CSCI446/946 Big Data Analytics

Week 5
Association Rules

School of Computing and Information Technology

University of Wollongong Australia

Task One

- Apply the Apriori algorithm to the "Groceries" dataset provided in R
  - Hint: Install `arules` and `arulesViz` packages
  - Load the dataset by using `data(Groceries)`
- Generate frequent itemsets satisfying
  - The minimum support threshold as 0.02
  - The minimum length of the itemsets as 1
  - The maximum length of the itemsets as 10

Task One

- Generate associate rules satisfying
  - The minimum support threshold as 0.001
  - The minimum confidence threshold as 0.6
- You shall be able to explain the involved commands and their outputs

Task Two

- Visualize the generated rules by plotting
  - Scatter plot of the generated rules
  - Scatter matrix on the support, confidence, and lift
  - Matrix-based visualization of the LHS and RHS of the rules whose confidence is above 0.9, coloured by lift and confidence
  - Graph visualization of the top five rules with the highest lift values
- You shall be able to explain these visualizations

Images Courtesy of Google Image