Comparison of Gaussian Process Regression Models for AgSe Flash Sintering

Ke Wang[1], Mortaza Saeidi-Javash[2], Minxiang Zeng[2], Zeyu Liu[2], Yanliang Zhang[2], Tengfei Luo[2], Alexander W. Dowling[1*]

[1]*Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA*
[2]*Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA*
adowling@nd.edu

Abstract

Engineered solid-state thermoelectric materials can significantly improve energy efficiency and reduce emissions in modern industry by converting waste heat into electricity. However, the performance of many state-of-the-art thermoelectric materials remains inadequate for adoption beyond niche applications. Current efforts to optimize flash sintering, an important step in additive manufacturing of thermoelectric devices, rely on intuition-driven Edisonian search which can be extremely time-consuming. The alternative way is using Bayesian optimization framework that leverages a probabilistic surrogate model to emulate an expensive objective function and an acquisition function to recommend future experiments that optimally balance exploitation and exploration.

In this paper, we develop a Gaussian Process Regression (GPR) for the surrogate model of the flash sintering of an n-type thermoelectric material, AgSe. It is part of a larger effort to leverage machine learning to optimize the entire additive manufacturing process. We explore whether different hyperparameters tuning methods are consequential for flash sintering datasets and perform retrospective analysis to prove the predictivity of GPR. Finally, we discuss the challenges and opportunities for applying BO for well-tuned GPR to manufacture high-performance solid-state thermoelectric devices.

Keywords: Additive Manufacturing; Data Analysis; Bayesian Optimization; Machine Learning

1. Introduction

1.1. Background

Discover functional material with desired properties is one of the key missions in the material science community, yet the process remains slow and expensive. For example, the discovery of drug-like molecules is facing approximately 10^{23} possibilities, while the synthesized ones are only 10^{8}. Computer-aided molecular design (CAMD) is frequently used to design new functional material, however, the success of that is limited by the accuracy of the description physical assumptions. In recent years, the surge of applying supervised machine learning to leverage data to overcome these challenges has demonstrated great promise for predicting physical properties and is

expected to revolutionize the design of functional materials, which is widely applied in chemical products, materials, and additive manufacturing. Gaussian process regression (GPR) is the most popular machine learning model for predicting material property and is usually applied with Bayesian optimization to accelerate the design of new functional material. It's a non-parametric model that easily fits any dataset with relative lack of prior knowledge and the conjugacy property makes it cost less computation time when comparing with CAMD. Numerous studies have proved the effectiveness of putting GPR under the Bayesian optimization framework boosting the performance of functional materials.

In this paper, we apply GPR to accelerate the development of the TE device. With the urgent demand for wearable electronic devices in daily life, developing high-performance thermoelectric (TE) materials has attracted much attention that fabricating TE devices to convert heat generated by the human body into energy for the electrical device. The physical performance of TE material is evaluated by a dimensionless figure of merit, $\mathrm{ZT} = \frac{\mathrm{S}^2\sigma\mathrm{T}}{\kappa}$ where $\sigma, \mathrm{S}, \kappa$, and T are the electrical conductivity, Seebeck coefficient, thermal conductivity of the material, and absolute temperature, respectively. However, it's relatively difficult to measure the thermal conductivity κ. For convenience, experimentalists usually set power factor $(\mathrm{S}^2\sigma)$, the numerator of ZT, as the evaluation metric.

2. Method

2.1. Decision Variables, Target and Data Pre-processing

Flash sintering is a well-known technique for fabricating high-performance thermoelectric material that contains multiple controllable experimental conditions. The previous study has indicated that the voltage (x_{i1}), pulse duration (x_{i2}), number of pulses (x_{i3}) and pulse delay (x_{i4}) are the most valuable experimental conditions, which we defined them as decision variables represents as $\boldsymbol{x_i}$ where i denotes the sample order. The experimentalist starts with 7 experiments to randomly fabricate TE device and based on the initials; they do single variable control to optimize the experimental conditions for getting the target (y_i), the power factor.

Let $\mathrm{D} = \{(\mathbf{x}_i, y_i), |\mathbf{x}_i \in \mathrm{R}^4, y_i \in \mathrm{R}, i \in 1, \ldots, 31\}$ be a collection of flash sintering dataset, where the vector $\boldsymbol{x_i}$ represents decision variables corresponding to the experimental result y_i. For convenience, we denote the data $\mathrm{D} = (\mathbf{X}, \mathbf{y})$ using matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ and vector $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$. $\boldsymbol{x_i}$, y_i each has different units that directly appliy them into GPR will fail the model performance. The simplest way to deal with it is by standardizing. Here, we use the standardization (1) to transform each decision variable and target centering around the mean equals zero with a unit standard deviation.

$$z = \frac{\big(\boldsymbol{y} - E(\boldsymbol{y})\big)}{Var(\boldsymbol{y})} \tag{1}$$

2.2. Gaussian Process Regression

During the experiment, randomness is unavoidable; it's natural to consider observed error (ε) that we assume it is normally distributed with zero mean and variance σ^2, $\varepsilon \sim N(0, \sigma^2)$ and incorporate it into the model through $y_i = f(\mathbf{x}_i) + \varepsilon$ where $f(.)$ is the objective function represents the behavior of different experimental conditions that can be defined as Gaussian Process (2).

$$f \sim \mathrm{GP}\big(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\big) \qquad \mathbf{x}, \mathbf{x}' \in \mathrm{R}^p \tag{2}$$

Here, for computation simplicity, we set $m(\mathbf{x})$ to zero. $k(\mathbf{x}, \mathbf{x}')$ is the kernel function that mainly determined the behaviors of the GPR and we use Radial Basis Function (k_{RBF}) for model fitting that $\boldsymbol{l} \in R^p$ is the length-scale for the feature $\boldsymbol{x}$. To reduce the number of parameters to address, $\boldsymbol{l}$ can simply set as a scaler for human intuition which means the equally important in every feature. However, such methodology can't satisfy the real-world application scenario that target is usually determined by few features among large multi-dimension features space. As (3) indicates, the length scale $\boldsymbol{l}$ represents the importance of the corresponding feature that the smaller the $\boldsymbol{l}$, the more important that feature. Thus, to find the optimal $\boldsymbol{l}$, log marginal likelihood and cross-validation are introduced in later section and the tuned result provides insight information for the most important feature of $\boldsymbol{x}$.

$$k_{RBF}(x, x') = e^{-\frac{1}{2}\Sigma_{j=1}^{p}\left(\frac{x_j - x_j'}{l_j}\right)^2} \qquad \theta = \boldsymbol{l} \tag{3}$$

We define new inputs values $\mathbf{X}_*$with corresponding prediction $\mathbf{f}_*$**.** Given training data $(\mathbf{X}, \mathbf{y})$ and values of the hyperparameters θ, we can write the outputs $\mathbf{y}$ and $\mathbf{f}_*$ as a multivariate normal (Gaussian) distribution (4) where $\mathbf{K}(\cdot,\cdot)$ is kernel function k evaluated elementwise. The conjugacy properties of it give the prediction results (5).

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} \boldsymbol{m}(\boldsymbol{X}) \\ \boldsymbol{m}(\boldsymbol{X}_*) \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{\sigma^2 I} & \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}_*) \\ \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}) & \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}_*) \end{bmatrix}\right) \tag{4}$$

$$E(\boldsymbol{f}_*) = \boldsymbol{m}(\boldsymbol{X}_*) + \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X})[\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}]^{-1}\big(\boldsymbol{y} - \boldsymbol{m}(\boldsymbol{X})\big) \tag{5a}$$

$$Var(\boldsymbol{f}_*) = \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}_*) - \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X})[\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}]^{-1}\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}_*) \tag{5b}$$

2.3. Hyperparameter Tuning

In this paper, we compare the performance of log marginal likelihood (LML) and cross-validation (CV) for illustrating GPR hyperparameters θ for flash sintering applications. LML uses all the training data $D = (\boldsymbol{X}, \boldsymbol{y})$ to find the θ maximizing the function, which is given by (6). CV reduces the variance of the prediction evaluation; the conjugacy property of GPR largely reduces the computation cost, thus leave-one-out cross-validation (Loo-CV) is adopted for evaluating optimal θ that maximize objective function L_{Loo-CV} (7) where training data is denoted as $D_{-i} = (\boldsymbol{X}_{-i}, \boldsymbol{y}_{-i})$ that $-i$ means all data except sample i.

$$\log p(\boldsymbol{y}|\boldsymbol{X},\theta) = -\frac{1}{2}\boldsymbol{y}^T[\boldsymbol{K}(\boldsymbol{X},\boldsymbol{X}|\theta) + \sigma^2\boldsymbol{I}]^{-1} - \frac{1}{2}\log|\boldsymbol{K}(\boldsymbol{X},\boldsymbol{X}|\theta) + \sigma^2\boldsymbol{I}| - \frac{n}{2}\log 2\pi \quad (6)$$

$$logP(y_i|X_{-i},y_{-i},\theta) = -\frac{1}{2}\log\sigma_i^2 - \frac{(y_i-\mu_i)^2}{2\sigma_i^2} - \frac{1}{2}log2\pi \ \#(7a)$$

$$L_{Loo-CV}(X,y,\theta) = \frac{1}{n}\sum_{i=1}^{n} logP(y_i|X_{-i},y_{-i},\theta) \quad (7b)$$

The length scale of four decision variables is denoted by l_1, l_2, l_3, l_4 respectively and observation error σ is set as a hyperparameter as well; each of them has a search region between $(0, 1)$. Here, the low-dimensional optimization (e.g 5 variables) makes its computationally acceptable for applying grid search; if higher features space is applied, advanced optimization methods (e.g gradient descent) are preferred.

3. Results

3.1. Log Marginal Likelihood (LML) and Leave-one-out Cross-Validation (Loo-CV) identify similar hyperparameter values

Table 1 has shown that we reached the same hyperparameter tuning results with either LML or Loo-CV. The first two rows give the optimal hyperparameters of $\boldsymbol{l}$ when setting $\sigma = 0.1$ as a fixed parameter. The optimal $\boldsymbol{l}$ in both frameworks are exactly same that indicates, under the scenario of fixing observation error, there is not much difference between two methods for hyperparameter tuning at least for flash sintering dataset. Conversely, the last two rows illustrate when add σ as a hyperparameter, there is no difference between two methodologies as well. Figure 2 gives a parity plot of σ tuned with LML optimal hyperparameters, $l_1 = 1, l_2 = 0.687, l_3 = 0.322, l_4 = 1, \sigma = 0.2$; it illustrates the capability of GPR using D_{-i} to predict $\boldsymbol{x}_i$. The x-axis and y-axis are experimental and predicted power factor respectively, the different labels correspond to different batch of experiments (details in 3.2), and error bars are given in one standard deviation.

One thing worth notice is that when adding σ as a tuneable hyperparameter, the l_2 increases from 0.635 of σ fixed model to 0.687, and σ itself increase from the prefixed 0.1 to 0.2. It's related to the trade-off between bias and variance and corresponding to the conclusion that relative complex model (e.g $l_2 = 0.635$) usually obtaining low observation error (e.g $\sigma = 0.1$); while a much simpler model (e.g $l_2 = 0.687$) has higher observation error (e.g $\sigma = 0.2$).

Table 1: LML and Loo-CV comparison

	l_1	l_2	l_3	l_4	σ
σ fixed: LML	1	0.635	0.322	1	0.1
σ fixed: Loo-CV	1	0.635	0.322	1	0.1
σ tuned: LML	1	0.687	0.322	1	0.2
σ tuned: Loo-CV	1	0.687	0.322	1	0.2

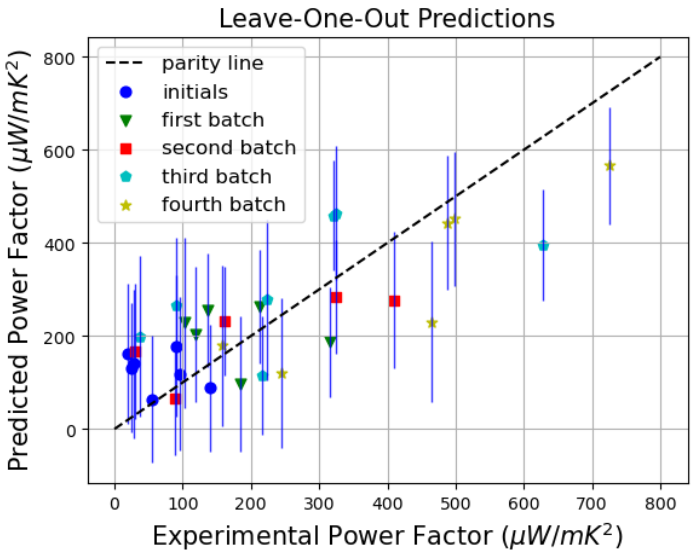

Figure 1: Parity plot of GPR in flash sintering

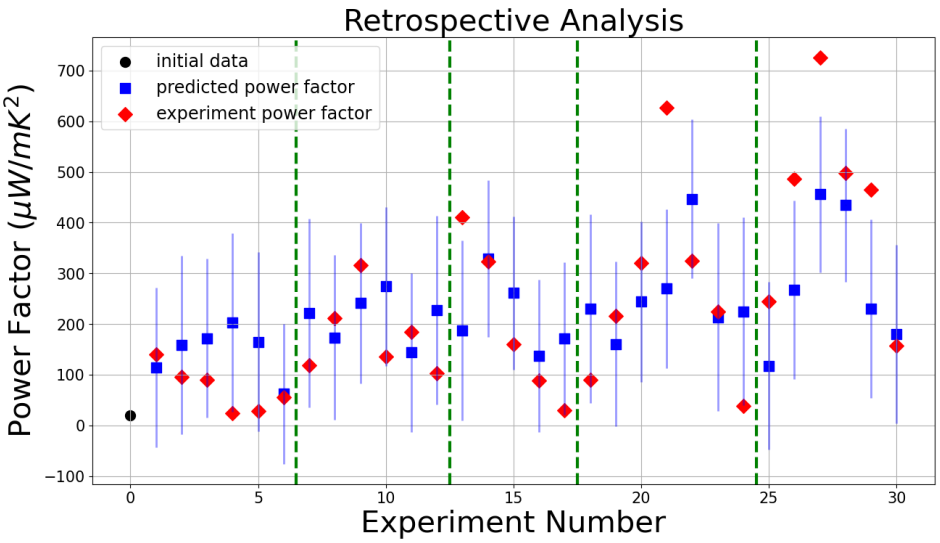

Figure 2: Retrospective analysis of GPR in flash sintering

3.2. Retrospective Analysis

Figure 2 illustrates the retrospective analysis of GPR prediction in flash sintering dataset that data is separated to five sections by the dashed line, each corresponds to an experimental batch where the first section are initials that experimentalist conducting experiments with randomly chosen experimental condition ($\boldsymbol{x}_i$) to form a batch and the rest are based on the initials to implement single variable control for optimizing the power factor (y_i). The experimentalists set different values of voltage (x_{i1}), pulse duration (x_{i2}), and pulse delay (x_{i4}) in each batch and optimize the number of pulses

(x_{i3}) and find out the highest power factor (y_i) is given by $x_1 = 2600V, x_2 = 1.5ms, x_3 = 4, x_4 = 400ms$.

The predicted power factor (squares) is generated iteratively that we first set the hyperparameters as $l_1 = 1, l_2 = 0.687, l_3 = 0.322, l_4 = 1, \sigma = 0.2$ and regards the first sample as training dataset to fit the model and next sample as the testing dataset for prediction; then, we added next sample into the training dataset and iteratively doing so. We plot experimental results (diamonds) and predict results while 25 out of 30 samples are fallen into the predicted (with one standard deviation) bounds. We notice that the GPR performs poorly to predict the power factor largely above the main portions. It's related to the smoothness of the GPR kernel function that most popular kernel functions (e.g RBF, Matern) are all hard to capture the drastic change of target values.

4. Conclusion

In this work, we successfully develop a GPR model for flash sintering of TE material that can predict the majority of power factor around the experimental results and illustrates the two most popular frameworks aren't consequential for identifying optimal hyperparameter θ, at least for our flash sintering dataset. Through both the parity plot (Figure 1) and retrospective analysis (Figure 2), we demonstrate the predictivity of the optimal θ in GPR; specifically, the retrospective analysis indicates the rationale to use GPR as surrogate model for Bayesian optimization framework that GPR precisely predict the behaviour of the objective function $f(.)$ that makes combining it with acquisition function to chooses the next experimental conditions that balance the trade-off between exploration and exploitation possible.

References

Bahk, J. et al. Flexible thermoelectric materials and device optimization for wearable energy harvesting. Journal of Materials Chemistry C 3, 10362-10374 (2015).

Kim, Y. J. et al. High-performance self-powered wireless sensor node driven by a flexible thermoelectric generator. Energy 162, 526-533 (2018).

Saeidi-Javash, Mortaza, et al. "3D Conformal Printing and Photonic Sintering of High-Performance Flexible Thermoelectric Films Using 2D Nanoplates." Advanced Functional Materials 29.35 (2019): 1901930.

Lookman, Turab, et al. "Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design." npj Computational Materials 5.1 (2019): 1-17.