Sparse Binary Distributed Encoding of Numeric Vectors

Article in Journal of Automation and Information Sciences · January 2005

DOI: 10.1615/JAutomatInfScien.v37.i11.60

5 authors, including:

Dmitri A. Rachkovskij
National Academy of Sciences of Ukraine

Serge V. Slipchenko
International Research and Training Center for Information Technologies and Syste…

Ernst Kussul
Universidad Nacional Autónoma de México

Please cite as:

Dmitri A. Rachkovskij, Sergey V. Slipchenko, Ivan S. Misuno, Ernst M. Kussul, Tatyana N. Baidyk. Sparse binary distributed encoding of numeric vectors. Journal of Automation and Information Sciences. 2005. Vol. 37, issue 11. pp. 47–61.

Abstract: Hypervectors (then known as codevectors) of vectors are considered.

Errata

Due to the translation of the original publication in Russian, which was done without consulting the authors, the translated article has several terminological errors:

–"unities", "unity" are met throughout the paper.

It means vector components with the value of 1.

– "rarefaction" means "sparsity".

– "flocet" means "float".

– "by module" means "modulo".

– "average of distribution", "AD" stand for "mean".

– "deterioration" means "distortion".

– "transposition" means "permutation".

– In Figure 7, all $\wedge$ above the plots should be read as $\vee$.

Sparse Binary Distributed Encoding of Numeric Vectors

D.A. RACHKOVSKIY, S.V. SLIPCHENKO, I.S. MISUNO,
E.M. KUSSUL, T.N. BAIDYK

Procedures of stochastic distributed binary sparse encoding of numeric vectors are considered. Some problems of such encoding and variants of their overcoming, including the way of binding by context-dependent thinning, are discussed. Characteristics of code overlapping for different dimensions of input space and encoding parameters are presented.

Key words: information processing, encoding, information capacity, distributed representation, thermometric encoding, flocet encoding, stochastic encoding.

Introduction

In distributed representation, information is encoded by multidimensional vectors, where individual elements, as usual, have no unique interpretation. Advantages of distributed representations are: efficient resource use [1–3]; simplicity of representation and estimation of degree of similarity (by scalar product); capability of working under conditions of noise, failures and indeterminacy; parallelism, neurobiological relevancy etc. [4].

There exist several approaches to representation of numeric information by binary codes, reflecting metric of input space of real vectors. For instance, the Gray code has no overlapping jumps, however only for neighboring numbers. Codes of thermometric and flocet encoding are not pseudorandom and possess linear information capacity, not exponential, like distributed codes [5]. Later we demonstrate, that generalization of these schemes for multidimensional case does not ensure binding information and provides no opportunity of linear solution of complex classification problems.

An attempt to overcome the mentioned problems by binding random binary codevectors without increasing their dimension are sparse codes [6, 7]. Sparse codes have exponential capacity and have no jumps of overlapping characteristics. For such codes, a single variant of encoding procedure of numeric vectors and binding by bitwise operation XOR was suggested [6]. However, fraction of unity bits in sparse codes is large, which reduces computational efficiency of operations, including use of associative memory [8]. Another adjoint class of stochastic encoding schemes, so called "rough" encoding [9], requires laborious calculation of distances from the input vector to the center or boundary of a large number $N >> A$ receptive fields. In approaches [10, 11], no proper attention to study of binary sparse codes (with small fraction of unities) and to problems of their binding, was paid.

Within the framework of paradigm of associative-projective neural networks (APNN) [1, 12], approaches to encoding of different information by sparse codevectors, binary vectors with small fraction of pseudorandomly located unities, are developed. One of directions was related to encoding of numeric information. Schemes of encoding of scalars were considered in [5]. In this paper we systematize and develop the approach to encoding of vector quantities, mainly based on bitwise operations on codevectors [12–15].

1. Characteristics and properties of codes

Consider encoding schemes, which perform transformation $\boldsymbol{\varphi}$ of input (signal) space $(0,1)^A$ into the space of codevectors B^N. Elements $x_j \in [0,1]$ of a real vector $\mathbf{x} = (x_1, x_2, \ldots, x_A)$ can be transformed into integer ones by the rule $q_i = \text{ent}(x_i Q)$. Usually, A and Q have order of several dozens.

The secondary space B^N is formed by multidimensional $(N >> 1)$, binary ({0, 1}), pseudorandom (with random, but the same placement of unities for identical encoded data) and sparse (number of unities is $M << N$) codevectors with large $(M >> 1)$ and approximately equal $(M \approx \text{const})$ number of unities. Typical values of N are hundred thousands, M are thousands. Such codevectors ensure statistical stability, high resolution, noise immunity and possibility of work with random subsets of code with preservation of properties of the whole code [2, 3]. Sparse binary vectors allow efficient implementation of vector and vector-matrix operations. Thanks to rarefaction, a higher informative capacity of associative memory, which may be used for storing and searching codevectors [1, 8], is ensured. Besides, such representations are neurobiologically relevant [16].

There exist the following methods of representing by means of such codevectors different types of data: scalars [5], numerical vectors (are studied in the present paper) and data with a more complicated structure [1]. Availability of such methods promotes development of a universal and unified approach to processing inhomogeneous information and creation of a series of new intelligent information technologies. Such technologies support both classification of vector information and processing complex structured information for systems of artificial intelligence (AI).

An important feature of the suggested methods of codevectors formation is as follows. Proximity of numeric quantities in the input space is transformed and reflected in similarity of codevectors. Transformation of metrics of input space can have nonlinear character and can be useful for solution of some problems of classification, approximation, control etc. with the use of methods and algorithms based on estimation of similarity through the scalar product [17].

We estimate degree of similarity of binary codevectors **X, Y** with elements from {0, 1} by their relative overlapping (fraction of coinciding unities):

$$V(X, Y) = \frac{|\mathbf{X} \wedge \mathbf{Y}|}{|\mathbf{X}|}, \tag{1}$$

where $\wedge$ is the elementwise conjunction, $|\mathbf{X}|$ is the number of unities in **X**, $V \in (p, 1)$. Besides, in the case of independent codevectors, by which, usually, dissimilar objects (for instance, distant numeric values) are encoded, $V = p$; and in the case of identical codevectors (for example, those encoding equal numeric values) $V = 1$; interim values of codevectors overlapping characterize objects with interim values of similarity.

Let us consider a series of encoding schemes (transformations $\boldsymbol{\varphi}$), possessing the property, that when encoded points are moving away from each other, overlapping of their codevectors decreases. The form of codes overlapping is reflected by the so called characteristic of overlapping, dependence of value of codes overlapping on coordinates of encoded points or some measure of their proximity. In further reasonings we shall consider average of distribution (AD) of codes overlapping; value of overlapping of actual codes approximates AD with increasing N.

Some of considered encoding schemes form code of the whole vector $\mathbf{x}$ on the basis of codes of its elements (scalars) $x_i \in R$ [5]. Properties of the resulting codevector are determined both by properties of codes of elements and by the technique of its forming from codes of elements.

Denote codevectors of input vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} = \mathbf{y} + \boldsymbol{\Delta} (\mathbf{x}, \mathbf{y}, \boldsymbol{\Delta} \in \mathbf{R}^A)$ by $\mathbf{X} = \boldsymbol{\varphi}(\mathbf{x}), \mathbf{Y} = \boldsymbol{\varphi}(\mathbf{y}), \mathbf{Z} = \boldsymbol{\varphi}(\mathbf{z} = \mathbf{y} + \boldsymbol{\Delta})$ and codevectors of their elements x_i, y_i, z_i by $\mathbf{X}_i = \boldsymbol{\varphi}(x_i), \mathbf{Y}_i = \boldsymbol{\varphi}(y_i), \mathbf{Z}_i = \boldsymbol{\varphi}(z_i = y_i + \boldsymbol{\Delta}_i)$. Notation $V(\mathbf{X}, \mathbf{Y})$ further will be used for AD of overlapping. Let us formalize some properties of overlapping of codes of such schemes.

For any points **x** and **y** characteristic of codes overlapping should be nonincreasing function under nondecreasing difference of their coordinates by each measurement:

$$\forall \mathbf{x}, \mathbf{y}, \boldsymbol{\Delta} : \text{if } |z_i - x_i| \geq |y_i - x_i| \; i = \overline{1, A}, \text{ then } V(\mathbf{X}, \mathbf{Z}) \leq V(\mathbf{X}, \mathbf{Y}). \tag{2}$$

Interesting are encoding schemes where overlapping is a decreasing function. Under increasing difference of coordinates by one or several dimensions and nondecreasing by remaining dimensions, codes overlapping of corresponding coordinates and respectively, overlapping of codes of the whole vector, decrease. For a stepwise decreasing function, equality of codes overlapping is allowed, however only for small variation of coordinates. The length of stage depends on encoding scheme (form of transformation $\boldsymbol{\varphi}$) and parameters of scheme. Parameters include codes resolution (the number of grades of codes overlapping is limited), the step of discretization of input space (actual accuracy of encoded values representation is finite, see also the end of this section) etc. If variation $\boldsymbol{\Delta}_i$ of coordinate i exceeds the length of the corresponding stage $d_i \in R$, then the codevector of this coordinate changes (see [5] and Section 2.2).

For sparse codes, significant distance from **y** by at least one measurement with a high probability results in change of the code of the whole vector:

$$\exists i^*, d_{i^*} > 0 \;\; \forall \mathbf{y}, i, \boldsymbol{\Delta}, \; |\Delta_{i^*}| \geq d_{i^*} : V(\mathbf{Y}, \mathbf{Z}) < 1. \tag{3}$$

Overlapping of **X** and **Z** is smaller than overlapping of **X** and **Y** of codevectors **x** and **y**, if at least one coordinate z_i is significantly farther from x_i, than y_i, and remaining z_i are not closer to x_i, than y_i:

$$\exists i^*, d_{i^*} > 0 \;\; \forall \mathbf{x}, \mathbf{y}, \boldsymbol{\Delta}, \;\; \mathrm{sgn}(\Delta_i) = \mathrm{sgn}(y_i - x_i), \; |\Delta_{i^*}| \geq d_{i^*} : V(\mathbf{X}, \mathbf{Z}) < V(\mathbf{X}, \mathbf{Y}). \tag{4}$$

For change of codes overlapping, the change of at least one coordinate of point is necessary:

$$\begin{gathered}\forall \mathbf{x}, \mathbf{y}, \boldsymbol{\Delta} : \text{ if } V(\mathbf{X}, \mathbf{Z}) < V(\mathbf{X}, \mathbf{Y}), \text{ then } \exists i : |z_i - y_i| > |y_i - x_i|, \\ \forall \mathbf{x}, \mathbf{y}, \boldsymbol{\Delta} : \text{ if } V(\mathbf{X}, \mathbf{Z}) > V(\mathbf{X}, \mathbf{Y}), \text{ then } \exists i : |z_i - y_i| < |y_i - x_i|.\end{gathered} \tag{5}$$

Some encoding schemes have a bounded radius of the decreasing portion of the overlap function ("effective" overlapping radius). In a general case, $\mathbf{r}(\boldsymbol{\varphi}) \in \mathbf{R}^A$ and for anisotropic schemes $r_i = r, i = \overline{1, A}$. For such schemes the following is also true. If **z** by at least one coordinate is significantly farther from **x**, than **y** from **x**, however the both points **y** and **z** are within the radius of overlapping, then overlapping codes are different:

$$\exists i^*, d_{i^*} > 0 \;\; \forall x, y, |y_{i^*} - x_{i^*}| < |z_{i^*} - x_{i^*}| < r, \; |\Delta_{i^*}| > d_{i^*} : V(\mathbf{X}, \mathbf{Z}) < V(\mathbf{X}, \mathbf{Y}). \tag{6}$$

This approach to stochastic encoding is close to ideology of so called randomized embeddings and sketches [10, 11]. However, in those approaches not only dependence on distances or other measures of proximity in the input space of AD of codes overlappings, but also probabilities of deviations (deteriorations) are analyzed. Some studied by us codes can be randomized embeddings for the original space $\mathbf{R}^A$ with Minkowsky metric D_d, if we choose distance f as some function of codes **X**, **Y** and their overlappings V. For them, with some probability the following inequality holds

$$D_d / k < f(V(\mathbf{X}, \mathbf{Y}), \mathbf{X}, \mathbf{Y}) < kD_d, \tag{7}$$

where k is the coefficient of deterioration, which depends on peculiarities of specific encoding schemes $\boldsymbol{\varphi}$, dimension of input space, characteristics of codevectors etc. For schemes, well approximating D_d, value of k is close to unity. In this paper we consider characteristics of AD $f(V(\mathbf{X}, \mathbf{Y}), \mathbf{X}, \mathbf{Y})$ without analysis of conditions and probability of fulfillment of (7).

2. Encoding vectors of numeric quantities

Consider approaches, methods and procedures of distributed encoding of vector numeric quantities, which ensure properties of codes, explained in the previous section. Encoding consists in transformation of an integer-valued vector **x** into a binary codevector **X** by constructing codevector of the latter or extracting from memory of the beforehand generated codevector. In many schemes, constructing codevector for **x** makes use of codevectors of its elements (scalars).

2.1. Encoding elements of vector, scalar quantities

Schemes of binary encoding of scalar numeric quantities are considered in papers [5, 18]. These are partially distributed schemes (thermometric and flocet encoding) as well as the developed by us schemes of distributed stochastic encoding. The latter give codevectors with approximately equal low density $p = M / N << 1$ and pseudorandom placement of unity elements. Strongly intersecting codes correspond to close numbers, to distant ones — weakly intersecting codes.

In Figure 1 we show examples of experimental (V, U) and theoretical (V_T, U_T) characteristics of codes overlapping of scalar quantities (with thinning-augmenting (*a*) and concatenation of parts of intersecting reference vectors (*b*)), obtained by subtractive-additive procedure of encoding and concatenation of parts of reference codevectors (methods of regulating form and steepness of overlapping characteristic and also encoding of cyclic quantities are considered in detail in [5]). We show dependence of characteristic of relative overlapping V on the difference of encoded quantities. The shown characteristic is obtained for codevectors of dimension $N = 10000$ with mean $M = 100$ and $Q = 20$ grades. It is seen, that the magnitude of codevectors overlapping decreases with increasing distance (difference) of encoded quantities.

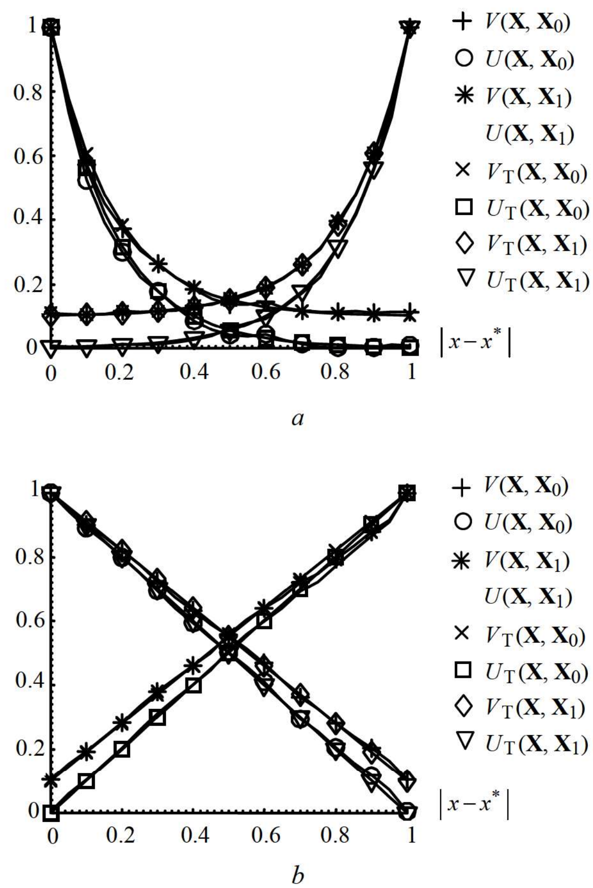


Figure 1

2.2. Union of codes of vector elements by concatenation and disjunction

Let A-dimensional vector $\mathbf{x}$ be at input, where each of A elements has Q grades. The output binary N-dimensional sparse random codevector $\mathbf{X}$ should satisfy requirements, listed in Section 1. In encoding input vectors we apply concatenation of codevectors of their elements [18]. In the result, dimension of the resulting codevector increases linearly with the number of elements of input vector $N^* = AN$, where N^* is dimension of pool, encoding one element of input vector.

Analytic form of AD of overlapping for concatenation of codes of elements is:

$$V(\mathbf{X},\mathbf{Y}) = \frac{1}{A}\sum_{i=1,A} V(\mathbf{X}_i,\mathbf{Y}_i). \tag{8}$$

If for each of coordinates, codes overlapping does not increase with increasing difference of coordinates, then under change of at least one coordinate overlapping of codes of vectors does not increase (conditions (2)–(6) are fulfilled). Specific form of dependence of overlapping decrease on distance is determined by the characteristic of overlapping decrease with respect to each coordinate.

Consider the form of overlapping characteristic for encoding procedure of elements of the input vector by concatenation of parts of reference vectors, neglecting discreteness of encoded quantities and deviation of overlapping from AD:

$$V(\mathbf{X},\mathbf{Y}) = \frac{1}{A}\sum_{i=1,\,A}(p+(1-p)(1-|x_i-y_i|)) = \frac{1}{A}\sum_{i=1,\,A}(1-(1-p)|x_i-y_i|) =$$

$$= 1-\frac{(1-p)}{A}\sum_{i=1,\,A}|x_i-y_i| = 1-\frac{(1-p)D_1(x,y)}{A}, \tag{9}$$

where D_1 is the Manhattan distance. Since $D_1(x,y) = A(1-V)/(1-p)$, putting $f(V(\mathbf{X},\mathbf{Y}),\mathbf{X},\mathbf{Y}) \equiv$ $\equiv A(1-V)/(1-p)$ for $\mathbf{X}$, $\mathbf{Y}$, which belong to the image of the studied mapping, we obtain f, for which properties of metric D_1 are fulfilled. If we use a steeper characteristic of codes overlapping with respect to each coordinate [5], codes overlapping becomes constant even within the interval of encoded quantities [0, 1], outside bounds of effective radius of overlapping r between encoded points (Section 1).

Consider the form of overlapping characteristic for subtractive-additive procedure:

$$V(\mathbf{X},\mathbf{Y}) = \frac{1}{A}\sum_{i=1,\,A}\left(p+(1-p)\left(\frac{(1-q)(1-p)}{1-p(1-q)}\right)^{|x_i-y_i|}\right). \tag{10}$$

Monotonous decreasing of V under increasing difference of coordinates of points conditions fulfillment of conditions (2)–(6) for overlapping of codes from Section 1.

For encoding elements of the input vector by sparse random vectors, union by not concatenation, but by bitwise disjunction (Figure 2) is possible. Such union of small number of codes, corresponding to elements of input vector, does not result in significant mutual influence due to their negligible intersection. To advantages also belong independence (relative) of codevector dimension on dimension of the input vector and natural possibility of representation of similar elements by correlated codevectors.

A recursive expression for calculation of value of overlapping in encoding codevectors of elements by disjunction with density p has the form:

$$V_{k+1} = P_{k+1}/Q_{k+1};$$

$$P_{k+1} = P_k + p_{k+1} - P_k p_{k+1} + 2(p-p_{k+1})(Q_k - P_k); \tag{11}$$

$$Q_{k+1} = 1-(1-p)^{k+1},$$

where P_k is the density of codes overlapping after disjunction of $k = 0,\ldots,A-1$ pairs of elements codevector; $p_k = pV_k$, $p_k \in [p^2, p]$, is the density of overlapping of codevectors of the k-th pair of vector elements; Q_k is the density of codevector of disjunction of k elements codevectors. We can demonstrate, that for $p \ll 1$,

$$P_A = 1-2\exp(-pA)+\exp\left(-\sum_{i=1,A}(2p-p_i)\right),$$

$$V = \frac{P_A}{1-(1-p)^A} \approx \frac{P_A}{1-\exp(-pA)}. \tag{12}$$

After estimation of approximation error (12), we can demonstrate, that conditions (2)–(6) are fulfilled.

Consider the case of encoding the elements by concatenation of parts of reference vectors. Therewith the result, the codevector of each element, should be subjected to independent transposition in order to eliminate unwanted correlation of resulting codevectors. Overlapping of codevectors of each element is determined by formula $V(\mathbf{X}_i, \mathbf{Y}_i) = 1 - (1-p)\left| x_i - y_i \right|$ [5]. Therefore

$$P_A = 1 - 2\exp(-pA) + \exp\left(-\sum_{i=1,\, A} (2p - p_i(\mathbf{X}_i, \mathbf{Y}_i)) \right) =$$

$$= 1 - 2\exp(-pA) + \exp(-pA)\exp(-p(1-p)\, D_1(\mathbf{x}, \mathbf{y})); \qquad (13)$$

$$V = \frac{P_A}{1 - \exp(-pA)}.$$

Thus, for this encoding scheme, function

$$f(V) = D_1(\mathbf{x}, \mathbf{y}) = \frac{\ln(\exp(-pA)(2 - E\{V(\mathbf{X}, \mathbf{Y})\}) + V(\mathbf{X}, \mathbf{Y}) - 1) + pA}{p(p-1)} \qquad (14)$$

corresponds to metric D_1 for images of $\mathbf{X}$, $\mathbf{Y}$ encoded points of input space $\mathbf{x}$ and $\mathbf{y}$.

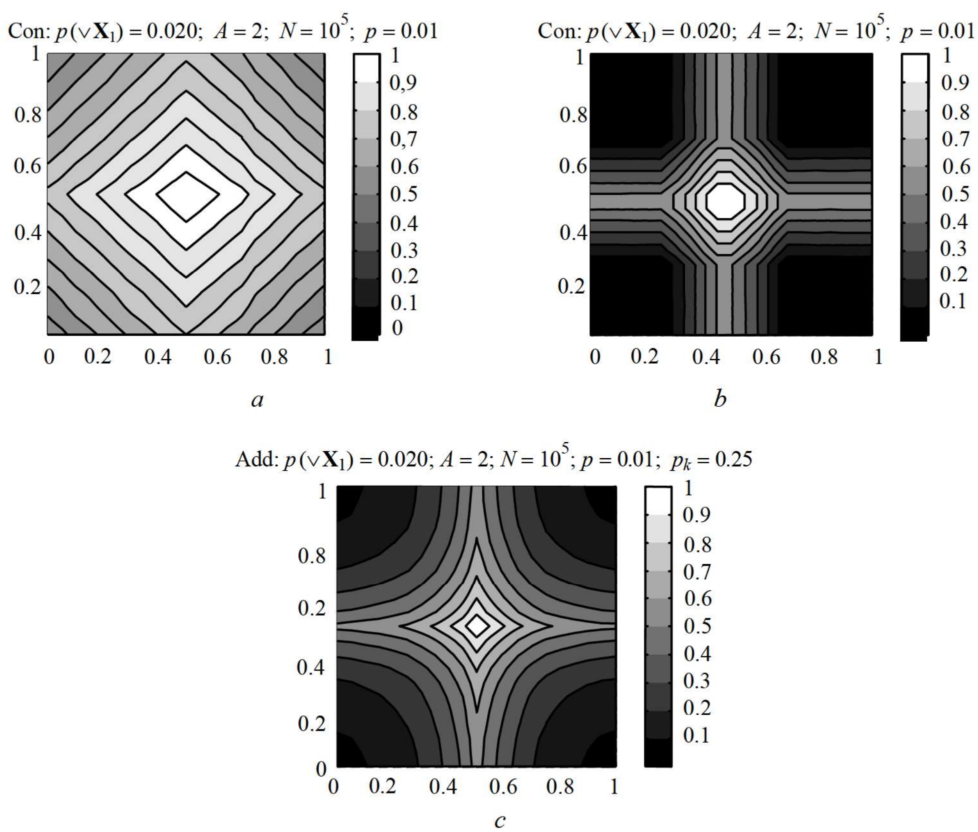


Figure 2

In Figure 2 we show characteristics of overlapping of disjunction of elements codevectors for concatenation of parts of reference vectors (a), concatenation of parts of reference vectors with bounded radius (b) and for subtractive-additive procedure (c). Pictures differ in different ways of encoding of input vector element and by different steepness of overlapping of element codes. Three-dimensional form of these characteristics is shown in Figure 3.

Here and later overlapping is calculated by formula (1). Magnitude of overlapping is reflected by shades of gray. The given characteristic is obtained for codevectors of dimension $N = 100000$ with mean number of unities $M = pN = 1000$ and $Q = 20$ grades by each dimension.

We see, that along coordinate axes magnitude of overlapping decreases slower, than along other directions. Besides, at some distance from the starting point it ceases to decrease and remains constant when distance increases. Such character of codes overlapping is called shade [6]. Similar effects exist also for union of codevectors by concatenation, and also for encoding more complex, than vector, types of data [1]. For reducing shades and control (modification) overlapping characteristics we apply approaches, described in the subsequent sections. This is the direct constructing the codes with isotropic characteristic of overlapping for representation of input vectors and codes binding.

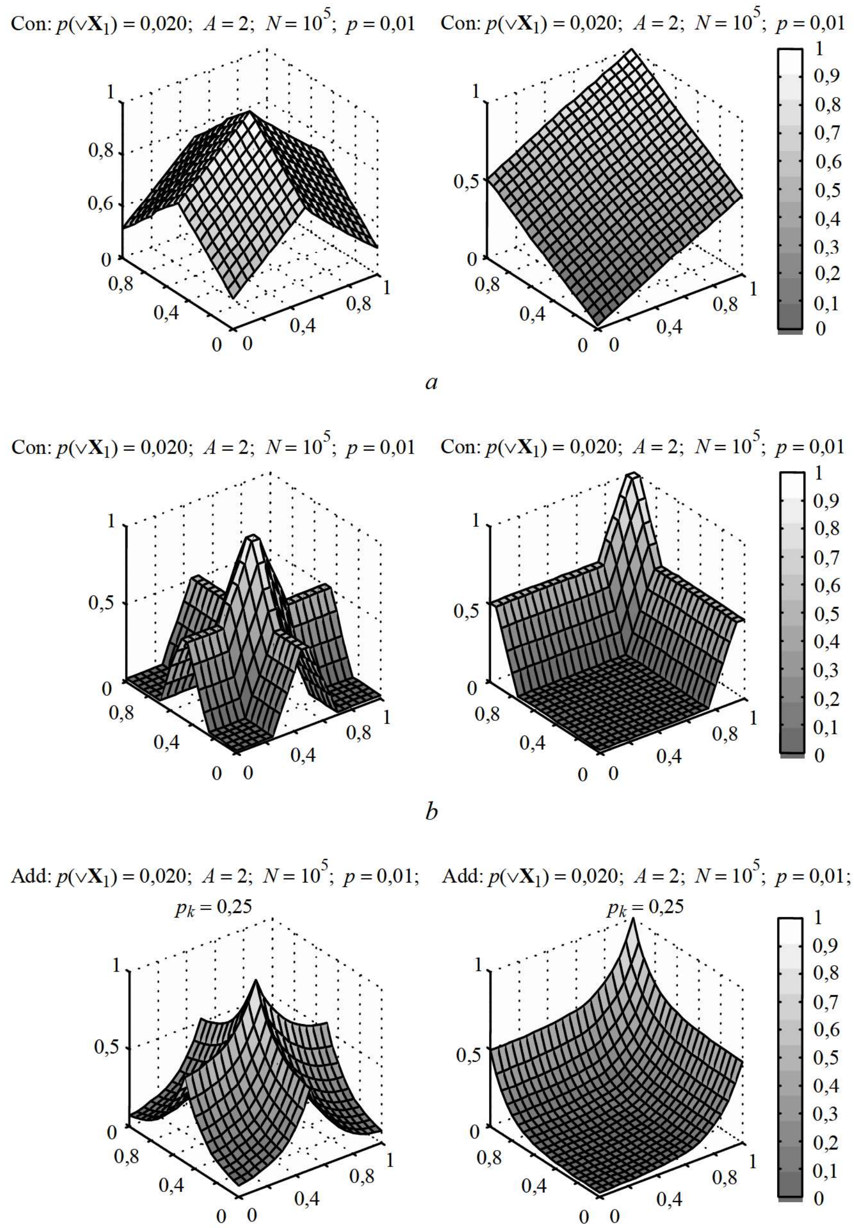


Figure 3

3. Constructing codes with isotropic overlapping characteristics

Let us consider procedures, which form for vectors of input space codevectors, that ensure isotropic overlapping characteristics without shades. In this section we consider a variant of such procedure, based on preliminary constructing the codevectors for all points of input space. Encoding is performed by reference to the corresponding code by values of elements of the input vector $x_1, x_2, \ldots, x_A$, used as indices of array of codevectors code $[x_1][x_2]\ldots[x_A]$.

For multidimensional cyclic codes, procedure of constructing the codevectors with isotropic overlapping characteristic is based on the idea of "reproduction" (for one-dimensional case this idea was considered in [5]). Some $\prod_{i=1,\,A} Q_i$ random binary reference vectors $\mathbf{Y}(q_1, q_2, \ldots, q_A)$ with small number of unities are generated. Here Q_i is the number of grades in dimension i. In Figure 4 we present multidimensional cyclic encoding $(A = 2)$; black circle denotes the encoded point $x = (4, 5)$ for $n = (6, 6)$; solid circles are codevectors, entering disjunction of the code for the encoded point; dashed circles are codevectors, which do not participate in disjunction. The resulting codevector **X**, encoding the input vector $(q_1, q_2, \ldots, q_A)$, is obtained by disjunction of $\prod_{i=1,\,A} n_i$ codevectors **Y** with numbers, preceding to the encoded number $(q_1, q_2, \ldots, q_A)$:

$$\mathbf{X}(q_1, q_2, \ldots, q_A) = \bigcup Y(j_1 \bmod Q_1, j_2 \bmod Q_2, \ldots, j_A \bmod Q_A),$$
$$j_1 = q_1, (q_1 + n_1 - Q_1),\ \ j_2 = q_2, (q_2 + n_2 - Q_2), \ldots,\ \ j_A = q_A, (q_A + n_A - Q_A). \tag{15}$$

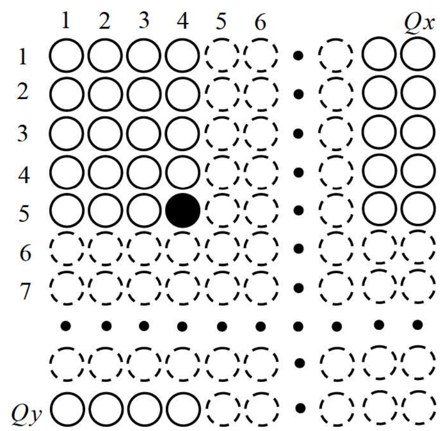
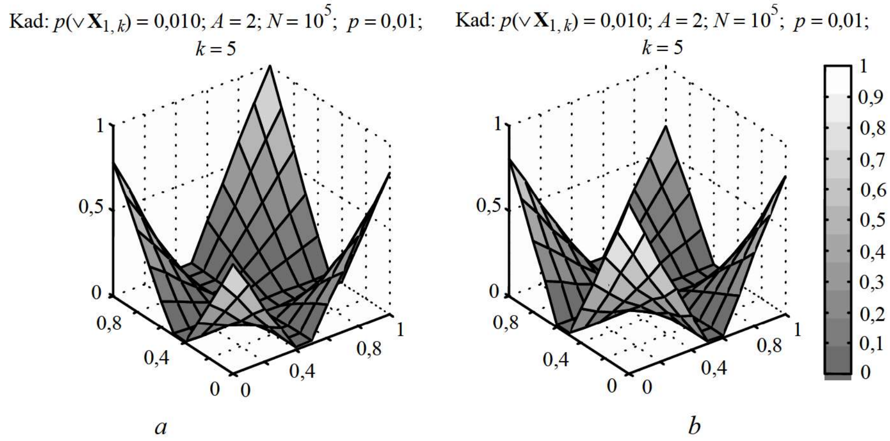


Figure 4

Cyclic encoding assumes, that the numbers of codes entering codes disjunction are taken by module Q_i. Parameters n_i determine steepness of decay of the codes overlapping characteristic. For isotropic overlapping characteristic, the number n is the same for all measurements. An example of the codevectors overlapping characteristic of "base" point (0.5; 0.5) with other space points for two-dimensional case is shown in Figure 5 (a is the base point (0; 0), b is the base point (0.5; 0.5)). It was obtained for codevectors of dimension $N = 100000$ with the mean number of unities $M = 1000$ and $Q = 10$ by each dimension. Characteristic has a similar form for another location of base point.

There are possible variants of procedure for generation of codevectors with noncyclical overlapping characteristic and also combined variants. Overlapping characteristic for isotropic variant of this procedure sufficiently well approximates Manhattan metric of the primary space (7) near the base point. For more remote portions, conditions (2)–(6) hold.

However, practical application of this procedure faces some difficulties. Disjunction of two sparse binary vectors requires $O(pN)$ operations. The total number of disjunctions is $\prod_{i=1,\,A} n_i$, i.e., $O(n^A)$. Then, encoding by generation of reference codevectors without storing generated codevectors requires $O(pNn^A)$ operations for each encoded vector. Analogously, amount of memory for storing preliminary generated codevectors also increases exponentially $O(NQ^A)$.



a *b*

Figure 5

4. Constructing codes of vector quantities using binding procedures

Binding provides an opportunity to solve the problem of storing in the codevector of information, in what combinations the elements of data structure were encountered. This is necessary for representation and processing by means of distributed representations of both simple vector representations, studied in this paper, and of more complex embedded structures [1]. In encoding vectors by union of codevectors of their elements via disjunction and concatenation, the binding procedures enable one to lose problems mentioned in subsection 2.2, and also problems of linear separability of codes (like problem XOR). Detailed discussion of binding procedures for random binary codevectors by context-dependent thinning (CDT) is given in [19].

4.1. Binding by conjunction

For binding by conjunction, obtaining of bound $\langle Z \rangle$ is performed as

$$\langle \mathbf{Z} \rangle = \bigcap_i \mathbf{X}_i, \tag{16}$$

where $i = \overline{1, A}$, A is the number of bound codevectors.

4.2. Binding by context-dependent thinning

Let us take bitwise disjunction S (S is small, usually 2–5) of pseudo-random input codevectors, which should be bound:

$$\mathbf{Z} = \bigcup_{i=1,\,S} \mathbf{X}_i. \tag{17}$$

Form thinned codevector $\mathbf{Z}$:

$$\langle \mathbf{Z} \rangle = \bigcup_{k=1,\,K} (\mathbf{Z} \wedge \mathbf{Z}_k^*) = \mathbf{Z} \wedge \bigcup_{k=1,\,K} \mathbf{Z}_k^*. \tag{18}$$

Here $\mathbf{Z}_k^*$ is $\mathbf{Z}$ with transposed elements. Each k-th transposition should be fixed, unique and independent. Random transpositions would be an ideal variant, however transpositions by cyclic shift with random number of shifts are sufficiently convenient in applications as well. The number K of disjunctively superimposed vectors with transposed elements is taken in such a way, that the number of unities in $\langle \mathbf{Z} \rangle$ would become close to the given value (as usual, of order of the number of unities in input codevectors $\mathbf{X}_i$). Density of the thinned vector $\mathbf{Z}$ is

$$p(\langle \mathbf{Z} \rangle) = p(\mathbf{Z}) p(\bigcup_{k=1,\,K} \mathbf{Z}_k^*) = p(\mathbf{Z})(1 - (1 - p(\mathbf{Z}))^K), \tag{19}$$

from where

$$K = \operatorname{ceil} \log \frac{1 - p(\langle \mathbf{Z} \rangle)}{p(\mathbf{Z})} / \log(1 - p(\mathbf{Z})). \tag{20}$$

If $\mathbf{Z}$ is disjunction of S independent codevectors with density p, then $p(\mathbf{Z}) = 1 - (1 - p)^S$.

4.3. Encoding vectors by binding by conjunction of codevectors of their elements

Consider encoding the vectors by binding by conjunction of codevectors of their elements. Density of the resulting vector is equal to p^A. Density of overlapping of two bound codevectors changes from p^A for identical codevectors to $(p^A)^2$ for independent ones. This corresponds to overlapping $V \in [p^A, 1]$. Analytic characteristic of overlapping for binding by disjunction has the following form:

$$V(\mathbf{X}, \mathbf{Y}) = \prod_{i=1,\,A} V_i(\mathbf{X}_i, \mathbf{Y}_i). \tag{21}$$

The overlapping characteristic of conjunction of codevectors of elements for concatenation of parts of reference vectors (a), concatenation of parts of reference vectors with bound radius (b) and subtractive-additive procedure (c) for two-dimensional case are represented in Figure 6. Characteristics has a similar form for other location of base point. Comparing Figure 6 with Figure 2 we see, that after binding the overlapping characteristic becomes more steep, and its "shelf" (random overlapping) becomes more low. Computational complexity of procedure $O(pNA)$ linearly depends on A, unlike exponential growth for procedures of Section 3. Memory expenditure for storing original codevectors is $O(AQpN)$.

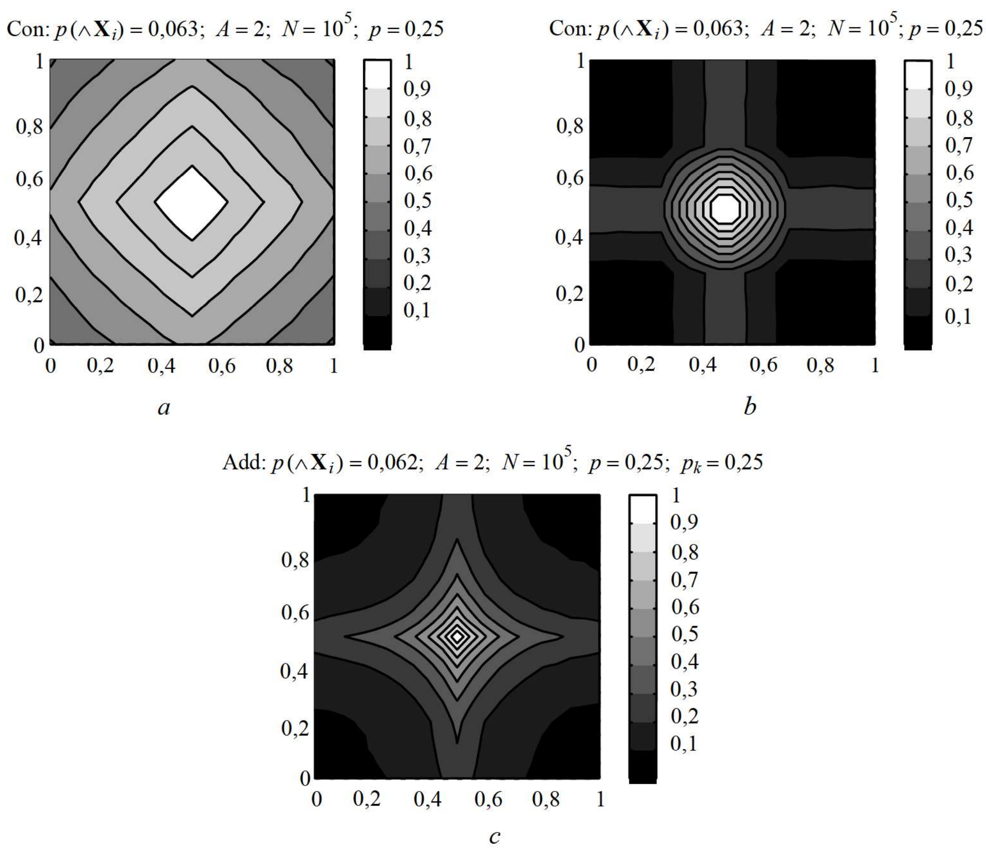


Figure 6

Examining "receptive fields" of individual bits of the resulting codevector (i.e., the domain of input space, where they have value 1) demonstrates, that vast majority of them has compact (hyper)rectangular shape. This enables one to solve, using such codes, classification problems with complex linearly inseparable domains of classes (see [9, 20], and also [13–15]).

4.4. Encoding vectors by binding codevectors of their elements by context-dependent thinning

For obtaining analytic characteristic of overlapping of such encoding, consider the vector of overlapping of thinned $\langle \mathbf{Z}1\rangle$ and $\langle \mathbf{Z}2\rangle$:

$$\langle \mathbf{Z}1\rangle \wedge \langle \mathbf{Z}2\rangle = \mathbf{Z}1 \wedge \bigcup_{k=1,\,K} \mathbf{Z}1_k^* \wedge \mathbf{Z}2 \wedge \bigcup_{k=1,\,K} \mathbf{Z}2_k^*.$$

Taking into account, that vectors $\mathbf{Z}1$ and $\mathbf{Z}2$ are independent of their transpositions, we obtain

$$p(\langle \mathbf{Z}1\rangle \wedge \langle \mathbf{Z}2\rangle) = p(\mathbf{Z}1 \wedge \mathbf{Z}2)\,p^*(\mathbf{Z}1, \mathbf{Z}2),$$

where $p^*(\mathbf{Z}1, \mathbf{Z}2) = p(\bigcup_{k=1,\,k} \mathbf{Z}1_k^* \wedge \bigcup_{k=1,\,K} \mathbf{Z}2_k^*)$ is calculated by formula (12), and all K components are independent and have overlapping density $p(\mathbf{Z}1 \wedge \mathbf{Z}2)$. Since transposition does not change codevector density, $p^*(\mathbf{Z}1, \mathbf{Z}2)$ varies from $p(\mathbf{Z}1 \wedge \mathbf{Z}2)$ $(K=1)$ to 1 $(K>>1)$. Thus

$$p(\mathbf{Z}1 \wedge \mathbf{Z}2)^2 \le p(\langle \mathbf{Z}1\rangle \wedge \langle \mathbf{Z}2\rangle) \le p(\mathbf{Z}1 \wedge \mathbf{Z}2). \tag{22}$$

Magnitude of overlapping of $\langle \mathbf{Z}1\rangle$ and $\langle \mathbf{Z}2\rangle$ is well approximated by formula

$$V(\langle \mathbf{Z}1\rangle \wedge \langle \mathbf{Z}2\rangle) = V(\mathbf{Z}1 \wedge \mathbf{Z}2)^{2-p(\langle \mathbf{Z}\rangle)/p(\mathbf{Z})}. \tag{23}$$

Thus, context-dependent thinning transforms overlapping as

$$V(\langle \mathbf{Z}1\rangle \wedge \langle \mathbf{Z}2\rangle) = V(\mathbf{Z}1 \wedge \mathbf{Z}2)^b,\ \ b \in [1, 2]. \tag{24}$$

This should be taken into account for modification of $f(V)$ for thinned codevectors $f(V) \to f'(V) = f(V^{1/b})$. Therewith, due to the increasing character of $f(V)$ we assume fulfillment of conditions (2)–(6).

Computational complexity of this form of encoding is $O(pNA(K+1))$, where complexity of disjunction of element codevectors is $O(pNA)$ and complexity of transpositions and disjunctions of CDT is $O(pNAK)$. Thus, complexity of encoding linearly depends on A, as for procedure of section 4.3, not exponentially, as in Section 3. Memory expenditures for storing original codevectors is $O(pNQA)$.

Characteristics of overlapping for two-dimensional case under different depth of thinning are shown in Figure 7 (with density of the resulting vector, equal to p (*a*) and $0.2p$ (*b*)). From comparison of Figure 7, *a* with Figure 7, *b* and Figure 6 we see, that increasing thinning depth (reducing density of the resulting codevector) results in a steeper characteristic of overlapping and a lower "shelf" (decreasing of shade effect). For deep thinning $(0, 2p)$, in characteristic of overlapping we clearly observe result of statistical instability of the number of unities in overlapping, since such thinning approaches thinning by conjunction (or reaches it). The character of receptive fields (see subsection 4.3) becomes more local compared to receptive fields of codevectors, obtained by disjunction (subsection 2.2), which also gives possibility of their application in solving complex problems of classification.

It is interesting to note, that the form of studied characteristics of overlapping is close to D_d with $d<1$. For $d<1$, Minkowsky distance D_d ceases to be a metric (due to breaking the triangle inequality), however such measure of proximity can adequately reflect similarity of data for some problems and subject fields.

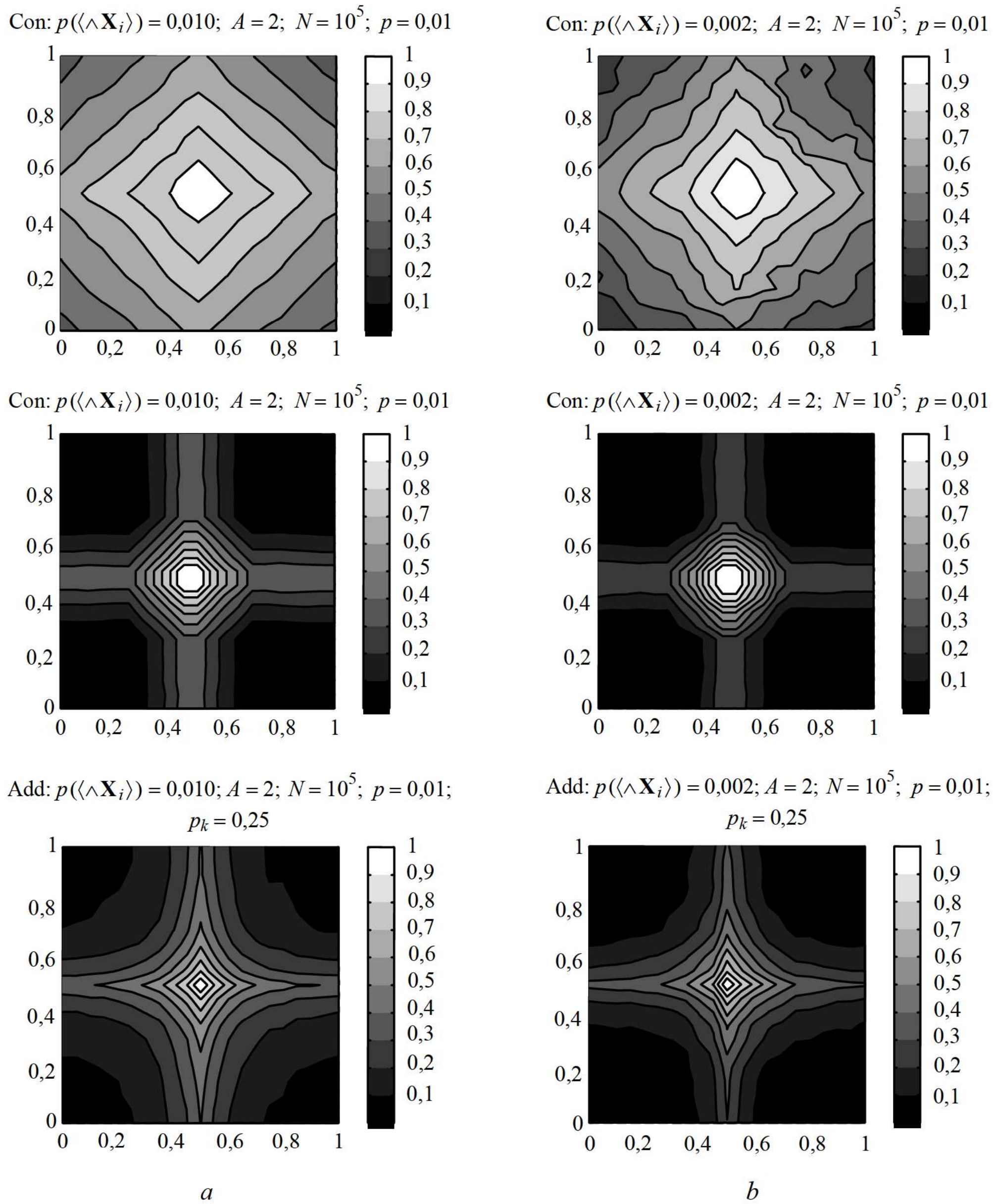


Figure 7

In Figure 8 we show overlapping characteristics for disjunction, conjunction and binding by context-dependent thinning of codevectors, obtained by concatenation of parts of reference vectors (*a*) and subtractive-additive procedure (*b*) for multidimensional case for different depths of thinning for input spaces of dimensions $A = 2$, 4, 16. Here V_1 is the dependence on change of one coordinate, V_A is the dependence on change of all coordinates simultaneously (movement along diagonal). The presented characteristics assume fulfillment of conditions (2)–(6).

4.5. Encoding by binding codevectors of elements with codevectors of values

Consider the encoding scheme with linear growth of memory for storing codevectors. For encoding of each of A elements of the input vector, an individual codevector $\mathbf{F}_i$, $i = 1, A$, is used. Numeric values of Q are represented by a set of codevectors $\mathbf{Z}_q$, $q = \overline{1, Q}$. Then input vector $\mathbf{x} = (q_1, q_2, \ldots, q_A)$ is encoded as follows:

$$\mathbf{X} = \bigcup_{i=1,\ A} \left\langle \mathbf{F}_i, \mathbf{Z}_{q(i)} \right\rangle, \qquad (25)$$

where $\left\langle \mathbf{F}_i, \mathbf{Z}_{q(i)} \right\rangle$ is the codevector of the i-th element $\mathbf{F}_i$, bound with the codevector of its value $\mathbf{Z}_{q(i)}$.

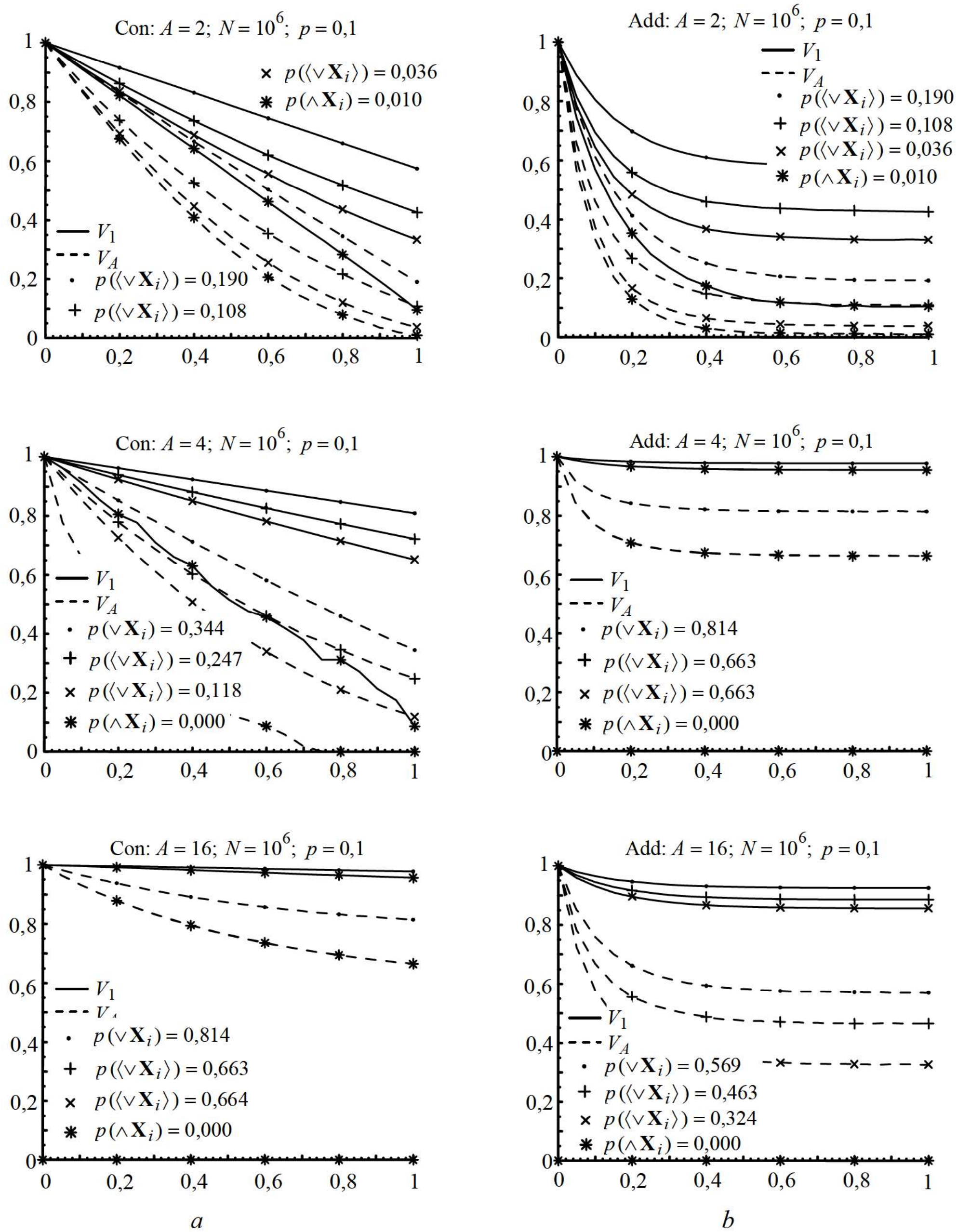


Figure 8

For binding we can use procedures of subsections 4.1, 4.2. The obtained codevector $\mathbf{X}$ is also subjected to binding by context-dependent thinning, thus resulting in $\langle \mathbf{X} \rangle$. This encoding scheme requires storage of $Q+A$ codevectors, each of which needs $O(pN)$ memory words. Computational complexity of forming codevector of the i-th dimension (disjunction plus transpositions and disjunctions CDT) is $O(pN(K+1))$ and complexity of forming codevector of the whole vector is $O(pNA(K+1))$. In the result, the whole procedure has complexity $O(pN(A+1)(K+1))$. The obtained characteristics of overlapping are analogous to characteristics for the scheme of subsection 4.4.

The suggested scheme also enables one to take into account dependence or similarity of elements of input vector by encoding such elements by correlated codevectors. Density of codevectors of elements $\mathbf{F}$ and their values $\mathbf{Z}$ should be selected in such a way, that the result of binding of each pair would have a low density. For each of elements of input vector we can use codevectors of numeric values with unique transposition (see [5]).

Conclusion

The studied in this paper bitwise-vector approaches to distributed encoding of numerical information by random binary codes use mainly computationally efficient bitwise operations over codevectors of scalars, elements of the input vector [5]. Characteristics of similarity (overlapping or Hamming distance) of the obtained codevectors can reflect aspects of similarity of their prototypes (vectors in the original space), useful for solving problems in some subject domains. Codevectors are sparse, this fact corresponds to the most important requirement from the viewpoint of efficiency of implementation, possibility of using effective associative memory [8] and also application of procedures of binding by context-dependent thinning [19] and methods of processing distributed information, suggested in APNN [1, 12, 15].

Recently, some papers were published (see review [10]), devoted to investigation of codes in the framework of direction, studying so called embeddings of spaces. The presence of embeddings of normalized spaces with metric $D_{d>1}$ in D_1 [10] in combination with the demonstrated in this paper possibility of approximation of D_1 by binary codevectors, opens opportunities of approximation of original metrics by functions of codes overlapping. Revealing parallels and the use of ideas and methods from this field can prove to be one of interesting directions of research.

The concept and assessment of similarity are important for modeling processes of natural intelligence in many methods and algorithms of AI, classification and recognition of images (associative memory, the method of the closest neighbor, methods and algorithms, based on scalar product etc.) This makes the suggested approaches, methods and procedures of encoding, generated by them codes and their properties an important object of studies and opens perspective of versatile use of the suggested schemes of representation of information in new highly intelligent information technologies.

References

1. Rachkovskij D.A., Kussul E.M., Binding and normalization of binary sparse distributed representations by context-dependent thinning, *Neural Computation*, 2001, **2**, No. 13, 411–452.
2. Thorpe S., Localized versus distributed representations, The Handbook of Brain Theory and Neural Networks, Ed. by M. Arbib, MIT Press, Cambridge, MA, 2003, 643–646.
3. Plate T., Holographic reduced representation: distributed representation for cognitive structures, Center for the Study of Language and Information Publications, Chicago, 2003.
4. Browne A., Sun R., Connectionist inference models, *Neural Networks*, 2001, **10**, No. 14, 1331–1355.
5. Rachkovskiy D.A., Slipchenko S.V., Kussul E.M., Baidyk T.N., Sparse binary distributed encoding of scalars, *Problemy upravleniya i informatiki*, 2005, No. 3, 89–102.
6. Stanford P.H., Smith D.J., The multidimensional scatter code: A data fusion technique with exponential capacity, *Proc. International Conference on Artificial Neural Networks*, Springer, Berlin, 1994, 1432–1435.
7. Kanerva P., Binary spatter-coding of ordered K-tuples, *Ibid.*, Springer, Berlin, 1996, 869–873.
8. Frolov A.A., Rachkovskij D.A., Husek D., On information characteristics of Willshaw-like autoassociative memory, *Neural Network World*, 2002, **2**, 141–157.
9. Rachkovskiy D.A., Slipchenko S.V., Kussul E.M., Baidyk T.N., Properties of codes of numeric quantities for a scheme of random subspaces RSC, *Kibernetika i sistemnyi analiz*, 2005, No. 4, 39–52.
10. Indyk P., Algorithmic applications of geometric embeddings, FOCS-01, *IEEE Computer Society*, 2001, 10–35.
11. Gormode G., Sequence distance embeddings: PhD Thesis, University of Warwick, 2003.
12. Kussul E.M., Associative neuron-like structures [in Russian], Naukova dumka, Kiev, 1991.
13. Artykutsa S.Ya., Baidyk T.N., Kussul E.M., Rachkovskiy D.A., Textures recognition with neurocomputer, Institut kibernetiki imeni V.M. Glushkova AN USSR, Kiev, 1991.
14. Rachkovskiy D.A., Fedoseeva T.V., Software-hardware neurocomputer complex for recognition of acoustic signals, *Neural-like networks and neurocomputers*, Institut kibernetiki imeni V.M. Glushkova AN USSR, Kiev, 1991, 62–68.
15. Baidyk T.N., Neural networks and problems of artificial intelligence [in Russian], Naukova dumka, Kiev, 2001.

16. Abeles M., Corticonics: neural circuits of the cerebral cortex, Cambridge University Press, Cambridge, UK, 1991.
17. Cristianini N., Shave-Taylor J., An introduction to support vector machines (and other Kernel-based learning methods), CUP, London, Paris, New York, 2000.
18. Goltsev A., An assembly neural network for texture segmentation, Neural Networks, 1996, 4, No. 9, 643–653.
19. Rachkovskiy D.A., Slipchenko S.V., Kussul E.M., Baidyk T.N., Binding procedure for binary distributed representation of data, Kibernetika i sistemnyi analiz, 2005, No. 3, 3–18.
20. Lukovich V.V., Rachkovskiy D.A., Study of characteristics of neural network classificator with random thresholds, Neirosetevyye tekhnologii i kompyutery, Kiev, Institut kibernetiki imeni V.M. Glushkova AN USSR, 1994, 27–36.