

Handling Ambiguity Issues In The Process Of Finding Association Rules

Dr. K. Karpagam¹, Dr. A. Kutralam²

¹Assistant Professor Of Computer Science, H.H. The Rajah's College (Autonomous), Pudukkottai, Tamil Nadu, India. (Affiliated To Bharathidasan University)

²Assistant Professor Of Computer Applications, Bishop Heber College (Autonomous), Tiruchirappalli, Tamil Nadu, India. (Affiliated To Bharathidasan University)

Email: Kkarpaga05@Gmail.Com¹, Kutralam.Ca@Bhc.Edu.In²

Abstract - Normally Association Rules Consider Only Items Enumerated In Transactions. Such Rules Are Termed As Positive Association Rules. In Case Of Negative Association Rules, They Too Consider The Same Items, But Also The Negated Items. While Performing Market-Basket Analysis, The Negative Association Rules Are Useful In To Identify Products That Conflict With Each Other Or Products That Complement Each Other. In This Paper The Two Main Issues In Generating Association Rules Are Addressed. One Is Minimizing Strong, Uninteresting Association Rules And The Other Is Identifying Misleading Association Rules With Strong Negative Correlation Between The Antecedents And Consequents. The Techniques Discussed In This Paper Are Found To Be More Effective And Precise. The Effectiveness Of The Mining Process Can Be Enhanced Significantly By Applying The Methods Discussed.

Key Words - Association Rule Mining, Uninteresting Rules, Negative Associated Rules, Correlation

1. INTRODUCTION

Data Mining Technology Has Emerged As A Means For Identifying Patterns And Trends From Large Quantities Of Digital Data. Data Mining Involves The Semi-Automatic Discovery Of Patterns, Associations, Changes, Anomalies, Rules And Statistically Significant Structures And Events In Data. The Goal In Data Mining Is To Extract Qualitative Models, Which Can Easily Be Translated Into Business Patterns, Logical Rules Or Visual Representations. Data Mining Techniques Are Able To Provide Companies A Greater Assistance In Optimizing Their Use Of Information.

Association Rule Mining

Association Rule Mining Finds Interesting Association Or Correlation Relationships Among A Large Set Of Data Items. Association Rule Identifies The Effect Of Occurrence Of One Event On The Occurrence Of Another Event. With Large Amount Of Data Being Collected And Stored Continuously, The Discovery Of Interesting Association Relationships Among Huge Amounts Of Business Transaction Records Can Help In Many Business Decision Making Processes, Such As Catalog Design, Cross-Marketing, And Loss-Leader Analysis. Association Rules Are Very Often Used By Retail Stores To Assist In Marketing,

Advertising, Floor Placement, And Inventory Control. Association Rules Are Useful In Identifying Frequent Item Sets Which Are Purchased Together. Identification Of Such Item Sets Helps The Retailers To Decide The Quantity Of Order Placement.

Association Rule Mining Uses Many Techniques Including Apriori Algorithm And FP-Growth Algorithm. They Discover Pattern Using Support-Confidence Framework. The Set Of All Sales Transactions Is Called The Population. Transactions Are Represented As One Record Per Transaction. The Transactions Are Represented By A Data Tuple.

II. CHALLENGES

A) STRONG UNINTERESTING RULES

Most Association Rule Mining Algorithms Employ A Support-Confidence Framework. Some Results Are Believable. Some Rules Need Additional Analysis. Some Rules Are Unbelievable. Though Association Rule Mining Uses Minimum Support And Confidence Thresholds To Exclude Uninteresting Rules, Many Rules That Are Not Interesting To The User May Still Be Produced.

Example – 1 Consider A Transaction Database That Contains Purchase Transactions In A Shop Including Brush And Paste. Assume By Applying A Data Mining Program, The Following Association Rule Is Obtained.

$\text{Buy}(X, \text{''Brush''}) \Rightarrow \text{Buy}(X, \text{''Paste''})$

Though This Rule Has High Level Of Support And Confidence, This Is A Common Occurrence Which Does Not Need The Support Of Association Rule Mining To Be Identified.

Sometimes, Even The Strong Rules Can Be Misleading, Makes The User To Take Wrong Decisions Based On The Association Rules Derived.

B) MISLEADING ASSOCIATION RULES

Example – 2 Consider A Shop That Sells Many Items That Includes Cricket Bat And Hockey Sticks. 5000 Transactions Were Analysed. 3000 Of The Customer Transactions Included Cricket Bat, 3750 Included Hockey Sticks, And 2000 Included Both Cricket Bat And Hockey Sticks. By Performing Association Rule Mining, The Following Association Rule Can Be Derived With The Support Of 40% And Confidence Of 66%.

$\text{Buy}(X, \text{''Cricket Bat''}) \Rightarrow \text{Buy}(X, \text{''Hockey Stick''})$

The Support For This Rule Is $2000/5000 = 40\%$
Confidence Value Is $2000/3000 = 66\%$

Though, The Above Rule Is Strong, It Is Misleading Since The Probability Of Purchasing Hockey Stick Is 75%, Which Is Even Larger Than 66%. Point To Be Noted Is That Cricket Bat And Hockey Sticks Are Negatively Associated Because The Purchase Of One Of These Items Actually Decreases The Likelihood Of Purchasing The Other. Without Fully

Considering This Phenomenon, Sometimes One Could Make Unwise Business Decisions Based On The Rule Derived.

I. SOLUTIONS

A) IDENTIFYING STRONG UNINTERESTING RULES USING CONSTRAINT-BASED ASSOCIATION MINING

For A Given Set Of Task-Relevant Data, The Mining Process May Generate Number Of Rules, Many Of Which Are Uninteresting To The User. Find All Patterns In A Data Set Is Unrealistic For There Are Too Many Patterns. In This Case, User Should Be Able To Direct What To Be Mined, Provide Constraints On What To Be Mined, To Make The Mining More Efficient. To Restrict Strong, Uninteresting Rules, We Do Apply **Constraint-Based Association Mining** Technique.

Constraint-Based Association Mining Includes Knowledge Type Constraints, Data Constraints, Dimension/Level Constraints, Interestingness Constraints And Rule Constraints. The Following Facts About Constrains Gets Us More Clarity On Constraint-Based Mining.

Antimonotonic - If An Itemset S Violates The Constraint, So Does Any Of Its Superset

Monotonic - If An Itemset S Satisfies The Constrains, So Does Any Of Its Superset

Succinct - Let A Be The Set Of Items Satisfying A Succinct Constraint C, Then Any Set S Satisfying C Must Contain A Subset Belongs To A

Convertible Constrains - Convert Tough Constraints Into Anti-Monotonic Or Monotonic By Properly Ordering Items.

B) IDENTIFYING MISLEADING RULES USING CORRELATION

Other Main Problem With Association Rule Mining Is That When It Uses Support-Confidence Framework Pattern, It Sometimes Produces Misleading Association Rules That Directs User To Take False Decisions. To Solve This Problem, Another Framework For Finding Interesting Relationships Between Data Item Sets **Based On Correlation** Is Discussed. A Correlation Rule Is Defined As A Set Of Itemsets That Are Correlated. The Aim Of Moving To Correlation Rule Is To Handle Negative Associated Rules.

Support And Confidence Are The Normal Factors Used To Measure The Quality Of An Association Rule:

$$\text{Support}(A \Rightarrow B) = P(A, B)$$

$$\text{Confidence}(A \Rightarrow B) = P(B/A)$$

However, Some Problems Arise With These Measurements. It Is Visible That Confidence Does Not Consider P(B). So, An Alternative Framework For Finding Interesting Relationships Between Data Itemsets Based On Correlation Is Considered.

If $P(A, B) = P(A)P(B)$, Then The Occurrence Of A Is Independent Of Occurrence Of B. Otherwise, A And B Are Dependent And Correlated.

The Correlation Between The Occurrence Of A And B Can Be Computed As

$$\text{Corr}_{A, B} = P(A, B) / P(A)P(B) \quad (\text{Equation} - 1)$$

If The Result Is Less Than 1, Then The Occurrence Of A Is Negatively Correlated With The Occurrence Of B. If The Resulting Value Is Greater Than 1, Then A And B Are Positively Correlated. If The Resulting Value Is 0, That Indicates That A And B Are Independent And There Exist No Correlation Between Them.

The Example – 2 Also Illustrates That The Confidence Of A Rule $A \Rightarrow B$ Can Be Deceiving In That It Is Only An Estimate Of The Conditional Probability Of Itemset B For Given Itemset A. It Does Not Measure The Real Strength Or Lack Of Strength Of The Implication Between A And B. Hence, Alternatives To The Support-Confidence Framework Can Be Useful In Mining Interesting Data Relationships.

To Help Filter Out Misleading “Strong” Associations Of The Form $A \Rightarrow B$, We Need To Study How The Two Itemsets, A And B, Are Correlated. The Transactions Are Summarized In The Table Below.

TABLE I PURCHASE OF CRICKET BAT AND HOCKEY STICK

	No. Of People Purchased Cricket Bat	No. Of People Not Purchased Cricket Bat	Total
No. Of People Purchased Hockey Stick	2000	175	3750
No. Of People Not Purchased Hockey Stick	1000	250	1250
Total	3000	2000	5000

From The Table, We Find That Probability Of Purchasing A Cricket Bat Is 0.60, The Probability Of Purchasing Hockey Stick Is 0.75, And The Probability Of Purchasing Both Is 0.40. By Equation – 1, The Correlation Value Obtained Is 0.89. Since The Value Is Less Than 1, It Is Detected That There Is A Negative Correlation Between The Two Occurrences.

2. CONCLUSION

In This Paper The Challenging Issues Such As Strong, Uninteresting Association Rules And Misleading Negatively Associated Rules During Generation Of Association Rules Were Discussed. As The Solution For The First Issue Mentioned, Using Of Constraint-Based Association Mining Is Proposed, Since They Minimize Strong But Uninteresting Association Rules. As The Solution For The Second Issue, That Is, Identification Of Misleading, Negatively Associated Rules Correlation Is Used. Advantage Of Correlation Is That It Is Upward Closed. That Means That If A Set Of Items Is Correlated, Then Every Superset Of Them Is Also Correlated.

3. REFERENCES

- [1] R. Agrawal And R. Srikant. Fast Algorithms For Mining Association Rules. VLDB 1994

- [2] B. Liu, W. Hsu, Y. Ma. Mining Association Rules With Multiple Minimum Supports. SIGKDD 1999
- [3] K. Wang, Y. He, And J. Han. Mining Frequent Itemsets Using Support Constraints. VLDB 2000
- [4] Jiawei Han, Micheline Kamber Data Mining Concepts And Techniques, MKP, 2001
- [5] G.K. Gupta, Introduction To Data Mining With Case Studies, PHI, 2006
- [6] Margeret H. Dunham, Data Mining Introductory And Advanced Topics, PE, 2006.