A simplified expert system for porosity classification using linguistic petrographical descriptions

Yilu Liu

College of Engineering and Computer Science
The Australian National University
Acton, Canberra, ACT, 2601
AUSTRALIA
u6270742@anu.edu.au

Abstract. Expert system is an interactive computer-based system that implements artificial intelligence technologies to simulate human decision-making progress. The purpose of an expert system is to solve data classification problem in analysis studies where linguistic descriptions are heavily used as those descriptions are hard to classify. This paper majorly implements a simplified expert system, along with experiments on artificial neural network and genetic algorithm, to classify porosity into four categories based on linguistic petrographical characteristics. Practically, the approach predicts categorical results from discrete and categorical linguistic groups. Limited by the dataset size and the knowledge base implemented, the expert system and the neural network does not generate a decent performance and therefore, further improvement could be made on the implementation of the three systems.

Keywords: Expert system; Neural Network; Genetic Algorithm; Porosity; Language classification

1 Introduction

Reservoir characterization is a complicated progress in geophysics fields that often requires quantitative geological data analysis. Being used in reservoir modelling and simulation and various enhanced recovery design process in the petroleum industry, reservoir characterization could also be used to for environment remediation [1]. Study in heterogeneity in clastic hydrocarbon reservoir is of great significance as to understand how to characterise non-uniform and non-linear spatial distribution of rock properties in reservoirs, such as porosity, oil and water saturation. To achieve this, well established analysis on the anatomy of reservoir heterogeneity, including component lithofacies analysis, needs to be conducted [2].

Geophysical well logging is a process that is used for collecting lithofacies features and calculate formation porosity. However, while this process has been implemented for decades, the calculations done in the process still majorly produces linear or modelled non-linear relationship between the rock properties, with other doubtful assumptions, e.g. uniformly distributed grain size/density, constant water resistivity. Furthermore, the mathematical modelled used to calculate porosity is not as accurate given that the calculation result is sometimes too absolute, which would not be representative under the real environments [1]. Therefore, new technologies could be implemented to improve the analysis level. One alternative way would be implementing artificial intelligence technologies to the petroleum fields studies. There are different categories of artificial intelligence, such as rule based expert systems and adaptive neural systems. This paper will focus on the first one [3]. The purpose of implementing the expert system is to observe, recognise and identify the underlying geological problems in a way that will be addressable by human decision-making process [2]. With rapid response speed and high performance, computers are capable of capturing hidden information from rocks' spatial distribution easily [4].

2 Methodology

Lithofacies recognition in heterogeneous reservoirs could be different as different geologist might come up with different lithofacies sets based from the same observed data and there is a lack of quantitative measurements towards those linguistic sets [2]. Therefore, this paper aims to introduce the expert system approach with different data pre-processing techniques to evaluate the linguistic descriptions used to describe the petrographical characteristics.

The general structure of the process in shown in Figure 1. Based on this structure, I will briefly cover the data pre-processing strategies that have been used to encode the input data and the experiment process that happened during knowledge base tuning for the expert system.

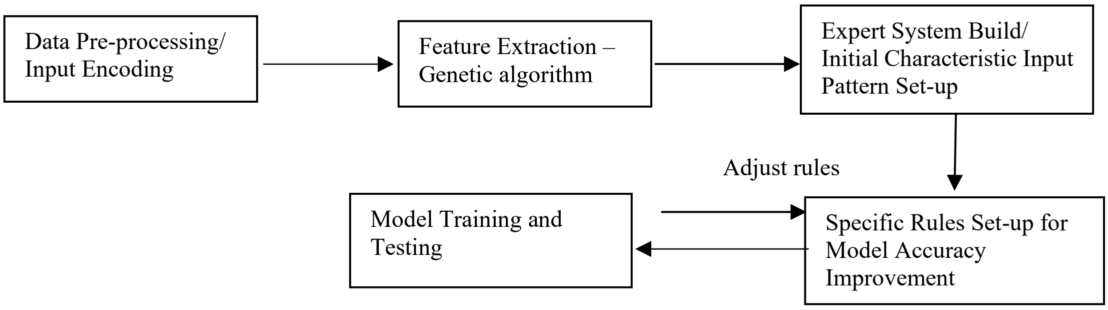

Fig. 1. General progress diagram.

2.1 Data Pre-processing

This paper uses a dataset that includes 226 core plug samples collected from an oil well in North West Shelf, Australia. The reservoir located there is mainly composed of carbonate cemented facies and sandstones [2]. The 226 samples contain six characters that relate to the samples' porosity, each with different numbers and types of character attributes:

- Grain Size
  With 12 attributes from *Medium, Fine* to *V. fine*. Since the 12 attributes are linguistically related to one another, they could be encoded as a continuous numerical sequence ranging from 0 to 1.
- Sorting
  With 8 attributes from *Poor* to *Well*. Similar to Grain Size, could be implemented as numerical sequence between 0 and 1.
- Matrix
  With 16 attributes that somewhat shows relativity with each other but it's not as obvious as Grain Size and Sorting. However, this input data has been divided into 16 categorical data attributes that are also normalised to 0 – 1.
- Sphericity & Roundness
  With 8 attributes showing a circular linguistic connection with the keywords 'Angular' and 'Rounded'. By generating a circular encoding, the values in Figure 2 (b) could then be normalised to 0 – 1 again, with accurate position presentation.

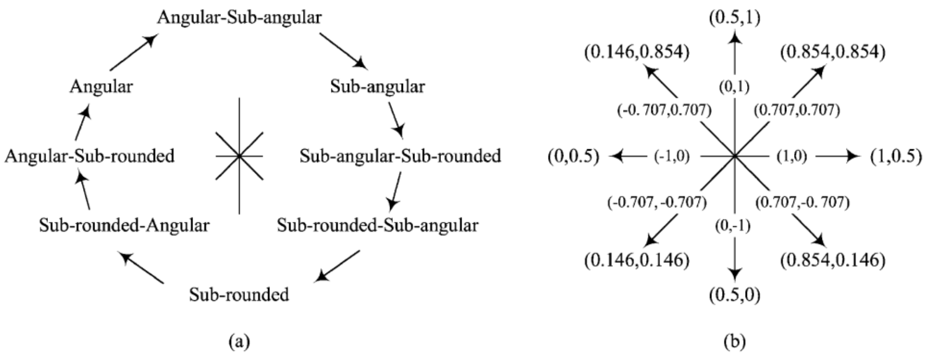

Fig. 2. (a). Circular encoding of Sphericity and Roundness

(b). Figure 2. (a) normalised to *sine* and *cosine* values [2].

- Bioturbation
  With 4 attributes from *Minor bioturbation* to *Abundant bioturbation*, been encoded into an ascending numerical sequence in 0 to 1.
- Laminae
  With 10 distinct attributes that don't show much similarity. I encoded it using a similar way with other characters, but in the later stage, the rules generated for Laminae in the expert system is different.

2.2 Expert System

In this case, the expert system is built by me and I'm also a Knowledge Engineer who generates rules for the expert system to do a specific task, e.g., translate input sequence into a score and find the connection between the score and the output data. To improve the accuracy of the predictions the system makes, I will manually check if the rules generated will be beneficial to the system using different training sets. Here during experiment, I combined the rules with characteristic input pattens so that each input pattern becomes a rule on its own. Some characteristic patterns could be configured through a number of statistical methods [5], but in this paper I have used the arithmetical mean and the standard deviation of the vector components extracted from the input data.

Afterwards, a test program, with the input data split into training set and test, will be written for the system to test on different scenarios and the response of the system will then be evaluated. During evaluation, the Knowledge Engineer needs to frequently go back and improves the rule set. Typically, the Knowledge Engineer has to model the tasks for the system, like conceiving the initial rules and knowledge structures, learn new rules and test them on different scenarios. Hence, the expert system required a lot of human interaction and maintenance.

During training, I have generated several rules based on each character's porosity distribution, e.g., I calculated the correlation coefficient between each character and porosity and gave limit to characters' value range to improve the estimation accuracy. Below is a typical rule for character Matrix, indicating its value should be within a certain range.

```
# Rules for Matrix #
if (row.Matrix < (vp_df['Matrix'].mean() + poor_df['Matrix'].mean())/2):
    matching_rule = "Matrix < " + str((vp_df['Matrix'].mean() + poor_df['Matrix'].mean())/2)
    matching_rules.append(matching_rule)
    vp_count += 1
```

Fig. 4. Sample rule, generated for one of the input values, character Matrix.

2.3 Neural Network Training

As a way to obtain rules for the expert system, neural network training could help with the automation process of the system that less human supervision will be needed. Here in this research, I trained a neural network using Sigmoid activation function and MSE Loss function as they have been widely used and offers decent performance in various cases, and based on the input data encoding, they would be a good choice for this research.

2.4 Feature Selection and Genetic Algorithm

For raw datasets we get, feature selection is a technique used to either automatically or manually select features that would benefit the most to our prediction variable and git rid of irrelevant features. Genetic algorithm would be an interesting one to be implemented in this research to help the system get better results while cut down the human effort. The algorithm contains several procedures, here's an ordered list: define fitness, select, crossover and mutate functions for the particular model, then translate population to DNA, select individuals with high fitness, implement the crossover and mutate function, then choose DNA with highest accuracy to get the features we want to have in the system. Due to the obstacles the research has met, further explanation of the system won't be the focus here.

3 Results and Discussion

Unfortunately, the research has encountered some difficulty and the three parts of the system, i.e., the expert system, the neural network and the genetic algorithm were not grouped together in use. With the 80% of the input data split into training data, the expert system predicted 180 samples and only guessed 76 of the porosity classes correctly, giving a 42.22% accuracy. From an external point of view, the size of the dataset could be one reason of the low accuracy. Nevertheless, as the key feature of expert system, the rules generated by me should be the main reason that caused the performance issue. One thing to note that is despite being encoded into sequences, some of the input attributes still have quite discrete relationships with the raw porosity values, as shown in Figure 5. This could be one possible reason that the system couldn't make accurate predictions as the casual relationship of the inputs to outputs are hard to obtain. With calculating the casual index (the rate of change of an output y with respect to an input x), the input characteristic patterns would be more accurate and affective [6].

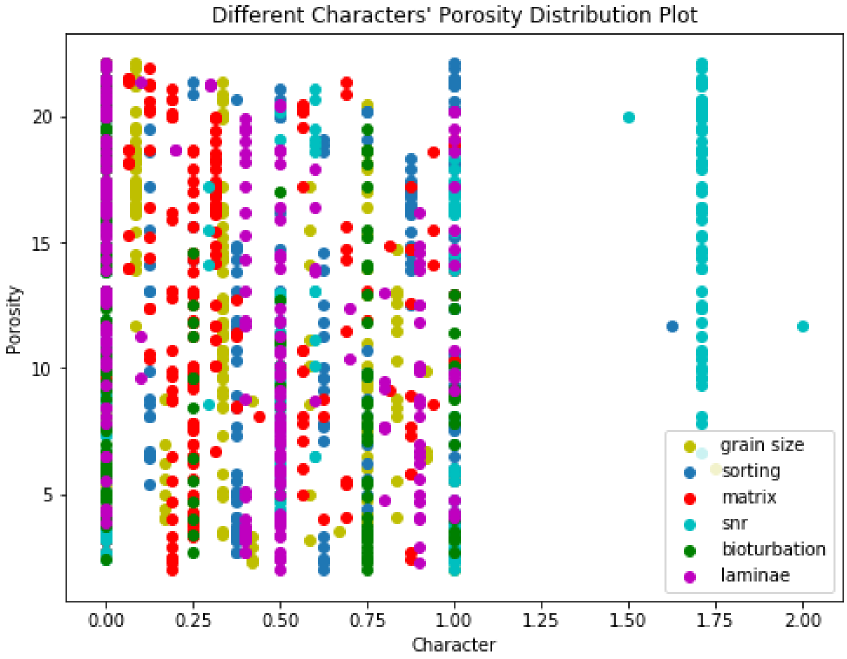

Fig. 5. Samples porosity distribution plot based on characters.

For the neural network model, research has suggested a clear limit of number of neurons in a hidden layer of neural network, which is within the range of *2/3 the size of input plus output layer* and *twice the size of input layer* [7]. Similarly, the learning rate of the model could be determined by the gradient of the MSE loss function [8]. Following this guidance, with hit-and-trail on epoch numbers, the network model reached peak performance when there are 12 hidden neurons, 5,000 epochs and when the learning rate is set to 0.1, and the I consider the result to be rather low-performing by having 20% - 40% training accuracy and only 10% - 20% testing accuracy. Below is a typical Loss change graph over the training. Obviously, as the loss converges from ~1000 epochs, the performance also stops improving, from this graph we could conclude that the model is not learning much for the rest of the epochs and this could be caused by the size of the dataset as well as potential misestimation of the loss function gradient. More experiments need to be conducted to improve the network's performance.

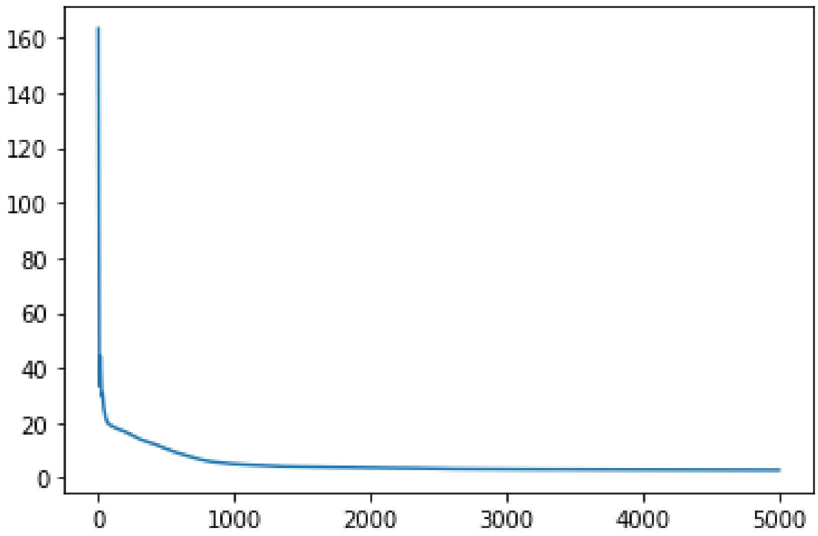

Fig. 6. Loss change (y-axis) diagram when number of epoch (x-axis) is 5000.

As for genetic algorithm, I wasn't able to obtain the target and fitness function, hence there's not valuable result generated from this side, which is very unfortunate.

4 Conclusion and Future Work

This paper provides a simplified version of expert system. The accuracy of the paper's approach is affected by the size of the network's training set. However, as we could see, compared with the original paper which has first used the technique to extract rules from neural networks and has over 60% accuracy on their model prediction, this research is considerably flawed as the implementation is not successful and better solutions should be designed. With the features simplifies, the system was not able to make effective and precise analysis on the data model based on limited rules. This, on the other hand, proves that the necessity of using a complete formula to calculate the Casual Index to enhance the accuracy of the prediction.

To further improve, using neural networks to help generate the rules in the expert system together would help the system make better decisions and the rules would become supportive. In detail, the activation values of a particular case from the neural network could be encoded into rules. Moreover, the trained network could be used as fitness function to calculate fitness in the genetic algorithm.

References

[1] S. Mohaghegh, R. Arefi, S. Ameri, K. Aminiand and R. Nutter, "Petroleum reservoir characterization with the aid of artificial neural networks", Journal of Petroleum Science and Engineering, vol. 16, no. 4, pp. 263-274, 1996. Available: 10.1016/s0920-4105(96)00028-9.

[2] T. Gedeon, D. Tamhane, T. Lin and P. Wong, "Use of linguistic petrographical descriptions to characterise core porosity: contrasting approaches", Journal of Petroleum Science and Engineering, vol. 31, no. 2-4, pp. 193-199, 2001. Available: 10.1016/s0920-4105(01)00130-9.

[3] S. Mohaghegh, and S. Ameri, "Artificial Neural Network As A Valuable Tool For Petroleum Engineers", Citeseerx.ist.psu.edu, 1995. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.112.4627&rep=rep1&type=pdf. [Accessed: 03- May- 2020].

[4] A. Ouenes, S. Richardson and W. Weiss, "Fractured Reservoir Characterization and Performance Forecasting Using Geomechanics and Artificial Intelligence", SPE Annual Technical Conference and Exhibition, 1995. Available: 10.2118/30572-ms

[5] T. D. Gedeon, S. Turner, "Explaining student grades predicted by a neural network", In Neural Networks, 1993. IJCNN'93-Nagoya. Proceedings of 1993 International Joint Conference, vol. 1, pp. 609-612. IEEE.

[6] H, Turner, T. D. Gedeon, "Extracting Meaning from Neural Networks", Proceedings 13th International Conference on AI, Avignon, vol. 1, pp. 243-252, 1993

[7] K. Sheela and S. Deepa, "Review on Methods to Fix Number of Hidden Neurons in Neural Networks", Mathematical Problems in Engineering, vol. 2013, pp. 1-11, 2013. Available: 10.1155/2013/425740.

[8] N. Buduma and N. Locascio, Fundamentals of deep learning, 1st ed. O'Reilly Media, Inc., 2017.