DOI: https:// 10.15407/kvt194.04.007

UDC 004.22 + 004.032.26

GRITSENKO V.I., Corresponding Member of NAS of Ukraine,
Director of International Research and Training
Center for Information Technologies and Systems
of the National Academy of Sciences of Ukraine
and Ministry of Education and Science of Ukraine
e-mail: vig@irtc.org.ua
RACHKOVSKIJ D.A., DSc (Engineering), Leading Researcher,
Dept. of Neural Information Processing Technologies
e-mail: dar@infrm.kiev.ua
REVUNOVA E.G., PhD (Engineering), Senior Researcher,
Dept. of Neural Information Processing Technologies
e-mail: egrevunova@gmail.com
International Research and Training Center for Information Technologies
and Systems of the National Academy of Sciences of Ukraine
and Ministry of Education and Science of Ukraine,
Acad. Glushkov av., 40, Kiev, 03187, Ukraine

NEURAL DISTRIBUTED REPRESENTATIONS OF VECTOR DATA IN INTELLIGENT INFORMATION TECHNOLOGIES

Introduction. *Distributed representation (DR) of data is a form of a vector representation, where each object is represented by a set of vector components, and each vector component can belong to representations of many objects. In ordinary vector representations, the meaning of each component is defined, which cannot be said about DR. However, the similarity of RP vectors reflects the similarity of the objects they represent.*

DR is a neural network approach based on modeling the representation of information in the brain, resulted from ideas about a "distributed" or "holographic" representations. DRs have a large information capacity, allow the use of a rich arsenal of methods developed for vector data, scale well for processing large amounts of data, and have a number of other advantages. Methods for data transformation to DRs have been developed for data of various types — from scalar and vector to graphs.

The purpose *of the article is to provide an overview of a part of the work of the Department of Neural Information Processing Technologies (International Center) in the field of neural network distributed representations. The approach is a development of the ideas of Nikolai Mikhailovich Amosov and his scientific school of modeling the structure and functions of the brain.*

Scope. *The formation of distributed representations from the original vector representations of objects using random projection is considered. With the help of the DR, it is possible to efficiently estimate the similarity of the original objects represented by numerical vectors.*

The use of DR allows developing regularization methods for obtaining a stable solution of discrete ill-posed inverse problems, increasing the computational efficiency and accuracy of their solution, analyzing analytically the accuracy of the solution. Thus DRs allow for increasing the efficiency of information technologies applying them.

Conclusions. *DRs of various data types can be used to improve the efficiency and intelligence level of information technologies. DRs have been developed for both weakly structured data, such as vectors, and for complex structured representations of objects, such as sequences, graphs of knowledge-base situations (episodes), etc. Transformation of different types of data into the DR vector format allows unifying the basic information technologies of their processing and achieving good scalability with an increase in the amount of data processed.*

In future, distributed representations will naturally combine information on structure and semantics to create computationally efficient and qualitatively new information technologies in which the processing of relational structures from knowledge bases is performed by the similarity of their DRs. The neurobiological relevance of distributed representations opens up the possibility of creating intelligent information technologies based on them that function similarly to the human brain.

Keywords: *distributed data representation, random projection, vector similarity estimation, discrete ill-posed problem, regularization.*

INTRODUCTION

The Department of Neural Information Processing Technologies of the International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences and the Ministry of Education and Science of Ukraine (International Center) is the heir to the Department of Biological and Medical Cybernetics, which was organized by Academician Amosov in 1962.

The main direction of research was considered by N.M. Amosov to be in the development of efficient neural network information processing technologies based on computer modeling of the principles of human thinking and features of the neural organization of the brain [1], [2]. The information technologies are intended for solving problems related to the field of Artificial Intelligence.

In this paper, we consider some of the research directions that have been developed over the past 20-ty years. Other areas are discussed in [3], [4]. Since the 1980s, the paradigm of associative-projective neural networks (APNNs) has been developed in the department [5], [2]. The idea of APNNs is to combine the hierarchical organization of the world model of Amosov with the advantages of structurally sensitive distributed representations as well as assemblies of Hebb.

APNNs are based on "distributed representations" of data of various types, nature, and complexity. Distributed representation [7], [8], [9], [10], [11] is a neural network approach based on modeling the representation of information in the brain that stemmed from the ideas of “distributed” or “holographic” representation of information.

Distributed representation (DR) of data is a form of a vector representation, where each object is represented by a set of vector components, and each vector component can belong to representations of many objects. In ordinary vector representations, the meaning of each component is defined, which cannot be said about DR. However, the similarity of the DR vectors reflects the similarity of the objects they represent.

Since DRs of various objects (from individual features to complex structured episodes of knowledge bases that are represented by hierarchically organized graphs) are vectors, a rich arsenal of methods developed for vector data can be applied to their processing. The computational complexity of these methods is usually not too high. For example, the complexity of calculating many measures of similarity of vectors is linear with respect to their dimension, while the complexity of calculating the editing distance between graphs is NP-complex in the general case. In addition, some types of DRs can reduce the complexity of computations compared to the vector representations from which they are obtained by reducing the dimension or using special formats for the DR vectors, such as binary and sparse ones. Distributed representations based on random projections are effectively used to regularize inverse machine learning problems, where the properties of the input-output transformation matrix lead to solution instability. Note that similar methods are being developed in areas known as "randomized algorithms", "random projections", "hyperdimensional computing" etc.

The DRs of APNNs use binary vectors with {0,1} components, which we call codevectors. Codevectors are sparse vectors, that is the proportion of non-zero components of the codevector is small. This data representation format is used in search engines, it allows one to achieve high efficiency of distributed auto-associative memory [12], [13], [14] and it is also required for the operation of APNNs. However, DRs can also be useful in the form of real-valued vectors (of small dimension), if they increase the processing efficiency relative to the initial representations of objects, such as vectors of high dimensionality, etc.

This article provides an overview of the distributed vector representations of source vector data on the basis of random projection, developed at the International Center both for efficient estimation of similarity of the initial vectors and for solving discrete inverse problems. The methods are protected by three patents and are used in information technologies for efficient processing of large data sets (Big Data) based on similarity as well as for efficient and accurate processing of signal information.

SIMILARITY ESTIMATION WITH DISTRIBUTED REPRESENTATIONS OF VECTOR DATA BASED ON RANDOM PROJECTIONS

Real-valued distributed representation of vector data is based on random projections. Most electronic digital data can be represented in the form of matrices or tables. For example, text corpuses for the purposes of search or classification are considered as word-text matrices, where the columns are texts, and the rows are words. The same information can be interpreted as a set of points in multidimensional space. The dimension of space can be, for example, hundreds of thousands (by the number of words in a language), and the number of points can be millions and billions (by the number of Internet web pages).

Many methods and algorithms of information retrieval, classification, clustering, approximation, learning, example-based reasoning, associative memory, etc. use measures of differences and similarities of vectors, such as Euclidean distance, scalar product, angle. Therefore, it would be useful to operate with transformed vector representations that have similarities consistent with the similarities in the original multidimensional vector space, but are more efficient in terms of saving memory, processing speed, the possibility of using special methods of data storage and processing.

Such a transformation can be performed by using a perceptron-like neural network. To solve the problems of classification, approximation, heteroassociative memory and others, the weights of the connections of such networks are usually tuned to the training set, starting with random weights. However, neural networks with a random structural organization have a number of useful properties.

We transform the input data, represented as the input matrix $\mathrm{A}(D \times N)$, where N is the number of vectors, into the matrix $\mathrm{U}(d \times N)$ by its feed, vector-by-vector, to a single-layer perceptron (Fig. 1) with random connections represented as a matrix $\mathrm{R}(d \times D)$, and thus performing $\mathrm{U} = \mathrm{RA}$. Note that for each vector-result of multiplication, component-based binarization operation can also be applied (see the next Subsection).

With a certain choice of **R**, by the resulting vectors (that is, by the d-dimensional column vectors of the matrix **U**), the distances between the original D-dimensional vectors in **A** can be calculated with high accuracy and computationally efficiently, even with $d << D$. For example, this is true for a random matrix **R** whose elements are formed as realizations of a Gaussian random variable.

Note that the components of the input vectors in our example with the presentation of texts had explicit semantics, i.e., the component corresponded to the word, and its value was a function of the occurrence frequency of the word in the text. The components of the output vectors corresponding to the texts no longer have such semantics of the components. However, similar output vectors correspond to similar input vectors. Such vectors are an example of distributed representations.

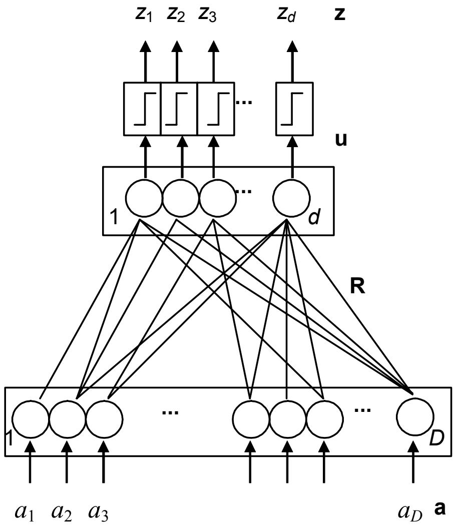

Fig. 1. Single-layer perceptron for transforming vector data by random projection

Random numbers in the floating point format required to represent Gaussian random variables are computationally difficult to generate, and they occupy a lot of space. Obviously, the simplest embodiment of the elements of random matrices are binary random variables with values of 0 and 1. They can be easily generated and stored. Then multiplication by a binary matrix is reduced to addition, that is, computationally simple. Note that binary random matrices in a neural network implementation are simply the set of present or missing links (of equal weight) between two pools of neurons. Therefore, of interest are methods of generating DRs using random binary matrices.

To transform the input vectors given in the floating-point format to the output vectors in the same format, we proposed to use projection by a random matrix with binary elements from the set {0, 1}. The element of the matrix takes value 1 with probability q, and the value 0 with probability $1 - q$. To center the result of the multiplication, we subtract $q \sum_{j = 1,D} a_j$ from it. Centering can also be provided by a binary matrix with the elements $\rho_{ij} = r_{ij} - q$. The analysis was carried out for such a matrix, and its results are close to the experimental results for the initial binary (sparse) matrix.

The output vectors allow estimating the scalar product and the Euclidean distance, as well as the Euclidean norm of the original vectors. The computational efficiency of the estimate increases as the dimension of the output vectors decreases. The error of estimating these similarity-difference measures was analyzed analytically and experimentally.

To compare the error of estimating the scalar product and the Euclidean distance by vectors after a random projection, it is informative to use the normalized standard deviation (coefficient of variation), i.e. the ratio of the root of the variance of the estimate to its expectation.

In [15], [16] it was shown that for the scalar product, the coefficient of variation $\mathrm{Var}^{1/2}\{\langle \mathbf{a},\mathbf{b}\rangle^*\} / \mathrm{E}\{\langle \mathbf{a},\mathbf{b}\rangle^*\}$ is equal to

$$(\langle \mathbf{a},\mathbf{b}\rangle\, d^{1/2})^{-1} \left[(\mathrm{E}\{\rho^4\} / \mathrm{E}^2\{\rho^2\} - 3) \textstyle\sum_{j=1,D} (a_j b_j)^2 + \langle \mathbf{a},\mathbf{b}\rangle^2 + ||\mathbf{a}||^2 ||\mathbf{b}||^2 \right]^{1/2}$$

and the square of the Euclidean distance $\mathrm{Var}^{1/2}\{||\mathbf{a} - \mathbf{b}||^{2*}\} / \mathrm{E}\{||\mathbf{a} - \mathbf{b}||^{2*}\}$ is equal to

$$(||\mathbf{a} - \mathbf{b}||^2\, d^{1/2})^{-1} \left[(\mathrm{E}\{\rho^4\} / \mathrm{E}^2\{\rho^2\} - 3) \textstyle\sum_{j=1,D} (a_j - b_j)^4 + 2\, ||\mathbf{a} - \mathbf{b}||^4 \right]^{1/2}.$$

Here ρ is a random variable, an element of the random matrix **R** (taking centering into account). Thus, for different distributions of ρ, we obtain various expressions for the coefficient of variation.

For a Gaussian random matrix i.i.d. elements, the value $\mathrm{E}\{\rho^4\} / \mathrm{E}^2\{\rho^2\} = 3$ [15].

For the binary random projection matrix under consideration (taking centering into account), it can be shown [16] that $\mathrm{E}\{\rho^4\} / \mathrm{E}^2\{\rho^2\} = 1/(q - q^2) - 3$. For a ternary matrix with elements from $\{-1/q^{1/2}, 0, +1/q^{1/2}\}$ with probabilities $\{q/2, 1 - q, q/2\}$ we get $\mathrm{E}\{\rho^4\} / \mathrm{E}^2\{\rho^2\} = 1/q$.

Let's compare the estimation errors obtained by projecting binary and ternary random matrices. The ratio $\mathrm{E}\{\rho^4\} / \mathrm{E}^2\{\rho^2\}$ for a binary random matrix is less than for a ternary one ($1/(q - q^2) - 3 < 1/q$) when $q < 2/3$. Therefore, for $q < 2/3$ the error of estimates obtained after the binary random projections that we propose is smaller than after ternary random projections at the same probability q of a nonzero matrix element (Fig. 2).

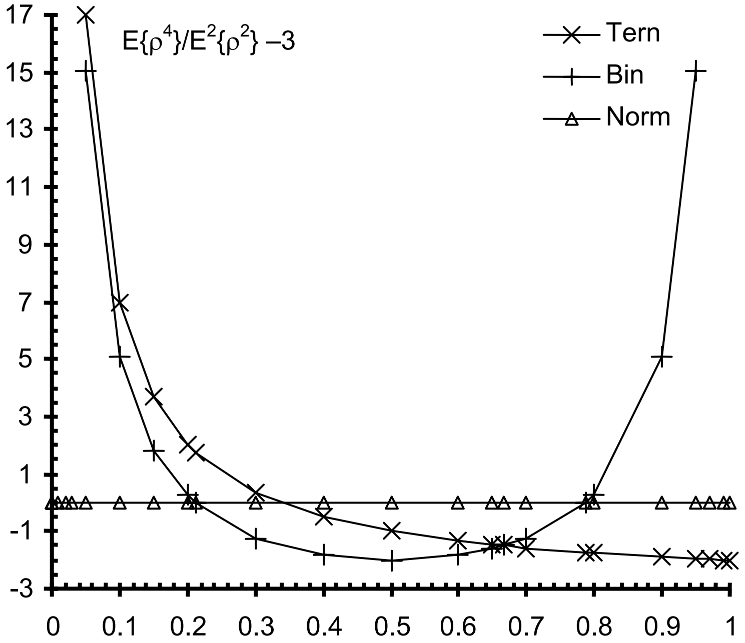

Fig. 2. The value $\mathrm{E}\{\rho^4\} / \mathrm{E}^2\{\rho^2\} - 3$ vs the probability q of a nonzero element of the matrix. Binary matrix (Bin), ternary matrix (Tern), Gaussian matrix (Norm)

Let us compare the error estimates for the binary and Gaussian random matrices. Since $1/(q-q^2)-6<0$ for $1/2-1/(2\sqrt{3})<q<1/2+1/(2\sqrt{3})$ (i.e., for $q \approx [0.2112, 0.7887]$), then binary random projections provide in this range the accuracy is higher than Gaussian (the greatest gain is achieved when $q=0.5$). On the other hand, projection acceleration requires $q \ll 0.5$, where the binary matrix loses due to the presence in the error of the terms with positive coefficients at $\sum_{j=1,D}(a_ib_i)^2$ and $\sum_{j=1,D}(a_i-b_i)^4$. However, when $D \gg 1$, their contribution is small (for data with a finite fourth moment), therefore, we obtain an accuracy comparable to the accuracy of Gaussian random projections, and for the case $q \ll 0.5$.

As for other types of random projection matrices, the error decreases with increasing dimension d of the output vectors $\sim 1/d^{1/2}$. Computational efficiency increases with decreasing q (with increasing "sparseness" of the binary projection matrix). In order to preserve the accuracy of estimates, the input vectors must have a sufficiently large dimension, as is assumed by the very formulation of the problem of efficiently evaluating the similarity of multidimensional vectors.

Binary distributed representation of vector data based on random projections. Let us apply in the output neurons of the perceptron network (Fig. 1) the binarizing threshold transformation $\mathbf{u} \to \mathbf{z}$: $z_i = 1$ when $u_i \geq t_i$ and $z_i = 0$ when $u_i < t_i$, where $t_i \geq 0$ is the threshold value for the i-th component of the output vector, $i = 1,...,d$.

The degree of sparseness of binary output vectors is governed by the threshold value. Moreover, the number of bits to represent binary vectors may be less than the number of bits per floating point representation of the vectors, even if the dimension of the binary vectors is larger.

After binarization $\mathbf{u} \to \mathbf{z}_1$ with a threshold t_1 and $\mathbf{v} \to \mathbf{z}_2$ with a threshold t_2, we determine the probability of coincidence of the unity components $z_{1,i} = 1$ and $z_{2,i} = 1$ of the codevectors $\mathbf{z}_1$ и $\mathbf{z}_2$. For standardized random variables (u_i,v_i), this probability is the probability of a simultaneous excess of the threshold values by the quantities u_i and v_i. The probability value is determined by the integral of the two-dimensional Gaussian distribution:

$$p_{\text{join}}(\theta) \equiv p(z_{1,i} = 1, z_{2,i} = 1 \mid \theta, t_1, t_2) = p(u_i \geq t_1, v_i \geq t_2 \mid \theta, t_1, t_2) =$$

$$= \frac{1}{2\pi(1-\cos^2\theta)} \int_{t_1}^{\infty}\int_{t_2}^{\infty} e^{-\frac{\eta_1^2 - 2\eta_1\eta_2\cos\theta + \eta_2^2}{2(1-\cos^2\theta)}} d\eta_1 d\eta_2 .$$

Thus p_{join} is a function of angle θ. From the p_{join} value, we can obtain the angle θ as $\theta = g(p_{\text{join}})$, where g is the function inverse to the function $p_{\text{join}}(\theta)$.

Therefore, θ can be estimated as follows:

— tabulate the $p_{\text{join}}(\theta)$ function;

— transform the input vectors into output codevectors $\mathbf{z}_1$ and $\mathbf{z}_2$, as indicated above;

— estimate p_{join} as $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle / d$;

— find in the tabulated table the value p_{join} closest to $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle / d$ and use the corresponding angle θ^* as the estimate of θ.

To standardize random variables (u_i,v_i) for matrices $\mathbf{R}$ from a symmetric distribution (e.g., Gaussian or ternary distribution with elements from {–1, 0,+1}, it is sufficient to apply scaling. For binary $\mathbf{R}$ it is also necessary to center, which is performed by subtracting $q \sum_{j=1,D} a_j$, and scaling is done by dividing by $(q - q^2)^{1/2}\ \|\mathbf{a}\|$.

The binarization threshold was chosen above under the assumption of a Gaussian distribution, and the relationship between p_{join} and $\cos\theta$ was made under the assumption of a two-dimensional Gaussian distribution. When projected by binary $\mathbf{R}$, the distributions are not Gaussian, but they converge to them. Analytically and experimentally we investigated the convergence of the distribution of the components of the real-valued vector, i.e., the result of the random projection, to the Gaussian distribution, and the rate of convergence. For this, the Lyapunov fraction of the third order (denoted by L_D) was used. For the sum of D random variables obtained by multiplying the input vector $\mathbf{a}$ by a row of the random matrix, the Lyapunov fraction can be represented as a product of two fractions, L_a and L_r:

$$L_D = \frac{\mathrm{E}\{|\, r - \mathrm{E}\{r\}\,|^3\}}{(\mathrm{E}\{|\, r - \mathrm{E}\{r\}\,|^2\})^{3/2}} \frac{\sum_{j=1}^{D} |\, a_j\,|^3}{\|\,\mathrm{a}\,\|_2^3} = L_r L_a,\ L_r \equiv \frac{\mathrm{E}\{|\, r - \mathrm{E}\{r\}\,|^3\}}{(\mathrm{E}\{|\, r - \mathrm{E}\{r\}\,|^2\})^{3/2}},$$

$$L_a \equiv \frac{\sum_{j=1}^{D} |\, a_j\,|^3}{\|\,\mathbf{a}\,\|_2^3} .$$

The fraction L_a depends on the input vector **a** and on D and does not depend on the type of random matrix **R**, and the fraction L_r depends on the distribution of random variables used in **R**.

According to the strong law of large numbers, the sample average of D realizations of independent and identically distributed random variables converges almost surely to the expected value (if it exists and finite) at $D \to \infty$. Apply this to L_a. Represent L_a as $L_a \equiv \frac{1}{D^{1/2}} \frac{\sum_{j=1}^{D} |a_j|^3 / D}{(\sum_{j=1}^{D} |a_j|^2 / D)^{3/2}}$.

We assume that the components a_j of the input vector **a** are realizations of random variables with a finite third absolute moment $\mathrm{E}\{|a_j|^3\} < \infty$. Then the moments of smaller orders are also finite, in particular, $\mathrm{E}\{|a_j|^2\} < \infty$. We use the strong law of large numbers for r.v. $|a_j|^3$ and for r.v. $|a_j|^2$. We obtain that when $D \to \infty$, then $\sum_{j=1,D} |a_j|^3 / D \to \mathrm{E}\{|a_j|^3\} < \infty$ and $\sum_{j=1,D} |a_j|^2 / D \to \mathrm{E}\{|a_j|^2\} < \infty$.

Therefore, L_a converges as follows:

$$L_a \equiv \frac{1}{D^{1/2}} \frac{\sum_{j=1}^{D} |a_j|^3 / D}{(\sum_{j=1}^{D} |a_j|^2 / D)^{3/2}} \to \frac{c}{D^{1/2}}, c \equiv \frac{\mathrm{E}\{|a_j|^3\}}{(\mathrm{E}\{|a_j|^2\})^{3/2}}.$$

Now consider L_r. It is easy to see [17] that for binary **R** the absolute central moments are equal to $\mathrm{E}\{|r - q|^3\} = (q - q^2)\ (1 - 2(q - q^2))$ and $\mathrm{E}\{|r - q|^2\} = q - q^2$. We obtain the expression for the fraction $L_r = (1 - 2q + +2q^2) / (q - q^2)^{1/2}$. Since $0 < q < 1$, then $1/2 \le 1 - 2q + 2q^2 < 1$ and $L_r \le 1 / (q - q^2)^{1/2}$.

Therefore, for a binary random matrix, the behavior of the entire Lyapunov fraction L_D is determined by the expression $(1 / (q - q^2) / D)^{1/2}$ when $D \to \infty$. Convergence to a Gaussian distribution occurs if the expression tends to zero, that is, if $1 / (q - q^2) = o(D)$.

The rate of convergence of the cumulative distribution function of the sum of independent random variables to the Gaussian cumulative distribution function can be estimated by the Berry-Esseen inequality. Its use in this problem is considered in [17]. The rate of convergence of the distribution of the components of a real-valued vector (the result of random projection) to the Gaussian distribution was studied analytically and experimentally for random matrices with discrete elements from $\{-1, 0, 1\}$ (ternary matrices) and from $\{0, 1\}$ (binary matrices) Using sparse random binary or ternary matrices instead of Gaussian random matrices allows obtaining output codevectors whose properties are similar to the properties of codevectors obtained using Gaussian matrices, when the dimension of input vectors is sufficiently high.

For binary and ternary random matrices, the experimental estimates of the difference between the distribution of u and the Gaussian distribution for the entire range of parameters studied are significantly less than the analytical limit values calculated by the right-hand side of the Berry-Esseen inequality. Experimental results for binary and ternary matrices are close to each other, the same is observed for analytical results.

Experimental results show that with the input vector dimension $D = 1000$, the difference between empirical and Gaussian distributions becomes close to the error level obtained due to the estimation of empirical distributions by the finite sample. This was observed for the studied probabilities $q = \{0.5, 0.1, 0.01\}$ of nonzero elements in random matrices.

The obtained results show that for parameters for which the empirical distributions are not very different from the Gaussian ones, the experimental and analytical errors in the angle determination by the output binary codevectors are also small. With the dimension of the input vector $D = 1000$, the experimental and analytical errors are close for all parameters studied. It follows that in applications it is necessary to comply with the condition $Dq > 10$.

Obviously, the smaller q, the greater the possibility of accelerating the implementation of vector projection. In this case, the projection by the binary matrix is potentially more efficient than by the ternary one, and even more so than by the Gaussian matrix.

We investigated [18] the estimates of similarity measures for real-valued vectors by binary codevectors obtained by projecting with a random binary matrix and then using the output threshold transform that allows us to adjust the degree of sparsity (the fraction of non-zero components) of binary codevectors. The similarity of binary codevectors was estimated by measures based on dot product (normalized to the codevector dimension).

The values of these codevector similarity measures decrease monotonically with increasing angle between the original real-valued vectors, and allow us to estimate the angle θ. The estimate together with the knowledge of the values of the Euclidean norms of the original real-valued vectors $||\mathbf{a}||_2$, $||\mathbf{b}||_2$ also made it possible to estimate their dot product $\langle \mathbf{a},\mathbf{b}\rangle^* = ||\mathbf{a}||\ ||\mathbf{b}||\ \cos\theta^*$ and Euclidean distance $||\mathbf{a} - \mathbf{b}||^2 = (||\mathbf{a}||^2 + ||\mathbf{b}||^2 - 2\ ||\mathbf{a}||\ ||\mathbf{b}||\ \cos\theta^*)^{1/2}$. The dependences of the error in estimating the angle, the scalar product, the Euclidean distance between the input real-valued vectors, on the angle value between them were analyzed analytically and experimentally.

To determine the expectation and variance of the angle estimate from the estimate $p_{\text{join}}^* = \langle \mathbf{z}_1, \mathbf{z}_2\rangle / d$ the linearization of the function of the random argument (delta-method) was used. The number of matching unit components $|\mathbf{A} \wedge \mathbf{B}|$ codevectors $\mathbf{A} = \mathbf{A}(a)$ and $\mathbf{B} = \mathbf{B}(b)$ of dimension d has a binomial distribution with probability of “success” p_{join} and d degrees of freedom, i.e. with the expectation value dp_{join} and the variance $dp_{\text{join}}(1 - p_{\text{join}})$. The estimate of $p_{\text{join}} = \text{sim}(a,b)$ by the empirical probability (or sample mean) $p_{\text{join}}^* = |\mathbf{A} \wedge \mathbf{B}| / d$ is unbiased: $\text{E}\{p_{\text{join}}^*\} = \text{E}\{|\mathbf{A} \wedge \mathbf{B}| / d\} = \text{E}\{|\mathbf{A} \wedge \mathbf{B}|\} / d = dp_{\text{join}} / d = p_{\text{join}}$.

Returning to the estimate sim* by p_{join}^*, we get $\text{E}\{\text{sim}^*\} \approx g(\text{E}\{p_{\text{join}}^*\})$ and $\text{Var}\{\text{sim}^*\} \approx (g'(\text{E}\{p_{\text{join}}^*\}))^2\ \text{Var}\{p_{\text{join}}^*\}$, with $\text{E}\{p_{\text{join}}^*\}$ and $\text{Var}\{p_{\text{join}}^*\}$ calculated as above. An approximate value of the derivative of $g'(\text{E}\{p_{\text{join}}^*\})$ can be determined for g specified using tabulation.

Errors of estimates by $|\mathbf{A} \wedge \mathbf{B}|/|\mathbf{A}|$ and by $1 - |\mathbf{A} \oplus \mathbf{B}|/d$ (where ⊕ is the component-wise XOR operation) are obtained similarly. The error values for binary and ternary random matrices are close for the studied parameter values. However, the proposed computational implementation of the transformation using a binary random matrix makes it easier.

A promising topic of further work is to study the effect of modifications of the proposed methods on the accuracy of similarity estimates, for example: using the real fraction of non-zero elements in the matrix as well as in its rows and columns instead of the probability of non-zero elements; the use of random matrices with a fixed number of randomly located non-zero elements in the entire matrix, as well as in its columns; taking into account the real fraction of non-zero components in the output binary codevectors when estimating the angle based on their dot product.

DISTRIBUTED REPRESENTATIONS BASED ON RANDOM PROJECTIONS FOR REGULARIZATION OF INVERSE PROBLEMS SOLVING IN MACHINE LEARNING

In problems of statistics and machine learning, a situation often arises when the solution by existing methods is unstable, i.e. small changes in the input data (conditions of the problem) lead to a large change in the solution. Such unstable solutions are inaccurate and cannot be used in practice. To remove the instability of the solution, the regularization approach is used.

Regularization imposes stability constraints on the sought solution. For example, a compromise of accuracy and stability is provided by choosing a regularization parameter that weights the ratio of the magnitude of the norm of the difference between the vectors of the reconstructed and the observed output, as well as the magnitudes of the norm of the solution vector (that is, the reconstructed input).

Our studies of the regularizing properties of random projection has begun since 2009 [19]. Later other researchers began to explore the regularizing properties of random projection, for example, for classification problems and machine learning [20], and, more recently, for solving inverse problems [21]. Since the approach of random projection, along with improving the accuracy of the solution by regularization, reduces the computational complexity of the solution, we have managed to develop algorithms that provide an accurate and fast solution for discrete inverse problems [22], [23], [24], [25], [26], [27], [28].

Let us consider in more detail the regularization of the inverse problem based on random projection. In many practical applications, signal transformation is described by a linear model of the form $\mathbf{y} = \mathbf{Ax} + \boldsymbol{\varepsilon}$, where the matrix $\mathbf{A} \in \Re^{N\times N}$ and the measurement vector $\mathbf{y} \in \Re^{N}$ ($\mathbf{y} = \mathbf{y}_0 + \boldsymbol{\varepsilon}$, $\mathbf{y}_0 = \mathbf{Ax}$) are known. The components of the noise vector $\boldsymbol{\varepsilon} \in \Re^{N}$ are realizations of independent Gaussian random variables with zero mean and variance σ^2. The signal vector $\mathbf{x} \in \Re^{N}$ has to be estimated.

In the case when $\mathbf{y}$ contains noise and the series of singular numbers of the matrix $\mathbf{A}$ smoothly drops to zero (with $\mathbf{A}$ having a high conditionality number), the problem of estimating $\mathbf{x}$ is called the discrete ill-posed problem (DIP) [29]. For DIP, the solution (estimate of signal $\mathbf{x}$) obtained on the basis of a pseudo-inversion as $\mathbf{x}^* = \mathbf{A}^+\mathbf{y}$, where $\mathbf{A}^+$ is a pseudoinverse [30], [31] is unstable and inaccurate. To overcome the instability and improve the accuracy of the solution, a regularization approach is used.

One of the approaches to ensuring the stability of solving ill-posed problems is the use of an integer regularization parameter, which is the number of summands in the model (linear with respect to parameters) approximating the original data. To obtain a stable solution (estimation $\mathbf{x}^*$), such methods as truncated singular value decomposition [32], truncated QR decomposition, and the method based on random projection [25], [26], [33] can be used.

To obtain solution based on random projection, both sides of the original equation are multiplied by the matrix $\mathbf{R}_k \in \Re^{k\times N}$ resulting in the equation

$$\mathbf{R}_k\mathbf{A}\mathbf{x} = \mathbf{R}_k\mathbf{y},$$

where $(\mathbf{R}_k\mathbf{A}) \in \Re^{k\times N}$, $(\mathbf{R}_k\mathbf{y}) \in \Re^{k}$. The vector of the recovered signal is obtained as

$$\mathbf{x}_k^* = (\mathbf{R}_k\mathbf{A})^+\mathbf{R}_k\mathbf{y}.$$

As a random matrix $\mathbf{R}$ we use:

- the matrix $\mathbf{G}_k\in\Re^{k\times N}$ whose elements are realizations of a random variable with a Gaussian distribution, zero mean and unit variance;
- the matrix $\mathbf{Q}_k\in\Re^{k\times N}$ obtained by QR decomposition of $\mathbf{GA}$ matrix ($\mathbf{GA}$=$\mathbf{QR}$);
- the matrix $\mathbf{\Omega}_k\in\Re^{k\times N}$ obtained by SVD decomposition of $\mathbf{G}$ matrix ($\mathbf{G} = \mathbf{\Omega\Sigma\Psi}^{\mathrm{T}}$).

Similar to estimating $\mathbf{x}$ based on truncated SVD $\mathbf{x}_{k\ SVD}^* = \sum_{i=1}^{k}\mathbf{v}_i s_i^{-1}\mathbf{u}_i^{\mathrm{T}}\mathbf{y}$ (where $\mathbf{u}_i \in \Re^N$, $\mathbf{v}_i \in \Re^N$ are left and right singular vectors, s_i are singular values), an estimate based on random projection can be represented by a linear model of the form [26]:

$$\mathbf{x}_{k\ R}^* = \sum_{i=1}^{k}\mathbf{h}_i\mathbf{r}_i^{\mathrm{T}}\mathbf{y}\ ,$$

where $\mathbf{r}_i \in \Re^N$ is the column of the matrix $\mathbf{R}_k = [\mathbf{r}_1,\ldots,\mathbf{r}_k]$, which is the result of a SVD of the matrix $\mathbf{R}$, whose elements are random variables with a normal distribution; $\mathbf{h}_i \in \Re^N$ is the column of the matrix $\left(\mathbf{Q}_k^{\mathrm{T}}\mathbf{A}\right)^+ = [\mathbf{h}_1,...,\mathbf{h}_k]$. Experimental studies have shown that there is an optimal number k ($k < N$) of the $\mathbf{R}$ rows, which minimizes the error $e_x = \left\|\mathbf{x} - \mathbf{x}_k^*\right\|^2$ of the true signal recovering. Fig. 3 shows an example of the $\mathbf{x}$ recovery error e_xQ (and its components e_xQ_1 and e_xQ_2) dependencies on k for the Phillips problem for three noise levels $\{10^{-2}, 10^{-3}, 10^{-4}\}$.

In reality, it is impossible to calculate the error $e_x(k)$ due to the lack of information about $\mathbf{x}$; therefore, it is impossible to directly determine the optimal k. To select k close to optimal, use the model selection criterion (MSC), i.e., a function that has an extremum when k is close or equal to optimal ([35], [36]).

When creating a MSC for solving DIP, it is required:

— to present an error in the form of the sum of two error components;

— to show increasing and decreasing of error components;

— to show that the dependence on k of the recovery error of $\mathbf{x}$ and of the recovery error of $\mathbf{y}_0$ has the global minima that coincide or are close;

— to obtain an expression for estimating the recovery error of $\mathbf{y}_0$ using the known measurement vector $\mathbf{y}$.

Search for the optimal number of rows of a random matrix. In [23], expressions for the recovery error of $\mathbf{x}$ were obtained for the random projection method:

$$e_x = \left\|(\mathbf{F}_k^+\mathbf{F}_k - \mathbf{I})\mathbf{x}\right\|^2 + \sigma^2 trace(\mathbf{F}_k^{+\mathrm{T}}\mathbf{F}_k^+)$$

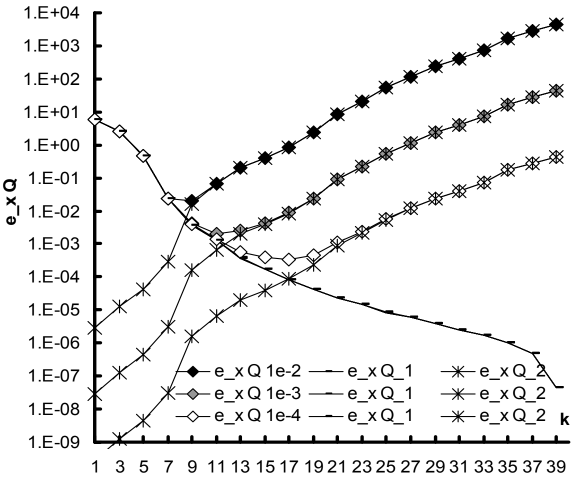

Fig. 3. Recovery error and its components vs the number *k* of the random matrix rows

as well as for the recovery error of $\mathbf{y}_0$:

$$e_y = \left\| (\mathbf{A}\mathbf{F}_k^+ \mathbf{Q}_k^{\mathrm{T}} - \mathbf{I})\mathbf{y}_0 \right\|^2 + \sigma^2 trace(\mathbf{F}_k^{+\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{A} \mathbf{F}_k^+) .$$

The number of columns N of the matrix $\mathbf{Q}_k$ is determined by the size of the original matrix $\mathbf{A}$. The number of rows k is not fixed a priori and can vary from 1 to N. The dependence of the error components (e_x, e_y) on the number of rows k of the matrix $\mathbf{Q}_k$ was analytically studied in [25]. Such a study is based on the representation of the matrix $\mathbf{F}_k = \mathbf{Q}_k^{\mathrm{T}} \mathbf{A}$ as the sum of the original matrix and the perturbation matrix.

In order to study the behavior of the components of the error e_x depending on k, we write the expression to get $\mathbf{F}_k^+$ in a recursive form. To do this, we use the representation of the perturbation of a pseudo-inverse matrix through the perturbation of the original matrix, proposed by Stewart:

$$\mathbf{B}^+ - \mathbf{A}^+ = -\mathbf{B}^+ \mathbf{P}_B \mathbf{E} \mathbf{R}_A \mathbf{A}^+ + \mathbf{B}^+ \mathbf{P}_B \mathbf{P}_A^\perp - \mathbf{R}_B^\perp \mathbf{R}_A \mathbf{A}^+ ,$$

where $\mathbf{B} = \mathbf{A} + \mathbf{E}$, $\mathbf{E}$ is the perturbation matrix, $\mathbf{P}_A = \mathbf{A}\mathbf{A}^+$ is the projector on the subspace of column vectors of the matrix $\mathbf{A}$, $\mathbf{R}_A = \mathbf{A}^+ \mathbf{A}$ is the projector on the subspace of row vectors of the matrix $\mathbf{A}$, $\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{P}_A$ and $\mathbf{R}_A^\perp = \mathbf{I} - \mathbf{R}_A$ are the projectors on the orthogonal complement of these subspaces, respectively.

Based on this representation, recursive expressions are obtained for the stochastic and deterministic components of the true signal recovery error. These expressions are a tool for studying the tendencies of (increasing, decreasing) behavior of the error components depending on k. The matrix $\mathbf{F}_k = \mathbf{Q}_k^{\mathrm{T}} \mathbf{A}$ is formed as:

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{F}_{k-1} \\ \mathbf{0}_k \end{bmatrix} + \begin{bmatrix} \mathbf{O}_{k-1} \\ \mathbf{f}_k \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{k-1} \\ \mathbf{f}_k \end{bmatrix},$$

where $\mathbf{F}_{k-1}\in\Re^{k-1\times N}$, the row vector $\mathbf{f}_k = \mathbf{q}_k\mathbf{A}$ has dimension k, $\mathbf{q}_k$ is the row of the matrix $\mathbf{Q}_k = [\mathbf{q}_1,\ldots,\mathbf{q}_k]^{\mathrm{T}}$, $\mathbf{O}_{k-1}$ is the zero submatrix size of $(k-1)\times N$. As a perturbation of the matrix $\mathbf{F}_{k-1}$, we consider a matrix $\mathbf{E}_k\in\Re^{k\times N}$ containing one nonzero k-th row $\mathbf{f}_k$, which is added at the k-th step.

In [25], recursive expressions for the stochastic and deterministic error components were obtained. For the stochastic component of the true signal recovery error, the recursive expression has the form:

$$e_{x\,s}(k)=\sigma^2 trace(\mathbf{F}_k^{+\mathrm{T}}\mathbf{F}_k^{+})=e_{x\,s}(k-1)+\sigma^2 trace(\mathbf{M}_{k-1}^{\mathrm{T}}\mathbf{M}_{k-1})+\sigma^2 d_k\,,$$

where $\mathbf{M}_{k-1}=\mathbf{f}_k^{+}\mathbf{f}_k\mathbf{F}_{k-1}^{+}$, $d_k=\mathbf{f}_k^{+\mathrm{T}}\mathbf{f}_k^{+}$. If $\mathbf{f}_k$ is nonzero then $d_k=\mathbf{f}_k^{+\mathrm{T}}\mathbf{f}_k^{+}>0$.

For a non-zero $\mathbf{M}_{k-1}$ we have $trace(\mathbf{M}_{k-1}^{\mathrm{T}}\mathbf{M}_{k-1})>0$. Therefore, the value of the stochastic component of the error increases with increasing k.

For the deterministic component of the true signal recovery error, the recursive expression has the form:

$$e_{xd}(k)=\mathbf{x}^{\mathrm{T}}\mathbf{x}-\mathbf{x}^{\mathrm{T}}\mathbf{F}_k^{+}\mathbf{F}_k\mathbf{x}=e_{xd}(k-1)-\mathbf{x}^{\mathrm{T}}\mathbf{f}_k^{+}\mathbf{f}_k(\mathbf{I}-\mathbf{F}_{k-1}^{+}\mathbf{F}_{k-1})\mathbf{x}\,.$$

As shown in [25], the matrix $\mathbf{f}_k^{+}\mathbf{f}_k(\mathbf{I}-\mathbf{F}_{k-1}^{+}\mathbf{F}_{k-1})$ can be obtained by the product of the column-vector $(\mathbf{f}_k(\mathbf{I}-\mathbf{F}_{k-1}^{+}\mathbf{F}_{k-1}))^{\mathrm{T}}$ and the same (not transposed) row vector divided by the square of its norm. Therefore, for non-orthogonal $\mathbf{x}$ and $\mathbf{f}_k(\mathbf{I}-\mathbf{F}_{k-1}^{+}\mathbf{F}_{k-1})$, the value of the deterministic component of the true signal recovery error decreases with increasing k.

An experimental study demonstrating a decrease in the deterministic component and an increase in the stochastic component of e_y, the proximity of the global minima of the $e_x(k)$ и $e_y(k)$ dependences, was carried out in [25]. The criterion for selecting the model $\mathbf{x}_{kR}$* when solving DIP based on random projection, which is an approximation of the output vector recovery error e_y, was obtained in [26]:

$$CR_Q=\mathrm{E}\left\|(\mathbf{A}\mathbf{F}_k^{+}\mathbf{Q}_k^{\mathrm{T}}-\mathbf{I})\mathbf{y}\right\|^2-\sigma^2 trace((\mathbf{A}\mathbf{F}_k^{+}\mathbf{Q}_k^{\mathrm{T}}-\mathbf{I})^{\mathrm{T}}(\mathbf{A}\mathbf{F}_k^{+}\mathbf{Q}_k^{\mathrm{T}}-\mathbf{I}))+$$
$$+\sigma^2 trace(\mathbf{F}_k^{+\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{F}_k)$$

Fig. 4 shows graphs of the values of the $\mathrm{CR_Q}$ criterion, the average output recovery error (e_y^{mean}), and the average true signal recovery error (e_x^{mean}) vs k for the Carasso problem. The graph lines corresponding to the $\mathrm{CR_Q}$ criterion and the average recovery error of $\mathbf{y}_0$ are close, so the MSC $\mathrm{CR_Q}$ well approximates the recovery error of the output $\mathbf{y}_0$. The positions of the minima are also close.

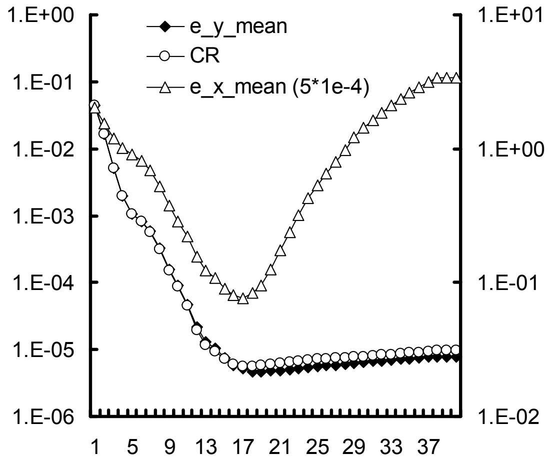

Fig. 4. Dependences of $\mathrm{CR}_Q(k)$, $e_x(k)$ and $e_y(k)$, for the noise level of 5×10^{-4}

Reducing the input vector recovery error. The accuracy of the DIP solution by the method of random projection depends on two independent random variables. The first one is the additive noise in the output vector (whose distribution is assumed to be Gaussian, and the variance is generally unknown) and the second one is the random variable that forms the random matrix (Gaussian distribution with the unit variance, in the studied case). Changing the number of rows k of the random matrix leads to a change in the accuracy of the DIP solution. In the absence of noise in the output vector, an increase in the number of rows of a random matrix leads to a decrease in the solution error. Noise in the output vector leads to the appearance of an error component, the value of which increases with increasing number of rows of a random matrix. Therefore, the dependence of the error of the DIP solution on the number of rows of the random matrix has a minimum at $k < N$ (at certain noise levels). In order to make the dependence of the error on the number of the random matrix rows more smooth and thereby facilitate the search for the minimum, in [22], [23], [24], [25], [26] we performed averaging over noise in the output vector. Experimental studies [23], [26] showed that averaging over random matrices leads to a smoothing of the $e_R(k)$ dependence and a decrease in the number of local minima. Analytic averaging over random matrices can lead to simpler expressions for $e_R(k)$, facilitate further analytical research and improve the accuracy of the method of DIP solving based on random projection.

In [34] important results were obtained related to averaging (finding the expectation) over random matrices $\mathbf{R}_k$ of expressions of the form:

$$\mathrm{E}_R\{\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{B}\mathbf{R}_k^{\mathrm{T}})^{-1}\mathbf{R}_k\},\ \mathrm{E}_R\{\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{Z}\mathbf{R}_k^{\mathrm{T}})^{-1}\mathbf{R}_k\},$$

where $\mathbf{B}\in\Re^{N\times N}$ is any symmetric positive semidefinite matrix, $\mathbf{Z}\in\Re^{N\times N}$ is the diagonal matrix of the eigenvalues of the matrix $\mathbf{B}$. In particular,

$$\mathrm{E}_R\{\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{B}\mathbf{R}_k^{\mathrm{T}})^{-1}\mathbf{R}_k\} = \mathbf{U}\mathrm{E}_R\{\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{Z}\mathbf{R}_k^{\mathrm{T}})^{-1}\mathbf{R}_k\}\mathbf{U}^{\mathrm{T}}.$$

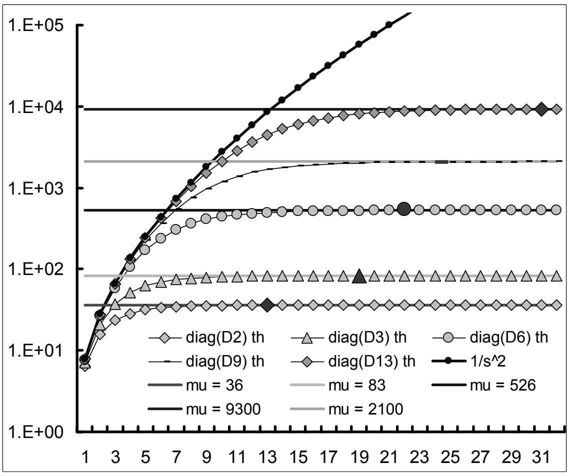

Fig. 5. Theoretical values of the elements of the diagonal of the matrix $\mathbf{D}_k$ for the Carasso problem

For the structure of the form $\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{Z}\mathbf{R}_k^{\mathrm{T}})^{-1}\mathbf{R}_k$, in [34] it is shown that

$$\mathrm{E}_R\{\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{Z}\mathbf{R}_k^{\mathrm{T}})^{-1}\mathbf{R}_k\} = \mathrm{diag}(\lambda_1,\ldots,\lambda_m,\ldots,\mu,\ldots,\mu) \equiv \mathbf{D}_k(\mathbf{Z}_m),$$

that is, $\mathbf{D}_k(\mathbf{Z}_m)$ $(N\times N)$ is diagonal, where $\lambda_i = \mu/(1+\mu s_i^2)$, $\mu = const.$

Averaging over random matrices leads to the diagonalization of the matrix included in both error components (deterministic and stochastic).

An experimental study of $\mathbf{D}_k(\mathbf{Z}_m)$ showed that the sequence $\lambda_1,\ldots,\lambda_m$ is bounded by the sequence $s_1^{-2},\ldots,s_m^{-2}$ from above and that several initial values of the diagonal of the matrix $\mathbf{D}_k$ approach s_i^{-2} with great accuracy. For example, for the Carasso problem the values of the diagonal elements of the matrix $\mathbf{D}_k$ for $k = \{2, 3, 6, 9, 13\}$ are shown in Fig. 5, where s_i^{-2} and μ are also shown. The values of the diagonal elements vary monotonically with k. This leads to a smoothing of the $e_x(k)$ and $e_y(k)$ dependences and to a decrease in the number of local minima. The results of an experimental study showed the connection of the element values of the diagonalized matrix with the singular values of the original one, which creates the basis for the study of the relationship of the truncated SVD and random projection.

To average the e_x error over random matrices, the error components were transformed so that they include the matrix structure of the form $\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{R}_k^{\mathrm{T}})^{-1}\mathbf{R}_k$ and then averaging was performed:

$$\mathrm{E}_R\{e_x\} = \mathrm{E}_R\{e_{x\,d}\} + \mathrm{E}_R\{e_{x\,s}\} = \mathbf{x}^{\mathrm{T}}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathrm{E}_R\{\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{R}_k^{\mathrm{T}})^{+}\mathbf{R}_k\}\mathbf{A}\mathbf{x} +$$
$$+\sigma^2 trace\ \mathrm{E}_R\{\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{R}_k^{\mathrm{T}})^{+}\mathbf{R}_k\},$$

$$\mathrm{E}_R\{e\} = \mathbf{x}^{\mathrm{T}}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{V}\mathbf{S}^2\mathbf{D}_k\mathbf{V}^{\mathrm{T}}\mathbf{x} + \sigma^2 trace(\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}).$$

The resulting expression for $\mathrm{E}_R\{e\}$ does not allow us to identify the entire structure of the error. The bias and variance of error that appear due to averaging over the realizations of the random matrix are not explicitly presented and therefore cannot be analyzed. In [33] expressions for the error were obtained in a form that allows us to propose a method for solving the DIP with a reduced error with respect to the random projection method [23], [25]. By analogy with the works that studied bias and variance of the error arising due to the presence of additive noise in the output vector, we call the component $e_{RB} = \|\mathbf{x} - \bar{\mathbf{x}}\|^2$ as the squared bias, and $e_{RV} = \|\bar{\mathbf{x}} - (\mathbf{R}_k\mathbf{A})^{+}\mathbf{R}_k\mathbf{b}\|^2$ as the variance of the vector **x** (input) recovery when averaged over random matrices $\mathbf{R}_k$. In [33], the bias and variance components were obtained by averaging the error e_x over the realizations of the random matrix $\mathbf{R}_k$. Bias and variance of the input vector recovery error are

$$e_{RV} = \mathbf{b}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b} - \|\mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}\|^2,\ e_{RB} = \|\mathbf{x} - \bar{\mathbf{x}}\|^2 = \|\mathbf{x} - \mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}\|^2,$$

where $\bar{\mathbf{x}} = \mathrm{E}_R\{(\mathbf{R}_k\mathbf{A})^{+}\mathbf{R}_k\mathbf{b}\} = \mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}$.

The input vector **x** recovery error, averaged over random matrices is

$$\mathrm{E}_R\{e_R\} = e_{RB} + e_{RV} = \|\mathbf{x} - \mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}\|^2 + \mathbf{b}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b} - \|\mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}\|^2.$$

From the expression for $\mathrm{E}_R\{e_R\}$ it can be seen that the recovery error of the input vector without averaging over noise can be reduced by the variance value e_{VR} (resulting from multiplication by a random matrix and subsequent averaging). To do this, the recovery of the input vector should be performed as $\tilde{\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}$, where **U** was obtained from SVD-decomposition $\mathbf{A}=\mathbf{A}\mathbf{S}\mathbf{V}^{\mathrm{T}}$, and $\mathbf{D}_k$ was obtained as $\mathrm{E}_R\{\mathbf{R}_k^{\mathrm{T}}(\mathbf{R}_k\mathbf{S}^2\mathbf{R}_k^{\mathrm{T}})^{-1}\mathbf{R}_k\} = \mathbf{D}_k$. Indeed, this gives us an error that coincides with the squared bias in $\mathrm{E}_R\{e_R\}$:

$$e_{DR} = \|\mathbf{x} - \tilde{\mathbf{x}}\|^2 = \|\mathbf{x} - \mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}\|^2.$$

Let us call the method of solving DIP according to $\tilde{\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}$ as the deterministic method based on analytical averaging of random projection (DRP). From expression e_{DR} it can be seen that the error of the input vector recovery, when noise-averaged

$$\mathrm{E}_\varepsilon\{\mathrm{E}_R\{e_R\}\} = \mathrm{E}_\varepsilon\{e_{RB}\} + \mathrm{E}_\varepsilon\{e_{RV}\} = \mathrm{E}_\varepsilon\{\|\mathbf{x} - \mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}\|^2\} +$$
$$+ \mathrm{E}_\varepsilon\{\mathbf{b}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b} - \|\mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}\|^2\}$$

is greater than the error e_{DR}, when noise-averaged averaged, i.e.

$$\mathrm{E}_\varepsilon\{e_{DR}\} = \mathrm{E}_\varepsilon\{\|\mathbf{x} - \tilde{\mathbf{x}}\|^2\} = \mathrm{E}_\varepsilon\{\|\mathbf{x} - \mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{D}_k\mathbf{U}^{\mathrm{T}}\mathbf{b}\|^2\}.$$

A function approximating the output vector recovery error of the DRP method was obtained in [33]:

$$CR_{DR} = \mathrm{E}_{\varepsilon}\{\|(\mathbf{AA}^{\mathrm{T}}\mathbf{UD}_k\mathbf{U}^{\mathrm{T}} - \mathbf{I})\mathbf{b}\|^2\} - \sigma^2 trace((\mathbf{AA}^{\mathrm{T}}\mathbf{UD}_k\mathbf{U}^{\mathrm{T}} - \mathbf{I})^{\mathrm{T}}(\mathbf{AA}^{\mathrm{T}}\mathbf{UD}_k\mathbf{U}^{\mathrm{T}} - \mathbf{I})) + \\ + \sigma^2 trace((\mathbf{AA}^{\mathrm{T}}\mathbf{UD}_k\mathbf{U}^{\mathrm{T}})^{\mathrm{T}}\mathbf{AA}^{\mathrm{T}}\mathbf{UD}_k\mathbf{U}^{\mathrm{T}}).$$

Averaging over random matrices leads to further smoothing the dependence of the error on the size *k* of the random matrix [23], [26].

Analytical averaging over random matrices allowed us to analyze the bias and variance of errors that appear due to averaging over the realizations of the random matrix, and to obtain an estimate of the input vector that provides greater accuracy of the DIP solution relative to the estimate obtained by averaging both over noise and over the matrices.

An experimental study [33] showed that the proposed estimate of the input vector provides a solution accuracy that is very close to the accuracy of the truncated singular value decomposition method. However the dependence of the error on *k* is smoother than for the truncated singular value decomposition. It is assumed that the resulting smoothness can be used to improve the accuracy of the DIP solution in real problems due to a more accurate choice of the optimal dimension of the model by the model selection criteria.

CONCLUSIONS

This article provides an overview of some of the research of the International Center in the field of neural network distributed representations. The formation of distributed representations from the original vector representations of objects using random projection is considered. Distributed representations allow one to efficiently estimate the similarity of the original objects. They can also be used in linear classifiers to perform an effective classification of objects whose representations are not linearly separable in the input space [37], [38], [39], [40]. The use of distributed representations formed by random projection allows increasing the computational efficiency and accuracy of information technologies based on solving discrete ill-posed problems [41], [42]. The solution accuracy of discrete ill-posed problems was investigated analytically, [25], [26], [33]. These developments are protected by three patents.

We note, however, that distribution representations include not only random projection based methods [15], [16], [17], [18], [43], [44], [45] but also a number of other representation schemes for vectors, such as those based on receptive fields of other types [46] or compositional methods [47], [48], [49], [50]. DRs can be used to represent certain types of images using a special type of LiRA receptive fields [51], [52], [53]. One of the promising research directions could consist in using DRs in the texture segmentation problem [54], [55], [56] and in classification of satellite optical and SAR images [57], [58], [59].

For a long time it was believed that the main drawback of distributed representations is the inability to represent structure. Recently, however, DRs have been developed for complexly structured representations of objects, such as sequences [60], [61], [62], [63], [64] or graphs of situations (episodes) of knowledge bases, etc., e.g. [65], [66], [7], [8], [3].

Thus, distributed representations built on the basis of ideas about the representation of information in the brain, when used in information technologies, increase their computational efficiency by converting data of different types — both unstructured information in the form of vector arrays and relational structures of knowledge bases — into a special format of vectors. In addition, distributed representations allow naturally combining information about structure and semantics, giving a basis for creating computationally efficient and qualitatively new information technologies for processing relational structures from data and knowledge bases. The neurobiological relevance of distributed representations opens the way to the creation on their basis of intelligent information technologies that function similarly to the human brain.

REFERENCES

1. Amosov N. M. Modelling of thinking and the mind. New York: Spartan Books, 1967. 192 p.
2. Amosov N.M., Baidyk T.N., Goltsev A.D., Kasatkin A.M., Kasatkina L.M., Rachkovskij D.A. Neurocomputers and Intelligent Robots. Kyiv: Nauk. Dumka. 1991. 269 p. (in Russian)
3. Gritsenko V.I., Rachkovskij D.A., Goltsev A.D., Lukovych V.V., Misuno I.S., Revunova E.G., Slipchenko S.V., Sokolov A.M., Talayev S.A. Neural distributed representation for intelligent information technologies and modeling of thinking. *Kibernetika i vyčislitel`naâ tehnika*. 2013. Vol. 173. P. 7–24. (in Russian)
4. Goltsev A.D., Gritsenko V.I. Neural network technologies in the problem of handwriting recognition. *Control Systems and Machines.* 2018. N 4. P. 3–20. (in Russian).
5. Kussul E.M. Associative neuron-like structures. Kyiv: Nauk. Dumka. 1992. 144 p. (in Russian)
6. Kussul E.M., Rachkovskij D.A., Baidyk T.N. Associative-Projective Neural Networks: Architecture, Implementation, Applications. *Proc. Neuro-Nimes'91.* (Nimes, 25–29th of Oct. 25–29, 1991). Nimes, 1991. P. 463–476.
7. Gayler R. Multiplicative binding, representation operators, and analogy. Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences. Edited by K. Holyoak, D. Gentner, and B. Kokinov. Sofia, Bulgaria: New Bulgarian University, 1998. P. 405.
8. Kanerva P. Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation.* 2009. Vol. 1, N 2. P. 139–159.
9. Goltsev A., Husek D. Some properties of the assembly neural networks. *Neural Network World*. 2002. Vol. 12, N. 1. P. 15–32.
10. Goltsev A.D. Neural networks with assembly organization. Kyiv: Nauk. Dumka. 2005. 200 p. (in Russian)
11. Goltsev A., Gritsenko V. Modular neural networks with radial neural columnar architecture. *Biologically Inspired Cognitive Architectures*. 2015. Vol. 13. P. 63–74.
12. Frolov A.A., Rachkovskij D.A., Husek D. On information characteristics of Willshaw-like auto-associative memory. *Neural Network World*. 2002. Vol. 12, No 2. P. 141–158.
13. Frolov A.A., Husek D., Rachkovskij D.A. Time of searching for similar binary vectors in associative memory. *Cybernetics and Systems Analysis*. 2006. Vol. 42, N 5. P. 615–623.
14. Gritsenko V.I., Rachkovskij D.A., Frolov A.A., Gayler R., Kleyko D., Osipov E. Neural distributed autoassociative memories: A survey. *Kibernetika i vyčislitel`naâ tehnika.* 2017. N 2 (188). P. 5–35.
15. Li P., Hastie T.J., Church K.W. Very sparse random projections. *Proc. KDD'06.* (Philadelphia, 20 – 23th of Aug.). Philadelphia, 2006. P. 287–296.
16. Rachkovskij D.A. Vector data transformation using random binary matrices. *Cybernetics and Systems Analysis*. 2014. Vol. 50, N 6. P. 960–968.

17. Rachkovskij D.A. Formation of similarity-reflecting binary vectors with random binary projections. *Cybernetics and Systems Analysis*. 2015. Vol. 51, N 2. P. 313–323.
18. Rachkovskij D.A. Estimation of vectors similarity by their randomized binary projections. *Cybernetics and Systems Analysis.* 2015. Vol. 51, N 5. P. 808–818.
19. Revunova E.G., Rachkovskij D.A. Using randomized algorithms for solving discrete ill-posed problems. *Intern. Journal Information Theories and Applications*. 2009. Vol. 16, N 2. P. 176–192.
20. Durrant R.J., Kaban A. Random projections as regularizers: learning a linear discriminant from fewer observations than dimensions. *Machine Learning*. 2015. Vol. 99, N 2. P. 257–286.
21. Xiang H., Zou J. Randomized algorithms for large-scale inverse problems with general Tikhonov regularizations. *Inverse Problems.* 2015. Vol. 31, N 8: 085008. P. 1–24.
22. Revunova E.G. Study of error components for solution of the inverse problem using random projections. *Mathematical Machines and Systems*. 2010. N 4. P. 33–42 (in Russian).
23. Rachkovskij D.A., Revunova E.G. Randomized method for solving discrete ill-posed problems. *Cybernetics and Systems Analysis*. 2012. Vol. 48, N. 4. P. 621–635.
24. Revunova E.G. Randomization approach to the reconstruction of signals resulted from indirect measurements. *Proc. ICIM'13* (Kyiv 16–20th of Sept., 2013). Kyiv, 2013. P. 203–208.
25. Revunova E.G. Analytical study of the error components for the solution of discrete ill-posed problems using random projections. *Cybernetics and Systems Analysis*. 2015. Vol. 51, N. 6. P. 978–991.
26. Revunova E.G. Model selection criteria for a linear model to solve discrete ill-posed problems on the basis of singular decomposition and random projection. *Cybernetics and Systems Analysis*. 2016. Vol. 52, N.4. P. 647–664.
27. Revunova E.G. Averaging over matrices in solving discrete ill-posed problems on the basis of random projection. *Proc. CSIT'17* (Lviv 05–08th of Sept., 2017). Lviv, 2017. Vol. 1. P. 473–478.
28. Revunova E.G. Solution of the discrete ill-posed problem on the basis of singular value decomposition and random projection. *Advances in Intelligent Systems and Computing II*. Cham: Springer. 2018. P. 434–449.
29. Hansen P. Rank-deficient and discrete ill-posed problems. Numerical aspects of linear inversion. Philadelphia: SIAM, 1998. 247 p.
30. Nowicki D., Verga P., Siegelmann H. Modeling reconsolidation in kernel associative memory. *PLoS ONE*. 2013. Vol. 8(8): e68189. doi:10.1371/journal.pone.0068189.
31. Nowicki D, Siegelmann H. Flexible kernel memory. *PLoS ONE.* 2010. Vol. 5(6): e10955. doi:10.1371/journal.pone.0010955.
32. Revunova E.G., Tyshchuk A.V. A model selection criterion for solution of discrete ill-posed problems based on the singular value decomposition. *Proc. IWIM'2015* (20–24th of July, 2015, Kyiv-Zhukin). Kyiv-Zhukin, 2015. P.43–47.
33. Revunova E.G. Improving the accuracy of the solution of discrete ill-posed problem by random projection. *Cybernetics and Systems Analysis.* 2018. Vol. 54, N 5. P. 842–852.
34. Marzetta T., Tucci G., Simon S. A random matrix-theoretic approach to handling singular covariance estimates. *IEEE Trans. Information Theory*. 2011. Vol. 57, N 9. P. 6256–6271.
35. Stepashko V. Theoretical aspects of GMDH as a method of inductive modeling. *Control systems and machines*. 2003. N 2. P. 31–38. (in Russian)
36. Stepashko V. Method of critical variances as analytical tool of theory of inductive modeling. *Journal of Automation and Information Sciences*. 2008. Vol. 40, N 3. P. 4–22.
37. Kussul E.M., Baidyk T.N., Lukovich V.V., Rachkovskij D.A. Adaptive neural network classifier with multifloat input coding. *Proc. Neuro-Nimes'93* (25–29th of Oct., 1993, Nimes). Nîmes, France, 1993 P. 209–216.
38. Lukovich V.V., Goltsev A.D., Rachkovskij D.A. Neural network classifiers for micromechanical equipment diagnostics and micromechanical product quality inspection. *Proc. EUFIT'97* (8–11th of Sept, 1997, Aachen). Aachen, Germany, 1997. P. 534–536.
39. Kussul E.M., Kasatkina L.M., Rachkovskij D.A., Wunsch D.C. Application of random threshold neural networks for diagnostics of micro machine tool condition. *Proc. IJCNN'01* (4–9th of May, 1998, Anchorage). Anchorage, Alaska, USA, 1998 P. 241–244.

40. Gol'tsev A.D. Structured neural networks with learning for texture segmentation in images. *Cybernetics and Systems Analysis*. 1991. Vol. 27, N 6. P. 927–936.
41. Rachkovskij D.A., Revunova E.G. Intelligent gamma-ray data processing for environmental monitoring. In: Intelligent Data Processing in Global Monitoring for Environment and Security. Kyiv-Sofia: ITHEA. 2011. P. 136–157.
42. Revunova E.G., Rachkovskij D.A. Random projection and truncated SVD for estimating direction of arrival in antenna array. *Kibernetika i vyčislitel`naâ tehnika.* 2018. N 3(193). P. 5–26.
43. Ferdowsi S., Voloshynovskiy S., Kostadinov D., Holotyak T. Fast content identification in highdimensional feature spaces using sparse ternary codes. *Proc. WIFS'16* (4–7th of Dec., 2016, Abu Dhabi) Abu Dhabi, UAE, 2016. P. 1–6.
44. Dasgupta S., Stevens C.F., Navlakha S. A neural algorithm for a fundamental computing problem. *Science.* 2017. Vol. 358(6364). P. 793–796.
45. Iclanzan D., Szilagyi S.M., Szilagyi L.. Evolving computationally efficient hashing for similarity search. *Proc. ICONIP'18.* 2. (Siem Reap, 15-18th of Dec., 2018). Siem Reap, Cambodia, 2018. 2018.
46. Rachkovskij D.A., Slipchenko S.V., Kussul E.M., Baidyk T.N. Properties of numeric codes for the scheme of random subspaces RSC. *Cybernetics and Systems Analysis*. 2005. Vol. 41, N. 4. P. 509–520.
47. Rachkovskij D.A., Slipchenko S.V., Kussul E.M., Baidyk T.N. Sparse binary distributed encoding of scalars. 2005. *Journal of Automation and Information Sciences*. Vol. 37, N 6. P. 12–23.
48. Rachkovskij D.A., Slipchenko S.V., Misuno I.S., Kussul E.M., Baidyk T. N. Sparse binary distributed encoding of numeric vectors. *Journal of Automation and Information Sciences*. 2005. Vol. 37, N 11. P. 47–61.
49. Kleyko D., Osipov E., Rachkovskij D.A. Modification of holographic graph neuron using sparse distributed representations. *Procedia Computer Science*. 2016. Vol. 88. P. 39–45.
50. Kleyko D., Rahimi A., Rachkovskij D., Osipov E., Rabaey J. Classification and recall with binary hyperdimensional computing: Tradeoffs in choice of density and mapping characteristics. *IEEE Trans. Neural Netw. Learn. Syst.* 2018.
51. Kussul E., Baidyk T., Kasatkina L. Lukovich V. Rosenblatt perceptrons for handwritten digit recognition. *Proc. IJCNN'01.* (Washington, 15–19 July, 2001). Washington, USA. 2001. P. 1516–1521.
52. Baidyk T, Kussul E., Makeyev O., Vega A., Limited receptive area neural classifier based image recognition in micromechanics and agriculture. *International Journal of Applied Mathematics and Informatics.* 2008.Vol. 2, N 3. P. 96–103.
53. Baydyk T., Kussul E., Hernández Acosta M. LIRA neural network application for microcomponent measurement. *International Journal of Applied Mathematics and Informatics.* Vol.6, N 4. 2012. P.173–180.
54. Goltsev A.D., Gritsenko V.I. Algorithm of sequential finding the textural features characterizing homogeneous texture segments for the image segmentation task. *Kibernetika i vyčislitel`naâ tehnika.* 2013. N 173. P. 25–34 (in Russian).
55. Goltsev A., Gritsenko V., Kussul E., Baidyk T. Finding the texture features characterizing the most homogeneous texture segment in the image. *Proc. IWANN'15.* (Palma de Mallorca, Spain, June 10-12, 2015). Palma de Mallorca, 2015. 2015. P. 287–300.
56. Goltsev A., Gritsenko V., Husek D. Extraction of homogeneous fine-grained texture segments in visual images. *Neural Network World.* 2017. Vol. 27, N 5. P. 447– 477.
57. Kussul N.N., Sokolov B.V., Zyelyk Y.I., Zelentsov V.A., Skakun S.V., Shelestov A.Y. Disaster risk assessment based on heterogeneous geospatial information. *J. of Automation and Information Sci.* 2010. Vol. 42, N 12. P. 32–45.
58. Kussul N., Lemoine G., Gallego F. J.,. Skakun S. V, Lavreniuk M., Shelestov A. Y. Parcel-based crop classification in Ukraine using Landsat-8 data and Sentinel-1A data. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens*. 2016. Vol. 9, N 6. P. 2500–2508.
59. Kussul N., Lavreniuk M., Shelestov A., Skakun S. Crop inventory at regional scale in Ukraine: developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery. *European Journal of Remote Sensing.* 2018. Vol. 51, N 1. P. 627–636.

60. Sokolov A., Rachkovskij D. Approaches to sequence similarity representation. *Information Theories and Applications.* 2005. Vol.13, N 3. P. 272–278.
61. Recchia G., Sahlgren M., Kanerva P., Jones M. Encoding sequential information in semantic space models: Comparing holographic reduced representation and random permutation. *Comput. Intell. Neurosci.* 2015. Vol. 2015. Art. 986574. P. 1–18.
62. Räsänen O.J., Saarinen J.P. Sequence prediction with sparse distributed hyperdimensional coding applied to the analysis of mobile phone use patterns. *IEEE Trans. Neural Netw. Learn.* Syst. 2016. Vol. 27, N 9. P. 1878–1889.
63. Gallant S.I., Culliton P. Positional binding with distributed representations. *Proc. ICIVC'16.* (Portsmouth, UK 3–5 Aug., 2016). Portsmouth, 2016. 2016. P. 108–113.
64. Frady E. P., Kleyko D., Sommer F. T. A theory of sequence indexing and working memory in recurrent neural networks. *Neural Comput.* 2018. Vol. 30, N. 6. P. 1449–1513.
65. Rachkovskij D.A. Some approaches to analogical mapping with structure sensitive distributed representations. *Journal of Experimental and Theoretical Artificial Intelligence.* 2004. Vol. 16, N 3. P. 125–145.
66. Slipchenko S.V., Rachkovskij D.A. Analogical mapping using similarity of binary distributed representations. *Int. J. Information Theories and Applications.* 2009. Vol. 16, N 3. P. 269–290.

Received 22.08.2018

Гриценко В.І., член-кореспондент НАН України,
директор Міжнародного науково-навчального центру
інформаційних технологій та систем НАН України та МОН України
e-mail: vig@irtc.org.ua
Рачковський Д.А., д-р техн. наук, пров. наук. співроб.
відд. нейромережевих технологій оброблення інформації,
e-mail: dar@infrm.kiev.ua
Ревунова О.Г., канд. техн. наук, старш. наук. співроб.
відд. нейромережевих технологій оброблення інформації,
e-mail: egrevunova@gmail.com
Міжнародний науково-навчальний центр інформаційних технологій
та систем НАН України та МОН України,
пр. Акад. Глушкова, 40, м. Київ, 03187, Україна

НЕЙРОМЕРЕЖЕВІ РОЗПОДІЛЕНІ ПОДАННЯ ВЕКТОРНИХ ДАНИХ У ІНТЕЛЕКТУАЛЬНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЯХ

Вступ. Розподілене подання (РП) даних — форма векторного подання, де кожний об'єкт надано безліччю компонентів вектора і кожний компонент вектора може належати поданню багатьох об'єктів. Це нейромережевий підхід, оснований на моделюванні подання інформації в мозку, до якого привели ідеї про "розподілене" або "голографічне" подання. РП мають велику інформаційну ємкість, дають змогу застосовувати багатий арсенал методів, розроблених для векторних даних, добре масштабуються для оброблення великих обсягів даних, мають низку інших переваг. Методи перетворення в РП розроблено для даних різного типу — від скалярних і векторних до графів.

Мета статті — надати опис частини робіт відділу нейромережевих технологій оброблення інформації Міжнародного Центру в галузі нейромережевого розподіленого подання. Підхід є розвитком ідей академіка М.М. Амосова і його наукової школи про моделювання структури і функцій мозку.

Результати. Розглянуто формування розподіленого подання з початкового векторного подання об'єктів за допомоги випадкового проектування. За допомоги РП можна ефективно оцінювати схожість початкових об'єктів — числових векторів, що дає змогу розробляти методи регуляризації для отримання стійкого рішення дискретних некоректних обернених задач, підвищити обчислювальну ефективність і точність їх розв'язання, аналітично досліджувати точність рішення.

Висновки. РП даних різних типів може бути використано для підвищення ефективності та рівня інтелектуальності інформаційних технологій. Розроблено РП як для слабо структурованих даних (вектори), так і для складно структурованого подання об'єктів (послідовності, графи ситуацій (епізодів) баз знань тощо). Перетворення різнотипних даних в векторний формат РП дає змогу уніфікувати базові інформаційні технології їх оброблення та домогтися надійної масштабованості зі збільшенням обсягів оброблюваних даних.

Ключові слова: *розподілене подання даних, випадкове проектування, оцінка подібності векторів, дискретна некоректна задача, регуляризація.*

Гриценко В.И., член-корреспондент НАН Украины,
директор Международного научно-учебного центра
информационных технологий и систем
НАН Украины и МОН Украины
e-mail: vig@irtc.org.ua
Рачковский Д.А., д-р техн. наук, вед. науч. сотр.
отд. нейросетевых технологий обработки информации
e-mail: dar@infrm.kiev.ua
Ревунова Е.Г., канд. техн. наук,
старш. науч. сотр.
отд. нейросетевых технологий обработки информации
e-mail: egrevunova@gmail.com
Международный научно-учебный центр информационных
технологий и систем НАН Украины и МОН Украины,
пр. Акад. Глушкова, 40, г. Киев, 03187, Украина

НЕЙРОСЕТЕВЫЕ РАСПРЕДЕЛЕННЫЕ ПРЕДСТАВЛЕНИЯ ВЕКТОРНЫХ ДАННЫХ В ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЯХ

В статье дан обзор части работ отдела нейросетевых технологий обработки информации Международного Центра в области нейросетевых распределенных представлений. Подход является развитием идей Н.М. Амосова и его научной школы о моделировании структуры и функций мозга. Распределенные представления данных различных типов могут быть использованы для повышения эффективности и уровня интеллектуальности информационных технологий. Разработаны РП как для слабо структурированных данных (векторы), так и для сложно структурированных представлений объектов (последовательности, графы ситуаций (эпизодов) баз знаний, и др.). Преобразование разнотипных данных в векторный формат РП позволяет унифицировать базовые информационные технологии их обработки и добиться масштабируемости с увеличением объемов обрабатываемых данных. В перспективе распределенные представления позволят соединить информацию о структуре и семантике для создания вычислительно эффективных и качественно новых ИТ, в которых обработка реляционных структур из баз знаний выполняется по сходству их РП.

Ключевые слова: *распределенное представление данных, случайное проецирование, оценка сходства векторов, дискретная некорректная задача, регуляризация.*