

云计算环境下计算机网络资源智能调度优化研究

季泽安

(阳江职业技术学院, 广东 阳江 529500)

摘要: 云计算环境的动态性与异构性特征使计算机网络资源的智能调度面临着多维度的优化挑战, 资源调度的效率直接影响着云服务的质量、成本与能效。云计算环境下计算资源、存储资源与网络带宽资源的调度涉及资源状态的实时感知、工作负载的精准预测、调度策略的自适应调整以及资源利用率的持续优化等多个环节, 是一个涉及多目标、多约束的复杂优化问题。传统资源调度方法在应对云环境的动态变化时表现出一定的局限性, 智能调度方法通过引入机器学习、深度强化学习以及元启发式优化算法等技术手段, 在资源需求预测、调度决策优化以及资源动态配置等方面展现出优于传统方法的适应性与优化能力。本研究聚焦于云计算环境下计算机网络资源的智能调度优化, 旨在系统分析云计算资源调度的架构与挑战、智能调度方法的技术体系、关键优化目标及其约束条件, 以及智能调度系统的实现路径与优化策略。

关键词: 云计算; 资源调度; 智能优化; 负载预测; 深度强化学习

中图分类号: TP393

文献标识码: A

文章编号: 3106-2709 (2026) 08-0013-05

DOI: 10.62022/NCAR.issn3106-2709.2026.08.003

Research on Intelligent Scheduling Optimization of Computer Network Resources in Cloud Computing Environments

Ji Ze'an

(Yangjiang Polytechnic, Yangjiang, Guangdong 529500)

Abstract: The dynamic and heterogeneous characteristics of cloud computing environments pose multidimensional optimization challenges to the intelligent scheduling of computer network resources. The efficiency of resource scheduling directly affects the quality, cost, and energy efficiency of cloud services. The scheduling of computing resources, storage resources, and network bandwidth resources in cloud computing environments involves multiple aspects, including real-time perception of resource states, accurate prediction of workloads, adaptive adjustment of scheduling strategies, and continuous optimization of resource utilization. It is a complex optimization problem involving multiple objectives and constraints. Traditional resource scheduling methods exhibit certain limitations in responding to dynamic changes in cloud environments. By introducing technologies such as machine learning, deep reinforcement learning, and metaheuristic optimization algorithms, intelligent scheduling methods demonstrate adaptability and optimization capabilities superior to those of traditional methods in resource demand prediction, scheduling decision optimization, and dynamic resource allocation. This study focuses on intelligent scheduling optimization of computer network resources in cloud computing environments, aiming to systematically analyze the architecture and challenges of cloud computing resource scheduling, the technical system of intelligent scheduling methods, key optimization objectives and their constraints, and the implementation paths and optimization strategies of intelligent scheduling systems.

Keywords: cloud computing; resource scheduling; intelligent optimization; load prediction; deep reinforcement learning

1 云计算环境下计算机网络资源调度的架构与挑战

1.1 云计算资源调度的基本架构与服务模型

云计算环境下的计算机网络资源调度建立在分层化的服务架构之上, 其基本架构涵盖了基础设施层、平台层与应用层三个主要层次。基础设施层 (IaaS) 提供虚拟化的计算资源、存储资源与网络资源, 用户可以根据需求动态申请与释放资源, 按使用量付费。基础设施层的资源调度涉

及物理服务器的资源分配、虚拟机的创建与迁移、存储空间分配以及网络带宽的预留等操作。基础设施层资源调度的目标是实现物理资源的高效利用与虚拟资源的按需供给, 其决策变量包括虚拟机的规格 (CPU核数、内存大小、存储容量)、部署位置 (物理服务器选择) 以及资源分配的时间周期。

平台层 (PaaS) 在基础设施层之上提供应用开发与部署的运行环境, 其资源调度关注应用实例的资源分配、服

作者简介: 季泽安, 硕士, 助教, 研究方向为云计算技术、网络资源调度。

务实例的动态扩缩容以及中间件服务的资源保障。平台层的资源调度需要考虑应用特征与服务等级协议 (SLA)，不同应用对资源的消耗模式与性能要求存在差异，调度决策需要根据应用的负载特征来进行资源的合理分配。平台层的资源调度还涉及服务发现、负载均衡以及故障恢复等功能的资源保障，这些功能与资源的可用性和调度策略之间存在紧密的耦合关系。

应用层 (SaaS) 面向终端用户提供软件服务，其资源调度关注多租户环境下的资源隔离与服务质量保障。应用层的资源调度决策直接影响用户体验，需要根据用户数量、请求频率以及响应时间要求等因素来动态调整资源分配。资源调度的精细化程度越高，资源浪费越少，但决策的复杂度也随之增加，在优化中需持续调整精细化的边界。

云计算的三种服务模型对资源调度提出了差异化的需求，IaaS层关注资源分配的灵活性与隔离性，PaaS层关注应用运行环境的稳定性与弹性，SaaS层关注多租户服务的性能与成本效益，各层的调度策略需要协同配合以实现云服务质量的整体优化。

1.2 云计算环境下的资源异构性与动态性特征

云计算环境的资源异构性体现在硬件平台、操作系统、虚拟化技术以及网络架构等多个层面。硬件平台的异构性包括不同型号的CPU (不同指令集架构、不同核心数与频率)、不同类型的存储设备 (SSD、HDD、NVMe等) 以及不同规格的内存与网络接口。虚拟化技术的异构性包括全虚拟化、半虚拟化与容器虚拟化等不同技术路线，以及VMware、KVM、Xen、Docker等不同实现方案。网络架构的异构性包括不同拓扑结构的数据中心网络、不同带宽与延迟的链路以及不同配置的网络设备。

云计算环境的动态性特征体现在工作负载的时间变化性、资源状态的实时波动性以及系统规模的动态伸缩性三个方面。工作负载的时间变化性表现为不同类型的工作负载在小时、天、周以及季节等不同时间尺度上的周期性变化，电子商务应用在购物节期间的负载可能达到平日水平的数倍至数十倍，数据分析任务在特定时间段可能产生明显的资源消耗高峰。资源状态的实时波动性表现为物理服务器的负载水平、网络链路的带宽利用率以及存储设备的I/O响应时间等指标随时间不断变化，这些变化受工作负载模式与资源竞争状态的影响，呈现出一定的规律性与随机性。

系统规模的动态伸缩性表现为云环境中资源的动态加入与退出，物理服务器的故障维护、新节点的动态加入以及虚拟机的创建与销毁都在持续发生，资源池的规模与构

成处于动态变化之中。资源异构性增加了调度决策中资源选择的维度，动态性则要求调度策略能够适应资源的实时变化，两者的叠加使资源调度问题的复杂度明显增加。

1.3 传统资源调度方法的局限性分析与智能调度需求

传统资源调度方法主要包括基于规则的方法、基于启发式的方法以及基于简单优化模型的方法，这些方法在应对云计算环境的动态性与异构性时存在一定的局限性。基于规则的调度方法 (如轮询、最少连接、加权轮询等) 实现简单、计算开销低，但规则一旦设定便难以适应工作负载的变化，在负载模式发生变化时需要人工调整规则，响应速度与灵活性存在优化空间。基于启发式的方法 (如最短作业优先、最佳适配、首次适配等) 在特定场景下可以获得较优的调度效果，但其性能依赖于启发式规则与场景特征的匹配程度，在资源类型多样、约束条件复杂的云环境中，简单启发式方法容易陷入局部最优。

基于传统优化模型的调度方法 (如整数规划、线性规划等) 可以形式化地描述资源调度问题并求解最优解，但随着问题规模的增大 (虚拟机数量、物理服务器数量、约束条件数量的增加)，求解时间急剧增长，难以满足云环境中实时调度的需求。传统优化方法通常假设系统参数是确定且已知的，而云环境的动态性与不确定性使这一假设难以成立。

传统方法面对云环境动态变化时的适应性限制推动了对智能调度方法的研究需求，智能调度方法通过机器学习来预测工作负载的变化趋势，通过深度强化学习来学习最优调度策略，通过元启发式算法来搜索大规模解空间中的优解，这些方法在适应动态环境、处理复杂约束以及实现多目标优化等方面展现出优于传统方法的特性，可以弥补传统方法在动态性与可扩展性方面的不足。

2 智能调度方法的技术体系与优化目标

2.1 基于机器学习的工作负载预测与资源需求估计

工作负载预测是智能资源调度的基础环节，准确的负载预测可以为资源分配的预决策提供依据，减少因资源供给滞后导致的性能下降或资源浪费。时间序列分析方法 (如ARIMA、SARIMA等) 利用负载数据的历史规律来预测未来趋势，适用于负载模式具有明显周期性的应用场景。时间序列模型的计算效率较高，但难以捕捉负载序列中的非线性关系，对于负载模式变化较为复杂的应用，预测准确率受到一定的限制。

机器学习方法 (如支持向量回归、随机森林回归、梯

度提升决策树等)可以通过学习负载数据与时间、事件以及其他相关变量之间的复杂关系来进行预测。机器学习模型的预测精度通常高于传统时间序列模型,但需要较多的历史数据进行训练,在负载模式发生突变时,模型的预测能力需要一定的适应时间才能恢复。深度学习模型(如长短期记忆网络(LSTM)、门控循环单元(GRU)、Transformer等)在处理时间序列数据方面具有独特的优势,可以自动学习负载数据中的长期依赖关系与复杂模式,适用于负载模式复杂、波动性大的应用场景。

资源需求估计是在负载预测的基础上将预测的负载水平转换为具体的资源需求规格,涉及CPU、内存、存储与网络带宽等多种资源类型的需求量化。资源需求估计需要考虑不同应用类型对各类资源的消耗特征以及资源之间的相互影响。CPU密集型应用的计算需求与内存访问模式密切相关,内存密集型应用对带宽的需求与内存容量成一定的比例关系,资源需求估计的准确性影响调度决策中资源分配的精确性。

2.2 深度强化学习在资源调度决策中的应用

深度强化学习为资源调度决策提供了一种从环境交互中学习最优策略的框架,其通过智能体与云环境的持续交互来学习在不同状态下采取何种行动可以获得最大的累积奖励。在资源调度问题中,智能体的状态可以定义为当前资源池的使用情况、工作负载的特征以及服务质量指标等;行动可以定义为虚拟机的创建、迁移、销毁以及资源分配的调整等操作;奖励函数可以设计为资源利用率的提高、能耗的降低、服务等级协议违反率的减少以及调度决策时间的缩短等优化目标的加权组合。

深度Q网络(DQN)是深度强化学习在离散行动空间中的经典算法,其通过深度神经网络来近似Q值函数,在高维状态空间中实现对不同行动价值的估计。在资源调度中,DQN可以用于虚拟机放置决策、任务分配决策等离散调度问题。DQN的训练需要大量与环境交互的数据,训练过程的收敛速度受网络结构、经验回放策略以及探索-利用平衡策略的影响。深度确定性策略梯度(DDPG)与近端策略优化(PPO)等算法适用于连续行动空间的资源调度问题,可以输出连续的资源分配比例、频率调整幅度等决策变量。

深度强化学习在资源调度中的应用价值在于其可以自适应地学习云环境的动态变化,当工作负载模式或系统状态发生变化时,智能体可以通过持续学习来更新其策略,无需人工重新设计调度规则。深度强化学习调度策略的一个局限性在于其决策过程的不可解释性,在需要明确决策

依据的运维场景中,这一点可能限制其应用范围。

2.3 智能调度优化的多目标权衡与约束处理

云计算资源调度优化通常涉及多个相互冲突的优化目标,需要在目标之间进行合理的权衡。常见的优化目标包括资源利用率的优化(提高CPU、内存、存储与网络带宽的利用率,减少资源闲置)、能耗的优化(通过服务器整合与负载迁移来减少活跃服务器数量,降低数据中心能耗)、服务质量的优化(确保应用的响应时间、吞吐量等性能指标满足服务等级协议要求)以及调度成本的优化(减少资源使用的经济成本)。四个目标之间存在不同程度的冲突关系,提高资源利用率与保障服务质量在配置低时可能相互制约,降低能耗与优化调度决策时间也存在一定的权衡。

多目标优化问题的求解需要在帕累托最优前沿上寻找合适的折中解,帕累托最优解是指在不使任何一个优化目标变差的情况下无法使另一个目标更好的解。多目标进化算法(如NSGA-II、MOEA/D等)可以在一轮优化中获得一组帕累托最优解,供决策者根据偏好选择。多目标优化方法的计算复杂度随着目标数量的增加而显著提高,目标数量超过3个时,帕累托前沿的表示与决策的难度都会明显增加,通常需要先进行目标相关性分析,减少冗余目标后再进行求解。

约束处理是资源调度优化中需要关注的方面,资源调度问题受到硬件资源总量、服务等级协议、安全隔离要求以及负载均衡等多重约束。约束可以分为硬约束(必须满足的约束,如资源分配不能超过可用总量)与软约束(尽可能满足的约束,如负载均衡的目标)两类。智能调度算法需要通过设计合适的约束处理机制来保证调度方案在硬约束下的可行性,并在软约束下寻求优化。

3 智能调度优化的核心算法与实现技术

3.1 基于深度强化学习的动态资源分配算法

基于深度强化学习的动态资源分配算法将资源分配问题建模为马尔可夫决策过程(MDP),通过智能体与云环境的交互来学习最优的资源分配策略。状态空间的设计需要全面反映资源池的状态、工作负载的特征以及服务质量指标,状态向量的维度过高可能导致学习效率下降,维度过低则可能遗漏关键信息,需要通过特征选择与降维技术在表征完整性与计算效率之间取得平衡。行动空间的定义需要根据资源分配的具体场景来确定,可以定义为虚拟机规格的选择、资源配额的比例调整以及实例数量的增减等。

奖励函数的设计对算法的学习方向具有决定性的影响,合理的奖励函数可以引导智能体学习到符合优化目标的调度策略。奖励函数的设计需要综合考虑多个优化目标的贡献,资源利用率提高时给予正奖励,服务等级协议违反时给予负奖励,能耗降低时给予正奖励。奖励函数的权重设置需要反映优化目标的优先级,权重的调整会影响学习到的策略的偏好方向。

深度强化学习算法的选择需要根据资源分配问题的具体特征来确定,离散行动空间适合采用DQN或其改进算法(如Double DQN、Dueling DQN等),连续行动空间适合采用DDPG、TD3或SAC等算法。算法的超参数(如学习率、折扣因子、探索率等)对训练效果有显著影响,需要经过系统的调优来确定。训练过程的稳定性与收敛速度可以借助经验回放、目标网络以及熵正则化等技术来改善,其中经验回放机制对打破连续样本间的相关性、稳定训练过程有直接的帮助。

3.2 基于元启发式优化的资源调度算法

元启发式优化算法通过模拟自然现象或生物行为来搜索调度问题的最优解或近优解,在处理大规模、多约束的资源调度问题时具有一定的应用价值。遗传算法通过模拟自然选择与遗传变异的过程来在解空间中搜索最优解,其将每个调度方案编码为染色体,通过选择、交叉与变异等操作来迭代优化种群。遗传算法适用于解空间较大、目标函数不连续或不可微的调度问题,其全局搜索能力较强,但收敛速度较慢,在大规模问题中计算开销较高。

粒子群优化算法模拟鸟群觅食行为中的信息共享与协作机制,每个粒子代表一个调度方案,通过跟踪个体最优位置与群体最优位置来更新粒子的速度与位置,逐步向最优解区域聚集。粒子群优化算法的收敛速度快于遗传算法,参数设置较少且实现简单,但容易陷入局部最优,需要通过惯性权重与收缩因子的适当设置以及多种群策略来增强全局搜索能力。模拟退火算法通过模拟固体物质退火过程中的热力学平衡来在解空间中搜索,其以一定概率接受劣化解的机制使其具备跳出局部最优陷阱的能力,适用于求解质量要求较高但时间约束较宽松的调度问题^[1]。

元启发式算法在资源调度中的应用优势在于其可以灵活地处理多目标与多约束问题,通过适应度函数的设计来整合多个优化目标与约束条件。适应度函数的设计需要将多目标转化为单目标或采用多目标适应度评估方法,转化过程中各项目标的权重设定会影响算法的搜索方向。元启发式算法与局部搜索方法的混合(如Memetic算法)可以结

合全局搜索与局部搜索的优势,在较短时间内获得质量较高的解。

3.3 多智能体协同调度架构与分布式实现

多智能体协同调度架构将资源调度的决策过程分布到多个智能体中,每个智能体负责一部分资源的调度决策,通过智能体之间的信息交互与协同来实现全局优化。多智能体架构的优势在于其可扩展性强、容错性好,可以适应大规模云环境的分布式部署需求。智能体的划分可以根据资源类型(计算智能体、存储智能体、网络智能体)、服务区域(不同可用区的调度智能体)或功能层次(全局调度智能体、局部调度智能体)等维度进行。

多智能体之间的协同机制包括集中式协同、分布式协同与混合式协同三种模式。集中式协同中一个全局协调智能体收集各智能体的决策信息并做出全局优化决策,适合中小规模系统但对全局协调智能体的性能与可靠性有较高要求。分布式协同中各智能体通过相互通信与协商来达成一致的调度决策,适合大规模分布式系统但通信开销与协商算法的设计复杂度较高。混合式协同结合集中式与分布式的优点,在局部范围内采用分布式决策,在全局范围内由协调节点进行统筹,适用于多层次的云环境调度需求。

多智能体协同调度在分布式实现中需要关注通信开销与决策一致性的权衡,通信频率的提高可以改善信息共享的效果但会增加系统开销,决策一致性要求各智能体的局部决策在全局层面上不产生冲突。多智能体强化学习(如QMIX、VDN等)可以学习多个智能体的协同策略,通过联合奖励函数的设计来促进智能体之间的协作,在多智能体调度场景中具有一定的应用潜力。

4 智能调度优化的评价体系与发展趋势

4.1 智能调度算法的性能评价指标与测试方法

智能调度算法的性能评价需要建立多维度、多层次的指标体系,以全面反映算法在不同场景下的表现。资源利用率指标关注CPU利用率、内存利用率、存储利用率以及网络带宽利用率的平均水平与分布均衡性,资源利用率的提高意味着资源浪费的减少,但过高的利用率可能导致服务质量下降。调度决策时间指标关注算法做出调度决策所需的时间,决策时间直接影响到调度策略的响应速度与适用场景的实时性要求,决策时间较长的算法更适合离线调度场景,决策时间较短的算法更适合在线实时调度场景。

服务质量指标关注应用的平均响应时间、请求的吞吐量以及服务等级协议违反率等,服务等级协议违反率是衡

量调度算法保障能力的重要指标,违反率越低表明调度算法对服务质量保障越有效。能耗指标关注数据中心的总能耗与单位计算量的能耗,能耗的降低对运营成本控制与环境保护均具有实际意义。调度成本指标关注资源使用的经济成本。

智能调度算法的测试方法包括仿真测试与真实环境测试两种途径。仿真测试通过构建云环境的仿真模型来评估算法性能,可以灵活地控制实验条件与重复实验,适合算法开发阶段的性能验证与参数调优。仿真模型需要合理抽象云环境的关键特征,模型与真实环境的差距会影响测试结论的适用性。真实环境测试在云平台部署算法并在真实工作负载下进行评估,测试结果具有较高的可信度,但实验成本较高且难以控制干扰变量。

4.2 智能调度系统的可解释性与可信性提升

智能调度系统的可解释性是其在生产环境中推广应用的重要条件,运维人员需要理解调度决策的依据以便在异常情况下进行人工干预。深度学习模型与强化学习模型的决策过程具有较为复杂的非线性特征,直接的模型解释存在一定的技术挑战。可解释性提升的方法包括对模型决策过程的可视化、对关键决策因素的识别与展示以及决策依据的自然语言描述等。

特征重要性分析可以识别对调度决策影响最大的输入特征,帮助运维人员理解模型关注的关键因素。注意力机制的可视化可以展示模型在做决策时对不同输入部分的关注程度,提供决策过程的直观解释。反事实解释通过分析输入特征的微小变化对决策结果的影响来揭示决策的边界条件与触发机制。

可信性提升关注调度系统在异常情况下的行为可靠性,包括对输入数据噪声的鲁棒性、对工作负载突变的适应性以及系统故障时的降级运行能力等。鲁棒性测试通过向输入数据中添加不同程度的噪声来评估模型性能的退化程度,帮助识别模型的脆弱点并采取针对性的加固措施。降级运行机制在系统部分功能异常时,能够通过人工规则接管或简化模型来维持基本调度功能,避免因系统完全失效导致的资源管理中断。

4.3 智能化资源调度的未来发展方向

智能化资源调度的未来发展方向将在异构计算资源调度、联邦学习辅助调度以及云-边-端协同调度等方向上持续深化。异构计算资源的调度需要同时管理CPU、GPU、FPGA、TPU等多种计算资源,不同计算资源的性能特征、

能耗特征与成本特征存在显著差异,调度决策需要根据任务的计算特征来选择最适宜的计算资源。异构资源调度还需处理资源之间的数据迁移与协同计算问题,跨资源类型的数据传输延迟对调度策略的效率有直接影响^[2]。

联邦学习辅助调度在保护数据隐私的前提下,允许多个云数据中心在不共享原始数据的情况下联合训练调度模型。各数据中心在本地数据上训练模型,将模型参数而非原始数据上传至中心服务器进行聚合,减少了数据传输的隐私风险。联邦学习框架中的通信效率与模型聚合策略直接影响联合训练的效果,不同数据中心的数据分布差异也对聚合后模型的适用性产生影响^[3]。

云-边-端协同调度将云计算、边缘计算与终端设备的计算资源进行统一调度,形成覆盖从端到云的计算连续体。协同调度的核心是任务的分割与分配,根据任务的延迟敏感性、计算需求以及数据位置来决定在端侧、边侧还是云侧执行。云-边-端协同调度需要处理网络连接的动态变化与计算资源的异构性,调度策略需要适应边缘节点与终端设备的资源限制,并考虑数据传输的成本与延迟。

5 结论

云计算环境下计算机网络资源的智能调度优化在基于机器学习的工作负载预测与资源需求估计、深度强化学习驱动的动态资源分配决策、元启发式算法支持的多目标优化求解以及多智能体协同调度架构的实现等方面取得了系统的研究进展,智能调度方法通过预测、学习与优化相结合的技术路径,在资源利用率提升、服务质量保障以及能耗控制等优化目标上展现出优于传统方法的适应性与可扩展性。云计算资源调度相关技术研发与运维主体在今后的工作中应当着力推进智能调度算法的可解释性与可信性提升、异构计算资源与云-边-端协同调度场景下的智能调度方法扩展以及智能调度系统性能评价体系的标准化与动态化完善,使智能化资源调度在云计算环境持续演进的背景下保持其适应性与优化效能,为云计算服务的高效、可靠与可持续运行提供更加坚实的技术支撑。

参考文献:

- [1] 郑钦润, 凌荣国. 计算机网络技术中的人工智能应用研究[J]. 人工智能研究, 2025, 2(3): 38-40.
- [2] 谢晓锋. 基于云环境的大数据计算机处理技术分析[J]. 教育教学研究前沿, 2025, 3(9): 65-67.
- [3] 郭丽娇, 王庆生. 云环境下的虚拟资源调度智能优化策略[J]. 计算机应用与软件, 2015, 32(6): 5.