COMP207 Exercises
Tutorial 10 (Week 12)

Frequent-Itemset Mining

This exercise is a minor variation of Exercise 22.1.1 in [2] and Exercise 28.14 in [1]. Consider the following table of transactions (purchases) by customers in a supermarket:

transaction_id	customer_id	items_bought
1	F	c, d
2	B	b, e
3	G	b, e
4	H	e, f, g
5	D	a, f
6	B	e, f
7	G	a, e, f
8	E	a, d
9	C	b, c, e, f
10	E	b, c, f
11	H	a, c
12	D	a, b, c, e
13	A	a, c, d
14	B	g

(a) What are the baskets with respect to the transactions (transaction_ids) and what are the baskets with respect to the customers (customer_ids)?

(b) What is the support with respect to the transactions and the support with respect to the customers of the following itemsets?

    (i) $\{a\}$

    (ii) $\{b\}$

    (iii) $\{e\}$

    (iv) $\{f\}$

    (v) $\{a, f\}$

    (vi) $\{b, e\}$

    (vii) $\{b, f\}$

    (viii) $\{e, f\}$

    (ix) $\{b, e, f\}$

(c) Let the support threshold be $s = 0.6$. Which of the itemsets in (b) are frequent with respect to the transactions, which are frequent with respect to the customers?

(d) What is the confidence of the following association rules with respect to the transactions and with respect to the customers?

  (i) $\{e\} \Rightarrow f$

  (ii) $\{f\} \Rightarrow a$

  (iii) $\{f\} \Rightarrow b$

  (iv) $\{b,e\} \Rightarrow f$

(e) The *lift* of an association rule is a measure of interestingness of that rule. The *lift* of an association rule $\{i_1, \ldots, i_n\} \Rightarrow j$ with respect to the customers is defined as

$$lift(\{i_1, \ldots, i_n\} \Rightarrow j) = \frac{conf(\{i_1, \ldots, i_n\} \Rightarrow j)}{supp(\{j\})}$$

where $conf(\{i_1, \ldots, i_n\} \Rightarrow j)$ is the confidence of $\{i_1, \ldots, i_n\} \Rightarrow j$ with respect to the customers, and $supp(\{j\})$ is the support of $\{j\}$ with respect to the customers. If we use the association rule for recommending items, then we want the lift of the rule to be as large as possible. A lift of 1 would mean that there is no difference between suggesting j to users who also bought $i_1, \ldots, i_n$ and suggesting a random item to those users. A lift larger than 1 indicates that the rule can suggest more useful items than choosing a random item. The larger the lift the better the recommendation.

Determine the lift of the association rules from (d).

(f) Use the A-Priori Algorithm to find all sets J of items such that the support of J with respect to the customers is at least $s = 0.5$.

References

[1] Ramez Elmasri and Shamkant B. Navathe. *Fundamentals of Database Systems*. Pearson Education, 7th edition, 2016.

[2] Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom. *Database Systems - The Complete Book*. Pearson Education, 2nd edition, 2009.