MACHINE LEARNING APPLICATIONS FOR DESIGN OF NEW MATERIALS: A REVIEW

Sr. Lecturer Kumar S. PhD., Prof. Nezhurina I. M. PhD.
College of Information Business Systems – National University of Science and Technology, MISiS
Leninsky Ave. 4, 119049, Moscow, Russian Federation

sachinagnihotri16@gmail.com

Abstract: *The importance of materials in our life is well known to the world. Material science is one of the crucial research domain which deals with planning, development, synthesis and analysis of materials and its use in the real applications. The scientists and researcher are using highly advance technologies for the production and design of new materials in present era; however the process is time consuming. Machine Learning can be defined as a set of techniques that learn from the large amount of available data and make predictions for the new data. Recently, machine learning has gained a great attraction in material science research and has been used in several research studies. Machine learning techniques have the potential to learn from the large amount of materials data and make predictions about different properties for new materials of good quality. These predictions can help in designing new materials. In this paper, we present a comprehensive review on the machine learning applications in different material science domain. We believe that the paper would be useful for the researchers, academicians and students who are involved in the design of new materials.*

Keywords: MACHINE LEARNING, MATERIAL SCIENCE, MATERIALS DESIGN, CLASSIFICATION AND PREDICTION, CLUSTERING

1. Introduction

Material science is an important research domain which is highly involved in planning, designing, synthesis and development of materials. The design of new materials is one of the hot topics at present among material scientists [1]. Scientists and researchers are working hard to design new materials with better efficiency, quality and life expectancy. Initially, the process of designing new materials involved 10-20 years from initial scratch to the production of materials. Fig. 1 illustrates the steps used in the traditional process of designing new materials. Present world is equipped with high tech equipment, world class laboratories and computationally efficient methods for the design and discovery of new materials. Hence, the process of developing new materials achieved some advancement in terms of time spent for analysis and processing of materials.

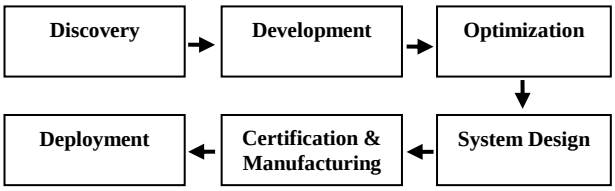

Figure 1: *Traditional steps in materials design*

The two most common approaches that have been widely used from decades in the material science research are laboratory experiments and computer simulations. In material science, laboratory experiments usually involved the study of microscopic structure of a material, analysis, synthesis and measurement of materials. All these required the use of specific instrument which is available in the laboratory only; however these processes are bound to certain temperature and pressure conditions and required a specific amount of time to complete the procedure along with expertise of the researcher involved in the study. Therefore, the importance of laboratory cannot be ignored. Computer simulation is an alternative approach in which a computer program is written to exploit the existing theoretical concepts to calculate electronics structure of the materials at macroscopic and macroscopic level. Density functional theory [2], Monte Carlo simulations [3] and Molecular dynamics [4] are the popular approaches on which computer simulations relied.

There is no doubt that computer simulations are much powerful and efficient in terms of processing time and provides full control over variables which is quite difficult to achieve with experimental analysis. However, certain challenges are still exist before computer simulation methods as well [5]. One of the issues is the dependency of microstructure of the materials. Second, these simulations are computationally expensive and require high computing clusters to run the simulations and achieve results on time which is quite expensive. Also, the previous simulations outcome does not help if a new system is introduced for study. Present demand in material science research needs a close collaboration between experimental and computational methods to provide fundamental facts about structure and properties of the materials to the scientific world. It is possible to conduct some of the experiments using computational tools due to rich technologies in present world and it has reduced the time spend on experiments from 10-20 years to 14-18 months [6]. Although, both experimental and computational methods has certain limitations that is not possible to overcome such as it is often difficult and not recommended to perform some experimental procedures over a wide temperature limit (for glass). Similarly, some the computational simulation is not possible exactly in the required form as they might be dependent on a variety of internal and external factors i.e. pressure, conformational features and structure [7].

In order to overcome these problems, several attempts have been made. MGI (Materials Genome Initiative) [8] is launched in 2011 which along with big data community provides a means to generate and store large amount of data related to material science. These initiatives made it possible for the engineers, researchers and scientists to access the properties of known materials. Several databases for materials properties are the results of these initiatives. These databases [9-11] are Inorganic Crystal Structure Database (ICSD), Super Conducting Critical Temperatures (SuperCon), Open Quantum Materials Database (OQMD) etc. Once we are equipped with huge amount of materials properties, the next task is to do some analysis to understand it and use this knowledge for decision making. The question is: How to get this information?

Machine learning [12] is a well-known powerful set of techniques that has shown its potential in different real world domains including material science. Machine learning methods enable a computer to learn from the huge amount of data and do predictions for unseen data. Recently, machine learning and big data methods have been effectively used to resolve complex issues in modeling the materials. The aim of the present study is to provide a review of the effectiveness of machine learning methods in the different domains of material science. In the next section, a brief introduction of machine learning with respect to material science is provided. In Section 3, application of machine learning in material's property prediction is provided. In Section 4, we discussed the machine learning applicability in molecular dynamics aspect of

material science. Finally, the most relevant conclusions are provided in conclusion section.

2. Machine Learning in Material Science

Fig. 2 illustrates the general idea of the knowledge discovery process in material science using machine learning. It shows that materials properties data can be generated through laboratory experiments and computer simulations. Next, data preprocessing methods should be performed to select the relevant materials properties and cleaning the data. Further, machine learning can be applied to develop a prediction model. This model can be used to predict the desired properties for unseen data of new materials properties.

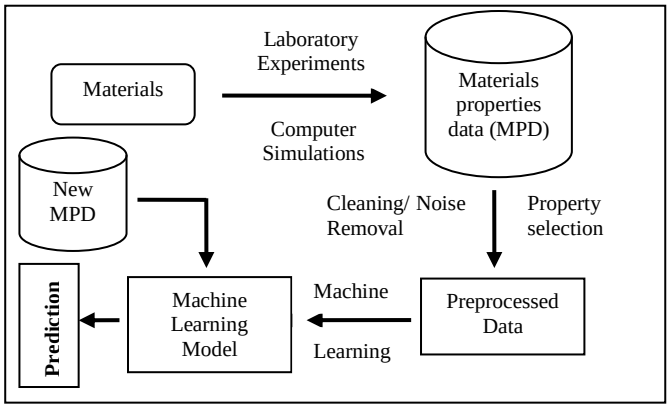

Figure 2: *Knowledge Discovery Process using Machine Learning in Material Science*

The crucial steps in knowledge discovery process for materials data analysis are data preparation and selection of suitable machine learning algorithm for model development. The properties of materials for the analysis must be efficiently done because not all properties are useful for specific analysis. Hence, the selection must be accurate. The next important task is to select the machine learning algorithm. This selection is based on the nature of the task to be performed and the features in the data set. Therefore, this has to be done carefully.

Machine learning has already shown its potential in material science both in terms of prediction accuracy and time [13]. In material science, machine learning can be used for property prediction (both macroscopic and microscopic level), discovery of new materials and force field prediction at atomic level for materials.

3. Materials Property Prediction using Machine Learning

Computer simulation and laboratory experiments are the two common approaches to study the materials properties. Melting point of a material, glass transition temperature, ionic conductivity, lattice constant and molecular atomization energy are some of the properties of materials which can be described both on macroscopic and microscopic levels. Although computational simulation and laboratory experiments are efficient enough to study the materials properties but if the results are not accurate as desirable then the time and resources spent in computation and laboratory experiments are wasted [14]. Moreover, in some cases it is very difficult and almost impossible to study the properties even with complex computer simulations and extensive efforts in laboratory experiments. Therefore, it is highly desirable to develop a system to understand the properties of materials with high efficiency and low computational cost.

Machine learning is a constructive way that enables the computer to learn from the huge amount of materials properties and create a linear or non-linear mapping (as applicable) between materials properties and related factors. Further, this model can be used for prediction of properties for new materials. The problem of materials property prediction can be divided in to macroscopic and microscopic level. At the macroscopic level, specifically we are interested in mechanical and physical properties of materials and its relationship with microstructure of materials [15]. Machine learning algorithms such as Support Vector Machine (SVM), Regression and Artificial Neural Network (ANN) with optimization algorithms is widely used for the study and analysis of macroscopic properties of materials. One of the popular versions of ANN is back propagation ANN that has been used to capture the relationship between materials properties and external factors [16]. Previously, few attempts have been made to predict the materials properties such as temperature responses, elongation, corrosion etc. using back propagation ANN [16-17]. Although back propagation ANN is an effective method to be used for property prediction in material science, slow convergence rate and local minima is a problem of concern.

Another version of ANN without such problems of local minima and low convergence rate is RBF-ANN (Redial Basis Function-ANN). A successful attempt with RBF-ANN has been made to investigate the crack propagation has been made in a study on bituminous layered pavement structure [18]. In another study [19], prediction of melting point of a material is investigated using ANN. There are other studies as well which tried to develop prediction models using ANN for excited state energy prediction [20], diffusion barriers [21] etc. Another aspect of ANN that must be taken into consideration is that ANN provides better results if the size of the data is relatively large enough. For relatively small samples, ANN results may not be reliable. Support vector machine (SVM) is another machine learning methods for data analysis and pattern recognition in the data. With a relatively small sample of data, SVM can be used for reliable results as it can efficiently handle large dimensions and over fitting problem. An approach based on Genetic algorithm (GA) and (Support Vector Regression) SVR has been proposed [22] with an objective to forecast the corrosion behaviors of zinc and steel. SVM has shown its potential in several domains and can be deployed for ionic conductivities prediction, glass transition temperature prediction [23-24].

On the other hand, microscopic properties of a material such as electronics crystal structure, lattice constant, band energy, molecular atomization energy etc. are also one of the important aspects in design of new materials. Logistic Regression (LR), SVR and ANN have been deployed for prediction of lattice constant in various studies [25-28]. In these studies, it was found that ANN has a better accuracy and performance than LR method; however the accuracy for training data was better than the accuracy on test data. Further, SVR found to achieve better accuracy and performance than ANN for lattice constant prediction. Also, SVR has the better learning and testing efficiency that ANN for the smaller datasets. In order to improve the accuracy of machine learning techniques, Ward et al. [29] used the concept of ensemble learning methods. They suggested that ensemble learning can overcome the shortcomings of machine learning algorithms and can provide improved efficiency and accuracy.

Therefore, it is evident that machine learning algorithms has played a big role in providing more accurate materials property predictions at both macroscopic and microscopic levels. However, not all machine learning methods are suitable in all cases. If the size of data sample of materials properties is not larger enough, the SVR performs better than ANN. On the other hand, if the size of the sample is high than ANN and RBF-ANN can be used to develop more effective and reliable prediction models.

4. Atomic Force Field Prediction using Machine Learning

Quantum based mechanical approaches also known as Ab-initio methods or first principle quantum calculations are very demanding at present in materials modeling. Based on first principles quantum mechanical simulations, a benchmark data set can be created which

consists of positions of the atoms (atomic configurations) and the corresponding forces working on them. This benchmark data set can be utilized to investigate the relationship between atoms positions and the potential energy using machine learning methods [30]. Many active research groups in the field have faith in machine learning that it can lead a pathway to create interatomic potentials with lower computational cost than quantum mechanical methods, with better accuracy and versatility than conventional interatomic potentials and provides better efficiency [31].

Zhenwei et al. [32] presented a machine learning approach along with first principle quantum calculations to perform molecular dynamic simulations. They mentioned that their proposed scheme is more accurate and takes lesser quantum mechanical calls as required in repeatedly encountered new chemical process. In recent studies, it is realized that there is a vectorial force experienced by every atom in the system, which can be learned from the given positions of the atom and can be predicted further for new positions [33, 34]. This approach is more appealing as the conventional methods tend to predict the potential energy of the system which is a global component of the systems, whereas atomic force on an atom is a local component which can be determined by the surrounding environment of the system.

Furthermore, Gaussian models, neural networks have already shown their potentials in the development of machine learning based models for molecular dynamic simulations [32-36]. Huan et al. [33] proposed a universal strategy to generate machine learning based atomistic force field in order to perform more accurate and powerful molecular dynamic simulations. They proposed to first create a dataset with low noise which can be obtained from density functional theory methods, second to create fingerprints by numerically representing the atomic configurations and corresponding forces. Next, select the optimal and non-redundant datasets from reference data for training purpose. Further, different prediction algorithms can be used to learn from the data and compute atomic forces directly from atomic configurations.

Assessment of Intra and inter-molecular energies can be done with quantum mechanical simulations which depend on the accuracy and preciseness of atomic force field. Friederich et al. [37] mentioned that non-local association between dihedral potentials which has been ignored in many state of the art dihedral force fields, has an important role to play in the description of total molecular energy. They used ANN to study the degrees of freedom of atoms within molecules of complex organic materials. Their results were quite promising in stating that multilayer ANN can provide better predictive results than the conventional approaches.

Therefore, it is worth saying that machine learning is being proved an important approach which has potential to develop fore field models which can perform better than conventional computational simulations.

4. Machine Learning: Is there any complexities?

There is no doubt that machine learning potentials in material science came up with several advantages, however certain complexities associated with machine learning potentials must be taken into consideration. One of the crucial factors is the preparation of the data that is the basic need for a machine learning task. So, most of the efforts must be done with preparation phase as if the data is not well prepared the machine learning will no longer be able to provide the reliable results. Furthermore, in material science expert knowledge is very important in preparation of the data i.e. selection of appropriate properties of materials that can be utilized in property prediction task. In atomic force field creation, fingerprinting is very important task in which atomic configurations and forces must be numerically presented for better analysis. Therefore, expert knowledge is required for effective data preparation.

Once the preparation of data is done effectively then the next important step is to select the efficient machine learning algorithms to develop prediction model. As shown in earlier sections, ANN is a reliable approach for materials property prediction but SVR performs better when the size of the data is relatively small [22, 23]. Furthermore, if the size of the data is relatively large that it is a better option to with ANN and its other version such as RBF-ANN. Multi-layered ANN also known as deep neural networks has been used recently in some studies related to material science. Mills et al. [38] used convolutional deep neural network to predict the ground state energy of electrons in two dimensional electrostatic potentials and the accuracy obtained are within chemical boundaries. In a study, Ryczko et al. [39] proposed an approach to compute the total potential energy for the atomic systems. They used deep convolutional neural networks and named this method as CNNAS (convolutional neural network for atomic systems). They mentioned that their method is able to predict energies for different types of 2D hexagonal lattices with a great accuracy. Therefore, if you have a large database of materials properties or DFT calculations, you may choose the deep neural networks as it is able to achieve a higher level of accuracy.

5. Conclusion

This study attempts to provide an understanding about the machine learning applicability in the materials science to the students, researchers and academicians engaged in the design of novel materials. The study took into consideration the traditional approaches such as experimental analysis in the laboratory and the computer simulation on powerful computers for the study of materials. Next, the issues associated with these conventional approaches such as their limitations and the computational cost etc. is discussed that required a new method to be developed to overcome these limitations. Machine learning is already being used by material scientists and engineers to study and investigate the materials. Different machine learning methods such as Genetic algorithms, Gaussian model, SVM, SVR, ANN and deep ANN has already been deployed to develop prediction models for different materials. Although, machine learning is contributing well in design and discovery of new materials; yet more extensive research is needed to fully understand its applicability in design of new materials with better quality.

Acknowledgement

The authors gratefully acknowledge the financial support of the Ministry of Education and Science of the Russian Federation in the framework of Increase Competitiveness Program of **NUST «MISiS» (№ К4-2017-052).**

References

1. About J., Materials genome initiative for global competitiveness, National Science and Technology Council, USA, 2011.
2. Hohenberg P., W. Kohn, Inhomogeneous electron gas, Phys Rev, 136, 1964, 864-871
3. Binder K., A. Baumgartner, The Monte Carlo method in condensed matter physics, 71, 1995.
4. Alder B. J., T. E. Wainwright, Studies in molecular dynamics. I. General method, J Chem Phys, 31 (2), 1959, 459-466.
5. Rajan K., Materials informatics, Mater TodCay, 8, 2005, pp. 38-45.
6. Hautier G., A. Jain, S. P. Ong, From the computer to the laboratory: materials discovery and design using first-principles calculations, J mater. Sci, 47, 2012, 7317-7340.

7. Yu X. L., B. Yi, X. Y. Wang, Prediction of the glass transition temperature for polymers with artificial neural network, J Theor Comput Chem, 7, 2008, 953-963.
8. Materials Genome Initiative: https://www.mgi.gov accessed on 30.07.2018.
9. Belsky A., M. Hellenbrandt, V. L. Karen, P. Luksch, New developments in the inorganic crystal structure database (ICSD): accessibility in support of materials research and design, Acta Crystallogr, Sect B Struct Sci, 58, 2002, 364-369
10. Super Conducting Critical Temperatures (SuperCon): https://supercon.nims.go.jp accessed on 30.07.18
11. Kirkli S., J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies, npj Comput Mater, 1, 2015, 15010.
12. Mueller T., A.G. Kusne, R. Ramprasad, Machine learning in materials science: recent progress and emerging applications, Rev Comput Chem, 29, 2016, 186.
13. Yue L., Z. Tianlu, J. Wangwei, S. Siqi, Materials discovery and design using machine learning, J Materiomics, 3, 2017, 159-177.
14. Ning X., M. Walters, G. Karypisxy, Improved machine learning models for predicting selective compounds, J Chem Inf Model, 52, 2012, 38-50.
15. Bertinetto C., C. Duce, A. Micheli, R. Solaro, A. Starita, M. R. Tine, Evaluation of hierarchical structured representations for QSPR studies of small molecules and polymers by recursive neural networks, J Mol Graph Model, 27, 2009, 797-802.
16. Guo Z., S. Malinov, W. Sha, Modelling beta transus temperature of titanium alloys using artificial neural network, Comput Mater. Sci, 32, 2005, 1-12.
17. Topçu B., M. Sarıdemir, Prediction of properties of waste AAC aggregate concrete using artificial neural network, Comput Mater Sci, 41, 2007, 117-125.
18. Altun F., Özgür Kişi, K. Aydin, Predicting the compressive strength of steel fiber added lightweight concrete using neural network, Comput Mater. Sci, 42, 2008, 259-265.
19. Gajewski J., T. Sadowski, Sensitivity analysis of crack propagation in pavement bituminous layered structures using a hybrid system integrating artificial neural networks and finite element method, Comput Mater Sci, 82, 2014, 114-117.
20. Häse F. , S. Valleau, E. Pyzer-Knapp, A. Aspuru-Guzik, Machine learning exciton dynamics, Chem Sci, 7 (8), 2016, 5139-5147.
21. Wu H., A. Lorenson, B. Anderson, L. Witteman, H. T. Wu, B. Meredig et al., Robust FCC solute diffusion predictions from ab-initio machine learning methods, Comput Mater Sci, 134, 2017, 160-165.
22. Fang S. F., M.P. Wang, W.H. Qi, F. Zheng, Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials, Comput Mater Sci, 44, 2008, 647-655.
23. Liu X., W.C. Lu, C.R. Peng, Q. Su, J. Guo, Two semi-empirical approaches for the prediction of oxide ionic conductivities in ABO_3 perovskites, Comput Mater Sci, 46, 2009, 860-868.
24. Ahmad A., A. Amjad, M. Denny, C. A. S. Bergström, Experimental and computational prediction of glass transition temperature of drugs, J Chem Inf Model, 54, 2014, 3396-3403.
25. Javed S. G., A. Khan, A. Majid, A. M. Mirza, J. Bashir, Lattice constant prediction of orthorhombic ABO_3perovskites using support vector machines, Comput Mater. Sci, 39, 2007, 627-634.
26. Li C. H., Thing YH, Zeng YZ, Wang CM, Wu P. Prediction of lattice constant in perovskites of $GdFeO_3$ structure. J Phys Chem Solids, 64, 2003, 2147-2156.
27. Majid A., A. Khan, T. Choi, Predicting lattice constant of complex cubic perovskites using computational intelligence, Comput Mater Sci, 50, 2011, 1879-1888
28. Majid A., A. Khan, G. Javed, A.M. Mirza, Lattice constant prediction of cubic and monoclinic perovskites using neural networks and support vector regression, Comput Mater Sci, 50, 2010, 363-372.
29. Ward L., A. Agrawal, A. Choudhary, C.Wolverton, A general purpose machine learning framework for predicting properties of inorganic materials, npj Comput Mater, 2, 2016, 16028.
30. Hastie T. R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed.; Springer: New York, 2009.
31. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. J. Chem. Phys. 2011, 134, 074106.
32. Li Z., J. R. Kermode, A. De Vita, Molecular Dynamics with On the-Fly Machine Learning of Quantum-Mechanical Forces. Phys. Rev. Lett. 2015, 114, 096405.
33. Huan T. D., R. Batra, J. Chapman, S. Krishnan, L. Chen & R. Ramprasad, A Universal strategy for the creation of machine learning-based atomistic force fields, npj Computational Materials 3, 37, 2017.
34. Botu V. R. Ramprasad, Learning scheme to predict atomic forces and accelerate materials simulations. Phys. Rev. B: Condens. Matter Mater. Phys. 92, 2015, 094306.
35. Behler, J., M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys. Rev. Lett. 98, 2007, 146401.
36. Bartók A. P., G. Csányi, Gaussian approximation potentials: a brief tutorial introduction. Int. J. Quant. Chem. 115, 2015, 1051–1057.
37. Friederich P., M. Konrad, T. Strunk, W. Wenzel, Machine learning of correlated dihederal potentials for atomistic molecular force fields, Scientific Reports 8, 2018, 2559.
38. Mills K., M. Spanner, I. Tamblyn, Deep learning and the Schrodinger equation, Phys. Rev. A, 96, 042113
39. Ryczko K., K. Mills, I. Luchak, C. Homenick, I. Tamblyn, Convolutional neural networks for atomistic systems, Comput Mater Sci, 149, 2018, 134-142.