



Part 2: Association Analysis

Abhinav Kumar (akumar39),
Sachin Kumar Kuppayya (skuppayy)
Vinooth Rao Kulkarni (vinoothr)

Data Mining & Bioinformatics
10/1/2018

Describe Apriori algorithm with your own words and the workflow on how you mine association rules briefly.

- ❖ Generate the frequent item sets based on the support values. i.e. filter out item sets that have support (less than the) $<$ threshold.
- ❖ Next, generate rules from the frequent item sets according to the confidence i.e. filter out item sets that have confidence $<$ threshold.
- ❖ Filter out the rules based on the queries provided in the templates by selecting only those rules that satisfy the query. Process the above rules to generate Rules, Body, Head based on the query and count the number of rules that match the query.

Support = 30, 40, 50, 60, 70

support = 30%

number of length-1 frequent item sets: 196
number of length-2 frequent item sets: 5340
number of length-3 frequent item sets: 5287
number of length-3 frequent item sets: 1518
number of length-3 frequent item sets: 438
number of length-3 frequent item sets: 88
number of length-3 frequent item sets: 11
number of length-3 frequent item sets: 1

number of all lengths frequent item sets: 12879

support = 40%

number of length-1 frequent item sets: 167
number of length-2 frequent item sets: 753
number of length-3 frequent item sets: 149
number of length-3 frequent item sets: 7
number of length-3 frequent item sets: 1

number of all lengths frequent item sets: 1077

support = 50%

number of length-1 frequent item sets: 109
number of length-2 frequent item sets: 63
number of length-3 frequent item sets: 2

number of all lengths frequent item sets: 173

support = 60%

number of length-1 frequent item sets: 34

number of length-2 frequent item sets: 2

number of length-3 frequent item sets: 0

number of all lengths frequent item sets: 36

support = 70%

number of length-1 frequent item sets: 7

number of length-2 frequent item sets: 0

number of length-3 frequent item sets: 0

number of all lengths frequent item sets: 7

Number of Rules generated for Support 30, 40 ,50, 60, 70 & Confidence = 70

support = 30% and confidence = 70%: 31579

support = 40% and confidence = 70%: 1137

support = 50% and confidence = 70%: 117

support = 60% and confidence = 70%: 4

support = 70% and confidence = 70%: 0