International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume No.3 Issue No.5, pp : 658-660 1 May 2014

Distributed Processing of Probabilistic Top-K Inquire in Wireless Sensor Networks

D.Dhayalan, Manikandan.S
Department Of MCA,Vel Tech Multi Tech Dr. RR and D r. SR Engg College, Avadi, Chennai, Tamilnadu
Email ID: dayalan.moorthy@rediffmail.com, manikandans28591@gmail.com

Abstract*—*The Top-k query processing is to evaluate the high ranked data tuples from the sensor node. Accordingly, the algorithms SSB, NSB for local Intercluster processing of the query. The data distribution that changes dynamically, the adaptive algorithm is used to decrease transmission cost and a constant round of data communication in the network.

***Keywords*—Data Pruning, Inter-cluster, Adaptive algorithm, Intra-cluster**

1. 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, science, industry, commerce and healthcare etc. In this networks, data insecurity and energy consumption is major issue when we consider in the sensor networks. New 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 later 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 expenditure will be high as it takes many rounds of communication. They have a number of retrieved tuples and materialized search states, as in [6]. So, it will take high memory to process each state. The ranking across the main models of indeterminate 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 huge 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 cluster of sensor nodes one of the nodes is selected as the cluster head in the zone. After composed 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.

2. 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. Following, 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(Ti) is found by evaluating the uncertain data set (Ti) from the cluster, there exist the tuple x.
It can be represented as,
S(Ti) ={ x|x = f xsb }

B. Necessary Set:

The necessary set N(Ti) is obtained by evaluating the local set (Ti) from the cluster, there exist the tuple x. it can be represented as,
N(Ti) ={ x|x Є Ti , x <f xnb }

3. 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 multi hop communications. Providing the base station receives all the candidate data tuples and supplementary tuples, we are able to calculate the final answer with a generic centralized algorithm

A. Sufficient Set-Based Algorithm

After collecting data tuples from its cluster, a cluster head calculates the sufficient set from the local composed tuples and drives it to the base station. If a sufficient set cannot be achieved, all the local data tuples are transferred to the base station. After receiving the transferred data tuples from all the cluster heads, the base station calculates the query answer.

Algorithm 1: *SSB ALGORITHM*

AT CLUSTER HEAD (ci):

1. ***if*** *SB(Ti) exits*

S(Ti) ← {x/x≤f SB(Ti) Λ x ϵTi }

Yi ←S(Ti)

Else

$Yi \leftarrow Ti$

2. Now, **Yi** is delivered to the base station.

AT BASESTATION:

1. It receive the tuples **Yi** from the cluster head.($1 \leq i \leq N$)
2.$T' \leftarrow U1 \leq i \leq N$ Yi

Where,
x is the tuples
ci is the cluster head
S(Ti) is the sufficient set
Ti is the records composed from the sensor
N is the number of clusters in the region
Ci is the cluster
Yi is the sufficient boundary for SSB
T′ is the combination of data sets received from the clusters

B. Necessary Set-Based Algorithm

After receiving all the necessary sets, the base station combines 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 thenecessary data have been delivered to the base station. Thus, the base station calculatesthe final answer. Otherwise, inflowing the second phase, the base station directs the GB back to the cluster heads, which return the additional 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(Ti),
$N(Ti) \leftarrow \{x | x \leq f\, NB(Ti)\ \Lambda\ x \in Ti\}$
2. Deliver N(Ti) to the base station
3. if cluster head receive GB from the base station
then
$N'(Ti) \leftarrow \{x | x \leq f\, GB\ \Lambda\ x \in [Ti - N(Ti)]\}$
Now, N′(Ti) is send to the base station.
end if

AT BASESTATION:

1. It receives the tuples **N (Ti)** from the cluster head. ($1 \leq i \leq N$)
$T' \leftarrow U1 \leq i \leq N\ N(Ti)$
2. Now, it will compute the global boundary.
3. if global boundary GB is less than that of NB(Ti), then
It compute the final necessary boundary
Else
It will transmission GB to ci and once again it collects necessary tuples
$T' \leftarrow U1 \leq i \leq N\ N'(Ti)$**)**
end if

Where,
x is the tuples
ci is the cluster head
N(Ti) is the necessary set
NB(Ti) is the necessary boundary
Ti is the records composed from the sensor
N is the number of clusters in the zone
T′ is the aggregation of data sets established from the clusters

C. Boundary-Based Algorithm

The boundary-based method first delivers the local information in clusters, in the form of sufficient boundary and necessary boundary, to the base station in order to deliver 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)
3. $Yi \leftarrow \{x | x \leq f\, GB\ x \in [Ti - N(Ti)]\}$
4. Now, Yi is delivered to the base station.

AT BASESTATION:
1. It will receive the NB and SB from cluster heads (ci),
2. Now, base station computes the (Sufficient Boundaryhigh and Necessary Boundarylow).
3. ifSBhigh<NBlow , then
SBhigh →GB
else
NBlow → GB
end if
3. Now, broadcast the global boundary to each Ci
4. $T' \leftarrow U1 \leq i \leq N\ Y(Ti)$
Where,
x is the tuple
ci is the cluster head
S(Ti) is the sufficient set
N(Ti) is the necessary set
Ti is the records composed from the sensor
N is the number of clusters in the region
Yi is the sufficient boundary for SSB
T′ is the aggregation of data sets established 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 examine the cost throughout data transmission a cost-based adaptive procedure is used. A cost-based adaptive algorithm that changes energetically among SSB, NSB, and BB as the data distribution within the network changes.

Algorithm 4: *Adaptive*

Count=0 ; ZSSB , ZNSB , ZBB =0
Where R is varied window size.
Then estimate the price of CSSB, CNSB, CBB
ZSSB ← ZSSB + CSSB
ZNSB ← ZNSB + CNSB

```
        ZBB ← ZBB + CBB
if count ≥ R then
if ZSSB = min{ ZSSB , ZNSB , ZBB}
        then switch to SSB
end if
if ZNSB = min{ ZSSB , ZNSB , ZBB}
        then switch to NSB
end if
if ZBB = min{ ZSSB , ZNSB , ZBB}
        then switch to SSB
end if
end if
```

4. Results

We evaluate the performance of our newly proposed algorithm with that of existing approach such as naive and iterative method. Both synthetic data and real data sets are used for performance evaluation. While estimating 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.

5. Conclusion

The approach shows that the algorithms reduce data transmissions significantly. In the proposed approach, 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

i. *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.*

ii. *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.*

iii. *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.*

iv. *Li, K. Yi, and J. Jestes, "Ranking Distributed Probabilistic Data," Proc. 35th SIGMOD Int'l Conf. Management of Data (SIGMOD '09), 2009.*

v. *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.*

vi. *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.*

vii. *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.*

viii. *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.*

ix. *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.*

x. *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.*

xi. *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.*

xii. *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.*