The International Journal Of Engineering And Science (IJES)
|| Volume || 3 || Issue || 3 || Pages || 39-42 || 2014 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Hybrid Model for Clinical Diagnosis and Treatment Using Data Mining Techniques

G Purusothaman[1], Dr Krishnakumari P[2],
[1,] Assistant Professor, [2,] Director,
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 veryprecious for a computeraided system that enables advances for both diagnosisand treatments. A Computer Aided System or Decision Support System thatcan suggestknowledgeable human logic or serve as an assistant to a physician in themedical domain is gradually important. In the medical area diagnostics, grouping and treatment are the key tasks for a physician. System developmentfor such a purpose is also a popular area in Artificial Intelligence research.Decision Support System that bearsresemblances 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 proposesthe plan of a real time adaptive framework thatshields the improvement of predicting, responding andmonitoring 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.*

Date of Submission: 14 March 2014 Date of Acceptance: 25 March 2014

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 datais generally analyzed with classier trees, clustering, orregression. For an excellent survey on these techniquesconsult [3].In this work we explore the idea of discovering fuzzy association rules in medical data, which we believe to bean untried approach. One of the most important features of association rules is that they are combinatorial in nature. This is particularly useful to discoverpatterns that appear in subsets of all the attributes.

However most designs normally discovered by currentalgorithms are not useful since they may contain redundant information, may be immaterial or describe trivialknowledge. The goal is then to find those rules whichare medically interesting besides having minimum support and confidence. In our investigation project the discovered rules have two purposes: endorse rules used by afuzzy based expert system to aid in various disease diagnoses[4] and learn new rules that relate causes toany kind of disease and thus can supplement the expert systemknowledge. At the instant all rules used by our knowledgeablesystem [4] were discovered and validated by a cluster offield experts. In our research we are mining data on clinical data warehouses.Clinical data warehouses simplify the study and contact to dataobtained in the patient care process which improvesthe value of decision making and timely processintervention [5].These are more than a large collection of clinicaldata and generally data comes from other initiativesystems, devices and sensors with the data beingcommonly integrated into data stores according to thedata warehouse architecture defined.According to [6], inaverage, patient'sinfluences have hundreds of differentfacts describing their current situation.Convolutedand unpredictable procedures require quick decisions.Thus, more advanced classification constructions arenecessary to provide nonstop data monitoring andmeasuring. In precipitate, it is necessary to accomplishlarge amount of mixed electronic heathrecords,and alsogive efficient provision to process query and analysis atany time. Examples for such thing are drug interactions, sensor measurementsor laboratory tests.

II. DEFINITIONS AND DATA PLOTTING

2.1 Fuzzy Association Rules

Here we give the classical definition of association rules. Let $\{s_1,s_2,\dots s_n\}$be a set of transactions, and let P be a set of Products, P = $\{ p_1,p_2 \dots p_n \}$.An association rule is an implication of the form X → Y, where X, Y $\in$P , $X \cap Y = \emptyset$.X is called the antecedent and Y is calledthe following of the rule.One vital topic in data mining research is concernedwith the discovery of exciting*association rules* [1]. Anexciting association rule describes an interesting relationshipamong dissimilar attributes and we refer to such relationship as an*association* in this paper.A boolean association includes binaryattributes.Ageneral association involves attributes that arehierarchically linked and a quantitative association includesattributes that can take on quantitative or categorical values.Present algorithms [7, 8] include discretizing thedomains of quantitative attributes into intervals so as to learnquantitative association rules. These intervals may not beshort and meaningful enough for human specialists to easilyobtain nontrivial knowledge from those rules discovered. Insteadof using intervals, we introduce a novel method, calledFuzzy intervals, which works linguistic terms to represent theexposedsymmetries and exclusions. The linguisticdepiction makes those rules discovered to be abundant naturalfor human specialists to understand. The definition of linguisticterms is founded on fuzzy set theory and hence we call the ruleshaving these terms fuzzy association rules. In statistic, the use of fuzzy methods has been measured as one of the keycomponents of data mining systems because of the attraction withthe human knowledge depiction [9].

2.2 General depiction of our medical data

The medical data set we are mining defines theoutlines of patients of a hospital being treated for any disease. Each record resembles to the most relevantinformation of one patient. This outline containspersonal 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 orexistences of assured diseases are stored. The diagnosticsmade by a medical doctor or specialist are included aswell. Time attributes mainly include medical historydates. Then we have a complex set of quantitiesthat estimate the degree of disease in convinced regionsof the patient's body, how healthy certain regions remain, andquality numbers that review the patient's bodyeffort under stress and relaxed conditions. Finallyperfusion information from some regions ofthe patient's body is deposited as binary data.The image data isjust a summarization of the patient's body divided into a fewregions.These number of regions varies between the age group 3 and 25. As we can understand this type of data is very rich ininformation comfortable.

2.3 Plotting attributes

The recorded medical data has to be distorted into a transaction format proper to discover association rules. Themedical data comprises categorical, numerical, time andimage attributes. To make the problem simpler wegive all the attributes as being either categorical ornumerical.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 incorrectwithout them. Also there is an interest by doctorsin examining missing evidence to find errors. Thistreatment of missingdata is not complete. Incertain cases a missing data may mean that the personhas no disease whereas in other cases it may be in applicable or not available. For some of the records almostall fields have missing data and then it becomesaencounter to get consistent rules involving them. Moreover,not all attributes are similarlyprobable to have missing data. In any case, since our data sets are so smalland it is important to take into account every samplewe do not discard records which have manymissing data. We need to conduct further research to find ruleswhich involve missing data.

The data table has categorical attributes that aresimplyplotted to items by associating an integer to eachdiverse categorical value. Each categorical value is adecent candidate to appear in a rule. Apparently this maybe a problem if the cardinality of the attribute domain ishigh. But that is uncommon with medical data. Binaryattributes are a special case in which sometimes boththe 0 and 1occurrences maybe exciting or only eitherof them is interesting. So we accept the medical doctordecides which categorical values are applicable.

The second type of attributes is numerical.To make simpler the problem time and image attributesare equally treated as numerical attributes. Touse association rules on numerical data attributesmust be divided into intervals, those intervals areindexed and the index is used to make rules. Animportant work that contracts with this problem is [12].

III. DISCOVERING FUZZY ASSOCIATION RULES IN MEDICAL DATA

Nevertheless of whether the association being considered isboolean, generalized or quantitative, existing algorithms [1-2, 8, 13-14] choose if it is interesting by having a user supplytwo thresholds support and confidence. Given two points Xand Y, the support is defined as the percentage of records havingboth points X and Y and the confidence is defined as thepercentage of records having Y given that they also have X. Iftogether support and confidence is greater than the usersuppliedthreshold, the association is measured interesting. A faintnessof these methods lies in the difficulty in determining what thesethresholds should be.To overawe this problem, our proposed method utilizes adjusteddifference [7, 15-16] analysis to classify interesting associationsbetweenpoints. Unlike former data mining algorithms [1-2,8, 13-14], the use of this technique has the benefit that itdoes not require any usersupplied thresholds which are frequentlyhard to determine. Furthermoreproposed method also has thebenefit that it allows us to learn both positive and negativeassociation rules. A positive association rule tells us that arecord having certain attribute value or linguistic term will alsohave another attribute value whereas anegative association rule expresses us that a record having certain linguistic termwill not have another attributevalue.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 characterizedby a table again.Columns and rows are labeled with identifiersof items and transactions respectively.The cell for item $i_k$and transaction $t_j$containsa (0, 1) value: the membership degree of $i_k$int_j, $t_j(i_k)$.

With all the above requirements we propose the following algorithm.

Let Δ be the maximum numberof items appearing in one rule. Let X_1, X2...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 Δ do

Extend frequent (k-1) item sets by one item belongingto any frequent (k-1) item set.

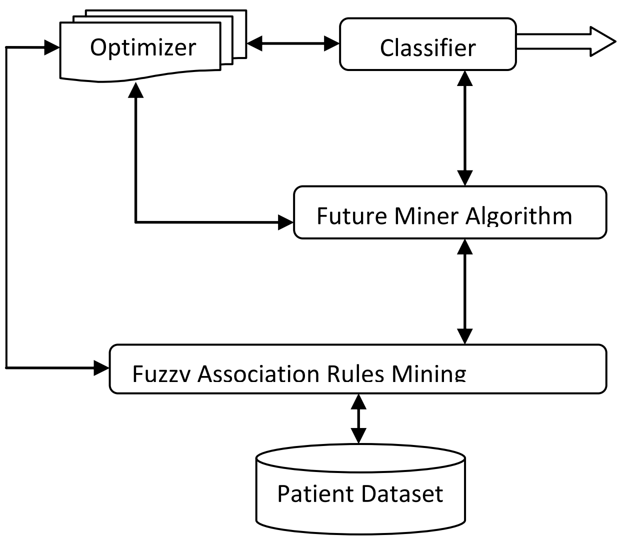
Let X = {i_1; $i_2 \ldots i_k$}be a k-item set.

If group(i_j) ≠ group(i_p) and group(i_j) * group(i_k) > 0 for j ≠k∩ 1<= j,p<= k then X is acandidate. Check support for all candidate kitem setsmaking one pass over the transactions. Those item sets X s.t. minsupport <= support(X) <= maxsupport will bethe input for the next iteration. If there is no frequentitem set stop (sooner) this step 1.

Step 2:

for j = 1 to M do for k = 1 to M do

Let X = X_j; Y = X_k, if X∩Y = ∅ and min support <= support(X union Y) <=maxsupport and (ac(i) ≠0 for alli€ X) and (ac(i) ≠1 for all i € Y) and (support(X union Y)/support(X)>= min confidence)then X→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

Ourresearch effort goes into applying fuzzy association ruleson 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 mainappeals of association rules is their simplicity. Mostof the developments we propose are simple but useful.Fuzzy Association rules have a combinatorial nature.Inthat essence we isolated those combinations that are interesting for our domain. We concisely address theproblem of mapping complex medical data to items. We make fuzzy associations to exclude certain combinationsof items. We constrain rules to have certain items inthe antecedent and certain items in the consequent.We boundary rule size to get higher confidence and highersupport rules. Our reformed algorithm is then fasterand finds fewer rules.But those rules tend to be conciseand 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

[1] RakeshAgrawal, Tomasz Imielinski, and ArunSwami.Mining association rules between sets of items in largedatabases. In ACM SIGMOD Conference, pages 207, 216, 1993.
[2] RakeshAgrawal and RamakrishnanSrikant. Fast algo-rithms for mining association rules in large databases.In VLDB, 1994.
[3] Norberto Ezquerra and RakeshMullick.Perfex: Anexpert system for interpreting myocardial perfusion.Expert Systems with Applications, 6:455{468, 1993.
[4] Usama Fayyad and G. Piateski-Shapiro.From DataMining to Knowledge Discovery.MIT Press, 1995.
[5] T.R. Sahama and P. R. Croll.A datawarehousearchitecture for clinical data warehousing. InProceedings of the fifth Australasian symposiumon ACSW frontiers, Australia, 2007, pp. 227 –232.
[6] P. Chountas and V. Kodogiannis.Developmentof a clinical data warehouse. Proceedings of theIDEAS Workshop on Medical InformationSystems, 2004, pp. 8–14.
[7] K.C.C. Chan, and W.-H. Au, “An Effective Algorithm forMining Interesting Quantitative Association Rules,” in*Proc. of the 12th ACM Symp. on Applied Computing(SAC’97)*, San Jose, CA, Feb. 1997.
[8] R. Srikant, and R. Agrawal, “Mining QuantitativeAssociation Rules in Large Relational Tables,” in *Proc. ofthe ACM SIGMOD Int’l Conf. on Management of Data*,Monreal, Canada, June 1996, pp. 1-12.
[11] A. Maeda, H. Ashida, Y. Taniguchi, and Y. Takahashi,“Data Mining System using Fuzzy Rule Induction,” in Proc.of 1995 IEEE Int’l Conf. on Fuzzy Systems, Yokohama,Japan, Mar. 1995, pp. 45-46.
[12] RamakrishnanSrikant and RakeshAgrawal. Miningquantitative association rules in large relational tables.In ACM SIGMOD Conference, 1996.
[13] J. Han, and Y. Fu, “Discovery of Multiple-LevelAssociation Rules from Large Databases,” in *Proc. of the21st VLDB Conf.*, Zurich, Switzerland, 1995, pp. 420-431.
[14] R. Srikant, and R. Agrawal, “Mining GeneralizedAssociation Rules,” in Proc. of the 21st VLDB Conf.,Zurich, Switzerland, 1995, pp. 407-419.
[15] K.C.C. Chan, and A.K.C. Wong, “APACS: A System for theAutomatic Analysis and Classification of ConceptualPatterns,” Computational Intelligence, vol. 6, pp. 119-131,1990.
[16] K.C.C. Chan, and A.K.C. Wong, “A Statistical Techniquefor Extracting Classificatory Knowledge from Databases,”in [14], pp. 107-123.