Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2024;10(4): 434-447

Original Article

Deep Neural Network for Cure Fraction Survival Analysis Using Pseudo Values

Ola Abuelamayem*

Department of Statistics, Faculty of Economics and Political Science, Cairo University, Giza, Egypt.

ARTICLE INFO

Received 13.06.2024
Revised 03.07.2024
Accepted 11.08.2024
Published 15.12.2024

Key words:
Cure fraction;
Deep learning;
Neural Network;
Pseudo Values

ABSTRACT

Introduction: The hidden assumption in most of survival analysis models is the occurrence of the event of interest for all study units. The violation of this assumption occurs in several situations. For example, in medicine, some patients may never have cancer, and some may never face Alzheimer. Ignoring such information and analyzing the data with traditional survival models may lead to misleading results. Analyzing long term survivals can be performed using both traditional and neural networks. There has been an increasing interest in modeling lifetime data using neural network due to its ability to handle complex covariates if any. Also, in several numerical results it provides a better prediction. However, for long-term survivors only one neural network was introduced to estimate the uncured proportion together with the EM algorithm to account for the latency part. Neural network in survival analysis requires special cost function to account for censoring.
Methods: In this paper, we extend the neural network using pseudo values to analyze cure fraction model. It neither requires the use of special cost function nor the EM algorithm.
Results: The network is applied on both synthetic and Melanoma real datasets to evaluate its performance. We compared the results using goodness of fit methods in both datasets with cox proportional model using EM algorithm.
Conclusion: The proposed neural network has the flexibility of analyzing data without parametric assumption or special cost function. Also, it has the advantage of analyzing the data without the need of EM algorithm. Comparing the results with cox proportional model using EM algorithm, the proposed neural network performed better.

Introduction

In most of the survival analysis models there is a hidden assumption of the occurrence of the event of interest for all study units. However, in some cases this assumption may not be satisfied. For example, in medical field, some patients are cured and never face the recurrence of a certain disease like melanoma. In economics, unemployed person may never find a job. In finance, some banks may never face bankrupt. In demography, one may never get married. (For more examples see, Amico and Keilegom).[1]

*.Corresponding Author: ola.abuelamayem@feps.edu.eg

In this case, standard survival models are no longer efficient for analyzing such data. To solve this problem cure fraction models were introduced. There are two main types, mixture and non mixture cure fraction models, the former is considered in our study. Boag[2] first proposed mixture cure fraction model (MC) to analyze breast cancer data. MC model assumes the population is a mixture of two groups cured (will never face the event of interest) and uncured (will face the event of interest). It consists of two components, incidence and latency. The former illustrates whether the subject is cured or not and the later represents the time of event for uncured group. Several parametric and semi parametric models have been presented to describe the time to event.

For example, in the parametric case, Farewell[3] applied Weibull distribution to analyze long term survivors for patients with three levels of zinc concentration. Yamaguchi[4] used the accelerated failure with generalized Gamma distribution to analyze the permanent employment in Japan. Yu et al[5] make a comparison using different distributions and found that generalized Gamma distribution is quite robust and applied the models to analyze cancer data. Kannan et al[6] applied the generalized exponential cure rate model to study relapse time for drug addicts. Martinez et al[7] analyzed gastric cancer data using cure fraction models with generalized modified Weibull distribution. Swain et al[8] and Omer et al[9] used mixture cure model with generalized Gompertz and exponentiated Weibull exponential distributions, respectively, to analyze melanoma patients.

In the semi parametric and non parametric cases, for example, Kuk and Chen[10] proposed a semiparametric model using a Cox proportional hazard. Peng and Dear[11] applied cox proportional hazard model to analyze long term survivors for breast cancer data. Sy and Taylor[12] developed maximum likelihood techniques for the joint estimation of the incidence and latency parameters using the nonparametric form of the likelihood and an EM algorithm. Othus et al[13] introduced a semi parametric class that accounts for dependent censoring and use it to analyze prostate cancer data. Xu and Peng[14] proposed a nonparametric estimator to estimate the cure rate with covariates.

Survival analysis can be done using statistical models or neural networks. Recently, the latest has attracted attention in the literature. Faraggi-Simon network (Faraggi and Simon[15]) was first introduced as a nonlinear extension of Cox proportional hazard model. However, the network failed to outperform the traditional Cox model. Katzman et al[16] modified Faraggi-Simon network by applying new deep learning techniques which outperforms the traditional Cox model. Different networks were developed keeping the basic assumption of proportional hazards, see for example Zhu and Huang[17] and Zhu et al[18]. Pawley[19] introduced a parametric neural network (DeepWeibull) based on Weibull distribution that allows to analyze continuous data. Zhao and Feng[20] introduced a novel neural network using pseudo value approach.

Although several neural networks were introduced to analyze survival data, only Xie and Yu[21] handled the case of cure fraction model. They introduced a neural network to estimate the probability of a subject in the non cured group, together with the EM algorithm to account for the latent variable in the proposed likelihood function. However, using EM

algorithm may be computationally expensive in high dimensional data or when bootstrap method is used in variance estimation (see, Su et al[22] for more details). Also, analyzing survival data requires a special cost function in neural network to deal with censoring (see, Zhao and Feng[20]). To solve this, we introduced a novel neural network using pseudo values to estimate the cure fraction models. The proposed neural network doesn't require the use of EM algorithm and special cost function. To the best of our knowledge, this is the first time a neural network using pseudo values is introduced to analyze cure fraction models.

Methods

We introduced a novel neural network based on pseudo values. It takes into consideration the long-term survival and efficiently analyze censored data without using special cost function.

We will first explain the likelihood function under survival analysis, illustrate the basic properties of cure fraction models and pseudo values. Then illustrate the derivation of the model.

Survival analysis

In this section, we illustrate the basic components of lifetime data, the main used functions and the form of the likelihood function under censoring.

Time to event data consists of three main elements: observed covariates (x), observed event time (T) and type of event indicator (δ) such that:

$$\delta = \begin{cases} 1, \textit{ if the event is observed} \\ 0, \textit{ if the event is censored}. \end{cases}$$

censoring occurs when time to event is not observed for all data units. Usually, the lifetime of censored data is only known to occur after the recorded time producing so called right censoring. The main concern in survival analysis is to study the relationship between different covariates and lifetime data taking into consideration the censored observations. To do so, one must reflect this in the used likelihood function.

For n data points under right censoring, the likelihood function can be expressed as follows

$$L_1 = \prod_{i=1}^{n} [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i}. \quad (1)$$

Where

S(t)=P(T>t), is the probability that an individual has survived up to time t.

f(t), is the probability density function which can be analyzed using parametric, non parametric or semi parametric as illustrated in section1.

This form of the likelihood function is written under the assumption that all units will face the event of interest if the study is long enough. To analyze data when this assumption is violated, we need to rewrite the likelihood function using cure fraction models which will be illustrated in the next section.

Cure fraction models

Cure fraction models handle the case of survival analysis when a portion of the population never face the event of interest. Such subjects are referred to as cured, immune or non-susceptible. The other part of the population is those who are subject to the event of interest, they are called non-cured or susceptible. To represent the two groups of the population, we define η as follows

$$\eta = \begin{cases} 1, & \textit{if the subject is suscepyible.} \\ 0, & \textit{if the subject is cured}. \end{cases}$$

Let P(η=1)=p, P(η=0)=1-p, F(t) be the cumulative distribution function of the entire population and F^* (t) is the cumulative distribution function of susceptible subjects. It is assumed that F^* (t) is a proper cumulative distribution function, thus

$$P(T \le t | \eta = 1) = F^*(t)$$
$$P(T \le t | \eta = 0) = 0.$$

Accordingly

$$F^*(t) = P(T \le t \,|\eta = 1)\, P(\eta = 1) + P(T \le t | \eta = 0)\, P(\eta = 0)$$
$$p\, F^*(t) + 0 = p\, F^*(t).$$

Thus, the survival function for the whole population S(t) can be written as follows

$$S(t) = 1 - F(t) = 1 - p\,F^*(t) = (1-p) + p\,S^*(t).$$

Where, S^*(t) is is the survival function of susceptible subjects.

The likelihood function in (1) can be rewritten under cure fraction model as follows

$$\prod_{i=1}^{n} [p_i\, f^*(t_i)]^{\delta_i} [(1 - p_i) + p_i\, S^*(t_i)]^{1-\delta_i}. \qquad (2)$$

Xie and Yu[21] used the neural network only in estimating p and then used the EM algorithm to account for the latency part. Instead of going through this process, one can use the pseudo values approach with no need to EM algorithm. Also, the cost function is handled as a simple regression problem by inherently accounting for censoring. This is explained in details in the next section.

Pseudo values

The pseudo observation approach presents a direct way to estimate the survival function and study the relation between survival probabilities and covariates taking into consideration the censoring information.

If there was no censoring and accordingly the data can be treated as complete, the survival probability P(T>t) can be obtained directly from the generalized linear model with a logit link. However, under censoring the survival probability is not observed for all items which leads to censoring phenomena. In this case, Kaplan-Meire estimator can be used to estimate the probability. The pseudo value approach applies Kaplan-Meier estimator to get the required estimates as illustrated in details in the upcoming paragraphs.

The pseudo value approach is based on the jackknife idea of leaving one out. The values are computed for all data (censored or uncensored) and then the data is treated as complete. It was first introduced by Andersen et al[23] to directly model the state probability in a multi-state model. After that, the pseudo values approach was used in survival analysis. For example,

Klein et al[24] applied it in analyzing survival curves at fixed point in time. Andrei and Murray[25] used it in regression models for the mean of the quality-of-life-adjusted restricted survival time. Nicolaie et al[26] applied this approach for prediction taking into consideration competing risks. Sabathe´ et al [27] used it in interval censored data. Su et al [28] applied it in the analysis of recurrent event data. Recently Su et al[22] introduced the usage of pseudo value approach in cure fraction.

Cure fraction models has two parameters to be estimated, uncured probability (p) and the survival function for susceptible subjects (S^* (t)). Su et al[22] introduced the pseudo values as

follows $\hat{p}_{KM}^{i} = n\,\hat{p}_{KM} - (n-1)\,\hat{p}_{KM}^{-i}$.

n: the number of observations.

$\hat{p}_{KM} = 1 - \hat{S}_{KM}(t_{max})$: is the Kaplan-Meier estimator for uncured rate.

$\hat{S}_{KM}$: is the Kaplan-Meier estimator.

t_{max} : is the maximum observed event time.

$\hat{p}_{KM}^{-i}$: the estimator of p when leaving subject i out of the sample.

The pseudo values for $S^{*}(t)$ is calculated as follows

$$\hat{S}_{KM}^{*i}(t) = n\,\hat{S}_{KM}^{*}(t) - (n-1)\,\hat{S}_{KM}^{*-i}(t).$$

Where

$$\hat{S}_{KM}^{*-i}(t) = \frac{\hat{S}_{KM}^{-i}(t) - \hat{S}_{KM}^{-i}(t_{max})}{1 - \hat{S}_{KM}^{-i}(t_{max})}:$$

is the estimator for $S^{*}(t)$ when leaving subject i out of the sample.

We will use the above methodology in preparing the data. Table 1 is an example output of a hypothetical subject. The subject is censored at time 0.286. The pseudo observation is calculated at 10 time points from the quantiles of observed event times between 0 and t_{max}. The pseudo values can be above 1 or below zero (see properties of pseudo values in Andersen and Perme[29]). The pseudo values for S will be used in the neural network illustrated in the next subject

Table 1. An example output of the Pseudo values.

	t	pseudo value for S
ID =1	0.109	1.013
	0.158	1.013
	0.207	1.013
	0.257	1.013
	0.306	0.844
	0.633	0.844
	0.961	0.844
	1.289	0.844
	1.617	0.844
	1.781	0.844

Model description

The goal of this model is to estimate the survival function for cure fraction model using the neural network with simple loss function and without EM algorithm. To do so, we will use the pseudo values illustrated in the previous section as the response variable. Hence, we can use the cost function as the mean square error between pseudo survival probabilities and predicted survival probabilities. First, a brief explanation of neural networks will be presented, then the structure of the presented network will be explained.

The main idea of neural network is to compute the output based on a functional relationship with the inputs. Neural network takes a weighted sum of the inputs with one additional term called a bias term illustrated as follows

Let $x_1, x_2, \ldots, x_n$ be a set of inputs with weights $w_1, w_2, \ldots w_n$, the neuron output is given by

$$z = b + \sum w_i x_i$$

Where: b is called a bias term

The output of the network depends on the objective of the study. So, the simple linear combination is not always the required output. Accordingly, instead of using z, a function $\sum = f(z)$ is considered and called an activation function. The choice of the activation function depends on the required range of the output. For example, if the output is a value between 0 and 1, sigmoid function could be a choice.

A neural network learns by adjusting the weights in order to minimize the observed errors. The prediction error is reflected through a function called the loss function. So, the network updates the weight by minimizing the loss function. Mean square error is a common choice for the loss function in most of the

applications. However, in survival analysis one needs to take censoring information into consideration. This is done usually through the likelihood function. So, the loss function is the negative of log likelihood function which is a special cost function to handle censoring. However, in our study, we applied pseudo values to account for censoring, so there is no need for a special cost function and mean square error can be used.

As illustrated in Figure 1, the introduced deep neural network takes as input the covariates, the time points at which the pseudo values are calculated and output the survival function. As the network directly estimates the survival function which is the main concern of practitioners, this makes it of great use in real life.

To select the hyper parameters, we tried several combinations and selected the one with the best performance. The neural network consists of two fully connected hidden layers of widths 1 and 2 of the covariates dimension with tanh activation function. The output layer has a sigmoid activation function to consist with the survival probability range [0,1]. The network was introduced shallow (only two layers) to avoid the overfitting problems for small datasets.

For training, we use back-propagation via the Adam optimizer. Each of the hidden layers has Xavier initialization. The network is implemented in a TensorFlow environment with the Keras API.

For model evaluation, both time dependent concordance index (C^{td})) and integrated brier score are considered and explained briefly as follows.

Time dependent concordance index (C^{td}): Is an index to evaluate the discrimination ability

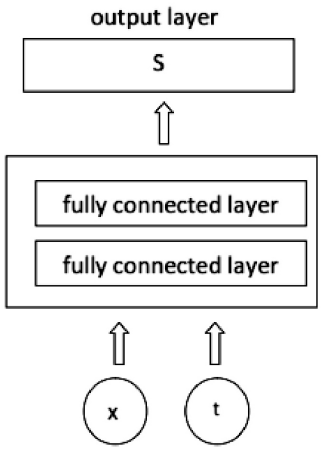

Figure 1. the architecture of the network

of survival model taking into consideration censored observations. The main advantage of C^{td}) over the usual concordance index is that there is no assumption for one-to-one correspondence between predicted survival probabilities and predicted times (i.e., no proportional hazard assumption). For a predicted survival probability $(\hat{S})$, the index is defined as follows

$$C^{td} = P(\hat{S}\left(T_i|X_i\left(t\right)\right) < \hat{S}\left(T_i|X_j\left(t\right)\right)|T_i < T_j \; and \; \delta_i = 1)$$

For more information see Antolini et al (23).

Brier score (BS(t)): Is a measure of the accuracy of probabilistic prediction. In survival analysis brier score is a measure of how well the model predicts the survival function. To account for censoring, inverse probability of censoring weighted Brier score is considered and defined as follows

$$BS\left(t\right) = \frac{1}{n}[\frac{\hat{S}_i\left(t\right)^2 .I\left(t_i \le t, \delta_i = 1\right)}{\hat{K}\left(t_i\right)} + \frac{\left[1-\hat{S}_i\left(t\right)^2\right].I\left(t_i > t\right)}{\hat{K}\left(t_i\right)}]$$

Where

n: Is the number of observations in the data.

$\hat{S}_i\left(t\right)$: is the predicted survival probability and i=1,…,n.

$\hat{K}\left(t_i\right)$: The estimated Kaplan-Meier survival

function.

However, brier score only gives a predictive performance at a given time point t. To overcome this disadvantage. Integrated brier score is introduced to average the brier score over time interval. It has the following formula

$$IBS(t) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} BS(t)\, dt$$

Usually, T_1 is set to zero and T_2 is the maximum value of t_i. For more details, see Pawley[19] and Håvard and Ørnulf.[30]

The model is examined on a synthetic and a real datasets. A brief description of the real dataset is given below. To evaluate the performance of the proposed network, we compare the results with the usual statistical cure fraction model using EM algorithm. The results illustrate that the proposed network has a good performance. However, as any neural network there is a heavier computational cost than the traditional statistical models.

Melanoma data e1684

e1684 dataset is an open access dataset freely available on R package. It aims to study melanoma disease for 155 patients below age 50 and the remaining at age 50 or above. Patients have been monitored for a range of 0.6 to 9.6 years, with a median follow up time among surviving participants more than 6.9 years (for more details, see Kirkwood et al (31)). In our study, we are interested in time till relapse of melanoma disease. Since not all participants will face the event of interest, we refer to this group as "cured".

Our study consists of 284 participants. Table 2 gives an overview of the dataset. The data consists of 3 features (gender, treatment and age). To test for the applicability of cure fraction models, one usually examines the Kaplan–Meier curve. If there is a long plateau at the later part of the curve, then there may be a subgroup of cured subjects. From Figure 2, It can be seen that there is a long plateau in the Kaplan-Meier curve. Hence, we can use cure fraction model to analyze the data.

Table 2. Descriptive statistics of melanoma dataset

	No. of uncensored	No. of censored	No. of features
Melanoma	196	88	3

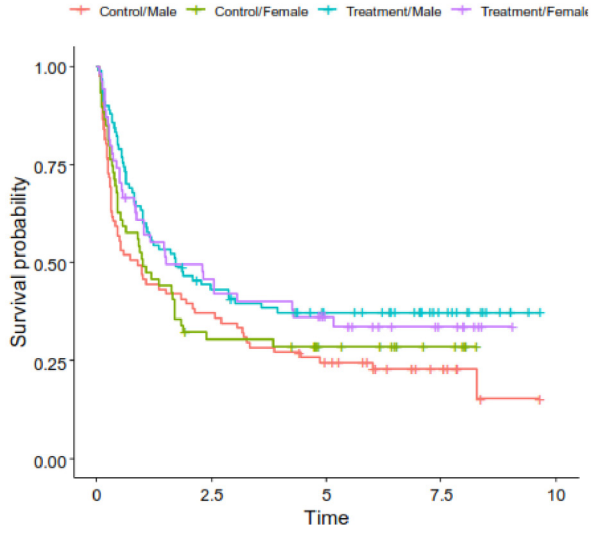

Figure 2. Kaplan-Meier curve for malenoma dataset

Synthetic

We generated 4 synthetic datasets using "penPHcure" library in R package that allows for data generation from mixture cure fraction model. The data consists of 4 covariates with different sample sizes and censoring percentages. From figures 3 till 6 it can be seen that there is a long plateau in the Kaplan–Meier curve. Hence, cure fraction model can be used to analyze the dataset.

In the analysis we compare the performance for the presented neural network and the cox proportional hazard model with EM algorithm that is heavily used in the literature. We will illustrate the evaluation techniques used in more details in the next subsection.

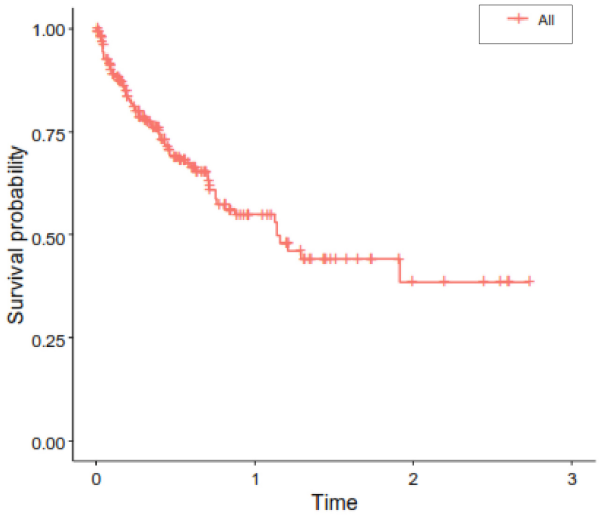

Figure 3. Kaplan-Meier curve for synthetic dataset (N=200, cen=0.7)

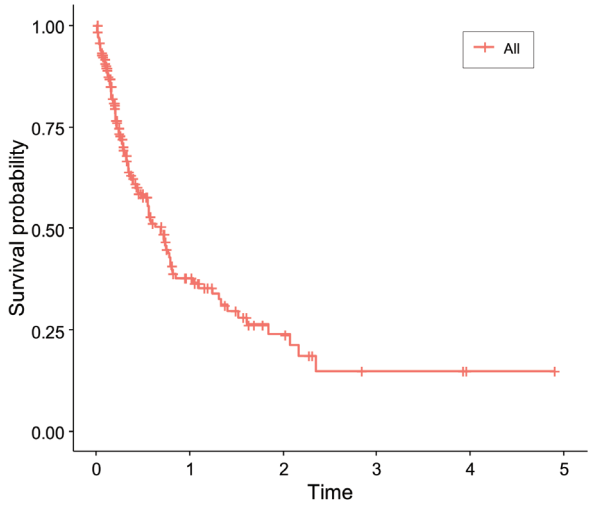

Figure 4. Kaplan-Meier curve for synthetic dataset (N=200, cen=0.5)

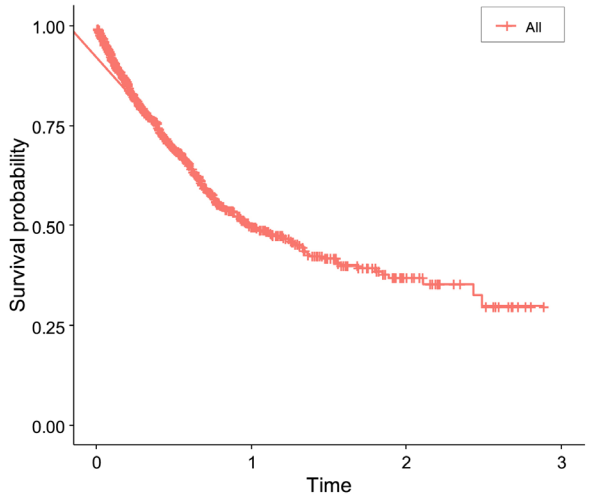

Figure 5. Kaplan-Meier curve for synthetic dataset (N=1000, cen=0.7)

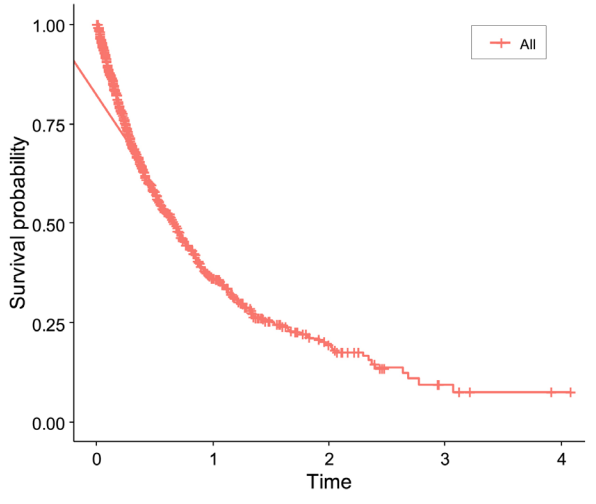

Figure 6. Kaplan-Meier curve for synthetic dataset (N=1000, cen=0.5)

Results

For evaluation, we split the data into two randomly train/test datasets, with 80 % of the units in each training set and the remaining 20% in the test set. To test the performance of the model, we used time dependent concordance index (C^{td}) and integrated brier score.

We applied the neural network using tenserflow environment with keras API. We used learning rate equals 0.0025, 1000 epochs in synthetic dataset, 100 epochs in real dataset and batch size equals 64. These hyper parameters are selected after trying several combinations and we selected the one with the best performance. Figures 4 till 8 illustrates the performance of the loss function through epochs for synthetic and melanoma datasets, respectively. It can be seen that; the loss function is minimized through our training model. For cox proportional hazard model, we applied the EM algorithm with 1000 and 100 iterations for synthetic and real datasets, respectively.

The concordance index and the integrated brier score for synthetic and real datasets are illustrated in Table 3. High C^{td} and low integrated brier score indicate better model performance in learning the patients' survival distribution.

In synthetic datasets, it is clear that the proposed neural network (NN) has higher C^{td} and slightly lower integrated brier score than that of the cox proportional model with EM algorithm (CP) except for the case with N=200 and censoring level is 50%, cox proportional model has lower brier score. Also, in melanoma data set, it can be seen that NN has much higher C^{td} than that for CP, while the integrated brier score is very

Table 3. Concordance Index (integrated brier score) for synthetic and melanoma datasets.

		Size	Censoring	Neural network	CP
Datasets	Synthetic	N=200	0.7	0.947 (0.277)	0.571 (0.279)
			0.5	0.760 (0.209)	0.649 (0.187)
		N=1000	0.7	0.859 (0.185)	0.791 (0.267)
			0.5	0.970 (0.194)_	0.658 (0.207)
	Melanoma	N=284	0.31	0.959 (0.235)	0.545 (0.229)

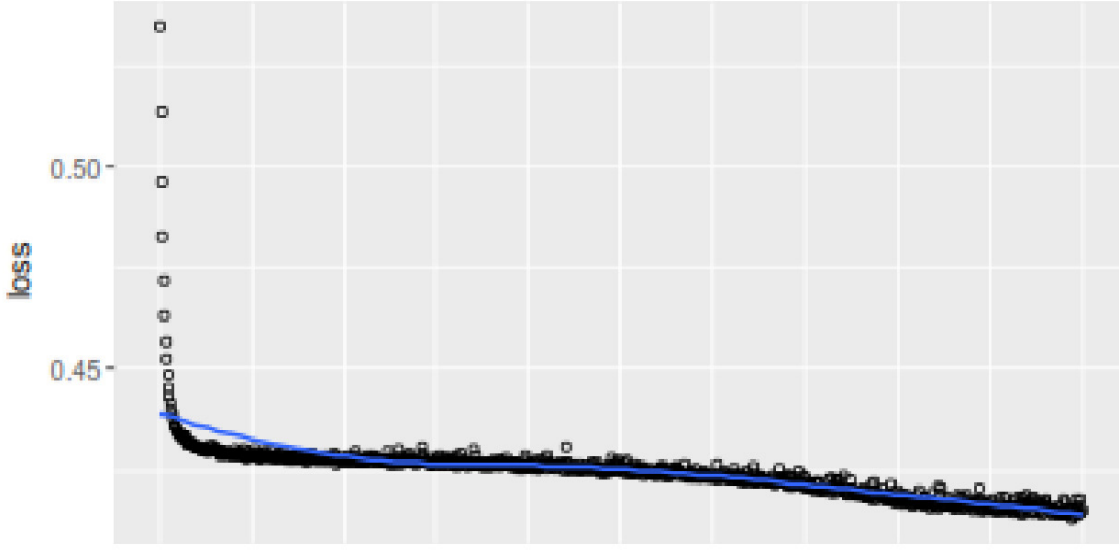

Figure 7. Loss function curve for synthetic dataset (N=200, cen=0.7)

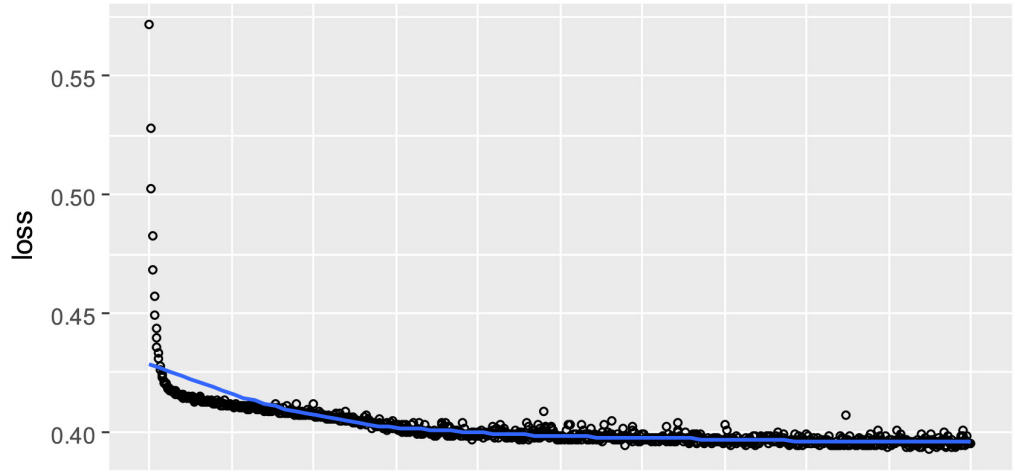

Figure 8. Loss function curve for synthetic dataset (N=200, cen=0.5)

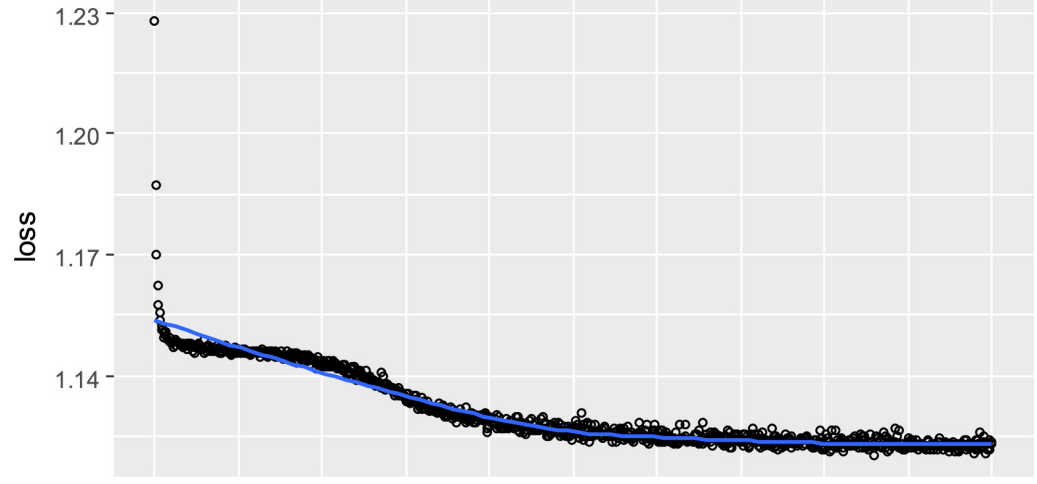

Figure 9. Loss function curve for synthetic dataset (N=1000, cen=0.7)

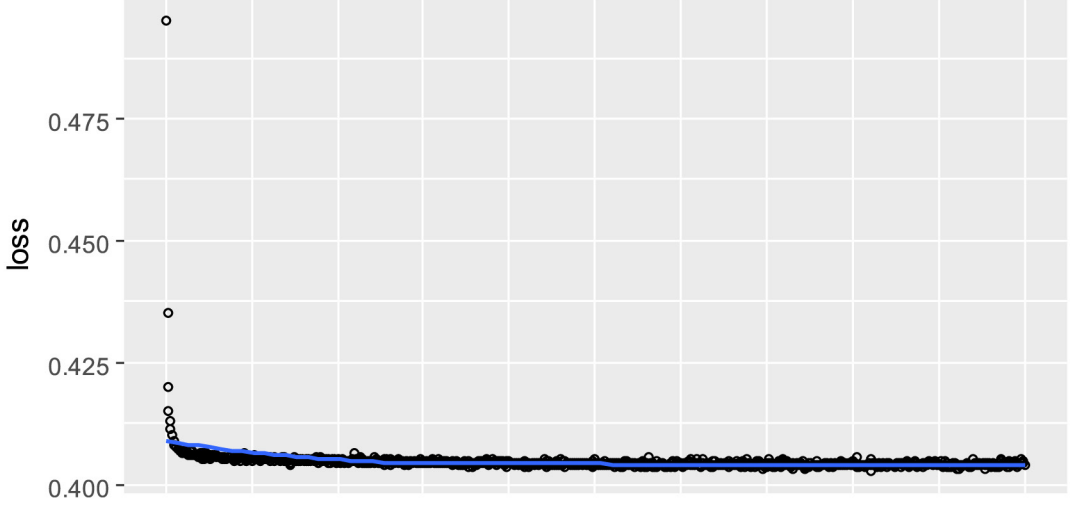

Figure 10. Loss function curve for synthetic dataset (N=1000, cen=0.5)

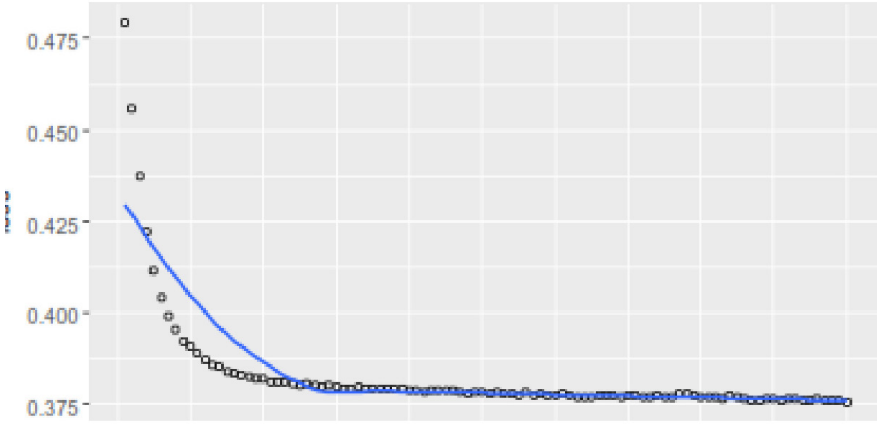

Figure 11. Loss function curve for Melanoma dataset

close to that of the CP model.

Discussion

Lifetime models have a hidden assumption of the occurrence of the event of interest for all observations. However, in some studies like Alzheimer, diabetic, Melanoma and cancer, part of the population may never face the event of interest. In this case, one need to take into consideration this information to assure accurate results.

In our study, a novel cure fraction pseudo neural network is introduced. It has the advantage of analyzing lifetime data without the need of special cost function. Also, we took into consideration the effect of long-term survivals. Besides, no need for applying EM algorithm.

To evaluate the network performance, both discriminative ability and accuracy are tested using time dependent concordance index and integrated brier score, respectively. The analysis was performed on both synthetic and real datasets. Comparing the results of the proposed network with cox proportional hazard model using EM algorithm, the former performed better.

Conclusion

This paper presents a novel model to be used in the mixture cure fraction survival analysis. It trains a neural network to estimate the survival function using pseudo values. To the best of our knowledge, this is the first time pseudo observation is used in neural network to analyze lifetime data with cure subgroup. The proposed neural network has the advantage of analyzing censored data without the need of special cost function. Also, it has the advantage of accounting for the latency part in mixture cure fraction model without the need of EM algorithm. As a test, we compared the performance of the proposed network with the performance of cox proportional hazard model with EM algorithm. The numerical results showed that, the presented neural network performed better.

Conflicts of interest

The authors declare that they have no conflicts of interest.

References

1. Amico, M. and Keilegom, I., V. Cure models in survival analysis. Annual Review of Statistics and Its Application. 2018; 5: 18.1-18.32.

2. Boag, J., W. Maximum likelihood estimates of patients cured by cancer therapy. Journal of the Royal Statistical Society. Series B. 1949; 11(1): 15-53.

3. Farewell, V., T. The use of mixture models for the analysis of survival data with long-term survivors. Biometrics. 1982; 38: 1041-1046.

4. Yamaguchi, K. Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of "permanent employment" in Japan. Journal of the American Statistical Association. 1992; 87(418): 284-292.

5. Yu, B., Tiwari, R., C., Cronin, K., A. and Feuer, E. J. Cure fraction estimation from the mixture cure models for grouped survival data. Statistics In Medicine. 2004; 23: 1733-1747.

6. Kannan, N., Kundu, D., Nair, P. and Tripathi, R., C. The generalized exponential cure rate model with covariates. Journal of Applied Statistics. 2010; 37(10): 1625-1636.

7. Martinez, E., Z., Achcar, J., A., Jacome, A., A. and Santos, J., S. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application on gastric cancer data. Computer Methods And Programs In Biomedicine. 2013, 112(3); 343-355.

8. Swain, P., K., Grover, G. and Goel, K. Mixture and non mixture cure fraction models based on generalized gompertz distribution under Bayesian approach. Tatra Mountains Mathematical Publication. 2016; 66: 11-135.

9. Omer, M., Abu Bakar, M., Adam, M. and Mustafa, M. Cure models with exponentiated Weibull exponential distribution for the analysis of melanoma patients. Mathematics. 2020; 8(11), 1926; https://doi.org/10.3390/math8111926.

10. Kuk, A., Y.,C. and Chen, C. A mixture model combining logistic regression with proportional hazards regression. Biometrica. 1992; 79(3), 531-541.

11. Peng, Y. and Dear, K., B., G. A nonparametric mixture model for cure rate estimation. Biometrics. 2000; 56, 237- 243.

12. Sy, P. G. and Taylor, J., M., G. Estimation in a Cox proportional hazards cure model. Biometrics. 2000; 56, 227- 236.

13. Othus, M., Li, Y. and Tiwari, R. C. A class of semiparametric mixture cure survival models With dependent censoring. Journal of the American Statistical Association. 2009; 104 (487), 1241- 1250.

14. Xu, J. and Peng, Y. Nonparametric cure rate estimation with covariates. The Canadian Journal of Statistics; 2014; 42(1), 1-17.

15. Faraggi, D., and Simon, R. A neural network model for survival data. Statistics in Medicine. 1995; 14: 73-82.

16. Katzman, J., Shaham, U., Cloninger, A., Bates, J., Jiang, T. and Kluger,'Y. Deep Survival: A Deep Cox Proportional Hazards Network. Stat. 2016, 1050 (2), 1–10; https://doi.org/10.1186/s12874-018-0482-1.

17. Zhu, X., Yao, J and Huang, J. Deep Convolutional Neural Network for Survival Analysis with Pathological Images. IEEE International Conference on Bioinformatics and Biomedicine. 2016: DOI: 10.1109/BIBM.2016.7822579.

18. Zhu,X., Yao, J., Zhu, F. and Huang, H. WSISA: Making Survival Prediction from Whole Slide Histopathological Images. IEEE Conference on Computer Vision and Pattern Recognition. 2017: DOI: 10.1109/CVPR.2017.725.

19. Pawley, M. DeepWeibull: a deep learning approach to parametric survival analysis. M.Sc. Thesis, Departement of Mathematics, Imperial College London. 2020.

20. Zhao, L. and Feng, D. Deep Neural Networks for Survival Analysis Using Pseudo Values. IEEE Journal of Biomedical and health informatics. 2020; 24(11), 3308 - 3314.

21. Xie, Y. and Yu, Z. Mixture cure rate models with neural network estimated nonparametric components. Computational Statistics. 2021; 36, 2467-2489.

22. Su, C., Chiou, S., H., Lin, F. and Platt, R., W. Analysis of survival data with cure fraction and variable selection: A pseudo-observations approach. Statistical Methods in Medical Research. 2022; 31(11), 2037- 2053.

23. Andersen, P., K., Klein, J., P. and Rosthoj, S. Generalized linear models for correlated pseudo- observations with applications to multi-state models. Biometrika. 2003; 90(1), 15-27.

24. Klein, J., P., Logan, B., Harhoff, M. and Andersen, P., K. Analyzing survival curves at a fixed point in time. Statistics in Medicine. 2007; 26, 4505- 4519.

25. Andrei, A. and Murray, S. Regression models for the mean of the quality-of-life-adjusted restricted survival time using pseudo-observations. Biometrics. 2007; 63, 398- 404.

26. Nicolaie, M., A., Houwelingen, J., C.V., Witte, T. M., D. and Putter, H. Dynamic pseudo-observations: A robust approach to dynamic prediction in competing risks. Biometrics. 2013; 69, 1043- 1052.

27. Sabathe´, C., Andersen, P., K., Helmer, C., Gerds, T., A., Jacqmin-Gadda, H. and Joly, P. Regression analysis in an illness-death model with interval-censored data: A pseudo-value approach. Statistical Methods in Medical Research. 2019; 29(3), 752- 764.

28. Su, C., Platt, R., W. and Plante, J. Causal inference for recurrent event data using pseudo-observations. Biostatistics. 2020; 23(1), 189- 206.

29. Andersen, P., K. and Perme, M., P. (2010). Pseudo-observations in survival analysis. Statistical Methods in Medical Research. 2010; 19, 71- 99.

30. Håvard Kvamme and Ørnulf Borgan (2023). The Brier Score under Administrative Censoring: Problems and Solutions. Journal of Machine Learning Research.2023; 24(2): 1-26.

31. Kirkwood, j., M., Strawderman, M., H., Ernstoff, M., S., Smith, T., J., Borden, E., C. and Blum, R., H. (1996). Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial EST 1684. Journal of Clinical Oncology. 1996; 14(1), 7-17.