Oil & Gas Research

Spain-A applied mathematics Learning Framework for Materials Science: Application to Elastic Moduli of k-nary Inorganic crystalline Compounds

Adegoke Sarajen
Simon Fraser University, Saudi Arabia

Abstract: Materials scientists increasingly employ machine or statistical learning (SL) techniques to accel- erate materials discovery and design. Such pursuits benefit from pooling training data across, and thus being able to generalize predictions over, k-nary compounds of diverse chemistries and structures. This work presents a SL framework that addresses challenges in materials science applications, where datasets are diverse but of modest size, and extreme values are often of interest. Our advances include the application of power or Hölder means to con- struct descriptors that generalize over chemistry and crystal structure, and the incorporation of mul- tivariate local regression within a gradient boosting framework. The approach is demonstrated by de- veloping SL models to predict bulk and shear moduli (K and G, respectively) for polycrystalline inorganic compounds, using 1,940 compounds from a grow- ing database of calculated elastic moduli for metals, semiconductors and insulators. The usefulness of the models is illustrated by screening for superhard materials. In recent years, first-principles methods for calculating properties of inorganic compounds have advanced to the point that it is now possible, for a wide range of chemistries, to predict many properties of a material before it is synthesized in the lab1. This achievement has spurred the use of high-throughput computing techniques as an engine for the rapid development of extensive databases of calculated material properties. Such databases create new opportunities for computationally-as- sisted materials discovery and design, providing for a diverse range of engineering applications with cus- tom tailored solutions. But even with current and near-term computing resources, high-throughput techniques can only analyze a fraction of all possible compositions and crystal structures. Thus, statistical learning (SL), or machine learning, offers an express lane to further accelerate materials discovery and inverse design. As statistical learning techniques ad- vance, increasingly general models will allow us to quickly screen materials over broader design spaces and intelligently prioritize the high-throughput anal- ysis of the most promising material candidates.

One encounters several challenges when applying SL to materials science problems. Although many elemental properties are available, we typically do not know how to construct optimal descriptors for each property, over a variable number of constitu- ent elements. For instance, if one believes that some average of atomic radii is an important descriptor, there are many different averages, let alone possible weighting schemes, that one might investigate. This challenge may be reduced by placing restrictions on the

number of constituent elements or types of chemistries or structures considered, but such re- strictions reduce the generalizability of the learned predictor. Materials science datasets are often also smaller than those available in domains where SL has an established history. This requires that SL be applied with significant care in order to prevent over-fitting the model. Over-fitting leads to predic- tions that are less generalizable to new data than anticipated, such that predictions are less accurate than expected. At the same time, smaller datasets challenge us to use the available data as wisely as possible. This may include leveraging observations related to the smoothness of the underlying physics