

# Bài tập Lý thuyết môn Khai phá Dữ liệu

Họ và tên: Huỳnh Nguyễn Thế Dân

MSSV: 21110256

Lớp: 21TTH1

## Questions

6. Represent the transaction database of Exercise 5 in vertical format.

7. Determine the confidence of the rules  $\{a\} \Rightarrow \{f\}$ , and  $\{a, e\} \Rightarrow \{f\}$  for the transaction database in Exercise 1.

8. Determine the confidence of the rules  $\{a\} \Rightarrow \{f\}$ , and  $\{a, e\} \Rightarrow \{f\}$  for the transaction database in Exercise 5.

9. Show the candidate itemsets and the frequent itemsets in each level-wise pass of the Apriori algorithm in Exercise 1. Assume an absolute minimum support level of 2.

---

## Answer

**\*For exercise 6:**

To represent the transaction database from Exercise 5 in a vertical format, we list each item along with the set of transaction IDs (tids) in which the item appears. This format is particularly useful for efficient computation of itemset supports in frequent pattern mining, allowing easier intersection of sets to find common transactions.

**Here is the vertical representation of the database:**

Items	Appears
a	{1, 2}
b	{3, 4, 5}
c	{1, 3, 6, 7}
d	{1, 2, 3, 4, 6, 8}
e	{1, 2, 3, 4, 5, 6, 7, 8}
f	{2, 3, 4, 5, 7, 8}

### Explanation:

- a: Appears in transactions 1 and 2.
- b: Appears in transactions 3, 4, and 5.
- c: Appears in transactions 1, 3, 6 and 7.
- d: Appears in transactions 1, 2, 3, 4, 6 and 8.
- e: Appears in transactions 1, 2, 3, 4, 6, 7 and 8.
- f: Appears in transactions 2, 3, 4, 5, 7 and 8.

This vertical format helps in quickly identifying the transactions associated with each item, facilitating operations like intersections for computing supports of larger itemsets.

---

### \*For exercise 7:

To determine the confidence of the association rules  $\{a\} \Rightarrow \{f\}$ , and  $\{a, e\} \Rightarrow \{f\}$  from the given transaction data, we'll first count the occurrences of each itemset in the transactions.

#### Association Rule Confidence Formula:

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{Support}(A,B)}{\text{Support}(A)}$$

where:

- $\text{Support}(A)$  is the proportion of transactions that contain the itemset A.
- $\text{Support}(A, B)$  is the proportion of transactions that contain both itemsets A and B.

#### The support counts for each itemset in the transactions are as follows:

- $\text{Support}(a) = 3$
- $\text{Support}(a, f) = 2$
- $\text{Support}(a, e) = 2$
- $\text{Support}(a, e, f) = 2$

### Confidence Calculations:

Rule:  $\{a\} \Rightarrow \{f\}$

$$\text{Confidence}(a \Rightarrow f) = \frac{\text{Support}(a,f)}{\text{Support}(a)} = \frac{2}{3} \approx 0.67 \text{ or } 67\%$$

Rule:  $\{a, e\} \Rightarrow \{f\}$

$$\text{Confidence}(a, e \Rightarrow f) = \frac{\text{Support}(a,e,f)}{\text{Support}(a,e)} = \frac{2}{2} = 1.0 \text{ or } 100\%$$

---

### \*For exercise 8:

**The support counts for each itemset in the transactions are as follows:**

- $\text{Support}(a) = 2$
- $\text{Support}(a, f) = 1$
- $\text{Support}(a, e) = 2$
- $\text{Support}(a, e, f) = 1$

## Confidence Calculations:

**Rule:  $\{a\} \Rightarrow \{f\}$**

$$\text{Confidence}(a \Rightarrow f) = \frac{\text{Support}(a,f)}{\text{Support}(a)} = \frac{1}{2} = 0.5 \text{ or } 50\%$$

**Rule:  $\{a, e\} \Rightarrow \{f\}$**

$$\text{Confidence}(a, e \Rightarrow f) = \frac{\text{Support}(a,e,f)}{\text{Support}(a,e)} = \frac{1}{2} = 0.5 \text{ or } 50\%$$

---

## \*For exercise 9:

Let's proceed with the Apriori algorithm for the given data, where each transaction includes a set of items and we need to find frequent itemsets with a minimum support threshold of 2.

### Step 1: Count the frequency of each item

- a: Transactions 1, 3, 4
- b: Transactions 1, 2, 5
- c: Transactions 1, 2
- d: Transactions 1, 3, 5
- e: Transactions 2, 3, 4
- f: Transactions 2, 3, 4, 5

### Step 2: Determine Frequent 1-itemsets (L1)

- $\{a\}$  appears 3 times.
- $\{b\}$  appears 3 times.
- $\{c\}$  appears 2 times.
- $\{d\}$  appears 3 times.
- $\{e\}$  appears 3 times.
- $\{f\}$  appears 4 times.

All items have a support of at least 2, so all are frequent.

### Step 3: Generate Candidate 2-itemsets (C2) from (L1)

- Possible combinations of the frequent 1-itemsets are:
  - $\{a, b\}, \{a, c\}, \{a, d\}, \{a, e\}, \{a, f\}$
  - $\{b, c\}, \{b, d\}, \{b, e\}, \{b, f\}$

- {c, d}, {c, e}, {c, f}
- {d, e}, {d, f}
- {e, f}

#### Step 4: Determine Frequent 2-itemsets (L2)

- Count the occurrences of each 2-itemset:
  - {a, b}: Transaction 1 (1 time)
  - {a, c}: Transaction 1 (1 time)
  - {a, d}: Transactions 1, 3 (2 times)
  - {a, e}: Transactions 3, 4 (2 times)
  - {a, f}: Transactions 3, 4 (2 times)
  - {b, c}: Transactions 1, 2 (2 times)
  - {b, d}: Transactions 1, 5 (2 times)
  - {b, e}: Transaction 2 (1 time)
  - {b, f}: Transactions 2, 5 (2 times)
  - {c, d}: Transaction 1 (1 time)
  - {c, e}: Transaction 2 (1 time)
  - {c, f}: Transaction 2 (1 time)
  - {d, e}: Transaction 3 (1 time)
  - {d, f}: Transactions 3, 5 (2 times)
  - {e, f}: Transactions 2, 3, 4 (3 times)
- *Frequent 2-itemsets:*
  - {a, d}, {a, e}, {a, f}, {b, c}, {b, d}, {b, f}, {d, f}, {e, f}

#### Step 5: Generate Candidate 3-itemsets (C3) from (L2)

- Possible combinations of the frequent 2-itemsets are:
  - {a, d, e}, {a, d, f}, {a, e, f}
  - {b, c, d}, {b, c, f}, {b, d, f}
  - {d, e, f}

#### Step 6: Determine Frequent 3-itemsets (L3)

- Count the occurrences:
  - {a, d, e}: Transaction 3 (1 time)
  - {a, d, f}: Transactions 3 (1 time)
  - {a, e, f}: Transactions 3, 4 (2 times)
  - {b, c, d}: Transaction 1 (1 time)
  - {b, c, f}: Transaction 2 (1 time)
  - {b, d, f}: Transactions 5 (1 time)
  - {d, e, f}: Transactions 3 (1 time)
- *Frequent 3-itemsets:*
  - {a, e, f}

## Conclusion

- Frequent 1-itemsets (L1): {a}, {b}, {c}, {d}, {e}, {f}
- Frequent 2-itemsets (L2): {a, d}, {a, e}, {a, f}, {b, c}, {b, d}, {b, f}, {d, f}, {e, f}
- Frequent 3-itemsets (L3): {a, e, f}

This breakdown covers each step of the Apriori algorithm, detailing the candidate and frequent itemsets at each level.

---

**End.**