A PRIMER IN MARKET BASKET ANALYSIS

TOBIAS PFAFF

(t.pfaff@foodtracks.de)

WHAT?

- Affinity analysis

- Association rule learning

Analyze associations and connections between objects

WHY?

Actionable insights how objects are related

HOW?
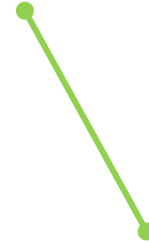
item (and itemsets)

transaction ID
(unique)

transaction

```
1   SalesRowId;ItemName
2   E3C7E1F6-B1DC-E611-9427-001E6797BCEE;PurDinkel   500g
3   E6C7E1F6-B1DC-E611-9427-001E6797BCEE;BIO-Röggelchen
4   E6C7E1F6-B1DC-E611-9427-001E6797BCEE;Kirschplunder           /=
5   E9C7E1F6-B1DC-E611-9427-001E6797BCEE;Sunny/Hähnchenbrust
6   EBC7E1F6-B1DC-E611-9427-001E6797BCEE;Käsestange              /-
7   EBC7E1F6-B1DC-E611-9427-001E6797BCEE;Becher Kaffee M
8   EEC7E1F6-B1DC-E611-9427-001E6797BCEE;Fünfkornfrischquark 750g
9   EEC7E1F6-B1DC-E611-9427-001E6797BCEE;Schoko-Croissant
10  EEC7E1F6-B1DC-E611-9427-001E6797BCEE;Schweineohren           /+
11  EEC7E1F6-B1DC-E611-9427-001E6797BCEE;Kaffeeglück             /+
```

ASSOCIATION RULES

{Rosinenschnecke, Striezel} $\Rightarrow$ {Kaffee}

antecedent / left-hand side / LHS   consequent / right-hand side / RHS

INTEREST MEASURES

Help focusing on more interesting association rules.

1. Support

2. Confidence

3. Lift

CONFIDENCE

```
         lhs                                          rhs                     support     confidence lift
[1]  {Mehrkornbrötchen,Vollkornbeisser}    => {Chia-Brötchen}      0.001380977 0.5714286  47.28980
[2]  {Roggenbrötchen,Vollkornbeisser}      => {Chia-Brötchen}      0.001208355 0.4666667  38.62000
[3]  {Chia-Brötchen,Vollkornbeisser}       => {Roggenbrötchen}     0.001208355 0.5000000  35.75926
```
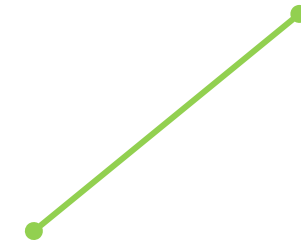
50 % chance that a transaction with {Chia-Brötchen, Vollkornbeisser} (LHS items) also contains {Roggenbrötchen} (RHS items)
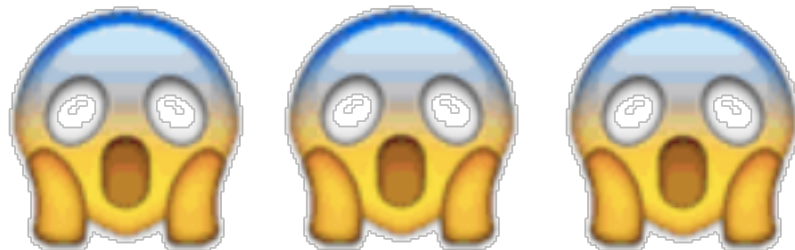
LIFT

```
       lhs                                          rhs                   support     confidence  lift
[1]    {Mehrkornbrötchen,Vollkornbeisser}     =>  {Chia-Brötchen}     0.001380977  0.5714286   47.28980
[2]    {Roggenbrötchen,Vollkornbeisser}       =>  {Chia-Brötchen}     0.001208355  0.4666667   38.62000
[3]    {Chia-Brötchen,Vollkornbeisser}        =>  {Roggenbrötchen}    0.001208355  0.5000000   35.75926
```

„ratio of the support of the items on the LHS of the rule co-occuring with items on the RHS divided by the probability that the LHS and RHS co-occur if the two are independent"

LIFT


```
    lhs                                          rhs                     support    confidence lift
[1] {Mehrkornbrötchen,Vollkornbeisser}    => {Chia-Brötchen}      0.001380977 0.5714286  47.28980
[2] {Roggenbrötchen,Vollkornbeisser}      => {Chia-Brötchen}      0.001208355 0.4666667  38.62000
[3] {Chia-Brötchen,Vollkornbeisser}       => {Roggenbrötchen}     0.001208355 0.5000000  35.75926
```

1. Greater lift, stronger associations

2. Look for rules with lift > 1

RESOURCES

- [R] http://discourse.snowplowanalytics.com/t/market-basket-analysis-identifying-products-and-content-that-go-well-together/1132

- [R] https://cran.r-project.org/web/packages/arulesViz/vignettes/arulesViz.pdf

- [R] https://github.com/mhahsler/arulesViz

- [R] https://github.com/mhahsler/arules

- [Python] http://aimotion.blogspot.de/2013/01/machine-learning-and-data-mining.html

- [Python] http://orange3-associate.readthedocs.io/en/latest/

- [Python] https://stackoverflow.com/questions/7047555/using-frequent-itemset-mining-to-build-association-rules

- [General] http://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html