
Optimizing of Support Vector Machine its Application in Oil And Gas Reservoir Classification Prediction

Xinsen Tian ^a, Huaping Yu ^b

College of Computer Science, Yangtze University, Jingzhou Hubei 434023, China

^atianxinsen@qq.com, ^byhjz@126.com

Abstract

Oil and gas reservoir classification prediction experiments, less effective samples, related factor is various reasons largely affect the accuracy of predicted results in laboratory and fitness. Support vector machine (SVM) is a kind of novel machine learning method based on small samples statistical. In this paper, we using LIBSVM toolbox of MATLAB to realize the model for oil and gas reservoir classification experiments. At the same time, this paper takes changqing oilfield logging data as test samples, through data visualization method and fractal dimension visualization (Box dimension method and Fractal dimension visual figure) parameter accuracy verification experiment, through the comparison of different kernel functions optimization experiment model parameters, through the heuristic algorithm (genetic algorithm and particle swarm algorithm) to optimize parameters of test model. Finally come to the conclusion through the experiment, for small samples of oil and gas reservoir classification support vector machine (SVM), with Radial basis kernel function as the kernel function is set up and through the genetic algorithm to optimize parameters of support vector machine (SVM) model can significantly improve the effectiveness and accuracy of the models.

Keywords

Oil and gas reservoir prediction, Parameter optimization, Support vector machine, Particle Swarm Optimization, Genetic Algorithm, Kernel function.

1. Introduction

The present oil enterprise, petroleum geology reservoir classification prediction in oil exploration and occupy an important position in high efficiency mining link. Logging is currently for petroleum geological exploration reservoir classification prediction sample data is the most effective means of exploration. Logging exploration this method usually refers to the well drilling a selected area detection, measuring the parameters of each sequence data collection, and to predict the sample analysis with regional stratigraphic categories. This method usually need to rely heavily on the expert's experience and background, at the same time work tedious time consuming, and there are a large number of not directional prediction results. Therefore, always cannot achieve extensive applicability in the oil and gas exploration industry. At the same time, each Wells in the oil and gas exploitation enterprise is not cheap, so precious exploration data, the effective samples are few and far between. Considering the whole process of reservoir classification prediction is a kind of very typical pattern recognition mathematical geology classification type, and oil industry factor many problems such as low effective samples, laboratory, classification of reservoir prediction in the process of information modeling method is a feasible laboratory.

Support Vector those(SVM) based on statistical learning theory (SLT) as the theoretical basis for development, and can be seen as a special case of machine learning problems based on data, based on

machine learning method has an important role in the current recognition of small sample. , therefore, will use the support vector machine (SVM) is applied to the petroleum geology reservoir classification prediction direction, can greatly improve reduce laboratory identification of forecasting process, and achieved a higher accuracy. In this experiment, through the changqing oil field actual project data experiment, obtained good experimental results.

2. The basic principle of support vector machine

SVM (Support Vector those, SVM) is thus Corinna ms Cortes and Vapnik is first proposed in 1995 a solving classification and regression problems of a new generation of universal learning machine. The SVM based on statistical learning theory (statistical learning, SLT) as the theoretical basis for development and can be regarded as a machine learning problems based on the data of a special case. Based on machine learning method has an important role in the current recognition of small sample, this research method refers to the study in the sample can't get through the analysis of the principle of the current law, the rule of how to use the machine learning method to analyze the objective object, to estimate and predict the unknown data. This new type of machine learning algorithms, classification and regression problems like multi-layer perceptron network and radial basis function (RBF) network, the technology has excellent learning performance, has now become a hotspot in the field of machine learning research, and in pattern recognition, regression estimation, probability density function is estimated, and other areas of the many successful applications.

Support vector machine (SVM) is a kind of pattern classification methods. The mechanism of this kind of classification method is: given a training set , the .support vector machine (SVM) need to be under the condition of the corresponding search the optimal solution to the problem. C - SVC model is more common binary classification support vector machine (SVM) model, the specific form is as follows:

1) Setting the known training set:

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l \quad (1)$$

Among equation (1), $x_i \in X = R^n$, $y_i \in Y = \{1, -1\} (i = 1, 2, \dots, l)$, x_i is the feature vector.

2) Selecting the proper kernel function $k(x, x')$ and parameter C , and then solving the optimization problem:

$$\min_a \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^l \partial_j, \quad \text{s.t.} \quad \sum_{i=1}^l y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, \dots, l \quad (2)$$

The optimum solution is $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$.

3) Selecting a positive component ($0 < \alpha_j^* < C$) of α^* is, and then calculating the threshold:

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i - x_j) \quad (3)$$

4) Structuring the decision function:

$$f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^*) \quad (4)$$

Support vector machine (SVM) is the core idea of a support vector institutions to build a data maximum separation hyperplane as decision surface, the vector is mapped to a high-dimensional space, namely vector X_i by function is mapped to a high-dimensional space, set up in the right direction of the separating hyperplane that two parallel maximum distance between the hyperplane. It is assumed to be, or the bigger the gap, the distance between parallel hyperplanes, classifier of total error is smaller; Support vector machine (SVM) based on the theory of statistics, which follow the structural risk minimization (SRM) principle of statistics, in order to experience and confidence risks while minimizing, make it has strong generalization ability. $C > 0$ is constant control punishment program.as the kernel function. The most common form of four basic kernel function is as follows:

Linear kernel function: $K(x, x_i) = x^T x_i$ (5)

Polynomial kernel function: $K(x, x_i) = (\gamma x^T x_i + r)^p, \gamma > 0$ (6)

Radial basis kernel function: $K(x, x_i) = \exp(-\gamma \|x - x_i\|^2), \gamma > 0$ (7)

Sigmoid kernel function: $K(x, x_i) = \tanh(\gamma x^T x_i + r)$ (8)

3. The experimental application

3.1 The raw data processing

For a complete set of effective sample data of oil field exploration, the property factor is various, discrete data is higher. Thus to predict before the official start of the lab, need to normalize data processing. Raw data to deal with the main purpose of this process there are two things: (1) excluding for petroleum geology reservoir classification prediction in the process of low correlation factor of laboratory. (2) to zoom the data, make the data more able to adapt to data processing of the model. Model used in this experiment is based on the predictions of a library of libsvm algorithm, thus eliminate low correlation between laboratory factors can improve the efficiency and accuracy of model prediction.

Because no one complete specification down logging data format, So you need to artificial data analysis, according to the oil industry background, this experiment selects the model reference attribute is 6 kinds, respectively is: shale content (Sh), acoustic time (Ac), oil saturation(So), permeability (permeability), Porosity (Porosity), the resistivity of the rock (Rt); At the same time, the oil and gas reservoir prediction results are divided into 5 types, respectively: reservoir, reservoir, oil and water, tree oil water layer, dry layer. In order to process the data of the results more intuitive and accurate, this experiment using the box dimension method and fractal dimension visual figure combination of the two ways for validation. The raw data used is 490 groups of changqing oilfield data of oil and gas, the capacity of 490 groups of sample data, the experimental results the figure below:

(1)The box dimension method ,see Fig.1.

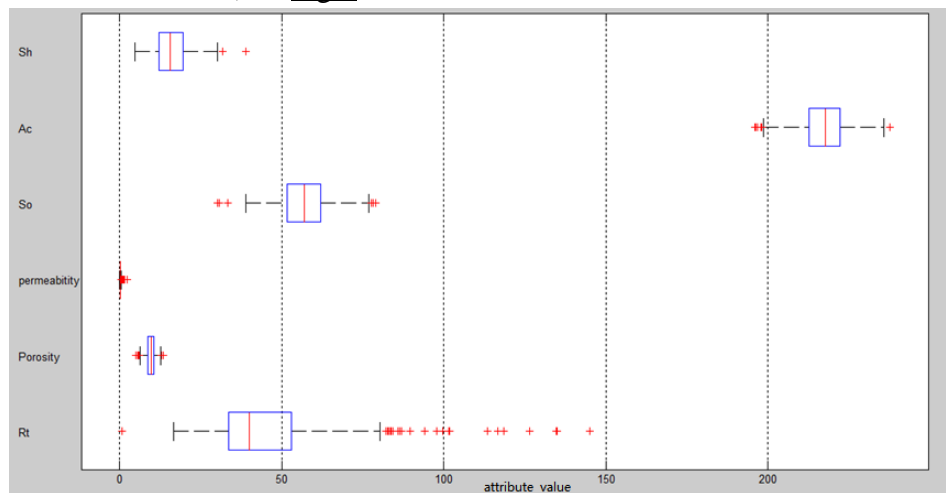


Fig.1 The box dimension method of the sample

(2)Fractal dimension visual figure,see Fig.2.

By comparison with the results found that the experimental sample data distribution in most experiments, decided to adopt the six properties of reservoir evaluation factors. Therefore, this experiment sample space is 6 d, is divided into five types. The support vector machine (SVM) in this experiment: six respectively reference attributes for shale content (Sh), acoustic time difference (Ac), the oil saturation (So), permeability, permeability and pore (Porosity), the resistivity of the rock (Rt). The prediction results of support vector machines (SVM) for five, namely oil reservoir with five oil level (reservoir, reservoir, oil and water, tree oil layer, dry layer).

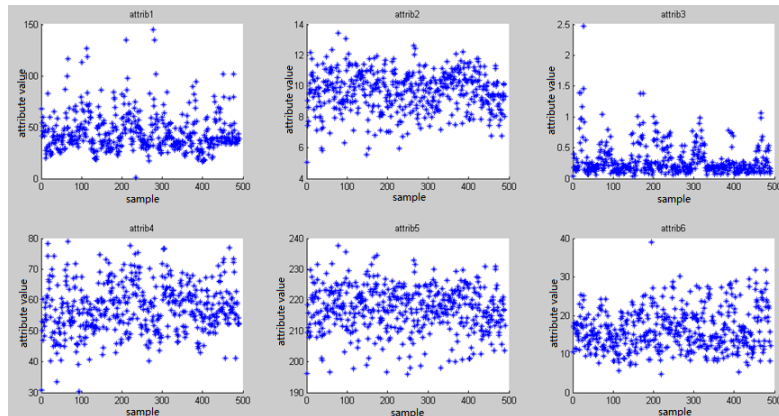


Fig.2 Fractal dimension visual figure of the sample

3.2 The choice of kernel function and determined

In the experiment of the SVM model, the selection of kernel function is largely determines the efficiency and accuracy of the model. Kernel function of SVM is divided into basic Linear kernel function and Polynomial kernel function, the Radial basis kernel function, the Sigmoid kernel function of 4 kinds of kernel function. Because data complexity is higher, at the same time as the parameter dimension is higher, therefore, the experiment selected Polynomial kernel function and Radial basis kernel function of two kinds of kernel function to experiment. The experimental results are as follows: Polynomial kernel function, see Fig.3.

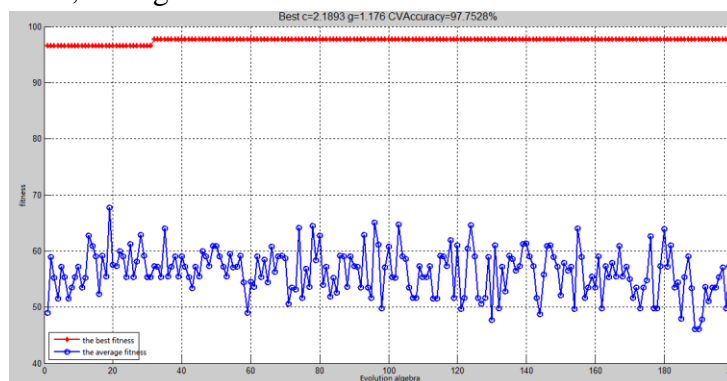


Fig.3 Based on the Polynomial kernel function of support vector machine (SVM) model fitness distribution

Radial basis kernel function, see Fig.4.

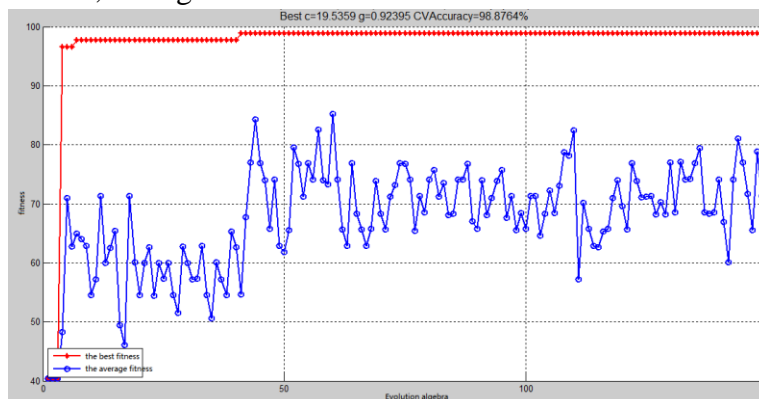


Fig.4 Based on the Radial basis kernel function of support vector machine (SVM) model fitness distribution

By comparison with the experimental data analysis, found that in selecting Radial basis kernel function for the kernel function model, with an average algebra, namely the increase of sample size, the best

fitness than choose Polynomial kernel function for the model when the kernel function is high, and better prediction results. Therefore, in to the prediction of reservoir classification model, experimental proof use Radial basis kernel function as the kernel function model of effectiveness is higher.

3.3 Support vector machine (SVM) parameters optimization

Article Swarm Optimization (PSO) is made by j. Kennedy and r. c. Eberhart originates from the behavior study of birds feed on, in 1995 developed a new Evolutionary Algorithm (Evolutionary Algorithm - EA). The algorithm is mainly refers to the starting from the random solution of initialization, through iteration to find the global optimal solution. On the multidimensional, relevance C&G oil complex data samples of SVM parameters, the PSO algorithm can seek through multiple iterations to the global optimal parameter set.

Based Algorithm is a kind of reference to the evolution rule of biology, by American J.H olland first proposed in 1975, a professor at random search method. Conceptual optimization method, GA can automatically acquire and to guide the optimization of search space, adaptively adjust the search direction, do not need to make sure the rules. On several uncertain item C&G solution of SVM parameters, GA also has better global optimization ability.

This experiment in the kernel function is selected as the Radial basis kernel function conditions, by PSO algorithm and GA algorithm optimizing the parameters of support vector machine (SVM) C&G respectively, and the PSO - SVM and GA - SVM experimental comparison, the experimental results are as follows:

Particle Swarm Optimization, see Fig.5.

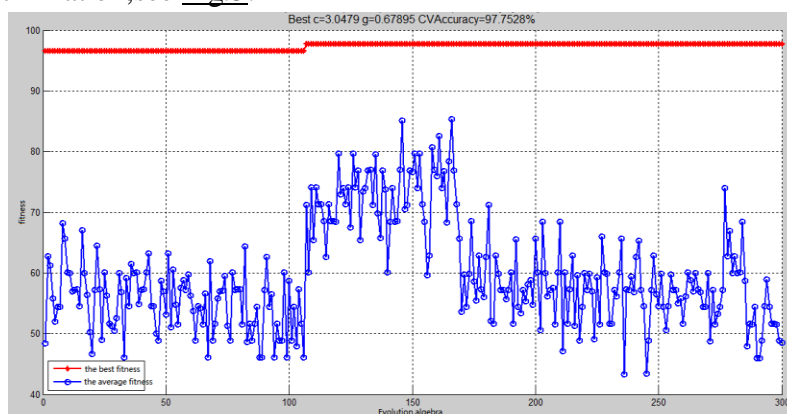


Fig.5 Based on Particle Swarm Optimization to optimize the parameters of support vector machine (SVM) model fitness distribution

Genetic Algorithm, see Fig.6.

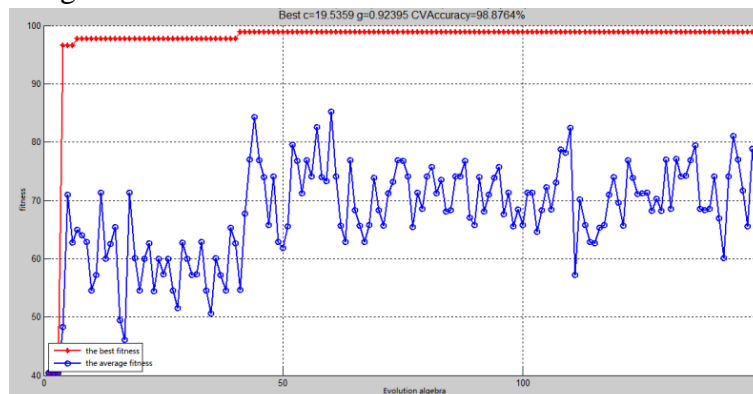


Fig.6 Based on Genetic Algorithm to optimize the parameters of support vector machine (SVM) model fitness distribution

By analyzing the results of the experiment, the experiment of genetic algorithm to optimize the SVM can get a higher fitness. Namely in changqing oil field oriented data, GA - SVM compared PSO - SVM has higher accuracy. At the same time, the experiment proved in the experiment of reservoir classification and prediction of practice, through the genetic algorithm to optimize the SVM has the feasibility and effectiveness.

4. Conclusion

This experiment in changqing oil field mining area of 490 groups of the same oil and gas data as sample data, and it can be divided into training set and testing set as sample data simulation model. Through the experimental results show that the classification for oil and gas reservoir prediction of this kind of correlation factor more, less effective samples prediction of small sample classification problem, support vector machine (SVM) model in oil and gas reservoir recognition experiment has a higher accuracy. At the same time, a support vector machine (SVM) to establish the simulation model is applied to predict oil and gas reservoir classification method has good adaptability and effectiveness. On a support vector machine (SVM) model in the oil and gas reservoir classification optimization, the SVM model are optimized by the genetic algorithms (GA - SVM) compared to the particle swarm algorithm to optimize the SVM model (PSO - SVM) has a higher fitness and accuracy.

Acknowledgements

The research is supported by College Students' Innovative Entrepreneurial Training Project of Yangtze University (Grant No. 2014025).

References

- [1] Cortes C and Vapnik V: Support-Vector network, Machine Learning, Vol. 20 (1995), p. 273-297.
- [2] Boser B, Guyon I. and Vapnik V: A training algorithm for optimal margin classifiers, ACM press: In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 1992.
- [3] HSU C W and LIN C J. : A simple decomposition method for support vector machines, Machine Learning, Vol.46 (2002), No. 1, p. 291-314.
- [4] Tran Quang-anh, ZHANG Qian-li and LI Xing: SVM classification-based intrusion detection system, Journal of china institute of communications, Vol. 23 (2002), No.5, p. 51-56.
- [5] Qingjie Zhu etc : Application of artificial neural network model in oil resources, Resource investigation and environment, Vol.23, No.4, 2002.
- [6] R.Rifkin, M.Pontil, and A.Verr : A note on support vector machine degeneracy, Proceedings of 10th International Conference on Algorithmic Learning Theory, (Lecture Notes in Artificial Intelligence Vol.1720), pp.252-263, 1999.
- [7] Xuegong Zhang : About the statistical learning theory and support vector machine (SVM), Acta automatic, Vol.26, 2000.1.
- [8] Zhaoqi Bian etc : Pattern recognition [M]. Version 2. Beijing: Tsinghua university press, 2004. 1.1.
- [9] Yu Zhao , Zhiliang Wang , Jiwei Liu : A kind of new logging lithology prediction method based on support vector machine (SVM), Science and Technology & Innovation , TP391 , 2004, 20(6).