

## Chapter 8

# Applications of Feature Selection and Regression Techniques in Materials Design: A Tutorial

**Partha Dey**

*Academy of Technology, Adisaptagram, Hooghly, India*

**Joe Bible**

*Eunice Kennedy Shriver National Institute of Child Health and Human Development, USA*

**Swati Dey**

*Indian Institute of Engineering, Science and Technology Shibpur, India*

**Somnath Datta**

*University of Florida, USA*

### ABSTRACT

*Feature selection is considered as an important preprocessing step to data mining and soft computing, whereas regression is a collection of methods to optimally assess the signal from a noisy output. Both seek to arrive at the dependence and relation between different attributes and a target material property. In the present chapter a flock of regression and feature selection techniques are discussed, and the kind of results that can be obtained with each of them has been illustrated with the help of a dataset on steel. The different methods are capable of abstracting data in different forms, thus revealing hidden knowledge from different perspectives. Choosing the most appropriate method depends on the application at hand and the kind of objective that one is looking for.*

DOI: 10.4018/978-1-5225-0290-6.ch008

## **INTRODUCTION**

Design of new and superior materials involves data with a large number of variables. It is very often advisable to assess the importance of these input variables (or predictors) and draw a quantitative interrelation between them and the output (or target) variable. This substantially improves any prediction or design model from different ways. Firstly it gives an overview of the role of different factors in the desired properties of the material and makes design and control of experiments easier by chopping off redundant variables. Secondly, in the case of predicting a material property from its composition and processing variables, selecting the most relevant variables renders the prediction model simpler, more efficient, and often even more accurate (by impeding over-fitting). Thirdly, where it is desirable to obtain a closed-form relationship between the input variables and the final property, a simple expression relating them makes the system and its underlying process explainable in physical terms.

Regression and feature selection have served as important tools to statistical learning and data mining. There are various feature selection and regression techniques for assessing the importance and interrelation of variables. In this chapter we discuss a number of techniques which will be useful for a practitioner or researcher in materials design and engineering, though they are equally applicable to a data set from any field of science, engineering as well as humanities. Finally, we illustrate the working and kind of results that these techniques generate, through a numerical example, focusing on the kind of inference that may be drawn from each method, along with a comment on choosing the most appropriate method.

## **REGRESSION TECHNIQUES BASED ON STATISTICAL LEARNING**

The most classical regression technique is the ordinary least squares (OLS), which works well when the number of predictors is small, the correlation between the predictors is weak and their contributions to the conditional mean response are linear and additive. Departure from these assumptions would require considering more sophisticated methods. As for example, when the number of predictors is greater than the number of observations, OLS is not applicable. The aforementioned task of addressing model creation in the presence of a large number of descriptors can be accomplished via the implementation of one or several applicable regression techniques; Principle Component Regression (PCR), Partial Least Squares (PLS), Sparse Partial Least Squares (SPLS), Least Absolute Shrinkage and Selection Operator (Lasso), Ridge Regression and Elastic Net are examples of such techniques. The first three belong to a class of techniques referred to as latent factor techniques, the last three a class referred to as regularized regression techniques. Both are capable of handling a large number of descriptors with the distinction being that latent factor regression accomplishes this end by performing the analysis using a set of derived components (latent factors) whereas regularization does so by allowing descriptors to contribute disproportionately to the analysis.

Principle Component Analysis constructs a set of vectors (principle components), each of which describes the projection of each descriptor onto a  $p$ -dimensional space. PCR performs regression analysis using this set of principle components. PLS unlike PCR takes into account the response (target variable/property) as well as the response variables in constructing the latent factors used in analysis. PLS derives the latent factors in a way that maximizes the covariance between the latent factors and the target vector. SPLS introduces a sparsity parameter into PLS analysis which adds another dimension to the investigation.

26 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/applications-of-feature-selection-and-regression-techniques-in-materials-design/156831](http://www.igi-global.com/chapter/applications-of-feature-selection-and-regression-techniques-in-materials-design/156831)

## Related Content

---

### Numerical Approach to Thermal Conductivity of Electrorheological Complex (Dusty) Plasmas

Muhammad Asif Shakoori, Misbah Khan, Haipeng Li, Maogang He, Aamir Shahzad and Syed Ali Raza (2024). *Emerging Applications of Plasma Science in Allied Technologies* (pp. 210-226).

[www.irma-international.org/chapter/numerical-approach-to-thermal-conductivity-of-electrorheological-complex-dusty-plasmas/338048](http://www.irma-international.org/chapter/numerical-approach-to-thermal-conductivity-of-electrorheological-complex-dusty-plasmas/338048)

### Effect of Microstructure on Chip Formation during Machining of Super Austenitic Stainless Steel

Mohanad Alabdullah, Ashwin Polishetty, Junior Nomani and Guy Littlefair (2017). *International Journal of Materials Forming and Machining Processes* (pp. 1-18).

[www.irma-international.org/article/effect-of-microstructure-on-chip-formation-during-machining-of-super-austenitic-stainless-steel/176058](http://www.irma-international.org/article/effect-of-microstructure-on-chip-formation-during-machining-of-super-austenitic-stainless-steel/176058)

### Optimization of Hot Extrusion Process Parameters Using Taguchi Based Grey Relation Analysis: An Experimental Approach

Sarojini Jajimoggala (2019). *International Journal of Materials Forming and Machining Processes* (pp. 1-18).

[www.irma-international.org/article/optimization-of-hot-extrusion-process-parameters-using-taguchi-based-grey-relation-analysis/221322](http://www.irma-international.org/article/optimization-of-hot-extrusion-process-parameters-using-taguchi-based-grey-relation-analysis/221322)

### Review of Application of Imidazole and Imidazole Derivatives as Corrosion Inhibitors of Metals

Khaoula Alaoui, Nadia Dkhireche, Mohamed Ebn Touhami and Younes El Kacimi (2020). *New Challenges and Industrial Applications for Corrosion Prevention and Control* (pp. 101-131).

[www.irma-international.org/chapter/review-of-application-of-imidazole-and-imidazole-derivatives-as-corrosion-inhibitors-of-metals/260771](http://www.irma-international.org/chapter/review-of-application-of-imidazole-and-imidazole-derivatives-as-corrosion-inhibitors-of-metals/260771)

### Characterization and Quantification of Polymer Content in Plastic Explosives Using FT-IR Techniques

Elizabeth da Costa Mattos and Rita de Cássia Lazzarini Dutra (2018). *Energetic Materials Research, Applications, and New Technologies* (pp. 288-321).

[www.irma-international.org/chapter/characterization-and-quantification-of-polymer-content-in-plastic-explosives-using-ft-ir-techniques/195310](http://www.irma-international.org/chapter/characterization-and-quantification-of-polymer-content-in-plastic-explosives-using-ft-ir-techniques/195310)