

文章编号: 1007-5321(2017)增-0063-05

DOI:10.13190/j.jbupt.2017.s.014

# 虚拟计算环境下基于聚类的资源匹配优化模型

秦童<sup>1</sup>, 孙斌<sup>1</sup>, 朱春鸽<sup>2</sup>, 刘悦<sup>1</sup>, 胡秀妮<sup>1</sup>

(1. 北京邮电大学 信息安全中心, 北京 100876; 2. 国家计算机网络应急技术处理协调中心, 北京 100029)

**摘要:** 虚拟计算环境中任务具有数量庞大、需求模糊、种类多样等特征,使得资源匹配面临巨大挑战。依据虚拟计算实验床平台公布数据,提出了一种融合虚拟资源与任务聚类的资源匹配优化模型。该模型通过分析任务需求、消耗等特征,基于改进二分 $K$ 均值进行任务聚类,并结合虚拟资源类型生成优化的资源匹配列表。经实验分析验证,该模型有效缩小资源匹配范围,提高任务运行成功率,为精准匹配提供基础。

**关键词:** 任务聚类; 二分 $K$ 均值; 资源匹配; 虚拟计算环境

**中图分类号:** TP311.13

**文献标志码:** A

## Resource Matching Optimization Model Based on Clustering under VCE

QIN Tong<sup>1</sup>, SUN Bin<sup>1</sup>, ZHU Chun-ge<sup>2</sup>, LIU Yue<sup>1</sup>, HU Xiu-ni<sup>1</sup>

(1. Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. National Computer Network Emergency Response Technical Team/Coordination Center, Beijing 100029, China)

**Abstract:** Under virtual computing environment(VCE), tasks have features of large quantity, ambiguous requirements, and various types. This makes resource matching face enormous challenges. According to the data published by VCE platform, a resource matching optimization model combined with resource and task clustering was proposed. By analyzing task requirements and consumption characteristics, the model clustered tasks based improved bisect  $K$ -means, and combined with the virtual resource types to generate optimized matching resource list. The experimental analysis verified that the model effectively reduced resource matching range and improved the successful rate of tasks to provide the foundation for precision matching.

**Key words:** task clustering; bisect  $K$ -means; resource matching; virtual computing environment

面向互联网的虚拟计算环境以资源虚拟化为核心,围绕资源的按需聚合和自主协同两个核心机制提供对互联网应用的基础服务支撑<sup>[1]</sup>。如何更好地适应互联网资源的自然特征、满足灵活多变的应用需求、实现资源的可信共享和高效利用是虚拟计算环境亟待解决的重要问题<sup>[2]</sup>。由于虚拟资源具有动态变化、种类多样、高度自治等特

征,海量任务具有需求模糊、类型多样等特征,使得资源匹配面临挑战。学者们大多从资源聚类<sup>[3-5]</sup>、任务偏好<sup>[6-7]</sup>等角度提供有效解决方案。笔者通过模拟虚拟计算环境中资源匹配的应用场景,对虚拟资源与任务进行聚类分析,提出了一种基于聚类的资源匹配优化算法,为精准匹配提供行之有效的解决方案。

收稿日期: 2016-05-18

基金项目: 国家242信息安全计划项目(2015A136)

作者简介: 秦童(1984—),女,博士生, E-mail: qintongbupt@163.com; 孙斌(1967—),女,副教授。

# 1 虚拟资源与任务的聚类模型

## 1.1 虚拟资源定义及类型

所谓虚拟计算环境(VCE, virtual computing environment)是指建立在开放的互联网基础设施之上,以网络资源的按需聚合与自主协同为核心机制,为终端用户或应用系统提供和谐、可信、透明的一体化服务环境,实现有效资源共享和便捷合作<sup>[1]</sup>. 其中,相关定义如下.

**定义1** 虚拟资源(VR, virtual resource)是虚拟计算平台中执行任务的基本单元,虚拟资源集合表示为  $VR = \{VR_1, VR_2, \dots, VR_n\}$ . 其中  $VR_i = (VRID, CA_{VR}, RA_{VR})$ ;  $CA_{VR}$  (configuration attributes of virtual resource)表示虚拟资源的物理配置属性,如物理机类型、CPU、内存、磁盘、网络带宽、地理位置等.  $RA_{VR}$  (running attributes of virtual resource)表示虚拟资源动态运行属性,如CPU、内存、磁盘等实时变化值.

资源匹配的大致步骤为,一是获取应用申请中的任务需求,找到符合任务需求的虚拟资源列表  $L_0$ ;二是针对不同任务类型,分配与之相对应的虚拟资源,生成满足匹配优化目标的列表  $L$ . 文献[3]中通过分析虚拟资源在CPU、内存、磁盘、网络带宽等信息,通过模糊聚类算法将虚拟资源分为3类,即计算型、存储型、带宽型. 其中,计算型代表了CPU类型、CPU利用率及操作系统的特征;存储型具备容量、存储类型及方式等特征;带宽型则突显了接入方式与网络带宽的主要特征.

## 1.2 任务定义及聚类模型

**定义2** 任务. 任务是指当用户期望获得某种应用服务而通过客户端向服务器发出的请求时,请求中包含与应用相关的需求等,通过任务管理层将请求解析为实现该应用而形成的程序指令,用符号  $T(\text{Task})$  加以标识. 虚拟计算环境的任务为元任务,即任务调度器进行任务分配的最小单位,且任务之间相互独立,不考虑任务的跨资源节点执行. 任务描述如下:

$T = \{t_1, t_2, \dots, t_n\}$  表示任务集合,  $n = |T|$  为任务集合的数量,  $t_i (i \in [1, n])$  表示第  $i$  个任务,  $t_i =$

$\{t_{ID}, d_{task}, t_{sev}, t_{status}, t_{data}\}$ , 各属性含义如下:

1)  $t_{ID}$  为任务标识;

2)  $d_{task} = \{d_1, d_2, \dots, d_k\}$ , 用以表示运行任务的资源应须具备的配置属性. 其中,  $d_k$  描述的是运行任务所须满足的第  $k$  个资源配置属性;

3)  $t_{sev}$  表示服务提供商应满足与任务相对应的服务相关属性;

4)  $t_{status}$  表示任务在虚拟计算环境中的状态;

5)  $t_{data}$  表示任务实际消耗虚拟资源的情况.

从任务定义中可看出,影响任务聚类的因素有很多,从运行需求、服务指标、任务状态与任务消耗等特征对任务进行聚类分析. 基于二分  $K$  均值的改进算法对任务特征进行聚类分析,步骤如下:

1) 数据标准化. 对原始数据  $x$  进行标准化处理,消除数据中的量纲,将数据压缩到  $[0, 1]$ .

$$x' = \frac{x - \bar{x}}{s}, \quad x'' = \frac{x' - x'_{\min}}{x'_{\max} - x'_{\min}} \quad (1)$$

2) 设置初始  $K$  值.  $K = 10^n$ , 其中,  $n$  为参数维度. 选取任务消耗参数 CPU 与内存作为维度,即  $n = 2$ ,  $K$  取 100. 为避免分类过细的情况,将内存与 CPU 消耗分为强、中、弱 3 种,除去噪声点,共计 10 种类型,即  $K$  值最小为 10.

3) 取值半径  $r$ . 计算参数在  $[0, 1]$  区间中,以 0.1 为间隔的概率,再得出累积分布函数(CDF, cumulative distribution function)  $F(X)$ . 当分布函数  $F(X) \geq 95\%$  时,对应的值设为  $r$ .

4) 将所有数据点作为一个簇,再选择能够最大程度降低误差平方和(SSE, sum of squared error)的簇划分为 2 个簇,直至簇的数目达到用户设定的  $K$  值.

5) 为避免二分  $K$  均值算法出现聚类过细的情况,这里设置合并参数  $\varphi, \varphi \in (0, 1]$ . 假设  $\alpha$  与  $\beta$  为任意 2 个质心,用  $d(\alpha, \beta)$  表示 2 个质心的距离,  $\bar{d}$  为  $K$  个子簇质心之间距离的平均值. 当 2 个簇之间的距离  $d(\alpha, \beta)$  不大于  $\varphi \times \bar{d}$  时,则 2 个簇合并;反之,则不合并. 可看出,  $\varphi$  值越小,簇的数量越小.

依据任务参数,即任务的内存与 CPU 消耗情况进行聚类研究,聚类分析结果图 3 描述了取值半径  $r$  为 1 和 0.5 时,当合并系数  $\varphi$  从 0.3 增加到 0.62 时,簇类数量减少。随着合并系数的增加,簇类数量也减少,当合并系数为 0.62 时,任务聚类数量为 10。也就是说,当取值半径减小、合并系数增加时,簇类数目也逐渐减少。

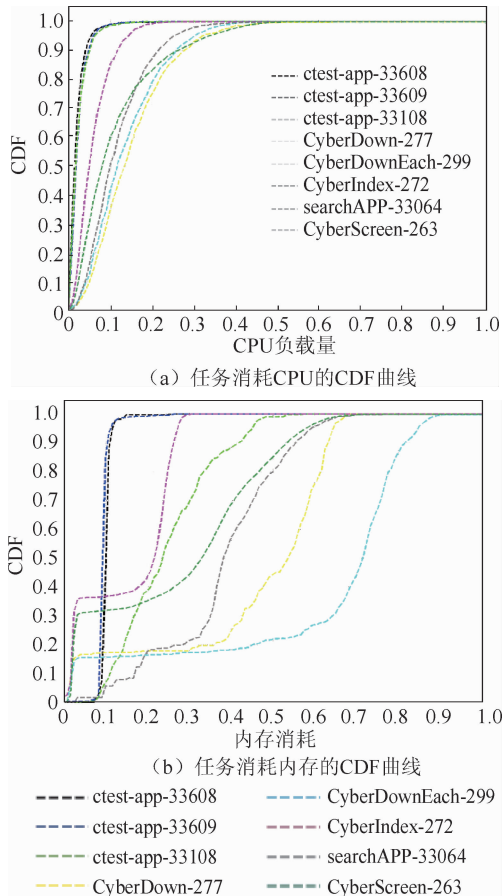


图 2 不同任务消耗 CPU 与内存的累积分布函数曲线

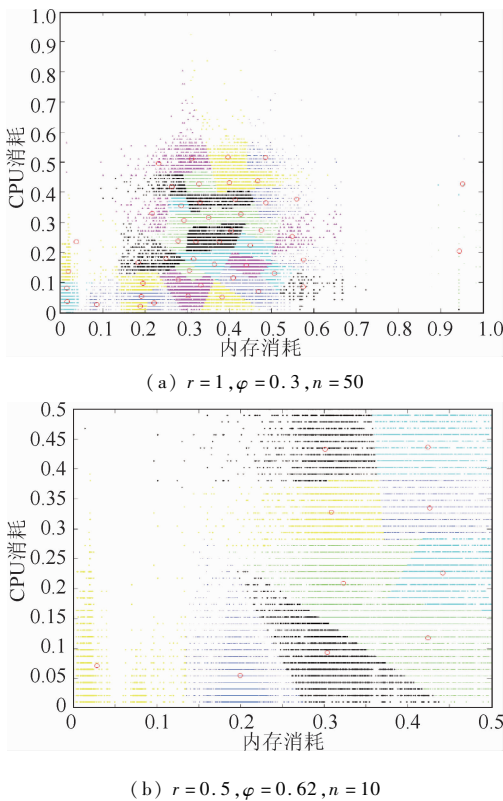


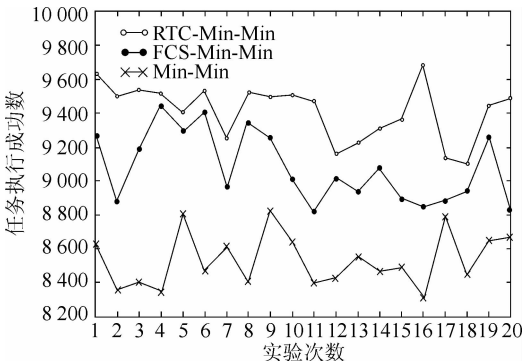
图 3  $r$  与  $\varphi$  不同取值的任务聚类结果

3.2 资源匹配优化

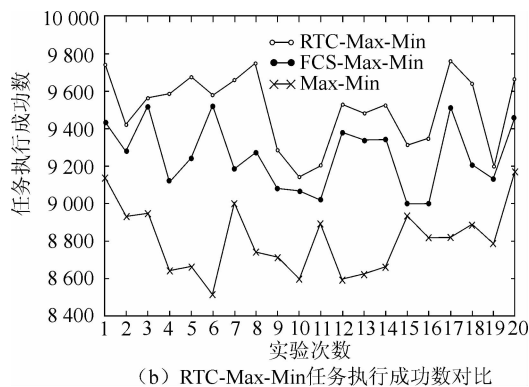
Min-Min 算法是一种实现简单、执行时间快的调度算法,其思想是首先映射小的任务,并且映射到执行快的机器上;Max-Min 算法与之不同的是先调度大任务,任务到资源的映射是选择最早完成时间最大的任务映射到所对应的机器上.这两种经典的任务调度被广泛应用到计算场景中,实验模拟在任务调度不变的前提下,将基于聚类的资源匹配优化模型应用到任务调度中验证模型有效性.不同于以往的资源匹配模型,基于虚拟资源与任务聚类的资源匹配优化模型(RTC-RMOM,resource matching optimization model based on resource and task clustering)进一步筛选初始匹配列表  $L_0$  中符合任务需求的虚拟资源,结合虚拟资源与任务的聚类结果,将任务类型相对应的虚拟资源进行匹配,这样可以缩减精准匹配范围.

实验选取 1 万个任务,经过 20 次验证,将经典任务调度算法 Min-Min、Max-Min,基于资源模糊聚类的调度优化算法 FCS-Min-Min<sup>[3]</sup>、FCS-Max-Min<sup>[3]</sup>与基于虚拟资源与任务聚类的资源匹配优化的任务调度算法 RTC-Min-Min 与 RTC-Max-Min 进行比较.

图 4(a) 的 Min-Min 任务成功率介于 83% ~ 90%,FCS-Min-Min 通过资源模糊聚类分析有效提高任务成功率,而 RTC-Min-Min 在 FCS-Min-Min 基础上,利用虚拟资源与任务双重聚类分析,任务成功率最高为 96.9%,与 FCS-Min-Min 相比,平均提高了近 3.7%.同样地,在图 4(b)中,经由多次实验结果表明,与经典算法 Max-Min 比较,FCS-Max-Min 与 RTC-Max-Min 仍在不同程度上提高了任务执行成功数,其中算法 RTC-Max-Min 较 Max-Min 平均提高了近 6.2%.



(a) RTC-Min-Min 任务执行成功数对比



(b) RTC-Max-Min任务执行成功数对比

图4 基于虚拟资源与任务聚类的任务成功数对比

## 4 结束语

虚拟计算所具有的开放、动态、异构等特性,海量资源的多样性与高度自治性,以及应用需求模糊、数量庞大等诸多不确定因素,使得资源匹配面临巨大挑战,从虚拟资源与任务双重角度进行聚类分析是非常必要而有效的解决方式. 在后续研究中,应建立线上实时分析与线下离线处理相结合的分析手段,挖掘虚拟资源与任务的多维特征,为资源匹配优化提供更精准的理论基础.

### 参考文献:

- [1] Lu Xicheng, Wang Huaimin, Wang Ji. Internet-based virtual computing environment: beyond the data center as a computer[J]. Future Generation Computer System-the International Journal of Grid Computing and Science, 2013, 29(1):309-322.
- [2] Liu Zhenhua, Wierman A, Chen Yuan, et al. Data cen-

ter demand response: avoiding the coincident peak via workload shifting and local generation[J]. Performance Evaluation, 2013, 70(10):770-791.

- [3] 朱春鸽,张哲宇,刘欣然,等. 虚拟计算环境下基于模糊聚类的资源调度算法[J]. 北京邮电大学学报, 2015,38(S1):6-10.  
Zhu Chungu, Zhang Zheyu, Liu Xinran, et al. Research on fuzzy clustering based on algorithm for resource scheduling in virtual computing environment [J]. Journal of Beijing University of Posts and Telecommunications, 2015, 38(S1):6-10.
- [4] 孙佳佳,王兴伟,高程希,等. 云环境下基于神经网络和群搜索优化的资源分配机制[J]. 软件学报, 2014, 25(8):1858-1873.  
Sun Jiajia, Wang Xingwei, Gao Chengxi, et al. Resource allocation scheme based on neural network and group search optimization in cloud environment [J]. Journal of Software, 2014, 25(8):1858-1873.
- [5] Guo Fengyu, Yu Long, Tian Shengwei, et al. A workflow task scheduling algorithm based on the resources' fuzzy clustering in cloud computing environment[J]. International Journal of Communication Systems, 2015, 28(6):1053-1067.
- [6] Hussain H, Malik S U R, Hameed A, et al. A survey on resource allocation in high performance distributed computing systems[J]. Parallel Computing (Systems & Applications), 2013, 39(11):709-736.
- [7] Di S, Kondo D, Cappello F. Characterizing and modeling cloud applications/jobs on a Google data center[J]. Journal of Supercomputing, 2014, 69(1):139-160.