Computational Materials Science 112 (2016) 364–367

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

Material synthesis and design from first principle calculations and machine learning

Keisuke Takahashi [a,*], Yuzuru Tanaka [b]

[a] Graduate School of Engineering, Hokkaido University, N-13, W-8, Sapporo 060-8628, Japan
[b] Meme Media Laboratory, Hokkaido University, N-13, W-8, Sapporo 060-8628, Japan

ARTICLE INFO

Article history:
Received 10 October 2015
Received in revised form 6 November 2015
Accepted 8 November 2015

Keywords:
Material informatics
Density functional theory
Machine learning
Material big data

ABSTRACT

Desired material synthesis and design can be directly predicted on the basis of first principle calculations and machine learning. Material big data is constructed based on density functional theory where every possible element combinations are considered and then used as training sets for support vector machines. The predicted material properties for common materials are successfully matched with experimental data. In addition, material combinations based on desired material properties are also able to be predicted. Thus, the proposed work flow becomes the bridge between the material database and designing materials. The approach enables efficient material mining from material big data and could potentially reveal undiscovered desired materials. This approach could also potentially enable targeted material mining from material big data, the unveiling of undiscovered desired materials, and the execution of targeted material synthesis in experiment.

Mining desired materials directly from the periodic table is the ultimate goal for material scientists. The complexity of designing materials in such a manner rests upon the extremely large number of possible combinations of elements where different crystal structures for each combination must be also accounted for. Such complexities of combinations of elements make scientists struggle to dig such a gold mine in experiment. The rise of first principle calculations based on quantum mechanical techniques takes a prominent place in computational material science with the help of supercomputers [1]. The movement towards creating a material database using simulation opens a new way of seeking and designing new materials [2–5]. Such movement is also occurring in the field of catalysts where desired chemical reactions are evaluated using first principle calculations [6,7]. With the rapid increase of supercomputer development, creating material databases using first principle calculations is gaining popularity, yet accessing and manipulating the data for predicting new undiscovered materials has yet to be realized. In other words, the creation of material databases has been limited to materials that have already been discovered and thus have not been used to their full potential.

Here, density functional theory and machine learning techniques are implemented for the creation of material databases and material predictions. In particular, support vector machines within machine learning is implemented in order to fill gaps between the material database and material design. Support vector machine is a particularly successful technique for recognizing patterns where it learns categorized data sets through so called training and reveals the overarching trends in the data set [8]. The trained support vector machine from material database is then applied towards searching and predicting desired material properties and material combinations that possess the desired material properties.

Desired materials can be directly mined from the periodic table with the aid of first principle calculations and machine learning, where the former provides the means of calculating big data and the latter predicts element combinations based on desired properties. Machine learning plays a particularly important role due to its ability to learn behaviors and trends present within the periodic table and make predictions based off the learned trends. The proposed outline shown in Fig. 1 demonstrates how material design and prediction can be achieved from elements in the periodic table. The periodic table is treated as an element pool where the individual elements act as building blocks for the modeling process. All possible combinations, including unrealistic cases, are constructed where different concentrations of elements and crystal structures are accounted for. First principle calculations are performed for each constructed model in order to acquire data on properties such as lattice constant, densities, formation energies, magnetic moment, and bulk modulus. The created database then becomes

* Corresponding author.
E-mail address: keisuke.takahashi@eng.hokudai.ac.jp (K. Takahashi).

http://dx.doi.org/10.1016/j.commatsci.2015.11.013
0927-0256/

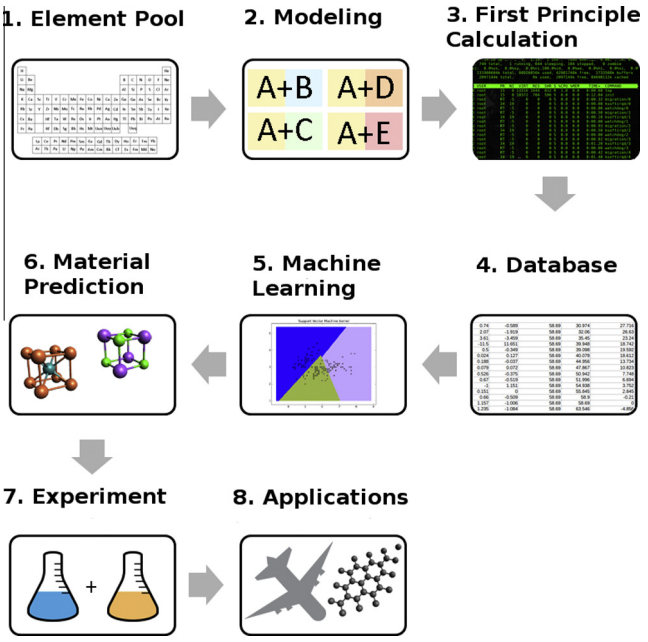

Fig. 1. Proposed work flow of material synthesis and design from first principle calculations and machine learning.

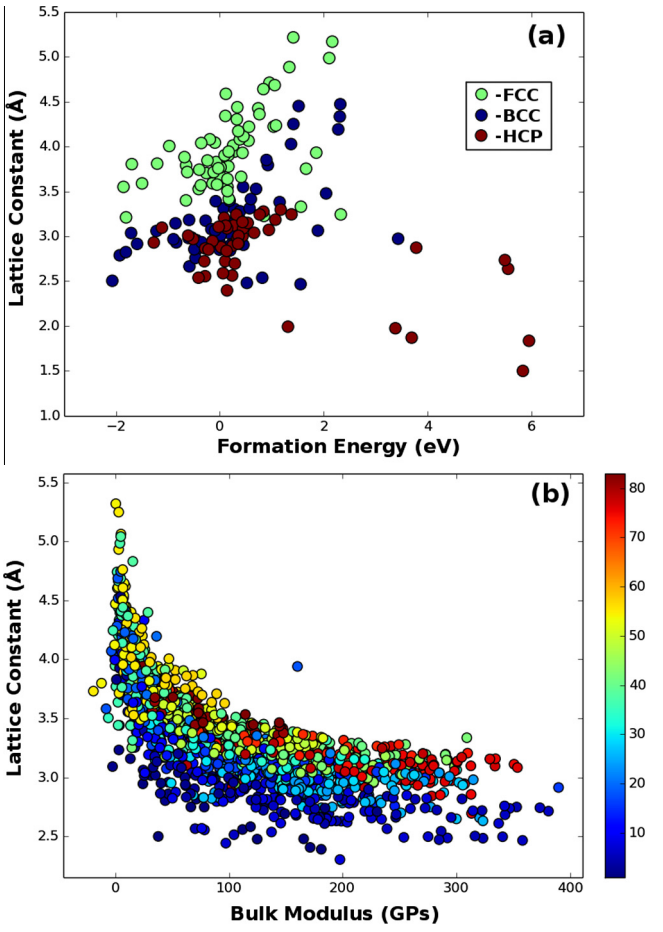

Fig. 2. (a) Lattice constant (Å) and formation of energy (eV) of Fe and all elements in the periodic table with atomic numbers within the range of 1–57 and 72–83 and with FCC, BCC, and HCP crystal structures. Note that the lattice constant of (a) is taken in the case of HCP. (b) Lattice constant (Å) and bulk modulus (GPa) of $X + Y$ where X and Y are all elements in the periodic table with atomic numbers within the range of 1–57 and 72–83 and with BCC crystal structures. The color bar indicates the atomic number of the element in Y. Please see Supporting Information for raw data used for these figures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the training data for support vector machines where various support vector machines are created depending on the material properties and material combinations. Material properties and material selection are predicted using the trained support vector machine based off the parameters put in by the inquirer. Based on the results, several predicted materials are tested and attempted to be synthesized in experimental settings. Synthesis of the desired material is then performed based on the results of the experimental investigation.

The material database in this work is created using the grid-based projector-augmented wave (GPAW) method within the density functional theory [9]. Grid spacing is set to 0.10 Å and 0.1 eV of smearing is applied. The exchange correlation of PBE exchange (PBE) and spin polarization calculations are implemented [10]. $4 \times 4 \times 4$ special k points of the Brillouin zone sampling is used within periodic boundary condition [11].

Support vector machine classification within scikit–learn is implemented for predicting the material properties and material combination [12]. Radial basis function kernel is used for kernel function and kernel coefficient is set to 0. Weight of classification is adjusted based on class frequencies and shrinking heuristic is also applied. Thus, support vector machine classification in this work is set to train the series of data containing multiple features and classes. Once the support vector machine is trained, it can predict the class based on user input.

The first database, Database 1, is created using Fe as the fixed element and the ratio of Fe and the variable element X is kept at 1:1 where X is every element of the periodic table with atomic numbers 1–57 and 72–83. The number of samples for Database 1 is 202. The calculated material properties and non-calculated features of all possible Fe + X combinations with body centered cubic (BCC), face centered cubic (FCC), and hexagonal close packed (HCP) structures are stored within the database. By organizing the data in this manner, the support vector machine can understand trends and behaviors of the materials based on their structures. In particular, Fig. 2(a) plots two features (in this case, lattice constant and formation energies) from the created database where it shows that each value is grouped according to their crystal structure. The figure demonstrates trends in behavior according to crystal structure, enabling the support vector machine to learn these trends in behavior. In the same fashion, increasing the number of features allows for a more accurate categorization within high dimensional features.

Separate support vector machines are created for each individual material property. In particular, five support vector machines are created for crystal structure, lattice constant, formation energy, magnetic moment, and bulk modulus. When creating a support vector machine, data from Database 1 is separated into training data and the target group which represents a particular material property. For instance, a support vector machine for lattice constant uses data such as element name, magnetic moment, formation energy, and bulk modulus as its training data and lattice constant is marked as the target group. The values for the target group are grouped together according to labels that are specific to each support vector machine. In the case of lattice constant, values are given a group name of 0 if the value falls between 0 and 2.5, 1 if the value is between 2.5 and 3.0, 2 if the value is between 3.0 and 3.5, 3 if the value is between 3.5 and 4.0, and 4 if the value is 4 or higher. Similarly, a support vector machine for formation energy would use the group values 0 and 1 to represent endothermic and exothermic values and organize the target group values by whether or not they are positive or negative. Support vector machi-