DOI 10.3233/THC-174141
IOS Press

Robust sparse accelerated failure time model for survival analysis

Haiwei Shen, Hua Chai, Meiping Li, Zhiming Zhou, Yong Liang*, Ziyi Yang, Haihui Huang, Xiaoying Liu and Bowen Zhang
Faculty of Information Technology and State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Macau 999078, China

Abstract. To identify the bio-mark genes related to disease with high dimension and low sample size gene expression data, various regression approaches with different regularization methods have been proposed to solve this problem. Nevertheless, high-noises in biological data significantly reduce the performances of methods. The accelerated failure time (AFT) modelwas designed for gene selection and survival time estimation in cancer survival analysis. In this article, we proposed a novel robust sparse accelerated failure time model (RS-AFT) through combining the least absolute deviation (LAD) and Lq regularization. An iterative weighted linear programming algorithm without regularization parameter tuning was proposed to solve this RS-AFT model. The results of the experiments show our method has better performancebothin gene selection and survival time estimationthan some widely used regularization methods such as lasso, elastic net and SCAD. Hence we thought the RS-AFT model may be a competitive regularization method in cancer survival analysis.

Keywords: AFT, survival analysis, regularization

1. Introduction

Accurate estimation for the cancer patients' survival time with high dimension and low sample size gene expression dataset is a significant challenge in survival analysis. The efficient method which can identify the relevant genes associated with tumours may be helpful for cancer research and treatment. In the last two decades, the Cox proportional hazards (Cox) model with the regularization approach has been widely used for the patient risk classification and relevant biomarkers identification [1–3]. However the Cox proportional hazards model may not be suitable if the data does not meet the proportional hazards assumption. Meanwhile, the patient's survival time estimationhas become a very important requirement in clinical treatment. Hence the accelerated failure time (AFT) model has already become one succedaneum of the Cox model in cancer survival analysis. Nevertheless the small sample size limits the performance of AFT model construction, such as the censored data in the cancer clinical data cannot be directly used in the model training. To increase the number of available data in the AFT model, some different imputation methods were proposed in cancer survival analysis. The most widely used one is Buckley-James estimation method [4–6], it estimates the censored data using the Kaplan-Meier

*Corresponding author: Yong Liang, Faculty of Information Technology and State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau 999078, China. Tel.: +853 63869506; Fax: +853 88972034; E-mail: yliang@must.edu.mo.

approach; the other one is called ranking based estimator, it estimated the survival time from computing the score function of the partial likelihood [7,8]. In our proposed RS-AFT model, we used Kaplan-Meier approach [9] to deal with the censored data.

The traditional ordinary least squares (OLS) approach has been used to construct the prediction model for a long time. However the OLS approach is sensitive to noise in the data, which significantly reduces its robustness in the practical application. Meanwhile the OLS estimation cannot achieve an unbiased solution under some certain conditions, and its estimated variance is quite large [10]. To improve the performance of the OLS estimation, the robust regression and the regularization methods were proposed. The least absolute deviation (LAD) is the kind of the robust regression method to confront the noise. The regularization approaches are widely used for variable selection in high dimensional data analysis. To overcome the shortcomings in the OLS method, Li et al. [11] proposed a RLAD method that combines the robust regression and regularization approach together. After that the LAD-lasso [12] and LAD-Adaptive lasso [13] were implemented. However compared to the L_1 type regularization, the Lq ($0 < q < 1$) type regularization can obtain more sparse result, and it has some attractive properties, such as unbiasedness, oracle properties and consistency of variable selection [14,15]. Therefore, Chang et al. [10] proposed LAD-L q regularization, which outperforms some existing methods based on the OLS with L_1 type regularization approaches in variable selection.

Considering the high dimensional and low sample size data in cancer survival analysis, many different kinds of regularization methods were used as the penalty function to combine with the regression loss function, such as lasso [16], elastic net [17], the smoothly clipped absolute deviation (SCAD) [18], $\mathrm{L}_{1/2}$ regularization [19]. These methods help the model predict the objective function value and select the feature genes related to the disease; however we found it was a difficult work to get a balance between the prediction accuracy and sparsity. Usually high prediction accuracy means large numbers of the selected genes; it means people have to waste much time for researching some unrelated genes. We considered the LAD-Lq regularization, which have the advantages of LAD and Lq ($0 < q < 1$), was a good choice to instead of these old regularization methods. Hence we proposed a robust sparse AFT model with LAD-Lq regularization approach (RS-AFT), we thought the new model can generate good performances in survival time estimation, and it has a powerful ability to find the cancer related genes because of its sparsity.

2. Method

Supposing the dataset included n samples, $(y_i, \delta_i, x_i)_{i=1}^{n}$ represents the single patient's sample, where y_i is the observed survival time pf the patient, $\delta_i = 0$ represents the sample is the censored data and if $\delta_i = 1$ means the sample is the completed data, $x_i = (x_{i1}, \ldots, x_{ip})$ indicate the p dimensional covariates.

The AFT model can be written as a linear regression model: $h(y_i) = x_i^T \beta + \varepsilon_i, i = 1, \ldots, n$ where $h(.)$ is the log transformation or some other monotone functions, ε_i is the independent random error with a normal distribution function, and $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ is the regression coefficient vector of p variables.

For estimating the censored time, we used the Kaplan-Meier weights estimation method because of its simple and fast [6]. The estimated value $h(y_i)$ of the censored time y_i can be written as:

$$h(y_i) = (\delta_i)h(y_i) + (1 - \delta_i)\{\hat{S}(y_i)\}^{-1} \sum\nolimits_{t_{(r)} > t} h(t_{(r)}) \Delta \hat{S}(t_{(r)}) \tag{1}$$

where $\Delta\hat{S}(t_{(r)})$ is the step of at time $t(r)$ [20].

As we know, the least squares approach method is widely used to find β:

$$\beta_{LS} = \arg\min \sum_{i=1}^{n} (h(y_i) - x_i^T\beta)^2 \tag{2}$$

To overcome the shortcomings of least squares approach, especially for data X with high noise, the least absolute deviation (LAD) was adopted:

$$\beta_{LAD} = \arg\min \sum_{i=1}^{n} |h(y_i) - x_i^T\beta| \tag{3}$$

In fact, not all genes in the microarray dataset may be associated with the patient's survival time, which means some coefficients β may be zero in the true model. A good method should select bio-mark genes consistently and efficiently. Some regularization methods have been widely used to find the true disease related genes. The different penalty function regularized AFT model using LAD approach will be written as:

$$\beta_{LAD} = \arg\min \left\{ \sum_{i=1}^{n} |h(y_i) - x_i^T\beta| + \lambda P(\beta) \right\} \tag{4}$$

The AFT model with the LAD-lasso regularization approach is:

$$\beta = \arg\min \left\{ \sum_{i=1}^{n} |h(y_i) - x_i^T\beta| + \sum_{j=1}^{p} \lambda|\beta_j|^1 \right\} \tag{5}$$

Trying to get more sparse solutions, we proposed the robust sparse AFT model with the LAD-Lq approach (RS-AFT):

$$\beta = \arg\min \left\{ \sum_{i=1}^{n} |h(y_i) - x_i^T\beta| + \sum_{j=1}^{p} \lambda|\beta_j|^q \right\} \tag{6}$$

3. Algorithm

Solving RS-AFT model is a non-convex optimization problem. We designed the weighted iterative algorithm to solve it. The regularization part $|\beta|^q$ in the RS-AFT model can be replaced by the first-order Taylor expansion:

$$|\beta|^q \approx |\beta_0|^q + \frac{1}{|\beta_0|^{1-q}}(|\beta| - |\beta_0|) = \frac{|\beta|}{|\beta_0|^{1-q}} \tag{7}$$

The minimization problem of the RS-AFT model will be shown:

$$\beta = \arg\min \left\{ \sum_{i=1}^{n} |h(y_i) - x_i^T\beta| + \sum_{j=1}^{p} \frac{\lambda}{|\beta_{0,j}|^{1-q}}|\beta_j| \right\} \tag{8}$$

In the literature [21], the BIC method was used to select the optimal regularization parameter λ. The likelihood function of the posterior probability by BIC is given by:

$$l(\beta) = \sum_{i=1}^{n} |h(y_i) - x_i^T\beta| + \sum_{j=1}^{p} (\lambda|\beta_j|^q - \log(\lambda)\log(n) - \log(n)) \tag{9}$$

Minimizing $l(\beta)$:

$$\lambda_{n,j} = \frac{\log(n)}{|\beta_j|^q} \tag{10}$$

$|\beta_{LAD}|^q$ (β_{LAD} is obtained by the least absolute deviation of the AFT model) can be seen as the estimator of $|\beta|^q$, λ can be written as:

$$\lambda = \frac{\log(n)}{|\beta_{LAD}|^q} \tag{11}$$

Since the variable selection consistency of the $\mathrm{L}_q(0 < q \leqslant 1)$ method has been proved in [1], we simply set $q = 1$ in the weighted iterative algorithm for turning the parameter λ. The detail procedure of the weighted iterative algorithm for the RS-AFT model is given:

```
The weighted iterative algorithm for the RS-AFT model
Input: The training dataset $(Y, \delta, X)$
Output: The AFT estimator
Initialize $\beta = 0$ $(j = 1, \ldots, p)$, compute the $\beta_{LAD}$ by using the least absolute deviation in the AFT model;
Set $\beta = \beta_{LAD}, t = 1$;
while $\sum_{j=1}^{p} |\beta_{tj} - \beta_{(t-1)j}| < 10E - 5$ do
    Update $\beta$:
    $\arg\min \left\{\sum_{i=1}^{n} |h(y_i) - x_i^T \beta| + \sum_{j=1}^{p} \frac{\lambda}{|\beta_{0,j}|^{1-q}} |\beta_j|\right\}$;
    $t = t + 1$;
end while
return $\beta$.
```

4. Results

4.1. Simulation experiments

In this section, we compared the AFT models with four different regularization approaches (LAD-Lq, lasso, SCAD, elastic net (EN)). Firstly we generated the vectors of independent standard normal distribution $\gamma_0, \gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{ip}, (i = 1, 2, \ldots, n)$ and set $x_{ij} = \gamma_{ij}\sqrt{1-c} + \gamma_{i0}\sqrt{c}, (j = 1, \ldots, p)$, where c is the correlationcoefficient [22], the patient's survival time was computed as: $t_i = \exp(\sum_{j=1}^{p} \beta_{ij} x_{ij}), (j = 1, 2, \ldots, p)$. The number of the censored data was decided by the censored rate r, and the censored time e_i were determined from a random distribution accordingly. The observed survival time in the simulated data was defined as: $y_i = \min(t_i, e_i)$ & $\delta_i = I(y_{ti} \leqslant y'_{ti})$. To test the performances of the different methods in the noise environment, we calculated $y_i = y_i + s \cdot \varepsilon$, where s is the noise control parameters and ε is the independent random errors from N(0, 1). Finally the simulated data were represented as (y_i, δ_i, x_i).

We set the dimension of the simulated datasets $p = 1500$. The coefficients of the 10 genes in these 1500 genes were nonzero, and the coefficients β of the remaining 1490 genes are zero. The right censored rate $r = 30\%$. We set training sample size $n = 150$, the correlation coefficient $c = 0$, 0.3 and the noise control parameter $s = 0$, 0.3 respectively. Each result obtained by different method was tested on a dataset including 50 samples, and the final outcomes were averaged over 100 repeats in the programme.

In this article, we used four evaluation parameters to compare the performances of different methods, the sensitivity, specificity, efficiency and absolute error E. The sensitivity, specificity, and efficiency parameters were used to test the gene selection performance. Supposing true positive (TP) is the number of selected correct genes, true negative (TN) is the number of the irrelevant genes which are selected, false negative (FN) is the number of the related genes to the disease which are not selected, and the false positive (FP) is the number of the irrelevant genes which are not selected by different methods.

$$\mathit{Sensitivity} = \frac{\mathit{TP}}{\mathit{TP} + \mathit{FN}} \tag{12}$$

Table 1
Theperformanceofgene selection obtainedby different AFT methods

Control parameter	Number of total selected genes				Number of correct genes			
	RS-AFT	Lasso	SCAD	EN	RS-AFT	Lasso	SCAD	EN
$c = 0.3$, $s = 0.3$	28.42	73.60	39.41	115.85	8.15	8.21	8.11	8.55
$c = 0.3$, $s = 0.0$	18.24	49.03	26.97	81.43	8.79	8.85	8.82	9.17
$c = 0$, $s = 0.3$	20.47	55.17	28.78	87.42	8.60	8.65	8.64	8.93
$c = 0$, $s = 0.0$	13.87	36.47	20.19	55.38	9.15	9.18	9.12	9.42

Table 2
The gene selection performancesofdifferent methods in simulation experiments

Control parameter	Sensitivity				Specificity				Efficiency			
	RS-AFT	Lasso	SCAD	EN	RS-AFT	Lasso	SCAD	EN	RS-AFT	Lasso	SCAD	EN
$c = 0.3$, $s = 0.3$	0.815	0.821	0.811	0.855	0.986	0.956	0.979	0.928	0.287	0.111	0.206	0.074
$c = 0.3$, $s = 0.0$	0.879	0.885	0.882	0.917	0.994	0.973	0.988	0.952	0.482	0.181	0.327	0.113
$c = 0$, $s = 0.3$	0.860	0.865	0.864	0.893	0.992	0.969	0.986	0.947	0.420	0.156	0.300	0.102
$c = 0$, $s = 0.0$	0.915	0.918	0.912	0.942	0.997	0.982	0.993	0.969	0.659	0.252	0.452	0.170

$$Specificity = \frac{TN}{TN + FP} \tag{13}$$

$$Efficiency = \frac{selected\ correct\ genes}{total\ selected\ genes} \tag{14}$$

The absolute error E was computed to test the ability of survival time estimation:

$$E = \frac{\sum_{i=1}^{n} |y_{ei} - y_i|}{n} \; \& \; \delta_i = 1 \tag{15}$$

where the y_i is the survival time of patient i in the dataset, and the y_{ei} is the estimated survival time of the patient using our model.

Tables 1 and 2 show gene selection performances of different methods in the different parameter settings. We found that with the decreasing of the noise parameter s and the correlation coefficient c, the models' performances become better. In Table 1 the RS-AFT always selected the least disease related genes in different datasets. Conversely, the AFT model with elastic net invariably selected most genes. The number of total genes selected by AFT model with SCAD was more than our model but less than lasso. Compared the number of selected correct genes, the elastic net selected most correct genes because its largest number of selected genes; the number of correct genes selected by remain three methods were much closed.

In Table 2, elastic net obtained the highest sensitivity because it selected most correct genes, but the specificity of elastic net was lowest because most irrelevant genes. The values of specificity obtained by our model were much closed to 1, it means RS-AFT model rarely selected irrelevant genes, we can say most of the selected genes obtained by RS-AFT were correct. And also we found the gabs between the values of specificity obtained by RS-AFT and SCAD were very small. Compared the efficiency, it is easy to find the gene selection efficiency of RS-AFT was highest, it means the users can easily find the true disease related genes in the RS-AFT model selected genes. These above results indicate that compared the gene selection performance, our RS-AFT model was better than the AFT models with lasso, elastic net and SCAD, it can help researchers find the real bio-mark genes fast.

The absolute errors obtained by different methods in simulation experiments were shown in Table 3, we can find the absolute errors obtained by elastic net model were always biggest, the SCAD was better

Table 3
The absolute error E obtained by different methods

Control parameter	RS-AFT	Lasso	SCAD	EN
$c = 0.3$, $s = 0.3$	1.95	3.85	2.56	4.37
$c = 0.3$, $s = 0.0$	1.16	2.43	1.75	2.94
$c = 0$, $s = 0.3$	1.28	2.61	1.83	3.12
$c = 0$, $s = 0.0$	0.73	1.74	1.17	2.05

Table 4
The detail information of four real gene expression datasets used in the experiments

Dataset	No. of genes	No. of samples	No. of censored	No. of training	No. of testing
DLBCL (2002)	7399	240	102	168	72
DLBCL (2003)	8810	92	28	64	28
Lung cancer	7129	86	62	60	26
AML	6283	116	49	81	35

Table 5
The number of selected genes obtained by different AFT models on the real datasets

Dataset	RS-AFT	Lasso	SCAD	EN
DLBCL (2002)	58.52	131.26	73.70	168.43
DLBCL (2003)	29.84	83.28	30.61	109.71
Lung cancer	28.43	86.51	39.42	102.46
AML	39.11	110.37	68.57	152.83

than lasso, and the RS-AFT model achieved the smallest absolute error. Hence we thought our method has best performance in survival time estimation compared other three methods.

From the above discussion, we thought the RS-AFT model was a more appropriate approach for can survival analysis in the microarray gene expression data because of its good performance of gene selection, and the high estimation precision for the patients' survival time.

4.2. Real data experiments

In this section, different methods were applied to the four real survival microarray datasets respectively, Diffuse large B-cell lymphoma dataset (DLBCL) 2002, DLBCL (2003), Lung cancer dataset and AML dataset. The DLBCL 2002 contains about 240 lymphoma patients' information and was first published in [23] by Rosenwald. Each patient sample includes the expression data of 7399 genes and the observed survival or censored time. Compared to DLBCL2002, DLBCL2003 only have 92 samples about the lymphoma patient, but the number of observed genes increased to 8810 [24]. The lung cancer dataset was published by van Beer [25], it has 86 cancer patients' samples which each sample include 7129 genes. The AML dataset was first mentioned by Bullinger [26], and has 116 patients which contains 6283 genes. A brief introduction of these datasets is summarized in Table 4.

Trying to compare the performance of four different AFT models, two thirds of the samples in the real dataset were used for the training and the other samples were seen as the data. The regularization parameters of different methods are tuned by the 5-fold cross validation.

The relevant gene selection performances of different AFT models in the four real datasets were shown in Table 5. The number of genes selected by our RS-AFT model was the least. The results of the SCAD were second-least and closed to the results of RS-AFT model. The third-least one is the number of genes selected by AFT model with lasso. The number of genes selected by AFT model with EN was much

Table 6
Absolute error obtained by different AFT models on the real microarray datasets

Dataset	RS-AFT	Lasso	SCAD	EN
DLBCL (2002)	0.65	1.31	0.84	1.87
DLBCL (2003)	1.16	2.28	1.41	2.83
Lung cancer	1.84	3.30	2.66	4.18
AML	3.11	4.93	3.74	6.08

Table 7
The disease related genes selected by different AFT methods in lung cancer dataset

Rank	RS	Lasso	EN	SCAD
1	SMAD4	WWP1	TRA2A	WWP1
2	ENPP2	HUWE1	WWP1	TRA2A
3	TRA2A	TRA2A	CCL21	HUWE1
4	LLGL1	CCL21	HUWE1	CCL21
5	WWP1	ADM	ADM	ADM
6	DYNLT3	PBXIP1	RPL36AL	PHKG1
7	DOC2A	RPS29	HLA-C	HLA-C
8	HUWE1	TNNC2	PEX7	RPS29
9	TEK	DOC2A	ZNF148	DOC2A
10	PHKG1	HLA-C	INHA	ATRX
11	PFN1	HTR6	RPS29	ENPP2
12	RPL23	TFAP2C	DOC2A	ZNF148
13	ENPP2	ZNF148	SERINC3	TFAP2C
14	POLR2A	HUMBINDC	GNS	TNNC2
15	CFTR	RPL36AL	ATRX	RAD23B

more compared with the other three methods. It means the researchers will pay much time to eliminate the irrelevant genes.

Table 6 describes the averaged absolute error E obtained by different AFT models in four datasets. It was obviously the performance of SCAD was better than lasso and the elastic net achieved the biggest absolute errors. And we can get the same conclusion as in the simulation experiments: the RS-AFT model achieved highest estimation precision with the least errors, which are much smaller than other method.

Comparing the performances in Tables 5 and 6, the results proved our RS-AFT model both have better performances in gene selection and survival time estimation. These are very important considerations in disease research and clinical application in cancer survival analysis. Hence we thought our method is more competitive than other regularization methods.

5. Conclusion

For biological analysis of the results, 15 top-ranked genes selected by the different AFT methods in Lung cancer dataset were shown in Table 7. Compared with the other AFT models based on the least squares approaches with different regularization methods, the RS-AFT model selected some unique genes, such as SMAD4, ENPP2, LLGL1. SMAD4 belongs to the member of Smad family which is one kind of signal transduction proteins. The Smad family proteins play a key role in transmitting the TGF-beta signals from the cell-surface receptor to cell nucleus, mutation or deletion of SMAD4, which has been proved to lead to the pancreatic cancer [27]. We think it may be strongly associated with the lung cancer. ENPP2 is also known as ATX, this gene can stimulate the motility of tumour cells. The expression of ENPP2 has been found to be up regulated in some different kinds of cancers [28]. The protein encoded by the gene LLGL1 was said to be very similar to the tumour suppressor of drosophila which is a highly relevant gene to cancer [29]. What is more, some relevant genes selected by other AFT models with lasso, elastic net and SCAD, were also found by the RS-AFT, for example, TRA2A, WWP1, DOC2A and HUWE1. They are significantly associated to the lung cancer which has been discussed [30].

We also obtained the similar experimental results from the analysis of the other three real datasets. The biological analysis showed that the RS-AFT model not only can find the relevant genes which were selected by AFT models with other regularization methods, but also can find some unique genes, which were not selected by other AFT models but also significantly associated to disease. Hence, we can say the RS-AFT model may find the disease related genes accurately and efficiently.

The experiment results show that the RS-AFT model outperforms some existing survival estimation approaches. It can effectively select the bio-mark genes and estimate the patients' survival time accurately in high dimensional and low sample size biological datasets. With the less mark genes and accurate survival time prediction, this method will be a more practical tool for cancer research and treatment.

In the data experiments we found that large number of the censored data great effect the accuracy of the RS-AFT model. The more censored data, the more difficulty we get in the experiments. Hence in the future work, we will try to combine the RS-AFT model with some machine learning methods, such as some semi supervised methods, we thought they may have strong ability to against with the censored data, the more completed data will improve the accuracy of our RS-AFT model obviously.

Acknowledgments

This work is supported by the Macau Science and Technology Development Funds (Grand No. 003/2016/AFJ) from the Macau Special Administrative Region of China.

Conflict of interest

None to report.

References

[1] Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997; 16: 385-395.
[2] Gui J, Li H. Penalized Cox regression analysis in the high-dimensional and low-sample size setting, with applications to microarray gene expression data. Bioinformatics 2005; 21: 3001-3008.
[3] Liu C, et al. The L1/2 regularization method for variable selection in the Cox model. Appl Soft Comput 2014; 14(c): 498-503.
[4] Buckley J, James I. Linear regression with censored data. Biometrika 1979; 66: 429-436.
[5] Tsiatis A. Estimating regression parameters using linear rank tests for censored data. Ann Stat 1990; 18: 354-372.
[6] Huang J, Ma S, Xie H. Regularized estimation in the accelerated failure time model with high dimensional covariates. Biometrics 2006; 62: 813-820.
[7] Cai T, Huang J, Tian L. Regularized estimation for the accelerated failure time model. Biometrics 2009; 65: 394-404.
[8] Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. Biometrika 2003; 90: 341-353.
[9] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958; 53: 457-481.

[10] Chang XY, Xu ZB, Zhang H, et al. Robust regularization theory based on Lq ($0 < q < 1$) regularization: The asymptotic distribution and variable selection consistence of solutions (in Chinese). Sci Sin Mat 2010; 40(10): 985-998. doi: 10.1360/012010-77.

[11] Li W, Michael DG, Ji Z. Regularized least absolute deviations regression and an efficient algorithm for parameter tuning. In: Proceedings of the Six International Conference on Data Mining Washington, IEEE Computer Society 2006; 690-700.

[12] Wang H, Li G, Jiang G. Robust regression shrinkage and consistent variable selection through the lad-lasso. J Business Economic Statist 2007; 25: 347-355.

[13] Xu JF, Ying ZL. Simultaneous estimation and variable selection in median regression using lasso-type penalty. Ann Inst Stat Math 2010; 62: 487-514.

[14] Xu ZB, Zhang H, Wang Y, et al. L1/2 regularization. Sci China Ser F 2010; 53: 1159-1169.

[15] Chartrand R, Staneva V. Restricted isometry properties and nonconvex compressive sensing. Inverse Problem 2008; 24: 1-14.

[16] Rajaratnam B, Sparks D. Fast Bayesian lasso for high-dimensional regression. Statistics 2015.

[17] Jing LI, Wang J, Hui LI, et al. Selection and classification of elastic net feature with fused electroencephalogram features. Journal of Biomedical Engineering 2016.

[18] Miao L, Zhou J, Naylor C, et al. Application of penalized linear regression methods to the selection of environmental enteropathy biomarkers. Biomarker Research 2017; 5(1): 9.

[19] Liu C, Liang Y, Luan XZ, et al. The L1/2, regularization method for variable selection in the Cox model. Applied Soft Computing 2014; 14(1): 498-503.

[20] Datta S. Estimating the mean life time using right censored data. Stat Methodol 2005; 2: 65-69.

[21] Hurvich CM, Tsai CL. Regression and time series model selection in small samples. Biometrika 1989; 76: 297-307.

[22] Sohn I, Kim J, Jung SH, Park C. Gradient lasso for Cox proportional hazards model. Bioinformatics 2009; 25(14): 1775-1781.

[23] Rosenwald A, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma. N Engl J Med 2002; 346: 1937-1946.

[24] Rosenwald A, et al. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. Cancer Cell 2003; 3: 185-197.

[25] Beer DG, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002; 8: 816-824.

[26] Bullinger L, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004; 350: 1605-1616.

[27] Boone BA, et al. Loss of SMAD4 staining in pre-operative cell blocks is associated with distant metastases following pancreaticoduodenectomy with venous resection for pancreatic cancer. J Surg Oncol 2014; 110(2): 171-5.

[28] Umezu-Goto M, et al. Autotaxin has lysophospholipase D activity leading to tumor cell growth and motility by lysophosphatidic acid production. J Cell Biol 2002; 158(2): 227-33.

[29] Schimanski CC, et al. Reduced expression of Hugl-1, the human homologue of Drosophila tumour suppressor gene lgl, contributes to progression of colorectal cancer. Oncogene 2005; 24(19): 3100-9.

[30] Chai H, et al. The L1/2 regularization approach for survival analysis in the accelerated failure time model. Comput Biol Med 2014.