iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** III **Month of publication:** March 2018

DOI: http://doi.org/10.22214/ijraset.2018.3181

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Prediction of Micro-RNA Diseases using Cross Ontology and Gene Ontology

Mecthaline Lithiya.T[1], Sandhya. R[2], Sushmitha. G[3], Deepa. J[4], Jerrin Simla. A[5], Panimalar.S[6]
[1, 2, 3]*Student IV Year CSE* , [4,5,6]*Assistant Professor, Panimalar Institute of Technology, Chennai.*

Abstract: ***Gene Ontology (GO) is a structured repository of conception that are associated to one or more gene products through a process referred to as annotation. There are different approaches of analysis to get bio information. One of the analysis is the use of Association Rules (AR) which discovers biologically relevant associations between terms of GO. In this gene ontology-based annotated datasets are associated for GO-WAR. The MOAL algorithms adapted to mine cross-ontology association rules, i.e. rules that involve GO terms present in the three sub-ontologies of GO. In this paper cross ontology is proposed to manipulate the Protein values from three sub ontologies for identifying the gene attacked disease. Also the proposed system, focus on intrinsic and extrinsic. Based on cellular component, molecular function and biological process values intrinsic and extrinsic calculation would be manipulated. In this Paper, the Co-Regulatory modules is done between miRNA (microRNA), TF(Transcription Factor) and gene on function level with multiple genomic data.. The regulations between miRNA-TF interaction, TF-gene interactions and gene-miRNA interaction are compared with the help of integration technique. These interactions could be taken the genetic disease like breast cancer, etc. Iterative Multiplicative Updating Algorithm is used in this paper to solve the optimization module function for the above interactions. After that interactions, the regulatory modules and protein value for gene are compared and generate the Bayesian rose tree for efficiency of the result.***

Keywords: Gene Ontology, Collaborative Filtering, Depth First Search, Regulatory modules, Integration Technique, Multiplicative Update Algorithm, Tree Representation.

I. INTRODUCTION

Ontologies are determining of a relational concordance, molecular functions in a species-independent manner. Bio- molecules of the cellular components and also living organs cells are the structural and functional. . While largest organs are accumulate of trillions cells and smallest organs are single cells. DNA is found in nearly all living cells, each cell carries chromosome having a distinctive DNA sequence.

The introduction of high-throughput technologies in molecular biology has produced the accumulation of a large set of experimental data. Such amount of experimental data has been integrated with additional information able to explain such data. For instance, genes and proteins have been accompanied by the storing of additional information used for the elucidation of the role of the investigated molecules. In order to systematize such knowledge, formal instruments such as controlled vocabularies and ontologies have been used to manage the used terms. Different ontologies have been proposed to elucidate different fields. For instance, the Gene Ontology (GO) is one of the frameworks that are largely used. Encompass of ontology have 3 sub ontology's i.e. BP, MF, CC. Each GO term is uniquely identified by a code, it belongs to only one ontology and for each GO Term a textual description is also available. For instance GO: 0006915 represent the apoptosis process. Increasingly large amounts of valuable, but heterogeneous and sparse, bimolecular data and information are characterizing life sciences [1]. In particular, semantic controlled annotations of bimolecular entities, i.e. the associations between bimolecular entities (mainly genes and their protein products) and controlled terms that describe the bimolecular entity features or functions, are of great value; they support scientists with several terminologies and ontologies describing structural, functional and phenotypic biological features of such entities (e.g. their sequences polymorphisms, expression in different tissues, and involvement in biological processes, biochemical pathways and genetic disorders).

II. RELATED WORK

Bio-molecules are the cellular components and these are living organs. In this organs are single as well as largest Cells, SO by using of DNA can easily found out nearby living cells. In the sequence of DNA distinctive having a chromosome cell carries. This type of cell is called biological matter or biological material, as some definitions of life exclude hypothetical types of biochemistry. BP is

made up of chemical reactions or other events that involved in the persistence and transformation of forms for example metabolism and homeostasis. Include the control of gene expression, interaction with a protein, protein modification, substrate molecule.

III. EXISTING SYSTEM

Existing system proposed association rules to support GO curators. It evaluates the annotation consistency in order to avoid possible inconsistent or redundant annotations. It uses the method called Classical association rules mining algorithms. In this base paper we used two techniques i.e., cross Ontology and Integration, in Ontology we existed patient gene ids as well as protein values. To calculate values of BP, MF, CC these three values can compare by using of ontology.

IV. PROPOSED SYSTEM

In this paper, the proposed co-regulatory modules between Transcription Factor , gene and Mi-RNA on functional level with genomic data. The integration technique supplemented between mi-RNA, Transcription Factor(TF) and gene. After integration, Iterative Multiplication update algorithm is used to check the optimization function between the regulatory modules. The expression or some value from this algorithm then compare to protein values. The protein value get from Biological Process(BP),Molecular Function(MF) and Cellular Component(CC) with the help of cross ontology technique. At last generate a Bayesians rose tree structure for the relation between regulatory modules and protein values of the gene. By this structure the disease which was affected in the chromosome are known and also know how to cure and the symptoms that are applicable for our gene by our web application.

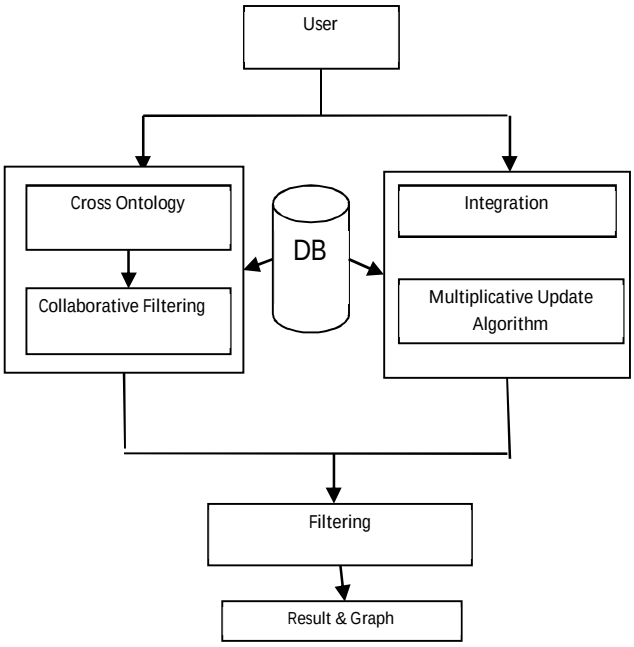


Fig. 1: Architecture Diagram

A. *Techniques And Description*

1) Gene Ontology
2) Collaborative Filtering
3) Depth First Search
4) Regulatory modules
5) Integration Technique
6) Multiplicative Update Algorithm
7) Tree Representation.

1) *Gene Ontology:* The Gene Ontology (GO) project is a co-operate effort to situation the need for consistent elucidation of gene products in different databases. In this paper the gene ontology is proposed , User login and register their details and get the gene id from Ontology base with the help of KNN algorithm. Full details of overall project are maintained our database

and ontology base. The proposed system includes cross ontology to manipulate the Protein values from three sub ontologies for identifying the gene attacked disease. Also the proposed system, focus on intrinsic and extrinsic. Based on cellular component, molecular *function* and biological process values intrinsic and extrinsic calculation would be manipulated

2) *Collaborative Filtering:* In this paper the semantic mining is used for logical analysis. User get the details from Ontology base with help of Collaborative filtering, also the gene disease and symptoms with the help of logical calculation for protein value of human and normal value for particular gene id, then cross ontology process to get the BP,CC&MF value for gene to identify the gene have Intrinsic or extrinsic.

a) *Intrinsic:* If the normal protein value of human is compare to lower than that of calculating cross ontology value (comparing BP&CC or MF&CC or MF&BP) is said to be Intrinsic.

b) *Extrinsic:* If the normal protein value of human is compare to higher than that of calculating cross ontology value (comparing BP&CC or MF&CC or MF&BP) is said to be extrinsic.

MOAL (Multi ontology data mining at all levels) algorithm for mines provides the cross ontology relationship between the ontologies. MOAL algorithm to mine cross-ontology association rules, i.e. rules that involve GO terms present in the three sub-ontologies of GO. By using collaborative filtering, user get the details about the gene id for cross ontology technique we have to compare the protein value and getting BP& MF value, or MF&CC value or CC&BP value getting the gene disease and symptoms for user requirements.

3) *Depth first Search:* Depth first search are used in relation to specific domains such as searching for solutions in artificial intelligence. The graph to be traversed is often either too large to visit in its entirety or infinite. In this cases search is performed to a depth due to limited resources. When DF search is performed to a limited depth the time is still linear in terms to number of expanded vertices. DF search also lends itself much better to heuristic methods for choosing a likely-looking branch. DFS is an iterative approach. The DFS implementations are as follows:

```
DFS(G, u)
u.visited = true
for each v ∈G.Adj[u]
ifv.visited == false
DFS(G,v)
init () {
For each u ∈ G
u.visited = false
For each u ∈ G
DFS(G, u)
}
```

4) *Regulatory Modules:* Regulatory modules are the sets of genes co-regulated to respond to different conditions. The possibility method is presented for identifying regulatory modules from gene expression data. The method saccharomyces cerevisiae expression data set ability coherent modules and their correct regulators.

In this gene ontology, an integrative framework modules from the cell cycle of cancer is used for incorporating multiple sources of biological data, molecular interaction. Among human being have 846 genes with putative roles in cell cycle regulation, here 46 transcription factors and 39 ontology groups.

5) *Integration Technique:* In this study, two approaches to the integration of mRNA, miRNA, and protein expression data is proposed, in order to identify cancer-related miRNAs and investigate relationships between miRNAs and the regulatory networks in cancer. A new computational method is presented for the ranking of cancer-related miRNAs based on the number of identified correlated genes, using both mRNA and protein datasets. valuate lists fabricate for each miRNA may advance our perceptive of the cancer-related miRNAs. Additionally, a method for the construction of modules containing mRNAs, miRNAs, and proteins is used. The modules were constructed based on the SAMBA bi-clustering algorithm and a Bayesian network model . Here mainly represent of correlated mRNAs, miRNAs, and also proteins.

6) *Multiplicative updating algorithm:* In this module, the optimization model function is presented effectively by iterative multiplicative updating algorithm. If “extra” information is given that one expert will be perfect find the best expert in long mistakes -multiplicative weights update rule says you’re not much worse than this algorithms are AHK,SDPs

7) *Tree Representation:* In this module, the tree consist of gene id as root element after that the leaf nodes contains its molecular function value, biological process value, cellular component values and also contains the symptoms ,diseases and curing possibilities of the related gene id. This tree representation is more useful to predict the details easily about the gene.

V. CONCLUSION

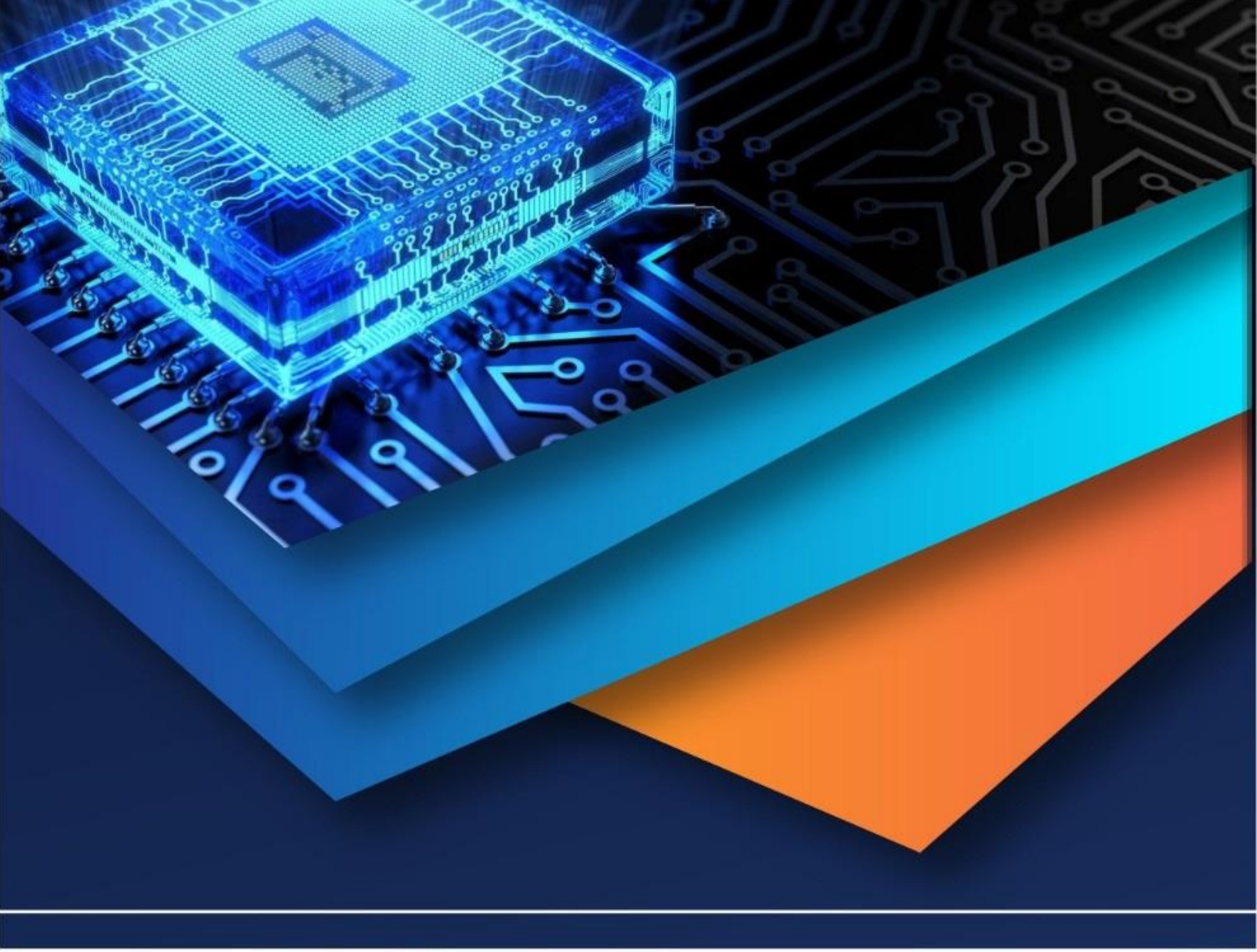
Relevant progresses in biotechnology and system biology are creating are remarkable amount of bio molecular data and semantic annotations; they increase in number and quality, but are dispersed and only partially connected. Integration and mining of these distributed and evolving data and information have the high potential of discovering hidden biomedical knowledge useful in understanding complex biological phenomena, normal or pathological, and ultimately of enhancing diagnosis, prognosis and treatment; but such integration poses huge challenges. The proposed system has tackled them by developing a novel and generalized way to define and easily maintain updated and extend an integration of many evolving and heterogeneous data sources; this approach proved useful to extract biomedical knowledge about complex biological processes and diseases.

VI. FUTURE WORK

Future work results found satisfactory in identifying and analyzing complex biological structures using graph clustering, collaborative filtering and depth first search. The intrinsic and extrinsic values are also calculated during gene ID analysis. Multiplicative algorithm is used for optimization. An objective function is initiated to after integrate the regulatory modules and gene ontology. After integration, optimal solution should be obtained. Hence to resolve the optimization concern, in this proposed system the multiplicative updating algorithm is used. Discuss the various data mining approaches to identify and provide diagnosis for breast cancer. This paper identifies the best predictor accuracy among all data mining techniques and states decision tree data mining technique provides 93.62% accuracy. Automatic decision tree based prediction with machine learning would be useful tool for medical research groups for predicting cancers. This paper also illustrates the scope of data mining technique in medical domain. Analyzing large volume of medical data, prediction of disease analyzing the historical data, association among the patterns would lead to huge demand for data mining techniques.

REFERENCES

[1] O. Hobert, “Gene regulation by transcription factors andmicrornas,” Science, vol. 319, no. 5871, pp. 1785–1786, 2008.
[2] L. He and G. J. Hannon, “Microrna: small rnas with a bigrole in gene regulation,” Nature Reviews Genetics, vol. 5,no. 7, pp. 522–531, 2004.
[3] J. LU, G. Getz, and E. A. Miska, “Microrna expressionprofiles classify human cancers,” nature, vol. 435, no. 9,pp. 834–838, Jun 2005.
[4] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, andN. M. Luscombe, “A census of human transcription factors: function, expression and evolution,” Nature ReviewsGenetics, vol. 10, no. 4, pp. 252–263, 2009.
[5] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph,G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, and I.Simon,” “Transcriptional regulatory networks insaccharomyces cerevisiae”, science, vol. 298, no. 5594, pp.799–804, 2002.
[6] R. Shalgi, D. Lieber, M. Oren, and Y. Pilpel, “Global and local architecture of the mammalian micrornatranscriptionfactor regulatory network,” PLoSComputBiol, vol. 3, no. 7, pp. 131, 2007.
[7] Y. Zhou, J. Ferguson, J. T. Chang, and Y. Kluger, “Inter- andintra- combinatorial regulation by transcription factors and microrna,” BMC genomics, vol. 8, no. 1, p. 1, 2007.
[8] C.-Y. Chen, S.-T.Chen, C.-S.Fuh, H.-F.Juan, and H.-C.Huang, “Coregulation of transcription factors and microrna in human transcriptional regulatory network,” BMCbioinformatics, vol. 12, no. 1, p. 1, 2011.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY