

К возможности прогнозирования значений параметров многокомпонентных неорганических соединений¹

Е.А.Ващенко*, М.А.Витушко*, В.А.Дударев**, Н.Н.Киселева**,
В.С.Переверзев-Орлов*

*Институт проблем передачи информации им. А.А. Харкевича, Российская академия наук, Москва, Россия

**Институт металлургии и материаловедения им. А.А. Байкова, Российская академия наук, Москва, Россия

Поступила в редколлегию 12.08.2019

Аннотация—Предложена модель для прогнозирования свойств неорганических веществ, основанная на “мягком” голосовании “кликковых” решателей, и разработаны алгоритмы ее реализации в приложении к химическим задачам. В качестве основы для исследований были выбраны два варианта кликовых методов, разработанных нами прежде в связи с задачами диагностики и прогнозирования в медицине – “Синдромный анализ” и “Фрагмент-Потенциал”, обладающих широким диапазоном возможностей для решения такого рода задач. Эти методы рассматривались в специфическом варианте обучения распознаванию, когда решение о прогнозируемом объекте принимается в ориентации на ближайшие по смыслу прецеденты, что реализуется голосованием решателей разной природы, формируемых при решении задачи обучения распознаванию прогнозируемого объекта при сопоставлении его с прецедентным множеством. Перспективность предложенного подхода была продемонстрирована на примере решения тестовой задачи прогнозирования параметров кристаллической решетки еще не полученных соединений с кристаллической структурой мелилита разного состава. Полученные результаты свидетельствуют о перспективности использования разработанных методов для оценки количественных свойств неорганических веществ.

КЛЮЧЕВЫЕ СЛОВА: машинное обучение, кликовые методы, синдромный анализ, фрагмент-потенциал, параметры неорганических соединений, параметры кристаллической решетки, мелилит.

1. ВВЕДЕНИЕ

В последние годы возрос интерес к методам машинного обучения как эффективному средству ускорения поиска и разработки новых материалов с заданными функциональными свойствами [1–15]. Одной из причин было то, что использование квантово-механических методов для расчетов еще не полученных многокомпонентных неорганических веществ, особенно в твердой фазе, связано с объективными трудностями. Термодинамические расчеты также неэффективны для прогноза еще не синтезированных неорганических веществ. С помощью же методов машинного обучения, проанализировав имеющуюся информацию об уже исследованных веществах, можно прогнозировать еще не синтезированные соединения и оценивать некоторые их свойства, зная только сравнительно хорошо известные параметры компонентов (химических элементов или более простых соединений).

¹ Работа выполнена при частичной финансовой поддержке гранта Российского фонда фундаментальных исследований № 18-07-00080.

Наиболее успешно применяемые в неорганической химии и материаловедении методы машинного обучения связаны с классификацией объектов, т.е. с прогнозированием их качественных (дискретных) характеристик. Использование такого подхода в неорганической химии дало возможность прогнозировать качественные свойства веществ (например, возможность образования соединений определенного состава и/или с заданным типом кристаллической структуры при определенных внешних условиях) со средней точностью выше 80% (по результатам сравнения прогнозов с новыми экспериментальными данными) [1, 3]. Однако прогнозирование дискретных свойств является лишь малой частью практических задач в химии. Подавляющее большинство проблем связано с предсказанием количественных характеристик объектов (например, параметров кристаллической решетки, температур плавления или кипения, температуры перехода в сверхпроводящее состояние, ширины запрещенной зоны, ударной прочности, упругости, электропроводности и т.п.). Несмотря на большую важность решения таких задач, методы машинного обучения для оценки количественных (численных, скалярных, векторных) свойств объектов, которые часто ассоциируются с задачами регрессии, используются значительно реже. В приложениях к неорганической химии этот парадокс связан, в частности, с ограничениями классического регрессионного анализа: проблемами мультиколлинеарности (и как следствие плохой обусловленностью матрицы признаков описания), негладкостью аппроксимируемых зависимостей, большой размерностью признаков описания в сочетании с малым количеством прецедентов, наличием ошибочных выбросов в данных и т.д. Использование регуляризации, различных методов отбора наиболее важных признаков, фильтрации (методы гребневой регрессии, LASSO, LARS, эластичных сетей, регуляризованных нейронных сетей и т.п.) во многом позволяют обойти некоторые из этих ограничений. Однако пока не разработан комбинированный подход, который преодолел бы все ограничения при решении задач восстановления многомерной регрессии. Один из путей создания такого метода связан с учетом особенностей предметной области (неорганической химии). Настоящее исследование посвящено поиску путей прогноза количественных параметров неорганических веществ.

Для тестирования разработанных методов выбраны соединения с ацентричной кристаллической структурой типа мелилита, представляющие интерес для поиска новых нелинейно-оптических [16] и пьезоэлектрических [17] материалов. Решается типичная для неорганической химии задача оценки значений параметров кристаллической решетки потенциально интересных новых мелилитов в системах $AO-B_2O_3-SiO_2(GeO_2)$ и $AO-BO-SiO_2(GeO_2)$, где А и В – разные химические элементы.

2. МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ

В качестве основы для исследований были выбраны два варианта кликовых методов, разработывавшихся нами прежде в связи с решением задач диагностики и прогнозирования в медицине, – “Синдромный анализ” и “Фрагмент-Потенциал”, обладающих широким диапазоном свойств и возможностей для решения задач в различных предметных областях.

“Синдромный анализ” (СА). Синдром рассматривается как сумма симптомов с порогом на число симптомов на входе синдрома. Симптом это – бинарный признак, соответствующий попаданию величины какой-то измеряемой сущности в какой-то характерный для нее интервал в решаемой задаче. Синдром может быть симптомом какого-то другого синдрома. Все это в совокупности порождает направленную с априорно не заданной сложностью пороговую синдромную сеть, выражающую некое содержательное понятие (нейрон), характеризующее некий класс или подкласс интересующих объектов [18–20].

“Фрагмент-Потенциал” (ФП). Метод основан на специальном виде голосующих кусочно-линейных правилах для одновременного разделения многих классов с последующим голосова-

нием таких правил и переводом решения исходной задачи в “пространство голосов”, в котором может одновременно учитываться сложная структура разделяемых классов, что создает дополнительные возможности для повышения качества решений [21–23]. Особенностью этого метода является то, что разделяющие плоскости в нем проводятся вдоль границ разделяемых классов.

Несмотря на внешние различия, оба этих метода в какой-то мере моделируют процесс рассуждения при принятии решений в сочетании с аргументацией, использующей ссылки на множество косвенных свидетельств в отсутствие ясного доказательного основания [20, 23].

Оба метода разрабатывались как:

- свободные от обычных требований о независимости первичных описаний,
- малокритичные в отношении соотношения объема выборки и количества характеристик, описывающих наблюдения в этой выборке,
- допускающие работу при значительной доле пропусков в описаниях,
- допускающие ошибки в описаниях,
- допускающие работу с сильно несбалансированными по объему классами распознаваемых объектов,
- ориентированные на работу с очень большим числом нечетко определяемых классов и использованием для их описания очень большого числа характеристик, в значительной мере субъективных,
- могущие быть использованными для анализа процессов.

В основу разработки этих методов была положена “идея кликовости”, в соответствии с которой одновременно должны удовлетворяться требования по составу подмножеств объектов и признаков (первичных описаний), на которых максимизируется некий критерий смысловой однородности состава клики. Клика понимается как группа сообщников, объединившихся для достижения своих целей [24].

При переходе к решению химических задач с использованием вышеуказанных методов были учтены следующие обстоятельства:

- При решении поставленных задач порождаются гипотезы о новом знании, относящемся к этим задачам, позволяя, в частности, выбирать наиболее информативные из исходных признаков, характеризующих распознаваемые объекты.
- Решающие правила указывают на “зону неопределенности” в “пространстве описаний” распознаваемых классов объектов.
- Множества решающих правил, порождаемых используемым алгоритмом машинного обучения при вариациях его параметров позволяют не только указать на пространственную локализацию области неопределенности, но и выделить ее как область повышенной концентрации разделяющих поверхностей в пространстве описаний, позволяя тем самым ввести локальную меру смысловых сходств с классами, отличную от обычно используемых, общих для всего пространства описаний и не зависящих от структуры классов в нем.
- Простейший вариант порождения такого рода смысловых метрик – голосование решающих правил.
- Если основываться на последнем, то появляется возможность создания “смысловых” процедур обучения распознаванию, более приближенных к смысловому разделению объектов на классы, чем широко известные и используемые. В частности, взяв за основу любой из известных методов кластеризации объектов в пространстве описаний и заменив в нем используемую обычно евклидову метрику на смысловую, мы сразу же получаем соответствующий вариант “смысловой кластеризации” и как следствие - повышение качества получаемых результатов. В методе ФП эта же идея используется для формирования “потенциала” сходства

с классами при использовании кусочно-линейных решающих правил [23], а в синдромном анализе потенциал сходства реализуется голосованием “симптомов” [20].

- Наконец, последнее в данном случае основание развиваемого здесь подхода к решению задач прецедентного сходства заключается в том, что СА, формируя синдром для отделения какого-то подкласса от остальных рассматриваемых в задаче объектов, тождественен некому смысловому кластеру-понятию [19].

Основываясь на сказанном и исходя из природы задач прогноза количественных свойств неорганических соединений, было решено, что:

1. Описания химических соединений в виде изолированных наборов физико-химических параметров компонентов сами по себе не могут в должной мере характеризовать свойства соединений как неких целостностей. Поэтому система описаний была усложнена введением новых признаков – неких функций от исходных, которые могли бы, хоть в какой-то мере улучшить эту ситуацию. На первом этапе в качестве простейших из такого рода функций было решено использовать попарные произведения значений однородных по физическому смыслу параметров химических элементов, образующих соединения. В дальнейшем этот выбор может быть расширен.

Единственным техническим недостатком такого рода операции для получения признаков, характеризующих неорганические соединения в целом, является степенной рост итоговой размерности.

2. Было решено, что признаковое пространство описаний следует сократить до приемлемых размеров, ориентируясь на возможности СА, совмещая это с некоторой оптимизацией его единственного в данном случае параметра – доли симптомного интервала в диапазоне размаха соответствующего ему признака.

Для каждого объекта выборки делается следующее. Сначала строится набор интервалов, каждый из которых включает значение соответствующего признака данного объекта. Число таких интервалов равно числу признаков. Далее, для каждого признака подсчитывается число объектов обучающей выборки, у которых значения соответствующих признаков попадают в соответствующий интервал. Чем меньше это число, тем уникальнее рассматриваемый объект с точки зрения данного признака.

В результате применения этой процедуры ко всем объектам, для каждого из них формируется набор чисел по всем признакам, после чего для каждого признака берется сумма всех таких чисел S , соответствующих рассматриваемому признаку:

$$S = \sum_{i=1}^N \sum_{j=1}^N s_i(v_j),$$

$$s_i(v) = \begin{cases} 1, & v \in [t_{i1}, t_{i2}]; \\ 0, & v \notin [t_{i1}, t_{i2}]; \end{cases}$$

где t_{i1}, t_{i2} – границы интервала i -го объекта; v_i – значение рассматриваемого признака i -го объекта, N – число объектов.

Отсортировав признаковый набор по полученным так суммарным значениям, получаем ранжированный список признаков, где признаки с меньшими значениями можно считать более информативными, чем признаки с большими значениями.

Фактически, так отбираются “симптомы”, характеризующие минимальным числом “ложных тревог”, а вместе с ними и наиболее информативные признаки. В качестве начального приближения для первых проб было решено ограничиться первыми 200-ми наиболее информативными признаками. В дальнейшем эта процедура может быть предметом самостоятельного исследования.

3. Предсказание по смыслу значения интересующей характеристики какого-то нового объекта из множества объектов должно осуществляться по ближайшим к нему прецедентам, когда кластер представляется именно этим объектом, а в качестве ближайших к нему прецедентов выбираются немногие (желательно один) близкие по смысловому сходству объекты в прецедентной базе. Эту эвристику назовем методом “один против всех”.

Кроме того, было решено начальную фазу исследований проводить с упрощенными версиями СА и ФП, чтобы иметь возможность оценить нижнюю границу этих подходов и сэкономить затраты времени и вычислительных ресурсов.

В отношении СА было решено, что можно попробовать начать с простейшего его варианта с формированием только лишь одноуровневых синдромов без учета возможных корреляций в исходных признаках и большого количества ложных тревог, а в отношении ФП было решено ограничиться неглубокими решающими правилами, допускающими достаточно большое число ложных тревог.

При этом исходили из того, что в результате голосования в каждом из этих случаев ложные тревоги в силу их хаотичности будут накапливать меньше голосов, позволяя отбирать в каждом случае лишь один ближайший прецедент (случаи равенства получаемых таким образом расстояний до одного и более, чем одного прецедента, предполагались редкими, что в дальнейшем подтвердилось), обеспечивая тем самым требования базовой эвристики.

3. ТЕСТИРОВАНИЕ РАЗРАБОТАННЫХ МЕТОДОВ НА ПРИМЕРЕ РЕШЕНИЯ ЗАДАЧ ПРОГНОЗА ПАРАМЕТРОВ КРИСТАЛЛИЧЕСКОЙ РЕШЕТКИ МЕЛИЛИТОВ

Прогнозы параметров кристаллической решетки соединений представляют большой интерес как для химических, так и материаловедческих исследований. Методы машинного обучения позволили оценить значения таких параметров с достаточно высокой точностью [7–15]. В большинстве расчетов использовались методы обучения нейронных сетей и опорных векторов [7–15], а также обучение случайного леса [12–14].

Рассматриваемые в данной работе методы были применены для решения задачи оценки параметров решетки мелилитов двух составов: $A_2^{+2}B_2^{+3}C^{+4}O_7$ и $A^{+2}B_2^{+2}C_2^{+4}O_7$, где А и В – разные химические элементы, а С – Si или Ge. Для соединений каждого из двух составов оценивались значения двух параметров решетки a и c . Таким образом, решались четыре задачи в зависимости от типа параметров и состава соединений: (1) для параметра a соединения состава $A_2^{+2}B_2^{+3}C^{+4}O_7$, (2) для параметра c соединения состава $A_2^{+2}B_2^{+3}C^{+4}O_7$, (3) для параметра a соединения состава $A^{+2}B_2^{+2}C_2^{+4}O_7$ и (4) для параметра c соединения состава $A^{+2}B_2^{+2}C_2^{+4}O_7$.

В качестве исходных свойств компонентов были выбраны свойства элементов и простых оксидов. В результате были сформированы четыре обучающие выборки для последующего компьютерного анализа. Для мелилитов состава $A_2^{+2}B_2^{+3}C^{+4}O_7$ выборки включали 37 примеров, а для $A^{+2}B_2^{+2}C_2^{+4}O_7$ – 28 примеров.

В связи с тем, что объемы обучающих выборок были весьма невелики, качество прогнозных моделей оценивалось только на обучающих данных в первую очередь критерием R^2 -score [25], приведенным в Таблице 1 для методов ФП и СА.

Для получения оценки качества возможного прогноза формируемыми моделями проводилась процедура скользящего контроля по отдельным объектам (leave-one-out) [26]. В данном случае это позволяло иметь максимально возможный объем обучающих выборок для каждого эксперимента обучение-экзамен. В результате каждого такого эксперимента получался прогноз по одному объекту обучающей выборки, а в итоге – прогноз по всем объектам исходной обучающей выборки. Полученные так значения сравнивались с истинными известными значениями целевого параметра путем применения выбранного критерия оценки качества моделей.

Таблица 1. Оценка моделей на обучающей выборке.

Задача	Объем выборки для обучения	ФП R^2 -score		Оценка	СА R^2 -score		Оценка
		n1	n2		n1	n2	
(1)	37	0.74	0.63	++	0.78	0.76	++
(2)	37	0.79	0.74	++	0.86	0.88	++
(3)	28	0.38	0.45	++	0.24	0.47	--
(4)	28	0.52	0.67	++	0.45	0.52	--

В столбцах n1 и n2 Таблицы 1 приведены значения критерия R^2 -score в случаях прогнозирования параметров кристаллической решетки (таргета) по одному и двум ближайшим прецедентам. Видно, что при вычислении предсказываемой величины неизвестного значения таргета по двум ближайшим интегральный результат оказывается несколько лучшим, чем в варианте с одним ближайшим. Однако анализ погрешностей предсказания и общие соображения показывают, что такого рода улучшения достигаются в основном за счет тех соединений, предсказываемые значения для которых лежат внутри диапазона изменений таргетов у прецедентов. В то же время для экстремальных по таргетам объектов погрешности оценивания заметно возрастают. Поэтому, ориентируясь на возможности поиска новых соединений с лучшими, чем у известных соединений свойствами, предпочтение следовало бы отдать варианту с одним ближайшим прецедентом.

Судя по приведенным в Таблице 1 оценкам, метод ФП в целом ведет себя несколько лучше, чем метод СА. В то же время явно видно, что оценки, формируемые обоими методами, сильно коррелированы, оказываясь лучшими для задач (1) и (2) и несколько худшими для задач (3) и (4). В последних у прогнозов методом СА появляются оценки “-”, характеризующие в этом случае сравнительно низкую степень доверия к формируемым прогнозам. Во всех остальных случаях, помеченных знаком “+”, значения “коэффициента детерминации” говорят о наличии заметной связи прогноза и прогнозируемого. Тем не менее, синхронность ухудшения обоих методов прогнозирования на задачах (3) и (4) представляется достаточно интересной и требующей дополнительного исследования, в частности, и в связи с тем, что метод СА был использован в этой задаче для отбора подпространства наиболее информативных признаков, в котором затем уже и строились модели для прогнозирования обоими методами.

Чтобы приблизиться к пониманию причин такого рода расслоения результатов, были построены графики для сопоставления прогнозов и истинных значений прогнозируемых величин для этих случаев, приведенные на Рис. 1–4, для ФП-метода, на Рис. 5–8 – для СА-метода, а также – соответствующие им графики погрешностей прогнозов, приведенные на Рис. 9i–12r и Рис. 13q–16x.

На Рис. 1–4 показаны результаты предсказания значений таргета в четырех рассматриваемых задачах на основе метода ФП. На Рис. 5–8 показаны результаты предсказания значений таргета в четырех рассматриваемых задачах на основе метода СА. По осям координат: по горизонтали – номера объектов, упорядоченных по таргету прецедентных данных, по вертикали – величины таргетов для рассматриваемой задачи.

figuremetafigure

В парах Рис. 9i–12r на рисунках а представлены погрешности предсказания значений таргета в четырех рассматриваемых задачах на основе метода ФП. По осям координат: по горизонтали на рисунках а – номера объектов, упорядоченных по таргету прецедентных данных, а по вертикали – величина погрешности предсказания. На рисунках b – гистограммы модуля погрешности. На гистограммах: по горизонтали – абсолютная величина погрешности, а по вертикали – число объектов, соответствующих интересующей погрешности.

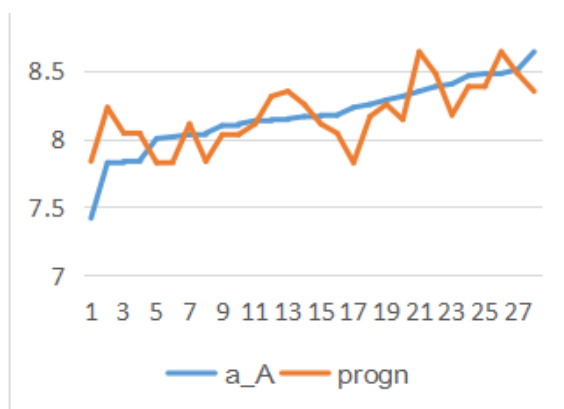


Рис. 1. Таргет-прогноз в ФП-задаче (3); таргет – голубая линия, прогноз – оранжевая.

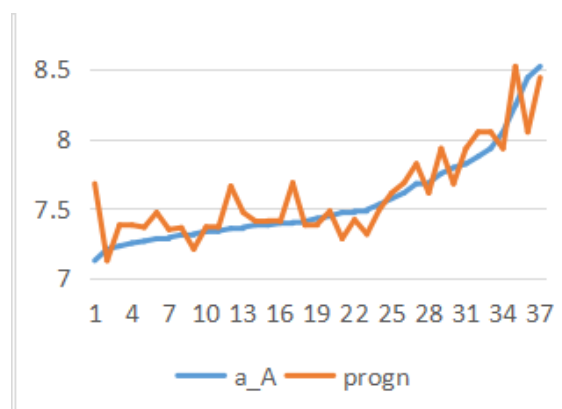


Рис. 2. Таргет-прогноз в ФП-задаче (1); таргет – голубая линия, прогноз – оранжевая.

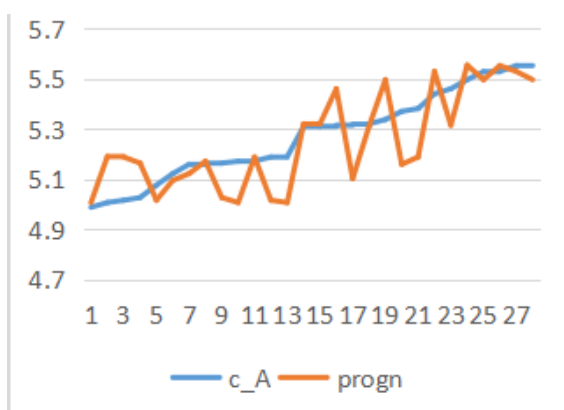


Рис. 3. Таргет-прогноз в ФП-задаче (4); таргет – голубая линия, прогноз – оранжевая.

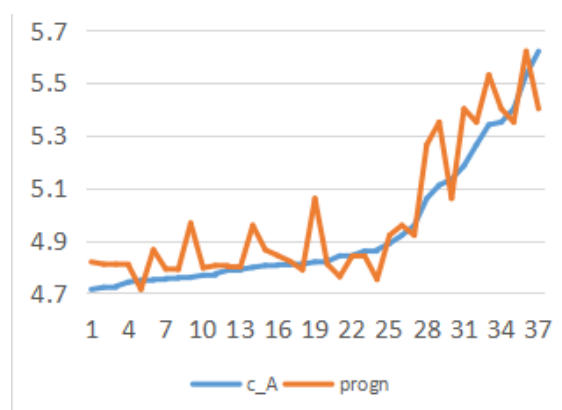


Рис. 4. Таргет-прогноз в ФП-задаче (2); таргет – голубая линия, прогноз – оранжевая.

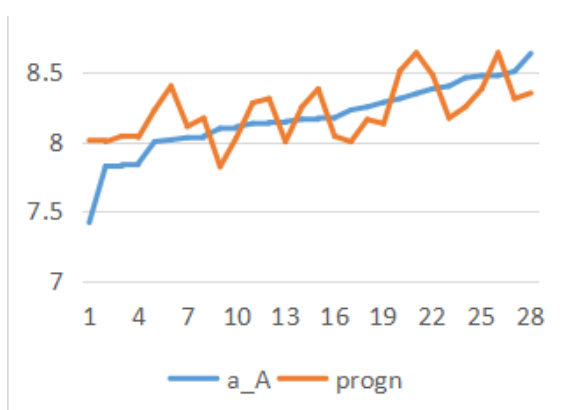


Рис. 5. Таргет-прогноз в СА-задаче (3); таргет – голубая линия, прогноз – оранжевая.

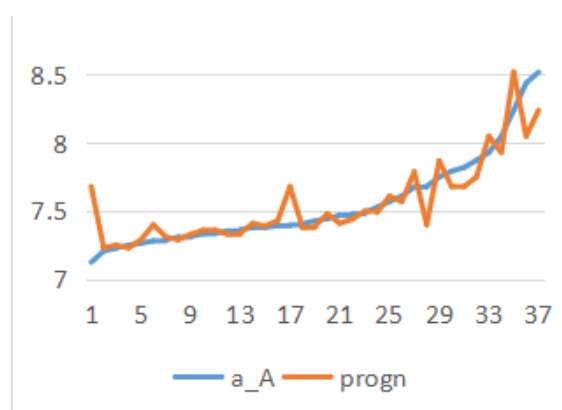


Рис. 6. Таргет-прогноз в СА-задаче (1); таргет – голубая линия, прогноз – оранжевая.

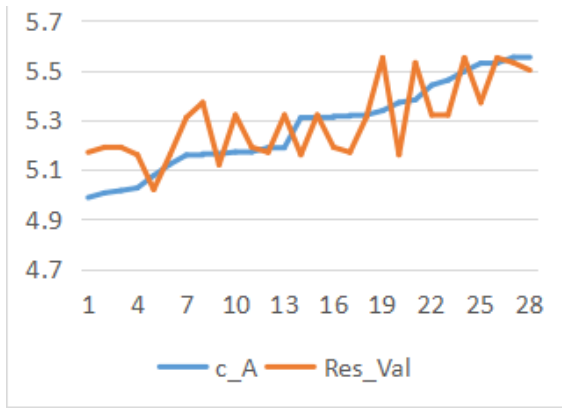


Рис. 7. Таргет-прогноз в СА-задаче (4); таргет – голубая линия, прогноз – оранжевая.

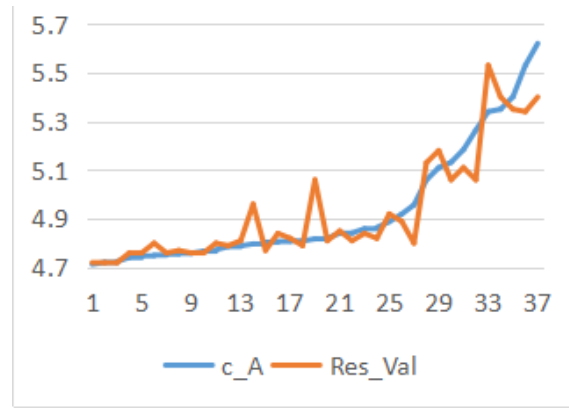


Рис. 8. Таргет-прогноз в СА-задаче (2); таргет – голубая линия, прогноз – оранжевая.

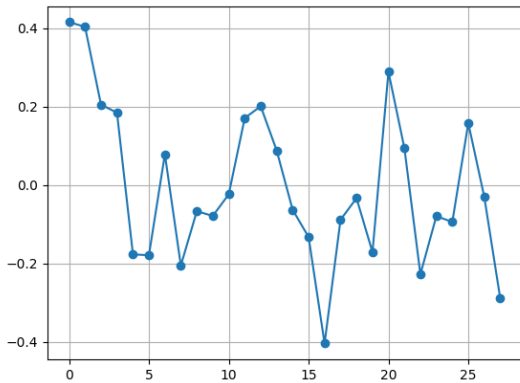


Рис. 9i. Графики погрешностей прогнозов (метод ФП, задача (3)).

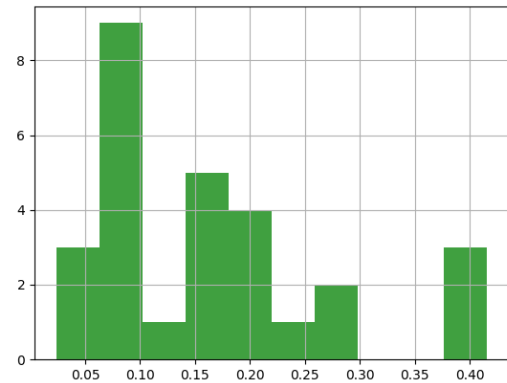


Рис. 9j. Гистограммы модуля погрешности прогнозов (метод ФП, задача (3)).

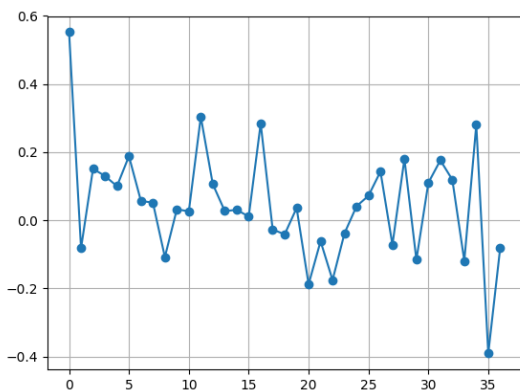


Рис. 10к. Графики погрешностей прогнозов (метод ФП, задача (1)).

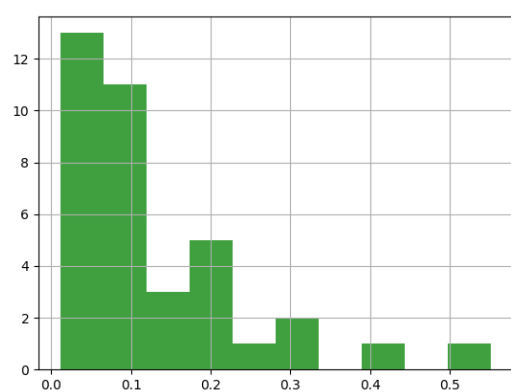


Рис. 10l. Гистограммы модуля погрешности прогнозов (метод ФП, задача (1)).

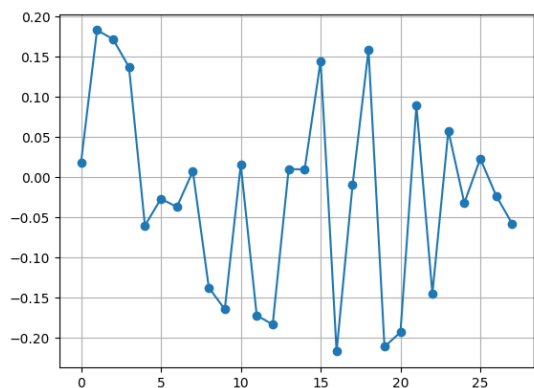


Рис. 11м. Графики погрешностей прогнозов (метод ФП, задача (4)).

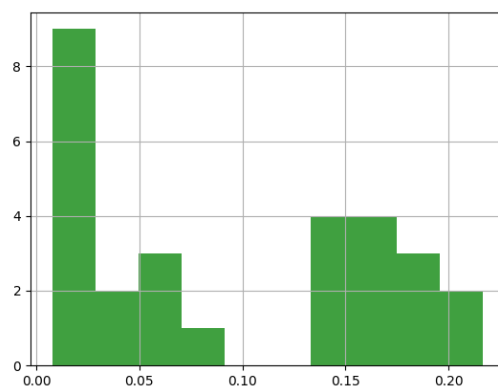


Рис. 11п. Гистограммы модуля погрешности прогнозов (метод ФП, задача (4)).

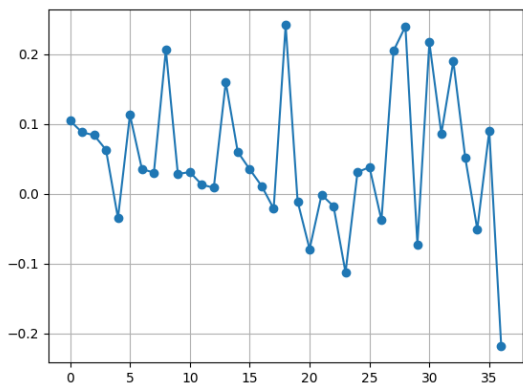


Рис. 12о. Графики погрешностей прогнозов (метод ФП, задача (2)).

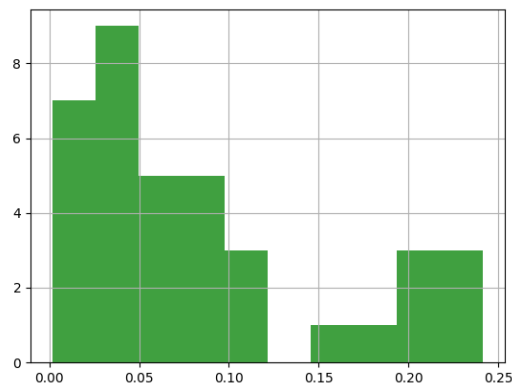


Рис. 12р. Гистограммы модуля погрешности прогнозов (метод ФП, задача (2)).

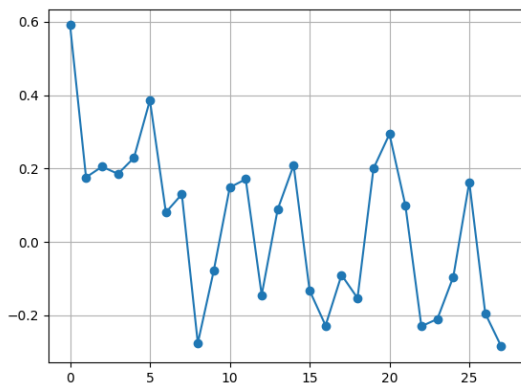


Рис. 13г. Графики погрешностей прогнозов (метод СА, задача (3)).

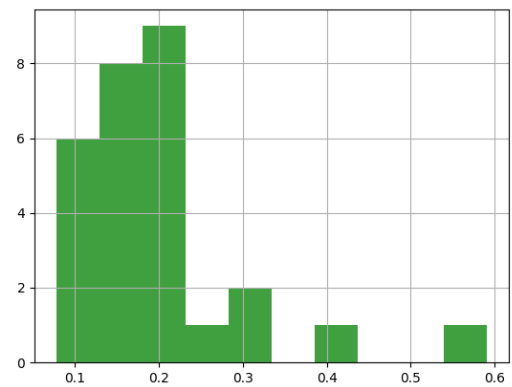


Рис. 13г. Гистограммы модуля погрешности прогнозов (метод СА, задача (3)).

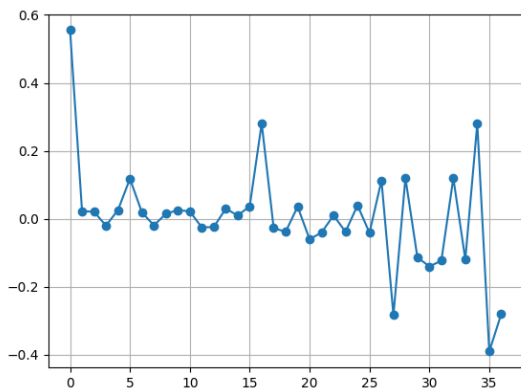


Рис. 14s. Графики погрешностей прогнозов (метод СА, задача (1)).

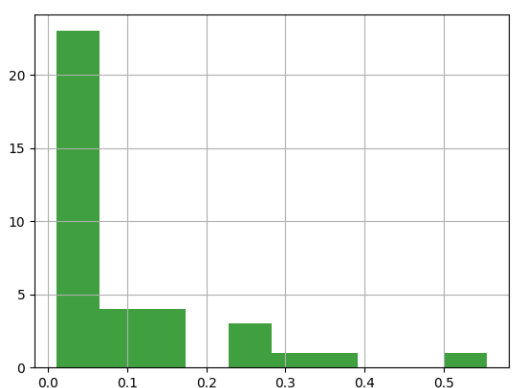


Рис. 14t. Гистограммы модуля погрешности прогнозов (метод СА, задача (1)).

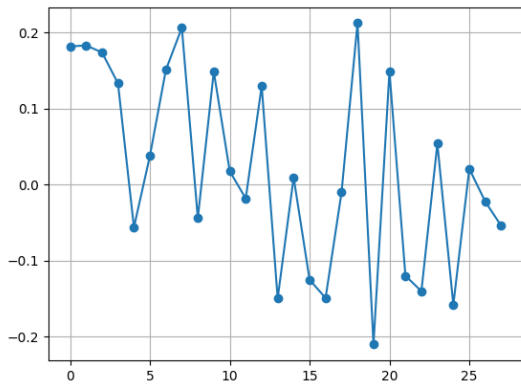


Рис. 15а. Графики погрешностей прогнозов (метод СА, задача (4)).

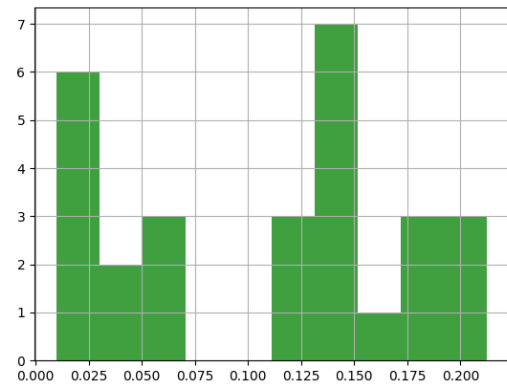


Рис. 15в. Гистограммы модуля погрешности прогнозов (метод СА, задача (4)).

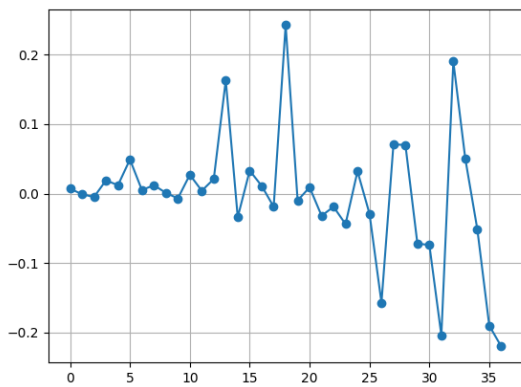


Рис. 16а. Графики погрешностей прогнозов (метод СА, задача (2)).

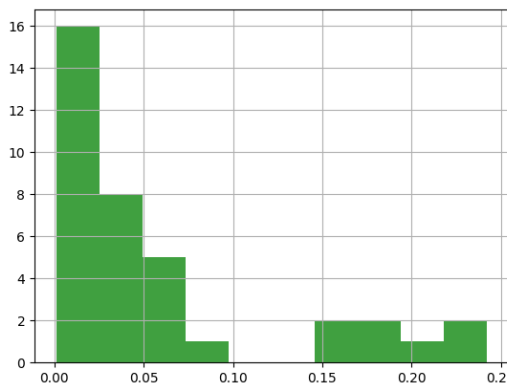


Рис. 16б. Гистограммы модуля погрешности прогнозов (метод СА, задача (2)).

В парах Рис. 13q–16x на рисунках а представлены погрешности предсказания значений таргета в четырех рассматриваемых задачах на основе метода СА. По осям координат: по горизонтали на рисунках а – номера объектов, упорядоченных по таргету прецедентных данных, а по вертикали – величина погрешности предсказания. На рисунках б – гистограммы модуля погрешности. На гистограммах: по горизонтали – абсолютная величина погрешности, а по вертикали – число объектов, соответствующих интересующей погрешности.

Из анализа этих графиков следует, что на материале формирования моделей:

1. В целом тренд прогноза соответствует требуемому при наличии сравнительно симметричных квазипериодических отклонений от него, укладывающихся в основном в диапазон ± 0.2 по таргету. В качестве доказательства этого были рассмотрены варианты сглаживания получаемых оценок таргета скользящим окном с шириной 3 и 5 отсчетов, интегральные оценки качества которых – R^2 -score приведены в Таблице 2. Видно, сколь сильно качество повышается даже при минимальном сглаживании. Однако это важно только для того, чтобы подчеркнуть то существенное обстоятельство, что рассматриваемые здесь подходы к моделированию таргетов действительно ориентированы на выделение трендовой компоненты даже при сильном ее квазипериодическом зашумлении, в природе которого, необходимо еще разбираться.
2. Хотя объем обучающей выборки очень мал, в погрешностях явно обнаруживаются три группы: с малым – порядка ± 0.1 разбросом, со средним – порядка ± 0.2 разбросом и малочисленная группа больших отклонений в диапазоне порядка ± 0.4 в основном на краях диапазона значений таргета.
3. Наличие квазипериодичности и расслоение погрешностей на три группы, судя по всему, связано с периодичностью свойств химических элементов, образующих исследуемые соединения. Это требует содержательного анализа информации о соединениях, попадающих в эти группы.
4. По-видимому, в ближайшей перспективе будет возможно установить формальные свойства этих трех групп, чтобы затем этим можно было бы воспользоваться для повышения точности прогнозирования параметров новых соединений.

Таблица 2. Выделение тренда таргета в результате подавления квазипериодических пульсаций прогнозов, формируемых ФП- и СА-моделями относительно таргетов обучающих данных.

Задача	ФП R^2 -score			СА R^2 -score		
	окно 1	окно 3	окно 5	окно 1	окно 3	окно 5
(1)	0.74	0.89	0.91	0.78	0.90	0.93
(2)	0.79	0.90	0.90	0.86	0.93	0.95
(3)	0.38	0.55	0.65	0.24	0.49	0.59
(4)	0.52	0.81	0.89	0.45	0.76	0.78

В Таблице 2, построенной для упорядоченных по возрастанию таргетов в решаемых задачах, представлены результаты сглаживания прогноза скользящими окнами с шириной 1, 3 и 5 отсчетов. Видно, что, начиная с окна шириной 3 отсчета, качество сглаженных прогнозных кривых заметно увеличивается, причем оно продолжает увеличиваться и при дальнейшем расширении окна до 5 отсчетов. Все это можно рассматривать в качестве объективного подтверждения того, что рассматриваемые здесь методы прогнозирования таргетов дают прогноз, тренд которого весьма близок к прогнозируемым величинам таргетов. К сожалению, это обстоятельство само по себе еще не указывает на то, каким способом эту особенность можно было бы использовать при прогнозировании, где такого рода операция выделения трендов не может

использоваться. Однако выявление факта такого рода позволяет прицельно заняться изучением свойств объектов, ориентируясь на особенности характера погрешностей прогнозирования, которые потенциально открывают возможности повышения точности прогнозирования. В дальнейшем мы планируем провести такого рода исследование.

В Таблице 3 представлены результаты оценки таргетов в рассматриваемых задачах как полусуммы оценок, независимо формируемых методами ФП и СА. Видно, что полусумма прогнозов во всех четырех задачах, как и следовало ожидать, приводит к улучшению качества прогнозирования, оцениваемому критерием R^2 -score. Причем, наибольшее улучшение происходит там, где каждый из применяемых нами методов сам по себе дает наихудшие результаты прогнозирования.

Таблица 3. Сопоставление оценок качества прогнозирования таргетов по методам ФП, СА и их полусумме при прогнозировании по одному ближайшему прецеденту (n1).

Задача	Объем выборки для обучения	ФП R^2 -score	СА R^2 -score	(ФП+СА)/2 R^2 -score
		n1	n1	n1
(1)	37	0.74	0.78	0.78
(2)	37	0.79	0.86	0.88
(3)	28	0.38	0.24	0.47
(4)	28	0.52	0.45	0.87

Разработанные методы были применены для оценки неизвестных значений параметров кристаллической решетки еще не полученных мелилитов. Результаты прогноза приведены в Таблицах 4 и 5.

Таблица 4: Результаты прогноза параметров кристаллической решетки еще не изученных мелилитов состава $A_2^{+2}B_2^{+3}C^{+4}O_7$.

Состав	a, ФП	a, СА	c, ФП	c, СА
Mg ₂ B ₂ SiO ₇	7.1248	8.52	4.8177	5.53
Sr ₂ B ₂ SiO ₇	7.93	7.1248	5.35	4.8177
Ba ₂ B ₂ SiO ₇	7.93	7.82	5.35	5.264
Sr ₂ Cr ₂ SiO ₇	8.44	8.52	5.62	5.53
Ba ₂ Cr ₂ SiO ₇	8.44	8.52	5.62	5.53
Sr ₂ Fe ₂ SiO ₇	8.48	8.24	5.575	5.34
Sr ₂ Ga ₂ SiO ₇	7.9345	7.82	5.3325	5.264
Ca ₂ Y ₂ SiO ₇	7.4264	7.291	4.8289	4.758
Sr ₂ Y ₂ SiO ₇	7.8673	7.315	5.0947	4.748
Be ₂ La ₂ SiO ₇	7.683	7.282	4.957	4.751
Be ₂ Ce ₂ SiO ₇	7.47	7.47	4.864	4.864
Be ₂ Pr ₂ SiO ₇	7.47	7.47	4.864	4.864
Ca ₂ Pr ₂ SiO ₇	7.4137	7.229	4.8293	4.721
Be ₂ Pm ₂ SiO ₇	7.47	7.47	4.864	4.864
Ca ₂ Pm ₂ SiO ₇	7.4137	7.207	4.8293	4.713
Sr ₂ Pm ₂ SiO ₇	8.52	8.24	5.53	5.34
Mg ₂ Sm ₂ SiO ₇	7.38	7.38	4.809	4.809
Ca ₂ Sm ₂ SiO ₇	7.38	7.38	4.809	4.809
Sr ₂ Sm ₂ SiO ₇	7.38	7.38	4.809	4.809
Mg ₂ Eu ₂ SiO ₇	7.38	7.38	4.809	4.809

$\text{Ca}_2\text{Eu}_2\text{SiO}_7$	7.38	7.3395	4.809	4.776
$\text{Sr}_2\text{Eu}_2\text{SiO}_7$	7.38	7.3935	4.809	4.8145
$\text{Mg}_2\text{Gd}_2\text{SiO}_7$	7.4137	7.338	4.8293	4.773
$\text{Ca}_2\text{Gd}_2\text{SiO}_7$	7.4646	7.309	4.8484	4.767
$\text{Sr}_2\text{Gd}_2\text{SiO}_7$	7.6405	8.24	4.962	5.34
$\text{Ca}_2\text{Tb}_2\text{SiO}_7$	7.4646	7.309	4.8484	4.767
$\text{Sr}_2\text{Tb}_2\text{SiO}_7$	7.6405	7.677	4.962	5.0594
$\text{Ca}_2\text{Dy}_2\text{SiO}_7$	7.541	7.677	4.877	5.0594
$\text{Sr}_2\text{Dy}_2\text{SiO}_7$	7.8673	7.8065	5.0947	5.198
$\text{Mg}_2\text{Ho}_2\text{SiO}_7$	7.4385	7.25	4.842	4.725
$\text{Ca}_2\text{Ho}_2\text{SiO}_7$	7.541	7.677	4.877	5.0594
$\text{Sr}_2\text{Ho}_2\text{SiO}_7$	7.867333	7.25	5.094667	4.725
$\text{Mg}_2\text{Er}_2\text{SiO}_7$	7.541	7.25	4.877	4.725
$\text{Ca}_2\text{Er}_2\text{SiO}_7$	7.541	7.229	4.877	4.721
$\text{Sr}_2\text{Er}_2\text{SiO}_7$	7.8673	7.677	5.0947	5.0594
$\text{Mg}_2\text{Tm}_2\text{SiO}_7$	8.44	7.229	5.62	4.721
$\text{Ca}_2\text{Tm}_2\text{SiO}_7$	8.44	8.44	5.62	5.62
$\text{Sr}_2\text{Tm}_2\text{SiO}_7$	8.44	7.229	5.62	4.721
$\text{Ca}_2\text{Yb}_2\text{SiO}_7$	8.44	7.611	5.62	4.919
$\text{Sr}_2\text{Yb}_2\text{SiO}_7$	8.44	7.611	5.62	4.919
$\text{Mg}_2\text{Lu}_2\text{SiO}_7$	7.647	7.207	4.938	4.713
$\text{Ca}_2\text{Lu}_2\text{SiO}_7$	7.541	7.57	4.877	4.888
$\text{Sr}_2\text{Lu}_2\text{SiO}_7$	7.8673	7.57	5.0947	4.888
$\text{Ca}_2\text{Bi}_2\text{SiO}_7$	8.44	8.049	5.62	5.401
$\text{Sr}_2\text{Bi}_2\text{SiO}_7$	8.44	8.049	5.62	5.401
$\text{Mg}_2\text{B}_2\text{GeO}_7$	7.1248	8.52	4.8177	5.53
$\text{Ca}_2\text{B}_2\text{GeO}_7$	7.84	7.75	5.23	5.11
$\text{Sr}_2\text{B}_2\text{GeO}_7$	7.93	7.93	5.35	5.35
$\text{Ba}_2\text{B}_2\text{GeO}_7$	7.93	7.93	5.35	5.35
$\text{Ba}_2\text{Al}_2\text{GeO}_7$	7.93	8.44	5.35	5.62
$\text{Ca}_2\text{Cr}_2\text{GeO}_7$	8.52	8.52	5.53	5.53
$\text{Sr}_2\text{Cr}_2\text{GeO}_7$	8.52	8.52	5.53	5.53
$\text{Ba}_2\text{Cr}_2\text{GeO}_7$	8.44	8.52	5.62	5.53
$\text{Ca}_2\text{Fe}_2\text{GeO}_7$	8.24	8.24	5.34	5.34
$\text{Be}_2\text{Ga}_2\text{GeO}_7$	7.1248	8.24	4.8177	5.34
$\text{Mg}_2\text{Ga}_2\text{GeO}_7$	7.873	8.24	5.184	5.34
$\text{Be}_2\text{Ce}_2\text{GeO}_7$	7.683	7.611	4.957	4.919
$\text{Be}_2\text{Pm}_2\text{GeO}_7$	7.5905	7.57	4.9035	4.888
$\text{Ca}_2\text{Sm}_2\text{GeO}_7$	8.44	7.49	5.62	4.841
$\text{Sr}_2\text{Sm}_2\text{GeO}_7$	8.44	7.49	5.62	4.841
$\text{Mg}_2\text{Eu}_2\text{GeO}_7$	8.44	7.49	5.62	4.841
$\text{Ca}_2\text{Eu}_2\text{GeO}_7$	8.44	7.484	5.62	4.836
$\text{Sr}_2\text{Eu}_2\text{GeO}_7$	8.44	7.4993	5.62	4.8473
$\text{Ca}_2\text{Gd}_2\text{GeO}_7$	7.541	7.428	4.877	4.809
$\text{Sr}_2\text{Gd}_2\text{GeO}_7$	8.52	8.24	5.53	5.34
$\text{Ca}_2\text{Tb}_2\text{GeO}_7$	7.541	8.24	4.877	5.34
$\text{Sr}_2\text{Tb}_2\text{GeO}_7$	8.52	7.75	5.53	5.11
$\text{Mg}_2\text{Dy}_2\text{GeO}_7$	7.541	7.428	4.877	4.809
$\text{Ca}_2\text{Dy}_2\text{GeO}_7$	7.541	7.75	4.877	5.11

Sr ₂ Dy ₂ GeO ₇	8.52	7.793	5.53	5.132
Mg ₂ Ho ₂ GeO ₇	7.541	7.393	4.877	4.789
Ca ₂ Ho ₂ GeO ₇	7.541	7.75	4.877	5.11
Sr ₂ Ho ₂ GeO ₇	8.52	7.355	5.53	4.759
Mg ₂ Er ₂ GeO ₇	7.541	7.355	4.877	4.759
Ca ₂ Er ₂ GeO ₇	7.541	7.3445	4.877	4.757
Sr ₂ Er ₂ GeO ₇	8.52	8.24	4.877	4.789
Mg ₂ Tm ₂ GeO ₇	7.541	7.355	4.877	5.11
Ca ₂ Tm ₂ GeO ₇	8.44	7.8775	5.53	4.759
Sr ₂ Tm ₂ GeO ₇	8.44	7.334	4.877	4.759
Ba ₂ Tm ₂ GeO ₇	8.44	7.334	4.877	4.757
Mg ₂ Yb ₂ GeO ₇	8.44	7.334	5.62	4.755
Ca ₂ Yb ₂ GeO ₇	8.44	7.315	5.62	4.748
Sr ₂ Yb ₂ GeO ₇	8.44	7.315	5.62	4.748
Ba ₂ Yb ₂ GeO ₇	8.44	7.346	5.62	4.762
Mg ₂ Lu ₂ GeO ₇	7.683	7.315	4.957	4.748
Ca ₂ Lu ₂ GeO ₇	7.541	7.38	4.877	4.786
Sr ₂ Lu ₂ GeO ₇	8.52	7.355	5.53	4.759
Ba ₂ Lu ₂ GeO ₇	8.52	8.24	5.53	5.34
Mg ₂ Bi ₂ GeO ₇	8.44	8.44	5.62	5.62
Ca ₂ Bi ₂ GeO ₇	8.44	8.44	5.62	5.62
Sr ₂ Bi ₂ GeO ₇	8.44	7.93	5.62	5.35
Ba ₂ Bi ₂ GeO ₇	8.44	7.82	5.62	5.264

Таблица 5: Результаты прогноза параметров кристаллической решетки еще не изученных мелилитов состава $A_2^{+2}B_2^{+2}C^{+4}O_7$.

Состав	a, ФП	a, СА	c, ФП	c, СА
Ca ₂ MgGe ₂ O ₇	7.8348	7.8348	5.0067	5.0067
Ca ₂ MnSi ₂ O ₇	7.8444	7.419	5.122	4.988
Ca ₂ FeSi ₂ O ₇	7.8444	7.8444	5.0265	5.0265
Ca ₂ CdSi ₂ O ₇	7.824	7.824	5.0159	5.0159
Pb ₂ MnSi ₂ O ₇	8.0138	8.31	5.1711	5.37
Ba ₂ FeSi ₂ O ₇	8.3230	8.1709	5.4167	5.3374
Pb ₂ FeSi ₂ O ₇	8.0138	8.1709	5.496	5.3374
Pb ₂ ZnSi ₂ O ₇	8.0138	8.4769	5.1711	5.381
Sr ₂ BeSi ₂ O ₇	7.419	8.1315	4.988	5.1589
Ba ₂ BeSi ₂ O ₇	8.2069	8.31	4.988	5.37
Eu ₂ BeSi ₂ O ₇	8.0138	8.139	5.1711	5.1635
Eu ₂ FeSi ₂ O ₇	8.139	8.139	5.496	5.1635
Eu ₂ CoSi ₂ O ₇	8.1709	8.1709	5.496	5.3374
Pb ₂ CoSi ₂ O ₇	8.0138	8.1709	5.33355	5.3374
Sr ₂ NiSi ₂ O ₇	8.0294	8.0294	5.1639	5.1639
Eu ₂ NiSi ₂ O ₇	8.1709	8.1709	5.496	5.3374
Ca ₂ CuSi ₂ O ₇	7.824	8.096	5.0159	5.076
Eu ₂ CuSi ₂ O ₇	8.0138	8.46023	5.1711	5.4396
Pb ₂ CuSi ₂ O ₇	8.0138	8.505	5.206367	5.528
Eu ₂ ZnSi ₂ O ₇	8.0138	8.4769	5.1711	5.381

$\text{Eu}_2\text{CdSi}_2\text{O}_7$	8.4769	8.4769	5.1711	5.381
$\text{Eu}_2\text{MgGe}_2\text{O}_7$	8.0138	8.0138	5.1711	5.1711
$\text{Ca}_2\text{MnGe}_2\text{O}_7$	7.8444	8.2833	5.3148	5.3096
$\text{Eu}_2\text{MnGe}_2\text{O}_7$	8.475	8.139	5.1635	5.1635
$\text{Pb}_2\text{MnGe}_2\text{O}_7$	8.0138	8.505	5.1711	5.528
$\text{Ca}_2\text{FeGe}_2\text{O}_7$	8.1616	8.2833	5.0265	5.3096
$\text{Eu}_2\text{FeGe}_2\text{O}_7$	8.475	8.475	5.496	5.496
$\text{Pb}_2\text{FeGe}_2\text{O}_7$	8.0138	8.475	5.496	5.496
$\text{Ca}_2\text{CoGe}_2\text{O}_7$	8.1616	7.8444	5.0265	5.0265
$\text{Eu}_2\text{CoGe}_2\text{O}_7$	8.381	8.381	5.496	5.554
$\text{Pb}_2\text{CoGe}_2\text{O}_7$	8.4	8.475	5.496	5.496
$\text{Ca}_2\text{NiGe}_2\text{O}_7$	8.1616	7.8444	5.0265	5.0265
$\text{Sr}_2\text{NiGe}_2\text{O}_7$	8.1616	8.1616	5.3103	5.3103
$\text{Ba}_2\text{NiGe}_2\text{O}_7$	8.381	8.381	5.554	5.554
$\text{Eu}_2\text{NiGe}_2\text{O}_7$	8.4602	8.4602	5.496	5.4396
$\text{Pb}_2\text{NiGe}_2\text{O}_7$	8.4	8.381	5.525	5.554
$\text{Ca}_2\text{CuGe}_2\text{O}_7$	8.1448	8.1448	5.3189	5.3189
$\text{Sr}_2\text{CuGe}_2\text{O}_7$	8.1616	8.096	5.3189	5.076
$\text{Eu}_2\text{CuGe}_2\text{O}_7$	8.4602	8.4602	5.4396	5.4396
$\text{Pb}_2\text{CuGe}_2\text{O}_7$	8.4	8.4602	5.46	5.4396
$\text{Eu}_2\text{ZnGe}_2\text{O}_7$	8.346	8.346	5.1711	5.5522
$\text{Pb}_2\text{ZnGe}_2\text{O}_7$	8.0138	8.635	5.1711	5.529
$\text{Ca}_2\text{CdGe}_2\text{O}_7$	7.824	7.824	5.0159	5.0159
$\text{Sr}_2\text{CdGe}_2\text{O}_7$	8.1448	8.1448	5.3189	5.3189
$\text{Eu}_2\text{CdGe}_2\text{O}_7$	8.635	8.635	5.3501	5.529
$\text{Pb}_2\text{CdGe}_2\text{O}_7$	8.0138	8.635	5.1711	5.529

4. ЗАКЛЮЧЕНИЕ

Исследованы возможности “мягкого” голосования “кликковых” решателей и разработаны алгоритмы реализации этого метода в приложении к химическим задачам. В качестве основы для исследований были выбраны два варианта кликовых методов, разработанных нами прежде в связи с задачами диагностики и прогнозирования в медицине – “Синдромный анализ” и “Фрагмент-Потенциал”, обладающих широким диапазоном возможностей для решения такого рода задач. Эти методы рассматривались в специфическом варианте обучения распознаванию, когда решение о прогнозируемом объекте принимается в ориентации на ближайшие по смыслу прецеденты. В разрабатываемой нами парадигме смысловой аспект – хотя бы в качестве нулевого приближения – подчеркивается голосованием решателей разной природы, формируемых при решении задачи обучения распознаванию исследуемого объекта при сопоставлении его с прецедентным множеством. Ближайшими по смыслу при этом считаются те из прецедентов, которые как ложные тревоги при распознавании набирают наибольшее число голосов. Так образуется минимальная клика, усреднение отклика по которой позволяет сформировать прогноз отклика для прогнозируемого объекта в качестве нулевого приближения. Последнее послужило основой для модификации уже разработанных методов, связанной с тем, что усреднение по клике является наиболее грубой возможностью формирования прогнозируемого. Существенно большего можно ожидать от поиска закономерности связи прогнозируемого на объектах клики с координатами, в которых она представлена. В этом можно видеть некий вариант многомерной регрессионной задачи, осложняемой небольшим объемом выборки, со-

ставляющей клику, и высокой размерностью ее описаний. Развитие в этом направлении можно рассматривать в качестве перспективного направления дальнейших исследований.

Результаты свидетельствуют о потенциальной применимости рассмотренных модификаций методов СА и ФП к решению регрессионных задач. При этом подтверждаются некоторые характерные свойства этих методов: способность работать с малыми объемами обучающей выборки, объекты которой содержат большое число характеристик, но с возможными ошибками в описаниях. Эксперименты показывают, что лучшие оценки для рассматриваемого типа данных как правило получаются в вариантах с поиском одного ближайшего объекта. Для таких вариантов легко оценить, действительно ли отбирались самые ближайшие в смысле целевого параметра. По результатам экспериментов видно, что так происходит не всегда, что теоретически открывает потенциальные возможности улучшения метода в части оценки близости объектов.

Таким образом, на примере решения тестовой задачи прогнозирования параметров кристаллической решетки соединений со структурой мелилита разного состава была продемонстрирована правильность предложенного подхода. Полученные результаты свидетельствуют о перспективности использования разработанных методов для оценки количественных свойств неорганических веществ. По оценкам качества результатов рассмотренные методы сопоставимы с известными методами при решении подобных задач и могут рассматриваться в качестве альтернативы им или дополнения.

СПИСОК ЛИТЕРАТУРЫ

1. Киселева Н.Н. *Компьютерное конструирование неорганических соединений. Использование баз данных и методов искусственного интеллекта*. М.: Наука, 2005.
2. Chen N.Y., Lu W.C., Yang J., Li G.Z. *Support vector machine in chemistry*. Singapore: World Scientific Publishing Co. Pte. Ltd., 2004.
3. Бурханов Г.С., Киселева Н.Н. Прогнозирование интерметаллических соединений. *Успехи химии*, 2009, т. 78, № 6, стр. 615–634.
4. Ghiringhelli L.M., Vybiral J., Levchenko S.V., Draxl C., Scheffler M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.*, 2015, vol. 114, no. 10, pp. 105503/1–5.
5. Mueller T., Kusne A.G., Ramprasad R. Machine Learning in Materials Science. Recent Progress and Emerging Applications. *Rev. Comp. Chem.*, 2016, vol. 29, pp. 186–273.
6. Ryan K., Lengyel J., Shatruk M. Crystal Structure Prediction via Deep Learning. *J. Amer. Chem. Soc.*, 2018, vol. 140, no. 32, pp. 10158–10168.
7. Aleksovska S., Dimitrovska S., Kuzmanovski I. Crystal structure prediction in orthorhombic ABO_3 perovskites by multiple linear regression and artificial neural networks. *Acta Chim. Sloven.*, 2007, vol. 54, no. 3, pp. 574–582.
8. Javed S.G., Khan A., Majid A., Mirza A.M., Bashir J. Lattice constant prediction of orthorhombic ABO_3 perovskites using support vector machines. *Comput. Mater. Sci.*, 2007, vol. 39, no. 3, pp. 627–634.
9. Khan A., Javed S.G. Predicting regularities in lattice constants of $GdFeO_3$ -type perovskites. *Acta Crystallogr.*, 2008, vol. B64, no. 1, pp. 120–122.
10. Li C., Thing Y., Zeng Y., Wang C., Wu P. Prediction of lattice constant in perovskites of $GdFeO_3$ structure. *J. Phys. Chem. Solids.*, 2003, vol. 64, no. 11, pp. 2147–2156.
11. Majid A., Khan A., Javed G., Mirza A.M. Lattice constant prediction of cubic and monoclinic perovskites using neural networks and support vector regression. *Comp. Mater. Sci.*, 2010, vol. 50, no. 2, pp. 363–372.
12. Dimitrovska S., Aleksovska S., Kuzmanovski I. Prediction of the unit cell edge length of cubic $A_2^{2+}BB'O_6$ perovskites by multiple linear regression and artificial neural networks. *Central Eur. J. Chem.*, 2005, vol. 3, no. 1, pp. 198–215.

13. Majid A., Khan A., Choi T.-S. Predicting lattice constant of complex cubic perovskites using computational intelligence. *Comp. Mater. Sci.*, 2011, vol. 50, no. 6, pp. 1879–1888.
14. Kockan U., Evis Z. Prediction of hexagonal lattice parameters of various apatites by artificial neural network. *J. Appl. Cryst.*, 2010, vol. 43, no. 4, pp. 769–779.
15. Zeng Y., Chua S.J., Wu P. On the Prediction of Ternary Semiconductor Properties by Artificial Intelligence Methods. *Chem. Mater.*, 2002, vol. 14, no. 7, pp. 2989–2998.
16. Becker P., Bohaty L., Liebertz J., Kleebe H.-J., Muller M., Eichler H.J., Rhee H., Hanuza J., Kaminskii A.A. Non-centrosymmetric tetragonal $\text{Sr}_2\text{ZnGe}_2\text{O}_7$ – a novel melilite-type nonlinear-laser crystal offering $\chi^{(2)}$ -, $\chi^{(3)}$ -, and cascaded $\chi^{(3)} \leftrightarrow \chi^{(2)}$ -interactions. *Laser Phys. Lett.*, 2010, vol. 7, no. 5, pp. 367–377.
17. Shen C., Zhang S., Cao W., Cong H., Yu H., Wang J., Zhang H. Thermal and electromechanical properties of melilite-type piezoelectric single crystals. *J. Appl. Phys.*, 2015, vol. 117, no. 6, pp. 064106/1–7.
18. Переверзев-Орлов В.С., Трунов В.Г. Синдромный анализ: новые вызовы. *Информационные Процессы*, 2008, том 8, № 4, стр. 235–239.
19. Vitushko M., Gurov N., Pereverzev-Orlov V. A Syndrome As a Tool for Presenting Concepts. *Pattern Recognition and Image Analysis*, 2002, vol. 12, no. 2, pp. 194–202.
20. Переверзев-Орлов В. *Советчик специалиста: опыт создания партнерской системы*. М: Наука, 1990.
21. Vaschenko E., Vitushko M., Pereverzev-Orlov V. Potentials of Learning on the Basis of Partner System. *Pattern Recognition and Image Analysis*, 2004, vol. 14, no. 1, pp.84–91.
22. Vashchenko E.A., Vitushko M.A., Gurov N.D., Pereverzev-Orlov V.S., Stenina I.I. Knowledge and Data Cooperation. *Pattern Recognition and Image Analysis*, 1998, vol.8, no. 2, pp. 25–41.
23. Левит В.Е., Переверзев-Орлов В.С. *Структура и поле данных при распознавании образов*. М: Наука, 1984
24. Ожегов С.И., Шведова Н.Ю. *Толковый словарь русского языка: 80 000 слов и фразеологических выражений. 4-е изд.* М.: Азбуковник, 1997.
25. Коэффициент детерминации.
URL: http://www.machinelearning.ru/wiki/index.php?title=Коэффициент_детерминации
26. Скользящий контроль.
URL: http://www.machinelearning.ru/wiki/index.php?title=Скользящий_контроль

On the possibility of predicting the parameter values of multicomponent inorganic compounds

Vashchenko E.A., Vitushko M.A., Dudarev V.A., Kiselyova N.N., Pereverzev-Orlov V.S.

A model for predicting the properties of inorganic substances based on the “soft” vote of “clique” solvers is proposed. Algorithms for its implementation in the application to chemical problems are developed. As a basis for research, we chose two variants of clique methods (“Syndrome Analysis” and “Fragment-Potential”) that we developed earlier in connection with the tasks of diagnosis and prognosis in medicine. These methods were considered in a specific version of learning recognition when the decision about the predicted object is made in the orientation on precedents that are closest in meaning. The methods are implemented by voting solvers of different nature, formed in solving the problem of learning to recognize the predicted object by comparing it with the training samples. The prospects of the proposed approach were demonstrated by solving the problem of predicting the crystal lattice parameters of compounds which are not yet obtained. Compounds with a crystalline structure of melilite of various compositions were considered.

KEYWORDS: machine learning, clique methods, syndrome analysis, fragment-potential, parameters of inorganic compounds, lattice parameters, melilite.