Data Mining and Knowledge Discovery

STRUCTURING OF RANKED MODELS

Leon Bobrowski

Abstract: *Prognostic procedures can be based on ranked linear models. Ranked regression type models are designed on the basis of feature vectors combined with set of relations defined on selected pairs of these vectors. Feature vectors are composed of numerical results of measurements on particular objects or events. Ranked relations defined on selected pairs of feature vectors represent additional knowledge and can reflect experts' opinion about considered objects. Ranked models have the form of linear transformations of feature vectors on a line which preserve a given set of relations in the best manner possible. Ranked models can be designed through the minimization of a special type of convex and piecewise linear (CPL) criterion functions. Some sets of ranked relations cannot be well represented by one ranked model. Decomposition of global model into a family of local ranked models could improve representation. A procedures of ranked models decomposition is described in this paper.*

Keywords: *Ranked regression, CPL criterion function, prognostic models, decomposition of ranked models*

ACM Classification Keywords: *Computing classification systems,*

Conference: *The paper is selected from International Conference "Classification, Forecasting, Data Mining" CFDM 2009, Varna, Bulgaria, June-July 2009*

Introduction

Linear regression models allow to predict the value of dependent variable as the weighted sum of the independent variables [1]. Parameters (weights) of such models can be estimated in a standard way from a set of feature vectors composed of independent variables values and accompanied by values of dependent variable.

Linear ranked models can also be used for the purpose of prognosis [2]. The ranked model is such a linear transformation of feature vector on a line which preserves in the best possible manner a given set of ranked relations defined on pairs of these vectors. Parameters (weights) of models are estimated on the basis of a set of ranked pairs of feature vectors. For this purpose, a special convex and piecewise linear (*CPL*) criterion functions is defined on a given family of ranked pairs of feature vectors. Parameters of the ranked line are found through minimization of a such *CPL* criterion function [3].

Some families of ordering relations between feature vectors can be fully preserved during adequate linear transformation of these vectors on the line. In such case, the ranked line represents all ordering relations between feature vectors. It has been proven that the linear model can reflect all the ranking relations between feature vectors if and only if the sets of positive and negative differences of these vectors, are linearly separable [4]. But there exist such families of order relations which cannot be fully represented by one ranked model. More than one ranked model could be needed for a satisfactory representation of ordering relations between feature vectors. Such problems are discussed in the presented paper.

Pairs of feature vectors with ranked relations

Let us take into consideration a set C of n-dimensional feature vectors $x_j[n] = [x_{j1},......,x_{jn}]^T$:

$$C = \{x_j[n]\} \ (j = 1,......., m) \tag{1}$$

Vectors $x_j[n]$ can be considered as points in the *feature space* $F[n]$ ($x_j[n] \in F[n]$). The component x_{ji} of the vector $x_j[n]$ is a numerical result of the *i*-th examination (*feature*) ($i = 1,.....,n$) of a given object or event O_j ($j = 1,.....,m$). The feature vectors $x_j[n]$ can be of a mixed type, and represent different types of measurements (for example: $x_{ji} \in \{0,1\}$ or ($x_{ji} \in R^1$). The symbol "$x_j[n] \prec x_k[n]$" means the ordering relation "*follows*", which is fulfilled for a pair of feature vectors $\{x_j[n], x_k[n]\}$ with the indices (j, k) from the set J_p:

$$(\forall (j, k) \in J_p) \ (x_j[n] \prec x_k[n]) \Leftrightarrow (x_k[n] \ \textit{follows} \ x_j[n]) \tag{2}$$

The relation "$\prec$" between feature vectors $x_j[n]$ and $x_k[n]$ ($(j,k) \in J_p$) means that the objects or events O_j and O_k could be in some causal dependence. This relation is determined on the basis of additional knowledge about some (not necessarily all) pairs of objects or events O_j and O_k. For example, a medical doctor who compares two patients O_j and O_k with the same disease can declare that the patient O_j is in a more serious condition than the patient O_k. A disease model can be designed on such basis and used for the purpose of prognosis. As another example let us consider a *causal sequence* of k events O_j:

$$O_{j(1)} \rightarrow O_{j(2)} \rightarrow \ldots\ldots\ldots. \rightarrow O_{j(k)} \tag{3}$$

were the symbol "$O_{j(k)} \rightarrow O_{j(k+1)}$" means that the event $O_{j(k+1)}$ is a consequence of the previous (earlier) event $O_{j(k)}$.

The causal sequence (2) of events O_j results in the below ordering relation among feature vectors $x_j[n]$:

$$x_{j(1)}[n] \prec x_{j(2)}[n] \prec \ldots\ldots\ldots. \prec x_{j(k)}[n] \tag{4}$$

The ordering relation (4) forms the *sequential pattern* $J_p(x)$ of feature vectors $x_j[n]$ [2].

Let us consider a linear transformation $y = w[n]^T x[n]$ of n-dimensional feature vectors $x_j[n]$ ($x_j[n] \in R^n$) on the points y_j of the line R^1 ($y_j \in R^1$):

$$(\forall j \in \{1,\ldots\ldots,m\}) \quad y_j = w[n]^T x_j[n] \tag{5}$$

where $w[n] = [w_1,......,w_n]^T$ is the weight vector.

The problem of how to design such a linear transformation $y = w[n]^T x[n]$ (5) which preserves the relation "$\prec$" for all or almost all pairs of indices (j,k) from some set J_p (2) has been analyzed in the paper [2].

Definition 1: Feature vectors $x_j[n]$ with indices j from the set J_p (2) constitute the *linear ranked pattern* $J_p(x[n])$ if and only if there exists such n-dimensional weight vector $w_p^*[n]$, that the below implication takes place for all ordering relations (2) defined by the set J_p (2):

$$(\exists w_p^*[n] \in R^n) \ (\forall (j,k) \in J_p) \ \ x_j[n] \prec x_k[n] \Rightarrow w_p^*[n]^T x_j[n] < w_p^*[n]^T x_k[n] \tag{6}$$

In this case, the ordering relations "$x_j[n] \prec x_k[n]$" are fully preserved on the ranked *line* $y = w_p^*[n]^T x[n]$.

Differential sets R^+ and R^-

The procedure of discovering the ranked linear patterns $J_p(x[n])$ (6) and the ranked line designing has been based on the concept of the positively and negatively oriented dipoles $\{x_j[n], x_{j'}[n]\}$, where $j < j'$ [2], [4].

Definition 2: The ranked pair $\{x_j[n], x_{j'}[n]\}$ of the feature vectors $x_j[n]$ and $x_{j'}[n]$ ($(j, j') \in J_p^+$, *where* $j < j'$) constitutes the *positively oriented dipole,* if and only if $x_j[n] \prec x_{j'}[n]$.

$$(\forall (j, j') \in J_p^+, \text{ where } j < j') \quad x_j[n] \prec x_{j'}[n] \tag{7}$$

Definition 3: The ranked pair $\{x_j[n], x_{j'}[n]\}$ of the feature vectors $x_j[n]$ and $x_{j'}[n]$ ($(j, j') \in J_p^-$, *where* $j < j'$ constitutes the *negatively oriented dipole* ($(j, j') \in J_p^-$), if and only if $x_{j'}[n] \prec x_j[n]$.

$$(\forall (j, j') \in J_p^-, \text{ where } j < j') \quad x_{j'}[n] \prec x_j[n] \tag{8}$$

Definition 4: The line $y(w[n]) = w[n]^T x[n]$ (5) is *fully ranked* if and only if

$$\begin{aligned} &(\forall (j, j') \in J_p^+, \text{ where } j < j') \quad w[n]^T x_j[n] < w[n]^T x_{j'}[n], \text{ and} \\ &(\forall (j, j') \in J_p^-, \text{ where } j < j') \quad w[n]^T x_{j'}[n] < w[n]^T x_j[n] \end{aligned} \tag{9}$$

where $J_p^+ \cup J_p^- = J_p$.

Let us introduce the positive set R^+ and the negative set R^- of the differential vectors $r_{jj'}[n] = x_{j'}[n] - x_j[n]$ on the basis of the sets of indices J_p^+ (7) and J_p^- (8).

$$\begin{aligned} R^+ &= \{r_{jj'}[n] = (x_{j'}[n] - x_j[n]) : (j,j') \in J_p^+\} \\ R^- &= \{r_{jj'}[n] = (x_{j'}[n] - x_j[n]) : (j,j') \in J_p^-\} \end{aligned} \tag{10}$$

We examine a separation of the sets R^+ and R^- (10) by such a hyperplane $H(w[n],\theta)$ in the feature space $F[n]$ that passes through the point 0 ($\theta = 0$), where:

$$H(w[n], \theta) = \{x[n]: w[n]^T x[n] = \theta\} \tag{11}$$

Definition 5: The differential sets R^+ and R^- (10) are linearly separable in the feature space $F[n]$ by the hyperplane $H(w[n], 0)$ with the threshold θ equal to zero ($\theta = 0$) if and only if the below inequalities hold:

$$\begin{aligned} &(\exists w'[n]) \ (\forall (j, j') \in J_p^+) \quad w'[n]^T r_{jj'}[n] > 0, \text{ and} \\ &(\forall (j, j') \in J_p^-) \quad w'[n]^T r_{jj'}[n] < 0 \end{aligned} \tag{12}$$

The hyperplane $H(w'[n],0)$ (11) separates the sets R^+ and R^- (10) if and only if all the above inequalities (12) with the vector $w'[n]$ are fulfilled.

Remark 1: All the implications (6) are fulfilled on the line $y(w'[n]) = w'[n]^T x[n]$ (5) if and only if the hyperplane $H(w'[n],0))$ (11) separates (12) the sets R^+ and R^- (10).

Convex and piecewise linear criterion function $\Phi(w[n])$

Designing the separating hyperplane $H(w[n], 0)$ (11) could be carried out through the minimisation of the convex and piecewise linear (*CPL*) criterion function $\Phi(w[n])$ similar to the perceptron criterion function [2]. Let us introduce for this purpose the positive penalty function $\varphi_{jj'}^+(w[n])$ and the negative penalty function $\varphi_{jj'}^-(w[n])$:

$$(\forall (j,j') \in J_p^+) \qquad \varphi_{jj'}^+(w[n]) = \begin{cases} 1 - w[n]^T r_{jj'}[n] & \text{if } w^T[n] r_{jj'}[n] < 1 \\ 0 & \text{if } w[n]^T r_{jj'}[n] \geq 1 \end{cases} \tag{13}$$

and

$$(\forall (j,j') \in J_p^-) \qquad \varphi_{jj'}^-(w[n]) = \begin{cases} 1 + w[n]^T r_{jj'}[n] & \text{if } w^T[n] r_{jj'}[n] > -1 \end{cases} \tag{14}$$

$$0 \qquad \textit{if} \ \ w[n]^T r_{jj'}[n] \ \leq -1$$

The criterion function $\Phi(w[n])$ is the sum of the penalty functions $\varphi_{jj'}^{+}(w[n])$ and $\varphi_{jj'}^{-}(w[n])$:

$$\Phi(w[n]) = \sum_{(j,j')\in J_p^+} \gamma_{jj'} \varphi_{jj'}^{+}(w[n]) + \sum_{(j,j')\in J_p^-} \gamma_{jj'} \varphi_{jj'}^{-}(w[n]) \qquad (15)$$

where $\gamma_{jj'}$ ($\gamma_{jj'} > 0$) is a positive parameter (*price*) related to the dipole $\{x_j[n], x_{j'}[n]\}$ ($j < j'$).

$\Phi(w[n])$ (14) is the convex and piecewise linear (*CPL*) criterion function as the sum of such type of penalty functions as $\phi_{jj'}^{+}(w[n])$ and $\phi_{jj'}^{-}(w[n])$ The basis exchange algorithms, similarly to linear programming, allow one to find the minimum of such function efficiently, even in the case of large multidimensional data sets R^+ and R^- (9) [3]:

$$\Phi^* = \Phi(w^*[n]) = \min_{w} \Phi(w[n]) \geq 0 \qquad (16)$$

The optimal parameter vector $w^*[n]$ and the minimal value Φ^* of the criterion function $\Phi(w[n])$ (15) can be applied to solving a variety of data mining tasks. In particular, the ranked line $y = (w^*[n])^T x[n]$ (5) can be found in this way. The below *Lemma* has been proved [2]:

Lemma 1: The minimal value $\Phi(w^*[n])$ (16) of the criterion function $\Phi(w[n])$ (15) is equal to zero if and only if all the inequalities (9) are fulfilled on the line $y(w^*[n]) = (w^*[n])^T x[n$ (5).

By taking into account *Remark* 1, we can prove that the minimal value $\Phi(w^*[n])$ (16) of the nonnegative criterion function $\Phi(w[n])$ (15) is equal to zero if and only if the differential sets R^+ and R^- (10) are linearly separable (12).

Linear models based on ranked relations family

Family F_p of ranked relations "$x_j(k) \prec x_{j'}(k)$" can be defined by the sets J_p^+ (7) and J_p^- (8) of pairs of indices (j, k).

$$F_p = \{x_j[n] \prec x_k[n] \colon (j, k) \in J_p\}, \ \textit{where} \ J_p = J_p^+ \cup J_p^- \qquad (17)$$

Definition 6: The family F_p is *transient* if the ranked relations "$x_j(k) \prec x_{j'}(k)$" from this family fulfill the following implication:

$$\textit{If} \ \text{"}x_j(k) \prec x_k(k)\text{"} \ \textit{and} \ \text{"}x_k[n] \prec x_l[n]\text{"}, \ \textit{then} \ \text{"}x_j[n] \prec x_l[n]\text{"} \qquad (18)$$

Definition 7: The family F_p the ranked relations is *complete* for the set C (1) if the ranked relations "$x_j[n] \prec x_k[n]$" is defined for each pair $\{x_j[n], x_k[n]\}$ of elements of this set.

Theorem 1: The complete family F_p (17) of relations "$x_j[n] \prec x_k[n]$" defines the linear ranked pattern $J_p(x[n])$ in the feature space $F[n]$ (*Definition* 1) if and only if this family is transient.

Proof: If the family F_p defines the linear ranked pattern $J_p(x[n])$, then there exists such weight vector $w_p^*[n]$ with the length equal to one ($||w_p^*[n]|| = 1$) that the below implication (6) takes place:

$$(\forall (j,k) \in J_p) \ \ x_j[n] \prec x_k[n] \Rightarrow y_j < y_k \qquad (19)$$

where $y_j = w_p^*[n]^T x_j[n]$ (5) is the point on the line $y = w_p^*[n]^T x[n]$ which is equal to the projection of the feature vector $x_j[n]$ on this line. The transient relation is fulfilled among all the points y_j on the line. Therefore, the transient relation (18) has to be fulfilled also among feature vectors $x_j[n]$. On the other hand, if the ranked relations "$x_j[n] \prec x_k[n]$" from the transient family F_p are defined for each pair $\{x_j[n], x_k[n]\}$ of elements $x_j[n]$ of the set C, then the projection points y_j fulfill the implication (6). □

Linearly separable learning sets C_k

We assume that each learning set C_k is composed of m_k labeled feature vectors $x_j(k)$ assigned in accordance with additional knowledge to the k-th category (class) ω_k ($k = 1,\ldots,K$):

$$C_k = \{x_j(k)\} \quad (j \in J_k) \tag{20}$$

where J_k is the set of indices j of the feature vectors $x_j(k)$ belonging to the class ω_k.

Vectors $x_j(k)$ can be treated as examples or prototypes for the category ω_k. The learning sets C_k (20) are *separable* in the feature space $F[n]$, if they are disjoined in this space. It means that the following rule is fulfilled: *if* $k \neq k'$, *then* $C_k \cap C_{k'} = \emptyset$.

Definition 8: The learning sets C_k (20) are *linearly separable* in the n-dimensional feature space $F[n]$ if each of the sets C_k can be fully separated from the sum of the remaining sets C_i by some hyperplane $H(w_k,\theta_k)$ (11):

$$(\forall k \in \{1,\ldots,K\})\ (\exists\, w_k,\theta_k)\ (\forall x_j(k) \in C_k)\quad (w_k)^T x_j(k) > \theta_k$$
$$\textit{and}\ \ (\forall x_j(k) \in C_i,\ i \neq k)\ \ (w_k)^T x_j(k) < \theta_k \tag{21}$$

In accordance with the relation (21), all the vectors $x_j(k)$ belonging to the learning set C_k are situated on the positive side ($(w_k)^T x_j(k) > \theta_k$) of the hyperplane $H(w_k,\theta_k)$ (11) and all feature vectors $x_j(i)$ from the remaining sets C_i are situated on the negative side ($(w_k)^T x_j(k) < \theta_k$) of this hyperplane. The linear separability (21) of the learning sets C_k (20) exists among others in the case of the linearly independent feature vectors $x_j(k)$ [2].

Definition 8: The family $F_{k,k'}$ of ordering relations "$x_j(k) \prec x_{j'}(k')$" ($(j, j') \in J_p$ (2)) among labeled feature vectors (17) from different learning sets C_k and $C_{k'}$ (20) is *consistent* with these sets, if and only if all the pairs $\{x_j(k),x_{j'}(k)\}$ are ordered in the same manner. This means that:

$$F_{k,k'} = \{\, x_j(k) \prec x_{j'}(k'):\ x_j(k) \in C_k \textit{ and } x_{j'}(k') \in C_{k'},\ \textit{where } k \neq k'\} \tag{22}$$

Let us remark that the above definition excludes ordering relations "$x_j(k) \prec x_{j'}(k)$" among labeled feature vectors $x_j(k)$ and $x_{j'}(k)$ (17) from the same learning sets C_k.

Definition 8: Two learning sets C_k and $C_{k'}$ are *linearly separable* (18) if there exists such hyperplane $H(w_k,\theta_k)$ (11) which separates these sets:

$$(\exists\, w_k,\theta_k)\ (\forall x_j(k) \in C_k)\quad (w_k)^T x_j(k) > \theta_k$$
$$\textit{and}\ \ (\forall x_{j'}(k') \in C_{k'})\quad (w_k)^T x_{j'}(k') < \theta_k \tag{21}$$

Lemma 2: If the learning sets C_k and $C_{k'}$ (20) are separated (23) by the hyperplane $H(w[n],\theta)$ (11) in the feature space $F[n]$, then the line $y(w[n]) = w[n]^T x[n]$ is *fully ranked* (9) in respect to an arbitrary consistent family $F_{k,k'}$ (22) of ordering relations "$x_j(k) \prec x_{j'}(k)$" between elements $x_j(k)$ and $x_{j'}(k')$ of these sets.

Lemma 3: If the line $y(w[n]) = w[n]^T x[n]$ is *fully ranked* (9) in respect to the consistent family $F_{k,k'}$ (22) of ordering relations "$x_j(k) \prec x_{j'}(k')$" (which are constituted by all elements $x_j(k)$ and $x_{j'}(k')$ of the learning sets C_k and $C_{k'}$ then these sets are linearly separable (23).

The above Lemmas point out the links between linear ranked models (9) and linear separability (23) of the learning sets C_k and $C_{k'}$ (20).

Decomposition of linear ranked models

As it results from the *Theorem* 1, the transient property of the complete family F_p (21) of ranked relations "$x_j[n] \prec x_k[n]$" assures that this family can be fully represented (6) on a line (5). The minimal value Φ^* (16) of the criterion function $\Phi(w[n])$ (15) is equal to zero in this case.

The minimal value Φ^* (16) of the criterion function $\Phi(w[n])$ (15) defined by arbitrary family F_p (17) of ranked relations allows to determine the degree of linearity of this family. The minimal value Φ^* (16) is greater than zero if the family F_p (21) is not linear (6). It has been proved that the minimal value Φ^* (16) of the criterion function $\Phi(w[n])$ (15) is *monotonical* in respect to reducing the relation family F_p (21) [4]. It means that:

$$(F_p \supset F_{p'}) \Rightarrow (\Phi_p^* \geq \Phi_{p'}^*) \qquad (24)$$

where Φ_p^* is the minimal value (16) of the criterion function $\Phi_p(w[n])$ (15) defined by ranked relations from the family F_p (17).

We can infer on the basis of the implication (24) that neglecting sufficient number of ranked relations "$x_j[n] \prec x_k[n]$" in the family F_p (17) allows to reduce to zero the minimal value Φ_p^* (16) of the criterion function $\Phi_p(w[n])$ (15). The multistage procedure of decomposing a global ranked model based on ranked relations family F_p (21) into a family of local ranked models can be based on the implication (24). During the first stage a possibly large subset F_1 ($F_1 \subset F_p$) of ranked relations is discovered. which can be represented in a satisfactory manner on some line (5) Then, the family F_p (17) is reduced to $F_{p'}$ by neglecting relations from the subset F_1 ($F_{p'} = F_p - F_1$). The reduced family $F_{p'}$ is then used to enhance the second linear model representing relations from the subset F_2. In this way the family F_p (21) can be reduced to zero after finite number stages and global ranked model can be replaced by a family of local ranked models.

Another procedure of decomposing the relations family F_p (21) and a global ranked model can be based on consistent subsets $F_{k,k'}$ (22) of ranked relations (2) between labeled feature vectors $x_j(k)$ and $x_{j'}(k')$ from selected learning sets C_k and $C_{k'}$ (20). In accordance with the *Lemma* 2, if the learning sets C_k and $C_{k'}$ are linearly separable, then the subset $F_{k,k'}$ (22) of relations (2) is linear and can be fully represented on the ranked line. Such conditions are shown on the Fig. 1.

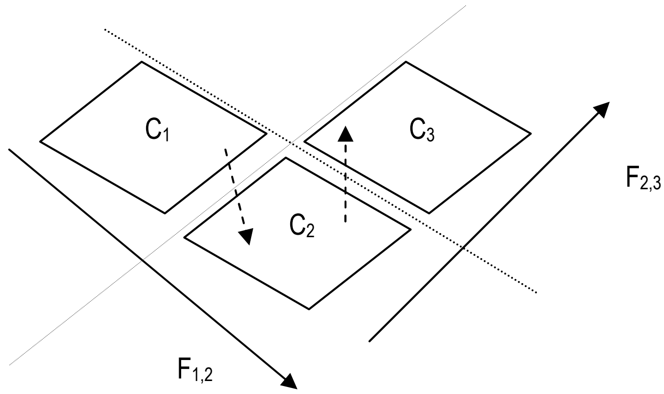

Fig. 1. An example of decomposition of nonlinear family F_p (21) of ranked relations into two linear subsets $F_{1,2}$ and $F_{2,3}$ defined by (25).

Three learning sets C_1, C_2 and C_3 are represented on the above Figure. Each learning set C_k is composed of a large number of two dimensional feature vectors $x_j(k) = [x_{j1}, x_{j2}]^T$ which can be visualized as points on the plane. We can assume that the vectors $x_j(k)$ has been generated in accordance with an uniform distribution with a specific *rhombus* shape for each learning set C_k.

Let us define the family $F_{k,\,k+1}$ (22) as a set of ranked relations "$x_j(k) \prec x_{j'}(k+1)$" between elements $x_j(k)$ and $x_{j'}(k+1)$ of the learning sets C_k and C_{k+1} ($k = 1, 2$):

$$F_{k,\,k+1} = \{ x_j(k) \prec x_{j'}(k+1), \textit{ where } k = 1 \textit{ or } k = 2 \qquad (25)$$

We can remark that the family F_p (25) is not linear, but the subsets $F_{1,2}$ and $F_{2,3}$ (22) of this set F_p are linear. As a result, the global linear model cannot represent all ranked relations from the family F_p (25), but two local models based the subsets $F_{1,2}$ and $F_{,2,3}$ allow to represent all ranked relations.

Concluding remarks

Linear ranked models can be applied for solving many problems of exploratory data analysis [2]. For example, this approach has been used for designing survival analysis models or in modeling causal sequence of liver diseases.

One of the important problems in ranked modeling is decomposing nonlinear family F_p (17) of ranked relations into linear subsets. The presented paper gives some theoretical insight into these problems where the family has the structure $F_{k,k'}$ (22) based on some learning sets C_k (20).

There are still many unanswered questions concerning decomposition of ranked models. Some of them concern the need for efficient and reliable procedures of local models enhancement when there is no specific assumption about the structure of the relations family F_p (17).

Acknowledgements

This work was supported by the by the KBN grant 3T11F01130, and partially financed by the grantS/WI/2/2009 from the Białystok University of Technology, and by the grant 16/St/2009 from the Institute of Biocybernetics and Biomedical Engineering PAS.

Bibliography

1. Duda O. R., Hart P. E., and Stork D. G.: *Pattern Classification,* J. Wiley, New York, 2001
2. L. Bobrowski, "Ranked modelling with feature selection based on the CPL criterion functions", in: *Machine Learning and Data Mining in Pattern Recognition*, Eds. P. Perner et al*., Lecture Notes in Computer Science* vol. 3587, Springer Verlag, Berlin 2005
3. Bobrowski L. and Niemiro W.: "A method of synthesis of linear discriminant function in the case of nonseparabilty". *Pattern Recognition* 17, pp.205-210,1984
4. Bobrowski : "Ranked linear models and sequential patterns recognition", pp. 1-7 in: *Pattern Analysis & Applications*, Volume 12, Issue1 (2009)

Authors' Information

Leon Bobrowski – Faculty of Computer Science, Białystok Technical University; Wiejska 45A, 15-351 Białystok, Poland, e-mail: leon@ibib.waw.pl