Midterm Project – Part 1

Create 30 items usually seen in Amazon, K-mart, or any other supermarkets (e.g. diapers, clothes, etc.).

(1) Create a database of 20 transactions each containing some of these items. The information can be stored in a file, or a DBMS (e.g. ORACLE).

(2) Repeat (1) by creating 4 additional, different databases each containing 20 transactions.

Using Apriori, generate and print out all the association rules and the input transactions for each of the 5 transaction databases you created (support and confidence should be user-determined parameter values, so the output should show different support and confidence values).

Midterm Project – Part 2

- Implement the brute force method and compare the brute force method with the Apriori algorithm on each of the 5 transaction databases you created. Present computation (CPU or clock) time to demonstrate that the Apriori algorithm is faster than the brute force method on each of the 5 transaction databases. The brute force method and Apriori algorithm should output the same association rules on each database.

- The brute force method for finding frequent itemsets works as follows. Enumerate and generate all possible 1-itemsets and 2-itemsets. There are 30 items, so there are 435 possible 2-itemsets totally. Check to see whether each possible 1-itemset/2-itemset is frequent. Then enumerate and generate all possible 3-itemsets. There are 4060 possible 3-itemsets totally. Check to see whether each possible 3-itemset is frequent. Keep on doing so until you see none of the possible k-itemsets is frequent for some k, at which point the brute force method terminates without generating $(k+1)$-itemsets.

Platforms are open

● Programming language is open; any one of the following is allowed: Python, Php, visual studio, PL/SQL, etc. Use   any programming language of your choice   (specify the programming language you use in the   project).

● Operating system is open; any one of the following is allowed: Windows, Solaris Unix, Linux, Mac OS, etc.

● Hardware is open; any one of the following is allowed: PC, Laptop, Sun Sparc, etc.

Project Grading

- There is a limit on the file size in Canvas. So, keep your project file small to avoid any problem that may occur when submitting the file in Canvas.

- The project file should contain the source code and documentation including **screenshots**. The screenshots are used to demonstrate the running situation of your program, particularly how the program executes and produces output based on different input data and user-specified parameter values.