



Applications of support vector machines in the exploratory phase of petroleum and natural gas: a survey

Yuri Ccoicca

Doctorate of Systems Engineering, Graduate Section of Industrial and Systems Engineering, National University of Engineering
Av. Túpac Amaru 210 - Rímac / Lima 25 - Perú
Email: jaccoiccap@uni.pe

Abstract

This paper presents an overview of support vector machines (SVM) as one of the most promising intelligent techniques for data analysis found in the published literature, as theoretical approaches and sophisticated applications developed for various research areas and problem domains. This work is an attempt to provide a survey of the applications of SVM for oil and gas exploration to professionals, researchers and academics involved with the hydrocarbons industry. The applications of SVM have been grouped and summarized in the different areas of the exploration phase, which can be used as a guide to assess the effectiveness of SVM over other data mining algorithms. It also provides a better understanding of the various applications that have been developed for an area that offers a glimpse of innovative applications in other domains of the industry.

Keywords: SVM, data mining, machine learning, hydrocarbon, petroleum and gas natural exploration.

1 Introduction

The exploration phase plays an important role in the industry of oil and natural gas. If successful, it will trigger all other phases: production, refining, etc.

It is widely known that there are a wide number of limitations in the traditional engineering methods, which are typically called knowledge-oriented, that is, depending greatly on the degree of the practitioner's knowledge. Unlike the above methods, the technique of data mining and especially machine learning, which are considered data-oriented, face a lot of data and never overlook any phenomenon of character that is unimportant, discovering hidden patterns and relationships in order to create successful predictive models. Data mining is an emerging domain that is based on analysis of historical data in order to ensure that the amount of available data is directly proportional to the quality of knowledge, derived from these data.

For the hydrocarbon industry, its problems are too complex to rely only in one discipline to solve them in the most effective way, which is why multidisciplinary approaches in this industry are becoming more frequent. Staying in the classical disciplines, such as the characterization of oil and gas reservoirs, geology, geophysics, geochemistry and drilling, as currently happens, will not lead to properly manage data of heterogeneous nature, already existing in the exploratory phase of oil and gas.

Since the prediction based on insufficient training samples produced unsatisfactory results, Vapnik et al., [55] in the decade of the 90s, proposed the algorithm SVM, which is based on the VC dimension theory of statistical learning and the principle of structural risk minimization. The VC dimension is a measure of the ability of statistical classification algorithms, defined as the cardinality of the largest set of points that the algorithm can separate.

The SVM has generated great expectations in recent years as they have been successful when used in classification problems, regression and forecasting, as they include aspects and techniques from machine learning, statistics, functional analysis and convex optimization. Aside from having a strong adaptability, global optimization, and a good generalization performance, the SVMs are suitable for classification of small samples of data.

Therefore, in this document, the application of the SVM is proposed to improve predictive indicator level, which results

in economic terms as a savings of several million dollars for the oil industry, specifically in the exploration phase, it increases the likelihood that the exploratory well penetrates a producing formation, indicating that the drilling program was successful.

Globally, the application of these techniques in the E&P has come a long way since the first articles began appearing in SPE conferences in early 1990s. From October 30, 2011, the technical section "Data for Action (D2A)", is dedicated to this technology and its advancement. This section was opened at the Annual Technical Conference of the SPE in Denver, USA.

The remainder of the article is organized as follows: the theoretical framework is developed on SVM and hydrocarbon exploration. Below there is a review of the literature and applications developed around the SVM classifier and predictor for the areas of E&P with a strong emphasis on the exploratory phase. The conclusions and references are to be found at the end of the present document.

2 Theoretical framework

SVMs are new discriminating techniques in the theory of the statistical training. For the data processing specialists, the SVM is a linear classifier with broad margin in a space with core. For the statisticians, the SVM is a nonparametric estimator.

SVMs are algorithms based on the three following mathematical principles:

- Principle of Fermat (1638)
- Principle of Lagrange (1788)
- Principle of Kuhn-Tucker (1951)

2.1 Classification SVM (SVC)

The theoretical foundation of the SVM is based on statistical learning theory developed by Vapnik et al, in the late 70s. The SVM model, as it is known today was introduced in the 90s by Vapnik, Boser, Guyon and Cortes, [3] allowing the transition from the theoretical formulation to actual applications of pattern recognition, proving to be a great performance when results were far superior to artificial neural networks (ANN) in handwriting recognition, using pixels as input, it was introduced as a powerful tool to solve classification problems in the disciplines of machine learning and data mining.

To understand the essence of SVM classification and hence the prediction, only need to understand four basic concepts: (i) the separation hyperplane, (ii) the maximum margin hyperplane, (iii) the soft margin and (iv) kernel function.

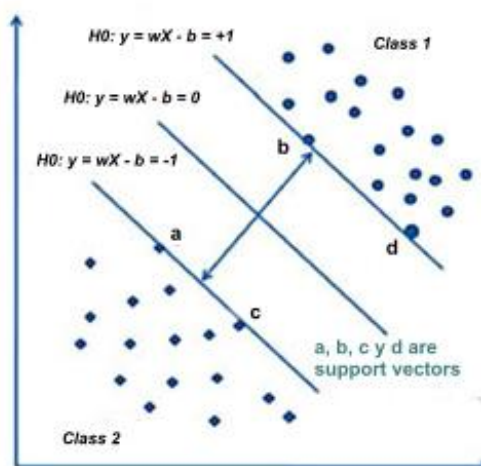


Fig.1: Maximizing the margin of separation

When data cannot be separated linearly, we have to make a change in the space through a mathematical function that transforms the data, so it can be achieved. This function is the kernel, which transforms input data to a multidimensional space (Hilbert's space) where the problem is solved. Finally, the mapping is returned to the original space and the classification is not linear in the input space. The kernel functions may be linear or nonlinear.

Most kernel methods for classification are based in the generalization of the hyperplane classifier to nonlinear cases. [46] The hyperplane binary classifier is a classification technique that finds the best hyperplane that separates the two distinct points in an n-dimensional space. The best separating hyperplane defines a classifier which separates the points of the two classes with more generalization, it's being understood that the greater generalization is obtained with the hyperplane that remains further away from the separating point sets. This distance is called the margin separating hyperplane. More formally, given a training sample $(x_i, y_i), i = 1, \dots, N$ where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$ (class label point), it can be solved:

$$\min \tau(w, b) = 1/2 \|w\|^2$$

$$w \in \mathbb{R}^d, b \in \mathbb{R} \quad \text{with}$$

$$y_i(\langle w, x_i \rangle + b) \geq 1, \quad \forall i=1, \dots, N \tag{1}$$

where τ is the function to be minimized with respect to the coefficient vector w and the constant b , defining the hyperplane that separates the classes with the same greater distance (i.e., with maximum margin). It is possible that this problem has no solution (i.e., that there is no separating hyperplane for a given set of points), so the condition is relaxed allowing some misclassification of points, leaving the problem as:

$$\min \tau(w, b, \xi) = 1/2 \|w\|^2 + C \sum_{i=1}^N \xi_i$$

$$w \in \mathbb{R}^d, b, \xi_i \in \mathbb{R} \quad \text{with}$$

$$\xi_i \geq 0$$

$$y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \quad \forall i=1, \dots, N \tag{2}$$

where, regarding the problem in Equation 1, we introduce slack variables $\xi_i, \forall i = 1, \dots, N$ and the parameter $C > 0$. C is the penalty factor (also called the regularization parameter), defined as a compromise between the empirical risk and confidence, to achieve risk minimization optimization, while ξ is the relaxation factor, which reflects the classification error of the sample.

This is an optimization problem solvable with quadratic programming using the method of Lagrange multipliers. Applying this method, we obtain the dual problem as:

$$\max J(\alpha) = \sum_{i=1}^N \alpha_i - 1/2 \sum_{j,i} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle,$$

$$\alpha_i \in \mathbb{R} \quad \text{with}$$

$$\sum_{i=1}^N \alpha_i y_i = 0 \tag{3}$$

$$\alpha_i \in [0, C], \quad \forall i=1, \dots, N$$

where α_i is the Lagrange multiplier associated with the input space (x_i, y_i) . Solving this problem the classifier can be formulated as:

$$f(x) = \sum_{i=1}^N \alpha_i y_i \langle x_i, x \rangle + b \tag{4}$$

where x is the entry point to be classified. As it is seen, everything is formulated in terms of scalar products of the input vectors. Remarkably, only the entry points that are on the margin have $\alpha_i > 0$ so that the rest can be ignored in the final model.

The SVM is based on making it to define a kernel function $k(x_i, x_j) = (\Phi(x_i), \Phi(x_j))$ where Φ is a function that maps the points of the input space to a feature space of higher dimensionality which is defined in the hyperplane [40]. This process is called Kernel Trick. The mapping Φ need not be explicitly defined, as it can be used as any kernel function k that fulfills Mercer's theorem [46].

A function $k(u, v)$ satisfies Mercer's theorem if for every square integrable function g :

$$\int_{u,v} k(u, v) g(u) g(v) du dv > 0 \tag{5}$$

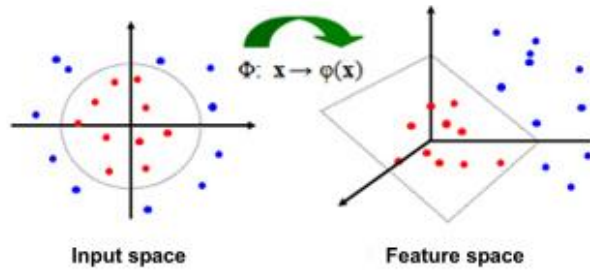


Fig.2: Transformation by a kernel to a high dimensional space to achieve separable classes

Examples of functions that meet this kernel are called linear (eq.6), polynomial (eq.7) and Gaussian (eq.8) (normally used as RBF):

$$k(x_i, x_j) = \langle x_i, x_j \rangle \tag{6}$$

$$k(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^d, d \in \mathbb{N} \tag{7}$$

$$k(x_i, x_j) = \exp\left[-\frac{1}{\sigma^2} \|x_i - x_j\|^2\right] \tag{8}$$

also be implemented as:

$$k(x_i, x_j) = \exp\left[-\lambda \|x_i - x_j\|^2\right] \tag{9}$$

λ parameter plays an important role in the performance of the kernel. It is notable that some kernel functions used in the literature do not fulfill Mercer's theorem, as the sigmoid kernel (known also as MLP, because of its similarity to the Multi-layer neural networks Perceptron (MLP) with signoid activation function):

$$k(x_i, x_j) = \tanh(a^T x_j + offset) \tag{10}$$

2.2 Regression SVM (SVR)

The SVM for regression (SVR) are a modification of a model for classification. Given a set of samples of a function $\{(x_1, y_1), \dots, (x_N, y_N)\} \subset X \times \mathbb{R}$, where X is a set in which there is defined a dot product operation between their elements, regression technique ϵ -SVR looks for a function $f(x)$ as smooth as possible and whose values differ from each actual output y_i no more than ϵ for all $x_i \in X$ (this is known as the error function ϵ -insensitive).

The function $f(x)$ is defined as:

$$f(x) = \langle w, x \rangle + b, w, x \in X, b \in \mathbb{R} \tag{11}$$

As in the case of SVC, the problem may be not resolvable, so that a greater amount of error ϵ to be tolerated. Introducing slack variables ζ_i, ζ_i^* :

$$\min_{w \in X, b, \zeta_i, \zeta_i^* \in \mathbb{R}} \tau(w, b, \zeta_i, \zeta_i^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\zeta_i + \zeta_i^*) \tag{12}$$

subject to:

$$y_i - (\langle w, x_i \rangle + b) \leq \epsilon + \zeta_i$$

$$(\langle w, x_i \rangle + b) - y_i \leq \epsilon + \zeta_i^*$$

$$\zeta_i, \zeta_i^* \geq 0, \forall i = 1 \dots N$$

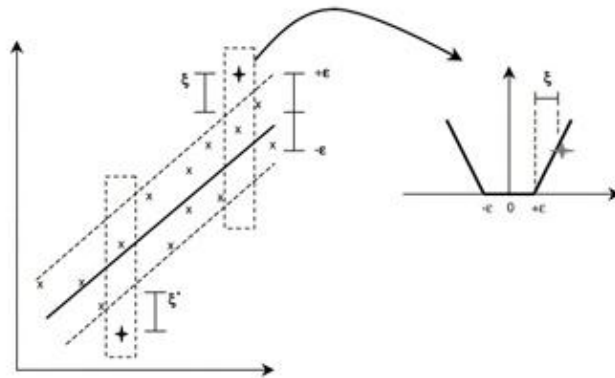


Fig.3: The soft margin lost setting for linear SVR. ζ is positive slack variable for measurements above and below the ϵ tube

The constant $C > 0$ represents a compromise between the smoothness and the number of errors greater than ϵ function. As in the case of the SVM for classification, is an optimization problem solvable quadratic programming using the method of Lagrange multipliers. Passing the dual problem is to solve it only remains to find Lagrange multipliers α_i and coefficient b . The solution (13) is written in terms of the α_i , α_i^* and b , as well as scalar products that can be replaced by a kernel function k .

The regression function is:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) k(x, x_i) + b \quad (13)$$

The major disadvantages of SVR against the SVC are:

- The introduction of an additional parameter ϵ .
- The calculation twice Lagrange multipliers (α_i , α_i^* , ie, 2 for each x_i).
- Not dispersed the resulting models, i.e., all points are used in the model training.

Despite its serious disadvantages compared to SVC (and, as will be seen, and other techniques for kernel regression), the SVR has been successfully employed in some work for time series modeling, for example, predicting loads of electrical networks, [37] or financial series. [52] Nowadays, it is less used taking into account that the LS-SVR gets better results with less computational complexity.

2.3 Optimization of parameters for classification and regression

In the field of pattern recognition and more specifically in its classification; in recent years, SVM has become a machine-learning algorithm with greater relevance than other very popular ones like K -th nearest neighbor (KNN), artificial neural networks (ANN) and classification trees (CART), since its mathematical apparatus is based on very solid foundations [55] that make having multiple advantages over other techniques mentioned. [58], [42] However, Lin et al. in 2003, presents the possibility of perspective, making SVM become the main method of classification (according KDNuggets 2002 Poll, neural networks and classification trees remain as major tools) arguing that the SVM problem was very likely caused by the misuse due to the lack of knowledge of the methodology. What usually unsuspecting users are doing is [<http://www.kernel-machines.org>] converting information to classify some applicative SVM format regardless available in most cases implications format, scaling, etc, and then treat with parameter values randomly and indiscriminately kernels without validation and without knowing in advance that the default settings in these programs are surprisingly important and the fact is that many users get results and generalization error values unsatisfactory.

Table 1 Compilation of some related works on SVM model hyper-parameters

Ref.	Kernel ^a	Método Búsqueda	Función objetivo
19	RBF	Grid-search (GS)	ν -Cross-validation
7	RBF	Gradient descent (GD)	Radius-margin, Span bounds, Leave-one-out
8, 26	RBF	Grid-search (GS)	ν -Cross-validation error (CV)
9	RBF	Genetic algorithm (GA)	Radius-margin bound
10	RBF	Genetic algorithm (GA)	ν -Cross-validation error (CV)
1	RBF	Gradient descent (GD)	Hold out error, radius-margin, Generalized Approximate CV error (GACV)
14	RBF	Gradient descent (GD)	Leave-one-out (LOO), span bound
58	RBF,POL	Grid-search (GS)	ν -Cross-validation
46	POL	Gradient descent (GD)	Generalization error estimation bound
52	RBF	Multi-objective GA (MOGA)	Modified radius-margin bounds
18	RBF	Particle Swarm Optimization (PSO), GS	Hold out error, $\xi\alpha$ -estimator
25	RBF	Uniform design (UD), Grid-search (GS)	ν -Cross-validation
28	RBF	Particle Swarm Optimization (PSO)	False Acceptance (FA)

^a RBF: Radial Basis Function kernel whose hyper-parameter is γ . POL: polynomial kernel which hyper-parameters are the degree u and coefficient r . Kernel hyper-parameters and the regularization parameter C are optimized simultaneously.

The minimum that is expected from the user according to Lin et al, is to scale the data for validation and training, consider the kernel RBF (Radial Basis Function) and find appropriate values for C and σ^2 (or λ). Now the finding of appropriate values is sometimes not an easy task, not to mention that the aim is not to find appropriate values but the best values for a given case. In general terms the Gaussian radial basis function (RBF) can show the best performance in most experiments. But in 2007, Chih-Jen Lin, who developed *LIBSVM*, maintained the view that linear kernel function is the optimal kernel function. However, this view needs more research to prove it.

To date, the hyper parameters selection techniques (the most popular is the Grid Search method) or the selection of the model as it is also called, are shown in Table 1, classified by the kernel, the search method and the objective function.

2.4 Hydrocarbon Exploration

Exploration is the term used in the oil and gas industry to designate the search and recognition of structural or stratigraphic traps where hydrocarbons may have accumulated. The methods used are varied: from the geological study of the rock formations that are flourishing in the observation surface to indirectly through various tools and techniques of exploration. One of the tools used in this stage are the maps. There are outcrops maps (showing the rocks on the surface), topographic maps and maps of the subsoil.

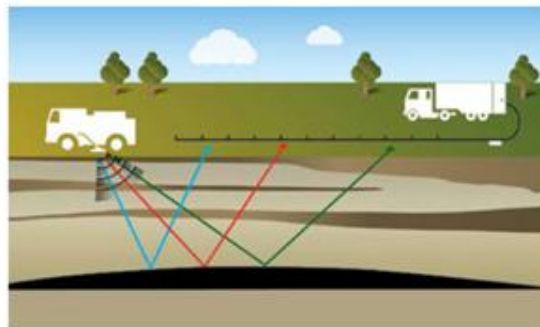


Fig. 4: Seismic reflection

The latter may be the most important because they show the geometry and position of a layer of rock in the subsoil, and are generated by seismic reflection, which is caused by a power source with explosives buried in the ground, -usually 3 to 9 m. deep- or vibrating trucks -these involve a significant reduction in the environmental impact- a front elastic waves travelling through the ground and are reflected at interfaces of the different layers. The surface is covered with an area of high sensitivity such devices also called "geophones" which are joined together and connected by cables to a receiving station. The waves produced by the explosion pass through the groundwater and return to the surface. The geophones capture and send those waves to the seismograph, where by special computer equipment, plotting goes underground. It can measure the time between the moment of the explosion and the arrival of the reflected waves, thereby being able to determine the position of the layers and depth, describing the location of the anticlines favourable for the accumulation of oil and gas. The final product is a representation of the subsoil, either two-dimensional (2D) or three dimensions (3D).

Moreover, the aeromagnetometry and gravimetry, two tools that are used in the early stages of the exploration and for determining the thickness of the sediment layer. Another technique is the geochemistry of the surface consisting of the detection of hydrocarbons in the subsurface accumulated through the measurement of the gas concentrated in the soil samples.

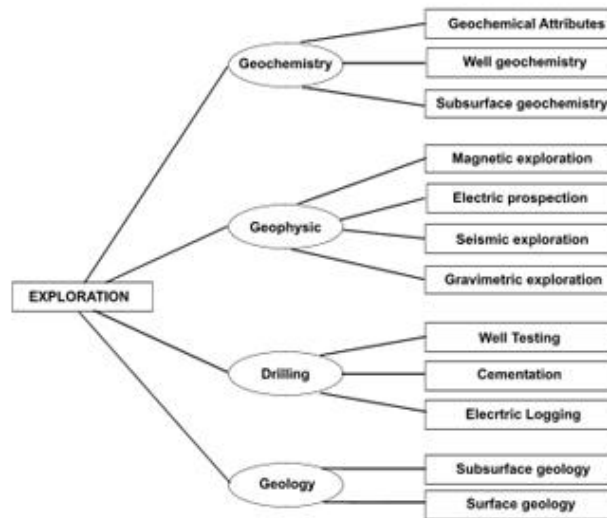


Fig 5: Domain area of hydrocarbon exploration

With the information collected, specialists produce different types of maps of the area examined. Provided information about the thickness, inclination, direction and nature of the strata, not only is useful to decide where to drill exploratory wells, but also to confirm the geological model, and most importantly, the existence of hydrocarbons in these structures. If exploration is successful and has made it through by a commercial discovery with a well, work began bounding the deposit discovered with the drilling of other wells, to perform after the evaluation of reserves. In oil exploration results are not always positive. Often wells are dry or water producers, causing losses of several million dollars, so this activity is considered a high risk investment.

3 Literature Review

Vapnik [56] proposed the SVM as a classification method that combines the use of the computer with the theoretical argument. These features are provided with a great reputation and have promoted their implementation in different areas (facial recognition, OCR, bioinformatics, text mining, time series, etc.). [23], [55]

Since the intelligent learning process of a machine has a number of limitations, SVM has become the most popular technique for data mining at the address offset bias or bias variance [20], the capacity control [22] and the data separation problems in the training samples. The advantage of SVM is that it can be used specifically to mitigate the problems of overfitting and underfitting. Generally, overfitting refers to blindly trust the knowledge base of a machine, while the low setting refers basically iterative access to a part of the knowledge base to make predictions. [44] With respect to support vector machines (SVM), are successfully overcome the limits of machine learning (machine learning) to improve the overall performance of pattern recognition learning. [4]

Due to its unique characteristics and appropriate, the SVM are growing in acceptance in the hydrocarbon industry. These are used individually or in combination with other intelligent techniques. Using kernels, it is possible to implicitly map the characteristics of the areas of high dimensional functions. The selection of kernel parameters in SVM is an important point for researchers. [48], [33], [7], [12] For example, Yao et al., Simulated kernels and successfully applied to a problem kernel of polynomial prediction hydrocarbons.

Application of kernel Fisher discriminant for predicting oil and gas, with the emergence of SVM gave good results. [62] The following year, SVM method was applied for the identification of sedimentary facies. [63] The results also showed that SVM in the case of small samples, accuracy is better than neural networks.

The expansion characteristics SVM training set by the linear SVM and the calculation of new weights characteristics. It is logical to conclude that the features with small weights are ignored. Yao et Al., in the first phase, collected data from reservoirs with low and high productivity [63], [5] and put this information to test using fuzzy math, back propagation neural networks and self-organizing maps, the second phase, these data were tested with SVM. The results were compared and found it more reliable to use SVM. [64] Two feature classification algorithms: Relief-F [29] and SVM-RFE [21] were used for the evaluation of complex formations. Later, Yao et Al, tried again using support vector machines to convert nonlinear problems in linear problems in new function spaces. For example, they found that if all the features of the function $f(x_1, x_2)$ be extended to $f(x_1, x_2, x_1x_2, x_1^2, x_2^2)$, the problem becomes in a linear

classification problem in a new feature space. Although the radial basis function (RBF) and polynomial methods are used for expansion feature, SVM is used for feature selection. Inappropriate features are removed in space. An algorithm for recursive feature elimination (RFE) was developed, which trains linear SVMs and calculates the weights of different features. The functions that correspond to the smallest weights are removed and the process continues iteratively until none remain significant feature. In terms of results, the method has been compared with two artificial intelligence techniques and SVM. [65] With the regression modeling support vector machines [50], Kok Wai Wang et al., applied Lagrange multiplier to the analysis of well logs (well logging) to address the problem of reservoir characterization. His application was formulated as a convex programming problem basic to generate an objective function, linked to a list of restrictions. In large databases, the model worked reasonably well. [68]

In another area, [35] SVM was used for developing linear and nonlinear models of solubility in electrolytes and non-electrolytes. The SVM and a heuristic method were combined to predict the solubility of hydrocarbons in electrolytes. This method determined the suitability of SVM models under real-time conditions, using data that were not pre-processed. The SVM were used in combination with radial basis functions, fuzzy neural interface systems and perceptrons coercive.

The seismic attribute selection is critical in the use of post-stack seismic data for the prediction of hydrocarbons. Presently, the seismic characteristics are broad and diverse and this aim in different ways [6], for the prediction of a gas-bearing reservoir (formation or interval containing gas, either dissolved in the formation fluid as free gas) proposed a new set of seismic attributes: seismic impressions, which enable you to find the cepstral coefficients of the prediction center. The essential contribution of Jun-Xing Cao, the selection is optimized and the subsequent use of SVM for predicting reservoir.

Regarding the elimination of signal noise in seismic prospecting can be considered a problem when applying the regression function [2]. The LS-SVR (support vector regression least squares) is based on the Ricker waved kernel, which is applied to remove noise from seismic exploration data. To adapt LS-SVR, so to correct the seismic data, the parameters include the kernel parameter wavy Ricker, f and the regularization parameter, γ , are automatically selected according to the characteristics of the data set in the window.

4 SVM based Implementations

This section is a review of papers in which the SVM method, were used for hydrocarbon exploration. In each study case, if it is possible, the results were contrasted with other methods to verify the effectiveness of the SVMs.

4.1 Discriminating hydrocarbons using C-SVM in the Exploration

Quanhai Wang and Fang Miao [39], make a theoretical review of the cases of linearity and nonlinearity with SVM and assert that compared with another pattern-recognition technique based on neural networks, SVM is provided a perfect mathematical foundation. They use a nonlinear SVM algorithm based on the radial basis function (RBF) so as to recognize use of oil and gas in the upper part of the Yanghsinian Series carbonate reservoir in Guan YinChang, Si Chuan province.

The result of the experiment showed that the technique based on the star schema to display the attributes of the well, give a good predictive level. Not only can the linear classifier properly address the problem, but also the nonlinear classifier can be successful using a kernel function. Many simulation results show that the algorithm classification C-SVM has a good predictive capability. However, how to construct the kernel function and its parameters in applications is critical, and how to choose data attribute is a research topic with practical value and theoretical significance, especially for different areas.

4.2 Comparing against three classifiers SVM used in Data Mining for E&P

Shi Guangren, [45] briefly reviews four classifiers: MRA, BAYD, BPNN and SVM, where MRA is a linear function while BAYD, BPNN and SVM are nonlinear functions. Because BAYD, MRA, and SVM are explicit, they can be concretely expressed as mathematical formulas in despite of BPNN, since it is implicit. An important contribution of the author is that it has responded to the question: What is the most applicable classifier for a specific application? It was answered in two case studies, using the same known parameters which share the same unknown to be predicted: 1) evaluation of the quality of the trap of Northern Kuqa Depression of the Tarim Basin in western China, and 2) the identification of oil anticlinal structure Xiefengqiao Jianghan Basin in central China.

Table 2 Comparison between applications of MRA, BAYD, BPNN, and SVM to Trap Quality Evaluation in the Northern Kuqa Depression of the Tarim Basin in Western China

Classifier	Fitting formula	Mean absolute relative residuals (%)		Dependence of the predicted value (y) on parameters (x_1, x_2, \dots, x_n) in decreasing order	Time consuming on PC (Intel Core 2)	Integrated evaluation
		Learning samples	Prediction samples			
MRA (successive regression analysis)	Linear, explicit	0	16.67	$x_{12}, x_{10}, x_9, x_4, x_3, x_{11}, x_8, x_6, x_{14}, x_1, x_2, x_7, x_{13}, x_3$	<1 s	Average
BAYD (successive Bayesian discrimination)	Nonlinear, explicit	0	0	$x_{12}, x_{10}, x_9, x_4, x_3, x_{11}, x_8, x_6, x_{14}, x_1, x_2, x_7, x_{13}, x_3$	3 s	Excellent
BPNN (back-propagation neural network)	Nonlinear, implicit	0	0	N/A	1 min 20 s	Excellent
SVM (C-SVM binary classifier)	Nonlinear, explicit	0	0	N/A	3 s	Excellent

Case 1 shows that the results of BAYD, BPNN and SVM are same and can have zero residuals, while MRA has unallowable residuals; but Case 2 shows that the results of only SVM have zero residuals, while BAYD, BPNN and MRA have unallowable residuals. The reasons are: a) since the two cases are nonlinear problems, the linear MRA is not applicable; b) since the nonlinearity of Case 1 is weak, the nonlinear BAYD, BPNN and SVM are applicable; and c) since the nonlinearity of Case 2 is strong, only nonlinear SVM is applicable.

Therefore, it is proposed that: we can adopt MRA when a problem is linear; adopt BAYD, BPNN, or SVM when a problem is weakly nonlinear; and adopt only SVM when a problem is strongly nonlinear. In addition, the predictions of the applicable classifiers coincide with real exploration results, and a commercial gas trap was discovered after the forecast in Case 1 and SVM can correct some erroneous well-log interpretations in Case 2.

Table 2 shows the comparison between the four classifiers MRA, Bayd, BPNN and SVM using 27 training samples and 3 testing samples for case 1 (trap Northern Kuqa Depression of the Tarim Basin in western China).

4.3 Using the classification algorithm SVM-RFE feature in the E&P

Hongqi Li et al, [24] selected feature selection algorithms which were categorized by: (1) ranking of features and (2) selection of feature subsets. First, in this study, two feature ranking algorithms characteristics were used: Relief-F and SVM-RFE. Second, subset selection searches the space of possible features using greedy hill climbing, exhaustive, and genetic algorithm (GA), etc. for the optimal subset, which can be broken into wrappers, filters and embedded.

The feature selection algorithms as CFS (Correlation-based Feature Selection) and LVF (Las Vegas Filter), which evaluated but not classified, were considered. Regarding the methods of classification, five methods are examined and described: (1) Decision tree (DT) with four algorithms: C4.5, CART, CHAID and LMT, (2) artificial neural networks (ANN) with two types: BP and RBFN, (3) Support Vector Machine (SVM) with two algorithms: C-SVC and v-SVC, (4) Bayesian network (BN) with two algorithms: NB and TAN and (5) Ensemble learning, with four algorithms: AC4.5, ABP, and ALMT y ARBFN.

Based on three complex problems in Chinese oilfields, this paper examined the effects of feature selection and learning algorithms on formation evaluation using well logs and the following conclusions can be drawn: (1) Two feature ranking approaches, Relief-F and SVM-RFE, can improve classification performances more effectively than other methods such as expert judgment and the filter approaches, and Relief-F may be more powerful when the underlying problem is relative complex. (2) In the context of formation evaluation, SVM are the best classifier with high robustness and prediction accuracy. In particular, it is really important to choose the kernel parameters when SVMs are applied in practice, and (3) For the hydrocarbon reservoir prediction, the choice of classification methods is more important than the feature selection algorithms, and the combination of SVMs and feature ranking should be preferred to the other approaches. In summary the SVM classification method of testing was victorious.

The tools used were the *Rapid-Miner* application for almost all tests, while for the SVM analysis, *LIBSVM* library was used, both based on Open Source.

4.4 Application of SVM for the prediction of hydrocarbons

According Tian-Ren Cao Fei and Jun-Xing, [54] because in the prediction of oil, there are some common algorithms, such as fuzzy mathematics [60], clustering analysis, neural networks and other methods, these algorithms need a greater quantity of training samples. If the prediction results were based on insufficient training samples, this could be unsatisfactory. Therefore, the contribution of the authors is to provide a reasonable selection method for selecting the attribute or attributes strongly related to oil and gas, which makes it easier and more efficient prediction of hydrocarbons. This article focuses on the parameters proposed by Cao JX [6], based on printing methods and optimization in order to predict the distribution of oil and gas from a place in Central Sichuan.

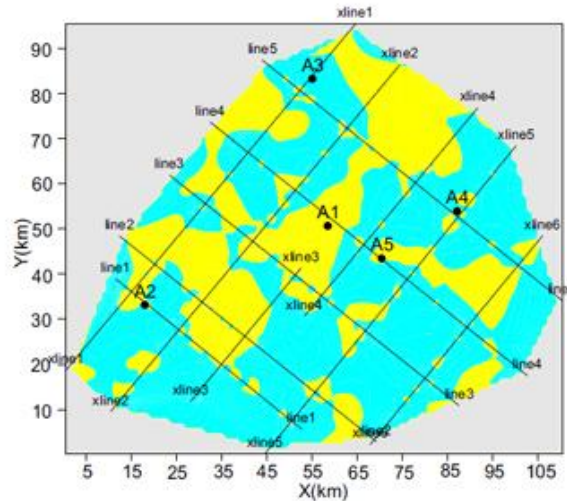


Fig.6: Using SVM to predict the distribution of hydrocarbons in a reservoir T_3X_2 (yellow for a possible distribution of hydrocarbons)

Experimental results show that the optimal parameters of seismic printing in combination with SVM, can help to improve prediction accuracy and reduce risk in the exploration phase of oil and gas, as well as offer a new way of managing projects in oil and gas. Analysis with SVM is improved by using an optimized version of **LIBSVM** Library.

4.5 Using LS-SVR with varying parameters and its application to seismic prospecting

Deng Xiaoying et al, [16] in his work, briefly review the theoretical framework of the SVM, as well as the SVR and specifically LS-SVR (support vector regression least squares). It mentions the contribution of Bing-Yu Sun et al. [2], which is applied to the signal with the objective of eliminating the Lidar signal under weak noise. SVM is named as SVR when applying the regression function. It is emphasized that LS-SVM is a modified form of SVM, just for simplicity. Signal denoising can be considered as a function regression problem, which involves estimating the underlying relationship from detected signal data.

LS-SVR embeds two tunable parameters including kernel parameter and regularization parameter that control the training setting, in which the later regulates the ratio between empirical risk and confidence interval of learning machine. These parameters may diminish the overall performance of LS-SVR if not well chosen.

They processed the data channel by channel. Firstly, the data of every channel is divided into windows every 100 data points. Then select two parameters by the aforementioned way and select every other data point as the training set to apply Ricker wavelet LS-SVR to remove random noise. Lastly, compute all points by LS-SVR model obtained.

The Ricker waved kernel is as follows:

$$K(x, x') = (1 - 2\pi^2 f^2 \|x - x'\|^2) e^{-\pi^2 f^2 \|x - x'\|^2} \quad (14)$$

where $f \in \mathbb{R}$ ($f > 0$) is the kernel parameter.

From the denoising experimental results of theoretical and practical seismic data, the performance of Ricker wavelet LS-SVR with variant parameters outperforms the one with invariant parameters in terms of the retrieved waveform in time domain and spectrum range in frequency domain.

5 Conclusions

In the phases of exploration and production of oil and natural gas, the volume of data can be divided into two categories. The first category is the same data type as petrophysical data, logging data, seismic data, etc. and the other category include diverse data types. For example, for the same geological body, different types of data must be processed and integrated with the characteristics of petrophysical data, logging data, seismic data and geological data in order to make a successful analysis.

To perform a successful classification and prediction of treatment from seismic exploration data and various sources, smart strategies will be defined, such as the choice of kernels or hybridization, or hybrid optimal method for tuning SVM hyperparameters, as well as pre-treatment of training and test sets (k-folds cross validation).

It's important to mention that the generalization ability of the SVM classification is directly related to the information implicit in the data used during training. Thus, it is desirable to train the machine if it has a considerable amount of data from other fields or deposits.

Publications reported to date have a strong interest in finding empirical risk levels so that it is not necessary to perform a validation to evaluate the solution obtained by a SVM for a given problem. While it is possible automatic selection of hyperparameters methods are being developed using those levels and optimization methods.

Data mining has been widely used in some fields of business and sciences, but the data mining application to hydrocarbon exploration and production is still in initial stage.

References

- [1] N. Ayat, M. Cheriet, C. Suen, "Automatic model selection for the optimization of SVM kernels", *Pattern Recognition* 38 (10) (2005) 1733–1745.
- [2] Sun Bing-Yu, D.S. Huang, H.T. Fang, "Lidar Signal Denoising Using Least-Squares Support Vector Machine", *IEEE Signal processing letters*, vol.12 (2) (2005) 101-104.
- [3] B. E. Boser, I. M. Guyon, V. N. Vapnik, "A training algorithm for optimal margin classifiers", 5th Annual ACM Workshop on COLT (ed. Haussler, D.) (1992) 144–152, ACM Press, Pittsburgh, PA.
- [4] C. J. Burges, "A tutorial on support vector machines for pattern recognition", *Data Min Know Disco* v2 (1998) 121–167.
- [5] Y. D. Cai, J. W. Gong, I. R. Gan, L. S. Yao, "Hydrocarbon reservoir prediction using artificial nerve network method", *Oil Geophys Prospect* 28 (1993) 634–638.
- [6] J. X. Cao, "Seismic Print Analysis", Twenty-sixth Annual Meeting of the Chinese Geophysical Society, 2010.
- [7] O. Chapelle, V. Vapnik, O. Bousquet, S. Mukherjee, "Choosing multiple parameters for support vector machines", *Machine Learning* 46 (1–3) (2002) 131–159.
- [8] Lin Chih-Jen et al, "Libsvm: A Library for Support Vector Machines", URL: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>. Revised 2005. Accessed April 13, 2013.
- [9] Z. Chunhong, J. Licheng, "Automatic parameters selection for SVM based on GA", *Proceedings of the 5th World Congress on Intelligent Control and Automation*, (2004) 1869–1872.
- [10] G. Cohen, M. Hilario, A. Geissbuhler, "Model selection for support vector classifiers via genetic algorithms. An application to medical decision support", *Proceedings of the 5th International Symposium on Biological and Medical Data Analysis*, (2004) 200–211.
- [11] S. Crisler, M. Morrissey, M. Anch, D. Barnett, "Sleep-stage scoring in the rat using a support vector machine", *Journal of Neuroscience Methods* 168 (2008) 524–534.
- [12] N. Cristianini, C. Campbell, J. Shawe-Taylor, "Dynamically adapting kernels in support vector machines", *Neural Inf Process Syst*, Vol. 2(1998) 204–210.
- [13] W. F. de la Vega, M. Karpinski, C. Kenyon, Y. Rabani, "Approximation schemes for clustering problems", In *Proc. 35th Annu. ACM Sympos. Theory Comput.*, (2003) 50–58.
- [14] X. Y. Deng, Y. Li, "Support Vector Regression Based on Ricker Wavelet Kernel Function and its Application to Seismic Prospecting Signal Denoising", *Journal of Jilin University (Earth Science Edition)*, Vol. 37(4) (2007) 821-827.
- [15] X. Y. Deng, Y. Li, "Study of parameters setting for least square support vector machine based on Ricker wavelet kernel in the denoising applications of seismic prospecting signals", *Progress in geophysics*, Vol.22(3) (2007) 953-959.
- [16] Xiaoying Deng, Dinghui Yang, Baojun Yang, "LS-SVR with Variant Parameters and its Practical Applications for Seismic Prospecting Data Denoising", *IEEE*, (2008) 1060–1063.
- [17] C. P. Diehl, G. Cauwenberghs, "SVM incremental learning, adaptation and optimization", *Proceedings of the International Joint Conference on Neural Networks*, (2003) 2685–2690.
- [18] B.F. de Souza et al., "Multiclass SVM model selection using particle swarm optimization", *Proceedings of the 6th International Conference on Hybrid Intelligent Systems*, (2006) 31–34.
- [19] C. Domeniconi, D. Gunopulos, "Incremental support vector machine construction", *Proceedings of the International Conference on Data Mining*, (2001) 589–592.
- [20] S. Geman, E. Bienenstock, "Neural networks and the bias-variance dilemma", *Neural Comput* 4 (1992)1–58.
- [21] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection", *J. Mach. Learn. Res.* 3 (2003) 1157–1182.
- [22] I. Guyon, V. Vapnik, B. Boser, L. Bottou, S. A. Solla, "Structural risk minimization for character recognition", *Adv Neural Inf Process Syst*, Vol 4 (1992) 471–479.
- [23] L. Hongdong, L. Yizeng, X. Qingsong, "Support vector machines and its applications in chemistry", *Chemometrics and Intelligent Laboratory Systems*, Vol 95 (2009) 188–198.

- [24] Li Hongqi, Guo Haifeng, Guo Haimin, Meng Zhaoxu, "Data Mining Techniques for Complex Formation Evaluation in Petroleum Exploration and Production: A Comparison of Feature Selection and Classification Methods", 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application, (2008) 37–43.
- [25] C. M. Huang, Y. J. Lee, D. K. Lin, V. Huang, "Model selection for support vector machines via uniform design", *Computational Statistics & Data Analysis* 52 (1) (2007) 335–346.
- [26] C. W. Hsu, C. J. Lin, "A comparison of methods for multiclass support vector machines", *IEEE Transactions on Neural Networks* 13 (2) (2002) 415–425.
- [27] Y. Jin, J. Branke, "Evolutionary optimization in uncertain environments - A survey", *IEEE Transactions on Evolutionary Computation* 9 (3) (2005) 303–317.
- [28] M. Jiang, X. Yuan, "Construction and application of PSO-SVM model for personal credit scoring", *Proceedings of the International Conference on Computational Science, Lecture Notes in Computer Science*, (2007) 158–161.
- [29] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF", *Proc. European Conf. on Machine Learning (Catania, April 1994)*
- [30] Kozielski, S., Wrembel, R. *New Trends in data Warehousing and Data Analysis*. USA. Springer; 2009.
- [31] L.C.F. da Silva, R.C.M. Portella, A.A. Emerick, N.F.F. Ebecken, "Predictive Data Mining Technologies for Oil Production Prediction in Petroleum", 2007 SPE Latin American and Caribbean Petroleum Engineering Conference, (2007) 1–5.
- [32] Jian Li, Haode Liao, Bing Wang, Bihua Xu, "The Design and Implementation of Web-based OLAP Drilling Analysis System", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery, (2010) 2570–2573.
- [33] Mingzhi Li, Yong Liu, Junhua Wang, "A New Parameter Optimization Algorithm of SVM", *Proceedings of AI 2011*, (2011) 174–178.
- [34] Hamel Lutz, *Knowledge Discovery with Support Vector Machines*, John Wiley & Sons, Inc., Canada, 2009.
- [35] W. Ma, X. Zhang, F. Luan, H. Zhang, R. Zhang, M. Liu, Z. Hu, B.T. Fan, "Support vector machine and the heuristic method to predict the solubility of hydrocarbons in electrolyte", *J Phys Chem A* 109 (2005) 3485–3492.
- [36] Moraes, DRS, Espíndola, RP, Evsukoff, AG, Ebecken, NEF "Cluster analysis of 3D seismic data for oil and gas exploration". *Data mining VII, data, text and web mining and their business applications*, Brazil, 2007.
- [37] P. Pai, W. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms", *Electric Power Systems Research*, 74(3) (2005) 417–425.
- [38] John Platt, "Using Analytic QP and Sparseness to Speed Training of Support Vector Machines", MIT Press, (1999) 557–563.
- [39] Wang Quanhai, Miao Fang, "The Support Vector Machine And Its Application To Hydrocarbon Discriminant In Oil And Gas Exploration", *International Conference Systems and Informatics (ICSAI 2012)*, (2012) 2337–2341.
- [40] A. Rukshan, T. Menik, F. Chandrika, "Data Mining Applications: Promise and Challenges", Chapter 2, *Data Mining and Knowledge Discovery in Real life Applications*. (2009); 201–214, Springer.
- [41] B. Scholkopf, C. J. C. Burges, A. J. Smola, *Advances in Kernel Methods: Support Vector Learning*, 1999.
- [42] B. Scholkopf, A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press Cambridge, Massachusetts, London, England, 2002.
- [43] M. Shaheen, M. Shahbaz, Z. Rehman, A. Guergachi, "Data mining applications in hydrocarbon exploration", *Artif Intell Rev*, (2011) 1–18.
- [44] H. Bhavsar, A. Ganatra, "Variations of Support Vector Machine Classification : A survey", *International Journal of Advanced Computer Research*, Volume 2, Number 4, Issue 6 (2012) 230–236.
- [45] Guangren Shi, "Four Classifiers Used in Data Mining and Knowledge Discovery for Petroleum Exploration and Development", *CS. Canada. Advances in Petroleum Exploration and Development*, (2011) 12–23.
- [46] A. Shilton, M. Palaniswami, D. Ralph, A. C. Tsai, "Incremental training of support vector machines", *IEEE Transactions on Neural Networks* ,16 (2005) 114–131.
- [47] Shigeo Abe., *Support Vector Machines for Pattern Classifications*, Second Edition. Springer, Londres; 2010.
- [48] Han Shunjie, Cao Qubo, Han Meng, "Parameter selection in SVM with RBF kernel function", *World Automation Congress (WAC) (2009)* 1–4.
- [49] A. J. Smola, B. Scholkopf, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press Cambridge, MA, USA, 2001.
- [50] A. J. Smola, B. Scholkopf, "A tutorial on support vector regression", *Stat Comput* 14 (2004) 199–222.
- [51] B.Y. Sun, D.S. Huang, H.T. Fang, "Lidar Signal Denoising Using Least-Squares Support Vector Machine", *IEEE Signal processing letters*, vol.12 (2) (2005) 101–104.
- [52] T. Sutorp, C. Igel, "Multi-objective optimization of support vector machines", *Multi-Objective, Machine Learning*, Vol. 16 of *Studies in Computational Intelligence*, Springer, (2006) 199–220.
- [53] F. Tay, L. Cao, "Application of support vector machines in financial time series forecasting", *Omega*, 29 (2001) 309–317
- [54] Ren-Fei Tian, Jun-Xing Cao, "Application of support vector machine method for predicting hydrocarbon in the reservoir", *Computational and Information Sciences (ICCIS)*, International Conference, (2011) 529–532.
- [55] S. Tripathi, V. Srinivas, R. Nanjundiah, "Downscaling of precipitation for climate change scenarios: A support vector machine approach", *Journal of Hydrology* 330, (2006) 621–640.
- [56] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer, N.Y., 1995, ISBN 0-387-94559-8.
- [57] V. N. Vapnik, *Statistical Learning Theory*. New York. Wiley, 1998.
- [58] G. Valentini, "An experimental bias-variance analysis of SVM ensembles based on resampling techniques", *IEEE Transactions on Systems Man and Cybernetics Part B* 35 (6) (2005) 1252–1271.
- [59] M. Venkatadri, C. Lokanatha, "A Review on Data mining from Past to the Future", *International Journal of Computer Applications*, Volume 15– No.7 (2011).
- [60] C. Y. Xiao, B. W. Zhu, "A fuzzy mathematical method for predicting hydrocarbon accumulation area by comprehensively analysing multiple kinds of seismic information", *OGP, Vol.25 (1990)* 191–200.
- [61] Li Xiongyan, Li Hongqi, Wu Zhuang, "A Model-Driven Data Mining in the Oil & Gas Exploration and Production", 2009 Second International Symposium on Knowledge Acquisition and Modeling, (2009) 20–24.
- [62] Jianhua Xu, Xuegong Zhang, Yanda Li, "Application of kernel Fisher discriminating technique to prediction of hydrocarbon reservoir", *OGP*, 37(2) (2002) 170–174.
- [63] J.H. Xu, R. Cai, "Application of supervised SOM neural network to oil and gas prediction", *Geophys Prospect Petrol* 37 (1996) 71–76.
- [64] K. Yao, W. Lu, S. Zhang, H. Xiao, Y. Li, "Feature expansion and feature selection for general pattern recognition problems", *IEEE international conference neural networks and signal processing Nanjing, China*, Vol. 1 (2003) 29–32.
- [65] K. Yao, W. Lu, S. Zhang, H. Xiao, Y. Li, "Hydrocarbon reservoir prediction using support vector machines", *LNCS, Springer, Berlin*, Vol 3173 (2004), 537–542.
- [66] P. Yu, J. Han, C. Faloutsos, *Link Mining: Models, Algorithms, and Applications*, Springer, 2010.

- [67] Shihong Yue, Ping Li, Peiyi Hao, "SVM Classification: Its Contents and Challenges", *Appl. Math. J. Chinese Univ. Ser. B*, Vol.18 (3) (2003) 332-342.
- [68] K.W. Wong, Y.S. Ong, D. Tamas et al., "Reservoir characterization using support vector machines", *Proceedings of the 2005 IEEE international conference on computational intelligence for modelling, control and automation, and international conference on intelligent agents, web technologies and internet commerce (CIMCA-IAWTIC'05)*.