

基于深度学习的云计算虚拟机分类算法^①

董 萍^{1,2}

1. 河南科技大学 应用工程学院, 河南 三门峡, 472000;

2. 三门峡职业技术学院, 信息传媒学院, 河南 三门峡, 472000

摘要: 随着云计算数据中心规模及复杂性的不断增长, 其监控和管理的可扩展性问题已逐渐成为一项挑战性任务. 对云计算平台中相似行为的虚拟机进行聚类处理, 可提升云计算平台监控和管理的可扩展性. 然而, 在准确率和时效性上, 现有的虚拟机聚类技术无法满足云计算平台的要求. 本文提出一种基于深度学习的虚拟机分类算法, 即对云计算平台中虚拟机的行为进行分析, 将具有相似行为的虚拟机进行分类, 然后将同类的虚拟机部署在尽可能少的服务器上、降低基础设施成本, 从而提升云计算数据中心监控和管理的可扩展性. 采用真实云数据中心的虚拟机数据对本文算法及同类算法进行对比实验, 结果表明本文算法准确率更高、实时性更好, 可满足云计算平台的要求.

关 键 词: 云计算; 虚拟机分类; 深度学习

中图分类号: TP391

文献标志码: A

文章编号: 1000-5471(2021)05-0110-05

云计算已成为通过网络交付复杂云服务被广泛采用的范例, 并为科学计算领域提供了更高效的技术与支撑环境^[1]. 它以网络化的方式聚合计算能力与通信资源, 使用虚拟化技术将大量成本低、计算能力弱的资源整合为一个强大的资源池. 因此, 云计算能够实现统一管理和资源调配, 提升云服务性能和价格优势^[2].

在过去几年里, 云服务器的数量及背后基础设施的复杂性都在迅速增加. 云计算基础设施在未来几年应对日益增长的资源需求的能力, 将对未来新兴数字技术的发展起到至关重要的作用. 在基础设施即服务(infrastructure as a service, IaaS)系统中, 资源监控和资源管理是特别关键的任务. 在这些系统中, 大量且不断增长的数据被收集, 用于管理客户委托的虚拟环境^[3-4]. 在这些云系统中, 数据中心管理员通常采用黑箱方法, 其中每个虚拟机(virtual machines, VM)都被认为是独立于其他虚拟机的, 这对监控和管理任务的可伸缩性造成了负面影响.

文献[5-6]研究表明, IaaS 云系统中的可伸缩性问题, 可以通过在资源使用时聚类具有相似行为的 VM 来改善, 即对相似行为的 VM 进行聚类化分组管理. 例如, 自动测定相似的 VM, 然后允许系统来识别每个类的几个虚拟机代表, 紧随其后的是同一个类的其他成员. 这些理论已被用来提高监控策略的可伸缩性, 最近还被应用到一个 VM 管理案例中, 即 IaaS 数据中心的服务器整合中. 在文献[7-8]中, 作者通过使用基于类的方法, 提出了云监控和管理可扩展性的建议. 在文献[9]中, 作者利用资源相关性和主成分分析对虚拟机行为进行建模, 并提出了一种 K 均值聚类算法. 在文献[10]中, 作者基于直方图的表示对 VMs 行为进行建模, 使用 Bhattacharyya 距离来度量 VMs 的相似性, 提出了一种用于聚类的谱算法. 但是, 上述聚类技术只能应用于超过 24 h 的时间序列, 即使在这种条件下它们也具有对 VM 错误分类的问题. 因此, 在 VM

① 收稿日期: 2020-04-21

基金项目: 河南省教育厅高等学校重点科研项目(15B520026); 河南省社科普及规划项目(0984); 三门峡市科学技术局科技发展规划项目(2020010125).

作者简介: 董 萍, 硕士, 副教授, 主要从事嵌入式计算机应用技术研究.

识别准确率和时效性上，现有聚类技术无法满足云计算数据中心监控及管理的可扩展性要求。

针对以上问题，本文提出一种基于深度学习的 VM 分类算法，通过提高 VM 分类的准确率和实时性，达到提升云数据中心监控及管理可扩展性的目的。

1 IaaS 云系统模型

IaaS 系统是利用 VM 自动聚类来实现云监控和管理的可伸缩方法。第一层由本地管理层组成，它在数据中心的每个物理服务器上执行：利用服务器上承载的 VM 资源度量实时监控过载情况，当监控到负载过重的服务器时利用动态 VM 迁移。第二层是全局管理层，它托管在管理节点之上：负责定期执行整合技术，将 VM 放在尽可能少的服务器上，降低基础设施成本，避免昂贵的资源过度供应。

基于 VM 聚类的云计算系统将运行相同客户应用程序的相似 VM 分组在一起。由于可伸缩性原因，聚类处理往往应用于同一客户的 VM。在聚类之后，每个已标识的类只选择很少的 VM 代表。一般选择 3 个 VM 代表，因为所选择的代表可能会意外地更改其与类相关的行为。对每个类的 VM 代表进行高采样频率监控，为周期性整合任务收集信息。假设非代表性 VM 遵循相同类代表的行为，使用更粗的粒度进行监控，可识别影响类更改的行为漂移。

由于 IaaS 云系统的提供者和管理员通常将每个 VM 视为一个黑箱，需要独立于其他 VM 进行监控和管理，从而加剧了这些任务的可伸缩性问题。因此，自动聚类管理的 IaaS 云系统具有以下优势：①监控系统可扩展性；②服务器整合过程的可扩展性；③系统资源需求估算效率。

2 本文算法

2.1 深度学习模型

卷积神经网络能够非常有效地解决涉及多媒体数据的复杂任务，例如图像和视频处理，它适用于识别输入数据中的简单模式，然后用于在后续操作中形成更复杂的模式，最终提供充分的信息进行输出和分类处理。

神经网络由几个卷积层和最后一个完全连接层组成。本文模型也由一系列可变块和一个最终完全连接层组成，用于输出分类。

VM 行为被描述为一组时间序列，每个时间序列描述特定度量（如存储器利用率、CPU 利用率、网络流量等）。由于每个度量是一系列跨时间采样的样本，本质上是一维信号，因此本文利用特定类型的卷积（即一维卷积）作为网络中的基本块，并且将每个度量视为独立卷积运算的输入通道。

一维卷积是深度学习线性操作，用于从一维数据中提取特征，目的是识别特定窗口内的局部模式，称为内核大小 (K_s)。内核中包含用于执行操作的可学习参数。设内核沿着时间维度移动的特定步幅为 s ，由于内核所关注的每个数据补丁执行相同的计算，因此在一个位置学习的模式也可以在另一个位置识别，这使得一维卷积具有平移不变性。

为简化模型，设 N 是批量大小， C 为通道数（对于第一层，通道对应于 VM 度量）， L 为信号序列的长度（称为输入），然后设一个带有输入大小 (N, C_{in}, L_{in}) 和输出 (N, C_{out}, L_{out}) 的卷积 1D 层，此时输出张量的值为

$$out[i, j] = b[j] + \sum_{k=0}^{C_m-1} \omega[j, k] * input[i, k] \quad (1)$$

式(1)中， $*$ 是卷积运算符， $input[i, k]$ 和 $out[i, j]$ 分别为模型输入和模型输出， ω 和 b 为学习参数。

步幅 s 是控制内核步长的一维卷积的参数：如果大于 1，则以更大的步长扫描数据，因此更省时，可使初始尺寸长度减小。此效果可以看作池化处理，即能够减少数据维度。

批量标准化^[11] 是一种流行的操作，它对每次考虑的批处理维度上的数据进行标准化，其中批处理是训练阶段用于加速基于梯度优化数据集的子样本。本文每个卷积层之后均使用批量标准化，这有助于深层网络更快地收敛。最后，采用非线性激活函数 ReLU(Rectified linear unit) 在训练期间稳定梯度。

2.2 基于深度学习的云计算虚拟机分类算法

在本文提出的深度学习网络中，每个块由一个内核大小为 3 且步幅为 2 的卷积 1D 层组成。然后，每个

块后面是批量标准化层和 ReLU 激活函数. 根据输入序列长度 W , 该块重复着可变的次数, 并且最少两次. 在最后一个块之后, 数据将被应用于全连接层, 然后输出分类结果.

模型的块数由下式计算得到

$$N_b = \max(\log_2(W) - 1, 2) \tag{2}$$

式(2)中, W 长度定义为窗口. 改变块的数量可简化模型: 模型需要有一个带有神经元的最终层, 感知域可以在整个输入序列上观察并利用信息. 鉴于这里使用不同的输入序列长度进行实验, 此时需要一个灵活的模型, 即可以根据输入数据的情况调整网络结构的深度. 一般神经元的接受场是神经元可访问的输入数据部分, 并且可以影响其激活.

输入序列长度为 32 个时间步长的通用模型如图 1 所示. 图 1 中概述了数据形状如何通过网络的每个块进行更改, 直到最终的全连接层. 每个度量标准都显示为一列值, 因此将 M 个度量标准放在一起, 可得到 (W, M) 输入形状. 同时添加 Batch-Size 维度 N , 它一次考虑批量大小的输入样本以优化网络, 可获得 (N, W, M) 训练的最终输入形状.

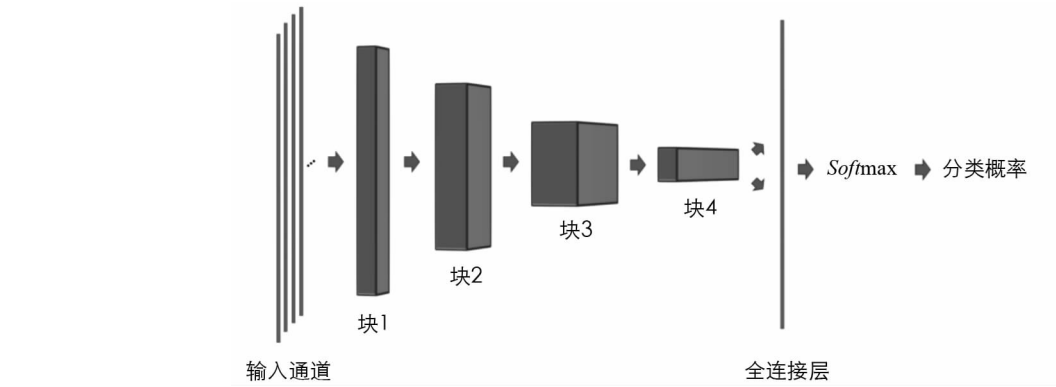


图 1 本文算法模型

此模型中给出的最终 out 分类的概率为

$$P(out) = \text{Softmax}(out) = \frac{e^{out}}{\sum e^{out}} \tag{3}$$

式(3)中, out 是模型的输出, 其计算方法为

$$out = (FC \circ B_1 \circ B_2 \circ \cdots \circ B_{N_b})(X) \tag{4}$$

式(4)中, $\circ$ 是神经网络块的连接运算符, N_b 是式(2)中定义的块数, FC 是最终的完全连接层, 每个块定义为

$$B_n = (A_{ReLU} \circ BN \circ C_{1D})(X) \tag{5}$$

式(5)中, A_{ReLU} 表示激活函数, BN 表示批量标准化层, C_{1D} 表示卷积 1D 层, X 表示输入张量.

3 实 验

在实验部分, 目的是验证本文算法对 VM 分类的准确率和实时性, 从而验证本文算法在提高云框架可扩展性方面的可行性.

3.1 实验数据

实验使用的数据集来自 Modena 大学和 Reggio Emilia 机构, 由 8 台真实的云虚拟机组成, 经过几年的监控, 分为 Web 服务器和 SQL 服务器.

3.2 实验设置

在输入数据的预处理中, 将每个通道的均值和单位方差标准化均为零. 由于初始数据流被划分为多个输入序列, 其长度定义为窗口 W , 因此可利用数据增强技术来处理大于 64 个时间步长的序列. 具体而言, 在序列之间应用 75% 的叠加, 这样可以得到更多的序列用于训练和评估.

以每个类中具有相同数量的样本为准则, 将数据集分成 3 个部分, 训练集、验证集和测试集, 分别选

择 0.7,0.2 和 0.1 比例的数据组成数据集.

对于训练阶段,使用交叉熵作为损失函数来评估模型的预测,并在训练阶段反向传播误差.在所有实验中,使用默认配置的 Adam 优化器^[12].在对学习速率和权重衰减超参数进行网格搜索后,可发现在大多数情况下,学习率为 0.000 3,权重衰减为 0.001 2.然而,当观察到 10 个连续时期的验证损失没有减少时,可将学习率降低 0.6 倍.

在每个训练阶段完成后,便在验证阶段(使用数据集的验证部分)评估整个模型的性能.在实验中,每个模型训练 110 次,观察所有模型在这个范围内的收敛情况.在验证阶段中,需使用之前从未使用过的测试数据进行评估,并得出最佳的性能模型.

3.3 实验结果与分析

为了比较本文算法的性能指标,这里选择了当前较优的 Ag 算法^[4]、PCA 集群算法^[9]相比较.由于我们的目标是快速准确地识别虚拟机,因此评估识别精度如何随窗口长度 W 而变化很重要.具体而言,实验中 W 的范围在 4~256 步长之间,这里的步长单位为 5 min,即 W 范围为 20~1 280 min 之间.

首先,对本文算法进行初步评估,即测试本文算法在不同步长情况下的精度,结果如图 2 所示.查看数据,可观察到每个步长的准确度始终高于 98.5%.特别是在短时间窗口的情况下,能够达到这种准确率,意味着本文算法近乎能够实时地对 VM 进行识别分类.另外,本文算法随着窗口增加而降低其性能,可能受到模型复杂度和卷积内核大小固定的影响.

在对本文算法初步评估之后,与 Ag 算法、PCA 集群算法进行比较,其错误率的实验结果如表 1 所示.表 1 给出了对于窗口 W 所考虑的替代方案的误差,窗口 W 的范围为 4~256 个时间步长.

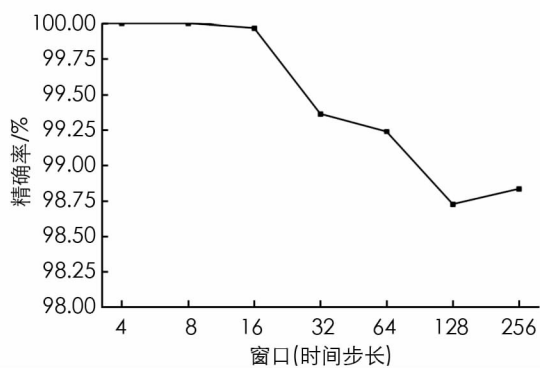


图 2 不同步长情况下的精度

表 1 错误率比较%

步长	4	8	16	32	64	128	256
算法							
PCA	—	17.9	17.5	16.1	15.4	15.2	15.1
Ag	—	1.8	2.8	2.3	1.9	1.8	2.4
灰度区域	—	47.8	41.1	27.8	22.3	19.9	18.7
本文算法	0	0	0.03	0.64	0.76	1.27	1.17

通过观察表 1 可知,对于 PCA 算法,错误率从 17.9%降低到 15.1%.对于 Ag 算法,错误率不是单调的,而是保持在[1.8%,2.4%]范围内.随着窗口增长,我们观察到灰色区域(未分类的虚拟机)明显减少,即由 47.8%降至 18.7%.进一步观察可以发现,本文提出的算法在每种窗口条件下,均表现出了最佳的性能.此外,对于小窗口($W\leqslant 8$),本文算法实现了零错误率,因此本文算法可满足云计算平台的准确性要求.

4 结 语

本文的研究重点是提升云基础架构的可扩展性问题,即通过对表现出相似行为的虚拟机进行分类,从而实现提高监控和管理可扩展性的解决方案.由于现有 VM 群集和分类方法的特点是在准确的 VM 识别和及时响应之间进行权衡,无法同时满足 VM 识别的准确率和实时性.为了解决这个问题,本文基于深度学习技术,提出了一种新颖的 VM 分类算法.通过来自真实云数据中心的数据验证了本文算法的有效性,在 VM 识别率和响应时间方面均具有较好的性能表现.

参考文献:

[1] 谢 兵. 基于移动云计算的计算迁移能效算法 [J]. 计算机应用研究, 2020, 37(10): 3014-3019.

- [2] 齐 平, 王福成, 王必晴, 等. 云计算环境下基于可靠性感知的任务调度算法 [J]. 计算机工程与科学, 2018, 40(11): 1925-1935.
- [3] DELFORGE P, WHITNEY J. Data Center Efficiency Assessment-Scaling up Energy Efficiency Across the Data Center Industry: Evaluating Key Drivers and Barriers [EB/OL]. (2014-08-01)[2020-03-18]. <https://www.nrdc.org/sites/default/files/data-center-efficiency-assessment-IP.pdf>.
- [4] CANALI C, LANCELLOTTI R. AGATE: Adaptive Gray Area-Based TEchnique to Cluster Virtual Machines with Similar Behavior [J]. IEEE Transactions on Cloud Computing, 2019, 7(3): 650-663.
- [5] 何远德, 黄奎峰. 移动云环境面向多重服务选择的计算卸载算法 [J]. 计算机应用研究, 2020, 37(6): 1633-1637, 1651.
- [6] HINZ M, KOSLOVSKI G P, MIERS C C, et al. A Cost Model for IaaS Clouds Based on Virtual Machine Energy Consumption [J]. Journal of Grid Computing, 2018, 16(3): 493-512.
- [7] CANALI C, LANCELLOTTI R. Exploiting Classes of Virtual Machines for Scalable IaaS Cloud Management [C]//2015 IEEE Fourth Symposium on Network Cloud Computing and Applications (NCCA). Munich: IEEE, 2015.
- [8] CANALI C, CHIARAVIGLIO L, LANCELLOTTI R, et al. Joint Minimization of the Energy Costs from Computing, Data Transmission, and Migrations in Cloud Data Centers [J]. IEEE Transactions on Green Communications and Networking, 2018, 2(2): 580-595.
- [9] CANALI C, LANCELLOTTI R. Improving Scalability of Cloud Monitoring through PCA-Based Clustering of Virtual Machines [J]. Journal of Computer Science and Technology, 2014, 29(1): 38-52.
- [10] CANALI C, LANCELLOTTI R. Automatic Virtual Machine Clustering Based on Bhattacharyya Distance for Multi-Cloud Systems [C]//Proceedings of the 2013 International Workshop on Multi-cloud Applications and Federated Clouds-Multi Cloud. New York: ACM Press, 2013.
- [11] CANALI C, LANCELLOTTI R. Exploiting Classes of Virtual Machines for Scalable IaaS Cloud Management [C]//2015 IEEE Fourth Symposium on Network Cloud Computing and Applications (NCCA). Munich: IEEE, 2015.
- [12] HAN Z D. Dyna: a Method of Momentum for Stochastic Optimization [EB/OL]. (2018-05-13)[2020-03-10]. <https://arxiv.org/abs/1805.04933>.

Cloud Computing Virtual Machine Classification Algorithm Based on Deep Learning

DONG Ping^{1,2}

1. College of Applied Engineering, Henan University of Science and Technology, Sanmenxia Henan 472000, China;

2. Information Institute of Media, Sanmenxia Polytechnic, Sanmenxia Henan 472000, China

Abstract: With the increasing scale and complexity of cloud computing data centers, the scalability of their monitoring and management has become a challenging task. Existing literature shows that clustering virtual machines (VM) with similar behaviors in cloud computing platform can improve the scalability of monitoring and management of cloud computing platform. However, in terms of accuracy and timeliness, existing virtual machine clustering technology cannot meet the requirements of cloud computing platform. In order to overcome this problem, a virtual machine classification algorithm based on deep learning is proposed. Virtual machines with similar behaviors are classified by analyzing their behaviors. Then, similar virtual machines are deployed on as few servers as possible to reduce infrastructure costs and improve the scalability of cloud computing data center monitoring and management. And finally, the virtual machine data from real cloud data center are used to compare the proposed algorithm and the similar algorithm. Experimental results show that the proposed algorithm has higher accuracy and better real-time performance, which can meet the requirements of cloud computing platform.

Key words: cloud computing; VM classification; deep learning