

УДК 632.08+004.8

## ПРОГНОЗИРОВАНИЕ СОСТАВА ШИРОКОЙ ФРАКЦИИ ЛЁГКИХ УГЛЕВОДОРОДОВ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ ПРИ ТРУБОПРОВОДНОЙ ТРАНСПОРТИРОВКЕ

© С. Н. Терещенко, А. Л. Осипов, Е. Д. Моисеева

*Новосибирский государственный университет экономики и управления,*

*630099, г. Новосибирск, ул. Каменская, 56*

*E-mail: alosip@mail.ru*

Исследован подход применения методов машинного обучения для автоматического прогнозирования компонентного состава широкой фракции лёгких углеводородов при транспортировке по трубопроводу. На базе сети библиотек CatBoost разработана модель машинного обучения, позволяющая с процентом ошибки 2,263 по метрике MAPE определять компонентный состав смеси.

*Ключевые слова:* широкая фракция лёгких углеводородов, искусственный интеллект, машинное обучение, градиентный бустинг, CatBoost, линейная регрессия, трубопровод.

DOI: 10.15372/AUT20220111

**Введение.** Широкая фракция лёгких углеводородов (ШФЛУ) — продукт переработки попутного нефтяного газа и газового конденсата — представляет собой смесь сжиженных углеводородных газов (пропана и бутана) и более тяжёлых углеводородов и используется в качестве сырья нефтехимическими предприятиями [1].

Трубопровод является одним из основных способов транспортировки широкой фракции лёгких углеводородов и наполняется содержимым на нескольких станциях подкачки, которые расположены последовательно на пути к потребителю. Состав ШФЛУ меняется на каждой станции с течением времени и таким образом меняется итоговый состав ШФЛУ, которую получает потребитель [2].

Прогнозирование состава ШФЛУ на конечном пункте трубопровода имеет важное промышленное и экономическое значение. Прогноз позволяет оптимизировать технологические процессы по переработке ШФЛУ.

Технологии искусственного интеллекта, в том числе и машинного обучения, активно используются во многих отраслях: химии, биологии, экономики [3], в промышленности [4] и при анализе графических изображений [5]. В данной работе рассмотрим использование технологий машинного обучения для прогнозирования состава ШФЛУ на выходе трубопровода в процессе транспортировки.

За последнее десятилетие было проведено много научных исследований проблем трубопроводной транспортировки нефтегазовых продуктов.

В [6] рассматривается проблема трубопроводной транспортировки широкой фракции лёгких углеводородов в сжиженном состоянии. Показано, что под действием конвективной и турбулентной диффузии компонентный состав ШФЛУ в процессе транспортировки непрерывно изменяется, поэтому в конце трубопровода приходит ШФЛУ иного состава, чем закачивается в начале трубопровода. Положение с прогнозом состава ШФЛУ ещё больше усложняется, если в промежуточном пункте трубопровода к транспортируемой подкачивают ШФЛУ от другого производителя. Предлагается метод расчёта компонентного состава ШФЛУ в конце трубопровода по известному составу в начале трубопровода. Работа [7] посвящена расчёту математической модели компонентного состава ШФЛУ

с использованием интеграла Дюгамеля. В ней рассмотрена типичная широкая фракция лёгких углеводородов (метан-этановая, пропан-бутановая и пентан-гексановая), которую транспортируют по трубопроводу в жидком состоянии под высоким давлением. В процессе транспортировки состав ШФЛУ постоянно изменяется вследствие продольного перемешивания её компонентов.

В [8] рассматриваются достижения в исследованиях влияния температуры на последовательную транспортировку по газо- и нефтепродуктовому трубопроводу.

Последние 10 лет широкое распространение для различных отраслей получили технологии искусственного интеллекта, в том числе и методы машинного обучения. Проведён анализ работ в сфере применения методов машинного обучения для задач нефтегазовой и нефтехимической отраслей. В [9] рассмотрена модель, разработанная на основе исторических данных и алгоритмов машинного обучения, которая решает проблемы низкой точности и высокой вычислительной сложности при расчёте длины смешанной нефти при транспортировке по трубопроводу. В [10] прогнозируется расход масла через расходомеры с диафрагмой на основе нескольких входных переменных с помощью гибридных моделей машинного обучения и оптимизатора для прогнозирования широкого диапазона дебитов нефти через счётчики с диафрагмой. В работе [11] сделан прогноз доставки для подземных хранилищ природного газа в соляных кавернах. В данной работе предлагается набор новых моделей прогнозирования для подземного хранения природного газа (Ungs) в соляных кавернах на основе алгоритмов машинного обучения. В [12] рассмотрено использование двух моделей скорости эрозии для создания базы данных, а также разработаны три подхода к машинному обучению в помощь инженерам для оценки оптимальной скорости потока в трубопроводах. В [13] сравниваются различные модели машинного обучения, такие как случайный лес, машина опорных векторов и искусственная нейронная сеть для прогнозирования давления разрыва трубы и получения моделей высокой точности. В [14] исследуется водовоздушный поток в горизонтальной трубе с использованием поглощения гамма-излучения. Анализ сигналов сцинтилляционных детекторов позволяет определить количество параметров потока и идентифицировать структуру потока.

В [15] рассматриваются текущие и инновационные технологические решения, связанные со снижением вязкости и трения для перемещения сырой нефти с места добычи на перерабатывающие предприятия.

В [16] проводится мониторинг протяжённых трубопроводов, используемых при транспортировке нефти и газа. Используется технология применения мобильного устройства для сбора сигналов, обработки, идентификации и диагностики параметров химического состава и температуры на основе беспроводных сетей датчиков.

В [17] предлагается комплексный метод оценки рисков, основанный на модели нечёткой сети Петри для магистральных нефте- и газотранспортных трубопроводов. Подтверждено, что такой метод обеспечивает некоторую поддержку принятия решений для управления рисками нефте- и газопроводов.

Данная работа использует современные подходы к проблеме прогнозирования методами машинного обучения химического состава ШФЛУ во время транспортировки трубопроводным способом и опирается на вышеописанные методы и подходы.

Цель представленной работы состоит в создании новой модели машинного обучения для прогнозирования состава широкой фракции лёгких углеводородов при транспортировке трубопроводным способом.

**Материалы и методы исследования.** В работе исходным материалом для проведения научных исследований послужили размеченные данные с параметрами химического состава, расхода сырья для станции подкачки, опубликованные в открытом доступе на платформе СИБУР Диджитал [2]. Методы исследования: проектирование и разработка информационных систем, программирование, обработка и подготовка датасетов для

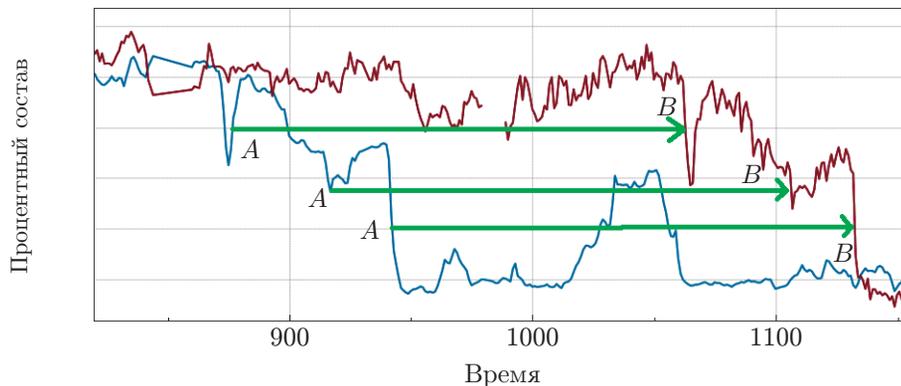


Рис. 1. Сопоставление временных графиков по компоненту C2H6 в точке A и в точке B

машинного обучения, использование алгоритмов настройки гиперпараметров в моделях нейронных сетей.

**Результаты исследования.** Открытый датасет представляет собой данные в CSV формате, в котором содержатся расходы сырья ШФЛУ в трубопроводе для исходной точки (станции подкачки) A и конечной точки (станции потребителя) B, а также массовые процентные доли различных химических компонентов в точке A (A\_C<sub>2</sub>H<sub>6</sub>, A\_C<sub>3</sub>H<sub>8</sub>, A\_iC<sub>4</sub>H<sub>10</sub>, A\_nC<sub>4</sub>H<sub>10</sub>, A\_iC<sub>5</sub>H<sub>12</sub>, A\_nC<sub>5</sub>H<sub>12</sub>, A\_C<sub>6</sub>H<sub>14</sub>) на время фиксации параметров (timestamp). Целевые переменные: массовые процентные доли нескольких компонентов в точке B (B\_C<sub>2</sub>H<sub>6</sub>, B\_C<sub>3</sub>H<sub>8</sub>, B\_iC<sub>4</sub>H<sub>10</sub>, B\_nC<sub>4</sub>H<sub>10</sub>) также на время фиксации параметров [2]. Объём составляет 5808 строк.

В датасете нет данных о протяжённости и размерах трубопровода. Учитывая, что скорость меняется в зависимости от химического состава и уровня расхода, одной из первых задач в исследовании было определение времени, которое необходимо для движения из точки в точку для соответствующего состава ШФЛУ с заданным уровнем расхода. На рис. 1 показано сопоставление временных графиков по компоненту C<sub>2</sub>H<sub>6</sub> в точках A и B. Из графика видно, что ключевые изменения по компоненту в исходной точке (станции подкачки) через определённый промежуток (около 92 ч) влияют на показания датчиков по этому же компоненту в точке B.

Горизонтальная ось показывает время замера компонента, а вертикальная ось характеризует его процентный состав.

При установке фиксированного промежутка времени на движение из точки A в точку B, например в значении 95 ч (подобрано опытным путём), были получены следующие результаты, представленные на рис. 2. Как видно на рис. 2, по некоторым ключевым изменениям компонента промежуток в 95 ч показывает почти полное совпадение точек A и B (выделено зелёными контурами). Но в то же время по другим колебаниям компонента на графике видно расхождение колебаний (жёлтые контуры), что указывает на ошибочное предположение определить время достижения конечной точки как константу.

Горизонтальная ось показывает время, а вертикальная ось характеризует давление ШФЛУ в трубопроводе.

Для определения времени прохождения между точками по заданным параметрам компонентов и общего расхода была обучена модель линейной регрессии. Обучающая выборка для модели формировалась по следующему алгоритму.

Каждому показателю давления компонента C<sub>2</sub>H<sub>6</sub> из точки A в заданный момент времени  $t$  сопоставлялись последовательно показатели в различные последующие моменты времени в точке B с  $t + 1$  до последней временной отметки в датасете.

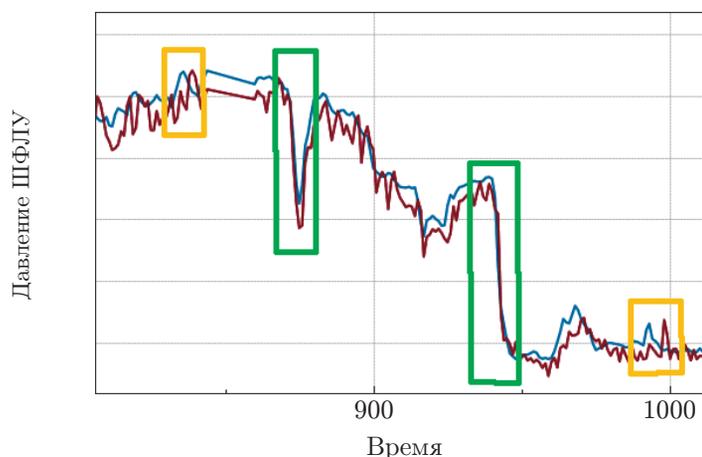


Рис. 2. Сопоставление временных графиков с заданным интервалом между точками в 95 ч по компоненту С2Н6

1. Для каждой сравниваемой пары показателей давления в трубопроводе для временных точек  $A$  рассчитывался коэффициент ранговой корреляции Спирмена.
2. Выбирается пара из  $k$  временных точек  $A$  и  $B$  с максимальным коэффициентом Спирмена.
3. Для пары рассчитывается расход компонента  $r$  как среднее арифметическое значение всех показателей расхода за интервал между показателями времени в точке  $A$  и точке  $B$ .
4. Для пары рассчитывается временной интервал  $d$  как разница между временными показателями точки  $B$  и точки  $A$ .
5. Показатели  $r$  (расход на точке  $B$ ) и  $d$  (временной интервал для прохождения ШФЛУ от точки  $A$  до точки  $B$ ) сохраняются как дополнительные параметры к точке  $A$  в обучающей выборке.

В модели линейной регрессии параметр  $r$  определяется как обучающая выборка, параметр  $d$  — как таргетируемый параметр для прогнозирования. Обученная модель линейной регрессии позволяет прогнозировать временной интервал для прохождения ШФЛУ от точки  $A$  до точки  $B$  на основе данных по точке  $A$ . В регрессионную модель для прогнозирования были поданы данные расхода в точке  $A$ .

Модель позволила прогнозировать соответствующие временные интервалы для прохождения ШФЛУ от точки  $A$  до точки  $B$  в зависимости от давления, по которым были определены таргетируемые параметры в точке  $B$ .

График прогнозируемых и фактических значений представлен на рис. 3. Из графика видно, что наблюдается корреляция между прогнозированием регрессионной модели и фактическими значениями обучающей выборки.

Таким образом, был обновлён датасет, где по каждой позиции расхода и компонентного состава в точке  $A$  был определён соответствующий расход и компонентный состав в точке  $B$ .

Для решения основной задачи по прогнозированию компонентного состава ШФЛУ в точке потребления по данным точки отправки была выбрана библиотека CatBoost — программная библиотека с открытым исходным кодом (от компании Яндекс) для построения моделей машинного обучения на основе градиентного бустинга.

Была выбрана модель CatBoostRegressor со следующими параметрами обучения модели: итераций обучения — 10000, шаг обучения — 0,05, функция потерь — MAPE (средняя абсолютная процентная ошибка).

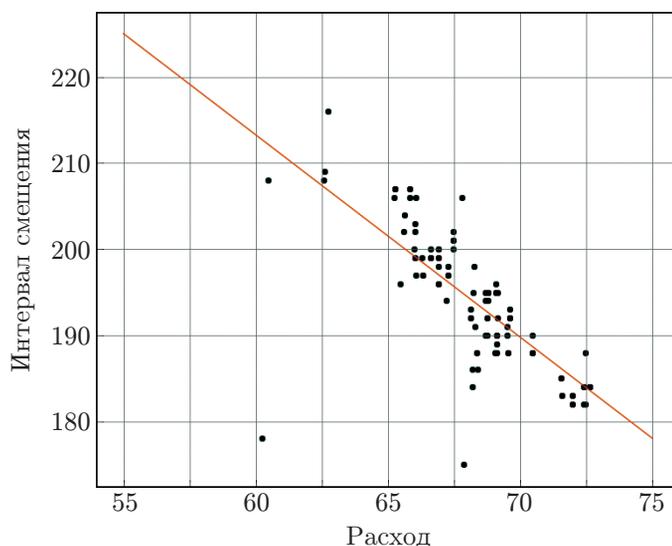


Рис. 3. Прогнозирование временного интервала для прохождения ШФЛУ от точки *A* к точке *B* (точки — фактические данные, сплошная линия — прогнозируемые данные)

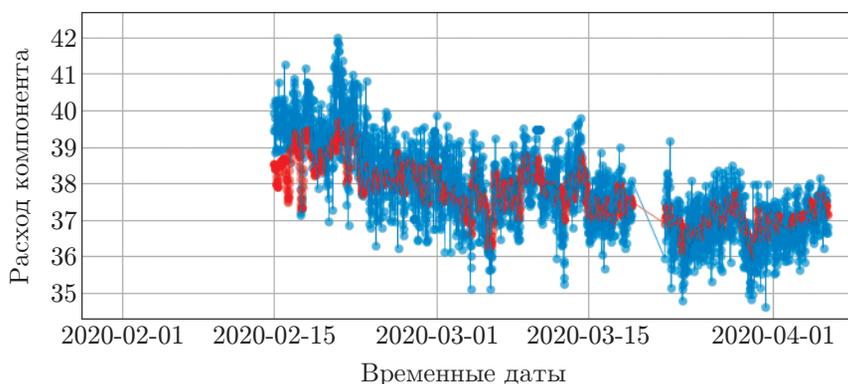


Рис. 4. Пример корреляции прогнозируемого (красный цвет) и фактических значений (голубой цвет) для компонента СЗН8 в различные моменты времени

Процесс обучения производился по каждому компоненту в отдельности. Пример результата в графическом виде по компоненту СЗН8 представлен на рис. 4.

Горизонтальная ось показывает временные даты, а вертикальная ось характеризует расход компонента.

Итоговая МАРЕ по всем компонентам по прогнозированию в точке *B* составила 2,263.

**Заключение.** Предсказание химического компонентного состава широкой фракции лёгких углеводородов при трубопроводной транспортировке является важной экономической и технологической составляющими в нефтехимической отрасли. В данной работе демонстрируется техническая осуществимость применения технологии машинного обучения для автоматического прогнозирования состава ШФЛУ после прохождения станций подкачки. Используется открытый набор данных в формате CSV из более 5800 позиций, содержащих значения компонентных параметров ШФЛУ при прохождении точек *A* и *B* на заданный момент времени. Была проведена предпроцессинговая подготовка обучающей выборки с помощью линейной регрессии к формату, позволяющему разметить параметры на обучающие и таргетируемые. Для обучения основной модели использовалась библио-

тека градиентного бустинга CatBoost. Точность средней абсолютной ошибки в процентах по всем компонентам ШФЛУ составила 2,263 %. В данной работе было исследовано применение методов машинного обучения для автоматического прогнозирования компонентного состава ШФЛУ при транспортировке по трубопроводу. Разработанная модель позволила прогнозировать состав ШФЛУ в точке *B* на основе данных в точке *A*. Была описана методика предпроцессинговой подготовки данных для обучения основной модели, процедура обучения модели библиотекой градиентного бустинга и настройкой гиперпараметров обучения. Разработанная модель может быть рекомендована для практического применения в нефтехимической отрасли при транспортировке трубопроводным способом ШФЛУ.

### СПИСОК ЛИТЕРАТУРЫ

1. **Широкая** фракция лёгких углеводородов. URL: <https://neftegaz.ru/tech-library/energoresursy-toplivo/141451-shirokaya-fraktsiya-legkikh-uglevodorodov/> (дата обращения: 12.05.2012).
2. **Сибур диджитал**. Sibur Challenge 2020. Прогноз состава сырья. URL: <https://sibur.ai-community.com/competitions/4/tasks/11> (дата обращения: 14.12.2020).
3. **Терещенко С. Н., Перов А. А., Осипов А. Л.** Определение заболеваний маниока методами компьютерного зрения // *Siberian Journ. Life Sci. Agriculture*. 2021. **13**, № 1. С. 144–155. DOI: 10.12731/2658-6649-2021-13-1-144-155.
4. **Ахвиев А. А., Шуршев В. Ф.** Применение машинного обучения в прогнозировании предаварийных ситуаций в системах теплоснабжения // *Вестн. Астраханского гос. техн. ун-та. Сер. Управление, вычислительная техника и информатика*. 2020. № 3. С. 74–81. DOI: 10.24143/2072-9502-2020-3-74-81.
5. **Терещенко С. Н., Перов А. А., Осипов А. Л.** Особенности применения предобученных свёрточных нейронных сетей к задачам стегаанализа графических изображений // *Автометрия*. 2021. **57**, № 4. С. 98–105. DOI: 10.15372/AUT20210412.
6. **Лаурье М. В., Диковская А. С.** Изменение состава ШФЛУ при транспортировании по магистральному трубопроводу // *Газовая промышленность*. 2012. **683**, № 12. С. 48–50.
7. **Лаурье М. В., Диковская А. С.** Перемешивание углеводородных составляющих широкой фракции лёгких углеводородов при транспортировании по трубопроводу // *Нефтяное хозяйство*. 2013. № 1. С. 100–103.
8. **Liang Y., He G., Fang L. et al.** Research advances in the influence of temperature on the sequential transportation in product pipeline // *Chinese Sci. Bull.* 2017. **62**, N 22. P. 2520–2533. DOI: 10.1360/N972016-00275.
9. **Chen L., Yuan Z., Xu J. et al.** A novel predictive model of mixed oil length of products pipeline driven by traditional model and data // *Journ. Petroleum Sci. Eng.* 2021. **205**. DOI: 10.1016/j.petrol.2021.108787.
10. **Farsi M., Barjouei H. S., Wood D. A. et al.** Prediction of oil flow rate through orifice flow meters: Optimized machine-learning techniques // *Measurement*. 2021. **174**. 108943.
11. **Ali A.** Data-driven based machine learning models for predicting the deliverability of underground natural gas storage in salt caverns // *Energy*. 2021. **229**. 120648.
12. **Liu G., Ayello F., Vera J. et al.** An exploration on the machine learning approaches to determine the erosion rates for liquid hydrocarbon transmission pipelines towards safer and cleaner transportations // *Journ. Cleaner Production*. 2021. **295**. 126478.
13. **Phan H. C., Dhar A. S.** Predicting pipeline burst pressures with machine learning models // *Intern. Journ. Pressure Vessels and Piping*. 2021. **191**. 104384. P. 1–32. DOI: 10.1016/j.ijpvp.2021.104384.

14. **Hanus R., Zych M., Kusy M. et al.** Identification of liquid-gas flow regime in a pipeline using gamma-ray absorption technique and computational intelligence methods // Flow Meas. and Instrum. 2018. **60**. P. 17–23.
15. **Martínez-Palou R., Mosqueira M. L., Zapata-Rendón B. et al.** Transportation of heavy and extra-heavy crude oil by pipeline: A review // Journ. Petroleum Sci. and Eng. 2011. **75**, N 3–4. P. 274–282.
16. **Junie P., Dinu O., Eremia C. et al.** A WSN based monitoring system for oil and gas transportation through pipelines // IFAC Proceedings. 2012. **45**, N 6. P. 1796–1801.
17. **Guo Y., Meng X., Wang D. et al.** Comprehensive risk evaluation of long-distance oil and gas transportation pipelines using a fuzzy Petri net model // Journ. Natural Gas Sci. Eng. 2016. **33**. P. 18–29.

*Поступила в редакцию 05.10.2021*

*После доработки 02.11.2021*

*Принята к публикации 07.12.2021*

---