IJCAT International Journal of Computing and Technology, Volume 1, Issue 4, May 2014
ISSN : 2348 - 6090
www.IJCAT.org

A Review on String Queries on Road Networks

[1] Sherlin Susan George, [2] Remya R.

[1] Department Of Computer Science and Engineering, Cusat, College Of Engineering Perumon.
Kollam, Kerala, India

[2] Department Of Information Technology ,Cusat, College Of Engineering Perumon
Kollam, Kerala, India

Abstract - Query processing on road networks has many applications. For solving the spatial range queries the baseline solution is based on the Dijkstra's algorithm. The earlier works provide efficient methods for exact keyword search only. But the RSASSOL algorithm returns the best objects that approximately match both the spatial predicate and string predicate. The OPRN algorithm retrieves points that have shortest distance to the query point and have textual similarity to the query keywords. Several string similarity functions are used here. The unnecessary points are pruned away in each stage. The results of extensive experiments show that this algorithm yields a flexible framework for the efficient processing of range queries.

Keywords - RSAS Queries, Dijkstra's algorithm , RSASSOL algorithm, String Similarity, Spatial Database, Range Queries, OPRN algorithm, Road networks, Approximate String Search

1. Introduction

In recent years, data mining has attracted a great deal of attention in the information industry. It is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. Data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses or other information repositories.

Today, an increasing number of usages of data sets have become available. Data mining is the practice of automatically searching large stores of data to discover patterns and trends that go beyond simple analysis. Data mining is also known as Knowledge Discovery in Data. Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. [6]. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Data mining is accomplished by building models. A model performs some actions on data based on some algorithm. The notion of automatic discovery refers to the execution of data mining models.

Approximate string matching is the technique of finding strings that match a pattern approximately rather than exactly [3]. The closeness of a match is measured in terms of the number of primitive operations necessary to convert the string into an exact match.This number is called the edit distance between the string and the pattern..Query processing on road networks has many applications such as online map services and mobile services .In a road network, each node represents a location , the edge between two nodes represents the path between them. Each node can be tagged with textual information such as school or hospital.

In this work, we focus on range queries in road networks known as spatial approximate string (SAS) queries in road networks (RSAS queries) [7].Spatial range queries inquire about certain spatial objects related to other spatial objects within a certain distance .Given a query ,this algorithm on road network returns the best objects with shortest path to the query location and textual relevance to the query keywords. For example given a query point q and a network distance r on a road network, we want to retrieve all objects within distance r to q and with the description similar to some keywords.

A key issue in SAS queries is to define the similarity between two strings .A straight forward solution to any existing techniques for answering the spatial component of an SAS query and verify the approximate string match predicate either in postprocessing or on the intermediate results of the spatial search. We refer to them as the spatial solution [5]. An approiximate string search is needed when spelling errors occurs while submitting the query.

For RSAS queries, the baseline spatial solution is based on the Dijikstras algorithm. Its performance degrades quickly when the query range enlarges and /or the data on the network increases. This motivates us to find a novel method to avoid the unnecessary road network expansions, by combining the pruning from the spatial and the string predicates simultaneously. This method partitions the road network, adaptively searches relevant sub graphs and prunes candidate points using both the string matching index and the spatial reference nodes. Lastly the distance formula and weight factors are used to verify the final set of candidates.

2. Literature Survey

Researchers are always being conducted to analyse the works associated with spatial approximate string queries. Some of the innovative approaches to road networks are:

2.1 Using Dijikstras Algorithm

Dijikstras algorithm is a graph search algorithm that solves the single-source shortest path problem for a graph with non-negative edge path costs, producing a shortest path tree. This algorithm is often used in routing and as a subroutine in other graph algorithms [2].For a given source vertex (node) in the graph, the algorithm finds the lowest cost (i.e. the shortest path) between that vertex and every other vertex. It can also be used for finding costs of shortest paths from a single vertex to a single destination vertex by stopping the algorithm once the shortest path to the destination vertex has been determined. For example, if the vertices of the graph represent cities and edge path costs represent driving distances between pairs of cities connected by a direct road, Dijkstra's algorithm can be used to find the shortest route between one city and all other cities. For RSAS queries, the baseline spatial solution is based on the Dijikstras algorithm.Its performance degrades quickly when the query range enlarges and/or the data on the network increases

2.2 Using RSASSOL Algorithm

The paper on spatial approximate string search [1] presents a comprehensive study for spatial approximate string queries in road networks. We use the edit distance, cosine similarity as the similarity measurement for the string predicate and focus on the range queries as the spatial predicate. .

Given a query, the RSASSOL algorithm on road network returns the best objects with shortest path to the query location and textual relevance to the query keyword. The RSASSOL method partitions the road network, adaptively searches relevant sub graphs and prunes candidate points using both the string matching index and the spatial reference nodes .Lastly the MPALT algorithm is used to verify the final set of candidates .This works returns only one facility which matches the string predicate. Future work includes finding several facilities together with least cost (shortest path).

Given a query point q and a network distance r on a road network, we want to retrieve all objects within distance r to q and with the description similar to "theatre," where the distance between two points is the length of their shortest path. The MPALT algorithm minimizes the access to the network by avoiding the nodes that will not be on any shortest path between s and any destination t_i. It also avoids repeatedly access to the explored part of the network when calculating multiple shortest paths to multiple destinations.

The basic idea works as follow: We start the expansion of the network from s with the two nodes from the edge containing s, and always expand the network from an explored node n (by adding adjacent nodes of n to a priority queue and checking points on corresponding edges) that has the shortest possible distance to any one of the destinations.

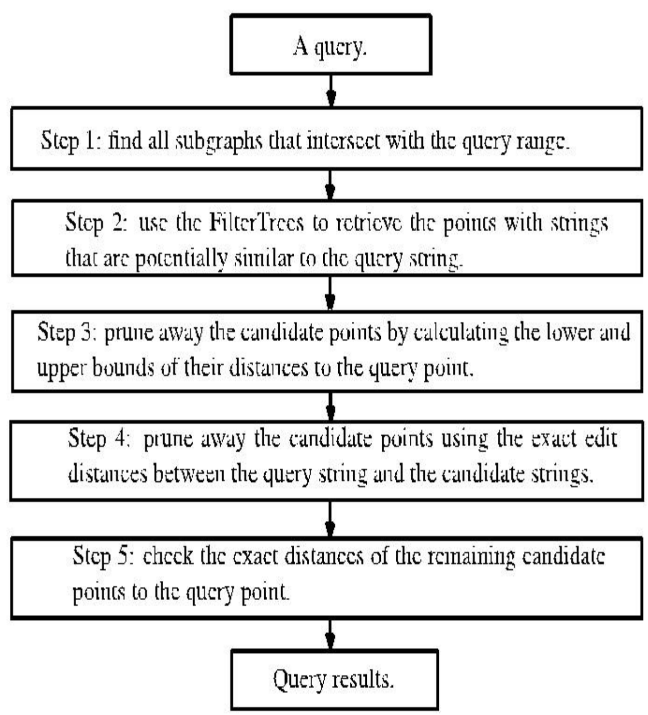

Fig 1 Overview of RSASSOL algorithm

We partition a road network G(V,E) into m edge-disjoint sub graphs $G_1; G_2; \ldots ; G_m$, where m is a user parameter, and build one string index (Filter Tree) for strings in each sub graph. We also select a small subset V_R of nodes from V as reference nodes: they are used to prune candidate points/nodes whose distances to the query point q are out of the query range r. conceptually, our RSAS query framework consists of five steps. Given a query, we first find all sub graphs that intersect with the query range.

Next, we use the Filter Trees of these sub graphs to retrieve the points whose strings are potentially similar to the query string. In the third step, we prune away some of these candidate points by calculating the lower and upper bounds of their distances to the query point, using V_R. The fourth step is to further prune away some candidate points using the exact edit distance between the query string and strings of remaining candidates. After this step, the string predicate has been fully explored. In the final step, for the remaining candidate points, we check their exact distances to the query point and return those with distances within r.

2.3 Using OPRN Algorithm

We model a road network as a graph G = (V,E) where V(E) denotes the set of nodes(edges) in G. We denote index nodes in G by unique ids and specify an edge by its two end nodes, placing the nodes with the smaller id first. A spatial approximate string query Q consists of two parts: the spatial predicate Qr and the string predicate Qs..

In road networks, Qr is specified by a query point q and a radius r and the string predicate Qs is defined by a set of strings and an edit distance threshold. The RSAS query framework consists of seven steps .Given a query, we first find all regions that intersect with the query range. Next, we use the similarity functions to retrieve the points whose strings are potentially similar to the query string. In the third step, we prune away some of these candidate points by calculating the lower and upper bounds of their distances to the query point. The fourth step is to further prune away some candidate points using the exact edit distance between the query strings and strings of remaining candidates. After this step, the string predicate has been fully explored. In the next step, for the remaining candidate points, we calculate distances to the query point and return those with shortest distances within r .Then we find the points with better cost values .In the last step display the optimal path using graph.

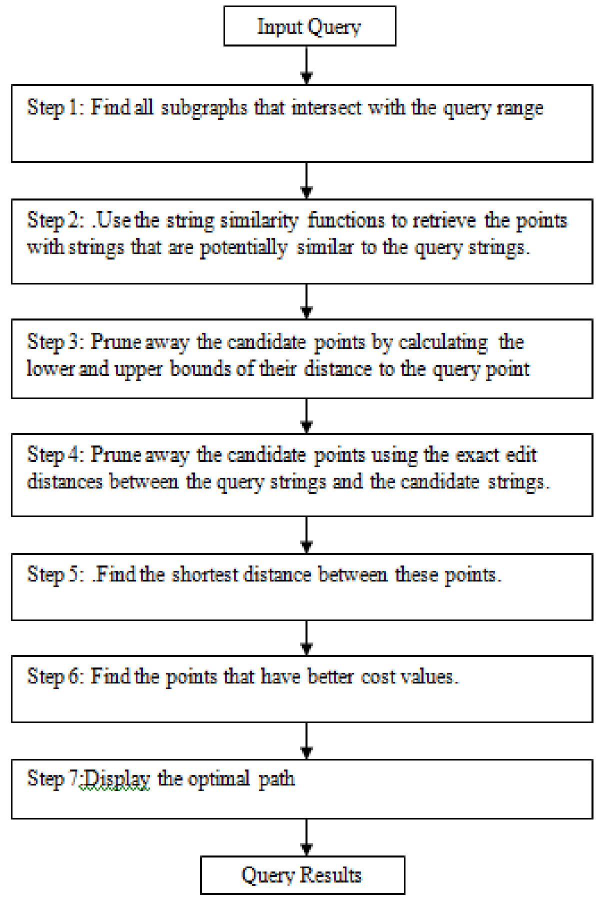

Fig 2.Overview of OPRN algorithm

3. Performance Analysis

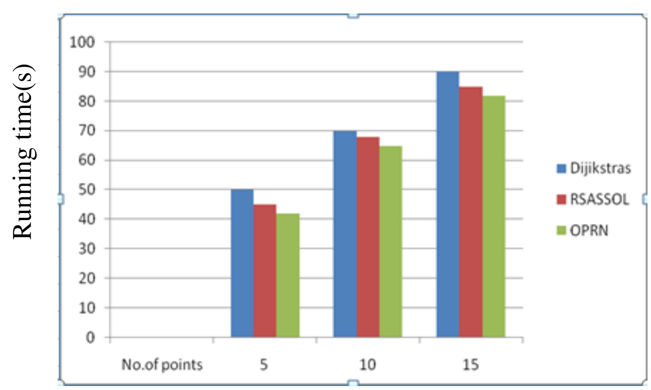

Fig3.Query performance for RSAS queries

RSAS queries, we studied the Dijikstras algorithm, RSASSOL algorithm and the OPRN method. We focus on using the spatial solutions and the string solutions. To test the RSAS queries we use real roadwork dataset.The road network in Germany.

We obtain a large number of real locations with text information in Germany from the open street map project and assign strings into the road network based on their coordinates .In the default dataset we randomly sample some points and assign them into the database. We study the effectiveness of these algorithms for the RSAS queries in this section. We investigate the effect of the number of points and the running time. The above figure shows the average running time when number of points varies from 5 to 15.Clearly OPRN outperforms RSASSOL and Dijikstras algorithm especially for more points.

The OPRN method has higher space consumption than the RSASSOL algorithm and the Dijikstras algorithm since it additionally stores the distances from the nodes to the reference nodes. The OPRN method displays the shortest path in a user-friendly manner.

4. Conclusions

This work presents a comprehensive study for spatial approximate string queries in road networks. We use the edit distance, cosine similarity as the similarity measurement for the string predicate and focus on the range queries as the spatial predicate. .Given a query, the Dijikstras, RSASSOL, OPRN algorithms on road network returns the best objects with shortest path to the query location and textual relevance to the query keyword.

Acknowledgments

We express my sincere gratitude to all the staff of Computer Science and Engineering Department in our college and beloved family members who helped me with their timely suggestions and support. I also express my sincere thanks to all my friends who helped me throughout the successful completion of the work. All glory and honor be to the Almighty God, who showered His abundant grace on me to make this work a success.

References

[1] Tang, and Mario Hadjieleftheriou "Spatial Approximate String Search" Proc. IEEE Int'l Conf. Knowledge and Data Engineering vol .5. No.6. June 2013.

[2] Joao B.Rocha-Junior and Kjetil Norvag,"Top-k-Spatial Keyword Queries on Road Networks" EDBT 26–30, Berlin, Germany 2012.

[3] Xin Caoy Gao Congy Christian S. Jensen "Retrieving Top k Prestige Based Relevant Spatial Web Objects", Proceedings of the VLDB Endowment, Vol. 3, No. 1,2010.

[4] D. Felipe, V.Hristidis, and N. Rishe. "Keyword Search on Spatial Databases" Proc.IEEE Int'l Conf. Data Eng.(ICDE), pp.656-665,2008.

[5] Oystein Egeland Carlsson "Keyword Search on Spatial Network Databases".This research was supported in part by NSF grants CNS-0320956, CNS-0220562, HRD-0317692, and IIS-0534530,2004.

[6] S.Alsubaiee, A. Behm, and C. Li, "Supporting Location-Based Approximate-Keyword Queries," Proc. SIGSPATIAL 18th Int'l Conf. Advances in Geographic Information Systems (GIS), pp. 61 -70, 2010.

[7] Ron Gutman "Reach-based Routing: A New Approach to Shortest Path Algorithms Optimized for Road Networks", proceedings of the sixth workshop on algorithm engineering and experiments,New York,2010

Sherlin Susan George Received the bachelor's degree in Computer Science and Engineering from Mahatma Gandhi University, Kerala in 2012. Presently she is pursuing her Mtech in the department of Computer Science and Engineering from Cochin University of Science and Technology, Kerala. Her research interests include Data Mining.

Remya R Received the bachelor's degree in Information Technology from University college Trivandrum, Kerala in 2004 and master's degree in Computer Science and Engineering from Anna University, Coimbatore in 2008.Currently working as an Assistant Professor in Information Technology department, College of Engineering Perumon, under Cochin university of Science and Technology. She has teaching experience of eight years. Research interests Includes Data Mining.