

A Survey on Identification of Motifs and Ontology in Medical Database

B. Lavanya*, T. Madhumitha

Department of Computer Science, University of Madras, Chennai, India.

*Email: lavanmu@gmail.com

ARTICLE INFORMATION

DOI: 10.15415/jotitt.2018.61003

Keywords: Motifs, Gene Network, Ontology Classification, Disease diagnosis, Data Mining

ABSTRACT

Motifs and ontology are used in medical database for identifying and diagnose of the disease. A motif is a pattern network used for analysis of the disease. It also identifies the pattern of the signal. Based on the motifs the disease can be predicted, classified and diagnosed. Ontology is knowledge based representation, and it is used as a user interface to diagnose the disease. Ontology is also used by medical expert to diagnose and analyse the disease easily. Gene ontology is used to express the gene of the disease

1. Introduction

The motifs are identified in the medical database, which is used for diagnose the disease. These diseases are caused due to loss of cells. The network motif is sub graph which is used as biological network. The network motif can also identify various diseases like coronary artery disease (CAD), micro biota related disease, like pneumonia and dental caries. The clustering analysis of CAD is made based on gene expression. The continuous glucose level for diabetes patient are analysed through contextual motifs. From the motif in the spikes, the level of hypo and hyper glycaemia event can also be identified.

Ontology is used in medical database for

disease diagnose which is web based application. Gene ontology annotation identifies the gene for clustering; the interrelation of gene is identified by ontology. Ontology will observe and explain the disease and are used for medical diagnostic system.

The mechanism of disease is important in bio-medical research and the disease can also be identified by the network motifs. The network is analysed by the interaction and association between the networks. The string representation in the ECG signal called motifs, detects normal and abnormal heart beat. SVM is used to classify the disease from the sequential outcomes of motifs. The motif mining tool is used to create the network motifs.

2. Literature survey

Rohit Gupta *et al* (2016), [1], proposed to identify the motifs to diagnose the cancer disease. This is done based on the cancer pathway and signal transduction network method. In these two methods, five algorithms are used which identify the gene network motif. The algorithms namely Auto Regulation Loop (ARL), Feed Forward Loop (FFL), Feed Backward Loop (FBL), Signal Input Motifs (SIM) and it uses the adjacency matrix, Bi-fan. The Topological behaviour of the network is identified by FANMOD. The drawback of FANMOD is only three node motifs are used to overcome using an adjacency matrix. Out of the five algorithms SIM and Bi-fan network is mostly used to identify the gene network motif. The oncologists can easily diagnose the cancer disease based on the network motifs.

Yin Li *et al* (2016), [2], describes the network motif for coronary artery disease. Differential integrated gene and protein-protein interaction gene are analyzed to interaction pattern is identified by screening of differential network. The network is to find the top 20 network, which is used to identify the coronary artery disease. For screening the network the R package global ancova software where used. The main advantage of screened network motif is, to give the accurate result to identify the coronary artery disease. This network motif method gives the accurate result.

Yin Wang *et. al.* (2016), [3], classify the disease based on microbial metagenome. These classifications are done by the method Phylogenetic tree based motif finding algorithm (PMF). The PMF algorithm has three parts that is motif finding, motif sorting and model

evaluations. This PMF classifies two diseases, pneumonia and dental caries based on the microbial meta-genome. The main advantage of using PMF is to find the motifs in the training data, from which disease is classified.

Ian Fox *et. al.* (2016), [4], in this system, contextual motifs for Continuous

Glucose Monitor (CGM) of type1 diabetes are identified. CGM is done based on the spikes in waves, from which glucose level is monitored. The two stage discovery method with data driven contextual motifs and join discovery method with generating contextual motifs are used. The first method is expert drive method, used by the domain expert and data driven is an unsupervised method. Join discovery method is used to identify, how long the patient suffered from diabetes by long term or short term event, and it uses AUC curve. The main advantage based on contextual motifs is, to identify hypo and hyper glyceic event.

Sakorn Mekruksavanich (2016), [5], discuss to diagnose the disease based on the diabetes ontology. This diagnose is done through the Medical Expert System. Ontology is done by sub-classes and in those subclasses the relationship is analyzed. The fuzzy logic is appeared for diagnose the disease by disease risk factor. The weighted similarity algorithm is used to show the result of the diagnosis. The proposed system is used for diabetes diagnosis. And the main advantage is, this is a web based application. So, the user can easily interface with this application and to get the best advice for the disease diagnosis.

Charles C.N. Wang *et. al.* (2016), [6], this proposed system, is used to cluster the cardiovascular disease (CAD) gene. Gene

ontology and Latent semantic analysis (LSA) method is used to analyze the gene for CAD. And the network pathway is created by clustering of genes and drug target is used for pathways. Jiawei Lu et al (2016), [7], this model identifies the functional similarity between miRNA and the similarity between the gene is identifies by the method MFS_GO, this method also identify the comparison between the miRNA. MFS_GO method is applied to identify the cancer that is breast, ovarian, colorectal neoplasm. The advantage of this method is to predict novel disease by miRNA using MFS_GO method.

S.Padmavathi *et. al.* (2015), [8], use the method Multivariant maximal time series motifs to identify the frequently occurring patterns and then it uses a Naive bases classifier to classify the normal and abnormalities signal, the accuracy is 93.33% and 98% of precision rate. This method is used in the application of Electrocardiogram (ECG) to classify the abnormality in ECG signals.

Firas Zekri *et. al.* (2016), [9], which identifies and diagnose Alzheimer disease (AD) based on the method AlzFuzzyOnt which relates to AD concept and mind ontology. Fuzzy logic utilities are used in AlzFuzzyOnt. The Hamilton scale will specify the mild depressive symptoms, moderate depressive symptoms, and severe depressive symptoms. And this system will suggest the domain specialist to diagnose disease.

Asma Khan *et. al.* (2017), [10], deals with the application of Hepatitis Disease by the method Hepatitis Ontology HEPO. The medical experts can use this ontology to diagnose the patient with hepatitis. Then the method called abdlicative reasoning in ontology is used for disease diagnosis. The

ontology in medical diagnose system will give a better idea for domain specialist to diagnose the disease.

Mario Cannataro *et. al.* (2015), [11], ranks the Gene based on disease. In the Gene Ranking Based on disease (GOD) algorithm, the prioritization method is used; gene priority method is used in ontology for post-processing step. The main advantage of this gene prioritization is legible in gene annotation with respect to disease. The main advantage of GOD algorithm is to prioritize the gene and it uses different types of annotation.

Duc-Hau Le *et. al.* (2016), [12], in this the network motifs is used for disease prediction. The Random walk restart on heterogeneous network (RWRH) algorithm is used in network motifs, which identify the similarity of network for Alzheimer's disease based on the network it gives the better functionality among the disease. Ontology is used to predict the network similarity.

Iyanuoluwa Emmanuel *et. al.* (2015), [13], the diabetes application is done by the method Diabetes based Biological Network (DBN) and network motifs are used which identify the diabetes and this disease by using the method called perturbation which is used to filter the high influence node by performing perturbation and the preliminary result will be obtained. Shameek Ghosh *et. al.* (2016), [14], which deals to detect the critical patient events like hypotension and septic shock based on the method, order sequential contrast pattern based classification in the time series sequence for detecting patient event. SVM and HMM is used to classify the disease and this use the arterial pressure series. And this will give the better prediction in ICU

outcomes which is the application of this system.

K. Shi *et. al.* (2016), [15], used the method called network motifs, the centrality for analysing the shortest path between the nodes. The highest the centrality scores the more significant motifs. This is the application based on colorectal cancer disease. The pathway in the disease, it is a significant pathway which enriches the gene reported related to cancer development.

Giuseppe Agapito *et. al.* (2016), [16], describe the method Gene Ontology Based Weighted Association Rule (GO-WAR) for ontology based annotation dataset. The GO-WAR use the Weighted Support and Confident, from which the analysis of research is better. The Performance is measured by memory consumption and execution time. This process is used as software tool to analyse the gene ontology.

Adnan Ferdous Ashrafi *et. al.* (2015), [17], which will find the motifs in DNA sequence by Integer Matching using Hash table indexing, and rank the motifs then calculate the fitness in DNA sequence. The main advantage is the DNA sequence will be accurate and effective.

J. Sivaranjani *et. al.* (2017), [18], which describes the motifs in medical ECG data and the performance of motif measured by F- Measures. And this motif is discovering time series method. This application uses the Hadoop environment to discover the motifs in ECG data.

3. Conclusion

From these types of method proposed, the clearly identified motifs and ontology is very helpful in medical database for identification of the disease. It helps to monitor the disease accurately and easily.

From this, the domain expert can easily use the application to the disease. The motifs are used to monitor the glucose for type1 diabetes patient and the network motif used the classifier to classify the disease. Web based application will give the solution for the patient with the disease and the ontology is mainly used for the knowledge based disease prediction and it can used even by domain experts. In future work, the motifs can identified in microarray database to classify the disease by using data mining techniques.

References

- [1] R. Gupta, S.M Fayaz and S. Singh "Identification of gene network motifs for cancer disease diagnosis", *IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, Mangalore, India, pp. 13-14, 2016.
- [2] Y. Li, Y. Cong and Y. Zhao, "Network motif-based for identifying coronary artery disease", *Experimental and Therapeutic Medicine*, vol. 12, no.1, pp. 257-261, 2016.
- [3] Y. Wang, *et. al.*, "Motif-based text mining of microbial metagenome redundancy profiling data for disease classification", *BioMed Research International*, 2016.
- [4] I. Fox, *et. al.*, "Contextual motifs- increasing the utility of motifs using contextual data", in *KDD '17 Proceeding of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax, NS, Canada – August 13 - 17, pp. 155 – 164, 2017.
- [5] S. Mekruksavanich, "Medical expert system based ontology for diabetes disease diagnosis", *IEEE 7th IEEE*

- International, Conference on Software Engineering and Service Science (ICESS)*, Beijing, China, pp. 26-28, Aug. 2016.
- [6] C. C. N. Wang, Y. Lee, P. C. Y. Sheu, J. J. P. Tsai, "Application of Latent Semantic Analysis to Clustering Cardiovascular Gene Ontology", *IEEE 16th International Conference on Bioinformatics and Bioengineering*, Taichung, Taiwan, pp. 363-368, 31 Oct. -2 Nov. 2016.
- [7] J. Lu, D. Dai, B. Cao, Y. Yin, "Inferring human miRNA functional similarity based on gene ontology annotation", *IEEE 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* Changsha, China, 13 - 15 Aug. 2016.
- [8] S. Padmavathi, E. Ramanujam, "Naïve bayes classifier for ecg abnormalities using multivariate maximal time series motif", *Elsevier Procedia Computer Science*, vol. 47, pp. 222 - 228, 2015.
- [9] Firas Zekri, Rafik Bouaziz, Emna Turki, Istanbul, "A fuzzy-based ontology for alzheimer's disease decision support", *IEEE International Conference on Fuzzy System (FUZZIEEE)*, Turkey, 2-5 Aug. 2015.
- [10] A. Khan, *et. al.*, "HEPO: The hepatitis ontology for abductive medical diagnostic systems", *IEEE International Conference on Communication, Computing and Digital Systems (C-CODE)*, Islamabad, Pakistan, 8-9 March 2017.
- [11] M. Cannataro, P. H. Guzzi and M. Milano, "GoD: An R-package based on ontologies for prioritization of genes with respect to diseases", *Journal of Computational Science*, Elsevier, pp. 7-13, 2015.
- [12] D. Le, V. Dang and S. B. Heidelberg, "Ontology-based disease similarity network for disease gene prediction", *Vietnam Journal of Computer science*, vol. 3, no. 3, 2016.
- [13] I. E. Obebode and A. Gangopadhyay, "Acquisition of diabetes-related biological associations using a motif based network: preliminary results", *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Washington, DC, USA, pp. 9-12, Nov. 2015.
- [14] S. Ghosh, H. Nguyen and J. Li, "Predicting short-term ICU outcomes using a sequential contrast motif based classification framework", *IEEE 38th Annual International Conference of the IEEE Engineering in Medical and Biology Society (EMBC)*, Orlando, FL, USA, 16-20 Aug. 2016.
- [15] K. Shi, L. Gao and B. Wang, "Systematic tracking of coordinate differential network motifs identifies novel disease-related genes by integrating multiple data", *Neurocomputing*, vol. 206, no. c, pp. 3-12, 2016.
- [16] G. Agapito, M. Cannataro, S. P. H. Guzzi and M. Milano, "Extracting Cross-Ontology Weighted Association Rules from Gene Ontology Annotations", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Mar-Apr 2016, vol. 13, no. 2, pp. 197-208.
- [17] A. F. Ashrafi, A. K. M. I. Newaz, R. A. Moin, M. Tanvee, M. A. Mottalib, "A modified algorithm for dna motif

- finding and ranking considering variable length motif and mutation” *IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*, Kolkata, India, 9-11 July 2015.
- [18] J. Sivaranjani and A. N. Madheswari, “A novel technique of motif discovery for medical big data using hadoop” *Conference on Emerging Devices and Smart Systems (ICEDSS)*, Tiruchengode, India, 3-4 March 2017.