

Calcul efficace du skyline basé sur l'indexation dimensionnelle

Rui Liu*, Dominique Li*,**

*Département d'Informatique, Université de Tours

**Laboratoire d'Informatique Fondamentale et Appliquée de Tours
{liurui, dominique.li}@univ-tours.fr

Résumé. Le calcul du skyline vise à trouver l'ensemble exact des tuples dominants parmi un ensemble de données multidimensionnelles, pour lequel de nombreux algorithmes ont été développés lors des deux dernières décennies. Cependant, la plupart des algorithmes existants ont été conçus et optimisés pour les données de faible dimensionnalité. Dans cet article, nous proposons SDI, un algorithme progressif basé sur l'indexation dimensionnelle pour le calcul du skyline. Notre analyse théorique et nos expériences montrent que SDI surpasse les algorithmes de calcul du skyline de l'état-de-l'art tant sur des données de haute dimensionnalité que sur des données de faible dimensionnalité.

1 Introduction

Le calcul du skyline vise à trouver l'ensemble complet de tuples dominants à partir des données multidimensionnelles. La figure 1 illustre un exemple du skyline de la relation prix-distance, où l'on considère les prix (axe Y) d'hôtels en fonction de leurs distances (axe X) à un endroit (le centre-ville, la plage ou la gare). Dans cet exemple, 6 hôtels sont représentés sous forme de points 2D dans lesquels $\{a, c, f\}$ forment le skyline puisque leurs valeurs, tant en prix qu'en distance, ne sont pas moins meilleures que les autres (*meilleur* est synonyme de plus petit dans notre exemple). Par exemple, l'hôtel a est meilleur que l'hôtel b , l'hôtel f est meilleur que les hôtels $\{d, e\}$ et l'hôtel c est incomparable avec tous les autres hôtels.

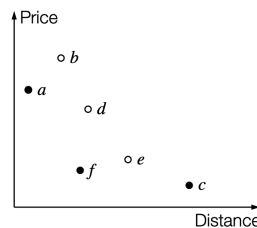


FIG. 1 – Un exemple de skyline.

Afin de réaliser les calculs du skyline, de nombreux algorithmes ont été développés au cours de plusieurs décennies. Cependant, la plupart des algorithmes existants ont été conçus et