ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Disperse Processin Of Presumption Top-K Inquire In Wireless Sensor Networks

M.P.Navaneedha[1], P. Matheswaran[2]

M.E. CSE Final Year, K.Ramakrishnan College of Technology, Trichy, India[1]

Assistant Professor, K.Ramakrishnan College of Technology, Trichy, India[2]

Abstract: WSN is a wider range of network to collect the data and extract information from the physical world. The new technology is implemented in various fields of military, science, industry, commerce etc. The sensing quality produces the good results of accuracy and tolerance to hardware and external noise. In existing system two techniques are used namely sufficient set and necessary set. But, it does not reduce the transmission cost. In the proposed system, we develop suitable algorithms namely SSB, NSB and BB for intercluster query processing with boundary rounds of communications. There are three algorithms which dynamically changes of data distribution in the network and minimize the transmission cost. In the wireless sensor networks, the sufficient set of data with two-tier hierarchical and tree structure network topologies. Hence, the experimental results show that the proposed algorithms reduce the data transmission significantly. According to various conditions which give the near optimal solutions with high performance.

Keywords*:* WSN, centralized, distributed, sensor network

I. INTRODUCTION

Wireless sensor networks are revolutionizing the ways to collect and use information from the physical world. This new technology has resulted in significant impacts on a wide array of applications in various fields, including military, science, industry, commerce, transportation, and health-care. However, the quality of sensors varies significantly in terms of their sensing precision, accuracy, tolerance to hardware/external noise, and so on. For example, studies show that the distribution of noise varies widely in different photovoltic sensors, precision and accuracy of readings usually vary significantly in humidity sensors, and the errors in GPS devices can be up to several meters. Thus, sensor readings are inherently uncertain. To facilitate management of uncertain data, research on probabilistic databases has received renewed attentions in the past few years. Most of the recent works on probabilistic data modeling propose to associate a confidence (in form of probability) with a data record/tuple to capture the data uncertainty and thus carry a possible world semantic. We first use an environmental monitoring application of wireless sensor network to introduce some basics of probabilistic databases. Consider 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.

Due to sensing imprecision and environmental interferences, the sensor readings are usually noisy. Thus, multiple sensors are deployed at certain zones in order to improve monitoring quality. 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. In this example, we assume that each tuple is comprised of tuple id, zone, a derived possible attribute value, along with a confidence that serves as a measurement of data uncertainty. Thus, the data tuples corresponding to the same zone collectively represent the probabilistic distribution of derived possible values for the zone. Since the existence of possible values in these tuples is exclusive to each other, they naturally form a logical tuple, called x-tuple. Top-k queries on

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

uncertain data have received increasing interests recently. Given a scoring function and a data set, a top-k query finds the highest ranked k tuples over all possible worlds. The answers to this query may vary in different possible worlds, making the problem interesting and challenging. Moreover, various semantics of top-k queries can be employed in different applications.

II. DESIGN GOAL

By using static methods, a cluster head may generate a set of data tuples for each zone within its monitored region. There are three proposed algorithm is to minimize the communication and energy overhead in responding to changing data distribution in the network. The algorithms are sufficient set-based (SSB) algorithm. The advantages of NSB and BB take advantage of the skewed necessary sets and necessary boundaries among local clusters to obtain their global boundaries, respectively, which are very effective for intercluster pruning. The transmission cost increases for all algorithms because the number of tuples needed for query processing is increased.

Fig 2.1 Architecture of Top-k process

2.1 Related work of top-K Query processing

Moreover, various semantics of top-k queries can be employed in different applications. There are several top-k query semantics and solutions proposed recently, including U-Topk and UkRank, PT-Topk in, PK-Topk in, expected rank and so on. A common way to process probabilistic top-k queries is to first sort all tuples based on the scoring attribute, and then process tuples in the sorted order (until some termination condition is met) to compute the final answer set. Answer sets for the abovementioned queries typically consist of highly ranked tuples in the sorted list because tuples positioned lowly usually do not have the required ranks and confidence to be included in the answer sets. While these works produce insightful results, they are conducted on a centralized database rather than in a distributed setting as we target on in this paper. Only recently, a study on processing a probabilistic query with "expected rank" semantic in distributed systems has appeared.

III. OVERVIEW

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Wireless sensor networks are revolutionizing the ways to collect and use information from the physical world. This new technology has resulted in significant impacts on a wide array of applications in various fields, including military, science, industry, commerce, transportation, and health-care. However, the quality of sensors varies significantly in terms of their sensing precision, accuracy, tolerance to hardware/external noise, and so on. For example, studies show that the distribution of noise varies widely in different photovoltic sensors, precision and accuracy of readings usually vary significantly in humidity sensors, and the errors in GPS devices can be up to several meters. Thus, sensor readings are inherently uncertain. To facilitate management of uncertain data, researches on probabilistic databases have received renewed attentions in the past few years. Most of the recent works on probabilistic data modeling propose to associate a confidence (in form of probability) with a data record/tuple to capture the data uncertainty and thus carry possible worlds semantic.

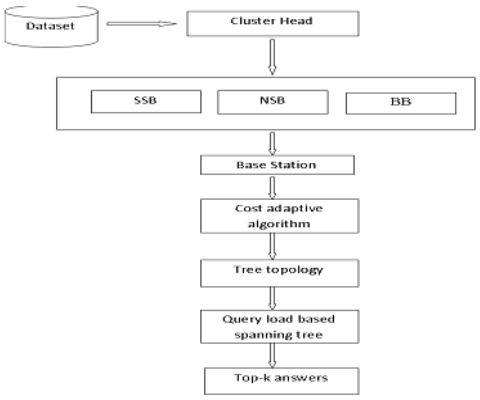


Fig.3.1 Data Set for System Design

The distinction between digital and analogue has become tired and inappropriate. This is also true in the world of architectural drawing, which paradoxically is enjoying a renaissance supported by the graphic dexterity of the computer. This new fecundity has produced a contemporary glut of stunning architectural drawings and representations that could rival the most recent outpouring of architectural vision.

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. Due to sensing imprecision and environmental interferences, the sensor readings are usually noisy. Thus, multiple sensors are deployed at certain zones in order to improve monitoring quality. 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. Thus, the data tuples corresponding to the same zone collectively represent the probabilistic distribution of derived possible values for the zone. Since the existences of possible values in these tuples are exclusive to each other, they naturally form a logical tuple, called x-tuple.

B Top-k Probability

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

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. The probability that wϵW exists is

$$P(W) = \prod_{t \in T \wedge t \in W} P(t) \prod_{t \in T \wedge t \notin W} (1 - P(t))$$

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 notion of sufficient set and necessary set, and describe how to identify them from local data sets at cluster heads. Next, 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. Given an uncertain data set Ti in cluster C_i. The sufficient boundary of T_i, $SB(T_i)$ is the highest ranked tuple t_{sb} such that the following condition is satisfied

$$\sum_{j=1}^{k} r_{sb,j-1}[D] < p$$

Given an uncertain data set T_i in cluster C_i, the necessary boundary $NB(T_i)$ is the lowest ranked tuple t_{nb} such that:

$$P_{L,topk}(t_{nb}) \geq p.$$

D Intercluster Query Processing

The notion of sufficient and necessary sets as a basis, we propose three distributed algorithms for processing probabilistic top-k queries in wireless sensor networks, namely 1) Sufficient Set-based method; 2) Necessary Set-based method; and 3) Boundary-based method. Here we focus on addressing the communication overhead which is critical for wireless sensor networks and their applications. For simplicity, we logically assume single-hop transmission in both intracluster and intercluster communications. Nevertheless, our algorithms are not restricted to this assumption and can be extended for the multihop communications. As long as the base station receives all the candidate data tuples and supplementary tuples, we are able to compute the final answer with a generic centralized algorithm. Note that the supplementary tuples refer to the unqualified data tuples needed for computing the confidence probability of final answer.

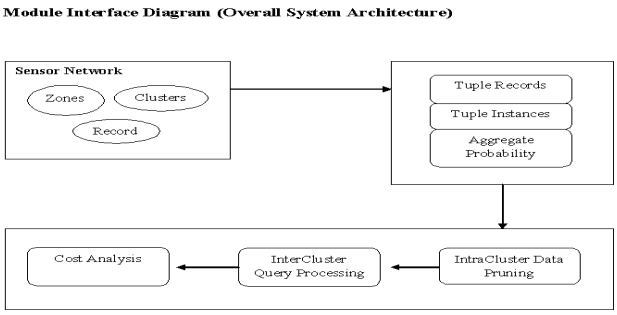


Fig 3.2 Module Interface diagram

E Cost Analysis

We first perform a cost analysis on data transmission of the three proposed methods. Let M denote the number of clusters in the network; and Sq,Sb, and Sd be the sizes of query message, boundary message, and data message, respectively. Also let |S(Ti)| and |N(Ti)| denote the cardinalities of the sufficient set and necessary set of the data set Ti in a cluster Ci(1<i<M), respectively.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

The total transmission cost for SSB

$$C_{SSB} = M.S_q + \sum_{i=1}^{M} |S(T_i)|.S_d.$$

The total transmission cost for NSB

$$C_{NSB} = \begin{cases} M.S_q + \sum_{i=1}^{M} |N(T_i)|.S_d \text{ (1 Phase)} \\ 2M.S_q + M.S_b \sum_{i=1}^{M} |N(T_i)|.S_d + \sum_{i=1}^{M} |ST(c_i)|.S_d \text{ (2 Phases)} \end{cases}$$

The total transmission cost for BB

$$C_{BB} = 2M.S_q \mid 3M.S_b \mid \sum_{i=1}^{M} |BB(c_i)|.S_d.$$

Input and Output information

- Sensor Network – Network Details; Top-k Probability – Answer Set Probability; Intracluster Pruning – SS & NS Boundaries; Intercluster Query Processing – Final answer set

Cost Analysis – Cost for all sets

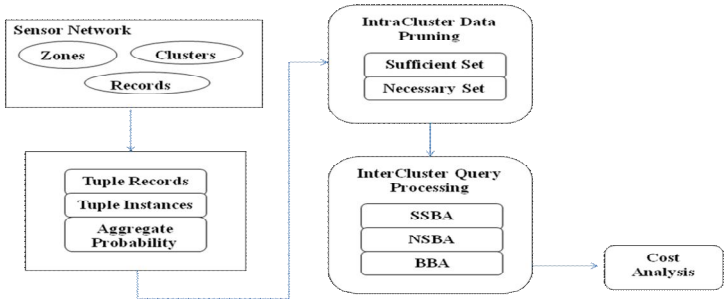


Fig 3.3 Flow Diagram

IV. OTHER ISSUES

Confidentiality, integrity, availability, and authentication are four important security issues to be addressed in clusters connected by an unsecured public network. Rather than addressing all the security aspects, particular attention is confidentiality services for messages passed among computing nodes in an unsecured cluster.

Sufficient Set-Based Algorithm

Compute the sufficient boundary $SB(T_i)$ of T_i
If $SB(T_i)$ exists **then**
$S(T_i) \leftarrow \{t \backslash t \leq_f SB(T_i) t \in T_i\}$
$T_i' \leftarrow S(T_i)$
else
$T_i' \leftarrow T_i$
end if
Deliver T_i' to the base station

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Necessary Set Boundary Algorithm

Compute the necessary boundary $NB(T_i)$ of T_i
$N(T_i) \leftarrow \{t \backslash t \leq_f NB(T_i) t \in T_i\}$
Deliver $N(T_i)$ to the base station
if Receive GB from the base station **then**
$N'(T_i) \leftarrow \{t \backslash t \leq_f GB\ t \in [T_i - N(T_i)]_i\}$
Send $N'(T_i)$ to the base station
end if

Boundary Based Algorithm

Compute $NB(T_i)$ and $SB(T_i)$of T_i
Send $NB(T_i)$ and $SB(T_i)$ to the base station
Receive GB from to the base station
$T_i' \leftarrow \{t \backslash t \leq_f GB\ )(t \in T_i)\}$
Deliver T_i' to the base station

/*Base Station Side*/
Collect $NB(T_i)$and $SB(T_i)$ from all c_i $(1 \leq i \leq M)$
Let $SB_{highest}$,NB_{lowest} be the highest ranked $SB(T_i)$and the lowest ranked $NB(T_i)$ respectively, where$(1 \leq i \leq M)$
if$SB_{highest} NB_{lowest}$**then,** $GB \leftarrow SB_{highest}$
else $GB \leftarrow NB_{lowest}$
end if
Broadcast GB to each cluster head
Collect T_i' from all $c_i(1 \leq i \leq M)$ $T_i' \leftarrow U_1 \leq i \leq M T_i'$
Execute centralized algorithm over T

V. EVALUATION

In this section, we first conduct a simulation-based performance evaluation on the distributed algorithms for processing PT-top k queries in two-tier hierarchical cluster- based wireless sensor monitoring system. As discussed, limited energy budget is a critical issue for wireless sensor network and radio transmission is the most dominate source of energy consumption. Thus, we measure the total amount of data transmission as the performance metrics. Notice that, response time is another important metrics to evaluate query processing algorithms in wireless sensor networks. All of those three algorithms, i.e., SSB, NSB, and BB, perform at most two rounds of message exchange, thus clearly outperform an iterative approach (developed based on the processing strategy in [14]), which usually needs hundreds of iterations. Note that, there is not much difference among SSB, NSB, and BB in terms of query response time, thus we focus on the data transmission cost in the evaluation. Finally, we also conduct experiments to evaluate algorithms, SSB-T, NSB-T, and NSB-T-Opt under the tree-structured network topology. A value generated based on the node is location in order to inject spatial locality and is a random variable ranging from 0 to 1. When ¼ 1, the data distribution is highly spatial-correlated. When? ¼ 0, the data distribution is totally random. The real sensor traces are from the Intel Lab Data [35]. We treat the readings for each sensor node in the traces as for an area of interest by assigning readings to sensor nodes in an area.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

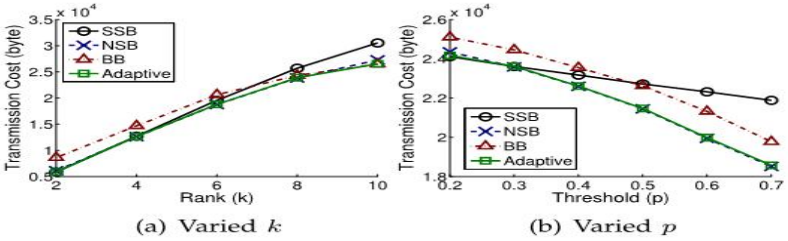
Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

VI. OVERALL PERFORMANCE

We first validate the effectiveness of our proposed methods in reducing the transmission cost against two baseline approaches, including 1) a naive approach, which simply transmits the entire data set to the base station for query processing; 2) an iterative approach devised based on the



(a) Varied k (b) Varied p

VII. CONCLUSION

We propose the sufficient set and necessary set for efficient in-network pruning of distributed uncertain data in probabilistic top-k query processing. Introduce new algorithms namely SSB, NSB, and BB for in-network processing of PT-Top k queries. We derive a cost model on communication cost of the three proposed algorithms and propose a cost-based adaptive algorithm that adapts to the application dynamics.

REFERENCES

[1] A.D. Sarma, O. Benjelloun, A. Halevy, and J. Widom, "Working Models for Uncertain Data,"Proc. 22nd Int'l Conf. Data Eng. (ICDE '06), p. 7, 2006.
[2] C. Re, N. Dalvi, and D. Suciu, "Efficient Top-k Query Evaluation on Probabilistic Data," Proc. Int'l Conf. Data Eng. (ICDE '07), pp. 896-905, 2007.
[3] 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.
[4] F. Li, K. Yi, and J. Jestes, "Ranking Distributed Probabilistic Data," Proc. 35th SIGMOD Int'l Conf. Management of Data (SIGMOD '09), 2009.
[5] Y. Diao, D. Ganesan, G. Mathur, and P.J. Shenoy, "Rethinking Data Management for Storage-Centric Sensor Networks,"Proc. Conf. Innovative Data Systems Research (CIDR '07),pp. 22-31, 2007.
[6] M.A. Soliman, I.F. Ilyas, and K.C. Chang, "Top-k Query Processing in Uncertain Databases,"Proc. Int'l Conf. Data Eng. (ICDE '07),2007
[7] 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.
[8] 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.
[9] 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.
[10] X. Lian and L. Chen, "Probabilistic Ranked Queries in Uncertain Databases,"Proc. 11th Int'l Conf. Extending Database Technology (EDBT '08), pp. 511-522, 2008.