Determination of representative features when building an extreme recognition algorithm

Sh Kh Fazilov, N M Mirzaev, S S Radjabov and G R Mirzaeva

Scientific and innovation center of information and communication technologies at the Tashkent university of information technologies named after Muhammad al-Khwarizmi, 17A, Buz-2, Tashkent, 100125, Uzbekistan
E-mail: nomazmirza@rambler.ru

Abstract. The problems of determining the representative features when constructing an extremal pattern recognition algorithm defined in a high-dimensional feature space are considered. As an initial model, a model of recognition algorithms based on radial functions was considered. The procedures for extracting subsets of interrelated features and selecting a set of representative features when constructing an extremal algorithm within the framework of the recognition model under consideration are presented. The main idea of the first procedure is to form a set of features that are unrelated (or are connected rather weakly) among themselves. The essence of it is as follows. The studied subsets of features are combined into one subset if they are close enough to each other in a sense. Otherwise, they belong to different subsets. The main idea of the second procedure is to define a representative element in each subset of tightly coupled features. The choice of an element from the considered subset of tightly coupled features is made on the basis of an assessment of the proximity of features. It is required that among the selected set of features there should not be tightly coupled features. The main advantage of the proposed procedures is to improve the accuracy of the results of allocating subsets of tightly coupled features when building a recognition algorithm under conditions of large dimensionality of the feature space and determining the quantitative evaluation of these subsets. In order to assess the performance of the developed procedures, experimental studies were carried out. The application of the developed procedures allows to more accurately determine the unknown parameters of recognition operators in the space of features of large dimension.

1. Introduction

Analysis of the literature in recent years, in particular [1–8], shows that a number of models of pattern recognition algorithms (ie, a family of algorithms for solving classification problems) have been built and studied: models based on the use of the separation principle; models based on the theory of statistical solutions; models built on the principle of potentials; models based on the calculation of estimates, etc. It is obvious that the developed recognition algorithms on one of the principles listed above can use ideas and other principles. For example, the method of potentials directly combines the use of the principles of potentials and separation, while the method of the committees uses the ideas of the principles of the division of classes by hyperplanes and voting [1].

However, the analysis of these models shows that they are mainly focused on the recognition of objects described in the space of features of small size. The task of constructing algorithms for recognition in the space of features of large dimension is associated with some computational
difficulties. In addition, in these conditions, many features are interrelated. In this regard, there is the problem of pattern recognition in the interconnectedness of features.

Some questions of the construction of recognition algorithms under conditions of interconnectedness of features are considered in [9–13]. But the task of constructing extremal recognition algorithms under conditions of interconnectedness of features has not been studied enough. Therefore, the task of identifying representative features, which is the first steps in the construction of extreme recognition algorithms in terms of interrelated features, is relevant.

The purpose of this work is to develop an algorithm for the selection of representative features in the problems of constructing extremal recognition algorithms based on the calculation of estimates. It uses a heuristic approach based on the assessment of the interconnectedness of features.

The proposed algorithm can be used to construct extremal pattern recognition algorithms defined in a space of large dimension.

2. Statement of the problem

Let \( x_1, \ldots, x_i, \ldots, x_n \) be the set of features forming the \( m \)-dimensional space \( X \). In the space \( X \) the function \( \mathcal{R}(x_i, x_j) \) is defined, which characterizes the strength of the pairwise connections between \( x_i \) and \( x_j \). It is assumed that \( \mathcal{R}(x_i, x_j) (i = 1, n; j = 1, \bar{n}) \) satisfies the conditions:

\[
\begin{align*}
\mathcal{R}(x_i, x_j) &= \mathcal{R}(x_j, x_i); \\
\mathcal{R}(x_i, x_i) &> \mathcal{R}(x_i, x_j) (i \neq j); \\
b_{\text{min}} &\leq \mathcal{R}(x_i, x_j) \leq b_{\text{max}}.
\end{align*}
\]

where \( b_{\text{min}}, b_{\text{max}} \) are real numbers.

The selection of representative features is based on the selection of “independent” and tightly coupled subsets of features. According to this method, the system of features under study belongs to one set if they are close enough (in a sense) to each other. Otherwise, they belong to different sets of features. Then, from each tightly coupled set of features, a feature is selected that is essential in combination with features from other groups. A formal description of the choice of strongly related features is as follows.

The task is to classify the set \( \{x_1, \ldots, x_n\} \) into independent subsets of strongly related features \( \{\Omega_1, \ldots, \Omega_{\bar{n}}\} \) based on the information in table \( T_{mn} \) [14], which consists of \( m \) columns and \( n \) rows.

Depending on the choice of \( \mathcal{R}(x_i, x_j) \) you can get a variety of algorithms for determining the interdependence of features.

3. Method of solution

The basic idea of identifying subsets of tightly coupled features is that the elements of each subset will be closer to other elements of this subset than to elements of other subsets. At the same time, this problem is considered to be solved if the centers of the clusters and the boundaries of the corresponding subsets of features have been determined by the set \( \{x_1, \ldots, x_n\} \).

Let \( \Lambda \) be the matrix of paired coupling coefficients for features \( \{x_i \} (i = 1, \bar{n}) \), given as follows

\[
\Lambda = \|\lambda_{ij}\|, \lambda_{ij} = \mathcal{R}(x_i, x_j)
\]

[13];

\[
\mathcal{R}(x_i, x_j) = \left(1 + \alpha \rho(x_i, x_j)\right)^{-1},
\]

where \( \alpha \) is the algorithm parameter.

The distance between the features \( x_i \) and \( x_j \) can be determined in various ways, for example:

a) \( \rho(x_i, x_j) = \left[\alpha_1 (r_i - r_j)^2 + \alpha_2 (\delta_i - \delta_j)^2\right] \),
b) $\rho(x_i, x_j) = 1 - |c_{ij}|$, $c_{ij} = \sum_{u=1}^{m}(\alpha_{iu} - r_i)(\alpha_{ju} - r_j)(\delta_i\delta_j)^{-1/2}$, where $\alpha_1$, $\alpha_2$ is the algorithm parameters; $r_i$, $\delta_i$ hematical expectation and dispersion of measurement results.

In the future, as a matrix of paired relations, we will consider an arbitrary matrix whose elements satisfy conditions (1):

$$\Lambda = \|\lambda_{ij}\|_{m \times n}. \quad (2)$$

Since the matrix is symmetric, only the set $\{\lambda_{ij} | i = 1, (n - 1); j = 1, (n)\}$ of the over-diagonal elements of the matrix (2):

$$\Lambda' = \{\lambda_{ij} | \lambda_{ij} = \mathcal{R}(x_i, x_j), x_i, x_j \in X\}. \quad (3)$$

3.1. Selecting a subset of tightly coupled features

Let be given the matrix of paired relations. Arranging all elements in descending order, we construct the following numerical sequence

$$\lambda_{i1} \geq \lambda_{i2} \geq \cdots \geq \lambda_{ik}, \quad k = m(m - 1)/2. \quad (4)$$

We introduce some definitions. The features $x_i$ and $x_j$ are called interconnected if they both belong to only one subset of strongly related features $\Omega_q$ $(x_i \in \Omega_q, x_j \in \Omega_q, \Omega_q \subset X)$, otherwise we call them loosely coupled. With the existence of subsets of tightly coupled features, estimates of measures of proximity of interconnected pairs of features are concentrated in the left part of the numerical sequence (4). Accordingly, in the right part of this series, estimates of measures of proximity of weakly coupled pairs of features are concentrated. Thus, the behavior of the numerical sequence (4) allows us to investigate the behavior of the initial set of features from given information about recognition objects.

To study the behavior of a numerical sequence (4), we construct its histogram. Obviously, one of the following cases occurs here:

1) there is at least one significant (from a statistical point of view) local minimum on the histogram of a numerical sequence (4);

2) there is no significant minimum on the histogram of the numerical sequence (4).

Let’s analyze each case separately. To obtain sufficiently substantiated conclusions, we construct an additional sequence of numbers, which characterizes a chain change in the estimates of proximity measures between adjacent pairs of features:

$$r_1, \ldots, r_j, \ldots, r_{k-1}, r_j = \lambda_{ij} - \lambda_{ij-1}, \quad k = n(n - 1)/2, \quad j = 1, (k - 1). \quad (5)$$

Obviously, in the case when the histogram has a significant local minimum, the numerical sequence (5) has large jumps. In this case, there is at least one number $j_0$, such that

$$\lambda_{ij_0} \gg \lambda_{ij_0+1} \quad \text{or} \quad r_{j_0} \gg r_{j_0-1}, r_{j_0} \gg r_{j_0+1}. \quad (6)$$

When condition (6) is fulfilled, the set of features $X$ has subsets of strongly related features. These subsets of features are sufficiently independent and are located far from each other. In such a situation, there exists at least one subset $\Omega_q$, $q \in \{1, 2, \ldots, n'\}$ such that the minimum estimate of proximity of $(\lambda_{uv}^{\min})$ interrelated pairs of features $x_u, x_v \in \Omega_q$ of the subset under consideration is much greater than the maximum estimate $(\lambda_{uv}^{\max})$ of weakly coupled pairs of features $x_u \in \Omega_i, x_v \in \Omega_j$:

$$\lambda_{uv}^{\min} > \lambda_{uv}^{\max}.$$ 

Consider the second case when the histogram has no significant local minima. Therefore, the sequence (5) does not have jumps. In this case, the set $X$ can be characterized by one of the following options:
1) all features \( x_i \) are interrelated \((x_i \in X, i = 1, n)\);  
2) the set of features \( X \) consists of disjoint, but closely spaced subsets of features.  

For a thorough analysis of the structural structure of the considered set of \( X \) we introduce auxiliary sequences of numbers that are formed as follows. First, estimates are obtained for the proximity of the first \( x_1 \) \((x_1 \in X)\) with all the other elements of the set as it is near \((1)\):  
\[ \lambda_{12}, \ldots, \lambda_{1j}, \ldots, \lambda_{1n}, \quad \lambda_{1j} = \mathcal{R}(x_1, x_j), \quad j = 2, (n). \]  

Further, from the numerical sequence \((7)\) we determine the maximum \((\lambda_{1}^{\text{max}})\) and minimum \((\lambda_{1}^{\text{min}})\) elements. For each feature \( x_i \) \((x_i \in X, i = 1, n)\) these operations are repeated:  
\[ \lambda_{i}^{\text{max}} = \max\{\lambda_{i1}, \ldots, \lambda_{in}\}, \]  
\[ \lambda_{i}^{\text{min}} = \max\{\lambda_{i1}, \ldots, \lambda_{in}\}, \]  
\[ \lambda_{ij} = \mathcal{R}(x_i, x_j), \quad i, j \in \{1, 2, \ldots, n\}, \quad i \neq j. \]

As a result, we get two sets \( \Lambda_{\text{max}} \) and \( \Lambda_{\text{min}} \):  
\[ \Lambda_{\text{max}} = \{\lambda_{1}^{\text{max}}, \ldots, \lambda_{n}^{\text{max}}\}, \quad \Lambda_{\text{min}} = \{\lambda_{1}^{\text{min}}, \ldots, \lambda_{n}^{\text{min}}\}. \]

It should be noted that many (possibly all) elements of \( \Lambda_{\text{max}} \) correspond to strongly coupled pairs of features. Moreover, it is impossible to say anything definite about the elements of a set \( \Lambda_{\text{min}} \). The elements of the set \( \Lambda_{\text{min}} \) can correspond to both a tightly coupled pair of features and loosely coupled. Obviously, in the formed sets \( \Lambda_{\text{max}} \) and \( \Lambda_{\text{min}} \) there are at least two identical (equal) elements. These elements correspond to one pair of features \((x_i, x_j)\) and \((x_j, x_i)\). In the considered sets \( \Lambda_{\text{max}} \) and \( \Lambda_{\text{min}} \) we leave only one of these identical elements. Further, by ordering (in descending order) the remaining elements \( \Lambda_{\text{max}} \) and \( \Lambda_{\text{min}} \) we form a sequence of numbers \( \mathcal{J}_{\text{max}} \) and \( \mathcal{J}_{\text{min}} \):  
\[ \mathcal{J}_{\text{max}} = \{\lambda_{1}^{\text{max}}, \ldots, \lambda_{u'}^{\text{max}}\}, \quad u \leq n' - 1 \]  
\[ \mathcal{J}_{\text{min}} = \{\lambda_{1}^{\text{min}}, \ldots, \lambda_{v'}^{\text{min}}\}, \quad v \leq n' - 1 \]  

From the construction of the sequence \( \mathcal{J}_{\text{max}} \) and \( \mathcal{J}_{\text{min}} \) it follows that their elements correspond to different pairs of features. To study the elements of the sequence \((8)\), we construct their histogram based on the elements of the \( \mathcal{J}_{\text{max}} \) and \( \mathcal{J}_{\text{min}} \) subsets and an auxiliary number sequence:  
\[ D_{\text{max}} = \{\Delta_{1}, \ldots, \Delta_{u-1}\}, \quad \Delta_i = \lambda_{i}^{\text{max}} - \lambda_{i+1}^{\text{max}} \]  
\[ D_{\text{min}} = \{\delta_{1}, \ldots, \delta_{v-1}\}, \quad \delta_i = \lambda_{i}^{\text{min}} - \lambda_{i+1}^{\text{min}} \]

According to the results of studies of the behaviors of \( D_{\text{max}} \) and \( D_{\text{min}} \) we can distinguish the strongly connected and weakly bound features of the initial set of features. If \( D_{\text{max}} \) does not have jumps, then the set of features is strongly connected or consists of closely spaced subsets of features. At the same time, the distribution density of the original features is constant. If \( D_{\text{max}} \) has jumps or its histogram contains at least one significant local minimum, then the set \( X \) has a subset of features with different densities. Among them may be such subsets of features that have one element at a time. They are called independent features. The number of independent features can be determined as follows. If \( k - 1 \) elements of the set \( D_{\text{max}} \) and \( D_{\text{min}} \) are equal, \( X \) then the set has at least \( k \) independent features. If the set \( D_{\text{min}} \) does not have jumps, then the set \( X \) consists of tightly coupled features or is formed from closely spaced subsets of features. Otherwise (i.e., when this sequence has jumps or its histogram has at least one significant local minimum), the set \( X \) consists of \( k \) subsets of features. At the same time, the minimum proximity of weakly dependent pairs of features has a different order.

As a result of using the considered algorithm \( n' \) subsets of features \( \{\Omega_1, \ldots, \Omega_{n'}\} \) are formed. Next, a representative feature is determined.
3.2. Search for a representative trait in the subset $\Omega_q$

It is obvious that each representative feature from the subset $\Omega_q$ ($q = 1, n'$) is a typical representative of the selected subset of tightly coupled features. You can use various ways to select unrelated (representative) traits from a subset of tightly coupled traits. The main idea of the choice is to identify the most independent (or weakly dependent) set of features [13].

Let $\Omega_q$ ($q = 1, n'$) be a subset of tightly coupled features. It is assumed that for each such set $N_q$ is calculated the number of elements (capacity) of these subsets: $N_q = \text{card}(\Omega_q)$, ($q = 1, n'$). Then the search procedure for representative features can be described as follows.

First, it is assumed that $q = 0$, $k = 1$. As a result of this procedure, $n'$ subsets are considered 3 times. To do this, use the variable $k$, which indicates the time at which the elements $\Omega_q$ are viewed.

First, isolated elements are selected (as representative features) of the subset $\Omega_q$, which differ sharply from other features. To do this, follow the procedure.

The value of $q$ is increased by one and the condition $q > n'$ is checked. If this condition is met, then $k$ is incremented by one. Check condition $k > 3$. If it is executed, this search procedure stops.

If $n = 1$, then the condition $N_q = 1$ is checked. If the conditions in question are fulfilled, then the element belonging to the subset $\Omega_q$, is among the representative features and the transition to the beginning of the procedure is carried out.

If $k = 2$, then the condition $N_q > 2$ is checked. If the conditions under consideration are satisfied, then for each element $\Omega_q$ an estimate of the proximity of each element to other elements of a given subset of features is calculated:

$$\mu_i = \sum_{j=1}^{i-1} \mathfrak{R}(x_i, x_j) + \sum_{j=i+1}^{N_q} \mathfrak{R}(x_i, x_j). \quad (9)$$

The element of the subset $\Omega_q$, that is as close as possible to the other elements is determined: $\mu_j = \max_{i \in [1, N_q]} \mu_i$. Then the feature $x_j$ is selected as a representative feature. After completing the selection of a representative feature, the system proceeds to the beginning of the procedure.

If $k = 3$, then the condition $N_q = 2$ is checked. If the conditions in question are satisfied, then for each element $\Omega_q$ an estimate of proximity to the representative elements of the other subsets, which were selected at the previous stages of selection, is calculated:

$$\mu_i = \sum_{j=1}^{N_0} \mathfrak{R}(x_i, x_j), \quad i = 1, 2, \ldots, 2k; \quad j = 1, N_0. \quad (10)$$

Here $k$ is the number of subsets that consist of two elements, $N_0$ – the number of separate elements and elements selected from subsets with a capacity of more than two.

The element of the subset $\Omega_q$, which is significantly different from other selected representative features $\mu_i = \min_{i \in [1, 2]} \mu_i$ is determined. Then the feature $x_j$ is selected as a representative feature. After completing the selection of a representative feature, the system proceeds to the beginning of the procedure.

As a result of this procedure, a set of features is formed, each of which is a typical representative of a selected subset of strongly related features. Obviously, as a result of these actions, a space is formed with smaller features ($n' < n$).

4. Experiment and results

For practical use of the considered algorithm, programs for the determination of representative features in the C++ language have been developed. The efficiency of the developed programs
will illustrate the solution of a number of problems, in particular the model and practical problems.

4.1. Model task
The source data for the model example was generated in the space of dependent features. In this case, the method of forming random vectors with given correlation properties was used [15]. The number of classes in this experiment is 2. The size of the analyzed sample is 300 implementations (150 implementations for each image). The number of features in this example is 120. The number of subsets of tightly coupled features is 5. There are 60 features in the first subset, 7 in the second, 2 in the third, 1 in the fourth, and 50 in the fifth. The distribution is normal.

Using the first procedure, 5 subsets of tightly coupled features \( \Omega_1 = \{ x_1, x_3, x_4, x_6, x_8, \ldots, x_{63} \} \), \( \Omega_2 = \{ x_{64}, \ldots, x_{70} \} \), \( \Omega_3 = \{ x_2, x_7 \} \), \( \Omega_4 = \{ x_5 \} \), \( \Omega_5 = \{ x_{71}, \ldots, x_{120} \} \) were identified. Then, using the formula (10), the following are determined in order: the first representative feature \( x_5 \) from \( \Omega_4 \), the second representative feature \( x_{25} \) from \( \Omega_1 \), the third representative feature \( x_{68} \) from \( \Omega_2 \), and the fourth representative feature \( x_{89} \) from \( \Omega_2 \). Finally, using formula (9) \( x_7 \) from \( \Omega_3 \).

Thus, as a set of representative features for objects belonging to the first class, we get \( x_5, x_{25}, x_{68}, x_{89}, x_7 \), which corresponds to the initial assumption of representative features.

Similarly, a set of representative features \( x_{31}, x_{58}, x_{86}, x_{115}, x_{29} \) is defined for objects that belong to the second class.

The experiment revealed all representative features. Further, on the basis of these features, it is possible to construct dependency models for each subset of tightly coupled features and on their base to build an extreme recognition algorithm.

4.2. Practical task
One of the most important tasks arising from the predictive studies of various territories is to determine the patterns of distribution of mineral deposits (MPI), which allow to identify stable spatial, temporal and genetic relationships of the ore object with various geological formations. The study of the identified empirical relationships, the analysis of their nature and properties allows us to form a set of geological ore-controlling factors and define a set of prediction criteria for IIP. Based on the use of these criteria, further processing of data on the entire territory as a whole, gradually (from small-scale to large-scale studies) areas that are most promising for a particular mineral are optimized.

Given a sample of 606 objects. Each object is described by 22 features - the results of chemical analysis of the composition of the soil sample of the earth. All these features characterize the proportion of the content of a certain set of metals in the sample. The place of sampling, the type of terrain and the tectonic location are also indicated. It is necessary for each of these three properties to identify a subset of tightly coupled features.

To solve the problem, the procedures described in section 2 were used. For example, in fig. Figure 1 shows all the steps for the formation of subsets of strongly related characteristics of the 2nd class of the sample, classified by tectonic dislocation (the number of objects in the class is 268). Here, at each step, one subset is indicated by one color (and number). As can be seen from fig. 1, the result is the following 4 subsets of tightly coupled features:

\[
\{ x_1, x_3, x_4, x_6, x_8, x_{10}, x_{12}, x_{15}, x_{16} \}, \{ x_2 \}, \{ x_5, x_9, x_{11}, x_{13}, x_{14}, x_{17}, x_{19}, x_{20}, x_{21}, x_{22} \}, \{ x_7 \}, \{ x_{18} \}.
\]

A meaningful analysis of the obtained results of solving this problem confirmed their considerable coincidence with the assumptions of subject matter experts.
5. Conclusions
One of the steps in defining a model of recognition algorithms based on an assessment of the interrelatedness of features is the selection of representative features.

To solve the problem of identifying representative traits, an algorithm is proposed that selects independent (or weakly dependent) traits from the initial set of traits. However, it is considered as two more simple tasks:

- selection of subsets of tightly coupled features;
- determination of a typical element of the subset under consideration.

The efficiency of the developed algorithm is shown when solving a model example.

It should be noted, although the proposed algorithm is oriented for building extremal recognition algorithms, it can be used in solving the following basic data mining tasks:

- reduction of dimensions of the considered models in order to reduce computational operations and facilitate the iteration of the results obtained;
- compression of volumes, stored source data.

The proposed algorithm can be used in the preparation of various programs aimed at solving applied problems of pattern recognition.

References
[1] Zhuravlev Y I and Gurevich I 1998 Selected scientific works (Moscow: Master)
[2] Zhuravlev Y I et al. 2006 Recognition. Mathematical methods. Software system. Practical application (Moscow: Fazis)
[3] Lbov G S and Startseva N G 1999 *Logical solving functions and the problem on solutions statistical stability* (Novosibirsk: Sobolev Institute of Mathematics)

[4] Haykin S 2009 *Neural Networks: A Comprehensive Foundation* (Moscow: Williams)

[5] R O Duda P E H and Stork D G 2001 *Pattern Classification* 2nd ed (New York: John Wiley Sons)

[6] Vapnik V N 1998 *Statistical learning theory* (New York: John Wiley Sons)

[7] Schlesinger M and Hlavc V 2002 *Ten lectures on statistical and structural pattern recognition* 1st ed (Springer)

[8] Fazilov S K, Mirzaev O N and Radjabov S S 2015 *Problems of computational and applied mathematics* 99

[9] Fazilov S K, Mirzaev N M and Mirzaev O N 2016 *Radio electronics, computer science, management* 58

[10] Kamilov M M, Fazilov S K, Mirzaev N M and Radjabov S S 2012 *Chemical technology, Control and management* 52

[11] Kamilov M M, Fazilov S K, Mirzaev N M and Radjabov S S 2012 *Chemical technology, Control and management* 45

[12] Kamilov M M, Fazilov S K, Mirzaev N M and Radjabov S S 2012 *Chemical technology, Control and management* 51

[13] Mirzaev N M and Saliev E A 2018 *Proc. Int. Conf. of the 3rd Russian-Pacific Conference on Computer Technology and Applications* (Vladivostok: FEFU) p 1

[14] Zhuravlev Y I, Kamilov M M and Tulyaganov S E 1974 *Algorithms for calculating estimates and their applications* (Tashkent: Fan)

[15] Devroye L and Gyrfi L 1985 *Nonparametric density estimation the L1 view* (New York: John Wiley)