Université de Monastir
Institut Supérieur d'Informatique et de Mathématiques de Monastir
Département des sciences de l'informatique
Filière : Ingénieur en Informatique
Responsable du cours : Mariem Gzara

TP structures de données avancées en PYTHON
Implémentation Python de l'algorithme a priori

1. **Objectif du TP :** L'objectif de ce TP est d'implémenter efficacement l'algorithme "a priori".
2. **Description détaillée de l'algorithme "a priori"**

L'algorithme "**a priori"** est un algorithme d'apprentissage non supervisé qui génère les "patterns" fréquents. Dans notre implémentation, nous voulons générer les ensembles d'items fréquents à partir d'une base de données transactionnelle.

Entrée de l'algorithme :

- Base de données transactionnelles avec chaque transaction est un ensemble d'items.
- K : taille maximum d'un ensemble d'items.

Sortie de l'algorithme : Ensemble des items fréquents de taille inférieur à k.

The Apriori Algorithm (Pseudo-Code)

C_k: Candidate itemset of size k
L_k : frequent itemset of size k

L_1 = {frequent items};
for (k = 1; L_k !=∅; k++) **do begin**
  C_{k+1} = candidates generated from L_k;
  for each transaction t in database do
    increment the count of all candidates in C_{k+1} that are contained in t
  L_{k+1} = candidates in C_{k+1} with min_support
  end
return $\cup_k L_k$;

November 13, 2023 Machine Learning Algorithms-Mariem Gzara 173

Implementation of Apriori

- How to generate candidates?
  - Step 1: self-joining L_k
  - Step 2: pruning
- The join step:
  - All items within a transaction or itemset are sorted in lexicographic order (or any kind of order)
    For the k-itemset l_i, the items are sorted such that $l_i[1]<l_i[2]<...l_i[k]$
  - $c_{k+1} = L_k \rhd\lhd L_k$: the joint operation
    - $l_1 \in L_k$, $l_2 \in L_k$, l_1 and l_2 are joined if $(l_1[1]=l_2[1]) \wedge (l_1[2]=l_2[2]) \wedge ... \wedge (l_1[k-1]=l_2[k-1]) \wedge (l_1[k-1]<l_2[k-1])$
    - The resulting itemset formed by joining l_1 and l_2 is l1[1],l1[2], ..., l1[K],l2[k]

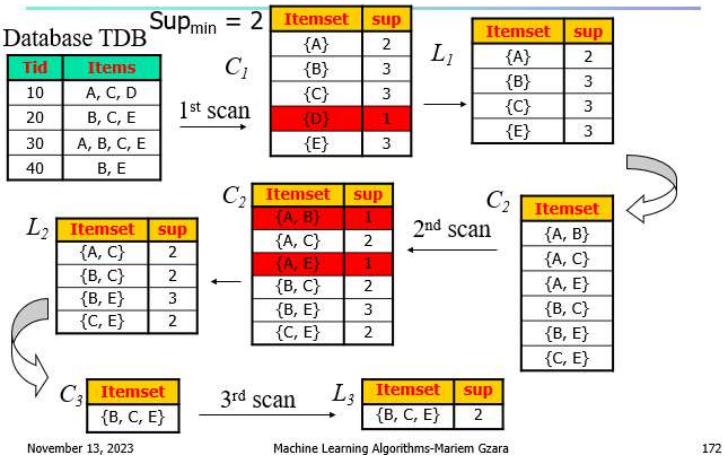
November 13, 2023 Machine Learning Algorithms-Mariem Gzara 174

Implementation of Apriori

- The prune step:
  - $L_{k+1} \subset C_{k+1}$: C_{k+1} is a superset of L_{k+1}
  - If any k-subset of a candidate (k+1)-itemset is not in L_k, then the candidate can be removed from C_{k+1}

3. Exécution à la main de l'algorithme "a priori"

The Apriori Algorithm—An Example



4. Travail demandé

a. Bien comprendre le fonctionnement de l'algorithme "a priori" et faire une exécution à la main.

b. Proposer les structures de données adéquates pour coder les transactions, les ensembles d'items fréquents, les listes L_k et C_k.

c. Proposer une implémentation efficace de l'algorithme "a priori".

Indication : dict, set, frozenset, orderedDict, list, chainMap.

BON courage