

一种可变尺度的 VANETs 无结构数据聚合算法

冯 诚, 李治军, 姜守旭

(哈尔滨工业大学 计算机科学与技术学院, 150001 哈尔滨, fengcheng7@126.com)

摘 要: 在 VANETs 的应用中, 为了满足应用目的节点对距其远近不同的节点对数据精度的不同要求, 聚合算法需要具有能够变尺度调节的能力, 形式化地定义了 VANETs 单节点上的可变尺度数据聚合问题并证明了该问题是 NP-难的, 同时提出了基于贪心思想的单节点变尺度数据聚合近似算法 ESAS. 模拟实验表明, 在保证可变尺度数据质量前提下, ESAS 与现有的考虑数据精度的聚合算法 (FBAS) 相比能显著减少传输数据量.

关键词: 车用自组织网络; 数据分发; 数据聚合; 可变尺度

中图分类号: TP393; TP311

文献标志码: A

文章编号: 0367-6234(2011)07-0096-06

An adjustable structure-free data aggregation approach on VANETs

FENG Cheng, LI Zhi-jun, JIANG Shou-xu

(School of Compute Science and Technology, Harbin Institute of Technology, 150001 Harbin, China, fengcheng7@126.com)

Abstract: In-network data aggregation is a useful technique for applications that require periodic dissemination of information to a large region. For applications on VANETs, destination node requires different degrees of data accuracy according to the distance from it to source node that generates data. Therefore, data aggregation scheme should adjust the aggregation factor to satisfy the different requirements of data accuracy from different applications. Based on this, we formalize adjustable data aggregation on one node on VANETs problem, and prove that it is NP-Hard, and we also propose an adjustable data aggregation approximate algorithm ESAS based on greedy. Simulation experiments make it clear that ESAS has a sharp decrease on the amount of communication data compared with that based on fuzzy logic aggregation scheme (FBAS).

Key words: VANETs; data dissemination; data aggregation; adjustable

车载自组织网络 (VANETs) 在智能交通系统中具有广泛应用前景, 成为近年来研究热点. 智能驾驶和线路导航是智能交通系统中 2 个重要应用^[1]. 对于这 2 个应用来说, 智能车需要通过 VANETs 周期性收集大量数据. 智能驾驶应用需要的数据是邻近区域内车辆较精确的速度信息, 而线路导航应用所需要的数据是大区域的可以用于提取路况信息的较粗糙的速度信息. 因此应用需要根据实际情况得到不同精确度这就要求数据聚合算法可以根据实际的数据情况调节聚合决策因子使其满足不同精度的数据需求.

对于 VANETs 本身来说, 它是车辆之间自组

织形成的无线网络. 它具有 3 个特点: 1) 动态变化频繁的网络拓扑结构; 2) 有限的网络资源; 3) 有限的节点存储能力. VANETs 本身的特点使得基于传感器网络的数据聚合算法不适合于 VANETs 环境. 在近几年, 国内外学者对于 VANETs 上的数据聚合模式已经做了一些研究. Wischhof 等^[2]提出 SOTIS 交通信息系统. 它的数据聚合是基于路段划分的, 因此数据的精确度是无法保证的. TrafficView^[3]提出数据聚合算法的目标是得到周围车辆的估计视图, 它的精确度测度是基于聚合数据的估计的位置而不是基于平均速度的偏差, 因此, 无法表明平均速度这一聚合值的精确度. 在文献[4]中提到的数据聚合方法是基于模糊理论的. 这一方法是根据具体的数值相似性进行聚合. 但该方法没考虑应用对数据精度

收稿日期: 2010-05-12.

基金项目: 国家自然科学基金资助项目 (60803148, 60973124).

作者简介: 冯 诚 (1986—), 女, 硕士研究生;

姜守旭 (1968—), 男, 教授, 博士生导师.

的不同需求,当进行大范围数据分发时,需要传输的数据量还是很大的.文献[5]是对空闲停车位信息的聚合,它是根据地理信息设定聚合层次区域,进行分层聚合.它处理的数据是空闲停车位的数量,是离散数值,而本文处理的数据为速度值,是连续数值,他们之间有着本质的区别,因此文献[5]提出的概率聚合算法在这里是不适用的.文献[6-7]研究的应用场景是带有路边设备的,解决的是车与路边设备之间进行信息交换问题,车与车之间的网内数据聚合问题得到了简化,但同时路边设备的引入也导致了额外的开销.

现有的VANETs上的各种数据聚合算法是针对不同的应用场景下不同种类的应用数据提出的.在无路边设备的应用场景下,对平均速度值进行数据聚合的算法主要有2类:1)基于由地理信息提取出的层次结构进行聚合.这样的聚合方法使得传输的数据量指数级减少,但聚合产生的误差无法保证;2)根据数据的相似程度进行聚合.这种聚合方法使得聚合结果引入的误差小,但是传输数据量的减少情况是和实际数据分布情况相关的,因此,聚合使得数据量的较少程度是未知的,不适用于大范围的数据分发.因此,本文对于整个网络上的数据聚合问题进行了形式化定义和描述,同时将该问题划分成单个节点上的聚合子问题进行了形式化定义并证明了该问题为NP-难的,提出了解决该问题的近似算法(ESAS),并通过模拟实验与基于模糊逻辑的聚合算法(FBAS)进行比较,得出ESAS算法可以达到数据精度的要求并且数据的分发率可以达到90%以上.

1 问题描述

1.1 数据对象

1) 原子数据.智能车在 t 时刻产生的原子数据为 $d_{\text{atom}} = (v, t, a_{\text{ID}}, c_{\text{ID}})$.其中: v 为车在 t 时刻采集的瞬时速度值; a_{ID} 为车在 t 时刻所在路段的ID; c_{ID} 为车的ID,它属于 C , C 是所有车形成的集合.

2) 聚合数据.它是由多个原子数据聚合得到的. $d_{\text{agg}} = \{\bar{v}, (v_{\min}, v_{\max}), nc, LA, (t_s, t_e)\}$.其中: $\bar{v}$ 为聚合得到的平均速度; $(v_{\min}, v_{\max})$ 为原子数据的速度值区域; nc 为进行聚合的原子数据的个数; LA 为原子数据所在路段的集合; (t_s, t_e) 为原子数据的时间区域.

1.2 网络环境

1) 道路网.由道路和交叉口形成的网络,用

有向图 $G = (X, A)$ 表示.其中: X 为交叉口的集合, $X = \{x_i \mid 1 \leq x_i \leq n_x\}$; n_x 为交叉口的个数; A 为路段有向边的集合, $A = \{x_i x_j \mid x_i, x_j \in X, 1 \leq i \leq n_a\}$; n_a 为有向路段的个数.

2) 通信网.路面上行驶着的车之间通过无线介质进行通信形成的网络. R 为车载通信系统的通信半径.通信网用无向图 CG 表示. $CG = (C, E)$,其中 C 为车的集合, $E = \{c_i c_j \mid c_i, c_j \in C, d(c_i, c_j) \leq R\}$.

3) 无线通信模型.采用节点排他干扰模型.即在节点的通信干扰范围内,一个传输时间片内只能有一个节点发送数据.

1.3 问题定义

已知道路网 G ,在 $t = 0$ 时刻任意的 $c \in C$,车 c 采集自己的原子数据 $d_{\text{atom}}^c = (v, t, a_{\text{ID}}, c_{\text{ID}})$,从时刻 $t = 0$ 到 $t = T$ 对这些数据进行分发,其中: $D(c, t)$ 为 t 时刻车 c 的数据集; d 为数据,即可以是原子数据,也可以是聚合数据; $h(d)$ 为数据 d 传输的跳数,给定误差界函数为 $\text{errB}(h(d))$,这里的误差界函数是随距离递增的函数,即体现了在数据分发过程中,距离数据源节点近的目的节点需要的数据的误差要小,而距离数据源节点远的目的节点需要的数据的误差可以适当地大些.本文中的误差界函数设为 $\text{errB}(h(d)) = 0.1 + 0.05 \times h(d)$.

定义1 数据集 $D(c, t)$ 的关于误差界函数 $\text{errB}(h(d))$ 的可聚合子集划分是满足下述条件的对 $D(c, t)$ 的划分 $\{B_1, B_2, \dots, B_m\}$ 为

$$\forall B_i, B_j, B_i \cap B_j = \emptyset,$$

$$\sum_{i=1}^m B_i = D(c, t),$$

$\forall B_i \{a_{\text{ID}} \mid d = (v, t, LA, c_{\text{ID}}), d \in B_i, a_{\text{ID}} \in LA\}$ 在 G 上的诱导子图是连通的.

$$\forall B_i, \text{err}(B_i) \leq \text{errB}(\min\{h(d) \mid d \in B_i\}).$$

其中 $\text{err}(B_i)$ 为 B_i 中数据进行聚合产生的误差值.

定义2 VANETs上变尺度数据聚合问题为给定在 $t = 0$ 时刻任意的 $c \in C$,车 c 采集的原子数据 $d_{\text{atom}}^c = (v, t, a_{\text{ID}}, c_{\text{ID}})$,数据分发的结束时刻 T ,误差界函数 $\text{errB}(h(d))$,求所有车的所有时刻的 $D(c, t)$ 的可聚合子集划分,使得 $\sum_{0 \leq t \leq T} \sum_{c \in C} |D(c, t)|$ 最小.

该问题的优化目标是在整个数据分发时间段 T 内,聚合算法在满足数据目的节点对数据的误

对 g_i 对应的数据集做聚合.

将聚合数据加入新数据集 D' .

将 g_i 从 G' 中删除.

Else

从 g_i 中,选出跳数最小的数据节点.

在 G' 中删去与该节点相连的边.

End if

End while

将 FG' 中的节点添加到数据集 D' .

2.3 可变尺度数据聚合的准则设定

决策准则是判断 2 个信息元素是否可以聚合的条件. 根据 VANETs 上应用的需求, 决定 2 个原子元素是否可以聚合的准则有 2 点:

1) 聚合元素的最大误差要小于原子元素的最小误差界. 聚合元素的最大误差是聚合值与形成该聚合元素的原子元素的差值的最大值. 原子元素的最小误差界是指根据各原子元素的跳数和给定的误差界函数求出的误差界值的最小值.

2) 产生原子数据的源节点在数据生成时刻所在的位置是相邻路段的.

2.4 数据聚合

数据聚合方法是根据应用的实际需要进行定义的. 本文采用的是取平均值. 聚合过程需要计算 4 个值: 平均值 $\bar{v}$, 最大误差值 $err(\bar{v})$, 聚合值包含的原子数据的个数 $size$ 和原子数据值的范围 $[v_{\min}, v_{\max}]$.

聚合过程根据处理的数据对象类型的不同, 分为 3 种情况.

1) 原子数据 v_1 和原子数据 v_2 聚合. 其中聚合数据的平均值为 $\bar{v} = (v_1 + v_2)/2$, 聚合数据包含的原子数据的个数 $size(\bar{v}) = 2$, 聚合数据相对原子数据的最大误差 $err(\bar{v}) = \max\{|\bar{v} - v_1|, |\bar{v} - v_2|\}$, 聚合数据的原子值范围 $[v_{\min}, v_{\max}] = [\min\{v_1, v_2\}, \max\{v_1, v_2\}]$.

2) 原子数据 v_1 和聚合数据 v' 聚合. 其中聚合数据的平均值为 $\bar{v} = \frac{(v_1 + \bar{v}' \times size(v'))}{(size(v') + 1)}$, 包含的原子数据的个数 $size(\bar{v}) = size(v') + 1$, 最大误差 $err(\bar{v}) = \max\{|\bar{v} - v_1|, |\bar{v} - v'_{\min}|, |\bar{v} - v'_{\max}|\}$, 聚合数据的原子值范围 $[v_{\min}, v_{\max}] = [\min\{v_1, v'_{\min}\}, \max\{v_1, v'_{\max}\}]$.

3) 聚合数据 v'_1 和聚合数据 v'_2 聚合. 其中平均值 $\bar{v} = \frac{(\bar{v}'_1 \times size(v'_1) + \bar{v}'_2 \times size(v'_2))}{(size(v'_1) + size(v'_2))}$, 包含原子数据的个数 $size(\bar{v}) = size(v'_1) + size(v'_2)$, 最大误差 $err(\bar{v}) = \max\{|\bar{v} - v'_{1\min}|, |\bar{v} - v'_{1\max}|, |\bar{v} - v'_{2\min}|, |\bar{v} - v'_{2\max}|\}$, 原子值区域 $[v_{\min}, v_{\max}] =$

$$[\min\{v'_{1\min}, v'_{2\min}\}, \max\{v'_{1\max}, v'_{2\max}\}].$$

2.5 节点上变尺度数据聚合算法复杂度分析

变尺度数据聚合算法的输入规模是数据集的大小 n . 根据道路网 G 构造 G' 需要时间 $O(n)$, while 循环的最多次数为 $n/2$, 在图 G' 上找最大连通分量需要时间 $O(n)$, 下一步调用聚合决策算法, 聚合决策算法如下:

输入: 原子元素集 $\{d_1, d_2, \dots, d_k\}$, 误差界函数 $errB(h(d))$, 道路网 G .

输出: 0 或 1.

If $\{d_1, d_2, \dots, d_k\}$ 产生时的路段集 $\{a_1, a_2, \dots, a_k\}$.

在 G 中是不连通时.

Return 0

Else

将 $\{d_1, d_2, \dots, d_k\}$ 做聚合, 得到聚合数据 agg 计算聚合数据的偏差 $err(agg)$

If $err(agg) \leq errB(\min\{h(d_i), 1 \leq i \leq k\})$

Return 1

End if

Return 0

End if

End if

聚合决策算法的输入规模是 g_i 的节点数, 记为 k . 算法中判断 k 条边在 G 中是否连通, 需要时间 $O(k)$, 做聚合计算聚合数据误差需要时间 $O(1)$, 判断 $err(agg)$ 与 k 个数据的最小误差界的关系需要时间 $O(k)$, 算法的总时间复杂度为 $O(k)$, g_i 的节点数 $\leq n$, 因此需要时间 $O(n)$, 当做聚合时需要时间 $O(1)$, 当不做聚合时从 g_i 中, 选出跳数最小的数据节点需要时间 $O(n)$. 综上, 该算法总的时间复杂度为 $O(n^2)$.

3 模拟实验

3.1 实验设计

在实验中模拟的是哈尔滨某区域交通高峰期的场景, 如图 2 所示. 地图信息采自 Google 地图. 车的瞬时速度信息是模拟产生的.



图2 仿真实验地图

实验中采用的是离散事件模拟器. 将时间划分成时间槽. 假设数据在节点上的一次聚合和分发过程在一个时间槽内完成.

表 1 列出了仿真中用到的系统参数. 实验中选取的地图是哈尔滨市某区域, 共有 23 个交叉路口, 72 条单向路段. 将 60 辆车随机分布在道路网中. 并设定工大校门前的西大直街上有堵车现象, 即在相应几个段路平均车辆负载数为 6~8, 路段的平均车速为 12.4 km/h. 实验中将 $t = 0$ 时产生的即时速度数据在 70 个时间槽内进行分发.

表 1 仿真参数设置

系统参数	道路数/ 条(单向)	交叉路口 数/个	车辆 数/辆	通信 半径/m
值	72	23	60	250

3.2 仿真实验结果

实验中将本文的可变尺度聚合算法(ESAS)与基于模糊逻辑的聚合算法(FBAS)进行了比较分析. 其 FBAS 是基于速度差值(SpeedD)进行聚合决策的. 在本文实验中设定速度差值(SpeedD)关系函数为

$$LOW_SpeedD:f(x) = \begin{cases} 1.0, & x < 3 \\ 6 - 2x, & 3 \leq x < 6 \end{cases};$$
$$HIGH_SpeedD:f(x) = \begin{cases} 0.5x, & -1.53 \leq x < 5 \\ 1.0, & x \geq 5 \end{cases}.$$

式中 x 为 2 个原子数据速度的差值. 实验中比较了 3 个性能参数: 分发效率、数据精确度和传输数据量. 其中分发效率用于衡量到数据分发结束时刻数据在网络中分发的范围; 数据精确度用于衡量分发到各节点数据的精确情况; 传输数据量用于衡量分发数据过程中需要的网络带宽情况.

1) 分发效率. 定义平均已获得信息比例(Average known situation)为

$$AKS(t) = \frac{\left(\sum_c \left(\frac{|knowledge_c^t|}{|C|}\right)\right)}{|C|},$$

式中 $knowledge_c^t$ 为车 c 到时刻 t 为止已收集到的信息. 数据的分发效率就用 AKS 这一指标进行衡量.

分发效率比较的结果在图 3 中展现出来. 2 种算法的性能在第 15 个时间槽之前是十分相近的. 在这以后, FBAS 几乎维持在 70% 左右就不再增长了, 而 ESAS 最终增长到 90% 多. 这说明 ESAS 比 FBAS 的数据分发效果更好, 即数据分发过程中采用 ESAS 聚合算法能在网络中对数据进行更广的分发.

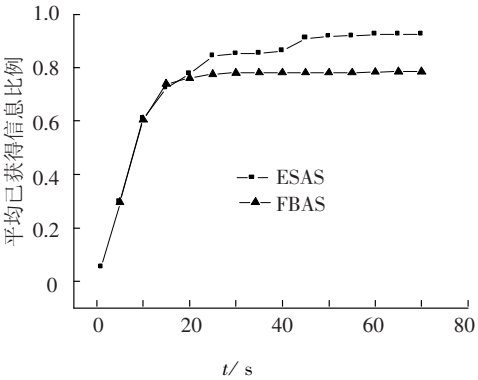


图 3 ESAS 和 FBAS 分发速度的比较

2) 数据精确度. 定义平均误差比率(Average inaccuracy percent)为 AIP , $AIP(h(d_c^T(c')))=\sum_c \frac{inaccuracy(d_c^T(c'))}{V(c')}$, 其中 $d_c^T(c')$ 为到时刻 T 车 c 收集到的关于车 c' 的速度值.

从图 4 中的曲线可以看出, 选取的误差比率界是关于跳数的线性函数. 而实际的平均误差比率一直在这一界线值之下, 说明 ESAS 可以满足给定误差界的条件. 还可以看出实际误差没有随着跳数的增加呈规律地增长. 这是因为聚合过程会有很多影响因素如可聚合数据同一时间槽出现在同一节点的几率, 即 2 个可聚合数据相遇是聚合的必要条件.

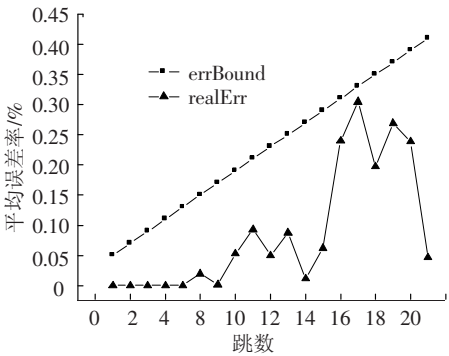


图 4 实际误差比率和给定误差比率界的比较

3) 传输数据量. 实验中本文的 ESAS 与无聚合模式和 FBAS 比较了传输的数据量. 从图 5 中可以得出, 采用 ESAS 算法的数据分发需要传输

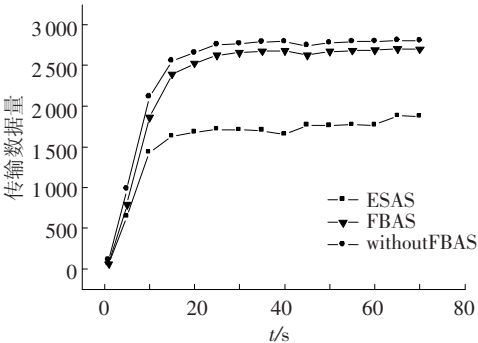


图 5 3 种模式下的数据传输量随时间的变化情况

的数据量明显小于利用 FBAS 的情况和不采用聚合模式的情形. FBAS 不能显著提高性能的原因是整个网络中所有数据的聚合是统一的标尺,没有考虑数据的精度需求是不同的这一事实. ESAS 的优势在于根据应用提出的不同精度需求,使得尽可能多的可聚合数据进行聚合.

4)道路网络规模对聚合模式性能的影响. 从表 2 中,可以看出来,随着网络规模的增大,需要传输的数据个数是显著增加的. 因此聚合模式的引入是非常必要的. 从 3 组数据上的对比可以看出,聚合模式的性能随着网络规模的增加是下降的. 因此随着网络规模的扩大,这种线性关系的数据误差界还是很难满足的.

表 2 网络规模和 ESAS 聚合模式下传输数据量的关系					
总传输数 据数	总减少传输 数据数	时间槽 数	网络规模		
			交叉口	单向道	车辆数
21 679	10 738	8	5	6	18
59 466	28 364	15	12	32	30
111 259	50 481	70	23	72	60

4 结 论

1)对 VANETs 上的数据聚合问题进行了形式化的描述,根据网络的实际情况,将核心聚焦在单节点上的数据聚合问题,并分析了该问题的复杂度是 NP - 难的.

2)设计节点上的可变尺度聚合算法. 该算法可以实现对聚合可满足条件的动态调节,使其满足应用对数据精确度的要求.

3)对 ESAS 算法与 FBAS 算法的性能进行了比较,结果表明,ESAS 与 FBAS 相比,数据的分发率更高,传输的数据量更小,该算法的可扩展性更好.

参考文献:

[1] HARTENSTEIN H, LABERTEAUX K P. A tutorial survey on vehicular ad hoc networks[J]. IEEE Communication Magazine, 2008, 46(6): 164 - 171.

[2] WISCHHOF L, EBNER A, ROHLING H, *et al.* SOTIS - a self-organizing traffic information system [C]//Proceedings of Vehicular Technology Conference The 57th IEEE Semiannual, New York, NY: ACM, 2003: 2442 - 2446.

[3] NADEEM T, DASHTINEZHAD S, LIAO Chunyuan, *et al.* TrafficView: Traffic data dissemination using car-to-car communication[J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2004, 8 (3): 6 - 19.

[4] DIETZEL S, BAKO B, SCHUCH E, *et al.* A fuzzy logic based approach for structure-free aggregation in vehicular ad-hoc networks[C]//Proceedings of the Sixth ACM International Workshop on VehiculAr InterNET Working. New York, NY: ACM, 2009: 79 - 88.

[5] LOCHERT C, SCHEUERMANN B, MAUVE M. Probabilistic aggregation for data dissemination in VANETs [C]//Proceedings of the Fourth ACM International Workshop on Vehicular ad Hoc Networks. New York, NY: ACM, 2007: 1 - 8.

[6] LOCHERT C, SCHEUERMANN B, WEWETZER C, *et al.* Data aggregation and roadside unit placement for a VANET traffic information system [C]//Proceedings of the Fifth ACM International Workshop on VehiculAr Inter-NET Workingl. New York, NY: ACM, 2008: 58 - 65.

[7] ARBABI M H, WEIGLE M. Using vehicular networks to collect common traffic data [C]//Proceedings of the Sixth ACM International Workshop on VehiculAr Inter-NET Working. New York, NY: ACM, 2009: 117 - 118.

[8] SCHEUERMANN B, LOCHERT C, RYBICKI J, *et al.* A fundamental scalability criterion for data aggregation in VANETs [C]//Proceedings of the 15th Annual International Conference on Mobile Computing and Networking. New York, NY: ACM, 2009: 285 - 296.

[9] IBRAHIM K, WEIGLE M C. Accurate data aggregation for VANETs [C]//Proceedings of the Fourth ACM International Workshop on Vehicular ad Hoc Networks. New York, NY: ACM, 2007: 71 - 72.

[10] YU Bo, GONG Jiayu, XU Chengzhong. Catch-Up: A data aggregation scheme for VANETs [C]//Proceedings of the Fifth ACM International Workshop on VehiculAr Inter-NET Working. New York, NY: ACM, 2008: 49 - 57.

(编辑 张 红)