Logical considerations in applying pattern recognition techniques on seismic data: Precise ruling, realistic solutions

Hosein Hashemi
Institute of Geophysics, University of Tehran, Tehran, Iran

Abstract

In this paper, comprehensive method and necessary considerations in seismic pattern recognition procedures are reviewed. In pattern recognition experiments, it is essential to decide about many parameters like picking representative objects with all possible spatial variations, minimum size of training set, knowledge on prior probabilities, seismic attribute redundancy reduction, choice of classifiers and posterior probabilities combinations. In order to follow the method better, two examples of seismic pattern recognition are discussed.

Introduction

The issue of categorizing data to different classes has significant and wide ranging of applications in analysis and interpretation of seismic data. Specialists are partly involved with these methods from first stages of seismic processing like automatic first break picking to final outputs such as AVO classification and seismic object detection. Lees (1996) and van der Bann & Jutten (2000) argued about applicability of neural networks in geophysical studies. Glinsky et al. (2001) used a trained probabilistic neural network for voxel classification using event times, subsurface points, and offsets. Meldahl (2001) showed the gas chimney representation in neural network on a standard set of input directive seismic attributes.

Different architectures of multilayer perceptron neural networks are described as good nonlinear classifiers, however they are not the only powerful ones in pattern recognition discipline. In order to accomplish a seismic classification routine with realistic solutions one shall consider the fact that the choice of classifier is just one important part of seismic pattern recognition procedure.

Labeling in physical domain

The first step in both seismic object detection and automatic first break picking is to choose a representative set. In recent years, geophysicists mostly practice problems with two classes e.g. fault vs. not a fault, first break vs. not a first break. This implies that prior to classification, a group of data shall be labeled as class 1 or class 2 in physical representation (i.e. seismic display). The theme of supervised learning is often used to describe learning classifiers based on labeled data. Each of the following procedures is calculated based on this representative labeled dataset, so choosing labeled dataset is a very sensitive task. It is highly recommended to choose data points with same desired class characteristics but possibly with different seismic responses and various inline/crossline locations. Subsequently trained classifier generalizes and adapts well to unseen data with different physical and geometrical characteristics but with the same seismic pattern specification. Figures 1 (a) & (b) show such a picking strategy in a seismic cube for the purpose of gas chimney detection and a seismic line for first arrival determination, respectively. It should be noted that the training set always forces a desired kind of output even if all the coming considerations are satisfied or not. This means the most important part of the seismic pattern recognition practice is choosing a well-defined representative set.

Role of prior probabilities

In seismic pattern recognition, the issue of considering prior knowledge is important. This is occurrence probability of a desired class that is already justifiable and lucid before solving the real classification. In most experiments it is not easy to set precise class prior probability, but on the other hand for others it is quite straightforward. In order to understand better first break picking and seismic object detection examples are considered again. Choice of a data point in first break picking is equivalent to selecting one sample from the total ensemble samples in a linear moveout (LMO) applied shot gather. This implies,

prior probabilities=1/(Number of samples in gather) × 100%.

On the other hand, it is not easy to guess the percentage of data samples that could be representative for a specific seismic object class such as a gas chimney, fault or salt dome in a seismic section. Wrong prior probabilities estimation yield to inaccurate posterior probabilities. Lawrence et al. (1998) explained the importance of using current prior probabilities in MLP networks. A wise choice in dealing the problems with unknown prior knowledge is to choose the class frequency evenly. Practically in a two class seismic object detection routine (like chimney and non-chimney), picking the same population of data points (objects) enables the classifier to act freely on the training dataset. The classifier primarily looks to each sample with the same possibility of belonging to each of two classes.

Minimum size of training set

The minimum size of the training set may change due to inherent complexity of the seismic pattern, type of classifier and different seismic attributes. Alternatively, using a lot of input picks to start classification exceeds the risk of wrong picking and increase the run time of the computer program. Figure 2 shows the behavior of misclassification percentage error versus the size of the training set in the gas chimney detection (study or example) using four different classifiers. These curves are well known as learning curve and are applicable in finding the minimum needed number of inputs.

Continued on Page 47

Logical considerations in applying pattern recognition...
Continued from Page 46

(a)

(b)

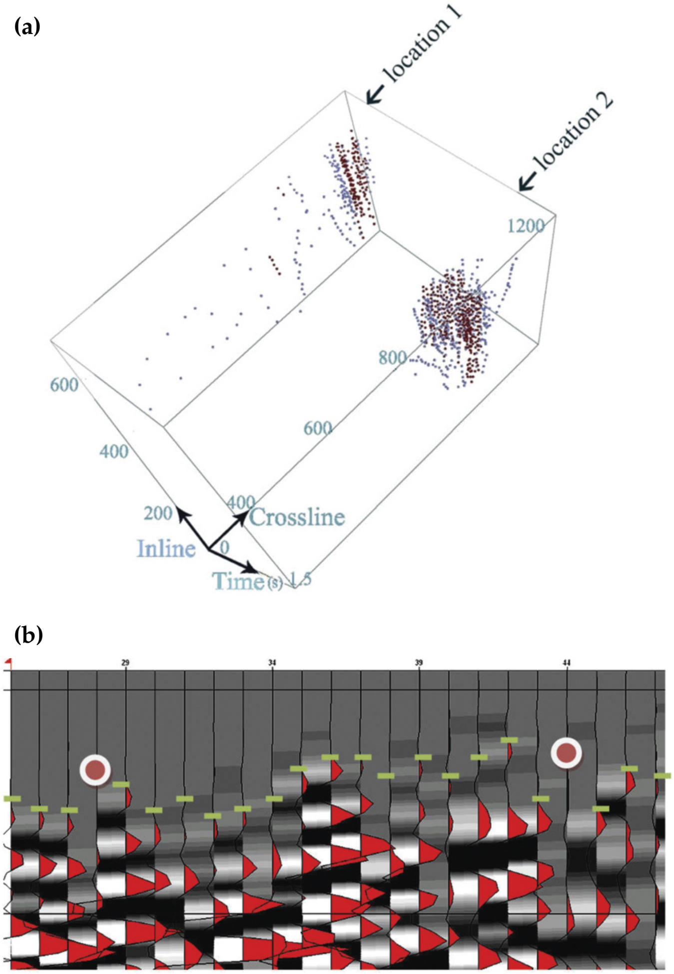

Figure 1. (a) Method for labeling seismic data in a 3D cube with no spatial preference in two separate locations. Training and test set for classifier learning procedure are chosen from mixed information of two locations. Blue picks represent chimney class while red one represents non-chimney class. (b) Method for labeling (manual picking) of a LMO shot gather. To form a representative training set, picks are expected to have different seismic amplitudes, but same class characteristics. Traces with red circles shall not be picked in any of the classes as a result of their reversed polarity. Picking traces with red circle lower the final classification performance.

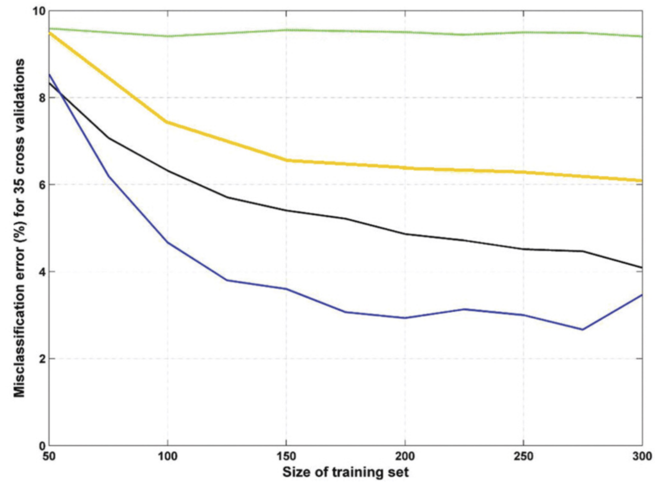

Figure 2. Learning curve for different classifiers. Dark Blue (support vector machine with radial kernel), Black (MLP neural network), Yellow (Linear discriminant analysis), Green (Nearest mean classifier). It is obvious that totally support vector can reach to the lower classification error. The minimum size of training set is where the steep trend is changed to horizontal one.

Nearest mean classifier (NMC), linear Bayes discriminant classifier (LDC), multilayer perceptron neural network (MLP), support vector machine with radial kernel (SVMRK) are used here. In order to avoid random errors, its average is computed on 35 repetitions using a cross-validation technique. Due to intrinsic mathematics behind each classifier (Webb, 2002), overall error trend does not follow a common behavior. The minimum number of required objects is where the error line flattens. For SVMRK, the minimum required objects is 120; for LDC, it is 150; for MLP, it is about 200 and it seems that error for NMC isn't affected by changing the training set size. Two immediate conclusions from figure 2 are inapplicability of NMC as a good classification method due to its relatively high error and preference of SVMRK in situations that extending confident representative training set for a specified seismic object is not easy.

Selecting relevant seismic attributes

In seismic object detection, relying on amplitude information in finding the exact location and acoustic response of a desired body does not suffice. One shall construct sections of different mathematical, physical and geometrical attributes to have better insight into the dilemma. On the other hand, redundant and useless seismic attributes shall be omitted before starting the classification (Hashemi & Javaherian, 2009). One simple reason is that similar (or quite similar) seismic attributes are not informative and they simply increase the dimension of the feature space and hence increase classifier complexity. The simple statistical tool for exploiting the relevance of the input features are scatter plots of principal component analysis (PCA) for different attributes. In this study, different seismic attributes are computed at each pick locations. A simple tool for finding good attributes is visual quality control of PCA scatter plots. Figure 3 illustrates PCA's of prominent seismic attributes in the gas chimney detection example. Those attributes which have better class separation shall be kept in feature space. It is obvious that time and energy are better in the sense of class separation; meanwhile all of these are selected as good attributes for gas chimney detection from a larger set.

Another criterion based method for selecting relevant attributes is training a non-linear fast classifier on *n*-D dimension (*n* is the number of calculated seismic attributes) feature space and use a search strategy to remove (or enter) the most relevant attributes to the feature space. Hashemi et al. (2008) discussed the issue of attributes redundancy reduction with such a technique. The algorithm stops upon reaching an optimum solution of relevant seismic attributes based on input seismic representative picks.

The procedure of first break picking relies on one attribute (feature) that is seismic amplitude. This simply means there is no need to go for feature extraction. Interested researchers may use various attributes like windowed energy in first break picking in conjunction with seismic amplitude.

Choice of classifier and combination

Reviewing previous works in seismic pattern recognition, most authors used classifiers in artificial neural network group like MLP, RBF, etc. Choice of classifier is a crucial step to be decided.

Continued on Page 48

Logical considerations in applying pattern recognition...

Continued from Page 47

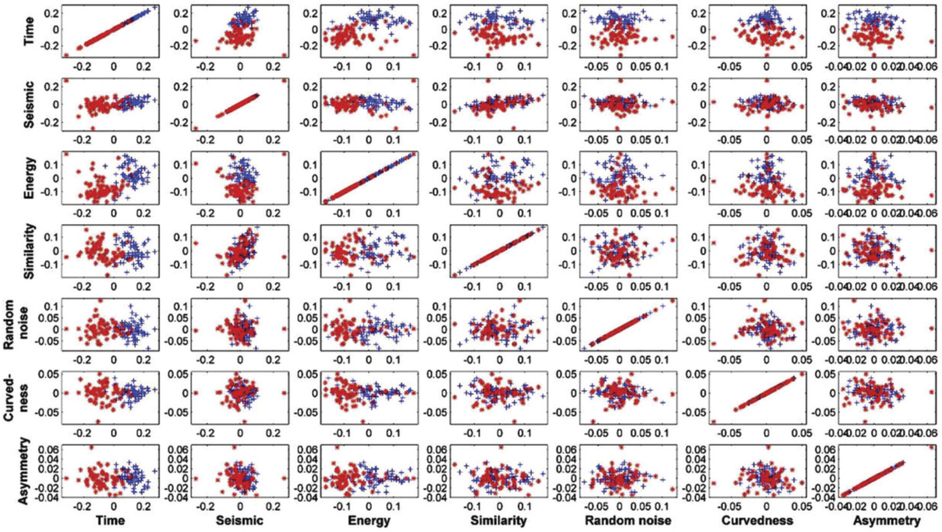

Figure 3. PCA scatter plots of selected five important attributes in gas chimney detection. Two class of chimney (red points) and not-chimney (blue points) are shown. Although major class overlap is completely visible, there is still enough class separation between some attributes. Relevant attribute pairs show higher-class separation with the others like time & energy.

Meantime, leaving other concepts discussed in this paper, it is highly probable that any powerful classifier fails in its output's performance. Two powerful nonlinear classifiers are support vector machine regularized with kernel trick and multilayer perceptron (Webb, 2002). This is also in accordance with averaged errors reported in figure 2. The essence of support vector classifier's architecture is like perceptron and both are linear, assuming separable data. It is visible from scatter plots in figure 3 that such a linear separation never happens in the complex seismic feature space with huge amount of class overlap. In order to accomplish this issue, either multiple perceptrons or support vector machine with radial kernel is used to find nonlinear discriminant function between classes. MLP have different initial weights in each run and due to its nature come with several output while SVMRK chooses one particular solution: the classifier which separates the classes with maximal margin.

When the knowledge from one classifier accumulates and becomes unchanged by increasing learning time, combining different ones leads to more logical answers in the sense of minimizing misclassification error and maximizing physical realization of the output. Hashemi et al. (2008) used stack combination of support vector classifier and multilayer perceptron neural network.

Posterior probabilities

In both seismic object detection and first break picking, the final stage is to map unlabeled data to classes using trained classifier(s). The degree on which a data object belongs to a specific class is called its posterior probabilities or classifier's output. Figure 4 (a) shows the output of MLP classification without considering a learning curve, seismic attribute selection and combining knowledge of different classifiers. In Figure 4 (b), all the steps are included and combination of MLP & SVMRK is illustrated. The main problem is to detect a or the gas chimney(s) in F3 block of North Sea data. The presence of gas is discussed with Schroot et al (1999) and other authors in this area. By definition, gas chimneys are vertical or semi-vertical subsurface paths for gas, they are useful in locating fluid migration pathways and hence minimizing risk while drilling.

Near surface cracked zones are mostly interpreted as high probability areas for gas chimneys in figure 4 (a), while in figure 4 (b) taking to account considerations of an integrated pattern recognition system is not included. Moreover, resolution and particularly vertical coherency is higher in figure 4 (b) which makes sense when fluid migration process is active in a reservoir. The misclassification error for MLP is 0.5% higher than the combinations of MLP&SVMRK, but their different posterior probabilities are mainly due to differences in redundancy reduction, classifiers specific strategies, considering learning curve and prior probabilities. This implies that most often sticking to the fact of lower misclassification error is not the only way forward to finalize the seismic pattern answer. The posterior shall satisfy the geophysical constraint and definitions of the problem under study.

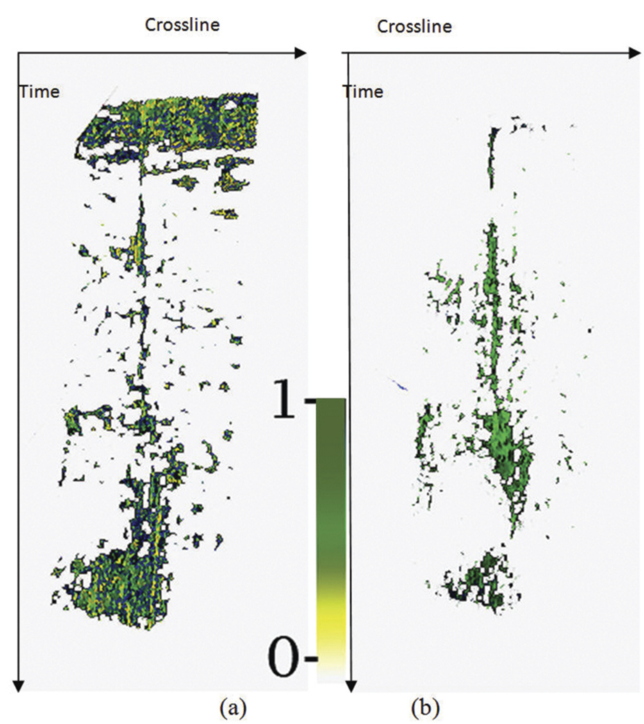

Figure 4. An overlay of gas chimney class posterior probabilities for successive crosslines in F3 data of the North sea, a) MLP classifier with no consideration in training set size, feature extraction and classifier combination; b) combination of SVMRK and MLP using optimal training set size using feature extraction technique. Shallow non-vertical high probabilities of (a) are not in accordance with definitions of gas chimney. Vertical directivity and higher resolution of posterior probabilities in (b) is evident. In order to visualize better a transparent color map is used in overlay mode. The color bar is same in two figures.

Continued on Page 49

Logical considerations in applying pattern recognition…

Continued from Page 48

Conclusions

It is discussed that there are some serious considerations that must be made in seismic pattern recognition to have reliable answers. The choice of classifier that is mentioned by several researchers is just one of those. It is discussed in this paper that each classifier yields a different error trends. Performance of the method is not just defined by statistical term i.e. misclassification error, the geological interpretation of posterior probabilities has equivalent importance. *R*

Acknowledgements

This research is facilitated by having F3 block data from dGB Earth sciences. I would like to acknowledge Dr. Paul de Groot and his team for facilitating access to data and developing Opendtect (an open source seismic interpretation package) that is partly used for generating seismic attributes and visualizations of this research. Many ideas of the current work is developed and shaped within the useful discussions with Dr. Robert Duin and Dr. David Tax from ICT department of University of Delft. I also acknowledge anonymous reviewers for their useful comments.

Hosein Hashemi is currently a faculty member in the Institute of Geophysics, University of Tehran. He received his Ph.D. in exploration seismology from the University of Tehran in 2009. He completed his M.Sc. in exploration seismology from the same faculty in 2004 and holds a B.Sc. degree in applied physics from the University of Isfahan (2001). To complete his Ph.D. project, Hosein stayed at Delft University, Electrical Engineering and computer science department, in the Information and communication theory group (2007-2008). There he studied the applicability of pattern recognition techniques on seismic data with a group of computer science specialists.

During that time, he also got the chance to work in the industry as a seismic processer and interpreter at Dana Geophysics Co. (2005-2008) and subsequently at Cadcam Iran PJS Co. (2008-2010).

References

Glinsky, M. E., Clark, G. A., Cheng, P. K. Z., SAndhya Devi, K. R., Robinson, J. H., Ford, G. E., 2001, *Automatic event picking in prestack migrated gathers using a probabilistic neural network,* Geophysics, 66, 5, 1488-1496.

Hashemi, H., Tax, D.M.J., Duin, R.P.W., 2008, Javaherian, A., and de Groot, P., *Gas chimney detection based on improving the performance of combined multilayer perceptron and support vector classifier,* Nonlinear Processes in Geophysics, 15, 863–871.

Hashemi, H. and Javaherian, A., 2009, *Seismic Attribute Redundancy Reduction Using Statistical Feature Extraction Technique,* 1st International Petroleum Conference & Exhibition, EAGE, Shiraz, Iran.

Lawrence, S., Burns, I., Back, A., Tsoi, A. C., Giles, C. L., 1998, *Neural Network Classification and Prior Class Probabilities, Lecture notes in computer science,* Springer Berlin / Heidelberg, 1524, 545.

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

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

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

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

Webb, A. R., 2002, *Statistical pattern recognition,* John Wiley & Sons Inc,. West Sussex, UK.