

Association Rules

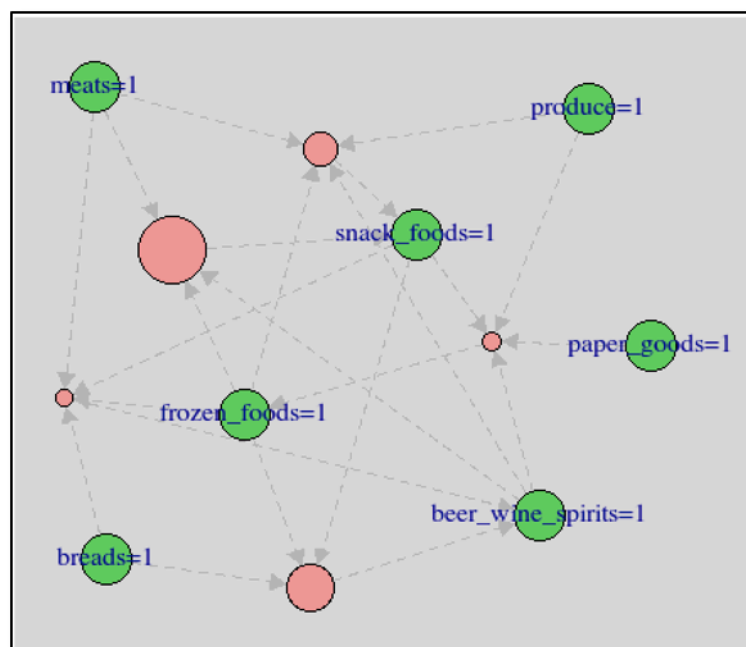
Cody Frisby

9/19/2017

Using R

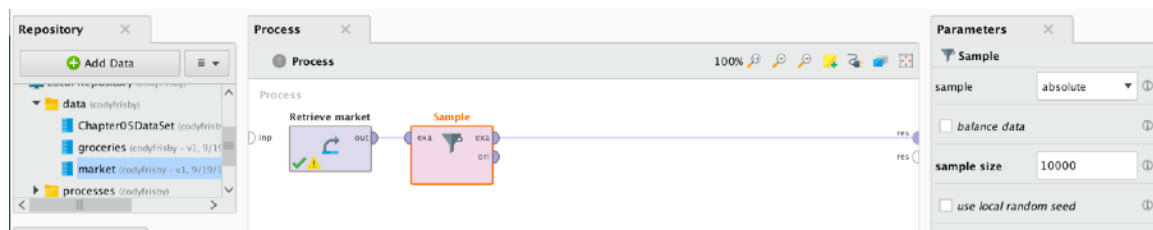
I used the dataset from canvas, *Supermarket Receipts*. Using the R package “arules” I set confidence = 0.5 and support = 0.01. I found over 240 rules. I study the top 5, based on confidence, for brevity.

Figure 1: Plot of the top 5 rules by confidence



Using RapidMiner

The data set that I used has 108,131 rows of data and 11 columns (excluding the ID column). This is too many rows for the version of RapidMiner that I have, so, after reading the data file in, I take a simple random sample of the data set (see screenshot below).



The model ran in R is the same in RapidStudio, namely support = 0.01 and confidence = 0.5 with all other values at the default. The top five, based on confidence are shown here.

Premises	Conclusion	Support	Confide... ↓
snack_foods, beer_wine_spirits, paper_goods, ...	frozen_foods	0.015	0.880
beer_wine_spirits, meats, paper_goods, frozen...	snack_foods	0.019	0.864
snack_foods, beer_wine_spirits, produce, breads	frozen_foods	0.011	0.860
snack_foods, beer_wine_spirits, meats, paper_...	frozen_foods	0.019	0.860
beer_wine_spirits, meats, frozen_foods	snack_foods	0.051	0.857

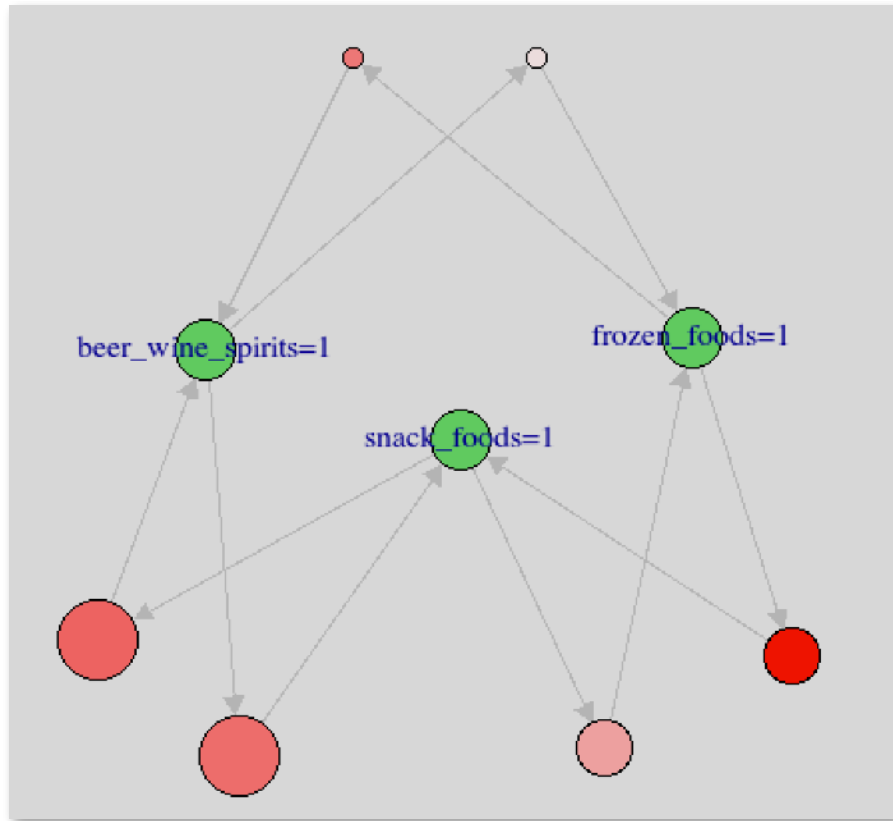
The two programs appear to match (at least the top 5) the same rules while their values for support and confidence are slightly different (this could be due to rounding differences and/or rounding errors) but they are very close so it is reasonable to assume they are producing the same result. Additionally, don't forget we are working with a subset of the data in RapidMiner due to the max dataset restriction.

The first rule, {paper goods, snack foods, beer/wine/spirits, product} => {frozen foods} makes some sense. Seems like this basket is picking up some supplies for a football watching party or similar.

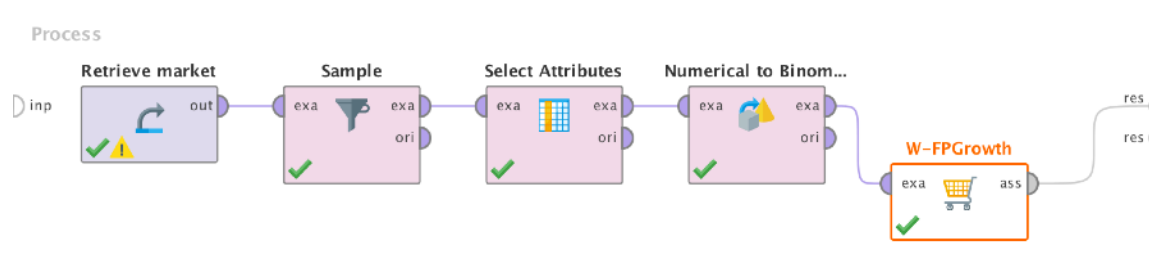
If we look at the top 6 based on support, we have simpler rules but lower confidence. But, a larger value for support indicates this rule appears in our dataset more than a rule with a lower value for support.

Premises	Conclusion	Support ↓	Confidence	LaPlace
snack_foods	frozen_foods	0.191	0.563	0.889
frozen_foods	snack_foods	0.191	0.587	0.899
snack_foods	beer_wine_spirits	0.189	0.557	0.888
beer_wine_spirits	snack_foods	0.189	0.562	0.890
beer_wine_spirits	frozen_foods	0.185	0.549	0.887
frozen_foods	beer_wine_spirits	0.185	0.568	0.894

We can visualize the 5 rules in another way using R and the "vizrules" package like we did above. As can be seen in this plot, the rules are a lot easier to understand visually. Seems as though we have some evidence that frozen foods, snacks, and/or beer/wine/spirits will be in baskets together.



My top 20 rules generated using the W-Fp Growth operator are shown here including a screenshot of my design. A quick glance at the first few rules it would appear that we found the same rules and that by default this design's output is ranking the rules by confidence.



W-FPGrowth

FPGrowth found 256 rules

1. [snack_foods=true, beer_wine_spirits=true, paper_goods=true, produce=true]: 167 ==> [frozen_foods=true]: 147 <conf:(0.88)> lift:(2.5)
2. [beer_wine_spirits=true, meats=true, paper_goods=true, frozen_foods=true]: 220 ==> [snack_foods=true]: 190 <conf:(0.86)> lift:(2.55)
3. [snack_foods=true, beer_wine_spirits=true, produce=true, breads=true]: 129 ==> [frozen_foods=true]: 111 <conf:(0.86)> lift:(2.64) le
4. [snack_foods=true, beer_wine_spirits=true, meats=true, paper_goods=true]: 221 ==> [frozen_foods=true]: 190 <conf:(0.86)> lift:(2.64)
5. [beer_wine_spirits=true, meats=true, frozen_foods=true]: 594 ==> [snack_foods=true]: 509 <conf:(0.86)> lift:(2.53) lev:(0.03) conv:f
6. [snack_foods=true, frozen_foods=true, breads=true]: 401 ==> [beer_wine_spirits=true]: 341 <conf:(0.85)> lift:(2.53) lev:(0.02) conv:
7. [beer_wine_spirits=true, meats=true, frozen_foods=true, desserts=true]: 146 ==> [snack_foods=true]: 124 <conf:(0.85)> lift:(2.5) lev
8. [beer_wine_spirits=true, frozen_foods=true]: 1849 ==> [snack_foods=true]: 1570 <conf:(0.85)> lift:(2.5) lev:(0.09) conv:(4.36)
9. [snack_foods=true, produce=true, frozen_foods=true, breads=true]: 131 ==> [beer_wine_spirits=true]: 111 <conf:(0.85)> lift:(2.52) le
10. [beer_wine_spirits=true, meats=true, produce=true, frozen_foods=true]: 286 ==> [snack_foods=true]: 242 <conf:(0.85)> lift:(2.49) le
11. [beer_wine_spirits=true, meats=true, frozen_foods=true, breads=true]: 149 ==> [snack_foods=true]: 126 <conf:(0.85)> lift:(2.49) lev
12. [snack_foods=true, beer_wine_spirits=true, produce=true]: 534 ==> [frozen_foods=true]: 449 <conf:(0.84)> lift:(2.58) lev:(0.03) cor
13. [beer_wine_spirits=true, frozen_foods=true, breads=true]: 406 ==> [snack_foods=true]: 341 <conf:(0.84)> lift:(2.48) lev:(0.02) conv
14. [snack_foods=true, beer_wine_spirits=true, paper_goods=true]: 538 ==> [frozen_foods=true]: 448 <conf:(0.83)> lift:(2.56) lev:(0.03)
15. [snack_foods=true, frozen_foods=true, juices=true]: 340 ==> [beer_wine_spirits=true]: 283 <conf:(0.83)> lift:(2.47) lev:(0.02) conv
16. [snack_foods=true, beer_wine_spirits=true]: 1891 ==> [frozen_foods=true]: 1570 <conf:(0.83)> lift:(2.55) lev:(0.1) conv:(3.96)
17. [snack_foods=true, beer_wine_spirits=true, breads=true]: 412 ==> [frozen_foods=true]: 341 <conf:(0.83)> lift:(2.54) lev:(0.02) conv
18. [snack_foods=true, beer_wine_spirits=true, meats=true, produce=true]: 293 ==> [frozen_foods=true]: 242 <conf:(0.83)> lift:(2.54) le
19. [snack_foods=true, meats=true, frozen_foods=true, breads=true]: 153 ==> [beer_wine_spirits=true]: 126 <conf:(0.82)> lift:(2.45) lev
20. [snack_foods=true, frozen_foods=true]: 1911 ==> [beer_wine_spirits=true]: 1570 <conf:(0.82)> lift:(2.44) lev:(0.09) conv:(3.71)