

Frequent Pattern (FP) Growth Algorithm

Groupe 1

UYI

FS

INFO M1

INF4117: Fouille de données

16 décembre 2021

TABLE – Membres Groupe I

Nom	Prénom	Matricule
BINELI OLA	Mathilda Maguy	17T2215
DJIEMBOU TIENTCHEU	Victor Nico	17T2051
KENFACK TEMGOUA	Vanessa	17J2871
Coordonateur :		
Dr TSOPZE NORBERT		

Introduction

L'extraction de connaissance dans les bases de données, également appelé **data mining**, désigne le processus permettant d'extraire des informations et des connaissances utiles qui sont enfouies dans les bases de données, les entrepôts de données (data warehouse) ou autres sources de données.[3]

Depuis sa création, le Data Analytique joue un rôle important dans le processus de prise de décision, du coup plusieurs algorithmes Frequent Pattern Mining (FPM) ont été développés pour améliorer les performances d'extraction.[3]

Durant cette présentation, nous nous intéresserons à l'algorithme de croissance de modèles fréquents dans l'extraction de connaissances.[3]

Le but de cette étude est de présenter tout d'abord ce que s'est que **l'Algorithme de croissance de modèles fréquents** (*Frequent Pattern (FP) Growth Algorithm*, en anglais), quels sont les atouts ? et quelles sont ses limites ?[3]

Définition

Items : Est tout objet, article, attribut, littéral appartenant à un ensemble fini d'éléments distincts $I = \{x_1, x_2, x_3, \dots, x_n\}$. En outre, il s'agit d'un ensemble d'attribut de la base de données transactionnelle.

Itemset : C'est un ensemble de **N** items.

Itemset Frequent : On dit qu'un itemset est fréquent si et seulement si son support est supérieur à un support minimum définit par l'utilisateur.

support minimal : Noté **Minsup** est le nombre minimum d'occurrence d'un itemset pour être considéré comme fréquent.

Définition

Transaction : Soit $I = \{i_1, i_2, \dots, i_m\}$ un ensemble d'items, D une BD de transactions où chaque transaction T est un sous ensemble de I.

classification d'algorithme d'extraction de motifs fréquents

Les algorithmes d'extraction de motifs fréquents peuvent être classés en deux grandes catégories :
Algorithmes générer et tester (candidate generate and test)
et en **croissance de Modèle (pattern growth)**.

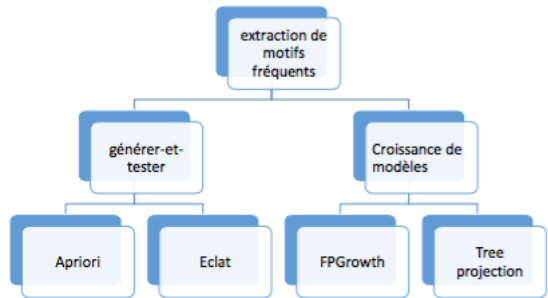


FIGURE – classification d'algorithme d'extraction de motifs fréquents [3]

Énoncé

Étant donnée un base de transactions T , un seuil minimum de support s , le problème est de pouvoir extraire l'ensemble des modèles ou motifs fréquents tels que le support de chaque itemsets frequents soit supérieur ou égal au minimum de support s définit par l'utilisateur [1].

Exemple d'application réel

Ce problème est communément rencontré dans les domaines tels :

- ▶ **Dans le domaine de l'éducation** : Extraction des règles d'association dans le data mining des étudiants admis à travers les caractéristiques et les spécialités.
- ▶ **Dans le domaine médical** : Par exemple, analyse de la base de données des patients.
- ▶ **En foresterie** : Analyse de la probabilité et de l'intensité des feux de forêt avec les données sur les feux de forêt.

Exemple d'application réel

- ▶ **la fonction de saisie semi-automatique** : par Google
- ▶ **Système de recommandation** : est utilisé par de nombreuses entreprises comme Amazon
- ▶ **Analyse du panier de la ménagère**
- ▶ **Détection des potentiels futurs clients**
- ▶ **Détection de fraude**

Principe

Le principe d'application de l'algorithme Frequent Pattern (FP) Growth est le suivant :

- ▶ On effectue un premier parcours de la base de transactions T , pour déterminer les items fréquents en fonction du support minimum fourni. Ces items seront triés par la suite par ordre décroissant de support dans une liste L . Les items ainsi triés seront traités dans cet ordre.[3]

Principe

- Un second parcours de T est alors effectué. Chaque transaction est alors trié selon l'ordre des items dans L . Le noeud racine de l'arbre $\{\text{null}\}$ est d'abord créé. Durant ce même parcours, une branche sera créée pour chaque transaction, mais les transactions ayant un même préfixe partageront le même début d'une branche de l'arbre, ainsi deux transactions identiques seront représentées par une et même branche.[3]

Algorithme FP Growth

Algorithm 1 Frequent Pattern (FP) Growth Algorithm

Require: un support seuil s et la base de transactions T .

Ensure: liste des itemsets fréquents.

- 1: $L \leftarrow$ liste de items de T dont les support sont $\geq s$ classée suivant s décroissant
- 2: trier T dans l'ordre décroissant de support dans L .
- 3: construire le FP-tree des données transactionnelles de T
- 4: de FP-tree, ressortir le FP-conditional tree pour chaque item (itemset)
- 5: déterminer les modèles fréquents.

Limites

- ▶ L'algorithme de Frequent Pattern (FP) Growth résout le problème de la nécessité de nombreuses analyses de la base de données, vue qu'il ne fait que deux balayages de la base de transactions. [4]
- ▶ Néanmoins, cela ne garantit pas, dans le cas où la base de transactions est trop volumineuse, que toute la structure du FP-tree tiendra en mémoire central [2]. De plus la construction du FP-tree peut s'avérer longue et pourrait consommer beaucoup de ressources système.[5]

Exemple

Soit donnée la base de transaction T ci-dessous et $s = 2$ le minimum de support. Trouvons tout les motifs fréquents en utilisant l'algorithme de croissance de modèles frequents.

Dans cette base, les différents caractères sont des substituants d'articles dans un carrefour.[4]

- ▶ A : bière
- ▶ B : vin
- ▶ C : chips
- ▶ D : oeuf

TABLE – Base de transactions

T	Items
1	A,B
2	B, C, D
3	A, C, D, E
4	A, D, E
5	A, B, C

Exemple : étape 1 et 2

TABLE – Liste des items fréquents de T

L	Support
A	4
B	3
C	3
D	3
E	2

TABLE – Base de transactions suivant L

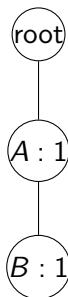
T	Ordered Frequent Items
1	A, B
2	B, C, D
3	A, C, D, E
4	A, D, E
5	A, B, C

Exemple : étape 3 i)

Constuire le FP-tree et ajouter T1.

TABLE – Base de transactions suivant L

T	Ordered Frequent Items
1	A,B

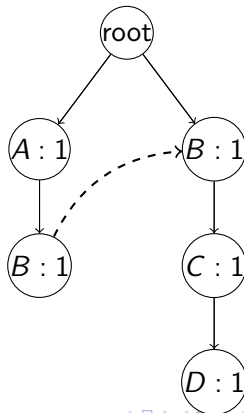


Exemple : étape 3 ii)

Ajouter T2 dans le FP-tree
transaction de la base T.

TABLE – Base de transactions
suivant L

T	Ordered Frequent Items
1	A,B
2	B,C,D

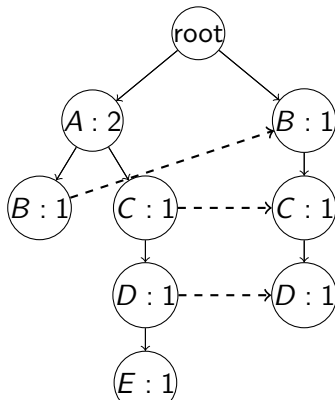


Exemple : étape 3 iii)

Ajouter T3 dans le FP-tree
transaction de la base T.

TABLE – Base de transactions
suivant L

T	Ordered Frequent Items
1	A,B
2	B,C,D
3	A,C,D, E

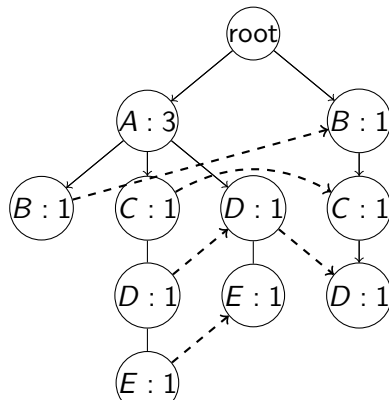


Exemple : étape 3 iv)

Ajouter T4 dans le FP-tree
 transaction de la base T.

TABLE – Base de transactions
 suivant L

T	Ordered Frequent Items
1	A,B
2	B,C,D
3	A,C,D, E
4	A,D, E

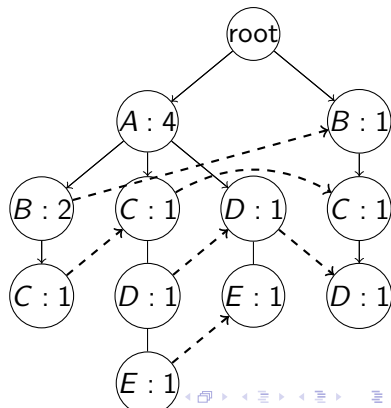


Exemple : étape 3 v)

Ajouter T5 dans le FP-tree
transaction de la base T.

TABLE – Base de transactions
suivant L

T	Ordered Frequent Items
1	A,B
2	B,C,D
3	A,C,D, E
4	A,D, E
5	A,B,C



Exemple : étape 4

Trouver les itemsets fréquents.

TABLE – Extraction de motifs fréquents

Item	Conditional Pattern base	Conditional FP-tree	Frequent Pattern Generated
E	$\{\{A, C, D : 1\}, \{A, D : 1\}\}$	$\langle A : 2 \rangle, \langle D : 2 \rangle$	$\{E, A : 2\}, \{E, D : 2\}, \{E, A, D : 2\}$
D	$\{\{A, C : 1\}, \{A : 1\}, \{B, C : 1\}\}$	$\langle A : 2 \rangle$	$\{D, A : 2\}$
C	$\{\{A, B : 1\}, \{A : 1\}, \{B : 1\}\}$	$\langle A : 2 \rangle$	$\{C, A : 2\}$
B	$\{\{A : 2\}\}$	$\langle A : 2 \rangle$	$\{B, A : 2\}$

Présenté Par :

Introduction

Définitions

classification d'algorithme d'extraction de motifs fréquents

Problème

Solution proposée

Exemple

Résumé

Références

Remerciement

Résumé





Département Informatique Doua IUT Lyon 1. *Problème.*

URL : <https://perso.liris.cnrs.fr/pierre-antoine.champin/enseignement/algo/cours/algo/probleme.html>.



SOFTWARETESTINGHELP 2021. *Frequent Pattern (FP) Growth Algorithm In Data Mining.* URL :

<https://www.softwaretestinghelp.com/fp-growth-algorithm-data-mining/>.



DAAOU Zeneddine et KERIOUI ABDERRAHIM.

Extractions des motifs fréquents orientés besoins du décideur.
République Algérienne Démocratique et Populaire, Ministère
de l'Enseignement Supérieur et de la Recherche Scientifique,
p. 50-55.



Joos KORSTANJE. *The FP Growth algorithm*. URL : <https://towardsdatascience.com/the-fp-growth-algorithm-1ffa20e839b8>.



Great LEARNING. *Understanding (Frequent Pattern) FP Growth Algorithm — What is FP Algorithm?* URL : <https://www.mygreatlearning.com/blog/understanding-fp-growth-algorithm/>.

	Présenté Par :
	Introduction
	Définitions
classification d'algorithme d'extraction de motifs fréquents	Problème
	Solution proposée
	Exemple
	Résumé
	Références
	Remerciement

Fin

**MERCI DE
VOTRE
ATTENTION**

