

利用实时路况数据聚类方法检测城市交通拥堵点

鲁小丫¹, 宋志豪², 徐柱^{2*}, 李木梓², 李婷², 孙维亚²

(1. 西南民族大学计算机科学与技术学院, 成都 610041; 2. 西南交通大学地球科学与环境工程学院, 成都 610031)

摘要: 城市交通拥堵严重制约其网络总体效率。开展检测交通拥堵点可有效识别网络瓶颈, 以整治交通拥堵现象。对此, 本文提出一种新的城市交通时空拥堵点检测的方法: 即采用实时路况数据, 通过定义时空关联, 检测时空意义上长期性、规律性交通拥堵点。本文基于 DBSCAN (Density-Based Spatial Clustering of Applications with Noise) 算法, 以成都市为试验区, 实现了这种拥堵点检测方法。试验表明, 该方法可快速、有效、准确地检测出城市道路严重拥堵路段, 并确定其拥堵时空范围, 为交通管理、交通拥堵机理分析、交通拥堵预测等提供参考。

关键词: 交通拥堵; 拥堵点检测; 时空关联; 聚类分析; DBSCAN 算法

DOI: 10.3724/SP.J.1047.2012.00775

1 引言

交通拥堵点检测和时空分布范围估计是研究拥堵机理、解决拥堵问题的前提^[1]。

交通拥堵是某一时空由于交通需求和供给产生矛盾所引起的交通滞留现象, 在国际上至今没有明确的标准。日本在 1994 年对交通拥堵的定义是一般道路长 1km 以上拥堵时间持续 10 分钟的状况^[2]。由此可见, 交通拥堵具有时间性和空间性。所以, 交通拥堵现象在城市中并非呈现在一段时间内的任何时刻和任何地点, 而是具有一定空间和时间范围的“时空拥堵点”。

目前, 交通拥堵检测方法主要考虑瞬时(或者说短时)局部拥堵的检测, 可分为直接检测和间接检测方法。直接检测方法是以监控视频图像的自动或人工分析来检测拥堵^[3-4]。间接检测方法通常以交通流 3 个基本参数(速度、流量、车流密度)来检测拥堵^[5]。其中, 拥挤度快速判别法、自适应神经模糊推理系统方法^[5]、Bayesian 算法^[3]、California 算法^[6]和 McMaster 算法^[7]等应用较广泛。

瞬时局部拥堵检测结果是道路的短时局部交

通状态, 而交通拥堵有不同的时空尺度, 所以, 不应局限在瞬时局部拥堵的检测, 而要能在更大尺度上检测较大规模、较长时间和规律性的拥堵, 获取时空意义上的拥堵点并确定拥堵时空范围。这可为交通规划、管理部门提供帮助, 进行交通拥堵机理分析。

本文检测时空拥堵点是以瞬时交通路况检测的结果, 从瞬时交通状况序列中挖掘出时空意义上的拥堵点。可将其建模为聚集分析^[8-9]问题。聚集分析应用于交通拥堵检测在文献^[10-13]中有过介绍并取得一定成果, 但这些方法都是面向瞬时局部检测的, 不能在同时空尺度下挖掘出时空拥堵点。此外, 时空拥堵具有时空并存性, 所以, 要采用时空聚集方法来检测。

本文将检测时空拥堵点问题建模为时空聚集的方法, 提出一种时空关联的定义方式, 并将 DBSCAN^[14] (Density-Based Spatial Clustering of Applications with Noise) 算法在时间上扩展, 选择成都市作为试验区, 将这种新的时空聚集方法应用到试验区的时空拥堵点检测当中。最后分析检测结果, 评价检测的正确性。

收稿日期: 2012-11-26; **修回日期:** 2012-12-03.

基金项目: 国土资源公益性行业科研专项经费(201111013); 国家自然科学基金项目(40971209); 中央高校基本科研业务费专项(SWJTU11CX059)。

作者简介: 鲁小丫(1975-), 女, 讲师, 主要从事计算机应用、计算机网络安全研究。E-mail: lu_xiaoya@163.com

*** 通讯作者:** 徐柱(1972-), 男, 博士, 教授, 博士生导师, 研究方向为空间数据挖掘、空间数据综合、空间数据共享、GIS 软件体系结构等。E-mail: xuzhucn@gmail.com

2 基于聚集分析的时空拥堵点检测方法

2.1 拥堵点检测的时空聚集分析建模

时空拥堵点检测旨在揭示时空意义上的大规模、长期性和规律性的拥堵,以获得不同规模的时空拥堵点。

瞬时路况数据序列是一种时态 GIS^[14]数据,空间上相同位置在不同时间片上的属性描述不尽相同,对于仅有一个描述属性的路况数据实际上是一个四维空间,分别是空间维(二维坐标或里程)、时间维、和属性维,如图 1 所示,图中的属性维是路况分类顺序量,共有 4 种。

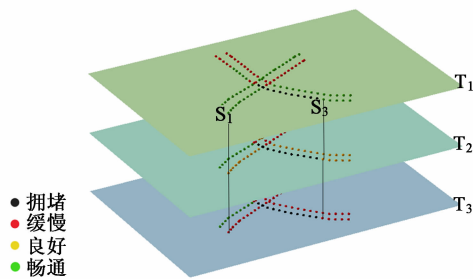


图 1 实时路况数据时空 4 维聚集空间

Fig. 1 The 4-D clustering space of the real time traffic data

交通拥堵具有时空性,故利用瞬时交通路况数据序列检测时空拥堵点可转化为时空聚集分析问题。而其可将道路网空间中具有相同或相似性质的时空关联对象聚集成一系列时空实体集合,以识别时空拥堵点。

目前,时空聚集的主要方法有时空扫描统计方法、密度方法和时空距离的方法^[15]。时空扫描统计方法通过定义圆柱体扫描窗口,用统计检验的方法确定集群。密度方法主要是 Birant 和 Wang 将 DBSCAN 进一步应用到时间维并提出的 ST-DBSCAN 方法^[16]。时空距离的方法主要有两种,一种是分别在时、空上定义邻近实体,一种是综合定义时空耦合距离^[17]。本文采用综合定义时空耦合距离并对聚集分析时空化扩展的方式检测交通拥堵点。

2.2 时空聚集——聚集分析的时空化扩展

道路实时路况数据是时空多维数据,用聚集分析方法检测交通拥堵点需要从时间、空间角度进行

搜索、延展,所以,对聚集分析方法需作时空化扩展。因此,有效地定义时空关联(时空距离)、确定时空邻近域,是解决时空聚集分析方法的核心问题。

某些时空聚集方法避免时空关联的定义^[15-16],只是确定一个时间窗口和空间邻近域,这种方法孤立地考虑时间邻近和空间邻近,缺乏对时空邻近的解释。这里给出一种时空关联的定义方式。

假设城市主干道拥堵状态的平均时速为 v ,单个时间片内两相邻实体间的平均点距,时间间隔为 $t_a - t_b$ 对应的空间点数为 p_{ab} ,任意两个时空对象 $A(x_a, y_a, t_a)$ 、 $B(x_b, y_b, t_b)$,则 A 、 B 之间的时空距离简化定义如式(1):

$$Dist(A, B) = \sqrt{\frac{(x_a - x_b)^2 + (y_a - y_b)^2}{d_{mean}^2} + \frac{v^2(t_a - t_b)^2}{d_{mean}^2}} \quad (1)$$

但通常情况下,在对大量数据进行时空聚集时考虑到数据量的问题,在时间维度上的采样密度相对于空间采样密度偏低,导致聚集时空对象比时间对象更“容易”聚集。如图 2 所示,当时间采样密度过低时,可在时间维上增加虚拟数据来达到时空采样密度的统一。从 DBSCAN 聚类方法的特点可知,只需实现上下两层真实数据的“连通”便可解决上述问题。所以引入参数重定义 A 、 B 之间的时空距离如式(2):

$$Dist(A, B) = \sqrt{\frac{(x_a - x_b)^2 + (y_a - y_b)^2}{d_{mean}^2} + \frac{v^2(t_a - t_b)^2}{d_{mean}^2 P_{ab}^2}} \quad (2)$$

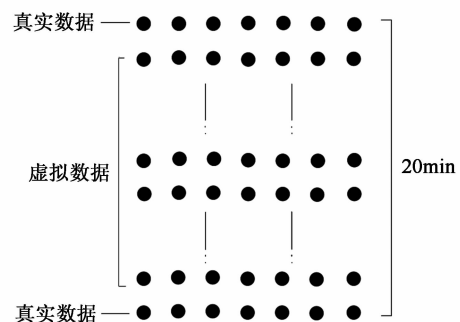


图2 通过增加虚拟数据处理低密度时间采样问题的示意图

Fig. 2 Illustration of dealing with low-density sampling in the time dimension by adding virtual data

通过对聚集算法进行时空化扩展,以获得一种可靠的时空聚集方法用于拥堵点检测。本文的时空聚集算法是将传统的 DBSCAN 算法在时间上进

行扩展,原理是在时空的三维欧氏空间中,将时空对象邻近区域的密度作为划分标准,在给定时空距离半径的邻域内至少包含最少量的点,即邻域的密度必须要超过某个阈值,使集群内的点集密度比集群外明显要高,从而区分集群与噪声。

算法有 2 个参数:时空邻域半径 Eps 、密度阈值 $MinPts$ 。如图 3(a)对象 q 满足 Eps 邻域内密度大于密度阈值 $MinPts$,那么, q 是一个核心点,查找其 Eps 邻域内所有核心点并作为种子,对种子递归进行如上步骤,若 p 恰在某一种子的 Eps 邻域内,那么, p 从 q 是密度可达。图 3(b) p 和 q 从 o 是密度可达,那么, p 和 q 密度相连。通过这种密度聚集分析,最终得到很多时空对象集群,每个时空对象集群都是一些密度相连点的集合。对于建立了三维格网索引的所有时空对象 O ,如果 O 是核心点,则通过时空查询得到该点的 Eps 邻域,如果该邻域中存在核心点,则把这些核心点作为种子点进行下一轮考察,寻找与其密度相连的点,不断进行时空查询来拓展当前集群。然后依照上述方法继续寻找其他集群,直至除了噪声点外的所有密度相连的点归并到不同的时空集群中。

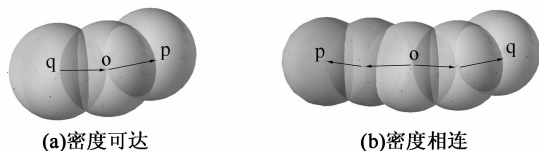


图 3 密度可达与密度相连示意图

Fig. 3 Illustration of density-reachability and density-connectivity

从对算法的描述来看,时空密度的定义已被简化,只是考虑对象 Eps 邻域内点的数目,这是因为采用全局性的 Eps 值定义。两个密度不小于最小密度集群的点集间最小距离必定大于 Eps ,所以,两个时空交通拥堵点在时空范围内将得以区分。下面将上述时空拥堵点检测方法应用于城市试验区。

3 时空拥堵点检测实验与分析

3.1 拥堵点检测的数据获取

本文使用的数据是成都市路网矢量数据和从 Google 地图上下载的成都市栅格路况数据。我国公安部在 2002 年颁布的“城市交通管理评价指标体系”中规定,交通拥堵程度可以根据城市主干道上的机动车平均行驶速度分为 4 个程度:畅通(\geq

30km/h)、轻度拥挤(≥ 20 km/h 且 < 30 km/h)、拥挤(≥ 10 km/h 且 < 20 km/h)、严重拥挤(< 10 km/h)^[1]。本文使用的 Google 地图实时路况数据,拥堵程度被定义为以上 4 个类别,其颜色按上述顺序分别对应绿色、黄色、红色和红黑相间。

在时间维上,由于多时间片数据的时间粒度(time granularity)大小对分析结果的影响较严重,时间粒度越小检测结果越精确,而过小的时间粒度又会导致数据量过大、计算开销增大,所以,本文把实时路况的时间粒度确定为 5min。在空间维上,将每个时间片的路网矢量数据离散化成点图层,然后从经过分类的栅格路况数据中提取描述信息,离散化间隔为 50m。

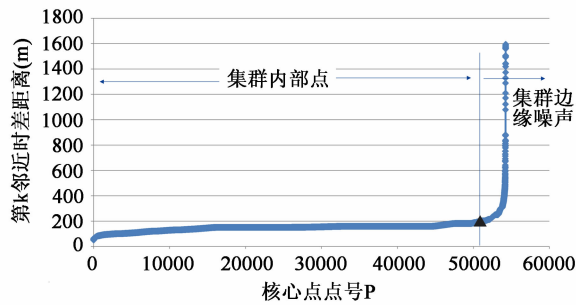
3.2 确定参数及算法评估

本文的时空聚集方法有两个参数:密度阈值 $MinPts$ 和时空邻域半径,通常需要人为确定最小但足以构成拥堵集群的点集内点的个数,即可以把此值作为算法的密度阈值 $MinPts$ 。根据程序可以计算出核心点第 $MinPts-1$ 邻近距离,这是邻域中与核心点相距最远点到核心点的距离。搜索的时空邻域半径的确定本文采用文献[14]中提供的一种 K-Dis 图方法,根据相应 $MinPts$ 值输出 K-Dis 图分析获得。所以,算法只需要用户预定义一个输入参数 $MinPts$,而所谓的 K-Dis 图是对每个时空对象的第 K 邻近距离排序后的示意图。

选取任意一个核心点 p ,将 $MinPts$ 设置为 $K+1$,用 $K_Dis(p)$ 表示第 K 邻近距离,对所有核心点进行第 K 邻近距离计算并输出 K-Dis 图。实验采用了成都市 2012 年 12 月 22 日 17:15-19:00 时间间隔为 5 分钟的 22 个时间片数据近 100 万条记录进行计算,得到 K-Dis 图,如图 4 所示。

从所有时空点的 K-Dis 图中可以看到存在一个突变转折点,将转折点所对应的第 K 邻近距离确定为对象的邻域半径 Eps 。所有第 K 邻近距离小于 Eps 的点便是拥堵集群中的点,而大于 Eps 的点全部被认为是集群边缘噪声,经过计算确定数据点的集群边缘噪声率约为 11.5%,而在多个单时间片下集群边缘噪声率平均约为 9.95%。

本文的拥堵点检测方法,通过建立时空三维格网索引来提高检索、查询效率,这对交通拥堵点检测方法极其重要,特别是在时空距离的计算和存储上,可极大减少算法的耗时和占用存储空间,本文

图 4 时空实时路况数据点 $k=8$ 时的 K-Dis 图Fig. 4 The K-Dis diagram of spatio-temporal points as $k=8$

用不同数量点数据进行实验并将其与图 5 中进行对比,建立索引之后的效率明显提升,数据量越大优势越为明显。

考虑到 K 值的选择具有主观性,且 Eps 值很大程度上决定于 K 值的选择,所以,需注意, K 不能取值过大,当 K 值在允许范围内,随 K 值增大,集群个数逐渐减少,因为 K 值增大导致少数小型集群在聚集过程中被忽略或者几个相距较近的集群合并为一个独立集群,以致于不能很好地反映实际拥堵情况。该方法对单时间片空间聚集和大量时间片的时空聚集都适用,算法运行时间与 K 取值、Eps

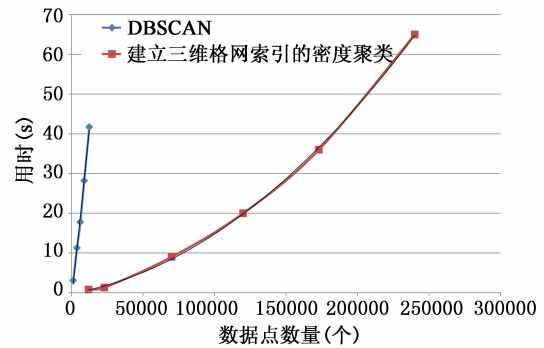


图 5 使用时空索引与未使用索引的计算用时对比

Fig. 5 Comparison of computing time used with and without index

值无关,只与计算数据量有关,在试验用计算机上约为 4min/100 万个数据点。

3.3 试验结果与分析

根据前述道路交通拥堵集群检测方法,对成都市 2012 年 12 月 22 日晚高峰时段 22 个时间片数据进行时空聚集分析。

对成都市 2012 年 12 月 22 日某一时刻的单时间片数据采用上述方法检测拥堵路段的结果及其栅格数据比较如图 6 所示。

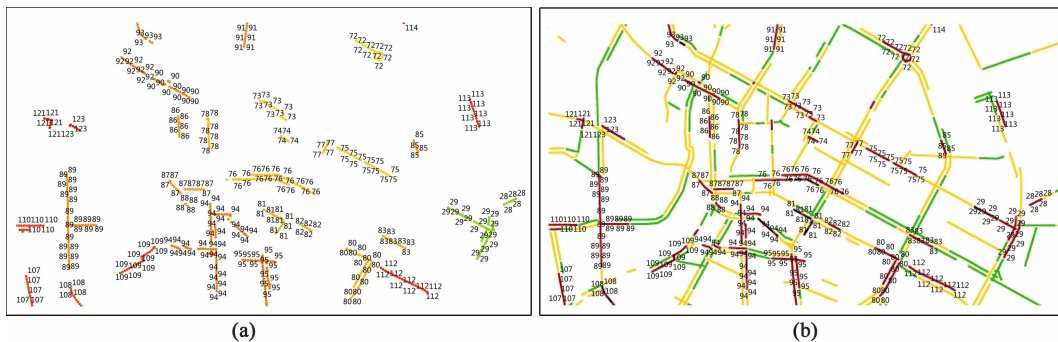


图 6 单时间片聚集结果(a)与原始数据(b)的对比

Fig. 6 The clustering result of single time slice (a) as compared to the source data (b)

对多时间片数据进行时空聚集,输出集群内时空点的个数,并根据此值大小决定拥堵点的拥堵严重性,图 7(a)为根据拥堵严重性进行分类后的时空拥堵点情况,从红色到绿色为拥堵严重性逐渐降低的过程,市中心区域拥堵严重,越往郊区方向绿色拥堵点越多,说明拥堵越为分散。对路段在时间维上拉伸以至在时间上构成连续拥堵,可以明显地得到路段的拥堵时间段,如图 7(b)所示。

交通拥堵发生时间具有规律性,根据拥堵的时间段可分为早高峰、午高峰、晚高峰、平峰等,本文

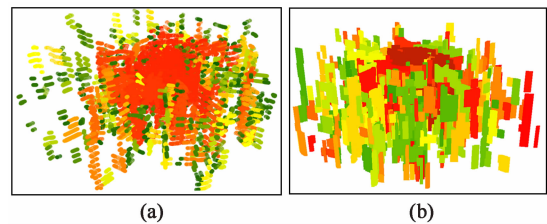


图 7 多时间片时空拥堵点检测结果的三维可视化

Fig. 7 Detecting results of multiple time slices in 3D view

采用傍晚时间段数据,主要观察成都市晚高峰时段交通拥堵在时间上的变化过程。对上述成都市傍

晚时段 22 个时间片数据的时空拥堵点聚集结果投影到一个时间片上,确定时空拥堵点的最大空间范围,并将成都市三环内的几个大型拥堵点分类生成矢量线图层,如图 8 所示。从图中可直观地发现并定性地得出结论:距离市中心愈远,拥堵路段数量愈少,交通拥堵点规模愈小,拥堵严重性愈轻;在一环内部形成一个大的拥堵路段集群,这是因为城市交通具有变动性和自适应性,所以,在道路网络中上一时刻的拥堵交通流量在下一时刻极有可能会蔓延或传递到网络中的其他相邻位置,造成时空中的大型拥堵。

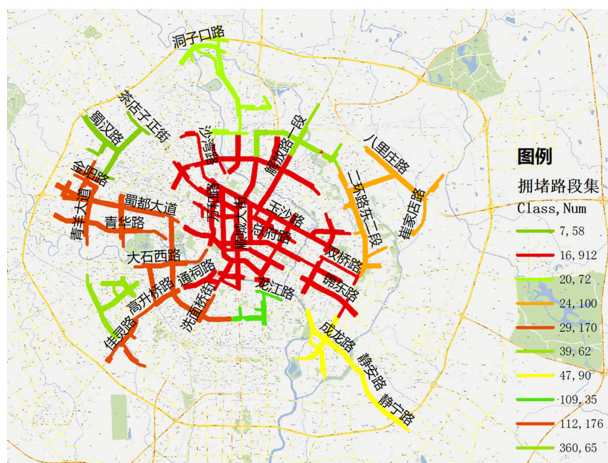


图 8 成都市三环内大型拥堵点的提取结果

Fig. 8 Detection results of Chengdu within the third ring road

成都市交管局在 2011 年 12 月发布的数据中显示了成都市拥堵最为严重的 12 个路段,其中有蜀都大道东段、东大街、高升桥段、二环路、蜀汉路、解放路、北站西二路等,这些拥堵路段与本文的检测结果吻合,图 9 是部分拥堵路段对应的检测结果图。

4 结束语

本文提出的瞬时路况数据序列检测交通时空拥堵点的方法,将时空拥堵点的检测建模为时空聚类问题,通过定义适用的时空关联,对空间聚集分析进行时空化扩展。试验表明,该方法可有效地从实时路况数据中检测出时空拥堵点,以获得拥堵的空间范围、持续时间、涉及路段等。

今后,对交通时空拥堵点检测与分析应深入考虑:(1)在更大时间尺度上检测拥堵,分析其时空动



图 9 部分交管部门发布的严重拥堵路段的拥堵检测结果

Fig. 9 Reported congestion areas as detected by the proposed method

态特性,揭示其演进规律;(2)道路实时路况数据实时产生,应进行实时分析,需研究增量式检测算法,提高分析效率。

参考文献:

- [1] Jiang G Y, Niu S F, An D C. The method of traffic congestion identification and spatial and temporal dispersion range estimation[C]. International Asia Conference on Informatics in Control, Automation and Robotics, 2010, 1: 36 - 39.
- [2] 祝付玲. 城市道路交通拥堵评价指标体系研究[D]. 南京: 东南大学, 2006.
- [3] 赵有婷, 李熙莹, 等. 基于视频全局光流场的交通拥堵检测[J]. 计算机应用研究, 2010, 27(11): 4355 - 4362.
- [4] Li L, Chen L, Huang X, Huang J. A traffic congestion estimation approach from video using time-spatial imagery[C]. ICINIS '08 Proceedings of the 2008 First International Conference on Intelligent Networks and Intelligent Systems, 2008, 465 - 469.
- [5] 石征华, 侯忠生. 城市快速路拥挤度判别方法研究[J]. 交通与计算机, 2006, 24(5): 20 - 23.
- [6] Payne H and Knoel H. Development and testing of incident-detection algorithm[J]. research methodology and detailed results[R]. FHWA-RD-06-20, 1976.
- [7] 刘伟铭. 高速公路系统控制方法[M]. 北京: 人民交通出版社, 1998.
- [8] 周成虎, 裴涛. 地理信息系统空间分析原理[M]. 北京: 科学出版社, 2011.

- [9] Michalopoulos P G. Vehicle detection video through image processing: The autoscope system[J]. IEEE Trans on Vehicular Technology, 1991, 40(1): 21 - 29.
- [10] Yang Y, Cui Z, Wu J. Fuzzy c-means clustering and opposition-based reinforcement learning for traffic congestion identification[J]. Journal of Information and Computational Science, 2012, 9(9): 2441 - 2450.
- [11] Lozano A, Manfredi G, Luciano N. An algorithm for the recognition of levels of congestion in road traffic problems[J]. Mathematics and Computers in Simulation, 2009, 79(6): 1926 - 1934.
- [12] Erman J, Arlitt M F, Mahanti A. Traffic classification using clustering algorithms[C]. Proceedings of the 2006 SIGCOMM Workshop on Mining Network Data, MineNet'06, 2006.
- [13] 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.
- [14] Ester M, Kriegel H, Sander J, and Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise[C]. KDD-96 Proceedings, 1996, 226 - 231.
- [15] 吴信才, 曹志月. 时态 GIS 的基本概念、功能及实现方法[J]. 地球科学—中国地质大学学报, 2002, 27(3): 241 - 245.
- [16] Birant D and Kut A. ST-DBSCAN: An algorithm for clustering spatial-temporal data[J]. Data and Knowledge Engineering, 2007, 60(1): 208 - 221.
- [17] 邓敏, 刘启亮, 王佳, 等. 时空聚类分析的普适性方法[J]. 中国科学: 信息科学, 2012, 42(1): 111 - 124.

Urban Traffic Congestion Detection Based on Clustering Analysis of Real-time Traffic Data

LU Xiaoya¹, SONG Zhihao², XU Zhu², LI Muzi², LI Ting² and SUN Weiya²

(1. Faculty of Computer Science and Technology, Southwest University for Nationalities, Chengdu 610041, China;

2. Faculty of Geosciences and Environment Engineering, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: Traffic congestion in urban road network heavily restricts transportation efficiency. Detecting traffic congestions in the spatio-temporal sense and identifying network bottlenecks become an important task in transportation management. Up to now, many traffic congestion detection methods have been proposed, which have focused on the detection of momentary local congestions. Larger-scale, longer-time and regular congestions can't be detected using these methods. That is because congestions have different tempo-spatial scales, and a characteristic is not considered in those methods. This paper proposes a new kind of urban traffic congestion detection method that deals with spatio-temporal extension of congestion. It is based on spatio-temporal clustering analysis of real-time traffic data. By defining a proper spatio-temporal correlation, the classic DBSCAN algorithm is adapted to tackle spatio-temporal clustering. With it we can detect longer time and regular traffic congestion in the spatio-temporal sense. Experiments have been conducted using real traffic condition data of Chengdu to validate the effectiveness of the method. The experiment shows that the proposed method can detect the congestion areas and identify the spatio-temporal extent of congestions accurately. The detected congestion areas were compared with congestion report from local traffic management authority and found to be consistent with the later.

Key words: traffic congestion detection; spatio-temporal correlation; spatio-temporal clustering analysis; DBSCAN algorithm