ISSN (Online): 2456-0448
International Journal of Innovative Research in Management, Engineering and Technology
Vol. 3, Issue 3, March 2018

Portable And Inexpensive Hemoglobin Testing Device Based On The Beer Lambert's Law

Proposed methodology and a comparative study on the basis on portability and cost

[1] K. Saranya, [2] P. Indira Priya
[1] [2] Department Of CSE, MNM Jain Engineering College, Chennai

Abstract: ***Gene ontology (GO) is a repository that is used to define the gene with predefined set of diseases in a genomic set of data. The repository contains list of gene with unique id to identify the disease. Each process in gene is referred to as annotation. The analysis used in the gene is Association Rules (AR) which discovers biologically data. In existing work we used gene ontology weighted association rule which is used to find the frequent sets of gene with weighted sets. We are proposing mine cross ontology to find the Protein from the average sets of protein identified from the dataset. Our proposed system focus on cross ontology to find the disease based on the values obtained and the regulatory modules contains miRna, transcription factor and gene are used to identify the disease based on particular gene. Depth first search is used in our proposed system to identify the symptoms and precautions at depth level. Finally Bayesian rose tree analysis will produce the result in hierarchical structure.***

Keywords: ***Gene Ontology, genomic data, microRNA(miRNA), regulatory module, transcription factor, regulatory module.***

I. INTRODUCTION

Data Mining or knowledge discovery is the computer- assisted process of data through the analysing and enormous sets of data and then extracting the meaning of the data. Data mining tools predict behaviour and future trends, allowing business to make proactive, knowledge-driven decisions. Data mining tools can answer business databases for hidden patterns. Data mining derives its name from similarities between searching for valuable information in a large database and mining mountain for a vein of valuable one. Gene Ontology is used for interpreting sets of genes making use of system classification, in which genes are assigned to a set of predefined bins depending on their functional characteristics. The Gene Ontology (GO) provides a system for hierarchically classifying genes or gene products to terms organized in a graph structure called ontology .Ontology describes sets of genes to provide symptoms and precautions for particular gene. The cross ontology technique is used to make the comparison between the proteins and produce the result based on the filtering. Regulatory modules are used to make the interaction between the miRNA, TFs and gene to produce the analysis for particular gene in hierarchical structure. This system is used to identify the disease based on miRNA and transcription factor in a particular gene. It also identifies symptoms and precaution at depth level.

II. RELATED WORK

Gene Ontology (GO) is a controlled vocabulary of concepts structured on three main ontologies. Each GO Term contains a description of a biological concept that is associated to one or more gene products through a process also known as annotation. The importance and the specificity of both GO terms and annotations are often measured by their Information Content (IC). Mining annotations and annotated data may extract meaningful knowledge from a biological stand point. An analysis of these annotated data using association rules provides evidence for the co-occurrence of annotations [1]. The methodology is used for extracting Weighted Association Rules from GO implemented in a tool named GOWAR. It is able to extract association rules with a high level of IC without loss of Support and Confidence from a dataset of annotated data.

Protein-protein interactions (PPIs) play a crucial part in cell functions. Interactions of these proteins will result in different biochemical processes that govern living cells. Gene code for proteins will establishing relationships between protein interactions and their related ontological definition would help predict unknown PPIs[8]. A novel method to

predict new PPIs using information from proteins annotated in Gene Ontology. Our work can also predict Gene Ontology based on the association rules and the semantic similarity of Gene Ontology annotated to the proteins. This method used to predict new PPIs using shrimp protein-
protein interactions data and found that Gene Ontology and interactions between proteins can be used as a key to predict unknown protein-protein interactions. Detecting protein interactions by means of biology experiments are costly and very time consuming. It is well known that regulators known as microRNA (miRNA) and transcription factor (TF) have been found to play an important role in gene regulation. However, there are few researches of collaborative regulatory (co-regulatory) mechanism between miRNA and TF on system level (function level). Meanwhile, recent advances in high-throughput genomic technologies have enabled researchers to collect diverse large-scale genomic data, which can be used to study the co-regulatory mechanism between miRNA and TF[3]. A novel method called Sparse Network regularized non-negative matrix factorization for co- regulatory modules identification which adopts multiple non- negative matrix factorization frameworks to identify co- regulatory modules including miRNAs, TFs and genes. We apply this method to multiple genomic data including the expression profiles of miRNAs, TFs and genes on breast cancer obtained from TCGA, priori miRNA-gene regulations, TF-gene regulations and gene-gene interactions. Gene Ontology (GO) is a structured repository of concepts (GO Terms) that are associated to one or more gene products through a process referred to as annotation. The analysis of annotated data is an important opportunity for bioinformatics[7]. There are different approaches of analysis, among those, the use of association rules (AR) which provides useful knowledge, discovering biologically relevant associations between terms of GO, not previously known. In a previous work, we introduced GO-WAR (Gene Ontology- based Weighted Association Rules), a methodology for extracting weighted association rules from ontology-based annotated datasets. We here adapt the GO-WAR algorithm to mine cross-ontology association rules, i.e., rules that involve GO terms present in the three sub-ontologies of GO. We conduct a deep performance evaluation of GO-WAR by mining publicly available GO annotated datasets, showing how GO-WAR outperforms current state of the art approaches. MicroRNA present a systems biology approach to the understanding of the miRNA-regulatory network in colon rectal cancer[5]. An initial set of significant genes in Colon Rectal Cancer (CRC) were obtained by mining relevant literature. An initial set of cancer-related miRNAs were obtained from three databases: miRBase, miRWalk, Targetscan and GEO microarray experiment. First principle
methods were then used to generate the global miRNA-gene network. Significant miRNAs and associated transcription factors in the global miRNA-gene network were identified using topological and sub-graph analyses. Eleven novel miRNAs were identified and three of the novel miRNAs, hsa-miR-630, hsa-miR-100 and hsa-miR-99a, were further analysed to elucidate their role in CRC. The proposed methodology effectively made use of literature data and was able to show novel, significant miRNA-transcription associations in CRC.

III. PROPOSED SYSTEM

i. Gene ontology is used for structure repository used to get list of unique gene.
ii. Collaborative filtering is used to filter the disease based on cross ontology base.
iii. Regulatory modules are used to make comparison
between the mi-rna, tf and gene.
iv. Depth first search is used to search the symptoms and precautions.
v. Integration technique is used to combine result
based on clustering.
vi. Bayesian rose tree analysis is used to produce the result in tree structure.

A. Architecture

The proposed system user collects the protein value from sample and search for particular disease specified. Collaborative filtering technique is used to filter the disease based on the cross ontology by collecting the molecular function, cellular

component, and biological process. Regulatory modules contain miRNA, TF and GENE that are used to identify the disease based on clustering. Depth first search algorithm is used to find the symptoms and precautions for particular gene. Finally Bayesian rose tree analysis is used to produce the result in hierarchical structure.

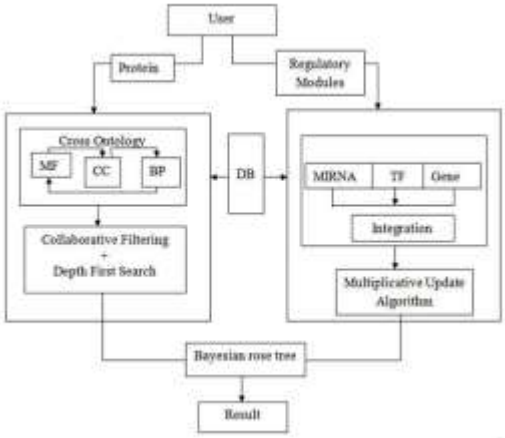

Fig. 1. Architecture Diagram.

A. Gene Ontology

Gene ontology (GO) is a repository that is used to define the gene with predefined set of diseases in a genomic set of data. The repository contains list of gene with unique id to identify the disease. Each process in gene is referred to as annotation. The analysis used in the gene is Association Rules (AR) which discovers biologically data. The gene ontology is grouped into three categories: molecular function, biological process and cellular component. For the particular gene id list of disease is listed and associated genes are specified.

B. Collaborative Filtering

The semantic mining for logical analysis is used to get the details from Ontology base with help of collaborative filtering. The cross ontology technique is an association rule mining to connect the goal with information content. Collaborative filtering is used to filter the disease based on the result evaluated from sub-ontology. The list of disease is listed from gene ontology. Two ontology is in a range the gene will be normal. Otherwise the normal values of protein are added and compare to lower than calculated ontology intrinsic type of disease is listed. If it is higher than calculated value extrinsic type of disease is listed.

C. *Depth First Search*

Depth first search is used for specific domains such as searching for solutions in artificial intelligence. The search is performed to a depth due to limited resources and displays the disease priority with high to low. DFS search also lends itself much better to heuristic methods for choosing a likely- looking branch. The algorithm is used to find the disease with symptoms and precautions.

D. *Regulatory Modules*

Cis-regulatory modules carry out their function by integrating the active transcription factors and the associated co-factors at a specific time and place in the cell where this information is read and an output is given. Gene regulatory modules that consist of miRNAs, TFs and their target genes based on the available predicted target. Clustering analysis or clustering is grouping of disease for particular gene that belongs to same group.

WORKING PRINCIPLE

a) The protein from gene list are collected and search for particular disease specified.

b) The gene with list of associated genes is displayed.

c) Cross ontology technique is used for comparing between each sub-ontology.

d) Collaborative filtering technique is used to filter the gene attacked diseases by collective the molecular function, cellular component, and biological process.

e) Regulatory modules are used to find the mi-rna, tf and gene to identify the gene attacked disease.

f) Multiplicative update algorithm is used to identify the disease

g) Depth first search algorithm is used to find the symptoms and precautions for the gene.

h) The final result is displayed in Bayesian rose tree analysis.

E. *Integration Technique*

Identify potential interactions between mi-RNAs TFs and Genes and pathways involved in development to identify the gene attacked diseases. The comparison is made between the mi-RNA TF and Gene and the result is produced by using cluster.

F. *Tree Representation*

Bayesian hierarchical clustering algorithm will produce trees with arbitrary branching structure at each node known as rose trees. Clustering is the task of grouping a set of objects in such a way that objects in the same group. It is often necessary to modify data pre-processing and model parameters until the result achieves the desired output.

IV. RESULT AND DISCUSSION

Result is proposed by using gene enrichment analysis. List of disease is identified from particular gene. The gene id is taken form list of gene list form gene ontology. The gene ontology covers cellular component, biological process, and molecular function.

Fig. 2. List of disease for particular gene.

Based on molecular function, biological process and cellular component the value is calculated. If the value is lower than the specific range defined intrinsic otherwise it will be extrinsic.

Fig. 3. Disease Identification.

Graphical representation of the disease will be done by using Bayesian rose tree hierarchical clustering. The tree is specified with ontology based on intrinsic or extrinsic and symptoms and precautions are analyzed based on gene.

Fig. 4. Tree Repersentation.

Conclusion

Relevant gene in biotechnology and system biology is creating a remarkable amount of bimolecular data and semantic annotations. The work has tackled them by developing a novel and generalized way to define and easily maintained updated and extend an integration of many evolving and heterogeneous data sources. The system is useful to extract biomedical knowledge about complex biological processes and diseases.

References

[1] D. Faria, A. Schlicker, C. Pesquita, H. Bastos, A. E. N. Ferreira, M. Albrecht, and A. O. Falco, "Mining go annotations for improving annotation consistency", 2012.

[2] P. Manda, F. McCarthy, and S. M. Bridges, "Interestingness measures and strategies for mining multi-ontology multi-level association rules from gene ontology annotations for the discovery of new {GO} relationships". J. Biomed. Informat, 2013.

[3] J Luo, Jiawei, Gen Xiang, and Chu Pan. "Discovery of microRNAs and Transcription Factors Co-Regulatory Modules by Integrating Multiple Types of Genomic Data" IEEE transactions on nanobioscience 16.1 (2017): 51-59.

[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] 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.

[6] M. Cannataro, P. H. Guzzi, and A. Sarica, "Data mining and-life sciences applications on the grid," Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery, vol. 3, no. 3, pp. 216–238, 2013.

[7] O. Hobert, "Gene regulation by transcription factors and micrornas," Science, vol. 319, no. 5871, pp. 1785–1786, 2008.

[8] Vaiwsri S, Prachumwat A, Ngamsuriyaroj S, Srisuphab A. Predicting shrimp protein-protein interactions and gene ontology terms using association rule and semantic similarity calculation. InComputer Science and Engineering Conference (ICSEC), 2014 International 2014 Jul 30 (pp. 278-283). IEEE.