Georgetown University

COSC 285: Introduction to Data Mining
Spring 2015
Faculty: Nazli Goharian

Task: Association Rule Mining

Date: March 24
Due Date & Time: April 24 (**No** grace period can be used !)

Grading: This assignment is 10 points out of the total 40 points allocated for all assignments in the semester. It will be graded on the scale of 100.

Assignment Description:

Use the provided data set, *small_basket.dat* for this assignment. This file contains a transactions and the list of items included in the transaction. Items are separated by integers (one integer for every distinct item). A zero entry indicates the item is not included in the transaction. A non-zero entry indicates the entry is included in this transaction. The separate file *products* indicates the name of each of the products. Use an Association Rule Mining algorithm that either uses anti-monotone property to avoid generating too many candidatess (such as A'Priori), or an algorithm that generates frequent itemsets without generating candidates (such as FP-Tree).

Runs and Results:

Perform your evaluation by picking different values for minimum support and minimum confidence thresholds. Create various combinations such that you can evaluate the effect of each these thresholds. For example (s=0.10, c=0.70) and (s=0.45, c=0.70) may show the affect of support when confidence is fixed. Plan at least 10 of these pairs of support, confidence to generate association rules. You may additionally calculate correlation to further filter out the rules that may have negative correlations, and only keep the positive correlations. List the rules that meet the thresholds. If based on your results you notice that you should change the min support and confidence, you should adjust and provide your results and analysis.

Report 1:

Min-support	Min-confidence	Number of frequent itemsets	Value of the largest K	Number of rules	Total Run Time (sec.)

Report 2:

Provide the top 10 rules (based on confidence & correlation) -- are they interesting?

Deliverables:

Cover page (1 pt): should contain the following in the exact order as specified:

a. Status of this assignment: Complete or Incomplete. If incomplete state clearly what is incomplete.
b.Time spent on this assignment. Approx. number of hours.
c.Things you wish you had been told prior to being given the assignment.

Design (10 pts): No code should be included in the design document. No specific template is provided. You may draw diagram(s) to show the architecture and the flow of your software components, and/or to provide the write-up of your software and design decisions.

Your working system, Results & Analysis (89 pts): See Report 1 & 2 for what you need to deliver. Your report should follow the template given under report 1.

You may be asked to give a demo and answer related questions.