**Nonlinear Processes
in Geophysics**

Gas chimney detection based on improving the performance of combined multilayer perceptron and support vector classifier

H. Hashemi[1,2], **D. M. J. Tax**[2], **R. P. W. Duin**[2], **A. Javaherian**[1], **and P. de Groot**[3]

[1]Institute of Geophysics, University of Tehran, P.O. Box 14155-6466, Tehran, I. R. Iran
[2]ICT, Faculty of EEMCS, Delft University of Technology, P.O. Box 5031, 2600 GA, Delft, The Netherlands
[3]dGB Earth Sciences BV, Nijverheidstraat 11-2, 7511 JM, Enschede, The Netherlands

Abstract. Seismic object detection is a relatively new field in which 3-D bodies are visualized and spatial relationships between objects of different origins are studied in order to extract geologic information. In this paper, we propose a method for finding an optimal classifier with the help of a statistical feature ranking technique and combining different classifiers. The method, which has general applicability, is demonstrated here on a gas chimney detection problem. First, we evaluate a set of input seismic attributes extracted at locations labeled by a human expert using regularized discriminant analysis (RDA). In order to find the RDA score for each seismic attribute, forward and backward search strategies are used. Subsequently, two non-linear classifiers: multilayer perceptron (MLP) and support vector classifier (SVC) are run on the ranked seismic attributes. Finally, to capitalize on the intrinsic differences between both classifiers, the MLP and SVC results are combined using logical rules of maximum, minimum and mean. The proposed method optimizes the ranked feature space size and yields the lowest classification error in the final combined result. We will show that the logical minimum reveals gas chimneys that exhibit both the softness of MLP and the resolution of SVC classifiers.

1 Introduction

When fluids migrate upwards through a sedimentary sequence, rocks are cracked or chemically altered and connate gas might stay behind after the fluids have passed. In processed seismic data these effects manifest themselves as subtle vertical noise trails. It is worth studying such trails as they reveal hydrocarbon migration paths and thus provide useful information about the petroleum system. On conventional

seismic displays only large vertical noise trails can be recognized as gas chimneys. Meldahl et al. (1999) developed a pattern recognition technique to facilitate the interpretation of gas chimneys. Their method transfers a seismic volume into a new volume that highlights only vertical disturbances. They refer to this new volume as "The Chimney Cube". The cube is generated by a neural network that was trained on multiple attributes extracted at positions labelled by a human expert. The target vectors for the neural network are (1,0) and (0,1) representing chimneys and non-chimney locations, respectively. In the application phase the node representing the chimney class is output. Values in the volume are representing chimney "probability", which ranges from approximately 0 to 1.

The chimney cube is used in the study of petroleum systems. Interpretation of fluid migration paths involves studying spatial relationships between chimneys, source rocks, reservoir traps, faults, hydrocarbon indicators (DHIs) and seepage-related features such as pock-marks and mud-volcanoes. The seismic evidence is combined with regional geological knowledge, well data, pressure data, basin models, geo-chemical measurements and other relevant information in an integrated study of the petroleum system. Since the first publications on chimney cubes (Meldahl et al., 1999 and Heggland et al., 1999) many cubes have been processed and interpreted around the world. Successful applications, revealing vertical hydrocarbon migration pathways between source, reservoirs and the seabed, fault seal analysis and prospect ranking have been reported by Heggland et al. (2000), Meldahl et al. (2001), Aminzadeh and Connolly (2002), Connolly et al. (2002), and Ligtenberg and Thomsen (2003).

The main purpose of this paper is to present an improved method for seismic object detection. Our objective is to enhance both classifier performance and the resolution of the final image. We demonstrate our method on a chimney detection problem. However, our method (like Meldahl et al.,

1999) can be applied to any seismic object by providing a set of locations labeled by an expert as "object" and "non-object" and tuning the input attributes for the classifier. In order to rank the relevant importance of each seismic attribute in the classification problem, we use regularized discriminant analysis (RDA) with forward and backward search strategy. This allows defining a rank for each seismic attribute that efficiently results in lower combined classification errors. Two well-known non-linear classifiers, namely multilayer perceptron (MLP) and support vector classifier (SVC) are used to find output posterior probabilities of chimney and non chimney class separately. These classifiers have different properties that become evident in their corresponding chimney prediction results. This implies that the different natural characteristics for finding multi dimensional hyper-plane boundary will appear in their output. In order to have a mixed sense of both, the stage of classifier combining will apply with three mean, minimum and maximum logical rules.

2   Attribute selection and feature extraction

Seismic attributes that are generated from the seismic data highlight special information relative to the propagated wave field. From a pattern recognition point of view, each computed seismic attribute is called a "feature". The procedure for finding appropriate features consists of two separate parts. Firstly a geophysicist has to choose an initial set of attributes from a seismic point of view and secondly a statistical feature extraction algorithm is applied to reduce this set by minimizing some class separability measure, for instance the classification error.

In the first stage, the seismic attributes are selected based on experience and knowledge of interpreter. Chopra and Marfurt (2005) extensively discussed different ideas about cataloging seismic attributes. Taner et al. (1994) state a useful taxonomy for seismic attributes, i.e. physical versus geometrical ones. Physical attributes give information about the physics of wave propagation in subsurface (e.g. phase, frequency and amplitude), while the geometrical attributes underscore shape and geometry of the reflection events (e.g. dip, azimuth and continuity). For the purpose of seismic object detection, it is often necessary to consider both physical and geometrical evidences of the desired object. Thus a seismic interpreter should choose meaningful and sufficient attributes from both the above categories for the classification task. Although the tuning and exact definition of the set is data dependent, Tingdahl et al. (2001) introduced a set of attributes for chimney detection.

Although the set of attributes should contain all the information required for the detection of gas chimneys, individual attributes may be too noisy, or may be much correlated with other attributes, making them less informative when they are used in conjunction with the correlated attributes. To define a concise and non-redundant attribute set, feature extraction

techniques have been developed. The idea is to construct several subsets of the original features and to estimate a class-separability criterion on that. In order to estimate the separation, one typically has to have labeled data available. For this application it means that a seismic interpreter should provide some chimney and non-chimney pick locations. Given the criterion values for all feature subsets, one can choose the feature subset with a maximum class-separabiliy. The criterion that is used in this paper is the classification performance obtained by RDA (Friedman, 1989).

Assume a k (any arbitrary integer greater than 1) class problem and p seismic attributes with values of x_p; then, vector $X=[x_1, x_2, x_3..., x_p]$ is defined as the selected seismic attribute values in every seismic trace sample, μ_k as the mean vector for class k, C_k as the covariance matrix of the k-th class, and P_k as the prior probability for class k. The log-probability for object X_i for n available seismic labeled picks (where, $i=1,...,n$) under the assumption that the classes are Gaussian distributed is,

$$d_k(X_i) = (X_i - \mu_k)^T C_k^{-1}(\lambda, \gamma)(X_i - \mu_k)$$
$$+ \ln |C_k(\lambda, \gamma)| - 2 \ln P_k \qquad (1)$$

Where λ and γ are the regularization parameters that determine the added value to the diagonal of the covariance matrix and how much it will therefore deviate from the maximum likelihood solution. In practice, values of λ and γ should be found by optimization techniques.

When the amount of labeled picks is relatively small it is hard to obtain reliable estimates for (in particular) the covariance matrix due to its singularity in inversion procedure. For these situations one regularizes the covariance matrix by enlarging the diagonal of the maximum likelihood solution $\hat{\Sigma}_k$ and by adding a fixed constant to the diagonal of the unity matrix (**I**),

$$C_k = (1 - \lambda - \gamma)\hat{\Sigma}_k + \lambda \, \text{diag} \, (\hat{\Sigma}_k) + \gamma \mathbf{I}. \qquad (2)$$

To classify a seismic pick X_i, the log-probabilities of the classes are compared and it is assigned to the class with the highest log-probability:

$$\hat{y}(X_i) = k \text{ when } d_k(X_i) \geq d_l(X_i), \ \forall k \neq l, \qquad (3)$$

where $\hat{y}(X_i)$ is the estimated label for X_i. The final criterion value is the fraction of well-classified picks among all the picks that have been supplied by the seismic interpreter. Assuming each n labeled picks has the true label of y_i, ν defines a subset of features and that X_i^ν indicates that pick X_i is represented by this subset of features. Then the criterion value is defined as:

$$J_\nu = \frac{1}{n} \sum_{i=1}^{n} K.i, \qquad \begin{cases} K = 1; \, if \, y_i = \hat{y}(X_i^\nu) \\ K = 0; \, if \, y_i \neq \hat{y}(X_i^\nu) \end{cases} \qquad (4)$$

The strategy of searching the above criterion through all the possible feature subsets can be done with different methods.

We used two algorithms "forward" and "backward" to find rankings of the features based on the RDA criterion (van der Heijden et al., 2004). In the forward feature selection method, the initial subset is empty. Features are added one by one and the feature for which the criterion J_v is increasing most is added to the set. This proceeds until a pre-defined number of features are added, or until the criterion value does not improve anymore. In the backward method, first all attributes are used; then, they are removed one by one keeping the class separability as large as possible. Note that both approaches are not guaranteed to obtain the optimal solution. For finding the optimal solution in principle all subsets have to be tested. Because this is very time extensive in practice, these sub-optimal feature selection approaches are often taken.

3 Classifications

The next step is to classify data, i.e. finding a decision boundary between two classes. In this problem, we use a neural network and a support vector classifier as two outstanding non-linear classifiers. The scheme of the total classification algorithm is presented in Fig. 1. The application of neural networks in geosciences is discussed by some authors in recent years (Lees, 1996; van der Baan and Jutten, 2000; Aminzadeh and de Groot, 2006). A typical neural network classifier is the MLP, the mathematical idea of perceptron is introduced by Rosenblatt, 1958. Tuning and parameterization of a neural network is a hard task, as one must decide about so many parameters, like the number of hidden layers, number of units in every layer, initial weights, method of training, network architecture, momentum term, activation function of neurons and so on. For some of them, some suggestions are given in the literature, for example Hornik et al. (1989) discussed the point that adding extra hidden layers to a MLP is not very fruitful in the network performance. Jang et al. (2005) fully discussed different single and hybrid strategies for supervised training of adalines, multilayer perceptron, radial basis and modular networks. They also mention about the problem of having no constraint on nodes (except differentiability) of adaptive neural networks, their further attempts to define such necessary constraints even makes the network structure more complex. Still, tuning a neural network is a crucial issue that most often cannot be done fully in practice. Although the parameterization is very sensitive; but regarding MLP's smooth boundary, it is popular for different classification purposes.

Another classifier which is used in this study is SVC. It is aimed to maximize the geometrical margin between classes for the situations that classes are linearly separable. The complete mathematical formalization of SVC is discussed by many authors (Corres and Varpnik, 1995; Varpnik, 1995; Kecman et al., 2001). Considering N data samples (z_i), each with a label $y_i \in \{1, -1\}$, $i=1, ..., N$, assume that a lin-

ear classifier $g(z) = \boldsymbol{w}^T \boldsymbol{z} + b$ (b is a constant) is able to separate the set of data samples perfectly meaning:

$$\begin{aligned} \boldsymbol{w}^T \boldsymbol{z}_i + b &\geq 1 \text{ when } y_i = +1 \\ \boldsymbol{w}^T \boldsymbol{z}_i + b &\leq -1 \text{ when } y_i = -1 \end{aligned} \tag{5}$$

It can be shown that the margin between the classes is inversely proportional to the norm of $\boldsymbol{w}$. Therefore, to maximize the margin we should minimize $\boldsymbol{w}^T \boldsymbol{w}$. For non-linear separable data (like what we deal in seismic object detection) a "kernel trick" will be applied to the maximum margin hyper plane. This transforms data to a higher dimensional space and finds linear hyper plane there, while in original data space a non-linear margin will be constructed. (for more detail, Varpnik, 1995). In this paper, we used the so-called Gaussian (or radial basis) kernel. This transformation contains a free parameter σ that controls the smoothness of the transformation. For smaller values it gives very detailed and sharp boundaries and for larger values smoother ones will be obtained.

MLP and SVC both have some advantages and disadvantages. The MLP is a flexible classifier that can efficiently train on most data distributions. Because of its random weight initialization, its output is not identical after each run. Furthermore, when the number of training samples is limited, the MLP tends to overfit. It adapts its weights so far that it also fits the noise in the data perfectly. In practice, MLP network should stop in a particular training time to avoid biasing the result and loosing the generality. This implies that MLP classification error which is very near to zero on training data does not give sense while applying on test data. On the other hand, the SVC is a deterministic procedure and will always obtain the same solution when the training samples are not changed. It appears that by maximizing the margin between the two classes, the SVC overfits much less than the MLP. A drawback of the SVC is that it can basically only predict the output label, only +1 or −1. To obtain a confidence of the classification output, it is possible to fit a logistic function to the (linear) output of the SVC (Platt et al., 1999). In the experiments shown in this paper it appeared that the output probabilities are still relatively crisp, i.e. the SVC outputs are rarely around 0.5.

In order to use the power of both MLP and SVC, the idea of combining classifiers is helpful to complete classification task. Kuncheva (2004) mentioned combining idea as a natural step when a critical mass of knowledge from a single classifier model has been accumulated, but the final performance.[1] is not satisfactory yet. To exploit the value of this approach in seismic object detection, we used minimum, maximum and mean logical rules for combing the results of MLP and SVC. Minimum criteria select a class with

---

[1] Performance here is referred to ability to interpret the results in true physical domain (inline, crossline and time slice) as well as average output error.
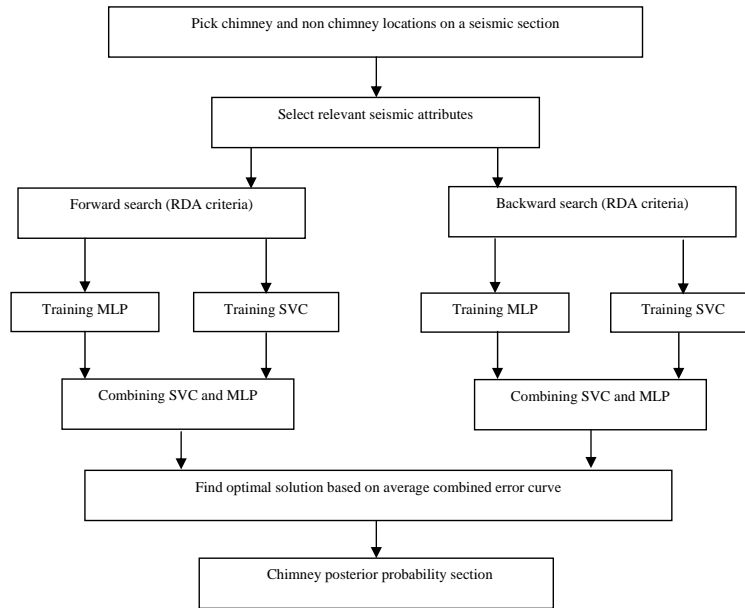
Fig. 1. The scheme of proposed classification algorithm.

Table 1. The output rank for seismic attributes with forward and backward search algorithms.

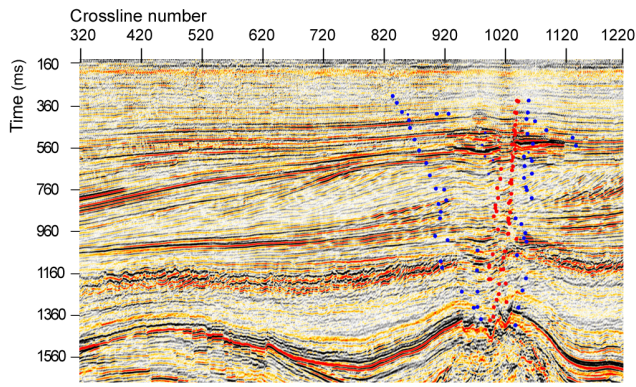| Seismic attributes (feature)	Forward rank	Backward rank
Reference time	7	5
Seismic	5	8
Energy (time window: $[-40, 40]$)	12	4
Similarity (time window: $[-120, -40]$, spatial trace positions: $(-1, 2) \times (1, -2)$)	3	9
Similarity (time window: $[-40, 40]$, spatial trace positions: $(-1, 2) \times (1, -2)$)	1	1
Similarity (time window: $[40, 120]$, spatial trace positions: $(-1, 2) \times (1, -2)$)	2	15
Similarity (trace window: $[-120, -40]$, spatial trace positions: $(1, 0) \times (0, 0)$)	13	16
Similarity (trace window: $[-40, 40]$, spatial trace positions: $(1, 0) \times (0, 0)$)	15	3
Similarity (trace window: $[40, 120]$, spatial trace positions: $(1, 0) \times (0, 0)$)	10	2
Polar dip	11	12
Curvedness	17	19
Curvedness (time shift: -80 (ms))	19	17
Curvedness (time shift: 80 (ms))	18	18
Seismic (low pass: 40 Hz)	8	10
Wavelet spectral decomposition (center frequency: 35 Hz)	6	14
Wavelet spectral decomposition (center frequency: 60 Hz)	4	7
Polar dip variance (time window: $[-40, 40]$)	14	11
Event asymmetry	16	13
Event zero crossing (negative-positive)	9	6

Fig. 2. A section from in-line 133 of F3 seismic data. Red picks are "chimney" and blue picks are "non-chimney" locations.
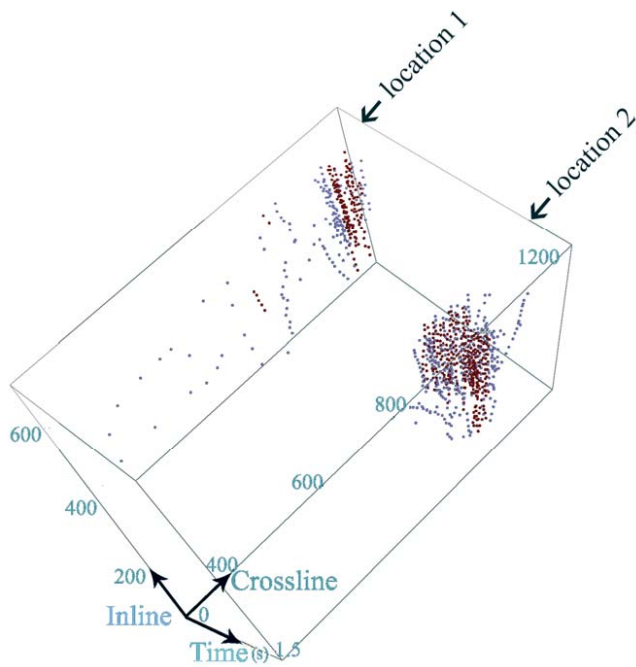

Fig. 3. Spatial and temporal distribution of pick locations in F3 seismic cube.

that gives the minimum output of the input classifiers, similarly two latter ones give maximum and mean output.

4   Experimental results

In this study, we used the seismic dataset from the F3 block in the Dutch sector of the North Sea. The presence of gas seepage is discussed in direct measurements (e.g. headspace gas analysis) of this area (Schroot, 2005). In the seismic data, there are evidences of wave scattering and loss of continuity. Meanwhile, it is not feasible to fully describe the shape of a chimney just on the seismic data or on a single relevant at-
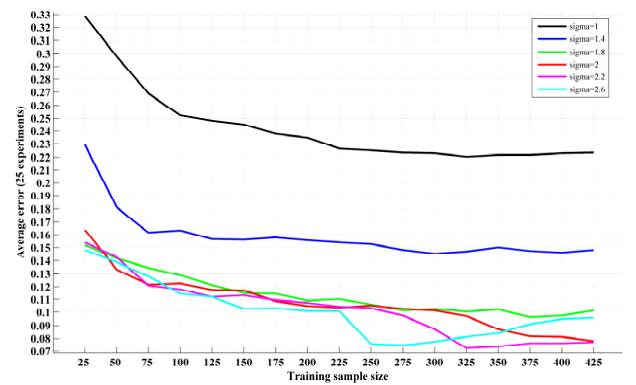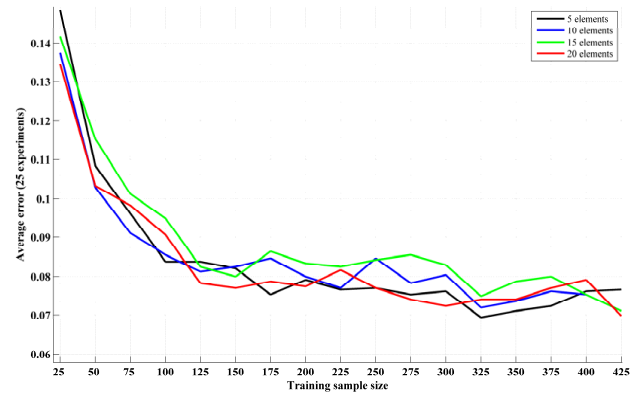

Fig. 4. Learning curve for different structures of MLP (top) and SVC (bottom) after 25 repetitions. There are two dominant linear trends in almost every diagram. This indicates having at least 150 objects for MLP and 75 for SVC are essential.

tribute. Figure 2 shows some locations in the seismic data labeled as chimneys (red) or as non chimneys (blue). In Fig. 3, the position of these picked locations in the original seismic cube is displayed and marked. We introduced 950 representative pick locations, with equal number of objects in each chimney and non-chimney class. In order to evaluate the generalization of trained classifiers in a proper way it is needed to have such a picking strategy. This shows training and evaluating classifiers on the same seismic data may cause a positive bias in the results even if the picks themselves are different. Through our experiment, using one spatial location for training and the other one for testing gives 2% higher average classification error with respect to the situation in which data from two locations are mixed with each other in the training and testing sets. We used the case in which picks from two different geometrical locations are mixed with each other and formed training and testing sets.

The results of the RDA criterion based on the forward and backward search algorithms for seismic attribute selection is shown in Table 1. The results are obtained after 50 cross validation tests within the training set which is a random subset (70%) of spatially mixed pick locations. In the other word,
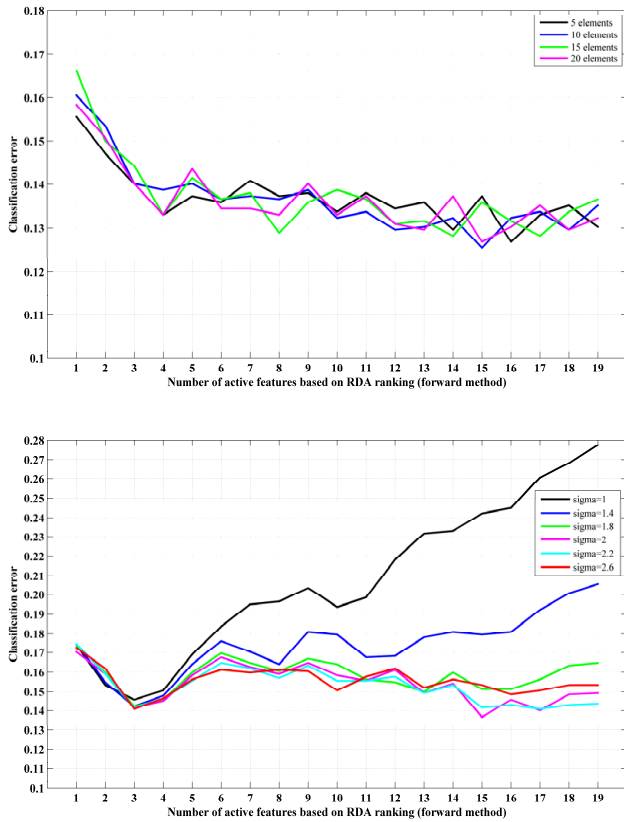
Fig. 5. Classification error based on forward search strategies for four MLP (top) and six SVC (bottom) structures versus number of active ranked features.

Fig. 6. Classification error based on backward search strategies for four MLP (top) and six SVC (bottom) structures versus number of active ranked features.

the routine is repeated 50 times within 670 chosen objects with random selection. As both search algorithms are suboptimal (a complete exhaustive search of all possible subsets of the seismic attributes is not practically feasible), the ranks obtained from backward and forward applications are not the same. In order to find a "best", but still suboptimal subset from the list, we will enter the attributes based on their ranks as the features for the MLP and the SVC.

The MLP structure used in this study is a feed forward architecture using the back propagation learning rule and one hidden layer with 5, 10, 15 and 20 elements. The target values for training are set to 0.1 and 0.9 to avoid over training. In the training phase of the back propagation procedure weight decay and the momentum rule are used for regularization. In our implementation of the SVC, the optimization of the radial basis kernel is done with the golden search algorithm and the parabolic interpolation for just on one feature space size (Brent, 1973). After determining the optimum sigma for the kernel width, 6 near sigma values are also used repetitively to evaluate the SVC on all possible feature space size. It is necessary to scale each input feature with respect to its variance in training and testing set for both MLP and SVC. Finally, a sigmoid function is applied on the SVC output optimizing
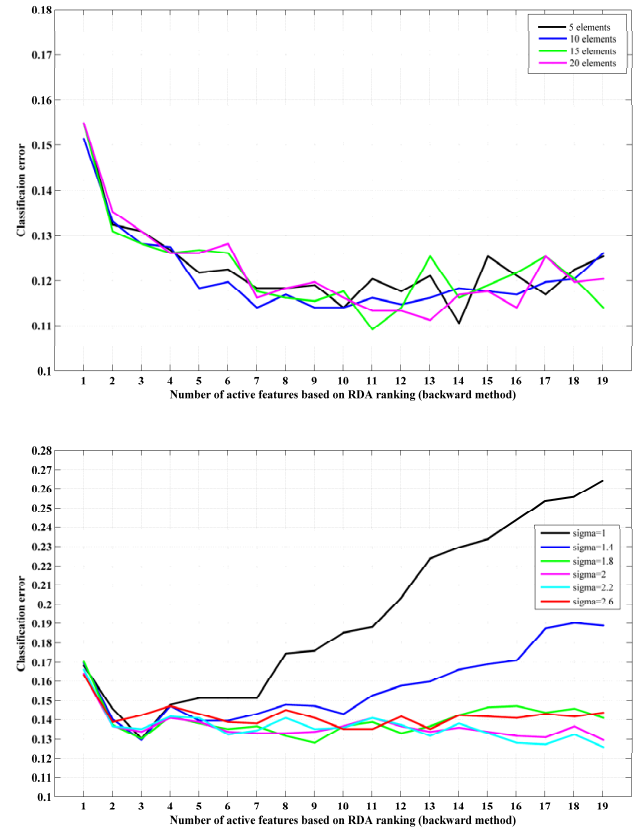
the likelihood of the posterior probabilities over the training set for achieving soft posterior probabilities.

Prior to building a final classifier, studying learning curves is a useful tool for judging the minimum number of required pick locations (training objects in the pattern recognition terminology). Figure 4 shows how increasing the number of training objects decreases the classification error of the MLP and the SVC (so called learning curves). Regarding two dominant apparent slopes, a promising minimum number of training objects is 150 pick locations for MLP and 75 for SVC.

Figures 5 and 6 show the average MLP and SVC classification errors versus ranked feature sizes based on the forward and backward selection procedures, respectively. These are computed over 5 repetitions of classification procedure with different random training sets. The idea for this repetition is to decrease noise in the classification error. The average classification errors of combining different structures of the MLP and the SVC are shown in Fig. 7. The role of feature space dimensionality is more evident here with respect to the single classifier case, so the 13 ranked features found by the forward algorithm are chosen as the optimum set of this object detection experiment.
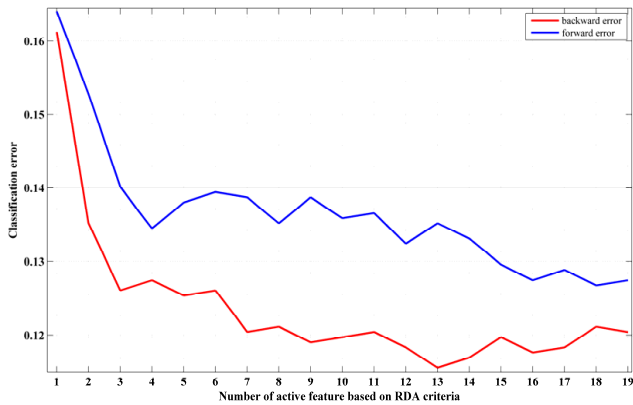
Fig. 7. Minimum averaged classification error for combined rules (minimum, maximum and mean) versus number of active ranked features for forward and backward methods. The figure shows backward ranking gives better performance rather than forward (i.e. lower error).

Figure 8 shows the posterior probabilities of the chimney class of the mentioned MLP and SVC structures. It is obvious that the output of the MLP is softer in the course of its variance inside the possible areas of "chimney" and "non chimney", whereas the result of the SVC is more likely to distinguish between areas with the same characteristics. The extra softness of the MLP makes tackling the near surface wave scattering ambiguity quite impossible, while the seemingly higher resolution of the SVC image helps to decide better in this part. As reported by Schroot (2005), this area is formed as a result of shallow gas packets. In the leaking reflector between time coordinates 1160–1360 (ms) with low continuity, the SVC gives slightly lower probability of chimney (yellow color) while the MLP reported it as high probable area. The result of the SVC is crisp inside the interested area of the chimney class (red color), which is softer in MLP one.

5 Discussion

The algorithm finally distinguishes seismic attributes with rank 14–19 in Table 1 based on the backward method to be excluded from the classifiers. This yields a better performance in a less complicated feature space. As stated earlier, performance means interpretability in physical domain as well as the average classification error. The corresponding average error of the combined classifier is 11.5%, which is acceptable regarding its corresponding MLP and SVC components. The average calculated error (Figs. 5 and 6) for the MLP with 20 elements is 11.1% and for SVC with the optimized kernel is 13.5% classification error on the final test set. As we mentioned above a random subset with the size of 70% of whole objects is devoted for the training and an independent test set with remnant 30% is used for testing the results.
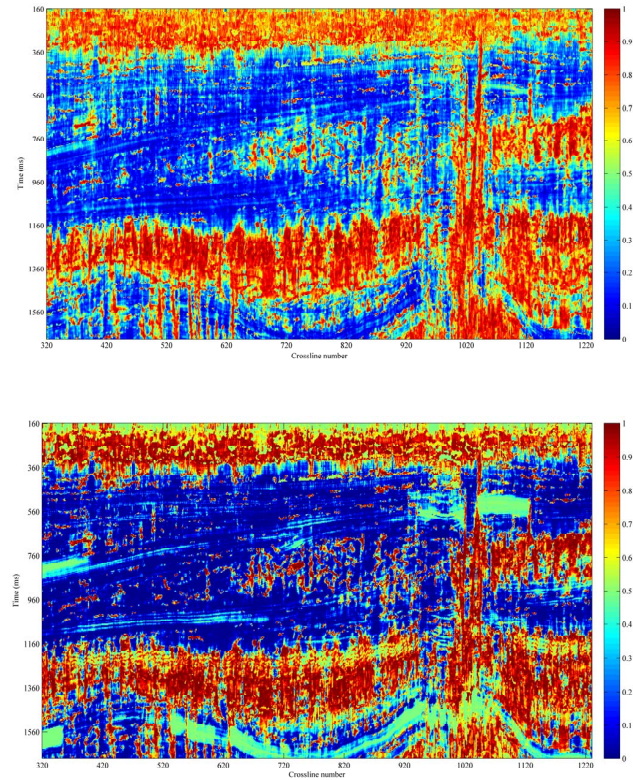

Fig. 8. Posterior probability of "chimney" class from MLP (top) and SVC (bottom). MLP has soft output with high chimney probability on leaking reflector (dark red), while the result of SVC is different for observed chimneys (red), high amplitudes (light green) and leaking reflector (yellow).

The second and most important parameter for evaluating the performance is the meaning of posterior probabilities in physical domain (confidences) and their consistency with the direct measurement experiments and other petroleum system intergradient's (e.g. fault cube, porosity, well logs).The confidences of the chimney class for the combined SVC and MLP classifiers by the above method are shown in Fig. 9. The minimum combining rule is a good choice, because it preserves the soft ability of a neural network in an appropriate manner. For this combiner, the extra softness of the blue area ("no chimney") is decreased while the softness of the red area ("chimney") is increased in comparison with the results of the MLP and the SVC. By the minimum combiner confidences, MLP output dominates inside the red area and the SVC mainly elsewhere. As a result, the minimum rule can highly constrain the softness of the MLP to a meaningful area. Mean and maximum combiner outputs are less useful as they have some disadvantages in proper imaging of the chimneys. Figure 10 compares the results of the MLP and the minimum combiner from a part of the seismic section, apparently in the case of resolution the minimum combiner shows better results in comparison with the MLP. The low
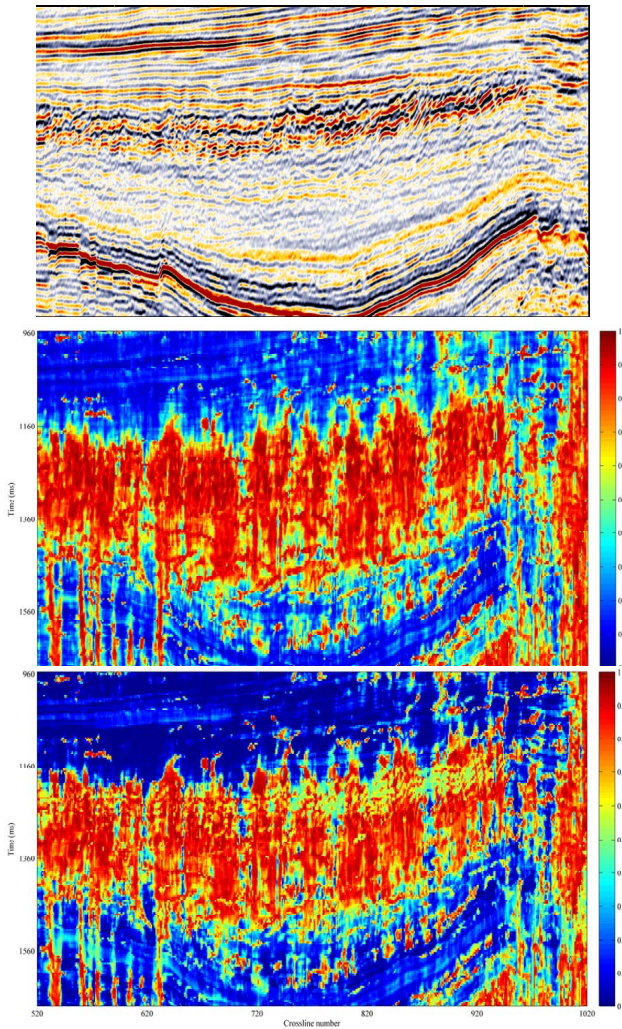
Fig. 9. Posterior probability of "chimney" class for combining MLP and SVC with minimum (top), maximum (middle) and mean (bottom) rules.
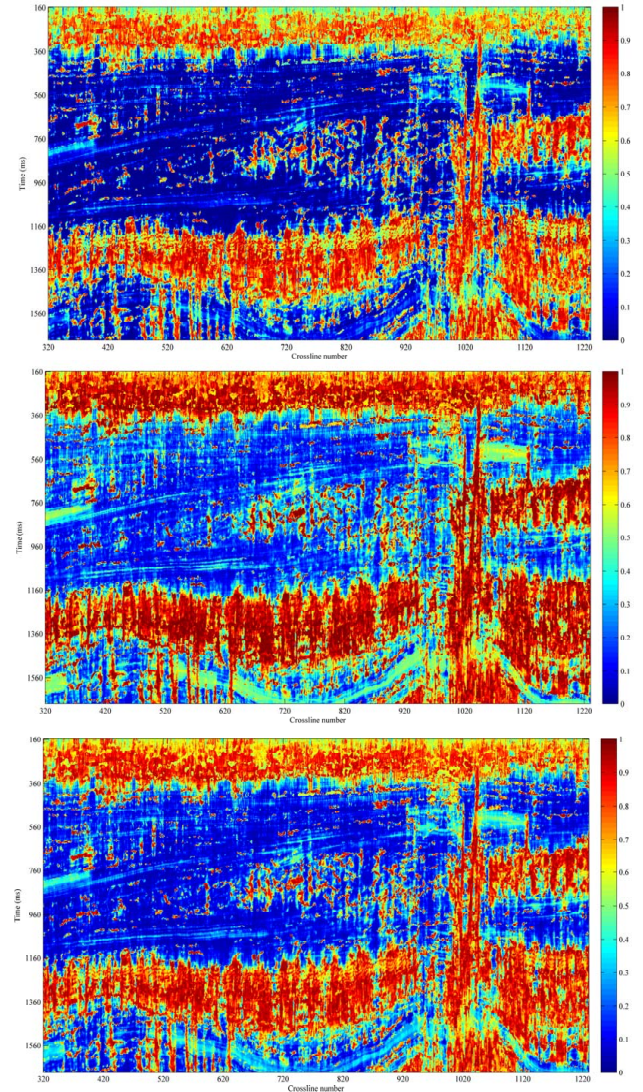

Fig. 10. Zoom section of seismic section (top), MLP section (middle) and minimum combiner section (bottom). In the bottom image, the leaky reflector (1160–1360 ms) is taken out from the most probable area (dark red) given by MLP. Resolution of combiner chimney section is more consistent with the original seismic section rather than soft MLP result.

coherent reflector between time coordinates 1160–1360 (ms) is taken out from the high probable area for chimney in the minimum combiner result, while the same area has a high chimney confidence in the MLP section.

6 Conclusions

Among the meaningful seismic attributes proposed by a seismic interpreter for the purpose of seismic object detection (chimney, fault, salt and ...), the user implicitly favors feature ranking to the classification task with "object" and "nonobject" picks. On the other hand, the classification with two potential non-linear methods (MLP and SVC) provides two different results consistent with their strategies: the MLP can handle very well overlapping class domains while the SVC searches for a "gap" between the classes. Combining is a hy-

brid tool for finding the lowest average error for an optimized feature space dimensionality and also using different strategies. It is concluded that a realistic image which is based on the softness of the MLP and the higher resolution of the SVC is obtained. The system is valuable especially when the interpreter does not have any insight for choosing the best attribute set for a specific seismic object detection problem. It is just needed to pick the suspicious locations on seismic data or one of the attribute sections, the algorithm then can suggest the most optimum attributes. It also guarantees to use both intrinsic property of MLP and SVC in an appropriate

way. We plan to analyze other seismic objects using the proposed algorithm in future studies.

Edited by: O. Talagrand
Reviewed by: D. Connolly and another anonymous referee

References

Aminzadeh, F. and Connolly, D.: Looking for gas chimneys and faults, AAPG Explorer, 23(12), 20–21, 2002.

Aminzadeh, F. and de Groot, P.: Neural networks and other soft computing techniques with applications in the oil industry, EAGE Publications BV, Houten, 161 pp., 2006.

Chopra, S. and Marfurt, K. J.: Seismic attributes: A new historical perspective, Geophysics, 70(4), 3SO–28SO, 2005.

Connolly, D. L., Aminzadeh, F., de Groot, P., Ligtenberg, H., and Sawyer, R.: Gas chimney processing as a new exploration tool: A west Africa example, Proc. of AAPG Annual meeting, Houston, 10–13 March 2002.

Corres, C. and Varpnik, V.: Support vector networks, Mach. Lear., 20(2), 165–194, 1995.

Friedman, J. H.: Regularized discriminant analysis, J. Am. Stat. Assoc., 84, 165–175, 1989.

Heggland, R., Meldahl, P., Bril, B., and de Groot, P.: The chimney cube, an example of semi-automated detection of seismic objects by directive attributes and neural networks: Part 2: Interpretation, SEG, Houston, Expanded Abstracts v. 1, 935–940, 1999.

Heggland, R., Meldahl, P., de Groot, P., and Aminzadeh, F.: Seismic chimney interpretation examples from the North Sea and the Gulf of Mexico, American Oil & Gas Reporter, February issue, 78–83, 2000.

Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural Networks, 2(4), 359–366, 1989.

Jang, J. S. R., Sun, C. T., and Mizutani, E.: Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence, Pearson Education Ltd., Delhi, 614 pp., 2005.

Kecman, V., Arthanari, T., and Hadzic, I.: LP and QP based learning from empirical data, IEEE Proceedings of IJCNN, 4, 2451–2455, 2001.

Kuncheva, L. I.: Combining pattern classifiers: methods and algorithms, John Wiley & Sons Inc., Hoboken, New Jersey, 374 pp., 2004.

Lees, B. G.: Neural network applications in the geosciences: an introduction, Computers and Geosciences, 22(9), 955–957, 1996.

Ligtenberg, J. H. and Thomsen, R. O.: Fluid migration path detection and its application to basin modeling, EAGE 65th conference, Stavanger, Expanded Abstract, C27, 4 pp., 2003.

Meldahl, P., Heggland, R., Bril, B. and de Groot, P.: The chimney cube, an example of semi-automated detection of seismic objects by directive attributes and neural networks: Part 1: methodology, Annual Meeting, SEG, Houston, Expanded Abstracts v. 1, 931–934, 1999.

Meldahl, P., Heggland, R., Bril, B., and de Groot, P.: Identifying faults and gas chimneys using multiattributes and neural networks, The Leading Edge, 20, 474–482, May 2001.

Platt, J. C., Smola, A., Bartlett, P., Scholkopf, B., and Schuurmans, D.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, advances in large margin classifiers, MIT Press, Cambridge, 61–74, 1999.

Rosenblatt, E.: The Perceptron: a probabilistic model for information storage and organization in the brain, Psychol. Rev., 65(6), 386–408, 1958.

Taner, M. T., Schuelke, J. S., O'Doherty, R., and Baysal, E.: Seismic attributes revisited: 64th Annual International Meeting, SEG, Los Angeles, 1104–1106, 1994.

Tingdahl, K. M., Bril, A. H., and de Groot, P. F.: Improving seismic chimney detection using directional attributes, J. Petrol. Sci. Eng., 29, 205–211, 2001.

Schroot, B. M., Klaver, G. T., and Schuttenhelm, R. T. E.: Surface and subsurface expression of gas seepage to the seabed-examples from the southern North Sea, Mar. Petrol. Geol., 22(3), 499–515, 2005.

van der Baan, M. and Jutten, C.: Neural networks in geophysical applications, Geophysics, 65, 1032–1047, 2000.

van der Heijden, F., Duin, R. P. W., de Ridder, D., and Tax., D. M. J.: Classification, parameter estimation and estate estimation: an engineering approach using Matlab, John Willey & Sons Inc., West Sussex, 2004.

Varpnik, V.: The nature of statistical learning theory, Springer-Verlag, New York, 314 pp., 1995.