

Code ▾

arules - Mining Association Rules and Frequent Itemsets with R

From <https://github.com/mhahsler/arules> (<https://github.com/mhahsler/arules>)

Hide

```
# install.packages("arules")
```

Load package and mine some association rules.

Hide

```
library("arules")
data("Adult")

rules <- apriori(Adult, parameter = list(supp = 0.5, conf = 0.9, target = "rules"))
```

Apriori

Parameter specification:

confidence	minval	s...	ar...	aval	originalSupport	maxti...	support	minlen
<dbl>	<dbl>	<dbl>	<chr>	<lgl>	<lgl>	<dbl>	<dbl>	<int>
0.9	0.1	1	none	FALSE	TRUE	5	0.5	1

1 row | 1-10 of 12 columns

Algorithmic control:

filter	tree	heap	memopt	load	sort	verbose
<dbl>	<lgl>	<lgl>	<lgl>	<lgl>	<int>	<lgl>
0.1	TRUE	TRUE	FALSE	TRUE	2	TRUE

1 row

Absolute minimum support count: 24421

```
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[115 item(s), 48842 transaction(s)] done [0.04s].
sorting and recoding items ... [9 item(s)] done [0.01s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 3 4 done [0.01s].
writing ... [52 rule(s)] done [0.00s].
creating S4 object ... done [0.01s].
```

Hide

Adult

transactions in sparse format with
48842 transactions (rows) and
115 items (columns)

Show basic statistics.

Hide

summary(rules)

set of 52 rules

rule length distribution (lhs + rhs):sizes
1 2 3 4
2 13 24 13

Min. 1st Qu. Median Mean 3rd Qu. Max.
1.000 2.000 3.000 2.923 3.250 4.000

summary of quality measures:
support confidence coverage lift count
Min. :0.5084 Min. :0.9031 Min. :0.5406 Min. :0.9844 Min. :24832
1st Qu.:0.5415 1st Qu.:0.9155 1st Qu.:0.5875 1st Qu.:0.9937 1st Qu.:26447
Median :0.5974 Median :0.9229 Median :0.6293 Median :0.9997 Median :29178
Mean :0.6436 Mean :0.9308 Mean :0.6915 Mean :1.0036 Mean :31433
3rd Qu.:0.7426 3rd Qu.:0.9494 3rd Qu.:0.7945 3rd Qu.:1.0057 3rd Qu.:36269
Max. :0.9533 Max. :0.9583 Max. :1.0000 Max. :1.0586 Max. :46560

mining info:

data <chr>	ntransactions <int>	support <dbl>	confidence <dbl>
Adult	48842	0.5	0.9
1 row			

Inspect rules with the highest lift.

Hide

inspect(head(rules, by = "lift"))

lhs <chr>	rhs <chr>
[1] {sex=Male,native-country=United-States}	=> {race=White}
[2] {sex=Male,capital-loss=None,native-country=United-States}	=> {race=White}
[3] {race=White}	=> {native-country=United-States}
[4] {race=White,capital-loss=None}	=> {native-country=United-States}

lhs	rhs
<chr>	<chr×chr>
[5] {race=White,sex=Male}	=> {native-country=U
[6] {race=White,capital-gain=None}	=> {native-country=U

6 rows | 1-4 of 8 columns

Visualize rules :

1/ install & import package

Hide

```
# install.packages("arulesViz", dependencies = TRUE)
```

Hide

```
# install.packages("grid")
```

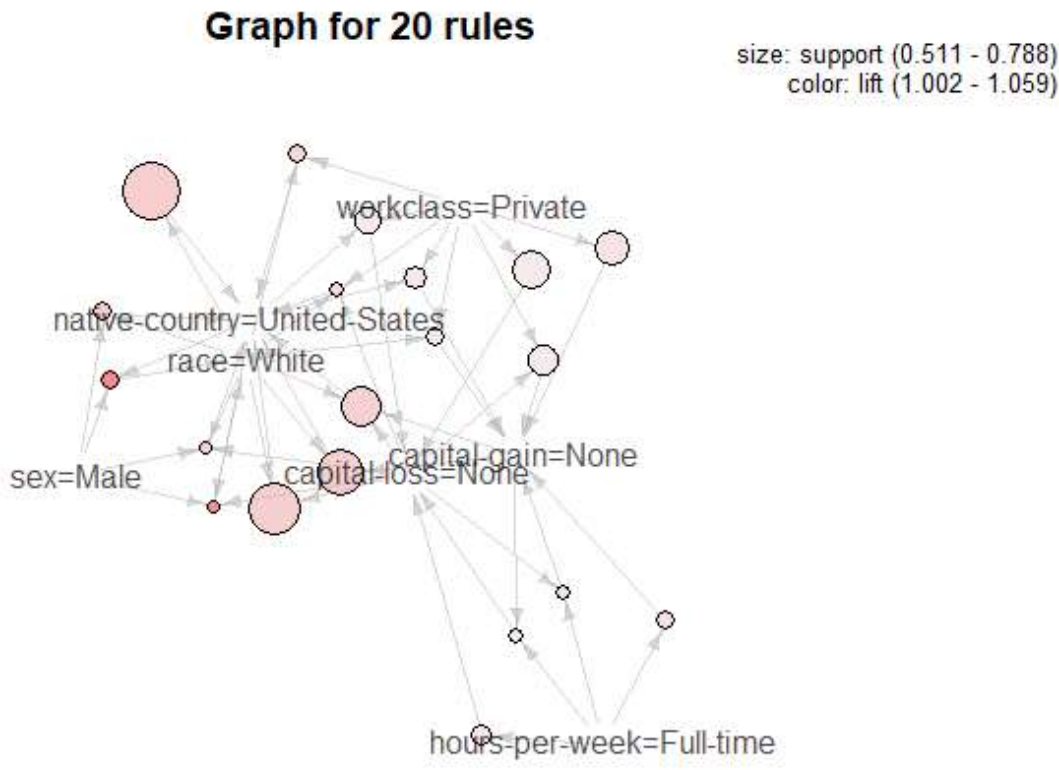
Hide

```
library(arulesViz)
```

2/ Plot a graph of the most important rules

Hide

```
plot(head(rules, 20, by = "lift"), method = "graph")
```



```
plot(head(rules, 20, by = "lift"), method = "matrix")
```

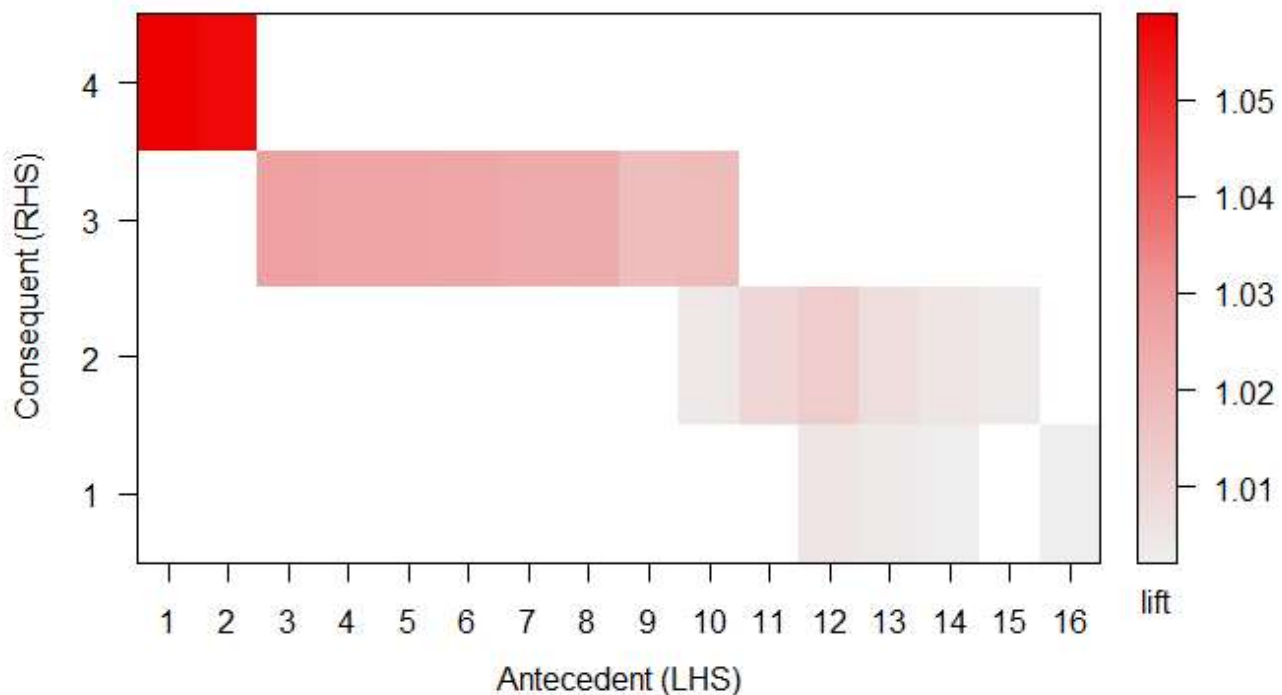
Itemsets in Antecedent (LHS)

```
[1] "{sex=Male,native-country=United-States}"
[2] "{sex=Male,capital-loss=None,native-country=United-States}"
[3] "{race=White}"
[4] "{race=White,capital-loss=None}"
[5] "{race=White,sex=Male}"
[6] "{race=White,capital-gain=None}"
[7] "{race=White,sex=Male,capital-loss=None}"
[8] "{race=White,capital-gain=None,capital-loss=None}"
[9] "{workclass=Private,race=White,capital-loss=None}"
[10] "{workclass=Private,race=White}"
[11] "{capital-loss=None,hours-per-week=Full-time}"
[12] "{hours-per-week=Full-time}"
[13] "{workclass=Private}"
[14] "{workclass=Private,native-country=United-States}"
[15] "{workclass=Private,capital-loss=None}"
[16] "{capital-gain=None,hours-per-week=Full-time}"
```

Itemsets in Consequent (RHS)

```
[1] "{capital-loss=None}"           "{capital-gain=None}"           "{native-country=United-States}"
[4] "{race=White}"
```

Matrix with 20 rules



Hide

```
rules <- apriori(Adult, parameter = list(supp = 0.5, conf = 0.9, target = "closed frequent itemsets"))
```

Apriori

Parameter specification:

confidence <dbl>	minval <dbl>	s... <dbl>	ar... <chr>	aval <lgl>	originalSupport <lgl>	maxti... <dbl>	support <dbl>	minlen <int>	
NA	0.1	1	none	FALSE	TRUE	5	0.5	1	

1 row | 1-10 of 12 columns

Algorithmic control:

filter <dbl>	tree <lgl>	heap <lgl>	memopt <lgl>	load <lgl>	sort <int>	verbose <lgl>
0.1	TRUE	TRUE	FALSE	TRUE	2	TRUE

1 row

Absolute minimum support count: 24421

```

set item appearances ...[0 item(s)] done [0.01s].
set transactions ...[115 item(s), 48842 transaction(s)] done [0.04s].
sorting and recoding items ... [9 item(s)] done [0.00s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 3 4 done [0.02s].
filtering closed item sets ... done [0.00s].
sorting transactions ... done [0.01s].
writing ... [49 set(s)] done [0.01s].
creating S4 object ... done [0.01s].

```

Hide

summary(rules)

```
set of 49 itemsets

most frequent items:
      capital-loss=None      capital-gain=None native-country=United-States
              23              21              21
              race=White      workclass=Private      (Other)
              19              14              20

element (itemset/transaction) length distribution:sizes
 1  2  3  4
 9 17 17  6

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000  2.000  2.000  2.408  3.000  4.000

summary of quality measures:
      support      transIdenticalToItemsets      count
Min.   :0.5051  Min.   :0.000e+00      Min.   :24671
1st Qu.:0.5434  1st Qu.:0.000e+00      1st Qu.:26540
Median :0.5942  Median :2.047e-05      Median :29024
Mean   :0.6449  Mean   :4.630e-04      Mean   :31497
3rd Qu.:0.7404  3rd Qu.:3.071e-04      3rd Qu.:36164
Max.   :0.9533  Max.   :7.309e-03      Max.   :46560

includes transaction ID lists: FALSE

mining info:
```

data <chr>	ntransactions <int>	support <dbl>	confidence <dbl>
Adult	48842	0.5	1

1 row

Hide

```
inspect(head(rules, by = "support"))
```

items <chr>	support <dbl>	transIdentical
[1] {capital-loss=None}	0.9532779	2.0
[2] {capital-gain=None}	0.9173867	0.0
[3] {native-country=United-States}	0.8974243	0.0
[4] {capital-gain=None,capital-loss=None}	0.8706646	2.0
[5] {race=White}	0.8550428	0.0
[6] {capital-loss=None,native-country=United-States}	0.8548380	0.0

6 rows

Hide

```
rules <- apriori(Adult, parameter = list(supp = 0.5, conf = 0.9, target = "maximally frequent itemsets"))
```

Apriori

Parameter specification:

confidence <dbl>	minval <dbl>	s... <dbl>	ar... <chr>	aval <lgl>	originalSupport <lgl>	maxti... <dbl>	support <dbl>	minlen <int>
NA	0.1	1	none	FALSE	TRUE	5	0.5	1

1 row | 1-10 of 12 columns

Algorithmic control:

filter <dbl>	tree <lgl>	heap <lgl>	memopt <lgl>	load <lgl>	sort <int>	verbose <lgl>
0.1	TRUE	TRUE	FALSE	TRUE	2	TRUE

1 row

Absolute minimum support count: 24421

```
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[115 item(s), 48842 transaction(s)] done [0.04s].
sorting and recoding items ... [9 item(s)] done [0.01s].
creating transaction tree ... done [0.01s].
checking subsets of size 1 2 3 4 done [0.00s].
filtering maximal item sets ... done [0.00s].
sorting transactions ... done [0.01s].
writing ... [11 set(s)] done [0.01s].
creating S4 object ... done [0.01s].
```

Hide

summary(rules)

set of 11 itemsets

most frequent items:

capital-loss=None	capital-gain=None	native-country=United-States
7	6	6
race=White	workclass=Private	(Other)
5	3	7

element (itemset/transaction) length distribution:sizes

1 2 3 4

2 1 2 6

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	2.500	4.000	3.091	4.000	4.000

summary of quality measures:

support	transIdenticalToItemsets	count
Min. :0.5051	Min. :0.000e+00	Min. :24671
1st Qu.:0.5099	1st Qu.:7.166e-05	1st Qu.:24904
Median :0.5181	Median :7.371e-04	Median :25307
Mean :0.5327	Mean :1.590e-03	Mean :26019
3rd Qu.:0.5259	3rd Qu.:1.443e-03	3rd Qu.:25686
Max. :0.6804	Max. :7.309e-03	Max. :33232

includes transaction ID lists: FALSE

mining info:

data <chr>	ntransactions <int>	support <dbl>	confidence <dbl>
Adult	48842	0.5	1

1 row

Hide

```
inspect(head(rules, by = "support"))
```


items	support	transIdenticalToItemsets	count
[1] {race=White, capital-gain=None, capital-loss=None, native-country=United-States}	0.6803980	0.0073092830	33232
[2] {workclass=Private, capital-gain=None, capital-loss=None, native-country=United-States}	0.5414807	0.0015969862	26447
[3] {race=White, sex=Male, capital-gain=None}	0.5313050	0.0002047418	25950
[4] {workclass=Private, race=White, capital-gain=None, capital-loss=None}	0.5204742	0.0007370706	25421
[5] {capital-gain=None, capital-loss=None, hours-per-week=Full-time}	0.5191638	0.0001433193	25357
[6] {workclass=Private, race=White, capital-loss=None, native-country=United-States}	0.5181401	0.0010851316	25307

Hide

```
rules <- apriori(Adult, parameter = list(supp = 0.5, conf = 0.9, target = "frequent itemsets")
))
```

Apriori

Parameter specification:

confidence	minval	s...	ar...	aval	originalSupport	maxti...	support	minlen
<dbl>	<dbl>	<dbl>	<chr>	<lgl>	<lgl>	<dbl>	<dbl>	<int>
NA	0.1	1	none	FALSE	TRUE	5	0.5	1

1 row | 1-10 of 12 columns

Algorithmic control:

filter	tree	heap	memopt	load	sort	verbose
<dbl>	<lgl>	<lgl>	<lgl>	<lgl>	<int>	<lgl>
0.1	TRUE	TRUE	FALSE	TRUE	2	TRUE

1 row

```
Absolute minimum support count: 24421

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[115 item(s), 48842 transaction(s)] done [0.05s].
sorting and recoding items ... [9 item(s)] done [0.01s].
creating transaction tree ... done [0.02s].
checking subsets of size 1 2 3 4 done [0.02s].
sorting transactions ... done [0.01s].
writing ... [49 set(s)] done [0.01s].
creating S4 object ... done [0.01s].
```

Hide

```
summary(rules)
```

```
set of 49 itemsets

most frequent items:
      capital-loss=None      capital-gain=None native-country=United-States
              23              21              21
      race=White      workclass=Private      (Other)
              19              14              20

element (itemset/transaction) length distribution:sizes
 1  2  3  4
9 17 17  6

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.000  2.000  2.000  2.408  3.000  4.000

summary of quality measures:
  support  transIdenticalToItemsets  count
Min.   :0.5051  Min.   :0.000e+00  Min.   :24671
1st Qu.:0.5434  1st Qu.:0.000e+00  1st Qu.:26540
Median :0.5942  Median :2.047e-05  Median :29024
Mean   :0.6449  Mean   :4.630e-04  Mean   :31497
3rd Qu.:0.7404  3rd Qu.:3.071e-04  3rd Qu.:36164
Max.   :0.9533  Max.   :7.309e-03  Max.   :46560

includes transaction ID lists: FALSE

mining info:
```

data <chr>	ntransactions <int>	support <dbl>	confidence <dbl>
Adult	48842	0.5	1
1 row			

Hide

```
inspect(head(rules, by = "support"))
```

Items <chr>	support <dbl>	transIdential
[1] {capital-loss=None}	0.9532779	2.0
[2] {capital-gain=None}	0.9173867	0.0
[3] {native-country=United-States}	0.8974243	0.0
[4] {capital-gain=None,capital-loss=None}	0.8706646	2.0
[5] {race=White}	0.8550428	0.0
[6] {capital-loss=None,native-country=United-States}	0.8548380	0.0
6 rows		