## HW3

**Due Date: Nov. 18, 2020**

**Submission requirements:**

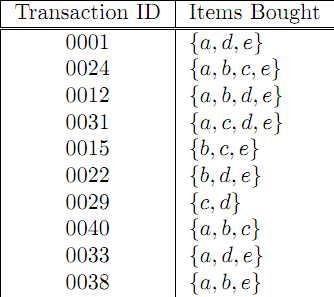
**Please submit your solutions to our class website.**

**Part I: written assignment**

1．Consider the data set shown in Table 1（min\_sup = 40%, min\_conf=75%）.

1. Find all frequent itemsets using Apriori by treating each transaction ID as a market basket.
2. Use the results in part (a) to compute the conﬁdence for the association rules {a, b}🡪{c} and {c}🡪{a, b}. Is conﬁdence a symmetric measure?
3. List all of the strong association rules (with support s and confidence c) matching the following Q, where X is a variable representing customers, and itemi denotes variables representing items (e.g. “A”, “B”, etc.):

Table 1. Example of market basket transactions.



2．Consider the data set shown in Table 1（min\_sup = 40%）.

(a) Find all frequent itemsets using FP-Growth. Please present all the FP-trees and all the conditional pattern bases.

(b) Compare the efficiency of Apriori and FP-Growth.

3. Table 2 gives a User-Product rating matrix.

Table 2. User-Product Rating Matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Product 1 | Product 2 | Product 3 | Product 4 |
| User 1 | 1 | 1 | 5 | 3 |
| User 2 | 3 | ？ | 5 | 4 |
| User 3 | 1 | 3 | 1 | 1 |
| User 4 | 4 | 3 | 2 | 1 |
| User 5 | 2 | 2 | 2 | 4 |

1. List the top 3 most similar users of user 2 based on Cosine Similarity.
2. Predict User 2’s rating for Product 2.

**Part II: Lab**

**Question 1**

Learn the use of market basket analysis for the purpose of making product purchase recommendations to the customers.

The data set contains transactions from a large supermarket. Each transaction is made by someone holding the loyalty card. We limited the total number of categories in this supermarket data to 20 categories for simplicity. The field value for a certain product in the transaction basket is 1 if the customer has bought it and 0 if he/she has not. The file named “Transactions” has data for 46243 transactions.

The data are available from the class web page.

Your submission should consist only of those deliverables marked indicated by “Hand-in”.

Market basket analysis has the objective to discover individual products, or groups of products that tend to occur together in transactions. The knowledge obtained from a market basket analysis can be employed by a business to recognize products frequently sold together in order to determine recommendations and cross-sell and up-sell opportunities. It can also be used to improve the efficiency of a promotional campaign.

Run Apriori on “transaction” data set. Set the “Type” of “COD” as “Typeless”, set the “direction” of all the other 20 categories as “Both”, set their “Type” as “Flag”. Set “Minimum antecedent support” to be 7%, “Minimum confidence” to be 45%, and “Maximum number of antecedents” to be 4 in the modeling node (Apriori node). In general you should explore by trying different values of these parameters to see what type of rules you get.

* **Hand-in**: The list of association rules generated by the model.

Sort the rules by lift, support, and confidence, respectively to see the rules identified. **Hand-in**: **For each case**, choose top 5 rules (note: make sure no redundant rules in the 5 rules) and give 2-3 lines comments. Many of the rules will be logically redundant and therefore will have to be eliminated after you think carefully about them.