

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

доктора химических наук, главного научного сотрудника Киселевой Надежды Николаевны

на диссертационную работу Шевченко Александра Петровича

«Теория и методы компьютерного геометрико-топологического анализа и прогнозирования строения и физических свойств координационных соединений», представленную на соискание ученой степени доктора химических наук по специальности 1.4.4. - физическая химия

Актуальность избранной темы диссертации

Диссертация А.П. Шевченко посвящена разработке теоретических моделей строения кристаллических веществ, алгоритмов и методов автоматизированной обработки больших объемов кристаллографических данных для поиска закономерностей «состав-структура-свойство» в кристаллах координационных соединений (КС) (в том числе и металл-органических координационных полимеров (МОКП)), позволяющих прогнозировать их строение и физические свойства.

Прогнозирование кристаллической структуры и физических свойств соединений является одной из важнейших проблем химии. Большие надежды ее решения были связаны с использованием квантово-механических методов. Однако уже в середине прошлого века стало ясно, что возможности этих методов недостаточны для точной априорной оценки кристаллической структуры и физических свойств еще не синтезированных соединений, особенно сложного состава. Как результат, в практических работах исследователи использовали, в основном, свой опыт и интуицию для поиска соединений с заданными свойствами. В связи с этим уже в шестидесятые годы были начаты работы по использованию методов искусственного интеллекта (ИИ) для моделирования процесса поиска химиками новых соединений (в первую очередь, органических). Несмотря на ограниченные возможности вычислительных машин того времени, результаты применения искусственного интеллекта для решения задач автоматизации поиска еще не синтезированных соединений с заданными свойствами оказались многообещающими. Новый всплеск интереса к использованию ИИ в химии был связан с разработкой свободно распространяемых пакетов программ машинного обучения, появлением мощных вычислительных машин и больших баз данных по свойствам веществ и материалов. Возможности ИИ в химии были оценены в США, где в 2011 г. правительство выдвинуло и профинансировало проект, названный «Инициативой Геном Материалов» (Materials Genome Initiative). Выполнение этого проекта позволило ускорить разработку новых материалов, обладающих заданными свойствами, что критично для достижения высокого уровня конкурентоспособности промышленности США и способствовало поддержке их лидирующей роли во многих секторах современного материаловедения и промышленности: от энергетики до электроники, от обороны до здравоохранения. В настоящее время актуальность разработки и использования методов компьютерного моделирования в химии и материаловедении не вызывает сомнения. Это тренд современной теоретической химии.

Научная новизна полученных результатов

Методы машинного обучения являются универсальным средством решения разнообразных задач. Сейчас опубликованы десятки тысяч статей, в которых представлены результаты успешного использования машинного обучения для решения самых разных задач химии и материаловедения. Однако машинное обучение имеет ряд ограничений. В первую очередь, его применение требует использования больших обучающих выборок, содержащих качественную информацию, что непосредственно связано с разработкой и использованием баз данных. Результат также зависит от правильности постановки задачи и выбора дескрипторов, от которых зависит прогнозируемое свойство. Последние задачи

решаются на основе тщательного анализа предметной области и использования теоретических моделей строения кристаллических веществ. В диссертационной работе А.П. Шевченко разработаны такие теоретические модели на основе геометрического и топологического моделирования атомных ансамблей и свободного пространства в кристаллических структурах с помощью их представления в виде периодических графов и разбиения Вороного. Диссертантом не только впервые объединены топологический и геометрический подходы при анализе и моделировании структур КС и МОКП, но и методы и модели кристаллохимии и материаловедения с информационными технологиями. В результате создано новое междисциплинарное направление, ориентированное на создание универсальных систем ИИ, пригодных для прогнозирования структуры и свойств кристаллических веществ любой природы.

Практическая значимость полученных результатов

Разработанный А.П. Шевченко геометрико-топологический подход к анализу и прогнозированию кристаллических структур и их свойств активно используется тысячами исследователей во всем мире. Большое практическое значение имеют созданные базы данных топологических типов и структурных строительных блоков, а также разработанные на их основе сервисы. Они применяются при топологическом и кристаллохимическом описании новых синтезируемых КС и МОКП, при поиске закономерностей строения кристаллических веществ, анализе свободного пространства и скрининге структурной информации для отбора потенциальных твердых электролитов.

Важно, что созданные топологические базы данных и использующие их сервисы, доступны пользователям в сети Интернет на сайте <https://topcryst.com>.

Важное практическое решение имеет разработанная в диссертации система прогнозирования, которая может использоваться специалистами для оценки структуры и свойств кристаллических веществ.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации, их достоверность

Основные положения и выводы диссертации обоснованы комплексом полученных расчетных данных и анализом литературы. Достоверность расчетов подтверждена статистическими методами. Точность прогнозов оценена с применением широко используемого в машинном обучении метода скользящего контроля и статистических критериев.

Результаты диссертации опубликованы в научных журналах, в которых статьи подвергаются тщательному рецензированию высококвалифицированными экспертами.

Публикации и апробация работы

Результаты диссертационной работы опубликованы в двух главах в монографиях, 33 статьях в журналах, индексируемых системами Web of Science и Scopus, а также 18 зарегистрированных программах и полностью соответствуют требованиям пункта 11 Положениях о присуждении ученых степеней. Результаты диссертации также прошли апробацию в форме 14 докладов на научных конференциях.

Следует отметить высокую цитируемость публикаций А.П. Шевченко по данным РИНЦ. Например, публикация 21 (автореферат диссертации) процитирована 2652 раз, а публикация 18 (автореферат диссертации) процитирована 1000 раз (данные от 21.02.25).

Рассмотрение публикаций диссертанта, представленных в автореферате, показало, что они полностью отражают основное содержание диссертации.

Соответствие паспорту специальности

Содержание диссертации полностью соответствует паспорту научной специальности 1.4.4. «Физическая химия» по следующему пункту:

10. Создание и разработка методов компьютерного моделирования строения и механизмов превращений химических соединений на основе представлений квантовой механики, различных топологических и статистических методов, включая методы машинного обучения, методов молекулярной механики и молекулярной динамики, а также подходов типа структура-свойства.

Структура и содержание работы

Диссертационная работа состоит из введения, обзора литературы, экспериментальной части, обсуждения результатов, выводов, заключения, списка литературы (284 источника) и приложения. Текст диссертационной работы изложен на 204 страницах, включает 36 таблиц и 91 рисунок.

Литературный обзор (Глава 1) включает информацию об имеющихся базах кристаллоструктурных данных (БД), а также подробный анализ публикаций о существующих моделях строения кристаллических соединений и об использовании методов машинного обучения для поиска закономерностей в структурных данных. На основе анализа публикаций диссертант отобрал дескрипторы, наиболее важные для описания кристаллической структуры КС и МОКП, и предложил общую схему обработки и применения кристаллографических данных.

Вторая глава содержит описание разработанного Шевченко универсального геометрико-топологического подхода к моделированию и анализу кристаллической структуры соединений любого химического состава и строения. Разработанный подход алгоритмизирован и реализован диссертантом в виде процедур программного пакета TorosPro, подробно рассмотренном в Главе 2. Разработанные программы пакета TorosPro и хранящаяся в его БД информация была использована А.П. Шевченко для создания общей схемы подготовки и анализа кристаллографических данных для дальнейшей обработки с применением геометрико-топологического подхода, статистических методов и программ машинного обучения, результаты которой приведены в Главе 3.

В Главе 3 подробно изложены постановка задач поиска закономерностей в информации о КС и МОКП и использования найденных закономерностей для прогнозирования свойств (размерности и топологического типа МОКП, пористости кристаллической структуры, степени окисления атомов металла в МОКП и т.д.) этих соединений разного состава. Статистический анализ полученных результатов показал их достаточно высокую точность.

Универсальность предложенного им нового геометрико-топологического подхода для поиска закономерностей «состав-структура-свойство» в кристаллах продемонстрирована А.П. Шевченко в Главе 4, в которой приведены результаты решения разнообразных задач прогнозирования свойств простых веществ, твердых электролитов и интерметаллидов. В частности, была решена актуальная и практически значимая задача прогнозирования новых потенциальных твердых электролитов, содержащих подвижные катионы натрия.

Замечания и вопросы по диссертации

1. Какая система управления базой данных была использована в комплексе TorosPro?

2. К сожалению, при использовании методов машинного обучения диссертант ограничился только программами из библиотеки scikit-learn на языке Python: классификаторов случайного леса - RF, К-ближайших соседей - KNN, дерева решений - CART, метода

опорных векторов - SVC, логистической регрессии - LR и гауссовского наивного байесовского классификатора - GNB). Хотелось бы порекомендовать диссертанту в дальнейших исследованиях расширить набор программ прогнозирования категориальных (дискретных) целевых параметров, включив, например, хорошо зарекомендовавшие себя программы на основе обучения нейронных сетей. Помимо этого, было бы целесообразно использовать для принятия коллективного решения ансамбли алгоритмов, позволяющие значительно увеличить точность прогнозов.

3. Объем данных для машинного обучения, используемых в диссертационной работе, достаточно велик, однако размеры классов различаются на два порядка. Естественно, что точность прогнозирования аналогов объектов малых классов будет невысокой. В таких случаях одним из решений является последовательные дихотомии - объекты целевого класса и все остальные. Далее проводится сравнение результатов многоклассового прогнозирования и серии дихотомий. Качество прогнозирования с использованием дихотомий можно оценить с помощью ROC-анализа. В дальнейшем было бы интересно использовать такой подход для повышения точности прогнозов.

4. При постановке задач с использованием машинного обучения диссертант ограничился прогнозированием только категориальных свойств, хотя можно было бы переформулировать задачи с целью оценки непрерывных свойств (например, ионной проводимости, пористости структуры). Многочисленный класс соответствующих программ для решения таких задач включен в библиотеку scikit-learn. Такие оценки физических свойств значительно более ценны для химиков и материаловедов, чем прогнозирование принадлежности к тому или иному классу веществ. Хотелось бы, чтобы диссертант при дальнейшем совершенствовании разработанной компьютерной системы расширил ее возможности за счет включения таких программ.

Следует отметить, что вышеуказанные замечания не ставят под сомнение достоверность и значимость основных результатов и выводов диссертации.

Заключение

Диссертация Александра Петровича Шевченко «Теория и методы компьютерного геометрико-топологического анализа и прогнозирования строения и физических свойств координационных соединений» является завершенной научно-квалификационной работой, в которой диссертантом создано новое междисциплинарное направление, ориентированное на создание универсальных компьютерных систем, пригодных для автоматизации прогнозирования структуры и свойств кристаллических веществ любой природы. Решение этой задачи имеет важное значение не только для развития теоретической химии, но и для интенсификации исследований при поиске новых материалов с заданными свойствами. Диссертационная работа по новизне, актуальности, практической значимости, достоверности результатов и обоснованности выводов удовлетворяет требованиям п.9-14 «Положения о присуждении ученых степеней (с изменениями на 25 января 2024 года) (утверждено Постановлением Правительства Российской Федерации от 24 сентября 2013 года N 842), а ее автор, Шевченко Александр Петрович, заслуживает присуждения ученой степени доктора химических наук по специальности 1.4.4. - «Физическая химия».

Официальный оппонент:
доктор химических наук
по специальности 02.00.21
«Химия твердого тела»,
главный научный сотрудник
Федерального государственного
бюджетного учреждения науки
Института металлургии и

материаловедения им. А.А. Байкова
Российской академии наук (ИМЕТ РАН)
Киселева Надежда Николаевна



подпись

24 февраля 2025 г.

Почтовый адрес:
119334, г. Москва, Ленинский проспект, д.49.
Институт металлургии и материаловедения
им. А.А. Байкова РАН, лаборатория
полупроводниковых материалов
тел.: 8(499)1352591,
e-mail: kis@imet.ac.ru

Подпись главного научного сотрудника Киселевой Н.Н. заверяю:

Заместитель директора ИМЕТ РАН, д.т.н.  В.С. Юсупов

24 февраля 2025 г.

