

Отчет

Агрелов И.Н., Саган В.С.

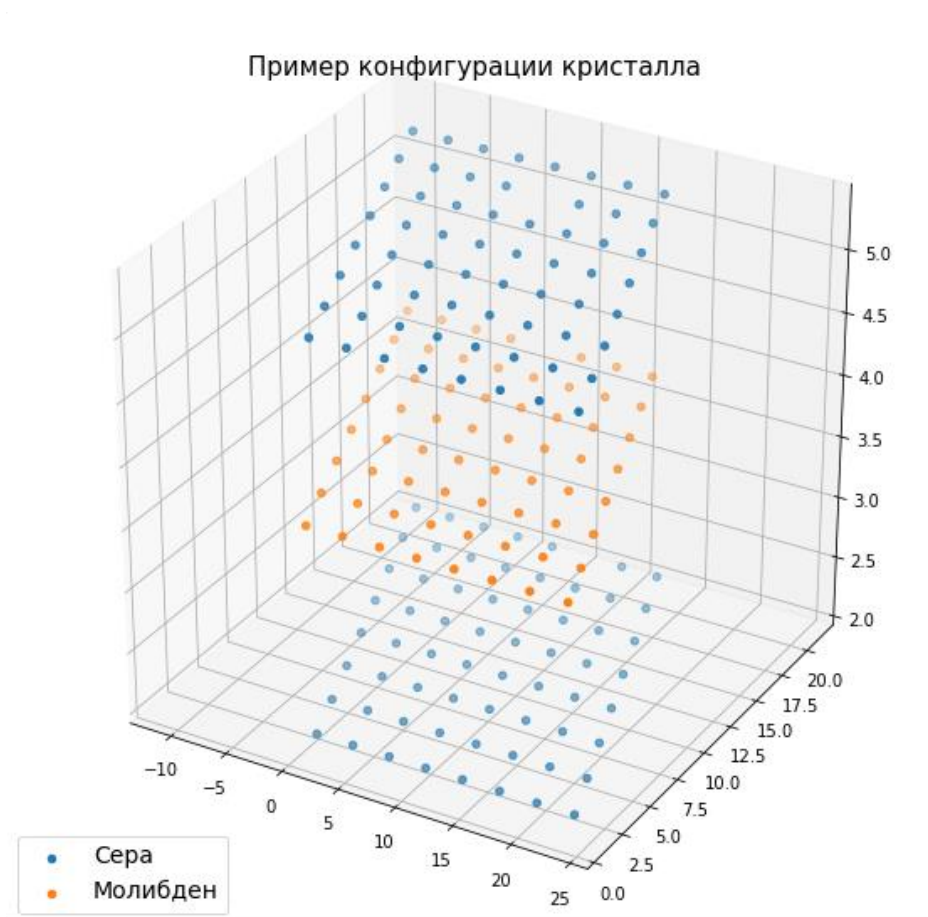
Постановка задачи

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 - MX_2 ; где М - группа переходных металлов, наиболее популярные из которых молибден и вольфрам, а Х - обычно сера или селен.

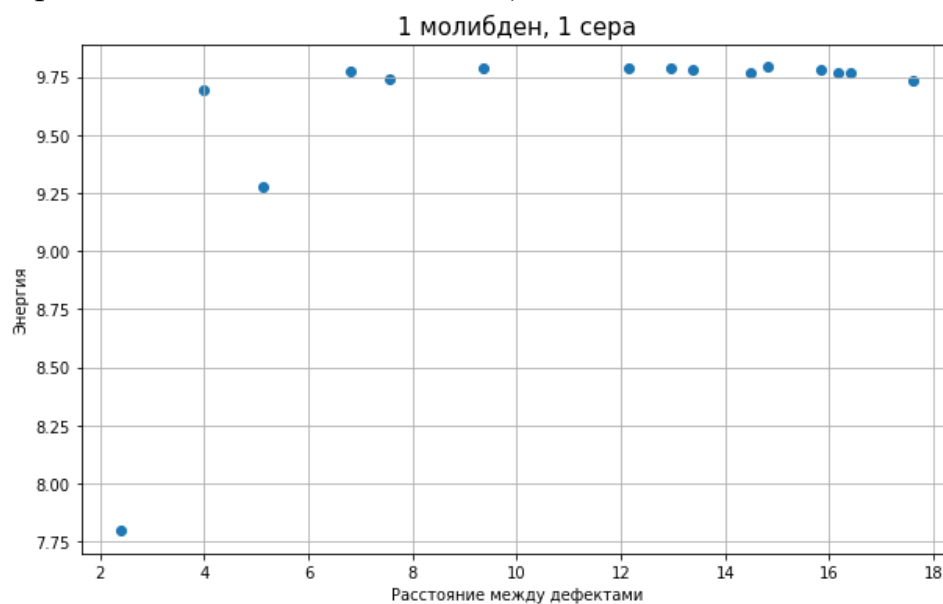
Молекулярная структура TMDCs может содержать различные дефекты, оказывающие влияние на свойства материала. Задача состоит в обучении модели, способной предсказывать свойства материалов, в зависимости от дефектов структуры.

Результаты

- Изучена библиотека rumatgen для хранения информации о структуре и свойствах кристаллов. Построен график, визуализирующий структуру изучаемого кристалла:

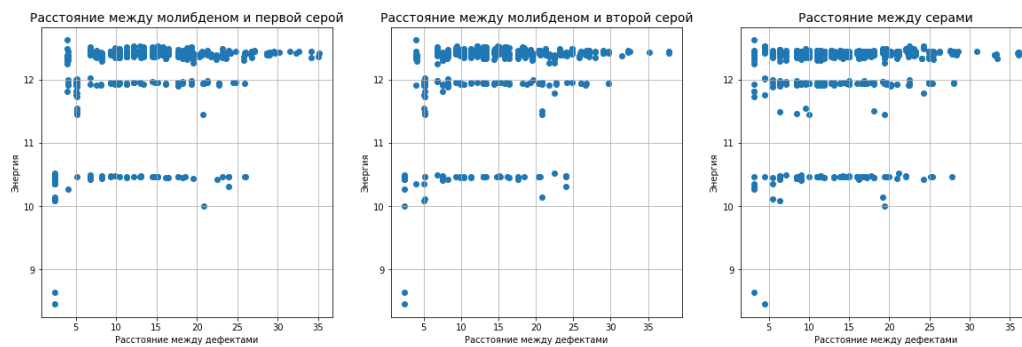


Построен график зависимости энергия(formation energy) от расстояния между дефектами (отсутствие в решетке одного атома серы и одного атома молибдена):



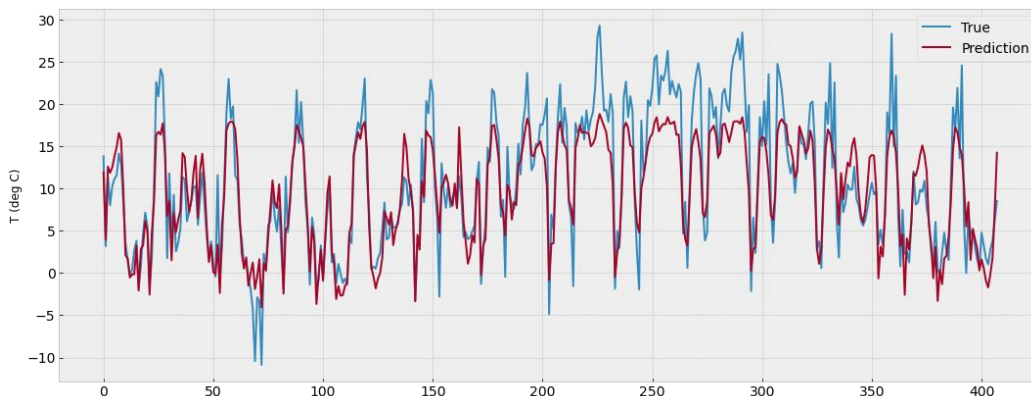
Построены также графики для случаев отсутствия двух атомов серы

и одного атома молибдена:

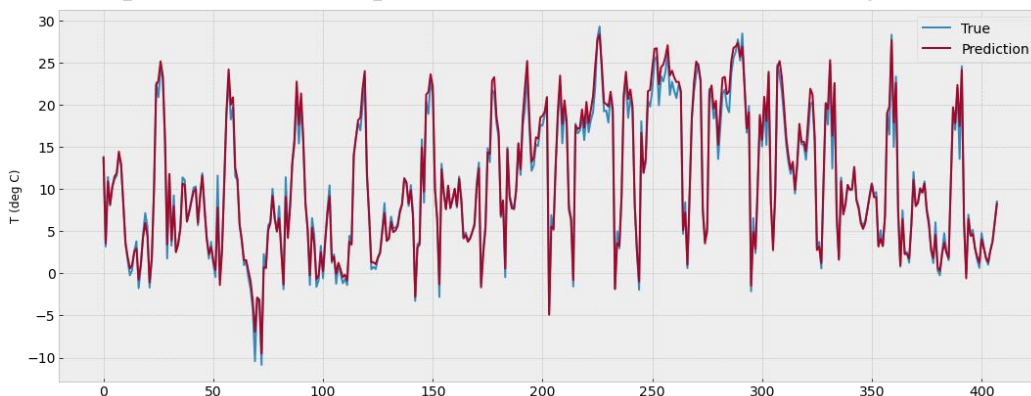


- Реализована базовая модель на основе случайного леса для сравнения с более совершенными моделями. Модель по заданным расстояниям между дефектами в кристалле позволяет предсказать энергию(formation energy). Модель показала точность 0.05(метрика MSE) для случая отсутствия в решетке кристалла двух атомов серы и одного атома молибдена.
- Выполнена работа по изучению модели глубокого обучения Трансформер. Сделано задание для закрепления знаний о Трансформерах, Positional Encoding.

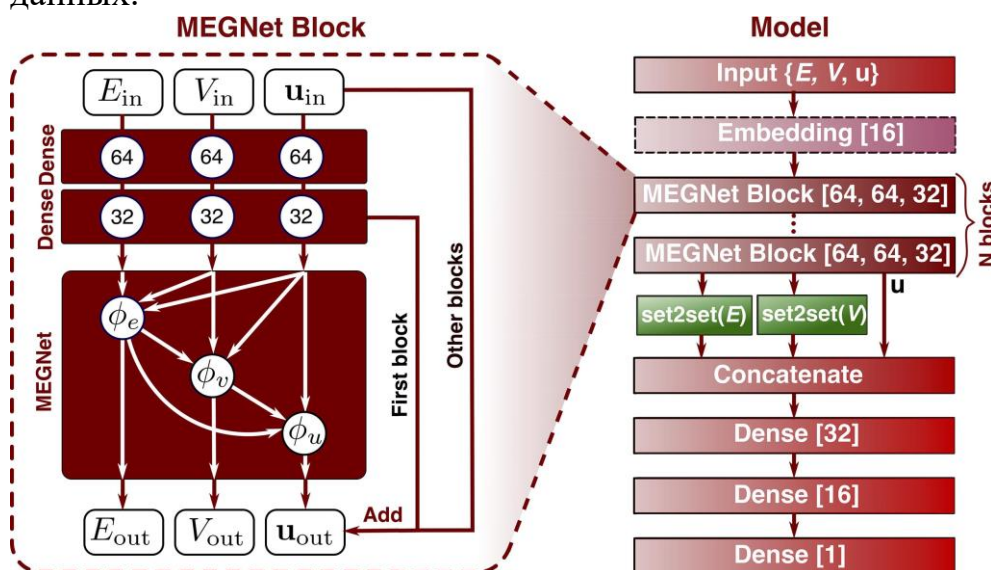
Предсказания без Positional Encoding:



Предсказания с применением Positional Encoding:



- Изучена архитектура MEGNet, произведен запуск на тестовых данных.



Каждому модулю графовой сети предшествует два мультислойных перцептрона, составляя блок MEGNet. Несколько блоков MEGNet могут быть объединены, позволяя информации быть переданной на большее расстояние. Число необходимых блоков зависит от числа параметров необходимых для предсказания целевой переменной. На последнем шаге используется set2set для того, чтобы отобразить вывод в скаляр/вектор.

Ссылки

<https://github.com/IlyaAgrelov/reprotech-2D-crystals-transformer> — наш репозиторий

https://github.com/RomanovIgnat/MEGNet_PyTorch — дружественная реализация MEGNet

<https://github.com/HSE-LAMBDA/IDAO-2022> — контекст с данными о структуре кристаллов

<https://maze-forsythia-33b.notion.site/Topic-2-5-Sequences-attention-mechanism-5fbbed7f71c34cfabccfc13b4a02ba84> — статья о трансформерах

<https://distill.pub/2021/gnn-intro/> — статья о графовых сетях

<https://arxiv.org/pdf/1812.05055.pdf> — статья про архитектуру MEGNet

<https://arxiv.org/pdf/0812.4970.pdf> — статья по физике процесса