EE4483/IM4483 Artificial Intelligence and Data Mining

Continuous Assessment - Homework

Due: Friday, 11 November 2016

You are to perform association analysis on a market basket transaction dataset "grocery.basket.txt". The dataset consists of a total of 150 transactions with 22 items {Apple, Banana, Coffee, Diaper, Egg, Fish, Ginger, Ham, IceCream, Jam, Ketchup, Lemon, Milk, Nuts, Olive, PeanutButter, Quiche, Rootbeer, Salad, Tea, Veg, Yogurt}. You may use any suitable data mining software tool(s) to analyze the data and answer the following questions.

(a) What is the maximum number of possible itemsets (including all 1-itemsets) and association rules (including all rules that have zero support and/or zero confidence) that can be generated from this dataset?

(b) What is the maximum size (in terms of number of items) of frequent itemsets (including those with *minsup* > 0) that can be extracted from this dataset?

(c) Which 2-itemset(s) and 3-itemset(s) have the largest support among all the frequent 2-itemsets and 3-itemsets, respectively?

(d) How many frequent itemsets have the minimum support of 12%, 10%, and 5%, respectively?

(e) What are the respective percentages of frequent 4-itemsets, 3-itemsets, and 2-itemsets, with respect to all possible itemsets, which have a minimum support of 2%?

(f) How many association rules have a minimum confidence of 50% and a minimum support of 5% and 10%, respectively?

(g) What are the two other items that customers most likely would buy when they buy Milk and Nuts?

(h) List three association rules that have the highest support with 100% confidence?

(i) Do you find any "interesting" rules? What are they? Briefly explain why.

(j) State what software tool(s) and functions you use to complete this homework.

Notes:

▪ You may use any suitable software tools for your work. Clearly state the software tools/functions used.

▪ You may consider using one of the following data mining software tools for your work.

  – Orange [http://orange.biolab.si]

  – Weka [http://www.cs.waikato.ac.nz/ml/weka/]

  – KNIME [https://www.knime.org]

  – Rapidminer [https://rapidminer.com]

  – R-Programming [https://www.r-project.org][http://www.rdatamining.com]

▪ If you couldn't find answers to any of the questions above, you may describe what you have done and attach the relevant screenshots or codes, if available.

▪ You are expected to uphold NTU Honour Code.

▪ Submit your work in hardcopy to Pauline Choo at S1-B1a-02 (Email: epchoo@ntu.edu.sg; Phone: 67905872) before the due date.