IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

EFFICIENT GROUP QUERY OPTIMIZATIONOF SPATIAL ASSIMILATION USING RANKING CONCEPTION

1st S.Sangeetha, 2nd Ariya Varghese, 3rd Shruthi

Assistant Professor, Department of Computer Science and Engineering, Al-AmeenEngineering College, Erode, Tamil Nadu

ABSTRACT: Interactive ranking techniques have substantially promoted analysts ability inmaking judicious and informed decisions effectively based on multiple criteria. However, the existing techniques cannot satisfactorily support the analysis tasksinvolved in ranking large- scale spatial alternatives, such as selecting optimal locations for chain stores, where the complex spatial contexts involved are essential to the decision-making process. Limitations observed in the prior attempts of integrating rankings with spatial contexts motivate us to develop a context integrated visual ranking technique. Based on a set of generic design requirements I summarized by collaborating with domain experts. I propose SRV is, a novel spatial ranking visualization technique that supports efficient spatial multi-criteria decision- making processes by addressing three major challenges in the aforementioned context integration, namely, a) the presentation of spatial rankings and contexts, b) the scalability of rankings' visual representations, and c) the analysis of context-integrated spatial rankings. Specifically, I encode massive rankings and their cause with scalable matrix- based visualizations and stacked bar charts based on a novel two-phase optimization framework thatminimizes the information loss, and the flexible spatial filtering and intuitive comparative analysis are adopted to enable the in-depth evaluation of the rankings and assist users in selecting the best spatial alternative. The effectiveness of the proposed technique has been evaluated and demonstrated with an empirical study of optimization methods, two case studies, and expert interviews.

1. INTRODUCTION

Nearest neighbor search is one of the research in Information Technology where several algorithms and theoretical performance bounds have been devised for exact and approximate processing in main memory. The nearest query discoversthe previously unknown, but potentially useful patterns from spatial databases, trying to find patterns in geographic data. Due to the popularity of search services on the Internet, users are allowed to provide a list of keywords besides the spatial Information of objects, which reduces Scalability and an increase of Query response time. Therefore there is a need for improved training methods, and virtual reality technology for processing this query, which is implemented by means of group nearest group (GNG) search.

ABOUT SPATIAL MINING

Spatial data mining is the process of discovering interesting and previously un-known, but potentially useful patterns from large spatial datasets. Extracting interesting and useful patterns from spatial datasets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation. Spatial data is about instances located in a physical space. When spatial information becomes dominant interest, spatial data mining should be applied. Spatial data structurescan facilitate spatial mining. Standarddata mining algorithms can be modified for spatial data mining, with a substantial part of preprocessing to take into account of spatial information

ABOUT NEAREST KEYWORD SEARCH

Nearest neighbor search (NNS),also known as proximity search, similarity search or closest point search, is an optimization problem for finding closestpoints in metric spaces. The problem is: given a set S of points ina metric space M and a query point q ∈ M, find the closest point in S to q. In metric space, there is a valid concept of distance between points. If I treat two features as points in space, I can determine the distance between them. The lesser the distance between them, the more similarthey are in appearance.

OBJECTIVE

The objective of Efficient Group Query Optimization is to provide the optimal solution for the nearest group point in the dataset. In nearest neighbor queries, an optimization problem is evaluated forfinding the closest points in metric spaces. Given a data point set D, a query point set Q and an integer k, the Group Nearest Group query finds a subset of points from D, ω (|ω| ≤ k), such that thetotal distance from all points in Q to thenearest point in ω is no greater than anyother subset of points in D.

The processing focus of our approachesis on minimizing the access and evaluation of subsets of cardinality k in D since the number of such subsets is exponentially greater than the dataset. To do that, the hierarchical blocks of data points at high level are used to find an intermediate solution and then refined by following the guided search direction at low level so as to prune irrelevant subsets. Exhaustive Hierarchical Combination algorithm and Subset Hierarchical Refinement canprovide an exact solution to find nearestgroup points. Thus, the focus of this experiment is to examine the efficiencythese algorithms and to provide the optimal solution.

LITERATURE REVIEW

RANGE NEAREST-NEIGHBOR QUERY

A range nearest-neighbor (RNN) query retrieves the nearest neighbor (NN) for every point in a range. I consider the ranges as (hyper) rectangles andpropose efficient in-memory processing and secondary memory pruning techniques for RNN queries in both 2D and high-dimensional spaces[2]. These techniques are generalized for kRNN queries, which return the k nearest neighbors for every point in the range[5].

In general, processing an NN query on a spatial index involves two interleaving phases:

- secondary memory pruning of distant index nodes &
- In-memory computation of the nearest neighbors

LOCATION-BASED INSTANT SEARCH

Location-based instant search that combines location based keyword search with instant search is formulated[11]. Initially the filtering-effective hybrid index (FEH) is evaluated. Then development of indexing and search techniques are utilized for the FEH index and store prefix information to efficiently answer instant queries[11].

I first present an index structure called “filtering-effective hybrid” (FEH) index. It judiciously uses two types of keywordfilters in a node of a spatial tree based on the selectiveness of each keyword[11]. One filter, called child filter, maps keywords and their corresponding children nodes. Another Filter, called “object filter”, maps keywords to their corresponding records in the sub tree of the node. During a traversal of the FEHindex tree, the object filter at each nodeallows us to directly retrieve records for these keywords in the filter, thus bypassing those intermediate nodes in the sub tree. Next is to find answers to aquery as the user is typing the keywords character by character. Existing index techniques are utilized and queries are answered using FEH.

HYBRID INDEX STRUCTURES FOR LOCATION BASED WEB SEARCH

Location-based instant search thatcombines location based keywordsearch with instant search is formulated[1]. Nearest neighbor (NN) queries on a spatial database is a classical problem. The k-NN algorithm for R-trees traverses an R-tree while maintaining a list of k potential nearest neighbors in a priority queue in a Depth-First (DF) manner[2]. The DFalgorithm is sub-optimal, i.e., it accesses more nodes than necessary. The Best-First (BF) algorithm achieves the optimal I/O performance by maintaining a heap with the entries visited so far, sorted by their mindist. DF can be more I/O consuming than BF. However, DF requires only bounded memory and at most a single tree path resides in memory during search[14].

The closest pair queries (CPQ) are a combination of spatial join and nearest neighbor queries, which find the pair with the minimum distance among all pairs from two data sets[15]. The difference between nearest neighbor queries and closest pair queries is that the algorithms of the latter access two index structures (one for each data set) and utilize the distance function of the two intermediate nodes to prune the pairs.NNK specifies only one query location specifies a set of query locations.

EXISTING SYSTEM

Most traditional spatial queries on spatial databases such as nearest neighbor queries, range queries use CLARNS (Clustering Large Applications based upon Randomized search) of GNG leads to gap of few percentage points missed. The existing system, takes long query processing time anddata accuracy problems were identified.

In nearest neighbor queries, an optimization problem is evaluated for finding the closest points in metric spaces .Given a set S ofpoints in a metric space M and a query point q ∈ M, finding the closest point in S to q. The informal observation usually referred to as the curse of dimensionality states that there is no general-purpose exact solution forNNS in high-dimensional Euclidean spaceusing polynomial preprocessing and poly logarithmic search time. The current system is unable to view the location of the spot in spatial data when new site is added. In practice, using local search heuristics for GNG query leads to a gap of a few percentage points between the obtainedsolution and the global optimum. In the worst case, the local search heuristics have been proved to achieve at most five times ofthe global optimum. The existing system reduced the cluster quality.

PROPOSED SYSTEM

The proposed system uses Exhaustive RRNull, RRClosed, SRV methods and Hierarchical Combination (EHC) algorithm and Subset Hierarchical Refinement (SHR) algorithm. Use hierarchical blocks instead of data points to optimize the number of subsets evaluated. This technique aims atminimizing the I/O accesses to the object and feature data sets.

ADVANTAGES OF PROPOSEDSYSTEM

Optimized version provides more efficient technique for computing the scores of the objects. It develops solutions for the top-k spatial preference query based on the temporal data. It Minimize access and Reduce Search Space. In this work, database techniques are explored to boost the GNG query processing of local search heuristics without any loss on clustering quality.Torefine the solution, the search space in lowerhierarchical level is minimized. In EHC, every set of k blocks is evaluated in high hierarchical level and the set with the currentbest value (i.e., the minimum total distance) are refined by visiting their children in next level.EHC is capable to provide the optimal solution

MODULE DESCRIPTION

The project is divided into four modules:

- Nearest Group
- Nearest Keyword Set Search
- Projection and Multi scale hashing
- Group Ordering

NEAREST GROUP

In spatial databases most of the work has focused on the point NN query that retrieves the k (≥1) objects from a dataset P that are closest (usually according to Euclidean distance) to a query point q. The real data setof points are collected which consists of the place with the longitude and latitude of the earth location. The synthetic data points were obtained containing the uniformly distributedpoints around the city. These data sets are unified into a unit region.

NEAREST KEWORD SET SEARCH

The data is represented by data blocks, e.g., using R-tree. The algorithm process Group Nearest query by treating the blocks as pointsto find an intermediate solution in higherhierarchical level first. To refine the solution,the search space in lower hierarchical level is minimized by following the guided search direction.

PROJECTION AND MULTI SCALE HASHING

ProMiSH –(Projection and Multiscale Hashing) that always retrieves the optimal top-k results, and an approximate ProMiSH is more efficient in terms of time and space, and is able to obtain near-optimal result. It isa local search heuristic with support of the database techniques. In higher hierarchical level, each block is treated as a point by ProMiSH to replace every element in the subset, and the resultant subset with the current best value is refined by visiting the children of the block.

GROUP ORDERING

A suitable ordering of the groups leads to an efficient candidate exploration by a multi- way distance join. First perform a pair wise inner joins of the groups with distance threshold rk. In inner join, a pair of points from two groups is joined only if the distance between them is at most rk.

SYSTEM TESTING

TESTING PROCESS

Testing is a process of checking whether the developed system is working according to the original objectives and requirements. Testing is a set of activities that can be planned in advance and conductedsystematically. Testing is vital to the success of the system. System testing makes a logical assumption that if all the parts of the system are correct, the global will be successfully achieved. In adequate testing if not testing leads to errors that may not appear even many months. This creates two problems,The time lag between the cause and the appearance of the problem . The effect of thesystem errors on the files and records within the system A small system error can conceivably explode into a much larger problem. Effective testing early in the purpose translates directly into long term cost savings from a reduced number of errors. Another reason for system testing is its utility, as a user- oriented vehicle before implementation. The best programs are worthless if it procedures the correct outputs. No other test can be more crucial. Followingthis step a variety of tests are conducted.

- Black box testing
- White box testing

BLACK BOX TESTING

Black box testing, also called behavioral testing focuses on the functional requirements of the software. That is, black box testing enables the software engineer to derive sets of input conditions that will fully exercise all functional requirements for a program. Black box testing is not alternative to white box techniques. Rather it is a complementary approach that is likely to uncover a different class of errors than whitebox methods. Black box testing attempts to find errors in the following categories.

White Box Testing:

White Box Testing is a testing in whichin which the software tester has knowledge of the inner workings, structureand language of the software, or at least itspurpose.

UNIT TESTING

A program represents the logical elements of a system. For a program to run satisfactorily, it must compile and test data correctly and tie in properly with other programs. Achieving an error free program is the responsibility of the programmer. Program testing check for two types of errors: Syntax and logical. Syntax error is a program statement that violates one or more rules of the language in which is written. An improperly defined field dimension or omitted keywords are common syntax errors. These errors are shown through error message generated by the computer. For logic errors the programmer must examine the output carefully.

CONCLUSION

We have introduced a new type of query called as Group Nearest Group Query with Ranking which retrieves number of objects from Query keyword Q with minimum sum of distances to its nearest Data points. We use exhaustive hierarchical combination andsubset hierarchical refinement algorithm, which prunes the query objects and finally the minimized summed distance is calculated.

The number of node accesses is also reduced which reduces the query response time, which exhibits good scalability with the query objects and the number of query keywords.

REFERENCES

[1] S. Al-Dohuki, Y. Wu, F. Kamw, J. Yang, X. Li, Y. Zhao, X. Ye, W. Chen, C. Ma, and F. Wang. Semantictraj: A new approach to interacting with massive taxi trajectories. IEEE TVCG, 23(1):11–20, 2017.

[2] K. Deng, X. Zhou, and H. Shen, “Multi- Source Skyline QueryProcessing in Road Networks,” Proc. 23th IEEE Int’l Conf. DataEng., 2007.

[3] M. Sharifzadeh and C. Shahabi, “The Spatial Skyline Queries,”Proc. 32nd Very Large Data Bases Conf., 2006.

[4] D. Papadias, Q. Shen, Y. Tao, and K. Mouratids, “Group NearestNeighbor Queries,” Proc. 20th IEEE Int’l Conf. Data Eng., 2004.

[5] D. Papadias, Y. Tao, K. Mourstidis, and C.K.Hui, “AggregateNearest Neighbor Queries in Spatial Databases,” ACM Trans.Database Systems, vol. 30, no. 2, pp. 529-576, 2005.

[6] P. Hansen, Systems of Cities and Facility Location. HarwoodAcademic Publishers GmbH, 1987.

[7] M. Garey and D. Johnson, Computers and Intractability: A Guide tothe Theory of NP-Completeness. Freeman and Company, 1979.

[8] Local Search in Combinatorial Optimization, E. Aarts andJ.K. Lenstra, eds. Princeton Univ. Press, 2003.

[9] K.E. Rosing, “An Empirical Investigation of

the Effectiveness of aVertex Substitution Heuristic," Environment and PlanningB,vol. 24, pp. 59-67, 1997.

[10] R. Whitaker, "A Fast Algorithm for the Greedy Interchange ofLarge-Scale Clustering and Median Location Problems," INFOR,vol. 21, pp. 95-108, 1983.

[11] V. Arya, N. Gary, R. Khandekar, A. Mayerson, K. Munagala, andV. Pandit, "Local Search Heuristics for k-Median and FacilityLocation Problems," Proc. 33rd ACM Symp. Theory of Computing,2001.

[12] L. Kaufman and P. Rousseeuw, Finding Groups in Data: AnIntroduction to Cluster Analysis. John Wiley & Sons, 1990.

[13] R. Ng and J. Han, "Efficient and Effective Clustering Method forSpatial Data Mining," Proc. 20th Very Large Data Bases Conf., 1994.

[14] K. Mouratidis, D. Papadias, and S. Papadimitriou, "Tree-BasedPartition Querying: A Methodology for Computing Medoids inLarge Spatial Datasets," The VLDB J., vol. 17, no. 4, pp. 923-945,2008.

[15] K. Deng, H. Xu, S. Sadiq, Y. Lu, G. Fung, and H. Shen, "ProcessingGroup Nearest Group Query," Proc. 25th IEEE Int'l Conf. DataEng., 2009.

[16] G. Hjaltason and H. Samet, "Distance Browsing in SpatialDatabases,"ACM Trans. Database Systems, vol. 24, no.2, pp. 265-318, 1999.

[17] C. Bohm, S. Berchtold, and D.Keim, "Searching in HighDimensional Spaces: Index Structures for Improving the Performanceof Multimedia Databases," ACM Computing Surveys,vol. 33, no. 3, pp. 322-373, 2001.

[18] K. Cheung and A.W.C. Fu, "Enhanced Nearest Neighbor Searchon the R-Tree," ACM SIGMOD Record, vol. 27, no. 3, pp. 16-21,1998