International Journal of Computational Intelligence and Informatics, Vol. 2: No. 1, April - June 2012

Biclustering Analysis of Coregulated Biclusters from Gene Expression Data

C. P. Chandran
Associate Professor Departement of Computer Science and Head, Post Graduate Departement of Bioinformatics,
Ayya Nadar Janaki Ammal College, Tamilnadu, India
drcpchandran@gmail.com

K. IswaryaLakshmi
Full-Time Research Scholar(M.Phil),
Post Graduate Departement of Computer Science and Information Technology,
Ayya Nadar Janaki Ammal College, Tamilnadu , India
isu.lakshmi@gmail.com

***Abstract*- In this paper, the Biclustering analysis of coregulated biclusters from gene expression data is carried out. Gene expression is the process, which produces functional product from the gene information. Data mining is used to find relevant and useful information from databases. Clustering groups the genes according to the given conditions. Biclustering algorithms belong to a distinct class of clustering algorithms that perform simultaneous clustering of both rows and columns of the gene expression matrix. In this paper a new algorithm, Enhanced Bimax algorithm is proposed. The normalization technique is included which is used to display a coregulated biclusters from gene expression data and grouping the genes in the particular order [1]. In this work, Synthetic Gene Expression dataset is used to display the coregulated genes, developed by Prelic *et.al.*, It contains constant values and coherent values over the conditions and non-overlapping and overlapping clusters. The data matrix contains 10 overlapping cluster and each cluster extends over 5 genes and 15 conditions.**

Keywords- Data mining, biclustering, enhanced bimax algorithm, coregulated biclusters, gene expression data.

I. INTRODUCTION

A. Data Mining

Data mining is the extraction of hidden predictive information from large databases. Data mining tools predict future trends and behaviors, helps organizations to make proactive knowledge-driven decisions [2].

Clustering and Biclustering

A clustering is essentially a set of clusters, usually containing all objects in the dataset [3]. Clustering is considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data [4]. The clustering is defined as the process of organizing objects into groups whose members are similar in some way. A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. A bicluster is defined as a set of genes whose expression profiles are mutually similar within a subset of experimental conditions/samples [5]. The goal of biclustering is to identify "homogeneous" submatrices. Given a gene expression data matrix $D=G\times C= \{d_{ij}\}$ (here $i \in [1, n]$, $j \in [1, m]$) is a real-valued $n\times m$ matrix, G is a set of n genes $\{g_1, g_2\ldots, g_n\}$, C a set of m biological conditions $\{c_1,c_2\ldots, c_n\}$. Entry d_{ij} means the expression level of gene g_i under condition c_j [6]. If there is a submatrix $B=g\times c$, where $g \in G$, $c \in C$, to satisfy certain homogeneity and minimal size of the cluster, we say that B is a bicluster. Biclustering [7] is a data mining technique which allows simultaneous clustering of the rows and columns of a matrix. Given a set of m rows in n columns, the biclustering algorithm generates the subset of rows which exhibit similar behavior. Biclustering algorithm generates a subset of rows which exhibit similar behavior across a subset of columns. Bicluster is a subset of genes that an exhibit similar expression patterns over a subset of conditions [8]. Here a bicluster as a subset of genes those jointly respond across a subset of conditions, where a gene is termed responding in some condition if its expression level changes significantly at that condition. It finds clusters of samples possessing similar characteristics together with features creating these similarities.

B. Genomics

Genomics aims to understand the structure of the genome, including the mapping genes and sequencing the DNA. Genomics examines the molecular mechanisms and the interplay of genetic and environmental factors in disease. The Genomics consists of 5 types of genomics [9]. They are Functional Genomics, Structural Genomics, Comparative Genomics, and Epigenomics. The Functional Genomics used to characterization of genes and their mRNA and protein products. Structural Genomics are used to dissection of the architectural features of genes and chromosomes. Comparative Genomics are used to evolutionary relationships between the genes and proteins of different species. Epigenomics (epigenetics) are used to DNA methylation patterns, imprinting and DNA packaging.

Gene Expression Data(GED)

Gene expression is the process by which information from gene is used in the synthesis of a functional gene product [9]. These products are often proteins, but in non-protein coding genes such as rRNA genes or tRNA genes, the product is a functional RNA. The process of gene expression is used by all known life - eukaryotes including multi cellular organisms), prokaryotes (bacteria and achaea) and viruses - to generate the macromolecular machinery for life. Several steps in the gene expression process may be modulated, including the transcription, RNA splicing, translation, and post-translational modification of a protein [10]. Gene regulation gives the cell control over structure and function, and is the basis for cellular differentiation, morphogenesis and the versatility and adaptability of any organism. Gene regulation may also serve as a substrate for evolutionary change, since control of the timing, location, and amount of gene expression can have a profound effect on the functions (actions) of the gene in a cell or in a multi-cellular organism. A gene is a unit of heredity in a living organism. The Fig. 1 represents the Gene expression which is the process by which the information encoded in a gene is converted into protein or some form of RNA [11]. The process of gene expression occurs in both prokaryotes and eukaryotes. Transcription is a process in which the DNA is converted into RNA. The production of RNA copies of the DNA is called transcription. The transcription process is carried out by the enzyme RNA polymerase. In the prokaryotes, the transcription is carried out by a single type of RNA polymerase. In eukaryotes, three types of RNA polymerases such as RNA polymerase I, RNA polymerase II and RNA polymerase III are involved.

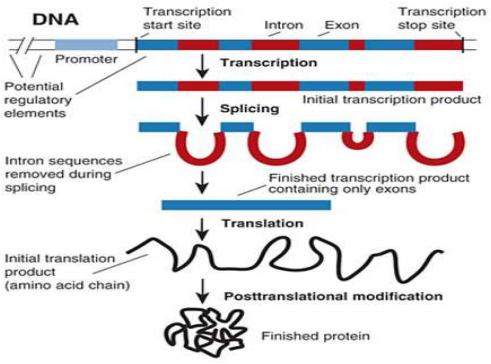

Figure 1. Basic process of Gene Expression

RNA splicing is a very important modification of eukaryotic pre m-RNA. The eukaryotic pre m-RNA consists of alternating segments called exons and introns. RNA splicing is a process in which introns are removed and then exons are joined together. It normally resides on a stretch of DNA that codes for a type of protein or for an RNA chain that has a function in the organism. All living things depend on genes, as they specify all proteins and functional RNA chains. Genes hold the information to build and maintain an organism's cells and pass genetic traits to offspring, although some organelles (e.g. mitochondria) are self-replicating and are not coded for by the organism's DNA. A modern working definition of a gene is "a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions, and or other functional sequence regions". A gene expression data (GED) is described as a table with 'n' rows corresponding to genes and 'm' columns corresponding to situations [12]. Gene expression data is obtained by extraction of quantitative information from the images or patterns resulting from the readout of fluorescent or radioactive hybridizations in a microarray chip.

A GED from microarray experiment is represented by a real valued matrix. $M = \{ A_{ij} | 1 \leq i \leq n, 1 \leq j \leq m \}$ where rows $G = \{ g_1, g_2, g_3, \ldots g_r \}$ represents the expression pattern of the genes and the column $S = \{ s_1, s_2,$

$s_3 \ldots s_c\}$ represents expression profiles for samples and each element w_{ij} is measured expression level, Where**,** r = no of genes, c = no of samples, M = gene expression data matrix, a_{ij} = element in the gene expression matrix [13]. GED has also specified only the seven samples and the given condition. The first samples are Anthocyaninless and describes that the Stems and leaves always lack anthocyanin and then specify the condition on 4.5. Second samples named as acroxantha described as smaller plant and then specify the condition on 5.7. Third samples named as acumbens-2 described rounded cotyledons and the last samples are named as apocarpous described as highly deformed multicarpellate and apocarpous fruits and then specify the condition on 3.3.

Coregulated bicluster

Coregulated biclusters are also often functionally associated, and such *a* priori known or pre-computed associations can provide support for appropriately grouping genes [14]. It is a measure of that describes a particular type of association between two genes. It is also the basis of several clustering methods. Most commonly used correlation statistic is Pearson Correlation Coefficient (PCC) which includes the measure to identify the coregulated bicluster, counts the number of times each gene in the same cluster. First, find the transcription sites in the coregulated genes specify the sequence and then locate the sites. Binding sites are often associated with specialized proteins known as transcription factors, and are thus linked to transcriptional regulation. It is used to analyze the co-regulation between pairs of genes [15]. Coregulated genes and their transcription factor binding sites are key steps towards understanding transcription regulation [16]. Coregulated genes should have at least one common sequence element [17] it is generally difficult to identify these genes from the presence of this element because it is very easily obscured by noise. To identify groups of coregulated genes are affected by different conditions and they are collectively affect an organism. Regulatory functions of different gene groups in specific situations. Coregulated biclusters from gene expression data is shown in Fig. 2.

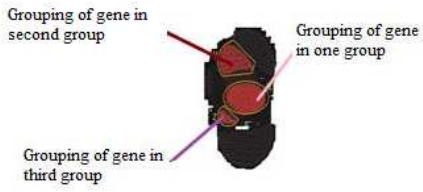

Figure 2. Coregulated Bicluster

To measure the relationship between one genes to another the KEGG Pathway is used. Analysis of transcription factor binding sites understand the mechanism of regulation, including coordinate regulation by multiple transcription factors acting together, and effectively identify and characterize mutations that disrupt the regulatory mechanism. Transcription factor is a protein that binds to specific DNA sequences, thereby controlling the flow of genetic information from DNA to mRNA. Transcription factors perform this function alone or with other proteins in a complex, by promoting, or blocking the recruitment of RNA polymerase to specific genes.

II. PREVIOUS WORKS

Mingoo Kim [18] has developed cMonkey, which detects computative co-regulated gene groupings by integrating the biclustering of gene expression data and various functional associations with the de novo detection of sequence motifs. Co regulated groups of genes for use in subsequent regulatory network inference procedures. Xin .Xu [19] said that reg-cluster model, to identify all the significant shifting-and-scaling co-regulation patterns. Reg-cluster algorithm is able to detect a significant amount of clusters are potentially of high biological significance. Dharan and Achuthsankar have demonstrated [20] that GRASP approach starts from small high quality bicluster seeds, which are tightly coregulated submatrices of the gene expression matrix. These seeds are further enlarged by adding more rows and columns to them. Bhattacharya and Rajat [21] have proposed a technique on Bi-Correlation Clustering Algorithm (BCCA) produces an diverse set of biclusters of co-regulated genes over a subset of samples where all the genes in a bicluster have a similar change of expression pattern over the subset of samples. Zhang. Zha [34] have proposed the Time-Series Biclustering algorithm (CC-TSB), restricting it to add and/or remove only columns that are contiguous to the partially constructed biclusters, thus forcing the resulting biclusters to have only contiguous columns. A predicted number of biclusters with contiguous columns is identified. The CC

TSB algorithm contains two major steps, an iterative deletion procedure and an iterative insertion procedure. Remondini [23] is developed a coregulated based model used to clear relationship between the network structures. Nurual Haque [24] is developed a robust coregulated method used to minimum the β-divergence based on robust correlation matrix. Tim Van den Bulcke [25] described that a network generator produces synthetic Transcriptional Regulatory Networks (TRNs) and corresponding microarray datasets. In these networks, the nodes represent the genes and the edges correspond to the regulatory interactions at transcriptional level between the genes. Cheng [26] said that BiVisu is an open-source software tool for detecting and visualizing biclusters embedded in a gene expression matrix and display the co -regulated genes. Bivisu clustering methods in which partition data based on whole set of genes or conditions, biclustering groups a subset of genes (rows) over a subset of conditions (columns). Genes which are co-regulated under certain biological processes can be identified. Peter Waltman[27] said that multiple-species datasets in order to identify conserved modules and the conditions under which these modules are active. The advantages of these method are that conserved modules more likely to be biologically significant than co-regulated gene groups lacking detectable conservation, and the identification of these conserved modules can provide a basis for investigating the evolution of gene regulatory networks.

III. METHODOLOGY USED

The methodology used in this work is shown in Fig.3. The input is gene sequence. Enhanced Bimax algorithm is used to display a maximal biclusters value and displays a coregulated biclusters. The Enhanced Bimax algorithm is used to measure a particular gene is present or not. It also finds the transcription sites of the coregulated biclusters. Normalization technique used to specify genes are presented in the particular group or not. The output is display the transcription factors.

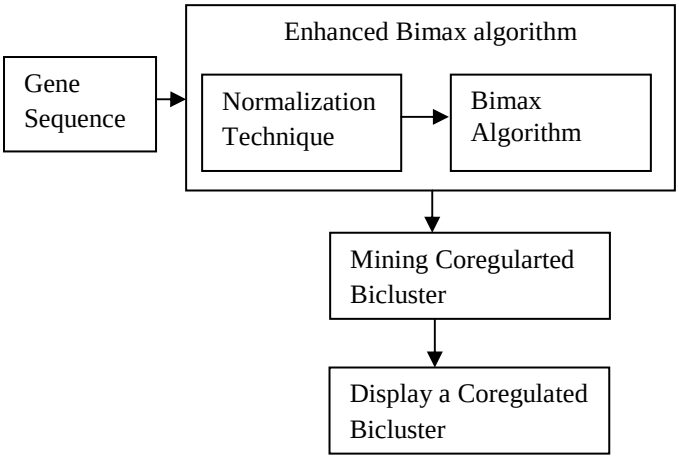

Figure 3. Block diagram of the methodology used for mining coregulated bicluster

IV . ALGORITHM USED

A. Bimax Algorithm

The Bimax algorithm needs to guarantee that only optimal, inclusion-maximal biclusters are generated [28]. The problem arises because V contains parts of the biclusters found in U, and as a consequence we need to ensure that the algorithm only considers those biclusters in V that extend over CV . The parameter Z serves this goal. It contains sets of columns that restrict the number of admissible biclusters. It is used to specify the genes and conditions. It is used to specify that analysis of DNA chips and gene networks. Bimax Biclustering based on the framework by Prelic *et. al.,* The algorithm realizes the divide-and-conquer strategy. Fig. 4 describes an original Bimax algorithm. It consists of three procedures. They are Enhanced Bimax, Conquer and Divide. Conquer function is call and check the condition is if the genes and conditions are equal then the partitioning is begin, otherwise it stop the process. Second step is split the data and normalization technique is used to group the splited data. It is used to find all add the maximum groups in general gene expression data. Each co regulated genes are grouping together the particular expression value and the particular situation.

```
Bimax ()
Begin
        Z ≠0
        S = divide (E, R,C, M)
        Return S
End
Conquer(E, (R, C), M)
Begin
    if R_i ϵ G, e_ij = 0 then
   return {(R,C)}
End if
  (G_U,R_U,C_V ) = divide (F, (G,C),Z)
M_U =0 M_V = 0
if G_U = 0 then
  M_U ← conquer (G_U ∪ G_W)
End if
End
Divide (F, (R, C), Z)
Begin
   G'=calculate (F,(R,C),Z)
If such an I ∪ G' exists then
  C_U = C_U +1;
End if
End
```

Figure.4 Original Bimax algorithm

B. Proposed Enhanced Bimax Algorithm

Enhanced Bimax algorithm can contain two procedures. Fig. 5 describes a flowchart for proposed Enhanced Bimax algorithm.

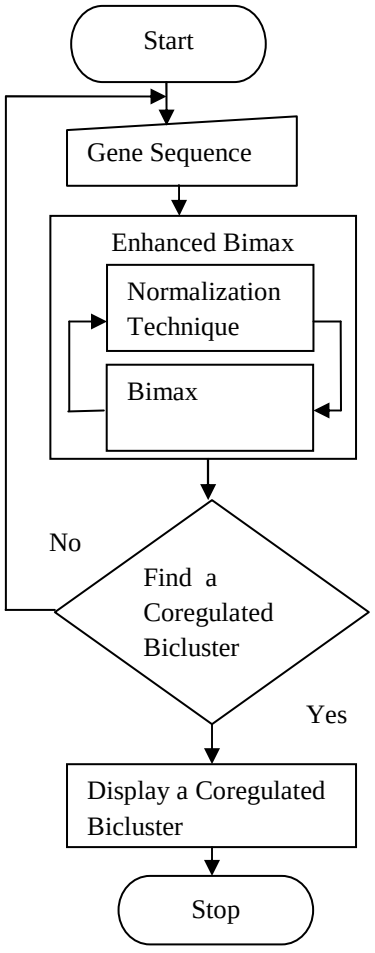

Figure 5. Flow Chart for Proposed Enhanced Bimax Algorithm

They are BFS and BPS. BFS uninformed search method that expands and examine all nodes of a graph or combination of sequences by systematically searching through every solution. In other words, it exhaustively searches the entire graph or sequence without considering the goal until it finds it. Binary Space Partitioning (BSP) is a method for recursively subdividing a space into convex sets by hyperplanes. This subdivision gives rise to a representation of the scene by means of a tree data structure known as a BSP tree. Normalization is the process of isolating statistical error in repeated measured data. Quintile normalization, for instance, is normalization based on the magnitude of the measures [29]. The goals in doing eliminate all the redundant data and ensure data dependencies. The numbers of genes that reproducibly showed and the unnormalized data and normalized data are displayed on the coregulated biclusters. Enhanced Bimax algorithm is applied data mining technique on clustering. In the clustering similar samples and similar gene probes are organized in a fashion so that they would lie close together. It consists of three procedures. They are Enhanced Bimax, BFS and BSP. First step is normalization technique used to remove the redundant data and then grouping genes in the specific conditions. Binary Space Partitioning function is call and check the condition is if the genes and conditions are equal then the partitioning is begin, otherwise it stop the process. It specifies that a particular gene is present in the given group then it is represents a one. With these maximum groups in general gene expression data can be found. Each co regulated genes are grouping together the particular expression value and the particular situation. Otherwise the gene is not present in the given group then it is representing as zero. Fig. 6 describes a proposed Enhanced bimax algorithm.

C. Dataset Used

In this work synthetic dataset used for GED. The creation of synthetic data is an involved process of data anonymization; that is to say that synthetic data is a subset of anonymized data [30]. Specify the given conditions and the genes are grouped together. In the partitioning strategy works on split the genes and then specify the situations then the given genes are grouped together until condition false. Specify the BFS technique works on the genes are grouped.

Table 1: Gene Expression Data [28]

Gene Name	Conditions		
	S_1	S_2	S_3
Mup1	4.5	5.7	2.3
Mup2	3.3	5.0	7.5
Serpina1d	4.5	8.5	5.0
Serpina3k	4.2	10.3	3.4
Mup3	5.0	8.2	5.0

Enhanced Bimax algorithm is applied to specify the condition on the 2.3 to 10.3. For example, in Table 1, the expression value g_1 has the conditions s_1 means given value is 4.5. The expression value g_2 has the situation s_2 means given value is 5.0. The expression value g_3 has the situation s_3 means given value is 5.0. The expression value g_4 has the situation s_1 means given value is 4.2. The expression value g_5 has the situation s_2 means given value is 8.2. Coregulated biclusters are presented in the particular group of genes (g_1 ... g_5), then it is represented as one or Otherwise zero. They are 'N' number of conditions are specified on s_1, s_2...s_n. Gene sequences are Mup1, Mup2, Serpina1d, and Serpina3k. It is validated on bench mark algorithm. Then algorithm steps are reduced. Synthetic gene expression dataset used to specify the gene sequence and condition. Each data matrix contains overlapping and non-overlapping structures. They are 10 non-overlapping clusters and overlapping clusters extends over a 5 genes 15 conditions. It can contains rows/columns clusters, and numbers of parameters. Synthetic gene expression dataset is freely available on [http://kdd.ics.uci.edu/]

V. RESULTS AND DISCUSSION

The Bimax algorithm includes the incremental algorithm to evaluate space complexity which leads to worst case. To overcome this drawback, Enhanced bimax algorithm is proposed. Space complexity required for the incremental algorithm is reduced which leads to simple and easy evaluation of GED. We maintain the set z is a lexicographic order according to the sets C. For each rows the algorithms performs the following: C is a column for gene expression data and U is a submatrices. Calculates the C* in O(n) time. Iterates through O(β) biclusters.

Suppose the function conquer calls in the tree that only has one child to which the submatrix U is passed. Then U has at least one row and one column. Row and column are to be partitioned parent and child is performed. Row U contains both 0s and 1s. The partitioning of U produces a non-empty set G_w and therefore submatrix resulting contains one's. Means the following conquer is a leaf node in the recursion tree. At least one half of all inner nodes have an out degree is greater than 1. Consider a tree where all inner nodes have an out degree of 2 and number of leaves equals β. Then total number of inner node is 2β. Recursion tree of a child node and conquer of overall number of nodes then conquer contains O(β). The dataset used in this work is synthetic dataset. The comparison of Enhanced bimax algorithms and Bimax algorithms is shown in Table 2.

In our Enhanced Bimax algorithm Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway is finds the coregulated genes in the particular groups and it occurs in the transcription binding sites. The results show that Enhanced bimax is comparatively faster than bimax algorithm. In the smaller and larger size used to display the pathway from the MPM and KEGG. MPM is compared with KEGG it is a smaller size. One conquer call needs O(nm min (n, m). maximum O(β) invocations conquer of the enhanced bimax. We show that the space complexity is O (n^2 log β), enhanced bimax number of conquer invocations is O(β) and the conquer call in the BFS of recursion tree is O(n^2). Z size less n*n total space complexity is O(n^2 log β). The quality of the bicluster is determined by the 'hscore' or Mean Squared Residue (MSR) value and to find maximal biclusters with low mean squared residue. Given the value of δ (δ>0), used to assess the quality of bicluster. Volume of the bicluster is determined by the number of elements e_{ij} such that i ∈ I & j ∈ J.

Table 2: Comparative analysis of GED using Coregulated Biclusters

Methods	Pathway Find			
	Shortest distance		Disconnected gene pairs	
	MPM	KEGG	CD3	DREB2A
	←	In seconds	→	
OPSMs	59.3	15.0	15.8	45.6
SAMBA	56.8	16.7	18.4	54.6
LEB	65.7	18.2	15.1	40
FABIA	45.6	13.4	13.4	34.5
Bimax	58.9	14.0	19.5	64.0
Enhanced Bimax	60.0	12.0	50.5	20.0

MPM: - Metabolic Pathway Map (MPM)
KEGG: - Kyoto Encyclopedia of Genes and Genomes (KEGG)
CD3:- T cell receptor/CD3 distinct intracellular signaling Pathway
DREB2A:- Dehydration-Responsive Element Binding Protein 2A

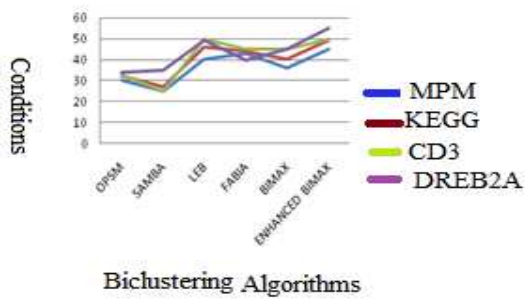

Figure 7. Comparsion of Enhanced Bimax algorithm with other algorithms

Fig. 7 describes comparison of Enhanced Bimax algorithm and Bimax algorithm. X axis specify the biclustering methods on OPSM, SAMBA and so on. Y axis specifies the biclustering conditions on 4.5 to 10.3. The results show that Enhanced bimax is comparatively faster than bimax algorithm. In the smaller and larger size used to display the pathway from the MPM and KEGG. MPM is compared with KEGG it is a smaller size. The inputs are GED and processing the data using enhanced bimax and display the coregulated genes. Fig.8 describes finding coregulated biclusters from GED. First specify the gene names and gene annotation used to specify the gene, nucleic acids and proteomics. Genomes used to specify the human and mouse and rat. Table. 2 describe comparison of various biclustering methods. CD3 is a homodimeric glycoprotein expressed on the surface of a major subset of human T cells that has been identified as a member of the immunoglobulin supergene family.

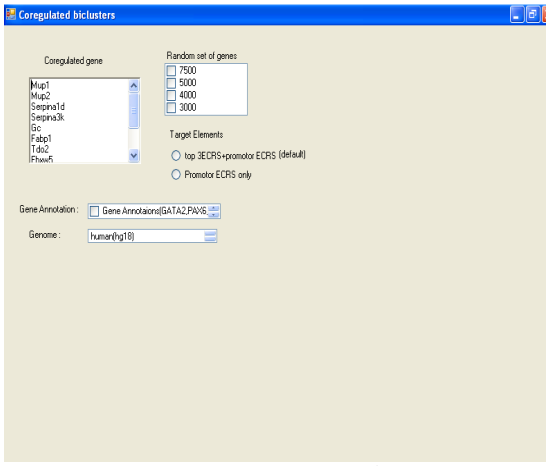

Figure 8. Coregulated Biclusters from GED

DREB1A and DREB2A proteins specifically bound to the DRE sequence in vitro and activated the transcription of the b-glucuronidase reporter gene driven by the DRE sequence in Arabidopsis leaf protoplasts. Expression of the DREB1A gene and its two homologs was induced by low-temperature stress, whereas expression of the DREB2A gene and its single homolog was induced by dehydration. Over expression of the DREB1A cDNA in transgenic Arabidopsis plants not only induced strong expression of the target genes under unstressed conditions but also caused dwarfed phenotypes in the transgenic plants. Comparative analysis of GED using several biclustering methods applied on OPSM, SAMBA, LEB, FABIA, Bimax and Enhanced Bimax. And then four pathways are used to find the coregulated biclusters from gene expression data, KEGG pathway gets optimal results.

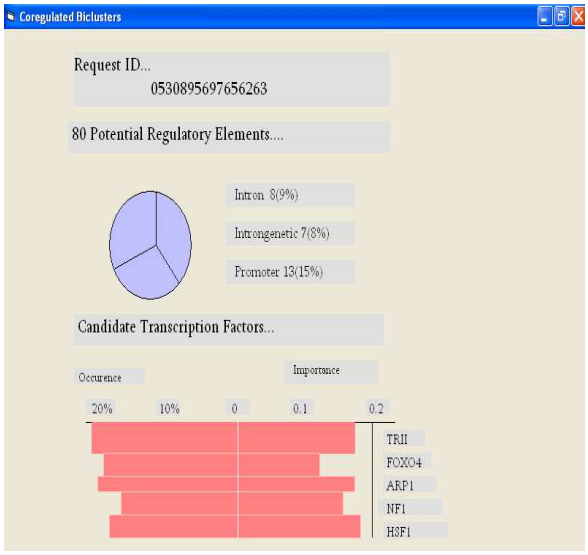

Figure 9a. Display the transcription factors from GED

Fig. 9a describes a display the transcription factors from GED. Gene name is matches from coregulated biclusters, then a given ID is generated and length is 25 digits. And then display the regulatory elements from expression data. It contains 3 types. They are Intron, Promoter and Intropromoter. Final one is display the transcription factors. ARP1 is occurring in 20% and the importance is 0.03456. And then display the transcription factors of the coregulated biclusters. Fig.9b describes Transcription Factor from GED. It contains the name of the transcription factor, occurrence and the importance. The name of transcription factor is TFIII, ARP1, NF1, HSF1 and HSF2 it will occur in the percentage is 9.52% and the Importance is 0.4434.

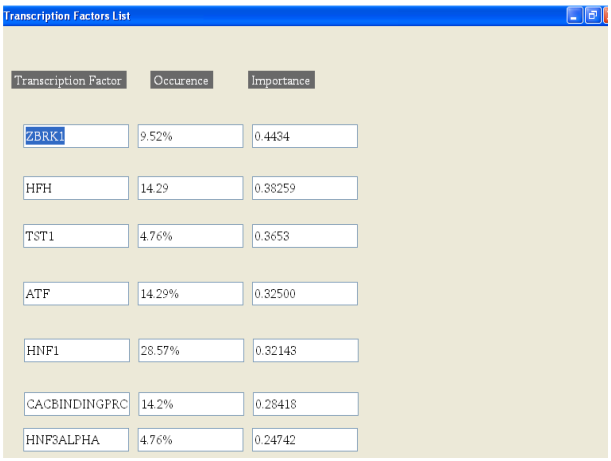

Figure 9b. Transcription Factor

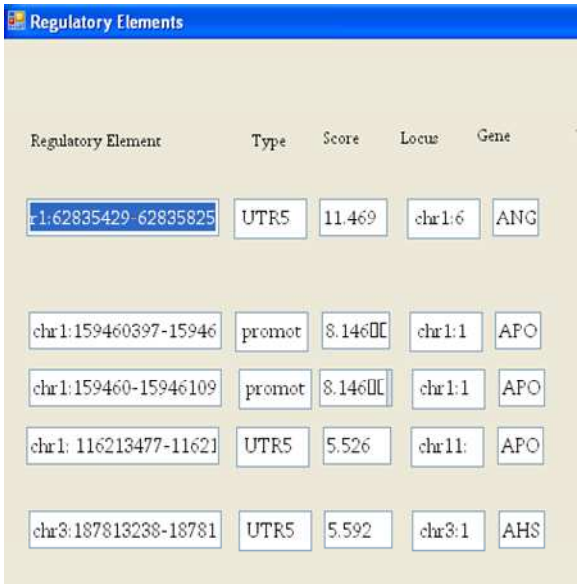

Figure 10. Regulatory elements from GED

Fig.10 describes Regulatory elements from GED. They are chromosomes, type, score, gene and Transcription factors binding sites from expression data. Type is an UTR5 and Promoter and gene is C4bp and Arg1 and Rdh7 and so on. Score is also defined as regulatory element execute the time on 0.248 seconds to complete the particular process. ARP1 occur in the percentage is 14.29% and the importance is 0.38529. NF1 occur in the percentage is 4.76% and the importance is 0.3653. HSF1 occur in the percentage is 28.57% and the importance is 0.32143.

VI. CONCLUSION

In this work, the transcription sites of coregulated biclusters are found. They are four pathways are applied on gene expression data. MPM pathway is used on analysis of DNA chips and gene networks. KEGG pathway is used on gene regulatory networks. CD3 used to find coregulated genes. DREB2A used on lymphoma 2-cell. Biclustering methods are used to compare the four pathways, KEGG got an optimal result. The proposed Enhanced Bimax algorithm is used and the coregulated biclusters are mined, to display the association group of patterns in the gene expression data. By comparing these pathways KEGG is an effective one. It is observed that the Enhanced Bimax is an effective one. The normalization technique is used to identify whether the particular gene is present or not. An experimental result on synthetic dataset for GED shows the reliability and efficiency of the proposed algorithm.

REFERENCES

[1] C.P. Chandran, and K. IswaryaLakshmi, "Mining Corelgulated Biclusters from Gene Expression Data", *Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, pp. 89-93, ISSN No. 978-1-4673-1038-3, 2012.

[2] H. Dunham and S. Sridhar, "Data mining introductory and advanced topics," *Pearson education*, Inc., 2006, ch.4, sec.4.1, pp. 123-125.

[3] J. Han. and M. Kamber, " Data Mining: Concepts and Techniques", *Elsevier*, 2007, ch.2, sec. 2.1, pp.12-14.

[4] K. Cios, W. Pedrycz, and R. Swiniarski, "Data Mining: A Knowledge Discovery Approach", *Springer*, 2007, ch.3, sec. 3.5, pp.56-58.

[5] H. Zhao and L. Cloots, "Query-based biclustering of gene expression data using Probabilistic Relational Models", *BMC Bioinformatics*, vol.12, Jan. 2011.pp.1471-2105, 2011, DOI: 10.2174/138920208786847935

[6] S. Busygin, and P. Pardalos "Biclustering in data mining", *Journal of Computers and Operations Research*, vol.35, Apr. 2008, pp.2964-2987, DOI: 10.1371/journal.pone.0032289

[7] A. Agra, "Biclustering of Gene Expressions", *ACM*, vol.2, Sep. 2006, pp.2-30, DOI: 10.1145/1183081.1183082

[8] R. Shamir, "Analysis of DNA Chips and Gene Networks Bioinformatics", *Springer*, vol.1, pp. 1-10, 2009.

[9] R. Guthke, "Gene Expression Data Mining for Functional Genomics using Fuzzy Technology", *Biomed Central*, vol.15, Aug. 2000, pp.23-36, DOI: 10.1093/bioinformatics/bti251.

[10] H. Banka, S. Mitra., "Evolutionary biclustering of gene expression data", *Proceedings of the 2nd international conference on Rough sets and knowledge technology*, Aug. 2000, pp. 284-291, DOI: 10.1504/IJBIC.2010.03702.

[11] Y. Luan and H. Li, "Clustering of time-course gene expression data using a mixed-effects model with B-splines", *Bioinformatics*, vol. 19, Oct. 2003, pp. 474–482, DOI: 10.1093/bioinformatics/btg014

[12] Y. Zhang and H. Zha, "A Time-Series Biclustering Algorithm for Revealing Co-regulated Genes", Bioinformatics, vol.23, pp.1-6, May. 2004. DOI: 10.1109/ITCC.2005.46

[13] K. Bloch, "Evolutionary biclustering of gene expression data", *2nd International conference on Rough sets and knowledge technology*, vol.3, Aug. 2007, pp. 284-291, DOI: 10.1504/IJBIC.2010.037021

[14] C. Yang and E. Zeng, "Clustering Genes using Gene Expression and Text Literature Data", *IEEE Computational Systems Bioinformatics Conference*, vol.12, Nov. 2008, pp. 21-33. DOI:10.1186/1471-2105-9-497.

[15] S. Das, "Greedy Search-Binary PSO Hybrid for Biclustering Gene Expression Data", *International Journal of Computer Applications*, vol. 2, May. 2010, pp.75 – 88, DOI: 10.1145/1722024.1722074.

[16] A. Gyenesei *et.al.*, "Mining co-regulated gene expression time series data for the detection of functional associations in gene expression data", *Bioinformatics*, vol.23, May. 2007, 1927–1935, DOI: 10.1093/bioinformatics/btm276.

[17] Q. Zhang, "Promoter Analysis of Co-regulated Genes in the Yeast Genome", *BMC Biomedcentral*, vol.1, Jun. 2000, pp.1-29, DOI: 10.1.1.86.6972.

[18] M. Kim, "Gene expression profiling in the lung tissue of cynomolgus monkeys in response to repeated exposure to welding fumes", *Biological Sciences*, vol.84, Jul.2010, pp. 191-203, DOI: 10.4172/2161-0495.S5-004.

[19] X. Xu, "Mining Shifting-and-Scaling Co-Regulation Patterns on Gene Expression Profiles", *Bioinformatics*, vol.1, Sep. 2009, pp.1-10, DOI: 10.1109/ICDE.2006.98.

[20] S. Dharan and S. Achuthsankar, "Biclustering of gene expression data using reactive greedy randomized adaptive search procedure", *Biomed Central*, vol.10, Jan.2009, pp. 13-19, DOI: 10.1145/1722024.1722041.

[21] A. Bhattacharya and K. Rajat, "Bi-correlation clustering algorithm for determining a set of co-regulated genes", *Bioinformatics*, vol. 25, Sep. 2009, pp. 2795–2801, DOI: 10.1093/bioinformatics/btp526.

[22] Y. Zhang *and* H. Zha, "A Time-Series Biclustering Algorithm for Revealing Co-regulated genes", *BMC Biomed Central*, vol.18, Jul. 2000, pp.1-6, DOI: 10.1109/ITCC.2005.46.

[23] S. Remondini, "Gene Expression Patterns", *Elsevier*, vol.12, Jan.2002, pp.12-20, DOI: 10.1371/journal.pgen.1002234.

[24] N. Haque, "Robust Hierarchical Clustering for Gene Expression Data Analysis", *Biostatistics*, vol.7, Jan. 2006, pp.1-5, DOI: 10.1186/1471-2105-9-497.

[25] T. Bulcke, “A benchmark algorithm for elastoplasticity with multiple yield surfaces”, *Soil Dynumics and Eurthquake Engineering*, vol.10, Jan. 1991, pp. 341-347, DOI: 10.1108/02644400110365842.

[26] Y. Cheng and G.M Church, “Biclustering of expression data”, *International Conference on Intelligent Systems for Molecular Biology*, vol.12, Jan. 2000, pp.93-103, DOI: 10.1186/1471-2105-7-78.

[27] P. Waltman, "Multiple species Integrative Biclustering", *Genome Biology*, vol.11, Dec. 2010, pp.1-10. DOI: 10.1186/1471-2105-7-280.

[28] A. Prelic Bleuler and S. Zimmermann, “A systematic comparison and evaluation of biclustering methods for gene expression data”, *bioinformatics*, vol.22, Aug. 2000, pp.112-212, DOI: 10.1093/bioinformatics/btl060.

[29] A. Ben-Dor, “Class Discovery in Gene Expression Data”, Bioinformatics, vol.23, Sep.2003, pp.34-42, DOI: 10.1145/369133.369167.

[30] R. Rathipriya, "Evolutionary of Biclustering algorithms", *International Journal of Computer Applications*, vol.8, Jan. 2011, pp.1-10, DOI: 10.1186/1756-0381-4-3.

Dr.C.P.Chandran is an Associate Professor of Computer Science and Head of Post-Graduate Department of Bioinformatics, Co-ordinator, Center for Technology Enhanced Learning (CTEL), Ayya Nadar Janaki Ammal College, Sivakasi, India, (Madurai Kamaraj University). He received his Doctoral degree in Computer Science from Madurai Kamaraj University, India. He has about 16 years of teaching experience in Computer Science. He is an Alumni of Department of Physics, NGM College, Pollachi and Department of Computer Science, Bharathiar University, Coimbatore, India. He is an editor/reviewer of various international journals. His research focuses on data mining in bioinformatics, rough sets, swarm intelligence and granular computing. He has authored 10 international journal papers. He has published 30 research papers in national and international conference proceedings.

K.IswaryaLakshmi was born in Sivakasi, Tamil Nadu (TN), India, in 1988. He received the Bachelor of Computer Science (B.Sc) degree from the Ayya Nadar Janaki Ammal College, Sivakasi, TN, India, in 2009 and the Master of Computer Science (M.Sc) degree from the Ayya Nadar Janaki Ammal College, Sivakasi, TN, India, and Master of Philosophy (M.Phil) of Computer Science degree from the Ayya Nadar Janaki Ammal College, Sivakasi in 2012. Her research interests include data mining and bioinformatics.