International Journal of Scientific Research in Computer Science and Engineering
Review Paper
Vol.4, Issue.3, pp.30-33, June (2016)
E-ISSN:2320-7639

An Optimal Route Search Using Spatial Keyword Query using Keyword Nearest Neighbor Expansion

H.K.N. Peddinti[1*], J.A. Lavanya[2], G.K. Chakravarthi[3]

[1]Dept. of CSE, Sanketika Vidya Parishad Engineering College, Visakhapatnam, India
[2]Dept. of CSE, Sanketika Vidya Parishad Engineering College, Visakhapatnam, India
[3]Dept. of CSE, Sanketika Vidya Parishad Engineering College, Visakhapatnam, India

Corresponding Author: harishpeddinti999@gmail.com

Available online at www.isroset.org

Received: 02 May 2016, Revised: 10 May 2016, Accepted: 17 Jun 2016, Published: 30 Jun 2016

Abstract— In our daily life the need to find optimal routes between the two points is critical, for instance finding the shortest distance to the nearest hospital. The internet based maps are now generally used for this purpose. Route search and optimal route queries are two significant classes of queries based on the road network concept. Route search queries find the route based on the given constraints. The optimal route queries find the optimum route from a set of specifications by a user. In the road map queries, users would have to give the specification of the starting point and the ending point of their travelling with or without the constraints. A few spatial features about the categories and the different locations should be specified along with this. If the travelling constraints are given then it should be unique. These constraints may be either a total order or a partial order. In this specification order there should be information about both the starting point and the destination point of the travelling. The optimal route queries optimize the possible routes and give the optimal route that meets all the constraints. This paper describes the survey on optimal route query processing, two categories namely optimal route query processing and spatial search with categorical information have been considered, an analysis on technique for optimal route query with constraints and without constraint is also included. The total order needs a specification of list of points and in the same order that they should be visited but that is not required for partial order constraints. Lastly this paper concludes with pros and cons of different techniques under optimal route queries.

Keywords- Spatial keyword query , spatial objects, spatial database, best keyword cover query.

I. INTRODUCTION

Data mining is the means of extracting data from a dataset for users to use it in various purpose. The purpose of such data plays an important role in keyword searching. Searching is a common activity happening in data mining. Searching for spatial objects from spatial database has recently sparked enthusiasm among researchers. This motivated to develop methods to retrieve spatial objects. Spatial objects consists of objects associated with spatial features. In other words, spatial objects involve spatial data along with longitude and latitude of location. Querying such data is called best keyword cover querying. Search is called best keyword cover search.

Existing method to such data consider either minimum inter objective distance and keyword search. As a result new methods for best keyword cover search was developed. Traditional nearest neighbour search compute nearest neighbour by considering distance as feature. In this context, nearest neighbour search focus on finding nearest neighbours where keywords and spatial data plays a major impact. It comes with algorithms to answer such query. (Size 10 & Normal) This document is a template. n electronic copy can be downloaded from the conference website. For questions on paper guidelines, please contact the conference publications committee as indicated on the conference website. Information about final paper submission is available from the conference website.

Query processing is the efficient retrieval of desired information from the database system. The various phases of query processing system are shown in the Fig 1. Four main steps are there in the query processing where first three steps are executed in compile time and last one in the runtime. Stering techniques are also been discussed. Most of the techniques have limitations in some particular area.

The main problems are some of the techniques may yield optimal route and others may not. Among this some may consider constraints and others without considering the constraints and these are the main problems that need to consider.

II. OPTIMAL ROUTE QUERY PROCESSING IN ROAD NETWORK

A. On Trip Planning Queries (TPQ) in Spatial Database

On Trip Planning Queries are well organized and have exact solutions for the general optimal route queries. A set of point of interest (POI) of various categories, starting point and destination point is given and TPQ finds the best trip starting from the particular source and will ends on the destination through some POI [1]. There are no ordered constraints here in this method. The actuality of multiple choices per category is the main difficulty of this technique and for solving this some of the estimation techniques are used. Mainly two greedy algorithms are available with the tight estimation ratios with respect to the total number of classifications. The first algorithm is the nearest neighbour algorithm. This algorithm find the best trip by visiting the nearest neighbour of the last category to be added and that have not been visited at that particular moment [2]. The route thus formed from source to destination point which is stated by the user. The second one is the minimum distance algorithm which introduces a novel greedy algorithm and while comparing with the first one this is having the better estimation bound [3]. This will find the set of vertices with minimum cost. In this paper the nearest neighbour algorithm is used to find the better route starting and ending at specific points. The advantages of the On Trip Planning Queries include 1. Approximation ratio is high, 2. Minimum cost for route finding. The disadvantage includes 1. No user defined constraints in the TPQ [4].

B. The Optimal Sequenced Route (OSR Query)

The Optimal Sequenced Route Query is a kind of Nearest Neighbor query and it finds the optimal route that starts from a specific location and passed through a number of typed location in some specific order. The shortest path issue is the basics of this technique. First of all the OSR problem is transformed into a simple shortest path problem in huge planar graph. For that the Dijkstra's algorithm is used [5]. The OSR query is given with starting point, a set of intermediary points and the sequence of the locations. The weighted directed graph is constructed from the given network. The starting point was connected to all the other vertices and the weight assigned to each edge of graph is the distance between two end vertices. Hence the optimal route of the OSR query is the route or set of points with minimum length. The shortest path finding is by travelling from starting point to all other points and returning the minimum path length and this will be done by the Dijkstra's algorithm. But this classic Dijkstra's algorithm is impractical because of the following reasons [6]. The first one is, in the real world dataset millions of possible edges have to be handled so that the time complexity is very high. Thus the complexity of the algorithm is also very huge. To improve the issue occurred due to the Dijkstra's algorithm the range query can be used. Even then the problem cannot be overcome. Therefore Light Optimal Route Discover (LORD) algorithm developed for handling OSR queries. This is an iterative and a light threshold based algorithm, which uses different thresholds to filter out the points that cannot be the optimal route. The memory requirement for LORD is very much less than Dijkstra's algorithm hence it is named as light. First this algorithm generates a set of partial sequenced routes in the opposite order. That is from the destination point to the starting point [8]. Attach each point to the recently added one and lastly forms the optimal route. The next one is R-LORD algorithm which is an enhanced version of LORD algorithm. It is based on the R-tree and it calculated the threshold value more efficiently. This is the first correct solution for the optimal route queries with total order travelling rules. R-LORD uses the greedy algorithm to find the optimal route and the threshold value. From the end class, finds the optimal points in the sequence and within the threshold distance to the query point. Iteratively finds the optimum route.

C. The Multi-Rule Partial Sequence Route Query

The optimal path computation is purely based on greedy [1] solutions in the premature stages. This uses two approaches. The first one is Nearest Neighbor Partial Sequenced Route which works similar to greedy approach. This will start from the query start point and find the point which is nearest to the start point and forms the edge. Then using the newly added point finds the nearest point and forms the route. The final route will be optimal through this greedy approach. That is it finds the most suitable at that moment. The second approach will find all the nearest points from each and every class and forms the route. This is also optimal route with in a specific range. Then another approach is the combination of NNPSR and RLORD [7]. The NNPSR is used to obtain the greedy route. From that it obtains the category of each and every point. This total order of classification is the input of R-LORD. And it will output the suboptimal routes.

III. SPATIAL SEARCH WITH CATEGORICAL INFORMATION

A. Top-K Spatial Preference Queries

The spatial preference query is used to rank the objects from the characteristic qualities of the neighborhood objects. Mainly Top-K spatial preference queries are used to rank the spatial objects effectively. Distinct types of algorithms are used for this ranking. First one is a baseline algorithm named simple probing algorithm and which applying the spatial queries on feature dataset and calculate the scores for every object.

Thus the ranking is done. The incremental computation techniques are used to advance the simple probing algorithm and which minimizes the number of computations of component scores. A variation of simple probing algorithm is the group probing algorithm and which computing the object scores in the specific leaf node frequently and thus reduces the I/O cost. Branch and Bound algorithm is the advanced version of group probing algorithm, which eliminates the entries other than the leaf nodes in the object tree that cannot yield better performance [8].

For this a new method is used by obtaining feature trees and deriving upper bound scores for the entities. The final one is feature join algorithm which performs over the feature tree by multi-way join used to find the group of feature points and then search for the objects using this classification.

B. Categorical Range Queries (CRQ) in Large Databases

The categorical range queries in the large database are handled through the paper. Mainly it is related to the spatial data like geostationary information systems. For this technique a multi tree indent is used and which is associated with an effective way of categorical data and spatial information. The main concept is the augmentation of categorical points with some information to accelerate the queries. The technical parts are a new method for spatial data structure and R-Tree based on the query processing of CRQs in the context of large databases. The new method is compared with two baseline algorithms. The first one is the regular range query which handles the query for the particular range of lower and upper boundary. The second one is the construction of R-Tree which contains the nodes that are the categorical attributes.

C. Query Processing in Spatial Network Databases

The query processing in spatial database is mainly based on Euclidean spaces. But in this paper a new architecture is proposed in which the road network is separated from the datasets. To gather connectivity and location, a disk-based representation is used. For handling the dynamic updates and Euclidean queries the spatial entities are scored by some of the corresponding spatial access methods. Based on the above architecture two techniques are developed which are Euclidean restriction and network expansion. The most common spatial queries are the range search queries, nearest neighbor queries, closest pair queries and distance join queries. These are processed by using the above mentioned frameworks. By using the location information and connectivity, the efficient pruning of the search space is possible. Through this the traditional processing methods can be expanded by the new algorithms. The query processing in the Spatial Network Databases (SNDB) in efficient and this paper introduces this advantage.

D. Optimal Route Query with Arbitrary Order Constraints

The optimal route query considers the partial order constraints for finding the optimal route. The user wants to specify starting point, destination point and a set of arbitrary order constraints. This is different from the total sequenced rules. The instance for the partial order constraints is “visit the pub before hotel”. Therefore any other categories can be included in between these two categories. But total order is the sequenced route conditions. For considering the partial order constraints two different types of techniques are developed namely backward search and forward search. Both the methods will take the same inputs and will produce the same output. The backward search algorithm finds the optimal route in the reverse manner that is starting from the destination and ends at the query starting point. This is similar to the R-LORD algorithm. First select the destination point and as per the partial sequenced route find out all the possible edges to the second last point. Then find the optimum edge from these and attached to the destination. Then repeat the process until the query starting point encountered. The second one is Forward search algorithm and which is similar to the greedy algorithm. First select the query starting point and then find the nearest neighbor point which satisfies the given constraint. The process repeats and finally one optimal route is obtained. Then this forward search algorithm will use the backward search algorithm for the backtracking process. This will eliminate the demerits of the greedy algorithm. That is it eliminates some points that will not be a part of the optimal route. Both the algorithms find the optimal routes and which satisfies all the given partial order constraints. The memory usage is reduced by using some pruning techniques. Thus the number of categories from the dataset is reduced and the memory usage will be reduced. This paper which uses both the optimal route query processing in road network and the spatial search with categorical information. Thus the methods used here is included in both the classifications. This technique solved the problem of optimal route query with partial order constraints. Another advantage is that several sub routes also can be obtained and are optimal. Therefore in the real world application some of the category points can be omitted to meet the particular cost or the time.

IV. CONCLUSION

The optimal route queries find the optimal route and this has greater applications in the road network. The spatial search with categorical information is used to consider the categorical points to be visited with better facilities. The initial solution of the optimal route query is based on the greedy solution. Some of the techniques considered total order constraints. The recent solutions of the optimal route query handle the arbitrary order constraints. All the methods will result in the optimal solutions with the given the starting and destination points. But many of the techniques do not consider constraints. Some may consider the total order constraints and some others use the partial order constraints. In future, some of the timing constraints can be included to the optimal route queries.

REFERENCES

[1]. C. Zeng, JF. Naughton, JY. Cai, “*On differentially private frequent itemset mining*”, VLDB Endowment, Vol.6, Issue.1, pp.25-36, 2012.

[2]. H. Chen, WS. Ku, MT. Sun, R. Zimmermann, “The partial sequenced route query with traveling rules in road networks”, GeoInformatica, Vol.15, Issue.3, pp.541-69, 2011.

[3]. J. Sander, M. Ester, HP. Kriegel, X. Xu, "*Density-based clustering in spatial databases: The algorithm gdbscan and its applications*", Data mining and knowledge discovery, Vol.2, Issue.2, pp.169-94, 1998.

[4]. J. Li, YD. Yang, N. Mamoulis, *"Optimal Route Queries with Arbitrary Order Constraints*", IEEE Trans. Computers, vol. 25, no. 5, pp. 1097- 1110, 2013.

[5]. A. Nanopoulos A, P. Bozanis, "*Categorical Range Queries in Large Databases*", Springer-Verlag, Berlin-Heidelberg, pp. 122-139, 2003.

[6]. M. Kolahdouzan, C. Shahabi, "*Voronoi-based k nearest neighbor search for spatial network databases*", International conference on Very large data bases, Vol.30, pp.840-851, 2004.

[7]. M. Sharifzadeh, MR. Kolahdouzan, C. Shahabi, *"The Optimal Sequenced Route Query",* International Journal of Very Large Data Bases, Vol.17, No.4, pp.765-787, 2008.

[8]. JB. Rocha-Junior, A. Vlachou, C. Doulkeridis, K. Nørvåg, *"Efficient processing of top-k spatial preference queries",* VLDB Endowment, Vol.4, Issue.2, pp.93-104, 2010.