

9. Malberet J. // L'Astronomie. – 1925. – V. 39. – P. 503.
10. Clayton H.H. // Smith. Misc. Coll. – 1947. – V.106. – Publ. 3870. – P. 1.
11. Condon J.J., Schmidt R.R. Planetary tides and sunspot cycles // Solar Physics. – 1975. – V. 42, № 2. – P. 529-532.
12. Stefani F., Stepanov R., Weier T. Shaken and Stirred: When Bond Meets Suess–de Vries and Gnevyshev–Ohl. // Solar Physics. – 2021. – V. 296, № 6. – Номер статьи 88. – DOI: 10.1007/s11207-021-01822-4.
13. Shlien S. Earthquake – tide correlation // Geophysical Journal of the Royal Astronomical Society. – 1972. – V. 28, № 1. – P. 27-34.
14. Авсюк Ю.Н. Приливные силы и природные процессы. – М.: ОИФЗ РАН, 1996. – С. 142-143.
15. Некрасова Л.В. Сейсмическая активность Западно-Уральского региона и сезонные вариации скорости вращения Земли // Горное эхо. – 2020. – № 4 (81). – С. 63-70. – DOI: 10.7242/echo.2020.4.13.
16. Некрасова Л.В. Ротационный режим Земли и сезонные особенности техногенных сейсмических процессов Уральского региона // Горное эхо. – 2019. – № 2 (75). – С. 71-74. – DOI: 10.7242/echo.2019.2.17.

УДК 550.8.052

DOI:10.7242/echo.2021.3.10

## МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ПРИ КЛАССИФИКАЦИИ ГЕОЛОГО-ГЕОФИЗИЧЕСКИХ ДАННЫХ

П.Н. Новикова

*Горный институт УрО РАН, г. Пермь*

**Аннотация:** В статье обсуждаются основные моменты применения методов интеллектуального анализа в области геологоразведки. Показаны основные этапы обучения моделей прогнозирования. Приводится пример применения методов машинного анализа для задачи классификации динамической физико-геологической модели: дерево решений, случайный лес и градиентный бустинг. Показаны результаты автоматической классификации данных.

**Ключевые слова:** машинное обучение, прогнозирование, классификация, деревья решений, физико-геологическая модель.

Методы интеллектуального анализа, связанные с поисками оптимальных решений, в явном и неявном виде давно используются на всех этапах геологоразведочного процесса: от проектирования работ до собственно интерпретации данных. При этом всегда стояла задача обработки существенных массивов данных, особенно при комплексировании различных поисковых методов.

На данный момент сформировано целое направление – машинное обучение, позволяющее автоматически выявлять многомерные закономерности в больших объемах данных, с помощью которых в дальнейшем возможно анализировать существующую информацию, принимать решения, а также выполнять прогнозы по новым наборам данных на основе предложенных моделей.

Современные модели прогнозирования хорошо адаптированы под информацию, которая может включать в себя не только числовые переменные, но и категориальные и бинарные переменные. Также машинное обучение успешно работает с графическими представлениями данных. Такая особенность является большим преимуществом при работе с разнородной геолого-геофизической информацией.

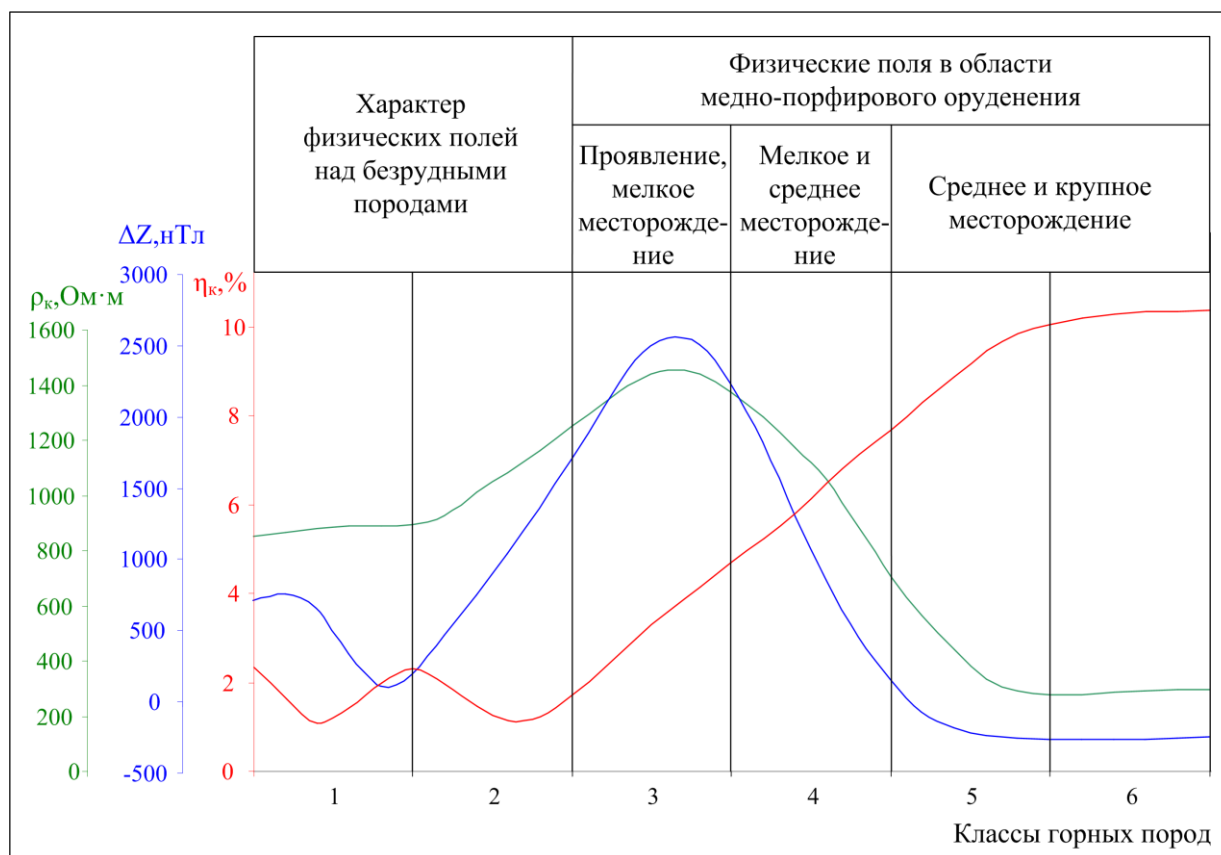
Ниже приведем процесс машинного обучения, который обычно состоит из следующих этапов:

- 1) сбор и подготовка данных, их проверка, обнаружение нехарактерных значений, поиск скрытых закономерностей между признаками, а также снижение размерности исходного набора данных;

- 2) определение типа модели и ее характеристик, которое обычно проводится экспериментальным путем;
- 3) обучение модели, т.е. поиск закономерностей в данных;
- 4) прогнозирование интересующих параметров;
- 5) оценка точности модели;
- 6) интерпретация результатов.

В данной статье приведен пример использования методов интеллектуального анализа для классификации модельных геолого-геофизических данных. В качестве модели обучения была выбрана динамическая физико-геологическая модель (ФГМ) медно-порфирового месторождения [1], демонстрирующая вариативность физических полей (электрического сопротивления  $\rho_k$ , модуля вертикальной составляющей магнитного поля  $\Delta Z$  и поляризуемости  $\eta_k$ ), а также наличия минералов-индикаторов на разных стадиях рудообразования в зависимости от масштаба проявления медно-молибденового оруденения (рис. 1, табл. 1, 2). В этой модели были выделены 6 стадий рудообразования, соответствующие 6 классам данных.

Исходная информация представлялась в виде таблицы числовых и категориальных переменных. В качестве входных данных для моделирования были выбраны статистические характеристики физических полей (максимальное, минимальное, среднее значения), категоризована морфология поля (табл. 1), включены условные значения содержания характерных минералов (табл. 2). Стоит отметить, что даже в упрощенном виде выделенные классы горных пород перекрываются по значениям, неоднозначны по морфологии поля и содержанию минералов.



**Рис. 1.** Динамическая физико-геологическая модель медно-порфирового месторождения, отражающая характер физических полей в зависимости от стадии рудообразования (по Никитину А.А., Хмелевскому В.К. [1])

Таблица 1

Пример табличного представления поля электрического сопротивления  $\rho_k$ 

Класс пород	Описание класса пород	Минимальное значение	Максимальное значение	Среднее значение	Вид аномалии
1	безрудные породы 1, предрудная стадия	853	893	873	спокойное поле
2	безрудные породы 2, предрудная стадия	894	1253	1073.5	градиентная зона
3	начальная рудная стадия	1254	1454	1354	положительная аномалия
4	средняя рудная стадия	705	1372	1038.5	градиентная зона
5	поздняя рудная стадия	275	704	489.5	градиентная зона
6	пострудная стадия	276	298	287	спокойное поле

Таблица 2

Табличное описание масштаба проявления медно-порфирового оруденения: Количественный параметр присутствия минералов в горных породах задан в условных единицах

Класс пород	Магнетит	Хлорит	Кварц	Молибденит	Халькопирит
1	2	2	3	0	0
2	0	0	0	0	0
3	4	0	4	1	1
4	1	0	3	1	2
5	0	0	2	1	4
6	0	0	1	1	1

Класс пород	Пирит	Серицит	Гематит-мартит	Сфалерит	Галенит
1	0	0	0	0	0
2	1	0	0	0	0
3	2	0	0	0	0
4	2	0	0	0	0
5	4	3	4	1	1
6	4	4	0	1	1

На основе представленной ФГМ была создана обучающая выборка, отражающая основные характеристики каждого класса пород, при этом в эти данные была внесена неоднозначность классов по всем признакам. Обучающая выборка была разделена на два набора данных – собственно обучающие и тестовые, по которым проводилась оценка точности моделей прогнозирования.

Для классификации представленных данных был создан алгоритм на языке программирования Python с использованием библиотеки scikit-learn. В качестве моделей обучения были выбраны методы обучения с учителем: дерево решений, случайный лес и градиентный бустинг, хорошо зарекомендовавшие себя при решении задач похожего типа [2, 3].

Обучение с учителем предполагает существенный набор данных для обучения, для которых уже известен искомый параметр. В рамках представленной задачи смоделирован ограниченный объем данных, тем не менее, достаточный для минимального обучения модели. Алгоритм дерева решений классифицирует объекты в соответствии с решающими правилами об их атрибутах, расположенных в узловых точках. В зависимости от ответа выбирается одна из ветвей, и так до тех пор, пока не будет достигнут «лист» — окончательный ответ. Регулировать такую модель можно, меняя глубину дерева, т.е. числа разбиений в ветвях. Методы случайного леса и градиентного бустинга являются ансамбльными. Случайный лес представляет собой множество решающих деревьев, в задаче классификации принимается решение голосованием по большинству. Градиентный бустинг также использует ансамбль из дерева решений, но в отличие от случайного леса, где деревья строятся независимо, бустинг использует последовательное улучшение моделей и обучается на ошибках предыдущих построений [3].

Эффективность моделей оценивалась через среднеквадратическую ошибку на тестовом наборе данных с известной классификацией. Параметры моделей прогнозирования, такие как глубина дерева, количество «листьев» дерева, количество деревьев в ансамбле, выбирались экспериментально.

При прогнозировании классов различных рудных стадий на небольшом наборе данных с использованием трех алгоритмов машинного обучения получились следующие результаты (табл. 3). В данной таблице приведены наиболее удачные реализации. Если рассмотреть вычислительный эксперимент в общем, то наиболее устойчивым по выделению оказались 4, 5 классы, что напрямую зависит от исходных данных, т.к. в этих классах меньше всего заложено неоднозначности, также это можно связать и с природой физических полей и процессов, происходящих на средней и поздней рудной стадии. Среднеквадратическая ошибка прогнозирования по наилучшим моделям составила: 0.29 – по методу дерева решений; 0.29 – по методу случайного леса; 0.37 – по методу градиентного бустинга.

Таблица 3

Результаты классификации модельных геолого-геофизических данных методами интеллектуального анализа

Фактические классы	Прогнозируемые классы		
	Дерево решений	Случайный лес	Градиентный бустинг
1	1	2	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	4	5	6

Применение методов интеллектуального анализа, в том числе машинного обучения, существенно упростит работу специалистов в геологоразведочной области, улучшит анализ больших данных, позволит быстро ориентироваться по разнообразным материалам и прогнозировать необходимые параметры.

*Работа выполнена при поддержке гранта РФФИ № 19-05-00654.*

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-05-00654.*

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Никитин А.А., Хмелевской В.К. Комплексование геофизических методов: учебник для вузов. – Тверь: ГЕРС, 2004. – 294 с.
2. Рашка С., Мирджалили В. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2 / пер. с англ. и ред. Ю.Н. Артеменко. – 3-е изд. – СПб.: Диалектика, 2020. – 846 с.: ил., табл.
3. Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. – СПб.: Питер, 2017. – 336 с.

УДК 550.83.017+550.837

DOI:10.7242/echo.2021.3.11

## РЕШЕНИЕ ОБРАТНОЙ ЗАДАЧИ ВЭЗ МЕТОДОМ ДИФФЕРЕНЦИАЛЬНОЙ ЭВОЛЮЦИИ

А.А. Тайницкий

*Горный институт УрО РАН, г. Пермь*

**Аннотация.** Решение обратной задачи ВЭЗ связано с поиском глобального минимума функции невязки, наиболее часто выраженной среднеквадратическим расхождением экспериментальных и теоретических значений кажущихся электрических сопротивлений кривых зондирования. Функция невязки сложным образом зависит от большого количества параметров, поэтому одной из проблем поиска минимума является ее сложный многоэкстремальный «овражный» характер, что может приводить к попаданию в локальный минимум. В рамках данной статьи рассмотрена возможность решения одномерной обратной задачи ВЭЗ стохастическим методом глобальной оптимизации широко известный как метод дифференциальной эволюции.

**Ключевые слова:** электроразведка, вертикальное электрическое зондирование, обратная задача, стохастические методы, метод дифференциальной эволюции.

### Введение

Решение обратной задачи ВЭЗ связано с определением параметров геоэлектрического разреза по существующему распределению значений физического поля, в данном случае кажущегося сопротивления. Для решения данной задачи наибольшее распространение получили методы подбора, основанные на сравнении наблюдаемой и теоретической кривой зондирования, рассчитанной для заданных значений параметров разреза [1]. В качестве минимизируемой (целевой) функции в большинстве случаев выступает среднеквадратическое расхождение экспериментальных  $\rho_k^{\text{э}}$  и теоретических  $\rho_k^T$  значений кажущихся электрических сопротивлений кривых зондирования на всех  $M$  разносах линии  $AB$ . Поиск минимума целевой функции выполняется с помощью алгоритмов многомерной оптимизации, среди которых наиболее популярны различные модификации метода наискорейшего спуска и метод Ньютона [2]. Сущность методов заключается в последовательном уточнении начального приближения к вектору параметров  $w = \{ \rho_1, \dots, \rho_N, h_1, \dots, h_N \}$  геоэлектрического разреза, обеспечивающего минимум целевой функции  $f$  при движении в направлении, определяемом антиградиентом функции  $f$ :

$$f = \frac{1}{M} \sum_{j=1}^M \left( 1 - \frac{\rho_{kj}^T}{\rho_{kj}^{\text{э}}} \right)^2, \quad (1)$$