SOCAR

SOCAR Proceedings

Oil and gas refining. Petrochemistry

journal home page: http://proceedings.socar.az

ПРИМЕНЕНИЕ ПРОГНОЗИРУЮЩИХ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ К ПРОЦЕССАМ НЕФТЕПЕРЕРАБОТКИ В РАМКАХ ИНТЕЛЛЕКТУАЛЬНОЙ АВТОМАТИЗАЦИИ

В. В. Бухтояров[1,2], И. С. Некрасов[1], В.С. Тынченко[1,2,3], К.А. Башмур*[1], Р.Б. Сергиенко[4]

[1]Институт нефти и газа, Сибирский федеральный университет, Красноярск, Россия; [2]Центр НТИ «Цифровое материаловедение: новые материалы и вещества», МГТУ им. Н.Э. Баумана, Москва, Россия; [3]Институт информатики и телекоммуникаций, СибГУ им. М.Ф. Решетнева, Красноярск, Россия; [4]Gini Gmbh, Мюнхен, Германия

Application of machine learning algorithms for refining processes in the framework of intelligent automation

V.V. Bukhtoyarov[1,2], I.S. Nekrasov[1], V.S. Tynchenko[1,2,3], K.A. Bashmur*[1], R.B. Sergienko[4]

[1]Institute of Petroleum and Natural Gas Engineering, Siberian Federal University, Krasnoyarsk, Russia; [2]Digital Material Science: New Materials and Technologies, Bauman Moscow State Technical University, Moscow, Russia; [3]Institute of Computer Science and Telecommunications, Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia; [4]Gini Gmbh, Munich, Germany

ABSTRACT

The oil refining industry is facing several challenges and issues in data handling. A large amount of data is generated by many different processes and equipment. This article is devoted to methods for efficient analysis of large amounts of data in an oil refinery. In particular, the effectiveness of machine learning methods for predicting failures of process equipment in the hydrocracking process is investigated. Machine learning, as an important element of digitalization, allows us to successfully solve many production problems. The article describes the application of some machine learning algorithms for solving problems of classifying and predicting failures of hydrocracking process equipment that occur during oil refining and diesel fuel production. The application of random forest methods, principal component analysis and hyperparameter tuning is considered. The effectiveness of these methods is compared on the basis of the Accuracy parameter. It is shown that the combination of these methods will improve the accuracy of the model by 2%.

KEYWORDS

Automation;
Machine learning;
Hydrocracking;
Simulation;
Oil refinery.

© 2022 «OilGasScientificResearchProject» Institute. All rights reserved.

Введение

В настоящее время компании рассматривают возможность интеграции машинного обучения и искусственного интеллекта с традиционными роботизированными процессами для обеспечения интеллектуальной автоматизации, чтобы оставаться конкурентоспособными и работать эффективно [1].

Роботизированная автоматизация процессов (RPA) получила известность во многих отраслях. Поскольку бизнес сосредоточен на цифровых инновациях, автоматизация повторяющихся задач для повышения эффективности и сокращения человеческих ошибок является приоритетом [2].

Интеллектуальная автоматизация помимо RPA может быть достигнута за счет интеграции машинного обучения и искусственного интеллекта в роботизированную автоматизацию процессов для автоматизации повторяющихся задач с дополнительным уровнем человеческого восприятия и прогнозирования.

Несмотря на почти вековую историю российской нефтегазовой отрасли, ее нынешнее состояние сопровождается множеством проблем, которые могут решить новейшие информационные технологии [3]. В частности, задачи оценки надежности и прогнозирования проблем в работе оборудования с помощью машинного обучения, выбора методов повышения нефтеотдачи, оптимизации транспортных маршрутов и цепочек поставок оборудования и продукции, а также прогнозирования эффективности освоения новых месторождений.

Эти направления обусловлены специфическими особенностями нефтегазовой отрасли:

- Непрерывность сложной технологической цепочки – от геологоразведки до транспортировки нефтегазовых продуктов.
- Географическое распределение нефтепромысловых и перерабатывающих предприятий.
- Непрерывная работа дорогостоящего оборудования в агрессивных условиях.

Распределенная структура и высокая капиталоемкость отрасли приводят к тому, что ежедневно появляется множество разноплановых данных, которые необходимо собирать и анализировать, чтобы снизить текущие затраты и увеличить прибыль в будущем. Это стимул

**E-mail:* bashmur@bk.ru

http://dx.doi.org/10.5510/OGP2022SI100665

для внедрения технологий больших данных и машинного обучения в нефтегазовой отрасли в производственные процессы, в частности, системы предиктивной аналитики.

Технологии гидрокрекинга находят всё большее развитие. В частности, технологии гидрокрекинга для получения биотопливных композиций из растительных масел относятся к перспективным направлениям, находящихся в фокусе внимания ученых по всему миру. Гидропереработка растительных жиров/жирных кислот под высокими давлением и температурой водорода в присутствии катализатора является наиболее предпочтительной для производства биотоплива [4]. В этом процессе ненасыщенные жирные кислоты сначала превращаются в насыщенные жирные кислоты посредством гидрирования, а затем гидрокрекингом жирных кислот в углеводороды с более короткой цепью [5].

Известны различные способы применения машинного обучения в нефтегазовой отрасли:

• В различных системах сбора и анализа данных так называемые концепции «цифровое месторождение» и «цифровой завод» [6]. При этом большая часть данных, собираемых в нефтегазовом секторе, всё ещё отбрасывается или анализируется поверхностно [7].

• При планировании бурения скважин на объектах разработки нефтяного месторождения [8].

• Для прогнозирования пластового давления при разработке нефтяных месторождений [9].

• Для моделирования процессов нефтепереработки и поиска оптимальных значений управляющих параметров [10, 11].

Но всё ещё недостаточно изучен вопрос применения машинного обучения для прогнозирования отказов оборудования в процессе нефтепереработки, в частности, в широко распространенном процессе гидрокрекинга.

Целью работы является исследование эффективности методов машинного обучения для прогнозирования отказов технологического оборудования гидрокрекинга.

Что требует решить следующие задачи:

• Провести моделирование установки гидрокрекинга.

• Смоделировать различные сценарии работы установки.

• Получить данные об отказах оборудования в процессе работы установки гидрокрекинга.

• С помощью методов машинного обучения построить модели прогнозирования отказов.

• Сравнить эффективность различных моделей машинного обучения.

Методика проведения моделирования

Моделирование процессов газонефтепереработки можно проводить с помощью математических моделей, например, с помощью универсальных программных средств MATLAB Simulink, но для воспроизведения и проектирования сложных технологических систем используют продукции таких компаний, как Aspen Technology Inc. (Aspen Plus, Aspen HYSYS), Simulation Science Inc. (PRO/II) и Chemstations Inc. (ChemCad). В данной работе использовалась программа Aspen HYSYS.

Существует множество алгоритмов машинного обучения, которые используются для решения достаточно сложных задач. Методы машинного обучения лежат в основе популярной в настоящее время дисциплины – интеллектуального анализа данных. Эти методы используются для решения таких задач, как классификация, регрессия, прогнозирование и ранжирование.

Наиболее распространенные алгоритмы машинного обучения:

• Рекомендательная система – системы контентной или коллаборативной фильтрации [12, 13].

• Линейная регрессия – прогнозирование некоторого значения y на основе некоторого набора значений x [14, 15].

• Логистическая регрессия похожа на линейную регрессию, но вместо вычисления значения y оценивает, к какой категории принадлежит данная точка данных, используемая для решения задач классификации [16, 17].

• K-Nearest Neighbours – решение задачи классификации при наличии более 2-х категорий [18, 19].

• Машины опорных векторов – это алгоритм разбиения, который делит набор данных на категории по наибольшим «промежуткам» между категориями [20].

• Кластеризация K-средних – неконтролируемый алгоритм машинного обучения, который принимает неразмеченные данные и пытается сгруппировать кластеры похожих наблюдений ваших данных [21, 22].

• Метод главных компонент (PCE) – используется для преобразования набора данных с большим количеством параметров в новый набор данных с меньшим количеством параметров, и каждый новый параметр этого набора данных представляет собой линейную комбинацию ранее существовавших параметров. Эти преобразованные данные призваны с гораздо большей легкостью обосновать большую часть дисперсии исходного набора данных [23].

• Деревья решений и «случайный лес» (RF) – древовидные алгоритмы машинного обучения [24, 25].

В данной работе используем алгоритм «случайный лес». Он основан на построении так называемого ансамбля деревьев решений. Алгоритм был получен путем смешения двух основных идей: метода бэггинга и метода случайных подпространств. Этот метод используется для решения задач классификации, регрессии и кластеризации.

Деревом решений обычно называют представление правил в иерархической последовательной структуре, где каждый объект дома соответствует узлу, который дает решение. Правило представляет собой логическую структуру «if… else».

Особенностью применения данного алгоритма является универсальность, скорость и случайность построения деревьев решений.

У этого метода есть ряд преимуществ:

• Встроенный контроль качества.

• Повышенная точность.

• Простота использования, так как основными параметрами являются количество деревьев в ансамбле и количество признаков, по которым эти деревья будут разбиваться.

• Обработка данных с большим количеством признаков и классов.

• Высокий параллелизм.

Удобство построения дерева решений заключается в удобном представлении знаний в экспертных системах, а именно после написания набора правил их легко изобразить в виде графика, это удобнее и нагляднее, чем перечисление правил.

Пусть имеется некоторый массив данных размерно-

стью m на n элементов. Можно построить на их основе дерево решений, произвольно взяв некоторые данные из этого массива. Этот принцип показан на рисунке 1.

Чтобы повысить производительность модели построенной с использованием алгоритма RF, можно использовать настройку гиперпараметров.

Кроме того, распространен подход, согласно которому данные перед передачей в модель обрабатываются методом главных компонент (PCA). Применение алгоритма PCA может привести к небольшому усложнению интерпретации каждого признака при анализе важности признаков радиочастотной модели. Однако алгоритм PCA приводит к уменьшению размерности пространства признаков, что может привести к уменьшению количества признаков, которые должны быть обработаны радиочастотной моделью. Объем вычислений – один из основных недостатков алгоритма случайного леса (то есть выполнение модели может занять много времени). Использование алгоритма PCA может стать очень важной частью моделирования, особенно при работе с сотнями или тысячами признаков. В результате, если важнее всего создать максимально эффективную модель, то можно пожертвовать точностью определения важности признаков.

Также для повышения производительности необходимо использовать метод оптимизации гиперпараметров модели – задачи машинного обучения выбора набора оптимальных гиперпараметров для обучающего алгоритма. Одни и те же виды моделей машинного обучения могут требовать различные предположения, веса или скорости обучения для различных видов данных. Эти параметры называются гиперпараметрами и их следует настраивать так, чтобы модель могла оптимально решить задачу обучения. Для этого находится кортеж гиперпараметров, который даёт оптимальную модель, оптимизирующую заданную функцию потерь на заданных независимых данных. Целевая функция берёт кортеж гиперпараметров и возвращает связанные с ними потери. Часто используется перекрёстная проверка для оценки этой обобщающей способности.

Одной из основных областей прогнозного моделирования в науке о данных является классификация. Классификация состоит в попытке предсказать, из какого класса поступает конкретная выборка параметров из генеральной совокупности. При введении двухклассового разделения при решении задачи прогнозирования состояния установки гидрокрекинга существует два возможных класса: работоспособна (положительный) и неработоспособна (отрицательный). Затем модель классификации пытается предсказать будет ли каждый набор параметров соответствовать одному из двух классов.

Наиболее распространенным показателем эффективности решения задачи классификации является точность, то есть доля правильно спрогнозированных выборок, что характеризуется формулой:

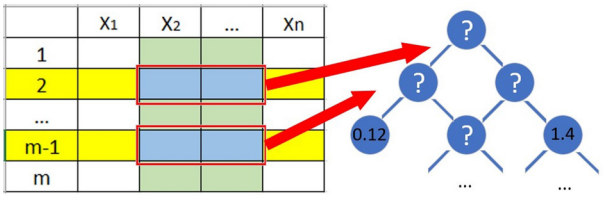

Рис. 1. Построение дерева решений на основе случайных данных

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

где TP – истинно положительный сегмент – фактически = 1, прогноз = 1; TN – истинно отрицательный сегмент – фактически = 0, прогноз = 0; FP – ложный положительный сегмент – фактически = 0, прогноз = 1; FN – ложно отрицательный сегмент – фактически = 1, прогноз = 0.

Для классификации установки гидрокрекинга обозначим выходной параметр работы установки как «failure», то есть отказ. Параметр может принимать 2 значения: 0 или 1. В соответствии со значением выходного параметра, каждый сценарий работы установки классифицируется: 0 – отказа нет, система работает; 1 – отказ системы, наличие неисправности.

С помощью программы Aspen HYSYS получена следующая модель установки гидрокрекинга (рис.2). Модель построена на основе принципиальной технологической схемы процесса гидрокрекинга. Процесс моделирования подробно описан в [26].

Описание производственного цикла работы установки:

Прямогонный вакуумный газойль (сырье) поступает на установку гидрокрекинга из резервуаров сырьевого парка или напрямую с вакуумной колонны.

После предварительного нагрева за счёт серии теплообменников в тройник смешения подаётся подпиточный

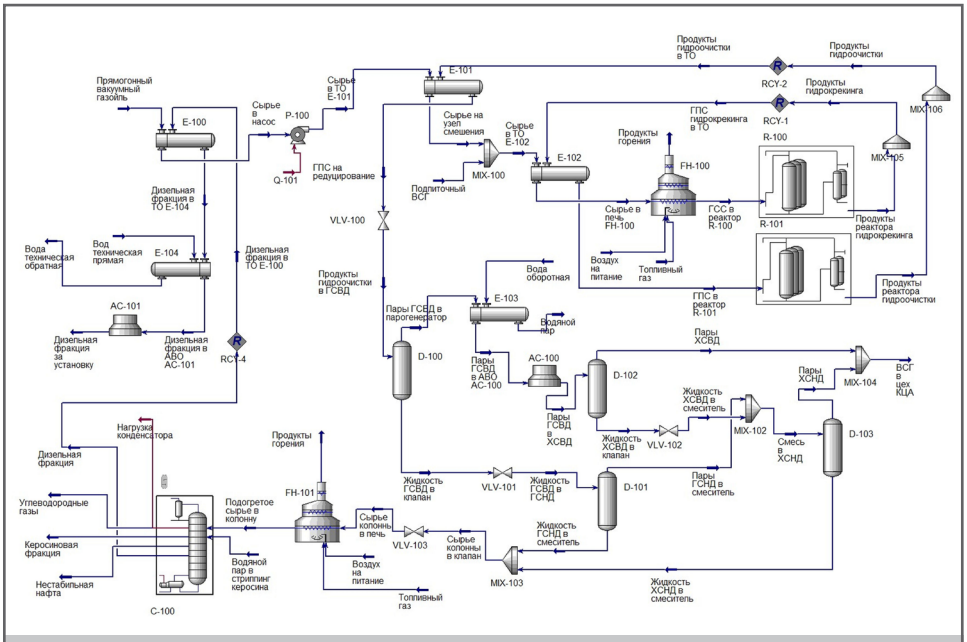

Рис.2. Технологическая схема установки гидрокрекинга

водородсодержащий газ.

Сырьё с температурой 333-391°C поступает в реактор гидрокрекинга поз. R-100. Реактор гидрокрекинга является одноступенчатым, с четырьмя неподвижными слоями катализатора гидрокрекинга, расположенными последовательно. По мере продвижения сырья через слои катализатора протекают реакции крекинга сырья. Экзотермические реакции крекинга высвобождают тепло и повышают температуры участвующих в реакции веществ. Для регулирования роста температуры и скорости реакции между слоями катализатора вводится циркулирующий водородсодержащий газ – квенч. Далее продукты реакции реактора гидрокрекинга R-100 поступают в реактор гидроочистки R-101.

В реакторе гидроочистки R-101 находится три слоя катализатора. В реакторе используются катализатор гидроочистки. Конструкция внутренних деталей реактора между слоями катализатора обеспечивает тщательное перемешивание реагентов и хорошее распределение паров и жидкости, стекающей на следующий слой. Пройдя реактор R-101, газопродуктовая смесь охлаждается рядом теплообменников.

Жидкие углеводороды смешиваются и направляются в колонну фракционирования C-100. Сырье колонны фракционирования поступает в печь нагрева, где нагревается до температуры 278 °C. После печи FH-101 сырье поступает в колонну. Колонна фракционирования C-100 содержит 30 тарелок: 22 тарелок над зоной испарения и 8 тарелок под ней. Пары выводится с верха колонны. Гидроочищенная керосиновая фракция отбирается с тарелки № 12. Очищенная дизельная фракция выводится с куба колонны фракционирования. Сырье подается в колонну в зону испарения между тарелками № 22 и № 23. Кубовый продукт колонны фракционирования – гидроочищенная дизельная фракция–охлаждается и затем уходит в парк готовой продукции.

Результаты

На основе полученной схемы технологического процесса гидрокрекинга с помощью встроенного менеджера Aspen Simulation Workbook смоделировано 1000 сценариев работы установки.

Данные импортированы, обработаны и получены в виде массива параметров с помощью программных библиотек Python – Pandas и NumPy.

Перед обучением моделей для имитации реальных условий проведено зашумление данных – 5% массива данных было обработано путем случайного умножения значения каждого параметра на коэффициент величиной от 0.9 до 1.1.

Все модели обучались на одном наборе данных: 50% набора данных определено как данных для обучения, остальные 50% – тестовые данные.

Также проведено масштабирование данных через нормализацию (изменение масштаба числовых атрибутов в диапазоне от 0 до 1) и стандартизацию (сдвиг распределения каждого атрибута, чтобы иметь среднее значение нуля и стандартное отклонение единицы).

Для обозначения технологических параметров использованы следующие обозначения:

- t1(Unit_1_temp) – температура потока входного сырья – вакуумного газойля (ВГ), °C.
- p1(Unit_1_press) – давление потока ВГ, МПа.
- f1(Unit_1_lvf) – объемный расход потока ВГ, м^3/ч.
- t2(Unit_2_temp) – температура потока второго компонента входного сырья – Водородсодержащего газа (ВСГ), °C.
- p2(Unit_2_press) – давление потока ВСГ, МПа.
- f2(Unit_2_lvf) – объемный расход ВСГ, м^3/ч.
- k1(Stable_prod_lvf) – объемный расход стабильного продукта в пламенный нагреватель FH-100, м^3/ч.
- k2(Gas_raw_mix_press) – давление газосырьевой смеси перед входом в реакторный блок R-100, МПа.

Failure – выходной параметр работы установки.

Общий вид полученного массива данных представлен на рисунке 3.

По результатам создания первой модели на основе RF-алгоритма получена оценка важности каждого параметра, представленная в таблице 1.

Точность первой модели составила 94.8%.

При создании второй модели на основе RF-алгоритма и метода PCA установлено, что использование 4-х компонент (комбинаций параметров) вместо 8 параметров объясняет 98% совокупной дисперсии данного распределения случайной величины (рис.4). Остальные 3 компоненты объясняют менее 2% дисперсии, поэтому ими можно пренебречь.

Точность второй модели составила 94.4%.

Третья модель построена с использованием оптимизации гиперпараметров (RF + PCA + HT).

На рисунке 5 показан широкий диапазон значений гиперпараметров, полученный с помощью функции RandomizedSearchCV.

Значение и анализ гиперпараметров модели:

- n_estimators – количество «деревьев» в случайном лесу. Значения 500, 700, 900, 1000 показывают лучшие средние результаты.
- min_samples_split – минимальное количество объектов, необходимое для разделения узла дерева. Значение 23 преимущественно выделяется, поэтому выбираем несколько значений в окружающем диапазоне: 22, 24, а также протестируем значения 2, 7 и 34.
- min_samples_leaf – минимальное количество объектов в листьях. Небольшие значения этого гиперпараметра дают лучшие результаты. Таким образом, используем значения от 2 до 7.
- max_features – число рассматриваемых возможностей при поиске наилучшего расщепления: если задано значение sqrt, используется max_features=sqrt(n_features). Опция sqrt дает самый высокий средний результат.

Таблица 1
Уровень влияния каждого параметра на общее распределение случайной величины

Параметр	Критерий Джини
Unit_2_lvf	0.468650
Stable_prod_lvf	0.272218
Unit_1_lvf	0.173638
Gas_raw_mix_press	0.070655
Unit_2_press	0.008469
Unit_2_temp	0.003121
Unit_1_temp	0.001675
Unit_1_temp	0.001675

	t1	p1	f1	t2	p2	f2	k1	k2	failure
0	38.0	0.3	35.0	95.0	17.798675	35.0	65.660215	16.255715	0
1	37.0	0.4	32.0	90.0	18.000000	32.0	62.660215	16.436470	0
2	37.1	1.0	40.5	90.9	17.468000	40.5	71.160215	15.586571	0
3	30.4	0.2	38.0	90.9	18.306000	38.0	68.660215	16.150799	0
4	31.1	0.5	25.5	93.3	19.424000	25.5	56.160215	16.651009	1
...	...	...	...	...	...	...	...	...	...
995	23.8	0.4	26.0	93.8	16.808000	26.0	56.660215	15.507289	1
996	22.0	0.2	30.0	101.5	19.521000	30.0	60.660215	16.502982	0
997	25.4	0.5	32.5	106.3	15.377000	32.5	63.160215	13.751663	0
998	31.2	0.2	29.5	88.2	18.105000	29.5	60.160215	16.523060	0
999	26.4	0.3	42.0	105.9	20.269000	42.0	72.660215	15.898547	0

Рис.3. Первые и последние 5 строк сгенерированного набора данных в выходной ячейке программы

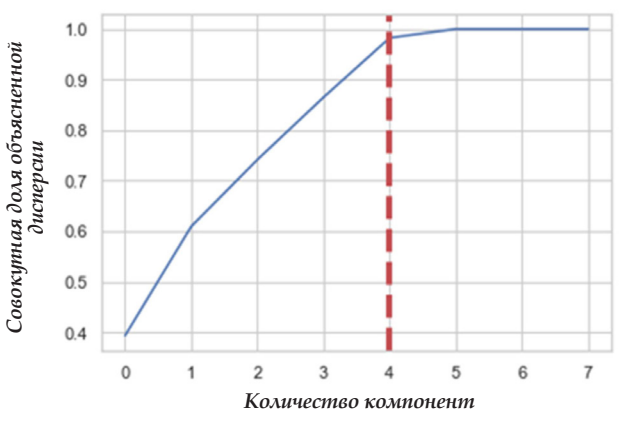

Рис.4. Влияние количества компонент на общую дисперсию

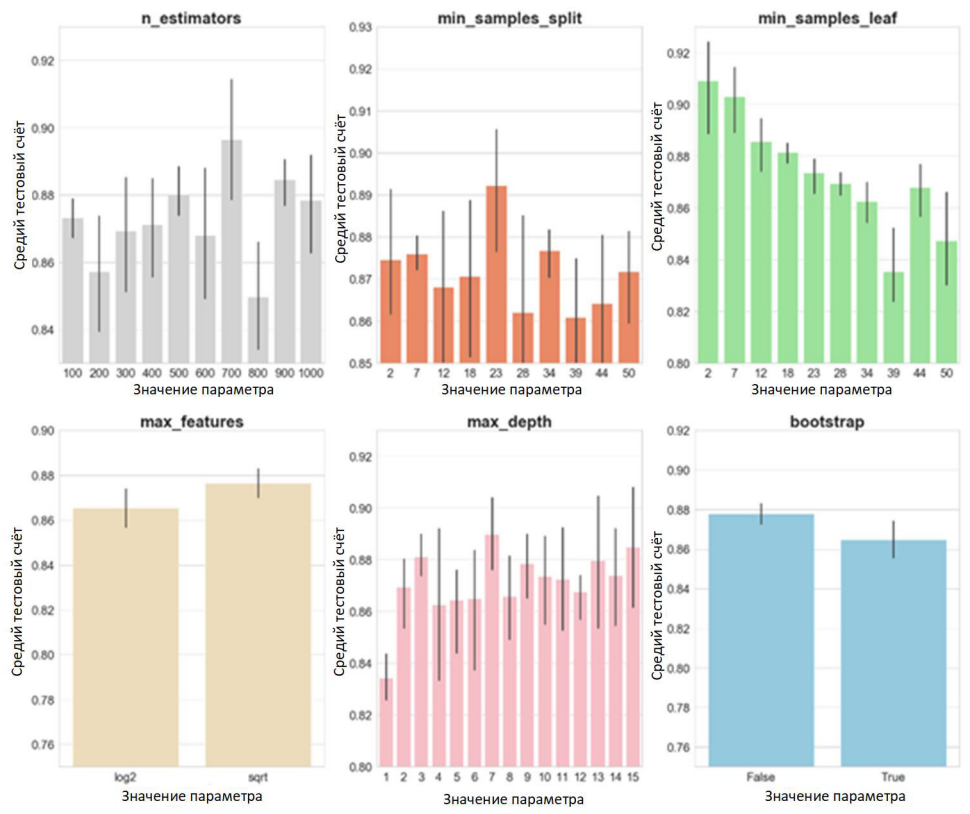

Рис.5. Визуализация широкого диапазона значений гиперпараметров

• max_depth – максимальная глубина деревьев. Следует использовать значения 3, 7, 9, 13, 15.

• bootstrap – начальная загрузка, определяет, нужна ли подвыборка с возвратом для построения деревьев. Если установлено значениие False, то для построения каждого дерева используется весь набор данных. Значение False показывает лучший средний результат.

Применена трёхблочная перекрестная проверка для 720 (4 × 6 × 6 × 1 × 5 × 1) сеансов обучения модели, в результате чего получили 2160 сеансов обучения модели.

Результат подбора оптимальных параметров, вызываемый с помощью параметра «best_params_», показан на рисунке 6.

Точность полученной модели составила 96.8%.

Подробные результаты обученных моделей показаны на общей матрице ошибок (табл.2).

В качестве показателя эффективности модели использована метрика точность («accuracy»), то есть отношение верных прогнозов к общему количеству прогнозов.

```
{'bootstrap': False,
 'max_depth': 7,
 'max_features': 'sqrt',
 'min_samples_leaf': 3,
 'min_samples_split': 7,
 'n_estimators': 700}
```

Рис.6. Результат подбора гиперпараметров

Расчёт точности каждой модели с использованием формулы (1) для удобства представим в процентном виде:

$$A_1 = \frac{(401+71)}{500} \cdot 100\% = 94.8\%$$

$$A_2 = \frac{(411+63)}{500} \cdot 100\% = 94.4\%$$

$$A_3 = \frac{(407+77)}{500} \cdot 100\% = 96.8\%$$

Точность базовой RF-модели составила 94.8%. Точность RF-модели после обработки PCA составила 94.4%.

Точность RF-модели с использованием PCA и оптимизации гиперпараметров составила 96.8%.

Таблица 2

Матрицы сопряженности для полученных прогнозных моделей

№ модели	Состояние системы	Прогноз 0	Прогноз 1
1	Фактически 0	401	10
	Фактически 1	18	71
2	Фактически 0	411	0
	Фактически 1	26	63
3	Фактически 0	407	4
	Фактически 1	12	77

Выводы

По результатам работы на основе принципиальной технологической схемы проведено моделирование установки гидрокрекинга с помощью программы Aspen HYSYS. Отработаны различные сценарии работы и отказов установки. Получены данные об отказах оборудования.

Созданы различные модели машинного обучения для прогнозирования отказов установки. Модель №1 обучена с использованием алгоритма «случайный лес». Модель №2 является улучшенной версией первой модели, доработанной с помощью метода главных компонент. Модель №3 является улучшенной версией второй модели с применением оптимизации гиперпараметров модели, то есть глубинных настроек.

Для оценки эффективности полученных моделей использован показатель Accuracy – точность модели. Точность базовой RF-модели (модель №1) составила 94.8%. Точность RF-модели после обработки PCA (модель №2) составила 94.4%. Точность RF-модели с использованием PCA и оптимизации гиперпараметров (модель №2) составила 96.8%.

На основе сравнения результатов точности моделей можно сделать важный вывод: улучшенная и оптимизированная модель не всегда может быть лучше базовой модели со стандартными настройками. Точность базовой модели (модель №1 – RF) была немного выше, чем у улучшенной модели (модель №1 – RF + PCA), но при этом процесс обучения и обработки данных занимает большее время и требует больше ресурсов в виде вычислительной мощности компьютера. Но в то же время модель №3 (модель №3 – RF + PCA + HT) все же эффективнее остальных. Поэтому для решения задач классификации и прогнозирования состояния технологического оборудования необходимо использовать и тестировать различные модели в зависимости от множества факторов. В случае с моделями, которые используются для прогнозирования отказов в нефтегазовой отрасли, можно сказать, что чем лучше модель, тем сильнее экономический эффект для предприятия.

В дальнейшем планируется исследовать вопрос прогнозирования отказов методами нейросетевого моделирования.

This study is carried out under the state assignment under the project "Development of a set of scientific and technical solutions in the field of creating biofuels and optimal biofuel compositions that provide the possibility of transforming consumed types of energy carriers in accordance with energy efficiency trends, reducing the carbon footprint of products and using alternative fuels to fossil fuels" (FSRZ Contract -2021-0012) in the scientific laboratory of biofuel compositions of the Siberian Federal University, created as part of the activities of the Research and Educational Center «Yenisei Siberia».

Литература

1. Zhang, X., Wen, Z. (2021). Thoughts on the development of artificial intelligence combined with RPA. *Journal of Physics: Conference Series*, 1883, 012151.

2. Plattfaut, R., Borghoff, V., Godefroid, M., et al. (2022). The critical success factors for robotic process automation. *Computers in Industry*, 138, 103646.

3. Pandey, Y. N., Rastogi, A., Kainkaryam, S., et al. (2020). Machine learning in the oil and gas industry. *New York, NY, USA: Apress*.

4. Jeong, H., Shin, M., Jeong, B., et al. (2020). Comparison of activity and stability of supported Ni_2P and Pt catalysts in the hydro-processing of palm oil into normal paraffins. *Journal of Industrial and Engineering Chemistry*, 83, 189-199.

5. Srihanun, N., Dujjanutat, P., Muanruksa, P., Kaewkannetra, P. (2020). Biofuels of green diesel–kerosene–gasoline production from palm oil: effect of palladium cooperated with second metal on hydrocracking reaction. *Catalysts*, 10, 241.

6. Еремин, Н. А., Столяров, В. Е. (2020). О цифровизации процессов газодобычи на поздних стадиях разработки месторождений. *SOCAR Proceedings*, 1, 59-69.

7. Perrons, R. K., Jensen, J. W. (2015). Data as an asset: What the oil and gas sector can learn from other industries about «Big Data». *Energy Policy*, 81, 117-121.

8. Ганиев, Б. Г., Насыбуллин, А. В., Саттаров, Р. З., и др. (2021). Применение методов машинного обучения при планировании бурения скважин на объектах разработки нефтяного месторождения. *Нефтяное хозяйство*, 7, 23-27.

9. Мартюшев, Д. А., Пономарева, И. Н., Захаров, Л. А., и др. (2021). Применение машинного обучения для прогнозирования пластового давления при разработке нефтяных месторождений. *Известия ТПУ. Инжиниринг георесурсов*, 10(332), 140-149.

10. Нужный, А. С., Однолько, И. С., Глухов, А. Ю., и др. (2021). Оптимизация содержания седиментов в процессе гидрокрекинга гудрона с использованием методов машинного обучения. *Прикладная математика и вопросы управления*, 1, 7-22.

11. Naveen, B., Ray, A. K., Rangaiah, G. P. (2006). Modeling, simulation, and multi-objective optimization of an industrial hydrocracking unit. *Industrial & Engineering Chemistry Research*, 45(4), 1354-1372.

12. Roy, G P. K., Chowdhary, S. S., Bhatia, R. (2020). A machine learning approach for automation of resume recommendation system. *Procedia Computer Science*, 167, 2318-2327.

13. Thangavel, S. K., Bkaratki P. D., Sankar, A. (2017, January). Student placement analyzer: A recommendation system using machine learning. In: Proceedings of the *4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India*.

14. Garcia-Ceja, E., Hugo, Å., Morin, B. et al. (2019, November). Towards the automation of a chemical sulphonation process with machine learning. In: *7th International Conference on Control, Mechatronics and Automation (ICCMA), Delft, The Netherlands*.

15. Lekbangpong, N., Muangprathub, J., Srisawat T., et al. (2019). Precise automation and analysis of environmental factor effecting on growth of St. John's Wort. *IEEE Access*, 7, 112848-112858.

16. Austin, P. C., Tu, J. V. (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of Clinical Epidemiology*, 57(11), 1138-1146.

17. Hoang, N.-D., Nguyen, Q.-L., Tran, X.-L. (2019). Automatic detection of concrete spalling using piecewise linear stochastic gradient descent logistic regression and image texture analysis. *Complexity*, 2019, 5910625.

18. Li, L., Zhang, Y., Zhao, Y. (2008). k-nearest neighbors for automated classification of celestial objects. *Science in China Series G: Physics, Mechanics and Astronomy*, 51(7), 916-922.

19. Cho, T.-H., Conners, R. W., Araman, P. A. (1991, October). A comparison of rule-based, K-nearest neighbor, and neural net classifiers for automated. In: *IEEE/ACM International Conference on Developing and Managing Expert System Programs, Washington, DC, USA*.

20. Begg, R. K., Palaniswami, M., Owen, B. (2005). Support vector machines for automated gait classification. *IEEE Transactions on Biomedical Engineering*, 52(5), 828-838.

21. Huang, J. Z., Ng, M. K., Rong, H., et al. (2005). Automated variable weighting in k-means type clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5), 657-668.

22. Chen, X., Xu, X., Huang, J. Z., et al. (2013). TW-k-means: Automated two-level variable weighting clustering algorithm for multiview data. *IEEE Transactions on Knowledge and Data Engineering*, 25(4), 932-944.

23. Liu, Y., Singleton, A., Arribas-Bel, D. (2019). A principal component analysis (PCA)-based framework for automated variable selection in geodemographic classification. *Geo-spatial Information Science*, 22(4), 251-264.

24. Tynchenko, V. S., Kurashkin, S. O., Tynchenko, V. V., et al. (2021). Software to Predict the Process Parameters of Electron Beam Welding. *IEEE Access*, 9, 92483-92499.

25. Ferreira, B., Silva, R. G., Pereira, V. (2017, September). Feature selection using non-binary decision trees applied to condition monitoring. In: *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus*.

26. Nekrasov, I., Tynchenko, V., Bukhtoyarov, V., et al. (2022, December). Simulation of the hydrocracking process to produce diesel fuel in the Aspen HYSYS system. In: *IV International Scientific Conference «Advanced Technologies in Aerospace, Mechanical and Automation Engineering» - «MIST: Aerospace-IV 2021», Krasnoyarsk, Russia*.

References

1. Zhang, X., Wen, Z. (2021). Thoughts on the development of artificial intelligence combined with RPA. *Journal of Physics: Conference Series*, 1883, 012151.

2. Plattfaut, R., Borghoff, V., Godefroid, M., et al. (2022). The critical success factors for robotic process automation. *Computers in Industry*, 138, 103646.

3. Pandey, Y. N., Rastogi, A., Kainkaryam, S., et al. (2020). Machine learning in the oil and gas industry. *New York, NY, USA: Apress*.

4. Jeong, H., Shin, M., Jeong, B., et al. (2020). Comparison of activity and stability of supported Ni_2P and Pt catalysts in the hydro-processing of palm oil into normal paraffins. *Journal of Industrial and Engineering Chemistry*, 83, 189-199.

5. Srihanun, N., Dujjanutat, P., Muanruksa, P., Kaewkannetra, P. (2020). Biofuels of green diesel–kerosene–gasoline production from palm oil: effect of palladium cooperated with second metal on hydrocracking reaction. Catalysts, 10, 241.

6. Eremin, N. A., Stolyarov, V. E. (2020). On the digitalization of gas production in the late stages of field development. *SOCAR Proceedings*, 1, 59-69.

7. Perrons, R. K., Jensen, J. W. (2015). Data as an asset: What the oil and gas sector can learn from other industries about «Big Data». *Energy Policy*, 81, 117-121.

8. Ganiev, B. G., Nasybullin, A. V., Sattarov, R. Z., et al. (2021). Application of machine learning methods in well drilling planning at oil field development objects. *Oil Industry*, 7, 23-27.

9. Martjushev, D. A., Ponovareva, I. N., Zaharov, L. A., et al. (2021). Application of machine learning to prediction of reservoir pressure in oil field development. *Bulletin of the Tomsk Polytechnic University, Geo Assets Engineering*, 10(332), 140-149.

10. Nuzhnyj, A. S., Odnol'ko, I. S., Glukhov, A. Yu., et al. (2021). Optimizing the content of the sediments in the process of hudron's hydrocracking with the use of machine learning methods. *Applied Mathematics and Control Sciences*, 1, 7-22.

11. Naveen, B., Ray, A. K., Rangaiah, G. P. (2006). Modeling, simulation, and multi-objective optimization of an industrial hydrocracking unit. *Industrial & Engineering Chemistry Research*, 45(4), 1354-1372.

12. Roy, G P. K., Chowdhary, S. S., Bhatia, R. (2020). A machine learning approach for automation of resume recommendation system. *Procedia Computer Science*, 167, 2318-2327.

13. Thangavel, S. K., Bkaratki P. D., Sankar, A. (2017, January). Student placement analyzer: A recommendation system using machine learning. In: Proceedings of the *4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India*.

14. Garcia-Ceja, E., Hugo, Å., Morin, B. et al. (2019, November). Towards the automation of a chemical sulphonation process with machine learning. In: *7th International Conference on Control, Mechatronics and Automation (ICCMA), Delft, The Netherlands*.

15. Lekbangpong, N., Muangprathub, J., Srisawat T., et al. (2019). Precise automation and analysis of environmental factor effecting on growth of St. John's Wort. *IEEE Access*, 7, 112848-112858.

16. Austin, P. C., Tu, J. V. (2004). Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *Journal of Clinical Epidemiology*, 57(11), 1138-1146.

17. Hoang, N.-D., Nguyen, Q.-L., Tran, X.-L. (2019). Automatic detection of concrete spalling using piecewise linear stochastic gradient descent logistic regression and image texture analysis. *Complexity*, 2019, 5910625.

18. Li, L., Zhang, Y., Zhao, Y. (2008). k-nearest neighbors for automated classification of celestial objects. *Science in China Series G: Physics, Mechanics and Astronomy*, 51(7), 916-922.

19. Cho, T.-H., Conners, R. W., Araman, P. A. (1991, October). A comparison of rule-based, K-nearest neighbor, and neural net classifiers for automated. In: *IEEE/ACM International Conference on Developing and Managing Expert System Programs, Washington, DC, USA*.

20. Begg, R. K., Palaniswami, M., Owen, B. (2005). Support vector machines for automated gait classification. *IEEE Transactions on Biomedical Engineering*, 52(5), 828-838.

21. Huang, J. Z., Ng, M. K., Rong, H., et al. (2005). Automated variable weighting in k-means type clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5), 657-668.

22. Chen, X., Xu, X., Huang, J. Z., et al. (2013). TW-k-means: Automated two-level variable weighting clustering algorithm for multiview data. *IEEE Transactions on Knowledge and Data Engineering*, 25(4), 932-944.

23. Liu, Y., Singleton, A., Arribas-Bel, D. (2019). A principal component analysis (PCA)-based framework for automated variable selection in geodemographic classification. *Geo-spatial Information Science*, 22(4), 251-264.

24. Tynchenko, V. S., Kurashkin, S. O., Tynchenko, V. V., et al. (2021). Software to Predict the Process Parameters of Electron Beam Welding. *IEEE Access*, 9, 92483-92499.

25. Ferreira, B., Silva, R. G., Pereira, V. (2017, September). Feature selection using non-binary decision trees applied to condition monitoring. In: *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus*.

26. Nekrasov, I., Tynchenko, V., Bukhtoyarov, V., et al. (2022, December). Simulation of the hydrocracking process to produce diesel fuel in the Aspen HYSYS system. In: *IV International Scientific Conference «Advanced Technologies in Aerospace, Mechanical and Automation Engineering» - «MIST: Aerospace-IV 2021», Krasnoyarsk, Russia*.

Применение прогнозирующих алгоритмов машинного обучения к процессам нефтепереработки в рамках интеллектуальной автоматизации

В. В. Бухтояров[1,2], И. С. Некрасов[1], В.С. Тынченко[1,2,3], К.А. Башмур[1], Р.Б. Сергиенко[4]
[1]Институт нефти и газа, Сибирский федеральный университет, Красноярск, Россия;
[2]Центр НТИ «Цифровое материаловедение: новые материалы и вещества», МГТУ им. Н.Э. Баумана, Москва, Россия; [3]Институт информатики и телекоммуникаций, СибГУ им. М.Ф. Решетнева, Красноярск, Россия; [4]Gini Gmbh, Мюнхен, Германия

Реферат

Нефтепереработка сталкивается с рядом вызовов и проблем при работе с данными. Большой объем данных генерируется множеством различных процессов и оборудованием. Данная статья посвящена методам эффективного анализа больших объемов данных на нефтеперерабатывающим предприятии. В частности, исследуется эффективность методов машинного обучения для прогнозирования отказов технологического оборудования процесса гидрокрекинга. Машинное обучение, как важный элемент цифровизации, позволяет успешно решать многие производственные задачи. В статье описано применение некоторых алгоритмов машинного обучения для решения задач классификации и прогнозирования отказов технологического оборудования гидрокрекинга, возникающих при переработке нефти и производстве дизельного топлива. Рассмотрено применение методов случайного леса, анализа главных компонент и настройки гиперпараметров. Сравнивается эффективность применения этих методов на основе параметра точности (Accuracy). Показано, что комбинация данных методов позволит повысить точность модели на 2%.

Ключевые слова: автоматизация; машинное обучение; гидрокрекинг; моделирование; нефтепереработка.

İntellektual avtomatlaşdırma çərçivəsində neft emalı proseslərinə maşın təliminin proqnozlaşdırıcı alqoritmlərinin tətbiqi

V. V. Buxtoyarov[1,2], İ. S. Nekrasov[1], V.S. Tınçenko[1,2,3], K.A. Başmur[1], R.B. Serqiyenko[4]
[1]Neft və Qaz İnstitutu, Sibir Federal Universiteti, Krasnoyarsk, Rusiya; [2]N.E.Bauman adına MDTU Milli texnoloji təşəbbüs Mərkəzi «Rəqəmsal materialşünaslıq: yeni materiallar və maddələr», Moskva, Rusiyas; [3]M.F.Reşetnev adına Sibir Dövlət Universiteti, İnformatika və Telekommunikasiya İnstitutu, Krasnoyarsk, Rusiya; [4]Gini Gmbh, Münhen, Almaniya

Xülasə

Neft emalında məlumatlarla işləyərkən bir sıra problemlərlə üzləşmək olar. Məlumatların böyük həcmi bir çox müxtəlif proseslər və avadanlıqlarla generasiya edir. Bu məqalə neft emalı müəssisəsində məlumatların böyük həcmlərinin səmərəli təhlili metodlarına həsr olunmuşdur. Xüsusilə, hidrokrekinq prosesinin texnoloji avadanlığının nasazlığının proqnozlaşdırılması üçün maşın təlimi metodlarının effektivliyi araşdırılır. Rəqəmsallaşmanın vacib elementi olan maşın təlimi, bir çox istehsal məsələlərini müvəffəqiyyətlə həll etməyə imkan verir. Məqalədə neftin emalı və dizel yanacağının istehsalı zamanı hidrokrekinq texnoloji avadanlığının yaranan nasazlığının təsnifatı və proqnozlaşdırılması məsələlərinin həlli üçün bəzi maşın təlimi alqoritmlərinin tətbiqi təsvir edilmişdir. Təsadüfi meşə metodlarının tətbiqi, əsas komponentlərin təhlili və hiperparametrlərin tənzimlənməsi nəzərdən keçirilir. Dəqiqlik parametri (Accuracy) əsasında bu metodların tətbiqinin effektivliyi müqayisə edilir. Göstərilmişdir ki, bu metodların kombinasiyası modelin dəqiqliyini 2% artırmağa imkan verəcək

Açar sözlər: avtomatlaşdırma; maşın təlimi; hidrokrekinq; modelləşdirmə; neft emalı.