Ranked linear modeling in survival analysis[1]

Leon Bobrowski[1,2], **Tomasz Łukaszuk**[1]

[1] *Faculty of Computer Science, Bialystok Technical University*

[2] *Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland*

1. Introduction

Survival analysis is a class of statistical methods for modeling timing of events [1], [2], [3]. These methods are first of all applied in medicine to study fatal cases. The results of new therapeutic treatments and the effects of introducing new drugs are evaluated in this way. But applications of survival analysis are much broader and include various types of events in social or natural science, e.g. the study of bankruptcy phenomenon in economy can be based on survival analysis techniques.

The common aim of the survival analysis is to design a causal or predictive model in which the risk of the event depends on the feature vector describing a given patient (object). Survival data sets often contain feature vectors linked to the survival time of particular patients. For example, feature vectors describing patients after heart surgery can be coupled with their survival times observed during the next months or years (retrospective data). Experimental data sets collected in survival analysis are characterized by the so called right or left censoring which means some kind of missing information. *Right censoring* means that observation is terminated before the event occurs and in the result we only know that the survival time is greater than the observation time. *Left censoring* occurs when we only know that the survival time is less than some value.

In this paper we examine the possibility of linear ranked models applications in the survival analysis. The ranked model has a form of a linear transformation of the feature vectors on the line which best preserves the known order between feature vectors [4], [5]. The line can reflect the order based on comparisons between the survival times. Designing ranked models with the feature selection based on the minimisation of the convex and piecewise linear (*CPL*) functions is described in the paper.

2. Censored survival data

Let us represent the objects (patients) O_j ($j = 1,2,.....,m$) by the n-dimensional feature vectors $\mathbf{x}_j = [x_{j1},\ldots,x_{jn}]^T$ The *feature* (*attribute*) x_i describes the numerical result of the i-th measurement taken on the given object O_j ($x_i \in \{0,1\}$ or $x_i \in R^1$).

For the purpose of the survival analysis, data about particular patients O_j is represented as the elements of data set C in the manner as below:

$$C = \{(\mathbf{x}_j, t_j, \delta_j)\} \quad (j = 1,2,.....,m) \tag{1}$$

where t_j is the survival time between the entry of the j-th patient into the study and the end of the observation and δ_j is an indicator of failure of these patients ($\delta_j \in \{0,1\}$): $\delta_j = 1$ - means the end of observation in the event of interest (*failure*), $\delta_j = 0$ - means that the follow-up on the

[1] This work was partially supported by the W/II/1/2005 grant from the Białystok University of Technology and by the 16/St/2005 grant from the Institute of Biocybernetics and Biomedical Engineering PAS.

j-th patient has ended before the event (*the right censored observation*). The vectors $\mathbf{x}_j$ could be represented as points in the n-dimensional feature space $\boldsymbol{X}$.

3. Ranked linear transformations

Let the symbol "$\prec$" mean that the ranked relation "*has longer survival time than*" which may be fulfilled between two feature vectors $\mathbf{x}_j$ and $\mathbf{x}_k$

$$\mathbf{x}_j \prec \mathbf{x}_k \Leftrightarrow \mathbf{x}_k \textit{ has longer survival time than } \mathbf{x}_j \tag{2}$$

An existence of the relation "$\prec$" between feature vectors x_j and x_k means that the pair (x_j,x_k) is *ranked*. The ranked relation between feature vectors x_j and x_k results from an additional information about the objects O_j and O_k and is based on the parameters t_j, δ_j and t_k (1).

Definition 1: The vector $\mathbf{x}_k$ "has longer survival time than" $\mathbf{x}_j$ ($\mathbf{x}_j \prec \mathbf{x}_k$) if and only if the below conditions are fulfilled

$$\delta_j = 1 \ \ \textit{and} \ \ t_j < t_k \tag{3}$$

Our aim here is to design such transformation of the feature vectors x_j on the *ranked line* $y = w^T x$ which preserves the relation "$\prec$" (2) as precisely as possible

$$y_j = y_j(\mathbf{w}) = \mathbf{w}^T \mathbf{x}_j \tag{4}$$

where $\mathbf{w} = [w_1,......,w_n]^T$ is the vector of parameters.

Definition 2: The relation "$\prec$" (2) is fully preserved by the transformation (4) if and only if the following implication holds

$$(\forall (j,k)) \quad \mathbf{x}_j \prec \mathbf{x}_k \Rightarrow y_j(\mathbf{w}) < y_k(\mathbf{w}) \tag{5}$$

This implication means that the linear model (4) preserves the all known relations (2) between survival times t_j (1).

4. Positively and negatively oriented dipoles

The ranked models can be designed on the basis of the concept of positively and negatively oriented dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$, where the index j is less than j' ($j < j'$) [4].

Definition 3: The ranked pair $(\mathbf{x}_j,\mathbf{x}_{j'})$ ($j<j'$) of the feature vectors $\mathbf{x}_j$ and $\mathbf{x}_{j'}$ constitutes the *positively oriented dipole* $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($(j, j') \in I^+$), if and only if $\mathbf{x}_j \prec \mathbf{x}_{j'}$

$$(\forall \ (j,j') \in I^+) \quad \mathbf{x}_j \prec \mathbf{x}_{j'} \tag{6}$$

where I^+ is the set of indices (j, j') of the positively oriented dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($j < j'$).

Definition 4: The ranked pair $(\mathbf{x}_j,\mathbf{x}_{j'})$ ($j < j'$) of the feature vectors $\mathbf{x}_j$ and $\mathbf{x}_{j'}$ constitutes the *negatively oriented dipole* $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($(j, j') \in I^-$), if and only if $\mathbf{x}_{j'} \prec \mathbf{x}_j$.

$$(\forall\ (j,j') \in I^-)\quad \mathbf{x}_{j'} \prec \mathbf{x}_j \tag{7}$$

where I^- is the set of indices (j, j') of the negatively oriented dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ $(j < j')$.

In accordance with the relation (6), the second vector $\mathbf{x}_{j'}$ in the pair $(\mathbf{x}_j,\mathbf{x}_{j'})$ “has longer survival time than” $\mathbf{x}_j$. The first vector $\mathbf{x}_j$ ”has longer survival time than” $\mathbf{x}_{j'}$ in the case of the relation (7).

Definition 5: The line $y(\mathbf{w}) = \mathbf{w}^T\mathbf{x}$ (4) is fully consistent (*ranked*) with the dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ orientations if and only if

$$(\forall\ (j,j') \in I^+)\quad y_j(\mathbf{w}) < y_{j'}(\mathbf{w}) \qquad \textit{and} \tag{8}$$
$$(\forall\ (j,j') \in I^-)\quad y_j(\mathbf{w}) > y_{j'}(\mathbf{w})$$

Let us introduce two sets C^+ and C^- of the differential vectors $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ which are given by

$$C^+ = \{\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j):\ (j,j') \in I^+\} \tag{9}$$
$$C^- = \{\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j):\ (j,j') \in I^-\}$$

We will examine the possibility of separating the sets C^+ and C^- by the hyperplane $H(\mathbf{w})$, which passes through the origin $\mathbf{0}$ of the feature space:

$$H(\mathbf{w}) = \{\mathbf{x}: \mathbf{w}^T\mathbf{x} = 0\} \tag{10}$$

where $\mathbf{w} = [w_1,......,w_n]^T$ is the vector of parameters.

Definition 6: The sets C^+ and C^- (9) are linearly separable with the threshold equal to zero if and only if there exists such a parameter vector $\mathbf{w}^*$ that

$$(\forall\ (j,j') \in I^+)\quad (\mathbf{w}^*)^T\, \mathbf{r}_{jj'} > 0 \tag{11}$$
$$(\forall\ (j,j') \in I^-)\quad (\mathbf{w}^*)^T\, \mathbf{r}_{jj'} < 0$$

The above inequalities can be represented in the following manner

$$(\exists \mathbf{w}^*)\ (\forall\ (j,j') \in I^+)\quad (\mathbf{w}^*)^T\, \mathbf{r}_{jj'} \geq\ 1 \tag{12}$$
$$(\forall\ (j,j') \in I^-)\quad (\mathbf{w}^*)^T\mathbf{r}_{jj'} \leq\ -1$$

Remark 1: If the parameter vector $\mathbf{w}^*$ linearly separates (12) the sets C^+ and C^- (8), then the line $y_j(\mathbf{w}^*) = (\mathbf{w}^*)^T\mathbf{x}_j$ is fully consistent (7) with the dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ orientations.

5. *CPL* criterion functions

Designing the separating hyperplane $H(\mathbf{w})$ (10) can be achieved through minimisation of the convex and piecewise linear (*CPL*) criterion function $\Phi(\mathbf{w})$ [4]. Let us introduce the positive $\varphi_{jj'}^+(\mathbf{w})$ and negative $\varphi_{jj'}^-(\mathbf{w})$ penalty functions for this function.

$(\forall\ (j,j') \in I^+)$

$$\varphi_{jj'}^+(\mathbf{w}) = \quad 1 - \mathbf{w}^T\mathbf{r}_{jj'} \qquad \textbf{\textit{if}}\ \mathbf{w}^T\, \mathbf{r}_{jj'} < 1 \tag{13}$$

$$
\text{and } (\forall (j,j') \in I^-) \qquad \varphi_{jj'}^-(\mathbf{w}) = \begin{cases} 0 & \textbf{\textit{if}}\ \mathbf{w}^T \mathbf{r}_{jj'} \geq 1 \\ 1 + \mathbf{w}^T \mathbf{r}_{jj'} & \textbf{\textit{if}}\ \mathbf{w}^T \mathbf{r}_{jj'} > -1 \\ 0 & \textbf{\textit{if}}\ \mathbf{w}^T \mathbf{r}_{jj'} \leq -1 \end{cases} \tag{14}
$$

The criterion function $\Phi(\mathbf{w})$ is the weighted sum of the above penalty functions

$$
\Phi(\mathbf{w}) = \sum_{(j,j')\in I^+} \gamma_{jj'} \varphi_{jj'}^+(\mathbf{w}) + \sum_{(j,j')\in I^-} \gamma_{jj'} \varphi_{jj'}^-(\mathbf{w}) \tag{15}
$$

where $\gamma_{jj'}$ ($\gamma_{jj'} \geq 0$) is a nonnegative parameter (*price*) related to the dipole $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($j<j'$).

The criterion function $\Phi(\mathbf{w})$ (14) is the convex and piecewise linear (*CPL*) function. The basis exchange algorithms, similar to the linear programming, allow to find a minimum of such functions efficiently, even in the case of large, multidimensional data sets C^+ and C^- (9):

$$
\Phi^* = \Phi(\mathbf{w}^*) = \min_{\mathbf{w}} \Phi(\mathbf{w}) \geq 0 \tag{16}
$$

The parameter vector $\mathbf{w}^*$ defines the line $y = (\mathbf{w}^*)^T\mathbf{x}$ (3), with the best ranking. The below Lemma can be proved.

Lemma 1: The minimal value Φ^* (15) of the criterion function $\Phi(\mathbf{w})$ (14) is equal to zero if and only if there exists such a line (4) which fully preserves the relation "$\prec$" (2).

The minimal value Φ^* of the criterion function $\Phi(\mathbf{w})$ (11) and the optimal parameter vector $\mathbf{w}^*$ can be applied in solving survival analysis problems. In particular, valuable prognostic models $y(\mathbf{w}) = (\mathbf{w}^*)^T\mathbf{x}$ (4) could be found this way. If the value Φ^* is equal to zero then such model preserves all the majority relations (2) between vectors $\mathbf{x}_j$ and $\mathbf{x}_{j'}$.

6. Linear separability of the positive C^+ and the negative C^- sets

The survival data C (1) can be used in designing linear ranked models (1). For this purpose the positive C^+ and the negative C^- sets (9) of the differential vectors $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ can be used. If these sets are linearly separable with the threshold equal to zero (*Def.* 6), then the transformation $y(\mathbf{w}^*) = (\mathbf{w}^*)^T\mathbf{x}$ preserves the all ranked relations (6) and (7) which are taken into account in the sets C^+ and C^- (9). In this sense, the linear separability of the sets C^+ and C^- (9) is a key factor in preserving on the designed line (4) defined a priori order between selected feature vectors.

The linear separability with the threshold equal to zero (11) of the sets C^+ and C^- (9) depends on the number of the linearly independent vectros $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ in these sets. The differential vectors $\mathbf{r}_{jj'}$ (9) can be seen as points in the n-dimensional feature space $F[n]$.

Lemma 2: An arbitrary set R_k of n linearly independent vectors $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ ($\mathbf{r}_{jj'} \in F[n]$) defines the line (4) passing through the origin of the n-dimensional feature space $F[n]$.

Proof: Given set R_k of n linearly independent vectors $\mathbf{r}_{jj'}$ defines the below equations (12)

$$
\begin{aligned}
(\forall (j,j') \in I_k^+) \quad (\mathbf{r}_{jj'})^T\mathbf{w} &= 1 \\
(\forall (j,j') \in I_k^-) \quad (\mathbf{r}_{jj'})^T\mathbf{w} &= -1
\end{aligned} \tag{17}
$$

The vector $\mathbf{w}_k'$ constituting solution of the equation (17) defines the line (4) in the space $F[n]$

$$y = y(\mathbf{x}) = (\mathbf{w}_k')^T\mathbf{x} \tag{18}$$

where

$$\mathbf{w}_k' = \mathbf{B}_k^{-1}\mathbf{1}' \tag{19}$$

and $\mathbf{B}_k$ is the matrix (*basis*) with rows which constitute of vectors $\mathbf{r}_{jj'}$ ($\mathbf{r}_{jj'} \in R_k$) and $\mathbf{1}'$ is the vector with the components equal to 1 or –1 in accordance with (17).

Remark 2: Such sets C^+ and C^- (9) which are built solely from the vectors $\mathbf{r}_{jj'}$ ($\mathbf{r}_{jj'} \in R_k$) constituting the basis $\mathbf{B}_k$ (17) are linearly separable with the threshold equal to zero (11).

Lemma 3: The sets C^+ and C^- (9) are linearly separable with the threshold equal to zero (1) if and only if it exists such set R_k of l ($l \leq n$) linearly independent vectors $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ ($\mathbf{r}_{jj'} \in F[n]$), that the following relations holds.

$$\begin{aligned} &(\forall\, (j,j') \in I^+) \quad (\mathbf{w}_k')^T\, \mathbf{r}_{jj'} > 0 \\ &(\forall\, (j,j') \in I^-) \quad (\mathbf{w}_k')^T\, \mathbf{r}_{jj'} < 0 \end{aligned} \tag{20}$$

where $\mathbf{w}_k'$ is the vector (*vertex*) given by the equation (19).

The proof of the lemma can be based on the property that the global minimum (15) of the *CPL* criterion function $\Phi(\mathbf{w})$ (15) has to be situated in one of the vertices $\mathbf{w}_k'$ (19) [5]. The minimal value Φ^*(16) of the function $\Phi(\mathbf{w})$ (15) can be found through directed search among the vertices $\mathbf{w}_k'$ (19) in accordance with the basis exchange algorithm [5] .

If the sets C^+ and C^- are not linearly separable (11) in the feature space $F[n]$, then the minimisation (16) of the function $\Phi(\mathbf{w})$ (15) gives the vertex $\mathbf{w}_k^*$ with the condition $\Phi(\mathbf{w}_k^*) > 0$. In this case, the optimal model (18) does not preserves the all ranked relations (6) and (7).

6. Experimental results

We have done experiments with a part of the *Echocardiogram* data set taken from the UCI repository. Each patient O_j in this set is described by 8 features (x_1, x_2,......, x_8), the survival time t_j in months after the heart attack and the indicator of failure δ_j (1). ($\delta_j \in \{0,1\}$): $\delta_j = 1$ - means the end of observation by the patient death, $\delta_j = 0$ - means that the observation t_j on the patient O_j has ended before his or her death and is censored).

The experimental data set (1) contained observations ($\mathbf{x}_j$, t_j, δ_j) on 15 patients O_j. The last 5 survival periods t_j given on the Fig.1 are censored ($\delta_j = 0$ for $t_j = 1, 5, 15, 21, 28$). The positive C^+ and the negative C^- sets (9) of the differential vectors $\mathbf{r}_{jj'} = (\mathbf{x}_{j'} - \mathbf{x}_j)$ have been composed on the basis of the 15 feature vectors $\mathbf{x}_j$. The sets C^+ and C^- (9) have been based on all the dipoles $\{\mathbf{x}_j,\mathbf{x}_{j'}\}$ ($j < j'$), oriented in accordance with the rule (3).

The ranked model $y = (\mathbf{w}^*)^T\mathbf{x}$ (3), obtained trough minimisation (16) has the form

$$y = - 0.2038\, x_1 - 3.0705\, x_2 - 37.4601\, x_3 - 0.2448\, x_4 - 1.5019\, x_5 - 1.2935\, x_6 + 14.66065\, x_7 \tag{21}$$

The above model is fully consistent (5) with the dipoles $\{\mathbf{x}_j, \mathbf{x}_{j'}\}$ orientations ($\Phi(\mathbf{w}^*) = 0$ (16)). In order to obtain the prognostic model $y' = y'(\mathbf{w}') = (\mathbf{w}')^T\mathbf{x}$ an additional model scaling has been applied.

$$y_j' = y_j'(\mathbf{w}') = (\mathbf{w}')^T\mathbf{x}_j = \alpha\, (\mathbf{w}^*)^T\mathbf{x}_j + \beta \qquad (22)$$

where α and β are the scaling parameters. The parameters α and β have been fixed through minimization of the sum of the differences $|\, t_j - \alpha\, (\mathbf{w}^*)^T\mathbf{x}_j + \beta|$ for all the uncensored times t_j. As a result, the following prognostic model has been obtained

$$y_j' = 90 + 2.3\, y_j \qquad (23)$$

The comparison of this model outputs y_j' with the uncensored times t_j ($\delta_j = 1$) is showed on the plot (Fig. 1).

t_j	y_j	y'_j
10	-33,1803	13,68522
16	-32,1803	15,98522
26	-31,1803	18,28522
29	-26,8588	28,22479
32	-23,6048	35,70893
32	-22,6048	38,00893
36	-22,6048	38,00893
40	-21,6048	40,30893
48	-20,6048	42,60893
53	-19,6048	44,90893
1	-24,7718	33,02478
5	-27,8842	25,86633
15	-25,9052	30,41814
21	-26,0824	30,01039
28	-25,7182	30,84818

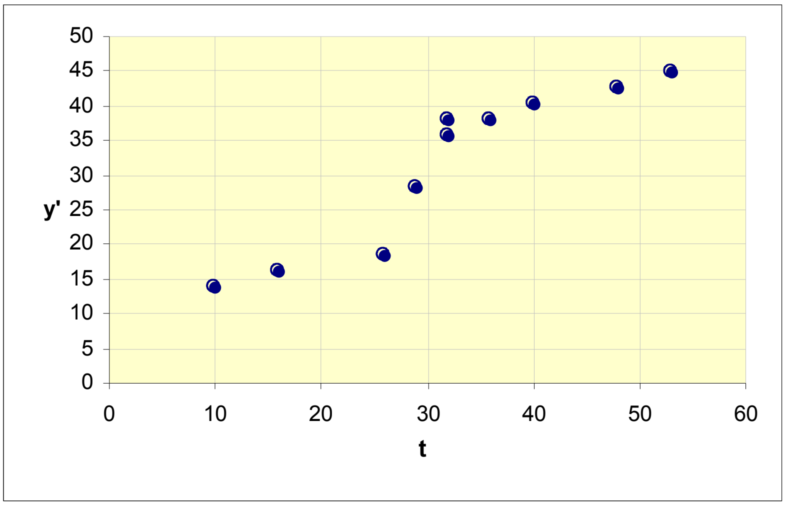

Fig. 1. Results of the experiment

Bibliography

[1] J. P. Klein, M. L. Moeschberger, *Survival Analysis*, *Techniques for Censored and Truncated Data*, Springer, NY 1997

[2] E. Biganzoli, P. Boracchi, E. Marubini: *Statistical modelling of censored survival data with artificial neural networks*, pp. 167-171 in: “Modelling and Simulation: a Tool for Next Millenium”, Edited by H. Szczerbicka, SCS, Delft 1999

[3] O. D.Allison, *Survival Analysis Using the SAS System. A Practical Guide*. SAS Institute Inc., Cary, NC. USA (1995)

[4] L. Bobrowski, *Linear ranked regression – designing principles*, CORES’05, IV International Conference on Computer Recognition Systems, Advances in Soft Computing,. Springer, 2005

[5] L. Bobrowski: *Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych* (*Data mining based on convex and piecewise linear (CPL) criterion functions*) (in Polish), Technical University Białystok, 2005