IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 6, June 2016

A New Approach to Rank Based Weighted Association Rule Mining Based on Fuzzy C- Means Algorithm

J.S. Esther Sylvia Jebarani [1], Mr. S. Saravana Kumar M.E[2]

M.E Computer Science and Engineering, SVS College of Engineering, Coimbatore, Tamilnadu, India [1]

Assistant Professor, SVS College of Engineering, Coimbatore, Tamilnadu, India [2]

Abstract: The Association Rule Mining is defined as a process of Finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transactional databases, relational databases, and other information repositories. This method is commonly used in bioinformatics for the ranking of genes and genomes. There is a drawback, which makes the decision maker more confusion due to huge number of evolved rules. To avoid this, a weighted association rule mining called RANWAR (or) Rank based Weighted Association Rule Mining which uses our proposed rule interestingness measures, viz., rank-based weighted condensed support (WCS) and weighted condensed confidence (WCC) is proposed in this paper. Based on these measures we assign weight to the each item, which generates less number of frequent item sets than state-of-the-art association rule mining. This process is run on Gene Expression and Methylation datasets. The resulted genes of the top rules are biologically validated by Gene Ontologies (GOs) and KEGG pathway analyses. The top ranked rules extracted from RANWAR that hold poor ranks in traditional Apriori, are highly biologically significant to the related diseases. This paper report the top rules evolved from RANWAR that are not in Apriori.

Keywords: Weighted Condensed Support (WCS), Weighted Condensed Confidence (WCC), limma.

1. INTRODUCTION

Association rule mining aims to explore large transaction databases for association rules. Classical Association Rule Mining (ARM) model assumes that all items have the same significance without taking their weight into account. It also ignores the difference between the transactions and importance of each and every itemsets. But, the Weighted Association Rule Mining (WARM) does not work on databases with only binary attributes. It makes use of the importance of each itemsets and transaction.

WARM requires each item to be given weight to reflect their importance to the user. The weights may correspond to special promotions on some products, or the profitability of different items. The concept of association rule mining proposes the support-confidence measurement framework and reduced association rule mining to the discovery of frequent item sets. WARM generalizes the traditional model to the case where items have weights. WARM requires for each item to be given weight to reflect their importance to the user. The weights may correspond to the profitability of different items. As more data is gathered, which are frequently getting updated, the construction of the graph should be dynamic instead of static. Using Online Hits algorithm, the graph can be constructed dynamically and the cost can be reduced by postponing updates whenever possible. By calculating Eigen values the mutual reinforcement relationship between the items are enforced. Limma is a package for differential expression analysis of data arising from microarray experiments. The package is designed to analyze complex experiments involving comparisons between many RNA targets simultaneously while remaining reasonably easy to use for simple experiments. The central idea is to fit a linear model to the expression data for each gene. Limma is designed to be used in conjunction with the affy or affy PLM packages for Affymetrix data. With two color microarray data, the m array package may be used for pre-processing. Limma itself also provides input and normalization functions which support features especially useful for the linear modeling approach. Microarray-based gene expression profiling experiments, which are routine today, allow researchers to identify, for instance, genes differentially expressed (DE) between diseased and normal patient samples or genes that change in expression over time during a treatment.

These aspects would include the position and role of each gene in a pathway, the types of signals between genes, the efficiency with which a signal travels from one gene to another, or the efficiency with which a certain reaction is carried out, rate limiting conditions, etc. Such methods have been proposed for both signalling pathways, and metabolic pathways, but no method is currently available to analyze both types of pathways taking into consideration all the information available. Hence, even though they do not use all information available, methods that treat the pathways as simple gene sets are still popular because they can be applied equally well to signalling pathways, metabolic pathways, GO terms, as well as arbitrary sets of genes.

DOI 10.17148/IJIREEICE.2016.4613

II.LITRATURE SURVEY

ARM is a popular technique to estimate interesting relationships among different items (genes).Apriori is the basic algorithm for learning basic association rules to control on database that have transactions. Apriori utilizes a bottom-up approach, when frequent subsets are extended one item at a time to generate each candidate and group of candidates are tested against the data.

In this paper different limitations have been found in the Apriori algorithm like generating a huge number o frequent item sets high elapsed time, multiple scan problems, importing same importance to all data sets. To reduce these problems ARM technique is introduced in this paper. ARM is a efficient technique, In this paper it mainly focus how to reduce elapsed time for rule mining in such a way that only top ranked items and their related highly significant rules will present in result for large transaction database.

A. Association Rule Mining In Genomics

Association rules, used widely in the area of market basket analysis, can be applied to the analysis of expression data as well. Association rules can reveal biologically relevant associations between different genes or between environmental effects and gene expression. An association rule has the form LHS→RHS, where LHS and RHS are disjoint sets of items, the RHS set being likely to occur whenever the LHS set occurs.

Items in gene expression data can include genes that are highly expressed or repressed, as well as relevant facts describing the cellular environment of the genes (e.g. the diagnosis of a tumour sample from which a profile was obtained). In this paper, association rule mining techniques that have been recently developed and used for genomic data analysis have been reviewed and discussed.

B. Mining Weighted Association Rules without Preassigned Weights

Association rule mining is a key issue in data mining. However, the classical models ignore the difference between the transactions, and the weighted association rule mining does not work on databases with only binary attributes. In this paper, we introduce a new measure w-support, which does not require pre assigned weights. It takes the quality of transactions into consideration using link-based models. A fast mining algorithm is given, and a large amount of experimental results are presented.

C. Mining Association Rules between Sets of Items In Large Databases

Here a large database is used of customer transactions. Each transaction consists of items purchased by a customer in a visit. Here an efficient algorithm that generates all significant association rules between items in the database. The algorithm incorporates buffer management and novel estimation and pruning techniques. These paper also present results of applying this algorithm to sales data obtained from a large retailing company, which shows the effectiveness of the algorithm.

III.PROPOSED ARM (RANWAR) TECHNIQUE

In this proposed technique, number of genes in data is large, the number of itemsets will be also large, thus, using Limma statistical test, an useful statistical process we have just taken into account the top differentially expressed (i.e., DE) or differentially methylated (i.e., DM) genes. Our proposed measures are basically rank-based weighted measures. Therefore, ranking of genes has a significant role here. Limma provides a rank-wise gene-list according to their p-values from best to worst cases. Thereafter, we assign weight to each item/gene with respect to their p-value ranking, and include these into the measures. Therefore, our measures give importance to each item (gene) by data discretization process which uses K-means Clustering process.

A. Identifying Differential Expressed/Methylated Items (Genes):

The first module is to identify the differential expressed or differential methylated genes. Here starting process pre-filtering process is applied on the data (viz., removal of genes having low variance which is insignificant for the further process). Thus, it is needed to check the overall variance of the data according for each gene and filter out the genes having very low variance. The filtered data should be normalized gene-wise as normalization (Zero Normalization) converts the data from different scales into a common scale. We use zero-mean normalization for converting the data into structural form where mean of each gene becomes zero and standard deviation becomes one. Then to identify DE/DM genes, a suitable non-parametric test should be applied correctly.

Thus, we choose Limma as it performs well for both normal and non-normal distributions for all sizes of data. The moderated t-statistic of Limma is stated respectively. However, from the resulting value of the t-statistic, corresponding p-value is calculated from t-table or cumulative distribution function (cdf). If p-value of a gene is less than 0.05, then the gene is called, otherwise not. The genes are then ranked with respect to their p-values.

B. Assigning Weight:

In our approach, Assigning weight is the second module; here all the genes have not same importance. To differentiate the genes, some weight is assigned to each gene with respect to their p-value ranking mentioned earlier. Here, the weights of the genes are calculated by the difference between the weights of any two consecutive ranked genes are same, and the weight of the first ranked gene is always 1 .The ranges of weight lie in between 0 and 1.

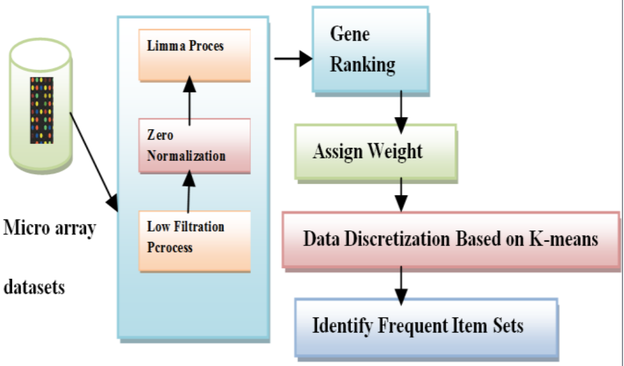
C. Data Discretization:

Data discretization is the third module in our approach. Here assume I[r, c] is input data matrix. Here, r denotes genes, c and denotes samples. First of all, the matrix I is transposed. Suppose, IT be the resulting matrix. Now, Discretization of the input data matrix is mandatory for applying association rule mining. For Discretization purpose we use K-means algorithms. For doing this we

DOI 10.17148/IJIREEICE.2016.4613

make the point (centroid) in the data set and make clusters based on those points. Then we have to run K-means algorithm in a sample wise on each row of IT.

D. Identifying Frequent Item Set And Rule Mining:
Identifying frequent item set is the final module in the framework. After the Data discretization, we need to identify frequent itemsets. For identifying the frequent itemsets, we evaluate of the 1-itemsets, and then identify the frequent singleton itemsets. Similarly, we calculate their supersets 2-itemsets and then determine frequent 2-itemsets. Then rules are extracted from the frequent 2-itemsets.Then, WCC of each rule is computed. The rules having greater than equal to minimum confidence value, are selected for resulting list of rules. Then, we determine their supersets 3-itemsets and then determine frequent 3-itemsets, and then extract significant rules from these, and so on. The algorithm will be stopped, if there is no further extension of frequent itemsets to be identified.



System Architecture

Algorithm used in RANWAR technique
The most common algorithm uses an iterative refinement technique. Due to its ubiquity it is often called the k-means algorithm; it is also referred to as Lloyd's algorithm, particularly in the computer science community. Given an initial set of k means $m_1,\dots,m_k$ (see below), the algorithm proceeds by alternating between two steps:

Assignment step: Assign each observation to the cluster whose mean yields the least within-cluster sum of squares (WCSS). Since the sum of squares is the squared Euclidean distance, this is intuitively the "nearest" mean. (Mathematically, this means partitioning the observations according to the Voronoi diagram generated by the means).

$$S_i^{(t)} = \{x_p : \|x_p - m_i^{(t)}\|^2 \le \|x_p - m_j^{(t)}\|^2 \ \forall j, 1 \le j \le k\},$$

where each x_p is assigned to exactly one $S^{(t)}$, even if it could be assigned to two or more of them.

Update step: Calculate the new means to be the centroid of the observations in the new clusters.

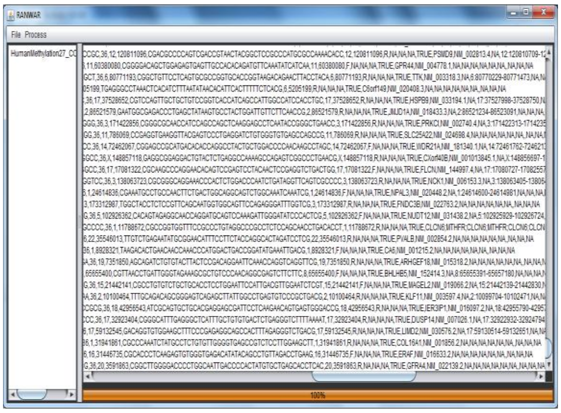
$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

Since the arithmetic mean is a least-squares estimator, this also minimizes the within-cluster sum of squares (WCSS) objective.

The algorithm has converged when the assignments no longer change. Since both steps optimize the WCSS objective, and there only exists a finite number of such partitionings, the algorithm must converge to a (local) optimum. There is no guarantee that the global optimum is found using this algorithm.
The algorithm is often presented as assigning objects to the nearest cluster by distance. The standard algorithm aims at minimizing the WCSS objective, and thus assigns by "least sum of squares", which is exactly equivalent to assigning by the smallest Euclidean distance. Using a different distance function other than (squared) Euclidean distance may stop the algorithm from converging. Various modifications of k-means such as spherical k-means and k-medoids have been proposed to allow using other distance measures.

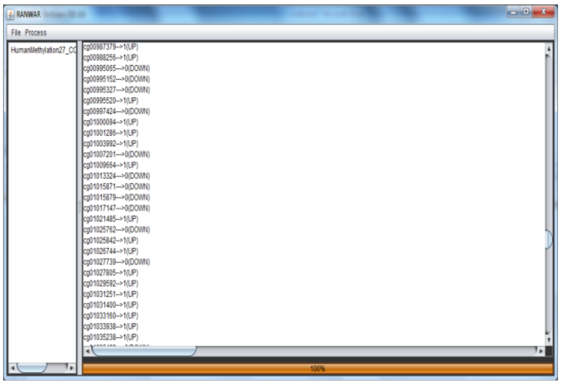
IV. IMPLEMENTATION AND PERFORMANCE EVALUATION

Data samples are collected to find the genes with high priority. Biological data sets are collected from the methylated datasets. The low variants are eliminated when the pre filtering process is applied. Limma statistical test is performed for the datasets. Assigning each gene a rank with respect to their weight. In this paper we apply Fuzzy c-means algorithm instead of k-means algorithm for the clustering process. RANWAR algorithm is used to extract the rules. Ranks are assigned to all the genes based on their weights.



```
CCGC,36,12,120811096,CGACGCCCCAGTCGACCGTAACTACGGCTCCGCCCATGCGCCAAAACACC,12,120811096,R,NA,NA,NA,TRUE,PSMD9,NM_002813.4,NA,12:120810709-12
,11,60380080,CGGGGACAGCTGGAGAGTGAGTTGCCACACAGATGTTCAAATATCATCAA,11,60380080,F,NA,NA,NA,TRUE,GPR44,NM_004778.1,NA,NA,NA,NA,NA,NA,NA,,NA,NA,NA
GCT,36,6,80771193,CGGCTGTTCCTCAGTGCGCCGGTGCACCGGTAAGACAGAACTTACCTACA,6,80771193,R,NA,NA,NA,TRUE,TTK,NM_003318.3,NA,6:80770229-80771473,NA,N
05199,TGAGGGCCTAAACTCACATCTTTAATATAACACATTCACTTTTTCTCACG,6,5205199,R,NA,NA,NA,TRUE,C6orf149,NM_020408.3,NA,NA,NA,NA,NA,NA,NA,,NA,NA,NA
,36,17,37528652,CGTCCAGTTGCTGCTGTCCGGTCACCATCAGCCATTGGCCATCCACCTGC,17,37528652,R,NA,NA,NA,TRUE,HSPB9,NM_033194.1,NA,17:37527998-37528750,N
2,86521579,GAATGGCGAGACCCTGAGCTATAAGTGCCTACTGGATTGTTCTTCAACCG,2,86521579,R,NA,NA,NA,TRUE,JMJD1A,NM_018433.3,NA,2:86521234-86523091,NA,NA,NA,
GGG,36,3,171422856,CGGGCGCAACCATCCAGCCAGCTCAAGGAGCCTCAATACCGGGCTGAACC,3,171422856,R,NA,NA,NA,TRUE,PRKCI,NM_002740.4,NA,3:171422313-1714235
GG,36,11,786069,CCGAGGTGAAGGTTACGAGTCCCTGAGGATCTGTGGGTGTGAGCCAGCCG,11,786069,R,NA,NA,NA,TRUE,SLC25A22,NM_024698.4,NA,NA,NA,NA,NA,NA,NA,,NA,NA,
CC,36,14,72462067,CGGAGCCGCATGACACACCAGGCCTACTGCTGGACCCCAACAAGCCTAGC,14,72462067,F,NA,NA,NA,TRUE,WDR21A,NM_181340.1,NA,14:72461762-724621
GCC,36,X,148857118,GAGGCGGAGGACTGTACTCTGAGGCCAAAGCCAGAGTCGGCCCTGAACG,X,148857118,R,NA,NA,NA,TRUE,CXorf40B,NM_001013845.1,NA,X:148856697-1
GCC,36,17,17081322,CGCAAGCCCAGGAACACAGTCCGAGTCCTACAACTCCGAGGTCTGACTGG,17,17081322,F,NA,NA,NA,TRUE,FLCN,NM_144997.4,NA,17:17080727-1708257
GGTCC,36,3,138063723,CGCGGGCAGGAAACCCACTCTGGACCCAATCTGATAGGTTCAGTCGCCCC,3,138063723,R,NA,NA,NA,TRUE,NCK1,NM_006153.3,NA,3:138063405-13806
5,1,24614836,CGAAATGCCTGCCAACTTCTGACTGGCAGGCAGTCTGGCAAATCAAATCG,1,24614836,F,NA,NA,NA,TRUE,NPAL3,NM_020448.2,NA,1:24614600-24614981,NA,NA,NA,
3,173312987,TGGCTACCTCTCCGTTCAGCAATGGTGGCAGTTCCAGAGGGATTTGGTCG,3,173312987,R,NA,NA,NA,TRUE,FNDC3B,NM_022763.2,NA,NA,NA,NA,NA,NA,NA,,NA,NA,NA
G,36,5,102926362,CACAGTAGAGGCAACCAGGATGCAGTCCAAAGATTGGGATATCCCACTCG,5,102926362,F,NA,NA,NA,TRUE,NUDT12,NM_031438.2,NA,5:102925929-102926724,
GCCCC,36,1,11788672,CGCCGGTGGTTTCCGCCCTGTAGGCCCGCCTCTCCAGCAACCTGACACCT,1,11788672,R,NA,NA,NA,TRUE,CLCN6;MTHFR;CLCN6;MTHFR;CLCN6;CLCN
6,22,35546013,TTGTCTGAGAATATGCGGAACATTTCCTTCTACCAGGCACTAGATCCTCG,22,35546013,R,NA,NA,NA,TRUE,PVALB,NM_002854.2,NA,NA,NA,NA,NA,NA,NA,,NA,NA,NA
36,1,8928321,TAAGACACTGAACAACCAAACCCATGGACTGACCGGATATGAAATTGACG,1,8928321,F,NA,NA,NA,TRUE,CA6,NM_001215.2,NA,NA,NA,NA,NA,NA,NA,,NA,NA,NA
A,36,19,7351850,AGCAGATCTGTGTACTTACTCCGACAGGAATTCAAACCAGGTCAGGTTCG,19,7351850,R,NA,NA,NA,TRUE,ARHGEF18,NM_015318.2,NA,NA,NA,NA,NA,NA,NA,,NA,NA,N
,65665400,CGTTAACCTGATTGGGTAGAAAGCGCTGTCCCAACAGGCGAGTCTTCTTC,8,65665400,F,NA,NA,NA,TRUE,BHLHB5,NM_152414.3,NA,8:65665391-65657180,NA,NA,N
G,36,15,21442141,CGCCTGTGTCTGCTGCACCTCCTGGAATTCCATTGACGTTGGAATCTCGT,15,21442141,F,NA,NA,NA,TRUE,MAGEL2,NM_019066.2,NA,15:21442139-21442830,N
A,36,2,10100464,TTTGCAGACAGCGGGAGTCAGAGCTTATTGGCCTGAGTGTCCCGCTGACG,2,10100464,R,NA,NA,NA,TRUE,KLF11,NM_003597.4,NA,2:10099704-10102471,NA,N
CGCG,36,18,42956543,ATCGCAGTGCTGCACGAGGAGCGATTCCTCAAGAACAGTGAGTGGGACCG,18,42956543,R,NA,NA,NA,TRUE,IER3IP1,NM_016097.2,NA,18:42955790-42957
CC,36,17,32923404,CGGGCATTTGAGGGCTCATTTGCTGTGTGACTCTGAGGGTCTTTTAAAAT,17,32923404,R,NA,NA,NA,TRUE,DUSP14,NM_007026.1,NA,17:32922932-32924794
6,17,59132545,GACAGGTGTGGAAGCTTTCCCGAGAGGCAGCCACTTTAGAGGGTCTGACG,17,59132545,R,NA,NA,NA,TRUE,LIMD2,NM_030576.2,NA,17:59130514-59132651,NA,N
6,1,31941861,CGCCCAAATCTATGCCTCTGTGTTGGGGTGAGCCGTCTCCTTGGAAGCTT,1,31941861,R,NA,NA,NA,TRUE,COL16A1,NM_001856.2,NA,NA,NA,NA,NA,NA,NA,,NA,NA,NA
6,16,31446735,CGCACCCTCAAGAGTGTGGGTGAGACATATACAGCCTGTTAGACCTGAAG,16,31446735,F,NA,NA,NA,TRUE,ERAF,NM_016633.2,NA,NA,NA,NA,NA,NA,NA,,NA,NA,NA
G,36,20,3591863,CGGCTTGGGGACCCCTGGCAATTGACCCCACTATGTGCTGAGCACCTCAC,20,3591863,R,NA,NA,NA,TRUE,GFRA4,NM_022139.2,NA,NA,NA,NA,NA,NA,NA,,NA,NA,NA
```

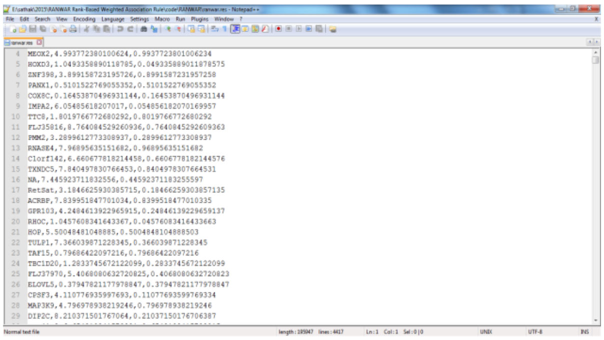
Fig1.shows all the datasets which are loaded for gene ranking



```
cg00987379-->1(UP)
cg00988256-->1(UP)
cg00995065--->0(DOWN)
cg00995152--->0(DOWN)
cg00995327--->0(DOWN)
cg00995520-->1(UP)
cg00997424--->0(DOWN)
cg01000094-->1(UP)
cg01001286-->1(UP)
cg01003992-->1(UP)
cg01007201--->0(DOWN)
cg01009664-->1(UP)
cg01013324--->0(DOWN)
cg01015871--->0(DOWN)
cg01015879--->0(DOWN)
cg01017147--->0(DOWN)
cg01021485-->1(UP)
cg01025762--->0(DOWN)
cg01025842-->1(UP)
cg01026744-->1(UP)
cg01027739--->0(DOWN)
cg01027805-->1(UP)
cg01029592-->1(UP)
cg01031251-->1(UP)
cg01031400-->1(UP)
cg01033160-->1(UP)
cg01033938-->1(UP)
cg01035238-->1(UP)
```

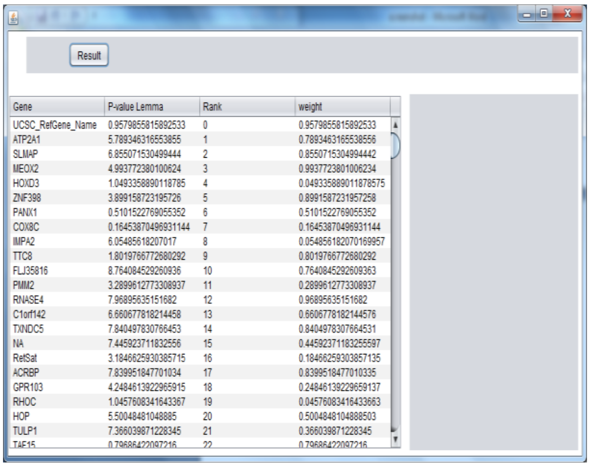
Fig2. All the low variants datasets are eliminated using the prefiltering process

DOI 10.17148/IJIREEICE.2016.4613


E:\sathak\2015\RANWAR Rank-Based Weighted Association Rule\code\RANWAR\ranwar.res - Notepad++
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
ranwar.res
Normal text file

```
MEOX2,4.993772380100624,0.9937723801006234
HOXD3,1.0493358890118785,0.049335889011878575
ZNF398,3.899158723195726,0.8991587231957258
PANX1,0.5101522769055352,0.5101522769055352
COX8C,0.16453870496931144,0.16453870496931144
IMPA2,6.05485618207017,0.054856182070169957
TTC8,1.8019766772680292,0.8019766772680292
FLJ35816,8.764084529260936,0.7640845292609363
PMM2,3.2899612773308937,0.2899612773308937
RNASE4,7.96895635151682,0.96895635151682
Clorf142,6.660677818214458,0.6606778182144576
TXNDC5,7.840497830766453,0.8404978307664531
NA,7.445923711832556,0.44592371183255597
RetSat,3.1846625930385715,0.18466259303857135
ACRBP,7.839951847701034,0.8399518477010335
GPR103,4.2484613922965915,0.24846139229659137
RHOC,1.0457608341643367,0.045760834164336663
HOP,5.50048481048885,0.500484810488503
TULP1,7.366039871228345,0.366039871228345
TAF15,0.79686422097216,0.79686422097216
TBC1D20,1.2833745672122099,0.2833745672122099
FLJ37970,5.4068080632720825,0.40680806327208823
ELOVL5,0.37947821177978847,0.37947821177978847
CPSF3,4.110776935997693,0.1107769359976934
MAP3K9,4.796978938219246,0.796978938219246
DIP2C,8.210371501767064,0.21037150176706387
```

Fig3.weights are assigned to all the genes.



Gene	P-value Lemma	Rank	weight
UCSC_RefGene_Name	0.9579855815892533	0	0.9579855815892533
ATP2A1	5.789346316553855	1	0.789346316553856
SLMAP	6.855071530499444	2	0.855071530499442
MEOX2	4.993772380100624	3	0.9937723801006234
HOXD3	1.0493358890118785	4	0.049335889011878575
ZNF398	3.899158723195726	5	0.8991587231957258
PANX1	0.5101522769055352	6	0.5101522769055352
COX8C	0.16453870496931144	7	0.16453870496931144
IMPA2	6.05485618207017	8	0.054856182070169957
TTC8	1.8019766772680292	9	0.8019766772680292
FLJ35816	8.764084529260936	10	0.7640845292609363
PMM2	3.2899612773308937	11	0.2899612773308937
RNASE4	7.96895635151682	12	0.96895635151682
C1orf142	6.660677818214458	13	0.6606778182144576
TXNDC5	7.840497830766453	14	0.8404978307664531
NA	7.445923711832556	15	0.44592371183255597
RetSat	3.1846625930385715	16	0.18466259303857135
ACRBP	7.839951847701034	17	0.8399518477010335
GPR103	4.2484613922965915	18	0.24846139229659137
RHOC	1.0457608341643367	19	0.045760834164336663
HOP	5.50048481048885	20	0.500484810488503
TULP1	7.366039871228345	21	0.366039871228345
TAF15	0.79686422097216	22	0.79686422097216

Fig4.Rank and weight is assigned to all the genes

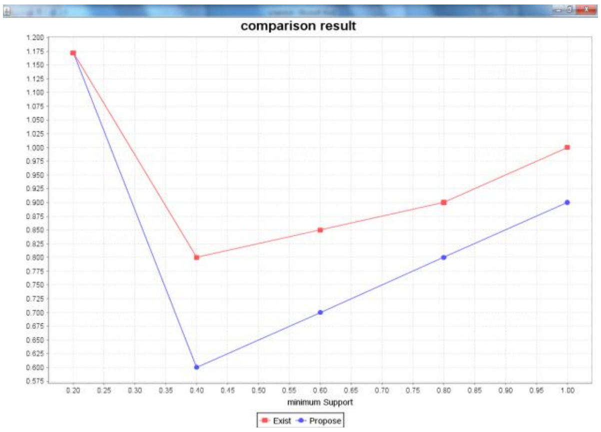


Fig5.This shows the comparison and evaluation result of the minimum support and number of frequent item datasets

V. CONCLUSION

There are huge numbers of evolved rules of items (or genes) by Association Rule Mining algorithms make confusion to choose the top genes for the decision maker. In this paper the two new rank based weighted condensed rule-interesting measures called weight condensed confidence and weight condensed support (WCC, WCS) are introduced. A weighted rule mining algorithm called RANWAR, which has been developed using the measures especially for micro array data. RANWAR uses a statistical process called Limma to compute p-value of each gene (item), and adding some weight to given genes. It saves time of execution of the algorithm. Finally top rules extracted from RANWAR that are not present in Apriori, which have high biological significance. Also in future work Fuzzy c-means clustering is used in the Data Discretization process, instead of K-means clustering process. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the centre of cluster. Our comparison results provide the efficient clustering process in RANWAR.

REFERENCES

[1] Thomas, "Expression profiling of cervical cancers in Indian women at different stages to identify gene signatures during progression of the disease," Cancer Med., vol. 2, no. 6, pp. 836–848, Dec. 2013.

[2] J. Liu, "Identifying differentially expressed genes and pathways in two types of non-small cell lung cancer: Adenocarcinoma and squamous cell carcinoma," Genet. Mol. Res., vol. 13, pp. 95–102, 2014.

[3] W. Wei, "The potassium-chloride cotransporter 2 promotes cervical cancer cell migration and invasion by an ion transport-independent mechanism," J Physiol., vol. 589, pp. 5349–5359, 2011.

[4] J. Pavon, S. Viana, and S. Gomez, "Matrix Apriori: Speeding up the search for frequent patterns," in Proc. IASTED, 24th Multi-Conf. Appl. Informat., Innsbruck, Austria, 2006.

[5] Y. Hong , "Incrementally fast updated frequent pattern trees," Expert Syst. Appl., vol. 34, pp. 2424–2435, 2008.

[6] D. Oguz and B. Ergenc, Incremental Item set Mining Based on Matrix Apriori Algorithm. Berlin/Heidelberg, Germany: Springer, 2012, pp. 192–204.

[7] S. Orlando, "Enhancing the apriori algorithm for frequent set counting," in Data Warehousing and Knowledge Discovery. Berlin/ Heidelberg, Germany: Springer , 2013, pp. 71–82.

[8] U. Yun, "WIP: Mining Weighted Interesting Patterns with a strong weight and/or support affinity," in Proc. SDM, 2006, vol. 6, pp. 3477–3499.

[9] K. M. Yu and J. L. Zhou, "A Weighted Load-balancing parallel apriori algorithm for association rule mining," in Proc. IEEE Int. Conf. Granular Computing (GrC 2008), pp. 756–761.

[10] K. Sun and F. Bai, "Mining weighted association rules without Preassigned weights," IEEE Trans. Know. Data Eng., vol. 20, no. 4, pp. 489–495, 2008.

DOI 10.17148/IJIREEICE.2016.4613