

М. Е. Маркелов (соиск.), Р. Ф. Ахметов (к.т.н., доц.),
А. Р. Набиева (к.т.н., доц.), Н. А. Шамова (к.т.н., доц.)

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПРЕДЕЛЕНИЯ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ПРОДУКЦИИ УСТАНОВКИ ГИДРООЧИСТКИ ДИЗЕЛЬНОГО ТОПЛИВА

Уфимский государственный нефтяной технический университет,
кафедра технологии нефти и газа
450064, г. Уфа, ул. Космонавтов, 1; e-mail: albena90@mail.ru

M. E. Markelov, R. F. Akhmetov, A. R. Nabieva, N. A. Shamova

MACHINE-LEARNING MATHEMATICAL MODEL FOR MEASUREMENTS OF PRODUCT QUALITY OF DIESEL FUEL HYDROTREATMENT PLANT

Ufa State Petroleum Technological University
1, Kosmonavtov Str., 450064, Ufa, Russia; e-mail: albena90@mail.ru

Рассмотрена задача повышения качества продукции установки гидроочистки дизельного топлива с помощью виртуального анализатора на основе модели машинного обучения. Разработан виртуальный анализатор содержания серы в продукте установки гидроочистки дизельного топлива на основе модели машинного обучения, представлена оценка его точности и проведено сравнение результатов полученной модели с предварительными оценками виртуального анализатора системы усовершенствованного управления технологическим процессом, внедренного на установке гидроочистки дизельного топлива Л-24-7.

Ключевые слова: виртуальный анализатор; качество продукции; математическое моделирование; машинное обучение; процесс гидроочистки дизельного топлива; регрессионный анализ; система усовершенствованного управления технологическим процессом; цифровизация производственных процессов.

Новые интеллектуальные системы способны мгновенно анализировать несколько сотен параметров и в зависимости от показаний виртуальных анализаторов автоматически регулировать режимы работы установок в установленных пределах. Цифровой комплекс обеспечивает автоматическое поддержание наиболее оптимального режима работы технологических установок на НПЗ, что позволяет увеличить выход продукции, сни-

The problem of improving the quality of diesel fuel hydrotreating unit products using a virtual analyzer based on a machine learning model is considered. A virtual analyzer of sulfur content in the product of a diesel fuel hydrotreating unit based on a machine learning model is developed, its accuracy is assessed, and the results of the obtained model are compared with preliminary estimates of the virtual analyzer of the advanced process control system implemented at the L-24-7 diesel fuel hydrotreating unit.

Key words: APC – advanced process control; diesel fuel hydrotreatment plant; index of determination; machine-learning; mean absolute error; mean-squared error; regression analysis; scikit-learn; virtual analyzer.

зить потребление энергоресурсов и, соответственно, повысить экономическую эффективность нефтепереработки. Внедрение систем усовершенствованного управления включает в себя цифровизацию производственных процессов, предиктивную аналитику работающего оборудования, дополненную виртуальную реальность в обучении студентов и работе сотрудников, а также другие проекты оптимизации производства.

Дата поступления 18.09.24

Классическим решением задачи оптимального регулирования технологического процесса для обеспечения оптимального качества продукции является внедрение систем усовершенствованного управления технологическим процессом (СУ-УТП), в которой используются методы управления по прогнозной модели, а также применяются виртуальные анализаторы качества (ВА). ВА дает предварительную оценку качества продукции на основе значений технологического режима и свойств сырья, решая задачу оптимизации ведения технологического режима ¹. Для предварительной оценки качества продукции технологического процесса предлагается математическая модель на основе машинного обучения. В качестве примера технологического процесса рассмотрим процесс гидроочистки дизельного топлива, сущность которого состоит в превращении соединений, содержащих S, N, O на катализаторе под давлением водорода в летучие соединения, такие как сероводород, аммиак, и воду.

Материалы и методы

Основными параметрами, характеризующими процесс гидроочистки, являются:

- температура газо-сырьевой смеси на входе в реактор;
- давление водорода в реакторе гидроочистки;
- соотношение водород:сырье;
- концентрация водорода;
- объемная скорость подачи сырья в реактор;
- фракционный состав сырья.

Основным показателем качества продукции процесса гидроочистки является содержание серы в гидроочищенном дизельном топливе, которое по современным стандартам ЕВРО не должно превышать 10 ppm (К5 ГОСТ Р 32511) ².

Для построения математической модели необходимо произвести идентификацию процесса гидроочистки на исторических данных работы установки. В качестве исходных данных используется отчет по мониторингу работы катализатора установки гидроочистки дизельного топлива Л-24-7. База данных содержит 2600 ежесуточных наблюдений технологических параметров (температуры, давление, расход сырья, концентрацию водорода, содержание серы в сырье, фракционный состав сырья). Для подготовки данных к алгоритмам машинного обучения применяется пакет *Scikit-Learn* – в т.ч. для стандартизации и нормализации данных, кодирования категориальных переменных, выбор методов регрессии, настройки гиперпараметров модели с помощью поиска по сетке, вычисления величин, которые показывают производительность математической модели (ко-

эффициент детерминации, среднеквадратичную ошибку и др.).

Основным методом определения взаимосвязей параметров в данной работе явился метод регрессионного анализа *Random Forest Regression* – это сумма «решающих деревьев», каждое из которых является кусочно-константной функцией. Регрессионный анализ встречается во всех областях науки и техники, когда требуется восстановить связь между признаками. Таким образом, одной из основных целей регрессионного анализа является предсказание значения зависимой переменной с помощью независимых признаков, называемых также стохастическими регрессорами. Для достижения данной цели требуется построить модель, содержащую вклад каждого отдельного признака. Наиболее популярный пример регрессионной модели – линейная регрессия, в которой объясняемая переменная зависит линейно от каждого отдельного признака ³. Для оценки параметров линейной регрессии используется метод наименьших квадратов (МНК), впервые примененный Гауссом в 1795 году. Его идея проста – требуется минимизировать сумму квадратов ошибок модели. Однако МНК является неустойчивым в случае, если данные содержат экстремальные значения, так называемые «выбросы», то есть наблюдения, сильно отличающиеся от остальных. Поэтому возникла нужда в робастных (от англ. *robust* – надежный) оценках, которые были бы устойчивы к выбросам в задачах многомерной линейной регрессии ⁴. В качестве метрик оценивания параметров будем использовать MSE (*mean squared error*) подсчитывает среднее значение квадратов ошибок предсказаний модели и MAE (*mean absolute error*) аналогична MSE, вычисляет среднее значение модулей ошибок предсказаний ⁵.

Достоверность построенных моделей регрессии проверяется по критерию их адекватности случайному среднеквадратичному отклонению зависимой величины. Адекватность модели определяет соответствие модели исходным данным и статистическую значимость уравнения регрессии. Основными шагами интерпретации являются ⁶:

1. *Анализ знаков перед коэффициентами регрессии*. В данном случае определяется степень влияния входной переменной на отклик системы. Опираясь на опыт экспертов, а также на исследования в данной области, можно сделать вывод об адекватности модели и возможности предсказания процесса с ее помощью.

2. *Анализ значений коэффициентов регрессии*. При необоснованно больших или малых значениях можно сделать вывод, что были допущены ошибки в расчетах и модель не пригодна к использованию.

3. Анализ расчета выхода модели

При помощи графика можно оценить поведение модели и оценить ее точность и адекватность. Если модель не повторяет исходные данные, то она не пригодна для описания и предсказания поведения объекта. В табл. 1 представлены статистические оценки математической модели.

Таблица 1
Оценка качества математической модели

Параметр	Математическая модель
Средняя абсолютная ошибка	0.93
Среднеквадратичная ошибка	1.58
Коэффициент детерминации	0.41

На рис. 1–3 показано содержание серы в продукте математической модели и измеренный показатель качества с применением лабораторных методов.



Рис. 1. Математическая модель определения серы в гидроочищенном дизельном топливе (тестовый сегмент №1)

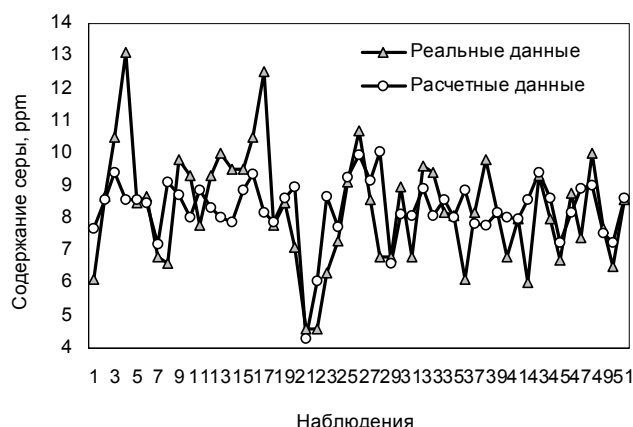


Рис. 2. Математическая модель определения серы в гидроочищенном дизельном топливе (тестовый сегмент №2)

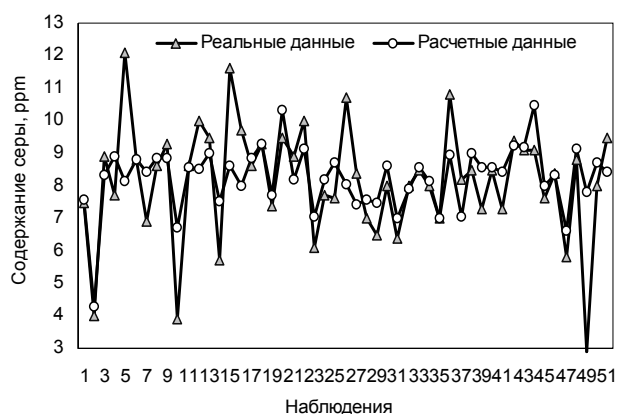


Рис. 3. Математическая модель определения серы в гидроочищенном дизельном топливе (тестовый сегмент №3)

Результаты исследования

В результате создания математической модели процесса гидроочистки дизельного топлива был выбран алгоритм для анализа содержания серы в продукте технологического процесса. Для подготовки данных к алгоритмам машинного обучения применялся пакет *Scikit-Learn*. Основным методом определения взаимосвязей параметров в данной работе явился метод регрессионного анализа: *Random Forest Regression*. Для оценки параметров линейной регрессии использовался метод наименьших квадратов (МНК). Достоверность построенной модели регрессии проверили по критерию ее адекватности случайному среднеквадратичному отклонению зависимой величины. При помощи графиков визуально оценили поведение модели и сделали оценку ее точности и адекватности. Модель повторяет исходные данные, что подтверждает ее пригодность для описания и предсказания поведения объекта. Наличие «выбросов» объясняется нестабильностью технологического режима реальной установки, а также сменой сырья, которое осуществляется по заданию производственно-диспетчерского отдела.

Одной из основных проблем процесса гидроочистки является снижение активности катализатора в следствии закоксовывания активных центров. С учетом ограничений по качеству продукта и плану производства это приводит к подъему температуры в системе реакторного блока и увеличению потребления энергоресурсов. Целью оптимизации технологического процесса является увеличение срока службы катализатора за счет снижения вовлечения продуктов вторичных процессов (ЛГ УЗК, ЛКГ КК) в сырье процесса гидроочистки, а также снижение температуры на входе в реактор при соблюдении границ по качеству продукта.

В рамках исследования технологического процесса гидроочистки дизельного топлива нами была проведена идентификация технологического процесса для получения математической модели на основе алгоритма машинного обучения при помощи применения методов регрессионного анализа исторических данных для предварительной оценки содержания серы в гидроочищенном ди-

зельном топливе. Модель была оценена при помощи статистических оценок и удовлетворяет требованием точности и адекватности. Дальнейшим этапом исследования является создание виртуального анализатора для определения значения содержания серы в продукте технологического процесса.

Литература

1. Хакимов Р.А. Идентификация математической модели процесса гидроочистки дизельного топлива для создания системы оптимизации группы технологических установок нефтеперерабатывающего завода // Омский научный вестник. – 2018. – №4(160). – С.174-178.
2. ГОСТ 32511-2013 (EN 590:2009) Топливо дизельное ЕВРО. Технические условия. – М.: Стандартинформ, 2019. – 15 с.
3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Вильямс, 2016. – 912 с.
4. Qadir M.F. Robust Method for Detection of Single and Multiple Outliers // Scientific Khyber. – 1996. – V.9. – Pp.135-144.
5. Дьяконов А.Г. Анализ малых данных // Известия РАН. Серия математическая. – 2012. – №76. – С.93-110.
6. Антонов В. А. Оценка адекватности регрессионной модели по погрешности экспериментальных измерений // Альманах современной науки и образования. – 2014. – №11(89). – С.27-32.

References

1. Khakimov R.A. *Identifikatsiya matematicheskoy modeli protsessa gidroochistki dizel'nogo topliva dlya sozdaniya sistemy optimizatsii gruppy tekhnologicheskikh ustanovok neftepererabatyvayushchego zavoda* [Identification of a mathematical model of the diesel fuel hydrotreating process for creating an optimization system for a group of process units at an oil refinery]. *Omskiy nauchnyi vestnik* [Omsk Scientific Bulletin], 2018, no.4(160), pp.174-178.
2. *GOST 32511-2013 (EN 590:2009) Toplivo dizel'noye EVRO. Tekhnicheskiye usloviya* [GOST 32511-2013 (EN 590: 2009) EURO diesel fuel. Specifications]. Moscow, Standartinform Publ., 2019, 15.
3. Dreyper N., Smit G. *Prikladnoy regressionnyy analiz* [Applied regression analysis]. Moscow, Vil'yams Publ., 2016, 912 p.
4. Qadir M. F. [Robust Method for Detection of Single and Multiple Outliers]. *Scientific Khyber*, 1996, vol.9, pp.135-144.
5. D'yakonov A.G. *Analiz malykh dannykh* [Small data analysis]. *Izvestiya RAN. Seriya matematicheskaya* [Bulletin of the Russian Academy of Sciences. Mathematical series], 2012, no.76, pp.93-110.
6. Antonov V. A. *Otsenka adekvatnosti regressionnoy modeli po pogreshnosti eksperimental'nykh izmereniy* [Assessment of the adequacy of the regression model by the error of experimental measurements]. *Al'manakh sovremennoy nauki i obrazovaniya* [Almanac of modern science and education], 2014, no.11(89), pp.27-32.