ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2023, том 514, № 2, с. 355–363

УДК 546

МАШИННОЕ ОБУЧЕНИЕ КАК ИНСТРУМЕНТ УСКОРЕНИЯ ПОИСКА НОВЫХ МАТЕРИАЛОВ ДЛЯ МЕТАЛЛ-ИОННЫХ АККУМУЛЯТОРОВ

© 2023 г. В. Т. Осипов^{1,*}, М. И. Гонгола^{2,3}, Е. А. Морхова¹,

член-корреспондент РАН А. П. Немудрый², А. А. Кабанов^{1,2}

Представлено академиком РАН А.И. Аветисяном Поступило 04.08.2023 г. После доработки 11.08.2023 г. Принято к публикации 24.10.2023 г.

Поиск новых кристаллических ионных проводников является важной задачей материаловедения, которая требует значительных ресурсов, но может быть ускорена с помощью методов машинного обучения (MO). В данной работе методы MO были применены для прогнозирования энергии миграции рабочих ионов. Обучающая выборка основана на данных о 225 каналах миграции ионов лития в 23 ионных проводниках. Дескрипторами выступали параметры свободного пространства в кристалле, полученные методом разбиения Вороного. Точность прогнозирования энергии миграции оценивалась путем сравнения с данными, полученными методами теории функционала плотности. В работе было применено два метода MO: регрессия методом опорных векторов и порядковая регрессия. Показано, что параметры свободного пространства в кристалле коррелируют с энергией миграции, при этом лучшие результаты дает порядковая регрессия. Разработанные модели MO могут применяться как дополнительный фильтр при анализе ионной проводимости в структурах.

Ключевые слова: ионные проводники, разбиение Вороного, ToposPro, машинное обучение, энергия миграции, ТФП расчеты

DOI: 10.31857/S2686954323601033, EDN: CXJJLK

1. ВВЕДЕНИЕ

В последние десятилетия мировое потребление лития резко увеличилось в связи с ростом производства литий-ионных аккумуляторов (ЛИА) для портативных электрических устройств и электромобилей [1], что приводит к истощению запасов лития и вызывает необходимость разработки новых систем хранения энергии и совершенствования существующих технологий ЛИА. Решение указанных вопросов неразрывно связано с использованием новых подходящих материалов. Основными элементами, определяющими емкостные свойства аккумулятора, являются активные материалы катода и анода, являющиеся кристаллическими ионными проводниками [1].

¹Самарский государственный технический университет, Самара, Россия

²Институт химии твердого тела и механохимии Сибирского отделения Российской академии наук, Новосибирск, Россия

³Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Большинство научных коллективов ищут решение сложившейся ситуации в двух направлениях: 1) поиск кристаллических ионных проводников для новых типов металл-ионных аккумуляторов, например на основе натрия, калия, магния, кальция, цинка и алюминия [1]; 2) повышение характеристик ЛИА за счет использования твердого электролита и новых материалов электродов, что повысит безопасность, энергоемкость и долговечность [2]. Оба варианта развития систем хранения энергии требуют новых материалов с высокими эксплуатационными показателями, поэтому их поиск является важной задачей современного материаловедения.

Исторически открытие новых материалов было основано на методе "проб и ошибок" или случайностей [3]. Данный подход является чрезвычайно затратным как с точки зрения времени, так и ресурсов, и поэтому неэффективен в условиях современного быстрого развития технологий. В настоящее время уровень развития теоретических методов в материаловедении позволяет уйти от концепции "проб и ошибок" к целенаправленному теоретическому (компьютерному) поиску материалов с требуемыми свойствами методами моделирования [1]. Для атомистического моде-

^{*}E-mail: vld.ospv@gmail.com

лирования необходимо знать состав и структуру вещества, и на основании этих данных можно прогнозировать различные физико-химические свойства с высокой точностью [1]. Более того, в последние десятилетия рост вычислительных мощностей позволил перейти от теоретического анализа свойств отдельных материалов к высокопроизводительному скринингу больших баз данных с целью выявления материалов с требуемыми свойствами [1]. Современный скрининг обычно проводится методами теории функционала электронной плотности (ТФП), обладающими хорошей точностью определения различных свойств: механических, электронных, магнитных. Одними из важнейших свойств материалов для металлионных аккумуляторов, которые могут быть рассчитаны в рамках ТФП, являются ионная проводимость, энергия активации диффузии, ширина запрещенной зоны, электродный потенциал, стабильность при (де)интеркаляции рабочего иона. Однако вычисления ТФП характеризуются большими затратами вычислительных ресурсов [1, 4]. Одним из вариантов снижения вычислительных затрат при скрининге баз данных является использовать системы фильтров, которые представляют собой последовательность различных методов теоретического анализа материалов, начинающихся от простых качественных методов к более сложным количественным. При анализе ионной проводимости в кристаллах можно использовать следующие фильтры: геометрико-топологический анализ (ГТ) кристаллической структуры, основанный на разбиении Вороного; расчеты с помощью метода валентных усилий связи (ВУС); кинетическое моделирование Монте-Карло; молекулярную динамику; ТФП расчеты [1]. Указанная совокупность методов позволяет достаточно эффективно выявлять вещества, в которых вероятна ионная проводимость. При этом один из важнейших критериев для оценки ионной проводимости — энергия миграции Е_m. Для соединений со щелочными рабочими ионами условно можно сказать, что значение Е_m менее 1 эВ свидетельствует о возможности диффузии, а выше 1 эВ – о ее отсутствии.

Другой вариант ускорения расчетов – использование методов машинного обучения для выявления новых закономерностей/взаимосвязей в больших многомерных данных, которые сложно анализировать "вручную". Способность МО добывать новые знания через обобщение и анализ существующих данных является ценным инструментом в разработке новых материалов. Кроме того, методы машинного обучения позволяют прогнозировать свойства новых материалов без проведения затратных вычислений. Таким образом, машинное обучение в настоящее время является важнейшим элементом ускорения теоретических разработок новых материалов и естественным образом дополняет упомянутые выше методы. Например, в работе [5] с помошью МО на сверточной нейронной сети предсказывались барьеры диффузии и энергии связи между соседними частицами на основе данных симуляций методом Монте-Карло с точностью порядка 10 мэВ. В работе [2] из базы данных Materials Project с помощью модели прогнозирования ионной проводимости было проанализировано более 12000 соединений, содержащих Li. Модель была обучена на нескольких десятках экспериментальных данных и позволила идентифицировать 21 соединение как новые ионные проводники. Последуюшее моделирование методами молекулярной динамики и теории функционала плотности для найденных соединений показало, что 10 из 21 соединения действительно обладают ионной проводимостью, о чем ранее в литературе не сообщалось. В исследовании [6] рассматривалась выборка из 15000 Li-содержащих соединений из базы данных Materials Project. С помощью машинного обучения прогнозировались потенциалы окисления и восстановления соединений с точностью $R^2 = 0.95$ и 0.92 соответственно. В недавнем исследовании [7] была предложена универсальная схема предсказания степеней окисления атомов металлов в ионных и координационных соединениях на основе параметров атомных полиэдров Вороного. Модель была обучена и проверена более чем на 35000 кристаллических структурах, содержащих более 90000 атомов металла в кислородной среде. Точность предсказания превысила 95%. В другой работе [8] была исследована взаимосвязь вида "химический состав-структура-ионная проводимость" в оксидах металлов типа граната методом регрессии опорных векторов (SVR). Прогностическая способность в отношении ионной проводимости для данной модели по статистическим параметрам R^2 , RMSE и MAE достигает 0.778, 0.372 и 0.283 соответственно. В работе [9] был проведен анализ значимости различных типов дескрипторов для предсказания энергий миграции лития в кубических аргиродитах. Было показано, что наибольшую значимость имеют структура, состав и путь миграции.

Настоящая работа направлена на создание математической модели на основе МО для оценки величины энергии миграции ионов лития в кристаллических соединениях разного структурного типа, используя в качестве исходных данных геометрические параметры свободного пространства в кристалле как основные дескрипторы для диффузии ионов. Известно, что диффузия ионов возможна только при наличии достаточного свободного пространства [1] и может быть охарактеризована в рамках разбиения Вороного такими дескрипторами, как радиус сферического домена R_{sd} , радиус элементарного канала r_{chan} , длина пути



Рис. 1. Фрагмент из четырех полиэдров Вороного, вершины которых соответствуют элементарным пустотам $R_{\rm sd}$, подходящим для размещения Li⁺ в структуре Li₄Ti₅O₁₂ (а), и основные геометрические критерии, используемые в ГТ анализе ионной проводимости: $R_{\rm sd}$ – радиус элементарной пустоты, $r_{\rm chan}$ – радиус канала диффузии в структуре (б).

диффузии *P*_{lenoth}. Прогностическая способность модели будет проверена на соединениях с найденными каналами миграции и посчитанными энергиями миграции ионов из литературных данных. Похожая работа с применением разбиения Вороного была сделана в [10], где рассматривали различные катионные и анионные проводники, включая литий-ионные проводники, основываясь на энергиях миграции Е_m ионов, полученных методом ВУС. Однако метод ВУС является приближенным и не позволяет рассчитать значение E_m с высокой точностью, в отличие от метода теории функционала плотности. В нашей работе исходные данные по Е_m были получены при помощи метода подталкивающей эластичной ленты (Nudged Elastic Band – NEB) в рамках ТФП подхода, что делает их более точными по сравнению с данными [10].

2. МЕТОДЫ

2.1. Формирование набора данных

2.1.1. Геометрико-топологический анализ. Метод ГТ анализа реализован в пакете структурнотопологических программ ToposPro [11], и основан на разбиении кристаллического пространства на полиэдры Вороного. ГТ метод может быть успешно использован для решения многих задач, связанных с диффузией ионов в твердых телах [12]. Для анализа миграции ионов в кристаллическом пространстве строится "двойной" граф: граф из атомов кристаллической решетки и разбиение Вороного, где вершины полиэдров символизируют пустоты, а ребра – каналы, соединяющие эти пустоты. Благодаря такому двойному графу описывается пустое пространство в кристаллической структуре, которое будет соответ-



Рис. 2. Длина пути P_{length} диффузии ионов лития в структуре Na_(2 – x)Li_xFePO₄F.

ствовать наиболее вероятным путям диффузии иона. Радиус сферического домена $R_{\rm sd}$ вычисляется как радиус сферы, объем которой равен объему полиэдра Вороного (рис. 1а) элементарной пустоты, построенного с учетом всех атомов структуры. Значение $r_{\rm chan}$ (рис. 1б) вычисляется как радиус окружности, описанной вокруг ребра полиэдра Вороного атома, так, чтобы все атомы, формирующие канал, лежали на данной окружности.

2.1.2. Расчеты ТФП. Для всех рассматриваемых структур барьеры миграции для ионов Li⁺ были рассчитаны в рамках ТФП с применением метода NEB. реализованного в пакете VASP. Исходные данные для расчета (список возможных путей миграции и соответствующие наборы промежуточных изображений для каждого пути) были получены с использованием скрипта РАТН-FINDER (https://pathfinder.batterymaterials.info). Во всех случаях использовались одинаковые настройки расчетов: энергия обрезания плоских волн 600 эВ, критерии сходимости при оптимизации структуры 10^{-6} эВ и 10^{-5} эВ/Å для энергии и величины межатомных сил соответственно. В расчетах был выбран обменно-корреляционный функционал в форме GGA-PBE. Для расчетов методом NEB критерий сходимости по межатомным силам был выбран на уровне 0.01 эB/Å, количество промежуточных точек пути миграции (N_{IMAGES}) равнялось 7, количество шагов оптимизации - 100.

2.1.3. Данные для обучения. Выборка данных для обучения включает информацию о 225 каналах миграции иона Li⁺ в 23 химических соединениях, для которых размеры свободного кристаллического пространства были определены в рамках ГТ подхода, а барьеры миграции лития для



Рис. 3. Корреляция между энергией миграции E_m , предсказанной двумя методами машинного обучения – методом опорных векторов и методом порядковой регрессии; и энергией миграции, рассчитанной с помощью ТФП. Данные приведены для обучающей и тестовой выборки. Пунктирная линия – идеальная зависимость. Слева – результат метода опорных векторов, справа – результат порядковой регрессии.

каждого из каналов вычислены методом ТФП. Более подробно информация указана в табл. 3.

2.2. Дескрипторы

Основной набор данных включает информацию о радиусе элементарной пустоты $R_{\rm sd}$, радиусе канала диффузии $r_{\rm chan}$ в структуре, длине канала $P_{\rm length}$ (рис. 2) и о величине энергетического барьера миграции E_m для рассматриваемого канала. Для обучения моделей использовался набор из 6 дескрипторов – $R_{\rm sd}$, $r_{\rm chan}$, $P_{\rm length}$, $R_{\rm sd}$ + $r_{\rm chan}$, $P_{\rm length}/R_{\rm sd}$ и $P_{\rm length}/r_{\rm chan}$.

2.3. Методы машинного обучения

Построение и обучение моделей было выполнено с помощью библиотеки Scikit-Learn [13] для языка программирования Python. В работе было применено два метода: регрессия методом опорных векторов и порядковая регрессия.

2.3.1. Регрессия методом опорных векторов (SVR). Базовая идея регрессии методом опорных векторов заключается в нахождении такой гиперплоскости в пространстве дескрипторов и решений, чтобы максимальное число точек данного пространства попадало в окрестность подбираемой гиперплоскости с шириной зазора є. Гиперпараметр регуляризации С позволяет управлять балансом между шириной зазора и количеством нарушений зазора. В случае решения задач нелинейной регрессии применяются параметрически редуцированные модели – искусственное добавление, в нашем случае, полиномиальных признаков *n*-й степени. Более подробно про методы регрессии опорных векторов можно узнать в работе

[14]. Подбор оптимальных гиперпарметров модели (параметра регуляризации С, ширины зазора є и степени полиномиального ядра n) был выполнен методом k-fold кросс-валидации с разбиением обучающего набора на 10 частей. При этом гиперпараметры перебирались из заданного дискретного набора. Качество моделей оценивалось средней абсолютной ошибкой.

2.3.2. Порядковая регрессия (ПР). Порядковая регрессия применяется в случаях, когда необходимо предсказать качественный результат, который можно ранжировать (например, плохо, средне, хорошо) [15], либо в задачах с нестационарной зависимостью цели от дескрипторов, которую сложно описать линейными или полиномиальными моделями [16]. В работе [17] показано, что данную задачу можно свести к решению множества задач бинарной классификации с оценкой точности каждого классификатора, в том числе на основе метода опорных векторов. Идея такого подхода заключается в большей точности ансамбля бинарных классификаторов (вследствие большей точности каждого классификатора по отдельности), по сравнению с алгоритмами мно-

Таблица 1. Значения метрик моделей. Обуч. — обучающая выборка, Тест — тестовая выборка

Метрика	SV	/R	Порядковая регрессия		
	Обуч.	Тест	Обуч.	Тест	
MAE	0.1421	0.1326	0.1211	0.1261	
RMSE	0.1897	0.1817	0.1667	0.1728	
R ²	0.7780	0.8043	0.8301	0.8231	



Рис. 4. Зависимость средних энергий миграции E_m , предсказанных моделями, от радиуса канала $r_{\rm chan}$, длины пути миграции $P_{\rm length}$ и радиуса элементарной пустоты $R_{\rm sd}$.

жественной классификации. В случае работы с непрерывными данными остается разбить набор целевых значений на фиксированное количество диапазонов с желаемым разрешением. Каждый классификатор из ансамбля далее обучается на всем наборе данных, однако отвечает за разделение данных только по одной границе. Из результатов всего ансамбля далее определяется наиболее вероятный целевой диапазон значения. В качестве бинарных классификаторов в данной работе был выбран метод опорных векторов с мягким зазором и полиномиальным ядром свертки *п*-й степени. Степень полиномиального ядра n, параметр регуляризации C и весовой параметр, который отвечает за степень влияния полиномов высокой степени на результат, подбирались методом k-fold кросс-валидации с разбиением обучающего набора на 5 частей. Гиперпараметры подбирались независимо для каждого бинарного

классификатора и перебирались из заданного дискретного набора. Качество моделей оценивалось метрикой "ассигасу" (точность). Точность каждого бинарного классификатора оценивалась как средняя точность кросс-валидации лучшей модели [18]. Таким образом, результатом работы каждого бинарного классификатора является вероятность нахождения образца в одном из двух диапазонов – ниже заданной границы вероятность P_i , выше – $(1 - P_i)$. *m* классификаторов дают информацию о положении каждого образца относительно *m* границ, т.е. m + 1 диапазонах. Вероятность нахождения образца в каждом конкретном диапазоне $P_{i,i}$ равна средней вероятности, предсказанной всеми классификаторами в рамках данного диапазона:

$$P_{i,j} = \frac{1}{m} \sum_{k=1}^{m} P(k),$$
 (1)

$$P(k) = \begin{cases} P_k, & k < j \\ 1 - P_k, & k > i \end{cases}$$
(2)

Результатом работы алгоритма является наиболее вероятный диапазон данных. При проверке модели в качестве результата бралось среднее значение диапазона, при этом использовались нижеописанные метрики.

2.3.3. Критерии оценки моделей. Для оценки качества моделей использовались три метрики – средняя абсолютная ошибка (MAE), квадратный корень из среднеквадратичной ошибки (RMSE) и коэффициент детерминации (*R*²). Указанные метрики определяются следующим образом:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y'_i|, \qquad (3)$$

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (y_i - y'_i)^2},$$
 (4)

Соединение	Путь в источнике	$P_{\text{length}}, \text{\AA}$	R _{sd} , Å	<i>r</i> _{chan} , Å	<i>E_m</i> , эВ (ПР)	E_m , $\Im B$ (SVR)	<i>E_m</i> , эВ (висточнике)	Источник
LiFePO ₄	1	3.003	1.385	1.743	0.4-0.5	0.592	0.2-0.27	[21]
LiFePO ₄	2	4.691	1.231	1.29	>1.4	1.6376	>2.5	[21]
$LiFe_{0.5}Mn_{0.5}PO_4$	А	3.025	1.396	1.76	0.4-0.5	0.56	0.59	[20]
$LiFe_{0.5}Mn_{0.5}PO_4 \\$	В	4.714	1.256	1.687	1.2-1.3	1.3436	2.86	[20]
$LiFe_{0.5}Mn_{0.5}PO_4 \\$	С	5.701	1.346	1.822	1.1-1.2	1.2314	3.58	[20]
LiCoO ₂	ODH	2.811	1.286	1.721	0.8-0.9	0.8219	0.49	[22]
Li ₂ FeSiO ₄	AB	3.097	2.614	3.856	0.1-0.2	-3.33	0.91	[23]
Li ₂ FeSiO ₄	CD	3.165	2.5	3.456	0.1-0.2	-2.89	1.55	[23]

Таблица 2. Сравнение энергии миграции рабочего иона, предсказанной двумя разными методами машинного обучения с литературными данными. ПР – метод порядковой регрессии, SVR – метод опорных векторов

ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ том 514 № 2 2023

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - y_{i}')^{2}}{\sum_{i=1}^{n} (y_{i} - y_{av})^{2}},$$
(5)

где y_i является рассчитанным значением энергии миграции, y'_i — значение энергии миграции, предсказанное моделью, y_{av} — среднее значение энергий миграции набора данных.

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1. Построение и обучение моделей для предсказания энергий миграции ионов лития

Для формирования разнообразной тестовой выборки, исходный набор данных (включая значения энергий миграции) был разделен на 6 поднаборов методом k-ближайших соседей [19], каждый из которых был разделен на обучающий (80%) и проверочный (20%) наборы данных. Было обучено две модели – "классическая" регрессионная модель (SVR) и модель порядковой регрессии. В случае порядковой регрессии обучалось 14 бинарных классификаторов, которые разделяли данные по границам от 0.1 до 1.4 эВ с шагом 0.1 эВ. Оптимальные гиперпараметры моделей приведены в табл. 4 и 5. На рис. 3 показано качество моделей на примере тренировочного и тестового набора данных соответственно. В табл. 1 приведены значения всех метрик для обучающего и тестового набора литий-ионных проводников.

Из полученных метрик можно сделать вывод, что модели для лития не переобучены и удовлетворительно описывают как обучающие, так и тестовые данные.

3.2. Анализ моделей

В случае моделей для лития, несмотря на дискретность, порядковая регрессия качественнее описывает наборы данных в целом, что свидетельствует о нестационарном поведении дескрипторов. Тем не менее классическая модель качественнее предсказывает диапазон энергий миграции от 0.7 до 1.4 эВ, который практически неразличим для модели порядковой регрессии. Для наглядности была построена зависимость средних энергий миграции, которые предсказывают две модели, от радиуса канала, длины пути миграции и радиуса элементарной пустоты (рис. 4).

Полученные результаты свидетельствуют, что геометрические параметры свободного кристаллического пространства — R_{sd} и r_{chan} коррелируют с энергией миграции иона; чем выше значение этих параметров, тем ниже энергия миграции иона. Кроме того, наблюдается качественная зависимость между длиной пути диффузии рабоче-

го иона *P*_{length} и его энергией миграции: более длинные пути соответствуют более высоким энергиям миграции.

3.3. Проверка моделей МО на литературных данных

Для дополнительного теста полученных моделей были отобраны литературные данные из работ [20–23], в которых применялись схожие методики ТФП расчетов. Были определены пустоты и каналы, доступные для миграции ионов Li⁺, по которым были вычислены параметры $R_{\rm sd}$ и $r_{\rm chan}$. Также из данных работ были взяты данные о расстоянии между началом и концом каналов диффузии (параметр $P_{\rm length}$). Сравнение опубликованных данных с результатами моделей приведено в табл. 2.

Из табл. 2 видно, что разработанные модели переоценили энергию миграции Е_т для пути 1 в соединении LiFePO₄, но недооценили для пути 2. Вместе с тем в статье [24] ТФП расчеты для путей 1 и 2 в том же соединении показывают значения 0.55 и 2.89 эВ, соответственно, и в этом случае спрогнозированные значения для канала 1 хорошо согласуются с результатами ТФП. Величина барьера миграции для канала 2 выходит за пределы обучающего набора для значений энергии миграции выше 1.5 эВ. Для пути А в структуре LiFe_{0.5}Mn_{0.5}PO₄ модели МО показывают хорошее согласие с литературными данными. а для путей В и С плохая прогностическая способность также объясняется выходом за пределы обучающего набора. Неточные результаты прогноза в структуре Li₂FeSiO₄ вероятно вызваны недостаточным набором кремний-содержащих соединений в обучающей выборке и выходом за пределы обучающего набора для пути CD. Отличие спрогнозированной энергии миграции для LiCoO₂ (0.8–0.9 эВ) от рассчитанной в литературе (0.49 эВ) может объясняться слоистой структурой соединения, которая недостаточно представлена в обучающей выборке. В целом полученные модели дают несколько заниженные значения энергии миграции, что объясняется как ограниченностью обучающей выборки (ограниченность по химическому составу соединений и их структуре), так и небольшим набором дескрипторов. Тем не менее разработанные модели МО в текущем виде уже позволяют на качественном уровне отделить пути, где миграция возможна ($E_m < 1$ эВ) от тех, где она маловероятна.

4. ВЫВОДЫ

Интуитивно понятный способ первичного скрининга баз данных структур по геометрикотопологическим критериям заслуживает особого внимания в качестве одного из фильтров, включенного в последовательность других методов теоретического анализа ионного транспорта в кристаллах. Было определено, что на энергию миграции влияет радиус элементарной пустоты: с возрастанием радиуса элементарной пустоты уменьшается энергия миграции.

Разработанные модели МО могут применяться в качестве дополнительного фильтра при анализе ионной проводимости в структурах. Однако данная работа показывает, что на качество прогнозирования математической модели энергии миграции существенное внимание оказывают разнообразие соединений по составу и структуре, объем выборки соединений и количество дескрипторов, описывающих диффузию иона. Так, в дальнейшем обучающий набор будет дополнен данными по другим ионным проводникам, а также увеличено количество дескрипторов, описывающих свободное кристаллическое пространство, структуру, состав и типы межатомных взаимодействий в кристалле.

Приложение 1

Таблица 3. Список кристаллических ионных проводников с рабочим ионом Li^+ , данные по которым были использованы для обучения моделей МО

Формула	Количество путей диффузии Li ⁺	Группа симметрии	Источник данных, DOI
Na _(2-x) Li _x FePO ₄ F	3	Pbcn	Собственные расчеты, будут опубликованы
$Li_7La_3Zr_2O_{12}$	4	<i>I</i> 41/ <i>acd</i>	10.1021/acsaem.1c03632
LiVPO ₄ F	4	<i>P</i> -1	10.1039/d0cp02204g
$Li_7La_3Zr_{0.5}Ta_{1.5}O_{12}$	6	Ia-3d	10.1021/acsaem.1c03632
$Li_7La_3Zr_{0.25}Ta_{1.75}O_{12}$	1	Ia-3d	10.1021/acsaem.1c03632
$Li_7La_3Zr_{0.75}Ta_{1.25}O_{12}$	3	Ia-3d	10.1021/acsaem.1c03632
$Li_7La_3Ta_2O_{12}$	6	Ia-3d	10.1021/acsaem.1c03632
$Li_7La_3Zr_{1.5}Ta_{0.5}O_{12}$	4	Ia-3d	10.1021/acsaem.1c03632
$Li_7La_3Zr_{1.25}Ta_{0.75}O_{12}$	7	Ia-3d	10.1021/acsaem.1c03632
$Li_7La_3Zr_{1.75}Ta_{0.25}O_{12}$	7	Ia-3d	10.1021/acsaem.1c03632
Li ₇ La ₃ ZrTaO ₁₂	2	Ia-3d	10.1021/acsaem.1c03632
$Li_{7}La_{3}Zr_{2}O_{12}(1)$	10	Ia-3d	10.1021/acsaem.1c03632
$Li_{7}La_{3}Zr_{2}O_{12}(2)$	9	Ia-3d	10.1021/acsaem.1c03632
$Li_{1.25}Ti_{0.5}Mn_{0.25}O_2(1)$	16	F <i>m</i> -3 <i>m</i>	Собственные расчеты, будут опубликованы
$Li_{1.25}Ti_{0.5}Mn_{0.25}O_2(2)$	3	F <i>m</i> -3 <i>m</i>	
$Li_{1.2}Nb_{0.2}Mn_{0.6}O_{2}$	28	F <i>m</i> -3 <i>m</i>	
$Li_{1.2}Ti_{0.4}Mn_{0.4}O_2$	26	F <i>m</i> -3 <i>m</i>	
$Li_{1.3}Nb_{0.3}Mn_{0.4}O_2(1)$	16	F <i>m</i> -3 <i>m</i>	
$Li_{1.3}Nb_{0.3}Mn_{0.4}O_2$ (2)	40	F <i>m</i> -3 <i>m</i>	
$Li_{1.5}Ge_{1.5}Al_{0.5}(PO_4)_3$	2	R-3 <i>c</i>	
$LiCoVO_4(1)$	7	F <i>d</i> -3 <i>m</i>	10.3390/pr11051427
$LiCoVO_4(2)$	14	F <i>d</i> -3 <i>m</i>	10.3390/pr11051427
LiZnPS ₄	1	<i>I</i> -4	Собственные расчеты
Li ₂ CoPO ₄ F	6	Pnma	10.1021/acs.jpcc.6b11027

Таблица 4. Гиперпараметры бинарных классификаторов модели порядковой регрессии. *С* – параметр регуляризации, *n* – степень полиномиального ядра, соеf0 – весовой параметр, отвечающий за степень влияния полиномов высокой степени на результат, Ассигасу – точность классификатора, оцененная методом кроссвалидации

N⁰	Граница (эВ)	С	n	coef0	Accuracy
1	0.1	1.0	1	1.0	0.977
2	0.2	1.0	1	1.0	0.875
3	0.3	1.0	1	1.0	0.886
4	0.4	1.0	1	1.0	0.892
5	0.5	1.0	1	1.0	0.926
6	0.6	1.0	1	1.0	0.903
7	0.7	3.0	2	7.0	0.937
8	0.8	8.0	4	7.0	0.864
9	0.9	10.0	1	1.0	0.791
10	1.0	2.0	4	1.0	0.695
11	1.1	1.0	3	1.0	0.870
12	1.2	3.0	3	1.0	0.904
13	1.3	1.0	2	1.0	0.955
14	1.4	1.0	3	1.0	1.0

Таблица 5. Гиперпараметры модели SVR. *С* – параметр регуляризации, *n* – степень полиномиального ядра, epsilon – ширина зазора

С	n	epsilon
0.1	1	0.034

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Исследования выполнены за счет гранта Российского научного фонда (проект 19-73-10026).

СПИСОК ЛИТЕРАТУРЫ

- 1. *Kabanov A.A. et al.* Computational design of materials for metal-ion batteries / ed. Reedijk J., Poeppelmeier K.R.B.T.-C.I.C.I.I.I. Third E. Oxford: Elsevier, 2023. P. 404–429.
- Sendek A.D. et al. Machine Learning-Assisted Discovery of Solid Li-Ion Conducting Materials // Chem. Mater. American Chemical Society, 2018. V. 31. № 2. P. 342–352.
- Carraher C.E. Columns: General Topics // Polym. News. Taylor & Francis, 2005. V. 30. № 5. P. 155–157.
- Lv C. et al. Machine Learning: An Advanced Platform for Materials Development and State Prediction in Lithium-Ion Batteries // Adv. Mater. 2022. V. 34, № 25. P. 2101474.
- Martynec T. et al. Machine learning predictions of surface migration barriers in nucleation and non-equilibrium growth // Commun. Mater. 2021. V. 2. № 1. P. 90.

- 6. *Honrao S.J. et al.* Discovery of novel Li SSE and anode coatings using interpretable machine learning and high-throughput multi-property screening // Sci. Rep. 2021. V. 11. № 1. P. 16484.
- Shevchenko A.P. et al. Mining Knowledge from Crystal Structures: Oxidation States of Oxygen-Coordinated Metal Atoms in Ionic and Coordination Compounds // J. Chem. Inf. Model. American Chemical Society, 2022. V. 62. № 10. P. 2332–2340.
- 8. *Kireeva N., Pervov V.S.* Materials space of solid-state electrolytes: unraveling chemical composition–structure–ionic conductivity relationships in garnet-type metal oxides using cheminformatics virtual screening approaches // Phys. Chem. Chem. Phys. The Royal Society of Chemistry, 2017. V. 19. № 31. P. 20904–20918.
- *Zhao Q. et al.* Machine learning prediction of activation energy in cubic Li-argyrodites with hierarchically encoding crystal structure-based (HECS) descriptors // Sci. Bull. Science China Press, 2021. V. 66. № 14. P. 1401–1408.
- *Zhang L. et al.* A Database of Ionic Transport Characteristics for Over 29 000 Inorganic Compounds // Adv. Funct. Mater. John Wiley & Sons, Ltd, 2020. V. 30. N
 № 35. P. 2003087.
- 11. Blatov V.A., Shevchenko A.P., Proserpio D.M. Applied Topological Analysis of Crystal Structures with the Program Package ToposPro // Cryst. Growth Des. American Chemical Society, 2014. V. 14. № 7. P. 3576–3586.
- Morkhova Y.A. et al. Computational Search for Novel Zn-Ion Conductors—A Crystallochemical, Bond Valence, and Density Functional Study // J. Phys. Chem. C. 2021.
- Pedregosa F. et al. Scikit-learn: Machine Learning in Python // J. Mach. Learn. Res. 2011. V. 12. P. 2825– 2830.
- 14. *Smola A.J., Scholkoph B.* A tutorial on support vector regression // Stat. Comput. 2004. V. 14. P. 199–222.
- Ananth C.V., Kleinbaum D.G. Regression models for ordinal responses: A review of methods and applications // Int. J. Epidemiol. 1997. V. 26. № 6. P. 1323–1333.
- Niu Z. et al. Ordinal Regression with Multiple Output CNN for Age Estimation // Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016. P. 4920– 4928.
- Li L., Lin H.T. Ordinal Regression by Extended Binary Classification // NIPS 2006 Proc. 19th Int. Conf. Neural Inf. Process. Syst. 2006. P. 865–872.
- Zhang Y., Yang Y. Cross-validation for selecting a model selection procedure // J. Econom. Elsevier B.V., 2015. V. 187. № 1. P. 95–112.
- Guo G. et al. KNN model-based approach in classification // Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2003. V. 2888. P. 986–996.
- 20. Gardiner G.R., Islam M.S. Anti-Site Defects and Ion Migration in the LiFe0.5Mn0.5PO4 Mixed-Metal

Cathode Material[†] // Chem. Mater. 2010. V. 22. P. 1242–1248.

- Morgan D., Ven A., Ceder G. Li conductivity in LixM-PO4 (M = Mn, Fe, Co, Ni) olivine materials // Electrochem. Solid-State Lett. 2004. V. 7. P. A30.
- Li J.-J., Dai Y., Zheng J.-C. Strain engineering of ion migration in LiCoO2 // Front. Phys. 2021. V. 17. № 1. P. 13503.
- 23. Armstrong A.R. et al. Structure and Lithium Transport Pathways in Li₂FeSiO₄ Cathodes for Lithium Batteries // J. Am. Chem. Soc. American Chemical Society, 2011. V. 133. № 33. P. 13031–13035.
- 24. *Islam M.S. et al.* Atomic-Scale Investigation of Defects, Dopants, and Lithium Transport in the LiFePO4 Olivine-Type Battery Material // Chem. Mater. American Chemical Society, 2005. V. 17. № 20. P. 5085–5092.

MACHINE LEARNING AS A TOOL TO ACCELERATE THE SEARCH FOR NEW MATERIALS FOR METAL-ION BATTERIES

V. T. Osipov^{*a*,[#]}, M. I. Gongola^{*b*,*c*}, E. A. Morkhova^{*a*}, Corresponding Member of the RAS A. P. Nemudrvi^{*b*}, and A. A. Kabanov^{*a*,*b*}

^aSamara State Technical University, Samara, Russia

^bInstitute of Solid State Chemistry and Mechanochemistry of the Siberian Branch of the RAS, Novosibirsk, Russia

^cNovosibirsk State University, Novosibirsk, Russia

[#]E-mail: vld.ospv@gmail.com

Presented by Academician of the RAS A.I. Avetisyan

The search for new solid ionic conductors is an important topic of material science that requires significant resources, but can be accelerated using machine learning (ML) techniques. In this work, ML methods were applied to predict the migration energy of working ions. The training set is based on data on 225 lithium ion migration channels in 23 ion conductors. The descriptors were the parameters of free space in the crystal obtained by the Voronoi partitioning method. The accuracy of migration energy prediction was evaluated by comparison with the data obtained by the density functional theory method. Two methods of ML were applied in the work: support vector regression and ordinal regression. It is shown that the parameters of free space in a crystal correlate with the migration energy, while the best results are obtained by ordinal regression. The developed ML models can be used as an additional filter in the analysis of ionic conductivity in solids.

Keywords: ionic conductors, Voronoi partitioning, ToposPro, machine learning, migration energy, DFT calculations