Report of Assignment Four

3140104750 黃霄

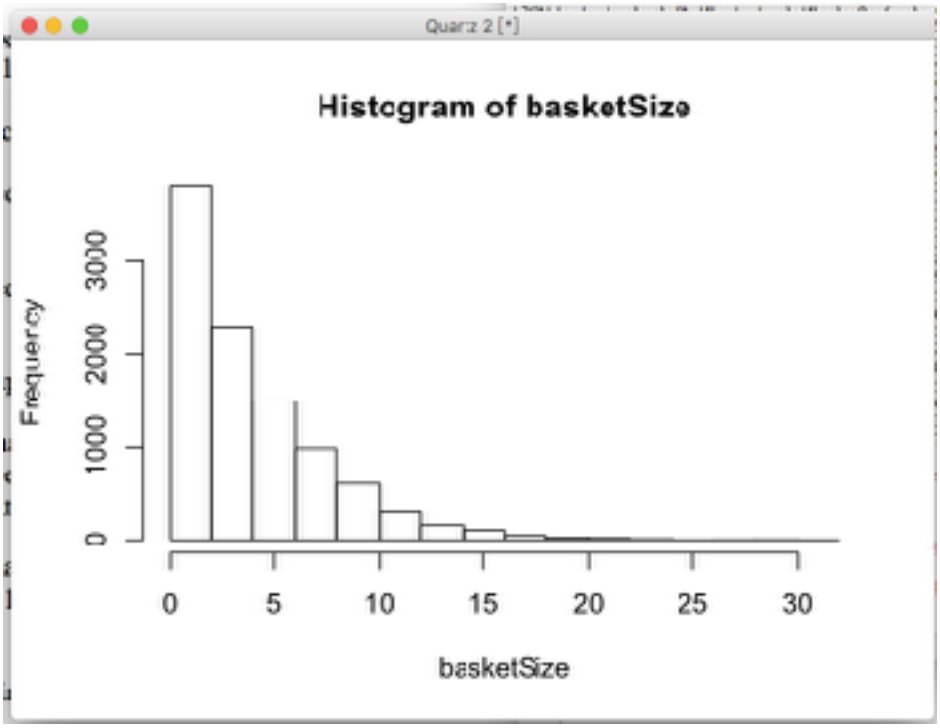
---

1. Plot a histogram of the number of items (categories) per transaction. What do you observe?

[Code]

```
install.packages("arules")
library("arules")
data(Groceries)
basketSize <- size(Groceries)
hist(basketSize)
```

[Result]



[Ans] I observe that most transaction has fewer than 5 items. The more the items are, the lower the frequency of transaction is.

---

2. How many frequent itemsets, closed frequent itemsets, and maximal frequent itemsets do you obtain with minimum support = 0.001?

[Code]

```
frequentSets=eclat(Groceries,parameter=list(support=0.001));
closedGroceries <- is.closed(frequentSets)
maxGroceries <- is.maximal(frequentSets)
sum(closedGroceries)
sum(maxGroceries)
```

[Result]

```
> frequentSets
set of 13492 itemsets
> closedGroceries <- is.closed(frequentSets)
> maxGroceries <- is.maximal(frequentSets)
> sum(closedGroceries)
[1] 13464
> sum(maxGroceries)
[1] 7794
```

[Ans] As shown in the image, the frequent itemsets, closed frequent itemsets, and maximal frequent itemsets are 13492, 13464 and 7794 respectively.

---

3. What is the result when minimum support = 0.01? List10 itemsets with the highest support and their support?

[Code]

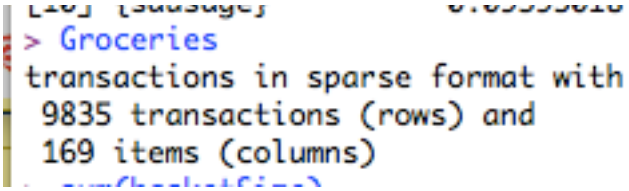
```
frequentSets1=eclat(Groceries,parameter=list(support=0.01));
closedGroceries1 <- is.closed(frequentSets1)
maxGroceries1 <- is.maximal(frequentSets1)
sum(closedGroceries1)
sum(maxGroceries1)
inspect(sort(frequentSets1,by="support")[1:10]);
```

[Result]

```
> frequentSets1
set of 333 itemsets
> closedGroceries1 <- is.closed(frequentSets1)
> maxGroceries1 <- is.maximal(frequentSets1)
> sum(closedGroceries1)
[1] 333
> sum(maxGroceries1)
[1] 243
> inspect(sort(frequentsets,by="support")[1:10]);
Error in sort(frequentsets, by = "support") : 找不到对象'frequentsets'
> inspect(sort(frequentSets1,by="support")[1:10]);
     items              support
[1]  {whole milk}       0.25551601
[2]  {other vegetables} 0.19349263
[3]  {rolls/buns}       0.18393493
[4]  {soda}             0.17437722
[5]  {yogurt}           0.13950178
[6]  {bottled water}    0.11052364
[7]  {root vegetables}  0.10899847
[8]  {tropical fruit}   0.10493137
[9]  {shopping bags}    0.09852567
[10] {sausage}          0.09395018
```

4. How do you explain the relatively small number of frequent itemsets for the already low minimum support of 0.01? How do you explain the observation that the numbers of frequent itemsets, closed frequent itemsets, and maximal frequent itemsets are so similar?

4.1



[Ans]From the image, we can see that there are only 169 kinds of items while there are 9835 transactions. And I think this(the low ratio of items number to transaction number) is the reason for relatively small number of frequent itemsets for the already low minimum support of 0.01.

4.2

[Code] # This is to see when mini support is lowered, whether the similarity still exists.

```
frequentSets2=eclat(Groceries,parameter=list(support=0.0005));
closedGroceries2 <- is.closed(frequentSets2)
maxGroceries2 <- is.maximal(frequentSets2)
sum(closedGroceries2)
sum(maxGroceries2)
```

[Result]

```
> closedGroceries2 <- is.closed(frequentSets2)
> maxGroceries2 <- is.maximal(frequentSets2)
> sum(closedGroceries2)
[1] 46267
> sum(maxGroceries2)
[1] 22725
>
> frequentSets2
set of 49580 itemsets
```

[Ans]

First, this is because of the small number of items per transaction. From problem 1, we can see that most transaction has fewer than 5 items. As a result, it is less likely that one itemset is the superset of another, which further results in the similarity between closed frequent itemsets and frequent itemsets.

Second, it is possibly because of the small number of frequent itemsets. From the image above, we can see that when the minimum support is decreased, the three attributes are not so similar then.

5. At minimum support = 0.01, how many association rules do you obtain with minimum confidence = 0.9? How far do you need to lower the minimum confidence to obtain more than 10 rules?

[Code]

```
Rules <- apriori(Groceries, parameter = list(support = 0.01, confidence = 0.9))
inspect(Rules)
Rules <- apriori(Groceries, parameter = list(support = 0.01, confidence = 0.8))
inspect(Rules)
Rules <- apriori(Groceries, parameter = list(support = 0.01, confidence = 0.7))
inspect(Rules)
Rules <- apriori(Groceries, parameter = list(support = 0.01, confidence = 0.6))
inspect(Rules)
Rules <- apriori(Groceries, parameter = list(support = 0.01, confidence = 0.5))
inspect(Rules)
```

[Result]

```
> inspect(Rules)
> Rules <- apriori(Groceries, parameter = list(support = 0.01, confidence = 0.5))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen
        0.5    0.1    1 none FALSE            TRUE       5    0.01      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: 98

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [88 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [15 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
> inspect(Rules)
     lhs                                rhs                support
[1]  {curd,yogurt}                   => {whole milk}       0.01006609
[2]  {other vegetables,butter}       => {whole milk}       0.01148958
[3]  {other vegetables,domestic eggs}=> {whole milk}       0.01230300
[4]  {yogurt,whipped/sour cream}     => {whole milk}       0.01087951
[5]  {other vegetables,whipped/sour cream} => {whole milk} 0.01464159
[6]  {pip fruit,other vegetables}    => {whole milk}       0.01352313
[7]  {citrus fruit,root vegetables}  => {other vegetables} 0.01037112
[8]  {tropical fruit,root vegetables}=> {other vegetables} 0.01230300
[9]  {tropical fruit,root vegetables}=> {whole milk}       0.01199797
[10] {tropical fruit,yogurt}         => {whole milk}       0.01514997
[11] {root vegetables,yogurt}        => {other vegetables} 0.01291307
[12] {root vegetables,yogurt}        => {whole milk}       0.01453991
[13] {root vegetables,rolls/buns}    => {other vegetables} 0.01220132
[14] {root vegetables,rolls/buns}    => {whole milk}       0.01270971
[15] {other vegetables,yogurt}       => {whole milk}       0.02226741
```

[Ans] When mini confidence=0.9, no association rule is obtained. Only when the mini confidence is lowered to around 0.5, there are more than 10 rules obtained.

6. For minimum support = 0.01 and minimum confidence = 0.5, plot only the rules that have "whole milk" in their right hand side.

[Code]

```
Rules <- apriori(Groceries, parameter = list(support =
 0.01, confidence = 0.5))
Rules.sub <- subset(Rules, subset = rhs %pin% "whole milk")
inspect(Rules.sub)
```

[Result]

```
> Rules.sub <- subset(Rules, subset = rhs %pin% "whole milk")
> inspect(Rules.sub)
     lhs                                rhs          support    confidence lift
[1]  {curd,yogurt}                   => {whole milk} 0.01006609 0.5823529  2.279125
[2]  {other vegetables,butter}       => {whole milk} 0.01148958 0.5736041  2.244885
[3]  {other vegetables,domestic eggs} => {whole milk} 0.01230300 0.5525114  2.162336
[4]  {yogurt,whipped/sour cream}     => {whole milk} 0.01087951 0.5245098  2.052747
[5]  {other vegetables,whipped/sour cream} => {whole milk} 0.01464159 0.5070423  1.984385
[6]  {pip fruit,other vegetables}    => {whole milk} 0.01352313 0.5175097  2.025351
[7]  {tropical fruit,root vegetables} => {whole milk} 0.01199797 0.5700483  2.230969
[8]  {tropical fruit,yogurt}         => {whole milk} 0.01514997 0.5173611  2.024770
[9]  {root vegetables,yogurt}        => {whole milk} 0.01453991 0.5629921  2.203354
[10] {root vegetables,rolls/buns}    => {whole milk} 0.01270971 0.5230126  2.046888
[11] {other vegetables,yogurt}       => {whole milk} 0.02226741 0.5128806  2.007235
```

7. Among the rules produced in task 6, which ones have the highest lift? Can you explain these rules? How interesting are they?

[Code]

```
inspect(sort(Rules.sub, decreasing=T, by="lift")[1])
```

[Result]

```
> inspect(sort(Rules.sub, decreasing=T, by="lift")[1])
    lhs                rhs          support    confidence lift
[1] {curd,yogurt} => {whole milk} 0.01006609 0.5823529  2.279125
```

[Ans] 7.1 For the rules, I can find two explanation:
One is that people are likely to buy similar things together. That is to say, when people buy yogurt or curd or butter or cream, they are more likely to buy whole milk.
The other is people would like to buy vegetable with milk together. Perhaps it is because people can make salad with these food together.
7.2 All the rules found are interesting. According to these rules, the grocery should put various kinds of vegetable and milk products close to each other.