DATE

PAGE

ASSIGNMENT-1

FOR

Web Data Mining

Submitted to

Gopeekrishnan R

Associate Professor

Dept. of Computer Applications

SCE, Kottayam

Submitted by

Ribin Sara John

S_5 - RMCA

Roll No: 19

SCE, Kottayam

Submitted on

30/09/2021

Simulation

Q- Simulate the algorithm genRules (F) for the just found out frequent item sets.

TID	list of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Algorithm

Algorithm genRules (F) // F is the set of all frequent itemsets

1 for each frequent k-itemset fk in F, k ≥ 2 do

2 output every 1-item consequent rule of fk with confidence ≥ minconf and

support ← f_k.count/n // n is the total numbe of transactions in T

3 H_1 ← {consequents of all 1-item consequent rules derived from f_k above};

4 ap-genRules(f_k, H_1);

5 endfor

Procedure ap-genRules(f_k, H_m) // H_m is the set of

1 if (k > m+1) AND ($H_m \neq \phi$) then m-item consequents

2 H_{m+1} ← candidate-gen(H_m);

3 for each h_{m+1} in H_{m+1} do

4 conf ← f_k.count / ($f_k - h_{m+1}$).count;

5 if (conf ≥ minconf) then

6 output the rule ($f_k - h_{m+1}$) → h_{m+1} with confidence = conf and support = f_k.count/n; // n is the total numbel of transactions in T

7 else

8 delete h_{m+1} from H_{m+1};

9 end-for

10 ap-genRules (fk, Hm+1);

11 endif

Generation of strong association rule

#3

Itemset	Support count
{I1,I2,I3}	2
{I1,I2,I5}	2

Since the frequent itemsets satisfies the Apriori property ie, if an itemset has minimum support, then every non empty subset of this itemset also has minimum support.

Hence for Itemset = {I1,I2,I3} with min sup = 2 have the following itemsets and support counts.

Case 1: {I1,I2,I3}

Itemset	Sup. count
{I1}	6
{I2}	7
{I3}	6
{I1,I2}	4
{I1,I3}	4
{I2,I3}	4

Following are the various association rules that can be generated from this itemset,

1. $\{I1\} \rightarrow \{I2, I3\}$
2. $\{I2\} \rightarrow \{I1, I3\}$
3. $\{I3\} \rightarrow \{I1, I2\}$
4. $\{I1, I2\} \rightarrow \{I3\}$
5. $\{I1, I3\} \rightarrow \{I2\}$
6. $\{I2, I3\} \rightarrow \{I1\}$
7. $\{I1, I2, I3\} \rightarrow \phi$ (Not considering since it is a nullset)

* Confidence of Rule 1. $[\{I1\} \rightarrow \{I2, I3\}]$

$$\text{Confidence} = \frac{\text{sup.count}(\{I1, I2, I3\})}{\text{sup.count}(\{I1\})}$$

$$= \frac{2}{6} = \frac{1}{3} = 0.33 = 33.33\%$$

* Confidence of Rule 2. $[\{I2\} \rightarrow \{I1, I3\}]$

$$\text{Confidence} = \frac{\text{sup.count}(\{I1, I2, I3\})}{\text{sup.count}(\{I2\})}$$

$$= \frac{2}{7} = 0.2857 = 28.7\%$$

* Confidence of Rule 3 [{I3} → {I1,I2}]

$$\text{confidence} = \frac{\text{sup.count}(\{I1,I2,I3\})}{\text{sup.count}(\{I3\})}$$

$$= \frac{2}{6} = 33.3\%$$

* Confidence of Rule 4 [{I1,I2} → {I3}]

$$\text{confidence} = \frac{\text{sup.count}(\{I1,I2,I3\})}{\text{sup.count}(\{I1,I2\})}$$

$$= \frac{2}{4} = 50\%$$

* Confidence of Rule 5 [{I1,I3} → {I2}]

$$\text{confidence} = \frac{\text{sup.count}(\{I1,I2,I3\})}{\text{sup.count}(\{I1,I3\})}$$

$$= \frac{2}{4} = 50\%$$

* Confidence of Rule 6 [{I2,I3} → {I1}]

$$\text{confidence} = \frac{\text{sup.count}(\{I1,I2,I3\})}{\text{sup.count}(\{I2,I3\})}$$

$$= \frac{2}{4} = 50\%$$

Since none of the generated association rules have confidence ≥ minconf ie, 70%. These rules are not strong association rules.

Case 2:

Frequent itemset = {I1, I2, I5} with support count = 2

Subsets of items	Support count
{I1}	6
{I2}	7
{I5}	2
{I1, I2}	4
{I1, I5}	4
{I2, I5}	2

Following are the various association rules that can be generated from this itemset

1. {I1} → {I2, I5}
2. {I2} → {I1, I5}
3. {I5} → {I1, I2}
4. {I1, I2} → {I5}

5. $\{I1, I5\} \rightarrow \{I2\}$.

6. $\{I2, I5\} \rightarrow \{I1\}$

$\{I1, I2, I5\} \rightarrow \phi$ (since it is nullset)

* Confidence of Rule 1. $[\{I1\} \rightarrow \{I2, I5\})$

$$\text{Confidence} = \frac{\text{sup.count}(\{I1, I2, I5\})}{\text{sup.count}(\{I1\})}$$

$$= \frac{2}{6} = \frac{1}{3} = 33.3\%$$

* Confidence of Rule 2 $[\{I2\} \rightarrow \{I1, I5\})$

$$\text{Confidence} = \frac{\text{sup.count}(\{I1, I2, I5\})}{\text{sup.count}(\{I2\})}$$

$$= \frac{2}{7} = 28.7\%$$

* Confidence of Rule 3 $[\{I5\} \rightarrow \{I1, I2\})$

$$\text{Confidence} = \frac{\text{sup.count}(\{I1, I2, I3\})}{\text{sup.count}(\{I5\})}$$

$$= \frac{2}{2} = 100\%$$

* Confidence of Rule 4 [{I1,I2} → {I5})

$$\text{Confidence} = \frac{\text{sup.count}(\{I1,I2,I3\})}{\text{sup.count}(\{I1,I2\})}$$

$$= \frac{2}{4} = 50\%$$

* Confidence of Rule 5 [{I1,I5} → {I2}]

$$\text{Confidence} = \frac{\text{sup.count}(\{I1,I2,I3\})}{\text{sup.count}(\{I1,I5\})}$$

$$= \frac{2}{2} = 100\%$$

* Confidence of Rule 6 [{I2,I5} → {I1}]

$$\text{Confidence} = \frac{\text{sup.count}(\{I1,I2,I3\})}{\text{sup.count}(\{I2,I5\})}$$

$$= \frac{2}{2} = 100\%$$

Minimum confidence of association rules are 70%
Association rules having min.conf ≥ 70% are strong association rules.

Here strong association rules that can be generated from the given market basket data frequent itemset F_3 is,

{I5} → {I1, I2} with conf = 100%

{I1, I5} → {I2} with conf = 100%

{I2, I5} → {I1} conf = 100%.

F2

Itemset	sup.count
{I1, I2}	4
{I1, I3}	4
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I2, I5}	2

i) {I1, I2}

Subsets are {I1} {I2} and {I1, I2}

① {I1} → {I2}

2. {I2} → {I1}

3 {I1, I2} → ϕ - Not possible

Confidence of Rule ① {I1} → {I2}

$$\text{Conf} = \frac{\text{sup.count}(\{I1, I2\})}{\text{sup count}(\{I1\})}$$

$= \frac{4}{6} = \frac{2}{3} = 66\%$

* Confidence of Rule ② $\{I2\} \rightarrow \{I1\}$

$Conf = \frac{4}{7} = 57\%$

ii) $\{I1, I3\}$

③ $\{I1\} \rightarrow \{I3\}$

④ $\{I3\} \rightarrow \{I1\}$

Confidence of ③ $\{I1\} \rightarrow \{I3\}$

$Conf = \frac{4}{6} = 66\%$

Confidence of ④ $\{I3\} \rightarrow \{I1\}$

$Conf = \frac{4}{6} = 66\%$

iii) $\{I1, I5\}$

⑤ $\{I1\} \rightarrow \{I5\}$

⑥ $\{I5\} \rightarrow \{I1\}$

Confidence of ⑤ $\{I1\} \rightarrow \{I5\}$

$Conf = \frac{2}{6} = 33.3\%$

iv) {I2, I3}

(7) {I2} → {I3}

(8) {I3} → {I2}

Conf. of Rule (7) {I2} → {I3}

$\text{Conf} = \frac{4}{7} = 57\%$

Confidence of Rule (8) {I3} → {I2}

$\text{Conf} = \frac{4}{6} = 66\%$

v) {I2, I4}

(9) {I2} → {I4}

(10) {I4} → {I2}

Confidence of Rule (9) {I2} → {I4}

$\text{Conf} = \frac{2}{7} = 28\%$

Confidence of Rule (10) {I4} → {I2}

$\text{Conf} = \frac{2}{2} = 100\%$

vi) {I2, I5}

(11) {I2} → {I5}

(12) {I5} → {I2}

Confidence of Rule (11) {I2} → {I5}

$conf = \frac{2}{7} = 28.7\%$

Confidence of Rule (12) {I5} → {I2}

$conf = \frac{2}{2} = 100\%$

minconf ≥ 70%. hence, strong association rules from F2 are

{I4} → {I2} conf = 100%

{I5} → {I2} conf = 100%

F1

Itemset	Support count
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2

Since all these itemsets are satisfying Apriori properties

their subsets will also have minimum support count. But here F_1 is 1-frequent itemset whose subsets are the same itemset itself which generates ϕ, Hence the algorithm is terminated.

Strong association rules generated for this market basket data is as follows,

$\{I4\} \rightarrow \{I2\}$

$\{I5\} \rightarrow \{I2\}$

$\{I5\} \rightarrow \{I1, I2\}$

$\{I1, I5\} \rightarrow \{I2\}$

$\{I2, I5\} \rightarrow \{I1\}$