DOI 10.1007/s10559-014-9687-4
Cybernetics and Systems Analysis, Vol. 50, No. 6, November, 2014

NEW MEANS OF CYBERNETICS, INFORMATICS, COMPUTER ENGINEERING, AND SYSTEMS ANALYSIS

VECTOR DATA TRANSFORMATION USING RANDOM BINARY MATRICES

D. A. Rachkovskij

UDC 004.22 + 004.93'11

Abstract. *This article proposes to use a binary random matrix with the elements {0,1} to project input floating-point vectors onto output floating-point vectors of smaller dimension. The accuracies of estimates of the scalar product, Euclidean distance, and norm of input vectors are analyzed with respect to output vectors. It is analytically and experimentally shown that an error for the proposed random projection is smaller than that for a ternary random matrix.*

Keywords: *binary random projection, decrease in dimension, estimate for the similarity of vectors.*

INTRODUCTION

A considerable part of digital data on objects different in nature can be represented in vector form, i.e., as a collection of features, fields, and components of vectors corresponding to the objects. The number of objects and dimension of vector descriptions of complicated objects can amount to thousands, millions, and billions. The presence of such data volumes allows one to search for similar data, to analyze, classify, and detect regularities, and to solve problems on the basis of precedents and analogies [1–5]. At the same time, a large number of features in descriptions of numerous complicated objects complicate data storage, access, analysis, and understanding.

The search for similar data (examples, precedents, and analogues) is of independent importance (for example, searching for information on the Internet), and also is the first stage of reasoning based on precedents and analogies and the application of this productive approach to the analysis and understanding of data in intelligent technologies and devices. One of directions of increasing the efficiency of searching for similar vector data and processing them consists of transformation of initial representations into vectors whose dimension is smaller and whose processing would yield results concordant with the results for the initial vectors but with smaller computational expenditures.

Methods that reduce dimensions, are adaptable to data, and do not use information obtained from a teacher (such as Principal Component Analysis (PCA) and others) or use such information (Linear Discriminant Analysis (LDA) and others) [6] are computationally intensive and transform similarity-difference measures of initial vectors in estimating them from vectors of smaller dimension.

The use of random projections is a computationally more efficient approach in which vectors are produced that make it possible to estimate similarity–difference measures of initial vectors. In this case, initial vectors are transformed into a secondary vector space by multiplication by projection matrices whose elements are randomly generated numbers from some distribution. (Note that here, unlike the traditional use of this mathematical term, such a random projection matrix is not idempotent, etc.). For a number of distributions of elements of random projection matrices, output vectors allow one to estimate some similarity and differences measures for initial vectors. This is shown for projecting with the help of random matrices with the Gaussian distribution [7, 8], ternary distribution with elements from $\{-1, 0, +1\}$ and also from $\{-1, +1\}$ [9–11], stable

International Research and Training Center of Information Technologies and Systems, NAS and MON of Ukraine, Kiev, Ukraine, *dar@infrm.kiev.ua*. Translated from Kibernetika i Sistemnyi Analiz, No. 6, pp. 157–166, November–December, 2014. Original article submitted December 24, 2013.

distribution [12], etc. Random projections are also used in other problems, for example, to efficiently reconstruct sparse signals (Compressed Sensing) [13] or to stably solve a discrete ill-posed inverse problem [14].

It is obvious that the simplest and most computationally efficient version of a random projection matrix from the viewpoint of generation and use is a binary random matrix with the elements $\{0,1\}$. However, as far as we know, the use of these matrices and properties of vectors obtained with their help have not been considered yet. This article investigates properties of retention of similarity–difference measures for vectors projected with the help of such matrices.

PROJECTION WITH THE HELP OF A RANDOM BINARY MATRIX

Let us consider the projection of vectors with the help of a random binary projection matrix **R** with elements r_{ij} from the set $\{0,1\}$. The distribution of elements (ones and zeros) from **R** is independent and identical (i.i.d.). Each r_{ij} assumes the value 1 with probability q and the value 0 with probability $1-q$. We denote by **x** and **y** input vectors (of dimension D) and by $\mathbf{u}^{\mathbf{R}} = \mathbf{Rx}$ and $\mathbf{v}^{\mathbf{R}} = \mathbf{Ry}$ (of dimension d) the result of their projection. Accordingly, the dimension of **R** equals $D \times d$. The problem is to estimate the measures of similarity–difference between **x** and **y** from **u** and **v**. As well as in [10] where a ternary projection matrix with elements from $\{-1,0,+1\}$ is considered, we will estimate the value of the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle$, square of the Euclidean distance $\| \mathbf{x} - \mathbf{y} \|^2$, and square of the Euclidean norm $\| \mathbf{x} \|^2$ but for the case of projection with the help of a binary random matrix.

In projecting, each component $u_i^{\mathbf{r}}$, $i = 1, \ldots, d$, of the vector $\mathbf{u}^{\mathbf{R}}$ is formed as a result of scalar multiplication of the $\mathbf{r}_i$th row of the matrix **R** by **x**, i.e., $u_i^{\mathbf{r}} = \langle \mathbf{r}_i, \mathbf{x} \rangle = \sum_{j=1}^{D} r_{ij} x_j$. We compute the expectation $E\{u_i^{\mathbf{r}}\}$, where the averaging is performed over different realizations of rows $\mathbf{r}_i$ for the same (constant) input **x** by the formula

$$E\{u_i^{\mathbf{r}}\} = E\left\{ \sum_{j=1}^{D} r_{ij} x_j \right\} = \sum_{j=1}^{D} x_j E\{r_{ij}\} = q \sum_{j=1}^{D} x_j$$

since $E\{r_{ij}\} = 1q + 0(1-q) = q$.

We define a centered random quantity u_i with $Eu_i = 0$ as follows: $u_i = u_i^{\mathbf{r}} - E\{u_i^{\mathbf{r}}\} = u_i^{\mathbf{r}} - q\sum_{j=1}^{D} x_j$. We can represent u_i in the form

$$u_i = \sum_{j=1}^{D} r_{ij} x_j - q \sum_{j=1}^{D} x_j = \sum_{j=1}^{D} (r_{ij} - q) x_j .$$

Thus, the same result $\mathbf{u}(\mathbf{v})$ will also be obtained for projecting $\mathbf{u} = \mathbf{Px}$ $(\mathbf{v} = \mathbf{Py})$ with the help of the centered random matrix $\mathbf{P} = \mathbf{R} - q$ with elements $\rho_{ij} = r_{ij} - q$, $E\{\rho_{ij}\} = E\{r_{ij} - q\} = (1-q)q + (0-q)(1-q) = 0$. The analysis of projection with the help of the centered **P** is simpler than with the help of **R**, and the results are identical if $u_i^{\mathbf{r}}$ is transformed into the form $u_i^{\mathbf{r}} - q\sum_{j=1}^{D} x_j \equiv u_i$ and then u_i is used. Therefore, we will analyze the projection of the centered **P**. Since ρ_{ij} are i.i.d., we will use ρ everywhere where it is pertinent, instead of ρ_{ij} for compactness. Since $u_i = \sum_{j=1}^{D} \rho_{ij} x_j$ also are i.i.d., we will sometimes consider $u = \sum_{j=1}^{D} \rho_j x_j$, where ρ_j are elements of a row that belongs to **P** and is multiplied by the input vector. We introduce random variables $\xi_j \equiv x_j \rho_j$ and $\zeta_j \equiv y_j \rho_j$. For them, we have $E\{\xi_j\} = x_j E\{\rho_j\} = 0$; similarly, $E\{\zeta_j\} = 0$. When $j \neq k$, the variables ξ_j and ζ_k are independent.

Scalar product. Let us find the expectation value (e.v.) and variance of the scalar product $\langle \mathbf{u}, \mathbf{v} \rangle$,

$$E\{\langle \mathbf{u}, \mathbf{v} \rangle\} = \sum_{i=1}^{d} E\{u_i v_i\}.$$

Since addends $u_i v_i$ are independent random quantities (r.q.) for different i, the variance

$$V\{\langle \mathbf{u}, \mathbf{v} \rangle\} = \sum_{i=1}^{d} V\{u_i v_i\}.$$

Let us consider $uv \equiv u_i v_i = \sum_{j=1}^{D} \xi_j \zeta_j + \sum_{j \neq k} \xi_j \zeta_k$. The expectation value

$$E\{uv\} = E\{\sum_{j=1}^{D} \xi_j \zeta_j\} = \sum_{j=1}^{D} x_j y_j E\{\rho_j^2\} = E\{\rho^2\} \langle \mathbf{x}, \mathbf{y} \rangle = q(1-q) \langle \mathbf{x}, \mathbf{y} \rangle$$

since, by virtue of independence, $\sum_{j\neq k} E\{\xi_j \zeta_k\} = \sum_{j\neq k} E\{\xi_j\} E\{\zeta_k\} = 0$ and

$$E\{\rho^2\} = (1-q)^2 q + (0-q)^2 (1-q) = q - q^2.$$

Thus, we have

$$E\{\langle \mathbf{u}, \mathbf{v}\rangle\} = \sum_{i=1}^{d} E\{u_i v_i\} = q(1-q)\langle \mathbf{x}, \mathbf{y}\rangle d. \tag{1}$$

Therefore, the estimate of $\langle \mathbf{x}, \mathbf{y}\rangle$ is found from the estimate $E^*\langle \mathbf{u}, \mathbf{v}\rangle$ as $E^*\langle \mathbf{u}, \mathbf{v}\rangle / (q(1-q)d)$.

We find the variance $V\{uv\} = E\{(uv)^2\} - E^2\{uv\}$ as follows:

$$(uv)^2 = \left(\sum_{j=1}^{D} \xi_j \zeta_j + \sum_{j\neq k} \xi_j \zeta_k\right)^2$$

$$= \left(\sum_{j=1}^{D} \xi_j \zeta_j\right)^2 + 2\left(\sum_{j=1}^{D} \xi_j \zeta_j\right)\left(\sum_{j\neq k} \xi_j \zeta_k\right) + \left(\sum_{j\neq k} \xi_j \zeta_k\right)^2,$$

$$\left(\sum_{j=1}^{D} \xi_j \zeta_j\right)^2 = \sum_{j=1}^{D} \xi_j^2 \zeta_j^2 + \sum_{j\neq k} \xi_j \zeta_j \xi_k \zeta_k,$$

$$\left(\sum_{j\neq k} \xi_j \zeta_k\right)^2 = \sum_{j\neq k} \xi_j^2 \zeta_k^2 + \sum_{j\neq k} \xi_j \zeta_j \xi_k \zeta_k.$$

We find $E\{(uv)^2\}$. Since $2\left(\sum_{j=1}^{D} \xi_j \zeta_j\right)\left(\sum_{j\neq k} \xi_j \zeta_k\right)$ contains a multiplicand in the form of an independent r.q. ζ_k $(k \neq j)$ with $E\{\zeta_k\} = 0$, we have $E\left\{2\left(\sum_{j=1}^{D} \xi_j \zeta_j\right)\left(\sum_{j\neq k} \xi_j \zeta_k\right)\right\} = 0$. Thus, we have

$$E\{(uv)^2\} = \sum_{j=1}^{D} E\{\xi_j^2 \zeta_j^2\} + \sum_{j\neq k} E\{\xi_j^2 \zeta_k^2\} + 2\sum_{j\neq k} E\{\xi_j \zeta_j \xi_k \zeta_k\}$$

$$= E\{\rho^4\} \sum_{j=1}^{D} x_j^2 y_j^2 + E^2\{\rho^2\} \sum_{j\neq k} x_j^2 y_k^2 + 2E^2\{\rho^2\} \sum_{j\neq k} x_j y_j x_k y_k.$$

We obtain

$$E^2\{\rho^2\}\left((E\{\rho^4\}/E^2\{\rho^2\}) \sum_{j=1}^{D} x_j^2 y_j^2 + \sum_{j\neq k} x_j^2 y_k^2 + 2\sum_{j\neq k} x_j y_j x_k y_k\right)$$

$$= E^2\{\rho^2\}\left((E\{\rho^4\}/E^2\{\rho^2\} - 3) \sum_{j=1}^{D} x_j^2 y_j^2 + 2\langle \mathbf{x}, \mathbf{y}\rangle^2 + \|\mathbf{x}\|^2 \|\mathbf{y}\|^2\right)$$

since

$$\|\mathbf{x}\|^2 \|\mathbf{y}\|^2 = \sum_{j=1}^{D} x_j^2 \sum_{j=1}^{D} y_j^2 = \sum_{j=1}^{D} x_j^2 y_j^2 + \sum_{j\neq k}^{D} x_j^2 y_k^2,$$

$$\langle \mathbf{x}, \mathbf{y}\rangle^2 = \left(\sum_{j=1}^{D} x_j y_j\right)^2 = \sum_{j=1}^{D} x_j^2 y_j^2 + \sum_{j\neq k}^{D} x_j y_j x_k y_k.$$

Thus, we have

$$V\{uv\} = E\{(uv)^2\} - E^2\{uv\} = E^2\{\rho^2\}\left((E\{\rho^4\}/E^2\{\rho^2\} - 3) \sum_{j=1}^{D} x_j^2 y_j^2 + \langle \mathbf{x}, \mathbf{y}\rangle^2 + \|\mathbf{x}\|^2 \|\mathbf{y}\|^2\right),$$

$$V\{\langle \mathbf{u}, \mathbf{v}\rangle\} = \sum_{i=1}^{d} V\{u_j v_j\} = V\{uv\} d. \tag{2}$$

Euclidean distance. Let us find the e.v. and variance of the square of the Euclidean distance $\|\mathbf{u}-\mathbf{v}\|^2$,

$$E\{\|\mathbf{u}-\mathbf{v}\|^2\}=\sum_{i=1}^{d}E\{(u_i-v_i)^2\},\ V\{\|\mathbf{u}-\mathbf{v}\|^2\}=\sum_{i=1}^{d}V\{(u_i-v_i)^2\},$$

$$(u-v)^2=\left(\sum_{j=1}^{D}(\xi_j-\zeta_j)\right)^2=\sum_{j=1}^{D}(\xi_j-\zeta_j)^2+\sum_{j\neq k}(\xi_j-\zeta_j)(\xi_k-\zeta_k),$$

$$E\{(u-v)^2\}=\sum_{j=1}^{D}E\{(\xi_j-\zeta_j)^2\}+\sum_{j\neq k}E\{(\xi_j-\zeta_j)(\xi_k-\zeta_k)\}.$$

Since $\xi_j-\zeta_j$ and $\xi_k-\zeta_k$ are independent when $j\neq k$ and $E\{(\xi_j-\zeta_j)\}=0$, we have

$$\sum_{j\neq k}E\{(\xi_j-\zeta_j)(\xi_k-\zeta_k)\}=\sum_{j\neq k}E\{(\xi_j-\zeta_j)\}E\{(\xi_k-\zeta_k)\}=0.$$

Therefore,

$$E\{(u-v)^2\}=\sum_{j=1}^{D}E\{(\xi_j-\zeta_j)^2\}=\sum_{j=1}^{D}E\{\rho_j^2\}(x_j-y_j)^2=(q-q^2)\|\mathbf{x}-\mathbf{y}\|^2,$$

$$E\{\|\mathbf{u}-\mathbf{v}\|^2\}=(q-q^2)\|\mathbf{x}-\mathbf{y}\|^2\,d. \tag{3}$$

We determine the variance $V\{(u-v)^2\}=E\{(u-v)^4\}-E^2\{(u-v)^2\}$ as follows:

$$(u-v)^4=\left(\sum_{j=1}^{D}(\xi_j-\zeta_j)^2+\sum_{j\neq k}(\xi_j-\zeta_j)(\xi_k-\zeta_k)\right)^2$$

$$=\left(\sum_{j=1}^{D}(\xi_j-\zeta_j)^2\right)^2+2\left(\sum_{j=1}^{D}(\xi_j-\zeta_j)^2\right)\left(\sum_{j\neq k}(\xi_j-\zeta_j)(\xi_k-\zeta_k)\right)$$

$$+\left(\sum_{j\neq k}(\xi_j-\zeta_j)(\xi_k-\zeta_k)\right)^2,$$

$$\left(\sum_{j=1}^{D}(\xi_j-\zeta_j)^2\right)^2=\sum_{j=1}^{D}(\xi_j-\zeta_j)^4+\sum_{j\neq k}(\xi_j-\zeta_j)^2(\xi_k-\zeta_k)^2,$$

$$\left(\sum_{j\neq k}(\xi_j-\zeta_j)(\xi_k-\zeta_k)\right)^2=2\sum_{j\neq k}(\xi_j-\zeta_j)^2(\xi_k-\zeta_k)^2+\sum\ldots.$$

Let us find $E\{(u-v)^4\}$. When $D\geq 2$, the expression

$$2\left(\sum_{j=1}^{D}(\xi_j-\zeta_j)^2\right)\left(\sum_{j\neq k}(\xi_j-\zeta_j)(\xi_k-\zeta_k)\right)$$

contains a multiplicand in each addend in the form of the independent r.q. $(\xi_k-\zeta_k)$ $(k\neq j)$ and, hence,

$$E\left\{2\left(\sum_{j=1}^{D}(\xi_j-\zeta_j)^2\right)\left(\sum_{j\neq k}(\xi_j-\zeta_j)(\xi_k-\zeta_k)\right)\right\}=0$$

since each addend includes the multiplicand $E\{(\xi_k-\zeta_k)\}=E\{\xi_k\}-E\{\zeta_k\}=0$. Similarly, we have $E\{\sum\ldots\}=0$. Thus, we obtain

$$E\{(u-v)^4\}=\sum_{j=1}^{D}E\{(\xi_j-\zeta_j)^4\}+3\sum_{j\neq k}E\{(\xi_j-\zeta_j)^2(\xi_k-\zeta_k)^2\}$$

$$= E\{\rho^4\}\sum_{j=1}^{D}(x_j - y_j)^4 + 3E^2\{\rho^2\}\sum_{j\neq k}(x_j - y_j)^2(x_k - y_k)^2$$

$$= E^2\{\rho^2\}\left((E\{\rho^4\}/E^2\{\rho^2\}-3)\sum_{j=1}^{D}(x_j - y_j)^4 + 3\|\mathbf{x}-\mathbf{y}\|^4\right),$$

since

$$\|\mathbf{x}-\mathbf{y}\|^4 = (\|\mathbf{x}-\mathbf{y}\|^2)^2 = \left(\sum_{j=1}^{D}(x_j - y_j)^2\right)^2$$

$$= \sum_{j=1}^{D}(x_j - y_j)^4 + \sum_{j\neq k}(x_j - y_j)^2(x_k - y_k)^2.$$

We obtain

$$V\{(u-v)^2\} = E\{(u-v)^4\} - E^2\{(u-v)^2\}$$

$$= E^2\{\rho^2\}\left((E\{\rho^4\}/E^2\{\rho^2\}-3)\sum_{j=1}^{D}(x_j - y_j)^4 + 2\|\mathbf{x}-\mathbf{y}\|^4\right),$$

$$V\{\|\mathbf{u}-\mathbf{v}\|^2\} = \sum_{i=1}^{d}V\{(u_i - v_i)^2\} = V\{(u-v)^2\}d. \tag{4}$$

Analysis of estimates. The expressions for the e.v. and variance of the square of the Euclidean norm of a vector are obtained from formulas (1), (2) or (3), (4) as $\langle \mathbf{u},\mathbf{u}\rangle$ or $\|\mathbf{u}-0\|^2$.

To compare errors of estimates of the scalar product, Euclidean distance, and norm of the initial vectors with respect to vectors after a random projection, the following relative standard deviation (variation coefficient) $V^{1/2}/E$ was used:

$$V^{1/2}\{\langle\mathbf{u},\mathbf{v}\rangle\}/E\{\langle\mathbf{u},\mathbf{v}\rangle\} = \frac{1}{\langle\mathbf{x},\mathbf{y}\rangle\sqrt{d}}\left((E\{\rho^4\}/E^2\{\rho^2\}-3)\sum_{j=1}^{D}x_j^2y_j^2 + \langle\mathbf{x},\mathbf{y}\rangle^2 + \|\mathbf{x}\|^2\,\|\mathbf{y}\|^2\right)^{1/2},$$

$$V^{1/2}\{\|\mathbf{u}-\mathbf{v}\|^2\}/E\{\|\mathbf{u}-\mathbf{v}\|^2\}$$

$$= \frac{1}{\|\mathbf{x}-\mathbf{y}\|^2\sqrt{d}}\left((E\{\rho^4\}/E^2\{\rho^2\}-3)\sum_{j=1}^{D}(x_j - y_j)^4 + 2\|\mathbf{x}-\mathbf{y}\|^4\right)^{1/2}. \tag{5}$$

For the binary random projection matrix being considered, we have

$$E\{\rho^4\} = (1-q)^4q + (0-q)^4(1-q) = (q-q^2)(1-3(q-q^2)),$$

$$E\{\rho^4\}/E^2\{\rho^2\} = (q-q^2)(1-3(q-q^2))/(q-q^2)^2 = 1/(q-q^2)-3.$$

For a ternary random projection matrix with elements $-1/\sqrt{q}$ (with probability $q/2$), $+1/\sqrt{q}$ (with probability $q/2$), and 0 (with probability $1-q$), we have

$$E\{\rho^2\} = (1/\sqrt{q})^2q/2 + (-1/\sqrt{q})^2q/2 = 1,$$

$$E\{\rho^4\} = (1/\sqrt{q})^4q/2 + (-1/\sqrt{q})^4q/2 = 1/q,\ E\{\rho^4\}/E^2\{\rho^2\} = 1/q$$

(see also [10]). When $q < 2/3$, we have $1/(q-q^2)-3 < 1/q$ and, hence, the error of estimates for the proposed binary random projections is smaller than for ternary ones with the same probability q of a nonzero matrix element.

For a random projection matrix with elements from the Gaussian distribution [8, 10], we have

$$V^{1/2}\{\langle\mathbf{u},\mathbf{v}\rangle\}/E\{\langle\mathbf{u},\mathbf{v}\rangle\} = \frac{1}{\langle\mathbf{x},\mathbf{y}\rangle\sqrt{d}}(\langle\mathbf{x},\mathbf{y}\rangle^2 + \|\mathbf{x}\|^2\,\|\mathbf{y}\|^2)^{1/2},\ V^{1/2}\{\|\mathbf{u}-\mathbf{v}\|^2\}/E\{\|\mathbf{u}-\mathbf{v}\|^2\} = \sqrt{\frac{2}{d}}. \tag{6}$$

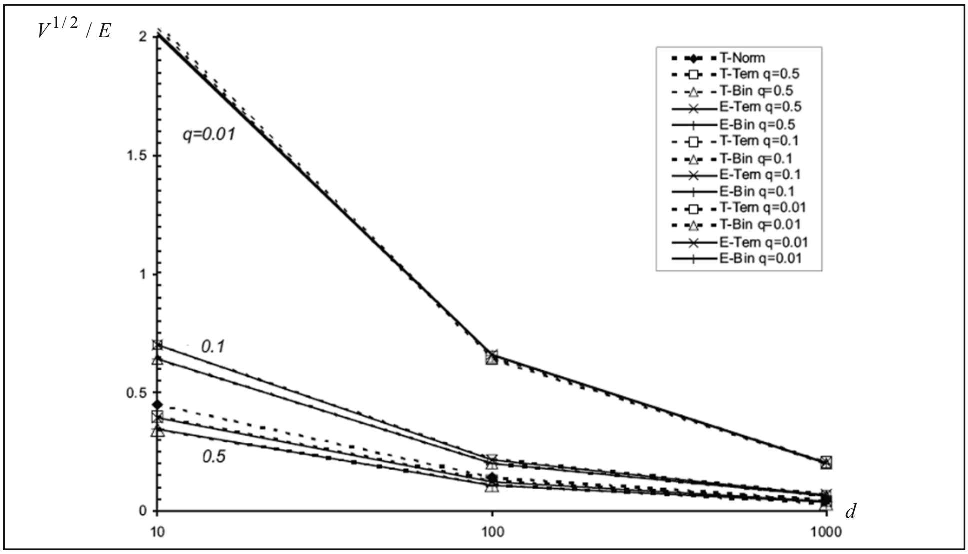

Fig. 1. Dependence of the error $V^{1/2}/E$ of the estimate for the square of the Euclidean distance between input vectors on the dimension d of vectors after projection. The dimension of input vectors equaled $D = 10$.

Comparing formulas (5) and (6), we can see that since $1/(q - q^2) - 3 - 3 < 0$ when $q \approx [0.2113;\ 0.7887]$ (i.e., when $1/2 - 1/(2\sqrt{3}) < q < 1/2 + 1/(2\sqrt{3})$), over this range, the accuracy provided by binary random projections exceeds that of Gaussian ones (the best result is reached when $q = 0.5$). On the other hand, to accelerate projecting, it is necessary to have $q << 0.5$ but, in this case, binary projections are not very efficient because of the presence of addends in an error with positive coefficients multiplying $\sum_{j=1}^{D} x_j^2 y_j^2$ and $\sum_{j=1}^{D} (x_j - y_j)^4$. However, when $D >> 1$, their contribution is small (for data with finite fourth moment) and, therefore, we also obtain an accuracy comparable with the accuracy of Gaussian random projections in the case when $q << 0.5$.

EXPERIMENTAL INVESTIGATION

The behavior of the error $V^{1/2}/E$ was investigated for the squares of the Euclidean norm and distance and also for the scalar product of vectors in projecting with the help of random binary (the elements $\{0, 1\}$) and ternary (the elements $\{-1,\ 0,\ +1\}$) matrices with different parameters. The experimentally obtained errors (computed from sample averages and variances) were compared with analytical expressions and included the error for Gaussian random projections.

Matrices were investigated for $q = \{0.5, 0.1, 0.01\}$ (for binary matrices, q was the probability of 1, and, for ternary matrices, the probability of 1 and -1 was the same and equaled $q/2$). Input vectors of length $D = \{10, 100, 1000, 10000\}$ were used. Their components were generated from the uniform distribution over $[-D, +D]$, and similarity was varied by concatenating different fragments of identical and different vectors. Note that the small $D = \{10, 100\}$ are usually uninteresting from the viewpoint of practical applications and were investigated to illustrate distinctive features of error behavior and to confirm the correctness of analytical expressions. Output vectors of dimension $d = \{10, 100, 1000\}$ were obtained. For each investigated set of projection parameters, experimental V and E are obtained by averaging over 10000 realizations of output vectors corresponding to 10000 realizations of random matrices (1000 realizations for both $D = 10000$ and $d = 1000$).

The results of all experiments for all parameter values (and for different values of similarity between input vectors) are close to the theoretical results. Figure 1 presents the dependences of the error of the estimate for the Euclidean distance between (two fixed) input vectors of length $D = 10$ on the dimension of vectors that were obtained after projection and from which this distance was estimated. Here, the lines whose denotations begin with T correspond to analytical expressions and with E correspond to the experimentally obtained results, Norm is a Gaussian projection matrix, Tern is a ternary matrix, Bin is the proposed binary matrix, and q is the probability of nonzero elements in a matrix.

TABLE 1

Types of Random Projections	Value of $V^{1/2}/E$		
	d =10	d =100	d =1000
T-Norm	0.447214	0.141421	0.044721
T-Tern q =0.5	0.447173	0.141409	0.044717
T-Bin q =0.5	0.447133	0.141396	0.044713
E-Tern q =0.5	0.442526	0.140829	0.04454
E-Bin q =0.5	0.448362	0.142213	0.04466
T-Tern q =0.1	0.447495	0.14151	0.044749
T-Bin q =0.1	0.447419	0.141486	0.044742
E-Tern q =0.1	0.441405	0.141205	0.044144
E-Bin q =0.1	0.446487	0.14008	0.044195
T-Tern q =0.01	0.451096	0.142649	0.04511
T-Bin q =0.01	0.451017	0.142624	0.045102
E-Tern q =0.01	0.452649	0.142631	0.044639
E-Bin q =0.01	0.44236	0.141008	0.045312

Errors considerably differ for different values of q for the same d. For $q = 0.5$, the error for a Gaussian projection matrix is higher than for ternary and binary ones, and, for $q = 0.01$, it is considerably lower. For $D = 100$, distinctions between values of errors decrease and, for $D = 1000$, they become insignificant. For $D = 10000$, distinctions between values of errors for Gaussian, ternary and binary random projection matrices are very small (Table 1). For all investigated q, the error for binary random projections is smaller than for ternary ones, which corresponds to the analytical estimates (distinctions are insignificant for the investigated parameters). Similar results are obtained for the scalar product and square of the norm of vectors.

In comparison with a Gaussian projection matrix, memory required for a binary matrix is smaller by a factor of 32–64 in using 1 bit per a matrix element instead of 32–64 bits for the representation of Gaussian random floating-point quantities.

CONCLUSIONS

To transform floating-point input vectors into floating-point output vectors, it is proposed to use the projection with the help of a random matrix with the binary elements $\{0, +1\}$. The output vectors allow one to estimate the similarity–difference measures (the Euclidean distance, scalar product, and also Euclidean norm) for the initial vectors; the computational efficiency of estimates increases with decreasing the dimension of output vectors.

The error of estimation of similarity–difference measures is analytically and experimentally investigated. As well as for other types of random projection matrices, this error decreases with increasing the dimension d of output vectors ($\sim 1/\sqrt{d}$). When d is fixed and the probability q is for the neighborhood of 0.5, in a binary projection matrix (the so-called "dense" binary matrices), the error of the estimate for similarity–difference measures is smaller than for a Gaussian random projection matrix. In this case, the generation of binary random quantities and the multiplication operation for the realization of a projection can be performed more efficiently than for Gaussian random quantities as a result of the absence of the need for multiplying floating-point numbers. The computational efficiency can be even higher with decreasing q (with increasing the "sparseness" of a binary projection matrix). In this case, to retain the accuracy of estimates, the dimension of input vectors must be rather large, which is what we have assumed in stating the problem of efficient estimation of the similarity of multidimensional vectors.

In comparison with a random matrix with ternary elements, a binary matrix gives a smaller error and, at the same time, provides a simpler generation of a random matrix and realization of a projection.

The investigation of questions of efficient realization of projections and also the applicability of binary projection matrices is promising when binary output vectors are obtained (by analogy with the investigation [11] for a ternary matrix). Such vectors reflecting similarity are examples of neural network-based randomized distributed representations that can be not only used for efficient similarity search [15–18] but also can be stored and processed in associative neural networks [19–29] and can also be elements of representations of hierarchically structured models of complicated objects (different in nature) of the outside world [30–37].

REFERENCES

1. V. I. Gritsenko, D. A. Rachkovskij, A. D. Goltsev, V. V. Lukovych, I. S. Misuno, E. G. Revunova, S. V. Slipchenko, and A. M. Sokolov, “Neural distributed representation for intelligent information technologies and modeling of thinking,” Cybernetics and Computer Engineering, **173**, 7–24 (2013).
2. D. A. Rachkovskij, “Representation and processing of structures with binary sparse distributed codes,” IEEE Trans. on Knowledge and Data Engineering, **13**, No. 2, 261–276 (2001).
3. D. A. Rachkovskij, “Some approaches to analogical mapping with structure sensitive distributed representations,” J. Experimental and Theoretical Artificial Intelligence, **16**, No. 3, 125–145 (2004).
4. S. V. Slipchenko and D. A. Rachkovskij, “Analogical mapping using similarity of binary distributed representations,” International J. Information Theories and Applications, **16**, No. 3 269–290 (2009).
5. D. A. Rachkovskij and S. V. Slipchenko, “Similarity-based retrieval with structure-sensitive sparse binary distributed representations,” Computational Intelligence, **28**, No. 1, 106–129 (2012).
6. C. J. C. Burges, “Dimension reduction: A guided tour,” Foundations and Trends in Machine Learning, **2**, No. 4, 275–365 (2010).
7. P. Indyk and R. Motwani, “Approximate nearest neighbors: Towards removing the curse of dimensionality,” in: Proc. 30th ACM Symposium on Theory of Computing (1998), pp. 604–613.
8. S. S. Vempala, The Random Projection Method, American Mathematical Society, Providence, RI (2004).
9. D. Achlioptas, “Database-friendly random projections: Johnson–Lindenstrauss with binary coins,” J. Computer and System Sciences, **66**, No. 4, 671–687 (2003).
10. P. Li, T. J. Hastie, and K. W. Church, “Very sparse random projections,” in: Proc. 12th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining, ACM Press, Philadelphia, PA, USA (2006), pp. 287–296.
11. D. A. Rachkovskij, I. S. Misuno, and S. V. Slipchenko, “Randomized projective methods for construction of binary sparse upsilon vector representations,” Cybernetics and Systems Analysis, **48**, No. 1, 140–150 (2012).
12. P. Li, “Very sparse stable random projections for dimension reduction in $l_\alpha\,(0<\alpha\le 2)$ norm,” in: Proc. 13th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining, ACM Press, San Jose, CA, USA (2007), pp. 440–449.
13. D. L. Donoho, “Compressed sensing,” IEEE Trans. Inform. Theory, **52**, No. 4, 1289–1306 (2006).
14. D. A. Rachkovskij and E. G. Revunova, “Randomized method for solving discrete ill-posed problems,” Cybernetics and Systems Analysis, **48**, No. 4, 621–635 (2012).
15. P. Kanerva, G. Sjodin, J. Kristoferson, R. Karlsson, B. Levin, A. Holst, J. Karlgren, and M. Sahlgren, “Computing with large random patterns,” in: Foundations of Real-World Intelligence, CSLI Publications, Stanford, California (2001), pp. 251–311.
16. I. S. Misuno, D. A. Rachkovskij, and S. V. Slipchenko, “Vector and distributed representations reflecting semantic relatedness of words,” Math. Machines and Systems, No. 3, 50–67 (2005).
17. A. Sokolov, “LIMSI: learning semantic similarity by selecting random word subsets,” in: Proc. 6th Intern. Workshop on Semantic Evaluation (SEMEVAL’12), Association for Computational Linguistics (2012), pp. 543–546.
18. A. Sokolov and S. Riezler, “Task-driven greedy learning of feature hashing functions,” in: Proc. NIPS of 13 Workshop “Big Learning: Advances in Algorithms and Data Management,” Lake Tahoe, USA (2013), pp. 1–5.
19. A. Frolov, A. Kartashov, A. Goltsev, and R. Folk, “Quality and efficiency of retrieval for Willshaw-like autoassociative networks. I. Correction,” Network: Computation in Neural Systems, **6**, No. 4, 513–534 (1995).
20. A. Frolov, A. Kartashov, A. Goltsev, and R. Folk, “Quality and efficiency of retrieval for Willshaw-like autoassociative networks. II. Recognition,” Network: Computation in Neural Systems, **6**, No. 4, 535–549 (1995).
21. A. A. Frolov, D. Husek, and I. P. Muraviev, “Informational capacity and recall quality in sparsely encoded Hopfield-like neural network: Analytical approaches and computer simulation,” Neural Networks, **10**, No. 5, 845–855 (1997).
22. A. A. Frolov, D. Husek, and P. Yu. Polyakov, “Recurrent-neural-network-based Boolean factor analysis and its application to word clustering,” IEEE Trans. on Neural Networks, 20, No. 7, 1073–1086 (2009).

23. A. A. Frolov, D. A. Rachkovskij, and D. Husek, "On informational characteristics of sparsely encoded binary auto-associative memory," in: Proc. 9th Intern. Conf. on Neural Information Processing ICONIP'02, Orchid Country Club, Singapore (2002), pp. 235–238.
24. A. A. Frolov, D. A. Rachkovskij, and D. Husek, "On information characteristics of Willshaw-like auto-associative memory," Neural Network World, **12**, No. 2, 141–158 (2002).
25. A. A. Frolov, D. Husek, and D. A. Rachkovskij, "Time of searching for similar binary vectors in associative memory," Cybernetics and Systems Analysis, **42**, No. 5, 615–623 (2006).
26. D. W. Nowicki and O. K. Dekhtyarenko, "Averaging on Riemannian manifolds and unsupervised learning using neural associative memory," in: Proc. ESANN 2005, Bruges, Belgium (2005), pp. 181–189.
27. D. Nowicki and H. Siegelmann, "Flexible kernel memory," PLoS ONE, **5**, No. 6 (2010): e10955. doi:10.1371/journal.pone.0010955
28. D. Nowicki, P. Verga, and H. Siegelmann, "Modeling reconsolidation in kernel associative memory," PloS ONE, **8**, No. 8 (2013): e68189. doi:10.1371/journal.pone.0068189
29. B. Emruli, R. W. Gayler, and F. Sandin, "Analogical mapping and inference with binary spatter codes and sparse distributed memory," in: Proc. Intern. Joint Conf. On Neural Networks (IJCNN) (2013), pp. 1–8.
30. E. M. Kussul and D. A. Rachkovskij, "Multilevel assembly neural architecture and processing of sequences," in: A. V. Holden and V. I. Kryukov (eds.), Neurocomputers and Attention, Vol. II, Connectionism and Neurocomputers, Manchester University Press, Manchester–New York (1991), pp. 577–590.
31. N. M. Amosov, T. N. Baidyk, A. D. Goltsev, A. M. Kasatkin, L. M. Kasatkina, E. M. Kussul, and D. A. Rachkovskij, Neurocomputers and Intelligent Robots [in Russian], Naukova Dumka, Kyiv (1991).
32. D. A. Rachkovskij, S. V. Slipchenko, E. M. Kussul, and T. N. Baidyk, "A binding procedure for distributed binary data representations," Cybernetics and Systems Analysis, **41**, No. 3, 319–331 (2005).
33. A. Letichevsky, A. Godlevsky, A. Letichevsky Jr, V. S. Potiyenko, and V. Peschanenko, "Properties of a predicate transformer of the VRS system," Cybernetics and Systems Analysis, **46**, No. 4, 521–532 (2010).
34. A. Letichevsky, A. Letychevsky Jr, and V. Peschanenko, "Insertion modeling system," Lecture Notes in Comput. Sci., **7162**, 262–274 (2011).
35. S. I. Gallant and T. W. Okaywe, "Representing objects, relations, and sequences," Neural Computation, **25**, No. 8, 2038–2078 (2013).
36. D. A. Rachkovskij, E. M. Kussul, and T. N. Baidyk, "Building a world model with structure-sensitive sparse binary distributed representations," Biologically Inspired Cognitive Architectures, **3**, 64–86 (2013).
37. A. A. Letichevsky, "Theory of interaction, insertion modeling, and cognitive architectures," Biologically Inspired Cognitive Architectures, **8**, 19–32 (2014).