Stability and Performances in Biclustering Algorithms

Maurizio Filippone[(1)], Francesco Masulli[(2,3,4)], Stefano Rovetta[(2,3)]

(1) Department of Computer Science, University of Sheffield, United Kingdom
(2) Department of Computer and Information Sciences, University of Genova, Genova, Italy
(3) CNISM Genova Research Unit, Genova, Italy
(4) Center for Biotechnology, Temple University, Philadelphia, USA
{rovetta,masulli}@disi.unige.it

Abstract. Stability is an important property of machine learning algorithms. Stability in clustering may be related to clustering quality or ensemble diversity, and therefore used in several ways to achieve a deeper understanding or better confidence in bioinformatic data analysis. In the specific field of fuzzy biclustering, stability can be analyzed by porting the definition of existing stability indexes to a fuzzy setting, and then adapting them to the biclustering problem. This paper presents work done in this direction, by selecting some representative stability indexes and experimentally verifying and comparing their properties. Experimental results are presented that indicate both a general agreement and some differences among the selected methods.

1 Introduction

Many bioinformatic data sets come from DNA microarray experiments and are normally given as a rectangular m by n matrix $X = (x_{ij})_{mn}$, where each column represents a feature (e.g., a gene) and each row represents a data point or condition (e.g., a patient), and value x_{ij} is the expression of i-th gene in j-th condition. The analysis of microarray data sets can provide valuable information on the biological relevance of genes and correlations among them.

Biclustering (also known under other names like *co-clustering* and *two-way clustering*) [17] is a methodology allowing for feature set and data points clustering simultaneously, i.e., to find clusters of samples possessing similar characteristics together with features creating these similarities. In other words, biclustering answers the question: *What characteristics make "similar" objects similar to each other?*

The output of biclustering is not a partition or hierarchy of partitions of either rows or columns, but a partition of the whole matrix into sub-matrices or patches. We can obtain different biclustering structures: single bicluster, different non-overlapping structures (as exemplified in Fig. 1), and overlapping with or without structure.

The goal of biclustering is to find as many patches as possible, and to have them as large as possible, while maintaining strong homogeneity within patches. This task is reported to be an NP-complete task [17,21].

In gene expression microarray data analysis biclustering methods allow us to identify genes with similar behavior with respect to different conditions. A single patch represents a given subset of genes in a given subset of conditions.

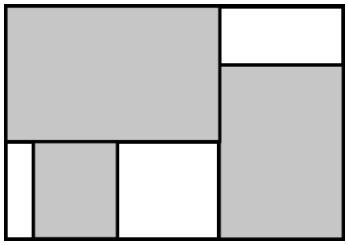

Fig. 1. Example of biclusters, shown as "patches".

Biclustering algorithms able to find largest biclusters from DNA microarray data that do not exceed an assigned homogeneity constraint [3] are necessary as they convey relevant biological information able to support important tasks, such as:

- *identification of coregulated genes* and/or *specific regulation processes* by identifying sets of genes that, under specific conditions, exhibit coherent activations;
- *gene functional annotation* by extending the class label shared by the the majority of genes in the bicluster to the remaining non-annotated genes of the same bicluster;
- *sample* and/or *tissue classification*, since considering the diagnosis of a specific pathology biclusters identify the different responses to treatment, and then the group of genes to be used as the most effective probe.

In the next section we introduce three non-pairwise indexes suited to studying the stability of biclustering algorithms. In Sect. 3 after stating the fuzzy framework for biclustering, we give a short overview of two biclustering algorithms based on this approach. Sect. 4 presents the experimental study of their stabilities. In Sect. 5 we draw the conclusions.

2 Stability Indexes

2.1 Stability of learning machines

In machine learning *stability* among solutions has been related to some important properties of learners, e.g., generalization [15,5].

When learning is formulated as an optimization process, the reliability of a solution can be inferred from its robustness with respect to perturbations in the data, parameters, or training process. A learning algorithm is *stable* if it produces robust solutions. If the analysis of a cost function landscape is possible, this relationship can be proved, but often we deal with difficult cost functions and then we must estimate the stability of a learning machine from empirical observations.

In case of biclustering algorithms we can evaluate their stability by means of indexes measuring the degree of similarity or overlap among solution sets, as for clustering, but preferring computationally parsimonious indexes, as the dimension of the solution space is very large (product of the row- and column-dimensions of the data matrix). We shall now describe the stability indexes we have considered.

2.2 Normalized Mutual Information between partitions

The Normalized Mutual Information between partitions is a pairwise index defined as [9]:

$$NMI = \frac{-2\sum_{i=1}^{|A|}\sum_{j=1}^{|B|} N_{ij}\log\left(\frac{N_{ij}N}{N_iN_j}\right)}{\sum_{i=1}^{|A|} N_i\log\left(\frac{N_i}{N}\right)+\sum_{j=1}^{|B|} N_j\log\left(\frac{N_j}{N}\right)} \tag{1}$$

where A and B are two partitions; $|A|$ is the number of biclusters in partition A; $|B|$ is the number of biclusters in partition B ($|A|$ and $|B|$ can be different); i a bicluster in A; j a bicluster in B; N_i and N_j are respectively the cardinalities of bicluster i and j; N_{ij} the cardinality of the intersection of bicluster i and bicluster j (i.e., number of data points which are assigned to bicluster i in A and to bicluster j in B); and N is the cardinality of the dataset.

Note that computation of NMI doesn't involve relabeling of biclusters, moreover we have NMI $= 1$ for perfect overlap and NMI $= 0$ (asymptotically) for completely independent partitions

2.3 Jaccard coefficient

Jaccard coefficient is another pairwise index and is defined as the ratio of cardinalities of the intersection of two sets to their union [12]:

$$J(A,B) = \frac{|A\cap B|}{|A\cup B|}. \tag{2}$$

After Jaccard's paper, this coefficient and some related ones have been proposed in several occasions, see, e.g., [18,2,4].

2.4 Entropy of Consensus Matrix

The *co-association matrix* [20] is a $N\times N$ matrix $\mathbf{M^{(s)}}$. Their elements $m_{ij}^{(s)}$ indicates whether two matrix elements i and j are in the same bicluster in experiment number s.

The *consensus matrix* $\mathbf{M}$ averages co-association matrixes over all experiments [20]. Then: (a) if all the clusterers agree on joining objects i and j in the same bicluster, $m_{ij} = 1$; (b) if all clusterers agree that objects i and j are in different clusters, $m_{ij} = 0$; (c) otherwise, if there is disagreement on joint membership of the two objects, m_{ij} between 0 and 1. Note that in the case of the largest disagreement, where i and j are in the same biclusters in exactly $L/2$ of the partitions $P_1,\ldots,P_L$, $m_{ij} = 0.5$.

We can define a global index called the *Entropy of Consensus Matrix* as the *averaged entropy* of the cells of $\mathbf{M}$ as [15]:

$$H(\mathbf{M}) = -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(m_{ij}\log\left(m_{ij}\right)+\left(1-m_{ij}\right)\log\left(1-m_{ij}\right)\right) \tag{3}$$

3 Fuzzy biclustering

3.1 Fuzzy framework for biclustering

Let x_{ij} be the expression level of the i-th gene in the j-th condition. A *bicluster* is defined as a subset of the $m \times n$ data matrix X, i.e., a bicluster is a pair $(\mathbf{g}, \mathbf{c})$, where $\mathbf{g} \subset \{1,\ldots,m\}$ is a subset of genes and $\mathbf{c} \subset \{1,\ldots,n\}$ is a subset of conditions [3,11,16,24].

The *size* (or volume) n of a bicluster is usually defined as the number of elements in the gene expression matrix X belonging to it, that is the product of the cardinalities $n_g = |\mathbf{g}|$ and $n_c = |\mathbf{c}|$:

$$n = n_g \cdot n_c \tag{4}$$

The *bicluster mean*, *bicluster row mean*, and *bicluster column mean* are defined as:

$$x_{IJ} = \frac{1}{n}\sum_{i\in\mathbf{g}}\sum_{j\in\mathbf{c}} x_{ij} \qquad x_{iJ} = \frac{1}{n_c}\sum_{j\in\mathbf{c}} x_{ij} \qquad x_{Ij} = \frac{1}{n_g}\sum_{i\in\mathbf{g}} x_{ij} \tag{5}$$

Hartigan [11] proposed the following definition for a biclustering *residue*:

$$d_{ij} = x_{ij} - x_{IJ}. \tag{6}$$

A residue suited for general "trends" of DNA microarray data analysis has been proposed by Cheng and Church [3,7]:

$$d_{ij} = x_{ij} - (x_{IJ} + \alpha_i + \beta_j) = x_{ij} + x_{IJ} - x_{iJ} - x_{Ij}. \tag{7}$$

This definition takes into account that for ideal constant row biclusters, each element x_{ij} is equal to the bicluster mean x_{IJ} plus an offset $\alpha_i = x_{iJ} - x_{IJ}$, and similarly, for ideal constant column biclusters, each element x_{ij} is equal to the bicluster mean x_{IJ} plus an offset $\beta_j = x_{Ij} - x_{IJ}$.

Following [3], we are interested in the largest biclusters from DNA microarray data that do not exceed an assigned homogeneity constraint. To this aim we can utilize one of those definitions of biclustering homogeneity, or better, of *biclustering (crisp) heterogeneity*:

$$H = \sum_{i\in\mathbf{g}}\sum_{j\in\mathbf{c}} d_{ij}^2 \qquad \text{Sum-squared residue [7];} \tag{8}$$

$$G = \frac{1}{n}\sum_{i\in\mathbf{g}}\sum_{j\in\mathbf{c}} d_{ij}^2 = \frac{H}{n} \qquad \text{Mean Squared residue [3].} \tag{9}$$

G measures the *bicluster heterogeneity*, i.e., the *difference between the actual value of an element x_{ij} and its expected value* as predicted from the corresponding row mean, column mean, and bicluster mean.

Hartigan's residue (Eq. 6) can capture constant biclusters only, while Cheng and Church's residue (Eq. 7) can capture biclusters with constant rows, constant columns, and even coherent values (addictive models) and is then best suited for gene expression data analysis [17].

In order to obtain a fuzzy formulation of the biclustering problem, we should first set biclustering in a (crisp) set theory framework, and then we can extend this setting

to a fuzzy set formulation. To this aim, for each bicluster we assign two membership vectors, one for the rows and one other for the columns, denoting them respectively **a** and **b**. In a *crisp sets framework* row i and column j can either belong to the bicluster or not. An element x_{ij} of X belongs to the bicluster if both $a_i = 1$ and $b_j = 1$, i.e., its membership to the bicluster is $u_{ij} = \text{and}(a_i, b_j)$. Therefore, we can define the cardinality of a bicluster as follows:

$$n = \sum_i \sum_j u_{ij} \tag{10}$$

To proceed toward a fuzzy set theory framework, we allow membership u_{ij}, a_i and b_j to belong in the interval $[0,1]$. The membership u_{ij} of an element x_{ij} of X to the bicluster can be obtained by the *aggregation* of row and column memberships, using, e.g..:

$$u_{ij} = a_i b_j \qquad \text{(product)} \tag{11}$$

or

$$u_{ij} = \frac{a_i + b_j}{2} \qquad \text{(average)}. \tag{12}$$

The fuzzy cardinality of the bicluster is defined as the sum of the memberships u_{ij} for all i and j and, formally, is still given in eq. 10. The same happens for *fuzzy residue* that is formally identical to the definitions in eq.s 6 and 7 (we will use the second one, as we will work with DNA microarray data), but variables included in them must be interpreted as *fuzzy bicluster mean, fuzzy bicluster row mean, fuzzy bicluster column mean*:

$$x_{IJ} = \frac{\sum_i \sum_j u_{ij} x_{ij}}{\sum_i \sum_j u_{ij}}, \; x_{iJ} = \frac{\sum_j u_{ij} x_{ij}}{\sum_j u_{ij}}, \; x_{Ij} = \frac{\sum_i u_{ij} x_{ij}}{\sum_i u_{ij}} \tag{13}$$

We can introduce now the definitions of *Fuzzy Sum-squared residue g* and *Fuzzy Mean Squared residue G* that generalize the *bicluster heterogeneity* concept:

$$g = \sum_i \sum_j u_{ij} d_{ij}^2 \qquad G = \frac{1}{n} \sum_i \sum_j u_{ij} d_{ij}^2 \tag{14}$$

3.2 Minimum Sum-squared Residue for Fuzzy Co-clustering Algorithm

We shall now present two fuzzy biclustering algorithms, namely the *Minimum Sum-squared Residue for Fuzzy Co-clustering* (MSR-FCC) algorithm [23] and the Possibilistic Biclustering (PCB) [8], that are inspired from the fuzzy central clustering algorithms. In the versions applied in this paper, both methods employ the product aggregator (11) in the computation of the membership of an element of the data matrix X to a bicluster. The former imposes the *probabilistic constraint* on memberships, used, e.g., in the Fuzzy C-Means [1], according to which the sum of the membership values of a matrix element to all the biclusters must be equal to one. PCB, instead, applies more relaxed constraints to the memberships, following the so-called possibilistic clustering framework [13].

The *Minimum Sum-squared Residue for Fuzzy Co-clustering* (MSR-FCC) algorithm [23] is based on the constrained minimization of a generalization of the fuzzy central clustering objective function:

$$\begin{aligned} J = & \sum_{cdij} a_{ci} b_{dj} d^2_{cdij} + T_a \sum_{ci} a^2_{ci} + T_b \sum_{dj} b^2_{dj} + \\ & \sum_i \lambda_i \left(\sum_c a_{ci} - 1 \right) + \sum_j \gamma_j \left(\sum_d b_{dj} - 1 \right), \\ c,q \in \{1,...,C\}, \quad & d,p \in \{1,...,D\}, \quad T_a, T_b, \lambda_i, \gamma_j \in \mathbb{R} \end{aligned} \tag{15}$$

where C is the number of data clusters; D is the number of feature clusters; the 1^{st} term is the Fuzzy Sum-squared residue g; the 2^{nd}, 3^{rd} terms are fuzzy Gini indexes; while the last two terms are Lagrange constraints due to the probabilistic constraints imposed for row- and column-memberships normalization.

By setting the derivatives of J with respect to the memberships a_i and b_j to zero we obtain these necessary conditions:

$$a_{ci} = \frac{1}{C} + \frac{1}{2T_a} \left(\frac{1}{C} \sum_{pdj} b_{dj} d^2_{pdij} - \sum_{dj} b_{dj} d^2_{cdij} \right) \tag{16}$$

$$b_{dj} = \frac{1}{D} + \frac{1}{2T_b} \left(\frac{1}{D} \sum_{cqi} a_{ci} d^2_{cqij} - \sum_{ci} a_{ci} d^2_{cdij} \right) \tag{17}$$

Fig. 2 shows a generic algorithm for fuzzy biclustering iterating these necessary conditions for minimizing the the objective function J. One thing that may be worth noticing is that, in the MSR-FCC optimization process, iterations are not necessarily contraction mappings, and hence convergence of the Picard iterations is not guaranteed as required by related fixed-point theorems [6,10].

1. Initialize ε and the memberships a_{ci} b_{dj} $\forall c,d,i,j$
2. Compute d^2_{cdij} $\forall c,d,i,j$
3. Update a_{ci} $\forall c,i$
4. Update b_{dj} $\forall d,j$
5. Compute $\Delta_{\max} = \max \left\{ \{|a_{ci} - a'_{ci}| \forall c,i\} \cup \{|b_{dj} - b'_{dj}| \forall d,j\} \right\}$
6. **if** $\Delta_{\max} < \varepsilon$ then **stop**
7. **else jump** to step 2

Fig. 2. Generic fuzzy biclustering algorithm.

3.3 Possibilistic Biclustering Algorithm

The Possibilistic Biclustering (PCB) algorithm proposed by our group [8] is based on the possibilistic clustering framework proposed by Krishnapuram and Keller in 1993 [13], that relaxes the clustering probabilistic constraint to:

$$u_{pq} \in [0,1] \quad \forall p,q; \tag{18}$$

$$0 < \sum_{q=1}^{r} u_{pq} < r \quad \forall p; \qquad \bigvee_{p} u_{pq} > 0 \quad \forall q. \tag{19}$$

These minimal constraints say that clusters cannot be empty and each pattern must be assigned to at least one cluster.

In PBC framework we can go to minimize this objective function [8,14]:

$$J = \sum_{ij} a_i b_j d_{ij}^2 + \lambda \sum_i (a_i \ln(a_i) - a_i) + \mu \sum_j (b_j \ln(b_j) - b_j) \tag{20}$$

where the first term is the *fuzzy squared residue H*, while the other two are penalization terms. The parameters λ and μ control the size of the bicluster. Setting the derivatives of J with respect to the memberships a_i and b_j to zero we obtain:

$$a_i = \exp\left(-\frac{\sum_j b_j d_{ij}^2}{\lambda}\right) \qquad b_j = \exp\left(-\frac{\sum_i a_i d_{ij}^2}{\mu}\right) \tag{21}$$

These necessary conditions for the minimization of J_B together with the definition of fuzzy residue d_{ij} can be used by an algorithm able to find a numerical solution for the optimization problem (Picard iteration), as show in Fig. 2.

4 Experimental analysis

We studied the stability and the performances of the Minimum Sum-squared Residue for Fuzzy Co-clustering (MSR-FCC) algorithm and of the Possibilistic Biclustering (PBC) algorithm using the Yeast data set [22] that is DNA microarray data set measuring the gene expression of 2879 genes in 17 conditions.

We used this data set and some its modifications obtained by adding to it uniform random noise of different levels. The two algorithms we initialized with random memberships, $\varepsilon = 10^{-6}$ and α-cut= .5 for defuzzification of data matrix elements' membership. The number of requested biclusters was 2×2 for MSR-FCC, and one for PBC. Stability indexes were all evaluated in pairwise way, to assess the overlap between each individual experiment and a reference solution (the one obtained without noise). The heterogeneity index is defined as $\Omega \equiv G$ (Eq. 9).

Table 1 reports the experimental results of stability analysis. All presented results are averaged on 10 runs. The Normalized Mutual Information (H), the Jaccard Coefficient (NMI), and the Entropy of Consensus Matrix (J) show a good concordance, confirming in such a way their usefulness, but at the same time suggesting that the information they provide is redundant. Concordance is to be expected at the extreme values (partitions matching completely, partitions completely independent); however, experimental results show that this holds even for intermediate values.

In general, the MSR-FCC method appears to be very stable. The possibilistic version has a certain dependence from user-defined parameters, but Table 1 shows that

Table 1. Results on yeast data. Average indexes with uniform noise added.

	Noise	*H*	NMI	*J*	*Largest n*	*avg.*Ω
	0%	0.000	1.00	1.00	23518	493.44
	1%	0.002	0.96	0.99	23485	491.75
PBC ($\mu = 0.6, \lambda = 80$)	2%	0.022	0.83	0.91	23309	479.78
	4%	0.057	0.59	0.78	18576	431.26
	8%	0.136	0.20	0.28	6820	271.55
	0%	0.000	1.00	1.00	15496	330.5
	1%	0.004	0.93	0.98	15470	329.32
PBC ($\mu = 0.34, \lambda = 120$)	2%	0.010	0.87	0.95	15236	326.81
	4%	0.022	0.75	0.88	14196	313.27
	8%	0.100	0.28	0.34	6320	220.66
	0%	0.000	1.00	1.00	15228	935.43
	1%	0.000	0.99	1.00	15228	935.94
MSR-FCC ($T_u = 1000, T_v = 1000$)	2%	0.001	0.99	1.00	15264	935.74
	4%	0.001	0.98	0.99	15273	936.35
	8%	0.002	0.97	0.98	15291	936.63

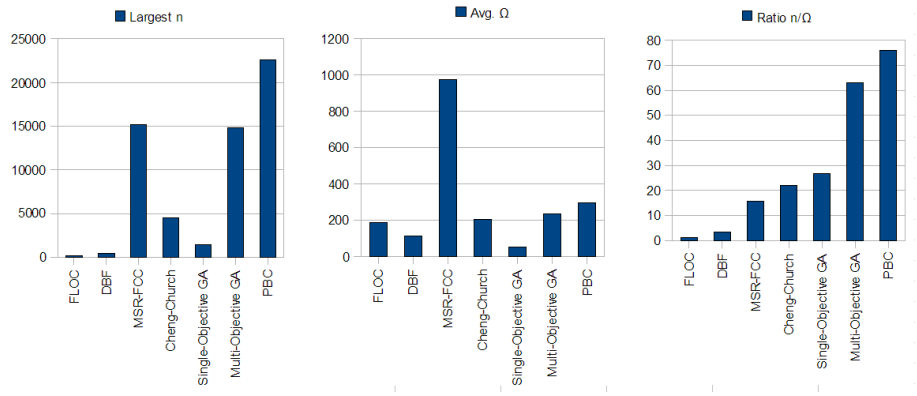

Fig. 3. Comparative performances of some biclustering techniques: heterogeneity Ω (left), size n (center), and synthetic indicator n/Ω (right).

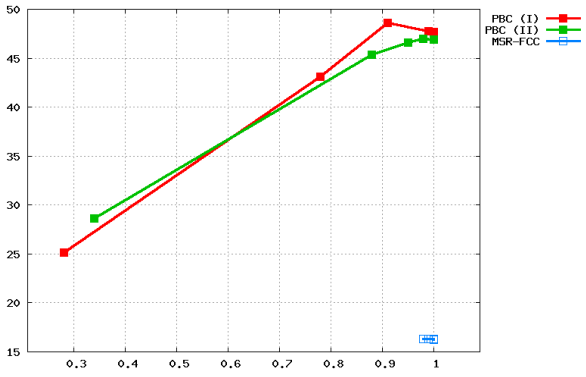

Fig. 4. Synthetic performance (n/Ω) as a function of Jaccard index J, computed from Table 1. PBC (I) and (II) indicate PBC with the first and second sets of parameters respectively. MSR-FCC is consistently worse.

heterogeneity is always better than for the competitive MSR-FCC method, even if stability is slightly lower. We can see from the table that the largest biclusters are obtained by PBC with the first set of parameters ($\lambda = 0.6$, $\mu = 80$); the heterogeneity of solutions obtained with the other set of parameters is larger, but this is obvious for smaller biclusters. On the other hand, MSR-FCC shows superior stability, but this is not associated to equally high performance.

Figure 3, based on data from [8], shows the relationship between the two quality criteria as obtained also on the other biclustering techniques outlined previously in this paper (namely, DBF [26], FLOC [25], Cheng-Church [3], Single-objective GA [19], and Multi-objective GA [19]). The first column graph shows the heterogeneity level (Ω) and the second graph shows bicluster size (n), computed, in case of fuzzy biclustering algorithms, after the final defuzzication. The third graph concerns the value of the performance index, defined as the ratio n/Ω: the largest the ratio, the better the result. Finally, as it is reasonable to expect, the value of the performance index shows low correlation with respect to stability.

The experiments reported in Figure 3 are different from those of the present paper, but the results on MSR-FCC and PBC are very similar. The competitive method is clearly the most stable. However, by taking into account performance along with stability, we see that PBC with both sets of parameters performs consistently better than MSR-FCC. This can be observed by computing the performance index n/Ω for the results in Table 1, as a function of stability. This is illustrated by the graphs in Figure 4, where n/Ω is plotted for the three methods as a function of J.

5 Conclusions

The stability of a learning algorithm is reported in the literature as related to is performance, e.g. to the generalization capabilities in supervised algorithms [5]. In ensemble

methods diversity of base learners is often exploited to increase stability of the ensemble.

In standard clustering complex relationships among stability indexes and performances have been experimental shown in [15] using k-means algorithm. In this study we have studied the relationships between similarity and performances in biclustering that is a generalization of clustering as it is a methodology allowing for feature set and data points clustering simultaneously.

While performances in clustering can be related to the *representation error* (that is the count of data points in each cluster disagreeing with the majority label in that cluster, summed over all clusters and expressed as a percentage), as done, e.g., in [15], performances for biclustering can be related to bicluster cardinality and heterogeneity.

In this study we have employed the Normalized Mutual Information, the Jaccard Coefficient, and the Entropy of Consensus Matrix as stability indexes and we have considered two fuzzy biclustering algorithms, namely the *Minimum Sum-squared Residue for Fuzzy Co-clustering* (MSR-FCC) algorithm [23] and the *Possibilistic Biclustering* (PBC) algorithm [8].

We noticed that for PBC there is a relationship between all stability indexes and the performance index, while MSR-FCC is more stable, but shows a worse performance index that does not depend on any stability index. Moreover, in general, the three considered stability indexes are strongly correlated.

Acknowledgments

We thank Luca Zini for programming support and Giorgio Valentini for the useful suggestions provided while discussing this work.

References

1. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Plenum, New York (1981).
2. P.A. Brivio , E. Binaghi , A. Rampini , P. Ghezzi, "A Fuzzy Set-Based Accuracy Assessment of Soft Classifications" *Pattern Recognition Letters*, vol. 20, pp. 935–948, 1999.
3. Y. Cheng and G. M. Church, "Biclustering of expression data." *Proc Int Conf Intell Syst Mol Biol*, vol. 8, pp. 93–103, 2000.
4. Russell G. Congalton,: Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. CRC Press (1999).
   K. Bryan, P. Cunningham, and N. Bolshakova, "Biclustering of expression data
5. O. Bousquet and A. Elisseeff, "Stability and generalization," *J. Mach. Learn. Res.*, vol. 2, pp. 499–526, 2002.
6. F. Browder, "Nonexpansive nonlinear operators in a banach space," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 54, pp. 1041–1044, 1965.
7. H. Cho, I.S. Dhillon, Y. Guan, and S. Sra, "Minimum Sum-Squared Residue Co-clustering of Gene Expression Data,' *Proc. Fourth SIAM Int. Conf. on Data Mining*, pp. 114–125, 2004.
8. M. Filippone, F. Masulli, S. Rovetta, S. Mitra, and H. Banka, "Possibilistic approach to biclustering: An application to oligonucleotide microarray data analysis." in Lecture Notes in Bioinformatics, C. Priami, Ed., vol. 4210. Springer, 2006, pp. 312–322.

9. A.L.N. Fred,A.K. Jain, "Robust Data Clustering," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03), vol. 2, pp.128–136, 2003.
10. D. Göhde, "Zum prinzip der kontraktiven abbildung," *Math. Nachr.*, vol. 30, pp. 251–258, 1965.
11. J. A. Hartigan, "Direct clustering of a data matrix," *Journal of the American Statistical Association*, vol. 67(337), no. 337, pp. 123–129, 1972.
12. P. Jaccard, "Étude comparative de la distribution florale dans une portion des alpes et des jura," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901
13. R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.
14. ——, "The possibilistic *C*-Means algorithm: insights and recommendations," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 385–393, August 1996.
15. L. I. Kuncheva and D. P. Vetrov, "Evaluation of stability of k-means cluster ensembles with respect to random initialization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 1798–1808, 2006.
16. S. Y. Kung, M. W. Mak, and I. Tagkopoulos, "Multi-metric and multi-substructure biclustering analysis for gene expression data," *Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference (CSB'05)*, 2005.
17. S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: A survey," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 1, pp. 24–45, 2004.
18. F. Masulli and A. Schenone, " A fuzzy clustering based segmentation system as support to diagnosis in medical imaging" *Artificial Intelligence in Medicine*, vol. 16, pp. 129–147, 1999.
19. S. Mitra and H. Banka, "Multi-objective evolutionary biclustering of gene expression data," *Pattern Recogn.*, vol. 39, no. 12, pp. 2464–2477, 2006.
20. S. Monti, P. Tamayo, J. Mesirov, and T. Golub, "Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data," *Machine Learning*, vol. 52, no. 1-2, pp. 91–118, July 2003.
21. R. Peeters, "The maximum edge biclique problem is NP-Complete," *Discrete Applied Mathematics*, vol. 131, pp. 651–654, 2003.
22. S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, "Systematic determination of genetic network architecture," *Nature Genetics*, vol. 22, no. 3, 1999.
23. W.-C. Tjhi and L. Chen, "Minimum sum-squared residue for fuzzy co-clustering," *Intelligent Data Analysis*, vol. 10, no. 3, pp. 237 – 249, 2006.
24. H. Turner, T. Bailey, and W. Krzanowski, "Improved biclustering of microarray data demonstrated through systematic performance tests," *Computational Statistics and Data Analysis*, vol. 48, no. 2, pp. 235–254, 2005.
25. J. Yang, H. Wang, W. Wang, and P. Yu, "Enhanced biclustering on expression data," in *BIBE '03: Proceedings of the 3rd IEEE Symposium on BioInformatics and BioEngineering*. Washington, DC, USA: IEEE Computer Society, 2003, p. 321.
26. Zhang, Z., Teo, A., Ooi, B.C.a.: Mining deterministic biclusters in gene expression data. In: Proceedings of the Fourth IEEE Symposium on Bioinformatics and Bioengineering (BIBE'04) (2004) pp. 283–292.