

## АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ ФАЦИАЛЬНЫХ КОМПЛЕКСОВ ПО ДАННЫМ ГЕОФИЗИЧЕСКИХ ИССЛЕДОВАНИЙ СКВАЖИН

**Ковтун Я.Д.**

**Научный руководитель – кандидат педагогических наук, доцент Прозорова Г.В.**

Тюменский индустриальный университет

yana-kholod@inbox.ru

### **Введение**

Геофизические исследования скважин (ГИС) играют ключевую роль в изучении горных пород, предоставляя данные о их составе и свойствах. Несмотря на наличие специализированных программ, интерпретация данных ГИС во многих случаях выполняется вручную на основе экспертных знаний из-за недостаточного объёма информации, изменчивости измерений и шумов в данных. По этим же причинам прямое применение методов машинного обучения (ML), включая классификацию, регрессию и кластеризацию для интерпретации ГИС часто оказывается неэффективным и ведет к недостоверным результатам. Цель работы заключается в подборе методов ML для одного из видов интерпретации ГИС - фациального анализа, и способов повышения его точности за счет использования синтетических данных при малом объёме размеченных данных и высокой изменчивости измерений.

### **Основная часть**

Фациальный анализ — это метод геологической интерпретации, направленный на выделение фаций и их закономерной смены по разрезу и площади, реконструкцию условий осадконакопления и прогноз строения песчаных тел и связанных с ними коллекторских свойств. Одной из методик для проведения фациального анализа является методика В.С. Муромцева, основанная на анализе формы кривых, полученных при геофизических исследованиях скважин [1]. Эта методика требует больших усилий специалистов, что делает автоматизацию важным шагом для повышения точности и снижения зависимости от человеческого фактора. В этой работе предлагается использовать методы машинного обучения (ML) для автоматической интерпретации данных.

Для автоматизации фациального анализа можно выделить два основных типа задач:

1) Задачи классификации фаций на основе данных ГИС. Эти задачи включают использование алгоритмов машинного обучения для классификации условий осадконакопления в пределах заданных интервалов глубин [2].

2) Задачи по улучшению качества классификации в условиях малых данных. Современные алгоритмы, такие как нейросети и ансамблевые методы, позволяют адаптировать модели даже при ограниченных объёмах размеченных данных, что значительно повышает точность анализа и устраняет проблемы, связанные с межскважинным смещением распределений [3]. Этот подход направлен на улучшение устойчивости и точности фациального анализа, даже при высокой изменчивости данных.

В частности, для фациальной классификации на основе формы кривой  $\alpha$ -ПС были опробованы различные алгоритмы машинного обучения, такие как метод опорных векторов (SVM), деревья решений и ансамблевые методы. Сравнение результатов показало, что при использовании ансамблевых методов можно добиться более высокой точности классификации и устойчивости модели [4].

В ходе экспериментов по улучшению качества классификации были протестированы алгоритмы RandomForest, SVM и MLP с использованием двух сценариев: Baseline (обучение

на реальных данных) и Generated (обучение с синтетическими данными, созданными с помощью GMM). Основная метрика для оценки —  $f1\_macro$ .

RandomForest показал стабильные результаты с  $f1\_macro$  в пределах 0.34–0.36, с незначительным улучшением при использовании синтетических данных.

SVM показал большую вариативность по фолдам, что свидетельствует о зависимости от конкретных скважин. Однако синтетика улучшила результат с 0.3068 до 0.3305.

MLP не показал хороших результатов:  $f1\_macro$  около 0.14, а использование синтетики ухудшило результат.

Для кластеризации с использованием KMeans синтетика также оказалась неэффективной. Метрики ухудшились: показатель силуэта снизился с 0.0963 до 0.0359, индекс Калицкого-Харабаша уменьшился с 5.6779 до 4.4568, а индекс Дэвиса-Боулдина увеличился с 1.2219 до 1.4408, что указывает на размывание структуры кластеров.

### **Выводы**

Применение методов машинного обучения, включая нейросети и ансамблевые алгоритмы, возможно для фациального анализа данных ГИС, позволяет автоматизировать процесс и снизить зависимость от человеческого фактора. Наиболее высокая точность классификации и устойчивость модели получена при использовании ансамблевых методов. В экспериментах по повышению качества классификации алгоритмы RandomForest и SVM показали улучшение с синтетическими данными, а кластеризация с помощью KMeans не дала улучшений с синтетикой.

### **Литература**

1. Муромцев В. С. Электрометрическая геология песчаных тел—литологических ловушек нефти и газа.—Л.: Недра, 1984.— 260 с.
2. Лебедев И. С. Адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации. Информационно-управляющие системы, 2022, № 3, С. 20–30.
3. Мелихова О. А., Григораш А.С., Джамбинов С.В., Чумичев В. С., Гайдуков А.Б. Методы обучения в системах искусственного интеллекта // Технические науки – от теории к практике № 11. С. 19-29.
4. Руйго Д.И., Литвинов В.Л. Разработка метода выделения коэффициентов спектрограмм электрофаций на базе технологии MFCC // Международная научная конференция по проблемам управления в технических системах, 2021 Т.1 С. 262-264.