Correlation–Based Scatter Search for Discovering Biclusters from Gene Expression Data

Juan A. Nepomuceno[1], Alicia Troncoso[2], and Jesús S. Aguilar–Ruiz[2]

[1] Department of Computer Science, University of Sevilla, Spain
janepo@us.es
[2] Area of Computer Science, Pablo de Olavide University of Sevilla, Spain
{ali,aguilar}@upo.es

Abstract. Scatter Search is an evolutionary method that combines existing solutions to create new offspring as the well–known genetic algorithms. This paper presents a Scatter Search with the aim of finding biclusters from gene expression data. However, biclusters with certain patterns are more interesting from a biological point of view. Therefore, the proposed Scatter Search uses a measure based on linear correlations among genes to evaluate the quality of biclusters. As it is usual in Scatter Search methodology an improvement method is included which avoids to find biclusters with negatively correlated genes. Experimental results from yeast cell cycle and human B-cell lymphoma datasets are reported showing a remarkable performance of the proposed method and measure.

Keywords: Gene Expression Data, Biclustering, Scatter Search, Evolutionary Computation.

1 Introduction

Nowadays, the study of the process of how proteins are coded by genes is one of the most important research topics in Biology. This codification process is known as *Gene Expression*. DNA microarrays technology enables us to mesure the gene expression level under a specific group of conditions. Data mining techniques are needed to analyze the huge volume of all this biological information [1]. The goal of *Biclustering* techniques is to discover transcription factors which determine that a group of genes is co-expressed under a set of conditions.

Several biclustering methods have been proposed in the last few years [2]. For example, in [3] an iterative hierarchical clustering is separately applied to each dimension and biclusters are built by the combination of the obtained results for each dimension. The Cheng and Church algorithm [4] builds biclusters by adding or removing genes or conditions in order to improve the measure of quality called Mean Squared Residue (MSR). In [5], it is proposed an exhaustive biclusters enumeration by means of a bipartite graph-based model, in which nodes were added o removed in order to find subgraphs with maximum weights. The FLOC algorithm [6] improved the method presented in [4] by obtaining a set of biclusters simultaneously and by adding missing values techniques. In [7], a simple linear model, in which normally distributed expression level for each gene or condition was supposed, for gene expression data was applied. Evolutionary computation techniques based on the MSR measure are used in [8,9] or Simulated Annealing in [10]. But, although MSR is used in many algorithms as merit function, it is not the most appropriate measure as the MSR measure can not find scaling patterns when the variance of gene values is high in the bicluster [11]. Recently, the study of the nature of different patterns in biclusters has motivated new techniques based on the search of hyperplanes in high–dimensional data space as interesting patterns share the geometry of linear manifolds [12,13,14]. Other measures to evaluate biclusters have been proposed as fitness function for optimization methods [15].

The gene expression level under a set of conditions can be seen as the values of a discrete random variable. Thus, the linear dependency between two genes can be studied by using the correlation coefficient between two random variables. This fact has motivated the use of the proposed measure in this paper. This measure based on correlations among genes is the main term of the fitness function proposed to evaluate the quality of biclusters in the Scatter Search. Scatter Search is a population-based method that emphasizes systematic processes against random procedures. Thus, the generation of the initial population is not random but a generation method based on diversification [16] is used to generate a set of diverse initial solutions. Moreover, Scatter Search includes an improvement method with the aim of exploiting the diversity provided by the generation and combination method.

This paper is organized as follows. The proposed Scatter Search is presented and different steps such as the improvement method and the fitness function are explained in details in Section 2. Some experimental results from two real datasets are reported in Section 3. Finally, Section 4 outlines the main conclusions of the paper and future works.

2 Description of the Algorithm

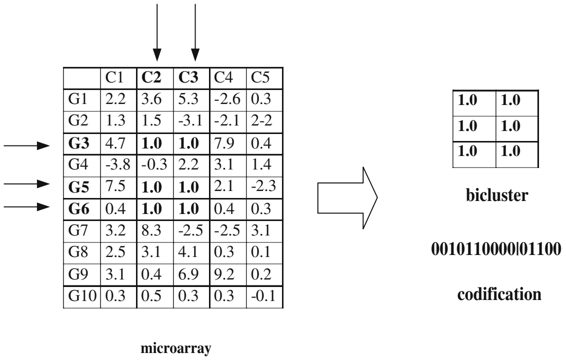
Scatter Search [16] is a population-based optimization metaheuristic which has recently been applied to combinatorial and nonlinear optimization problems. Optimization algorithms based on populations are search procedures where a set of individuals that represent trial solutions evolves in order to find optimum solutions of the problem. On the opposite to other evolutionary heuristics, Scatter Search emphasizes systematic processes against random procedures. Scatter Search uses strategies to diversify and to intensify the search in order to avoid local minima and to find quality solutions.

Basically, the optimization process consists in the evolution of a set called *Reference Set*. This set is initially built with the best solutions from the population, according to the value of their fitness function, and the most scattered ones from the population regarding the previous best solutions. This set is updated by using the *Combination Method* and the *Improvement Method* until it does not change. When the *Reference Set* is stable, it is rebuilt again. That is, the building of the *Reference Set* is based on quality and diversity, but its updating is only guided by quality. Thus, diversity is introduced in the evolutionary process when the initial population is generated and, mainly, when the reference set is rebuilt in each step. The search intensification is due to the improvement method where the solutions are improved by exploiting the knowledge of the problem.

The pseudocode of the proposed Scatter Search for biclustering is presented in Algorithm 1. The Scatter Search process is repeated $numBi$ times where $numBi$ is the number of biclusters to be found and the best solution of the reference set is stored in a set called $Results$ for each iteration. Thus, the $Results$ set is formed by $numBi$ biclusters and it is the output of the Algorithm 1.

2.1 Initialization Phase

Formally, a microarray is a real matrix M composed of N genes and L conditions. The element (i, j) of the matrix means the level of expression of gene i under the condition j. A bicluster B is a submatrix of the matrix M composed of $n \leq N$ rows or genes and $l \leq L$ columns or conditions. Biclusters are encoded by binary strings of length $N + L$. Each of the first N bits of the binary string is related to the genes and the remaining L bits to the conditions from microarray as it can be be seen in Fig. 1.



	C1	**C2**	**C3**	C4	C5
G1	2.2	3.6	5.3	-2.6	0.3
G2	1.3	1.5	-3.1	-2.1	2-2
G3	4.7	**1.0**	**1.0**	7.9	0.4
G4	-3.8	-0.3	2.2	3.1	1.4
G5	7.5	**1.0**	**1.0**	2.1	-2.3
G6	0.4	**1.0**	**1.0**	0.4	0.3
G7	3.2	8.3	-2.5	-2.5	3.1
G8	2.5	3.1	4.1	0.3	0.1
G9	3.1	0.4	6.9	9.2	0.2
G10	0.3	0.5	0.3	0.3	-0.1

microarray

1.0	**1.0**
1.0	**1.0**
1.0	**1.0**

bicluster

0010110000|01100

codification

Fig. 1. Microarray and bicluster {G3,G5,G6|C2,C3} with its codification

The initial population is generated with solutions as diverse as possible. Thus, the diversification generation method [16] takes a binary string, x_i with $i = 1, \ldots, n$ where n is the number of bits, as a seed solution and generates solutions x'_i by the following rule:

$$x'_{1+kh} = 1 - x_{1+kh} \text{ for } k = 0, 1, 2, 3, \ldots, \lfloor n/h \rfloor \tag{1}$$

where $\lfloor n/h \rfloor$ is the largest integer less or equal than n/h and h is an integer less than $n/5$. All remaining bits of x' are equal to that of x.

Algorithm 1. Scatter Search Algorithm for Biclustering

INPUT microarray M, number of biclusters to be found $numBi$, penalization factors M_1 and M_2, maximum number of iterations $numIter$, size of the initial population and size S of the reference set.

OUTPUT Set $Results$ with $numBi$ biclusters.

```
begin
  $num \leftarrow 0$, $Results \leftarrow \emptyset$
  while ($num < numBi$) do
    Initialize population $P$
    $P \leftarrow$ Improvement Method ($P$)
    //Building Reference Set
    $R_1 \leftarrow S/2$ best biclusters from $P$ (according to the fitness function)
    $R_2 \leftarrow S/2$ most scattered biclusters, regarding $R_1$, from $P \setminus R_1$ (according to a distance).
    $RefSet \leftarrow (R_1 \cup R_2)$
    $P \leftarrow P \setminus RefSet$
    //Initialization
    stable $\leftarrow$ FALSE, $i \leftarrow 0$
    while ($i < numIter$) do
      while (NOT stable) do
        $A \leftarrow RefSet$
        $B \leftarrow$ Combination Method($RefSet$)
        $B \leftarrow$ Improvement Method($B$)
        $RefSet \leftarrow S$ best biclusters from $RefSet \cup B$
        if ($A = RefSet$) then
          $stable \leftarrow TRUE$
        end if
      end while
      //Rebuilding Reference Set
      $R_1 \leftarrow S/2$ best biclusters from $RefSet$
      $R_2 \leftarrow S/2$ most scattered biclusters from $P \setminus R_1$
      $RefSet \leftarrow (R_1 \cup R_2)$
      $P \leftarrow P \setminus RefSet$
      $i \leftarrow i + 1$
    end while
    //Storage in Results
    $Results \leftarrow$ the best one from $RefSet$
    $num \leftarrow num + 1$
  end while
end
```

After generating all posible solutions with that seed, if more solutions are necessary, the diversification generation method is applied again by using the last solution as new seed.

2.2 Biclusters Evaluation: Fitness Function

In this work, biclusters with shifting and scaling patterns are desired. A group of genes has a *shifting pattern* when the expression values vary in the addition of a fixed value for all the genes. A group of genes has a *scaling pattern* when the expression values vary in the multiplication of a fixed value for all the genes. Two genes show a shifting and scaling pattern if they are described from (2).

$$g_Y = \alpha g_X + \beta \quad \alpha, \beta \in \mathbb{R} \tag{2}$$

Consequently, two genes with shifting and scaling patterns are linearly dependent.

The correlation coefficient between two variables X and Y measures the grade of linear dependence between them. It is defined by:

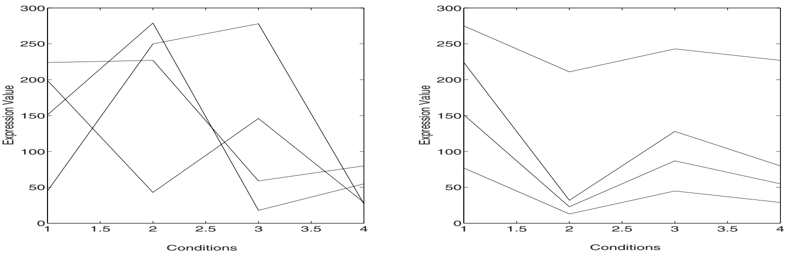
$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_i^n (x_i - \overline{x})(y_i - \overline{y})}{n \sigma_X \sigma_Y} \tag{3}$$

where $cov(X,Y)$ is the covariance of the variables X and Y, $\overline{x}$ and $\overline{y}$ are the average of the values of the variables X and Y and σ_X and σ_Y are the standard deviations of X and Y, respectively. The values for the correlation coefficient vary between -1 and 1. If $\rho(X,Y) = 0$, the variables X and Y are linearly independent, and if $\rho(X,Y) = \pm 1$ the variables are linearly dependent. When the correlation value is equal to -1, the variables X and Y are dependent with negative correlation, that is, when the values of the variable X increase the values of the variable Y decrease linearly.

Given a bicluster B composed of N genes, $B = [g_1, \ldots, g_N]$, the average correlation of B, $\rho(B)$, is defined as follows,

$$\rho(B) = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \rho_{g_i g_j} \tag{4}$$

where $\rho_{g_i g_j}$ is the correlation coefficient between the gene i and the gene j. Note that $\rho_{g_i g_j} = \rho_{g_j g_i}$, therefore, only $\binom{N}{2} = \frac{N(N-1)}{2}$ elements have been considered in the aforementioned sum.



$$\begin{bmatrix} 151 & 279 & 18 & 55 \\ 199 & 43 & 146 & 29 \\ 224 & 227 & 59 & 80 \\ 45 & 250 & 278 & 27 \end{bmatrix} \Rightarrow \rho(B) = 0.17 \qquad \begin{bmatrix} 151 & 23 & 87 & 55 \\ 77 & 13 & 45 & 29 \\ 224 & 32 & 128 & 80 \\ 275 & 211 & 243 & 227 \end{bmatrix} \Rightarrow \rho(B) = 1$$

Fig. 2. Biclusters with lowly–correlated and highly–correlated genes

Fig. 2 presents two biclusters along with their average correlations. It can be observed that the bicluster with perfect shifting and scaling patterns has an average correlation of 1 while that the bicluster without patterns has an average correlation close to 0.

In this work, biclusters with highly–correlated genes and high volume are preferred. Therefore, the fitness function used to evaluate the quality of biclusters is defined by:

$$f(B) = (1 - \rho(B)) + \sigma_\rho + M_1 \left(\frac{1}{nG}\right) + M_2 \left(\frac{1}{nC}\right) \tag{5}$$

where nG and nC are the number of genes and conditions of the bicluster B, respectively, M_1 and M_2 are penalization factors to control the volume of the bicluster B and σ_ρ is the standard deviation of the values ρ_{g_i,g_j} from (4). The standard deviation is included in order to avoid that the value of the average correlation can be high for a bicluster and this bicluster can contain several non–correlated genes with the remaining ones of the bicluster. Best biclusters are the ones with the lowest value for the fitness function. Thus, it has been considered $(1 - \rho(B))$ to evaluate biclusters with highly–correlated genes as good biclusters.

2.3 Improvement Method

Scatter Search uses improvement methods when the solutions have to fulfill some constraints or simply to improve them in order to intensify the search process. This method depends on the problem under study and usually it consists in classical local searches for continuous optimization problems.

The goal of this work is to find biclusters with shifting and scaling patterns. Thus, biclusters with positively–correlated genes are only searched for. Therefore, the proposed improvement method aims at extracting positively–correlated genes either from biclusters of the initial population or from biclusters obtained by the combination method. The pseudocode of the improvement method is presented in the Algorithm 2.

2.4 Building and Rebuilding Method of the Reference Set

Reference set is initially built with the best solutions, according to the value of their fitness function, and the most scattered ones from the population regarding the previous best solutions. The *Hamming* distance is used to measure the distance among biclusters in this work. After getting the stability of reference set in the updating process, it is rebuilt to introduce diversity in the search process. Thus, the reference set is rebuilt with the best biclusters from the updated reference set, according to the fitness function, and the most distant from the population regarding the previously chosen best solutions.

The initial population has to be updated too in the evolutionary process by removing solutions which have already been considered in the building or rebuilding of the reference set. When the initial population is empty, a new

Algorithm 2. Improvement Method

INPUT Bicluster $B = [g_1, \ldots, g_N]$
OUTPUT Bicluster $B' \subseteq B$ such that $\rho_{g_i,g_j} \geq 0 \quad \forall g_i, g_j \in B'$

```
begin
  $i \leftarrow 1$, $B' \leftarrow \{g_i\}$, $R \leftarrow \{\}$
  while $(i \leq N)$ do
    $j \leftarrow i + 1$
    while $(j \leq N)$ do
      if $(\rho_{g_i,g_j} > 0)$ then
        if $(g_j \notin R)$ then
          $B' \leftarrow B' \cup \{g_j\}$
        end if
      else
        $R \leftarrow R \cup \{g_j\}$
        $B' \leftarrow B' \setminus \{g_j\}$
      end if
      $j \leftarrow j + 1$
    end while
    $i \leftarrow i + 1$
  end while
end
```

population is created by using the diversification generation method previously explained.

2.5 Combination Method and Reference Set Updating

New solutions are introduced in the search process by the combination method. Two solutions are combined by using a uniform crossover operator and a new one is generated. All pairs of biclusters in the reference set are combined, generating thus, $S * (S - 1)/2$ new biclusters where S is the size of the reference set. This crossover operator generates randomly a mask and the child is composed of values from the first parent when there is a 1 in the mask, and from the second parent when there is a 0.

After combining all pairs of biclusters, the best solutions from the joining of the previous reference set and the new solutions are chosen. Hence, best solutions according to the value of their fitness function remain in the reference set.

3 Experimental Results

Two well known data sets have been used to show the performance of the proposed algorithm. The first data set is the *human B-cells lymphoma* expression data with 4026 genes and 96 conditions [17]. The second one is the *yeast Saccharomyces cerevisiae* cell cycle expression with 2884 genes and 17 experimental conditions [18]. Both data sets were used in [4] where original data were processed by replacing missing values with random values.

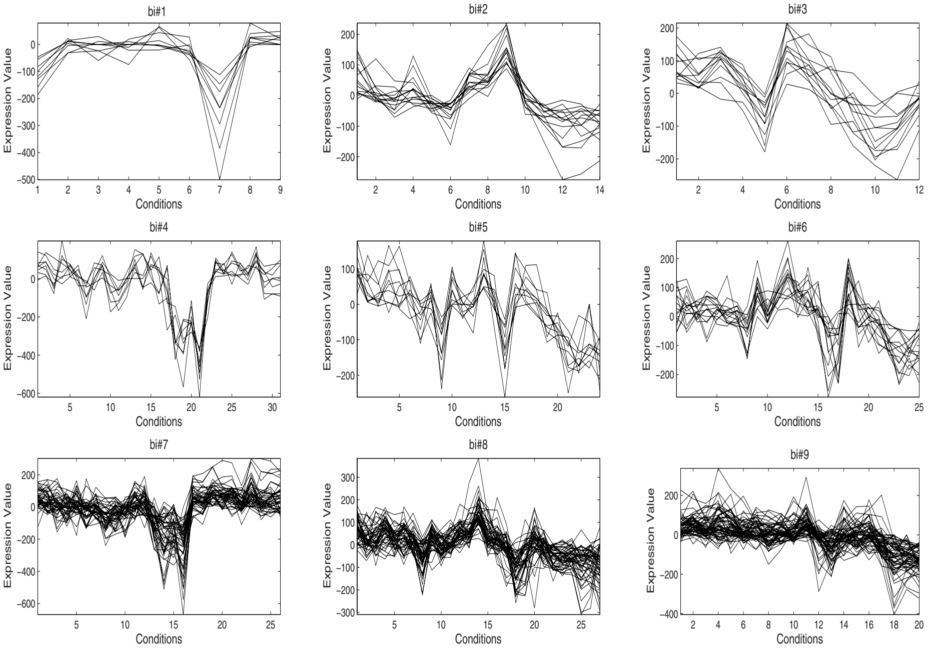


Fig. 3. Biclusters found by Algorithm 1 from lymphoma data set

The main parameters of the Algorithm 1 are as follows: 20 for the maximum number of iterations of the Scatter Search, 10 for the size of the reference set, 200 for the number of solutions of the initial population and 10 for the number of biclusters to be found in each execution. M_1 and M_2 are weights in order to drive the search depending on the required size of biclusters. High values of M_1 and M_2 may be used when biclusters with a lot of genes and conditions are desired.

Fig. 3 presents several biclusters obtained by the application of the Scatter Search to the Lymphoma dataset. These biclusters have been obtained with different values for the weights M_1 and M_2 in order to test their influence. The biclusters $bi\#1$, $bi\#2$ and $bi\#3$ have been found with $M_1 = 1$ and $M_2 = 1$, $bi\#4$, $bi\#5$ and $bi\#6$ with $M_1 = 1$ and $M_2 = 10$ and $bi\#7$, $bi\#8$ and $bi\#9$ with $M_1 = 10$ and $M_2 = 10$. It can be observed how these weights determine the number of genes and conditions of the biclusters. Note that shifting and scaling patterns can clearly be appreciated in all biclusters.

Table 1 shows the information about the biclusters represented in Fig. 3. For each bicluster an identifier of the bicluster, the number of genes, the number of conditions, the volume, the average correlation, the standard deviation, the MSR measure and the variance of gene values are presented. The variance of gene values measures how different the values for the gene expression level are and if this value is high then the MSR measure is high too [11]. It can be observed that

Table 1. Information about biclusters found by Algorithm 1 from lymphoma dataset

id.	Genes	Conditions	Volume	$\rho(B)$	$\sigma_\rho(B)$	MSR	Genes Variance
bi#1	8	9	72	0.92	0.5	2124.6	9742.3
bi#2	14	14	196	0.85	0.4	1657.4	6468.9
bi#3	12	12	144	0.84	0.4	1608.7	8184.2
bi#4	8	31	248	0.82	0.5	3337.4	20042.2
bi#5	10	24	240	0.77	0.4	1963.3	8007.5
bi#6	15	25	375	0.70	0.4	2389.5	7620.4
bi#7	44	26	1144	0.62	0.4	4800.0	11558.8
bi#8	47	27	1269	0.60	0.3	2779.3	6219.8
bi#9	58	20	1160	0.59	0.3	2878.0	6157.4

the higher volume for biclusters, the smaller value for the average correlation. However, all biclusters present a high value for the average correlation which indicates that such biclusters present shifting and scaling patterns. Moreover, the standard deviation is low, that is, the correlation coefficients of each pair of genes have similar values and they are close to the average correlation of the bicluster. Therefore, all biclusters with a high average correlation do not contain non–correlated genes as it can be observed in Fig. 3. Most of papers published in the literature present algorithms based on the MSR measure and a bicluster is considered a high—quality bicluster for the Lymphoma dataset if the value

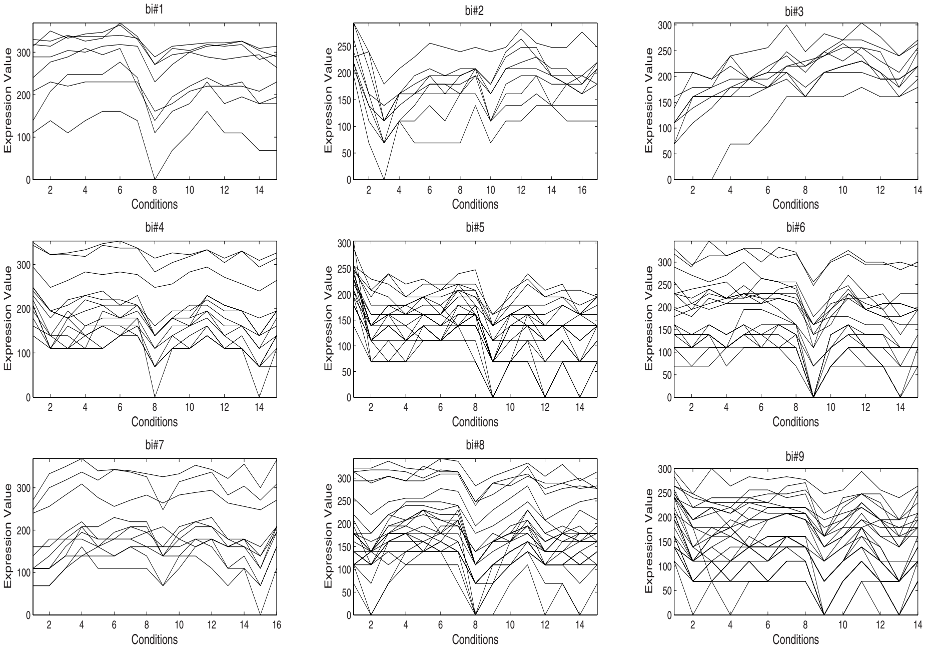


Fig. 4. Biclusters found by Algorithm 1 from yeast dataset

of its MSR is less than 1200 [4,8]. It can be noticed that the bicluster *bi#4* has a value of MSR much greater than 1200 since its variance of gene values is high. However, it can be considered a high–quality bicluster because it presents shifting and scaling patterns as its average correlation is 0.82.

In Fig. 4 several biclusters found by the proposed Scatter Search from Yeast dataset are shown. These biclusters have been obtained with values $M_1 = 1$ and $M_2 = 10$ in order to obtain biclusters with a moderate volume. From a geometrical point of view, it can be noticed that the genes present a similar behavior under a set of conditions. Information about these biclusters is presented in Table 2. In the literature, a bicluster is considered a high—quality bicluster for the Yeast dataset if the value of its MSR is less than 300 [4,8]. However, it can be appreciated how several biclusters that have values of MSR greater than 300 present high average correlations, and therefore, shifting and scaling patterns.

Table 2. Information about biclusters found by Algorithm 1 from yeast dataset

id.	Genes	Conditions	Volume	$\rho(B)$	$\sigma_\rho(B)$	MSR	Genes variance
bi#1	8	15	120	0.83	0.3	273.4	1028.5
bi#2	9	17	153	0.74	0.4	306.6	1230.9
bi#3	9	14	126	0.84	0.4	367.8	1740.7
bi#4	14	15	210	0.76	0.4	257.7	854.0
bi#5	20	15	300	0.75	0.3	437.9	1367.5
bi#6	19	15	285	0.75	0.4	342.9	1119.4
bi#7	11	16	176	0.70	0.4	245.7	842.2
bi#8	23	15	345	0.68	0.4	369.4	991.7
bi#9	27	14	378	0.72	0.3	332.2	1038.0

Figs. 5 and 6 show the performance of the Scatter Search with and without improvement method from Lymphoma dataset (Figures. 5a) and 5b), respectively) and Yeast dataset (Figures. 6a) and 6b), respectively). The evolution of the average fitness function, average correlation and average volume for all biclusters of the reference set throughout the iterations is presented. It can be observed how the average correlation increases when the fitness function decreases during the evolutionary process. When the improvement method is not included in the Scatter Search scheme, the convergence of the fitness function to an optimum solution is very slow and it is necessary more iterations. Obviously, the improvement method leads to lower volumes for both datasets due to this method just selects the positively–correlated genes by removing the negatively–correlated genes.

Table 3 presents the average correlation and average volume for ten biclusters obtained by the Algorithm 1, ignoring the improvement method and considering it in order to show the importance of such method in the proposed Scatter Search. Notice that the average correlation is lower when the improvement method is not included due to the negatively–correlated genes can be considered.

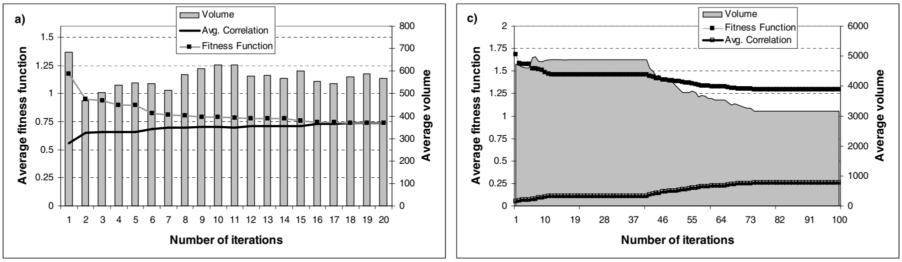


Fig. 5. Evolutionary process for Lymphoma dataset: a)improvement method considered; b) no improvement method considered

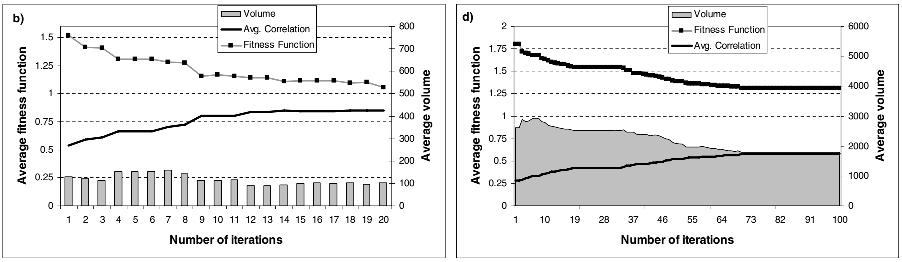


Fig. 6. Evolutionary process for Yeast dataset: a)improvement method considered; b) no improvement method considered

Table 3. Optimal solution with and without improvement method

	With Improvement Method		Without Improvement Method	
	Correlation	Volume	Correlation	Volume
lymphoma	0.74	600	0.02	14110
yeast	0.80	154	0.16	7065

4 Conclusions

In this paper, a Scatter Search for finding biclusters from gene expression data has been presented. The proposed Scatter Search has used as merit function to evaluate biclusters a new measure based on correlations among genes with the aim of obtaining biclusters with shifting and scaling patterns. Moreover an improvement method, which consist in removing negatively–correlated genes from biclusters, has been incorporated to intensify the search. Experimental results from human B-cell lymphoma data set and yeast cell cycle data set have been reported revealing the good convergence and remarkable performance of the proposed method and measure.

Future works will focused on some improvements for the proposed algorithm with regard to the overlapping among genes and the comparison with other biclustering techniques using Gene Ontology Database [14].

Acknowledgments

The financial support given by the Spanish Ministry of Science and Technology, project TIN2007-68084-C00 and by the Junta de Andalucía, project P07-TIC-02611 is acknowledged.

References

1. Larranaga, P., et al.: Machine learning in bioinformatics. Briefings in Bioinformatics 7(1), 86–112 (2006)
2. Busygin, S., Prokopyev, O., Pardalos, P.: Biclustering in data mining. Computers and Operations Research 35(9), 2964–2987 (2008)
3. Getz, G., Levine, E., Domany, E.: Couple two-way clustering analysis of gene microarray data. In: Proceedings of the National Academy of Sciences (PNAS) of the USA, pp. 12079–12084 (2000)
4. Cheng, Y., Church, G.: Biclustering of Expression Data. In: Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology, vol. 8, pp. 93–103 (2000)
5. Tanay, A., Sharan, R., Shamir, R.: Discovering statistically significant biclusters in gene expression data. Bioinformatics 18(90001), 136–144 (2002)
6. Yang, J., Wang, H., Wang, W., Yu, P.: Enhanced biclustering on expression data. In: 3rd IEEE Simposium on Bioinformatics and Bioengeneering, pp. 321–327 (2003)
7. Bergmann, S., Ihmels, J., Barkai, N.: Iterative signature algorithm for the analysis of large-scale gene expression data. Physical Review E 67(031902) (2003)
8. Divina, F., Aguilar-Ruiz, J.: Biclustering of Expression Data with Evolutionary Computation. IEEE Transactions on Knowledge and Data Engineering 18(5), 590–602 (2006)
9. Mitra, S., Banka, H.: Multi-objective evolutionary biclustering of gene expression data. Pattern Recognition 39(12), 2464–2477 (2006)
10. Bryan, K.: Biclustering of Expression Data Using Simulated Annealing. In: Proceedings of the 18th IEEE International Symposium on Computer-Based Medical Systems, USA, pp. 383–388 (2005)
11. Aguilar-Ruiz, J.: Shifting and scaling patterns from gene expression data. Bioinformatics 21(20), 3840–3845 (2005)
12. Harpaz, R., Haralick, R.: Mining Subspace Correlations. In: IEEE Symposium on Computational Intelligence and Data Mining, pp. 335–342 (2007)
13. Zhao, H., Liew, A., Xie, X., Yan, H.: A new geometric biclustering algorithm based on the Hough transform for analysis of large-scale microarray data. Journal of Theoretical Biology 251(2), 264–274 (2008)
14. Gan, X., Liew, A., Yan, H.: Discovering biclusters in gene expression data based on high-dimensional linear geometries. BMC Bioinformatics 9(209), 1–15 (2008)
15. Nepomuceno, J.A., Troncoso Lora, A., Aguilar-Ruiz, J.S., García-Gutiérrez, J.: Biclusters Evaluation Based on Shifting and Scaling Patterns. In: Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X. (eds.) IDEAL 2007. LNCS, vol. 4881, pp. 840–849. Springer, Heidelberg (2007)
16. Marti, R., Laguna, M.: Scatter Search. In: Methodology and Implementation in C. Kluwer Academic Publishers, Boston (2003)
17. Alizadeh, A., et al.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403, 503–511 (2000)
18. Cho, R., et al.: A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle. Molecular Cell 2(1), 65–73 (1998)