Contemporary Engineering Sciences, Vol. 4, 2011, no. 4, 187 – 200

Theoretical Grounds of Information Representation and Processing by Using the Potential Codes

Raimy Abdourahmane[a], Nickolay M. Bykov[b] and Aliona E. Segeda[b]

[a]Cheikh Anta-DiopUniv., Dakar, Senegal
araimy16@yahoo.fr

[b]Vinnitsa State Tech. Univ., 95 Khmelnitskoe shosse, Vinnitsa, Ukraine 21021

Abstract

The method of unified representation of the systems knowledge about objects of an external world by rank transformation of their descriptions, made in the different features spaces - deterministic, probabilistic, fuzzy and other, and also models of rank configurations are resulted. The method of the rank configuration description by the DRP - code (distance rank preserving code) is developed. The problems of its completeness, information capacity, noise immunity and secrecy are reviewed.

Keywords: Information representation and processing, potential codes, model, rank configuration, making decision

1. INTRODUCTION

Evolution of intelligent checking and control systems leads to constant involving of the increasing quantity of approaches to a data analysis, their representation, processing and decision making. It is called by variety of objects under control and control goals and often produces large difficulties in designing effective methods of a data analysis, knowledge discovering and their usage for decision making on control. Problem of understanding of these approaches and reductions them to «a general denominator» appeared on this reason. Authors in the given activity are attempting to solve this problem ground of such fundamental concept, as a degree (rank) of object states remoteness in the features space. The beginning of this idea put those fact, demonstrated in the literature, that the identification of states of the object under control and its performance optimization can be executed ground usage of the information about distances between its states in the descriptions space. The evidence of this fact for the spaces with the deterministic and probabilistic descriptions can be found, for example, in [1, 2], and for the spaces with the fuzzy and possibility descriptions - in [3, 4]. In development of the mentioned idea in the given activity authors demonstrates that the sufficient information for implementation of process of states identification and optimizations of system operation are rather rank relation between these states but not values of distances. The given approach can be applicable also to the non-metric spaces in case of their preliminary transformation to metric by known methods of multidimensional scaling [5]. The concept of rank configuration of the space of states to describe the rank relation between distances is entered. The operability of the rank configurations for computerized data analysis and decision making is determined by a method of their description. In this connection in activity the computationally effective method of binary coding is offered, which one simultaneously determines both the object (state), and the information on distance ranks to it from all other objects (states). Such code authors call a potential code, or DRP-code (distance rank preserving code). Due to its manner of construction DRP-code is friend for a grouping of points and revealing of the main patterns in the parametric space that allows to design unified and improved data clustering algorithms.

2. MAIN THEORETICAL GROUNDS

We shall first accept following definitions. Let us $\{C_i(S)\}$, $i=\overline{1,k}$ denotes the set of the object (or system) states, where $C_i(s)$ represents one of these states and is a concatenation of some members $s_j \in S$, $,j=\overline{1,m}$; k is a cardinal number of the set of states and m is a number of members in the set S. Assume s_j to be described as a point or a cluster of the n-dimensional parametric space. Practical examples of unit s_j can be sounds or phonotypes of speech, primitives of the images, vertices of the cognitive maps, neurons of a neural network, patterns inside the database, members of an object or some system. Examples of states can be speech words or sentences, structural combination of the image primitives, combination of the neuron reactions, combination of the members of some object or systems, etc. For the convenience we shall call further the member as a character or symbol and the state as a string. As the influencing of an environment distorts the description space of characters, the decision making about system states (its identification) implements via the minimal distance rule

$$d[C_i(s),C_j(s)] = \min \to C_i(s) = C_j(s)\,, \tag{1}$$

$$d[C_i(s),C_j(s)] = \sum_l d(s_l^i, s_l^j)\,, \tag{2}$$

$C_i = (s_1^i, s_2^i, \dots, s_l^i, \dots, s_q^i)$, $C_j = (s_1^j, s_2^j, \dots, s_l^j, \dots, s_q^j)$, $l=\overline{1,q}$, q – length of a string.

The description of characters in different parameter spaces (deterministic, probabilistic, fuzzy and other) generates a diversity of the similarity measures in these spaces and, accordingly, algorithms of calculus of distances $d(s_l^i, s_l^j)$ in the formula (2) between characters s_l^i, s_l^j.

Characters s_j in the generally accepted approach are represented in the computer storage by binary codes, thus the coding is designed only with allowance to requirements of their discrimination and noise immunity. For preserving of the information about the system components spatial configuration, which one characterizes its states, it is necessary to store except of m members codes also $m\ (m\ -1)/2$ codes distances between members in space of parameters. The process of finding of distances between strings of characters demands fulfillment of arithmetic summation of distances between pairs of characters from reference and recognizing strings. Thus it is necessary each time to read out from memory the numbers which describes distances between pairs of characters. Obviously, that the generally accepted approach to coding has such lacks, as a necessity for expenditure of the memory resource on $m\ (m\ -1)/2$ distance codes storing, and also reducing the rate of decision marking because of time loss on reading of these codes from memory. In this connection we offer such way of a binary representation a component of vectors, at which one the information on distances (or similarities) between components of these vectors would be placed in their codes [2]. Thus, as will be proved, for preservation of the decision making adequacy the space of binary codes should be isomorphous to initial space of symbols to within ranks of distances. Such codes are called as DRP-codes (distance rank preserving codes), or potential (by analogy with a field of electric charges, in which one the value of charges determines force of interplay between them).

Definition 1. The binary representation of character string $C_i = (s_1^i, s_2^i, ..., s_l^i, ..., s_q^i)$, or binary string is the sequence $u_1 = (b_1, b_2, ..., b_l, ..., b_q)$ of binary codes $b_l = \{0,1\}^n$ of length n such, that

$$B = \varphi(S). \qquad (3)$$

In expression (3) φ is some mapping of character set S to set of binary codes $B = \{b_1, b_2, ..., b_q\}$. It is necessary to determine the kind of this mapping.

Definition 2. The difference (distance) h_{ij} between two binary codes b_i and b_j of length n we call the binary value obtained by fulfillment of definite logic operation $\oplus$ above these codes:

$$h_{ij} = b_i \oplus b_j. \qquad (4)$$

The kind of mapping: φ: $s_i \to b_i$ can be determined taking into account those fact, that the recognition of system states is carried out on the rule of minimum distance to one of reference states from the set C, namely:

$$C_i(s) = C_j(s), \text{ if } d_{ij}(C) = \min(d_{il}(C)), \qquad l = \overline{1,k}, \qquad l \neq i. \qquad (5)$$

From here by natural way the *statement P1* follows: the set of differences H between binary strings should be isomorphous mapping of the ordered set D between the symbols string:

$$H = f(D), \tag{6}$$

where $H = \{h(u_{ij})\}$, $h(u_{ij}) = \sum_{l=1}^{q} h(b_l^i, b_l^j)$, $D = \{d(C_{ij})\}$, $d(C_{ij}) = \sum_{l=1}^{q} d(s_l^i, s_l^j)$.

We assume further $h(b_l^i, b_l^j) = h_{ij}$, $d(s_l^i, s_l^j) = d_{ij}$.

On the basis of this statement we can determine the kind of mapping φ in expression (3), by proving previously such theorem:

Theorem 1:

$$\mathrm{T1} \overset{df}{\equiv} \mathrm{P1} \Rightarrow \underset{i,j}{\forall}((d_{i1j1} \le d_{i2j2}) \Rightarrow (h_{i1j1} \le h_{i2j2})), \qquad i \in \{1,...,m-1\}, j \in \{1,...,m\}, \tag{7}$$

which one says, that if has a place the statement P1, the ordered set of differences between binary codes of symbols should be an isomorphous mapping of the ordered set of distances between symbols..

Proof of necessity. Of the statement P1 follows, that for any concatenation of q symbols of m possible in words C_i, C_j, C_k, and concatenation of binary codes $b_i \in B$, which corresponds to these symbols in binary words u_i, u_i i u_i, the following relation should be fulfilled:

$$d(C_i, C_j) \le d(C_i, C_k) \Rightarrow h(u_i, u_j) \le h(u_i, u_k) .$$

As the statement P1 should be real for any pair of strings, which are arbitrary concatenations in a symbols subset of set S, we'll define some set of words C_s of power m in following manner. Let's construct m strings $C_i = (s_{i1},...,s_{iv},...,s_{ik},...,s_{iq})$ with length of q such, that the symbols, which occupies in strings an identical v-th position, are identical, except for some k-th position, for which $i_k = \gamma$, $\gamma = \{1,2,...,m\}$. This position is engaged sequentially by all symbols of the set S . Similarly we shall construct set U_b of binary strings $u_i = (b_{i1},...,b_{iv},...,b_{ik},...,b_{iq})$, where $b_{iv} = \varphi(s_{iv})$ is a binary representation of a character s_{iv}. For set of words C_s we shall construct a matrix $\left\|d_{ij}^C\right\|$ of distances between words, which members d_{ij} are calculated pursuant to expression (2). For set of binary strings U_b we shall construct a matrix of differences $\left\|h_{ij}^u\right\|$, the members h_{ij} of which are calculated on the basis of expression (4). Then it is possible to write following expression to be equivalent to the statement P1:

$$\mathrm{P1} \Rightarrow \underset{i,j}{\forall}((d_{i1j1}{}^C \le d_{i2j2}{}^C) \Rightarrow (h_{i1j1}{}^U \le h_{i2j2}{}^U)), \qquad i \in \{1,...,m-1\}, j \in \{1,...,m\}. \tag{8}$$

But due to used principle of the string construction are justified following equations:

$$\left\|d_{ij}^C\right\| = \left\|d_{ij}\right\|, \tag{9}$$

$$\left\| h_{ij}^{U} \right\| = \left\| h_{ij} \right\|, \tag{10}$$

where $\left\| d_{ij} \right\|$ and $\left\| h_{ij} \right\|$ - distance matrices constructing on the sets S and B accordingly. Taking into account the relations (9) ,(10) and (8), we can write following:

$$\underset{i,j}{\forall}((d_{i1j1}{}^{C} \le d_{i2j2}{}^{C}) \Rightarrow (h_{i1j1}{}^{U} \le h_{i2j2}{}^{U})) \equiv \underset{i,j}{\forall}((d_{i1j1} \le d_{i2j2}) \Rightarrow (h_{i1j1} \le h_{i2j2}))\,. \tag{11}$$

From (8) and (11), by equitable of equivalence relation transitivity, it is possible to write:

$$\mathrm{P1} \equiv \underset{i,j}{\forall}((d_{i1j1} \le d_{i2j2}) \Rightarrow (h_{i1j1} \le h_{i2j2}))\,. \tag{12}$$

Let us introduce (12) in the form:

$$\mathrm{P1} \Leftrightarrow \mathrm{P2} = \Im\,, \tag{13}$$

where $\mathrm{P2}$ - expression in a right side of relation (11), and $\Im$ - character of the logical verity. From the logical formula (13) follows $\mathrm{P1} \Rightarrow \mathrm{P2}$, or: $\mathrm{P1} \Rightarrow \underset{i,j}{\forall}((d_{i1j1} \le d_{i2j2}) \Rightarrow (h_{i1j1} \le h_{i2j2})), \quad i \in \{1,...,m-1\}, j \in \{1,...,m\}$,

that we just need to proof.

Proof of sufficiency. From the Boolean formula (13) follows $\mathrm{P2} \Rightarrow \mathrm{P1}$:

$$\underset{i,j}{\forall}((d_{i1j1} \le d_{i2j2}) \Rightarrow (h_{i1j1} \le h_{i2j2})) \Rightarrow \mathrm{P1}\ ,$$

so this is a proof of sufficiency.

Consequent 1. For achievement of identity of decision making results on set of binary strings to decision making results on set of character strings it is necessary and enough the mapping of expressions (6) to be isomorphism.

Consequent 2. Because of all calculations at proving of the *theorem 1* were conducted only with allowance for arrangement of distances in serially ordered sets, instead of their absolute values, it is possible to affirm, that the mapping $\varphi : S \to B$ saves the rank orders of distances. In other words the space of binary codes should be isomorphous to characters space to within ranks of distances.

Definition 3. Code B, which preserves ranks of distances (DRP code), is mapping $i \to B_i$ of the set $M = \{1, 2,..m\}$ in set $\{0,1\}^n$ of binary sequences of length n such, that

$$\underset{i,j}{\forall}(R(d_{ij}) = r \Rightarrow R(h_{ij}) = r), \;\; r = \overline{1, m_r,}\;\; i, j \in M\ . \tag{14}$$

In expression (14) $R(d_{ij})$ - rank of distance d_{ij} between symbols i and j in space of objects, $R(h_{ij})$ - rank of difference h_{ij} in space of binary codes, r - integer, concrete value of a rank, m_r - maximum value of a rank.

Let us define properties of the initial space of symbols, which provide an ability of DRP coding. Obviously, that due to adopted definition of distances between binary codes (formula (4)), the configuration of initial space of symbols which are being a subject to coding, should be such, that the set D of distances between symbols was linearly ordered with relation of the stringent order. It is possible to show three cases resulting in disturbance of the above mentioned order relation.

1.There are ties in the chain which presents ordered set of distances:

$$d_{i1j1} = d_{i2j2} = d_{i3j3},$$

where one of the indexes is identical in all these distances, for example:

$$i_1 = i_2 = i_3 \ .$$

In this case the symbols j_1, j_2 and j_3 have equal distances to the symbol i_1, and in space of binary codes from the expression for binary differences $h_{ij} = b_i \oplus b_j$ follows as a result the upsetting of uniqueness of mapping $\varphi : S \to B$.

Described situation can be corrected by taking into account the fact, that in this case a classifier usually does arbitrary choice. It allows to make stringent ordering of the given chain section of distances by assigning them arbitrary different ranks in the range $[k+1;\ l-1]$, where k - rank of that chain distance, which precedes a tie, and l - rank of that chain distance, which follows a tie. If we have some a priori information about encoded symbols, for example, we know probability of symbols occurrence, an arbitrariness of distance ranks determining can be reduced by way of assigning ranks, for example, pursuant to value of probability.

2. Case similar to case 1, exception that all distance indexes are different. It corresponds to a presence of different equidistant pairs of objects in the parameter space; it means that ranks of equal distances are not incidental to the same symbol. The given case demands more serious special analysis on the following reasons. Firstly, under certain conditions in a described situation it is possible to construct such code words on set U_b, which ones are different, but give identical distances (for example, one of equal spacing intervals is received from a pair of codes b_i, b_j, and other - from a pair of their inversions $\overline{b_i}, \overline{b_j}$). However capability of covering of the initial space configurations by a binary DRP-code decreases in this case. Secondly, attempt of an ordering of distances set by the same way as in first case can lead to distortion of the recognition outcomes in space of binary codes, that it is not difficult to show.

3. There are infinitely removed symbols in space. This case is similar to partial orderliness of set, as there are incomparable members. Such situation, which one can meet in the non-metric spaces, is unrecoverable, in this connection these spaces can not be mapped by a DRP code.

Summing the above-stated reasons, we can say, that at the given stage of analysis the speech will go about the binary mapping by a DRP code of the metric spaces with given relation of the stringent order on the linearly ordered set of distances between symbols, or spaces with relation of the weak order in the event that ranks of equal distances are adjacent to the same symbol.

*Definition 4***.** The rank configuration K_m of space of m symbols is the set of the $(m-1)$ - element subsets, the members of which are the ranks of distances, adjacent the same symbol. Configuration can be represented by geometrical, combinatorial, algebraic or topological model. To geometrical models it is possible to relate the graphs and multidimensional simplexes. On fig.1 the images of the rank configuration as the three-dimensional simplex and the graph are shown.

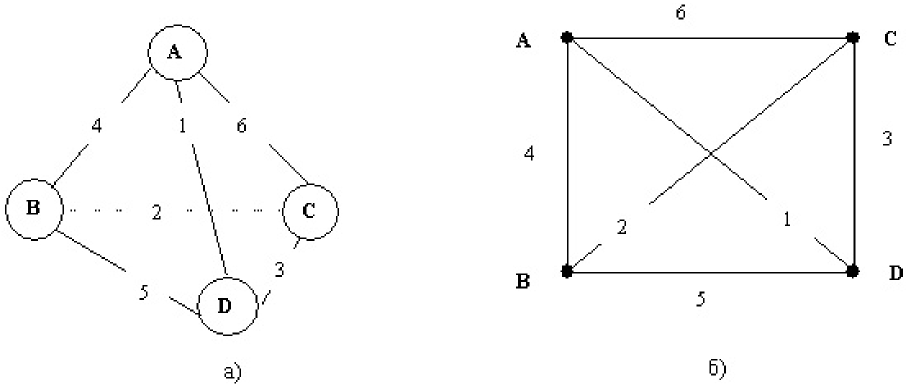


Figure 1: Geometrical models of rank configurations: a) - three-dimensional simplex, b) - complete regular graph

The algebraic model enables describing of the rank configuration pursuant to *Definition 4* by the way of set of rank subsets. For example, the rank configuration on fig.1 can be written to an algebraic view as

$$K_4 = \{\{1,4,6\},\{2,4,5\},\{2,3,6\},\{1,3,5\}\} \qquad (15)$$

The combinatorial models of rank configurations can be described as the designs or block designs or group of permutations on the rank adjacent matrix.

The $t-$design with parameters (n,w,l) (or $t-(n,w,l)-$scheme) calls the collection B of subsets (blocks) of set R, which one contains of n points, such, that any of subsets contains of w points, and any set of t points is contain exactly in l subsets from B [6]. The block design is a type of the t - design (t – scheme). Concretely block design (or 2-scheme) is set of n members, located in m blocks, each block consist of w members, each of pairs of different members is placed in l units, and each member occurs exactly in v blocks. In [6] is proved, that any complete regular graph is described by the block design. For the block design the following equations are implemented:

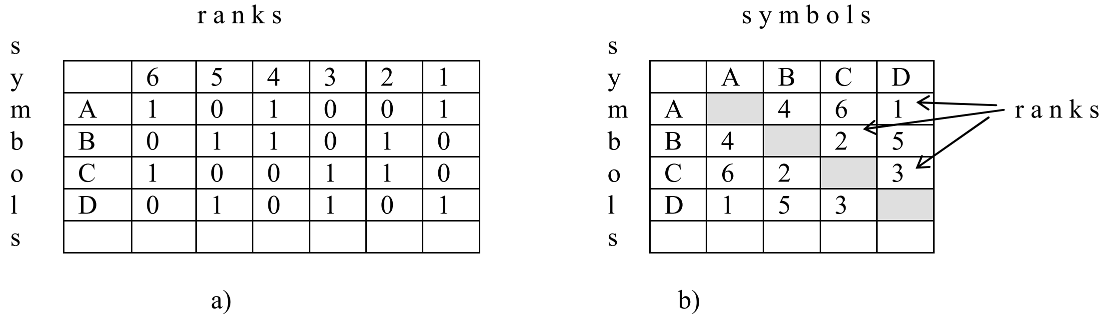
$$nv = mw \qquad (16)$$

$$l(n-1) = v(w-1)\,. \qquad (17)$$

Variable m in the expression (16) denotes number of blocks in set B and corresponds number of encoding symbols; quantity of members n corresponds total number of ranks; w corresponds quantity of ranks, which ones are incident to one symbol. Design of the rank configuration on fig.1 has such parameters:

$$m = 4; n = \frac{m(m-1)}{2} = 6; w = m-1 = 3; l = 1; v = 2\,.$$

Substitution of this value to the expressions (16) and (17) shows default of condition (17) to rank configuration. It means, that the block design can not be adopted as combinatorial model of the rank configuration, example of which is represented in fig.1 and expression (15). Such model is convenient for presenting by the adjacency matrix of the complete regular graph of the rank configuration or the rank adjacency matrix. For the configuration, shown in a figure 1, the incidence matrix and the rank adjacency matrix are shown accordingly in a fig. 2,a) and 2,b).



	6	5	4	3	2	1
A	1	0	1	0	0	1
B	0	1	1	0	1	0
C	1	0	0	1	1	0
D	0	1	0	1	0	1

a)

	A	B	C	D
A		4	6	1
B	4		2	5
C	6	2		3
D	1	5	3	

b)

Figure 2: Figure: combinatorial representation of the rank configuration: a) - incidence matrix of the graph in fig. 1,b); b) – rank adjacency matrix of simplex in fig.1,a)

The description of the rank configuration by the rank adjacency matrix allows drawing a conclusion, that the rank configuration can be represented by group on a set of permutations. This property allows to encode not only symbols, but also configurations, and to store the information on them in a convolute kind.

Application of the rank adjacency matrix allows to determine cardinality of set of the configurations for a given number m of symbols without considerable complexities, and also to elaborate algorithms of symbols encoding by a DRP-code.

Theorem 2. Total number M_c of rank configuration (cardinality of the set of rank configurations) for the unlabelled m - dimensional simplex is equal

$$M_c = \frac{[m(m-1)/2]!}{m!}. \tag{18}$$

Proof. The number n of different distance ranks for $(m-1)$ - dimensional simplex (see fig. 1, b) is equal the number of different pairs of vertices in this simplex:

$$n = \frac{m(m-1)}{2}. \tag{19}$$

Consequently, total number of these ranks permutation over the ranks adjacent matrix is equal $n!$, that is $[m(m-1)/2]!$. This produces $n!$ combination of $(m-1)$ ranks from n in the row of ranks adjacency matrix. There are precisely $m!$ identical combinations of ranks among them, which corresponds to all possible permutations of m rows of mentioned matrix.

Therefore, the cardinality M_c of the configuration set of unlabelled $(m-1)$-dimensional simplex is equal

$$M_c = \frac{[m(m-1)/2]!}{m!}.$$

Theorem is proved.

The concept of a unlabelled simplex will be used when DRP codes are designed for the purposes of an information transmission. If the developed codes are supposed to be applied at problem of decision making, it is necessary to use concept of a labelled simplex. Obviously, that in this case the cardinality of the set of the rank configurations is equal to number of permutations of ranks:

$$M_c = n!. \tag{20}$$

The growth of quantity of the configurations under condition of ascending of number m of symbols corresponds to the law of ”combinatorial explosion”. For example, if $m = 4$, the quantity of the rank configurations of an unlabelled simplex according to expression (19) is equal:

$$M_4 = \frac{6!}{4!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4} = 30,$$

for 5 symbols $M_5 = 30240$, for 6 symbols $M_6 = 1{,}8 \cdot 10^9$. For a labelled simplex the quantity of the configurations increases even faster with growth of the number of encoded symbols, what follows from expression (20). This fact illustrates large information capacity of the DRP code.

3. BASIC PROPERTIES OF DRP CODES

Basic properties of codes are a code length, information capacity, noise immunity and privacy. In this activity a rank code designed for the clustering purposes are considered. Such code we use for structuring of the data represented by the points in space of parameters. In this case a DRP code is used for comparing of distances ranks between different points hence the fulfillment of the axiom of a reflexivity is not required, as the distance and its rank for the same point are not determined. In the given consideration the rank $R(h_{ij})$ of distance between binary words b_i and b_i of a DRP code is determined via operation of logical multiplying AND in according to expression:

$$R(h_{ij}) = \log_2(b_i \Lambda b_j), \tag{21}$$

where the symbol Λ denotes operation AND.

The possibility of construction of the complete DRP code, which is capable to map in a binary kind any rank configuration, follows directly from an example on a figure 2 and particular evidence does not demand. In a figure 2, a) the rows of the graph adjacency matrix will derivate the words of DRP code of the given rank configuration. For example we will now define the rank of distance between the code words of symbols A and B by using an expression (21):

$$R(h_{AD}) = \log_2(b_A \Lambda b_D) = \log_2(101001\Lambda 010101) = \log_2 2^1 = 1 .$$

In practice the operation of the logarithm taking for definition of a precise rank is not necessary, because the ranking of distances between the code words implements by operation AND.

The digit capacity of the DRP code in a considered case is equal to quantity of ranks of the $(m-1)$ - dimensional simplex:

$$n = \frac{m(m-1)}{2} .$$

As the DRP code is a permutation code with a constant weight the defining of its noise immunity can be made pursuant to known relations for these codes [7] . The probability of k errors occurrence in n positions for a symmetrical transmit ting channel is equal:

$$P(k,n) = \binom{n}{k} p^k (1-p)^{n-k} ,$$

where p - probability of one error, and $\binom{n}{k}$ - combination of k symbols from n . Then the probability of not detection of one error in the code word (the joint probability of conversion of one unit in zero point and one zero point in unit) for a DRP code is equal:

$$P_n = \binom{m-1}{1} p(1-p)^{m-2} \binom{n-m+1}{1} p(1-p)^{n-m} , \tag{22}$$

where m - quantity of words in a code. For a code of an example in a figure 2, a) we have:

$$P_n = \left(\binom{3}{1} p(1-p)^2 \binom{4}{1} p(1-p)^3 \right)^2 . \tag{23}$$

The degree 2 in expression (23) corresponds to the that fact, that the error in code does not find out only in that case, when units occupying identical position in two words are simultaneously distorted. By accepting value $p = 1 \cdot 10^{-4}$, we shall receive numerical $P_n = 12 \cdot 10^{-16}$.

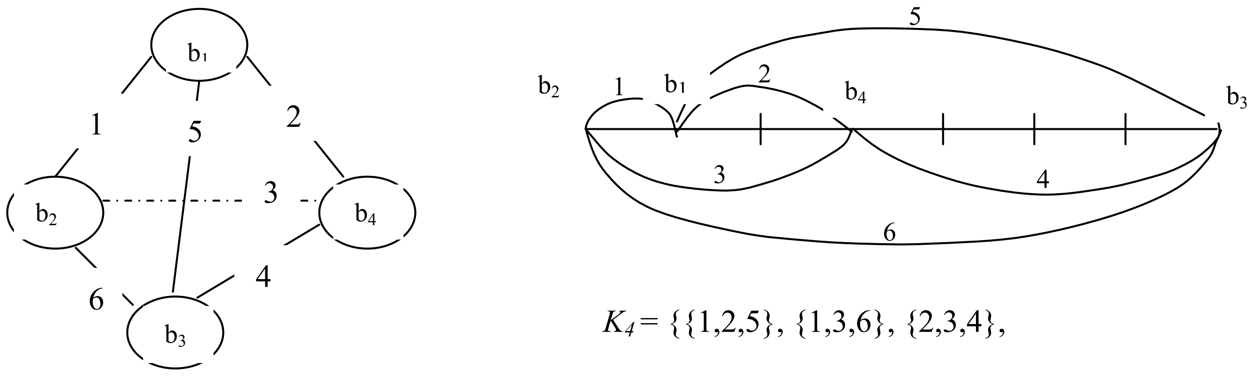
Information capacity of a DRP code we shall define as a ratio of quantity Q_r of the received data words to quantity Q_T of the transferred code words:

$$I_C = \frac{Q_r}{Q_T} = \frac{m + K_m \cdot n}{m \cdot K_m}. \tag{24}$$

For example, transmission by DRP - code of 30 different rank configurations of a 3 –dimensional simplex (the figure 1,a) requires to transmit of $Q_T = 4 \times 30 = 120$ words, from which ones it is possible to extract $Q_r = 4 + 30 \cdot 6 = 184$ of data words, whence $I_C = 184/120 \approx 1{,}53$. That is, the channel capacity in this case exceeds 1. The expressions (24) and (19) demonstrate the increasing of a code information capacity when ascending of dimension m of an encoded simplex. The channel capacity can be more increased by transmitting of not all rank configuration but its code corresponding to a sequence number of rank permutation. The privacy of a code thus can be ensured via selection of one of the possible orders of ranks permutation on a adjacency matrix and is determined by probability of guessing of this number:

$$C_R = \frac{1}{n!}. \tag{25}$$

For those applications that require the existence of the axiom of identity, the code should be designed taking into account the additive operation "exclusive OR" (XOR). These applications include, for example, the problem of recognizing strings of characters, the problem of noise immune communication and others. The advantages of these codes (for example, compared with the difference-preserving codes [8]), are reviewed in [9,10,11]. For prooving of codes completeness in this case, the authors developed a topological interval model of rank configuration, an example of which is shown in Figure 3.



K_4= {{1,2,5}, {1,3,6}, {2,3,4},

а)

б)

Fig.3 - interval model of rank configuration

This model is a linear segment of length *2q*, which is divided into unit intervals, the boundaries of the intervals correspond to an integer in the range [0...2q]. Here q is the selected digit of code. On this segment enclosed the intervals that are scaled to correspond to a rank configuration (depicted by the arcs in Fig. 3b). For the model are valid such axioms:

Axioms of difference operations:

$$\underset{i,j\in M}{\forall} \quad (b_i \oplus b_j) = (b_j \oplus b_i) = \left|b_i \oplus b_j\right| = h_{ij}; \text{ - the axiom of symmetry,} \tag{26}$$

$$\underset{i,j\in M}{\forall} \quad (b_i \oplus b_j) = (b_j \oplus b_i) = \left|b_i \oplus b_j\right| = h_{ij}; \text{ - axiom of operation reversibility.} \tag{27}$$

The axioms of rank intervals:

$$h_{ij}(r_k) + h_{ij}(r_{k+1}) = h_{ij}(r_k + r_{k+1}) \text{ - axiom of intervals rigidity in the cycle,} \tag{28}$$

$$r_k + r_{k+1} \geq r_c \text{ - the triangle axiom for the ranks in the loop,} \tag{29}$$

where r_k and r_{k+1} - ranks of the k-th and k +1- th intervals on the model (the ranks of the neighboring intervals), r_c - rank of the closing cycle interval. The indicated ranks are depicted by the arcs in the topological graph model (Fig. 3b). The concept of cycle interval corresponds to the concept of the model cycle at the topological graph.

On the basis of the interval model is formulated and proved the following theorem.

Theorem 3: Valid are only those rank configurations for which the axioms (28) and (29) are fulfilled.

This theorem shows that the number of rank configurations provided the use of XOR operations is limited, that is not possible to construct full potential code. For example, the number of allowed configurations of rank for the 3-dimensional simplex is 7 out of 30 possible, and the coefficient of completeness of the code, is 7/30, approximately 23.3%.

To determine the of allowed rank configurations and construct the corresponding potential codes an algorithm was developed that is based on the simplex method and takes into account the expression (26) - (29), as well as algorithms for checking the adequacy of the proposed model.

4. CONCLUSIONS

The method of unitized representation of the systems knowledge about objects of an external world by rank transformation of their descriptions, made in the different features spaces - deterministic, probabilistic, fuzzy and other, and also models of rank configurations are resulted. The method of the rank configuration description by the DRP - code (distance rank preserving code) is developed. The problems of its completeness, information capacity, noise immunity and secrecy are reviewed.

This problem is solved on the basis of such fundamental concepts as the degree (rank) of the distance between the states of objects in the parameter space. It is shown that important information for the identification of states and the optimization of the system are not the actual distance between these states but their rank relationships. An effective with respect to computational cost binary encoding method, which determines both the objects themselves (state), as well as information on the ranks of the distances between them. Such a code is called potential or DRP-code (code that preservs the ranks of distances). The questions of its completeness, information capacity, noise immunity and secrecy were reviewed. We propose a interval model of rank configuration, which allows to decide whether the definition of completeness of the rank code for the case of using of additive operations for rank determining.

REFERENCES

1. S. Shteinberg, *Identification in Control Systems,* Energoatomizdat, Moscow, 1987. (in Russia)

2. J. Tou, R. Gonzalez, *Pattern Recognition Principle*, Addison-Wesley, London-Amsterdam-Don Mills, 1974.

3. B. Kosko, "Fuzzy entropy and conditioning", *Information sciences*, **2**, pp.165-174, 1966.

4. D. Dubois, H. Prade, "Twofold fuzzy sets: An approach to the representation of sets with fuzzy boundaries based on possibility necessity measures", *Fuzzy Mathematics* (Huazhong, China), **4,** pp. 53-76, 1983.

5. J. Kruskal, "Multidimensional scaling by optimizing goodness-of-fit to an nonmetric hypothesis", *Psychometrika*, **29**, pp. 1-28, 115- 129,1964.

6. P. Cameron, J. Van Lint, *Graph, Codes, Designs*, Cambridge University Press, London, 1980.

7. B. Zlotnik, *Noise immunity coding in communication systems,* Radio and Communications, Moscow, 1989. (in Russia)

8. F. Preparata, J. Nievergelt, "Difference-Preserving codes", *IEEE Trans. Information Theory*, **IT-20**, pp. 643-649, 1974.

9. N.Bykov, I. Kuzmin, O. Kobersky, "Comparable characterization of structural code efficiency in speech recognition systems", *Proceeding of the 5-th International Conf. On Checking and Control Complex Systems (CCCS-99)*, pp. 191-194, Vinnitsa, 1999. (in Russian)

10. N.Bykov, I. Kuzmin, A. Iakovenko, Data clastering with using of potential codes, *Вісник Вінницького політехнічного інституту.* – 2001. - №6. - P. 61-64. (In Ukrainian)

11. N.Bykov, The unified method of information representation in evolutionary intelligence systems, *Відбір і обробка інформації*, 2006.- Вип. 24(100). P. 35-42. (In Ukrainian)

Received: March, 2011