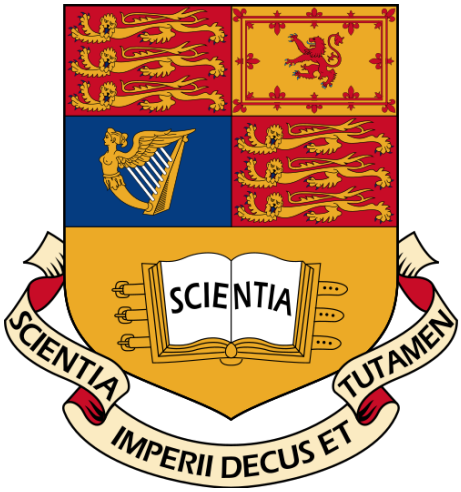

APPLIED COMPUTATIONAL SCIENCE AND ENGINEERING - Course 2018/2019

EARTH SCIENCE DEPARTMENT

Application of Machine Learning to Unbiased Classification of Well Log Data

PROJECT PLAN

Author: Runzhi Zhou
Supervisor: Dr. Gerard Gorman, Dr. Navjot Kukreja and Dr. Peter Fitch

Contents

1 Introduction

1.1 Background

As the "Age of Big Data" kicks in, understanding the underlying meaning of the data becomes an essential task for a wide range of scientific and commercial applications, for example, online shopping and bioscience [1, 2].

There hasn't been much success of machine learning data analysis in geoscience-related area because of the diversity of the geological environment and limitation of data. However, well-log data could be a good source for detecting complexity and characteristic (e.g.lithologic variation) of geological environment by using machine learning techniques [3]. More details is included in the literature review section.

This project aims to achieve an unbiased classification (of lithology, fluid and facies) for wells in the same oil field and for wells across other oil fields.

1.2 Data

There are three sets of data from different sources. In this project, I will spend more time on data set two (from wells in the Wytch Farm oilfield) as it contains more labels information than the other two data sets. More details of time plan are included in the time plan section. The table below labels the origin and description of the data.

Data	**Origin of the data**	**Description of the data**
Data set 1	the Hugoton and Panoma Fields [2]	9 wells with lithology labels with no fluid labels with facies labels
Data set 2	the Wytch Farm oilfield	35 wells 2 wells with lithology labels 2 wells with fluid labels 2 wells with facies labels
Data set 3	Ovation Data Services's data from many oil fields (Location unknown)	Contain about 1000 wells all wells need to be labelled

Table 1: data description

NOTE: there are three different kinds of labels: lithology (e.g. sandstone, mudstone and Limestone) labels, fluid labels (e.g. gas, oil and water) and facies labels. Facies labels are the hardest to predict.

1.3 The rationale and project objectives

Both the workflow and method selection could impact the result for lithology, fluid or facies clustering [4]. Supervised learning could be unreliable due to limited number of labelled data set. Preprocessing more training samples could overfit the training networks. This project will therefore focus more on unsupervised methods. It is worth to try different workflow and data analysis methods.

This project aims to achieve an unbiased classification (of lithology, fluid and facies) for wells in the same oil field and for wells across other oil fields. There are **TWO main objectives** in this project.

1. **Firstly**, I would like to use unsupervised learning to assign labels(lithology, fluid and facies) to wells that have not been described in the same oil field.
2. **Secondly**, I would like to try other methods (including lowering the dimensions with unsupervised learning and supervised learning) to assign SOME OF THE labels (lithology, fluid or some of the facies) to wells that have not been described across different oil fields, which aims to achieve better efficiency.

1.4 Proposed approach

To achieve **the FIRST objective** (use unsupervised learning to assign labels to wells that have not been described in the same oil field), there are four important steps:

1. **Choose Input**: Using methods (e.g. PCA, t-SNE, SOM) to determine the importance of each input dimension and relationship between each input dimension.
2. **Clustering**: Using unsupervised methods (K-means or others) to cluster the training data into different categories.
3. **Label**: Label the clusters based on matching values with original labels (as a classifier).
4. **Application**: Use the labelled clusters to identify the categories in unlabelled wells data sets (in the same field).

To achieve **the SECOND objective** (to assign some of the labels to wells that have not been described across different oil fields), there are some other methods worth trying:

1. Unsupervised clustering with lowering the dimensions;
2. Artificial neural network (ANN)/ logistic regression/ decision tree.

1.5 literature review

1.5.1 Big Data and Geo-science

Min et al. mention that big data could be applied to commercial analysis, science analysis, structured data analysis and text data analysis [1]. All these historical data-based knowledge can help human and industries to make faster and more accurate decisions [5].

In geo-science, well data has a large variety of data types, and therefore could be useful to perform correlation analysis using machine learning techniques [4].

1.5.2 Machine Learning

Supervised learning and unsupervised learning are two categories of machine learning. If the data are given with labels then the learning is supervised. In contrast, unlabeled data can only do unsupervised learning [6].

Noticeably, Scikit-learn is a widely used Python module with machine learning algorithms for medium-scale supervised and unsupervised problems [7].

1.5.3 Applications

By using data analysis and machine learning methods, new information and new business values could be developed [4].

In the area of geoscience, industries have large data sets in many formats. For example, well logs, core fluid analysis, drilling parameters, seismic surveys, etc. In specific, well logs data can provide new and better information on porosity, lithology and fluid properties among others [4].

In 1982, Martin Wolf and Jacques Pelissier-Combescure first enabled clustering into electrofacies (for geological facies determination and well to well correlation). They defined Electrofacies as a set of log responses characterizing a sediment [8]. In 1987, J.M. Busch et al. used statistical analysis of wire-line log measurements to enable determination of lithology [9]. In 1997, Jong-Se Lim et al. applied hierarchical cluster analysis on well log measurements [10]. In 2001, Trond Mathisen et.al successfully applied non-parametric regression techniques to characterize electrofacies with permeability prediction [11]. Recently, Y. Zee Ma analysed the popular artificial neural network (ANN) for lithology or facies clustering with improved solutions and workflows [12].

In conclusion, both the workflow and method selection could impact the result for lithology, fluid or facies clustering. It is worth to try different workflow and machine learning methods on well log data [3].

2 Time plan and milestons

2.1 Gantt chart

Figure 1: Gantt chart

Figure 1 is the gantt chart for this project. The project starts from 17th June and ends at 30th August with 11 weeks in total. There are five categories in this gantt chart. The blue, red, orange, purple, and green lines represent the literature reviews, basic methods applications, methods development, data testing, and report writing respectively.

2.2 The milestones are listed with expected and date

- **Week 2 (This week)**

✓ Finish reviewing the wells&geophysics related and clustering related literatures; (Finished)
✓ Attempt k-means and PCA methods on dataset 1; (Finished)
✓ Attempt the correlation between each input dimension;(Finished)
✓ Finish the planing report; (Finished)

- **Week 3 (05 JUL)**

o Finish reviewing all the useful clustering methods;
o Attempt other basic methods in SK-learn library on dataset 1;

- **Week 4 (12 JUL)**

o Finish reviewing all the possible methods for mashine learning on well-log;
o Attempt lower dimension methods on data set 1 and 2;

- **Week 5 (19 JUL)**

o Keep testing data set 1 and data set 2;
o start testing data set 3 (the final data set);
o write comparision and final outline for the final report;

- **Week 6 (26 JUL)**

o Finish all the basic testing on data set 1;
o Start testing data set 3;

- **Week 7 (02 AUG)**

o Finish all the basic testing on data set 2;

- **Week 8 (09 AUG)**

o Final test the unsupervised methods and try other methods;

- **Week 9 (16 AUG)**

o Stop attempting new methods;
o Making conclusion and develope tried methods;
o Finish a first draft of the final report;

- **Week 10 (23 AUG)**

o Finish all the testing on the data set 3 (the final data);

- **Week 11 (30 AUG)**

o Finish the final report;

Figure 2: milestones

NOTE: There will be a meeting with company (Ovation Data Service) between the third week and fourth week. An updated plan will be finished after the meeting.

References

[1] Min Chen, Shiwen Mao, and Yunhao Liu. Big data: A survey. *Mobile networks and applications*, 19(2):171–209, 2014.

[2] Brendon Hall. Facies classification using machine learning. *The Leading Edge*, 35(10):906–909, 2016.

[3] Joel Gevirtz, Adriana Ovalle, et al. Lithofacies classification for earth modeling: Letting the data speak for itself. In *SPE Unconventional Resources Conference*. Society of Petroleum Engineers, 2017.

[4] J Johnston, A Guichard, et al. New findings in drilling and wells using big data analytics. In *Offshore Technology Conference*. Offshore Technology Conference, 2015.

[5] Martin Hilbert. Big data for development: A review of promises and challenges. *Development Policy Review*, 34(1):135–174, 2016.

[6] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160:3–24, 2007.

[7] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[8] Martin Wolf, Jacques Pelissier-Combescure, et al. Faciolog-automatic electrofacies determination. In *SPWLA 23rd Annual Logging Symposium*. Society of Petrophysicists and Well-Log Analysts, 1982.

[9] JM Busch, WG Fortney, LN Berry, et al. Determination of lithology from well logs by statistical analysis. *SPE formation evaluation*, 2(04):412–418, 1987.

[10] Jong-Se Lim, Joe M Kang, Jungwhan Kim, et al. Multivariate statistical analysis for automatic electrofacies determination from well log measurements. In *SPE Asia Pacific Oil and Gas Conference and Exhibition*. Society of Petroleum Engineers, 1997.

[11] Trond Mathisen, Sang Heon Lee, Akhil Datta-Gupta, et al. Improved permeability estimates in carbonate reservoirs using electrofacies characterization: a case study of the north robertson unit, west texas. In *SPE Permian Basin Oil and Gas Recovery Conference*. Society of Petroleum Engineers, 2001.

[12] Y Zee Ma. Lithofacies clustering using principal component analysis and neural network: applications to wireline logs. *Mathematical Geosciences*, 43(4):401–419, 2011.