

---

---

**СИСТЕМНЫЙ АНАЛИЗ ПРИБОРОВ  
И ИЗМЕРИТЕЛЬНЫХ МЕТОДИК**

---

---

УДК 681.586.67

© Т. В. Осипова, А. М. Баранов, И. И. Иванов, 2023

**ОЦЕНКА ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ КЛАССИФИКАТОРОВ  
МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧЕ РАСПОЗНАВАНИЯ  
ТИПА ГАЗА И ЕГО КОНЦЕНТРАЦИИ**

В данной работе были рассмотрены методы классификации в машинном обучении, проведена оценка применимости классификаторов в задаче распознавания типа газа и его концентрации в смеси. В ходе исследования были сформированы теоретические значения водорода, углеводородов и их смесей в соотношении 20, 50 и 80%, и оценка классификаторов проводилась на экспериментальных данных, полученных с термokatалитического сенсора. Представленные классификаторы позволили определить тип газа с точностью до 87.5%.

*Кл. сл.:* термokatалитический сенсор, метод главных компонент, классификация, определение концентрации, водород, обработка данных

**ВВЕДЕНИЕ**

На сегодняшний день проведение измерений и анализа состава газовой смеси играет ключевую роль в различных областях, включая мониторинг окружающей среды и промышленное производство. Способность точно идентифицировать, количественно определять и разделять по типу газообразные вещества играет важную роль в обеспечении безопасности жизнедеятельности, оптимизации технологических процессов и диагностике окружающей среды. Традиционные методы газового анализа в значительной степени опираются на стандартные методы измерений и ручную интерпретацию, часто сталкиваясь с присущими им ограничениями в чувствительности обнаружения, трудоемкостью анализа и подверженностью человеческим ошибкам.

Среди огромного количества разнообразных углеводородов особое место занимает водород. Имея самый большой диапазон взрывоопасных концентраций (от 4 до 75% об.), а также являясь крайне летучим, водород считается одним из самых взрывоопасных газов. Также водород рассматривается в качестве перспективного вида топлива (или компонента топлива, в частности, в смесях с углеводородами) [1, 2], разработка эффективных сенсоров водорода и методик его мониторинга в воздухе является актуальной задачей.

В последние годы с появлением машинного обучения произошло ключевое обновление в применении искусственного интеллекта, развитии экспертных систем и появлении новых подходов

к обработке больших массивов данных. Эта область искусственного интеллекта позволяет разрабатывать инновационные подходы для автоматизации сложных задач, повышения эффективности и получения ценной информации из обширных наборов данных. Используя возможности методов машинного обучения, исследователи модернизируют методы газового анализа, способствуя развитию этой важной научной области.

Используя алгоритмы машинного обучения, газовый анализ может выйти за рамки стандартных практик и преобразоваться в более точный, эффективный и комплексный подход. Способность быстро обрабатывать огромные массивы данных и извлекать закономерности в сочетании со скоростью принятия решений алгоритмами приведет к модернизации или к разработке новых способов в методологиях газового анализа. Более того, модели машинного обучения могут значительно повысить точность распознавания горючего газа и количественной оценки состава газовой смеси, что позволяет проводить более надежные и точные измерения.

Инструментом для проведения измерений при газовом анализе является сенсор газа. В настоящее время существует ряд сенсоров, основными из которых считаются полупроводниковые, оптические и термokatалитические [3]. В данной работе исследование проводилось на данных, полученных с помощью термokatалитического сенсора. Такой тип сенсора является широко применяемым при измерениях концентраций горючего газа или смеси. Но ввиду того, что на поверхности сенсора

окисляются все присутствующие в окружающей среде газы, одним из недостатков является низкая избирательность к искомому типу газа. Предполагается, что применение методов машинного обучения позволит усовершенствовать существующие методики измерений и анализа газовой смеси.

В наших предыдущих работах [4, 5] представлены результаты, показавшие, что метод главных компонент (МГК) как способ, применяемый в области машинного обучения, позволяет различать отклики сенсоров при разной концентрации без использования дополнительных математических операций подготовки полученных данных. Также установлено, что при обработке исходных данных при помощи МГК вне зависимости от количества сенсоров полученные значения главных компонент образуют линейные зависимости концентрации, которые пропорциональны между собой. МГК позволяет как визуально различать отклики сенсоров при разной концентрации, так и с использованием дополнительных математических операций получить значение концентрации.

Другим направлением в области машинного обучения является классификация — получение однозначного ответа на основе набора признаков. Классификация необходима для газового анализа, в первую очередь для определения типа и состава различных газов. Это важно по нескольким причинам:

- безопасность (разные газы обладают разным уровнем взрывоопасности, благодаря точной классификации анализируемого газа можно принять надлежащие меры предосторожности для предотвращения несчастных случаев, утечек или вредного воздействия);

- воздействие на окружающую среду (некоторые газы оказывают значительное воздействие на окружающую среду и способствуют загрязнению воздуха, анализ и классификация газов помогают оценить присутствие и концентрацию таких газов, поддерживая усилия по снижению их выбросов и контролю за ними);

- оптимизация технологических процессов (во многих промышленных процессах газы используются в качестве реагентов, катализаторов или носителей, понимание свойств и состава этих газов с помощью классификации позволяет лучше оптимизировать процесс, что приводит к повышению эффективности и производительности).

В статье [6] предлагается метод классификации газов для системы электронного носа, для которого используются комбинированные признаки, полученные с помощью линейно-дискриминантного анализа. Результаты проведенного эксперимента показали, что для восьми типов данных газа, которые были измерены с помощью сенсорной матрицы, состоящей из 16 каналов, метод показал хоро-

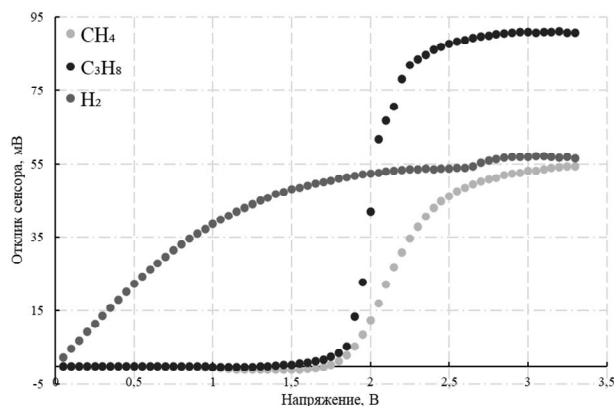
шую эффективность классификации. В частности, когда в процессе измерения генерируется шум, предложенный метод показал результаты классификации, которая является более эффективной, чем у других методов. Учитывая возможность тестирования на множестве сложных данных, в будущем будут проводиться эксперименты, сочетающие различные типы характеристик.

В статье [7] представлен метод, основанный на алгоритме ядерного дискриминантного анализа для идентификации типов газа. Было обнаружено, что точность алгоритма была на 4.17% выше, чем та, которая была получена с использованием МГК. В случае стандартного отклонения представленный алгоритм обладает самой высокой скоростью распознавания и наименьшими затратами времени.

Целью данной работы является изучение алгоритмов классификации машинного обучения, а также оценка точности применимых классификаторов. Важность данного исследования связана с оптимизацией параметров и характеристик термодаталитических сенсоров, в частности чувствительности и селективности, а также с необходимостью изменения методик измерения для получения большего числа данных для последующей математической обработки и повышения селективности сенсоров.

## МЕТОДИКА ЭКСПЕРИМЕНТА

Многомерный сигнал сенсора представляет собой совокупность значений отклика сенсора, измеренных при разном приложенном напряжении питания ( $U_{пит}$ ). На рис. 1 представлены экспериментально полученные S-образные зависимости отклика сенсора от напряжения питания для водорода, пропана и метана в чистом виде.

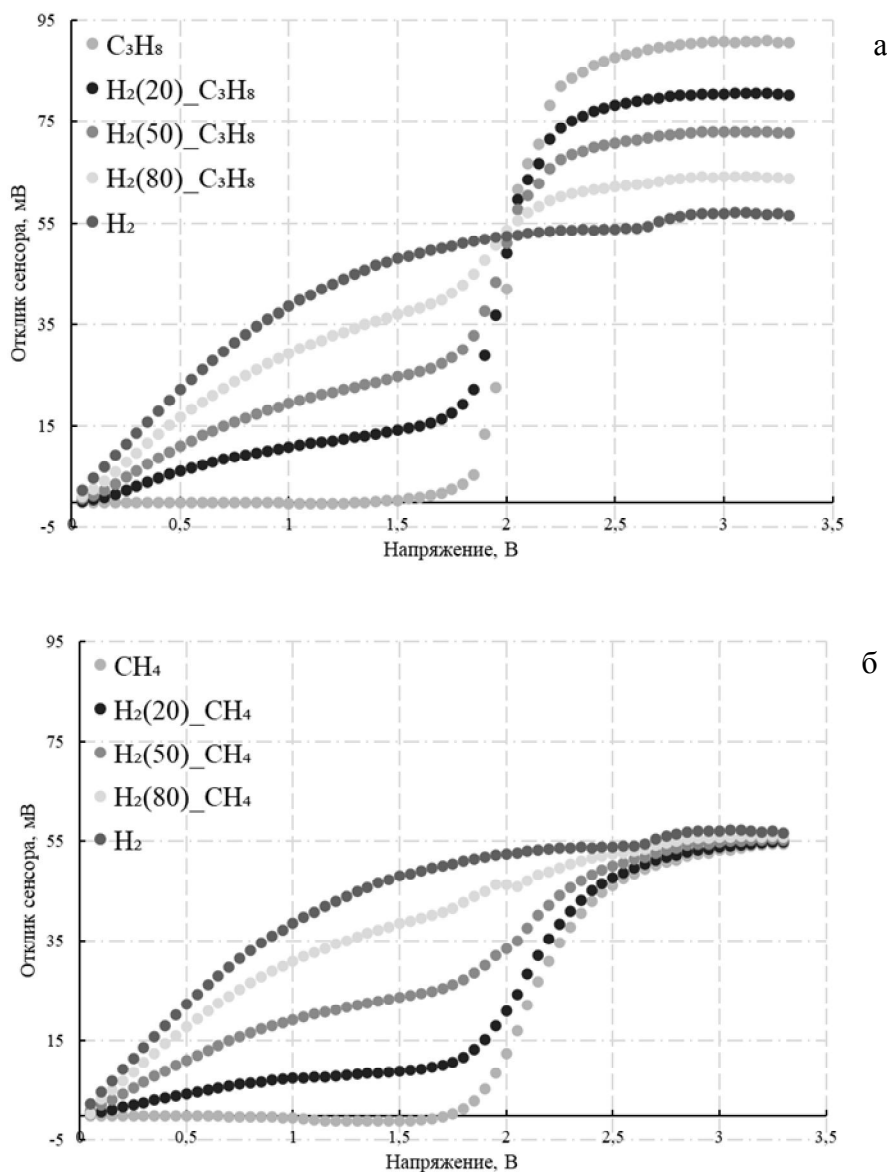


**Рис. 1.** Экспериментальные S-образные зависимости отклика сенсора от напряжения питания для водорода, пропана и метана

Измерения сигнала осуществлялись при сканировании с изменением напряжения от 0.05 В до 3.3 В с шагом 0.05 В. В качестве сенсоров были взяты промышленные термокаталитические сенсоры, которые серийно выпускаются фирмой "НТЦ ИГД" [8]. Термокаталитический сенсор был включен в мостовую измерительную схему, в которой в одном плече вместе с ним размещался сенсор сравнения (аналогичный по конструкции сенсор, но без нанесенного катализатора), а во втором плече — два прецизионных резистора с номиналом 1 кОм.

В соответствие с работой [5] и на основе полученных экспериментальных S-образных зависимостей были сформированы значения отклика сенсора на двухкомпонентные смеси (водород – пропан) и (водород – метан) с процентным содержанием  $H_2$ : 20, 50 и 80% (в дальнейшем — теоретические значения). Теоретические значения отклика сенсора представлены на рис. 2. На рис. 2, а, представлены значения для водорода, пропана и их смесей, на рис. 2, б, — для водорода, метана и их смесей.

Теоретические значения были обработаны при помощи МГК, результат обработки представлен на рис. 3.



**Рис. 2.** Теоретические значения отклика сенсора. а — для водорода, пропана и их смесей; б — для водорода, метана и их смесей

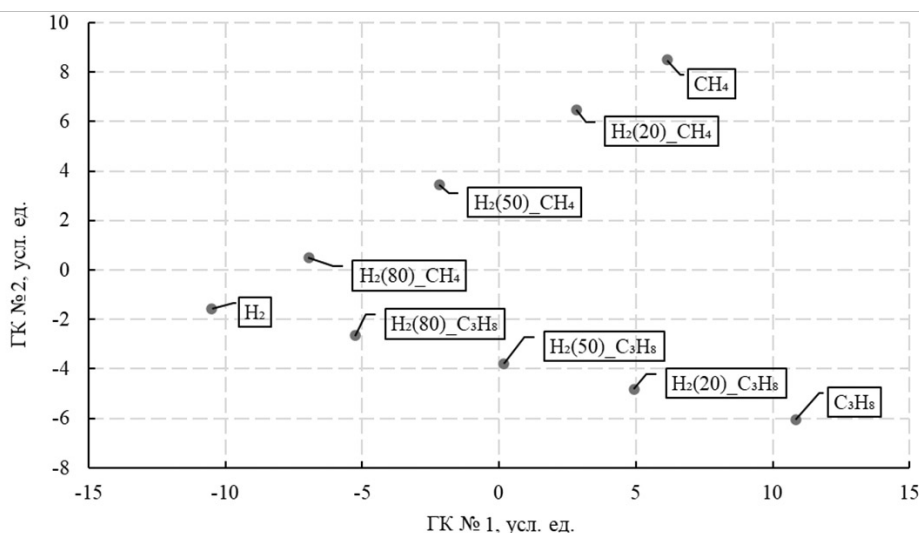


Рис. 3. Теоретические значения после применения МГК

Отмечено, что значения главных компонент образовали линейные зависимости и расположены в порядке возрастания концентрации водорода в смеси (слева — максимум концентрации (чистый водород), справа — минимум концентрации), формируя свои шкалы распределения концентрации отдельно для смесей (водород – метан) и (водород – пропан).

В контексте машинного обучения классификация относится к обучению с учителем. Данный тип обучения подразумевает, что на этапе обучения исходным данным, подаваемым на вход, уже присвоен класс, а важная часть признаков уже разделена на отдельные категории.

Существует несколько алгоритмов, используемых для задач классификации, в том числе:

- наивный байесовский алгоритм (вероятностный алгоритм предполагает независимость между объектами и вычисляет вероятность принадлежности экземпляра данных к каждому классу);

- метод опорных векторов (метод находит гиперплоскость, которая наилучшим образом разделяет точки данных разных классов в многомерном пространстве);

- метод *k*-ближайших соседей (предсказывает класс точки данных на основе большинства классов ее *k* ближайших соседей в пространстве объектов);

- логистическая регрессия (выводит прогнозы о точках в бинарном масштабе — нулевые или единичные; если значение чего-либо равно либо больше 0.5, то объект классифицируется в большую сторону (к единице), если значение меньше 0.5 — в меньшую (к нулю)).

Чтобы выполнить классификацию, модель машинного обучения обучается на помеченных дан-

ных, где каждая точка данных связана со своим правильным классом. Во время обучения модель изучает закономерности и взаимосвязи в данных, что позволяет ей делать точные прогнозы для испытательных примеров.

В данной работе мы рассмотрим методы классификации, перечисленные выше, и оценим их производительность по точности прогноза.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Применение классификации к результатам, полученным после обработки МГК, выполнено с помощью библиотеки *scikit-learn* [9] средствами языка Python 3.11. В качестве проверочных данных были использованы экспериментальные данные, полученные по методике, описанной в работе [10], для водорода, метана, пропана и их смесей.

После обработки экспериментальных откликов, полученных системой сенсоров, при помощи МГК были получены следующие результаты, представленные на рис. 4.

Как видно из полученного результата, экспериментальные данные образовали области локализации вокруг теоретических значений чистых углеводородов, водорода и их смесей. Разброс между полученными значениями главных компонент связан с разбросом экспериментальных данных (отклонения в снятых значениях обусловлены допуском, закладываемым на каждый сенсор в рамках производства). Также отмечено, что некоторые значения водорода и его смесей с пропаном в соотношении 50% и 80% находятся в окрестности теоретических точек смеси (водород – метан) в тех же соотношениях.

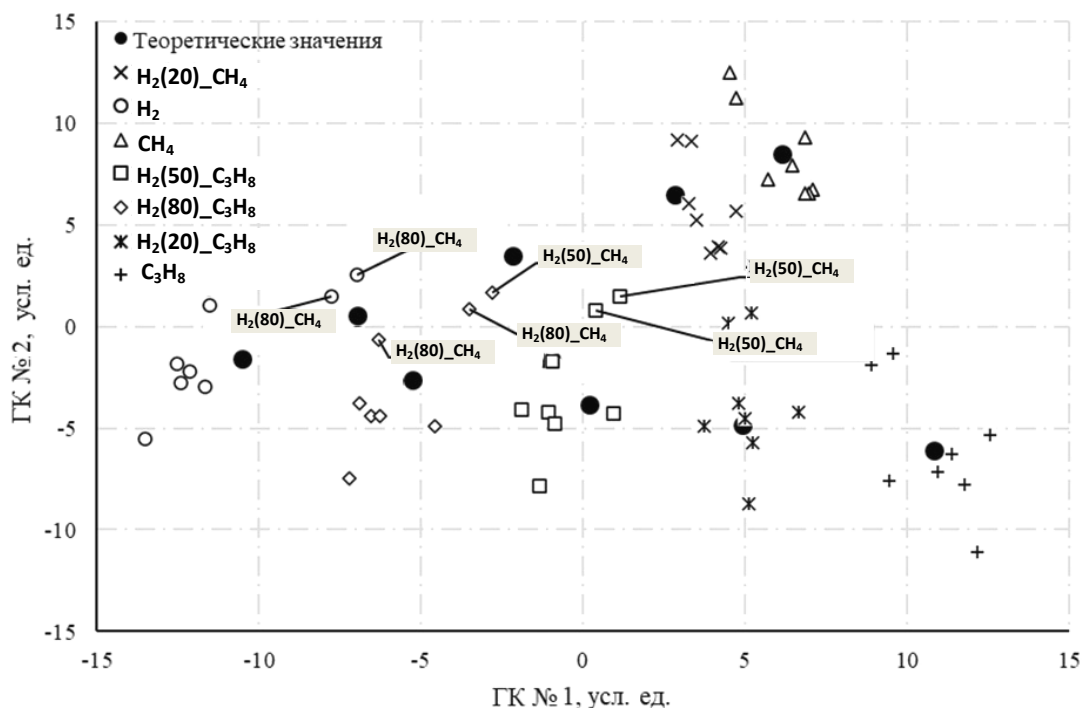


Рис. 4. Результат обработки экспериментальных данных при помощи МГК

Табл. Точность прогнозирования каждого метода классификации

№ п/п	Метод	Точность, %
1	Метод k-ближайших соседей	87.5
2	Логистическая регрессия	85.71
3	Наивный байесовский алгоритм	87.5
4	Метод опорных векторов	87.5

Предполагается, что при применении классификаторов возможно неправильное определение типа газа. Результат определения представлен на рис. 4, где неверно классифицированные значения обозначены подписями на сером фоне. Точность прогнозирования составила 87.5%. Аналогично были рассмотрены остальные методы, результаты прогнозирования которых сведены в таблицу.

Как видно из полученных результатов, большинство алгоритмов корректно определяет тип газа. Различие по точности логистической регрессии объясняется тем, что данный классификатор является линейным и чаще используется на тех данных, у которых есть линейная зависимость. Ввиду того, что исходные данные имеют нелинейную зависимость, использование логистической

регрессии в задаче классификации является нецелесообразным, хотя и имеет достаточную точность.

Стоит отметить, что точность классификаторов возможно повысить путем применения большего количества экспериментальных данных или предварительной подготовкой снятых данных с термокаталитического сенсора (приведение исходных данных к одному общему формату данных). Также возможным вариантом повышения точности классификаторов является уточнение и переопределение теоретических значений.

#### ЗАКЛЮЧЕНИЕ

В работе было проведено исследование применения классификаторов для определения типа газа

на основе данных, полученных с термодаталитических сенсоров, при помощи МГК в мультисенсорной системе. Применение классификаторов позволяет с точностью до 87.5% распознать тип газа или двухкомпонентную смесь.

Показано, что при обработке исходных данных при помощи МГК полученные значения главных компонент образуют линейные зависимости концентрации, которые пропорциональны между собой, а также не имеют пересечений для разных углеводородов. В данной работе представлены такие углеводороды, как метан и пропан, обработка значений которых сформировала две линейные отдельные зависимости. Это позволило визуально различать отклики сенсоров при разной концентрации.

Полученные результаты имеют важное значение для разработки и оптимизации методов анализа состава воздуха, в частности позволяют перейти от пересчетов концентрации "внутри" сенсора к математической обработке получаемых с сенсора данных и на ее основе проводить определение концентрации поданного газа.

Также, ввиду того что водород становится топливом будущего, новым энергоносителем (его роль будет сопоставима с ролью угля, нефти или газа), разработанный метод общеприменим и перспективен для дальнейшего использования в сенсорах газа, не только сенсорах водорода, которые будут актуальны в ближайшем будущем, но и при распознавании других целевых газов.

## СПИСОК ЛИТЕРАТУРЫ

1. Попадьюко Н.В., Рожнятовский Г.И., Дауди Д.И. Водородная энергетика и мировой энергопереход // Инновации и инвестиции. 2021. № 4. С. 59–64. URL: <https://elibrary.ru/item.asp?id=45723109>
2. Грушевенко Е., Капитонов С., Мельников Ю. и др. Декарбонизация нефтегазовой отрасли: международный опыт и приоритеты. Центр энергетике МШУ СКОЛКОВО, 2021. 158 с. [Электронный ресурс]. URL: [https://energy.skolkovo.ru/downloads/documents/SEneC/Research/SKOLKOVO\\_EneC\\_Decarbonization\\_of\\_oil\\_and\\_gas\\_RU\\_22032021.pdf](https://energy.skolkovo.ru/downloads/documents/SEneC/Research/SKOLKOVO_EneC_Decarbonization_of_oil_and_gas_RU_22032021.pdf) (дата обращения: 01.08.2023).
3. Баранов А.М., Осипова Т.В. Современные тенденции в развитии сенсоров взрывоопасных концентраций горючих газов и паров горючих жидкостей (Краткий

- обзор) // Научное приборостроение. 2021. Т. 31, № 4. С. 3–29. URL: <http://iairas.ru/mag/2021/abst4.php#abst1>
4. Осипова Т.В., Баранов А.М., Иванов И.И. Метод главных компонент как альтернативный алгоритм обработки данных термодаталитических сенсоров // Научное приборостроение. 2022. Т. 32, № 1. С. 77–92. URL: <http://iairas.ru/mag/2022/abst1.php#abst8>
  5. Осипова Т.В., Баранов А.М., Иванов И.И. Метод главных компонент как способ определения концентрации водорода в многокомпонентных смесях // Научное приборостроение. 2023. Т. 33, № 2. С. 24–34. URL: <http://iairas.ru/mag/2023/abst2.php#abst3>
  6. Choi S.-I., Eom T., Jeong Gu-Min. Gas classification using combined features based on a discriminant analysis for an electronic nose // Journal of Sensors. 2016. Vol. 2016. Id. 9634387. DOI: 10.1155/2016/9634387
  7. Li H., Luo D., Sun Y., Gholamhosseini H. Classification and identification of industrial gases based on electronic nose technology // Sensors. 2019. Vol. 19. Id. 5033. DOI: 10.3390/s19225033
  8. Карпов-сенсоры. Производство термодаталитических сенсоров горючих газов. [Электронный ресурс] URL: <http://karpov-sensor.com/> (дата обращения 01.08.2023).
  9. Scikit-learn: machine learning in Python. [Электронный ресурс] URL: <https://scikit-learn.org> (дата обращения 09.08.2023).
  10. Иванов И.И., Баранов А.М., Талипов В.А., Мионов С.М., Колесник И.В., Напольский К.С. Разработка эффективных сенсоров обнаружения взрывоопасных концентраций H<sub>2</sub> // Научное приборостроение. 2021. Т. 31, № 3. С. 25–36. URL: <http://iairas.ru/mag/2021/abst3.php#abst4>

**Московский авиационный институт (национальный исследовательский университет), Москва, Россия**

Контакты: Осипова Татьяна Владиславовна,  
t.osipova.95@mail.ru

Материал поступил в редакцию 11.09.2023

# EVALUATION OF THE EFFECTIVENESS OF MACHINE LEARNING CLASSIFIERS FOR GAS TYPE AND CONCENTRATION RECOGNITION

T. V. Osipova, A. M. Baranov, I. I. Ivanov

*Moscow Aviation Institute (National Research University), Moscow, Russia*

The paper highlights the artificial intelligence (AI) classification methods in machine learning for recognizing the gas type and its concentration in a mixture. The applicability of classifiers is assessed. During the study, theoretical values of hydrogen, hydrocarbons, and their mixtures with a hydrogen fraction of 20, 50 and 80% were calculated, and the AI classifiers were evaluated using experimental data obtained from a catalytic sensor. The presented classifiers made it possible to determine the type of gas with an accuracy of up to 87.5%.

*Keywords:* catalytic sensor, principal component analysis, classification, determining concentration, hydrogen, data processing

## REFERENCES

1. Popadko N.V., Rozhnyatovsky G.I., Daudi D.I. [Hydrogen energy and the global energy transition]. *Innovatsii i investitsii* [Innovation and Investment], 2021, no. 4, pp. 59–64. URL: <https://elibrary.ru/item.asp?id=45723109> (In Russ.).
2. Grushevenko E., Kapitonov S., Mel'nikov Yu., et al. *De-karbonizatsiya neftegazovoi otrasli: mezhdunarodnyi opyt i priority* [Decarbonization of the oil and gas industry: international experience and priorities]. Tsentr ehnergetiki MSHU SKOLKOVO, 2021. 158 p. URL: [https://energy.skolkovo.ru/downloads/documents/SEneC/Research/SKOLKOVO\\_EneC\\_Decarbonization\\_of\\_oil\\_and\\_gas\\_RU\\_22032021.pdf](https://energy.skolkovo.ru/downloads/documents/SEneC/Research/SKOLKOVO_EneC_Decarbonization_of_oil_and_gas_RU_22032021.pdf) (accessed 01.08.2023). (In Russ.).
3. Baranov A.M., Osipova T.V. [Recent trends in the development of sensors for pre-explosive concentrations of flammable gases and vapors of flammable liquids (review)]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2021, vol. 31, no. 4, pp. 3–29. DOI: 10.18358/np-31-4-i329 (In Russ.).
4. Osipova T.V., Baranov A.M., Ivanov I.I. [Principal component analysis as an alternative algorithm for processing data of thermocatalytic sensor]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2022, vol. 32, no. 1, pp. 77–92. DOI: 10.18358/np-32-1-i7792 (In Russ.).
5. Osipova T.V., Baranov A.M., Ivanov I.I. [The principal component analysis as a method for determining the hydrogen concentration in multicomponent mixtures]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2023, vol. 33, no. 2, pp. 24–34. (In Russ.). URL: <http://iairas.ru/mag/2023/abst2.php#abst3>
6. Choi S.-H., Eom T., Jeong Gu-Min. Gas classification using combined features based on a discriminant analysis for an electronic nose. *Journal of Sensors*, 2016, vol. 2016, id. 9634387. DOI: 10.1155/2016/9634387
7. Li H., Luo D., Sun Y., GholamHosseini H. Classification and identification of industrial gases based on electronic nose technology. *Sensors*, 2019, vol. 19, id. 5033. DOI: 10.3390/s19225033
8. Karpov-sensors. *Proizvodstvo termokataliticheskikh sensorov goryuchikh gazov* [Production of thermocatalytic sensors of combustible gases]. URL: <http://karpov-sensor.com/> (accessed 01.08.2023). (In Russ.).
9. Scikit-learn: machine learning in Python. URL: <https://scikit-learn.org> (accessed 09.08.2023). (In Russ.).
10. Ivanov I.I., Baranov A.M., Talipov V.A., Mironov S.M., Kolesnik I.V., Napolskii K.S. [Development of effective sensors for detecting pre-explosive H<sub>2</sub> concentrations]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2021, vol. 31, no. 3, pp. 25–36. DOI: 10.18358/np-31-3-i2536 (In Russ.).

Contacts: *Osipova Tat'yana Vladislavovna*,  
t.osipova.95@mail.ru

Article received by the editorial office on 11.09.2023

## INTRODUCTION

To date, measurements and analysis of gas mixture composition play a key role in various fields, including environmental monitoring and industrial production. The ability to accurately identify, quantify, and classify gaseous substances plays an important role in life safety, process optimization, and environmental diagnostics. Traditional gas analysis methods rely heavily on standard measurement techniques and manual interpretation, often facing inherent limitations in detection sensitivity, labor intensity of analysis, and susceptibility to human error.

Among the huge number of different hydrocarbons, hydrogen occupies a special place. Having the widest range of explosive concentrations (from 4 to 75% vol.), and also being extremely volatile, hydrogen is considered one of the most explosive gases. Hydrogen is also considered as a promising type of fuel (or fuel component, in particular, in mixtures with hydrocarbons) [1, 2], development of effective hydrogen sensors and methods for its monitoring in the air is an urgent task.

In recent years, with the advent of machine learning, there has been a key update in the application of artificial intelligence, the development of expert systems, and the emergence of new approaches to processing large amounts of data. This area of artificial intelligence enables the development of innovative approaches to automate complex tasks, increase efficiency, and gain valuable insights from vast data sets. Using the power of machine learning methods, researchers are modernizing gas analysis methods, contributing to the development of this important scientific field.

Using machine learning algorithms, gas analysis can move beyond standard practices into a more accurate, efficient, and comprehensive approach. The ability to quickly process huge amounts of data and extract patterns, combined with the decision-making speed of algorithms, will lead to the modernization or development of new methods in gas analysis methodologies. Moreover, machine learning models can significantly improve the accuracy of flammable gas recognition and quantitative assessment of the composition of the gas mixture, which allows for more reliable and accurate measurements.

The instrument for carrying out measurements in gas analysis is a gas sensor. Currently, there are a number of sensors, the main ones of which are considered semiconductor, optical, and thermocatalytic [3]. In this work, the study was carried out on data obtained using a thermocatalytic sensor. This type of sensor is widely used when measuring the concentrations of a flammable gas or mixture. But due to the fact that all gases present in the environment are oxidized on the surface of the sensor, one of the disadvantages

is low selectivity for the desired type of gas. It is expected that the use of machine learning methods will improve existing methods for measuring and analyzing a gas mixture.

Our previous works [4, 5] presented results showing that principal component analysis (PCA), as a method used in the field of machine learning, makes it possible to distinguish sensor responses at different concentrations without the use of additional mathematical operations for preparing the obtained data. It was also found that when processing initial data using PCA, regardless of the number of sensors, the resulting values of the principal components form linear concentration dependences that are proportional to each other. PCA allows you to both visually distinguish sensor responses at different concentrations and, using additional mathematical operations, obtain the concentration value.

Another direction in the field of machine learning is classification: obtaining an unambiguous answer based on a set of features. Classification is necessary for gas analysis, primarily to determine the type and composition of various gases. This is important for several reasons:

- safety (different gases have different levels of explosiveness; thanks to the precise classification of the gas being analyzed, proper precautions can be taken to prevent accidents, leaks, or harmful effects);
- environmental impact (some gases have a significant impact on the environment and contribute to air pollution; gas analysis and classification help assess the presence and concentration of such gases, supporting efforts to reduce and control their emissions);
- optimization of technological workflow (many industrial processes use gases as reagents, catalysts, or carriers; understanding the properties and composition of these gases through classification allows for better optimization of the process, resulting in increased efficiency and productivity).

In article [6], a method for classifying gases using the electronic nose system is proposed. This method makes use of combined features obtained by linear discriminant analysis. The experiment's findings indicated that a sensor array with 16 channels performed well in classifying the eight different categories of collected gas data. In particular, when noise is generated during the measurement, the proposed method has shown classification results that are more efficient than other methods. Given the ability to test on a variety of complex data, future experiments will combine different types of characteristics.

Article [7] presents a method based on the nuclear discriminant analysis algorithm for identifying gas types. It was found that the accuracy of the algorithm was 4.17% higher than that obtained using PCA. In the case of standard deviation, the presented algorithm



has the highest recognition speed with the least amount of time.

The purpose of this work is to study machine learning classification algorithms as well as assess the accuracy of applicable classifiers. The importance of this study is related to the optimization of the parameters and characteristics of thermocatalytic sensors, in particular sensitivity and selectivity, as well as the need to change measurement techniques to obtain more data for subsequent mathematical processing and increase the selectivity of sensors.

## EXPERIMENTAL PROCEDURE

The multidimensional sensor signal is a set of sensor response values measured at different applied supply voltages ( $U_{\text{пит}}$ ). Fig. 1 shows the experimentally obtained S-shaped dependences of the sensor response on the supply voltage for hydrogen, propane, and methane in their pure form.

**Fig. 1.** Experimental S-shaped dependences of the sensor response on supply voltage for hydrogen, propane, and methane

The signal measurements were carried out by scanning with a voltage change from 0.05 V to 3.3 V with a step of 0.05 V. Industrial thermocatalytic sensors, which are commercially produced by the company STC IGD LLC (NTC IGD) [8], were used. The thermocatalytic sensor was included in a bridge measuring circuit, in which a reference sensor (a sensor similar in design but without an applied catalyst) was placed in one branch along with it, and two precision resistors with a nominal value of 1 k $\Omega$  were placed in the second branch.

In accordance with work [5], and on the basis of the obtained experimental S-shaped dependencies, the values of the sensor response to two-component mixtures (hydrogen - propane) and (hydrogen - methane) with a percentage of H<sub>2</sub>: 20, 50 and 80% (below — theoretical values) were calculated. The theoretical values of the sensor response are presented in Fig. 2. In Fig. 2, a, values for hydrogen, propane and their mixtures are presented, in Fig. 2, b, — for hydrogen, methane, and their mixtures.

**Fig. 2.** Theoretical sensor response values. a — for hydrogen, propane, and their mixtures; b — for hydrogen, methane, and their mixtures

Theoretical values were processed using PCA, the result is presented in Fig. 3.

**Fig. 3.** Theoretical values after applying PCA

It was noted that the values of the main components create a linear dependence and are located in increasing order of hydrogen concentration in the mixture (on the left is the maximum concentration (pure hydrogen), on the right is the minimum concentration), forming its own concentration distribution scale, separately for mixtures: (hydrogen – methane) and (hydrogen – propane).

In the context of machine learning, classification refers to supervised learning. This type of training implies that at the training stage, the initial data supplied to the input has already been assigned a class, and an important part of the features has already been divided into separate categories.

There are several algorithms used for classification problems, including:

- naive Bayes classifier (the probabilistic algorithm assumes independence between objects and calculates the probability of a data instance belonging to each class);
- support vector machine (the method determines the hyperplane that best separates data points of different classes in multidimensional space);
- $k$ -nearest neighbors algorithm (predicts the class of a data point based on the majority of classes of its  $k$  nearest neighbors in object space);
- logistic regression (predicts points on a binary scale — zero or one; if the value of something is equal to or greater than 0.5, then the object is classified upward (towards one), if the value is less than 0.5 — downwards (toward zero)).

To perform classification, a machine learning model is trained on labeled data, where each data point is associated with its correct class. During training, the model learns patterns and relationships in the data, which allows it to make accurate predictions for test cases.

In this paper, we review the classification methods listed above and evaluate their performance in terms of prediction accuracy.

## RESULTS AND DISCUSSION

The application of classification to the results obtained after the PCA was carried out using the scikit-learn library [9] and the Python 3.11 language. Experimental data obtained using the method described in [10] for hydrogen, methane, propane, and their mixtures were used as verification data.

The following results were obtained (Fig. 4) after PCA processing the experimental responses received by the sensor system.

**Fig. 4.** The result of processing experimental data using PCA

As can be seen from the obtained result, the experimental data formed localization regions around the theoretical values of pure hydrocarbons, hydrogen, and their mixtures. The spread between the obtained values of the principal components is associated with the spread of experimental data (deviations in the measured values are due to the tolerance provided for each sensor as part of production). It is also noted that some values of hydrogen and its mixtures with propane in a ratio of 50% and 80% are in the vicinity of the theoretical points of the mixture (hydrogen - methane) in the same ratios.

It is assumed that when using classifiers, an incorrect determination of the type of gas is possible. The result of the determination is presented in Fig. 4, where incorrectly classified values are indicated by labels on a gray background. The prediction accuracy was 87.5%. The other methods were considered similarly, the prediction results of which are presented in Tab.

**Tab.** Prediction accuracy of each classification method

As can be seen from the results obtained, most algorithms correctly determine the type of gas. The deviation in accuracy of logistic regression is explained by the fact that this classifier is linear and is more often used on data that has a linear relationship. Due to the fact that the source data has a nonlinear relationship, the use of logistic regression in the classification

problem is inappropriate, although it has sufficient accuracy.

It is worth noting that the accuracy of classifiers can be increased by using a larger amount of experimental data or by preliminary preparing the data recorded by a thermocatalytic sensor (bringing the original data into one common data format). Another possible option for increasing the accuracy of classifiers is to refine and redefine the theoretical values.

## CONCLUSION

The technique described in the paper uses PCA and data from thermo-catalytic sensors in a multisensor system to apply classifiers to identify the type of gas. The use of classifiers makes it possible to recognize the type of gas or two-component mixture with an accuracy of 87.5%.

It is shown that when using PCA to process the initial data, the obtained values of the principal components form linear dependences of concentration, which are proportional to each other and also do not have intersections for various hydrocarbons. This work presents hydrocarbons such as methane and propane, the processing of the values of which formed two linear separate dependencies. This made it possible to visually distinguish sensor responses at different concentrations.

The results obtained are important for the development and optimization of methods for analyzing air composition; in particular, they allow us to move from recalculating the concentration "inside" the sensor to mathematical processing of data received from the sensor and, on its basis, to determine the concentration of the supplied gas.

Also, due to the fact that hydrogen is becoming the fuel of the future, and a new energy carrier (its role will be comparable to that of coal, oil, or gas), the developed method is generally applicable and promising for further use in gas sensors, not only hydrogen sensors, which will be relevant in the near future, but also when recognizing other target gases.