

基于循环生成式对抗网络实现停车场 时空数据的修复

孙玉强^{1,2} 彭 磊¹ 李慧云¹

¹(中国科学院深圳先进技术研究院 深圳 518055)

²(中国科学技术大学 苏州 215123)

摘 要 停车诱导技术在一定程度上缓解了高峰时段无序停车问题,并减少了司机寻找车位的时间,但停车诱导系统对实时数据和历史数据有较高的依赖。如果缺少相应数据,那么诱导系统的准确性将大打折扣。针对这一问题,该文通过挖掘停车场周围的空间数据,提出了一种停车场空间相似度量,并计算出停车场空间相似情况下其数据的相似条件概率。当条件概率足够大时,以已知数据为学习样本,使用循环生成式对抗网络获得修复数据。实验结果表明,当停车场空间具有较高空间相似性时,其数据同样有大概率的相似性,使用循环生成式对抗网络生成的数据与真实数据具有相同的分布。该文提出的方法可在短时间内生成大量的合理数据,实现停车场数据的修复,提高诱导系统的可靠性。

关键词 停车诱导系统; 停车数据修复; 数据挖掘; 时空相似性; 循环生成式对抗网络

中图分类号 TP 181 **文献标志码** A

Spatiotemporal Data Repairing of Parking Lots Based on Recurrent Generative Adversarial Networks

SUN Yuqiang^{1,2} PENG Lei¹ LI Huiyun¹

¹(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

²(University of Science and Technology of China, Suzhou 215123, China)

Abstract The parking guidance system (PGS) can alleviate the disordered parking problem in the peak time, and reduce the time to find parking space. But existing PGS techniques are quite dependent to the real-time data and historical data, its performance will be greatly degraded when the data is insufficient. To solve this problem, a data repairing based PGS method was presented. First, by mining the spatial data around the parking lots, a spatial similarity metric of parking lot was proposed. Then, the possibility of parking data similarity was calculated when the parking lots had spatial similarity. If the conditional probability was large enough, the known data of parking lots would be used as the learning samples. Finally, the reparative data

收稿日期: 2018-07-10 修回日期: 2018-08-02

基金项目: 广东省科技计划重大项目(2015B010106004)

作者简介: 孙玉强, 硕士研究生, 研究方向为机器学习、数据挖掘、交通大数据等; 彭磊(通讯作者), 高级工程师, 研究方向为物联网、车联网、智能交通、智能驾驶等, E-mail: lei.peng@siat.ac.cn; 李慧云(通讯作者), 研究员, 研究方向为密码安全芯片设计、车联网等, E-mail: hy.li@siat.ac.cn.

could be generated by recurrent generative adversarial networks. Experimental results show that when the parking lots have high spatial similarity, the data of parking lots have high similarity probability also. The data generated by the recurrent generative adversarial networks also have the same distribution with real data. By the proposed method, a large number of reasonable data can be generated efficiently, and the PGS performance can be improved while only few parking data is available.

Keywords parking guidance system; parking data repairing; data mining; spatiotemporal similarity; recurrent generative adversarial networks

1 引 言

随着私家车数量的激增, 停车问题逐渐成为我国亟待解决的一个交通问题。针对这一问题, 近年来许多停车诱导系统^[1]被开发出来用以缓解停车位的供需矛盾。停车诱导系统主要基于停车场的历史停车数据进行建模, 使用长短期记忆(Long-Short Term Memory, LSTM)^[2]网络等神经网络模型可以对未来较长时间内停车场剩余车位做较精准的预测, 司机则可以根据预测的结果来对自己的行程做规划。通过在停车场周围铺设传感器^[3], 还可以根据停车场的实时数据^[4]对剩余车位做更精准的短时预测^[5]。此外, Rajabioun 和 Ioannou^[6]提出了时空结合预测模型, 通过结合时间序列和地理空间信息, 能够得到更好的预测效果。但是, 由于我国城市化进程起步较晚, 早期的城市规划缺乏经验, 对停车场的建设也不够完善, 使得我国目前多数城市仍有大量的停车场处于缺少停车数据的状态^[7]。在这种情况下, 停车诱导系统便难以起到有效的预测效果。为了解决这一问题, 我们希望建立一个既可以对停车场的历史数据进行修复, 又无需加装额外的数据采集设备与联网设施的模型。因为建设一个完备的城市级停车数据网联系统会耗费高昂的经济成本和大量的时间成本。但目前相关领域的研究相当匮乏, 尤其是在无先验知识的情况下对停车数据进行修复。同时, 与图像修复相比, 对时间序列数

据的修复在一定程度上更具挑战性。这是因为有更多的数学工具和方法可以对修复数据的效果做检验, 评判标准更严苛, 而对图像修复的评判则更多的是基于主观评判。本文提出了一种停车数据的修复方法试图解决上述问题, 该方法在一定程度上可以缓解停车场的数据残缺问题, 同时能够节约经济成本和时间成本。

要修复一个停车场的的数据, 即生成一个停车场的历史数据, 需要找到与这个停车场具有相似特征的停车场数据作为学习样本。在数据残缺的情况下, 无法对停车场间的数据相似性做出判断, 但停车场间的空间信息是较容易获取的, 因此可以方便地对停车场间的空间相似性作出判断。那么在停车场具有空间相似的情况下, 是否意味着其停车数据也具有相似性呢? 事实上, 除了经度、纬度信息外, 每个停车场周围还有非常多的地理信息点, 如医院、商场、餐厅、公交站点等。每一个地理信息点都被称为一个地理兴趣点(Point of Interest, POI), 不同的 POI 会对停车场产生不同的影响。如果停车场周围有若干大型商场, 巨大的人流量会使该停车场的停车位非常稀缺; 而如果停车场周围有许多公交站、地铁站, 便利的交通会使该停车场的需求量降低。因此, 如果两个停车场周围具有相似的 POI 分布, 那么二者受到的影响也会非常相似, 即二者的停车数据会具有一定的相似性。根据这一猜想, 本文首先提出了停车场空间相似性的计算公

式, 然后提出了计算在停车场空间相似条件下其数据相似的概率的方法。如果猜想成立, 就可以使用生成模型生成修复数据。循环生成式对抗网络 (Recurrent Generative Adversarial Networks, RGAN) 非常适合于时间序列数据的生成。考虑到 RGAN 训练过程的不稳定性, 本文对其进行了部分改进, 使其更适合停车场数据的生成。最后, 使用卡方检验对生成的结果做评估, 如果生成的结果和真实数据属于同一分布, 说明该修复方法是有效的。

2 相关研究进展

生成式对抗网络 (Generative Adversarial Nets, GAN) [8] 自被提出以来, 已经成为人工智能学界最热门的研究方向之一。GAN 由两个网络构成, 分别为生成器和鉴别器, 其训练过程可视为两个网络间的对抗学习过程: 在每一次迭代中, 生成器期望生成一个与真实数据相差无几的数据; 而鉴别器则要尽可能地将生成器产出的数据与真实数据区分开。目前, GAN 主要被应用于图像生成领域 [9,10], 深度卷积生成式对抗网络 (Deep Convolutional Generative Adversarial Networks, DCGAN) 第一次将 GAN 与卷积神经网络 (Convolutional Neural Network, CNN) 结合在一起, 在图像生成方面取得了非常好的效果。在 DCGAN 的基础上, 通过文字描述生成图片也得到了令人惊艳的结果 [11]。在序列数据生成方面 [12], 目前主要的研究都集中在文本生成领域, 其核心思想是将生成器的生成过程视为强化学习中策略选择问题, 同时将鉴别器的鉴别结果视为回馈奖励。在连续值数据生成方面 [13], 目前最好的方法是将 GAN 和 LSTM 网络结合在一起, 其网络模型与原始 GAN 无异, 只是将生成器与鉴别器替换为了 LSTM 网络。由于 LSTM 网络在处理时间序列数据方面表现十分优异, 循环生成式对抗

网络 [14] 也采用了这样一种网络结构, 不同之处在于循环生成式对抗网络鉴别器无需是双向性的, 且生成器的输出也不用作为反馈值。RGAN 在生成时间序列数据, 如医疗数据上取得了较好的结果。本文也以 RGAN 为基础, 实现停车场数据的生成和修复。RGAN 的网络结构如图 1 所示。

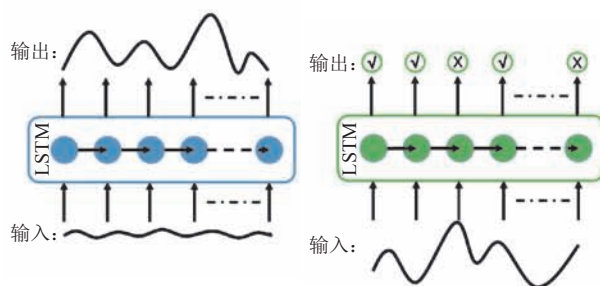


图 1 循环生成式对抗网络结构

Fig. 1 Architecture of recurrent generative adversarial networks model

3 时空相似度检验模型

每个停车场周围都存在着十分丰富的空间信息, 因此任意停车场间的空间相似度很容易做定量分析。而当停车距离十分相近或停车场周围的 POI 分布十分相近时, 其停车数据也可能会十分相似, 如图 2 所示。而本文关心的是, 在停车场间空间相似的情况下, 其数据相似的可能性有多大呢?

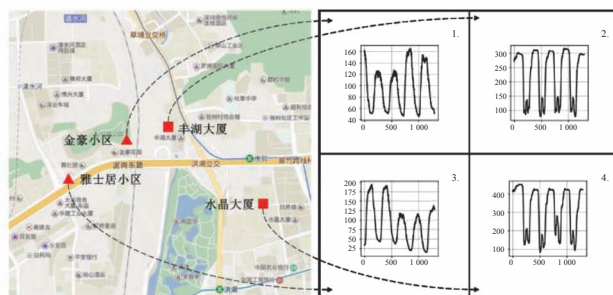


图 2 深圳市罗湖区部分停车场分布及其停车数据

Fig. 2 Part of parking lots and their data in Luohu, Shenzhen

首先, 需要建立一个停车场间空间相似度的度量模型。每个停车场本身会具备一个经度和纬度信息, 可以记为一个 2 维向量 \boldsymbol{l} 。此外, 我们知道每个停车场周围都有许多 POI 信息, 这些 POI 信息也需要被定量分析。事实上, 当司机达到目的地周边时会选择离目的地最近的停车场停车, 之后司机需要从停车场步行至目的地。因此, 当停车场与目的地的距离超过司机可接受范围时, 司机便不会在此处停车。基于此, 我们称停车场与目的地间的最远距离为容忍度半径, 记为 R_t 。这是一个统计量, 它代表了城市中多数司机可接受的停车场到目的地的最远距离。假设一个城市中有 h 种 POI, 则对任意一个停车场 \boldsymbol{l}_p 来说, 其周围产生影响的 POI 可用集合 $S_p = \{\boldsymbol{l}_i \mid \text{distance}(\boldsymbol{l}_p, \boldsymbol{l}_i)\} \leq R_t$ 来表示, 即以停车场为圆心, 以 R_t 为半径的圆内 POI 都会对该停车场产生影响。之后统计 S_p 中每种 POI 的个数, 便可得到关于这个停车场的一个 h 维向量, 记为 \boldsymbol{s} 。因此, 每个停车场的地理空间信息都可以用一个 $(2+h)$ 维的向量 $(\boldsymbol{l}, \boldsymbol{s})$ 来进行标定。那么对任意两个停车场 p_1 和 p_2 来说, 可以定义其空间相似度为:

$$S_s(p_1, p_2) = a \frac{\boldsymbol{l}_1 \cdot \boldsymbol{l}_2}{\|\boldsymbol{l}_1\| \cdot \|\boldsymbol{l}_2\|} + b \frac{\boldsymbol{s}_1 \cdot \boldsymbol{s}_2}{\|\boldsymbol{s}_1\| \cdot \|\boldsymbol{s}_2\|} \quad (1)$$

其中, $(\boldsymbol{l}_1, \boldsymbol{s}_1)$ 为 p_1 的向量; $(\boldsymbol{l}_2, \boldsymbol{s}_2)$ 为 p_2 的向量; a 和 b 为影响因子, 分别代表经、纬度相似性和 POI 分布相似性对空间相似的影响程度。 S_s 值介于 $[-1, 1]$, 越接近 -1 表示停车场间越不相似, 越接近 1 表示二者空间越相似。系数 a 和 b 值可以通过最小二乘法来确定, 但由于本身缺少关于停车场间空间相似度的客观评价标准, 因此首先需要对城市中部分停车场间的相似度做人工标注。在计算出 a 和 b 值后, 再对城市中其余停车场间的相似度做定量计算。

在得到了停车场间的空间相似度度量模型后, 需要确定在空间相似条件下停车场间数据

相似的概率。令 $P(S)$ 为停车场间空间相似的概率, $P(D, S)$ 为停车场间时空相似的联合分布概率, 则在空间相似条件下停车场数据相似的概率 $P(D|S)$ 可由如下贝叶斯公式计算得到。

$$P(D|S) = \frac{P(D, S)}{P(S)} \quad (2)$$

显然, 可以首先计算出这个城市的 $P(S)$ 和 $P(D, S)$ 值, 然后再计算出整个城市的 $P(D|S)$ 值。但这会带来两个问题: 对整个城市的数据做计算会消耗较多的计算资源和时间; 此外, 本文更关心的是空间相似停车场间的数据相似性, 将所有停车场纳入计算可能会影响 $P(D|S)$ 的值。因此, 可以根据公式 (1) 对停车场做聚类, 然后在每个簇中分别计算 $P(S)$ 和 $P(D, S)$ 值, 以此来节约计算时间。

假设经过聚类后, 在簇 φ 中有 m 个停车场。根据人工标注的结果, 统计出空间相似度为“非常相似”的停车场对数, 由集合 U_s 表示, 则在簇 φ 中, $P(S)$ 的计算公式如下:

$$P(S) = \frac{U_s}{C_m^2} \quad (3)$$

其中, C 表示组合数; m 表示簇 φ 中停车场的个数。

停车场数据间的相似度可以通过皮尔逊相关系数来表示。簇 φ 中任意两个停车场间 p_1 和 p_2 的数据相似度由皮尔逊相关系数计算如下:

$$S_d(p_1, p_2) = \frac{E[(d_1 - Ed_1)(d_2 - Ed_2)]}{\sqrt{D(d_1)}\sqrt{D(d_2)}} \quad (4)$$

其中, d_1 为 p_1 的数据; d_2 为 p_2 的数据; E 表示数据的期望值; D 表示数据的方差。设阈值为 ρ , 即当 d_1 和 d_2 的相似度超过 ρ 时, 二者具有强正相关性。聚类后, 理想的结果是簇 φ 中多数停车场的的数据都具有一定的相似性, 称该部分停车场为 φ 的主体部分, 记为集合 U_n 。假设主体部分包括 n 个停车场, 计算其两两之间的皮尔逊相关系数, 并统计出数据具有强相关性的停车场

对数, 由集合 U_{sp} 表示。之后取集合 U_s 和 U_{sp} 的交集, 即具有空间相似性和数据强相关性的停车场对数, 则 $P(D,S)$ 可计算如下:

$$P(D,S) = \frac{U_n \cdot U_s \cap U_{sp}}{m \cdot C_n^2} \quad (5)$$

其中, n 为主体停车场的个数。根据公式(3)和(5), 在簇 φ 中, 可以得到 $P(D|S)$ 的计算公式为:

$$P(D|S) = \frac{U_n \cdot U_s \cap U_{sp} \cdot C_m^2}{m \cdot C_m^2 \cdot U_s} \quad (6)$$

当 $P(D|S)$ 值足够大时, 可以说, 在簇 φ 中当停车场间空间相似时, 其数据也是相似的。

4 生成模型搭建

对簇 φ 中任意一个缺失数据的目标停车场, 都可以找到一个与其空间最相似的样本停车场, 且在较高的概率下, 可以认为二者的数据也非常相似。接下来需要对样本停车场的数据作清洗。通常来说, 在没有突发事件的影响时, 每天停车场的的数据应该符合正态分布, 因此需要对样本停车场的的数据做正态性检验。首先, 获取样本停车场任意月份的数据, 按当月天数分为若干条; 然后, 对不同天数中同一时刻的数据做正态性检验; 最后, 检验完一天中每个时间节点的数据后, 统计出不符合正态分布的数据组数。如果不符正态性的数据占比较高, 则可认为样本停车场的的数据是不符合正态分布的。检验数据的正态性可以使用 W 检验 (Shapiro-Wilk Test)、 D 检验 (D'Agostino-Pearson Test) 和 KS 检验 (Kolmogorov-Smirnov Test) 等方法。其中, 在数据量较小的情况下, D 检验效果较好, 其计算公式如下:

$$D = \frac{\sum \left(i - \frac{n+1}{2} \right) / x_i}{\sqrt{n^3 \left[\sum x^2 - (\sum x)^2 / n \right]}} \quad (7)$$

其中, x 表示停车数据; n 表示样本含量; i 是将数据按从小到大排列后所得的秩次。计算出 D 值后, 通过查询 D 临界值表可以确定 P 值。当 P 值低于显著性水平时, 说明数据不具有正态性; 否则就有充分的理由认为该组数据是符合正态分布的。

如果样本停车场的的数据符合正态分布, 就可以使用该数据作为学习样本生成新数据, 为目标停车场做数据修复。选用 RGAN 作为生成模型, 其生成器和鉴别器均为一个 LSTM 网络, 非常适合于时间序列数据的生成。在每一轮迭代中, RGAN 都会学习样本集中任一条数据的特征, 经过多轮迭代后, 生成的结果会具备样本数据的特征。但是 GAN 的训练过程经常会出现不稳定的情况, 因此会导致生成结果异常而无法用于数据修复。同时, 对于 GAN 生成结果的评估也较为困难, 对时间序列数据的评估尤为如此。在图像生成领域, 通常会采用“视觉评估”的方式来判断生成结果的好坏, 但这具有强烈的主观性, 而且对时间序列数据来说很不可靠。为此, 本文提出了一种贝叶斯过滤器模型, 将其添加在 RGAN 的输出层之后, 其模型结构如图 3 所示。在进行数据生成之前, 已经得到了簇 φ 的 $P(D|S)$ 值, 在得到了生成结果后, 将其添加进原始样本数据集中, 重新计算得到簇 φ 的 $P(D|S)'$ 值。如果 $P(D|S)'$ 小于初始的 $P(D|S)$ 值, 则需要将生成结果丢弃。

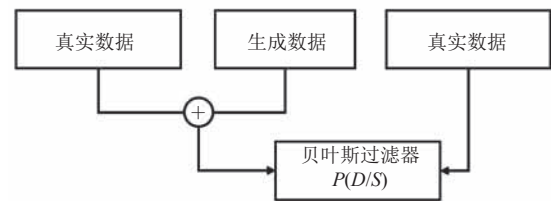


图 3 贝叶斯过滤器

Fig. 3 Architecture of Bayes filter

当生成足够多的数据后, 需要验证生成数据的合理性。使用卡方检验可以验证生成数据与真

实数据是否符合同一分布,其计算公式如下:

$$X^2 = \sum \frac{(A-T)^2}{T} \quad (8)$$

其中, A 表示生成数据; T 表示真实数据。 X^2 的值越接近 0, 说明生成的结果和真实数据越相似。

5 根据空间相似性验证数据相似性

以深圳市罗湖区为例,首先通过爬虫程序获取罗湖区 402 个停车场和 7 种主要 POI 的经、纬度值,且获取到其中 100 个停车场的停车数据,部分停车场与 POI 的地理分布如图 4 所示。由于缺乏相关的统计数据,容忍度半径 R_i 值无法确定。为了方便计算,这里假定 R_i 值为经度 1", 约等于 308 m。根据爬取到的地理信息数据,统计出罗湖区每个停车场周围 308 m 范围内的 7 种 POI 数量,然后得到每个停车场的一个 9 维向量。根据人工标注的结果,对罗湖区的数据使用最小二乘法计算得影响系数 a 和 b 的值分别为 0.71、0.29,说明对于该区域来说,经、纬度对于停车场空间相似性的影响程度更大。在确定了空间相似度的计算公式后,接下来需要对停车场做聚类。为了尽可能减少参数并保证算法的收敛速度足够快,本文采用 K -Means 算法来对停车场做高维聚类。本次实验中, k 值设为 6。经过聚



图 4 深圳市罗湖区部分 POI 分布

Fig. 4 Distribution of part of POIs in Luohu, Shenzhen

类运算后,在 100 个停车场中,有 37 个停车场被分在同一个簇中,是样本数最多的类型,将该簇作为样本集完成后续实验。

首先对 37 个停车场的数据做预处理。原始停车场的的数据是一个时间表,它记录了每一辆车进出停车场的准确时间,但时间表数据无法作为 RGAN 的学习样本。取 5 min 为一个时间步长,统计在每个时间节点处停车场的车辆数,这样时间表数据就被转换成了一维时间序列数据。考虑到停车场间规模的不同会影响到数据生成过程,该一维时间序列数据还需要进一步做归一化。在每个时间节点,用停车场的规模减去剩余车辆数再除以停车场的规模,就得到了每个时间节点的的空车率。本文选取样本集 37 个停车场于 2016 年 7 月的空车率数据,如图 5 所示。由图 5 可以看出,有 3 个停车场因数据缺失而导致空车率一直为 100%,如图中绿色直线所示,这部分数据应该作为脏数据被剔除。在其余 34 个停车场中,有 30 个停车场的的数据彼此间较为相似,这部分停车场可以作为主体部分,如图 5 中蓝色曲线所示。剩余 4 个停车场的的数据与主体部分明显呈现负相关性,如图 5 中红色曲线所示。本次实验中设置阈值 ρ 为 0.5,分别计算主体部分停车场两两之间的皮尔逊相关系数,在 435 个结果中有 386 个结果的值大于或等于阈值 ρ 。根据公式 (3)~(5),计算得到该簇中 $P(S)$ 、 $P(D,S)$ 和

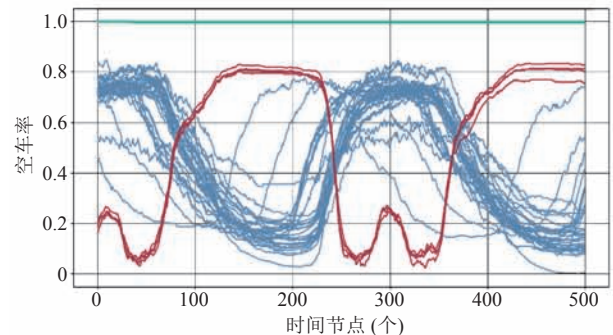


图 5 37 个停车场的空车率数据

Fig. 5 Unoccupied parking space rate of 37 parking lots

$P(D|S)$ 的值如表 1 所示。计算结果表明, 对该簇来说, 当停车场间具有空间相似性时, 可以认为其停车数据也是具有一定相似性的。

表 1 概率计算结果

Table 1 Probability table

概率类型	计算结果
$P(S)$	0.904
$P(D,S)$	0.723
$P(D S)$	0.800

6 测试样本数据正态性

假设该簇中物资大厦停车场为缺失数据的停车场, 其 9 维向量为 $[(22.560\ 64, 114.127\ 5), (3, 2, 3, 0, 7, 4, 4)]$ 。根据公式 (3) 计算得到与其空间相似度最高的停车场为百汇大厦停车场, 其 9 维向量为 $[(22.561\ 8, 114.113\ 3), (2, 2, 3, 1, 4, 4, 5)]$, 可将该停车场数据作为样本数据。2016 年 7 月共有 31 天, 每天按 5 min 为时间步长划分出 288 个时间节点, 即要对 31 条数据在 288 个时间节点上分别做正态性检验, 如图 6 所示。设置显著性水平为 0.05, 当计算得到的 P 值低于 0.05 时, 说明该组数据不符合正态分布, 结果如表 2 所示。

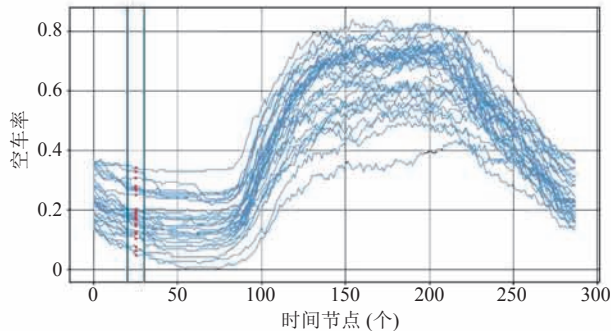


图 6 百汇大厦停车场数据正态性检验

Fig. 6 Normality test of Baihui building parking lot's data

表 2 正态性检验结果

Table 2 Normality test result

所有组数	非正态性组数	比例 (%)
288	15	5.2

显然, 根据计算结果有充分理由认为样本数据是符合正态分布的, 因此可以将该数据作为生成模型的学习样本。

7 生成修复数据

使用 RGAN 作为生成模型, 在每一次迭代学习的过程中, RGAN 都会学习样本集中一条数据的特征, 样本集中的数据将会被反复地学习, 生成器则要尽可能地生成与样本集非常相似的数据。在本次实验过程中, 部分参数如表 3 所示, 生成器生成数据的变化过程如图 7 所示。

表 3 神经网络参数

Table 3 Neural networks parameters

隐层神经元	批处理大小	时间步长	学习率
50	1	288	0.003

每生成一条数据后, 将其放入贝叶斯过滤器进行判断。如果生成的结果不会降低样本集的条件概率值, 则该生成结果将会被保留, 部分被保留的生成结果如图 8 所示。

最后, 使用同样的方法生成 31 条数据作为物资大厦停车场 2016 年 7 月的修复数据 (图 9), 并使用卡方检验验证生成数据是否有效。设置显著性水平为 0.05, 在每一个时间节点, 将生成数据与真实数据用公式 (8) 做卡方检验, 结果如表 4 所示。

表 4 卡方检验

Table 4 Chi-square test

所有组数	非同分布组数	比例 (%)
288	15	5.21

结果表明, 生成数据与真实数据是较为相似的, 即使用本方法生成的修复数据是有效的。但从图 9 可以看出, 修复的结果仍然存在一些相位偏差和噪声, 这些问题可能是由 RGAN 本身的缺陷所造成的。同时, 贝叶斯过滤器的设置也不足以使训练过程更加稳定。此外, 生成结果的峰

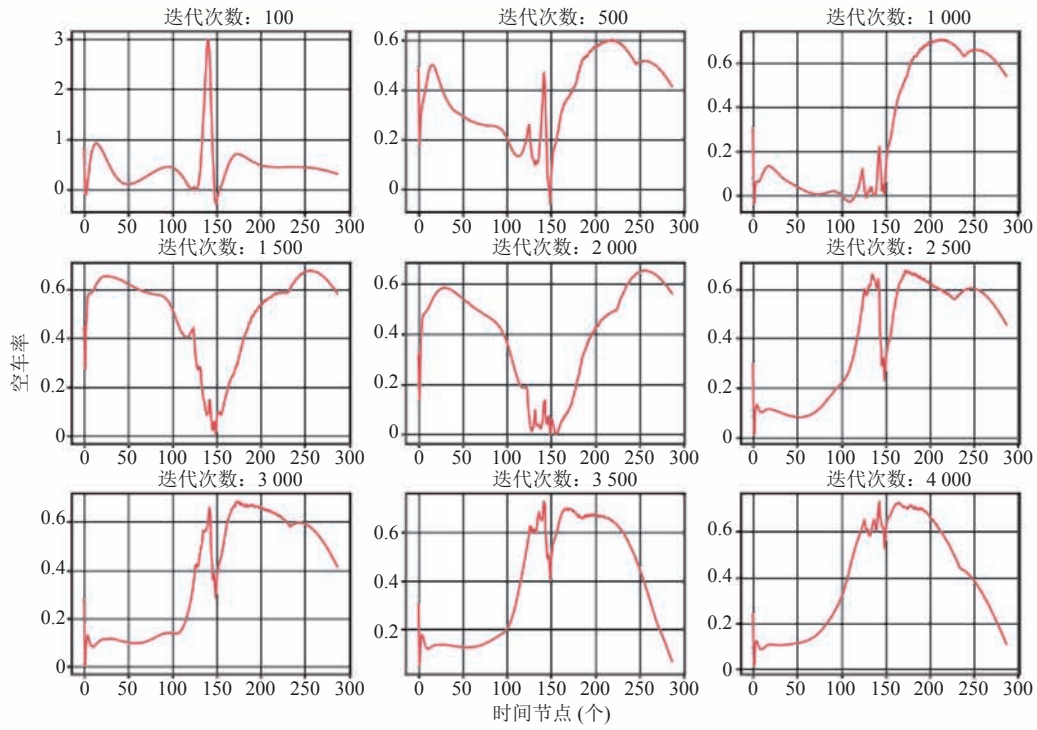


图7 循环生成式对抗网络训练过程

Fig. 7 Training process of recurrent generative adversarial networks

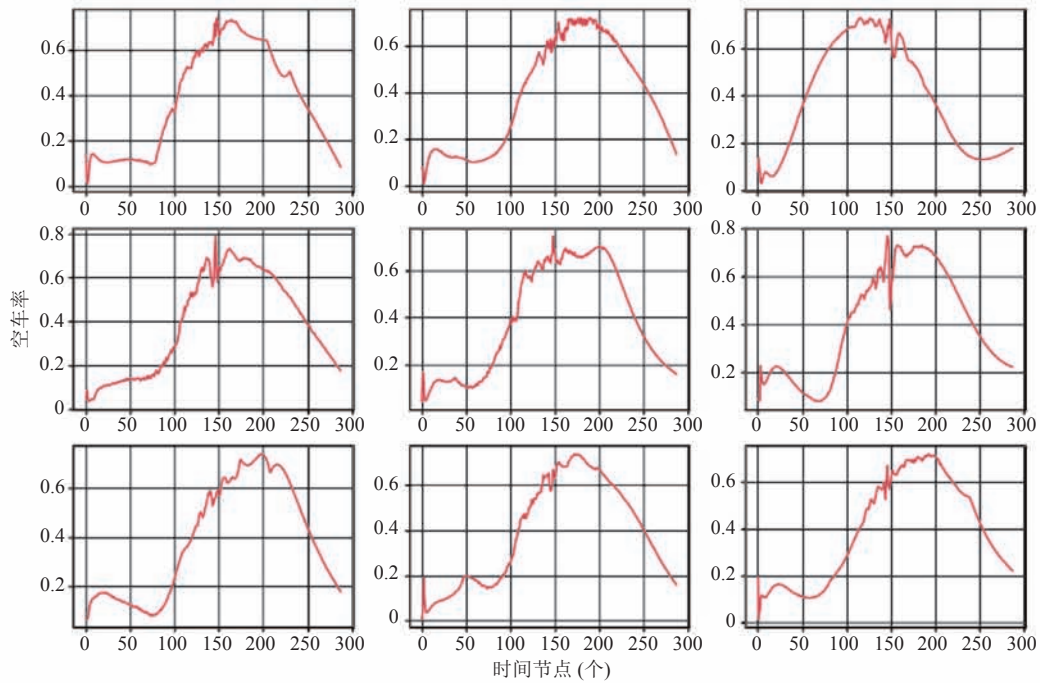


图8 贝叶斯过滤器筛选结果

Fig. 8 Generated data filtered by Bayes filter

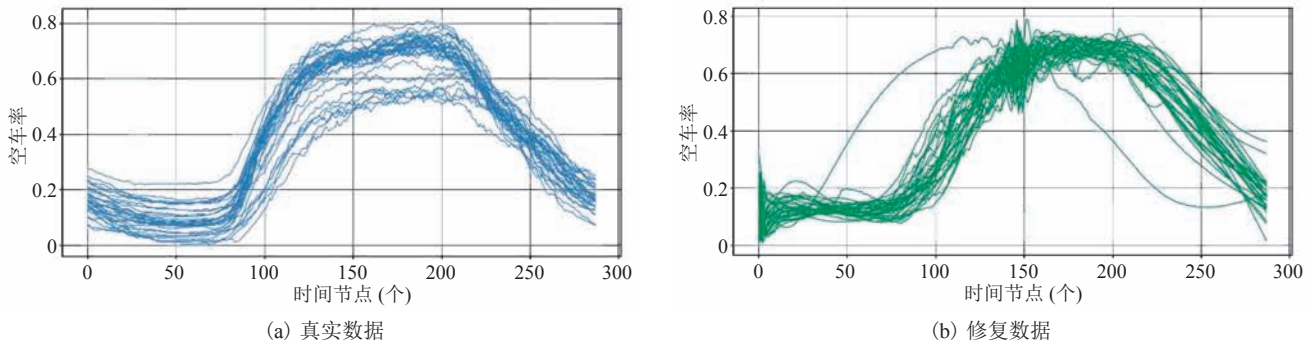


图 9 物资大厦停车场真实数据与修复数据对比图

Fig. 9 Real data and generated data of Wuzi building

值主要集中在 0.65~0.7, 与真实数据相比过于集中, 这也可能是 RGAN 对样本学习过于充分而导致的泛化性能不佳。考虑到数据修复本身的挑战性, 在一定程度上, 生成的结果是可以接受的。

8 与国内外相似研究的对比分析

在数据修复领域, 常用的方法主要有线性插值、基于期望最大化估计的多项式回归和核回归等。这些算法能够对 5 min 以内的缺失数据进行补齐, 但不能解决更大间隔的数据缺失问题。其中, 在完全无先验知识的情况下, 使用插值法更加无法完成数据修复工作。在动态交通流数据修补方面, 基于空间相关性的数据修复考虑到了相同道路横断面的不同车道之间、道路纵向上下游断面之间存在着的相关关系。其利用了相同断面不同车道之间历史参数的比例关系和已知车道的交通数据来推算未知车道的交通数据。该方法可以避免采用历史数据进行修复时不能反应实际交通状态的问题, 以提高修复数据的实时变化特性。相对于交通流数据来说, 由于停车场在地理空间上更加分散, 使得停车场间的数据并不具有连续性, 因而也无法采用类似于动态交通流数据修复的方法来实现停车场间的数据修复。

目前, 国内外对于在完全无先验知识的情况下进行停车场出入数据修复领域的研究非常匮

乏, 本文则针对该领域研究空白进行了一次全新尝试, 以当下最热门的生成式对抗网络作为生成模型来生成修复数据, 并在此基础上对生成式对抗网络进行改进。从实验结果来看, 本文提出的方法在一定程度上可以较为准确地实现停车场缺失数据的修复, 并且相对于在停车场部署数据采集设备的方法可以节省大量的时间与经济成本。

9 结 论

本文提出了一种停车场数据修复的方法, 可以在短时间内生成大量修复数据。首先, 本文提出停车场空间相似性的度量公式; 然后, 验证在停车场空间相似条件下停车场数据相似的概率; 最后, 当条件概率足够大时, 在同一簇中, 找到与目标停车场具有相似空间特征的停车场作为样本停车场, 并使用该停车场数据作为学习样本生成新数据完成目标停车场的的数据修复。仿真结果表明, 当停车场间具有空间相似性时, 其数据在很大概率上也具有相似性。使用 RGAN 作为生成模型生成的修复数据与真实数据有相同的分布特征。由于现阶段针对停车场数据修复的相关研究较少, 针对停车场数据缺失的情况, 常用的方法是通过在停车场加装额外数据采集设备来弥补数据量的不足。而本文提出的数据修复方法则可以在完全无先验知识和不加装设备的情况下, 在

短时间内生成大量合理数据, 因此能够节省大量的时间成本和经济成本。但目前本方法仍存在一些不足, 未来将会在 RGAN 结构上做进一步改进, 同时会将更多的时空特征纳入考虑范围, 以提升数据修复的效果。

参 考 文 献

- [1] Jaurkar HV, Mulay GN, Gohokar V. Parking guidance system using Internet of Things [C] // International Conference on Inventive Computation Technologies, 2016: 1-6.
- [2] A Graves. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [3] Zheng Y, Rajasegarar S, Leckie C. Parking availability prediction for sensor-enabled car parks in smart cities [C] // IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 2015: 1-6.
- [4] Yu F, Guo J, Zhu X, et al. Real time prediction of unoccupied parking space using time series model [C] // International Conference on Transportation Information and Safety, 2015: 370-374.
- [5] 韩超, 宋苏, 王成红. 基于 ARIMA 模型的短时交通流实时自适应预测 [J]. 系统仿真学报, 2004, 16(7): 1530-1532.
- [6] Rajabioun T, Ioannou PA. On-street and off-street parking availability prediction using multivariate spatiotemporal models [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(5): 2913-2924.
- [7] DongS, Chen M, Peng L, et al. Parking rank: a novel method of parking lots sorting and recommendation based on public information [C] // The 23rd IEEE International Conference on Industrial Technology, 2018: 1381-1386.
- [8] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C] // International Conference on Neural Information Processing Systems, 2014: 2672-2680.
- [9] Im DJ, Kim CD, Jiang H, et al. Generating images with recurrent adversarial networks [J]. arXiv: 1602.05110, 2016.
- [10] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [J]. arXiv: 1511.06434, 2015.
- [11] Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis [C] // Proceedings of the 33rd International Conference on Machine Learning, 2016: 1060-1069.
- [12] Yu L, Zhang W, Wang J, et al. SeqGAN: sequence generative adversarial nets with policy gradient [J]. arXiv: 1609.05473, 2017.
- [13] Mogren O. C-RNN-GAN: continuous recurrent neural networks with adversarial training [J]. arXiv: 1611.09904, 2016.
- [14] Esteban C, Hyland SL, Rättsch G. Real-valued (medical) time series generation with recurrent conditional GANs [J]. arXiv: 1706.02633, 2017.