

人工智能驱动的云计算资源智能能耗管理研究

薛国东¹ 陈向荣² 何珩³ 凌佳生³

1 金投健康（杭州）科技有限公司，浙江杭州，310014；

2 浙江移动有限公司，浙江杭州，310014；

3 浙江移动信息产业集成有限公司，浙江杭州，310014；

摘要：随着云计算基础设施在全球范围内的迅猛扩展，数据中心能耗问题日益凸显，成为制约其可持续发展的重要因素。传统能耗管理手段多依赖静态配置与人工干预，难以应对复杂多变的资源负载和动态调度需求。人工智能技术凭借其强大的预测能力、优化能力和自适应特性，为云计算资源的智能能耗管理提供了全新路径。本文从云计算能耗现状入手，分析能耗管理中的关键挑战，系统阐述人工智能技术在能耗感知、智能调度、节能优化等方面的应用实践与创新模式，并探讨构建绿色智能云计算系统的发展趋势。研究表明，融合深度学习、强化学习、多目标优化等技术的智能管理系统，已成为提升资源利用效率与降低碳排放的核心驱动力。

关键词：人工智能；云计算；能耗管理；智能调度；绿色计算；数据中心

DOI：10.69979/3029-2727.25.08.031

引言

云计算作为信息技术发展的重要成果，推动了社会各行业的信息化转型。随着业务需求的爆炸式增长，海量数据的存储与处理依赖于高密度部署的数据中心。当前，全球数据中心的能耗已占据整个 ICT 能耗的显著比例，不仅造成运维成本上升，也带来了严重的环境影响。据统计，某些超大型云服务运营商的数据中心年耗电量已接近中等城市用电水平。尽管近年来在硬件能效、冷却系统、虚拟化调度等方面取得了不少成果，但依赖静态策略的能耗管理方式仍难以适应实时性高、波动性强的资源使用场景。

人工智能技术的发展为云计算资源管理带来了范式转变。基于 AI 的系统能够实时感知系统运行状态，预测资源负载变化趋势，动态调整计算资源与能耗策略，实现节能与性能之间的平衡。AI 算法在任务调度、资源预测、冷却控制、能效评估等方面的应用不断深入，为实现“绿色云计算”提供了坚实技术基础。本文从能耗挑战、AI 驱动技术路径和未来趋势三个维度，全面探讨云计算在能耗智能管理方面的研究现状与发展前景。

1 云计算能耗管理的现状与挑战

1.1 高密度资源部署导致能耗激增

随着数字经济的加速发展，云计算平台在处理能力、服务规模及数据吞吐量方面的需求持续上升，促使数据中心向高密度部署演进。现代数据中心普遍配置大规模 CPU 集群、GPU 服务器、FPGA 加速器及高速存储设备，

以支持深度学习、大数据分析、实时计算等复杂业务场景。然而，虽然硬件设备的单体能效有所提升，但整体能耗却呈指数级增长，尤其是在单位面积内部署密集设备后，功耗密度极高，致使散热压力剧增。

由于服务的高可用性要求，大量服务器即使在业务低谷时也保持运行状态，这种“负载低、能耗高”的非线性矛盾导致资源利用效率低下，形成大量能源浪费。更为严重的是，部分云平台为应对流量突发，在资源冗余配置上采用“过度保守”的策略，进一步推高了闲置设备的电力消耗。在这种背景下，传统以物理扩容为主的建设思路已难以为继，数据中心必须从基础架构规划、部署密度控制、动态调度机制等多方面着手，探索更科学的能耗管理体系，以实现资源性能与能源效率的双重提升。

1.2 能耗管理策略缺乏动态适应性

当前大多数数据中心仍采用静态阈值控制和手工调度策略进行能耗管理，这种模式在面对云计算平台高并发、高弹性、多租户等动态复杂环境时显得力不从心。由于传统方法大多依赖历史经验设定规则参数，在实际运行中难以及时响应资源需求的剧烈波动，尤其在业务高峰期或突发任务集中时，原定能耗控制机制往往失效，甚至可能因超负载运行引发硬件异常或系统宕机。

而在平峰期，静态调度策略又无法有效识别低效运行节点，导致部分设备长时间处于“待机能耗”状态，造成能源浪费。此外，在面对诸如虚拟机的动态迁移、冷热数据的弹性分布、临时任务的抢占调度等关键操作

时,能耗管理系统若无法智能预测和联动响应,不仅难以实现节能目标,反而可能因频繁调度引发资源冗余与负载抖动,引起系统整体能效的下降。因此,要实现真正意义上的节能优化,必须摒弃传统的静态管理方式,引入具备实时感知、智能预测与自适应控制能力的策略系统,以匹配现代云环境对灵活性与精准性的双重需求。

1.3 冷却系统能耗成为关键瓶颈

在数据中心总能耗结构中,冷却系统长期占据着极高比例,其能耗甚至可能超过核心计算设备本身,成为制约整体能效提升的主要瓶颈之一。传统空调系统普遍采用全场统一冷却、固定风量输送的策略,缺乏对实时热负载变化的感知和响应能力,导致部分区域过冷、而热点区域得不到有效散热的现象频繁发生。这种冷热分布失衡不仅导致能耗过高,还可能引发硬件故障率上升、设备寿命缩短等隐患。

尽管近年数据中心逐步引入液冷、自然冷却、冷通道封闭等新型技术方案,提升了冷却效率,但大多数冷却系统仍以手动调节和规则设定为主,缺乏数据驱动和智能调控机制。尤其在多区域数据中心、季节变换或设备状态多变的环境中,单一冷却参数很难兼顾局部与整体的热需求,从而使得冷却系统整体运行在“冗余功耗”状态。构建能够基于热力模型、环境参数、服务器运行状态进行实时计算与动态控制的冷却智能调度系统,是破解能效瓶颈的关键路径之一。唯有实现“按需制冷、分布调节、智能响应”的目标,才能彻底摆脱冷却系统能耗持续高企的困局,推动数据中心朝着绿色化、低碳化方向转型。

2 人工智能赋能的能耗智能感知与调度技术

2.1 基于深度学习的能耗预测模型

深度学习技术用在云计算能耗预测上,已经从理论的研究转到了实际的工程应用,变成了能耗感知系统中的重要核心部分。以前的预测方法因为模型对复杂非线性关系抓取能力不够,所以只能给出比较粗略和滞后的结果,很难适应实时动态调度的需要。而深度学习建立的预测模型可以通过大量历史数据自动提取特征,发现服务器功耗、网络传输、冷却温度等不同因素之间的相互联系,实现更精确的负载和能耗预测。比如用 CNN 模型来分析数据中心内部的热力分布图,就能很快找到热点区域和制冷有问题的地方,提前发出过热警告;要是用 RNN 或者 LSTM 网络来对任务负载的变化趋势建立模型,就能更好预测未来的功耗变化情况,给任务转移和资源分配提供参考依据。更先进的端到端模型还能直接把能耗、负载、设备状态等各种数据输入到一个统一

的神经网络里,输出实时预测结果,加上注意力机制来优化计算资源的分配顺序。深度学习的应用不仅让预测准确度和响应速度有很大提高,还为智能化的能耗调度和节能控制提供了可靠的数据基础和感知能力的支持。

2.2 强化学习驱动的自适应资源调度

强化学习最大的好处是有个“学习-反馈-优化”这样循环的过程,特别适合用在云计算平台里资源分配这种常常变化又有不确定的情形。在省电优化的场景里,系统要处理很多状态、连续的动作和多个目标的回报函数组成的复杂情况,老办法很难在短时间内找到最好的解决办法。但是强化学习让智能体通过不断和环境互动,不用精确的数学模型就能自己调整策略。具体来说,把虚拟机、任务列表、CPU 使用率、周围温度这些数据当作输入状态,把调度行为比如移动虚拟机、调整负载平衡、让服务器睡觉或者启动当作动作输出,然后通过降低能耗和保证性能这两方面作为奖励机制,训练出能应对不同工作量的策略网络。当突然流量变大或者某些服务器出问题的时候,强化学习可以靠着之前的经验快速给出调度办法,防止用电量突然增加或者资源堵车。还有现在深度强化学习(DRL)里厉害算法例如 DQN、PPO 这些,更能应对复杂的策略空间和大型任务队列,让处理复杂环境的能力变得更好。用了强化学习之后,能源调度不需要再按照固定规则来,而是有了“实时监控—自动调整—更新策略”这样的智能决策流程,彻底推进了能源管理智能化的进展。

2.3 多目标优化算法提升能效平衡

在实际的云平台运行过程中,能耗并不能单独考虑,还必须考虑例如系统运行的性能、服务质量好坏、响应时间长短、设备寿命多久等因素,所以单一目标的优化方法就显得不够全面。通过多目标优化算法的话,需要建立一个包含多个维度的收益函数模型,然后找出帕累托最优解的集合,这样就可以在多个目标之间找到平衡点。比如例如遗传算法就是模仿生物进化的过程,通过筛选、交换和突变这些步骤,让调度策略像生物种群那样进化,从而更快找到最优的能耗调度方案;而粒子群优化算法是模仿群体智能来找解空间中找到既能省电又能保证性能的妥协方案;至于蚁群算法的话就比较适合用来在任务分配的网络上找那些能耗低的路径,让资源分布的拓扑结构更合理。

最近这几年把人工智能和多目标优化结合起来成了热门方向,比如比如用深度强化学习去帮助遗传算法做策略的引导,或者用图神经网络让粒子群算法在初始化阶段做得更好,这样算法就能更快算出结果并且更好

找到局部最优解。具体应用时可以拿这些算法去处理比如比如控制风扇转多快、怎么划分冷却区域、优化服务器启动的策略这些重要问题,通过实时调整让系统既不冷又不超负荷运转。另外多目标优化还能预测未来可能要做的任务,配合 AI 预测结果提前调整资源分配和功耗布局,这样系统就能从被动处理变成主动规划,整体的运营效率和节能效果都能有很大提高。

3 构建面向未来的绿色智能云计算架构

3.1 云边协同的能耗分布式管理机制

云边端一体化架构的发展让能耗管理变得更加灵活和可控了,使得能耗优化从过去那种集中式的调控方式扩展成了多节点、分布式的共同管理。在这种结构里,边缘节点不再只是简单的传输工具,而是变成了有自己算力和自主能力的智能管理模块,能够自己进行任务分配、资源转换还有感知耗电这些操作。AI 在这里的作用特别重要,它能根据网络延迟、计算量、节点状况这些信息即时做出调配决定,比如在访问高峰时候把实时分析任务分配到用户附近边缘节点,这样可以避免远距离传数据导致耗电太多,同时在低负载时把资源集中到能效更高的核心云节点处理,减少边缘节点待机的电量消耗。

另外 AI 还能结合地理环境、电力价格、天气状况这些多维度的数据,制定跨区域的能耗调配方案,让绿色能源和任务执行智能配合起来。比方说太阳能发电充足的地方可以优先运行高计算任务,遇到雨雪天气就把本地负荷减少转移到供电稳定的地区,真正把能源感知结合到计算调度里面。随着 5G 和物联网技术的支持,云边协作体系会变得更加细化,最终形成以 AI 为核心的多个层次、低碳环保、高效节能的能源智能管理网络。

3.2 基于数字孪生的能耗虚拟仿真平台

数字孪生通过在虚拟世界里构建真实对应的虚拟模型,给云计算平台的节能改造提供了完整的建模和可视化工具。对能耗管理来说,这个技术不仅能展示现在系统的运转情况,还能用模拟测试预估后面系统会产生多少耗能。比如当要增加新服务器或者调整冷却方案的时候,管理员可以先在数字孪生系统里运行看看对耗能和性能有什么影响,然后比较不同参数的影响,就不用靠实际测试来浪费能源了。更厉害的是,这个技术还能结合人工智能技术搞出可以自己学习的平台,用真实的数据不断调整模型,让预测更准和处理突发情况的能力变强。优化空调系统时,数字孪生能模拟冷气怎么流动、温度变化情况、哪些部位容易发热这些细节,然后灵活调整空调的工作效率;分配计算资源的时候,数字孪生

系统可以同时模拟多个任务运行的情况,帮助人工智能比较不同方案,找到最省电的那个方案。现在随着计算机性能和监测技术的进步,数字孪生现在变成了把虚拟和现实结合起来的東西,让能源管理从过去的分散和滞后变成有计划和预测性的操作。

3.3 构建 AI 驱动的能耗智能决策系统

传统能耗管理主要靠运维经验和固定规则,实时性和系统性不足,现在已经很难适应云计算里大量任务和复杂资源的匹配需要了。开发用 AI 驱动的智能能耗管理系统,变成实现绿色云计算的重要方法。这种系统一般有四个部分:感知、分析、预测和控制。感知部分要装传感器网络、能耗监测组件和日志收集器,实时收集服务器负载、温度、电压、空气湿度等不同维度的数据;分析部分用机器学习技术对数据进行分类和发现异常情况,动态评估能源消耗结构;预测部分用深度学习、线性回归这些算法来预测资源使用量变化、任务数量增长和对应的能耗走势;控制模块根据这些结果调整策略,比如重新分配资源、限制功率、调节风扇速度这些操作,最终形成智能化的闭环管理。系统还能和 SLA 协议结合,在满足要求的前提下尽量省电,还能通过反馈数据不断自我改进,慢慢达到最好的能耗效率。以后这种智能系统还会增加碳排放跟踪和环境成本计算功能,让云计算公司在赚钱的同时也能尽到环保责任,推动整个行业把数字经济和低碳发展更好地结合起来。

4 结语

面对不断增长的计算需求和日益严峻的环保压力,云计算平台的能耗管理已不再是单一技术问题,而是企业、技术、社会多方协同的系统性挑战。人工智能的引入不仅提供了强大的技术支撑,更开启了云计算从“节能化”向“智能化”的转型路径。通过深度融合 AI 与能耗调控机制,构建基于数据驱动、自适应决策、智能反馈的绿色计算架构,已成为云服务产业迈向高质量发展和双碳目标的重要突破口。未来,随着更多智能算法的应用与平台能力的提升,一个更加低碳、高效、智能的云计算时代正加速到来。

参考文献

- [1] 王伟. 人工智能在数据中心资源分配中的应用研究[J]. 计算机应用研究, 2021, 38 (12): 3601-3607.
- [2] 刘洋, 李明. 深度学习在虚拟机调度中的应用[J]. 计算机科学, 2020, 47 (6): 88-94.
- [3] 陈杰, 黄婷. 基于强化学习的资源优化方法[J]. 软件学报, 2022, 33 (8): 2400-2410.