Iran University of Science and Technology

School of Computer Engineering

Second Assignment

ADVANCED DATA MINING, SPRING 2025

DR. MINAEI

Asma Rashidian, Zahra Amiri

Due Sunday ,Ordibehesht 7, 1404

Problem 1)

a) Explain the brute-force method for generating frequent itemsets, including its inherent limitations. Additionally, discuss how modern algorithms such as Apriori and FP-Growth overcome these limitations to improve efficiency and scalability.

b) Explain the key idea of the ECLAT algorithm and how its data representation differs from Apriori. Additionally, define 'itemset lattice traversal' and describe its role in ECLAT and Apriori.

Problem 2)

Given the following transaction dataset:

TID	Items
T1	A, B, C
T2	A, D
T3	B, C, D
T4	A, C, D
T5	A, B, C, D

Use Apriori algorithm with minsup = 60% and minconf = 70% to:

a) Find all frequent itemsets (up to size 3).

b) Generate all strong association rules.

Problem 3)

a) Given a transactional database and a minimum support threshold of 33.34%, construct the Frequent Pattern Tree (FP-Tree). For each transaction, illustrate how the tree evolves step by step as transactions are inserted.

b) Apply the FP-Growth algorithm to the constructed FP-Tree to identify all frequent itemsets that meet the specified support threshold.

TID	Items
T1	Milk, Bread, Eggs
T2	Milk, Diapers, Beer, Bread
T3	Milk, Diapers, Beer, Cola
T4	Bread, Diapers, Beer
T5	Bread, Milk, Cola
T6	Diapers, Cola

Problem 4)

Most frequent pattern mining algorithms consider only distinct items in a transaction. However, multiple occurrences of an item in the same shopping basket, such as four cakes and three jugs of milk, can be important in transactional data analysis. How can one mine frequent itemsets efficiently considering multiple occurrences of items? Propose modifications to the well-known algorithms, such as Apriori and FP-growth, to adapt to such a situation.

Problem 5) Analyzing Customer Purchase Patterns using Association Rules and Frequent itemsets with Orange:

To analyze customer purchase patterns in a retail store using the "Market Basket" dataset. This exercise will employ association rules and frequent itemsets to identify patterns such as "if a customer purchases product X, how likely are they to purchase product Y?".

Steps:

a. Load and Examine the "Market Basket" Dataset:

- Import the "Market Basket" dataset into a suitable programming environment.
- Explore the dataset's structure, including the list of products purchased by each customer.

b. Data Visualization and Preprocessing:

- Visualize the distribution of products purchased across customers.
- Identify and handle any missing values or inconsistencies in the data.

c. Apply Association Rule Mining Algorithms:

- Set appropriate support and confidence thresholds to filter out irrelevant or insignificant patterns.

d. Analyze and Evaluate Association Rules:

- Interpret the extracted association rules, identifying frequent product co-occurrences.
- Evaluate the business implications of these rules, such as cross-selling opportunities or product substitution strategies.

e. Explore Frequent Itemsets:

- Identify frequent itemsets, which represent groups of products frequently purchased together.
- Analyze these frequent itemsets to gain insights into customer preferences and purchase behaviors.

Note:

Please present your answers in the following format:

1. Step-by-step explanation for each of the five steps (a–e), including screenshots using Orange GUI.
2. Visual outputs:
    - Charts (e.g., bar plots for product frequency)
    - Frequent itemset tables
    - Rule visualizations (e.g., rule graphs or lists with metrics)
3. Interpretation Section:
    - Brief summary of findings
    - Key takeaways for business decision-making
4. Optional Code Snippets (if using Python or another coding tool along with Orange)