

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Математико-Механический факультет
Кафедра информационно-аналитических систем

Применение машинного обучения для анализа угольных скважин

Дипломная работа студента 546 группы
Чурикова Никиты Сергеевича

Научный руководитель:
Доцент ГРАФЕЕВА Н. Г.

Заведующий кафедрой:
Доцент МИХАЙЛОВА Е. Г.

Санкт-Петербург
2017 г.

Содержание

1 Введение	1
2 Обзор литературы	1
3 Заключение	2

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 литологии на различной глубине в почве. 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

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 по тому, как начинать с работать данными скважин, они выкладывают открытый датасет, на котором можно потренироваться, а также объясняют и показывают что литологии можно спутать с их "соседями" т.е. литологиями, которые трудно различить между собой даже геофизикам.

Также по результатам этого соревнования были написаны интересные статьи. Эти замечательные работы кратко описывают научные результаты конкурса. Статья [1] подводит итоги и рассказывает о том, как генерировать новые атрибуты используя глубину, а также показывает, что лучшим алгоритмом соревнования были деревья основанные на градиентном бустинге [?]. В статье [2] приведена попытка применить популярный алгоритм convolutional neural network (CNN) [?]. 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 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. — 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>.