Hybrid Model for Clinical Diagnosis and Treatment Using Data Mining Techniques

G Purusothaman¹, Dr Krishnakumari P²,
¹Assistant Professor, ²Director,
Department of Computer Applications (MCA), RVSCAS

ABSTRACT

Medical special treatment is today expanding so quickly to the amount that even experts have difficulties in following the latest new results, changes and treatments. Computers improve on humans in their ability to remember and such property is very precious for a computer-aided system that enables advances for both diagnosis and treatments. A Computer Aided System or Decision Support System that can suggest knowledgeable human logic or serve as an assistant to a physician in the medical domain is gradually important. In the medical area, diagnostics, grouping and treatment are the key tasks for a physician. System development for such a purpose is also a popular area in Artificial Intelligence research. Decision Support System that bears resemblances with human reasoning have benefits and are often easily accepted by physicians in the medical domain. Electronic health data supervision is characterized by high pressure and timely access. Retrieving patient data needs that all services and objects are connected to make data from different health care bases available. Clinical data warehouses in this contextual facilitate the analysis, link and access of the data obtained in the patient care process to progress the quality of decision making. This paper proposes the plan of a real-time adaptive framework that shields the improvement of predicting, responding and monitoring strange behaviors in patient data in adata warehouse environment.

Keywords: Health care data, data mining, clinical data ware house, Computer Aided Diagnosis, Artificial Intelligence, Fuzzy Association Rules, Optimization.

I. INTRODUCTION

Data Mining is an active research area. One of the most popular approaches to do data mining is discovering association rules [1, 2]. Association rules are generally used with basket, census data. Medical data is generally analyzed with classifier trees, clustering, or regression. For an excellent survey on these techniques consult [3]. In this work we explore the idea of discovering fuzzy association rules in medical data, which we believe to be an untried approach. One of the most important features of association rules is that they are combinatorial in nature. This is particularly useful to discover patterns that appear in subsets of all the attributes.

However most designs normally discovered by current algorithms are not useful since they may contain redundant information, may be immaterial or describe trivial knowledge. The goal is then to find those rules which are medically interesting besides having minimum support and confidence. In our investigation project the discovered rules have two purposes: endorse rules used by a fuzzy based expert system to aid in various disease diagnoses [4] and learn new rules that relate causes to a kind of disease and thus can supplement the expert system knowledge. At the instant all rules used by our knowledge-based system [4] were discovered and validated by a cluster of field experts. In our research we are mining data on clinical data warehouses. Clinical data warehouses simplify the study and contact to data obtained in the patient care process which improves the value of decision making and timely data processing intervention [5]. These are more than a large collection of clinical data and generally data comes from other initiatives systems, devices and sensors with the data being commonly integrated into data stores according to the data warehouse architecture defined. According to [6], in average, patient’s influences have hundreds of different facts describing their current situation. Convolute and unpredictable procedures require quick decisions. Thus, more advanced classification constructions are necessary to provide nonstop data monitoring and measuring. In precipitate, it is necessary to accomplish a large amount of mixed electronic heath records and also give efficient provision to process query and analysis at any time. Examples for such thing are drug interactions, sensor measurements or laboratory tests.
II. DEFINITIONS AND DATA PLOTTING

2.1 Fuzzy Association Rules

Here we give the classical definition of association rules. Let \( \{s_1, s_2, \ldots, s_n\} \) be a set of transactions, and let \( P \) be a set of Products, \( P = \{ p_1, p_2, \ldots, p_m \} \). An association rule is an implication of the form \( X \rightarrow Y \), where \( X, Y \subseteq P \), \( X \cap Y = \emptyset \). \( X \) is called the antecedent and \( Y \) is called the consequent of the rule. One vital topic in data mining research is concerned with the discovery of exciting association rules [1]. An exciting association rule describes an interesting relationship among dissimilar attributes and we refer to such relationship as an association in this paper. A boolean association includes binary attributes. A general association involves attributes that are hierarchically linked and a quantitative association includes attributes that can take on quantitative or categorical values. Present algorithms [7, 8] include discretizing the domains of quantitative attributes into intervals so as to learn quantitative association rules. These intervals may not be beshort and meaningful enough for human specialists to easily obtain nontrivial knowledge from those rules discovered. Instead of using intervals, we introduce a novel method, called Fuzzy intervals, which works linguistic terms to represent the exposed symmetries and exclusions. The linguistic depiction makes those rules discovered to be abundant natural for human specialists to understand. The definition of linguistic terms is founded on fuzzy set theory and hence we call the rules having these terms fuzzy association rules. In statistic, the use of fuzzy methods has been measured as one of the key components of data mining systems because of the attraction with the human knowledge depiction [9].

2.2 General depiction of our medical data

The medical data set we are mining defines the outlines of patients of a hospital being treated for any disease. Each record resembles to the most relevant information of one patient. This outline contains personal information such as age, race, smoker or non-smoker. Dimensions on the patient such as weight, heart rate, blood pressure, etc. are included. Pre-existence of assured diseases are stored. The diagnostics made by a medical doctor or specialist are included as well. Time attributes mainly include medical history dates. Then we have a complex set of quantities that estimate the degree of disease in convinced regions of the patient’s body, how healthy certain regions remain, and quality numbers that review the patient’s body under stress and relaxed conditions. Finally perfusion information from some regions of the body is deposited as binary data. The image data is just a summarization of the patient’s body divided into a few regions. These number of regions varies between the age group 3 and 25. As we can understand this type of data is very rich in information comfortable.

2.3 Plotting attributes

The recorded medical data has to be distorted into a transaction format proper to discover association rules. The medical data comprises categorical, numerical, time and image attributes. To make the problem simpler we give all the attributes as being either categorical or numerical. It is vital to create an item for missing information in each medical variable for two reasons. The missing values are shared and planning would be incorrect without them. Also there is an interest by doctors in examining missing evidence to find errors. This treatment of missing data is not complete. Incertain cases a missing data may mean that the person has no disease whereas in other cases it may be in applicable or not available. For some of the records almost all fields have missing data and then it becomes a counter to get consistent rules involving them. Moreover, not all attributes are similarly probable to have missing data. In any case, since our data sets are so small it is important to take into account every sample even do not discard records which have many missing data. We need to conduct further research to find rules which involve missing data.

The data table has categorical attributes that are simply plotted to items by associating an integer to each diverse categorical value. Each categorical value is adequate candidate to appear in a rule. Apparently this maybe a problem if the cardinality of the attribute domain is high. But that is uncommon with medical data. Binary attributes are a special case in which sometimes both the 0 and 1 occurrences maybe exciting or only either of them is interesting. So we accept the medical doctor decides which categorical values are applicable. The second type of attributes is numerical. To make simpler the problem time and image attributes are equally treated as numerical attributes. Touse association rules on numerical data attributes must be divided into intervals, those intervals are indexed and the index is used to make rules. An important work that contracts with this problem is [12].
III. DISCOVERING FUZZY ASSOCIATION RULES IN MEDICAL DATA

Nevertheless of whether the association being considered is boolean, generalized or quantitative, existing algorithms [1-2, 8, 13-14] choose if it is interesting by having a user supply two thresholds for support and confidence. Given two points X and Y, the support is defined as the percentage of records having both points X and Y and the confidence is defined as the percentage of records having Y given that they also have X. If together support and confidence is greater than the usersupplied threshold, the association is measured interesting. A faintness of these methods lies in the difficulty in determining what these thresholds should be. To overawe this problem, our proposed method utilizes adjusted difference [7, 15-16] analysis to classify interesting associations between points. Unlike former data mining algorithms [1-2, 8, 13-14], the use of this technique has the benefit that it does not require any usersupplied thresholds which are frequently hard to determine. Furthermore, our proposed method also has the benefit that it allows us to learn both positive and negative association rules. A positive association rule tells us that a record having certain attribute value or linguistic term will also have another attribute value whereas a negative association rule expresses us that a record having certain linguistic term will not have another attribute value. The general view of proposed method is as follows:

Item: label/fuzzy subset, One vs. several items for attribute, Predetermined labels, Labels obtained from a cluster-partition process, Horizontal attributes.

IV. THE PROPOSED METHOD

A set of fuzzy transactions may be characterized by a table again. Columns and rows are labeled with identifiers of items and transactions respectively. The cell for item $i$ and transaction $t$ contains a $(0, 1)$ value: the membership degree of $i$ in $t$, $\mu(t, i)$. With all the above requirements we propose the following algorithm.

Let $\Delta$ be the maximum number of items appearing in one rule. Let $X_1, X_2, \ldots, X_M$ be all frequent item sets obtained in step 1.

Step 1:
Generate all item sets as candidates and make one pass over $t_1, t_2, \ldots, t_n$ to compute their supports.
for $k = 0$ to $\Delta$ do
Extend frequent $(k-1)$ item sets by one item belonging to any frequent $(k-1)$ item set.
Let $X = \{i_1; i_2; \ldots; i_k\}$ be a $k$-item set.
If $\text{group}(i_j) \neq \text{group}(i_p)$ and $\text{group}(i_j) \ast \text{group}(i_k) > 0$ for $j \neq k\cap 1 \leq j, p \leq k$ then $X$ is a candidate. Check support for all candidate item sets making one pass over the transactions. Those item sets $X$ s.t. min support <= support($X$) <= max support will be the input for the next iteration. If there is no frequent item set stop (sooner) this step 1.

Step 2:
for $j = 1$ to $M$ do for $k = 1$ to $M$
Let $X = X_j; Y = X_k$, if $X \cap Y = \emptyset$ and min support <= support($X$ union $Y$) <= max support and $(ac(i) \neq 0$ for all $i \in X)$ and $(ac(i) \neq 1$ for all $i \in Y)$ and (support($X$ union $Y$)/support($X$) >= min confidence) then $X \rightarrow Y$ is valid. The proposed hybrid model is shown in the figure.
In this model, Fuzzy association rules are created with Future miner algorithms based on the patient data set. The optimizer provides checks optimization at each level. Finally classifier categorizes the pattern associated with dataset. The identified pattern is useful for further decision making in the hospital. These patterns can be used to take treatment on next level. Patient can be directed into right level treatment for particular disease by the physician.

V. CONCLUSIONS AND FUTURE ENHANCEMENT

Our research effort goes into applying fuzzy association rules on medical data which is located in Clinical Data Warehouse. Initially patient data is recorded and compared with existing history of medical records to discover new pattern or interesting data to further diagnosis and treatment. One of the main appeals of association rules is their simplicity. Most of the developments we propose are simple but useful. Fuzzy Association rules have a combinatorial nature. In that essence we isolated those combinations that are interesting for our domain. We concisely address the problem of mapping complex medical data to items. We make fuzzy associations to exclude certain combinations of items. We constrain rules to have certain items in the antecedent and certain items in the consequent. We boundary rule size to get higher confidence and high support rules. Our reformed algorithm is then faster and finds fewer rules. But those rules tend to be concise and relevant. Features which deserve further research. Automate mapping of attributes relating machine-generated partitions back to domain knowledge. Examine problems with noisy data more closely. Identify other useful constraints besides grouping and antecedent and consequent. Extend grouping constraints to include several groups.

REFERENCES