

Requêtes skyline en présence d'exceptions

Hélène Jaudoin, Olivier Pivert, Daniel Rocacher

Université de Rennes 1 – Irisa

{jaudoin@enssat.fr, pivert@enssat.fr, rocacher@enssat.fr}

Résumé. Dans cet article, nous nous intéressons à la recherche des points les plus intéressants au sens de l'ordre de Pareto, i.e., à l'évaluation de requêtes « skyline », dans des jeux de données présentant des anomalies. Il n'est pas rare que les données, de petites annonces par exemple, soient peuplées d'erreurs ou d'exceptions qui peuvent perturber la recherche des meilleurs points car celles-ci sont susceptibles de dominer les autres points. L'approche présentée vise à calculer les requêtes skyline malgré la présence de ces exceptions, sans pour autant les écarter définitivement, et à présenter graphiquement les résultats de façon à identifier rapidement les points d'intérêt et les anomalies potentielles.

1 Introduction

In database research, the last two decades have witnessed a growing interest in preference queries on the one hand. Motivations for introducing preferences inside database queries are manifold Hadjali et al. (2008). First, it has appeared to be desirable to offer more expressive query languages that can be more faithful to what a user intends to say. Second, the introduction of preferences in queries provides a basis for rank-ordering the retrieved items, which is especially valuable in case of large sets of items satisfying a query. Third, a classical query may also have an empty set of answers, while a relaxed (and thus less restrictive) version of the query might be matched by items in the database.

Approaches to database preference queries may be classified into two categories according to their qualitative or quantitative nature Hadjali et al. (2008).

Dans la dernière, les préférences sont exprimées quantitativement grâce à une fonction de score monotone, le score global étant positivement corrélé avec les scores partiels. Dans les approches qualitatives, les préférences sont définies au travers de relations binaires. Comme ces relations peuvent être définies en termes de fonctions de score, cette famille est plus générale que la précédente.

Dans cet article, une vision qualitative est adoptée, à savoir l'approche dite Skyline, introduite dans (Börzsönyi et al. (2001)). Étant donné un ensemble de points dans l'espace, une requête skyline retrouve les points qui ne sont dominés par aucun autre au sens de l'ordre de Pareto. Ce problème correspond à la recherche des extrema dans un ensemble de vecteurs (Kung et al. (1975)). Quand le nombre de dimensions sur lesquelles les préférences sont exprimées devient grand, de nombreux tuples peuvent être incomparables. Quelques approches ont été proposées pour définir un ordre entre deux tuples incomparables dans le contexte des requêtes skyline, fondées sur :