Iranian Journal of Oil & Gas Science and Technology, Vol. 4 (2015), No. 2, pp. 01-14
http://ijogst.put.ac.ir

Separating Well Log Data to Train Support Vector Machines for Lithology Prediction in a Heterogeneous Carbonate Reservoir

Mohammad Ali Sebtosheikh[1], Reza Motafakkerfard[1], Mohammad Ali Riahi[2], and Siyamak Moradi[1,*]

[1] Department of Petroleum Exploration, Petroleum University of Technology, Abadan, Iran
[2] University of Tehran, Geophysics Institute, Tehran, Iran

Received: July 16, 2013; *revised*: June 29, 2014; *accepted*: August 23, 2014

Abstract

The prediction of lithology is necessary in all areas of petroleum engineering. This means that to design a project in any branch of petroleum engineering, the lithology must be well known. Support vector machines (SVM's) use an analytical approach to classification based on statistical learning theory, the principles of structural risk minimization, and empirical risk minimization. In this research, SVM classification method is used for lithology prediction from petrophysical well logs based on petrographic studies of core lithology in a heterogeneous carbonate reservoir in southwestern Iran. Data preparation including normalization and attribute selection was performed on the data. Well by well data separation technique was used for data partitioning so that the instances of each well were predicted against training the SVM with the other wells. The effect of different kernel functions on the SVM performance was deliberated. The results showed that the SVM performance in the lithology prediction of wells by applying well by well data partitioning technique is good, and that in two data separation cases, radial basis function (RBF) kernel gives a higher lithology misclassification rate compared with polynomial and normalized polynomial kernels. Moreover, the lithology misclassification rate associated with RBF kernel increases with an increasing training set size.

Keywords*:* Lithology Prediction, Support Vector Machines, Kernel Functions, Heterogeneous Carbonate Reservoirs, Petrophysical Well Logs

1. Introduction

Lithology prediction is one of the most important issues in all fields of petroleum engineering such as reservoir characterization, formation evaluation, geological studies, reservoir modeling, enhanced oil recovery processes, and well planning including drilling and well completion management. It is absolutely necessary to identify the exact lithology of a predetermined depth, especially in heterogeneous carbonate reservoirs, in order to make petroleum engineering related decisions Lithology prediction from drilling cuttings is not accurate due to problems associated with depth matching of cuttings; lithology determination from core plugs is not also economic because of operation costs. Accordingly, petrophysical well logs are used for lithology identification as a more efficient and cheaper approach than drilling cutting and core analysis (Rider, 2002). There are some traditional lithology identification methods that have been developed from petrophysical well logs by combining them and using cross plots. These methods are still useful today for quick evaluations (Ellis and Singer, 2008). However, traditional cross plotting methods have lost their efficiency in

* Corresponding Author:
Email: moradi.s@put.ac.ir

large data sets of heterogeneous reservoirs. Several approaches have been introduced for lithology classification such as cross plots interpretation and statistical analysis (Delfiner et al., 1987), statistical analysis based on histogram plotting (Busch et al., 1987), associating analysis by fuzzy logic, neural network and multivariable statistical methodologies (Carrasquilla et al., 2008), artificial intelligence approach and multivariate statistical analysis (Lim et al., 1999), fuzzy logic technique (Cuddy, 2000), artificial neural network methodologies (Chang et al., 2002; Chikhi and Batouche, 2005; Katz et al., 1999; Raeesi et al., 2012; Tang, 2009; Chikhi and Batouche, 2007), multi-agent collaborative learning architecture approach (Gifford and Agah, 2010), multivariate statistical method (Tang and White, 2008), facies classification using seismic attributes by SVM (Bagheri and Riahi, 2013), aggregation of principal components, clustering and discriminate analysis (Teh et al., 2012), and statistical characterization, discrimination, and stratigraphic correction methodologies (Borsaru et al., 2006).

The performance of artificial neural networks and fuzzy logic approaches are better than statistical analyses (Busch et al., 1987; Carrasquilla et al., 2008; Chang et al., 2002; Katz et al., 1999; Raeesi et al., 2012; Tang, 2009; Tang and White, 2008). Self-organized Map (SOM) neural network method shows a better performance in lithology classification compared with other methods (Chikhi and Batouche, 2005). Probabilistic neural network involves more computational steps and thus is slower than other kinds of neural networks (Tang, 2009). The minimum misclassification rate of the said methods is 19% which belongs to fuzzy logic techniques (Cuddy, 2000) and the maximum misclassification rate is 26% belonging to probabilistic neural networks (PNN) (Tang, 2009).

Support vector machine (SVM) is based on the statistical learning theory and was first introduced by Boser, Guyon, and Vapnik at the Computational Learning Theory conference where they presented their paper in 1992 (Boser et al., 1992). SVM has shown good performance in classification tasks. This is attributed to the fact that SVM's minimize an upper bound of the generalization error through maximizing the margin between the separating hyper planes (Amari and Wu, 1999). Recently, SVM's have successfully been applied to a number of applications such as drug design (Burbidge et al., 2001), fault diagnosis in power transmission systems (Ravikumar et al., 2008), microarray data classification (Huerta et al., 2006), protein structure prediction (Hua and Sun, 2001), text detection in digital videos (Shin et al., 2000), microarray data and satellite radiance data classification (Lee et al., 2004), speaker identification (Mezghani et al., 2010), document classification (Wang and Sun, 2011), and hyper spectral images classification (Ding, 2011).

This study examines the performance of different kernel functions of SVM in the lithology prediction. It uses a well by well data separation technique in a heterogeneous carbonate reservoir in southwestern Iran by applying petrophysical well logs. To choose the well logs which are most effective on the SVM performance, an attribute selection approach was used. In order to find the best SVM parameters, a grid search technique has been utilized. The effect of different kernel types on the SVM performance was investigated.

2. Methodology

2.1. Support vector machines

Support vector machines represent a machine learning algorithm for both classification and regression tasks (Alpaydin, 2010; Hamel, 2009). Support vector classifiers are maximum margin classifiers that find a decision function for pattern vectors X of dimension n attributes belonging to either of the two classes (Boser et al., 1992). Maximum margin classifiers construct decision surfaces that are

equidistant to the class boundaries called hyper planes. They maximize the margin between the two classes supporting hyper planes (Hamel, 2009).

An SVM, as a maximum margin classifier, is an optimization routine. It represents an ideal method to select the best solution from a number of feasible or possible solutions. That solution has the maximum distance from the two-class hyper planes in the presence of limitative instances of the two classes which are called support vectors as constrain conditions of the optimization problem. Support vectors are instances that classes hyper planes cannot cross over in order to maximize the margin (Hamel, 2009). To find an optimal decision function, first the margin between the boundaries of the classes is formulated in the direct space and then it is transformed into the dual space by means of the Lagrangian (Boser et al., 1992; Hamel, 2009).

2.2. SVM's as linear classifiers

A data set, in its simplest form, is assumed to be linearly separable. A training data point, $X_i \in R^n$, $i = 1,...,l$, where l is the total number of the training instances, and n is the dimension of the input attributes. Every instance has a class which is labeled with $y_i \in \{-1,+1\}$. The decision function is given as:

$$f(X) = W^T X + b, \tag{1}$$

where, $W, X \in R^n$; b is a scalar bias and W is a weight vector that is called the normal vector. A standard hyper plane to facilitate the computation of the support vectors is given by:

$$\min_{X_i \in \mathbb{R}^n} \left| W^T X_i + b \right| = 1 \tag{2}$$

The normal vector is obtained from the following optimization problem (Kecman, 2005):

$$\text{Minimize } \frac{1}{2} W^T W \tag{3}$$

$$\textit{subject to } y_i \left| W^T X_i + b \right| \geq 1, \; i = 1,...,l, \tag{4}$$

where, Equation 4 is the constraint of Equation 3. Equation 3, subject to Equation 4, has a solution in its saddle point which can be determined by using the following Lagrangian functional:

$$L(W, b, \alpha) = \frac{1}{2} W^T W - \sum_{i=1}^{l} \alpha_i (W^T X_i + b) - 1, \tag{5}$$

where, α_i are Lagrangian multipliers as the dual parameters (Kecman, 2005).

As the saddle point, the Lagrangian *L* is minimized with respect to *W* and *b* and is maximized with respect to non-negative α_i.

To solve the optimization problem, the dual problem is formed and from the Karush-Kuhn-Tucker (KKT) conditions, the following relationship must be satisfied at the saddle point so as to differentiate with respect to W_s and b_s; setting the derivatives equal to zero yields (Kecman, 2005):

$$\frac{\partial L}{\partial W_s} = 0 \Rightarrow \quad W_s = \sum_{i=1}^{l} \alpha_i y_i X_i \tag{6}$$

$$\frac{\partial L}{\partial b_s} = 0 \Rightarrow \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{7}$$

where, the *s* subscript denotes the values at the saddle point. In order to find b_s, by substituting Equations 6 and 7 into Equation 5, the following dual variables Lagrangian $L_d(\alpha)$ is obtained:

$$maximize\ L_d(\alpha) = -\frac{1}{2}\alpha^T H \alpha + e^T \alpha \tag{8}$$

$$subject\ to\ \hat{f}(\vec{x}) = \begin{cases} y^T \alpha = 0 \\ \alpha_i \geq 0\ ,\ i = 1,...,l \end{cases} \tag{9}$$

where, $\alpha = [\alpha_1, \alpha_2, ..., \alpha_i]^T$; H is the Hessian matrix ($H_{ij} = y_i y_j X_i^T X_j$) and e is the unit vector. The results of solving Equation 8 are the optimal values of α_i which are denoted by α_{si}. The optimal values W_s and b_s are then found as follows (Kecman, 2005):

$$W_s = \sum_{i=1}^{l} \alpha_{si} y_i X_i \tag{10}$$

$$b_s = \frac{1}{N_{sv}} \sum_{t=1}^{N_{sv}} y_t - X_t^T W_s\ ,\ t = 1,..., N_{sv} \tag{11}$$

where, N_{sv} is the number of support vectors. The decision function $f(x)$ and the indicator labeling function, i_F are then given as:

$$f(X) = \sum_{i=1}^{l} W_{si}^T X_i + b_s \tag{12}$$

$$i_F = sign(f(X)), \tag{13}$$

where, the indicator labeling function by performing the labeling task defines the classification between categories $i_F = -1$ or $i_F = +1$.

2.3. SVM's as nonlinear classifiers

Very few data sets in the real world are linearly separable. The remarkable characteristic of the support vector classifiers is that the basic linear framework is extended to the case where the data is not linearly separable. The fundamental idea behind this extensibility is to transform the input space where the data set is not linearly separable into a higher dimensional space called a feature space (Hamel, 2009). The input vectors $X \in R^n$ are projected onto vectors $\Phi(X)$ of a higher dimensional feature space, where the data set is linearly separable and then the SVM separates new images of the projections of X using the linear classifier formulation. This transformation results in a quadratic programming optimization problem with constraint in the feature space (Kecman, 2005). The indicator labeling functions becomes:

$$i_F = sign\left(\sum_{i=1}^{l} \alpha_i y_i \Phi^T(X_i) \Phi(X) \right) = sign\left(\sum_{i=1}^{l} v_i k(X_i, X) + b \right) \tag{14}$$

where, v_i are the weights and $k(X_i, X)$ is the kernel function. Kernel functions evaluate a dot

product in feature space and the defining characteristic of a kernel is that the value of this dot product is actually computed in the input space (Hamel, 2009). Some common kernel functions and their associated mathematical formulas are listed in Table 1.

Table 1
Common kernel functions and their associated mathematical formulas.

Kernel Function	Mathematical Formula
Linear kernel	$k(X_i, X) = X_i \bullet X$
Polynomial kernel	$k(X_i, X) = (X_i \bullet X)^e$
Normalized Polynomial kernel	$k(X_i, X) = \left[(X_i \bullet X)^e \right] \Big/ \sqrt{(X_i \bullet X_i)^e (X \bullet X)^e}$
Radial basis kernel	$k(X_i, X) = e^{-\left(\lvert X_i - X \rvert^2 / 2\sigma^2 \right)}$

Following the linear support vector classifier, the formulation for the nonlinear support vector classifier is a generalization on the linear SVM and it is given by (Kecman, 2005):

$$Maximize\ L_d(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(X_i, X_j) \tag{15}$$

$$subject\ to \begin{cases} \alpha_i \geq 0\ ,\ i = 1, ..., l \\ \sum_{i=1}^{l} y_i \alpha_i = 0 \end{cases} \quad for\ a\ separable\ nonlinear\ classifier \tag{16}$$

or

$$subject\ to \begin{cases} C \geq \alpha_i \geq 0\ ,\ i = 1, ..., l \\ \sum_{i=1}^{l} y_i \alpha_i = 0 \end{cases} \quad for\ an\ overlapping\ nonlinear\ classifier. \tag{17}$$

where, C is a positive nonzero value called penalty parameter (or cost factor) and determines the trade-off between the training error and Vapnik–Chervonenkis (VC) dimension of the model. The VC dimension of the classifier in a model class is the number of instances in training data set that can be separated by the classifier (Hamel, 2009). More precisely, a large value for C forces the optimization to consider solutions with small margins (Hamel, 2009). The penalty parameter C is an important parameter on which the performance of SVM extremely depends (Witten et al., 2011; Hamel, 2009). The penalty parameter was chosen by grid search technique using the Weka software (Hall et al., 2009).

After the optimization of the training data in order to determine the Lagrangian multipliers, the decision function and the indicator labeling function are obtained by (Kecman, 2005):

$$f(X) = \sum_{i=1}^{l} y_i \alpha_i k(X_i, X) + b \tag{18}$$

$$i_F = sign\left(\sum_{i=1}^{l} y_i \alpha_i k(X_i, X) + b \right) \tag{19}$$

Choosing the appropriate kernel function, it is possible to classify nonlinearly separable data sets. The evaluation criterion for a nominal prediction work is defined as the number of testing set instances

which were predicted correctly divided by the total number of testing set instances. This criterion is called the misclassification rate.

3. Data preparation

The lithologies of 427 instances from petrographic analysis with the exact core depth were prepared. These instances are anaclitic to three wells of a heterogeneous carbonate reservoir in southwestern Iran. Three diagnosed lithology from petrographic analysis were used. All the instances had been matched with the log depth, whereas every instance has exact values of petrophysical well logs as its associated attributes. Deep latero log (LLD), shallow latero log (LLS), micro spherically focused log (MSFL), sonic transit time log (DT), neutron porosity log (NPHI), caliper log (CALI), photoelectric factor log (PEF), density log (RHO) and gamma ray log (GR) were nine petrophysical well log measurements in the three wells used for lithology prediction. Since the learning process is easier on equal limited range input data, all the input petrophysical well log data of every well were normalized separately in the range of [-1, +1].

4. Implementation

4.1. Cross-validation

The result credibility of a data mining process depends on choosing an efficient method for data partitioning. In practical terms, it is common to hold-out one-third of the data for testing and use the remaining two-thirds for training. When trying to perform an accurate grid search technique for the optimization of kernel function parameters, the instance that is used for training or testing might not be representative and all instances with a certain class might be omitted from the training set. To overcome these shortcomings of hold-out method in grid search, an important statistical technique, called cross-validation, was used. In cross-validation, the data set is divided into a fixed number of folds or partitions. Since it has experimentally been proved that a 10-fold cross-validation gives the best estimate of misclassification rate error, using it as a standard cross-validation technique is recommended. In a 10-fold cross-validation, the whole of the data set is randomly separated into ten equal partitions. Each part is held out in turn and then the learning scheme is trained on the remaining nine; then, this procedure is executed on the nine remaining different partitions (Witten et al., 2011).

4.2. Attribute selection

In practice, adding irrelevant or distracting attributes to the input data set often confuses machine learning systems such as SVM's (Witten et al., 2011).

The Weka software (Hall et al., 2009) was used to perform attribute selection against the whole data set, where all the petrophysical well logs were ranked. Table 2 shows the ranked well logs during the attribute selection operation. The top six relevant attributes of RHOB, LLS, NPHI, LLD, DT, and PEF were selected as the input data.

Table 2
Well logs ranking.

Ranking	Ranked Attributes
1	RHOB
2	LLS
3	NPHI
4	LLD
5	DT
6	PEF
7	CALI
8	MSFL
9	GR

4.3. Data separation

Well by well data separation was performed as a data partitioning criterion for training and testing SVM models. Thus, as shown in Table 3, three cases were formed with each one having its own associated training set and testing set. They were used for training and testing SVM models for grid search and SVM performance investigation. Furthermore, cases index, sizes of training and testing data sets, and the ratio of training set size to the whole data set size are tabulated in Table 3.

Table 3
Cases index, their associated training and testing sets, data sets sizes, and the ratio of training set size to the whole data set size.

Case Index	Training Set	Testing Set	Training Set Size	Testing Set Size	Training Set Size to Whole Data Set Size Ratio
1	Well 2 and well 3	Well 1	237	190	55.50%
2	Well 1 and well 3	Well 2	295	132	69.09%
3	Well 1 and well 2	Well 3	322	105	75.40%

4.4. Kernel parameters optimization by grid search technique

The accuracy of a SVM model operation extremely depends on the selection of suitable kernel function parameters. In a grid search process, two parameters of a kernel function are determined with search range and step size incrimination of each parameter. Then, the accuracy of the classifier with each pair is evaluated. Afterwards, a pair that causes minimum misclassification rate error is chosen as the optimal pair of kernel function parameters (Witten et al., 2011). A 10-fold cross-validation grid search, based on maximum accuracy as the evaluation criterion, was performed to find the optimal values of the SVM model parameters using the training set data of each data separation case by the Weka software (Hall et al., 2009). Search range, the optimum value of each parameter, and the total number of evaluated kernel functions of all the data separation cases for different kernel types are shown in the Table 4.

Table 4
Search ranges, optimal parameter values, and the total number of evaluated kernel functions obtained from grid search of all the data separation cases for different kernel types.

Kernel Function		Parameters	Search Range	Optimum Values for Parameters	Total Number of Evaluated Kernel Functions
Case 1	Polynomial	Penalty parameter, *C*	0.1-40000.0	18000.1	29067
		Exponent parameter, e	2.0-8.0	2.0	
	Normalized polynomial	Penalty parameter, *C*	0.1-40000.0	0.1	30053
		Exponent parameter, e	2.0-8.0	7.5	
	Radial basis function	Penalty parameter, *C*	0.1-40000.0	0.1	30300
		RBF parameter, σ	0.01-5.01	4.073	
Case 2	Polynomial	Penalty parameter, *C*	40000.0	6000.1	29145
		Exponent parameter, e	2.0-8.0	2.0	
	Normalized polynomial	Penalty parameter, *C*	0.1-40000.0	16000.1	30120
		Exponent parameter, e	– 8.0	2.0	
	Radial basis function	Penalty parameter, *C*	0.1-40000.0	1000.1	30246
		RBF parameter, σ	0.01-5.01	1.023	
Case 3	Polynomial	Penalty parameter, *C*	40000.0	29000.1	29033
		Exponent parameter, e	2.0 – 8.0	5.0	
	Normalized polynomial	Penalty parameter, *C*	0.1 – 40000.0	2000.1	30012
		Exponent parameter, *e*	– 8.0	4.0	
	Radial basis function	Penalty parameter, *C*	0.1-40000.0	9680.1	30345
		RBF parameter, σ	0.01-5.01	1.023	

5. Results and discussion

Application and assessment of different kernel functions of SVM

The data was first prepared and preprocessed through a) the normalization of numeric input data separately for each petrophysical well log, b) attribute selection process, and c) the selection of the optimal values of the parameters based on grid search techniques for radial basis function, polynomial, and normalized polynomial kernels. The effect of kernel types on the SVM performance was then examined.

Three kernel types, namely radial basis function (RBF), polynomial, and normalized polynomial kernels, with the associated optimal values of the parameters were applied to the data set using the Weka software (Hall et al., 2009). For each of case of Table 3, the SVM model was trained with the training data set with its associated optimal values of the parameters obtained from grid search, and then the trained SVM model was applied to the lithology prediction of the testing set. The results for each case are as follows:

Case 1: Prediction of well 1 instances

Three kernel types with their associated optimal parameter values were tested separately to predict all 190 instances of well 1 against training with 237 instances of well 2 and well 3. In this case, the ratio of the training set size to the whole data set size is 55.50%. Tables 5-7 show the confusion matrices

for RBF, polynomial, and normalized polynomial kernel types respectively. Total lithology misclassification rates associated with well 1 predictions for RBF, polynomial, and normalized polynomial kernels were 15.26%, 14.21%, and 17.36% respectively as depicted in Figure 1.

Table 5
Confusion matrix of well 1 instances using RBF kernel.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	109	4	0
	Limestone	21	52	0
	Anhydrite	3	1	0

Table 6
Confusion matrix of well 1 instances using polynomial kernel function.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	104	4	5
	Limestone	12	59	2
	Anhydrite	3	1	0

Table 7
Confusion matrix of well 1 instances using normalized polynomial kernel function.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	106	7	0
	Limestone	22	51	0
	Anhydrite	3	1	0

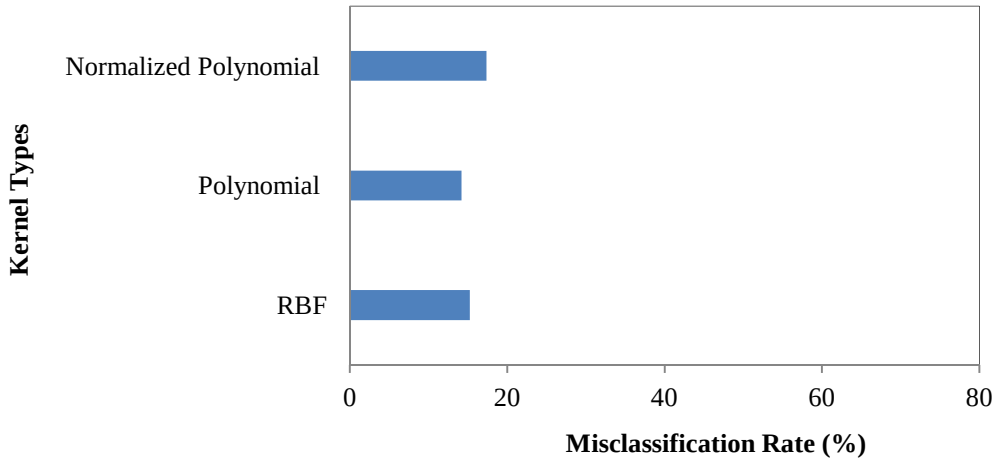

Figure 1
Total of lithology misclassification rate of well 1 instances using RBF, polynomial, and normalized polynomial kernels.

Case 2: Prediction of well 2 instances

Three kernel types with their associated optimal parameter values were tested separately to predict all

132 instances of well 2 against training with 295 instances of well 1 and well 3. In this case, the ratio of the training set size to the whole data set size is 69.09%. Tables 8-10 show the confusion matrices for RBF, polynomial, and normalized polynomial kernel types respectively. Total lithology misclassification rates associated with well 2 predictions for RBF, polynomial, and normalized polynomial kernels were 17.42%, 16.66%, and 14.39% respectively as illustrated in Figure 2.

Table 8
Confusion matrix of well 2 instances using RBF kernel.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	34	11	0
	Limestone	10	75	1
	Anhydrite	1	0	0

Table 9
Confusion matrix of well 2 instances using polynomial kernel function.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	36	9	0
	Limestone	11	74	1
	Anhydrite	1	0	0

Table 10
Confusion matrix of well 2 instances using normalized polynomial kernel function.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	36	9	0
	Limestone	9	77	0
	Anhydrite	1	0	0

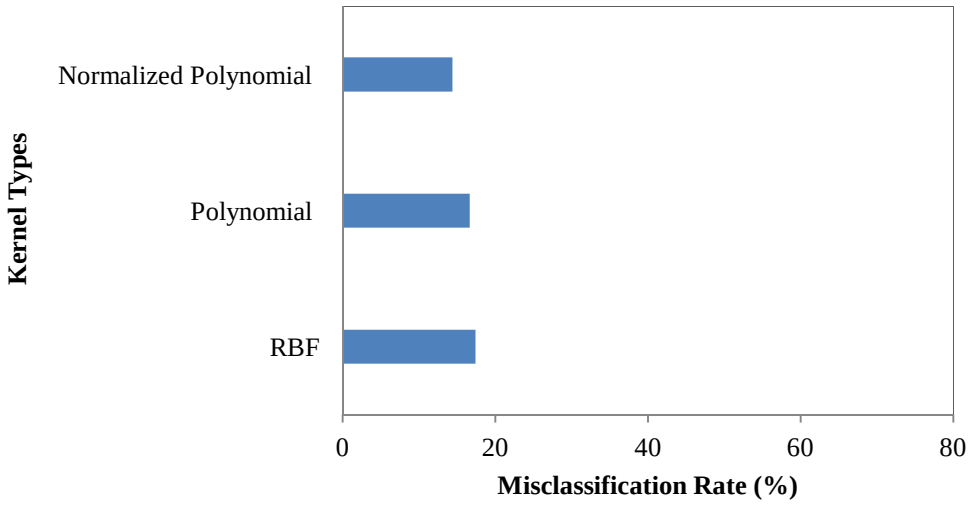

Figure 2
Total of lithology misclassification rate of well 2 instances using RBF, polynomial, and normalized polynomial kernels.

Case 3: Prediction of well 3 instances

Three kernel types with their associated optimal parameter values were tested separately to predict all 105 instances of well 3 against training with 322 instances of well 1 and well 2. In this case, the ratio of the training set size to the whole data set size is 75.40%. Tables 11-13 show the confusion matrices for RBF, polynomial, and normalized polynomial kernel types respectively. Total lithology misclassification rates associated with well 3 predictions for RBF, polynomial, and normalized polynomial kernels were 18.09%, 17.14%, and 17.14% respectively as displayed in Figure 3.

Table 11
Confusion matrix of well 3 instances using RBF kernel.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	28	11	0
	Limestone	7	58	0
	Anhydrite	1	0	0

Table 12
Confusion matrix of well 3 instances using polynomial kernel function.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	29	10	0
	Limestone	7	58	0
	Anhydrite	1	0	0

Table 13
Confusion matrix of well 3 instances using normalized polynomial kernel function.

		Predicted		
		Dolomite	**Limestone**	**Anhydrite**
Actual	**Dolomite**	30	9	0
	Limestone	8	57	0
	Anhydrite	1	0	0

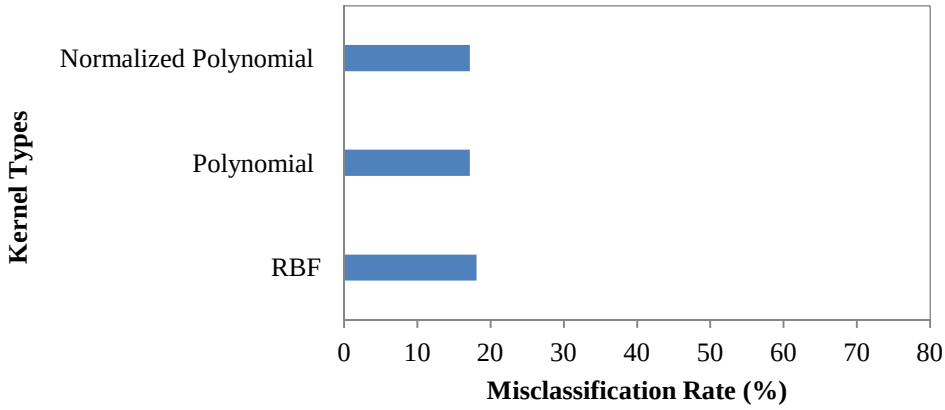

Figure 3
Total of lithology misclassification rate of well 3 instances using RBF, polynomial, and normalized polynomial kernels.

6. Conclusions

A support vector machine, as a supervised learning classifier, was used to predict the lithology from petrophysical well logs in a heterogeneous carbonate reservoir in southwestern Iran based on core lithology verification. In order to remove irrelevant or distracting input well logs, an attribute selection approach was employed to rank the input well logs. Three well logs with less dependency were therefore omitted from input well logs. aRHOB, LLS, NPHI, LLD, DT, and PEF logs are most lithology affected well logs in the investigated reservoir. Well by well data separation as a data partitioning criterion was performed for generating training and testing data sets. Because of the dependency of SVM on its associated parameters, a grid search algorithm was used to characterize parameter optimization for each kernel function and each data separation case. It is concluded that the SVM is capable of predicting lithology in heterogeneous carbonate reservoirs. All of the misclassification rates of this study are less than those in previous works. The results show that the SVM performance in the lithology prediction of wells by applying well by well data partitioning technique is good and that, in two data partitioning cases, radial basis function (RBF) kernel gives more lithology misclassification rate than polynomial and normalized polynomial kernels. In addition, the lithology misclassification rate associated with RBF kernel function increases with increasing training data set size. Using these kernels with their associated optimal parameter values, it is possible to predict lithology in the investigated reservoir. In most cases, the anhydrite instances were predicted to be dolomite. It seems that this occurred because of the similarity of the petrophysical properties of these two lithology types. Therefore, using more lithology affected well logs is recommended to overcome this shortcoming.

Acknowledgements

The authors would like to thank the Exploration Directorate of National Iranian Oil Company (NIOC) for providing the data for this research.

Nomenclature

b	: Bias constant
C	: Penalty parameter
e	: Exponent parameter of polynomial kernel
H	: Hessian matrix
k	: Kernel function
l	: Number of instances
L	: Lagrangian equation
X	: Input vector of attributes
y	: Output vector of class label
α, β	: Lagrangian multipliers
σ	: RBF parameter
ξ	: Slack variable
W	: Normal vector
Φ	: Mapping function from input space to feature space

Superscripts and subscripts

d	: Dual
i, j	: Indices
n	: Input space dimension

s	: Optimal
p	: Primal
SV	: Support vectors
T	: Transpose

References

Alpaydin, E., Introduction to Machine learning, The Massachusetts Institute of Technology Press, 537 p., 2010.

Amari, S. and Wu, S., Improving Support Vector Machine Classifiers by Modifying Kernel Functions, Neural Networks, Vol. 12, No. 6, p. 783-789, 1999.

Borsaru, M., Zhou, B., Aizawa, T., Karashima, H., and Hashimoto, T., Automated Lithology Prediction from PGNAA and other Geophysical Logs: Applied Radiation and Isotopes, Vol. 64, No. 2, p. 272-282, 2006.

Boser, B. E., Guyon, I. M., and Vapnik, V. N., A Training Algorithm for Optimal Margin Classifiers, in Proceedings Proceedings of the Fifth Annual Workshop on Computational learning Theory1992, ACM, p. 144-152, 1992.

Burbidge, R., Trotter, M., Buxton, B., and Holden, S., Drug Design by Machine Learning: Support Vector Machines for Pharmaceutical Data Analysis: Computers & Chemistry, Vol. 26, No. 1, p. 5-14, 2001.

Busch, J. M., Fortney, W. G., and Berry, L. N., Determination of Lithology from Well Logs by Statistical Analysis: SPE Formation Evaluation, Vol. 2, No. 4, p. 412-418, 1987.

Carrasquilla, A., Silvab, J. D., and Flexa, R., Associating Fuzzy Logic, Neural Networks and Multivariable Statistic Methodologies in the Automatic Identification of Oil Reservoir Lithologies through Well Logs: Revista de Geologia, Vol. 21, No. 1, p. 27-34, 2008.

Chang, H. C., Kopaska-Merkel, D. C., and Chen, H.-C., Identification of Lithofacies using Kohonen Self-organizing Maps: Computers & Geosciences, Vol. 28, No. 2, p. 223-229, 2002.

Chikhi, S. and Batouche, M., Using Probabilistic Unsupervised Neural Method for Lithofacies Identification: The International Arab Journal of Information Technology, V. 2, No. 1, p. 58-66, 2005.

Chikhi, S. and Batouche, M., Hybrid Neural Network Methods for Lithology Identification in the Algerian Sahara: World Academy of Science, Engineering and Technology, Vol. 4, p. 774-782, 2007.

Cuddy, S. J., Litho-facies and Permeability Prediction From Electrical Logs Using Fuzzy Logic: SPE Reservoir Evaluation & Engineering, Vol. 3, No. 4, p. 319-324, 2000.

Delfiner, P., Peyret, O., and Serra, O., Automatic Determination of Lithology from Well Logs: SPE Formation Evaluation, Vol. 2, No. 3, p. 303-310, 1987.

Ding, S., Spectral and Wavelet-based Feature Selection with Particle Swarm Optimization for Hyperspectral Classification: Journal of Software, Vol. 6, No. 7, p. 1248-1256, 2011.

Ellis, D. V. and Singer, J. M., Well Logging for Earth Scientists, Dordrecht, Netherlands, Springer, 692 p., 2008.

Gifford, C. M. and Agah, A., Collaborative Multi-agent Rock Facies Classification from Wireline Well Log Data: Engineering Applications of Artificial Intelligence, Vol. 23, No. 7, p. 1158-1172, 2010.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H., The WEKA Data Mining Software: An Update, SIGKDD Explorations, Vol. 11, No.1, 2009.

Hamel, L., Knowledge Discovery with Support Vector Machines, Hoboken, New Jersey, John Wiley

& Sons, Inc., 246 p., 2009.

Hua, S., and Sun, Z., A Novel Method of Protein Secondary Structure Prediction with High Segment Overlap Measure: Support Vector Machine Approach: Journal of Molecular Biology, Vol. 308, No. 2, p. 397-408, 2001.

Huerta, E. B., Duval, B., and Hao, J.-K., A Hybrid GA/SVM Approach for Gene Selection and Classification of Microarray Data, Applications of Evolutionary Computing, Springer, p. 34-44, 2006.

Katz, S. A., Vernik, L., and Chilingar, G. V., Prediction of Porosity and Lithology in Siliciclastic Sedimentary Rock using Cascade Neural Assemblies: Journal of Petroleum Science and Engineering, Vol. 22, No. 1, p. 141-150, 1999.

Kecman, V., Support Vector Machines-An Introduction, *in* Wang, L., ed., Support Vector Machines: Theory and Applications, Springer Berlin Heidelberg, Chap. 1, p. 1-47, 2005.

Lee, Y., Lin, Y., and Wahba, G., Multicategory Support Vector Machines: Theory and Application to the Classification of Microarray Data and Satellite Radiance Data: Journal of the American Statistical Association, Vol. 99, No. 465, p. 67-81, 2004.

Lim, J.-S., Kang, J. M., and Kim, J., Interwell Log Correlation Using Artificial Intelligence Approach and Multivariate Statistical Analysis, SPE Asia Pacific Oil and Gas Conference and Exhibition: Jakarta, Indonesia, Society of Petroleum Engineers, 1999.

Mezghani, D., Boujelbene, S., and Ellouze, N., Evaluation of SVM Kernels and Conventional Machine Learning Algorithms for Speaker Identification: International Journal of Hybrid Information Technology, Vol. 3, No. 3, p. 23-34, 2010.

Raeesi, M., Moradzadeh, A., Doulati Ardejani, F., and Rahimi, M., Classification and Identification of Hydrocarbon Reservoir Lithofacies and Their Heterogeneity using Seismic Attributes, Logs Data and Artificial Neural Networks: Journal of Petroleum Science and Engineering, Vol. 82, p. 151-165, 2012.

Ravikumar, B., Thukaram, D., and Khincha, H., Application of Support Vector Machines for Fault Diagnosis in Power Transmission System: Generation, Transmission & Distribution, IET, Vol. 2, No. 1, p. 119-130, 2008.

Rider, M., The Geological Interpretation of Well Logs, Sutherland, Scotland, Rider-French Consulting Ltd., 280 p., 2002.

Shin, C., Kim, K., Park, M., and Kim, H., Support Vector Machine-based Text Detection in Digital Video, *in* Proceedings Neural Networks for Signal Processing X., Proceedings of the 2000 IEEE Signal Processing Society Workshop2000, Vol. 2, p. 634-641, 2000.

Tang, H., Successful Carbonate Well Log Facies Prediction using an Artificial Neural Network Method: Wafra Maastrichtian Reservoir, Partitioned Neutral Zone (PNZ), Saudi Arabia and Kuwait, SPE Annual Technical Conference and Exhibition: New Orleans, Louisiana, Society of Petroleum Engineers, 2009.

Tang, H. and White, C. D., Multivariate Statistical Log log-facies Classification on a Shallow Marine Reservoir: Journal of Petroleum Science and Engineering, Vol. 61, No. 2, p. 88-93, 2008.

Teh, W., Willhite, G. P., and Doveton, J., Improved Reservoir Characterization in the Ogallah Field using Petrophysical Classifiers within Electrofacies, in Proceedings SPE Improved Oil Recovery Symposium, Tulsa, Oklahoma, USA, 2012.

Wang, Z. and Sun, X., An Efficient Discriminant Analysis Algorithm for Document Classification, Journal of Software, Vol. 6, No. 7, p. 1265-1272, 2011.

Witten, I. H., Frank, E., and Hall, M. A., Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, Elsevier, 630 p., 2011.