

Construction Incrémentale de Graphes de Voisinage avec Accès Réduit aux Disques

Hakim Hacid*, Tetsuya Yoshida**

* Alcatel-Lucent Bell Labs France
Centre de Villarceaux Route de Villejust, 91620 Nozay, France
hakim.hacid@alcatel-lucent.com,

**Hokkaido University,
Graduate School of Information Science and Technology,
N-14 W-9, Sapporo 060-0814, Japan
yoshida@meme.hokudai.ac.jp

Résumé. La recherche efficace de voisinage dans un espace multidimensionnel a été étudiée dans plusieurs domaines tels que la reconnaissance de formes et l'exploration de données. Les graphes de voisinage sont des structures géométriques basées sur le concept de proximité des éléments de données, et ils peuvent être utilisés pour déterminer les plus proches voisins de données multidimensionnelles. Le problème est que, outre la construction couteuse de ces graphes, leur mise à jour est aussi difficile en raison de l'insertion ou la suppression d'un élément. Dans cet article nous proposons d'adapter les graphes de voisinage pour l'indexation de plus grandes quantités de données multidimensionnelles. Nous proposons une modification d'une méthode existante de mise à jour locale de graphes de voisinage de telle sorte à ce qu'elle puisse considérer des accès réduits aux disques afin de prendre en compte l'évolution des bases de données. Cette proposition est étendue pour faire de la construction incrémentale de graphes de voisinage. Des évaluations de l'approche ont été effectuées et les résultats montrent que notre approche est prometteuse.

1 Introduction

Une des exigences pour traiter des données multidimensionnelles à grande échelle telles que des données multimédias, des données temporelles, des données spatiales, etc., est de fournir une solution efficace pour rechercher le voisinage. En plus de l'hétérogénéité des données, leur grande quantité et leur haute dimensionnalité rendent cette tâche de recherche assez difficile. Rechercher le voisinage efficacement est une question clé dans le traitement automatique de données multidimensionnelles. Divers algorithmes et techniques dans différents domaines de recherche ont été proposés. Par exemple, *k-means* (MacQueen (1967)), Cartes de Kohonen (Somervuo and Kohonen (1999)), et *k-NN* sont des solutions largement utilisées dans l'exploration de données ou dans la majorité des techniques d'indexation de bases de données (Gaede and Günther (1998)).