

智慧城市网格管理事件模式挖掘与预测

Urban Grid Management Incidents Pattern Mining and Prediction

吴俊 王杰艺 金耀辉 WU Jun, WANG Jieyi, JIN Yaohui

摘要 城市化进程的快速发展给城市管理带来了巨大的挑战,城市问题事件的管理和预警已成为城市可持续发展的一个重要组成部分。多变量时间序列分析预测模型RBTA可以挖掘城市管理事件发展的基本趋势、周期性、异常事件以及不同城市管理事件之间的耦合关系。使用来自于上海市徐汇区网格中心的真实城市管理事件数据及对模型进行评估,该模型的平均拟合均方根误差为0.12,平均预测均方根误差为0.15,预测准确率比现有方法中最好的方法高4.9%。

Abstract The rapid development of urbanization has brought great challenges to the management of cities. The management and early warning of urban incidents have become an important part of urban sustainable development. This paper proposes RBTA, a multivariate time-series model, to find the patterns including basic trend, seasonality, irregular components and relationship among different incidents. We evaluate our model on the real dataset from the downtown area of Shanghai, one of the biggest metropolitan of the world. The average forecasting root mean squared error (RMSE) is 0.15, which decreases 4.9% comparing to the best one of the existing methods.

关键词 智慧城市 | 时间序列 | 预测

Keywords Smart city | Time-series | Forecast

文章编号 1673-8985 (2018) 01-0051-06 中图分类号 TU981 文献标志码 A

作者简介

吴俊

上海信投建设有限公司

副总经理,高级工程师

王杰艺

上海交通大学电子信息与电气工程学院

硕士研究生

金耀辉

上海交通大学人工智能研究院

教授,博士生导师,博士

0 引言

工业革命以后,城市化的速度呈指数形式的增长,城市中每天都有数以千计的新住宅建成,全球范围内每周有数百万的人口迁入城市。快速城市化带来了经济发展的同时出现了许多问题,例如城市人口密度过高、交通堵塞、严重的环境污染等。

原有的人工管理方式已经无法覆盖飞速增加的城市面积与城市问题,需要引入更加智能化的方式对城市进行管理,智慧城市的概念应运而生。智慧城市是城市使用先进的信息与通信手段采集多种数据,通过对数据进行分析处理以对城市管理资源配置提供有效的支持^[1]。

智慧城市的主要应用领域集中在城市公共安全、城市基础设施、城市交通、能源及公共事

业、环境保护、智慧治理等方面。本文的研究内容属于智慧治理范畴。政府推进信息通信技术的构建,提高公民决策的参与程度,提高公共社会服务的效率,实现透明治理,更加有效的管理城市资源,这就是“智慧治理”^[2]。考虑到城市管理的重要性以及各利益相关方协作以达到理想的城市发展水平、经济增长率以及城市的宜居性,智慧治理是智慧城市研究发展的重要方向之一^[3]。

当前在智慧城市和智慧治理领域国内外已有了一系列的研究工作,网格化管理是城市管理的电子化平台,将整个城市分区按照一定的规则划分成网格,由专门的网格中心监管人员对所管辖范围内的网格进行巡查,当发现所辖区域内出现问题后上报网格中心。同时,网格

表1 时间序列分析算法比较

	一元时间序列模型				多元时间序列模型	
	DFT	Wavelet	SARIMA	TBATS	CompCube	RBTA
基本趋势				✓	✓	✓
季节性	✓	✓	✓	✓	✓	✓
异常	✓	✓		✓	✓	✓
预测			✓	✓	✓	✓
竞争					✓	✓
多重关系						✓

资料来源:作者自制。

中心也收集来自于市民的投诉或建议,并统一上报处置。网格化管理的核心功能主要包括:基础数据的采集与存储、各类数据的智能化汇总统计与分析、各组织机构绩效的考核与评比等。

城市各类问题事件都在不同程度上影响了人们的正常生活或财产安全,导致城市各类问题发生的影响因素可能众多而复杂。本文主要围绕时空行为信息学,充分利用互联网、大数据等创新技术,基于网格化管理系统中采集得到的各类投诉数据,重点挖掘城市中各类问题事件的发生规律、可能的影响因素,全面、精确、实时地掌握各类风险动态,同时基于挖掘结果建立预测与预警模型。一方面能够对各类事件在未来一段时间内最可能发生的时间进行预测,提前预防、控制可能发生的危险事故和突发事件。另一方面能够对各类事件发生和演化过程进行监控和预警,因此能够在危机事件发生后,做到有效的信息共享和协调联动。核心目标是基于城市网格化管理数据,旨在推进智慧城市与智慧治理建设过程中,从城市与社区管理角度出发,有效地掌握数据采集、分析、预测的方法,实现人与技术的充分融合,在危机发生前及发生后均能做出有效的决策或行动,使城市与社区的管理更智慧、更高效、更安全。

本文研究包含了数据采集、数据分析、事件预测与预警等整套挖掘过程。从网格化管理系统中采集相应数据,并对包含居民投诉建议及各类基础设施的基本信息建立多级数据仓库,对多源数据进行数据关联与融合;针对各类问题相关事件,结合基础设施基本信息等多源数据,详细分析导致每类事件发生的触发因素,以及导致各类事件严重程度增加的演化因素;基

于触发因素关联分析结果,建立多维度融合预测模型,对各类事件在未来一段时间的发展进行预测;基于演化因素关联分析结果,对各类事件在发生和演化过程进行监控,并在事件急速恶化之前及时预警。

1 国内外研究现状

1.1 智慧治理

巴布尔^[4]认为市政府对解决全球性问题至关重要,并指出“市长统治世界”。目前,对城市智慧治理的行政重视已经与学术关注相平行。城市治理已经发展成为一个相对成熟的学术领域^[5]。但是最近,电子政务和创新研究正在与它联系起来,制定可使城市更加智能的方法^[6]。吉尔^[7]指出,智慧治理是一种新型电子治理形式,其利用先进的信息技术互联和整合信息、流程、机构和物质基础设施,使政府可以更好地为公民和社区服务。

KORTUEM G等^[8]研究了智慧治理中公共服务以及社会管理方面的内容,在日常生活,对于群众的投诉、求助、社会事务等方面的处理可以镶嵌在全市智慧管理和市场运作的社会服务体系之中,并在此基础上为城市综合规划、应急响应、社区管理提供基础平台服务,实现政府一站式服务体系。在这种情况下,政府可以收集和分析城市各个地区的实时数据,为公众提供更快捷、更灵活的服务。公众可以通过手机、个人电脑等方式上传信息,实现对事件的实时查询。

1.2 时间序列模型

现有的大多数时间序列预测方法是不考虑

不同城市事件之间关系的单一模型。小波变换、傅立叶变换和AutoPlait^[9]可以检测季节性,但是它们不能预测未来的事件。AR、ARIMA++、PLif^[10]和Trimine^[11]能够挖掘时间序列的发展模式和预测未来的发展情况,但是由于它们使用线性模型,因此无法处理非线性情况。FUNNEL^[12]、TBATS^[13]可以捕获单变量时间序列方法的所有模式,但是它们不能捕捉不同时间序列事件之间的关系。EcoWeb^[14]是一种多元时间序列方法,但它只能处理竞争关系,而不能处理其他类型的关系。深度学习算法如LSTM等模型不具有可解释性,无法解释时间序列发展的内在原因,对实践应用的帮助不大。表1展现了RBTA模型与其他时间序列分析模型的对比结果。

2 案例:上海市徐汇区网格中心

上海市徐汇区网格中心打造了具有徐汇特色的城区综合治理体系,即1+13+X体系。1个区级平台:在网格化管理的基础上,融入12345市民服务热线、联动联动、应急处置等职能,拓展社会管理和公共服务事项,构建一体化运行的综合枢纽。13个街镇中心:通过重心下移、资源下沉和权力下放,让最靠近服务对象的管理主体真正有职有权有人有物,使街道镇成为城区管理和社会治理的重要平台。延伸X个工作站:306个居民区、5个园区、1个地下轨交设立工作站,做到管理空间全覆盖。徐汇区通过上述的体系标准,形成了一套行之有效的城市综合管理方法,配套信息化手段,提升了智慧城市的能力水平。在推进过程中发现,传统的信息化手段能够提高网格化工作的效率,起到了事中事后监管的作用,但是把握不准城市管理事件发展的基本趋势、周期性、异常事件和不同事件之间关系,仅能通过个人经验对一些问题进行预测,存在较大的错误率。

本文使用徐汇区网格中心的实际业务数据来评估我们提出的模型RBTA,本次使用的徐汇区网格中心业务数据跨度从2016年1月1日到2016年12月31日,事件记录93万条,事件记录数据包括发生的时间、事件所

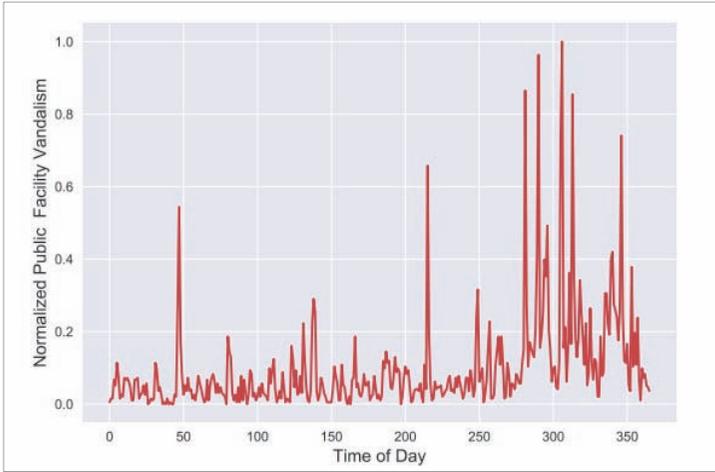


图1 公共设施损坏在全年内的发生次数
资料来源:作者自绘。

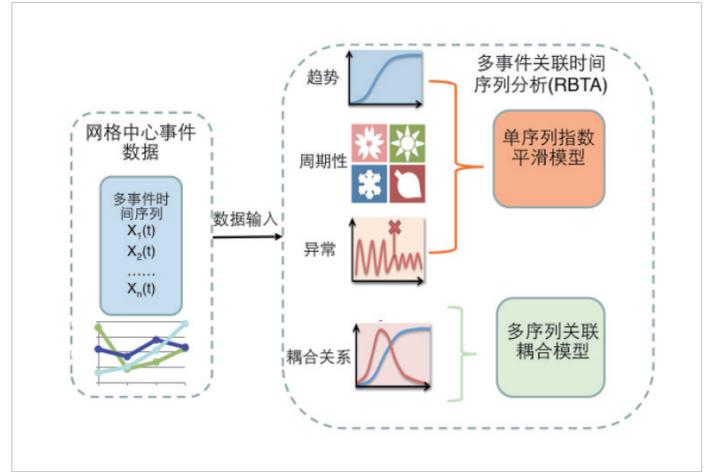


图2 RBTA时间序列分析模型框架
资料来源:作者自绘。

表2 城市问题事件的分类及其所占比例

问题类别	所占比例 (%)	问题类别	所占比例 (%)
市容市貌	69.8	设施管理	5.8
街道秩序	12.3	社区管理	1.0
投诉电话	10.5	紧急事件	0.6

资料来源:作者自制。

归属的类别、事件发生的地点(城建地理坐标、所在街道、所在社区)等。

图1描述了公共设施损坏这一问题事件在全年内的发生发展情况,横轴是以天为单位的时间变化,纵轴是公共设施损坏发生的归一化次数。

徐汇区的网格数据被人工分为6个一级分类和46个小类,表2描述了6个一级分类以及他们所占的比例。从中我们可以看出,市容市貌所占的比例最高,达到了50%以上,说明上海市徐汇区所受的最大困扰是市容市貌问题。

3 管理事件的模式挖掘及预测

3.1 模式挖掘

本文提出名为RBTA的多元时间序列模型,其框架如图2所示。RBTA是基于一元时间序列模型TBATS,增加了不同事件间相互影响的因子。该模型背后的直觉逻辑是,不同的问题事件并不是相互独立的,它们会

影响一些事件的发生,同时受到其他事件的影响。例如,垃圾问题和涂鸦问题经常同时出现在一些整洁程度较差的街道上。同时一些事件之间存在竞争性的关系,比如街边的违法设摊和非法停车,街边的位置有限,当违法设摊被清除后会出现非法停车的情况。如图3所示的事件间的竞争关系。

RBTA通过对来自于上海市徐汇区网格中心的实际数据进行评估,结果表明RBTA具有以下优点:

(1) 该算法可以很好地获取单个事件时间序列发展的全部规律,包括其发展的基本趋势,即共同演化的非线性动态、季节性和特异事件,以及不同事件间的相互影响。算法所提取的时间序列发展规律是符合人类理解和逻辑的,相对于其他深度学习算法如LSTM等算法具有很强的可解释性,强大的算法可解释性可以为业务部门带来对数据更深入的理解。算法的拟合均方根误差是0.12,比现存的其他算法的误差小。

(2) 该算法可用于时间序列的长期预测,结果准确度相对于其他算法有明显的提升。预测的均方误差为0.15,比现有的最好算法准确率提高4.9%。

本文模型RBTA的数学公式如下所示^[13]。

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^{(\omega)} - 1}{\omega} & \omega \neq 0 \\ \log(y_t) & \omega = 0 \end{cases}$$

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-1}^{(i)} + d_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t$$

$$s_{t-1}^{(i)} = \sum_{j=1}^{k_i} s_{j,t-1}^{(i)}$$

$$d_t = \sum_{i=1}^p \phi_i d_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t$$

$$y_t^i = \sum_{l=1}^n \sum_{j=1}^m c_{jil} \hat{y}_{t-1}^j$$

3.1.1 基本趋势

在图4中,观测数据(图中红色的线条)是原始数据经过Box-Cox变换后的结果,Box-Cox变换的主要作用是将非稳态数据标准化以方便处理。图中黑色的线条为RBTA模型从问题事件中所提取出的发展趋势,综合观测数据和基本趋势图,可以看出RBTA模型很好地捕获了问题事件协同演化发展中的长期和短期趋势,从图中可以看出在上半年问题事件出现小的波动但整体变化不剧烈,下半年问题事件发生的次数出现指数式增长。

3.1.2 季节性和特殊情况

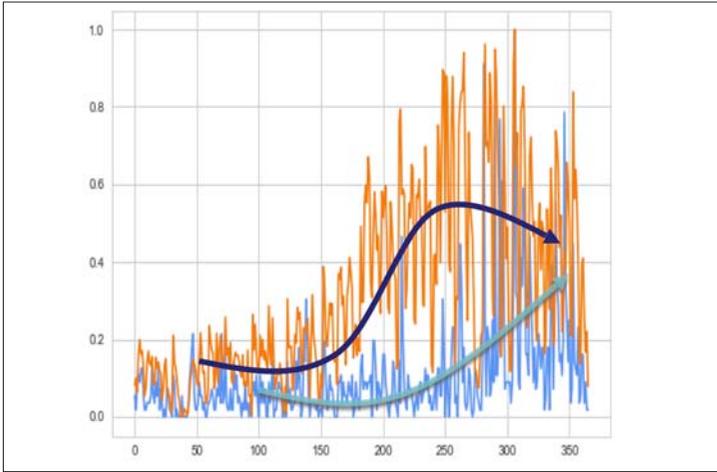


图3 事件间的相互影响
资料来源: 作者自绘。

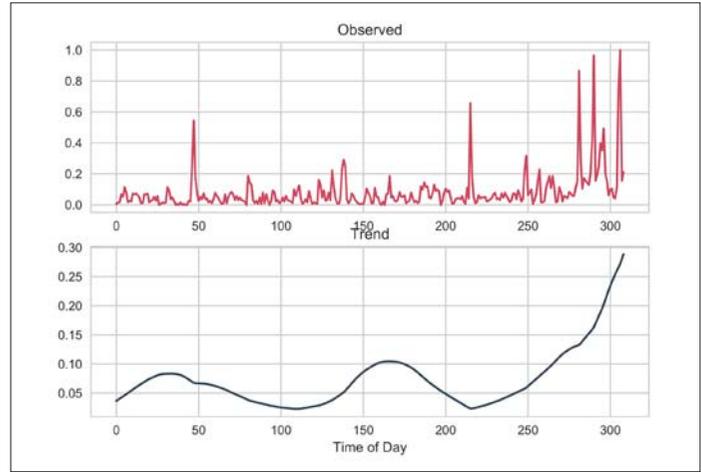


图4 问题事件的原始数据和基本趋势
资料来源: 作者自绘。

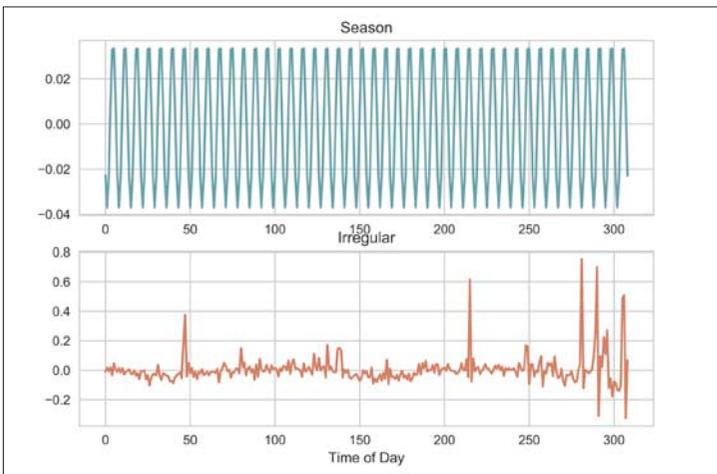


图5 问题事件的季节性和特异事件
资料来源: 作者自绘。

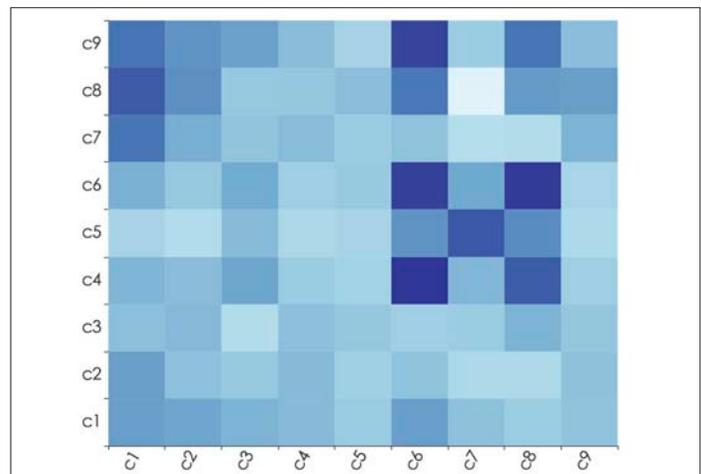


图6 不同事件间的相互关系
资料来源: 作者自绘。

图5中,绿色线条描述了RBTA模型从问题事件中所提取出的季节性特征,可以看出该季节性特征是以7天为周期,受工作日划分的影响较大,在工作日时问题事件的发生较多而非工作日时问题事件发生的数目较少。图中红色线条描述了问题事件发生的特异情况,相对于其他时间序列模型假设误差服从高斯白噪声分布, RBTA模型描述了噪声的ARMA过程。

3.1.3 事件间的耦合关系

图6描述了不同问题事件间的相互关系,由于问题事件的种类过多,为方便展示这里只选取了9类问题事件研究它们之间的相互影响,并设时间窗口 $\lambda=1$ 。图中的c1对应于环卫环

保类事件, c2对应于园林绿化类事件, c3对应于废弃车辆类事件, c4对应于暴露垃圾类事件, c5对应于道路保洁类事件, c6对应于乱设户外设施类事件, c7对应于墙面污损类事件, c8对应于跨门营业类事件, c9对应于乱晾晒类事件。

图6中横轴的事件是受影响事件,纵轴的事件是影响因素,颜色越深表示纵轴的事件对横轴事件的正向促进作用越大,即纵轴事件发生会很大程度上导致横轴事件发生,颜色越浅说明两个事件间的影响关系很小,当颜色浅至白色时说明纵轴事件对横轴事件的抑制作用越大,即纵轴事件的发生会抑制横轴事件的发生。从图中可以看出,颜色并不是呈对称分布的,这

说明在多数情况下事件间的相互影响并不是堆成的。事件a会促进事件b的发生,但事件b并不能促使事件a发生。图中对角线上的颜色也并不都呈现深色,这是因为一个事件的发展受其他事件影响的程度多于其自身过去的发展。这也是合理的,当今社会是一个十分复杂的结构,一个事件的发展受众多因素的影响,不同事件间的关系呈现复杂的分布。

另外,还可以看出c8对c1有促进作用,即跨门营业类事件会促使环卫环保类问题的出现,跨门营业是占用经营场所外的公共空间进行摆摊、放置货物的行为,街道上如果很多商家都跨门经营,会在公共空间产生很多垃圾,并且因为街道被商家占用很难及时清理,在一段时间后

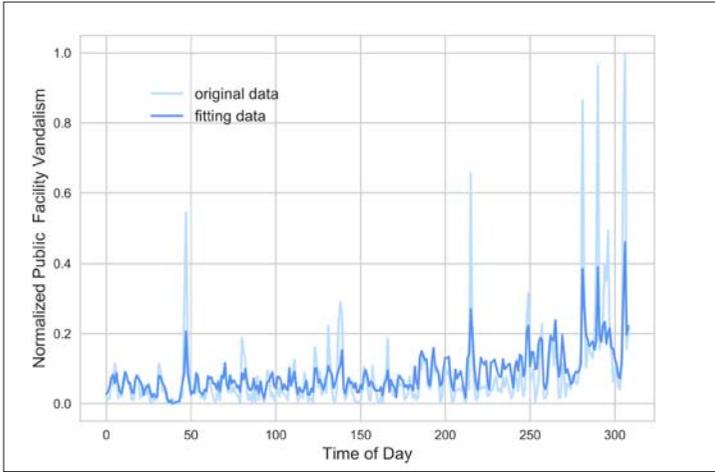


图7 RBTA模型拟合结果
资料来源:作者自绘。

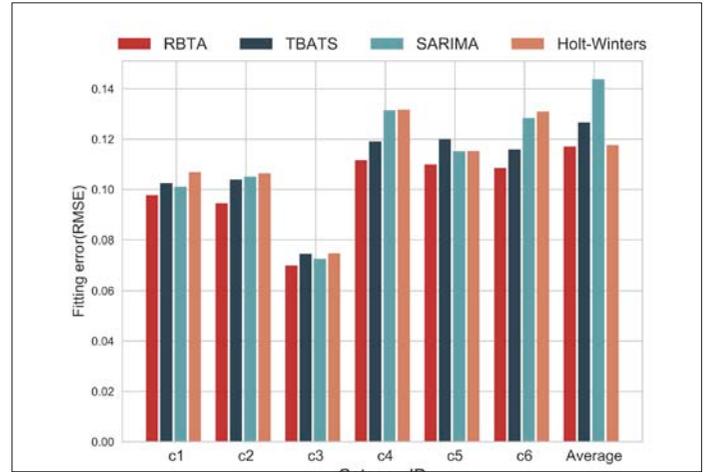


图8 拟合结果对比
资料来源:作者自绘。

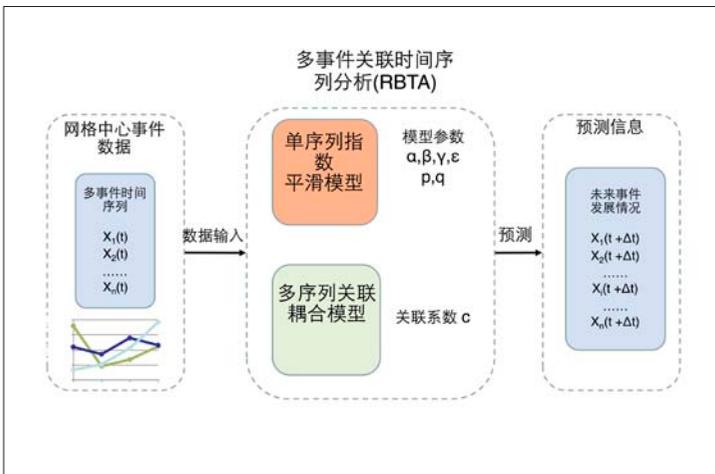


图9 RBTA预测模型
资料来源:作者自绘。

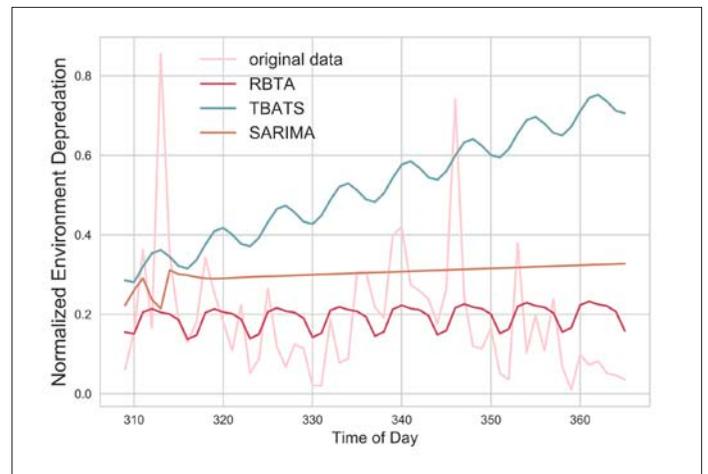


图10 RBTA拟合结果
资料来源:作者自绘。

会演化成严重的环卫环保类问题。c5对c7也有促进作用,即道路保洁问题会对墙面污损问题有促进,可以理解当城市某条街道的道路保洁出现问题,即该街道整体干净卫生情况较差,会继续催生街道墙面的污损情况。c3、c7、c8是受其他事件影响较大,但自身过去的发展影响较小的事件,对于这一类事件未来发展情况的预测需要投入更多的精力研究其他事件的发展情况,而调低其自身过去发展的权重。c8对c7的发生发展存在抑制作用,而c7的发生对c8的影响作用不大。

3.1.4 拟合

图7展示了RBTA模型的拟合结果,拟合的

均方误差 (RMSE) 是0.094,图中浅色的线条是原始数据,深色的线条是RBTA的拟合结果,可以看出RBTA模型很好的抓住了原始数据的基本趋势、季节性和特异情况,拟合的结果和原始数据相差很小。

为了证明RBTA算法的拟合结果强于现有的其他算法,我们将RBTA与其他算法进行比较,比较的算法包括TBATS算法、SARIMA算法以及Holt-Winters算法(图8)。对比试验的数据集在上海市徐汇区的网格数据中选取6个类型的数据作为展示,并在所有的共计46种类型的数据上进行拟合,将结果取平均值作为最终的结果。6个作为展示的问题事件类型分别

是c1公共设施损坏事件,c2道路交通问题事件,c3园林绿化问题事件,c4垃圾问题事件,c5道路保洁问题事件,c6墙面乱涂写问题事件。可以看出RBTA在这6类事件以及所有事件的均值结果上的均方拟合误差都最小,拟合效果比TBATS、SARIMA和Holt-Winters的结果都好。

3.2 预测

预测算法的框架如图9所示,在获取RBTA的各项参数后,我们可以对不同的城市管理事件未来的发展情况进行预测。

图10展示了问题事件发展的预测结果,其

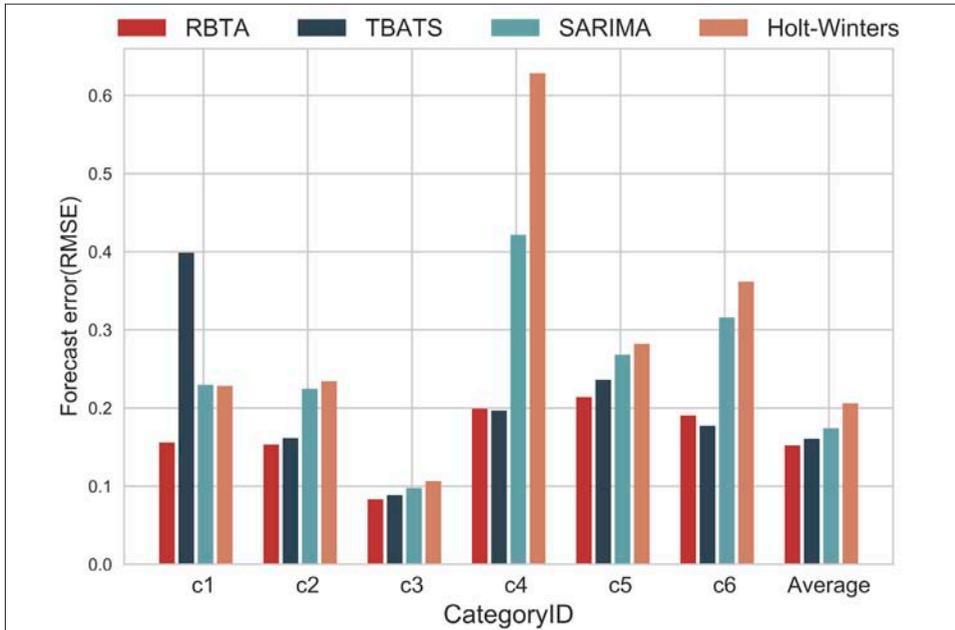


图11 预测结果对比
资料来源:作者自绘。

中淡红色的线条是真实数据,深红色的线条是RBTA模型的预测结果,绿色的线条是TBATS模型的预测结果,橙色的线条是SARIMA模型的预测结果。可以看出,相对于其他算法的预测结果,RBTA模型的预测结果与真实数据最接近,预测误差较小。且RBTA模型很好地捕获了问题事件未来发展的趋势、季节性以及特异情况,达到了很好的预测结果。TBATS模型在建模期间获取了单个事件的发展规律,但对之后的发展趋势预测有误,导致预测结果与真实数据相差较大。SARIMA算法对于这种波动性较大的非平稳序列,倾向于使用最近几次观测值的平均值作为预测值以消除噪声,尤其在预测时间较长时,SARIMA的预测误差会更差。

图11展现了不同时间序列预测模型对于上海市徐汇区网格中心的问题事件数据预测的均方误差,我们将RBTA与其他算法的预测结果进行比较,比较的算法包括TBATS算法、SARIMA算法以及Holt-Winters算法。与拟合实验类似,对比试验的数据集在上海市徐汇区的网格数据中选取6个类型的数据作为展示,并在所有的共计46种类型的数据上进行拟合,将结果取平均值作为最终的结果。6个作为展示的问题事件类型分别是c1公共设施损坏事件,

c2道路交通问题事件, c3园林绿化问题事件, c4垃圾问题事件, c5道路保洁问题事件, c6墙面乱涂写问题事件。可以看出RBTA在这6类事件以及所有事件的均值结果上的预测误差都最小,预测效果比TBATS、SARIMA和Holt-Winters的结果都好。

4 结语

本文研究了时间序列的模式挖掘与预测问题,创新性的提出了一个新的时间序列挖掘与预测的模型RBTA,该模型是在TBATS模型的基础上引入不同时间序列相互影响的因素。RBTA具有很强的可解释性,可以完整地获取时间序列发展的基本趋势、季节性、特异情况以及不同时间序列间的相互关系,并根据这些获取的模式进行准确的拟合和预测,准确度优于现存的时间序列分析模型。本文用上海市徐汇区网格中心的真实业务数据对我们提出的模型RBTA进行评估,拟合与预测的误差都小于现存的其他算法,预测的均方误差为0.15,比现存最好的算法准确率提高了4.9%。

参考文献 References

- [1] HAMBLE M. Just what is a smart city?[EB/OL]. (2015-10-01) [2018-03-06]. <https://www.computerworld.com/article/2986403/internet-of-things/just-what-is-a-smart-city.html>.
- [2] GIFFINGER R, FERTNER C, KRAMAR H, et al. City-ranking of European medium-sized cities[R]. 2014.
- [3] DEAKIN M. Smart cities: governing, modeling, and analyzing the transition[M]. Oxford: Routledge, 2013.
- [4] BARBER B R. If mayors ruled the world: dysfunctional nations, rising cities[M]. New Haven: Yale University Press, 2013: 186.
- [5] PIERRE J. Models of urban governance: the institutional dimension of urban politics[J]. Urban Affairs Review, 1999, 34(3): 372-396.
- [6] NAM T, PARDO T A. Smart city as urban innovation: focusing on management, policy, and context[C]// Proceedings of the International Conference on Theory and Practice of Electronic Governance. Tallinn, Estonia, September, 2011: 185-194.
- [7] GIL-GARCIA J R. Enacting electronic government success: an integrative study of government-wide websites, organizational capabilities, and institutions[J]. Integrated, 2012, 22(21): 3144-3150.
- [8] KORTUEM G, KAWSAR F, SUNDRAMOORTHY V, et al. Smart objects as building blocks for the Internet of things[J]. IEEE Internet Computing, 2009, 14(1): 44-51.
- [9] MATSUBARA Y, SAKURAI Y, FALOUTSOS C. AutoPlait: automatic mining of co-evolving time sequences[C]//ACM SIGMOD International Conference on Management of Data. 2014: 193-204.
- [10] LI L, PRAKASH B A, FALOUTSOS C. Parsimonious linear fingerprinting for time series[J]. Proceedings of the VLdb Endowment, 2010, 3(1-2): 385-396.
- [11] MATSUBARA Y, SAKURAI Y, FALOUTSOS C, et al. Fast mining and forecasting of complex time-stamped events[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2012: 271-279.
- [12] MATSUBARA Y, SAKURAI Y, PANHUIS W G V, et al. FUNNEL: automatic mining of spatially coevolving epidemics[C]//Proceedings of the 20th ACM SIGKDD Conference, New York City, August 24-27, 2014: 105-114.
- [13] LIVERA A M D, HYNDMAN R J, SNYDER R D. Forecasting time series with complex seasonal patterns using exponential smoothing[J]. Monash Econometrics & Business Statistics Working Papers, 2011, 106(496): 1513-1527.
- [14] MATSUBARA Y, SAKURAI Y, FALOUTSOS C. The web as a jungle: non-linear dynamical systems for co-evolving online activities[J]. 2015, 24(3): 721-731.