Local Acquisition Function for Active Level Set Estimation

Yuta Kokubun KOKUBUN.Y.AA@M.TITECH.AC.JP
Department of Mathematical and Computing Science, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552 Japan

Kota Matsui MATSUI.K@MED.NAGOYA-U.AC.JP
Department of Biostatistics, Nagoya University
65 Tsurumai-cho, Showa-ku, Nagoya, Japan

Kentaro Kutsukake KENTARO.KUTSUKAKE@RIKEN.JP
RIKEN Center for Advanced Intelligence Project
Nihonbashi 1-chome, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, Japan

Wataru Kumagai KUMAGAI@WEBLAB.T.U-TOKYO.AC.JP
Graduate School of Engineering, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan

Takafumi Kanamori KANAMORI@C.TITECH.AC.JP
Department of Mathematical and Computing Science, Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552 Japan

Abstract

In this paper, we propose a new acquisition function based on local search for active super-level set estimation. Conventional acquisition functions for level set estimation problems are considered to struggle with problems where the threshold is high, and many points in the upper-level set have function values close to the threshold. The proposed method addresses this issue by effectively switching between two acquisition functions: one rapidly finds local level set and the other performs global exploration. The effectiveness of the proposed method is evaluated through experiments with synthetic and real-world datasets.

Keywords: Level Set Estimation, Active Learning, Gaussian Processes

1. Introduction

In the field of materials science, it is often common to assess the quality of manufactured materials based on whether the measured physical properties exceed a certain threshold. For example, in the manufacturing process of multicrystalline silicon ingots for solar cells, only the region where a physical property called "carrier lifetime", measured by microwave photoconductivity decay, exceeds a certain threshold is used in the actual product (Hozumi et al. (2023)). Here, we represent experiments like the one mentioned above as an unknown black-box function f, where the input is the experimental conditions and the output is the physical property value. Then this problem can be seen as the task of identifying a set of candidate conditions for which the value of f exceeds a certain threshold. Such a problem is referred to as a level set estimation problem, and an approach called active level set estimation (Willett and Nowak (2007); Gotovos (2013); Zanette et al. (2019); Inatsu et al.

(2020)) has been proposed as a method to perform accurate level set estimation with as few experiments as possible.

In active level set estimation, some acquisition functions such as Straddle (Bryan et al. (2005); Gotovos (2013)) and RMILE (Zanette et al. (2019)) are primarily used. Straddle is an acquisition function that specifies the next condition based on the magnitude of prediction uncertainty and the proximity to a threshold for function values. On the other hand, RMILE is an acquisition function that specifies the next condition based on the expected improvement in the super-level set. However, these acquisition functions tend to exhibit unstable behavior in problems where the threshold is close to the maximum value of the black-box function and the upper-level set is significantly smaller than the lower-level set.

In this paper, to address the above problem, we propose Localized MILE acquisition function, which is a modification of the MILE acquisition function. Localized MILE is a so-called "local" acquisition function that excels in discovering subsets of the upper-level set. In our proposed method, we efficiently estimate the super-level set by combining Localized MILE, which performs local search, with Uncertainty Sampling, which performs global exploration. We demonstrate the effectiveness of the proposed method through experiments using synthetic data and real-world data.

2. Active Level Set Estimation

Let us formulate the active level set estimation problem. The target function is $f : \mathcal{X} \to \mathbb{R}$. For a given threshold $h \in \mathbb{R}$, the *level set* of f is defined by $\{\boldsymbol{x} \in \mathcal{X} | f(\boldsymbol{x}) > h\}$. More concretely, we say the super level set of f. The observation at $\boldsymbol{x} \in \mathcal{X}$ is given by $y = f(\boldsymbol{x}) + \epsilon$, where ϵ is a Gaussian noise with mean 0 and variance σ_ϵ^2. The problem is to find the level set using the observation $(\boldsymbol{x}_i, y_i), i = 1, \ldots, n$. In practice, the function f is unknown, and the observation is costly. Such a function is called the black-box function. We want to suppress the observation cost as low as possible while keeping the estimation accuracy of the level set of the black-box function. When we select the observation points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in an active way, the problems called the active level set estimation.

2.1 Gaussian Process Model as a Prior of Black-Box Function

The Gaussian Process (GP) model is a versatile method of dealing with black-box functions. For the function f, the prior distribution $\mathrm{GP}(\mu_0, k_0)$ is assumed. This means that for any points $\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_m \in \mathcal{X}$, the prior distribution of $(f(\boldsymbol{x}'_1), \ldots, f(\boldsymbol{x}'_m))$ is given by the multinomial normal distribution with the mean vector $(\mu_0(\boldsymbol{x}'_1), \ldots, \mu_0(\boldsymbol{x}'_m))$ and covariance matrix $(k_0(\boldsymbol{x}'_i, \boldsymbol{x}'_j))_{i,j=1,\ldots,m}$. See (Rasmussen and Williams, 2006) for details of GP modeling.

We assume that the black-box function f as a realization from $\mathrm{GP}(\mu_0, k_0)$. Given the observation $(\boldsymbol{x}_i, y_i), i = 1, \ldots, n$ with $y_i = f(\boldsymbol{x}_i) + \epsilon_i$, the posterior distribution of $(f(\boldsymbol{x}'_1), \ldots, f(\boldsymbol{x}'_m))$ is the Gaussian distribution with mean $(\mu_n(\boldsymbol{x}'_1), \ldots, \mu_n(\boldsymbol{x}'_m))$ and the variance-covariance matrix $(k_n(\boldsymbol{x}'_i, \boldsymbol{x}'_j))_{i,j=1,\ldots,m}$, where $\mu_n(\boldsymbol{x}) = \mu_0(\boldsymbol{x}) + \boldsymbol{k}_0(\boldsymbol{x}, \boldsymbol{x}_{1:n})(\boldsymbol{K}_n + \sigma_\epsilon^2 \boldsymbol{I})^{-1}(\boldsymbol{y}_{1:n} - \boldsymbol{\mu}_0(\boldsymbol{x}_{1:n}))$ and $k_n(\boldsymbol{x}, \boldsymbol{x}') = k_0(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{k}_0(\boldsymbol{x}, \boldsymbol{x}_{1:n})(\boldsymbol{K}_n + \sigma_\epsilon^2 \boldsymbol{I})^{-1}\boldsymbol{k}_0(\boldsymbol{x}_{1:n}, \boldsymbol{x}')$ for $\boldsymbol{y}_{1:n} = (y_1, \ldots, y_n)^T, \boldsymbol{\mu}_0(\boldsymbol{x}_{1:n}) = (\mu_0(\boldsymbol{x}_1), \ldots, \mu_0(\boldsymbol{x}_n))^T, \boldsymbol{k}_0(\boldsymbol{x}, \boldsymbol{x}_{1:n}) = \boldsymbol{k}_0(\boldsymbol{x}_{1:n}, \boldsymbol{x})^T = (k_0(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, k_0(\boldsymbol{x}, \boldsymbol{x}_n))$, and $\boldsymbol{K}_n = (k_0(\boldsymbol{x}_i, \boldsymbol{x}_j))_{i,j=1,\ldots,n}$. An estimator of the level set is given by $\{\boldsymbol{x} \in \mathcal{X} \mid \mu_n(\boldsymbol{x}) - \beta\sqrt{k_n(\boldsymbol{x}, \boldsymbol{x})} > h\}$ with an appropriate $\beta > 0$. In the next subsection, we introduce some existing acquisition functions for the active level set estimation with GP modeling.

2.2 Acquisition Function for Active Level Set Estimation

Let us briefly introduce Maximum Improvement in Level-Set Estimation (MILE) acquisition function tailored for active super-level set estimation. Suppose that input points are selected from a predefined finite subset Ω of $\mathcal{X}$. Remember that the posterior mean is $\mu_n(\boldsymbol{x})$ and the posterior standard deviation is $\sigma_{\mathrm{GP}}(\boldsymbol{x}) = \sqrt{k_n(\boldsymbol{x},\boldsymbol{x})}$. When an additional observation $(\boldsymbol{x}^+, y^+)$ is obtained, the posterior standard deviation is updated to σ_{GP^+}, which does not depend on y^+.

Let us define I_{GP} as the subset in Ω currently classified in the level set of the black-box function and let $|I_{\mathrm{GP}}|$ be the cardinality of I_{GP}. Suppose that I_{GP} is updated to I_{GP^+} when the outcome y^+ is observed at an additional input point $\boldsymbol{x}^+$. The MILE acquisition function (Zanette et al., 2019) is then defined as $\alpha_{\mathrm{MILE}}(\boldsymbol{x}^+) = \mathbb{E}_{y^+|\boldsymbol{x}^+}[|I_{\mathrm{GP}^+}|] - |I_{\mathrm{GP}}|$. The input point that maximizes the MILE attains the maximum expected improvement of the level set. Under the GP model, the above conditional expectation has a simple analytic expression, further investigated in the next section. To study theoretical properties of the MILE, Zanette et al. (2019) introduced the robust variant of MILE (RMILE) as the MILE with uncertainty sampling based on σ_{GP}. See Section 4 of (Zanette et al., 2019) for details.

3. Proposed Method: Localized MILE

In the MILE, the conditional expectation of the size of the updated level set $\mathbb{E}_{y^+|\boldsymbol{x}^+}[|I_{\mathrm{GP}^+}|]$ is used for the input selection. The MILE computes the contribution of all points $\boldsymbol{x} \in \Omega$ to the conditional expectation using the posterior correlation between $\boldsymbol{x}$ and the candidate point $\boldsymbol{x}^+$. The conditional expectation is computed by

$$\mathbb{E}_{y^+|\boldsymbol{x}^+}[|I_{\mathrm{GP}^+}|] = \sum_{\boldsymbol{x}\in\Omega} \int_{y^L_{\mathrm{GP}}(\boldsymbol{x},\boldsymbol{x}^+)}^{\infty} p(y^+|\boldsymbol{x}^+)\mathrm{d}y^+, \tag{1}$$

where $y^L_{\mathrm{GP}}(\boldsymbol{x},\boldsymbol{x}^+) = \frac{\sigma^2_{\mathrm{GP}}(\boldsymbol{x}^+)+\sigma^2_\epsilon}{k_n(\boldsymbol{x},\boldsymbol{x}^+)}\{h + \beta\sigma_{\mathrm{GP}^+}(\boldsymbol{x}) - \mu_n(\boldsymbol{x})\} + \mu_n(\boldsymbol{x}^+)$; See Appendix A in (Zanette et al., 2019).

Though the GP model enables us to compute the conditional expectation $\mathbb{E}_{y^+|\boldsymbol{x}^+}[|I_{\mathrm{GP}^+}|]$, the contribution from the point $\boldsymbol{x}$ far from $\boldsymbol{x}^+$ is thought to be hard to evaluate. For this reason, we propose a localized variant of the MILE as

$$\begin{aligned}\alpha_{\text{loc-MILE}}(\boldsymbol{x}^+) &= \sum_{\substack{\boldsymbol{x}\in\Omega\\ k_n(\boldsymbol{x},\boldsymbol{x}^+)>\delta}} \int_{y^L_{\mathrm{GP}}(\boldsymbol{x},\boldsymbol{x}^+)}^{\infty} p(y^+|\boldsymbol{x}^+)\mathrm{d}y^+ - |I_{\mathrm{GP}}| \\ &= \sum_{\substack{\boldsymbol{x}\in\Omega\\ k_n(\boldsymbol{x},\boldsymbol{x}^+)>\delta}} \Phi\left(\frac{\sqrt{\sigma^2_{\mathrm{GP}}(\boldsymbol{x}^+)+\sigma^2_\epsilon}}{k_n(\boldsymbol{x},\boldsymbol{x}^+)}\{\mu_{\mathrm{GP}}(\boldsymbol{x}) - \beta\sigma_{\mathrm{GP}^+}(\boldsymbol{x}) - h\}\right).\end{aligned} \tag{2}$$

The second expression in the above is obtained by following the computation by Zanette et al. (2019). Note that the original MILE (1) takes the sum over all points in Ω. On the other hand, in the localized MILE, only the input points $\boldsymbol{x}$ satisfying $k_n(\boldsymbol{x}, \boldsymbol{x}^+) > \delta$ contribute to the acquisition function. In such a case, the observation y^+ at $\boldsymbol{x}^+$ will provide reliable information about whether $\boldsymbol{x}$ belongs to the level set or not. The point $\boldsymbol{x}^+$ with a small $k_n(\boldsymbol{x}, \boldsymbol{x}^+)$ tends to be far from $\boldsymbol{x}$. For such a point, y^+ may not be informative to determine whether $f(\boldsymbol{x}) \geq h$. In numerical experiments, we examined the localized MILE with $\delta = 0$ and $\delta = -0.005$.

Let us think of an interpretation for the localized MILE. The data distribution is assumed to be $p(y^+, \boldsymbol{x}^+, \boldsymbol{x}) = p(y^+|\boldsymbol{x}^+, \boldsymbol{x})p(\boldsymbol{x}^+)p(\boldsymbol{x}) = p(y^+|\boldsymbol{x}^+)p(\boldsymbol{x}^+)p(\boldsymbol{x})$, where the marginal probability $p(\boldsymbol{x})$ and $p(\boldsymbol{x}^+)$ are the uniform distribution on Ω. Let us define the event, $R = \{(y^+, \boldsymbol{x}^+, \boldsymbol{x}) \,|\, y^+ \in [y^L_{\mathrm{GP}}(\boldsymbol{x}, \boldsymbol{x}^+), \infty)\}$. Then, a simple calculation yields that the integral in (2) equals the conditional expectation $|\Omega|P(y^+ \in R, k_n(\boldsymbol{x}, \boldsymbol{x}^+) > \delta \mid \boldsymbol{x}^+)$, while the conditional expectation in MILE equals $|\Omega|P(y^+ \in R \mid \boldsymbol{x}^+)$. The input point $\boldsymbol{x}^+$ selected by optimizing $\alpha_{\text{loc-MILE}}$ is expected to take not only the expected improvement but also the reliability of the estimated improvement into account.

The naive localization of the MILE tends to seek only a local region in Ω. That is, a kind of mode collapse can occur. To enhance the global search, we introduce the uncertainty sampling to the localized MILE as follows. Let us define I_{GP^-} as the estimated level set at one step before the currently obtained I_{GP}. Then, the acquisition function of the localized MILE with historical dependency is defined by

$$\alpha^{\text{hist}}_{\text{loc-MILE}}(\boldsymbol{x}^+) = \begin{cases} \alpha_{\text{loc-MILE}}(\boldsymbol{x}^+), & |I_{\mathrm{GP}}| > |I_{\mathrm{GP}^-}|, \\ \sigma_{\mathrm{GP}}(\boldsymbol{x}^+), & \text{o.w.} \end{cases}$$

By introducing the historical dependency, the exploration effect is boosted.

4. Numerical Experiments

In this section, we will conduct a comparative evaluation between the proposed method and existing methods through experiments using both synthetic and real-world data.

4.1 Experimental Setup

Synthetic Data: We consider the following one-dimensional function as a true black-box function, $f(x) = 5 + 1.5\sin(5x) - 2.2\cos(2.3(x-1)) + \frac{\exp(x)}{20}$, $0 \leq x \leq 8$. We used each grid point obtained by dividing the interval [0, 8] into 200 equal parts as a candidate set of inputs. Furthermore, we conducted the analysis under experimental settings consisting of combinations of high threshold value ($h = 8.0$) and low threshold value ($h = 5.0$), as well as large observation error variance ($\sigma^2_\epsilon = 0.5$) and small observation error variance ($\sigma^2_\epsilon = 0.1$).

Real World Data: In this study, we use quality data of silicon (Si) crystals used in solar photovoltaic modules (Miyagawa et al., 2021a,b). This data consists of 24 pairs of $(\boldsymbol{x}, y)$, where $\boldsymbol{x}$ is a 7-dimensional vector representing process temperature, process time, H_2 pressure, H_2 flow rate, RF power, electrode distance, and ALD cycle count, and y is a 1-dimensional continuous quantity called carrier lifetime. We specifically focused on process temperature and H_2 pressure and at first fitted a Gaussian process model using the data of 24 points to predict the carrier lifetime values from these two features. We treated the predictive mean function of this GP model as a pseudo ground truth function

Table 1: F-values and their standard deviations of each method after conducting 100 observations. The alphabets following the dataset names correspond to the settings in the figures.

data \ method	loc-MILE w/hist.dep. ($\delta = 0$)	loc-MILE w/hist.dep. ($\delta = -0.005$)	RMILE	Random
synthetic (a)	0.710 ± 0.031	0.705 ± 0.021	**0.721 ± 0.024**	0.701 ± 0.024
synthetic (b)	0.659 ± 0.066	**0.723 ± 0.077**	0.710 ± 0.039	0.645 ± 0.066
synthetic (d)	**0.974 ± 0.012**	0.969 ± 0.004	0.970 ± 0.003	0.965 ± 0.005
synthetic (e)	**0.945 ± 0.009**	0.943 ± 0.008	0.943 ± 0.013	0.938 ± 0.0
real-world (a)	0.902 ± 0.001	0.903 ± 0.002	**0.905 ± 0.001**	-
real-world (b)	**0.872 ± 0.008**	0.865 ± 0.006	0.869 ± 0.006	-

and conducted experiments by generating pseudo data from it. Specifically, we took inputs corresponding to process temperature on 21 equally spaced grids within the range of 50-300 [°C] and inputs corresponding to pressure on 19 equally spaced grids within the range of 100-1000 [Pa]. By arranging these inputs into 2-dimensional vectors, we generated a total of 399 2-dimensional input candidate points. Furthermore, we used the neural network's output for these candidate points as the pseudo y-values. The thresholds were determined using the 50-percentile and 80-percentile relative to the maximum obtained y-values.

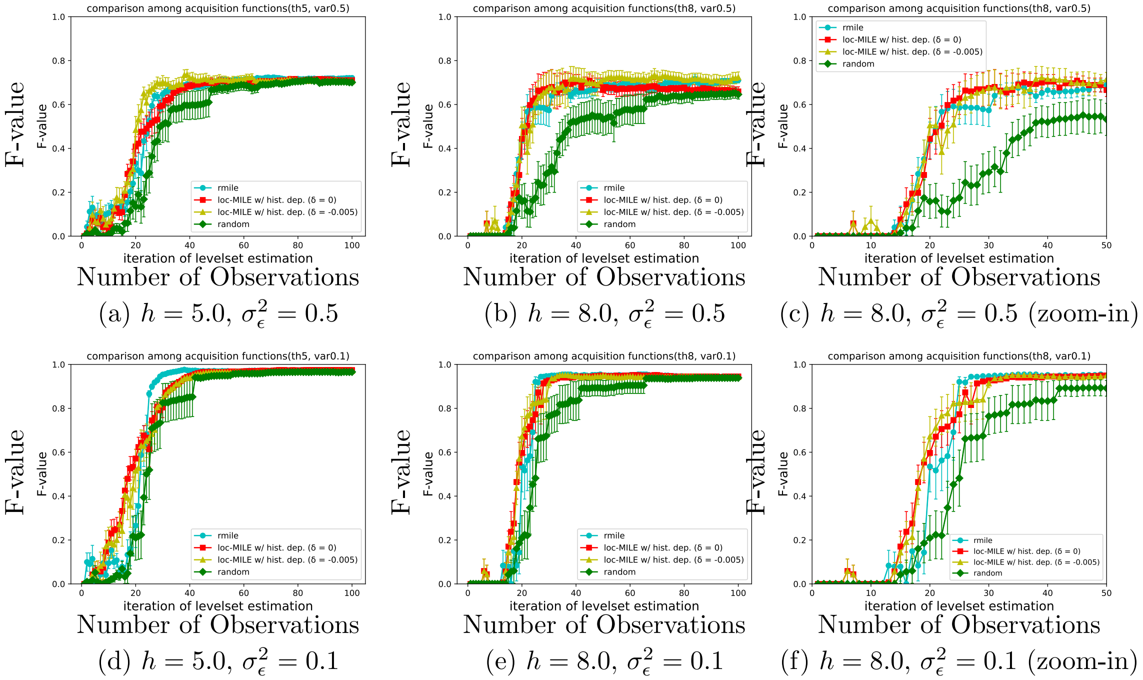
4.2 Results

We compared our proposed method, localized MILE with historical dependency (loc-MILE w/hist.dep. ($\delta = 0, -0.005$)), with RMILE and random search (Random). We conducted 10 trials for each configuration and summarized the results.

Figure 1 shows the evolution of F-values for four methods across different settings of synthetic data. A notable point is that in problems with a high threshold and a large observation error variance (as shown in plots (b) and (c)), the proposed method exhibits significantly better performance. Furthermore, given that the F-value rises quickly with a small number of observations, it is considered that the loc-MILE effectively captures local level sets. It also can be observed that in other settings as well, the proposed method exhibits performance equal to or better than existing methods. Figure 2 shows the evolution of F-values for three methods (other than Random) across different settings of real-world data. For this data as well, the localized MILE shows performance comparable to the RMILE, but the results indicate that there is not much difference between the methods. This is likely because, as shown in Figure 2(c), the generated pseudo-data formed relatively simple level sets. The F-values for level set estimation obtained after 100 observations are shown in Table 1. Overall, the results suggest that the proposed method can estimate the level set more efficiently compared to existing methods.

5. Concluding Remarks

In this paper, we proposed the Localized MILE acquisition function for active level set estimation. The proposed method excels in discovering local level sets quickly, based on the idea of local search. Furthermore, combining with uncertainty sampling, which performs global exploration, the proposed method can efficiently perform level set estimation without getting trapped in "local solutions". As a future work, we will conduct more comprehensive



(a) $h = 5.0$, $\sigma_\epsilon^2 = 0.5$

(b) $h = 8.0$, $\sigma_\epsilon^2 = 0.5$

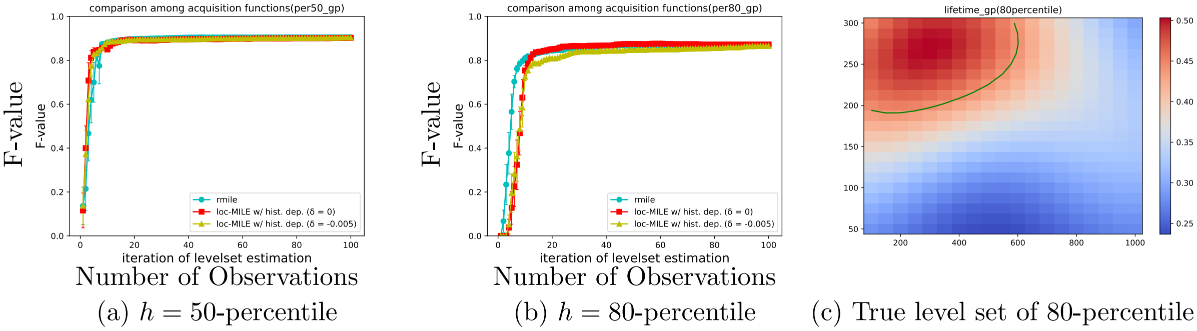
(c) $h = 8.0$, $\sigma_\epsilon^2 = 0.5$ (zoom-in)

(d) $h = 5.0$, $\sigma_\epsilon^2 = 0.1$

(e) $h = 8.0$, $\sigma_\epsilon^2 = 0.1$

(f) $h = 8.0$, $\sigma_\epsilon^2 = 0.1$ (zoom-in)

Figure 1: Results of synthetic data. The red, yellow, cyan and blue lines represent the F-values of loc-MILE w/hist.dep.($\delta = 0$), loc-MILE w/hist.dep.($\delta = -0,005$), RMILE and Random.



(a) $h = 50$-percentile

(b) $h = 80$-percentile

(c) True level set of 80-percentile

Figure 2: Results of real-world data. The red, yellow, cyan lines in (a) and (b) represent the F-values of loc-MILE w/hist.dep.($\delta = 0$), loc-MILE w/hist.dep.($\delta = -0,005$), RMILE. (c) shows the true level sets of pseudo real-world data with 80-percentile as threshold.

experiments to provide a detailed evaluation of the proposed method. Additionally, we will explore theoretical properties such as convergence.

Acknowledgments

This work was supported by Grants-in-Aid from the Japan Society for the Promotion of Science (JSPS) for Scientific Research (KAKENHI grant nos. 20K19871 and 19H04071 to K.M., and 19H04071, 20H00576, and 23H03460 to T.K.).

References

Brent Bryan, Robert C Nichol, Christopher R Genovese, Jeff Schneider, Christopher J Miller, and Larry Wasserman. Active learning for identifying function threshold boundaries. *Advances in neural information processing systems*, 18, 2005.

Alkis Gotovos. Active learning for level set estimation. Master's thesis, Eidgenössische Technische Hochschule Zürich, Department of Computer Science,, 2013.

Shota Hozumi, Kentaro Kutsukake, Kota Matsui, Syunya Kusakawa, Toru Ujihara, and Ichiro Takeuchi. Adaptive defective area identification in material surface using active transfer learning-based level set estimation. *arXiv preprint arXiv:2304.01404*, 2023.

Yu Inatsu, Masayuki Karasuyama, Keiichi Inoue, and Ichiro Takeuchi. Active learning for level set estimation under input uncertainty and its extensions. *Neural Computation*, 32 (12):2486–2531, 2020.

Shinsuke Miyagawa, Kazuhiro Gotoh, Kentaro Kutsukake, Yasuyoshi Kurokawa, and Noritaka Usami. Application of bayesian optimization for improved passivation performance in tio x/sio y/c-si heterostructure by hydrogen plasma treatment. *Applied Physics Express*, 14(2):025503, 2021a.

Shinsuke Miyagawa, Kazuhiro Gotoh, Kentaro Kutsukake, Yasuyoshi Kurokawa, and Noritaka Usami. Application of bayesian optimization for high-performance tiox/sioy/c-si passivating contact. *Solar Energy Materials and Solar Cells*, 230:111251, 2021b.

Carl Edward Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*, volume 1. Springer, 2006.

Rebecca M Willett and Robert D Nowak. Minimax optimal level-set estimation. *IEEE Transactions on Image Processing*, 16(12):2965–2979, 2007.

Andrea Zanette, Junzi Zhang, and Mykel J Kochenderfer. Robust super-level set estimation using gaussian processes. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part II 18*, pages 276–291. Springer, 2019.