Tania Soutonglang

CS 422-01

Oct. 24, 2022

Homework 5

* 1. Tan, Ch. 5
     1. Exercise 15

Answer the following questions using the data sets shown in Figure 5.34. Note that each data set contains 1,000 items and 10,000 transactions. Dark cells indicate the presence of items and white cells indicate the absence of items. We will apply the Apriori algorithm to extract frequent itemsets with minsup=10% (i.e., itemsets must be contained in at least 1,000 transactions).

1. Which data set(s) will produce the most number of frequent itemsets?  
   A or E – A has item sets with 200 items and 200 transactions, whereas E has many item sets that are also overlapping .
2. Which data set(s) will produce the fewest number of frequent itemsets?  
   D – the item sets only have 1 transaction.
3. Which data set(s) will produce the longest frequent itemset?  
   E – a lot of the item sets are overlapping.
4. Which data set(s) will produce frequent itemsets with highest maximum support?  
   B – item 100 occurs has almost 8,000 transactions.
5. Which data set(s) will produce frequent itemsets containing items with wide-varying support levels (i.e., items with mixed support, ranging from less than 20% to more than 70%)?  
   E – the item sets have a lot of varying transaction lengths.
   1. Zaki, Ch. 8
      1. Exercise 1(a)  
         Given the database in Table 8.2. Using minsup = 3/8, show how the Apriori algorithm enumerates all frequent patterns from this dataset.

|  |  |
| --- | --- |
| **itemset** | **support** |
| {A} | 5 |
| {B} | 4 |
| {C} | 5 |
| {D} | 6 |
| ~~{E}~~ | ~~1~~ |
| {F} | 4 |
| {G} | 5 |

|  |  |
| --- | --- |
| **itemset** | **support** |
| {A, B} | 3 |
| {A, C} | 3 |
| {A, D} | 4 |
| ~~{A, F}~~ | ~~2~~ |
| ~~{A, G}~~ | ~~2~~ |
| ~~{B, C}~~ | ~~2~~ |
| ~~{B, D}~~ | ~~2~~ |
| ~~{B, F}~~ | ~~1~~ |
| ~~{B, G}~~ | ~~2~~ |
| {C, D} | 4 |
| ~~{C, F}~~ | ~~2~~ |
| {C, G} | 3 |
| {D, F} | 4 |
| {D, G} | 3 |
| ~~{F, G}~~ | ~~2~~ |

|  |  |
| --- | --- |
| **itemset** | **support** |
| ~~{A, B, C}~~ | ~~1~~ |
| ~~{A, B, D}~~ | ~~2~~ |
| {A, C, D} | 3 |
| ~~{C, D, G}~~ | ~~2~~ |
| ~~{D, F, G}~~ | ~~2~~ |

The Apriori algorithm determines {A, C, D} is the most frequent pattern in the dataset

* + 1. Exercise 4  
       Given the database in Table 8.4. Show all rules that one can generate from the set ABE.

|  |  |
| --- | --- |
| **itemset** | **support** |
| {A} | 4 |
| {B} | 5 |
| {E} | 4 |

|  |  |
| --- | --- |
| **itemset** | **support** |
| {A, B} | 3 |
| {A, E} | 2 |
| {B, E} | 4 |

|  |  |
| --- | --- |
| **itemset** | **support** |
| {A, B, E} | 2 |

|  |  |  |
| --- | --- | --- |
| **itemset** | **work** | **confidence** |
| {A → BE} |  | 0.50 |
| {B → AE} |  | 0.40 |
| {E → AB} |  | 0.50 |
| {BE → A} |  | 0.50 |
| {AE → B} |  | 1.00 |
| {AB → E} |  | 0.67 |