Submitted to the 14th International Symposium on Process Systems Engineering, PSE 2021+. (c) Wang et al., October 30, 2021. Preprint downloaded from dowlinglab.nd.edu

Gaussian Process Regression Machine Learning Models for Photonic Sintering

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

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

Abstract

Novel solid-state thermoelectric (TE) materials have the potential to improve energy efficiency by converting waste heat into electricity. However, the performance of many state-of-the-art TE materials remains inadequate for adoption beyond niche applications. Current efforts to optimize photonic sintering, an important step in additive manufacturing of TE devices, rely on expert-driven trial-and-error search which is often extremely time-consuming and without the guarantee of improvement. Emerging Bayesian optimization frameworks offer a principled approach to intelligentially recommend optimized experimental conditions by balancing exploitation and exploration. In this paper, we develop a Gaussian Process Regression (GPR) machine learning model to predict the outcomes of photonic sintering of aerosol jet printed n-type $Bi_2Te_{2.7}Se_{0.3}$ TE films. We compare hyperparameter tuning methods and perform retrospective analysis to quantify the predictivity of GPR. Finally, we discuss the challenges and opportunities for Bayesian optimization of both photonic sintering process and large additive manufacturing processes to create novel high-performance solid-state TE devices.

Keywords: Additive Manufacturing; Data Science; Bayesian Optimization; Machine Learning; Gaussian Process Regression

1. Introduction

1.1. Background

Discover functional materials with desired properties is a central goal of material science and engineering; yet materials discovery and optimization is often slow and expensive. For example, out of the 10^{23} possible drug-like molecules, only 10^{8} have been synthesized. (Elton et al., 2019) Computer-aided molecular design (CAMD) is frequently used to design new functional material, however, its success is usually limited by the accuracy and efficiency of the physical models. (Austin et al., 2016) Supervised machine learning has demonstrated great promise for predicting the physical properties of material and revolutionizing the design process. (Lookman et al., 2019) For example, Gaussian Process Regression (GPR) and Bayesian optimization have been shown over the past decade to accelerate the design and manufacturing of new functional material. (Wang et al., 2022)

In this paper, we apply GPR to model photonic sintering of aerosol jet printed n-type $Bi_2Te_{2.7}Se_{0.3}$ TE films. With the urgent demand for waste energy recovery and wearable electronic devices, developing high-performance TE materials has attracted much attention. For example, TE devices can harvest heat from the automobile exhaust to improve its overall efficiency. (Han et al., 2018) They can also convert heat generated by the human body into energy for the electronic device, e.g., wearable medical monitors. (Jiang et al., 2020) The thermoelectric performance of TE materials is typically evaluated by a dimensionless figure of merit, $zT = \frac{S^2\sigma T}{\kappa}$, which depends on the electrical conductivity (σ), Seebeck coefficient (S), thermal conductivity (κ), and absolute temperature (T). Among these, the power factor, $(S^2\sigma)$, is usually of the utmost importance in developing new TE materials platforms.

2. Method

2.1. Decision Variables and Data

Photonic sintering is a well-known technique for fabricating high-performance thin-film TE material. (Yu et al., 2017) For photonic sintering, there are four key process variables to optimize: voltage (x_{i1}), pulse duration (x_{i2}), number of pulses (x_{i3}), and pulse delay (x_{i4}). For each experiment i, photonic sintering is performed at conditions $\boldsymbol{x}_i$ and the power factor (y_i) is measured. Saeidi-Javash et al. (2019) described the experimental procedure used to collect the thermoelectric properties under various photonic sintering conditions in this paper. Figure 1 summarizes the photonic sintering dataset. In Group 0, seven experiments were performed with one pulse, and voltage and pulse duration systematically varied based on expert intuition. The remaining groups, 1 – 4, are one dimension sensitivity analysis in which voltage (x_{i1}), pulse duration (x_{i2}), and pulse delay (x_{i4}) are fixed in each group and the number of pulses (x_{i3}) is systematically varied. The GPR machine learning models are then trained using these data in this work.

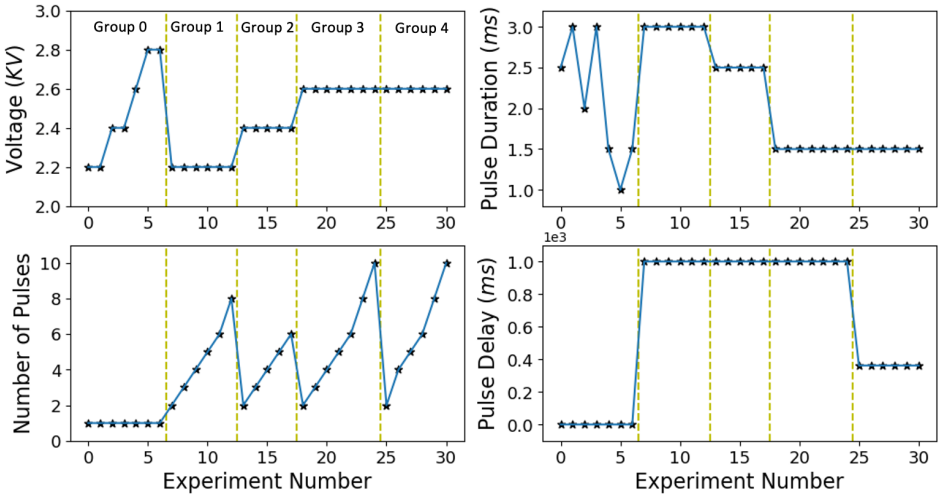

Figure 1: The four photonic sintering optimization variables were systematically varied, based on expert intuition, across 31 experiments which are divided into five groups.

Let $\mathrm{D} = \{(\boldsymbol{x}_i, y_i), |\boldsymbol{x}_i \in \mathrm{R}^4, y_i \in \mathrm{R}, i \in 1,\ldots,31\}$ be a collection of 31 photonic sintering experiments (Figure 1). For convenience, we denote the data $D = (\boldsymbol{X}, \boldsymbol{y})$ using matrix $\boldsymbol{X} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^T$ and vector $\mathbf{y} = (y_1,\ldots,y_n)^{\mathrm{T}}$. However, each dimension of $\boldsymbol{x}_i$, as well as y_i, has different units. To address this, each dimension of *D* is standardized using the mean (expected value) and standard deviation where $\boldsymbol{x}_j$ is the *j*th column of $\boldsymbol{X}$:

$$\boldsymbol{y} \leftarrow \frac{\big(\boldsymbol{y} - E(\boldsymbol{y})\big)}{\sqrt{Var(\boldsymbol{y})}}, \qquad \boldsymbol{x}_j \leftarrow \frac{\big(\boldsymbol{x}_j - E(\boldsymbol{x}_j)\big)}{\sqrt{Var(\boldsymbol{x}_j)}} \tag{1}$$

2.2. Gaussian Process Regression

Gaussian Processes (GPs) are non-parametric probabilistic models that are well-known to emulate expensive continuous functions, $f(.)$, by interpolating between training data.

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

The output of a GP model is a normally distributed random variable fully specified by the mean function, $m(\boldsymbol{x}) = E[f(\boldsymbol{x})]$, and the kernel function $k(\boldsymbol{x}, \boldsymbol{x}') = E\big[\big(f(\boldsymbol{x}) - m(\boldsymbol{x})\big)\big(f(\boldsymbol{x}') - m(\boldsymbol{x}')\big)\big]$. (Rasmussen, 2003) The kernel function determines how the GP model interpolates between the encoded data D. In doing so, the kernel function also specified the uncertainty in GP predictions. $k(.,.)$ contains hyperparameters which often include the length-scales, denoted by $\boldsymbol{l} \in R^p$, for each dimension of the input data $\boldsymbol{X}$. The smaller the l_j, the more important that corresponding feature (x_j). Training the hyperparameters is often performed by using log marginal likelihood or cross-validation methods. The optimal length-scales $\boldsymbol{l}$ help identify which features are most important. For simplicity, we set $m(\boldsymbol{x})$ to zero and use Radial Basis Function (k_{RBF}) defined in Eq. (3), where $\boldsymbol{l} = (l_1, l_2, l_3, l_4)^T$.

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

We define new inputs values $\boldsymbol{X}_*$ with corresponding prediction $\boldsymbol{f}_*$. Given training data $(\boldsymbol{X}, \boldsymbol{y})$ and values of the hyperparameters θ, we can write the outputs $\boldsymbol{y}$ and $\boldsymbol{f}_*$ as a multivariate normal (Gaussian) distribution, Eq. (4), where $\boldsymbol{K}(.,.)$ kernel function $k(.,.)$ is evaluated elementwise. Moreover, we assume each measurement is corrupted by normally distributed observation error ε with zero mean and variance σ^2, $\varepsilon \sim N(0, \sigma^2)$.

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

The conjugacy properties of multivariate Gaussian distribution give (Bishop, 2006):

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

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

2.3. Hyperparameter Tuning

A key step in GP modeling is training the hyperparameters. We start by comparing the performance of log marginal likelihood (LML) Eq.(6) and cross-validation (CV) Eq.(7) for training the length scales of each dimension (l_1, l_2, l_3, l_4) and the optional observation

error σ as hyperparameters. LML (a.k.a. maximum likelihood estimation, MLE) uses all the training data D to find the hyperparameter values which maximize the log-likelihood function:

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

In contrast, CV uses only a subset of the data to reduce the variance of the prediction evaluation. The LML is computed with data $D_{-i} = (\boldsymbol{X}_{-i}, \boldsymbol{y}_{-i})$ where $-i$ denotes all data except sample i:

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

The conjugacy property of GPR greatly reduces the computation cost of evaluating Eq. (7a). The overall leave-one-out CV (Loo-CV) likelihood function is computed by averaging all the leave-one-out samples:

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

The domain knowledge of experimentalists believed that the four proposed variables are all influential for determining power factor (y_i). To incorporate this prior knowledge, we bounded hyperparameter values, including $\boldsymbol{l}$ and σ, between 0 and 1 for this preliminary analysis, since a large value for l_j would imply that dimension j is not important.

3. Results

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

We start by comparing LML and Loo-CV hyperparameter training approaches for the photonic sintering data. Table 1 shows LML and Loo-CV identify identical optimal hyperparameters using grid search. The first two rows correspond to optimizing $\boldsymbol{l}$ with fixed $\sigma = 0.1$ which is informed by the experimental observation error. The optimal $\boldsymbol{l}$ obtaining with LML and Loo-CV methods are the same which suggests the simpler method, LML, is adequate for this photonic sintering dataset. Conversely, the third and fourth rows consider both $\boldsymbol{l}$ and σ as optimized hyperparameters. With σ considered as a tuneable hyperparameter, σ increases from 0.1 to 0.2, and l_2 increases from 0.635 to 0.687. These changes reflect the trade-off between bias and variance (Bishop, 2006) and correspond to the conclusion that relatively more complicated model (e.g., $l_2 = 0.635$) usually obtaining low observation error (e.g., $\sigma = 0.1$), while a simpler model (e.g., $l_2 = 0.687$) has higher observation error (e.g., $\sigma = 0.2$).

Next, we consider the predictive uncertainty of the GP model. Figure 2 is a parity plot for LML optimal hyperparameters ($l_1 = 1, l_2 = 0.687, l_3 = 0.322, l_4 = 1, \sigma = 0.2$). This plot shows the leave-one-out predictions with the GP model. The x-axis and y-axis are experimental and predicted power factor, respectively. The five symbols demark groups of experiments. The error bars show one prediction standard deviation from the GP model. We observed that 27 out of 31 predictions are within one prediction standard deviation of the parity line, suggesting the GPR model successfully emulates the relation between experimental conditions $\boldsymbol{x}_i$ and target y_i.

Table 1: Comparison of hyperparameter values from LML and Loo-CV training

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

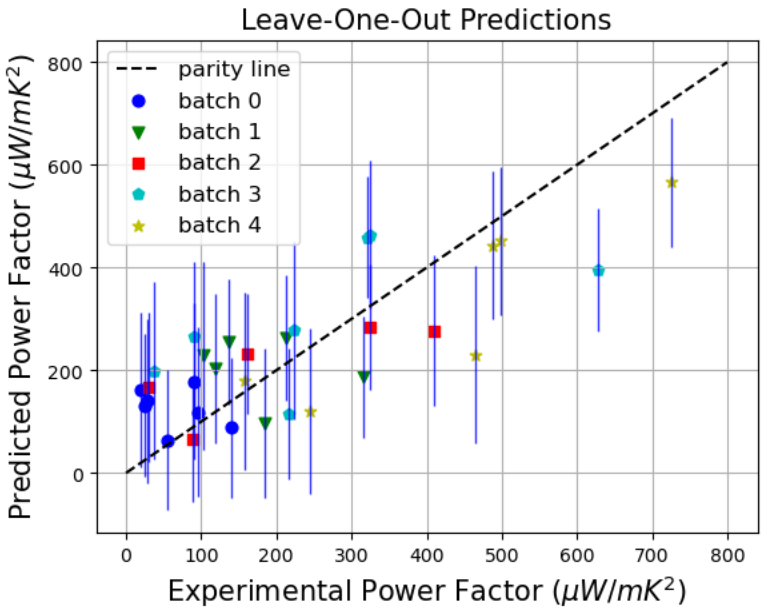

Figure 2: Parity plot of GPR prediction and experimental power factor R in photonic sintering.

3.2. Retrospective Analysis

Figure 3 illustrates the retrospective analysis of GPR prediction in the photonic sintering dataset. The predicted power factors (squares) are generated iteratively (with hyperparameters $l_1 = 1, l_2 = 0.687, l_3 = 0.322, l_4 = 1, \sigma = 0.2$ fixed) using all previous data (to the left of each square). For example, the GPR prediction for experiment 5 uses 5 prior observations for training. The red diamonds show the experimentally measured power factor, and the dashed lines demark each experimental group. In this analysis, we observed 25 out of 30 samples fallen into the predicted (within one standard deviation) bounds. This result demonstrates the GPR model predictions improve as additional data are incorporated into the model.

4. Conclusion

In this work, we successfully develop a GP model to predict the power factor of sintered n-type $Bi_2Te_{2.7}Se_{0.3}$TE films as a function of four photonic sintering variables. This analysis shows that LML and Loo-CV hyperparameter tuning methods identify the same optimal hyperparameters. Through both the parity plot (Figure 2) and retrospective analysis (Figure 3), we show the accuracy of the GPR predictions (both mean and uncertainty estimates). These results suggest that the GP models can be integrated into a

Bayesian optimization framework to identify photonic sintering experimental conditions to maximize the thermoelectric power factor.

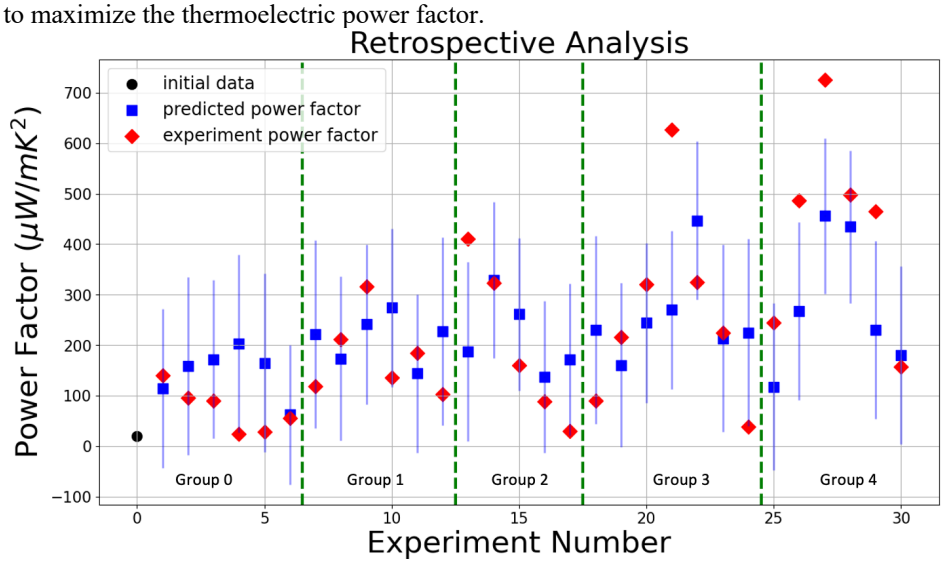

Figure 3: Retrospective analysis of GPR in photonic sintering

References

Elton DC, Boukouvalas Z, Fuge MD, Chung PW. Deep learning for molecular design—a review of the state of the art. Molecular Systems Design & Engineering. 2019;4(4):828-49.

Austin ND, Sahinidis NV, Trahan DW. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. Chemical Engineering Research and Design. 2016 Dec 1;116:2-6.

Jiang C, Ding Y, Cai K, Tong L, Lu Y, Zhao W, Wei P. Ultrahigh performance of n-type Ag2Se films for flexible thermoelectric power generators. ACS applied materials & interfaces. 2020 Feb 3;12(8):9646-55.

Saeidi-Javash M, Kuang W, Dun C, Zhang Y. 3D Conformal Printing and Photonic Sintering of High-Performance Flexible Thermoelectric Films Using 2D Nanoplates. Advanced Functional Materials. 2019 Aug;29(35):1901930.

Bishop CM. Pattern recognition. Machine learning. 2006 Feb;128(9).

Rasmussen CE. Gaussian processes in machine learning. InSummer school on machine learning 2003 Feb 2 (pp. 63-71). Springer, Berlin, Heidelberg.

Yu M, Grasso S, Mckinnon R, Saunders T, Reece MJ. Review of flash sintering: materials, mechanisms and modelling. Advances in Applied Ceramics. 2017 Jan 2;116(1):24-60.

Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Computational Materials. 2019 Feb 18;5(1):1-7.

Ke Wang, Alexander W Dowling, Bayesian optimization for chemical products and functional materials, Current Opinion in Chemical Engineering, Volume 36, 2022, 100728

Han C, Tan G, Varghese T, Kanatzidis MG, Zhang Y. High-performance PbTe thermoelectric films by scalable and low-cost printing. ACS Energy Letters. 2018 Feb 23;3(4):818-22.