Multimedia Tools and Applications (2020) 79:14811–14824
https://doi.org/10.1007/s11042-019-7656-7

A novel biclustering of gene expression data based on hybrid BAFS-BSA algorithm

Yan Cui[1] · Rui Zhang[1] · Huacheng Gao[1] · Yuanyuan Lu[1] · Yinqiu Liu[1] · Guangwei Gao[2]

Received: 4 January 2019 / Revised: 22 March 2019 / Accepted: 15 April 2019 /
Published online: 9 May 2019

Abstract
As one of usual concepts, co-expressed genes can represent co-regulated genes in gene expression data. This strategy can be refined further because co-expression of the genome may be the result of independent activation under same experimental samples, rather than the same regulatory regime. Therefore, traditional clustering techniques are proposed to find significant clusters, especially, the biclustering technology. By combining Binary Artificial Fish Swarm (BAFS) with Binary Simulated Annealing (BSA) algorithms, the hybrid algorithm named BAFS-BSA-BIC was proposed in this paper. When this method of biclustering was applied to several datasets, lots of biological significant bifclusters were searched, and the results demonstrate the promising clustering performance of our method. The proposed technology was also compared to classical biclustering technologies-CC, QUBIC, FLOC and original BAFS algorithm, and its robustness and quality are better than these algorithms in searching optimal biclusters of co-expressed genes.

Keywords Microarray · Gene expression data · Biclustering · Binary artificial fish swarm algorithm · Binary simulated annealing algorithm

1 Introduction

DNA Microarray technology signs gene expression in a cell, which is implemented by gene expression data [19, 20]. Gene expression data matrix illuminates expression profiles of thousands of genes under a series of different sample groups. The matrix is applied to

✉ Yan Cui
cuiyan@njupt.edu.cn

[1] National Engineering Reaearch Center of Communication and Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

[2] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

the orthodox clustering technology, whose rows, columns and elements are made up of genes, samples and gene expression level values along certain samples severally. Clustering technology is used to merge genes (rows) based on all samples (columns) or vice versa. In some researches, the guilt-by-association heuristic is hypothesized that genomes have the same regulatory scheme with similar gene expression level, which have the same or similar heuristic [13, 16, 17, 21, 28]. This viewpoint is recently being reformulated, because the co-expression of a set of genes may be independently activated under some certain sample groups, rather than the same regulatory mechanism is determined [14, 18, 23, 33]. In this case, the researchers have proposed the concept of biclustering technology based on clustering technology [24, 30, 32]. Biclustering of gene expression data that gathers genomes under a subset of samples and not all samples or cluster sample groups under partial genome and not all genes was proposed. In other words, the difference between orthodox clustering and biclustering is that the biclustering concerns clustering genes and samples simultaneously.

The advantage of biclustering approach is to seek sets of locally co-expressed genes efficiently without acting on the entire gene expression matrix. The first successful implementation of biclustering on gene expression data matrix was completed by Cheng and Church [8]. So far, large amounts of biclustering algorithms have been reported with two typical methods: different heuristic and search norm to seek biclusters [6, 7, 29]. Several classic biclustering algorithms including Cheng and Church's algorithm (CC), Order-preserving Submatrix Algorithm (OPSM) [3], Flexible Overlapped biclustering Algorithm (FLOC) [31], Iterative Signature Algorithm (ISA) [4], Factor Analysis for Bicluster Acquisition (FABIA) [10] and QUBIC [15], are commonly used as standards for evaluating the other biclustering algorithms. CC, a greedy iterative search algorithm, generates a bicluster by adding and deleting rows or columns alternately. OPSM algorithm based on biclustering enumeration have been demonstrated as a high-efficiency method in biclustering family [2]. FLOC, based on the method described by J.Yang et al., has the advantages of being able to obtain a set of optimal biclusters simultaneously and combining methods for dealing with missing values. Besides, ISA, FABIA and QUBIC are also effective biclustering algorithms that have various advantages in processing different gene expression datasets [30].

The gene expression level based on a set of samples can be considered as discrete random variable values. Therefore, it is reasonable to predict the degree of linear dependence between two genes by calculating the correlation value of two random variables. In this paper, the concept is used to explore the optimal biclusters in gene expression data matrix, which acts as an important factor in the construction of the fitness function in the biclustering algorithm in this paper.

BASF algorithm is enlightened by the foraging behavior of fish in the search for high-quality food sources and has been applied to the analysis of optimization problems [9]. The foraging behavior is specially divided into four modules: the preying, aggregating, following and random behavior. In order to improve the ability of BASF algorithm to search the global optimum, the Binary Simulated Annealing algorithm is introduced to form the hybrid BAFS-BSA algorithm. In this work, we presented a new biclustering technology based on the hybrid BAFS-BSA algorithm.

The remainder of the paper is organized as follows. Section 2 presents the biclustering technology based on the hybrid BAFS-BSA algorithm. Section 3 evaluates the performance of the proposed method based on several commonly used gene expression datasets. Finally, Section 4 concludes this paper.

2 Methodology

2.1 Initialization

The initialization phase of the algorithm was composed of two parts. First, *K*-Means clustering method was adopted to cluster genes under samples and gather samples based on genes respectively. Therefore, *m* gene clusters and *n* sample clusters were produced. Then the gene clusters and sample clusters were combined to form $m \times n$ co-clusters. Second, these co-clusters were encoded into binary form in order to unify their formats. The length of each of encoded co-cluster was *p*, where *p* standed for the sum of the number of genes and samples in the gene expression data matrix. When a gene or a sample was in a co-cluster, the corresponding bit was set to 1, otherwise, it was set to 0. Moreover, the input datum of the Hybrid BAFS-BSA algorithm were these encoded co-clusters, namely Artificial Fishes (*AF*s). Figure 1 depicts the general format of *AF* [26].

2.2 Fitness function in BASF-BSA-BIC

In the BAFS-BSA-BIC algorithm, firstly, an objective function as fitness function was defined in order to find biologically relevant biclusters. The fitness function we used reflected the relationship between the linear correlation coefficients of all the genes in a bicluster and the volume of a bicluster. By determining the relationship between these two aspects, we looked for related genes with shift and scale patterns.

To calculate the fitness function value of a bicluster, a bicluster *B* needed to be determined first. Each *AF* represents a bicluster in BAFS-BSA-BIC algorithm. For example, an *AF* is 10,101|101, and *M* is a 5×3 gene expression data matrix, containing 5 rows (genes) and 3 columns (samples). However, the size of actual gene expression data matrix is much larger than the data matrix in this example. As shown in Fig. 2, a 3×2 gene expression data matrix is extracted, namely bicluster *B*.

The proposed fitness function to estimate bicluster *B* is defined as follows.

$$f(B) = (1-\rho(B)) + \frac{1}{N_G} + \frac{1}{N_C} \tag{1}$$

where

$$\rho(B) = \frac{1}{\binom{N_G}{2}} \sum_{i=1}^{N_G-1} \sum_{j=i+1}^{N_G} \left|\rho_{ij}\right| \tag{2}$$

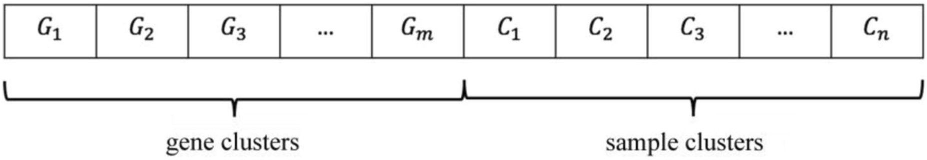


Fig. 1 The general format of *AF*

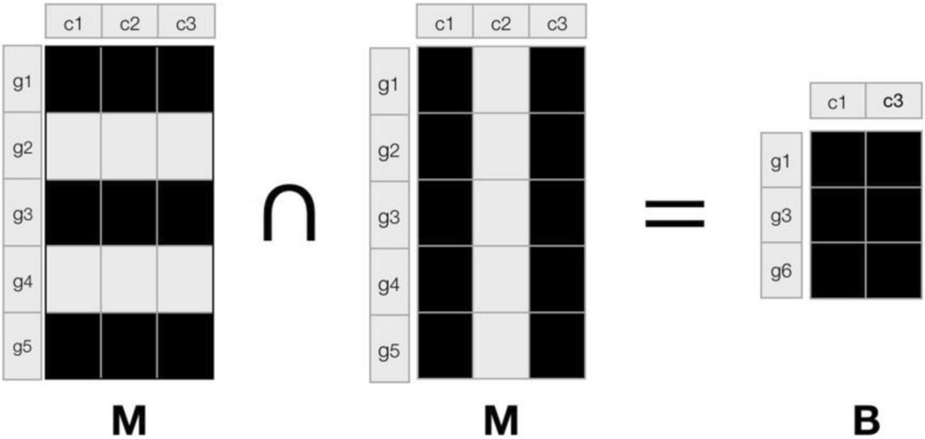


Fig. 2 Conversion method between AF and bicluster in the gene expression data matrix

where $|\rho_{ij}|$ represents the Pearson correlation coefficient of gene G_i and gene G_j. The number of genes and samples of B are N_G and N_C [30]. The best bicluster was searched by the BSFS-BSA algorithm, which has the lowest value of the fitness function.

2.3 Description of BAFS-BSA-BIC

BAFS-BSA-BIC is a hybrid BIClustering algorithm that combines a Binary Artificial Fish Swarm with a Binary Bimulated Annealing algorithm. This discussion was divided into two parts. First, the BAFS algorithm was described briefly. Second, the Binary SA (BSA) was introduced in detail, which was different from the traditional SA algorithm. BSA was combined with the BAFS algorithm to form the hybrid BAFS-BSA-BIC algorithm and enriched the biclustering technology system.

2.3.1 The BAFS algorithm

The implementation of BAFS algorithm was assembled by four main parts: following behavior, aggregating behavior, preying behavior and random behavior. In addition, there were some important details in this algorithm that we needed to consider, such as determination of the visual range of an artificial fish, calculation of crowding factor of fish swarm and setting criteria for performing four different behaviors.

The differences between BAFS and traditional artificial fish swarm algorithms are as follows:

(i) Each fish in a shoal is represented as a Boolean vector rather than traditional values;
(ii) The visual range is constantly changing, which is not a predetermined fixed value, according to the position of each fish, achieved by the hamming distance [8].
(iii) The crowding factor that is not a fixed value varies with the visual range.

These improvements make the artificial fish swarm algorithm more efficient and more widely used. In this paper, the further improved BAFS algorithm, namely BAFS-BSA, is applied in the biclustering experiment of genes expression data.

2.3.2 The BSA and BAFS-BSA-BIC algorithm

SA is a local search method inspired by the physical annealing process studied in statistical mechanics [1]. SA approach follows search directions and executes thousands of iteration generation procedures that perfects the objection function of algorithm, which reduces the fitness function value in BAFSA-BS-BIC algorithm. While exploring solution space, the possibility of accepting poorer neighbor solutions in a controlled manner is offered by SA algorithm to escape from local minima [5]. The SA approach we used had some differences from the traditional SA approach. To enable SA algorithm to process the gene microarray or gene expression data matrix, the binary form was introduced. More precisely, in each iteration, a current solution vector *AF* is a Boolean vector of length *p*, where *p* is the sum of the number of genes and conditions.

$$AF = \left(g_1, g_2, ..g_m, \dots, s_1, s_{2,} \dots, s_N\right), g_i, s_j \in (0,1), i \in (1,2,\dots,m), j \in (1,2,\dots,n) \qquad (3)$$

The *1-opt* neighborhood *N(AF)* could be obtained by changing an element in the solution *AF*, i.e., the hamming distance d_H between the current solution *AF* and each neighbor *AF*′ (*AF*′ ∈ *N*(*AF*)) is equal to 1 [12]. Therefore, the size of *N(AF)* was *p*. In this iteration, for a current *AF* characterized by the fitness function value *f(AF)*, a neighbor *AF'* was randomly selected from the neighborhood *N(AF)* and *f(AF')* was calculated. Then the objection difference $\Delta = f(AF') - f(AF)$ was evaluated. When $\Delta \le 0$, *AF* is replaced by *AF'*. Otherwise, *AF'* would be accepted with a probability $P = e^{(-\Delta)/T}$ [5].

2.4 Mean squared residue (MSR) and average correlation value (ACV)

MSR [25] is one of the most popular measures of homogeneity. Given a bicluster B with I rows and J columns, let a_{ij} represent the element of the bicluster. MSR of this bicluster can be defined as

$$MSR = \frac{1}{|I||J|}\sum_{i\in I, j\in J}\left(a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}\right)$$

where

$$a_{iJ} = \frac{1}{|J|}\sum_j a_{ij}\ a_{Ij} = \frac{1}{|I|}\sum_i a_{ij}a_{IJ} = \frac{1}{|I||J|}\sum_{i,j} a_{ij} \qquad (4)$$

Here a_{iJ} denotes the mean of the *i*-th row and a_{Ij} denotes the mean of the *j*-th column. Besides, a_{IJ} refers to overall means of the bicluster.

ACV is another useful measure of the coherence of the genes and conditions in a bicluster. ACV of a bicluster B(I,J) can be expressed as

$$\text{ACV} = \max\left\{\frac{\sum_{i_1=1}^{|I|}\sum_{i_1=2}^{|J|}|r\_row_{i_1 i_2}| - |I|}{|I|^2 - |I|}, \frac{\sum_{j_1=1}^{|J|}\sum_{j_1=2}^{|J|}|r\_col_{j_1 j_2}| - |J|}{|J|^2 - |J|}\right\} \qquad (5)$$

Where $r\_row_{i_1 i_2}$ and $r\_col_{j_1 j_2}$ stands for the Pearson coefficient of rows i_1, i_2 and columns j_1, j_2 , respectively.

3 Experiments

The purpose of experiments was to verify the performance of the proposed algorithm. The BAFS-BSA algorithm was applied to multiple datasets and desirable results were obtained. Then these results were compared with the classical biclustering methods such as CC, QUBIC, FLOC and the original BAFS biclustering algorithm.

3.1 Datasets

The hybrid BAFS-BSA algorithm for biclustering has been applied to four datasets: complete_DTT, elutriation, cdc_15 and Mice Protein Expression datasets. The first three

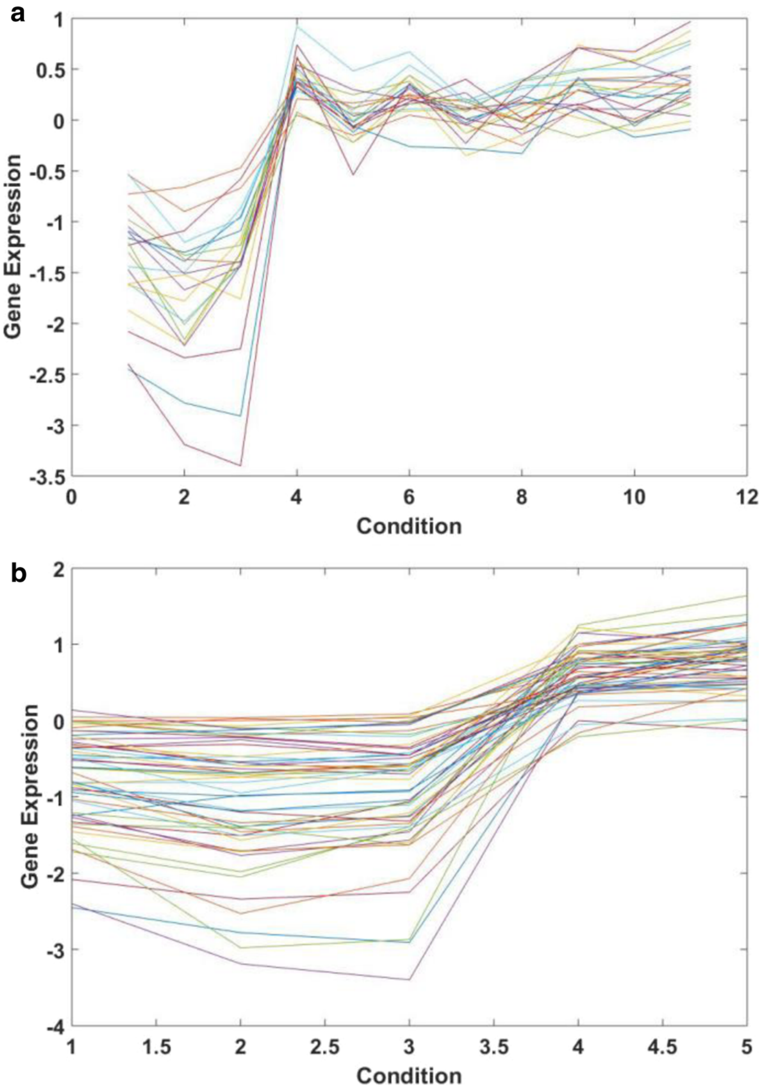


Fig. 3 Several biclusters results with the BASF-BSA method on cdc_15 dataset

datasets respectively include 962 genes and 7 samples, 935 genes and 14 samples, 1087 genes and 24 genes, which have been downloaded from the supplementary information provided in [11]. The last dataset is composed of 1080 genes and 51 samples. Because a small part of the value of the downloaded original data set is missing, we adopted a method to improve this data set, which has been realized by J. A. Nepomuceno et al. [22].

3.2 Results

In complete_DTT, elutriation, cdc_15 and Mice Protein Expression datasets, 10 optimum biclusters were produced by utilizing the BAFS-BSA-BIC algorithm respectively. Most of these biclusters have better performance and several of them are selected randomly to show in Figs. 3, 4, 5 and 6. Each bicluster is represented by a graph, where the x-coordinate represents the sample and the y-coordinate represents the gene expression value. From Figs. 3, 4, 5 and 6,

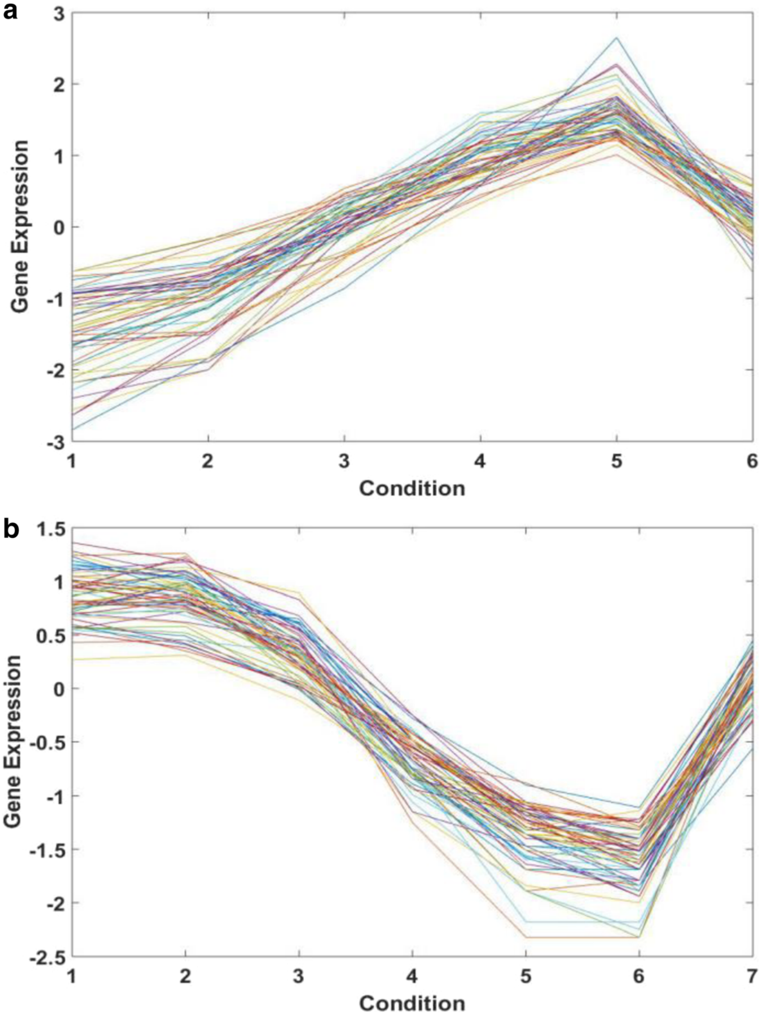


Fig. 4 Several biclusters results with the BASF-BSA method on complete_DTT dataset

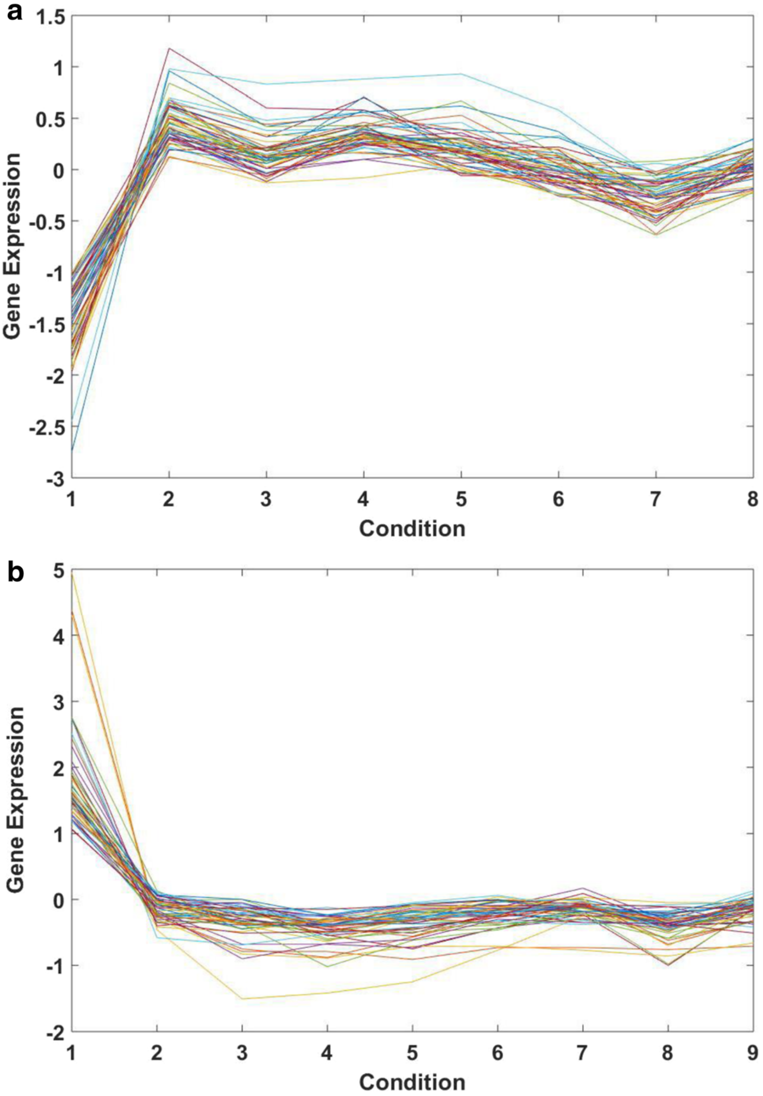


Fig. 5 Several biclusters results with the BASF-BSA method on elutriation dataset

we can intuitively find that each gene in the bicluster searched by the BAFS-BSA-BIC algorithm has similar gene expression in corresponding samples. In addition, this algorithm worked better in processing complete_DTT, elutriation and Mice Protein Expression datasets, because most of the gene expression profile trends in each bicluster searched from these datasets are the same. According to the bicluster graphs obtained from the cdc_15 dataset, we can see these gene expression profile fluctuate a little in the trend, but remain stable overall.

To confirm the degree of correlation between genes in these biclusters, The MSR and ACV and Volume (V) were introduced as three different intra-biclusters evaluation criteria [27]. The value of MSR represents the change related to the interaction between genes (rows) and sampes (columes) in each biclutser. From this it can be concluded that the lower MSR value,

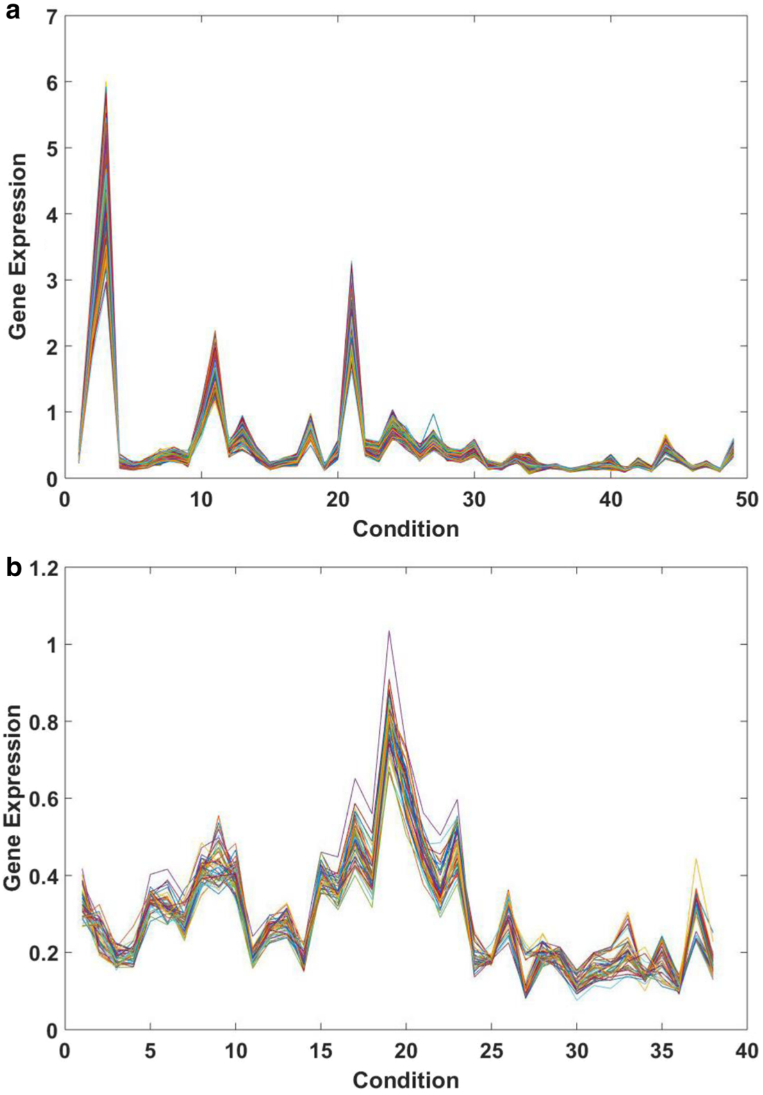


Fig. 6 Several biclusters results with the BASF-BSA method on Mice Protein Expression dataset

Table 1 The quality of biclusters in four datasets

Datasets	Biclusters	Genes	Samples	Volume	$\rho(B)$	MSR	ACV
D1	B1	52	5	260	0.97	0.10	0.95
	B2	60	7	420	0.98	0.04	0.96
D2	B1	21	11	231	0.94	0.10	0.89
	B2	54	5	270	0.97	0.16	0.96
D3	B1	52	8	416	0.96	0.03	0.94
	B2	49	9	441	0.96	0.11	0.95
D4	B1	412	49	20,188	0.99	0.01	0.99
	B2	46	40	1840	0.96	0.01	0.94

Table 2 Comparison of performance of proposed method with other algorithms in *complete*_DTT datasets

	Ave-Volume	Ave- $\rho(B)$	Ave-MSR	Ave-ACV
BAFSA-SA	340	0.98	0.06	0.96
BAFSA	8	1	0.02	0.66
CC	225	0.71	0.20	0.69
QUBIC	4	0.47	0.69	0.80
FLOC	2300	0.09	0.41	0.74

the stronger coherence in the bicluster, while the judgment criterion of ACV value is the opposite. The details are as follows: a bicluster with smaller MSR value, larger ACV value and volume value suggests that all its genes are more coherent and similar genes are more numerous [27]. Therefore, these standards can not only define the quality of biclusters, but also can be used to compare the performance among these biclustering algorithms.

Table 1 summarizes the quality of biclusters acquired by BAFS-BSA-BIC algorithm for the proposed four datasets. In Table 1, datasets are composed among complete_DTT (D1), elutriation (D2), cdc_15 (D3) and Mice Protein Expression (D4). The number of genes (Genes) and samples (Samples), the volume (Volume), the average correlation (ρ(B)), the MSR and ACV are presented to identify the quality of each bicluster. From Table 1, the biclusters with strong coherence from these four datasets can be obtained with our method. Each bicluster in this table has a large volume, a large ACV value and a small MSR value. At the same time, the correlation value ρ(B) close to 1 can also prove the strong coherence of the genes in these biclusters.

3.3 Discussion

The performance of BAFS-BSA for bicluster has been compared with several classical biclustering algorithms such as CC [8], QUBIC [15], FLOC [31] and the original BAFS algorithm for biclustering. The parameters, which are average volume (Ave-Volume), average correlation (Ave- ρ(B)), average MAR (Ave-MSR) and average ACV (Ave-ACV), were

Table 3 Comparison of performance of proposed method with other algorithms in *elutriation* datasets

	Ave-Volume	Ave- $\rho(B)$	Ave-MSR	Ave-ACV
BAFSA-SA	385	0.96	0.07	0.94
BAFSA	1082	0.87	0.12	0.89
CC	215	0.97	0.19	0.72
QUBIC	116	0.40	0.47	0.76
FLOC	2718	0.01	0.31	0.62

Table 4 Comparison of performance of proposed method with other algorithms in *cdc*_15 datasets

	Ave-Volume	Ave- $\rho(B)$	Ave-MSR	Ave-ACV
BAFSA-SA	250	0.96	0.13	0.93
BAFSA	2742	0.84	0.08	0.83
CC	12	0.83	0.02	0.60
QUBIC	120	0.80	0.25	0.80
FLOC	3330	0.01	0.28	0.48

Table 5 Comparison of performance of proposed method with other algorithms in *Mice Protein Expression* datasets

	Ave-Volume	Ave- $\rho(B)$	Ave-MSR	Ave-ACV
BAFSA-SA	11,014	0.97	0.01	0.96
BAFSA	5266	0.99	0.01	0.98
CC	912	0.83	0.11	0.74
QUBIC	27	0.84	0.18	0.84
FLOC	12,027	0.90	0.04	0.72

proposed to record the performance of these optimal biclusters. Tables 2, 3, 4 and 5 records the parameter values of biclusters in complete DTT dataset, elutriation dataset, cdc_15 dataset and Mice Protein Expression dataset respectively.

A bicluster with some characteristics including larger volume, correlation coefficient value, ACV value and the smaller MSR value can be defined as a better bicluster [27]. Figure 7 is the comparison of Ave-Volume, Ave-$\rho(B)$, Ave-MSR and Ave-ACV of different biclustering algorithm based on the same datasets. Compared with other algorithms, BAFS-BSA-BIC has no obvious advantage of acquiring a bicluster with a larger volume. A bigger volume of each bicluster was obtained by FLOC and BAFS algorithm. However, according to the calculation results of $\rho(B)$, ACV and MSR, the biclusters obtained by the proposed method has the best quality. The gene expression level is very similar in each bicluster. Moreover, BAFS-BSA-BIC can obtain satisfactory results when processing different datasets, indicating its wide applicability.

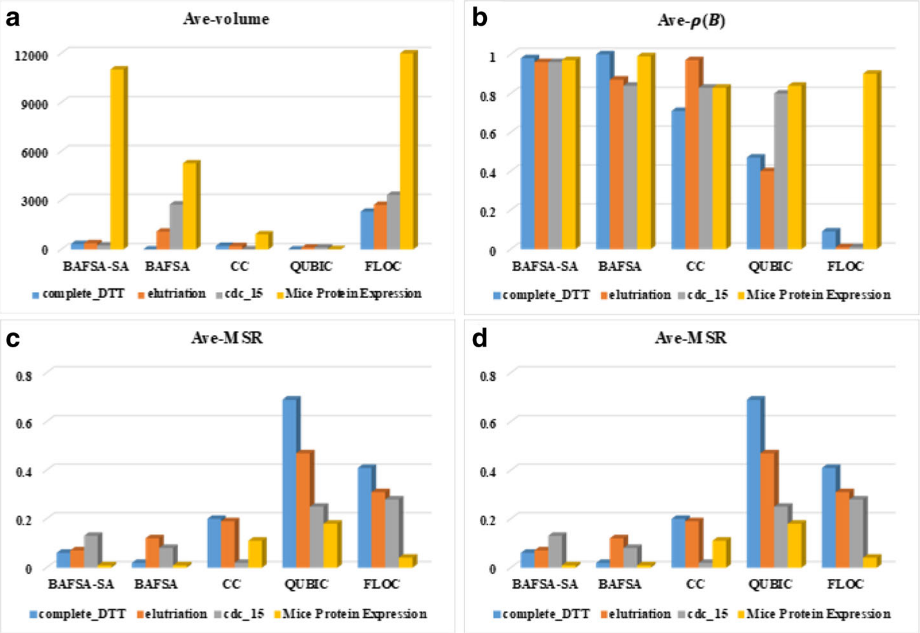


Fig. 7 Comparison of Ave-Volume, Ave-$\rho(B)$, Ave-MSR and Ave-ACV of five biclustering algorithms in four datasets

In particular, both CC and QUBIC perform the worst in the volume of these datasets, extracting biclusters with smaller volume. The correlation coefficient $\rho(B)$ of biclusters acquired from FLOC is much smaller, which means that the degree of linear correlation between each two genes of the biclusters obtained is lower. However, BAFS-BSA-BIC has prominent advantages in ρ(B), ACV and MSR aspects despite no obvious deficiency in the volume, which indicates that our algorithm has better stability than other algorithms.

4 Conclusions

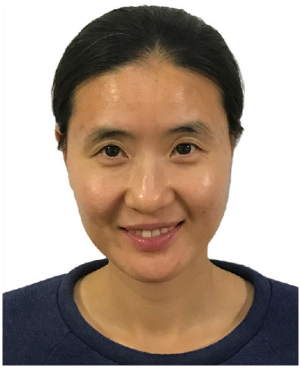
The Hybrid BASF-BSA algorithm for biclustering that integrates biological information and revels new molecular functions of the organism has been implemented in this paper. The BAFS algorithm has the characteristic of fast convergence speed, but it is prone to fall into local optimal problem. In order to avoid this risk, the BSA algorithm has been combined with BAFS algorithm. Experiments on several real datasets show superior performance of our method.

Acknowledgements We acknowledge the financial support from the National Natural Science Foundation of China (61402240,61502245,61772568), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX18_0921).

References

1. Aarts E, Korst J (1989) Simulated annealing and boltzmann machines. Handbook of brain theory & neural networks
2. Banka H, Mitra S (2006) Evolutionary biclustering of gene expressions. Ubiquity 10:5
3. Ben-Dor A, Chor B, Karp R, Yakhini Z (2003) Discovering local structure in gene expression data: the order-preserving submatrix problem. J Comput Biol 10(3):373–384
4. Bergmann S, Ihmels J, Barkai N (2003) Iterative signature algorithm for the analysis of large-scale gene expression data. Phys Rev E 67(3):031902
5. Bouleimen K, Lecocq H (2003) A new efficient simulated annealing algorithm for the resourceconstrained project scheduling problem and its multiple mode version. Eur J Oper Res 149(2):268–281
6. Bryan K, Cunningham P, Bolshakova N (2005) Biclustering of expression data using simulated annealing, in: Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on, IEEE, 383–388
7. Busygin S, Prokopyev O, Pardalos PM (2008) Biclustering in data mining. Comput Oper Res 35(9):2964–2987
8. Cheng Y, Church GM (2000) Biclustering of expression data. International Conference on Intelligent Systems for Molecular Biology 8: 93–103.
9. Cheng Y, Jiang M, Yuan D (2009) Novel clustering algorithms based on improved artificial fish swarm algorithm. In: Fuzzy Systems and Knowledge Discovery 3: 141–145
10. Hochreiter S, Bodenhofer U, Heusel M, Mayr A, Mitterecker A, Kasim A, Khamiakova T, Van Sanden S, Lin D, Talloen W (2010) Fabia: factor analysis for bicluster acquisition. Bioinformatics 26(12):1520–1527
11. Jaskowiak PA, Campello RJ, Costa Filho IG (2013) Proximity measures for clustering gene expression microarray data: a validation methodology and a comparative analysis. IEEE ACM T Comput Bi 10(4):845–857
12. Katayama K, Narihisa H (2001) Performance of simulated annealing-based heuristic for the unconstrained binary quadratic programming problem. Eur J Oper Res 134(1):103–119
13. Lan R, Zhou Y, Liu Z, Luo X (2018) Prior knowledge based probabilistic collaborative representation for visual recognition. IEEE T CYBERNETICS: 1–11
14. Lan R, Li Z, Liu Z, Gu T, Luo X (2019) Hyperspectral image classification using k-sparse denoising autoencoder and spectral-restricted spatial characteristics. Appl Soft Comput 74:693–708
15. Li G, Ma Q, Tang H, Paterson AH, Xu Y (2009) Qubic: a qualitative biclustering algorithm for analyses of gene expression data. Nucleic Acids Res 37(15):e101–e101
16. Lu H, Li Y, Mu S, Wang D, Kim H, Serikawa S (2018) Motor anomaly detection for unmanned aerial vehicles using reinforcement learning. IEEE Internet Things 5(4):2315–2322

17. Lu H, Li Y, Chen M, Kim H, Serikawa S (2018) Brain intelligence: go beyond artificial intelligence. Mobile Netw Appl 23:368–375
18. Lu H, Li Y, Uemura T, Kim H, Serikawa S (2018) Low illumination underwater light field images reconstruction using deep convolutional neural networks. Futur Gener Comput Syst 82:142–148
19. Ma PC, Chan KC (2009) A novel approach for discovering overlapping clusters in gene expression data. IEEE T Bio Med Eng 56(7):1803–1809
20. Madeira SC, Oliveira AL (2004) Biclustering algorithms for biological data analysis: a survey. IEEE ACM T Comput BI 1(1):24–45
21. Markowetz F, Spang R (2007) Inferring cellular networks–a review. BMC Bioinformatics 8(6):S5
22. Nepomuceno JA, Troncoso A, Aguilar-Ruiz JS (2011) Biclustering of gene expression data by correlation-based scatter search. Biodata Min 4(1):3
23. Nepomuceno JA, Troncoso A, Nepomuceno-Chamorro IA, Aguilar-Ruiz JS (2015) Integrating biological knowledge based on functional annotations for biclustering of gene expression data. Comput Meth Prog Bio 119(3):163–180
24. Panteli A, Boutsinas B, Giannikos I (2019) On solving the multiple p-median problem based on biclustering. Oper Res: 1–25
25. Pontes B, Girldez R, Aguilar-Ruiz JS (2015) Quality measures for gene expression biclusters. PlOS ONE 10(3):e0115497
26. Rathipriya R, Thangavel K, Bagyamani J. Binary particle swarm optimization based biclustering of web usage data. arXiv preprint arXiv:1108.0748
27. Saber HB, Elloumi M (2015) Dna microarray data analysis: a new survey on biclustering. Int J Comput Bi 4(1):21–37
28. Serikawa S, Lu H (2014) Underwater image dehazing using joint trilateral filter. Comput Eletr Eng 40(1):41–50
29. Tanay A, Sharan R, Shamir R Handbook of bioinformatics, chapter biclustering algorithms: a survey, To appear
30. Xie J, Ma A, Fennell A, Ma Q, Zhao J (2018) It is time to apply biclustering: a comprehensive review of biclustering applications in biological and biomedical data. Brief Bioinform: 1–16
31. Yang J, Wang H, Wang W, Yu PS (2005) An improved biclustering method for analyzing gene expression profiles. Int J Artif Intell T 14(5):771–789
32. Yoon S, Nguyen HCT, Jo W (2019) Biclustering analysis of transcriptome big data identifies conditionspecific microRNA targets. Nucleic Acids Res: 1–10
33. Zhang Y, Gravina R, Lu H, Villari M, Fortino G (2018) PEA: Parallel electrocardiogram-based authentication for smart healthcare systems. J Netw Comput Appl 117:10–16



Yan Cui received the B. Eng. degree in Computer Science from the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, in 2006, and the PhD in pattern recognition from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2013. She is currently the assistant professor in Nanjing University of Posts and Communications, Nanjing, China. Her research interests are pattern recognition and bioinformatics.

Rui Zhang is currently the graduate student in School of Internet of Things, Nanjing University of Posts and Communications, Nanjing, China. Her interests are pattern recognition andbioinformatics.