International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Survey of Approximate Shortest Distance Computing Using Improved Shortest Path Tree

Sonali Malode[1], Mansi Bhonsle[2]

[1]Department of Computer Engineering, GHRCEM Wagholi, Pune University, India

[2]Assistant Prof. Department of Computer Engineering, GHRCEM Wagholi, Pune University, India

Abstract: ***In large scale network, shortest distance query is a primary operation. In literature there isa number of existingsystems presented that take a landmark embedding approach. These systems select a set of graph nodes as landmarks. Then itcomputes the shortest distances from each landmark to all nodes as an embedding. All the existing methods follow the triangulation based distance estimationthatcalculates approximately the shortest distance among a pair of query nodes as the sum of their distances to a landmark. The landmark set gives a lone global view for all possible queries that could be diameter apart or close by because the landmark selection stage is query independent. Hence it is difficult to accomplishconsistently good performance on all queries as well as the landmark embedding approach may introduce a large relative error. Recently new method presented which is called as a query-dependent Local Landmark Scheme (LLS), which identifies a local landmark specific to a pair of query nodes. In these approach first needs to find query-dependent "local landmark" which is close to both query nodes for more accurate distance estimation. Then, the distance between the two query nodes is estimated as the sum of their shortest distances to the local landmark that is nearer than the any other. This method is only focusing on reducing the distance estimation error.***

Keywords: Shortest Distance Query, Landmark selection, Local Landmark Scheme

1. Introduction

These days the size of graphs that rise up out of different application domains is drastically expanding; the quantities of nodes achieve the scale of many millions or significantly more.Because of the enormous size, even straightforward graph queries becomingdifficult tasks. Over last four decades, the shortest distance query has been widely studied over.

Querying shortest paths or shortest distances between nodes in a huge graph has paramount applications in numerous areas including street systems, interpersonal organizations, communication networks, the World Wide Web, etc.Case in point, in street systems, the objective is to discover shortest path between source and destination; in interpersonal organizations, the objective is to discover the closest social relations; while in the World Wide Web, the objective is to discover the closest server to diminish access latency for customers.

Albeit traditional algorithms like breadth-first search (BFS), Dijkstra's algorithm [1], and A* search [2], [3], [4] can calculate the precise shortest paths in a network, the huge size of current information systems and the online nature of such queries make it infeasible to apply the traditional algorithms on the web.

Then again, it is space wasteful to pre-compute and store the shortest paths between all sets of nodes, as it obliges O(n3) space to store the shortest paths and O(n2) space to store the distances for a graph with n nodes. As of late, there have been numerous distinctive systems [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15] for assessing the shortest distance between nodes focused around graph embeddings. An ordinarily utilized embedding strategy is landmark embedding, where a set of graph nodes is chosen as landmarks and the shortest distances from a landmark to the various hubs in a graph are pre-computed. These distances can be utilized online to give an estimated distance between two graph nodes focused around the triangle inequality.

The SPT can be constructed by the well-known static algorithm, the Dijkstra algorithm. However, this well studied static algorithm becomes very inefficient when only a small part of the SPT needs to be updated for link state changes in a network. This is because one link change results in re-computation of the whole tree in every router and entries updating in its routing table. Most of the time, the new SPT shows a little modification compared with the old SPT or no difference at all. The method using the static algorithm for the SPT update incurs a lot of unnecessary computing and routing table entry updates. Thus, it is very important that algorithms for dynamically updating SPT are introduced to handle the link state changes in a network efficiently.

2. Literature Review

P.E. Hart, N.J. Nilsson, and B. Raphael [2] had presented how heuristic information from the problem domain can be joined into a formal numerical hypothesis of graph searching and show an optimality property of class of search strategies.

A.V. Goldberg and C. Harrelson [3] had proposed shortest path algorithms whichutilizesA^* search in combination with a new graph-theoretic lower-bounding method based on landmarks and the triangle inequality. At the point when their research was about to complete, they researched the work of Gutman [16], who studied the P2P issue in a similar context to them. Gutman's algorithms were focused around the concept of reach and oblige to store a single "reach value" and Euclidean coordinates of each vertex. In view of indirect comparison, performance of his quickest algorithm is superior to that of theirs with one landmark and more regrettable than that of theirs with sixteen landmarks. Gutman's methodology obliges a bigger number of suppositions about the input domain than theirs, his preprocessing was more tedious, and his methodology did not appear to adjust to dynamic settings and in theirs too. Then again, his results were extremely fascinating. Specifically, Gutman observed that his concepts can be joined with A^* search. It would be fascinating to check whether utilizing Gutman's reach-based pruning in ALT algorithms will discernibly enhance their effectiveness.

It has ended up progressively apparent that a distance map service is fundamental for distributed applications in the Internet. Be that as it may, the problem of how to construct such a distance map remains to a great extent unexplored. P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang [5] propose a global distance measurement infrastructure called as IDMaps and tackle the problem of how it can be put on the Internet to gather distance data. In the setting of nearest mirror determination for customers, they demonstrated that huge enhancement over arbitrary selection can be attained utilizing placement heuristics that do not oblige a full knowledge of the hidden topology. Additionally, they demonstrated that IDMaps overhead can be minimized by gathering Internet addresses into APs to decrease the quantity of measurements, the quantity of Tracers needed to give helpful distance estimations is fairly little, and applying *t*-spanner to the Tracer-Tracer VLs can bring about linear measurement overhead concerning the quantity of Tracers in the normal case. Generally, this study has given positive results to show that a valuable Internet distance map service can in reality be constructed scalably.

T.S.E. Ng and H. Zhang [6] had examined another class of solutions for the Internet distance prediction issue that is focused around end hosts- maintained coordinates, to be specific the beforehand proposed triangulated heuristic and their new approach called Global Network Positioning (GNP). They propose to apply these arrangements in the setting of a peer-to-peer architecture. These arrangements permit end hosts to carry out distance predictions in a convenient manner and are profoundly adaptable. Utilizing measured Internet distance data, they have directed a reasonable Internet study of the distance prediction precision of the triangulated heuristic, GNP and also IDMaps. They have demonstrated that both the triangulated heuristic and GNP out-perform IDMaps fundamentally. Specifically, GNP is most exact and powerful. They have additionally investigated various key issues identified with the GNP methodology to amplify execution. The primary finding is that a relative error measurement function consolidated with a Euclidean space model of a proper number of measurements accomplishes great performance. They will keep on developing solutions around the GNP system later on.

C. Shahabi, M. Kolahdouzan, and M. Sharifzadeh [7] concentrated on the class of K nearest neighbor (KNN) queries for moving objects in road networks. They proposed to apply an embedding strategy to a road system (RNE) with a specific end goal to transform over its points to a higher dimensional space and utilize the Chessboard metric for distance measurements in the new space. They presented the concept of truncated-RNE that has less computation complexity than RNE yet at the same time gives an adequate accuracy. They proposed two extensions to RNE: 1) D-RNE to embed new points into the embedding space and 2) SP-RNE to discover the shortest path between points in the first road network utilizing their conversion in the embedding space. They want to expand their study in two ways. To begin with, they want to adapt the current Euclidean-based storing strategies for KNN queries to work for the Chessboard metric in the embedding space. Second, they expect to formalize the tradeoffs between the Euclidean metric in the first space and the Chessboard metric in the embedding space to use these tradeoffs inside a query optimizer for picking one methodology over the other.

M. Thorup and U. Zwick [9] proposed approximate distance oracles with quick preprocessing times, basically ideal space requirements, and consistent query time. It yields, as side effects, enhanced algorithms for developing sparse spanners, more minimal tree covers, and more short distance labeling. Because of their basic nature, they anticipate that their concepts will be valuable in numerous different settings. To start with, their fundamental preprocessing algorithm is randomized. While it was not difficult to derandomize it when the full distance matrix was accessible, it is not clear how to obtain it in $o(mn)$ time in the graph context. It appears that new concepts would be required to accomplish that. Their oracles are very nearly ideal, in all regards, when the parameter k is extensive. It remains a fascinating open issue, then again, to minimize the preprocessing times of little extend oracles. The circumstances for stretch 3 are particularly interesting. They demonstrate here that an stretch 3 oracle with a space necessity of $O(n^{3/2})$ can be built in $O(mn^{1/2})$ time. Cohen and Zwick [17] have demonstrated that an stretch 3 oracle that uses $O(n^2)$ space can be built in $O(n2 \log n)$ time.

M.J. Rattigan, M. Maier, and D. Jensen [10] have distinguished a number of strategies for making a network structure index. Two of these strategies, ZONE and DTZ, permit proficient and exact estimation of path lengths between random nodes in a system. Utilizing these files, we can estimate path length between a couple of nodes in constant time, and we can evaluate the closeness centrality of a given hub in time that is direct in the quantity of distant hubs. Likewise, they have indicated experimentally that we can assess the betweenness centrality of a given node by getting fewer than 5% of the nodes examining by breadth-first search. Such proficient calculations of network statistics open up another scope of potential methodologies to knowledge discovery. First and foremost, distinguishing short paths and central nodes are paramount sorts of knowledge discovery for networks. Closeness centrality can administer consideration regarding the center of a network, and betweenness centrality can help distinguish key connectors of generally separate groups.

3. Proposed System

To overcome the limitations of existing methods presented for landmark embedding approach, recently new method presented which is called as a query-dependent local landmark scheme (LLS), which identifies a local landmark specific to a pair of query nodes. In these approach first needs to find query-dependent "local landmark" which is close to both query nodes for more accurate distance estimation. Then, the distance between the two query nodes is estimated as the sum of their shortest distances to the local landmark that is nearer than the any other. This method is only focusing on reducing the distance estimation error. However the other parameters like computation time, processing speed which is very vital in online social networking.

Thus the main objective of this project is to present efficient, scalable and faster method for landmark embedding framework. In a interpersonal network with a huge number of nodes, the technique with the entire SPT re-computation by conventional static algorithms is not adequate. It will take huge amount of computation time. Therefore we tend to extend previously presented robust method with improved SPT approach.

4. Conclusion

Anew shortest path tree-based local landmark scheme has been presented, which discovers a node near the querynode as a question particular nearby point of interest for a triangulation-based shortest distance estimation. Particularly, a local landmark point is characterized as the LCA of the query nodes in a shortest path tree established at a worldwide landmark. Proficient algorithms for indexing and recovering LCAs are presented, which accomplish low offline indexing complexity and online query complexity. This system altogether decreases the distance estimation error, contrasted and global landmark embedding.

References

[1] E.W. Dijkstra, "A Note on Two Problems in Connexion with Graphs," NumerischeMathematik, vol. 1, no. 1, pp. 269-271, 1959.
[2] P.E. Hart, N.J. Nilsson, and B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths," IEEE Trans. Systems Science and Cybernetics, vol. SSC-4, no. 2, pp. 100-107, July 1968.
[3] A.V. Goldberg and C. Harrelson, "Computing the Shortest Path: A* Search Meets Graph Theory," Proc. 16th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA '05), pp. 156-165, 2005.
[4] A.V. Goldberg, H. Kaplan, and R.F. Werneck, "Reach for A*: Efficient Point-to-Point Shortest Path Algorithms," Proc. SIAM Workshop Algorithms Eng. and Experimentation, pp. 129-143, 2006.
[5] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang, "IDMaps: A Global Internet Host Distance Estimation Service,"IEEE/ACM Trans. Networking, vol. 9, no. 5, pp. 525-540, Oct. 2001.
[6] T.S.E. Ng and H. Zhang, "Predicting Internet Network Distance with Coordinates-Based Approaches," Proc. IEEE INFOCOM,pp. 170-179, 2002.
[7] C. Shahabi, M. Kolahdouzan, and M. Sharifzadeh, "A Road Network Embedding Technique for K-Nearest Neighbor Search in Moving Object Databases," Proc. 10th ACM Int'l Symp. Advances in Geographic Information Systems (GIS '02), pp. 94-100, 2002.
[8] J. Kleinberg, A. Slivkins, and T. Wexler, "Triangulation and Embedding Using Small Sets of Beacons," Proc. IEEE 45th Ann.Symp. Foundations of Computer Science (FOCS), pp. 444-453, 2004.
[9] M. Thorup and U. Zwick, "Approximate Distance Oracles,"J. ACM, vol. 52, no. 1, pp. 1-24, 2005.
[10] M.J. Rattigan, M. Maier, and D. Jensen, "Using Structure Indices for Efficient Approximation of Network Properties," Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '06), pp. 357-366, 2006.
[11] H.-P. Kriegel, P. Kro¨ger, M. Renz, and T. Schmidt, "Hierarchical Graph Embedding for Efficient Query Processing in Very Large Traffic Networks," Proc. 20th Int'l Conf. Scientific and Statistical Database Management (SSDBM '08), pp. 150-167, 2008.
[12] M. Potamias, F. Bonchi, C. Castillo, and A. Gionis, "Fast Shortest Path Distance Estimation in Large Networks," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM '09), pp. 867- 876, 2009.
[13] A.D. Sarma, S. Gollapudi, M. Najork, and R. Panigrahy, "A Sketch-Based Distance Oracle for Web-Scale Graphs," Proc. Third ACM Int'l Conf. Web Search and Data Mining (WSDM'10), pp. 401-410, 2010.
[14] A.Gubichev, S. Bedathur, S. Seufert, and G. Weikum, "Fast and Accurate Estimation of Shortest Paths in Large Graphs," Proc. 19th ACM Int'l Conf. Information and Knowledge Management (CIKM '10), pp. 499-508, 2010.
[15] M. Qiao, H. Cheng, and J.X. Yu, "Querying Shortest Path Distance with Bounded Errors in Large Graphs," Proc. 23rd Int'l Conf. Scientific and Statistical Database Management (SSDBM '11), pp. 255- 273, 2011.
[16] R. Gutman. Reach-based Routing: A New Approach to Shortest Path Algorithms Optimized for Road Networks. In Proc. 6th International Workshop on Algorithm Engineering and Experiments, pages 100{111, 2004.
[17] Cohen E., and Zwick U. 2001. All-pairs small-stretch paths. J. Algor. 38, 335–353.