International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 25 Issue 5 – APRIL 2018 (SPECIAL ISSUE).

DISCOVERYOFMICRORNASAND TRANSCRIPTIONFACTORSCO-REGULATORYMODULESBYINTEGRATIN GMULTIPLETYPESOFGENOMICDATA

S.Aruna[1],X.JebaAncySindhuja[2] , J.Banupriya[3] andK.G.Saravanan[4]

[1,2,3]*U.G. Student, Department of Computer Science and Engineering, Kings Engineering College, Chennai.*
[4]*Assistant Professor, Department of Computer Science and Engineering, Kings Engineering College, Chennai*
.

***ABSTRACT*-Gene Ontology (GO) is a structured repository of concepts 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 used GO-WAR (Gene Ontology-based Weighted Association Rules) for extracting Weighted Association Rules from ontology-based annotated datasets. 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 biological process values intrinsic and extrinsic calculation would be manipulated.**

In this Project, We done the Co-Regulatory modules between miRNA (microRNA), TF (Transcription Factor) and gene on function level 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 project 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.

I. INTRODUCTION

Ontologies are specifications of a relational vocabulary. Gene ontology (GO) is a major bioinformatics initiative to unify the representation of gene and gene product attributes across species.

The GO project has developed three structured controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. The ontology covers three domains: cellular component, the parts of a cell or its extracellular environment; molecular function, the elemental activities of a gene product at the molecular level, such as binding or catalysis; and biological process, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms.

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 used to vocabularies and ontologies have been 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 main 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 represents the apoptosis process.

Increasingly large amounts of valuable, but heterogeneous and sparse, biomolecular data and information are characterizing life sciences [1]. In particular, semantic controlled annotations of biomolecular entities, i.e. the associations between biomolecular entities (mainly genes and their protein

products) and controlled terms that describe the biomolecular 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 polymorphisms, expression in different t issues, or involvement in biological processes,biochemical pathways and genetic disorders).

These semantic annotations can effectively support the interpretation of genomics and proteomics test results a nd the extraction of biomolecular information, which can be used to formulate and validate biological hypotheses and possibly discover new biomedical knowledge.

1.Gene Ontology

Gene Ontology is the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects: molecular function molecular activities of gene products, cellular component wheregene products are active, biological process pathways and larger processes made up of the activities of multiple gene products. The Gene Ontology (GO) project is a collaborative effort to address the need for consistent descriptions of gene products in different databases In our project 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 project 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.

2.Collaborative Filtering

In our Project 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 ofhuman 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.

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.

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 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.

1.Depth first search

Depth first search in relation to specific domains such as searching forsolutions 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.Such as memory or disk space one typically does not use data structures to keep track the set of all previous visited vertices.

When DF search is performed to a limited depth the time is still linear in terms to number of expanded vertices. Edges although this number is not the same as the capacity of the entire graph because some vertices may be searched more than once and others not at all but the space complexity of this variant of DFS is only proportional to the depth limit. As a result is much smaller than the space needed for searching to the same depth using BFS. DF search also lends itself much better to heuristic methods for choosing a likely-looking branch.

1.Regulatory modules

Much of a cell's activity is organized as a network of interacting modules: sets of genes co-regulated to respond to different conditions. We present a probabilistic method for identifying regulatory modules from gene expression data. Our procedureidentifies 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. We present microarray experiments supporting 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 wasexamined by training a recurrent neural network whose topology mimics 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.

CONCLUSION

Relevant progresses in biotechnology and system biology are creating a remarkable amount of biomolecular data and semantic annotations; they increase in number and quality, but are dispersed and only partially connected. Integration and mining of these distributed and eolving data and information have the high potential of discovering hidden bio medical knowledge useful in understanding complex biological phenomena, normal or pathological, and ultimately of enhancing diagnosis, prognosis and treatment; but such integration poses hug e challenges.

Our work has tackled the m by developing a novel and generalized way to define and easily maintain up dated and extend an integration of man y evolving and heterogeneous data sour ces; our approach proved useful to extract biomedical knowledge about complex biological processes and diseases.

REFERENCES

[1] O. 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 fac- tors: function, expression and evolution,” Nature Reviews Genetics, vol. 10, no. 4, pp. 252–263, 2009.