Pattern recognition in spatial data: A new method of seismic explorations for oil and gas in crystalline basement rocks

Vadim Mottl, Sergey Dvoenko
Tula State University, Russia
mottl@atm.tsu.tula.ru
dvoenko@atm.tsu.tula.ru

Vladimir Levyant
Central Geophysical Expedition
Moscow, Russia

Ilya Muchnik
Rutgers University, USA
muchnik@dimacs.rutgers.edu

Abstract

The problem of prospecting oil and gas reserves in the crystalline basement of the Earth mantle by way of a combined interpretation of seismic data registered on the daylight surface and direct information from a sparse net of exploratory wells is considered as pattern recognition problem in which the role of objects whose class membership is to be recovered is played by points of the three-dimensional underground medium. Local properties of reflected seismic signals serve as features of the membership of the respective rock mass zones in the class of collectors, i.e. spatial areas capable of accumulating fluids, whereas direct data obtained from exploratory wells serve as trainer's information. A new spatial approach to supervised pattern recognition is proposed which makes use of the fact that objects to be recognized are arranged in an array in space. Along with the additional assumption that immediately adjacent points offer a tendency to belong to the same class, this fact allows for drawing reliable decisions from relatively unreliable features.

1. Introduction

All the classical statements of the pattern recognition problem deal with indivisible objects each of which is assumed to belong, as a whole, to one of a finite set of classes. The observer has to discover the hidden class membership of the given object by a vector of its features.

However, there are many practical problems in which it is required to make a decision on the classes, at once, of all the objects arranged in an array under the additional a priori assumption that immediately adjacent objects offer a tendency to belong to the same class. It is clear that the availability of a priori information on the interdependence of classes in the array must contribute essentially to the accuracy of recognition in comparison with the case when the classes are independent.

In this work, we consider a practical problem in which objects of recognition are arranged in an array in multidimensional space. This is the problem of prospecting oil and gas reserves in the so-called crystalline basement of the Earth mantle by way of a combined interpretation of direct information on the location of reservoirs obtained from a rare net of highly expensive exploratory wells and relatively cheap spatially complete seismic data registered on the daylight surface. Local properties of reflected seismic signals serve as features of the membership of the respective rock mass zones in the class of collectors, i.e. underground areas capable of accumulating fluids, whereas direct data obtained from exploratory wells serve as trainer's information.

The unknown combination of classes at the points of the medium being examined is assumed to be realization of a hidden Markov random field. But in contrast to the one-dimensional case, the spatial Markov assumption does not lead automatically to a simple recognition algorithm [1]. Therefore, we resort to an approximation of the lattice-like neighborhood relation in the array by a combination of tree-like ones, which allows for an effective algorithmic solution [2,4].

2. Seismic and drilled information on local properties of the rock mass

A seismic exploratory data set consists of synchronous records of reflected seismic signals registered by a large number of geophones (seismic sensors) placed in the nodes of a rectangular lattice on the Earth surface. As the source of the initial seismic energy, usually serves a series of explosions. After quite a complicated processing, it becomes possible to identify the time axis with depth under the respective sensor, so that the resulting seismic data array gives a three-dimensional model of the hidden space being studied, as it is shown in Fig. 1.

The sedimentary rocks forming the upper coat of the Earth mantle have served, up to now, as the main source of oil and gas. However, the reserves of hydrocarbons in this relatively thin layer covering the massive body of the crystalline basement are quickly getting exhausted. A significant increase of prospected oil and gas reserves in the coming century cannot be provided without assimilating the basement interval of the mantle thickness.

Geophones

Proceedings of the 15th International Conference on Pattern Recognition, September 3-8, 2000, Barcelona, Spain. Track B, pp. 315-318.

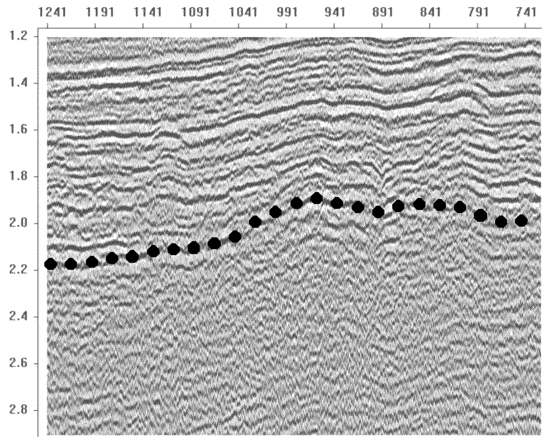

Figure 1. Fragment of a vertical seismic section through upper sedimentary cover and underlying crystalline basement.

These two intervals of the depth scale are essentially different from the viewpoint of hydrocarbon prospecting. The most preferable way of mining information on the structure of gas- and oil-promising objects are seismic explorations, that allow for examining the underground medium from the daylight surface without drilling highly expensive exploratory wells and give a one-to one correspondence between elements of the data array and points of the respective spatial interval. But the presently existing principles of seismic data analysis are essentially oriented to recovering quasi-horizontal reflecting boundaries between sedimentary rock strata, and are not applicable to the basement massif where stratification is practically absent.

One of the main geological factors that determine the ability of a basement fragment to serve as collector of fluids is its increased hollowness caused by presence of fractures and caverns. Because of the unavailability, at present, of less expensive means, the main source of information on the local hollowness of the basement rock mass are data acquired from exploratory wells. At the same time, the inevitably sparse net of wells does not allow for extending, with a sufficient reliability, the results of such point-wise estimates onto the entire volume of the potential reservoir.

The aim of this work is to fill in this gap. The central idea is to use the seismic image of the basement fragment for interpolation of the sparse drilled information on the location of collectors in the crystalline rock.

Results of our preliminary investigations say that the differences in physical properties of the basement rock cause marked distinctions in the three-dimensional local texture of the seismic picture around the respective elements of the data array [3].

Fractures, pores and caverns create in a crystalline massif a great number of microinhomogeneities, hence, an incident seismic wave will generate in such a medium a great number of micro-diffracted waves or scatter essentially. On the contrary, a monolith homogeneous crystalline body must be almost transparent in the seismic sense. So, the high-frequency component of the seismic field, or, in other words, its texture, must be different in monolith and substantially fractured zones.

Let $X = \big(x_t,\, t=(t_1,t_2,t_3)\in T\big)$ be the original seismic data array, where (t_1,t_2) are horizontal coordinates and t_3 is time interpreted as depth under the Earth surface. As mathematical model of the seismic texture, we use the spatial autoregression equation whose coefficients are assumed to be different at different points $x_t = \sum_{s\in S} a_t^s x_{t+s} + b_t \xi_t$, where S is a three-dimensional autoregeression mask and ξ_t is white noise with variance taken equal to unity. We estimated the autoregression parameters $\mathbf{a}_t = (a_t^s,\, s\in S;\, b_t)$ at each point of the space by a special technique [3], but it is enough to conventionally assume that they are found as constants in sufficiently small volumes. As features of the local spatial texture, we used four secondary parameters $\mathbf{y}_t = (y_t^1,...,y_t^4)$ calculated from the autoregression model, namely, the local irregularity of signal oscillations $y_t^1 = b_t$, full local variance of the signal y_t^2, energy of the seismic field in a narrow frequency band in the vertical direction y_t^3, and that in the horizontal direction y_t^4.

Having been quantitatively evaluated, the local texture characteristics $\mathbf{y}_t$ will show the basement areas with similar mechanical properties, which can serve as pathways of spatial interpolation of the drilled direct information on collector zones between the exploratory wells.

By its spirit, such an idea falls into the competence area of supervised pattern recognition, where the points of the basement medium fragment are considered as entities to be allocated over a finite number of rock types. Each point is characterized by the seismic image in some immediate vicinity of its match in the data array, whereas the part of trainer's data is played by the information on the rock type acquired from the exploratory wells only for a part of the points. It is required to extend the known local class membership of the points that happen just at the vertical wells drilled through the basement rock onto the rest of the spatial interval being examined.

However, the essence of the pattern recognition problem of such a kind substantially differs from the traditional classical version of this problem.

3. Problem and principles of spatial pattern recognition

Let T be set of elements $t\in T$ that are to be considered jointly. Let, further, $G\subset T\times T$ be an arbitrary undirected graph without loops interpreted as relation of immediate adjacency between elements $(t',t'')\in G$.

In our geophysical application, the role of $t\in T$ will be played by elements of the seismic data array which are associated with discrete points of the respective three-dimensional underground interval $t=(t_1,t_2,t_3)\in T=\{t_i=1,...,N_i;\ i=1,2,3\}$, and graph G will be taken in the form of the rectangular lattice that represents the natural neighborhood relation in space.

Let each element $t\in T$ be assigned a random index $k_t\in M=\{1,...,m\}$ of its hidden membership in one of m classes and random vector $\mathbf{y}_t\in R^n$ of its observable features. Hence, being considered jointly, these variables will form a two-component random field (K,Y), $K=(k_t,t\in T)$, $Y=(\mathbf{y}_t,t\in T)$.

In particular, in the problem of finding collector areas in the basement medium we shall deal with two classes of crystalline rocks, namely, collectors and non-collectors.

Let the probabilistic properties of the hidden component K be completely known to the observer in the form of a priori probabilities $p(K)$ for all the combinations of the hidden classes. Then, the individual (marginal) a priori probabilities of classes $q_t(k_t)$ are also known for each of the elements of the field $t\in T$. As to the observable feature vectors, they will be assumed to be conditionally independent $\varphi(Y\,|\,K)=\prod_{t\in T}\psi(\mathbf{y}_t\,|\,k_t)$ with the same unknown individual density $\psi(\mathbf{y}\,|\,k)$.

Our ultimate aim is to restore the class indices of all the elements by processing the entire field of their feature vectors $\hat{K}(Y)$ on the basis of information given by the trainer, who is assumed to have indicated the class memberships for a sparse subset of elements k_{t_j}, $j=1,...,N$. For this purpose, it is completely enough to infer, from the trainer's data, the a posteriori probabilistic properties of the hidden random field $\pi(K\,|\,Y)$, then the solution of the recognition problem can be found as result of its maximization

$$\hat{K}(Y)=\arg\max_K \pi(K\,|\,Y). \qquad (1)$$

From the theoretical viewpoint, the joint a posteriori probabilities are completely determined by the relation $\pi(K\,|\,Y)\propto p(K)\varphi(Y\,|\,K)$, but the intent to estimate $\psi(\mathbf{y}\,|\,k)$ would be a poor choice, because a posteriori probabilities of classes are, as a rule, much simpler functions of features than conditional densities in the feature space.

If we observe only one of the feature vectors $\mathbf{y}_t$, the a posteriori probabilities of the hidden class of this element are proportional to the a priori probabilities and conditional densities $p_t(k_t\,|\,\mathbf{y}_t)\propto q_t(k_t)\psi(\mathbf{y}_t\,|\,k_t)$. But if all the feature vectors $Y=(\mathbf{y}_t,t\in T)$ are observed at once, the conclusion on the class of an element should be inferred, on the force of the assumed probabilistic interdependence of class memberships, from the feature vectors of all the elements. Nevertheless, in accordance with the following theorem, the individual posterior probabilities $p_t(k_t\,|\,\mathbf{y}_t)$, $k_t=1,...,m$, contain all the required information.

Theorem 1. *The a posteriori hidden random field of class memberships is completely determined by the joint a priori and individual a posteriori probabilities, respectively,* $p(K)$ *and* $p_t(k_t\,|\,\mathbf{y}_t)$*:*

$$\pi(K\,|\,Y)\propto\frac{p(K)}{\prod_{t\in T}q_t(k_t)}\prod_{t\in T}p_t(k_t\,|\,\mathbf{y}_t).$$

In most applications, it is hardly reasonable to take a priori models $p(K)$ with different marginal probabilities of classes for different elements of the array. If the hidden random field is homogeneous in the sense that $q_t(k)=q(k)$ for all $t\in T$, we have also $p_t(k\,|\,\mathbf{y})=p(k\,|\,\mathbf{y})$, and the training problem for spatial data will differ not at all from its classical version for independent objects of recognition. It is sufficient to estimate, using the trainer's data, the vector function of a posteriori probabilities in the feature space $p(k\,|\,\mathbf{y})$, $k=1,...,m$.

We do not consider how to do this, the respective methods are commonly adopted. What is new here is the problem of recognition, which, actually, does not exist in the classical case, because the a posteriori probabilities $p(k_t\,|\,\mathbf{y}_t)$ immediately lead to the decision if the class-memberships k_t are independent in T.

If K is nontrivial random field, the hidden classes can be estimated only jointly $\hat{K}(Y)$. It is clear that the recognition procedure will essentially depend on the assumption about the a priory hidden field. In this work, we assume $p(K)$ to be a Markov random field with respect to the adjacency graph G in the sense that the conditional a priori probabilities of classes at an element $q_t(k_t\,|\,K_{(t)})$, where $K_{(t)}=(k_s,s\in T:\ s\neq t)$ is the rest of the field, depend only on those at immediately adjacent elements $q_t(k_t|K_{(t)})=q_t(k_t|K^0_{(t)})$, $K^0_{(t)}=\big(k_s,s\in T:\ (s,t)\in G\big)$.

In our application, the sought-for collector areas are expected to be stretched in space, what can be expressed in Markov terms by high values of $q_t(k_t\,|\,K^0_{(t)})$ if the current index k_t is the same as the majority of indices within the immediate vicinity $K^0_{(t)}$.

Theorem 2. *If the assembly of hidden variables* K *is a Markov random field with conditional local distributions* $q_t(k_t\,|\,K^0_{(t)})$ *in accordance with adjacency graph* G*, and elements of the observable random field* $(Y\,|\,K)$ *are conditionally independent with densities* $\psi(\mathbf{y}_t\,|\,k_t)$*, the a posteriori random field* $(K\,|\,Y)$ *is also a Markov one with respect to the same adjacency graph:*

$$\pi_t(k_t|K_{(t)},Y)=\pi_t(k_t|K^0_{(t)},\mathbf{y}_t)\propto q_t(k_t|K^0_{(t)})\,p_t(k_t|\mathbf{y}_t).$$

Thus, given the field of observable features Y, the local Markov properties of the a posteriori hidden random field $(K \mid Y)$ are immediately defined by the respective a priori conditional probabilities $q_t(k_t \mid K^0_{(t)})$ and local a posteriori ones $p_t(k_t \mid \mathbf{y}_t)$. But for the adjacency graph G of general kind, finding the combination of class indices $\hat{K}(Y)$ that maximizes the a posteriori probability (1) is a hard computational problem of global optimization.

As it was shown in our previous research [4], an algorithm of global optimization exists for the case when the adjacency graph is a tree. Therefore, in this work, we substitute the rectangular lattice of spatial neighborhood by a system of trees as it is shown in Fig. 2. In this case, the number of elementary operations required for finding the solution of the recognition problem is proportional to the number of elements in the data array.

4. Recognition of collector zones in the basement of the oil field Bombay High

On the basis of this approach, we attempted to forecast the location of collector zones in the crystalline basement of the oil field Bombay High on the western coastal shelf of India in the Arabian Sea. We took for the analysis quite an old data set obtained in 1972 as a system of vertical seismic cross-sections covering a part of the field in the form of a grid 5×7 km as it is shown in Fig. 3. The texture characteristics of the seismic data array were estimated in the cross-sections and interpolated at all the points of the three-dimensional fragment.

As the source of trainer's data, served 19 wells that reached the basement rocks. A group of experts indicated the boundaries of collector zones in each of the wells on the basis of studying log data, results of hydraulic testing, and the measured yields of fluids (oil, gas, or water) from some intervals of depth. The location of the wells and trainer's data at the depth of 300 m under the upper boundary of the basement are shown in Fig. 3.

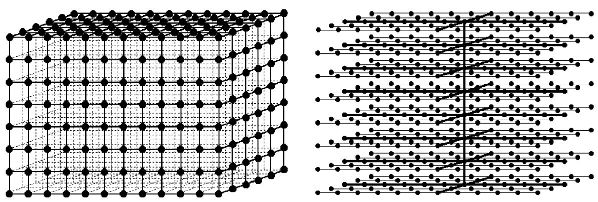

Figure 2. Graph of the spatial neighborhood between points of the underground medium and the tree used instead of the lattice for calculating marginal posterior probabilities of classes in one vertical row.

As result of training and recognition, we built a three-dimensional model of collector zones in the basement space. The shaded zones in Fig. 3 represent the map of the found spatial area at the same level of depth under the basement roof at which the trainer's data are shown. The

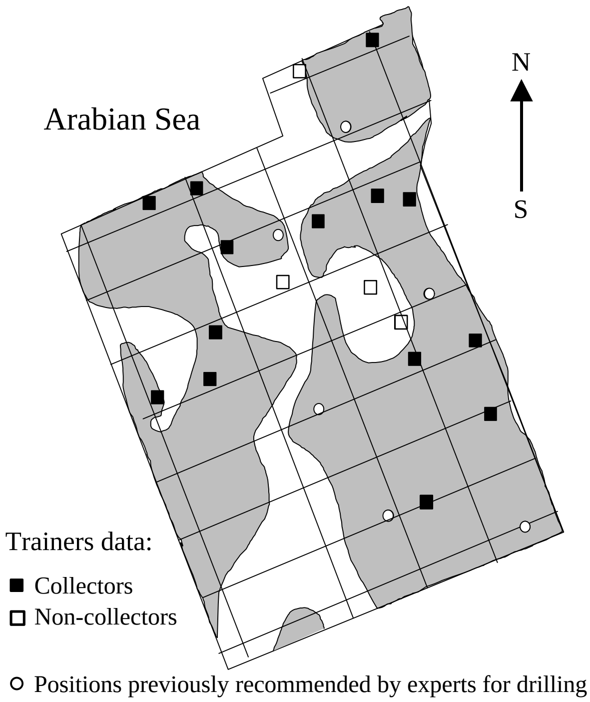

Figure 3. Trainer's data and map of estimated collector zones of the oil field Bombay High at the depth of 300 m under the basement's upper boundary.

map covers all the positions that had been previously recommended by experts for drilling from geological considerations [3].

References

[1] S. Geman, and D. Geman. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. on PAMI*, Vol. 6, November 1984, pp. 721-741.

[2] C.-H. Wu, P.C. Doershuk. "Tree approximation to Markov random fields", *IEEE Trans. on PAMI*, Vol. 17, No. 4, April 1995, pp. 391-402.

[3] V. Leviant, V. Mottl, N.K. Lal, R.S. Dirghangi, and T.R. Varun. "Experience of reservoir zone identification in crystalline basement of Bombay High field from seismic data", *Proceedings of the 3rd International Petroleum Conference PETROTECH-99*, New Delhi, India, January 9-12, 1999.

[4] V. Mottl, I. Muchnik. *Bellman functions on trees for segmentation, generalized smoothing, matching and multi-alignment in massive data sets*. DIMACS Technical Report 98-15, February 1998. Center for Discrete Mathematics and Theoretical Computer Science, Rutgers University, USA.