

基于 GPS 数据的交通拥堵检测

张亚茹,赵海涛,刘南杰,黄波,李大鹏
(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

摘要:随着交通拥堵问题的日益突出,如何有效检测区域内实时的交通状态是城市交通发展中的研究热点。为了实现区域内交通状态进行有效的检测,提出了一种基于多源 GPS 数据交通拥堵的检测方法。首先,剔除原始 GPS 数据中的异常数据并且对缺失的数据进行修复,以保证 GPS 数据的合理性,对采集的原始 GPS 数据进行了预处理;其次,在保证采集的交通信息完整性的基础上,利用 GPS 数据的时空特性,使用 k -means 算法对 GPS 数据进行聚类分析,实现了不同簇下的研究区域的划分;最后,为获取每个簇的实时交通状态,采用了一种基于簇内样本数量进行随机抽样的交通状态检测方法。仿真结果表明,所提算法可以实现对区域内交通状态的有效检测且算法时效性得到了有效提升。

关键词:GPS;交通拥堵;预处理; k -means;抽样

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2017)01-0139-04

doi:10.3969/j.issn.1673-629X.2017.01.031

Detection of Traffic Congestion Based on GPS Data

ZHANG Ya-ru, ZHAO Hai-tao, LIU Nan-jie, HUANG Bo, LI Da-peng

(College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: With traffic congestion becoming increasingly serious, how to effectively detect the real-time traffic status has become a research hotspot in the field of urban transportation. For effective detection of traffic area, a detection method based on multi-source GPS data is proposed. Firstly, in order to eliminate abnormal data and ensure the rationality of data, the method makes a preprocessing of GPS data to eliminate abnormal data and repair missing data. Secondly, on the basis of ensuring the integrity of data, the method uses k -means algorithm to cluster GPS data in accordance with the temporal-spatial characteristics of GPS data, and then divides the study area into different clusters. Finally, for obtaining the real-time traffic conditions of each cluster, a sampling algorithm based on the number of samples within each cluster is used to detect the traffic conditions. Simulation results show that the proposed algorithm with simple process can effectively detect the state of traffic within the region.

Key words: GPS; traffic congestion; preprocessing; k -means; sampling

0 引言

道路拥堵早已成为一个不容忽视的问题,尤其在大城市拥堵则更为严重。随着人们生活水平的快速提高及城市化进程的不断推进,城市机动车的数量随之急剧增加,这与滞后的城市基础设施建设和城市管理水平形成了剧烈的矛盾和冲突,导致道路拥堵问题日益频繁与严重^[1]。严重的道路拥堵问题会带来很多的负面问题,比如增加燃料的消耗、浪费出行人员的时间以及带来的环境污染问题等,这些问题都会阻碍城市的发展,影响人们的日常生活。这些现象在一些大城

市则更为突出,比如北京、上海以及广州等地,每天都会因为拥堵问题而带来巨大的经济损失。因此,解决交通拥堵问题迫在眉睫。

如果人们在出行的过程中能够及时了解前方道路的拥堵情况,那么出行者就可以选择其他线路。这样不仅可以有效缓解交通拥堵,而且还可以节约出行者的出行时间。现在社会各界越来越注重对交通拥堵问题的研究,而且已经提出了很多对道路拥堵情况进行检测的技术。文献[2]介绍了智能实时路况系统建设过程中的关键技术。在之前对路况的信息采集技术主

收稿日期:2016-03-07

修回日期:2016-06-16

网络出版时间:2017-01-04

基金项目:国家自然科学基金资助项目(61302100,61471203);教育部博士点基金(20133223120002);江苏省产学研项目(BY2015011-1)

作者简介:张亚茹(1992-),女,硕士研究生,研究方向为移动通信与无线技术;赵海涛,博士,副教授,研究方向为车联网资源优化;刘南杰,博士,教授,研究方向为泛在通信、车联网、智能交通;黄波,博士,讲师,研究方向为卫星通信与导航、信号处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170104.1028.042.html>

要有波频检测、磁频检测和视频检测等^[3]。然而近年来 GPS 技术的迅速发展给交通拥堵检测带来了数据量庞大的时空数据。文中提出了一种基于 GPS 数据的拥堵检测方法,通过预处理—聚类—抽样来完成对道路拥堵的检测,该方法过程简单且具有一定的可行性。

1 数据预处理

要判断一个道路是否拥堵,可以根据行驶在这条道路上的汽车的行驶数据来判断。然而 GPS 数据的来源各式各样,而人们所需要的只是行驶在道路上的汽车数据,所以要去除那些非行驶在道路上的汽车数据。2G 和 3G 的发展促进了 GPS 设备的发展,可以采集到大量的时空数据集^[4],而且采集到的数据中还会存在一些异常数据,对于这些数据要进行一定的舍弃和修复工作。文献[5]是基于 GPS 数据对高速公路的实时路况状态进行研究。

1.1 道路上车辆行驶信息的获取

假设要获取某路段上的道路拥堵情况,就要获取该路段上车辆的行驶信息。地球上的每一个位置都有自己的经纬度,同样每一条道路也都有它自己的经纬度范围,而要获取到这些信息也是轻而易举的。同时获取到的 GPS 数据 G_i 中包括 GPS 数据源所处位置的经纬度。这样可以把被检测道路的经纬度范围看作是一个集合 R ,然后通过判断所获取的 GPS 的经纬度是否属于这个集合来判断数据源是否来自于这条道路,以确保所获得的 GPS 数据都属于这条道路,保证最后结果的可靠性^[6]。如果 GPS 的经纬度包含于 R ,则保留该条数据,否则去除这条数据。

对于上面获取的数据,还需要做进一步地处理。因为 GPS 数据来源于不同的设备,比如汽车、手机或者平板电脑等。假如当 GPS 数据来自于路上行人的手机时,而这些人正处于追赶公交车的状态,这时获取的 GPS 数据会对最终的判断结果有一定的影响。这是因为这些人的交通模式改变了,从起初的步行模式到后来的乘车模式,其中步行时期的数据会对拥堵判断造成一定的误差。那么如何来去除那些步行数据呢?首先需要根据获取到的 GPS 数据得到每一个数据源的总体运动轨迹 C 。总体来说,由于运动模式的不同,获得的运动轨迹应该是不连续且分段的。然后得到每一个轨迹的运动区域,假如该区域在道路内则保留,否则丢弃。最后再对上面保留的 GPS 数据日志进行处理,若交通模式发生改变,则求出它的每一段轨迹的平均速度,若速度小于设定值 T ,则判别为步行。这样就能去除其中的步行数据,使结果更准确。

具体处理流程如图 1 所示。

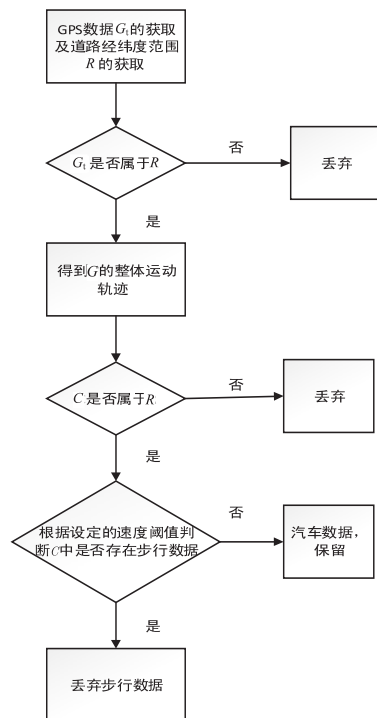


图 1 行驶在道路上的车辆信息的获取

1.2 异常数据的处理

上面获取的数据中可能会存在一些速度为负值的数据,很明显这些数据肯定是错误的,这时就要把这些数据筛选出来,然后这些负值用 0 来代替。有时由于交通事故和天气等突发性因素,采集到的车辆速度可能会偏离正常的速度范围,因为异常数据的存在在一定程度上会影响后面交通状态的检测精度,因此在数据的预处理过程中必须舍弃此类异常数据。

在数据的采集过程中经常会出现数据丢失的现象,再加上对异常数据的丢弃,这些都导致了数据的不全面性,因此需要对缺失的数据进行修复。文中采用权值平均方法,分别针对孤立的数据丢失和连续的数据丢失两种情况进行修复。具体公式^[7]如下:

$$\hat{x} = \frac{1}{w} \sum_{k=-T}^T w_k \cdot x_{t+k} \quad (k \neq 0) \quad (1)$$

其中, w_k 为权值系数; W 为所有权值系数之和; T 为需要进行数据修复的最大时间间隔。

式(1)对孤立的丢失数据 x_t 的修复。值得注意的是,与数据修复点的时间相差越远, w_k 越小。连续丢失的数据可以通过如下的二次指数平滑法进行修复:

$$\hat{x}_{t+r} = a_t + b_t r \quad (r = 1, 2, \dots) \quad (2)$$

其中, a_t 和 b_t 可以通过下式确定:

$$\begin{cases} a_t = 2Q_t^{(1)} - Q_t^{(2)} \\ b_t = \frac{\alpha}{1-\alpha} (Q_t^{(1)} - Q_t^{(2)}) \end{cases} \quad (3)$$

其中, $Q_t^{(1)}$ 为一次指数平滑值; $Q_t^{(2)}$ 为二次指数平滑值; 平滑系数 $\alpha \in (0, 1)$ 。

$$\begin{cases} Q_t^{(1)} = \alpha x_t + (1 - \alpha) Q_{t-1}^{(1)} \\ Q_t^{(2)} = \alpha Q_t^{(1)} + (1 - \alpha) Q_{t-1}^{(2)} \end{cases} \quad (4)$$

此外,可以通过交通信息的空间相似性改善数据修复的质量。

2 聚类分析

很早就有对交通模式的研究。在 1967 年,文献[8]就提出了对交通信号灯的系統控制。2002 年,文献[9]使用地图划分法将一个城市分为不同的集群来对用户的位置进行预测。文献[10]使用机器学习方法来估计交通条件和速度模式。现在在对数据的挖掘过程中提出了很多方法,而且学者们对这些方法都进行了改进。文献[11-12]使用了聚类方法对交通状态进行检测。聚类分析是数据挖掘中的一种重要方法,它可以把具有相同特征的事物聚为一个类,能够更简洁清晰地了解到整体的一个状态。

为了简单起见,这里以一段路段来进行聚类分析。获取的 GPS 中包括的参数有经度 L 、纬度 W 、时间 t 和行驶速度 v ,则取向量 $G = (L, W, t, v)$ 。以 G 为样本,根据 k -means 算法,首先确定簇的个数 k 。但是 k 往往不好确定,这也是 k -means 算法的一个局限。然后随机选取 k 个质心,初始地代表一个簇,对于剩下的样本则要根据它们与这些聚类中心的相似度,分别将它们分配给与其最相似的聚类^[13]。再计算每个所获得的新聚类的聚类中心(该聚类中所有对象的均值),不断重复这一过程,直至新的质心与原质心相等或小于指定阈值,则算法结束。 k -means 算法的最终结果是要使每个簇内的元素尽量紧凑,簇之间要尽可能分开。这里也可以根据实际情况设定每个簇的大小,比如设定半径值。该算法的一个缺点是对一些孤立点等脏数据比较敏感,因此一定要做好数据的预处理。某一时刻路段的聚类效果如图 2 所示。使用算法的公式如下:

$$V = \sum_{j=1}^k \sum_{G_i \in R} (\vec{G}_i - \vec{G}_j)^2 \quad (5)$$

其中, k 为簇的个数; G_i 为一个 GPS 数据; R 为道路的经纬度集合; V 为非质心元素与质心元素的距离。



图2 聚类效果图

3 拥堵等级判别

目前常用于衡量路段拥堵情况的参数主要有车流密度、路段平均速度。文中通过计算路段平均速度实现对交通拥堵等级的划分。根据不同等级道路的平均速度,可以将道路的拥堵状况分为 3 个等级:

(1) 畅通:城市快速路和联络线平均速度高于 50 km/h,主干道平均交通速度大于 20 km/h;

(2) 缓行:城市快速路和联络线平均速度高于 20 km/h,低于 50 km/h,主干道平均交通速度高于 10 km/h,低于 20 km/h;

(3) 拥挤:城市快速路和联络线平均速度低于 20 km/h,主干道平均交通速度低于 10 km/h^[14]。

以南京市内的某一主干道路段进行研究。如图 2 所示,为了得到较精确的结果,首先获取每个簇内车辆的加权平均速度,然后把该速度作为衡量交通拥堵状况的主要参数。由于城市道路网中车辆分布的不均匀性,因此道路上每个簇内采集的样本点分布也不均匀,即有的簇内样本点较少,有的则较多。对样本点较少的簇,通过对簇内所有样本点求加权平均来表示该簇的速度;对样本点较多的簇,通过随机取样的方法来获取该簇的速度。因为由簇的定义可知,同一个簇内样本点的相似度很高,因此可以认为抽取的部分样本能够真实地反映整体水平。

若簇内的样本数量小于等于 10,则对所有的样本点速度加权求平均值;若簇内样本数量大于 10 且小于 30,则应随机抽取样本数量的 80% 来求得平均速度;若簇内的样本数量大于 30 且小于 50,则应随机抽取样本数量的 75% 进行评估;若簇内的样本数量大于 50,则应随机抽取样本数量的 70% 进行计算。从整个城市道路网的交通拥堵检测的层次上来看,提出的基于簇内样本数量计算平均速度的方法,可以在保证检测结果精度的前提下达到简化计算的目的。

4 仿真分析

将获取的 GPS 数据经过预处理和聚类之后,再计算每个簇内的平均速度,然后根据平均速度来判断路段的拥堵情况。文中模拟了某一时刻道路通畅、缓行和拥堵的三种场景,而且对这三种场景进行了仿真,如图 3 所示。区域 1 中汽车平均速度大于 20 km/h,代表的是通畅,其中有两个异常数据在预处理之后被丢弃;区域 2 中汽车平均速度大于 10 km/h 且小于 20 km/h,代表的是缓行;区域 3 中汽车平均速度小于 10 km/h,表示的是拥堵。“*”表示每个簇的质心,对应的横坐标是其编号;折线表示的是处理后的平均速度。图 4 为图 3 对应的速度折线图,该图可以较好地显示道路的拥堵等级。

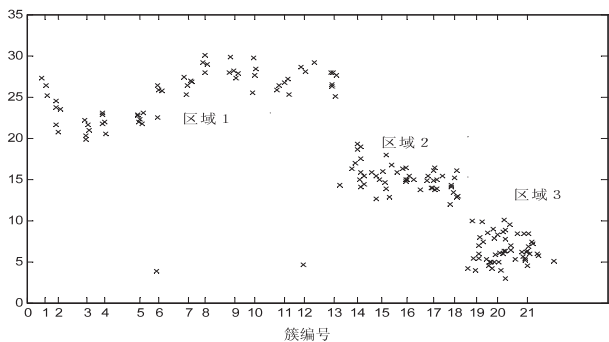


图 3 交通拥堵分布图

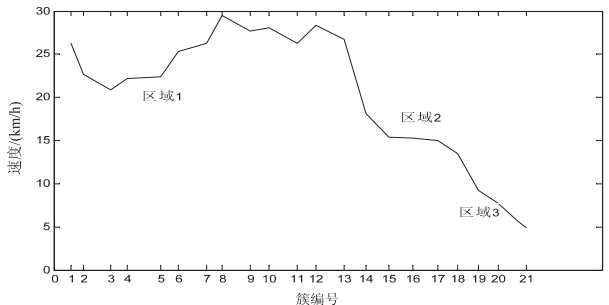


图 4 车辆平均速度折线图

5 结束语

在 GPS 技术快速发展和 GPS 数据广泛应用的基础上,提出了一种基于多源 GPS 数据的道路拥堵检测方法。由于道路交通的突发性和偶然性,采集的交通信息中存在一定的异常数据。为剔除其中的异常数据,提高城市道路拥塞区域的检测精度,对采集的样本数据进行预处理。在保证数据完整性的基础上,使用 k -means 算法对样本数据进行聚类。针对每个簇中的样本数据,提出了一种基于簇内样本数量计算簇内交通情况的方法,以获取的簇内平均速度作为衡量每个簇的交通状况的主要参数。仿真结果表明,该方法可以准确检测交通的拥堵状况。

参考文献:

[1] 林树宽,于伶姿,乔建忠,等. 基于 GPS 轨迹数据的拥堵路段的检测[J]. 东北大学学报:自然科学版,2015,36(11): 1530-1534.

[2] 吴 飞. 智能实时路况系统的关键技术研究[J]. 测绘与空间地理信息,2013,36(8):95-97.

[3] 姚敏杰. 基于车辆联网的实时路况信息系统研究[D]. 杭州:浙江工业大学,2012.

[4] Wei S, Wynter L. Real-time road traffic fusion and prediction with GPS and fixed-sensor data[C]//International conference on information fusion. [s. l.]: [s. n.], 2012: 1468-1475.

[5] 周 洋. 基于 GPS 浮动车的高速公路实时路况系统的研究[D]. 南昌:南昌航空大学,2012.

[6] Gupta A, Choudhary S. DTC: a framework to detect traffic congestion by mining versatile GPS data[C]//Emerging trends and applications in computer science. [s. l.]: [s. n.], 2013: 97-103.

[7] Zhang Z, Yang D, Zhang T, et al. A study on the method for cleaning and repairing the probe vehicle data[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(1):419-427.

[8] Inose H, Fujisaki H, Hamada T. Theory of road-traffic control based on macroscopic traffic model[J]. Electronics Letters, 1967, 3(8):385-386.

[9] Ashbrook D. Learning significant locations and predicting user movement with GPS[C]//Sixth international symposium on wearable computers. [s. l.]: IEEE, 2002: 101-108.

[10] Ho Y H, Wu Y C, Chen M C, et al. GPS data based urban guidance[C]//International conference on advances in social networks analysis and mining. [s. l.]: IEEE, 2011: 703-708.

[11] Amelia A, Saptawati G A P. Detection of potential traffic jam based on traffic characteristic data analysis[C]//International conference on data and software engineering. [s. l.]: IEEE, 2014: 1-5.

[12] 刘瑶杰. 基于实时路况的交通拥堵时空聚类分析[D]. 北京:首都师范大学,2014.

[13] 鲁小丫,宋志豪,徐 柱,等. 利用实时路况数据聚类方法检测城市交通拥堵点[J]. 地球信息科学学报,2012, 14(6):775-780.

[14] 贾 森. 基于实时信息的城市道路交通状态判别方法研究[D]. 北京:北京交通大学,2007.