

Bordures de motifs fréquents et transversaux minimaux d'hypergraphes

Aghilas Smail Idriss Felloussi

Travail d'Etudes et de Recherche
Master 2 Informatique (SID) 2023 - 2024

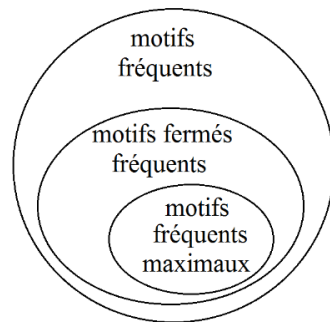
February 26, 2024

Table des matières

- 1 Introduction
- 2 Méthode existante
- 3 Comparatif expérimental
- 4 Méthode probabiliste proposée
- 5 Experimentations avec l'approche probabiliste
- 6 Remerciments

Introduction

- Identification des motifs fréquents et leurs dérivés.
- La bordure positive et négative.
- Problèmes rencontrés :
 - Vaste espace de recherche.
 - Nombre élevé de motifs produits.



- Découverte de motifs fréquents.
- Contexte de fouille de données.
- Trouver l'ensemble S des motifs fréquents :

Id	Items				
t1	A	B	E	G	
t2	B	C	E	G	
t3	A	B	E	H	
t4	A	B	D	H	
t5	B	C	F	H	
t6	B	C	E	F	H

minsup = 3

AB frequent (support=3)

BC frequent (support=3)

EG et FH infrequent (support=2)

Méthode existante

Définition 1 (Hypergraphe) :

Un hypergraphe $H = (V, E)$ est constitué d'un ensemble V de sommets et d'un ensemble E d'hyperarêtes, avec $\forall e \in E, e \subseteq V$.

Définition 2 (Traverse et traverse minimale) :

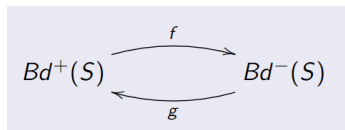
Soit un hypergraphe H , les traverses, notées $Tr(H)$, sont définies comme l'ensemble de tous les sous-ensembles τ de sommets V tels que chaque hyper-arête e_i possède une intersection non vide avec τ .

Définition 3 (Bordures positive et négative)

La bordure positive (resp. négative) de S , notée $Bd^+(S)$ (resp. $Bd^-(S)$), est constituée par les motifs fréquents maximaux.

La bordure négative de S , notée $Bd^-(S)$, est constituée par les motifs non fréquents minimaux.

- La méthode d'approximation des bordures repose sur le concept de dualisation, impliquant deux fonctions, f et g



- Nouvelle fonction f' qui utilise un calcul de traverses minimales approchées.

- La méthode proposée est constituée de 2 phases :
 - Réduction de l'hypergraphe H . avec algorithme glouton conçu pour le calcul de traverses minimales (se base sur les intersections des hyperarêtes et le nombre d'occurrences de chaque sommet).
 - Calcul des traverses minimales approximatifs de l'hypergraphe réduit H_r .

Comparatif expérimental

- Une version multithreadée (HRmulti).
- Une version monothreadée (HR).
- 6 jeux de donnée on été exploité sur les 2 version HR et HRmulti.
 - 1 ac (accidents)
 - 2 bms (BMS-WebView2)
 - 3 Mushroom
 - 4 Chess
 - 5 Connect
 - 6 Kosarak

- Les figures 1, 2, 3, 4, 5 et 6 présentent les résultats pour les jeux de données accidents ,bms ,chess,connect ,kosarak et mushroom.

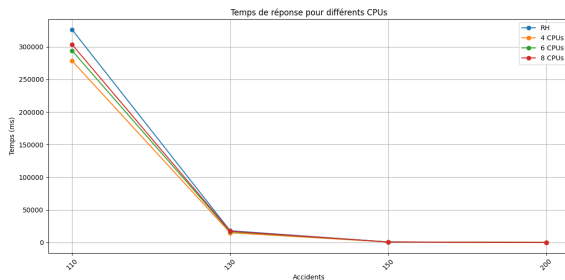


Figure 1 : Accidents

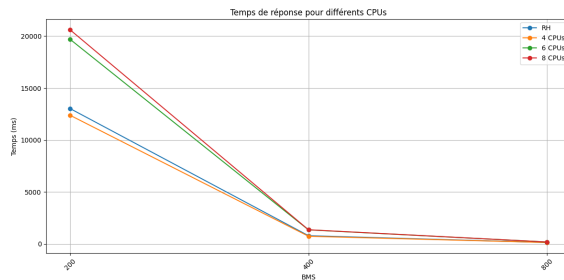


Figure 2 : BMS

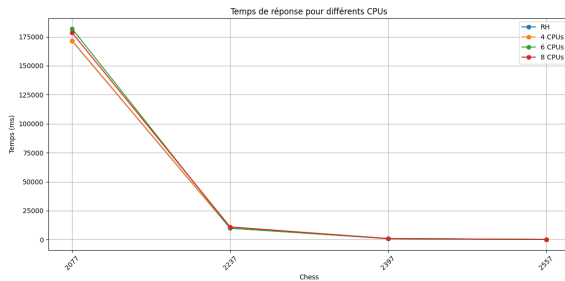


Figure 3 : Chess

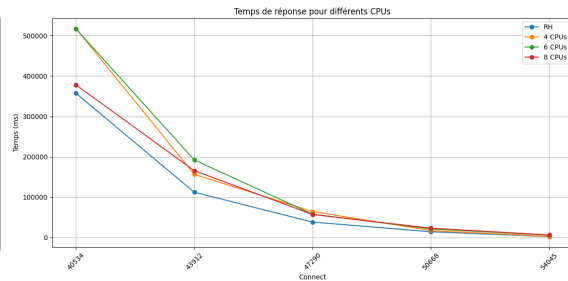


Figure 4 : Connect

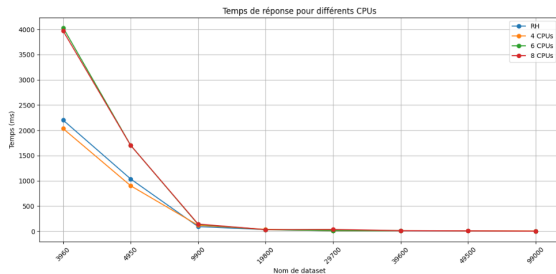


Figure 5 : kosarak

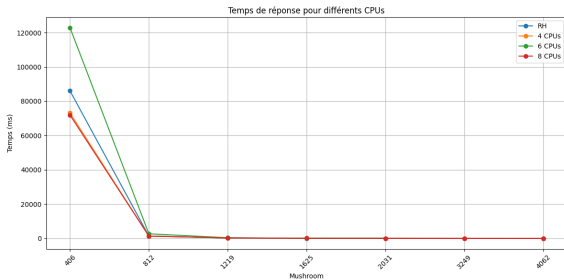


Figure 6 : Mushroom

Matériel utilisée : ordinateur portable à 8 coeurs physiques avec une RAM de 1638 Mo avec le système d'exploitation Windows avec Java comme langage de programmation.

Méthode probabiliste proposée

La méthode probabiliste consiste à appliquer une réduction ciblée de l'hypergraphe initial par des pourcentages différents.

Cette approche permet de réduire significativement la complexité du calcul des transversaux.

L'utilisation d'une approche stochastique pour la réduction d'hypergraphes est particulièrement utile dans plusieurs contextes:

- Grande variabilité des propriétés.
- Préservation des propriétés statistiques.
- Obtention de différentes instances.

Méthodologie :

- Un algorithme de réduction d'hypergraphe basé sur l'échantillonnage aléatoire de sommets.
- Un algorithme d'estimation de transversales minimales basé sur une approche stochastique.

Experimentations avec l'approche probabiliste

- Le résultat final est une liste de sommets représentant la couverture minimale de sommets dans l'hypergraphe.

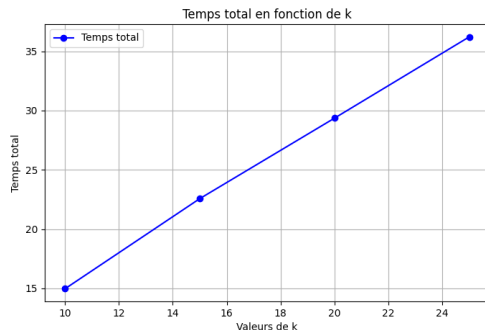


Figure: Temps d'exécution en fonction de k en secondes

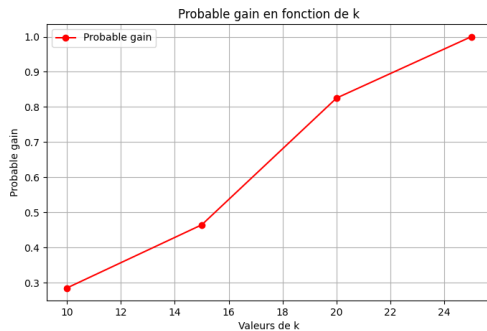


Figure: Gain probable en fonction du nombre d'itérations k

- Pour $k = 25$, le gain probable est de 1.0, indiquant que toutes les transversales minimales trouvées avec une réduction de 20% sont présentes dans la réduction de 10%.
- La méthodologie offre une perspective sur la sensibilité de l'algorithme aux variations de seuil et sa robustesse dans la découverte de transversaux minimaux.

Remerciements

Merci de votre attention !