             | 10       | 8.8 ± 0.4  | 11.4 ± 0.8 | 13.2 ± 1.3 |
| 60          | 0.1      | 10.2 ± 0.4 | 11.2 ± 0.4 | 14.2 ± 0.4 |
|             | 10       | 10.0 ± 0.0 | 13.4 ± 0.5 | 19.6 ± 1.6 |
| 40          | 0.1      | 12.0 ± 0.6 | 14.4 ± 0.8 | 18.6 ± 1.9 |
|             | 10       | 12.6 ± 0.8 | 17.6 ± 0.9 | 24.0 ± 0.9 |

$P_w = 40°C$ is shown in Figure 4. Animation of the entire search process is provided in the Supplemental Material. For lengthscale = 1 and variance = 1, the maximum value was found in the 13th search, and the shape of the obtained prediction curve was close to that of the true model function (Figure 4(a-c)). In contrast, for lengthscale = 5 and variance = 10, the shape of the prediction curve after the 13th search is quite different from that of the model function, and the global maximum is not predicted well (Figure 4(d-g)).

These results indicate the importance of setting a smaller initial lengthscale than $P_w$, which is expected from the experimental experience of material scientists. For example, if the initial value of the lengthscale is set to 5 (= 100°C), then the influence of a certain data point will affect the prediction of the data within a range of ±100°C. The resulting loose prediction curve is not suitable for predicting steeply rising peaks, such as the global maximum peak at $P_w = 40°C$. It should be noted that the lengthscale also depends on the size of the grid. In this study, the temperature range of 200–1200°C was divided by 51 points; thus, 1 grid spacing = 20°C and therefore, lengthscale = 1 corresponds to 20°C. In contrast, if the range of 200–1200°C is divided by 101 points, lengthscale = 1 corresponds to 10°C. If the search space of the explanatory variables is divided by an extremely fine grid, the search space expands unnecessarily. Therefore, it is necessary to set the grid size and lengthscale properly according to the materials and $P_w$.

For the synthesis parameter treatment, explanatory variables that vary exponentially, such as pressure, should not be used directly in Bayesian optimization. Rather, these variables should be logarithmically transformed, and then the parameter range is divided into a grid because a linearly changing lengthscale cannot follow the exponential changes in parameters such as pressure. Thus, after Bayesian optimization, the parameters are exponentially transformed back to the original values. The logarithmic transformation is an important preprocessing step in the application of Bayesian optimization to material synthesis, as we successfully demonstrated in our previous work on TiO$_2$ thin-film synthesis [2].

We now comment on the effect of the variance. When the lengthscale and variance are set to 5 and 10, respectively, the maximum value is found in the 25th search. However, the prediction curve is not appropriate and tends to ignore each data point (Figure 4(g–i), blue curves) because the initial value of the variance is extremely large compared to the changes in the physical properties. For example, if the scale of the vertical axis of the model function is increased by a factor of 10, the minimum $N_{\text{search}}$ value is obtained when the initial variance is 10 (Table S2). Therefore, it is important to pre-process the measured data by adjusting the initial variance according to the expected range of the changes in the physical properties (for example, 0–1 for transmittance, ~10 digits for electrical conductivity by taking the logarithm), or by normalizing and logarithmically transforming the values of the physical properties.

4. Comparison of human search and Bayesian optimization

We discuss the relationship between our model function and the properties-temperature data obtained from actual experiments. In this study, model functions with $P_w$ of 100, 60, and 40°C were created to simulate actual material synthesis. The electrical conductivity [27–29], phosphorescence [30], catalytic property [31], coercivity (of a ferromagnetic material) [32], and so on were reported in actual experiments can be regarded as the physical properties of this model function. Although the maximum value of the physical properties was set to about 1, the current results can be viewed as a simulation of a wide range of material synthesis, considering that the present model function is concerned with normalizing the objective variable.

Next, we discuss the number of trials required to find the best synthesis condition when using Bayesian and human optimization. If a human researcher performs an experiment represented by the model function (the optimization of 1D synthesis conditions in the range of 200–1200°C), the optimization flow would be as follows, based on our experience: First, a candidate for the best synthesis temperature is estimated based on 11 experiments performed in 100°C increments. Next, the global optimization is completed by searching for two or three additional points in 50°C increments. Thus, the number of syntheses required to search for the global maximum is supposed to be 13–14 for human researchers. Despite the severe conditions of $P_w = 40°C$, Bayesian optimization can
complete the optimization in a comparable number of trials (13 trials) as a human. Furthermore, Bayesian optimization enables the global maximum to be found in fewer trials (~10 trials) than in human experiments in the cases of $P_w = 100$ and $60^\circ\text{C}$. This result indicates that rapid material exploration and data accumulation can be performed using a fully automated and autonomous material synthesis system by combining robotics and Bayesian optimization with appropriate acquisition functions and hyperparameters. It is necessary to find the optimal initial values of the hyperparameters for the global maximum in a 2D, 3D, or higher-dimensional search space. When hyperparameters are appropriately selected, the best synthesis conditions can be obtained in fewer syntheses than in human-guided experiments.

To apply actual experiments, Bayesian optimization may need to be performed on noisy data. In such cases, the following kernel function $\kappa(x, x')$, which is the sum of the RBF kernel and a white noise kernel, can be used:
\[ \kappa(x, x') = \sigma_{RBF}^2 \exp \left( -\frac{(x-x')^2}{2l^2} \right) + \sigma_{wh}^2 \delta(x, x') \] 

(4)

where \( \sigma_{RBF}^2 \) is the variance of noise-free signal, \( l \) is the lengthscale, \( \sigma_{wh}^2 \) is the variance of the noise, and \( \delta(x, x') \) is the Kronecker delta function \([7,33]\). The first term corresponds to the RBF kernel, and the second term is the white noise kernel to capture the noise in data. If the data is heavily noisy, tuning of \( \sigma_{wh}^2 \) will also be necessary. It should be noted that achieving autonomous experiments requires both the adjustment of hyperparameters in Bayesian optimization as well as the development of high-precision automated measurement methods without human researchers \([34–37]\). In addition, for Bayesian optimization in search spaces of higher dimensions, it is necessary to consider whether condition optimization can be completed in a practical number of experiments. The higher dimensional case will be thoroughly discussed in a subsequent paper. The results obtained from the 1D simulation are essential for future discussions and will greatly contribute to the exploration of materials using fully automated and autonomous experiments.

5. Conclusion

In this study, we optimized the hyperparameters in Bayesian optimization for material synthesis using 1D model functions with different process windows \( (P_w = 100, 60, \text{ and } 40\% \text{C}) \) and investigated the number of syntheses required to determine the optimal synthesis conditions. For small \( P_w \), the number of trials required to search for the global maximum using the EI was smaller than that for the UCB. In Bayesian optimization with the optimized hyperparameters (acquisition function EI, lengthscale \( = 1 \text{ (i.e., } 20\text{C}) \), variance \( = 1 \text{), the material showing the maximum physical properties is expected to be obtained with 90% probability after 13 syntheses. These results contribute to the application of Bayesian optimization for material exploration and pave the way for fully automated and autonomous material synthesis using a combination of Bayesian optimization and robots.

Acknowledgments

This research was supported by JST-CREST (Grant No. JPMJCR1523), JST-Mirai Program (Grant Nos. JPMJMI19G4 and JPMJMI21G2), and JST-SCORE (University Promotion Type: Developing the Environment for Creation of Startup Ecosystem in Startup Cities Type; Grant No. JPMJST2074). T.H. and R.N. acknowledge funding from JSPS Kakenhi Grant Nos. 18H03876, 19K15672, 22K14692. T. H. acknowledges support from the MEXT Project of the Tokyo Tech Academy for Convergence of Materials and Informatics (TAC-MI). Numerical calculations were carried out on the TSUBAME3.0 supercomputer at the Tokyo Institute of Technology. We would like to thank Editage www.editage.jp for English language editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ryo Nakayama (http://orcid.org/0000-0001-6284-9884)
Ryota Shimizu (http://orcid.org/0000-0001-9600-7255)
Yasunobu Ando (http://orcid.org/0000-0003-3702-034X)
Shigeru Kobayashi (http://orcid.org/0000-0003-4997-8782)
Nobuaki Yasuo (http://orcid.org/0000-0002-3114-7895)
Masakazu Sekijima (http://orcid.org/0000-0002-3806-9535)
Taro Hitosugi (http://orcid.org/0000-0002-7795-0683)

References

[1] Kirkpatrick P, Ellis C. Chemical Space. Nature. 2004;432:823.
[2] Shimizu R, Kobayashi S, Watanabe Y, et al. Autonomous materials synthesis by machine learning and robots. APL Mater. 2020;8:111110.
[3] Yeh JW, Chen SK, Lin SJ, et al. Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes. Adv Eng Mater. 2004;6:299.
[4] Kato Y, Hori S, Saito T, et al. High-power all-solid-state batteries using sulfide superionic conductors. Nat Energy. 2016;1:16030.
[5] Kushner HJ. A versatile stochastic model of a function of unknown and time varying form. J Math Anal Appl. 1962;5:150–167.
[6] Kushner HJ. A new method of locating the maximum point of an arbitrary multiplet function in the presence of noise. J Basic Eng. 1964;86:97–106.
[7] Shahriari B, Swersky K, Wang Z, et al. Taking the human out of the loop: a review of Bayesian optimization. Proc IEEE. 2015;1:148.
[8] Häse F, Roch LM, Kreisbach C, et al. Phoenics: a Bayesian optimizer for chemistry. ACS Cent Sci. 2018;4:1134.
[9] Harada M, Takeda H, Suzuki S, et al. Bayesian-optimization-guided experimental search of NASICON-type solid electrolytes for all-solid-state Li-ion batteries. J Mater Chem A. 2020;8:15103.
[10] MacLeod BP, Parlane FGL, Morrissey TD, et al. Berlinguette CP. Self-driving laboratory for accelerated discovery of thin-film materials. Sci Adv. 2020;6:eaaz8867.
[11] Li C, de Celis Leal DR, Rana S, et al. Bayesian optimisation for synthesis of short polymer fiber materials. Sci Rep. 2017;7:5683.
[12] Paulson NH, Libera JA, Stan M. Frame spray pyrolysis optimization via statistics and machine learning. Mater Des. 2020;196:108972.
[13] Paulson NH, Yanguas-Gil GA, Aboumar OY, et al. Intelligent agents for the optimization of atomic layer deposition. ACS Appl Mater Interfaces. 2021;13:17022.
[14] Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Proc 25th Int Conf Neural Inf Process Syst. 2012;2:2951–2959.

[15] Letham B, Karrer B, Ottoni G, et al. Constrained Bayesian optimization with noisy experiments. Bayesian Anal. 2019;14:495–519.

[16] Wu CM, Schulz E, Speekenbrink M, et al. Generalization guides human exploration in vast decision spaces. Nat Hum Behav. 2018;2:915–924.

[17] Burger M, Maffettone PM, Gusev VV, et al. A mobile roboticist. Nature. 2020;583:237–241.

[18] Haber A, Aykol M, Suram SK, et al. Benchmarking the acceleration of materials discovery by sequential learning. Chem Sci. 2020;11:2696.

[19] Snoek J, Rippel O, Swersky K, et al. Proc. 32nd Int. Conf. Mach. Learn. 2015;37:2171–2180.

[20] Tsuruhama T, Hitosugi T, Oki H, et al. Preparation of layered-rhombohedral LiCoO2 epitaxial thin films using pulsed laser deposition. Appl Phys Express. 2009;2:085502.

[21] Yamada N, Hitosugi T, Kasai J, et al. Direct growth of transparent conducting Nb-doped anatase TiO2 polycrystalline films on glass. J Appl Phys. 2020;105:123702.

[22] Bradley K, Giaglogliou K, Hayden BE, et al. Reversible perovskite electrocatalysts for oxygen reduction/oxygen evolution. Chem Sci. 2019;10:4609.

[23] Rasmussen CE, Williams KL. Gaussian Process for Machine Learning. MIT Press; 2006.

[24] Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. J Glob Optim. 1998;13:455–492.

[25] Brochu E, Cora VM, de Freitas NA Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. arXiv:1012.2599;2010.

[26] GPy. GPy: A Gaussian process framework in python. 2012. Available from: http://github.com/SheffieldML/GPy.

[27] Kim Y, Ballantyne AM, Nelson J, et al. Effects of thickness and thermal annealing of the PEDOT: PSS layer on the performance of polymer solar cells. Org Electron. 2009;10:205–209.

[28] Natsume Y, Sakata H. Zinc oxide films prepared by sol-gel spin-coating. Thin Solid Films. 2000;372:30–36.

[29] Kuo SY, Chen WC, Lia IF, et al. Effects of doping concentration and annealing temperature on properties of highly-oriented Al-doped ZnO films. J Cryst Growth. 2006;287:78–84.

[30] Zhang X, Zhang J, Ren X, et al. The dependence of persistent phosphorescence on annealing temperatures in CaTiO3:Pr3+ nanoparticles prepared by a coprecipitation technique. J Solid State Chem. 2008;181:393–398.

[31] Jiang Z, Yang H, Wei Z, et al. Catalytic properties and structures of nano-amorphous Ni–B alloys affected by annealing temperatures. Appl Catal A Gen. 2005;279:165–171.

[32] Ai L, Jiang J. Influence of annealing temperature on the formation, microstructure and magnetic properties of spinel nanocrystalline cobalt ferrites. Curr Appl Phys. 2010;10:284–288.

[33] Valldarees H, Li T, Zhu L, et al. Gaussian process-based prognostics of lithium-ion batteries and design optimization of cathode active materials. J Power Sources. 2022;323:10126.

[34] Nagata K, Sugita S, Okada M. Bayesian spectral deconvolution with the exchange Monte Carlo method. Neural Netw. 2012;25:82–89.

[35] Matsumura T, Nagamura N, Akaho S. Spectrum adapted expectation-maximization algorithm for high-throughput peak shift analysis. Sci Tech Adv Mater. 2019;20:733–745.

[36] Miyazaki Y, Nakayama R, Yasuo N, et al. Bayesian statistics-based analysis of AC impedance spectra. AIP Adv. 2020;10:045231.

[37] Ueno T, Ishibashi H, Hino H, et al. Automated stopping criterion for spectral measurements with active learning. NPJ Comput Mater. 2021;7:139.