Big Data Analytics

ESSEC

Olga Klopp

Solution Home work 6 (Frequent Itemsets)

1. (**Exercise 6.1.1 MMDS book)** Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, also numbered 1 to 100. Item i is in basket b if and only if i divides b with no remainder. Thus, item 1 is in all the baskets, item 2 is in all fifty of the even-numbered baskets, and so on. Basket 12 consists of items $\{1, 2, 3, 4, 6, 12\}$, since these are all the integers that divide 12. Answer the following questions:

   (a) If the support threshold is 5, which items are frequent?

   (b) what is the confidence of the following association rules?

      i. $\{5, 7\} \rightarrow 2$

      ii. $\{2, 3, 4\} \rightarrow 5$

   Solution:

   (a) The items that are frequent are $\{1, 2, 3, 4, 5, \ldots, 20\}$ because these all appear in at least 5 baskets.

   (b) what is the confidence of the following association rules?

      i. Confidence= 1/2 as 5 and 7 will appear together in basket no. 35 and 70 and 2 will appear along with 7 and 5 in basket no. 70.

      ii. Confidence= 1/8 as $\{2, 3, 4\}$ appear in baskets having basket number multiple of 12 i.e in baskets no $\{12, 24, 36, 48, 60, 72, 84, 96\}$ and $\{2, 3, 4, 5\}$ appear together only in basket no 60.

2. (**Exercise 6.1.3 MMDS book)** Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, also numbered 1 to 100. Item i is in basket b if and only if b divides i with no remainder. For example, basket 12 consists of items

$$\{12, 24, 36, 48, 60, 72, 84, 96\}$$

   (a) If the support threshold is 5, which items are frequent?

   (b) what is the confidence of the following association rules?

      i. $\{24, 60\} \rightarrow 8$

   Solution:

   (a) All numbers that have at least 5 different dividers (including 1 and itself). For example 16: it has 3 different dividers (2,4, 8) and it's in the baskets 1,2,4,8 and 16.

   (b) what is the confidence of the following association rule?

      i. Confidence = 3/6 = 1/2 as 60 appears in basket no. 1,2,3,4,5,6,10,12,15,20,30,60 then 60 and 24 will appear together in basket no. 1,2,3,4,6,12 and 8 will appear along with 60 and 24 in basket no. 1,2,4.

3. (**Apriori algorithm) Solution:**
   We have the support threshold 1/3 which implies that the threshold is at least 2 transactions. Applying Apriori:

Pass (k)	Candidate k-itemsets and their support	Frequent k-itemsets
$k = 1$	milk(4), bread(2), juice(2), Coke(3), Chips(4)	milk, bread, juice, Coke, Chips
$k = 2$	$\{milk, bread\}(2)$, $\{milk, juice\}(1)$, $\{milk, coke\}(2)$, $\{milk, chips\}(2)$ $\{bread, juice\}(1), \{bread, Coke\}(0), \{bread, Chips\}(0)$ $\{juice, chips\}(1), \{juice, Coke\}(0), \{Coke, Chips\}(3)$	$\{milk, bread\}$ $\{milk, coke\}$ $\{milk, chips\}$, $\{Coke, Chips\}$
$k = 3$	$\{milk, Coke, Chips\}(2)$	$\{milk, Coke, Chips\}$
$k = 4$		

Note that there is no need to go to $k = 4$ since the longest transaction has only 3 items.

Association rules:

- $\{milk, bread\}$ would generate: $milk \rightarrow bread$ $(2/6 = 0.33, 2/4 = 0.5)$ and $bread \rightarrow milk$ $(2/6 = 0.33, 2/2 = 1)$;
- $\{milk, Coke\}$ would generate: $milk \rightarrow Coke(0.33, 0.5)$ and $Coke \rightarrow milk(2/6 = 0.33, 2/3 = 0.66)$;
- $\{Milk, Chips\}$ would generate: $Milk \rightarrow Chips(0.33, 0.5)$ and $Chips \rightarrow milk(2/6 = 0.33, 2/4 = 0.5)$;
- $\{Milk, Coke, Chips\}$ would generate: $Milk \rightarrow \{Coke, Chips\}(2/6 = 0.33, 2/4 = 0.5), Coke \rightarrow \{Chips, Milk\}(2/6 = 0.33, 2/3 = 0.66), Chips \rightarrow \{Coke, Milk\}(2/6 = 0.33, 2/4 = 0.5), \{Milk, Coke\} \rightarrow Chips(2/6 = 0.33, 2/2 = 1), \{Milk, Chips\} \rightarrow Coke(2/6 = 0.33, 2/2 = 1)$ and $\{Coke, Chips\} \rightarrow Milk(2/6 = 0.33, 2/3 = 0.66)$.

With the confidence threshold set to 60%, the Strong Association Rules are (sorted by confidence):

(a) Coke → Chips (0.5, 1)

(b) Bread → Milk (0.33, 1);

(c) $\{Milk, Coke\}$ → Chips(0.33, 1)

(d) $\{Milk, Chips\}$ → Coke(0.33, 1)

(e) Chips → Coke (0.5, 0.75);

(f) Coke → Milk (0.33, 0.66);

(g) Coke → $\{Chips, Milk\}$ (0.33, 0.66)

(h) $\{Coke, Chips\}$ → Milk(0.33, 0.66).