

# 基于dtw-prophet算法的模型在服务器分区使用率性能数据预测中的应用

康凯<sup>1</sup>, 赵以爽<sup>2</sup>, 荆亚<sup>3</sup>, 郭锋<sup>3</sup>, 王晓通<sup>3</sup>

1.中国联通集团云网运营中心, 河南郑州, 450000; 2.中讯邮电咨询设计院有限公司, 河南郑州, 450000;  
3.济南浪潮数据技术有限公司, 山东济南, 250000

**摘要:** 服务器是计算机的一种, 它比普通计算机运行更快、负载更高、价格更贵。服务器在网络中为其其他客户机提供计算或者应用服务。在运行过程中, 服务器各项性能指标能够很好地反馈其运行状态、使用场景。因此, 对服务器性能数据的研究能够很好地分析服务器的工况、潜在风险, 提前规避进而提升服务器的稳定性、可靠性。传统的服务器分区使用率预测主要包括Istm、arima、svm、prophet、三次平滑等模型, 而本文构建的dtw-prophet模型, 将prophet算法作为基础, 使用dtw算法对数据的不同周期进行最优化求解, 从而寻找prophet的最优季节项周期。结果表明, 该模型对时序序列数据有较强的适应力, 对时序数据的预测效果较佳。该模型在数据处理、模型构建方向提出了新的思路, 提升了自身的适应力、稳定性。

**关键词:** prophet算法; dtw算法; 服务器性能预测; 数据处理

**中图分类号:** G623.58

**文献标志码:** A

## 0 引言

服务器性能是服务器应用的重要指标, 性能稳定的服务器能够提供高效迅速、稳定及时的访问响应, 因此, 对服务器性能进行预测评估以及运维, 能确保服务器的可靠稳定运行, 而如何对服务器性能进行预测评估已成为当前研究的重点<sup>[1]</sup>。

服务器性能数据主要有内存使用率、CPU使用率、分区使用率和系统负载等。现阶段, 服务器性能预测的研究重点是对各服务器性能数据进行预测和对场景、故障、需求进行分析, 从而对潜在的风险进行提前删减以及规避, 降低运维风险, 提升系统的稳定性<sup>[2]</sup>。

## 1 背景

服务器是数字化转型和云计算、互联网等技术发展的基础, 其技术直接影响了各行业的发展

速度<sup>[3]</sup>。如今, 服务器智能、硬件加速、分布存储、安全稳定等成为服务器发展的重点方向。服务器稳定性的研究包含自动纠稳、提前预防等<sup>[4]</sup>。提前预防指对服务器状态进行模型预测, 分析潜在的风险, 并增删操作以规避风险。因此, 服务器性能预测已成为提升服务器稳定性的一种思路和各厂商研究的突破点。对服务器性能数据, 人们应根据其时序特性采用时序预测算法模型进行研究<sup>[5]</sup>。

目前, 国内外主流的预测算法有: (1) 时间序列基本规则法, 即周期因子法; (2) 线性回归, 即利用时间特征做线性回归; (3) 传统时序建模方法, 即ARMA/ARIMA等线性模型; (4) 时间序列分解, 即使用加法模型或乘法模型将原始序列拆分为4部分; (5) 从机器学习特征工程着手, 采用时间滑窗改变数据的组织方式, 使用XGBoost/LSTM模型/时间卷积网络等; (6) 转化为监督学习数据集, 使用XGBoost/LSTM模型/时间卷积网络/Seq2Seq; (7) Facebook-prophet, 其

**作者简介:** 康凯, 男, 汉族, 山西朔州, 硕士研究生, 高级工程师, 研究方向: 云计算基础设施架构、云平台开发。

在控制程度和可解释性上比传统时序模型更有优势，所以单独列出；（8）深度学习网络，结合CNN+RNN+Attention，作用各不相同互相配合<sup>[6]</sup>。

针对服务器性能预测，国外主流厂家采样的算法模型有svm、lstm、prophet等。起初，svm、lstm算法在性能预测中占据主流地位，但prophet的横空出世，很快占据了时序数据预测的主流，现阶段已成为时序预测的主流选择。国内该方面的研究起步较晚，目前主要应用的算法模型有ES和prophet。鉴于prophet的易用性、稳定性，越来越多的厂商着手研究prophet模型，并基于此模型衍生出众多新型算法模型<sup>[7]</sup>。

## 2 模型构建

### 2.1 prophet算法模型原理

时间序列预测是指将获得的数据按时间顺序排成序列，分析其变化方向和程度，从而对未来若干时期可能达到的水平进行推测，即其是利用历史数据对未来数据进行预测的一种研究。

prophet算法是基于时间序列分解和机器学习的拟合来做的，适用于具有强烈季节性和有多个季节历史数据的时间序列。prophet对缺失数据和趋势变化具有鲁棒性，并且通常可以很好地处理异常值。

prophet算法将时间序列拆分为 $s(t)$ （季节项）、 $g(t)$ （趋势项）、 $h(t)$ （节假日项）、 $\varepsilon(t)$ （剩余项）四项，一般有加法模型和乘法模型。

$$y(t) = s(t) + g(t) + h(t) + \varepsilon(t) \quad (1)$$

式(1)为加法模型；

$$y(t) = s(t) \times g(t) \times h(t) \times \varepsilon(t) \quad (2)$$

式(2)为乘法模型，同时取对数可转换为加法模型： $\ln(y) = \ln(s) + \ln(g) + \ln(h) + \ln(\varepsilon)$

趋势项： $g(t)$ 为趋势函数，表示时间序列上的非周期变化；趋势项有两个重要函数：逻辑回归函数和线性回归函数。

(1) 逻辑回归函数：

$$g(t) = \frac{c(t)}{1 + \exp(-(k + a(t)^T \delta) \times (t - (m + a(t)^T r)))}$$

逻辑回归函数主要用于拟合非线性变化的

趋势， $c(t)$ 表示增长的饱和值， $k + a(t)^T \delta$ 表示增长率， $m + a(t)^T r$ 表示线性偏移。

(2) 线性回归函数：

$$g(t) = (k + a(t)^T \delta) \times t + (m + a(t)^T r)$$

季节项：时间序列通常会随天、周、月、年等呈现季节性的变化，即周期性变化，模型中采用傅里叶级数变化计算。

$$s(t) = \sum_{n=1}^N (a_n \cos\left(\frac{2\pi n t}{p}\right) + b_n \sin\left(\frac{2\pi n t}{p}\right))$$

节假日项：

$$h(t) = z(t)_k = \sum_{i=1}^l k_i \times l \{i \in d_i\} \quad k \sim \text{Normal}(0, v^2)$$

### 2.2 dtw算法原理

dtw（动态时间规整算法）最初用于识别语音的相似性，是一种计算两个时间序列数据（尤其是两个时间长度不同的序列数据）相似度的一种动态规划算法。此方案对原始数据采用dtw算法寻找适用于prophet算法的周期序列。

假设有两个时序序列 $q$ 和 $c$ ，长度分别为 $m$ 和 $n$ ，

$$c = c_1 c_2 \cdots c_m$$

$$q = q_1 q_2 \cdots q_n$$

为了对齐两个序列，算法中构建了个 $m \times n$ 矩阵，矩阵的位置 $(i, j)$ 用来存储点 $c_i$ 与 $q_j$ 的距离（通常为欧式距离，即 $d_{(i,j)} = \sqrt{(c_i - q_j)^2}$ ），dtw算法就是从矩阵中找一条从 $(1,1)$ 到 $(m, n)$ 开销最小的道路：

$$w = w_1 w_2 \cdots w_k \text{ 其中, } \max\{m, n\} \leq k \leq m+n-1$$

根据时间序列的特点，路径应满足以下特征：

(1) 边界条件：

$w_1 = (1,1)$ 和 $w_k = (m, n)$ ，即开始和最后时刻的对齐是确定的，路径必须从矩阵的左下角到右上角结束。

(2) 连续性：

如果 $w_{k-1} = (a', b')$ ，则对于下一路径点 $w_k = (a, b)$ ，需满足：

$$0 \leq (a - a')$$

$$0 \leq (b - b')$$

即路径上的点是随着时间单调进行的。根据上述，dtw算法就是要找出一条从左下角到右上

角累加距离最小的一条连续路径,定义 $r(i, j)$ 为矩阵左下角到矩阵 $c_{(i, j)}$ 的最小累加距离,则

$$r(i, j) = d(c_i, c_j) + \min \{r(i-1, j), r(i, j-1), r(i-1, j-1)\}$$

### 2.3 dtw-prophet模型构建

dtw-prophet模型结合了dtw算法与prophet算法,使用dtw算法对预测原始数据集进行运算,寻找最优prophet使用训练周期,其工作结构的模型流程为:原始数据→预处理→dtw算法→prophet算法,具体如下:

(1) 模型获取原始数据后,对原始数据进行预处理,主要包括重采样、数据有效性分析、数据补全、初始周期提取、输出重采样及补全后的训练数据;

(2) dtw算法对1中输出的训练数据进行最接近周期查找,获得使距离最近的数据周期;

(3) prophet依据1中输出的训练数据、2中输出的最优数据周期,构建自定义季节项,构建prophet算法模型进行预测。

## 3 模型应用

### 3.1 数据来源

将近6个月服务器分区使用率作为原始数据对模型进行验证。该模型适用于对具有时序、周期特性的事件、场景进行预测。因此,具有时序周期特性的服务器性能数据(内存使用率、分区使用率、系统负载等)皆适用于此模型的预测验证<sup>[8]</sup>。

该数据以小时(H)为采样周期,对服务器分区使用率进行采样,获取近6个月的服务器分区使用率数据。该数据共包含4103个数据点、近6个月的时序数据,满足prophet算法的应用环境<sup>[9-10]</sup>。

### 3.2 数据分析

对数据文档进行分析可知,该服务器分区使用率呈现递增趋势,且最低使用率为2.33,出现在2022-01-14 11:00:00;最大使用率为6.07,出现于2022-06-23 16:00:00。服务器分区使用率受节假日等突变点影响较小,我们以[24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 264, 288, 312, 336, 360, 384, 408, 432, 456, 480,

504, 528, 552, 576, 600, 624, 648, 672, 696, 720]为间隔对数据进行周期分割,采用dtw算法对各分割后的数据矩阵进行运算(表1):

表1 运算数据

| fre | dtw    | fre | dtw    | fre | dtw    |
|-----|--------|-----|--------|-----|--------|
| 24  | 4.1543 | 264 | 4.1015 | 504 | 4.1332 |
| 48  | 4.1328 | 288 | 4.2004 | 528 | 4.1912 |
| 72  | 4.0349 | 312 | 4.181  | 552 | 4.2008 |
| 96  | 4.0036 | 336 | 4.1616 | 576 | 4.2011 |
| 120 | 3.9906 | 360 | 4.0325 | 600 | 4.1889 |
| 144 | 4.0015 | 384 | 4.0028 | 624 | 4.1518 |
| 168 | 4.0013 | 408 | 4.1126 | 648 | 4.1601 |
| 192 | 4.0893 | 432 | 4.0002 | 672 | 4.1255 |
| 216 | 4.11   | 456 | 4.0129 | 696 | 4.0884 |
| 240 | 4.1413 | 480 | 4.0966 | 720 | 4.0802 |

依据dtw分析,可知数据存在120个采样点,即5天的最优周期。因此,在prophet模型新建自定义周期:

```
m.add_seasonality(name='self5days',
                  period=5,
                  fourier_order=3,
                  prior_scale=1.0,
                  mode='additive',
                  )
```

分别以无自定义周期与有自定义周期构建prophet模型,对数据进行训练预测。

### 3.3 结果分析

对比图1可知,在prophet模型使用dtw算法计算季节性周期最优值后,预测稳定性MES:由1.84808提升为1.73462。由此可见,该模型能够有效适用于时序数据预测,提升数据预测的稳定性和可靠性。

## 4 结语

prophet模型在时序预测中能够提供高效稳定的预测效果。基于该模型衍生的dtw-prophet模型能够提升模型对数据的适应性,提升预测的稳定性和契合度,且能够很好地筛选出适用于模型的季节性周期参数。自定义季节模型能够有效地解决原始数据周期不显性的问题,并在对数据的季节性参数的提取、模型的构建上提供了新的思路。

(下转第012页)



行为轨迹,形成了全流程、全量、全要素的客群画像。

通过监测商场整体及每个商铺的客流漏斗,运营者可以洞察顾客客群需求,清楚地了解客群流量、流向、流速和场域客群流量的分布及相关商业属性标签,从而知道顾客客群关注哪些商铺,更愿意去哪些区域,什么时间段客流的进店率高,甚至知道不同的商铺组合对顾客客群的吸引度高。

AI精准客流系统可以对商铺的经营能力进行精准评估诊断,锁定所属客群标签,帮助店铺经营得到改善和提升,降低空置率。

总之,通过对AI精准客流数据的采集及运用,经营者的管理能力和水平得到了明显的提高;精准客流系统为商业地产带来了真正的基于顾客全流程数字化的深度运营。

### 参考文献

- [1] 业界领先算法,客流数据福音: 趋视智能视频客流统计分析仪[J].中国公共安全,2015(22):188-189.
- [2] 袁帅.具有客流统计功能的商场视频监控系统设计与实现[D].青岛:中国海洋大学,2013.
- [3] 李超.红外传感器客流计数系统的设计[D].石家庄:河北科技大学,2010.

(上接第007页)

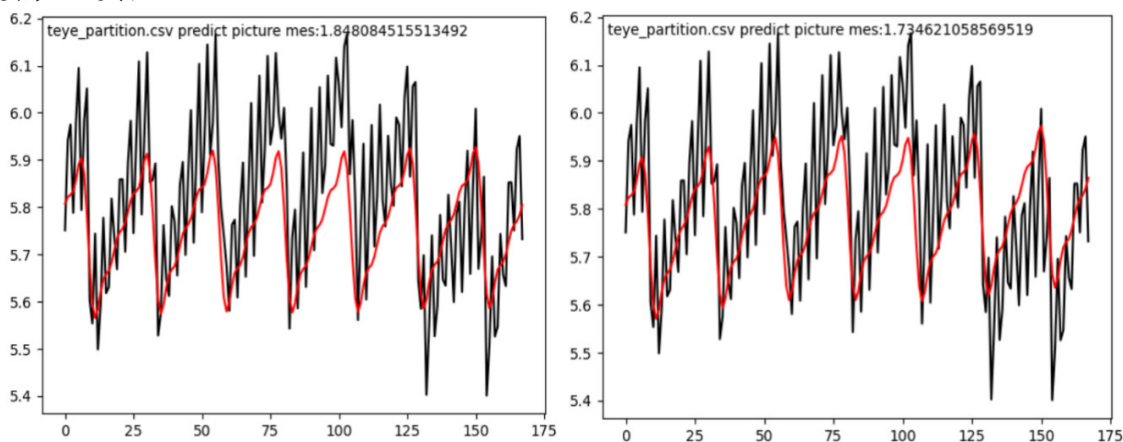


图1 prophet模型无自定义和有自定义周期预测

### 参考文献

- [1] 郇松桦,刘秀丽.一种改进的CEEMDAN-LSTM-Neural Prophet Net模型:用于COVID-19背景下我国月度消费预测[J].计量经济学报,2022,2(3):620-643.
- [2] 鲁铁定,李祯.基于Prophet-XGBoost模型的GNSS高程时间序列预测[J].大地测量与地球动力学,2022,42(9):898-903.
- [3] 杨正宇,韩彤,赵振刚.基于ARIMA-GWO-SVR组合模型的线损率时序预测[J].供用电,2022,39(7):58-63.
- [4] 黄心裕,唐军,王晓宇.基于Prophet算法的海南近海波浪长时段时序分析与预测[J].海洋学报(中文版),2022,44(4):114-121.
- [5] 姜洪军.浪潮积极应对服务器市场新变局[N].中国信息化周报,2014-11-17(10).
- [6] 金鹏.中国IA服务器发展趋势[J].微电脑世界,2001(10):38-39.
- [7] 陈慧.基于SVM算法的服务器性能分析与预测研究[D].南昌:南昌大学,2019.
- [8] 李丹青,杨超宇.基于ARIMA和LSTM的瓦斯浓度预测研究[J].新乡学院学报,2022,39(9):45-48.
- [9] 佟喜峰,樊鑫.基于DTW的手势识别算法[J].计算机与数字工程,2022,50(8):1782-1786.
- [10] 宗文泽,吴永明,徐计,等.基于DTW-kmedoids算法的时间序列数据异常检测[J].组合机床与自动化加工技术,2022(5):120-124+128.