Analysis method of the information value of indicators

Y Artamonov¹ and I Kamanin²

¹Federal State budget scientific institution "State scientific-methodological center", 51
Lyusinovskaya Street, Moscow 115998, Russian Federation
²Department of Economics of urban and housing rights, Moscow Metropolitan
Governance University, 28 Sretenka Street, Moscow 109045, Russian Federation

¹E-mail: junaart@mail.ru

Abstract. A new method for assessing the information significance of a group of indicators is
proposed based on minimizing the matrix norm of deviations of the evaluation matrix from the
matrix of the most informative indicators. The proposed method allows to combine the classes
of tasks of autoinformativity and external informative characteristics. Comparison of the
method with existing methods makes it possible to assert that in the proposed approach the
information loss of the initial evaluation matrix is minimized. The analysis of regression
dependencies on various subsets of indicators made it possible to conclude that in the
developed method the construction of the most accurate linear regression dependencies is
possible on the selected subset of informative indicators.

1. Introduction
A wide range of management tasks in education requires the selection of the most informative
indicators. At the present day, the indicative method is widely used for managerial decision-making:
the group of the most important in some sense indicators is selected, on which threshold values are
being set. Then the objects of analysis are evaluated in a variety of ways according to the values of the
selected indicators. The assessment of performance of universities, scientific organizations is
conducted this way. A similar methodology is used to evaluate scientific and technical projects under
various state programs in the field of science and education.

The task of determining the most informative indicators is similar to the problem of dimension
reduction. However, in this case, it is required to select from the wide range of indicators the most
important indicators, and not to suggest various transformations of the initial indicators, for example,
in the form of linear convolution (as in the principal component method). In general, the scientific
approaches used of dimension reduction can be divided into two classes:

- class of methods based on the autoinformativity criterion aimed at maximum preservation of the
  information contained in the original data array for which each indicator is estimated by the level of
  influence on the overall variability of the data;
- class of methods based on external information, aimed at maximum "extrusion" the information
  contained in the source array with respect to some other (external) indicators.

The first class of methods is based on various matrix decompositions. The key method in this case
is principal component analysis (PCA) and its improvements [1-3]. An important and one of the first
results in this direction was the theorem of C. Eckart, G. Young [4], in which it is shown that for a
fixed rank of an approximation matrix the minimum of the Frobenius norm ensures the corresponding
singular-value decomposition of the original matrix. Recently, methods of approximation of the initial
evaluation matrix by matrices of smaller ranks have become increasingly popular: non-negative matrix factorization (NMF), and various tensor decompositions options as well [5-8]. However, it is possible to use these methods for the task of selecting the most informative indicators only after an appropriate interpretation, which is often ambiguous. For example, for PCA as measure of the importance of the indicator factor loading of corresponding component is traditionally used.

The second class of methods is traditionally used to determine the most significant factor characteristics that are subsequently used to obtain predictive estimates of other external characteristics or to classify the objects of analysis. Here are traditionally used: canonical correlation analysis (CCA) [9, 10], discriminant analysis (DA) [11, 12], and regression analysis [13, 14]. In a number of cases, direct estimates of the informative value of the indicators are proposed on a basis of a posteriori data [15].

The array of the examined approaches is rather heterogeneous. Often for the same task of determining the most informative attributes, different methods provide different results. At the same time, the question remains open: whether the drift of results is conditioned by the imperfection of methods or whether it reflects the specific features of a particular problem. The proposed research considers a generalization of a number of approximation methods of the initial evaluation matrix with the interpretation of the drift of the results.

2. Analysis method of the information significance of indicators based on the approximation of the evaluation matrix

The task of analyzing the informational significance of the indicators is as follows: let there be given a system (set) of indicators $P = \{p_1, p_2, \ldots, p_n\}$, characterizing the analysis objects $O = \{o_1, o_2, \ldots, o_m\}$. Without loss of generality, we can consider all indicators of the set to be unidirectional in improving the set of objects. This means that the trajectory of the development of the system of objects leads to progress (improvement) if at the time $t_1 > t_0$ the system of inequalities $\forall i = 1, m, j = 1, n: p_{ij}(t_0) \leq p_{ij}(t_1)$ is fulfilled, and at least one indicator shows a strict inequality. It is required to choose some minimal subset $B \subset P$ of the most significant indicators in some sense preserving or allowing to recover the greatest amount of initial information from the system of parameters $P$.

Let the information about the objects of analysis $O$, characterized by the set of indicators $P$, be represented by the evaluation matrix:

$$A = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,1} & p_{m,2} & \cdots & p_{m,n} \end{pmatrix}$$  \hspace{1cm} (1)

In the research [16] the author proposed a new approach to the solution of the problem of group selection, based on the use of different matrix norms. This approach can be extended to the task of selecting the most informative indicators.

For this, first of all, let us perform the normalization of the matrix $A$. Likewise [16] rank normalization will be examined by replacing all indicator values $p_{ij}$ with their ranks. Thus we proceed from matrix (1) to matrix $A'$:

$$A' = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \cdots & r_{m,n} \end{pmatrix}$$  \hspace{1cm} (2)

Let us introduce the matrix of a given rank $B$, approximating the matrix $A'$. In order to do this, we proceed from the general idea that with the help of the most informative indicators we can perform a more accurate approximation. If the matrix $B$ has rank $1$, then, as an approximation, we consider the matrix $B$, all columns of which are identical and equal to a column of the $A'$. Let us denote such a matrix by $B^{(i)}$, $i = 1, m$, where $i$ – is the selected column of the matrix $A'$.
To select this column, we consider a function $f(A', B^{(i)})$, or which we obtain a minimum of a matrix norm:

$$
\text{min}\{\|f(A', B^{(i)})\|: i = 1,n\}
$$

(3)

As a matrix norm further we shall consider element-wise matrix norm $L_p$, and as a function $f(A', B^{(i)})$ we shall use matrices difference function $f(A', B^{(i)}) = A' - B^{(i)}$:

$$
L_p(f(A', B^{(i)})) = \left( \sum_{k=1}^{m} \sum_{j=1}^{n} |r_{kj} - r_{ki}|^p \right)^{1/p}
$$

(4)

In general it is possible to consider other matrix norms (for example, the Schatten norm $S_p$). As a function $f(A', B^{(i)})$ it is as well possible to use [16] the difference of covariance matrices:

$$
f(A', B^{(i)}) = A' \cdot (A')^T - B^{(i)} \cdot (B^{(i)})^T
$$

By ranging columns $i = 1,n$ by matrix norm value (4), we get the ranking of each separate indicator by information significance (the lower value of the matrix norm corresponds to the greater information importance). In this case it is only necessary to go through $n$ options. The parameter of matrix norm $p$, as shown in [16], reflects ratio to risk: at larger value of $p$ large deviations $|r_{kj} - r_{ki}|$ get larger penalties. Thus by changing the parameter $p$, it is possible to see the drift of priorities of indicators of information significance in relation to risk to get significant discrepancies with at least one of the objects.

With increasing rank of selecting matrix $B$ till the value $s$ from matrix (2) it is necessary to choose $s$ columns. Let us denote such a matrix by $B^{(I_1,I_2:\cdots,I_s)}$.

For the element-wise matrix norm, the selection of the matrix $B^{(I_1,I_2:\cdots,I_s)}$ is simplified by the fact that in order to find the minimum of the matrix norm $L_p(f(A', B^{(I_1,I_2:\cdots,I_s)}))$ one can search for a minimum separately for each column $i$ of the matrix $A'$:

$$
\sum_{j=1}^{m} |r_{ij} - r_{ik}|^p \rightarrow \text{min}
$$

where $p$ - the parameter of the element-wise matrix norm $L_p$, $k \in \{i_1,i_2,\cdots,i_s\}$ is the set of selected columns of the matrix $A'$ (variable selection parameter). If $\text{Rank}(B^{(I_1,I_2:\cdots,I_s)}) = s$, then it is required to sort out ${n \choose s}$ sets $\{i_1,i_2,\cdots,i_s\}$ (for large rank it is reasonable to perform a random search).

Let us note that within the proposed approach it is easy to combine the two classes of autoinformativity and external informativity problems. In the conditions of the criterion of external informativeness, it is necessary to assign a ranking of the objects external to the evaluation matrix. These can be expert rankings or rankings obtained by some methods on the basis of the estimated matrix $A$. In this case, the ranking can be either fully consistent or contradictory. Let $D$ be a matrix of dimension $m \cdot r$ such rankings from $r$ experts or external ranking methods. Then the task of evaluating the external informativeness of the characteristics consists in choosing a subset of such columns of the estimated matrix $A'$ and forming from them a matrix $B$, which will provide the minimum element-wise matrix norm $\|D - B\|_p \rightarrow \text{min}.

We also note that in the matrix $B$, as in the case of the autoinformativity criterion, and in the case of the criterion of external informativity, the chosen indicators can meet several times. This quantity can be a priori considered as the importance of the exponent relative to other exponents of the matrix $B$. From the primary indicators of matrix $A'$ you can build derived indicators by adding them to the initial set of indicators and evaluating their informativeness.

Finally, in the proposed method, it is possible to combine the autoinformativity criterion with the criterion of external informativity, for this it is sufficient to use the method on combining the columns of the matrices $D, A'$. 

3
3. Substantiation of the proposed approach for analysis of the information significance of indicators

To substantiate the proposed approach for assessing the informativeness of indicators, a number of interrelated statistical indicators were modeled and method performance on a priori known relationships was considered. Standard methods were as well used for comparison to analyze the relationship of a number of characteristics based on the principal component method, and the construction of regression dependencies. In the modeling, some indicators were taken independent, other indicators were a function (linear or nonlinear) of independent indicators with the addition of noise:

- random variable normally distributed with the expected value of 12 units, standard deviation 1 unit: \( p_1 \sim N(12,1) \);
- random variable normally distributed with the expected value of 10 units, standard deviation 3 units: \( p_2 \sim N(10,3) \);
- random variable normally distributed within the range of 1 to 5 units: \( p_3 \sim U(1,5) \);
- random variable distributed by Weibull distribution with scale parameter \( \lambda = 1 \) and shape parameter \( k = 1 \): \( p_4 \sim W(1,1) \);
- random variable exponentially distributed with rate parameter \( \lambda = 10 \): \( p_5 \sim Exp(10) \);
- random variable \( p_6 \sim 5p_1 + 7p_2 + N(0,1) \);
- random variable \( p_7 \sim 3p_1 + p_1p_2 + N(0,1) \);
- random variable \( p_8 \sim 4p_3^3 + p_1p_2 + N(0,1) \);
- random variable \( p_9 \sim p_4p_5 + N(0,1) \);
- random variable \( p_{10} \sim p_1p_2p_3 + N(0,1) \).

For modeling, the statistical computation environment R was used, for the repetition, a given pseudo-random sequence was chosen. Program code for generating the source data is presented below.

```r
N<-100
set.seed(777)
p1<- rank(rnorm(N,12,1))
p2<- rank(rnorm(N,10,3))
p3<- rank(runif(N,1,5))
p4<- rank(rweibull(N,4,20))
p5<-rank(rexp(N,10))
p6<-rank(5*p1+7*p2+rnorm(N,0,1))
p7<-rank(3*p1^2+p2*p1+rnorm(N,0,1))
p8<-rank(4*p3^3+p1*p2+rnorm(N,0,1))
p9<-rank(p4*p5+rnorm(N,0,1))
p10<-rank(p1*p2*p3+rnorm(N,0,1))
Quest <- data.frame(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10)
```

In total \( N = 100 \) copies of each random variable were generated, which allowed forming an estimated matrix of size \( 100 \cdot 10 \) (100 objects, each of which is characterized by 10 indicators). For the comparability of methods in the future, the normalized values of these indicators were used, the rationing was carried out by ranks. Thus, the task was to select the most informative signs set for the objects, characterized by 10 signs.

Initially, the correlation of the obtained indices was evaluated. Figure 1 shows the cross-correlation between the initial indicators.
Figure 1. Correlogram of initial indicators (Spearman's rank correlation coefficient is used).

It appears from Figure 1 that independent indicators are characterized by:
- indicator $p_1$ has strong correlation relationship with $p_7$, moderate relationship with $p_2, p_6$;
- indicator $p_2$ has strong correlation relationship with $p_6$ and moderate relationship with $p_7, p_{10}$;
- indicator $p_3$ has strong correlation relationship with $p_8, p_{10}$;
- indicator $p_4$ does not have significant relationship with other indicators;
- indicator $p_5$ has moderate relationship with $p_9$.

 Altogether, the revealed correlation dependencies reflect the above models of constructing indicators. The results of revealing the most informative subset of indicators from the proposed 10 indicators using the developed method are presented below. When selecting one most informative indicator, all 10 indicators were enumerated at different values of the element-wise matrix norm $p=2, 3, 4, 10, 100$. Calculations have shown that with the increase in the indicator $p$, ranking of indicators by informativeness in the choice of one indicator does not change. Figure 2 shows the deviations of the matrix approximated by one index from the evaluation matrix. It appears from Figure 2 that the most informative indicator is $p_{10}$, second most informative is $p_6$; the least informative are $p_5, p_9$. In general, the estimates obtained correspond to the correlation estimate: indicator $p_{10}$ has maximum correlation relationships with other indicators, indicator $p_6$ has strongest correlation relationship with $p_2$. 
Figure 2. The informative value when choosing one indicator.

During the analysis of the information value of subsets of two indicators, a complete enumeration of 45 options must be performed. The results of such enumeration at the parameter of matrix norm $p=2$ are shown in Figure 3.

Figure 3. The informative value when choosing two indicators.

It appears from Figure 3 that the most informative set of two indicators is the set of $p_9, p_{10}$. Closest sets to them are: $p_6, p_9$, as well as $p_5, p_{10}$. The least informative sets are $p_5, p_9$, as well as $p_3, p_8$. In the obtained results it should be noted that in the most informative set of the two indicators, the most informative set of one indicator ($p_{10}$) was combined with the least informative set of one indicator ($p_9$). Considering the structure of the indicator models it is clear that in the set $p_9, p_{10}$ all five independent indices are combined $p_1, p_2, p_3, p_4, p_5$, of which others are derivatives. Calculations also showed that an increase in the parameter of the matrix norm $p$ smoothes out differences in the informativeness of the indicator sets, but generally does not change the qualitative picture.
Further expansion of the set of indicators leads to a rapid increase in the number of options. For practical purposes, in order to search for an acceptable solution, a random search was implemented. At selecting the three indicators, the most informative set became $p_3, p_6, p_9$, as well as the set $p_6, p_9, p_9$. Thus, it is evident again that to one of the most informative sets of two indicators the indicator from the least informative set of two indicators is added: either $p_3$, or $p_9$. In addition, the analysis of the structure of models also indicates that by sets $p_3, p_6, p_9$, or $p_6, p_9, p_9$ all independent indicators of the array of indicators of interest are covered.

To compare the obtained results the existing methods were used. Figure 4 shows absolute values of the factor loadings of each of the characteristics when the estimated matrix is decomposed into principal components (the first principal component is taken). As the figure shows, the greatest factor loading is observed in indicators $p_{10}, p_6$, the least factor loading is observed in indicators $p_9, p_4$. These results almost coincide with the isolation of the most informative indicators by the developed method at one indicator in the approximating subset. However, in this case the developed method does not repeat completely the principal component method (for example, in the developed method the worst-case in approximation of the estimated matrix is the indicator $p_5$).

![Figure 4. Decomposition by first principal component.](image)

The further use of the principal component method in order to isolate the most informative indices from more than one element becomes complicated, since in the principal component method each main component represented as a linear combination of the sought parameters retains a certain portion of the information of the estimated matrix proportional to the eigenvalue of the principal component. Thus, the developed method extends the method of principal components in this direction.

An important criterion for the information significance of indicators is the construction of the most accurate regression dependencies on their basis. To perform such a comparative evaluation, all possible subsets of the three indicators were analyzed (their number amounted to 120 options). The calculations were carried out in R environment. The program code for such enumeration is presented below.

```r
count<-1
main_matrix<-matrix(0,nrow = 120, ncol=5)

for (i in c(1:8))
{
    for (j in c((i+1):9))
    {
        for (k in c((j+1):10))
```
\[
\begin{align*}
X & \leftarrow \text{Quest}[c(i)] \\
Y & \leftarrow \text{Quest}[c(j)] \\
Z & \leftarrow \text{Quest}[c(k)] \\
S & \leftarrow c() \\
\text{for } l \in \text{setdiff}(c(1:10), c(i,j,k)) & \\
\{ \\
\text{mydata} & \leftarrow \text{data.frame}(X,Y,Z,\text{Quest}[c(l)]) \\
\text{my_model} & \leftarrow \text{lm(} \\
& \quad \text{as.formula(paste(colnames(mydata)[4], "~",} \\
& \quad \quad \quad \text{paste(colnames(mydata)[c(1,2,3)],} \\
& \quad \quad \quad \quad \text{collapse = "+"},} \\
& \quad \quad \quad \text{sep = "")}, \text{data=mydata}) \\
S & \leftarrow c(s, \text{summary(my_model)}[\text{r.squared}]) \\
\}\ \\
\text{main_matrix}[\text{count,1}] & \leftarrow i \\
\text{main_matrix}[\text{count,2}] & \leftarrow j \\
\text{main_matrix}[\text{count,3}] & \leftarrow k \\
\text{main_matrix}[\text{count,4}] & \leftarrow \text{mean}(s) \\
\text{main_matrix}[\text{count,5}] & \leftarrow \text{sd}(s) \\
\text{count} & \leftarrow \text{count+1} \\
\}
\]

\text{write.csv(}\text{main_matrix, file = "~/inform.csv", row.names=TRUE})
\

For the evaluation, linear regression dependencies of the performance indicators (all indicators that were not included in the subset of the most informative ones were used as such) from factor characteristics (the indicators, considered the most informative were taken as such) were constructed. Each subset of the most informative features for each obtained model was compared with the coefficient of determination of the model. Eventually, each subset of characteristics taken as the most informative was compared with the average value of the coefficient of determination for all models, as well as the standard deviation of the determination coefficient for all the models built.

Table 1. Analysis of the informativeness of indicators in view of prognosis composition.

| \(p_i\) | \(p_j\) | \(p_k\) | mean\(R^2\) | sd\(R^2\) |
|--------|--------|--------|------------|----------|
| 3      | 6      | 9      | 0.6509213449 | 0.2000243582 |
| 6      | 8      | 9      | 0.6495901871 | 0.1989376844 |
| 3      | 7      | 9      | 0.6168643024 | 0.3155501039 |
| 7      | 8      | 9      | 0.616395132  | 0.3149012883 |
| 3      | 9      | 10     | 0.6120201159 | 0.2022198237 |
| 8      | 9      | 10     | 0.6064224195 | 0.2029571474 |
| 1      | 3      | 9      | 0.582801548  | 0.334085641  |
| 3      | 4      | 6      | 0.5826783665 | 0.300481417  |
| 1      | 8      | 9      | 0.5826300283 | 0.333471353  |
Table 1 shows a fragment of the calculation results, which are sorted in descending order of the average determination coefficient. Out of the obtained results it follows that the chosen by the developed methods subsets of the most informative indicators $p_6, p_8, p_9$, as well as $p_2, p_6, p_9$ have a maximum average coefficient of determination. Thus, the subset proposed by the developed method, in addition to the features considered, tends to maximize the predictive power of informative indicators, providing as a whole more accurate linear approximation of other indices.

4. Approbation of the method of analyzing the informativeness of indicators

Let us demonstrate the use of the proposed method on the example of the analysis of the information significance of the indicators of the performance of scientific organizations. Estimates were made based on data monitoring for the element-wise matrix norm for different values of the parameter $p$. The analysis showed that for small values of $p$ preference is given mainly to the financial indicator unit. So, for $p = 1$ the most informative are: $6_3_3$ - The share of internal costs for research and development of funds received to fulfill orders of economic entities; $6_1_3$ - Share of income from the rent of machinery and equipment for rent in the total income of the scientific organization (%); $5_1_7$ - Average annual cost of fixed assets of research and development per one employee of a scientific organization; $5_1_10$ - Average monthly salary of researchers of a scientific organization; $3_1_1$ - The volume of funds received from the transfer of technology, attributed to the number of workers in the scientific organization; $1_1_3$ - The number of doctoral and candidate dissertations defended during the evaluation period by employees of the scientific organization, referred to the number of researchers.

The least informative in this case are: $6_2_1$ - - The share of expenses from ordinary activities in the total actual expenditures of the scientific organization; $5_1_5$ - - Share of areas not leased out in the total area of buildings; $1_1_1$ - The share of internal costs for research and development in the total volume of works performed by the scientific organization, services.

At $p = 2$ the priorities between the indicators are somewhat equalized. And the priorities are shifted to the area of the scientific component. So the most informative are recognized: $5_1_2$ - Share in fixed assets: machinery and equipment; $4_1_3$ - Number of researchers under the age of 39 years, referred to the number of researchers; $1_2_5$ - Impact factor of publications of scientists in the Web of Science; $4_1_1$ - The proportion of researchers in the total number of workers in a scientific organization; $3_3_1$ - List of elements of the innovation infrastructure created by the scientific organization or with its participation.

At the same time, the least informative are: $5_1_5$ - Share of areas not leased in the total area of buildings (preserved from $p = 1$); $1_1_4$ - List of state and international awards, prizes, rewards, honorary titles received by a scientific organization; $2_2_1$ - The number of scientific and educational structures (departments, laboratories, scientific and educational centers, etc.), created in cooperation with higher educational institutions.

Finally, at $p = 10$, priorities are gradually shifted to the assessment of the human, technological and publication component of the scientific organization: $4_1_1$ (preserved from $p = 2$), $5_1_2$ (preserved from $p = 2$); $4_1_2$ - Number of highly qualified specialists (candidates and doctors of science); $4_1_3$ - Number of researchers under the age of 39 years, referred to the number of researchers; $5_1_6$ - The share of machinery and equipment up to 3 years of age, inclusive, in the total cost of machinery and equipment; $1_2_6$ - Number of published reports, abstracts of papers submitted by scientists at major conferences, symposiums and readings, attributed to the number of researchers; $1_2_5$ - Impact factor of publications of scientists in the Web of Science. At the same time, most of the financial indicators accepted as the most priority become the least priority for $p = 1$: $6_3_3;$ $6_1_3;$ $5_1_7;$ $5_1_10$.

Thus we shall conclude that for $p = 1$ and $p = 10$, two diametrically opposite views on the informational significance of the indicators are obtained: an increase in $p$ contributes to giving preference to primary driving factors (human, technological); a decrease in $p$ shifts priorities to the area of stimulation of primary factors (income level).
5. Conclusions
The analysis, allowed to reveal the features of the proposed method in comparison with the existing approaches, is presented in Table 2.

Table 2. Advantages and disadvantages of the proposed method.

| Property                                      | The developed method | PCA | Regression |
|-----------------------------------------------|----------------------|-----|------------|
| Analysis of autoinformativity                 | +                    | +   | -          |
| Analysis of information significance in relation to external features | + | - | + |
| Analysis of information significance of one indicator | + | + | + |
| Analysis of information significance of group of indicators | + | - | + |
| Analysis of the drift of priorities of information significance | + | - | - |
| Presence of an effective computational algorithm | - | + | + |

The analysis of the correlation relationships of the indicators made it possible to conclude that the developed method is oriented towards indicators that have the maximum number of the strongest statistical links with the rest of the indicators.

Analysis of data by the method of principal components made it possible to conclude that the developed method tends to select a subset of informative indicators, minimizing the information loss of the initial evaluation matrix, on one indicator the developed method and the method of principal components practically coincide; however, for the developed method, it is possible to expand a subset of informative indicators without a linear aggregation.

Analysis of regression dependencies on various subsets of indicators made it possible to conclude that using the developed method on the selected subset of informative indicators the construction of the most accurate linear regression dependencies is possible.

References
[1] Rico-Sulayes A 2017 Reducing Vector Space Dimensionality in Automatic Classification for Authorship Attribution Revista Ingeniería Electrónica, Automática y Comunicaciones 38 (3) 26–35
[2] Wang Y, Liu Y and Wu L 2011 Study on background modeling method based on robust principal component analysis Annual Conference on Electrical and Control Engineering ICECE 6787–6790
[3] De La Torre F and Black M 2003 A framework for robust subspace learning International Journal on Computer Vision 117–142
[4] Eckart C and Young G 1936 The approximation of one matrix by another of lower rank Psychometrika 1 (3) 211–218
[5] Gillis N 2012 Sparse and unique nonnegative matrix factorization through data preprocessing Journal of Machine Learning Research 13 3349–3386
[6] Liu J, Wang C, Gao J and Han J 2013 Multi-View Clustering via Joint Nonnegative Matrix Factorization Proceedings of SIAM Data Mining Conference 252–260
[7] Zhang T and Golub G H, 2001 Rank-one approximation to high-order tensors SIAM J Matrix Analys 23 534–550
[8] Tyrtyshnikov E 2003 Tensor Approximations of Matrices Generated by Asymptotically Smooth Functions Sb. Math. 194:6 147-160
[9] Hotelling H 1936 Relations Between Two Sets of Variates *Biometrika* **28** (3–4) 321–377

[10] Hardoon D R, Szedmak S and Shawe-Taylor J 2004 Canonical correlation analysis: an overview with application to learning methods *Neural Computation* **16** 2639–2664.

[11] Fisher R A 1936 The Use of Multiple Measurements in Taxonomic Problems *Annals of Eugenics* **7** (2) 179–188

[12] Asghari O and Hezarkhani A 2008 Applying discriminant analysis to separate the alteration zones within the Sungun porphyry copper deposit *Journal of Applied Sciences* **24** 4472-4486

[13] Fornalski K W 2015 Applications of the robust Bayesian regression analysis *International Journal of Society Systems Science* **7** (4) 314–333

[14] Andersen R 2008 Modern Methods for Robust Regression *Sage University Paper Series on Quantitative Applications in the Social Sciences* 07-152

[15] Siddiqi N 2015 SAS Institute Inc., Cary, North Carolin, *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring* 208 p

[16] Artamonov Yu 2016 Group Selection Using Matrix Norms *Izv. Irkutsk State University. University. Ser. Mathematics* **18** 3-20