

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Математико-Механический факультет
Кафедра информационно-аналитических систем

Применение методов машинного обучения для автоматической разметки результатов геофизического исследования скважин

Курсовая работа студента 546 группы
Чурикова Никиты Сергеевича

Научный руководитель:
Доцент ГРАФЕЕВА Н. Г.

Заведующая кафедрой:
Доцент МИХАЙЛОВА Е. Г.

Санкт-Петербург
2017 г.

Содержание

1 Введение	1
2 Обзор литературы	1
3 Постановка задачи	2
4 Решение	2
4.1 Анализ данных угольных шахт	3
5 Заключение	3

1 Введение

Машинное обучение проникает во многие сферы нашей жизни [?] автоматизируя различные рутинные процессы, вроде поездок на машине [?] и обработки рутинных документов. Поэтому у профессионалов из различных областей естественно возникает желание сократить время работы на не столь увлекательных задачах.

В данном тексте пойдет речь о применении машинного обучения в области геофизики. У специалистов в этой области есть очень трудоемкая задача по выделению на различной глубине в почве пород, основываясь на так называемых методах каротжа. Будет показано, что представляют из себя данные скважин, которые геофизики анализируют, какие наработки, продукты и технологии в данной области уже есть, а также будут приведены наработки и идеи автора по данной задаче.

2 Обзор литературы

Идея применения методов машинного обучения к задаче выделения пород в скважине не новая. Существует достаточно много литературы и статей на эту тему.

Начать разбираться в области применения машинного обучения к классификации литологии стоит с соревнования по данному вопросу [3], которое проводилось сообществом SEG [5]. В этом контексте приводят отличный пример того, как начинать с работать данными по скважинам. Также благодаря этому конкурсу, существует открытый датасет с разметкой пород. Они также объясняют и показывают, что породы бывают трудноразличимыми и потому ошибка в одну похожую породу допустима.

Также по результатам этого соревнования были написаны интересные статьи, которые кратко описывают научные результаты конкурса.

Статья [1] подводит итоги и рассказывает о том, как генерировать новые атрибуты. Помимо стандартных подходов, вроде попарных перемножений фичей, они также предлагают считать градиент от атрибутов, воспринимая фичу, как функцию от глубины. Данный подход оказался достаточно удачным и был использован во всех лучших решениях. Помимо этого, работа показывает, что, что лучшим алгоритмом соревнования были деревья основанные на градиентном бустинге [?].

В статье [2] приведена попытка применить популярный алгоритм convolutional neural network (CNN) [?] к данным соревнования. Но, несмотря на то, что они популярны, и то что атрибуты являются вещественными значениями, на этих данных алгоритм не попал даже в десятку лучших решений. Авторы статьи утверждают, что проблема заключается в недостаточном количестве данных.

Неплохой литературой для начала погружения в геофизику и машинное обучение является книга Мухамедиева Р.И. [4]. В этой работе приведено хорошо описание методов каротажа, базовых алгоритмов машинного обучения, а также приводятся рекомендации по подготовке таких специфичных данных. В частности, они не рекомендуют использовать вейвлет преобразования [?], а советуют обратить внимание на следующие этапы предобработки данных:

1. Удаление аномальных значений;
2. Линейная нормировка;
3. Очистка данных по методу «ближайших соседей»;
4. Формирование плавающего окна данных.

3 Постановка задачи

Дана информация об обработке одной или нескольких скважин в некотором месторождении и известно, что на глубине d_i^j встретилась порода y_i^j , где j – номер скважины. Также для каждой скважины j и для каждой глубины i известны значения применявшихся методов исследования скважин x_i^j – **методов каротажа**.

По данным X необходимо сделать прогноз в новой скважине о том, какие породы в ней встретились для каждой глубины i .

Получается, что данную задачу можно интерпретировать, как задачу *классификации*, где X – наблюдаемые значения, а y – целевая переменная.

4 Решение

В данной статье были использованы размеченные данные, полу

4.1 Анализ данных угольных шахт

В рамках проверки концепции, что можно делать прогноз о типе литологии по данным каротажа, были использованы данные, полученные в результате исследования наличия угля в месторождении. Нам были предоставлены данные по двум месторождениям: 24 скважины из первого месторождения и 7 скважин из второго. На обоих месторождениях использовались, в большинстве своем, одинаковые методы каротажа, однако есть различия два метода.

На Рис. 1 можно увидеть визуализацию первого месторождения, используя метод понижения размерности TSNE [?]. Видно, что некоторые литологии можно выделить разделяющей поверхностью, но также некоторые встречаются значительно чаще других, а некоторые литологии вообще не видно. Таким образом, встает проблема *несбалансированных классов*, что подтверждается также графиком распределения классов.

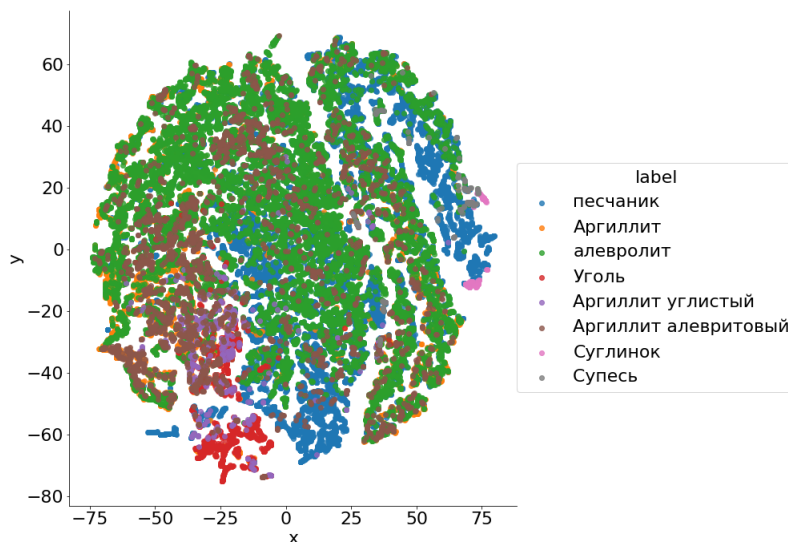


Рис. 1: Визуализация данных каротажа в двухмерном пространстве с использованием метода TSNE.

5 Заключение

Список литературы

- [1] Bestagini, P. A Machine Learning Approach to Facies Classification Using Well Logs / Paolo Bestagini, Vincenzo Lipari, Stefano Tubaro // SEG Technical Program Expanded Abstracts 2017. —

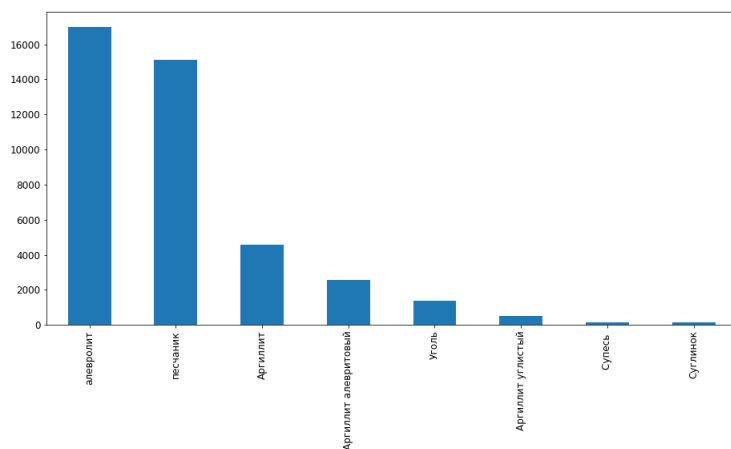


Рис. 2: Распределение литологий

2017. — P. 2137–2142. — <http://library.seg.org/doi/abs/10.1190/segam2017-17729805.1>.

- [2] Facies classification from well logs using an inception convolutional network / Valentin Tschannen, Matthias Delescluse, Mathieu Rodriguez, Janis Keuper. — 2017. — <http://arxiv.org/abs/1706.00613>.
- [3] Geophysics machine learning contest. — 2017. — oct. — <https://github.com/seg/2016-ml-contest>.
- [4] Muchamediev, R. I. Machine Learning methods applied to geophysics research / Ravil Ilgizovich Muchamediev. — Riga, 2016. — Vol. 200 p.
- [5] Society of explorational geophysicists. — 2017. — oct. — <http://seg.org>.