Expected Ranking Method for Probabilistic Top-k Queries in Wireless Sensor Network

K.Shanmuga priya, M.Manopriya

Abstract— **A wireless sensor Network consists of spatially dispersed self-directed sensors to monitor the environment conditions. In probabilistic Top-k, a new approach such as Expected ranking method is used to get accurate result in finding Top-k results and as well as accurate probability. For intercluster processing three algorithm is used namely Sufficient-set based, Necessary-set based and Boundary based algorithm. For reducing transmission cost adaptive algorithm is used and this gives the efficient result in the bounded rounds of communication. This gives least transmission cost and not exceeds two rounds of communication.**

Index Terms—**Expected Ranking, Adaptive, Top-K, cluster.**

I. INTRODUCTION

A wireless Sensor Networks consists of number of nodes that is used in different applications such as military, health care, commerce, etc. Usually a sensor is used for sensing precision to monitor environmental conditions. This will be varies in sensing precision. Every sensor will be varying in the sensing quality. So, whatever the values i.e. raw sensor readings that are collected from sensor is of data uncertainty and energy consumption. Inorder to remove the data uncertainty many approach has been used, but that gives inefficient results.

A data uncertainty is removed by placing more sensor and as well as by calculating the probability i.e. aggregate probability. For obtaining Top-k results aggregate probability does not give accurate results. So, in order to avoid this problem Expected Ranking Method is used after calculating the probability. It is easy to obtain when we have ranking method to get Top-k results from the sensor.

A two-layer approach to managing uncertain data is proposed. An underlying logical model that is complete and it have result of low quality. Process the large amounts of probabilistic data. In order, to avoid the uncertainty and energy consumption here we introducing three new algorithms and an adaptive algorithm are introduced for the dynamic changes in the network. From the group of sensor nodes one of the nodes is selected as the cluster head in the zone. After collected the reading, the sensor nodes send the values to the cluster head for pruning operation. Here sufficient set and necessary set are two important approach used for pruning in the cluster head. The communication cost is also estimated for three proposed algorithms. Finally, we estimate which algorithm is better to obtain Top-k queries in wireless sensor network. Then the Top-k data is obtained by using 3 algorithms.

II. EARLIEST TECHNIQUES

A new algorithm called "Three-Phase Uniform Threshold" (TPUT). TPUT reduces network bandwidth consumption by pruning away ineligible objects, and terminates in three round-trips regardless of data input. The paper presents two sets of results about TPUT. First, trace-driven simulations show that, de-pending on the size of the network, TPUT reduces network traffic by one to two orders of magnitude compared to existing algorithms. Second, TPUT is proven to be instance-optimal on data series that satisfy a lower bound on the slope of decreases in values.

The problem is a generic one that can be found in almost any monitoring network or collaborative distributed system. When the lists are short, the simple method where each node sends its list to the manager works fine. However, when the lists are long, more sophisticated solutions are needed. Such solutions are useful in many systems besides content distribution networks, for example, sensor networks and spam detection networks, as in Ref [6].

Ref [7] shows that, a two-layer approach to managing uncertain data is proposed. An underlying logical model that is complete and one or more working models that are easier to understand. The two-layer approach description is being used in our prototype DBMS for uncertain data, and we identify a number of interesting open problems to fully realize the approach. A novel approach, which computes and ranks efficiently the top-k answers to a SQL query on a probabilistic database. The restriction to top-k answers is natural, since imprecision in the data often lead to a large number of answers of low quality, and users are interested only in the answers with the highest probabilities.

Ranking queries are essential tools to process large amounts of probabilistic data that encode exponentially many possible deterministic instances. The sorted access framework based on expected scores. Propose the notion of approximate distributions in probabilistic data used for ranking. Ref [4], a new approach of expected rank is proposed. The expected rank satisfies all the required properties for a ranking query. Efficient solutions to compute the ranking across the major models of uncertain data, such as attribute-level and tuple-level uncertainty are done.

III. OUR CONTRIBUTION IN PROPOSED SYSTEM

In proposed approach, expected ranking is estimated after aggregate probability is calculated. Then sufficient set and necessary set is calculated to obtain Top-k result. Then the

Manuscript received March, 2014.

K.Shanmuga priya*, Department of Computer Science and Engineering, K.S.R College of Engineering, Erode, India, Mobile: +91-8870161217.*

M.Manopriya*, Assistant Professor Department of Computer Science and Engineering, K.S.R College of Engineering, Erode, India,*

ISSN: 2278 – 7798

three algorithms are used to send those set to base station for finding Top-k.

A. Sensor Network

A wireless sensor network that consists of a large number of sensor nodes deployed in a geographical region. Feature readings (e.g., moisture levels or speed of wind gust) are collected from these distributed sensor nodes. In this network, sensor nodes are grouped into clusters, within each of which one of sensors is selected as the cluster head for performing localized data processing. By using statistic methods a cluster head may generate a set of data tuples for each zone within its monitored region. Each tuple is comprised of tupleid, zone, a derived possible attribute value, along with a confidence that serves as a measurement of data uncertainty.

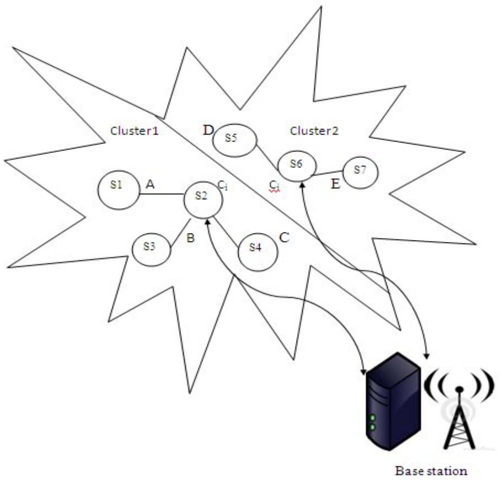

Figure 1: Wireless sensor network with 5 zones (A, B, C, D, and E) and with cluster head (c_i) and two clusters C_i.

B. Tuple Probability & Ranking

We define the tuple structure for each Zone. Then calculate the aggregate probability for all zones. Let W denote a possible world which consists of a subset of tuples in T and W denote the set of all possible worlds. One tuple of each x-tuple exists in any w϶W and tuples belonging to different x-tuples are independent of each other. After that we perform the expected ranking method to rank the tuple.

C. Intracluster Pruning

In a cluster-based wireless sensor network, the cluster heads are responsible for generating uncertain data tuples from the collected raw sensor readings within their clusters. We propose the method of sufficient set and necessary set, and describe how to identify them from local data sets at cluster heads. We use the PT-Topk query as a test case to derive sufficient set and necessary set and show that the top-k probability of a tuple t obtained locally is an upper bound of its true top-k probability.

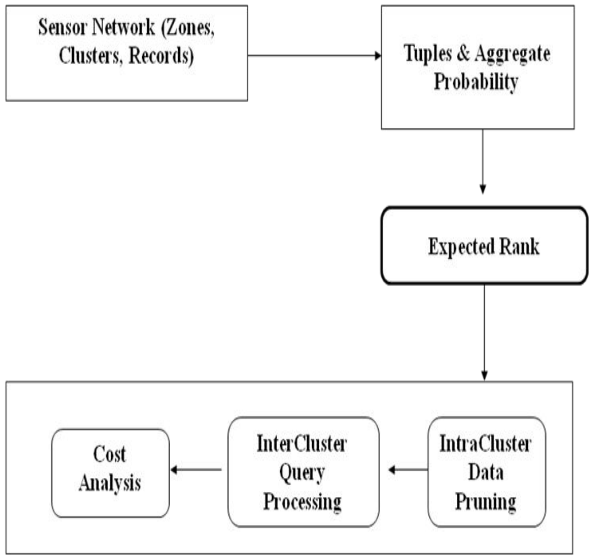

Figure 2: Architecture of overall analysis

D. Intercluster Query Processing

The propose three distributed algorithms for probabilistic top-k queries in wireless sensor networks, such as sufficient set based, necessary set based and boundary based algorithm.

Sufficient Set Method:

After collecting data tuples from its cluster, c_i computes the $S(T_i)$ from the locally collected tuples and sends it to the base station. If a sufficient set cannot be obtained, then all the tuples are transmitted to the base station. After receiving the transmitted data tuples from all the cluster heads, they compute final answer.

Necessary Set Method:

After receiving all the necessary sets, the received tuples are merged into a table in a base station and finds the necessary boundary called the global boundary (GB)). If GB is ranked higher than the highest ranked necessary boundary, all the necessary data have delivered to the base station. Otherwise, it entering the second phase, it sends the GB back to the c_i, which return the supplementary data tuples ranked between its local necessary boundary and GB. Then, the base station computes the final answer.

Boundary Based Method:

The boundary-based method first delivers the local knowledge in clusters, in the form of NB and SB, to the base station in order to provide a refined global data pruning among clusters. It is done instead of directly delivering data tuples to the base station.

E. Cost Analysis

We perform a cost analysis on data transmission of the three proposed methods by using adaptive algorithm.

Adaptive Algorithm:

The performance of the data transmission using proposed method is affected by factors such as the skewness of data distribution among clusters which may change continuously over time. A cost-based adaptive algorithm that is used dynamically Sufficient Set Based, Necessary Set

Based, and Boundary Based as the data distribution within the network changes.

IV. CONCLUSION

The concepts of sufficient set and necessary set are universal and can be easily extend to a network with tree topology. The approach is evaluated and shows that the algorithms reduce data transmissions significantly. In the proposed approach, there may be a variation in distributing the workload among the sensors. It leads to data collision and there is wastage of energy. It is also not guaranteed that the routing tree provide a faster path to the sensors with higher query load. To overcome this problem, we enhance a query load-based spanning tree construction method. It reduces the query response delay as well as energy consumption in query execution and provides query response with the best possible accuracy. Then through expected ranking method we almost find Top-k data from the network.

REFERENCES

1) M. Hua, J. Pei, W. Zhang, and X. Lin, "Ranking Queries on Uncertain Data: A Probabilistic Threshold Approach,"Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08), 2008.
2) F. Li, K. Yi, and J. Jestes, "Ranking Distributed Probabilistic Data," Proc. 35th SIGMOD Int'l Conf. Management of Data (SIGMOD '09), 2009.
3) C. Jin, K. Yi, L. Chen, J.X. Yu, and X. Lin, "Sliding-Window Top-k Queries on Uncertain Streams,"Proc. Int'l Conf. Very Large Data Bases (VLDB '08), 2008.
4) G. Cormode, F. Li, and K. Yi, "Semantics of Ranking Queries for Probabilistic Data and Expected Ranks,"Proc. IEEE Int'l Conf. Data Eng. (ICDE '09), 2009.
5) X. Liu, J. Xu, and W.-C. Lee, "A Cross Pruning Framework for Top-k Data Collection in Wireless Sensor Networks,"Proc. 11th Int'l Conf. Mobile Data Management,pp. 157-166, 2010.
6) Pei Cao and Zhe Wang, "Efficient Top-K Query Calculation in Distributed Networks", Proc.Int'l conf.very large Data Bases, 2010.
7) Anish Das Sarma, Omar Benjelloun, Alon Halevy and Jennifer Widow, "Working Models for Uncertain Data", 2005.
8) J. Li, B. Saha, and A. Deshpande, "A Unified Approach to Ranking in Probabilistic Databases," Proc. Int'l Conf. Very Large Data Bases (VLDB),vol. 2, no. 1, pp. 502-513, 2009.
9) K. Yi, F. Li, G. Kollios, and D. Srivastava, "Efficient Processing of Top-k Queries in Uncertain Databases with X-Relations,"IEEE Trans. Knowledge and Data Eng.,vol. 20, no. 12, pp. 1669-1682, Dec. 2008.
10) D. Wang, J. Xu, J. Liu, and F. Wang, "Mobile Filtering for ErrorBounded Data Collection in Sensor Networks,"Proc. 28th Int'l Conf. Distributed Computing Systems (ICDCS '08),pp. 530-537, 2008.

K.Shanmuga priya received her B.E degree in Computer Science and Engineering from Anna University, Chennai and currently pursuing her M.E degree in Computer Science and Engineering from K.S.R. College of Engineering under Anna University, Chennai. Her research interest includes Networking and Data Mining.

Ms.M.Manopriya received B.E degree in CSE from Anna University, Chennai. She received M.E degree in Computer Science and Engineering, under Anna University of Technology, Coimbatore. She is currently working as Assistant Professor in Department of CSE in K.S.R. College of Engineering, Erode. Her teaching experience is of 2 years. Her research interest includes Computer Networks and Data Mining.

ISSN: 2278 – 7798