

УДК 543.42.062

DOI: 10.17223/00213411/68/11/3

Нейросетевая предиктивная модель для количественного анализа данных абсорбционной ИК-спектроскопии атмосферных газов*

Д.А. Вражнов¹, А.И. Князькова¹, А.К. Третьяков¹, С.Э. Шипилов¹, Ю.В. Кистенев¹

¹ *Национальный исследовательский Томский государственный университет, г. Томск, Россия*

Предложены предиктивные модели для оценки концентрации малых газовых примесей атмосферы на основе обучающей выборки модельных данных абсорбционной ИК-спектроскопии. Спектры целевых газов взяты из базы HITRAN. Исследованы следующие методы машинного обучения: регрессии опорных векторов и на основе случайного леса, LASSO, искусственная нейронная сеть прямого распространения. Предложен новый конвейер машинного обучения на основе искусственной нейронной сети, превосходящий по точности остальные методы. Полученные модели могут быть использованы для оценки концентрации целевых газов в атмосфере по данным абсорбционной ИК-спектроскопии.

Ключевые слова: ИК-спектроскопия атмосферы, машинное обучение, количественный анализ, малые газовые составляющие атмосферы, промышленные выбросы.

Введение

Промышленные поллютанты антропогенного происхождения, попадая в атмосферный воздух, водные ресурсы и почвенный покров, оказывают прямое негативное воздействие на здоровье населения. Высокие концентрации опасных соединений коррелируют с ростом риска возникновения ряда патологий, а также со снижением показателей продолжительности и качества жизни [1]. Примерами технологических газовых загрязнителей, образующихся при сжигании ископаемых видов топлива, являются CO, CO₂, CH₄, SO₂, NO₂, N₂O [2, 3]. Примером промышленных загрязнителей, эмитируемых сельскохозяйственными предприятиями, является карбонилсульфид (COS) [4].

Разработка экономичных и простых в применении методов детекции и количественного определения таких веществ представляет собой актуальную научно-практическую задачу. Одним из перспективных направлений в данной области является абсорбционная ИК-спектроскопия [5]. Одной из проблем количественного анализа данных ИК-спектроскопии атмосферы является перекрытие линий поглощения различных компонент. При этом задача разрешения спектральных кривых становится крайне трудной из-за наличия множества решений, сильной корреляции между значениями интенсивности сигнала на разных частотах. Для решения таких задач перспективны методы машинного обучения (МО), позволяющие предсказать по снятому спектру смеси и известным спектрам чистых веществ концентрацию отдельной компоненты смеси [6]. Однако вопрос выбора оптимального метода с точки зрения точности нахождения концентрации малых газовых компонент остается открытым.

В данной работе были исследованы следующие методы МО: регрессия опорных векторов, регрессия на основе случайного леса, Least Absolute Shrinkage and Selection Operator (LASSO), искусственная нейронная сеть прямого распространения (табл. 1).

Таблица 1

Преимущества и недостатки исследованных методов МО

Название метода	Преимущества	Недостатки
Регрессия на основе случайного леса (РСЛ) [7, 8]	Ансамбли регрессоров позволяют достигать высокой точности моделей и обладают хорошей робастностью к выбросам и шуму. Возможность работы с нелинейными зависимостями между входными и выходными данными. Встроенная оценка важности входных данных. Метод позволяет работать с мультиколлинеарными данными	Высокие вычислительные затраты при больших данных. Тенденция к переобучению на шумных данных. Чувствительность к несбалансированному данным. Требует оптимизации гиперпараметров

* Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации (грант № 075-15-2024-557 от 25.04.2024 г.).