

ОТЗЫВ официального оппонента
на диссертацию на соискание ученой степени
кандидата химических наук Осипенко Сергея Владимировича
на тему: «Прогнозирование хромато-масс-спектрометрических
характеристик химических соединений в нецелевом анализе с
применением методов машинного обучения»
по специальности 1.4.2 Аналитическая химия

Диссертация подготовлена автором в Сколковском институте науки и технологий. Диссертация представлена на соискание ученой степени кандидата химических наук по специальности 1.4.2 Аналитическая химия (химические науки) в Диссертационный совет МГУ 014.5.

В диссертационной работе предложены подходы к оценке времен/индексов удерживания в жидкостной/газовой хромато-масс-спектрометрии, а также предложен способ прогнозирования масс-спектров электронной ионизации. Разработанные подходы основаны на применении машинного обучения, точность их оценивалась с помощью кросс-валидации, а также с использованием независимых выборок. В работе рассматривается совместное использование предложенных моделей и метода изотопного обмена $^{16}\text{O}/^{18}\text{O}$, для чего изучена селективность последнего. Результаты работы могут быть использованы для нецелевого химического анализа.

Диссертационная работа включает введение, шесть глав, заключение, выводы, списки цитируемой литературы и сокращений. Объем работы составляет 163 страницы, включая 57 рисунков, 27 таблиц и одно приложение.

Во введении рассмотрена актуальность проблемы прогнозирования хромато-масс-спектрометрических характеристик, поставлена цель и задачи диссертационного исследования. Введение также содержит формулировку научной новизны, практической значимости. Изложены положения, выносимые на защиту, сведения об апробации работы. Указан личный вклад

автора, а также соответствие паспорту специальности, обоснована степень достоверности результатов.

Первая глава является обзором литературы, посвященной прогнозированию времен и индексов хроматографического удерживания, а также масс-спектральных характеристик и характеристик, измеряемых в спектрометрии ионной подвижности. Дана краткая характеристика основных методов машинного обучения. Рассмотрены примеры применения предсказанных значений аналитических характеристик для идентификации химических соединений в нецелевом анализе.

Вторая глава включает описание использованных реагентов, материалов и оборудования. Детально описаны условия хромато-масс-спектрометрического анализа образцов, приведен список использованного программного обеспечения и библиотек.

Третья глава описывает три подхода для предсказания времен удерживания в жидкостной хроматографии на основе различных алгоритмов машинного обучения и представлений молекул. Предложено использование метода «обучения с переносом» для получения оценок в различных условиях разделения. Наиболее точная из моделей характеризовалась средним отклонением порядка 30 с, при этом при переносе оценок на различные системы точность ожидаемо снижалась.

В четвертой главе разработанные модели прогнозирования времен удерживания в жидкостной хроматографии использовали совместно с методом изотопного обмена $^{16}\text{O}/^{18}\text{O}$, для чего изучена селективность последнего и предложено программное обеспечение для обработки соответствующих данных. Применение совместной фильтрации по данным изотопного обмена и предсказанным временам удерживания в модельном эксперименте показало сокращение пространства поиска при идентификации в среднем более чем на 70%.

В пятой главе предложены подходы к прогнозированию газохроматографических индексов удерживания применительно к

соединениям из списков Конвенции по запрещению химического оружия. Предсказания отличаются хорошей точностью, которую дополнительно увеличить при использовании инкрементного метода оценки.

В главе 6 рассмотрено прогнозирование масс-спектров электронной ионизации методами машинного обучения. В работе создана модель на основе алгоритма градиентного бустинга, которая характеризовалась удовлетворительным сходством предсказанных и измеренных спектров со средней косинусной мерой более 0.75. В качестве основы для обучения модели использовали данные из библиотеки NIST 20.

Хромато-масс-спектрометрия является одним из наиболее информативных методов анализа многокомпонентных образцов. Идентификация соединений, входящих в состав образца, требует проведения измерений стандартного образца, или по крайней мере совпадения измеренных хроматографических и масс-спектральных характеристик рассматриваемого компонента со значениями из специализированных баз данных. Однако для большинства известных химических соединений стандартные образцы труднодоступны, и даже наиболее полные базы масс-спектров предоставляют справочные значения лишь для ограниченного числа веществ. Поэтому задача оценки хромато-масс-спектрометрических характеристик с помощью вычислительных методов представляется весьма **актуальной**. Построенные в рамках работы модели машинного обучения являются новыми и составляют **научную новизну** работы. Основная **практическая значимость** состоит в повышении информативности нецелевого химического анализа с использованием предложенных автором подходов. **Научные положения, выводы и рекомендации, сформулированные в диссертации**, получены с использованием современного хромато-масс-спектрометрического оборудования и программного обеспечения. Оценку результатов проводили в соответствии с общепринятыми подходами с использованием кросс-валидации и независимых выборок, сформулированные цели исследования достигнуты,

задачи диссертационного исследования решены. **Личный вклад** автора сомнений не вызывает. По результатам работы опубликовано 6 статей. Статьи опубликованы в ведущих отраслевых международных изданиях, в том числе в таких изданиях как *Analytical and Bioanalytical Chemistry*, *Journal of the American Society for Mass Spectrometry*, *Journal of Chromatography A*, *The Analyst*. Результаты работы прошли апробацию на российских и международных конференциях. Таким образом, диссертационная работа С.В. Осипенко является законченным исследованием, выполненным на высоком научном уровне.

Тем не менее, по работе имеется ряд вопросов и замечаний:

1. Автор не использует для идентификации одновременно масс-спектрометрические и хроматографические данные, что представляется важным при идентификации изомеров.
2. Предложенные автором вычислительные подходы реализованы в виде исходного кода приложений. У квалифицированных химиков-аналитиков использование подобных приложений может вызвать определенные затруднения. Разработка интерфейса или подробных инструкций по использованию подходов может упростить практическое внедрение разработанных подходов.
3. Может ли рассмотренный в главе 4 метод изотопного обмена $^{16}\text{O}/^{18}\text{O}$ быть применен в сочетании с газовой хромато-масс-спектрометрией, и соответствующими моделями, предложенными автором по прогнозированию масс-спектров электронной ионизации, и индексов удерживания?

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В.Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 1.4.2 Аналитическая химия (по химическим наукам), а также критериям, определенным Положением о присуждении

ученых степеней в Московском государственном университете имени М.В.Ломоносова.

Таким образом, соискатель Осипенко Сергей Владимирович заслуживает присуждения ученой степени кандидата химических наук по специальности 1.4.2 Аналитическая химия.

Официальный оппонент:

доктор химических наук, профессор, член-корреспондент РАН
директор Федерального государственного бюджетного учреждения науки
Институт физической химии и электрохимии им. А.Н. Фрумкина Российской
академии наук

БУРЯК Алексей Константинович

04 марта 2024 г.

Контактные данные:

тел.: 7(495)9554487, e-mail: dir@phycbe.ac.ru

Специальность, по которой официальным оппонентом
защищена диссертация:
02.00.04 – Физическая химия

Адрес места работы:

119071, Москва, Ленинский проспект, д. 31 корп. 4.,
Федеральное государственное бюджетное учреждение науки Институт
физической химии и электрохимии им. А.Н. Фрумкина Российской академии
наук
Тел.: 7(495)9554487; e-mail: dir@phycbe.ac.ru

Подпись А.К. Буряка удостоверяю:
Секретарь Учёного совета Института

к.х.н.


И.Г. Варшавская
ПОДПИСЬ
М.П.