

ПРИМЕНЕНИЕ ДЕРЕВЬЕВ РЕШЕНИЙ ДЛЯ КЛАССИФИКАЦИИ ПОЖАРОВЗРЫВООПАСНЫХ МАТЕРИАЛОВ, ИСПОЛЗУЕМЫХ В ХИМИЧЕСКОЙ ПРОМЫШЛЕННОСТИ

Попок В. Н. ORCID ID 0000-0001-7800-7032

*Федеральное государственное бюджетное образовательное учреждение высшего образования
«МИРЭА – Российский технологический университет», Москва, Российская Федерация,
e-mail: vnpopok@mail.ru*

В статье представлены результаты классификации группы органических соединений, в том числе легковоспламеняющихся жидкостей, как пожаровзрывоопасных материалов с использованием разных методов. Решается задача классификации объектов по набору признаков – свойств в сравнении с искусственной классификацией по одному признаку. Выбраны статистические методы группировки и классификации растворителей по набору неэмпирических и эмпирических признаков и задаваемым классам опасности – деревья решений и линейная регрессия. В качестве базовой искусственной классификации используется классификация по температуре вспышки широкой группы органических соединений – близкая к стандартной. Показана эффективность и приемлемая погрешность метода деревьев решений при многомерной – с обучением на искусственной группировке, классификации соединений различных классов по группам опасности. Установлено, что применение линейной регрессии к рассматриваемому набору данных дает неудовлетворительные результаты классификации – большая доля ошибочных прогнозов (до 25 %). Показано влияние выбора групп признаков на результаты классификации и приемлемая точность прогноза для базовой группы неэмпирических признаков. Отражены возможности расширения номенклатуры признаков-свойств и наблюдаемых корреляций между ними для повышения качества и информативности классификации с использованием деревьев решений.

Ключевые слова: деревья решений, объекты, признаки, группы опасности, регрессия

APPLICATION OF TREES OF SOLUTIONS FOR CLASSIFICATION OF FIRE- AND EXPLOSION-HAZARDOUS MATERIALS USED IN THE CHEMICAL INDUSTRY

Popok V. N. ORCID ID 0000-0001-7800-7032

*Federal State Budgetary Educational Institution of Higher Education
“MIREA – Russian Technological University”, Moscow, Russian Federation,
e-mail: vnpopok@mail.ru*

The article presents the results of the classification of a group organic compound, including flammable liquids, as fire- and explosion-hazardous materials using various methods. The article analyzes the important task of classifying objects based on a set of features and properties, in comparison with artificial classification based on a single feature. The article selects statistical methods for grouping and classifying solvents based on a set of non-empirical and empirical features and predefined hazard classes, such as decision trees and linear regression. The flash point classification of a wide group of organic compounds, which is close to the standard classification, is used as the basic artificial classification. The paper demonstrates the effectiveness and acceptable error of the decision tree method in multidimensional classification of compounds of various classes (including flammable liquids) into hazard groups, with training on artificial grouping. Applying linear regression to the data set under consideration results in unsatisfactory classification outcomes, with a high percentage of incorrect predictions (up to 25 %). The paper also highlights the impact of feature group selection on classification results and the acceptable accuracy of predictions for the base group of non-empirical features. The article reflects the possibilities of expanding the nomenclature of feature-properties and the observed correlations between them to improve the quality and informativeness of classification using decision trees.

Keywords: decision trees, objects, features, danger groups, regression

Введение

Классификация и оценка пожаровзрывоопасных материалов (ПВОМ) по их чувствительности или опасности базируется на использовании сравнительного анализа с практическими (опорными) рядами чувствительности или с классификационными рядами, построенными по какому-то па-

раметру чувствительности или опасности [1; 2; 3, с. 159–187]. Для легковоспламеняющихся жидкостей (ЛВЖ) или пылевоздушных смесей (ПВС) классификационные ряды построены на основе выделения интервалов изменения одного, стандартизованного параметра чувствительности [3, с. 180–212; 4; 5]. Классификации ЛВЖ

и ПВС являются внешними (искусственными) и требуют установления взаимосвязи искусственных классификационных рядов с комплексом свойств используемых веществ и материалов, что приводит к необходимости использовать методы классификации с обучением алгоритма распознавания по искусственному признаку [6–8]. К настоящему времени разработаны математические методы, алгоритмы и комплексы программ анализа, классификации, группировки, ранжирования многомерных данных разной природы [9; 10, с. 280–310; 11]. Среди этих методов, учитывая природу физико-химических свойств ЛВЖ, ПВС и задач по обеспечению безопасности работ с этими классами материалов, следует акцентировать внимание на методах построения деревьев решений [11–13]. А также на методах, использующих нейронные сети, логико-структурные и другие методы с алгоритмами обучения [12–14].

Метод деревьев решений, алгоритмы обработки и визуализации данных детально рассмотрены в многочисленных публикациях, реализованы в большом количестве пакетов программ, например Deductor, Minitab и др., включающих обширный арсенал статистических методов анализа данных [6; 8; 13].

В статье представлены результаты классификации объектов с использованием деревьев решений из выборки данных для ряда широко используемых органических соединений (условно – растворителей) [3, с. 112–134; 14], преимущественно ЛВЖ.

Цель исследования – классификация пожаровзрывоопасных материалов, применяемых в химической промышленности, с использованием деревьев решений

Материалы и методы исследования

В качестве искусственного (внешнего) признака использовалась классификация растворителей по температуре вспышки (Твсп) [3, с. 96–112; 14] на три класса опасности: первый класс – $T_{всп} < 0$ °С; второй класс – $0 \leq T_{всп} \leq 50$ °С; третий класс – $T_{всп} > 50$ °С. Эти классы опасности частично соответствуют (перекрываются) стандартизованным классам опасности (по Твсп) ЛВЖ [3, с. 134–157].

Статистическая обработка данных проводилась с использованием доступного программного обеспечения пакета Deductor [13]. Значения молекулярных дескрипторов [14] для растворителей определялись в ряде случаев с использованием программного комплекса Dragon. Коэффициенты корреляции пар признаков и кластерный анализ

использовались при выборе групп свойств-признаков для построения деревьев решений. Основной целью анализа блока данных для растворителей было построение деревьев решений, обеспечивающих с приемлемой точностью классификацию объектов анализа по степени опасности без привлечения эмпирически определяемых параметров.

Группа растворителей включает 36 объектов, характеризующихся 9 признаками [14], определяемыми структурой, эмпирическими свойствами соединений и одним искусственным признаком деления объектов на три класса опасности (Кл. оп.) по значениям температуры вспышки. Классификация растворителей по классам опасности по значениям Твсп проведена выше. Используется алгоритм выбора C4.5 разделяющего признака на каждом шаге ветвления (модификация алгоритма ID3 [9; 11]), реализованный в пакете программ Deductor. При обучении алгоритма [9; 11; 12] правило ветвления определяется по частотному признаку – область определения какого-то признака разбивается на подобласти с определением количества объектов, попадающих в каждый интервал и фиксацией класса опасности. В качестве признаков (физико-химических свойств) используются: индекс Винера (W), индекс электронной плотности Бончева (J), значение полной электронной энергии молекулы (E, эВ), энергия высшей заселенной молекулярной орбитали (ВЗМО, эВ), энергия нижней свободной молекулярной орбитали (НСМО, эВ), молекулярная масса (М.М.) [14]. В качестве эмпирически определяемых признаков рассматриваются: температура кипения (Ткип), температура вспышки (Твсп), температура самовоспламенения (Тс.в.) растворителей [3, с. 234–250; 14]. Значения температуры для всех признаков указаны в °С. Численные значения признаков были заимствованы из справочной литературы, публикаций [3, с. 278–300; 14] или определялись с использованием пакета Dragon.

Результаты исследования и их обсуждение

Номера объектов анализа – растворителей и значения признаков приведены в табл. 1. Выборка результатов классификации растворителей по деревьям решений с указанием ошибок прогноза на разном наборе признаков приведена в табл. 2 – ошибки решений выделены знаком «*». Представлены два набора неэмпирических признаков, приводящих к разному количеству ошибок прогноза класса опасности.

Таблица 1

Сводные данные по физико-химическим свойствам объектов анализа
и их искусственной классификации

№	W	J	E	ВЗМО	НСМО	Ткип	Твсп	Тс.в.	М.М	Кл. оп.
1	18	57,5	3415	-11,49	3,50	27,9	-52	427	72,2	1
2	27	69	4275	-11,29	3,47	80,8	-18	260	84,2	1
3	35	69	4299	-11,27	3,36	68,7	-20	234	86,2	1
4	56	80,5	5337	-11,27	3,01	98,4	-4	223	100,2	1
5	39	69	3171	-9,75	0,40	80,1	-11	562	78,1	1
6	58	80,5	4179	-9,44	0,37	110,6	4	536	92,1	2
7	81	92	5335	-9,30	0,39	144,4	32	464	106,2	2
8	83	92	5267	-9,18	0,36	138,4	26	595	106,2	2
9	84	92	5271	-9,40	0,39	136	20	420	108,2	2
10	113	103,5	6535	-9,53	0,38	152,4	34	424	120,2	2
11	201	138	8693	-9,14	0,14	255	113	566	154,2	3
12	431	161	12891	-9,48	0,37	272	129		182,3	3
13	91	92	4882	-9,13	-0,12	146	30	530	104,2	2
14	161	115	7619	-9,50	0,37	182	71	412	134	3
15	153	115	6639	-8,84	-0,41	218	80	530	128,2	3
16	383	161	11317	-8,69	-0,39	345	121	472	178,4	3
17	84	98,5	5481	-9,11	0,35	156	41	485	108,1	2
18	83	176	5341	-9,23	0,24	173,4	66	648	147	3
19	58	154	4213	-9,81	-0,05	156	30	545	157	2
20	18	61,5	3456	-9,45	3,08	63	-9		73,1	1
21	16	61,5	3514	-9,47	3,01	43,8	-12	380	73,1	1
22	32	73	4585	-10,78	3,15	114	39	347	88	2
23	16	131	3520	-10,79	-0,15	51	-25	602	92,6	1
24	21	99	3347	-10,41	1,23	78	-6,6	460	92,6	1
25	20	64	3522	-10,41	2,86	35,6	-41	164	74,1	1
26	52	75,5	5301	-10,50	0,86	136	41		102,2	2
27	12	52,5	2293	-10,77	0,79	56,2	-18	465	58,1	1
28	22	70,5	3414	-11,26	1,01	57,3	-15	470	74,1	1
29	37	82	4410	-11,25	1,03	77,1	2	400	88,1	2
30	22	68	3276	-9,55	1,04	153	59	420	73,1	3
31	8	111,5	992	-9,83	-1,61	46,3	-43	90	76,1	1
32	10	129,8	2469	-10,67	0,53	83,5	9	413	99	2
33	27	82	4666	-10,45	2,84	105,4	11	340	88	2
34	58	122,5	4215	-9,39	0,06	132	29	593	112,6	2
35	118	120,5	6513	-10,60	-1,13	211	87	445	123,1	3
36	91	98,6	5022	-10,05	-0,48	179,2	64	205	106,1	3

Примечание: составлена автором на основе полученных данных в ходе исследования и источников [3; 14].

Таблица 2

Результаты классификации растворителей по степени опасности

№	Растворитель	Класс опасности по Твсп – внешний	Признаки: W, J, ВЗМО, НСМО, E, М.М., прогноз класса опасности		Признаки: W, E, М.М.; прогноз класса опасности (Кл. оп.), № правила – табл. 3.	
			Класс опасности		Класс опасности	№ правила
			Дерево решений	Линейная регрессия		
1	2-Метилбутан	1	1	1	1	1
2	Циклогексан	1	1	1	1	1
3	<i>n</i> -Гексан	1	1	1	2*	3
4	<i>n</i> -Гептан	1	1	2*	1	4
5	Бензол	1	1	2*	1	1
6	Толуол	2	2	2	2	3
7	<i>o</i> -Ксилол	2	2	2	1*	4
8	<i>n</i> -Ксилол	2	2	2	2	3
9	Этилбензол	2	2	2	2	3
10	<i>изо</i> -Пропилбензол	2	2	3*	2	5
11	Дифенил	3	3	3	3	6
12	1,2-Дифенилэтан	3	3	3	3	6
13	Стирол	2	2	2	2	3
14	Бутилбензол	3	3	3	3	6
15	Нафталин	3	3	3	3	6
16	Антрацен	3	3	3	3	6
17	Анизол	2	2	2	2	5
18	1,4-Дихлорбензол	3	2*	2*	1*	4
19	Бромбензол	2	2	2	2	3
20	<i>втор</i> -Бутиламин	1	1	1	1	1
21	<i>трет</i> -Бутиламин	1	1	1	1	1
22	<i>изо</i> -Амиловый спирт	2	2	1*	2	3
23	<i>трет</i> -Бутилхлорид	1	1	2*	1	2
24	1-Хлорбутан	1	1	1	1	2
25	Диэтиловый эфир	1	1	1	1	1
26	3,3-Диметил-2-бутанон	2	2	2	2	3
27	Ацетон	1	1	1	1	1
28	Метилацетат	1	1	1	1	1
29	Этилацетат	2	2	2	2	3
30	Диметилформамид	3	1	2*	1*	1
31	Сероуглерод	1	1	1	1	1
32	1,2-Дихлорэтан	2	2	1*	1*	2
33	1,4-Диоксан	2	2	2	2	3
34	Хлорбензол	2	2	2	2	3
35	Нитробензол	3	3	3	3	6
36	Бензальдегид	3	1*	2*	2*	3
Ошибки классификации			2 ошибки	9 ошибок	6 ошибок	

Примечание: составлена автором на основе полученных данных в ходе исследования.

Таблица 3

Правила решений (6 правил) для дерева решений и результаты прогноза

№	Правила решений	Класс опасности (прогноз)	Достоверность	
			%	Количество
1	$W < 115,5; M.M < 86,1$	1	90,00	9
2	$W < 115,5; M.M \geq 86,1; W < 24$	1	66,67	2
3	$W < 115,5; M.M \geq 86,1; W \geq 24; E < 5319$	2	90,91	10
4	$W < 115,5; M.M \geq 86,1; W \geq 24; E \geq 5319; W < 83,5$	1	50,00	1
5	$W < 115,5; M.M \geq 86,1; W \geq 24; E \geq 5319; W \geq 83,5$	2	100,00	2
6	$W \geq 115,5$	3	100,00	6

Примечание: составлена автором на основе полученных данных в ходе исследования.

Значения коэффициентов парной корреляции ($abs(R) = 0,2-0,9$) признаков изменяются в широких пределах, однозначного влияния корреляции неэмпирических признаков на результаты классификации не выявлено, как и однозначного влияния предварительной группировки признаков с использованием кластерного анализа [5–7]. При выборе признаков для построения деревьев решений также не учитывалась наблюдаемая корреляция между группами признаков [14]. Среди эмпирических признаков температура вспышки (Твсп) хорошо коррелирует (нелинейная зависимость) с температурой кипения (Ткип): $R^2 = 0,96$.

Минимальная погрешность прогноза классификации – два случая из 36 ($< 6\%$), получена для всей совокупности неэмпирических признаков. Учет только трех признаков (W, E, M.M.) приводит к увеличению количества ошибок прогноза до 6. При этом реализуется разное количество правил решения, то есть набора условий «если – то».

Для случая трех признаков (W, E, M.M.) набор правил, решений и их достоверность приведены (для иллюстрации) в табл. 3. При изменении набора признаков получена максимальная ошибка классификации – 7 ошибок. Для сравнения с результатами классификации с использованием деревьев решений в табл. 2 представлены результаты классификации на основе линейной регрессии для всех признаков. Количество ошибок классификации возросло до 9 из 36 случаев (до 25%). Снижение ошибок прогноза можно получить при учете новых признаков, например, некоторых новых молекулярных дескрипторов [13; 14].

Точность прогноза – 2 ошибки из 36 объектов, полученная с использованием деревьев решений, является типичной

и, по-видимому, корректируемой в сторону уменьшения погрешности путем расширения количества признаков и их сортировки. Предварительный анализ показывает, что в качестве основного источника погрешности классификации по некоторым группам признаков является значимый вклад частичного наложения областей определения некоторых признаков в разных классах искусственной классификации [6; 10, с. 23–74; 14]. Этот вывод подтверждают результаты кластерного анализа – объекты анализа группируются в три кластера, включающих объекты из разных классов опасности.

Таким образом, представленные результаты показывают в целом высокую эффективность деревьев решений и правил для классификации объектов, предварительно классифицированных по внешнему признаку, по набору неэмпирических свойств-признаков. Следует отметить, что пакеты Deductor, Minitab содержат опцию тестирования нового объекта на обученных деревьях решений и определенных кластерах.

Использование других подходов и методов, включая логико-структурный, корреляционный, факторный и дискриминантный анализы, нейросетевые алгоритмы, расширит, как показывают результаты ряда работ, информативность проведенной группировки и классификации, а также позволит получать корреляционные и другие соотношения, в том числе инвариантные, с низким значением дисперсии результатов [5; 15].

Выводы

Проведенные исследования позволяют сделать следующие выводы:

1. Выбраны статистические методы группировки и классификации растворителей (в том числе ЛВЖ) по набору неэмпирических признаков и задаваемым классам опас-

ности: деревья решений – базовый метод, линейная регрессия, кластерный анализ.

2. Показана эффективность и приемлемая погрешность (6%) в сравнении с погрешностью линейной регрессии (25%) метода деревьев решений при многомерной – с обучением на искусственной группировке, классификации соединений различных классов (в том числе ЛВЖ) по группам опасности.

3. Установлено влияние выбора групп признаков на результаты классификации методом деревьев решений – с увеличением в ряде случаев погрешности классификации до 20 %, и получена приемлемая точность прогноза (6 %) для базовой группы неэмпирических признаков.

Список литературы

1. Назин Г. М., Корсунский Б. Л. Определение термической стабильности взрывчатых веществ методом опорного ряда // *Химическая физика*. 2021. Т. 40. № 3. С. 53–59. URL: <https://www.sciencejournals.ru/cgi/getPDF.pl?jid=khimfiz&year=2021&vol=40&iss=3&file=KhimFiz2103009Nazin.pdf> (дата обращения: 19.01.2026). DOI: 10.31857/S0207401X21030092.
2. Назин Г. М., Корсунский Б. Л., Казаков А. И. Чувствительность взрывчатых веществ к удару и скорость реакции термического разложения // *Химическая физика*. 2023. Т. 42. № 3. С. 49–57. URL: <https://www.sciencejournals.ru/cgi/getPDF.pl?jid=khimfiz&year=2023&vol=42&iss=3&file=KhimFiz2303012Nazin.pdf> (дата обращения: 19.01.2026). DOI: 10.31857/S0207401X23030123.
3. Пожарная опасность веществ и материалов, применяемых в химической промышленности // *Справочник под ред. Н. В. Рябова*. М: Химия, 1970. 336 с. URL: <https://djvu.online/file/HuFWpGWayqqih> (дата обращения: 19.01.2026).
4. Попок В. Н. Применение кластерного анализа для классификации, сравнительного анализа, построения опорных рядов чувствительности и опасности групп пожаровзрывоопасных материалов // *Современные наукоемкие технологии*. 2025. № 12. С. 145–150. URL: <https://top-technologies.ru/ru/article/view?id=40616> (дата обращения: 20.01.2026). DOI: <https://doi.org/10.17513/snt.40616>.
5. Попок В. Н. Некоторые результаты статистической группировки и классификации энергетических материалов по характеристикам термического разложения и теплового взрыва // *Бутлеровские сообщения*. 2019. Т. 57. № 2. С. 41–49. URL: <https://butlerov.com/files/reports/2019/vol57/2/41/19-57-2-41.pdf> (дата обращения: 19.01.2026). EDN: YZJGUX.
6. Possolo A., Koepe A., Newton D., Winchester M. R. Decision Tree for Key Comparisons // *Journal of Research of National Institute of Standards and Technology (USA)*. 2021. Vol. 126. № 126007. P. 1–36. URL: <https://nvlpubs.nist.gov/nist-pubs/jres/126/jres.126.007.pdf> (дата обращения: 19.01.2026). DOI: 10.6028/jres.126.007.
7. Niazi S. K., Mariam Z. Recent Advances in Machine-Learning-Based Chemoinformatics: A Comprehensive Review // *Int. J. Mol. Sci.* 2023. Vol. 24. 11488. URL: <https://www.mdpi.com/1422-0067/24/14/11488> (дата обращения: 19.01.2026). DOI: 10.3390/ijms241411488.
8. Sarker I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions // *SN Computer Science*. 2021. Vol. 2. 160. <https://link.springer.com/article/10.1007/s42979-021-00592-x#citeas> (дата обращения: 19.01.2026). DOI: 10.1007/s42979-021-00592-x.
9. Mowbray M., Vallerio M., Perez-Galvan C., Zhang D. Industrial data science – a review of machine learning applications for chemical and process industries // *Reaction Chemistry&Engineering*. 2022. Vol. 7. P. 1471–1509. URL: https://www.researchgate.net/publication/360109573_Industrial_data_science_-_a_review_of_machine_learning_applications_for_chemical_and_process_industries (дата обращения: 19.01.2026). DOI: 10.1039/D1RE00541C.
10. Попок В. Н. Сокристаллизация компонентов смесевых энергетических материалов. Казань: Издательство ООО «Инновационно-издательский дом «Бутлеровское наследие», 2023. 324 с. [Электронный ресурс]. URL: <https://koha.benran.ru/cgi-bin/koha/opac-detail.pl?biblionumber=2250073> (дата обращения: 19.01.2026). ISBN 978-5-6047662-3-1.
11. Bahzad T. J., Adnan M. A. Classification Based on Decision Tree Algorithm for Machine Learning // *Journal of Applied Science and Technology Trends*. 2021. Vol. 2. Is. 1. P. 20–28. URL: https://www.researchgate.net/publication/350386944_Classification_Based_on_Decision_Tree_Algorithm_for_Machine_Learning (дата обращения: 19.01.2026). DOI: 10.38094/jast20165.
12. Цветков А. М. Разработка алгоритмов индуктивного вывода с использованием деревьев решений // *Кибернетика и системный анализ*. 1993. № 1. С. 174–178. URL: <https://lib.iis.nsk.su/node/226616> (дата обращения: 19.01.2026).
13. Дмитриев А. А., Михеев Г. М., Каландаров Х. У. Применение алгоритма дерева решений для оценки результатов хроматографического анализа трансформаторного масла // *Вестник Чувашского университета*. 2023. № 4. С. 74–84. URL: <https://cyberleninka.ru/article/n/primenenie-algoritma-dereva-resheniy-dlya-otsenki-rezultatov-hromatograficheskogo-analiza-transformatornogo-masla/viewer> (дата обращения: 19.01.2026). DOI: 10.47026/1810-1909-2023-4-74-84.
14. Урядов В. Г., Курдюков А. И., Аристова Н. В., Официров Е. Н. Топологический подход к описанию термодинамических свойств органических соединений, содержащих гетероатомы // *Химия и компьютерное моделирование. Бутлеровские сообщения*. 2000. Т. 1. № 3. С. 67–74. URL: <https://butlerov.com/files/reports/2000/vol1/3/67/67-74.pdf> (дата обращения: 19.01.2026). EDN: MPONHU.
15. Попок В. Н. Корреляция характеристик теплового взрыва и термического разложения компонентов и композиций смесевых энергетических материалов // *Бутлеровские сообщения*. 2018. Т. 56. № 12. С. 71–78. URL: <https://butlerov.com/files/reports/2018/vol56/12/71/18-56-12-71.pdf> (дата обращения: 19.01.2026). EDN: YTAETZ.

Конфликт интересов: Авторы заявляют об отсутствии конфликта интересов.

Conflict of interest: The authors declare that there is no conflict of interest.

Финансирование: Авторы заявляют об отсутствии внешнего финансирования.

Financing: The research was performed without external funding.