Ensemble of Dipolar Neural Networks in Application to Survival Data

Małgorzata Krętowska

Faculty of Computer Science
Białystok Technical University
Wiejska 45a, 15-351 Białystok, Poland
mmac@wi.pb.edu.pl

Abstract. In the paper the ensemble of dipolar neural networks (EDNN) for analysis of survival data is proposed. The tool is build on the base of the learning sets, which contain the data from clinical studies following patients response for a given treatment. Such datasets may contain incomplete (censored) information on patients failure times. The proposed method is able to cope with censored observations and as the result returns the aggregated Kaplan-Meier survival function. The prediction ability of the received tool as well as the significance of individual features is verified by the Brier score, $\tilde{D}_{S,x}$ and $\hat{D}_x$ measures of predictive accuracy.

1 Introduction

The main objective of regression methods is to predict the value of dependent variable y by using the set of independent features, that would be observed in the future. The regression models are usually built by minimization the sum of squares of differences between empirical (y) and theoretical ($\hat{y}$) values over all the observations from the learning set. The problem arises when the dataset does not contain the exact values of y. Such a situation is very common in survival data, in which the time of a given failure (i.e. death, disease relapse) is under investigation. The lack of knowledge of exact failure times is caused, on the one hand, by unpredictable failures being the results of other, not investigated diseases or accidents, on the other hand, by the end of follow-up time. The follow-up time in clinical trials, in which the patients response for a given treatment is studying, is determined in advance. If the failure did not occur before the end of follow-up time, the observation is cut exactly at this time. In such *censored cases* we only know that the failure time is not less than their follow-up time.

In figure 1, two described above situations are presented. Assuming that the follow-up time is a one year interval, from 01.01.2005 to 31.12.2005, the patients are included into the study just after they underwent a given treatment, often surgery. The beginning of the treatment is the starting point of their follow-up $t = 0$ (Fig. 1b). As we can see in figure 1a for patients A and D the failure occurred during the follow-up time, patient C was lost to follow-up before 31.12.2005, and patient B was observed to the end of follow-up time and during

L. Rutkowski et al. (Eds.): ICAISC 2008, LNAI 5097, pp. 78–88, 2008.

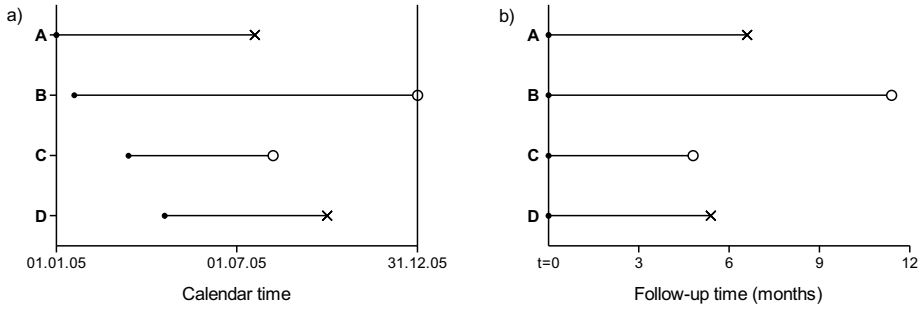

Fig. 1. Clinical trial in two points of reference: a) calendar time, b) follow-up time; x means uncensored observation, o - censored observation

this time the failure did not occur. Therefore, observations B and C are censored - the exact failure time is unknown for them.

Since the survival data is to a large extent censored, the crucial point of methods for failure time prognosis, is using the information from censored cases. The well known and widely used tool is Cox's proportional hazards model [3]. In its basic form it assumes, among other things, that covariates are independent of time and of themselves as well. If the conditions are impossible to fulfill, other, non-parametric techniques are adopted to the problem. The most common are regression trees and artificial neural networks. Recently, also methods concerning the use of random forests, bagging and boosting techniques in prognosis of survival time appear. Their application allows receiving the tool unaffected by small changes in dataset, what is particularly important in discovering the risk factors. Hothorn *et al.* [6] proposes boosting survival trees to create aggregated survival function. Krętowska [12] developed the approach by using the dipolar regression trees instead of the structure proposed in [6]. The technique proposed by Ridgeway [13] allows minimizing the partial likelihood function (boosting Cox's proportional hazard model). The Hothorn *et al.* [7] developed two approaches for censored data: random forest and gradient boosting. Breiman [2] provided the software that allows induction of the random forest for censored data.

In the paper the ensemble of dipolar neural networks (EDNN) is proposed. The individual DNN [11] is build by minimization a dipolar criterion function [1] and, by appropriate formation of the function, is able to cope with censored data. As the result of DNN a set of Kaplan-Meier survival functions is received. Each function represents the survival of an individual subgroup of observations, which is characterized by similar survival experience. A new patient may be classified to the appropriate subgroup, without exact prediction of his own failure time. The proposed algorithm enables receiving the aggregated Kaplan-Meier survival function [6] precisely for analyzed patient and predicts his survival time as the median value. The predictive ability of the proposed technique as well as the significance of individual features is assessed by using measures which were developed to cope with censored data: the Brier score [5], indirect and direct estimator of absolute predictive error and explained variation ([15,14]).

The paper is organized as follows. Section 2 describes the survival data and introduces the idea of Kaplan-Meier survival function. In Section 3 induction of dipolar neural network is presented. Section 4 contains the algorithm how to build the aggregated survival function based on ensemble of DNN and Section 5 introduces the measures of predictive accuracy. Experimental results are presented in Section 6. They were carried out on the base of two real datasets. The first one contains the feature vectors describing the patients with primary biliary cirrhosis of the liver [4], the other includes the information from the Veteran's Administration lung cancer study [8]. Section 7 summarizes the results.

2 Introduction to Survival Data

Let T^0 denotes the true survival time and C denotes the true censoring time with distribution functions F and G respectively. We observe random variable $O = (T, \Delta, \mathbf{X})$, where $T = \min(T^0, C)$ is the time to event, $\Delta = I(T \leq C)$ is a censoring indicator and $\mathbf{X} = (X_1, ..., X_N)$ denotes the set of N covariates from a sample space χ. We have learning sample $L = (\mathbf{x}_i, t_i, \delta_i)$, $i = 1, 2, ..., n$, where $\mathbf{x}_i$ is N-dimensional covariates vector, t_i - survival time and δ_i - failure indicator, which is equal to 0 for censored cases and 1 for uncensored ones.

The distribution of random variable T may be described by the marginal probability of survival up to a time $t > 0$ ($S(t) = P(T > t)$). The estimation of survival function $S(t)$ may be done by using the Kaplan-Meier product limit estimator [9], which is calculated on the base of learning sample L and is denoted by $\hat{S}(t)$:

$$\hat{S}(t) = \prod_{j|t_{(j)} \leq t} \left(\frac{m_j - d_j}{m_j} \right) \tag{1}$$

where $t_{(1)} < t_{(2)} < \ldots < t_{(D)}$ are distinct, ordered survival times from the learning sample L, in which the event of interest occurred, d_j is the number of events at time $t_{(j)}$ and m_j is the number of patients at risk at $t_{(j)}$ (i.e., the number of patients who are alive at $t_{(j)}$ or experience the event of interest at $t_{(j)}$).

The 'patients specific' survival probability function is given by $S(t|\mathbf{x}) = P(T > t|\mathbf{X} = \mathbf{x})$. The conditional survival probability function for the new patient with covariates vector $\mathbf{x}_{new}$ is denoted by $\hat{S}(t|\mathbf{x}_{new})$.

3 Dipolar Neural Network - DNN

A dipolar neural network model, considered in the paper, was proposed by Krętowska and Bobrowski [11]. The network consists of two layers: input and output layer. The output layer is build from neurons with binary activation function:

$$z = f(\mathbf{x}, \mathbf{w}) = \begin{cases} 1 & \text{if } \mathbf{w}^T\mathbf{x} \geq \theta \\ 0 & \text{if } \mathbf{w}^T\mathbf{x} < \theta \end{cases} \tag{2}$$

where $\mathbf{x}$ is a feature vector, $\mathbf{w}$ - a weight vector and θ is a threshold. From the geometrical point of view a neuron divides a feature space into two subspaces by using hyperplane $H(\mathbf{w}, \theta) = \{\mathbf{x} : \mathbf{w}^T\mathbf{x} = \theta\}$. If the vector $\mathbf{x}$ is situated on the positive side of the hyperplane the neuron is activated and $z = 1$. The layer of L binary neurons divided the N-dimensional feature space into disjoint regions - *active fields* (AF). Each region is represented by L-dimensional output vector: $\mathbf{z} = [z_1, z_2, \ldots, z_L]^T$, where $z_i \in \{0, 1\}$.

In case of survival analysis, the aim of learning algorithm is to receive such active fields which would contain observations with similar failure times. Then, on the base of Kaplan-Meier survival functions ($S_i(t)$) connected with individual active fields (AF_i), the survival probability at a given time t may be predicted and compared among all the subgroups (Fig. 2).

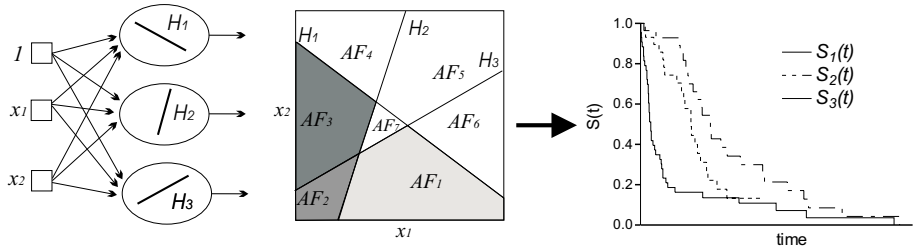

Fig. 2. Working of individual DNN

The described above objective of neural network learning procedure is realized by minimization of dipolar criterion function [1] that is built on the base of dipoles. Dipoles - pairs of feature vectors - are formed according to the following rules:

1. a pair of feature vectors $(\mathbf{x}_i, \mathbf{x}_j)$ forms the pure dipole, if
   - $\delta_i = \delta_j = 1$ and $|t_i - t_j| < \eta$
2. a pair of feature vectors $(\mathbf{x}_i, \mathbf{x}_j)$ 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 $(\delta_i = 1, \delta_j = 0$ and $t_j - t_i > \zeta)$

where η and ζ are equal to quartiles of absolute values of differences between uncensored survival times. Based on earlier experiments, parameter η was fixed as 0.2 quartile and ζ - 0.6.

As we can see, the pure dipoles are formed between feature vectors for which the difference between failure times is small, and mixed dipoles between pairs with distant failure times. In the latter case we can use the information from censored cases.

Two types of piece-wise linear and convex (CPL) penalty functions $\varphi_j^+(\mathbf{v})$ and $\varphi_j^-(\mathbf{v})$ are defined:

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

$$\varphi_j^-(\mathbf{v}) = \begin{cases} \delta_j + < \mathbf{v}, \mathbf{y}_j > & \text{if} < \mathbf{v}, \mathbf{y}_j > \geq -\delta_j \\ 0 & \text{if} < \mathbf{v}, \mathbf{y}_j > < -\delta_j \end{cases} \tag{4}$$

where δ_j is a margin ($\delta_j = 1$, for each j), $\mathbf{y}_j = [1, x_1, \ldots, x_N]^T$ is an augmented covariate 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 functions 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:

$$\Psi_d(\mathbf{v}) = \sum_{(j,i)\in I_p} \alpha_{ij}\varphi_{ij}^p(\mathbf{v}) + \sum_{(j,i)\in I_m} \alpha_{ij}\varphi_{ij}^m(\mathbf{v}) \tag{5}$$

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. Based on earlier experiments the value of α_{ij} for pure dipoles was fixed as 1 and for the mixed ones as 1000. The neurons weight values are obtained by sequential minimization of the dipolar criterion functions. The function is built from all the pure dipoles and those mixed dipoles which were not divided by previous neurons. The learning phase is finished when all the mixed dipoles are divided.

To improve the generalization ability of the network the second phase of learning procedure - optimization - is applied. The optimization phase consists of two steps. The first step is aimed at distinguishing and enlargement of prototypes(i.e. active fields which contain the largest number of feature vectors $\mathbf{x}$) and the other at reduction of redundant neurons. More detailed description can be find in [11].

4 Ensembles of DNN

Ensemble of dipolar neural networks (EDNN) is a set of DNN_i, $(i = 1, 2, \ldots, k)$, generated on base of k learning samples $(L_1, L_2, \ldots, L_k)$ drawn with replacement from the given sample L. As the result of each DNN_i, the set of active fields $SAF_i = \{AF_i^1; AF_i^2; \ldots, AF_i^{k_i}\}$ is received. Each active field AF_i^j contains the subset of observations from the learning sample L_i. Having a new covariate vector $\mathbf{x}_{new}$, each DNN_i, $i = 1, 2, \ldots, k$ returns the active field $AF_i(\mathbf{x}_{new})$, which the new observation belongs to. Let $L_i(\mathbf{x}_{new})$ denotes the set of observation covered by active field $AF_i(\mathbf{x}_{new})$. Having k sets $L_i(\mathbf{x}_{new})$, aggregated sample $L_A(\mathbf{x}_{new})$ is built [6]:

$$L_A(\mathbf{x}_{new}) = \{L_1(\mathbf{x}_{new}); L_2(\mathbf{x}_{new}); \ldots; L_k(\mathbf{x}_{new})\}$$

The aggregated conditional Kaplan-Meier survival function, calculated on the base of set $L_A(\mathbf{x}_{new})$ can be referred to as $\hat{S}_A(t|\mathbf{x}_{new})$.

To summarize the above considerations, the algorithm leading to receive the aggregated survival function is as follows:

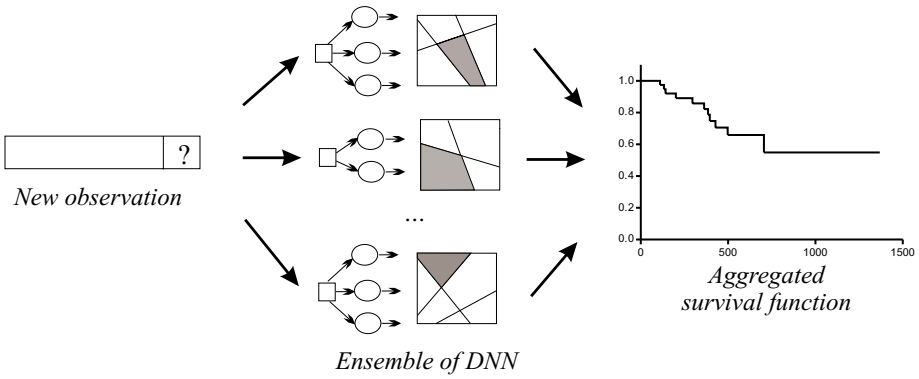

Fig. 3. EDNN in prediction of survival time for a new observation

1. Draw k bootstrap samples $(L_1, L_2, \ldots, L_k)$ of size n with replacement from L
2. Induction of dipolar neural network DNN_i based on each bootstrap sample L_i, $i = 1, 2, \ldots, k$
3. Build aggregated sample $L_A(\mathbf{x}_{new}) = \{L_1(\mathbf{x}_{new}); L_2(\mathbf{x}_{new}), \ldots, L_k(\mathbf{x}_{new})\}$
4. Compute the Kaplan-Meier aggregated survival function for a new observation $\mathbf{x}_{new}$: $\hat{S}_A(t|\mathbf{x}_{new})$ (Fig. 3).

5 Evaluation of Prediction Ability

Beside the problems concerning the use of censored data in the process of building the prediction tool, the question how to evaluate the prediction ability of received models appears. The lack of exact failure times for a part of data causes that the classical measures based on difference between empirical and theoretical values can not be used. Instead of them, other, censoring oriented, measures are proposed.

One of them is the Brier score introduced by Graf *at al.* [5]. The Brier score as a function of time is defined by

$$BS(t) = \frac{1}{n}\sum_{i=1}^{N}(\hat{S}(t|\mathbf{x}_i)^2 I(t_i \leq t \wedge \delta_i = 1)\hat{G}(t_i)^{-1} + (1 - \hat{S}(t|\mathbf{x}_i))^2 I(t_i > t)\hat{G}(t)^{-1}) \quad (6)$$

where $\hat{G}(t)$ denotes the Kaplan-Meier estimator of the censoring distribution. It is calculated on the base of observations $(t_i, 1 - \delta_i)$. $I(condition)$ is equal to 1 if the condition is fulfilled, 0 otherwise. The BS equal to 0 means the best prediction.

The Brier score belongs to direct estimators of prediction ability, because it uses the information explicitly from the data. Another direct approach is proposed by Schemper and Henderson [15]. The predictive accuracy (without

covariates), expressed by absolute predictive error (APE), at each distinct failure time $t_{(j)}$ is defined as:

$$\hat{M}(t_{(j)}) = \frac{1}{n}\sum_{i=1}^{n}\Big[I(t_i > t_{(j)})(1-\hat{S}(t_{(j)})) + \delta_i I(t_i \le t_{(j)})\hat{S}(t_{(j)}) + (1-\delta_i)I(t_i \le t_{(j)})\Big\{(1-\hat{S}(t_{(j)}))\frac{\hat{S}(t_{(j)})}{\hat{S}(t_i)} + \hat{S}(t_{(j)})(1-\frac{\hat{S}(t_{(j)})}{\hat{S}(t_i)})\Big\}\Big] \quad (7)$$

The measure with covariates ($\hat{M}(t_{(j)}|\mathbf{x})$) is obtained by replacing $\hat{S}(t_{(j)})$ by $\hat{S}(t_{(j)}|\mathbf{x})$ and $\hat{S}(t_i)$ by $\hat{S}(t_i|\mathbf{x})$. To receive overall estimators of APE with ($\hat{D}_x$) and without covariates ($\hat{D}$) the weighed averages of estimators over failure times are calculated:

$$\hat{D} = w^{-1}\sum_j \hat{G}(t_{(j)})^{-1} d_j \hat{M}(t_{(j)}) \quad (8)$$

$$\hat{D}_{\mathbf{x}} = w^{-1}\sum_j \hat{G}(t_{(j)})^{-1} d_j \hat{M}(t_{(j)}|\mathbf{x}) \quad (9)$$

where $w = \sum_j \hat{G}(t_{(j)})^{-1} d_j$, d_j is the number of events at time $t_{(j)}$ and $\hat{G}(t)$ denotes the Kaplan-Meier estimator of the censoring distribution (see equation 6).

The indirect estimation of predictive accuracy was proposed by Schemper [14]. In the approach the estimates (without $\tilde{M}(t_{(j)})$ and with covariates $\tilde{M}(t_{(j)}|\mathbf{x})$) are defined by

$$\tilde{M}(t_{(j)}) = 2\hat{S}(t_{(j)})(1-\hat{S}(t_{(j)})) \quad (10)$$

$$\tilde{M}(t_{(j)}|\mathbf{x}) = 2n^{-1}\sum_i \hat{S}(t_{(j)}|\mathbf{x}_i)(1-\hat{S}(t_{(j)})|\mathbf{x}_i) \quad (11)$$

The overall estimators of predictive accuracy with ($\tilde{D}_{S,\mathbf{x}}$) and without ($\tilde{D}_S$) covariates are calculated similarly to the estimators $\hat{D}_{\mathbf{x}}$ and $\hat{D}$. The only change is replacing $\hat{M}(t_{(j)})$ and $\hat{M}(t_{(j)}|\mathbf{x})$ by $\tilde{M}(t_{(j)})$ and $\tilde{M}(t_{(j)}|\mathbf{x})$ respectively.

Based on the above overall estimators of absolute predictive error, explained variation can be defined as:

$$\tilde{V}_S = \frac{\tilde{D}_S - \tilde{D}_{S,\mathbf{x}}}{\tilde{D}_S} \quad (12)$$

and

$$\hat{V} = \frac{\hat{D} - \hat{D}_{\mathbf{x}}}{\hat{D}} \quad (13)$$

6 Experimental Results

The analysis was conducted on the base on two datasets. The first one is from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984 [4]. 312 patients participated in the randomized trial. Survival time was taken as a number of days between registration and death,

transplantation or study analysis time in July 1986. Patients are described by the following features: age(AGE), sex, presence of edema, logarithm of serum bilirubin [mg/dl] ($LOGBILL$), albumin [gm/dl] ($ALBUMIN$), logarithm of prothrombin time [seconds], histologic stage of disease. Dataset contains 60 per cent of censored observations.

All the experiments were performed using the ensemble of 200 DNN. The measures of predictive accuracy were calculated on the base of learning sample L. To calculate the aggregated survival function for a given example $\mathbf{x}$ from the learning set L, only such DNN_i ($i = 1, 2, \ldots, 200$) were taken into consideration, for which $\mathbf{x}$ was not belonged to the learning set L_i (i.e. $\mathbf{x}$ did not participate in the learning process of the DNN_i).

Table 1. Measures of predictive accuracy for PBC dataset

Model	BS (12years)	Indirect APE/ Explained variation	Direct APE/ Explained variation
K-M Estimator	0.23	0.37	0.37
Ensemble of DNN			
all covariates	0.17	0.29/0.22	0.27/0.26
AGE	0.22	0.36/0.036	0.36/0.038
$LOGBILL$	0.17	0.28/0.25	0.28/0.25
$ALBUMIN$	0.22	0.33/0.11	0.33/0.12

In table 1 the Brier score as well as direct and indirect estimators of absolute prediction error (APE) and explained variation for /emphPBC dataset are presented. The absolute prediction error without covariates is equivalent to the error for K-M estimator and is equal to 0.37. The indirect and direct estimators of the predition errors for ensemble of DNN with all the covariates are $\tilde{D}_{S,\mathbf{x}} = 0.29$ and $\hat{D}_{\mathbf{x}} = 0.27$ respectively. It means that the knowledge of the prognostic factors reduces the absolute error of prediction of survival probability in the first 12 years after registration by 0.1 (or 0.08 in direct approach). The variation explained by the model is equal to 22 per cent in indirect approach and 26 per cent according to direct approach. More detailed analysis of individual covariates shows that the logarithm of serum bilirubin is the most important prognostic factor with $\tilde{D}_{S,\mathbf{x}} = \hat{D}_{\mathbf{x}} = 0.28$ and $\tilde{V}_S = \hat{V} = 0.25$. The influence of age and albumin for prediction of survival probability is less important. Similar conclusions can be draw from the analysis of the Brier score. The BS after 12 years of follow-up for Kaplan-Meier estimator is equal to 0.23. Similar values are for AGE and ALBUMIN ($BS = 0.22$). The BS(12 years) is smaller and equals to 0.17 in two cases: for the ensemble with all the covariates and for the model with $LOGBILL$ only.

In figure 4 we can see Kaplan-Meier survival functions received for three different values of AGE and LOGBILL, together with predicted failure times (median values). Estimated failure times for LOGBILL (Fig. 4a) equal to 0.7 and 1.5 are 2796 and 1427 [days] respectively. The median value for LOGBILL equal to -0.5 is greater than 4000 days. We can say that greater values of LOGBILL are

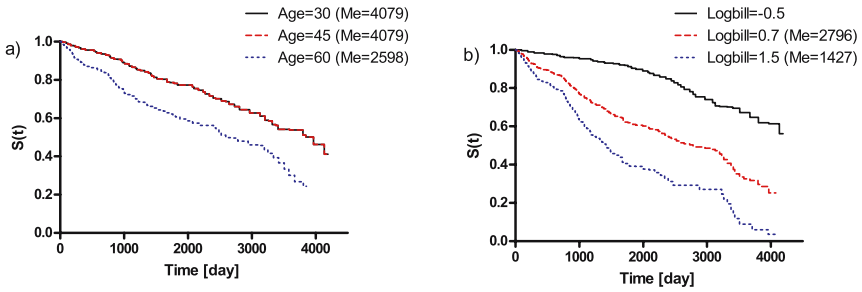

Fig. 4. Kaplan-Meier survival functions received for *PBC* dataset for different values of AGE (a) and LOGBILL (b)

connected with worse survival prediction. The failure times for three different values of AGE (30, 45, 60) are 4079, 4079 and 2598 respectively.

The other analyzed dataset contains the information from the Veteran's Administration (*VA*) lung cancer study [8]. 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 (0 - squamous, 1 - small, 2 - adeno, 3 - large) - CELL TYPE, prior therapy, performance status at baseline (Karnofsky rating - KPS), disease duration in months (TIME) and age in years at randomization (AGE), was available.

Table 2. Measures of predictive accuracy for *VA lung cancer* data

Model	*BS* (100 days)	Indirect *APE*/ Explained variation	Direct *APE*/ Explained variation
K-M Estimator	0.24	0.335	0.335
Ensemble of DNN			
all covariates	0.18	0.3/0.11	0.29/0.14
AGE	0.24	0.32/0.034	0.33/0.013
CELL TYPE	0.24	0.33/0.002	0.33/0.006
KPS	0.19	0.3/0.11	0.29/0.13
TIME	0.24	0.33/0.003	0.33/0.003

The measures of predictive accuracy for *VA lung cancer* data was shown in table 2. The unconditional absolute predictive error is 0.335. The ensemble of DNN, built on the base of all the covariates, reduces the error by 0.035 or 0.045 for indirect and direct approach respectively. The variance explained by the model is equal to 11 (14) per cent. The most important prognostic factor is KPS with the error equal to 0.3. Explained variation is 11 (13) per cent. Other variables have the marginal influence on the prediction of survival probability. Taking into account the values of Brier score after the first 100 days of follow-up the best prediction ability have the model built on the base of all the covariates

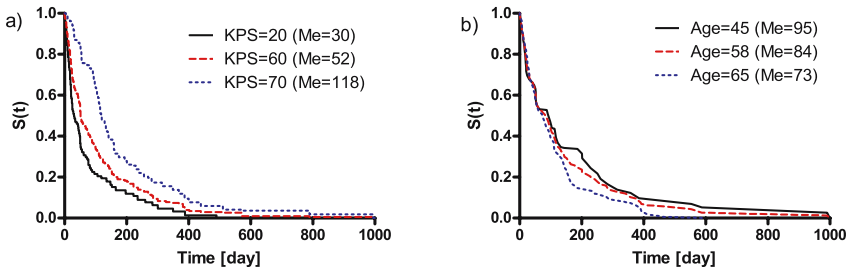

Fig. 5. Kaplan-Meier survival functions received for *VA lung cancer* data for different values of KPS (a) and AGE (b)

($BS = 0.18$). Similar value is received for the KPS feature ($BS = 0.19$) what makes it the strongest prognostic factor. For other covariates the BS value do not differ from the value for K-M estimator.

In figure 5 we can see Kaplan-Meier survival functions received for three different values of AGE and KPS together with predicted failure times. Estimated failure times for KPS (Fig. 5a) equal to 20, 60 and 70 are 30, 52 and 118 [days] respectively. We can noticed that greater values of LOGBILL are connected with better survival prediction. The failure times for three different values of AGE (45, 58, 65) are 95, 84 and 73. The functions received for KPS are more diverse than the estimators obtained for different values of AGE.

7 Conclusions

In the paper the ensemble of dipolar neural networks for prediction of survival time is proposed. The method is able to cope with censored observations, for which the exact failure time is unknown. The method, based on results of individual DNNs, produces the aggregated Kaplan-Meier survival function for a new patient described by $\mathbf{x}$. The unknown failure time for $\mathbf{x}$ may be estimated by median value of the received function.

The prediction ability of the model was verified by several measures, such as the Brier score and direct and indirect estimators of absolute predictive errors: $\tilde{D}_{S,x}$, $\hat{D}_x$. The direct comparison of the received assessments is rather difficult. We can only noticed, that all the measures distinguished the same risk factors - the features that influence the survival the most: Karnofsky rating in case of *VA lung cancer* data and serum bilirubin for *PBC* dataset. The results were confirmed by graphical representation of survival function for different feature values.

Acknowledgements. This work was supported by the grant W/WI/4/05 from Białystok Technical University.

References

1. Bobrowski, L., Krętowska, M., Krętowski, M.: Design of neural classifying networks by using dipolar criterions. In: Proc. of the Third Conference on Neural Networks and Their Applications, Kule, Poland, pp. 689–694 (1997)
2. Breiman, L.: How to use survival forest, `http://stat-www.berkeley.edu/users/breiman`
3. Cox, D.R.: Regression models and life tables (with discussion). Journal of the Royal Statistical Society B 34, 187–220 (1972)
4. Fleming, T.R., Harrington, D.P.: Counting Processes and Survival Analysis. John Wiley & Sons, Inc., Chichester (1991)
5. Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M.: Assessment and comparison of prognostic classification schemes for survival data. Statistics in Medicine 18, 2529–2545 (1999)
6. Hothorn, T., Lausen, B., Benner, A., Radespiel-Troger, M.: Bagging survival trees. Statistics in Medicine 23, 77–91 (2004)
7. Hothorn, T., Buhlmann, P., Dudoit, S., Molinaro, A.M., van der Laan, M.J.: Survival ensembles. Berkeley Division of Biostatistics Working Paper Series, vol. 174 (2005), `http://www.bepress.com/ucbbiostat/paper174`
8. Kalbfleisch, J.D., Prentice, R.L.: The statistical analysis of failure time data. John Wiley & Sons, New York (1980)
9. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 5, 457–481 (1958)
10. Krętowska, M.: Dipolar regression trees in survival analysis. Biocybernetics and biomedical engineering 24(3), 25–33 (2004)
11. Krętowska, M., Bobrowski, L.: Artificial Neural Networks in Identifying Areas with Homogeneous Survival Time. In: Rutkowski, L., Siekmann, J.H., Tadeusiewicz, R., Zadeh, L.A. (eds.) ICAISC 2004. LNCS (LNAI), vol. 3070, pp. 1008–1013. Springer, Heidelberg (2004)
12. Krętowska, M.: Random Forest of Dipolar Trees for Survival Prediction. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) ICAISC 2006. LNCS (LNAI), vol. 4029, pp. 909–918. Springer, Heidelberg (2006)
13. Ridgeway, G.: The state of boosting. Computing Science and Statistics 31, 1722–1731 (1999)
14. Schemper, M.: Predictive accuracy and explained variation. Statistics in Medicine 22, 2299–2308 (2003)
15. Schemper, M., Henderson, R.: Predictive accuracy and explained variation in Cox regression. Biometrics 56, 249–255 (2000)