

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ «САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

(СПбГУ)

Институт Наук о Земле

Кафедра геологии месторождений полезных ископаемых

Жилкова Анна Анатольевна

**Особенности применения технологий машинного обучения при моделировании
геологических систем, прогнозировании и поиске месторождений полезных
ископаемых.**

Магистерская диссертация по направлению 05.04.01 «Геология»

(шифр образовательной программы ВМ.5515.2021)

Научный руководитель:

Алексеев Иван Александрович, кандидат геолого-минералогических наук, доцент
кафедры геологии месторождений полезных ископаемых.

Санкт-Петербург

2023

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
1 ГЛАВА. ТЕОРИТИЧЕСКИЕ ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ	6
1.1 Основные понятия о машинном обучении.....	7
1.2 Типы реальных данных и способы их использования.....	10
1.2.1 Категории алгоритмов машинного обучения.....	12
1.2.2 Задачи и алгоритмы машинного обучения.....	15
1.2.3 Различие между алгоритмами кластеризации.....	18
1.3 Сферы использования машинного обучения	20
1.3.1 Машинное обучение в геологии	21
1.3.2 Преимущества машинного обучения (самообучающихся алгоритмов)	23
1.3.3 Недостатки и сложности машинного обучения	25
1.4 Пути развития технологий машинного обучения в поисках полезных ископаемых	27
2 ГЛАВА. ГЕОЛОГИЯ КАЙЭНМЫВААМСКОГО РУДОПРОЯВЛЕНИЯ	29
2.1 Физико-географические сведения о районе исследований	30
2.2 Рудопроявление Кайэнмываам	32
2.3 Магматизм и метаморфизм Кайэнмываам	35
2.4 Условия формирования золото-серебряной эпиптермальной минерализации.....	37
2.5 Вмещающие породы Кайэнмываам.....	38
3 ГЛАВА. МЕТОДИКА СЕТИ КОХОНЕНА И СТАТИСТИЧЕСКИХ МЕТОДОВ ОБРАБОТКИ ГЕОХОМИЧЕСКИХ ДАННЫХ	41
3.1 Методика работы: Сеть Кохонена.....	42
3.1.1 Практическое применение сети Кохонена для моделирования геохимических данных.....	46
3.2 Статистический метод обработки геохимических данных	50
3.3 Интерпретация и сравнение полученных результатов на основе двух методов.	57
ЗАКЛЮЧЕНИЕ.....	68
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	69
Приложение 1.....	74
Приложение 2.....	75
Приложение 3.....	76

ВВЕДЕНИЕ

Актуальность.

В условиях растущей конкуренции и уменьшения количества доступных месторождений, применение технологий машинного обучения может стать важным фактором для оптимизации производственных процессов и повышения эффективности поиска, моделировании и прогнозировании месторождений полезных ископаемых. Использование машинного обучения позволит сократить время, затрачиваемое на анализ и обработку больших объемов данных. Кроме того, применение технологий машинного обучения может помочь выявить неочевидные закономерности в геологических данных.

Цель.

Целью магистерской диссертационной работы является создание оптимизированного алгоритма, на основе самоорганизующихся карт Сети Кохонена для обработки геохимических данных.

Задачи:

1. Изучить актуальные направления применения технологий обработки больших данных (Big Data), машинного обучения (Machine Learning), глубокого обучения (Deep Learning), а также рассмотреть современные методы хранения больших объемов данных.
2. Исследовать применение методов машинного обучения в геологии и других отраслях, проанализировать опыт применения этих методов и оценить их эффективность.
3. Выполнить многомерный статистический анализ и выделить перспективные зоны на участке.
4. Обработать полученные данные с помощью алгоритма кластеризации - сети Кохонена для выявления особенностей распределения полезных ископаемых и прогнозирования месторождений.
5. Сравнить результаты, полученные с помощью искусственных нейронных сетей (ИНС), с данными, полученными с помощью статистических методов, и оценить точность прогнозирования месторождений.
6. Описать алгоритм кластеризации - сети Кохонена, используемый в данной работе, с приведением подробных формул и описаний.

Объект исследования.

Объектом исследования является участок Кайэнмываам расположен в пределах Центрально-Чукотского сектора Охотско-Чукотского вулканогенного пояса и имеет размеры 20 на 53 км. Для изучения геохимических характеристик данного участка были собраны и проанализированы геохимические пробы методом индуктивно-связанной плазмы с использованием масс-спектрометрии.

Фактический материал.

Фактическим материалом, полученным в ходе исследования, являются данные Масс-спектрометрия с индуктивно-связанной плазмой. Также использовались карты предшественников для получения дополнительных данных об изучаемом участке.

Методика исследования.

Методика исследования включает в себя многомерный статистический анализ и машинное обучение на основе алгоритма кластеризации сети Кохонена, которые были применены для обработки и анализа геохимических данных и поиска возможных признаков наличия эпitherмальных золото-серебряных месторождений на участке.

Предмет исследования.

Являются алгоритмы машинного обучения, в частности, сети Кохонена, которые будут использованы для интерпретации данных и создания точных и геологически значимых прогнозов.

Практическая значимость исследования.

Практическая значимость исследования заключается в потенциале повышения эффективности и точности геологического моделирования, прогнозирования и разведки месторождений полезных ископаемых. Используя алгоритмы машинного обучения, можно лучше понять геологические процессы и делать прогнозы с более высокой точностью. Результаты этого исследования могут привести к принятию более обоснованных решений в горнодобывающей промышленности, что имеет решающее значение для устойчивого управления ресурсами и экономического роста. Кроме того, применение машинного обучения в геологии может служить моделью для применения этих технологий в смежных областях.

Алгоритмы машинного обучения предназначены для точного и эффективного прогнозирования многомерных данных. Данные предоставляются аналитикам для моделирования и прогнозирования взаимодействия между данными и изучаемыми объектами. Таким образом, анализ больших объемов разрозненных, многомерных, геопространственных данных с использованием алгоритмов машинного обучения открывает большие перспективы для промышленности и исследований в области наук о Земле.

В данной работе я тестирую новый подход к интерпретации данных при помощи алгоритмов машинного обучения. Работа выполняется на основе геохимических данных полученных с перспективного участка золото-серебряной минерализации на территории Чукотского вулканического пояса. Практическое применение алгоритмов машинного обучения требует реализации трех ключевых этапов: (1) предварительная обработка данных; (2) обучение алгоритму; и (3) оценка прогноза. Эта методология обеспечивает основу для

создания точных и геологически значимых прогнозов с минимальным вмешательством пользователя.

Для более точного и эффективного анализа геохимических данных, я буду использовать алгоритм самоорганизующейся карты Кохонена, который позволит произвести кластеризацию данных и выделить характерные особенности распределения геохимических параметров на исследуемом участке. С помощью этого алгоритма, можно будет обнаружить структуры в данных, которые не были заметны при первоначальном визуальном анализе.

После этого, я сравню результаты, полученные с помощью алгоритма Кохонена, с результатами, полученными при помощи статистического анализа, который является классическим методом обработки геохимических данных. Это позволит оценить эффективность использования алгоритма Кохонена в данном контексте и сравнить его с другими методами анализа геохимических данных.

Таким образом, использование алгоритма Кохонена и сравнение его с статистическим анализом позволит получить более полное представление о геохимических процессах на исследуемом участке и выделить наиболее значимые признаки, которые будут использоваться для прогнозирования золото-серебряной минерализации в будущем.

1 ГЛАВА. ТЕОРИТИЧЕСКИЕ ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ

Машинное обучение - это одно из самых востребованных и быстро развивающихся областей современной науки. Алгоритмы искусственного интеллекта находят применение во многих сферах деятельности человека. С каждым годом количество данных, которые мы генерируем и собираем, растет в геометрической прогрессии, и возможности их обработки и анализа без помощи машинного становится невозможным.

В первой главе работы будут рассмотрены основные понятия и термины, связанные с машинным обучением, а также будут описаны различные алгоритмы машинного обучения, их преимущества и недостатки. Одной из главных причин популярности машинного обучения является то, что оно может быть использовано во многих сферах деятельности. Каждый алгоритм машинного обучения имеет свои особенности и может быть применен для решения определенного набора задач. Для правильного выбора алгоритма машинного обучения, соответствующего конкретной задаче, необходимо иметь понимание его принципов функционирования и возможных способов применения. Таким образом, изучение различных алгоритмов машинного обучения и сфер их применения является важной задачей для всех, кто хочет понимать современные технологии и использовать их в своей работе.

1.1 Основные понятия о машинном обучении

Первая и самая простая нейронная сеть «перцептрон» была предложена американским психологом и нейрофизиологом Фрэнком Розенблаттом в 1957 г. Она состояла из одного скрытого слоя и сетью искусственных нейронов, способной к обучению с учителем (Казначев, 2016).

Термин «машинное обучение» ввел исследователь искусственного интеллекта Артур Самуэль еще в далеком 1959 году. На основе игры в шахматы он построил первую модель, где дальнейшее поведение компьютера не было запрограммировано (Kauri, 2020).

Искусственный интеллект (ИИ) относится к разработке компьютерных систем, которые могут выполнять задачи, обычно требующие человеческого интеллекта, такие как распознавание речи, понимание естественного языка, принятие решений. Искусственный интеллект достигается путем обучения алгоритмов машинного обучения большим объемам данных, что позволяет алгоритму распознавать закономерности и делать прогнозы.

Существует два основных типа ИИ: узкий или слабый ИИ, который предназначен для выполнения конкретной задачи, и общий или сильный ИИ, который обладает способностью выполнять любую интеллектуальную задачу, на которую способен человек. Область искусственного интеллекта постоянно развивается и продвигается вперед, исследователи разрабатывают новые алгоритмы и методы, чтобы сделать искусственный интеллект более эффективным.

Машинное обучение - это область искусственного интеллекта, которая использует алгоритмы для получения информации из данных и создания моделей, способных решать задачи без явного программирования. Искусственный интеллект, в свою очередь, описывает создание систем, способных имитировать человеческую интеллектуальную и творческую деятельность. В этом контексте, машинное обучение представляет собой имитацию процесса обучения, который мы можем наблюдать у людей и животных.

Искусственная нейронная сеть - это один из подходов к машинному обучению, который использует математическую модель, основанную на организации и функционировании биологических нейронных сетей. Эти модели используются для решения различных задач, например, распознавания образов, классификации данных и прогнозирования результатов.

Deep learning - это глубокое обучение, используемое для решения более сложных задач в машинном обучении. Он использует многослойные нейронные сети для обработки и анализа больших объемов данных. Это может быть полезно для распознавания речи,

обработки естественного языка, анализа изображений и других задач, которые требуют более высокой степени сложности.

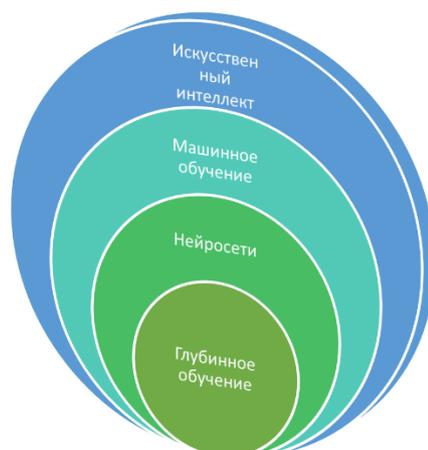


Рис. 1.1 Схематическое объяснение искусственного интеллекта. [Составлено автором]

Нейросеть – это взаимосвязь простейших вычислений. Входные сигналы являются вершинами и обрабатываются по определенной формуле, далее приобретенный новый сигнал направляется далее на другие нейроны. Эта взаимосвязь обладает определенным уровнем значимости. В результате получаем параметр, называемый – весовым коэффициентом. Входные сигналы суммируются, применяется функция активации $f(S)$, далее нейроны выводятся или же алгоритм повторяется. Из подобных несложных компонентов строится – граф (нейросеть) (Алексеев, 2022).

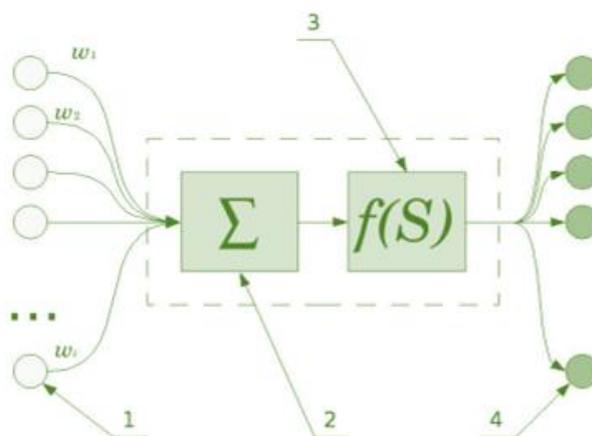


Рис. 1.2 Модель искусственного нейрона (1 – входные сигналы с весовым коэффициентом w_i , 2 – сумматор, 3 – функция активации, 4 – выходной сигнал). Источник: [Алексеев, 2022].

Данный раздел является вводным и посвящен основным понятиям машинного обучения, описаны все возможные виды алгоритмов и их использование в разных сферах, а так же основной акцент на использовании машинного обучения в науках о Земле. Описываются недостатки и преимущества искусственной нейронной сети.

1.2 Типы реальных данных и способы их использования

Мы живем в век данных, где все вокруг нас подключены к источнику данных, и все в нашей жизни записано в цифровом виде (Сао L, 2017). Например, в современном электронном мире существует множество различных типов данных, таких как данные Internet of Things (IoT), данные кибербезопасности, интеллектуальные городские данные, бизнес-данные, данные смартфона, данные социальных сетей, данные о здоровье, данные о COVID-19, неструктурированные данные месторождений полезных ископаемых и т.д

Все данные можно подразделить на структурированные, полуструктурированные и неструктурированные. Также используются метаданные “metadata” т.е данные о данных (Sarker IH, 2020).

- Структурированные данные – имеют четкую структуру, стандарты и очень легки в использовании. Обычно представляют собой таблицу. Примерами структурированных данных в геологии могут служить базы данных о месторождениях полезных ископаемых, геофизические карты и другие. Такие данные могут использоваться для построения моделей и прогнозирования поведения геологических систем.

- Неструктурированные данные – не имеют определенной структуры и формата, что делает данные сложными в использовании и понимании пользователями и программами. В основном содержат текстовые и мультимедийные материалы. Примерами неструктурированных данных в геологии могут служить геологические карты, фотографии горных пород и месторождений, записи геологических наблюдений и другие. Такие данные могут быть использованы для анализа текста и изображений, а также для обучения моделей машинного обучения.

- Полуструктурированные данные – имеют определенные структурные свойства, облегчающие его анализ HTML, XML, JSON документы, базы данных NoSQL и т.д (Iqbal, 2021). Пример: данные о результатах геологических исследований, которые могут содержать как структурированные данные (например, результаты анализа образцов горных пород), так и неструктурированные данные (например, описания геологических объектов и примечания).

- Метаданные – «данные о данных». Главное различие между «данными» и «методанными» заключается в том, что данные – это простой материал, который можно классифицировать, измерять, изменять. Метаданные – описывают основную информацию о данных, пример это размер файла, дата создания и редактирования, автор.

Big Data - это технологический термин, который описывает огромные объемы данных, которые накапливаются и обрабатываются в реальном времени (Victor M., 2013). Эти данные могут быть структурированными, например, данные из баз данных и таблиц, или

неструктурированными, например, данные из социальных сетей, датчиков и медиафайлов. Для работы с такими объемами данных необходимы специальные программы и технологии, такие как: Hadoop (Tom White, 2015) - это фреймворк с открытым исходным кодом, который позволяет распределять обработку больших объемов данных на кластеры серверов, Apache Spark (Chambers, 2018) - это высокопроизводительный движок для обработки больших объемов данных в реальном времени, NoSQL (Fowler, 2012) базы данных - это базы данных, которые позволяют хранить и обрабатывать неструктурированные данные, например, документы, графы и временные ряды, Apache Kafka (Shapira, 2017) - это распределенная платформа, которая позволяет обрабатывать потоковые данные в реальном времени, Tableau, QlikView и другие инструменты визуализации данных - позволяют создавать отчеты и дашборды для анализа больших объемов данных. Одним из примеров использования BigData в геологических исследованиях является проект "Deep Carbon Observatory" (DCO). DCO - это междисциплинарный проект, который изучает процессы, связанные с углеводородами. Для обработки больших объемов данных, собранных в рамках проекта, используются высокопроизводительные вычислительные системы и специальные инструменты для анализа данных.

Эффективность и результативность решений для машинного обучения зависит от типа, характера, структурированности данных и эффективного обучения нейронной сети. В областях машинного обучения данные описанные выше используются в различных областях.

1.2.1 Категории алгоритмов машинного обучения

В данном параграфе рассмотрим основные четыре категории алгоритмов машинного обучения.



Рис. 1.3 Схема категорий алгоритмов машинного обучения. [Составлено автором]

- Обучение с учителем – это тип машинного обучения, при котором машина обучается на основе положительного примера и на основе этих данных прогнозируется результат (Han J, 2021). Машинное обучение с учителем является синонимом классификации. Как правило задаются правильные значения на входные и выходные, нейросеть строит функции с входной переменной (x) с выходной переменной (y). При обучении с учителем для начала помечается набор данных, где модель узнает о каждом типе данных, после модель тестируется на тестовых данных, а затем прогнозируются выходные данные (Sarker IH, 2020). Можно использовать для оценки рисков, классификации изображений, фильтрации спама и т.д. Некоторые известные алгоритмы обучения с учителем включают в себя линейную регрессию, метод опорных векторов (SVM), случайный лес и градиентный бустинг.

Примеры применения контролируемого обучения в геологии:

1. Классификация геологических данных: использование методов машинного обучения для классификации геологических данных может помочь в определении типов пород и нахождении месторождений полезных ископаемых. В статье "Application of supervised learning methods to geological data classification" были использованы алгоритмы, такие как метод опорных векторов (SVM) и алгоритм случайного леса (Random Forest) для классификации геологических данных.

2. Предсказание свойств горных пород: контролируемое обучение может использоваться для предсказания свойств горных пород, таких как плотность, прочность и

твердость. В статье "Predicting rock strength using artificial neural networks" была использована нейронная сеть для предсказания прочности горных пород на основе данных о плотности, водонасыщенности и других параметрах.

3. Поиск месторождений полезных ископаемых: контролируемое обучение может помочь в поиске месторождений полезных ископаемых путем анализа геохимических данных. В статье "Geochemical prospecting using supervised learning algorithms" были использованы алгоритмы машинного обучения, такие как метод опорных векторов и нейронные сети, для анализа геохимических данных и поиска месторождений полезных ископаемых.

4. Предсказание риска землетрясений: контролируемое обучение может использоваться для предсказания риска землетрясений на основе геологических данных. В статье "Supervised learning approaches for earthquake risk prediction" (Main, I. G, 2019) были использованы алгоритмы машинного обучения для анализа данных о геологических структурах, глубинах и других факторах, связанных с землетрясениями.

- Обучение без учителя – или кластеризация. Ставится задача достичь определенной цели, задается определенный набор входных данных (Han J, 2021). Данные не маркируются, модели сами находят скрытые закономерности из предоставленных данных. Основная задача найти базовую структуру данных, сгруппировать по сходствам и представить в сжатом формате. Некоторые известные алгоритмы обучения без учителя включают в себя кластеризацию и снижение размерности. Кластеризация может использоваться для группировки геологических образцов на основе их сходства по геохимическим параметрам. В дальнейшей работе будет рассмотрена данная категория алгоритмов.

-Полуконтролируемое обучение – это обучение на основе проб и ошибок системы, для достижения результата (Kaelbling, 1996). Что бы алгоритм работал, пользователю необходимо указывать на правильные и неправильные подходы. Никаких подсказок к решению задачи не дается, модель сама на основе положительных и отрицательных результатов выполняет поставленную задачу. ИИ на протяжении всего обучения и множества решенных задач накапливает опыт. В статье "Semi-Supervised Learning for Fault Detection in Seismic Images" (Deng et al., 2021), где авторы использовали полуконтролируемое обучение для обнаружения дефектов на сейсмических изображениях. Они использовали ограниченный набор размеченных данных и большое количество не размеченных данных для обучения модели, которая могла обнаруживать дефекты на не размеченных данных.

- Активное обучение – позволяет пользователям активно влиять на ход обучения. Машина может попросить промаркировать пример из выборки или данные синтезируемой обучаемой программой. Преимущество модели в том, что данные постоянно

оптимизируются, получая знания от исследователя. Поиск нефтяных месторождений: Алгоритмы активного обучения могут быть использованы для определения оптимальных мест для бурения на месторождениях нефти и газа. "Active Learning of Rock Typing Models from Multiscale Measurements" (Zhao et al., 2018). В этой статье авторы использовали активное обучение для создания моделей типов пород на основе измерений на нескольких масштабах. Алгоритм выбирал примеры, которые могли наилучшим образом уточнить модель, и предлагал их для анализа экспертам. Таким образом, эксперты использовали информацию, полученную из предложенных примеров, для уточнения модели.

1.2.2 Задачи и алгоритмы машинного обучения

Задачи машинного обучения:

Классификация: определение принадлежности объекта к определенному классу.

Примеры из статей: классификация геологических объектов по типу, классификация геологических формаций по типу (Sun, J., 2021).

Регрессия: предсказание числовых значений в зависимости от входных данных.

Пример: прогнозирование объемов запасов нефти и газа в месторождениях, предсказание стоимости добычи рудных полезных ископаемых (Demiral, B., 2020).

Кластеризация: группировка объектов на основе сходства между ними. Примеры из статей: кластеризация геологических формаций по их геохимическим свойствам, кластеризация геологических объектов по их физическим свойствам (Zhang, H., 2020).



Рис. 1.4 Схема алгоритмов машинного обучения. [Составлено автором]

К алгоритмам классификации относятся:

- Бинарная классификация – применяется к задачам, где два взаимоисключающих себя класса. это номинальный атрибут только с двумя категориями или состояниями: 0 или 1, где 0 обычно означает, что атрибут отсутствует, а 1 означает, что он присутствует. Бинарные атрибуты называются булевыми, если два состояния соответствуют истине и лжи (Bishop, C. M., 2006).

- Мультиклассовая классификация – данная задачи классификации имеет более двух видов классов. В отличие от бинарной классификации не имеет нормальных или аномальных исходов (Shalev-Shwartz, 2014). Например, имеет возможность классифицировать различные сетевые атаки DoS, U2R, R2L.

- Классификация с несколькими метками – является значимым фактором, когда данные связаны с несколькими классами. Каждый образец данных имеет метку m и может быть

отнесен к нескольким классам n (Kaufmann, 2005). Значение m может принадлежать более чем одному классу. Многоклассовая классификация включает передовые алгоритмы машинного обучения, которые позволяют предсказывать несколько взаимно непересекающихся классов или меток, в отличие от традиционных задач классификации, где классы являются взаимоисключающими.

- Байесовский классификатор – основан на теореме Байеса. Данная классификация может использоваться как в бинарных, так и в мультиклассовых алгоритмах. Алгоритм определяет класс к которому принадлежит объект, расчет вероятности (Pedregosa, 2011). Одним из примеров использования классификации в геологии является определение типа пород на основе геофизических данных. В статье " Microstructure Investigation of Oil-Bearing Rhyolites: A Case Study from the Nailar Basin " (Zhang, H., 2020), авторы применяют алгоритмы деревьев решений, Байесовского классификатора и опорных векторов для классификации типов пород на основе геофизических данных в бассейне Хайлаэр в Китае.

Алгоритмы кластеризации - это методы группировки объектов на основе их сходства в наборы, называемые кластерами. Они используются в различных областях, включая машинное обучение, статистику, биоинформатику, анализ данных и многие другие. Существует множество различных алгоритмов кластеризации, и каждый из них подходит для различных типов данных и задач.

Некоторые из наиболее распространенных алгоритмов кластеризации включают в себя:

- K-средних - этот алгоритм является одним из самых популярных и простых методов кластеризации. Он начинается с задания случайных центроидов кластеров, затем объекты данных присваиваются к ближайшим центроидам, и центры кластеров обновляются до тех пор, пока не будет достигнута сходимость. Это задача разделения объектов на группы (кластеры) на основе сходства между ними. Одним из примеров использования кластеризации в геологии является выявление типов пород на основе их химического состава.

- DBSCAN - алгоритм основан на плотности данных. Он начинается с выбора случайной точки данных и нахождения всех точек данных, находящихся на заданном расстоянии от нее. Затем алгоритм расширяет кластеры путем добавления точек, находящихся в пределах заданного радиуса, и перехода к соседним кластерам (Ester, M., 1996).

- Агломеративная кластеризация - алгоритм начинается с того, что каждый объект данных рассматривается как отдельный кластер (Murtagh, F., 2012). Затем он объединяет наиболее близкие кластеры, пока не будет достигнуто заданное количество кластеров или определенный уровень сходства.

- Спектральная кластеризация - алгоритм основан на разбиении матрицы сходства между объектами данных на собственные векторы и использует их для кластеризации (Jordan, M., 2002). Он хорошо работает для данных с ярко выраженными структурами.

- Сеть Кохонена (Kohonen network), также известная как самоорганизующаяся карта Кохонена (Self-Organizing Map, SOM) (Kohonen, T., 1982, 1990, 1995), это нейронная сеть без учителя, которая используется для кластеризации и сжатия данных. Данный алгоритм кластеризации будет подробно описан в этой работе.

Это только некоторые из множества алгоритмов кластеризации, которые используются в практике. Выбор конкретного алгоритма зависит от характеристик данных и задачи, которую необходимо решить.

Алгоритмы регрессии являются одними из наиболее популярных алгоритмов машинного обучения, используемых в геологии.

- Линейная регрессия является одним из наиболее простых алгоритмов регрессии, который можно использовать для моделирования линейной зависимости между независимыми и зависимой переменными. В геологии линейная регрессия может быть использована для моделирования зависимости между геохимическими параметрами и другими геологическими показателями.

- Алгоритмы регрессии на основе решающих деревьев, такие как CART (Classification and Regression Trees) и Random Forest, также широко используются в геологии для моделирования нелинейной зависимости между независимыми и зависимой переменными. Эти алгоритмы могут использоваться для моделирования сложных взаимодействий между геохимическими параметрами и другими геологическими показателями, такими как тектонические структуры и геологические формации. Еще одним примером использования алгоритмов регрессии в геологии является исследование, опубликованное в журнале *Geophysical Journal International* (Li, Y., 2017). В нем был использован алгоритм случайного леса (Random Forest Regression) для прогнозирования значения крикетного звука на основе геологических характеристик местности, таких как геологический тип и тип грунта.

- Градиентный бустинг - это алгоритм машинного обучения, который комбинирует несколько слабых моделей, чтобы получить более сильную модель (Friedman, J. H., 2001). В геологии градиентный бустинг может использоваться для построения моделей, которые могут учитывать множество геологических параметров, таких как расстояние до зоны минерализации, геохимические и гравитационные параметры.

1.2.3 Различие между алгоритмами кластеризации

K-means и сеть Кохонена - это два разных алгоритма машинного обучения, используемые для кластеризации данных. Оба алгоритма пытаются разбить данные на группы или кластеры на основе их сходства (Е. Мамонтов, 2015). Однако, у них есть различия в методе работы.

Алгоритм K-means работает следующим образом:

1. Выбирается количество кластеров (K), которое мы хотим получить.
2. Выбираются случайные центры для каждого кластера.
3. Каждый элемент данных относится к ближайшему центру кластера.
4. Вычисляется среднее значение для каждого кластера и перемещается центр кластера к новому местоположению.
5. Шаги 3-4 повторяются до тех пор, пока центры кластеров не стабилизируются.

У алгоритма есть некоторые недостатки: он слишком чувствителен к выбросам, что может исказить среднее значение; он работает медленно на больших базах данных; требуется задание количества кластеров; невозможно применение алгоритма к данным, в которых присутствуют пересекающиеся кластеры.

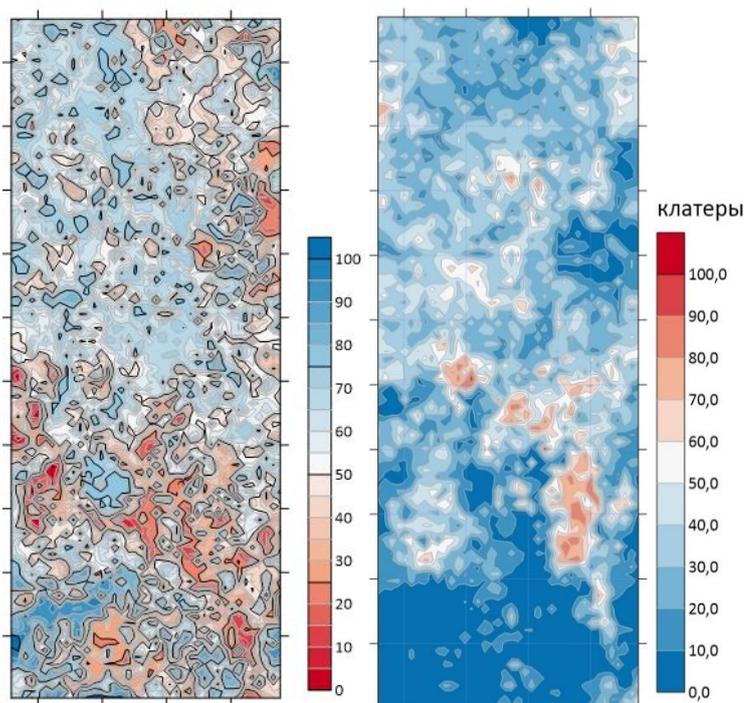


Рис. 1.5 Сравнение карт, полученных алгоритмом кластеризации k-means и сетью Кохонена. [Составлено автором]

С другой стороны, сеть Кохонена работает так:

1. Случайным образом инициализируются веса нейронов, соответствующих каждому кластеру.
2. Каждый элемент данных сравнивается с каждым нейроном, и выбирается нейрон с наименьшим расстоянием.
3. Нейрон с наименьшим расстоянием становится победителем, и его веса обновляются, чтобы сделать его ближе к этому элементу данных.
4. Соседние нейроны также обновляются, чтобы они стали ближе к выбранному элементу данных.
5. Шаги 2-4 повторяются для каждого элемента данных.

Алгоритм Expectation-Maximization (EM) и сеть Кохонена - это два разных алгоритма машинного обучения, используемые для кластеризации данных. Оба алгоритма пытаются разбить данные на группы или кластеры на основе их сходства. Однако, у них есть различия в методе работы.

Алгоритм EM работает следующим образом:

1. Выбирается количество кластеров (K), которое мы хотим получить.
2. Инициализируются параметры распределения для каждого кластера.
3. Каждый элемент данных присваивается вероятность принадлежности к каждому кластеру.
4. Используя оценку вероятности, вычисляются новые параметры распределения для каждого кластера.
5. Шаги 3-4 повторяются до тех пор, пока параметры не стабилизируются.

Математические различия между этими алгоритмами заключаются в способе моделирования кластеров. В алгоритме EM каждый кластер представлен параметрами вероятностного распределения, в то время как в сети Кохонена каждый кластер представлен нейроном и его соответствующими весами.

1.3 Сферы использования машинного обучения

Машинное обучение – это, быстро развивающаяся область, которая оказывает значительное влияние на многие отрасли. Он используется для анализа больших наборов данных и выявления закономерностей, составления прогнозов и автоматизации процессов. Рассмотрим примеры использования алгоритмов машинного обучения в таких областях, как медицина, финансы, обработка естественного языка и компьютерное зрение. Кроме того, мы рассмотрим программы, используемые для реализации этих алгоритмов.

Алгоритмы машинного обучения используются в медицине для анализа медицинских данных, диагностики заболеваний и прогнозирования результатов лечения. Примеры таких алгоритмов включают нейронные сети, деревья решений и метод опорных векторов. Для реализации этих алгоритмов используются программы, такие как Python и R, а также специализированные библиотеки, например, TensorFlow и Keras.

Алгоритмы машинного обучения также используются в финансовой отрасли для прогнозирования рыночных трендов, определения рисков и оценки активов (Yang, X., 2019). Некоторые из наиболее часто используемых алгоритмов включают линейную регрессию, случайный лес и градиентный бустинг. Для реализации этих алгоритмов используются программы, такие как Python, R и MATLAB.

Алгоритмы машинного обучения также находят применение в обработке естественного языка, например, в анализе тональности текста, определении языка и машинном переводе. Примеры таких алгоритмов включают рекуррентные нейронные сети, сверточные нейронные сети и методы анализа тональности (Jurafsky, D., 2019). Для реализации этих алгоритмов используются программы, такие как Python, Java и C++. Специализированные библиотеки, такие как NLTK, Gensim и SpaCy, также используются для обработки естественного языка.

Алгоритмы машинного обучения находят широкое применение в компьютерном зрении, например, для распознавания объектов, классификации изображений и обнаружения лиц. Некоторые из наиболее часто используемых алгоритмов включают сверточные нейронные сети, методы глубокого обучения и метод опорных векторов (Russakovsky, 2015). Для реализации этих алгоритмов используются программы, такие как Python, C++ и MATLAB, а также специализированные библиотеки, например, TensorFlow, PyTorch и OpenCV.

1.3.1 Машинное обучение в геологии

В последние годы машинное обучение применяется в широком спектре отраслей промышленности, включая геологию. Использование ML в геологии может помочь геологам делать более точные прогнозы, находить новые месторождения полезных ископаемых и более эффективно изучать геологические системы.

Статья "Artificial Intelligence in Geology" МакИнерни и Чен, опубликованная в журнале Earth-Science Reviews, представляет собой обзор применения техник искусственного интеллекта (ИИ) в геологии. В статье рассматриваются различные алгоритмы машинного обучения, такие как искусственные нейронные сети, деревья решений, метод опорных векторов и байесовские сети, которые в настоящее время используются в геологии.

Авторы обсуждают, как эти техники ИИ применяются в различных областях геологии, включая минеральную эксплуатацию, нефтегазовую геологию, инженерную геологию и экологическую геологию. Например, искусственные нейронные сети были использованы для прогнозирования свойств резервуаров в нефтегазовой геологии, а метод опорных векторов был использован для прогнозирования месторождений руд в минеральной эксплуатации.

Одним из ключевых применений машинного обучения в геологии является прогнозирование месторождений полезных ископаемых. Это включает в себя использование алгоритмов машинного обучения для выявления закономерностей и взаимосвязей в геологических данных, таких как геофизические исследования, геохимический анализ и данные керн бурения. Например, исследователи использовали методы машинное обучение для прогнозирования наличия месторождений полезных ископаемых на основе анализа спутниковых снимков. Используя алгоритмы машинное обучение, геологи могут анализировать большие объемы данных и делать прогнозы быстрее и точнее, чем с помощью традиционных методов.

Статья "Predicting Porosity and Permeability in Reservoir Rocks Using Machine Learning Techniques" Суна, Чжао и Чжанга, опубликованная в журнале Energy Exploration & Exploitation, рассматривает применение методов машинного обучения для прогнозирования пористости и проницаемости в горных породах месторождений.

Авторы обсуждают, как машинное обучение может помочь в предсказании свойств горных пород месторождений, которые являются ключевыми факторами при оценке и разработке нефтегазовых месторождений. Они также описывают различные методы машинного обучения, которые могут быть использованы для анализа геологических данных, такие как нейронные сети, случайный лес и метод опорных векторов.

Основной задачей исследования было разработать модель для прогнозирования добычи нефти и газа на основе геологических данных. Для этого авторы использовали нейронную сеть, обученную на исторических данных о производительности скважин. В результате была разработана модель, которая позволила предсказывать добычу нефти и газа с высокой точностью. Кроме того, авторы рассмотрели возможность использования нейронных сетей для прогнозирования других геологических параметров, таких как структура горных пород, качество нефти и газа, проницаемость и т.д. Для этого были использованы данные, полученные с помощью бурения скважин и геофизических исследований.

В итоге, авторы пришли к выводу, что нейронные сети могут быть эффективным инструментом для прогнозирования геологических параметров месторождений и оптимизации процесса добычи нефти и газа. Однако, необходимо учитывать особенности каждого конкретного месторождения и подбирать соответствующие модели и параметры обучения нейронных сетей.

Машинное обучение также используется в геологии для анализа и интерпретации данных. Например, исследователи используют алгоритмы машинного обучения для классификации и интерпретации сейсмических данных. Алгоритмы машинного обучения также могут быть использованы для анализа геохимических данных и выявления закономерностей, которые нелегко увидеть человеческому глазу, таких как распределение различных типов минералов. Это позволяет геологам принимать более обоснованные решения о составе и структуре геологических систем.

Следует отметить, что использование машинного обучения в геологии становится все более популярным и потенциально способно революционизировать подход геологов к изучению геологических систем. Благодаря своей способности анализировать большие объемы данных и делать прогнозы быстро и точно, машинное обучение обладает потенциалом значительно улучшить наше понимание недр Земли.

1.3.2 Преимущества машинного обучения (самообучающихся алгоритмов)

Машинное обучение является одним из наиболее перспективных направлений в области искусственного интеллекта. Оно позволяет создавать самообучающиеся алгоритмы, которые могут адаптироваться к новым данным и изменять свою модель поведения на основе этих данных. Это дает им ряд преимуществ по сравнению с традиционными алгоритмами, которые работают только в рамках заранее заданной модели.

Одним из главных преимуществ машинного обучения является его адаптивность. Самообучающиеся алгоритмы могут адаптироваться к новым данным и изменять свою модель поведения на основе этих данных. Это позволяет им быть более гибкими и эффективными, чем традиционные алгоритмы.

Машинное обучение может использоваться в различных областях, таких как медицина, финансы, производство, геология и т.д. Самообучающиеся алгоритмы могут обрабатывать и анализировать большие объемы данных, что позволяет выявлять скрытые зависимости и тренды в данных, которые могут быть незаметны для человека.

Самообучающиеся алгоритмы могут обучаться на большом количестве данных и использовать эти данные для создания более точных моделей. Это позволяет им быть более точными, чем традиционные алгоритмы, которые могут иметь ограниченную точность из-за недостатка данных.

Некоторые компании и организации опубликовали статистику по времени, затрачиваемому на обработку геохимических данных. Например, в исследовании, опубликованном в журнале *Exploration Geophysics*, ученые из компании Petrobras описали процесс обработки геохимических данных для поиска нефтяных месторождений. Они отмечают, что в зависимости от объема данных и сложности их обработки, процесс может занимать от нескольких дней до нескольких месяцев.

Также, другие исследования показывают, что использование методов машинного обучения может существенно сократить время обработки геохимических данных. Например, в статье "Application of machine learning to geochemical data for mineral exploration: A case study from Dongping gold deposit, (China" Li, X., 2020). авторы отмечают, что использование методов машинного обучения сократило время обработки данных в 2-3 раза по сравнению с традиционными методами.

Машинное обучение позволяет автоматизировать многие задачи, которые ранее требовали участия человека. Например, классификация изображений или распознавание речи могут быть выполнены с помощью самообучающихся алгоритмов без участия человека.

Самообучающиеся алгоритмы могут оптимизировать свою модель на основе обратной связи от данных. Это означает, что они могут учиться на своих ошибках и улучшать свою модель с каждым новым примером обучения. Это может привести к более эффективному использованию ресурсов и улучшению результатов.

1.3.3 Недостатки и сложности машинного обучения

Хотя машинное обучение является мощным инструментом, у него есть и ряд недостатков и сложностей, которые следует учитывать при его применении. Некоторые из основных недостатков и сложностей машинного обучения включают в себя:

1. Необходимость больших объемов данных

Для создания точных моделей машинного обучения требуются большие объемы данных. В некоторых областях, таких как медицина и финансы, доступ к данным может быть ограничен из-за конфиденциальности или других причин. Это может затруднить обучение моделей машинного обучения в этих областях.

2. Неполные данные и шум

Нередко данные, с которыми работает модель машинного обучения, содержат ошибки, выбросы или пропущенные значения. Это может привести к неточности модели и ухудшению качества прогнозов. Неполные или некачественные данные могут также привести к переобучению модели.

3. Переобучение

Переобучение - это явление, когда модель слишком точно подстраивается под тренировочные данные и теряет способность к обобщению на новые данные. Это может привести к неправильным прогнозам и низкой точности модели на новых данных.

4. Интерпретируемость моделей

Некоторые модели машинного обучения, такие как нейронные сети, могут быть очень сложными и трудно интерпретируемыми. Это может затруднить объяснение принятых моделью решений и усложнить процесс анализа и отладки.

5. Этические и социальные вопросы

Машинное обучение может иметь этические и социальные последствия. Например, модели машинного обучения могут учитывать предвзятость данных и принимать неправильные решения, которые могут привести к дискриминации или другим негативным последствиям. Также возникают вопросы, связанные с конфиденциальностью и безопасностью данных.

6. Необходимость высококвалифицированных специалистов

Машинное обучение требует высококвалифицированных специалистов, которых не всегда легко найти и привлечь. Для создания моделей машинного обучения требуется знание математики, статистики, программирования и понимание специфики предметной области, на которую направлена модель. Кроме того, требуется опыт работы с инструментами и библиотеками машинного обучения.

7. Затраты на оборудование и вычислительные ресурсы

Обучение моделей машинного обучения требует значительных вычислительных ресурсов. Это может быть затратно для компаний и организаций, которые не имеют доступа к достаточным вычислительным мощностям. Кроме того, для обучения моделей требуются специализированные инструменты и программное обеспечение, которые также могут быть дорогостоящими.

8. Низкая устойчивость к изменениям

Модели машинного обучения могут быть чувствительны к изменениям в данных, на которых они были обучены. Это может привести к необходимости переобучения модели при изменении условий или вводе новых данных. Кроме того, если модель используется в долгосрочной перспективе, то необходимо учитывать изменения в окружающей среде и обновлять модель, чтобы она продолжала давать точные прогнозы.

Таким образом, машинное обучение имеет некоторые ограничения и сложности, которые могут затруднить его применение в некоторых областях. Однако, с правильным подходом, машинное обучение может стать мощным инструментом для решения различных задач и принятия решений в реальном мире.

1.4 Пути развития технологий машинного обучения в поисках полезных ископаемых

Искусственный интеллект активно внедряется в производство и практику, стремительно увеличивается применение искусственных нейронных сетей в геологоразведке и нефтяной отрасли. Объяснение геологических данных включает анализ таких геофизических исследований как сейсморазведка, электроразведка, каротаж, определение коллекторских свойств породы, геохимия. Для успешного внедрения необходима четкая последовательность, где последующие шаги опираются на результаты предыдущих.

В использовании алгоритмов для добычи углеводородов снижаются капитальные затраты и повышается эффективность выполняемых задач. Так же с помощью искусственного интеллекта можно спрогнозировать цены на сырье на фондовых рынках в ближайшие годы. Алгоритмы машинного обучения могут быть использованы для разработки прогностических моделей, которые можно использовать для определения областей, наиболее вероятно содержащих полезные ископаемые. Это может значительно сократить время и затраты, связанные с разведкой полезных ископаемых. Алгоритмы могут автоматизировать многие задачи, связанные с разведкой полезных ископаемых, освобождая геологов и других специалистов, чтобы сосредоточиться на более сложных задачах.

Это лишь некоторые из способов, с помощью которых разрабатываются технологии машинного обучения для поиска полезных ископаемых. Поскольку эта область продолжает развиваться, вполне вероятно, что машинное обучение будет играть все более важную роль в открытии новых месторождений полезных ископаемых.

Заключение.

Одним из наиболее важных преимуществ машинного обучения является возможность обработки большого объема данных, которые собираются и генерируются в различных сферах деятельности. Алгоритмы машинного обучения позволяют анализировать эти данные, выделять из них закономерности и строить прогнозы на основе полученных результатов.

Однако, при применении алгоритмов машинного обучения необходимо учитывать, что это лишь инструмент, который не может заменить человеческий опыт и экспертизу. Также необходимо помнить о том, что любой алгоритм машинного обучения может давать неточные результаты в случае, если входные данные недостаточно качественные или неполные.

В целом, применение алгоритмов машинного обучения является важным шагом в развитии современных технологий, и его значение будет только расти в будущем. Однако, для того чтобы максимально использовать потенциал машинного обучения, необходимо продолжать исследования и развитие в этой области, а также обучать специалистов, которые

смогут правильно применять алгоритмы машинного обучения в различных сферах деятельности.

2 ГЛАВА. ГЕОЛОГИЯ КАЙЭНМЫВААМСКОГО РУДОПРОЯВЛЕНИЯ

Глава 2 посвящена изучению Кайэнмываамского рудопроявления, которое было выделено в пределах Центрально-Чукотского сектора Охотско-Чукотского вулканогенного пояса. Для данного участка характерны эпитермальные золото-серебряные месторождения, что делает его особенно интересным для изучения в геологическом аспекте.

В данной главе будет проведен анализ геологической структуры участка и изучены основные геологические процессы, которые привели к формированию эпитермальных золото-серебряных месторождений. Будут рассмотрены особенности геологической структуры, горных пород, минерального состава и структуры месторождений на данном участке. Цель данной главы заключается в более глубоком изучении геологических особенностей и месторождений на данном участке.

2.1 Физико-географические сведения о районе исследований

Охотско-Чукотский вулканогенный пояс (ОЧВП) занимает площадь примерно 500 тысяч квадратных километров и простирается более чем на 3 тысячи километров вдоль Тихоокеанского сектора России (Лебедев, 1987), от реки Киран на юго-западе до Чукотского полуострова на северо-востоке. Впервые эту структуру выделил в 30-е годы С.В. Обручев.

ОЧВП — один из крупнейших структурных элементов Северо-Востока Азии — представлен мощными (до 5000 м) субаэральными вулканогенными образованиями. Исследования стратиграфии ОЧВП проводились геологами различных организаций, включая Севвостокгеологию, Дальгеологию, ПГО Аэрогеология, СВКНИИ ДВНЦ АН СССР и другие. Кратко они изложены при описании отдельных секторов пояса, а до 1966 г. — в сводной работе по геологии Северо-Востока СССР (Геология СССР, 1970).

Многие ученые отмечают сходство ОЧВП с вулканическим поясом Анд и указывают на его связь с зоной субдукции океанской литосферы.

В самом центре Чукотского автономного округа располагается Кайэнмываамское рудное поле на расстоянии 330 км от города Анадырь, 240 км от поселка Марково, 350 км от города Билибино, 350 км от города Певек и 130 км от рудника Купол (Прокофьев и др., 2019).

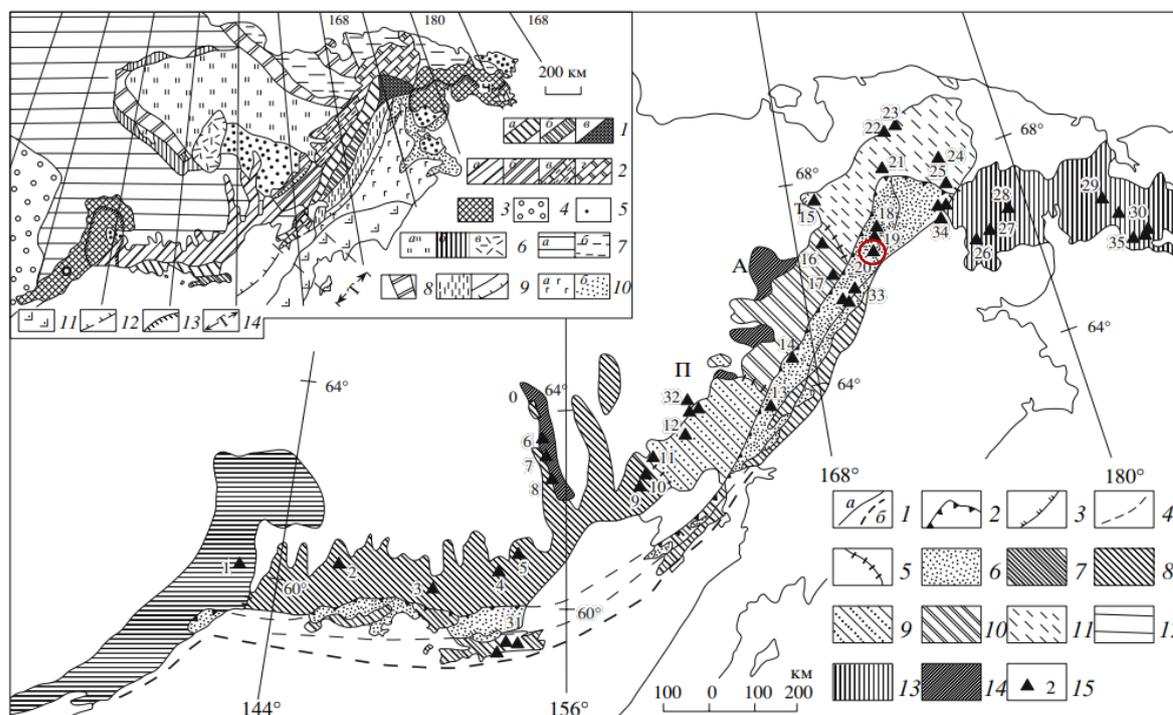


Рис. 2.1 Схема районирования ОЧВП и размещения в нем золото-серебряных месторождений. 1- обобщенная граница распространения магматических образований ОЧВП: альб-сантон (а - на суше, б - в акватории); 2 - граница внешней и внутренней зон ОЧВП; 3 - граница подзон грабен-прогибов и магматогенных поднятий (штрихи

ориентированы в сторону подзоны грабен-прогибов); 4 - продолжение границ 2, 3 в акватории; 5 - границы секторов внешней зоны и фланговых зон; 6 - вулканиды внутренней зоны; 7 - магматогенные поднятия ОЧВП, в которых обнажены структурно-формационные комплексы Тайгоносской вулканической дуги: верхний палеозой-низы альба; 8-11 - внешняя зона ОЧВП (секторы: 8 - Охотский, 9 - Пенжинский, 10 - Анадырский, 11 - Центрально-Чукотский); 12, 13 - фланговые зоны (12 - Западно-Охотская, 13 - Восточно-Чукотская); 14 - позднеорогенные (конец баррема-начало альба) структуры мезозой (О - Омсукчанский грабен, П - Верхне-Пенжинская, У - Умкувеевская, А - Айнахургенская, Т - Ты-тыльвеевская впадины); 15 - золото-серебряные месторождения (1 - Хаканджа, 2 - Бургаглыкан, 3 - Карамкен, 4 - Нявленга, 5 - Джульетта, 6 - Арылах, 7 - Лунное, 8 - Дукал, 9 - Сопка Кварцевая, 10 - Ирбычан, 11 - Ороч, 12 - Ке-гали, 13 - Сергеевское, 14 - Иргувеев, 15 - Двойной, 16 - Купол, 17 - Горностаевый, 18 - Энмываам, 19 - Арыкваам, **20 - Кайэнмываам**, 21 - Кытлатап, 22 - Промежуточный, 23 - Сопка Рудная, 24 - Телевеев, 25 - Провальные озера, 26 - Жильное, 27 - Валунистое, 28 - Тэркней, 29 - Коррида, 30 - Пепенвеев; 31-35 - медно-порфиновые группы месторождений: 31 - Конипьягинская группа, 32 - Эргувеевская группа, 33 - Ольховская группа, 34 - Танюрерская группа, 35 - Провиденская группа).
Источник: [А.А.Сидоров 2009].

2.2 Рудопроявление Кайэнмываам

Чукотский автономный округ (ЧАО) представляет собой наиболее богатую золотом и серебром северо-восточную часть Арктической зоны России. Он охватывает весь Чукотский полуостров и несколько островов. В 1955 году А. А. Сидоровым были обнаружены первые золото-серебряные месторождения эпитеpмального типа в ЧАО и на Северо-Востоке России. В последующие 65 лет в округе были выявлены многочисленные эпитеpмальные золото-серебряные месторождения и рудопроявления, приуроченные к меловому окраинно-континентальному Охотско-Чукотскому (ОЧВП) и позднеюрско-раннемеловому, островодужному Олойскому (ОВП) вулканическим поясам, а также постколизиионным, раннемеловым вулканическим впадинам. На территории ОЧВП эпитеpмальные Au-Ag месторождения обнаружены в терригенных флишевых толщах и останках вулканических покровов (Волков и др., 2021).

Кайэнмываам состоящей из вулканических образований среднего и кислого состава и относится к Энмываамской вулкано-тектонической депрессии.

Таблица 2.1
Таблица свит и формаций Кайэнмываама. [Составлено автором].

Свиты.	Формации.
Эмунеретская	риолит-базальтовая
Эргываамская	риолитовая
Энмываамская	базальт-трахибазальтовая

Главными рудоконтролирующими структурами на территории являются Анадырский разлом северо-восточного простирания и оперяющие его северо-западные разломы (Хорин Г.И., Соколов С.В., 1985).

На данной площади известны проявления и пункты минерализации различных металлов, включая золото, серебро, медь, молибден и полиметаллы. Среди наиболее исследованных объектов на территории находятся месторождение Арыкэваам и проявление Тэлевеем.

Кратко можно охарактеризовать вулканические пояса и эпитеpмальные месторождения Au-Ag на Чукотке.

На Дальнем Востоке России сосредоточен ряд промышленных руд, средних и крупных Au - Ag эпитеpмальных месторождений. Данные руды относятся к LS-типу, где доминирует кварцево-жильное оруденение. Богатые золотом руды могут простираться более чем на

километр и иметь мощность десятки метров, содержание золота будет варьироваться от 10 до 20 тонн. Экономически выгодно разрабатывать месторождения, даже в том числе если они перекрыты сотнями метров массивами вулканических пород.

Меловые, окраинно-континентальные Охотско-Чукотский (ОЧВП) и позднеюрско-раннемеловые, островодужные Олойский (ОВП) вулканические пояса, а также постколизийные, раннемеловые вулканические впадины, принадлежащие Чукотскому полуострову, содержат многочисленные эпитермальные золото-серебряные месторождения и рудопроявления. Размещение месторождений контролируется вулканотектоническими депрессиями, кальдерами и вулканическими куполами (Волков и др., 2021).

Были описаны случаи когда рудные кварцево-жильные тела формировали помимо протяженных жильных зон, штокверковые линейные и изометрические залежи. В большинстве месторождений эпитермального типа наблюдаются сложные, комбинационные брекчиево-крустификационные текстуры. Жильные руды характеризуются распределением рудных вкрапленников в виде ритмично- и колломорфно-полосчатых текстур с тонким характером, часто с полосчатыми скоплениями рудных минералов (гингуро). От низкой до высокой пробностью может варьироваться самородное золото. Акантит – минерал серебра (Ag_2S) часто встречается в рудах и достигает максимальных концентраций в рудных телах с многократным перераспределением веществ, таких как Коррида и Валунистое (Волков и др., 2022).

Незначительное количество Ag-Au эпитермальных месторождений находящихся на Чукотке содержат селен и теллур-содержащие минералы, таким образом основываясь на минералогических исследованиях проявления и месторождения можно отнести к Se-типу. Например: обогащённые теллуридами, отдельные участки месторождений Телевеем и Сентябрьское расположены в удалении от главных рудных тел.

Для большего количества Ag-Au эпитермальных месторождений имеют схожие характеристики: преобладание углекислоты над метаном, низкие и средние температуры, пониженная солёность флюида.

Повышенные значения температуры и солёности флюидов отмечаются в Центральном-Чукотском секторе ОЧВП и Баимской зоне ОВП, в то время как пониженные значения наблюдаются во Восточно-Чукотской фланговой зоне и внутренней зоне ОЧВП.

Содержание сульфатов в флюидах внутренней зоны ОЧВП в среднем в два раза выше, чем во Восточно-Чукотской фланговой зоне ОЧВП. Также флюиды внутренней зоны ОЧВП содержат больше углекислоты и бикарбонат-ионов по сравнению с флюидами Восточно-Чукотской фланговой зоны ОЧВП. Термобарогеохимические данные указывают на то, что месторождение Весеннее в Баимской зоне относится к среднесульфидизированному классу, а

также предсказывают возможное открытие высокосульфидизированных эпитермальных месторождений во внутренней зоне ОЧВП.

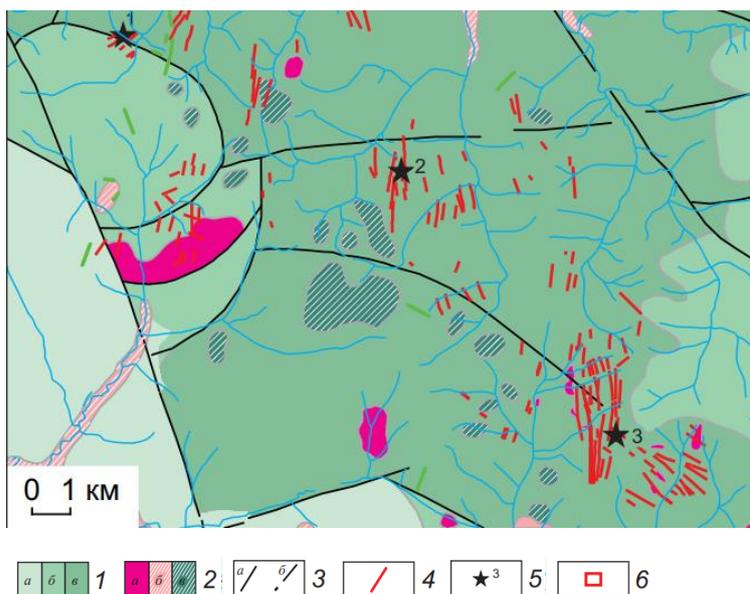


Рис. 2.2 Схематическая геологическая карта Кайэнмываамского рудного поля: 1 – верхнемеловые стратифицированные вулканические породы ОЧВП (от молодых до более древних свит); 2 – диориты (а), субвулканические риолиты (б), базальты (в); 3 – установленные (а), предполагаемые (б) разрывные нарушения; 4 – кварц-карбонатные жилы; 5 – перспективные участки (1 – Левый Кайэнмываам, 2 – Средний Кайэнмываам, 3 – Телевеем); 6 – район работ. Источник: [Прокофьев, 2019]

Экономически район практически не освоен и скуден. Транспортное сообщение летом возможно вертолётom и гусеничным транспортом; в зимний период возможно движение и осуществление основных грузоперевозок колёсным транспортом только по автозимникам.

2.3 Магматизм и метаморфизм Кайэнмываам

По одной из теорий Охотско-Чукотский вулканический пояс (ОЧВП) является крупнейшим крупным вулканическим всех окраинно-континентальных поясов фанерозоя, также крупным поясом Восточной Азии. Общая протяженность этой структуры превышает 3000 км, а современная площадь распространения вулканитов (без учета эродированных) приближается к 400 тыс. км². В семидесятые годы ОЧВП рассматривался как яркий представитель «краевых вулканических поясов», а после, с утверждением плейт-тектонической парадигмы, как надсубдукционный пояс андийского типа (Тихомиров и др., 2020).

За последние десятилетия и первых работ о представлении геологии и возрасте ОЧВП сильно изменяются (Устиев, 1959, 1963). В шестидесятых годах некоторые геологи полагали, что формирование пояса происходило на протяжении всего мелового периода, (Устиев, 1963), в то время как другие указывали на возраст формирования от альба до сеномана. В восьмидесятые годы было предложено исключить вулканогенные толщи верхней юры и нижнего мела, которые формировались в субаквальной обстановке и отделены от молодых вулканитов поверхностью несогласия. Удско-Мургальская вулканическая дуга (или вулканический пояс) связывает связывающая верхнеюрско-нижнемеловые вулканиты и коагматичные им интрузивы. Согласно современным точкам зрения и датированию магматических пород формирование ОЧВП приходится на интервал 106—74 млн лет.

Важный вклад в решение вышеперечисленных вопросов внесли результаты работ, проведенных в начале 21 века в северной части Удско-Мургальского пояса, в бассейне реки Убиенка (левого притока р. Анадырь). Исследование включало структурные наблюдения, отбор проб и комплекс лабораторных процедур, такие как: анализ валовых проб на содержание петрогенных и примесных элементов, петрографическое изучение пород, оценку U-Pb возраста цирконов и ⁴⁰Ar/³⁹Ar возраста биотита (Акинин, 2011). Нехватка данных датирования магматических пород, оставляет открытыми вопросы о степени сходства УМВП и ОЧВП, а так же времени формирования УМВП.

Если рассматривать весь ОЧВП, устанавливается прерывистый характер надсубдукционного магматизма от среднего альба до среднего кампана (106-76 млн. лет). По латерали пояса вулканизм асинхронен. Отмечается несколько пиков вулканизма с пиками около 105, 100, 96, 92.5, 87, 82 и 77 млн. лет. Коньяк-сантонские пики относятся к максимально объемным стадиям среднего и позднего циклов кислого вулканизма. Реконструируется спад\перерыв магматической активности в конце сеномана-начале турона (95-91 Ma). Завершают извержения ОЧВП платобазальты с возрастом около 76-78 млн. лет,

которые фиксируют изменение геодинамической обстановки с фронтальной субдукции на режим трансформной окраины с локальным растяжением в поперечных к сдвигу зонах. В основании ОЧВП вскрыты вулканические породы более древних, верхнеюрско-раннемеловых вулканических цепей и дуг, таких как Удско-Мургальской, Балыгычано-Сугойской, Нутесынский, Аргытский и др. Геометрия и поперечное простирание последних по отношению к меловой континентальной палеоокраине (и общему простиранию ОЧВП), далеко вдающиеся в континент линейные структуры выдают возможное происхождение их в результате растяжений/расколов вдоль трансформной окраины. Эволюция мелового магматизма ОЧВП на Чукотке особенно важна для нашего начального понимания того, как надсубдукционный и синсдвиговый магматизм менялся и мигрировал в пространстве и времени, как это коррелирует с глобальным взаимодействием плит, которое привело к формированию современной неактивной арктической окраины и активной тихоокеанской окраины в северной Пацифике.

2.4 Условия формирования золото-серебряной эпипермальной минерализации

Рудные тела месторождений представлены кварц-адуляровыми жилами, иногда собранными кулисообразно в протяженные жильные зоны, а также изометричными и линейными штокверковыми залежами. Для рудоносных жил большинства месторождений характерны сложные, комбинационные брекчиевокрустификационные текстуры. Вулканотектонические депрессии, кальдеры и вулканические купола контролируют размещение месторождений. В жильных рудах широко развиты ритмично- и колломорфно-полосчатая текстуры, обладающие преимущественно тонким характером распределения рудных вкрапленников, нередко с полосчатыми скоплениями рудных минералов (гингуро).

Проявление Кайенмываам (Ш-4-2) связано с зоной и включает кварцевые жилы (Волков и др., 2012), маломощные зоны прожилкования и хлоритовые метасоматиты. Промышленное оруденение месторождения связано преимущественно с кварцевыми и кварц-адуляровыми жилами и концентрируется в эндоконтактовых частях жил. Рудные минералы включают самородное золото, сульфосоли серебра, халькопирит, галенит и сфалерит. Среднее золото-серебряное отношение составляет 1:11. Оценка прогнозных ресурсов рудопроявления Кайенмываам по категории P2 составляет 0,36 млн т руды, 9,6 т золота, 76,5 т серебра (Чукотприродресурсы, 1998).

Золото-серебряное оруденение связано с наклонными субмеридиональными кварцевыми жилами, а также кварцевыми и кварц-карбонатными брекчиями и прожилково-жильными зонами. Мощность этих зон составляет от 1,6 до 16 м, протяженность - от 50 до 900 м, а вертикальный размах - от 25 до 180 м. Они развиты вдоль тектонических нарушений среди метасоматитов на основе вулканитов нижеэмунеретской подсвиты.

В рамках исследований было обнаружено 6 рудоносных зон с содержанием золота от 1 до 3,9 г/т и серебра от 28,7 до 393,5 г/т. Месторождение относится к типу эпипермальных золото-серебряных рудных формаций с преобладанием золота и кварца и наличием убогосульфидных минералов. Доля свободного и связанного самородного золота примерно одинакова.

Крупность золота относится преимущественно к классу менее 0,1 мм.

2.5 Вмещающие породы Кайэнмываам

Кайэнмываамское рудное поле расположено в пределах Маюлервеевской вулканоструктуры Охотско-Чукотского вулканогенного пояса (ОЧВП).

Месторождения наиболее изученного Кайэнмываамского рудного района находятся соответственно на южном и северном склонах (крыльях) реликтового вулканического поднятия между прогибами (Волков и др., 2020). Ядро поднятия представлено массивными порфирокластическими риолитовыми игнимбритами пыкарваамской свиты, протяженность которого достигает 10 км. Окружающие его области содержат породы коэквиуньской (андезитовая формация) и эргываамской (риолитовая формация) свиты. В пределах поднятия также широко развиты аргиллизит-вторичнокварцитовые шляпы [Белоусов и др., 2020].

Месторождение Арыкэваам сосредоточено в купольной структуре на северном замыкании поднятия, в то время как Кайэнмываамское и Тэлевемское рудопоявления связаны сближенными отрицательными вулканотектоническими структурами в центральной и южной части поднятия. Потенциальные рудные тела представлены жилами мощностью до 20 м (в раздувах) и длиной до 300 м. В пределах внутренней зоны ОЧВП прогнозируется открытие высокосульфидизированного класса эпитермальных Золото-серебряных месторождений (Волков и др., 2021).

В геологическом строении участвуют стратифицированные игнимбриты кислого состава, андезибазальты, андезиты, базальты, туфы среднего-кислого состава, прорванные дайками, штоками трахиандезитов, андезитов, субщелочных диоритов и монзонитов. распространение субвулканических тел, кварцевых жил и полей метасоматитов соответствует радиально-кольцевому рисунку трещиноватости вулканоструктуры. На территории рудного поля широко распространены эпитермальные кварцевые жилы с минерализацией золота и серебра. Выделяются три перспективных участка с максимальной концентрацией жил: Левый Кайэнмываам, Средний (Правый) Кайэнмываам и Телевеем.

Исследования геологической структуры Кайэнмываамского рудного поля позволили выявить взаимосвязь между распределением месторождений и региональными структурами. Размещение месторождений на южных и северных склонах реликтового вулканического поднятия указывает на влияние тектонических процессов. Купольная структура на северном конце поднятия, где находится месторождение Арыкэваам, представляет особый интерес для дальнейших исследований.

В ходе проведения поисковых работ Анадырской геологоразведочной экспедицией (1989–1994 гг.) на перечисленных участках было обнаружено множество эпитермальных

золото-серебряных жил и жильных зон, в которых выявлены несколько рудных столбов. Самый крупный из них имеет мощность 2,2 м и простирается по канавам на расстоянии 330 м, а по развалам рудного кварца – более 500 м. По геофизическим данным, вертикальный размах оруденения превышает 150 м. Жилы характеризуются полосчатыми текстурами, обусловленными чередованием мелкозернистого и крупнозернистого халцедоновидного кварца с мозаичной текстурой и друзовидного кварца. Обычно, мелкозернистые разности слагают внешние части жил, в то время как друзовидные разности характерны для их центральных частей. В жилах, помимо кварца, присутствуют адуляр, хлорит, каолинит, гипс, гидрослюда и кальцит.

В последнее время геохимические исследования в пределах рудного поля позволили охарактеризовать площадь рудопроявлений по вторичным ореолам рассеяния, а колонковое бурение подтвердило наличие богатого оруденения на глубине. Минералогические исследования подтвердили, что в ранней минеральной ассоциации присутствуют пирит, молибденит, халькопирит, галенит, сфалерит, борнит, тетраэдрит-теннантит, а в поздней (золото-серебряной) – минералы Au-Ag, пирсеит-полибазит, акантит, агвиларит, хемусит, тетрадимит, кабацулит и теллуриды, а также самородный теллур (Власов и др., 2016). Пробность золото-серебряных минералов составляет 550–655‰, в коре выветривания достигает 963‰. Типичными примесями являются медь (0,03–0,14%) и, реже, ртуть (до 0,18% по массе). На участке Телевеем основными концентраторами золота и серебра являются теллуриды (петцит, сильванит, гессит), в то время как самородные минералы Au-Ag имеют второстепенное значение.

В центральной части чукотского отрезка внутренней зоны Охотско-Чукотского вулканогенного пояса (ОЧВП) располагается рудопроявление "Комплексное". Оно находится на южном фланге крупного Кайэнмываамского вулканического поднятия, внутри которого простираются обширные поля риолитового вулканического комплекса (верхний альб) (Мальшева и др., 2012). В этом рудопроявлении обнаружены не только отложения пыкарваамской свиты, но и сопряженные с ними субвулканические куполообразные тела, штоки, силлы и дайки. Мощность пыкарваамской свиты варьирует от 800 до 1200 м. В нижней части (на глубине 500-800 м) преобладают игнимбриты дацитов, которые тонко чередуются с игнимбритами риолитов и редкими туфами и витрофирами дацитов и риолитов. В верхней части (на глубине 300 м) преобладают риолиты, иногда встречаются туфы риолитов и игнимбриты дацитов. Глубинные северо-восточные и северо-западные разломы служат тектоническими границами Кайэнмываамского поднятия.

Рудопроявление "Комплексное" охватывает площадь примерно 2 км² и представлено серией овальных аргиллизитовых "шляп" диаметром от 200 до 900 м. Эти "шляпы" состоят из поверхностных красных и светло-серых глин (Белоусов, 2020).

Таблица 2.2

Таблица концентраций элементов аргиллитовых «шляп». [Составлено автором].

Элемент	Концентрация
Золото (Au)	От 0,05 до 0,1 г/т,
Серебро (Ag)	От 0,04 до 0,17 г/т
Медь (Cu)	От 10 до 35,8 г/т
Молибден (Mo)	От 7,56 до 37,68 г/т.

Заметно, что области глинистых отложений окружаются вытянутыми линзовидными телами вторичных кварцитов, которые, вероятно, образуют внешнюю границу кольцевой палеовулканогенной структуры. Эта структура контролирует центры эксплозивной и гидротермальной активности (палеократеры) во внутренней зоне и представлена пологими куполами, состоящими из аргиллизитов в верхней части.

Рудообразующая гидротермальная система отражается в геохимических аномалиях: золота, серебра, меди, цинка, молибдена и связана с вторичными кварцитами и аргиллизитами «шляпами». На расстоянии нескольких км от аргиллизитовых "шляп" находятся эпитеральные кварцевые жилы, преимущественно северо-западного простирания, содержащие значительные количества золота (Au) и серебра (Ag) (Власов и др., 2016; Прокофьев и др., 2019).

В заключение можно отметить, что изучаемый участок имеет большую площадь и очень сложное геологическое строение. Проведение геохимического анализа на данном участке является важным и оправданным, так как предполагаемый промышленный тип месторождений – это, эпитеральные золото-серебряные месторождения, которые хорошо отражаются в геологических полях. Результаты геохимического анализа позволят определить наличие полезных ископаемых на участке, а также провести более детальное изучение геологического строения данного региона.

3 ГЛАВА. МЕТОДИКА СЕТИ КОХОНЕНА И СТАТИСТИЧЕСКИХ МЕТОДОВ ОБРАБОТКИ ГЕОХОМИЧЕСКИХ ДАННЫХ

В данной главе будут проведены два анализа: статистические методы и сеть Кохонена на основе геохимических данных по участку Кайэнмываам. Анализы будут направлены на выявление геохимических аномалий, которые могут указывать на наличие месторождений полезных ископаемых в изучаемом участке.

Первый анализ будет проведен с использованием статистических методов, которые позволяют получить пространственное распределение значений геохимических элементов на участке. Для проведения анализа будут использованы программные средства, такие как Surfer, QGIS, Statistica и другие.

Второй анализ будет основан на применении сети Кохонена, которая является одним из методов машинного обучения. Сеть Кохонена позволяет проводить кластерный анализ данных и выявлять скрытые зависимости между ними. В данной работе сеть Кохонена будет использоваться для анализа геохимических данных. Сеть Кохонена была построена при помощи программного обеспечения Loginom.

Сравнение результатов, полученных с помощью статистических методов и сети Кохонена, позволит оценить эффективность и применимость каждого из методов для выявления геологических аномалий и определения перспективных зон на изучаемом участке.

3.1 Методика работы: Сеть Кохонена

Сеть Кохонен, также известная как самоорганизующаяся карта (SOM), была изобретена Теуго Кохоненом, финским профессором информатики, в начале 1980-х годов. Он ввел концепцию SOM как способ организации многомерных данных в низкоразмерное представление.

Одно из первых практических применений сети Кохонен было в области распознавания речи. В конце 1980-х и начале 1990-х годов исследователи использовали SOM для классификации звуков речи в фонемы, которые являются основными единицами звука в языке (Kohonen, T., 1982, 1990, 1995). Это был один из ранних примеров того, как сеть Кохонена может быть использована для решения сложных проблем в таких областях, как машинное обучение и искусственный интеллект.

Сеть Кохонена является одним из методов машинного обучения, который может использоваться для кластеризации и визуализации геохимических данных.

Алгоритм начинается с случайной инициализации весов каждого узла. Затем для каждого входного вектора в наборе данных алгоритм вычисляет расстояние между входным вектором и каждым узлом в сетке. Узел, который находится ближе всего к входному вектору, называется "best-matching unit" (BMU).

Затем алгоритм корректирует веса BMU и соседних с ним узлов, причем величина корректировки уменьшается по мере увеличения расстояния от BMU. Этот процесс корректировки веса называется "тренировкой", и он продолжается в течение определенного количества итераций.

Как только обучение завершено, каждый входной вектор присваивается кластеру, соответствующему ближайшему к нему узлу. Результирующие кластеры обычно визуализируются на двумерной сетке узлов, причем аналогичные кластеры сгруппированы вместе.

Стоит подробнее рассмотреть все выше перечисленные параметры алгоритма сети Кохонена.

В процессе обучения алгоритм итеративно обновляет весовые векторы, связанные с каждым узлом в сетке, чтобы лучше представлять распределение входных данных. Правило обновления было введено Теуго Кохоненом (Teuvo Kohonen) в 1982 году в статье "Self-organized formation of topologically correct feature maps". Векторы веса обновляются в соответствии со следующей формулой:

$$w(t + 1) = w(t) + \eta(t)h(i, j, t)(x(t) - w(t))$$

где:

$w(t+1)$ - обновленный вектор веса для узла в момент времени $t+1$

$w(t)$ - вектор веса для узла в момент времени t

$\eta(t)$ - скорость обучения в момент времени t

$h(i,j,t)$ - функция окрестности в момент времени t , которая управляет масштабом обновления в сетке

$x(t)$ - входной вектор в момент времени t

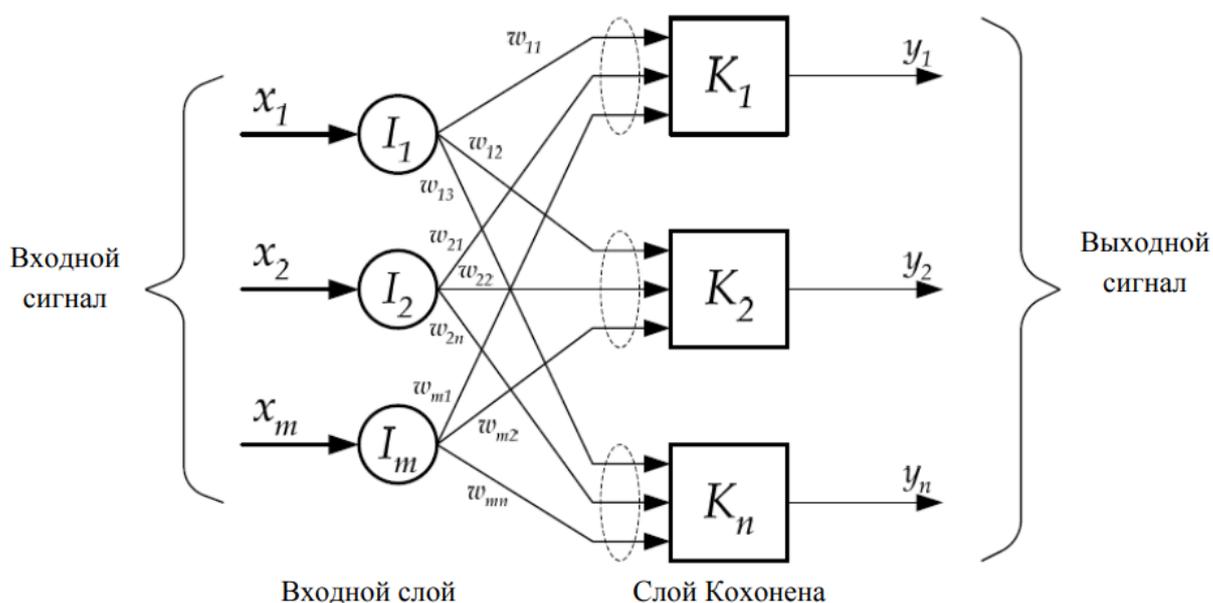


Рис. 3.1 Структура самоорганизующейся сети Кохонена: x_1, x_2, \dots, x_m – множество входных характеристик; w_{ij} – весовые коэффициенты сети; y_1, y_2, \dots, y_n – множество выходных характеристик, соответствующих n кластерам. Источник: [Ю.А. Келлер, 2014]

Функция окрестности $h(i, j, t)$ - это гауссова функция, центрированная вокруг единицы наилучшего соответствия (ВМУ), которая является узлом с вектором веса, ближайшим к входному вектору $x(t)$. Значение функции Гаусса уменьшается со временем, так что обновления становятся более локализованными по мере прохождения обучения.

Количество кластеров в алгоритме Кохонена равно количеству узлов в сетке. Каждый входной вектор присваивается кластеру, связанному с узлом, чей весовой вектор находится ближе всего к нему.

Расстояние между входным вектором x и узлом с вектором веса w обычно измеряется с использованием евклидова расстояния:

$$d(x, w) = \|x - w\|$$

где $\|x - w\|$ - евклидова норма разности между входным вектором и вектором веса.

Таким образом, алгоритм Кохонена включает в себя создание двумерной сетки узлов, обновление весовых векторов, связанных с каждым узлом, в ответ на входные данные и присвоение каждого входного вектора кластеру, связанному с узлом с ближайшим весовым вектором.

Математически евклидово расстояние между двумя векторами x и w , представляющими концентрации элементов в образце и вектор веса, связанный с узлом в алгоритме Кохонена, соответственно, вычисляется как:

$$d(x, w) = \sqrt{\sum (w_i - x_i)^2}$$

где:

i - варьируется по элементам

x_i - концентрация элемента i в выборке

w_i - вес, связанный с элементом i для узла.

Квадрат разности между концентрацией каждого элемента и его соответствующим весом в узле суммируется, и квадратный корень из этой суммы берется для получения евклидова расстояния.

В этом контексте расстояние до узла представляет сходство между концентрациями элементов в образце и вектором веса, связанным с узлом. Образцы с аналогичными концентрациями элементов назначаются узлам с аналогичными весовыми векторами, в то время как образцы с различными концентрациями элементов назначаются узлам с различными весовыми векторами.

Количество узлов или кластеров в сети Кохонена выбирается пользователем и представляет желаемый уровень детализации или разрешения в кластеризации. Узлы или кластеры можно рассматривать как представителей различных групп образцов, которые имеют сходные концентрации элементов. Как и описывалось ранее каждый узел или кластер связан с вектором веса, который представляет средние концентрации элементов в образцах, назначенных этому узлу или кластеру.

Математически назначение выборки x узлу j в сети Кохонена определяется путем нахождения узла с вектором веса, ближайшим к вектору выборки x . Это может быть выражено математически как:

$$j = \operatorname{argmin}(k) \|x - w_k\|$$

где:

k распространяется на все узлы сети,

w_k - вектор веса, связанный с узлом k ,

$\| \cdot \|$ представляет евклидово расстояние между вектором выборки x и вектором веса w_k , $argmin(k)$ указывает значение k , которое минимизирует расстояние.

В геохимии сетевой алгоритм Кохонена может быть использован для группировки образцов горных пород или почвы на основе данных об их элементном составе. Параметры сети, которые необходимо установить, включают количество узлов или кластеров в сети, скорость обучения.

Выбор оптимальных параметров сети важен для достижения точной и значимой кластеризации выборок. Если в сети слишком мало узлов, она может быть не в состоянии охватить всю сложность данных, в то время как если в сети слишком много узлов, она может перегружать данные и плохо обобщаться на новые выборки. Скорость обучения и размер окрестности влияют на скорость и стабильность процесса обучения, и их необходимо тщательно настраивать, чтобы избежать колебаний или сходимости к неоптимальным решениям.

Один из подходов к настройке параметров сети заключается в использовании комбинации метода проб и ошибок и визуализации. Пользователь может попробовать различные значения параметров сети и изучить результирующую кластеризацию, либо визуализируя весовые векторы, либо исследуя качество кластеризации с использованием таких показателей, как ошибка квантования или топографическая ошибка. Затем пользователь может настроить параметры сети на основе наблюдаемых результатов и повторять процесс до тех пор, пока не будет получена оптимальная кластеризация.

Другой подход заключается в использовании автоматизированных методов, таких как поиск по сетке или случайный поиск, для изучения пространства параметров и нахождения оптимальных параметров сети. Эти методы включают систематическое опробование большого количества комбинаций сетевых параметров и оценку результирующей кластеризации и могут быть трудоемкими с точки зрения вычислений, но могут быть более эффективными в поиске оптимального решения, чем ручная настройка.

Таким образом, настройка параметров сети важна для достижения точной и значимой кластеризации геохимических данных с использованием алгоритма Кохонена. Выбор параметров может повлиять на скорость, стабильность и качество процесса обучения, и его следует тщательно настраивать, используя комбинацию проб и ошибок и автоматизированных методов.

3.1.1 Практическое применение сети Кохонена для моделирования геохимических данных

Алгоритм Кохонена является одним из самоорганизующихся алгоритмов, используемых в машинном обучении. Для применения данного алгоритма были получены геохимические данные из лаборатории, в которой использовался метод ICP-MS. Было получено 6873 точки наблюдения, каждая из которых содержала информацию о 51 элементе.

ICP-MS (Inductively Coupled Plasma Mass Spectrometry) - это метод анализа химических элементов, основанный на измерении масс-спектров атомов и молекул в плазме, возникающей при воздействии индуктивно связанного плазменного разряда на анализируемый образец. Данный метод позволяет получить точные и достоверные результаты анализа для широкого диапазона элементов, что делает его полезным инструментом в геохимии.

Первоначально данные были обработаны в программе Excel, где были удалены контрольные точки, элементы с содержанием ниже порога определяемости (B, Be, Cd, Ce, Hf, P, In, La, Re, Ta, Te, Tl, U, W, Y). Пропуски и аномальные значения не заполнялись и не удалялись, так как это занимает много времени, а с помощью Loginom, данная задача выполняется очень быстро. Loginom - Low-code платформа для реализации всех аналитических процессов: от интеграции и подготовки данных до моделирования, развертывания и визуализации.

После предварительной обработки загружаю Excel файл в программу Loginom, предварительно конвертировать в иную базу данных не нужно. Loginom поддерживает файлы Excel, Loginom data, Tableau, Deductor Warehouse, различные базы данных и текстовые файлы. Из возможных сценариев в разделе предобработка выбираю «заполнение пропусков», в настройках пропущенные значения можно заменить средним или удалить столбец или строку. Далее в том же разделе добавляю на рабочее поле «редактирование выбросов». Обработчик предназначен для автоматической корректировки выбросов и экстремальных значений в наборах данных. Для каждого поля исходного набора данных критерии определения выбросов и экстремальных значений. Экстремальные значения задаются самостоятельно с помощью указания допустимого стандартного отклонения или интерквартильного размаха. Выбросы – значения содержания элемента, которого достаточно сильно преувеличивает среднее значения всех содержаний данного элемента. Экстремальные значения – содержание элемента настолько сильно отклоняются от типичных значений, что перестают соответствовать логике исследуемых процессов и явлений.

Завершив всю предобработку данных, запускаем алгоритм Кохонена. На панели управления «Самоорганизующаяся сеть», так же выводим на рабочий лист. В настройках назначаем элементы, которые будут служить входными данными для нашего алгоритма. При использовании сети Кохонена для кластеризации геохимических данных, входные данные представляются в виде многомерных векторов, каждый из которых соответствует содержанию 36 элементов в одной точке наблюдения.

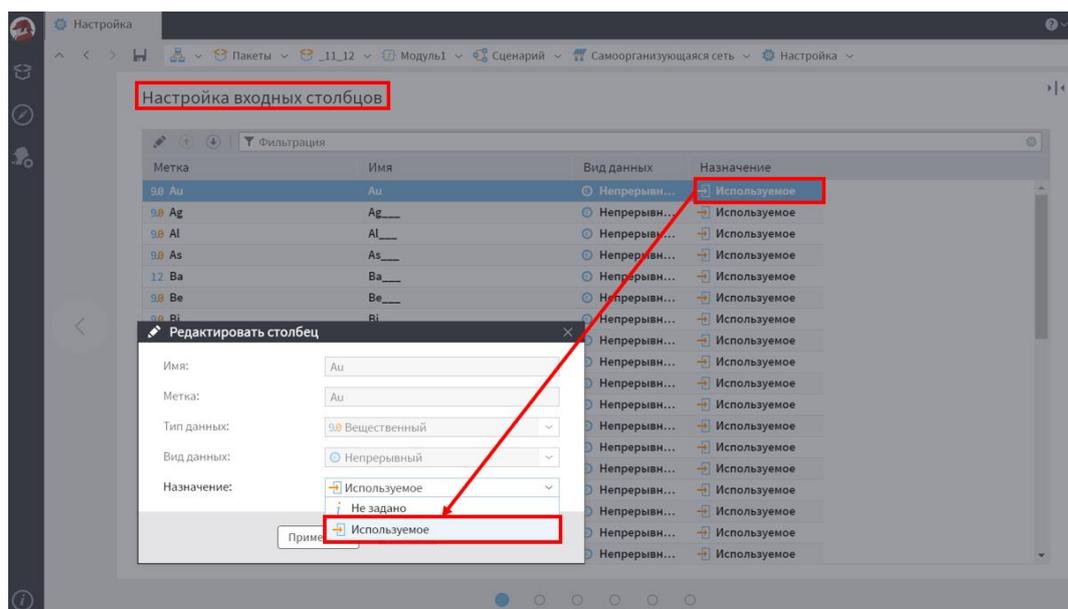


Рис. 3.2 Настройка входных столбцов. [Составлено автором на основе программного обеспечения Loginom].

Нажимаем на значок далее, где необходимо задать нормализатор – достаточно его активировать и программа сама подберет необходимые параметры для каждого входного столбца (пользователь может настроить вручную, любой нормализатор подходящий для выполнения конкретной задачи).

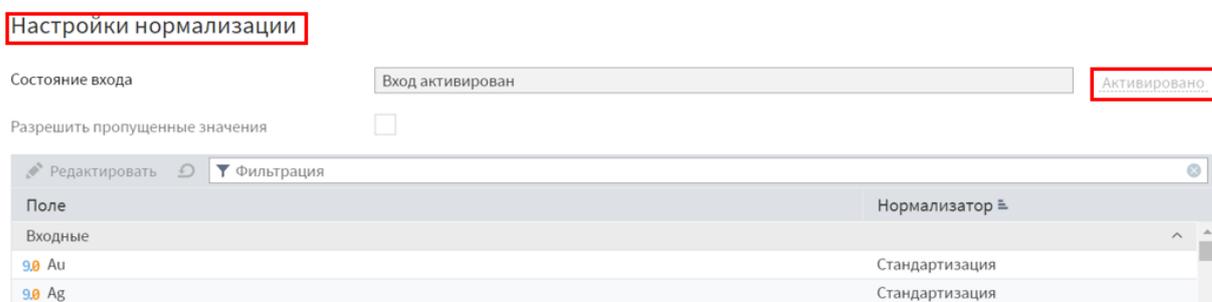


Рис. 3.3 Настройка нормализации. [Составлено автором на основе программного обеспечения Loginom].

Кластер - это группа нейронов, которые имеют схожие весовые вектора. Число кластеров также называют размерностью карты Кохонена.

Разница между сетью Кохонена с 1 кластером и сетью Кохонена с 100 кластерами заключается в количестве кластеров, которые используются для кластеризации входных данных.

Сеть Кохонена с 1 кластером означает, что все входные данные будут принадлежать одному кластеру. Это означает, что весовые вектора всех нейронов будут близки или совпадать, что может быть полезно в определенных ситуациях, например, когда необходимо произвести бинарную классификацию входных данных.

Сеть Кохонена с 100 кластерами означает, что входные данные будут разбиты на 100 кластеров, каждый из которых будет соответствовать группе входных данных с схожими характеристиками. Это может быть полезно в тех случаях, когда входные данные имеют множество различных характеристик или когда необходимо производить более детальную классификацию данных.

Количество кластеров на сетке может варьироваться, и оптимальное количество зависит от размерности входных данных. В общем случае, число кластеров должно быть меньше, чем количество точек данных. Для 36-мерных данных можно начать с 100-200 узлов на сетке, а затем настраивать это число для оптимальной кластеризации. Программа автоматически подберет количество оптимальных узлов (кластеров), но также этот процесс можно сделать вручную методом проб и ошибок.

Самоорганизующаяся нейронная сеть

Тип самоорганизующейся сети	Сеть Кохонена
Параметры сети	
Топология сети	2D прямоугольная
Размер сети по X	10
Размер сети по Y	10
Размер сети по Z	3

Рис. 3.4 Оптимальные параметры для имеющихся данных. [Составлено автором на основе программного обеспечения Loginom].

Важно отметить, что оптимальные параметры для сети Кохонена могут варьироваться в зависимости от данных и целей исследования. Чтобы найти оптимальные параметры, может

потребуется провести несколько экспериментов с различными параметрами и оценить качество кластеризации с помощью метрик, таких как SSE (sum of squared errors) или silhouette score.

Настроив оптимально все параметры, получаем итоговый результат для дальнейшей визуализации в QGIS.

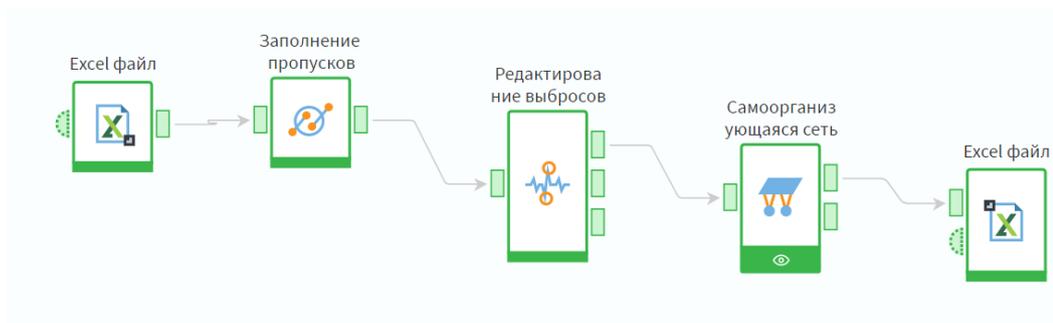


Рис. 3.5 Алгоритм в программе Loginom. [Составлено автором на основе программного обеспечения Loginom].

Выводы:

1. Применение метода сети Кохонена для геохимических данных позволяет классифицировать элементы в соответствии с их химическими свойствами. Для применения этого метода необходимо иметь набор данных, в котором указаны концентрации различных химических элементов.

2. Метод сети Кохонена для геохимических данных широко используется в геологии и геохимии для выделения геохимических аномалий и кластеризации элементов по геохимическим свойствам. Например, данный метод может быть использован для определения зон минерализации в рудных месторождениях или для анализа качества грунтов и воды.

3. Метод сети Кохонена для геохимических данных имеет следующие преимущества: он не требует заранее заданных классов и позволяет определить группы элементов на основе сходства их химических свойств. Однако, недостатком данного метода является то, что он не дает полной информации о причинах образования групп и не является универсальным методом для всех типов геохимических данных. Кроме того, результаты метода могут быть зависимы от выбора параметров и настроек сети.

Таким образом, использование сети Кохонена позволяет сократить размерность данных и выделить группы объектов с похожими характеристиками. Это может быть полезно для идентификации геохимических аномалий, выявления закономерностей и установления причинно-следственных связей.

3.2 Статистический метод обработки геохимических данных

Как и в предыдущем методе, была проведена предварительная обработка данных, полученных из лаборатории. Данные после обработки стали более структурированными. Удалены или заменены аномальные значения. Для выполнения задания использовались данные ISP. По данному методу было определено 51 элемент, по 6873 точкам наблюдения. Изучаемый объект находится на Чукотке в 100 км от месторождения Купол. Предполагаемая рудная площадь (Au-Ag минерализация), имеет размеры 55 на 20 км. Для выполнения факторного анализа были выбраны самые значимые элементы. Всего 36 элементов. В программе Statistica, был проведен корреляционный, факторный анализ и была посчитана элементарная статистика (Маштакова, И.Б., 2013).

При предварительной обработке данных были удалены контрольные точки, и элементы на уровне порога чувствительности (V, Be, Cd, Ce, Hf, P, In, La, Re, Ta, Te, Tl, U, W, Y). Для начала, нужно выполнить некоторые предварительные операции, такие как удаление выбросов и заполнение пропущенных значений. Затем можно вычислить различные статистические показатели для каждого элемента. Далее данные загружаются в программу Statistica, где был проверен закон распределения для каждого элемента. Рассмотрев их нормальное и логнормальное распределение, было прологарифмировано 32 элемента. Выбор между нормальным и логнормальным распределениями для описания концентраций геохимических элементов зависит от конкретных данных и целей анализа. В общем, если данные имеют более нормальное распределение, то нормальное распределение может быть более подходящим, а если данные имеют скошенное распределение, то логнормальное распределение может быть более подходящим. Также можно использовать статистические методы, такие как тест Колмогорова-Смирнова, чтобы определить, какое распределение лучше соответствует данным.

Было проведено статистическое исследование (ПРИЛОЖЕНИЕ 2), Статистические методы анализа данных играют важную роль в геохимических исследованиях, позволяя получить информацию о распределении и характеристиках геохимических элементов. Рассмотрим Au и Ag из данной таблицы:

Для золота валидных наблюдений было 6641. Среднее значение (Mean) равно 0,0024, что означает, что в среднем на каждую точку наблюдения приходится 0,0024 грамма золота. Геометрическое среднее (Geometric) составляет 0,0009, а гармоническое среднее (Harmonic) - 0,0007. Медиана (Median) равна 0,0005, что означает, что половина точек имеют значения золота меньше 0,0005, а другая половина - больше 0,0005. Мода (Mode) равна 0,0005, а частота ее появления (Frequency) составляет 4499. Сумма значений (Sum) золота равна 16

граммам. Минимальное (Min) значение золота - 0,0005 грамма, а максимальное (Max) - 0,0005 грамма. Стандартное отклонение (Std.Dev.) равно 0,33, а коэффициент асимметрии (Skewness) - 16,47. Стандартная ошибка (Std.Err.) составляет 0,01, а эксцесс (Kurtosis) - 372,62.

Для серебра также было 6641 валидных наблюдений. Среднее значение (Mean) равно 0,08, что означает, что в среднем на каждую точку наблюдения приходится 0,08 грамма серебра. Геометрическое среднее (Geometric) составляет 0,05, а гармоническое среднее (Harmonic) - 0,03. Медиана (Median) равна 0,05, что означает, что половина точек имеют значения серебра меньше 0,05, а другая половина - больше 0,05. Мода (Mode) равна 0,04, а частота ее появления (Frequency) составляет 889. Сумма значений (Sum) серебра равна 553 граммам. Минимальное (Min) значение серебра - 0,01 грамма, а максимальное (Max) - 6,44 грамма. Стандартное отклонение (Std.Dev.) равно 0,17, а коэффициент асимметрии (Skewness) - 16,36. Стандартная ошибка (Std.Err.) составляет 0,03, а эксцесс (Kurtosis) - 431.

Была вычислена корреляция между геохимическими элементами. (ПРИОЖЕНИЕ 3). (Леонтьев, В.М., 2013, Витязев, А.В., 2011, Гордеев, Е.В., 2015). Значения корреляции варьируются от -1 до 1. Коэффициент корреляции равный 1 означает идеальную положительную линейную зависимость, когда значения двух переменных растут вместе. Коэффициент корреляции равный -1 означает идеальную отрицательную линейную зависимость, когда значения двух переменных убывают вместе. Коэффициент корреляции равный 0 означает отсутствие линейной зависимости между переменными.

На основании данной таблицы корреляции можно сделать следующие выводы:

Корреляция между золотом (Au) и серебром (Ag) является средней положительной (0,37), что может указывать на наличие связи между содержанием этих элементов в геохимических образцах, так же Au имеет положительные связи с Висмут (Bi), Молибден (Mo), Мышьяк (As), Свинец (Pb), Цинк (Zn).

Алюминий (Al) имеет положительную корреляцию с кальцием (Ca), кобальтом (Co), хромом (Cr), железом (Fe), галлием (Ga), германием (Ge), ртутью (Hg), натрием (Na), свинцом (Pb), титаном (Ti), цинком (Zn) и цирконием (Zr). Значение коэффициента корреляции между алюминием (Al) и другими элементами изменяется в диапазоне от -0,18 до 0,90. Это говорит о том, что связь между алюминием (Al) и другими элементами может быть как слабой, так и сильной.

Были посчитаны факторные нагрузки. Факторные нагрузки - показывают степень взаимосвязи соответствующих переменных и факторов.

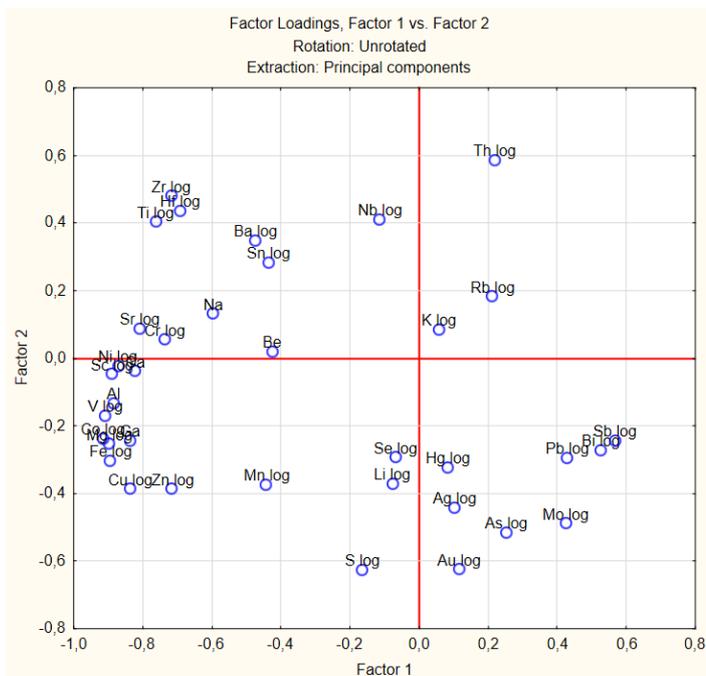


Рис. 3.6 Диаграмма нагрузок Фактора 1 от Фактора 2. [Составлено автором на основе программного обеспечения Statistica].

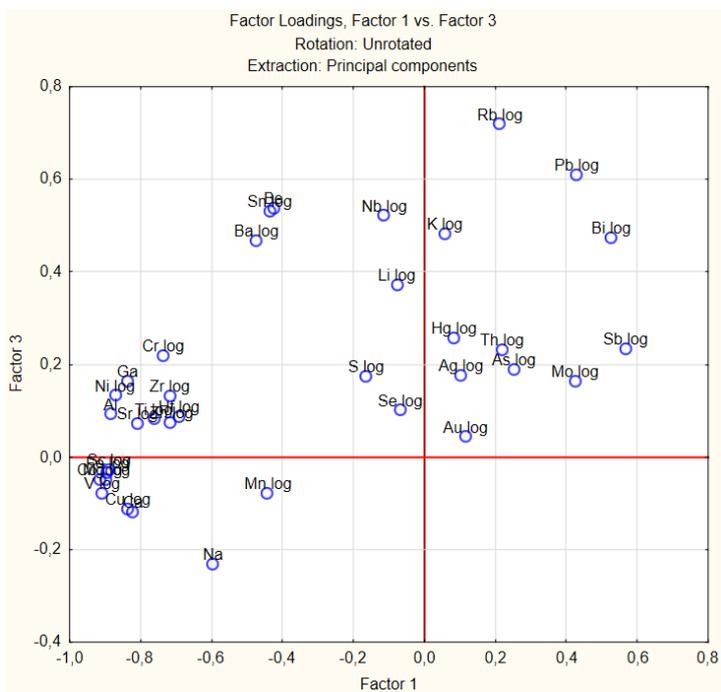


Рис. 3.7 Диаграмма нагрузок Фактора 1 от Фактора 3. [Составлено автором на основе программного обеспечения Statistica].

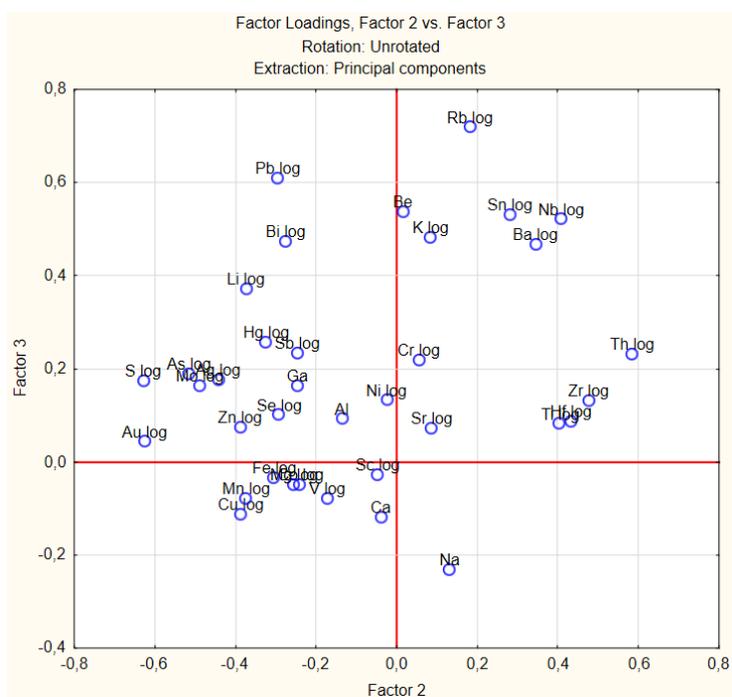


Рис. 3.8 Диаграмма нагрузок Фактора 2 от Фактора 3. [Составлено автором на основе программного обеспечения Statistica].

Таблица 3.1

Таблица нагрузок факторов. [Составлено автором на основе программного обеспечения Statistica].

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
	Au log	0,12	-0,62	0,04	0,02
Ag log	0,10	-0,44	0,18	-0,01	0,10
Al	-0,88	-0,13	0,09	-0,08	-0,09
As log	0,25	-0,52	0,19	0,07	0,34
Ba log	-0,47	0,35	0,47	-0,21	-0,21
Bi log	0,53	-0,27	0,47	-0,20	0,14
Be	-0,42	0,02	0,54	0,06	-0,20
Ca	-0,82	-0,04	-0,12	0,29	-0,06
Co log	-0,91	-0,24	-0,05	0,08	0,08
Cr log	-0,73	0,06	0,22	-0,29	0,00
Cu log	-0,84	-0,39	-0,11	-0,06	0,12
Fe log	-0,89	-0,30	-0,03	-0,01	0,08
Hf log	-0,69	0,43	0,09	-0,12	0,37
Ga	-0,84	-0,25	0,16	-0,11	0,04
Hg log	0,08	-0,32	0,26	-0,43	-0,17
K log	0,06	0,08	0,48	0,65	-0,27
Li log	-0,07	-0,37	0,37	0,23	0,27
Mg log	-0,90	-0,25	-0,05	0,14	-0,02

Mn log	-0,44	-0,37	-0,08	0,56	-0,10
Mo log	0,43	-0,49	0,16	0,09	0,09
Nb log	-0,11	0,41	0,52	-0,17	-0,19
Ni log	-0,87	-0,02	0,13	-0,10	-0,09
Pb log	0,43	-0,29	0,61	-0,06	0,12
Rb log	0,21	0,18	0,72	0,36	-0,13
S log	-0,17	-0,63	0,17	-0,38	-0,40
Sb log	0,57	-0,25	0,23	0,09	0,32
Sc log	-0,89	-0,05	-0,03	-0,14	0,30
Se log	-0,07	-0,29	0,10	-0,18	0,02
Na	-0,60	0,13	-0,23	0,29	-0,08
Sn log	-0,44	0,28	0,53	-0,05	0,06
Sr log	-0,81	0,09	0,07	0,01	-0,11
Th log	0,22	0,58	0,23	0,14	0,57
Ti log	-0,76	0,41	0,08	0,09	0,06
V log	-0,91	-0,17	-0,08	-0,12	0,16
Zn log	-0,71	-0,39	0,07	0,35	-0,06
Zr log	-0,72	0,48	0,13	-0,10	0,32
Expl.Var	13,29	4,15	3,10	1,94	1,59
Prp.Totl	0,37	0,12	0,09	0,05	0,04

Наглядно значения каждого фактора можно отобразить в виде дробей.

В числителе факторы с положительной нагрузкой, в знаменателе с отрицательной.

$$I_{37} = \frac{Sb_{57}Bi_{53}Pb_{43}Mo_{43}As_{25}Th_{22}Rb_{21}}{Co_{91}V_{91}Mg_{90}Fe_{89}Sc_{89}Al_{88}Ni_{87}Ga_{84}Cu_{84}Ca_{82}Sr_{81}Ti_{76}Cr_{73}Zr_{72}Zn_{71}Hf_{69}Na_{60}Ba_{47}Mn_{44}Sn_{44}Be_{42}}$$

$$II_{12} = \frac{Th_{58}Zr_{48}Nb_{41}Ti_{41}Ba_{35}Sn_{28}}{S_{63}Au_{62}As_{52}Mo_{49}Ag_{44}Zn_{39}Cu_{39}Mn_{37}Li_{37}Hf_{32}Fe_{30}Pb_{29}Se_{29}Bi_{27}Mg_{25}Ga_{25}Sb_{25}Co_{24}}$$

$$III_9 = \frac{Rb_{72}Pb_{61}Be_{54}Sn_{53}Nb_{52}K_{48}Bi_{47}Ba_{47}Li_{37}Hg_{26}Th_{23}Cr_{22}}{Na_{23}}$$

$$IV_5 = \frac{Hg_{43}S_{38}Cr_{29}Ba_{21}}{K_{65}Mn_{56}Rb_{36}Zn_{35}Ca_{29}Na_{29}Li_{23}}$$

$$V_4 = \frac{Th_{57}Hf_{37}As_{34}Sb_{32}Zr_{32}Sc_{30}Au_{30}Li_{37}}{S_{40}K_{27}Ba_{21}Be_{20}}$$

Интерпретация первого фактора несет в себе 36,5% информации о рассматриваемом геологическом объекте. Первый фактор отражает кислые породы и основные породы. Основные породы – ассоциация элементов V Mg Fe Ni Ti Cr, кислые Th, Rb - у этих элементов нагрузки незначимые. Но скорее по первому фактору отражены основные породы и спутники золота Sb, Bi, Mo, As.

По второму фактору рудная минерализация. Золото имеет равные нагрузки с S и As и близкие с Mo, Ag, Zn, Cu. Наличие таких элементов, как Th, Zr, Nb и Ti, которые обычно связаны с магматическими процессами, предполагает, что в регионе могло наблюдаться значительное количество вулканической активности. Относительные концентрации таких элементов, как Au, Ag, Mo и Zn, которые обычно ассоциируются с минеральными месторождениями, позволяют предположить, что в регионе может существовать значительный потенциал для эпitherмальных кварцевых жил с золото-серебряной минерализацией.

Ведущая роль в 1 и 2 факторе породообразующий процесс- гранитный магматизм и на них наложена золото-серебряная минерализация. Фактор 3 несет в себе небольшую долю информации 9% и его интерпретация достаточна сложна. Можно предположить, что в контексте эпitherмальных кварцевых жил с золото-серебряной минерализацией присутствие таких элементов, как Rb, Pb, Sn и Ba, может указывать на высокотемпературную систему минерализации, в то время как такие элементы, как Be, Li, Hg и Th, могут быть связаны с низкотемпературной минерализацией. Высокое соотношение Nb/Na может указывать на магматическое происхождение некоторой минерализации, а наличие Cr может указывать на влияние гидротермальных изменений.

Заключение:

Статистические методы позволяют проводить анализ и оценку данных, определять статистические параметры, такие как среднее значение, медиана, стандартное отклонение и т.д., а также строить гистограммы и другие графики. Это помогает понять распределение данных и определить выбросы, что может быть полезным при интерпретации данных и выявлении причин аномальных значений.

Корреляционный анализ позволяет оценить степень взаимосвязи между различными параметрами, что может помочь определить физические и геологические процессы, ответственные за формирование данных параметров. Например, корреляция между

содержанием двух элементов может указывать на наличие минерала, содержащего оба элемента.

Факторный анализ позволяет выделить наиболее важные параметры, которые вносят наибольший вклад в общую дисперсию данных. Это может быть полезно для определения основных факторов, влияющих на формирование геологических объектов, и для улучшения точности прогнозирования параметров на основе полученных данных.

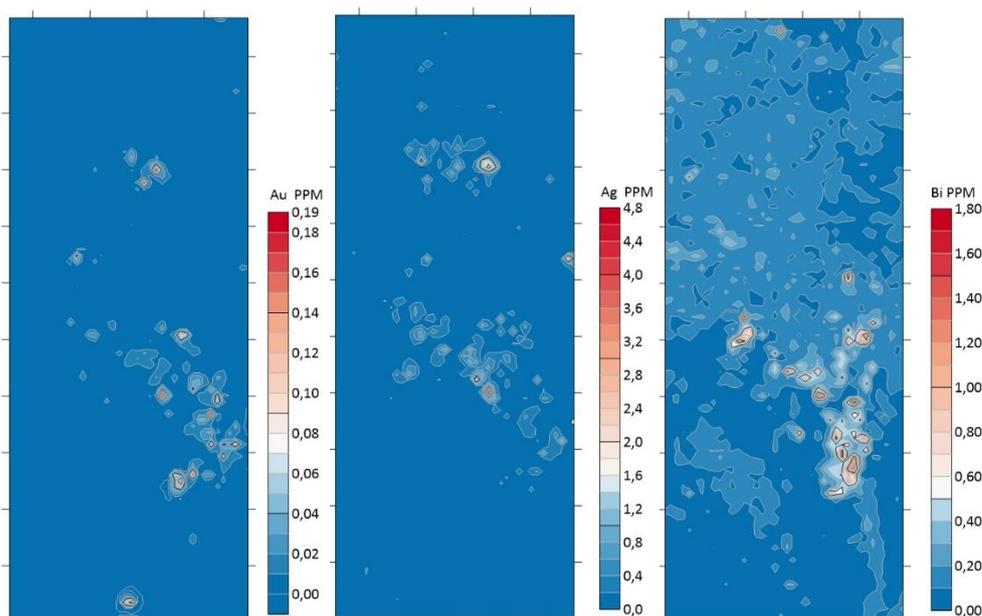
Таким образом, применение статистических методов может помочь в понимании геологических процессов, улучшении точности прогнозирования и выявлении необычных аномалий.

3.3 Интерпретация и сравнение полученных результатов на основе двух методов.

Для создания карт содержания элементов Au, Ag, Bi, Mo и Pb использовались данные, полученные в ходе предыдущих исследований. Эти данные были обработаны и преобразованы в формат, необходимый для использования программой Surfer (Levitt, J., 2014). С помощью этой программы были созданы карты, на которые нанесены данные о содержании каждого из элементов на исследуемой территории.

На данных картах можно увидеть, что на выбранной территории присутствует высокое содержание золота (Au) и серебра (Ag), а также значительное количество висмута (Bi), молибдена (Mo) и свинца (Pb). Особенно высокое содержание золота и серебра отмечается в районе Кайэнмываамского рудного поля, которое расположено в пределах Маюлервеемской вулканоструктуры Охотско-Чукотского вулканогенного пояса.

В центре карты повышенные содержания золота и его спутников, так же данные элементы имеют положительную корреляцию. Значения золота и серебра совпадают с картами предшественников, которые указывали на наличие золото-серебряных эпитермальных проявлений на данном участке.



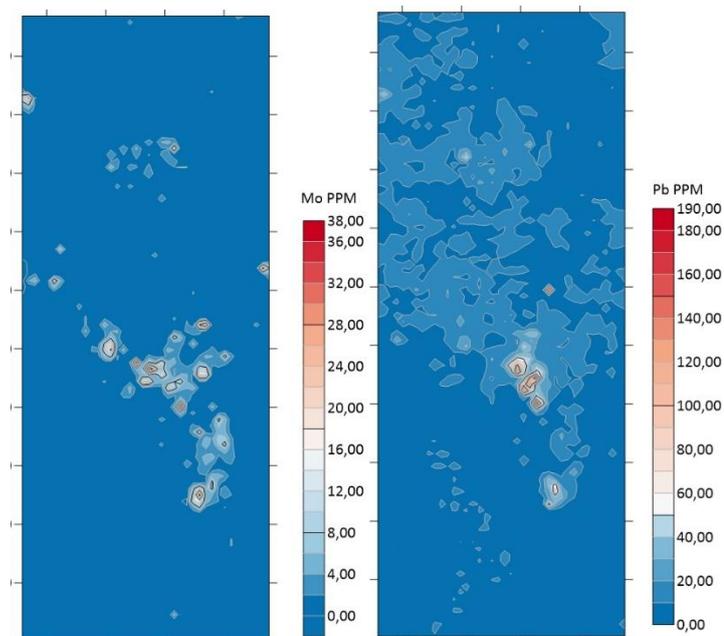


Рис. 3.9 Моноэлементные карты. [Составлено автором на основе программного обеспечения Surfer].

На карту так же были нанесены значения факторов.

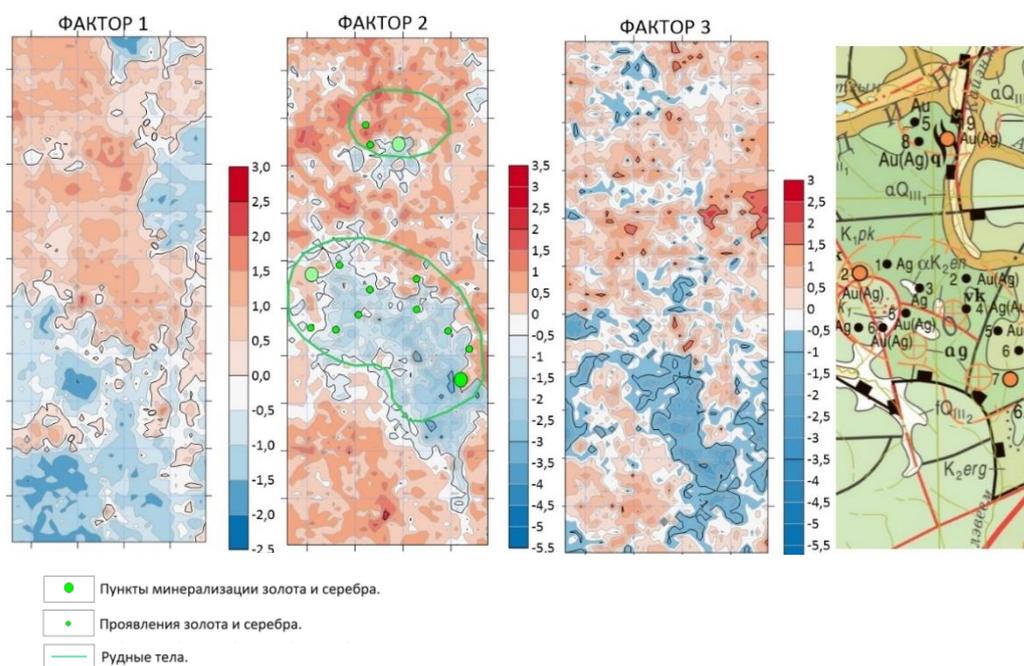


Рис. 3.10 Нанесение факторных нагрузок на площадь. [Составлено автором на основе программного обеспечения Surfer]

По значениям первого фактора и при сравнение его с геологической картой можно сделать следующие выводы: Положительные значения покрывают всю Пыкарваамску свиту (дациты, туфы риолитов, игнимбриты). Так же Пыкарваамская свита включает в себя

оруденение золота, серебра и свинца. Отрицательные значения первого фактора связаны с Энмываамской свитой (базальты, туфы, риолиты, алевролиты, риодациты), свита так же благоприятна для локализации золото-серебряной минерализации в пределах кольцевых вулканических структур и зонах поперечных глубинных разломов. Можно сделать вывод что первый фактор отражает различие между Пыкарваамской свитой и Энмываамской.

Второй фактор вызывает особый интерес, так как отражает эпитермальные золото-серебряные проявления. Отрицательные значения совпадают с картами предшественников, где отмечена рудная минерализация. В самом южном участке имеется значительная положительная аномалия, которая проходит через пересечение тектонических разломов. Данный участок связан с повышенными концентрациями элементов.

Интерпретация третьего фактора достаточна сложна, так как он не отражает значительных геологических процессов, можно предположить, что он связан с высоко и низкотемпературной минерализацией.

Далее полученные данные с помощью машинного обучения, при помощи самоорганизующихся сетей (Сети Кохонена) были нанесены на карту.

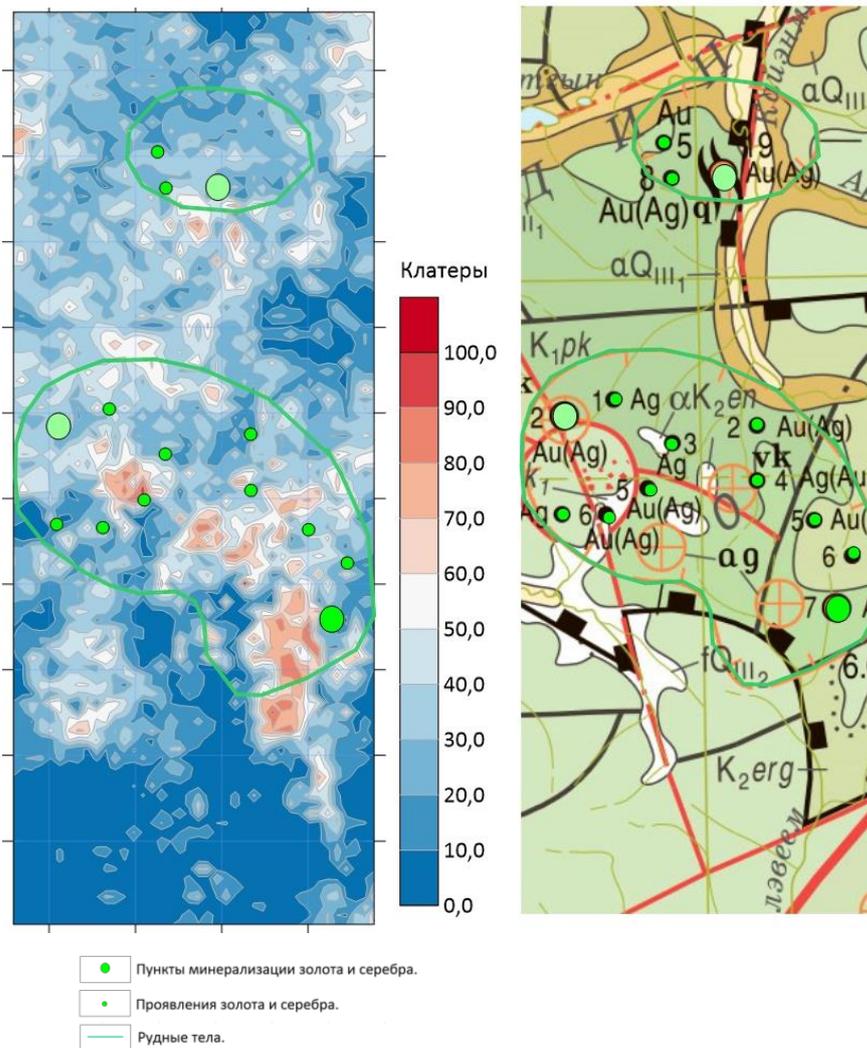


Рис. 3.11 Сравнение карты полученной при помощи алгоритма Сети Кохонена и геологической карты. [Составлено автором на основе программного обеспечения Surfer]

Подробнее разберем почему значения кластеров более 60, являются значимыми и совпадают с геологической картой предшественников.

Параметр значимости (importance score) позволяет оценить важность каждого входного признака (feature) в работе искусственной нейронной сети (ИНС). Этот параметр показывает, как изменение значения признака влияет на выходное значение ИНС.

Один из способов вычисления параметра значимости заключается в анализе весов (weights) ИНС. Веса представляют собой числовые значения, которые ИНС определяет для каждого входного признака, и они играют важную роль в принятии решений ИНС. Более высокие веса указывают на более важные признаки.

Для вычисления параметра значимости можно использовать следующий алгоритм:

1. Необходимо обучить ИНС на тренировочных данных, используя все входные признаки.

2. Вычислить выходное значение ИНС для тестовых данных.
3. Оценить вклад каждого входного признака в выходное значение ИНС, используя абсолютное значение веса. Чем выше абсолютное значение веса, тем больший вклад вносит признак.
4. Нормализация значения весов, чтобы они суммировались до 1.
5. Умножить нормализованные веса на соответствующие им значения признаков. Это даст оценку вклада каждого признака в выходное значение ИНС.
6. Просуммировать оценки вклада всех признаков, чтобы получить параметр значимости.

Математически этот алгоритм можно записать следующим образом:

Пусть w_i - вес i -го признака, x_i - значение i -го признака в тестовых данных, y - выходное значение ИНС. Тогда параметр значимости S_i для i -го признака можно вычислить следующим образом:

$$S_i = |w_i| * \left(\frac{x_i}{\sum_j |w_j|} * x_j \right)$$

где $\sum_j |w_j|$ - сумма абсолютных значений всех весов.

Таким образом, параметр значимости оценивает, насколько каждый признак важен для прогнозирования выходного значения ИНС. Он позволяет исследовать, как входные признаки взаимодействуют между собой и как они влияют на результаты ИНС.

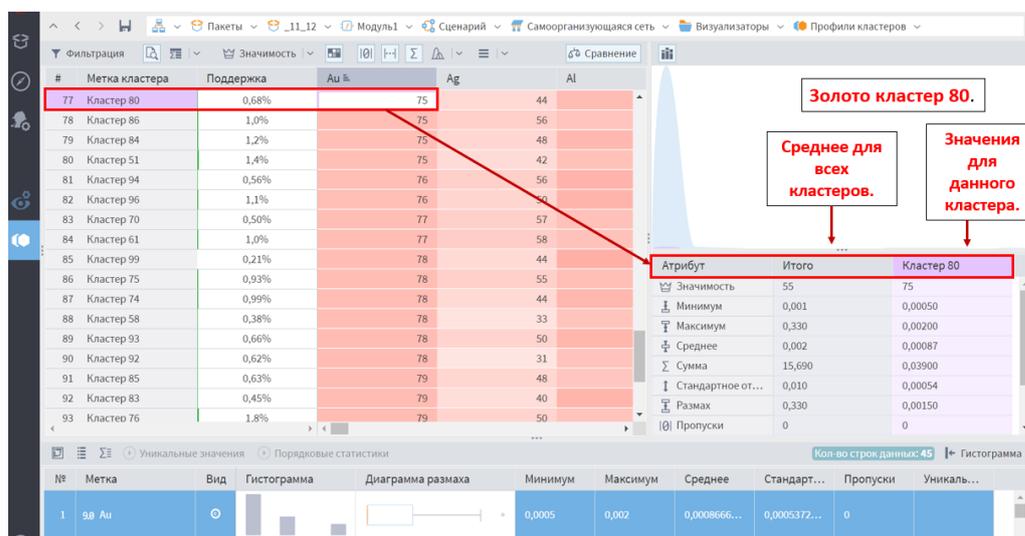


Рис. 3.12 Параметр «значимость» для золота. [Составлено автором на основе программного обеспечения Loginom]

Рассмотрим более подробно кластер 80.

Таблица атрибутов кластера 80 по золоту. [Составлено автором на основе программного обеспечения Loginom]

Атрибут	Итого	Кластер 80
👑 Значимость	55	75
⚓ Минимум	0,001	0,00050
⚓ Максимум	0,330	0,00200
⚖ Среднее	0,002	0,00087
∑ Сумма	15,690	0,03900
↑ Стандартное от...	0,010	0,00054
⚓ Размах	0,330	0,00150
∅ Пропуски	0	0
# Значения	6642	45
S Количество ун...		
⚓ Центр кластера		0,00084

В графе «Итого» - общая значимость параметра золота равна 55, что может означать, что этот параметр оказывает влияние на разделение объектов на кластеры. Значимость параметра для кластера 80 равна 75, что может указывать на то, что этот параметр является более важным для разделения объектов на этом конкретном кластере. В целом, более высокая значимость параметра может указывать на то, что этот параметр имеет большее влияние на разделение объектов на кластеры.

В результате можно сделать вывод: что параметр значимости является важным для дальнейшего построения геохимических карт, кластеры 60 и более отличаются значительными параметрами значимости золота и серебра. Делаем вывод что будем рассматривать данные кластеры, как золото-серебряные эпипермальные проявления.

Получаем карту достаточно схожую со второй фактор. Сеть Кохонена - это тип неконтролируемой нейронной сети, которая идентифицирует закономерности в данных и отображает их в пространство меньшей размерности. Преимущество использования сетей Кохонена заключается в том, что они могут выявлять сложные нелинейные взаимосвязи между переменными и представлять их в виде простой, легко интерпретируемой карты.

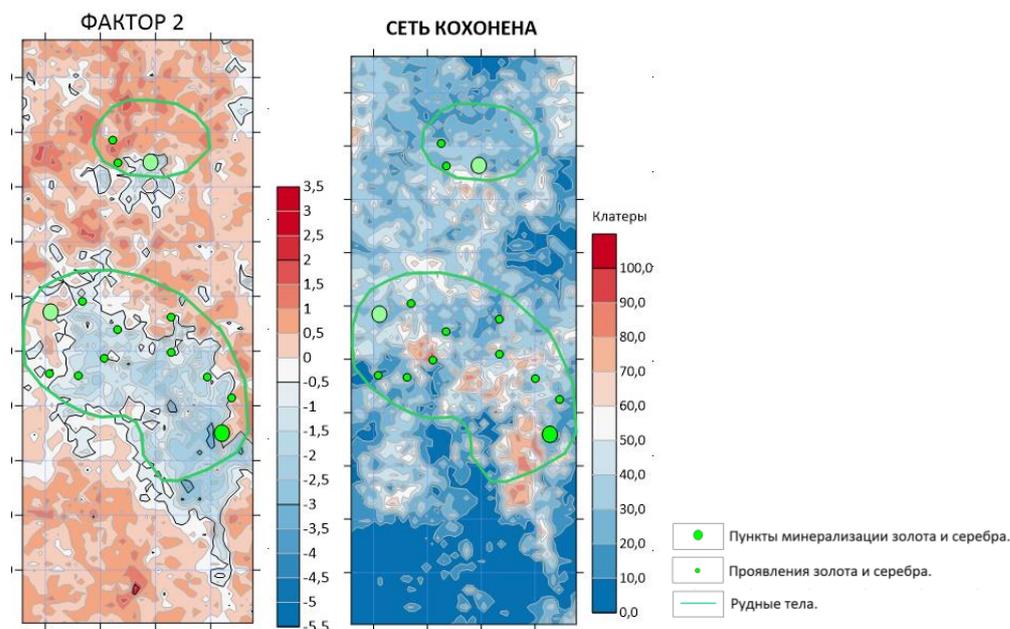


Рис. 3.13 Сравнение Фактора 2 и Сети Кохонена. [Составлено автором на основе программного обеспечения Surfer]

При сравнении второго рудного фактора и самоорганизующихся карт Кохонена, можно сделать визуально вывод, что карты очень схожи. В карте составленной с помощью машинного обучения аномалии более локальные.

Взяв точки наблюдения с наиболее перспективного полигона. Кросс-валидация (cross-validation) - это метод оценки качества обучаемых моделей машинного обучения, который позволяет оценить способность модели обобщаться на новые данные. Он заключается в разбиении исходного набора данных на несколько частей (фолдов), после чего обучение модели производится на одной из частей, а оценка - на другой. Такое разбиение повторяется несколько раз, при этом каждый раз используется другая часть данных для обучения и оценки.

Кросс-валидация может быть использована при прогнозировании месторождений полезных ископаемых на основе геохимических данных. Для этого можно разбить исходные геохимические данные на обучающую и тестовую выборки, обучить модель на обучающей выборке и оценить ее точность на тестовой выборке. Если модель показывает хорошую точность на тестовой выборке, это говорит о ее способности обобщаться на новые данные.

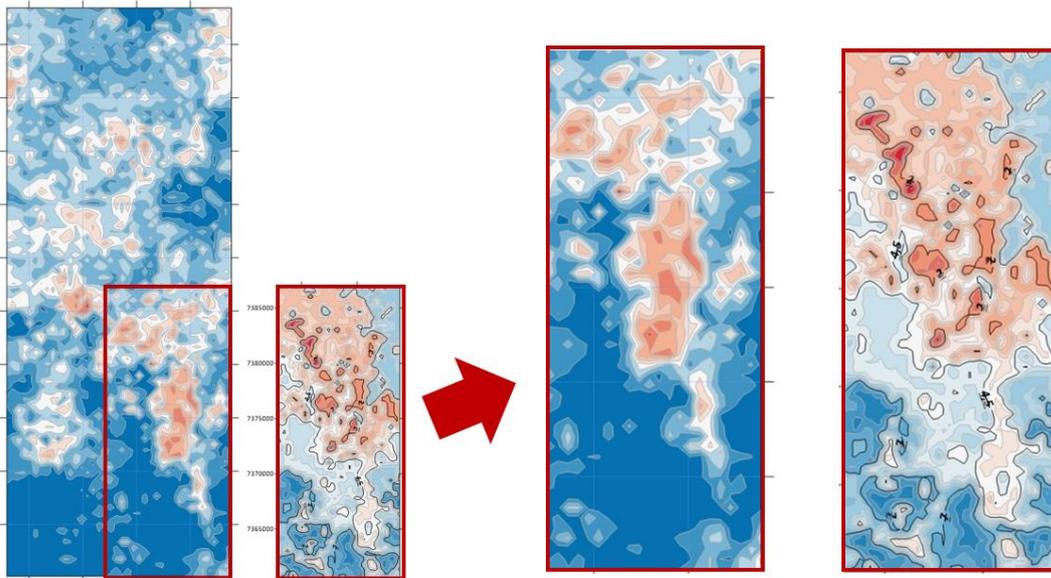


Рис. 3.14 Cross-validation на тестовой выборке. [Составлено автором на основе программного обеспечения Surfer]

Для оценки применимости сети Кохонена в выборке было исключено наличие значений золота и серебра. Как результат, была получена карта, на которой отсутствуют местные аномалии, связанные с эпитермальными проявлениями золота и серебра. Эта карта явно показывает различия между Пыкарваамской и Энмываамской свитами. Как и ожидаемо, пропадают участки оруденения, для работы данного метода чисто по элементам спутникам – вопрос требует до изучения.

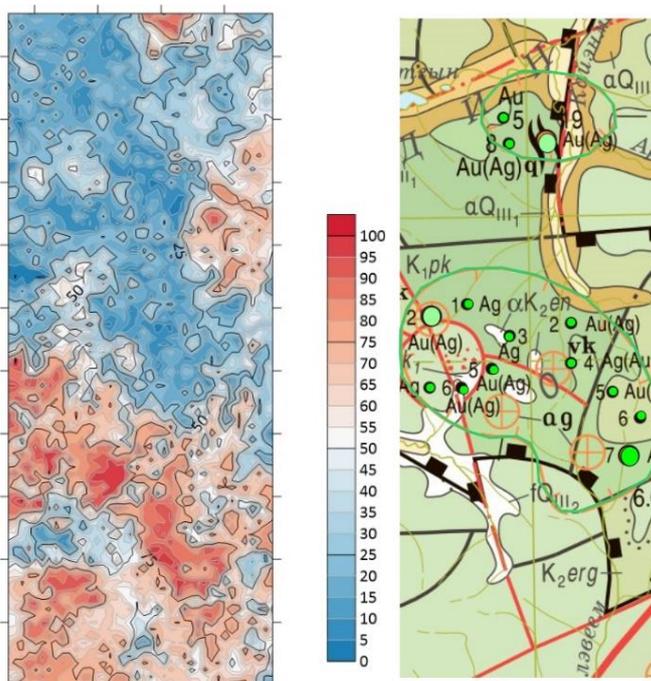


Рис. 3.15 Тестовая выборка, не включающая в себя золото и серебро. [Составлено автором на основе программного обеспечения Surfer]

Для оценки применимости, мы уменьшаем плотность начальных данных в два раза, а расстояние между точками наблюдения увеличивается до 400 на 800 метров. При использовании искусственной разреженной сети и картировании, мы получаем результаты, которые можно сравнить с исходными картами, однако, в центре наблюдается существенная аномалия, связанная с локализацией золото-серебряных эпитермальных проявлений. Однако, эта аномалия «расплывается» и даёт менее четкие тренды.

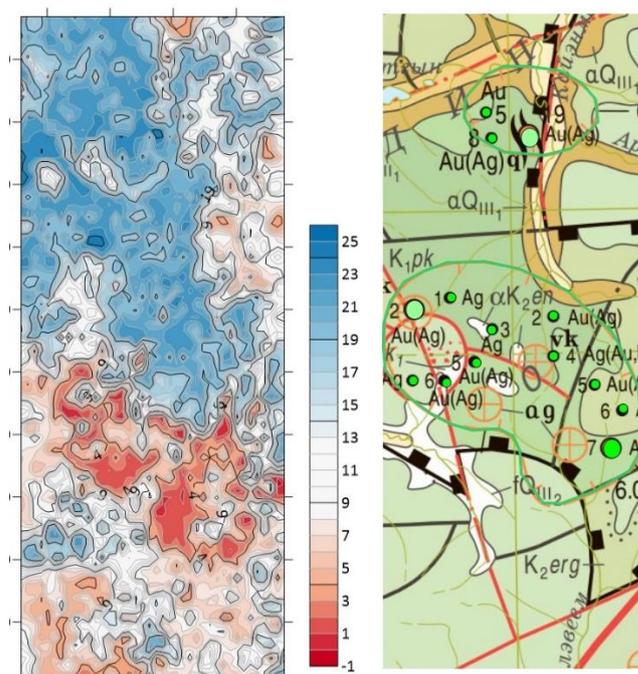


Рис. 3.16 Тестовая выборка, 400 на 800 м. [Составлено автором на основе программного обеспечения Surfer]

При разрежении сети до 800 на 800 метров видим основные черты геологического строения, в силу своих размеров аномальные черты теряются и можно сделать вывод для данных геологических условий сеть 400x400 является оптимальной.

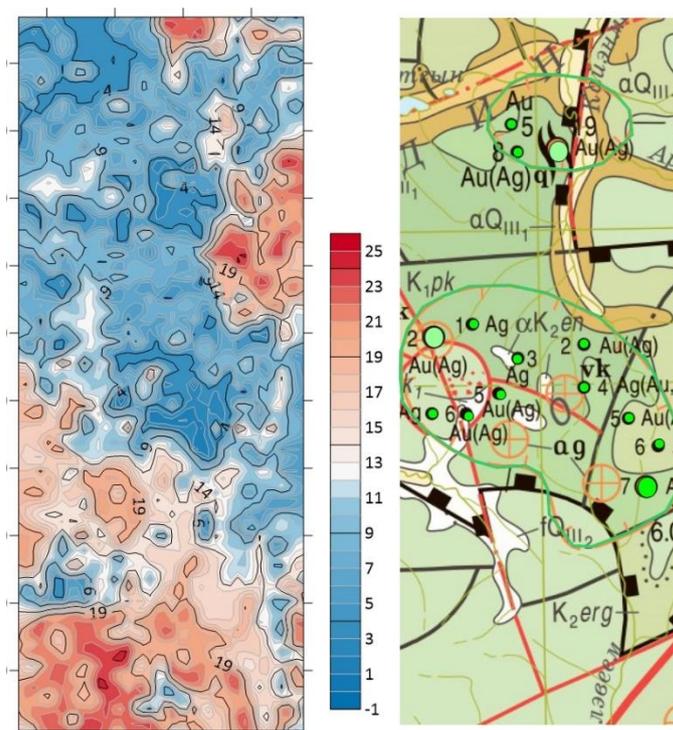


Рис. 3.17 Тестовая выборка, 800 на 800 метров. [Составлено автором на основе программного обеспечения Surfer].

Выводы:

1. По результатам исследования актуальных направлений применения технологий обработки больших данных, машинного и глубокого обучения, а также методов хранения больших объемов данных можно сделать вывод о том, что эти технологии имеют широкое применение в различных отраслях, включая науку, медицину, банковское дело, маркетинг, логистику, производство и др. Применение этих технологий позволяет существенно увеличить эффективность бизнеса, оптимизировать процессы и принимать более обоснованные решения.

2. Изучив применение методов машинного обучения в геологии и других отраслях, можно сделать вывод о том, что эти методы являются эффективными инструментами для анализа данных и построения прогнозов. Машинное обучение позволяет автоматически обрабатывать большие объемы данных, выделять закономерности, которые могут быть незаметны для человеческого восприятия. Однако, для успешного применения методов машинного обучения необходимо правильно выбирать модели и алгоритмы, обучать их на соответствующих данных, и проводить качественную проверку результатов.

3. Выполнение многомерного статистического анализа и выделение перспективных зон на участке является важной задачей. По результатам анализа были выделены наиболее перспективные участки для дальнейшего изучения и разработки.

4. Обработка полученных данных с помощью алгоритма кластеризации - сети Кохонена, позволяет выявлять особенности распределения полезных ископаемых и прогнозировать месторождения. Этот алгоритм обеспечивает высокую точность кластеризации и позволяет выделять схожие объекты в данных. На основе данного алгоритма были построены и проанализированы карты.

5. Сравнение результатов, полученных с помощью искусственных нейронных сетей и статистических методов, позволяет оценить точность прогнозирования месторождений и выбрать наиболее эффективный подход для решения задачи. Сеть Кохонена при сравнении с результатами, полученными с помощью статистических методов, продемонстрировала высокую точность прогнозирования месторождений и выделения перспективных зон на участке. Благодаря своей способности кластеризации и определению схожих объектов в данных, сеть Кохонена позволяет эффективно анализировать большие объемы данных и находить скрытые закономерности. Поэтому данный метод может быть рекомендован для дальнейшего использования на практике.

6. Алгоритм кластеризации - сети Кохонена используется в данной работе для выявления особенностей распределения полезных ископаемых и прогнозирования месторождений. Он основан на принципе самоорганизации и представляет собой нейронную сеть, которая обучается на основе входных данных и позволяет выделять схожие объекты в данных. Алгоритм состоит из следующих шагов: инициализация весов, выбор случайного образца, определение ближайшего вектора, обновление весов соседних нейронов, и повторение этих шагов до сходимости. Формулы и подробное описание алгоритма были представлены в данной работе.

ЗАКЛЮЧЕНИЕ

В работе была рассмотрена проблема прогнозирования месторождений полезных ископаемых, и были описаны основные этапы геологического моделирования и поиска месторождений.

Были проанализированы существующие методы обработки геохимических данных и использованы стандартные статистические методы для их обработки.

Для решения задачи кластеризации геохимических данных была использована сеть Кохонена, которая позволяет определить закономерности в данных и классифицировать их по категориям.

Были проведены эксперименты с использованием различных параметров сети Кохонена и геохимических данных для оценки качества моделирования.

Результаты экспериментов показали, что сеть Кохонена может успешно использоваться для кластеризации геохимических данных и предсказания месторождений полезных ископаемых.

Были предложены рекомендации по улучшению качества моделирования, включая использование дополнительных данных и уточнение параметров сети Кохонена.

Таким образом, результаты исследования подтверждают эффективность применения сети Кохонена и стандартных статистических методов при моделировании геологических систем, прогнозировании и поиске месторождений полезных ископаемых. Однако, для улучшения точности прогнозирования необходимо проводить дополнительные исследования и использовать дополнительные данные.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. A density-based algorithm for discovering clusters in large spatial databases with noise / M. Ester et al. // In Kdd. – 1996. – Vol. 96, №. 34. – P. 226-231.
2. Active learning of rock typing models from multiscale measurements / Y. Zhao // IEEE Transactions on Geoscience and Remote Sensing. – 2018. – № 56(5). – P. 2994-3004.
3. Bishop C. M. Pattern Recognition and Machine Learning / C. M. Bishop. Information Science and Statistics. - 2nd ed. Springer, 2006. - 758 p.
4. Cao L. Data science: a comprehensive overview / L. Cao // ACM Comput Surv (CSUR). – 2017. –. № 50(3):43. -P. 1-40
5. Data mining: concepts and techniques / Han J et al. -3th ed. -Amsterdam: Elsevier, 2011. - 740 p.
6. Friedman J. H. Greedy function approximation: a gradient boosting machine / J. H. Friedman // Annals of Statistics. – 2001. – № 29(5). – P. 1189-1232.
7. Haykin S. Neural networks: a comprehensive foundation / S. Haykin, Pearson. - 2nd ed. Prentice Hall, 1999. - 842p.
8. Iqbal H. Machine Learning: Algorithms, Real World Applications and Research Directions / H. Iqbal, S.N. Sarker// Computer Science. – 2021. – P. 4.
9. Jing M. Application of machine learning in predicting coal and gas outburst danger: a review / M. Jing, Z. Chong // Journal of Natural Gas Science and Engineering. – 2020. –№ 103329. – P.78.
10. Jurafsky D. Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition / Jurafsky D. et al. -3rd ed. Prentice Hall, 2019. -636 p.
11. Kaelbling L. P. Reinforcement learning: a survey / L.P. Kaelbling, M.L. Littman, A.W. Moore// J Artif Intell Res. – 1996. – № 4:237. – P. 85.
12. Kohonen T. Self-organized formation of topologically correct feature maps / T. Kohonen // Biological cybernetics. – 1982. – № 43(1), – P. 59-69.
13. Kohonen, T. Self-organizing maps / T. Kohonen. // Springer-Verlag, 1995. - P. 7-12
14. Kohonen, T. The self-organizing map / T. Kohonen. // Proceedings of the IEEE. – 1990. – № 78(9). – P. 1464-1480.
15. Levitt J. Using Surfer Software in Geomorphology / J. Levitt // In Geomorphological Techniques. – Springer, Dordrecht. – 2014. – P. 177-189.

16. Li X. Application of machine learning to geochemical data for mineral exploration: A case study from Dongping gold deposit, China / X. Li, Q. Li, J. Li, M. Li // *Ore Geology Reviews*. – 2020. – № 103612. – P. 125
17. Li, Y. Predicting porosity of marine carbonate rocks using random forest regression from seismic inversion attributes / Y. Li, B. Feng // *Geophysical Journal International*. – 2017. – № 211(3). – P. 1413-1433.
18. Main I. G. Machine learning for earthquake risk prediction. / I. G. Main, L. J. Donnelly // *Earth-Science Reviews*. – 2019. – № 102896. – P. 197.
19. Mobile data science and intelligent apps: concepts, ai-based modeling and research directions / I.H. Sarker et al. // *Mob Netw Appl*. – 2020. – P. 1–19.
20. Murtagh F. Algorithms for hierarchical clustering: an overview / Murtagh F., & Contreras P. // *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. – 2012. – № 2(1). – P. 86-97.
21. Ng A. On spectral clustering: analysis and an algorithm / A. Ng, M. Jordan, Y. Weiss // In *Advances in neural information processing systems*. – 2002. – P. 849-856.
22. Pedregosa F. Scikit-learn: machine learning in python / Pedregosa F et al. // *J Mach Learn Res*, 2011. - № 12:2825-30. -P. 234-265
23. Russakovsky O. ImageNet large scale visual recognition challenge / Russakovsky O. et al. // *International Journal of Computer Vision*. – 2015. – № 115(3). – P. 211-252.
24. Sahin H. O. Prediction of coal quality parameters using machine learning algorithms: A comparative study / H.O. Sahin, B. Demiral // *Fuel*. – 2020. – № 116377. – P. 261.
25. Sarker I.H. Cybersecurity data science: an overview from machine learning perspective / I.H. Sarker // *Big Data*. – 2020. – № 7(1):1. – P. 29.
26. Semi-Supervised Learning for Fault Detection in Seismic Images / J. Deng et al. // *IEEE Transactions on Geoscience and Remote Sensing*. – 2021. – № 59(2). – P. 1052-1065.
27. Shalev-Shwartz, S. Understanding Machine Learning: From Theory to Algorithms / Shalev-Shwartz, S., Ben-David, S. - 1st ed. Cambridge University Press, 2014. - 449 p.
28. Skvortsov D. A. Neural network application for geological parameters forecasting of oil and gas deposits / D.A. Skvortsov, T.A. Burova, M. V Kuznetsov // *Geofizika*. – 2019. – № 36(1). – P. 45-56.
29. Sun J. Geological structure classification using deep learning: A case study of the Wufeng–Longmaxi shale gas area, southeast Sichuan Basin / J. Sun, Y. Wang, X. Sun // *Journal of Natural Gas Science and Engineering*. – 2021. – № 103896. – P. 98.

30. Sun X. Predicting porosity and permeability in reservoir rocks using machine learning techniques / X. Sun, Y. Zhao, X. Zhang // Journal of Petroleum Science and Engineering. – 2019. – № 178. – P. 485-499.
31. Witten I.H. Data Mining: Practical machine learning tools and techniques / I.H. Witten, E. Frank. -2nd ed. Morgan Kaufmann, 2005. -558 p.
32. Zhang H. An automatic method for clustering lithological units based on geochemical data: A case study in the Qimantage area, China / H. Zhang, S. Jiang, S. Li // Computers & Geosciences. – 2020. – № 104483. – P. 144.
33. Zhang Y. Machine learning in finance: status and challenges /Y. Zhang, Y. Qi, X. Yang, // Journal of Business Research. – 2019. – № 98. – P. 365-380.
34. Zheng H. Microstructure Investigation of Oil-Bearing Rhyolites: A Case Study from the Hailar Basin / M.Anqi, H. Zheng, S. Xiaomeng // Geofluids. -2020. - №10(8).
35. Алексеев П. А. Алгоритмы классификации и идентификации аудиозаписей // Тульский государственный педагогический университет им. Л.Н. Толстого факультет математики, физики и информатики (магистратура, 1-ый год обучения) / под науч. ред. д. ф.-м. н., доцент Балаба И. Н. Тула, 2022.
36. Золото-серебряное эпитермальное месторождение Двойное (п-ов Чукотка, Россия) / А.В.Волков и др. // Геология рудных месторождений. -2018. -Т.60. -№ 6. -С. 590–609.
37. Акинин В.В. Эволюция известково-щелочных магм Охотско-Чукотского вулканогенного пояса / В.В. Акинин, Э.Л. Миллер // Петрология. -2011. -Т.19, № 3. -С. 249—290.
38. Аргиллизитовые “шляпы” рудопроявления “Комплексное”, Кайэнмываамское вулканическое поднятие (Центральная Чукотка) / П. Е Белоусов и д.р // Вулканология и Сейсмология. -2020. -№5. -С. 5-12.
39. Витязев А.В. Методы геохимического анализа и обработки геохимических данных / А.В. Витязев // Изд-во Института геологии и минеральных ресурсов СО РАН. -2011.
40. Гордеев Е.В. Статистические методы обработки и интерпретации геохимических данных. / Е.В. Гордеев // Изд-во Нижегородского государственного университета им. Н.И. Лобачевского. -2015.
41. Золото-теллуридная минерализация Западной Чукотки: минералогия, геохимия и условия образования / Ю.Н. Николаев и др // Геология рудных месторождений. -2013. - Т.55, -№2. -С. 114–144.
42. Казначеев П.Ф. Применение методов искусственного интеллекта для повышения эффективности в нефтегазовой и других сырьевых отраслях. / П.Ф Казначеев, Р.В. Самойлова, Н.В. Курчиски // Экономическая политика. -2016. -Т.11. -№5. -С. 190-192.

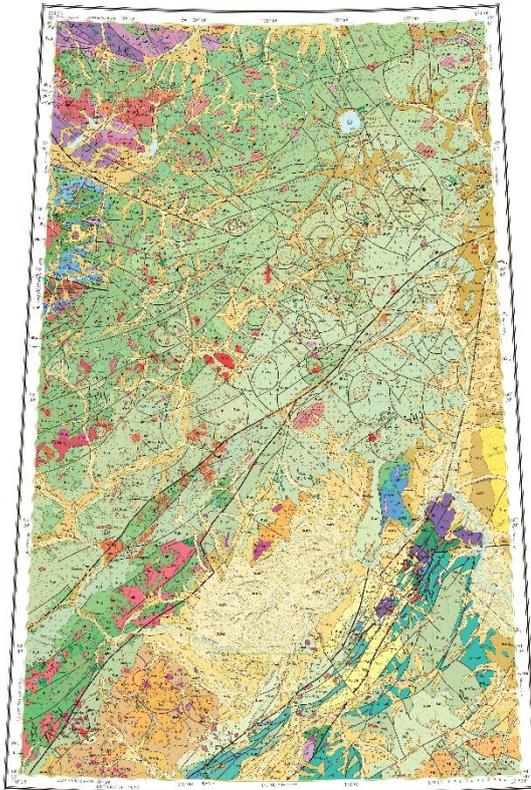
43. Кряжев С.Г. Использование метода ICP MS при анализе состава рудообразующих флюидов / С.Г. Кряжев, В.Ю. Прокофьев, Ю.В. Васюта // Вестн. МГУ. Сер. 4. Геология. - 2006. -№ 4. -С. 30–36.
44. Леонтьев В.М. Геохимические методы поисков рудных месторождений. / В.М. Леонтьев, В.А. Петров, Г.Ф.Митрофанов // Новосибирск: Изд-во Института геологии и минеральных ресурсов СО РАН. -2013. -С. 4-12
45. Маштакова И.Б. Статистические методы обработки геохимических данных в прикладной геохимии: учебное пособие. / И.Б. Маштакова, М.А. Чернобыльская // Изд-во МГУ. - 2013. -№2
46. Новая находка золото-теллуридной минерализации на Чукотке: минералогия и условия формирования рудопоявления Телевеем / Е.А. Власов и др. // Руды и металлы. -2016. - № 4. -С. 48–50.
47. Особенности рудообразования на золото-серебряном месторождении Купол, Северо-Восток России (по данным исследования флюидных включений) / А.В.Волков и др. // Геология рудных месторождений. -2012. -Т.54. -№ 4. -С. 350– 359.
48. Первые данные об условиях формирования эпитермальной золото-серебряные минерализации месторождения Жильное (Восточная Чукотка) / А.В. Волков и др. // Доклады академии наук. -2018. -Т.480. -№ 6. -С. 693–697.
49. Рудообразование на Au-Ag месторождении Купол, по данным изучения флюидных включений (СевероВосток России) / А.В. Волков и др. // Геология рудных месторождений. -2012. -Т. 54. -№ 4. -С. 350-359.
50. Стружков С.Ф. Провинция Охотско-Чукотского вулканогенного пояса / С.Ф. Стружков // Золоторудные месторождения России. М.: Акварель. -2010. -С. 213-242.
51. Тихомиров П.Л. Чукотский сегмент Удско-мургальского и Охотско-чукотского вулканических поясов: геодинамическая природа и вопросы возраста / П.Л. Тихомиров, Н.В. Правикова, Я.В. Бычкова // Геология и геофизика. – 2020. -Т.61. -№4
52. Устиев Е.К. Охотский структурный пояс и проблемы вулканоплутонических формаций / Е.К. Устиев // Проблемы магмы и генезис изверженных пород. М., Изд-во АН СССР. - 1963. -С. 161—182.
53. Устиев Е.К. Охотский тектономагматический пояс и некоторые связанные с ним проблемы / Е.К. Устиев // Советская геология. -1959. -№ 3. -С. 3-26.
54. Хорин Г.И. Прогнозная оценка ресурсов полезных ископаемых по данным литохимической съемки / Г.И. Хорин, С.В. Соколов// Геохимические методы при геологическом картировании. Сбор. ИМГРЭ. -1985. -С. 50-70.

Электронные ресурсы:

55. Low-code платформа для реализации всех аналитических процессов [Электронный ресурс] // Loginom 2023. - Режим доступа: <https://loginom.ru/>, свободный. (дата обращения: 25.04.2023).
56. Академия Google [Электронный ресурс] // Google. - Режим доступа: <https://scholar.google.com/>, свободный. (дата обращения: 25.04.2023).
57. КиберЛенинка предоставляет возможность читать тексты научных статей бесплатно [Электронный ресурс] // Научная электронная библиотека «КиберЛенинка». - Режим доступа: <https://cyberleninka.ru/>, свободный. (дата обращения: 25.04.2023).
58. Научно исследовательский геологический институт [Электронный ресурс] // Всероссийский научно-исследовательский геологический институт им. А.П. Карпинского. - Режим доступа: <https://vsegei.ru/ru/>, свободный. (дата обращения: 25.04.2023).
59. Росгеология [Электронный ресурс] // 2023 АО «Росгео». - Режим доступа: <https://rusgeology.ru/>, свободный. (дата обращения: 25.04.2023).

Приложение 1

ГОСУДАРСТВЕННАЯ ГЕОЛОГИЧЕСКАЯ КАРТА РОССИЙСКОЙ ФЕДЕРАЦИИ масштаба 1:1 000 000
 ЧУКОТСКАЯ СЕРИЯ
 Q-59 (Менделеевский лист)



Масштаб 1:1 000 000
 Численный масштаб 1 см = 100 км
 Иллюстрация масштаба 1:100 000

У С Л О В Е Н И К

1	Горные породы (символы)
2	Горные породы (символы)
3	Горные породы (символы)
4	Горные породы (символы)
5	Горные породы (символы)
6	Горные породы (символы)
7	Горные породы (символы)
8	Горные породы (символы)
9	Горные породы (символы)
10	Горные породы (символы)
11	Горные породы (символы)
12	Горные породы (символы)
13	Горные породы (символы)
14	Горные породы (символы)
15	Горные породы (символы)
16	Горные породы (символы)
17	Горные породы (символы)
18	Горные породы (символы)
19	Горные породы (символы)
20	Горные породы (символы)
21	Горные породы (символы)
22	Горные породы (символы)
23	Горные породы (символы)
24	Горные породы (символы)
25	Горные породы (символы)
26	Горные породы (символы)
27	Горные породы (символы)
28	Горные породы (символы)
29	Горные породы (символы)
30	Горные породы (символы)
31	Горные породы (символы)
32	Горные породы (символы)
33	Горные породы (символы)
34	Горные породы (символы)
35	Горные породы (символы)
36	Горные породы (символы)
37	Горные породы (символы)
38	Горные породы (символы)
39	Горные породы (символы)
40	Горные породы (символы)
41	Горные породы (символы)
42	Горные породы (символы)
43	Горные породы (символы)
44	Горные породы (символы)
45	Горные породы (символы)
46	Горные породы (символы)
47	Горные породы (символы)
48	Горные породы (символы)
49	Горные породы (символы)
50	Горные породы (символы)
51	Горные породы (символы)
52	Горные породы (символы)
53	Горные породы (символы)
54	Горные породы (символы)
55	Горные породы (символы)
56	Горные породы (символы)
57	Горные породы (символы)
58	Горные породы (символы)
59	Горные породы (символы)
60	Горные породы (символы)
61	Горные породы (символы)
62	Горные породы (символы)
63	Горные породы (символы)
64	Горные породы (символы)
65	Горные породы (символы)
66	Горные породы (символы)
67	Горные породы (символы)
68	Горные породы (символы)
69	Горные породы (символы)
70	Горные породы (символы)
71	Горные породы (символы)
72	Горные породы (символы)
73	Горные породы (символы)
74	Горные породы (символы)
75	Горные породы (символы)
76	Горные породы (символы)
77	Горные породы (символы)
78	Горные породы (символы)
79	Горные породы (символы)
80	Горные породы (символы)
81	Горные породы (символы)
82	Горные породы (символы)
83	Горные породы (символы)
84	Горные породы (символы)
85	Горные породы (символы)
86	Горные породы (символы)
87	Горные породы (символы)
88	Горные породы (символы)
89	Горные породы (символы)
90	Горные породы (символы)
91	Горные породы (символы)
92	Горные породы (символы)
93	Горные породы (символы)
94	Горные породы (символы)
95	Горные породы (символы)
96	Горные породы (символы)
97	Горные породы (символы)
98	Горные породы (символы)
99	Горные породы (символы)
100	Горные породы (символы)

Государственная геологическая карта Российской Федерации (Чукотская серия). Карта составлена: ФГБУ «ВСЕГЕИ», 2016 г.

Приложение 2

	Valid N	Mean	Geometric	Harmonic	Median	Mode	Frequency	Sum	Min	Max	Std.Dev.	Skewness	Std.Err.	Kurtosis
Au	6641	0,0024	0,0009	0,0007	0,0005	0,0005	4499	16	0,0005	0,33	0,01	16,47	0,03	372,62
Ag	6641	0,08	0,05	0,03	0,05	0,04	889	553	0,01	6,44	0,17	16,36	0,03	431,38
Al	6641	2,85	2,49	2,13	2,59	1,54	33	18921	0,14	8,95	1,43	0,64	0,03	-0,19
As	6641	9,7	4,7	2,45	4,6	101	125	64451	0,10	321	18,49	5,66	0,03	47,75
B	6641	5,03	5,02	5,01	5	5	6621	33402	5,00	27	0,60	24,6	0,03	695,59
Ba	6641	143,79	123,17	104,56	125	77	56	954880	5,00	1664	88,16	3,09	0,03	26,70
Be	6641	0,64	0,55	0,42	0,57	0,44	120	4242	0,03	3,28	0,34	1,33	0,03	3,53
Bi	6641	0,13	0,1	0,08	0,1	0,09	477	859	0,01	3,59	0,16	9,26	0,03	133,59
Ca	6641	0,63	0,41	0,23	0,44	0,12	133	4208	0,01	4,39	0,57	1,47	0,03	2,67
Cd	6641	0,06	0,04	0,02	0,05	0,005	764	410	0,01	2,86	0,09	14,55	0,03	337,19
Ce	6641	34,29	31,1	28,15	30,91	Multiple	8	227731	1,34	215,48	16,19	1,77	0,03	6,83
Co	6641	11,90	8,69	6,05	8,9	3	85	79043	0,20	124,3	8,95	1,33	0,03	6,17
Cr	6641	21,34	16,28	12,5	15	9	344	141711	0,50	151	16,62	1,6	0,03	3,13
Cs	6641	2,36	1,56	1,07	1,55	1,07	41	15693	0,03	56,14	2,84	4,77	0,03	42,28
Cu	6641	23,96	16,07	10,4	17,2	4,5	60	159089	0,40	194	21,21	1,64	0,03	4,01
Fe	6641	3,14	2,73	2,34	2,86	1,44	35	20821	0,23	18,38	1,57	0,51	0,03	0,42
Ga	6641	7,62	6,9	6,16	7,33	4,76	16	50583	0,39	20,8	3,22	0,47	0,03	-0,28
Ge	6641	0,06	0,04	0,04	0,03	0,025	3851	367	0,03	0,47	0,05	2,37	0,03	8,38
Hf	6641	0,15	0,08	0,04	0,09	0,01	1124	965	0,01	1,03	0,15	1,66	0,03	3,33
Hg	6641	0,04	0,03	0,02	0,03	0,005	1166	288	0,01	1,82	0,06	11,23	0,03	260,59
In	6641	0,03	0,02	0,02	0,03	0,0025	225	200	0,0025	2,27	0,03	42,84	0,03	2730,62
K	6641	0,12	0,1	0,09	0,1	0,1	660	767	0,01	0,92	0,06	3,24	0,03	21,11
La	6641	15,84	14,43	13,11	14,6	11,6	58	105186	0,70	120,4	7,32	2,13	0,03	13,13
Li	6641	12,92	11,02	9,37	10,6	8,9	82	85769	0,50	75,1	8,34	2,32	0,03	7,46
Mg	6641	0,69	0,47	0,32	0,43	0,17	149	4577	0,01	4,06	0,61	1,44	0,03	1,93
Mn	6641	629,97	508,98	383,16	569	485	16	4183627	18	9537	451,08	4,71	0,03	57,5
Mo	6641	0,76	0,36	0,23	0,33	0,19	159	5042	0,03	68,48	2,46	14,64	0,03	289,42
Na	6641	0,07	0,05	0,04	0,05	0,03	1177	468	0,01	1,77	0,07	4,64	0,03	72,36
Nb	6641	0,95	0,77	0,54	0,88	0,75	71	6341	0,03	5,63	0,6	1,52	0,03	4,33
Ni	6641	14,91	8,81	5,73	7,6	3,3	94	99020	0,10	184,4	18,89	3,07	0,03	12,73
Pb	6641	9,27	8,01	7,25	7,9	7,2	105	61579	1,3	341,8	10,15	16,98	0,03	418,86
Rb	6641	8,85	7,78	6,75	8,1	6	85	58752	0,4	76,3	4,76	2,23	0,03	12,97
Re	6641	0,0008	0,0006	0,0005	0,0005	0,0005	6061	5	0,001	0,07	0,0018	21,12	0,03	591,33
S	6641	0,04	0,02	0,02	0,02	0,02	1894	248	0,01	1,51	0,07	10,03	0,03	144,58
Sb	6641	0,21	0,11	0,07	0,13	0,025	1574	1421	0,03	15,17	0,45	13,66	0,03	318,90
Sc	6641	6,24	4,5	3,01	5	1,5	129	41410	0,1	27,20	4,62	0,84	0,03	-0,04
Se	6641	0,54	0,25	0,17	0,2	0,1	3305	3565	0,1	27	1,12	8,89	0,03	127,84
Sn	6641	0,67	0,62	0,57	0,6	0,6	1327	4425	0,1	3,9	0,24	1,24	0,03	8,07
Sr	6641	162,71	85,92	49,76	81,40	Multiple	13	1080579	3,40	4611,3	247,89	5,37	0,03	51,47
Te	6641	0,07	0,01	0,01	0,01	0,005	3523	483	0,01	10,72	0,34	17,17	0,03	420,02
Ta	6641	0,01	0,01	0,01	0,01	0,005	6560	36	0,01	0,19	0,01	22,11	0,03	593,46
Th	6641	1,89	1,43	0,95	1,5	0,8	315	12540	0,10	10,7	1,33	1,09	0,03	1,09
Ti	6641	0,11	0,07	0,03	0,08	0,0025	111	745	0,003	0,63	0,1	1,21	0,03	0,89
Tl	6641	0,08	0,07	0,05	0,07	0,05	725	539	0,01	1,34	0,06	6,01	0,03	90,93
U	6641	0,83	0,67	0,53	0,71	Multiple	76	5481	0,03	9,03	0,58	2,98	0,03	22,86
V	6641	78,75	61,01	44,78	70	22	114	522963	4	236	50,69	0,52	0,03	-0,74
W	6641	0,07	0,05	0,04	0,03	0,025	3599	485	0,03	1,97	0,1	6,79	0,03	84,40
Y	6641	8,94	7,85	6,81	7,95	4,58	16	59380	0,44	49,63	4,62	1,12	0,03	2,02
Zn	6641	49,69	44,93	40,09	46	42	149	329984	2	497	24,03	3,75	0,03	47,09
Zr	6641	7,34	3,48	1,39	3,9	0,25	564	48720	0,25	64	8,39	1,81	0,03	3,88

Приложение 3

	Au	Ag	Al	As	Ba	Bi	Ca	Co	Cr	Cs	Cu	Fe	Ga	Ge	Hg	K	Li	Mg	Mn	Mo	Na	Nb	Ni	Pb	Rb	S	Sb	Sc	Se	Sn	Sr	Th	Ti	V	Zn	Zr
Au	1,00	0,37	-0,05	0,12	-0,12	0,20	-0,05	0,00	-0,04	0,10	0,11	0,02	0,01	-0,02	0,10	0,01	0,13	0,03	0,03	0,34	-0,06	-0,12	-0,03	0,28	0,00	0,12	0,12	-0,02	0,12	-0,06	-0,06	-0,06	-0,10	-0,02	0,22	-0,09
Ag	0,37	1,00	-0,09	0,11	-0,10	0,19	-0,08	-0,05	-0,07	0,13	0,07	-0,04	-0,05	-0,02	0,11	0,08	0,11	-0,04	0,03	0,44	-0,11	-0,09	-0,06	0,47	0,05	0,22	0,13	-0,09	0,13	-0,10	-0,07	-0,03	-0,13	-0,10	0,12	-0,11
Al	-0,05	-0,09	1,00	0,02	0,41	-0,18	0,78	0,77	0,57	-0,10	0,66	0,80	0,90	0,26	0,02	0,06	0,12	0,75	0,29	-0,13	0,48	0,19	0,63	-0,16	-0,07	0,04	-0,18	0,77	0,02	0,37	0,55	-0,37	0,67	0,79	0,51	0,53
As	0,12	0,11	0,02	1,00	-0,03	0,12	0,03	0,06	0,02	0,12	0,10	0,09	0,04	0,06	0,12	0,17	0,08	0,09	0,07	0,17	-0,02	-0,08	0,08	0,10	0,07	0,23	0,21	-0,01	0,12	-0,04	0,06	-0,07	-0,02	-0,01	0,11	-0,03
Ba	-0,12	-0,10	0,41	-0,03	1,00	-0,06	0,24	0,27	0,46	-0,18	0,09	0,29	0,34	0,22	0,01	0,27	-0,19	0,21	0,06	-0,09	0,21	0,29	0,37	-0,02	0,30	-0,02	-0,14	0,31	-0,01	0,46	0,34	0,09	0,42	0,26	0,19	0,51
Bi	0,20	0,19	-0,18	0,12	0,24	1,00	-0,25	-0,21	-0,17	0,11	-0,06	-0,13	-0,12	-0,11	0,15	0,11	0,06	-0,17	-0,08	0,26	-0,20	-0,11	-0,19	0,35	0,11	0,34	0,18	-0,20	0,33	-0,06	-0,16	0,03	-0,29	-0,21	-0,03	-0,21
Ca	-0,05	-0,08	0,78	0,03	0,00	-0,25	1,00	0,73	0,44	-0,04	0,66	0,72	0,69	0,35	-0,08	0,22	0,09	0,77	0,35	-0,13	0,69	0,04	0,58	-0,20	-0,10	-0,05	-0,14	0,70	-0,03	0,23	0,55	-0,30	0,73	0,71	0,54	0,50
Co	0,00	-0,05	0,77	0,06	0,46	-0,21	0,73	1,00	0,60	-0,16	0,74	0,89	0,74	0,30	-0,05	0,00	0,11	0,81	0,55	-0,07	0,49	0,03	0,66	-0,17	-0,18	0,02	-0,17	0,78	0,03	0,23	0,47	-0,37	0,65	0,83	0,62	0,47
Cr	-0,04	-0,07	0,57	0,02	-0,18	-0,17	0,44	0,60	1,00	-0,25	0,36	0,58	0,52	0,32	-0,04	-0,02	-0,06	0,56	0,13	-0,09	0,31	0,32	0,74	-0,12	-0,05	0,00	-0,18	0,51	0,02	0,40	0,44	-0,16	0,65	0,56	0,42	0,59
Cs	0,10	0,13	-0,10	0,12	0,09	0,11	-0,04	-0,16	-0,25	1,00	-0,06	-0,16	-0,01	-0,10	0,13	0,16	0,58	-0,11	0,04	0,04	-0,19	-0,01	-0,21	0,11	0,27	0,03	0,41	-0,18	-0,01	0,00	-0,16	0,03	-0,22	-0,20	-0,04	-0,24
Cu	0,11	0,07	0,66	0,10	0,29	-0,06	0,66	0,74	0,36	-0,06	1,00	0,77	0,68	0,19	0,02	-0,03	0,15	0,74	0,35	0,02	0,44	-0,16	0,41	-0,07	-0,25	0,19	-0,11	0,76	0,14	0,08	0,32	-0,41	0,43	0,77	0,58	0,27
Fe	0,02	-0,04	0,80	0,09	0,34	-0,13	0,72	0,89	0,58	-0,16	0,77	1,00	0,79	0,28	-0,01	0,00	0,11	0,79	0,44	-0,03	0,50	0,01	0,55	-0,13	-0,21	0,15	-0,15	0,83	0,11	0,27	0,42	-0,41	0,66	0,91	0,66	0,48
Ga	0,01	-0,05	0,90	0,04	0,22	-0,12	0,69	0,74	0,52	-0,01	0,68	0,79	1,00	0,24	0,04	-0,02	0,27	0,71	0,29	-0,10	0,36	0,21	0,51	-0,12	-0,07	0,06	-0,14	0,78	0,06	0,44	0,41	-0,38	0,60	0,79	0,53	0,48
Ge	-0,02	-0,02	0,26	0,06	0,43	-0,11	0,35	0,30	-0,10	0,19	0,28	0,24	1,00	-0,05	0,17	-0,03	0,32	0,11	-0,04	0,29	0,12	0,35	-0,06	0,06	-0,03	-0,08	0,26	0,04	0,20	0,28	-0,07	0,41	0,24	0,26	0,36	
Hg	0,10	0,11	0,02	0,12	0,16	0,15	-0,08	-0,05	-0,04	0,13	0,02	-0,01	0,04	-0,05	1,00	-0,01	0,06	-0,07	0,01	0,07	-0,11	-0,01	-0,09	0,10	0,04	0,19	0,15	-0,05	0,09	-0,01	-0,06	-0,11	-0,16	-0,04	-0,02	-0,13
K	0,01	0,08	0,06	0,17	0,42	0,11	0,22	0,00	-0,02	0,16	-0,03	0,00	-0,02	0,17	-0,01	1,00	0,00	0,15	0,08	0,09	0,17	0,07	0,18	0,13	0,64	0,21	0,09	-0,10	0,06	0,12	0,27	0,11	0,19	-0,15	0,14	0,10
Li	0,13	0,11	0,12	0,08	0,21	0,06	0,09	0,11	-0,06	0,58	0,15	0,11	0,27	-0,03	0,06	0,00	1,00	0,17	0,16	0,02	-0,14	0,01	-0,06	0,04	0,07	0,01	0,18	0,06	0,00	0,07	-0,09	-0,13	-0,06	0,05	0,18	-0,12
Mg	0,03	-0,04	0,75	0,09	0,06	-0,17	0,77	0,81	0,56	-0,11	0,74	0,79	0,71	0,32	-0,07	0,15	0,17	1,00	0,35	-0,06	0,48	0,04	0,73	-0,17	-0,12	0,06	-0,16	0,67	0,06	0,18	0,53	-0,40	0,64	0,70	0,63	0,40
Mn	0,03	0,03	0,29	0,07	-0,09	-0,08	0,35	0,55	0,13	0,04	0,35	0,44	0,29	0,11	0,01	0,08	0,16	0,35	1,00	0,05	0,24	-0,12	0,19	0,00	-0,06	0,02	-0,03	0,31	0,02	0,04	0,13	-0,26	0,22	0,36	0,47	0,05
Mo	0,34	0,44	-0,13	0,17	0,21	0,26	-0,13	-0,07	-0,09	0,04	0,02	-0,03	-0,10	-0,04	0,07	0,09	0,02	-0,06	0,05	1,00	-0,11	-0,13	-0,08	0,40	0,03	0,26	0,10	-0,12	0,17	-0,11	-0,10	0,00	-0,16	-0,11	0,18	-0,14
Na	-0,06	-0,11	0,48	-0,02	0,29	-0,20	0,69	0,49	0,31	-0,19	0,44	0,50	0,36	0,29	-0,11	0,17	-0,14	0,48	0,24	-0,11	1,00	-0,06	0,38	-0,17	-0,08	-0,04	-0,15	0,52	-0,04	0,15	0,34	-0,14	0,57	0,54	0,36	0,40
Nb	-0,12	-0,09	0,19	-0,08	0,37	-0,11	0,04	0,03	0,32	-0,01	-0,16	0,01	0,21	0,12	-0,01	0,07	0,01	0,04	-0,12	-0,13	-0,06	1,00	0,28	-0,05	0,19	-0,05	-0,07	0,00	-0,04	0,46	0,19	-0,01	0,40	0,00	0,00	0,34
Ni	-0,03	-0,06	0,63	0,08	-0,02	-0,19	0,58	0,66	0,74	-0,21	0,41	0,55	0,51	0,35	-0,09	0,18	-0,06	0,73	0,19	-0,08	0,38	0,28	1,00	-0,14	0,02	-0,01	-0,17	0,43	-0,01	0,27	0,66	-0,22	0,65	0,45	0,42	0,51
Pb	0,28	0,47	-0,16	0,10	0,30	0,35	-0,20	-0,17	-0,12	0,11	-0,07	-0,13	-0,12	-0,06	0,10	0,13	0,04	-0,17	0,00	0,40	-0,17	-0,05	-0,14	1,00	0,11	0,25	0,14	-0,17	0,13	-0,03	-0,12	0,09	-0,19	-0,20	0,17	-0,12
Rb	0,00	0,05	-0,07	0,07	-0,02	0,11	-0,10	-0,18	-0,05	0,27	-0,25	-0,21	-0,07	0,06	0,04	0,64	0,07	-0,12	-0,06	0,03	-0,08	0,19	0,02	0,11	1,00	-0,04	0,11	-0,21	-0,03	0,26	0,10	0,26	0,00	-0,28	-0,07	0,05
S	0,12	0,22	0,04	0,23	-0,14	0,34	-0,05	0,02	0,00	0,03	0,19	0,15	0,06	-0,03	0,19	0,21	0,01	0,06	0,02	0,26	-0,04	-0,05	-0,01	0,25	-0,04	1,00	0,08	0,01	0,38	-0,14	0,00	-0,22	-0,11	0,02	0,08	-0,12
Sb	0,12	0,13	-0,18	0,21	0,31	0,18	-0,14	-0,17	-0,18	0,41	-0,11	-0,15	-0,14	-0,08	0,15	0,09	0,18	-0,16	-0,03	0,10	-0,15	-0,07	-0,17	0,14	0,11	0,08	1,00	-0,19	0,05	-0,05	-0,13	0,06	-0,19	-0,18	-0,05	-0,19
Sc	-0,02	-0,09	0,77	-0,01	-0,01	-0,20	0,70	0,78	0,51	-0,18	0,76	0,83	0,78	0,26	-0,05	-0,10	0,06	0,67	0,31	-0,12	0,52	0,00	0,43	-0,17	-0,21	0,01	-0,19	1,00	0,05	0,32	0,33	-0,22	0,64	0,88	0,50	0,60
Se	0,12	0,13	0,02	0,12	0,46	0,33	-0,03	0,03	0,02	-0,01	0,14	0,11	0,06	0,04	0,09	0,06	0,00	0,06	0,02	0,17	-0,04	-0,04	-0,01	0,13	-0,03	0,38	0,05	0,05	1,00	-0,04	-0,02	-0,09	-0,06	0,04	0,05	-0,04
Sn	-0,06	-0,10	0,37	-0,04	0,34	-0,06	0,23	0,23	0,40	0,00	0,08	0,27	0,44	0,20	-0,01	0,12	0,07	0,18	0,04	-0,11	0,15	0,46	0,27	-0,03	0,26	-0,14	-0,05	0,32	-0,04	1,00	0,20	0,09	0,47	0,28	0,26	0,50
Sr	-0,06	-0,07	0,55	0,06	0,01	-0,16	0,55	0,47	0,44	-0,16	0,32	0,42	0,41	0,28	-0,06	0,27	-0,09	0,53	0,13	-0,10	0,34	0,19	0,66	-0,12	0,10	0,00	-0,13	0,33	-0,02	0,20	1,00	-0,16	0,56	0,36	0,29	0,40
Th	-0,06	-0,03	-0,37	-0,07	0,42	0,03	-0,30	-0,37	-0,16	0,03	-0,41	-0,41	-0,38	-0,07	-0,11	0,11	-0,13	-0,40	-0,26	0,00	-0,14	-0,01	-0,22	0,09	0,26	-0,22	0,06	-0,22	-0,09	0,09	-0,16	1,00	-0,15	-0,36	-0,36	0,12
Ti	-0,10	-0,13	0,67	-0,02	0,18	-0,29	0,73	0,65	0,65	-0,22	0,43	0,66	0,60	0,41	-0,16	0,19	-0,06	0,64	0,22	-0,16	0,57	0,40	0,65	-0,19	0,00	-0,11	-0,19	0,64	-0,06	0,47	0,56	-0,15	1,00	0,66	0,50	0,76
V	-0,02	-0,10	0,79	-0,01	-0,02	-0,21	0,71	0,83	0,56	-0,20	0,77	0,91	0,79	0,24	-0,04	-0,15	0,05	0,70	0,36	-0,11	0,54	0,00	0,45	-0,20	-0,28	0,02	-0,18	0,88	0,04	0,28	0,36	-0,36	0,66	1,00	0,56	0,49
Zn	0,22	0,12	0,51	0,11	0,51	-0,03	0,54	0,62	0,42	-0,04	0,58	0,66	0,53	0,26	-0,02	0,14	0,18	0,63	0,47	0,18	0,36	0,00	0,42	0,17	-0,07	0,08	-0,05	0,50	0,05	0,26	0,29	-0,36	0,50	0,56	1,00	0,31
Zr	-0,09	-0,11	0,53	-0,03		-0,21	0,50	0,47	0,59	-0,24	0,27	0,48	0,48	0,36	-0,13	0,10	-0,12	0,40	0,05	-0,14	0,40	0,34	0,51	-0,12	0,05	-0,12	-0,19	0,60	-0,04	0,50	0,40	0,12	0,76	0,49	0,31	1,00