Mining magnetic materials data

DAM Hieu Chi
(Japan Advanced Institute of Science and Technology)

The most important underlying hypothesis of materials researches is that the features of the structure of materials, as well as its derived physical properties has strong multivariate correlations. The task of materials design is to make these correlations clear and to determine a strategy to modify the materials to obtain desired properties. However, such correlations are usually hidden and difficult to uncover or predict by experiments or experience.

For dealing with this issue, data mining methods which can extracting meaningful information and knowledge from large data sets, are attracted a great deal of interest. Motivated by using data mining to solve data-intensive problems in materials science, we develop a method to quantitatively model the multivariate correlations between physical properties of materials and their structures by using sparse modeling. The key idea of our method is to use advanced statistical mining algorithms, in particular multiple linear regression and non-linear regression regularized least-squares [1, 2] to solve the sparse approximation problem on the space of structural and physical properties of materials. We use cross-validation to consistently and quantitatively evaluate the conditional relations of physical properties to all the structural features of the materials in terms of prediction. We apply the method to a data set of more than four thousand transition rare-earth metal alloys. We demonstrate that the obtained sparse model is not only significant for the comprehension of the physics relating to the materials, but also valuable for the guidance of effective material design.

Reference

1) R. Tibshirani, J. R. Statist. Soc. B 58, 267 (1996). B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, Annals of Statistics 32, 409 (2004).
2) C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, MIT Press (2006).