International Journal of Computer Science
Oddity...Probe...Reviste...

http://www.ijcsjournal.com Volume 2, Issue 2, No 2, 2014. ISSN: 2348-6600
Reference ID: IJCS-058 PAGE NO: 336-342.

Generation of Concise Clusters Using Fuzzy Combined Pattern Association Rules

S.Maheswari*
System Programmer,
Alagappa University-Karaikudi
maheswaris@alagappauniversity.ac.in

Dr.S.S.Dhenakaran
Assistant Professor,
Alagappa University-Karaikudi
ssdarvind@yahoo.co.in

Abstract*—Data mining is the process of extracting interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from large information repositories such as: relational database, data warehouses, XML repository, etc. Combined mining is one of the ordinary methods for analysing complex data for identifying complex knowledge. In this research we process a new technique called Fuzzy Combined Pattern Mining (FCPM) for Domain Driven Data Mining. 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*— Combined Pattern, Association rules, clustering, Data sets, cluster reduction.**

I. INTRODUCTION

The traditional data mining research concentrates more on the developing, demonstrating and pushing the use of the specific algorithms and models(1). Mining fuzzy rules is one of the best ways to summarize large databases while keeping information as clear and understandable as possible for the end-user. Several approaches have been proposed to mine such fuzzy rules, in particular to mine fuzzy association rules. However, we argue that it is important to mine rules that convey information about the order. For instance, it is very interesting to convey the idea of time running in rules, which is done in fuzzy sequential patterns. Regarding the evaluation results, the knowledge can be presented if the result is satisfactory, otherwise we have to run some or all of those processes again until we get the satisfactory result.

The combined mining technique is used for handling the complexity of employing multi feature sets, multi information sources, constraints, multi methods and multi models in data mining and for the analysing complex relations between objects or descriptors (attributes, sources, methods, constraints, labels and impacts) or between identified patterns during the learning process(2). Combined patterns are formed with the analysis of the internal relations between objects or pattern which constitutes and obtained by a single method on a single dataset.

In this research a set of non numeric data for the clustering is collected. The frequent item sets are collecting using apriori algorithm. The item set are collected with the help of Weka tools. Using the item sets, the related record sets from the datasets are collected. The collected record sets are clustered by grouping similar records. The attribute values are numeric or non-numeric data. This paper only considers the non-numeric data for this analysis.

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 for classification, clustering and association. Weka supports ARFF and CSV format data sets. Weka works the algorithms like Apriori, Filtered associative, predictive Apriori, Teritus and generalized sequential patterns. This work supports Apriori algorithm for association rule and finding the frequent item sets with the help of weka tool.

Published by SK Research Group of Companies (SKRGC).

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 2, No 2, 2014. ISSN: 2348-6600
Reference ID: IJCS-058 PAGE NO: 336-342.

II. RELATED WORKS

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.

Precede(3) Combined Pattern Mining: From Learned Rules to Actionable Knowledge; this paper presents a new idea of combined patterns. The concepts of combined association rules, combined rule pairs and combined rule clusters are defined; the interestingness of each is designed; and two kinds of redundancy are analysed(5). The proposed combined 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.

III. RESULTS AND DISSCUSSIONS

In this analysis, we take the datasets of contact lenses. This datasets is based on the patient’s eye problem. It is an ARFF format. Based on the dataset, we have four attributes like age, spectacle prescription, astigmation and tear production rate. The age attribute is classified into young, presbyopic, and pre-presbyopic. The spectacle prescription attribute have myope and hypermetrope. The astigmation have No and Yes. The tear production rates attribute has reduced and normal values. The class contact lenses have soft, hard and none.

For fuzzy based combined pattern mining, first of all we find the fuzzy association rule. Then the fuzzy association rule merge with combined pattern mining.

The following algorithm steps used to find out the fuzzy data sets.

Step 1: Select the non numerical data (Categorical 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 combined patterns
Step 5: Check the confidence level using the Fuzzy data set.
The following datasets has 24 instances or record sets.

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 2, No 2, 2014. ISSN: 2348-6600
Reference ID: IJCS-058 PAGE NO: 336-342.

TABLE I
RECORD SETS OF CONTACT LENSES

S.NO	AGE	SPECTACLE PRESCRIPION	ASTIGMATION	TEAR PRODUCTION RATE	CONTACT LENSES	CHOLESTEROL
1	Young	Myope	No	Reduced	None	174
2	Young	Myope	No	Normal	Soft	150
3	Young	Myope	Yes	Reduced	None	170
4	Young	Myope	Yes	Normal	Hard	200
5	Young	Hypermetrope	No	Reduced	None	200
6	Young	Hypermetrope	No	Normal	Soft	168
7	Young	Hypermetrope	Yes	Reduced	None	175
8	Young	Hypermetrope	Yes	Normal	Hard	230
9	Pre-presbyopic	Myope	No	Reduced	None	163
10	Pre-presbyopic	Myope	No	Normal	Soft	172
11	Pre-presbyopic	Myope	Yes	Reduced	None	201
12	Pre-presbyopic	Myope	Yes	Normal	Hard	204
13	Pre-presbyopic	Hypermetrope	No	Reduced	None	180
14	Pre-presbyopic	Hypermetrope	No	Normal	Soft	158
15	Pre-presbyopic	Hypermetrope	Yes	Reduced	None	200
16	Pre-presbyopic	Hypermetrope	Yes	Normal	None	200
17	Presbyopic	Myope	No	Reduced	None	200
18	Presbyopic	Myope	No	Normal	None	208
19	Presbyopic	Myope	Yes	Reduced	None	210

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 2, No 2, 2014. ISSN: 2348-6600
Reference ID: IJCS-058 PAGE NO: 336-342.

20	Presbyopic	Myope	Yes	Normal	Hard	205
21	Presbyopic	Hypermetrope	No	Reduced	None	220
22	Presbyopic	Hypermetrope	No	Normal	Soft	160
23	Presbyopic	Hypermetrope	Yes	Reduced	None	205
24	Presbyopic	Hypermetrope	Yes	Normal	None	210

TABLE II
RECORD SETS OF CONTACT LENSES

This dataset is applied in the Weka tool for finding frequent item sets with the help of association algorithm. 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 0.9.

TABLE II
IIItem sets satisfying for Association Rule

Attribute 1	Attribute 2	Attribute 3	Attribute 4	Class
	Myope		Reduced	None
	Hypermetrop		Reduced	None
		No	Reduced	None
		Yes	Reduced	None
		No		Soft
			Normal	Soft
		No	Normal	Soft

This item sets are collected from association rules, that rules are matched with originally data sets. The item sets matched with each rule are stored in the separate table. Now the analysis will combine two or more rules. That is combining the datasets of the two association rules by a confidence value.

Now collecting the data sets of the association rule 1 is matching the original data sets. The resultant data sets are stored in a new table. Likewise for each association rule we are getting match 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.

International Journal of Computer Science
Oddity...Probe...Reviste...

http://www.ijcsjournal.com Volume 2, Issue 2, No 2, 2014. ISSN: 2348-6600
Reference ID: IJCS-058 PAGE NO: 336-342.

TABLE III
Data set for association Rule 1

AGE	SPECTACLE PRESCRIPION	ASTIGMATION	TEAR PRODUCTION RATE	CONTACT LENSES	CHOLESTEROL
Young	Myope	No	Reduced	None	174
Young	Myope	Yes	Reduced	None	180
Young	Hypermetrope	No	Reduced	None	172
Young	Hypermetrope	Yes	Reduced	None	170
Pre-presbyopic	Myope	No	Reduced	None	180
Pre-presbyopic	Myope	Yes	Reduced	None	183
Pre-presbyopic	Hypermetrope	No	Reduced	None	185
presbyopic	Myope	No	Reduced	None	189
presbyopic	Myope	No	Normal	None	178
presbyopic	Myope	Yes	Reduced	None	190
presbyopic	Hypermetrope	No	Reduced	None	175
presbyopic	Hypermetrope	Yes	Reduced	None	188

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 2, No 2, 2014. ISSN: 2348-6600
Reference ID: IJCS-058 PAGE NO: 336-342.

TABLE IV
Data set for association Rule 2

AGE	SPECTACLE PRESCRIPION	ASTIGMATION	TEAR PRODUCTION RATE	CONTACT LENSES	CHOLESTEROL
Young	Myope	No	Reduced	None	174
Young	Myope	Yes	Reduced	None	180
Young	Hypermetrope	No	Reduced	None	172
Young	Hypermetrope	Yes	Reduced	None	170
Pre-presbyopic	Myope	No	Reduced	None	180
Pre-presbyopic	Myope	Yes	Reduced	None	183
Pre-presbyopic	Hypermetrope	No	Reduced	None	185
presbyopic	Myope	No	Reduced	None	189
presbyopic	Myope	No	Reduced	None	178
presbyopic	Myope	Yes	Reduced	None	190
Presbyopic	Hypermetrope	No	Reduced	None	175

The similar record sets are taken from two data sets and the counting is stored. The value is 11. The number of records in the first data set is 12. The number of records in the second data set is 11. By the equation 1,
Confidence value = 11/ (12+11-11)*100
= 91.66%

IV. Conclusions

A key reason for clustering rules is to obtain more concise and abstract descriptions of the data. In this analysis, the researcher considers only the non numerical data, the main scope of this work to reduce the number of clusters. So the reduction of the drawback is rectifying using the iterative process. 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 reduction analysis. The cluster reduction 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 the combined mining using fuzzy, for discovering informative knowledge in complex data.

In future, the main aim of this research will find an accurate solution for checking previous order clustering to proceed for the higher order clustering. And also this research is extending to the numeric data.

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 2, No 2, 2014. ISSN: 2348-6600
Reference ID: IJCS-058 PAGE NO: 336-342.

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{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.

