International Journal of Physical Sciences Vol. 7(44), pp. 5892-5899, 23 November, 2012
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS12.231
ISSN 1992-1950 ©2012 Academic Journals

Full Length Research Paper

The algorithm of Fuzzy C-Means clustering based on non-negative matrix factorization

Wang Nian[1]*, Su Liangliang[1], Tang Jun[1], Liang Dong[1] and Zeng Yanjun[2]

[1]Key Laboratory Intelligent Computing and Signal Ministry of Education,Anhui University, Hefei, Anhui, China.
[2]Biomedical Engineering Center, Beijing University of Technology, Beijing, 100022, China.

Accepted 28 May, 2012

Clustering analysis is an effective method to discover and identify tumor classes. So, this paper proposes a Fuzzy C-Means clustering (FCM) algorithm based on Non-negative matrix factorization (NMF). Firstly, gene expression profiling (GEP) is simply processed through mean and variance of gene expression, which can then be mapped into a low dimensional space by NMF method. Finally, for discovering and identifying cancer classes, the FCM algorithm is adopted to cluster the GEP. Experimental results show that the NMF reduction dimension method has the capability to resist noise. Compared with Principal component analysis (PCA) method, the NMF reduction dimension method also shows certain advantage.

Key words: Fuzzy C-Means clustering, gene expression profiling, cancer, non-negative matrix factorization.

INTRODUCTION

The DNA microarray technique has given rise to a revolutionary influence to traditional cancer treatment methods (Rui et al., 2008), which can detect simultaneously tens of thousands of gene expression level in different samples. Now, studies based on the GEP have attracted more and more attention. However, analyzing GEP still faces many challenges; a typical one is “fewer samples and higher dimension”. So, how to effectively reduce dimension of the GEP is becoming a research hot spot. Many researchers have focused on reduction dimension methods (Tusher et al., 2001; Wang et al., 2006; Yeung et al., 2009), most of which are supervised reduction dimension methods, while a little attention was paid to unsupervised reduction dimension methods (Tang and Zhang, 2003). In fact, unsupervised reduction dimension methods are necessary because a lot of unknown class data are in existence in the real world.

Currently, most of unsupervised reduction dimension methods are based on statistical knowledge or clustering algorithm to find a subset of gene (information gene) (He et al., 2003; Zhu et al., 2005). Talavera (2000) proposed dependency-based feature selection method. The method was based on the correlation between information genes under the assumption that the other genes does not exist. Pena et al. (2001) constructed a related measurement standard and computed correlation threshold to achieve dimension reduction. Regarding this, traditional methods ignore the fact that the GEP is often “highly connected” (Jiang et al., 2003). Su et al. (2003) utilized Rank gene, a program which contains a series of common genetic ranking criterion, to extract feature genes in 2003. Ding (2003) presented a two-way ordering method to select feature gene in the same year under the assumption that some genes may correspond to a new kind of unknown expression class. The algorithm proposed by Watson (2006), which used co-Xpress as a means of identifying groups of genes, can overcome the shortcoming of traditional methods that may miss groups of genes from differential co-expression patterns under different subsets of experimental conditions. Experimental results showed that the methods were effective.

The non-negative matrix factorization (NMF) method (Lee and Seung, 1999, 2001), a recent method for compressing data scale, is a linear, non-negative approximate data representation, and should be noted that negative often does not has meaning in reality and

*Corresponding author. E-mail: wn_xlb@ahu.edu.cn.

the non-negative is more close to reality (Paatero and Tapper, 1994). This paper proposes the NMF dimension reduction method to map the samples with high dimension into low dimensional space. Meanwhile, the class information of all the samples is effectively preserved. The Fuzzy C-Means clustering (FCM) algorithm (Dembele and Kastner, 2003) is adopted to cluster all the samples, meanwhile, the Principal component analysis (PCA) (He and He, 2007) is used to compare with the proposed algorithm.

NON-NEGATIVE MATRIX FACTORIZATION (NMF) METHOD

The NMF method (Lee and Seung, 1999), a recent method for compressing data scale, is a linear, non-negative approximate data representation. Suppose that $\boldsymbol{S}=\left(s_{ij}\right)_{M\times N}$ is a non-negative matrix subject to $s_{ij}\geq 0$, $i=1,2,\cdots,M$, $j=1,2,\cdots,N$, the NMF method means that the matrix $\boldsymbol{S}$ is approximately factorized into two sub non-negative factors $\boldsymbol{W}$ with size $M\times r$ and $\boldsymbol{H}$ with size $r\times N$. Usually, $r<<\min\{N,M\}$, so that $\boldsymbol{W}$ and $\boldsymbol{H}$ are smaller than the original matrix $\boldsymbol{S}$. Then, an approximate factorization $\boldsymbol{S}\approx\boldsymbol{WH}$ is obtained. The approximate degree is quantified by computing Kullback-Leibler Divergence function (Cover and Thomas, 1991):

$$D(\boldsymbol{S},\boldsymbol{WH})=\sum_{i,j}\left(s_{ij}\log\frac{s_{ij}}{(\boldsymbol{WH})_{ij}}-s_{ij}+(\boldsymbol{WH})_{ij}\right) \tag{1}$$

This is lower bounded by zero, and clearly vanishes if and only if $\boldsymbol{S}=\boldsymbol{WH}$. In addition, the divergence $D(\boldsymbol{S},\boldsymbol{WH})$ is non increasing under the following update rules (Lee and Seung, 2001):

$$h_{st}=h_{st}\frac{\sum_i w_{is}s_{it}/(\boldsymbol{WH})_{it}}{\sum_j w_{js}} \tag{2}$$

$$w_{ik}=w_{ik}\frac{\sum_j h_{kj}s_{ij}/(\boldsymbol{WH})_{ij}}{\sum_j h_{kj}} \tag{3}$$

Fuzzy C-Means clustering

The FCM algorithm (Dembele and Kastner, 2003) allows a sample belonging to one or more classes. Assume that there are M samples which were assigned into C classes ($\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_C$), then u_{ij} is defined to describe the correlation degree between each sample x_j and class $\boldsymbol{P}_i$. So, the correlation matrix $\boldsymbol{U}$ can be obtained as follows:

$$\boldsymbol{U}=\begin{bmatrix}u_{11} & u_{12} & \cdots & u_{1M}\\ u_{21} & u_{22} & \cdots & u_{2M}\\ \vdots & \vdots & \ddots & \vdots\\ u_{c1} & u_{c2} & \cdots & u_{cM}\end{bmatrix} \tag{4}$$

Where $u_{ij}\in[0\quad 1]$, $i=1,2,\cdots,C$, $j=1,2,\cdots,M$. For a given sample x_j, u_{ij} indicates a strong association to class $\boldsymbol{P}_i$ if u_{ij} is close to 1 and a lower association if u_{ij} is close to 0. Meanwhile, the u_{ij} must meet the following constraints:

$$\sum_{i=1}^{C}u_{ij}=1,\qquad \forall j=1,...,M \tag{5}$$

$$0<\sum_{j=1}^{M}u_{ij}<M\qquad \forall i=1,2,\cdots,C \tag{6}$$

The optimal clustering results can be obtained by minimizing the cost function of Equation 7, which includes the correlation degree between samples and classes, and distance information of samples and class centroids.

$$J(\boldsymbol{U},v_1,...,v_C)=\sum_{i=1}^{C}\sum_{j=1}^{M}u_{ij}^m d_{ij}^2\quad i=1,2,\cdots,C\quad j=1,2,\cdots,M \tag{7}$$

$$d_{ij}^2=\left\|x_j-v_i\right\|^2 \tag{8}$$

Where v_i indicates the centroid of class $\boldsymbol{P}_i$, $m\in(0\quad+\infty)$ is a weight index, d_{ij} is the Euclidean norm. through the introduction of Lagrange multiplier, the constraint condition of Equation 7 optimization problem can be transform into an optimization problem with no constraint conditions.

$$\begin{aligned}\bar{J}(\boldsymbol{U},v_1,...,v_C,\lambda_1,...,\lambda_M)&=J(\boldsymbol{U},v_1,...,v_C)+\sum_{j=1}^{M}\lambda_j\left(\sum_{i=1}^{C}u_{ij}-1\right)\\&=\sum_{i=1}^{C}\sum_{j}^{M}u_{ij}^m d_{ij}^2+\sum_{j=1}^{M}\lambda_j\left(\sum_{i=1}^{C}u_{ij}-1\right)\end{aligned} \tag{9}$$

Where $\lambda_j(j=1,2,\cdots,M)$ is a Lagrange multiplier. For Equation 9, we can solve the partial derivative of the objective function and get v_i and u_{ij}. Finally, Equation 6 can reach an ideal value through continuously updating v_i and u_{ij}.

$$v_i=\frac{\sum_{j=1}^{M}u_{ij}^m x_j}{\sum_{j=1}^{M}u_{ij}^m} \tag{10}$$

$$u_{ij}=\frac{1}{\sum_{k=1}^{C}\left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \tag{11}$$

Clustering validity

The cluster valid index plays a very important role to clustering

analysis. In this paper, Xie-Beni index (Xie and Beni, 1991) that is proposed in accordance to the FCM algorithm is applied to assess the capability of clustering.

Xie-Beni index: $V_{xb}(\boldsymbol{U};\boldsymbol{V};\boldsymbol{X}) = \frac{\sum_{i=1}^{C}\sum_{l=1}^{M} u_{ik}^{h}\left\|x_l - v_i\right\|^2}{M * d_{\min}(v_i - v_j)}$ (12)

Here, $h \in [0, -\infty)$, $d_{\min}(v_i - v_j)$ is the shortest Euclidean distance between the classes' centroids. The V_{xb} includes not only the information of the correlation degree but also that of samples, which can be seen that the smaller the V_{xb} , the better the clustering result.

RESULTS

Experimental procedure

The GEP can be expressed in a matrix $\boldsymbol{G}=\left(g_{ij}\right)_{M\times N}$, *M* and *N* indicates the number of samples and the number of genes. g_{ij} shows the expression level of gene g_j in sample x_i . Generally, $N >> M$, which is the so-called "dimension disaster". Reducing dimension of the GEP is necessary, while the information of samples class can be remained as much as possible.

Data preprocessing

Previous studies have shown that sample classes can be discriminated through only a small subset of genes whose expression levels strongly correlated with the class distinction (Golub et al., 1999). This means that the GEP often contains a huge amount of noise. Then a simple method is implemented to preprocess the GEP.

$$AVE_j = \frac{1}{M}\sum_{i=1}^{M} g_{ij}, \quad i = 1,2,\cdots,M \qquad j = 1,2,\cdots N \tag{13}$$

$$Gen_j = \frac{Max_j - Min_j}{AVE_j}, j = 1,2,\cdots,N \tag{14}$$

Here, AVE_j , Max_j and Min_j represent mean, maximum and minimum of the ***i***th gene in *M* samples, respectively. These genes with smaller Gen_j will be eliminated so that the subsequent processing complexity can be reduced.

Extract the basis factors

The data matrix $\boldsymbol{G}'$ of size $M \times L$ will be obtained through preprocessing the GEP, then we perform NMF on $\boldsymbol{G}'$ and get an approximate expression according to the updated rules of Equations 2 and 3.

$$\boldsymbol{G}' \approx \boldsymbol{W}\boldsymbol{H} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1r} \\ w_{21} & w_{22} & \cdots & w_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ w_{M1} & w_{M2} & \cdots & w_{Mr} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \\ \vdots \\ \boldsymbol{H}_r \end{bmatrix} \tag{15}$$

Here, the *i*th sample $\boldsymbol{Y}_i \approx w_{i1}\boldsymbol{H}_1 + w_{i2}\boldsymbol{H}_2 + \cdots + w_{ir}\boldsymbol{H}_r$, $i = 1,2\cdots,M$. $\boldsymbol{H}_j (j = 1,2,\cdots,r)$ represents a basis vector (that is, basis factor). That is, each sample is approximated by a linear combination of $\boldsymbol{H}_j$, weighted by the components of ***W***. Generally, the number of the basis factors is relatively few, so all the samples can be mapped into a low dimensional space. Then ***W*** is a compressed version of the data matrix $\boldsymbol{G}'$ in the space.

Fuzzy C-Mean clustering

Clustering analysis based on the GEP can be transform into ideal with the row vectors of $\boldsymbol{W} = (w_{ij})_{M\times N}$. In order to obtain the optimal $\boldsymbol{U}$ and clusters centroids $V = (v_1, v_2, \cdots, v_C)$, the following steps are performed:

Step 1: Randomly initialize matrix $\boldsymbol{U}$ from 0 to 1 according to Equations 5 and 6, and subject to $u_{ij} \in [0 \quad 1]$;

Step 2: Calculate centroids of clusters by using Equation 10, and obtain v_i and d_{ij} $i = 1,2,\cdots,C; j = 1,2,\cdots,M$;

Step 3: Compute the cost function according to Equation 7, Stopping criteria is given by predefined value or $\left\|J^k - J^{k-1}\right\| \le T$. J^k is the value of the cost function in the *k*th iteration, T is a given value. Otherwise, go to Step 4;

Step 4: Update $\boldsymbol{U}$ with Equation 11, go to Step 2.

Simulation experiment

Simulation data is divided into three classes of curves, which includes three lines, three sine curves and three parabolas. As shown in Figure 1, 21 discrete points are selected according to x as varied in some intervals from 0 to 2π , then simulation data can be organized as a matrix $\boldsymbol{V}_{9\times 21}$. $\boldsymbol{V}$ is factorized by the NMF method into two-dimensional and three-dimensional spaces when the number of basis factors is $r = 2$ and 3; the effect of reduction dimension by NMF is shown in Figures 2 and 3. Here, blue, red and green points represent sine curves, lines and parabolas, respectively. Compared with the Figure 1, the different classes of curves became easier to recognize.

Curve cross phenomenon can be observed in Figure 1. Curve classes cannot be an accurate judgment if we only

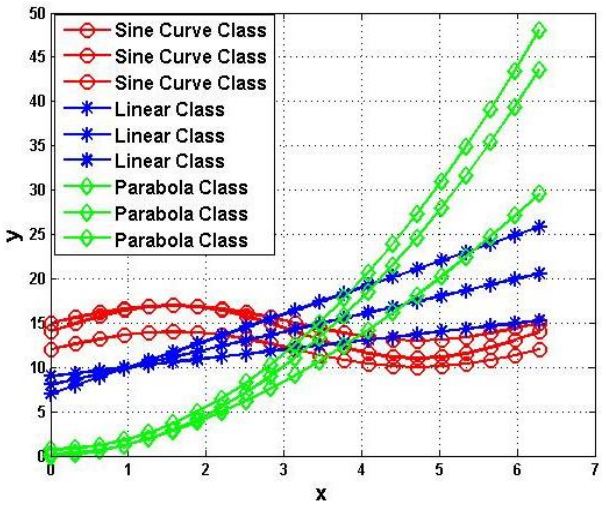

Figure 1. Simulation dataset.

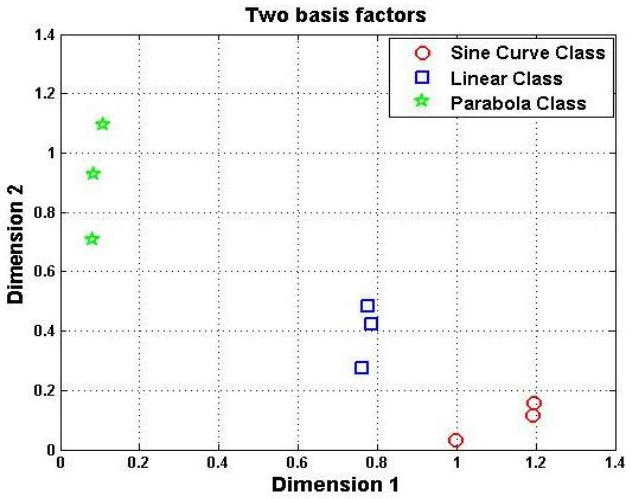

Figure 2. Three classes curves in the two-dimensional space.

use the points of intersection; these points cannot very well reflects curve feature and may lead to a bad clustering result. Nevertheless, in Figures 2 and 3, the influence of the points is effectively eliminated by NMF. What is more? the points from the same curve have strong correlation, that is, a lot of noise exists in those points. Analogously, the noise has been reduced observably in Figures 2 and 3.

Cancer data experiments

The results of simulation experiment show that NMF can effectively reduce dimension and reserve category information. Then the NMF reduction dimension method is used in the next two typical datasets-leukemia dataset and colon dataset (Table 1). Colon dataset URL:

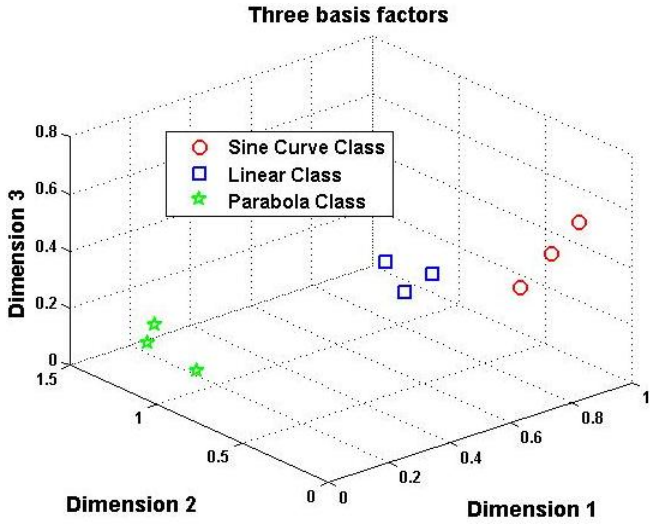

Figure 3. Three classes curves in the three-dimensional space.

http://linus.nci.nih.gov/~brb/DataArchive_New.html, leukemia dataset URL: http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi.

Leukemia dataset

Firstly, the V_{xb} is used to analyze the clustering validity. Let $r=1,3,5,7$, leukemia samples are clustered into 2, 3, 4, 5 and 6 classes. Considering the initialization of NMF and $\boldsymbol{U}$ is random, the instability of the clustering result is existent. The solution is obtained by testing repeatedly to record the optimal V_{xb} index.

In Figure 4, V_{xb} shows the clustering validity of the leukemia dataset. In the case of $r=1,3,5,7$, we find the minimum value of the V_{xb} in $C=3$ and the sub-minimum value of the V_{xb} in $C=2$ from four sub-figures, which illustrates that the best number of clusters is 3 and the second good number of clusters is 2. This result is in accordance with the practical situation of 3 classes (that is, AML, ALL_B and ALL_T) and 2 classes (that is, AML and ALL). The fact confirms that the NMF reduction dimension method does not change the class information of sample.

In follow-up experiments, the dimension reduction effect of the NMF method will be in the spot light. Compared with the PCA method (He and He, 2007), the results are given as follows:

When $r=3$ and $C=2$ or 3, there is only one AML sample and is misclassified into ALL class by the NMF method, which illustrates that this method has a good performance in keeping classification information. Table 2 shows that this method is better than the PCA method.

Table 1. Two cancer datasets.

Dataset	Samples number	Class 1	Class 2	Genes number
Leukemia	38	27 (19 ALL_B and 8 ALL_T)	11 (AML)	5000
Colon	62	22 (Normal)	40 (Cancer)	2000

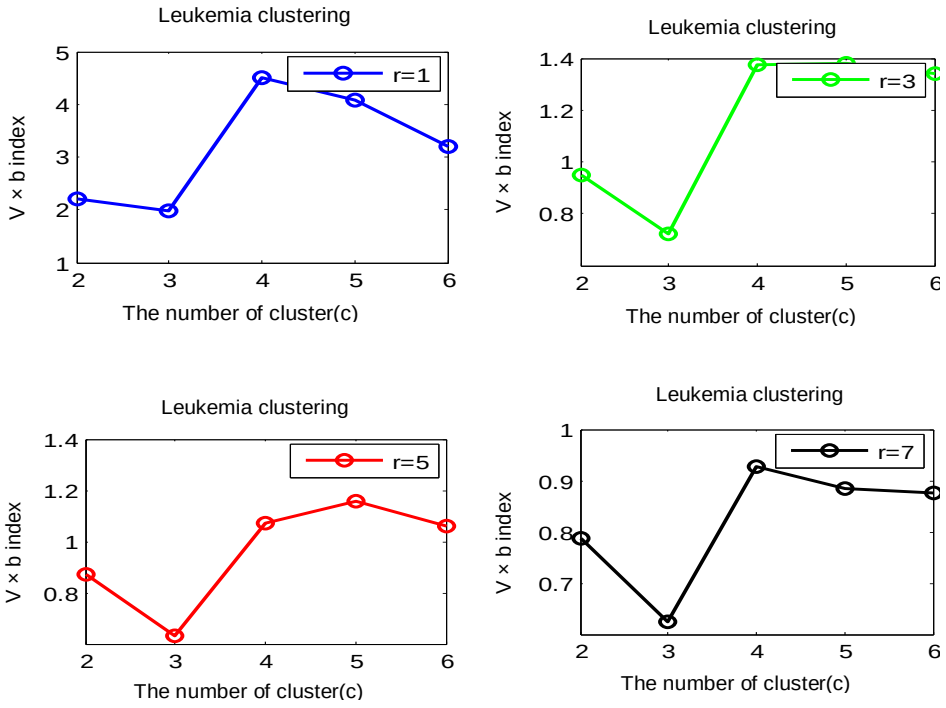

Figure 4. The V × b index (r = 1, 3, 5, 7).

Table 2. The contrast experiment results of Leukemia dataset.

Class number(*C*)	Method	Clustering accuracy (%)
2	NMF	97.37
	PCA	89.47
3	NMF	97.37
	PCA	78.95

The reduction dimension effect of the NMF method is vividly and directly observed in Figure 5 when $r=2$. 38 samples were divided into two classes in Figure 5a, while the clustering result through the FCM algorithm is shown in Figure 5b when $C=2$.

The clustering result of 38 samples is shown in three-dimensional space in Figure 6. The distinction between classes is mainly embodied in each dimension when $C=3$, as shown in Figure 6b. The value of AML samples in the first dimension is significantly greater than that of ALL_T and ALL_B samples, and the value of ALL_T and ALL_B samples are the biggest in the second and third dimension, respectively. Here, only one error is that one ALL_B sample is assigned into class ALL_T.

Colon dataset

The comparison results about the NMF and PCA methods are shown in Table 3. It is clear that the NMF

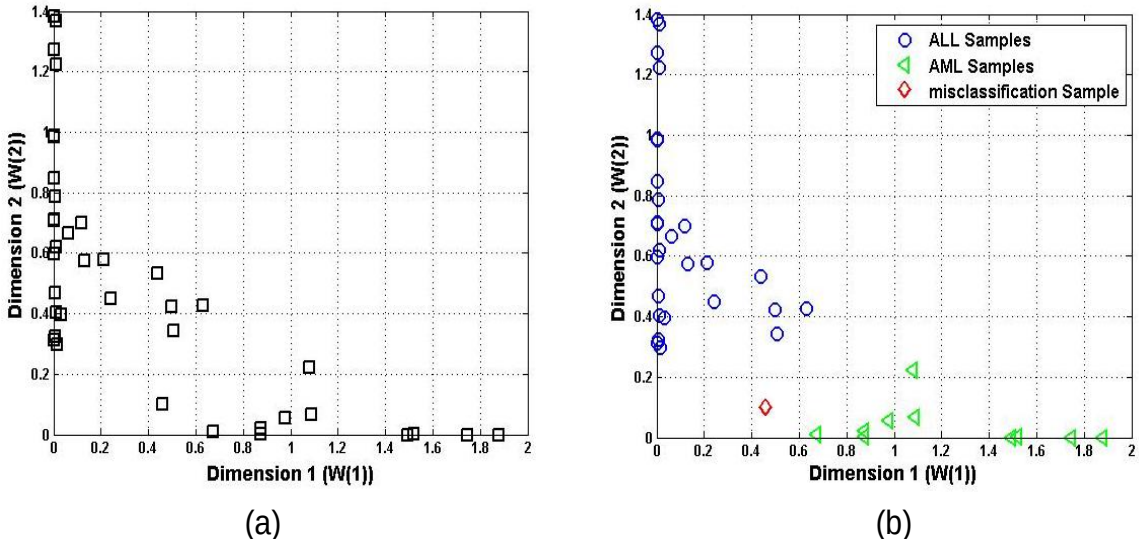

Figure 5. The G' obtained after preprocessing Leukemia dataset, then factorized by NMF when $r=2$ and gain $W_{38\times2}$. Each row of $W_{38\times2}$ corresponds to a sample. The effect of before clustering and after clustering are (a) and (b), respectively.

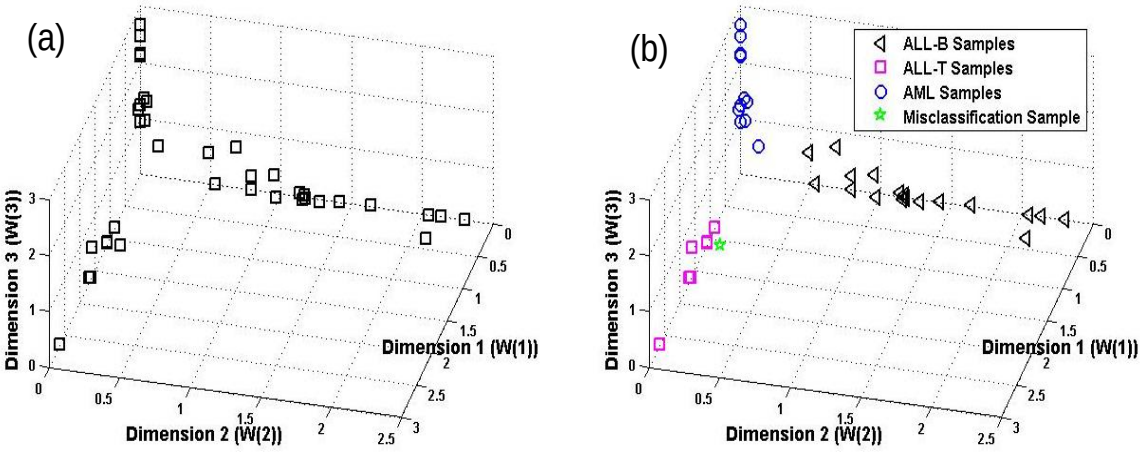

Figure 6. The G' obtained after preprocessing Leukemia dataset, then factorized by NMF when $r=2$ and gain $W_{38\times3}$. Each row of $W_{38\times3}$ corresponds to a sample. The effect of before clustering and after clustering are (a) and (b), respectively.

method is superior to the PCA method. If about 250 genes were chosen in the pretreatment stage, the correct rate of the clustering result has reached 88.71%. We find that the correct rate drops when the numbers of gene in the pretreatment stage are more than 300 or less than 200 on the basis of many times experiments. The possible reason is that the noise and un-information genes lowered the NMF effect when more than 300. In the case of less than 200, the effect is also deteriorated due to the loss of information genes. Thus, the number of information genes can be estimated.

Analogously, to observe more vividly the effect of reduction dimension by the NMF method, colon samples are clustered in Figures 7 and 8.

Colon data is clustered into 2 classes, and clustering effect is observed directly from Figure 8b, and about 7 samples are misclassification samples, the results show that the class information of the samples is effectively reserved after the NMF method factorizes the G'.

The results of the simulation dataset and the real

Table 3. The contrast experiment results of colon dataset.

Class number(*C*)	Method	Clustering accuracy (%)
2	NMF	88.71
	PCA	80.65

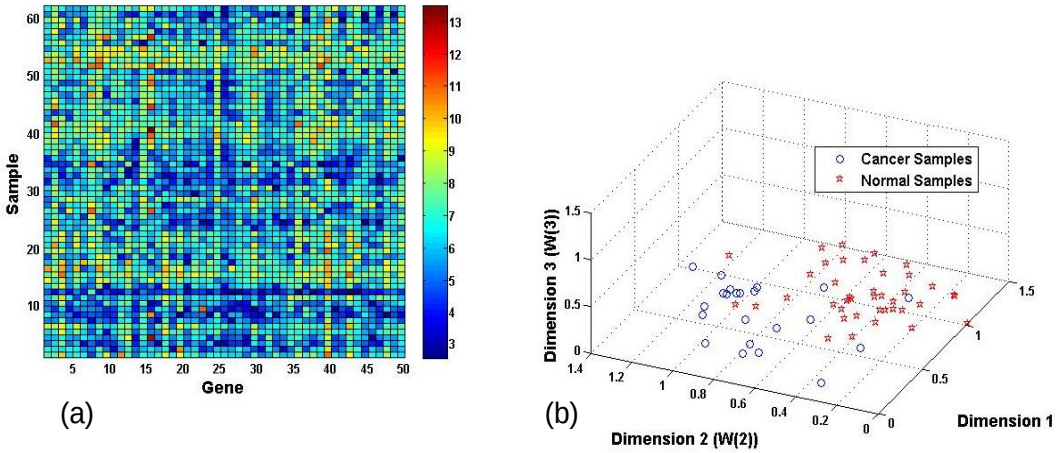

Figure 7. The G' with 50 genes obtained after preprocessing Leukemia dataset in (a), then factorized by NMF when $r=3$ and gain $W_{62\times3}$. Each row of $W_{62\times3}$ corresponds to a sample. The compression effect of G' is shown in (b).

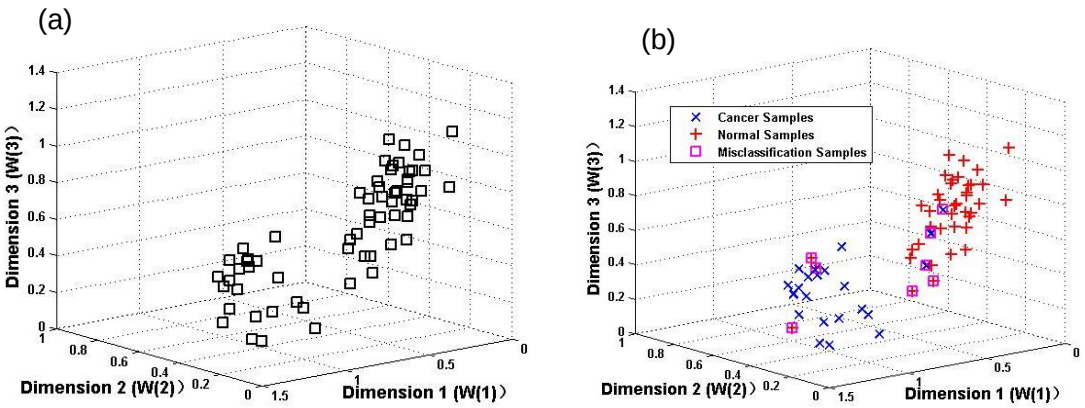

Figure 8. The G' obtained after preprocessing Leukemia dataset, then factorized by NMF when $r=3$ and gain $W_{62\times3}$. Each row of $W_{62\times3}$ corresponds to a sample. The effect of before clustering and after clustering are (a) and (b), respectively.

datasets verify that the NMF method is a feasible and valid reducing dimension.

DISCUSSION

Many methods reduce the GEP dimension by selecting information genes, whose shortcoming is the lack of considering the correlation between genes. However, the methods, like PCA and Independent component analysis (ICA), have some restrictive conditions on datasets. For instance, the datasets must be linearly separable or with Gaussian distribution. In generally, these conditions are not necessarily reasonable.

In this paper, the datasets need to meet the negative, which is consistent with the reality, so the NMF method has certain rationality. But the clustering result appeared unstable due to the initialization of the NMF method that is random. So how to initialize is significant in a follow-up research. Meanwhile, the clustering accuracy can be further enhanced through combining with other classifiers (as K Nearest Neighbor classifier, Support Vector Machine and so on). All the experiments were completed with Maltab software (version 7.0).

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No: 61172127), the Anhui Provincial Natural Science Foundation of China (Grant No. 1208085MF93) and the Innovative research team of 211 project of Anhui University (Grant No: KJTD007A).

REFERENCES

Cover TM, Thomas JA (1991). Elements of Information Theory. New York: Wiley.

Dembele D, Kastner P (2003). Fuzzy C-means method for clustering microarray data. Bioinformatics 19(8):973-980.

Ding HQ (2003). Unsupervised feature selection via two-way ordering in gene expression analysis. Bioinformatics 19(10):1259-1266.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999). Molecular Classification of Cancer: Class Discovery and Prediction by Gene Expression Monitoring. Science 286(5439):531-537.

He J, Tan AH, Tan CL (2003). Self-organizing Neural Networks for Efficient Clustering of Gene Expression Data, Proceedings of the International Joint Conference on Neueal. Networks 3:1684-1689.

He XY, He QH (2007). Application of PCA method and FCM clustering to the fault diagnosis of excavator's hydraulic system. Proc. IEEE Int. Conf. Autom. Logist. JiNan, China. pp. 1635-1639.

Jiang D, Pei J, Zhang AD (2003). Interactive Exploration of Coherent Patterns in Time-Series Gene Expression Data. Proc. Ninth ACM SIGKDD int. Conf. Knowl. Discov. Data Min. Washington DC, USA. pp. 565-570.

Lee DD, Seung HS (1999). Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788-791.

Lee DD, Seung HS (2001). Algorithms for non-negative matrix factorization. In Advances in Neural Information Processing 13 (Proc. NIPS*2000). MIT Press, 2001.

Paatero P, Tapper U (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5(2):111-126.

Pena JM, Lozano JA, Larranaga P, Inza I (2001). Dimensionality Reduction in Unsupervised Learning of Conditional Gaussian Networks. IEEE Trans. Pattern Anal. Mach. Intell. 23(6):590-603.

Rui X, Steven D, Donald CW (2008). Clustering of Cancer Tissues Using Diffusion Maps and Fuzzy ART with Gene Expression Data, IEEE Int. Joint Conf. Neural Networks, HongKong, China. pp. 183-188.

Su Y, Murali TM, Pavlovic V, Schaffer M, Kasif S (2003). RankGene: identification of diagnostic genes based on expression data. Bioinformatics 19(12):1578-1579.

Talavera L (2000). Dependency-based feature selection for clustering symbolic data. Intell. Data Anal. 4(1):19-28.

Tang C, Zhang AD (2003). Interrelated Two-way Clustering and Its Application on Gene Expression Data, Int. J. Artificial Intell. Tools 14(4): 577-597.

Tusher VG, Tibshirani R, Chu G (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA 98(8):5116-5121.

Wang S, Wang J, Chen HW, Tang WS (2006). The Classification of Tumor Using Gene Expression Profile Based on Support Vector Machines and Factor Analysis. Sixth International Conference on Intelligent Systems Design and Applications, JiNan, CHINA: 471-476.

Watson M (2006). CoXpress: Differential co-expression in gene expression data. BMC Bioinformatics 7(1):509.

Xie XL, Beni G (1991). A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 13(8):841-847.

Yeung CW, Leung FHF, Chan KY, Ling SH (2009). An Integrated Approach of Particle Swarm Optimization and Support Vector Machine for Gene Signature Selection and Cancer Prediction. Proc. Int. Joint Conf. Neural Netw. Atlanta, Ga. USA. pp. 3450-3456.

Zhu DX, Hero AO, Cheng H, Khanna R, Swaroop A (2005). Network constrained clustering for gene microarray data. IEEE Int. Conf. Acoust. Speech Signal Process. 21(21):4014-4020.