Dipolar criterion in survival time prediction

Krętowska M

Białystok Technical University, Faculty of Computer Science, Białystok, Poland

Introduction

Analysing the survival data may focus on two main tasks: to predict the failure time corresponding to a new patient and to understand the relationships between measured variables and the survival time. In both cases the problem how to treat censored cases arises. Censoring is the characteristic feature of survival data. It is connected with lack of knowledge of the failure (e.g death) occurrence. In censored observations there is no exact information about the failure time, we only know that it is greater than patient's follow-up time.

Incomplete information causes many problems in understanding the nature of the relationships among variables in the data. The basic question is how to use this particular knowledge in the prediction tools. Simultaneously with statistical methods other techniques, which try to cope with the problem, are developed. In the paper the possibilities of using piecewise linear criterion functions in the survival time prediction are presented.

Dipolar criterion

The dipolar criterion function is based on the concept of dipoles [2]. The dipole is a pair of different covariate vectors (χ_i, χ_j) from the learning set. Mixed and pure dipoles are distinguished. Mixed dipoles are formed between objects that should be separated, while pure ones between objects that are similar from the point of view of an analyzed criterion.

The aim is to find such a hyper-plane $H(v)$ that divides possibly high number of mixed dipoles and possibly low number of pure ones. It is done by minimization of the dipolar criterion function. Two types of piece-wise linear and convex (CPL) penalty functions $\varphi_i^+(v)$ and $\varphi_j^-(v)$ are considered:

$$\varphi^+_j(v)=\begin{cases}\delta^j-\langle v,y_j\rangle & \text{if}\ \langle v,y_j\rangle<\delta^j\\ 0 & \text{if}\ \langle v,y_j\rangle\geq\delta^j\end{cases}$$

$$\varphi^-_j(v)=\begin{cases}\delta^j-\langle v,y_j\rangle & \text{if}\ \langle v,y_j\rangle>\delta^j\\ 0 & \text{if}\ \langle v,y_j\rangle\leq\delta^j\end{cases}$$

where $y = [1, \chi^T]^T$ is an augmented covariate vector and $v = [-\theta, w_1, w_2, \ldots, w_N]^T$ is an augmented weight vector. Each mixed dipole (y_i, y_j), which should be divided, is associated with a function being a sum of two function with opposite signs

$$\varphi_{ij}^m(v)=\varphi_i^-(v)+\varphi_j^+(v)\ \text{ or }\ \varphi_{ij}^m(v)=\varphi_i^+(v)+\varphi_j^-(v)$$

With pure dipoles, which should remain undivided, we associate a function:

$$\varphi_{ij}^p(v)=\varphi_i^-(v)+\varphi_j^-(v)\ \text{ or }\ \varphi_{ij}^p(v)=\varphi_i^+(v)+\varphi_j^+(v)$$

The dipolar criterion function is a sum of penalty functions associated with each dipole:

$$\Psi(v)=\sum_{(i,j)\in I_p}\alpha_{ij}\varphi_{ij}^p(v)+\sum_{(i,j)\in I_m}\alpha_{ij}\varphi_{ij}^m(v)$$

where α_{ij} determines relative importance (price) of the dipole (y_i, y_j), I_p and I_m are the sets of pure and mixed dipoles, respectively.

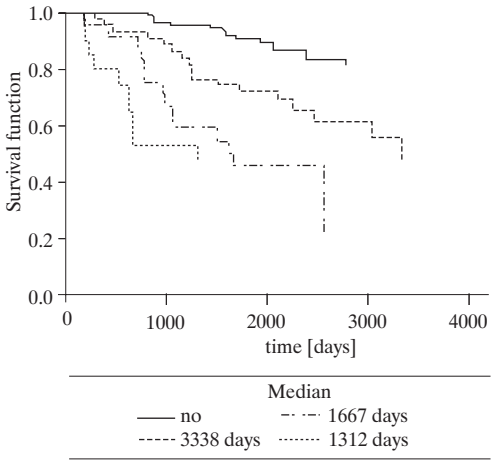
Because $\Psi(v)$ is the convex, piece-wise linear function, basis exchange algorithms, similar to linear programming, are used as a minimization method.

Prediction tools

Dependently on the presumed task the algorithm may lead to different solutions. The difference is connected with the way the dipoles are formed and the tool used. The above procedure may be used to search parameters of individual neurons in the

ADDRESS FOR CORRESPONDENCE:
Małgorzata Krętowska
Białystok Technical University, Faculty of Computer Science,
Wiejska 45a, Białystok, Poland
e-mail: mmac@ii.pb.bialystok.pl

Figure 1. **Kaplan-Meier estimates of the survival functions for distinguished subgroups obtained for *Malignant melanoma* data set**



artificial neural network or to search the splits in internal nodes of the regression tree.

Survival data: The i-th observation (patient) in survival data is described by a set (χ_i, t_i, δ_i) i=1,2,…,M, where χ_i is N-dimensional covariate vector, t_i – survival time, and δ_i – failure indicator. Failure indicator is equal to 1 for patients for whom the event of interest occurred (uncensored cases) and 0 otherwise (censored cases). Survival time may be also considered as a discrete variable. In that case the survival time t_i is divided into K disjoint intervals I_k (k=1, 2, ...,K), where $I_k=[t_{k-1}, t_k)$ and $0<t_1<t_2<...<t_K$, $t_0=0$, $t_K<\infty$. Each patient O_i is then described by a set (χ_i, I_i, δ_i), where I_i is the last time interval in which the subject O_i was observed.

Regression tree: Hierarchical and sequential structure of a tree recursively partition the feature space. The tree consists of terminal nodes (leaves) and internal (non-terminal) nodes. An internal node contains a split, which tests the value of an expression of the covariates. Each distinct outcome of the test generates one child node, which means that all non-terminal nodes have two or more child nodes. A terminal node generates no descendant [1].

The proposed method of regression tree induction aims at dividing the feature space into such areas, which would include the patients with similar survival time. It may be done by appropriate rules of dipoles formation. Pure dipoles are created between pairs of feature vectors, for which the difference of failure times is small, mixed dipoles – between pairs with distant failure times. Taking into account censored cases the following rules of dipole construction can be formulated:

a) a pair of feature vectors (χ_i, χ_j) forms the pure dipole, if
- $\delta_i=\delta_j=1$ and $|t_i-t_j|<\eta$;

b) a pair of feature vectors (χ_i, χ_j) forms the mixed dipole, if
- $\delta_i=\delta_j=1$ and $|t_i-t_j|>\zeta$;
- $(\delta_i=0, \delta_j=1$ and $t_i-t_j>\zeta)$ or $(\delta_i=1, \delta_j=0 \wedge t_j-t_i>\zeta)$.

Parameters η and ζ are equal to quartiles of absolute values of differences between uncensored survival times. The parameter η is fixed as 0.2 quartile and ζ – 0.6. The hyper-planes in the internal nodes of the tree are computed by minimization of dipolar criterion function. Each terminal node is described by median survival time and Kaplan-Meier estimator of the survival function.

Neural network: The analysis is focused on the prediction of the conditional failure probabilities (discrete hazards) in separate time intervals. The hazard rate in the k-th time interval is defined as $h_k=P(t\in I_k / t>t_{k-1})$. Taking into account the likelihood function one can show that for the i-th patient the censoring indicator d_{ki} can be treated as an estimator of the hazard h_{ki}. d_{ki} is equal to 1 for the last time interval in which the uncensored subject O_i was observed and equal to 0 otherwise.

Prediction of the conditional failure probabilities is done by using a modular neural network [3]. The network consists of K-1 ordered neural networks NN_k. Each network NN_k is trained to differentiate patients with the failure time belonging to the k-th time interval (the output equal to 1) from other patients being at risk in this time interval (0 at the output). The dipoles are formed according to following rules:

a) a pair of feature vectors (χ_i, χ_j) forms the pure dipole, if $d_{ki}\neq d_{kj}$,

b) a pair of feature vectors (χ_i, χ_j) forms the mixed dipole, if $d_{ki}=d_{kj}$.

Experimental results

The first analyzed data set contains the information on 205 patients (14 censored cases) with malignant melanoma following radical operation. The data was collected at Odense University Hospital in Denmark by Drzewiecki KT. Each patient is described by 4 features: sex, age, tumor thickness [cm] and ulceration.

The regression tree received for *Malignant melanoma* data consists of three internal nodes, which divide the feature space into four areas. The Kaplan-Meier survival functions are shown in *Fig. 1*.

The other data set contains the information from the Veteran's Administration (VA) lung cancer study. In this trial, male patients with advanced inoperable tumors were randomized to either standard (69 subjects) or test chemotherapy (68 subjects). Only 9 subjects from 137 were censored. Information on cell type, prior therapy, performance status at baseline, disease duration in months, and age in years at randomization, was available.

The survival time was divided into 3 time intervals [days]: <0,31); <31, 100);<100, ...). The modular neural network contains two networks: NN_1 and NN_2. The accuracy, sensitivity and specificity with 95% confidence interval of each network are [%]: NN_1 {acc: 78.8 (71.1; 85.4); sen: 73.2 (57.1; 85.8); sp: 81.3 (72; 88.5)}, NN_2 {acc: 73 (64.7; 80.3); sen: 74.7 (63.6; 83.8); sp: 70.7 (57.3; 82)}.

Conclusions

The dipolar criterion function is a flexible method to create different tools for survival time prediction. The main advantage of the technique is its ability to cope with censored cases, which are taken into account while dipoles construction.

Acknowledgements

The work was supported by the grant W/WI/05 from Białystok Technical University.

References

1. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees, Wadsworth, 1984.
2. Bobrowski L, Krętowska M, Krętowski M. Design of neural classifying networks by using dipolar criterions. Proceedings of the Third Conference "Neural Networks and their Applications", Częstochowa, Poland, 689-694, 1997.
3. Krętowska M, Bobrowski L. Evaluation of dipolar neural networks in survival time prediction. Advances in Soft Computing, Physica-Verlag, pp. 230-5