MACHINE LEARNING METHODS FOR PREDICT ELECTROPHYSICAL PROPERTIES OF SEMICONDUCTOR MATERIALS FOR OPTOELECTRONIC AND ENERGY STORAGE DEVICES

A. Khinevich, A. Stsiapanau, A. Smirnov
Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus

hinevics@gmail.com

I. INTRODUCTION

There were several notable attempts at utilizing Machine Learning to predict physical properties of various materials. Huang et al. reported prediction of band gap properties for ternary metal nitride compounds using ML approach based on the calculated data using Heyd–Scuseria–Ernzerhof (HSE) hybrid functionals and Perdew–Burke-Ernzerhof (PBE) DFT methods. In that study electronegativity, valence and covalent radius were used as feature for the training of the ML algorithm and prediction. In another study, high accuracy of the prediction was achieved for the ML algorithm trained on the dataset with 3 only features such as ionic radius, electronegativity and number of row associated with position of specific element in the periodic table [1].

II. DEFINING THE MACHINE LEARNING PROBLEM

Firstly, earth abundant and non-toxic ternary metal nitrides were pre-selected. This is followed by screening of computational databases for the predicted ternary metal nitride compounds suitable for solar energy harvesting with formation energy of < 0 eV/atom, energy above hull of < 0.3 eV/atom, band gap of 0.5 – 3.2 eV. Generalized gradient approximation (GGA) and PBE DFT methods that are known to underestimate the actual value of band. However, they methods are also widely used in materials databases [2]. Therefore, the compounds with band gap of over 0.5 eV were selected for the further processing.

Prior to DFT analysis, the screened nitride compounds were also sorted according to their effective masses for electrons and holes and ratio of direct and indirect band gaps available in [3] to identify materials with the most promising electronic and optical properties for further study.

The algorithm was trained using a dataset of experimentally reported band gap values of about 360 ternary metal nitrides, oxides, sulfides and phosphides compounds. A lack of data is a common problem in machine learning. This usually imposes limitations on various parts of machine learning pipeline, which we will discuss below in details.

III. FEATURE EXTRACTION AND PROCESSING

Having low amounts of data also imposes certain limits on the amount of features that can be used without causing the curse of dimensionality [4]. This suggests us to pick a number of features $d << N$, where $N=360$ is the total number of data samples available in the training split. Some models, like deep neural networks are able to extract features from the raw data fed as input. In our case of small data however, we resorted to hand picking useful features by utilizing our domain knowledge.

The previous works that used machine learning for band gap prediction have prioritized electronegativity, covalent radius and valence of each element in the compound as the key features. For example, authors in [4] have successfully used this feature combination and highlighted that those properties of an element are the most important from the physical viewpoint. A similar set of features including electronegativity, ionic radius, and row in the periodic table was found in the band gap prediction study done by [5]. Some additional physical information in the form of ionic charge, period in the periodic table, atomic number, atomic mass, van der Waals radius, and the first ionization energy was used as features for band gap prediction in [6].

Thus, our choice of a list of characteristic element-specific features used for the machine learning prediction included detailed electronic configuration, maximum valence, atomic mass, electronegativity, atomic and covalent radius, ionization potential, electron affinity, period, group and block of the constituent elements. We later processed the electronic configuration and extracted a set of features out of it, such as the number of electrons on the last orbital, the number of electrons before the last orbital, number of electrons at last level, and number of vacancies at the outer orbital.

After removing some of the redundant features and highly correlated features, we ended up with $d=51$ total features for our experiments.

IV. MODEL

Limited dataset size also affects the choice of the model and limits our choices depending on the model's complexity, since a very sophisticated model, e.g. deep neural network with many parameters, will tend to overfit when the number of training samples is too low. We used support vector regression (SVR) with

the nonlinear radial basis function (RBF) kernel as a machine learning model in our experiments. The model of choice was implemented with the use of the Scikit-learn framework and Python 3.7.

V. RESULTS

ML predictions of band gap values were used in combination with data available in open access databases to reduce the number of potential candidates for subsequent theoretical modeling with hybrid HSE functionals. In addition, the remaining compounds were sorted based on the value of effective electron and hole masses.

REFERENCES

[1] Venkatraman V., The utility of composition-based machine learning models for band gap prediction., Department of Chemistry, Norwegian University of Science and Technology, Norway, 2021, 0927-0256.

[2] Liu, Yue, et al. "Materials discovery and design using machine learning." Journal of Materiomics 3.3 (2017): 159-177.

[3] Li, Yuheng, and Kesong Yang. "High-throughput computational design of organic–inorganic hybrid halide semiconductors beyond perovskites for optoelectronics." Energy & Environmental Science 12.7 (2019): 2233-2243.

[4] Trunk, G. V. (July 1979). ”A Problem of Dimensionality: A Simple Exam-ple”. IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1(3): 306–307.

[5] L. Weston and C. Stampfl, Machine learning the band gap properties of kesteriteI2–II–IV–V4quaternary compounds for photovoltaics applications, Phys. Rev. Mater., 2018, 2(8), 085407.

[6] J. Lee, A. Seko and K. Shitara, et al., Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques, Phys. Rev. B, 2016, 93(11), 115104.