

# Добыча знаний из кристаллических структур: определение степени окисления атомов металлов

М.И. Смольков, А.Ф. Крутов

Поволжский государственный университет телекоммуникаций и информатики, Самара, Россия

**Обоснование.** В настоящее время машинное обучение приобрело огромную популярность благодаря способности находить решения сложных и нетривиальных задач. С ростом объема учитываемых данных качество найденных решений, как правило, улучшается. Кристаллохимия не является исключением, в настоящее время изучено строение более миллиона структур, которые хранятся в кристаллографических базах данных [1, 2]. Используя накопленные данные, авторы [3] с помощью машинного обучения создали подход к автоматической классификации структур по симметрии кристаллов. А в статье [4] с помощью нейронной сети проанализировали шаблоны строения известных кристаллических структур с целью прогнозирования вероятности образования новых соединений. В статье [5] предлагают способ определения степени окисления металлов в металлорганических каркасах.

**Цель** — создание и реализация сервиса для определения степени окисления атом металлов в координационных и ионных соединениях с O окружениями.

**Методы.** Для создания модели машинного обучения необходимо: выбрать набор данных для обучения, отобрать дескрипторы структуры, разработать классификационную модель. Далее описывается процесс создания модели обучения на примере кислородного окружения.

Было отобрано 4 набора данных, составленных из баз CSD [1] и ICSD [2]. Всего — 33,253 соединения со степенью окисления от 1 до 7 (рис. 1).

Данные соответствовали следующим критериям: положительные и отрицательные заряды сбалансированы; структура полностью определена; отсутствуют статистически неупорядоченные атомы, отсутствуют дубликаты, тестовые и тренировочные данные не пересекаются.

Для предсказания степени окисления атомов металлов были отобраны следующие дескрипторы соединений, дающие лучший результат предсказания, — оптимальный набор признаков:

- $N_{at}$  — порядковый номер металла в Периодической системе;
- $Group$  — номер группы металла в Периодической системе;
- $CN$  — координационное число;
- $G_3$  — второй момент инерции полиэдра Вороного;
- $R_{SD}$  — радиус сферического домена, объем которого равен объему полиэдра Вороного (Å);
- $\Sigma\Omega(X)$  — доля телесных углов граней, которые образованы соседними атомами  $X$  определенного химического сорта (O) в полиэдре Вороного атома.

$CN$ ,  $G_3$ ,  $R_{SD}$  и  $\Sigma\Omega(X)$  были рассчитаны с помощью пакета ToposPro [6]. В качестве классификатора выбран алгоритм Random Forest из библиотеки Scikit-Learn [7], который на оптимальном наборе признаков дает лучший результат.

**Результаты.** Был проведен расчет гиперпараметров для классификатора, и в результате была создана модель предсказания, дающая лучшую оценку предсказательной способности на оптимальном наборе признаков (табл. 1).

**Таблица 1.** Оценка качества предсказания на наборах данных модели машинного обучения, основанной на алгоритме Random Forest, с использованием лучшего набора признаков и оптимальных значений гиперпараметров

| На чем обучено         | На чем тестировалась | Качество, % | Точность, % | Полнота, % | F1-мера, % |
|------------------------|----------------------|-------------|-------------|------------|------------|
| Тренировочный набор I  | Тестовый набор I     | 98,7        | 98,4        | 96,4       | 97,3       |
|                        | Тестовый набор II    | 95,9        | 91,8        | 89,1       | 90,3       |
| Тренировочный набор II | Тестовый набор I     | 99,0        | 98,5        | 99,1       | 98,8       |
|                        | Тестовый набор II    | 98,1        | 94,9        | 95,2       | 95,0       |

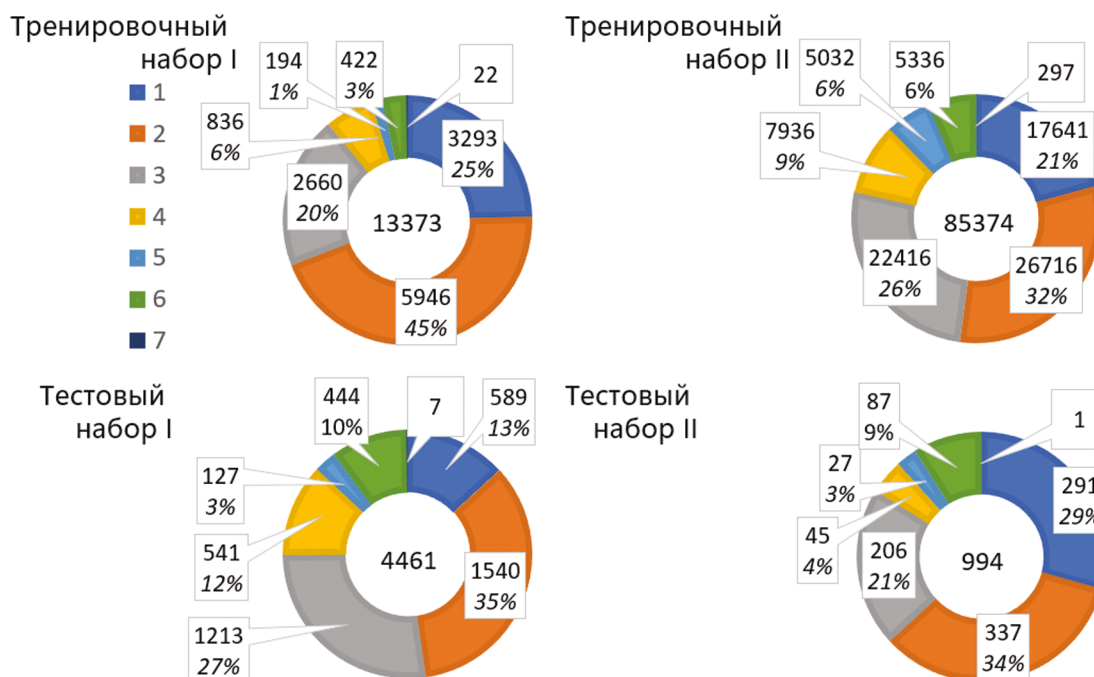


Рис. 1. Статистические сведения о составленных для обучения и тестирования наборах данных

ICSD Code: 113572  
Mn<sub>2</sub>F<sub>5</sub>:

Из первоисточника:  
Mn(1) - +3  
Mn(2) - +2

Источник:  
Bandemehr, Jascha et al.  
Syntheses and  
Characterization of the  
Mixed-Valent  
Manganese(II/III) Fluorides  
Mn<sub>2</sub>F<sub>5</sub> and Mn<sub>3</sub>F<sub>8</sub>. Inorganic  
Chemistry, 2021.

Прогнозирование с  
помощью сервиса:

Atom Mn (1)  
Coordinates (0.00000, 0.00000, 0.00000)  
ΣΩ<sub>0</sub> = 0.000 < 0.98 → using «Non Oxygen» classification model.

| Oxidation State | Prediction |
|-----------------|------------|
| 0               | 6.0%       |
| 1               | 5.6%       |
| 2               | 17.6%      |
| 3               | 44.4%      |
| 4               | 26.4%      |

Atom Mn (2)  
Coordinates (0.00000, 0.46580, 0.25000)  
ΣΩ<sub>0</sub> = 0.000 < 0.98 → using «Non Oxygen» classification model.

| Oxidation State | Prediction |
|-----------------|------------|
| 0               | 2.8%       |
| 1               | 2.3%       |
| 2               | 56.9%      |
| 3               | 35.2%      |
| 4               | 2.8%       |

ICSD Code: 113664  
Mn<sub>2</sub>O<sub>3</sub>:

Из первоисточника:  
Mn(1) - +4  
Mn(2) - +2

Источник:  
Sergey V. Ovsyannikov, et  
al.: Synthesis of Ilmenite-  
type ε-Mn<sub>2</sub>O<sub>3</sub> and Its  
Properties.  
Inorganic Chemistry, 2021.

Прогнозирование с  
помощью сервиса:

Atom Mn (1)  
Coordinates (0.66667, 0.33333, 0.67408)  
ΣΩ<sub>0</sub> = 0.997 > 0.98 → using «Oxygen» classification model.

| Oxidation State | Prediction |
|-----------------|------------|
| 3               | 13.7%      |
| 4               | 86.3%      |

Atom Mn (2)  
Coordinates (0.33333, 0.66667, 0.52976)  
ΣΩ<sub>0</sub> = 0.976 < 0.98 → using «Non Oxygen» classification model.

| Oxidation State | Prediction |
|-----------------|------------|
| 2               | 98.6%      |
| 3               | 1.4%       |

Рис. 2. Результаты предсказания степеней окисления атомов Mn в структуре Mn<sub>2</sub>O<sub>3</sub> [9] с помощью веб-сервиса Crystal Predictor [8]

**Выводы.** Данная модель легла в основу сервиса Crystal Predictor [8]. Вычисления для структуры Mn<sub>2</sub>O<sub>3</sub> [9], не входящей в использованные наборы данных, демонстрируют способность системы адекватно предсказывать степени окисления атомов металлов (рис. 2). На настоящее время сервисом обработано более 648 уникальных структур, загруженных пользователями из всего мира.

**Ключевые слова:** кристаллография; кристаллическая структура; металлы; степень окисления; машинное обучение; тестирование и оценка.

## Список литературы

1. ccdc.cam.ac.uk [Электронный ресурс]. Cambridge Structural Database (CSD). Доступ по: <https://www.ccdc.cam.ac.uk/solutions/software/csd/>
2. icds.products [Электронный ресурс]. Inorganic Crystal Structure Database (ICSD). Доступ по: <https://icsd.products.fiz-karlsruhe.de/>
3. Ziletti A., Kumar D., Scheffler M., Ghiringhelli L.M. Insightful classification of crystal structures using deep learning // Nat Commun. 2018. Vol. 9. ID. 2775. DOI: 10.1038/s41467-018-05169-6
4. Ryan K., Lengyel J., Shatruk M. Crystal structure prediction via deep learning // J Am Chem Soc. 2018. Vol. 140, No. 32. P. 10158–10168. DOI: 10.1021/jacs.8b03913
5. Jablonka K.M., Ongari D., Moosavi S.M., Smit B. Using collective knowledge to assign oxidation states of metal cations in metal-organic frameworks // Nat Chem. 2021. Vol. 13. P. 771–777. DOI: 10.1038/s41557-021-00717-y
6. Blatov V.A., Shevchenko A.P., Prosperio D.M. Applied topological analysis of crystal structures with the program package ToposPro // Cryst Growth Des. 2014. Vol. 14, No. 7. P. 3576–3586. DOI: 10.1021/cg500498k
7. Pedregosa F., Varoquaux G., Gramfort A., et al. Scikit-learn: Machine learning in Python // JMLR. 2011. Vol. 12, No. 85. P. 2825–2830.
8. crystalpredictor.com [Электронный ресурс]. Crystal Predictor. Доступ по: <https://crystalpredictor.com/>
9. Bandemehr J., Zimmerhofer F., Ivlev S.I., et al. Syntheses and characterization of the mixed-valent manganese (II/III) fluorides  $Mn_2F_5$  and  $Mn_3F_8$  // Inorg Chem. 2021. Vol. 60, No. 17. P. 12651–12663. DOI: 10.1021/acs.inorgchem.1c01833

### *Сведения об авторах:*

**Михаил Игоревич Смольков** — аспирант 3-го года обучения, кафедры высшей математики; Поволжский государственный университет телекоммуникаций и информатики, Самара, Россия. E-mail: [m.smolkov97@gmail.com](mailto:m.smolkov97@gmail.com)

**Александр Федорович Крутов** — научный руководитель, доктор физико-математических наук, профессор; Поволжский государственный университет телекоммуникаций и информатики, Самара, Россия. E-mail: [a\\_krutov@rambler.ru](mailto:a_krutov@rambler.ru)