Computational Materials Science 156 (2019) 17–25

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

An efficient machine learning approach to establish structure-property linkages

Jaimyun Jung[a], Jae Ik Yoon[a], Hyung Keun Park[a], Jin You Kim[b], Hyoung Seop Kim[a,c,*]

[a] Department of Materials Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea
[b] Pohang Research Lab. Steel Products Research Group 1, POSCO, Pohang 790-785, Republic of Korea
[c] Center for High Entropy Alloys, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea

ARTICLE INFO

Keywords:
Microstructure
Machine learning
Gaussian process regression
Optimization

ABSTRACT

Full-field simulations with synthetic microstructure offer unique opportunities in predicting and understanding the linkage between microstructural variables and properties of a material prior to or in conjunction with experimental efforts. Nevertheless, the computational cost restrains the application of full-field simulations in optimizing materials microstructures or in establishing comprehensive structure-property linkages. To address this issue, we propose the use of machine learning technique, namely Gaussian process regression, with a small number of full-field simulation results to construct structure-property linkages that are accurate over a wide range of microstructures. Furthermore, we demonstrate that with the implementation of expected improvement algorithm, microstructures that exhibit most desirable properties can be identified using even smaller number of full-field simulations.

1. Introduction

While traditional trial and error has served well in designing material microstructures with desired or even enhanced properties and performance, computational means to aid microstructure design are highly desirable for further accelerating materials design. Unfortunately, the linkage between microstructure and properties is a vastly complex one. The sheer complexity in quantifying the geometry of the microstructure coupled with the uncertainty brought about by how different phases mechanically interact given specific structures are only two of many frustrations in computationally aided microstructure design.

There are several approaches available for establishing structure-property linkages. Some of the conventional approaches include analytical approach based on statistical continuum theories [1–3], mean-field approach based on Eshelby's inclusion problems [4–6], numerical approach based on finite element method (FEM). The approach based on statistical continuum theories, when successfully established, is computationally very cheap. Nonetheless, the approach is hindered by difficulties regarding the derivation of analytic expressions for Green's function kernels and the convergence of the series expansions employed in the approach [7,8]. Mean-field approach is also known to be computationally cheap. However, the approach is unable to predict stress or strain localization and often cannot take into account complex morphology and spatial distribution of differing phases. The numerical approach involving FEM is often limited by its computational cost. To date, several works that employed a framework that establishes process-structure-property (PSP) linkages using low dimensional representation of microstructures and regression models as surrogate models are available [9–14]. In the framework, microstructures are first quantified using two-point correlation statistics. Then, principal component analysis (PCA) is applied to the statistics to reduce its dimension. Finally, a surrogate model is constructed using the low dimensional representation of microstructures as input variables. Some of these works employed full-field simulations to capture the PSP linkages [9–14]. In particular, Latypov et al. [9] captured structure-property (SP) relations using microstructure based simulation results on synthetic microstructures for multivariate regression.

The advantage of full-field simulations with synthetic microstructures is that these simulations can quantitatively evaluate how microstructural variations affect the mechanical properties of materials [9,15]. This type of modelling approach already demonstrated its effectiveness in describing and predicting mechanical behavior of composites [16], steels [17–20], and other alloys [21–23]. Despite their merits, microstructure based modelling techniques suffer from their high computational cost. In the framework introduced above, the issue

* Corresponding author at: Department of Materials Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea.
E-mail address: hskim@postech.ac.kr (H.S. Kim).

https://doi.org/10.1016/j.commatsci.2018.09.034
Received 28 May 2018; Received in revised form 13 September 2018; Accepted 15 September 2018

is mainly addressed by constructing an approximate model. Approximate model, or surrogate model, is any model that can mimic the behavior of full-field models at a fraction of their computational cost. Nevertheless, any surrogate model to replace full-field simulations will still require output from the simulations. Consequently, the time it takes to develop an accurate surrogate will be proportional to the number of the simulations necessary. Therefore, robust surrogate modelling with an efficient sampling scheme is necessary to expedite the process of constructing SP linkages over a wide range of microstructures.

New opportunities for surrogate modelling have opened up with the recent surge of interest in machine learning (ML) techniques. Many ML methods such as artificial neural network models [24], Gaussian process (GP) based methods [25], decision-tree and random forest models [26], and kernel-based methods [27] can serve as accurate surrogate models. Among these methods, GP methods offer a principled approach in dealing with model uncertainty. In particular, Gaussian process regression (GPR) can predict unobserved values as well as their uncertainty. This allows one to selectively conduct full-field simulations where high uncertainty in the ML model is expected. Furthermore, GPR is known to be suitable for optimization of expensive black-box simulations via Efficient Global Optimization (EGO) method based on expected improvement (EI) algorithm [28], making GPR a very attractive approach for microstructure optimization using full-field simulations. In this work, we extended the recently developed data-driven frameworks for SP linkages [9–14] by combining synthetic microstructure based simulations and GP based ML to construct SP linkages. Furthermore, we implemented the EI algorithm to find optimal microstructures within a large microstructure database. We first demonstrate effectiveness of the approach in constructing an accurate SP linkages over 1100 synthetic two-phase microstructures using only a fraction of the microstructures for full-field simulations. Afterwards, by implementing the EI algorithm, we show that the approach can search for an optimal microstructure that maximizes specific property within the dataset with even fewer full-field simulations.

2. Methods

2.1. Summary of the method

The overall schematic for the proposed approach is presented in Fig. 1. Firstly, a database consisting of large amount of synthetic microstructures are generated (Section 2.2). The microstructures are then quantified using two-point correlation statistics, which has been successfully used in microstructure quantifications for surrogate models [9–14]. Following the works that establish PSP linkages [9–14], the two-point correlation statistics of synthetic microstructures is projected into a low-dimensional space. In our work, multi-dimensional scaling (MDS) instead of PCA was used as dimensionality reduction technique for better between-data distances preservation (Section 2.3). Microstructure based simulations were conducted with randomly sampled microstructure dataset. The GPR model was trained with the simulated results using low-dimensional microstructural variables as input features. Afterwards, full-field simulations were preferentially conducted with microstructure data with highest predicted variance (Section 2.5), or highest expected improvement for optimization (Section 2.6), until a specified stopping criteria was reached. There are limited work regarding the stopping criteria [29–32]. These works typically utilized 0.001 to 0.002 as stopping criteria. In this study, we adopted a stopping criteria that terminated the GPR training when the maximum predicted variance, or EI, reached below 0.0015. GPflow [33] was used to implement GPR and Matlab 2015a was used to perform MDS and PCA. A Python script was written to implement the overall algorithm shown in Fig. 1.

2.2. Microstructure generation

In order to construct SP linkages over a wide range of microstructural features, a database filled with microstructures reflecting those features is necessary. In this work, a database consisting of two-phase microstructures composed of $27 \times 27 \times 27$ voxels were generated with periodic boundary condition using the open-source software DREAM.3D [34]. The size of the synthetic microstructures were selected based on the works by Latypov et al. [9].

We adopted a similar technique used by Latypov et al. [9] to generate different classes of microstructures by controlling the volume fraction, size distribution, and aspect ratio of second phase inclusions. Because inclusions are not allowed to overlap in the inclusion/matrix microstructures in DREAM.3D, as specified in [9], inclusion/matrix microstructures with inclusion volume fractions in the range of 5% to 35% and 65% to 95% were generated. Equiaxed two phase

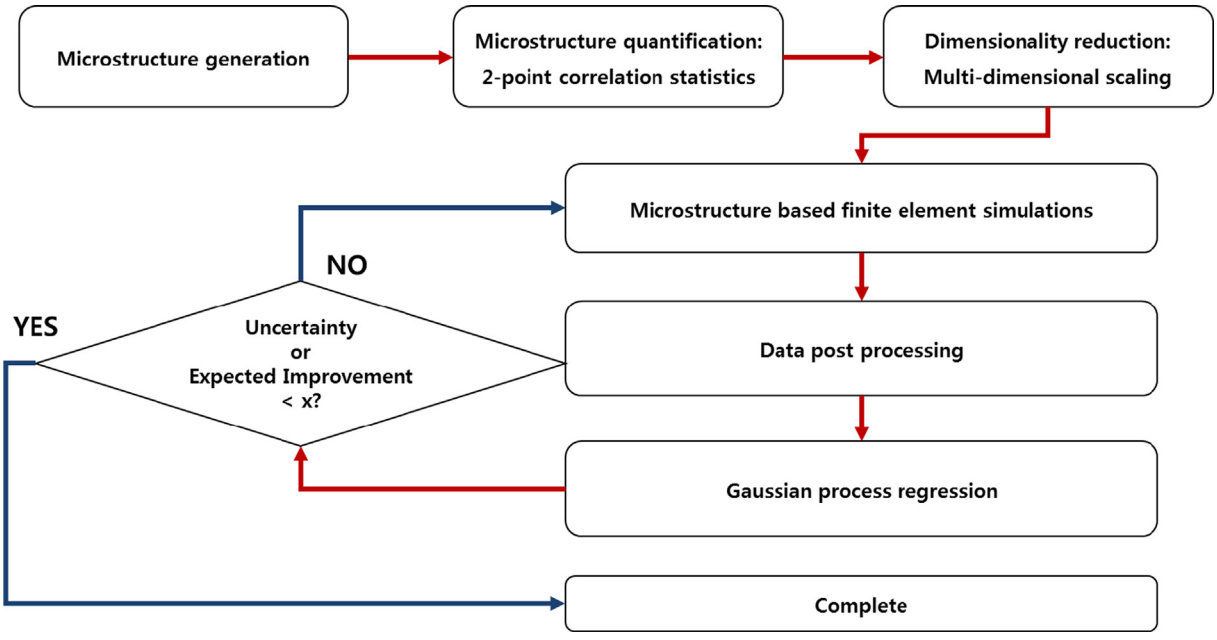

Fig. 1. Schematic of the proposed framework.