Future Prediction For Heart Attack Problem Based on Most Appropriate Attribute value

Amit Kisan Pagare¹, Vijay Kumar Verma²

¹ Lord Krishna College of Technology Indore  M.P. India, M.Tech IV Sem
² Assistant Professor CSE, Loed Krishna College of Technology Indore M.P India,

Abstract: -Superior data mining techniques are developed and used to discover hidden pattern form historical data. New Models are developed from these techniques will be useful for medical practitioners to take successful decision. Diagnosis of heart attack is a significant task in medical science. The term Heart attack includes the various diseases that involve the heart attack problem. The exposure of heart attack problem from different symptoms is an important issue for predicting heart attack problem. In this paper have taken 10 attribute which are responsible for the heart attack problem. We convert the given test data set into binary format with a possible conditions for heart attack. In seconds step we divide the data set and apply most appropriate condition on each attribute Find pair for each attribute which satisfy the condition. We repeat the process for grouping the attribute until no more grouping is possible. At last we find the most common attribute and calculate how much percentage data is accurately classified and confidence that satisfies the validation of downward closure property[4,5].

In 2011 Mrs.G.Subbalakshmi and Mr. K. Ramesh Proposed “Decision Support in Heart Disease Prediction System using Naive Bayes”. They proposed a Decision Support in Heart Disease Prediction System (DSHDPS) using data mining modeling technique, namely, Naive Bayes. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. They implement the system by using web based questionnaire application. This system helps to train nurses and medical students to diagnose patients with heart disease. Decision Support in Heart Disease Prediction System is developed using Naive Bayesian Classification technique. The system extracts hidden knowledge from a historical heart disease database. This is the most effective model to predict patients with heart disease. This model could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy. DSHDPS can be further enhanced and expanded[6,7,15].

In 2011 Mai Shouman, Tim Turner, Rob Stocker proposed “Using Decision Tree for Diagnosing Heart Disease Patients “Heart disease is the leading cause of death in the worldover the past 10 years. Researchers have been using several data mining techniques to help healthcare professionals in the diagnosis of heart disease. DecisionTree is one of the successful data mining techniques used. However, most research has applied J4.8 Decision Tree, based on Gain Ratio and binary discretization. Gini Index and Information Gain are two other successful types of Decision Trees that are less used in the diagnosis of heart disease. Also other discretization techniques, voting method, and reduced error pruning are known to produce more accurate Decision Trees. This research investigates applying a range of techniques to different types of Decision Trees seeking better performance in heart disease diagnosis. [7,12]

Keyword: fitness value, heart disease, attribute, data mining, classifier.

1. Introduction

In our everyday life there are several example exit where we have to analyze the historical data for example a bank loans officer needs analysis of her data in order to learn which loan applicants are “safe” and which are “risky” for the bank[1,2]. Similarly for a medical researcher it is necessary to analyze breast cancer data in order to predict specific treatments for a patient. These are some examples where the data analysis task required before taking any decision. Classification is a data analysis process, where a classifier is constructed to predict class, for bank loan example prediction class is “yes” or “no” Similarly for medical researcher prediction class is “treatment A”, “treatment B”, or “treatment C” for the medical data[3,4],

Classification process can be divided into two parts
(1) Learning: Training data are analyzed by a classification algorithm. Here, the class label attribute is loan decision, and the learned model or classifier is represented in the form of classification rules. (2) Classification: Test data are used to estimate the accuracy of the classification rules. If the accuracy is considered acceptable, the rules can be applied to the classification of new data tuples.

2. Lecturer Review

In 2010 Sunita Soni and O.P. Vyas proposed “Associative Classifiers for Predictive Analysis in Health Care Data Mining”. They used a combined approach that integrates association rule mining and classification rule mining. The integration is done by focusing on mining a special subset of association rules and then classification is being performed using these rules. Given the readability of the associative classifiers, they are especially fit to applications were the model may assist domain experts in their decisions. They also introduce that combining the advanced association rule mining with classifiers gives a new type of associative classifiers with small refinement in the definition of support

In 2011 Sunita Soni and O.P. Vyas proposed “Associative Classifiers for Predictive Analysis in Health Care Data Mining”. They used a combined approach that integrates association rule mining and classification rule mining. The integration is done by focusing on mining a special subset of association rules and then classification is being performed using these rules. Given the readability of the associative classifiers, they are especially fit to applications were the model may assist domain experts in their decisions. They also introduce that combining the advanced association rule mining with classifiers gives a new type of associative classifiers with small refinement in the definition of support

and confidence that satisfies the validation of downward closure property[4,5],
3. Problem Statement

The heart is very important part of human body. Which pumps blood into the entire body? If circulation of blood in body is inefficient the organs like brain suffer and if heart stops working altogether, death occurs within minutes. Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it. Some of the risk factors for heart disease are Smoking, Cholesterol, and Blood pressure. Diabetes, Sedentary life style, Eating Habits, Stress etc[10,11].

4. Proposed Method

In the proposed method we use the concepts of fitness number. In the proposed approach we convert the given data set into binary format as per the given condition for heart attack. We divide the large data set into number of parts. We perform simple intersection calculation for every part of data set and finally we consider only those item set which satisfy the given minimum support fitness value.

Consider a simple example

<table>
<thead>
<tr>
<th>S.NO</th>
<th>Age</th>
<th>Sex</th>
<th>Blood pressure</th>
<th>Cholesterol</th>
<th>Fasting bloodsugar</th>
<th>Resting ECG</th>
<th>Thalach value</th>
<th>Old peak</th>
<th>Slope</th>
<th>Thal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Possible conditions for heart attack
Age>45, BP>120, Cholesterol range>200, FBS>120, Resting ECG>1, Thalach value>100 Beats/Minute, Old peak>0, Slope>2, Thal Value>3. As per the given conditions we convert the test data set into binary format. 1 represent for condition true and 0 for condition false.

Table 3: binary conversion as per the given condition

<table>
<thead>
<tr>
<th>S.NO</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Output:
Pair of Attribute satisfy the given conations for heart attack

Method:
1) Scan the database D and partition the transaction table into equal size.
2) Find Appropriate Condition (AC) for each attribute.
3) Consider only those attribute which satisfy the given minimum Appropriate Condition remaining attribute.
4) Convert the dataset into binary format 1 for Appropriate Condition and 0 for others.
5) To discover the pair of two attribute use Join and perform logical AND.
6) To determine attribute set, join them and perform logical AND operation.
7) The algorithm iterates to find up to pair of n-attribute itemsets.
8) From each pair find out pair of n-attribute itemsets. These pair of attribute are said to local attribute which satisfy Appropriate Condition.

6. Architecture the Proposed Method
First we convert data set into binary format according to Appropriate Condition conditions. In second step we divide the data set into number of parts. Find pair for each attribute which satisfy the condition. We repeat the process for grouping the attribute until no more grouping is possible. At last we find the most common attribute in all parts and calculate how much percentage data is accurately classified. The working process of proposed model is shown in figure 1.

7. Experimental Evaluation
We use VB dot net 2010 as front end and SQL server as back end for data base. All the experiments were performed on a i3 4M Cache, 2.50 GHz Intel PC machine with 2 gigabyte main memory, running Microsoft Windows 7. To evaluate the performance Real life dataset is used. We have implemented three algorithms first one Bayesian classification algorithm, second is weight associated classifiers and third is our proposed method. We have taken 10 attribute which are mainly responsible for heart attack problem. We have taken the data from a pathology laboratory. We perform experiments using 1000, 2000 and 5000 records.

8. Comparison and Graphs
We have taken number of record and percentage of record classify accurately. In figure 2 show comparison graph. It is clearly show that the percentage of classify correct records by proposed method is more as compared to the previous methods.

9. Conclusion and Future Work
From the experiment it is clear that proposed method is more accurately classify the records as compared to previous method. Proposed method considers all attribute given to heart attack condition. Proposed method is also simple to understand and calculation is also easy. We have taken only ten attribute which are mainly responsible for heart attack, in future we have consider more than ten attribute which are also responsible for heart attack.

References


[10] V. Jaya Rama Krishnaiah, V. Chandra Sekaran, and Dr. K. Ramachand. Predicting the Heart Attack Symptoms using Biomedical Data Mining Techniques. Volume 1, No. 3, May 2012 ISSN2278-1080 The International Journal of Computer Science & Applications (TIJCSA)


