

Preface

Integration, handling data of immense size and uncertainty, and dealing with risk management are among crucial issues in petroleum geosciences. The problems one has to solve in this domain are becoming too complex to rely on a single discipline for effective solutions, and the costs associated with poor predictions (e.g., dry holes) increase. Therefore, there is a need to establish new approaches aimed at proper integration of disciplines (such as petroleum engineering, geology, geophysics, and geochemistry), data fusion, risk reduction, and uncertainty management.

This book presents several artificial intelligent approaches¹ for tackling and solving challenging practical problems from the petroleum geosciences and petroleum industry. Written by experienced academics, this book offers state-of-the-art working examples and provides the reader with exposure to the latest developments in the field of artificial intelligent methods applied to oil and gas research, exploration, and production. It also analyzes the strengths and weaknesses of each method presented using benchmarking, while also emphasizing essential parameters such as robustness, accuracy, speed of convergence, computer time, overlearning, or the role of normalization.

The reader of this book will benefit from exposure to the latest developments in the field of modern heuristics applied to oil and gas research, exploration, and production. These approaches can be used for uncertainty analysis, risk assessment, data fusion and mining, data analysis and interpretation, and knowledge discovery, from diverse data such as 3-D seismic, geological data, well logging, and production data. Thus, the book is intended for petroleum scientists, data miners, data scientists and professionals, and postgraduate students involved in the petroleum industry.

Petroleum Geosciences are—like many other fields—a paradigmatic realm of difficult optimization and decision-making real-world problems. As the number,

¹ Artificial Intelligence methods, some of which are grouped together in various ways, under names such as *Computational Intelligence*, *Soft Computing*, *Meta-heuristics*, or *Modern heuristics*.

difficulty, and scale of such specific problems increase steadily, the need for diverse, adjustable problem-solving tools can hardly be satisfied by the necessarily limited number of approaches typically included in a curriculum/syllabus from academic fields other than Computer Science (such as Petroleum Geology). Therefore, the first three chapters of this volume aim at providing working information about modern problem-solving tools, in particular in machine learning and in data mining, and also at inciting the reader to look further into this thriving topic.

Traditionally, solving a given problem in mathematics and in sciences at large implies the construction of an abstract model, the process of proving theoretical results valid in that model, and eventually, based on those theoretical results, the design of a method for solving the problem. This problem-solving paradigm has been and will continue to be immensely successful. Nevertheless, an abstract model is an approximation of the real-world problem; there have been failures triggered by a tiny mismatch between the original problem and the proposed model for it. Furthermore, a problem-solving method developed in this manner is likely to be useful only for the problem at hand. While, ultimately, any problem-solving technique may be—in various degrees—subject to these two observations, some relatively new approaches illustrate alternative lines of attack; it is the editors' hope that the first three chapters of the book illustrate this idea in a way that will prove to be useful to the readers.

In the first chapter, Simovici presents some of the main paradigms of intelligent data analysis provided by machine learning and data mining. After discussing several types of learning (supervised, unsupervised, semi-supervised, active, and reinforcement learning), he examines several classes of learning algorithms (naïve Bayes classifiers, decision trees, support vector machines, and neural networks) and the modalities to evaluate their performance. Examples of specific applications of algorithms are given using System R.

The second and third chapters, by Luchian, Breaban, and Bautu, are dedicated to *meta-heuristics*. After a rather simple introduction to the topic, the second chapter presents, based on working examples, evolutionary computing in general and, in particular, genetic algorithms and differential evolution; particle swarm optimization is also extensively discussed. Topics of particular importance, such as multimodal and multi-objective problems, hybridization, and also applications in petroleum geosciences are discussed based on concrete examples. The third chapter gives a compact presentation of genetic programming, gene expression programming, and also discusses an R package for genetic programming and applications of GP for solving specific problems from the oil and gas industry.

Ashena and Thonhauser discuss the Artificial Neural Networks (ANNs), which has the potential to increase the ability of problem solving in geosciences and in the petroleum industry, particularly in case of limited availability or lack of input data. ANN applications have become widespread because they proved to be able to produce reasonable outputs for inputs they have not learned how to deal with. The following subjects are presented: artificial neural networks basics (neurons, activation function, ANN structure), feed-forward ANN, back-propagation and learning, perceptrons and back-propagation, multilayer ANNs and back-propagation

algorithm, data processing by ANN (training, overfitting, testing, validation), ANN, and statistical parameters. An applied example of ANN, followed by applications of ANN in geosciences and petroleum industry complete the chapter.

Al-Anazi and Gates present the use of support vector regression to accurately estimate two important geomechanical rock properties, Poisson's ratio and Young's modulus. Accurate prediction of rock elastic properties is essential for wellbore stability analysis, hydraulic fracturing design, sand production prediction and management, and other geomechanical applications. The two most common required material properties are Poisson's ratio and Young's modulus. These elastic properties are often reliably determined from laboratory tests by using cores extracted from wells under simulated reservoir conditions. Unfortunately, most wells have limited core data. On the other hand, wells typically have log data. By using suitable regression models, the log data can be used to extend knowledge of core-based elastic properties to the entire field. Artificial neural networks (ANN) have proven to be successful in many reservoir characterization problems. Although nonlinear problems can be well resolved by ANN-based models, extensive numerical experiments (training) must be done to optimize the network structure. In addition, generated regression models from ANNs may not perfectly generalize to unseen input data. Recently, support vector machines (SVMs) have proven successful in several real-world applications for its potential to generalize and converge to a global optimal solution. SVM models are based on the structural risk minimization principle that minimizes the generalization error by striking a balance between empirical training errors and learning machine capacity. This has proven superior in several applications to the empirical risk minimization principle adopted by ANNs that aims to reduce the training error only. Here, support vector regression (SVR) to predict Poisson's ratio and Young's modulus is described. The method uses a fuzzy-based ranking algorithm to select the most significant input variables and filter out dependency. The learning and predictive capabilities of the SVR method is compared to that of a back-propagation neural network (BPNN). The results demonstrate that SVR has similar or superior learning and prediction capabilities to that of the BPNN. Parameter sensitivity analysis was performed to investigate the effect of the SVM regularization parameter, the regression tube radius, and the type of kernel function used. The result shows that the capability of the SVM approximation depends strongly on these parameters.

The next three chapters introduce the active learning method (ALM) and present various applications of it in petroleum geosciences.

First, Cranganu, and Bahrpeyma use ALM to predict a missing log (DT or sonic log) when only two other logs (GR and REID) are present. In their approach, applying ALM involves three steps: (1) supervised training of the model, using available GR, REID, and DT logs; (2) confirmation and validation of the model by blind-testing the results in a well containing both the predictors (GR, REID) and the target (DT) values; and (3) applying the predicted model to wells containing the predictor data and obtaining the synthetic (simulated) DT values. Their results indicate that the performance of the algorithm is satisfactory, while the performance time is significantly low. The quality of the simulation procedure was assessed by

three parameters, namely mean square error (MSE), mean relative error (MRE), and Pearson product momentum correlation coefficient (R). The authors employed both the measured and simulated sonic log DT to predict the presence and estimate the depth intervals where overpressured fluid zone may develop in the Anadarko Basin, Oklahoma. Based on interpretation of the sonic log trends, they inferred that overpressure regions are developing between $\sim 1,250$ and $2,500$ m depth and the overpressured intervals have thicknesses varying between ~ 700 and $1,000$ m. These results match very well previous published results reported in the Anadarko Basin, using the same wells, but different artificial intelligent approaches.

Second, Bahrpeyma et al. employed ALM to estimate another missing log in hydrocarbon reservoirs, namely the density log. The regression and normalized mean squared error (MSE) for estimating density log using ALM were equal to 0.9 and 0.042, respectively. The results, including errors and regression coefficients, proved that ALM was successful in processing the density estimation. In their chapter, the authors illustrated ALM by an example of a petroleum field in the NW Persian Gulf.

Third, Bahrpeyma et al. tackled the common issue when reservoir engineers should analyze the reservoirs with small sets of measurements (this problem is known as the small sample size problem). Because of small sample size problem, modeling techniques commonly fail to accurately extract the true relationships between the inputs and the outputs used for reservoir properties prediction or modeling. In this chapter, small sample size problem is addressed for modeling carbonate reservoirs by using the active learning method (ALM). Noise injection technique, which is a popular solution to small sample size problem, is employed to recover the impact of separating the validation and test sets from the entire sample set in the process of ALM. The proposed method is used to model hydraulic flow units (HFUs). HFUs are defined as correlatable and mappable zones within a reservoir controlling the fluid flow. This research presents quantitative formulation between flow units and well log data in one of the heterogeneous carbonate reservoirs in Persian Gulf. The results for R and $nMSE$ are 85 % and 0.0042, respectively, which reflect the ability of the proposed method to improve generalization ability of the ALM when facing with sample size problem.

Dobróka and Szabó carried out a well log analysis by global optimization-based interval inversion method. Global optimization procedures, such as genetic algorithms and simulated annealing methods, offer robust and highly accurate solution to several problems in petroleum geosciences. The authors argue that these methods can be used effectively in the solution of well-logging inverse problems. Traditional inversion methods are used to process the borehole geophysical data collected at a given depth point. As having barely more types of probes than unknowns in a given depth, a set of marginally overdetermined inverse problems has to be solved along a borehole. This single inversion scheme represents a relatively noise-sensitive interpretation procedure. To reduce the noise, the degree of overdetermination of the inverse problem must be increased. This condition can be achieved by using a so-called interval inversion method, which inverts all data from a greater depth interval jointly to estimate petrophysical parameters of hydrocarbon reservoirs to

the same interval. The chapter gives a detailed description of the interval inversion problem, which is then solved by a series expansion-based discretization technique. The high degree of overdetermination significantly increases the accuracy of parameter estimation. The quality improvement in the accuracy of estimated model parameters often leads to a more reliable calculation of hydrocarbon reserves. The knowledge of formation boundaries is also required for reserve calculation. Well logs contain information about layer thicknesses, which cannot be extracted by the traditional local inversion approach. The interval inversion method is applicable to derive the layer boundary coordinates and certain zone parameters involved in the interpretation problem automatically. In this chapter, the authors analyzed how to apply a fully automated procedure for the determination of rock interfaces and petrophysical parameters of hydrocarbon formations. Cluster analysis of well-logging data is performed as a preliminary data-processing step before inversion. The analysis of cluster number log allows the separation of formations and gives an initial estimate for layer thicknesses. In the global inversion phase, the model including petrophysical parameters and layer boundary coordinates is progressively refined to achieve an optimal solution. The very fast simulated reannealing method ensures the best fit between the measured data and theoretical data calculated on the model. The inversion methodology is demonstrated by a hydrocarbon field example, with an application for shaly sand reservoirs.

Finally, Mohebbi and Kaydani undertake a detailed review of meta-heuristics dealing with permeability estimation in petroleum reservoirs. They argue that proper permeability distribution in reservoir models is very important for the determination of oil and gas reservoir quality. In fact, it is not possible to have accurate solutions in many petroleum engineering problems without having accurate values for this key parameter of hydrocarbon reservoir. Permeability estimation by individual techniques within the various porous media can vary with the state of in situ environment, fluid distribution, and the scale of the medium under investigation. Recently, attempts have been made to utilize meta-heuristics for the identification of the relationship that may exist between the well log data and core permeability. This chapter overviews the different meta-heuristics in permeability prediction, indicating the advantages of each method. In the end, some suggestions and comments about how to choose the best method are presented.

December 2014

Constantin Cranganu
Henri Luchian
Mihaela Elena Breaban

Artificial Intelligent Approaches in Petroleum
Geosciences

Cranganu, C.; Luchian, H.; Breaban, M.E. (Eds.)

2015, XII, 290 p. 126 illus., 81 illus. in color., Hardcover

ISBN: 978-3-319-16530-1