ARTICLE OPEN

Automated stopping criterion for spectral measurements with active learning

Tetsuro Ueno¹✉, Hideaki Ishibashi², Hideitsu Hino³ and Kanta Ono^{4,5}

The automated stopping of a spectral measurement with active learning is proposed. The optimal stopping of the measurement is realised with a stopping criterion based on the upper bound of the posterior average of the generalisation error of the Gaussian process regression. It is revealed that the automated stopping criterion of the spectral measurement gives an approximated X-ray absorption spectrum with sufficient accuracy and reduced data size. The proposed method is not only a proof-of-concept of the optimal stopping problem in active learning but also the key to enhancing the efficiency of spectral measurements for high-throughput experiments in the era of materials informatics.

npj Computational Materials (2021)7:139; <https://doi.org/10.1038/s41524-021-00606-5>

INTRODUCTION

Machine learning and artificial intelligence (AI)-related techniques have been rapidly implemented in daily lives and industries, such as electronic commerce, manufacturing, and automated driving. This situation is not an exception for various scientific fields; in particular, materials science combined with information science is referred to as ‘materials informatics’^{1–7}. Materials informatics has been successful in predicting and finding new functional materials^{8–13} or optimising devices and material microstructures^{14,15}.

In general, machine learning techniques, such as deep learning, require substantial data to learn. Therefore, in materials informatics, theoretical calculations, such as first-principles calculations or molecular dynamics simulations, are useful methods for accumulating materials property data to construct a materials database^{16–18}. However, substantial problems in materials science, e.g., the coercivity in permanent magnets, catalytic reactions, and the charge and discharge of batteries, are difficult to completely describe with theoretical calculations. These physical and chemical phenomena have complexities that span multi-temporal and spatial scales. These problems are difficult to model and require a tremendous computational cost. Therefore, an experimental approach is essential for understanding these phenomena. First, data acquisition by experiments incurs various costs related to real specimen preparation, human-operated measurement equipment, trial-and-error analyses, etc. There is a strong demand to promote the efficiency of experimental methods to accelerate materials science beyond traditional processes. In addition to materials informatics, ‘measurement informatics’, a research area that integrates measurements and machine learning, has been explored to date. With this approach, the measurement^{19–21} and data analysis efficiency can be enhanced^{22–26}.

Active learning (AL) is a machine learning scheme used to obtain predictive models with high precision at a limited cost through the sophisticated selection of samples for labelling²⁷. Similar to other machine learning methods, AL is utilised in materials informatics^{28–34} and measurement informatics^{19,20}. Ueno et al. developed a method for X-ray magnetic circular dichroism

(XMCD) spectral measurement with AL¹⁹. XMCD spectroscopy is a variation of X-ray absorption spectroscopy (XAS) using circularly polarised X-rays to investigate the element-specific magnetic properties of materials³⁵. In particular, magneto-optical sum rules relate the XMCD spectrum to the spin and orbital magnetic moments, which are the fundamental physical parameters of elements^{36–38}.

Measurements of XAS and XMCD spectra are usually performed in a step-by-step manner, i.e., the measurement of the X-ray absorption intensity and tuning of the X-ray energy is repeated over the energy range of the spectrum (Fig. 1b). In conventional XAS and XMCD measurements, the numbers of energy points and energy steps are predetermined by the experimenter, and usually, several hundred points are measured. These measurement conditions are set based on experimenters’ experience and intuition. This type of conventional DoE has possibility to be optimised by a machine learning technique. The XMCD spectral measurement with AL uses Gaussian process regression (GPR)³⁹ to predict the whole spectral shape from the measured data (Fig. 1a). The standard deviation of the prediction is used to determine the optimal energy point to measure. The measurement stops if the convergence criterion is satisfied. Compared to conventional measurements, this method succeeded in reducing the number of measurement energy points.

However, the XMCD spectral measurement with AL depends on the magnetic moment evaluated from the predicted spectrum as the stopping criterion. This condition means that the method is only applicable when the relation between the spectrum and the physical parameter is known. In addition, the physical parameter must be evaluated quickly from the spectrum to fit in the AL cycle of the measurement and GPR. In the common situation of a spectral measurement, the evaluation of physical parameters from the spectra is not straightforward; e.g., a comparison between the experimental and theoretical spectra is often needed. Therefore, a stopping criterion without the physical parameter is required to apply the spectral measurement with AL to general types of spectra.

¹Synchrotron Radiation Research Center, Kansai Photon Science Institute, Quantum Beam Science Research Directorate, National Institutes for Quantum and Radiological Science and Technology, Sayo, Hyogo, Japan. ²Department of Human Intelligence Systems, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan. ³The Institute of Statistical Mathematics, Research Organization of Information and Systems, Tachikawa, Tokyo, Japan. ⁴Department of Applied Physics, Graduate School of Engineering, Osaka University, Suita, Osaka, Japan. ⁵Center for Integrative Quantum Beam Science, Institute of Materials Structure Science, High Energy Accelerator Research Organization, Tsukuba, Ibaraki, Japan. ✉email: ueno.tetsuro@qst.go.jp

Fig. 1 Concept of the spectral measurement with active learning. **a** Schematic of the spectral measurement with active learning. **b** Schematic of the spectral measurement with conventional design of experiments. Red markers, blue curves, and grey bars indicate the sampled data, predicted mean of Gaussian process regression, and energy points, respectively. Green arrows represent the sampling order that is determined with active learning in **a** or sequential sampling in **b**. **c** Flowchart of the spectral measurement with active learning. (1) A sampling of the initial data. (2) Gaussian process regression (GPR) is used to predict the spectrum and evaluate the standard deviation. The measurement is finished when the stopping criterion is satisfied. (3) If the stopping criterion is not satisfied, then the next sampling point is determined based on the acquisition function. (4) A sampling of new data and back to the process (2).

The optimal stopping problem is a long-standing problem in AL²⁷. Without an appropriate stopping criterion, the active learner would require many unnecessary samples to be labelled or too few samples, resulting in poor predictive performance. Despite its importance, there have been few studies on timing considerations for AL^{40–42}, mainly because of its problem dependency. From the viewpoint of materials science, an optimal stopping rule is highly desired to avoid useless and costly experiments. In the typical problem setting of AL in materials and measurement informatics, the experiment is stopped when the budget is exhausted or the experimenter is satisfied with the results²⁸. The former assumes the experiment of one who needs the best result in a limited time. In such a situation, the experimenter will set the maximum iteration of experiments within the limited time⁴³. The latter suppose the situation that the experiment is iterated until the experimenter satisfies with the target property, e.g. search for material with the highest melting point⁴⁴ and convergence of the magnetic moments within the predetermined threshold¹⁹.

In this paper, we applied a universal stopping criterion for XAS spectral measurement with AL. The stopping criterion is based on the stability of the expected generalisation errors⁴⁵. This stopping criterion is completely evaluated on a mathematical basis; therefore, it is applicable to general spectral measurements whose relation to the physical parameter is unknown. We applied the method to the spectral measurement with AL of several types of simulated and experimental XAS spectra. It is revealed that the automated stopping criterion of the spectral measurement gives an approximated XAS spectrum with sufficient accuracy. The proposed method can be applied not only to spectral measurements but also to other types of measurements and improves the

efficiency of high-throughput experiments in the era of materials informatics.

RESULTS

Spectral measurements with AL

Figure 1c shows the flowchart of a spectral measurement with AL. Spectral measurement with AL includes the following steps: (1) First, initial spectral data $\mathbf{Y}_0(\mathbf{X}_0)$ are sampled. Here, $\mathbf{Y}_0 = (y_1, \dots, y_l)$ represents the spectral intensity at energy points $\mathbf{X}_0 = (x_1, \dots, x_l)$. (2) Subsequently, GPR is applied to the initial data, and we obtain a mean μ_n and standard deviation σ_n for each energy point. $n = 0, 1, \dots$ represents the number of samplings ($n = 0$ for the initial sampling). The mean μ_n is regarded as a predicted spectrum. The stopping criterion is evaluated with the result of the GPR fitting. (3) If the stopping criterion is not satisfied, then the next sampling point x_{next} is automatically determined based on an acquisition function. (4) Subsequently, a new energy point is sampled and added to the measured data $\mathbf{Y}_n = (y_1, \dots, y_{\text{next}}, \dots, y_l)$. Finally, repeating process (2), GPR is applied to $\mathbf{Y}_n(\mathbf{X}_n)$ again. Processes (2) to (4) are repeated until the stopping criterion is satisfied.

In our previous work¹⁹, a physical parameter of interest was evaluated by comparing the fitted spectrum and the reference spectrum of a known material. Then, the physical parameter converges to a predetermined accuracy range, and the measurement is terminated. However, this method has some limitations for general use, i.e., only applicable to materials whose physical parameter can be evaluated with the similarity measure, and the reference value of the physical parameter is available. Moreover, different results are generated for different similarity measures, and the threshold must be properly determined. Therefore, we

Fig. 2 Schematic of the stopping criterion with the error ratio. The data size increases with measurement time t . The error ratio λ_t is evaluated after every Gaussian process regression. The measurement is stopped when λ_t falls below the threshold λ . The original conceptual diagram appears in ref. ⁴⁵.

adopt the criterion that depends only on the fitted spectra to promote the application of spectral measurement with AL.

The choice of a covariance function (kernel) $K(x, x')$ is essential in GPR. We adopt the standard power exponential correlation function defined by

$$K(x, x') = \theta_1 \exp\left(-\left(\frac{|x - x'|}{\theta_2}\right)^2\right) \quad (1)$$

with the amplitude parameter θ_1 and bandwidth parameter θ_2 . The power exponential or Gaussian correlation function performs better than the Matérn correlation function for the fitting of the XAS spectra¹⁹.

Stopping criterion

Approximating the spectrum by the GPR is considered a problem of supervised learning. In this setting, the goodness of approximation is evaluated by the average prediction error of the intensity at an unseen energy point, which is called the generalisation error defined by

$$\mathcal{L}(f) = \int dy \int dx (y - f(x))^2 p(x, y). \quad (2)$$

where $p(x, y)$ is a joint density function of (x, y) . Note that $f(x)$ is a stochastic predictor sampled from the fitted Gaussian process; hence, $\mathcal{L}(f)$ is a random variable. In a problem setting of the spectral measurement, x , y and $f(x)$ are the energy, the ground truth (unknown) spectrum and the fitted curve by the GPR model, respectively.

Let $p_t(f)$ be the posterior distribution of the predictive model $f(x)$ obtained by fitting a Gaussian process to the observation up to time t . Then, the posterior average of the generalisation error is defined by

$$\mathcal{L}_t = \int df p_t(f) \mathcal{L}(f). \quad (3)$$

If the gap $|\mathcal{L}_t - \mathcal{L}_{t+1}|$ of the posterior average of generalisation errors at t and $t + 1$ is small enough, then there is only a small gain by performing an additional observation, and the experiment should be stopped. It is not possible to directly calculate the generalisation error because we do not access the distribution $p(x, y)$. A method to estimate the upper bound of $\mathcal{L}_t - \mathcal{L}_{t+1}$ is proposed in ref. ⁴⁶, and a stopping criterion of AL based on the convergence of the gap $|\mathcal{L}_t - \mathcal{L}_{t+1}|$ is developed in ref. ⁴⁵. Suppose $\mathcal{L}(f) \in [a, b]$, then

$$|\mathcal{L}_t - \mathcal{L}_{t+1}| \leq (b - a)\{r(p_t, p_{t+1}) + r(p_{t+1}, p_t)\} \quad (4)$$

holds where

$$r(p, q) = \exp\left\{W_0\left(\frac{\text{KL}(p||q) - 1}{e}\right) + 1\right\} - 1. \quad (5)$$

In the above equation, W_0 is the main branch of the Lambert W -function⁴⁷, $\text{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx$ is the Kullback–Leibler divergence⁴⁸, and e is the base of the natural logarithm. The Kullback–Leibler divergence between posterior distributions p_t and p_{t+1} is shown to be exactly calculated by using the observed points $\{(x_i, y_i)\}_{i=1}^t$; hence, the upper bound of $|\mathcal{L}_t - \mathcal{L}_{t+1}|$ is computable from the data at hand. To align the scale of $r(p_t, p_{t+1}) + r(p_{t+1}, p_t)$ and remove the constant term $b - a$, we consider the ratio $\lambda_t = \frac{r(p_t, p_{t+1}) + r(p_{t+1}, p_t)}{r(p_1, p_2) + r(p_2, p_1)}$. When this ratio is smaller than a certain threshold $\lambda \in [0, 1]$, the experiments are stopped. Figure 2 shows the schematic of the stopping criterion with the error ratio. Intuitively, the threshold λ is regarded as the expected rate of the improvement of fitting by one additional observation. The stopping time is not very sensitive to the value of λ when it is set to be small enough, but it is possible to determine this parameter by a simulation study in advance. We take this strategy and report the experimental results in the following section.

Application to the simulated spectrum

First, we applied the method to the simulated noise-free Ni $L_{2,3}$ XAS of divalent nickel ions (Ni^{2+}) to verify the effectiveness of the overall strategy. Details of the simulation are described in the Methods section. All spectra used in the study are presented in Supplementary Fig. 1. Figure 3a–f shows snapshots of the GPR fitting of the simulated $L_{2,3}$ XAS of Ni^{2+} . In Fig. 3a, randomly selected initial data points (data size = 10) and GPR fitting are shown. The L_3 peak is measured occasionally in the initial sampling. The covariance function exhibits a large standard deviation between the sampled data points, and the next sampling point is chosen from these x values with a large value of the acquisition function (Eq. (11)). The GPR fitting after several samplings (data size = 20) is shown in Fig. 3b. The whole spectral shape appears, but the standard deviation is still large, approximately one-third of the intensity at the L_3 peak between the sampled data points. Fig. 3c shows the GPR fitting after 50 samplings (data size = 60). Satellite peaks around the L_3 main peak and multiplet structure around the L_2 peak appear at this degree of sampling density. Figure 3d–f shows the results of the GPR fitting for the different stopping timings, i.e., different thresholds. The standard deviation is relatively small compared to the intensity of the L_3 and L_2 peaks in the whole energy range. The intensity relations of the multiplet structure around the L_2 peak are correctly approximated. In Fig. 3d, e, the difference between the GPR fitting for different thresholds only appears in the standard deviations around non-peak regions. In Fig. 3f, the GPR fitting for a data size = 178 is almost the same as the GPR fitting for a data size = 116, but an increase in sampling points around the peak region is visible.

To visualise the progress of the spectral measurement and the stopping timing, the error ratio and the test error versus data size are shown in Fig. 3g, h, respectively. In Fig. 3g, the error ratio λ_t is plotted as a function of the data size, i.e., the number of measurement points x . The error ratio decreases with increasing data size and converges to an almost constant value. Figure 3h shows the test error, i.e., the posterior average of the generalisation errors (Eq. (3)), versus the data size. The test error steeply decreases with increasing data size as compared to the error ratio in the initial stage of the measurement and also converges to a constant value. The spikes in error ratio in Fig. 3g come from the subtraction of the test errors at steps to evaluate the upper bound of the test error in Fig. 3h. The spikes in error ratio coincide with the spikes in the difference of test error between the n -th and

Fig. 3 Spectral measurement of simulated Ni $L_{2,3}$ XAS for Ni^{2+} with active learning. Gaussian process regression (GPR) **a** for data size = 10 (initial sampling), **b** data size = 20, **c** data size = 60, **d** data size = 94 (stopping criterion with $\lambda = 0.1$), **e** data size = 116 (stopping criterion with $\lambda = 0.05$) and **f** data size = 178 (stopping criterion with $\lambda = 0.025$). Black full circles, blue solid curves and blue shades indicate the measured data points, mean function of the GPR, and covariance function of the GPR, respectively. **g** Error ratio versus data size. Black and red curves indicate error ratio and their minimum value, respectively. **h** Test error versus data size. Black and blue curves indicate test error and difference of test error between n -th and $(n - 1)$ -th sampling. In **g** and **h**, red, green and blue vertical dashed lines indicate stopping timings for λ of 0.1, 0.05 and 0.025, respectively. **i** Heat maps of the predicted mean and **j** standard deviation versus data size. Red, green and blue horizontal dashed lines indicate stopping timings for λ of 0.1, 0.05 and 0.025, respectively. Orange markers indicate sampled energy points in each AL cycle.

$(n - 1)$ -th sampling. The spikes are inevitable because the discontinuous improvement of the test error occasionally arises in the AL. Alternatively, the minimum value of the error ratio is plotted in Fig. 3g. It is revealed that the minimum value of the

error ratio gradually decreases with the increase of data size. Therefore, immoderate early stopping of a measurement can be avoided. The vertical dashed lines in Fig. 3g, h indicate the stopping timings for different thresholds that correspond to the

GPR fitting shown in Fig. 3d–f. These results indicate that the automated stopping of the XAS measurement based on the generalisation error gives the GPR fitting with small errors from the ground truth spectrum with the reduced data size.

To perceive the overall picture of the spectral measurement with AL, the predicted mean μ and standard deviation σ versus data size are visualised as heat maps in Fig. 3i, j, respectively. The stopping timings for different thresholds are represented as horizontal dashed lines. In Fig. 3i, the peak structures around L_3 and L_2 regions are shown as thick coloured vertical lines. The peak structures seem to be explicit with increasing data size. On the other hand, the standard deviation decreases with increasing data size, as shown in Fig. 3j. Sampled energy points indicated by orange markers in Fig. 3i, j, are nearly uniformly distributed in the whole spectral energy range. We confirmed a nearly uniform distribution by plotting a histogram of sampled energy points as shown in Supplementary Fig. 6a. Thus, it seems that biased sampling, e.g., intensive sampling around the peak region, does not occur in the spectral measurement with AL with the acquisition function with the form of Eq. (11).

Application to the experimental spectrum

Next, we applied the method to the experimental data to demonstrate its applicability to actual spectral measurements. Experimental data inherently include measurement noise; therefore, it is essential to ascertain the noise tolerance of the present method. Moreover, the experimental XAS spectrum of Ni metal has a finite background that comes from a continuum state approximated by the double step-like function. Figure 4a–f shows snapshots of the GPR fitting of the experimental Ni $L_{2,3}$ XAS of nickel metal. In Fig. 4a, randomly selected initial data points and the GPR fitting are shown. The standard deviation σ is large in sparsely sampled energy regions, as is the usual behaviour of the GPR. The overall spectral shape appears after 50 samplings, as shown in Fig. 4c. In addition to the L_3 and L_2 main peaks, the so-called 6 eV satellite⁴⁹ can be seen at ~858 eV at this sampling density. Figure 4d–f shows the results of the GPR fitting for the different stopping timings with a specific threshold. Similar to the results for the simulated Ni^{2+} XAS spectrum, the L_3 main peak is properly approximated by the GPR fitting with a stopping timing of $\lambda = 0.1$. The standard deviation around the L_2 peak region and other non-peak energy regions decreases with increasing measurements, as shown in Fig. 4e, f.

Figure 4g, h shows the error ratio and the test error versus the data size. As shown in Fig. 4g, the error ratio decreases with increasing data size and converges to constant values. Occasionally appearing spikes originate likewise the results for Ni^{2+} shown in Fig. 4g. Figure 4h shows the data size dependence of the test error. The test error decreases with increasing data size and converges to a constant value, as in the case of Ni^{2+} . The stopping timing (vertical dashed lines in Fig. 4g, h) appears in a data size from 69 to 84. The GPR fittings shown in Fig. 4d–f indicate that relevant stopping timing gives accurate spectral shapes, including main peaks and multiplet structures. These results indicate that the present method also works for the experimental XAS spectrum with noise. The animation of GPR fittings and the evolution of the error ratio and the test error presented above are shown in the Supplementary Movie.

Heat maps of the predicted mean μ and standard deviation σ versus the data size are shown in Fig. 4i, j. In Fig. 4i, peak structures around the L_3 and L_2 regions appear as thick coloured vertical lines and become explicit with increasing data size as with the case of the simulated Ni^{2+} XAS. The sampling tendency is also similar to the case of the simulated Ni^{2+} XAS measurement, and sampling energy points are uniformly distributed in the whole spectral energy range (see also a histogram in Supplementary Fig. 6d).

Time-cost evaluation

It is essential to estimate the time cost, i.e. total measurement time, with the spectral measurement with AL for realistic application to measurements. The time cost for the spectral measurement with AL t_{AL} is defined as

$$t_{\text{AL}} = \sum_n (t_{\text{ene}} \times \Delta E_n^{\text{Init}} + t_{\text{meas}} + t_{\text{GP}}) + \sum_n (t_{\text{ene}} \times \Delta E_n^{\text{AL}} + t_{\text{meas}} + t_{\text{GP}} + t_{\text{SC}}), \quad (6)$$

where t_{ene} , t_{meas} , t_{GP} and t_{SC} are the time to change unit energy, e.g. 1 eV, time to measure single spectral intensity, time to compute the GPR and time to evaluate the stopping criterion, respectively. ΔE_n^{Init} is a distance between energies in the initial sampling and ΔE_n^{AL} is a distance between energies of the n -th and the $(n-1)$ -th sampling. Thus, t_{AL} consists of time costs for the initial sampling and the sampling in AL. On the other hand, the time cost for the conventional DoE t_{CDoE} is defined as

$$t_{\text{CDoE}} = \sum_n (t_{\text{ene}} \times \Delta E^{\text{CDoE}} + t_{\text{meas}}). \quad (7)$$

For simplicity, let ΔE^{CDoE} be a constant by assuming a measurement that constant energy step is used in all spectral energy ranges.

Figure 5 shows the time cost evaluated for $\text{Ni}^{2+}L_{2,3}$ XAS spectral measurement with various $t_{\text{meas}}/t_{\text{ene}}$ ratios, which is variable with the experimental conditions. In this evaluation, we assume the signal-to-noise ratio (SNR) of the XAS spectrum does not depend on t_{meas} , i.e. we ignore the effect of SNR on the spectral measurement with AL, which will be explored in a further study. Meanwhile, it is supposed that $t_{\text{GP}}/t_{\text{ene}} = 0.1$ for all evaluation because the computational time for GPR t_{GP} is much shorter than t_{ene} and t_{meas} , when the number of measurements is in the order of <1000. We note that the computational cost for evaluating the stopping criterion is negligible. Time cost in Fig. 5 exhibits arbitrary units, one can read vertical axis in seconds, for example, $t_{\text{meas}} = t_{\text{ene}} = 1$ [sec] and $t_{\text{GP}} = 0.1$ [sec] in case of Fig. 5a. Both in the spectral measurement with AL and with the conventional DoE, time cost monotonically increases with data size. The slope of the time cost dependence of the conventional DoE is constant because ΔE^{CDoE} is constant as mentioned above. On another hand, the slope of the time cost dependence of the spectral measurement with AL changes because ΔE_n^{AL} changes in each sampling. At the stopping timing of the spectral measurement with AL, an experimenter can observe the whole spectral shape as shown in Fig. 3d–f. In the conventional DoE, the whole spectral shape appears only after the measurement is finished. The spectral measurement with AL outperforms the conventional DoE, i.e. lower time cost is achieved in most cases, except for the case of $t_{\text{meas}}/t_{\text{ene}} = 1$ with the threshold $\lambda = 0.025$. A similar tendency is observed for the measurement of $\text{Co}^{2+}L_{2,3}$ XAS; however, in other cases, the spectral measurement with AL always realises lower time cost than that of the conventional DoE as shown in Supplementary Figs. 7–11.

DISCUSSION

In this paper, we proposed the application of an automated stopping criterion for spectral measurement with AL to enhance the efficiency of spectral measurements. The method was applied to the simulated and experimental Ni $L_{2,3}$ XAS spectra. Predicted spectra with GPR fitting demonstrate satisfactory accuracy at different stopping timings with several thresholds. The method was applied to XAS spectra other than Ni. The GPR fittings and data size dependency of the error ratio and the test error for simulated and experimental Mn and Co $L_{2,3}$ XAS are shown in Supplementary Figs. 2–5. These results show a similar tendency to the simulated and experimental Ni $L_{2,3}$ XAS results. Time costs for various XAS spectral measurements with $t_{\text{meas}}/t_{\text{ene}} = 1$ are

Fig. 4 | Spectral measurement of experimental Ni $L_{2,3}$ XAS with active learning. Gaussian process regression (GPR) **a** for data size = 10 (initial sampling), **b** data size = 20, **c** data size = 60, **d** data size = 69 (stopping criterion with $\lambda = 0.1$), **e** data size = 75 (stopping criterion with $\lambda = 0.05$) and **f** data size = 84 (stopping criterion with $\lambda = 0.025$). Black full circles, blue solid curves and blue shades indicate the measured data points, mean function of the GPR, and covariance function of the GPR, respectively. **g** Error ratio versus data size. Black and red curves indicate the error ratio and their minimum value, respectively. **h** Test error versus data size. Black and blue curves indicate test error and difference of test error between n -th and $(n - 1)$ -th sampling. In **g** and **h**, red, green and blue vertical dashed lines indicate stopping timings for λ of 0.1, 0.05 and 0.025, respectively. **i** Heat maps of the predicted mean and **j** standard deviation versus data size. Red, green and blue horizontal dashed lines indicate stopping timings for λ of 0.1, 0.05 and 0.025, respectively. Orange markers indicate sampled energy points in each AL cycle.

summarised in Fig. 6. The GPR fittings give a reasonable approximation of the XAS spectra in the automated stopping timing whose number of measurements is dramatically reduced as compared to the conventional experimental design. It is revealed

that the automated stopping criterion works well for the spectral measurement with AL in general.

Based on the estimation of time cost, the spectral measurement with AL outperforms that with the conventional DoE in many

Fig. 5 Time cost estimation of the Ni $L_{2,3}$ XAS spectral measurement for simulated Ni^{2+} . Time cost versus data size is plotted for various ratio between time to measure single spectral intensity t_{meas} and time to change of unit energy t_{ene} **a** $t_{\text{meas}}/t_{\text{ene}} = 1$, **b** $t_{\text{meas}}/t_{\text{ene}} = 10$, **c** $t_{\text{meas}}/t_{\text{ene}} = 100$ and **d** $t_{\text{meas}}/t_{\text{ene}} = 1000$, respectively. Red and grey curves indicate time cost for the spectral measurement with AL and that with conventional DoE, respectively. Red, green, blue and grey circles indicate time cost at stopping timings of λ of 0.1, 0.05 and 0.025, and the conventional DoE respectively.

cases as shown in Figs. 5 and 6. In particular, the advantage of the spectral measurement with AL is emphasised in the cases of large $t_{\text{meas}}/t_{\text{ene}}$ ratio. Therefore, the spectral measurement with AL especially is effective in measurements with long measurement time per energy or other scanning parameters such as XAS measurement of a very dilute system, e.g. single molecule on the surface or inelastic neutron scattering, generally known as a neutron-hungry experiment. It is also effective to experiment like scanning transmission X-ray microscopy (STXM). In the STXM experiment, two-dimensional spatial scan is performed at each X-ray energy, so the measurement time per energy becomes much longer than that of changing X-ray energy. Thus, advantage of the reduction of the measurement point by the spectral measurement with AL becomes prominent for experiments with long measurement time per scanning parameter. Note that the so-called ‘on-the-fly’ scan is used as a very quick measurement technique in particular for a single XAS spectral measurement. It is important to use properly such techniques and the spectral measurement with AL depending on the experimenter’s purpose to improve the efficiency of the measurement.

Here, we discuss why the automated stopping criterion for the spectral measurement with AL works. In Fig. 1a, the measurement energy points in the spectral measurement with AL jump backward and forward in the whole spectral energy range. Thus, GPR fitting can approximate the rough shape of the spectrum in the early stage of the measurement, and the generalisation error becomes small. Fine spectral features, such as satellite peaks, appear as the measurement progresses; however, these measurements are not very relevant to the improvement in the test error between different thresholds. In other words, the method works with a balance between ‘exploration’ and ‘exploitation’. Exploratory sampling is more important than exploitative sampling to reduce the generalisation error and stop the experiment with a minimum number of measurements in the spectral measurement

with AL. Alternatively, exploitative sampling becomes effective when one wants to measure detailed spectral features. This type of sampling becomes possible by using an acquisition function proposed in the literature²⁰. The utilisation of prior knowledge regarding spectra has the potential to design an acquisition function; however, this subject is a topic for future research.

In conclusion, we applied the stopping criterion based on the stability of the expected generalisation errors for the XAS spectral measurement with AL. This stopping criterion can be evaluated from the self-contained information of the GPR fitting. It is revealed that the automated stopping criterion of the spectral measurement gives an approximated XAS spectrum with sufficient accuracy. The implementation utilises the application of the state-of-the-art theory of the optimal stopping problem in AL to actual measurements. The proposed method can be applied not only to spectral measurements but also to other types of measurements. An enhancement of the spectral measurement efficiency enables the high-throughput characterisation of materials for the construction of an experimental materials database in the era of materials informatics.

METHODS

Simulation and measurement of X-ray absorption spectra

The simulation of XAS spectra was performed using CTM4XAS software⁵⁰. Ni, Mn and Co $L_{2,3}$ XAS spectra were calculated for Ni^{2+} , Mn^{2+} , and Co^{2+} ions. Crystal field parameters were set to O_h symmetry with $10Dq = 1.0$. The calculated multiplets were broadened with Lorentzian and Gaussian functions of 0.2 eV half-width at half-maximum each.

The XAS experiment was performed at the BL-19B at the Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization, Japan⁵¹. Three types of samples, manganese dioxide (MnO_2) powder and pieces of bulk cobalt (Co) and nickel (Ni), were mounted at the sample manipulator in the vacuum chamber. Mn, Co and Ni $L_{2,3}$ XAS were

Fig. 6 Time costs at stopping timings for various XAS spectral measurements. Comparison of time costs for the spectral measurement with AL at automated stopping timing with several thresholds and the conventional DoE for **a** simulated $L_{2,3}$ XAS spectra of Ni^{2+} , Mn^{2+} and Co^{2+} and **b** experimental $L_{2,3}$ XAS spectra of Ni, MnO_2 , and Co. The ratio between time to measure single spectral intensity t_{meas} and time to change unit energy t_{ene} is set to $t_{\text{meas}}/t_{\text{ene}} = 1$.

obtained at room temperature by the total electron yield method, which measures the sample drain current. XAS spectra were obtained by dividing the sample current I by the mirror current I_0 to negate the intensity variation in the incident X-ray. All simulated and experimental spectra used in this study are shown in Supplementary Fig. 1.

Gaussian process regression

The fundamental idea of spectral measurement with AL is identifying the spectral measurement problem with supervised curve fitting or the regression problems. Energy point x is considered as an explanatory variable in the regression, and the corresponding intensity y is the response variable in terms of the regression analysis. In this section, we briefly explain GPR, which was adopted in the present study to realise AL. The details of GPR are thoroughly described in ref. ³⁹.

For energy points x_n , we assume that the intensity y_i at x_i is modelled as $y_i = f(x_i) + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, \xi^2)$ is the observation noise. In Gaussian process modelling, function $f(x)$ itself is assumed to be a random variable under a Gaussian distribution with mean $\mu_0(x)$ and covariance function $\mathbf{K}(x, x')$; hence, $\mathbf{y}$ is also a realisation of the Gaussian random variable with mean $\mu_0(x)$ and variance $\sigma^2(x) = \mathbf{K}(x, x) + \xi^2$. Given a collection of observations $\mathbf{y}$ corresponding to the collection of energy points $\mathbf{X}_n = (x_1, \dots, x_n)$, the mean function and the variance function of the Bayesian posterior distribution are denoted by $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$, respectively. To obtain the mean and variance values at a newly observed point x^* , we consider the joint distribution of $\mathbf{y}$ and $f(x^*)$, which is expressed as

$$\begin{pmatrix} \mathbf{y} \\ f(x^*) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_0(\mathbf{X}_n) \\ \mu_0(x^*) \end{pmatrix}, \begin{pmatrix} \mathbf{K}_{n,n} + \xi^2 \mathbf{I}_n & \mathbf{K}_{n,*} \\ \mathbf{K}_{n,*}^\top & \mathbf{K}(x^*, x^*) \end{pmatrix}\right) \quad (8)$$

where $\mathbf{x}$, $\mathbf{K}_{n,n} = \mathbf{K}(\mathbf{X}_n, \mathbf{X}_n) \in \mathbb{R}^{n \times n}$, and $\mathbf{K}_{n,*} = \mathbf{K}(\mathbf{X}_n, x^*) \in \mathbb{R}^n$. The mean function value of the posterior distribution of $f(x^*)$ is obtained as

$$\hat{\mu}(x^*) = \mu_0(x^*) + \mathbf{K}_{n,*}^\top (\mathbf{K}_{n,n} + \xi^2 \mathbf{I}_n)^{-1} (\mathbf{y} - \mu_0(\mathbf{X}_n)) \quad (9)$$

where $\mathbf{k}_n = (\mathbf{K}(x_1, x^*), \dots, \mathbf{K}(x_n, x^*))$. Moreover, the posterior variance at the new energy point x^* is obtained as

$$\hat{\sigma}^2(x^*) = \mathbf{K}(x^*, x^*) - \mathbf{k}_n (x^*)^\top (\mathbf{K}_{n,n} + \xi^2 \mathbf{I}_n)^{-1} \mathbf{k}_n (x^*). \quad (10)$$

The acquisition function is defined as follows

$$a(\hat{\sigma}, \hat{\mu}, t) = \frac{\hat{\sigma}(x^*)}{\sigma_{\max}} + \frac{1}{t} \sqrt{\frac{\hat{\mu}(x^*)}{\mu_{\max}}}, \quad (11)$$

where $\sigma_{\max}$ and $\mu_{\max}$ are maximum standard deviation and mean among consequent measurements at time $1, \dots, t$.

The amplitude θ_1 and the bandwidth θ_2 of the covariance function $\mathbf{K}(x, x')$ (Eq. (1)) and noise variance are predetermined by maximising the marginal likelihood of the Gaussian process model for a similar dataset measured in the past by the same device.

Both the simulated and experimental XAS spectra used in the study have 2000 data points in total. In the present implementation of AL, the data points were divided into three parts: the initial sampling, the pool data and the test data for evaluating the generalisation error, those data sizes were set to 10, 900 and 1090, respectively. Therefore, we assumed the total energy points measured in the conventional DoE are same as the size of the pool data ($N = 900$).

DATA AVAILABILITY

The data obtained in this study are available from the authors upon reasonable request.

CODE AVAILABILITY

The computer codes developed in this study are available from the authors upon reasonable request.

Received: 7 April 2021; Accepted: 5 August 2021;

Published online: 25 August 2021

REFERENCES

- Rajan, K. Materials informatics. *Mater. Today* **8**, 38–45 (2005).
- Mueller, T., Kusne, A. G. & Ramprasad, R. In *Reviews in Computational Chemistry Vol. 29* (eds Parrill, A. L. & Lipkowitz, K. B.) (Wiley, 2016).
- Lookman, T., Alexander, F. J. & Rajan, K. *Information Science for Materials Discovery and Design* (Springer, 2016).
- Agrawal, A. & Choudhary, A. N. Perspective: Materials informatics and big data: realisation of the ‘fourth paradigm’ of science in materials science. *APL Mater.* **4**, 053208 (2016).
- Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M. & Fazzio, A. From DFT to machine learning: recent approaches to materials science—a review. *J. Phys. Mater.* **2**, 032001 (2019).
- Agrawal, A. & Choudhary, A. Deep materials informatics: applications of deep learning in materials science. *MRS Commun.* **9**, 779–792 (2019).
- Schmidt, J., Marques, M. R. G., Botti, S. & Marques, A. L. Recent advances and applications of machine learning in solid state materials science. *npj Comput. Mater.* **5**, 83 (2019).
- Yamawaki, M., Ohnishi, M., Ju, S. & Shiomi, J. Multifunctional structural design of graphene thermoelectrics by Bayesian optimisation. *Sci. Adv.* **4**, eaar4192 (2018).
- Iwasaki, Y. et al. Machine-learning guided discovery of a new thermoelectric material. *Sci. Rep.* **9**, 2751 (2019).
- Iwasaki, Y. et al. Identification of advanced spin-driven thermoelectric materials via interpretable machine learning. *npj Comput. Mater.* **5**, 103 (2019).
- Chen, L. et al. Frequency-dependent dielectric constant prediction of polymers using machine learning. *npj Comput. Mater.* **6**, 61 (2020).
- Zhong, M. et al. Accelerated discovery of CO_2 electrocatalysts using active machine learning. *Nature* **581**, 178–183 (2020).
- Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).
- Liu, P. et al. Machine learning assisted design of γ -strengthened Co-base superalloys with multi-performance optimization. *npj Comput. Mater.* **6**, 62 (2020).
- Mao, Y., He, Q. & Zhao, X. Designing complex architected materials with generative adversarial networks. *Sci. Adv.* **6**, eaaz4169 (2020).

16. Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials property repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
17. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
18. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: The Open Quantum Materials Database (OQMD). *JOM*. **65**, 1501–1509 (2013).
19. Ueno, T. et al. Adaptive design of an X-ray magnetic circular dichroism spectroscopy experiment with Gaussian process modelling. *npj Comput. Mater.* **4**, 4 (2018).
20. Wakabayashi, Y. K. et al. Improved adaptive sampling method utilizing Gaussian process regression for prediction of spectral peak structures. *Appl. Phys. Express* **11**, 112401 (2018).
21. Saito, K. et al. Accelerating small-angle scattering experiments on anisotropic samples using kernel density estimation. *Sci. Rep.* **9**, 1526 (2019).
22. Suzuki, Y., Kotsugi, M., Hino, H. & Ono, K. Automated estimation of materials parameter from X-ray absorption and electron energy-loss spectra with similarity measures. *npj Comput. Mater.* **5**, 39 (2019).
23. Matsumura, T., Nagamura, N., Akaho, S., Nagata, K. & Ando, Y. Spectrum adapted expectation-maximization algorithm for high-throughput peak shift analysis. *Sci. Technol. Adv. Mater.* **20**, 733–745 (2019).
24. Shinotsuka, H. et al. Development of spectral decomposition based on Bayesian information criterion with estimation of confidence interval. *Sci. Technol. Adv. Mater.* **21**, 402–419 (2020).
25. Ozaki, Y. et al. Automated crystal structure analysis based on blackbox optimisation. *npj Comput. Mater.* **6**, 75 (2020).
26. Suzuki, Y. et al. Symmetry prediction and knowledge discovery from X-ray diffraction patterns using an interpretable machine learning approach. *Sci. Rep.* **10**, 21790 (2020).
27. Hino, H. Active Learning: problem settings and recent developments. Preprint at <https://arxiv.org/abs/2012.04225> (2020).
28. Lookman, T. et al. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput. Mater.* **5**, 21 (2019).
29. Terayama, K. et al. Efficient construction method for phase diagrams using uncertainty sampling. *Phys. Rev. Mater.* **3**, 033802 (2019).
30. Tian, Y. et al. Role of uncertainty estimation in accelerating materials development via active learning. *J. Appl. Phys.* **128**, 014103 (2020).
31. del Rosario, Z. et al. Assessing the frontier: Active learning, model accuracy, and multi-objective candidate discovery and optimization. *J. Chem. Phys.* **153**, 024112 (2020).
32. Pestourie, R. et al. Active learning of deep surrogates for PDEs: application to metasurface design. *npj Comput. Mater.* **6**, 164 (2020).
33. Tian, Y. et al. Efficient estimation of material property curves and surfaces via active learning. *Phys. Rev. Mater.* **5**, 013802 (2021).
34. Xie, Y. et al. Bayesian force fields from active learning for simulation of interdimensional transformation of stanene. *npj Comput. Mater.* **7**, 40 (2021).
35. van der Laan, G. & Figuerola, A. I. X-ray magnetic circular dichroism – A versatile tool to study magnetism. *Coord. Chem. Rev.* **277–278**, 95–129 (2014).
36. Thole, B. T., Carra, P., Sette, F. & van der Laan, G. X-Ray circular dichroism as a probe of orbital magnetism. *Phys. Rev. Lett.* **68**, 1943–1946 (1992).
37. Carra, P., Thole, B. T., Altarelli, M. & Wang, X. X-ray circular dichroism and local magnetic fields. *Phys. Rev. Lett.* **70**, 694–697 (1993).
38. Chen, C. T. et al. Experimental confirmation of the X-ray magnetic circular dichroism sum rules for iron and cobalt. *Phys. Rev. Lett.* **75**, 152–155 (1995).
39. Rasmussen, C. E. & Williams, C. K. I. *Gaussian Processes for Machine Learning* (MIT Press, 2006).
40. Schohn, G. & Cohn, D. Less is more: Active Learning with support vector machines. In *Proc. of the Seventeenth International Conference on Machine Learning (ICML2000)* (ed. Langley, P.) (Morgan Kaufmann Publishers Inc, 2000).
41. Krause, A. & Guestrin, C. Nonmyopic Active Learning of Gaussian processes: an exploration-exploitation approach. In *Proc. of the 24th International Conference on Machine Learning (ICML2007)* (Oregon State University, Corvallis, 2007).
42. Altschuler, M. & Bloodgood, M. Stopping Active Learning Based on Predicted Change of F Measure for Text Classification. *IEEE 13th International Conference on Semantic Computing (ICSC2019)* (Newport Beach Marriott Bayview, Newport Beach, CA, 2019).
43. Balachandran, P. V. et al. *Materials Discovery and Design* (Springer, 2018).
44. Seko, A., Maekawa, T., Tsuda, K. & Tanaka, I. Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids. *Phys. Rev. B* **89**, 054303 (2014).
45. Ishibashi, H. & Hino, H. Stopping criterion for active learning based on error stability. Preprint at <https://arxiv.org/abs/2104.01836> (2021).
46. Ishibashi, H. & Hino, H. Stopping criterion for active learning based on deterministic generalization bounds. In *Proc. of the Artificial Intelligence and Statistics (AISTATS2020)* (AISTATS 2020, 2020).
47. Corless, M. R., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J. & Knuth, D. E. On the Lambert W function. *Adv. Comput. Math.* **5**, 329–359 (1996).
48. Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).
49. Jo, T. & Sawatzky, G. A. Ground state of ferromagnetic nickel and magnetic circular dichroism in Ni 2p core x-ray-absorption spectroscopy. *Phys. Rev. B* **43**, 8771–8774 (1991).
50. Stavitski, E. & de Groot, F. M. F. The CTM4XAS program for EELS and XAS spectral shape analysis of transition metal L edges. *Micron* **41**, 687–694 (2010).
51. Fukushi, K. et al. Photon Factory BL-19: a new STXM beamline with wide energy range for aquaplanetology. In *Asteroid Science in the Age of Hayabusa2 and OSIRIS-REx* (Tucson, 2019).

ACKNOWLEDGEMENTS

This work was supported by JST-Mirai Program Grant Numbers JPMJM19G1 and JPMJM21G2. T.U. acknowledges the support of JSPS KAKENHI Grant Number JP18K13984 and QST President's Strategic Grant (Exploratory Research). H.H. acknowledges the support of NEDO Grant Number JPNP18002 and JST CREST Grant Number JPMJCR1761. This work was carried out under the ISM Cooperative Research Program (H30-J-4302 and 2019-ISMRP-4206). The XAS experiment was performed under the approval of the Photon Factory Program Advisory Committee (Proposal No. 2018MP001). The authors thank Dr. Yasuo Takeichi for the support of the experiments at the Photon Factory.

AUTHOR CONTRIBUTIONS

T.U., H.H., and K.O. conceived the project. T.U. H.I. and H.H. wrote the manuscript. H.I. and H.H. performed the computation of the AL. H.I. and H.H. wrote the computer codes. T.U. performed the XAS experiment. All authors discussed the results and reviewed the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-021-00606-5>.

Correspondence and requests for materials should be addressed to T.U.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021