

城市场景中车联网时空数据分析及其通达性方法

程久军¹, 原桂远¹, 崔杰², 周爱国³, 吕博⁴, 李光耀⁵

(1. 同济大学嵌入式系统与服务计算教育部重点实验室, 上海 200092; 2. 安徽大学计算机科学与技术学院, 安徽 合肥 230601;
3. 同济大学机械与能源工程学院, 上海 200092; 4. 上海师范大学天华学院, 上海 201815;
5. 同济大学电子与信息工程学院, 上海 200092)

摘 要: 为了解决城市场景中车联网时空数据异构以及单个基础设施范围内存在连通效率低下的问题, 提出一种车联网时空数据分析及其通达性方法。首先, 给出基于噪声去除和数据填充的时空数据分析方法, 构建基于张量因子聚合的神经网络预测车辆之间的连通强度; 然后, 基于车联网连通强度给出有基础设施车联网的通达性方法。仿真实验结果表明, 基于张量因子聚合的神经网络可以有效预测车辆之间的连通强度, 所提方法可以有效减少连通冗余和路边基础设施负载。

关键词: 车联网; 时空数据分析; 通达性; 城市场景

中图分类号: TP391.9

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021110

Spatio-temporal data analysis and accessibility method for IoV in an urban scene

CHENG JiuJun¹, YUAN Guiyuan¹, CUI Jie², ZHOU Aiguo³, LYU Bo⁴, LI Guangyao⁵

1. Ministry of Education Key Laboratory of Embedded System and Service Computing, Tongji University, Shanghai 200092, China
2. School of Computer Science and Technology, Anhui University, Hefei 230601, China
3. School of Mechanical Engineering, Tongji University, Shanghai 200092, China
4. Tianhua College, Shanghai Normal University, Shanghai 201815, China
5. College of Electronic and Information Engineering, Tongji University, Shanghai 200092, China

Abstract: In order to solve the problems of diversity spatio-temporal data and low connectivity efficiency in a single road side unit for Internet of vehicles (IoV) in an urban scene, a spatio-temporal data analysis and accessibility method was presented. First, a spatio-temporal data analysis method based on de-noising and data filling was introduced, and a tensor factor aggregation-based neural network was constructed to predict connectivity intensity among vehicles. Then, a connectivity intensity prediction-based accessibility method was proposed. The simulation results demonstrate that the proposed connectivity intensity prediction method can accurately predict connectivity intensity among vehicles, and the proposed accessibility method can effectively reduce connectivity redundancy and loads of road side units.

Keywords: Internet of vehicles, spatio-temporal data analysis, accessibility, urban scene

收稿日期: 2020-11-02; 修回日期: 2021-04-19

通信作者: 周爱国, zhouaiguo@tongji.edu.cn

基金项目: 国家自然科学基金资助项目(No.61872271); 中央高校基本科研业务费重点领域学科交叉重大基金资助项目(No.22120190208); 网络与交换技术国家重点实验室(北京邮电大学)开放课题基金资助项目(No.SKLNST-2020-1-20)

Foundation Items: The National Natural Science Foundation of China (No.61872271), The Fundamental Research Funds for the Central Universities (No.22120190208), Open Foundation of State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) (No.SKLNST-2020-1-20)

1 引言

车联网是一种可以实现车辆与其他网络通信的动态网络。在车联网中,车辆通过传感器收集自身及周围车辆、道路和环境信息,对采集的时空数据进行分析并发布,为用户提供各种应用服务,例如车辆远程诊断^[1]、视频下载^[2]、隐私保护^[3-4]、数据传输^[5-6]等功能。因此,车联网时空数据分析及其通达性至关重要。

随着传感器技术的发展,车辆可以在运动过程中获取自身状态并感知周围环境。同时,车辆可以通过车车通信获取其他车辆的速度、加速度、地理位置等信息。然而,目前缺少统一的车辆数据采集存储标准,不同传感器存储的数据结构可能相差很大。另外,在城市场景中,由于建筑物的遮挡和其他因素干扰,时空数据可能存在噪声和数据丢失的问题。需要有效的车联网时空数据分析方法进行数据的整合、噪声去除和填充,为车联网应用提供数据支持。

按照有无路边基础设施支持,车联网可分为有基础设施的车辆网络和车辆自组织网络。与车辆自组织网络不同,在城市场景中车联网具有路边基础设施支持,车辆之间不仅可以通过车车通信实现连通,也可以跨路边基础设施实现连通。然而,路边基础设施带宽有限,不能为所有车辆提供数据转发服务。因此,需要实时准确地检测车联网中不能通过车车通信实现连通的子网,并通过路边基础设施实现车辆之间的连通,保证拓扑结构发生变化时车联网的通达性。

与高速公路场景下的车联网相比,城市场景中的车联网具有以下特点。1) 车辆运动受到其他车辆和红绿灯的影响,行车环境更加复杂。2) 车辆数量较多,车联网规模较大。同时,由于红绿灯和路网结构的影响,车辆可能存在分布不均匀的现象。3) 由于建筑物的遮挡,距离小于通信范围的车辆之间可能不能直接建立通信链路。同时,受各种干扰因素的影响,车辆之间的通信链路会频繁地断开重连,预估车辆之间在未来一段时间内保持连通的可能性更加困难。因此,高速公路场景下的车联网通达性不完全适用于城市场景中的车联网。

本文提出一种城市场景中车联网时空数据分析及其通达性方法,主要贡献包括以下两点:1) 提

出一种车联网时空数据分析方法;2) 提出一种城市场景中车联网通达性方法。

2 相关工作

在车联网时空数据分析方面,文献[7]总结了当前车联网时空数据分析面临的挑战:找到合适的过滤器提取有价值信息;将无用和冗余信息剔除;利用有效的方式对数据进行表示和分析;建立有效的预测模型进行交通管理。当车联网网络负载不足以支撑车辆间数据传输时,文献[8]将路网进行划分,将车辆映射到划分出的格子中,以交通密度、带宽、时延和花费作为指标对车辆的网络接入进行评价,从而对车辆接入传统网络或车联网以及车联网中哪个节点进行决策,最终使车辆获得较好的服务质量。针对传统的关系型数据库和数据库管理系统对给定时间周期和数据维度的查询支持度不够的问题,文献[9]在静态R树结构上使用希尔伯特曲线和HBase技术,使模型对范围查询和k近邻查询都有较好的性能表现。文献[10]分析了车联网的特性,将车联网中的数据流比拟为流体,推导出网络特性描述方程,然后,基于网络特性描述方程推导出车联网网络容量计算模型,该模型分析了网络规模和网络时延对网络容量的影响,得出了网络容量下降的原因。文献[11]提出了一种基于并行数据挖掘的轨迹数据模式发现方法,并构建了知识模型捕捉用户移动行为。该轨迹数据模式发现方法可用于车辆位置预测、旅行推荐、智能交通管理等应用和服务。针对车辆自组织网络中网络拓扑结构多变和时空数据复杂的问题,文献[12]提出了一种基于贝叶斯联盟博弈和学习自动机的车辆节点联合处理时空数据的方法。

车辆快速移动导致车联网拓扑结构频繁变化和节点之间通信链路不稳定,这些特性使车联网通达性面临2个挑战:构建准确描述车辆运动的模型;基于车辆运动模型构建车联网通信协议实现车辆之间的互联互通。文献[13]研究了高速公路场景中车联网的连通性,分析了在特定路段上车辆之间连通概率与连通集直径、连通集数目、车辆密度和车辆传输距离之间的关系。分析结果表明,当车辆进入高速公路是泊松过程时,车联网中车辆位置满足伽马分布。通过统计道路的交通密度、交叉路口和路段等道路信息,文献[14]提出了一种道路感知路由协议,并研究了路由协

议的恢复机制和性能指标。实验结果表明,即使在交通密度较高的情况下,文献[14]协议依然无法避免网络分裂的现象,也不能保证网络整体连通;要达到减少网络分裂的目的,可借助反向运动车辆实现信息转发和增加节点通信距离。文献[15]分析了高速公路场景中车辆之间的连通性,研究了车辆进入高速公路的速率、车辆速度以及车辆离开高速公路的概率等参数对连通性的影响。针对大规模车联网互联互通耦合度低的问题,文献[16]通过一种车联网连通基的网络拓扑结构,提出了分布式连通基构造方法,针对大规模车联网通达性问题,作者结合平滑高斯-半马尔可夫移动模型研究了连通基的内部结构属性和动态特性,对动态环境下车联网的通达性进行评估。在大规模车联网中,连通基可用于转发数据包,进而实现车辆节点之间的连通。针对车辆自组织网络拓扑结构频繁变化导致的连通易变问题,文献[17]提出了一种基于自编码网络和循环神经网络的车辆自组织网络连通预测方法和一种基于连通预测的动态分簇方法,细化了簇内角色,研究了节点之间数据转发的代价,给出了车辆之间连通路径构造方法。

在车联网时空数据分析方面,大多数研究对车联网实时性以及拓扑结构变化快的特性考虑不足。另外,在车联网通达性研究方面,大多数研究集中在网络整体通达性,只总结各个参数对通达性的影响,缺少对节点间通达性的研究。

3 城市场景中车联网时空数据分析

3.1 相关定义

定义 1 城市场景中的车联网可表示为

$$G=(V,E,W,P,U,\varphi) \quad (1)$$

其中, $V=\{v_i|i=1,2,\dots\}$ 表示车辆集合; $E=\{e_j|j=1,2,\dots\}$ 表示车辆之间的连通关系集合; W 表示车辆之间连通强度集合,连通强度表示车辆之间在未来一段时间内保持连通的可能性; $P=\{\rho_k|k=1,2,\dots\}$ 表示路段集合; $U=\{u_l|l=1,2,\dots\}$ 表示路边基础设施集合; $\varphi:E \rightarrow W$ 表示连通强度映射函数。

定义 2 车辆 v_i 在 t 时刻的特征可表示为

$$f(v_i,t)=[s,a,p,\rho,u] \quad (2)$$

其中, s 表示 v_i 的速度, a 表示 v_i 的加速度, p 表示 v_i 的位置, ρ 表示 v_i 所处的路段, u 表示 v_i 关联的路边基础设施。

3.2 基于噪声去除和数据填充的时空数据处理

3.2.1 噪声去除

假设 v_i 在 t 时刻的位置为 p , 在 \hat{t} 时刻的位置为 \hat{p} , 如果 v_i 的速度与加速度方向相同, 则

$$\hat{p}=p+s \times(\hat{t}-t)+\frac{1}{2} a \times(\hat{t}-t)^2 \quad (3)$$

然而, 车辆的加速度与速度之间可能存在夹角, 式(3)估计的车辆位置可能不够准确。为了判断传感器采集的车辆位置是否准确, 给出车辆位置可信区域。可信区域的 4 个顶点的坐标为

$$\left\{\begin{array}{l} p_1=\left(\left[\begin{array}{cc} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{array}\right] p^T\right)^T \frac{|p+s t|-t}{|p|} \\ p_2=\left(\left[\begin{array}{cc} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{array}\right] p^T\right)^T \frac{|p+s t|+t}{|p|} \\ p_3=\left(\left[\begin{array}{cc} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{array}\right] p^T\right)^T \frac{|p+s t|-t}{|p|} \\ p_4=\left(\left[\begin{array}{cc} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{array}\right] p^T\right)^T \frac{|p+s t|+t}{|p|} \end{array}\right. \quad (4)$$

其中, p_1 、 p_2 、 p_3 、 p_4 分别表示可信区域的左下角、左上角、右下角、右上角坐标, θ 表示 s 与 a 之间的夹角, t 表示车辆在 t 时刻与 \hat{t} 时刻位置之间的距离。

若位置传感器采集的车辆位置位于可信区域内, 则保留数据; 否则判定采集的数据为噪声, 将其去除。

3.2.2 时间性自相关数据填充

本文针对噪声去除导致少量数据丢失的问题, 从时间角度分析车辆的数据, 对丢失数据进行估计, 进而达到数据填充的目的。传统的指数平滑方法^[18]在进行数据填充时仅考虑过去一段时间传感器的采样值, 没有考虑未来一段时间传感器的采样值。为此, 本文改进了指数平滑方法, 将其用来填充车联网数据, 填充时不仅考虑填充时间前传感器采集的数据, 而且考虑数据丢失时间后传感器采集的数据。本文将该数据填充方法定义为时间性自相关数据填充, 即

$$F(v_i, t) = \frac{\alpha \left(\sum_{j=1}^{\frac{\eta}{2}} f(v_i, t - j\Delta t)(1-\alpha)^j \right)}{\sum_{i=1}^{\frac{\eta}{2}} \alpha(1-\alpha)^j} + \frac{\alpha \left(\sum_{j=1}^{\frac{\eta}{2}} f(v_i, t + j\Delta t)(1-\alpha)^j \right)}{\sum_{i=1}^{\frac{\eta}{2}} \alpha(1-\alpha)^j} \quad (5)$$

其中, Δt 表示传感器采样的时间间隔, η 表示数据填充窗口的大小, α ($\alpha \in (0,1)$) 表示平滑系数, $f(v_i, t - j\Delta t)$ 表示数据丢失时间 t 之前传感器采集的数据, $f(v_i, t + j\Delta t)$ 表示数据丢失时间 t 之后传感器采集的数据。

3.2.3 时空性协同过滤数据填充

路网匹配^[19]将前后轨迹数据映射到路网中, 对可能的轨迹进行评价, 从而选取概率最大的一条作为填充数据, 但存在对应时间点的参数预测困难的问题。虽然节点状态变化频繁, 拓扑结构每时每刻都在更新, 但由于路网的存在, 最佳路线选取的策略几乎一致, 在某时间点连通的两辆车在过去和未来都可能存在相似的轨迹和状态。尤其是在高速路段以及工作日早晚高峰, 车辆的目的地极其相近, 所以可以借助这些轨迹相似的车辆对目标车辆的空值数据段进行填充。协同过滤^[20]的核心思想是用向量描述用户的历史信息, 然后计算用户之间的相似性, 再通过与目标用户相似性较高的邻居对其产品的评价, 从而得到目标用户对特定产品的潜在需求程度, 系统根据计算到的结果进行针对性推荐。将协同过滤的思想用到车联网数据填充上, 关键是对车辆之间的相似性进行计算。针对传感器故障等问题导致大量数据丢失的问题, 本文提出一种基于路网匹配的时空性协同过滤数据填充方法。

假设 v_i 在 t 到 \hat{t} 之间的特征需要进行填充, \hat{t} 时刻 v_i 的邻近节点集为

$$\Gamma(v_i, \hat{t}) = \{v_j \mid v_j \in V \wedge \|p_i - p_j\| \leq R\} \quad (6)$$

其中, $\|p_i - p_j\|$ 表示 v_i 与 v_j 之间的距离, R 表示 v_i

的通信范围。

v_i 与 v_j ($v_j \in \Gamma(v_i, \hat{t})$) 在 \hat{t} 时刻的相似度为

$$S(v_i, v_j, \hat{t}) = \frac{c_1 - c_2}{c_3 c_4} + 1 \quad (7)$$

其中, $c_1 = n \sum_{k=1}^n f_k(v_i, \hat{t}) f_k(v_j, \hat{t})$, $c_2 = \sum_{k=1}^n f_k(v_i, \hat{t}) \sum_{l=1}^n f_k(v_j, \hat{t})$,

$$c_3 = \sqrt{n \sum_{k=1}^n (f_k(v_i, \hat{t}))^2 - \left(\sum_{k=1}^n f_k(v_i, \hat{t}) \right)^2}, \quad c_4 = \sqrt{n \sum_{k=1}^n (f_k(v_j, \hat{t}))^2 - \left(\sum_{k=1}^n f_k(v_j, \hat{t}) \right)^2},$$

n 表示车辆特征向量的长度, $f_k(v_i, \hat{t})$ 和 $f_k(v_j, \hat{t})$ 分别表示 \hat{t} 时刻 v_i 和 v_j 的第 k 个特征。

$S(v_i, v_j, \hat{t})$ 取值范围为 $[0, 2]$, 其值越大表示 v_i 与 v_j 的相似度越高。

基于 \hat{t} 时刻 v_i 与邻近节点的相似度, 以及 v_i 邻近节点的时空数据, 给出时空性协同过滤数据填充方法, 即

$$I(v_i, \hat{t}) = \sum_{v_j \in \Gamma(v_i, \hat{t})} \frac{S(v_i, v_j, \hat{t}) f(v_j, \hat{t})}{\sum_{v_j \in \Gamma(v_i, \hat{t})} S(v_i, v_j, \hat{t})} \quad (8)$$

3.3 基于特征压缩的时空数据降维

1) 特征无量纲转化

假设 $f(v_i, t)$ 中的第 j 个特征为 $f_j(v_i, t)$, 使用 max-min 方法将 $f_j(v_i, t)$ 缩放至 $[0, 1]$, 即

$$\hat{f}_j(v_i, t) = \frac{f_j(v_i, t) - f_j^{\min}(v_i, t)}{f_j^{\max}(v_i, t) - f_j^{\min}(v_i, t)} \quad (9)$$

其中, $\hat{f}_j(v_i, t)$ 表示无量纲处理后的车辆特征, $f_j^{\min}(v_i, t)$ 表示车辆第 j 个特征的最小值, $f_j^{\max}(v_i, t)$ 表示车辆第 j 个特征的最大值。

2) 基于主成分分析的特征降维

假设所有车辆在 t 时刻的特征为 $f(t) = [f(v_1, t), \dots, f(v_{|V|}, t)]$, 其协方差矩阵为

$$X = \frac{1}{|V|} f(t) (f(t))^T \quad (10)$$

其中, $|V|$ 表示车辆的个数, T 表示向量的转置。

协方差矩阵的特征向量为

$$Y = [z_1, \dots, z_{|V|}] \quad (11)$$

将车辆特征向量与协方差矩阵特征向量相乘后，选择值最大的 m 个维度作为车辆的特征向量，即可实现特征降维。

3.4 基于张量因子聚合的神经网络构建

张量神经网络^[21]结合了单层模型和潜在因子模型，是一种将两者特点结合在一起扩展产生的模型。数据抽取过程及基于张量因子聚合的神经网络结构如图 1 所示。

在数据抽取过程中，首先从车辆维度对数据进行抽取。由于连通的时效性，需要时间一致，因此本文从时间维度对数据进行进一步选择，以得到网络训练所需要的输入数据。

通过改进张量神经网络，给出适用于车联网的张量因子聚合层，张量因子聚合的输出为

$$o = \left[f(v_i, t)W_1f(v_i, t)^T, f(v_j, t)W_1f(v_j, t)^T \right] + W_2 \left[f(v_i, t), f(v_j, t) \right]^T + W_3 \quad (12)$$

其中， $f(v_i, t)$ 和 $f(v_j, t)$ 分别表示车辆 v_i 和 v_j 的特征向量， $f(v_i, t)$ 和 $f(v_j, t)$ 的维度为 $1 \times n$ ， W_1 、 W_2 和 W_3 为权重矩阵， W_1 的维度为 $n \times n$ ， W_2 的维度为 $1 \times n$ ， W_3 的维度为 1×1 。

基于张量因子聚合构建神经网络，用于预测车辆之间的连通强度，其目标函数为

$$J(w) = \arg \min_w \log(y) + \left\| W_4 f(v_i, t) + W_5 - W_6 f(v_j, t) \right\|_2^2 + \beta \|w\|_2 \quad (13)$$

其中， y 表示连通强度预测值， $\left\| W_4 f(v_i, t) + W_5 - W_6 f(v_j, t) \right\|_2^2$ 表示 $f(v_i, t)$ 和 $f(v_j, t)$ 之间的相似度， $\beta \|w\|_2$ 表示正则化项。

对时空数据去噪、填充和降维后，整合存储为训练数据集。使用训练数据集和梯度下降优化目标函数，通过梯度下降迭代求解目标函数，得到最小化的目标函数和基于张量因子聚合的神经网络的参数值。

4 城市场景中车联网通达性

4.1 相关定义

定义 3 一个路边基础设施 u_l 覆盖范围内的车联网可表示为

$$\hat{G} = (\hat{V}, \hat{E}, W, \rho_k, u_l, \varphi) \quad (14)$$

其中， ρ_k 表示路边基础设施所处的路段， $\hat{V} \subseteq V$ 表示处于路段 ρ_k 的车辆集合， $\hat{E} = (\hat{V} \times \hat{V}) \cap E$ 表示车辆之间的连通关系， W 和 φ 的含义与定义 1 相同。

定义 4 \hat{G} 的拉普拉斯矩阵为

$$L(\hat{G}) = D(\hat{G}) - A(\hat{G}) \quad (15)$$

其中， $A(\hat{G})$ 表示 \hat{G} 的邻接矩阵， $D(\hat{G}) = \text{diag} \cdot \left[|\Gamma(v_1, \hat{t})|, |\Gamma(v_2, \hat{t})|, \dots, |\Gamma(v_{|V|}, \hat{t})| \right]$ 表示车辆节点度

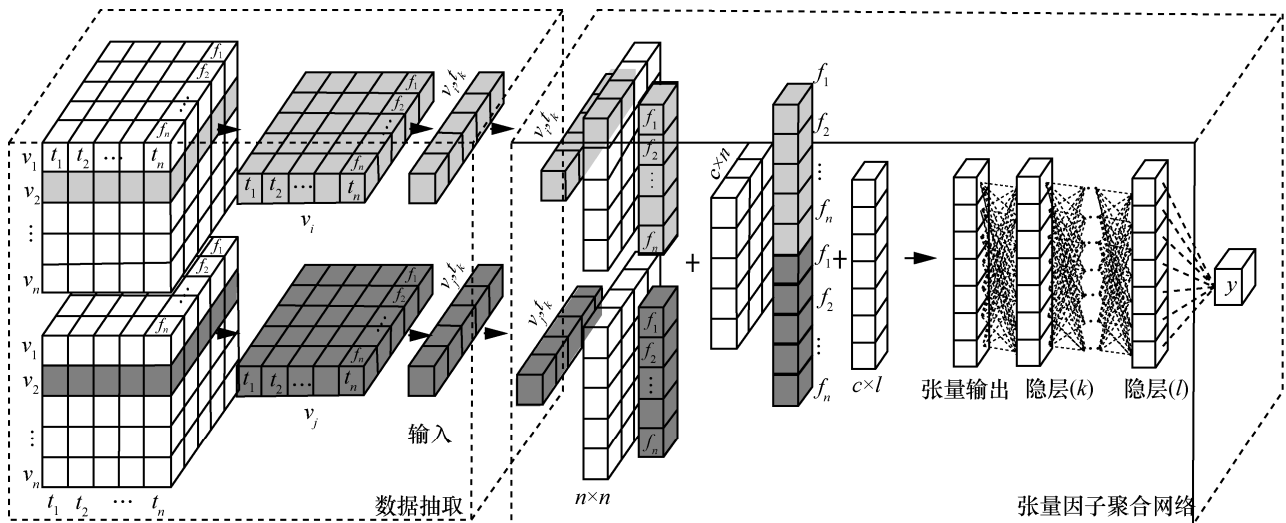


图 1 数据抽取过程及基于张量因子聚合的神经网络结构

的向量, $|\Gamma(v_i, \hat{t})|$ 表示 \hat{t} 时刻 v_i 邻近节点个数。

定理 1^[22] 对于 $|\hat{V}| \times 1$ 维向量 \mathbf{r} , $\mathbf{L}(\hat{G})$ 满足

$$\mathbf{r}^T \mathbf{L}(\hat{G}) \mathbf{r} = \frac{1}{2} \sum_{i,j=1}^{|\hat{V}|} \varphi(v_i, v_j) (\mathbf{r}_i - \mathbf{r}_j)^2 \quad (16)$$

4.2 基于拉普拉斯矩阵的弱连接检测

假设 \hat{G} 划分为 \hat{G}_1, \hat{G}_2 , \hat{G} 的子网之间的权重之和可表示为

$$H = \sum_{\hat{G}_1, \hat{G}_2} \sum_{e_k \in \hat{V}_1 \times \hat{V}_2 \cap \hat{V}} \varphi(e_k) \quad (17)$$

假设 $\mathbf{r} = (r_1, r_2, \dots, r_{|\hat{V}|})^T$, 其中

$$r_i = \begin{cases} \sqrt{\frac{|\hat{G}_2|}{|\hat{G}_1|}}, & v_i \in \hat{G}_1 \\ -\sqrt{\frac{|\hat{G}_1|}{|\hat{G}_2|}}, & v_i \in \hat{G}_2 \end{cases}$$

则有

$$\begin{aligned} \mathbf{r}^T \mathbf{L}(\hat{G}) \mathbf{r} &= \\ & \frac{1}{2} \sum_{v_i \in \hat{G}_1, v_j \in \hat{G}_2} \delta(v_i, v_j) \left(\sqrt{\frac{|\hat{G}_2|}{|\hat{G}_1|}} + \sqrt{\frac{|\hat{G}_1|}{|\hat{G}_2|}} \right)^2 + \\ & \sum_{v_i \in \hat{G}_2, v_j \in \hat{G}_1} \delta(v_i, v_j) \left(-\sqrt{\frac{|\hat{G}_2|}{|\hat{G}_1|}} - \sqrt{\frac{|\hat{G}_1|}{|\hat{G}_2|}} \right)^2 \end{aligned}$$

假设 $\mathbf{L}(\hat{G})$ 的特征值为 $\mathbf{z} = [z_1, \dots, z_{|\hat{V}|}]$, 且

$\mathbf{r}^T \mathbf{L}(\hat{G}) \mathbf{r} = \mathbf{r}^T \mathbf{z} \mathbf{r}$, 则

$$\mathbf{r}^T \mathbf{r} = \|\mathbf{r}\|^2 = |\hat{G}_1| \sqrt{\frac{|\hat{G}_2|}{|\hat{G}_1|}} + |\hat{G}_2| \sqrt{\frac{|\hat{G}_1|}{|\hat{G}_2|}} = |\hat{V}| \quad (18)$$

所以, 最小化特征值 \mathbf{z} 并找到其对应的特征向量即可找到车联网中的弱连接。根据之前提到的性质, $\mathbf{L}(\hat{G})$ 的最小特征值 \mathbf{z} 对应特征向量为 $\bar{\mathbf{1}}$ 。根据 Rayleigh-Ritz 理论^[22], 不符合 $\mathbf{z} \perp \bar{\mathbf{1}}$ 条件, 故取第二小特征值。如果将 \hat{G} 划分为 l 组, 则变成取前 l 个特征值及其特征向量。

4.3 连通候选节点集构建

弱连接检测可以发现车联网中容易断开的边。

为了解决弱连接检测的边断开而导致整个网络出现不连通的问题, 本文提出一种连通候选节点集构建方法。首先, 给出当前不连通的 2 个车辆在未来时刻可能连通的度量方法, 即

$$Q(v_i, v_j) = \frac{\|\mathbf{p}_i - \mathbf{p}_j\| |\Gamma(v_i, t)| + |\Gamma(v_j, t)|}{R \Gamma_{\max}} \cdot \frac{1}{\|\mathbf{W}_4 \mathbf{f}(v_i, t) + \mathbf{W}_5 - \mathbf{W}_6 \mathbf{f}(v_j, t)\|_2^2} \quad (19)$$

其中, Γ_{\max} 表示车辆邻近节点的最大个数。

在弱连接检测的边断开后, 一个路边基础设施下的车联网划分为若干不连通的子网, 从不连通的子网中选择 $Q(v_i, v_j)$ 最大的 2 个节点作为连通候选节点。

4.4 基于启发式搜索的通达性算法

本节给出一种基于启发式搜索的通达性 (HA, heuristic-search-based accessibility) 算法, 在通达性路径搜索过程中, 消耗函数、未完成度函数和评价函数分别为

$$C(\mathcal{P}) = \frac{|\mathcal{P}|}{\sum_{e_j \in \mathcal{P}} \varphi(e_j)} \quad (20)$$

$$B(\mathcal{P}) = \frac{\sigma(\mathcal{P}, v_2)}{R} \quad (21)$$

$$K(\mathcal{P}) = C(\mathcal{P}) + B(\mathcal{P}) \quad (22)$$

其中, $C(\mathcal{P})$ 表示消耗函数, $B(\mathcal{P})$ 表示未完成度函数, $K(\mathcal{P})$ 表示评价函数, \mathcal{P} 表示连通路程, $\sigma(\mathcal{P}, v_2)$ 表示连通路程的最后一个路段与数据接收节点所处路段之间的距离。

HA 算法的具体过程如算法 1 所示。首先, 遍历数据发送节点周围的连通候选节点, 若它周围没有连通候选节点, 则使用路边基础设施进行数据转发 (步骤 5)~步骤 6); 否则, 将连通候选节点加入连通路程集合 (步骤 7)~步骤 8)。然后, 重复选择连通路程集合中的节点, 使用评价函数选择连通路程的下一跳节点 (步骤 11)~步骤 21)。最后, 基于连通路程集合构建发送节点到接收节点的连通路程 (步骤 23)~步骤 26)。

算法 1 HA 算法

输入 发送节点 v_1 , 接收节点 v_2

1) 初始化连通路程集合 list

- 2) while $v_2 \notin \text{list}$
- 3) if list 为空
- 4) 寻找并遍历连通候选节点集合
- 5) if 无连通候选节点
- 6) 使用路边基础设施连通
- 7) else
- 8) 将连通候选节点放入 list
- 9) end if
- 10) end if
- 11) for $v_i \in \text{list}$
- 12) for $v_j \in \Gamma(v_i, t)$
- 13) if $K(v_j) < K(v_i)$
- 14) v_j 为最新节点
- 15) v_j 的父节点为 v_i
- 16) end if
- 17) if $v_j \notin \text{list}$
- 18) v_j 放入 list
- 19) end if
- 20) end for
- 21) end for
- 22) end while
- 23) while v_2 的父节点不是 v_1
- 24) $\mathcal{P} = \mathcal{P} \cup (v_2, v_2 \text{ 的父节点})$
- 25) $v_2 = v_2$ 的父节点
- 26) end while
- 27) 输出 \mathcal{P}

5 仿真实验与结果分析

5.1 仿真实验环境

仿真实验采用 NGSIM 项目采集的数据集^[23]。该数据收集位于美国加利福尼亚州洛杉矶的 Lankershim Boulevard 路段的车辆行驶轨迹数据。仿真实验使用 2005 年 6 月 16 日 8:28—8:45 时间段的车辆及道路信息。仿真实验参数如表 1 所示。

5.2 时空数据分析仿真实验结果

为了评价车联网时空数据分析方法，本文对比了使用原始数据与处理后数据的网络性能。从传统反向传播 (BP, back propagation) 神经网络和基于张量因子聚合的神经网络的数据体积、平均错误率、训练速度 3 个指标进行评价，结果分别如图 2 和图 3 所示。与使用原始数据相比，BP 神经网络使用处理后的数据可以降低 18% 的数据体积和 21%

的平均错误率。在对比神经网络训练速度时，保持网络的优化方法、学习率、批处理规模等参数不变，对比神经网络使用原始数据与处理后数据的训练速度。由图 2 可知，BP 神经网络使用处理后的数据可以提高 12% 的训练速度。与使用原始数据相比，基于张量因子聚合的神经网络使用处理后的数据可以降低 20% 的数据体积和 30% 的平均错误率，可以提高 40% 的训练速度。

表 1 仿真实验参数

参数名称	参数值
道路长度/m	487
数据采集时间间隔/s	0.1
最大同向车道数量	6
数据记录数目	902 025
车辆数量/辆	1 601
通信距离/m	50
采样频率/Hz	10
路段个数	5
RSU 辐射范围	一个路段

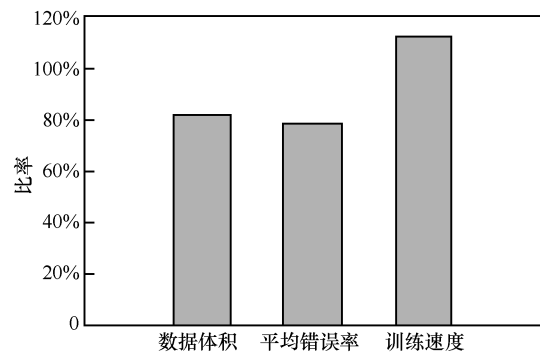


图 2 时空数据处理前后 BP 神经网络的性能对比

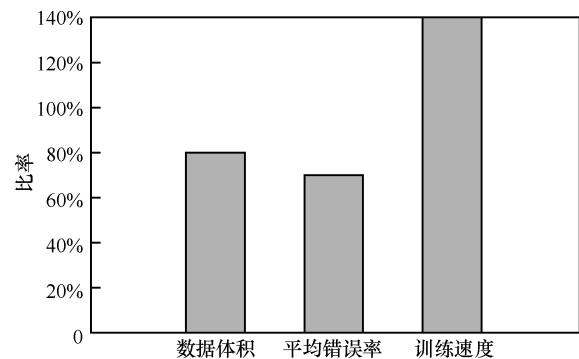


图 3 时空数据处理前后基于张量因子聚合的神经网络的性能对比

基于张量因子聚合的神经网络输出 2 个车辆之间的连通强度，1 000 对车辆的误差分布如图 4 所

示。横坐标表示连通强度预测值与实际样本值的误差，纵坐标表示落在相应误差范围内的样本占比。可以看出，基于张量因子聚合的神经网络在测试集实验结果中，有 81% 的输出结果误差在 10% 内，90% 的结果误差小于 20%。

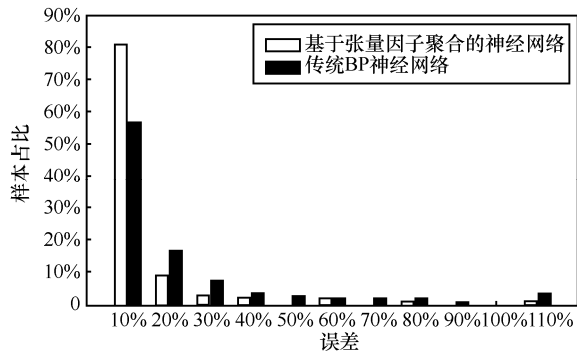


图 4 误差分布

5.3 车联网通达性方法仿真实验结果

图 5 展示了路段划分方法。图 6 展示了路网中车辆数目变化情况，横坐标为时间，纵坐标为车辆数目。由图 6 可知，车辆数目在 30~100 辆波动。图 7 展示了不同路段车辆数目变化。由图 7 可知，不同路段之间车辆分布不均匀，同一路段不同时刻的车辆数目变化也非常大。图 8 展示了不同路段车辆密度变化。由图 8 可知，车辆分布是不均匀的。



图 5 Lankershim Boulevard 路段

为了评价本文所提出的车联网通达性方法，本节对比了贪婪周边无状态路由（GPSR, greedy perimeter stateless routing）算法^[24-26]和路边基础设施转发（RT, RSU transmission）算法。在 RT 算法中，一个路边基础设施负责转发其覆盖范围内车辆的数据，而车辆间不直接通信。

仿真实验选取路段 3 车辆数目变化最频繁的

200 s 数据。弱连接检测的准确性如图 9 所示。不同时间的车辆连通数目如图 10 所示。3 种通达性算法中车辆之间的平均时延如图 11 所示。

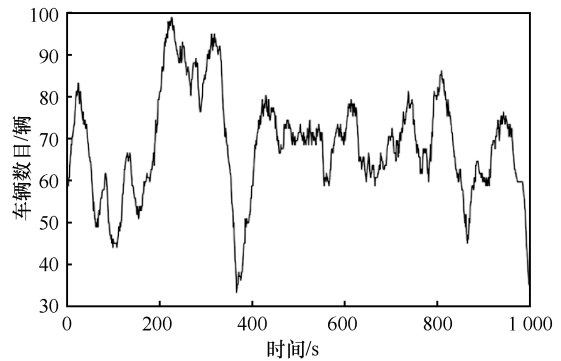


图 6 路网中车辆数目变化

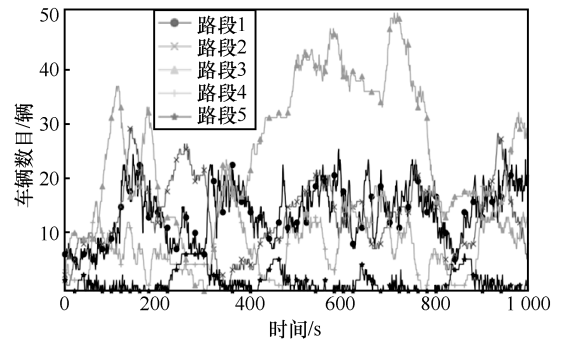


图 7 不同路段车辆数目变化

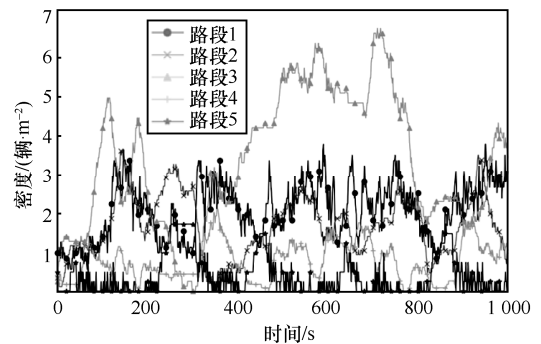


图 8 不同路段车辆密度变化

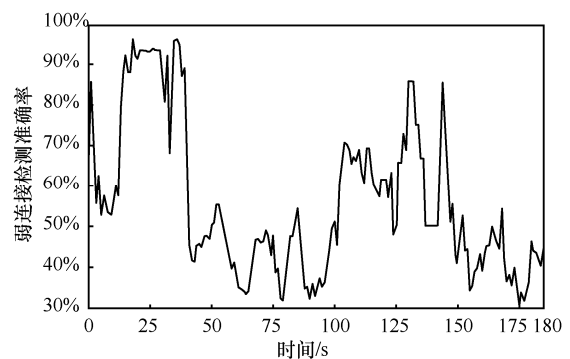


图 9 弱连接检测准确率

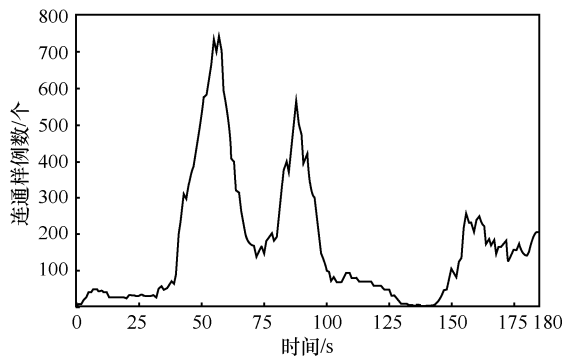


图 10 连通数目

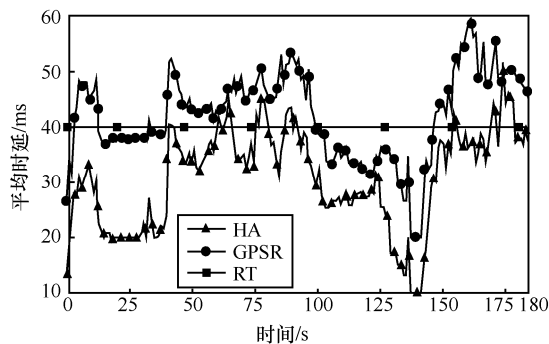


图 11 平均时延

HA 算法的弱连接检测准确率和平均时延成反相关, 即弱连接检测准确率越高, 平均时延越低。弱连接检测后连通候选节点集构建可以减少由于连接断开导致网络割裂, 进而解决车辆依赖路边基础设施进行数据中转导致时延上升的问题。

由图 10 和图 11 可知, HA 算法和 GPSR 算法的平均时延在某些情况下高于直接利用路边基础设施转发的 RT 算法。然而, 路边基础设施带宽有限, 当车辆所需带宽大于路边基础设施带宽时, 路边基础设施不能为所有车辆提供数据转发服务。因为 HA 算法考虑了车辆之间的连通强度, 能够检测弱连接并构建连通候选节点集, 避免了因为车辆间连接断开所造成的路边基础设施重新调度及通信带来的时延, 因此 HA 算法效果要优于 GPSR 算法。

6 结束语

本文研究了城市场景中车联网时空数据分析及其通达性方法, 提出了基于噪声去除和数据填充的时空数据分析方法, 构建了基于张量因子聚合的神经网络预测车辆之间的连通强度; 提出了基于连通强度预测的车联网通达性方法。仿真结果表明, 本文所提出的车联网时空数据分析及其通达性方法具有较好的性能。

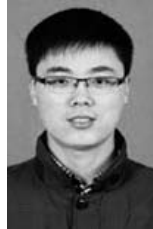
参考文献:

- [1] 焦政达, 马建峰, 孙聪, 等. 新的车辆远程诊断授权协议[J]. 通信学报, 2014, 35(11): 146-153.
JIAO Z D, MA J F, SUN C, et al. New remote authorization protocol for vehicle diagnosis[J]. Journal on Communications, 2014, 35(11): 146-153.
- [2] 陈亮, 王军, 陈蓉, 等. 车载自组织网视频流媒体协助下载研究[J]. 通信学报, 2019, 40(1): 51-63.
CHEN L, WANG J, CHEN R, et al. Research on video streaming media cooperative downloading in vehicular ad hoc network[J]. Journal on Communications, 2019, 40(1): 51-63.
- [3] ZHANG J, CUI J, ZHONG H, et al. PA-CRT: Chinese remainder theorem based conditional privacy-preserving authentication scheme in vehicular ad-hoc networks[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 18(2): 722-735.
- [4] CUI J, ZHANG X Y, ZHONG H, et al. Extensible conditional privacy protection authentication scheme for secure vehicular networks in a multi-cloud environment[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 1654-1667.
- [5] CHEN C, HU J N, QIU T, et al. CVCG: cooperative V2V-aided transmission scheme based on coalitional game for popular content distribution in vehicular ad-hoc networks[J]. IEEE Transactions on Mobile Computing, 2019, 18(12): 2811-2828.
- [6] CHEN C, WANG C, QIU T, et al. Caching in vehicular named data networking: architecture, schemes and future directions[J]. IEEE Communications Surveys & Tutorials, 2020, 22(4): 2378-2407.
- [7] CONTRERAS-CASTILLO J, ZEADALLY S, GUERRERO IBAÑEZ J A. Solving vehicular ad hoc network challenges with Big Data solutions[J]. IET Networks, 2016, 5(4): 81-84.
- [8] LIU Y S, CHEN X, CHEN C L, et al. Traffic big data analysis supporting vehicular network access recommendation[C]//2016 IEEE International Conference on Communications. Piscataway: IEEE Press, 2016: 1-6.
- [9] DU N B, ZHAN J F, ZHAO M, et al. Spatio-temporal data index model of moving objects on fixed networks using HBase[C]//2015 IEEE International Conference on Computational Intelligence & Communication Technology. Piscataway: IEEE Press, 2015: 247-251.
- [10] CHENG J J, YUAN G Y, ZHOU M C, et al. A fluid mechanics-based data flow model to estimate VANET capacity[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(6): 2603-2614.
- [11] ALTOMARE A, CESARIO E, COMITO C, et al. Trajectory pattern mining for urban computing in the cloud[J]. IEEE Transactions on Parallel and Distributed Systems, 2017, 28(2): 586-599.
- [12] KUMAR N, MISRA S, RODRIGUES J J P C, et al. Coalition games for spatio-temporal big data in Internet of vehicles environment: a comparative analysis[J]. IEEE Internet of Things Journal, 2015, 2(4): 310-320.
- [13] 刘业, 吴国新. 基于 802.11p/WAVE 的车联网连通性模型及其应用研究[J]. 通信学报, 2013, 34(6): 85-91.
LIU Y, WU G X. Study on the connectivity model of VANET based on 802.11p/WAVE and its application[J]. Journal on Communications, 2013, 34(6): 85-91.
- [14] QURESHI K N, ABDULLAH A H, LLORET J, et al. Road-aware routing strategies for vehicular ad hoc networks: characteristics and comparisons[J]. International Journal of Distributed Sensor Networks, 2016, 12(3): 1605734.

- [15] ZHENG J, WANG Y. Connectivity analysis of vehicles moving on a highway with one entry and exit[J]. IEEE Transactions on Vehicular Technology, 2018, 67(5): 4476-4486.
- [16] CHENG J J, YUAN G Y, ZHOU M C, et al. Accessibility analysis and modeling for IoV in an urban scene[J]. IEEE Transactions on Vehicular Technology, 2020, 69(4): 4246-4256.
- [17] CHENG J J, YUAN G Y, ZHOU M C, et al. A connectivity-prediction-based dynamic clustering model for VANET in an urban scene[J]. IEEE Internet of Things Journal, 2020, 7(9): 8410-8418.
- [18] GARDNER E S J. Exponential smoothing: the state of the art—part II[J]. International Journal of Forecasting, 2006, 22(4): 637-666.
- [19] BRAKATSOULAS S, PFOSER D, TRYFONA N. Practical data management techniques for vehicle tracking data[C]//21st International Conference on Data Engineering. Piscataway: IEEE Press, 2005: 324-325.
- [20] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th International Conference on World Wide Web. New York: ACM Press, 2001: 285-295.
- [21] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[J]. Advances in Neural Information Processing Systems, 2013, 26: 926-934.
- [22] FIEDLER M. Algebraic connectivity of graphs[J]. Czechoslovak Mathematical Journal, 1973, 23(2): 298-305.
- [23] ALEXIADIS V, COLYAR J, HALKIAS J, et al. The next generation simulation program[J]. ITE Journal, 2004, 74(8): 22-26.
- [24] CHENG J J, CHENG J L, ZHOU M C, et al. Routing in Internet of vehicles: a review[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(5): 2339-2352.
- [25] 吴振华, 胡鹏. VANET 中路由协议分析[J]. 通信学报, 2015, 36(S1): 75-84.
WU Z H, HU P. Analysis on VANET routing protocols[J]. Journal on Communications, 2015, 36(S1): 75-84.
- [26] KARP B, KUNG H T. GPSR: greedy perimeter stateless routing for wireless networks[C]//Proceedings of the 6th Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2000: 243-254.



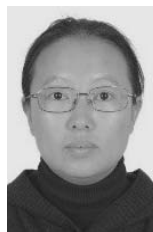
原桂远（1993-），男，山东烟台人，同济大学博士生，主要研究方向为车联网、无人驾驶等。



崔杰（1980-），男，河南淮阳人，博士，安徽大学教授、博士生导师，主要研究方向为车联网安全、物联网安全等。



周爱国（1973-），男，湖南株洲人，博士，同济大学副教授、博士生导师，主要研究方向为机器人控制、无人驾驶等。



吕博（1977-），女，辽宁抚顺人，博士，上海师范大学副教授，主要研究方向为计算机视觉、机器学习等。

[作者简介]



程久军（1974-），男，安徽怀宁人，博士，同济大学教授、博士生导师，主要研究方向为车联网、无人驾驶等。



李光耀（1965-），男，安徽安庆人，博士，同济大学教授、博士生导师，主要研究方向为图形图像、数据挖掘及人工智能等。