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, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Disperse Processing of Presumption Top-K Inquire in Wireless Sensor Networks

K.Shanmuga priya[1], M.Manopriya[2]

M.E, Department of CSE, K.S.R.College of Engineering, Tiruchengode, Tamilnadu, India[1]

Assistant Professor, Department of CSE, K.S.R.College of Engineering, Tiruchengode, Tamilnadu, India[2]

ABSTRACT: The Top-k query processing in sensor network is to evaluate the high ranked data tuples from the sensor node. Data pruning operations are performed on the cluster head for uncertain data. Accordingly, the new algorithm such as Sufficient Set-Based (SSB), Necessary Set-Based (NSB) for local Intercluster processing of the query. The Boundary-Based (BB) algorithm is proposed to process the query between two different boundary regions. There is the data distribution that changes dynamically, the adaptive algorithm is used to reduce transmission cost and there is a constant round of data communication in the network. The performance is evaluated in two-tier hierarchical network.

KEYWORDS— Data Pruning, Intercluster, Adaptive algorithm, Intracluster.

I. INTRODUCTION

Wireless Sensor Network usually collects the information from the physical environment. The information that is collected from those physical environments is of uncertain data and there is a presence of noise. Sometimes, many sensor nodes are deployed in an environment to avoid the data uncertainty for sensing precision. This is used in various applications such as military, commerce and healthcare etc. In this networks, data uncertainty and energy consumption is main issue when we consider in the sensor networks. Recent research on probabilistic data has received renewed attentions and they are measured by confidence values that are associated with it. It is measured by using fixing the threshold limit for removing uncertainty. Then after the process in sensor nodes the information is delivered to the base station, sometimes it takes many rounds of communication to complete the process. So, the energy consumption will be high as it takes many rounds of communication. They have anumber of accessed tuples and materialized search states, as in [6]. So, it will take high memory to process each state. The ranking across the major models of uncertain data, such as attribute and tuple-level uncertainty are done, as in [8]. The imprecision in the data often lead to a large number of answers of low quality, as in [2].

Reference [1] shows, 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, as in [4]. 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.

II. INTRACLUSTER PROCESSING

In the intracluster processing, we perform a data pruning operations in the cluster heads. It performs pruning operation based on the two important approaches such as by having sufficient set and Necessary set, 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 tuples obtained locally is an upper bound of its true top-k probability.

A. Sufficient Set:

The sufficient set $S(T_i)$ is obtained by evaluating the uncertain data set (T_i) from the cluster, there exist the tuple x. It can be represented as,

$$S(T_i) = \{ x|x =_f x_{sb} \}$$

B. Necessary Set:

The necessary set $N(T_i)$ is obtained by evaluating the local set (T_i) from the cluster, there exist the tuple x. it can be represented as,

$$N(T_i) = \{ x|x \in T_i, x <_f x_{nb} \}$$

III. INTERCLUSTER PROCESSING

In the intercluster processing, the sufficient and necessary sets are the basis and three distributed algorithms such as Sufficient Set-based algorithm (SSB), Necessary Set-based algorithm (NSB), Boundary-based algorithm (BB) for processing probabilistic top-k queries in wireless sensor networks are going to implemented in both sets. 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

A. Sufficient Set-Based Algorithm

After collecting data tuples from its cluster, a cluster head computes the sufficient set from the local collected tuples and sends it to the base station. If a sufficient set cannot be obtained, all the local data tuples are transmitted to the base station. After receiving the transmitted data tuples from all the cluster heads, the base station computes the query answer.

Algorithm 1: *SSB ALGORITHM*

AT CLUSTER HEAD (c_i):

1. ***if*** *$SB(T_i)$ exits*

$S(T_i) \leftarrow \{x|x \leq_f SB(T_i) \wedge x \in T_i \}$

$Y_i \leftarrow S(T_i)$

else

$Y_i \leftarrow T_i$

2. Now, $\mathbf{Y_i}$ is delivered to the base station.

AT BASESTATION:

1. It receive the tuples $\mathbf{Y_i}$ from the cluster head.$(1 \leq i \leq N)$
2. $T' \leftarrow \bigcup_{1\leq i\leq N} Y_i$

Where,

x is the tuples
c_i is the cluster head
$S(T_i)$ is the sufficient set
T_i is the records collected from the sensor
N is the number of clusters in the zone
C_i is the cluster
Y_i is the sufficient boundary for SSB
T' is the aggregation of data sets received from the clusters

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, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

B. Necessary Set-Based Algorithm

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

Algorithm 2: *NSB ALGORITHM*

AT CLUSTER HEAD:
1.Compute the necessary boundary $NB(T_i)$,
$N(T_i) \leftarrow \{x|x \leq_f NB(T_i) \wedge x \in T_i\}$
2. Deliver $N(T_i)$ to the base station
3. if cluster head receive GB from the base station
then
$N'(T_i) \leftarrow \{ x|x \leq_f GB \wedge x \in [T_i - N(T_i)]\}$
Now, $N'(T_i)$ is send to the base station.
end if

AT BASESTATION:
1. It receives the tuples **N (T_i)** from the cluster head. $(1 \leq i \leq N)$
$T' \leftarrow U_{1 \leq i \leq N} N(T_i)$
2. Now, it will calculate the global boundary.
3. if global boundary GB is less than that of $NB(T_i)$, then
It calculate the final necessary boundary
else
It will broadcast GB to c_i and once again it collects necessary tuples
$T' \leftarrow U_{1 \leq i \leq N} N'(T_i)$
end if

Where,
x is the tuples
c_i is the cluster head
$N(T_i)$ is the necessary set
$NB(T_i)$ is the necessary boundary
T_i is the records collected from the sensor
N is the number of clusters in the zone
T′ is the aggregation of data sets received from the clusters

C. Boundary-Based Algorithm

The boundary-based method first delivers the local knowledge in clusters, in the form of sufficient boundary and necessary boundary, 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

Algorithm 3: *BB Algorithm*

AT CLUSTER HEAD:
1. Calculate the Necessary Boundary (NB) and Sufficient Boundary (SB) and send it to the base station.
2. Base station receive Global Boundary (GB)

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, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

3. $Y_i \leftarrow \{ x | x \leq_f GB \; x \in [T_i - N(T_i)]\}$
4. Now, Y_i is delivered to the base station.

AT BASESTATION:
1. It will receive the NB and SB from cluster heads (c_i),
2. Now, base station computes the (Sufficient Boundary_{high} and Necessary Boundary_{low}).
3. if $SB_{high} < NB_{low}$, then

    $SB_{high} \rightarrow GB$

  else

    $NB_{low} \rightarrow GB$

  end if
4. Now, broadcast the global boundary to each C_i

    $T' \leftarrow U_{1\leq i\leq N} \; Y(T_i)$

Where,

x is the tuple
c_i is the cluster head
$S(T_i)$ is the sufficient set
$N(T_i)$ is the necessary set
T_i is the records collected from the sensor
N is the number of clusters in the zone
Y_i is the sufficient boundary for SSB
T′ is the aggregation of data sets received from the clusters

D. 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. So, to analyze the cost during data transmission a cost-based adaptive algorithm is used. A cost-based adaptive algorithm that switches dynamically among SSB, NSB, and BB as the data distribution within the network changes.

Algorithm 4: *Adaptive*

Count=0 ; Z_{SSB}, Z_{NSB}, Z_{BB} =0
Where R is varied window size.
Then estimate the cost of C_{SSB}, C_{NSB}, C_{BB}
$Z_{SSB} \leftarrow Z_{SSB} + C_{SSB}$
$Z_{NSB} \leftarrow Z_{NSB} + C_{NSB}$
$Z_{BB} \leftarrow Z_{BB} + C_{BB}$
if count $\geq$ R then
if $Z_{SSB} = \min\{ Z_{SSB}, Z_{NSB}, Z_{BB}\}$
    then switch to SSB
end if
if $Z_{NSB} = \min\{ Z_{SSB}, Z_{NSB}, Z_{BB}\}$
    then switch to NSB
end if
if $Z_{BB} = \min\{ Z_{SSB}, Z_{NSB}, Z_{BB}\}$
    then switch to SSB
end if
end if

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, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

IV. RESULTS

We evaluate the performance of our newly proposed algorithm with that of existing approach such as naive and iterative approach. Both synthetic data and real data sets are used for performance evaluation. While evaluating performance of iterative approach it takes more than 60-200 rounds of communication (i.e) at one round of communication only one data tuple is send to the base station. So, it takes many rounds of communication to complete the process. Comparing our newly proposed algorithm, it reduces data transmission against that of two approaches. Our newly proposed algorithms complete within the two rounds. Then the adaptive algorithm gives the least transmission cost

V. 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.

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.
11) 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.
12) 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.
13) 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.