The Set of Data Transformations for Dimensionality Reduction in Technical and Biomedical Diagnosis

Sergey A. Subbotin

Abstract—The set of rapid transformations from the original multi-dimensional space into one-dimensional axis was firstly proposed. They provide a solution of the feature extraction and feature selection problems. The set of criteria allow to compare transformations to the generalized axis is proposed.

Keywords—Sample, feature, locality-preserving transformation, dimensionality reduction.

I. INTRODUCTION

The computational diagnostics is widely used in technology and medicine, where the expert knowledge is missing or insufficient. The theoretical basis of computational diagnostics is the theory of pattern recognition.

The urgent problem in computational diagnosis and pattern recognition is a model complexity reduction and a model construction speed increasing. It is caused by that the samples frequently characterized by a big number of features or contain a big number of instances (precedents). So to reduce the model complexity and to speed-up it's training process we need to reduce the data dimensionality [1, 2].

The data dimensionality reduction usually performed by a feature selection [1]. It needs, as a rule, an exhaustive search of possible feature combination with the least number of features that provides an acceptable accuracy. When the number of features and instances are big such technique require a lot of time. As an alternative to this way the sample selection [3] can be used. Unfortunately, it involves necessity of computation and using a distance matrix in N-dimensional feature space, that causes a combinatorial complexity of large data processing.

The way to solve these problems is the using of transformation from the multidimensional space of initial features to a one-dimensional axis for data dimensionality reduction [2, 3].

There are various methods of transformation for data dimension reduction (Principal Component Analysis, Semidefinite Embedding, Multifactor Dimensionality Reduction, Non-linear Dimensionality Reduction, Partial Least Squares, Independent Component Analysis) [2–11], which, however, require the calculation of distances between instances or feature correlation coefficients and for a large-scale problem they are hardly applicable in practice due to big requirements of time and computer memory in the process of determining the transformation parameters and in the process of transformation execution. This situation is additionally compounded by that the number of known transformations and their modifications is very big and there are no any formal criteria to analyze their quality, as well as to select the best available transformation for a particular task [3].

The purpose of this work is the development of rapid transformations from a multidimensional feature space to a one-dimensional axis, that can be used for a sample selection and the development of criteria for the transformation selection to use in a particular problem solving.

S. A. Subbotin, Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine (e-mail: subbotin@zntu.edu.ua)

II. PROBLEM STATEMENT

Suppose we have an initial (original) sample $X = <x, y>$ the set of S precedents describing dependence $y(x)$, $x = \{x^s\}$, $y = \{y^s\}$, $s = 1, 2, ..., S$, characterized by a set of N input features $\{x_j\}, j = 1, 2, ..., N$, where j is the number of feature, and an output feature y. Each s-th precedent can be represented as $<x^s, y^s>$, $x^s = \{x_j^s\}$, where x_j^s is the value of j-th input feature and y^s is the value of output feature for the s-th precedent (instance) of the sample, $y^s \in \{1, 2, ..., K\}$, where K is the number of classes, $K > 1$.

Then the problem of the sample X dimensionality reduction can be formally represented as follows: find a transformation $H: X \to I$, which for each instance $x^s = \{x_j\}$ determine the coordinate I^s on the generalized axis I and thus provides a mapping of instances of different classes to the different intervals of the generalized axis.

Since, as a rule, known transformations do not guarantee an exact solution of this problem, further problem arises of designing of indicators to quantify the quality of the transformation and to compare the results of the various transformations between themselves to choose the best transformation of the set.

III. TRANSFORMATIONS TO THE GENERALIZED AXIS

For large-scale problems it is advisable to ensure the creation of such transformations, which would allow the mapping of individual instances without loading of whole initial sample, as well as taking into account the feature informativity in the process of transforming and to provide a generalization of data.

To ensure the generalization of close located data points (instances) we propose to replace feature values to numbers of feature value interval. For this we need previously to discretize the features by partitioning them into intervals of values.

To partitioning the features into intervals the number if interval (term), which hits the s-th instance on the j-th feature is proposed to determine as (1):

$$\hat{x}_j^s = \begin{cases} \mathrm{round}\left(1 + \dfrac{x_j^s - x_j^{\min}}{\theta_j}\right), \theta_j > 0; \\ 1, \theta_j = 0, \end{cases} \tag{1}$$

where

$$\theta_j = \frac{x_j^{\max} - x_j^{\min}}{k_j},$$

$$k_j = \begin{cases} K, K > \mathrm{round}(\ln S), K < \sqrt[N]{S}; \\ \max\{2, \mathrm{round}(\sqrt[N]{S})\}, K > \mathrm{round}(\ln S), K > \sqrt[N]{S}; \\ \max\{2, \mathrm{round}(\ln S)\}, K < \mathrm{round}(\ln S) < \sqrt[N]{S}; \\ K, \mathrm{round}(\ln S) \leq K, K < \sqrt[N]{S}; \\ \max\{2, \mathrm{round}(\sqrt[N]{S})\}, \mathrm{round}(\ln S) \leq K, K \geq \sqrt[N]{S}, \end{cases}$$

$x_j^{\min}$, $x_j^{\max}$ are the minimum and maximum values of j-th feature, respectively.

To ensure the generalization of close located data points (instances) we propose to replace feature values to numbers of feature value interval. For this we need previously to discretize the features by partitioning them into intervals of values.

For the mapping of instances from the original multi-dimensional feature space to the one-dimensional generalized axis is suggested to use the following transformations.

Transformation 1 (see Fig.1). For each number of interval of j-th feature get it binary representation (binary numbers padded with zeros from the left to c_j – the number of digits in k_j). Set the coordinate of s-th instance on the generalized axis $I^s = 0$, set the position (bit) number of generalized axis coordinate $p = 1$. Going by the feature numbers j s in descending order of

their rank and by the group of digits in the interval number $c = 1, 2, ..., c_j$ perform in a cycle: if $p \leq d$, where d is a number of bits in a computer bit grid, then record at p-th position (with numbering from left) of the binary representation of a generalized feature I^s the c-th position value (with numbering from left) of interval number, in which the s-th instance hit on the j-th feature, and set: $p=p+1$. As a result, we will obtain a generalized axis coordinate of instance with the implicit ranking and selection of features.
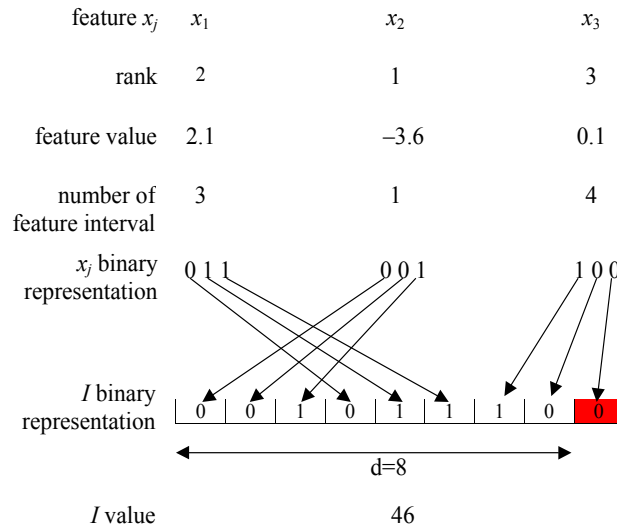

Fig. 1 Illustration of first transformation

Transformation 2 (see Fig. 2). It is an alternative format of constructing a generalized feature for transformation 1. If the total number of bits to represent interval numbers of all features $c_j k_j N$ does not exceed the number of bits in a binary bit grid d when the values c_j are equal for all features: for each interval number of j-th feature obtain its binary representation (binary numbers padded with zeros from the left to c_j – the number of digits in k_j) set the coordinate of s-th instance on the generalized axis $I^s = 0$, set the position number of coordinate on a generalized axis $p = 1$; looking in a cycle on a group of digits in the interval number $c = 1, 2, ..., c_j$ and on feature numbers j in the descending order of their ranks: put to the p-th bit position (numbering from the left) of the binary representation of the generalized feature I^s the c-th bit (numbering from the left) of interval number, in which the s-th instance hits on the j-th feature and set: $p=p+1$. As a result, we obtain the generalized axis coordinate with implicit ranking of features.
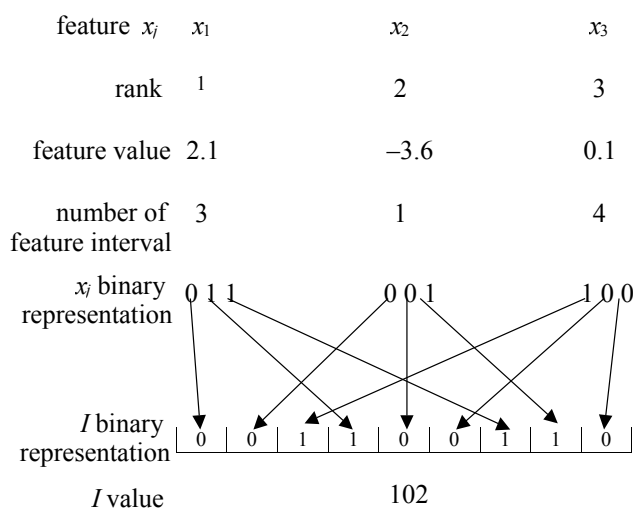

Fig. 2 Illustration of second transformation

Transformation 3. The generalized feature formed on the basis of locality-preserving hashing [12–15]. The initial feature space is divided into 2^k equal hypercubes, each of which identified by the key I^s of a k bit length, where k is a number of feature partitions. After the i-th partition the initial feature space split to 2^i N-dimensional cubes, wherein the i-th partition is carried out on the j-th dimension: $j = i \bmod N$. At the i-th partition if hypercube located in the top half of the partitioned range, then set to one the i-th bit of its key, and otherwise set the i-th bit of its key to zero (set to one the bit in the i-th position of k-bit identifier, extended by zeros from the left, if the length is less than k). The key I^s algorithmically can be generated as follows: set: $I^s = 0$, $x_j^{\min'}=x_j^{\min}$, $x_j^{\max'} = x_j^{\max}$, then for $i = 1, 2, ..., k$ do: set: $j = i \bmod N$, $x_j^{\mathrm{mid}} = (x_j^{\min'}+x_j^{\max'})/2$, $I^s = 2I^s$; if $x_j^s > x_j^{\mathrm{mid}}$, then set: $x_j^{\min'} = x_j^{\mathrm{mid}}$, $I^s = I^s + 1$, else set: $x_j^{\max'} = x_j^{\mathrm{mid}}$.

Transformation 4. The above-described transformations provide mapping to the discrete generalized axis. If the total number of bits to represent numbers of all feature intervals exceeds the number of bits in a bit grid of computer, it is possible to use a transformation to the generalized real axis with partial information loose: add to the real coordinate on generalized feature I^s the c-th bit (numbering from the left) of interval number, in which the s-th instance hits on the j-th feature (2):

$$I^s = \sum_{j=1}^{N} \frac{w_j \hat{x}_j^s}{r_j k_j},\qquad(2)$$

where

$$w_j = \frac{1}{k_j}\sum_{k=1}^{k_j} w_{j,k}\,,\quad w_{j,k} = \left\{ \left. \frac{\max\limits_{q=1,2,...K}\{S_{j,k}^q\}}{S_{j,k}}\right| S_{j,k} > 0 \right\},\quad S_{j,k} = \sum_{q=1}^{K} S_{j,k}^q,$$

$S_{j,k}^q$ is a number of instances of q-th class located in the k-th interval of j-th feature, r_j is a rank of j-th feature (the number of j-th feature in decreasing order of individual feature importance).

Transformation 5. Define the distance from the s-th instance to the unit vector in the normalized coordinate system (3):

$$d^s = \sqrt{\sum_{j=1}^{N}\left(\hat{x}_j^s - 1\right)^2},\qquad(3)$$

and the angle between the instance as a vector and the unit vector (4):

$$\varphi^s = \arccos\left(\frac{\sum_{j=1}^{N}\hat{x}_j^s}{\sqrt{N\sum_{j=1}^{N}\left(\hat{x}_j^s\right)^2}}\right).\qquad(4)$$

Thus we map the s-th instance from the N-dimensional space into two-dimensional space. Next for coordinates of s-th instance in formed two-dimensional space by analogy with the first transformation obtain coordinate of s-th instance on the generalized axis I^s.

Transformation 6. Generate Q support vectors – the centers of pseudo-clusters $C^q = \{C^q_j\}$, $q=1, 2, ..., Q, K \leq Q \ll S, j = 1, 2, ..., N$. In the simplest case their coordinates can be set as random taking into account dimensionality and feature scales ($x^{\min}_j \leq C^q_j \leq x^{\max}_j$), or by setting $Q=K$ to determine the center of each its class (5):

$$C_j^q = \frac{1}{S^q} \sum_{s=1}^{S} \{x_j^s \mid y^s = q\}, \tag{5}$$

where $j = 1, 2, ..., N$, $q=1, 2, ..., K$.

After this calculate the clusters based on their proximity and position in feature space relative to the smallest feature values:

− find the distance from the cluster centers to the point with the lowest feature values (6):

$$R_{\min}(C^q) = \sum_{j=1}^{N} \left(C_j^q - x_j^{\min}\right)^2 ; \tag{6}$$

− find the distance between the cluster centers (7):

$$R(C^q, C^p) = \sum_{j=1}^{N} \left(C_j^q - C_j^p\right)^2 ; \tag{7}$$

− find the center of cluster closest to the point with the lowest feature values (8):

$$q = \arg \min_{g=1,2,...Q} \{R_{\min}(C^g)\}; \tag{8}$$

− set this center as the current, set a new number of current cluster $t=1$, put current cluster in the set of centers with a new index ($C^*=C^* \cup C^{*1}$, $C^{*1} = C^q$) and delete it from the set of centers without a new index ($C=C / C^q$);

− while exist at least one cluster without a new index (i.e. $C \neq \varnothing$) perform: among the remaining clusters without a new index in C find the closest cluster to the current cluster (9):

$$p = \arg \min_{\substack{g=1,2,...Q; \\ C^g \in C}} \{R(C^q, C^g)\}, \tag{9}$$

then increase $t = t+1$, put the current cluster to the set of centers with a new index ($C^*=C^* \cup C^{*t}$, $C^{*t} = C^p$) and remove it from the set of centers without a new index ($C=C / C^p$).

As a result we will receive C^* − a set of cluster centers with numbers corresponding to their proximity to the point with the lowest values of features, and also allowing to determine qualitatively the proximity of the cluster centers.

Further for each instance of the initial sample x^s, $s=1, ..., S$ do:

− define the distance from it to each cluster center, $q=1,2,...,Q$, by (10):

$$R(x^s, C^{*q}) = \sum_{j=1}^{N} \left(x_j^s - C_j^{*q}\right)^2 ; \tag{10}$$

− find the index of the nearest cluster center (11):

$$p = \arg \min_{q=1,2,...Q} \{R(x^s, C^{q*})\} ; \tag{11}$$

− find the angle between the vectors x^s and C^{p*} relative to the point with the lowest feature values (12):

$$\varphi = \arccos \frac{\sum_{j=1}^{N}(x_j^s - x_j^{\min})(C_j^{p*} - x_j^{\min})}{\sqrt{\sum_{j=1}^{N}\left(x_j^s - x_j^{\min}\right)^2}\sqrt{\sum_{j=1}^{N}\left(C_j^{p*} - x_j^{\min}\right)^2}};\qquad(12)$$

– assign the s-th instance with the coordinate on the generalized axis (13):

$$I^s = p + \frac{\varphi}{\pi}.\qquad(13)$$

IV. CHARACTERISTICS OF TRANSFORMATIONS TO THE GENERALIZED AXIS

The introduced above transformations are encouraged to use the following characteristics of instance mapping process:

– t^s is the time of transforming of one instance from the original feature space to the generalized axis for the sequential computations;

– m^s is the computer memory volume used by the transformation method for processing one instance;

– λ is the number of adjustable parameters of transformation needed for its implementation;

– t is the time of calculation of transformation parameters based on the training sample;

– m the computer memory volume used to calculate the transformation parameters on the basis of training sample.

Situations where several instances have equal coordinates may occur in the original and in the synthesized feature spaces. Such situations are called collisions. Under the collision point we will understand the point in the feature space, in which there is a collision.

The collision is quite admissible and even desirable in problems of automatic classification on condition that all instances located at the point of collision belongs to the same class. However, if the instances located at the point of collision, belongs to different classes, the used feature set does not provide a good separability of instances.

Denote the set of points of collision $\{g_v\}$, $v=1, 2, ..., V$, where g_v is a set of instances belonging to a v-th point of collision, V is the number of points of collision, which obviously can not exceed $0,5S$.

To estimate the quality of the results of considered transformations we propose to use the following indicators.

The number of points of collisions in which instances belongs to different classes, after the transformation of the sample to the generalized axis can be defined as (14):

$$\cdot \qquad E_{<I,y>}^{*} = \sum_{v=1}^{V}\{1\mid \exists s, p = 1,2,...,S, p \neq s : I^s \in g_v, I^p \in g_v, I^s = I^p, y^s \neq y^p\}\qquad(14)$$

This indicator in the best case will be equal to zero when there is no collision points, and in the worst case it maximum value will not exceed $0,5S$.

The probability estimation (frequency) of the collision points in which instances belonging to different classes, after the transformation of the sample to the generalized axis can be expressed by (15):

$$P_{<I,y>}^{*} = \frac{2E_{<I,y>}^{*}}{S}.\qquad(15)$$

The corrected number of points of collisions in which instances belong to different classes, after the transformation of the sample to the generalized axis is defined as (16):

$$E^{*'}_{<I,y>} = E^{*}_{<I,y>} - E^{*}_{<x,y>},$$ (16)

where $E^{*}_{<x,y>}$ − is the number of collision points in which the instances belongs to different classes in the initial sample (17):

$$E^{*}_{<x,y>} = \sum_{v=1}^{V} 1 \left\{ \begin{array}{l} \exists s, p = 1,2,...,S, p \neq s : x^s \in g_v, x^p \in g_v, \\ y^s \neq y^p, \forall j = 1,2,...,N : x^s_j = x^p_j \end{array} \right\} .$$ (17)

The indicator $E^{*'}_{<I,y>}$ more accurately characterizes the quality of transformation to the generalized axis because it eliminates the errors present in the sample. In the best case it will be equal to zero when there is no collisions, and in the worst case it maximum value will not exceed 0,5S.

The corrected probability estimation (frequency) of the collision points in which the instances belong to different classes after the sample transformation to the generalized axis can be obtained by (18):

$$P^{*'}_{<I,y>} = \frac{2E^{*'}_{<I,y>}}{S} .$$ (18)

The total number of instances in the collision points in which the instances belong to different classes after the sample transformation to the generalized axis is suggested to calculate by (19):

$$E^{\Sigma}_{<I,y>} = \sum_{v=1}^{V} \left\{ |g_v| \left| \begin{array}{l} \exists s, p = 1,2,...,S, p \neq s : I^s \in g_v, \\ I^p \in g_v, I^s = I^p, y^s \neq y^p \end{array} \right. \right\} .$$ (19)

The more will be value of this indicator, the worse separability of instances on the generalized axis. In the best case it will be equal to zero and in the worst case it will not exceed the number of instances in the sample S.

The probability estimation of instance hitting to the collision point in which instances belong to different classes after the sample transformation to the generalized axis can be obtained by (20):

$$P^{\Sigma}_{<I,y>} = \frac{E^{\Sigma}_{<I,y>}}{S} .$$ (20)

The total number of instances in the collision points of the initial sample in which the instances belong to different classes it is proposed to define as (21):

$$E^{\Sigma}_{<x,y>} = \sum_{v=1}^{V} \left\{ |g_v| \left| \begin{array}{l} \exists s, p = 1,2,...,S, p \neq s : x^s \in g_v, x^p \in g_v, \\ y^s \neq y^p, \forall j = 1,2,...,N : x^s_j = x^p_j \end{array} \right. \right\} .$$ (21)

The more will be the value of this indicator, the worse the separability of instances of the initial sample will be. In the best case it will be equal to zero and in the worst case it not exceed the number of instances in the sample S.

The probability estimation of instance collision in the sample in which instances belong to different classes can be obtained from (22):

$$P_{<x,y>}^{\Sigma} = \frac{E_{<x,y>}^{\Sigma}}{S} \quad . \tag{22}$$

The number of pairwise collision of instances of different classes after the sample transformation to the generalized axis is proposed to determine by (23):

$$E_{<I,y>} = \sum_{s=1}^{S} \sum_{p=s+1}^{S} \{y^s \neq y^p \mid I^s = I^p\}. \tag{23}$$

In the best case, this indicator is zero when there is no any collision, and in the worst case its value will not exceed $S(S-1)$.

The probability estimation (frequency) of pairwise collision of instances of different classes after the sample transformation to the generalized axis can be calculated by (24):

$$P_{<I,y>} = \frac{E_{<I,y>}}{S(S-1)}. \tag{24}$$

The corrected number of pairwise collision of instances of different classes after the transformation of training and (or) test sample to the generalized axis is proposed to determine as (25):

$$E_{<I,y>}^{'} = E_{<I,y>} - E_{<x,y>}, \tag{25}$$

where $E_{<x,y>}$ is a number of pairwise collision of instances of different classes in the original sample (26):

$$E_{<x,y>} = \sum_{s=1}^{S} \sum_{p=s+1}^{S} \{y^s \neq y^p \mid \forall j = 1,2,...,N : x_j^s = x_j^p\}. \tag{26}$$

This indicator $E_{<I,y>}^{'}$ in comparison with the previous indicator more accurately characterizes the quality of the transformation to the generalized axis, because it eliminates the errors present in the original sample. In the best case, it would be equal to zero when there is no any collision, and at worst case, it maximum value will not exceed $S(S-1)$.

The corrected probability estimation (frequency) of pairwise collisions of instances of different classes after sample transformation to the generalized axis can be defined by (27):

$$P_{<I,y>}^{'} = \frac{E_{<I,y>}^{'}}{S(S-1)}. \tag{27}$$

The average number of clusters per class on a generalized axis can be calculated by (28)

$$\overline{k} = \frac{k}{K}, \tag{28}$$

where k is a number of clusters of different classes on a generalized axis.

To determine k, we need order the instances $<I^s, y^s>$ in ascending order on the generalized axis. Then, looking from left to right, we need to identify clusters – the intervals of one-dimensional axis, all instances of each of which belong to only one class.

The less will be the number of such clusters, the simply is partition of generalized axis.

In the best case when the classes are compact, i.e. $k = K$, this indicator is equal to one.

The more will be value of this indicator, the worse the separability of instances will be on the generalized axis.

In the worst case where each instance falls into a single cluster its value will be $\bar{k} = S / K$.

The minimum distance between instances of different classes on the generalized axis is offered to determine by (29):

$$R'_{\min} = \min_{\substack{s=1,\dots,S; \\ p=s+1,\dots,S}} \left\{ I^s - I^p \,\big\|\, y^s \neq y^p \right\}. \tag{29}$$

The more will be value of this ratio, the better classes will be separated on the generalized axis.

The maximum distance between instances of one class on the generalized axis is offered to determine by (30):

$$R'_{\max} = \max_{\substack{s=1,\dots,S; \\ p=s+1,\dots,S}} \left\{ I^s - I^p \,\big\|\, y^s = y^p \right\}. \tag{30}$$

The less will be this indicator value, the more compact instances of each class will be positioned on the generalized axes.

The average ratio of distances on the generalized axis and in the original feature space is proposed to calculate by (31):

$$\overline{\Delta} = \frac{R_{\max}}{0,5S(S-1)R^*_{\max}} \sum_{s=1}^{S} \sum_{p=s+1}^{S} \left\{ \left| \frac{|I^s - I^p|}{R(x^s, x^p)} \right| R(x^s, x^p) > 0 \right\}, \tag{31}$$

where

$$R^*_{\max} = \max_{\substack{s=1,\dots,S; \\ p=s+1,\dots,S}} \left\{ I^s - I^p \,| \right\}, \quad R_{\max} = \max_{\substack{s=1,\dots,S; \\ p=s+1,\dots,S}} \left\{ R(x^s, x^p) \,| \right\}, \quad R(x^s, x^p) = \sqrt{\sum_{j=1}^{N} (x_j^s - x_j^p)^2} \,.$$

The more will be value of this indicator, the better on the average the transformation on the generalized axis reflects location of instances in the original space and features the better separability of instances on the generalized axis;

Average of the relative distance products on the generalized axis and in the original feature space is defined as (32):

$$\Delta = \frac{\displaystyle\sum_{s=1}^{S} \sum_{p=s+1}^{S} | I^s - I^p | \, R(x^s, x^p)}{0,5S(S-1)R^*_{\max} R_{\max}}. \tag{32}$$

This indicator will vary from zero to one: The more will be its value, the better on the average the transformation on the generalized axis reflects the location of instances in the original feature space.

The indicator of generalized axis feasibility of establishing we define as (33):

$$G = \frac{\min\limits_{j=1,2,\dots N}\left\{k_{<x_j,y>}\right\}}{k}.$$ (33)

where $k_{<x_j,y>}$ is the number of intervals of different classes on the axis of feature x_j.

This indicator in the best case will be equal to S/K, and in the worst case will be equal to K/S. If this indicator will be greater than one, the use of the generalized axis will be feasible, otherwise it can be replaced with the original feature, characterized by the smallest number of intervals of different classes.

V. The Comparison Criteria of Generalized Axis Transformations

On the basis of the indicators characterizing the basic properties of on the generalized axis transformations introduced in the previous section it is possible to determine the criteria for comparison, the criteria for performing and criteria for evaluating the quality of the results of transformations.

The criteria for evaluation of the transformation process is proposed to define as the following:

– the combined criterion of the minimum of time and memory on the instance transformation: $F_1 = t^s m^s \xrightarrow{} \min$;

– the combined criterion of the minimum of time and memory to determine the transformation parameters for the training sample: $F_2 = \lambda\, tm \xrightarrow{} \min$;

– the integral criterion: $F_3 = t^s m^s + \lambda S^{-1} tm \xrightarrow{} \min$.

The criteria for evaluating the quality of results of transformations:

– the criterion of the minimum of probability of instance group collisions: $F_4 = P'_{<I,y>} \rightarrow \min$;

– the criterion of the minimum of probability of instance pair collisions: $F_5 = P^{*'}_{<I,y>} \rightarrow \min$;

– the combined criterion of the minimum probability of pair and group collisions: $F_6 = 0.5(P'_{<I,y>} + P^{*'}_{<I,y>}) \rightarrow \min$;

– the maximum of class compactness-separability: $F_7 = \bar{k} \rightarrow \min$;

– the integral criterion of minimum of collisions-compactness-separability of classes calculated by (34):

$$F_8 = \frac{\bar{k}}{2}\left(P'_{<I,y>} + P^{*'}_{<I,y>}\right) \rightarrow \min,\ \bar{k} > 0;$$ (34)

– the integral criterion of minimum of collision-maximum of compactness-separability of classes and maximum of average of relative distance products on the generalized axis and in the original feature space (35):

$$F_9 = \frac{\bar{k}\left(P'_{<I,y>} + P^{*'}_{<I,y>}\right)}{1 + \Delta e^{-\bar{\Delta}+1}} \rightarrow \min.$$ (35)

– the integral criterion of the minimum of collisions-maximum of a generalized axis establishing feasibility-compactness-separability of classes and the maximum of average of relative distances products on the generalized axis and in the original feature space (36):

$$F_{10} = \frac{\left(P'_{<I,y>} + P^{*'}_{<I,y>}\right)}{G + \Delta e^{-\bar{\Delta}+1}} \rightarrow \min.$$ (36)

VI. EXPERIMENTS AND RESULTS

The proposed transformations to the generalized axis have been implemented as software and experimentally studied in practical problem solving of technical and medical diagnosis, and of automatic classification [3, 16] (see Table 1). The conducted experiments confirmed the efficiency and the practical suitability of the developed mathematical tools. The experiments

TABLE I

THE CHARACTERISTICS OF INITIAL DATA SAMPLES AND THE FRAGMENT OF THE EXPERIMENTAL RESULTS TO STUDY THE TRANSFORMATIONS ON GENERALIZED AXIS

Characteristics	Task			
	Gas-turbine air-engine blade diagnosis	Chronic obstructive bronchitis diagnosis	Agricultural plant recognition on the remote sensing data	Fisher Iris classification
S	32	205	3226	150
N	513	28	256	4
K	2	2	3	3
Best transformation:				
number	2	1	4	1
$E^*_{<I,y>}$	1	0	0	1
$P^*_{<I,y>}$	0.0625	0	0	0.013333
$E_{<I,y>}$	1	0	0	9
$P_{<I,y>}$	0.0010081	0	0	0.00040268
$E^*_{<x,y>}$	0	0	0	0
$E^{*'}_{<I,y>}$	1	0	0	1
$E_{<x,y>}$	0	0	0	0
$E'_{<I,y>}$	1	0	0	9
$E^{\Sigma}_{<x,y>}$	0	0	0	0
$P^{\Sigma}_{<x,y>}$	0	0	0	0
$E^{\Sigma}_{<I,y>}$	2	0	0	10
$P^{\Sigma}_{<I,y>}$	0.0625	0	0	0.066667
k	15	72	1773	11
R_{max}	70.114	918.99	7.5551	7.0852
R^*_{max}	$1.5228 \cdot 10^9$	$1.6927 \cdot 10^9$	3.3217	$1.5126 \cdot 10^9$
R'_{min}	0	1056	$1.1437 \cdot 10^{-7}$	0
R'_{max}	$1.5228 \cdot 10^9$	$1.6833 \cdot 10^9$	3.318	609746944
$\overline{\Delta}$	0.82534	0.69432	0.37993	0.88456
Δ	0.095099	0.098493	0.028541	0.18511
G	0.2	0.69444	0.56007	1
t^s	0.00097501	0.00053269	$3.8686 \cdot 10^{-5}$	0.000416
m^s	4309	262.83	2057.9	82.133
t	0.2184	0.093601	4.524	0.1092
m	333484	109644	6764132	20340
λ	1539	84	1024	12
F_1	4.2013	0.14001	0.079613	0.034168
F_2	$1.1209 \cdot 10^8$	$8.6207 \cdot 10^5$	$3.1336 \cdot 10^{10}$	$0.26654 \cdot 10^5$
F_3	$3.5028 \cdot 10^6$	4205.4	$9.7134 \cdot 10^6$	177.73
F_4	0.0010081	0	0	0.00040268
F_5	0.0625	0	0	0.013333
F_6	0.031754	0	0	0.006868
F_7	7.5	36	591	3.6667
F_8	0.23816	0	0	0.025183
F_9	0.42786	0	0	0.041701
F_{10}	0.20274	0	0	0.011373

have shown that the proposed transformations allow to significantly reduce the data sample dimensionality.

The Table 1 shows the best transformations for different practical tasks according to proposed criteria.

The proposed transformations can be recommended for use in the construction of diagnostic and recognizing models by precedents, as well as for the formation of the training samples from the source samples of large number of instances.

VII. CONCLUSION

The urgent problem of the development of mathematical support for data dimensionality reduction was solved in the paper.

The scientific novelty of results consists in that the set of rapid transformations from the original multi-dimensional space into one-dimensional axis was firstly proposed. The proposed transformations provide a solution both the problem of constructing of artificial features (feature extraction), and the problem selection of the most significant features (feature selection).

− the set of rapid transformations from the original multi-dimensional space into one-dimensional axis was firstly proposed. It is based on the principles of hashing and provides taking into account the instance locations in the feature space with respect to the class centers of gravity, and also allows to determine and to take into account the feature weights and thereby implicitly solves the problem of feature selection. Thus, the proposed transformations provide a solution both the problem of constructing of artificial features (feature extraction), and the problem selection of the most significant features (feature selection);

− the complex of indicators characterizing the properties of transformations from multidimensional space to generalized axis was firstly proposed. On the basis of the proposed indicators the set of criteria is defined. It facilitates comparison and selection of the best transformations and results of their work at diagnosis and recognition problems solving by precedents.

The practical significance of obtained results is that:

− the software realizing proposed transformations and indicators characterizing their properties was developed. Its usage allows to automate the data dimensionality reduction and analysis of its results;

− the experimental investigation of the proposed transformations and the indicators characterizing them was conducted at practical problem solving. The results of research allow to recommend the proposed transformations for use in practice for diagnosis and pattern recognition problem solving.

The prospects of further utilization of the results obtained in this work consists in the possibility of their use for automation of training and testing samples formation (data dimensionality reduction through decrease of the number of precedents by the most important precedents extraction).

REFERENCES

[1] R. Jensen, and Q. Shen, *Computational intelligence and feature selection: rough and fuzzy approaches*, Hoboken: John Wiley & Sons, 2008.

[2] O. V. Babak, "The solution of some problems of data processing on the basis of the general generalized variable" ["Reshenie nekotorykh zadach obrabotki dannykh na osnove metoda generalnoy obobshchennoy peremennoy"], *Problemy upravleniya i informatiki*, 2002, no. 6, pp. 79–91. (in Russian)

[3] S. A. Subbotin, An. A. Oleinik, E. A. Gofman et al., *Intelligent information technologies of automated diagnostic and pattern recognition systems design : monograph [Intellektualnye informatsionnye tekhnologii proektirovaniya avtomatizirovannykh sistem diagnostirovaniya i raspoznavaniya obrazov : monografiya]*. Ed. by S. A. Subbotin. – Kharkov: Company "SMIT", 2012.. (in Russian)

[4] T. W. Lee, *Independent component analysis: theory and applications*, Berlin: Springer, 2010.

[5] J. A. Lee, and M. Verleysen, *Nonlinear dimensionality reduction*, New York: Springer, 2007.

[6] C. Padraig, "Dimension reduction," Technical report UCD-CSI-2007-7, Dublin: University College Dublin, 2007.

[7] Y. Jiang, R. Zhang, G. Liu et al., "Multifactor dimensionality reduction for detecting haplotype-haplotype interaction," in *Proc. Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, Tianjin, 2009, pp. 241 –245.

[8] B. Kulis, A. C. Surendran, and J. C. Platt, "Fast low-rank semidefinite programming for embedding and clustering," in *Proc. Eleventh International Conference on Artificial Intelligence and Statistics*, San Juan, 2007, vol. 2, pp. 235–242.

[9] O. V. Babak, and A. E. Tatarinov, "An approach to solving classification problems under incomplete information," Cybernetics and Systems Analisys, 2005, vol. 41, no. 6, pp 891–897.

[10] V. I. Vasilev, A. I. Shevchenko, and S. N. Esh, *Principle of reductions it the problems of dependence detection: monograph [Printsip reduktsii v zadachakh obnaruzheniya zakonomernostey : monografiya]*, Donetsk: Nauka i osvita, 2009. (in Russian)

[11] S. Yu, "Feature selection and classifier ensembles: a study on hyperspectral remote sensing data," Proefschrift ... doctor in de wetenschappen. Antwerpen: Universitaire Instelling Antwerpen, 2003.

[12] J. Jianqiu, J. Li, Sh. Yany, B. Zhang et al., "Super-bit locality-sensitive hashing," in *Advances in Neural Information Processing Systems*, P. Bartlett et al. Eds. 2012, vol. 25, pp. 108–116.

[13] A. Andoni, and P. Indyk, "*Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions*," Communications of the ACM, 2008, vol. 51, no. 1, pp. 117–122.

[14] X. Yang and Y. Hu, "A scalable index architecture for supporting multi-dimensional range queries in peer-to-peer networks," in *Proc. International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom-2006)*, Atlanta, 2006, pp. 1–10.

[15] A. Andoni, M. Datar, N. Immorlica, P. Indyk, and V. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Nearest neighbor methods in learning and vision: theory and practice*, T. Darrell, P. Indyk, G. Shakhnarovich Eds. Massachusetts: MIT Press, 2006, pp. 55–67.

[16] R.A. Fisher, Iris Data Set. Available on-line: https://archive.ics.uci.edu/ml/datasets/Iris