Practical 7

Aim: Implement Apriori Algorithm.

Transaction ID	onion	Potato	Burger	milk	Beer
t1	1	1	1	0	0
t2	0	1	1	1	0
t3	0	0	0	1	1
t4	1	1	0	1	0
t5	1	1	1	0	1

Date :

Practical No. 7

Aim: Implement Apriori Algorithm.

Theory:

In general association rule mining can be viewed as a two-step process:

(1) Find all frequent itemset: By definition each of these itemsets will occur at least as frequently as a predetermined minimum support count min sup.

(2) Generate strong association rules from the frequency itemsets: By definitions these rule must satisfy minimum support and minimum confidence.

Let $I = \{i1, i2, i3, \dots, in\}$ be a set of n attributes called items and $D = \{t1, t2, \dots, tn\}$ be the set of transaction. It is called a database. Every transaction, ti in D has a unique transaction ID, and it consist of subsets of item set in I.

A rule can be defined as an implication, $X \to Y$ where X and Y are subsets of I $(X, Y \subseteq I)$ and they have no element in common i.e. $X \cap Y$, X and Y are the antecedent and the consequent of the rule, respectively.

Let's take an easy example from the Supermarket Sphere. The example that we are considering is quite small and in practical situation, datasets contain millions or billions of transactions. The set of itemsets, I = {onion, Burger, potato, milk, Beer} and a database consisting of six transactions. Each transaction is a tuple of 0's and 1's where 0 represents the absence of an item and 1 the presence.

An example for a rule in this scenario would be {onion, potato} => {Burger}, which means that if onion and potato are bought, customers also buy a burger.

There are multiple rules possible even from a very small db, so in order to select the interesting ones, we use constraints on various measures of interest and significance. We will look at some

Date :

of these useful measures such as support, confidence, lift and conviction.

Conclusion:

Hence implementation of Apriori algorithm for association rule mining is studied.

Viva Questions:

① Define association rule mining?

→ Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently a itemset occur in a transaction. A typical example is market Based Analysis. market based analysis is one of the key technique used by large relation to show associations betn items. It allows retailers to identify relationship betn the items that people buy together frequently.

② Define apriori algorithm?

→ The Apriori algo. is used for data mining frequent itemsets & devising association rules from a transactional database. The parameters 'support' and 'confidence' are used. Support refers to items' frequency of occurrence, confidence is conditional probability.

③ What is meant by frequent itemset mining?

→ Frequent pattern mining is a data mining subject with the objective of extracting frequent itemsets from a db. Frequent itemsets plays an essential role in many data mining tasks and are related to interesting patterns in data, such as Association Rules.

Date :

(4) Define support and confidence ?

→ The number of transactions that include items in the {x} and {y} parts of the rule as a percentage of the total no. of transaction. It is a measure of how frequently the collection of items occur together as a percentage of all transactions.

Page No.

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Associator

Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Result list (right-...

16:47:09 - Apriori

Associator output

```
=== Run information ===

Scheme:       weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     soybean
Instances:    683
Attributes:   36
              date
              plant-stand
              precip
              temp
              hail
              crop-hist
              area-damaged
              severity
              seed-tmt
              germination
              plant-growth
              leaves
              leafspots-halo
              leafspots-marg
              leafspot-size
              leaf-shread
              leaf-malf
              leaf-mild
              stem
              lodging
              stem-cankers
              canker-lesion
              fruiting-bodies
              external-decay
              mycelium
              int-discolor
              sclerotia
              fruit-pods
              fruit-spots
              seed
              mold-growth
              seed-discolor
              seed-size
              shriveling
              roots
              class
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.8 (546 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 4

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 6
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Screenshot captured
You can paste the image from the clipboard.

Associator

Choose | Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start | Stop

Result list (right-...

16:47:09 - Apriori

Associator output

```
              plant-growth
              leaves
              leafspots-halo
              leafspots-marg
              leafspot-size
              leaf-shread
              leaf-malf
              leaf-mild
              stem
              lodging
              stem-cankers
              canker-lesion
              fruiting-bodies
              external-decay
              mycelium
              int-discolor
              sclerotia
              fruit-pods
              fruit-spots
              seed
              mold-growth
              seed-discolor
              seed-size
              shriveling
              roots
              class
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.8 (546 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 4

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 2

Best rules found:

 1. int-discolor=none 581 ==> sclerotia=absent 581    conf:(1)
 2. mycelium=absent int-discolor=none 575 ==> sclerotia=absent 575    conf:(1)
 3. leaves=abnorm sclerotia=absent 548 ==> mycelium=absent 547    conf:(1)
 4. sclerotia=absent 625 ==> mycelium=absent 619    conf:(0.99)
 5. int-discolor=none 581 ==> mycelium=absent 575    conf:(0.99)
 6. int-discolor=none sclerotia=absent 581 ==> mycelium=absent 575    conf:(0.99)
 7. int-discolor=none 581 ==> mycelium=absent sclerotia=absent 575    conf:(0.99)
 8. leaf-malf=absent 554 ==> mycelium=absent 548    conf:(0.99)
 9. mycelium=absent 639 ==> sclerotia=absent 619    conf:(0.97)
10. leaves=abnorm mycelium=absent 567 ==> sclerotia=absent 547    conf:(0.96)
```