JOURNAL OF PURE AND APPLIED MICROBIOLOGY, Nov 2015. Vol. **9**(Spl. Edn. 2), p. 611-618

Maximize Pair Genes from Microarray using the Enhanced Fuzzy Clustering Algorithm

K. Vengatesan[1] and S. Selvarajan[2]

[1]Department of Computer Science and Engineering,
Muthayammal Engineering College, Tamilnadu, India.
[2]Muthayammal College of Engineering, Rasipuram-637408, Tamilnadu, India.

(Received: 10 August 2015; accepted: 03 September 2015)

The clustering is an important task in data mining, the functional analysis of gene clustering investigation can be perfomed by using various algorithms. The clustering techniques are useful to realize gene functions, cellular process, gene regulation and subtypes of cells. The different techniques are used to measure the performance of the gene overlapping such as CLICK, SOM, and rRFCM. The proposed ErRFCM increase the probability membership of the clusters and also handle the overlapping gene clusters effectively.It is also useful in dealing with probabilistic lower approximation and possibility lower approximation. The proposed methods are used to identify the strong group of Co expressed genes and produce the best result. The gene clusters produced are HCM, FCM, RFCM, SOM, CLICK and rRFCM algorithms, and visualized by Tree View software for 14 Microarray dataset.

Keywords: Gene Clustering ,Rough fuzzy clustering, Dunn Index, Silhouette Index.

In biological domain, the gene based on the special characteristics of genes,clustering has several new challenges, due to special characteristics of genes. The retrieval natural data structure of the gene and data distribution, carried by gene expression data by clustering. The clustering finds natural group present in the gene set[1]. It divides the genes into two category either same cluster or different cluster. The similar cellular function, can be cluster with similar functions of the gene, are called as co-expressed gene also understand the functionality of many genes. The co-expressed gene has strong correlation between the pair of genes[10]. Regularity motifs used to search a common DNA sequence in promoter regions of genes within the same cluster. The co expressed genes are cluster using the gene clustering techniques also consider the co functions and co regulation. The gene expression data are grouped using clustering algorithms such as, model based method, Graph theoretic methods, soft computing density based methods, hierarchical methods and partitioning methods[20]. The rough clusters are defined as similar to rough set using lower and upper boundary approximation, another important distinction of rough cluster is one cluster are overlapped with another cluster. The rough clustering allows grouping of related object based on the similarity relation equivalence[15].

The Fuzzy set clustering same like Fuzzy C-mean of that data object belong to multiple clusters according to degree of membership. Rough set based clustering proves to provide a solution that is less restrictive than the traditional clustering algorithms like k-means and less descriptive than the fuzzy clustering methods. The gene clustering searches the exclusive partitions of objects that are similar to each other, also called as biclustering. In which gene is either homogeneous or heterogeneous groups of objects and extracted based on the certain features[9].In

* To whom all correspondence should be addressed.
E-mail: vengicse2005@gmail.com

the biotechnological field monitoring the thousand of gene expression data is an important task can be done using microarray. The microarray is finding the hidden patterns from the functional genomics. A common representation of microarray is n*m matrix, from that n represents the number of gene and m represents the particular sample at the point. When the number of genes increases the gradually, the complexity of the biological networks will also increase. Clustering is the general technique used to group the related gene based on the genetic similarity[17]. The cluster group have gene with same cluster or different cluster based on the gene regulation, cellular processes and subtypes of cells. The Co expressed gene has similar appearance of cellular function and strong correlation between the genes. This approach also helpful for the functional analysis of many genes to performs various hypotheses mechanisms[22].The Co expressed gene are grouped together using gene clustering, the gene clustering has various challenges due to special characteristics of genetic data. In data mining and knowledge discovery ,the clustering is one of the most important step, there are various clustering algorithms used to form the group, such as hierarchical clustering, self organizing map, graph theoretical approached, K-mean algorithm, density based approach and model based clustering are applied to find the Co expressed group of gene from the micro array[23].

The clustering algorithm for gene clustering will handle the huge amount of data even background noise, the general study about gene expression data is highly connected and cluster is overlapped one with another, so handle the effective algorithm to filter the most relevant data from the data set. The uncertainty in the data can be captured using the rough set theory and fuzzy set theory that has mathematical framework[18]. The fuzzy c-mean algorithm uses the prototype based clustering algorithm to find the association between the data. It will assign the membership to each gene how it related to another gene, sometime a gene may be related to more than one cluster, because of gene are overlapped with one another, based on the certain boundaries the gene group is formed[19]. The fuzzy clustering methods used to identify the information regarding overlapped cluster and cellular pathways of each gene. If the membership value of the fuzzy c-means is low means, then there is a noisy environment, weakness in gene expression data. The possibilistic c-Mean algorithm produces the best coincidence clusters.The efficient method used in the pattern reorganization algorithm that follows the support vector machine concept to predict the functional sequence of proteins. The string kernel function is based on the conventional bio basis follows the fisher ratio and novel concepts for the degree of resemblance. It also uses the new quantitative indices to evaluate the quality of the identified bio basic sequence, it also demonstrates the different protein data set effectively[1].

The supervised attribute clustering is a method used to find the group in the gene expression data set. In which the similarity between the genes are calculated and also removes the redundancy among the genes. It's also suitable for various data set such as cancer and arthritis, which follows the classification algorithm such as K-nearest neighbor, naïve Bayes classifier and support vector machine that are effectively separate the clusters. The biological significance of the cluster is improved using gene ontology mechanism[2].The rough set theory, used to maximize the relevance based on selected features and improve the significance. It will select reduced feature also produce higher predictive accuracy. The performance of this method when compared to support vector machine, K-nearest neighbor and QSAR are increased for cancer arthritis microarray data sets[3].

The gene selection for real value data set is a highly challenging task, the novel feature selection method is based on the fuzzy rough set while they maximize the relevance and minimizing the repeated features. The shannons entropy based method is a fuzzy equivalence partition matrix for gene selection, which is suitable for real value data set also increased the performance of the gene cluster using f-information measures[4].Even though the large amount of genes are represented as microarray data, only a small portion of data is effectively handled and applied for various test to enhance the performance of the genetic similarity. Consider two genes that are completely independent of each other, the joint distribution function is used to calculate the relationship between the genes, and also various predictive accuracy measuring techniques such as support

vector machine, k-nearest neighbor rule and naïve Bayes classifiers are applied. The f-information measures are efficient for breast cancer data, leukemia data sets[5].

The rough fuzzy, c-mean is technically used to handle the rough set and fuzzy set, which effectively work on lower and upper approximation values and efficient to applicable for overlapped partitions. It also avoids the noise sensitivity problem occurred in fuzzy, c-mean algorithm[6].The second order fuzzy measure and weighted co-occurrence matrix are another method used to measure the gene co-occurrence based on the threshold value. The efficiency is more when compared to fuzzy entropy, fuzzy correction methods and local information are more accurately measured[7].

Integrating the merits of fuzzy sets and rough sets, different rough-fuzzy clustering algorithms such as the rough fuzzy, c-mean, rough fuzzy possibilistic c-mean and rough possibilistic c-mean in which each cluster is represented by a cluster prototype with possibilistic boundary and lower approximation[12]. Various clustering algorithms are used in microarray to calculate the Co expression of gene expression data sets, from that crisp lower approximation and fuzzy boundary are generally assumed the spherical in shape, which find the arbitrary shape of gene clustering. The strong associated Co expressed genes are calculated using the fuzzy rough supervised gene clustering algorithm[14]. The rough fuzzy, c-mean was derived from a rough fuzzy clustering algorithm, which is efficient to handle the micro array gene expression data even though of overlapped partition and noisy data. There are three basic parameters needed to form the clustes namely, possibility lower approximation, probability boundary and cluster prototype or centroid. The cluster centroid depends on the weighting average of the probabilistic boundary and possibility lower approximation. A main problem of an existing method is to detect the efficient method to find the prototype of the gene at initial stage. The effectiveness of the algorithm is compared with existing algorithm along to 14 microarray data set

Robust Rough fuzzy C-Mean Algorithm

It is an efficient method to handle the cluster in both possibilistic and probabilistic fuzzy sets, and upper and lower approximation of rough sets into C-Mean algorithm, while integrating both techniques, it will handle the overlapping cluster in noisy environments, also deal with vagueness, incompleteness uncertainty in cluster definition. The objective function is let $Y=\{y_1.....y_j.....y_n\}$ be a set of n objects and $C=\{c_1....c_i.....c_c\}$ be the set of centroid, where $y_j \varepsilon R^m$ and $v_j \varepsilon R^m$. Each of the clusters β_i is represented by a cluster center υ_i which follows both lower and upper approximation of the cluster. The minimization function of the proposed C cluster is written as

$$J=\begin{cases} wA_1+(1-w)\beta_1 & if\ A(\beta)\neq\theta, B(\beta_i)\neq\theta \\ A_1 & if\ A(\beta)\neq\theta, B(\beta_i)\neq\theta \\ B_1 & if\ A(\beta)\neq\theta, B(\beta_i)\neq\theta \end{cases} \quad ...(1)$$

In which $A(\beta)$ is the lower approximation and $B(\beta_i)$ is the probability boundary where, A_1, B_1 are represented as

$$A_1=\sum_{i=1}^{c}\sum_{x_j\varepsilon\ A(B_i)}(v_{ij})^{m_2}\left\|x_j-v_i\right\|^2+\sum_{i=1}^{c}\eta_i\sum_{x_j\varepsilon\ A(B_i)}(1-v)^{m_2}$$

$$B_1=\sum_{i=1}^{c}\sum_{x_j\varepsilon\ B(B_i)}(\mu_{ij})^{m_1}\left\|x_j-v_i\right\|^2$$

The relative important of lower boundary is represented by the parameter w and (1-w), while 1<=m_1<" and 1<=m_2<" are the probability functions. The centroids of the cluster should be independent to the lower approximation along with memberships of the objects. The membership function between the object is represented as following equation

$$\mu_{ij}=\left[\sum_{k=1}^{c}\left(\frac{\left\|x_j-v_i\right\|^2}{\left\|x_j-v_k\right\|^2}\right)^{\frac{1}{m_1-1}}\right]^{-1}$$

$$v_{ij}=\left[1+\left\{\frac{\left\|x_j-v_i\right\|^2}{\eta_i}\right\}^{\frac{1}{m_2-1}}\right]^{-1}$$

where the scale parameter

$$\eta_i=K.\frac{\sum_{j=1}^{n}(v_{ij})^{m_2}\left\|x_j-v_i\right\|^2}{\sum_{j=1}^{n}(v_{ij})^{m_2}}$$

which represents the size of the cluster . The

centroids of the cluster are calculated based on the weighting average of the probabilistic boundary and possibility lower approximation.

Table 1. The performance analysis of 14 microarray dataset with average threshold value.

Micro array Data set	Number of Clusters	Different validity index	Average Threshold value »
GDS608	26	Dunn Index	0.5
		DB Index	1.33
		Silhouette Index	0.18
GDS759	25	Dunn Index	0.54
		DB Index	1.05
		Silhouette Index	0.35
GDS1013	18	Dunn Index	2.3
		DB Index	0.48
		Silhouette Index	0.57
GDS1550	21	Dunn Index	1.54
		DB Index	0.33
		Silhouette Index	0.35
GDS1611	26	Dunn Index	0.95
		DB Index	0.89
		Silhouette Index	0.27
GDS2002	25	Dunn Index	0.86
		DB Index	0.96
		Silhouette Index	0.48
GDS2003	23	Dunn Index	0.64
		DB Index	1.06
		Silhouette Index	0.42
GDS2196	24	Dunn Index	0.7
		DB Index	0.16
		Silhouette Index	0.27
GDS2267	14	Dunn Index	3.6
		DB Index	0.64
		Silhouette Index	0.6
GDS2318	21	Dunn Index	0.78
		DB Index	0.67
		Silhouette Index	0.43
GDS2347	18	Dunn Index	1.15
		DB Index	0.622
		Silhouette Index	0.55
GDS2712	15	Dunn Index	0.29
		DB Index	0.59
		Silhouette Index	0.35
GDS2713	14	Dunn Index	0.34
		DB Index	0.65
		Silhouette Index	0.35
GDS2715	16	Dunn Index	0.28
		DB Index	0.48
		Silhouette Index	0.38

RESULTS AND DISCUSSION

In this work, the performance of the proposed algorithm enhanced robust rough fuzzy, c-mean algorithms are compared with the rough-fuzzy, c-mean, hard c-mean, cluster identification via connectivity kernel, self organizing map and robust rough fuzzy c-mean algorithm with several microarray gene expression data set. Assume there are six parameters to consider two analysis various algorithms such as Dunn index, Eisen plot, Davies-Bouldin index, Sihouette index, â index and execution time. The Gene ontology Term Finder is another method to analysis the biological significance of the generated gene clusters. The CLICK algorithm is used to decide the number of cluster C in the microarray dataset. By using the rough fuzzy clustering, calculate the weight parameter that is assumed as 0.99 and two falsifiers value are m1=2.0 and m2=2.0. The Gene expression data for this work is assumed publicly from available 14 yeast microarray time series data and the performance of the dataset is compared with a different algorithm. The Davies-Bouldin Index is one of the quantitative measures are relationships between the clusters, if the DB Index is low then good clustering are formed.

$$\text{DB}=\frac{1}{c}\sum_{i=1}^{c}\max_{i\neq k}\left\{\frac{S(v_i)+S(v_k)}{d(v_i,v_k)}\right\})$$

For 1<=i,k<=C, The DB index minimizes the within cluster distance $S(v_i)$ mutually maximizes the between cluster separation

The Dunn Index is used to identify the set of cluster that are well separated and compact. In which the good clustering produce the high Dunn index, for 1<=i,k<=C, can written as

$$D=\min_i\left\{\min_{i\neq k}\left\{\frac{d(v_i,v_k}{\max_l S(v_l)}\right\}\right\}$$

The â index is an another technique used to measure the relationship between the genes, it ratio of total variation and within cluster variation can be written as

$$\beta=\frac{N}{M}\text{ where }N=\sum_{i=1}^{c}\sum_{j=1}^{n}\left\|x_{ij}-\bar{v}\right\|^2 \quad ;$$

$$M = \sum_{i=1}^{c}\sum_{j=1}^{n}\left\|x_{ij} - \bar{v}\right\|^2 \text{ and } \sum_{i=1}^{c} n_i = n;$$

where n_i is the number of objects in the ith cluster (i=1,2,3 …,c) and n si the total number of objects , x_{ij} is the jth object in the cluster i, v_i is the mean or centroid of the cluster i, and is the mean of the n objects. The initial cluster is selected from different gene cluster based on the optimum value of the threshold ë also control the redundancy among the initial prototypes. It will direct influence to the performance of c-mean algorithm. The Silhouette index used to find the optimal threshold value. Consider the gene g_i• $\hat{a}_r$, i=1,2….n_r is the cardinality of the cluster $\hat{a}_r$. For each gene g_i, let x_i be the average distance between gene g_i, and rest of the gene $\hat{a}_r$, that can be written as

$$x_i = d_{avg}(g_i, \beta_r - \{g_i\})$$

Where average distance measure between a gene is denoted by d_{avg}. The Silhouette width of gene g_i is defined as

$$S(g_i) = \frac{b_i - a_i}{\max(b_i, a_i)}$$

where -1<=s(g_i)<=1. When the distance of the S(g_i) nearby 1 means the distance of gene g_i and $\hat{a}_r$ significantly well formed cluster. On the other hand, the value of S (GI) is close to -1 the distance between the cluster are increasing. Finally, when the value of S (GI) is 0 then both are same border between the two clusters.

The clusters are produced by different algorithms and evaluated by functional analysis

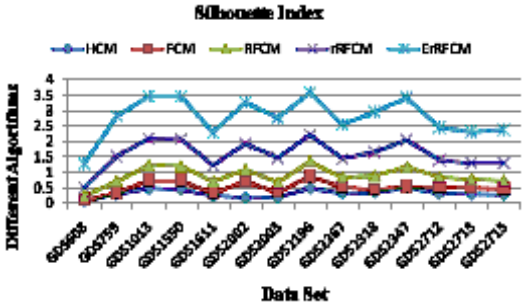

Fig. 1. Performance Analysis of Silhouette Index with Yeast Microarray Data Sets

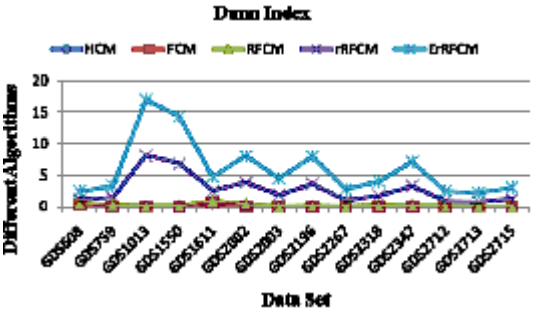

Fig. 2. Performance Analysis of Dunn Index with Yeast Microarray Data Sets

Table 2. Performance Analysis of Different Algorithms

Micro Array Data Sets	Silhouette Index		DB Index		Dunn Index		β Index	
	rRFCM	ErRFCM	rRFCM	ErRFCM	rRFCM	ErRFCM	rRFCM	ErRFCM
GDS608	0.27	0.52	0.92	1.17	0.55	0.8	4.68	4.93
GDS759	0.81	1.06	0.22	0.47	1.11	1.36	4.03	4.28
GDS1013	0.87	1.12	0.11	0.36	8.05	8.3	41.4	41.65
GDS1550	0.88	1.13	0.08	0.33	6.69	6.94	56.34	56.59
GDS1611	0.54	0.79	0.42	0.67	1.51	1.76	23.01	23.26
GDS2002	0.85	1.1	0.19	0.44	3.42	3.67	3.43	3.68
GDS2003	0.8	1.05	0.21	0.46	1.8	2.05	3.25	3.5
GDS2196	0.87	1.12	0.08	0.33	3.53	3.78	102.91	103.16
GDS2267	0.61	0.86	0.32	0.57	0.97	1.22	39.51	39.76
GDS2318	0.79	1.04	0.15	0.4	1.53	1.78	3.42	3.67
GDS2347	0.88	1.13	0.09	0.34	3.1	3.35	3.31	3.56
GDS2712	0.56	0.81	0.33	0.58	0.81	1.06	34.78	35.03
GDS2713	0.52	0.77	0.35	0.6	0.67	0.92	34.26	34.51
GDS2715	0.58	0.83	0.3	0.55	1.13	1.38	32.14	32.39

of gene ontology. The term annotation ratio is used to find wheather ,genes are related to each other or not, if the annotation ratio value is high, the genes are more related to each others, otherwise the genes are irrelevant or noisier. The sum of the entire annotation ratio is considered as the final annotation ratio. The comparative performance analysis of different algorithms performed with three common terms such as Molecular Function, Biological Process and Cellular Components. The significant clusters p-value is always considered as less than 0.05.

The Gene Ontology Term Finder is used to detect the biological significance of gene clusters. The GO is based on the three processes such as biological process, cellular components and molecular function that generate the tree structure of gene clusters. The hypergeometirc distribution used to calculate the Bonferroni multiple hypothesis correction between the genes, denoted by

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

Where ,the total number of gene sin the background distribution is denoted by N, The number of genes within distribution list can be either indirectly or directly represented as M, The size of the gene list is n, The amount of the list within gene are annotated is denoted by k. Whenever the p value is closer to zero that gene

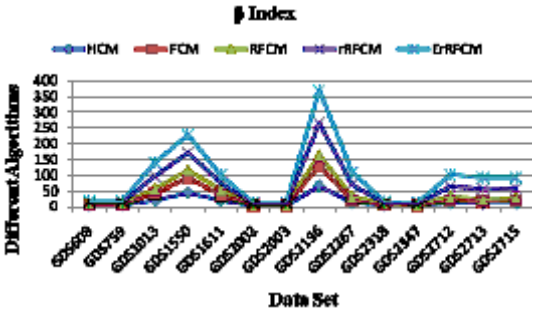

Fig. 3. Performance Analysis of β Index with Yeast Microarray Data Sets

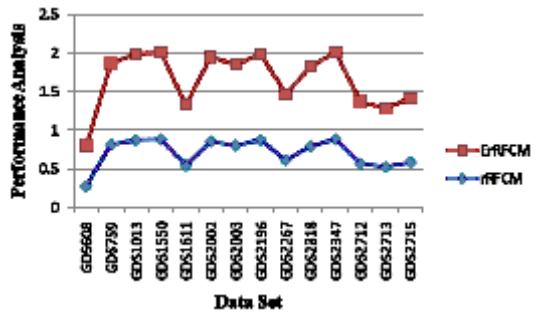

Fig. 4. Performance comparison of Silhouette Index with various algorithms

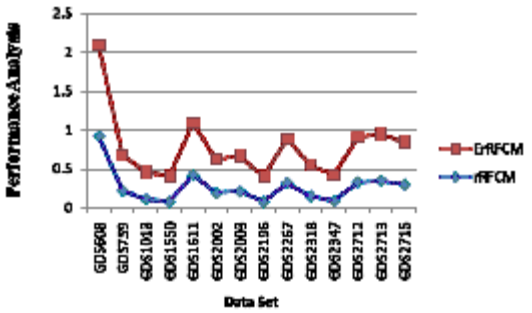

Fig. 5. Performance comparison of DB index various algorithms

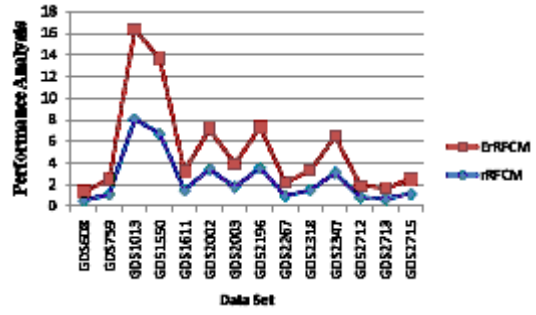

Fig. 6. Performance comparison of Dunn Index with various algorithms

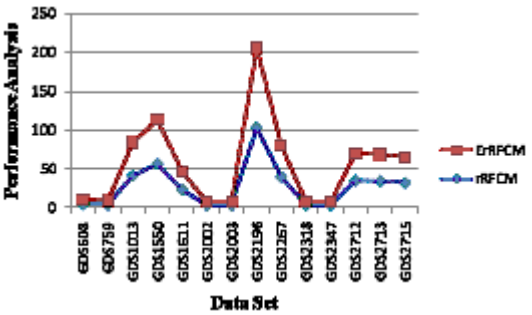

Fig. 7. Performance comparison of â Index with various algorithms

are more associated with each others. The error in the multiple hypothesis tests is measure using false discovery rate technique which is mainly used in microarray gene expression data.

Biologically significant gene clusters

In this section the performance analysis of different gene clustering algorithms are discussed , Fig.1 presents the result for the HCM, FCM, RFCM and ErRFCM ontology's on 14 yeast microarray data sets for Silhouette index. The statistically significant gene cluster is produced by GO finder. Similarly the fig. 2 shows the performance analysis for a Dunn index of microarray dataset. The â Index for micro array is represented using the Fig. 3.

The biological interpretation of gene clusters is generated using ErRFCM algorithm the Fig. 4 represents the Silhouette Index of microarray dataset and produced better result, compared to rRFCM algorithm. Similarly the Fig. 5 shows the DB Index with Yeast Microarray Data Sets that also produce the best result. The Fig. 6 represents the Performance Analysis of Dunn index finally the Fig.7 shows the **β Index** with Yeast Microarray Data Sets

CONCLUSION

In this paper, we proposed a new gene clustering algorithm, which is a combination of various algorithms such as rough set, c-mean algorithm and possibility, probabilistic memberships of fuzzy sets, which produced the maximum result for rough and fuzzy sets, the effects of our algorithm is compared to other algorithms, to 14 yeast microarray gene expression data set using gene ontology and standard cluster validity indices. The Enhanced Fuzzy Clustering is applied to gene expression data that produce significantly better result, irrespective of the quantitative indices and microarray data sets using a biological process, molecular function and cellular components for lower approximations. The result reported in the table, shows that proposed gene clustering algorithm is better than other cluster algorithms. It extracts the highly similar genes from lower approximation , upper approximation, any shaped gene clusters and handles efficiently overlapped gene cluster

REFERENCES

1. P. Maji and C. Das, Protein Functional Sites Prediction Using Modified Bio-Basis Function and Quantitative Indices, *IEEE Transactions on NanoBioscience*, 2010; **9**(4): 250—257.
2. P. Maji and C. Das, Relevant and Significant Supervised Gene Clusters for Microarray Cancer Classification, *IEEE Transactions on NanoBioscience*, 2012; **11**(2): 161—168.
3. P. Maji and S. Paul, Rough Sets for Selection of Molecular Descriptors to Predict Biological Activity of Molecules, *IEEE Transactions on System, Man and Cybernetics, Part C, Applications and Reviews*, 2010; **40**(6): 639—648.
4. P. Maji and S. K. Pal, Feature Selection Using *f*-Information Measures in Fuzzy Approximation Spaces, *IEEE Transactions on Knowledge and Data Engineering*, 2010; **22**(6): 854—867.
5. P. Maji, *f*-Information Measures for Efficient Selection of Discriminative Genes from Microarray Data, *IEEE Transactions on Biomedical Engineering*, 2009; **56**(4): 1063-1069.
6. P. Maji and S. K. Pal, Rough Set Based Generalized Fuzzy *C*-Means Algorithm and Quantitative Indices, *IEEE Transactions on System, Man and Cybernetics, Part B, Cybernetics*, 2007; **37**(6): 1529-1540.
7. P. Maji, M. K. Kundu, and B. Chanda, Second Order Fuzzy Measure and Weighted Co-Occurrence Matrix for Segmentation of Brain MR Images, *Fundamenta Informaticae*, **88**(1-2): 161-176.
8. Pradipta Maji, and Sushmita paul, Rough-fuzzy clustering for grouping functionality similar gene from microarray data, IEEE/ACM Transctions on computational biology and bioinformatics, 2013; **10**; 286-299.
9. H.Causton,J.Quackenbush, and Brazama, Microarray Gene Expression Data Analysis, A Beginners guide, Wiley-Blackwell,2003.
10. E. Domany, Cluster Analysis of gene Expression data", *J. Statistical Physics*, 2003; **110**: nos.3-6: 1117-1139.
11. D.Jiang, C.Tang, and A.Zhang, "Cluster Analysis of gene Expression Data: A survey,*IEEE Trans. Kwnowledge anda data Eng*, 2004; **16**(11): 1370-1386.
12. M.B.Eisen, P.T.Spellman,O.Patrick,and D.Botstein,"Cluster Analysis and Display of Genome-Wide Expression Patterns,"*Proc. Nat'l Academy of science USA* 1998; **95**(25): 14-863-14-868.

13. A.Brazma and J.Vilo,"Minireview:Gene Expression Data Analysis", *Federation of European Biochemical Societies letters,* 2000; (1): 17-24
14. L.Heyer, S.Kruglyak,and S.Yooseph,"Exploring Expression Data:Identification and analysis of Coexpressed genes", *Genome Research,* 1999;**11**: 1106-1115.
15. P.Tamayo, D.Slonim,J.Mesirov, Q.Zhu, S. Kitareewan, E.Dmitrovsky,E.S.Lander,and T.R.Golub, "Interpreting Patterns of Gene Expression with Self Organizing Maps: Methods and Application to Hematopoietic Differentiation," *Proc Nat'l Academy of science USA,* **9**(6),pp.2007-2912.
16. E.Hartuv and R.Shamir, "A Clustering Algorithm based on Graph Expression Patterns", *J. Computational Biology,* 1999; **6:** 281-297.
17. R.Sharmir, and R.Sharan,"Click :A Clustering Algorithm for Gene Expression Analysis", *Proc. Eight Int'lConf. Intelligent System for Molecular Biology,* 2000.
18. D.Ghosh and A.M.Chinnaiyan,"Mixture Modelling of gene Expression Data from Microarray Experiments", Bioinformatics, 2002; **18**(2): 275-286.
19. D.Jiang,J.Pei and A.Zhang, "DHC:A Density Based Hierarchical Clustering Methods for Time Series Gene Expression Data," *Proc,IEEE Third Int'l Symp, Bioinformatics and BioEng,* 2003; 393-400.
20. J.C.Bezdek,Pattern Recognition with fuzzy objective function Algorithm,Plnum,1981.
21. P.J.Woolf and Y.Wang,"Fuzzy logic approach to Analyzing Gene Expression Data", *Physiological Genomics,* 2000; **3**: 9-15.
22. D.Dembele and P.Kastner,Fuzzy C-Mean Methods for Clustering Microarray Data," *Bioinformatics,* 2003; **19**(8): 973-980.
23. A.P.Gasch and M.B.Eisen,"Exploring the Conditional Coregulation of yeast Gene Expression through Fuzzy K-Means Clustering", *Genome Biologym,* 2002; **3**(11): 1-22.
24. R.Krishnapuram and J.M.Keller,"A Possibilistic Approach to Clustering ", *IEEE Trans, Fuzzy System,* 1993; **1**(2): 98-110.
25. N.Belacel,M.Cuperlovic-Culf,M.Laflamme,and R.Quellette,"Fuzzy J-Means and VNS methods for clustering Gene from Microarray Data," *Bioinformatics,* 2004; **20**(11): 1670-1701.