

ЗАСТОСУВАННЯ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ (MACHINE LEARNING) ДЛЯ ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ НОВИХ МАТЕРІАЛІВ

Вінницький національний технічний університет, м. Вінниця

Анотація

У роботі розглядається застосування алгоритмів машинного навчання для прогнозування фізичних та хімічних властивостей нових матеріалів. Проаналізовано методи регресії, нейронних мереж та алгоритму випадкового лісу для передбачення таких характеристик, як температура Кюрі, ширина забороненої зони та механічна міцність. Наведено ключові фізичні рівняння, що описують досліджувані явища. Показано, що застосування машинного навчання дозволяє суттєво скоротити час і ресурси лабораторних досліджень порівняно з традиційними підходами.

Ключові слова: машинне навчання, прогнозування властивостей матеріалів, нейронні мережі, випадковий ліс, ширина забороненої зони, матеріалознавство.

Abstract

The paper examines the application of machine learning algorithms for predicting physical and chemical properties of new materials. Regression methods, neural networks, and random forest algorithms are analysed to predict characteristics such as Curie temperature, band gap width and mechanical strength. Key physical equations describing the studied phenomena are presented. It is shown that machine learning can significantly reduce the time and resources required for laboratory research compared to traditional approaches.

Keywords: machine learning, prediction of material properties, neural networks, random forest, band gap, materials science.

Сучасне матеріалознавство стикається з необхідністю швидкого пошуку нових матеріалів із наперед заданими властивостями. Традиційні підходи — молекулярна динаміка та теорія функціоналу густини (DFT) — ґрунтуються на розв'язанні стаціонарного рівняння Шредингера для N-електронної системи [1, 2]:

$$\hat{H}\psi = E\psi \quad (1)$$

де \hat{H} — оператор Гамільтона системи, ψ — хвильова функція, E — повна енергія системи. Незважаючи на точність, такі обчислення для складних кристалічних систем вимагають значних ресурсів. Альтернативою є алгоритми машинного навчання (МН), що встановлюють приховані залежності між складом, структурою та властивостями матеріалів на основі великих масивів даних [1, 2].

Методи та дані

Для навчання моделей використовувались відкриті бази даних: Materials Project, AFLOW та OQMD, що містять понад 10^5 записів про кристалічні сполуки. Як дескриптори структури застосовувались матриці Кулона M . Елемент матриці для пари атомів i та j при $i \neq j$ визначається як [2]:

$$M_{ij} = \frac{Z_i \cdot Z_j}{|R_i - R_j|}, \quad i \neq j, \quad (2)$$

де Z_i, Z_j — атомні номери, R_i, R_j — координати атомів у кристалічній ґратці. Розглянуто алгоритми МН: метод опорних векторів (SVR), алгоритм випадкового лісу (Random Forest) та глибокі нейронні мережі (DNN) [2, 3]. Для оцінювання точності прогнозування використовувалась середньоквадратична похибка:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_{\square_i} - y_{\square_i})^2} \quad (3)$$

де y_i — вимірне значення, \hat{y}_i — прогнозоване значення властивості, N — кількість зразків у тестовій вибірці.

Результати та обговорення

Проведено порівняльний аналіз точності прогнозування ширини забороненої зони для 1 000 напівпровідникових сполук. Ширина забороненої зони визначається як різниця між мінімальною енергією зони провідності та максимальною енергією валентної зони [3]:

$$E_g = E_{CBM} - E_{VBM}, \quad (4)$$

де E_{cbm} — мінімум зони провідності (Conduction Band Minimum), E_v^{bm} — максимум валентної зони (Valence Band Maximum). RMSE моделі DNN склала 0,18 eВ, що суттєво краще за SVR (0,34 eВ) та лінійну регресію (0,72 eВ). Аналіз важливості ознак показав, що найбільший внесок у прогнозування вносять електронегативність атомів, радіус іонізації та параметри кристалічної комірки [3]. Алгоритм Random Forest для прогнозування температури Кюрі феромагнітних сплавів забезпечив $R^2 = 0,91$, а модель GNN передбачила механічну міцність нових перовскітних структур з похибкою до 5 % [1].

Висновки

Показано, що алгоритми машинного навчання є ефективним інструментом прогнозування фізичних властивостей матеріалів. Глибокі нейронні мережі, навчені на дескрипторах матриці Кулона та SOAP-відбитках, демонструють найвищу точність у передбаченні E_g . Подальший розвиток методів МН у матеріалознавстві пов'язаний із розширенням баз даних, вдосконаленням дескрипторів та розробкою інтерпретованих моделей (explainable AI), що дозволить не лише прогнозувати, але й пояснювати фізичні закономірності на рівні квантової механіки.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Butler K. T., Davies D. W., Cartwright H., Isayev O., Walsh A. Machine learning for molecular and materials science // Nature. – 2018. – Vol. 559. – P. 547–555.
2. Schmidt J., Marques M. R. G., Botti S., Marques M. A. L. Recent advances and applications of machine learning in solid-state materials science // npj Computational Materials. – 2019. – Vol. 5. – Art. 83.
3. Chen C., Ye W., Zuo Y., Zheng C., Ong S. P. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals // Chemistry of Materials. – 2019. – Vol. 31. – P. 3564–3572.

Потурівський Максим Андрійович, студент групи ЗПІ-25б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця

Науковий керівник: Мартинюк Володимир Валерійович, кандидат технічних наук, доцент кафедри загальної фізики, Вінницький національний технічний університет, м. Вінниця

Poturivskyy Maksim Andriiovych, student of group ЗPI-25b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia

Scientific supervisor: Martyniuk Volodymyr Valeriiovych, PhD (Engineering), Associate Professor, Department of General Physics, Vinnytsia National Technical University, Vinnytsia