Gene Selection for Cancer Classification Using Improved Group Lasso

Juntao Li[1] Wenpeng Dong[1] Deyuan Meng[2] Huimin Xiao[3]

1. Henan Engineering Laboratory for Big Data Statistical Analysis and Optimal Control, School of Mathematics and Information Science, Henan Normal University, Xinxiang 453007, P.R.China
E-mail: juntaolimail@126.com
E-mail: wenpengdongmail@126.com

2. The Seventh Research Division, Beihang University (BUAA), Beijing 100191, P.R.China
E-mail: dymeng23@126.com

3. College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450002, P.R.China
E-mail: huiminxiao@126.com

Abstract: An improved group lasso is proposed for simultaneous cancer classification and gene selection. A new criterion is firstly proposed to evaluate the individual gene importance by using the conditional mutual information. Then the weights with biological explanation are constructed and the improved group lasso is presented. A blockwise descent algorithm for solving the proposed model is also developed. The experimental results on lung cancer and prostate cancer data sets demonstrate that the proposed method can effectively perform classification and gene selection.

Key Words: Cancer classification, Gene selection, Conditional mutual information, Group lasso.

1 INTRODUCTION

With the quick development of gene microarray technology, cancer classification based on microarray gene expression data, which is essential for cancer diagnosis and treatment, has gained much attention in machine learning and bioinformatics [1–5]. Biologists and medical scientists are interested in finding the important genes related to cancer. Hence, how to effectively identify a small number of discriminatory genes is a great challenge.

Statistical machine learning methods have been widely applied to cancer classification and gene selection. As the most popular method, support vector machine and its extensions [1, 2] have been successfully applied to gene selection for cancer classification. According to different penalty strategies, many new learning machines emerged. Tibshirani [3] proposed the lasso by using 1-norm penalty. Zou and Hastie [4] proposed the elastic net by combining 1-norm penalty and 2-norm penalty and so on. To select variables in groups, Yuan and Lin [5] proposed group lasso. It is noted that the group lasso can identify the important groups, but not differentiate the important variables from these selected groups. To produce the both groupwise sparsity and within group sparsity, Simon et al. [6] proposed sparse group lasso and developed the accelerated generalized gradient descent algorithm.

In recent years, researchers have proposed many adaptive shrinkage methods [7–11], which can adaptively select genes. In particular, Li et al. [8] proposed the partly adaptive elastic net by using the initial estimator. Fang et al. [10] proposed the adaptive sparse group lasso by using the group bridge estimator. Vincent et al. [11] proposed the multinomial sparse group lasso by using the least squares estimator. Note that the weights in the aforementioned methods are constructed by using statistical estimator. Hence, the constructed weights can not provide the biological explanation.

The information theoretic criterion quantifies the uncertainty among variables. It also can measure non-linear dependency in multidimensional feature space by exploiting information entropy. Especially, mutual information (MI) and conditional mutual information (CMI) from information theory have been successfully used to infer gene regulatory networks [12] and protein modulation [13] from gene expression data. Motivated by these methods, we use the conditional mutual information to construct the weights with biological significance and introduce the weights to sparse group lasso. Therefore, we propose the improved group lasso and develop the blockwise descent algorithm.

The rest of this article is organized as follows. Section 2 gives a brief description for research problem and preliminary. The improved group lasso and its corresponding solving algorithm are shown in Section 3. Simulation results on lung cancer and prostate cancer data sets are provided in Section 4. Finally, we conclude the paper with a brief summary in Section 5.

2 PROBLEM STATEMENT AND PRELIMINARIES

Given a training data set $\{(x_i, y_i)\}_{i=1}^n$, where $x_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$ is a multidimensional input vector with dimension p and $y_i \in R$ is the output response. Tibshi-

This work is supported by Natural Science Foundation of China (61203293, 61473010, 61374079, 60850004), the Beijing Natural Science Foundation (4162036), Key Scientific and Technological Project of Henan Province (122102210131), Program for Science and Technology Innovation Talents in Universities of Henan Province (13HASTIT040), Henan Higher School Funding Scheme for Young Teachers (2012GGJS-063). Backbone Teachers Program of Henan Normal University.

rani [3] proposed the popular lasso estimator

$$\min_{\beta}\left\{\sum_{i=1}^{n}(y_i-\beta^T x_i)^2+\lambda\left\|\beta\right\|_1\right\}, \tag{1}$$

where $\beta=(\beta_1,\cdots,\beta_p)^T$ is the regression coefficient of the variables, and $\|\beta_1\|=\sum_{j=1}^{p}\beta_j$ is called lasso penalty which stands for the 1-norm of the regression coefficients, λ is the regularization parameter whose size determines the sparsity of the solution. In fact, the lasso is degraded into an ordinary least square regression estimator when $\lambda=0$.

Note that variable selection typically amounts to the selection of important factors (groups of variables) rather than individual derived variables. Hence, it is interesting to find important explanatory factors in predicting the response variable, where each explanatory factor may be represented by a group of derived input variables. Following the aforementioned idea, Yuan and Lin [5] proposed the group lasso (GL) estimator

$$\min_{\beta}\left\{\frac{1}{2}\left\|y-\sum_{l=1}^{m}X^{(l)}\beta^{(l)}\right\|_2^2+\lambda\sum_{l=1}^{m}\sqrt{p_l}\left\|\beta^{(l)}\right\|_2\right\}, \tag{2}$$

where $X^{(l)}$ is the submatrix of X whose columns correspond to the predictors in group l, $\beta^{(l)}$ is the subvector of β corresponding coefficients in group l, p_l is the length of $\beta^{(l)}$, and λ is the regularization parameter. If the size of each group is 1, it will degenerate to the lasso.

Although the group lasso gives a sparse set of groups, it does not produce sparsity within a group. In order to deal with this problem, Simon et al. [6] proposed a sparse group lasso (SGL)

$$\begin{aligned}\min_{\beta}\Bigg\{&\frac{1}{2n}\left\|y-\sum_{l=1}^{m}X^{(l)}\beta^{(l)}\right\|_2^2+(1-\alpha)\lambda\\&\times\sum_{l=1}^{m}\sqrt{p_l}\left\|\beta^{(l)}\right\|_2+\alpha\lambda\left\|\beta\right\|_1\Bigg\},\end{aligned} \tag{3}$$

where λ and α are regularization parameters. $\alpha\in[0,1]$ determines a convex combination of the lasso and group lasso penalties. The sparse group lasso will be transformed into lasso when $\alpha=1$, and group lasso when $\alpha=0$. In fact, the sparse group lasso can produce both the groupwise sparsity and within group sparsity, i.e., the groupwise sparsity refers to the number of groups with at least one nonzero coefficient, and within group sparsity refers to the number of nonzero coefficients within each nonzero group. An algorithm to fit this model via accelerated generalized gradient descent is also proposed. This algorithm can also be used to solve the general form of the group lasso, with non-orthonormal model matrices. A publically available R implementation of this algorithm in the package SGL is available on request.

Although it can identify important variables within each selected group, the sparse group lasso adopts the same penalty coefficient to all variables without considering relative importance among genes in selected groups. To adaptively select grouped variables, the adaptive sparse group lasso [10] has been proposed by using weights to adjust penalty. However, the method only uses the statistical estimator to construct weights and less biological explanations. This paper is devoted to solving the aforementioned problem by introducing information theory.

In the following, we present some fundamental concepts of information theory [14] as preliminary knowledge. Let X, Y and Z be three discrete random variables, the entropy $H(X)$ of variable X is defined as:

$$H(X)=-\sum_{x\in X}p(x)log(x), \tag{4}$$

where $p(x)$ is the probability distribution of each x. The entropy of a random variable is a average measure of its uncertainty.

The conditional entropy $H(X|Y)$ is defined as:

$$H(X|Y)=-\sum_{y\in Y}p(y)\sum_{x\in X}p(x|y)logp(x|y), \tag{5}$$

where $p(x|y)$ is the conditional probability distribution. The conditional entropy $H(X|Y)$ represents the entropy of a random variable X conditional on the knowledge of another random variable Y.

Mutual information (MI) is introduced to measure the amount of information shared by two random variables and is defined as:

$$\begin{aligned}I(X;Y)&=H(X)-H(X|Y)\\&=\sum_{x\in X,y\in Y}p(x,y)log\frac{p(x,y)}{p(x)p(y)},\end{aligned} \tag{6}$$

where $p(x,y)$ denotes the joint probability of x and y.

Conditional mutual information (CMI) of variables X and Y, when Z is given, is defined by:

$$I(X;Y|Z)=H(X|Z)-H(X|Y,Z) \tag{7}$$

where CMI represents the amount of information shared by variables X and Y given variable Z.

Entropy and mutual information provide intuitive tools to measure the uncertainty of random variables and the amount of information shared by them. To calculate entropy and mutual information, Brown et al. [15] has developed the MIToolbox for C and MATLAB, which is available on-line at http://www.cs.man.ac.uk/∼pococka4/MIToolbox.html.

3 MAIN RESULTS

3.1 Improved Group Lasso

For microarry data, mutual information generally represents the degree of mutual dependence between X_i and X_j which respond to two different genes. Conditional mutual information measures conditional dependency between two genes when the other genes are given. Given gene k, we let

$$r_k=\frac{1}{(p-1)^2}\sum_{i=1,i\neq k}^{p}\sum_{j=1,j\neq k}^{p}I(X_i;X_j|X_k), \tag{8}$$

where X_i, X_j, X_k respectively represent the ith, jth, kth gene expression level among p genes. r_k represents the average of the shared information between all the other pairwise genes when gene k is fixed. Obviously, the bigger r_k is, the more relevant all the other pairwise genes will be. In fact, r_k can be viewed as the criterion to evaluate the individual importance of kth gene. When all the other pairwise genes are conditionally independent for the fixed k, we get $r_k = 0$.

According to the equation (8), we assume that p genes are divided into m no-overlapping groups. The following weight coefficients in the lth group are constructed by

$$w_k^{(l)} = r_k^{-1},\ k = 1, \cdots, p_l, l = 1, \cdots, m. \tag{9}$$

According to the above weight coefficients, we construct the following weight matrix

$$\begin{aligned} W = diag\{w^{(1)}, \cdots, w^{(m)}\} = diag\{w_1^{(1)}, \cdots, \\ w_{p_1}^{(1)}, \cdots, w_1^{(m)} \cdots w_{p_m}^{(m)}\}. \end{aligned} \tag{10}$$

Introducing the weight matrix into the sparse group lasso penalty, we propose the following improved group lasso penalty

$$\begin{aligned} (1-\alpha)\lambda \sum_{l=1}^{m} \sqrt{p_l}\|\beta^{(l)}\|_2 + \alpha\lambda\|W\beta\|_1 = (1-\alpha)\lambda \\ \times \sum_{l=1}^{m} \sqrt{p_l}\|\beta^{(l)}\|_2 + \alpha\lambda \sum_{l=1}^{m} \|w^{(l)}\beta^{(l)}\|_1. \end{aligned} \tag{11}$$

Applying the penalty (11) to squared error loss, we propose the improved group lasso (IGL)

$$\begin{aligned} \min_{\beta} \Bigg\{ \frac{1}{2n}\|y - \sum_{l=1}^{m} X^{(l)}\beta^{(l)}\|_2^2 + (1-\alpha)\lambda \\ \times \sum_{l=1}^{m} \sqrt{p_l}\|\beta^{(l)}\|_2 + \alpha\lambda \sum_{l=1}^{m} \|w^{(l)}\beta^{(l)}\|_1 \Bigg\}, \end{aligned} \tag{12}$$

where α and λ are the regularization parameters. Obviously, it will be transformed into sparse group lasso when the weight matrices $w^{(l)}, l = 1, \cdots, m$ are identity matrix.

Due to the group lasso penalty $\sum_{l=1}^{m} \sqrt{p_l}\|\beta^{(l)}\|_2$, IGL can automatically select the important groups, i.e., the groupwise sparsity. Since the weights which evaluate the importance of genes are introduced to the 1-norm penalty, IGL can adaptively identify the significant genes within the selected groups, i.e., within group adaptive sparsity.

3.2 Improved Blockwise Descent Algorithm

Note that the objective function in (12) is convex, Hence, the optimal solution of IGL is characterized by the subgradient equations. Let $\hat{\beta}$ be the optimal solution of the IGL. Considering the kth group, the solution $\hat{\beta}^{(k)}$ satisfies the following subgradient equatiuon

$$\begin{aligned} \frac{1}{n} X^{(k)T}(y - \sum_{l=1}^{m} X^{(l)}\hat{\beta}^{(l)}) = \sqrt{p_k}(1-\alpha)\lambda u^{(k)} \\ + \alpha\lambda w^{(k)} v^{(k)}, \end{aligned} \tag{13}$$

where u_k and v_k are subgradients of $\|\hat{\beta}^{(k)}\|_2$ and $\|\hat{\beta}^{(k)}\|_1$, respectively. From the [6], it is obvious that $u_k = \hat{\beta}^{(k)}/\|\hat{\beta}^{(k)}\|_2$ if $\hat{\beta}^{(k)} \neq 0$, otherwise $\|u_k\|_2 \leq 1$. $v_{kj} = \text{sign}(\hat{\beta}_j^{(k)})$ when $\hat{\beta}_j^{(k)} \neq 0$, otherwise $|v_{kj}| \leq 1$.

Let $\gamma_{(-k)}$ denote the partial residual of y, that is, subtracting all group fits other than group k

$$\gamma_{(-k)} = y - \sum_{l \neq k} X^{(l)}\hat{\beta}^{(l)}. \tag{14}$$

Similar to [6], a necessary and sufficient condition for $\hat{\beta}^{(k)} = 0$ is

$$\|S(X^{(k)T}\gamma_{(-k)}/n, \alpha\lambda w^{(k)} e_k)\|_2 \leq \sqrt{p_k}(1-\alpha)\lambda, \tag{15}$$

where e_k is a $k \times 1$ vector with each element 1, and S is defined as

$$(S(a,b))_j = \text{sign}(a_j)(|a_j| - b_j)_+. \tag{16}$$

The unpenalized loss function is denoted as

$$\ell(\gamma_{(-k)}, \beta) = \frac{1}{2n}\|\gamma_{(-k)} - X^{(k)}\beta^{(k)}\|_2^2. \tag{17}$$

The cyclical coordinate-wise algorithm is used to fit the proposed model within group. It is equivalent to minimize the following optimization problem

$$\ell(\gamma_{(-k)}, \beta) + (1-\alpha)\lambda\sqrt{p_k}\|\beta^k\|_2 + \alpha\lambda\|w^{(k)}\beta^{(k)}\|_1. \tag{18}$$

By using the majorization minimizing scheme at a point $\beta_0^{(k)}$, we can obtain

$$\begin{aligned} \ell(\gamma_{(-k)}, \beta^{(k)}) \leq \ell(\gamma_{(-k)}, \beta_0^{(k)}) + (\beta^{(k)} - \beta_0^{(k)})^T \\ \times \nabla\ell(\gamma_{(-k)}, \beta_0^{(k)}) + \frac{1}{2t}\|\beta^{(k)} - \beta_0^{(k)}\|_2^2, \end{aligned} \tag{19}$$

where t is sufficiently small and $\nabla\ell(\gamma_{(-k)}, \beta_0^{(k)})$ is the gradient taken over group k.

After some similar algebraic deductions, we get

$$\begin{aligned} \hat{\beta}^{(k)} = S\left(\beta_0^{(k)} - t\nabla\ell(\gamma_{(-k)}, \beta_0^{(k)}), t\alpha\lambda w^{(k)} e_k\right) \\ \times \left(1 - \frac{t\sqrt{p_k}(1-\alpha)\lambda}{\|S(\beta_0^{(k)} - t\nabla\ell(\gamma_{(-k)}, \beta_0^{(k)}), t\alpha\lambda w^{(k)} e_k)\|_2}\right)_+. \end{aligned} \tag{20}$$

Similar to [6], by cyclically iterating algorithm through the blocks, we can get the overall optimum of the IGL.

4 EXPERIMENTS

To illustrate the effectiveness of the proposed method, we conduct experiments on lung cancer data and prostate cancer data.

4.1 Experiments On Lung Data

The lung cancer data set consists of gene expression profiles of 12,626 genes for 197 lung tissue samples, which includes 139 adenocarcinomas(AD), 21 squamous cell carcinomas(SQ), 20 carcinoids (COID) and 17 normal lung (NL). The data set is

available on-line at http://www.broadinstitute.org/cgi-bin/cancer/publications/view/87. To distinguish lung adenocarcinomas from the normal lung tissues, we consider the diagnosis of lung cancer as a binary classification problem. Let 17 normal lung be positive class and 139 lung adenocarcinomas be negative class. Following the idea of [16], we reserve the 1000 most significant genes after the preprocessing. In the experiment, we compare

Table 1: Experimental results on lung cancer data over 10 runs (the standard deviations are reported in parentheses).

Method	Average classification accuracy	Average number of genes
GL	0.802(0.031)	67.30(3.13)
SGL	0.827(0.027)	58.30(2.71)
IGL	0.832(0.024)	56.10(2.19)

IGL with SGL in [6] and GL in [5] from the two aspects: average classification accuracy and gene selection performance. The lung data set is randomly divided into two parts: two-thirds for training and one-third for testing. To avoid the contingency of single experiment, each process is repeated 10 times. Experiment results are shown in Table 1.

As is shown in Table 1, IGL achieves higher classification accuracy than GL and SGL. Note that the standard deviation of average classification accuracy for IGL is also the smallest among three methods. Hence, it just shows that classification property is more stabler for the proposed method. In our experiment, IGL selects the least number of genes and achieves the smallest standard deviation of average number of genes among all the methods. It implies that the property of gene selection for the proposed model is the most stable among three methods. Due to the different number of the randomly selected genes, sometimes average number of genes will probably be a non-integer.

Some key genes for lung cancer selected by IGL are shown in Table 2, which are believed to be highly related to lung cancer. To illustrate the effectiveness, we search the related function of these genes in NCBI database. For example, VEGFA is upregulated in many known tumors and its expression is correlated with tumor stage and progression. TP53BP1 is important for transcription, DNA-templated and protein binding. PURA plays a vital role in transcription factor activity and RNA polymerase II distal enhancer sequence-specific binding. PRDX2 may have a proliferative effect and play a role in cancer development or progression. This also explains that these selected key genes are highly consistent with the actual function of genes.

Table 2: Some key genes selected in lung cancer data via IGL.

Gene symbol	Gene title	Possible function
VEGFA	vascular endothelial growth factor A	VEGFA is a cellular response to vascular endothelial growth factor stimulus and protein binding.
RGS1	regulator of G-protein signaling 1	RGS1 regulates G-protein coupled receptor protein signaling pathway and GTPase activity.
BENE	T-cell differentiation protein-like	BENE encodes an element of the machinery for raft-mediated trafficking in endothelial cells.
TP53BP1	tumor protein p53 binding protein 1	TP53BP1 is a positive regulation of transcription from RNA polymerase II promoter and sequence-specific DNA binding transcription factor activity.
CLP36	PDZ and LIM domain 1	CLP36 is a regulation of transcription, DNA-templated and encodes a member of the enigma protein family.
PRDX2	peroxiredoxin 2	PRDX2 is a transcription initiation from RNA polymerase II promoter and regulates hydrogen peroxide metabolic process.
FBLN1	fibulin 1	FBLN1 regulates gene expression, substrate-dependent cell migration and cell attachment to substrate.
PODXL	podocalyxin-like	PODXL is a positive regulation of cell migration and cell-cell adhesion mediated by integrin.
PURA	purine-rich element binding protein A	PURA regulates cell proliferation and transcription from RNA polymerase II promoter.

4.2 Experiments On Prostate Data

Prostate cancer data [17] are used to identify novel biomarkers related to prostate cancer. The data set includes the expression profiles of 54,675 genes in 13 prostate cancer samples and 8 normal samples. The raw and unprocessed microarray data are available in the NCBI GEO database (accession number : GSE55945). In order to reduce noise and computational burden, the original gene expression data set is preprocessed before experiment analysis. We only select 3,500 most significant genes after the preprocessing and simultaneously standardize the data. Let 13 prostate cancer samples be negative class and 8 normal samples be positive class. Similar to section 4.1, we compare the IGL with SGL and GL from the above two aspects: average classification accuracy and gene selection performance. aspects. The prostate data set is randomly divided into two parts: two-thirds for training and one-third for testing. To avoid the contingency of single experiment, each process is repeated 10 times. Experiment results are shown in Table 3.

As is shown in Table 3, IGL achieves the highest classification accuracy among three methods. It is also shown that the standard deviation of average classification accuracy for IGL is almost the same as SGL, which is much

Table 3: Experimental results on prostate cancer data over 10 runs (the standard deviations are reported in parentheses).

Method	Average classification accuracy	Average number of genes
GL	0.833(0.023)	60.10(3.39)
SGL	0.839(0.019)	53.90(2.91)
IGL	0.847(0.017)	49.80(2.35)

smaller than GL. Compared with other methods, IGL selects the least number of genes and achieves the smallest standard deviation of average number of genes among three methods. It demonstrates that the property of gene selection for the proposed model is the most stable among three methods. Similar to section 4.1, average number of genes is not an integer.

Some key genes for prostate cancer are selected by the improved group lasso, which are believed to be closely related to prostate cancer. For example, CAPN1 plays a vital role in regulating cell proliferation, protein binding and receptor catabolic process. NBL1 is a secreted protein that is highly restricted to the prostate [18]. CD268 is important for tumor necrosis factor-mediated signaling pathway and stimulates B cell growth. MDM2 positively regulates mitotic cell cycle, cell proliferation and proteasomal ubiquitin-dependent protein catabolic process. This also shows that the actual function of these selected key genes is related to prostate cancer.

5 CONCLUSION

The improved group lasso has been proposed in this paper and the solving algorithm has been developed. By using the conditional mutual information, we propose a new criterion to evaluate the individual gene importance. Experiments performed on two cancer data sets demonstrate that the proposed model achieves the better properties of classification and gene selection. Meanwhile, it has also been verified that these selected genes are highly related to cancer. Furthermore, it is important to further test the proposed method on additional data sets and give reasonable biological interpretations. On the other hand, note that group lasso and sparse group lasso highly depend on the division of the groups. Hence, it is desirable to divide genes into different groups by using the reasonable strategy. We leave these issues for future research.

REFERENCES

[1] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machine, Machine Learning, Vol.46, No.1, 389-422, 2002.

[2] L. Wang, J. H. Zhu, Hybrid huberized support vector machines for microarray classification and gene selection, Bioinformatics, Vol.24, No.3, 412-419, 2008.

[3] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society, Vol.58, No.1, 267-288, 1996.

[4] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society, Vol.67, No.2, 301-320, 2005.

[5] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, Journal of the Royal Statistical Society, Series B, Vol.68, No.1, 49-67, 2006.

[6] N. Simon, J. Friedman, T. Hastie, R. Tibshirani, A sparse-group lasso, Journal of Computational and Graphical Statistics, Vol.22, No.2, 231-245, 2013.

[7] H. Zou, The adaptive lasso and its oracle properties, Journal of the American Statistical Association, Vol.101, No.476, 1418-1429, 2006.

[8] J. Li, Y. Jia, Z. Zhao, Partly adaptive elastic net and its application to microarray classification, Neural Computing and Applications, Vol.22, No.6, 1193-1200, 2013.

[9] H. Wang, C. Leng, A note on adaptive group lasso, Computational Statistics and Data Analysis, Vol.52, No.12, 5277-5286, 2008.

[10] K. Fang, X. Wang, S. Zhang, et al, Bi-level variable selection via adaptive sparse group Lasso, Journal of Statistical Computation and Simulation, Vol.85, No.13, 1-11, 2014.

[11] M. Vincent, N. R. Hansen, Sparse group lasso and high dimensional multinomial classification, Computational Statistics and Data Analysis, Vol.71, No.1, 771-786, 2014.

[12] X. Zhang, X. M. Zhao, K. He, et al, Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information, Bioinformatics, Vol.28, No.1, 98-104, 2012.

[13] F. M. Giorgi, G. Lopez, J. H. Woo, et al, Inferring protein modulation from gene expression data using conditional mutual information, Plos One, Vol.9, No.10, 2014.

[14] T. M. Cover, J. A. Thomas, Elements of Information Theory, NewYork: Wiley, 1991.

[15] G. Brown, A. Pocock, M. J. Zhao et al, Conditional likelihood maximisation: a unifying framework for information theoretic feature selection, Journal of Machine Learning Research, Vol.13, No.1, 27-66, 2012.

[16] S. Monti, P. Tamayo, J. Mesirov, T. Golub, Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data, Machine Learning, Vol.52, No.12, 91-118, 2003.

[17] M. S. Arredouani, L. Bin, B. Manoj, et al, Identification of the Transcription Factor Single-Minded Homologue 2 as a Potential Biomarker and Immunotherapy Target in Prostate Cancer, Clinical Cancer Research, Vol.15, No.18, 5794-5802, 2009.

[18] T. Hayashi, S. Ohara, J. Teishima, et al, The search for secreted proteins in prostate cancer by the escherichia coli ampicillin secretion trap: expression of nbl1 is highly restricted in prostate and related in progression, Pathobiology, Vol.80, No.2, 60-69, 2013.