

基于动态系数的三次指数平滑算法负载预测

罗辰辉¹, 揭晶方², 张伟¹, 沈琼霞³

(1. 湖北大学 计算机与信息工程学院, 武汉 430062; 2. 湖北大学 楚才学院, 武汉 430062; 3. 烽火通信科技股份有限公司 业务与终端产出线, 武汉 430073)

摘要: 数据中心是企业信息化的重要组成部分, 云计算的核心思想就是把数据中心整成一个资源池, 对资源池进行统一的调度与管理; 随着虚拟化技术的发展, 目前对数据中心的资源利用率越来越高, 但是还是存在大量资源浪费的情况, 其原因在于当前对数据中心未来负载预测的算法还存在一定的局限性, 如果对未来负载预测值远远大于实际负载情况, 则导致大量的虚拟机资源利用率不高, 反之则会导致虚拟机的资源使用率消耗增大, 云平台中不同物理服务器之间的负载情况不平衡, 一部分物理服务器负载过大, 导致云计算平台响应时间过长; 因此云计算平台选取一个合适的负载预测算法显得越发重要, 如何权衡以上问题, 是云计算里面的一个重点研究方向; 负载预测选取时间序列预测算法中的三次指数平滑法, 在该算法原有的静态系数基础之上, 设计了一种动态系数提取方法; 通过等距法把静态系数分成若干份进行训练, 然后在预测过程中提取该时段误差最小值所对应的系数; 在预测结束后, 重新计算其误差, 并通过均值法覆盖旧误差; 实验结果表明, 基于自适应三次指数平滑算法其预测误差明显小于静态系数所预测的误差, 计算复杂度低, 具有一定的应用价值。

关键词: 数据中心; 指数平滑算法; 虚拟化; 负载预测

Load Prediction of Three Exponential Smoothing Algorithm Based on Dynamic Coefficients

Luo Chenhui¹, Jie Jingfang², Zhang Wei¹, Shen Qionxia³

(1. School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China; 2. Chucai Honors College, Hubei University, Wuhan 430062, China; 3. Service and CPE Business Unit, Fiberhome Telecommunication Technologies Co., Ltd., Wuhan 430073, China)

Abstract: data center is an important part of enterprise informatization, and the core idea of cloud computing is to integrate data center into a resource pool, and to conduct unified scheduling and management of resource pools. With the development of virtualization technology, the current resource utilization of the data center is higher and higher, but there are still a lot of waste of resources situation, the reason for this is that the current load forecasting algorithm for data center in the future also has certain limitation, if the load forecast for the future is far greater than the actual load, leading to a large number of the virtual machine resources utilization rate is not high, the opposite can lead to increase consumption of resource utilization, the virtual machine between different physical servers in the cloud platform of load imbalance, part of the physical server load is too large, lead to cloud computing platform response time is too long. Therefore, it is more and more important to select a suitable load prediction algorithm for cloud computing platform. How to balance the above problems is a key research direction in cloud computing. Based on the static coefficients of the algorithm, a dynamic coefficient extraction method is designed. By equidistant method, the static coefficients are divided into several parts for training, and then the corresponding coefficients of the minimum error of the time period are extracted in the prediction process. After the prediction is over, the error is recalculated and the old error is covered by means of means. The experimental results show that the prediction error is significantly less than the prediction error of the static coefficient based on the adaptive cubic index smoothing algorithm, and the computational complexity is low, which has certain application value.

Keywords: data center; exponential smoothing algorithm; virtualization; load prediction

0 引言

2006年亚马逊推出了云计算平台之后, 云计算就如火

如荼地发展着。国外的如微软、oracle、IBM 等公司, 国内的阿里、腾讯、新浪等, 在云计算上都取得一定的研究成果。云计算的核心技术就是虚拟化技术, 其特点就是把数据中心的物理服务器虚拟成一个庞大的资源池, 供用户使用。虚拟化技术让数据中心具有可伸缩的特点, 提高数据中心的资源利用率, 对应用隔离, 提高数据中心灾难恢复能力, 减少碳排放量等起到重要作用。但是如何高效地利用数据中心资源池里面的 CPU、内存、硬盘、带宽资源, 已经成为很多互联网企业所面临的问题。但是目前数据中

收稿日期: 2018-02-12; 修回日期: 2018-03-26。

基金项目: 国家自然科学基金(61301144; 61601175)。

作者简介: 罗辰辉(1991-), 男, 湖北省黄冈市人, 研究生, 主要从事物联网、大数据方向的研究。

通讯作者: 张伟(1979-), 男, 湖北省武汉市人, 博士, 讲师, 硕士研究生导师, 主要从事无线通信、物联网、嵌入式系统方向的研究。

心存在很多宿主机资源使用率不高的情况,比如天猫双十一负载达到顶峰时期,宿主机资源使用率高,其他时期基本上很多宿主机都处于比较空闲的状态,大量的资源使用率不高的宿主机导致一些不必要的资源浪费。如何均衡数据中心虚拟机负载,现在已成为云计算当前的一个重点研究方向。云平台如何负载均衡很大程度上取决于虚拟机未来负载的情况,由于数据中心的用户访问量每时每刻都在发生变化,当前的负载情况往往不能决定数据中心未来的负载均衡策略,因此既要保证用户的云服务体验,又要避免大量的资源浪费,云平台的负载均衡策略往往起到重要作用,而虚拟机负载预测是云平台负载均衡策略的重要依据,因此虚拟机在未来时期的负载预测^[1]对整个云平台负载均衡调控尤为重要。当虚拟机的实际负载高于预测负载时,由于资源不足导致整个云平台的响应速度过慢,以至于影响用户体验。当虚拟机的实际负载又远远小于预测负载,则会导致一部分物理服务器过多的资源处于空闲状态。因此,数据中心的负载预测对减少整个数据中心资源的消耗以及提升云服务体验都起到重要的作用。

基于这种情况,本文采用指数平滑法^[2-5]预测数据中心的未来负载情况,该算法在时间序列预测模型中得到了大量的实验证明,其预测结果有一定的说服力。然而该算法有一个缺点,其预测系数没有一个明确的求解方法,一般是由人为经验估算得出,并且不能随着实际情况动态地改变其预测系数,导致预测负载与真实情况存在较大的误差。在这种背景下,本文设计了一种基于动态系数^[6-8]的三次平滑指数来预测数据中心负载走向。

1 虚拟机负载预测

在云计算平台中,随着虚拟机的负载越来越大,就意味着对虚拟机对宿主机的资源使用率越来越高。如果宿主机的资源使用率(包括计算、I/O等)达到一定范围,则会导致该宿主机上的虚拟机出现对资源竞争的情况,导致云服务的响应速度变慢。为了避免这种情况的出现,VMM(虚拟机管理器)能够及时调控能起到关键作用,因此虚拟机负载预测对整个云平台的负载均衡调控十分重要。

时间序列预测算法是基于时间序列的历史信息来完成预测,其基本思路是根据已有的序列数据,发现某种隐含的规则,来预测序列后面的走向。以下是本文对时间序列预测算法进行一个简单的介绍。

1.1 时间序列预测算法

时间序列预测算法包含移动平均法^[9]、指数平滑法、BP神经网络预测算法^[10-11]、回归预测法^[12]等。移动平均法将近期和远期数据进行平均化,因此只适合近期变化不大的时间序列进行预测。如果序列处于某种上升或者下降比较明显的趋势时,就不适合作为该序列的预测算法。指数平滑法根据上期的预测值与真实值,通过加权的方式预测下一期的走势,其优点就是不用保存大量历史数据,节省很多内存空间,是一种广泛使用的预测算法。BP神经网络

预测算法其基本原理是输入信号 X_i 通过隐层点作用于输出层点,经过非线性转换,产生输出信号 Y_k 。如果在输出层得不到期望的输出,则使误差往梯度方向下降,经过反复的学习训练,最终得到误差最小的网络系数。该算法优点就是非线性序列预测精准度高,具有自学习与自适应能力,有一定的容错能力,缺点就是其算法本质上是梯度下降法,需要进行复杂的科学计算,导致其学习速度很慢。回归预测法是在自变量与因变量相关关系基础之上,建立变量之间的相关方程,并将方程作为预测模型。根据自变量个数的不同,分为一元回归预测法和多元回归预测法。该算法的优点是在分析多因素模型中,更加简单方便,缺点就是需要大量的历史数据计算才能得到回归因子。

1.2 指数平滑法

指数平滑法由布朗提出,它是一种特殊的加权移动平均法,通过加权平滑系数,在时间序列模型中通过上一期预测值与真实值加权平均得到下一期的预测值。其数学表达式为:

$$y_{t+1} = y_t + \alpha(x_t - y_t) \quad (1)$$

式中, x_t 是上期的观测值, y_t 是上期的预测值, y_{t+1} 是下阶段的预测值,指数平滑算法只需要上期的预测值与观测值,不需要大量的数据计算,减少对服务器的负载压力。在指数平滑算法中,预测成功的关键是 α 的选择。 α 的大小规定了在新预测值中新数据和原预测值的比例。 α 值越大,新数据所占的比例就越大,原数据所占比重就越小,反之亦然。

1.3 三次指数平滑法

根据平滑次数的不同,指数平滑法分为一次指数平滑法、二次指数平滑法、三次指数平滑法等^[13]。二次指数平滑法是建立在一次指数平滑法基础之上的,适合预测线性时间序列,所预测的效果也优于一次指数平滑法。若时间序列的走向呈现出抛物线式曲线,则需要采用三次指数平滑法进行预测,三次指数平滑法是在二次指数平滑法基础之上再进行一次平滑^[14]。由于数据中心的虚拟机负载时间序列呈抛物线状,所以采用三次指数平滑法更为合适。不同次数的指数平滑法适用于何种类型的序列如表 1 所示。

表 1 不同平滑次数所适用序列

| 指数平滑法次数 | 适用于类型 |
|---------|-------|
| 一次指数平滑法 | 平滑序列 |
| 二次指数平滑法 | 线性序列 |
| 三次指数平滑法 | 抛物线序列 |

设数据中心虚拟机负载值为 $\{X_t\}$ ($t=1, 2, 3\cdots$), 第 n 时刻虚拟机负载一次、二次、三次指数平滑值分别记为 $S_t^{(1)}$ 、 $S_t^{(2)}$ 、 $S_t^{(3)}$, 各指数平滑值计算公式如下:

$$\begin{cases} S_t^{(1)} = \alpha X_{t-1} + (1-\alpha)S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)} \end{cases} \quad (2)$$

采用三次指数平滑法对 t 期后的 q 期数据进行预测得到,公式为:

$$X_{t+q} = a_t + b_t q + c_t q^2 \quad (3)$$

其中: a_t, b_t, c_t 为第 t 期的预测参数, 计算公式为:

$$a_t = 3s_t^1 - 3s_t^2 + s_t^3$$

$$b_t = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)s_t^1 - 2(5-4\alpha)s_t^2 + (4-3\alpha)s_t^3]$$

$$c_t = \frac{\alpha^2 (s_t^1 - 2s_t^2 + s_t^3)}{2(1-\alpha)^2} \quad (4)$$

1.4 平滑系数的选择

由于平滑系数的选择对负载预测的影响很大, 因此对于整体的预测模型十分重要。 α 取值范围为 $0 \sim 1$, 一般情况下, 根据经验, 其取值范围 $0.30 \sim 0.70^{[6]}$ 。 α 的取值一旦固定, 则加权系数也随之固定, 即无法修改。由式 (1) 得知, 在一次指数平滑预测模型中, α 越大则历史数据占的比例越高, α 越小则预测数据占的比例就越高。因此 α 的取值大小, 取决于模型对时间序列的变化速度, α 取较小值时, 预测序列平滑能力较强, 反之模型对时间序列的变化反应速度较快。当时间序列呈现水平趋势时, 应选取较小的序列值, 常在 $0.10 \sim 0.30$ 之间, 序列有波动且长期趋势变化不大时, 可以选稍大的系数值, 常在 $0.30 \sim 0.50$ 之间取值, 序列波动较大, 呈现明显上升或者下降时, 可选择较大的系数值, 可以从 $0.60 \sim 0.80$ 之间取值^[15]。

2 基于自适应三次指数平滑系数负载预测

2.1 模型缺陷

传统的指数平滑法能够简单高效地预测时间序列的走势, 但是模型有一个明显缺陷, 预测系数一旦确定就无法改变, 无法根据时间序列的波动而动态变化。这种静态的平滑系数导致预测过程中, 无法适应时间序列的动态变化的情况, 对数据跳跃较为明显的序列预测效果不佳。

在预测过程中, 平滑系数反应的是数据的变化趋势。传统模型中, 平滑系数一旦确定即为一个常数, 使得整个预测模型中自适应能力低。在大多数应用场景中, 时间序列往往随着时间的变化而发生动态变化, 难以捕捉到其变化趋势。本文所预测的数据中心负载波动往往较大, 传统的静态三次指数平滑法无法适应环境的变化, 导致预测效果不佳。这也表明, 动态系数在指数平滑预测模型中显得越来越重要, 能够适应过去数据变化的能力, 采用动态系数对过去的系数进行一个修正, 实验表明动态系数在整个预测模型中起到决定性作用。

2.2 动态系数求取

在实际应用中, 为了让三次指数平滑系数能更好的适应环境的变化, 达到最佳的负载预测效果, 本文采用能够动态变化的平滑系数。首先考虑到云平台负载情况往往在某个时间段内变化不大, 根据负载变化将云平台负载预测划分为若干个时间段。对于不同时段, 平滑系数 α 采用步长为 0.01 的迭代, 通过反复训练 α , 提高预测的精准度。具体步骤如下:

1) 选定某个时段的历史数据, 采用等距法进行搜索最

佳值。设定距长 λ 为 0.01 , 则将 α 分为 0.01 相同的距离区间, 记为 $\alpha_k (k = 1, 2, \dots)$ 。

2) 根据误差平方和最小原则选取最佳平滑系数, 即在这个时间段选取误差所对应最小的平滑系数 α 。

$$\sigma = \min_i \sum_{i=1}^{100} (x_i - \bar{x})^2 \quad (5)$$

3) 通过步骤 (2) 得到最小误差所对应的系数 α , 新系数所预测的负载数据跟实际数据存在误差, 那么新的误差值为之前误差与新误差二者的均值, 并且覆盖原来的误差, 即:

$$\sigma = \frac{\sigma_1 + \sigma_2}{2} \quad (6)$$

4) 重复步骤 2)、3) 的操作, 选取误差所对应最小的三次平滑系数, 并将新预测的误差与旧误差二者的均值重新覆盖原来的旧误差。整个算法流程如图 1 所示。

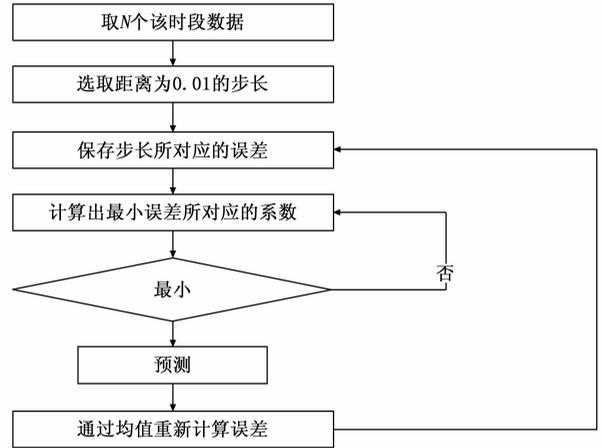


图 1 动态平滑系数求取流程图

2.3 基于动态三次平滑系数负载预测

通过式 (5) 与式 (6) 得到最优平滑系数, 再通过式 (2) ~ 式 (4) 完成负载预测。本文采用误差比对方式获取最佳系数值, 通过新一轮的预测再重新校对误差, 相比传统的静态系数预测更加精准, 此种方式不仅提高了预测的精准度, 同时也有效解决了负载暴增或者暴减的突发情况。

3 实验及分析

为了验证文中提出的基于动态系数的三次指数平滑算法预测的精准度, 本文采用某智慧农业云平台虚拟机负载时间序列作为本文的实验数据。

3.1 验证过程

假设在需要预测未来负载走向的数据中心上, 其运行在物理服务器之上的虚拟机负载 L_i 表示形式为:

$$L_i = u_i * p_i \quad (7)$$

其中: u_i 表示虚拟机 CPU 的使用率, p_i 表示虚拟机 CPU 核心数量。若 h_0 是一台双核心的虚拟云服务器, $p_0 = 2$, 现在 CPU 的使用率是 0.3 , 那么 $L_0 = 2 * 0.3 = 0.6$ 。本文采用的智慧农业云平台某时期负载作为实验数据对动态预测进行验证, 如表 2 所示。

表 2 智慧农业云平台某时期不同时段负载情况

| 时间段 | 负载 |
|----------|------|
| T_0 | 0.64 |
| T_1 | 0.72 |
| T_2 | 0.90 |
| T_3 | 1.16 |
| T_4 | 1.32 |
| T_5 | 1.10 |
| T_6 | 0.84 |
| T_7 | 0.78 |
| T_8 | 0.68 |
| T_9 | 0.84 |
| T_{10} | 1.20 |
| T_{11} | 1.48 |
| T_{12} | 1.27 |
| T_{13} | 0.80 |
| T_{14} | 0.71 |

通过对数据分析发现，智慧农业云平台虚拟机负载数据呈抛物线变化情况，可以用三次指数平滑算法进行预测。本文首先采用两种静态系数来预测，分别为 0.30 和 0.50，用 python 语言绘制出初始值之后时段的预测值，对比两种静态系数所预测的效果，其结果如图 2 所示。

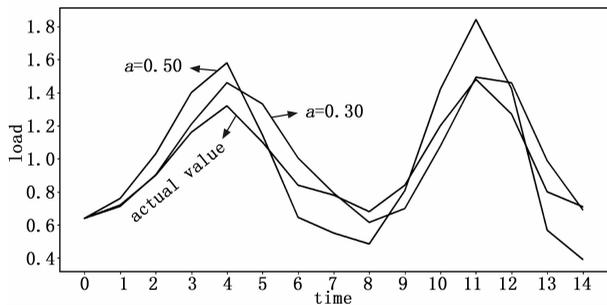


图 2 静态系数与真实值对比结果图

接下来采用动态系数进行预测，其预测结果与实际情况对比如图 3 所示。

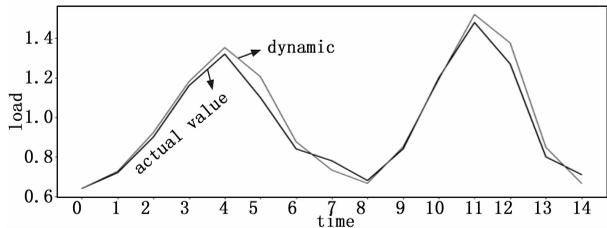


图 3 动态系数与真实值对比结果图

3.2 结果分析

由图 2 可知，静态平滑系数为 0.3 所预测负载的准确度比系数为 0.5 要高，更贴近真实值。而图 3 所采用的动态系数是通过反复迭代寻找最小误差所对应的最优解得到的，其预测负载准确度较以上两个静态系数更高。实验结果表

明，本文所采用的动态系数三次指数平滑法预测负载相比传统的静态系数准确度更高，预测系数能够随着环境的变化而发生动态改变，计算复杂度低，在工程上具有一定的应用价值。

4 结束语

本文采用基于动态系数三次指数平滑法预测数据中心虚拟机负载，传统方法是采用静态系数去预测未来的时间序列，其缺点是不能根据环境变化动态调整其预测系数。本文通过在不同时段采用等距迭代的方式反复测量不同系数的误差情况来求解最佳系数，新系数预测出来的误差又重新与旧误差产生一个均值并覆盖原来的误差。实验表明，本文提出的动态系数三次指数平滑法预测负载的误差小于静态系数，具有更好的实用性。

参考文献:

- [1] 刘晓艳, 王 颖. 基于改进云模型的云计算负载预测 [J]. 计算机应用研究, 2015, 32 (10): 3124 - 3127.
- [2] 唐炎森. 指数平滑预测公式与平滑系数 [J]. 统计与信息论坛, 1998 (1): 39 - 44.
- [3] 张忠平. 指数平滑法 [M]. 北京: 中国统计出版社, 1996.
- [4] 黎锁平, 刘坤会. 动态指数平滑优化模型及其应用 [J]. 系统与工程学报, 2013, 18 (2): 163 - 167.
- [5] 徐大江. 预测模型参数的指数平滑估计法及其应用的进一步研究 [D]. 浙江: 浙江财经学院经济数学教研室, 1999.
- [6] 沈海迪, 万振凯. 基于指数平滑法的动态预测机制 [J]. 计算机技术与发展, 2017, 27 (07): 6 - 9.
- [7] 杨济瑞, 赵海涛, 刘南杰. 改进的三次指数平滑法及其在车联网中的应用 [J]. 计算机技术与发展, 2017, 26 (11): 164 - 167.
- [8] 张 丽. 基于动态指数平滑模型的粮食价格预测方法研究 [D]. 河南: 河南工业大学, 2011.
- [9] 刘剑宇, 熊允发. 移动平均法在公安情报分析中的应用 [J]. 中国人民公安大学学报, 2007, 54 (4): 54 - 56.
- [10] 许兴军, 颜钢锋. 基于 BP 神经网络的股价趋势分析 [J]. 浙江金融, 2011 (11): 57 - 59, 64.
- [11] 王晨辉, 张晓亮, 梁晓传. 云计算架构下基于 BP 神经网络负载预测策略的研究 [J]. 电子信息与通信技术, 2016, 14 (11): 46 - 50.
- [12] 李永娜. 基于支持向量机的回归预测综述 [J]. 信息通信, 2014, 143 (11): 32 - 33.
- [13] 贾学力, 焦 帅, 李建军. 二次指数平滑在交通信息短期预测中的应用 [J]. 公路交通科技, 2011, 28 (S1): 101 - 104, 109.
- [14] 冯金巧, 杨兆升, 张 林, 等. 一种自适应指数平滑动态预测模型 [J]. 吉林大学学报 (工学版), 2007 (6): 1284 - 1287.
- [15] 张嘉望, 郭军献, 李福松. 基于最优平滑系数三次指数平滑法的转速预测 [J]. 探测与控制学报, 2015, 37 (5): 43 - 46.