УДК 004.89

DOI: 10.15827/0236-235X.140.715-728

Дата подачи статьи: 30.06.22, после доработки: 04.07.22 2022. Т. 35. № 4. С. 715–728

Информационная и алгоритмическая поддержка интеллектуальной системы экологического мониторинга воздуха на основе нейронных сетей

 Γ .А. Ярыгин 1 , д.т.н., профессор, научный руководитель, office@diem.ru

M.В. Баюкин 1 , к.т.н., зам. директора, bayukin@diem.ru

 $B.\Phi.$ Корнюшко 2 , $\partial.m.н.$, профессор, vfk256@mail.ru

Е.Г. Шмакова ², к.т.н., доцент, 89254225204@mail.ru

Λ.В. Садеков ², acnupaнm, leosadekov@yandex.ru

- ¹ Научно-производственная фирма «ДИЭМ», г. Москва, 117485, Россия
- 2 МИРЭА Российский технологический университет, г. Москва, 119454, Россия

Работа посвящена алгоритмической и информационной поддержке интеллектуальной системы управления современными газоанализаторами, применяемыми в системах экологического мониторинга воздуха «электронный нос». Основу информационной поддержки составляют нейронные сети. Описана современная технология автоматического распознавания запахов на основе измерений с помощью низкоселективных датчиков в мультисенсорных системах обнаружения компонентов газовых смесей в атмосферном воздухе. Показано преимущество ее применения по сравнению с традиционными системами с высокоселективными чувствительными элементами. На основе записанных откликов с матрицы сенсоров создается библиотека образов запахов, которая хранится в БД интеллектуальной системы. Затем отклики анализируемого газа сравниваются с откликами отдельных веществ из библиотеки образов.

Для обработки информации предложен двухступенчатый метод кластеризации данных. Сначала данные наблюдений нормализуются, чтобы каждый входной параметр имел одинаковое влияние на систему. Затем они собираются в кластеры с использованием самоорганизующихся карт Кохонена и алгоритма k-средних. При этом каждый отдельный кластер представляет вещество со схожим запахом. Получение конкретных оценок было выполнено на основе экспериментальных данных, собранных в системе экологического мониторинга в районе мусоросжигательного завода в Кожухово. Рассмотрен выбор критерия идентификации запаха, которым в дальнейшем руководствуются эксперты при принятии решения. Обоснован выбор в качестве критерия значения метрики близости аналитических проб как нормы расстояния между векторами запаха в каждой пробе.

Разработаны алгоритм идентификации газовой аналитической пробы вещества с использованием нейронных сетей и выбранного критерия подготовки принятия решения, а также программный продукт на языке R, позволяющий оценить принадлежность полученных с прибора данных к определенному запаху с последующим получением визуальных результатов динамики распространения запахов в реальном времени. Приведены результаты применения разработанных алгоритмов в системе экомониторинга мусоросжигательного завода в районе Косино-Ухтомский Московской области.

Ключевые слова: интеллектуальная система, экологический мониторинг, мультисенсорная система газоанализатора, нейронная сеть, кластеризация данных, карты Кохонена, алгоритм *к-средних, метрика близости аналитических проб, программная среда R, система принятия решений.*

Одной из основных задач, направленных на улучшение качества атмосферного воздуха как на самих промышленных предприятиях, так и в окружающих их среде, является выявление источников загрязнения и проведение мероприятий по уменьшению концентраций загрязняющих веществ в атмосферном воздухе. Для ее решения в современных системах промышленного экологического мониторинга и в системах научных лабораторных исследований все более широкое распространение получают

технологии автоматического распознавания запахов с использованием многоканальных газоанализаторов на основе малогабаритных сенсоров и систем интеллектуального управления для принятия решений. Автоматический многоканальный анализатор предназначен для непрерывного контроля атмосферного воздуха в режиме реального времени [1].

Анализатор имеет встроенную библиотеку запахов отдельных объектов и способен измерять их интенсивность. Он может быть вклю-

чен в систему сети территориального контроля качества воздуха или использоваться как отдельный анализатор для мониторинга интенсивности запаха и идентификации его источника. Мультисенсорная система обнаружения компонентов газовых смесей в атмосферном воздухе названа «Электронный нос» [2]. Подобная технология автоматического распознавания запахов работает на основе измерений с низкоселективных датчиков в отличие от более традиционных систем с высокоселективными чувствительными элементами.

Запах образуется смесью пахучих веществ неизвестного состава, а не отдельным веществом, следовательно, требуется контролировать концентрацию выбросов газовых смесей, а не отдельного пахучего вещества. Ольфактометрические исследования позволяют получить полноценную количественную картину запаха в атмосферном воздухе или выбросах газовых смесей с последующим моделированием распространения запаха. Без проведения количественных химических анализов в стационарной химической лаборатории выделить конкретные соединения и источник запаха из смеси веществ практически невозможно [3].

Матрица сенсоров часто обладает перекрестной чувствительностью, в большинстве случаев их количество избыточно, поэтому необходимо уменьшать размерность пространства измерений за счет выделения наиболее информативных элементов. Подобные операции осуществляются с использованием хемометрического анализа для выделения существенной информации из набора данных. Например, метод принципиального компонентного анализа определяет направления максимальной дискриминации картин сенсорных откликов и наиболее часто используется в линейном приближении. Однако для решения классификационных задач данный метод недостаточно эффективен. Среди других методов следует отметить метод линейного дискриминационного анализа. Он позволяет находить направление, по которому достигаются наибольшие различия между образцами с различными запахами и при этом минимизируются различия между образцами с одинаковыми запахами. Однако в силу того, что при функционировании мультисенсорной системы газоанализатора вероятны достаточно большие отклонения от линейности, для получения корректной количественной информации требуются методы, способные обрабатывать данные без априорного знания функциональных зависимостей между входными сигналами и выходными параметрами, то есть методы нелинейные и непараметрические [4].

С принятием международных решений о постепенном переходе целых отраслей промышленности на зеленые технологии требуется более точный и широкий спектр идентифицируемых загрязнений. Кроме того, отслеживание изменений состава атмосферного воздуха на протяжении большого периода (от нескольких месяцев до года) вблизи источника выбросов требует периодической проверки корректности работы мультисенсорной системы. Необходимы периодическая перенастройка и проверка работоспособности с учетом активно меняющихся погодных условий. Для обработки данных требуются методы обработки без априорного знания функциональных зависимостей между входными сигналами и выходными параметрами [5]. Поэтому в целях корректного определения изменения запаха близ источников производства с высоким содержанием газовых выбросов предлагается использовать двухступенчатый метод кластеризации наблюдений, собранных газовым анализатором. Снятые с прибора данные моделируются с помощью самоорганизующихся карт Кохонена, построенных на основе цифровых сигналов от матрицы сенсоров газового анализатора, позволяя определить повторяющиеся паттерны, которые собираются в двумерную карту схожести среди нейронов в п-мерном пространстве, а затем разносятся в кластеры запахов с использованием алгоритма k-средних. Это дает возможность визуализировать и оценить характеристики образца запаха близ его источника.

Научная новизна настоящей исследовательской работы заключается в формировании методики применения системного анализа для разработки и реализации информационной поддержки интеллектуальной системы распознавания аналитических проб в многоканальном газоанализаторе. Методика включает иерархическую систему критериев и теоретико-множественные модели на стадии разработки, а также информационные и математические модели на основе нейронных сетей на стадии реализации с применением двухступенчатой кластеризации. Разработанный программный продукт визуализирует кластерные наблюдения в виде цветовых образов и сводит пространство собранных экспериментальных данных в компактные структуры нескольких кластеров (запахов) с отсчетом отклонений относительно среднего значения.

Основное преимущество интеллектуальных систем управления на основе нейронных сетей – принятие решений после соответствующего обучения системы. Для этого записываются отклики массива чувствительных сенсоров, через которые прокачивается газ, содержащий пары вещества. На основе записанных откликов с матрицы сенсоров создается библиотека образов запахов, которая хранится в БД интеллектуальной системы. Затем отклики анализируемого газа сравниваются с откликами отдельных веществ из библиотеки образов. Данный процесс называют распознаванием газовой пробы. В случае обнаружения схожих откликов или их комбинации появляется сообщение о нахождении в анализируемом газе паров исследуемого вещества или набора веществ.

Следует отметить, что интеллектуальная система газового анализатора не является измерителем концентрации. Определяется не количественная характеристика анализируемой пробы (концентрация), а качественная: проводится идентификация пробы, устанавливается класс образов, к которому она относится. По своей сути это специализированная информационная система, включающая подсистемы сбора, передачи, хранения, обработки и визуализации данных [6].

Целью настоящего исследования является повышение эффективности процессов распознавания образов запаха за счет информационной поддержки интеллектуальной системы принятия решений на основе нейронных сетей. В ходе ее достижения были решены следующие задачи.

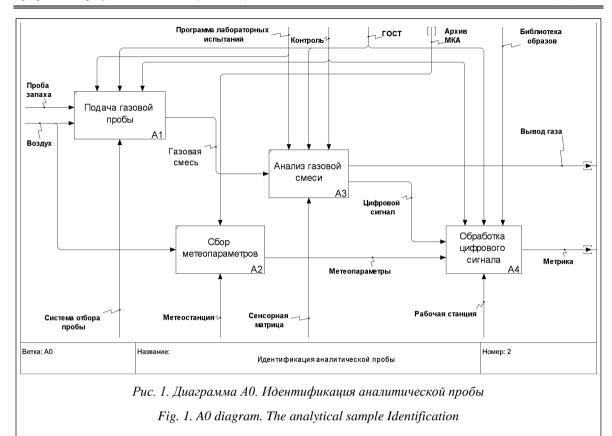
- 1. Разработка математического и функционального описания процесса идентификации газовой аналитической пробы вещества.
- 2. Разработка информационного обеспечения поддержки принятия решений исследователем при проведении анализа аналитической пробы вещества.
- 3. Разработка архитектуры интеллектуальной информационной системы обработки информации распознавания аналитических проб в многоканальном газоанализаторе, содержащей библиотеку аналитических проб, подсистему идентификации, подсистему информационной и алгоритмической поддержки и подсистему экспертного оценивания.
- 4. Разработка алгоритмического обеспечения идентификации аналитической пробы с применением средств искусственного интеллекта.

- 5. Разработка программного продукта для реализации выбранного алгоритмического решения
- 6. Апробация разработанных методов и алгоритмов идентификации в реальной интеллектуальной системе экологического мониторинга.

Для анализа информационных потоков в интеллектуальной системе были применены инструменты функционального моделирования в нотации IDEF0 [7].

Функционально-технологическая процесса идентификации аналитической пробы вещества в атмосферном воздухе отражает последовательный процесс поступления газовой смеси в газоанализатор, преобразование ее в цифровой сигнал и вывод значимого критерия системы в виде значения метрики, определяющей степень схожести образов запаха. На рисунке 1 представлена декомпозиция контекстной диаграммы в нотации IDEF0 описания процесса идентификации аналитической пробы многоканального газоанализатора с управлением на основе системы искусственного интеллекта. Диаграмма отражает процесс обработки цифрового сигнала, поступающего с сенсорной матрицы на рабочую станцию, в которую запрограммирован алгоритм распознавания аналитической пробы. Декомпозиция функциональных блоков позволяет исследовать информационные потоки внутри интеллектуальной системы управления многоканальным газоанализатором.

Газовая смесь образа запаха с окружающим воздухом поступает внутрь системы для передачи на сенсорную матрицу. Частицы газа оседают на сенсорах, что повышает их электрочувствительность и формирует выходной цифровой сигнал анализируемой пробы, или, как называют эту величину в теории распознавания образов, вектор признаков. Любой вектор в п-мерном пространстве можно охарактеризовать модулем вектора и его направлением. Для аналитических проб модуль вектора трактуется как интенсивность запаха или индекс загазованности. При разбавлении аналитической пробы чистым воздухом в 2 раза концентрация примесей уменьшается в 2 раза, и модуль вектора признаков на выходе прибора (индекс загазованности) падает тоже примерно в 2 раза. Различия направления двух векторов в теории распознавания образов оцениваются через расчет значения метрики. Она не зависит от модулей векторов, а характеризует именно разность их направлений. Меньшее значение метрики



означает высокую сонаправленность векторов, то есть высокую схожесть между двумя аналитическими пробами [8].

Непрерывное наблюдение за чистотой атмосферного воздуха в системе экологического мониторинга приводит к накоплению огромных объемов информации, хранящейся в БД. Для сжатия этой информации применяются различные методы, при этом одним из наиболее эффективных является метод нейронных сетей. В работе были использованы самоорганизующиеся карты Кохонена, которые преобразуют п-мерное пространство измерений в двухмерное координатное пространство обрабатывающих элементов или нейронов.

Карты Кохонена позволяют значительно сократить объем хранящейся в БД информации за счет ее разбиения на кластеры. При этом схожие наблюдения разбиваются на выборки, что облегчает обработку данных и процесс принятия решений, так как каждый кластер можно проанализировать отдельно. Для интеллектуальной системы газоанализатора это означает определение моментов времени, когда система фиксировала схожие образы запахов в атмосферном воздухе [9].

Самоорганизующиеся карты Кохонена представляют собой однослойные сети, где от-

дельные нейроны соединены с каждым элементом n-мерного входного значения вектора. Вектор весов описывается формулой $w = [w_1, w_2, ..., w_n]$.

Вектор весов рассчитывается для каждого нейрона. Исходя из расчета весов выбирается нейрон-победитель — такой нейрон, расстояние которого от входного элемента наименьшее и для которого выполняется следующее отношение: $d(x, w) = \min d(x, w_i)$; 1 < i < n, где d(x, w) — расстояние между входным и весовым векторами; n — количество нейронов.

Особенность расчета весовых векторов заключается в том, что соседние нейроны адаптируют значения своих весов в соответствии с выбранным нейроном-победителем, следуя направлению его вектора по следующему правилу: $w(k+1) = w_i(k) + \mu_i(k)[x-w_i(k)]$, где $\mu_i(k)$ – коэффициент обучения соседнего нейрона в области S(k).

Разработанная алгоритмическая и информационная поддержка принятия активно применяется в экологическом мониторинге района на границе мусоросжигательного завода в Кожухово. Наблюдение за атмосферным воздухом осуществлялось с 01.12.2019 по 15.01.2020. Экспериментальные данные описывают изменение концентрации летучих примесей в атмо-

сферном воздухе за указанный период. Для проведения нейронного анализа были использованы данные с матрицы сенсорных датчиков, разделенные на три подвыборки.

- 1. Training Dataset. Представляет 70 % от общих наблюдений. На основе этих данных создается сеть Кохонена.
- 2. Testing Dataset 15 % от общего количества наблюдений. С помощью этой подвыборки проводится уточнение параметров сгенерированной сети.
- 3. Validation Dataset 15 % от общих наблюдений. С помощью данной подвыборки проводится дополнительное тестирование сети, чтобы подтвердить ее приемлемость.

Сначала данные наблюдений нормализуются, чтобы каждый входной параметр имел одинаковое влияние на систему. Затем они собираются в кластеры с использованием самоорганизующихся карт Кохонена и алгоритма k-средних. Однослойные карты позволяют визуализировать данные в группы наблюдений и определять их в фактические кластеры по характеру запаха. Для задачи наблюдения за атмосферным воздухом в Кожухово выбрана размерность карты в 20×20 единиц согласно эвристическому правилу, что ширина и высота карты выходят из формулы $m = 5\sqrt{n}$ [10]. Распределение наблюдений по кластерам карт Кохонена представлено в таблице 1.

Созданная сеть распределила наблюдения по топологической сетке размером 20×20, где каждый отдельный кластер представляет вещество со схожим запахом, зафиксированное в разный момент времени. Сеть отличается малой ошибкой выборок, что говорит о достаточной корреляции собранных наблюдений.

Недостаток сетей Кохонена заключается в том, что нейрон, изначально обученный на принадлежность к одному классу, во время обучения сети может изменить свой вектор направления и начать ассоциироваться с совершенно другим классом. По этой причине сеть Кохонена может стать нестабильной в период полного обучения. Во избежание перезаписи

Частота наблюдений каждого кластера на карте Кохонена

Таблица 1 Table 1

Monitoring frequency of each Kohonen map cluster

n/n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	66	190	106	287	54	12	62	100	24	26	26	22	29	55	94	45	11	40	39	18
2	33	35	146	111	89	16	26	45	13	17	14	10	28	15	5	29	10	42	34	35
3	22	343	238	92	18	76	61	99	31	40	17	39	20	6	34	16	24	34	21	43
4	35	223	90	66	72	56	250	42	75	50	26	23	28	42	37	25	81	62	76	71
5	242	244	157	99	64	48	134	34	20	26	27	53	55	30	45	53	74	61	90	20
6	52	67	66	40	53	1	19	29	5	12	41	43	44	43	12	21	12	21	13	30
7	183	69	30	68	40	41	54	41	39	51	43	46	38	42	7	16	15	13	17	32
8	30	47	76	23	36	17	7	55	21	16	46	34	46	55	47	5	9	16	12	58
9	55	134	87	100	75	123	32	42	35	47	21	81	99	46	69	43	14	6	9	113
10	32	96	141	119	84	88	43	44	52	30	26	90	51	52	35	35	43	50	36	9
11	76	97	40	32	54	56	75	46	56	40	45	59	49	53	25	75	78	63	32	31
12	52	51	118	53	49	101	96	68	75	68	75	95	59	38	22	56	22	11	18	84
13	44	70	68	74	37	89	41	70	77	83	55	49	59	30	34	42	48	26	61	46
14	62	45	52	79	70	62	26	66	44	82	34	56	46	63	43	25	11	40	66	86
15	44	81	38	68	58	67	69	101	71	71	49	47	57	37	45	10	32	33	72	80
16	29	46	78	76	102	66	56	80	97	41	35	60	58	53	33	32	26	81	67	46
17	54	57	76	71	55	89	90	63	83	95	74	57	54	70	42	49	28	78	58	108
18	62	23	82	53	57	62	83	119	107	98	40	49	43	45	67	42	63	0	39	13
19	75	78	82	33	70	45	132	69	159	120	117	74	123	77	69	119	44	12	22	105
20	38	48	59	51	51	86	51	87	35	78	140	82	79	30	47	36	58	13	46	37

нейрона добавлен второй слой кластеров, основанный на алгоритме k-средних, благодаря чему схожие кластеры собираются в один общий кластер образа запаха, позволяя игнорировать возможные отклонения в процессе обучения карт Кохонена (рис. 2).

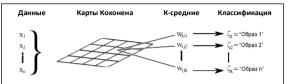


Рис. 2. Описание алгоритма кластеризации

Fig. 2. A clustering algorithm description

Второй слой сети строится на основе весов нейронов каждого из кластеров первого слоя, не добавляя большого количества лишних вычислений. Поскольку при методе k-средних количество кластеров необходимо указывать вручную, разделим кластеры предыдущего слоя на 5 больших кластеров, то есть k=5. Расчет весов кластеров второго слоя происходит следующим образом: $w_i(t) = w_i(t-1) + l(x_i - w_i(t-1))$, где x – вес нейрона первого слоя.

Проведя серию экспериментов с кластеризацией с изменением k от 2 до 7 и изучив нормативные документы района исследования, было принято решение разбить топологическую карту на 5 кластеров. В случае k > 5 начинали выделяться достаточно малые группы наблюдений, которые при отсутствии соответствующей временным меткам нормативной информации считались незначительными и нивелировались. В таблице 2 собрана результирующая статистика обработки данных газоанализатора для того, чтобы попытаться сделать промежуточный вывод о характере наблюдаемых запахов. Далее приведены усредненные показатели сенсорных датчиков каждого кластера.

Разбиение собранных наблюдений на кластеры позволяет оценить частоту появления определенного образа запаха в атмосферном воздухе. Пока основные высокочастотные кластеры описывают характерный для точки взятия наблюдений запах, малые кластеры описывают редкие или аномальные случаи появления отдельных летучих примесей в воздухе. Данные кластеров необходимо дополнительно исследовать для принятия конечного решения о характере запаха. Измерение атмосферного воздуха на предмет присутствия летучих примесей за отведенный период времени, жалобы граждан и волонтеров, метеопараметры и т.д. могут служить сопроводительной информацией для принятия решения о соответствии запаха отдельного кластера измерений. Затем результат может быть добавлен в библиотеку образов для сравнения с новыми наблюдениями, используя расчет значения метрики через евклидово расстояние между двумя векторами признаков.

Следующей важнейшей задачей интеллектуальной системы является выбор критерия идентификации запаха, которым в дальнейшем руководствуются эксперты для принятия решения об идентификации запахов. В общем виде задача математически решается с использованием критериального подхода, основанного на формировании множества локальных критериев управленческих задач для разработки механизма корректной оценки построения информационной поддержки интеллектуальной системы управления и подготовки, на основе анализа параметров системы $\Pi = \{n_i\}$, где i = 1, ..., p; p – количество параметров. Отдельный вариант управленческого решения Х состоит из набора критериев К, характеризующих качество выбранного решения: $K = \{k_i\}$, где i = 1, \dots , n; n – количество критериев. Интегральный

Таблица 2 Средние показатели наблюдений всех датчиков для каждого кластера

Table 2

Шаатата	Среднее значение показателя										
Частота	TGS2600	TGS2602	TGS2603	NO	CO	SO ₂	NO_2	IPZ			
2920	0,8314	0,6807	0,7867	0,0475	0,5299	0,0432	0,0357	0,0506			
6998	0,7963	0,7820	0,7885	0,0443	0,4680	0,0429	0,0165	0,0512			
17778	0,8310	0,8147	0,8014	0,0447	0,4973	0,0797	0,0362	0,0520			
17656	0,8109	0,7504	0,7960	0,0557	0,3443	0,0391	0,0165	0,0502			
30770	0,7821	0,6851	0,7624	0,0558	0,3577	0,0213	0,0167	0,0471			

Mean observation values of all sensors for each cluster

критерий K_g определяет глобальную цель проводимых исследований и характеризует направление изменения характеристик для принятия управленческого решения.

Матрица $K(\mathbf{m})$ экспертных оценок c_{ij} определяет влияние исполнения операции на формирование критериев управленческих оценок. Значение критерия под воздействием руководящего фактора в форме выполненных операций определяется алгебраической суммой: $k_i^t = k_i^0 + \Delta k_i$, где k_i^0 — текущее значение критерия; Δk_i — приращение критерия в результате изменения интенсивности выполнения операций; значение критерия формулы описывается

уравнением
$$k_i^t = k_i^0 \left(1 + \sum_{j=1}^q c_{ij} \frac{\Delta \mathbf{M}_j}{\mathbf{M}_j} \right)$$
.

Критерии развития определяют целевые функции. Решая непосредственно многокритериальные задачи, следует

- учитывать влияние отдельного критерия при решении контекстной задачи, поскольку каждый критерий может обладать индивидуальной значимостью или приоритетом;
- стандартизировать критерии; данные наблюдений могут приводиться в различных единицах измерений, поэтому необходимо провести их нормализацию;
- подобрать метод оптимизации ресурсов, выделенных на решение поставленной задачи.

Профессиональным экспертом или группой экспертов оцениваются значения критериев. Затем, руководствуясь полученными оценками, определяется глобальный критерий — многокритериальная экспертная оценка. Глобальный критерий G формируется из множества частных критериев K_j^m :

$$G = \left\{S_i^m \mid i = \overline{1,M}\right\} \xrightarrow{f^p} K^M = \left\{K_j^m, D_j^m \mid j = \overline{1,N}\right\},$$
 где S_i^m — элемент формальной модели; f^p — функция отображения; N — число частных критериев; D_j^m — множество атрибутов частного критерия.

Локальные критерии содержат экономические, экологические и технологические критерии. Разработка интеллектуальных систем управления с применением экспертных систем необходима для реализации управления с учетом этих критериев [11].

В качестве технологических критериев при поэтапном контроле процесса идентификации аналитической пробы исследуемого вещества в интеллектуальной системе многоканального газоанализатора применяются критерии (оцен-

ки) технологического контроля ($k_{\scriptscriptstyle T}$). Технологический контроль за процессом идентификации запаха проводится на всех стадиях протекания процесса: на стадии подачи газовой смеси воздуха и аналитической пробы создается контролируемое окружение; при анализе газовой смеси на сенсорной матрице система должна быть откалибрована должным образом, чтобы датчики фиксировали оседание частиц корректно; при обработке получившихся цифровых сигналов данные служат входом для обучения нейросети кластеризации наблюдений и расчета близости вектора признаков образа запаха – получения значения метрики. Конечный локальный критерий - значение метрики близости аналитических проб. Это один из важнейших критериев процесса управления интеллектуальной системой многоканального газоанализатора. На его основе изучается близость исследуемых аналитических проб и принимается решение о необходимости внесения изменений в технологический процесс идентификации пробы [12].

Образ аналитической пробы представляет собой представленную в виде массива данных совокупность сигналов массива датчиков с одинаковой временной меткой, которой с математической точки зрения соответствует некоторый вектор в многомерном пространстве. Координаты этого вектора соответствуют нормированным сигналам отдельных датчиков, а размерность многомерного пространства – количеству каналов измерений.

Длина (модуль) вектора соответствует интенсивности запаха, а его направление индивидуально для каждого отдельного источника.

Алгоритм идентификации аналитической пробы сводится к вычислению метрики между цифровым кодом пробы и цифровыми кодами всех стандартных образцов из БД. Наименьшее значение метрики указывает на идентифицированный образец из БД, а количественное значение метрики служит оценкой достоверности этой идентификации: чем меньше значение метрики, тем выше достоверность идентификации. При лабораторных испытаниях с применением стандартных образцов, сравнивая значение модуля вектора анализируемой пробы и концентрацию идентифицированного образца, записанного в БД, можно оценить содержание компонента в пробе. Введем обозначения: і – номер образца; j – номер пробы; k – номер канала сенсора; V_i – вектор признаков образцов запахов $V_i(X_1, X_2, ..., X_k)$; V_i – вектор признаков

пробы воздуха; $M(V_i, V_j) = M_{ij}$ – метрика расстояния между i-м образцом и j-й пробой.

Значение метрики оценивается через евклидово расстояние между нормированными векторами, например, для двумерного случая (табл. 3).

Координаты нормированного вектора $v_i(x_i, y_i)$ одинаковы для всех векторов данного направления независимо от его длины, а норма такого вектора $v_i(x_i, y_i) = 1$. Метрика не зависит от модулей векторов, а характеризует именно разность их направлений. Чем меньше значение параметра метрики, тем ближе направления векторов, то есть тем ближе запахи исследуемой пробы и образца. Рассмотрим расчет метрик на примере двух векторов признаков образцов $V_1(180,40)$ и $V_2(20,140)$ и некоторой пробы $S_1(50,25)$ (табл. 4).

Визуализация исходных и нормированных векторов признаков образцов и пробы приведена на рисунке 3, результаты расчетов метрик — в таблице 5.

Для наглядности использован 20-кратный масштаб для нормированных векторов.

Результаты расчета метрик подтверждают вывод, очевидный из графической визуализации исходных данных примера: вектор признаков пробы S_1 намного ближе к вектору признаков образца V_1 ($M_{11}=0.2444$), чем к вектору признаков образца V_2 ($M_{12}=0.9282$). Графическая визуализация применима при размерности

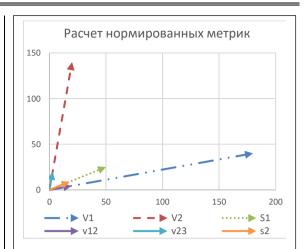


Рис. 3. Исходные и нормированные векторы признаков образцов V_1 и V_2 и пробы S_1

Fig. 3. Initial and normalized feature vectors of V_1 and V_2 samples and S_1 sample

задачи не более 3, то есть для двух- и трехканальных сенсоров. В случае использования восьмиканальных сенсоров графический метод неприменим и расчет метрики является единственным способом оценки близости векторов признаков.

Для выбора алгоритма расчета метрики было проведено сопоставление нескольких вариантов расчета с целью определения наиболее подходящего варианта. После анализа полученных результатов окончательно была вы-

Таблииа 3

Нормировка координат вектора

Table 3

Normalization of vector coordinates

Образец	Dawron	Коорди	наты	Норма		
Ооразец	Вектор	X_i	Y_i	$ V_i $		
Исходный образец 1	V_1	180	40	184,4		
Нормированный образец 1	v_1	0,9762	0,2169	1,0		
Половинный образец 1б	$0,5*V_1$	90	20	92,2		
Нормированный образец 16	v_1	0,9762	0,2169	1,0		

Таблица 4

Исходные и нормированные данные

Table 4

Source data

		Исх	одные дан	іные	Нормированные данные			
Образец	Вектор	Коорди	наты	Норма	Коорд	инаты	Норма	
		X_i	Y_i	$ V_i $	X_i	Y_i	$ V_i $	
Образец 1	V_1	180	40	184,4	0,9762	0,2169	1,0	
Образец 2	V_2	20	140	141,4	0,1414	0,9899	1,0	
Проба 1	S_1	50	25	55,9	0,8944	0,4472	1,0	

Таблица 5

Результаты расчета метрик сравнения векторов признаков

Table 5

The calculation	n data	of feature	vector	comparing	metrics
-----------------	--------	------------	--------	-----------	---------

Вектор 1	Вектор 2	<i>x</i> ₁	y 1	<i>X</i> 2	y 2	Метрика
v_1	v_2	0,9762	0,2169	0,1414	0,9899	1,1
v_1	s_1	0,9762	0,2169	0,8944	0,4472	0,2444
v_2	<i>S</i> ₁	0,1414	0,9899	0,8944	0,4472	0,9282

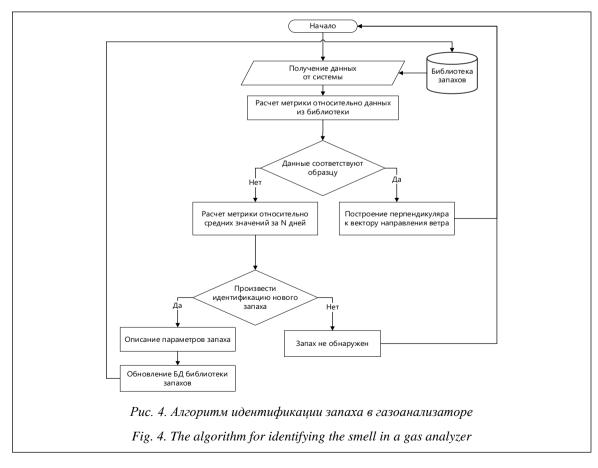
брана метрика:
$$M = \arccos\left(\frac{\vec{s}\vec{v}}{\left|\vec{S}\right|\left|\vec{V}\right|}\right)\frac{200}{\pi}$$
, где \vec{V} —

вектор текущего измерения; \vec{S} — вектор образа запаха в БД.

В расчете значения метрики M соответствия вектора измерений и образца используется отношение угла между векторами измерения и образца к максимально возможному значению угла в пространстве положительных значений, которое рассчитывается как арккосинус отношения скалярного произведения векторов к произведению их модулей (взято в единицах процентов).

Совокупность сигналов восьми датчиков с одинаковой временной меткой, представленная в виде массива данных, которой с матема-

тической точки зрения соответствует некоторый вектор в многомерном пространстве, соответствует запаху пробы. Алгоритм вычисления значения метрики не является универсальным, различные подходы при распознавании образов отличаются именно алгоритмами вычисления метрики. Сущность алгоритма заключается в расчете метрики между вектором признаков пробы (а) и вектором всех стандартных образцов, находящихся в БД (b). Вектор признаков источников запаха описывает образ состава анализируемой пробы. БД рабочей станции содержит векторы признаков шести подготовленных образцов с концентрациями на уровне 0,5-1 ррт. Поток данных поступает на рабочую станцию, где сравниваются сигналы от сенсорных модулей с БД образцов (рис. 4).



Разработка алгоритмов велась применительно к аппаратной и интеллектуальной симногоканального газоанализатора фирмы «ДИЭМ» (рис. 5). При этом на вход подсистемы анализа цифровых сигналов поступали нормированные данные восьми датчиков: PID, TGS2600, TGS2602, TGS2603, NO, CO, SO2, NO2. Анализатор предназначен для выполнения непрерывных автоматических измерений в приземном слое атмосферного воздуха массовых (или объемных) концентраций загрязняющих веществ, а также метеорологических параметров, уровня шума, гамма-фона. Приоритетными местами его размещения являются границы санитарно-защитной зоны предприятий, вблизи границ жилой застройки, населенные пункты (для целей государственного мониторинга).

Для программной реализации разработанного алгоритма идентификации наблюдений была использована программная среда R, поскольку она включает в себя широкий набор готовых библиотек для статистического анализа данных [13].

Разработанный программный продукт визуализирует работу сенсоров в реальном времени в виде цветовых образов (рис. 6), которые позволяют продемонстрировать динамику попадания частиц вещества на отдельном сенсорном датчике. Совмещение карт образов всех входных элементов позволяет оценить корреляцию между ними, например, то, что запах в отдельный период времени формирует четкий отклик на одних сенсорах, в то время как другие показывают малое значение отклика. Сопоставив техническую информацию с картой жалоб жителей, можно составить картину запаха для отдельного промежутка времени.

Программа сводит набор наблюдений в компактную структуру данных нескольких кластеров, где отклонения в реальном времени сравниваются со средним значением отдельного кластера (запаха). Однослойные карты позволяют визуализировать данные в группы наблюдений и определить их в фактические кластеры по характеру запаха. Размерность топологической карты, как и остальные параметры, передаются в программу. Для задачи наблюдения за атмосферным воздухом в Кожухово, как уже упоминалось, выбрана размерность карты в 20×20 единиц. Полученная карта разграничена на 5 кластеров запахов (рис. 7).

Формируя библиотеку запахов, необходимо охарактеризовать кластеры групп наблюдений

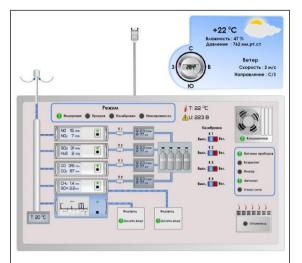


Рис. 5. Газоанализатор «НПФ ДИЭМ»

Fig. 5. The gas analyzer by NPF DIEM

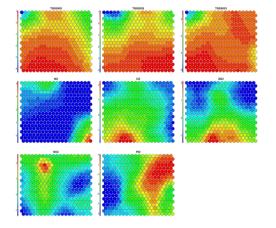


Рис. 6. Цветовые образы сенсорных датчиков

Fig. 6. Color images of sensors

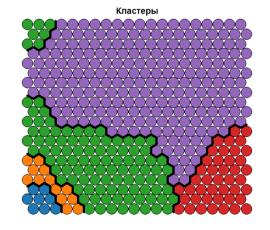


Рис. 7. Кластеры запахов

Fig. 7. Odor clusters

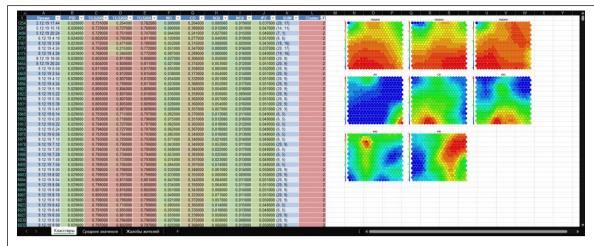


Рис. 8. Распределение собранных наблюдений по кластерам

Fig. 8. Distribution of collected observations into clusters

(рис. 8). Помимо применения технологии искусственного интеллекта, для разбиения данных на фактические кластеры запахов необходимо ориентироваться на нормативные документы: жалобы жителей ближайших районов, их опросы, сведения волонтеров о времени и месте появления запаха, метеопараметры с информацией о направлении ветра и т.д. Например, наиболее вероятно, что превалирующий по числу наблюдений четвертый кластер ассоциируется с безвредным воздухом, поскольку время наблюдений распределено равномерно и с этими показателями не ассоциируются нормативные сведения. Малые кластеры (3 и 5) соотносятся с аномальными запахами, не присущими точке наблюдения: проехала машина, подул ветер со стороны, противоположной заводу, и т.д. Наблюдения в первом и втором кластерах часто фиксировались с разной временной периодичностью. Дополнительно с указанными временными метками совпадают жалобы жителей на присутствие гари и сероводорода в воздухе и т.д. Наиболее вероятно, что они ассоциируются с запахом с мусоросжигательного завода [14].

При принятии решений по анализу атмосферного воздуха эксперт-исследователь использует таблицу расчета двух типов метрики: между соседними записями в БД и между средними показаниями кластеров и их статистическими оценками — минимальным, средним и максимальным. Он отслеживает динамику изменения состава воздуха и способен определить, когда содержание газовых примесей в воздухе отклоняется от средних значений объекта библиотеки. На основе количественных

данных, которые фиксируются в реальном времени, и алгоритма идентификации пробы, заложенного в рабочей станции, эксперт способен эффективно принимать решения об изменении состава веществ в атмосферном воздухе в реальном времени.

Заключение

Выполненный комплекс исследований ориентирован на информационное и алгоритмическое обеспечение процесса идентификации аналитической пробы исследуемого вещества с использованием инструментов системного анализа и технологий искусственного интеллекта.

Разработано математическое и функциональное описание процесса идентификации аналитической пробы вещества.

Сформировано множество критериев оценки качества идентификации аналитической пробы исследуемого вещества с выводом глобального эколого-технологического критерия на основе критериального подхода.

Разработана информационная поддержка этапа идентификации аналитической пробы образа вещества, начиная от закачки газовой пробы до расчета технологического критерия, определяемого в виде значения некоторой метрики, с применением средств системного анализа создания функциональных моделей процессов в нотации IDEF0.

Разработан программный продукт анализа экспериментальных данных, сводящий снимаемые с прибора наблюдения в компактные кластеры для оценки принадлежности их к определенному запаху с последующей визуализацией.

Выведен способ расчета значения метрики из множества вариантов с проведением эмпирического эксперимента. Проведен анализ подсистем поддержки интеллектуальной информационной системы. Разработана реализация самоорганизующейся нейронной сети Кохонена с методом расчета k-средних для кластеризации собранных наблюдений многоканаль-

ным газоанализатором. На основе полученных результатов могут приниматься решения о природе запаха в отдельном кластере.

Разработано алгоритмическое обеспечение идентификации запаха на основе критерия, представляющего собой метрику для оценки разности цифровых кодов пробы и цифровых кодов стандартных образцов из БД.

Литература

- 1. Городнова Н.В. Применение искусственного интеллекта в проектах «Smart-экология» // Дискуссия. 2021. Т. 106. № 2. С. 34–48.
- 2. Ганшин В.М., Фесенко А.В., Чебышев А.В. От обонятельных моделей к «электронному носу». Новые возможности параллельной аналитики. URL: http://www.vrsystems.ru/stati/ot_obonyatelnix_modelei k elektronnomu nosu.htm (дата обращения: 20.06.2022).
- 3. Ярыгин Г.А., Баюкин М.В., Корнюшко В.Ф., Садеков Л.В. «Электронный нос» как элемент искусственного интеллекта в задачах экологического мониторинга // ИТТ. 2021. № 50. С. 22–29.
- 4. Marco S. The need for external validation in machine olfaction: Emphasis on health-related applications. Analytical and Bioanalytical Chemistry, 2014, vol. 406, no. 16, pp. 3941–3956. DOI: 10.1007/s00216-014-7807-7.
- 5. Capelli L., Sironi S., Del Rosso R. Review: Electronic noses for environmental monitoring applications. Sensors, 2014, vol. 14, no. 11, pp. 19979–20007. DOI: 10.3390/s141119979.
- 6. Рыбина Г.В. Основы построения интеллектуальных систем. М.: Финансы и статистика, 2014. 432 с.
- 7. Корнюшко В.Ф., Николаева О.М., Золотов С.Ю., Крюкова О.В., Садеков Л.В. Интеллектуальные информационные технологии в экспериментальных исследованиях разработки готовых лекарственных форм (ГЛФ) при производстве лекарственных средств // Наука России: Цели и задачи: сб. тр. 2021. С. 96–104. DOI: 10.18411/sr-10-02-2021-22.
- 8. Ярыгин Г.А., Баюкин М.В., Корнюшко В.Ф., Шмакова Е.Г., Садеков Л.В. Информационная поддержка системы распознавания компонентов газовых смесей «электронный нос» с искусственным интеллектом в задачах экологического мониторинга // Вестн. технологического университета. 2022. Т. 25. № 1. С. 81–87.
- 9. Licen S., Barbieri G., Fabbris A., Brigglio S.C., Pillon A., Stel F., Barbieri P. Odor control map: Self-Organizing map built from electronic nose signals and integrated by different instrumental and sensorial data to obtain an assessment tool for real environmental scenarios. Sensors and Actuators B: Chemical, 2018, vol. 263, pp. 476–485. DOI: 10.1016/j.snb.2018.02.144.
- 10. Вагин В.Н., Ганишев В.А. Кластеризация пользователей по голосу с помощью улучшенных самоорганизующихся растущих нейронных сетей // Программные продукты и системы. 2015. № 3. С. 136–142. DOI: 10.15827/0236-235X.111.136-142.
- 11. Мелентьев В.А. Использование метода Мелентьева представления графов для выявления клик и анализа топологий вычислительных систем // ISJ Theoretical & Applied Science. 2018. Т. 68. № 12. С. 201–211. DOI: 10.15863/TAS.2018.12.68.28.
 - 12. Мешалкин В.П. Экспертные системы в химической технологии. М.: Химия, 1995. 367 с.
- 13. Богданов П.Ю., Краева Е.В., Веревкин С.А., Пойманова Е.Д., Татарникова Т.М. Программные среды для изучения основ нейронных сетей // Программные продукты и системы. 2021. Т. 34. № 1. С. 145–150. DOI: 10.15827/0236-235X.133.145-150.
- 14. Варшавский П.Р., Кожевников А.В. Реализация программных средств для классификации данных на основе аппарата сверточных нейронных сетей и прецедентного подхода // Программные продукты и системы. 2020. Т. 33. № 4. С. 591–598. DOI: 10.15827/0236-235X.132.591-598.

Software & Systems DOI: 10.15827/0236-235X.140.715-728 Received 30.06.22, Revised 04.07.22 2022, vol. 35, no. 4, pp. 715–728

Informational and algorithmic support of an environmental air monitoring intelligent system based on neural networks

G.A. Yarygin ¹, Dr.Sc. (Engineering), Professor, Scientific Supervisor, office@diem.ru

M.V. Bayukin ¹, Ph.D. (Engineering), Deputy Director, bayukin@diem.ru

V.F. Kornyushko², Dr.Sc. (Engineering), Professor, vfk256@mail.ru

E.G. Shmakova², Ph.D. (Engineering), Associate Professor, 89254225204@mail.ru

L.V. Sadekov², Postgraduate Student, leosadekov@yandex.ru

Abstract. The article discusses algorithmic and informational support of an intelligent control system for modern gas analyzers used in environmental air monitoring systems called the Electronic nose. Neural networks form the base of information support. The paper describes a modern automatic odor recognition system based on measurements using low-selective sensors in multi-sensor systems for detecting components of gas mixtures in ambient air. It also shows the advantage of the proposed system compared with traditional systems with highly selective sensing elements. There is a library of smell images based on a series of prerecorded responses from the sensor matrix. It is stored in the intelligent system database. Then the responses of an analyzed gas are compared with the responses of individual substances from the image library.

The authors propose a two-stage data clustering method for information processing. First, observational data is normalized so that each input parameter equally affects the system. Then the data are assembled into clusters using self-organizing Kohonen maps and the k-means algorithm. Each cluster represents an odor with a similar smell. Specific assessments are based on experimental data collected in the environmental monitoring system in the area of the waste incineration plant in Kozhukhovo. The paper considers the choice of an odor identification criteria, which will be used by experts in deciding on odor identification. There is a substantiation of choosing the proximity metric of analytical samples as the norm of the distance between the odor vectors in each sample as a criterion.

The authors have developed an algorithm for identifying a substance's gas analytical sample using neural networks and the selected criterion for decision-making support. There is also a developed (using R programming language) software product that allows assessing data membership obtained from a device to a certain smell followed by providing visual results of a odors' spread dynamics in real-time. The paper presents the application results of the developed algorithm in the eco-monitoring system of the incinerator plant in the Kosino-Ukhtomsky district of the Moscow region.

Keywords: intelligent systems, environmental air monitoring, gas analyzer multisensory systems, neural networks, data clustering, Kohonen maps, k-means algorithm, proximity metric of analytical samples, R programming environment, decision-making system.

References

- 1. Gordonova N.V. Applying of artificial intelligence in "SMART-ecology" projects. *Discussion*, 2021, vol. 106, no. 2, pp. 34–48 (in Russ.).
- 2. Ganshin V.M., Fesenko A.V., Chebyshev A.V. From Olfactory Models to the "Electronic Nose". New Features of Parallel Analytics. Available at: http://www.vrsystems.ru/stati/ot_obonyatelnix_modelei_k_elektronnomu_nosu.htm (accessed June 20, 2022) (in Russ.).
- 3. Yarygin G.A., Bayukin M.V., Kornyushko V.F., Sadekov L.V. "Electronic nose" as an element of artificial intelligence in tasks of environmental monitoring. *Information and Telecommunication Technologies*, 2021, no. 50, pp. 11–25 (in Russ.).
- 4. Marco S. The need for external validation in machine olfaction: Emphasis on health-related applications. *Analytical and Bioanalytical Chemistry*, 2014, vol. 406, no. 16, pp. 3941–3956. DOI: 10.1007/s00216-014-7807-7.
- 5. Capelli L., Sironi S., Del Rosso R. Review: Electronic noses for environmental monitoring applications. *Sensors*, 2014, vol. 14, no. 11, pp. 19979–20007. DOI: 10.3390/s141119979.
 - 6. Rybina G.V. Fundamentals of Building Intelligent Systems. Moscow, 2014, 432 p. (in Russ.).

¹ SPF DIEM, Moscow, 117485, Russian Federation

² MIREA – Russian Technological University, Moscow, 119454, Russian Federation

- 7. Kornyushko V.F., Nikolaeva O.M., Zolotov S.Yu., Kryukova O.V., Sadekov L.V. Intelligent information technology in experimental research of development of finished dosage forms (GLF) in the production medicines. *Collection of Scientific Papers*, 2021, pp. 96–104. DOI: 10.18411/sr-10-02-2021-22 (in Russ.).
- 8. Yarygin G.A., Bayukin M.V., Kornyushko V.F., Shmakova E.G., Sadekov L.V. Information support of the "electronic nose" gas mixture components recognition system with artificial intelligence in environmental monitoring tasks. *Herald of the Technological University*, 2022, vol. 25, no. 1, pp. 81–87 (in Russ.).
- 9. Licen S., Barbieri G., Fabbris A., Brigglio S.C., Pillon A., Stel F., Barbieri P. Odor control map: Self-Organizing map built from electronic nose signals and integrated by different instrumental and sensorial data to obtain an assessment tool for real environmental scenarios. *Sensors and Actuators B: Chemical*, 2018, vol. 263, pp. 476–485. DOI: 10.1016/j.snb.2018.02.144.
- 10. Vagin V.N., Ganishev V.A. Speaker clustering using enhanced self-organizing incremental neural networks. *Software & Systems*, 2015, no. 3, pp. 136–142. DOI: 10.15827/0236-235X.111.136-142 (in Russ.).
- 11. Melent'ev V.A. Use of Melentiev's graph representation method for detection of cliques and the analysis of topologies of computing systems. *ISJ Theoretical & Applied Science*, 2018, vol. 68, no. 12, pp. 201–211. DOI: 10.15863/TAS.2018.12.68.28 (in Russ.).
 - 12. Meshalkin V.P. Expert Systems in Chemical Technology. Moscow, 1995, 367 p. (in Russ.).
- 13. Bogdanov P.Yu., Kraeva E.V., Verevkin S.A., Poymanova E.D., Tatarnikova T.M. Software environments for studying the basics of neural networks. *Software & Systems*, 2021, vol. 34, no. 1, pp. 145–150. DOI: 10.15827/0236-235X.133.145-150 (in Russ.).
- 14. Varshavskiy P.R., Kozhevnikov A.V. Implementation of data classification software based on convolutional neural networks and case-based reasoning approach. *Software & Systems*, 2020, vol. 33, no. 4, pp. 591–598. DOI: 10.15827/0236-235X.132.591-598 (in Russ.).

Для цитирования

Ярыгин Г.А., Баюкин М.В., Корнюшко В.Ф., Шмакова Е.Г., Садеков Л.В. Информационная и алгоритмическая поддержка интеллектуальной системы экологического мониторинга воздуха на основе нейронных сетей // Программные продукты и системы. 2022. Т. 35. № 4. С. 715–728. DOI: 10.15827/0236-235X.140.715-728.

For citation

Yarygin G.A., Bayukin M.V., Kornyushko V.F., Shmakova E.G., Sadekov L.V. Informational and algorithmic support of an environmental air monitoring intelligent system based on neural networks. *Software & Systems*, 2022, vol. 35, no. 4, pp. 715–728 (in Russ.). DOI: 10.15827/0236-235X. 140.715-728.