

УДК 577.29

ПРАКТИКА И ПОТЕНЦИАЛ РАЗВИТИЯ ИНТЕРПРЕТИРУЕМОСТИ МОДЕЛЕЙ КОЛИЧЕСТВЕННОГО ОТНОШЕНИЯ «СТРУКТУРА-АКТИВНОСТЬ» (QSAR)

Аддылова Ф. Т., Давронов Р. Р., Жамилов У. У., Каюмов О. А.
fatadilova@gmail.com; rifqat@gmail.com, jamilovu@gmail.com,
oybekonlineuz@mail.ru

Институт математики им. В. И. Романовского АНРУз, г. Ташкент, ул. Мирзо Улугбека, 81

Достоинство модели QSAR состоит в понимании механизма изучаемого действия, поэтому проблема интерпретируемости модели приобретает особую актуальность. Сегодня есть тысячи молекулярных дескрипторов, доступных для построения моделей QSAR, но дать физико-химическую интерпретацию для многих из них часто невозможно. До недавнего времени применялась схема интерпретации «модель-дескриптор-структура», зависящая от набора дескрипторов и методов машинного обучения, что затрудняло интерпретацию моделей, построенных на новейших методах машинного обучения. Выходом стала новая парадигма «модель-структура», обеспечивающая понимание смысла любой модели QSAR. В статье анализируются плюсы и минусы подходов к интерпретации, и в вычислительном эксперименте обосновывается эффективность нового универсального подхода к интерпретации моделей QSAR.

Ключевые слова: модель «структура-активность», интерпретируемость модели, машинное обучение

Цитирование: Аддылова Ф. Т., Давронов Р. Р., Жамилов У. У., Каюмов О. А. Практика и потенциал развития интерпретируемости моделей количественного отношения «структура-активность» (QSAR) // Проблемы вычислительной и прикладной математики. — 2018. — № 5(17). — С. 7–26.

1 Введение

Одним из наиболее популярных подходов к моделированию физических и биологических свойств химических веществ является моделирование отношения «структура–активность» или QSAR моделирование, поскольку модели QSAR используют для оценки потенциального воздействия химических веществ, материалов и наноматериалов на здоровье человека и экологические системы. Эта область вычислительной химии прошла 60–летний период успехов и неудач, но в 2014 году была опубликована статья «QSAR Modeling: Where Have You Been? Where Are You Going To?» авторами которой являлись 20 руководителей научных школ в этой области. В статье представлена история QSAR, понятия, важные этапы и тренды развития, нерешенные и насущные проблемы, новые приложения [1]. В число насущных проблем вошла интерпретируемость модели QSAR.

Одно из главных достоинств модели QSAR, - понимание механизма изучаемого действия. Несмотря на то, что сегодня есть тысячи молекулярных дескрипторов, доступных для построения моделей QSAR, дать физико-химическую интерпретацию для многих из них часто невозможно. Считалось, что понимание решения модели обусловлено интерпретируемостью выбранных дескрипторов и метода машинного обучения. Интерпретация через дескрипторы была общим ограничением, начиная с

первых моделей QSAR, которые строились только на дескрипторах, имеющих четкий структурный или физико-химический смысл. В то же время понятно, что использование не-интерпретируемых или трудно интерпретируемых дескрипторов может улучшить прогнозную способность модели [2, 3].

С другой стороны, модели, построенные современными методами машинного обучения, такими как SVM, нейронные сети (NN), глубокие нейронные сети (DNN), Random Forest (RF), не интерпретируются. Бытует мнение о том, что есть компромисс между прогнозируемостью и интерпретируемостью моделей QSAR [4–6]. Однако, недавние работы изменили парадигму интерпретации моделей, сделав все модели интерпретируемыми, независимо от дескрипторов или методов машинного обучения [7–11]. Целью данной статьи является анализ подходов к интерпретации QSAR моделей и обоснование эффективности новой, универсальной парадигмы интерпретации в вычислительном эксперименте.

2 Парадигмы интерпретации моделей QSAR

Очевидно, что следует интерпретировать только высоко прогностичные модели, поскольку такие модели дают разумные отношения структура–свойство [12, 13]. Но если даже слабые модели удастся объединить в ансамбль (консенсус) [14], то окончательная модель также может быть интерпретирована.

Интерпретация должна учитывать область применимости (Application Domain, AD) модели, т.к. модель, обученная на ограниченной выборке, охватывает только небольшую часть химического пространства [15]. Область применимости может быть определена с использованием различных подходов [16]. Результаты интерпретации не могут быть экстраполированы на соединения, которые находятся слишком далеко от области применимости, поэтому, если мы получаем определенную связь структуры и свойства, это не означает, что она применима к любому соединению в химической вселенной.

Результаты интерпретации зависят от соединений обучающей выборки, используемой для моделирования, поскольку они могут меняться для одной и той же конечной точки; отсюда и следует изменение области применимости модели.

2.1 Парадигма интерпретации «модель → дескрипторы → (структура)»

Традиционно, понимаемые химиками модели были построены только на интерпретируемых дескрипторах. Парадигма интерпретации «модель → дескрипторы → (структура)» была сформулирована неявно и состояла из двух последовательных шагов. На первом, вклады дескрипторов вычисляются в зависимости от интерпретируемости метода машинного обучения, используемого для моделирования. На втором шаге интерпретируются рассчитанные вклады «как есть» или переводятся на структурный уровень для оценки вкладов подструктур. Ясно, что этот шаг зависит от интерпретируемости дескрипторов.

Для специалистов computer science, наибольший интерес представляет первый шаг, -оценка зависимости интерпретируемости модели от метода машинного обучения. Поэтому анализируя роль машинного обучения в интерпретации, рассматривают две группы подходов машинного обучения: метод-специфичные, применяемые только к моделям, основанным на конкретных методах машинного обучения, и метод-независимые, которые могут применяться к моделям, основанным на любом методе машинного обучения.

2.2 Метод-специфичные подходы Множественная линейная регрессия (MLR, модели Hansch)

Первые интерпретируемые модели QSAR были разработаны Hansch и др.; в моделях тщательно подбирали химические дескрипторы для объяснения активности ингибирования роста растений феноксиуксусных кислот.

$$\frac{1}{C} = 4.08\pi - 2.14\pi^2 + 2.78\sigma + 3.38 \quad (1)$$

Результаты моделирования были полезны для разработки новых активных соединений, однако уравнение (1) оказалось неприменимым вне серии феноксиуксусной кислоты.

Bhatarai et al. показал, что результаты интерпретации моделей MLR, полученные на основе различных наборов дескрипторов, и с использованием различных методов выбора переменных сходятся [17]. В литературе и сегодня преобладают два параллельных подхода построения MLR моделей QSAR, один из которых основан на вручную выбранных дескрипторах, а другой - на использовании статистических методов. Чтобы проверить основные различия между этими методами, было проведено сравнительное их исследование на 158 циклоалкилпиранонах, биологически активных по ингибированию (Ki) ВИЧ-протеазы. Для этого вначале разработали модели MLR на ограниченном количестве «понятных» молекулярных дескрипторов. Затем, высоко прогностические модели MLR были получены на том же наборе дескрипторов, но с использованием двух статистических подходов с отбором дескрипторов разными способами. Все модели признали гидрофобность и поляризуемость как важные факторы, что доказывало гипотезу Bhatarai et al. о том, что результаты интерпретации могут быть одинаковыми, независимо от используемых дескрипторов и программ их отбора.

2.3 Модели Free-Wilson

Free и Wilson предложили другой подход, основанный на предположении об аддитивном характере количественных отношений структуры-свойство [18]. Линейное уравнение связывает биологическую активность соединений с наличием/отсутствием конкретных заместителей в их структурах.

Относительная ингибирующая активность соединений против *Staphylococcus aureus* представлена уравнением 2.

$$\begin{aligned} Act = 75R_H - 112R_{CH_3} + 84X_{Cl} - 16X_{Br} - 26X_{NO_2} + \\ + 123Y_{NH_2} + 18Y_{NHC(=O)CH_3} - 218Y_{NO_2} \end{aligned} \quad (2)$$

где R , X и Y - индикаторные переменные, равные 1, если заместитель присутствует в этой позиции, и равны 0, если его нет. Из-за того, что модели линейной регрессии не работают с взаимно коррелированными переменными и высоко размерными данными, в модели ограничено число переменных (в 5-10 раз меньше количества соединений) и исключены коррелированные дескрипторы. Эти проблемы решаются различными методами регуляризации, например, гребневой регрессией, или методом частичных наименьших квадратов (PLS) [19].

Коэффициенты уравнения (2) показывают вклад заместителей. Модель была обучена и протестирована на небольшой выборке (10 соединений в обучении, 8 в тестировании прогнозной способности модели). Очевидно, что модели имеют ограниченную область применимости, - в основном, для аддитивных конечных точек, которые бывают редко, но модели понятны химикам из-за четкого структурного объяснения

отношений структура-активность. Отметим, что идея новой парадигмы интерпретируемости модели очень похожа на подход Free и Wilson.

2.4 Частичные наименьшие квадраты (PLS)

PLS является обобщением множественной линейной регрессии и может работать с коррелированными и многомерными X -переменными и несколькими выходами (Y). Модель PLS находит новые дескрипторы (X -оценки, T), которые являются оценками скрытых переменных. X -оценки ортогональны, они оцениваются как линейная комбинация исходных переменных X с коэффициентами W^* (уравнение 3) и являются хорошими предикторами Y (уравнение 4). Метод был предложен в 1982г. на Gordon Conference, где S.Wold [19] впервые изложил подход PLS.

$$T = XW^* \quad (3)$$

$$Y = TC' + F = XW^*C' + F = XB + F \quad (4)$$

Анализ коэффициентов регрессии в модели PLS (уравнение 3) аналогичен нормальной линейной регрессии (модели Hansch).

Примеры моделей PLS можно найти в Stanton [20]. Другие методы PLS, такие как ортогональные PLS (OPLS), O2PLS, L-образные PLS и ортогональные LPLS, также нашли применение в хемоинформатике. Их можно интерпретировать аналогично обычным PLS-моделям на основе коэффициентов регрессии, оценок и нагрузки [21, 22].

2.5 Деревья решений

Прогностическая способность линейных моделей ограничена, если нет возможности установить линейный тренд отношений структуры и свойства. В этих случаях предпочтительными являются методы, которые работают с нелинейностью. Одним из таких ранних подходов являются деревья решений, которые не являются параметрическими и могут применяться к задачам регрессии и классификации [23]. Модели дерева решений могут работать с многомерными данными и коррелированными дескрипторами, поскольку шаг выбора переменных здесь реализуется в процессе каждого расщепления веток дерева, - модель выбирает оптимальное значение дескриптора для разделения. Модели дерева решений подвержены переобучению и восприимчивости к шуму, что существенно снижает их прогностическую способность. Основным недостатком деревьев решений в интерпретации является их нестабильность, - небольшие изменения в обучающей выборке могут привести к существенно разным правилам разделения в узлах дерева.

2.6 Модели «черного ящика»

Для преодоления проблем низкой прогностической способности и нестабильности модели были применены более сложные подходы к моделированию: нейронные сети (NN), векторная машина (SVM) и случайный лес (RF) [24]. Они могут работать в многомерном дескрипторном пространстве даже без предварительного выбора переменных, обеспечивая высокую прогностичность и разумную вычислительную скорость. В то время как модели NN подвержены переобучению, модели SVM и RF более устойчивы и толерантны к шуму во входных данных. Эти методы стали золотыми стандартами в химической информатике, но, в отличие от линейных моделей и деревьев решений, их интерпретация, - непростая задача. Поэтому эти модели называют моделями «черного ящика».

Не останавливаясь на описании архитектур и функций нейронных сетей, проанализируем новое направление, -глубокое обучение нейронных сетей [25]. Глубокие нейронные сети имеют несколько скрытых слоев и способны вычислять слои адаптивных нелинейных признаков, которые охватывают более сложные шаблоны данных с каждым дополнительным слоем. Основной архитектурой глубокой нейронной сети является полностью связанная сеть прямой передачи с более чем одним скрытым слоем. Термин «полностью подключенный» относится к архитектуре, в которой каждый нейрон на сетевом уровне подключен ко всем нейронам без смещения на следующем уровне, а «подача вперед» указывает на отсутствие циклов или обратных связей в сетевой архитектуре.

Проблема обучения состоит из аппроксимации многомерной нелинейной функции $f(x)$. Это приближение реализуется несколькими нелинейными взвешенными преобразованиями вектора входных данных в скрытых слоях. Нелинейность реализуется в общей сети путем определения соответствующей архитектуры и использования нелинейных функций активации. Общим выбором для функций активации являются сигмоидальные функции, такие как логистическая функция

$$a(z) = \frac{1}{1+e^{-z}}$$

и функция линеаризации

$$a(z) = \max(0, z).$$

Примером частоиспользуемой функции оценки является сумма квадратов ошибок (sum of squared errors, SSE),

$$C(O_i^s) = \frac{1}{2} \sum_s \sum_l (t_l^s - O_l^s)^2,$$

где S - индекс, обозначающий один образец обучения (вектор молекулярного дескриптора), подаваемый в сеть. Как только вычисляется функция оценки, можно попытаться свести ее к минимуму с помощью веса, чтобы повысить производительность сети. Таким образом, вычисление градиента функции оценки и последующая корректировка весов обеспечивают оптимизацию веса:

$$\Delta w_{qp}^{(n)} = -\eta \frac{\partial C}{\partial w_{qp}^{(n)}} = \begin{cases} -\eta \delta_q^{(n)} x_p, & n = 1 \\ -\eta \delta_q^{(n)} h_p^{(n-1)}, & n \neq 1, \end{cases}$$

$$\delta_q^{(n)} = \frac{\partial C}{\partial z_q^{(n)}} \begin{cases} E(o_q) a'(z_q^{(n)}), & n = N \\ a'(z_q^{(n+1)}) \sum_r \delta_r^{(n+1)} w_{rq}^{(n+1)}, & n \neq N, \end{cases}$$

где $\Delta w_{qp}^{(n)}$ обозначается изменение одного из весов сети (p, q и r - произвольные индексы для нейронов внутри вовлеченных слоев), η - скорость обучения $n \in \{1, \dots, N\}$, l представляет собой индекс по числу сетевых уровней (исключая входной уровень, который содержит только модули разветвления), $h_p^{(n-1)}$ является выходным значением скрытого нейрона p в слое $(n - 1)$, $z_q^{(n)}$ - предварительная активация нейрона q в слое n , $a(\dots)$ - нелинейная активационная функция с $a'(\dots)$, обозначая его производную. Функция ошибки косвенно зависит от всех весов сети через выходные значения. $o_l(w_{lk}^{(N)}, h_k^{(N-1)})$

Существует несколько преимуществ глубоких нейронных сетей, которые, как предполагают, должны заменить мелкие NN:

- Глубокая, иерархическая архитектура позволяет сети выполнять множество нелинейных преобразований, что приводит к изучению более абстрактных функций по сравнению с неглубокими сетями [26]. Это позволяет реализовать более сложную комбинацию низкоуровневых функций (т.е. нелинейных комбинаций дескрипторов) с тем, чтобы добиться лучшего молекулярного представления и, следовательно, классификации соединений. Уменьшается потребность в априорной разработке сложных дескрипторов. Абстрактные нелинейные функции имеют тенденцию быть инвариантными к локальным изменениям входных данных, что приводит к шумоподавлению и повышению надежности сети. Поэтому архитектура глубокого обучения реализует семейства функций, экспоненциально более эффективных, чем мелкие архитектуры.
- Глубокие архитектуры дают возможность повторного использования признаков. Это имеет значение для отсутствующих классов в наборе обучения, что типично для больших наборов данных QSAR. Многозадачное обучение также имеет преимущества совместно используемого конвейера выделения признаков для нескольких задач. Dahl предположил, что эта концепция может иметь эффект регуляризации, поскольку веса, как правило, определяются так, чтобы быть полезными для многих целей, а не для переобучения одной цели [27]. Это обещает одновременное тестирование биоактивности по многим связанным с ними объектам и поэтому такие приложения могут быть эффективными в области полифармакологии [28]. В этом случае глубокое обучение может использовать методы ядра, чтобы найти дополнительные приложения для лекарств («повторное назначение») и помочь в выявлении нежелательных побочных активностей [29].
- Как было подчеркнуто на конкурсе Merck QSAR [30], глубокие многозадачные нейронные сети могут быть успешно применены к QSAR-моделированию [31]. Генеративные глубокие архитектуры дополняют эволюционные алгоритмы, которые часто используются для поиска новых соединений с определенными функциями, идентифицированными ранее обученной машинной архитектурой [32]. Обученная глубокая генеративная нейронная сеть научилась отделять сложные классы активности, имея возможность генерировать выходные данные в соответствии с полученным представлением, т.е. эта сеть содержит обе функции в одном алгоритме. Поэтому глубокое обучение может дать новый подход к решению «обратной задачи QSAR» [33].

Однако простое расширение сети только в глубину не приводит автоматически к лучшим моделям. Поэтому в работе [34] исследовали: (1) большое количество конфигураций DNN с гиперпараметрами, чтобы понять их влияние на производительность сети при их настройке; и (2) эффективность глубоких сетей относительно широко используемых методов, - Naive Bayes (NB), kNN, случайный лес (RF), векторные машины (SVM). Кроме того, была оценена устойчивость методов машинного обучения к различным уровням искусственно введенного шума. Для проведения исследования применили систему глубокого изучения Caffe с открытым исходным кодом и современные графические процессоры NVidia, позволяющие генерировать большое количество конфигураций DNN. Оказалось, что глубокие feed-forward нейронные сети способны обеспечить высокую эффективность классификации и при оптимизации превзойти обычные методы в разных классах активности. Гиперпараметры, - функция активации, регуляризация выпадения, число скрытых слоев и количество нейронов в слое играют важную роль. DNN дает лучшие QSAR модели, но всё равно

они являются «черными ящиками» как обычные нейронные сети, и потому вопрос интерпретации остается для них актуальным [35].

Был предложен ряд подходов к интерпретации сетевых моделей, идея которых состояла в извлечении правил. Они извлекают правила вида «если ... то ... иначе», как в деревьях решений. Первоначально они были разработаны для интерпретации моделей NN, однако позже алгоритмы были разработаны и для моделей SVM. Одним из ранних подходов, который был применен для интерпретации моделей NN, был КТ [36], который находит отдельные функции (нейроны) или их подмножества, гарантирующие превышение порога активации в нейроне следующего слоя NN. Другой подход, - анализ интервалов достоверности (Validity Interval Analysis, VIA), итеративно находит интервалы входных переменных, которые последовательно активируют выходные нейроны [37]. Подходы генерируют множество правил, более устойчивых к шуму, чем деревья решений, и могут превзойти их, демонстрируя лучшие возможности обобщения. Недавно были разработаны подходы к извлечению правил из глубоких нейронных сетей. [38]. Barakat и Bradley [39] предложили несколько подходов получения правил для моделей SVM.

На самом деле, правила не следует рассматривать как интерпретацию моделей «черного ящика», а только как их приближение к читаемому пользователем результату, но они эффективны, если в модели относительно небольшое число важных дескрипторов. В целом, применимость подходов извлечения правил ограничена.

2.7 Метод-независимые подходы машинного обучения

Эти подходы не опираются на внутреннюю структуру моделей. Они исследуют поведение модели, - как изменяется выход модели с изменением набора дескрипторов и работают со всеми моделями как с «черными ящиками», и потому подходят для интерпретации любой модели, независимо от используемого метода машинного обучения. Этими подходами являются анализ чувствительности, значимость переменных и частные производные.

Анализ чувствительности исследует зависимость выходных значений от систематических изменений в значениях одного дескриптора, при условии, что значения всех остальных дескрипторов остаются неизменными [40]. Поскольку число всех возможных комбинаций значений дескриптора несчетно, было предложено сохранить несколько фиксированных значений дескрипторов: минимум, первый квартиль, медиану, третий квартиль и максимум [41]. В одном из первых исследований по чувствительности, авторы построили NN-модель для 68 ингибиторов дигидрофолатредуктазы: были обнаружены четыре наиболее важных дескриптора, - молярное преломление заместителей в 3, 4 и 5 положениях (MR3, MR4, MR5) и гидрофобная константа заместителя в положении 3 (π_3) [42]. Однако, если модель, как обычно, включает в себя большое количество дескрипторов, то анализ чувствительности становится неэффективным и потому редко используется (см. обзор Iooss и Saltelli [43]).

Оценка важности (значимости) дескриптора говорит об его относительной важности для построения модели, но не о направлении его влияния на конечную точку (положительном или отрицательном). Чаще всего оценка используется для отбора переменных, но возможны и другие приложения. Существует несколько подходов, которые могут быть применены к любой модели [44]. Оценка важности дескриптора является относительно простой методикой и широко используется, но имеет некоторые недостатки, связанные с интерпретацией моделей RF. Алгоритм CART классификации и регрессии дерева (Classification and Regression Tree, CART), используемый в обычной Random Forest для построения отдельных деревьев, предпочтителен для

выбора переменных с большим числом категорий, что приводит к переоценке важности таких переменных [45]. Исходная бутстреп-выборка, используемая в RF, также может иметь отрицательный эффект. Значимость переменной может быть переоценена для высоко коррелированных переменных [46]. Чтобы преодолеть эти проблемы, Strobl et al. предложил: 1) строить деревья на основе условной схемы вывода вместо критерия Gini, используемого в CART, и 2) оценивать значимость переменной по условным значениям, когда исходные объекты сначала разбиваются на группы, согласно значению дескриптора, и работают со значениями дескриптора внутри отдельных групп.

Методом локального анализа чувствительности исключительной мощности и широкого использования является метод частных производных. Aoyama and Ichikawa предложили вычислять частные производные для моделей NN, которые можно получить в аналитической форме [47]. Например, для трехслойной модели NN частные производные i -го дескриптора вычисляются согласно уравнению 5:

$$\frac{\partial Y}{\partial X_i} = \sum_j f'(y_j) I_{ij} g'(y) H_j \quad (5)$$

где $f'(y_j)$ и $g'(y)$ - производные активационных функций во втором (скрытом) и третьем (выходном) слоях модели NN, I_{ij} и H_j - веса соединений между первым и вторым, и между вторым и третьим слоями, соответственно. В случае нескольких выходов в задачах классификации или многозадачных моделях частные производные рассчитываются отдельно для каждого выходного нейрона.

Первые частные производные обозначают вклады дескриптора, а Baskin et al. предложили рассчитать их дисперсии как меру нелинейности вкладов, и дисперсии вторых частных производных для различения зависимостей более высокого порядка, таких как параболические или гиперболические [48]. Marcou et al., предложил схему вычисления частных производных для задач классификации и продемонстрировал их применимость на нескольких примерах [49].

Метод конечных разностей может быть использован для вычисления первых частных производных для моделей, которые не поддерживают аналитические производные. An et al. использовали Canvas 2D отпечатки пальцев и PLS для моделирования многих конечных точек с последующей визуализацией вкладов атомов [50]. Для вычисления частных производных они использовали метод прямой разности с размером шага $\delta = 0,05$.

Существует множество исследований, в которых подход с частными производными использовался для интерпретации моделей QSAR, многие из которых использовали визуализацию расчетных вкладов [51, 52]. Несмотря на простоту вычислений и широкую применимость, подход с частными производными может не работать для некоторых методов машинного обучения, например, для моделей k -ближайших соседей (k -NN), поскольку небольшое изменение одного значения дескриптора вряд ли изменит соседей соединения и, следовательно, прогнозируемое значение. Другой неисследованной проблемой этого подхода является оценка ошибки числового дифференцирования, - уменьшение размера шага снизит погрешность аппроксимации, но увеличит ошибку расчета. Поэтому должно быть найдено оптимальное значение шага δ , которое минимизирует общую ошибку.

Визуализация вкладов дескрипторов очень эффективна для интерпретации моделей, построенных на фрагментных дескрипторах, при этом вклад дескриптора может быть рассчитан с использованием любого доступного подхода, например, это могут

быть коэффициенты регрессии от линейных моделей или частные производные от нелинейных моделей, которые распределены на соответствующих атомах. Вычисленный вклад каждого дескриптора делится на значение дескриптора и число атомов в фрагменте и присваивается соответствующим атомам. Полный вклад атомов представляют собой сумму баллов, полученных из разных дескрипторов. Они могут быть визуализированы на молекулярной структуре исследованных соединений путем цветового кодирования.

В работе [53] Kuz'min et al. моделировали аффинность рецептора 5 HT1A и 347 соединений. Были построены две модели PLS и RF, а для оценки вклада дескриптора были использованы два метод-специфичных метода машинного обучения, чтобы установить тренд отношения структура-активность. Первый метод был основан на коэффициентах регрессии модели PLS как вкладах дескрипторов, а второй - на оценке вкладов дескрипторов из моделей RF, описанных в предыдущем разделе. Обе модели обучались на симплексных дескрипторах (количество тетраатомных фрагментов). Для создания модели PLS было отобрано 72 дескриптора, тогда как модель RF была обучена по всем 6460 дескрипторам. Обе модели имели разумную прогностическую мощь. Суммированные вклады атомов были визуализированы на обучающих наборах молекул для представления относительной важности различных частей молекулы - локальной интерпретации.

Рассчитанные вклады атомов можно суммировать для выбранных фрагментов, чтобы получить общую тенденцию отношения структура-активность, что выявляет важную особенность интерпретации модели: результаты структурной интерпретации правильных моделей не должны существенно зависеть от выбранного метода машинного обучения, процедуры интерпретации или выбора переменных.

И тем не менее, общим ограничением всех подходов, используемых в традиционной парадигме «модель → дескрипторы → (структуры)», является необходимость использования только интерпретируемых дескрипторов.

2.8 Универсальная парадигма интерпретации: «модель → структура»

Подходы, построенные на этой парадигме, появились недавно и еще не стали популярными. Однако они обладают большим потенциалом, т.к. позволяют интерпретировать любую модель независимо от метода машинного обучения и используемых дескрипторов, за счет оценки вклада отдельных структурных мотивов непосредственно из модели, которая устраняет промежуточный этап расчета вкладов дескрипторов.

Идея подхода очень простая. Если имеется молекула AB , состоящая из двух фрагментов A и B , то вклад фрагмента B можно рассчитать, как разность между прогнозируемыми значениями активности для соединения AB и соответствующего фрагмента A (гипотетическая молекула, которая получается из AB путем удаления фрагмента B).

Существует несколько реализаций парадигмы «модель → структура». Первые два подхода, которые похожи, были предложены независимо Riniker и Landrum [54], а также Polishchuk et. al. [55]. Riniker и Landrum предложили оценить вклад отдельных атомов, удаляя их из структуры и вычисляя разницу между предсказанными значениями исходной структуры и структуры с удаленным атомом. Это дает информацию о вкладе отдельных атомов в каждую молекулу – локальную интерпретацию. Однако этот подход ограничен использованием дескрипторов, которые могут быть рассчитаны для молекул, состоящих из нескольких фрагментов. После удаления одного атома довольно часто появляются множественные фрагменты, и поэтому мо-

дели, основанные, например, на дескрипторах Dragon, не могут быть истолкованы таким образом.

Polishchuk et al. предложил более общий подход, – оценивать вклады целых фрагментов или групп атомов с помощью той же процедуры удаления. Это позволяет оценить вклад целых фрагментов (функциональных групп, каркасов, линкеров), а также эффект удаления их сочетания, если две или несколько групп удаляются одновременно из структуры. Этот подход может использоваться для локальной и глобальной интерпретации. Последняя означает ранжирование серии отдельных фрагментов в соответствии с их средними вкладами для представления взаимосвязанных отношений структуры и активности. Распределение вкладов одного фрагмента может также давать дополнительную полезную информацию: низкая дисперсия вкладов означает, что вклад фрагментов почти не зависит от молекулярного контекста, а высокая дисперсия предполагает сильное влияние молекулярного контекста на вклад фрагмента.

Применимость разработанного метода структурной интерпретации была продемонстрирована на двух задачах моделирования: растворимости (регрессия) и мутагенности Ames (классификация) [55]. Позднее этот подход к интерпретации был успешно применен в других исследованиях: моделировании антагонистов рецепторов аденозинового A2B (ансамбль линейных моделей, построенных на дескрипторах Dragon и MOE) [56] и моделировании проницаемости гематоэнцефалического барьера (модели RF и SVM на основе ECFP) [57].

Описанный подход к структурной интерпретации был расширен с тем, чтобы получить возможность оценить не только общий вклад фрагментов, но и вклад их различных физико-химических свойств, чтобы объяснить важность этих фрагментов [10]. Расширенный подход структурной и физико-химической интерпретации (SPCI) использует интерпретируемые дескрипторы, которые имеют ясный физико-химический смысл, но не относятся к парадигме «модель→дескрипторы→(структура)», потому что вклады отдельных дескрипторов не вычисляются.

Проиллюстрируем подход SPCI, используя простой пример. Предположим, что активность описывается уравнением $F = \alpha\pi + p\sigma + i$, где π - гидрофобный член, например, липофильность, σ - электронный член, например, константа Гамметта, а i – свободный член. Пусть есть молекула AB , состоящая из двух фрагментов. Общий вклад фрагмента B можно рассчитать, как разность между $F_{AB} = \alpha\pi_{AB} + p\sigma_{AB} + i$ и $F_A = \alpha\pi_A + p\sigma_A + i$ (структурная интерпретация). Вклад отдельных физико-химических признаков фрагмента B , например, гидрофобного фактора, рассчитывается как разность между $F_{AB} = \alpha\pi_{AB} + p\sigma_{AB} + i$ и $F_{\pi A} = \alpha\pi_A + p\sigma_{AB} + i$ (физико-химическая интерпретация). Фрагмент B был удален с точки зрения одного физико-химического фактора. Это можно повторить отдельно для всех других членов.

Не останавливаясь на деталях, отметим, что подход SPCI реализуется в программном обеспечении с открытым исходным кодом и графическим пользовательским интерфейсом http://qsar4u.com/pages/sirms_qsar.php [58]. Однако подход не ограничивается использованием только симплексных дескрипторов; могут использоваться и другие подходящие дескрипторы.

SPCI был успешно применен для интерпретации RF, SVM, PLS, Gradient Boosting Machines и ансамблевых моделей проницаемости гематоэнцефалического барьера, сродства антагонистов рецептора фибриногена и острой токсичности *in vivo* у крыс, а также на одном из классических наборов данных модели Free-Wilson [10]. Все ре-

зультаты интерпретации отдельных моделей сходились и соответствовали экспериментально наблюдаемому соотношению структура-активность.

Ещё один подход структурной интерпретации был предложен Sushko et al. [59]. Авторы предложили рассчитать разницу между прогнозируемыми значениями активности двух соединений, которые образуют согласованную пару (ММР), для оценки вклада преобразования; например, разница в предсказанных значениях соединений AB и AC приводит к вкладу трансформации $B > C$. Такая информация может быть полезна для оптимизации хита. Отсутствующие члены согласованных пар в обучающем наборе могут быть фактически созданы для увеличения охвата пар и повышения значимости результатов по сравнению с ММР. Предлагаемый подход не имеет ограничений на используемые дескрипторы. Он был реализован как часть портала OCHEM (<http://ochem.eu>).

Сравнение подходов ММР и SPCI в интерпретации моделей QSAR было проведено в [10]. Оба подхода ММР и SPCI имеют свои преимущества и недостатки. Некоторые из характеристик перечислены в таблице 1. Выбор подходящего подхода зависит от набора данных и задачи. Однако авторы из [10] считают, что объединение подходов может дать различную информацию об исследуемой проблеме.

Таблица 1

Сравнение ММР и SPCI подходов

	SPCI	ММР
Задачи	Регрессия и классификация	Регрессия и классификация
Результат	Структурная и физико-химическая интерпретация	Структурная интерпретация только
Размер выборки	Любая, на ней можно строить модель	Достаточно большая
Разнообразие выборки		Должна содержать МР, которые могут лимитировать разнообразие
Точность	Зависит от экспериментальных и прогнозных ошибок	Зависит от экспериментальных и ошибок и числа пар
Ранжирование фрагментов	Все фрагменты могут быть ранжированы одновременно	Только фрагменты, появляющиеся в МР (или сериях) могут быть ранжированы
Зависимость от контекста	Да, если используются неаддитивные модели	Да

Структурно-физическая интерпретация моделей QSAR работает на любом наборе данных, независимо от размера и структурного разнообразия, если могут быть построены прогностические модели. ММР может применяться, в основном, для больших наборов данных, содержащих достаточное число молекулярных пар, что неявно ограничивает разнообразие набора данных.

Точность интерпретации QSAR зависит от ошибок эксперимента и прогнозирования. На результаты MMP-подхода влияют экспериментальные ошибки и количество пар [60]. Важно отметить, были ли измерены экспериментальные значения с использованием тех же анализов или условий, поскольку в случае, если они различны, это может существенно повлиять на конечные значения в результате систематических ошибок и введения дополнительных помех в наборе данных.

Такой шаг проверки данных часто отсутствует в анализе MMP, что может исказить его результаты. Поэтому данные должны быть проверены до анализа, или их неоднородность должна учитываться при оценке статистической значимости результатов [60]. Ниже представлены результаты вычислительных экспериментов, тестирующие новую парадигму интерпретации «модель → структура».

3 Материал, результаты и обсуждение

По современной химической номенклатуре алкалоиды гармалы обыкновенной хиназолинового строения и их производные, исследованные в [61], разделяются на две группы: 1) соединения собственно хиназолинового строения или дезоксипеганинового ряда (соединения от 1 до 43); и 2) соединения хиназолонного строения или дезоксивазиционного ряда (соединения от 46 до 65). Этот материал мы использовали в своих ранних работах [62–65], применяя разные наборы дескрипторов и методы машинного обучения. Потому представляется интересным сравнить эти результаты с новыми, полученными в рамках структурной и физико-химической интерпретации, предлагаемой в подходах P.Polishchuk, [10, 55] и Yu.Sushko и др. [60]. После фильтрации 55 соединений в формате sdf и их активности в форме $\log(\frac{1}{C})$ были загружены в программу SPCI.

Схема интерпретации: пусть есть два соединения $A - B$ и $A - C$ с общим каркасом A . Если мы предскажем с помощью модели QSAR значение свойства для фрагмента $A(P^{pred}(A))$ и вычтем его из предсказанных значений свойства соединений $A - B(P^{pred}(AB))$ и $A - C(P^{pred}(AC))$, то мы сможем оценить и сравнить вклад фрагментов $B(P_{AB}(B))$ и $C(P_{AC}(C))$ в исследуемое свойство.

$$P_{AB(i)} = p^{pred}(AB(i)) - p^{pred}(B) \quad (6)$$

$$P(A) = \frac{1}{n} \sum_i^n P_{AB(i)}(A) \quad (7)$$

где $P_{AB(i)}(A)$ является локальным вкладом фрагмента A в значение свойства соединения AB_i , $P^{pred}(AB_i)$ является прогнозируемым значением активности соединения $AB(i)$, содержащим фрагмент A , $P^{pred}(B)$ является предсказанной активностью для молекулярной части B_i соединения $AB(i)$, оставшейся после удаления фрагмента A , $P(A)$ является глобальным вкладом фрагмента A в свойство, а n есть число соединений, содержащих фрагмент A .

Алгоритм интерпретации состоит из следующих шагов: (I) удаление выбранного молекулярного фрагмента из молекулы; (II) обработка оставшейся части (частей) молекулы и расчет дескрипторов для них; (III) прогнозирование (с учетом области применимости, AD) значений активности для оставшейся части (ей) со стадии (II); (IV) оценка вклада фрагмента как разность между предсказанной активностью молекулы и предсказанными значениями фрагментов, полученных на этапе (III) (уравнение 6); (V) повторение этапов (I) - (IV) для всех молекул, содержащих выбранный

фрагмент; и (VI), усреднение полученных вкладов фрагментов (уравнение 7). Полученное среднее значение представляет собой глобальный вклад фрагмента в целевое свойство.

Мы использовали симплекс-дескрипторы, помеченные частичным зарядом атомов, липофильностью, рефрактивностью и H-связыванием. Все эти параметры были рассчитаны с использованием программного инструментария ChemAxon *sxcalc* [66]. В структурной интерпретации удалялись дескрипторы всех четырех групп, упомянутых выше, для каждого выбранного фрагмента. Для физико-химической интерпретации удаляли одну группу дескрипторов за один раз для каждого фрагмента и каждый раз повторяли процедуру для каждой группы дескрипторов независимо.

Для моделирования использовали пакет Python *sklearn*. SVM с ядром радиальной базисной функции (Radial Basis Function, RBF). RF и метод бустинга градиента (Gradient Boosting Method, GBM) были построены для того, чтобы решить задачу классификации. Для регрессионных моделей были использованы методы SVM, RF, GBM и частичных наименьших квадратов (PLS). Прогнозы на основе ансамбля были получены путем усреднения предсказаний регрессионных моделей или путем выбора класса, который имел наибольшее число голосов среди классификационных моделей. Производительность всех моделей оценивали путем пятикратной перекрестной проверки. Параметры моделей были оптимизированы перебором по сетке. Статистические параметры для оценки прогнозирующей эффективности моделей регрессии и классификации получены по результатам кросс-валидации, - квадрат коэффициента R^2 корреляции между предсказанной и наблюдаемой активностью, среднеквадратичная ошибка RMSE. В таблице 2 приведены результаты вычислительного эксперимента, из которого следует, что наилучший выход имеет модель RF случайного леса, хотя цифры отнюдь нельзя назвать адекватными по сравнению с оценками, полученными нами на модели kNN-QSAR.

Таблица 2

Модели регрессии на симплекс дескрипторах

Модель	R^2	RMSE
GBM	0.42	0.32
RF	0.54	0.29
SVM	0.50	0.30
PLS	0.31	0.35

На рисунке 1 представлены результаты моделирования по методу SPCI на 4 моделях и их ансамбле.

Из рисунка 1 можно видеть, какие вклады фрагментов увеличивают, а какие уменьшают активность. Характер изменения вкладов независимо от модели машинного обучения, имеет одинаковые тренды. Визуализация интерпретируемости моделей еще раз демонстрирует, что лучшей по определенности вкладов фрагментов является модель RF.

Дополнением к этому является представленная ниже диаграмма медиан фрагментов разных физико-химических факторов.

Таким образом, вычислительный эксперимент еще раз показал, что баланс между точностью модели и её интерпретируемостью имеет место. Фактически, потеряв в

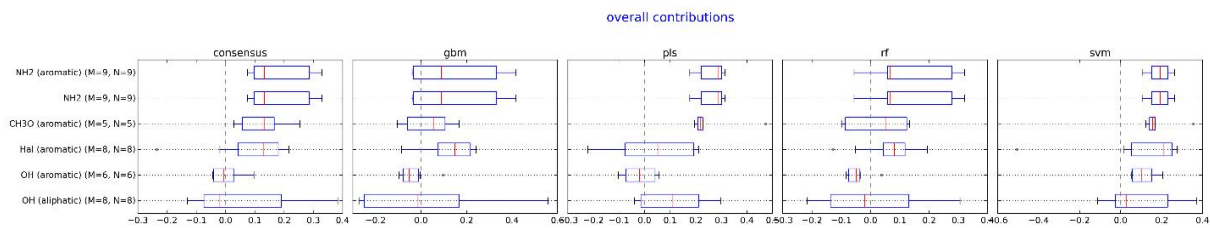


Рис. 1. Распределение вкладов фрагментов, рассчитанных с использованием моделей RF, GBM, SVM, PLS и ансамбля. Показаны только фрагменты, встречающиеся, по меньшей мере, в 10 соединениях. M - количество соединений, содержащих фрагмент, а N - количество фрагментов в наборе данных (некоторые соединения имеют несколько одинаковых фрагментов, и их вклады оценивались отдельно).

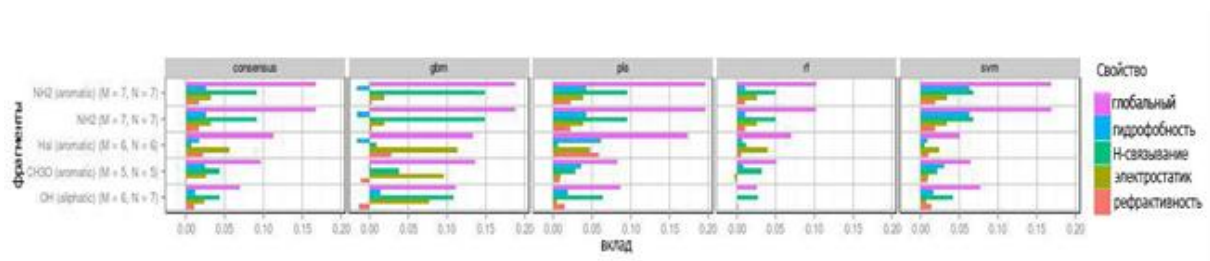


Рис. 2. Медианы фрагментов разных физико-химических факторов (ось Y), оцененных по модели ансамбля, моделям GBM, RF, SVM, PLS. Показаны только фрагменты, встречающиеся, по меньшей мере, в 10 соединениях. Значения M и N - те же, что и на рисунке 1. По оси X представлены вклады физико-химических свойств в каждый из фрагментов.

точности модели, получили (впервые) наглядно физико-химические свойства соединений, которые определяют её активность.

К сожалению, мы не можем разделить оптимизма по поводу интерпретируемости в рамках подхода MMP, поскольку в исходном наборе соединений просто не нашлось достаточное число пар, с которыми можно работать. И это еще раз подтверждает рациональность подхода на симплекс-дескрипторах, предложенного Pavel G. Polishchuk. В заключение, - несколько общих замечаний о потенциале направления интерпретируемости в QSAR моделировании.

4 Заключение

Результаты интерпретации во всех случаях имеют отношение к существующему знанию структурных свойств в соответствующих областях; другими словами, все модели основаны на знаниях. Однако следует отметить, что результаты интерпретации зависят от данных, используемых для моделирования, и они могут быть необъективными. Отдельные модели, построенные с использованием разных статистических методов, дают аналогичные результаты интерпретации, но во избежание смещения и шума, создаваемых отдельными моделями, использование ансамблевых моделей представляется предпочтительным.

Также важно правильно понимать значимость результатов SPCI и MMP, т.к. использование стандартных статистических тестов является недостаточным, - на результаты влияют моделирование и ошибка эксперимента. Однако точно оценивать экспериментальные ошибки в больших наборах данных, где данные собираются из разных источников очень сложно. MMP, и SPCI модели QSAR могут быть использо-

ваны для интеллектуального анализа наборов химических данных, но SPCI можно считать хорошей альтернативой анализу ММР, поскольку (i) подход SPCI обеспечивает тренды отношения структуры - активности, которые можно рассматривать как «бесконечная молекулярная серия» и (ii) подход SPCI больше подходит для анализа небольших или разнообразных наборов данных.

Вероятно, появятся новые подходы к интерпретации, при этом акцент будет сделан на разработку подходов, ориентированных на набор данных типа «черного ящика», где основной механизм (механизмы) неизвестен.

Другим важным направлением является разработка простых в использовании интерпретационных рабочих процессов и инструментов для анализа огромных объемов информации при интерпретации моделей QSAR и их визуализации. Здесь представляется привлекательной визуализация отношений структура-активность в виде (не)направленных графов. Однако, такие сети могут быть чрезвычайно сложными и нуждаться в инструменте для извлечения из них соответствующей информации [67, 68]. Интерпретация сложна даже при использовании автоматизации, потому человеческое вмешательство по-прежнему будет необходимо.

Но самое перспективное продолжение направления интерпретируемости моделей QSAR,- её потенциал в решении так называемой обратной задачи QSAR,-когда требуется строить структуру соединений, обеспечивающей требуемую исследователю активность.

Литература

- [1] *Cherkasov A. et.all.* QSAR Modeling: Where Have You Been? Where Are You Going To? // *Med Chem*, 2014. Т. 57. № 12. С. 4977–5010. doi: <http://dx.doi.org/10.1021/jm4004285>.
- [2] *Novotarskyi S. et.all.* A comparison of different QSAR approaches to modeling CYP450 1A2 inhibition // *J Chem Inf Model*, 2007. Т. 51. № 6. С. 1271–1280. doi: <http://dx.doi.org/10.1021/ci200091h>
- [3] *Zhu H. et.all.* Combinatorial QSAR Modeling of Chemical Toxicants Tested against *Tetrahymena pyriformis* // *J Chem Inf Model*, 2008. Т. 48. № 4. С. 766–784. doi: <http://dx.doi.org/10.1021/ci700443v>.
- [4] *Guha R.* On the interpretation and interpretability of quantitative structure–activity relationship models // *J Comput - Aided Mol Des*, 2008. Т. 22. № 12. С. 857–871. doi: <http://dx.doi.org/10.1007/s10822-008-9240-5>.
- [5] *Fujita T. Winkler D.* Understanding the Roles of the “Two QSARs” // *J Chem Inf Model*, 2016. Т. 56. № 2. С. 269–274. doi: <http://dx.doi.org/10.1021/acs.jcim.5b00229>.
- [6] *Johansson U. et.all.* Trade-off between accuracy and interpretability for predictive in silico modeling // *Future Med Chem*, 2011. Т. 3. № 6. С. 647–663. doi: <http://dx.doi.org/10.4155/fmc.11.23>.
- [7] *Kuz'min V. E.* Hierarchic system of QSAR models (1D – 4D) on the base of simplex representation of molecular structure // *J Mol Model*, 2005.Т. 11. № 6. С. 457–467. doi: <http://dx.doi.org/https://doi.org/10.1007/s00894-005-0237-x>.
- [8] *Kuz'min V. E., Artemenko A. G., Muratov E. N.* Hierarchical QSAR technology based on the Simplex representation of molecular structure // *J Comput Aided Mol Des*, 2008. Т. 22. № 6–7. С. 403–421. doi: <http://dx.doi.org/10.1007/s10822-008-9179-6>.
- [9] *Polishchuk P. G. et.all.* Universal approach for structural interpretation of QSAR/QSPR models // *Mol Inf*, 2013. Т. 32. № 9–10. С. 843–853. doi: <http://dx.doi.org/10.1002/minf.201300029>.

- [10] *Polishchuk P. G. et. al.* Structural and physico-chemical interpretation (SPCI) of QSAR models and its comparison with matched molecular pair analysis // *J Chem Inf Model*, 2016. Т. 56. №8. С. 1455–1469. doi: <http://dx.doi.org/10.1021/acs.jcim.6b00371>.
- [11] *Polishchuk P. G.* Interpretation of QSAR models: past, present and future // *J Chem Inf Model*, 2017. Т. 57. №11. С. 2618–2639. doi: <http://dx.doi.org/10.1021/acs.jcim.7b00274>.
- [12] *Tropsha A., Gramatica P., Gombar V. K.* The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models // *Mol Inf*, 2003. Т. 22. №1. С. 69–77. doi: <http://dx.doi.org/10.1002/qsar.200390007>.
- [13] *Tropsha A.* Best practices for QSAR model development, validation, and exploitation // *Mol Inf*, 2010. Т. 29. №6–7. С. 476–488. doi: <http://dx.doi.org/10.1002/minf.201000061>.
- [14] *Адылова Ф. Т., Икрамов А. А.* Ансамблевые методы в биоинформатике: опыт их применения в геномике и QSAR моделировании // *Проблемы вычислительной и прикладной математики*, 2017. №3. С. 87–95.
- [15] *Polishchuk P. G., Madzhidov T. I., Varnek A.* Estimation of the size of drug-like chemical space based on GDB-17 data // *J Comput-Aided Mol Des*, 2013. Т. 27. №8. С. 675–679. doi: <http://dx.doi.org/10.1007/s10822-013-9672-4>.
- [16] *Roy K., Ambure P., Aher R. B.* How important is to detect systematic error in predictions and understand statistical applicability domain of QSAR models? // *Chemom Intell Lab Syst*, 2017. Т. 162. №8. С. 44–54. doi: <http://dx.doi.org/10.1016/j.chemolab.2017.01.010>.
- [17] *Bhatarai B., Garg R., Gramatica P.* Are mechanistic and statistical QSAR approaches really different? MLR Studies on 158 Cycloalkyl-Pyranones // *Mol Inf*, 2010. Т. 29. №3. С. 511–522. doi: <http://dx.doi.org/10.1002/minf.201000011>.
- [18] *Free S. M., Wilson J. W.* A mathematical contribution to structure-activity studies // *J Med Chem*, 1964. Т. 7. №4. С. 395–399. doi: <http://dx.doi.org/10.1021/jm00334a001>.
- [19] *Wold S., Sjöström M., Eriksson L.* PLS-regression: a basic tool of chemometrics // *Chemom Intell Lab Syst*, 2001. Т. 58. №2. С. 109–130. doi: [http://dx.doi.org/10.1016/S0169-7439\(01\)00155-1](http://dx.doi.org/10.1016/S0169-7439(01)00155-1).
- [20] *Stanton D. T.* QSAR and QSPR model interpretation using partial least squares (PLS) analysis // *Curr Comput-Aided Drug Des*, 2012. Т. 8. №2. С. 107–127. doi: <http://dx.doi.org/10.2174/157340912800492357>.
- [21] *Eriksson L. et. al.* Orthogonal PLS (OPLS) modeling for improved analysis and interpretation in Drug design // *Mol Inf*, 2012. Т. 31. №6–7. С. 414–419. doi: <http://dx.doi.org/10.1002/minf.201200158>.
- [22] *Hasegawa K., Funatsu K.* Data mining of chemogenomics data using bi-modal PLS methods and chemical interpretation for molecular design // *Mol Inf*, 2014. Т. 33. №11–12. С. 749–756. doi: <http://dx.doi.org/10.1002/minf.201400061>.
- [23] *Breiman L.* Classification and regression trees // Wadsworth: Belmont CA, 1984. С. 368.
- [24] *Breiman L.* Random forests // *Machine learning*, 2001. Т. 45. №1. С. 5–32. doi: <http://dx.doi.org/10.1023/A:1010933404324>.
- [25] *Goodfellow I., Bengio Y., Courville A.* Deep Learning // The MIT Press London England, 2016. С. 800.
- [26] *Bengio Y., Courville A., Vincent P.* Representation learning: a review and new perspectives // *IEEE Trans Pattern Anal Mach Intell*, 2013. Т. 35. №8. С. 1798–1828. doi: <http://dx.doi.org/10.1109/TPAMI.2013.50>.

- [27] *Dahl G.* Deep learning approaches to problems in speech recognition. // Computational chemistry and Natural language text processing PhD thesis University of Toronto , 2015. C. 92. doi: <http://dx.doi.org/1705.10874>.
- [28] *Schneider G.et.all.* Coping with polypharmacology by computational medicinal chemistry // *Chimia* , 2014. T.68. №9. C. 648–653. doi: <http://dx.doi.org/10.2533/chimia.2014.648..>
- [29] *Reutlinger M.et.all.* Multi-objective molecular de novo design by adaptive fragment prioritization // *Agnew Chem Int Ed Engl* , 2014. T.53. №16. C. 4244–4248. doi: <http://dx.doi.org/10.1002/anie.201310864>.
- [30] *Ma J.et.all.* Deep neural nets as a method for quantitative structure–activity relationships // *J Chem Inf Model*, 2015. T.55. №2. C. 263–274. doi: <http://dx.doi.org/10.1021/ci500747n>.
- [31] *Erhan D.et.all.* Collaborative filtering on a family of biological targets // *J Chem Inf Model*, 2006. T.46. №2. C. 626–635. doi: <http://dx.doi.org/10.1021/ci050367t>.
- [32] *Schneider G.et.all.* Virtual screening for bioactive molecules by evolutionary de novo design // *Agnew Chem Int Ed Engl*, 2000. T.39. №22. C. 4130–4133. doi: <http://dx.doi.org/10.1002/1521-3773>.
- [33] *Wong W. W., Burkowski F. J.* A constructive approach for discovering new drug leads: Using a kernel methodology for the inverse-QSAR problem // *J Cheminform*, 2009. T.1. №4. doi: <http://dx.doi.org/10.1186/1758-2946-1-4..>
- [34] *Koutsoukas A.et.all.* Deep learning: investigating deep neural networks hyperparameters and comparison of performance to shallow methods for modeling bioactivity data // *J Cheminform*, 2017. T.9. №42. doi: <http://dx.doi.org/10.1186/s13321-017-0226-y>.
- [35] *Guha R, Jurs P.* Interpreting computational neural network QSAR models: a measure of descriptor importance // *J Chem Inf Model*, 2005. T.45. №3. C. 800–806. doi: <http://dx.doi.org/10.1021/ci050022a>.
- [36] *Fu L.* Rule learning by searching on adapted nets . // *Proceedings of the 9th National conference on Artificial intelligence Anaheim California: AAAI Press*, 1991. C. 590–595. doi: <http://dx.doi.org/AAAI/1991/AAAI91-092>.
- [37] *Thrun S.* Extracting Provably Correct Rules from Artificial Neural Networks // *Technical report University of Bonn Germany*, 1993. doi: <http://dx.doi.org/=10.1.1.2.2110&rep=rep1&type=...>
- [38] *Zilke J. Mencía E. Janssen F.* DeepRED – rule extraction from deep neural networks // In: *Calders T., Ceci M., Malerba D. (eds) Discovery Science.*, 2016. T.9956. doi: http://dx.doi.org/10.1007/978-3-319-46307-0_29.
- [39] *Barakat N., Bradley A.* Rule extraction from support vector machines: A review // *Neurocomputing* , 2010. T.74. №1–3. C. 178–190. doi: <http://dx.doi.org/10.1016/j.neucom.2010.02.016>.
- [40] *So S., Richards W.* Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl)pyrimidines as DHFR inhibitors // *J Med Chem*, 1992. T.35. №17. C. 3201–3207. doi: <http://dx.doi.org/10.1021/jm00095a016>.
- [41] *Lek S.et.all.* Role of some environmental variables in trout abundance models using neural networks // *Aquat Living Resour*, 1996. T.9. №1. C. 23–29. doi: <http://dx.doi.org/10.1051/alr:1996004>.
- [42] *Selassie D. C.et.all.* Optimization of hydrophobic and hydrophilic substituent interactions of 2,4-diamino-5-(substituted-benzyl)pyrimidines with dihydrofolate reductase // *J Med Chem*, 1991. T.34. №1. C. 46–54. doi: <http://dx.doi.org/10.1021/jm00105a008>.

- [43] *Iooss B., Saltelli A.* Introduction to sensitivity analysis // In Handbook of Uncertainty Quantification, Ghanem, R.; Higdon, D.; Owhadi, H., Eds. Springer International Publishing: Cham, 2016. С. 1–20. doi: http://dx.doi.org/10.1007/978-3-319-11259-6_31-1.
- [44] *Адылова Ф. Т., Давронов Р. Р., Кузиев Б. Н.* Методы отбора дескрипторов в моделировании отношения структура-активность (QSAR) // Кимёвий технология назорат ва бошқарув, 2017. № 1. С. 82–92.
- [45] *Strobl C.et.all.* Bias in random forest variable importance measures: Illustrations, sources and a solution // BMC Bioinf, 2007. Т. 8. № 25. doi: <http://dx.doi.org/10.1186/1471-2105-8-25>.
- [46] *Strobl C.et.all.* Conditional variable importance for random forests // BMC Bioinf, 2008. Т. 9. № 307. doi: <http://dx.doi.org/10.1186/1471-2105-9-307>.
- [47] *Aoyama T., Ichikawa H.* Neural networks as nonlinear structure-activity relationship analyzers. Useful functions of the partial derivative method in multilayer neural networks // J Chem Inf Comput Sci, 1992. Т. 32. № 5. С. 492–500. doi: <http://dx.doi.org/10.1021/ci00009a015>.
- [48] *Baskin I. I.et.all.* An approach to the interpretation of backpropagation neural network models in QSAR studies // SAR QSAR Environ Res, 2002. Т. 13. № 1. С. 35–41. doi: <http://dx.doi.org/10.1080/10629360290002073>.
- [49] *Marcou G.et.all.* Interpretability of SAR/QSAR models of any complexity by atomic contributions // Mol Inf, 2012. Т. 31. № 9. С. 639–642. doi: <http://dx.doi.org/10.1002/minf.201100136>.
- [50] *An Y., Sherman W., Dixon S.* Kernel-based partial least squares: application to fingerprint-based QSAR with model visualization // J Chem Inf Model, 2013. Т. 53. № 9. С. 2312–2321. doi: <http://dx.doi.org/10.1021/ci400250c>.
- [51] *Chen H.et.all.* Beyond the scope of Free-Wilson analysis: building interpretable QSAR models with machine learning algorithms // J Chem Inf Model, 2013. Т. 53. № 6. С. 1324–1336. doi: <http://dx.doi.org/10.1021/ci4001376>.
- [52] *Stålring J.et.all.* Localized heuristic inverse quantitative structure activity relationship with bulk descriptors using numerical gradients // J Chem Inf Model, 2013. Т. 53. № 8. С. 2001–2017. doi: <http://dx.doi.org/10.1021/ci400281y>.
- [53] *Kuz'min V. E.et.all.* Interpretation of QSAR models based on random forest method // Mol Inf, 2011. Т. 30. № 6–7. С. 593–603. doi: <http://dx.doi.org/10.1002/minf.201000173>.
- [54] *Riniker S., Landrum G.* Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods // J Cheminf, 2013. Т. . № 43. doi: <http://dx.doi.org/10.1186/1758-2946-5-43>.
- [55] *Polishchuk P. G. et.all.* Universal approach for structural interpretation of QSAR/QSPR models // Mol Inf, 2013. Т. 32. № 9–10. С. 843–853. doi: <http://dx.doi.org/10.1002/minf.201300029>.
- [56] *Pérez-Garrido A.et.all.* Latest QSAR study of adenosine A2B receptor affinity of xanthines and deazaxanthines // Mol Diversity, 2015. Т. 19. № 4. С. 975–989. doi: <http://dx.doi.org/10.1007/s11030-015-9608-0>.
- [57] *Zhang Y. Y.et.all.* Integrating in silico and in vitro approaches to predict drug accessibility to the central nervous system // Mol Pharmaceutics, 2016. Т. 13. № 5. С. 1540–1550. doi: <http://dx.doi.org/10.1021/acs.molpharmaceut.6b00031>.
- [58] *Krysko A. A.et.all.* Synthesis, biological evaluation and molecular docking studies of 2-piperazin-1-yl-quinazolines as platelet aggregation inhibitors and ligands of integrin

- α Иб β 3 // Bioorg Med Chem Lett, 2016. Т. 26. № 7. С. 1839–1843. doi: <http://dx.doi.org/10.1016/j.bmcl.2016.02.011>.
- [59] *Sushko Y.et.all.* Prediction driven matched molecular pairs to interpret QSARs and aid the molecular optimization process // J Cheminf, 2014. Т. 6. № 48. doi: <http://dx.doi.org/10.1186/s13321-014-0048-0>.
- [60] *Kramer C.et.all.* Matched molecular pair analysis: significance and the impact of experimental uncertainty // J Med Chem, 2014. Т. 57. № 9. С. 3786–3802. doi: <http://dx.doi.org/10.1021/jm500317a>.
- [61] *Tulyaganov N.* Pharmacological studies of alkaloids of Peganum Harmala L. quinazoline and quinazolone structure and their derivatives // The dissertation author's abstract on the scientific degree of Doctor of Chemical Sciences, Moscow, 1981. doi: <http://dx.doi.org/10.1016/j.fitote.2017.05.001>.
- [62] *Адъилова Ф. Т. et.all.* Прогноз ЛД₅₀ активности алкалоидов гармалы обыкновенной хиназолинового, хиназолонного строения и их производных на основе QSAR моделей // Кимёвий технология назорат ва бошқарув, 2015. № 5. С. 21–27.
- [63] *Адъилова Ф. Т. et.all.* Ошибки в QSAR моделировании // Журнал теоретической и клинической медицины, 2015. № 5. С. 140–143.
- [64] *Адъилова Ф. Т. et.all.* Сравнительный анализ разработки моделей структура-активность (QSAR) для ряда дитерпеновых алкалоидов: традиционный и новый инструментарий // Тошкент тиббиёт академияси ахборотномаси, 2016. № 2. С. 44–48.
- [65] *Адъилова Ф. Т. et.all.* Оценка эффективности подхода kNN-QSAR моделирования в хемоинформатике // Mol Inf, 2017. № 4. С. 3–6.
- [66] cxcalc, version 5.4; Chemaxon: Budapest, Hungary // ,
- [67] *Wawer M., Bajorath J.* Local structural changes, global data views: graphical substructure-activity relationship trailing // J Med Chem, 2011. Т. 54. № 8. С. 2944–2951. doi: <http://dx.doi.org/10.1021/jm200026b>.
- [68] *Maggiore G., Bajorath J.* Chemical space networks: a powerful new paradigm for the description of chemical space // J Comput-Aided Mol Des, 2014. Т. 28. № 8. С. 795–802. doi: <http://dx.doi.org/10.1007/s10822-014-9760-0>.

Поступила в редакцию 02.08.2018

UDC 577.29

PRACTICE AND POTENTIAL FOR THE DEVELOPMENT OF THE INTERPRETABILITY OF QUANTITATIVE "STRUCTURE-ACTIVITY" MODELS (QSAR)

Adilova F. T., Davronov R. R., Jamilov U. U., Kayumov O. A.
fatadilova@gmail.com; rifqat@gmail.com, jamilovu@gmail.com,
oybekonlineuz@mail.ru

Institute of Mathematics named after V. I. Romanovskiy NAS RUz, 81 Mirzo Ulugbek st.,
Tashkent, Uzbekistan

The advantage of the QSAR model is the revealing of the action mechanism being studied, so the problem of the interpretability of the model is very essential. Today there are thousands of molecular descriptors available for constructing QSAR models, but it is often impossible to give a physic-chemical interpretation to many of them. Until recently,

the "model-descriptor-structure"interpretation scheme was used, depending on the set of descriptors and machine learning methods, which made it difficult to interpret models based on the advanced methods of machine learning. The solution of this problem was the new "model-structure"paradigm, which provides understanding of the meaning of any QSAR model. The article analyzes the pros and cons of approaches to interpretation, and in the computational running, the effectiveness of a new universal approach to the interpretation of QSAR models is justified.

Keywords: structure-activity model, model interpretability, machine learning

Citation: Adilova F. T., Davronov R. R., Jamilov U. U., Kayumov O. A. 2018. Practice and potential for the development of the interpretability of quantitative "structure-activity"models (QSAR). *Problems of Computational and Applied Mathematics*. 5(17):7–26.