DOI: 10.3979/j.issn.1673-825X.2018.01.005

面向公共安全的时空数据挖掘综述

王永坤[1],王海洋[2],潘平峻[2],李龙元[2],金耀辉[1,2]
(1. 上海交通大学 中国城市治理研究院,上海 200240; 2. 上海交通大学 光纤通信国家重点实验室,上海 200240)

摘 要: 随着各种手持无线设备及传感器的普及,大量的具有时间和空间属性的轨迹数据在不间断地产生。这些不同来源的轨迹数据记录了个体在时间和空间上的活动,从微观和宏观揭示出个人和团体的活动规律,对研究人群行为及城市管理,特别是城市公共安全管理方面,具有重要的意义。以公共安全管理为主要目标,分 4 个方面调研了相关的研究工作,并分别给出了笔者的研究进展。使用了 2 类比较有代表性的数据,第 1 类是智能手机的时间、空间轨迹数据;第 2 类是城市公共交通卡的换乘数据。第 1 类是从"点"上分析挖掘个体或者群体的活动规律,而第 2 类数据则是从"线"上发现人群的聚散规律。基于第 1 类数据,针对"个体的发现"介绍了相关工作;对于第 2 类数据,分别从短时和突发 2 个方面,发现具有潜在危害性的事件,从而向有关部门提供预测和预警,防范该区域可能出现的公共安全事件。比较了各类模型包括经典的时序数学模型 ARIMA(autoregressive integrated moving average model)和 SARIMA(seasonal autoregressive integrated moving average)、机器学习和神经网络模型 SVR(support vector re-gression)、NN(neural networks)、和 LSTM(long short-term memorg),发现笔者的模型在短时客流预测方面可以最多提高 27.78%,突发客流预测精度可以最高提高到 14.68 倍。

关键词: 时空分析;大数据;异常发现;数据预测
中图分类号: TP399 **文献标志码:** A **文章编号:** 1673-825X(2018)01-0040-13

A survey of data mining on spatial-temporal user behavior data for public safety

WANG Yongkun[1], WANG Haiyang[2], PAN Pingjun[2], LI Longyuan[2], JIN Yaohui[1,2]
(1. China Institute for Urban Governance, Shanghai Jiao Tong University, Shanghai 200240, P. R. China;
2. State Key Lab of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, Shanghai 200240, P. R. China)

Abstract: With the popularity of smart phones and wireless sensors, large amount of data with timestamps and geo-locations (spatial-temporal) has been produced continuously. This spatial-temporal data records individual behaviors by time and locations, shows macro and micro behavior patterns of people by statistical methods, which is very important for studying the human behavior, especially significant for managing the public safety for city administrators. In this paper, we survey the state-of-the-art research of the human behavior mining for public safety on spatial-temporal data in four aspects, and provide our work in each aspect respectively. We discussed two types of spatial-temporal data, one is smartphone data, and the other is smart card data of public transit. The former shows the individual and crowd behavior from "point" view, and the latter shows the crowd behavior pattern from "line" view. With the former data, we discussed how to discover suspect individ-

收稿日期: 2017-09-28 **修订日期:** 2018-01-10 **通讯作者:** 金耀辉 jinyh@sjtu.edu.cn
基金项目: 国家自然科学基金(61371084)
Foundation Item: The National Natural Science Foundation of China (61371084)

uals; with the latter data, we introduced how to find harmful events from short-term and burst passenger traffic, so as to provide the early warning to administration if necessary. We compared our model with existing ones such as ARIMA, SARIMA, SVR, NN, and LSTM. The result shows that our model can reduce the error up to 27.78% for short-term traffic prediction, and up to 14.68x for burst traffic prediction.

Keywords: spatial-temporal analysis; big data; outlier detection; prediction

0 前　言

随着手持设备及传感器技术的发展，大量用户的时间和空间轨迹数据被记录了下来。例如，智能手机已经普及到了城市的绝大部分人群，智能手机中内置的软件和硬件，可以非常准确地记录用户在某一时刻的位置，进而可以记录用户较完整的时间、空间轨迹。另外，各种车载设备也可以记录用户的时空轨迹，比如出租车以及各种新能源车，都搭载了传感器，可以以近似实时的速率上传地理位置信息。还有最近兴起的共享单车也搭载了传感器，可以比较精确地知道单车及用户的时空轨迹情况。在一些大城市比如上海，城市公共交通系统中的地铁、公共汽车、出租车等，在计费系统方面已经连通起来，所以，用户可以使用公共交通卡在各种交通工具中切换，从而用户的轨迹也被各种交通工具记录了下来。

在保证用户隐私的前提下，对用户的时间、空间轨迹数据进行挖掘，可以发现很多个体和集体的行为规律，从而为更多的上层应用和决策提供参考。已经有很多研究者对用户时空数据从不同方面入手开展了大量的研究。由于出发点或者目标不同，研究者们使用的数据也多种多样，研究方法、研究目标也有各自的特色。

本文尝试从公共安全管理的角度出发，调研并讨论相关的研究，期待将用户的时空数据应用于目前城市管理者和市民最关心的公共安全方面。公共安全管理是当前城市管理的一个非常重要而且迫切的方面。随着经济的高速增长和城市化的快速发展，危害城市公共区域的事件和危机也不断出现，给城市管理者带来了巨大的压力[1]。以往的人工方法的危机监测和应对方式既被动又低效，因此，使用大数据及人工智能技术，利用个体移动设备数据以及各种传感器监测信息，挖掘出潜在的风险，及时预警相关管理部门，成为当前研究的热点。

本文以公共安全管理为主要目标，分 4 个方面调研了相关的研究工作，并分别给出了我们的研究进展。我们使用了 2 类比较有代表性的数据，第 1 类是智能手机的时间、空间轨迹数据; 第 2 类是城市公共交通卡的换乘数据。第 1 类是从“点”上分析挖掘个体或者群体的活动规律，而第 2 类数据则是从“线”上发现人群的聚散规律。基于第 1 类数据，针对个体的发现介绍了相关工作; 对于第 2 类数据，我们分别从短时和突发 2 个方面，发现具有潜在危害性的事件，从而向有关部门预警，防范该区域可能出现的公共安全事件。

1 基于个体移动模序异常的大规模活动识别

随着世界人口的不断增多，公共安全已经日渐成为了一个需要重点关注的问题。古往今来，由于人群一段时间内聚集地过于密集、拥堵而引起的灾祸事件屡见不鲜。而近年来，有关公共场所大规模人群活动的不安全事件更是层出不穷。可以说对大规模活动的监管控制是公共安全治理的一个重大话题。

移动通讯技术和定位技术的发展使得研究者可以获取大量人群的移动数据、捕捉用户的移动轨迹。通过挖掘这些带有用户行为特征的轨迹数据可以得出很多有意义的结论。对大规模活动的监管也开始趋向于多时间维度、更广泛空间维度和方法多样化。

目前针对人类移动性研究和大规模活动检测研究现状有如下几个方面。

1.1 人类移动性研究现状

在文献 [2] 中，研究者发现人类移动模式并不是随时间随机分布的，而是在一个时间周期内由大量重复性事件再加上一些少量的突发事件组成。这种非泊松性质的人类活动性是对人类移动性分析的基础。人类移动是由大部分的常规性规律性移动和一部分差异化的突发性移动组成。在常规性研究方面，文献 [3] 证实了人类移动具有很高的可预测性，约有 70% 的时间用户都基本处于其该时段最常访问的地方，这种高可预测性根源于人类空间运动的高度规则性。文献 [4] 同样证实了这一点，每个个体都具有高度的时空维度的规律性。每个个体都具有一定的与时间无关的迁移距离特征，并且总是有很大可能性会回到一些高频率出现过的地区。文献 [5] 的研究者使用移动模序对人类移动模式进行了

探索。通过对使用个人问卷调查和匿名的移动电话数据获得的日常移动数据进行网络性分析,研究者发现17种不同的移动模序就可以覆盖来自不同国家的90%调研者的移动出行模式。同时,文献[5]中提出可以使用移动距离分布$p(r)$、回转半径$r(t)$、一段时间内的访问位置数$S(t)$这3个指标来描述人类移动模式。基于人类移动模式的常规性,研究者通过定量研究发现在大量人群中这3个指标存在一个普遍结论:都近似服从各自参数的幂律分布。这无疑对人类移动的常规性是一个有力的说明。

同时人类移动也由一部分带有差异化的突发性事件组成。即使面对相同原因的外界刺激,不同个性的人也会做出不同的移动性反应,因此,这部分突发性事件在数据中的存在可以帮助我们区分不同用户。文献[6]指出不同个体移动性上的差别是由个体移动模式和一定的群体异质性相互卷积造成的。而文献[7-8]都指出了人类移动行为很大程度上会受到彼此之间的社会交流影响。在社交关系中,双向边的好友关系对于空间位置移动轨迹的影响更大。文献[8]详细对比了社交关系对移动模式影响在不同用户上的差别,发现性别、年龄的不同使得用户在空间移动签到上也有很大差异。这种差异性使得在进行移动预测时需要考虑用户的自身属性。文献[9]则使用社交媒体签到数据以及由签到数据获得的移动轨迹数据反向推断用户的一些固有属性,进行用户画像。通过移动性的分析在推断用户年龄层次、婚姻状况、教育程度上都取得了较好的成果,这从另一个方向证明了人类移动会受个性化因素影响。

1.2 大规模活动检测现状

大规模活动或是一些异常活动在时空层面上,会导致一个区域在一个时间段内反常地聚集大量的人群;在人群整体层面上,会导致一批过去时空距离没有相关性的人在一个时间段内聚集在一起;在个体层面上,会导致一个有很强移动规律性的个体产生有异于往常的移动行为。大规模活动的监测在时间尺度上可以分成3个部分:事前预测、事中监控和事后检测挖掘。其中,事前预测主要使用交通数据、社交签到数据,在已有大量用户移动轨迹数据和活动记录的基础上,对未来可能发生的大规模活动进行预测[10-12]。而事中监控则主要使用一些活动、人流量较大较密集区域的现场监控设备进行人流量的自动识别统计[13-14]。文献[13]提出了一种基于多特征融合的人数统计算法,可以快速而准确地计算出视频中大规模活动人数,以方便管理者进行现场监管。文献[14]利用多个监控器即时地检测人群活动,将多个监视器结果进行汇总评判并可以即时地给出对活动时间的检测结果。事后检测挖掘部分与本文主题契合,是本文的研究重点,在该领域同样有大量研究成果[15-18]。文献[15]把聚集的概念定义为大量个体持续、稳定地以高密度状态聚集在一个区域里,并研究了如何从轨迹数据库中发现这种聚集模式。文献[16]介绍了一种使用海量移动手机数据作为数据源,基于贝叶斯位置推断框架的社会事件检测方法,并讨论了一些未来可行的时间检测技术。文献[17]通过分析非洲一些国家的手机信令定位数据,定义了一种“圆柱聚类法”来捕捉处理这种稀疏的数据,并通过一系列方法从轨迹数据中提取出异常聚集人群。最终该方法在稀疏数据集上识别异常聚集人群上的效果被证明比简单的聚集算法更好。文献[18]则使用了隐马尔科夫模型来寻找具有相同时空移动模式的用户并检测其聚集行为。上述大规模活动检测的相关工作数据来源与事前预测和事中监控较为不同,大多数使用的是移动设备的位置记录功能,包括基站定位,社交网络位置签到、GPS(global positioning system)定位以及WIFI热点定位。

最后文献[19]使用移动手机数据分析了用户移动性和大规模社会事件的关系,发现举办地点距离用户的家越近,越会吸引用户参加从而改变其移动性。该研究证明了大规模活动和用户移动性之间的相互关联和互相影响。本文将继续深入探究这种影响在数据中的表现。

1.3 笔者的相关工作

进行大规模活动检测一方面可以帮助我们对人群移动模式有更深入的了解,深入地探索大规模活动对人类移动性的影响,以及人类移动性变化对这种影响的反映情况。

使用移动模序(motif)[5]作为研究方法,该方法可以抽象地表现单个个体的移动结构。我们已知大规模活动会对单个个体的单次移动活动或移动选择产生巨大影响,而借助移动模序的变化可以探索大规模活动对个体的整体移动性结构造成的影响。

单个个体的移动结构变化很难影响整体移动结构分布的变化,而大规模活动会造成相当大一批人的移动结构都产生变化,因此这种变化会展示在整体移动结构分布上,从而使得我们可以在整体移动

结构分布的层面上研究变化趋势，发现大规模活动。本研究从多源时空数据中提取模序，并通过检测个体的异常模序来推测大型活动。图 1 给出了我们的工作流程。

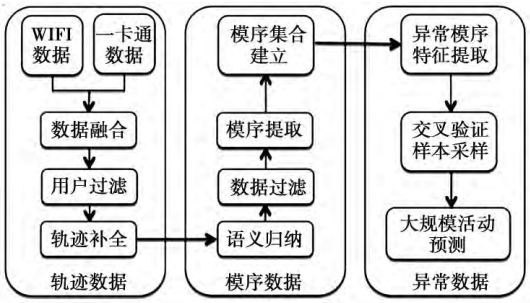


图 1　工作流程框架

Fig. 1　Flowchart of system

本研究场景下的校园内大规模活动事件，映射到社会的场景下就是一些社会事件，这些大规模活动或社会事件的存在可能带有一些危害性，需要监管者对其发生发展进行足够的监管。在历史数据上进行大规模活动检测，可以帮助管理者发现一些始终没有被记录在案的大规模活动。如果能够从历史数据中挖掘检测到这些无备案的大规模活动，则日后对该类活动可以形成及时而有效的备案。而对有备案记录的大规模活动的检测挖掘，可以研究该活动的参与度以及该活动在人群中的影响力，帮助监管者对大规模活动有更深入的掌握。通过在历史数据上的活动检测，可以对大规模聚集事件进行事前预测、事中监控、事后挖掘 3 个维度的协同管理，完善大型事件的监管机制。

2　基于时空轨迹检索的团伙发现

在实际应用中，常常有这样的问题: 已知一个嫌疑人的轨迹，但是不知道他有没有其他团伙。或者给定一个人的轨迹，如何找到和他一起移动过的人。

用户的时空轨迹可以看做是用户位置随时间变化的时间序列。基于时空轨迹检索的团伙发现定义为以下问题: 在一群人的时空轨迹数据库中，给定一个人的轨迹，如何找到与其具有一定相似子轨迹的人。实际的轨迹数据相似度度量需要考虑如下问题。

1) 异频采样性: 轨迹的时间序列并不能保证按照固定的间隔采样。例如手机信令产生的轨迹数据可以由用户主动的行为产生，也可以由手机与基站的通信、切换等被动产生。这种采样的不均匀性影响了相似度的度量。

2) 时间序列性: 轨迹点的产生伴随着时间戳，2 条轨迹的相似应当不仅仅在形状上相似，还应该在时间上相似。

3) 异常点: 由于传感器的失效，或者设备的故障，可以会带来时间或空间上的异常点。常规的欧氏距离的方法可能会因为异常点的存在计算出非常大的距离。

4) 不等长度: 在实际应用中，2 条轨迹中包含的轨迹点数量并不能保证是一样的。因此，相似度度量要截断长的轨迹或者填充短的轨迹。

5) 效率: 为了保证实际应用中的可用性，需要保证相似度的计算要相对简单。

6) 相似子轨迹: 2 条轨迹相似并不一定处处相似，若 2 条轨迹存在一定长度的子轨迹互相相似，那么也应当给予合适的相似度。

2.1　轨迹检索的研究现状

轨迹的相似性检索过程中，需要计算轨迹和轨迹之间的相似度。2 条轨迹的相似度通常用某种聚合的距离来表示。对于 2 条轨迹 A，B，分别由点集 p，p' 组成，传统的方法为最近点对法(closest-pair-distance) ，即使用 2 个轨迹间最相近的一对点的距离来表示相似度，$CPD(A,B) = \min_{p\in A,p'\in B} D(p,p')$，其中，$p$，$p'$ 分别为 2 条轨迹上的点。假设 2 条轨迹的长度相等，那么可以用点对和距离(sum-of-pairs distance) 来表示，$SPD(A,B) = \sum_{i=1}^{n} D(p_i,p'_i)$，其中，$p_i$，$p'_i$ 分别为 2 条轨迹 A 和 B 上的点。2 点的距离常常用欧氏距离表示。

但在实际应用中，2 条轨迹的长度不一定相等，因此，Agrawal 等提出动态时间弯曲 (dynamic time warping，DTW) 距离，DTW 允许一些点重复使用，将相似度度量转化为最优化问题，来获得满足约束条件的代价最小的路径，使得得到的总距离最小[20-21]。DTW 的算法为

$$DTW(A,B) = \begin{cases} 0, & m = n = 0 \\ \infty, & m = 0 \text{ or } n = 0 \\ dist(p_1,p'_1) + \min\{DTW(Rest(A), \\ \quad Rest(B))\}, & \text{其他} \\ DTW(Rest(A),B), DTW(A,Rest(B)) \end{cases} \quad (1)$$

(1) 式中: m，n 分别为轨迹 A，B 的长度; $Rest(A)$ 指轨迹 A 去除首项后的剩余轨迹。

然而，由于轨迹数据中常常存在大量的噪声，使

得 DTW 无法找到合适的匹配。为了解决这个问题,Michail 等研究者提出了基于公共子序列(longest common subsequence,LCSS)的方法来度量轨迹的相似度[22],基于 LCSS 的方法允许跳过某些噪声点,并简化的计算的复杂度。使用阈值 ε 来控制 2 点匹配时允许的距离,使用阈值 δ 来控制轨迹中 2 点匹配允许的时间差。基于 LCSS 的度量算法为

$$LCSS_{\varepsilon,\delta}(A,B) = \begin{cases} 0, A \text{ or } B \text{ is empty} \\ 1 + LCSS_{\varepsilon,\delta}(Head(A),Head(B)), \\ \quad |a_{x,n} - b_{x,m}| < \varepsilon, |a_{y,n} - b_{y,n}| < \\ \quad \varepsilon \text{ and } |n-m| < \delta \\ \max(LCSS_{\varepsilon,\delta}(Head(A),B), \\ \quad LCSS_{\varepsilon,\delta}(A,Head(B))), \text{otherwise} \end{cases} \quad (2)$$

(2) 式中,$Head(A)$ 指轨迹 A 去除尾项后的剩余轨迹。

类似于最大公共子序列的方法,Chen 等[23]提出了基于编辑距离的实序列编辑距离(edit distance on real sequence,EDR)[21]。编辑距离又称 Levenstein 距离,指两字符串之间由其中一个转化为另一个所需的最少编辑次数。EDR 使用阈值 ε 来控制匹配的过程,并为子序列的匹配赋予惩罚。EDR 的算法为

$$EDR(A,B) = \begin{cases} n, & m=0 \\ m, & n=0 \\ \min\{EDR(Rest(A),Rest(B)) + \\ \quad subcost, EDR(Rest(R),S)+1, \\ \quad EDR(R,Rest(S))+1\}, \text{otherwise} \end{cases} \quad (3)$$

Chen 等[23]还提出了实数代价编辑距离,通过给定一个固定的参考点来结合 DTW 和 EDR 的优势。

除了确定的轨迹外,对于不确定的轨迹,即轨迹 X 是确定的轨迹 O 和一系列的概率分布函数的组合。Chunyang 等[24]研究者提出了 KSQ(top-k similarity query),KSQ 集中于在不确定的轨迹数据库中,找到最相似的 k 条轨迹,又称 Top-k 查询。其中,最关键的部分是如何近似地量化 2 条不确定轨迹的相似度。KSQ 采用了一个新的距离测度和一个可扩展的索引架构来支持查询。

对于语义的轨迹,即以语义的位置来代替实际经纬度位置的轨迹。Xiao 等[25]研究者将相似度度量从实际的物理位置扩展到了语义空间。在没有物理空间限制的情况下,可以度量不同城市中生活的用户的相似度。在语义空间中,从轨迹中生成驻留点,使用驻留点对应兴趣点(point of interes,POI)的分布来表示一条轨迹。这样不同用户的驻留点就可以聚类为层次结构。

以上的研究大多是对于整条轨迹的查询,没有考虑其中具有相似子轨迹的情况。对于相似子轨迹的距离测量,Jeung 等[26]研究者提出了基于最小外接矩形(minimum bounding rectangle, MBR)的子轨迹度量。MBR 将子轨迹用最小外接矩形包裹,然后使用矩形的端点坐标 (x_u,y_u),(x_l,y_l) 描述轨迹。基于 MBR 的相似度 $D_{\min}(B_1,B_2)$ 定义为点(B_1,B_2)中最小的距离,计算公式为

$$\sqrt{(\Delta([x_l,x_u],[x'_l,x'_u]))^2 + (\Delta([y_l,y_u],[y'_l,y'_u]))^2} \quad (4)$$

(4) 式中:$x'_l, x'_u y'_l, y'_u$ 为轨迹 B 最小外接矩形包裹的 4 个端点坐标。2 个区间的距离定义为

$$(\Delta([x_l,x_u],[x'_l,x'_u])) = \begin{cases} 0, & [x_l,x_u] \cap [x'_l,x'_u] = \varnothing \\ x'_l - x_u, & x'_l > x_u \\ x_l - x'_u, & x_l > x'_u \end{cases} \quad (5)$$

Lee 等[27]研究者提出了另一种相似度测量,叫做轨迹-豪斯多夫距离(Trajectory-Hausdorff distance),D_{haus} 是 3 种距离的加权和,距离测量的公式为

$$D_{\text{haus}} = w_1 d_\perp + w_2 d_\parallel + w_3 d_\theta, \quad (6)$$

(6) 式中:w_1, x_2, x_3 是权重,根据实际应用取不同的值;$d_\perp$ 表示轨迹之间的分离程度的聚合垂直距离;$d_\parallel$ 表示轨迹长度区别的聚合平行距离;d_θ 表示轨迹方向上的区别的角距离,其计算方法分别为

$$d_\perp = \frac{d_{\perp,a}^2 + d_{\perp,b}^2}{d_{\perp,a} + d_{\perp,b}} \quad (7)$$

$$d_\parallel = \min(d_{\parallel,a}, d_{\parallel,b}) \quad (8)$$

$$d_\theta = \|L_2\| \cdot \sin\theta \quad (9)$$

(9) 式中:θ 表示角度;L_2 表示轨迹夹角的对边长度。

2.2 笔者的工作

对于手机基站提取的用户轨迹,与 GPS 等轨迹

数据不同，我们从中观察到一种现象: 基站中提取的用户轨迹，大部分用户都具有自己固定的活动空间，用户绝大多数轨迹都在小范围空间里产生。

根据这个现象，我们发现在相似性查询的过程中，大部分的相似性比对计算都是没有必要的。如果 2 个用户的活动空间根本没有交集，那么也就没有必要比对 2 人的轨迹。因此，设计了基于区域驻留时长的无关用户过滤算法。将一个城市中所有基站的位置，通过 mean-shift 算法聚类为上百个区域，然后为每个用户统计其在各个区域的驻留时长。这样就得到了用户的驻留特征。

得到了每个用户的驻留特征后，就可以通过余弦相似度的方法过滤掉无关用户。具体来讲，当查询某个用户的相似轨迹时，首先根据查询的用户的驻留特征，与其他用户的特征进行余弦相似度计算。该过程的计算复杂度为 $O(n)$，远小于轨迹相似度比对的 $O(n^2)$。实验中保留相似度大于 0.1 的用户。

这样，在实际 30 万人的手机基站轨迹数据库中，每个用户只需要比对几百个用户的轨迹，而不用一一比对。这样就加快了相似轨迹检索的速度。

对于余下的用户，我们提出了基于弗雷歇距离的滑动窗口算法。其中，弗雷歇距离 $F(A,B)$ 定义为

$$F(A,B) = inf_{\alpha,\beta}\max_{t\in[0,1]}\{d(A(\alpha(t))),B(\beta(t))\} \tag{10}$$

(10) 式中: α,β 分别为单位区间上的 2 个重新参数化(reparameterization) 函数。

3　城市有轨交通网络短时客流预测

短时交通流预测是智能交通系统(intelligent traffic system，ITS) 的重要方面，其能够缓解交通拥堵，减少交通拥堵带来的污染和能源消耗。同时，如果没有交通预测，人们只能依靠现有的交通状态推断未来的交通情况，这样的推断是不能满足实际需求的。相反地，交通流预测可以利用历史交通流和现有交通情况来预测未来交通情况[28]。

由于短时交通预测的重要性，大量的研究者一直在探索交通流预测的理论并提出了相应的预测方法。虽然交通流数据复杂多变，影响其的因素众多，如天气状况、节假日、特殊的事件等，但是纵观现有的交通流预测方法，可以发现，交通流预测方法大致可以分为以下 2 个方面。

1) 时间序列预测模型: 研究者通过构造数学模型或者利用神经网络等方法，捕捉交通流数据预测未来与过去之间交通状态的关系

2) 多模式预测模型: 研究者利用交通流数据的多模式特性来预测，如空间相似性和传播性、周与天等模式下的周期性等。其中，空间相似性和传播性是指: 交通流数据在邻近的路段和截面，表现出一定空间相似性和传播性，这一信息有助于更加准确地预测未来的交通状态。对于周、天等模式下的周期性是指: 由于人们出行的习惯与大部分工作的性质，交通流数据在天与天之间、周与周之间呈现了强烈的周期性，有效利用周期性能够提高预测的准确度[29]。

下面我们将具体分析现有交通流预测方法。

3.1　时间序列预测模型

由于交通流数据随时间变化，故早期的预测方法都将其构建成时间序列，再利用数学模型，挖掘交通流数据的时间模式变化特征来预测未来的交通流量，其预测模型可以抽象为

$$[w_t,w_{t+1},\cdots,w_{t+d}] = f([w_{t-1},w_{t-2},\cdots,w_{t-h}]) \tag{11}$$

(11) 式中: d 表示预测区间的长度; h 表示历史数据的长度; $f()$ 表示未来预测数据与历史数据之间的关系函数。这种预测模型，数据只有一个时间维度，数据的表达形式呈向量流。

在时间序列预测模型中，最常见的是自回归积分滑动平均模型(autoregressive integrated moving average model，ARIMA)[30]，这种方法是基于时间序列的自相关分析来捕捉交通流数据未来与历史的关系。由于交通流数据在时间上表现为强烈的非线性，而 ARIMA 等线性模型无法捕捉交通流数据的非线性变化，因此，一些研究者提出了大量的非线性预测方法，如 M. Castro-Neto 等[31] 提出了自向量回归模型(support vector regression) 来挖掘交通流数据的非线性变化。为了减少自向量回归模型的计算复杂度，James Haworth 等[32] 提出了线核岭(online kernel ridge) 回归模型。同时，一些研究者也开始采用神经网络(neural network，NN) 来挖掘未来交通状况和历史数据之间的非线性关系。其中比较典型的一个工作是文献[33]，在文中，他们提出了一种基于神经网络的交通流模型，其目的是将其纳入实时自适应城市交通控制系统。建模分为 2 个部分，首先，交通流由局部神经网络在单个信号链路上建模; 其

次,基于局部神经网络之间的通信,交通流量在广泛的接口网络上建模。同时,基于模拟数据,文章也总结了应用于交通流量建模的神经网络的潜力。

文献[34]比较和结合了2种典型的时间序列预测方法:ANN(artificial neural network)和ARIMA。在ANN模型中,过去的事件能够被分析并且模式能够被推断出,利用这些模式就可以进一步预测出未来的交通流量。该工作指出了,在ARIMA或ANN模型的传统结构中,通常都是假设先前的模式将会被延续到未来的交通行为中,但是如果这个假设不成立,则预测效果将会不佳。因此,该工作引入一种判断调整机制来影响纠正少量和不规则的未来事件。实验证明判断调整技术有助于减少预测误差。此外,ANN和ARIMA结合的模型明显优于他们各自的基础模型。与其他研究不同的是,该工作指出了ARIMA模型优于ANN模型。

3.2 多模式预测模型

对于多模式预测方法,研究者在原来的时间序列数据基础上增加了空间信息或周期性信息。这是因为在交通流中邻近的交通流数据存在着一定的关联性,如在交通路网中道路上游的交通流量往往决定着下游的交通流量,而当交通流量较大时,下游交通流量又会反过来影响上游交通流量[34]。因此,单靠时间维度上的预测模型往往会忽略空间上的特性,导致交通流数据信息的缺失,使得预测准确率下降。故有效地利用时空特性能够得到较好的预测性能,相应的预测模型为

$$\begin{bmatrix} w_{s_1,t} & \cdots & w_{s_1,t+d} \\ \vdots & & \vdots \\ w_{s_j,t} & \cdots & w_{s_j,t+d} \end{bmatrix} = f\left(\begin{bmatrix} w_{s_1,t-h} & \cdots & w_{s_1,t-1} \\ \vdots & & \vdots \\ w_{s_j,t-h} & \cdots & w_{s_j,t-1} \end{bmatrix}\right) \tag{12}$$

(12)式中:s_i表示第i个预测位置;d表示预测区间的长度;h表示历史数据的长度。这样的数据表现形式即为矩阵流数据。

典型的多模预测方法是Van Lint等[35]提出的状态空间神经网络,在基础的神经网络模型中加入了空间信息,即路段各个截面的交通状态,以提高预测的准确度。文献[36]分析了交通流量的天模式,并确定天模式是否能够改善流量预测。另外,该项工作比较了利用传统模型预测交通流量与减去天模式残差后的流量的性能,发现后一种情况预测效果有着明显的改善。

文献[37]提出了一种基于移动平均(moving average,MA),指数平滑(exponential smoothing,ES),ARIMA和神经网络模型的交通流预测组合方法。该方法将原始的交通流时间序列构造为以周为周期的时间序列、以天为周期的时间序列以及以小时为周期的时间序列,然后分别利用MA,ES和ARIMA来对3个相关时间序列进行预测。最后将预测出来的序列输入神经网络中得出最后的预测效果。该项工作证明了组合模型在提高交通流预测方面可以带来实质性的好处,同时多模式的预测方式提高了预测的准确度。

文献[34]提出一个基于深度学习的端到端结构的模型来预测城市区域中进客流和出客流。具体来说,该方法利用残差神经网络来建模人流的时间接近度、周期性和趋势特征。对于每个属性,他们都设计了一个残差卷积单元分支,每个单元都模拟人群流量的空间属性。该方法基于数据动态聚合了3个残差神经网络的输出并给不同分支和区域分配了不同的权重。同时,该方法还加入额外的影响因子,如天气和星期的影响,来进一步预测最终的人流量以提高算法的鲁棒性和准确度。

表1总结了已有工作考虑的因素。

表1 客流预测的相关工作

Tabl. 1 Related work on traffic prediction

考虑的因素	相关工作	解释
不同的周期性	文献[34]	
天气	文献[39]	
节假日	文献[29]	周末与工作日的区别
特殊事件	文献[38]	大型活动或事故的影响
道路结构	文献[35]	道路上下游的影响
一天内的周期性	文献[36-37]	一天的数据可分成几个时间段

3.3 笔者的工作

对于复杂的交通网络产生的交通流预测问题,只依靠交通流时间序列的特征已经不能满足现在的预测精度需求。大多数研究者会偏向提取交通流数据的多个特征并加入其他相关信息以提高预测的精度。当加入越多信息时,精度可以相应得到提高,但是训练的复杂度也会相应变大,需要大量的计算时间去完成一次预测。这无法满足实际的实时预测需求。所以,研究者都尝试寻求一种均衡时间复杂度和预测精度的方法。另一方面,现在的研究很多都是基于单个站点或者小块路段,对整个交通路网的

交通流的精准预测还比较困难。因此,能在解决复杂度的情况下,完成整个路网的交通流实时预测,将是交通流预测研究的一个重要进展。

笔者在这些方面也做了相应的研究。区别于以往的向量形式和矩阵形式交通流预测算法,设计了一个新型的交通流预测算法,能充分利用交通流量的多模特性来提高预测准确度。具体来讲,针对复杂的交通流数据,构建了一个不同于以往向量流或矩阵流形式的数据模型(张量模型),能够有效表征交通流数据的多模特性,充分利用数据中所包含的空间信息、周期信息以及时间变量信息。利用张量分解技术挖掘交通流张量的多模特性,以证明张量模型的有效性和可靠性,为进一步预测提供理论分析。对于交通流量预测问题,笔者将其视作张量填充,即利用张量中已知数据对未知数据进行预测。我们的模型主要是针对大客流预测,因此,本文中将其称为大客流张量模型。

我们在上海地铁数据上评估了大客流张量模型,并与已有的数学模型(ARIMA,SARIMA)和机器学习模型(SVR,NN,LSTM)进行了比较。

选用平均绝对百分比误差(mean absolute percentage error,MAPE)和平均绝对误差(mean absolute error,MAE)作为评估指标。实验结果如表 2 所示。表 2 中第 2 列是正则化的 MAE,可以看到大客流张量模型比已有性能最好的 SRIMA 模型仍然低 34%。对于 MAPE,张量填充模型也提高了 1.04% ~27.78%。

表 2　预测性能对比

Tab. 2　Forecasting performance

模型	*MAE*	*MAPE*/%
大客流张量模型	1.00	8.65
SARIMA	1.34(+34%)	9.69(+1.04)
LSTM	2.32(+132%)	12.47(+3.82)
历史平均	1.90(+90%)	14.61(+5.96)
NN	1.53(+53%)	15.59(+6.94)
ARIMA	3.74(+274%)	25.40(+16.75)
SVR	4.94(+394%)	36.43(+27.78)

4　城市公交系统突发客流早期预警

城市地铁中的突发客流是指短时间内聚集大量的乘客,这会给交通系统带来不良影响并诱发踩踏事件发生。对突发客流早期预警可以使城市管理部门提前做出应急准备,防止紧急事件发生。目前城市地铁管理系统仍然缺乏有效的预警工具[39]。

针对交通(乘客)流量已有很多相关研究,然而更多的研究专注于交通流量短时预测和交通流量监控。交通流量短时预测方法对于常规交通(乘客)流量有较好的性能,然而对于突发交通(乘客)流量预测精度较差。交通流量监控方法仅对当下的情景做出感知,并假设当前的情况能够对短期的将来情况提供最好的估计,然而这一假设不适合突发交通流量。已有研究缺乏针对突发交通流量的早期预警。

以上研究方法已在许多研究问题中应用,例如交通系统客流量、高速公路车流量、区域人群流量。相关研究运用了多种数据源,例如公交卡交易数据、GPS 轨迹数据、视频监控数据等。

下面将具体分析现有交通流预测方法。

4.1　短时预测方法

短时预测多使用数学模型挖掘模式,再根据捕捉的模式预测未来的交通流量。Wei 等[40]使用一个包括经验模态分布和反向传播神经网络的混合模型预测地铁系统中的乘客流量。Li 等[41]提出了多尺度径向基(multi scale radial basis function,MSRBF)网络预测特殊事件场景下的客流量。Sun 等[42]提出了混合模型(小波支持向量机),把客流量分解成不同频率的流量序列,并针对各自特征分别建模预测地铁系统中的客流量。

针对城市中的区域人流量,也有相关的短时预测研究,例如文献[34,43-44]使用 GPS 轨迹数据利用深度残差网络等模型建模预测城市每个区域的进出人流量。Zhang 等[45]使用移动终端(手机)采集数据,提出了一个混合模型预测城市区域的人群流量。

对于高速公路车流量的短时预测,也有一些相关研究。Abadi 等[46]利用交通数据提出了一个基于自回归模型算法预测交通网络中所有路段的车流量。Lv 等[47]提出了一个基于深度学习的车流量短时预测方法,该方法同时考虑了时间和空间的相关性。Hou 等[48]则关注一城市工作区的车流量预测。

4.2　监控方法

很多研究专注于对交通系统客流量、区域人群客流量、高速公路车流量的监控。

Liu 等[49]使用视屏监控数据,利用计算机视觉、数字图像处理等技术来监控客流量,以防止踩踏事

件的发生。同时利用 GIS(geographic information systems)技术提出了一个客流监控算法,该算法可以动态地展示城市交通网络中实时的客流量分布,并刻画其变化趋势。并分析不同指标的适用条件和使用范围,最后提出客流预警阈值方面的参考建议。

Liang 等[50]对上海的一条老街的人群流量进行监控,并通过图像处理方法使用视频监控数据计算不同时间段的人群客流量。Xu 等[51]提出了一个考虑多因素的区域人群流量监控方法,考虑区域人群密度、区域人群更迭速率、区域进出人流比、平均速度等多因素对大型商业区的人群流量进行监控。

Quinn 等[52]使用视频监控数据对道路的车流量数据进行监控,提出一种基于概率推理的方法来使交通监控的标准方法变得更加健壮。此外该方法可以不需要进行车辆分割过程,将交通道路视为流体,并估计流量,而不是跟踪单个车辆。该方法即使在有噪音的情况下也可以准确地监控交通流量。

在以上的监控方法中,当流量超过了阈值系统会发出报警。但是这些方法可以做到实时监控却难以提前发出报警,而突发交通流量多是很短时间之内发生的,因此,监控方法难以给城市管理人员足够的准备时间对突发交通流量采取措施。

4.3 早期预警方法

有关交通流量早期预警的研究较为有限。Zhou 等[53]设计了一个新颖的方法可以提前预警某区域的大量人群。他们利用百度地图中查询数量与定位用户数量之间的强相关模式,发现当某地址有大量的查询数据时,一段时间后往往会有大量人群聚集,并基于此提出区域人群早期预警模型。但他们的工作集中在固定区域的人流预测,而不适合轨道交通系统中的客流。

表3是对以上工作的总结。

4.4 笔者的工作

我们定义了交通系统突发客流早期预警问题,针对突发客流,希望能够实现以下2个目标。

1)提前告警。提前足够长的时间发出公共交通系统中突发客流即将到来的告警,使城市管理相关人员有足够长的时间采取措施做出防范。

2)定量预估。在发出预警时或其后的一段时间内,作出对突发客流峰值时间、峰值数量的预估。

然后我们对公交卡刷卡数据进行了初步探索,不仅对客流量(进客流)进行了分析,也同时也考虑了出客流,进出信息可以通过交易金额来推断。图3展示了“上海体育场”地铁站某一天的进出客流量,浅色线为出客,深色线为进客流。如图2所示,我们得知,有一场足球赛于当天17:00—18:50在该地铁站附近举办,这是导致突发客流的原因。

表3 交通早期预警相关研究

Tab. 3 Related work about early warning on traffic

研究方法	研究问题	数据源
短期预测	交通系统客流量	公交卡交易数据[40-42]
	区域人群流量	GPS 轨迹数据[34,43-44] 移动设备数据[45]
	高速公路车流量	交通流量仿真数据[46] 高速公路车流量数据[47-48]
监控	交通系统客流量	监控视频数据[49] 公交卡交易数据[54]
	区域人群流量	监控视频数据[50-51]
	高速公路车流量	高速公路车流量数据[55] 监控视频数据[52]
早期预警	区域人群流量	电子地图查询数据[53]

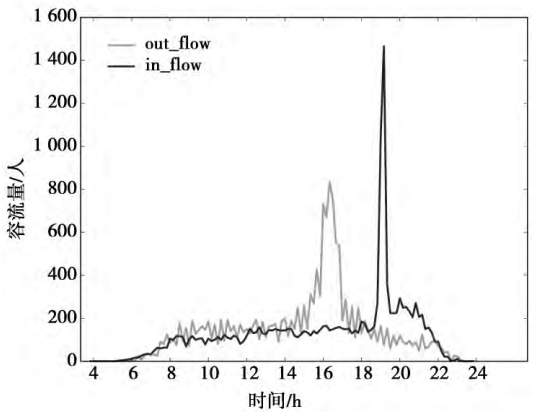


图2 上海体育场地铁站的进出客流量

Fig. 2 Passengers of Shanghai stadium

可以看到,在当有大型活动举办时,离举办场馆较近的地铁站的进出客流量在活动前后有2个“尖峰”—异常出客流量和突发进客流量。通常情况下,前者持续时间较长,峰值通常较小,对交通系统几乎没有负面影响。相比之下,后者在较短的时间内有非常高的峰值,并可能对公共安全造成危害。此外,两者之前通常有2小时以上的延迟。

由此可见,如果仅靠历史客流信息结合数学模型很难提前预知突发客流(深色线)。但是,考虑了出客流信息以及人群行为之后,可以较明显地看到

突发客流的模式，并基于此提出突发客流早期预警框架。

笔者的框架考虑了宏观乘客流量和微观个人出行行为。它由 2 个模型组成: 异常流量检测告警模型和突发流量峰值估计模型。异常流量检测告警的目标是通过弹性滑动窗口在线检测异常出客流量，基于移位小波树(shifted wavelet tree，SWT) 的时间序列模型[56]。当检测到异常出客流量时，进行告警，提醒未来一段时间可能会有大量突发的客流。同时，突发流量峰值估计模型开始计算即将到来的突发客流的峰值时间和峰值数量。突发流量峰值估计模型使用多元回归模型计算峰值时间，并结合流量预测模型和物理模型所组成的混合模型来估计峰值量。

笔者使用上海市 2015 年 4 月的公卡交易数据评估笔者的框架和模型。正则化的实验结果如表 4 所示。使用了 3 个典型地铁站来对笔者的模型进行评估，并与其他代表性模型如 SARIMA，SVR，NN 进行比较，选用 RMSE(root mean square error) 误差作为评估指标。从表 4 可见，其他模型峰值 RMSE 误差是笔者模型的 1.28 ~ 14.58 倍。

表 4　峰值流量预估的 RMSE

Tab. 4　Normalized RMSE of peak volume prediction

	我们的模型	SARIMA	SVR	NN
上海体育馆	1	14.58	14.68	11.38
上海体育场	1	8.87	8.70	8.79
中华艺术宫	1	1.36	1.33	1.28

5　结束语

笔者调研了面向公共安全的时空数据挖掘的研究进展，并针对 2 类数据集，分别介绍了基于个体移动模序异常的大规模活动识别、城市有轨交通网络短时客流预测以及城市公交系统突发客流早期预警，同时介绍了笔者的最新研究进展。公共安全管理对城市的发展和市民生活无比重要，利用时空数据可以真切地感知并预警潜在的危险事件，因此该研究具有重要的现实意义。

参考文献:

[1] 赵汗青. 中国现代城市公共安全管理研究 [D]. 长春: 东北师范大学，2012.
ZHAO Hanqing. Study of China's Modern City Public Safety Management [D]. Changchun: Northeast Normal University, 2012

[2] BARABASI A L. The origin of bursts and heavy tails in human dynamics [J]. Nature, 2005, 435(7039): 207-211.

[3] HADDADI H, HUI P, BROWN I. MobiAd: private and scalable mobile advertising [C] //Proceedings of the fifth ACM international workshop on Mobility in the evolving internet architecture. New York, NY, USA: ACM, 2010: 33-38.

[4] GONZALEZ M C, HIDALGO C A, BARABASI A L. Understanding individual human mobility patterns [J]. Nature, 2008, 453(7196): 779-782.

[5] SCHNEIDER C M, BELIK V, COURONNé T, et al. Unravelling daily human mobility motifs [J]. Journal of The Royal Society Interface, 2013, 10(84): 246-253.

[6] GONZALEZ M C, HIDALGO C A, BARABASI A L. Understanding individual human mobility patterns [J]. Nature, 2008, 453(7196): 779-782.

[7] YANG S, YANG X, ZHANG C, et al. Using social network theory for modeling human mobility [J]. IEEE network, 2010, 24(5): 6-13.

[8] 卢扬. 人类移动行为模式研究 [D]. 成都: 电子科技大学，2015.
LU Yang. Study of Human Mobility Pattern [D]. Chengdu: University of Electronic Science and Technology of China, 2015.

[9] ZHONG Y, YUAN N J, ZHONG W, et al. You are where you go: Inferring demographic attributes from location check-ins [C] //Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2015: 295-304.

[10] 戴蓉蓉，朱海红，李霖. 基于 ARIMA 模型的市内人群移动预测 [J]. 测绘工程，2016，25(2): 38-41.
DAI Rongrong, ZHU Haihong, LI Lin. Intra-urban human mobility prediction based on ARIMA model [J]. Engineering of Surveying and Mapping, 2016, 25(2): 38-41.

[11] 杨喜平，方志祥，赵志远，等. 城市人群聚集消散时空模式探索分析——以深圳市为例 [J]. 地球信息科学学报，2016，18(4): 486-492.
YANG X, FANG Z, ZHAO Z, et al. Exploring Urban Human Spatio-temporal Convergence-Dispersion Patterns: A Case Study of Shenzhen City [J]. Journal of Geo-Information Science, 2016, 18(4): 486-492.

[12] ADAM A, RIVLIN E, SHIMSHONI I, et al. Robust real-time unusual event detection using multiple fixed-location monitors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(3): 555-560.

[13] 郭婷. 大规模群体人数检测算法研究[D]. 西安: 西安科技大学,2014.
GUO T. Research on Algorithm of Counting High-Density Crowd [D]. Xi'an: Xi'an University of Science and Technology, 2014.

[14] HAGBERG A, SWART P, CHULT D S. Exploring Network Structure, Dynamics, and Function Using NetworkX [C] // Proceedings of the 7th Python in Science conference (SciPy 2008). [S. l.]: Conference Publication, 2008: 11-15.

[15] ZHENG Yu, YUAN N J,ZHENG K, et al. On discovery of gathering patterns from trajectories [C] // Proceeding ICDE '13 Proceedings of the 2013 IEEE International Conference on Data Engineering (ICDE 2013) Washington, DC, USA: IEEE Computer Society, 2013: 242-253.

[16] TRAAG V A, BROWET A, CALABRESE F, et al. Social event detection in massive mobile phone data using probabilistic location inference [C] //Privacy, security, risk and trust (PASSAT) and 2011 IEEE Third International conference on social computing (SocialCom), 2011 IEEE Third International Conference on. Boston, MA, USA: IEEE, 2011: 625-628.

[17] DONG Y, PINELLI F, GKOUFAS Y, et al. Inferring unusual crowd events from mobile phone call detail records [C] // Proceeding ECMLPKDD'15 Proceedings of the 2015th European Conference on Machine Learning and Knowledge Discovery in Databases. Switzerland: Springer, 2015: 474-492.

[18] WITAYANGKURN A, HORANONT T, SEKIMOTO Y, et al. Anomalous event detection on large-scale gps data from mobile phones using hidden markov model and cloud platform [C] //Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication. New York, NY, USA: ACM, 2013: 1219-1228.

[19] CALABRESE F, PEREIRA F C, DI LORENZO G, et al. The geography of taste: analyzing cell-phone mobility and social events [C] //International Conference on Pervasive Computing. Berlin Heidelberg: Springer, 2010: 22-37.

[20] AGRAWAL R, FALOUTSOS C, SWAMI A N. Efficient similarity search in sequence databases [C] //Proceeding FODO '93 Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms. London, UK: Springer-Verlag, 1993: 69-84.

[21] CHEN Lei, ÖZSU M T, ORIA V. Robust and fast similarity search for moving object trajectories [C] //Proceedings of the 2005 ACM SIGMOD international conference on Management of data. New York, NY, USA: ACM, 2005: 491-502.

[22] VLACHOS M, KOLLIOS G, GUNOPULOS D. Discovering similar multidimensional trajectories [C] //Data Engineering, 2002. Proceedings. 18th International Conference on. San Jose, CA, USA, USA: IEEE, 2002: 673-684.

[23] CHEN Lei, RAYMOND Ng. On the marriage of Lp-norms and edit distance [C] // Proceeding VLDB '04 Proceedings of the Thirtieth international conference on Very large data bases. Toronto, Canada: VLDB Endowment 2004: 792-803.

[24] MA C, LU H, SHOU L, et al. KSQ: Top-k similarity query on uncertain trajectories [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(9): 2049-2062.

[25] LI Quannan, ZHRNG Yu, XIE Xing, et al. Mining user similarity based on location history [C] //Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems. New York, NY, USA: ACM, 2008.

[26] JEUNG H, YIU M L, JENSEN C S. Trajectory pattern mining [M] //Computing with spatial trajectories. New York: Springer, 2011: 143-177.

[27] LEE, Jae-Gil, HAN Jiawei, WHANG Kyu-Young. Trajectory clustering: a partition-and-group framework [C] // Proceedings of the 2007 ACM SIGMOD international conference on Management of data. New York, NY, USA: ACM, 2007:594-604.

[28] BOLSHINSKY E, FREIDMAN R. Traffic flow forecast survey [R]. Hefa, Israel: Technion-Israel Institute of Technology. Technical Report, 2012.

[29] 伍元凯. 基于动态张量填充的短时交通流预测研究[D]. 北京: 北京理工大学,2015.
WU Yuankai. Short-tem Traffic Prediction based on Dynamic Tensor Completion [D]. Beijing: Beijing Institute of Technology, 2015.

[30] van der VOORT M, DOUGHERTY M, WATSON S. Combining Kohonen maps with ARIMA time series models to forecast traffic flow [J]. Transportation Research Part C: Emerging Technologies, 1996, 4(5): 307-318

[31] CASTRO-NETO M, JEONG Y S, JEONG M K, et al. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions [J]. Expert systems with applications, 2009, 36(3): 6164-6173.

[32] HAWORTH J, SHAWE-TAYLOR J, CHENG T, et al. Local online kernel ridge regression for forecasting of urban travel times [J]. Transportation Research Part C: Emerging Technologies, 2014(46): 151-178.

[33] LEDOUX C. An urban traffic flow model integrating neural networks [J]. Transportation Research Part C: Emerging Technologies, 1997, 5(5): 287-300.

[34] ZHANG J, ZHENG Y, QI D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction [C] // Proceeding of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17). arXiv preprint arXiv: 1610.00081. [S. l.]: AAAI, 2017, 2016.

[35] van LINT J W C, HOOGENDOORN S P, van ZUYLEN H J. Accurate freeway travel time prediction with state-space neural networks under missing data [J]. Transportation Research Part C: Emerging Technologies, 2005, 13(5): 347-369.

[36] CHEN C, WANG Y, LI L, et al. The retrieval of intra-day trend and its influence on traffic prediction [J]. Transportation research part C: emerging technologies, 2012(22): 103-118.

[37] TAN M C, WONG S C, XU J M, et al. An aggregation approach to short-term traffic flow prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2009, 10(1): 60-69.

[38] CHANG S C, KIM R S, KIM S J, et al. Traffic-flow forecasting using a 3-stage model. In Intelligent Vehicles Symposium [C] // Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE, Dearborn, MI, USA: IEEE, 2000: 451-456.

[39] YANG J T. Safety risk analysis and countermeasures study on regular mass passenger flow of china's urban subway [J] Procedia Engineering, 2016(135): 175-179.

[40] WEI Y, CHEN M C. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks [J]. Trans-portation Research Part C: Emerging Technologies, 2012, 21(1): 148-162.

[41] LI Y, WANG X, SUN S, et al. Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks [J]. Transportation Research Part C: Emerg- ing Technologies, 2017(77): 306-328.

[42] SUN Y, LENG B, GUAN W. A novel wavelet-svm short-time passenger flow prediction in beijing subway system [J]. Neurocomputing, 2015(166): 109-121.

[43] ZHANG J, ZHENG Y, QI D, et al. Dnn-based prediction model for spatio-temporal data [C] // Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York, NY, USA: ACM, 2016: 92.

[44] HOANG M X, ZHENG Y, SINGH A K. Fccf: forecasting citywide crowd flows based on big data [C] //Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York, NY, USA: ACM, 2016: 6.

[45] ZHANG K, WANG M, WEI B, et al. Identification and prediction of large pedestrian flow in urban areas based on a hybrid detection approach [J]. Sustainability, 2016, 9(1): 36.

[46] ABADI A, RAJABIOUN T, IOANNOU P A. Traffic flow prediction for road transportation networks with limited traffic data [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 653-662.

[47] LV Y, DUAN Y, KANG W, et al. Traffic flow prediction with big data: a deep learning approach [C] //IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 865-873.

[48] HOU Y, EDARA P, SUN C. Traffic flow forecasting for urban work zones [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(4): 1761-1770.

[49] LIU S, ZHU Z, CHENG Q, et al. Analysis and design of public places crowd stampede early-warning simulating system [C] //Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), 2016 International Conference on. Wuhan, China: IEEE, 2016: 210 - 213.

[50] LIANG J, YANG J T, WU P Y. A graded pedestrian flow early warning for an ancient street [C] //Procedia Engineering, 2016(135): 118-122.

[51] XU X, MA Y, LI T, et al. Risk early-warning study of passenger flow in business district [C] //Emergency Management and Management Sciences (ICEMMS), 2010 IEEE International Conference on. Beijing, China: IEEE, 2010: 310-313.

[52] QUINN J A, NAKIBUULE R. Traffic flow monitoring in crowded cities [C] //2010 AAAI Spring Symposium Series, Artificial Intelligence for Development. [S. l.]: AAAI Publications, 2010: 73-78.

[53] ZHOU J, PEI H, WU H. Early warning of human crowds based on query data from Baidu map: Analysis based on shanghai stampede [EB/OL]. (2016-03-22). https://arxiv.org/abs/1603.06780.

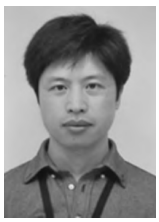
[54] BAI Li, WANG Fuzhang, ZHANG Ming. Urban rail transit network passenger flow monitoring and early warning system based on GIS [J]. Urban Rapid Rail Transit, 2013, 26(6): 56-59.

[55] GALLO M, SIMONELLI F, de LUCA G, et al. An artificial neural network approach for spatially extending road traffic monitoring measures [C] //Environmental, Energy, and Structural Monitoring Systems (EESMS), 2016

IEEE Workshop on. Bari, Italy: IEEE, 2016: 1-5.

[56] ZHU Y, SHASHA D. Efficient elastic burst detection in data streams [C] //Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, 2003: 336-345.

作者简介:



王永坤(1977-) ,**男,山东诸城人,博士,主要研究方向为大规模分布式系统设计与实践、可扩展非结构化数据库理论、多源数据的情报综合与分析等**。E-mail: ykw@ sjtu. edu. cn。

王海洋(1990-) ,**男,黑龙江哈尔滨人,博士生,研究方向为多源异构时空数据挖掘**。E-mail: 010350180@ sjtu. edu. cn。

潘平峻(1994-) ,**男,浙江台州人,硕士生,研究方向为多源异构时空数据挖掘**。E-mail: panpingjun@ sjtu. edu. cn。

李龙元(1993-) ,**男,陕西汉中人,博士生,研究方向为多源异构时空数据挖掘**。E-mail: jeffli@ sjtu. edu. cn。

金耀辉(1971-) ,**男,安徽安庆人,教授,博士,博士生导师,研究方向为云计算网络架构、数据管理与机器学习、时空数据挖掘与应用、公众参与的开放创新等**。E-mail: jinyh@ sjtu. edu. cn。

(编辑: 魏琴芳)