MATRIX "FEATURE VECTORS" IN GROUPING INFORMATION PROBLEM: LINEAR DISCRIMINATION

Volodymyr Donchenko, Fedir Skotarenko

Abstract: *The problem of classification, clusterization or patterns recognition is one the manifestation of grouping information problem (GIP) in applied researching. It involves, beside mentioned above, the problem of recovering function, represented by empirical data (observations). Solutions of GIP largely depend of on the choice of "math representatives" of the objects under investigation. It`s usual to use a collection of real valued characteristics – "feature vector", in classification form of the GIP. Feature vector is in the essence a vector from Euclidean space* R^n *. This choice is due to the highly advanced ties - and correspond techniques- in mathematical structure of such type. This technique includes, particularly, spectrum of linear operator (SVD), Moore-Penrose inversion, orthogonal projectors operators for fundamental subspaces of the linear operator, Grouping operators and so on. Euclidean spaces* $R^{m\times n}$ *of all matrixes of fixed dimension are natural spaces of "representatives" for a great many important applied fields of investigations: speech recognition, image processing and so on. In the paper SVD and Moore – Penrose technique for* $R^{m\times n}$ *, proposed and developed in the earlier paper of the authors published in 2012 is used for formulating and solution of linear discrimination of two classes, represented by matrix learning samples.*

Keywords: *Feature vectors, information aggregating, matrix corteges, matrix corteges operators, Single Valued Decomposition for cortege linear operators, linear discrimination.*

ACM Classification Keywords: *G.2.m. Discrete mathematics: miscellaneous, G.1.6. Numerical analysis, I.5.1. Pattern Recognition, H.1.m. Models and Principles: miscellaneous*

Introduction

Grouping information problem (GIP) is fundamental problem in applied investigations. There are two main form of it, namely: the problem of recovering the function, represented by their observations, and the problem of clustering, classification and pattern recognition. Examples of approaches in the field are represented perfectly in [Kohonen, 2001], [Vapnik, 1998], [Haykin, 2001], [Friedman, Kandel, 2000], [Berry, 2004]. It is opportune to notice, that math modeling is the representation of an object structure by the means of mathematical structuring. A math structure after Georg Cantoris is a set plus "ties" between its elements. Only four fundamental types of "ties" (with its combination as fifth one) exist: relations, operations, functions and collections of subsets. Thus, the mathematical description of the object (mathematical modeling) can not be anything other than representing the object structure by the means of mathematical structuring. It refers fully to so call "complex system". A "complex system" should be understanding and, correspondingly, determined, as an objects with complex structure (complex "ties"). Namely, when reading attentively manuals by the theme (see, for example, [Yeates, Wakefield, 2004], [Forster, Hölzl, 2004]) one could find correspondent allusions. "Structure" understanding is reasonable determining of a "complex systems" instead of defining them as the "objects, consisting of numerous parts, functioning as an organic whole".

In the essence, math modeling is representing by math "parts plus ties" of the object in applied field.

It is usual in GIP to represent object under consideration by the ordered collection of characteristics: quantitative (numerical) or qualitative (non numerical). Such ordered collection with real numbered characteristics is called feature vector and thus can be considered, naturally, as element of R^n . Sometimes such collection is not collection of numbers and called cortege this case. In clustering and classification problem the collection may be both qualitative and quantitative. Feature vector case is more attractive since it allows using structural diversity of Euclidean space R^n , namely: linear operations (addition and scalar multiplying), scalar product and orthogonality, norm and distance.

Euclidean space R^n is not unique, which naturally appears in applications: the space $R^{m\times n}$ of all matrixes of a fixed $m\times n$ dimension is another example. Using off R^n in applied researches is determined largely by sophisticated techniques developed for R^n - vectors handling. Namely, these are: matrix algebra, spectrum technique (Single Valued Decomposition – SVD), Pseudo Inverse by Moore – Penrose (PIM-P) [Nashed, 1978] (see, also, [Albert, 1972], [Ben-Israel, Greville, 2002]. One cannot mention in this context the outstanding contribution of N.F. Kirichenko in development of PIM-P – technique for R^n (especially, [Кириченко, 1997] [Kirichenko, 1997], see also [Кириченко, Лепеха, 2002]). Greville's formulas: forward and inverse -for PIM-P matrixes, formulas of analytical representation for disturbances of PIM-P, - are among them. Additional results in the theme as to further development of the technique and correspondent applications one can find in [Кириченко, Лепеха, 2001], [Donchenko, Kirichenko, Serbaev, 2004], [Кириченко, Крак, Полищук, 2004], [Kirichenko, Donchenko, Serbaev, 2005], [Кириченко, Донченко, 2005], [Donchenko, Kirichenko, Krivonos, 2007], [Кириченко, Донченко, 2007], [Кириченко, Кривонос, Лепеха, 2007], [Кириченко, Донченко, Кривонос, Крак, Куляс, 2009].

As to technique designing for the Euclidean space $R^{m\times n}$ as "environmental" math structure first steps have been made for example, by [Донченко, 2011], [Donchenko, Zinko, Skotarenko, 2012]. Speech recognition with the spectrograms as the representative and the images in the problem of image processing and recognition are the natural application areas for the correspond technique.

As to the choice of the collection (design of cortege or vector) it is necessary to note, that good "feature" selection (components for feature vector or cortege or an arguments for correspond functions) determines largely the efficiency of the problem solution. This phase in solving the grouping information problem is the special step of the investigation. Experience indicates that this step should be arranged in the form of recurrent selection procedures: pre-selection and subsequent improvement of the feature characteristics. Vivid examples of such approach are the next publications on [Ivachnenko, 1995] (also [Ivachnenko, 1969] with Ivachnenko's GMDH (Group Method Data Handling) and [Vapnik, 1998] with Vapnik's Support Vector Machine. Further development of the recurrent approach in feature selection through the development and systematical application of advanced PIMP technique with criteria for estimation of feature informative significance one can find in [Donchenko, Kirichenko, Serbaev, 2004], [Кириченко, Крак, Полищук, 2004], [Kirichenko, Donchenko, Serbaev, 2005], [Кириченко, Донченко, 2005], [Donchenko, Kirichenko, Krivonos, 2007], [Кириченко, Донченко, 2007], [Кириченко, Кривонос, Лепеха, 2007], [Donchenko, Krak, Krivonos, 2012]. The idea of nonlinear recursive regressive transformations (generalized neuron nets or neurofunctional transformations) due to Professor N.F Kirichenko is represented in the works referred earlier.

Correspondent technique has been designed in this works separately for each of two its basic form f the grouping information problem. The united form of the grouping problem solution is represented here in further consideration. The fundamental basis of the recursive neurofunctional technique includes the development of pseudo inverse theory in the publications mentioned earlier first of all due to Professor N.F. Kirichenko and his disciples.

The essence of the idea mentioned above is thorough choice of the primary collection and changing it if necessary by standard recursive procedure. Each step of the procedure include detecting of insignificant components, excluding or purposeful its changing, control of efficiency of changes has been made. Correspondingly, the means for implementing the correspondent operations of the step must be designed. Methods of neurofunctional transformation (NfT) (generalized neural nets, nonlinear recursive regressive transformation: [Donchenko, Kirichenko, Serbaev, 2004], [Кириченко, Крак, Полищук, 2004], [Кириченко, Донченко, Сербаєв, 2005]).

There are two basic approaches in solving to solving classification- clusterization form of GIP when Euclidean space is “environmental” space: using of recurrent procedure of k-means type and discrimination with linear discrimination as a base. First approach needs and use so called distance of conformity with classes or clusters. Variants of such distances based on advanced PIMP-technique one can find, for example, in [Кириченко, Донченко, 2007], [Донченко, 2011]. In one can find the development of the “distance of conformity” approach for $R^{m\times n}$, based on developing of PIMP-technique for Euclidean spaces of $R^{m\times n}$ - type.

The linear discrimination (LD) form for classification - clusterization variant of GIP for $R^{m\times n}$ is formulated below in the proposed paper and solved fully on the base of PIMP-technique developed the [Donchenko, Zinko, Skotarenko, 2012], has been cited earlier.

Linear discrimination as a form classification- clusterization in GIP – problem - formalization

Linear discrimination as a form of clusterization and classification of GIP-problem (Cl-Cl GIP) for Euclidean spaces of R^n - type has been discussed and solved fully on the base of PIMP – technique in [Кириченко, Кривонос, Лепеха, 2007], [Donchenko, Krak, Krivonos, 2012] including designing of recurrent selection procedure as well as criteria of informative significance components of feature vector.

In this paper we apply the ideas of papers just have been cited for formulating and solving fully linear discrimination problem for Euclidean spaces of $R^{m\times n}$ - type for two classes, represented by learning samples.

We will reference these classes by Cl_1, Cl_2 with united learning sample $X(j)\in R^{m\times n}, j=\overline{1,N}$ and with J_1, J_2 - partition of index set $\{1,...,N\}$ which corresponds leaning samples for each of the classes:

$$J_1, J_2 \subseteq \{1,...,N\}:\ J_1\cap J_2=\varnothing,\ J_1\cup J_2=\{1,...,N\},$$
$$X(j)\in Cl_k \Leftrightarrow j\in J_k,\ k=1,2,$$

$j=\overline{1,N}$

We mean by LD - problem in $R^{m\times n}$ (linear discrimination problem) of two classes Cl_1, Cl_2 represented by the parts $X(j), j\in J_1,\ X(j), j\in J_2$ of a united learning sample $X(j), j=\overline{1,N}$, the problem of designing linear functional $\varphi : R^{m\times n}\to R^1$ (discrimination function) $\Delta>0$, which would “$\Delta-$ differentiate” classes for some $\Delta>0$, in the sense, that:

$$\begin{aligned} y_j &= \varphi(X(j)) \ge \Delta, j\in J_1, \\ y_j &= \varphi(X(j)) \le -\Delta, j\in J_2 \end{aligned} \tag{1}$$

Linearity for functional φ means, that it can be uniquely represented through the inner product i.e. that $m\times n$ - matrix A exists such, that

$$\varphi(X) = (A,X)_{tr} \tag{2}$$

Dot product $(A,B)_{tr}, A = (a_{ij}), A = (a_{ij}) \in R^{m\times n}$ - is a trace inner product, determined in the standard way by the equation

$$(A,B)_{tr} = \sum_{i=\overline{1,m}, j=\overline{1,n}} a_{ij} b_{ij} ,$$

or, equivalently, by the sum of diagonal elements (trace) of matrix product $A^T B$:

$$(A,B)_{tr} = A^T B .$$

We will denote by $\Omega(\Delta)$ for any $\Delta > 0$ the subset of all vectors $y \in R^N : y^T = (y_1,\dots,y^N)$ such, that its components $y(j), j = \overline{1,N}$ satisfy inequalities from (1):

$$\Omega(\Delta) = \{ y \in R^N : y^T = (y_1,...,y_N), y_j > \Delta, j \in J_1, y_j < -\Delta, j \in J_2 \} .$$

We will use also denotation $\wp_\alpha^*$ with matrix cortege

$$\alpha = (A_1,\dots,A_N), A_j \in R^{m\times n}, j = \overline{1,N} \tag{3}$$

for linear operator from $R^{m\times n}$ to R^N, defined by the equation

$$\wp_\alpha^* Y = \begin{pmatrix} (A_1, Y)_{tr} \\ \cdots \\ (A_N, Y)_{tr} \end{pmatrix} = \begin{pmatrix} trA_1^T \\ \cdots \\ trA_N^T Y \end{pmatrix} .$$

It has been proven in [Donchenko, Zinko, Skotarenko, 2011] that $\wp_\alpha^*$ is conjugate to a so called cortege operator $\wp_\alpha : R^N \to R^{m\times n}$, defined for a matrix cortege from (3 by the equation

$$\wp_\alpha x = \sum_{j=1}^{N} x_j A_j , x \in R^N , x^T = (x_1,...,x_N) \tag{4}$$

Thus, in the notations, have been introduced earlier, the text theorem is true.

Theorem 1. In the notation introduced previously LD-problem in $R^{m\times n}$ is equivalent to the solving of conditional system of linear equations

$$\wp_{\alpha_L}^* A = y, y \in \Omega(\Delta) \tag{5}$$

with cortege α_L, designed from the matrixes of united Learning sample:

$$\alpha_L = \left(A(1),\dots,A(N) \right) .$$

Prove. Indeed, system of the inequalities (1) is equivalent, that real-valued vector *y* with the components from (1) and functional φ from (2) is valid next statement

$$y = \begin{pmatrix} y_1 \\ \cdots \\ y_N \end{pmatrix} = \begin{pmatrix} (A,A(1))_{tr} \\ \cdots \\ (A,A(N))_{tr} \end{pmatrix} \in \Omega(\Delta) .$$

Then, by mentioned above theorem from [Donchenko, Zinko, Skotarenko, 2011] formula

$$\begin{pmatrix} (A, A(1))_{tr} \\ \cdots \\ (A, A(N))_{tr} \end{pmatrix}$$

define a linear operator from $R^{m\times n}$ to R^N, which are conjugate to cortege operator $\wp_{\alpha_L}$ from $R^{m\times n}$ to R^N with cortege α_L, defined by united learning sample:

$$\alpha_L = \left(A(1),\ldots,A(N)\right)$$

$$\wp^*_{\alpha_L} A = \begin{pmatrix} (A, A(1))_{tr} \\ \cdots \\ (A, A(N))_{tr} \end{pmatrix}.$$

Thus

$\wp^*_{\alpha_L} A \in \Omega(\Delta)$

We denote by X Gramian matrix for the united collection of Learning sample matrixes:

$X = \left(\left(A(i), A(j)_{tr}\right)\right)_{i,j=\overline{1.N}}$.

The next theorem is valid in the notations have been introduced.

Theorem 2. LD-problem is equivalent to solvability of quadratic optimization problem for $y^T Z(X) y$ in domain $\Omega(\Delta)$ i.e. it is necessary and sufficient for existence of LD-problem solution that minimum $y_* \in R^N$ of quadratic form $y^T Z(X) y$ belongs to $\Omega(\Delta)$:

$$y_* = \arg\min_{y\in R^N} y^T Z(X) y, y_* \in \Omega(\Delta) \tag{3}$$

where

$$Z(X) = E_N - X^+ X ,$$

and X^+ - Moore-Penrose pseudo inverse for matrix X as linear operator from R^N in R^N (see, for example, [Albert, 1972]).

Prove. Indeed, condition $\wp^*_{\alpha_L} A = y \in \Omega(\Delta)$ indicate, that for some $y \in \Omega(\Delta)$ linear equation $\wp^*_{\alpha_L} A = y$ is solvable for some $y \in \Omega(\Delta)$. This means, that y belongs to range of $\wp^*_{\alpha_L}$: $y \in \Re\left(\wp^*_{\alpha_L}\right)$.It is obvious, that

$\Re\left(\wp^*_{\alpha_L}\right) = \Re\left(\wp_{\alpha_L}\wp^*_{\alpha_L}\right) = \Re(X)$.

Belonging to linear subspace or range means that it is a fixed point of the correspond orthogonal projector. As the $\Re\left(\wp^*_{\alpha_L}\right) = \Re(X)$ correspond orthogonal projectors coincides $P_{\Re\left(\wp^*_{\alpha_L}\right)} = P_{\Re(X)}$, so

$$P_{\Re(X)} y = y \tag{6}$$

Consequently

$y = P_{\Re(X)} y$,

or

$\left(E_N - P_{\Re(X)}\right)y = 0$.

Orthogonal projector $P_{\Re(X)}$ uniquely determined by pseudo inverse for X according to the next equality

$P_{\Re(X)} = X^+X$.

Thus (6) one can rewrite

$\left(E_N - X^+X\right)y = 0$,

or, equivalently, in notation of, for example [Kirichenko,1997]

$$Z(X)y = 0. \tag{7}$$

In its turn, last equality is equivalent to $y^T Z(X)y = 0$.

Last equality means that absolute minimum of nonnegative quadratic form $y^T Z(X)y, y \in R^N$ is achieved in domain $\Omega(\Delta)$. It is equivalent, that there exists a $y_* \in \Omega(\Delta)$ which is minimum of $y^T Z(X)y, y \in \Omega(\Delta)$ and the minimum value is zero:

$y_*^T Z(X)y_* = 0, y_* \in \Omega(\Delta)$.

This is the finish if the prove.

Insolvability of the optimization problem with constraints from *Theorem 2* means insolvability LD-problem with the feature matrixes of the model. So the features need purposeful change, for the matrixes feature (matrixes "feature vector") now. So, criteria for the choice of correspondent components and means for correspondent changes must be available, just as that was exposed in [Кириченко, Кривонос, Лепеха, 2007], [Donchenko, Krak, Krivonos, 2012] for feature vector from R^n.

Conclusion

Conception of enriching the standard considering the "representatives" in Applied Math to be the feature vectors: elements from Euclidean space R^n - has been further developed in the paper (see, also, [Donchenko, Zinko, Skotarenko, 2012]). Using matrixes as the "representatives" of the real objects is main idea of the conception. This mean, that matrix instead vector represents all principal features of the objects in applied fields. Support of this concept requires the development of technologies handling with matrixes similar techniques operating with vectors from Euclidean spaces R^n. SVD-technique as well as PIMP - technique are the priority among them. The results of such type are represented in the paper. These results demanded a generalization of matrix algebra and transforming it in algebra of matrix and vector cortege as well as definition and using the linear cortege operator. Correspond results are represented in the paper of the authors [Donchenko, Zinko, Skotarenko, 2012]. Using that handling technique for matrix features ("matrix feature vectors") make it possible to put and fully solute the Linear Discrimination problem for two collection of matrixes. Correspond solution uses standard SVD and PIMP for Gramian matrix of united collections and solution of quadratic optimization in a domain of appropriate R^n. Thus, the development of matrix technique manages to reduce to existing technique for real valued vectors. Solution of Linear Discrimination Problem for matrixes is similar to correspond result for real-valued vectors in [Кириченко, Кривонос, Лепеха, 2007] or [Donchenko, Krak, Krivonos, 2012]. The two obvious application areas are worth mentioning within the context of the application of these results. These are: speech recognition and image processing. Matrixes naturally represent the objects under consideration, namely, spectrograms and digital images.

Bibliography

[Albert, 1972] Albert A.E. Regression and the Moore-Penrose pseudoinverse.-Academic Press., 1972, 180 p.

[Ben-Israel, Greville, 2002] Adi Ben-Israel, Thomas N.E. Greville Generalized Inverses. Theory and Applications.- Second Edition.- Springer-Verlag New York, Inc. -2003.-420 p

[Berry, 2004] Michael W. Berry Editor Survey of Text Mining: Clustering, Classification, and Retrieval Springer - Verlag New York, Inc. -2004.-244 p.

[Donchenko, Kirichenko, Krivonos, 2007] Donchenko V. Kirichenko M., Krivonos Yu. Generalizing of neural nets: Functional nets of special type// International Journal Information Models and Analysis. – N 3. –V.14. – 2007. – P. 259-266.

[Donchenko, Kirichenko, Serbaev, 2004] Donchenko V.S. Kirichenko M.F., Serbaev D.P. Recursive regression transformation and dynamical systems// Proceedings: of the Seventh International Conference "Computer Data analysis and Modeling: robustness and computer intensive methods". – V.1. – September 6-10, 2004. – Minsk. – P. 147-151

[Donchenko, Krak, Krivonos, 2007] Donchenko V. Krak Ju.., Krivonos Yu. Recurrent procedure in solving the Grouping Information Problem in Applied Mathematics//ITHEA: International Journal Information Theories and Applications. – V.1.– N 1. – 20012. – P.62 -77.

[Donchenko, Zinko, Skotarenko, 2012] Donchenko V, Zinko T, Skotarenko F. "Feature Vectors" in Grouping Information Problem in Applied Mathematics: Vectors and Matrixes. - Vitalii Velichko, Alexey Voloshin, Krassimir Markov(ed.)- Problems of Computer Intellectualization.- Institute of Cybernetics NASU, ITHEA.-Kyiv, Ukraine -Sofia, Bulgaria.- 2012.- P.111-124.

[Forster, Hölzl, 2004] Forster J., Hölzl V. Applied Evolutionary economics and complex systems.-2004-- 293 p.

[Friedman, Kandel, 2000] Menahem Friedman Abraham Kandel Introduction to pattern recognition statistical, structural, neural and Fuzzy logic approaches.- World scientific publishing co. Pte. Ltd. - reprinted 2000.-329 p.

[Haykin, 2001] Haykin S. Neural Networks. A Comprehensive Foundation.-Prentice Hall, Upper Saddle River, New Jersey 07458. – 1999.– 842 p

[Ivachnenko, 1969] Івахненко А.Г. Системи розпізнавання і автоматичного управління, що самоорганізуються. – К.: Техніка.-1969.–395 с.

[Ivachnenko, 1995] Ivachnenko A.G and Muller J.-A. Present state and new problems of further GMDH development.-SAMS, 1995.- pp.3-16.

[Kirichenko, 1997] Kirichenko N.F. Analytical Representation of Perturbation of Pseudo inverse Matrices.// Cybernetics and Systems Analysis. -. March-April 1997. - Vol.33. - Number 2.- P.230-239.

[Kirichenko, Donchenko, Serbaev, 2005] Kirichenko N.F., Donchenko V.S. Serbaev D.P. Nonlinear recursive nonlinear Transformations: Dynamic Systems and Optimizations.//Cybernetics and System Analysis. - May 2005. – V.41. - № 3. – P. 364-373.

[Kohonen, 2001] Kohonen T. Self-Organizing Maps -3-d ed. - Tokyo: Springer, 2001.- 501 p

[Nashed, 1976] Nashed M. Zuhair, Votruba G.F., editors. A Unified Operator Theory of Generalized Inverse.//Proceedings of an Advanced Seminar Sponsored by the Mathematical Research Center, The University of Wisconsin, Madison, October 8-10, 1973. – New York, Academic Press, 1976.–1200 p.

[Vapnik, 1998] Vapnik, V.N. Statistical Learning Theory. New York: Wiley, 1998

[Yeates, Wakefield, 2004] Yeates D., Wakefield T. Systems Analysis and Design. - Pearson Education. – 2004. - 499 p.

[Донченко, 2011] Владимир Донченко Евклидовы пространства числовых векторов и матриц: конструктивные методы описания базовых структур и их использование.// International Journal "Information technologies & Knowledge".- 2011.- Vol. 5.- Number 3.-P.203-216.

[Кириченко, Донченко, 2005] Кириченко М.Ф., Донченко В.С. Задача термінального спостереження динамічної системи: множинність розв'язків та оптимізація//Журнал обчислювальної та прикладної математики. – 2005. –№5– С.63-78.

[Кириченко, Донченко, 2007] Кириченко Н.Ф., Донченко В.С. Псевдообращение в задачах кластеризации.//- Кибернетика и системный анализ.- 2007.- № 4.- с. 73 –92

[Кириченко, Донченко, Кривонос, Крак, Куляс, 2009] Кириченко М.Ф., Донченко В.С Кривонос Ю.Г., , Крак Ю.В.,., Куляс А.І. Аналіз та синтез ситуацій в системах прийняття рішень. - Київ: Наукова думка.-2009. 336с.

[Кириченко, Донченко, Сербаев, 2005] Кириченко Н.Ф., Донченко В.С., Сербаев Д.П. Нелинейные рекурсивные регрессионные преобразователи: динамические системы и оптимизация.// Кибернетика и системный анализ.– №3, 2005.– С. 58-68.

[Кириченко, Крак, Полищук, 2004] Кириченко Н. Ф., Крак Ю. В., Полищук А.А. Псевдо обратные и проекционные матрицы в задачах синтеза функциональных преобразователей.// Кибернетика и системный анализ –2004.–№3.

[Кириченко, Кривонос, Лепеха, 2007] Кириченко Н.Ф., Кривонос Ю.Г., Лепеха Н.П. Синтез систем нейрофункциональных преобразователей в решении задач классификации. // Кибернетика и системный анализ. – 2007. – №3. С. 47-57

[Кириченко, Лепеха, 2002] Кириченко Н.Ф., Лепеха Н.П. Применение псевдо обратных и проекционных матриц к исследованию задач управления, наблюдения и идентификации.// Кибернетика и системный анализ. – 2002. – № 4. – С. 107-124.

[Кириченко, 1997] Н.Ф. Кириченко Н.Ф. Аналитическое представление возмущений псевдо обратных матриц.// Кибернетика и системный анализ. – 1997. – № 2. – С. 98-107.

Authors' Information

Volodymyr Donchenko *– Professor, National Taras Shevchenko University of Kyiv. 60, Volodymyrs'ka street, Kyiv, 03680, Ukraine; e-mail: voldon@bigmir.net.*

Fedir Skotarenko – *post-graduate student, National Taras Shevchenko University of Kyiv. 60, Volodymyrs'ka street, Kyiv, 03680, Ukraine.*