Abstract

Cox regression model is most often used in the analysis of survival data, when we want to know the relationship between the incidents by the predictor variables. But at the assumption of this regression are not met, another model is needed as an alternative. Neural networks are artificial nonparametric regression model is proposed as an alternative model because it requires only minimal assumptions. To compare these two models, made four schema data simulation using Monte Carlo simulation, then the censoring of data as much as 20%, 40%, 60% and 80%, further analysis of the prediction accuracy of both models using the concordance index, the result accuracy Artificial Neural Network model is better than the Cox regression model accuracy. so it can be concluded that the model of neural networks can be used as an alternative model in survival data analysis.

Keywords: Survival data, regression, cox regression, neural networks, concordance index.

1. Introduction

Survival analysis is a set of statistical procedures used to analyze the data in which variable to consider is the time to occurrence of an event (Kleinbaum and Klein, 2005). Survival analysis is also called time analysis of events (time to event analysis), in which the time until occurrence of an event that chill-called survival time or failure time(Danardono, 2012).

One objective analysis of survival data is to determine the relationship between the incident (time to failure) and predictor variables were measured at the time of the study. It can be made by regression method, one of which is a Cox regression. Model Cox stated hazard rate of one individual at time t with a known predictor variables. However, the use of regression Cox proportional hazard must meet, if this assumption is not met in the Cox regression model, mean linear components that make up the model in a variety of time is not appropriate, consequently Cox regression modeling was not right. According to Collet (1994), stated that if this assumption is not met means a linear component in Cox models change depending on the time and said non-proportional hazard. Therefore, to overcome this, it takes a more flexible model.

Model Artificial Neural Network (ANN) can be seen as a flexible model for multivariate nonlinear problems and have a lot of attention over the past few years as a mathematical tool that can be used to solve the problem of nonlinear regression or classification problems. Neural
Network Model potentially produce a more accurate predictor of survival time than traditional models, because one of the main advantages of artificial neural network model is able to dig up hidden information recorded without constraints on the properties of the data.

2. Methodology

2.1 Cox Regression

Cox regression was first introduced by Cox, is one of the survival analysis of the most frequently used. This regression has no assumptions about the nature and form in accordance with the normal distribution as in other forms of regression; the distribution used is in accordance with the dependent variable used. Cox Regression The general form is:

\[ h(t) = h_0(t) \exp(\beta_1 X_1(t) + \beta_2 X_2(t) + \ldots + \beta_p X_p(t)) \]  

To determine the best model, it is required coefficient estimates \( X_1, X_2, \ldots, X_p \) i.e. \( \beta_1, \beta_2, \ldots, \beta_p \), coefficient \( \beta \) in the proportional hazard model can be estimated using the maximum likelihood method.

\[ L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta' x_j)}{\sum_{l \in R(t_j)} \exp(\beta' x_l)} \]  

and its likelihood log function is

\[ \log L(\beta) = \sum_{j=1}^{r} \beta' x_j - \sum_{j=1}^{r} \log(\sum_{l \in R(t_j)} \exp(\beta' x_l)) \]

To obtain the estimator of the parameters used numerical iterative method (Newton-Raphson iteration) to obtain the values of these parameters are approximations;

\[ \beta' = \beta'^{-1} - H^{-1} \beta'^{-1} g(\beta'^{-1}), \quad l = 1, 2, \ldots \]

(3)

2.2 Artificial Neural Networks

Artificial neural network is an information processing system that has characteristics similar to biological neurons. Neural network contains a large number of processing elements called neurons. Each neuron is connected to one another by a link, any link associated with weight. Weight percentages information used by the network to solve the problem. The presence of hidden layer with nonlinear activation function provides the ability to solve more problems than can be solved only with a network that consists of input and output only. Therefore, we will use the model MLP neural network method for the analysis of survival data.

In these discussions, the activation function is given in the hidden and output layers. Then the nonlinear function can be written

\[ y_i = g_o(\beta_0 + \sum_{j=1}^{p} H_i \beta_j) + \epsilon_i \]  

\[ H_i = h_h(\alpha_0 + \sum_{h=1}^{n} X_i \alpha_h) \]

\[ y_i = g_o(\beta_0 + \sum_{j=1}^{p} \beta_j g_h(\alpha_0 + \sum_{h=1}^{n} X_i \alpha_h)) + \epsilon_i \]

By entering the sigmoid activation function then the above equation can be written as follows:

\[ y_i = [1 + \exp(-\beta_0 - \sum_{j=1}^{p} H_i \beta_j)]^{-1} + \epsilon_i \]

\[ = [1 + \exp(-\beta_0 - \sum_{j=1}^{p} \beta_j [1 + \exp(\alpha_0 + \sum_{h=1}^{n} X_i \alpha_h)]^{-1})]^{-1} + \epsilon_i \]

\[ = g(X_i, \beta, \alpha_1, \alpha_2, \ldots, \alpha_n) + \epsilon_i \]  

(5)
In MLP models, a problem often encountered is to control overfitting, but it can be overcome by giving a penalty is added to the optimization criteria. In this case the penalty imposed on the criterion of least squares to estimate the parameters are,

\[ E^* = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 + \lambda (\sum \beta_i^2 + \sum \alpha_j^2) \]

In the model likelihood is formulated with,

\[ E^* = -(log\text{likelihood}) + \lambda (\sum \beta_i^2 + \sum \alpha_j^2) \]

Penalty \( \lambda \) set between over and under fitting. The best value \( \lambda \) is range from 0,001 and 0,1 selected by cross validation. To obtain the right model, to match the number of hidden layer, then we have to minimize the error, using the model Likelihood, i.e.

\[ E^* = -(log(L)) + \lambda (\sum \beta_i^2 + \sum \alpha_j^2) \]

To identify the number of hidden layer and model selection, use Bayesian Information Criterion (BIC) and Mean Square Error (MSE). The best model is the model with the minimum result of these criteria.

3. Simulation

To compare the accuracy of the prediction model of neural network and Cox regression, simulation scheme set up four different data, based on Monte Carlo simulation. That is:

1. \( \lambda = \exp(X_1 + 0.2X_2), X_1 \sim Ber(0.5), X_2 \sim N(0,1) \)
2. \( \lambda = \exp(X_1 + 0.2X_2 + 0.2X_1X_2), X_1 \sim Ber(0.5), X_2 \sim N(0,1) \)
3. \( \lambda = \exp(X_1 + X_2 + 0.5X_1 + 0.5X_2 + 0.25X_1X_2 + 0.25X_1X_3), X_1 \sim Ber(0.25), X_2 \sim Ber(0.5), X_3 \sim N(0,1) \)
4. \( \lambda = \exp(2X_1 + X_2 + 0.5X_3 + X_1X_2 + X_1X_3 + X_1X_4 + 0.5X_2X_3 + X_1X_2X_3 + X_1X_3X_4 + 0.5X_1X_2X_3X_4), X_1 \sim Ber(0.25), X_2 \sim Ber(0.5), X_3 \sim N(0,1), X_4 \sim N(0,1) \)

In each scheme, Hazard function of time t is considered as an exponential form, namely \( \lambda \). Wherein for each scheme raised 1.000 data for any random variable. To obtain the prediction accuracy, each random sample was divided into two parts. The first part is the training group consisted of 700 observations, and the remaining 300 observations in the second group is the group testing. Furthermore, in all simulations, the average rate sensor is 20%, 40%, 60%, 80%. Based on the parameters of the exponential relationship between the independent variables, survival time was formed. After a survival time was obtained, censoring the right data, in the context if the data survival time \( t \) greater than quintile exponential function with parameters \( \lambda \), it is considered as a sensor, the process is repeated 100 times, from the simulation data that has been formed is processed to select the best model is to calculate the BIC and MSE, and results can be seen in the following table:

<table>
<thead>
<tr>
<th>M</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>20</td>
<td>40</td>
<td>60</td>
<td>80</td>
</tr>
<tr>
<td>H</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

Furthermore, for comparing model predictions neural network with Cox regression model, the concordance index was calculated. Concordance index is known as a generalization of the area below the characteristic curve for the data censored. This index shows that the proportion of cases were classified appropriately in the event of not censored (event) and censored and values between 0-1 indicates as the ability of the model’s accuracy.
Table 3.2 The average results for Concordance Index Calculation Simulation To Study Data Testing (300 cases with 1000 repetitions).

<table>
<thead>
<tr>
<th>model</th>
<th>JST</th>
<th>Reg. Cox</th>
<th>model</th>
<th>JST</th>
<th>Reg. Cox</th>
</tr>
</thead>
<tbody>
<tr>
<td>I:</td>
<td></td>
<td></td>
<td>III:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0.586</td>
<td>0.356</td>
<td>20%</td>
<td>0.607</td>
<td>0.269</td>
</tr>
<tr>
<td>40%</td>
<td>0.647</td>
<td>0.354</td>
<td>40%</td>
<td>0.684</td>
<td>0.269</td>
</tr>
<tr>
<td>60%</td>
<td>0.638</td>
<td>0.354</td>
<td>60%</td>
<td>0.718</td>
<td>0.268</td>
</tr>
<tr>
<td>80%</td>
<td>0.468</td>
<td>0.352</td>
<td>80%</td>
<td>0.548</td>
<td>0.263</td>
</tr>
<tr>
<td>II:</td>
<td></td>
<td></td>
<td>IV:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>0.578</td>
<td>0.346</td>
<td>20%</td>
<td>0.722</td>
<td>0.154</td>
</tr>
<tr>
<td>40%</td>
<td>0.649</td>
<td>0.346</td>
<td>40%</td>
<td>0.7</td>
<td>0.159</td>
</tr>
<tr>
<td>60%</td>
<td>0.645</td>
<td>0.346</td>
<td>60%</td>
<td>0.808</td>
<td>0.149</td>
</tr>
<tr>
<td>80%</td>
<td>0.477</td>
<td>0.344</td>
<td>80%</td>
<td>0.515</td>
<td>0.142</td>
</tr>
</tbody>
</table>

From the table 3.2 above, can be seen the value of test results of prediction accuracy using the concordance index for Cox regression models, only between 0.142-0.356 (close to the value of 0.5), which means that the prediction accuracy of Cox regression models are less good. While the value of the test results of prediction accuracy of ANN models ranged between 0.488 to 0.808 (close to the value 1) which means that the prediction accuracy of ANN models is very good if compared with Cox regression models. In other words, the more complex the model, the prediction accuracy of the model JST is getting better.

4. Conclusion

In this study used two models to analyze survival data with different levels of sensor, namely Cox Regression and Neural Network. The data used in this study is data simulation by using Monte Carlo simulation, where the data is formed of four different data models with a level sensor 20%, 40%, 60%, and 80% of the data were used to compare the accuracy of prediction models Regression Cox and Neural Network. From the test results of prediction accuracy using the concordance index, the value of concordance index of models of neural networks are much better than the model Regression Cox, other than that the more complex the model the better the prediction accuracy of model of neural networks, it is inversely proportional to Cox regression models. However, in the case of Cox regression model estimation is still better than the Artificial Neural Network model, especially for simple models, because the components of non lineara on Artificial Neural Network model are more complex. Finally has found that there are several alternative methods that flexible that it can be used for the analysis of survival data one model of Neural Network.

5. References

Fox. (2002). Cox Proporsional Hazard Regresion for Survival Data (Appendix to an R and S-Plus companion to Aplied Regression)
Qudratullah. (2007). *Bayesian Information Criterion (BIC) dalam pemilihan model terbaik Feed Forward Neural Network (FFNN)*, Tesis Universitas Gadjah Mada, Yogyakarta

Raykar., et al. (no year). *On Rangking In Survival Analysis: Bounds on the Concordance Index*.