

## Сравнение алгоритмов G-Means и сети Кохонена в решении задач кластеризации

О.В. Судаков<sup>1</sup>, Д.В. Дмитриев<sup>1</sup>

<sup>1</sup> *Нижегородский государственный технический университет им. Р.Е. Алексева, ул. Минина, 24, корпус 1, Нижний Новгород, 603950, Россия*

### Аннотация

В данной работе рассмотрен вопрос о том, как улучшить самоорганизующуюся нейронную сеть, состоящую из связки алгоритма кластеризации и многослойного перцептрона для задач верификации данных в условиях отсутствия обучающих пар. Самым популярным алгоритмом кластеризации для данной задачи является сеть Кохонена, однако на сегодняшний день это не единственный алгоритм способный быстро и точно выполнять поставленную задачу. В работе проводится сравнение сети Кохонена и алгоритма G-Means. Кратко разобран принцип работы этих двух алгоритмов. Проводится сравнение точности данных алгоритмов и скорости их обучения. Актуальность данной работы заключается в том, что подготовка обучающих пар для интеллектуальных систем и сам процесс обучения с учителем является ресурсоемкой задачей. Рассматриваемые системы самообучающихся алгоритмов позволят значительно увеличить скорость обучения, а также избавит от необходимости ручной классификации данных и создания обучающих пар для перцептрона что в свою очередь позволит создать самообучающуюся систему с возможностью обобщения и прогнозирования.

### Ключевые слова

Нейронные сети, обучение, самообучающиеся алгоритмы, обучение без учителя, G-Means, сеть Кохонена, гибридные системы.

## Comparison of G-Means Algorithms and Kohonen Network in Solving Clustering Problems

O.V. Sudakov<sup>1</sup>, D.V. Dmitriev<sup>1</sup>

<sup>1</sup> *Nizhny Novgorod State Technical University named after R.E. Alekseev, 24 building 1, Minina str., Nizhny Novgorod, 603950, Russia*

### Abstract

**Purpose:** In this paper, the question of how to improve a self-organizing neural network consisting of a bundle of clustering algorithm and a multilayer perceptron for data verification tasks in the absence of training pairs is considered.

**Design/methodology/approach:** The most popular clustering algorithm is the Kohonen network, but today it is not the only algorithm capable of performing the task quickly and accurately. The paper compares the Kohonen network and the G-Means algorithm. The principle of operation of these two algorithms is briefly analyzed. The accuracy of these algorithms and the speed of their learning are compared.

**Findings:** By conducting experiments, conclusions were drawn about the speed and accuracy of the algorithms.

---

*ГрафиКон 2022: 32-я Международная конференция по компьютерной графике и машинному зрению, 19-22 сентября 2022 г.,*

*Рязанский государственный радиотехнический университет им. В.Ф. Уткина, Рязань, Россия*

EMAIL: Sudakov.oleg123@yandex.ru (О.В. Судаков); dmitdmit@mail.ru (Д.В. Дмитриев)

ORCID: 0000-0003-3723-5973 (О.В. Судаков); 0000-0001-9481-0968 (Д.В. Дмитриев)



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Originality/value: The relevance of this work lies in the fact that the preparation of training pairs for intelligent systems and the process of learning with a teacher is a resource-intensive task. The systems of self-learning algorithms under consideration will significantly increase the learning rate, as well as eliminate the need for manual classification of data and the creation of training pairs for the perceptron, which in turn will allow you to create a self-learning system with the ability to generalize and predict.

### Keywords

Neural networks, learning, self-learning algorithms, Unsupervised learning, G-Means, Kohonen network, hybrid systems.

## 1. Введение

Современные интеллектуальные системы активно используются в обществе для решения различных задач. Если некоторое время назад, это были простые алгоритмы, которые могли решать узкий спектр задач, то сегодня это сложные системы, состоящие из множества эвристических и линейных алгоритмов. Данные системы, имеют широкую область применения в бизнесе, автомобилях, играх и т.д. [16-20].

На данный момент активно используется интеллектуальная система, состоящая из сети Кохонена и персептрона для решения задач верификации и прогнозирования в условиях отсутствия предварительной обработки данных с целью выделения обучающих пар. Эта система хорошо зарекомендовала себя как самообучающаяся система, не требующая дополнительной обработки данных перед обучением системы. Но со времен изобретения алгоритма Кохонена было разработано множество новых алгоритмов, которые также могут выполнять кластеризацию данных и разведывательный анализ с целью автоматической подготовки обучающих пар для многослойного персептрона.

Одним из таких алгоритмов является G-Means. Он позволяет классифицировать данные на произвольное количество кластеров, которые определяются самостоятельно в отличие от алгоритма K-Means, его предшественника, где требуется использование труда аналитика для определения конечного количества кластеров.

Алгоритм G-Means решает ту же задачу, что и сеть Кохонена, но в отличие от него имеет ряд важных улучшений.

Современное научное сообщество используют связку алгоритма Кохонена и сеть прямого распространения также известную как персептрон, что позволяет автоматически, опираясь только на исходные данные обучить алгоритм [23]. Однако поскольку данная система ещё не может являться полностью автоматической, т.к. сеть Кохонена требует тонкой настройки перед использованием. G-Means может решить данную проблему.

В данной работе, в рамках жизненного цикла программного обеспечения, а именно, этапа дизайна архитектуры программы для анализа клавиатурного почерка, были кратко разобраны и реализованы алгоритмы G-Means и Кохонена. Каждый из них работает в связке с персептроном прямого распространения. Данные алгоритмы выступают интерфейсом между входными данными и персептроном для того, чтобы создать обучающие пары. Также было проведено тестирование реализованных алгоритмов самообучения по критериям быстродействия и точности.

## 2. Сеть Кохонена

Карты Кохонена были описаны в его трудах в 1984 году [15]. Основной его идеей было создание самоорганизующейся сети способной обучаться без учителя. Использование такого обучения позволяет использовать только значения входных переменных, а в процессе обучения не происходит сравнения выходных значений с эталонными. Таким образом сеть развивает способность понимать структуру данных.

Самоорганизующиеся карты Кохонена – это особый вид нейронных сетей, разработанный для решения задач кластеризации данных. Основное нововведение в работе сети является

добавление в нейрон топологических свойств, то есть информации позволяющей судить о его расположении.

Самоорганизующиеся карты признаков - это эффективный алгоритм кластеризации, позволяющий проецировать многомерное пространство в пространство с более низкой размерностью. При этом создаются двумерные карты, на которых отображаются объекты, соответствующие векторам, а расстояния между ними показывают расстояние между векторами в многомерном пространстве. Сами значения признаков обычно отображаются различными цветами, что в свою очередь позволяет за счет различных цветовых оттенков показывать три измерения.

На сегодняшний день, алгоритм имеет множество модификаций и используется также в сочетании с другими алгоритмами для поиска закономерностей в больших массивах данных, в задачах прогнозирования [1], моделирования, а также выявления независимых признаков и т.д. в задачах безопасности [3], планирования [2], медицины [4] и других. Большое распространение сети Кохонена получили в задачах кластеризации «без учителя». Данная задача подразумевает использование строк из таблицы с данными как входных векторов и распределение строк по соответствующим кластерам.

Это позволяет алгоритму выполнять задачи классификации, требующие проведение некоторого статистического анализа, к примеру, региональной кластеризации по параметрам человеческого капитала или верификация личности по динамическим биометрическим параметрам. Простота реализации и использования, а также скорость его обучения и хорошая точность, делают сеть Кохонена хорошим не гибридным решением для эвристического решения поставленных задач.

Основными недостатками данного алгоритма является не полная автономность, т.к. требуется труд аналитика для расчета итогового значения кластеров. Однако решение данной проблемы было предложено в работе по методике определения оптимального числа нейронов выходного слоя сети Кохонена при решении задач кластеризации [22]. В данной работе предложен математический способ определения, основанный на вычислении среднего максимального значения для каждого кластера. С использованием следующей формулы:

$$mid = \frac{\sum_{i=0}^{k_n} Out_{in}}{\sum_{n=1} k_n} \quad (1)$$

где  $k_n$  – количество образцов, принадлежащих  $n$ -му кластеру, а  $Out_{kn}$  – максимальное значение выхода нейронов последнего слоя для  $n$ -го образца, принадлежащему к  $i$ -му кластеру.

Высокая точность данного метода, достигаемая за счет переобучения сети до тех пор, пока не будет найдено оптимальное значение, приводит к значительным временным затратам на обучение.

Можно сократить временные затраты в ущерб точности. Количество кластеров при инициализации сети Кохонена, должно быть установлено в количестве равным половине доступных данных. В процессе обучения сети, часть кластеров не будут использованы, что позволит исключить их. Это определит количество выходов нейронной сети, достаточного для работы классификатора. Данным фактом можно воспользоваться для сокращения количества выходов будущей нейронной сети прямого распространения, выполняющей глубокий анализ данных, до действительного используемого.

Вторым важным недостатком алгоритма является то, что конечный результат во многом зависит от начальной настройки сети. Так точность используемой в последующих экспериментах сети зависит от количества эпох, которое уменьшается с увеличением данных. Однако число итераций зачастую не оптимально. Это может увеличить время обучения если количество эпох выше оптимального или уменьшить точность работы сети в ином случае.

Для настройки сети используются следующие формулы.

- Формула квадрата Евклидова расстояния:

$$d_{pk} = \sum_{i=0}^n (x_{pi} - x_{ki})^2 \quad (2)$$

где  $d_{pk}$  - квадрат расстояния между точкой и кластером,  $x_{pi}$  – координаты точки P,  $x_{ki}$  – координаты центра кластера k.

- Формула коррекции весовых коэффициентов (Формула 3)

$$w_{ij}(t + 1) = w_{ij}(t)(1 - \alpha(t)) + \alpha(t)(x_i) \quad (3)$$

где  $w_{ij}$  – вес  $i$  входа для  $j$  выхода, а  $\alpha(t)$  – коэффициент обучения.

### 3. K-Means и G-Means

Одной из популярных и часто используемых методик кластеризации является разделительная кластеризация, где на  $n$  элементов задается  $k$  кластеров, при этом  $k < n$ . К простым и эффективным алгоритмам относится K-Means. Он активно используется в задачах медицины [5-6], анализа текста [7-8], информационной безопасности [9] и других.

Данный тип кластеризации зарекомендовал себя как имеющий высокую степень надёжности и возможность дальнейшей высокопроизводительной обработки информации для применения методов прогнозирования, что открывает дополнительные возможности для анализа данных [21].

K-Means в качестве основной составляющей использует принцип оптимального разбиения множества данных на  $k$  кластеров. При этом, количество кластеров, как и в случае сети Кохонена изначально рассчитывается аналитиком. Далее определяются центры кластеров. Для это вычисляется среднее значения переменных объектов, которые входят в кластер. Однако до начала обучения алгоритма невозможно предсказать, какие объекты к какому кластеру будут принадлежать.

Для решения данной проблемы, алгоритм берет число объектов равное числу определенных аналитиком кластеров и объявляет их центрами. Далее алгоритм итеративно производит вычисление новых центров кластеров и распределяет в них объекты из обучающей выборки. Процесс продолжается до тех пор, пока объекты выборки не перестанут изменять свои кластера на новых итерациях.

Однако в данной реализации алгоритм K-Means является альтернативой сети Кохонена с теми же недостатками, а именно важностью присутствия аналитика и зависимости от изначальной настройки сети.

Данным недостатком лишён алгоритм G-Means. Он включает в себя алгоритм K-Means, но добавляет возможность автоматического определения количества кластеров, что позволяет полностью автоматизировать процесс кластеризации. Для реализации подобного функционала G-Means последовательно выполняет статистический тест. Этот тест проверяет, чтобы данные подчинялись определенному Гауссовскому закону распределения (откуда и название алгоритма). Тест проходят все кластера на каждой итерации. Если тест не выполняется, то кластер необходимо разбить на два отдельных кластера. Число кластеров будет оптимальным, когда тест станет выполняться для каждого кластера.

Важно отметить что основой алгоритмов K-Means и G-Means является гипотеза о компактности, которая говорит, что информацию о объектах можно представить, как многомерные вектора, которые образуют в пространстве своеобразный комок.

### 4. Гибридные системы

Рассмотренные алгоритмы имеют широкую область применения в различных задачах. Но обычно, данные алгоритмы используются как часть более сложной системы. К примеру алгоритм K-Means в работах [10-11] используется совместно с генетическим алгоритмом, а сеть Кохонена в [12-14] с персептроном.

Не смотря на то что рассмотренные алгоритмы уже является самообучающимся, у них отсутствует одно из важнейших свойств – обобщение. Свойством обобщения обладает персептрон. Поэтому для решения задачи обобщения в рамках данного проекта реализуется гибридная система в основе которой лежит персептрон и один из алгоритмов кластеризации.

Чаще всего, для создания такой системы используют самоорганизующиеся карты Кохонена [23]. Однако цель данной статьи сравнить эффективность работы персептрона, обученного не только с данным алгоритмом, но и с классификатором G-Means.

При проведении эксперимента будет реализована система, способная автоматически распределять данные по кластерам и проводить обучение персептрона. Её задачей будет не только сгруппировать данные, но и выявлять к какой группе относятся новые входящие объекты для проверки их достоверности.

Для создания системы используются следующий общий алгоритм, основанный на известных методах интеллектуального анализа данных:

- собрать обучающую выборку, состоящую из достоверных данных;
- провести разведочный анализ собранных данных;
- провести глубокий анализ собранных данных;
- оценить достоверность каждой из строк таблицы;
- исключить возможность взаимодействия с недостоверными данными.

Первый этап необходимый для работы системы - выборка, состоящая из достоверных данных, это повысит качество обучения. Невыполнение данного этапа может сильно снизить точность работы персептрона в реальных задачах.

Второй этап – разведывательный анализ используется для создания обучающих пар из полученного массива данных. Для этого применяются самообучающиеся алгоритмы, которые проводят кластеризацию. Поскольку понятие кластеризация субъективно, стоит учитывать, что распределение по кластерам может не совпадать с мнением аналитика. Однако обычно важно наличие общего элемента у объектов, а не принадлежность к конкретному кластеру в задачах подобного типа.

После окончания анализа, данные отправляются персептрону, который использует входящие вектора аналогичные тем, что использовались при разведывательном анализе, а в качестве эталонных значений используются присвоенные кластера. Для настройки весов персептрона используется классический метод коррекции ошибок с помощью алгоритма обратного распространения.

Когда персептрон обучен, ему передаются данные ранее не содержащиеся в обучающей выборке. Персептрон выбирает к какому кластеру принадлежит объект и выдаёт в качестве результата соответствующие вероятности.

Проверка будет проводиться с помощью двух баз с данными. Первая база включает в себя 8 строк с данными которые можно разделить на несколько подгрупп одним логическим образом. Вторая база будет содержать 400 строк данных, которые могут быть распределены различным способом.

Первым критерием работы системы является предсказуемость кластеризации. Проверка проводится путем проверки алгоритмов на выборке, данные которой можно распределить на кластеры однозначно

Вторым критерием работы системы является скорость кластеризации выборки. Используются данные большой выборки и измеряется скорость работы самообучающихся алгоритмов.

Третьим критерием успешности выполнения поставленной задачи является вероятность, которую персептрон выдаёт на данные относящиеся к выборке, но не входящие в неё (ожидается высокое значение) и к объекту, не относящемуся к выборке вовсе (ожидается отрицательная реакция на все кластеры).

## **5. Результаты эксперимента**

В качестве инструментов, рассматриваемых для реализации алгоритмов, использовался язык Python с библиотекой, применяемой для высокопроизводительных вычислений – TensorFlow. Данный набор также планируется применить при разработке программного обеспечения. Таким образом код, полученный в результате эксперимента, можно переиспользовать на других этапах жизненного цикла.

Предоставим алгоритмам для обучения следующую базу данных (таблица 1).

Таблица 1 - База данных для проверки точности

ID	VAL1	VAL2	VAL3	VAL4
1	1	1	0	0
2	0	0	1	1
3	1	1	0	0
4	0	0	1	1
5	1	1	0	0
6	0	0	1	1
7	1	1	0	0
8	0	0	1	1

В таблице приведены чередующиеся векторы. Значения каждого элемента вектора представлены в полях VAL.

Очевидно, что вектора можно разделить на группы следующим единственным образом (таблица 2).

Таблица 2 - Ожидаемый результат

Кластер 1	Кластер 2
1100	0011
1100	0011

Основываясь на этом, проверяется точность работы алгоритмов. Было проведено 100 итераций в каждой из которых сети обучались данными векторами, пока не выдали ответ. Количество правильных разделений показывает процент точности работы алгоритмов. В таблице 3 указан диапазон точности работы алгоритмов.

Таблица 3 - Результаты экспериментов на проверку точности

G-Means	Kohonen
97-100%	25-33%

По результатам тестирования выяснилось, что точность G-Means намного выше чем сети Кохонена. Однако, точность сети Кохонена возможно увеличить путём повышения количества итераций, но это скажется на скорости обучения.

Все эксперименты проводились на машине со следующими характеристиками:

- процессор Intel Core i9 9900;
- ОЗУ 16 Гб, 4200 MHz

Для проверки скорости работы алгоритмов используется массив данных следующего вида (таблица 4).

Таблица 4 - Пример данных для проверки скорости

Gender	Age	Salary
Male	19	19000
Male	35	20000
Female	26	43000
Female	27	57000
Male	19	76000

Каждая строка содержит три поля: пол человека, его возраст и заработную плату. Количество строк в базе данных сравнительно велико. Благодаря этому можно сделать вывод о скорости кластеризации засекая время, за которое алгоритм предоставит готовый результат.

В таблице 5 приведена скорость кластеризации данных для представленной выборки.

**Таблица 5 - Результаты экспериментов на проверку скорости**

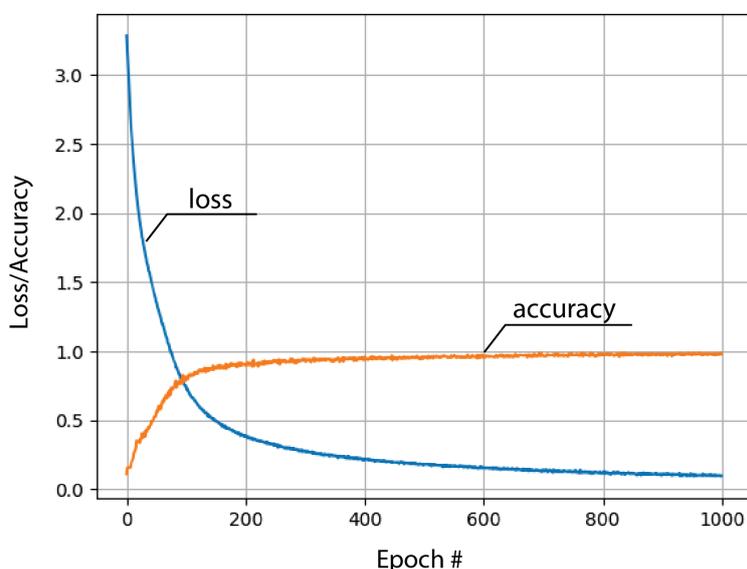
G-Means	Kohonen
< 1 сек.	10-30 минут

Сеть Кохонена работает во многом медленнее из-за невозможности точно определить необходимое число итераций. В итоге, сеть делает множество лишних циклов обучения, что расходует дополнительное время. Для проверки работоспособности системы с каждым из алгоритмов, были обучены 2 персептрона. Для обучения использовались результаты работы G-Means и сети Кохонена. Нейронная сеть прямого распространения получала две строки:

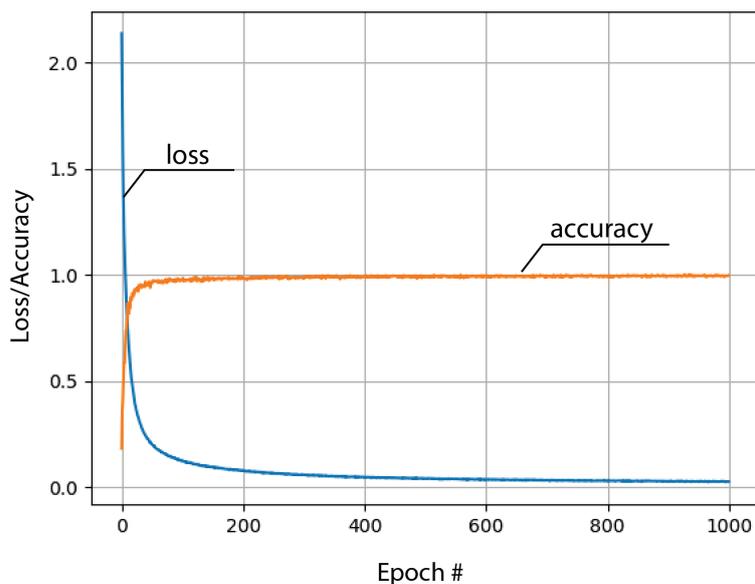
- данные из тестовой выборки;
- данные которые являются неверными и не подходят ни к одному кластеру.

Ожидается, что персептрон покажет к какому кластеру вероятнее всего принадлежит строка в первом случае, и маленькие вероятности для каждого кластера во втором.

Результаты обучения персептрона (рисунки 1-2).



**Рисунок 1 – График обучения на данных от сети Кохонена**



**Рисунок 2 – График обучения на данных от алгоритма G-Means**

Результаты эксперимента для G-Means приведены в таблице ниже. Поле «Expected» содержит ожидаемое эталонное значение, а поле «Result» полученное значение (таблица 6).

**Таблица 6** -Результаты экспериментов с перцептроном и G-Means

Gender	Age	Salary	Expected	Result
Male	22	19000	100%	97%
Unknown	7	20000	0%	35%

Результаты эксперимента для сети Кохонена (таблица 7).

**Таблица 7** - Результаты экспериментов с перцептроном и сетью Кохонена

Gender	Age	Salary	Expected	Result
Male	22	19000	100%	82%
Unknown	7	20000	0%	48%

Из полученных данных можно сделать вывод, что алгоритм G-Means является хорошей альтернативой алгоритму Кохонена. G-Means быстрее справляется с поставленными задачами, за счёт автоматического определения необходимого количества итераций (что одновременно упрощает его настройку), более точно выполняет кластеризацию. Благодаря этому, перцептрон быстрее обучается и показывает лучшие результаты при определении принадлежности новых данных к кластеру. Стоит учитывать, что есть другие реализации алгоритма K-Means с автоматическим определением количества кластеров [24], которые потенциально могут показать результаты ещё выше. Алгоритм G-Means был выбран благодаря его эффективной и быстрой работе, а также большой популярности.

## 6. Заключение

В данной работе проведен эксперимент, в результате которого произведено сравнение двух алгоритмов, а именно сети Кохонена и G-Means. Результаты эксперимента показали, что для самообучающейся интеллектуальной системы наиболее эффективным алгоритмом является G-Means. Данный алгоритм показал высокую точность и скорость работы по сравнению с сетью Кохонена. Также проведено сравнение двух алгоритмов на реальных данных где G-Means доказал свою эффективность. Данный алгоритм выгоднее использовать не только по причине его высоких показатели скорости и эффективности, но и благодаря его способности автоматически определять число необходимых кластеров, и итераций для обучения что исключает необходимость тонкой настройки перед началом работы всей системы.

## 7. Список источников

- [1] Guo A. et al. Data mining algorithms for bridge health monitoring: Kohonen clustering and LSTM prediction approaches // The Journal of Supercomputing. 2020. Т. 76. №. 2. С. 932-947.
- [2] Pukkala T. Can Kohonen networks delineate forest stands? // Scandinavian Journal of Forest Research. 2021. Т. 36. №. 2-3. С. 198-209.
- [3] Barletta V. S. et al. Intrusion detection for in-vehicle communication networks: An unsupervised kohonen som approach // Future Internet. 2020. Т. 12. №. 7. С. 119.
- [4] Mallick P. et al. Kohonen's self-organizing map optimizing prediction of gene dependency for cancer mediating biomarkers // Emerging Technologies in Data Mining and Information Security. – Springer, Singapore, 2019. С. 863-870.
- [5] Hassan N. S. et al. Medical Images Breast Cancer Segmentation Based on K-Means Clustering Algorithm: A Review // Ultrasound. 2021. Т. 27. С. 28.
- [6] Armya R. E. A., Abdulazeez A. M. Medical images segmentation based on unsupervised algorithms: a review // Qubahan Academic Journal. 2021. Т. 1. №. 2. С. 71-80.

- [7] Liu W. et al. Study on Chinese Text Clustering Algorithm Based on K-Mean and Evaluation Method on Effect of Clustering for Software-intensive System // 2020 International Conference on Computer Engineering and Application (ICCEA). IEEE, 2020. С. 513-519.
- [8] Rong Y. et al. Staged text clustering algorithm based on K-Means and hierarchical agglomeration clustering // 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2020. С. 124-127.
- [9] Jain T., Gupta C. Multi-Agent Intrusion Detection System Using Sparse PSO K-Mean Clustering and Deep Learning // Proceedings of 2nd International Conference on Artificial Intelligence: Advances and Applications. Springer, Singapore, 2022. С. 91-102.
- [10] Haq E. U. et al. Performance evaluation of novel selection processes through hybridization of K-Means clustering and genetic algorithm // Applied Ecology and Environmental Research. – 2019. Т. 17. №. 6. С. 14159-14177.
- [11] Mustafi D., Sahoo G. A hybrid approach using genetic algorithm and the differential evolution heuristic for enhanced initialization of the K-Means algorithm with applications in text clustering // Soft Computing. 2019. Т. 23. №. 15. С. 6361-6378.
- [12] Belej O., Halkiv L. Using Hybrid Neural Networks to Detect DDOS Attacks // 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP). IEEE, 2020. С. 61-66.
- [13] Ghaleb M. S. et al. Image retrieval based on self-organizing feature map and multilayer perceptron neural networks classifier // 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS). IEEE, 2019. С. 189-193.
- [14] Djuris J. et al. Application of machine-learning algorithms for better understanding of tableting properties of lactose co-processed with lipid excipients // Pharmaceutics. 2021. Т. 13. №. 5. С. 663.
- [15] Kohonen T. The Self-Organizing Map // Invited Paper – 1990.
- [16] Климин, А. А., Гилева С.С., Отекина Н.Е. Использование искусственного интеллекта в автомобилях // Мир Инноваций. 2020. № 4. С. 48-51.
- [17] Коваленко, Т. А., Никифоров Р. В. Искусственный интеллект в современном мире // Научное пространство: актуальные вопросы теории и практики: материалы II Всероссийской научно-практической конференции (Чебоксары, 24 сентября 2021 года.) / Чебоксары: Негосударственное образовательное частное учреждение дополнительного профессионального образования "Экспертно-методический центр", 2021. С. 11-15.
- [18] Кононенко Д. В., Хахина А. М. Развитие искусственного интеллекта в играх и применение в других отраслях // Информационные технологии: прошлое, настоящее, будущее: сборник статей по материалам межинститутской студенческой научно-практической конференции (Санкт-Петербург, 20 мая 2021 года.) / Санкт-Петербург: Санкт-Петербургский государственный лесотехнический университет имени С.М. Кирова, 2021. С. 172-177.
- [19] Бахтаирова Е. А., Вихорева М. В. Нейронные сети и искусственный интеллект в работе с обращениями граждан в органы власти // Развитие российского общества: вызовы современности: Материалы национальной научно-практической конференции с международным участием, посвященной 90-летию Байкальского государственного университета (Иркутск, 15–16 октября 2020 года.) / Иркутск: Байкальский государственный университет, 2021. С. 120-125.
- [20] Короткова Е. В. Машинное обучение и нейронные сети. Важный шаг в развитии человечества // Современные вопросы естествознания и экономики: Сборник трудов III Международной научно-практической конференции (Прокопьевск, 18 марта 2021 года.) / Прокопьевск: Филиал Федерального государственного бюджетного образовательного учреждения высшего профессионального образования "Кузбасский государственный технический университет имени Т. Ф. Горбачева" в г. Прокопьевске, 2021. С. 240-242.
- [21] Антропов В.В. Применение алгоритмов кластеризации K-Means и G-Means в задачах распознавания воздушных объектов // МНИЖ. 2017. №7-3 С. 61.
- [22] Фатхи Д. В., Галушка В. В. Методика определения оптимального числа нейронов выходного слоя сети Кохонена при решении задач кластеризации // Информационная безопасность регионов. 2011. №2. С. 41-44.

- [23] Фатхи, Д.В., Применение методов интеллектуального анализа данных для тестирования баз данных систем информационной безопасности // Информационная безопасность регионов, 2014. №1. -С. 48-51.
- [24] Фролов В.В., Слипченко С.Е., Приходько О.Ю. Метод расчета числа кластеров для алгоритма К-Means // Экономика. Информатика. 2020. №1. С. 213-225.