ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

IMUA approach to Identify mRNA's and Transcription Factor Modules of Genomic Data

[1] P.Deepa, [2] R.Baghyalakshmi, [3] G.Hemantha Sree, [4] J.Jayasankari
[1] Associate Professor, [2 3 4] UG Scholars
Department of Computer Science & Engineering
Panimalar Engineering College, Chennai-600 l23,India

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 existing work we extract weighted Association Rules from Ontology Based annotated datasets by using GO- W AR(Gene Ontology-based Weighted Association Rules).We here adapt the MOAL algorithm to mine cross-ontology association rules, i.e. rules that involve GO terms present in the three sub-ontologies of GO. We are proposing cross ontology to manipulate the Protein values from three sub ontologies for identifying the gene attacked disease. Also our 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, We done the Co-Regulatory modules between miRNA (microRNA), TF(Transcription Factor) and gene on functionlevel with multiple genomic data.. We compare the regulations between miRNA-TF interaction, TF-gene interactions and gene-miRNA interaction with the help of integration technique. These interaction could be taken the genetic disease like breast cancer, etc.. Iterative Multiplicative Updating Algorithm is used in our paper to solve the optimization module function for the above interactions. After that interactions, we compare the regulatory modules and protein value for gene and generate Bayesian rose tree for efficiency of our result.**

Keywords**-Gene Ontology-based Weighted Association Rules, Co-Regulatory modules, Iterative Multiplicative Updating Algorithm, Bayesian Rose Tree.**

I. INTRODUCTION

Ontologies are determining of a relational concordance. Molecular functions in a species-independent manner. In ontology , it contains three domains biological process for encrust, event with its defined beginning and end by molecular function, cells, tissues, organs & organisms are integrated living units by using cellular component. Cellular components are the complex bio molecules and structures of which cells, and thus living organisms are composed. Cells are the structural and functional units of life. Single cells are considered as smallest organisms, while trillions of cells are said to be a largest organisms. 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. Gene Ontology includes three mam sub-ontologies: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). Each ontology stores and organizes biological Concepts, called GO Terms, used for describing functions, processes and localization of biological molecules. 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 sequence polymorphisms,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

expression in different tissues, or involvement in biological processes, biochemical pathways and genetic disorders).

II. RELATED WORK

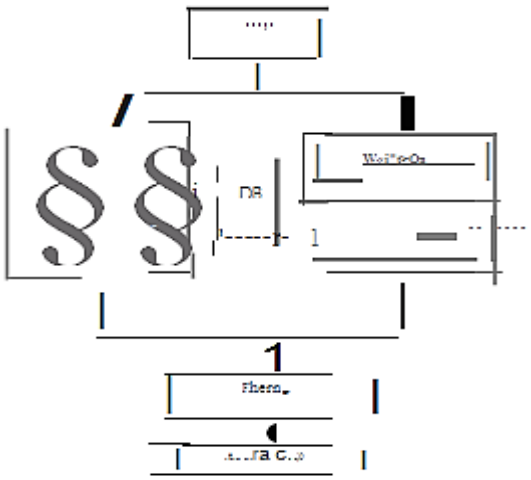
Biological matter and biological material are otherwise called as Cellular Component. Soft matter one of the characteristics in biological matter. In biological matter made by everyone life such as Carbon-based, Organic Biological or even simple living things. These definitions of life excluded in hypothetical types of biochemistry. The processes of vital information in living organism to live by Biological Processes. These process are involved in persistence and transformation of life forms and are also made up of many Chemical reaction. The two examples of Biological process are Metabolism and homeostasis. Frequency, rate or extent modulated with the help of the regulation of biological process. These Processes are regulated by the control of gene expression, protein modification or interaction with protein & Molecules.

III. EXISTING SYSTEM

The ex1stmg 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 mmmg algorithms. In this base paper we used two techniques i.e. is Cross Ontology and Integration, in Ontology we existed patient gene id's 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 our Paper, we 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. We get 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 we generate a Bayesians rose tree structure for the relation between regulatory modules and protein values of our gene. By this structure we know our disease which was affected in our chromosome and also know how to cure? What are the symptoms are applicable for our gene by our web application.



V. TECHNIQUES AND DESCRIPTION

5.1 Gene Ontology

It classifies functions along three aspects: molecular function molecular activities of gene products, cellular component where gene products are active, biological process pathways and larger processes made up of the activities of multiple gene products. The Gene Ontology (GO) paper is a cooperate effort to situation the need for consistent elucidation of gene products in different databases. In our paper we are proposing gene ontology , User login and register their details and get the gene id from Ontology base with the help of KNN algorithm. Full details of overall paper are maintained our database and ontology base. We are proposing cross ontology to manipulate the Protein values from three sub ontologies for identifying the gene attacked disease. Also our proposed system, focus on intrinsic and extrinsic. Based on cellular component, molecular function and biological process values intrinsic and extrinsic calculation would be manipulated.

5.2 Collaborative Filtering:

In our Paper we used semantic mining 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 we get the BP,CC&MF value for gene to identify the gene have Intrinsic or extrinsic.

5.2.1 Intrinsic

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

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

5.2.2 Extrinsic

If the normal protein value of human is compare to higher than that of calculating cross ontology value (comparing BP&CC or JI.1F&CC or JI.1F&BP) is said to be extrinsic. MOAL (Multi ontology data mining at all levels) algorithm for mines 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& JI.1F value, or JI.1F&CC value or CC&BP value getting the gene disease and symptoms for user requirements.

5.3 Regtdatory modules

Regulatory modules: sets of genes co•regulated to respond to different conditions. We present a possibility method for identifying regulatory modules from gene expression data. Our gimmick describe modules of co-regulated genes, their regulators and the conditions under which regulation occurs, generating testable hypotheses in the form 'regulator X regulates module Y under conditions W'. We applied the method to a Saccharomyces cerevisiae expression data set, showing its ability to identify functionally coherent modules and their correct regulators. Three novel predictions, suggesting regulatory roles for previously uncharacterized proteins. We propose an integrative framework that infers gene regulatory modules from the cell cycle of cancer cells by incorporating multiple sources of biological data, including gene expression profiles, gene ontology, and molecular interaction. Among 846 human genes with putative roles in cell cycle regulation, we identified 46 transcription factors and 39 gene ontology groups. We reconstructed regulatory modules to infer the underlying regulatory relationships. Four regulatory network motifs were identified from the interaction network. The relationship between each transcription factor and predicted target gene groups was examined by training a recurrent neural network whose topology m1m1cs the network motif(s) to which the transcription factor was assigned. Inferred network motifs related to eight well-known cell cycle genes were confirmed by gene set enrichment analysis, binding site enrichment analysis, and comparison with previously published experimental results.

5.4 Integration Technique

In this study, we propose two approaches to the integration of mRNA, miRNA, and protein expression data, in order to identify cancer-related miRNAs and investigate relationships between miRNAs and the regulatory networks in cancer. We present a new computational method 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, we present a method for the construction of modules containing mRNAs, miRNAs, and proteins. The modules were constructed based on the SAMBA hi-clustering algorithm and a Bayesian network model . To construct these modules, we extended the approach proposed by Jin and Lee by adding a step in which the proteins are included into mRNA- sample modules prior to the inclusion of miRNAs. The identified modules represent subgroups of highly correlated mRNAs, miRNAs, and proteins, and may explain regulatory networks between miRNAs and genes.

5.5 Multiplicative updating algorithm

In this module, we solve the optimization model function effectively by iterative multiplicative updating algorithm. If "extra" information is given that one expert will be perfect find the best expert in login mistakes -multiplicative weights update rule says you're not much worse than this synopsis , in more ordinary cases applications to algorithms • AHK - generalize the losses to matrix losses to solve SDPs- 0 n 2 time • KRV - "cut- matching" game to solve sparsest cuts. 2 players: a cut player, and a matching player.

5.6 Tree Representation

Binary tree representation of trees. (data structure) Definition: A way to represent a multiway tree as a binary tree. 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.

VI. CONCLUSION AND 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 we propose multiplicative updating algorithm discuss the various data mining approaches to identify and provide .

REFERENCES

[1] 0. Hobert, "Gene regulation by transcription factors and micrornas," Science, vol. 319, no. 5871, pp. 1785-1786, 2008.

[2] L. He and G. J. Hannon, "Micrornas: small rnas with a big role in gene regulation," Nature Reviews Genetics, vol. 5, no. 7,pp. 522-531,2004.

[3] J. LU, G. Getz, and E. A. Miska, "Microrna expression profiles classify human cancers," nature, vol. 435, no. 9, pp. 834-838, Jun 2005.

[4] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe, "A census of human transcription factors: function, expression and evolution," Nature Reviews Genetics, 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. Thomp- son, and I. Simon, "Transcriptional regulatory networks in Saccharomyces 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 mammalian micro rna transcription factor regulatory network," PLoS Comput Bioi, vol. 3, no. 7, p. el31, 2007.

[7] Y. Zhou, J. Ferguson, J. T. Chang, and Y.Kluger,"Inter-andintra-combinatorial regulation by transcription factors and micrornas," 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, "Co-regulation of transcnptwn factors and micrornas in human transcriptional regulatory network," BMC bioinformatics, vol. 12, no. 1, p. 1, 2011.