

# 2D Materials

## Постановка задачи

Двумерные дихалькогениды переходных металлов (TMDCs) - это относительно новые типы материалов, которые обладают различными свойствами, начиная от полупроводниковых, металлических, магнитных, сверхпроводящих и заканчивая оптическими. Химический состав TMDCs -  $\text{MX}_2$ ; где М - группа переходных металлов, наиболее популярные из которых молибден и вольфрам, а Х - обычно сера или селен. Молекулярная структура TMDCs может содержать различные дефекты, оказывающие влияние на свойства материала. Задача состоит в обучении модели, способной предсказывать свойства материалов, в зависимости от дефектов структуры.

## Актуальность задачи

В области квантовых материалов развитие новых экспериментальных и вычислительных методов машинного обучения увеличило объем и скорость сбора данных. Искусственный интеллект способен повлиять на исследования новых материалов, таких как сверхпроводники, спиновые жидкости и топологические изоляторы. Использование этих материалов необходимо для развития квантовых вычислений[1], устройств связи и высокоскоростных электронных схем[2].

## Данные, планируемые для использования

Датасет из примерно 3000 TMDCs(двумерные дихалькогениды переходных металлов):

- Структура
- Энергия
- Энергетический спектр

## Планируемые результаты

- Цель: проектирование материалов с заданными свойствами
- Прямая задача: предсказание свойств по структуре.
- Обратная задача: создание структуры по заданным свойствам.

## Методы

- Прямая задача:
  - Свёрточные нейронные сети на графах
  - графовая нейронная сеть MEGNet
  - Set transformer
  - Expert features
- Обратная задача:
  - Байесовская оптимизация
  - Приближенное вычисление градиента

## Литература

1. *Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Das Sarma, S. NonAbelian anyons and topological quantum computation. Rev. Mod. Phys. 80, 1083–1159 (2008).*
2. *Ponomarenko, L. A. et al. Chaotic dirac billiard in graphene quantum dots. Science 320, 356 (2008).*
3. *Stanev, V., Choudhary, K., Kusne, A. G., Paglione, J., & Takeuchi, I. Artificial intelligence for search and discovery of quantum materials. Communications Materials, 2(1), 1-11 (2021).*

## Литературный обзор задачи

- В статье [1] была построена модель машинного обучения, предсказывающая топологию материала на основе его состава. В качестве данных использовался датасет из более чем 37000 материалов, среди которых было около 6 тысяч топологических изоляторов и 14000 полуметаллов. Модель – градиентный бустинг, построенный на деревьях. Для оценки результата использовались такие метрики как Accuracy и F-score. В результате довольно простая модель дала неплохой (Accuracy около 90%) результат. Однако данная модель не может гарантировать, что материал на самом деле будет иметь топологические особенности.
- В работе [2] случайные деревья использовались для предсказания MCE (magnetocaloric effect). Входными данными являлись свойства и количества атомов, содержащиеся в химической формуле, полученные из базы данных MagneticMaterials, а также экспериментальные значения. В качестве модели использовалась реализация градиентного бустинга из библиотеки XGBoost. В результате, с помощью машинного обучения был получен ферромагнетик, демонстрирующий сильный MCE.
- В статье [3] использовалась модель на основе нейронных сетей для предсказания кристаллической структуры материалов. Входными данными являлись химические формулы, решётки Браве, кристаллографические группы и другие параметры решётки. В качестве источника данных использовалась база данных ICSD. Авторы статьи объединили несколько многослойных перцептронов, и получили модель, способную предсказывать различные особенности кристаллической решетки. Качество оценивалось через accuracy. В результате, был разработан инструмент CRYSPNet, превосходящий по точности тривиальные предсказательные стратегии. Тем не менее, результат работы инструмента сильно зависит от тренировочной и валидационной выборки, показывая наилучшие результаты на выборках со сплавами металлов и с высоко симметричными кристаллическими структурами.
- В работе [4] авторы применяют сверточные нейронные сети для предсказания топологии материалов по данным рентгеновской абсорбционной спектроскопии (XAS). Данные XAS были получены с использованием базы данных вычисленных спектров рентгеновского поглощения вблизи краевой структуры, распространяемой в рамках Material Project [5-8]. Модель состояла из

нескольких сверточных слоев, также применялся dropout. Основные используемые метрики - precision, recall, F-score. Сравнив эффективность кластеризации без учителя с классификатором нейронной сети, авторы пришли к выводу, что результаты обеих моделей сильно зависят от предсказываемого элемента. Таким образом, XAS с помощью машинного обучения может стать простым, но мощным экспериментальным инструментом для топологической классификации.

Проанализировав статьи, мы приходим к выводу, что методы машинного обучения являются крайне перспективными при разработке новых квантовых материалов. Имеющиеся модели уже показывают неплохую точность, однако остаются не исследованными более сложные модели, например сверточные сети на графах.

## Литература

1. Claussen, N., Bernevig, B. A., & Regnault, N. (2020). *Detection of topological materials with machine learning*. Physical Review B, 101(24), 245117.
2. de Castro, P. B., Terashima, K., Yamamoto, T. D., Hou, Z., Iwasaki, S., Matsumoto, R., ... & Takano, Y. (2020). *Machine-learning-guided discovery of the gigantic magnetocaloric effect in HoB 2 near the hydrogen liquefaction temperature*. NPG Asia Materials, 12(1), 1-7.
3. Liang, H., Stanev, V., Kusne, A. G., & Takeuchi, I. (2020). *CRYSPNet: Crystal structure predictions via neural networks*. Physical Review Materials, 4(12), 123802.
4. Andrejevic, N., Andrejevic, J., Rycroft, C. H., & Li, M. (2020). *Machine learning spectral indicators of topology*. arXiv preprint arXiv:2003.00994.
5. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. a. Persson, "The Materials Project: A materials genome approach to accelerating materials innovation," APL Materials, vol. 1, no. 1, p. 011002, 2013.
6. C. Zheng, K. Mathew, C. Chen, Y. Chen, H. Tang, A. Dozier, J. J. Kas, F. D. Vila, J. J. Rehr, L. F. J. Piper, K. A. Persson, and S. P. Ong, "Automated generation and ensemble-learned matching of X-ray absorption spectra," npj Computational Materials, vol. 4, p. 12, 03 2018.
7. S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, "Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis," Computational Materials Science, vol. 68, pp. 314–319, Feb. 2013.
8. S. P. Ong, S. Cholia, A. Jain, M. Brafman, D. Gunter, G. Ceder, and K. A. Persson, "The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles," Computational Materials Science, vol. 97, pp. 209–215, feb 2015.

