Assignment 1 Question 1

Find a data science application published in an online journal (which one does not matter) and summarize the paper. Including an overview of the approach, a link to the online version of the paper, a reference to the paper, a description of the dataset used, as well as strength and weaknesses (450 to 550 words). Your submission has to be in your own words, copying and pasting the abstract or anything else from the paper will not gain you any marks.

Journal Paper – Data Science Application

Title – Structure-property linkages using a data science approach: Application to a non-metallic inclusion/steel composite system

Summary

The paper discusses the feasibility of a data science approach in the field of mechanical and material linkages for certain inelastic properties of these composite materials. The process of extracting features and attributes of these composite materials that affect the linkages revolve around the Homogenization theories. Choosing the right and adequate set of microstructure measures of the material is important to ensure the accuracy of this theory. This is the reason, that sometimes practicing the homogenization theories are not enough and thus there is a need to resort to numerical approaches for exploring complex linkages between a microstructure of a material and the other related effective properties. This is where there is a need for data science approach because identification of the prominent microstructure measures that influence properties of the material on a large basis.

Kalidindi and his team designed a mathematical framework for the systematic quantification of microstructure based on the concept of n-spatial correlations. There are issues with this method as the dataset is very large and unwielded. This large dataset leads to requirement of a data analytic tool for computationally efficient mining. The technique decided and used is “Principal Component Analysis (PCA) dimensionality reduction techniques” to extract high fidelity structure-property linkages. This paper explored the approaches for non-linear attributes like plasticity in a steel composite material system.

Approach

The data science approach mainly consists of three steps -

1. Generating a standardization dataset that comprises of a test data of microstructures with the attributes of interest and suitable physics-based numerical models.

2. Extracting the useful attributes – for instance, the quantitative measures of the microstructures – from the dataset using the data analytic approaches
3. And using the algorithm to establish the linkages of authenticated structure-property

The theory being used is the homogenization theory to formulate the equation for the calculations. The beginning series expression used to derive the final equation is:

$$C^* = \overline{C} - \overline{C'\Gamma C'} + \overline{C'\Gamma C'\Gamma C'} - \dots$$

Where, C* denotes the effective property and each term in this infinite series is the integration over the microstructure volume.

For the spatial statistics being used, the series need to include the higher-order spatial statistics. The purpose of the data science approach is to identify the important terms that affect the microstructure properties and the coefficients. The final formula derived is a polynomial combination of principal components that identify the dependencies between them is the following:

$$C^* \approx A_o + \sum_i A_i \tilde{\alpha}_i$$

The approach for linear linkage is – first PCA is applied on the spatial correlations in an unsupervised dimensionality reduction based on structure metrics – which means no assumptions are made of the properties of the data. Therefore, for the non-linear linkages, **the approach of cross-validation** is used since the data is a large dataset and it has high cost for generating data and a lot of computationally power as well. This method ensures maximum utilization of the available information in making the model.

Dataset

Generating the dataset is important for a data science approach to train and work. The data comprises of microstructures that can be extracted from measurements or generated manually. They need to be discretize and digitized and then the required properties can be obtained via mechanical response directly from the measurements and numerical models.

When the approach was evaluated for nonlinear phenomenon such as plasticity, shapes, sizes, and distributions in a steel matrix with certain limited dataset (900 microstructures) , the approach proved to be versatile. The dataset seemed to yield adequate results given the fact that the microstructures can have a variety of values.

Strengths and Weaknesses of the approach

Strengths:

- Using data science approach, gives a higher accuracy to the results
- Even though requires extra effort, using data science approach ensures proper objective selection of the structure measures. The automated code leads to selecting correct and adequate measures.
- The model is highly flexible and can use both simple and complex microstructure

Weaknesses:

- The amount of effort needed for establishing the data science linkages is higher than when using simple conventional approaches. Combining various traditional approaches with simple conventional methods can give equally good models with less effort
- The training data used is a set of 900 microstructures which limit the assumption and can lead to inaccurate results when faced with a new kind of microstructure

Reference

Gupta, A., Cecen, A., Goyal, S., Singh, A. K., & Kalidindi, S. R. (2015). Structure–property linkages using a data science approach: application to a non-metallic inclusion/steel composite system. *Acta Materialia*, *91*, 239-254.

Link - *Link to paper*