

Прогнозирование процесса растворимости вещества в сверхкритических флюидах на основе машинного обучения

Т.Р. Билалов, Д.А. Лаврухина, А.Д. Павлов, М.П. Шлеймович

*Казанский национальный исследовательский технический университет
им. А.Н. Туполева-КАИ*

Аннотация: Применяемые на практике подходы к определению параметров процессов, протекающих в среде сверхкритических флюидов, базируются на экспериментальных данных, в первую очередь, растворимости ключевых компонентов системы в сверхкритических флюидных растворителях. Применяемые на практике математические модели не всегда обеспечивают точное описание растворимости, и не способны прогнозировать её за пределами исследованного диапазона параметров состояния. При этом проведение экспериментальных исследований может быть невозможно или затруднено. Однако в области сверхкритических флюидных технологий оптимизировать процессы и предсказывать их свойства можно на основе моделей и методов машинного обучения с использованием как накопленных экспериментальных, так и рассчитанных данных. Исследованию этого подхода посвящена данная статья. В работе рассмотрены входные параметры системы, свойства растворителя, свойства растворяемого и выходной параметр – растворимость. Проведенные исследования показали эффективность подхода к прогнозированию процесса растворимости на основе машинного обучения.

Ключевые слова: сверхкритические флюиды, растворимость веществ, факторы растворимости веществ, прогнозирование растворимости, машинное обучение, анализ остатков, анализ важности признаков.

Введение

В настоящее время исследования процессов растворимости веществ в среде сверхкритических флюидах осуществляются экспериментальными методами, которые дополняются различными математическими моделями, призванными обобщить полученные результаты и интерполировать их в рамках исследованного диапазона таких параметров состояния, как температура и давление [1, 2]. В качестве примеров можно указать методы Соаве-Редлиха-Квонга, Видаля и Микельсена, Вонга-Сэндлера [3], Кристила, дель Валле и Агилеры, Адачи и Лу, Мендеса-Сантьяго и Теджи, Бартла [4], Пенга-Робинсона [5].

В результате экспериментов, проведенных в течение длительного времени, а также их математического моделирования, накоплен значительный объем данных по растворимости различных веществ в

сверхкритических флюидных растворителях, в частности в сверхкритическом диоксиде углерода. На основе этих данных представляется целесообразным применить для анализа и прогнозирования параметров рассматриваемых процессов методы машинного обучения [6], которые активно используются в различных сферах человеческой деятельности [7-9].

Применение методов машинного обучения позволяет с относительно небольшими затратами определить с некоторым допуском значения параметров процесса, которые затем можно будет достаточно быстро уточнить экспериментальным путем [10].

Таким образом, актуальны и практически значимы исследования возможности прогнозирования растворимости на основе моделей машинного обучения с использованием как накопленных в результате физических экспериментов данных, так и рассчитанных с применением имеющихся математических моделей.

Постановка задачи

Как было сказано выше, современные информационные технологии, которые в настоящее время базируются на моделях и методах машинного обучения, позволяют обеспечить эффективный предварительный анализ процессов растворимости без проведения трудоемких и затратных физических экспериментов.

Для решения задачи прогнозирования параметров указанных процессов необходимо выполнить следующие этапы:

1. Сбор исходных данных, полученных экспериментально или посредством расчетов с применением имеющихся математических моделей;
 2. Предварительный анализ собранных данных;
 3. Определение количественных признаков, вычисленных по собранным данным;
-

4. Выбор моделей машинного обучения;
5. Обучение классификаторов, базирующихся на выбранных моделях машинного обучения;
6. Оценка классификаторов;
7. Выбор оптимального классификатора.

На первом этапе необходимо сформировать набор данных, содержащих информацию о растворимости различных веществ одного класса в сверхкритическом флюидном растворителе. Эти данные могут включать химический состав и структуры молекул веществ, физико-химические параметры, такие как температура и давление, и результаты экспериментов по измерению растворимости. Необходимость второго этапа связана с тем, что данные могут быть зашумлены или содержать отсутствующие значения. Поэтому необходимо провести их предварительную обработку, включая очистку, нормализацию и заполнение пропущенных значений. Важным этапом является создание хороших признаков для описания входных данных с точки зрения решаемой задачи. Поэтому на третьем этапе выполняются действия по формированию количественных характеристик, отражающих наиболее существенные физические и химические свойства процесс растворимости. На четвертом этапе определяются модели машинного обучения, на которых будут строиться классификаторы для прогнозирования параметров анализируемого процесса. Пятый этап связан с обучением исследуемых классификаторов на подготовленных данных. На шестом этапе осуществляется оценка качества обученных классификаторов, по результатам которой на последнем этапе выбирается наилучший из них.

Результаты исследования

При проведении исследований были рассмотрены параметры системы (температура, давление), свойства растворителя (критическая температура, критическое давление, приведенная плотности, химический потенциал и др.),

свойства растворяемого (критическая температура, критическое давление, температура кипения, молярная масса и др.). Прогнозируемым параметром является растворимость.

Собранный набор данных содержит 1107 объектов в виде 17 входных признаков и одного целевого значения. Предварительный анализ показал, что пропуски отсутствуют, но дублируется одна строка. Кроме того, 3 входных признака T_k , $R_{кр}$, w растворяемого были удалены, т.к. они являются константами.

Исследовательский анализ признаков показал, что все признаки растворяемого являются числовыми, но относятся скорее к дискретным переменным, т.к. имеют ограниченный диапазон конкретных значений. Можно попробовать модели с разным подходом – передавать такие признаки как числовые или категориальные (номинальные или порядковые), либо синтезировать из них новые признаки.

По графикам в количественных признаках (таких как приведенная плотность, химический потенциал и энтропия растворителя) заметно довольно большое число аномальных значений (рис. 1). Однако их удалять нецелесообразно, поскольку эти признаки получены из натуральных экспериментов, отражают характеристики реальных объектов и поэтому удаление таких аномальных значений может привести к переобучению модели на наиболее усредненных значениях.

У целевого признака также довольно большой разброс значений (рис. 2) и множество аномалий, которые можно удалить для повышения точности модели. Кроме того, возможным решением является создание отдельных моделей для аномальных значений.

Распределение признаков не является нормальным. Все признаки растворяемого являются мультиколлинеарными (рис. 3). Это может быть существенной проблемой для линейных моделей. Целевой признак также

больше коррелирует с признаками растворяемого, чем с признаками системы или растворителя.

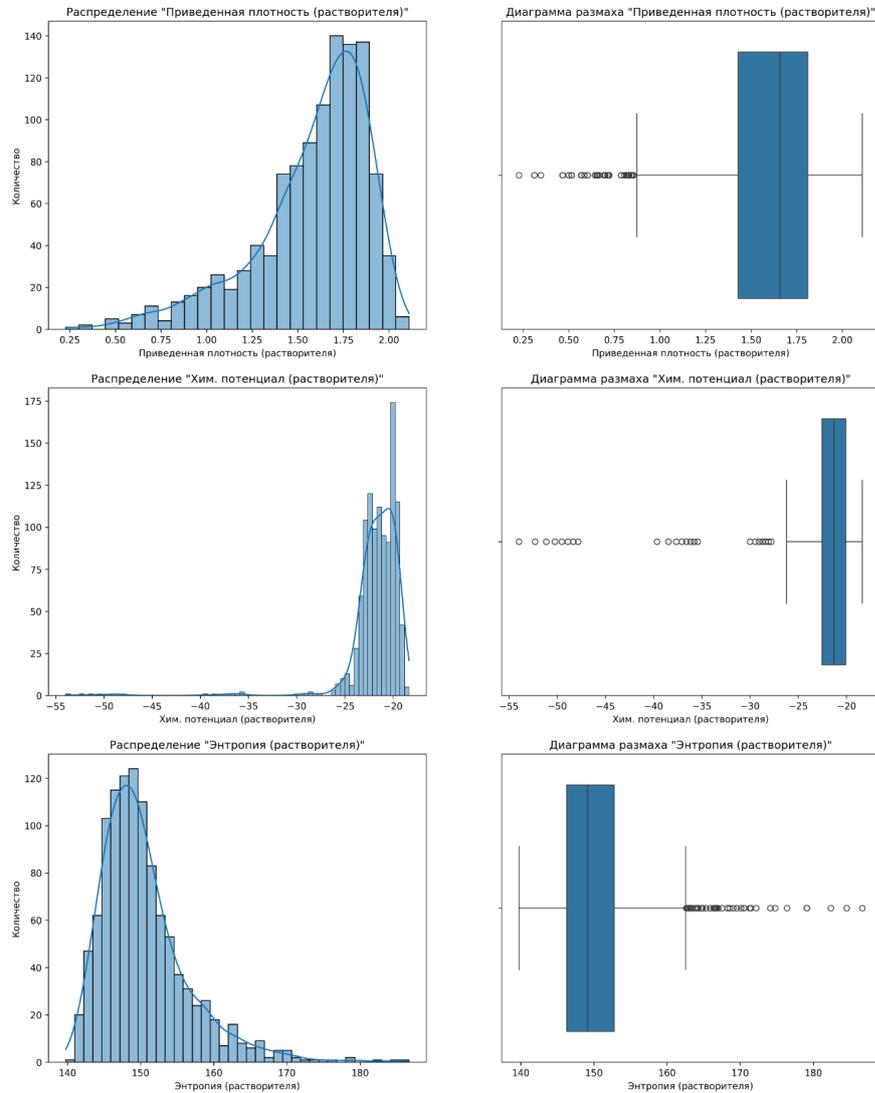


Рис. 1. – Распределение и диаграмма размаха входных признаков

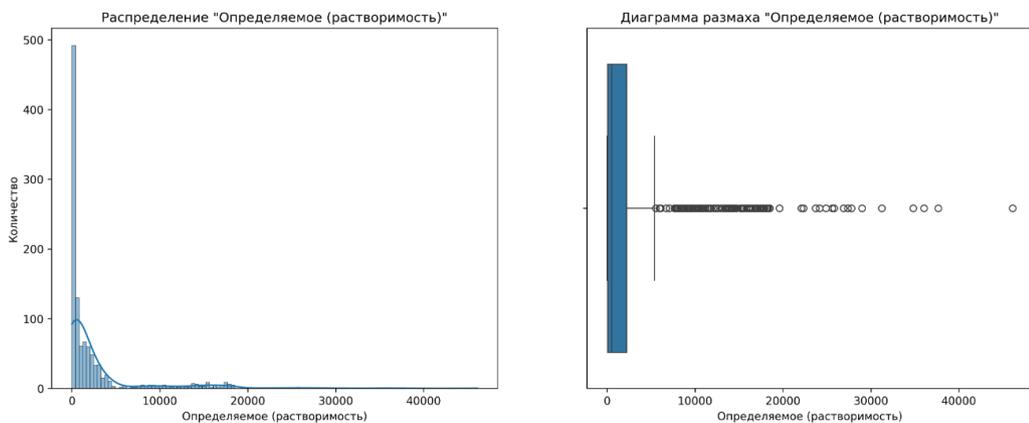


Рис. 2. – Распределение и диаграмма размаха целевого признака

При выполнении исследований были рассмотрены следующие базовые модели машинного обучения: LinearRegression, Lasso, Ridge, ElasticNet, SVR, KNeighborsRegressor, DecisionTreeRegressor, RandomForestRegressor и GradientBoostingRegressor. По итогам их обучения и оценки наилучшие результаты показала модель RandomForestRegressor, для которой получены следующие значения метрик:

$$RMSE = 236,56308;$$

$$MAE = 106,83315;$$

$$R_2 = 0,95496.$$

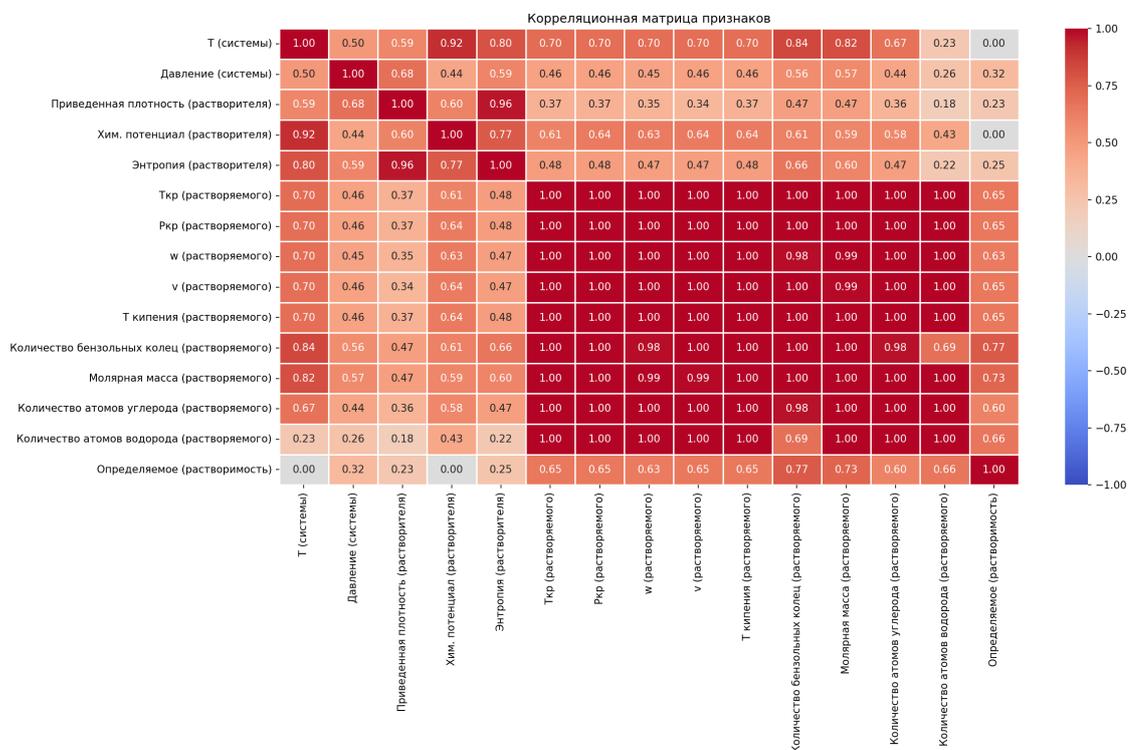


Рис. 3. – Корреляционная матрица признаков

Анализ остатков модели на рис. 4 показывает зависимость, похожую на квадратическую – значения ошибки увеличиваются со значением целевого признака. Это говорит о том, что модель машинного обучения, скорее всего, уловила не все зависимости в исходных данных. Это связано с использованием ограниченного набора гиперпараметров и ограниченных диапазонов из значений.

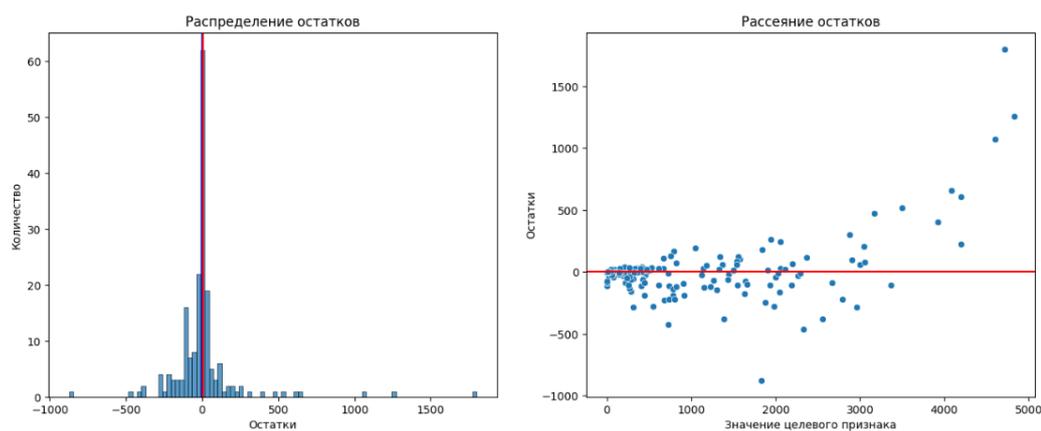


Рис. 4. – Анализ остатков

Отметим, что полученная точность является достаточно высокой, что говорит о перспективности рассматриваемого подхода. Для ее увеличения предполагается в дальнейшем использовать расширенный диапазон гиперпараметров и вероятностный подход к их подбору.

После построения модели машинного обучения важно провести также анализ важности входных признаков – т.е. проанализировать, какие входные признаки оказывают наибольшее влияние на целевой признак (рис. 5). Для наилучшей модели исследования наиболее значимыми признаками оказались $T_{кр}$ и T кипения растворяемого и давление системы.

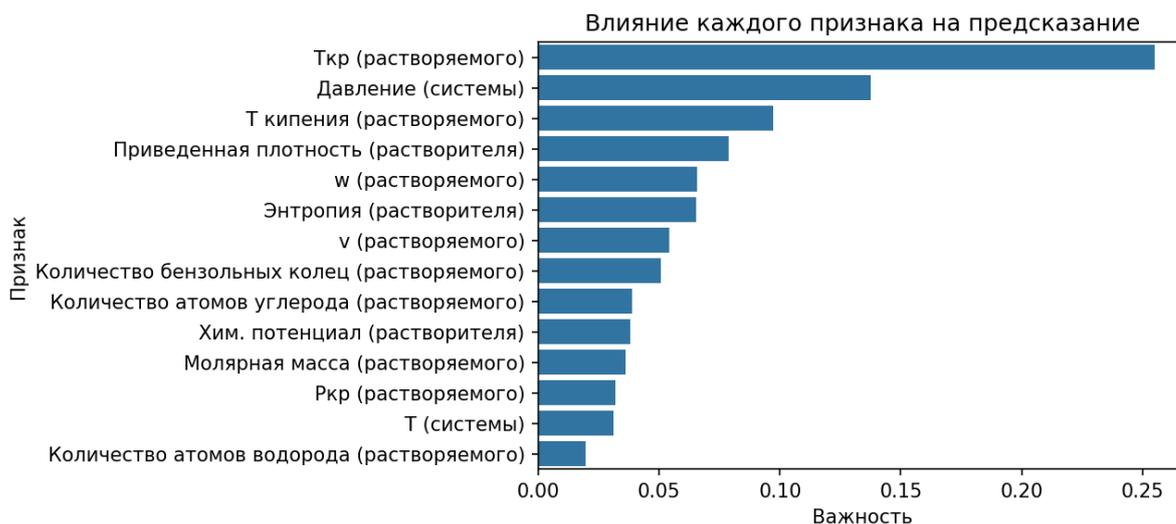


Рис. 5. – Анализ важности признаков

Заключение

Проведенные исследования показали эффективность подхода к прогнозированию процесса растворимости на основе машинного обучения. Данный подход не отменяет необходимости в физических экспериментах. Он позволяет быстро, относительно точно и с малыми затратами получить ориентировочные значения растворимости при заданных параметрах, чтобы затем уточнить их или получить представления о скрытых закономерностях.

Литература

1. Nasri L., Bensaad S., Bensefati Z. Correlation and Prediction of the Solubility of Solid Solutes in Chemically Diverse Supercritical Fluids Based on the Expanded Liquid Theory // *Advances in Chemical Engineering and Science*. 2013. Vol. 3. No 4. P. 255.
2. Rathnam V.M., Lamba N., Madras G. Evaluation of new density based model to correlate the solubilities of ricinoleic acid, methyl ricinoleate and methyl 10-undecenoate in supercritical carbon dioxide // *J. Supercrit. Fluids*. 2017. Vol. 130. P. 357.
3. Vazquez da Silva M., Barbosa D. Prediction of the Solubility of Aromatic Components of Wine in Carbon Dioxide // *J. Supercrit. Fluids*. 2004. Vol. 31. No 1. P. 9.
4. Sparks D.L., Hernandez R., Estevez L.A. Evaluation of Density-Based Models for the Solubility of Solids in Supercritical Carbon Dioxide and Formulation of a New Model // *Chem. Engineering Science*. 2008. Vol. 63. No 17. P. 4292.
5. Manohar B., Sankar K.U. Prediction of solubility of Psoralea corylifolia L. Seed extract in supercritical carbon dioxide by equation of state models // *Theor Foun Chem Eng*. 2011. Vol. 45. P. 409.
6. Kubat M. *An Introduction to Machine Learning*. Springer, 2015. 291 p.



7. Стебаков И.Н. Шутин Д.В. Марахин Н.А. Машинное обучение в реабилитационной медицине и пример классификатора движений пальцев для кистевого тренажера // Инженерный вестник Дона, 2020, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2020/6514

8. Федутинов К.А. Машинное обучение в задачах поддержки принятия решений при управлении охраной природы // Инженерный вестник Дона, 2021, № 9. URL: ivdon.ru/ru/magazine/archive/n9y2021/7186.

9. Горлатов Д.В. Машинное обучение прогнозных моделей на несбалансированных данных по опасным астероидам // Инженерный вестник Дона, 2023, № 5. URL: ivdon.ru/ru/magazine/archive/n5y2023/8394.

10. Лаврухина Д.А., Павлов А.Д., Шлеймович М.П., Билалов Т.Р. Возможности методов машинного обучения в задачах прогнозирования растворимости веществ в сверхкритическом диоксиде углерода // Сверхкритические флюиды: теория и практика. 2024. Т. 19. № 3. С. 4-12.

References

1. Nasri L., Bensaad S., Bensetiti Z. Advances in Chemical Engineering and Science. 2013. Vol. 3. No 4. P. 255.
 2. Rathnam V.M., Lamba N., Madras G. j. Supercrit. Fluids. 2017. Vol. 130. P. 357.
 3. Vazquez da Silva M., Barbosa D. j. Supercrit. Fluids. 2004. Vol. 31. No 1. P. 9.
 4. Sparks D.L., Hernandez R., Estevez L.A. Chem. Engineering Science. 2008. Vol. 63. No 17. P. 4292.
 5. Manohar B., Sankar K.U. Theor Foun Chem Eng. 2011. Vol. 45. P. 409.
 6. Kubat M. An Introduction to Machine Learning. Springer, 2015. 291 p.
 7. Stebakov I.N., Shutin D.V., Marakhin N.A. Inzhenernyj vestnik Dona, 2020, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2020/6514
-



8. Fedutinov K.A. Inzhenernyj vestnik Dona, 2021, № 9. URL: ivdon.ru/ru/magazine/archive/n9y2021/7186.

9. Gorlatov D.V. Inzhenernyj vestnik Dona, 2023, № 5. URL: ivdon.ru/ru/magazine/archive/n5y2023/8394.

10. Lavrukhina D.A., Pavlov A.D., Shleimovich M.P., Bilalov T.R. Sverhkriticheskie Flyuidy: Teoriya i Praktika. 2024. Vol. 19. № 3. pp. 4-12.

Дата поступления: 13.02.2025

Дата публикации: 27.03.2025