Contribution au calcul du skyline par réduction de l'espace candidat

Lougmiri Zekri*,Hadjer Belaicha*

*Déparetement d'Informatique Université d'Oran,
BP 1524 EL-Mnaouer, Maraval Oran, Algérie
{lougmiri, belaichahadjer}@gmail.com

Résumé. L'opérateur skyline est devenu un paradigme dans les bases de données. Il consiste à localiser Sky l'ensemble des points d'un espace vectoriel qui ne sont pas dominés. Cet opérateur est utile lorsqu'on n'arrive pas à se décider dans les situations conflictuelles. Le calcul des requêtes skyline est pénalisé par le nombre de points que peuvent contenir les bases de données. Dans ce papier, nous présentons une solution analytique pour la réduction de l'espace candidat et nous proposons une méthode efficace pour le calcul de ce type de requêtes

1 Introduction

Les requêtes skyline sont importantes dans les applications qui nécessitent la localisation des réponses selon plusieurs critères. Ayant un ensemble de points dans un espace vectoriel de d dimensions, un algorithme traitant ce type de requêtes doit retourner l'ensemble des points de S dits non dominés. Il est meilleur pour ce type d'algorithmes de fonctionner progressivement Kossmann et al. (2002) car les utilisateurs sont souvent impatients de recevoir des réponses.
La relation de dominance se définit comme suit Börzsonyi et al. (2001) : Soit S un ensemble de données de d dimensions (d critères) sur lequel va porter l'opérateur skyline. Soit D l'ensemble de toutes les dimensions $D = \{d_1, , d_d\}$. Soient p et q deux points de S. La relation de dominance ($\prec$) suivant D est pour $1 \leq i, j \leq d$,p domine $q \Longleftrightarrow \{\forall d_i \in D, p(i) \leq q(i)\}$ et $\{\exists d_j \in D, p(j) < q(j)\}$. Lorsque aucun point ne domine l'autre on dit qu'ils sont non dominés ou concurrents. L'opérateur skyline renvoie l'ensemble des points concurrents, suivant toutes les dimensions D : $SkyD(S) = \{p \in S / \nexists q \in S : q \prec p\}$
Dans ce papier, nous présentons une solution analytique pour déduire l'ensemble de points candidats afin d'éviter de parcourir l'ensemble S entièrement. Nous donnons un nouveau théorème pour l'élimination des points non candidats. Notre méthode est basée sur le tri Tan et al. (2001) et à la différence avec ce travail, où les tests entre les points balayent tout l'ensemble S, nous donnons des théorèmes pour la déduction des points les plus évidents et qui constituent les premières solutions à présenter. Nous montrerons que la combinaison de DC Divide-and-Conquer avec notre méthode fournit des résultats meilleurs que lorsqu'il est appliqué tout seul.
Le reste de ce papier se présente comme suit. La section 2 présente les travaux liés à cette problématique. La section 3 donne notre approche. La section 4 présente les résultats des expérimentations. La conclusion et les travaux futurs sont donnés dans la section 5.