

Exercise 1

Apriori for the given dataset in Canvas Numerical manual solve:

Apriori Algorithm Numerical Experiment 1

Trans-id	Itemlist
1	A, B, D, K
2	A, B, C, D, E
3	A A, B, C, E
4	A, B, D

We have to take minimum support as $N = 60\%$
 $= 0.6$

minimum confidence as $Z = 80\%$
 $= 0.8$

ID	A	B	C	D	E	K
1	1	1	0	1	0	1
2	1	1	1	1	1	0
3	1	1	1	0	1	0
4	1	1	0	1	0	0

$K=1$

$C_1 \Rightarrow$ Candidate itemsets with 1-itemsets with their respective minsup value

Set	A	B	C	D	E	K
SupCount	4	4	2	3	2	1

Only considering ones which are greater than ≥ 2

$L_1 \Rightarrow \{A, B, D\}$ [Frequent 1-itemsets from C_1]

$L_2 \Rightarrow \{ \{A, B\}, (A, C), (A, D), (A, E), (B, C), (B, D), (B, E), (C, E) \}$ [Frequent 2-itemsets]

$\begin{matrix} 4 & 2 & 3 & 2 & 2 & 3 \\ 2 & 2 \end{matrix}$

$$I_2 \quad b \rightarrow c \quad (b, c) \quad \frac{2}{a}$$

$$M_2 = \{1/1 \text{ or } 1/2\}$$

$K=2$

Similarly for L_3

Itemset	Support Count
$\{A, B, C\}$	2
$\{A, B, D\}$	3
$\{A, B, E\}$	2
$\{A, C, E\}$	2
$\{B, C, E\}$	2

Similarly for L_4

Itemset	Support Count
$\{A, B, C, E\}$	2

Frequent itemsets $L = L_1 \cup L_2 \cup L_3 \cup L_4$

$\{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{A, B\}, \{A, C\}, \{A, D\}, \{A, E\},$
 $\{B, C\}, \{B, D\}, \{B, E\}, \{C, E\}, \{A, B, C\}, \{A, B, D\},$
 $\{A, B, E\}, \{A, C, E\}, \{B, C, E\}\}$

$i = \{A, B, C, D, E, K\}$

Marking the lattice structure with min conf of 0.8
 Steps \Rightarrow

For each $i \in L = L_1 \cup L_2 \cup L_3 \cup L_4$

Generate all the rules in the form of

$R = \{L \rightarrow R \mid R \in i, |R| \geq 1, L = (i - R), \sigma(L \rightarrow R) \geq \text{min conf}\}$

For 2-itemssets

~~Rule $A \rightarrow B$~~

Rules	Confidence	Good Rule
$A \rightarrow B$	$\frac{4}{4} = 1$	$[\geq 0.8] \checkmark$
$B \rightarrow A$	1	$[\geq 0.8] \checkmark$
$A \rightarrow C$	0.5	X
$C \rightarrow A$	1	$[\geq 0.8] \checkmark$
$A \rightarrow D$	0.75	X
$D \rightarrow A$	1	$[\geq 0.8] \checkmark$
$A \rightarrow E$	0.5	X
$E \rightarrow A$	1	$[\geq 0.8] \checkmark$
$B \rightarrow C$	0.5	X
$C \rightarrow B$	1	$[\geq 0.8] \checkmark$
$B \rightarrow D$	0.75	X
$D \rightarrow B$	1	$[\geq 0.8] \checkmark$
$B \rightarrow E$	0.5	X
$C \rightarrow E$	1	$[\geq 0.8] \checkmark$
$E \rightarrow B$	1	$[\geq 0.8] \checkmark$

As we can see a total of 9 rules with one item in the head.

Some will be repeated for other rules and the number some of the other best rules are

$b, d \rightarrow a$	Confidence 1
$a, d \rightarrow b$	confidence 1
$d \rightarrow a, b$	confidence 1
$b, c \rightarrow a$	Confidence 1
$a, c \rightarrow b$	Confidence 1

and so on.

$b, c, e \rightarrow a$	Confidence	1
$a, c, e \rightarrow b$	Confidence	1
$a, b, c \rightarrow e$	Confidence	1
$c \rightarrow a, b, e$	"	1
$e \rightarrow a, b, c$	"	1
$c, e \rightarrow a, b$	"	1
$c \rightarrow b, e$	"	1
and so on.		

Exercise 2

Apriori for the given dataset in Canvas

The 'database' below has four transactions. What association rules can be found in this set, if the minimum support (i.e coverage) is 60% and the minimum confidence (i.e. accuracy) is 80% ?

Trans_id	Itemlist
T1	{K, A, D, B}
T2	{D, A, C, E, B}
T3	{C, A, B, E}
T4	{B, A, D}

Solution:

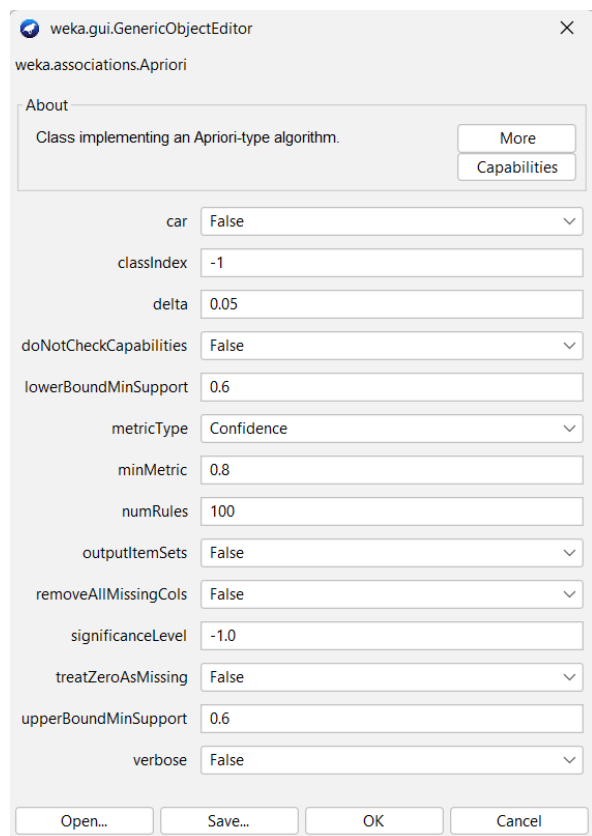
Data in Weka:

Viewer						
Relation: exercise						
No.	1: exista Nominal	2: existb Nominal	3: existc Nominal	4: existd Nominal	5: existe Nominal	6: existk Nominal
1	TRUE	TRUE		TRUE		TRUE
2	TRUE	TRUE	TRUE	TRUE	TRUE	
3	TRUE	TRUE	TRUE		TRUE	
4	TRUE	TRUE		TRUE		

Configuration:

MinSup set to 0.6

MinConf set to 0.8



Log Output after running Weka:

=== Run information ===

Scheme: weka.associations.Apriori -N 100 -T 0 -C 0.8 -D 0.05 -U 0.6 -M 0.6 -S -1.0 -c -1

Relation: exercise

Instances: 4

Attributes: 6

exista

existb

existc

existd

existe

existk

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

Apriori

=====

Minimum support: 0.6 (2 instances)

Minimum metric <confidence>: 0.8

Number of cycles performed: 8

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. existc=TRUE 2 ==> exista=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. existe=TRUE 2 ==> exista=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. existc=TRUE 2 ==> existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. existe=TRUE 2 ==> existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. existe=TRUE 2 ==> existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
6. existc=TRUE 2 ==> existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
7. existb=TRUE existc=TRUE 2 ==> exista=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. exista=TRUE existc=TRUE 2 ==> existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. existc=TRUE 2 ==> exista=TRUE existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. existb=TRUE existe=TRUE 2 ==> exista=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
11. exista=TRUE existe=TRUE 2 ==> existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
12. existe=TRUE 2 ==> exista=TRUE existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
13. existc=TRUE existe=TRUE 2 ==> exista=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
14. exista=TRUE existe=TRUE 2 ==> existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
15. exista=TRUE existc=TRUE 2 ==> existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
16. existe=TRUE 2 ==> exista=TRUE existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
17. existc=TRUE 2 ==> exista=TRUE existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
18. existc=TRUE existe=TRUE 2 ==> existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
19. existb=TRUE existe=TRUE 2 ==> existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
20. existb=TRUE existc=TRUE 2 ==> existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
21. existe=TRUE 2 ==> existb=TRUE existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
22. existc=TRUE 2 ==> existb=TRUE existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
23. existb=TRUE existc=TRUE existe=TRUE 2 ==> exista=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
24. exista=TRUE existc=TRUE existe=TRUE 2 ==> existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
25. exista=TRUE existb=TRUE existe=TRUE 2 ==> existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
26. exista=TRUE existb=TRUE existc=TRUE 2 ==> existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
27. existc=TRUE existe=TRUE 2 ==> exista=TRUE existb=TRUE 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
28. existb=TRUE existe=TRUE 2 ==> exista=TRUE existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
29. existb=TRUE existc=TRUE 2 ==> exista=TRUE existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
30. exista=TRUE existe=TRUE 2 ==> existb=TRUE existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
31. exista=TRUE existc=TRUE 2 ==> existb=TRUE existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
32. existe=TRUE 2 ==> exista=TRUE existb=TRUE existc=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)
33. existc=TRUE 2 ==> exista=TRUE existb=TRUE existe=TRUE 2 <conf:(1)> lift:(2) lev:(0.25) [1] conv:(1)

Conclusion:

We can see that from the numerical we solved. Some of the rules we wrote matched the console log output and a total of 33 rules were generated. The confidence value of itemsets shown above match our calculations.

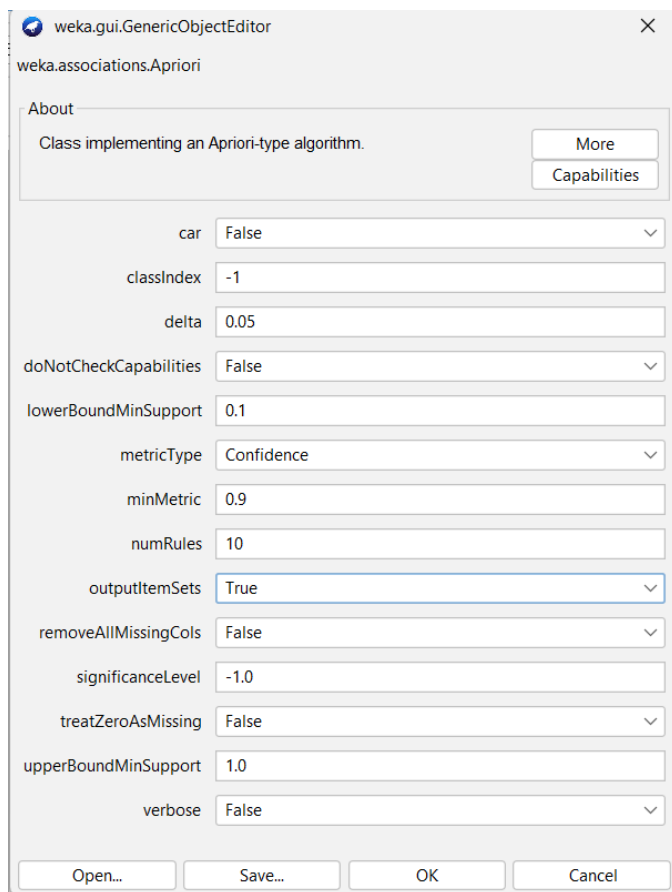
Exercise 3

Apriori for the nominal weather dataset

Data in Weka:

Relation: weather.symbolic					
No.	1: outlook Nominal	2: temperature Nominal	3: humidity Nominal	4: windy Nominal	5: play Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Configuration 1:



=== Run information ===

Scheme: weka.associations.Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

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

Apriori

=====

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): 12

Large Itemsets L(1):

outlook=sunny 5
outlook=overcast 4
outlook=rainy 5
temperature=hot 4
temperature=mild 6
temperature=cool 4
humidity=high 7
humidity=normal 7
windy=TRUE 6
windy=FALSE 8
play=yes 9
play=no 5

Size of set of large itemsets L(2): 47

Large Itemsets L(2):

outlook=sunny temperature=hot 2
outlook=sunny temperature=mild 2
outlook=sunny humidity=high 3
outlook=sunny humidity=normal 2
outlook=sunny windy=TRUE 2
outlook=sunny windy=FALSE 3
outlook=sunny play=yes 2
outlook=sunny play=no 3
outlook=overcast temperature=hot 2
outlook=overcast humidity=high 2
outlook=overcast humidity=normal 2
outlook=overcast windy=TRUE 2
outlook=overcast windy=FALSE 2
outlook=overcast play=yes 4
outlook=rainy temperature=mild 3
outlook=rainy temperature=cool 2
outlook=rainy humidity=high 2
outlook=rainy humidity=normal 3
outlook=rainy windy=TRUE 2
outlook=rainy windy=FALSE 3
outlook=rainy play=yes 3
outlook=rainy play=no 2
temperature=hot humidity=high 3
temperature=hot windy=FALSE 3
temperature=hot play=yes 2
temperature=hot play=no 2
temperature=mild humidity=high 4
temperature=mild humidity=normal 2
temperature=mild windy=TRUE 3
temperature=mild windy=FALSE 3

temperature=mild play=yes 4
temperature=mild play=no 2
temperature=cool humidity=normal 4
temperature=cool windy=TRUE 2
temperature=cool windy=FALSE 2
temperature=cool play=yes 3
humidity=high windy=TRUE 3
humidity=high windy=FALSE 4
humidity=high play=yes 3
humidity=high play=no 4
humidity=normal windy=TRUE 3
humidity=normal windy=FALSE 4
humidity=normal play=yes 6
windy=TRUE play=yes 3
windy=TRUE play=no 3
windy=FALSE play=yes 6
windy=FALSE play=no 2

Size of set of large itemsets L(3): 39

Large Itemsets L(3):

outlook=sunny temperature=hot humidity=high 2
outlook=sunny temperature=hot play=no 2
outlook=sunny humidity=high windy=FALSE 2
outlook=sunny humidity=high play=no 3
outlook=sunny humidity=normal play=yes 2
outlook=sunny windy=FALSE play=no 2
outlook=overcast temperature=hot windy=FALSE 2
outlook=overcast temperature=hot play=yes 2
outlook=overcast humidity=high play=yes 2
outlook=overcast humidity=normal play=yes 2
outlook=overcast windy=TRUE play=yes 2
outlook=overcast windy=FALSE play=yes 2
outlook=rainy temperature=mild humidity=high 2
outlook=rainy temperature=mild windy=FALSE 2
outlook=rainy temperature=mild play=yes 2
outlook=rainy temperature=cool humidity=normal 2
outlook=rainy humidity=normal windy=FALSE 2
outlook=rainy humidity=normal play=yes 2
outlook=rainy windy=TRUE play=no 2
outlook=rainy windy=FALSE play=yes 3
temperature=hot humidity=high windy=FALSE 2
temperature=hot humidity=high play=no 2
temperature=hot windy=FALSE play=yes 2
temperature=mild humidity=high windy=TRUE 2
temperature=mild humidity=high windy=FALSE 2
temperature=mild humidity=high play=yes 2
temperature=mild humidity=high play=no 2

temperature=mild humidity=normal play=yes 2
 temperature=mild windy=TRUE play=yes 2
 temperature=mild windy=FALSE play=yes 2
 temperature=cool humidity=normal windy=TRUE 2
 temperature=cool humidity=normal windy=FALSE 2
 temperature=cool humidity=normal play=yes 3
 temperature=cool windy=FALSE play=yes 2
 humidity=high windy=TRUE play=no 2
 humidity=high windy=FALSE play=yes 2
 humidity=high windy=FALSE play=no 2
 humidity=normal windy=TRUE play=yes 2
 humidity=normal windy=FALSE play=yes 4

Size of set of large itemsets L(4): 6

Large Itemsets L(4):

outlook=sunny temperature=hot humidity=high play=no 2
 outlook=sunny humidity=high windy=FALSE play=no 2
 outlook=overcast temperature=hot windy=FALSE play=yes 2
 outlook=rainy temperature=mild windy=FALSE play=yes 2
 outlook=rainy humidity=normal windy=FALSE play=yes 2
 temperature=cool humidity=normal windy=FALSE play=yes 2

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
2. temperature=cool 4 ==> humidity=normal 4 <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
3. humidity=normal windy=FALSE 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
4. outlook=sunny play=no 3 ==> humidity=high 3 <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
5. outlook=sunny humidity=high 3 ==> play=no 3 <conf:(1)> lift:(2.8) lev:(0.14) [1] conv:(1.93)
6. outlook=rainy play=yes 3 ==> windy=FALSE 3 <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3 <conf:(1)> lift:(1.56) lev:(0.08) [1] conv:(1.07)
8. temperature=cool play=yes 3 ==> humidity=normal 3 <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
9. outlook=sunny temperature=hot 2 ==> humidity=high 2 <conf:(1)> lift:(2) lev:(0.07) [1] conv:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2 <conf:(1)> lift:(2.8) lev:(0.09) [1] conv:(1.29)

Configuration 2:

weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm.

More

Capabilities

car: False

classIndex: -1

delta: 0.05

doNotCheckCapabilities: False

lowerBoundMinSupport: 0.2

metricType: Confidence

minMetric: 0.9

numRules: 10

outputItemSets: True

removeAllMissingCols: False

significanceLevel: -1.0

treatZeroAsMissing: False

upperBoundMinSupport: 1.0

verbose: False

Open... Save... OK Cancel

Log output:

=== Run information ===

Scheme: weka.associations.Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

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

Apriori

=====

Minimum support: 0.2 (3 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Large Itemsets L(1):

outlook=sunny 5

outlook=overcast 4

outlook=rainy 5

temperature=hot 4

temperature=mild 6

temperature=cool 4

humidity=high 7

humidity=normal 7

windy=TRUE 6

windy=FALSE 8

play=yes 9

play=no 5

Size of set of large itemsets L(2): 26

Large Itemsets L(2):

outlook=sunny humidity=high 3

outlook=sunny windy=FALSE 3

outlook=sunny play=no 3

outlook=overcast play=yes 4

outlook=rainy temperature=mild 3

outlook=rainy humidity=normal 3

outlook=rainy windy=FALSE 3

outlook=rainy play=yes 3

temperature=hot humidity=high 3

temperature=hot windy=FALSE 3

temperature=mild humidity=high 4

temperature=mild windy=TRUE 3

temperature=mild windy=FALSE 3

temperature=mild play=yes 4

temperature=cool humidity=normal 4

temperature=cool play=yes 3

humidity=high windy=TRUE 3

humidity=high windy=FALSE 4

humidity=high play=yes 3

humidity=high play=no 4

humidity=normal windy=TRUE 3

humidity=normal windy=FALSE 4

humidity=normal play=yes 6

windy=TRUE play=yes 3

windy=TRUE play=no 3

windy=FALSE play=yes 6

Size of set of large itemsets L(3): 4

Large Itemsets L(3):

outlook=sunny humidity=high play=no 3

outlook=rainy windy=FALSE play=yes 3

temperature=cool humidity=normal play=yes 3

humidity=normal windy=FALSE play=yes 4

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
2. temperature=cool 4 ==> humidity=normal 4 <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
3. humidity=normal windy=FALSE 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
4. outlook=sunny play=no 3 ==> humidity=high 3 <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
5. outlook=sunny humidity=high 3 ==> play=no 3 <conf:(1)> lift:(2.8) lev:(0.14) [1] conv:(1.93)
6. outlook=rainy play=yes 3 ==> windy=FALSE 3 <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3 <conf:(1)> lift:(1.56) lev:(0.08) [1] conv:(1.07)
8. temperature=cool play=yes 3 ==> humidity=normal 3 <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)

We can see that the number of best rules decreased in the second configuration

Configuration 3:

weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm.

More

Capabilities

car: False

classIndex: -1

delta: 0.05

doNotCheckCapabilities: False

lowerBoundMinSupport: 0.2

metricType: Confidence

minMetric: 0.3

numRules: 10

outputItemSets: True

removeAllMissingCols: False

significanceLevel: -1.0

treatZeroAsMissing: False

upperBoundMinSupport: 1.0

verbose: False

Open... Save... OK Cancel

Log output:

=== Run information ===

Scheme: weka.associations.Apriori -I -N 10 -T 0 -C 0.3 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

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

Apriori

=====

Minimum support: 0.3 (4 instances)

Minimum metric <confidence>: 0.3

Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Large Itemsets L(1):

outlook=sunny 5

outlook=overcast 4

outlook=rainy 5

temperature=hot 4

temperature=mild 6

temperature=cool 4

humidity=high 7

humidity=normal 7

windy=TRUE 6

windy=FALSE 8

play=yes 9

play=no 5

Size of set of large itemsets L(2): 9

Large Itemsets L(2):

outlook=overcast play=yes 4

temperature=mild humidity=high 4

temperature=mild play=yes 4

temperature=cool humidity=normal 4

humidity=high windy=FALSE 4

humidity=high play=no 4

humidity=normal windy=FALSE 4

humidity=normal play=yes 6

windy=FALSE play=yes 6

Size of set of large itemsets L(3): 1

Large Itemsets L(3):

humidity=normal windy=FALSE play=yes 4

Best rules found:

1. outlook=overcast 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
2. temperature=cool 4 ==> humidity=normal 4 <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
3. humidity=normal windy=FALSE 4 ==> play=yes 4 <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
4. humidity=normal 7 ==> play=yes 6 <conf:(0.86)> lift:(1.33) lev:(0.11) [1] conv:(1.25)
5. play=no 5 ==> humidity=high 4 <conf:(0.8)> lift:(1.6) lev:(0.11) [1] conv:(1.25)
6. windy=FALSE 8 ==> play=yes 6 <conf:(0.75)> lift:(1.17) lev:(0.06) [0] conv:(0.95)
7. play=yes 9 ==> humidity=normal 6 <conf:(0.67)> lift:(1.33) lev:(0.11) [1] conv:(1.13)
8. play=yes 9 ==> windy=FALSE 6 <conf:(0.67)> lift:(1.17) lev:(0.06) [0] conv:(0.96)
9. temperature=mild 6 ==> humidity=high 4 <conf:(0.67)> lift:(1.33) lev:(0.07) [1] conv:(1)
10. temperature=mild 6 ==> play=yes 4 <conf:(0.67)> lift:(1.04) lev:(0.01) [0] conv:(0.71)

Exercise 4

Vote dataset

Dataset in WEKA:

Viewer							
Relation: vote							
No.	1: handicapped-infants Nominal	2: water-project-cost-sharing Nominal	3: adoption-of-the-budget-resolution Nominal	4: physician-fee-freeze Nominal	5: el-salvador-aid Nominal	6: religious-groups-in-schools Nominal	7: anti-nuclear-energy Nominal
1	n	y	n	y	y	y	n
2	n	y	n	y	y	y	n
3	y	y	y	y	y	y	n
4	n	y	y	n	y	y	n
5	y	y	y	n	y	y	n
6	n	y	y	n	y	y	n
7	n	y	n	y	y	y	n
8	n	y	n	y	y	y	n
9	n	y	n	y	y	y	n
10	y	y	y	n	n	n	y
11	n	y	n	y	y	n	n
12	n	y	n	y	y	y	n
13	n	y	y	n	n	n	y
14	y	y	y	n	n	y	y
15	n	y	n	y	y	y	n
16	n	y	n	y	y	y	n
17	y	n	y	n	n	y	n
18	y	y	y	n	n	n	y
19	n	y	n	y	y	y	n
20	y	y	y	n	n	n	y
21	y	y	y	n	n	y	y
22	y	y	y	n	n	n	y
23	y	y	y	n	n	n	y
24

Configuration:

weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm.

More

Capabilities

car False

classIndex -1

delta 0.05

doNotCheckCapabilities False

lowerBoundMinSupport 0.1

metricType Confidence

minMetric 0.9

numRules 10

outputItemSets True

removeAllMissingCols False

significanceLevel -1.0

treatZeroAsMissing False

upperBoundMinSupport 1.0

verbose False

Open... Save... OK Cancel

Log output:

=== Run information ===

Scheme: weka.associations.Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: vote

Instances: 435

Attributes: 17

handicapped-infants
water-project-cost-sharing
adoption-of-the-budget-resolution
physician-fee-freeze
el-salvador-aid
religious-groups-in-schools
anti-satellite-test-ban
aid-to-nicaraguan-contras
mx-missile
immigration

synfuels-corporation-cutback
education-spending
superfund-right-to-sue
crime
duty-free-exports
export-administration-act-south-africa
Class

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

Apriori

=====

Minimum support: 0.45 (196 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20

Large Itemsets L(1):

handicapped-infants=n 236

adoption-of-the-budget-resolution=y 253

physician-fee-freeze=n 247

el-salvador-aid=n 208

el-salvador-aid=y 212

religious-groups-in-schools=y 272

anti-satellite-test-ban=y 239

aid-to-nicaraguan-contras=y 242

mx-missile=n 206

mx-missile=y 207

immigration=n 212

immigration=y 216

synfuels-corporation-cutback=n 264

education-spending=n 233

superfund-right-to-sue=n 201

superfund-right-to-sue=y 209

crime=y 248

duty-free-exports=n 233

export-administration-act-south-africa=y 269

Class=democrat 267

Size of set of large itemsets L(2): 17

Large Itemsets L(2):

adoption-of-the-budget-resolution=y physician-fee-freeze=n 219

adoption-of-the-budget-resolution=y anti-satellite-test-ban=y 201

adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y 215
 adoption-of-the-budget-resolution=y education-spending=n 201
 adoption-of-the-budget-resolution=y Class=democrat 231
 physician-fee-freeze=n anti-satellite-test-ban=y 197
 physician-fee-freeze=n aid-to-nicaraguan-contras=y 211
 physician-fee-freeze=n education-spending=n 202
 physician-fee-freeze=n Class=democrat 245
 el-salvador-aid=n aid-to-nicaraguan-contras=y 204
 el-salvador-aid=n Class=democrat 200
 el-salvador-aid=y religious-groups-in-schools=y 197
 religious-groups-in-schools=y crime=y 214
 anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 210
 anti-satellite-test-ban=y Class=democrat 200
 aid-to-nicaraguan-contras=y Class=democrat 218
 education-spending=n Class=democrat 213

Size of set of large itemsets L(3): 6

Large Itemsets L(3):

adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198
 adoption-of-the-budget-resolution=y physician-fee-freeze=n Class=democrat 219
 adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 203
 physician-fee-freeze=n aid-to-nicaraguan-contras=y Class=democrat 210
 physician-fee-freeze=n education-spending=n Class=democrat 201
 el-salvador-aid=n aid-to-nicaraguan-contras=y Class=democrat 197

Size of set of large itemsets L(4): 1

Large Itemsets L(4):

adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y
 Class=democrat 198

Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219 <conf:(1)> lift:(1.63) lev:(0.19) [84] conv:(84.58)
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> Class=democrat 198 <conf:(1)> lift:(1.63) lev:(0.18) [76] conv:(76.47)
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210 <conf:(1)> lift:(1.62) lev:(0.19) [80] conv:(40.74)
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 <conf:(1)> lift:(1.62) lev:(0.18) [77] conv:(39.01)
5. physician-fee-freeze=n 247 ==> Class=democrat 245 <conf:(0.99)> lift:(1.62) lev:(0.21) [93] conv:(31.8)
6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 <conf:(0.98)> lift:(1.77) lev:(0.2) [85] conv:(22.18)
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 <conf:(0.98)> lift:(1.76) lev:(0.2) [88] conv:(18.46)

10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210 <conf:(0.96)>
lift:(1.7) lev:(0.2) [86] conv:(10.47)

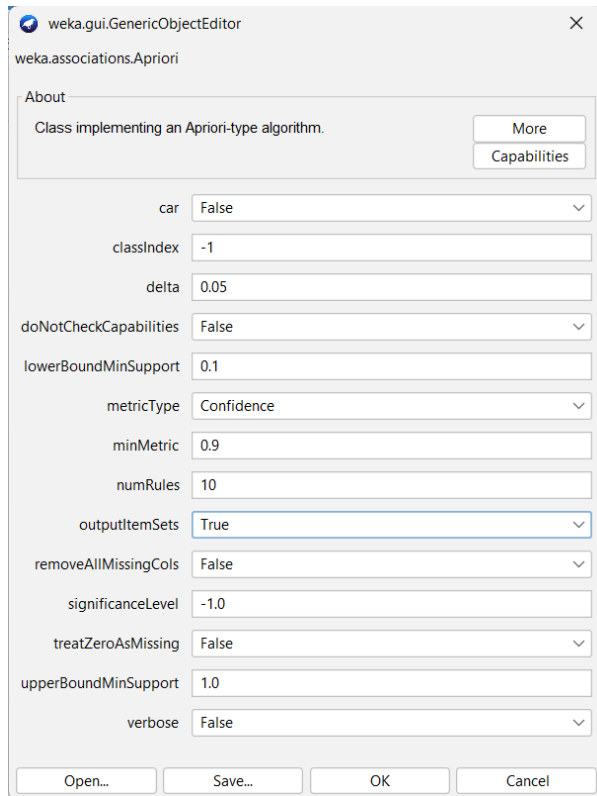
Viewer

Relation: supermarket

No.	1: department1 Nominal	2: department2 Nominal	3: department3 Nominal	4: department4 Nominal	5: department5 Nominal	6: department6 Nominal	7: department7 Nominal	8: department8 Nominal	9: department9 Nominal	10: grocery misc Nominal	11: c
1											
2	t										
3											
4	t										
5											
6			t				t				
7	t										
8											
9	t		t								
10											
11											
12	t										
13	t	t								t	
14											
15											
16	t				t		t				
17											
18	t		t								
19	t									t	
20	t										
21		t			t					t	
22	t	t									
23											
24											

Add instance Undo OK Cancel

Configuration 1:



weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm. [More](#) [Capabilities](#)

car: False

classIndex: -1

delta: 0.05

doNotCheckCapabilities: False

lowerBoundMinSupport: 0.1

metricType: Confidence

minMetric: 0.9

numRules: 10

outputItemSets: True

removeAllMissingCols: False

significanceLevel: -1.0

treatZeroAsMissing: False

upperBoundMinSupport: 1.0

verbose: False

Open... Save... OK Cancel

Log output:

=== Run information ===

Scheme: weka.associations.Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: supermarket

Instances: 4627

Attributes: 217

[list of attributes omitted]

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

Apriori

=====

Minimum support: 0.15 (694 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 44

Large Itemsets L(1):

department1=t 1047

bread and cake=t 3330

baking needs=t 2795

juice-sat-cord-ms=t 2463

.
.
.

Size of set of large itemsets L(2): 380

Large Itemsets L(2):

department1=t bread and cake=t 794

department1=t milk-cream=t 699

department1=t fruit=t 731

bread and cake=t baking needs=t 2191

bread and cake=t juice-sat-cord-ms=t 1869

.
.

.[omitted]

Size of set of large itemsets L(3): 910

Large Itemsets L(3):

bread and cake=t baking needs=t juice-sat-cord-ms=t 1291

bread and cake=t baking needs=t biscuits=t 1456

bread and cake=t baking needs=t canned fruit=t 762

bread and cake=t baking needs=t canned vegetables=t 939

bread and cake=t baking needs=t breakfast food=t 1074

.
.

.[omitted]

Size of set of large itemsets L(4): 633

Large Itemsets L(4):

bread and cake=t baking needs=t juice-sat-cord-ms=t biscuits=t 916

bread and cake=t baking needs=t juice-sat-cord-ms=t breakfast food=t 704

bread and cake=t baking needs=t juice-sat-cord-ms=t sauces-gravy-pkle=t 771

bread and cake=t baking needs=t juice-sat-cord-ms=t frozen foods=t 942

bread and cake=t baking needs=t juice-sat-cord-ms=t party snack foods=t 855

bread and cake=t baking needs=t juice-sat-cord-ms=t tissues-paper prd=t 820

.
.

.[omitted]

Size of set of large itemsets L(5): 105

Large Itemsets L(5):

bread and cake=t baking needs=t juice-sat-cord-ms=t biscuits=t frozen foods=t 730

bread and cake=t baking needs=t juice-sat-cord-ms=t biscuits=t fruit=t 706

bread and cake=t baking needs=t juice-sat-cord-ms=t frozen foods=t milk-cream=t 694

bread and cake=t baking needs=t juice-sat-cord-ms=t frozen foods=t fruit=t 716

bread and cake=t baking needs=t juice-sat-cord-ms=t frozen foods=t vegetables=t 720

bread and cake=t baking needs=t juice-sat-cord-ms=t milk-cream=t fruit=t 695

.

.

.[omitted]

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

Large Itemsets L(6):

bread and cake=t baking needs=t biscuits=t frozen foods=t fruit=t vegetables=t 716

Best rules found:

1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 <conf:(0.92)> lift:(1.27)
lev:(0.03) [155] conv:(3.35)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf:(0.92)> lift:(1.27)
lev:(0.03) [149] conv:(3.28)
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 <conf:(0.92)>
lift:(1.27) lev:(0.03) [150] conv:(3.27)
4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746 <conf:(0.92)> lift:(1.27)
lev:(0.03) [159] conv:(3.26)
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779 <conf:(0.91)> lift:(1.27)
lev:(0.04) [164] conv:(3.15)
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 <conf:(0.91)>
lift:(1.26) lev:(0.03) [151] conv:(3.06)
7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701 <conf:(0.91)>
lift:(1.26) lev:(0.03) [145] conv:(3.01)
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf:(0.91)> lift:(1.26) lev:(0.04) [179]
conv:(3)
9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757 <conf:(0.91)> lift:(1.26)
lev:(0.03) [156] conv:(3)
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf:(0.91)> lift:(1.26) lev:(0.04)
[179] conv:(2.92)

Configuration 2:

weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm.

More

Capabilities

car False

classIndex -1

delta 0.05

doNotCheckCapabilities False

lowerBoundMinSupport 0.2

metricType Confidence

minMetric 0.8

numRules 10

outputItemSets False

removeAllMissingCols False

significanceLevel -1.0

treatZeroAsMissing False

upperBoundMinSupport 0.2

verbose False

Open... Save... OK Cancel

Output:

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.8 -D 0.05 -U 0.2 -M 0.2 -S -1.0 -c -1

Relation: supermarket

Instances: 4627

Attributes: 217

[list of attributes omitted]

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

Apriori

=====

Minimum support: 0.2 (925 instances)

Minimum metric <confidence>: 0.8

Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 38

Size of set of large itemsets L(2): 225

Size of set of large itemsets L(3): 302

Size of set of large itemsets L(4): 80

Size of set of large itemsets L(5): 2

Best rules found:

1. biscuits=t margarine=t vegetables=t 1054 ==> bread and cake=t 925 <conf:(0.88)>
lift:(1.22) lev:(0.04) [166] conv:(2.27)
2. juice-sat-cord-ms=t frozen foods=t milk-cream=t 1104 ==> bread and cake=t 925
<conf:(0.84)> lift:(1.16) lev:(0.03) [130] conv:(1.72)
3. pet foods=t tissues-paper prd=t 1120 ==> bread and cake=t 925 <conf:(0.83)> lift:(1.15)
lev:(0.03) [118] conv:(1.6)
4. juice-sat-cord-ms=t cheese=t 1148 ==> bread and cake=t 925 <conf:(0.81)> lift:(1.12)
lev:(0.02) [98] conv:(1.44)

Conclusion: there was a significant decrease in the number of good rules when we set the minimum support to 20 percent.