Asian Journal of Multidisciplinary Research (AJMR) 2017

E-ISSN: 2395-1710 P-ISSN: 2395-1729
Volume 03- Issue02 -, pp-27-29

Research Paper Open Access

SMART DISEASE PREDICTION USING CLUSTERING

[1]Vagulamaliga.K, [2]Mirudhula. S, [3]Pavithra S.R, [4]Rama.K , [5]Kodhai. E
[1,2,3,4,5]Department of CSE, Sri Manakula Vinayagar Engineering College
Madagadipet, Puducherry, India.
[1]vagulamaligak@gmail.com,[2]mirudhi@gmail.com,[3]pavis621@gmail.com,
[4]ramakmrs@gmail.com

Abstract

Data mining is theprocess of extracting hidden prognosticinformation from large databases and is a powerful new technology with great potential. Mining useful knowledge from corpus of data has become a paramount application in many fields. There are many data mining algorithms for which performs operations like clustering, classification work on the data and provide crisp information for analysis. As these data are available through various channels into public domain, privacy for the owners of the data is significant. The healthcare industry contains large information, which is tedious to process by manual methods. In the existing system, they proposed a graph-based, semi-supervised learning algorithm called SHG-Health (Semi-supervised Heterogeneous Graph) for risk predictions. But there are two issues, performance issue and data source issue in the system. Medical datasets are often not balanced in their class labels. Most of the existing classification methods tend to perform poorly on dataset which is extremely imbalanced. Hence, an alternative method of modelling the objects is required .If we propose an efficient algorithm then the above said issues can be avoided. So in this paper we propose an algorithm for healthcare system to accurately predict the result from the large amount of data.

Introduction

Apart from extraction of information, data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Pattern Evaluation and Data Presentation. The diagnosis of Disease is an important as well as tedious task. Detecting heart disease using various factors or symptoms is a multilayered issue which is not free from fallacious surmises and often accompanied by unpredictable effects. So in order to make this process easier, classifying the large data sets based the causing factor range obtained from medical databases may be a better option. The healthcare data are often not exploited. The existing methodologies have many issues such as redundancy of data and are out dated approaches. They cannot be used for effective decision making. Hence an alternative method is required and which could be done by forming clusters. Using clusters we could easily classify the datasets. So in this paper proposes an algorithm to accurately predict the result from large medical dataset.

Related work

M. F. Ghalwash, et al. [1] proposed a model for the extraction of interpretable multivariable patterns for early diagnostics. First, the time series data is transformed into a binary matrix representation suitable for

application of classification methods. Second, a novel convex-concave optimization problem is defined to extract multivariate patterns from the constructed binary matrix. Then, a mixed integer discrete optimization formulation is provided to reduce the dimensionality and extract interpretable multivariate patterns. Finally, those interpretable multivariate patterns are used for early classification in challenging clinical applications.Y. Zhao, G. Wang [2] proposed a model for advanced microarray technologies have enabled to simultaneously monitor the expression levels of all genes. An important problem in microarray data analysis is to discover phenotype structures. The goal is to 1) find groups of samples corresponding to different phenotypes (such as disease or normal), and 2) for each group of samples, find the representative expression pattern or signature that distinguishes this group from others.Shekar B and Natarajan R[3], proposed a transaction-based neighbourhood-driven approach to quantifying association rules. And Radial Basis Function Neural network (RBFN) to categorize mind MRI pictures to either cancer or noncancerous growth instantly. BPN and RBF classifiers are used to categorize and section the growth section in irregular pictures. Both the examining and training stage gives the amount of precision on each parameter in sensory systems, which gives the idea to choose the best one to be used in further, performs.

Research direction

- In the graph based approach they use extraction methods which tend to perform poorly on dataset which is extremely imbalanced.
- In SVM methods picking/finding the right kernel can be a challenge and Results/output are incomprehensible.
- In neural network picking the correct topology is difficult and Training takes a long time/requires a lot of data.
- Using Decision trees and divide-and-conquer algorithms may over fit data
- If any noisy or irrelevant data are found during data processing the entire process will often reinitialize from the initial stage thereby increasing the process time.

Discussion

In the existing system, accuracy of the prediction is increased through adding additional attributes. Using the basic attributes such as blood pressure, cholesterol, Pulse rate, age and gender the accuracy level of the prediction and when two additional attributes for any disease were added the accuracy level of the prediction would be increased. The predictive accuracy of the developed model was evaluated using precision, recall. They are used to evaluate the correctness of the partition of extracted samples, the computations of which follow a common evaluation. Proposed framework which demonstrates that the models were indeed able to make accurate predictions. Having even more accurate predictions would further support tactical decision-making, and could further improve the care process. The drawbacks of the existing system can be avoided by using the above said algorithm.

Conclusion

Data mining has great importance for area of medicine, and it represents comprehensive process that demands thorough understanding of needs of the healthcare organizations. Knowledge gained with the use of techniques of data mining can be used to make successful decisions that will improve success of healthcare organization and health of the patients. Mining health examination data is

challenging especially due to its heterogeneity, intrinsic noise, and particularly the large volume of unlabeled data. The main techniques included in the survey are decision tree, knowledge discovery and booster algorithm. Through the SVM algorithm decisions can be easily predicted and it is a time efficient process. The results evaluated and finally our current system will accurately predict the result from the large amount of data. Through this proposed algorithm decisions can be easily predicted and it is a time efficient process. The results evaluated and finally our current system will accurately predict the result from the large amount of data.

References

[1] M. F. Ghalwash, V. Radosavljevic, and Z. Obradovic, "Extraction of interpretable multivariate patterns for early diagnostics," IEEE International Conference on Data Mining, pp. 201–210, 2013.

[2] Y. Zhao, G. Wang, X. Zhang, J. X. Yu, and Z. Wang, "Learning phenotype structure using sequence model," IEEE Trans. Knowl. Data Eng., vol. 26, no. 3, pp. 667–681, Mar. 2014.

[3] Shekar B, Natarajan R 2004b A transaction-based neighbourhood-driven approach to quantifying interestingness of association rules. Proc. Fourth IEEE Int. Conf. on Data Mining (ICDM 2004) (Washington, DC: IEEE Computer. Soc. Press) pp 194–2014.

[4] Yan Li, Changxin Bai, Chandan K. Reddy," A distributed ensemble approach for mining healthcare data under privacy constraints", Information Sciences,vol. 330, no. 20, pp.245-259, 2015.

[5] L. Chen, X. Li, S. Wang, H.-Y. Hu, N. Huang, Q. Z. Sheng, and M. Sharaf, "Mining personal health index from annual geriatric medical examinations," in Proc. IEEE Int. Conf. Data Mining, 2014, pp. 761–766.

[6] H. Huang, J. Li, and J. Liu, "Gene expression data classification based on improved semisupervised local Fisher discriminant analysis," Expert Syst. Appl., vol. 39, no. 3, pp. 2314–2320, 2012.

[7] C.-H. Jen, C.-C. Wang, B. C. Jiang, Y.-H. Chu, and M.-S. Chen, "Application of classification techniques on development an early-warning system for chronic illnesses," Expert Syst. Appl.,vol. 39, no. 10, pp. 8852–8858, Aug. 2012.

[8] E. Kontio, A. Airola, T. Pahikkala, H. Lundgren-Laine, K. Junttila, H. Korvenranta, T. Salakoski, and S. Salanter€a, "Predicting patient acuity from electronic patient records," J. Biomed. Informat. Vol. 51, pp. 8–13, 2014.