Practical -3

3) Apply Pre-Processing on Data set using weka.

title: Apply Pre-Processing on Data set

Description:- Real world database are highly influenced to noise, missing and inconsistancy due to their queue size so the data can be Pre-processed to improve the quality of data and missing results and it also improves the efficiency.

Pre-processing Techniques:-

a) missing value:- This situation arises when some data is missing in data set. It can be handle in various ways

a.1) For numerical attribute we can use attribute mean to fill in the missing value.

Steps : 1) open notepad.
2) create training data set
@ relation Practical-3
@ attribute variable_1 numeric
@ attribute variable_2 string
@ data
12.34 , "aa"
34 , ?
44 , "cc"

-433 , "ff"
? , ?
6743 , "df"
3 , "fg"
4 , ?
3 , "g"
34 , "dd"

3) open weka
4) click on work bench / explore
5) click on open file
6) select "Practical 3 - arff" file
7) click on choose button
8) select unsupervised filter
9) select attribute
10) select Replace missing value
11) make ignore class is true
12) click on ok button
13) click on apply button
14) click on edit button to view the result

| "elation : Practical_3 | | |
|---|---|---|
| No. | 1: variable-1 numeric | 2: variable 2 String |
| 1 | 12.34 | aa |
| 2 | 34.0 | |
| 3 | 44.0 | cc |
| 4 | -433.0 | ff |
| 5 | 716.037777... | |
| 6 | 6743.0 | df |

VISION

| | | | |
|---|---|---|---|
| 7 | 3.0 | fg | |
| 8 | 4.0 | | |
| 9 | 3.0 | | |
| 10 | 34.0 | dd | |

Q.2) Replace all missing values for nominal, string numeric and data attributes in the dataset with user-supplied constant values.

Steps : 1) open notepad

2) create training data set

ⓐ Solution Practical-3

ⓑ attribute variable-1

ⓒ attribute variable-2

ⓓ data

| | |
|---|---|
| 12.34 | "aa" |
| 34 | "g" |
| 44 | "cc" |
| -433 | "ff" |
| ? | ? |
| 6743 | "df" |
| 3 | "fg" |
| 4 | ? |
| 3 | "g" |
| 34 | "dd" |

3) open weka

4) click on workbench / explorer

5) click on openfile

6) select "practical-3.arff" file.

7) click on choose button.

8) select unsupervised filter
9) select attribute
10) select Replace Missing With User Constant
11) modify attribute option, with value 1,2
12) modify ignore class option is true.
13) modify nominal String Replacement value option with value "sb"
14) modify numeric Replacement value option with value "555"
15) click on ok button
16) click on "Class: variable_2 (Str)" drop down menu.
17) click on No Class
18) click on Apply button
19) click on Edit button to view the result.

Q.3) A filter that can be used to introduce missing values in a dataset.

Steps: 1) open notepad
2) create training dataset
@ relation Practical-3
@ attribute v-1 numeric
@ attribute v-2 string
@ attribute v-3 {y,n}
@ data
12.34, "aa", y
34, ?, y
44, "sd", n
-433, "ff", y

?, ?, ?
6743, "ff", ?
3, "fg", n
4, ?, ?
3, ?, n
34, "dd", ?

3) open weka
4) click on workbench / explorer
5) click on open file
6) select "Practical-3.arff" file
7) click on choose button.
8) select unsupervised filter.
9) select attribute
10) select Replace with missing Value
11) modify attribute Indies option with value 1,2,3
12) modify ignoreClass option with true
13) click on ok button
14) click on apply button.
15) click on Edit button to view the result.

Relation :- Practical-3

| No. | 1: V-1 Numeric | 2: V-2 string | 3: V-3 nominal. |
|---|---|---|---|
| 1 | 12.34.0 | aa | y |
| 2 | 34.0 | | y |
| 3. | 44.0 | sd | n |
| 4 | -433.0 | ff | y |
| 5 | | | |
| 6 | 6743.0 | df | n |

| | | | |
|---|---|---|---|
| 7 | 3.0 | fg | n |
| 8 | 4.0 | | |
| 9 | 3.0 | | n |
| 10 | 34.0 | dd | |

4.2) Result table

Relation : Practical-3

| No | 1: variable-1 numeric | 2: variable-2 string |
|---|---|---|
| 1 | 12.34.0 | aa |
| 2 | 34.0 | sb |
| 3 | 44.0 | acc |
| 4 | -433.0 | ff |
| 5 | 555.0 | sb |
| 6 | 6743.0 | df |
| 7 | 3.0 | fg |
| 8 | 4.0 | sb |
| 9 | 3.0 | g |
| 10 | 34.0 | dd |

b) Transformation (Normalization):

Normalizes all numeric values in the given dataset from (start from the class attribute, if set)

Steps: 1) open notpad

2) Create training data set

@ relation practical-3

@ attribute v-1 numeric

@ attribute v-2 string

@ attribute v-3 {y, n}
@ data

| | | |
|---|---|---|
| 12.34 | "aa" | y |
| 34 | ? | y |
| 44 | "sd" | n |
| -433 | "ft" | y |
| ? | ? | ? |
| 6743 | "df" | n |
| 3 | "fg" | n |
| 4 | ? | ? |
| 3 | ? | n |
| 34 | "dd" | ? |

3) Open weka
4) Click on workbench / explorer.
5) Click on open file
6) Select "Practical-3.arff" file
7) Click on choose button
8) Select unsupervised button.
9) Select attribute
10) Select Normalize
11) Modify attribute ignoreClass is True
12) Modify attribute scale is with value 1.0
13) Modify translation attribute with value 0.0
14) Click on ok button
15) Click on "Class: v-3 (nominal)" dropdown menu
16) Click on No Class
17) Click on Apply button
18) Click on Edit button to view the result.

Relation: Practical-3

| No. | 1: V_1 | 2: V_2 | 3: V_3 |
|---|---|---|---|
| | Numeric | String | Nominal |
| 1 | 0.0620.... | qq | y |
| 2 | 0.0650..... | | y |
| 3 | 0.0664.... | sd | n |
| 4 | 0.0 | ff | y |
| 5 | | | |
| 6 | 1.0 | df | |
| 7 | 0.0607.... | fg | n |
| 8 | 0.0608.... | | |
| 9 | 0.0607.... | | n |
| 10 | 0.0650.... | dd | |

Relation: weather

| No. | 1: outlook | 2: temparature | 3: humidity | 4: windy | 5: **play** |
|---|---|---|---|---|---|
| | Nominal | Numeric | Numeric | Nominal | Nominal |
| 1 | sunny | 85.0 | 85.0 | false | no |
| 2 | overcast | 80.0 | 90.0 | true | no |
| 3 | sunny | 83.0 | 86.0 | false | yes |
| 4 | rainy | 70.0 | 86.0 | false | yes |
| 5 | rainy | 68.0 | 80.0 | false | yes |
| 6 | rainy | 65.0 | 70.0 | true | no |
| 7 | overcast | 64.0 | 65.0 | false | yes |
| 8 | sunny | 72.0 | 95.0 | true | no |
| 9 | sunny | 69.0 | 70.0 | false | yes |
| 10 | rainy | 75.0 | 80.0 | false | yes |

Viewer

Relation: weather-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0

| No. | 1: outlook | 2: temparature | 3: humidity | 4: windy | 5: **play** |
|---|---|---|---|---|---|
| | Nominal | Numeric | Numeric | Nominal | Nominal |
| 1 | sunny | 1.0 | 0.66666... | false | no |
| 2 | overcast | 0.76190476... | 0.83333... | true | no |
| 3 | sunny | 0.90476190... | 0.7 | false | yes |
| 4 | rainy | 0.28571428... | 0.7 | false | yes |
| 5 | rainy | 0.19047619... | 0.5 | false | yes |
| 6 | rainy | 0.04761904... | 0.16666... | true | no |
| 7 | overcast | 0.0 | 0.0 | false | yes |
| 8 | sunny | 0.38095238... | 1.0 | true | no |
| 9 | sunny | 0.23809523... | 0.16666... | false | yes |
| 10 | rainy | 0.52380952... | 0.5 | false | yes |

# Practical - 4

4) Implement Association rule mining on dataset using apriori algorithm.

Title : Apriori algorithm

Description : Iteratively reduces the minimum support until it finds the required number of rules with the given minimum confidence. The algorithm has an option to mine class association rules. It is adapted as explained in the second reference.

Procedure :-

Steps :- 1) open weka tutorial
2) write training dataset
   ① relation Practical-4
   ② ~~relation~~ attribute $I_1$ {T, ?}
   ③ attribute $I_2$ {T, ?}
   ④ attribute $I_3$ {T, ?}
   ⑤ attribute $I_4$ {T, ?}
   ⑥ attribute $I_5$ {T, ?}
   ⑦ data:

T, T, ?, ?, T
?, T, ?, T, ?
?, T, T, ?, ?
T, T, ?, T, ?
T, ?, T, ?, ?
?, T, T, ?, ?

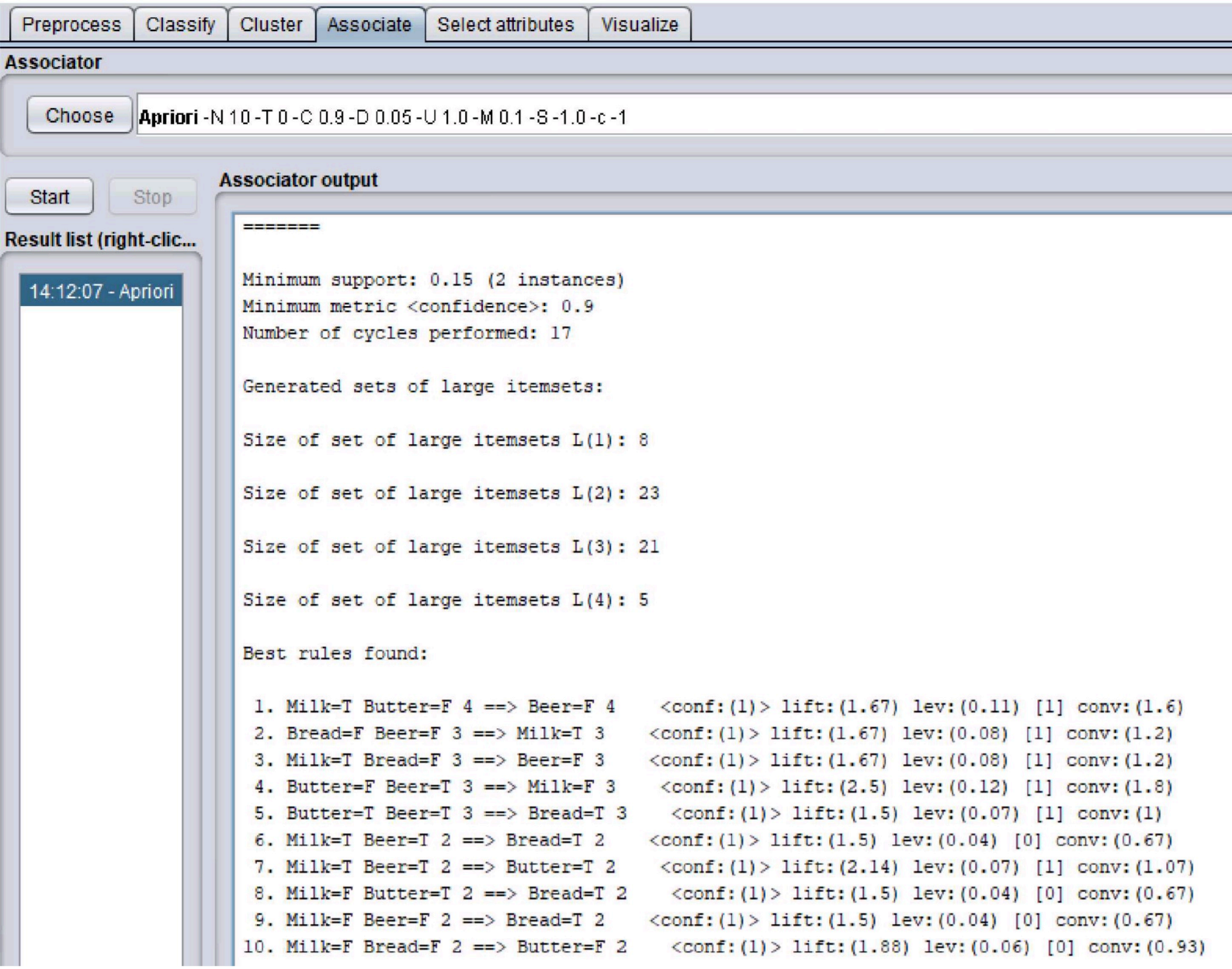
T, ?, T, ?, ?
T, T, T, ?, T
T, T, T, ?, ?

3) after that the file is saved with .arff file format
4) open weka
5) click on workbench / explore
6) Click on open file
7) select "practical-4.arff" file
8) click on Associate
9) click on choose button.
10) click on Apriori
11) modify lowerBoundMinSupport option with value 0.1
12) modify minmetric option with value 0.9
13) modify outputitemsets option with value true
14) click on ok button.
15) click on start button to view the result.

Run information:
Minimum support: 0.16 (1 instances)
minimum metric <confidence>: 0.9.
Number of cycles performed: 17
Generated sets of large itemsets:
size of set of large itemsets ($L_1$): 5
size of set of large itemsets ($L_2$): 8
size of set of large itemsets ($L_3$): 5
size of set of large itemsets ($L_4$): 1.

Best Rules found:

1. IS = T2 ⇒ E1 = T2 <conf:(1)> Lift:(LS) lev:(0.07)(D)
2.
3.
4.-
5. The list of items as the input given.
6.
7.
8.
9.


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-clic...
14:12:07 - Apriori
Associator output
=======
Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17
Generated sets of large itemsets:
Size of set of large itemsets L(1): 8
Size of set of large itemsets L(2): 23
Size of set of large itemsets L(3): 21
Size of set of large itemsets L(4): 5
Best rules found:
1. Milk=T Butter=F 4 ==> Beer=F 4 <conf:(1)> lift:(1.67) lev:(0.11) [1] conv:(1.6)
2. Bread=F Beer=F 3 ==> Milk=T 3 <conf:(1)> lift:(1.67) lev:(0.08) [1] conv:(1.2)
3. Milk=T Bread=F 3 ==> Beer=F 3 <conf:(1)> lift:(1.67) lev:(0.08) [1] conv:(1.2)
4. Butter=F Beer=T 3 ==> Milk=F 3 <conf:(1)> lift:(2.5) lev:(0.12) [1] conv:(1.8)
5. Butter=T Beer=T 3 ==> Bread=T 3 <conf:(1)> lift:(1.5) lev:(0.07) [1] conv:(1)
6. Milk=T Beer=T 2 ==> Bread=T 2 <conf:(1)> lift:(1.5) lev:(0.04) [0] conv:(0.67)
7. Milk=T Beer=T 2 ==> Butter=T 2 <conf:(1)> lift:(2.14) lev:(0.07) [1] conv:(1.07)
8. Milk=F Butter=T 2 ==> Bread=T 2 <conf:(1)> lift:(1.5) lev:(0.04) [0] conv:(0.67)
9. Milk=F Beer=F 2 ==> Bread=T 2 <conf:(1)> lift:(1.5) lev:(0.04) [0] conv:(0.67)
10. Milk=F Bread=F 2 ==> Butter=F 2 <conf:(1)> lift:(1.88) lev:(0.06) [0] conv:(0.93)

# Practical-5

5) Implement Association rule mining on dataset using Fp-Tree Growth algorithm and compare with apriori algorithm

title : Fp-Tree Growth Algorithm

Description: Allows frequent itemset discovery without candidate itemset generation.
- It is a two step method.

Step-1 : Build a compact data structure called the Fp-Tree.

Step-2 : Extracts frequent item sets directly from the Fp-tree.

Procedure :

Steps- 1) Open notepad
2) write training dataset.

@relation Practical-5
@attribute $I_1$ {T, ?}
@attribute $I_2$ {T, ?}
@attribute $I_3$ {T, ?}
@attribute $I_4$ {T, ?}
@attribute $I_5$ {T, ?}
@data

T, T, ?, ?, T
?, T, ?, T, ?
?, T, T, ?, ?
T, T, ?, T, ?
T, ?, T, ?, ?
?, T, T, ?, ?
T, ?, T, ?, ?
T, T, T, ?, T
T, T, T, ?, ?

3) After that the file is saved with .arff
4) file format
5) open weka
6) click on workbench / explore
7) click on open file
8) select "Practical_5.arff" file
9) click on Associate
10) click on choose button
11) click on FP Growth option
12) click on tab beside choose button then dialog box, opened then set values according to your need and click ok
13) click on start button and output will be displayed on the right side of the window.

Result table:

=== Run Information ===

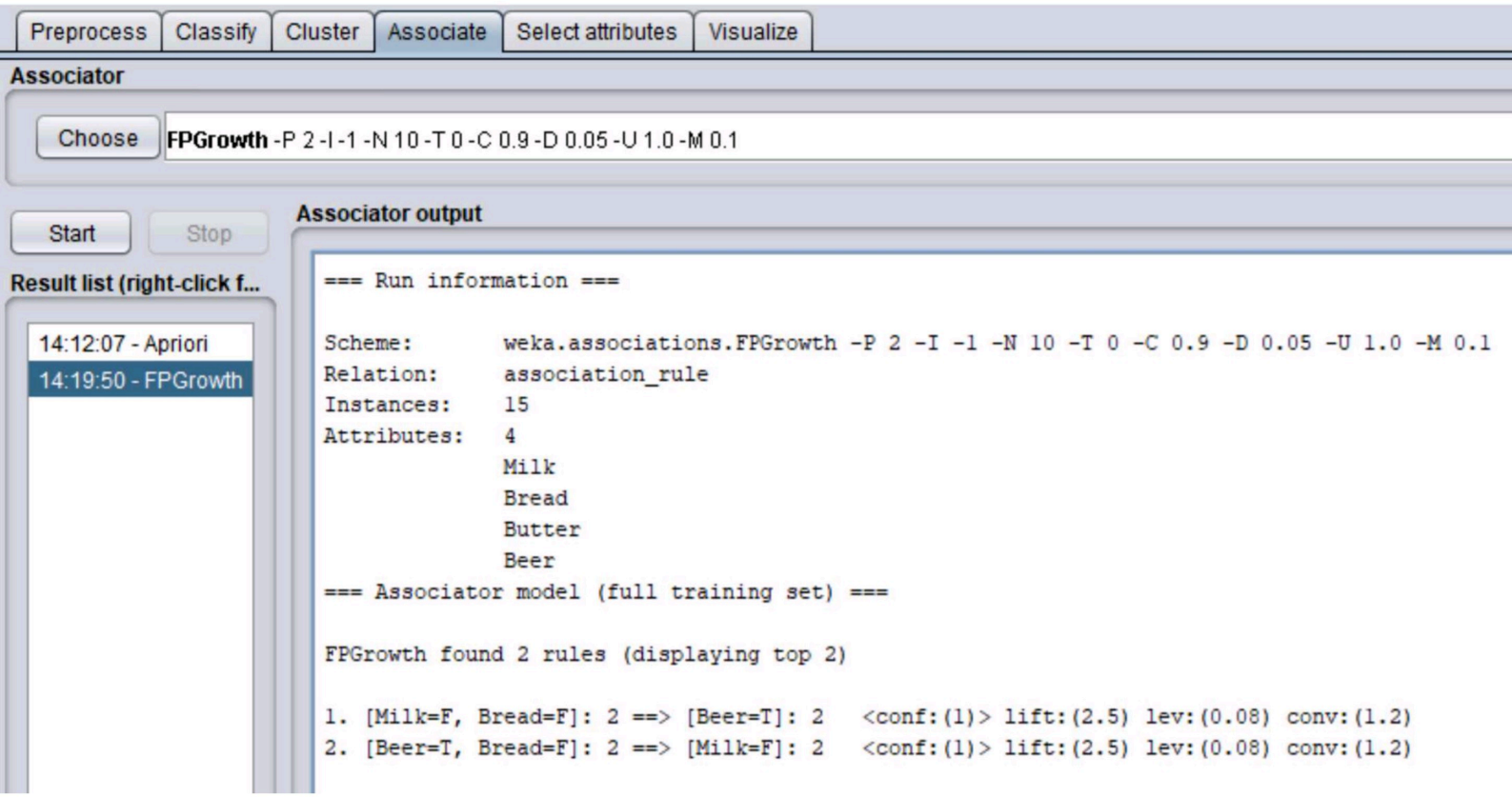
Scheme: weka.associations.FPGrowth -P 2 -I -N 10
Relation: Itemset
Instances: 9
Attribute: 5
$I_1$
$I_2$
$I_3$
$I_4$
$I_5$

=== Associator model (full training set) ===

FP Growth found 6 rules (displaying top 6)

1. $[I_5 = T]$: 2 ⇒ $[I_2 = T]$: 2 <conf(1)> lift (1, 2.9) lev (0, 0.5)
2. $[I_4 = T]$:
3. $[I_3 = T]$:
4. $[I_5 = T]$:
5. $[I_2 = T, I_5 = T]$:
6. $[I_1 = T, I_5 = T]$:

other Items


Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Start
Stop
Result list (right-click f...
14:12:07 - Apriori
14:19:50 - FPGrowth
Associator output
=== Run information ===
Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation: association_rule
Instances: 15
Attributes: 4
Milk
Bread
Butter
Beer
=== Associator model (full training set) ===
FPGrowth found 2 rules (displaying top 2)
1. [Milk=F, Bread=F]: 2 ==> [Beer=T]: 2 <conf:(1)> lift:(2.5) lev:(0.08) conv:(1.2)
2. [Beer=T, Bread=F]: 2 ==> [Milk=F]: 2 <conf:(1)> lift:(2.5) lev:(0.08) conv:(1.2)