Machine Learning for the Computational Materials Scientist

Chris Fischer,[1] Kevin Tibbetts,[1] Dane Morgan,[2] and Gerbrand Ceder[1,*]

[1]*Massachusetts Institute of Technology*

[2]*University of Wisconsin-Madison*

(Dated: September 26, 2007)

Abstract

Computational materials science (CMS) is currently undergoing an expansion from a focus on understanding and optimizing the properties of known materials to one encompassing the computational design of materials before experimental synthesis. As CMS increasingly turns its eye towards this problem, both the design space and the sheer abundance of data will provide an arena in which machine learning techniques can have significant impact. This talk will outline areas in which there exists a need for machine learning techniques to intelligently make use of both experimental and computed data.

Structure Prediction Commodity computational resources can now be used to search for new compounds with targeted properties in a space of tens of thousands of chemistries and compositions. Modern methods of quantum mechanics based on Density Functional Theory are remarkably accurate in reproducing, and in some cases predicting, the properties of materials ranging from bulk moduli to ionic conductivity. However, calculating properties of materials for which no experimental information is available requires knowledge of structure (i.e., the calculations must be performed on the correct structure to yield an accurate prediction). A key problem in CMS is therefore the prediction of crystal structure as it will play a critical role in materials design. Predicting crystal structure requires a search to be performed in the space of all possible structures, which unfortunately, is both infinite and riddled with complexity precluding the use of a brute force optimization. Search strategies are traditionally split into two different camps; one directly optimizing structure and another based on heuristic rules useful for suggesting likely phases and/or eliminating physically irrelevant states. In this talk we describe a technique leveraging the information content of a large database of experimental crystal structures to drive structure prediction with quantum mechanics. Our data mining technique is used to extract structure correlations, depicted in Figure 1, from a large database of binary metallic alloys that are subsequently used in a predictive setting. These correlations are used to encode historical knowledge and significantly reduce the computational overhead associated with making accurate material property predictions.

High Throughput Computing Owing to continuing improvements in computing power and the availability of robust, thoroughly tested *ab initio* electronic structure codes, it is now possible to compute a detailed electron's eye view of nearly every known inorganic compound. The field of computational materials science is rapidly expanding in both the size of systems as well as the

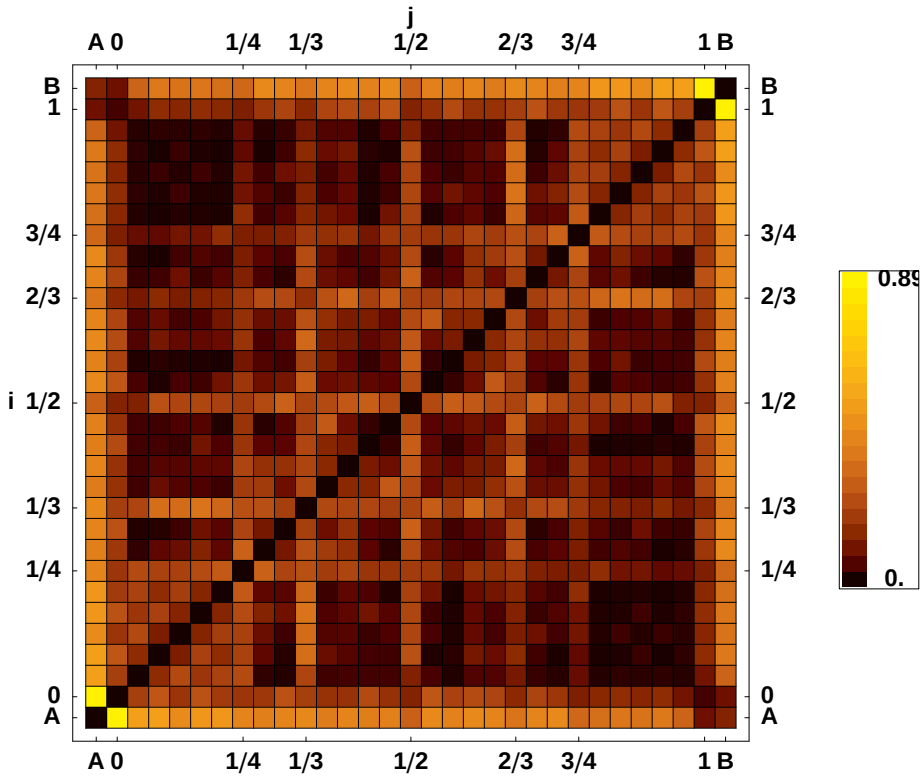

FIG. 1: Crystal structure correlation (mutual information) between pairs of binary alloy compositions extracted from a database of binary metallic alloys. Combinations include pairs of variables where each represents a structure forming at intermediate compositions, and combinations involving an element and the structure forming at a particular composition. Lighter colors indicate stronger correlation.

number of chemistries that it can tackle. In both areas, machine learning can play a key role in distilling the complexity of many-particle systems and extracting useful information from a detailed view of thousands of chemistries. The field is rapidly approaching a data glut in which raw data with very high information content is produced before methods are available to intelligently analyze it. Both data management and analysis of quantum calculations are quickly becoming an area in which machine learning can provide the necessary tools for the computational materials scientist. Managing large amounts of data in a coherent way will play an increasingly important role for the exchange and mining of materials information. Data reporting and representation in materials science, especially with regard to property values, is far from systematic and must be addressed.

* gceder@mit.edu; http://burgaz.mit.edu