

基于时空循环卷积网络的都市区域人口流量预测

郭晟楠 林友芳 金文蔚 万怀宇

(北京交通大学计算机与信息技术学院 北京 100044)

(北京交通大学交通数据分析与挖掘北京市重点实验室 北京 100044)

(北京交通大学综合交通运输大数据应用技术交通运输行业重点实验室 北京 100044)

摘要 都市区域人口流量的准确预测可以为交通监管和市民出行提供有效的决策支持。城市各区域人口流量同时具有时间维度上的变化规律和空间维度上的相关性,这给流量的精准预测带来了极大的挑战。文中提出了一种基于注意力机制的时空循环卷积网络(ASTRCNs)模型,可以全面地对影响区域人口流量的多种因素进行统一建模。ASTRCNs 共包含 3 个组件,分别用于描述人口流量的短时依赖关系、日周期规律、周周期规律。在真实的北京市人口流量数据集上进行了实验,结果表明 ASTRCNs 模型的预测效果优于传统的时间序列预测模型以及其他现有的基于深度学习的人口流量预测模型。

关键词 时空数据,人口流量预测,深度学习,循环卷积网络

中图法分类号 TP181 文献标识码 A

Citywide Crowd Flows Prediction Based on Spatio-Temporal Recurrent Convolutional Networks

GUO Sheng-nan LIN You-fang JIN Wen-wei WAN Huai-yu

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

(Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China)

(Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China)

Abstract Accurately forecasting the crowd flows in urban areas can provide effective decision-making support for traffic management and citizens' travel. The crowd flows in each urban region have strong correlations in both temporal dimensions and spatial dimensions. These complex factors bring great challenges to accurate predictions. A novel neural network structure named attention-based spatio-temporal recurrent convolution networks (ASTRCNs) was proposed, which can simultaneously model various factors that affect the crowd flows. ASTRCNs consists of three components, which can respectively capture the short-term dependences, the daily periodicity influence and the weekly patterns of the crowd flows. Experimental results on a real data set of crowd flows in Beijing demonstrate that the proposed ASTRCNs outperforms the classical time series methods and the existing deep-learning based prediction methods.

Keywords Spatio-Temporal data, Crowd flows prediction, Deep learning, Recurrent convolutional networks

1 引言

在智能交通系统(Intelligent Transportation System, ITS)中,预测城市未来各个区域的人口流量对加强城市公共安全和提高城市交通效率具有重要的作用,并能有效减少城市交通拥堵造成的经济损失。例如,2012年北京城市发展研究中心年度报告^[1]显示,该年度北京市因为交通延误造成的经济损失高达 1056 亿元。又如,在 2015 年上海外滩跨年夜活动中,大量人群短时间内涌入陈毅广场,造成了 36 人死亡、49 人受伤的惨烈事故。如果能够对城市任意区域未来的人口流量进行提前感知和预测,相关部门就可以采取提前预警、交通流管控等方法有效地避免此类事故的发生。

根据现实需求,本文将研究同时预测两种类型的区域人

口流量:流入人口数量(inflow)和流出人口数量(outflow)。如图 1 所示,给定一个时间区间,某区域在该时间段内的流入人口数量定义为在这段时间内从其他区域流入到该区域的人口总数;相应地,流出人口数量定义为在这段时间内从该区域流出到其他区域的人口总数。这两种类型的流量反映了城市区域之间的人口流动模式。在实际应用中,多种类型的城市数据都可以用来计算区域的人口流入、流出量,例如手机信令数据、出租车 GPS 轨迹数据、公共交通数据(公交车或地铁刷卡数据)等。

预测城市各区域的人口流入、流出量是一个非常具有挑战性的研究课题,因为每个区域的人口流量受到许多复杂因素的影响。1)在空间维度上相关联的区域的人口流量会相互影响。例如,图 1 中区域 r_2 的人口流入量受它临近区域(如 r_1

本文受国家自然科学基金项目(61603028)资助。

郭晟楠(1992-),女,博士生,CCF 会员,主要研究方向为时空数据挖掘、深度学习;林友芳(1971-),男,博士,教授,主要研究方向为数据挖掘、智能系统;金文蔚(1992-),男,硕士生,主要研究方向为交通数据挖掘;万怀宇(1981-),男,博士,副教授,CCF 会员,主要研究方向为交通数据挖掘、社交网络分析,E-mail:hywan@bjtu.edu.cn(通信作者)。

和 r_2) 的人口流出量的影响;与此同时,区域 r_2 的人口流出量又会影响到其周围区域的人口流入量,甚至 r_2 的人口流入量也会影响到其自身的人口流出量。随着时间的推移,较远区域的人口流量变化也会影响到区域 r_2 的人口流量。2) 在时间维度上,某区域一段时间后的人口流量与其当前时间以及前一段时间的人口流量都有很大的关系。另外,由人的作息规律可知,人口流量具有很明显的以日和周为单位的周期性规律,利用这些规律都可以帮助我们有效提高人口流量预测的准确性。

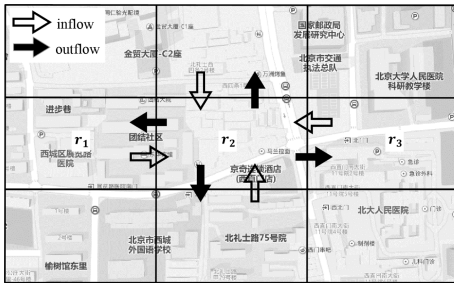


图1 给定区域的人口流入量和人口流出量

近年来,深度学习在很多应用领域都取得了很大的成功。卷积神经网络(Convolutional Neural Network, CNN)可以有效捕获网格类型数据的空间特征,因此被广泛应用于计算机视觉领域。长短期记忆模型(Long Short-Term Memory, LSTM)作为一种改进的循环神经网络,有效地缓解了传统循环神经网络在处理长序列数据时存在的梯度消失问题,被广泛应用于时间序列学习任务上。受此启发,为了同时刻画城市区域人口流量在时间和空间维度上的依赖特性,本文提出了一种端到端的预测模型——基于注意力机制的时空循环卷积神经网络(Attention-based Spatio-Temporal Recurrent Convolutional Networks, ASTRCNs),用于解决城市区域人口流量预测问题。

ASTRCNs 模型首先利用卷积神经网络捕获相邻区域之间人口流量的相互影响,然后利用长短期记忆模型刻画随着时间的推移人口流量的变化规律,最后该模型利用注意力机制动态地描述历史不同时刻对未来预测的不同重要程度。ASTRCNs 全面建模了人口流量的动态时空属性,提高了城市区域人口流量预测的准确性。

2 相关工作

近年来,交通预测已经成为智能交通系统中的一个核心问题。从研究对象的角度来看,交通预测问题可以分为路段导向和区域导向两类。在路段导向问题中,研究人员主要关心的是城市主干道或高速公路的交通状况检测和预测^[2-7],而区域导向问题更关心整个城市所有区域的交通状况。本文研究的问题属于区域导向的预测问题,这类问题受到城市计算领域学者们的高度重视^[8-11]。

交通预测问题的核心研究对象是时空序列交通数据。时空序列数据是一种特殊的时间序列数据,因此在交通领域的预测问题上,早期的研究主要采用经典的时间序列预测模型,如自回归滑动平均模型(Autoregressive Moving Average, ARMA)^[12]、自回归积分滑动平均模型(Autoregressive Integrated Moving Average, ARIMA)^[13-15]。基于 ARIMA 模型,一些扩展模型如 SARIMA^[16]、KARIMA^[17]、VAR^[18]以及 STARIMA^[19]等被提出以适应不同的预测问题。然而,此类时间序列预测模型在描述时空序列数据的空间依赖性时具有

很大的局限性,其预测结果通常很难满足人们的要求。

随着大数据技术越来越成熟,学者们在解决交通领域的预测问题时,开始关注并重点研究数据驱动模型。其中应用得最多的模型包括支持向量回归模型(Support Vector Regression, SVR)^[20]和人工神经网络(Artificial Neural Networks, ANNs)模型。其中,SVR 模型通过核函数将数据映射到高维空间以描述交通数据的非平稳变化特征^[21-23],但其预测结果的好坏在很大程度上取决于核函数的选择。很多学者针对不同的需求设计了不同的 ANNs 模型^[24-25]用于特定场景下的交通数据预测问题。

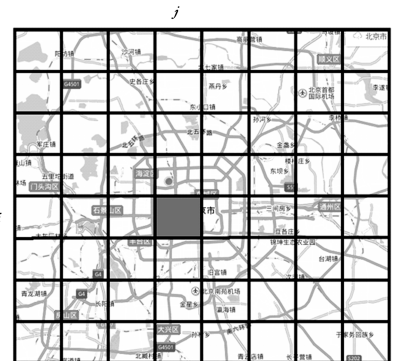
近年来,因为神经网络强大的表示能力,基于深度学习的模型在交通预测问题中的应用越来越广泛,此类模型中的两个重要分支分别是基于卷积神经网络的模型和基于循环神经网络(Recurrent Neural Networks, RNNs)的模型。卷积神经网络通过多次卷积操作和非线性变换,可以很好地描述数据的空间信息,在很多领域都取得了广泛的应用^[26-28]。在交通数据预测领域,Zhang 等^[8]利用网格化的方法对城市进行划分之后,提出了一种基于 CNN 的深度时空残差网络(ST-ResNet),用于预测整个城市的区域人口流量,但该模型没有充分挖掘人口流量数据的时间维度的特征。作为第二类重要分支的循环神经网络常被用于序列数据学习任务^[29-31]。其中,长短期记忆模型^[32]在文本分析^[33]、语音识别^[34]以及机器翻译^[35]等序列数据学习任务中都取得了巨大的成功。在交通数据预测领域,Zhao 等^[36]利用级联的 LSTM 模型进行交通流的短期预测,但该模型不能自动学习交通流数据的空间依赖关系。

在城市区域人口流量预测问题中,我们希望预测模型能够自动学习人口流量数据中的动态时间和空间依赖特征,而不需要人为处理数据提取特征。因此,本文提出了一种基于注意力机制并同时融合 CNN 和 LSTM 的深度学习网络结构来解决这一问题。

3 问题定义

本节对城市区域人口流量预测问题进行了形式化描述,首先给出城市区域人口流量的形式化表示,然后在此基础上对问题进行形式化定义。

如图 2 所示,根据经纬度将一个城市均匀划分为 $I \times J$ 的地理网格,其中 I 和 J 分别表示行数和列数,网格中的一个单元 (i, j) 表示一个区域。



注: (i, j) 表示一个区域

图2 京市网格划分示意图

城市区域人口流量预测的目标是根据已观测到的城市区

域历史人口流量数据来预测未来一段时间内各区域的人口流入量和流出量。显然,城市区域人口流量预测问题是一个时空序列预测问题。

不同来源的城市数据均可以用来计算城市区域的人口流入量和流出量,其中轨迹数据是最重要的一类数据。轨迹数据是人或者物移动生成的三元组 (x, y, τ) 序列,其中 (x, y) 是移动物体在时间 τ 的地理坐标,对于每一个移动物体 r ,其轨迹 T_r 可以表示为一条按照时间顺序排列的点序列:

$$T_r: p_1 \rightarrow p_2 \rightarrow \dots \rightarrow p_k \rightarrow \dots \rightarrow p_l$$

其中, l 表示轨迹 T_r 的总长度,第 k 个轨迹点 $p_k = (x_k, y_k, \tau_k)$ 表示物体 r 在时间 τ_k 的地理坐标为 (x_k, y_k) 。

定义 1 给定某一时间区间 t ,某一移动物体 r 生成的轨迹为 T_r , $q_{T_r}^t$ 表示轨迹 T_r 在时间区间 t 内经过的区域集合:

$$q_{T_r}^t = \{(i, j) | \tau_k \in t \wedge (x_k, y_k) \in (i, j)\} \quad (1)$$

其中, (i, j) 表示一个区域。对于一个移动物体 r ,当选择不同大小的时间区间时,其所经过的区域集合可能是不同的,用 $|q_{T_r}^t|$ 表示该物体在时间区间 t 内经过的区域数。 $q_{T_r}^t$ 在理论上最少包含 1 个区域,无上限。

定义 2 假设 S 表示连续 3 个时间区间 $t-1, t$ 和 $t+1$ 内所有移动物体产生的轨迹的集合,对于区域 (i, j) ,其在时间区间 t 内产生的人口流入量 $x_t^{in,i,j}$ 和人口流出量 $x_t^{out,i,j}$ 分别定义为:

$$x_t^{in,i,j} = |\{T_r | T_r \in S \wedge (i, j) \notin q_{T_r}^{t-1} \wedge (i, j) \in q_{T_r}^t\}| \quad (2)$$

$$x_t^{out,i,j} = |\{T_r | T_r \in S \wedge (i, j) \in q_{T_r}^t \wedge (i, j) \notin q_{T_r}^{t+1}\}| \quad (3)$$

其中, $|\cdot|$ 表示集合的基数,式(2)中 $t > 1$,式(3)中 $t \geq 1$ 。

基于以上形式化定义,区间 (i, j) 在时间区间 t 内的人口流量可以表示为一个向量 $\mathbf{X}_t^{i,j} \in \mathbb{R}^2$,其中 $(\mathbf{X}_t^{i,j})_0 = x_t^{in,i,j}$, $(\mathbf{X}_t^{i,j})_1 = x_t^{out,i,j}$ 。一个被划分为 $I \times J$ 个区域的区域在时间区间 t 内的全部区域人口流量可以表示为一个张量 $\mathbf{X}_t \in \mathbb{R}^{2 \times I \times J}$,其中 $(\mathbf{X}_t)_{i,j} = \mathbf{X}_t^{i,j}$ 。

人口流量预测问题的定义为:给定城市人口流量历史观测值 $\{\mathbf{X}_t | t = 1, 2, \dots, n\}$,预测 $\mathbf{X}_{t+\Delta t}$ 。其中, $\Delta t \in \{1, 2, \dots\}$ 表示待预测的时间区间与当前时间区间 t 之间的跨度。

4 基于注意力机制的时空循环卷积神经网络模型

本节将详细介绍基于注意力机制的时空循环卷积神经网络 ASTRCNs 的结构,该网络通过有机融合 CNN 和 LSTM 来同时捕获人口流量预测问题中的时间和空间依赖关系。

4.1 问题分析

在城市区域人口流量预测问题中,存在 3 类时间依赖特性,分别为:短时依赖(Closeness)特性、日周期依赖(Daily influence)特性和周周期依赖(Weekly influence)特性。例如,图 3 为一个由出租车轨迹数据生成的某段时间内北京市所有区域人口流量均值统计图,横坐标表示以小时为单位的时间区间,纵坐标表示所有区域人口流入量均值。图 3 很明显地展示了城市区域人口流量的 3 类时间依赖特性。

1) 短时依赖特性:指某个区域在刚刚过去的时间区间内的人口流量对预测该区域在未来某时间区间人口流量的影响,即相邻时间区间的区域人口流量之间存在很强的相关性。

2) 日周期依赖特性:体现了区域人口流量以天为单位的周期性规律。如图 3 中人口流量的早高峰常常发生在每天上午 8 点—10 点,即人口流量变化会受到人们的每日作息规律的影响。

3) 周周期依赖特性:体现了区域人口流量以星期为单位的周期模式。如图 3 中城市区域人口流量在周末和工作日具有明显不同的模式。

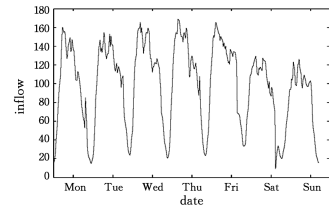
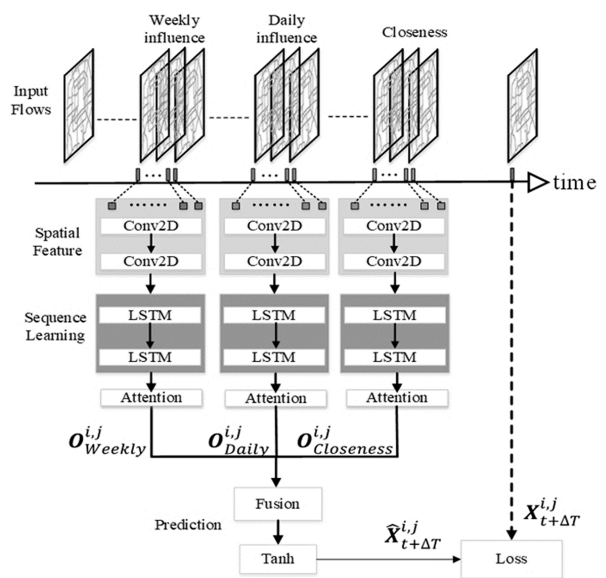


图 3 北京市所有区域人口流入量均值统计图

4.2 模型结构

图 4 给出了本文提出的基于注意力机制的时空循环卷积神经网络模型的完整网络结构。ASTRCNs 一共包含 3 个组件,分别是短时依赖组件、日周期依赖组件和周周期依赖组件。



注:Conv2D 表示 2 维卷积,LSTM 表示长短期记忆模型,Attention 表示注意力层

图 4 ASTRCNs 网络结构

图 4 的顶端表示模型的输入。本文中,给定任意时间区间,根据定义 1 和定义 2,我们可以将每个时间区间的城市区域人口流入量、流出量转换为一个类似于图片的两通道张量。因此输入历史观测值 $\{\mathbf{X}_t | t = 1, 2, \dots, n\}$,我们可以从中提取出体现 3 类时间依赖特性的子序列。

短时子序列:

$$[\mathbf{X}_{t-(l_c-1)}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_t] \quad (4)$$

日周期子序列:

$$[\mathbf{X}_{t-d-(l_d-1)}, \dots, \mathbf{X}_{t-d-1}, \mathbf{X}_{t-d}] \quad (5)$$

周周期子序列:

$$[\mathbf{X}_{t-w-(l_w-1)}, \dots, \mathbf{X}_{t-w-1}, \mathbf{X}_{t-w}] \quad (6)$$

其中, l_w, l_d, l_c 表示周周期依赖、日周期依赖和短时依赖子序列长度, w 表示一周包含的时间区间个数, d 表示一天包含的时间区间个数。

基于深度学习的 ASTRCNs 模型中有很多需要学习的参数,因此需要大量训练样本。本文提出了一种局部化样本构造方法,在给定历史时间区间的全市区域人口流量数据上,可以构造出更多的训练样本。这种样本构造方法尤其适用于历史数据量不充分的情况。对于城市某一区域 (i, j) ,为了预测

该区域 $t+\Delta t$ 时的人口流量 $\mathbf{X}_{t+\Delta t}^{i,j}$, 我们可以以区域 (i,j) 为中心, 提取其周围 $N \times N$ 范围内的所有区域的历史人口流量 $\{\mathbf{X}_t^{i,j,N} | t=1, 2, \dots, n\}$ 作为样本的已知信息, 其中 $\mathbf{X}_t^{i,j,N} \in \mathbb{R}^{2 \times N \times N}$, $N < I$, $N < J$, $\mathbf{X}_t^{i,j,N}$ 的中心为 $\mathbf{X}_t^{i,j}$ 。如图 5 所示, 取 N 等于 5, 同一时刻每个区域都可以构造出一个独立的样本。

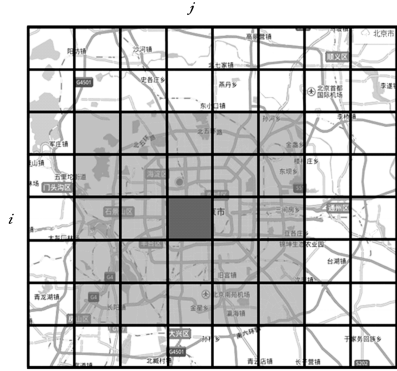


图 5 以区域 (i,j) 为中心取 N 等于 5 时构造的样本

图 4 中的 3 个子网络分别为对周周期依赖、日周期依赖和短时依赖 3 类依赖于子序列的处理组件。这 3 个组件具有相同的网络结构, 都是先利用二维卷积神经网络捕获相邻区域的人口流量在空间维度上的相互影响关系, 然后利用长短期记忆模型捕获区域人口流量在时间维度上的变化规律。需要注意的是, 历史不同时刻的流量对未来预测目标的影响是不同的。为了动态调整历史不同时刻对预测结果的影响程度, 我们引入了注意力机制。以短时依赖组件为例, 输入 $[\mathbf{X}_{t-l_c-1}^{i,j,N}, \dots, \mathbf{X}_{t-1}^{i,j,N}, \mathbf{X}_t^{i,j,N}]$, 经过多层卷积和 LSTM 神经网络处理后, 输出记为 $\mathbf{R}_{\text{Closeness}}^{i,j} = [\mathbf{R}_{t-l_c-1}^{i,j,N}, \dots, \mathbf{R}_{t-1}^{i,j,N}, \mathbf{R}_t^{i,j,N}] \in \mathbb{R}^{l_c \times u}$, 其中 u 表示最后一层 LSTM 的神经元个数。通过如下注意力机制, 动态求解每个历史时刻的信息的重要度:

$$\mathbf{a} = \text{softmax}(\tanh(\mathbf{R}_{\text{Closeness}}^{i,j} \mathbf{W}_a + \mathbf{b})) \mathbf{u}_a \quad (7)$$

$$\tilde{\mathbf{O}}_{\text{Closeness}}^{i,j} = (\mathbf{R}_{\text{Closeness}}^{i,j})^T \mathbf{a} \quad (8)$$

其中, $\mathbf{b}, \mathbf{u}_a \in \mathbb{R}^u$, $\mathbf{W}_a \in \mathbb{R}^{u \times u}$, $\mathbf{a} \in \mathbb{R}^{l_c}$ 。 $\tilde{\mathbf{O}}_{\text{Closeness}}^{i,j}$ 再经过全连接网络的处理, 输出与目标输出等维度的张量 $\mathbf{O}_{\text{Closeness}}^{i,j}$ 。

以上 3 个组件的输出分别为 $\mathbf{O}_{\text{Weekly}}^{i,j}$, $\mathbf{O}_{\text{Daily}}^{i,j}$ 和 $\mathbf{O}_{\text{Closeness}}^{i,j}$ 。在预测城市区域未来人口流量时, 需要同时考虑这 3 种依赖, 但是这 3 种因素对于预测结果的影响程度不一样, 因此 ASTRCNs 将采用加权融合的方法, 来对 3 个组件的输出进行加权融合:

$$\mathbf{X}_{\text{output}}^{i,j} = \mathbf{W}_w \circ \mathbf{O}_{\text{Weekly}}^{i,j} + \mathbf{W}_d \circ \mathbf{O}_{\text{Daily}}^{i,j} + \mathbf{W}_c \circ \mathbf{O}_{\text{Closeness}}^{i,j} \quad (9)$$

其中, \circ 表示哈达马乘积 (Hadamard Product), $\mathbf{W}_w, \mathbf{W}_d$ 和 \mathbf{W}_c 是 3 组需要学习的参数, 分别表示每种影响因素对预测结果的不同重要程度。

最后, ASTRCNs 模型在时间区间 $t+\Delta t$ 的预测结果用 $\hat{\mathbf{X}}_{t+\Delta t}^{i,j}$ 表示, 如式 (10) 所示, 我们用 \tanh 函数对 $\mathbf{X}_{\text{output}}^{i,j}$ 进行激活, 保证了输出结果的范围为 $-1 \sim 1$ 。

$$\hat{\mathbf{X}}_{t+\Delta t}^{i,j} = \tanh(\mathbf{X}_{\text{output}}^{i,j}) \quad (10)$$

ASTRCNs 模型的训练以最小化损失函数为目标。我们将损失函数定义为区域人口流量的真实值和预测值之间的均方误差:

$$\mathcal{L}(\boldsymbol{\theta}) = \|\mathbf{X}_{t+\Delta t}^{i,j} - \hat{\mathbf{X}}_{t+\Delta t}^{i,j}\|_2^2 \quad (11)$$

其中, $\boldsymbol{\theta}$ 包括 ASTRCNs 模型中所有需要训练的参数。

4.3 模型训练

算法 1 展示了 ASTRCNs 模型的参数训练过程, 本文采用误差逆传播算法进行模型的训练, 并配合 Adma 优化算法^[37] 来提高训练效率。

算法 1 ASTRCNs 训练算法

输入: 区域历史人口流量观察值 $\{\mathbf{X}_t | t=1, 2, \dots, n\}$; Weekly influence, Daily influence 及 Closeness 子序列的长度 l_w, l_d 和 l_c ; 周时间区间个数 w ; 天时间区间个数 d ; 预测时间区间跨度 $\Delta t \geq 1$

输出: ASTRCNs 模型参数

1. $\emptyset \rightarrow S$
2. for 所有可用时间区间 t do
3. for 所有区域 (i,j) do
4. $\mathbf{I}_w = [\mathbf{X}_{t-w-1}^{i,j,N}, \dots, \mathbf{X}_{t-1}^{i,j,N}, \mathbf{X}_t^{i,j,N}]$
5. $\mathbf{I}_d = [\mathbf{X}_{t-d-1}^{i,j,N}, \dots, \mathbf{X}_{t-1}^{i,j,N}, \mathbf{X}_t^{i,j,N}]$
6. $\mathbf{I}_c = [\mathbf{X}_{t-l_c-1}^{i,j,N}, \dots, \mathbf{X}_{t-1}^{i,j,N}, \mathbf{X}_t^{i,j,N}]$
7. $\mathbf{X}_{t+\Delta t}^{i,j}$ 是待预测的在 $t+\Delta t$ 时目标区域的人口流量
8. 将样本 $(\{\mathbf{I}_w, \mathbf{I}_d, \mathbf{I}_c\}, \mathbf{X}_{t+\Delta t}^{i,j})$ 加入训练集
9. end for
10. // 模型训练
11. 初始化 ASTRCNs 中所有需要学习的参数 $\boldsymbol{\theta}$
12. repeat
13. 从训练集中随机选择一个子集 S_b
14. 用子集 S_b 寻找最优参数 $\boldsymbol{\theta}$ 以最小化式 (11)
15. Until 满足停止条件

算法 1 中的第 1-8 行是构造训练样本集 S 的过程, 第 11-14 行采用批处理的方式在训练样本上不断训练模型, 直到满足终止条件为止。

5 实验与结果分析

我们在真实数据集上测试了 ASTRCNs 模型, 并分析了参数设置对模型预测效果的影响。

5.1 数据集

数据集的统计信息如表 1 所列, 其中 MobileBJ 是基于某移动运营商提供的经过脱敏的手机与基站通信的信令信息生成的北京市区域人口流量数据。

表 1 数据集统计信息

数据集	MobileBJ
数据类型	基于手机信令的人口流量数据
城市	北京
时间跨度	2016-08-01 至 2016-08-05
时间区间长度/min	5
时间区间总数	1440
区域网格数量	(32,32)
区域人口流量取值范围	[0,2020]
外部数据	无
节假日	无
天气(晴、雨等)/种	23
温度	[-27,4.42,6]
风速/级	[0,12]

5.2 数据预处理

MobileBJ 是一个时间跨度为 5 天的北京市区域人口流量数据集, 我们将前 4 天的数据当作训练集, 最后一天的数据当作测试集。在训练过程中采用早停策略来避免模型过拟合。

在模型训练之前, 需要对数据进行必要的预处理。本文采用 Min-Max 方法将区域人口流量归一化到 $[-1, 1]$ 之间。

最终在评价模型的预测效果时,再将预测值重新还原到原来的大小再与真实值进行比较。

可以根据不同的预测需求将城市划分为不同大小的区域,本文使用的数据集 MobileBJ 为北京市的数据,为了进行有效的实验对比,依据文献[8],我们将北京市 6 环以内的区域划分为 32×32 的网格,每个网格的长和宽大约为 1.8 km。

5.3 基准方法与评价指标

本文将 ASTRCNs 与以下 6 种方法进行比较。

1) PreHA: 人口流量数据具有很强的短时依赖性,因此该方法直接将当前时间区间的人口流量作为下一时间区间的预测值;

2) HA: 人口流量数据具有以天为单位的周期性规律,因此该方法用每天对应时间区间的历史人口流量的平均值作为预测值;

3) ARIMA: 自回归滑动平均求和模型,是一种应用广泛的时间序列预测方法;

4) LSTM: 长短期记忆模型,是一种循环神经网络模型,擅长处理序列类型的数据;

5) Conv1DNet: 一维卷积神经网络(在时间维度上进行卷积),可以用于时间序列预测问题;

6) ST-ResNet: ST-ResNet 模型是文献[8]提出的一种基于残差学习的神经网络模型。

本文使用均方误差 RMSE(Root Mean Square Error)作为模型预测效果的评价指标。

$$RMSE = \sqrt{\frac{1}{Z * T} \sum_i (x_i - \hat{x}_i)^2} \quad (12)$$

其中, Z 表示城市的区域总数, T 表示预测的时间区间总数, \hat{x}_i 是预测值, x_i 是真实值。

5.4 实验分析

我们在 MobileBJ 数据集上验证了 ASTRCNs 预测下一个时间区间城市区域人口流量的效果。模型的超参数设置如表 2 所列,预测结果如表 3 所列。

表 2 ASTRCNs 中的超参数设置

参数	值	备注
卷积核数目	64	卷积层中的卷积核数目
卷积核大小	(3,3)	每个卷积核的大小
Dropout 大小	0.25	Dropout
时间依赖长度	3	LSTM 中的时间区间步数
学习速率	0.0002	学习速率
激活函数	ReLU	卷积操作后的激活函数
激活函数	Tanh	输出层的激活函数
优化器	Adam	训练时用的优化器
批大小	1024	训练模型时的批大小

表 3 MobileBJ 数据集下的实验结果

模型	备注	RMSE
PreHA		61.43
HA		61.81
ARIMA		35.47
LSTM		43.65
Conv1DNet		33.76
ST-ResNet		122.31
ASTRCN_1	Closeness	30.32
ASTRCN_2	Closeness+Daily	31.41

实验使用了两个版本的 ASTRCNs 模型:ASTRCN_1 只考虑了时间依赖当中的 Closeness,ASTRCN_2 则同时考虑

到了 Closeness 和 Dailyinfluence。由于 MobileBJ 只有连续 5 天的数据,因此 Weekly influence 在本次实验中没有被考虑。表中展示的实验结果均是 10 次实验的平均值。

如表 3 所列,只包含短时依赖的 ASTRCN_1 取得了最好的预测效果。同时考虑短时依赖和日周期特征的 ASTRCN_2 在下一时刻人口流量预测任务上的结果略低于 ASTRCN_1。在 MobileBJ 数据集中,ST-ResNet 模型的效果很差,其主要原因是 MobileBJ 数据集只有 5 天的人口流量数据,按照 ST-ResNet 模型构造样本的方式,仅能构造 1440 个样本数据用于训练,但作为一种深度神经网络模型的 ST-ResNet 需要大量的样本来进行训练,因此 MobileBJ 的数据远不够训练出好的参数。而我们提出的 ASTRCNs 可以只考虑区域周围的局部空间依赖,每个区域可以当作一个独立的样本,因此可以构造出更多的训练样本。从这个角度来说,ASTRCNs 大大降低了对可利用数据的时间跨度要求。

在实际应用当中,预测更久时间之后的人口流量更具有应用价值,也更具有挑战性。因此,我们又连续预测了 6 个时间区间(共 30 min)的区域流量。如图 6 所示,由于没有足够的样本数据,ST-ResNet 的结果较差,因此没有将其作为对比模型。

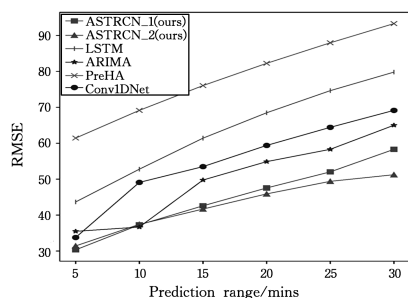


图 6 不同方法在数据集 MobileBJ 上的实验结果

由图 6 可以看出,ASTRCNs 比所有基准模型的表现都好,且随着预测时间的加长,ASTRCNs 的 RMSE 增长速度更缓慢。对于长时间跨度的预测,ASTRCN_2 的优势更加明显,说明人口流量数据以日为周期的规律对提高长时间跨度预测结果的准确率起到了更重要的作用。

5.6 ASTRCNs 在不同参数下的实验结果

神经网络模型的训练需要人工设定一些超参数,本小节将分析 ASTRCNs 在不同超参数条件下的实验结果对比(本小节的实验只利用近期组件),结果如表 4 所列。

由表 4 可知,LSTM 模块在时间依赖长度设为 3 步时表现最好。这表明在城市区域人口流量预测问题中,适当的时间依赖长度对模型来说是比较重要的。

表 4 中随着 LSTM 层数的增加,RMSE 先减少后增加。这表明 LSTM 的层数越多,ASTRCNs 模型能够刻画的信息就越多,但是 LSTM 层数过多时,训练变得困难,预测结果也不再提升。

不同的卷积核刻画了数据不同的特征,增加卷积核个数,就增加了卷积神经网络自动学习到的特征数量。表 4 中的结果显示,随着卷积核个数的增加,RMSE 先减少后增加,这表明卷积核个数越多,模型对数据的特征刻画就越全面,但是卷积核过多,也可能造成过拟合。

表4 不同超参数下 ASTRCNs 模型在 MobileBJ 上的实验结果比较

模型	时间依赖	卷积核	卷积核	LSTM	二维卷积	
	长度	大小	个数	层数	层数	
时间依赖长度						
ASTRCNs_TS1	1	33	64	2	2	33.07
ASTRCNs_TS2	2	33	64	2	2	31.12
ASTRCNs_TS3	3	33	64	2	2	30.32
ASTRCNs_TS6	6	33	64	2	2	31.24
ASTRCNs_TS9	9	33	64	2	2	32.00
ASTRCNs_TS12	12	33	64	2	2	32.86
LSTM 层数						
ASTRCNs_LN1	3	33	64	1	2	31.20
ASTRCNs_LN2	3	33	64	2	2	30.32
ASTRCNs_LN3	3	33	64	3	2	40.24
ASTRCNs_LN4	3	33	64	4	2	45.58
卷积核个数						
ASTRCNs_FN16	3	33	16	2	2	32.54
ASTRCNs_FN32	3	33	32	2	2	31.37
ASTRCNs_FN64	3	33	64	2	2	30.32
ASTRCNs_FN128	3	33	128	2	2	31.34
Conv2D 层数						
ASTRCNs_CN1	3	33	64	2	1	30.75
ASTRCNs_CN2	3	33	64	2	2	30.32
ASTRCNs_CN3	3	33	64	2	3	31.31
ASTRCNs_CN5	3	33	64	2	4	33.77

注:加粗超参数为变化的,其他超参数不变

深度卷积神经网络模型中,随着卷积层数的增加,模型表达能力更强。表4中的结果说明,适当增加卷积层数可以获得更好的结果,但是卷积层数也不宜过深。

结束语 本文提出了一种基于注意力机制的时空循环卷积网络模型(ASTRCNs)用于解决城市区域人口流量预测问题。ASTRCNs模型可以同时影响区域人口流量变化的两类因素(即空间依赖和时间依赖)进行建模。ASTRCNs结合卷积神经网络和长短期记忆模型自动学习人口流量数据的时空特征,并结合注意力机制动态调整不同历史数据对预测目标的重要性,最终通过自适应学习权重的方式,对模型的各个组件的输出结果进行加权融合。我们将ASTRCNs模型和经典的时间序列预测模型、最新的基于深度学习的预测模型在真实数据集上进行了对比实验,实验结果表明ASTRCNs取得了最优的预测效果。

实际上,许多类型的城市数据都可以用于区域人口流量预测,比如公共交通的刷卡数据、道路上的传感器数据等,但当前大部分预测模型都只考虑了单一的数据集。在未来的工作中,我们将尝试融合多源数据的预测模型,以达到更精准的预测效果。

参考文献

[1] 北京交通发展研究中心. 2012北京市交通发展年度报告[R]. 北京:北京交通发展研究中心,2012.

[2] SUN Z, FOX G. Traffic flow forecasting based on combination of multidimensional scaling and SVM[J]. International Journal of Intelligent Transportation Systems Research, 2014, 12(1): 20-25.

[3] JARV O, AHAS R, SALUVEER E, et al. Mobile phones in a traffic flow: a geographical perspective to evening rush hour traffic analysis using call detail records[J]. PloS One, 2012, 7(11): e49171.

[4] JIN B, CUI Y, ZHANG F. Fusing static and roving sensor data for detecting highway traffic conditions in real time[C]//Proc of

the 40th IEEE Conference on Computer Software and Applications Conference. NJ:IEEE,2016,1:807-816.

[5] FOULADGAR M, PARCHAMI M, ELMASRI R, et al. Scalable deep traffic flow neural networks for urban traffic congestion prediction[C]//Proc of International Joint Conference on Neural Networks. NJ:IEEE,2017:2251-2258.

[6] SHANG J, ZHENG Y, TONG W, et al. Inferring gas consumption and pollution emission of vehicles throughout a city[C]//Proc of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York:ACM,2014: 1027-1036.

[7] YANG B, GUO C, JENSEN C. Travel cost inference from sparse, spatio temporally correlated time series using Markov models[J]. Proceedings of the VLDB Endowment, 2013, 6(9): 769-780.

[8] ZHANG J, ZHENG Y, QI D. Deep spatio-temporal residual networks for citywide crowd flows prediction[C]//Proc of the 31st AAAI Conference on Artificial Intelligence. California: AAAI, 2017:1655-1661.

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

[10] HOANG M, ZHENG Y, SINGH A. FCCF: forecasting citywide crowd flows based on big data[C]//Proc of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York:ACM,2016:6.

[11] CHEN H, GRANT-MULLER S. Use of sequential learning for short-term traffic flow forecasting[J]. Transportation Research Part C: Emerging Technologies, 2001, 9(5): 319-336.

[12] HAMILTON J. Time series analysis[J]. Economic Theory, 1995, 11: 625-630.

[13] BOX G, JENKINS G, REINSEL G, et al. Time Series Analysis: Forecasting and Control[M]. Hoboken: John Wiley & Sons, 2015.

[14] MOORTHY C, RATCLIFFE B. Short term traffic forecasting using time series methods[J]. Transportation planning and technology, 1988, 12(1): 45-56.

[15] LEE S, FAMBRO D. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting[J]. Transportation Research Record: Journal of the Transportation Research Board, 1999, 1678: 179-188.

[16] WILLIAMS B, HOEL L. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results[J]. Journal of Transportation Engineering, 2003, 129(6): 664-672.

[17] VAN D, 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.

[18] KAMARIANAKIS Y, PRASTACOS P. Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches[J]. Transportation Research Record: Journal of the Transportation Research Board, 2003, 1857: 74-84.

[19] KAMARIANAKIS Y, PRASTACOS P. SpaceCtime modeling of

- traffic flow [J]. Computers and Geosciences, 2005, 31(2): 119-133.
- [20] SMOLA A, SCHÖLKOPF B. A tutorial on support vector regression [J]. Statistics and computing, 2004, 14(3): 199-222.
- [21] WU C, HO J, LEE D. Travel-time prediction with support vector regression [J]. IEEE Transactions on Intelligent Transportation Systems, 2004, 5(4): 276-281.
- [22] ZHANG Y, LIU Y. Traffic forecasting using least squares support vector machines [J]. Transportmetrica, 2009, 5(3): 193-213.
- [23] WANG J, SHI Q. Short-term traffic speed forecasting hybrid model based on chaos-wavelet analysis-support vector machine theory [J]. Transportation Research Part C: Emerging Technologies, 2013, 27: 219-232.
- [24] YIN H, WONG S, XU J, et al. Urban traffic flow prediction using a fuzzy-neural approach [J]. Transportation Research Part C: Emerging Technologies, 2002, 10(2): 85-98.
- [25] ZHENG W, LEE D, SHI Q. Short-term freeway traffic flow prediction; Bayesian combined neural network approach [J]. Journal of transportation engineering, 2006, 132(2): 114-121.
- [26] WU C, HO J, LEE D. Travel-time prediction with support vector regression [J]. IEEE Transactions on Intelligent Transportation Systems, 2004, 5(4): 276-281.
- [27] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Imagenet classification with deep convolutional neural networks[C]//Proc of the 26th Annual Conference on Neural Information Processing Systems. Massachusetts; MIT Press, 2012: 1097-1105.
- [28] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436.
- [29] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [30] WILLIAMS R, ZIPSER D. A learning algorithm for continually running fully recurrent neural networks [J]. Neural Computation, 1989, 1(2): 270-280.
- [31] SUTSKEVER I, VINYALS O, LE Q. Sequence to sequence learning with neural networks[C]//Proc of the 28th Annual Conference on Neural Information Processing Systems. Massachusetts; MIT Press, 2014: 3104-3112.
- [32] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [33] SUTSKEVER I, MARTENS J, HINTON G. Generating text with recurrent neural networks[C]//Proc of the 28th International Conference on Machine Learning. New York; ACM, 2011: 1017-1024.
- [34] VINYALS O, RAVURI S, POVEY D. Revisiting recurrent neural networks for robust ASR[C]//Proc of 2012 IEEE International Conference on Acoustics, Speech and Signal Processing. NJ; IEEE, 2012: 4085-4088.
- [35] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [EB/OL]. (2016-05-19)[2018-07-05]. <https://arxiv.org/abs/1409.04736>
- [36] ZHAO Z, CHEN W, WU X, et al. LSTM network; a deep learning approach for short-term traffic forecast [J]. IET Intelligent Transport Systems, 2017, 11(2): 68-75.
- [37] KINGMA D, BA J. Adam: A Method for Stochastic Optimization[C]//Proc of the 3rd International Conference for Learning Representations. 2015.

(上接第364页)

结束语 本文介绍了一种针对无线传感器网络的轻量级加密算法,通过对WEP加密协议进行优化,用混沌序列优化了密钥产生器,较高的敏感度生成不重复的子密钥,解决了RC4应用过程中密钥空间不足的问题,同时没有在RC4的基础上增加计算时间和复杂度,可以适应无线传感器CPU有限的运算速度。握手和通讯协议针对常见攻击增加了保护,可以有效应对攻击者的窃听、干扰和伪造通讯。

参 考 文 献

- [1] 郭海霞,裘雪红. 嵌入式系统安全的研究与设计[J]. 电子科技, 2005, 8: 49-54.
- [2] 路安平,杨济民,李峰. 几种轻量级加密算法的比较研究[J]. 现代电子科技, 2014, 37(12): 37-41.
- [3] 刘宴兵,田四梅,唐浩坤,等. 基于混沌的+RC4+流加密算法[J]. 计算机工程, 2011, 37(2): 136-138.
- [4] 刘坤. 无线局域网的安全性研究[D]. 上海: 复旦大学, 2008.
- [5] LORENZ E N. Deterministic nonperiodic flow[J]. Journal of atmosphere science, 1963, 20: 130-141.
- [6] FREY D Q. chaotic digital encoding; an approach to secure communication[J]. IEEE Trans Circuit and Systems, 1993, 40(40): 660-666.
- [7] DEDIEU H, KENNEDY M P, HASLER M. chaos shift keying: modulation and demodulation of a chaotic carrier using self-synchronizing chua's circuit[J]. IEEE Trans Circuit and Systems, 1993, 40(40): 634-642.
- [8] 胡岗,萧井华,郑志刚. 混沌控制[M]. 上海: 上海科技教育出版社, 2000: 167-193.
- [9] 孟庆树,王丽娜,傅建明. 密码编码学与网络安全[M]. 北京: 电子工业出版社, 2006.
- [10] LIU M S, ZHANG Y, LI J H. Research on Improving Security of DES by Chaotic Mapping[C]//Proc. of the 8th International Conference on Machine Learning and Cybernetics. Baoding, China, 2009.
- [11] QIAN G B, JIANG Q F, QIU S S. A New Image Encryption Scheme Based on DES Algorithm and Chua's Circuit [C] // Proc. of International Workshop on Imaging Systems and Techniques. Shenzhen, China, 2009.
- [12] 廖晓峰,肖迪,陈勇,等. 混沌密码学原理及其应用[M]. 北京: 科学出版社, 2009.