

# FCP-Growth, une adaptation de FP-Growth pour générer des règles d'association de classe

Emna-Bahri\*, Stéphane-Lallich\*

\*Laboratoire ERIC, Université de Lyon ; 5, avenue Pierre Mendès-France, 69500, Bron  
emna.bahri|stephane.lallich@univ-lyon2.fr ; <http://eric.univ-lyon2.fr>

**Motivations.** La classification associative (Liu et al., 1998) prédit la classe à partir de règles d'association particulières, dites règles d'association de classe. Ces règles, dont le conséquent doit être la variable indicatrice de l'une des modalités de la classe, s'écrivent  $A \rightarrow c_i$ , où  $A$  est une conjonction de descripteurs booléens et  $c_i$  est la variable indicatrice de la  $i_e$  modalité de classe. L'intérêt des règles de classe est de permettre la focalisation sur des groupes d'individus, éventuellement très petits, homogènes du point de vue des descripteurs et présentant la même classe. Pour extraire les règles de classe, les méthodes de classification associative procèdent par filtrage des règles générées par les algorithmes d'extraction de règles d'association développés en non-supervisé. Dans une première étape, ces algorithmes extraient tous les itemsets plus fréquents que le seuil, puis ils en déduisent toutes les règles dont la confiance dépasse le seuil de support, ce qui pose différents problèmes. Dans la première étape, on extrait des itemsets fréquents inutiles, ceux qui ne contiennent pas la classe, alors que la seconde étape peut être simplifiée, puisqu'un itemset contenant la classe ne donne lieu qu'à une seule règle de classe. Afin de pouvoir travailler avec des seuils de support le plus bas possible, nous proposons FCP-Growth une adaptation de FP-Growth qui élimine les itemsets fréquents ne contenant pas de classe. En outre, pour ne pas désavantager la classe la moins nombreuse, le seuil de support utilisé dans chaque classe est proportionnel à la taille de la classe.

**Etat de l'art.** A l'opposé d'Apriori qui génère des itemsets candidats et qui les teste pour ne conserver que les itemsets fréquents, FP-Growth (Han et al. (2000)) construit les itemsets fréquents sans génération de candidats. Tout d'abord, il compresse les itemsets fréquents représentés dans la base de données à l'aide des FP-Tree (*frequent-pattern tree*) dont les branches contiennent les associations possibles des items. Chaque association peut être divisée en fragments qui constituent les itemsets fréquents. La méthode FP-Growth transforme le problème de la recherche de l'itemset fréquent le plus long par la recherche du plus petit et sa concaténation avec le suffixe correspondant (le dernier itemset fréquent de la branche aboutissant à l'item considéré). Ceci permet de réduire le coût de la recherche. Dans notre étude, nous avons retenu FP-Growth, en raison de sa structuration (FP-Tree) qui le rend plus efficace qu'Apriori.

**Contribution : FCP-Growth.** FCP-Growth, l'algorithme que nous proposons pour construire directement les itemsets de classe fréquents, repose sur plusieurs principes :

- au cours de la construction du FP-Tree, il élimine les itemsets qui ne sont pas de classe, Le gain de temps d'exécution et de stockage obtenu doit permettre de diminuer le seuil de support, ce qui nous aidera à trouver des pépites de classe).
- il utilise un seuil de support adaptatif au sens où le seuil utilisé dans chaque classe est proportionnel à la taille de la classe, dans le but de ne pas pénaliser les petites classes.