Laboratory Manual

Subject: DWM Lab | Class/Sem / Batch: T.E. /V

Name of Faculty: Prof. Archana Kotangale | Department: Computer Engineering

Academic Year: 2022-2023

Experiment – 9

Title: Implementation of Association Rule Mining algorithm (Apriori)

Objective:

- To learn association rule for Apriori algorithm

Reference:

- Data Mining Introductory & Advanced Topic by Margaret H. Dunham
- Data Mining Concept and Technique By Han & Kamber

Pre-requisite:

- Fundamental Knowledge of Database Management

Theory:

Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given data base. The problem is usually decomposed into two sub problems.

- Find those item sets whose occurrences exceed a predefined threshold in the database; those item sets are called frequent or large item sets.

- Generate association rules from those large item sets with the constraints of minimal confidence.

Suppose one of the large item sets is L k = {I1, I2,...,Ik}; association rules with this item sets

A regenerated in the following way: the first rule is {I1, I2,...,Ik−1}=>{Ik}.By checking the confidence this rule can be determined as interesting or not. Then, other rules are generated by deleting the last items in the antecedent and inserting it to the consequent , further the confidences of the new rules are checked to determine the interestingness of them. This process iterates until the antecedent becomes empty.
Since the second sub problem is quite straight forward, most of the research focuses on the first sub problem. The Apriori algorithm finds the frequent sets Lin Database D.

Laboratory Manual

Subject: DWM Lab | Class/Sem / Batch: T.E. /V

Name of Faculty: Prof. Archana Kotangale | Department: Computer Engineering

Academic Year: 2022-2023

Find frequent set Lk − 1.
Join Step.
Ck is generated by joining Lk − 1with itself
Prune Step.
Any (k − 1)-item set that is not frequent cannot be a subset of a frequent k-item set, hence should be removed. where

- (Ck: Candidate item set of size k)
- (Lk: frequent item set of size k)

Apriori Pseudo code

Apriori Pseudocode

$Apriori\ (T, \varepsilon)$

$L_1 \leftarrow \{$ large 1-itemsets that appear in more than ε transactions $\}$

$k \leftarrow 2$

while $L_{k-1} \neq \varnothing$

$C_k \leftarrow$ Generate(L_{k-1})

for transactions $t \in T$

$C_t \leftarrow$ Subset(C_k,t)

for candidates $c \in C_t$

$\text{count}[c] \leftarrow \text{count}[c] + 1$

$L_k \leftarrow \{c \in C_k | \ \text{count}[c] \geq \varepsilon\}$

$k \leftarrow k + 1$

return $\bigcup_k L_k$

Input :

A large super market tracks sales data by SKU (Stoke Keeping Unit) (item),and thus is able to know what items are typically purchased together. Apriori is a moderately efficient way to build a list of frequent purchased item pairs from this data Let the database of transactions consist of the sets {1,2,3,4}, {2,3,4}, {2,3}, {1,2,4}, {1,2,3,4}, and {2,4}.

Output

Each number corresponds to a product such as "butter" or "water". The first step of

Laboratory Manual

Subject: DWM Lab | Class/Sem / Batch: T.E. /V

Name of Faculty: Prof. Archana Kotangale | Department: Computer Engineering

Academic Year: 2022-2023

Apriori to count up the frequencies, called the supports, of each member item separately:

Item	Support
1	3
2	6
3	4
4	5

We can define a minimum support level to qualify as "frequent," which depends on the context. For this case, let min support=3.Therefore, all are frequent. The next step is to generate a list of all 2-pairs of the frequent items. Had any of the above items not been frequent, they wouldn't have been included as a possible member of possible 2-item pairs.

In this way, Apriori prunes the tree of all possible sets.

Item	Support
{1,2}	3
{1,3}	2
{1,4}	3