Computational Materials Science 150 (2018) 454–463

Contents lists available at ScienceDirect

Computational Materials Science

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

Predicting the thermodynamic stability of perovskite oxides using machine learning models

Wei Li, Ryan Jacobs, Dane Morgan*

Department of Materials Science and Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

ARTICLE INFO

Keywords:
Perovskite oxides
Thermodynamic stability
Materials discovery
Machine learning
Density Functional Theory

ABSTRACT

Perovskite materials have become ubiquitous in many technologically relevant applications, ranging from catalysts in solid oxide fuel cells to light absorbing layers in solar photovoltaics. The thermodynamic phase stability is a key parameter that broadly governs whether the material is expected to be synthesizable, and whether it may degrade under certain operating conditions. Phase stability can be calculated using Density Functional Theory (DFT), but the significant computational cost makes such calculation potentially prohibitive when screening large numbers of possible compounds. In this work, we developed machine learning models to predict the thermodynamic phase stability of perovskite oxides using a dataset of more than 1900 DFT-calculated perovskite oxide energies. The phase stability was determined using convex hull analysis, with the energy above the convex hull (E_{hull}) providing a direct measure of the stability. We generated a set of 791 features based on elemental property data to correlate with the E_{hull} value of each perovskite compound, and found through feature selection that the top 70 features were sufficient to produce the most accurate models without significant overfitting. For classification, the extra trees algorithm achieved the best prediction accuracy of 0.93 (± 0.02), with an F_1 score of 0.88 (± 0.03). For regression, leave-out 20% cross-validation tests with kernel ridge regression achieved the minimal root mean square error (RMSE) of 28.5 (± 7.5) meV/atom between cross-validation predicted E_{hull} values and DFT calculations, with the mean absolute error (MAE) in cross-validation energies of 16.7 (± 2.3) meV/atom. This error is within the range of errors in DFT formation energies relative to elemental reference states when compared to experiments and therefore may be considered sufficiently accurate to use in place of full DFT calculations. We further validated our model by predicting the stability of compounds not present in the training set and demonstrated our machine learning models are a fast and effective means of obtaining qualitatively useful guidance for a wide-range of perovskite oxide stability, potentially impacting materials design choices in a variety of technological applications.

1. Introduction

The discovery of novel functional materials is central to the continuing development of materials technologies. Recently, high-throughput DFT methods have been used to guide the discovery of new compounds for numerous applications, including: perovskite oxides for solid oxide fuel cell (SOFC) cathodes [1,2], thermochemical water splitting [3], half-heusler and sintered compounds for thermoelectrics [4,5], oxides and oxynitrides for light harvesting [6] and photoelectrochemical water-splitting [7,8], and binary metal alloys for electrocatalytic hydrogen evolution [9] and oxygen reduction [10]. While high-throughput DFT studies are valuable for discovering new functional materials, they suffer from the high computational cost required to conduct hundreds to thousands of DFT calculations.

In an effort to reduce the large amount of time required to conduct large-scale screening studies, either computational or experimental, we here apply machine learning approaches that have been demonstrated to efficiently predict many properties of materials given only relatively easily obtained structural or compositional information. Examples of properties predicted using machine learning approaches include: relative permittivity and oxygen diffusion properties of ceramic materials [11], band gap of inorganic materials [12], formation energy of elpasolite structures [13], molecular electronic properties in chemical compound space [14], density of electronic states at the Fermi energy [15], molecular atomization energies of molecules [16], Curie temperature of high-temperature piezoelectric perovskites [17], thermodynamic stability of ternary oxide compounds [18], and band gap energy of crystalline compounds and metallic glass-forming ability of ternary amorphous alloys [19]. Accurate machine learning model predictions for a material can be orders of magnitude faster than the

* Corresponding author.
E-mail address: ddmorgan@wisc.edu (D. Morgan).

https://doi.org/10.1016/j.commatsci.2018.04.033
Received 26 January 2018; Received in revised form 30 March 2018; Accepted 15 April 2018

corresponding DFT simulations or experiments, allowing them to be used to quickly understand trends in materials properties and inform materials discovery.

Of the numerous materials families investigated with high-throughput DFT methods, perovskite materials stand out as a particularly challenging class of materials for computational screening and property evaluation. When one accounts for the large number of different A- and B-site elements, as well as different typical dopant ratios and combinations, the potential number of unique perovskite compositions may be easily greater than 10^7 materials (assuming 18 possible A-site species, 31 possible B-site species, and possibly mixing up to 3 components on each site with composition restricted to increments of 0.25). This compositional flexibility of the perovskite structure enables an array of complex functional properties, including active catalysis of many reactions, ferroelectricity, piezoelectricity, superconductivity and efficient light-to-energy conversion. This flexibility also creates a significant challenge to predicting the thermodynamic stability, as stoichiometric alloying information needs to be taken into account for the different sublattices of the ABX_3 structure (where A and B are one or more cations and X is one or more anions). Recently, Schmidt, et al. reported their work on the stability prediction of ternary perovskite and anti-perovskite compounds, which used a DFT-generated dataset of about 250,000 ABX_3 compounds. The A, B, and X species were chosen from a pool of more than 60 elements ($64 \times 63 \times 62 = 249{,}984$) and a achieved mean absolute error of 121 meV/atom for regression of energy above the convex hull [20]. However, there are a large number of quaternary or quinary perovskite materials with doped elements in the A- and B- sites in an array of technologically relevant applications, so it is important to also explore the use of machine learning approaches on perovskites which have alloying on the A- and B-sites.

Recently, Jacobs, et al. used high-throughput DFT methods to screen the catalytic activity and thermodynamic phase stability of 2145 perovskite oxides for use as SOFC cathodes [2]. In general, the thermodynamic phase stability of a perovskite is a key materials property, the value of which may determine the utility of the perovskite in the given application of interest. The stability typically correlates at least loosely with whether a perovskite is synthesizable, as well as whether it may be expected to degrade (or remain stable) over time under some operational environment, such as a specific working temperature or partial pressure of oxygen [2,17]. In the work of Jacobs, et al., the stability of perovskite oxides was evaluated by using the phase diagram tools contained within the Pymatgen toolkit. The phase diagram tools in Pymatgen enable one to perform convex hull analysis, where the stability of a particular material composition (e.g. $LaFeO_3$) within a user-provided composition space (e.g. all inorganic crystalline compounds comprising the La-Fe-O system) can be performed. The main parameter governing stability is the energy above the convex hull (E_{hull}) [21]. The value of E_{hull} is a measure of the decomposition energy of the compound into a linear combination of the stable phases present on the phase diagram. Thermodynamically stable compounds exhibit an E_{hull} of zero (i.e., they are on the convex hull and are stable, equilibrium phases present on the phase diagram, at least at near zero temperature), and more positive values of E_{hull} indicate decreasing stability [22]. Based on the provided example above for $LaFeO_3$, this material is thermodynamically stable and has $E_{hull} = 0$ meV/atom. However, if one were to dope Sr on the A-site and Co on the B-site of $LaFeO_3$ to create $La_{0.375}Sr_{0.625}Co_{0.25}Fe_{0.75}O_3$ (LSCF, a well-studied commercial SOFC cathode material), then the convex hull analysis of this compound in the La-Sr-Fe-Co-O system results in $E_{hull} = 47$ meV/atom, where the energy is relative to the more stable decomposition products of $LaFeO_3$, $Sr_2Co_2O_5$, $Sr_2Fe_2O_5$, and O_2. This analysis indicates that LSCF is less stable than $LaFeO_3$, as the E_{hull} value of LSCF is larger. The pool of approximately 2145 perovskite materials calculated by Jacobs, et al. represents a very small fraction of the composition space of possible perovskite oxide compositions. Thus, data-driven methodologies based on machine learning would be beneficial to predict the stability of many additional perovskite oxide compounds.

In this work, we predict the thermodynamic phase stability of perovskite oxides using machine learning models and a subset of the perovskite stability data from Jacobs, et al. [2] of 1929 compounds (these 1929 were the subset of the 2145 compounds available at the time of writing this paper). The model can serve as a screening tool for fast discovery of potential stable compounds, significantly reducing DFT computational time and effort. We have trained several machine learning models for both classification and regression. For classification of determining stable versus unstable compounds, we found that the extra trees classifier (also known as extremely randomized trees) [23], resulted in the best classification model as determined by its calculated precision, recall and F_1 score of stable/unstable predictions. For regression of the E_{hull} values, we found the kernel ridge regression model [24] after parameter optimization gave the best regression fitting performance as determined by its calculated R^2 score and RMSE of predicted E_{hull} values. Overall, our model can predict the thermodynamic phase stability of perovskite oxide materials with uncertainties that are within typical DFT energy error bars compared to experiments.

2. Methods

The construction and validation of our machine learning models to predict perovskite stability involved five steps: (i) Generation of a feature set that can describe the thermodynamic properties of perovskite oxides. (ii) Identification of relevant features that show high correlation with stability through feature selection. (iii) Selection of the best machine learning model from the set of candidate machine learning algorithms. (iv) Examination of the model validity for different perovskite composition spaces, based on the frequency each element occurs in the training dataset. (v) Prediction of thermal stability of new perovskites outside of the dataset and comparison of the predicted values with DFT calculations. In the following sections, we detail each of the above steps needed to construct our machine learning models.

In this work, we have used the python library scikit-learn [25] for all machine learning models, feature selection methods and model evaluations. Scikit-learn is an open source machine learning package distributed under BSD license. A summary of all scikit-learn routines and function calls used in this work is provided in the Data in Brief (DiB) [26]. The training dataset of perovskite oxide compositions and DFT-calculated E_{hull} values, as well as the project source code and best models are also provided in the **DiB**.

2.1. Dataset and feature generation

The training dataset was comprised of 1929 perovskite oxide compositions from the work of Jacobs, et al. [2] These perovskite materials were simulated using DFT methods, and the stability of each compound was analyzed using the Pymatgen toolkit and all DFT-calculated materials present in the Materials Project online database as of December 2016 [22]. The E_{hull} values were obtained under environmental conditions of $T = 1073$ K, $p(O_2) = 0.2$ atm (this corresponds to an oxygen chemical potential of -6.25 eV/O, which is -1.31 eV/O relative to the O_2 molecule energy calculated in the Materials Project (material identification number mp-12957)), which represents the approximate working conditions of SOFC cathodes. Additionally, H_2 was present in the phase stability calculations via equilibrium with O_2 and H_2O gas, and a relative humidity of 30%. Additional computational details can be found in Jacobs, et al. [2] We note here that based on our choice of T and $p(O_2)$ conditions, the present model is suitable for predicting the stability of perovskites at elevated temperature at approximately room $p(O_2)$ conditions. We believe this choice of thermodynamic conditions does not overly limit the general applicability of our model in predicting perovskite stability because (1) many technological applications involving the use of perovskite oxides operate at elevated temperatures or in environments that are otherwise more reducing that standard