

УДК 004.896

Разработка виртуального анализатора качества нефтепродуктов на основе методов машинного обучения

Кобаченко Александра Андреевна

Мурашов Михаил Владимирович (научный руководитель)

МГТУ им. Н.Э. Баумана, Москва, Россия

Рассмотрена задача разработки виртуального анализатора для контроля содержания общей серы в бутан-бутиленовой фракции в процессе очистки сжиженных углеводородных газов. Актуальность обусловлена необходимостью повышения точности и оперативности контроля качества продукции в условиях высокой динамичности технологического процесса и ограниченности традиционных лабораторных методов анализа. Виртуальные анализаторы, построенные на основе алгоритмов машинного обучения, представляют собой математические модели, способные по текущим параметрам с датчиков прогнозировать значения показателей качества. Целью работы является построение и сравнительная оценка моделей виртуального анализа с применением современных методов машинного обучения. Для объективной оценки эффективности моделей использованы унифицированные метрики качества: средняя абсолютная ошибка и коэффициент детерминации.

Ключевые слова: виртуальный анализатор, машинное обучение, контроль качества, нефтепереработка, моделирование

Введение. Современные технологии переработки углеводородного сырья, в том числе процессов очистки сжиженных углеводородных газов (СУГ), предъявляют повышенные требования к точности и оперативности контроля качества продукции [1]. Одним из критически важных показателей качества в данных процессах является содержание общей серы, превышение которой приводит к ускоренной коррозии оборудования, отравлению катализаторов и нарушению экологических регламентов.

На практике контроль содержания серы осуществляется с использованием лабораторных анализов, которые обладают высокой точностью, но сопряжены с существенными временными задержками, невозможностью непрерывного мониторинга и дополнительными затратами на обслуживание. В условиях высокой динамичности технологического процесса подобные ограничения становятся критическим фактором, препятствующим эффективной оптимизации режимов работы.

В связи с этим в последние годы возрос интерес к разработке виртуальных анализаторов (ВА) — программно-математических моделей, способных в режиме реального времени на основе текущих значений измеряемых параметров (расход, давление, температура и др.) прогнозировать ключевые показатели качества [2].

Наряду с классическими методами регрессии в работе предлагается использование ансамблевых алгоритмов и глубоких нейронных сетей, производится их оценка и сравнение, рассматриваются их преимущества и ограничения. Применение методов машинного обучения (МО) позволяет моделировать сложные нелинейные зависимости между входными и выходными параметрами технологической системы, обеспечивая высокую точность при отсутствии необходимости в дополнительном физическом оборудовании [3].

Обзор существующих методов контроля качества. В России для контроля серы в СУГ применяют как лабораторные, так и поточные анализаторы. Лабораторные методы (энергодисперсионная рентгеновская флуоресценция — (ЭД-РФА), УФ-флуоресценция, хемилюминесценция) дают высокую точность (предел обнаружения до 0,1 ppm серы), но требуют отбора и подготовки проб и занимают некоторое время [4, 5]. Например, анализ одной пробы на занимает порядка 3–10 минут. Дополнительно: пробоотбор, доставка, подготовка, ввод в прибор, охлаждение и очистка. Итоговое время полного цикла может достигать 60 минут. Также лабораторные анализы требуют наличие обслуживающего персонала, работы в несколько смен, затрат на реагенты, пробоотборные устройства.

Поточные анализаторы дают данные в режиме реального времени: обновление результатов происходит каждые 2–6 мин, а время отклика — менее 100 с. Такие системы позволяют непрерывно следить за содержанием серы, но требуют регулярной дорогостоящей калибровки и дороже в обслуживании. Их чувствительность обычно ниже по сравнению с лабораторными установками. Стоимость установки: от 5 до 20 млн рублей, не считая затрат на эксплуатацию.

Появляется необходимость в математических моделях, которые по текущим технологическим параметрам могут предсказать содержание серы, обеспечивая при этом почти мгновенную оценку без проведения анализа. Для данного технологического процесса необходима модель, способная «понимать» нелинейные связи между переменными процесса и качеством продукта. Решение — разработка виртуальных анализаторов.

Описание предлагаемого решения. Процесс очистки СУГ нелинеен и динамичен. Простые линейные регрессии (например, метод наименьших квадратов) не учитывают сложные взаимодействия переменных и инерционность системы, поэтому плохо описывают динамику серы. Появляется необходимость в использовании нейросетевых моделей. Современные методы МО способны улавливать нелинейные зависимости и поэтому именно они являются перспективным решением для виртуальных анализаторов серы в сложных нефтеперерабатывающих процессах [6].

В данной работе рассматриваются три подхода к построению ВА для оценки содержания серы в процессе очистки СУГ: модель взвешенных наименьших квадратов (WLS) в качестве базовой модели для сравнения, ансамблевая модель (WLS + Random Forest) и модель на основе глубокой

нейронной сети (DNN). Все модели обучались на исторических данных с лабораторными измерениями, собранных в реальном производственном цикле.

Входными параметрами модели выступают три технологических параметра:

- x_1 — расход щелочного раствора на входе;
- x_2 — давление в регенераторе щелочи;
- x_3 — расход гидроочищенного бензина.

Первая модель реализует базовый метод, чаще всего использующийся для построения ВА на реальных производствах, — метод взвешенных наименьших квадратов (WLS). Основное преимущество данной модели заключается в ее интерпретируемости и устойчивости к выбросам за счет применения индивидуальных весов к наблюдениям [7]. Наблюдения, по которым хотя бы один признак отклонялся от среднего более чем на три стандартных отклонения (т. е. $|z| > 3$), получили пониженные веса.

Согласно результатам обучения на исторических данных, итоговое регрессионное уравнение для прогнозирования содержания серы имеет вид

$$y = 725,78 - 58,23 \cdot x_1 - 454,03 \cdot x_2 - 0,13 \cdot x_3.$$

График соответствия лабораторных данных и прогнозных значений модели WLS представлен на рис. 1.

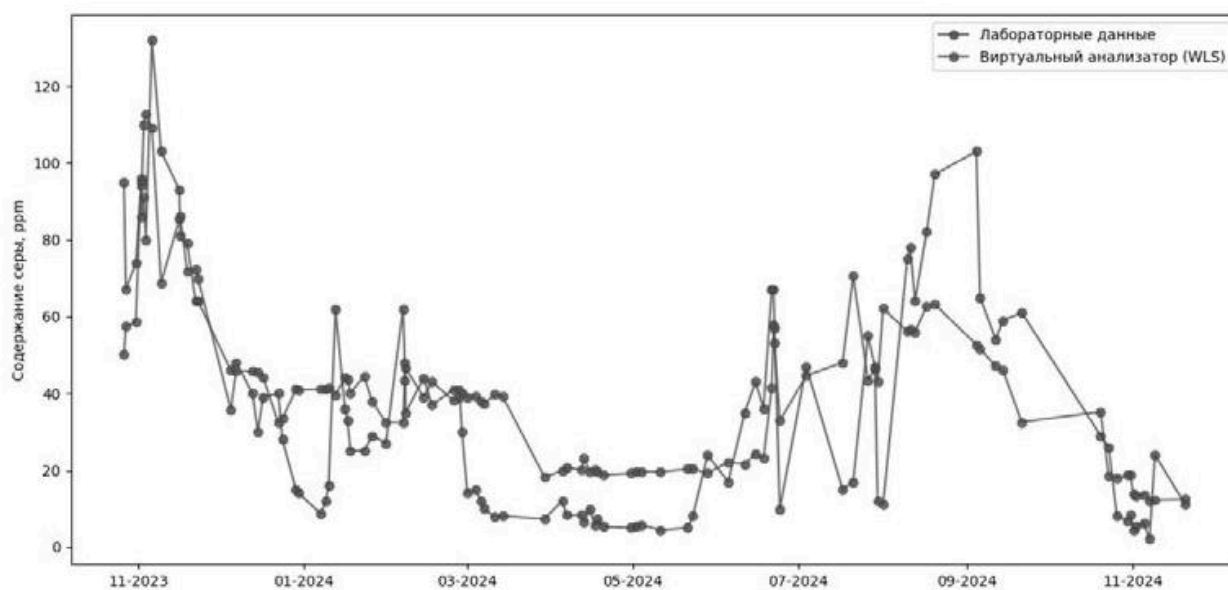


Рис. 1. Результат работы модели WLS

Из рис. 1 можно сделать вывод, что модель WLS продемонстрировала ограниченную точность: средняя абсолютная ошибка составила 16,71 ppm, коэффициент детерминации $R^2 = 0,55$. Это значит, что линейная аппроксимация действительно не способна в полной мере отразить сложные нелинейные зависимости между параметрами процесса и выходным значением серы, необходимо использование методов МО.

Для повышения точности и устойчивости ВА была реализована ансамблевая модель. В качестве базовых компонентов ансамбля были использованы WLS-модель и модель случайного леса (Random Forest, RF) — ансамблевый метод построения деревьев решений, устойчивый к переобучению и способный моделировать сложные нелинейные зависимости.

Пусть $\hat{y}_{WLS}$ — прогноз, полученный с использованием модели WLS, $\hat{y}_{RF}$ — прогноз, полученный с помощью модели Random Forest, $\hat{y}_{ens}$ — итоговый ансамблевый прогноз. Тогда итоговый прогноз ансамбля вычисляется по формуле

$$\hat{y}_{ens} = \alpha \cdot \hat{y}_{WLS} + (1 - \alpha) \cdot \hat{y}_{RF}.$$

Гиперпараметр α был выбран эмпирически для достижения баланса между интерпретируемостью и точностью.

Графики (рис. 2) отображают сравнение лабораторных данных и прогнозов ансамблевой модели.

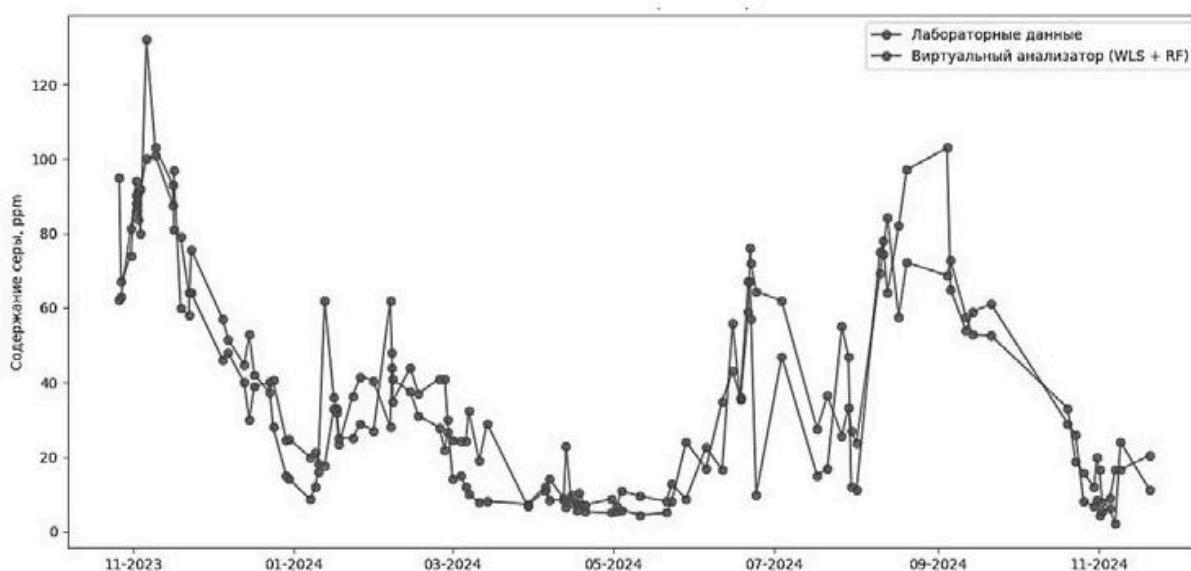


Рис. 2. Результат работы ансамблевой модели

Из рис. 2 можно сделать вывод, что ансамбль продемонстрировал заметное улучшение по сравнению с WLS: средняя абсолютная ошибка составила 10,77 ppm, $R^2 = 0,77$. Такой прирост точности свидетельствует о высокой значимости нелинейных эффектов в процессе гидроочистки СУГ.

Также был разработан виртуальный анализатор на основе модели глубокой нейронной сети (DNN). Модель построена как последовательная нейронная сеть с тремя скрытыми слоями:

- первый скрытый слой: 128 нейронов, функция активации ReLU, L2-регуляризация с коэффициентом 0,001, пакетная нормализация и дропаут 0,3 для борьбы с переобучением;
- второй скрытый слой: 64 нейрона, ReLU, L2-регуляризация с коэффициентом 0,01, пакетная нормализация, дропаут 0,2;

- третий скрытый слой: 32 нейрона, ReLU;
- выходной слой: 1 нейрон, линейная активация (регрессия).

Финальная формула предсказания не имеет аналитического выражения, так как нейросеть аппроксимирует зависимость в виде комплексной нелинейной комбинации весов и активаций

$$\hat{y} = f_{\text{DNN}}(X) = W_3 \cdot \phi(W_2 \cdot \phi(W_1 \cdot X + b_1) + b_2) + b_3,$$

где ϕ — функция активации ReLU; W_i, b_i — веса и смещения на соответствующих слоях.

Оптимизация функции потерь производилась с помощью алгоритма Adam с адаптивным изменением скорости обучения и использованием механизмов ранней остановки для предотвращения переобучения. Количество эпох обучения — до 100, с ранней остановкой при отсутствии улучшений.

На рисунке 3 представлены сравнительные графики прогнозируемых значений и лабораторных измерений.

Полученные метрики качества на тестовой выборке:

- средняя абсолютная ошибка: 9,20 ppm;
- коэффициент детерминации (R^2): 0,82.

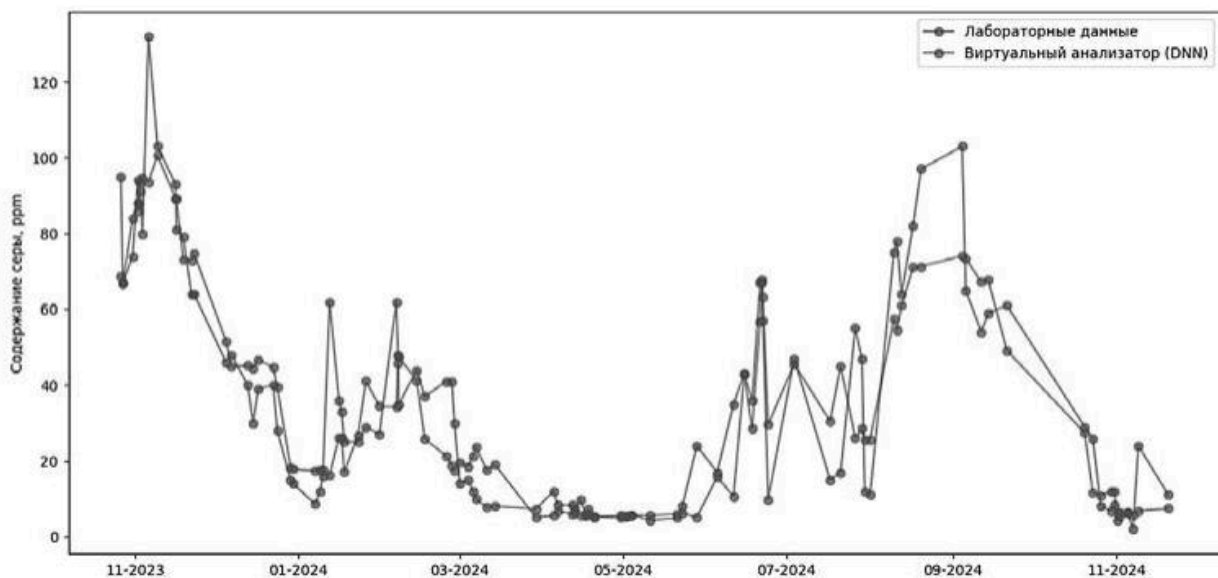


Рис. 3. Результат работы модели на основе глубокой нейронной сети

Данные результаты свидетельствуют о наилучшей точности среди рассмотренных моделей. Однако более высокая вычислительная нагрузка и необходимость тщательной настройки гиперпараметров требуют дальнейших исследований для промышленного внедрения.

Заключение. В настоящей работе решена задача построения ВА содержания серы в потоке после блока очистки СУГ. Результаты тестирования продемонстрировали, что линейная модель не обеспечивает требуемой точности. Использование ансамблевой схемы позволило существенно повысить

точность предсказания за счет повышения устойчивости к выбросам и более гибкого аппроксимирования зависимостей. Наибольшей точностью обладает модель на основе глубокой нейронной сети, что подтверждает целесообразность использования современных методов глубокого обучения для моделирования виртуальных анализаторов. Но данная модель имеет ограничение в виде более высокой вычислительной нагрузки.

Разработанный виртуальный анализатор позволяет оперативно и непрерывно оценивать содержание серы в продукте, что способствует снижению зависимости от лабораторных измерений, минимизации технологических рисков и повышению общей эффективности системы управления качеством.

В перспективе планируется реализация адаптивных механизмов перенастройки под изменяющиеся условия производства, а также интеграция модели в контур реального времени в составе АСУ ТП.

Литература

- [1] Капустин В.М., Рудин М.Г. *Химия и технология нефтепереработки*. Москва, Химия, 2013, 496 с.
- [2] Шевлягина С.А., Торгашов А.Ю. *Методы построения виртуальных анализаторов в системах усовершенствованного управления технологическими процессами*. Владивосток, Изд-во Дальневост. федерал. ун-та, 2024, 86 с.
- [3] Rožanec J., Trajkova E., Jinzhi L., Sarantinoudis N., et al. Cyber-Physical LPG Debutanizer Distillation Columns: Machine-Learning-Based Soft Sensors for Product Quality Monitoring. *Applied Sciences*, 2021, vol. 11, art. no. 11790.
- [4] ГОСТ 32139–2019. *Нефть и нефтепродукты. Определение содержания серы методом энергодисперсионной рентгенофлуоресцентной спектроскопии*. Москва, Стандартинформ, 2019, 24 с.
- [5] ГОСТ Р 56866–2016. *Углеводородные газы (газообразные и сжиженные). Определение содержания серы методом ультрафиолетовой флуоресценции*. Москва, Стандартинформ, 2016, 19 с.
- [6] Tan Q., Li B. Soft Sensor Modeling Method for Sulfur Recovery Process Based on Long Short-Term Memory Artificial Neural Network (LSTM). *E3S Web of Conferences*, 2023, 406 p. <https://doi.org/10.1051/e3sconf/202340604028>
- [7] Кобаченко А.А., Паротькина М.А., Егоров С.В., Волынчиков М.А. Предобработка данных, поступающих на виртуальный анализатор нефтеперерабатывающего предприятия. *Робототехника и искусственный интеллект. XVI Всерос. науч.-техн. конф. с междунар. участ.: материалы*. Красноярск, ЛИТЕРА-принт, 2024, с. 130–133.

Development of a soft sensor for petroleum product quality based on machine learning

Kobachenko Aleksandra Andreevna

Murashov Mikhail Vladimirovich (*scientific advisor*)

BMSTU, Moscow, Russia

This paper addresses the development of a soft sensor (virtual analyzer) for monitoring the total sulfur content in the butane-butylene fraction. The relevance of this task is driven by the need for improved accuracy and real-time quality control in the context of highly dynamic technological processes and the limitations of traditional laboratory-based measurements. Soft sensors based on machine learning algorithms are mathematical models capable of predicting quality indicators from current process parameters.

Keywords: soft sensor, virtual analyzer, machine learning, quality control, oil refining, process modeling, regression analysis