Rapport sur l'Algorithme de Recherche du Voisin le Plus Proche Approximatif (ANN) Basé sur le Graphe

Université d'Ottawa - Faculté de génie

CSI2110 - Algorithmes et Structures de Données

Automne 2023

Introduction

Dans le cadre de l'assignation de programmation de CSI2110, nous avons exploré l'implémentation et l'évaluation d'un algorithme de recherche du voisin le plus proche approximatif (ANN) en utilisant une méthode basée sur le parcours de graphes. L'objectif principal était de trouver des vecteurs dans un ensemble de données qui sont parmi les voisins les plus proches d'un point de requête, sans pour autant être le plus proche absolu.

Méthodologie

Le processus s'est articulé en deux étapes principales :

Construction du Graphe :

Utilisant la classe GraphA1NN, un graphe non dirigé a été construit à partir de l'ensemble de données. Chaque vecteur de données a été représenté comme un sommet dans le graphe.

Parcours du Graphe pour la Recherche ANN :

La recherche commence par sélectionner un sommet aléatoire, calculer la distance jusqu'au point de requête, et poursuivre le parcours en insérant les sommets dans un tableau trié selon leur distance par rapport au point de requête.

Expérimentation

Les expériences ont été menées sur un ensemble de données de 10 000 points. Les paramètres clés ajustés étaient k (le nombre de voisins dans le graphe k-NN) et S (la capacité du tableau pour le parcours de graphe). Les performances de l'algorithme ont été évaluées en termes de précision et de temps d'exécution.

Résultats

Les résultats ont révélé des tendances intéressantes :

Précision : La précision de l'algorithme, mesurée en pourcentage, variait avec les paramètres k et S. En moyenne, la précision augmentait avec des valeurs plus élevées de k. Cela suggère que plus le graphe est dense (plus de voisins par sommet), plus l'algorithme a de chances de trouver un voisin proche du point de requête.

Temps d'Exécution : Le temps d'exécution, mesuré en millisecondes, a également montré une corrélation avec les paramètres k et S. Des valeurs plus élevées de k entraînaient généralement une augmentation du temps d'exécution, probablement en raison de la complexité accrue du parcours de graphe.

Discussion

L'approche du parcours de graphe pour la recherche ANN a démontré son efficacité, en particulier dans des contextes où une solution approximative est acceptable. Cela est particulièrement pertinent dans des applications telles que la recherche d'images ou les recommandations musicales, où la rapidité de suggestion est plus cruciale que la précision absolue.

Cependant, il est important de noter un équilibre entre la précision et le temps d'exécution. Un graphe plus dense (avec un k plus élevé) améliore la précision mais augmente également le temps de calcul. Par conséquent, le choix de k et S doit être soigneusement considéré en fonction des exigences spécifiques de l'application.

Conclusion

L'algorithme de recherche ANN basé sur le graphe offre un compromis intéressant entre précision et rapidité, le rendant approprié pour des applications où les réponses rapides sont plus valorisées que la précision absolue. Cette étude a permis de mieux comprendre les nuances et les compromis impliqués dans la conception et l'optimisation de telles méthodes de recherche.

Gabriel Zohrob 300309391