ФГБУ «ВИМС»

ГЕОХИМИЧЕСКИЕ И ДР. ТЕХНОЛОГИИ, МЕТОДЫ И МЕТОДИКИ ПРИ ПРОГНОЗИРОВАНИИ И ПОИСКАХ МЕСТОРОЖДЕНИЙ

(преимущественно «скрытого» типа)

№ 4

Редактор-составитель: В.В. Коротков

СОДЕРЖАНИЕ:

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ГЕОХИМИЧЕСКИХ ИССЛЕДОВАНИЙ

(Классификация металлогенического плодородия в дуговых магмах и понимание формирования медно-порфировых месторождений) *[1]*

1. Введение

Свиты магматических пород, связанные с порфирово-медными рудами, обычно характеризуются отчетливой цельной геохимической сигнатурой, которая была разработана в качестве индикатора металлогенической "плодородности", что означает, что магмы с такими сигнатурами могут быть предрасположены к образованию порфирово-медной минерализации. Было предложено несколько признаков плодородия магмы, включая высокий Sr / Y, высокий La / Yb, высокий Eu / Eu *, высокий Sr / MnO ивысокий Al_2O_3/ TiO_2, которые все чаще используются при разведке порфировой меди. Считается, что особый химический состав обусловлен процессами, которые могут быть важны для образования магм, образующих порфировые отложения меди, включающими сильно сжимающие тектонические режимы, которые способствуют утолщению коры и длительному хранению магмы на глубоких уровнях. Это вызывает дифференциацию магм при высоком давлении и расплаве, содержание H_2O, стабилизирующее амфибол ± гранат, в котором совместимы Y, MREEs и HREEs, но подавляющее плагиоклаз, в котором совместимы Sr и Eu. Это приводит к образованию пород, обладающих указанными выше характеристиками, которые затем можно отличить от обычных дуговых пород, используя двумерные пороговые значения. Такие методы классификации полезны, но ограничены, поскольку они игнорируют дополнительные переменные, которые могут содержать сигналы фертильности и приводить к ложноположительным результатам, поскольку другие процессы могут генерировать аналогичные геохимические сигнатуры. Некоторые параметры также подвержены изменению в результате гидротермальных изменений. Кроме того, ложноотрицательные результаты также распространены в порфировых породах, например, в менее развитых составах (<65 мас.% SiO_2) и в более изменчивых тектонических условиях, связанных со щелочными магмами богатыми золотом.

Машинное обучение - это наука об использовании компьютеров для обучения на основе данных. За последние несколько десятилетий были разработаны алгоритмы машинного обучения для выявления закономерностей и тенденций в различных наборах данных и составления прогнозов. Контролируемое классификационное обучение является ответвлением этого, где входным данным присваивается метка класса и машина обучается предсказывать метку класса с использованием входных данных. Геологическим примером является классификация тектонических условий

базальтов с использованием химии цельных пород. Такие методы имеют значительный потенциал в разведке полезных ископаемых, поскольку наборы данных становятся все более большими, со значительным количеством наблюдений (т.е. анализов) и характеристик (т.е. анализируемых веществ). Ранее подчеркивалось применение таких методов к различным типам данных для прогнозирования перспективности полезных ископаемых, таких как гиперспектральное картирование, литологическое картирование, структурное картирование, геохимия почв и литогеохимия. Существует множество контролируемых алгоритмов машинного обучения, но наиболее успешными и широко используемыми алгоритмами в разведке полезных ископаемых и геохимии являются логистическая регрессия, деревья решений (включая случайный лес), машины опорных векторов и искусственные нейронные сети.

В этом исследовании четыре контролируемых алгоритма машинного обучения (логистическая регрессия, искусственные нейронные сети, машины опорных векторов и случайный лес) были применены для классификации металлогенического плодородия горных пород при компиляции глобальных данных о всей породе, чтобы отличить образцы, пространственно и временно связанные с порфировым медным оруденением, от образцов, не связанных с минерализацией. Целью данного исследования является:

(1) продемонстрировать потенциал таких методов для разведки порфирового оруденения в магматических дугах и количественно оценить любое улучшение по сравнению с существующими, в основном двумерными методами;

(2) сравнить эффективность каждого метода классификации;

(3) установить, можно ли создать высокопроизводительный классификатор независимо от сродства к магме; и

(4) определить наиболее важные параметры различия плодородия магмы и обсудить их влияние на формирование порфирового оруденения.

2. Методы

2.1 Сбор данных и контроль качества

Геохимические данные по порфировым месторождениям были собраны из литературы (таблица 1). Был включен диапазон размеров и типов депозитов, чтобы гарантировать, что модели машинного обучения способны обучаться и прогнозировать рудоносность независимо от размера или типа депозита. Данные были случайным образом отобраны с уменьшением выборки, чтобы гарантировать, что ни один отдельный вклад не составлял более 100 наблюдений в наборе данных, чтобы

уменьшить смещение из-за чрезмерной репрезентации. В собранных данных анализам цельных пород было присвоено сродство магмы (от известково-щелочной к шошонитовой), чтобы можно было сравнить химический состав их цельных пород. Эти сродства к магме были получены из литературы.

Таблица 1 **Месторождения порфировой меди, включенные в набор геохимических данных по целым породам, используемый в наборе обучающих данных для методов машинного обучения.** *(CA = связан с кальциево-щелочными магмами, K = связан с кальциево-щелочными магмами с высоким содержанием K до шошонитовых. Возрастные диапазоны включают магматизм , который предшествует минерализации > 2 млн лет)*

Deposit	Country	Magma affinity	Age range (Ma)	Tonnage	Cu grade (%)	Mo grade (%)	Au grade (g/t)
Altar	Argentina	K	12–10	802	0.42	-	0.06
Almalyk	Uzbekistan	K	326–312	6080	0.39	0.0023	0.37
Andacolla	Chile	K	104	417	0.34	-	0.12
Balsapamba	Ecuador	CA	22	-	<0.1	-	-
Batu Hijau	Indonesia	K	15–13	1644	0.44	0	0.35
Bingham Canyon	USA	K	38–37	3230	0.882	0.053	0.38
Canicapa	Ecuador	CA	20	-	-	-	-
Chaucha	Ecuador	CA	24–10	363	0.4	0.03	0
Chuquicamata	Chile	CA	35–33	21,277	0.592	0.04	0.013
Corocco-huayco	Peru	CA	40–35	155	1.57	0	0.33
Cuellaje	Ecuador	CA	NA	-	-	-	-
Don Manuel	Chile	CA	4–3	-	-	-	-
Dos Amigos-Tricolor	Chile	K	107	36	0.36		
El Abra	Chile	CA	63–37	1779.4	0.494	0.0058	0
El Salvador	Chile	CA	44–42	3836.3	0.447	0.022	0.1
El Teniente	Chile	CA	24–3	20,731	0.62	0.019	0.005
Escondida	Chile	CA	268–37	11,158	0.769	0.0062	0.25
Gaby-Papa Grande	Ecuador	CA	21	308	0.09	0.025	0.73

Junin	Ecuador	CA	9	319	0.71	0.026	0
Kadjaran	Armenia	K	34–21	1700	0.27	0.055	0.65
La Colosa	Colombia	K	8	821	0.11	0.017	0.8
Ministro Hales	Chile	CA	210–33	1249	0.68	-	-
Northparkes	Australia	K	452–436	472	0.56	0	0.19
Ok Tedi	Papau New Guinea	K	1.5–1.1	854	0.64	0.011	0.78
Pebble	Alaska (USA)	K	100–41	7510	0.416	0.024	0.33
Productora	Chile	K	130	214.3	0.48	0.0138	0.1
Qulong	Tibet (China)	K	18–15	1517	0.52	0.032	0
Radomiro Tomic	Chile	CA	39–34	4980	0.39	0.015	0
Relincho	Chile	CA	100–64	581	0.43	0.018	0
Rio Blanco-Los Bronces	Chile	CA	19–4	16,816	0.601	0.02	0
Sarycheku	Uzbekistan	K	338–313	200	0.5	0	0.1
Tampakkan	Phillipines	K	14–0.2	2500	0.48	0	0.2
Telimbela	Ecuador	CA	22	-	-	-	-

Данные не связанные с месторождениями порфировой меди, были проанализированы из базы данных GEOROC (http://georoc.mpch-mainz.gwdg.de/georoc) для дуг Анд, Сулавеси, Лусона, Банды и Соломона. Эти дуги были выбраны потому, что они, как известно, содержат порфировые м-ния и перспективы различных типов и представляют собой целый ряд тектонических условий и магматических сродств. Эти наборы данных были отфильтрованы, чтобы исключить любые наблюдения, помеченные как осадочные породы, метаморфические породы, перидотиты или жильные породы.

Основными и микроэлементами, выбранными для обобщенных данных, были Si, Al, Fe (в пересчете на Fe^{2+}), Mg, Ca, Na, K, Ti, Mn, Sr, Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb и Lu. Эти признаки были выбраны на основе: (1) знания о том, что они проявляют высокую вариабельность во время петрогенетических процессов; (2) того факта, что они эффективно различают порфирово-медно-магматические породы; и/или (3) о них часто сообщается в литературе. Дополнительные элементы (V, Sc, Nb и Zr) действуют как полезные дискриминанты. Список ограничен элементами, которые

имеют наиболее полные записи в базе данных и в наименьшей степени подвержены неполному анализу и недостающим значениям.

Использование составных наборов данных (GEOROC), содержащих многочисленные источники данных, требует контроля качества, поскольку каждый набор данных был собран с использованием различных методов (различная подготовка и инструментарий). Проанализированные породы также будут содержать различные степени гидротермальных изменений, особенно для пород, полученных из медных порфировых систем. Чтобы гарантировать, что компиляция содержала приемлемые данные, были установлены критерии отбора <3,5 мас.% потерь при воспламенении (LOI) и аналитические итоги 97,5–101,5 мас.% (c.f.). Данные также были отфильтрованы для удаления анализов, которые были нормативными для полевых шпатов или содержали> 3% нормативного корунда, поскольку они, вероятно, указывают на значительные гидротермальные изменения или, в случае высокого нормативного корунда, могут указывать на накопление плагиоклаза. После фильтрации данных компиляция из порфировых отложений Cu (не включая данные GEOROC) включала 555 наименее измененных наблюдений из 41 порфировой системы Cu (таблица 1; Электронные дополнительные материалы (ESM), рис. 1). Отфильтрованная база данных, полученная из GEOROC, состояла из 3 559 наблюдений.

2.2 Обработка выбросов

Рассмотрение обработки выбросов необходимо при обучении алгоритмов машинного обучения, поскольку они могут быть очень чувствительны к диапазону и распределению данных. Однако выбор метода обработки удаленных точек данных является сложной задачей, поскольку выбросы могут иметь несколько источников. Выбросы, возникающие из-за аналитической или человеческой ошибки, не будут отражать закономерности в наборе данных и их лучше всего отбросить. Однако статистические выбросы также могут возникать как естественный продукт различных геологических процессов и их отбрасывание может в конечном итоге привести к искажению моделей. Распространенные методы идентификации выбросов, такие как стандартное отклонение или метод Тьюки, основаны на нормальном распределении данных и поэтому плохо подходят для геохимических данных, которые редко демонстрируют нормальное или логарифмически нормальное распределение. Например, отбрасывание данных выше пороговых значений потенциального выброса и экстремальных выбросов для Sr (т.е. точек данных, которые в три раза больше межквартильного расстояния от медианы по методу Тьюки) приведет к удалению

многих наблюдений Sr, которые могут отражать менее распространенный, но, тем не менее, важный геологический процесс. Поскольку для магматических пород, связанных с порфировым оруденением, может быть характерен высокий Sr, это внесло бы смещение в модели машинного обучения. Это может дополнительно снизить способность моделей выявлять геологически значимые процессы посредством анализа признаков. Следовательно, мы не отфильтровываем выбросы из данных, а редкие или экстремальные выбросы рассматриваются как "шум", которому алгоритмы машинного обучения должны придавать небольшое значение.

2.3 Обработка пропущенных значений

Другой проблемой является обработка пропущенных значений в наборах данных, поскольку большинство алгоритмов машинного обучения не могут быть применены в таких случаях. Определение оптимальной стратегии работы с недостающими данными является сложной задачей, поскольку они могут возникать по множеству причин. Например, в геохимических наборах данных могли использоваться различные аналитические пакеты, которые, возможно, не все были способны определить полный диапазон элементов (отсутствующие значения), или могут быть данные ниже предела обнаружения (цензурированные значения).

Простой подход заключался бы в удалении наблюдений, содержащих пропущенные значения. Однако это может значительно уменьшить размер набора данных, поскольку пропущенные значения для нескольких элементов часто встречаются в геохимических компиляциях с многочисленными переменными источниками данных. Кроме того, удаление наблюдений с отсутствующими данными может привести к предвзятости, например, из-за того, что некоторые источники данных (например, отраслевые или академические) занижают конкретные элементы. Другим распространенным методом является замена отсутствующих значений (‘вменение’ их) средним / медианным значением, полученным из остальных данных или из заданного подмножества данных. Существуют более сложные методы вменения, такие как те, которые используют регрессию с несколькими переменными, подходы ближайших соседей и непараметрические методы, подходящие для композиционных данных. Важно отметить, что в настоящее время существуют методы классификации, которые способны обрабатывать пропущенные значения в наборах данных, такие как древовидная система XGBoost. Для простоты не включали наблюдения, если они содержали пропущенные или подвергнутые цензуре значения для выбранных элементов, за исключением частичных пробелов в данных о редкоземельных элементах, которые могут быть точно интерполированы из соседних

редкоземельных элементов. Однако надлежащая обработка пропущенных значений имеет решающее значение для отраслевых задач классификации или для небольших обучающих наборов данных, где потеря данных невозможна.

2.4 Определение и классификация фертильности.

Дуговые магмы, которые предрасположены к образованию порфировых м-ний меди, называются ‘плодородными’. Тем не менее, исследования районного масштаба показали, что химические признаки фертильности характерны не только для синеминерализационных интрузий, но и могут быть обнаружены в породах, которые непосредственно предшествуют и после минерализации, а также в породах, которые образовались за несколько миллионов лет до минерализации, таких как вмещающие батолиты (рис. 1). Поэтому в идеале ‘фертильность’ следует рассматривать как вероятностный показатель. Однако для этого первоначального анализа выбрали более простой подход с бинарной маркировкой ("плодородный" или "неплодородный"), но это требует объективного определения в процессе обучения, чтобы обеспечить эффективную модель классификации, подходящую для приложения, для которого она разрабатывается. Поэтому классифицировали любую магматическую породу из наборов образцов, связанных с порфиром (из порфировых районов в масштабе ~ 10 км), которые образовались за 2 млн. лет до или после минерализации, как "плодородные", а породы из того же района, которые старше или моложе этого окна, как ‘неплодородные’. Этот интервал в 2 млн. лет был выбран потому, что, согласно недавним исследованиям, развитие "плодородного" химического состава магмы происходит в течение 2 млн. лет после начала минерализации в порфировых районах Cu, например, в Квеллавеко, Перу (рис. 1). Этот временной масштаб также эквивалентен максимальной продолжительности минерализации порфировой меди в отдельных центрах. Временная схема классификации была реализована с использованием опубликованной геохронологии для возрастов магматических событий и минерализации для каждого включенного месторождения (Таблица 1). Основываясь на этом критерии, набор данных porphyry Cu состоял из 440 наблюдений за плодородием и 115 наблюдений за бесплодием. В базе данных GEOROC все наблюдения, полученные из известных систем porphyry, были перенесены в базу данных porphyry Cu, а остальные наблюдения были помечены как ‘бесплодные’. Предполагается, что это возрастное окно для образцов порфирного района не определяет точное начало развития "фертильных" сигналов, в чем трудно быть уверенным на основе текущих исследований и которые могут отличаться от участка к участку, а скорее обеспечивает объективный критерий для процесса

обучения. Результатами моделей машинного обучения являются различия, которые относятся к тому же определению "плодородия", то есть модели, в случае успеха, должны быть способны распознавать породы, которые имеют умеренно тесную пространственную и временную связь с потенциальным событием минерализации. Сам по себе этот подход не может предсказать, действительно ли образовалось месторождение, поскольку многие дополнительные факторы, которые не включены в этот подход, могут влиять на формирование порфировой медной минерализации. Эти факторы включают тип и доступность структурных каналов, реологию и состав земной коры, а также объем и продолжительность магматической активности.

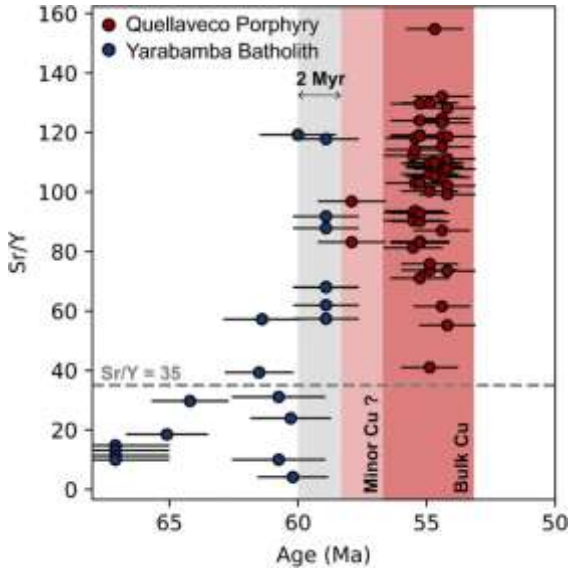

Рис. 1. Диаграмма рассеяния, показывающая Sr / Y цельной породы в районе Квеллавеко, на юге Перу в зависимости от возраста залегания, определенного по данным U–Pb LA ICP-MS циркона. Столбики ошибок - это 2σ ошибок взвешенного среднего с распространенной систематической неопределенностью 2%. Пунктирная линия указывает ‘плодородный’ порог для Sr / Y. Серые области указывают на периоды времени основной минерализации меди (более темный красный) и предполагаемой незначительной минерализации меди (более бледно-красный). График показывает, что высокие составы Sr / Y ("плодородные") присутствовали в ~ 2 млн. лет до начала минерализации Cu (область серым цветом) и достигли максимума во время основной минерализации Cu

3. Полноразмерное изображение.

3.1 Проблема композиционных данных.

Геохимические данные являются примером данных о составе, где каждый компонент не является абсолютной величиной, а представляет собой относительные значения, которые в сумме составляют константу C (например, 100%). Данные о составе обычно выражаются в виде суммы k отдельных компонентов xi, и для данной композиции, если бы один компонент xi должен был увеличиться, это потребовало бы уменьшения других компонентов xk−1, чтобы сохранить постоянную сумму. Классическим примером являются геохимические данные по целым породам, которые обычно выражаются в процентах или в частях на миллион, что в сумме составляет

константу (например, 100% или 106). Эффект постоянной суммы приводит к ложным корреляциям, таким как смещение в сторону отрицательных корреляций между компонентами, которые в противном случае положительно коррелируют. Например, традиционная ‘диаграмма изменения Харкера SiO_2 по сравнениюс Al_2O_3 для магматических пород обычно показывает отрицательную корреляцию (рис. 2), традиционно интерпретируемый как указывающий на фракционирование плагиоклаза. Однако эта тенденция может быть частично или полностью ложной, поскольку увеличение SiO_2 от 50 до 75 мас.% должно сопровождаться уменьшением вдвое других компонентов, чтобы сохранить постоянную сумму в 100%. Кроме того, состав магматических пород обычно дает ненормальные распределения. Эти свойства исключают применение традиционных статистических методов к необработанным композиционным данным, поскольку многие статистические методы предполагают, что переменные изменяются независимо друг от друга и что они удовлетворяют нормальным распределениям.

Признание взаимозависимого и непараметрического характера композиционных данных привело к введению логарифмических преобразований. Центрированное логарифмическое отношение (clr) для отдельного компонента композиции xi получается путем деления этого компонента на среднее геометрическое значение, g(x), всех компонентов композиции x и взятия натурального логарифма этого отношения (таблица 1 для примера расчета):

$$clr(x_i) = \ln\left(\frac{x_i}{g(x)}\right) \quad (1)$$

Преобразование clr позволяет переопределенным компонентам изменяться в неограниченном реальном пространстве, а не ограничиваться постоянной суммой, и, таким образом, позволяет применять методы многомерной статистики к преобразованным данным. Такая предварительная обработка особенно ценна в этом исследовании, где различные степени гидротермальных изменений могут вызвать значительную потерю / увеличение массы, что приводит к дополнительным ложным корреляциям в наборе данных. Еще одна полезная характеристика clr-координаты заключаются в том, что они чувствительны к относительной, а не абсолютной дисперсии. Например, в свите пород, где SiO_2 изменяется от 60 до 65 мас.%, а CaO - от 1 до 6 мас.%, оба компонента имеют абсолютную дисперсию 5 мас.%, но относительные отклонения составляют 1,08 и 5 соответственно. Последнее фиксируется с помощью clr-координат. Хотя случайный лес является непараметрическим (т. е. может обрабатывать взаимозависимые и ненормально

распределенные данные), в этом исследовании преобразовали все данные в clr для обеспечения согласованности, используя пакет pyrolite на Python.

Обычно наборы данных масштабируются перед обучением моделей машинного обучения, чтобы каждый признак имел среднее значение 0 и стандартное отклонение 1. Это может быть выгодно тем, что позволяет каждой функции вносить пропорциональный вклад в процессе обучения и может сократить время вычислений, необходимое для определенных методов (например, алгоритмы градиентного спуска в искусственных нейронных сетях). Однако здесь мы не масштабируем объекты, потому что все объекты имеют одинаковые единицы измерения, и устранение дисперсии может привести к потере информации.

3.2 Классовый дисбаланс.

Многие проблемы классификации связаны с классами, которые не содержат равного количества наблюдений; это, вероятно, часто встречается при разведке полезных ископаемых, поскольку рудные месторождения представляют собой редкие геохимические аномалии в земной коре. В этом исследовании доля данных из порфировых систем (9%) намного ниже, чем из ‘неплодородных’ магматических пород (91%). Это может привести к искажению точности классификации, потому что, в нашем случае, если классификатор всегда предсказывал "неоплодотворенный", то классификатор вводил бы в заблуждение точность классификации в 91%, несмотря на то, что никогда не классифицировал правильно какие-либо наблюдения за плодородием. Одним из способов избежать этой проблемы является использование альтернативных показателей производительности, таких как матрицы путаницы и кривые рабочих характеристик приемника. Однако классовый дисбаланс все еще может привести к моделям, которые благоприятствуют прогнозированию класса большинства и, следовательно, имеют более высокую вероятность неправильной классификации меньшинства. Было показано, особенно для небольших и сложных наборов данных, что дисбаланс классов может снизить производительность деревьев решений, нейронных сетей и машин с опорными векторами.

Хотя дисбаланс классов в нашем наборе данных (91% бесплодных, 9% плодородных) не является экстремальным по сравнению с возможным в других задачах машинного обучения, сбалансированное распределение классов является оптимальным. Простейшими подходами к достижению этой цели являются занижение выборки класса большинства или завышение выборки класса меньшинства для создания сбалансированной по классам базы данных. Хотя было показано, что чрезмерная выборка повышает производительность и позволяет избежать потери

данных, присущей недостаточной выборке, в этом исследовании использовалась недостаточная выборка большинства, поскольку было обнаружено, что чрезмерная выборка приводит к переоснащению моделей (рис. 3).

3.3 Обобщение модели

В конечном счете, контролируемые модели машинного обучения нацелены на то, чтобы хорошо работать при прогнозировании результатов на основе данных, не встречающихся в процессе обучения. Это достигается, когда модель может обобщать данные обучения, а не переобучать или недооценивать. Переобучение обучающих данных происходит, когда модель фиксирует дисперсию или "шум" в наборе данных, а не базовое распределение данных, что означает, что она не сможет надежно предсказать будущие наблюдения (рис. 4). Такие модели имеют высокую погрешность "дисперсии", что означает, что они чувствительны к небольшим колебаниям в обучающих данных. Недостаточная подгонка возникает, когда модель не полностью отражает базовое распределение данных, поскольку она недостаточно сложна (рис. 4). Такие модели имеют высокую ошибку ‘смещения’, что означает, что модель чрезмерно упрощает проблему из-за неверных допущений в процессе обучения. Уменьшение дисперсии (уменьшение сложности модели) приводит к увеличению смещения, и наоборот (известный как "компромисс между отклонением и отклонением"); следовательно, необходимо искать оптимальный баланс, чтобы уменьшить общую ошибку в модели. Если рассматривать двумерный фиктивный набор данных, содержащий два класса данных (рис. 4), оптимальный классификатор будет лучше классифицировать невидимые тестовые данные по сравнению с классификатором overfit, который моделирует шум, или классификатором underfit, который упрощает. Выбор характеристик, уменьшение размерности и настройка гиперпараметров являются примерами методов, которые могут быть использованы в процессе обучения для предотвращения переобучения и недообучения и, таким образом, уменьшения общей ошибки.

3.4 Уменьшение размерности и анализ главных компонентов.

Хотя в центре внимания этого исследования находится контролируемое машинное обучение, обычно контролируемым методам предшествует неконтролируемый этап уменьшения размерности. Неконтролируемые методы - это те, которые изучают шаблоны из немаркированных данных. Уменьшение размерности уменьшает набор данных до меньшего числа параметров, которые могут представлять ковариации в исходном наборе данных. Этот метод ценен тем, что с увеличением числа признаков (таких как композиционные переменные) полные данные, как правило, становятся все

более разреженными по мере увеличения евклидова расстояния между точками данных. Разреженность обычно увеличивается экспоненциально с увеличением количества объектов, что требует чрезвычайно большого количества наблюдений для покрытия пространства высокой размерности. Применение методов контролируемого машинного обучения к таким наборам данных может быть сложной задачей, а сгенерированные модели очень подвержены переоснащению. Хотя такие эффекты могут быть небольшими при наборе данных рассматриваемого здесь размера, и аналогичные точные результаты получаются с уменьшением размерности и без него (рис. 5), мы предпочитаем включить этот этап в качестве наилучшей практики. Кроме того, геологическая интерпретируемость также может быть получена на основе этого этапа (нагрузки основных компонентов и оценки).

Используем анализ основных компонентов (PCA) в clr-преобразованные данные для сокращения объектов в геохимической базе данных до меньшего числа объектов, которые являются репрезентативными для ковариационной структуры набора данных. Главные компоненты (ПК) представляют собой линейные комбинации исходных переменных. Первый ПК вычисляется с использованием метода наименьших квадратов, чтобы найти плоскость, которая учитывает максимальную величину дисперсии в наборе данных. Следовательно, второй ПК ортогонален первому компоненту, чтобы показать размерность второй по величине дисперсии. Дальнейшие ПК рассчитываются таким же образом. Эти операции увеличивают соотношение сигнал / шум в наборе данных, что может помочь в более эффективной классификации. Кроме того, они ортонормированы (т.е. статистически независимы) и отражают линейные процессы, которые могут быть отнесены к геологическим процессам. Каждому наблюдению присваивается факторный балл для каждого ПК, который представляет степень дисперсии наблюдения в измерении ПК. Эти коэффициенты используются в качестве исходных данных для машинного обучения. PCA был реализован с использованием функции PCA в scikit learn на Python. Только первые 6 ПК, на долю которых приходится ~ 90% дисперсии обучающего набора данных (рис. 6), включены в модели. Удаление дополнительных ПК уменьшает шум в обучающем наборе данных и, следовательно, делает модели менее подверженными переобучению (рис. 7).

3.5 Контролируемые методы машинного обучения

Описания методов машинного обучения, используемых здесь, представляют собой краткое изложение тех, которые были найдены в Hastie et al. (2009) Alpaydın (2014), Bell (2014) and Kubat (2017). Для полного обсуждения этих методов необходимо

обратиться к этим текстам и ссылкам, приведенным в них. Все методы машинного обучения были применены с использованием scikit learn, пакета машинного обучения, закодированного на Python.

3.6 Логистическая регрессия.

Логистическая регрессия - один из самых простых методов контролируемого машинного обучения. Он использует логистическую (сигмоидальную) функцию для прогнозирования метки двоичного класса на основе линейной комбинации одной или нескольких независимых переменных. Для примера набора данных (рис. 2a) вероятность наблюдения X (например, состава цельной породы), принадлежащего классу Y (где Y может быть либо плодородным, либо неплодородным), будет смоделирована с использованием логистической функции для вычисления вероятности от 1 до 0:

$$P(Y|X) = \frac{1}{1 + e^{-f(x)}} \quad (2)$$

где линейная комбинация признаков f(x) представляет собой сумму каждого отдельного признака x (например, каждого компонента в объемном составе породы), умноженную на весовой коэффициент w с добавленным членом смещения, или перехват b:

$$f(x) = \sum_{i=1}^{n} w_i x_i + b \quad (3)$$

Алгоритм численной оптимизации (который минимизирует целевую функцию) используется для выбора значений весовых коэффициентов для каждого признака, чтобы минимизировать ошибку классификации в процессе обучения. Полученная логистическая функция "наилучшего соответствия" затем может быть использована для прогнозирования вероятности неизвестного наблюдения, принадлежащего к классу Y (например, вероятности фертильности), и если вероятность превышает "пороговое значение’ (обычно 0,5—рис. 2a), он классифицируется как положительный (например, фертильный). Логистическая регрессия использовалась в целом ряде геологических исследований, с соответствующими примерами, включая картирование перспективности, картирование гидротермальных изменений с использованием литогеохимических данных и исследования происхождения обломочных пород.

3.7 Искусственные нейронные сети.

Искусственные нейронные сети (ANN) получили свое название из-за их архитектуры, аналогичной мозгу животных. Отдельным элементом искусственной нейронной сети является нейрон, который, подобно биологическому нейрону,

преобразует серию входных сигналов в выходной сигнал. Нейрон сначала принимает взвешенную сумму всех входных данных (в которой каждому входному элементу присваивается вес w в зависимости от его желаемого влияния на результат) и добавляет член смещения b (уравнение 3). Затем это проходит через функцию f (такую как сигмовидная, ReLU или tanh функция), называемую функцией активации, который преобразует значение в выходной сигнал, определяющий, насколько "активен" нейрон (вставка рис. 2в). Пример нейрона с xn входными функциями с соответствующими wn весами и членом смещения b может использовать сигмовидную функцию активации, такую как логистическая регрессия (уравнение 3). По сути, это преобразует входные данные в значение между заданным диапазоном (например, между 0 и 1, где 1 полностью активен, а 0 неактивен). Многослойные нейронные сети с прямой связью состоят из слоев, где каждый слой состоит из множества нейронов (рис. 2в). Сначала данные поступают с входного уровня (рис. 2c) и затем проходят через один или несколько "скрытых слоев", прежде чем попасть в выходной слой, где может быть произведена классификация (рис. 2c; плодородный или неоплодотворенный). Во время обучения модель нацелена на определение значений для w и b в каждом нейроне, который дает наиболее успешный результат классификации. Когда необученный ANN впервые получает обучающие данные и производит начальную классификацию, он вычисляет ошибку классификации с использованием функции стоимости. Алгоритм работает в обратном направлении по сети с помощью процесса, известного как обратное распространение, для корректировки весов и смещений, чтобы минимизировать затраты. Это достигается путем вычисления градиента для функции затрат и корректировки весов и отклонений в сторону минимумов для функции затрат (с использованием самого крутого градиента) с помощью градиентного спуска.

Процесс обратного распространения имеет значительное количество весов и смещений, которые должны быть скорректированы в каждом нейроне для достижения низкой ошибки классификации. ANN могут создавать чрезмерно сложные модели, что делает их склонными к переобучению, особенно там, где количество скрытых нейронов велико, или где задействованы зашумленные наборы данных, поэтому обычно требуются методы регуляризации для уменьшения переобучения (например, отсев). Нейронные сети были в центре внимания ряда исследований картирования перспективности полезных ископаемых с использованием различных данных, включая геохимию почвы, информацию о геологических картах и геофизику (Porwal et al. 2003; Rodriguez-Galiano et al. 2015; Zhang et al. 2019; Li et al. 2020, 2021)..

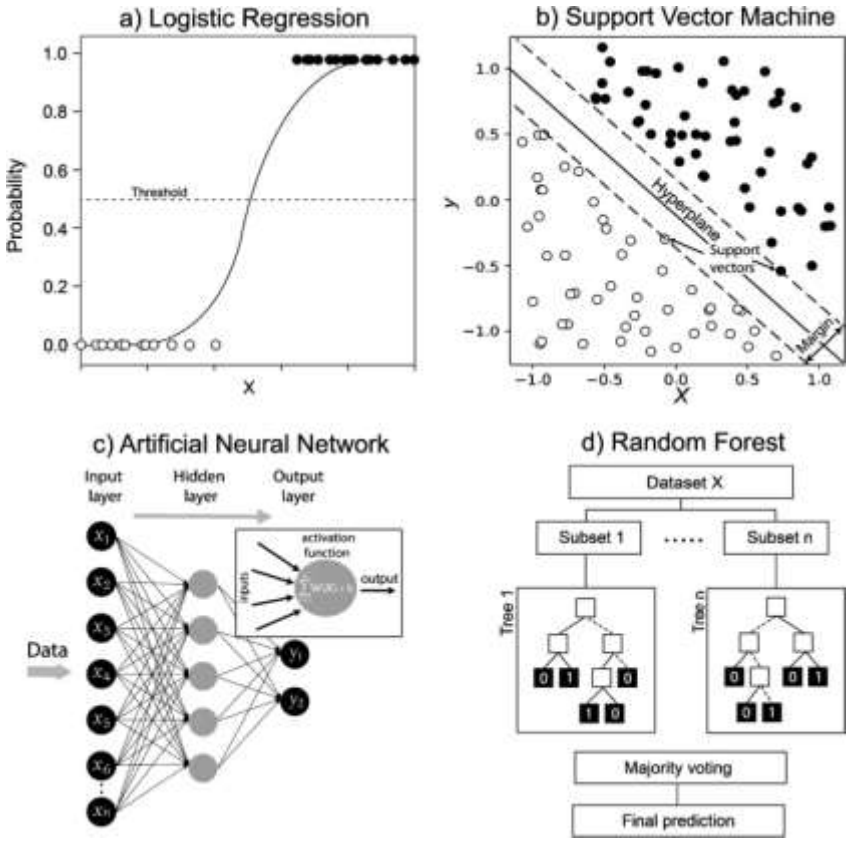

Рис. 2. Схематические иллюстрации четырех контролируемых методов машинного обучения, a. логистическая регрессия, показывающая логистическую функцию, соответствующую двоичным данным на основе переменной x; b. машина опорных векторов (SVM), показывающая линейную гиперплоскость между двоичными классами на двумерном графике, показывающая местоположения опорных векторов и границы; c. искусственная нейронная сеть (ANN), показывающая входной слой, состоящий из n различные функции (т.е. ‘лементы), которые транслируются через ‘скрытый слой’, состоящий из пяти нейронов, производя двоичный вывод y (то есть фертильный или неоплодотворенный). На вставленном рисунке показан отдельный нейрон, который содержит логистическую функцию в качестве функции активации; d. Случайный лес — где набор данных выбирается n раз и для каждого подмножества строится дерево классификации для прогнозирования метки (т. е. фертильности). В качестве окончательного прогноза берется средний результат по всем деревьям

3.8 Машины опорных векторов/

Машины опорных векторов - это контролируемый метод машинного обучения, используемый для классификации многомерных нелинейных данных. Метод преобразует данные в пространство высокой размерности с целью их разделения с использованием поверхности принятия решений высокой размерности, называемой гиперплоскостью. В простом случае два класса двумерных данных (рис. 2b) могут быть разделены линией, причем точки данных, ближайшие к этой линии, являются опорными векторами. Оптимальной позицией классификации линии является та, которая находится дальше всего от опорных векторов и, следовательно, имеет максимальный запас.

Во многих реальных наборах данных невозможно полностью разделить классы данных, что требует формулировки мягких границ. Здесь налагается штраф за данные, которые неправильно классифицированы или находятся в пределах допустимого

предела, при этом неправильно классифицированным точкам данных, расположенным дальше от гиперплоскости, назначается больший штраф. Это важно для шумных наборов данных; однако это также может привести к переобучению, когда модель становится слишком склонной к шуму в обучающих данных. Пользователь может указать важность, придаваемую ошибкам классификации (C), где низкая важность фокусируется на максимизации запаса (т. е. расстояния между опорными векторами), тогда как высокая важность фокусируется на уменьшении ошибочных классификаций — за счет уменьшения запаса.

Во многих задачах классификации может оказаться невозможным использовать линейную или плоскую функцию для классификации данных. В этих ситуациях линейно неразделимые данные сопоставляются с пространством более высокой размерности с помощью функции ядра, с помощью которой они могут быть лучше разделены. Этот процесс преобразования переменных в более высокие измерения для классификации является более дорогостоящим с точки зрения вычислений; следовательно, машины опорных векторов используют трюк ядра, который позволяет алгоритму работать в пространстве высокой размерности без преобразования данных.

Ярким примером геохимического применения машин с опорными векторами является тектоническая дискриминация вулканических пород с использованием геохимии и изотопов цельных пород Петрелли и Перуджини (2016). Авторы показали, что производительность модели улучшается, когда модели обучаются на все большем числе измерений (анализируемых объектов) и при использовании нелинейной функции ядра вместо линейной. Было также показано, что машины с опорными векторами эффективны при составлении карт перспективности полезных ископаемых, литологической классификации и различении измененных фаций.

3.9 Деревья классификации.

Деревья классификации, подмножество деревьев принятия решений, используют наблюдения за элементом для прогнозирования его значения класса с использованием повторяющегося набора двоичных разделов на основе пороговых значений наблюдений. Они представлены в виде "деревьев", в которых классификация набора данных X инициируется в "корневом узле" и передается вниз по дереву, где при каждом разделении или "узле" в X выполняется разделение на два дочерних подмножества на основе разделения по одной переменной (больше или меньше заданного значения концентрации для данного элемента). В конце концов, элемент достигает "терминального узла", где ему присваивается значение класса (например, фертильный или неоплодотворенный). Каждый узел стремится выбрать переменную и

значение, которые наилучшим образом разделяют набор элементов, минимизируя вероятность неправильной классификации (т.е. примеси). Это может быть количественно определено примесью Джини G, которая представляет собой вероятность p неправильной классификации случайно выбранного наблюдения в наборе данных, если оно было случайным образом помечено в соответствии с распределением классов в наборе данных, где C – общее количество классов:

$$G = \sum_{i=1}^{C} p(i) \times (1 - p(i)) \quad (4)$$

Классификационные деревья обладают некоторыми преимуществами по сравнению с другими методами машинного обучения. Первое, что является особым преимуществом для изучения композиционных данных, заключается в том, что деревья классификации являются непараметрическими и, следовательно, не предполагают, что данные соответствуют определенному распределению, в отличие от логистической регрессии, искусственных нейронных сетей и некоторых машин опорных векторов. Второе преимущество заключается в том, что интерпретируемость классификационных деревьев намного выше, поскольку они могут быть визуализированы в двух измерениях. Однако учащиеся, изучающие дерево решений, могут создавать слишком сложные деревья, которые недостаточно обобщают данные (переобучение). Чтобы избежать этой проблемы, необходимы такие механизмы, как обрезка (удаление некритичных или избыточных участков деревьев) и установка максимальной глубины дерева. Продемонстрировали, что классификационные деревья, построенные на основе 51 основных, второстепенных и микроэлементов и изотопных соотношений, могут успешно классифицировать тектоническую близость базальтов с вероятностью 89%, и описали несколько полезных свойств этого подхода по сравнению с типичными двух- или трехвариантными диаграммами тектонической дискриминации.

3.10 Ансамблевые методы и случайный лес.

Случайный лес является примером метода ансамблевой классификации, который использует комбинацию многих предикторов (деревьев классификации) и выбирает прогноз большинства в качестве конечного результата. Случайный лес уменьшает дисперсию усредненного результата по сравнению с одним деревом решений и, следовательно, значительно снижает частоту ошибок. Дисперсия уменьшается двумя способами, направленными на минимизацию корреляции между отдельными деревьями. Во-первых, каждое дерево в Случайном лесу строится на основе случайной выборки наблюдений (начальной загрузки) в данных (рис. 2d), в котором субвыбор

происходит путем замены. Это означает, что наблюдения могут повторяться в многочисленных бутстрапах. Во-вторых, в каждом узле дерева разделение производится с использованием случайного подбора объектов в данных. Наблюдение классифицируется по каждому из этих некоррелированных деревьев в Случайном лесу, и результатом большинства является предсказанный класс. Этот процесс агрегирования нескольких множественных версий предиктора из подвыборок набора данных известен как пакетирование. Еще одним преимуществом процесса пакетирования является то, что наблюдения, которые не включены в подвыборку для каждого дерева, данные "из пакета", могут быть использованы для оценки ошибки классификации при добавлении деревьев в лес, что означает, что набор данных для проверки не требуется.

Случайный лес особенно эффективен для наборов данных со многими характеристиками, даже когда значительная часть неважна или когда наблюдения ограничены или зашумлены. Также считается, что этот подход уменьшает переобучение обучающих данных, поскольку по мере добавления в лес большего количества деревьев функция ошибок сходится к минимуму, что делает их гораздо более надежными, чем одно дерево решений. Например, использовали ряд алгоритмов машинного обучения и химический состав расплава клинопироксена для определения температур и давлений кристаллизации ряда магматических пород, а их основанные на дереве методы ансамбля (включая случайный лес), как правило, давали меньшие ошибки, чем модель с одним деревом решений.

Случайный лес все чаще используется для решения задач разведки полезных ископаемых. Например, использовали его для эффективной классификации типов рудных месторождений, используя 11 различных микроэлементов в пирите. Популярным использованием случайного леса является картирование литологии и перспективности полезных ископаемых с использованием комбинации геохимических и геофизических данных.

3.11 Валидация модели.

Тестирование моделей на невидимых данных имеет решающее значение для оценки способности моделей машинного обучения под наблюдением к обобщению. Модели обычно тестируются путем удержания части набора данных из данных, используемых для обучения модели, и использования удерживаемых данных для тестирования производительности модели. Эффективность разработанных здесь моделей была проверена с использованием десятикратной методики перекрестной проверки (рис. 3). Это включает в себя разделение данных на десятикратное (или

подмножества), причем девятикратное используется для обучения модели, а удержанное - для тестирования модели. Это повторяется 10 раз, пока каждая складка не появится один раз в качестве тестового набора, и для оценки производительности модели берется среднее значение показателей. Преимущество этого подхода, в отличие от одного набора последовательностей / тестов, заключается в том, что он снижает вероятность высокой погрешности, которая может возникнуть из-за одного набора последовательностей / тестов, и помогает обеспечить лучшее обобщение моделей на невидимых данных. Дополнительный ‘тестовый’ набор данных, с которым алгоритм никогда не сталкивается в процессе обучения, сохраняется в стороне для дальнейшего тестирования производительности модели. Важно убедиться, что этапы предварительной обработки, такие как PCA, подходят только для обучающих данных после разделения перекрестной проверки, а не для всего набора данных, а затем применяются как к наборам обучения, так и к наборам проверки. Это связано с тем, что этапы предварительной обработки должны быть изучены только из обучающего набора данных; в противном случае обучающий набор данных будет преобразован на основе информации, хранящейся в наборе данных проверки, что в конечном итоге приведет к искажению процесса перекрестной проверки. Поэтому реализуем этапы предварительной обработки, используя ‘конвейер’, который последовательно применяет преобразования данных и оценщик для каждого этапа перекрестной проверки.

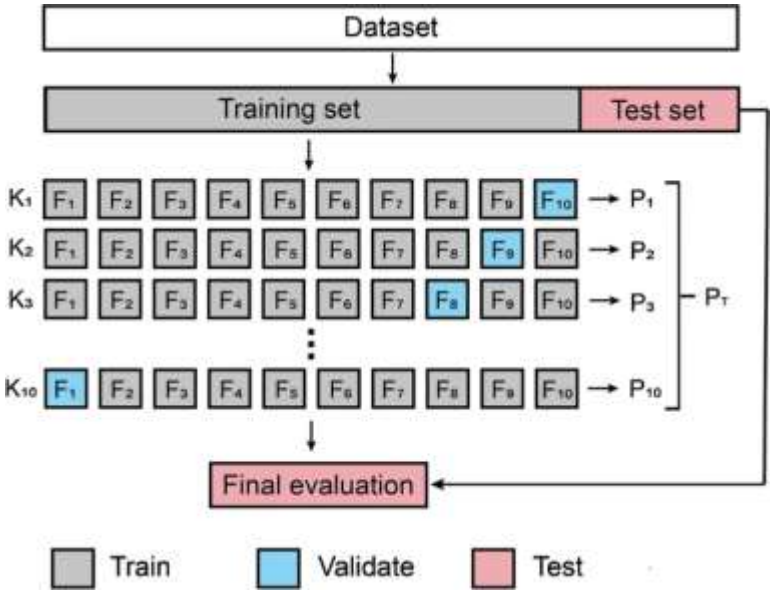

Рис. 3. Схематическая иллюстрация десятикратного рабочего процесса перекрестной проверки. Набор данных разбивается на обучающий набор и тестовый набор. Обучающий набор далее разбивается на десятикратный, после чего процесс перекрестной проверки выполняется 10 раз, причем каждый шаг включает в себя обучение модели с девятикратным и тестирование с исключенным сгибом. Показатели производительности (P) рассчитываются для каждого сгиба и среднего показателя (PT) рассчитывается как общая производительность. Настройка гиперпараметров происходит во время этого процесса перекрестной проверки. Затем выполняется окончательная оценка настроенной модели с использованием тестового набора, где на этом этапе дальнейшая настройка модели не происходит.

Существует несколько показателей для оценки эффективности классификации двоичных классификаторов. Наиболее популярным показателем является точность, которая представляет собой просто долю правильных прогнозов, сделанных классификатором. Однако, поскольку точность менее полезна для наборов данных с несбалансированным классом и задач классификации нескольких классов, часто используются дополнительные показатели, которые обеспечивают лучшее представление о производительности модели. Показатель истинного положительного результата (TPR), также известный как чувствительность или отзыв, является мерой того, сколько положительных наблюдений были правильно классифицированы (истинно положительные, TP) по сравнению с положительными наблюдениями, которые были классифицированы неправильно (ложноотрицательные, FN). Частота ложноположительных результатов (FPR) - это показатель того, сколько отрицательных наблюдений было неправильно классифицировано (ложноположительный результат, FP) по сравнению с количеством истинно отрицательных наблюдений (истинно отрицательный результат, TN). Точность (PPV) - это показатель того, сколько положительных наблюдений было правильно классифицировано. Обычно среднее гармоническое значение отзыва и точности, известное как оценка F1, используется в качестве альтернативного показателя точности. Эти параметры определяются как:

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

$$FPR = \frac{FP}{FP + TN} \quad (6)$$

$$PPV = \frac{TP}{TP + FP} \quad (7)$$

$$F1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} \quad (8)$$

Распространенным методом оценки эффективности двоичной классификации является использование кривой рабочих характеристик приемника (ROC). Бинарные классификаторы обеспечивают вероятность того, что наблюдение принадлежит к классу, и если эта вероятность превышает пороговое значение (обычно 0,5), оно классифицирует наблюдение как принадлежащее к положительному классу. Кривая ROC отображает частоту истинных положительных результатов в сравнении с частотой ложных положительных результатов (FPR) при изменении порога распознавания для классификатора. Классификатор без навыков показал бы одинаковый TPR и FPR при изменении порога, в то время как опытный классификатор

имел бы высокую частоту истинных положительных результатов и низкую частоту ложных положительных результатов. Площадь под кривой ROC для классификатора известна как площадь под кривой (AUC) и обычно используется в качестве показателя производительности классификатора. AUC указывает вероятность того, что классификатор оценит случайно выбранное положительное наблюдение выше, чем случайно выбранное отрицательное наблюдение, где AUC = 1 будет указывать на идеальный классификатор, а AUC = 0,5 будет классификатором без навыков.

3.12 Оптимизация модели.

Параметры, которые не изучаются непосредственно во время машинного обучения, но которые управляют тем, как сам алгоритм строит модель и извлекает уроки из данных, называются гиперпараметрами и задаются ‘заранее’ во время инициализации модели. Примерами гиперпараметров являются количество и максимальная длина деревьев в алгоритме случайного леса, количество и размер слоев в искусственной нейронной сети или коэффициент регуляризации и функция ядра для машины опорных векторов. Чтобы выбрать оптимальные гиперпараметры для каждого алгоритма, используется "поиск по сетке", который подгоняет модель к набору данных, используя различные комбинации заданных гиперпараметров, и возвращает оптимальные гиперпараметры для модели с использованием десятикратной перекрестной проверки. Полный список настроенных гиперпараметров из этого исследования приведен в таблице 2.

4. Результаты и обсуждение.

4.1 Двумерные графики дискриминации.

Понимание петрогенеза дуговых магм, связанных с порфировыми отложениями, обычно опиралось на двумерные графики, в основном включающие соотношения Sr, Y и REE. Двумерные графики могут частично разделить собранный здесь набор данных, демонстрируя различия в химическом составе пород плодородной дуги по сравнению с неплодородными породами дуги (рис. 4 и 5), что согласуется с предыдущими исследованиями (высокий Sr /Y,высокий Al_2O_3/TiO_2, высокий Sr / MnO и высокий La / Yb). Такие участки могут быть полезны при разведке террейнов, предрасположенных к м-ниям порфировой меди. Однако проблема с такими двумерными схемами классификации заключается в том, что существует много неправильных классификаций из-за перекрытия между этими двумя классами. Например, отдельные графики плотности ядра для классов в наборе данных демонстрируют, что двумерные сигналы плодовитости бесполезны для менее развитых композиций (<60 мас.%;

рис. 4). Кроме того, отличительные признаки более выражены для порфировых м-ний Cu, обнаруженных в более толстых дугах и связанных с известково-щелочными магмами (обычно Cu- и Mo-богатые порфиры), чем для многих порфировых объектах, которые обнаружены в более тонких дугах и которые обычно связаны с известково-щелочными и шошонитовыми магмам, как правило, богатыми золотом. Таким образом, ложноотрицательные сигналы плодородия часто наблюдаются в магматических породах, связанных с Cu–Au порфировыми системами. Кроме того, ложноположительные результаты могут быть общими при использовании таких схем классификации отчасти потому, что двумерные сигнатуры, такие как высокий Sr / Y и высокий La / Yb, не являются исключительными для магм, образующих порфировое оруденение Cu, и могут образовываться в результате различных петрологических процессов.

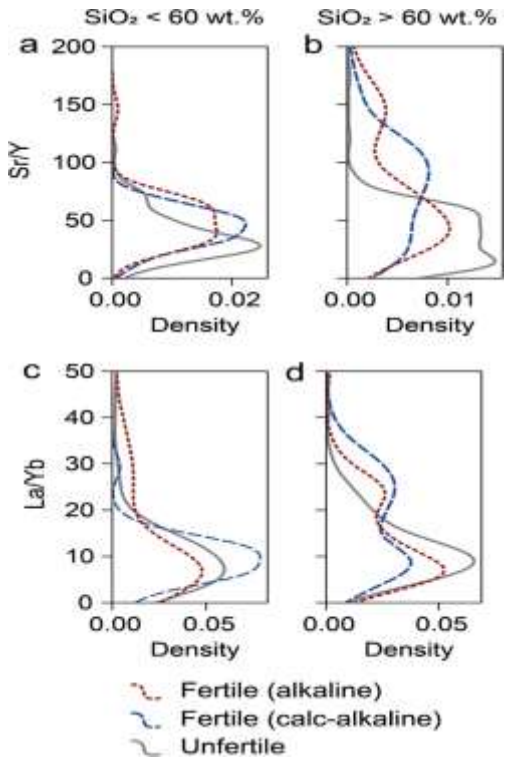

Рис. 4. Графики плотности ядра, показывающие частотное распределение Sr / Y (a и b) и La / Yb (c и d) в обучающем наборе данных. Отдельные графики приведены для менее выделившегося (< 60 мас.% SiO_2) и более развитые (> 60 мас.%) композиции, чтобы проиллюстрировать, что ранее определенные "показатели фертильности" дают больше ложноотрицательных результатов в менее развитых композициях. Отдельные кривые для известково-щелочных (CA)- и щелочно-родственных (K) порфировых систем демонстрируют, что некоторые признаки (например, высокий Sr/Y) менее выражены в щелочно-родственных порфировых системах. Небольшой скачок при ~ 150 Sr / Y для щелочных порфировых систем вызван порфировым м-нием Cu-Mo в Кулунге, которое демонстрирует нетипичный состав с высоким содержанием Mg, ультракалий и высоким содержанием Sr/Y

Протестировали производительность схем Loucks SiO_2 против Sr / Y Sr / Y против Sr / MnO в нашем наборе данных до внедрения контролируемых методов машинного обучения. Loucks предположил, что магматические комплексы, имеющие Sr / Y> 35 при SiO_2> 57 мас.%, следует считать Cu-плодородными (рис. 5). Основываясь на

нашем наборе данных, этот критерий дает оценку точности 69% для порфировых отложений Cu, связанных с известково-щелочными магматическими свитами (TPR = 77%, FPR = 35%), и точность 65% для порфировых отложений Cu, связанных с высококалорийной известково-щелочной и шошонитовой магмой (TPR = 72%, FPR = 34%). Классификация данных, собранных здесь, с использованием диаграммы классификации рождаемости Sr / Y по сравнению с Sr / MnO (рис. 6) возвращает точность классификации 48% (TPR = 68%, FPR = 25%) для порфировых отложений Cu, связанных с известково-щелочными магматическими свитами, и точность 45% (TPR = 56%, FPR = 26%) для порфировых отложений Cu, связанных с высококалорийными известково-щелочными и шошонитовыми магмами (рис. 6).

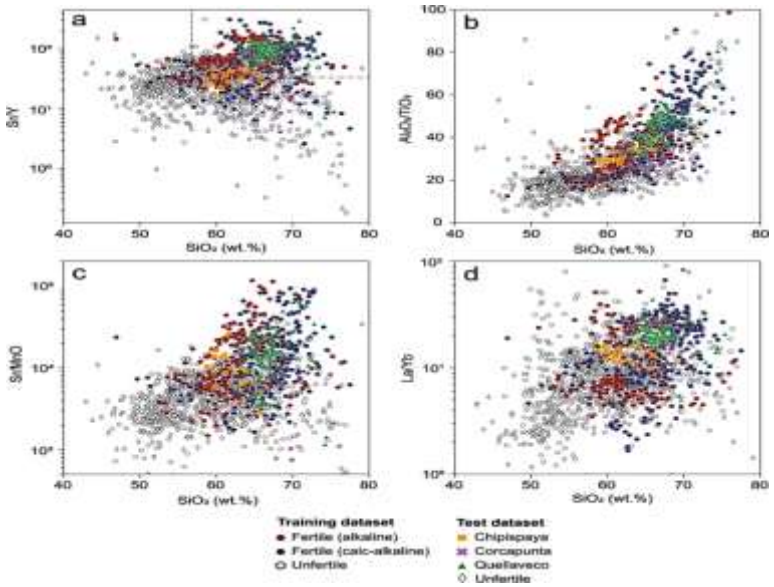

Рис. 5. Диаграммы рассеяния соотношений микроэлементов в зависимости от SiO_2. (a) график Sr / Y с пунктирной линией, указывающей критерий фертильности Меди Sr / Y> 35 при SiO_2> 57 мас.%. (b) Al_2O_3/TiO_2 ; (c) Sr/MnO. (d) La/Yb. Порфировые Cu породы разделяются на основе их сродства к магме, где щелочные относятся к породам, которые классифицируются как высококалорийные известково-щелочные до шошонитового состава.

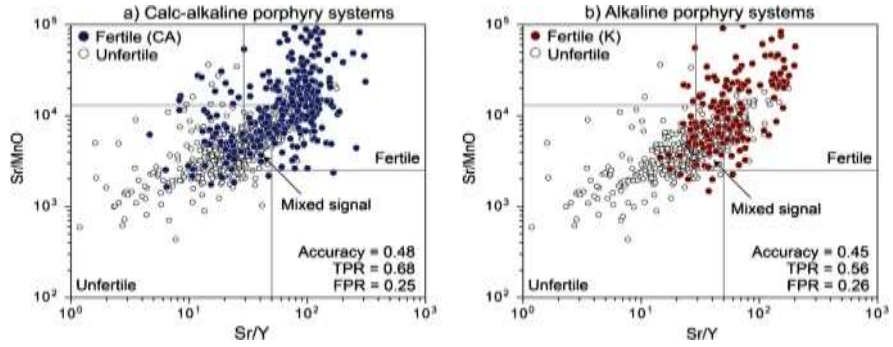

Рис. 6 Обобщение геохимических данных по целым породам, построенных на основе двумерной схемы классификации плодородия для порфировых отложений, связанных с известково-щелочными магматическими свитами (слева) и высококалорийными известково-щелочно-шошонитовыми магматическими свитами (справа).). Точность, частота истинных положительных результатов (TPR) и частота ложных положительных результатов (FPR) рассчитываются на основе этой схемы классификации. Эти показатели рассчитываются с использованием тех же данных, которые использовались для обучения моделей машинного обучения, что позволяет сравнивать двумерные классификаторы и классификаторы машинного обучения. Данные в перекрывающейся ‘смешанной зоне’ присваиваются как неправильные классификации

Эти тесты показывают, что такие схемы полезны, но могут быть ограничены. В обоих случаях для щелочных пород наблюдается более низкая эффективность классификации, и первый тест дает примерно 1 из 3 ложноположительных результатов. Предполагается, что это в основном является следствием небольшого числа элементов, используемых для классификации данных, которые не учитывают полную дисперсию наборов данных и, следовательно, не могут охватить все основные различия между популяциями. Вместо этого эти совокупности данных могут быть лучше разделены в пространстве высокой размерности, где можно смоделировать большую долю дисперсии данных.

Двумерные графики все еще можно использовать для визуализации пространства данных с более высокой размерностью путем построения оценок ПК. Нагрузки ПК, связанные с этими оценками, могут отражать ключевые геологические процессы, которые могут быть важными признаками эволюции магмы и потенциала залежей порфировой меди (рис. 7 и 8). График зависимости PC1 от PC2 для всего набора данных (рис. 7a) позволяет представить 70% дисперсии набора данных, что в данном случае является максимально возможным на двумерном графике. График соответствующих нагрузок на ПК (рис. 7b) показывает, какие элементы осуществляют контроль над каждым ПК, позволяя визуализировать отклонения данных, связанные с геологическими процессами. Здесь дифференциация магмы, по-видимому, является процессом, контролирующим PC1, потому что Mg, Ca, Fe, Mn и Ti имеют большие положительные нагрузки, тогда как K, LREEs и Si имеют отрицательные нагрузки. PC2, по-видимому, преимущественно отражает специфические для минералов процессы фракционирования, типичные для нижней дуговой коры (например гранат, амфибол и отсутствие плагиоклаза), потому что MREEs, HREEs и Y (совместимые с амфиболом и гранатом) имеют высокие положительные нагрузки, но Sr, Al, Na, Si и Eu (совместимые с плагиоклазом) имеют отрицательные нагрузки.

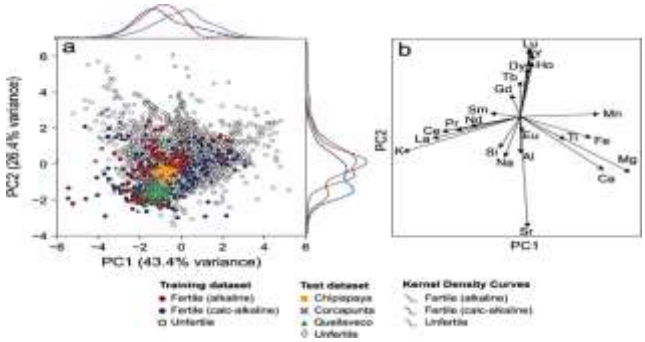

Рис. 7 Графики анализа основных компонентов обучающих и тестовых наборов данных. (a) Оценки PC1 против PC2, которые иллюстрируют максимальную дисперсию, возможно представленную на 2D графике (70% дисперсия). (b) График загрузки ПК для PC1 против PC2, где векторы для каждого преобразованного clr элемента показывают их относительные нагрузки на каждом из двух ПК, показанных в (a)

Из рис. 7 можно сделать несколько ключевых выводов.. Во-первых, плодородные магмы падают дальше вдоль тенденции магматической дифференциации (более низкие показатели PC1), чем неплодородные породы, что согласуется с ассоциацией порфировых отложений Cu с магмами средней кислой дуги. Во-вторых, плодородные магмы подверглись фракционированию глубже в коре (более низкие показатели PC2) по сравнению с неоплодотворенными магмами в соответствии с их высокими соотношениями Sr / Y и La / Yb и общей ассоциацией с дугами, где кора толстая. Обнаружено, что порфировые руды Cu, связанные с известково-щелочными магмами, развивались на более глубоких уровнях земной коры (более низкие показатели PC2), чем отложения, связанные со щелочными магмами. Это соответствует связи между отложениями порфира в более толстой дуговой коре (обычно богатой Cu и бедной Au) с известково-щелочными магмами и между отложениями порфира в более тонкой дуговой коре (обычно богатой Au) с более щелочными магмами.

5. Результаты машинного обучения

Средние показатели производительности для каждого классификатора представлены в результате десятикратного процесса перекрестной проверки (таблица 2), при этом отдельные показатели для каждого сгиба представлены в таблице 3. Все алгоритмы демонстрируют одинаково высокие показатели точности (0,84–0,88), отзыва (0,75–0,78), оценки F1 (0,79–0,80), FPR (0,10-0,14) и точности (0,81–0,83). Эти средние баллы являются значительным улучшением по сравнению с теми, которые были получены в результате тестирования классификаторов Loucks и Ахмед. Для каждого метода контролируемого машинного обучения, изученного здесь, кривая ROC была вычислена для каждого сгиба, выполненного в десятикратном процессе перекрестной проверки (рис. 9). Средняя кривая ROC для каждого классификатора была определена с использованием вертикального усреднения, что позволило сравнить производительность четырех классификаторов (рис. 8). Для всех четырех использованных методов классификации AUC варьировался между 0,87 и 0,89, что указывает на то, что существует 87-89% вероятность того, что классификаторы ранжируют случайно выбранную плодородную породу выше, чем случайно выбранную неплодородную породу.

Таблица 2. **Средние показатели производительности после десятикратной перекрестной проверки обучающего набора данных** *(SVM = машина опорных векторов, ANN = искусственная нейронная сеть, LR = логистическая регрессия, RF = Случайный лес). Значения в круглых скобках представляют собой стандартные отклонения в 1 сигму*

Точность	Оценка F1	Точность	TPR (Отзыв)	ROC-AUC	FPR
SVM 0.81 (0.11)	0.79 (0.14)	0.86 (0.06)	0.75 (0.21)	0.88 (0.11)	0.12 (0.05)
ANN0.83 (0.09)	0.80 (0.12)	0.88 (0.05)	0.76 (0.18)	0.89 (0.10)	0.10 (0.04)
ЛР0.81 (0.09)	0.80 (0.12)	0.84 (0.06)	0.77 (0.16)	0.87 (0.09)	0.14 (0.04)
РФ0.82 (0.10)	0.80 (0.13)	0.87 (0.05)	0.76 (0.18)	0.89 (0.11)	0.11 (0.04)

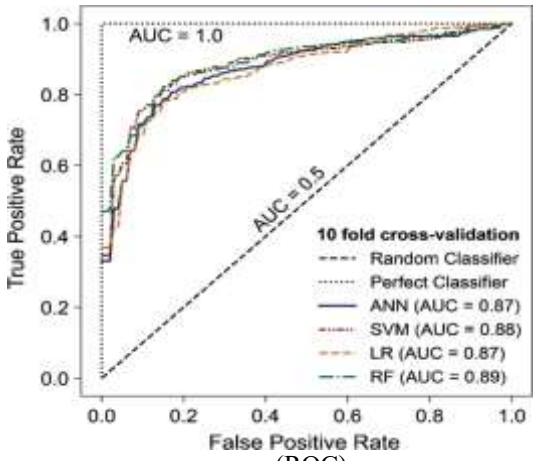

Рис. 8 Кривые рабочих характеристик приемника (ROC) для каждого метода машинного обучения, показывающие истинную положительную частоту и частоту ложных срабатываний, поскольку порог классификации варьируется от 0 до 1. Кривая ROC для каждой модели представляет собой среднее значение из 10 кривых десятикратной перекрестной проверки, определенных по правилу трапеции. Для справки показаны случайный (AUC = 0,5) и идеальный классификатор (AUC = 1,0).

5.1 Независимые тесты производительности модели.

Ограничение использования такой проверки модели заключается в том, что при тестировании методов классификации могут использоваться данные из месторождений или проявлений, на которых были обучены модели. Для дальнейшей проверки прочности методов классификации модели были протестированы на четырех независимых наборах данных, содержащих данные из трех систем porphyry Cu, которые вообще не отображаются в обучающем наборе данных, плюс дополнительные удерживаемые неоплодотворенные данные GEOROC. Три исследованных месторождения порфировой меди имеют разный размер, тип и тектоническую обстановку, все они найдены в Перуанских Андах: (1) Квеллавеко — гигантское месторождение палеоцен-эоценовой порфировой меди и молибдена (3000 тонн с содержанием меди 0,57%); (2) Коркапунта – месторождение миоценовой порфировой меди и молибдена; и (3) Чиписпайя — перспектива миоценового порфира Cu–Au. Данные из Коркапунты и Чиписпайи были получены в ходе тех же аналитических исследований и в них можно найти контроль качества данных. Распределения элементов в обучающих и тестовых наборах данных отслеживались, чтобы убедиться в отсутствии искажений из-за различных аналитических методов (рис. 10).

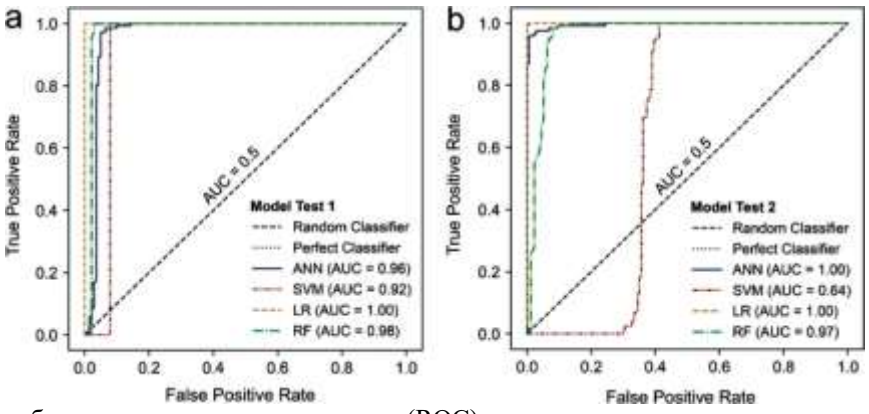

Рис. 9 Кривые рабочих характеристик приемника (ROC) для модельных тестов, которые выполнялись на данных, невидимых в процессе обучения. кривые ROC для каждой модели при выделении данных из месторождений порфировой меди Quellaveco, Chipispaya и Corcapunta из данных GEOROC, которые были случайным образом отобраны и сохранены до обучения модели. b Кривые ROC для каждой модели при выделении данных по месторождениям порфировой меди Квеллавеко, Чиписпайя и Коркапунта из данных прибрежных Кордильер (Чили)

Классификация наборов данных тестов на фертильность по сравнению с удержанными неоплодотворенными данными с использованием четырех обученных моделей классификации показала хорошую эффективность классификации (рис. 9a; таблица 4) с моделями, дающими оценки точности 0,93–0,97 (TPR = 0,95–1,00 и FPR = 0,06–0,13), за исключением опорных векторов, которые дали точность 0,46, поскольку они не смогли правильно классифицировать ни одно из плодотворных наблюдений (TPR = 0). Все модели дали оценки AUC от 0,91 до 1,00.

Ограничение такого модельного тестирования заключается в том, что использование данных GEOROC в качестве отрицательного класса может быть не совсем реалистичным, поскольку эта база данных включает в себя в основном вулканические породы, многие из которых имеют промежуточный состав, обычно не связанный с порфировыми отложениями Cu. В качестве дополнительного теста оценили способность моделей отличать тестовые наборы данных по порфиру Cu от гранитоидов из прибрежных Кордильер, Чили, которые, возможно, более реалистично представляют неплодородные литологии, с которыми можно было бы столкнуться при разведке порфировой меди. Для этого теста эффективность классификации слабее: точность = 0,68–0,71, TPR = 1,00 и FPR = 0,36–0,54. Это указывает на то, что все плодородные породы были правильно классифицированы как плодородные, но присутствует много ложноположительных результатов (или предполагаемых ложноположительных результатов, учитывая, что в наборе данных могут присутствовать необнаруженные плодородные системы). Однако, за исключением метода опорных векторов (SVM), оценки AUC по–прежнему высоки (0,97-1,00), что указывает на то, что, несмотря на ухудшение двоичной классификации, модели по-прежнему способны надежно прогнозировать более высокую вероятность плодородия

для пород, связанных с порфиром Cu, по сравнению с неоплодотворяющими данными. Это подчеркивает, что вероятности, извлеченные из таких моделей, могут быть наиболее полезными, а не только двоичный вывод. В целом, высокая эффективность классификации подтверждает вывод о том, что магматические процессы, связанные с образованием порфировой меди, отличаются от процессов в типичных магматических дугах, иллюстрирующий дополнительные знания, которые могут быть извлечены путем валидации процесса.

Важно отметить, что существует небольшая разница в показателях классификации между различными месторождениями порфировой меди в тестовом наборе данных (рис. 11); следовательно, модели могут использоваться независимо от сродства к магме или размера / типа месторождения. Многие программы разведки направлены на то, чтобы отличать плодородные породы от конкретных литологий, и в этом случае обучающий набор данных может быть уточнен, или в процессе обучения могут быть назначены веса образцов, в результате чего более взвешенные наблюдения более сильно контролируют положение границы принятия решения.

Хотя случайный лес, логистическая регрессия и искусственные нейронные сети демонстрируют сравнительно высокую производительность, предполагается, что случайный лес предоставляет наиболее доступный и эффективный инструмент для геохимических исследований из изученных здесь классификаторов. Непараметрические свойства случайного леса, встроенное предсказание ошибок, оценка важности объектов и простота визуализации - все это желательные свойства. В случайном лесу важность признака может быть оценена с использованием среднего значения среднего уменьшения примеси, т.е. насколько каждая переменная уменьшает неопределенность при классификации данных в дереве (рис. 10a). Компоненты состава, которые появляются в верхней части дерева классификации, имеют более высокое значение, поскольку они приводят к наибольшему уменьшению примесей. Второй метод заключается в том, что объекты в тестовом наборе данных переставляются, а тестовые данные реклассифицируются; любое снижение производительности модели после перестановки объекта указывает на важность объекта (рис.10b). Наконец, случайный лес демонстрирует наименьшую изменчивость точности прогнозирования при настройке гиперпараметров и хорошо работает с использованием гиперпараметров по умолчанию, что делает его эффективным для решения целого ряда проблем.

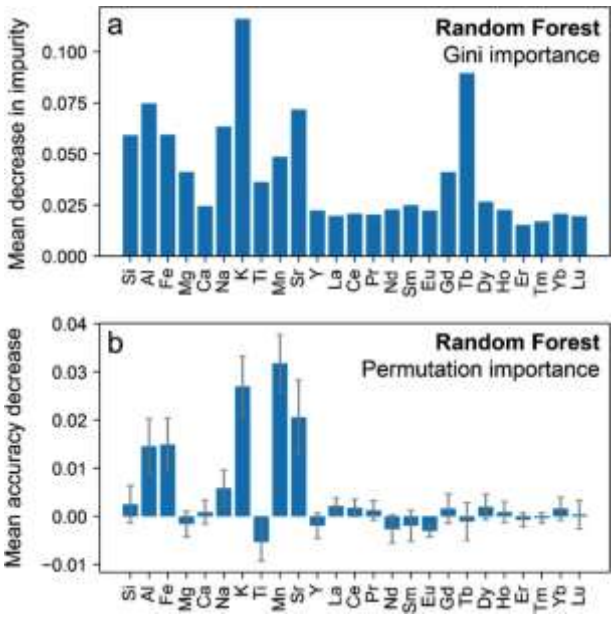

Рис. 10 Нормализованные оценки важности признаков, генерируемые алгоритмом случайного леса при различении плодородных и неплодородных данных. a среднее уменьшение примеси для каждого элемента. b Важность перестановки, когда данные из тестового набора данных были переклассифицированы моделью случайного леса с перестановкой каждого признака для оценки их влияния на результат. Элементы были изменены по 10 раз каждый, и показано среднее снижение точности, где полоса ошибок указывает на стандартное отклонение в 1 сигму. Отрицательная средняя точность указывает на то, что перестановка элемента улучшила производительность модели

5.2 Петрогенетические последствия.

Часто упоминаемым ограничением методов контролируемого машинного обучения является их природа ‘черного ящика’, в котором процедура, используемая для получения результата классификации, является непрозрачной. Это может привести к необъяснимым методологиям, которые не улучшают знания о процессах, способствующих разделению данных. Контролируемые модели машинного обучения, которые являются ‘черным ящиком’, могут затруднить осознанный пересмотр несовершенных классификаций и могут повлиять на более широкое внедрение производных моделей. Одним из способов решения этой проблемы является использование анализа важности объектов, при котором оценки присваиваются объектам на основе их важности для прогнозирования целевой переменной. Было показано, что количественная оценка важности объектов, полученная на основе петрологических данных с использованием машинного обучения, позволяет получить представление о петрогенетических процессах.

Оценки важности объектов для опрошенного здесь набора данных были рассчитаны с использованием sci-kit learn реализации случайного леса на Python. Это показывает, что наиболее важными признаками, используемыми для классификации плодородия магмы в породах дуги с помощью обученной модели, являются K, Tb, Sr и Mn (рис. 10a), и Mn, K, Sr и Fe для важности перестановки (рис. 10b).

Другие классификаторы, использованные в этом исследовании, не содержат встроенных функций для определения важности признаков. К счастью, библиотеки объяснения моделей, такие как SHAP, могут получать оценки важности функций из многих контролируемых методов машинного обучения. Здесь используется SHAP (Shapely Additive exPlanations) для объяснения отдельных прогнозов с использованием библиотеки shap для Python. Значения SHAP вычисляются с использованием теории коалиционных игр, в которой различные коалиции набора признаков (т.е. многочисленные итерации моделей со всеми возможными комбинациями элементов) используются для переоценки прогноза класса, и разница в прогнозе при наблюдении определенного признака по сравнению с исключенным усредняется. Отдельные композиционные параметры вводятся в модель SHAP, а не в баллы PC, чтобы обеспечить лучшее распознавание относительной важности каждого параметра. Ввод элементов как clr-координаты вместо оценок PC приведут к созданию моделей, немного отличающихся от тех, которые были проверены / протестированы; однако предпочтительна дополнительная интерпретируемость в том, что оценки важности признаков присваиваются отдельным химическим компонентам, и не ожидается больших различий между этими моделями. Подчеркнуто, что такие оценки важности признаков отражают важность признака для модели, а не прямую важность этого признака в природе.

Моделирование важности объектов особенно сложно для объектов с высокой мультиколлинеарностью. В качестве примера, поскольку REE являются коллинеарными, модель может преимущественно использовать Tb, а не соседние REE, поскольку они не предоставляют никакой дополнительной информации и, таким образом, делают их избыточными. Это может привести к большим различиям в важности объектов для таких коллинеарных объектов.

Используя библиотеку SHAP, оценки важности признаков были рассчитаны для каждой модели во время классификации тестового набора данных, и приведены их средние значения (рис. 11). Как правило, пять признаков с наивысшим рейтингом схожи между классификаторами: машина опорных векторов (Mn, Sr, Al, Tb и K), искусственная нейронная сеть (Sr, Mn, Al, Ca и Tb), логистическая регрессия (Al, Tb, Mn, Sr, Ce и La). и Случайный лес (Al, K, Mn, Sr и Ti). Ключевым преимуществом SHAP является то, что оценки важности признаков могут быть рассчитаны для отдельных композиций. Анализ этих показателей показывает, что компонентами, которые демонстрируют высокие концентрации в плодородных породах по сравнению

с неплодородными породами, являются Al, Sr, K, LREEs и Ti, тогда как компонентами с низким содержанием в плодородных породах являются Mn, MREE-HREEs и Ca.

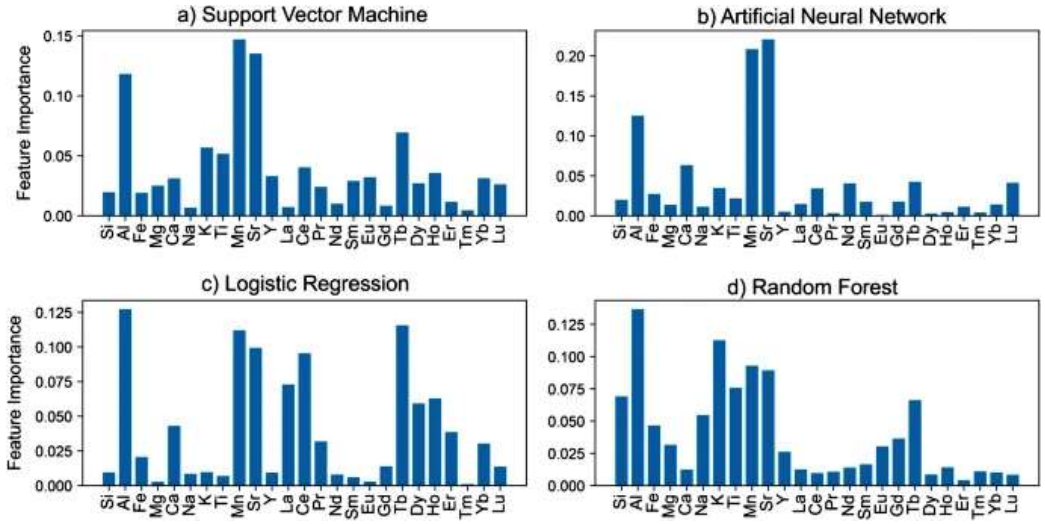

Рис. 11 Нормализованные оценки важности признаков (SHAP) для классификации тестовых наборов данных по каждому классификатору. Более высокие оценки важности признаков указывают на большую важность выделения данных из неплодородных / плодородных пород дуги, как определено теорией коалиционных игр. Помимо случайного леса, они были рассчитаны на основе случайного подмножества из 10 наблюдений из каждого тестового набора данных, поскольку вычисления SHAP являются более дорогостоящими в вычислительном отношении для этих моделей

Полученные важные признаки могут быть использованы для интерпретации петрогенетических процессов, которые являются ключевыми для формирования магм, образующих порфировые отложения меди. Низкое содержание Mn согласуется с предыдущей работой, в которой предполагалось, что раннее фракционирование фаз, таких как амфибол и гранат, в которых Mn совместим во время дифференциации магмы высокого давления в утолщенных дугах (например, D амфибол / расплавMn = 1-28), может привести к образованию расплавов с низким содержанием Mn. Альтернативно, низкий уровень Mn может быть результатом потери гидротермального флюида во время внедрения порфира. Высокое содержание Al и Sr в магматических свитах, связанных с порфиром, также наблюдалось ранее и было связано с подавлением фракционирования плагиоклаза в водных расплавах при высоких давлениях.

Как отмечалось выше, интерпретация оценок важности отдельных признаков для REEs является сложной задачей из-за мультиколлинеарности. Тем не менее, все модели идентифицировали Tb (и умеренную важность для Dy и Ho) как наиболее полезные РЗЭ для классификации, что в целом совпадает с пиковыми коэффициентами разделения амфибола и расплава для РЗЭ. Это подразумевает фракционирование амфибола (± титанита) в нижней коре во время образования магм плодородной дуги, что согласуется с многочисленными исследованиями, которые выявили листрические кривые REE в магматических породах, связанных с порфировыми отложениями Cu.

Наши модели не выявили сильной роли HREEs в классификации плодородных магм, что может свидетельствовать о меньшей роли граната (± циркон) в производстве этих магматических композиций. Гранат стабилен в магмах при более высоких давлениях (> 0,8 ГПа) и / или более высоком содержании воды (> 8 мас.% H_2O), которые часто связаны с развитием порфирово-медно-фертильной магмы в нижней коре. Предполагаемая большая важность фракционирования амфибола в формировании состава плодородной магмы, полученная из наших моделей, может отражать наиболее распространенные внутрикоровые условия (давление и содержание воды), при которых образуются исходные магмы для порфировых отложений Cu. В целом, наиболее важные геохимические особенности согласуются с предыдущими петрологическими исследованиями образования порфировых отложений меди, которые подчеркивают утолщение коры, вызванное сильно сжимающими тектоническими условиями и последующей эволюцией магмы на глубине.

5.3 Предубеждения в машинном обучении и дополнительные влияния.

Разработка, оценка и применение алгоритмов машинного обучения при разведке полезных ископаемых требуют учета предубеждений, которые могут повлиять на принятие решений в пользу конкретной группы. Многие из этих искажений являются искажениями данных, поскольку большие наборы данных часто неоднородны по многим подгруппам. Например, месторождения порфировой меди в изученном здесь наборе данных могут быть нерепрезентативными в глобальном масштабе, поскольку исследования, как правило, сосредоточены на крупнейших месторождениях в зрелых разведанных террейнах, таких как Анды. Кроме того, наши модели предполагают одинаковое значение и важность различных типов отложений, размеров и экономической значимости при различных тектонических режимах (например, смещение накопления), которые неравномерно представлены в обучающем наборе данных (таблица 1). Существуют также ключевые различия между двумя классами данных. Набор данных о плодородии в значительной степени основан на известных залежах полезных ископаемых, где отбор проб в первоначальных исследованиях был сосредоточен на слабо или сильно измененных порфировых интрузиях, в основном расположенных на глубине <6 км, с плотной пространственно-временной выборкой. Напротив, неоплодотворенный набор данных, в основном полученный из GEOROC, содержит спектр в основном неизмененных типов и составов магматических пород с редким географическим распределением, расположенных на разных глубинах или извергнутых на поверхность. Эти отклонения не являются исчерпывающими, частично неизбежными и имеют различную важность в зависимости от приложения, но могут

быть частично смягчены с помощью более тщательной выборки, подвыборки и взвешивания подгрупп в процессе обучения.

Несмотря на эти ограничения, предполагается, что этот подход верен, поскольку признаки плодородия магмы (Sr / Y, La / Yb и Sr / MnO) были обнаружены как в вулканических породах, связанных с порфировыми отложениями, так и в более глубоких исходных гранитоидах порфировых системах. Для контролируемого машинного обучения для геохимических исследований необходим широкий базовый набор данных, позволяющий классификаторам распознавать диапазон типов горных пород, которые могут встретиться во время применения. Повторный запуск моделей, в которых набор данных GEOROC был уточнен, чтобы включать только плутонические породы, привел лишь к небольшому (~ 0,05 снижение ROC-AUC) снижению производительности модели и отсутствию систематических изменений в важности объектов.

Несмотря на методы фильтрации данных и уменьшения размерности, использованные для подготовки набора данных, возникает дополнительная погрешность, поскольку многие породы из порфировых отложений имеют некоторые гидротермальные изменения. Добавление и удаление элементов во время гидротермальных изменений, связанных с минерализацией, может означать, что классификаторы частично различают различия в степени изменения, а не различия в первичной магматической геохимии. Как правило, модель классификации, частично подверженная гидротермальному воздействию, все еще эмпирически полезна, но затуманивает понимание основного процесса и может снизить точность классификации из-за потенциальной значительной пространственной изменчивости эффектов изменения. Кроме того, эти эффекты могут частично сильно скрывать определенные магматические сигнатуры; например, подвижность Sr может препятствовать его использованию для определения плодородия магмы. Эти гидротермальные эффекты более вероятны для наиболее подвижных компонентов жидкости, таких как K, Sr, Mn, Ca и Fe. Несмотря на это, обнаруживается, что большая часть различий, проявляемых этими элементами, объясняется магматическими процессами, о чем свидетельствует PCA (рис. 7 и 8).

Выводы.

Продеманстрировано, что четыре контролируемых метода машинного обучения (логистическая регрессия, искусственные нейронные сети, машины опорных векторов и случайный лес) могут быть обучены отличать магматические породы,

пространственно и временно связанные с медно-порфировыми месторождениями, от тех, которые не связаны с минерализованными системами, с высокой производительностью, независимо от типа или размера магматического месторождения. Эта методология превосходит более традиционные двумерные схемы классификации. Этот обобщенный подход может быть адаптирован для более индивидуальных приложений разведки, таких как определение более конкретного типа или размера месторождения, использование более точного набора обучающих данных или применение взвешиваний в процессе обучения. Многие из этих методов требуют критических этапов предварительной обработки для учета многих свойств геохимических наборов данных, таких как эффекты состава данных, разреженность, высокая мультиколлинеарность и дисбаланс классов. Случайный лес потенциально обеспечивает наиболее прозрачную и простую модель, требующую небольшой предварительной обработки данных.

Оценки важности признаков, полученные на основе классификаторов, обеспечивают определенную степень интерпретируемости, которая может помочь стимулировать внедрение модели. Эти оценки также дают полезную информацию о петрогенетических процессах, связанных с образованием магм, образующих порфировые отложения меди. Наиболее важными компонентами, которые помогают отличить плодородные породы дуги от неплодородных пород, являются образцы с высоким содержанием Al, низким содержанием Mn, высоким содержанием Sr, высоким содержанием K и listric REE. Некоторые из них (Al, Mn, MREEs и Sr) согласуются с геохимическими признаками, полученными во время водной магматической эволюции под высоким давлением в нижней коре, где фракционирование плагиоклаза подавлено, а амфибол (± гранат) является распространенной фракционирующей фазой.

Таким образом, полученные важные признаки дополнительно подтверждают предыдущие петрологические исследования магм, связанных с порфировыми отложениями меди.

В целом, этот подход демонстрирует мощь использования пространства геохимических данных высокой размерности для составления обоснованных классификаций для эффективной разведки полезных ископаемых.

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ СТРУКТУРНЫХ ПОСТРОЕНИЙ
(Структурный контроль минерализации меди, Восточный Китай) *[2]*

1. Введение

В глобальном масштабе гидротермальные системы обычно формируются в определенных тектонических условиях, например, порфировые системы в основном встречаются в условиях магматических дуг. В региональном масштабе гидротермальные месторождения демонстрируют близость к региональной системе разломов или зонам сдвига, которые служат путями для транспортировки рудообразующих флюидов из глубинных источников в места рудоотложения. В масштабе месторождения брекчии и жилы, которые связаны с оперяющими зонами разломов региональных структур, служат благоприятными местами для сосредоточения и отложения рудоносных флюидов и интерпретируются как ответственные за локализацию рудных тел. Однако такие элементы контроля оруденения могут быть неоднозначными в различных источниках геологических данных, потому что:

(1) структуры, особенно крупномасштабные, могут иметь переменные выражения от глубины к поверхности (например, зона милонита на глубине и зона разлома вблизи поверхности);

(2) пространственные ассоциации между обобщенными на карте структурами и оруденением, спроецированными на поверхность на 2D-картах могут привести к неточному представлению или даже неправильному пониманию в отношении контроля минерализации; и

(3) структурные особенности вместе со структурно контролируемой минерализацией могут формироваться в результате последовательной деформации и многофазной тектоники.

Таким образом, выявление структурных особенностей, связанных с рудой, и выяснение структурных элементов управления, а также измерение их вклада в формирование месторождений полезных ископаемых является сложной задачей.

С помощью количественных методов и простого в использовании программного обеспечения ГИС определение пространственных закономерностей залежей полезных ископаемых и их связей с геологическими особенностями (например, структурными) может, в дополнение к полевым наблюдениям, геохимическим и др. методам, могут дать представление о механизмах контроля, действующих в различных масштабах. Кроме того, распознавание структурных особенностей, контролирующих

минерализацию, имеет решающее значение для определения прогнозно-поисковых комплексов.

Поскольку проявления полезных ископаемых упрощены для представления в виде точек на различных картах, методы пространственного анализа точечных моделей чаще всего используются при изучении пространственного распределения и геологического контроля месторождений полезных ископаемых, в основном с использованием фрактальной геометрии и анализа Фрай. С помощью статистических расчетов фрактальный анализ и анализ Фрай способны выявить структуру распределения месторождений полезных ископаемых, которую может быть трудно распознать, полагаясь исключительно на логическую интерпретацию. Более того, анализ распределения расстояний и анализ весов доказательств (WofE) могут дополнительно количественно оценить силу пространственной связи между залежами полезных ископаемых и геологическими особенностями, которые, как считается, благоприятны для прогнозирования местоположения минерализации. Совместное применение этих методов необходимо, поскольку отдельные методы характеризуют только конкретный аспект, такой как неслучайная кластеризация месторождений или преимущественное направление распределения месторождений, сложных пространственных особенностей систем минерализации.

Рудный район Тунлин (TOD) является одним из наиболее важных производителей меди в Китае с общими оценочными запасами более 5 млн тонн меди. Крупные пластовые месторождения меди составляют большую часть запасов меди в этой области, например, месторождение Дунгушань с 1 млн куб.м при 1,01% и месторождение Синьцяо с 0,5 млн куб. м при 0,71%, которые привлекли множество исследований, посвященных их происхождению.

Хотя генезис оруденения все еще остается спорным, большинство исследователей склонны соглашаться с эпигенетическим происхождением или, по крайней мере, доминирующим вкладом Яншаньской магма-гидротермальной деятельности в наложенные процессы рудообразования. В генетической модели, связанной с магматизмом, стратиформные рудные тела образовались в результате прогрессирующего взаимодействия флюида с породой вдоль структурно контролируемых каналов, параллельных слоям, и были неотъемлемой, но удаленной частью крупной гидротермальной системы, которая породила проксимальные скарновые рудные тела в зонах контакта и порфировые рудные тела в Яншаньских интрузиях. Такая гидротермальная система и стратифицированные отложения аналогичны своим аналогам в других местах, среди которых двумя репрезентативными

примерами являются порфирово-скарновые полиметаллические месторождения в районе Эртсберг в Индонезии и месторождения меди типа манто в Чили. В районе Эртсберг восточно-скарновое рудное тело Эртсберг, одно из крупнейших рудных тел, расположено в зоне, ограниченной параллельным разломом, между известняком формации Фаумай и доломитовым карбонатом формации Варипи. В районе Пунта-дель-Кобре в Чили стратифицированные табличные рудные тела залегают в горизонтах андезитовой брекчии между нижележащим массивным андезитом и вышележащим сланцем, в то время как участки экономической концентрации меди, по-видимому, контролируются разломами. Поскольку структура является важным контролирующим фактором этих слоистых отложений, в TOD были проведены соответствующие исследования, включая пространственные модели деформаций и формирования рудовмещающих структур. Однако эти исследования в основном были сосредоточены на теоретической дедукции и качественном анализе. В данном исследовании предпринята попытка разграничить структурный контроль как качественными, так и количественными аналитическими методами, сосредоточив внимание на механизмах структурного контроля, действующих в разных масштабах, что может облегчить понимание формирования месторождений меди и обеспечить критерии для будущей разведки в TOD.

2. Материалы и методы

2.1. Область исследования

TOD расположен в центральной части Средне-Нижнего Cu-Au-Fe металлогенического пояса Янцзы (MLYMB) вдоль северной окраины кратона Янцзы, граничащего с орогенным поясом Цинлин-Дабишань и Северо-Китайским кратоном на севере (рис. 1a).

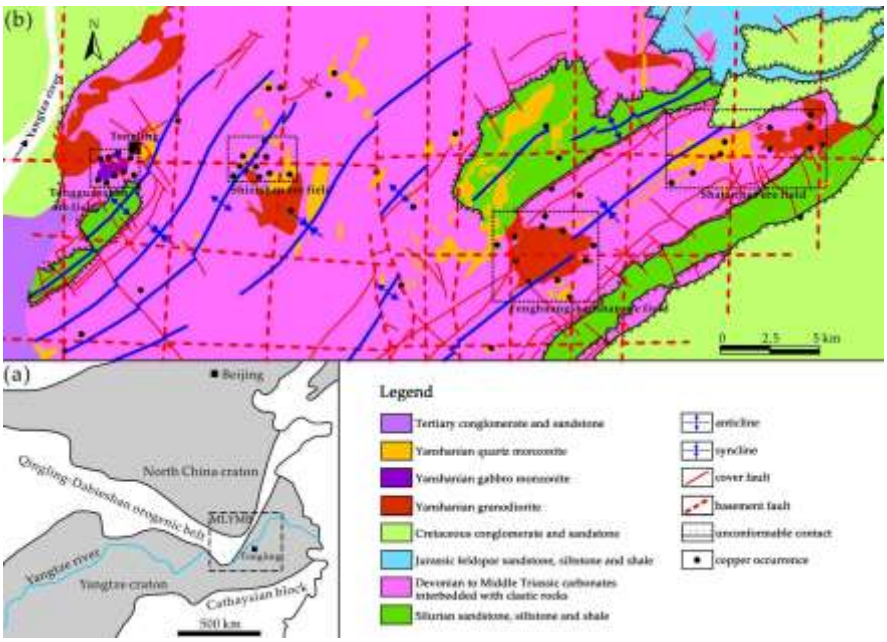

Рис. 1. Карта района исследований: (a) - упрощенная тектоническая карта, показывающая расположение ТОД; и (b) геологическая карта TOD, показывающая местонахождение месторождений меди,

Северный кратон Янцзы подстилается тоналитово-тронджемитово-гранитными (TTG) гнейсами возрастом от 3,45 до 2,87 млрд лет. Фундамент ТОД составляют гнейсы и несогласно перекрывающие их метаморфические породы от архея до палеопротерозоя. От кембрия до среднего триаса ТОД представлял собой стабильный желоб, заполненный карбонатными и обломочными породами мелководной морской фации. В этот период развились две осадочные толщи, включая нижнесилурийско-верхнедевонскую регрессивную батиальную с прибрежными обломочными породами и вышележащие карбонаты от верхнего карбона до среднего триаса от прибрежных до неритических карбонатов, переслаивающиеся с батиальными и альтернативными морскими и континентальными обломочными породами. От юрского до мелового периода этот регион пережил событие, которое интерпретировалось как стадия внутриконтинентальной активизации с обильным магматизмом, Были развиты мощные наземные вулканогенно-осадочные толщи, которые перекрывают породы от силура до триаса (таблица 1).

Таблица 1. **Стратиграфия и тектонические события в рудном районе Тунлин**

Epoch	Lithostratigraphic Unit	Code	Lithological Description	Tectonic Activity
Upper Cretaceous	Xuannan Formation	K_2x	Conglomerate and sandstone	Yanshanian movement (ca. 135 Ma)
Middle Jurassic	Luoling Formation	J_2l	Feldspar sandstone, siltstone and shale	
Lower Jurassic	Moshan Formation	J_1m	Feldspar sandstone with interlays of silty shale and coal, conglomerate at bottom	Indosinian movement (ca. 195 Ma)
Middle Triassic	Tongtoujian Formation	T_2t	Siltstone with interlays of sandy shale	
	Yueshan Formation	T_2y	Limestone, dolomite in upper and siltstone in lower	
Lower Triassic	Nanlinghu Formation	T_1n	Limestone	
	Helongshan Formation	T_1h	Limestone	
	Yingkeng Formation	T_1y	Limestone with interlays of silt shale	
Upper Permian	Dalong Formation	P_2d	Siliceous shale with interlays of limestone	
	Longtan Formation	P_2l	Fine sandstone and silt shale with interlays of coal	
Lower Permian	Gufeng Formation	P_1g	Siliceous slate and siliceous shale	
	Qixia Formation	P_1q	Bioclastic limestone in upper and carbonaceous shale in lower	
Upper Carboniferous	Chuanshan Formation	C_2c	Orbicular limestone and bioclastic limestone	
	Huanglong Formation	C_2h	Bioclastic limestone and dolomite	
Upper Devonian	Wutong Formation	D_3w	Quartz sandstone and silty shale	
Middle Silurian	Fentou Formation	S_2f	Sandstone, siltstone and sandy shale	
Lower Silurian	Gaojiabian Formation	S_1g	Black shale	
Upper Ordovician	Wufeng Formation	O_3w	Black siliceous shale	
	Tangtou Formation	O_3t	Calcareous shale with interlayers of limestone	
Middle Ordovician	Tangshan Formation	O_2t	Limestone with interlayers of thin slate	
Lower Ordovician	Lunshan Formation	O_1l	Limestone in upper and dolomite in lower	
Cambrian	Huangjiabang Formation	ϵ	Limestone	
Precambrian	Dongling Group	Pt_3d	Biotite quartz schist and gneiss	Jinning movement (ca. 850–800 Ma)

В региональных структуре TOD преобладают складки с мелкими осевыми поверхностями сигмоидальной формы (рис. 1b). Вторичные структуры включают надвиговые разломы северо-западного направления. Региональные гравитационные аномалии и профили глубокого сейсмического отражения указывают на наличие разломов в фундаменте EW- и NS-трендов. Яншаньский магматизм привел к образованию более 70 интрузий, которые в основном состоят из гранодиорита, кварцевого монцонита, габбромонцонита и их гипабиссальных эквивалентов. Результаты U-Pb датирования циркона показали, что интрузии были сформированы в раннем мезозое (в основном 145-129 млн лет). В медно-полиметаллических месторождениях преобладают месторождения скарнового типа, при этом в более глубоких частях некоторых скарновых месторождений встречаются незначительные залежи меди порфирового типа. В ТОД было обнаружено более 60 месторождений медно-полиметаллических скарнов, в основном сосредоточенных в четырех рудных полях с запада на восток: Тунгуаньшань, Шизишань, Фэнхуаншань и рудное поле Шатаньцзяо (рисунок1b).

Исходные данные были изучены перед вводом в пространственную базу данных. В анализ были включены только медные и полиметаллические месторождения с преобладанием меди. Структурные особенности были классифицированы по трем

категориям: разломы фундамента, разломы чехла и складки. Все изученные данные были скомпилированы в векторные форматы и импортированы в платформу ArcGIS 10 для последующего пространственного анализа.

2.2. Фрактальный анализ

Фракталы - это объекты, которые имеют сходные геометрические узоры при наблюдении в различных масштабах. Эта масштабная инвариантность может быть описана пропорциональной зависимостью между измерением целевого шаблона и его масштабом. Были предложены различные методы для оценки фрактальной размерности данного паттерна, каждый из которых раскрывает аспект геометрической сложности целевого паттерна. Методы подсчета квадратов и радиальной плотности, которые являются наиболее часто используемыми методами при анализе геологических точечных моделей, были использованы для оценки соответствующих фрактальных измерений. Фрактальный анализ может выявлять статистические законы распределения медных залежей, связанных с масштабом и типом структурно контролируемых процессов.

В методе подсчета ячеек область исследования, включающая представляющие интерес геологические объекты (например, месторождения), перекрывается сеткой, которая содержит квадратные ячейки или ячейки с длиной стороны δ, а затем подсчитывается количество N(δ) этих ячеек, содержащих части целевых объектов (рис. 2a).. Описанный выше процесс повторяется с использованием другого размера ячейки δ для получения соответствующего номера ячейки N(δ) (рис. 2b, c). Если анализируемый паттерн относится к фрактальному паттерну, то соотношение между N(δ) и δ должно соответствовать степенной функции, как показано ниже:

$$N(\delta) \propto A\delta^{-D_B} \quad (1)$$

где Db - фрактальная размерность, учитывающая количество ячеек, а A - константа. Практически, строится график зависимости логарифма (N(δ)) от логарифма (δ), а затем методом наименьших квадратов проводится линия регрессии, наиболее подходящая, в то время как наклон линии регрессии представляет фрактальную размерность с подсчетом квадратов (рис. 2d).

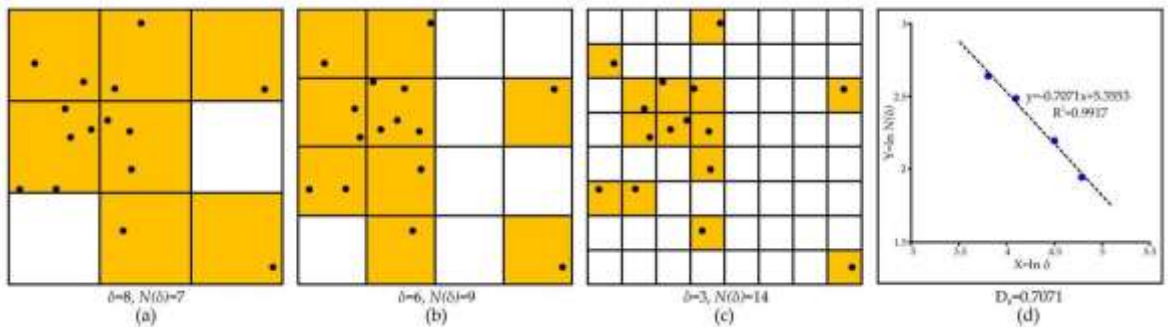

Рисунок 2. Схематическая диаграмма анализа подсчета ячеек: (a) 7 ячеек, содержащих целевые точки с размером ячейки. δ = 8; (b) 9 ячеек, рассчитанных с размером δ = 6; (c) 14 ячеек, подсчитанных с размером δ = 3; и (d) логарифм график, показывающий степенную зависимость количества подсчитанных ячеек N (δ) и размера δ, получая фрактальная размерность подсчета ячеек DB = 0,7071.

В методе радиальной плотности было продемонстрировано, что фрактальные точки, также называемые фрактальной пылью, удовлетворяют соотношению радиальной плотности, которое можно описать как:

$$d \propto Br^{D_R-2} \quad (2)$$

где d - средняя плотность точек окружностей с радиусом r, центр которых находится в каждой точке, а B - константа, в то время как Dr - фрактальная размерность радиальной плотности. Аналогично, Dr обычно получается путем вычисления наклона линии регрессии, которая представляет линейную зависимость d и r на логарифмическом графике.

2.3. Анализ Фрай.

Анализ Фрай - это геометрический метод пространственной автокорреляции для анализа точечных паттернов, который реализуется путем построения диаграммы автокорреляции, называемой графиком Фрая. На рисунке 3 показана базовая процедура создания графика Фрай:

(1) подготавливаются два аналоговых листа, включая оригинальный лист с записью необработанных точек (рис. 3a) и чистый лист трассировки;

(2) начало исходного листа O помещается на одну из необработанных точек, таким образом сохранение ориентации и расстояний всех остальных точек (рис. 3b);

(3) точки на исходном листе затем переносятся на лист трассировки, при этом O совпадает с началом листа трассировки O’ (рис. 3c);

(4) начало O перемещается в другую исходную точку (рис. 3d), и новая схема распределения - необработанные точки копируются на лист трассировки после шага (3) (рис. 3f).

Эта процедура повторяется до тех пор, пока каждая точка на исходном листе не будет использована в качестве начала координат O (рис. 3e,f), в результате чего на

листе трассировки будет (n2-n) точек для n необработанных точек (рис. 3г). Окончательный лист трассировки называется графиком Фрай, а точки на этом листе называются точками Фрай.

Fry plot записывает расстояния и ориентации каждой исходной точки относительно любой другой точки, тем самым улучшая тонкие шаблоны объектов целевой точки, на основе которых обычно строятся диаграммы розы для анализа предпочтительных ориентаций пар точек на определенных расстояниях, которые показывают направленный контроль минерализации в разных масштабах карты.

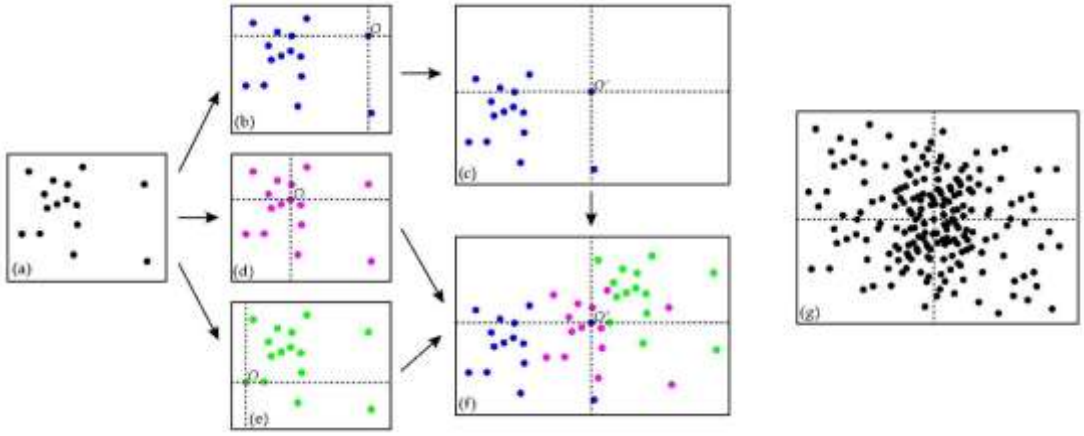

Рисунок 3. Схематическая диаграмма для построения графика Фрая: (a) исходный лист записывает необработанные точки; (b) начало координат O расположено в одной из исходных точек; (c) картина распределения необработанных точек в соответствии с происхождение O переносится на кальку; (d,e) начало координат O переставляется в каждой исходной точке; (e) в трассировочном листе фиксируются все закономерности распределения исходных точек по разным источникам; и (g) Фрай участок построен.

2.4. Анализ распределения расстояний.

Анализ распределения расстояний - это основанный на пространственном буфере метод количественной оценки пространственной связи между набором точек (например, рудных залежей) и другим набором пространственных объектов. Этот анализ включает в себя вычисление и сравнение совокупной относительной частоты в соответствии с заданными расстояниями от определенного набора геологических объектов до:

(1) мест залегания полезных ископаемых (обозначаемых как Dm) и

(2) мест залегания (обозначаемых как Dn).

Dn указывает естественно случайное распределение плотности вероятности регулярных шаблонов в пределах заданного буферного расстояния, в то время как Dm представляет неслучайное распределение плотности вероятности минерализованных структур, которое характеризуется неравномерной кластеризацией залежей полезных ископаемых в соответствующем буфере. Разница D, которая вычисляется по формуле (Dm − Dn), представляет, насколько совокупная частота

встречаемости полезных ископаемых (т.е. Dm) выше или ниже, чем ожидаемая из-за случайности (т.е. Dn), измеряя интенсивность пространственной связи между анализируемым геологическим объектом и минерализацией.

Чтобы показать статистически, если Dm значительно больше, чем Dn, верхняя доверительная полоса для кривой Dn (обозначаемая как uc) может быть задана:

$$uc = D_N + \sqrt{9.21(M+N)/4MN} \quad (3)$$

где M - количество залеганий полезных ископаемых, которые использовались для оценки Dm, в то время как N - количество мест, не залегающих, используемых для вычисления Dn, а 9.21 - константа для уровня значимости α = 0,01.

2.5. Анализ весомости доказательств (WofE).

Анализ Wo***f***E - это байесовский статистический метод, основанный на данных, который предлагает количественное измерение пространственной связи между набором заданных геологических особенностей и целевыми проявлениями (например, рудными залежами, перспективами или геологическими аномалиями).

В ГИС-приложении для анализа, связанного с залеганием полезных ископаемых, анализ Wo***f***E реализуется на основе нескольких бинарных прогнозирующих карт геологических объектов. Во-первых, исследуемая область подразделяется на T квадратных ячеек одинакового размера, среди которых D ячеек занимают залежи минералов. Априорная вероятность может быть определена как:

$$P_{prior} = P(D) = \frac{D}{T} \quad (4)$$

и относительная важность пространственной связи между геологическим объектом Bi и минерализацией оценивается парой весов, а именно положительным весом W+ и отрицательным весом W−, которые могут быть заданы:

$$W^+ = \ln\left\{\frac{P(B|D)}{P(B|\overline{D})}\right\},\ W^- = \ln\left\{\frac{P(\overline{B}|D)}{P(\overline{B}|\overline{D})}\right\} \quad (5)$$

где P обозначает соответствующую вероятность; B и $\overline{B}$ являются наличие и отсутствие геологических особенностей; D и $\overline{D}$ являются наличие и отсутствие залежей полезных ископаемых. Например, P(B|D) представляет вероятность появления B при наличии D. Контраст C определяется как общее измерение пространственной корреляции, которое задается:

$$C = W^+ - W^- \quad (6)$$

Чтобы оценить значимость контраста C, здесь используется достоверность контраста (обозначаемая как CS), полученная с помощью t-критерия Стьюдента, и определяется как:

$$C_S = \frac{C}{S(C)} = \frac{C}{\sqrt{S^2(W^+) + S^2(W^-)}} \quad (7)$$

где S обозначает стандартное отклонение соответствующего параметра.

3. Результаты и обсуждение

3.1. Пространственные закономерности залегания меди

Логарифмический график подсчета ячеек показывает, что распределение залежей меди в исследуемой области имеет бифрактальный характер, т.е. логарифмический график зависимости номера ячейки N(δ) от размера δ может быть снабжен двумя линиями регрессии (рис. 4a), что приводит к двум фрактальным измерениям 0,2468 (δ ≤ 1,6 км) и 0,75 (δ> 1,6 км). Напротив, анализ радиальной плотности дает трифрактальную картину, на что указывают три линии регрессии, которые представляют три фрактальных измерения, варьирующихся от 0,796 (r ≤ 1,4 км), 1,1722 (между 1,5 и 4,5 км) до 0,8092 (r> 4,5 км) (рисунок 4б). Одна линия регрессии представляет степенную (фрактальную) зависимость между измерениями и их масштабами, подразумевая модель масштабной инвариантности, возникающую в результате нелинейного процесса. В этом исследовании различные фрактальные закономерности залегания меди, указанные многострочной фрактальной моделью, могут быть приписаны различным процессам контроля руды, работающим в разных масштабах. Примечательно, что логарифмический график, особенно для анализа радиальной плотности, по-видимому, альтернативно снабжен одной единственной линией регрессии. Однако двухлинейные и трехлинейные фрактальные модели, показанные на рисунке 4 являются оптимальными, поскольку они достигают максимальных коэффициентов регрессии (R2) для подогнанных линий, что означает, что уменьшение любой линии регрессии приведет к снижению коэффициентов регрессии.

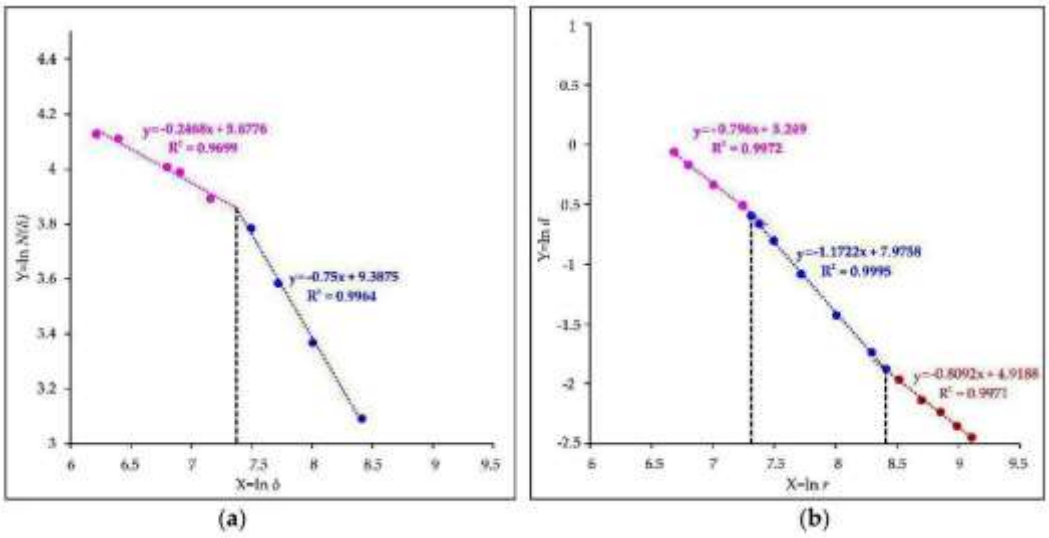

Рис. 4. Логарифмический график, определяющий фрактальные размерности пространственной структуры месторождений меди в TOD: (a) линейная зависимость подсчета ячеек; и (b) линейная зависимость радиальной плотности.

Несмотря на очевидные различия во фрактальных измерениях, отмеченные в приведенных выше анализах, существует общее согласие между результатами метода подсчета квадратов и анализа радиальной плотности, так что изменения фрактальных измерений, на которые указывает пересечение соседних регрессирующих линий, происходят на расстоянии около 1,5 км, что позволяет предположить, что различные фрактальные структуры существуют в пределах идентичных диапазонов (в пределах 1,5 км и за пределами 1,5 км) как для подсчета боксов, так и для фрактальной зависимости радиальной плотности. Также отмечается, что на фрактальном графике радиальной плотности есть пересечение в 4,5 км; однако неясно, существует ли еще одно фрактальное измерение в анализе подсчета ячеек, когда размер ячейки превышает 4,5 км, поскольку количество ячеек в такой ситуации недостаточно велико для статистического подсчета.

Результаты фрактального анализа в этом исследовании, включая модель мультифрактальной размерности и фрактальные структуры, встречающиеся в идентичных диапазонах, согласуются с результатами некоторых предыдущих исследований. Считается, что расхождения во фрактальных измерениях правдоподобно связаны с различными геологическими контролями, действующими в различных масштабах, например, в региональном, локальном и перспективном масштабе. Тем не менее, такой изменяющийся в масштабе геологический контроль все еще остается загадочным и нуждается в дальнейшем анализе.

Анализ Fry был выполнен для изучения ориентации вероятных средств контроля минерализации меди. 3906 точек были получены из 63 залежей меди в ТОД (рис. 5a),

на основе которых были построены диаграммы розы. Диаграмма розы для всех точек иллюстрирует просто доминирующую тенденцию РЭБ (рис . 5b), что предполагает фундаментальный контроль за минерализацией меди в региональном масштабе. Поскольку фрактальный анализ показывает различия во фрактальных измерениях в пределах 1,5 и 4,5 км, мы также проанализировали характеристики точек в пределах этих диапазонов. Диаграмма розы для точек Срю в пределах 4,5 км друг от друга указывает на преимущественный тренд NNE с подчиненными трендами NE и EW (рис. 5c). Диаграмма розы для точек Срю, расположенных на расстоянии 1,5 км друг от друга, показывает основную тенденцию NE-NE, с дополнительными тенденциями в направлениях EW и NS (рис. 5d).

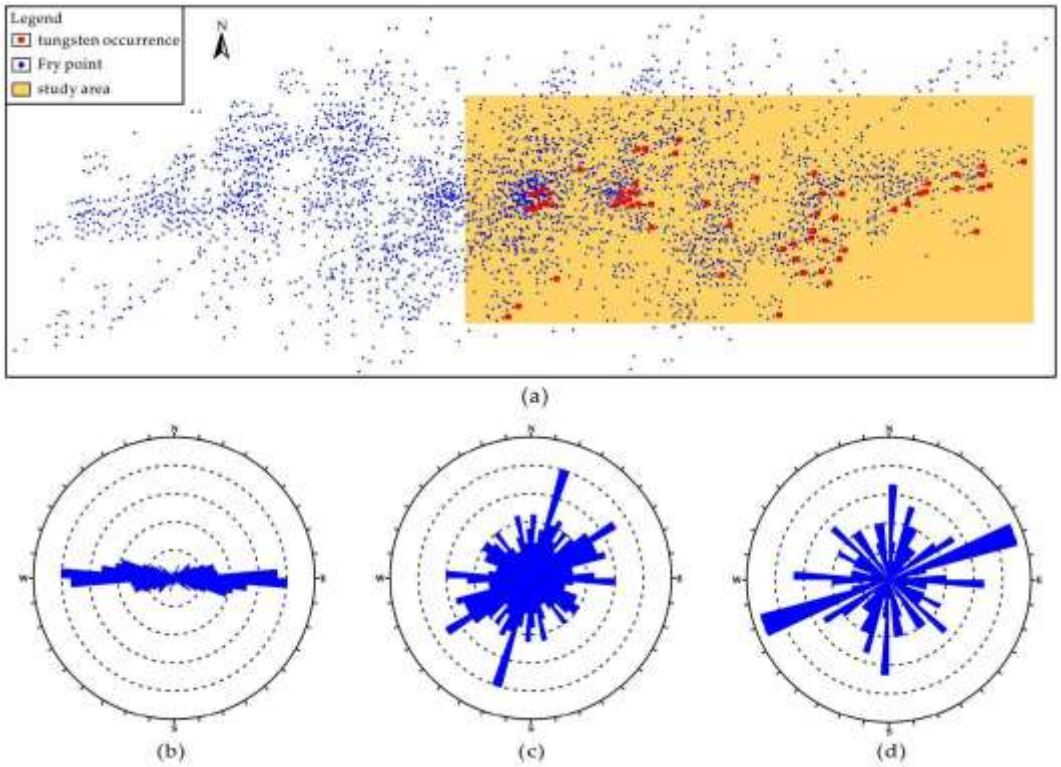

Рис. 5. (a) График Фрая, показывающий пространственное распределение точек Фрая, полученных по 63 месторождениям меди; и розовая диаграмма для (б) всех точек Фрая; c) точки в пределах 4,5 км; (d) Точки в пределах 1,5 км.

Результаты анализа Fry делают вывод о различном контроле направления в региональном (> 4,5 км) и мелком (<4,5 км) масштабах, что может быть связано с подробными структурными особенностями в TOD. Однако такая корреляция не является специфической. Например, управление мелким трендом в мелком масштабе может быть связано с мелкими трендовыми разломами или быть связано со складками с мелкими выступающими осями. Необходим дальнейший анализ, чтобы определить взаимно однозначное соответствие между элементами управления масштабируемыми переменными и подробными структурными особенностями.

3.2. Пространственная корреляция структурных особенностей с медным оруденением.

Структурные особенности являются результатом разнообразных геологических процессов, лишь немногие из которых связаны с рудообразующими процессами и являются структурным контролем минерализации. Чтобы выявить тонкие структурные элементы управления в TOD, структурные особенности были сгруппированы с точки зрения их ориентации, и пространственные ассоциации этих особенностей с медной минерализацией были количественно оценены с помощью анализа распределения расстояний. Область исследования была разделена на 20250 квадратных ячеек с длиной стороны 200 м, среди которых 63 ячейки, содержащие залежи меди, представляют собой образцы залежей, а остальные ячейки взяты как безрудные образцы.

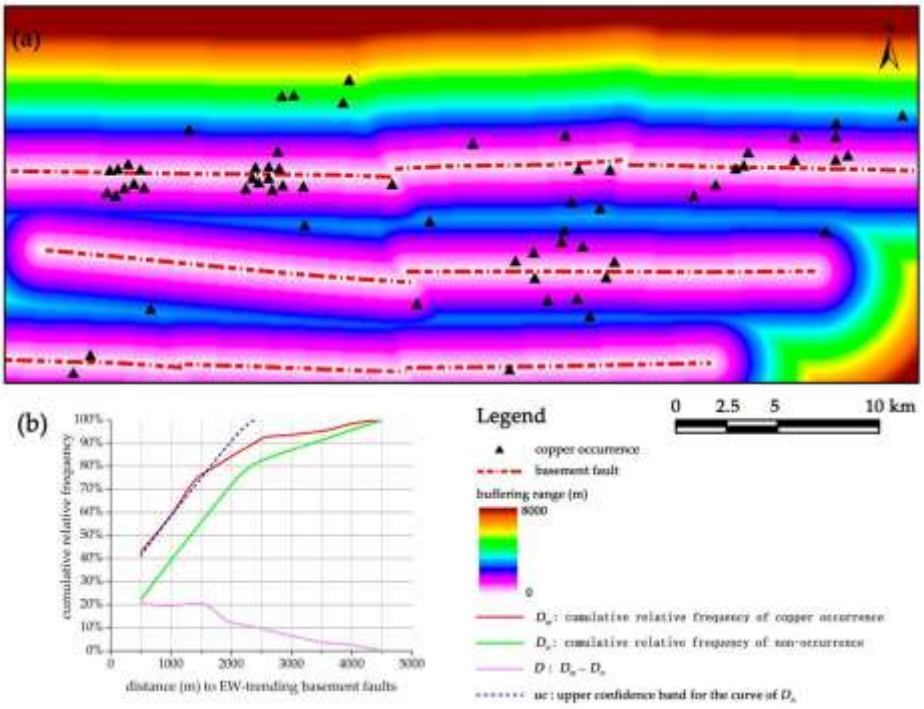

Рисунок 6. (а) Анализ буфера и (б) график кумулятивной относительной частоты относительно расстояния до разломов EW-тренда.

Разломы фундамента с тенденцией EW демонстрируют положительную корреляцию с залеганием меди в соответствии с кривой D (рис. 6). В пределах оптимального буферного расстояния в 1,5 км частота встречаемости меди не более чем на 21% выше, чем можно было бы ожидать из-за случайности. Проверяется, что такая корреляция является статистически значимой (при $\alpha = 0{,}01$), поскольку кривая Dm нанесена выше верхней доверительной полосы Dnв пределах 1,5-километрового буфера (рис. 6b).

Разломы фундамента с NS-трендом имеют положительную, но слабую связь с залеганием меди за пределами буферного расстояния в 1 км, достигая лишь на 2% большей частоты, чем можно было бы ожидать (рис. 7). Однако кривая Dm нанесена ниже верхней доверительной полосы Dn во всем диапазоне буферного анализа (рис. 7b), что указывает на то, что слабая связь между разломами с NS-трендом и медной минерализацией не имеет статистической значимости.

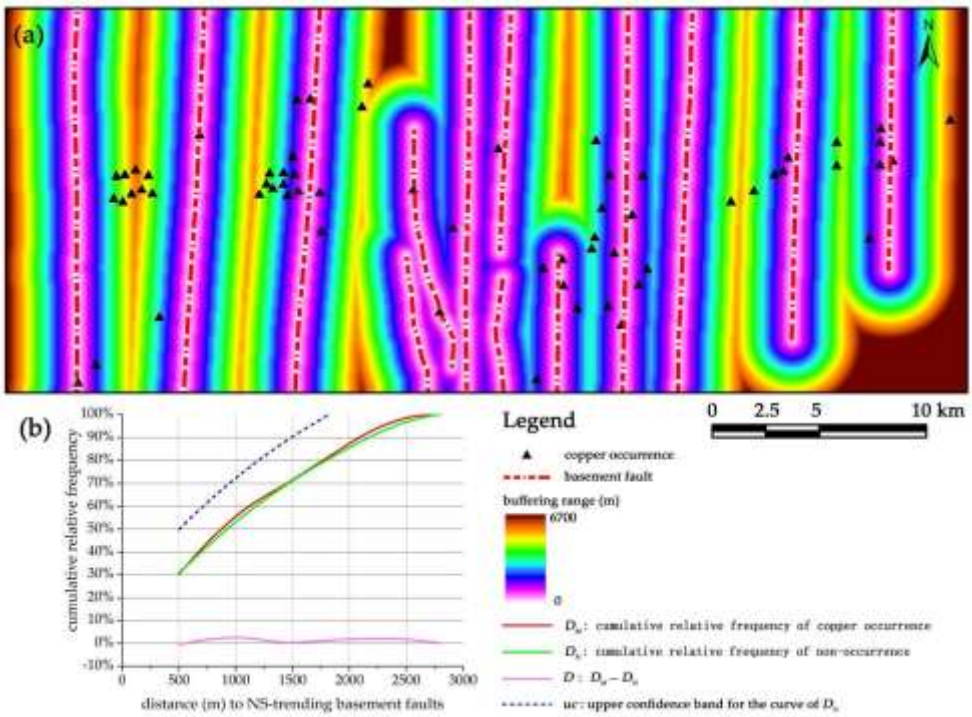

Рисунок 7. (а) Анализ буфера и (б) график кумулятивной относительной частоты относительно расстояния до NW разломы.

Пересечения разломов фундаментов имеют статистически значимую положительную корреляцию с залеганием меди между буферными расстояниями 2 и 3 км (рис. 8). При оптимальном буферном расстоянии 2,5 км частота встречаемости меди на 23% выше, чем можно было бы ожидать (рис. 8б).

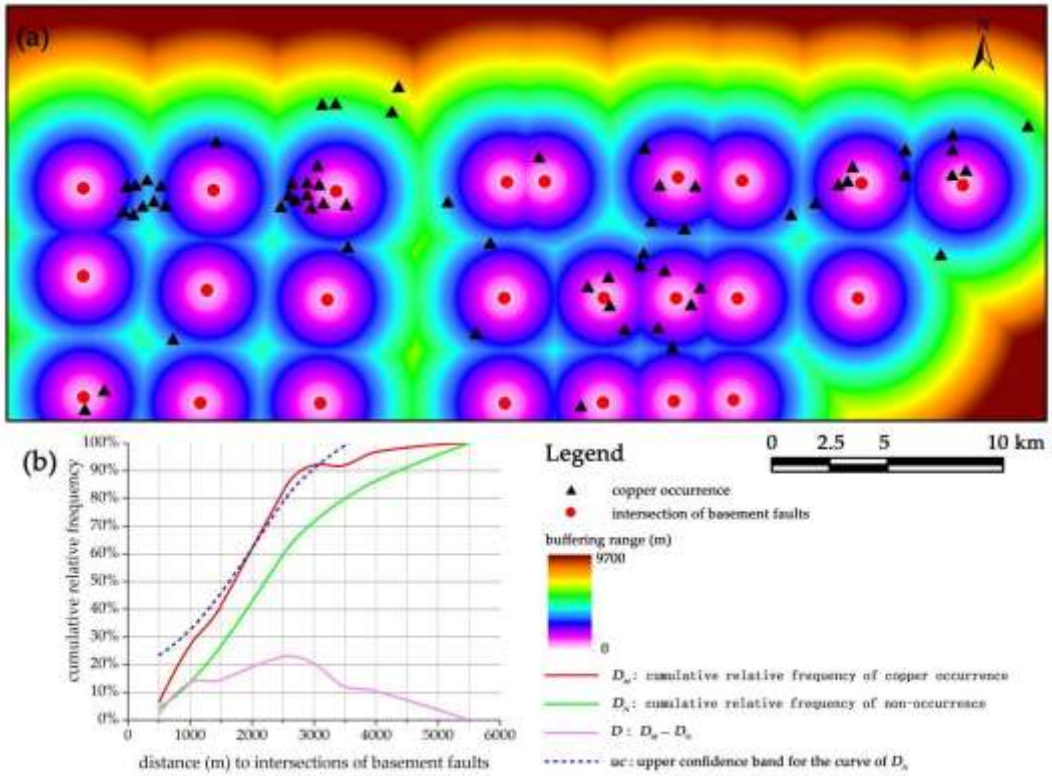

Рисунок 8. (а) Анализ буфера и (б) график кумулятивной относительной частоты относительно расстояния до пересечения разломов фундамента.

Складки демонстрируют статистически значимую положительную корреляцию с содержанием меди в буферах в диапазоне от 1,5 до 3 км (рис. 9). Частота встречаемости меди на 22% выше, чем можно было бы ожидать при 2,5-километровом буфере (рис. 9б).

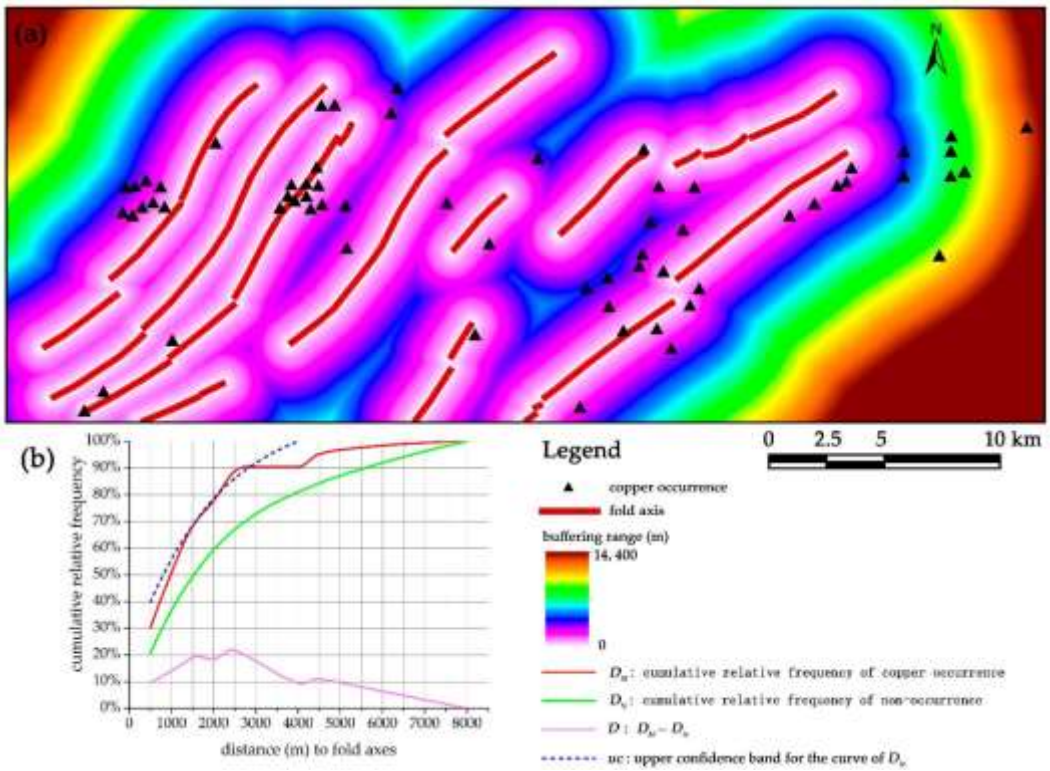

Рисунок 9. (а) Анализ буфера и (б) график кумулятивной относительной частоты относительно расстояния к складкам.

Покровные разломы, состоящие из разломов северо-западного и северо-западного направлений, а также пересечения этих разломов демонстрируют положительную пространственную связь с залежами меди. Частота встречаемости не более чем на 11%, 10% и 9% выше, чем можно было бы ожидать в пределах оптимальных буферов разломов северо-западного и северо-западного направлений и их пересечений соответственно (рис. 10, 11 и 12). Тем не менее, ни одна из этих структурных особенностей не имеет статистически значимой связи с появлением меди на любом буферном расстоянии (рис. 10b, 11b и 12b).

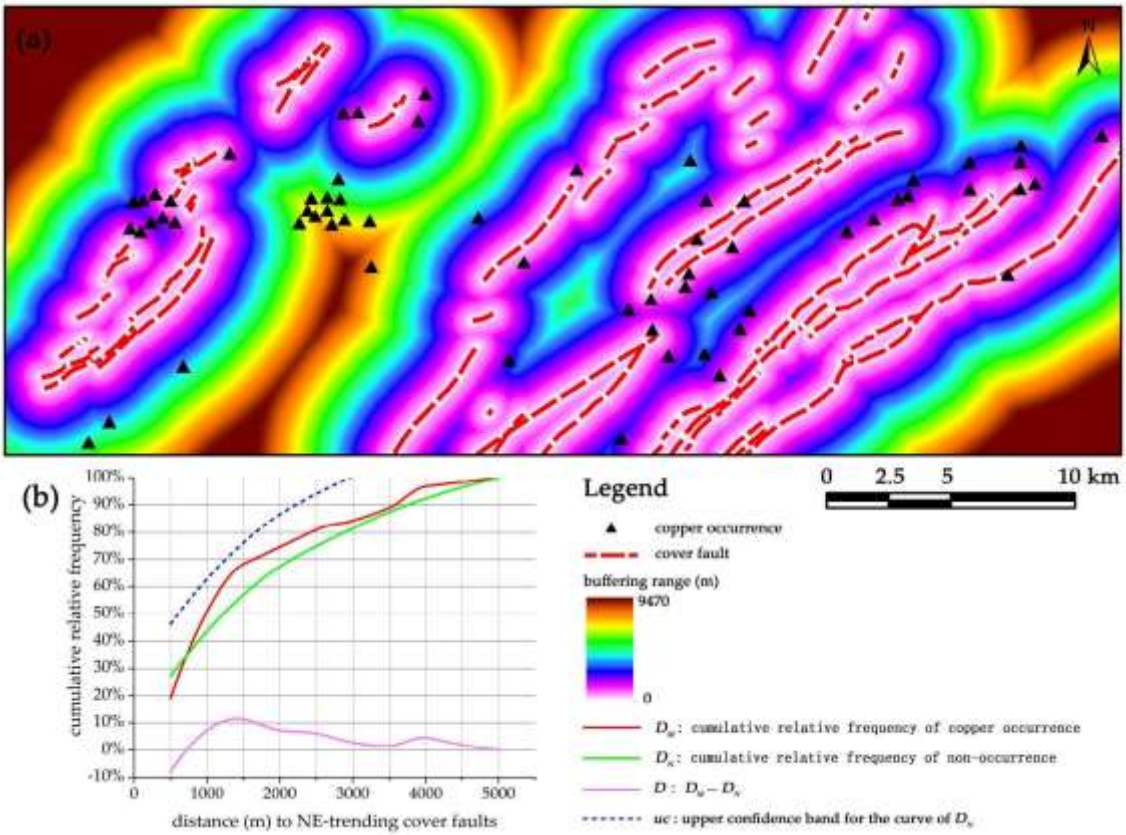

Рисунок 10. (а) Анализ буфера и (б) график кумулятивной относительной частоты относительно расстояния до разломов NS простирания.

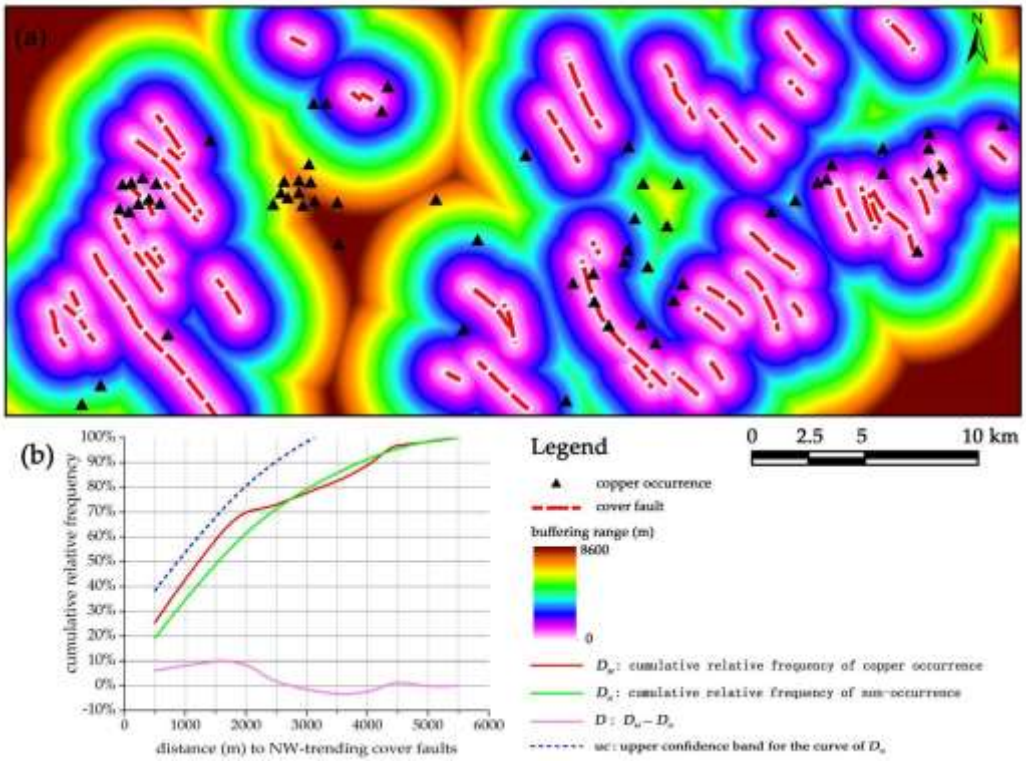

Рисунок 11. (а) Анализ буфера и (б) график кумулятивной относительной частоты в зависимости от расстояния до разломов NW простирания.

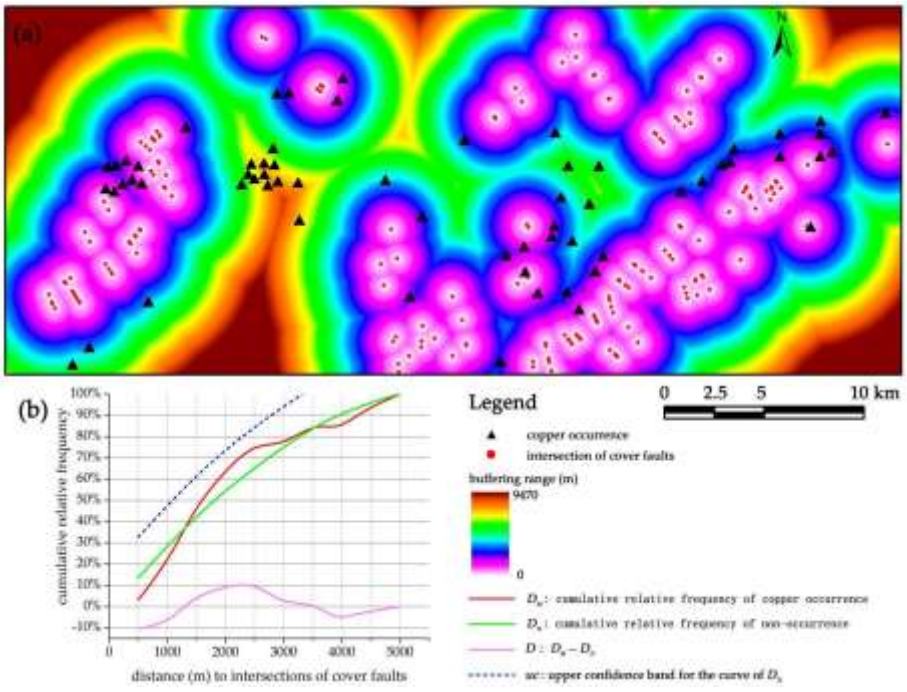

Рисунок 12. (а) Анализ буфера и (б) график кумулятивной относительной частоты относительно расстояния до пересечения покровных разломов.

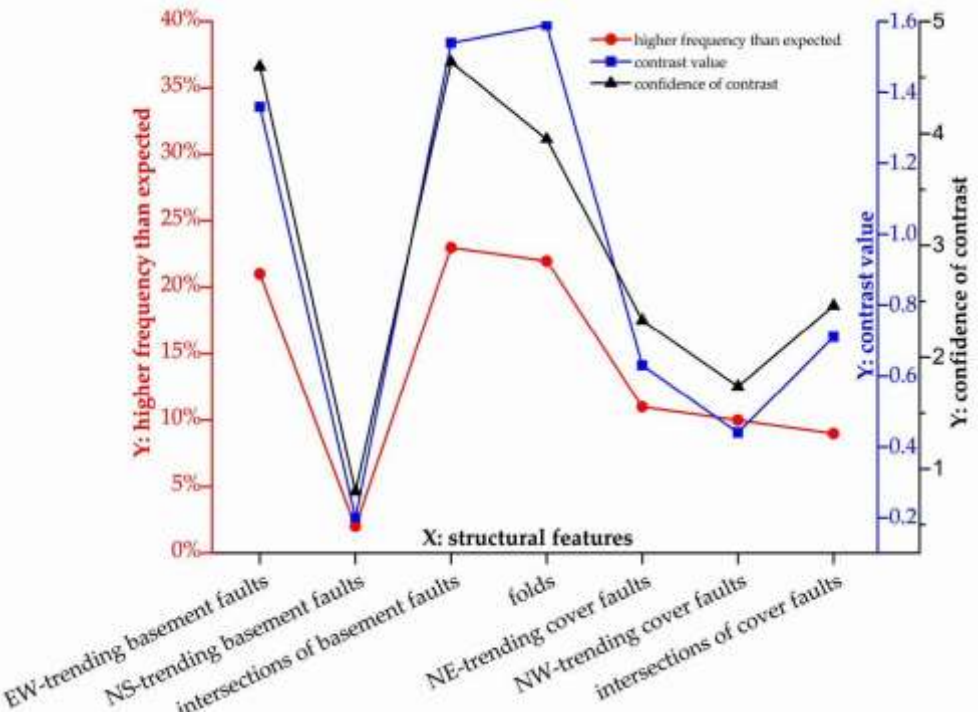

Рисунок 13. График, показывающий вариации более высоких частот, чем ожидалось, контрасты и достоверность контраст подробных структурных особенностей в TOD.

Таблица 2. **Результаты распределения по расстоянию и анализа WifE.**

Structural Features	Optimal Buffer Distance (m)	Distance Distribution Analysis				WofE Analysis	
		D_M	D_N	D	uc	C	C_s
EW-trending faults	1500	76%	55%	21%	75%	1.36	4.6
NS-trending faults	1000	55%	53%	2%	72%	0.2	0.8
intersections of basement faults	2500	83%	60%	23%	79%	1.54	4.64
folds	2500	89%	67%	22%	86%	1.59	3.95
NE-trending faults	1500	68%	57%	11%	76%	0.63	2.33
NW-trending faults	1500	59%	49%	10%	68%	0.44	1.74
intersections of cover faults	2500	74%	65%	9%	84%	0.71	2.46
contact of Yanshanian intrusion	350	87%	21%	55%	52%	3.04	8.03

D_M: cumulative relative frequency of copper occurrence; D_N: cumulative relative frequency of non-occurrence; D: $D_N - D_N$; uc: upper confidence band for the curve D_N; C: contrast value; and C_s: confidence of contrast.

Анализ Wo*f*E был также применен для изучения связи структурных особенностей с содержанием меди. При соответствующих оптимальных расстояниях буфера были рассчитаны значения контрастности и доверительные значения контрастности. Как показано на рис. 13 и в таблице 2, разломы с тенденцией РЭБ, пересечения разломов фундамента и складок имеют три самых высоких значения как контрастности, так и достоверности контраста, которые значительно выше, чем у других структурных элементов. Контрасты и достоверность контрастности, которые могут помочь в оценке интенсивности пространственной ассоциации, демонстрируют точно такие же вариации, что и результаты, полученные в результате анализа распределения расстояний, подразумевая, что разломы с обратным направлением, пересечения разломов фундамента и складок, вероятно, являются основными структурными регуляторами минерализации меди в TOD.

3.3. Пространственная корреляция разломов с интрузиями.

Учитывая, что в месторождениях меди в ТОД преобладают скарновые отложения, яншаньская интрузия является ключевым рудоуправляющим фактором, и ее контакт со стеновой породой можно рассматривать как особую структуру. Результат анализа распределения расстояний показывает самую сильную связь контакта с медными объектами. В пределах буфера в 1,7 км от контакта частота встречаемости меди не более чем на 55% выше, чем можно было бы ожидать, и такая сильная связь оказывается статистически значимой (рис. 14). Анализ Wo*f*E дает значение контраста 3,04 и доверительное значение 8,03, которые заметно выше, чем соответствующие значения других структурных особенностей (таблица 2), что подтверждает наиболее значительную связь контакта с медной минерализацией.

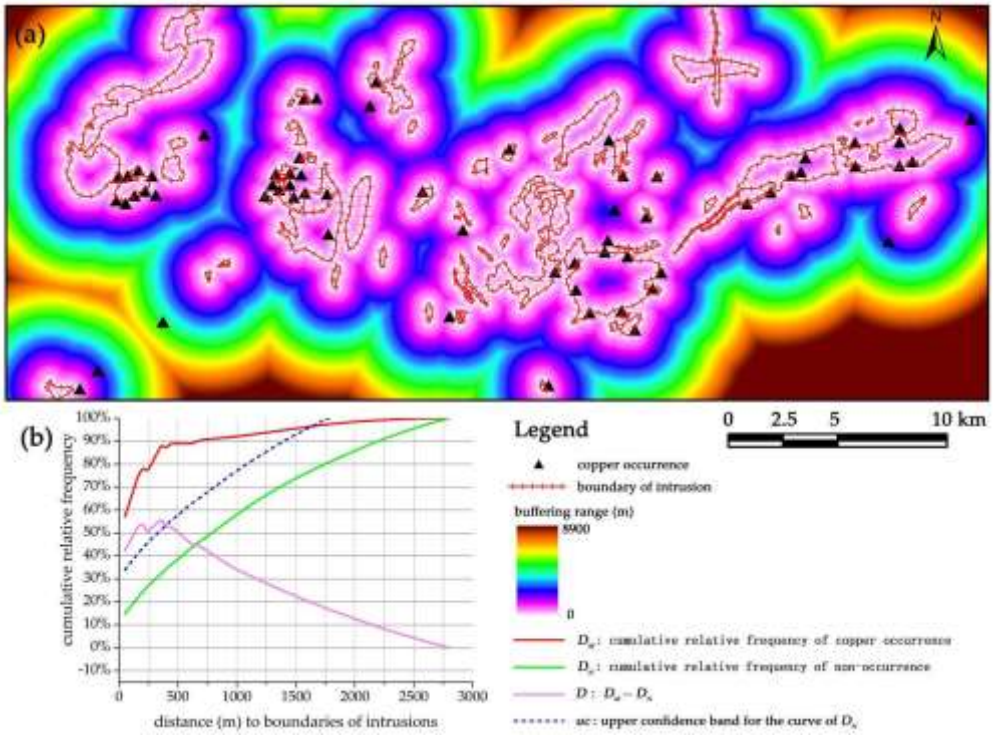

Рисунок 14. (а) Анализ буфера и (б) график кумулятивной относительной частоты относительно расстояния до границы вторжений.

Поскольку считается, что региональные разломы контролируют размещение интрузий согласно многим предыдущим исследованиям, мы также провели анализ распределения расстояний, чтобы исследовать корреляцию интрузии с разломами различной ориентации. Результаты показывают, что EW-, NS-трендовые разломы и их пересечения имеют статистически значимые положительные корреляции с областями вторжения на большинстве буферных расстояний (рис. 15a–c). Частота областей вторжения на 26% и 17% выше, чем можно было бы ожидать при оптимальных буферах пересечений разломов фундамента и разломов РЭБ, соответственно (рис. 15a, c), что предполагает сильную связь этих структурных особенностей с вторжениями.

Разломы фундамента с NS-трендом имеют умеренную корреляцию с интрузией, очерченную на 11% более высокой частотой областей интрузии, чем можно было бы ожидать (рис. 15b). Напротив, разломы покрытия и их пересечения показывают отрицательную корреляцию с областями вторжения в пределах 1,5-километрового буфера (рис. 15д–ф). За пределами буферного расстояния в 1,5 км они демонстрируют положительные, но слабые ассоциации с вторжениями. Частота областей вторжения на 6%, 9% и 5% выше, чем можно было бы ожидать при оптимальных буферах разломов северо-западного и северо-западного направлений и их пересечений соответственно (рис. 15d–f).

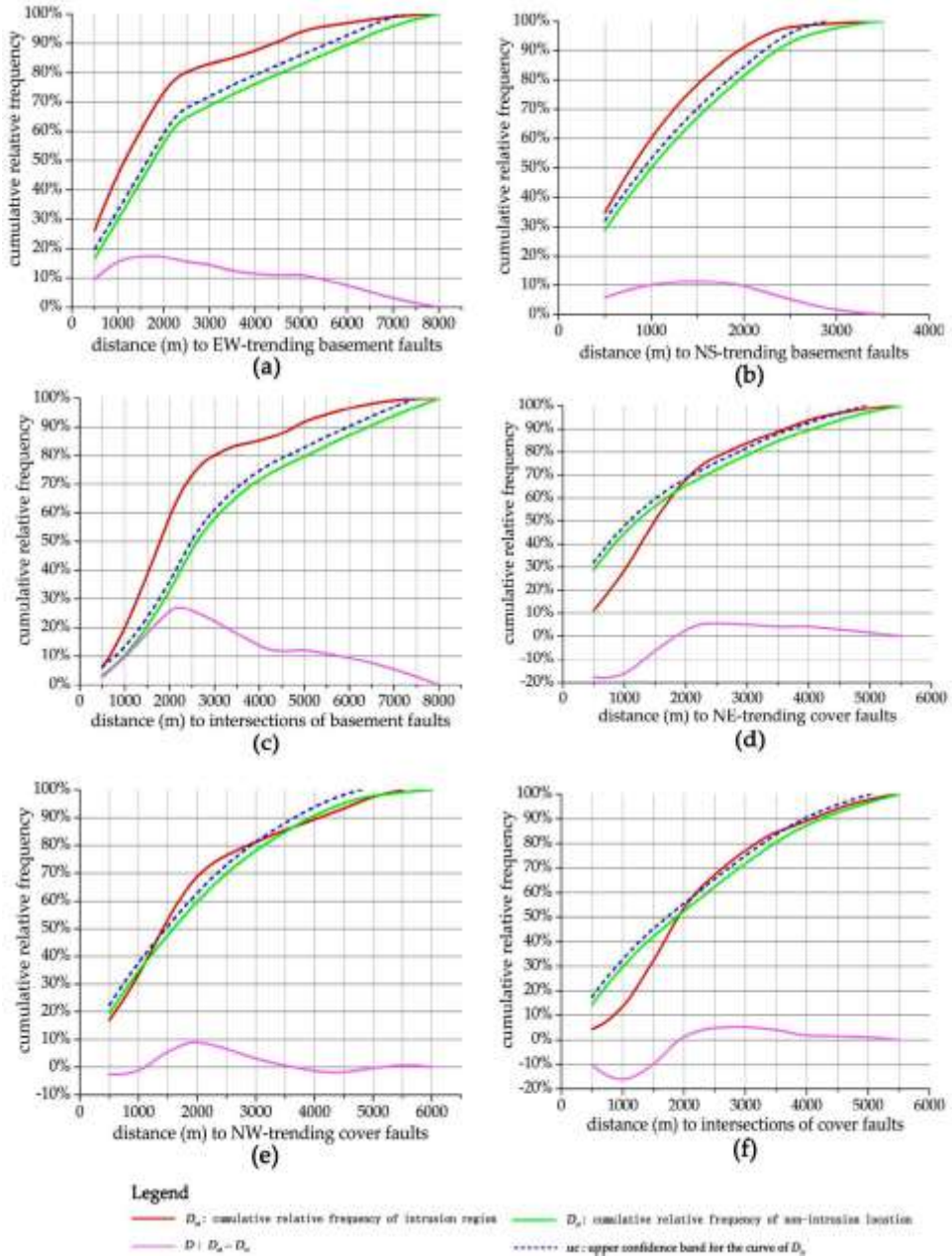

Рис. 15. График кумулятивной относительной частоты в зависимости от расстояния до (а) разломов северо-западного простирания; (б) разломы субмеридионального простирания; в – пересечения разломов фундамента; г – разломы северо-восточного простирания; (е) СЗ простирание разломов и (е) пересечения покровных разломов.

Примечательно, что разломы с EW-трендом и пересечения разломов фундамента, которые демонстрируют самые сильные корреляции с интрузиями, также демонстрируют значительные ассоциации с залежами меди в предыдущем анализе распределения расстояний. Необходимо оценить, какая степень этого структурного контроля над интрузией определяет их сильную корреляцию с медной минерализацией. Разломы с РЭБ-трендом и яншаньские интрузии были буферизованы с их оптимальными расстояниями, и были подсчитаны залежи меди, расположенные в соответствующих буферных зонах. По-видимому, 98% (47 из 48) залежей меди, распределенных в пределах буферов разломов с тенденцией РЭБ, расположены в перекрывающихся зонах буферных разломов и интрузий с тенденцией РЭБ, на долю которых приходится 33,58% от общей площади. В буферные зоны, где интрузии отсутствуют, включено только одно месторождение (занимающее 66,42% от общей площади). (рис. 16). Аналогичным образом, 96% (49 из 51) залежей меди, расположенных в пределах буферов пересечений разломов фундамента, включены в перекрывающиеся зоны буферных пересечений разломов и интрузий, которые занимают 37,11% от общей площади (рис. 17). Делается вывод о том, что значительно сильные ассоциации разломов EW-trending и пересечений разломов фундамента с медной минерализацией объясняются контролем этих структурных особенностей интрузий.

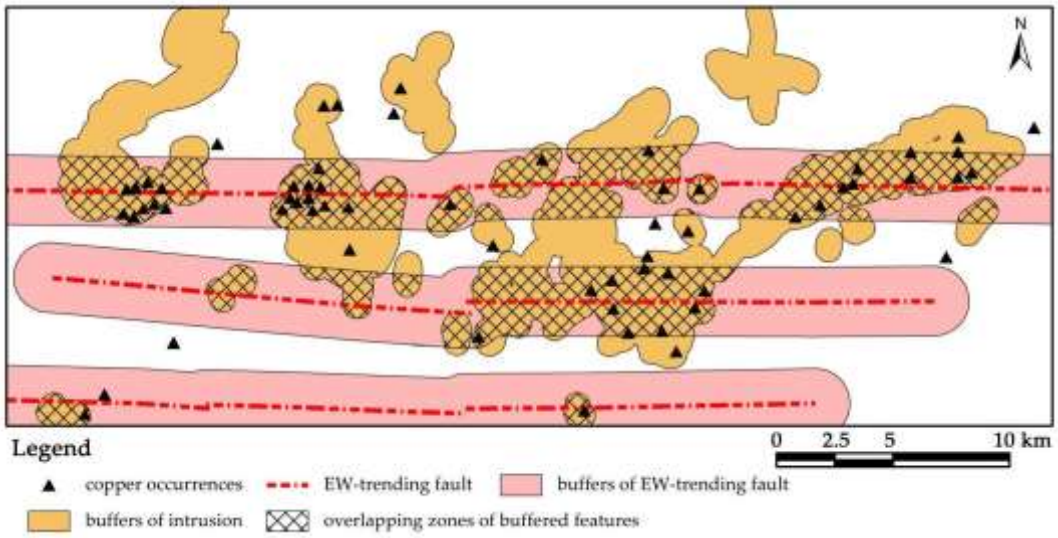

Рисунок 16. Анализ буфера, показывающий распределение проявлений меди в буферизованном EW-тренде неисправности и вторжения.

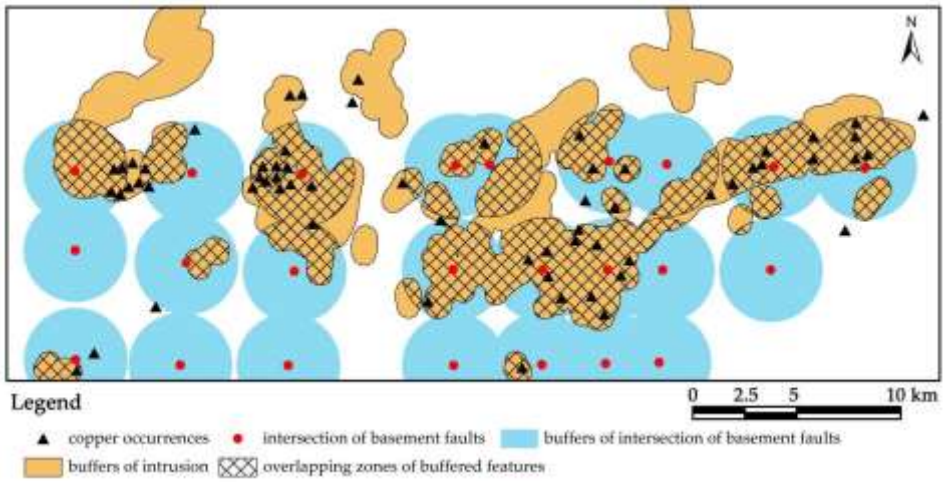

Рисунок 17. Анализ буфера, показывающий распределение медных вхождений в буферизованных пересечениях разломы и интрузии фундамента.

3.4. Интерпретация структурных регуляторов минерализации меди

Тектоническую эволюцию ТОД можно разделить на четыре этапа. Первый этап - это формирование и развитие фундамента Нижнего рельефа Янцзы (LYT) до движения Цзиньнин (около 850-800 млн лет назад), когда ТОД все еще был неотъемлемой частью LYT. Во-вторых, после движения Цзиньнин и до индосинского движения (около 195 млн лет назад) ЛИТ постепенно превратился в архипелажную океаническую стадию, и в этом регионе сформировался основной осадочный покров. В то же время катайский блок и LYT постепенно дрейфовали к Северо-Китайскому кратону, что привело к нескольким мягким столкновениям. На этом этапе преобладало вертикальное движение, вызванное эффектом открывания-закрывания, связанным с мягким столкновением, что привело к некоторым несоответствиям (таблица 1). В-третьих, конвергенция кратона Янцзы и Северо-Китайского кратона (называемая индосинским горообразованием) началась в конце триаса, что привело к формированию ряда существенных структурных особенностей, включая угловое несоответствие между триасовыми и юрскими слоями (таблица 1), складками и разломами. Считается, что индосинское движение создало нынешнюю структурную структуру в ТОД и даже в Южном Китае. В конце концов, TOD испытал яншаньское движение (около 135 млн лет), характеризующееся переходом от сжатия к расширению с раннего мела, что привело к образованию широко распространенных промежуточно-кислых интрузий и связанной с ними минерализации. Многоступенчатая тектоническая эволюция ответственна за структурные особенности как фундамента, так и осадочного чехла, которые связаны с эпигенетической медной минерализацией.

В структурах фундамента преобладают разломы EW- и NS-трендов. Считается, что эти разломы, полностью перекрытые мезозойскими слоями, образовались до

индосского периода и активизировались в мезозое, хотя подробные геометрические и кинематические характеристики этих разломов до сих пор не ясны. В предыдущих исследованиях было доказано, что разломы действуют как благоприятные пути для транспортировки связанной с рудой магмы и рудообразующих флюидов из глубоких источников в неглубокие зоны ловушек, что приводит к тесной связи этих разломов с гидротермальными месторождениями. В районе исследования петрологические данные и геофизические профили свидетельствуют о том, что в мезозое на расстоянии около -10 км от поверхности сформировался магматический очаг. Интерпретируется, что разломы фундамента с EW-трендом играют жизненно важную роль в направлении магмы из магматической камеры в неглубокие зоны ловушек в яншаньский период. Этот значительный контроль разломами, простирающимися в направлении EW, на яншаньских интрузиях подтверждается анализом распределения расстояний и Wo*f*E, который полностью отвечает за сильную корреляцию между разломами, простирающимися в направлении EW, и минерализацией меди. Эта интерпретация может объяснить результат анализа Fry, который демонстрирует преобладающую тенденцию РЭБ в региональном масштабе.

Известные месторождения меди расположены в осадочном чехле, где доминирующими структурами являются складки с осями сигмоидальной формы, поэтому определение процесса формирования складок имеет решающее значение для понимания структурного каркаса и минерализации меди на уровне чехла. Поскольку самый молодой слой, вовлеченный в складки, относится к среднему триасу, делается вывод, что складки образовались во время индосинского движения, что привело к угловому несоответствию между средним триасом и нижней юрой (таблица 1). Классическая модель деформации правого простого сдвига в зоне сдвигового разлома вводится для иллюстрации формирования складок и разломов при деформационном режиме индосинского движения, в котором преобладают сжатие северо-запада и юго-востока и сдвиг вправо (рис. 18 и 19). По мере возникновения зоны разлома формируется структурная система, состоящая из (1) сопряженных сдвиговых разломов, (2) складок, (3) обратных разломов и (4) нормальных разломов (рис. 19a). Первоначально сформированные складки и обратные разломы имеют тенденцию, перпендикулярную направлению наибольшего укорочения, в то время как нормальные разломы имеют тенденцию, параллельную направлению наибольшего укорочения. Впоследствии продолжающийся сдвиг при ударе может привести к вращению элементов в этой системе. Оси ранее сформированных складок приобретают сигмоидальную форму. Ранее сформированные нормальные разломы

приспосабливаются к синестральному движению удара-скольжения, а обратные разломы приспосабливаются к правостороннему движению удара-скольжения (рис. 19б). Надвиговые разломы с мелким трендом, наблюдаемые в полевых условиях и смещения СЗ-трендовых разломов, выявленных на геологической карте (рис. 1), подтверждают рациональность этой модели.

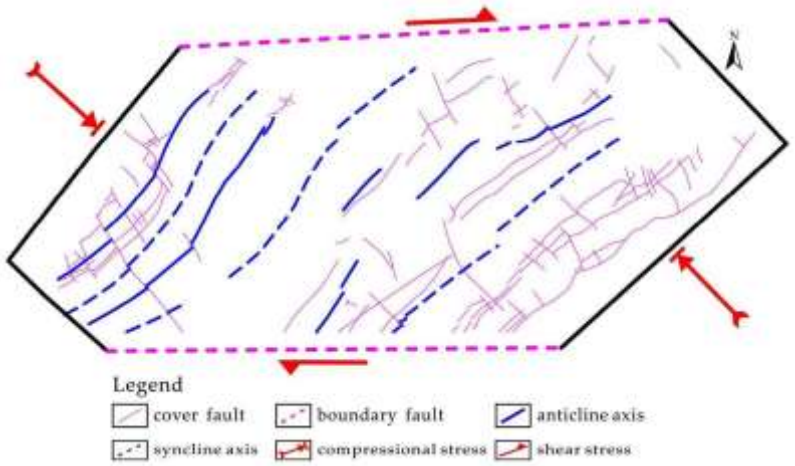

Рис. 18. Напряженный режим в процессе формирования складок с сигмоидальными осями.

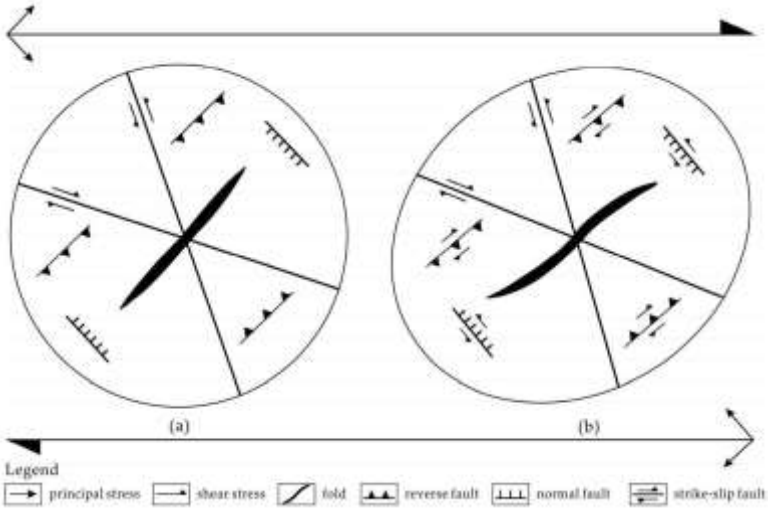

Рис. 19. Деформационная модель правостороннего сдвига в зоне сдвигового разлома, а) структурная система, сформированная на начальной стадии деформации; и (б) поворот структурных элементов при продолжающемся сдвиге.

В мезозойских слоях существовало несколько границ раздела между двумя соседними слоями, которые имеют различные механические свойства, некоторые из которых также представляли собой границы несоответствия, например, граница раздела между кварцевым песчаником верхнего девона и известняком верхнего карбона. Во время процесса формирования складок в индосский период вышеупомянутые границы раздела подвергались прогрессирующей деформации складчатости и сдвига, что приводило к образованию обширных параллельных слоям зон сдвига (рис. 20 и 21).

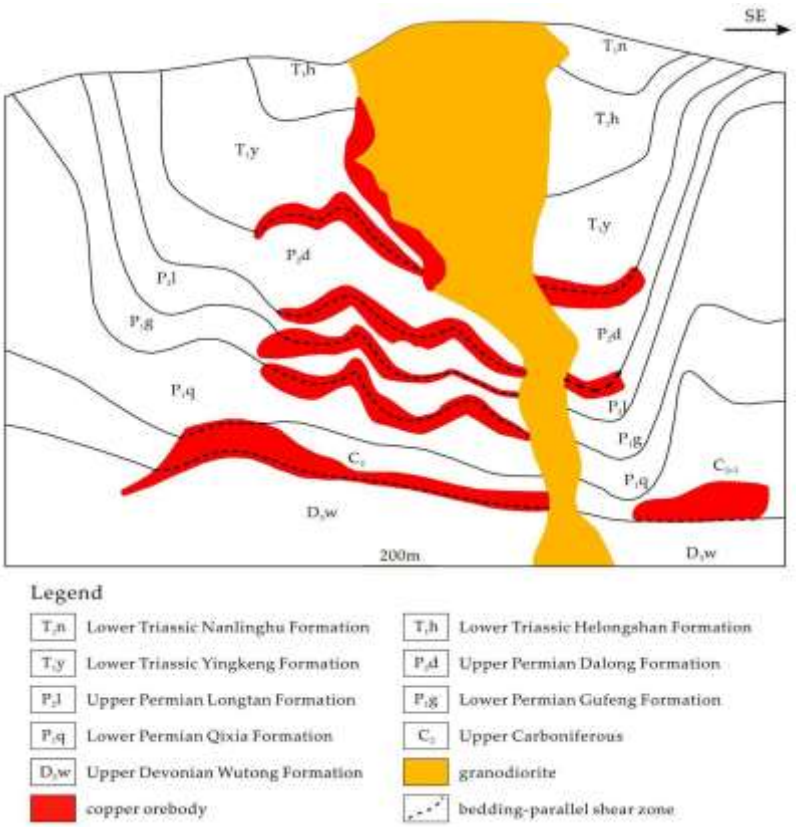

Рисунок 20. Типичный поперечный разрез Шизишанского рудного поля, показывающий характерный стратифицированный скарн и рудные тела, залегающие в складках.

Рис. 21. Полевая фотография напластованно-параллельной сдвиговой зоны между известняками и кварцевыми песчаниками месторождения Синьцяо.

В частности, отсоединения подстилки происходят в сердцевинах складок из-за проскальзывания параллельно слоям в процессе формирования складок. Эти зоны сдвига были перекрыты деформацией растяжения в меловой период, когда тектонический режим в этом регионе изменился от сжатия к растяжению, что благоприятствовало улавливанию и локализации минерализованных флюидов. Этот вывод подтверждается:

1. На рисунке 22 четко различимые границы между слоистыми рудными телами и вмещающими породами предполагают, что руды были отложены в пространствах механического расширения

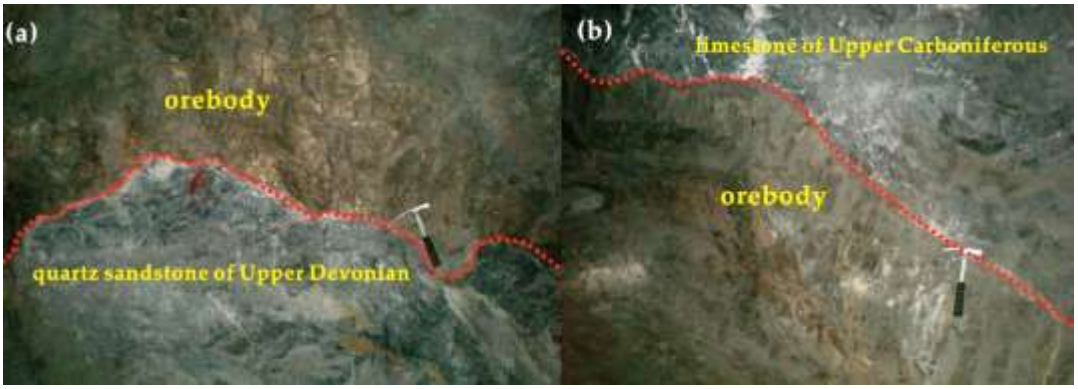

Рис. 22. Фотографии, показывающие несогласующиеся границы между стратиформными рудными телами и породами месторождения Синьцяо. a – граница между рудным телом и подстилающим верхнедевонским кварцем, песчаником; и (b) граница между рудным телом и вышележащим верхнекаменноугольным известняком.

2. Результатом численного моделирования на месторождении Донгуашань, которое демонстрирует, что стратиформные зоны, вызванные напряжением растяжения, благоприятны для рудных флюидов и согласуются с рудными телами. Кроме того, зоны ловушек, параллельные слоистому слою, расположены вблизи контактов интрузий, где имеются достаточные источники тепла и жидкости, и расположены в наборе карбонатных слоев, подходящих для образования скарна (рис. 20). Следовательно, структуры, параллельные слоям, в складчатых слоях являются благоприятными для размещения, сосредоточения и осаждения рудоносных флюидов, способствуя образованию слоистых рудных тел. Утолщение рудных тел в ядрах складок объясняется происходящими там отслоениями (например, крупное рудное тело в пределах C_2 на рисунке 20). Эта интерпретация подтверждается анализами распределения расстояний и WofE, которые демонстрируют сильную пространственную связь складок с медной минерализацией. Также предполагается, что доминирование трендов NE, NNE и NEE на диаграммах роз точек Фрай в мелком масштабе (<4,5 км) объясняется контролем складок с осями простирания NE, а не разломов с NE, которые показывают плохую корреляцию с минерализацией меди. Более того, ни покрывающие разломы различной ориентации, ни пересечения этих разломов не демонстрируют статистически значимой корреляции с медной минерализацией, предполагая, что они могут играть роль только в миграции рудоносных флюидов к благоприятным вмещающим структурам (т.е. многослойным слоисто-параллельным зонам сдвига), где произошла концентрация флюидов и отложение минералов.

4. Выводы

1. Фрактальные размеры, полученные в результате подсчета квадратов и анализа радиальной плотности, предполагают, что различные структурные элементы управления работают в разных масштабах <1,5 км, 1,5-4,5 км и> 4,5 км. Это управляемое поведение, зависящее от масштаба, подтверждается и изучается результатами анализа Fry, который иллюстрирует доминирующую тенденцию РЭБ в региональном масштабе (> 4,5 км) и преимущественные тенденции NE-NNE-NEE в мелком масштабе (<4,5 км).

2. Пространственные ассоциации детальных структурных особенностей с медной минерализацией дополнительно исследуются с помощью количественного пространственного анализа. Яншаньские интрузии, разломы EW-trending, пересечения разломов фундамента и складки имеют значительные ассоциации с медным оруденением, на что указывают их высокие значения количественных параметров как в анализе распределения расстояний, так и в анализе WofE.

3. Интерпретация структурных регуляторов минерализации меди производится в сочетании с приведенными выше аналитическими результатами. Изменяющиеся в масштабе закономерности залегания полезных ископаемых объясняются различными структурными регуляторами, действующими в фундаменте и осадочном чехле. В фундаменте разломы с обратным направлением служат путями для направления магмы из магматической камеры в зоны ловушек в яншаньский период. Значительный контроль разломов с тенденцией EW яншаньской интрузии полностью отвечает за сильную корреляцию между разломами с тенденцией EW и минерализацией меди. Этот вывод подтверждается результатами анализа Fry, который показывает доминирующую тенденцию РЭБ в региональном масштабе (> 4,5 км).

В осадочном чехле зоны сдвига, параллельные слоям, образовавшиеся во время складчатости и деформаций растяжения в яншаньский период, действуют как рудолокализующие структуры, что соответствует доминированию трендов NE-NNE-NEE в мелком масштабе (<4.5 м). Такие слоисто-параллельные структуры вместе с контактными зонами интрузии оказывают важное влияние на формирование характерных слоистых рудных тел скарнов в TOD.

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ПРОГНОЗИРОВАНИЯ ЦЕЛИ ПОИСКА
(На основе сети выборочного переноса) *[3].*

1. Введение

В последние годы развитие технологий машинного обучения достигло значительного успеха во многих областях, например, в компьютерном зрении, обработке естественного языка, радаре с синтезированной апертурой и др. Методология также используется при различных геологических исследованиях - традиционный метод геологоразведки постепенно превращается в интеллектуальный метод, основанный на машинном обучении.

В частности, технологии машинного обучения могут использоваться для определения корреляции между геологическими и металлогеническими переменными характеристиками и прогнозирования металлогенических целей. Хотя количественное прогнозирование целей поиска на основе технологий машинного обучения находится в зачаточном состоянии, был достигнут определенный прогресс. Эти методы можно разделить на следующие три категории в зависимости от способа выполнения:

(1) Методы, основанные на ансамблевом обучении, объединяют несколько алгоритмов контролируемого обучения для поиска целевого прогноза. Например, определяются гиперпараметры «случайного леса» путем моделирования процесса естественной эволюции, которые были использованы для повышения точности модели при прогнозировании целевой области поиска. Использовали алгоритм «изолирующего леса» для прогнозирования выбросов для определения целевой области поиска. Было предложено использовать метрическое обучение в случайном лесу для проецирования выборочных объектов в пространство объектов, разделяя фоновые и целевые объекты интеллектуального анализа данных, чтобы повысить точность прогнозирования модели.

(2) Методы, основанные на опорных векторах (SVM), разделяют металлогенические и др. выборки через гиперплоскость. Например, отделение образцов “мин” (рудных) от образцов “без мин” (безрудных) через оптимальную гиперплоскость и определение трех областей поиска целей по обе стороны от гиперплоскости. Используется генетический алгоритм для оптимизации гиперпараметров SVM, чтобы уменьшить его влияние на прогнозирование цели поиска.

(3) Методы, основанные на глубинных нейронных сетях, проецируюшие геонаучные данные в одно и то же пространство глубинных нейронных сетей и извлекающие эффективные характеристики с помощью множества нелинейных

преобразований для прогнозирования целей поисков. Например, используется трехслойная свертка для извлечения особенностей распределения и концентрации Zn-элемента для прогнозирования целевой поиска. Используется AlexNet для извлечения особенностей нескольких схем рудообразующих факторов для определения четырех целевых областей поиска.

Несмотря на то, что вышеупомянутые методы достигли успеха, такие проблемы, как небольшое количество геологических данных и нерегулярные особенности районов поисков при прогнозировании, не были решены должным образом. В последние годы по этим проблемам был проведен ряд исследований. Эти работы можно разделить на следующие две категории в зависимости от способа выполнения:

(1) Методы, основанные на дополнении данных, увеличивающие разнообразие за счет обрезки, изменения хроматической аберрации и размера, а также искажения объектов. Например, добавление случайных шумов в геологические данные, чтобы предсказать поисковую цель с помощью глубокой сверточной сети. Сначала производится передискретизация признаков оруденения, а затем использование случайного леса для определения целевых зон. Предложено рекомбинировать пары пикселей геологических данных, чтобы помочь в прогнозировании целей. Эти методы увеличивают количество геологических данных, но не могут справиться с нерегулярными особенностями районов поисков.

(2) Методы, основанные на многомасштабном преобразовании признаков разного масштаба за счет нерегулярной выборки для обучения. Например, предложено извлекать нерегулярные объекты разных масштабов с помощью многогрупповой свертки или операции объединения для прогнозирования целей. Используются четыре операции свертки с различными размерами для извлечения и объединения нерегулярных элементов геологических данных, для повышения точности прогнозирования.

Однако эти методы не учитывают небольшое количество данных в целевой зоне поиска, поэтому дальнейшая разработка в двух направлениях:

(1) Выборочная передача информации о метках, полученных из больших задач, в небольшие задачи, что может повысить эффективность обобщения (необходимо выборочно передавать данные из хорошо подготовленной задачи прогнозирования крупных поисковых целей в целевую задачу для помощи в обучении).

(2) Расширение восприимчивого поля ядра свертки без увеличения количества параметров. Таким образом, можно его использовать для решения нестандартных особенностей районов поисков.

Необходимо объединять эти два мотива и предлагать новую структуру глубокого обучения, основанную на *выборочной передаче знаний (SKT)* для прогнозирования целей поиска.

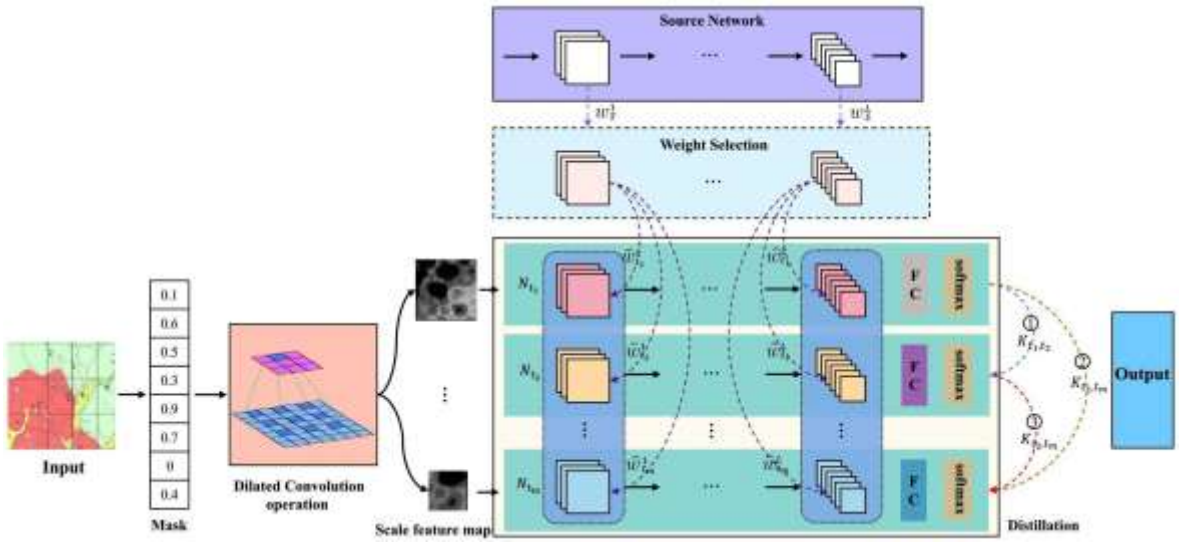

Рисунок 1. Схематическая архитектура платформы SKT. w l S обозначает ядра свертки l-го слоя свертки исходной сети, Nt v, v = 1, 2, . . . , m – целевые сети, ˆw l t v обозначает ядра свертки, перенесенные в Nt v из исходной сети, а Kt i, t j обозначает направление дистилляции знаний, что означает, что Nt i направляет Nt j к обучению. Дистилляция знаний идет сверху вниз, например, дистилляция знаний осуществляется в порядке 1, 2, 3. Выбор делает веса исходной сети разреженными через продукт Адамара, а затем передает в целевую сеть.

Как показано на рисунке 1, сначала вводим большую хорошо обученную сеть прогнозирования целей поиска в качестве исходной сети и определяем несколько целевых сетей с одинаковой структурой. Затем используем мягкую маску, чтобы смягчить различия в металлогенических эффектах, которые могут вызывать различные элементы. Затем используем расширенную свертку для захвата объектов разного масштаба в качестве входных данных целевых сетей. Основываясь на этом, выборочно переносим весовые параметры из исходной сети для обучения различных целевых сетей. Наконец, выполняем нисходящую дистилляцию в крупномасштабной и мелкомасштабной целевой сети, чтобы извлечь скрытые параметры между картами объектов разного масштаба для обучения мелкомасштабной целевой сети. Входной размер W×H×n получается путем обработки геохимических данных, где W обозначает ширину, H обозначает высоту, а n обозначает количество геохимических элементов. Результат генерируется путем голосования по всем целевым сетям и содержит два значения, представляющие голоса, полученные за и против, соответственно. Вся структура SKT работает комплексно. Обширные экспериментальные результаты показывают, что этот метод в значительной степени конкурирует с самыми современными методами.

Основные преимущества:

1. Предложена структура глубокого обучения для прогнозирования целей поиска, которая обеспечивает новый способ прогнозирования.

2. Разработан новый механизм выборочной передачи данных, который передает их из исходной сети в целевые сети, что повышает производительность целевых сетей при прогнозирования во время тестирования без добавления дополнительных вычислительных затрат.

3. Предлагается стратегия мягкой маски для поддержания консистенции сопутствующих минерагенических характеристик. Ее цель состоит в том, чтобы использовать металлогенически ориентированную значимость основных и сопутствующих минеральных параметров для выполнения задачи прогнозирования.

2. Область исследования и данные

2.1. Область исследования

Исследуемым является район Пангсидун в провинции Гуандун, Китай. На рис. 2 показана его геологическая карта, где красные точки обозначают м-ния полезных ископаемых. Он является частью металлогенической системы Циньчжоу–Ханчжоу, которая начиная с позднего палеозоя, испытывала длительное воздымание и являлась важной металлогенической областью для формирования м-ний драгоценных и цветных металлов, включая золотые, серебряные, свинцово–цинковые, вольфрамоввые, молибденовые, железные и др. руды. Типы месторождений в основном включают серебряно–золотые месторождения зоны пластичного сдвига, молибденово–вольфрамово–медные полиметаллические месторождения порфирового типа, осадочно–реформационного типа, магматического гидротермального типа и контактно-метасоматические свинцово-цинковые полиметаллические месторождения. Полезные ископаемые в основном распространены на северо–западе и юго-востоке, вдоль зоны разлома Пангситонг и зоны разлома Гученг-Шачан, в северо-восточном направлении.

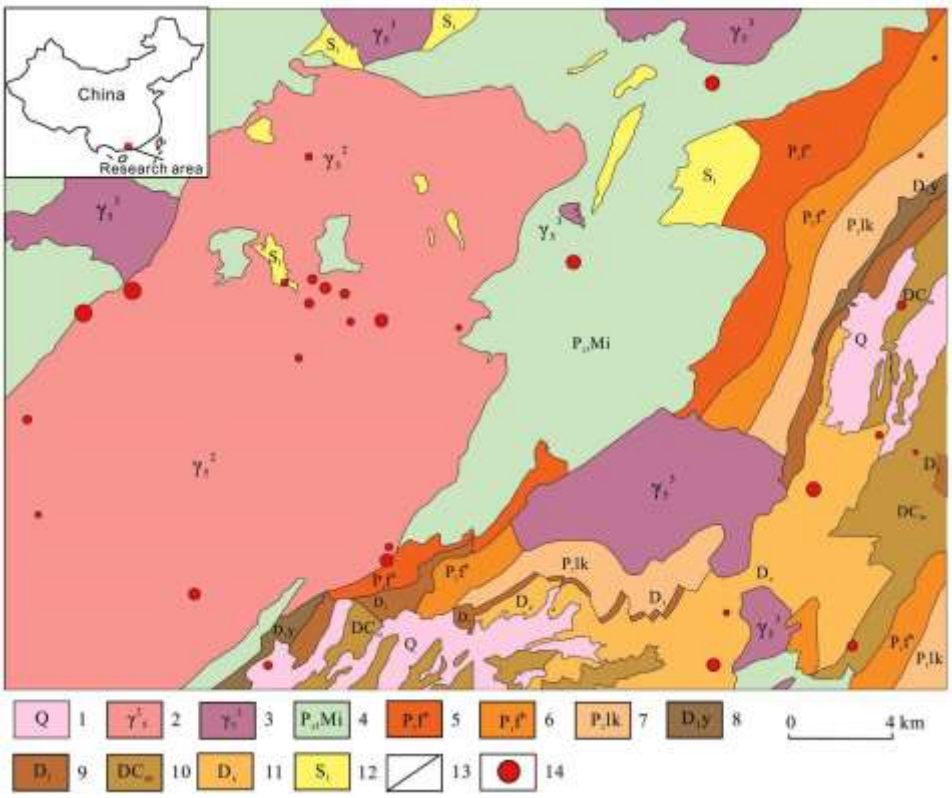

Рис. 2. Геологическая карта района исследований (1 – четвертичные; 2 – раннеяншаньские граниты; 3 – позднеяншанские граниты; 4 – верхнепротерозойские мигматиты; 5—нижняя часть средне-верхней протерозойская формация Фэндункоу; 6 – верхняя пачка среднего-верхнего протерозоя Фэндункоу. формирование; 7 – средне-верхнепротерозойская ланкенгская свита; 8 – девонская свита Янси; 9— девонская формация Лаохутоу; 10 – девонско-каменноугольная свита Маоцзыфэн; 11 — девон формация Синду; 12 – силурийская лиантаньская свита; 13 — разломы; 14—месторождения).

2.2. Обработка данных

Экспериментальные данные состоят из геохимических данных донных отложений ручьев в масштабе 1:50 000. Площадь отбора проб донных отложений составила 1694 km^2, а средняя плотность выборки составила 4,27 на km^2. Шестнадцать химических элементов, включая Au, B, Sn, Cu, Ag, Ba, Mn, Pb, Zn, As, Sb, Bi, Hg, Mo, W и F, были проанализированы из проб донных отложений. В таблице 1 приведены некоторые из исходных данных, где X и Y - координаты точек отбора проб, а остальные значения - содержания геохимических элементов, в общей сложности 7237 точек. Площадь под кривой (AUC) определяется как площадь под кривой рабочей характеристики приемника (ROC), ограниченной осями координат. На основе тезиса, использовали SVM и статистические методы для отсеивания геохимических элементов, которые указывают на минерализацию. В частности, рассчитали AUC каждого элемента с использованием алгоритма SVM, а затем стандартное отклонение AUC было рассчитано с помощью

$$S_{AUC} = \sqrt{\frac{AUC(1-AUC) + (C_p - 1)\left(Q_1 - AUC^2\right) + (C_n - 1)\left(Q_2 - AUC^2\right)}{C_p \times C_n}} \quad (1)$$

где S_{AUC} обозначает стандартное отклонение AUC; Cp и Cn номера образцов с рудой и без соответственно; и Q_1 и Q_2 являются временными переменными. Чтобы определить, значительно ли отличался AUC от 0,5, определили следующую случайную величину:

$$Z_{AUC} = \frac{AUC - 0.5}{S_{AUC}} \quad (2)$$

Таблица 1. **Геохимические данные Панксидонга.**

X	Y	Au	B	Sn	Cu	Ag	Ba	Mn	Pb	Zn	As	Sb	Bi	Hg	Mo	W	F
422.24	2418.80	0.9	3	8.7	4	0.025	33	147	27	26	1.17	0.31	0.23	0.04	2.67	0.79	212
421.37	2418.80	0.54	4	2.56	7	0.078	88	209	12	23	0.9	0.29	0.13	0.04	0.82	1.16	204
419.76	2418.25	0.81	3	1.52	5	0.043	1111	423	42	14	0.51	0.35	0.06	0.07	0.59	0.38	101
420.12	2418.40	0.37	2	1.65	6	0.046	941	498	38	17	0.53	0.31	0.1	0.02	0.57	0.33	111
420.55	2418.60	1.09	4	1.53	8	0.033	427	338	37	29	0.74	0.28	0.09	0.07	1.68	0.73	186
433.81	2397.92	2.31	121	2.2	4	0.075	365	239	16	18	4.31	0.96	0.43	0.066	0.77	3.01	186
424.17	2415.02	0.43	5	2.18	4	0.069	42	250	13	14	1.21	0.33	0.4	0.031	1.04	1.53	201
423.74	2415.31	0.51	5	4.85	7	0.004	30	242	47	31	0.5	0.26	0.32	0.016	1.02	3.07	210
425.14	2414.87	0.46	6	2.08	7	0.061	28	298	15	25	1.49	0.35	0.24	0.075	1.75	1.3	217
425.14	2415.15	0.47	6	1.95	7	0.055	54	420	9	18	1.07	0.37	0.14	0.042	1.08	0.85	108
424.86	2414.76	0.5	5	1.46	4	0.036	21	355	6	11	1.1	0.33	0.17	0.022	0.98	1.14	130
424.47	2414.47	0.59	6	2.6	4	0.038	29	170	6	21	1.01	0.33	0.2	0.039	1.32	2.02	192
424.82	2414.37	0.43	11	2.26	2	0.027	22	210	6	19	1.19	0.31	0.2	0.03	1.5	1.76	177
425.22	2414.46	1.05	45	4.2	3	0.065	39	125	6	25	1.94	0.39	0.54	0.046	2.17	3.04	396
424.41	2414.11	0.4	6	1.84	3	0.054	16	231	6	10	0.77	0.28	0.11	0.016	0.83	0.91	93
424.72	2413.83	0.9	7	3.87	9	0.094	135	231	48	45	2.15	0.41	0.73	0.069	1.82	2.54	327
424.35	2413.78	0.68	6	2.68	3	0.059	26	130	5	25	2.32	0.34	0.37	0.045	1.09	1.69	201
431.88	2411.24	0.81	4	3.58	15	0.053	140	143	83	34	2.76	0.36	2	0.052	2.05	8.94	241
432.90	2411.89	0.39	3	3.4	10	0.077	121	133	93	33	1.73	0.34	4.24	0.061	2.33	6.41	230
433.60	2410.63	0.42	5	2.62	1	0.042	81	152	12	21	1.55	0.31	0.3	0.054	0.77	0.94	135
433.91	2411.37	0.8	4	3.49	2	0.025	120	88	31	33	3.17	0.33	0.72	0.062	0.85	1.48	231
434.07	2410.91	0.42	6	2.99	4	0.057	109	117	12	22	2.38	0.32	0.27	0.048	0.63	2.19	210
434.73	2410.27	0.37	4	3.11	9	0.045	171	132	12	21	1.93	0.31	0.12	0.042	0.69	1.98	180
434.09	2409.69	0.36	5	2.92	6	0.044	135	123	5	21	1.76	0.29	0.09	0.046	0.71	0.6	156
432.45	2414.37	1.86	46	1.82	11	0.059	96	278	13	22	11.58	0.42	0.35	0.017	0.78	3.05	150
432.29	2414.59	3.76	65	6.04	4	0.036	41	164	67	20	9.37	0.79	0.67	0.047	2.42	7.75	486
432.60	2414.75	2.63	83	2.09	22	0.049	112	208	30	17	26.06	0.54	0.64	0.033	1.18	3.09	201

Случайная величина Z_{AUC} удовлетворяет стандартному нормальному распределению. Критическое значение получается путем сравнения Z_{AUC} со стандартной таблицей нормального распределения. Основываясь на таблице 2, выбрали Z_{AUC} больше критического значения 2,58, когда уровень значимости составлял 0,01. В общей сложности восемь элементов, включая Au, Sn, Cu, Ag, Ba, Sb, Hg и Mo, были выбраны в качестве ориентировочных элементов для прогнозирования целей разведки.

Таблица 2. **Результаты расчетов AUC и ZAUC для 16 геохимических элементов.**

Element	AUC	Z_{AUC}	Element	AUC	Z_{AUC}
Au	0.6024	2.8395	B	0.5901	2.4839
Sn	0.6065	2.9595	Cu	0.6311	3.6977
Ag	0.6762	5.1563	Ba	0.6147	3.2020
Mn	0.5573	1.5617	Pb	0.5778	2.1341
Zn	0.5450	1.2232	As	0.5655	1.7893
Sb	0.5942	2.6017	Bi	0.5901	2.4839
Hg	0.6393	3.9516	Mo	0.5983	2.7203
W	0.5778	2.1341	F	0.5696	1.9037

Объединили результаты экспериментов и геологическую среду исследуемого района для анализа. Исследуемый район Пангсидонг имеет разломы, развитые

складчатые структуры и сильную магматическую активность. Металлогеническая геологическая среда аналогична многим районам с богатыми минеральными ресурсами в металлогеническом поясе Циньчжоу–Ханчжоу. Речь идет в основном о металлических рудах Au, Ag, Pb, Zn, W, Mo, Fe и Mn. Затем в девонских пластах на исследуемой территории видны металлические руды Pb, Zn и Cu. Кроме того, северо–восточные разломы Гучен-Шачан и Пангситонг в исследуемой области являются основными рудоформирующими структурами для Au, Ag, Sn и других полезных ископаемых. Таким образом, большинство из восьми выбранных индикаторных элементов соответствуют металлическим рудам в исследуемом районе. Это показывает обоснованность выбора элементов индикаторов.

Ссылаясь на метод обработки данных об осадках потока, использовали обратную интерполяцию веса расстояния для создания сеток 3072 × 3072 (карта элементного содержания) в соответствии со значениями геохимического элементного содержания. В частности, вычислили расстояние между соседними дискретными точками и сеткой (x0,y0):

$$D_i = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2} \quad (3)$$

где Di обозначает расстояние от i-й дискретной точки вблизи (x0,y0), и (xi,yi) обозначает координаты i-й дискретной точки.

Исходя из этого, оценили значение сетки (x0,y0) следующим образом:

$$Z(x_0, y_0) = \sum_{i=1}^{N} \frac{1}{(D_i)^2} Z_i / \sum_{i=1}^{N} \frac{1}{(D_i)^2} \quad (4)$$

где Z(x0,y0) обозначает оценочное значение в сетке (x0,y0), Zi обозначает наблюдаемое значение в i-й дискретной точке, а N обозначает количество дискретных точек, участвующих в вычислении.

На рисунке 3 показаны карты элементного состава восьми элементов. Нормализовали значения карты содержания элементов, сделав ее среднее значение 0, а дисперсию 1. Для получения значения x в карте элементного содержания нормализованное значение выглядит следующим образом:

$$\hat{x} - \frac{x - \bar{x}}{\sigma} \quad (5)$$

где x^ обозначает нормализованное значение, $\bar{x}$ обозначает среднее значение значений карты элементного содержания, и σ обозначает отклонение значений карты элементного содержания.

Разделили карту элементного содержимого размером 3072×3072 на две части, где верхняя часть (2560×3072) использовалась для создания обучающего набора данных, а

нижняя часть (512×3072) использовалась для создания тестового набора данных. Определили два окна размером 256×256 для перемещения по верхней и нижней картам содержимого элементов с шагом 128. В обучающем наборе данных было 437 исправлений, из которых 78 исправлений были с рудой, а 359 исправлений были без. С другой стороны, тестовый набор данных содержал 69 исправлений, в том числе 17 исправлений с рудой и 52 исправления. Кроме того, для увеличения данных был добавлен гауссовский шум со средним значением 0 и дисперсией 0,01. В итоге сгенерированный обучающий набор данных содержал 2169 исправлений, в том числе 1092 исправления с рудой и 1077 исправлений без, а сгенерированный тестовый набор данных содержал 275 исправлений, в том числе 119 исправлений с рудой и 156 исправлений без. Объединили карты элементного состава восьми элементов, то есть экспериментальный набор данных состоял из 2444 участков размером 256×256×8. Кроме того, случайным образом обрезали данные до размера 224×224 перед каждой итерацией в качестве входных данных модели.

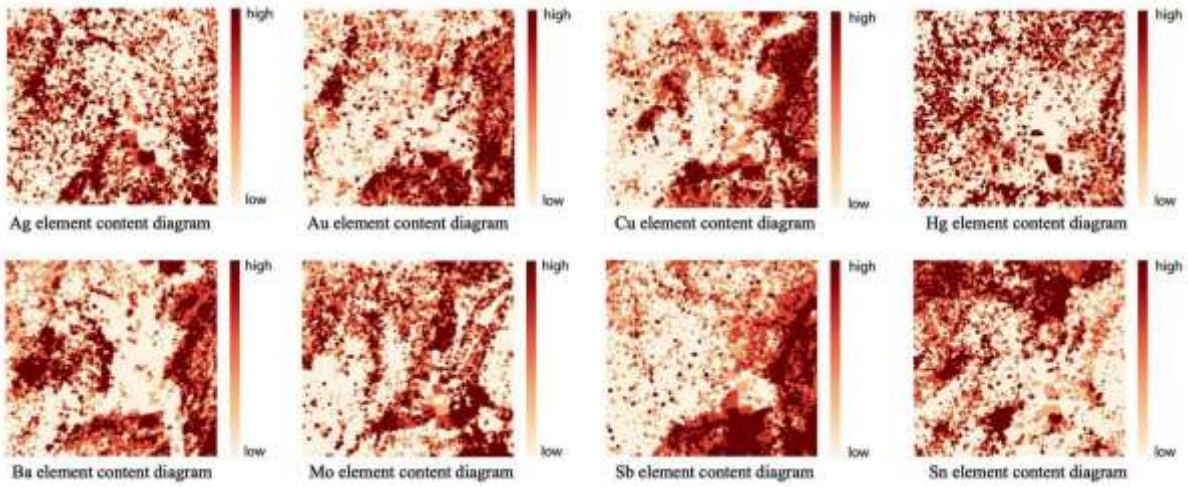

Рисунок 3. Диаграммы элементного содержания.

3. Методология

Предлагается метод выборочной передачи знаний для прогнозирования целей разведки. В частности, в разделе 3.1 формально рассматривается этот вопрос. В разделе 3.2 объясняется, как использовать мягкую маску, чтобы поддерживать соответствие между соответствующими минеральными элементами и основными минеральными элементами, и использовать расширенную свертку для получения карт объектов разного масштаба. В разделе 3.3 обсуждается выборочная передача знаний. В разделе 3.4 рассматривается саморазведка для извлечения скрытых знаний о целевых сетях с различными многомасштабными функциями. Раздел 3.5 экспортирует целевую функцию SKT, используемую для развертывания этого механизма.

3.1. Постановка задачи

Преобразуем геохимические данные в карту содержания геохимических элементов с помощью обратной интерполяции веса и расстояния. Затем определяем скользящее окно размером 256×256. В то же время набор геохимических данных $D = \{x_h^i, y_h^i\}_{h=1,i=1,2,\ldots,n}^{N}$ создается путем вырезания карты элементного содержимого с шагом размером 128. Здесь N обозначает количество образцов, i обозначает класс геохимического элемента, и $x_h \in \mathbb{R}^d$ and $y_h \in \{0,1\}$ обозначьте вектор признаков и соответствующую метку, где 0 и 1 означают “без руды” и “с рудой” соответственно, а d обозначает размер пространства объектов выборки. Предполагаем, что структура SKT состоит из хорошо обученной исходной сети задач NS и сети с несколькими целевыми задачами $N_T=\{N_{t1},N_{t2},\ldots,N_{tm}\}$. Эти сети имеют одинаковую структуру и L уровней свертки, где числа входных и выходных каналов в l-м уровне свертки равны M_l и M_{l+1}, соответственно. Ядро свертки на l-м уровне в исходной сети NS определяется как $W^l_S=[w^l_{S1},w^l_{S2},\ldots,w^l_{SMl+1}]$, где $W^l_S \in R^{Ml+1\times Ml\times K\times K}$, $w^l_{Sj} \in R^{Ml\times K\times K}$, и $^{K\times K}$ является размером ядра свертки. Для всех ядер свертки на l-м уровне NT, он определяется как $W^l_T=[W^l_{t1},W^l_{t2},\ldots,W^l_{te}] \in Re^{\times Ml+1\times Ml\times K\times K}$, где ядра свертки сети задач tv являются $W^l_{tv}=\{w^l_{tv1},w^l_{tv2},\ldots,w^l_{tv}{}^{Ml+1}\}$, и настройка ядра свертки согласуется с настройкой в NS. Предлагается использовать выборочные знания о передаче для прогнозирования целей поиска.

3.2. Соответствие родственных минеральных элементов.

Считаем, что концентрация геохимических элементов имеет ориентировочное значение для минерализации, и разные элементы имеют различия в прогнозировании минерализации. Как правило, полезные ископаемые содержат сопутствующие элементы. Их основные минеральные элементы имеют важное ориентировочное значение для минерализации, в то время как сопутствующие минеральные элементы также имеют определенное ориентировочное значение. Поэтому стараемся вводить мягкую маску M=[M1,M2,…,Mn] - стратегия, основанная на вышеизложенном соображении. Его цель состоит в том, чтобы сделать соответствующий вес сопутствующих минеральных элементов как можно более согласованным с весом основных минеральных элементов и увеличить разнообразие эффективных образцов следующим образом:

$$\hat{x}_h^i = x_h^i \times \mathcal{M}_i \quad (6)$$

где x^i_h обозначает h-й образец в i-м геохимическом элементе, и M_i обозначает вес, соответствующий i-му геологическому элементу в M, который является скалярным; x^i_h

обозначает h-ю пробу в i-м геохимическом элементе с помощью операции маски. Таким образом, получаем карты характеристик различных геохимических элементов.

Чтобы решить проблему, связанную с нерегулярными особенностями горных районов, выполняем операцию расширенной свертки для характеристик различных геохимических элементов для создания карт объектов разного масштаба. В частности, определяем набор расширенных коэффициентов $\rho=\{\rho_1,\rho_2,\ldots,\rho_m\}$, где каждый коэффициент соответствует карте объектов и целевой сети в разных масштабах. Затем выполняем операцию расширенной свертки на карте объектов v-го масштаба в качестве входных данных целевой сети N_{tv}, где значение позиции (p_{row},p_{col}) на карте объектов вычисляется по

$$\mathbf{f}_{t_v}^{i} = \sum_{u=-r}^{r} \sum_{g=-r}^{r} \hat{x}_{t_v}^{i}(\rho_{t_v} u + p_{\mathrm{row}}, \rho_{t_v} g + p_{\mathrm{col}}) \times o(u,g) \tag{7}$$

где f^{i}_{tv} обозначает t_v- t_h характеристик i-го геохимического элемента в масштабе выборки с помощью расширенной свертки, ρ_{tv} обозначает соответствующий расширенный коэффициент , равный N_{tv}, $x^{\wedge i}_{h}(\cdot,\cdot)$ обозначает вес h-го положения образца i-го геохимического элемента, и $o(\cdot,\cdot)$ обозначает вес позиции в расширенном ядре свертки. u,g обозначает ближайшие соседние единицы измерения позиции (prow,pcol), и r=S−1, где S - размер расширенного ядра свертки. Например, для расширенного ядра свертки размером 3 ближайший сосед u и g равен 1.

3.3. Выборочная передача знаний.

Из-за небольшого количества выборок это легко приводит к таким проблемам, как недостаточное соответствие или несоответствие в процессе обучения модели прогнозирования цели поиска. Таким образом, выборочно переносим элементы в ядрах свертки из хорошо обученной исходной сети задач N_S чтобы помочь целевому обучению сети, как показано на рисунке 4. В частности, сначала определяем матрицу P^{l}_{tvj} с тем же размером, что и ядро свертки в целевой сети Ntv. Затем вычисляем произведение Адамара на W^{l}_{S} следующим образом:

$$\hat{w}_{t_v j}^{l} = w_{Sj}^{l} \odot \mathrm{P}_{t_v j}^{l} \tag{8}$$

где $w^{\wedge l}_{Sj}$ обозначает выбранные ядра свертки, а $\odot$ обозначает произведение Адамара.

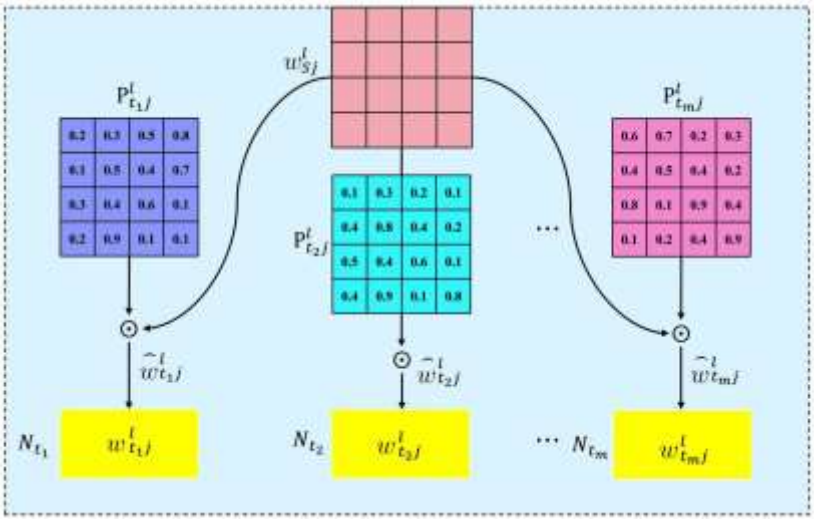

Рисунок 4. Описание алгоритма выбора веса. wlSj обозначает ядра свертки l-й сверточный слой исходной сети, j = 1, 2, . . . , M^{l+1} – выходные каналы, $P^{l}_{t\,v\,j}$ – матрица, соответствующая целевой сети $N_{t\,v,}$ $_{v}$ = 1, 2, . . . , m обозначает произведение Адамара, $\hat{}w^{l}_{t\,v\,j}$ обозначает выбранные ядра свертки, а $w^{l}_{t\,v\,j}$ обозначает ядра свертки N_{tv}

Чтобы еще больше помочь в обучении сети целевых задач, передаем w^lSj Для NT для обучения с помощью

$$f_{t_v} = \left(\hat{w}^l_{Sj} \times w^l_{t_v j}\right) * \mathbf{f}^i_{t_v} + b \quad (9)$$

где f_{tv} обозначает выходные данные операции свертки, ∗ обозначает операцию свертки, b обозначает смещение свертки и f^{i}_{tv} обозначает tv-карта характеристик i-го геохимического элемента в масштабе выборки.

3.4. Самоотгонка.

Чтобы добыть скрытые знания между картами объектов разного масштаба, выполняем перегонку знаний сверху вниз в соответствии с целевой сетью, соответствующей размеру карт объектов. Например, существует три целевые сети, N_{t1}, N_{t2}, и N_{t3}, и размер их входных карт объектов последовательно уменьшается. Используем N_{t1} для руководства N_{t2} и N_{t3}, и N_{t2} для руководства N_{t3}. В частности, используем дивергенцию Кульбака–Лейблера (KL) для вычисления распределения вероятностей выходных данных softmax между каждой парой целевых сетей путем

$$\mathcal{L}_{t_v KD} = \sum_{j=1}^{v-1} \sum_{i=1}^{n} f\left(\mathbf{f}^i_{t_j}, \theta^{N_{t_j}}\right) \log\left(\frac{f\left(\mathbf{f}^i_{t_j}, \theta^{N_{t_j}}\right)}{f\left(\mathbf{f}^i_{t_v}, \theta^{N_{t_v}}\right)}\right) \quad (10)$$

где L_{tvKD} обозначает потери при самоотводе целевой сети N_{tv}; размер входных карт объектов N_{tj}больше, чем у N_{tv}; θN_{tj} и θN_{tv} обозначают параметры целевых сетей N_{tj} и N_{tv}, соответственно; f(·)обозначает операцию softmax; и f_{itj} и f_{itv} обозначим tj-й и tv-й масштаб выборки показывает карты i-го геохимического элемента с помощью операции расширенной свертки.

3.5. Целевая функция.

Целевая сеть сопоставляет объекты с соответствующим пространством меток через слой с полным подключением. Во время обучения разработали классификационные потери для каждой целевой сети следующим образом:

$$\mathcal{L}_{t_v} = \sum_{i=1}^{n} \mathrm{L}\left(\theta^{N_{t_v}}, \mathbf{f}_{t_v}^{i}, \mathrm{P}_{t_v}^{1:L}, \mathrm{y}_{\mathrm{h}}^{i}\right) \quad (11)$$

где L_{tv} обозначает потерю классификации целевой сети N_{tv}, $L(\cdot)$ обозначает перекрестную энтропию, θN_{tv} обозначает параметр N_{tv}, y^i_h является меткой для ввода x^i_h, f^i_{tv} обозначает tv- карта признаков масштаба выборки i-го геологического элемента, полученная с помощью операции маски и операции расширенной свертки, и $P^1{:}L_{tv}$ обозначает матрицу, связанную с выборочной передачей знаний.

В конечном счете, общая цель оптимизации SKT заключается в минимизации потерь при классификации и самоотгонке путем

$$\mathcal{L}_{total} = \sum_{v=1}^{m} \mathcal{L}_{t_v} + \beta \sum_{v=1}^{m} \mathcal{L}_{t_v KD} \quad (12)$$

где L_{total} обозначает целевую функцию, и β обозначает коэффициент потерь при самоотгонке.

4. Эксперименты

В этом разделе - эксперименты для проверки метода. В частности, в разделе 4.1 описаны экспериментальная среда и настройки. В разделе 4.2 представлено сравнение метода с современными методами на основе геохимических данных и анализ экспериментальных результатов. В разделе 4.3 представлены экспериментальные результаты, основанные на соответствующих модулях и параметрах метода. Раздел 4.4 визуализирует результат прогнозирования для исследуемого района Пангсидун в провинции Гуандун.

4.1. Экспериментальные настройки

Оценочные показатели, использованные в экспериментах, включают точность, отзыв и оценку F1. Все эксперименты запрограммированы и реализованы с помощью платформы PyTorch и одного графического процессора GeForce RTX 3090.

Платформа SKT реализована на основе архитектуры ResNet-18. Во время обучения модель использует оптимизатор SGD с импульсом 0,01, весовой спад 1×10^{-4}, и мини-пакет размером 128. Скорость обучения изначально равна 0,1 и уменьшается наполовину каждые 30 эпох. Устанавливаем расширенные коэффициенты $\rho=\{\rho_1,\rho_2,\rho_3,\rho_4,\rho_5,\}=\{1,6,12,18,24\}$, ссылаясь на DeepLab. Модель

обучается для 150 эпох на основе геохимического набора данных. Кроме того, коэффициент β потери при самоотгонке устанавливаются равными 0,7.

4.2. Экспериментальные результаты и анализ.

В этом разделе сравниваем алгоритмы машинного обучения и некоторые современные методы классификации с SKT, чтобы продемонстрировать, что он превосходит другие модели в задачах прогнозирования поиска целей. В частности, сравниваем следующие методы, включая традиционные методы SVM, KNN, RandomForest и Decisiontree, а также методы глубокого обучения ResNet-18, ShufflenetV2, GoogLeNet, MobileNetV2, Mnasnet, SCNet, Efficientnet-b0, T2T-vit-14и SNL. SCNet и SNL реализованы на основе архитектуры ResNet-18. В алгоритме машинного обучения сжимаем каждую точку данных в геохимическом наборе в одномерный тензор в качестве входных данных для алгоритма.

Как показано в таблице 3, модель работает лучше, чем ResNet-18. В частности, точность модели увеличивается на 12,30%, отзыв увеличивается на 15,83%, а оценка F1 улучшается на 11,79%. Кроме того, SKT превосходит другие методы с точки зрения точности, отзыва и оценки F1. Это указывает на то, что SKT обладает отличными показателями в прогнозировании целей разведки и имеет наибольшее улучшение в прогнозировании оруденения. Это также доказывает, что метод может эффективно решать такие проблемы, как небольшое количество геологических данных и нерегулярные особенности горных районов при прогнозировании целей разведки.

Таблица 3. **Экспериментальные результаты.**
Оптимальные характеристики выделены жирным шрифтом.

Methods	Accuracy	Recall	F1-Score
SVM	49.51	17.64	43.73
KNN	51.45	35.29	50.09
RandomForest	59.70	25.49	54.27
Decisiontree	58.73	39.21	57.03
ResNet-18	56.79	24.50	53.21
ShufflenetV2	57.45	17.64	48.24
GoogLeNet	61.81	31.09	56.51
MobilenetV2	55.82	16.66	47.74
Mnasnet	59.22	17.64	50.61
SCnet	58.73	30.39	55.05
Efficientnet-b0	57.28	23.52	51.70
T2T-vit-14	57.76	39.21	56.19
SNL	59.70	35.29	57.07
Ours	69.09	40.33	65.00

Рисунок 5 представляет собой матрицу путаницы, показывающую количество истинно отрицательных (TN), ложноположительных (FP), ложноотрицательных (FN) и истинно положительных (TP) выборок. Этот рисунок показывает, что:

(1) среди этих четырех значений TN является самым высоким, т.е. число правильно предсказанных выборок без руды является наибольшим. В то же время FP является

самым низким, т.е. количество неправильно предсказанных выборок без руды является наименьшим. Перед использованием гауссовского шума для увеличения данных выборки без руды намного больше, чем рудные выборки, а выборки без руды обладают богатыми возможностями, которые полезны для STS framework для прогнозирования.

(2) TP обозначает количество правильно предсказанных рудных выборок. Напротив, FN обозначает количество неправильно предсказанных рудных выборок. TP - это третье из четырех значений, ниже, чем FN, но намного выше, чем FP. Причиной такого результата является небольшое количество рудных выборок и их нерегулярные характеристики, что негативно влияет на структуру STS для прогнозирования, но все же имеет определенную производительность. SKT может хорошо прогнозировать безрудные образцы, а также обладает определенной способностью прогнозировать рудные.

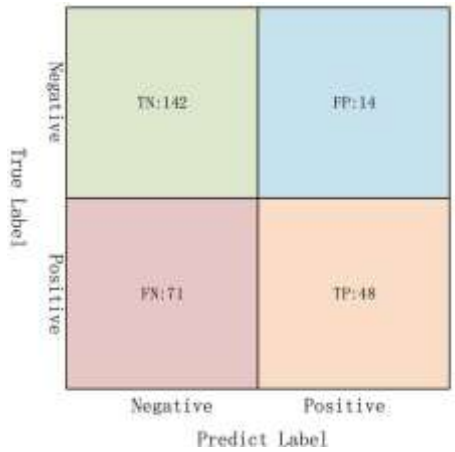

Рисунок 5. Матрица путаницы.

В рамках SKT окончательный результат прогнозирования получается путем голосования по нескольким целевым сетям. Таблица 4 показывает прогнозируемые результаты для нескольких целевых сетей и голосования.

Таблица 4. **Экспериментальные результаты целевых сетей и голосования.**

Target Network	Accuracy	Recall	F1-Score
$\rho = 1$	57.45	31.93	53.30
$\rho = 6$	65.45	41.17	62.08
$\rho = 12$	61.45	35.29	57.38
$\rho = 18$	70.18	47.05	67.34
$\rho = 24$	64.72	37.81	60.70
Voting	69.09	40.33	65.00

Из этой таблицы видно следующее:

(1) По мере увеличения коэффициента расширения производительность целевых сетей сначала увеличивается, а затем уменьшается.

(2) По сравнению с лучшей целевой сетью голосование имеет лишь небольшой разрыв в точности и результатах F1, но есть определенный разрыв в отзывах.

Приведенный выше результат вызван тем, что увеличение коэффициента расширения влияет на производительность модели, но в то же время самоотвод может улучшить производительность целевой сети, приводя сначала к увеличению, а затем к снижению. Кроме того, из-за противоречия между коэффициентом расширения и самоотдачей для улучшения производительности модели мы не можем определить оптимальную целевую сеть. Однако производительность голосования аналогична производительности оптимальной целевой сети, поэтому конечный результат прогнозирования, полученный в результате голосования, в определенной степени является разумным.

4.3. Эксперимент с корреляционным анализом.

Представлены экспериментальные результаты, основанные на соответствующих модулях SKT, в разделе 4.3.1. Затем, в разделе 4.3.2, проверяем влияние коэффициента потерь при самоотгонке на SKT.

4.3.1. Эксперименты по абляции.

Чтобы оценить эффективность метода, проводим эксперименты по абляции SKT и используем различные варианты SKT. В частности, разрабатываем следующие эксперименты по абляции:

(1) удаляем мягкую маску (R-S-Mask),

(2) удаляем расширенную свертку (R-D-Convolution),

(3) удаляем выборочную передачу знаний (R-Sk-Transfer) и

(4) удаляем самоотгонку (R-S-Дистилляция).

Основываясь на SKT, применяем подход с управляющими переменными к мягкой маске, расширенной свертке и выборочной передаче знаний по очереди. Устанавливаем расширенные коэффициенты $\rho=\{\rho_1,\rho_2,\rho_3,\rho_4,\rho_5,\}=\{1,1,1,1,1\}$ во втором эксперименте. Наконец, SKT сравнивается с вышеуказанными методами.

Таблица 5. **Результаты эксперимента по абляции.**
Оптимальные характеристики выделены жирным шрифтом.

Methods	Accuracy	Recall	F1-Score
R-S-Mask	64.72	29.41	55.43
R-D-Convolution	61.09	29.41	58.29
R-Sk-Transfer	62.54	30.25	56.83
R-S-Distillation	65.81	31.09	59.72
Ours	69.09	40.33	65.00

В таблице 5 представлены результаты эксперимента:

1. Мягкая маска обеспечивает максимальное соответствие соответствующего веса сопутствующих минеральных элементов весу основных минеральных элементов. Расширенная свертка имеет дело с нерегулярными особенностями горных районов с помощью различных восприимчивых полей. Выборочная передача знаний повышает производительность обобщения модели для решения проблемы небольшого числа выборок. Самоочистка добывает скрытые знания между картами объектов разного масштаба. Все вышеупомянутое может улучшить точность, отзыв и оценку F1 прогноза.

2. Вклад этих приемов в SKT различен. В соответствии с вкладом от большого к малому они ранжируются следующим образом: расширенная свертка, выборочная передача знаний, мягкая маска и самоотгонка.

4.3.2. Эксперименты по анализу параметров

Целевая функция SKT включает в себя потери при самоотгонке. Для оценки влияния коэффициента потерь при самоотгонке β на SKT устанавливаем β=0.1,0.2,…,1 при проведении 10 экспериментов. На рис. 6 показаны результаты эксперимента.

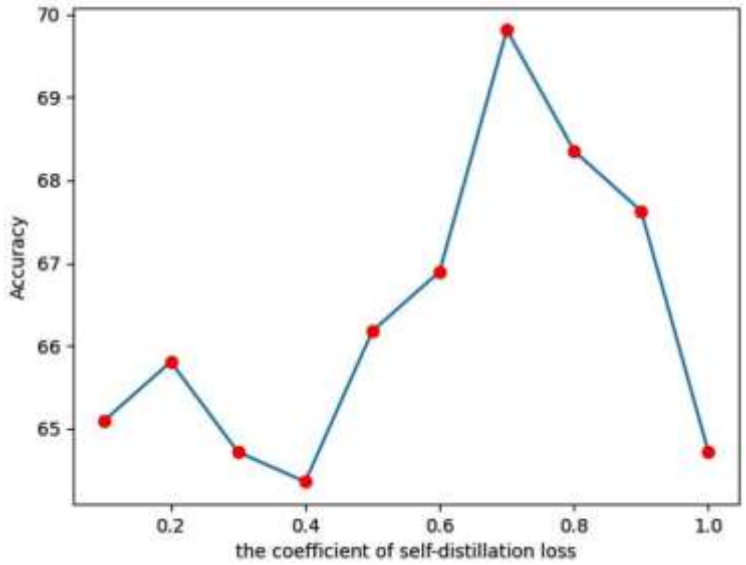

Рис. 6. Коэффициент потерь при самоперегонке. Результаты эксперимента.

Из этого рисунка можем найти следующее:

(1) β оказывает определенное влияние на производительность SKT.

(2) Наилучшие результаты прогнозирования целей разведки могут быть получены, когда β составляет от 0,6 до 0,8.

4.4. Визуализация.

В этом разделе используем SKT, обученный в разделе 4.2, для прогнозирования целевого района поиска в районе исследования Пангсидонг и визуализации

результатов прогнозирования. В частности, сначала разрезаем карту содержания элементов размером 3072×3072, полученную в разделе 2.2, на участки размером 12×12, каждый размером 256×256. Затем используем обученный SKT для прогнозирования и визуализации. Рисунок 7 показывает результат визуализации.

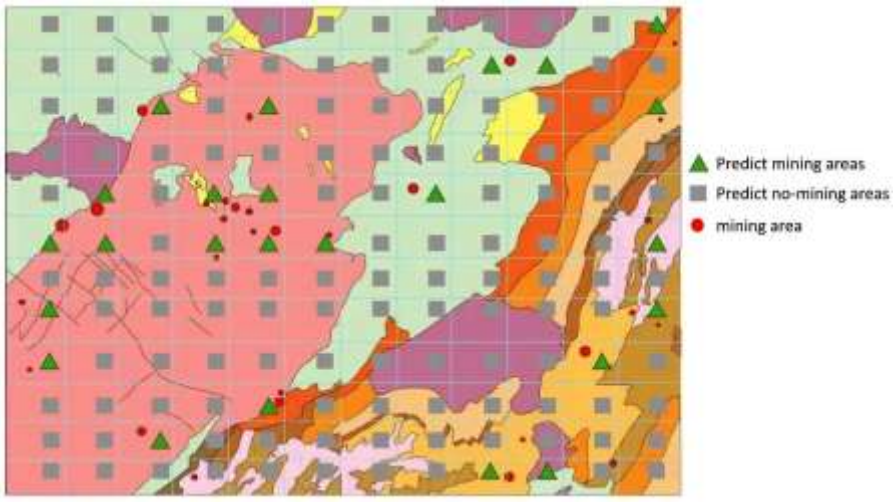

Рисунок 7. Визуализация результатов экспериментов в районе исследования Панксидонг.

Основываясь на визуализации, приходим к следующим выводам:

(1) Прогнозируемый целевой район разведки в основном соответствует фактическому району добычи.

(2) Результаты прогнозирования (строка 7, столбец 4), (строка 11, столбец 10), (строка 12, столбец 4) и (строка 12, столбец 11) несовместимы с фактической площадью добычи.

Кроме того, используем метод анализа главных компонентов (PCA), чтобы уменьшить размерность геохимических данных в эксперименте до одного измерения и визуализировать его (рис. 8).

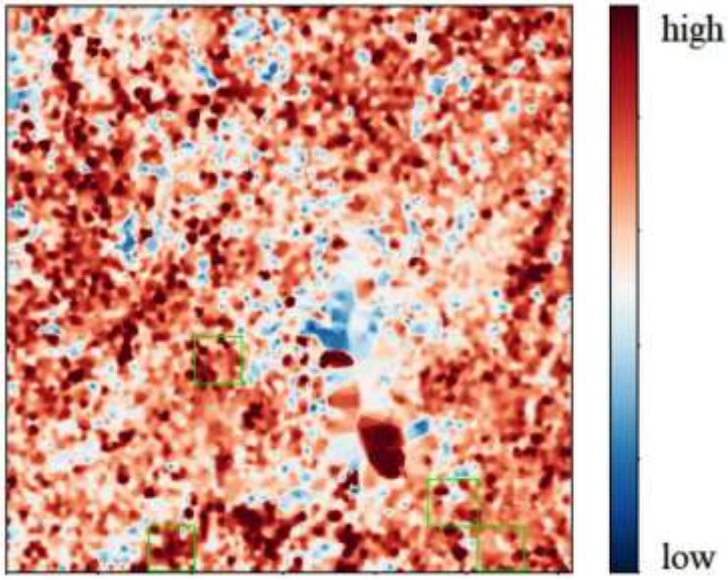

Рисунок 8. Визуализация геохимических данных.

Как видно из рисунка 8:

(1) большинство районов добычи полезных ископаемых соответствуют высоким значениям, т.е. богаты этими восемью геохимическими элементами. Это доказывает, что геохимические элементы оказывают важное влияние на формирование оруденения.

(2) Нерегулярные особенности геохимически обогащенных районов (зеленые кадраты) затрудняют прогнозирование STS.

(3) SKT может соответствовать распределению целевой области поиска в районе исследования Пангсидонг.

5. Выводы

Платформа глубокого обучения (SKT), основанная на выборочной передаче знаний, решения проблем при малом количестве геологических данных и нерегулярных особенностях рудных районов, предлагается для интеллектуального прогнозирования поисковых работ. Взяв в качестве примера исследуемый район Пангсидун в провинции Гуандун, можно разумно спрогнозировать целевой район поисков, используя геохимические данные этого района. По сравнению с другими методами доказана эффективность этого метода.

Основными выводами являются:

(1) Для прогнозирования поисков при небольшом количестве геологических данных и нерегулярных особенностях рудных районов, рекомендуется платформа глубокого обучения (SKT) на основе выборочной передачи знаний Она превосходит другие методы и улучшает эффективность прогнозирования оруденения.

(2) *Мягкая маска* максимально приближает соответствующий вес сопутствующих минеральных элементов к весу основных минеральных элементов; *свертка с расширением* обогащает нерегулярные особенности рудных районов за счет захвата объектов в разных масштабах; *выборочная передача знаний* улучшает производительность обобщения модели и решает проблему небольшого количества даннах; а *самоочистка* добывает скрытые знания между картами объектов разного масштаба.

(3) Эксперименты по анализу параметров показывают, что свертка с расширением, выборочная передача знаний, мягкая маска и самоотгонка могут повысить точность прогнозирования SKT.

ИСТОЧНИКИ:

1. *Четан Л. Натвани, Джейми Джей Уилкинсон, Джордж Фрай, Робин Н . Армстронг, Дэниел Дж. Смит. Mineralium Deposita 2022.*
2. *Дао Сун Ин Сюй, Сюхуэй Юй, Вэймин Лю, Жуйсюэ Ли, Цзыцзюань Ху и Юнь Ван. Minerals 2018.*
3. *Юнцзе Хуан, Цюань Фэн, Ли Чжан, Ле Гао Полезные ископаемые 2022.*