

面向视频智能分析的商业街行人交通流预测建模

倪慧荟¹, 吴波鸿^{2,3}

1. 北京市劳动保护科学研究所安全与应急管理研究室, 北京 100054

2. 中国科学院科技战略咨询研究院, 北京 100190

3. 中国科学院大学, 北京 100049

摘要 阐述了利用面向视频智能分析技术,对商业区行人交通流数据进行样本提取、预处理和建模分析的方法全过程。以北京西单商业区为例,构建了包含不同类型监测点、不同时间点的日期分组式纵向时间序列,并完成了预测建模和效果对比。研究表明,所有序列均为平稳非白噪声序列,具有相似的自回归移动平均(ARMA)模型形式,能较好地实现对行人流量的预测。

关键词 视频智能分析;商业街行人交通流;预测建模

借助技术系统做好密集人群风险预警已在国内得到了广泛应用,目前主流的技术分为视频智能分析和手机定位两大类。要实现对人群聚集风险的预警,就必然要对客流数据做分析和预测建模。早期对于客流数据的研究大多集中于两类,一是研究交通站点的行人交通流。例如,以地铁站视频资料为基础,人工观测行人交通行为并进行计数,分析速度、候车位置分布、密度时空分布等^[1];二是对客流交通行为进行观测,例如研究步幅、步速、流量、步行设施等使用情况,总结出特定区域的行人交通现状和特性^[2-4]。由于对数据的分析来源于人

工测算,很难积累起长时间的大规模数据,因此类似的研究主要应用于步行路线的规划以及疏散通道的设计。

技术系统的应用,改变了原本人工观测的困境。以视频智能分析为例,它可以持续地从目标图像中提取流量、密度、速度等人群运动参数并进行存储记录,使得人群聚集状态的实时监测成为可能,同时也为行人交通流的规律提取奠定了基础。初期,各类客流预警系统普遍采用了“静态阈值预警法”,即为人群运动参数(如流量)设置一个静态的控制数(即阈值),当监测到流量超越该阈值时,

收稿日期:2019-02-21;修回日期:2019-05-12

基金项目:北京市科学技术研究院创新工程预研项目(PXM2018_178304_000011)

作者简介:倪慧荟,助理研究员,研究方向为安全与应急管理,电子信箱:nihuihui@kingehs.com

引用格式:倪慧荟,吴波鸿.面向视频智能分析的商业街行人交通流预测建模[J].科技导报,2019,37(16):74-82;doi:10.3981/j.issn.1000-7857.2019.16.009

就发出报警。应用实践证明,该方法有很显著的局限性。一是阈值的设定较为主观;二是通过一个绝对数控制警情导致波动性较大,频繁误报^[5]。

因此后续系统逐步改进为结合场景分析和算法设计,将视频智能分析得到的行人流数据转化为人群规律,基于人群规律得到的预测值与实时数据的偏离比较,从侧面度量人群聚集风险,取得了较好的效果^[6]。该方法的核心在于构造具有良好预测效果的行人交通流模型。但是传统的行人交通流建模方法大多采用元胞自动机^[7]、磁力^[8]、社会力^[9]等模型实现物理建模,并应用 Legion、STEPS、BuildingEXODUS 等模拟软件验证^[10]。这种并不直面数据,而只从微观角度探讨人群流动物理特性的建模方法无助于提升客流预警效果。

面向客流数据预测的研究大多以轨道交通作为研究对象,运用神经网络^[11]、SVM^[12]、云模型^[13]等进行建模和分析。一方面对客流数据来源要求较高,以出入口闸机为主;另一方面因模型算法较为复杂,多以离线分析为主要应用场景,难以直接应用于基于视频智能分析的客流预警系统。

本研究基于北京西单商业区人群预警系统^[6]

(以下简称预警系统)采集到的人群流量数据,构建了一整套从数据筛选到预测建模的方法,最终得到的模型可直接嵌套至人群预警系统实现无缝对接,为提升系统预测效果提供了可行的解决方案。方法原理同样也适用于具备相似数据条件的其他人员密集场所(例如交通枢纽、旅游景区等),可切实提升城市热点区域的大客流管理科技水平。

1 数据样本选取

1.1 监测点选择

商业街的行人交通视频监测点大致可分为天桥、街面、交通站点3种类型。其中交通站点类监测点由于缺乏垂直照射靶面(图1(a)),导致数据平均准确率在60%左右,可靠性不足,只能舍去。剩余2类监测点的数据质量也受到照射靶面、人流线和光照条件等因素的影响,呈现不同的完整度和准确性。本研究从中挑选了数据质量最优的第31号监测点(横二条东侧步行道,如图1(b)所示)作为街面代表,第4号监测点(1号桥西端,如图1(c)所示)作为天桥代表。



(a) 交通站点类监测点

(b) 街面类监测点

(c) 天桥类监测点

图1 3类典型监测点

Fig. 1 Three typical surveillance points

1.2 数据类型与组织方式

预警系统采集的数据包括人群流量、速度、密度三大类。由于智能分析技术成熟度的限制,流量数据是目前准确率相对最高的^[5],因此选择行人流量作为数据类型。预警系统的初版预测模型采用了基于横向时间序列(即同一监测点自某日0:00

至次日0:00之间所有时间点的监测数据,按照时间顺序排列而得到的序列)的移动平均模型^[14]。此外,若将同一监测点每天同一时间点的流量数据按照日期先后顺序进行排列,就可得到纵向时间序列。两者的直观区别如图2所示。

纵向时间序列与横向时间序列相比(图3),具

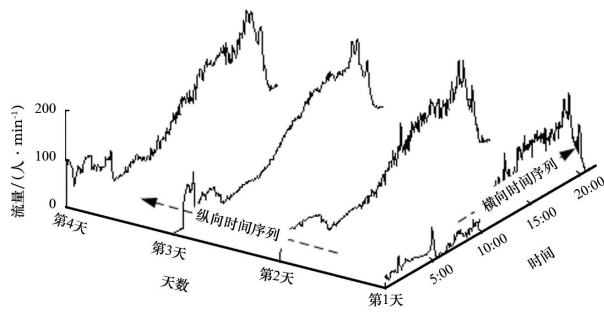


图2 横向时间序列与纵向时间序列

Fig. 2 Horizontal time series and vertical time series

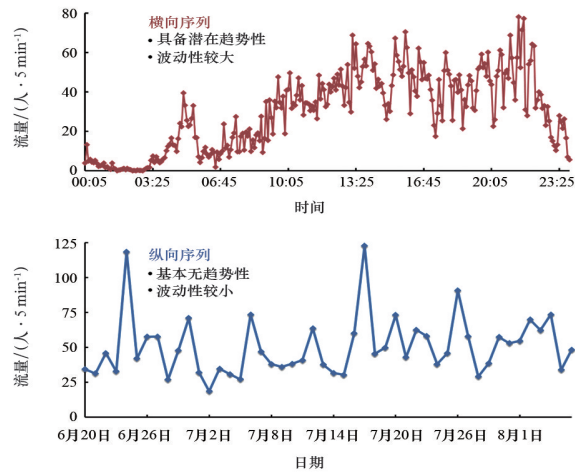


图3 横向时间序列与纵向时间序列对比示例

Fig. 3 Comparison between horizontal time series and vertical time series

有“基本无趋势性、波动性小”的显著优点,因此,本研究采用纵向时间序列的形式组织数据。观察历年的数据发现,2011年度目标监测点整体的数据完整性最优;同时,每日8:00前,因客流稀少,流量或趋于0,或波动性过大;20:00点后因夜晚光照条件不良而准确度不足。所以最终筛选出目标监测点2011年度每日8:00—20:00的流量数据作为研究对象。

1.3 时间点与日期处理

流量原始数据在系统中以6 s的时间间隔进行保存。为了降低波动性,将其合成为5 min尺度(根据系统试运行期间的测试,该时间尺度可以比较恰当地反映人群运动状态的变化^[14])的合成数

据,即每一个元数据都代表了5 min内某监测点的累积流量。从8:00—20:00,以5 min为刻度单元,共计145个时间点,即每一监测点共有145个纵向时间序列。考虑到很多邻接时间点具有类似的特征,所以只抽取少数典型时间点进行分析。

首先通过实地抽样调查发现,前往西单商业区的人群中,95%以上是去各大商场购物。各商场10:00开始正式营业。所以早8:00—10:00的流量数据不具有典型性,暂可不予考虑。其次,分析单日横向时间序列可以发现,人群流量从10:00起呈不断上升趋势,在14:00—15:00出现一个小高峰,之后平稳波动,在17:00—19:00出现又一个长时段高峰,之后呈逐步下降趋势。鉴于此,不失一般性,抽取10:30、14:30、17:30这3个典型时间点,构建纵向时间序列。

理论上,人们通常更倾向于在非工作日购物消费,因此周末和节假日的数据值普遍比工作日多一个增量,两者的数据形态区别显著。以图4为例,从数据形态看,工作日的数据极值基本均低于节假日(春节除外,春节期间北京大量外来人口流出,春节期间的行人流量甚至低于工作日数据)。

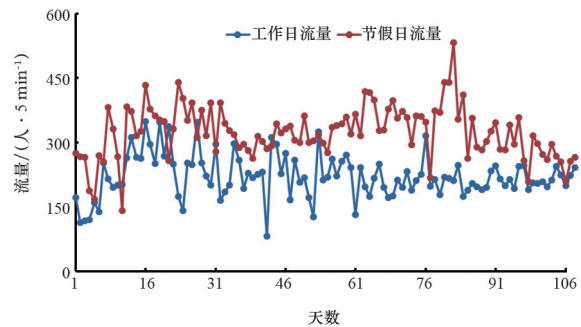
图4 工作日与节假日数据区别示例
(横二条10:30截取部分)

Fig. 4 Data variety between workday and holiday (data from Hengertiao at 10:30)

因此,为了保证分析的准确性,将工作日数据和周末、节假日数据进行分组处理。工作日数据作为一组,周末与节假日合并作为另一组。春节数据因过于特殊,所以暂时剔除。为叙述方便,定义以下序列符号:

$$HW_t^p, p = 10, 14, 17, t = 1, 2, \dots, 250 \quad (1)$$

$$HH_t^p, p = 10, 14, 17, t = 1, 2, \dots, 108 \quad (2)$$

$$BW_t^p, p = 10, 14, 17, t = 1, 2, \dots, 250 \quad (3)$$

$$BH_t^p, p = 10, 14, 17, t = 1, 2, \dots, 108 \quad (4)$$

式中, HW 、 HH 、 BW 、 BH 是序列名称, 其中首字母表示监测点: H 表示横二条, B 表示一号桥; 第 2 个字母表示日期组: W 表示工作日, H 表示周末与节假日; p 是时间点标志, 取值 10、14、17 分别对应抽样时间点为 10:30、14:30、17:30; t 是日期标志, 代表了不同组别 2011 年的不同日期 (按时序先后排列), 其中工作日组自 1 月 4 日—12 月 31 日, 共计 250 个数据点, 周末与节假日组自 1 月 1 日—12 月 25 日, 共计 108 个数据点。例如, HH_3^{14} 表示横二条节假日组第 3 个日期 (2011 年 1 月 3 日) 下午 14:30 对应的行人流量数据。后文中表示整个序列时, 通常省略下标 t 。

2 数据预处理

2.1 缺失值修复

观察发现, 部分序列中存在连续 0 值的情况 (表 1), 经视频验证, 该连续 0 值是由于智能分析设备或网络通信故障导致。为保证分析的准确性, 需修复这些异常 0 值。

常见修复方法包括个案剔除法、均值替换法、热卡填充法、回归替换法、多重替代法等^[15]。本研究待分析序列缺失数据比例低于 5%, 两组样本序列的数据量分别为 250 和 108, 属于中等偏小的规

表 1 缺失数据示例
Table 1 Example of missing data

日期	星期	10:30	14:30	17:30
2011-03-01	星期二	0	0	0
2011-03-02	星期三	0	0	0
2011-03-03	星期四	0	0	0
2011-03-04	星期五	0	255	348

模, 不适合使用个案剔除法; 由于缺失数据对应单一变量 (行人流量), 没有其他可比变量来计算关系矩阵, 也没有对应的自变量及数据构建回归方程, 所以热卡填充法与回归替换法同样不适用; 样本数据的统计特性并不具有先验知识, 需要在修复后才能进行相应的分析, 因此多重替代法也不适合。因此, 本研究使用均值替代法对缺失数据进行修复。

考虑到西单商业区行人流动受到季节、促销活动等因素的影响, 单纯用非缺失数据的样本均值修复 0 值并不完全合理。调查发现, 特定日期的数据与过去紧邻该日期的数据具有更强的相似性。因此, 本研究采用缺失数据对应日期的一周数据的算术平均值对 0 值进行修复, 即采用如下公式:

$$F_t^p = \frac{1}{7} \sum_{i=1}^7 F_{t-i}^p, p = 10, 14, 17 \quad (5)$$

式中, F 表示行人流量, 不再区分具体是那组序列, p 和 t 的含义同前。

2.2 异常极值修复

由于视频智能分析算法的局限性, 序列中还存在极高或极低的异常值, 如图 5 所示 (经视频验证)。为保证分析的准确性, 还需要修复这些异常极值。

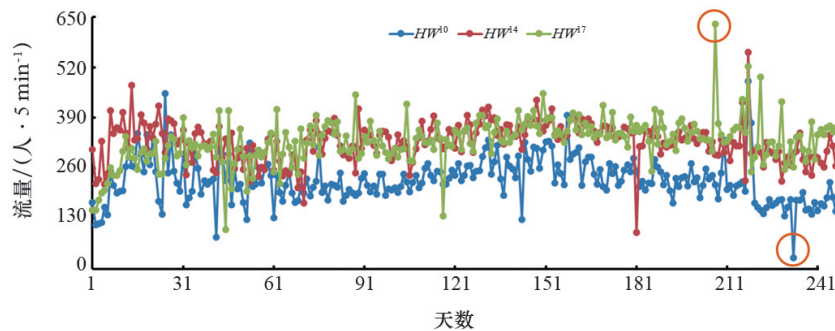


图 5 HW 组序列可能的异常值 (图中圈出的点)

Fig. 5 Possible extreme data of HW series (marked with circle)

检验极值的常见方法包括标准化分数法、Hampel identifier法、盒状图法和截尾平均值法等^[16]。标准化分数法要求数据变量符合正态分布假设,目前暂不适用。盒状图法和截尾平均值法计算过程相对繁琐,而且对于偏离程度并没有统一标准,在实际操作中存在困难。本研究采用Hampel identifier法进行异常极值检验。该方法中利用的 Z 值的计算公式为

$$\begin{cases} Z_i^p = \frac{|F_i^p - \text{Median}(F_i^p)|}{\left(\frac{M}{0.6745}\right)} \\ M = \text{Median}\left(|F_i^p - \text{Median}(F_i^p)|\right) \end{cases} \quad (6)$$

式中,Median表示取中位数, F 、 p 和 t 的含义同前。

按照式(6)计算后,各组序列的初步结果如表

表2 Hampel identifier法对各序列的检验结果
Table 2 Results of Hampel identifier applied to series

序列	极端值个数			序列	极端值个数		
	$Z > 2.24$	$Z > 3.5$	$Z > 4$		$Z > 2.24$	$Z > 3.5$	$Z > 4$
HW^{10}	13	4	3	BW^{10}	6	0	0
HW^{14}	12	3	2	BW^{14}	18	6	5
HW^{17}	17	7	4	BW^{17}	17	6	3
HH^{10}	4	1	0	BH^{10}	6	0	0
HH^{14}	21	9	5	BH^{14}	8	4	4
HH^{17}	7	3	3	BH^{17}	12	3	2

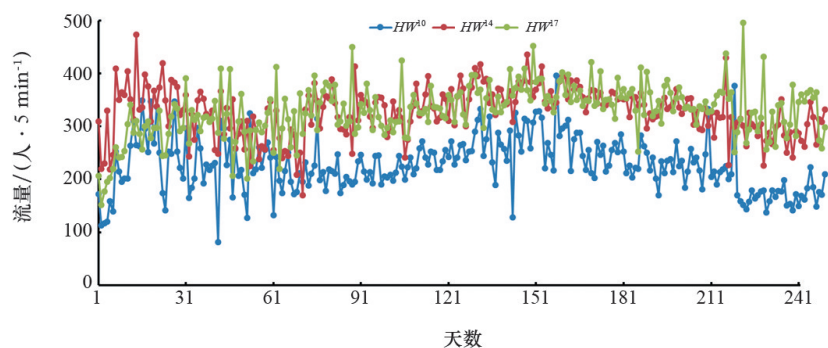


图6 HW 组序列的折线图(异常极值修复后)

Fig. 6 HW series with extreme data fixed

3 序列建模分析

3.1 平稳性检验

经预处理后的纵向时间序列首先应检验其平

2所示。通过观察可以发现,当 Z 值取2.24(即假定序列数据服从近似正态分布)时,各序列极值数量存在明显高估;当 Z 值取3.5(Hampel本人拟定的临界值)时,极值已明显减少,但对比历年数据发现,部分极值并未超越合理范围;于是继续逐步提升 Z 值并对比录像计数,直到 Z 值进一步提升至4以后,极值个数进一步被压缩,对比录像证实其确实已处于合理范围之外,可以被认定为异常极值。

上述异常极值同样采用均值替换法修复。不断重复“修复-Hampel identifier检验”循环,直至检验不到任何异常极值为止。至此,数据预处理工作基本完成。图6展示了 HW 组序列修复后的折线图,已基本不存在缺失值与异常极值,数据分布形态更为合理。

稳性和纯随机性,根据检验结果将之分为平稳序列和非平稳序列,进而采用不同的方法建模^[17]。本文使用具有广泛适用性的ADF检验(augmented Dickey-Fuller test)实施平稳性检验。以 HW^{10} 序列为

例,利用SAS程序处理后,可得到如图7所示结果。可见,在0.05的显著性水平下, HW^{10} 序列当采用常数均值后显著平稳。对其他所有序列同样执行以上操作,可知所有序列当采用常数均值后均显著平稳。

3.2 纯随机性检验

纯随机序列不具备任何分析的价值。因此,在

平稳性检验的基础上,本研究使用较可靠的LB(Ljung-Box)统计量进一步进行纯随机性检验。同样以 HW^{10} 序列为例,利用SAS程序处理后,可得到如图8所示的结果, HW^{10} 序列在显著性水平99.99%以上可判定为非纯随机序列。对其他所有序列同样执行以上操作。可知所有序列均显著为非白噪声序列。

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-6.8870	0.0694	-1.85	0.0619		
	1	-2.4693	0.2800	-1.06	0.2594		
Single Mean	0	-145.603	0.0001	-10.13	<.0001	51.29	0.0010
	1	-94.0104	0.0015	-6.93	<.0001	24.04	0.0010
Trend	0	-147.434	0.0001	-10.21	<.0001	52.15	0.0010
	1	-96.5907	0.0006	-7.05	<.0001	24.98	0.0010

注:检验结果中lags表示滞后阶数;Rho、Tau、F均为ADF检验统计量;ADF检验中的3类模型形式:Zero Mean—零均值;Single Mean—常数均值;Trend—趋势项。

图7 HW^{10} 序列 ADF 检验结果
Fig. 7 Results of ADF test for HW^{10} series

To Lag	Chi-Square	DF	Pr > ChiSq	-----Autocorrelations-----					
6	176.41	6	<.0001	0.415	0.351	0.350	0.303	0.301	0.301
12	264.99	12	<.0001	0.264	0.255	0.314	0.278	0.085	0.146
18	311.13	18	<.0001	0.183	0.219	0.118	0.160	0.124	0.189
24	344.16	24	<.0001	0.203	0.163	0.102	0.170	0.107	0.042

注:检验结果中6、12、18等代表残差滞后阶数;Chi-Square代表检验统计量

图8 HW^{10} 序列 LB 检验结果
Fig. 8 Results of LB test for HW^{10} series

3.3 建模分析

鉴于所有序列均为平稳非白噪声序列,本研究采用自回归移动平均(autoregressive moving average, ARMA)这一经典模型,对所有序列进行建模^[7]。自回归移动平均模型ARMA(p, q)通常定义为

$$\begin{cases} x_t = \theta_0 + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \dots + \theta_p x_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} \\ \quad - \varphi_2 \varepsilon_{t-2} - \dots - \varphi_q \varepsilon_{t-q} \\ \theta_p \neq 0, \varphi_q \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (7)$$

式中, $x_t, x_{t-1}, \dots, x_{t-p}$ 为序列及其自回归滞后项; θ_0 为常数项, $\theta_1, \theta_2, \dots, \theta_p$ 是序列自回归滞后项系数; $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ 是残差及其滞后项; $\varphi_1, \varphi_2, \dots, \varphi_q$ 是残差滞后项系数; $E(\varepsilon_t)$ 与 $\text{Var}(\varepsilon_t)$ 分别代表残差的均值和方差; $E(\varepsilon_t \varepsilon_s) = 0$ 说明残差值互相独立, $E(x_s \varepsilon_t) = 0$ 说明残差值与序列值互相独立。

对已判定为平稳非白噪声的序列,通常的建模步骤如下(图9)。

- 1) 求出样本序列的自相关系数(ACF)和偏自相关系数(PACF)。
- 2) 根据ACF和PACF性质,选择合适阶数用

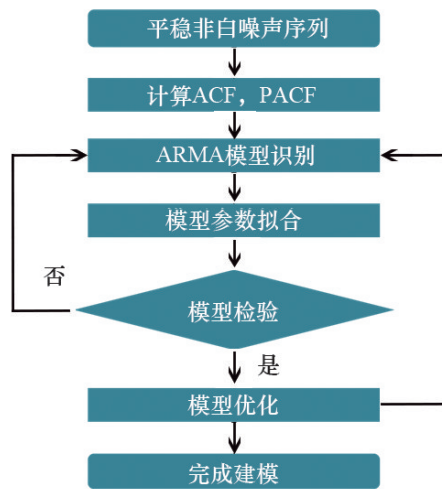


图9 平稳非白噪声序列的建模步骤

Fig. 9 Steps of modeling stationary & non-white noise series

ARMA(p, q)模型进行拟合。

- 3) 估计模型中未知参数的值。
- 4) 检验模型有效性。如果未通过,则返回步骤2,重新选择模型再拟合。
- 5) 模型优化。如果通过检验,仍然返回步骤2,充分考虑各种可能,建立多个拟合模型,从所有通过检验的拟合模型中选择最优模型。

针对所有序列执行上述建模过程。可得到最终的模型分别如下:

$$HW_t^{10} = 178.2652 + 0.98878HW_{t-1}^{10} + \varepsilon_t - 0.79723\varepsilon_{t-1}$$

$$HW_t^{14} = 325.3555 + 0.25707HW_{t-1}^{14} + 0.195HW_{t-2}^{14} + 0.23601HW_{t-3}^{14} + \varepsilon_t$$

$$HW_t^{17} = 224.4041 + 0.99998HW_{t-1}^{17} + \varepsilon_t - 0.85634\varepsilon_{t-1}$$

$$HH_t^{10} = 320.8186 + 0.34985HH_{t-1}^{10} + 0.207HH_{t-2}^{10} + \varepsilon_t$$

$$HH_t^{14} = 520.1308 + 0.5428HH_{t-1}^{14} + 0.41775HH_{t-4}^{14} + \varepsilon_t$$

$$HH_t^{17} = 394.8288 + 0.91211HH_{t-1}^{17} + \varepsilon_t - 0.5784\varepsilon_{t-1}$$

$$BW_t^{10} = 481.9447 + 0.90809BW_{t-1}^{10} + \varepsilon_t - 0.63028\varepsilon_{t-1}$$

$$BW_t^{14} = 811.5528 + 0.37228BW_{t-1}^{14} + 0.15552BW_{t-2}^{14} + 0.1777BW_{t-3}^{14} + \varepsilon_t$$

$$BW_t^{17} = 655.5046 + 0.99563BW_{t-1}^{17} + \varepsilon_t - 0.75106\varepsilon_{t-1}$$

$$BH_t^{10} = 764.4702 + 0.69706BH_{t-1}^{10} + \varepsilon_t$$

$$BH_t^{14} = 1132.679 + 0.28604BH_{t-1}^{14} + 0.29989BH_{t-4}^{14} + \varepsilon_t$$

$$BH_t^{17} = 997.0689 + 0.99889BH_{t-1}^{17} + \varepsilon_t - 0.72163\varepsilon_{t-1}$$

所有序列的模型形式整理后如表3所示。当取相同时间点和相同组别时,2个监测点的数据序列模型非常相似。其中,17:30时间点对应的4个序列甚至具有完全相同的模型形式ARMA(1,1),只是在估计参数方面存在差别。这说明不同监测点行人流量的纵向时间序列具有显著规律可以挖掘,同时也从侧面验证了本研究分组处理方式的合理性。

表3 所有序列模型形式

Table 3 Models for all selected series

序列	HW^{10}	HW^{14}	HW^{17}	HH^{10}	HH^{14}	HH^{17}
模型	ARMA(1,1)	AR(3)	ARMA(1,1)	AR(2)	AR(4)	ARMA(1,1)
序列	BW^{10}	BW^{14}	BW^{17}	BH^{10}	BH^{14}	BH^{17}
模型	ARMA(1,1)	AR(3)	ARMA(1,1)	AR(1)	AR(4)	ARMA(1,1)

注:当残差项不存在滞后项时,移动平均项消失,ARMA模型退化为自回归(AR)模型。

3.4 预测效果分析

考虑到预测步数(即滞后日期的个数)的增加对于时间序列模型的预测误差存在放大效应,因此本研究只额外采集2个监测点后续5个工作日

(2012年1月4—6日,1月9—10日)和5个节假日(2012年1月1—3日,1月7—8日)的数据简单进行1步预测和5步预测的效果对比分析。为确保可比性,在进行5步预测时新的真实数据不对模型进行

任何修正。可得到表4所示的结果。表中的原预测模型是指预警系统内置的横向时间序列移动平均模型(不含经验参数修正项),新预测模型使用3.3节建模后得到的公式。

从表4可知,无论是1步预测还是5步预测,新模型在误差优势计数(即准确率占优一方的累计个数)以及误差幅度方面,均比原模型具有显著的优势。这说明,来自于视频智能分析技术的商业街行人交通流纵向时间序列,经预处理后建模得到的预测模型在预测效果上具有很好的潜力。

表4 模型预测误差对比
Table 4 Error comparison of forecast models

序列	1步预测误差		5步预测平均误差	
	原预测模型	新预测模型	原预测模型	新预测模型
HW^{10}	0.0828	0.0084	0.0639	0.0119
HW^{14}	0.0586	0.0091	0.0444	0.0236
HW^{17}	0.1257	0.0346	0.1702	0.0428
HH^{10}	0.2264	0.0547	0.0969	0.1050
HH^{14}	0.0006	0.0226	0.0488	0.0430
HH^{17}	0.0342	0.1186	0.0973	0.1114
BW^{10}	0.1011	0.0196	0.2261	0.0048
BW^{14}	0.6978	0.0792	0.4838	0.0612
BW^{17}	0.4647	0.1672	0.2901	0.0279
BH^{10}	0.0329	0.0178	0.3273	0.1188
BH^{14}	0.2907	0.0102	0.2794	0.1145
BH^{17}	0.0253	0.0125	0.1120	0.1123
优势计数	2	10	3	9
平均误差	0.1784	0.0462	0.1867	0.0648

4 结论

基于视频智能分析技术的行人数据无法使用传统的行人交通流物理建模方法,也难以直接沿用基于轨道交通数据的各类离线分析技术实现高度集成化的预测预警。本研究面向视频智能分析技术,以西单商业区为例,构造了一整套数据筛选、预处理、检验、建模的方法过程。研究表明,采用分组式纵向时间序列建模能提供较好的预测效果。模型形式简单,易嵌入系统实现实时监测与预警,通过后续的研究,亦可进一步引入自学习机制,实现模型的适时更新与优化。

参考文献(References)

- [1] 唐明, 贾洪飞, 杨丽丽. 城市轨道交通车站站台行人交通特性[J]. 城市交通, 2010, 8(4): 47-51.
Tang Ming, Jia Hongfei, Yang Lili. Characteristics of riders' behavior at urban rail transit stations[J]. Urban Transport of China, 2010, 8(4): 47-51.
- [2] 金晓琼, 韩萍, 左忠义, 等. 大连市西安路商业区行人交通特性分析[J]. 大连交通大学学报, 2008, 29(2): 27-31.
Jin Xiaojiong, Han Ping, Zuo Zhongyi, et al. Pedestrian traffic characteristics of Xi'an commercial in Dalian[J]. Journal of Dalian Jiaotong University, 2008, 29(2): 27-31.
- [3] 王秋平, 张译, 孙皓. 历史街区行人交通特性研究[J]. 公路工程, 2016, 41(6): 53-58.
Wang Qiuping, Zhang Yi, Sun Hao. Research on the historical block pedestrian traffic characteristics[J]. Highway Engineering, 2016, 41(6): 53-58.
- [4] 叶建红, 陈小鸿. 行人交通流三参数基本关系式适用性研究[J]. 西南交通大学学报, 2016, 51(1): 138-144.
Ye Jianhong, Chen Xiaohong. Applicability analysis of tri-parametric fundamental equations for pedestrian traffic flow[J]. Journal of Southwest Jiaotong University, 2016, 51(1): 138-144.
- [5] 倪慧荟. 城市人群聚集风险预警及客流安全管理[J]. 现代职业安全, 2015, 165: 30-33.
Ni Huihui. Urban overcrowding forecast and pedestrian safety management[J]. Modern Occupational Safety, 2015, 165: 30-33.
- [6] 胡成. 人员密集场所预警系统研究[J]. 安全, 2009, 30(11): 57-58.
Hu Cheng. Research on the early warning system for assembly occupancies[J]. Safety, 2009, 30(11): 57-58.
- [7] Blue V J, Adler J L. cellular automata microsimulation for modeling bi-directional pedestrian walkways[J]. Transportation Research, Part B, 2001(35): 293-312.
- [8] Okazaki S. A study of pedestrian movement in architectural space, part 1: Pedestrian movement by the application of magnetic models[J]. Journal of Architecture and Planning, 1979, 283(3): 111-119.
- [9] Helbing D, Molnar P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51: 4282-4286.
- [10] 迟菲, 胡成, 李凤, 等. 密集人群流动规律与模拟技术[M]. 北京: 化学工业出版社, 2012.
Chi Fei, Hu Cheng, Li Feng, et al. Mass crowd flow and simulation technology[M]. Beijing: Chemical Industry

- Press, 2012.
- [11] 张亚运. 基于客流短时预测的城市轨道交通运营组织 [D]. 西安: 长安大学, 2016.
Zhang Yayun. Operation organization of urban rail transit based on the passenger flow short-term prediction [D]. Xi'an: Chang'an University, 2016.
- [12] 郭文, 肖为周, 秦菲菲. 基于支持向量机模型的地铁进站客流量预测[J]. 河北工业科技, 2019, 36(1): 31-35.
Guo Wen, Xiao Weizhou, Qin Feifei. Prediction of subway entry flow based on support vector machine model [J]. Hebei Journal of Industrial Science & Technology, 2019, 36(1): 31-35.
- [13] 付保明, 王健, 张宁, 等. 基于云模型的城市轨道交通短时客流预测[J]. 城市轨道交通研究, 2018, 21(4): 61-65.
Fu Baoming, Wang Jian, Zhang Ning, et al. Forecast of short-term passenger flow in urban rail transit based on cloud model[J]. Urban Mass Transit, 2018, 21(4): 61-65.
- [14] 北京市劳动保护科学研究所. 基于组合多步预测的行人交通拥堵预警方法及系统: ZL 2011 1 0109527.4 [P]. 2014-05-28.
Beijing Municipal Institute of Labour Protection. An overcrowding forecast method and system based on the combined multi-step prediction: ZL 2011 1 0109527.4 [P]. 2014-05-28.
- [15] 张朝雄, 沈旻, 张慧. 缺失数据插补方法比较研究[J]. 市场研究, 2007(9): 33-35.
Zhang Chaoxiong, Shen Min, Zhang Hui. Missing data's interpolation method[J]. Marketing Research, 2007(9): 33-35.
- [16] 谭克平. 极端值判断方法简介[J]. 台东大学教育学报, 2008, 19(1): 131-150.
Tam Hakping. Introduction to the identification of extreme value[J]. NTU Educational Research Journal, 2008, 19(1): 131-150.
- [17] 王燕. 应用时间序列分析[M]. 2版. 北京: 中国人民大学出版社, 2005.
Wang Yan. Applied time series analysis[M]. 2nd ed. Beijing: China Renmin University Press, 2005.

Forecast models for commercial street pedestrian traffic flow data based on intelligent video analysis

NI Huihui¹, WU Bohong^{2,3}

1. Research Laboratory of Safety and Emergency Management, Beijing Municipal Institute of Labour Protection, Beijing 100054, China
2. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China
3. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract The sampling, the preprocessing, and the modeling for the commercial street pedestrian traffic flow data based on the intelligent video analysis are presented in this paper. The Xidan mall is taken as an example, the by-date grouping-style vertical time series are established, consisting of multi surveillance points and different points-in-time. The modeling and forecast results show that all vertical time series are stationary and of non-white noise with similar ARMA expression formulas, which can be well applied to the forecast of pedestrian traffic flow data.

Keywords intelligent video analysis; commercial street pedestrian traffic flow; forecast model ●



(责任编辑 傅雪)