

## СОДЕРЖАНИЕ

Аннотация.....	1
ВВЕДЕНИЕ .....	2
The background of the study.....	2
The problem statement .....	3
The professional significance.....	4
ОБЗОР ЛИТЕРАТУРЫ.....	4
МЕТОДОЛОГИЯ .....	8
ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ .....	8
СПИСОК ЛИТЕРАТУРЫ.....	9

### Аннотация

Система «электронный нос» получает все больше приложений во многих областях, от оценки качества пищевых продуктов до обнаружения запрещенных грузов и диагностики некоторых заболеваний. Благодаря развитию электроники, появляются портативные приборы, пригодные для экспресс-анализа. Однако, интерпретация результатов требует значительного времени и присутствия эксперта.

Исследование методов интеллектуального анализа данных, которые позволили бы автоматически интерпретировать показания сенсоров – цель данного проекта.

## ВВЕДЕНИЕ

### The background of the study

В настоящее время в аналитической химии для исследования многокомпонентных смесей летучих соединений значительное развитие получает технология «электронный нос», благодаря низкой цене (в сравнении с газовыми хроматографами) и универсальности прибора. Такие приборы строятся на основе массива пьезосенсоров, высокочувствительных к наиболее распространённым соединениям. Количество сенсоров в этих системах может варьироваться в зависимости от решаемой задачи. В качестве выходных данных, доступных для дальнейшей обработки и анализа, прибор предоставляет матрицу откликов сенсоров, время исследования и некоторую служебную информацию.

Зарубежный рынок систем «электронный нос» представлен следующими приборами: FOX с массивом из 6 металл-оксид-электронные сенсоров, и более продвинутая версия GEMINI (до 18 сенсоров), комбинированная система «электронный нос» с газовым хроматографом HERACLES производства компании Alpha MOS, портативный газовый хроматограф zNose GS/SAW от Electronic Sensor Technology Inc, MOSES II немецкой компании GSG Meß- und Analysengeräte Vertriebsgesellschaft mbH и KAMINA (Германия). В России подобный прибор разработан в Воронежском государственном университете инженерных технологий группой под руководством доктора химических наук, профессора Татьяны Анатольевны Кучменко и называется МАГ-8.

Большой интерес в настоящее время представляет разработка интегрированных аналитических систем, образующих единый конвейер, начиная с измерения, через обработку и интеллектуальный анализ данных и заканчивая принятием решения. При этом, необходимо учитывать так же специфику подобных исследований: сначала исследователь определяется с характером веществ-маркеров, с которыми он планирует работать, исходя из этого подбирает селективные покрытия электродов пьезокварцевых

резонаторов, затем тестирует этот массив, проводя первичные пробы отдельных веществ и их возможных смесей, и только после этого приступает к оценке тех объектов, которым посвящено исследование. Таким образом, несмотря на то, что методы анализа данных могут многое рассказать о каждом веществе из набора, применять их к отдельным матрицам откликов нецелесообразно из-за большого числа этих матриц. Этот этап должен быть либо скрыт от конечного пользователя, либо от него следует отказаться в пользу более высокоуровневых подходов. В качестве основного подхода здесь может быть использована классификация на  $N$  классов.

### The problem statement

По практическим причинам, настоящее исследование ограничено узким кругом методов интеллектуального анализа данных: **анализ временных рядов, поскольку матрицу откликов можно рассматривать как развивающийся во времени процесс, подбор авторегрессионных моделей, алгоритмы машинного обучения и нейронных сетей.** Данное исследование не нацелено на разработку строго графических методов, так как подобный метод разработан, апробирован и успешно используется группой профессора Кучменко уже несколько лет. Визуальная составляющая анализа будет представлена в виде **обобщенных** графиков. Кроме того, целью работы не является разработка принципиально новых математических методов анализа матриц откликов сенсоров.

Основная проблема применения традиционных алгоритмов анализа данных состоит в том, что мы имеем дело с чрезвычайно малыми и несбалансированными выборками данных: несколько десятков первичных проб и значительное количество новых объектов, которым необходимо назначит  $N$  меток, соответственно тому, какие веществам-маркеры обнаружил алгоритм в матрице откликов. В виду специфичности сорбентов, обучать алгоритм на выборках, подготовленных для других экспериментов, невозможно, поскольку, очевидно, наборы веществ-маркеров для оценки качества столовых вин и для оценки качества полимеров, из которых

изготовлены детские игрушки, будут существенно отличаться. Отсюда возникает необходимость понимания природы этих матриц, которое в результате получить возможность генерировать искусственные данные и эффективно обучать алгоритм. Важное замечание: поскольку смесей представлено 4 штуки, вопрос генерации смесей останется в стороне.

Таким образом, основная цель моей работы состоит в том, чтобы подобрать способ генерации достоверных искусственных данных и обучить алгоритм из матриц проб, взятых с 75 детских игрушек, обнаруживать присутствие наиболее полного набора веществ-маркеров. Оценка результатов будет проведена путем сравнения полученных меток классов с результатами исследования группы профессора Кучменко.

### **The professional significance**

Для достижения поставленной цели, необходимо решить следующие задачи: запустить основной набор алгоритмов машинного обучения на начальных данных и выяснить, на какой результат можно рассчитывать, с помощью методов анализа временных рядов выяснить основные свойства рассматриваемых данных (как веществ и смесей, так и проб игрушек), построить модель опорных веществ и обучиться на сгенерированных данных. В случае неудачи – выяснить причины, которые не позволяют генерировать синтетические данные.

Поставленная задача может представлять интерес как для специалистов в области анализа данных как иллюстрация универсальности применяемых методов, так и для специалистов-химиков, которые получают модель работы с подобными данными.

### **ОБЗОР ЛИТЕРАТУРЫ**

Следующий обзор основан на изучении исследований ученых-химиков, описывающих применение системы «электронный нос» и анализ результатов, и статьях из области анализа данных, описывающих общие подходы к работе в системах искусственного обоняния.

Основной подход, применяемый в данный момент профессором Кучменко, состоит в анализе визуальных отпечатков откликов сенсоров (кинетических и максимумов) в равновесной газовой фазе. Эти графики имеют вид полярной диаграммы, где осями являются временные метки, а факторами – значения сенсоров в момент времени  $t$ . В [1] делается вывод о том, что «электронный нос» «МАГ-8» превосходит физико-химические показатели в задаче оценки органолептических характеристик вина.

Сравнительный анализ возможностей интегрального анализатора газа “VOCmeter” (Германия) и дифференциального анализатора «МАГ-8» приведен в [2] на примере задачи количественной и качественной оценки легколетучей фракции ароматических добавок для мясного сырья. Сделан вывод о том, что результат, получаемые с использованием отечественной разработки превосходят результаты “VOCmeter” и в большей степени коррелируют с результатами газохроматографии.

Для идентификации веществ по визуальным отпечаткам используется расчет таких геометрических параметров фрагментов фигуры визуальных отпечатков, как площади под кривыми  $i$ -х пьезосенсоров  $Si$ , площадь «визуального отпечатка» массива сенсоров, соотношение проекций сигналов сенсоров  $i$  и  $j$  на сигнал сенсора  $n$  и угол между этими проекциями (в радианах). Диссертация [3] на соискание степени кандидата химических наук Дроздовой Е.В. под руководством Кучменко Т.А. полностью посвящена апробации данного подхода в задаче оценки безопасности изделий из полимерных материалов на основе проб воздуха в локальных точках вблизи их поверхности. Кроме того, в данной работе показана возможность применения PCA и кластеризации как методов обработки данных, получаемых с помощью электронного носа «Маг-8».

Большое количество работ [4][5] посвящено подбору массивов сенсоров для решения определенных задач, что говорит о специфичности этих сенсоров и позволяет предположить, что обучение универсального алгоритма

классификации невозможно – для решения каждой отдельной задачи он должен обучаться на отдельном наборе.

Основной упор в рассмотренных выше работах делается на применение данной системы в экспресс-анализе, что так же обосновывает необходимость разработки такой системы анализа данных, которая бы позволила ускорить процесс принятия решения исследователем.

Такие системы, построенные на основе двух компонентов - «электронного носа» и автоматизированной системы распознавания, нашли применение в медицине, охране окружающей среды и пищевой промышленности. В работе Келлера [6] описан прототип такой системы и показан успешный пример применения как нейронных сетей, обученных методом обратного распространения ошибки, так и сетей fuzzy ARTMAP, сочетающих в себе аппарат нечеткой логики и адаптивной резонансной теории. Обе архитектуры показали близкую точность – 92.9% и 93.4%, соответственно. Необходимо так же заметить, что обучение проводилось на сравнительно небольшой для нейронных сетей выборке в 619 объектов, а тестирование – на 196 объектах.

Еще в 1989 году в своей работе [7] Hoffheins показал, что, благодаря использованию многоканальных массивов сенсоров, нейронные сети успешно решают задачу распознавания паров летучих соединений, поскольку количество распознаваемых химических веществ в общем случае больше числа сенсоров. Эта фундаментальная работа дала важные результаты как по организации массива сенсоров, так и основу для дальнейших исследований архитектур нейронных сетей и способов представления входных данных. В работе предложен способ представления данных и показано, что худший результат показали сети Хопфилда, вероятно, и-за малой размерности входных данных, а наилучший – сеть Больцмана, которая не только возвращала лучшую метку, но также показывала следующего подходящего кандидата, что позволяло использовать эту сеть в задаче отображения концентрация газа. Сеть Хэмминга показала наилучший результат в

распознавании смесей многих компонентов. Кроме того, показано, что алгоритмы кластеризации способны успешно разделять гексан и этанол, а также высокие и низкие концентрации смесей воды и этанола. Еще один важный результат данной работы состоит в том, что было найдено важное ограничение: сеть не способна распознавать неизвестные ей смеси веществ, присутствовавших в обучающей выборке.

Еще один успешный пример применения системы «электронный нос» описан в [8]: рассматривается целый набор задач по проверке грузов, которые каждый день решают сотрудники службы безопасности и таможенной службы в портах – обнаружение наркотических веществ, споров грибов и плесени, которые могут угрожать сельскохозяйственным культурам, опасных химикатов. В качестве метода выбрано построение ароматических профилей каждого контейнера в виде полярных графиков. Интересным так же является предложение использовать эти профили как своеобразные «контрольные суммы» контейнеров, изменение которых можно было бы отслеживать на протяжении всего маршрута и таким образом выявлять, в каком из портов к содержимому контейнеров был добавлен контрабандный товар. Важным отличием от прочих работ является использование единственного сенсора, что существенно сокращает стоимость применения такого устройства.

Работа [9] описывает все существующие подходы к анализу данных в задаче распознавания запахов, включая широкий спектр графических методов таких, как полярные диаграммы и иерархическая кластеризация, РСА, алгоритмы кластеризации и классификации, линейный и квадратичный дискриминантный анализ, нейронные сети, методы нечеткой логики и генетические алгоритмы. Делается особый акцент на необходимость нормализации данных и приводится около десятка способов нормализации, рассматриваются способы отбора признаков (сенсоров). В целом, данная работа скорее обобщает существующие подходы, чем привносит что-то принципиально новое.

Недостатком предыдущих работ по анализу данных является то, что они не освещают возможности многослойных нейронных сетей. Статья [10] восполняет этот пробел. Авторы рассматривают влияние смесей на качество распознавания (для простоты, берут смеси только двух веществ) и сравнивают данные от двух видов сенсоров (пьезокварцевых и металл-оксидных) и делают вывод о том, что электронные носы можно применять во многих областях (не может быть!).

## МЕТОДОЛОГИЯ

В файле [Statistics.pdf](#) лежит анализ временных рядов и в файле [Nose\\_report.pdf](#) лежат алгоритмы машинного обучения и генерация данных.

## ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ

Результаты данного исследования затрагивают вопросы генерации искусственных матриц откликов сенсоров на представленные в тренировочном наборе вещества, чтобы охватить как можно больше различных концентраций и возможных состояний. В качестве ожидаемого результата будет предложен подход к генерации таких данных и наиболее подходящий алгоритм распознавания запахов и их смесей в ароматических профилях детских игрушек.

Проверка достоверности полученных результатах будет проводиться сравнением полученных меток с результатам исследования группы под руководством Т.А. Кучменко. Данный метод позволит оценить качество сгенерированных данных и способность алгоритма классификации обнаружить как можно больше составляющих в матрице.



## СПИСОК ЛИТЕРАТУРЫ

1. Т.А. Кучменко, Р.П. Лисицкая, А.А. Шуба, Информативность анализатора газов «электронный нос» для оценки качества вина. - Аналитика и контроль. - №4. – 2014.
2. Т.А. Кучменко, Д.А. Погребная, Сравнительная оценка возможностей интегрального и дифференциального анализаторов газа типа «электронный нос» для исследования мясных продуктов. - Аналитика и контроль. - №3. – 2011.
3. Е.В. Дроздова, Определение органических легколетучих токсикантов массивом пьезосенсоров для оценки безопасности полимерных материалов: диссертация ... кандидата химических наук: 02.00.02 / Дроздова Евгения Викторовна; [Место защиты: Воронеж. гос. ун-т]. - Воронеж, 2016. - 263 с.: ил
4. An integrated chemical sensor array using carbon black polymers and a standard cmos process, **год??**
5. Chemical Sensors for Electronic Nose Systems, 2004
6. Electronic noses and their applications, 1995
7. Using Sensor Arrays and Pattern Recognition to Identify Organic Compounds, 1989
8. Homeland Security, Olfactory Images, and Virtual Chemical Sensors
9. Data analysis for electronic nose systems, 2005
10. Intelligent Classification of Odor Data Using Neural Networks, 2012