

一种基于超节点和能量优先的 无线传感器网络的高效查询算法

胡 华^{1,2}, 苗媛媛¹, 庄 毅¹

(1 浙江工商大学 计算机与信息工程学院, 浙江 杭州 310018; 2 杭州电子科技大学, 浙江 杭州 310018)

摘 要: 文中提出一种基于超节点和能量优先的无线传感器网络的高效查询算法. 该算法包括传感器节点的层次聚类算法及基于能量代价模型等支撑技术, 主要解决了以下两个问题: (1) 数据如何从传感器节点传送到汇聚节点; (2) 通过对传感器节点进行聚类, 形成超节点, 使得在查询过程中减少对无关节点的访问. 实验表明该算法在提高无线传感器网络查询效率的情况下, 延长网络的使用寿命.

关键词: 无线传感器网络; 高维查询; 超节点

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2011)03-0115-04

A Hierarchical Clustering and Energy Efficient Query Algorithm over Wireless Sensor Network

HU Hua^{1,2}, MIAO Yuan-yuan¹, ZHUANG Yi¹

(1 College of Computer Science and Information Engineering, Zhejiang Gongshang University,
Hangzhou 310018, China; 2 Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: In this paper we proposed a efficient algorithm based on hierarchical clustering, this algorithm include hierarchical clustering algorithm, energy-distance algorithm. We propose to solve two problems: (1) how to send data from sensor node to sink node; (2) how to implement hierarchical clustering algorithm to produce super sensor nodes. Visit frequency to unrelated nodes is decreased. Both improve query efficiency and the network lifetime.

Key words: wireless sensor networks; high-dimensional queries; super sensor node

1 引言

无线传感器网络中数据的存取主要是研究感知数据的存储和访问机制, 关键是确定感知数据在传感器网络中的存放位置, 即如何将数据存放在适当的位置, 查询如何路由到存放位置并获取相关的数据. 传感器节点的能量有限, 存取策略的选取对存取代价, 节点能量的消耗, 以及整个网络的寿命起决定性作用.

文献[1-5]对 WSN 的数据存贮及索引技术做了研究. 研究表明, 传感器网络中的数据存储与索引

技术主要需要考虑以下几个因素: 良好的存储能力, 能够利用有限的资源存储大量的感知数据; 减少算法的复杂度, 有效地支持对数据的访问; 存储过程中通信代价小.

近年来, 无线传感网络数据查询处理技术得到了广泛的研究. TinyDB 具有一套完整的数据查询处理技术. Cougar 系统的查询处理方法, 提出了网内查询处理的思想, 给出了相应的查询处理器的结构, 并讨论了查询语言、多查询优化、catalog 管理等问题, 但是没有给出具体解决方法. 由于传感器网络本身的特点, 所以其查询的表示以及处理过程均有

收稿日期: 2010-05-04; 修回日期: 2010-07-12

基金项目: 国家自然科学基金项目(60873022, 60903053); 浙江省自然科学基金项目(Y1080148, Y1090165); 浙江省科技厅重大科技项目(2008C13082); 浙江工商大学青年人才基金重点项目(Q09-7)

其特殊性,而且最重要的是节省能量.

2 查询算法

2.1 符号及定义

表 1 文中使用的符号及其定义

符号	含义
M	传感器节点集
N	sink 节点集
SN_i	第 i 个 sink 节点
sN_i	第 i 个传感器节点
UN_i	第 i 个用户节点
SN_{ij}	第 i 个 sink 节点产生的第 j 个节点
K	与存储查询数据的传感器节点距离最近的节点数目
SSN_j	第 j 个超节点
Q	查询向量
r	查询半径
E_{init}	单位数据从节点 i 传送到节点 j 所消耗的能量
E_{ij}	单位数据从节点 i 传送到节点 j 所消耗的能量
R_i	第 i 个节点的当前能量

定义 1 无线传感器网络 (Wireless Sensor Network, WSN) 表示为一张图 $G = (V, E)$. 其中 V 表示节点, E 表示对应节点间的带宽. 节点 V 由用户节点 (User Node, UN)、汇聚节点 (Sink Node, SN)、传感器节点 (Sensor Node, sN) 组成. 形式化表示为: $V = UN + SN + sN$.

定义 2 查询超球 (Query hypersphere) 是指在多维空间以高维查询向量 Q 为圆心, r 为查询半径所形成的一个查询搜索区域 $O(Q, r)$.

定义 3 相似查询 (Similarity query) 是指在多维空间与查询向量 Q 的距离小于或等于查询半径 r 即 $d(SN_{ij}, Q) \leq r$ 的传感器网络中传感器节点的有效数据.

定义 4 超节点 (Super Sensor Node, SSN) 表示由一些具有相关性的传感器节点聚类后所形成的父查询节点, 记为 $SSN_j(sN_1, sN_1, \dots, sN_i)$, 其中 j 为第 j 个超节点, i 为具有相关性的传感器节点个数且 $i \in [1, m]$.

2.2 算法

在无线传感器网络中, 无线传感器网络节点能量有限, 因此需要考虑节点的能量消耗, 以及整个网络的能量均衡使用^[6]. 查询过程由汇聚节点发送查

询消息, 找到用户需要的数据存放的传感器节点, 选择汇聚节点与目的传感器节点之间的最优路径来进行数据传送. 本文中提出的聚类算法和能量代价模型主要是针对上述过程来进行优化的. 聚类算法可以快速确定存放用户需要的数据的传感器节点, 减少对无关节点的访问, 加快查询的效率. 能量代价模型 (如图 1) 在源节点与目的节点之间选择最优路径, 代价与节点之间的距离成正比, 节点之间代价越小, 该节点被选择的概率就越小. 能量是传输路径的剩余能量, 路径剩余能量越小, 则该路径被选择的概率就越小, 本算法在提高查询效率的情况下, 提高了网络的寿命和保证了网络的公平性.

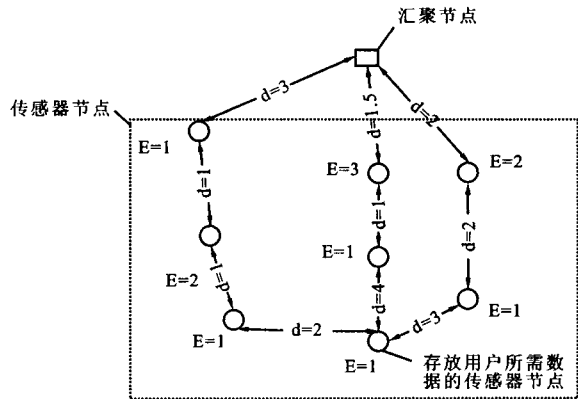


图 1 能量代价模型

能量代价模型通过对存放用户需要数据的传感器节点与汇聚节点之间的路径进行优化分析, 把能量消耗控制到较低程度, 同时选择能量剩余较多的路径, 这样就避免了过早的出现能量空洞, 延长传感器网络的寿命.

路径的选择分为几步, 首先就是要确定源节点与汇聚节点之间的路径, 在确定了路径之后, 根据, 代价与节点之间的距离成正比, 节点之间代价越小, 该节点被选择的概率就越小. 能量是传输路径的剩余能量, 路径剩余能量越小, 则该路径被选择的概率就越小, 传送代价与路径剩余能量计算出路径选择概率, 概率越大的路径被选择的可能性就越大.

路径的建立: 汇聚节点发送查询信息到目的节点, 当某个节点距离发送信息的节点更近, 而距离目的节点更远时, 该节点转发该信息, 否则丢弃该信息, 并且记录节点之间通信所消耗的能量, 以及节点的剩余能量. 其中 i 为发送信息节点与目的节点之间的点, a 为发送查询信息的节点, b 为目的节点. 式 (1) 描述了路径建立的过程.

$$i \in \{i \mid \text{dis}(a, i) < \text{dis}(i, b)\} \quad (1)$$

Path(a, b) 为从 a 到 b 之间的所有有效的路径

集合, i 为网络中任意一点, k 为路由表中 i 的下一跳节点的备选节点, P_k 为路由表中下一节点为 k 的选择概率, E_k 为节点 i 与节点 k 之间通信所消耗的能量, R_k 为节点 k 的剩余能量. 式(2)是节点 i 的下一跳节点为 j 的选择概率.

$$P_k = P\{\text{Min}(E_k) \& \& \text{Max}(P_k)\} \quad (2)$$

式中节点 k 被选择的概率与节点 i 与 k 之间的通信能量消耗和节点 k 的剩余能量, 这两个条件是相互独立的所以得出式(3).

$$P_k = P\{\text{Min}(E_k)\} \times P\{\text{Max}(R_k)\} \quad (3)$$

算法1通过综合考虑通信路径上的通信代价和节点的剩余能量来平衡传感器网络的能量消耗, 动态的选择传送节点, 最大限度的延长传感器网络的寿命.

算法1 基于能耗代价模型与最近邻的 sink 节点负载均衡算法

输入: n 个 sensor node 产生的高维数据

输出: sink 节点获得的用户需要有效数据

(1) 确定用户所需信息存放节点 i ;

(2) 统计节点 i 的下一跳备选节点集合 $\text{Next}_i[k]$;

(3) 则下一条节点的 j 被选中的概率为 P_{ij} , E_{ij} 为节点 i 与节点 j 之间的通信能量消耗, R_j 为节点 j 的当前剩余能量;

(4) 根据式(3), 计算出 P_{ij} , P_{ij} 与节点 i 和节点 j 之间的通信能量消耗成反比, 与节点 j 的剩余能量成正比, 所求出的值即为节点 j 被选中的概率, P_{ij} 的值越大, 节点 j 被选中的概率就越大.

传统的无线传感器网络对数据的查询, 需要逐个访问所有与查询相关的传感器节点, 这样大大降低查询效率. 对传感器节点中的数据事先采用聚类算法, 可以快速定位到存放用户所需数据的传感器节点, 减少对无关传感器节点的访问, 提高查询效率. 如图2所示, 如果对查询相关的传感器节点进行聚类, 形成超节点, 当对超节点所属的传感器节点进行数据查询时, 只需要访问超节点, 通过对超节点的访问, 可以判断是否需要继续向下访问, 缩小搜索范围, 大大减少节点的访问次数, 从而有助于减少对传感器网络的访问量, 减少能量消耗.

算法2 基于超节点的传感器节点聚类

输入: n 个传感器节点产生的高维数据

输出: m 个超节点存储的高维数据

(1) 对于任意两个传感器节点 SN_i 和 SN_j ;

(2) 若这两个节点产生的高维数据存在相

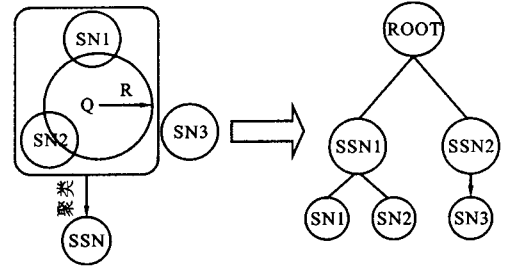


图2 传感器网络层次模型

关性;

(3) 则合并 SN_i 和 SN_j , 构成一个超节点 SSN_i .

算法实现过程如图3所示. 首先用户的查询通过 sink node 发送到无线传感器网络中, 对无线传感器网络内部传感器节点进行聚类, 形成了超节点. 组织好的数据信息通过基于能耗代价模型与最近邻的 sink 节点负载均衡算法, 选择 sink 节点把有效数据信息传送给用户.

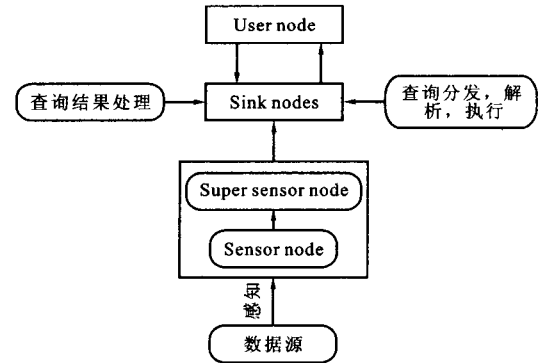


图3 查询框架

算法3 查询算法

输入: 用户提交的查询请求

输出: 用户需要的查询结果

(1) 将用户的查询请求发送到汇聚超节点 SSN_i ;

(2) 根据超节点快速对应汇聚节点;

(3) 根据确定的汇聚节点得到用户所需信息存放的传感器节点 i ;

(4) 在传感器节点上进行相似查询;

(5) 把用户所需信息从节点 i 传送到汇聚节点, 最终传送到给用户.

3 实验分析

针对影响高维相似查询操作算法的各种参数, 文中主要进行了三组模拟实验. 实验结果证明, 该算法在减少网络通信开销、增加 I/O 和 CPU 并行、降低响应时间方面具有较好的性能. 所有实验的模拟

运行环境为局域网(可以看作是一种传感器网络)。

文中采用两组数据作为测试数据:(1) UCI 提供的颜色直方图数据作为实验数据,它包含了从 Corel 图片库提取 68040 个 32 维的颜色直方图特征,每一维的值的范围都在 0 和 1 之间。(2) 计算机随机产生的 1 000 000 个 100 维的均匀分布的合成数据,其中每一维值的范围都在 0 和 1 之间。实验表明,在向量总数一定的情况下,随着传感器网络连接速度的增加,结果向量传输时间占总响应时间百分比都在减少。如图 4、图 5 所示。

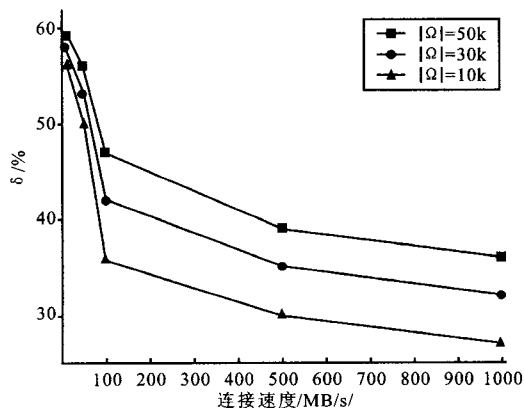


图 4 δ 与传输速率

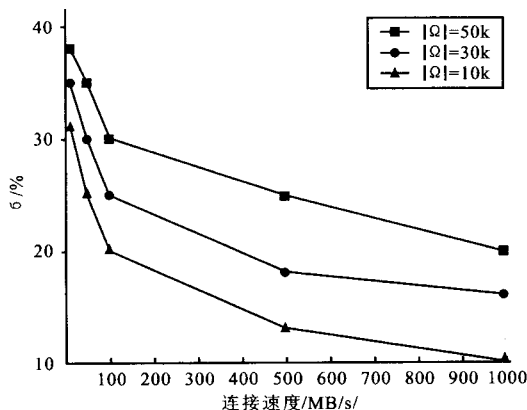


图 5 σ 与传输速率

实验对相似查询中的数据量、半径及传感器节点个数(h)对性能加速比的影响做一个评估。加速比表示为采用传统查询方式所需的时间与文中方法查询的时间的比值,用式(4)表示:

$$\text{Speed up} =$$

$$\frac{\text{The total time for conventional query over sensor network}}{\text{The total time for the query based on our approach}}$$

(4)

从图 6 看出,随着数据量的增加,其查询加速比也随之提高。当数据量为 100 000 时,其加速比小于 1。这是因为对于数据量不大的基于传感器网络的相似查询,其网络传输代价往往大于执行节点上的

CPU 距离计算代价,因此它非常适合海量高维数据的相似查询。图 7 显示了半径对加速比的影响。假定在数据量和传感器节点个数一定的条件下,当半径增加时,其对查询加速比的影响不是很大。从图 8 看出,当传感器节点个数(h)增加时,相似查询的性能加速比在提高,但提高的幅度较缓慢。这是因为在传感器节点个数增加的同时,却增加了从数据节点发送到执行节点的代价,会抵消一部分性能的提高。

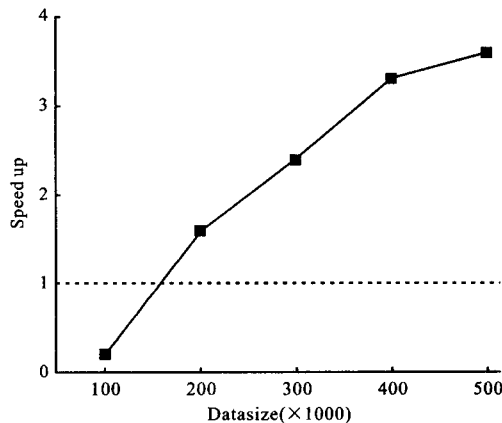


图 6 数据量对加速比的影响

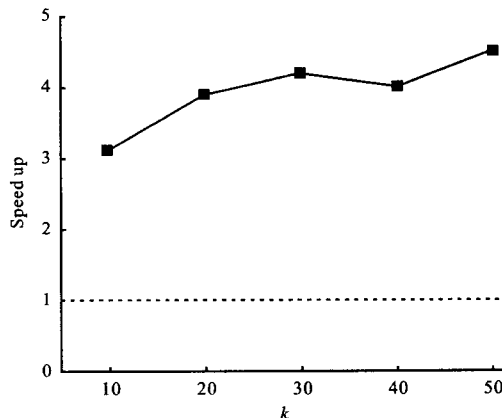


图 7 半径对加速比的影响

4 结束语

文中提出了一种基于树聚类的无线传感器网络的高效查询算法。该算法通过对传感器网络中的传感器节点进行聚类,形成超节点,来减少对传感器节点的访问,从而提高查询效率。同时,综合考虑了能耗代价和访问代价提出了基于能耗代价模型与最近邻的 sink 节点负载均衡算法,在查询的过程中在通信代价和能量代价中得到一个平衡。实验表明该算法能有效地提高无线传感器网络的数据查询效率。

此外,在本算法的基础上可以进一步考虑对传

(下转第 122 页)

实际应用也说明开发人员可以在版本计划阶段考虑不同构件的可靠性分配。

5 结束语

基于构件的敏捷软件开发过程是一个增量开发、持续测试的过程,没有独立的测试阶段且受构件固有特性的影响,传统构件软件系统可靠性模型目前还无法使用到敏捷开发过程中。因此,开发基于敏捷开发过程和测试实践的新的构件化软件可靠性估计模型是有一定价值的。敏捷实践是测试驱动的开发方法,在开发过程中测试用例是不断调整的,测试用例揭示错误的能力和改错能力是随着开发活动提升的,所以应考虑测试充分度随开发周期的变化,这将在后续工作中考虑这一变化。

参考文献:

- [1] 周国强,黄达明,曾庆凯. 基于安全属性分解的构件安全性量化和计算[J]. 微电子学与计算机,2008(10): 43—46.
- [2] 张驰. 软件构件匹配技术研究[J]. 微电子学与计算机,2009(3): 109—112.
- [3] 潘悦,沈备军. 基于构件的敏捷软件开发方法[J]. 计算

机工程,2005,31(15):68—70.

- [4] Yi Ding, Nannette Napier. Measurement framework for assessing risks in component—based software development [C] // Proceedings of the 39th Hawaii International Conference on System Sciences. Washington D C: IEEE Computer Society, 2006.
- [5] Arie van Deursen, Customer involvement in extreme programming[J]. Software Engineering Notes, 2001, 26(6): 70—73.
- [6] Kuppuswami S, Vivekanandan K, Ramaswamy P, et al. The effects of individual XP practices on software development effort[J]. ACM SIGSOFT Software Engineering Notes, 2003, 28(6): 1—6.
- [7] Nancy Van S, Ron Morssicarto. Taming the Embedded Tiger—Agile test techniques for Embedded software [C] // Proceedings of the Agile Development Conference (ADC'04). Washington, DC, USA: IEEE Computer society, 2004: 120—126.

作者简介:

王世安 男,(1971—),硕士,副教授,高级工程师。研究方向为算法设计与分析、计算机应用与软件。

(上接第 118 页)

传感器节点数据进行打包压缩,以期获得更好的查询效果。

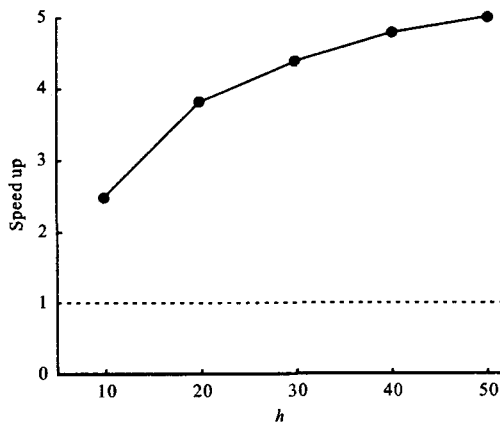


图 8 h 对加速比的影响

参考文献:

- [1] Sylvia Ratnasamy, Brad Karp. GHT: A geographic hash table for data centric storage [C] // The 1st ACM Int'l Workshop on Wireless Sensor Networks and Applications (WSNA 2002), Atlanta, USA: ACM, 2002.
- [2] Deepak Ganesan, Deborah Estrin, John Heidemann. Dimensions: why do we need a new data handling architecture for sensor networks[J]. ACM Sigcomm Com-

puter Communication Review, 2003, 33(1): 143—148.

- [3] A Deshpande, C Guestrin, W Hong, et al. Exploiting correlated attributes in acquisitional query processing [C] // IEEE Int'l Conf on Data Engineering (ICDE), Tokyo, Japan: IEEE, 2005.
- [4] Peter Desnoyers, Deepak Ganesan, Prashant Shenoy. TSAR: A two tier sensor storage architecture using interval skip graphs [C] // ACM Conf on Embedded Networked Sensor Systems (SENSYS'05), San Diego, USA: ACU, 2005.
- [5] Raghavendra C S, Krishna M Sivalingam, Taieb Zhati, et al. Wireless Sensor Networks [M]. Berlin: Kluwer Academic Publishers, 2004: 185—252.
- [6] 亢晓丽,王换招,张鹏. EERPLS: 一个大规模传感器网络中的高能效率路由协议[J]. 微电子学与计算机, 2009(5): 242—247.

作者简介:

胡 华 男,(1964—),教授,博士。研究方向为移动计算、软件工程。

苗媛媛 女,(1984—),硕士。研究方向为无线传感网络的数据管理。

庄 毅 男,(1978—),博士,副教授。研究方向为数据管理。