

基于反馈的无线传感器网络 Top-K 查询处理算法

郑 瑾 吴关仁

(中南大学 信息科学与工程学院 湖南 长沙 410083)

摘 要 无线传感器网络是一种应用非常广泛的网络。Top-K 查询是无线传感器网络应用的基础,基于某种比较规则,其返回网络数据的前 K 个值。文章提出一种基于转发树结构的反馈式 Top-K 查询处理算法,该算法利用父亲节点与孩子节点的交互结果。在每一轮交互中,孩子节点根据父亲节点的反馈信息有选择性地返回部分数据,以达到减少发送数据量的目的。模拟实验表明该算法具有较好的节能效果。

关键词 无线传感器网络;反馈;Top-K 查询;转发树

中图分类号 TP393 **文献标识码** A **文章编号** 1006-8937(2009)09-0004-03

A feedback-based Top-K query processing algorithm in wireless sensor networks

ZHENG Jin, WU Guan-ren

(School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, China)

Abstract: Wireless sensor networks (WSNs) can be used in many applications. Top-K query is the base of many applications in WSNs, which returns the earlier K data items of the data set produced by WSNs under some rules of comparison. With the help of forwarding tree, this paper proposes a novel feedback-based Top-K query processing algorithm, which makes the best of the results produced in the course of interaction between a parent node (PN) and its children nodes (CNs). During every round of interaction, CNs return some selective data according to the feedback received from PN for reducing the amount of data packets to be transmitted. Simulation results show that this algorithm is energy effective.

Keywords wireless sensor networks; feedback; Top-K query; forwarding tree

1 引言

近年来,针对无线传感器网络的研究及应用都得到了飞速发展,已越来越受到工业界、商业界、学术界等的重视^[1]。Top-K 查询要求返回网络产生的数据中的前 K 个结果^[2],支持很多应用场景。如管理者希望知道火灾监测时网络中哪些地方的温度最高,以此为依据进行及时、有针对性地火灾预防;又如通过网络监测来确定某种鸟类的主要栖息地^[2],生物学家以此为依据来跟踪研究该鸟类以达到保护生态的目的。

Top-K 查询按照对结果要求是否精确分为精确查询和近似查询。针对精确型查询,文献[2]基于 TAG 思想,每个孩子节点(假如有 N 个)返回 K 个数据,父亲节点处理 N*K 个数据后得到新的 Top-K 结果,并将此结果采用相同方法返回给上层父亲节点,如此直到基站(Sink)。该算法简单,但产生的(N-1)*K 个冗余数据浪费了大量能量。文献[3]提出了两种精确查询算法,一种是采用集中式处理思想,传感器节点返回所有数据给基站,基站对这些数据集中处理后得出最终结果。该算法简单并能快速响应后续查询,但对于单次型 Top-K 查询是无效的。另一种

算法(FILA)在传感器节点中设置过滤器以有选择地返回数据,查询结果的正确性取决于过滤器的优劣和更新情况。该算法要求基站的通信半径覆盖整个网络以方便更新过滤器,并且基站保存所有过滤器信息,过滤器需要基站收集所有节点的部分数据来完成初始化而代价太高。

针对近似型查询,文献[2]利用线性规划策略建立针对精确度的目标函数以及针对能量消耗的约束条件,最后通过数据采样来确定最终结果,该算法因没有考虑所有传感器的数据而只能得出近似结果,在要求结果精确时耗能太大。

针对多次 Top-K 查询,文献[4]提出算法 FbUA, Sink 采集所有节点的样本数据为每个节点初始化过滤器并逐步更新过滤器, Sink 根据全局信息选择部分节点返回最终结果。该算法的过滤器初始化及更新代价太高而不适用于单次 Top-K 查询。文献[5]提出的算法首先基于 TAG 思想执行第一个 Top-K 查询,对于后续 Top-K 查询,根据所需结果是否在历史查询结果中来判断是否向网络索取所缺数据。此算法的能量有效性取决于对历史查询的冗余结果的利用率。

上述精确型查询算法由于冗余数据量太大而浪费大量的能量,而近似查询算法的查询结果精确性无法保证,因此无法满足要求查询单次且精确的情况。本文提出一种基于反馈的 Top-K 精确查询处理算法,此算法在网内分

收稿日期 2009-02-15

作者简介:郑瑾(1970—),女,四川巴中人,博士研究生,副教授,主要研究方向:无线传感器网络数据管理。

布式进行,利用节点的本地计算能力来减少数据的传输。

2 网络模型

网络由基站和 n 个传感器节点组成,传感器节点都为同构的随机部署的普通节点。假设网络中维持一棵转发树(Sink 为根),用于查询包的分发和数据的返回。每个传感器节点只拥有局部信息,即孩子节点与父亲节点的部份信息。用户从 Sink 发出 Top-K 查询请求,要求返回网络某一时刻的 Top-K 结果。

3 基于反馈的 Top-K 查询处理算法

基站将接收到的查询请求包(QRP)沿着转发树分发到网络中,数据按反向路径返回。此过程包括 QRP 分发阶段和反馈式数据收集阶段。

3.1 QRP 分发

基站首先将接收到的 QRP 沿着转发树分发给所有孩子节点,接收到 QRP 的节点判断其缓存的数据中是否包含以其为根的子树的 Top-K 结果,是则与父亲节点进行交互,即进入反馈式数据收集阶段,否则节点将往下一层孩子节点分发 QRP,直至叶子节点,叶子节点将直接进入反馈式数据收集阶段。

QRP 包含三方面信息:K 值,父亲节点期望每个孩子节点返回的满足条件的数据的个数 $NofReq$;父亲节点当前所拥有的数据集中的第 K 个数据值 $KthVal$ 。

3.2 反馈式数据收集

在此阶段中,节点首先根据某种比较规则以及父亲节点反馈的 $KthVal$ 返回数据集中的前 $NofReq$ (小于 K)个数据。如果满足条件的数据的个数 C 小于 $NofReq$,则返回该 C 个数据,并附加空消息用于告诉父亲节点将不参与后续交互。父亲节点对收集到的数据排序,并将更新过的 $KthVal$ 反馈给需要参与下一轮交互的孩子节点以进行新一轮数据收集。若干轮交互后,当没有孩子节点需要再访问时,父亲节点完成以自己为根的子树的 Top-K 查询,进而将收集的 Top-K 数据交互式返回给上一层父亲节点,如此直到 Sink 完成整个网络的 Top-K 查询。

QRP 分发和返回数据处理算法由父亲节点来执行,算法如图 1 所示。

```
GetTopK( K )
{
    KthVal $\leftarrow$  $-\infty$ ; NofAcc $\leftarrow$  N; NofReq $\leftarrow$   $\lceil K/N \rceil$ ;
    while (NofAcc > Th)
    {
        SendReq( K, NofReq, KthVal );
        Update( NofAcc, KthVal );
    }
    GetAllVal( KthVal );
}
```

图 1 QRP 分发和返回数据处理算法

其中 $NofAcc$ 为需要交互的孩子节点个数, N 为孩

子节点总个数, $SendReq$ 发送数据请求给需要交互的孩子节点, $Update$ 根据孩子节点返回信息更新 $NofAcc$ 和 $KthVal$, $GetAllVal$ 收集需要交互的孩子节点的所有满足条件的数据。孩子节点返回数据算法如图 2 所示。

```
ReturnData()
{
    if (不要求返回全部满足条件的数据)
    {
        C  $\leftarrow$  NofValid();
        if (C < NofReq)
            Return( C, KthVal );
        EMsg();
    }
    else
        Return( NofReq, KthVal );
}
else
    ReturnAll( KthVal );
}
```

图 2 孩子节点返回数据算法

其中 $NofValid$ 计算上次交互后剩余数据中满足条件的数据的个数, $Return$ 返回 $NofReq$ 个或 C 个满足条件的数据, $EMsg$ 将向父亲节点发送空消息, $ReturnAll$ 表示返回所有满足条件的数据。

当 Top-K 结果不均匀分布在孩子节点中时,由于结果集中的孩子节点需要参与较多轮次的交互,所以结果收集缓慢,这将增加延迟和能量消耗。为了优化这种情况,算法设置阈值 Th ,当父亲节点需要访问的孩子节点个数小于 Th 时,则通知那些孩子节点在下轮交互中返回所有满足条件的数据,并终止交互,这将减少交互的轮次从而降低查询延迟和能量消耗。

发送空消息给父亲节点以退出后续的交互是因为随着交互的进行,父亲节点反馈的 $KthVal$ 值(Top-K 中最小的数据)越来越大,孩子节点中本轮不满足条件的数据不可能在下一轮满足条件而返回,这样将减少剩余数据不必要的传输和交互的轮数以减少能量消耗。自底向上进行的反馈式数据收集过程如图 3 所示,孩子节点完成以自己为根的子树的 Top-K 结果收集才与父亲节点进行交互。

如图 3 所示,假设用户执行 Top-7 查询,阈值 Th 为 2, $S1, S2, S3, S4$ 为 P 节点的孩子节点。第一轮交互如图 A 所示,每个孩子节点向 P 返回 2 个数据。在第二轮交互中(如图 B 所示), $S4$ 接收到反馈数据 80 后将数据 82 返回,其他孩子节点由于没有满足条件的数据而返回空消息以退出后续交互。在第三轮交互中,需要交互的孩子节点个数为 1,小于阈值 Th ,则 $S4$ 在接收到反馈数据 81 后返回所有满足条件的数据并终止交互,此时 P 完成以自己为根的子树的 Top-7 查询。在此层的交互中,基于反馈的算法传输的数据包个数为 $2*4+4+1+1$ 共 14 个,而基于 TAG 思想的算法为 $4*7$ 共 28 个。

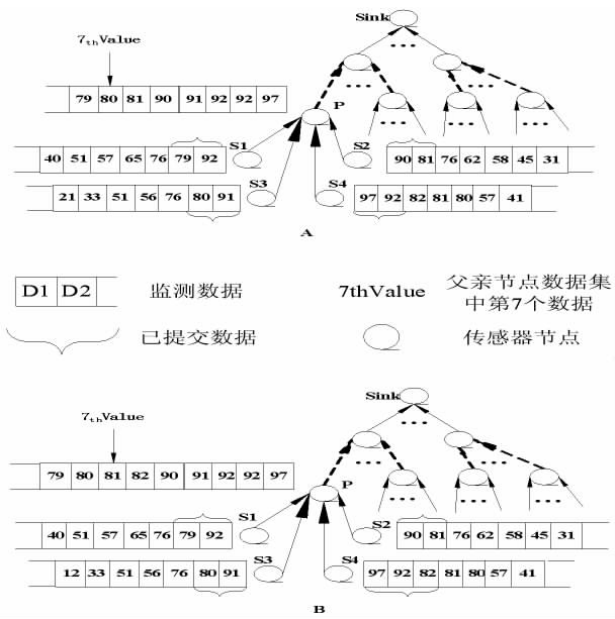


图 3 反馈式数据收集过程

4 性能分析与模拟仿真分析

本文基于 OMNET++模拟节点数为 200、500、1000 的三种规模的网络，选取正整数的大小比较作为比较规则，K 值取 50，节点随机产生一些数据，模拟使用数据包个数来衡量能量的消耗。

基于反馈的算法在反馈的同时增加了额外的数据包，但如果 Top-K 数据在孩子节点中分布比较均匀，则在首轮交互后就完成了大部分的 Top-K 数据收集，只需再进行少数的交互就能完成整个查询，算法较基于 TAG 的算法更具能量有效性。当 Top-K 数据集中在一棵子树上时，基于阈值 T_h 的优化可以减少延迟和能量的消耗，如图 4 所示。算法的交互过程将产生额外的延迟，如图 5 所示。

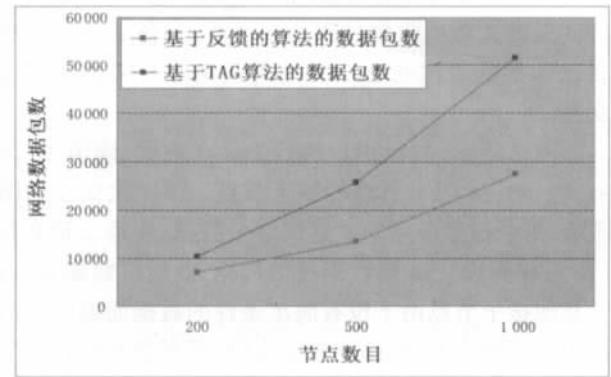


图 4 不同网络规模下的数据包数 (K=50)

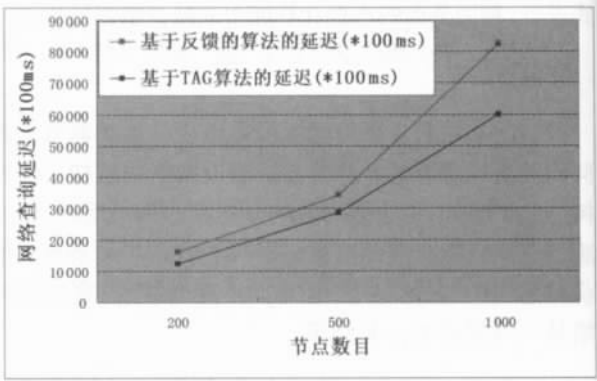


图 5 不同网络规模下查询的延迟 (K=50)

5 结语

能量问题是 WSNs 中影响网络生命期的重要问题。本文提出了一种基于反馈的 Top-K 查询处理算法，模拟结果表明该算法相比基于 TAG 思想的算法和集中式处理算法都是能量有效的，且能返回精确结果，但反馈式交互过程带来了较大延迟。

实际中用户可能提出连续查询（要求返回某个时间段的 Top-K 结果），怎样快速并能量有效地处理连续查询是以后研究工作的一个方面。已有研究工作针对多次查询主要分两个阶段处理，在第一阶段中执行第一次查询，此阶段产生较大的数据冗余，在第二阶段中，通过挖掘历史查询结果来得到后续查询的 Top-K 结果。优化第一阶段的数据冗余是以后研究的另一主要方面。

参考文献：

[1] 任丰原, 黄海宁, 林闯. 无线传感器网络[J]. 软件学报, 2003, 14(7):1282-1291.

[2] A. Silberstein, R. Braynard, C. Ellis, K. Munagala, and J. Yang. A sampling-based approach to optimizing top-k queries in sensor networks[C]. In ICDE, 2006:68-79.

[3] M. Wu, J. Xu, X. Tang, and W. C. Lee. Monitoring top-k query in wireless sensor networks[C]. In IEEE ICDE, April 2006: 143-146.

[4] 赵志滨, 李斌阳, 姚兰, 等. 一种基于过滤器的无线传感器网络复杂查询优化算法[J]. 东北大学学报, 2007, 28(7):965-968.

[5] Qunhua Pan, Minglu Li, Min-You Wu. History-sensitive Based Approach to Optimize Top-k Queries in Wireless Sensor Networks[M]. Springer Berlin Heidelberg, 2006.