Clustering using Mining Fuzzy Association Rules

Dr. S. Maheswari

System Programmer, Alagappa University, Karaikudi, Tamil Nadu, India

ABSTRACT

Data mining play an important role in extracting information or patterns from large database such as datawarehouse and XML repository. In this research we process a technique Clusters are used in fuzzy association rule. It was used to find all the rules that satisfy the minimum support and minimum confidence constraints. In this proposed work new patterns match technique to group association rules, based on the similar attributes, pattern matching clustering algorithm is used to cluster the rules. This research work is used to combine more number of rules with a conditional value. Based on the conditional value, the result will be declared whether the rules or cluster or not.
Keywords: Association rules, Mining Fuzzy, clustering, Data sets.

I. INTRODUCTION

Many Researchers concentrates more on the developing, demonstrating and pushing some of the specific algorithm [1]. In order to make the end user to understand the information clearly from large database fuzzy is applied in database. Many approaches are proposed to mine the association rules but it is important to mine the rules which will satisfies the business needs. Time running rules which is done in fuzzy sequential patterns. The knowledge which is extracted should be satisfactory the end user.

For handling the complexity fuzzy is applied in multi feature sets, multi information sources, constrains and multi model of data mining for analysing complex relations between object [2].

The datasets are collected from the UCI machinery Datasets. The database contain both numeric data and non-numeric data. The non-numeric data that is categorical data are converted into numerical data. The famous Apriori algorithm is used to collect a frequent itemsets. The collected records are clustered by grouping similar records.

1.1 Datasets
The data are collected in the ARFF or CSV format. This data like climate, contact lenses, animal, nursery, car, etc are gathered in this research. The data sets contain attributes and instances. Each record set is a collection of value of attributes. Each dataset have more number of instances.

1.2 Weka

Weka (Waikota Environment for Knowledge Analysis) is one of the useful tools for data mining. Weka is widely used clustering. Weka supports ARFF and CSV format data sets. This work supports Apriori algorithm for association rule and finding the frequent item sets with the help of weka tool.

II. METHODS AND MATERIAL

Related Works

Precede [3] From Learned Rules to Knowledge; this paper presents a new idea of cluster patterns. The concepts of cluster association rules are defined; the interestingness of each is designed; and two kinds of redundancy are analysed [5]. The proposed patterns are more useful and actionable than traditional simple association rules. And our technique, which has been tested with real world data, has provided some interesting and helpful results.

Analysis in this paper [4], we surveyed the list of existing association rule mining techniques. Various data mining techniques are applied to the data source; different knowledge comes out as the mining result. That knowledge is evaluated by certain rules, such as the domain knowledge or concepts. After we get the knowledge, the final step is to visualize the results. They can be displayed as raw data, tables, decision trees, rules, charts, data cubs or 3D graphics. This process is tried to make the data mining results easier to be used and more understandable.

An Approach [6] to Medical Image Classification Using Neuro Fuzzy Logic and ANFIS Classifier; Experimental result indicates that the technique is workable with accuracy greater than 90%. This technique is fast in execution, efficient in classification and easy in implementation. As an overall conclusion, this paper is successful as it met the objectives of the paper and successfully developed, run and optimized the performance of the classification technique.

III. RESULTS AND DISCUSSION

The heart attack datasets are taken. It is an ARFF format. Based on the dataset, we have Seventeen attributes like age, sex, smoking, overweight, hereditary, Badcholesterol, (BC) blood sugar, alcohol, high salt, Blood Pressure, Sedentary, Exercise, Heart rate, High Fat, Stress, Sugar and Salt.

For fuzzy mining, first of all we find the fuzzy association rule. Then the fuzzy association rule merge are cluster.

The following algorithm steps used to find out the fuzzy Association with cluster data sets.
Step 1: Select the non-numerical data (Categorical data) and converted into Numerical data
Step 2: Convert the data set to fuzzy data set
Step 3: Extract rules (Fuzzy Rules) from fuzzy data set to rule1 and rule2 data set.
Step 4: Extract fuzzy Association rule patterns
Step 5: Check the confidence level using the Fuzzy data set.
Step 6: Cluster the Fuzzy Association Rule

Some of datasets has 24 instances or record sets.

TABLE I
SAMPLE RECORD SETS OF HEART ATTACK

Age	Smoking	Hereditary	BC	Alcohol	Exercise
45	Never	No	140	Never	Current
35	Current	Yes	340	Past	Never
25	Past	No	240	Current	Current
55	Never	No	200	Never	Pat
35	Current	No	123	Past	Past
65	Never	No	156	Never	Current
42	Past	Yes	345	Current	Never
44	Never	No	260	Never	Current

Frequent item sets of the database are found with the help of Weka tool. One of the known algorithms for association rule called Apriori algorithm is applied in these datasets by giving appropriate option values. The item sets for the ten association rules are collect and these frequent itemsets are shown in the following figure with minimum confidence of 0.91.

TABLE II
IItem sets satisfying for Association Rule

Smoking	Blood cholesterol	Alcohol	Exercise	Class
Yes	180	Current	Never	None
Yes	147	Past	Past	None
No	133	Never	Current	Soft
		Never	Normal	Soft
		Past	Normal	Soft

From the Association Rules the item sets are collected and matched with original data sets and it is stored in separate table. The analysis will cluster two or more rules. Clustering the datasets of the two association rules by the confidence value. The collected data sets of association rule 1 is matched the original data sets. The resultant data sets are stored in a new table. Same wy each association rule are feat equal record sets and stored in a separate table.

The following table shows the record sets of the matching record of the association rule1 and rule 2 respectively.

TABLE III
Data set for association Rule 1

Age	Smoking	Heredity	Alcohol	Exercise	Bad h Cholestero
45	Never	No	Never	Current	140
35	Current	Yes	Current	Never	340
25	Past	No	Never	Current	240
55	Never	No	Past	Pat	200
35	Current	No	Never	Past	123
65	Never	No	Never	Current	156
42	Past	Yes	Current	Never	345
44	Never	No	Never	Current	260
43	Past	Yes	Past	Past	245
32	Never	No	Never	Current	167
24	Never	No	Never	Current	142

TABLE IV
Data set for association Rule 2

Age	Smoking	Heredity	Alcohol	Exercise	Bad h Cholestero
45	Never	No	Never	Current	156
35	Past	Yes	Current	Never	335
25	Never	No	Never	Current	250
55	Past	Yes	Past	Past	240
35	Never	No	Never	Current	187
65	Never	No	Never	Current	166
42	Past	Yes	Current	Never	365
48	Never	No	Never	Current	280
50	Never	No	Never	Current	140
33	Current	Yes	Past	Never	340
46	Past	No	Current	Current	240

Data pre-processing is to remove the unwanted data. Non numerical data is converted into numerical data by Data Conversion Algorithm. The Categorical data are converted in numeric then to fuzzy.

Algorithm: DCM

Input: Catorigal data;
1. Read Catorigal Data (CD)
2. For each A value in CD
3. String a_1=""; (Contains binary   value)
4. For each B value in CD
   a. If (A= =B)
   $a_1 = a_1 + 1$
   b. Else
   $a_1 = a_1 + 0$;
5. End if
6. Convert s1 to Decimal Value
7. End for

IV. CONCLUSION

A tactical reason for clustering rules is to obtain more concise and abstract descriptions of the data. In this analysis, the researcher considers only the both numeric and non-numerical data; the main scope of this work is to build a number of clusters. When number of iteration is increased, then less number of clusters is get. The most challenging problem in the data mining research and development is the mining complex data for complex knowledge.

The fixed confidence value plays a key role in the cluster. The cluster only depends on the fixed confidence value. As the fixed confidence value varies the number of cluster also varies. It means that whenever fixed confidence value decreases, it automatically decreases the number of clusters. This paper has presented the most comprehensive and a general approach called Fuzzy Association rules in cluster for discovering informative knowledge in complex data.

In future, the main aim of this research will find an accurate solution for checking the order of clustering by giving the priory to the cluster as high low and medium.

V. REFERENCES

[1] Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., and Verkamo, A. I. 1994. Finding interesting rules from large sets of discovered association rules. In Third International Conference on Information and Knowledge Management (CIKM'94), N. R. Adam, B. K. Bhar-gava, and Y. Yesha, Eds. ACM Press, 401{407.

[2] Koperski, K. and Han, J. 1995. Discovery of spatial association rules in geographic information databases. In Proc. 4th Int. Symp. Advances in Spatial Databases, SSD, M. J. Egenhofer and J. R. Herring, Eds. Vol. 951. Springer-Verlag, 47{66.

[3] Mannila, H., Toivonen, H., and Verkamo, A. I. 1995. Discovering frequent episodes in sequences. In Interantional Conference on Knowledge Discovery and Data Mining. IEEE Com-puter Society Press.

[4] Pei, J. and Han, J. 2000. Can we push more constraints into frequent pattern mining? In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM Press, 350 to 354.

[5] Nandagopal, S., V.P. Arunachalam and S. Karthik, 2012. Mining of datasets with an enhanced apriori algorithm. J. Comput. Sci., 8: 599-605. DOI: 10.3844/jcssp.2012.599.605

[6] Martino, F.D. and S. Sessa, 2012. Detection of fuzzy association rules by fuzzy transforms. Adv. Fuzzy Syst., 2012: 258476-258487. DOI: 10.1155/2012/258476.