ISSN No. 2395-0900
International Journal of Engineering Studies and Technical Approach
IJESTA

Behavior-Based Clustering and Analysis of Interestingness Measures for Association Rule Mining

Mr. Abhinay Vilas Kulthe *& Prof. Asra Anjum**

**M.Tech.,Department of Computer Science And Technology,M.I.T,Aurangabad, Maharashtra, India.*
***M.Tech.,Department of Computer Science And Technology,M.I.T.,Aurangabad, Maharashtra, India*

ABSTRACT

A large range of studies have been performed to concentrate on insight structure of properties and behavior of interestingness measures for association rule mining.This study analyze rule-ranking behavior of interesting measures tested on rule generated from different datasets.In context of clustering,domain knowledge is key issue to undiscovered bias within interestingness measure gets confirm and extends performance.Study confirms performances and extend it to a logical equivalences among measures to get successed in business objective.

Keywords: *Interestingness measures ,Clustering ,Behavior analysis ,Association rule mining.*

INTERESTINGNESS MEASURE

Interestingness Measure are used to rank association rule the paper is used to analysis the rule ranking behaviour of 61 interestingness measure .By clustering based on rule ranking behaviour .we can higher previously unreported equivalence among interestingness measure.interestingness measure is commonly used confidence for measure and ranking of rules generated from dataset. this can used by sort out by keywords within rules of dataset.The measures that makes the rule interesting ,then produced rules get satisfy conditions that those measure is supposed to be interesting . But sometimes produced rules are not supposed to be interesting so they are supposed to hold some of rules were discovered and they found distributional assumptions.Interestingness Measure uses domain knowledge to produce interesting rules and remove meaningless knowledge associated with rules . coverage is perform after rules formation to remove bias some use raw correlation while today uses rank correlation.

CLUSTERING

Clustering can be enlarge defined as nested subsequence of tree structure .clustering can be made through as per [fig 1. Behaviour based clustering of interestingness measure] as interestingness measure is nested substructure where cut point of each interestingness measure is element to gather similar structure as per distance metrics.the graph correlation of interestingness measure is P+ where distance is less then under root 0.025 .the graph correlation of interestingness measure is P- where distance is greater than under root 0.975.the graph uncorrelation of interestingness measure is found as distance between under root 0.4875 and under root 0.5125 .

BEHAVIOUR ANALYSIS

Raw score is assign to each rule generated in dataset.the rank assigned as per rule by confidence .bias reduced by Fisher's Z Transform and Averaging then bias in those elements

is removed by Fisher's Back Transform .similarity of rule ranking behaviour of interestingness measure.

ASSOCIATION RULE MINING

Association rule mining algorithm view data record as sets of attribute value pair or items, the rule extract of association will have features values of association between attribute value pair of data records.the brief rules for association abstract by single general form is

P(X Y|)=P(X) - P(XY)

P(X| Y)=P(Y) - P(XY)

P(XY|)= 1 - P(X) -P(y) + P(XY)

RANKING

Ranking rules can be made in association rules as per the support of interestingness measure like confidence, lift ,leverage ,conviction this normally uses the confidence as the ranking measure for generated associated rules .this highest confidence can be moved to first rank is series and lowest confidence can be moved to last rank in series. the number of rule generated is depend on simulation ranking always get advanced as the meaningfull rules get sorted in set and domain knowledge is to be get sorted out in ranking order.

ATTRIBUTE VALUE PAIR

Number of Examples

Dataset	Number of Examples
35	1 to 200
36	201 to 500
18	501 to 1000
21	1001 to 48842

Number of Attributes

Dataset	Attributes
36	1 to 20
30	21 to 100
20	101 to 200
19	201 to 300
5	301 to 1559

ISSN No. 2395-0900
International Journal of Engineering Studies and Technical Approach
IJESTA

Number of Attribute Value Pair

Dataset	Attribute Value Pair
23	8 to 30
28	31 to 100
25	101 to 500
31	501 to 1000
3	1001 to 3121

ALGORITHM

Standard procedure to run and understand (as per specified)[1] the normal working of association rule mining which reduces amount of bias ,we

1). run the standard Apriori algorithm as implemented in Weka with a minimum confidence threshold of 0.0.

2). set the minimum support threshold low enough for each dataset to generate at least 1000 rules;and

3). randomly select 100 rules from the result set.

we would keep computational cost resonable while increasing diversity and reducing bias in the set of rules used in our analysis.

```
File : 1 to 110
Association rule : 1 to 100
Interestingness Measure : Confidence
Algorithm : Distance Matrix  computation
for file i is element of FILE (110) do
{
        100 association rules in file;
        for conf is element of CONF(100) do
        {
                for every association rule do
                {
                        Vc(a) <- value of c on a;
                }
                order the pair [a , Vc(a) ] by decreasing Vc(a)
                create new pair [a , Rc (a) ] by replacing Vc(a) by Rc(a).
                order the pair [ a , Rc(a) ] by label increasing value
        }
        for conf1 is element of CONF(100) do
        {
                for conf2 is element of CONF , c1 =! c2 do
```

ISSN No. 2395-0900
International Journal of
Engineering Studies and Technical Approach
IJESTA

```
                {
                        [S 12 on d]  <-  [a,R1(a)]   [a,R2(a)]
                        [Z 12 on d]  <-  Fisher Z transform S 12 on d
                }
        }
}
for c1 is element of CONF 100 do
{
        for c2 is element of CONF , c1 =! c2 do
        {
                Z 12 <- Average D over Z12 on d
                S 12 <- Fisher Back Transform Z12
                M12 <- under root ((1-S 12)% 2)
        }
        M 11 <- 0;
}
```

DISTANCE COMPUTATION

Algorithm explanation
IM = interestingness measure
1). IM to each rule assign raw score.
2). compare score and range of score ranking for rule by place.
3). raw score replace by place.
4). assign label value
5). created matrix of rule by IM . assign rule for each rule.
6). Spearman rank correlation coefficient

```
        score  =  +1     identical ranking
                  -1     reversed ranking
                  -1 to +1        other cases
```

not additive , so create bias in averaging correlation coefficient.

```
bias reduce by = {      Fisher Z transform
                        Averaging Transform Value
                        Back Transform
                }
```

7).|I| * |I| similarity / dissimilarity of rule ranking behaviour among IM.
8).Mij <- under root ((1-Sij) /2)

```
        1 <- identical   distance matrix Mij
        0 <- opposite    true of IM final matrixs unbias in I.
```

ANALYSIS

Fisher's Transform
Z = 0.5 log ((1+S) / (1-S))
rank correlation values are subjected.

Averaging

The purpose is to limit the impact of choice of datasets on our results, and so we will wish to average over these datasets.

Fisher's Back Transform
S = (e ^ 2Z -1)/(e ^ 2z + 1)
Minor bias in Fisher's Transform is eliminated in Back Transform

Spearman Rank formulae[4]
R=1-((6 { d^2)/(n^3 - n))

R = coefficient.
d = difference
n = number of values
+1 Perfect Positive Correlation
0 No Correlation
-1 Perfect Negative Correlation

Spearman rank correlation coefficient is computed for each pair of distinct interestingness measures

Difference

distance -> rank -> average rank -> difference

distance is been fixed in I * I matrix as per similarity/dissimilarity of IM.The ranking is achive by rule ranking among IM.Average Rank would be decided by taking average of sub group IM rule rank as per similarity.Difference would be count by distance between dissimilarity within rank or subgroup rank of IM.

CONCLUSION

As per our study we drawn inferences summary is

1).Defined each interestingness measure in terms of its formula and name.

2).Spot relatively redundant measures and discover equivalences among them focus work based on empirical results.

3).Concentrated on ranking behaviour and reduces number of Interestingness Measure from 61 to 21.

4).Strong corelation,anti-corelation and independence among Interestingness Measure,which permits to consider which Interestingness Measure in what context.

REFERENCES

i. C Tew,C Giraud-Carrier,K Tanner,S Burton(2013) Behavior-based clustering and analysis of interestingness measures for association rule mining.

ii. Abe H, Tsumoto S (2008) Analyzing behavior of objective rule evaluation indices based on a correlation coefficient. In: Proceedings of the 12th international conference on knowledge-based intelligent information and engineering systems (LNAI 5178), pp 758–765.

iii. Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of the 20th international conference on very large data bases, pp 487–499.

iv. Hill T, Lewicki P (2007) Statistics: methods and applications. StatSoft, Tulsa. http://www.statsoft.com/ textbook/.

v. Jain A, Dubes R (1988) Algorithms for clustering data. Prentice-Hall, Inc., Englewood Cliffs.

vi. Johnson S (1967) Hierarchical clustering schemes. Psychometrika 2:241–254.

vii. Padmanabhan B (2004) The interestingness paradox in pattern discovery. J Appl Stat 31(8):1019–1035.

viii. Piatetsky-Shapiro G (1991) Discovery, analysis, and presentation of strong rules. In: Piatetsky-Shapiro G.

ix. Frawley WJ (eds) Knowledge discovery in databases. AAAI Press, Cambridge, pp 229–248.

x. Sahar S (2010) Interestingness measures—on determining what is interesting. In: Maimon O, Rokach L (eds) Data mining and knowledge discovery handbook, 2nd edn. Springer, New York, pp 603–612.

xi. Yao Y, Zhong N (1999) An analysis of quantitative measures associated with rules. In: Proceedings of the 3rd Pacific-Asia conference on knowledge discovery and data mining (LNCS 1574), pp 479–488.

xii. Zhang T (2000) Association rules. In: Proceedings of the 4th Pacific-Asia conference on knowledge discovery and data mining (LNAI 1805), pp 245–256.