Evaluation of Dipolar Neural Networks in Survival Time Prediction

Małgorzata Krętowska[1] and Leon Bobrowski[1,2]

[1] Faculty of Computer Science, Technical University of Białystok, Wiejska 45a, 15-351 Białystok, Poland
[2] Institute of Biocybernetics and Biomedical Engineering PAS, Ks. Trojdena 4, Warsaw, Poland

Abstract. In this paper a dipolar neural network designed for prediction of survival time is presented. The proposed method is based on minimization of piece-wise linear criterion functions. Basis exchange algorithms are exploited as a optimization technique. The method allows to estimate of discrete hazards as conditional probabilities of failure occurrence. The prediction capacity of neural network models is evaluated by accuracy, sensitivity, and specificity measures, which are compared to results of logistic regression.

1 Introduction

Artificial neural networks can be regarded as an alternative method to the statistical models [4,11], which often require some prior assumptions. In this paper a neural network application for prediction of the time of a specific event occurrence (survival time) is considered. Survival data [5] are a set of covariates vectors with an additional variable. In case of patients, for whom the event of interest (e.g. death, disease relapse) occurred, the variable represents failure time. If the event did not occur, the variable expresses follow-up time. For the latter, so called censored cases, we do not know the exact failure time, we only know that it is no less than their follow-up time. Such incomplete information causes several problems in analysis of survival data. As ignoring the information from censored cases may bias the outcome, techniques which are able to cope with censoring are needed.

Several approaches to adapt artificial neural networks to the analysis of survival data have been proposed so far. The most common idea is to divide the survival time into disjoint intervals and train a neural network model to classify patients failure times into appropriate time intervals. Depending on the structure of a neural network, the output may be regarded as failure probability [7], hazard function [1], or cumulative failure probability [8] in ordered time intervals.

In this paper we evaluate the capacity of dipolar neural networks [6] for prediction of discrete hazards as conditional probabilities of failure. The approach is compared to logistic regression, which is widely used in medical applications.

2 Grouped survival data

Grouped survival time is considered in the paper. The failure time is divided into K disjoint intervals I_k $(k = 1, 2, \ldots, K)$, where $I_k = [t_{k-1}, t_k)$ and $0 < t_1 < t_2 < \ldots < t_K$, $t_0 = 0$, $t_K < \infty$. Each patient O_i $(i = 1, 2, \ldots, M)$ is described by $(\mathbf{x}_i, \delta_i, t_i)$, where $\mathbf{x}_i$ is N-dimensional covariates vector, δ_i - failure indicator (is equal to 1 for uncensored cases and 0 otherwise) and t_i - survival time.

The analysis is focused on prediction of the conditional failure probabilities (dicrete 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 built on a given data set we receive:

$$-\log L = -\sum_{i=1}^{M}\sum_{k=1}^{K_i}\{d_{ki}\log h_{ki} + (1 - d_{ki})\log(1 - h_{ki})\} \tag{1}$$

where K_i specifies the last time interval in which the subject O_i was observed, d_{ki} is a censoring indicator, which is equal to 1 for time interval I_{K_i} for uncensored subject O_i and equal to 0 otherwise. The equation shows that the censoring indicator d_{ki} can be treated as an estimator of the hazard h_{ki}.

3 Modular neural network

Prediction of the conditional failure probabilities is done by using a modular neural network [6]. The network consists of $K - 1$ ordered neural networks NN_k (Fig. 1). 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).

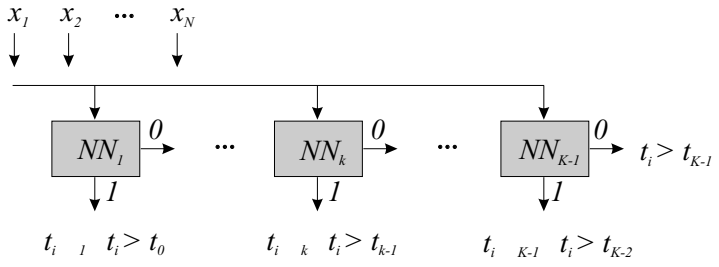

Fig. 1. The structure of modular neural network

A rule for prediction of the hazard rates for a new patient O_i is shown in Fig. 1. A new covariates vector $\mathbf{x}_i$ (describing the patient O_i) is given for the input of the network NN_1. If the module returns 1, the patient is supposed to fail in the first time interval. The output equal to 0 means, that the patient

survives the first interval and the vector $\mathbf{x}_i$ is given for the input of the next network. The procedure is repeated until one of the modules returns 1 or the last network NN_K is reached.

3.1 Learning of dipolar neural network

Each dipolar neural network NN_k consists of three layers: input, hidden and output layer. The hidden layer is built from neurons with the binary activation function.

The proposed learning procedure is based on the concept of dipoles [2]. The dipole is a pair of different covariates vectors $(\mathbf{x}_i, \mathbf{x}_j)$ from the learning set. We distinguish mixed and pure dipoles. Mixed dipoles are formed between objects, which should be separated while pure ones between objects which are similar from the point of view of an analyzed criterion.

Considering the k-th time interval we take into account only those patients, who are at risk in this interval. The main goal is to separate patients who are supposed to fail in the k-th time interval from those, who fail later. In this case the mixed dipoles are formed between each two objects O_i and O_j with $d_{ki} \neq d_{kj}$. If $d_{ki} = d_{kj}$ objects O_i and O_j constitute a pure dipole.

Let us introduce two types of piece-wise linear and convex (CPL) penalty functions $\varphi_j^+(\mathbf{v})$ and $\varphi_j^+(\mathbf{v})$:

$$\varphi_j^+(\mathbf{v}) = \begin{cases} \delta_j - < \mathbf{v}, \mathbf{y}_j > \text{ if } < \mathbf{v}, \mathbf{y}_j > \leq \delta_j \\ 0 \qquad\qquad\qquad\quad \text{if } < \mathbf{v}, \mathbf{y}_j > > \delta_j \end{cases} \tag{2}$$

$$\varphi_j^-(\mathbf{v}) = \begin{cases} \delta_j + < \mathbf{v}, \mathbf{y}_j > \text{ if } < \mathbf{v}, \mathbf{y}_j > \leq -\delta_j \\ 0 \qquad\qquad\qquad\quad \text{if } < \mathbf{v}, \mathbf{y}_j > > -\delta_j \end{cases} \tag{3}$$

where $\mathbf{y}_j = [1, x_1, \ldots, x_N]^T$ is an augmented covariates vector and $\mathbf{v} = [-\theta, w_1, \ldots, w_N]^T$ is an augmented weight vector. Each mixed dipole $(\mathbf{y}_i, \mathbf{y}_j)$, which should be divided, is associated with a function $\varphi_{ij}^m(\mathbf{v})$ being a sum of two function with opposite signs ($\varphi_{ij}^m(\mathbf{v}) = \varphi_j^+(\mathbf{v}) + \varphi_i^-(\mathbf{v})$ or $\varphi_{ij}^m(\mathbf{v}) = \varphi_j^-(\mathbf{v}) + \varphi_i^+(\mathbf{v})$). For pure dipoles, which should stay undivided, we associate a function $\varphi_{ij}^p(\mathbf{v})$ ($\varphi_{ij}^p(\mathbf{v}) = \varphi_j^+(\mathbf{v}) + \varphi_i^+(\mathbf{v})$ or $\varphi_{ij}^c(\mathbf{v}) = \varphi_j^-(\mathbf{v}) + \varphi_i^-(\mathbf{v})$). A dipolar criterion function is a sum of penalty functions associated with each dipole:

$$\Psi_d(\mathbf{v}) = \sum_{(j,i) \in I_p} \alpha_{ij} \varphi_{ij}^p(\mathbf{v}) + \sum_{(j,i) \in I_m} \alpha_{ij} \varphi_{ij}^m(\mathbf{v}) \tag{4}$$

where α_{ij} determines relative importance (price) of the dipole $(\mathbf{y}_i, \mathbf{y}_j)$, I_p and I_m are the sets of pure and mixed dipoles, respectively. The parameters of neurons in a hidden layer are obtain by sequential minimization of the dipolar criterion function. The minimization is done by using basis exchange algorithms [3], similar to linear programming. The hidden layer divides the whole N-dimensional feature space into disjoint areas. Each area is connected

with one of the two distinguished classes: patients who fail in the k-th interval, and patients who survive. A decision element in the output layer returns 1 if the input vector belongs to the area connected with first class and 0 otherwise.

4 Logistic regression

One of the most common methods used in medical applications for classification is logistic regression [11]. It can be also used to predict hazard rates for sequential time intervals I_k. For each module NN_k we could calculate (if the assumptions are fulfilled) an alternative logistic regression model LR_k:

$$\frac{P(t \in I_k | \mathbf{y}, t > t_{k-1})}{P(t > t_k | \mathbf{y}, t > t_{k-1})} = \exp(\beta^T \mathbf{y}) \tag{5}$$

where $\beta = [\beta_0, \beta_1, \ldots, \beta_N]^T$ is a vector of regression coefficients. The LR_k model should separate patients with different d_{ki} values. Thus only patients who are at risk in the k-th time interval are taken into account while calculating the regression coefficients.

5 Experimental results

The data set from the Veteran's Administration (VA) lung cancer study [5] is considered. 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 performance status at baseline (Karnofsky rating - KPS), disease duration in months, age in years at randomization, prior therapy (yes, no), and cell type (large, squamous, small, adeno), was available. The survival time was divided into three intervals (see table 1).

Table 1. Description of survival time intervals

Time interval	Uncensored cases (n)	Censored cases (n)
$I_1 : \langle 0, 31)$	39	1
$I_2 : \langle 31, 100)$	39	3
$I_3 : \langle 100, \ldots)$	50	5

The quality of the models is evaluated by accuracy, sensitivity and specificity measures. A comparison of the above measures was done by using McNemar's test for no association for matched pairs [10].

Logistic regression analysis and McNemar's test were performed using SAS STAT package. A p-value of less than 0.05 was regarded as statistically significant. Selection of the appropriate structure of the NN_k networks

(the number of neurons in hidden layers) was done by using a nested cross-validation method [9]. We received three neurons for NN_1 module and four neurons for NN_2.

Table 2. Evaluation of quality of dipolar neural networks NN_k and logistic regression LR_k

	NN_1	LR_1	NN_2	LR_2
TP (n)	30	26	23	19
FN (n)	11	15	15	19
FP (n)	18	9	17	9
TN (n)	78	87	40	48
Accuracy (%)	78.8	82.5	66.3	70.5
(95% CI)	(71.1; 85.4)	(75; 88.5)	(55.9; 75.7)	(60.3; 79.4)
Sensitivity (%)	73.2	63.4	60.5	50
(95% CI)	(57.1; 85.8)	(46.9; 77.9)	(43.4; 75.9)	(33.4; 66.6)
Specificity (%)	81.3	90.6	70.2	84.2
(95% CI)	(72; 88.5)	(83; 96)	(56.6; 81.5)	(72.1; 92.5)
McNemar's test (p)	0.194	0.004	0.724	0.059

The quality measures of dipolar neural networks and logistic regression models are shown in table 2. Taking into consideration dipolar neural networks, the differences between capacity for failures and survivors prediction are not significant ($p = 0.194$ and $p = 0.724$). In case of logistic regression models the differences are significant for LR_1 ($p = 0.004$) and marginally significant for LR_2 ($p = 0.059$).

Table 3. Comparison of dipolar neural networks NN_k to the results obtained for logistic regression models LR_k (p-values for McNemar's test)

	NN_1 vs. LR_1	NN_2 vs. LR_2
Accuracy	0.166	0.346
Sensitivity	0.046*	0.206
Specificity	0.003*	0.005*

* - statistically significant differences at 0.05 significance level

The accuracy of neural network modules is lower than logistic regression models (table 2), but the differences are not statistically significant (see table 3: $p = 0.166$ and $p = 0.206$). Statistically significant differences are between sensitivity, of NN_1 and LR_1 models ($p = 0.046$) and between specificity of the two approaches ($p = 0.003$ and $p = 0.005$ for NN_1 vs. LR_1 and NN_2 vs. LR_2 respectively).

6 Conclusions

The artificial neural network approach was proposed for prediction of discrete hazards as conditional probabilities of failure occurrence. Prediction quality of each single dipolar neural network, which is included in a modular network, was compared to a relative logistic regression model. The accuracy of the models was comparable. Statistically significant differences were found between sensitivity and specificity for the two approaches. The capacity of correct failures prediction (sensitivity) of logistic regression models was worse than the capacity of dipolar neural networks. On the other hand, the ability of correct survivors prediction (specificity) was better for logistic regression models. The differences between specificity and sensitivity for dipolar neural networks models were not very high (about 10%). In case of logistic regression models the sensitivity was about 30% lower than specificity. As from medical point of view correct failures prediction is more important, the modular neural network can be regarded as a useful tool for survival time prediction.

Acknowledgements This work was supported by the grant St/II/1/98 from Technical University of Białystok.

References

1. Biganzoli E., Boracchi P., Mariani L., Marubini E. (1998) Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. Statistics in Medicine **17**(10), 1169–1186
2. Bobrowski L., Kręţowska M., Kręţowski M. (1997) Design of neural classifying networks by using dipolar criterions. Proc. of the Third Conference on Neural Networks and Their Applications, 689–694
3. Bobrowski L., Niemiro W. (1984) A method of synthesis of linear discriminant function in the case of nonseparability. Pattern Recognition **17**, 205–210
4. Cheng B., Titterington D. M. (1994) Neural Networks: A Review from a Statistical Perspective. Statistical Science **9**(1),2–30
5. Kalbfleisch J. D., Prentice R. L. (1980) The Statistical Analysis of Failure Time Data. John Wiley & Sons
6. Kręţowska M. (2001) Dipolar neural networks applied to censored survival data. Biocybernetics and Biomedical Engineering **21**(2), 55–65
7. Ohno-Machado L. (1997) A comparison of Cox proportional hazards and artificial neural network models for medical prognosis. Comput. Biol. Med. **27**(1)
8. Ohno-Machado L., Musen M. (1997) Modular neural networks for medical prognosis: Quantifying the benefits of combining neural networks for survival prediction. Connection Science **9**, 71–96.
9. Ripley R. M. (1998) Neural networks for breast cancer prognosis. PhD Thesis, Department of Engineering Science, University of Oxford
10. SAS Institute Inc. (1996) SAS/STAT Software: Changes and Enhancements Release 6.11. Cary, NC: SAS Institute Inc.
11. Schumacher M., Rossner R., Vach W. (1996) Neural networks and logistic regression: Part I. Computational Statistics & Data Analysis **21**, 661–682