Reservoir Characterization Using Support Vector Machines

Kok Wai Wong[1], Yew Soon Ong[2], Tamás D. Gedeon[3], Chun Che Fung[1]

[1]*Murdoch University School of Information Technology South St, Murdoch Western Australia 6150*

[2]*Nanyang Technological University School of Computer Engineering #N4-02a-32, Nanyang Ave, Singapore 639798*

[3]*Department of Computer Science The Australian National University Acton, ACT 0200, Australia*

Abstract

Reservoir characterization especially well log data analysis plays an important role in petroleum exploration. This is the process used to identify the potential for oil production at a given source. In recent years, support vector machines (SVMs) have gained much attention as a result of its strong theoretical background. SVM is based on statistical learning theory known as the Vapnik-Chervonenkis theory. The theory has a strong mathematical foundation for dependencies estimation and predictive learning from finite data sets. This paper presents investigation on the use of SVM in reservoir characterization. Initial results show that SVM can be an alternative intelligent technique for reservoir characterization.

1. Introduction

Well logging plays an essential role in the determination of the production potential of a hydrocarbon reservoir [1]. It is a geophysical prospecting technique that has been in use since 1927. The process involves lowering a number of instruments into a borehole with the purpose of collecting data at different depth intervals. The measurements broadly fall into three categories: electrical, nuclear and acoustic. A log analyst is one who interprets the data with an objective to translate the log data into petrophysical parameters of the well. To obtain an accurate picture of the important petrophysical parameters, extensive analysis of the core has to be carried out. This will provide answers to questions on the petrophysical properties of the particular borehole such as lithology, porosity, amount of clay, grain size, water saturation, permeability and many others. All these answers are essential to the evaluation of the reservoir formation [2]. One of the key issues in reservoir evaluation using well log data is the prediction of petrophysical properties such as porosity and permeability. Over the life of the reservoir, many crucial decisions depend on the ability to accurately estimate the formation permeability and porosity. However, the prediction of such properties is complex, as the measurement sites available are limited to isolated well locations.

Although core data obtained from the detailed laboratory analysis are deemed to be most accurate, the analysis process is an expensive and lengthy exercise. Usually, limited core data are available at certain intervals. They are used as the basis to establish an interpretation model for other zones with similar log responses. Ideally, the model could be used to interpret log data from wells within the neighbouring region without the need to carry out further core analysis. This requires an integrated knowledge of the tool responses and understanding of the geology of the region, together with various mathematical techniques in order to derive an interpretation model which relates the log data to the petrophysical properties. However, the establishment of an accurate well log interpretation model is not an easy task due to the complexity of different factors that influence the log responses.

In order to perform a reasonable petrophysical properties determination, log analysts have to perform some form of initial preprocessing on the raw data. The preprocessing involved is normally similar to those used for the correction of environmental effects, used to flag special minerals, used to correct resistivity

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 **IEEE**

logs for invasion and so on [2]. For multi-well analysis, further preprocessing such as recalibrating the logs is also required.

A large number of techniques have been introduced in order to establish an adequate interpretation model over the past fifty years [3]. The way that petrophysical properties determination is carried out has also changed considerably over the years due to the development in logging tools and methodologies. The analysis process has also undergone substantial changes due to the development and understanding of the physics of porous media and the rapid development of computer technology. In the past decade, beside the conventional empirical and statistical techniques, another technique that has emerged as an option for predicting petrophysical properties is the Artificial Neural Network (ANN). Research has shown that an ANN can provide an alternative approach to predicting petrophysical properties with improvement over the traditional methods [4, 5]. Most of the ANN based petrophysical properties determination models have used the Multi-layer Neural Network (MLNN) utilizing the backpropagation learning algorithm [6, 7, 8]. Such networks are commonly known as Backpropagation Neural Networks (BPNNs). A BPNN is suited to this application, as it resembles the characteristics of regression analysis in statistical approaches. ANNs perform analysis in a fundamentally different way from the traditional empirical and statistical approaches. ANNs can be used to address most of the mentioned factors that could possibly affect the accuracy of the model. An ANN does not require a prior assumption of the functional form of the dependency. It also offers a numerical model free of estimators and dynamic systems. In addition, an ANN possesses the capability to model complex nonlinear processes with acceptable accuracy and has the ability to reject noise.

Beside applications that use BPNN directly, there are some applications where other techniques are used to enhance the performance of the BPNN. For example, Arpat [9] proposed using the neighboring log data point relations to predict petrophysical properties with only limited core. Fung et. al [10] make use of Self-Organizing Map (SOM) and Learning Vector Quantization (LVQ) to identify the electrofacies and then build a BPNN for each electrofacies for predicting petrophysical properties. Wong [11] makes use of adjacent core data using an improved windowing technique such that the scales of the well log and core are matched.

In recent years, another machine learning approach, support vector machines (SVMs) have gained much attention as a result of its strong theoretical background based on statistical learning theory. [12]. The Vapnik-Chervonenkis theory has a strong mathematical foundation for dependencies estimation and predictive learning from finite data sets. The objective of the SVM is to minimize both the empirical risk and the complexity of the model, thus enabling high generalization abilities. This paper investigates the use of SVM for reservoir characterization and proposes it as an alternative intelligent for reservoir characterization.

2. Petrophysical properties determination model

The petrophysical properties determination problem in reservoir characterization falls into the category of function approximation problems. In function approximation, the objective is to build a model to represent the relationship between the input well logs x and the core petrophysical property y without any assumed prior parameters. Given the well logs vector X and the petrophysical property vector Y, the following expression can be used to describe the relationship:

$$Y = g(X) \qquad (1)$$

When obtaining the training set, there will be some environmental factors that affect the measurements. In well log analysis, these could be due to the mud used, the logging instruments used, the lab technician errors etc. Therefore it is not possible to define an exact function, $g(\,)$, that describes the relationship between X and Y. However, a probabilistic relationship governed by a joint probability law $P(\nu)$ can be used to describe the relative frequency of occurrence of vector pairs (X_n, Y_n) for n training patterns. The joint probability law $P(\nu)$ can be further separated into an environmental probability law $P(\mu)$ and a conditional probability law $P(\gamma)$. For notation expression, the probability law is expressed as:

$$P(\nu) = P(\mu)P(\gamma) \qquad (2)$$

The environmental probability law $P(\mu)$ describes the occurrence of the input well logs X. The conditional probability law $P(\gamma)$ describes the occurrence of the petrophysical properties Y based on the given input well logs X. A vector pair (X, Y) is considered as noise if X does not follow the environmental probability law $P(\mu)$, or the output Y based on the given X does not follow the conditional probability law $P(\gamma)$.

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

From (1), the relationship *g(X)* based on the available training set can be assumed to be analogous to the conditional probability law *P(γ)*. Therefore, it is the role of estimating *P(γ)* that any determination model is performing. It can also be denoted as *E(Y|X)* as the expectation of *Y* given *X*. Therefore:

$$g(X) = E(Y \mid X) \qquad (3)$$

In most models, *g(X)* cannot be obtained directly from the training set (X_n, Y_n). Models have to undergo certain training processes in realizing the best *g(X)*. Normally the best *g(X)* will be an approximation of the function including some error:

$$g(X) = E(Y \mid X) + \theta \qquad (4)$$

where θ denotes the error.

The generalization ability of the determination model is the most important feature in most practical applications. It is used to measure how close the final model *g(X)* is to the expected model *E(Y|X)*. As the realization of the best-fit model is dependent on the available training data, it is also regarded as a measure of how well the model can provide reasonable predictions from 'unseen' input logs other than the training data set.

3. Support vector machines for regression

The Support Vector Machines (SVM) [12], derived from Vapnik's statistical learning theory has become a popular technique among machine learning models. These algorithms create a sparse decision function expansion by choosing only a selected number of training points, known as support vectors. Through the use of kernel, linear function approximation algorithms involving explicit inner products between data points in an input space can be conveniently and efficiently transformed into their nonlinear generalizations. SVMs approximately implement Vapnik's structural risk minimization principle through a balanced tradeoff between empirical error (risk) and model complexity (measured through the VC dimension).

We consider the problem of SVM regression modeling given observational data of the form $(x_i, y_i)_{i=1}^{l}$ where $x_i \in \Re^p$ denotes the input and y_i as a real valued target. SVM seeks to model the relationship between the inputs and the output. Assume that the functional form that SVM is seeking is the familiar linear function, $f(x,w,b) = \langle w, x \rangle + b$, where $w \in \Re^p$, denotes a p dimensional vector of unknown coefficients and $b \in \Re$ is an unknown but constant bias term. Then it tries to find w, b such that empirical risk $\Re_{emp}$ is minimized; simultaneously, it tries to minimize the L_2 norm of the weight vector w for capacity control. Formally, the following basic convex programming problem is posed as:

Minimize:

$$(1/2)\langle w, w \rangle$$

subject to constraints:

$$y_i - \langle w, x_i \rangle - b \le \varepsilon_i$$
$$\langle w, x_i \rangle + b - y_i \le \varepsilon_i$$
$$(5)$$

Since a feasible solution may not exist satisfying the above optimization problem (or we may want to tolerate some noise), it is necessary to introduce slack variables $\xi_i, i = 1, \cdots, l$ to relax the constraints in the original optimization problem. An equivalent optimization problem with quadratic penalization on ξ_i s can be formulated as follows:

Minimize

$$F(\xi) = (1/2)\langle w, w \rangle + (C/2)\sum_{i=1}^{l} (\xi_i)^2 ,$$

subject to the constraints:

$$y_i(\langle w, x_i \rangle + b) \ge 1 - \xi_i \qquad (6)$$
$$\xi_i \ge 0$$

The desired weight vector has the form: $w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i$, where α_i , α_i^* are non-negative Lagrange multipliers required to solve the above optimization problem. The parameter C measures a *trade-off* between empirical error and model complexity and is usually set *a priori* (through cross validation, for example). A nonlinear generalization is effected by simply noting that the resulting solution $f(x)$ can be explicitly written in terms of inner products between data points; these inner products are then replaced by a Mercer kernel $k(x, x_i)$ and the resulting solution has the form:

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

$$f(x)=\sum_{i=1}^{l}\left(\alpha_i-\alpha_i^*\right)k(x,x_i)+b$$

(7)

4. Case study and discussions

The data set for this study is obtained from a real reservoir located in the North West Shelf, offshore Western Australia. The training well has a total of 112 data points and the testing well has a total of 158 data. The well logs available are: GR (gamma ray), RDEV (deep resistivity), RMEV (shallow resistivity), RXO (flushed zone resistivity), RHOB (bulk density), NPHI (neutron porosity), PEF (photoelectric factor) and DT (sonic travel time). The petrophysical property that needs to be determined is Phi (porosity). As the reservoir is heterogeneous, no depth information is used in determining the porosity.

For comparison, two determination models are constructed. In the first model, we construct a porosity determination model using the guassian radial basis function SVM as discussed in Section III. In our experiments, we used the ε-SVR regression machine from the LIBSVM library [13] of support vector machine techniques. The tradeoff parameters in the SVM regression scheme were based on the recommended defaults. In the next model, we construct a BPNN determination model. The BPNN configuration is 8-16-1. BPNN is chosen for comparison as it is the most commonly used intelligent technique for reservoir characterization.

The results in the form of mean square errors (M.S.E.) of the test well are presented in Table 1. Two sets of mean square errors are presented in the table. The normalized M.S.E.s are the errors based on the errors calculated when the data are normalized between 0 and 1. The raw M.S.E.s are errors measured in the actual range of the porosity. The results plot of the porosity determination models for the test well is presented in Figure 1.

Table 1: Comparison results for the porosity determination models.

Determination Model	**M.S.E. (Normalised)**	**M.S.E. (Raw)**
SVM	0.026347	23.7126
BPNN	0.074090	66.6817

From Table 1, the BPNN model is less accurate as compare to the SVM. This is mainly due to the techniques used to ensure the generalization capability of the BPNN. As the wells used in the case study are from a real world reservoir and it is noisy and heterogeneous, the accuracy of the prediction using BPNN depends very much on the generalization ability of the determination model. Hence, one way to obtain improvement in the accuracy would be to integrate further forms of preprocessing and postprocessing in the BPNN model [14]. However, the trade-off is of course an increase in computational time and complexity.

When using SVM, our preliminary accuracy results show that it performed much better than the model uses random 2-fold cross validations. It is also worth noting that the SVM used in this investigation does not incorporate any optimization or cross-validation process to define the tradeoff parameter in the SVM regression scheme that reflects the balance between model complexity and empirical errors. This is motivated by our curiosity on the competency of raw SVM. A list of other kernels such as linear, polynomial, or sigmoid may be employed to improve the technique, and will be carried on in our further investigations. It is worth noting that even with the default parameters, the SVM can perform much better than BPNN. This shows that SVM could be a good alternative intelligent technique for reservoir characterization.

5. Conclusion

A petrophysical properties determination model based on the use of SVM is investigated in this paper. SVM has gained much attention in machine learning circles as a result of its strong theoretical foundation for dependencies estimation and predictive learning from finite data sets. The preliminary results have been reported and compared to the BPNN for petrophysical properties determination model. In the construction of the determination models, the generalization capability of the model poses as a critical factor on the prediction accuracy. Our empirical results show that SVM can produce promising results to the BPNN model, which is used commonly as the intelligent technique in reservoir characterization. It is also worth noting the default parameters of the SVM are used in this test. SVMs provide a generic mechanism to fit the surface of the hyper plane to the data through the use of a kernel function. In this initial study, we have used the SVM based on the commonly preferred gaussian radial basis function kernel. In most SVM research, different kernels are found work best in different applications. A list of other kernels such as linear, polynomial, or sigmoid may be employed. The choice of kernels is an old question but remains to be open as it is often problem dependent. Hence, there is much evidence to warrant further investigations on the best choice of

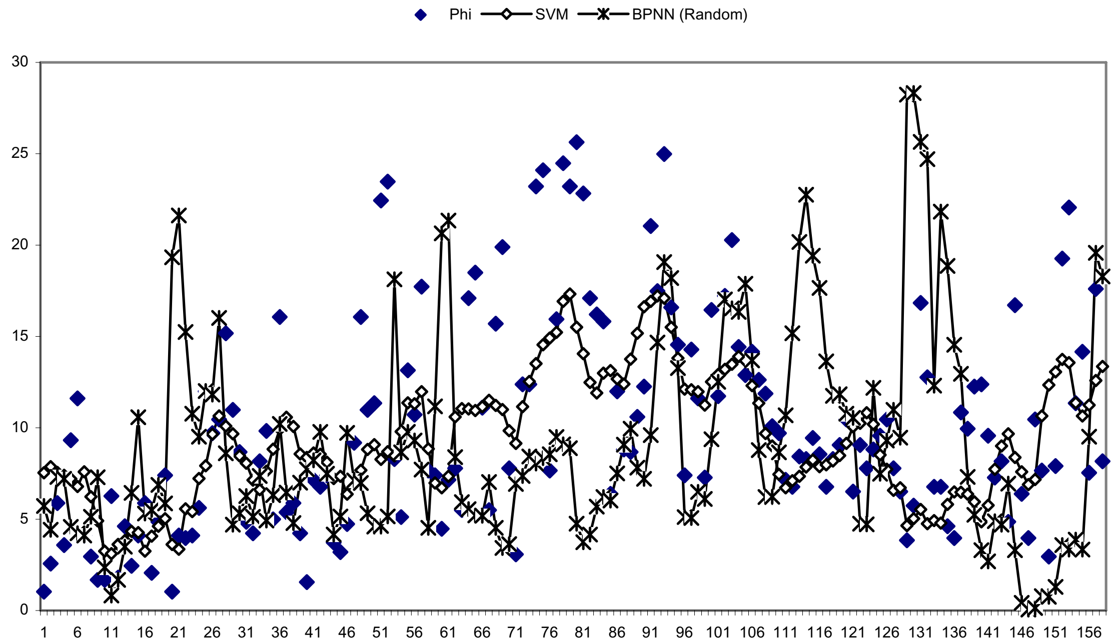

Figure 1: Results plot of the different porosity determination models for the test well.

kernels for accurate petrophysical properties prediction. A probable other approach may involve adaptive selection of kernels via some intelligent means during SVM training. Our future work will be to explore SVM determination models used for reservoir characterization and to search for a set of guiding conditions where SVMs will work best in this application domain. This will introduce SVMs to be used as an alternative method for petrophysical properties prediction in addition to existing intelligent approaches.

.

6. References

[1] D.V. Ellis, Well Logging for Earth Scientists, Elsevier Science Publishing Co. 1987.

[2] M. Rider, The Geological Interpretation of Well Logs, Second Edition, Whittles Publishing, 1996.

[3] B. Balan, S. Mohaghegh, S., and S. Ameri, "State-Of-The-Art in Permeability Determination From Well Log Data: Part 1 - A Comparative Study, Model Development," *SPE Technical Report 30978*, 1995.

[4] D.A. Osborne, "Neural Networks Provide More Accurate Reservoir Permeability", *Oil and Gas Journal*, 28, 1992, pp. 80-83.

[5] P.M. Wong, F.X. Jian, and I.J. Taggart, "A Critical Comparison of Neural Networks and Discriminant Analysis in Lithofacies, Porosity and Permeability Predictions," *Journal of Petroleum Geology*, vol. 18(2), 1995, pp. 191-206.

[6] 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*, 16 (4), 1996, pp. 263-274.

[7] Z. Huang, J. Shimeld, M. Williamson, and J. Katsube, "Permeability Prediction with Artificial Neural Network Modeling in the Venture Gas Field, Offshore Eastern Canada," *Geophysics*, 61(2), 1996, pp. 422-436.

[8] H. Crocker, C.C. Fung, and K.W. Wong, "The Stag Oil Field Formation Evaluation: A Neural Network Approach," *The APPEA Journal*, 39, 1999, pp. 451-459.

[9] G.B. Arpat, "Prediction of Permeability from Wire-line Logs Using Artificial Neural Networks," *Proceedings of SPE Annual Technical Conference and Exhibition v Omega n Part 2*, 1997, pp. 531-538.

[10] C.C. Fung, K.W. Wong, H. Eren, R. Charlebois, and H. Crocker, "Modular Artificial Neural Network for Prediction of Petrophysical Properties from Well Log

Data," *IEEE Transactions on Instrumentation & Measurement, 46(6)*, 1997, pp. 1259-1263.

[11] P.M. Wong, "Permeability Prediction from Well Logs Using An Improved Windowing Technique," *Journal of Petroleum Geology*, 22(2), 1999, pp. 215-226.

[12] V. Vapnik, Statistical Learning Theory, Wiley, 1998.

[13] C.C. Chang and C.J. Lin, LIBSVM: a library for support vector machines, 2001, Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[14] K.W. Wong, T.D. Gedeon, and C.C. Fung, "The Use of Soft Computing Techniques as Data Preprocessing and Postprocessing in Permeability Determination from Well Log Data," in Wong, P.M., Aminzadeh, F., and Nikravesh, M. (Eds.) *Soft Computing for Reservoir Characterisation and Modeling*, Studies in Fuzziness and Soft Computing, Physica-Verlag, Springer-Verlag, 2002, pp. 243-271.

0-7695-2504-0/05 $20.00 © 2005 IEEE