

Market Basket Analysis

pranit shinde

10 December 2018

Markert basket analysis

load the librarires

```
library(arules)
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'arules'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      abbreviate, write
```

```
library(arulesViz)
```

```
## Loading required package: grid
```

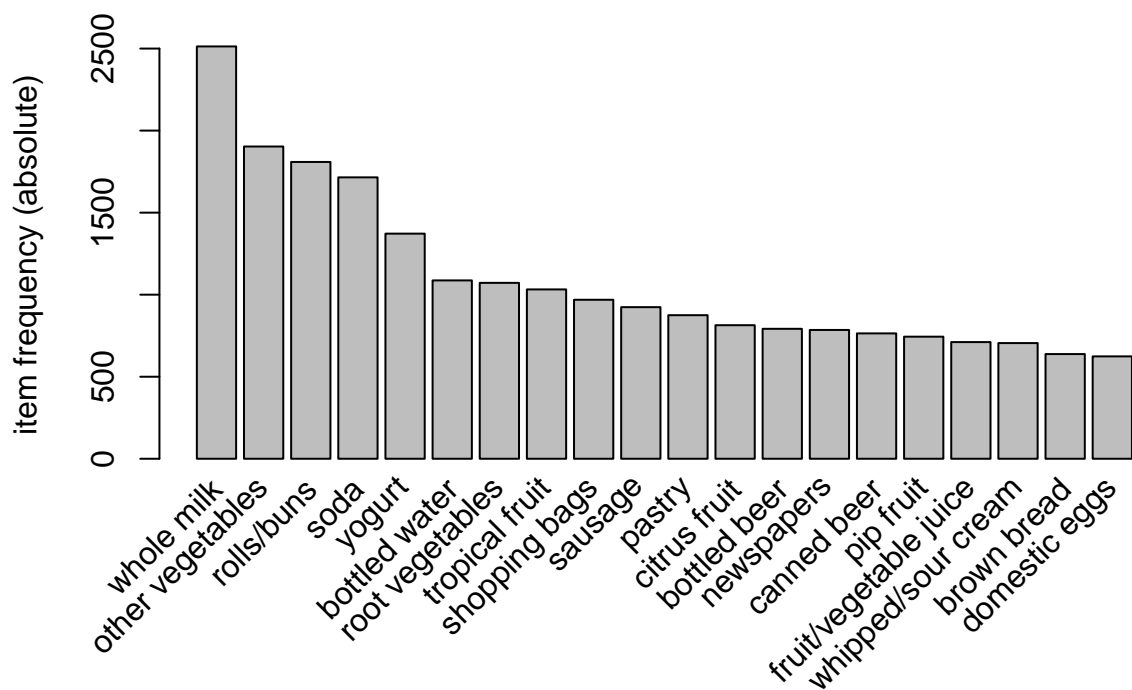
```
library(datasets)
```

load data

```
data("Groceries")
```

create an item frequency plot for the top 20 items

```
itemFrequencyPlot(Groceries, topN = 20, type = "absolute")
```



get the rules

```
rules <- apriori(Groceries, parameter = list(supp = 0.001, conf = 0.8))
```

```
## Apriori
##
## Parameter specification:
## confidence minval smax arem aval originalSupport maxtime support minlen
##          0.8    0.1    1 none FALSE                TRUE     5   0.001     1
## maxlen target   ext
##          10 rules FALSE
##
## Algorithmic control:
## filter tree heap memopt load sort verbose
##    0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 9
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [157 item(s)] done [0.00s].
## creating transaction tree ... done [0.01s].
## checking subsets of size 1 2 3 4 5 6 done [0.02s].
```

```
## writing ... [410 rule(s)] done [0.00s].
## creating S4 object ... done [0.00s].
```

show the top 5 rules but only 2 digits

```
options(digits = 2)
inspect(rules[1:5])
```

```
##      lhs                      rhs      support confidence lift
## [1] {liquor,red/blush wine} => {bottled beer} 0.0019  0.90    11.2
## [2] {curd,cereals}          => {whole milk}  0.0010  0.91     3.6
## [3] {yogurt,cereals}        => {whole milk}  0.0017  0.81     3.2
## [4] {butter,jam}            => {whole milk}  0.0010  0.83     3.3
## [5] {soups,bottled beer}    => {whole milk}  0.0011  0.92     3.6
##      count
## [1] 19
## [2] 10
## [3] 17
## [4] 10
## [5] 11
```

```
summary(rules)
```

```
## set of 410 rules
##
## rule length distribution (lhs + rhs):sizes
##   3   4   5   6
## 29 229 140  12
##
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.0    4.0    4.0    4.3    5.0    6.0
##
## summary of quality measures:
##      support      confidence      lift      count
## Min.   :0.00102  Min.   :0.80  Min.   : 3.1  Min.   :10.0
## 1st Qu.:0.00102  1st Qu.:0.83  1st Qu.: 3.3  1st Qu.:10.0
## Median :0.00122  Median :0.85  Median : 3.6  Median :12.0
## Mean   :0.00125  Mean   :0.87  Mean   : 4.0  Mean   :12.3
## 3rd Qu.:0.00132  3rd Qu.:0.91  3rd Qu.: 4.3  3rd Qu.:13.0
## Max.   :0.00315  Max.   :1.00  Max.   :11.2  Max.   :31.0
##
## mining info:
##      data ntransactions support confidence
## Groceries      9835    0.001      0.8
```

Sorting stuff out

```
rules <- sort(rules, by = "confidence", decreasing = T)
rules[1:5]
```

```
## set of 5 rules
```

```
rules
```

```
## set of 410 rules
```

```
inspect(rules[1:20])
```

##	lhs	rhs	support	confidence	lift	count
## [1]	{rice, sugar}	=> {whole milk}	0.0012	1	3.9	12
## [2]	{canned fish, hygiene articles}	=> {whole milk}	0.0011	1	3.9	11
## [3]	{root vegetables, butter, rice}	=> {whole milk}	0.0010	1	3.9	10
## [4]	{root vegetables, whipped/sour cream, flour}	=> {whole milk}	0.0017	1	3.9	17
## [5]	{butter, soft cheese, domestic eggs}	=> {whole milk}	0.0010	1	3.9	10
## [6]	{citrus fruit, root vegetables, soft cheese}	=> {other vegetables}	0.0010	1	5.2	10
## [7]	{pip fruit, butter, hygiene articles}	=> {whole milk}	0.0010	1	3.9	10
## [8]	{root vegetables, whipped/sour cream, hygiene articles}	=> {whole milk}	0.0010	1	3.9	10
## [9]	{pip fruit, root vegetables, hygiene articles}	=> {whole milk}	0.0010	1	3.9	10
## [10]	{cream cheese , domestic eggs, sugar}	=> {whole milk}	0.0011	1	3.9	11
## [11]	{curd, domestic eggs, sugar}	=> {whole milk}	0.0010	1	3.9	10
## [12]	{cream cheese , domestic eggs, napkins}	=> {whole milk}	0.0011	1	3.9	11
## [13]	{pip fruit, whipped/sour cream, brown bread}	=> {other vegetables}	0.0011	1	5.2	11
## [14]	{tropical fruit, grapes, whole milk, yogurt}	=> {other vegetables}	0.0010	1	5.2	10
## [15]	{ham, tropical fruit, pip fruit,					

```

##      yogurt}                => {other vegetables}  0.0010          1  5.2    10
## [16] {ham,
##      tropical fruit,
##      pip fruit,
##      whole milk}            => {other vegetables}  0.0011          1  5.2    11
## [17] {tropical fruit,
##      root vegetables,
##      yogurt,
##      oil}                   => {whole milk}        0.0011          1  3.9    11
## [18] {root vegetables,
##      other vegetables,
##      yogurt,
##      oil}                   => {whole milk}        0.0014          1  3.9    14
## [19] {root vegetables,
##      other vegetables,
##      butter,
##      white bread}           => {whole milk}        0.0010          1  3.9    10
## [20] {pork,
##      other vegetables,
##      butter,
##      whipped/sour cream} => {whole milk}        0.0010          1  3.9    10

```

Targeting Items

```

rules <- apriori(data = Groceries, parameter = list(supp = 0.001, conf = 0.08),
  appearance = list(default = "lhs", rhs = "whole milk"),
  control = list(verbose = F))
rules <- sort(rules, by = "confidence", decreasing = T)
inspect(rules[1:20])

```

```

##      lhs                rhs      support confidence lift count
## [1] {rice,
##      sugar}              => {whole milk}  0.0012          1  3.9    12
## [2] {canned fish,
##      hygiene articles}   => {whole milk}  0.0011          1  3.9    11
## [3] {root vegetables,
##      butter,
##      rice}               => {whole milk}  0.0010          1  3.9    10
## [4] {root vegetables,
##      whipped/sour cream,
##      flour}              => {whole milk}  0.0017          1  3.9    17
## [5] {butter,
##      soft cheese,
##      domestic eggs}      => {whole milk}  0.0010          1  3.9    10
## [6] {pip fruit,
##      butter,
##      hygiene articles}   => {whole milk}  0.0010          1  3.9    10
## [7] {root vegetables,
##      whipped/sour cream,
##      hygiene articles}   => {whole milk}  0.0010          1  3.9    10
## [8] {pip fruit,
##      root vegetables,

```

```

## hygiene articles} => {whole milk} 0.0010      1  3.9   10
## [9] {cream cheese ,
##      domestic eggs,
##      sugar}      => {whole milk} 0.0011      1  3.9   11
## [10] {curd,
##      domestic eggs,
##      sugar}      => {whole milk} 0.0010      1  3.9   10
## [11] {cream cheese ,
##      domestic eggs,
##      napkins}    => {whole milk} 0.0011      1  3.9   11
## [12] {tropical fruit,
##      root vegetables,
##      yogurt,
##      oil}        => {whole milk} 0.0011      1  3.9   11
## [13] {root vegetables,
##      other vegetables,
##      yogurt,
##      oil}        => {whole milk} 0.0014      1  3.9   14
## [14] {root vegetables,
##      other vegetables,
##      butter,
##      white bread} => {whole milk} 0.0010      1  3.9   10
## [15] {pork,
##      other vegetables,
##      butter,
##      whipped/sour cream} => {whole milk} 0.0010      1  3.9   10
## [16] {other vegetables,
##      butter,
##      whipped/sour cream,
##      domestic eggs}  => {whole milk} 0.0012      1  3.9   12
## [17] {citrus fruit,
##      whipped/sour cream,
##      rolls/buns,
##      pastry}      => {whole milk} 0.0010      1  3.9   10
## [18] {pip fruit,
##      root vegetables,
##      other vegetables,
##      bottled water} => {whole milk} 0.0011      1  3.9   11
## [19] {sausage,
##      tropical fruit,
##      root vegetables,
##      rolls/buns}   => {whole milk} 0.0010      1  3.9   10
## [20] {tropical fruit,
##      root vegetables,
##      other vegetables,
##      yogurt,
##      oil}          => {whole milk} 0.0010      1  3.9   10

```

Whole milk in lhs

```

rules <- apriori(data = Groceries,
                  parameter = list(supp = 0.001, conf = 0.08),

```

```

        appearance = list(default = "rhs", lhs = "whole milk"),
        control = list(verbose = F)
    )
rules <- sort(rules, by = "confidence", decreasing = T)
inspect(rules[1:20])

```

##	lhs	rhs	support	confidence	lift	count
## [1]	{whole milk}	=> {other vegetables}	0.075	0.29	1.5	736
## [2]	{whole milk}	=> {rolls/buns}	0.057	0.22	1.2	557
## [3]	{whole milk}	=> {yogurt}	0.056	0.22	1.6	551
## [4]	{}	=> {other vegetables}	0.193	0.19	1.0	1903
## [5]	{whole milk}	=> {root vegetables}	0.049	0.19	1.8	481
## [6]	{}	=> {rolls/buns}	0.184	0.18	1.0	1809
## [7]	{}	=> {soda}	0.174	0.17	1.0	1715
## [8]	{whole milk}	=> {tropical fruit}	0.042	0.17	1.6	416
## [9]	{whole milk}	=> {soda}	0.040	0.16	0.9	394
## [10]	{}	=> {yogurt}	0.140	0.14	1.0	1372
## [11]	{whole milk}	=> {bottled water}	0.034	0.13	1.2	338
## [12]	{whole milk}	=> {pastry}	0.033	0.13	1.5	327
## [13]	{whole milk}	=> {whipped/sour cream}	0.032	0.13	1.8	317
## [14]	{whole milk}	=> {citrus fruit}	0.031	0.12	1.4	300
## [15]	{whole milk}	=> {pip fruit}	0.030	0.12	1.6	296
## [16]	{whole milk}	=> {domestic eggs}	0.030	0.12	1.9	295
## [17]	{whole milk}	=> {sausage}	0.030	0.12	1.2	294
## [18]	{}	=> {bottled water}	0.111	0.11	1.0	1087
## [19]	{}	=> {root vegetables}	0.109	0.11	1.0	1072
## [20]	{whole milk}	=> {butter}	0.028	0.11	1.9	271

visualization

```

library(arulesViz)
plot(rules[1:10], method = "graph", interactive = TRUE, shading = NA)

```

```

## Warning in plot.rules(rules[1:10], method = "graph", interactive = TRUE, :
## The parameter interactive is deprecated. Use engine='interactive' instead.

```