Journal of Information Technology and Telecommunications
Received: Nov 24, 2015, Accepted: Dec 26, 2015, Published: Dec 29, 2015
J Inform Technol Telecomm, Volume 1, Issue 1
http://crescopublications.org/pdf/JITT/JITT-1-R005.pdf
Article Number: JITT-1-R005

Review Article Open Access

Fast Nearest Neighbour Search with Keywords

Venkateswarlu Ch

KS Raju Institute of Engineering and Technology, Moinabad, Rangareddy, Telangana, India

Citation: Venkateswarlu Ch (2015) Fast Nearest Neighbour Search with Keywords. J Inform Technol Telecomm 1: R005.

Abstract

Conventional spatial queries, such as range search and nearest neighbour **retrieval, involve only conditions on objects'** geometric properties. Today, many modern applications call for novel forms of queries that aim to find objects satisfying both a spatial predicate, and a predicate on their associated texts. For example, instead of considering all the restaurants, a nearest neighbour **query would instead ask for the restaurant that is the closest among those whose menus contain "steak, spaghetti, brandy" all at the same time. Currently, the best solution to such** queries is based on the IR2 -tree, which, as shown in this paper, has a few deficiencies that seriously impact its efficiency. Motivated by this, we develop a new access method called the spatial inverted index that extends the conventional inverted index to cope with multidimensional data, and comes with algorithms that can answer nearest neighbour queries with keywords in real time. As verified by experiments, the proposed techniques outperform the IR2 -tree in query response time significantly, often by a factor of orders of magnitude.

Existing System

- Spatial queries with keywords have not been extensively explored. In the past years, the community has sparked enthusiasm in studying keyword search in relational databases.
- It is until recently that attention was diverted to multidimensional data. The best method to date for nearest neighbour search with keywords is due to Felipe et al.. They nicely integrate two well-known concepts: R-tree, a popular spatial index, and signature file, an effective method for keyword-based document retrieval. By doing so they develop a structure called the IR2 -tree, which has the strengths of both R-trees and signature files.
- Like R-trees, the IR2 - **tree preserves objects' spatial proximity, which is the key to solving sp**atial queries efficiently. On the other hand, like signature files, the IR2 -tree is able to filter a considerable portion of the objects that do not contain all the query keywords, thus significantly reducing the number of objects to be examined.

Disadvantages of Existing System

- Fail to provide real time answers on difficult inputs.
- The real nearest neighbour lies quite far away from the query point, while all the closer neighbours are missing at least one of the query keywords.

Proposed System

- In this paper, we design a variant of inverted index that is optimized for multidimensional points, and is thus named the spatial inverted index (SI-index). This access method successfully incorporates point coordinates into a conventional inverted index with small extra space, owing to a delicate compact storage scheme.
- Meanwhile, an SI-index preserves the spatial locality of data points, and comes with an R-tree built on every inverted list at little space overhead. As a result, it offers two competing ways for query processing.
- We can (sequentially) merge multiple lists very much like merging traditional inverted lists by ids. Alternatively, we can also leverage the R-trees to browse the points of all relevant lists in ascending order of their distances to the query point. As demonstrated by experiments, the SI-index significantly outperforms the IR2 -tree in query efficiency, often by a factor of orders of magnitude.

Advantages of Proposed System

- ✓ Distance browsing is easy with R-trees. In fact, the best-first algorithm is exactly designed to output data points in ascending order of their distances
- ✓ It is straight forward to extend our compression scheme to any dimensional space

System Requirements

Hardware Requirements:

➢ System	:	Pentium IV 2.4 GHz.
➢ Hard Disk	:	40 GB.
➢ Floppy Drive	:	1.44 Mb.
➢ Monitor	:	15 VGA Colour.
➢ Mouse	:	Logitech.
➢ Ram	:	512 Mb.

Software Requirements:

➢ Operating system	:	Windows XP/7.
➢ Coding Language	:	JAVA/J2EE
➢ IDE	:	Netbeans 7.4
➢ Database	:	MYSQL

Reference

Yufei Tao and Cheng Sheng **"Fast Nearest Neighbor Search with Keywords"** IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, APRIL 2014.