

基于大数据和机器学习的 5G 通信基站流量的分析与长短期预测建模

刘天然 李明臻 宋睿萌

(暨南大学经济学院, 广东 广州 515400)

摘要: 近年来, 随着移动通信技术的发展, 4G、5G给人们带来了极大便利, 产生的移动流量也呈现爆炸式增长, 仅2021年春节期间的移动数据流量就高达357.3万TB, 同比增长23.4%。因此, 对基站流量数据的研究变得越来越重要。另一方面, 在大数据的背景下, 机器学习、数据挖掘等技术的发展, 能够帮助5G流量解决高维的计算和优化问题。本文将在5G移动通信单基站上下行流量的大数据条件下, 分析流量数据周期性波动特征。并结合实际应用需求, 对上下行流量进行长、短期两个维度的预测建模。最后根据分析和预测的结果, 提出利用Hadoop分布式数据库为底座设计单个通信基站载频的可视化和自动开关机制的设想, 为5G基站的建设部署和基站节能机制的优化提出参考依据。

关键词: 5G流量预测; 时间序列分析; LSTM神经网络; 数据挖掘

本研究将主要从数据角度出发, 尽可能多地挖掘数据时空信息, 并将流量预测分为长期和短期。短期预测能更好地捕捉每日流量的高、低峰值和周期性变动; 长期预测更适用于研究流量最大值趋势的变化, 以便更早地进行规划和设计建设新的基站。让“大数据+通信工程”得以强强联手, 探讨5G未来的发展方向。

1 短期预测模型的建立

1.1 Holt-Winters指数平滑模型的建立与预测

鉴于数据量非常庞大, 故选择了其中编号#186基站作为例子进行建模说明。经检验, 训练集中无缺失值和异常值, 数据集有4月14日一天的缺失值。

通过对比三个指数模型的RMSE等指标, 发现三指数模型的RMSE最小。考虑优先使用三指数模型进行预测^[1]。此时, 模型可表示为:

$$Y_t = level + slope * t + S_t + irregular_t \quad (1)$$

其中, S_t 代表时刻的季节效应。

采用Holt-Winters指数平滑模型进行拟合, 并利用测试集判断拟合效果。结果给出了时序值、预测值以及80%和95%的置信区间(见图1)。

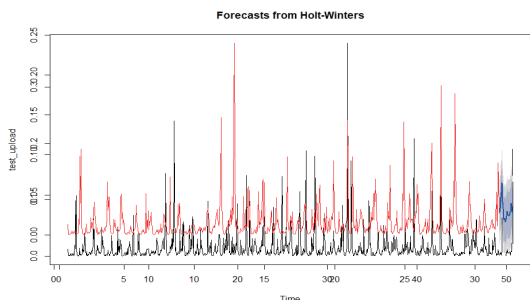


图1 Holt-Winters模型对#186小区上行流量的预测

1.2 SARIMs模型的建立和预测

1.2.1 ADF单位根检验和序列白噪声

这里使用ADF单位根检验来判别序列平稳性。根据结果, 检验的p-value小于显著性水平0.05, 可以认为两个序列均为平稳非白噪声序列, 不需要差分处理。故 $d=D=0$ 。同时对数列进行了白噪声检验, p-value小于0.05, 序列是非白噪声序列。

1.2.2 模型定阶

从上一步对时序平稳性的分析中可以确定, 非季节差分

阶数 $d=0$, 季节差分阶数 $D=1$ 和季节周期 $S=24$ 。取 $p=1, q=2$ 。由于模型的季节部分难以从图中直接识别, 故 P 和 Q 可取0、1、2, 从低阶到高阶搭配, 逐个建立备选模型, 并计算模型的BIC值^[2]。依照BIC最小准则选择最优模型为SARIMA(1,0,2) $\times(1,0,1)_{24}$, 此时的BIC值最小, 为-615.25。

1.2.3 诊断检验

结果显示两个模型的白噪声检验均通过, 表明序列信息已被充分提取, 模型残差为白噪声, 模型合理有效, 可用于流量数据的预测。

1.2.4 模型预测

采用SRIMA模型进行拟合, 并利用测试集数据判断拟合效果。结果给出了时序值、预测值以及80%和95%的置信区间, 结果见图2:

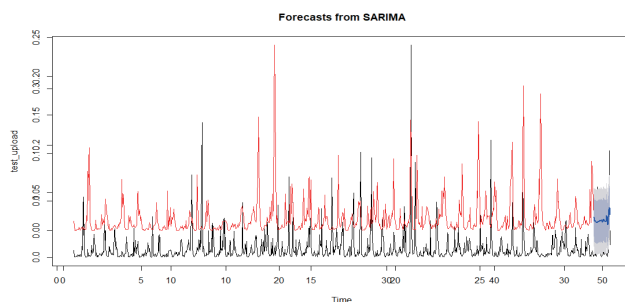


图2 SARIMA模型对#186小区的下行流量预测

2 长期预测模型的建立

2.1 数据预处理

本文选择数据集中前31天(2018/3/1-2018/3/31)为训练集, 后19天的数据为测试集进行预测。针对空缺字符采用临近平均插值方法完成数据填补后, 将数据标准化处理。

2.2 定义参数

数据集只有一个变量, 故采用1个输入层神经元, 1个输出层神经元, 2层LSTM循环神经网络模型。以(RMSE)和(MAPE)作为损失误差函数, 内部激励函数为sigmoid函数, 返回[0,1]之间的数, 学习速率选取区间[0.01,0.20]的中值0.1, 批处理数量为1。并采用Adam一阶优化算法, 来更新迭代过程中的权重。本文的一维5G流量时间序列数据, 数据维度体量相对较小, 为预防过拟合现象发生, 需要在隐藏层输出之后添加

(下转第214页)

像的依赖性越来越强,这是人们逐渐缺乏想象力和创造力而导致的问题。而VR与AR技术的出现则让图像艺术逐渐向视觉艺术转变,具体表现为图像观感的整体升级,从形式上来说,图像往往是二维的,具有静态效果,但VR与AR使图像能够联系现实,在现实中投放出立体图像,并给予其动态效果使静态图像变为动态图像,让人能够全方位对对象进行观察,从而大大提高图像的视觉效果,增强其视觉艺术性。数字新媒体的发展使人们在创作渠道上更加自由,使艺术具有整体性、完整性和创造性的特点。它的发展不是对现实环境的模仿,使人们能够从不同的角度直观地体验,拓宽了人们的体验感受。通过虚拟意境的运用,人们重构幻想世界将更加简单,创造内容的重构度也更加贴近现实。数字化不仅体现在技术的传播上,也体现在新的生存方式和思维方式上,数字媒体艺术的发展使人们的生活变得更加丰富多彩,为人们带来了更好的感官效果,同时也使人们的生活更加方便,能够随时随地获取信息,观赏艺术作品,大大促进了视觉艺术的发展。

3.3 技术推动跨界合作发展

视觉世界不断扩大,视觉体验更加丰富多彩,在这种情况下,艺术也发生了变化,开放性更加明显。艺术所创造的景观体验在各类前卫的设计影响下,全面地体现在公共空间中。相

关研究设计工作人员在艺术创新的过程中进行植入、转化途径等操作,可全面将现代媒体墙、广告商位等内容进行有效的结合。数字媒体艺术在建筑设计中渗透的具体表现,也是跨界合作的产物,可以展现技术发展的具体情况^[4]。

4 结语

随着数字时代的到来,VR和AR技术的发展取得了长足的进步,数字媒体使人们的日常生活更加舒适高效。数字时代下的VR和AR技术改变了传统的艺术思想和观念。数字媒体艺术使艺术设计进行了全方面的创新,为数字媒体产业的发展和数字媒体质量的可持续发展开辟了广阔的前景,创造了更多经济效益和社会效益,促进了技术进步。

参考文献

- [1] 谭文才,金笑语.VR技术在工业产品研发评审中的应用探索[J]. 新型工业化,2020,10(5):107-109.
- [2] 孟庆波.VR与AR对数字媒体艺术的影响探索[J]. 明日风尚,2018(7):78-79.
- [3] 孙熙茜.数字媒体艺术设计专业VR/AR教学的设计与研究[J]. 科教导刊:电子版,2018(7):105.
- [4] 张继江.基于VR技术的数字媒体艺术专业课程建设与实践[J]. 艺术科技,2018,31(12):235-236.

(上接第210页)

DropOut函数。

2.3 模型的训练

设置网络的训练步数为10,预测步长为一小时开始训练模型。先基于RMSE指标调整隐藏层神经元的个数^[3]。

从结果中可以看到,第一层和第二层神经元分别为200和300时,误差总和最小。因此,最佳隐含层神经元个数为,网络的最佳结构为1-200-300-1。

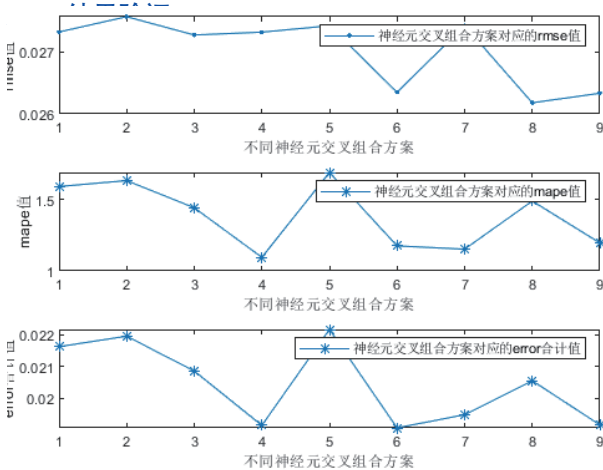


图3 交叉验证结果

3 论文总结

本文利用2018年3月1日至2018年4月19日的各基站(扇区)的小时级上下行流量数据进行分析。通过每周内同一天的纵向对比,发现每周的同一天基本呈现相似的波动规律,说明流量的预测研究有很大价值。长期和短期的预测模型结果良好。SARIMA模型更适合用于5G基站流量的短期预测,长期预测模型选择了LSTM神经网络变体进行预测,训练结果更好。

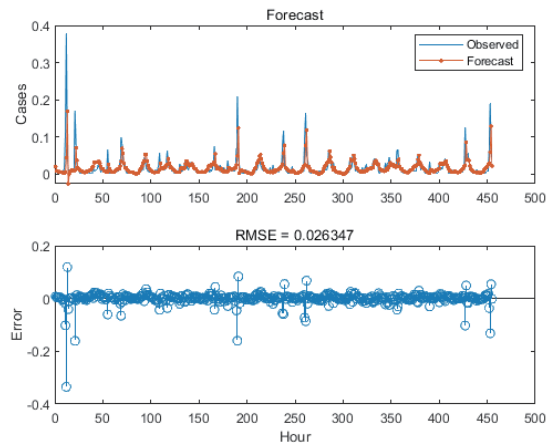


图4 模型预测情况与结果

4 结语

可以搜集更多的支撑数据,作为流量预测建模的补充变量,提高模型的精度和实际效用。在更多其他支撑数据下,时间序列建模可以使用更多的现代方法,考虑多方方程的建模。本次LSTM长短记忆神经网络的使用中,对损失函数的选择不同,隐含层神经元的个数也不同,收敛的速度和模型精度也不同。这里选择了RMSE和MAPE两个系数之比作为损失函数,还可以探讨不同的系数比,对隐含层神经元数,以及预测结果的影响。

参考文献

- [1] 杨欢.数据挖掘技术的分析与应用探析[J]. 电子元器件与信息技术,2021,5(6):71-72.
- [2] 周凌翔,车金庆.改进BP神经网络的研究及应用[J]. 价值工程,2012,31(34):209-210.
- [3] 赵庆.面向短时交通流分析与预测的循环神经网络算法研究[D]. 西安理工大学,2019.