Indonesian Journal of Data and Science

Volume 5 Issue 3
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v5i3.177

Research Article

Probabilistic Graphical Models for Predicting Properties of New Materials Based on Their Composition and Structure

Vusumuzi Malele [1,*], Ashley Phala [2]

[1] North West University, South Africa, vusi.malele@nwu.ac.za
[2] North West University, South Africa, mohlagashanephala@gmail.com

Correspondence should be addressed to Vusumuzi Malele; vusi.malele@nwu.ac.za

Received 17 October 2024; Accepted 20 December 2024; Published 31 December 2024

Abstract:

Probabilistic graphical model (PGMs) offer a powerful framework for modeling complex relationships between different components. By integrating information on the element composition and structural features, these models enable the inference of materials properties with a probabilistic perspective. This approach holds promising efforts towards accelerating materials discovery design, as it facilitates the predication of diverse materials characteristics, ranging from electronic and mechanical properties to thermal and optical behavior. The use of PGMs in materials science represents a sophisticated methodology for harnessing data-driven insights to guide the exploration of innovative materials with tailored functionalities. The purpose of this paper is to investigate literature for the exploitation of the data science concepts, big data and machine learning that yields computational intelligence. A literature review approach to understand the exploitation and use of computational intelligence in the leading-edge research and innovation of materials science. The findings illustrate that machine learning can be used to intricate chemical problems that otherwise would not be tractable. Leveraging PGMs presents a promising avenue for predicting the properties of new materials based on their composition and structure.

Keywords: Machine Learning, Probabilistic Graphical Models, Big Data, Research on Discovering New Properties of Materials
Dataset link: -

1. Introduction

In materials science, the quest for discovering novel materials with desired properties has been a persistent challenge. Conventional methods that have been used over the years to discover new materials, such as empirical trial and error and the density functional theory(DFT)-based methods are incapable of keeping up with the fast pace of the development of new materials today because of their long development cycles, low efficiency, and high cost [1].

Machine learning has become a topic of interest in many industrial fields such as material science and chemistry, used for the prediction of materials properties, accelerate simulations, design new structures, for the prediction synthesis routes of new materials [2].

PGMs are one of the fast-growing classes of machine learning models [3]. They are commonly used in chemistry and materials science due to their direct work on graph or structural representations of molecules and materials as such have full access to relevant information required to characterize materials [4]. Materials scientists are continuously striving to improve their ability to understand, predict and improve materials properties [5]. Over the past years materials scientists have relied highly on simulation and modeling methods to understand and predict materials properties. Due to the imitations of the traditional trial-and-error method in materials research, scientists have come up with new methods to predict and improve materials properties more affordable and less time consuming.

Recently, data-driven methods such as machine learning have been adopted for the prediction of new materials which lead to a paradigm shift in the way materials science research is conducted [6].

Machine learning and data science have become an important part of natura science, considered as the fourth pillar in science, next to the experiment, theory, and simulations. Machine learning techniques are increasingly applied in all steps of the materials development cycle, from the database screening, finding initial candidate materials property predictions, for material designs, for the prediction of the synthesis conditions and automated experimental data analysis as well as experimental planning [7]. Machine learning techniques applied in material science ranges from a wide range of classical machine learning algorithms such as decision trees, convolutional neural networks, and probabilistic graphical models.

PGMs have garnered significant attention for their potential to predict the properties of emerging materials based on their intrinsic composition and structural characteristics. PGMs represent a rapidly evolving class of machine learning models particularly well-suited for applications in chemistry and materials science [8]. Their inherent ability to operate directly on graph or structural representations of molecules and materials affords them comprehensive access to pertinent information crucial for characterizing materials. By leveraging the complex relationships between different material components, PGMs enable probabilistic inference of material properties, offering a nuanced understanding of diverse characteristics spanning electronic and mechanical properties to thermal and optical behaviour [9], [10].

The graphical models have been applied to a large number of fields, including bioinformatics, social science, control theory, image processing, marketing analysis, among others [11], [12]. However, structure learning for graphical models remains an open challenge since one must cope with a combinatorial search over the space of all possible structures.

The quest for discovering and designing novel materials with specific and desirable properties has been a longstanding challenge in the field of materials science. Traditionally, the process of discovering novel materials relied heavily on empirical trail and error methods, which were guided by intuition and a limited understanding of the underlying principles governing materials behaviour [13]. However, as the materials science has evolved, the demand for efficient and systematic approaches to accelerate the discovery and design of new materials has become increasingly pronounced. In the past decades, density functional theory (DFT) and other theoretical methods played a very important role in predicting materials propertied by simulating the behaviour of atoms and electrons within materials. While these methods provided valuable insights, they also had a lot of limitations such as being computationally expensive, limiting their application to relatively small-scale and impeding their ability to keep pace with the rapidly advancing landscape of materials development [14], [15].

With the advent of machine learning (ML) techniques, a paradigm shift occurred in materials science. ML, described by its ability to discover patterns and relationships within large datasets, presents a promising avenue for predicting material properties, accelerating simulations, and guiding the design for new materials [16]. Researchers started leveraging ML algorithms to discover hidden correlations between materials composition, structure, and properties allowing for more informed and efficient exploration of the vast materials space.

Within the domain of ML, PGMs emerged as a sophisticated and versatile tool for predicting materials properties [17]. PGMs provide a framework for representing and reasoning about uncertain information, which makes them well-suited for capturing complex relationships within materials. PGMs directly work with graph or structural representations of molecules and materials, PGMs offer a unique advantage in extracting relevant information important for material characterization [18].

The use of PGMs in material sciences science has gained traction due to their ability to integrate different sources of information, consider uncertainties, and facilitate probabilistic inference of materials properties. This approach enables a nuanced understanding of the intricate relationship between different material components, allowing the prediction of a broad spectrum of properties, including electronic, mechanical, and thermal [19].

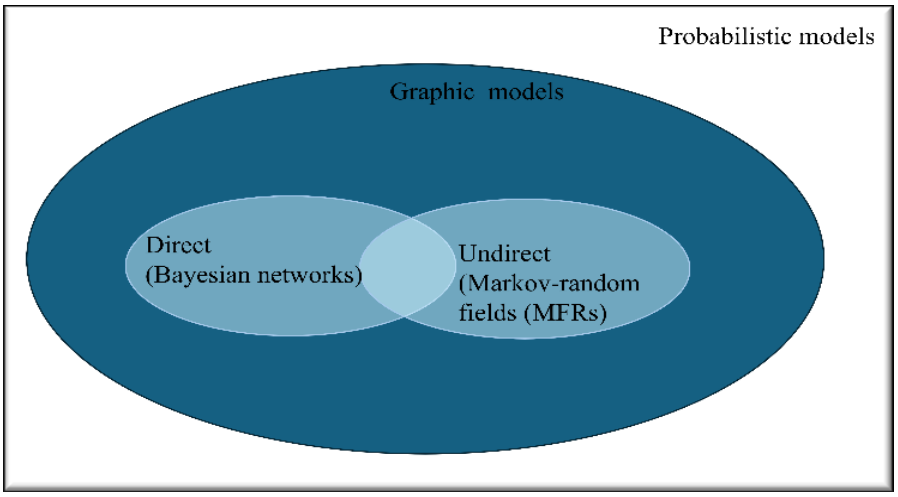

Figure 1. Flavours of probabilistic graphical models

Figure 1 illustrates the Probabilistic graphical models as a model with two sub-categories: (i). Direct and Undirected graphic models. The Direct graphic models are described as Bayesian networks also known as belief networks. They are probabilistic graphical models that represent a set of variables and their conditional dependencies via a direct acrylic graph (DAG). In a Bayesian network, each node represents a random variable, and the directed edges between the nodes encode probabilistic dependencies, indicating the casual relationships between variables.

The Undirected graphic models are described as Markov networks, represents a set of variables with pairwise Markovian dependencies via an undirected graph. In a Markov random field (MRF), each node represents a random variable, and the absence of direct edges implies that the relationship between the variables are undirected and typically encode the notion of local interactions or spatial proximity [20], [21].

In this paper, we present a comprehensive survey of the existing structure learning algorithms. This research aims to harness the inherent probabilistic nature of PGMs to decipher intricate relationships between material compositions, structural attributes, and resulting properties. By integrating diverse datasets encompassing material compositions, structural configurations, and corresponding properties, this study seeks to construct and refine PGMs capable of accurately predicting material characteristics. Such predictive models hold the potential to accelerate the design and discovery of innovative materials. Despite this introduction, this paper comprises four other sections: method, data collection, recommendations, and conclusion.

2. Method:

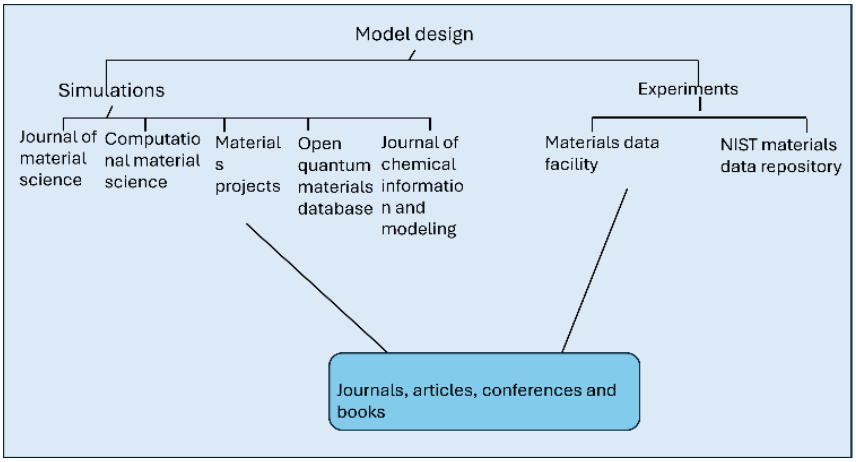

Figure 2. Different sources used to conduct this work.

Figure 2 illustrates the number sources which were consulted. The search criteria used to find out information was "PGMs for prediction of materials properties" AND "machine learning for prediction of materials properties", from

Materials project, Open Quantum materials database, journals of materials science, computational materials science, journals of chemical information and modelling and ACS applied materials and interfaces. This includes conferences, journal, early access articles and magazines were selected used in this research.

A lot of research has been done in using data-driven and machine learning for the prediction of materials properties, but very limited research has been done in the use of PGMs to predict or design novel materials for specific functionalities. The need for the discovery of novel material properties for specific functionalities is a challenge experienced by many researchers in different sectors such as energy, computer and communication, aerospace, and medicine. Researcher from different backgrounds should work together to maximize their potential, to improve the short comes of predicting new properties of materials, efficiently and less cost effective. The sources consulted and used to conduct this work were Materials project, Open Quantum materials database, journals of materials science, computational materials science, journals of chemical information and modelling and ACS applied materials and interfaces. These sources combined a comprehensive, expertly curated abstract and citation database. It has enriched data and scholarly literature from a wide range of disciplines. It provides more diverse range of results.

3. Results and Discussion

Researchers in materials science, chemistry and engineering usually use data-driven models and machine learning for the prediction of materials properties. No research documents results were found for PGMs for use of prediction of materials properties but over a million results were found for machine learning and data driven models for prediction of materials properties. The most commonly used machine learning algorithms for the prediction of materials properties are random forest, Support Vector Machines, Convolutional Neural Networks, Gradient Boosting Machines and Gaussian Process Regression.

A lot of literature has been published on the probabilistic graphical models and their impact in the materials science field, but little research has been done to use these models to actually predict any materials properties. There is more research that needs to be done in exploring PGMs for use in prediction of materials properties. In order to advance the utilization of probabilistic graphical models (PGMs) in predicting properties of new materials, several recommendations emerge from our analysis. First and foremost, the promotion of data sharing and collaboration among researchers is paramount. Open-access repositories and collaborative platforms can facilitate the exchange of materials data and foster interdisciplinary collaboration within the materials science community. Additionally, investing in the development of user-friendly software tools and platforms for building, training, and deploying PGMs is essential to empower researchers to effectively harness the capabilities of these models. Moreover, providing training and educational resources can equip researchers with the requisite knowledge and skills to leverage PGMs optimally in materials science research. Standardized protocols and benchmarks for validating the performance of PGMs in predicting material properties should be established to ensure rigor and comparability across studies. Collaboration between computational and experimental researchers to integrate PGMs with advanced experimental techniques can enhance the reliability and accuracy of material property predictions. Furthermore, ethical considerations, long-term funding and support, and engagement with stakeholders from academia, industry, government, and the broader community are crucial factors to consider in advancing the field. By embracing these recommendations, the materials science community can propel the use of PGMs towards innovative discoveries and advancements in materials design and engineering.

4. Conclusion

The exploration of PGMs for predicting properties of new materials based on their composition and structure represents a promising frontier in material science research. From Bayesian networks to Markov random fields, PGMs offer a versatile framework for modelling the dependencies between material components, providing insight into the underlying mechanisms governing material behaviour. By integrating information on element composition, structure motifs, and environmental factors, PGMs enable the prediction of diverse materials characteristics spanning electronic, mechanical, thermal, and optical domains. It is evident that the application of PGMs holds immerse promise for accelerating materials discovery and design. By leveraging data-driven insight and computational intelligence, researchers can explore innovative materials with tailored functionalities, driving advancement across various industrial sectors and scientific disciplines. Looking ahead, the future of PGMs in materials science lies in their continued refinement and application to emerging challenges and opportunities.

Future research directions in the realm of PGMs for predicting new properties of materials offer a promising avenue for exploration. One of the avenues includes the development of dynamic graphical models capable of

capturing temporal dependencies and changes in material properties over time, particularly relevant for material subjected to varying environmental conditions or undergoing phase transition. Hierarchical modelling approach could be investigated to capture the multiscale nature of material behaviour, from atomic and molecular scale to macroscale properties, providing insight into the complex material systems. Integrating uncertainty quantification techniques into PGMs would enable researchers to quantify and incorporate uncertainty into model predictions, enhancing the robustness and reliability of materials property predictions. The future research direction holds the potential to advance the predictions capabilities of PGMs and drive innovation in materials science and engineering.

References:

[1] Wei, J., Chu, X., Sun, X.Y., Xu, K., Deng, H.X., Chen, J., Wei, Z. and Lei, M., 2019. Machine learning in materials science. InfoMat, 1(3), pp.338-358, doi: 10.1002/inf2.12028.

[2] Wang, A.Y.T., Murdock, R.J., Kauwe, S.K., Oliynyk, A.O., Gurlo, A., Brgoch, J., Persson, K.A. and Sparks, T.D., 2020. Machine learning for materials scientists: an introductory guide toward best practices. Chemistry of Materials, 32(12), pp.4954-4965, doi: 10.1021/acs.chemmater.0c01907.

[3] Mueller, T., Kusne, A.G. and Ramprasad, R., 2016. Machine learning in materials science: Recent progress and emerging applications. Reviews in computational chemistry, 29, pp.186-273, doi: 10.1002/9781119148739.ch4.

[4] Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O. and Walsh, A., 2018. Machine learning for molecular and materials science. Nature, 559 (7715), pp.547-555, doi: 10.1038/s41586-018-0337-2.

[5] Morgan, D. and Jacobs, R., 2020. Opportunities and challenges for machine learning in materials science. Annual Review of Materials Research, 50, pp.71-103, doi: 10.1146/annurev-matsci-070218-010015.

[6] El-Shafie, M., 2023. Hydrogen production by water electrolysis technologies: A review. Results in Engineering. Natural gas, 240, p.48, doi: 10.1016/j.rineng.2023.101426.

[7] Koller, D. and Friedman, N., 2009. Probabilistic graphical models: principles and techniques. MIT press.

[8] Reiser, P., Neubert, M., Eberhard, A., Torresi, L., Zhou, C., Shao, C., Metni, H., van Hoesel, C., Schopmans, H., Sommer, T. and Friederich, P., 2022. Graph neural networks for materials science and chemistry. Communications Materials, 3(1), p.93, doi: 10.1038/s43246-022-00315-6.

[9] Bhadeshia, H.K.D.H. 1999. Neural networks in materials science. ISIJ international, 39(10), pp.966-979, doi: 10.2355/isijinternational.39.966.

[10] Ferguson, A.L., 2017. Machine learning and data science in soft materials engineering. Journal of Physics: Condensed Matter, 30(4), p.043002, doi: 10.1088/1361-648x/aa98bd.

[11] Sucar, L.E., 2015. Probabilistic graphical models. Advances in Computer Vision and Pattern Recognition. London: Springer London. doi, 10(978), p.1, doi: 10.1007/978-3-030-61943-5.

[12] Pernkopf, F., Peharz, R. and Tschiatschek, S., 2014. Introduction to probabilistic graphical models. In Academic Press Library in Signal Processing (Vol. 1, pp. 989-1064). Elsevier, 10.1007/978-1-4615-1539-5_2.

[13] Martin, D. Schmidt, M.W. and Hillerbrand, R. Implementing AI Ethics in the Design of AI-assisted Rescue Robots", 2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS), West Lafayette, IN, USA, i2023, pp. 1-1, doi: 10.1109/ETHICS57328.2023.10155062.

[14] Bagayoko, D., 2014. Understanding density functional theory (DFT) and completing it in practice. AIP Advances, 4(12), doi: 10.1063/1.4903408

[15] Burke, K. and Wagner, L.O., 2013. DFT in a nutshell. International Journal of Quantum Chemistry, 113(2), pp.96-101, doi: 10.1002/qua.24259.

[16] Jordan, M.I., 2003. An introduction to probabilistic graphical models.

[17] Larrañaga, P., Karshenas, H., Bielza, C. and Santana, R., 2012. A review on probabilistic graphical models in evolutionary computation. Journal of Heuristics, 18, pp.795-819, doi: 10.1007/s10732-012-9208-4.

[18] Frey, B.J. and Jojic, N., 2005. A comparison of algorithms for inference and learning in probabilistic graphical models. IEEE Transactions on pattern analysis and machine intelligence, 27(9), pp.1392-1416, doi: 10.1109/TPAMI.2005.169.

[19] Ankan, A. and Panda, A., 2015. Probabilistic graphical models using python. In Proceedings of the 14th python in science conference (scipy 2015) (Vol. 10). Citeseer, doi: 10.25080/Majora-7b98e3ed-001.

[20] Bernardo, J.M. and Smith, A.F., 2009. Bayesian theory (Vol. 405). John Wiley & Sons.

[22] Li, Stan Z. "Markov random field models in computer vision." In Computer Vision—ECCV'94: Third European Conference on Computer Vision Stockholm, Sweden, May 2–6 1994 Proceedings, Volume II 3, pp. 361-370. Springer Berlin Heidelberg, 1994, doi: 10.1007/978-4-431-66933-3.