Prediction for Common Disease using ID3 Algorithm in Mobile Phone and Television

L.Sathish Kumar,   **A.Padmapriya M.C.A**

Department of Comp. Sci & Engg, Alagappa University, Karaikudi – 630001

Abstract

The data mining has become a unique tool in analyzing data from different perspective and converting it into useful and meaningful information. Now we have a lot of known diseases and unknown diseases around the world. The healthcare has big challenge to predict the kind of disease and the solution for that disease. In India illiteracy rate is high, so that most of the people are scared about these diseases become of thesis ignorance. Hence they may take wrong decision regarding the disease that they have been affected problem. Considering this serious issue we have used data mining as a tool to overcome this issue. We have already created the prediction for common disease [17]. And we are in the process implementing of mobile phone and television because all category people can used easily find and predicted what kind of disease through television and mobile phones.

Keywords

ID3 algorithm, Data mining, Common Disease, Prediction, Television, Mobile Phones.

1. Introduction

The healthcare domain have a lot of challenges and difficult task its one of the main difficult challenge is in disease diagnosis. The data mining is the process of analyzing a huge data from different perspective and summarizing it into useful information. The information can be converted into knowledge about historical patterns and future trends. Health care industry today generates large amounts of complex data about patients, hospitals resources, disease, diagnosis methods, electronic patients records, medical devices etc..

Medical history data comprises a number of tests essential to diagnose a particular disease [2]. Clinical databases are elements of the domain where the procedure of data mining has develop into an inevitable aspect due to the gradual incline of medical and clinical research data. It is possible for the healthcare industries to gain advantage of Data mining by employing the same as an intelligent diagnostic tool. It is possible to acquire knowledge and information concerning a disease from the patient specific stored measurements as far as medical data is concerned. Therefore, the data mining has been developed into a vital domain in healthcare [19]. It is possible to predict the efficiency of medical treatments by building the data

mining applications. Data mining can deliver an assessment of which courses of action prove effective [13] by comparing and evaluating causes, symptoms, and courses of treatments. The real-life data mining applications are attractive since they provide data miners with varied set of problems, time and again. Working on heart disease patients databases is one kind of a real-life application. The detection of a disease from several factors or symptoms is a multi-layered problem and might lead to false assumptions frequently associated with erratic effects. Therefore it appears reasonable to try utilizing the knowledge and experience of several specialists collected in databases towards assisting the diagnosis process [8], [3].

The researchers in the medical field identify and predict the diseases besides proffering effective care for patients [8, 19, 10, 16, 11] with the aid of data mining techniques. The data mining techniques have been utilized by a wide variety of works in the literature to diagnose various diseases including: Diabetes, Hepatitis, Cancer, Heart diseases and the like [4, 1, 15, 7]. Information associated with the disease, prevailing in the form of electronic clinical records, treatment information, gene expressions, images and more; were employed in all these works. In the recent past, the data mining techniques were utilized by several authors to present diagnosis approaches for diverse types of Common diseases [12, 18, 22, 6, 9, 5].

So the data mining used we have used can easily prediction to the disease based on its symptoms. This paper proposes the idea of publication through the mass media like mobile phones and televisions where this communication devices usage exceeds 94 million in India. India has in second place at watching television. Mobile phones and television has become a most common communication devices even among the low class people, regardless of the educated or uneducated, so disease diagnosis method implemented to television and mobile phone. The people easily predict which one disease we have afire and what kind of treatment we have taken from television prediction and mobile phone prediction application. It method can easily use and predict uneducated people also.

2. Background Literature

A model Intelligent Heart Disease Prediction System (IHDPS) built with the aid of data mining techniques like Decision Trees, Naïve Bayes and Neural Network was proposed by Sellappan Palaniappan et al. [18]. The results illustrated the peculiar strength of each of the methodologies in comprehending the objectives of the specified mining objectives. IHDPS was capable of answering queries that the conventional decision support systems were not able to. It facilitated the establishment of vital knowledge, e.g. patterns, relationships amid medical factors connected with heart disease. IHDPS subsists wellbeing web-based, user-friendly, scalable, reliable and expandable.

The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al. [22]. Experiments were carried out on a sample database of patients' records. The Neural Network is tested and trained with 13 input variables such as Age, Blood Pressure, Angiography's report and the like. The supervised network has been recommended for diagnosis of heart diseases. Training was carried out with the aid of back-propagation algorithm. Whenever unknown data was fed by the doctor, the system identified the unknown data from comparisons with the trained data and generated a list of probable diseases that the patient is vulnerable to.

The problem of identifying constrained association rules for heart disease prediction was studied by Carlos Ordonez [6]. The assessed data set encompassed medical records of people having heart disease with attributes for risk factors, heart perfusion measurements and artery narrowing. Three constraints were introduced to decrease the number of patterns. First one necessitates the attributes to appear on only one side of the rule. The second one segregates attributes into uninteresting groups. The ultimate constraint restricts the number of attributes in a rule. Experiments illustrated that the constraints reduced the number of discovered rules remarkably besides decreasing the running time. Two groups of rules envisaged the presence or absence of heart disease in four specific heart arteries.

Data mining methods may aid the clinicians in the predication of the survival of patients and in the adaptation of the practices consequently. The work of Franck Le Duff et al. [24] might be executed for each medical procedure or medical problem and it would be feasible to build a decision tree rapidly with the data of a service or a physician. Comparison of traditional analysis and data mining analysis illustrated the contribution of the data mining method in the sorting of variables and concluded the significance or the effect of data and variables on the condition of the study. A chief drawback of the process was knowledge acquisition and the need to collect adequate data to create an appropriate model.

A novel heuristic for efficient computation of sparse kernel in SUPANOVA was proposed by Boleslaw Szymanski et al. [5]. It was applied to a benchmark Boston housing market dataset and to socially significant issue of enhancing the detection of heart diseases in the population with the aid of a novel, non-invasive measurement of the heart activities on basis of magnetic field generated by the human heart. 83.7% predictions on the results were correct thereby outperforming the results obtained through Support Vector Machine and equivalent kernels. The spline kernel yielded equally good results on the benchmark Boston housing market dataset.

In [11] Latha Parthiban et al. projected an approach on basis of coactive neuro-fuzzy inference system (CANFIS) for prediction of heart disease. The CANFIS model diagnosed the presence of disease by merging the neural network adaptive capabilities and the fuzzy logic qualitative approach and further integrating with genetic algorithm. On the basis of the training performances and classification accuracies, the performances of the CANFIS model were evaluated. The CANFIS model is promising in the prediction of the heart disease as illustrated by the results.

In [14] Kiyong Noh et al. put forth a classification method for the extraction of multi-parametric features by assessing HRV from ECG, data preprocessing and heart disease pattern. The efficient FP-growth method was the basis of this method which is an associative. They presented a rule cohesion measure that allows a strong push of pruning patterns in the pattern generating process as the volume of patterns created could possibly be huge. The multiple rules and pruning, biased confidence (or cohesion measure) and dataset consisting of 670 participants, distributed into two groups, namely normal people and patients with coronary artery disease, were employed to carry out the experiment for the associative classifier.

Here upon the all papers concluded a particular disease prediction like heart attack, blood pressure, sugar, breast cancer also but we propose this paper prediction for common disease and that method implemented television and mobile phone.

3. Data mining in common disease prediction

A. ID3 Algorithm

Itemized Dichotomozer 3 algorithm or better known as ID3 algorithm [20] was first introduced by JR.Quinlan in the late 1970's. It is greedy algorithm that selects the next attributes. The information gain associated with the attributes. The information gain is measured by entropy ID3 algorithm. Refers that the generated tree is shorter and the attributes with lower entropies are near the top of the tree.

B. Prediction Procedure

- ✓ Select the dataset for which the test to be retrieved.
- ✓ By using the ID3 algorithm sort the specific pattern and classify the datasets based on the symptoms [5].
- ✓ Then pre-process the fields of dataset based on "Symptom" field and then diagnosis the causes and treatment of the disease also.
- ✓ The paper focuses, the retrieval of dataset, based on the ID3 algorithm that result in the specific dataset fields retrieval.

C. ID3 Algorithm Approach

Function ID3 (part, symptoms, symptoms set, disease, treatment, sample)

```
{/*        part is the set of input attributes
/*         symptoms is the set of input attributes
/*         symptoms set is the set of input attributes
/*         disease is the set of output attributes
/*         treatment is the set of output attributes
/*         sample is a set of training data
/*         Function ID3 returns a decision tree
*/         if (sample is empty) {
```

Return a single node with the value of the most frequent value disease in sample.
/*   Now handle the case where we can't return a signal node */
Compute the information gain for each attribute in part relative to sample;
Let x be the attributes with largest gain (x, sample) at this attributes in parts;
Trees ID3 (part-{x}; Symptoms (Sample-1)... ID3 (test-(sample-n)} ;
Output: Dataset value will be retrieved.

4. Disease Prediction from Television and Mobile Phone

A. Television
Today television has become an integral part of our lives. We have any one information and any one something we got from television. Now a days, people are spending more time on televisions, and no people are there without mobile phone so, through there devices the information can be easily delivered to the people. Now the disease diagnosis method can be easily prediction through television.

Figure 1. Prediction Common Disease form Television

B. Mobile Phone
Its component is cheap though mobile phone has limited capacity and speed. It is handy in dialing, talking, making video film, e-mailing, sending pictures etc. Mobile phones are being extensively used by students at all levels, doctors, engineers, service man, jobbers and common man and woman in their day-to-day activities. Today business is next to impossible without a mobile phone. Starting from aviation industry to service sector, the mobile phones are playing a vital role. Here the mobile phones have lot of application so now we can implement disease prediction method to mobile phone. So we can easily predict our disease very simple at the movement.

Figure 2. Prediction Common Disease form Mobile Phone

5. CONCLUSION

This paper proposes the idea of implementing the disease diagnosis method and prediction of common diseases in common media, television and mobile phones. As, the illiteracy rate is India is high the people are not aware of the disease and its effects. So to avoid theis ignorance, it is implemented is common media like television and mobile

phone. This proposed method phone. This proposed method not only helps the people to know about the diseases but also to avoid the death rate and disease affected people count.

REFERENCE

[1] Abdelghani Bellaachia and David Portnoy, "E-CAST: A Data Mining Algorithm for Gene Expression Data", 2nd Workshop on Data Mining in Bioinformatics at the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, pp. 49 – 54, July 23rd, 2002.

[2] Anamika Gupta, Naveen Kumar, and Vasudha Bhatnagar, "Analysis of Medical Data using Data Mining and Formal Concept Analysis", Proceedings Of World Academy Of Science,Engineering And Technology,Vol. 6, June 2005,.

[3] Andreeva P., M. Dimitrova and A. Gegov, "Information Representation in Cardiological Knowledge Based System", SAER'06, pp: 23-25 Sept, 2006.

[4] A. Bellaachia and Erhan Guven, "Predicting Breast Cancer Survivability using Data Mining Techniques", Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining (SDM 2006), Saturday, April 22, 2006

[5] Boleslaw Szymanski, Long Han, Mark Embrechts, Alexander Ross, Karsten Sternickel, Lijuan Zhu, "Using Efficient Supanova Kernel For Heart Disease Diagnosis", proc. ANNIE 06,intelligent engineering systems through artificial neural networks, vol. 16, pp:305-310, 2006.

[6] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules,"Seminar Presentation at University of Tokyo, 2004.

[7] Florin Gorunescu, "Data Mining Techniques in Computer-Aided Diagnosis: Non-Invasive Cancer Detection," International Journal of Bioengineering, Biotechnology and Nanotechnology, Vol.1, No.2, pp.105 - 108, 2008.

[8] Frank Lemke and Johann-Adolf Mueller, "Medical data analysis using self-organizing datamining technologies," Systems Analysis Modelling Simulation, Vol. 43, No. 10, pp: 1399 -1408, 2003.

[9] Franck Le Duff, Cristian Munteanb, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, Vol. 107, No. Pt 2, pp. 1256-9, 2004.

[10] Fu-ren Lin, Shien-chao Chou, Shung-mei Pan, Yao-mei Chen, "Mining time dependency patterns in clinical pathways", Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Vol. 1, pp. 8, 4-7 January 2000.

[11] L.Goodwin, M. VanDyne, S. Lin, S. Talbert, "Data mining issues and opportunities for building nursing knowledge" Journal of Biomedical Informatics, vol. 36, pp: 379-388, 2003.

[12] Heon Gyu Lee, Ki Yong Noh, Keun Ho Ryu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.

[13] Hian Chye Koh and Gerald Tan, "Data Mining Applications in Healthcare", Journal of healthcare information management, Vol. 19, No. 2, pp. 64-72, 2005.

[14] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, pp: 721- 727, 2006

[15] KS Leung,YT Ng, KH Lee, LY Chan, KW Tsui, Tony Mok, CH Tse, Joseph Sung, "Data Mining on DNA Sequences of Hepatitis B Virus by Nonlinear Integrals" Proceedings Taiwan-Japan Symposium on Fuzzy Systems & Innovational Computing, Japan, pp.1-10, 18-22 Aug 2006.

[16] Margaret R. Kraft, Kevin C. Desouza, Ida Androwich, "Data Mining in Healthcare Information Systems: Case Study of a Veterans' Administration Spinal Cord Injury Population",Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03), pp.9, 6-9 January 2003.

[17] L.Sathish Kumar and A.padmapriya,"ID3 Algorithm Performance of Diagnosis for common disease", Vol. 2, Issue 5, May 2012.

[18] Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008

[19] S Stilou, P D Bamidis, N Maglaveras, C Pappas, "Mining association rules from clinical databases: an intelligent diagnostic process in healthcare", Stud Health Technol Inform 84: Pt 2. 1399-1403, 2001.

[20] Quinlan J.R., INDRODUCTION OF DECISION TREES Machine learning. VOL 1986, 81-106

L.Sathish Kumar is with Department of Computer Science, Alagappa University, Karaikudi-630003, Tamil Nadu, India. As Research Scholar. He is currently pursuing Master of Philosophy in Computer Science at Alagappa University. He has published 2 papers in National and International Conference and Published 1 Paper in ijarcsse international Journal.

Dr. A.Padmapriya is with Department of Computer Science & Engineering, Alagappa University, Karaikudi – 630 003, Tamil Nadu, India as Assistant Professor. She has received her Ph.D in Computer Science and Engineering from Alagappa University, Karaikudi, Tamil Nadu. She has been a member of IACSIT. She published 6 papers in National and International conferences.