Bull. Nov. Comp. Center, Math. Model. in Geoph., 6 (2000), 131–136

On preference of pattern recognition algorithms in geosciences

Yu.A. Voronin, E.N. Cheremisina

> The real science starts not with application of mathematical simulation but with its reasonable application, i.e., in the absence of theory, unknown when.
>
> *A.A. Markov*

1. There are grounds to consider that pattern recognition in geological sciences [1] is a challenge to those engaged in applied mathematics, which was cast in the 50-s of the XX century by the practice of scientific investigations, to which no worth answer was found [2, 3]. As it seems, this is caused to a considerable extent by the fact that schemes and norms of reasonings and algorithms of applied mathematics which turned out to be not much appropriate outside theoretical physics [6, 11], are employed in pattern recognition [4, 5].

The application of the pattern recognition results indicates to the fact that up to now there has been no answer to such questions as: when pattern recognition has scientific sense and when not?, when such a sense can be given to pattern recognition and how to do this without changing the system of concepts and the scheme of data acquisition?

One can assume that the pattern recognition is, in a sense, not a construction of an optimal algorithm, but a construction due to accumulation of empirical data, of an algorithms sequence, each subsequent one being more preferable than the previous one. In other words, the pattern recognition is, first of all, the preference of pattern recognition algorithms.

The pattern recognition does not imply immediate substitution of special automated system, but their subsequent and substantiated ousting.

The pattern recognition should be finished with the evident evaluation of the advantages to be gained from the use of expensive and sophisticated automated systems instead of the conventional speculations as well as the estimation of losses due to a dishonest advertising of these automated systems [3].

2. Let a set of $A = (a_i)$, $i = 1, \ldots, N$, and a number of properties $F = (f_p)$, $p = 1, \ldots, m$, be given, where each property f_p takes q_p different values. Let us denote the value of property f_p by f_p^i on the element a_i of A, and let us assign to the elements a_i of A the vectors $F_i = (f_p^i)$, $p = 1, \ldots, m$, of property values.

Let use assume that by considering Ψ with respect to any element a_j, to which the vector F_i is assigned, it is possible to determine whether it belongs to A or not. ("A is formally closed.")

In addition, we assume that investigating Φ with respect to any element a_i of A, to which the vector F_i is assigned, we can determine exactly whether it belongs to the pattern A_1 or A_2 [2]. ("The errors of pattern recognition are really determined in A.")

3. Suppose that the material for the pattern recognition is given:

$$A_o \subset A: \quad \begin{cases} A_{01} \subset A_1 : (a_i F_i, \varphi_1), & i = 1, \ldots, n_{01}, \\ A_{02} \subset A_2 : (a_j F_j, \varphi_2), & j = 1, \ldots, n_{02}. \end{cases} \tag{1}$$

Let, on the basis of (1) and explicitly stated assumptions two *admissible* pattern recognition algorithms $R(F, A_o, a_k)$ and $S(F, A_o, a_k)$ be constructed. By means of these algorithms any element a_k of A, to which the vector F_k is assigned, can be assigned probably in a wrong way either to the pattern A_1 or to the pattern A_2. We shall consider these algorithms $R(F, A_o, a_k)$ and $S(F, A_o, a_k)$ as *admissible* in the sense that they are *universal* (i.e., they are applicable with any vector F_e), *adaptive* (i.e., do not they produce errors for any test), *transparent* (that is they allow reasonable interpretation), and *simple* (with respect to their explicit construction).

Our goal is:

1) to specify the similarity measure between $R(F, A_o, a_k)$ and $S(F, A_o, a_k)$,
2) to construct a rule that gives a possibility to determine the preference relation between nonsimilar $R(F, A_o, a_k)$ and $S(F, A_o, a_k)$.

It is essential for $R(F, A_o, a_k)$ and $S(F, A_o, a_k)$ to be equivalent from the methodological requirements and correspondence to the experimental data.

4. Let us consider (F_k) – the set of all logically possible vectors F_k. It contains $N = \prod_{p=1}^{m} q_p$ such vectors. Let (F_o) be a set of vectors F_e, which are represented in A_o.

A vector F_h in (F_k) is called bi-permissible if all its pairs of values f_p^h and f_r^h are at least in one F_e in (F_o). Let us denote the set of all be-permissible vectors (F_h) by $(F_h)_{02}$ and a subset of elements a_h of A, to which the vectors F_h in $(F)_{02}$ are assigned, by A_{02}. It is clear that $A_o \subset A_{02} \subset A$. The subset of elements A_{02} is called a bi-subset of elements A_o assuming A_o.

We consider A_{02} as the set of elements a_k of A for which it is reasonable to recognize on the bases of A_o by any algorithms $Q(F, A_o, a_k)$.

5. Using the pattern recognition algorithms $R(F, A_o, a_k)$ and $S(F, A_o, a_k)$ in A_{02} we arrive to its subdivisions on the following patterns:

$$A_{02}^R \Rightarrow (A_{02}^R)_1 \text{ and } (A_{02}^R)_2, \quad A_1 \supset (A_{02}^R)_1, \quad A_2 \supset (A_{02}^R)_2,$$
$$A_{02}^s \Rightarrow (A_{02}^s)_1 \text{ and } (A_{02}^s)_2, \quad A_1 \supset (A_{02}^s)_1, \quad A_2 \supset (A_{02}^s)_2.$$

We define the similarity measure between these algorithms as follows:

$$\Lambda_{A_{02}}(R,S) \equiv \frac{N_{A_{02}}(R,S)}{N(A_{02})}, \tag{2}$$

where $N_{02}(R,S)$ is a number of all elements of A_{02} which are equally recognized by these algorithms and $N(A_{02})$ is a number of all elements in A_{02}. Suppose that $R(F,A_o,a_k)$ and $S(F,A_o,a_k)$ are different from (2), i.e., they are different from the point of view of our main goal.

6. Let us consider the description of subdivisions of A_{02}, that are generated by the pattern recognition algorithms $R(F,A_o,a_k)$ and $S(F,A_o,a_k)$:

$$(A_{02}^R)_1, (A_{02}^R)_2 \Rightarrow \omega_\pi(A_{02}R) \quad \pi = 1,\dots,\pi_o,$$
$$(A_{02}^s)_1, (A_{02}^s)_2 \Rightarrow \omega_\pi(A_{02}S) \quad \pi = 1,\dots,\pi_o.$$

We shall assume the hypothesis that establishing the preference relation between pattern recognition algorithms $R(F,A_o,a_k)$ and $S(F,A_o,a_k)$ can be reduced to that of the analogous relation between the generated subdivisions $(A_{02}^R)_1, (A_{02}^R)_2$ and $(A_{02}^S)_1, (A_{02}^S)_2$ with assuming there a priori choosing description.

It is essential that it is known *a priori*, if these descriptions are reasonable, they should be "multiparameter" and "multiscale" [12]. Almost everything depends on how this description is constructed.

The above said confirms that in the pattern recognition "the practice can serve as criterion of truth only conditionally" [3, 10, 11].

7. Let A_{02}^1 and A_{02}^2 be a subdivision of A_{02} in to two patterns, and $\Lambda_F(a_i,a_j,\theta)$ be a "reference" similarity measure between the elements a_i and a_j of A with respect to the set of properties $F = (f_p), p = 1,\dots,m$, a particular choice of which can be left apart [6]. Let us assume that:

(1) $0 \le \Lambda_F(a_i,a_j,\theta) \le q$,

(2) $\Lambda_F(a_i,a_j,\theta) = \Lambda_F(a_i,a_j,\theta)$,

(3) $\Lambda_F(a_i,a_j,\theta) = 1 \Leftrightarrow a_i \underset{F}{=} a_j$,

(4) $\Lambda_F(a_i,a_j,\theta) = 0 \Leftrightarrow a_i \underset{F}{=} a^* \in A_1,\ a_j \underset{F}{=} a^{**} \in A_2$.

It is important that there exist only two elements, from different patterns, the measure of similarity between which is equal to zero. It is also of importance that one can speak of the compactness of the set of subdivisions only with the fixed similarity measure [12].

8. Let us turn to the description of the subdivision of A_{02} onto A_{02}^1 and A_{02}^2. We shall introduce $\omega_1(A_{02}^1, A_{02}^2)$ – the *expanding* index for A_{02}^1 and A_{02}^2:

$$\omega_1(A_{02}^1, A_{02}^2) \equiv 1 - \max \Lambda_F(a_i, a_j, \theta), \tag{3}$$

where a_i is in A_{02}^1 and a_j is in A_{02}^2.

If for any a_i in A_{02}^e, $e = 1, 2$, the most similar to its element a_k in A_{02} also belongs to A_{02}^e, we say that A_{02}^e is *compact*, otherwise we call A_{02}^e as *non-compact*.

Let us introduce $\omega_2(A_{02}^1, A_{02}^2)$ – the indicator of *simplicity* for A_{02}^1 and A_{02}^2:

$$\omega_2(A_{02}^1, A_{02}^2) \equiv \begin{cases} 1, & \text{if both } A_{02}^1 \text{ and } A_{02}^2 \text{ are compact,} \\ 1/2, & \text{if } A_{02}^e \text{ is compact, but } A_{02}^{e'} \text{ is non-compact,} \\ 0, & \text{if both } A_{02}^1 \text{ and } A_{02}^2 \text{ are non-compact.} \end{cases} \tag{4}$$

If a_i in A_{02} is not the most similar to any a_k in A_{02}, we consider a_i to be *isolated* in A_{02}.

Let us introduce $\omega_3(A_{02})$ as an indicator of *compactness* for A_{02}:

$$\omega_3(A_{02}^1) \equiv 1 - \frac{n(A_{02})}{N(A_{02})}, \tag{5}$$

where $n(A_{02})$ is the number of all isolated a_j in A_{02}, and $N(A_{02})$ is the number of all a_i in A_{02}. In a similar way, one can introduce the indicators of "compactness" for A_{02}^1 and A_{02}^2:

$$\omega_3(A_{02}^1) \equiv 1 - \frac{n(A_{02}^1)}{N(A_{02}^1)}, \qquad \omega_3(A_{02}^2) \equiv 1 - \frac{n(A_{02}^2)}{N(A_{02}^1)}. \tag{5)$'$(}$$

If for a_i and a_j in A_{02} the following equality is satisfied:

$$\Lambda(a_i, a_j, \theta) > \max \Lambda_F(a_e, a_k, \theta),$$

where a_e – in A_{02}^1 and a_k – in A_{02}^2, then we say that they "are connected in one connection". The subset A_{02}' of set A_{02} is called "homogeneous component of connection", if any a_i and a_j in A_{02}' can be connected with one another by means of some quantity of bundles provided that there is no a_k out of A_{02}' connected with any a_e from A_{02}' by means of one bundle.

Let us introduce $\omega_4(A_{02}^1, A_{02}^2)$ as an indicator of "homogeneous coherence" for A_{02}^1 and A_{02}^2:

$$\omega_4(A_{02}^1, A_{02}^2) \equiv \Big(\frac{2}{H(A_{02})} - \frac{2}{N(A_{02})}\Big), \tag{6}$$

where $H(A_{02})$ is the number of homogeneous components of coherence in A_{02}. It is clear that $2 \le H(A_{02}) \le N(A_{02})$.

For any a_i in A_{02}^e, $e = 1, 2$, one can assign $n(a_i, A_{02}^e)$, which is the number of "relatives" that is the number a_j from A_{02}^e which are similar to a_i more than the most similar a_k with $A_{02}^{e'}$, $e' = 2, 1$.

Let us introduce $\omega_5(A_{02}^1, A_{02}^2)$ as an indicator of *density* for A_{02}^1 and A_{02}^2:

$$\omega_5(A_{02}^1, A_{02}^2) \equiv \frac{1}{2}\left[\frac{1}{N(A_{02}^1)}\sum_i \frac{n(a_i, A_{02}^1)}{N(A_{02}^1) - 1} + \frac{1}{N(A_{02}^2)}\sum_j \frac{n(a_j, A_{02}^2)}{N(A_{02}^2) - 1}\right]. \quad (7)$$

It is clear that only the availability of the descriptions of the set subdivisions allows to say about the theory of classifications from mathematical point of view [2].

A lot of such descriptions ("the descriptions of the location of different colour points in the multidimensional cube" [3]) can be constructed, a part of which has visual sense.

9. We call the subdivision A_{02}^1, A_{02}^2 *qualitatively the best* if

$$\omega_\pi(A_{02}^1, A_{02}^2) = 1, \qquad \pi = 1, \ldots, 5 \quad (8)$$

and the *worst in quality* if

$$\omega_\pi(A_{02}^1, A_{02}^2) = 0, \qquad \pi = 1, \ldots, 5. \quad (9)$$

Let us compare two subdivisions $A_{02}^{1\alpha}$, $A_{02}^{2\alpha}$ and $A_{02}^{1\beta}$, $A_{02}^{2\beta}$ in quality by means of lexicographical rule, for example [13]:

"A subdivision $A_{02}^{1\alpha}$, $A_{02}^{2\alpha}$ is better than a subdivision $A_{02}^{1\beta}$, $A_{02}^{2\beta}$ if

$$\begin{aligned}
&\text{either} \quad \omega_1(A_{02}^{1\alpha}, A_{02}^{2\alpha}) > \omega_1(A_{02}^{1\beta}, A_{02}^{2\beta})\\
&\text{or} \quad \omega_1(A_{02}^{1\alpha}, A_{02}^{2\alpha}) = \omega_1(A_{02}^{1\beta}, A_{02}^{2\beta}) \quad \text{and} \quad \omega_2(A_{02}^{1\alpha}, A_{02}^{2\alpha}) > \omega_2(A_{02}^{1\beta}, A_{02}^{2\beta})\\
&\text{or} \quad \omega_2(A_{02}^{1\alpha}, A_{02}^{2\alpha}) = \omega_2(A_{02}^{1\beta}, A_{02}^{2\beta}) \quad \text{and} \quad \omega_3(A_{02}^{1\alpha}, A_{02}^{2\alpha}) > \omega_3(A_{02}^{1\beta}, A_{02}^{2\beta})\\
&\text{or} \quad \omega_3(A_{02}^{1\alpha}, A_{02}^{2\alpha}) = \omega_3(A_{02}^{1\beta}, A_{02}^{2\beta}) \quad \text{and} \quad \omega_4(A_{02}^{1\alpha}, A_{02}^{2\alpha}) > \omega_4(A_{02}^{1\beta}, A_{02}^{2\beta})\\
&\text{or} \quad \omega_4(A_{02}^{1\alpha}, A_{02}^{2\alpha}) = \omega_4(A_{02}^{1\beta}, A_{02}^{2\beta}) \quad \text{and} \quad \omega_5(A_{02}^{1\alpha}, A_{02}^{2\alpha}) > \omega_5(A_{02}^{1\beta}, A_{02}^{2\beta}).
\end{aligned}$$

Provided $\omega_\pi(A_{02}^{1\alpha}, A_{02}^{2\alpha}) = \omega_\pi(A_{02}^{1\beta}, A_{02}^{2\beta})$, $\pi = 1, \ldots, 5$, the subdivisions $A_{02}^{1\alpha}$, $A_{02}^{2\alpha}$ and $A_{02}^{1\beta}$, $A_{02}^{2\beta}$ are equal."

10. Let us consider bi-subset A_{02} of the set A_o assuming pattern recognition A_o. Let us introduce the indicator of representation A^* with respect to A_{02}, thus:

$$æ_F(A_0/A_{02}) = \min_{a_i \in A_{02}} \max_{a_j \in A_o} \Lambda(a_i, a_j, \theta). \quad (10)$$

For any a_i in A_{02} one can find the most similar to it a_j in A_o and choose among these measures the minimal one. It is not of big importance how many a_j are in A_o, the most important is how they are located with respect to a_i in A_{02} assuming $\wedge_F$. This limits strongly traditional statistical approach to pattern recognition [1, 3, 14]. Let us choose $\Lambda_F(a_i, a_j, \theta)$ from the condition $æ_F(A_0/A_{02}) = \max$.

11. Basing on above, one can specify questions mentioned in Paragraph 1 under the particular situations of pattern recognition and try to answer them. In particular, one can specify when the subdivision A into A_1 and A_2 on the basis of the investigation of Φ is considered to be the subdivision in two patterns with allowance for A_o.

12. The major difficulties of computer application in pattern recognition problems which have been revealed up to now in geosciences, can be considered as inherent to any other problem of applied mathematics [6, 8]. The main difficulties are in the problem formulation but not in their solution [7]. Overcoming of these difficulties by use of supercomputers and artificial intelligence is hardly perspective [11, 13]. One should know for that purposes, who and how will use the solutions of these problems.

References

[1] Zagoruiko N.G. Pattern Recognition Methods and their Application. – Moscow: Nauka, 1972.

[2] Voronin Yu.A. Theory of Classification and its Applications. – Novosibirsk, 1985.

[3] Voronin Yu.A. The Future of Geological Prospecting: Expert or Theoretical Systems. – Novosibirsk, 1987.

[4] Blehman, Mishkis A.D. Applied Mathematics: Subject, Logics, Features of Approach. – Kiev, 1976.

[5] Tolstova Yu.N. Logics of Mathematical Data Analysis. – Moscow, 1991.

[6] Voronin Yu.A. Principles of the Similarity Theory. – Novosibirsk, 1991.

[7] Voronin Yu.A. To the problem of "closure" of the problems of geological interpretation of complex geophysical data // Geophysical methods of the Earth core exploration. – Novosibirsk, 1998.

[8] Vorovin Yu.A. On geological problems and on similarity in connection with complex data interpretation // Geology and geophysics. – 1977. – № 6.

[9] Shafarevich I.R. Mathematical Mentality and Nature. – Vitebsk, 1996.

[10] Kapitza S.P., Kurdjumov S.P., Malinetzkij G.G. The Synergetics and Future Prediction. – Moscow, 1997.

[11] Dzigalin V.A. Mental automatization as the cultural-social problem // Questions of methodology. – 1994. – № 1–2.

[12] Zagoruiko N.G. The compact and λ-compact hypothesis in pattern recognition // Industrial mathematics methods. – 1998. – № 1.

[13] Lorjer G.-L. Artificial Intelligence System. – Moscow, 1991.

[14] Ljubarski Yu.A. Intellectual Information Systems. – Moscow, 1990.