

基于时间序列的流量预测算法综述

杨国丽

(北方工业大学, 北京 100043)

摘要:随着我国城市化进程的加快,越来越多问题的出现对我国城市的市政系统有了更高的要求,为给相关单位的管控工作与规划提供可靠的依据,提高基于时间序列数据的流量预测的准确性是目前研究的重中之重。本文介绍了基于时间序列流量预测的传统方法以及它们的发展趋势,在此基础上指出了基于LSTM与传统神经网络的组合模型,组合模型作为一组相对比较新颖的模型,通过组合多个特点模型以提高,要比单一预测模型有着更高的预测精度与效率,并为未来研究方向指明了方向。

关键词: 流量预测; 时间序列; 长短期记忆神经网络

中图分类号: TP301

文献标志码: A

DOI: 10.19772/j.cnki.2096-4455.2021.3.073

本文引用格式: 杨国丽. 基于时间序列的流量预测算法综述 [J]. 电子元器件与信息技术, 2021, 5(3): 165-167.

0 引言

随着我国经济的不断发展,我们的生活也变得越来越便捷,生活水平的显著提高,使得对城镇的管道系统的要求逐渐加大^[1];同时交通事业的迅猛发展,使得交通道路的堵塞情况愈加严峻,造成通行率低下、交通安全事故频频发生的问题;并且计算机网络技术的不断发展也促进了各类应用的产生,这对网络的性能及服务质量有了更高的要求等,于是各种问题也渐渐凸显了出来。

1 研究背景及意义

不论是管道流量、交通流量还是网络流量,全面准确地流量分析与预测,对于之后的优化及其服务质量抑或是管控工作都是极其重要的,是实现服务质量保证的一种重要基础技术,也是发挥大数据辅助决策功能的重要基础,所以我们进行实时、准确的流量预测研究是具有非常重要的现实意义。为给排水管道、交通路线优化,抑或是网络基站站址的规划提供依据,我们对传统的时间序列预测进行了介绍和对比,基于深度学习的建模的准确性是流量预测的关键。

不论哪种流量形式我们都可从时间维度来考虑将其看作时间序列问题相关的预测,而不管交通流量、管道流量还是网络流量,这些流量的变化形式都是一个实时、非平稳的一个随机的过程,并且都具有复杂性和不稳定性等时间序列数据的基本特征,所以针对序列数据的特征如何提高流量预测精度一直是众多学者一直关注并在研究的问题。

Kalpakis等人提出了运用自回归移动平均模型求和来表示时间序列,并为这个表示方法定义了一个相似性度量方式^[2];王婷婷等人提出了利用灰色马尔科夫模型对机场的进出流量进行了预测^[3];R.Vinayakumar, K.P.Soman and P. Poornachandran提出了应用LSTM长短期记忆网络模型对骨干网流量去进行预测^[4];覃明贵是提出了将神经网络与交通流量序列分割组合起来的一种组合预测模型,及时

准确的预测出道路交通状态^[5];芮兰兰,李钦铭设计了基于极限学习机和交通流量序列分割的一种组合模型^[6],他们是先用K-means对交通流量数据序列在时间上进行一个序列分割,其次利用ELM预测了短时交通流量序列进行;总之,现阶段时间序列数据它正以一种难以估计的速度产生于生活的每个领域,国内外学者关于时间序列数据的预测方法一直不断进步与更新,随着社会与工业的不断发展,时间序列数据在不同环境,不同状况下存在很大的差异,且大数据时代的急剧增长,数据规模的越来越大,这使得我们对时间序列数据的预测精度有了更高的要求,我们需要建立一套准确的模型。

2 时间序列

我们传统的流量预测方式主要都是将流量数据整形成一个流量的时间序列,即将流量预测问题规划成基于时间序列的预测问题。时间序列预测主要是根据已有的时间序列数据预测该数据未来某个时间段的变化趋势,通过对拥有的时间序列数据进行分析以及特征的提取,提取出来其中的变化特征,建立一个合适的数学模型,并将其运用到我们要预测的数据中去,由此规律来预测未来的数据可能会发生的变化趋势^[7]。

针对流量预测问题,已有国内外学者展开了深入的研究,且取得不错的工作成果,针对流量预测这类时间序列的预测问题,我们传统的预测方法大致有两种^[8],一种是经典的线性回归模型,比如差分自回归移动平均模型,另一种是非线性预测模型,主要是基于机器学习的预测模型,比如神经网络,梯度提升回归树,支持向量机等,且众多实验显示线性预测方法的预测效果远没有非线性预测方法好。

2.1 经典线性流量预测模型

时间序列就是按照时间的顺序记录的一系列有序的数据,这些有序数又可以分为平稳不平稳,基于时间序列的线性模型有比较多的种类,例如自回归滑动平均模型

基金项目: 基于物联网的云端市政排水管网自动清洁系统原型设计(项目编号: NLX0211)

ARMA、移动平均模型(MA)、自回归模型AR等平稳序列模型以及非平稳序列ARIMA模型等^[9-10]。

线性预测模型的理论基础是对于平稳非白噪声序列,我们通常会构建一个线性序列模型,接着要去拟合它的发展,然后以此来提炼并获取到序列数据中蕴含的有用信息。在针对平稳序列 $\{x_t\}$ 进行预测时,其自回归移动平均(ARMA)模型可表示为:

$$x_t = \sum_{j=1}^{\infty} \phi_j x_{t-j} + \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j} \quad (1)$$

上式中 $\sum_{j=1}^{\infty} \phi_j x_{t-j}$ 为序列 $\{x\}$ 的确定部分,即时间序列在当前时刻 t 的值可以用之前时刻自回归(AR)表示,如果添加与以前 t 时刻无关的随机干扰项,那么它就是AR模型; $\sum_{j=0}^{\infty} \phi_j \epsilon_{t-j}$ 为序列的随机性部分,用随机误差项的移动平均(MA)表示。

我们在运用ARMA模型进行预测的前提是序列要求是稳定的,并且序列的均值和方差不会随着时间变化。在许多实际问题中,观察到的时间序列不是稳定的序列,不过可以通过以下方式进行转换:对不平稳的序列进行差分处理,就能转化为相对稳定的序列。我们常使用的方法是先差分在进行ARMA建模,这就是ARIMA模型^[11]。

但此种预测模型通常在短时间序列数据中计算比较简单,随着我们需要预测的数据量逐渐增大,数据也愈加复杂,数据中时间序列的非线性特征更加明显,由于流量的复杂性和不确定性,传统方法的预测精度远远不够。

2.2 基于机器学习预测模型

机器学习其实就是人们让机器去解析数据并找到规律,然后通过它找到的规律去对新的序列数据进行处理,从中提取到某种有用的信息和规律,并利用学习到的规律对未知的数据进行预测。作为一种数据驱动算法,它具有高度拟合以及非线性估计能力的机器学习模型正成为流量预测的主流技术,比如循环神经网络(RNN)、人工神经网络(ANN)、支持向量机(SVR)等,其中神经网络是一种大的非线性机器学习模型。在神经网络中,循环神经网络(RNN)的最主要特点就是在时间维度上建模,每个输入变量都对应一个时间步和多个特征,它是通过在网络中构建循环来对数据的时间维度建模,这样一来它在时间序列数据中就可以展现出来更好的适应性^[12]。

3 LSTM原理及特点

长短期记忆模型(LSTM)它是一种特殊的RNN模型,它主要是为了解决RNN模型梯度消失这一问题而提出的^[13],其主要特点是它的时间记忆单元中具有多层隐藏层,通过门机制来控制信息传递的累积速度,且可以有选择地从前一时间状态中去遗忘或加入新的信息,可以有效地改进原始RNN存在的长依赖问题,以此来增强数据拟合的效果。

它是在RNN的隐藏层中添一个长短期记忆单元,循环结构中有着一组记忆单元与三种门结构来对历史信息进行添加或删除,这三个门分别称为遗忘门、输出门和输入门,LSTM神经网络如图1所示:

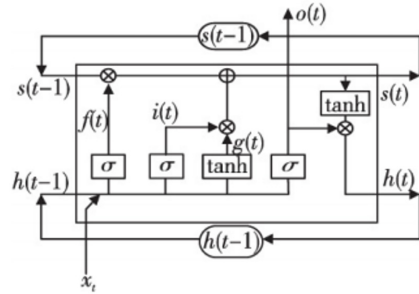


图1 LSTM结构图

在上图所示结构里, $f(t)$ 是遗忘门的输出信息遗忘门功能是来决定在上一时间状态中传递过来的信息有哪些信息是可以被丢弃掉的。输入门是决定添加哪些信息,它有两部分组成:一部分是使用了sigmoid函数,它的输出为 $i(t)$;而另一部分是使用tanh激活函数得到备选信息 $g(t)$ 。计算公式如下:

$$f(t) = \sigma(W_f h(t-1) + U_f x_t + b_f)$$

$$g(t) = \tanh(W_g h(t-1) + U_g x_t + b_g)$$

下面再对 $s(t-1)$ 更新,就构成了新的长期状态信息 $s(t)$,它是通过遗忘门来选择忘记从上一步传来的旧信息,通过输入门选择更新添加一部分新的信息获取到的。 $S(t)$ 的计算公式为:

$$s(t) = S(t-1) \odot f(t) + i(t) \odot g(t)$$

输出门的作用在于有它控制的记忆有多少可以用于下一状态的信息更新中 $o(t)$ 和将要传递给下一步的短期隐藏状态信息 $h(t)$,当前神经元 $o(t)$ 的计算公式为:

$$o(t) = \sigma(W_o h(t-1) + U_o x_t + b_o)$$

最终输出结果则要利用上一时间状态的短期状态值,长期状态值和当前输入值的保留部分来综合计算得出。公式如下:

$$h(t) = o(t) \odot \tanh(s(t))$$

其中 W_f, U_f, b_f 遗忘门隐藏层的神经元权重和偏置量, $W_i, U_i, b_i, W_g, U_g, b_g$ 为输入门的隐藏层权重和偏置量, W_o, U_o, b_o 为输出层的隐藏层权重和偏置量,用 \odot 表示Hadamard乘积,在一定程度上LSTM的门机制可以过滤掉一些与常规信息不同的内容,由此我们就可以消除掉训练数据中那些随机的偶然信息对测试数据预测精度的影响。

4 基于LSTM的组合预测模型

由于流量的复杂和不规律的曲线形态以及其突发性,采用单一模型进行预测的效果不会很理想,目前能够对流量进行一个好的预测的算法大都是基于深度学习与传统

神经网络的算法的混合模型，能够更加有效的利用各种模型的优点^[14]。

黄国伦^[15]等人新提出一种基于支持向量回归（SVR）和长短期记忆（LSTM）的新型组合预测模型，这一模型主要针对网络流量数据的随机性和波动性特点，来提高流量预测的精度。这一模型的预测过程主要是，它首先通过变分模态分解（VMD）对流量序列的固有模态进行提取，以此来减少噪声的随机性影响，再利用 LSTM模型对各本征模态进行拟合，学习到流量数据在时间上的分布规律，最后再分析它在流量方面的影响因素，利用 SVR 去拟合残差余量，让所有子预测模型进行一个整合叠加，如此一来该模型就可以有效将模型预测精度提高。

针对流量的突发性以及随机性特征，黄林、王电钢等人提出了一种基于LSTM的并行LSTM预测器。他们主要针对的是在具有强突发性特征的短期流量预测中预测性能明显下降的问题，他们让其中一个作为主预测器，另一个用于检测流量数据变化中的突发时刻，然后让这两个预测器进行内部状态的一个交换，由此来对预测精度进行一个改进。现阶段在时间序列预测问题中，有越来越多的学者使用了LSTM与RNN或其他传统神经网络混合起来的模型，比如江务学针对网络流量进行预测提出了改进的双线性化递归神经网络模型方法^[16]。总而言之，多个实验表明组合模型预测作为一组相对比较新颖的模型，它要比单个模型的预测效果要更好，组合模型预测方法也将是未来提高流量预测精度的趋势。

5 总结

随着我们对流量预测工作的研究，本文阐述并对比了经典的线性流量预测方法与基于机器学习算法的非线性预测模型算法。时间序列与方法虽然经历了长时间的发展，但数据规模的急剧增长，给传统的预测方法带来了极大的挑战，流量序列的复杂性和不确定性以及随机突发性等基本特征，传统的时间序列预测方法的预测精度远远不够，近年来的研究成果表明，基于LSTM的预测模型与传统神经网络的组合模型较于单一预测模型有着更理想的预测效果，组合模型的预测已然成为新的预测趋势，也为进

一步的研究指明了方向，其势必将得到越来越广泛的研究与应用，也需要研究者进一步努力。

参考文献

- [1] 宋磊,杜龙刚,黄涛,等.基于管道流量监测数据的城市雨洪模型实用性研究[J].水文,2019,39(03):22-28.
- [2] K. Kalpakis, D. Gada and V. Puttagunta, "Distance measures for effective clustering of ARIMA time-series," Proceedings 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 2001, pp. 273-280, doi: 10.1109/ICDM.2001.989529.
- [3] 王婷婷,王紫晨.基于灰色马尔科夫对龙洞堡机场旅客吞吐量预测[J].交通科技与经济,2017,19(05):48-51.
- [4] R. Vinayakumar, K. P. Soman and P. Poornachandran, "Applying deep learning approaches for network traffic prediction," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 2353-2358, doi: 10.1109/ICACCI.2017.8126198.
- [5] 覃明贵.城市道路交通数据挖掘研究与应用[D].复旦大学,2010.
- [6] 芮兰兰,李钦铭.基于组合模型的短时交通流量预测算法[J].电子与信息学报,2016,38(05):1227-1233.
- [7] 黄林,王电钢,刘萧,等.基于LSTM的网络流量预测方法[J].计算机应用研究,2020,37(S1):264-265,272.
- [8] 顾建强,刘海星,张朝,等.基于自适应卡尔曼滤波的用水量预测和爆管诊断[J].中国给水排水,2019,35(07):62-67.
- [9] 王海宁,袁祥枫,杨明川.基于LSTM与传统神经网络的网络流量预测及应用[J].移动通信,2019,43(08):37-44.
- [10] 汪志勇,邱晓红.基于多重分数差分AR模型的网络流量预测[J].计算机技术与发展,2009,19(03):84-86.
- [11] 张冉,赵成龙.ARIMA模型在网络流量预测中的应用研究[J].计算机仿真,2011,28(02):171-174.
- [12] 崔毓伟,卜世衍.高速公路收费站通行流量预测方法[J].上海船舶运输科学研究所学报,2020,43(02):63-67.
- [13] 郑友康,王红蕾.基于深度学习的短时道路交通流量预测研究[J].软件,2020,41(05):72-74.
- [14] 钱伟,车凯,李冰锋.基于组合模型的短时交通流量预测[J].控制工程,2019,26(01):125-130.
- [15] 黄国伦,诸葛丽强,杨乐,等.基于LSTM的电力通信流量预测[J].中国科技信息,2020(Z1):105-108.
- [16] 江务学.基于结构优化递归神经网络的网络流量预测[J].西南大学学报(自然科学版),2016,38(02):149-154.