chap8

Satoshi Koiso

12/28/2021

Chapter 8

This is an R Markdown document about Chapter 8 of this book.

Evaluation indicator in association analysis

This is a great reference.

Support

The ratio of transaction including product X and Y to all transactions.
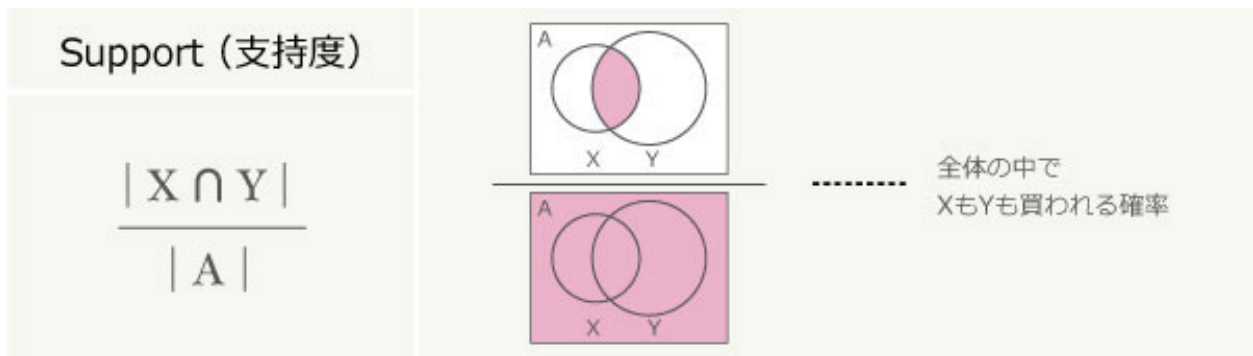

Figure 1: support

Confidence

The ratio of transaction including product X and Y to transaction of X.

This indicator decides the direction of recommendation.

For example, if X is printer and Y is ink cartridge,

X -> Y : reasonable recommendation
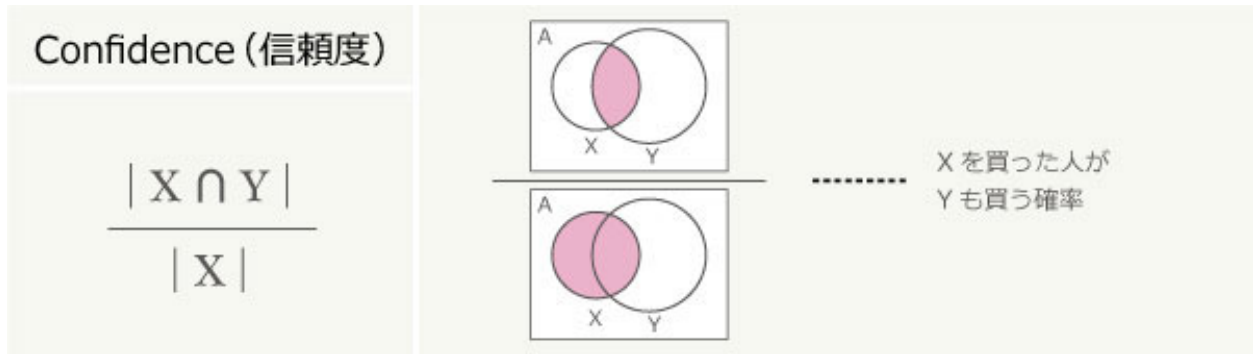
Y -> X : questionable recommendation

Figure 2: conficence

Lift

The ratio of confidence to the ratio of transaction of Y to all transactions. More than 1 is considered a good recommendation rule.
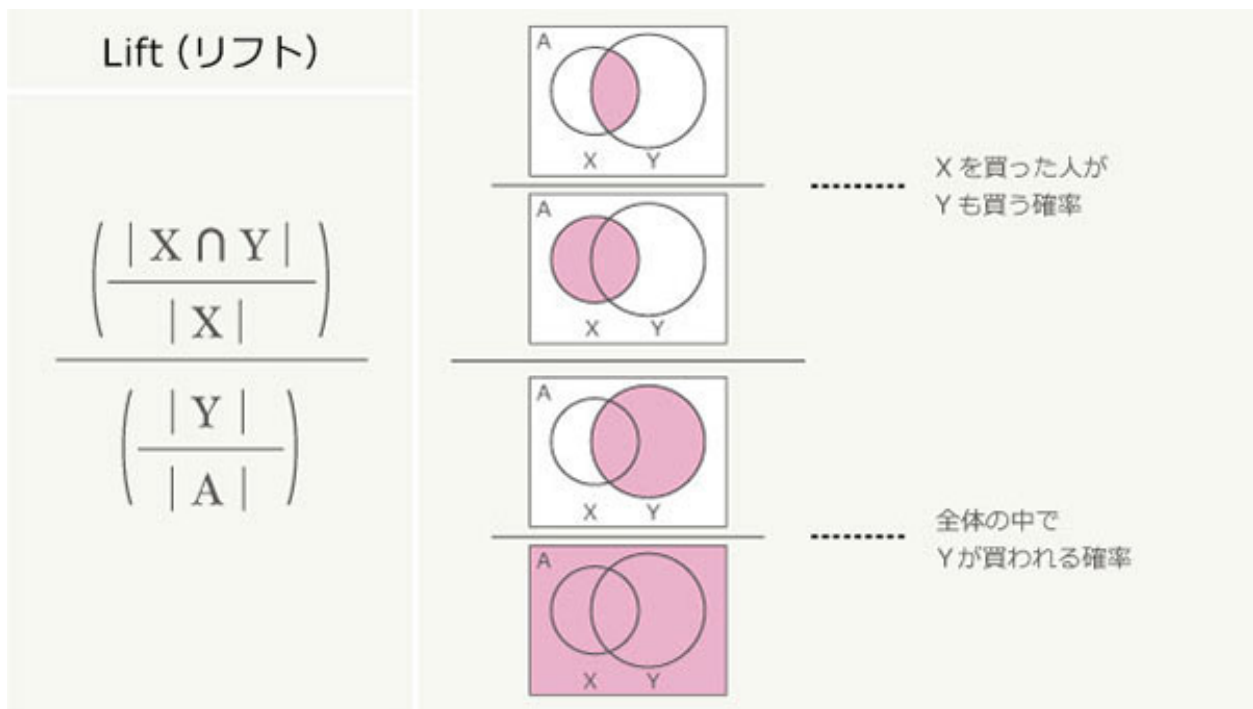

Figure 3: lift

8.3 Try arules package

summary statistics

```
inspect(head(Groceries))
```

```
##     items
```

```
## [1] {citrus fruit,
##      semi-finished bread,
##      margarine,
##      ready soups}
## [2] {tropical fruit,
##      yogurt,
##      coffee}
## [3] {whole milk}
## [4] {pip fruit,
##      yogurt,
##      cream cheese ,
##      meat spreads}
## [5] {other vegetables,
##      whole milk,
##      condensed milk,
##      long life bakery product}
## [6] {whole milk,
##      butter,
##      yogurt,
##      rice,
##      abrasive cleaner}
```

```
summary(Groceries)
```

```
## transactions as itemMatrix in sparse format with
##  9835 rows (elements/itemsets/transactions) and
##  169 columns (items) and a density of 0.02609146
##
## most frequent items:
##       whole milk other vegetables       rolls/buns           soda
##             2513             1903             1809           1715
##          yogurt          (Other)
##             1372            34055
##
## element (itemset/transaction) length distribution:
## sizes
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
## 2159 1643 1299 1005  855  645  545  438  350  246  182  117   78   77   55   46
##   17   18   19   20   21   22   23   24   26   27   28   29   32
##   29   14   14    9   11    4    6    1    1    1    1    3    1
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   3.000   4.409   6.000  32.000
##
## includes extended item information - examples:
##         labels  level2          level1
## 1 frankfurter sausage meat and sausage
## 2     sausage sausage meat and sausage
## 3  liver loaf sausage meat and sausage
```

```
itemFrequencyPlot(Groceries, support=0.04, cex.names=0.8)# support: set threshold of support, cex.names
```
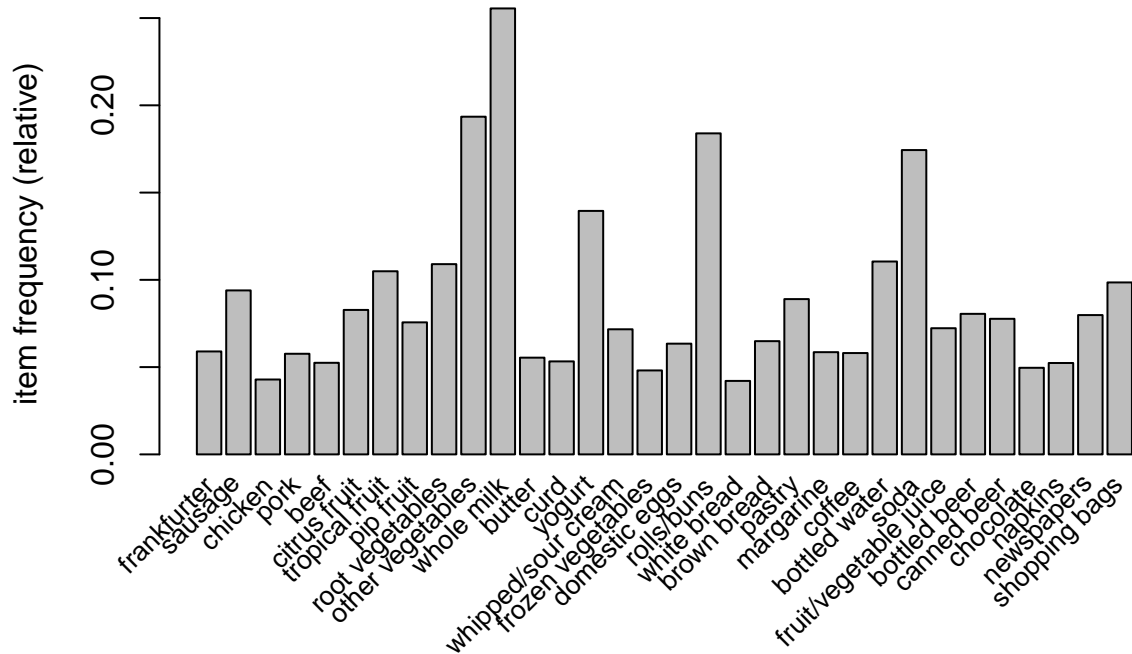
apriori function

```r
rules <- apriori(Groceries, parameter = list(support = 0.005, confidence = 0.01)) # we can set threshol
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##        0.01    0.1    1 none FALSE            TRUE       5   0.005      1
##  maxlen target   ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 49
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
## sorting and recoding items ... [120 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 done [0.00s].
## writing ... [2138 rule(s)] done [0.00s].
```

```
## creating S4 object  ... done [0.00s].
```

If we want to get associated transactions of beef, we can find products often bought with beef by sorting the data by `lift`.

```
beefRules <- subset(rules,subset= rhs %in% "beef")
inspect(head(sort(beefRules,by= "lift")))
```

```
##      lhs                                 rhs     support     confidence
## [1] {root vegetables, other vegetables} => {beef} 0.007930859 0.1673820
## [2] {root vegetables, whole milk}       => {beef} 0.008032537 0.1642412
## [3] {root vegetables}                   => {beef} 0.017386884 0.1595149
## [4] {other vegetables, rolls/buns}      => {beef} 0.005795628 0.1360382
## [5] {pork}                              => {beef} 0.007625826 0.1322751
## [6] {other vegetables, whole milk}      => {beef} 0.009252669 0.1236413
##      coverage   lift     count
## [1] 0.04738180 3.190313  78
## [2] 0.04890696 3.130449  79
## [3] 0.10899847 3.040367 171
## [4] 0.04260295 2.592898  57
## [5] 0.05765125 2.521174  75
## [6] 0.07483477 2.356613  91
```

However, based on support, the 3rd combination–{root veg} -> {beef}– might be more insightful.

```
write(beefRules, file="data.csv", sep=",", col.names=NA)
```