

Relaxation des Requêtes Skyline : Une Approche Centrée Utilisateur

Djamal Belkasmi^{*,**}, Allel Hadjali^{**}, Hamid Azzoune^{***}

^{*}DIF-FS UMBB Boumerdes, Algérie
djamal.belkasmi@ensma.fr

^{**}LIAS-ENSMA Poitiers, France
allel.hadjali@ensma.fr

^{***}LRIA/USTHB, Alger, Algérie
azzoune@yahoo.fr

Résumé. Les requêtes skyline constituent un outil puissant pour l'analyse de données multidimensionnelles et la décision multicritère. En pratique, le calcul du skyline peut conduire à deux scénarios : soit (i) un nombre important d'objets sont retournés, soit (ii) un nombre réduit d'objets sont retournés, ce qui peut être insuffisant pour la prise de décisions. Dans cet article, nous abordons le second problème et proposons une approche permettant de le traiter. L'idée consiste à rendre le skyline plus permissive en lui ajoutant les objets, non skyline, les plus préférés. L'approche s'appuie sur une nouvelle relation de dominance floue appelée «Much Preferred». Un algorithme efficace pour calculer le skyline relaxé est proposé. Une série d'expériences sont menées pour démontrer la pertinence de l'approche et la performance de l'algorithme proposé.

1 Introduction

Les requêtes skyline Börzsönyi et al. (2001) constituent un outil puissant d'analyse de données en vue de prendre des décisions intelligentes face à des données à grande échelle. Elles permettent d'extraire l'ensemble des points les plus intéressants quand différents critères, souvent conflictuels, sont pris en compte. Elles s'appuient sur le principe de dominance de Pareto. Soit D un ensemble de points à d dimensions, une requête skyline calcule l'ensemble des points non dominés dans D . Un point p domine (au sens de Pareto) un point q si et seulement si p est meilleur ou égal à q sur toutes les dimensions et strictement meilleur que q sur au moins une dimension. Par conséquent, les points skyline sont incomparables. Plusieurs études ont été menées pour développer des algorithmes efficaces et introduire des variantes pour les requêtes skyline Chomicki et al. (2013); Yiu et Mamoulis (2007); Khalefa et al. (2008); Pei et al. (2007); Hadjali et al. (2010). Toutefois, l'interrogation d'un ensemble de données multidimensionnelles à l'aide de l'opérateur skyline peut conduire à deux scénarios possibles : soit (i) un nombre important de réponses sont retournées, ce qui est généralement peu informatif et