



UNIVERSITÉ  
DE MONTPELLIER

# Extraction de Bases pour les Règles d'Association à partir des Itemsets Fermés Fréquents

Etudiant : TRAN Thi Tra My - 21511002

21 Mai 2020

## **I. Résumé l'article :**

Le problème de l'utilité et de la pertinence des règles d'association extraites est primordial dans Data mining. Les règles d'association ont été bien appliquées dans plusieurs domaines (*commercial, diagnostic, médicale, etc*). Afin d'extraire des règles d'association, il est nécessaire d'effectuer 4 phases (*Préparation des données, Extraction des ensembles fréquents d'attributs, Génération des règles d'association, Interprétation des résultats*). Mais il existe 2 problèmes majeurs qui ont donné lieu à de nombreuses recherches : *Problème des temps et Problème de la pertinence et l'utilité*.

Pour résoudre le premier problème, l'auteur propose d'utiliser des algorithmes d'extraction des itemsets fréquents maximaux et d'extraction des itemsets fermés fréquents. L'algorithme Close a été utilisé par l'auteur comme exemple pour l'extraction des itemsets fermés fréquents. Quant au deuxième problème, les méthodes permettant de réduire le nombre de règles d'association extraites, ou bien de sélectionner un sous-ensemble de règles ont été proposées.

En utilisant les itemsets fermés fréquents et leurs générateurs, la base générique pour les règles d'association exactes et la base informative pour les règles d'association approximatives sont définies. L'union de ces bases donne un ensemble générateur non redondant pour tous les règles d'association valides, leurs supports et leurs confiances. Au point de vue d'utilisateur, des règles d'association non redondantes minimales sont les plus utiles et plus pertinentes.

L'auteur présente au dernier des applications des algorithmes au jeu de données afin de montrer l'efficacité et l'utilité des bases proposées.

## **II. La différence entre des itemsets clos et des itemsets maximaux :**

Pour les itemsets maximaux, tous les sur-ensembles de ces itemsets sont inféquents. Mais tous les sur-ensembles des itemsets clos sont moins fréquents (pas inféquents).

## **III. Le nombre d'itemsets maximaux est supérieur ou inférieur au nombre d'itemsets clos :**

A mon avis, le nombre d'itemsets maximaux est inférieur au nombre d'itemsets clos car selon cet article, "*Les itemsets fréquents maximaux sont des itemset fermés fréquents maximaux*".

C'est à dire, les itemsets fermés fréquents contiennent les itemsets maximaux.

## **IV. A partir des itemsets maximaux d'obtenir l'ensemble des itemsets clos :**

Oui, C'est possible d'obtenir de l'ensemble des itemsets clos à partir des itemsets maximaux parce qu'on sait déjà que les itemsets maximaux sont aussi les itemsets clos et on connaît aussi leur support. En réalisant un balayage, on aurait pu déterminer le support de tous les itemsets fréquents à partir de cela on pourrait déterminer l'ensemble des itemsets clos. D'après l'article "*À partir des itemsets*

*fréquents maximaux, tous les itemsets fréquents sont dérivés et leurs supports sont déterminés en réalisant un balayage du contexte”.*

## **V. Application sur projet**

L'utilisation des règles d'association dans un classification nous permettra de trouver la présence des règles d'association utilisant les mots et les classes préparés dans la phase pré-traités. Nous pouvons donc trouver tous les règles ce qui nous permettra de mieux prédire en utilisant les méthodes de classification plus efficaces. Avec les règles d'association nous allons pouvoir mieux prédire la vérité d'un article en appliquant les règles d'association utiles et pertinents trouvés. La précision serait donc plus correcte.