Computers in Biology and Medicine 100 (2018) 203–208

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

The $L_{1/2}$ regularization network Cox model for analysis of genomic data

Hong-Kun Jiang, Yong Liang*

Faculty of Information Technology & State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, 999078, China

ARTICLE INFO

Keywords:
Cox proportional hazards model
$L_{1/2}$ penalty
Network

ABSTRACT

Methods based on a $L_{1/2}$ penalty have been utilized to solve the variable selection problem associated with the Cox proportional hazards model. One limitation of the existing methods for survival analysis is that these ignore the regulatory networks and pathways information. To merge prior pathway information into the analysis of genomic data, we proposed a network-based regularization method for the $L_{1/2}$ penalty and applied it to high-dimensional survival analysis data. This method used a $L_{1/2}$ regularized solver and network that penalizes a Cox proportional hazards model with respect to the sparsity of the regression and the smoothness between the coefficients in a given network. Based on the limited simulation studies and real breast cancer gene expression datasets, the experimental results showed that our method achieves a higher predictive accuracy than previous methods. Even though fewer genes were selected compared to those using previous methods, results showed stronger associations with cancer. The results of the analysis were also validated using GeneCards.

1. Introduction

With the development of technology, thousands of gene expression levels are able to be measured. The Cox model is essentially a regression model commonly used statistical in medical research for exploring the relationship between the survival of a patient and several variables. In the study of the dependence of survival time T on covariates $X = (x_1, ...x_p)^T$, the Cox proportional hazards model [1] includes a hazard function $h(t|X)$ of a subject with covariates z of the form

$$h(t|\mathrm{X}) = h_0(t)exp(\beta^T \mathrm{X}) \quad (1)$$

where $h_0(t)$ is a baseline hazard function and $\beta = (\beta_1, ... \beta_p)^T$ is regression coefficients.

Gene data has a feature of high-dimensional and small samples that can be a solution to the p≫n problem. Moreover, only a small portion of genes are disease-related among a vast number of genes, thus making the genes hard to find. Based on the above data features, the traditional methods [32] can not satisfy the features of high-dimensional gene data [38]. On the other hand, $Lq(0 < q \leq 1)$ regularizations have sparsity property and can prevent overfitting, so a series of regularization methods have been proposed to improve the performance of the high-dimensional Cox model. One of the popular directions is the Lasso method [2,3], and alternative methods include the elastic net [4] and Lasso network (L_1 + *Net*) [5], which could achieve some grouping and network effects. The series of the Lasso methods are based on the L_1 penalty. However, the L_1 penalty may not yield sufficiently sparsity variable selection in real applications.

In 2010, Xu et al. proposed $L_{1/2}$ regularization [6] and proved that the $L_{1/2}$ penalty is sparser than the L_1 penalty in regression models. Some researchers [7–9] have also used the $L_{1/2}$ regularization in the Cox model. However, the pathway information has been ignored in this existing $L_{1/2}$ method.

Here, we purposed a $L_{1/2}$ regularization network Cox model algorithm. $L_{1/2}$ regularization can be taken as a representative of L_q $(0 < q < 1)$ regularizations and demonstrated various attractive properties, such as unbiasedness, sparsity and oracle properties. Furthermore, the sparsity is better than that of L_1 regularization. Based on the $L_{1/2}$ penalized Cox regression model, the network structure is used to describe the gene pathways for gene expression data. Extensive simulation studies indicated that the effect of our method is better than existing methods in terms of variable selection accuracy and stability. The effectiveness of our method was further validated with real breast cancer gene expression data. In general, our method demonstrated a greater selectivity than the existing methods, including the ability to select a higher level of relevant genes.

* Corresponding author.
E-mail address: yliang@must.edu.mo (Y. Liang).

https://doi.org/10.1016/j.compbiomed.2018.07.009
Received 19 March 2018; Received in revised form 13 July 2018; Accepted 14 July 2018

2. Methods

2.1. Network-regularization

In this paper, we have n samples. Let $X = (x_1, \ldots x_p)^T$ be a vector of covariates and $\beta = (\beta_1, \ldots \beta_p)^T$ is the regression coefficient vector of p variables. The Cox partial log–likelihood can be written as

$$\ell(\beta) = \sum_{i=1}^{n} \delta_i \left\{ x_i^T \beta - \log\left(\sum_{j \in R_i} \exp(x_j^T \beta) \right) \right\} \tag{2}$$

where R_i is the set of the individuals at risk at time t_i.

The network-constrained estimation of the regression coefficients [2,10,11] denoted by $\hat{\beta}$

$$\hat{\beta} = \arg\min\{Q(\beta)\} \tag{3}$$

where $Q(\beta) = -\frac{1}{n}\ell(\beta) + P(\beta)$.

The $P(\beta)$ is the penalty function to obtain a sparsity estimate with the structure of a given network and it can be written as

$$P(\beta) = \lambda_1 \|\beta\|_{\frac{1}{2}} + \lambda_2 \beta^T L \beta \tag{4}$$

where $||\cdot||_{1/2}$ is a $L_{1/2}$ norm. The tuning parameters λ_1 and λ_2 control the amount of regularization for sparsity and smooth.

L_1 is a popular regularization technique that was used in previous method and it has less sparsity than L_q $(0 < q < 1)$. However, when q lies closer to zero, L_q experiences more sparser and thus more difficult to convergence. Some researchers have studied the properties of L_q $(0 < q < 1)$ regularization and proved that the $L_{1/2}$ regularization is extremely important and plays a special role [6,15,16]. Therefore, the $L_{1/2}$ regularization can be expressed as L_q $(0 < q < 1)$ regularization and also possesses unbiasedness and oracle properties [15,16]. In genetic data, only a few genes are relevant to diseases; therefore, in application, the $L_{1/2}$ penalty approach would be more appropriate than L_1 approach. Consequently, the $L_{1/2}$ penalty was chosen in our Cox regression model.

The graph structure has been represented the p-dimensional Laplacian matrix $L = \{l_{ij}\}$ when the network information of predictors [12,13] is provided as follows:

$$l_{ab} = \begin{cases} 1 & if \ \text{a= } b \text{ and } d_a \neq 0 \\ -(d_a d_b)^{-1/2} & if \ a \text{ and } b \text{ are linked with each other} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where d_a is the total number of links of the node i in graph theory. Based on the works of predecessors, the net penalty can be written as:

$$\beta^T L \beta = \sum_{a=1}^{p} \sum_{a \sim b} \left(\frac{\beta_a}{\sqrt{d_a}} - \frac{\beta_b}{\sqrt{d_b}} \right)^2$$

Li [13] earlier pointed out that the performance of the penalty function is reduced when two negatively correlated predictors are linked to each other. Subsequently, the adaptive Laplacian net has been proposed [5,13] as a solution to this problem. In this situation, the corresponding regression coefficients have different signs, so they are not expected to be locally smooth. In order to overcome this problem, the signs of coefficients are estimated first before adding it to the Laplacian matrix:

$$l_{ab}^* = \begin{cases} 1 & if \ a = b \ and \ d_a \neq 0 \\ -\text{sgn}(\beta_a^*)\text{sgn}(\beta_b^*)(d_a d_b)^{-1/2} & if \ a \ and \ b \ are \ linked \ with \ each \ other \\ 0 & otherwise \end{cases}$$

The net penalty can be written as:

$$\beta^T L^* \beta = \beta^T (S^T L S) \beta = (\beta^T S^T) L (S\beta)$$

where $L^* = \{l_{ab}^*\} = S^T L S$, with the $S = diag(sign(\beta^*))$ and $\text{sgn}(\beta^*)$ is the estimated sign of coefficients by preliminary regression analysis.

Based on Eq (5), the adaptive net penalty function can be write as:

$$\beta^T L^* \beta = \sum_{a=1}^{p} \sum_{a \sim b} \left(\frac{\text{sgn}(\beta_a^*)\beta_a}{\sqrt{d_a}} - \frac{\text{sgn}(\beta_b^*)\beta_b}{\sqrt{d_b}} \right)^2$$

The penalty function can be changed to:

$$P(\beta) = \lambda_1 \sum_{j=1}^{p} |\beta_j|^{\frac{1}{2}} + \lambda_2 \sum_{a=1}^{p} \sum_{a \sim b} \left(\frac{\text{sgn}(\beta_a^*)\beta_a}{\sqrt{d_a}} - \frac{\text{sgn}(\beta_b^*)\beta_b}{\sqrt{d_b}} \right)^2 \tag{6}$$

Based on $|\beta_a| \approx \text{sgn}(\beta_a^*)\beta_a$ for $\beta_a \approx \beta_a^*$, the Eq (6) can be rewritten as:

$$P(\beta) = \lambda_1 \sum_{j=1}^{p} |\beta_j|^{\frac{1}{2}} + \lambda_2 \sum_{a=1}^{p} \sum_{a \sim b} \left(\frac{|\beta_a|}{\sqrt{d_a}} - \frac{|\beta_b|}{\sqrt{d_b}} \right)^2 \tag{7}$$

2.2. The coordinate descent algorithm

The coordinate descent algorithm [13,14] is an efficient method for solving regularization models. Its basic procedure can be described as follows: for each coefficient, the target function with respect to β_i is partially optimized with the remaining elements of β fixed at their most recently updated values.

In previous work [14], the approximate ℓ was written by Taylor expansion at current estimates $\tilde{\beta}$ as that

$$l(\beta) \approx \tilde{l}(\beta) = \sum_{i=1}^{n} w_i (z_i - x_i^T \beta)^2 \tag{8}$$

Where $w_i = \ell''(x_i^T \tilde{\beta})$ and $z_i = x_i^T \tilde{\beta} - \frac{\ell'(x_i^T \tilde{\beta})}{\ell''(x_i^T \tilde{\beta})}$

The exact solution is

$$\hat{\beta}_j = \frac{HALF\left(\frac{1}{n} \sum_{i=1}^{n} w_i x_{ia} (z_i - \sum_{k \neq a} x_{ik}^T \beta_k) - \lambda_2 \sum_{a \sim b} \frac{|\beta_b|}{\sqrt{d_a d_b}}, \lambda_1 \right)}{1 + \lambda_2} \tag{9}$$

$HALF(\sigma, \gamma)$ is an enhanced $L_{1/2}$ thresholding operator [15,16] for the coordinate descent algorithm.

$$HALF\left(\sigma, \eta \right) = \begin{cases} \frac{2}{3}\sigma \left(1 + \cos\left(\frac{2\left(\pi - \phi_\eta(\sigma)\right)}{3} \right) \right) & \text{if } |\sigma| > \frac{\sqrt[3]{54}}{4}(\eta)^{\frac{2}{3}} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $\phi_\eta(\sigma) = across\left(\left(\frac{\eta}{8}\right)\left(\frac{|\sigma|}{3}\right)^{-\frac{2}{3}} \right)$.

Thus, the algorithm is:

Step 1 Initialize $\tilde{\beta}$.
Step 2 Compute w_i and z_i.
Step 3 Update $\hat{\beta}$ using Eq (9) and set $\tilde{\beta} = \hat{\beta}$.
Step 4 Repeat steps 2–4 until convergence of $\hat{\beta}$.

3. Results and discussion

3.1. Analyses of simulated data

The goal of this section is to evaluate the performance of our proposed method in the simulation study. We simulated gene expression data within an assumed network by works of predecessors. In model 1 and 2 [13,17,33], there were 100 networks respectively in model 1 and 2 where each has one transcription factor gene (TF) and 10 regulated genes, thus making a total of 1100 genes. We set $w_{ij} = 1$ between the TFs and their regulated genes and $w_{ij} = 0$ otherwise. The value of each TF was based on standard normal distribution and the 10 regulated gene values were generated from a conditional normal distribution as $N(\rho * x_{TFi}, 0.51)$. A total of 100 networks were divided into 25 groups. The correlation ρ between the regulated genes and TF in each group