

基于信息-能源-时间三元组的 算力资源综合度量与优化方法研究



Comprehensive Measurement and Optimization Method of Computing Power Resources Based on Information-Energy-Time Triplet

刘超清/Liu Chaoqing¹, 周丹媛/Zhou Danyuan²,
缪政/Miao Zheng³, 丁亦志/Ding Yizhi¹, 武振宇/Wu Zhenyu¹

(1. 中国移动通信集团设计院有限公司, 中国 北京 100080;

2. 中国移动通信集团有限公司, 中国 北京 100033;

3. 中移(苏州)软件技术有限公司, 中国 苏州 215000)

(1. China Mobile Group Design Institute Co., Ltd., Beijing 100080, China;

2. China Mobile Communications Group Co., Ltd., Beijing 100033, China;

3. China Mobile (Suzhou) Software Technology Co., Ltd., Suzhou 215000, China)

DOI: 10.12142/ZTETJ.202602011

网络出版地址: <https://link.cnki.net/urlid/34.1228.tn.20260422.1724.003>

网络出版日期: 2026-04-24

收稿日期: 2025-11-20

摘要: 现有算力度量方法以浮点运算次数(FLOPS)为单一指标, 忽视了能效与时间效率, 导致资源调度不合理。提出一种基于“信息-能源-时间”三元组的算力资源综合度量与优化方法。构建了以最优硬件为基准的归一化评分模型, 将算力、能效比、时间3个维度归一化至[0,1]区间, 并支持性能优先、节能优先、响应优先等场景化权重配置。基于“算力大脑”的仿真验证表明, 该方法能够有效提升资源利用率, 降低能耗并优化任务执行时间, 为算力网络的精细化资源管理提供了新方案。

关键词: 算力网络; 算力度量; 信息-能源-时间三元组; 归一化评分; 能效比; 动态调度

Abstract: Existing computing power metrics rely solely on Floating Point Operations Per Second (FLOPS) as a performance indicator, neglecting energy efficiency and time efficiency, which leads to suboptimal resource scheduling. To address this issue, a comprehensive measurement and optimization method for computing power resources based on the “information-energy-time” triplet is proposed. A normalized scoring model benchmarked against optimal hardware is constructed, which normalizes the three dimensions of computing power, energy efficiency ratio, and time into the [0, 1] range, and supports scenario-specific weight configurations such as performance, energy-saving, and response priorities. Simulation verification based on the “Computing Network Brain” shows that the proposed method can effectively improve resource utilization, reduce energy consumption, and optimize task execution time, providing a new solution for fine-grained resource management in computing power networks.

Keywords: computing power network; computing power measurement; information-energy-time triplet; normalized scoring; energy efficiency ratio; dynamic scheduling

引用格式: 刘超清, 周丹媛, 缪政, 等. 基于信息-能源-时间三元组的算力资源综合度量与优化方法研究 [J]. 中兴通讯技术, 2026, 32(2): 73-80. DOI: 10.12142/ZTETJ.202602011

Citation: Liu C Q, Zhou D Y, Miao Z, et al. Comprehensive measurement and optimization method of computing power resources based on information-energy-time triplet [J]. ZTE technology journal, 2026, 32(2): 73-80. DOI: 10.12142/ZTETJ.202602011

1 概述

1.1 算力网络发展背景

随着国家一体化大数据中心算力枢纽项目(即“东数西算”工程)的逐步推进,“算力网络”作为一种新兴网络概念应运而生。算力网络的核心在于借助先进的网络技术,实现不同地理分布的算力中心之间的互联互通,并对上述资源进行实时监控。基于该机制,算力网络能够在全局范

围内智能地分配与调度各类任务及数据传输,从而提升算力资源的利用效率,促进算力、数据与应用资源的共享与整合。简而言之,算力网络是一种具备全局感知、分配与调度能力的智能网络系统,其体系架构如图1所示。

基于“算力可用、可控、可计量”的一体化算力交易调度平台,目前常用的算力系统调度方法包括集中式、分布式、层次式和混合式调度等,主要通过动态精准调度算力供



图1 算力网络体系架构示意图

给侧的资源，以调整资源分配。

云计算、大数据和人工智能技术的迅速发展，计算任务的复杂性不断提升，都为算力网络调度带来了更为严峻的挑战。算力调度问题本质上是一个多约束、多目标的优化问题，涉及资源分配、任务调度和成本效益分析等多个方面。在云计算环境中，计算任务通常需要在多个物理或虚拟服务器上执行，每个任务均具有特定的计算需求和截止时间。资源分配的目标是在满足所有任务需求的前提下，最小化资源使用成本或最大化资源利用率。由于任务和资源的多样性以及需求的动态变化，算力调度问题通常具备大规模与高复杂性的特征。

1.2 现有算力度量的不足

算力是指服务器通过对数据进行处理并输出结果的计算能力，是衡量服务器或芯片计算能力的关键指标，数值越大表示计算能力越强。算力主要包括以中央处理器（CPU）芯

片为代表的通用算力，以及以图形处理器（GPU）/人工智能（AI）加速芯片为代表的高性能算力（即智能算力）。常用的计量单位为每秒所执行的浮点运算次数（FLOPS）。

传统算力性能指标存在以下不足：

- 1) 单一度量指标。现有技术通常仅以 FLOPS 作为算力的主要计量单位，忽视了算力资源的多样性与复杂性，难以全面反映不同计算任务与算法的实际需求。
- 2) 缺乏综合考量。FLOPS 作为单一指标，无法兼顾能源效率、计算任务特性（如并行性、内存需求等）以及软件架构的先进性，而这些因素对算力性能具有重要影响。
- 3) 适应性不足。随着人工智能与大数据等技术的发展，现代计算任务日趋多样化，单一的 FLOPS 指标难以适应新场景下算力资源的复杂需求。
- 4) 资源调度局限。仅依赖 FLOPS 指标进行资源调度，容易造成资源分配不均衡，难以实现系统性能与能源效率的

整体优化。

5) 无法反映实际应用性能。FLOPS 指标主要衡量处理器的峰值性能，而在实际应用中，受内存带宽、输入输出(I/O) 瓶颈等因素影响，往往难以达到该峰值。

1.3 相关研究工作

在算力度量与优化方面，学术界和工业界已进行诸多探索。Google 提出“Carbon-Aware Computing”框架，根据碳排放强度动态调度数据中心负载；Microsoft Azure 采用“Resource Central”系统，通过机器学习预测资源使用并优化能效。在中国，阿里云提出“绿色计算”模型，综合考虑电源使用效率(PUE)、碳排放与任务延迟；华为提出“全栈 AI 能效优化”方案，将 AI 训练任务调度至可再生能源富集区域。然而，这些方法大多聚焦于能源或性能的单维度，尚未形成综合考虑能源、环境、计算效率及用户需求的统一模型。

本文将在第2章提出基于“信息-能源-时间”三元组的算力资源综合度量方法，相比上述单维度方法，能够更全面地评估和调度算力资源。

2 基于信息-能源-时间三元组的算力优化方法

2.1 度量标准设定

本节中我们定义算力资源综合度量的3个核心维度及其量化指标。为构建归一化综合评分模型，需对3个维度进行准确定义和量化。

1) 运算能力 (I): 以 FLOPS 为基本单位进行衡量。为实现跨平台可比性，采用每秒万亿次浮点运算 (TFLOPS) 作为量化指标。该指标通过标准化基准测试获取。针对异构算力芯片类型，本方案通过统一基准测试套件进行测评，按 AI 训练、推理、高性能计算 (HPC) 等任务类型进行分类评估，确保跨平台的一致性。

2) 能源效率 (η): 以能效比 (FLOPS/W) 为关键指标进行衡量。该指标反映单位能耗下的运算能力，计算公式为 $\eta = I/P$ ，其中 I 为算力， P 为功耗。能效比越高，表示能源利用效率越优。能源效率体现了工艺的先进性和能源利用效率，是评估算力可持续性的关键因素。

3) 时间效率 (T): 以特定任务的执行时间 (s) 进行衡量。通过标准化基准测试任务 (如 ResNet-50 图像识别) 记录各硬件平台的执行时间。该指标反映软件架构先进性和软硬件协同效率。时间越短，执行效率越高。

为实现综合得分的统一量化处理，以上3个维度均采

用相对归一化方法，以当前云化资源池中的最优硬件性能作为基准进行标准化。

2.2 信息-能源-时间三元组的归一化模型

以资源池内最优硬件为基准的归一化综合评分模型如公式 (1):

$$S = w_1 \times \frac{I}{I_{\max}} + w_2 \times \frac{\eta}{\eta_{\max}} + w_3 \times \frac{T_{\min}}{T} \quad (1)$$

1) 参数说明

- I : 当前硬件的算力 (TFLOPS);
- I_{\max} : 当前资源池中的最大算力 (TFLOPS);
- η : 当前硬件的能效比 (TFLOPS/W), $\eta = I/P$;
- η_{\max} : 当前资源池中的最高能效比 (TFLOPS/W);
- T : 当前硬件的任务执行时间 (s);
- T_{\min} : 当前资源池中的最短任务执行时间 (s);
- w_1, w_2, w_3 : 权重因子, 满足 $w_1 + w_2 + w_3 = 1, w_i \in [0, 1]$ 。

2) 模型特点

- 得分范围: $[0, 1]$, $S=1$ 表示在当前资源池中达到最优性能;
- 算力维度: I/I_{\max} , 体现相对于最优硬件的算力水平, 越高越好;
- 能效维度: η/η_{\max} , 体现相对于最优硬件的能效水平, 越高越好;
- 时间维度: T_{\min}/T , 体现相对于最短时间的效率水平, 时间越短得分越高。

3) 权重因子设定原则

- w_1, w_2, w_3 的值根据应用场景需求确定及调整, 且总和等于 1, 以确保综合得分 S 的有效性;
- w_1 代表运算能力的权重, w_2 代表能效优化的权重, w_3 代表时间效率的权重。

w 权重因子的设定应根据实际应用场景的具体需求及数据中心的具体目标, 进行细致的调整与优化。

- 算力优先场景: 对于大规模 AI 训练、科学计算等对算力需求敏感的场景, 可设置较高 w_1 值;
- 能效优先场景: 对于绿色数据中心、边缘计算等对能耗敏感的场景, 可设置较高 w_2 值;
- 时间优先场景: 对于实时推理、在线服务等对延迟敏感的场景, 可设置较高 w_3 值;
- 均衡场景: 对于通用计算场景, 可采用相对均衡的权重配置。

4) 模型适用范围

当前模型适用于软件栈统一、同类型硬件资源池。在跨硬件类型进行比较时， T 维度因软件适配度的差异而存在不可公度性问题。建议在实际应用中限定比较范围或引入软件适配系数进行修正。

2.3 模型计算示例

下面我们通过具体示例说明归一化模型的使用方法（实际应用时需结合现有数据中心的硬件配置、系统软件架构及性能等）。

2.3.1 基础数据假设

假设某数据中心资源池包含两种异构服务器，需为图像识别任务选择最优算力资源。基准测试采用 ResNet-50 模型在 ImageNet 数据集上的推理任务，资源池硬件规格如表 1 所示。

资源池基准值为：

- $I_{\max} = \max(1.89, 312) = 312$ TFLOPS (GPU 服务器)；
- $\eta_{\max} = \max(0.0092, 1.04) = 1.04$ TFLOPS/W (GPU 服务器)；
- $T_{\min} = \min(1800, 300) = 300$ s (GPU 服务器)；
- 图像识别任务推荐使用 FP16 混合精度，可显著提升 GPU 效率。

2.3.2 归一化计算

假设应用场景以平衡性能与能效为目标，采用权重配置 $w_1=0.4, w_2=0.4, w_3=0.2$ 。

1) CPU 服务器综合得分

算力归一化： $I/I_{\max} = 1.89/312 = 0.0061$ ；能效归一化： $\eta/\eta_{\max} = 0.0092/1.04 = 0.0089$ ；时间归一化： $T_{\min}/T = 300/1800 = 0.1667$ 。

$$S_{\text{CPU}} = 0.4 \times 0.0061 + 0.4 \times 0.0089 + 0.2 \times 0.1667 = 0.0024 + 0.0036 + 0.0333 = 0.0393 \quad (2)$$

2) GPU 服务器综合得分

算力归一化： $I/I_{\max} = 312/312 = 1.0$ ；能效归一化： $\eta/\eta_{\max} = 1.04/1.04 = 1.0$ ；时间归一化： $T_{\min}/T = 300/300 = 1.0$ 。

$$S_{\text{GPU}} = 0.4 \times 1.0 + 0.4 \times 1.0 + 0.2 \times 1.0 = 1.0 \quad (3)$$

3) 结果分析

GPU 服务器在 3 个维度上均达到资源池最优水平，综合得分为满分 1.0，因此应作为图像识别任务的优选资源。与 CPU 服务器相比，GPU 服务器的算力提升 165 倍 (312/1.89)，能效提升 113 倍 (1.04/0.0092)，执行时间缩短 83% (从 1800 s 降至 300 s)。从综合权重得分来看，GPU 服务器与 CPU 服务器的比值约为 25 : 1 (1/0.0393)。具体情况如表 2 所示。

2.3.3 资源分配优化

基于综合得分可指导算力资源的智能调度：

- 高分资源 ($S \rightarrow 1$)：优先分配计算密集型任务；
- 中等得分资源 ($0.5 < S < 1$)：适合平衡型任务；
- 低得分资源 ($S < 0.5$)：仅作为弹性扩展资源；
- 优化效果预期：通过将图像识别任务优先调度至 GPU 服务器，相比 CPU 服务器可实现约 25 倍的综合能力提升。

2.4 对综合得分 S 值任务依赖性的分析

1) S 值与任务相关

S 值是“硬件 × 任务”的联合属性，同一硬件在不同任务下对应的 S 值可能不同。以 CPU 和 GPU 为例：

- 在 AI 训练任务中，GPU 凭借其大规模并行计算能力，

表 2 资源池硬件综合得分及调度建议示例

服务器	S 值	排名	调度建议
GPU 服务器	1.0000	1	优先分配任务
CPU 服务器	0.0393	2	仅在 GPU 资源不足时使用

CPU：中央处理器 GPU：图形处理器

表 1 资源池硬件配置及基准测试结果示例

参数	CPU 服务器	GPU 服务器	说明
处理器	Intel Xeon Gold 6248R	NVIDIA A100 80GB PCIe	—
核心规格	24核/48线程	6912 Shaders	—
算力 I (TFLOPS)	1.89	312 (FP16 Tensor)	CPU 实测 SGEMM、GPU FP16 性能
功耗 P (W)	205	300	官方 TDP 值
能效比 η (TFLOPS/W)	0.0092	1.04	$\eta = I/P$
执行时间 T (s)	1800	300	ResNet-50 推理时间
软件架构	标准框架	优化 CUDA 库	—

CPU：中央处理器 CUDA：统一计算设备架构 GPU：图形处理器 SGEMM：单精度通用矩阵乘法 TFLOPS：每秒万亿次浮点运算

S值通常远高于CPU；

- 在逻辑推理任务（如复杂的条件分支、递归调用）中，CPU凭借其强大的分支预测与低延迟特性，S值反而可能高于GPU。

2) 时间维度 T 具有任务敏感性

T 是 S 值任务依赖性的主要来源。同一硬件的时间维度得分取决于所执行的具体任务：

- AI 训练任务：GPU 的 T_{norm} 通常远优于 CPU（矩阵运算效率高）；

- 编译构建任务：CPU 的 T_{norm} 可能优于 GPU（编译器优化成熟）；

- 数据库查询：涉及大量分支判断，CPU 可能表现更优。

T 本质上是“软硬件协作效率”的度量，而非单纯的硬件执行时间。在软件栈统一的场景下，T 反映硬件架构差异；在软件生态存在差距的场景下，T 会放大生态成熟度的影响。

3) 等价性的任务相对性

2.3 节模型计算示例中“25 台 CPU \approx 1 台 GPU”的等价性具有任务相关性：该等价关系适用于 AI 训练类并行计算任务，但对于串行计算或逻辑推理任务则不成立。

因此，在进行算力评估和调度时，必须明确所针对的任务类型，否则 S 值的比较将失去意义。

3 基于“大云天穹”算网大脑的模拟验证

“大云天穹”算网大脑是中国移动算力网络的核心中枢系统（如图 2 所示），向下拉通算力网络全域资源，向上支撑算网融合类全业务，旨在最大化算力供给能力、数据处理能力、网络连通能力与业务供给能力。

算网大脑构建了“5 个中心、2 个网关”的整体架构。其中，5 个中心包括设计中心、编排调度中心、感知接入中心、能力接入中心与智能中心。设计中心负责算网业务拓扑、流程及端到端解决方案的设计，形成算网产品方案；编

排调度中心承担业务需求分析、算网产品方案加载、一体化方案生成与交付、资源能力调度及实例全生命周期管理等职能；感知接入中心实现算网多专业、多维度数据的感知接入，完成算网全域数据的汇聚与管理；能力接入中心负责基础设施层能力的接入与管理，包括接入认证、能力封装、能力管理与执行监控等；智能中心提供业务需求解析、方案设计、业务与资源调整调度等过程中的智能化能力。

2 个网关分别为网络域能力网关与算力域能力网关。网络域能力网关实现对网管系统实时感知、敏捷开通、动态优化能力的纳管与开放；算力域能力网关实现对异构算力服务，以及基础设施即服务（IaaS）、平台即服务（PaaS）、软件即服务（SaaS）各层服务的集中接入与开放。

本研究基于中国移动算网大脑，通过引入基于信息、能

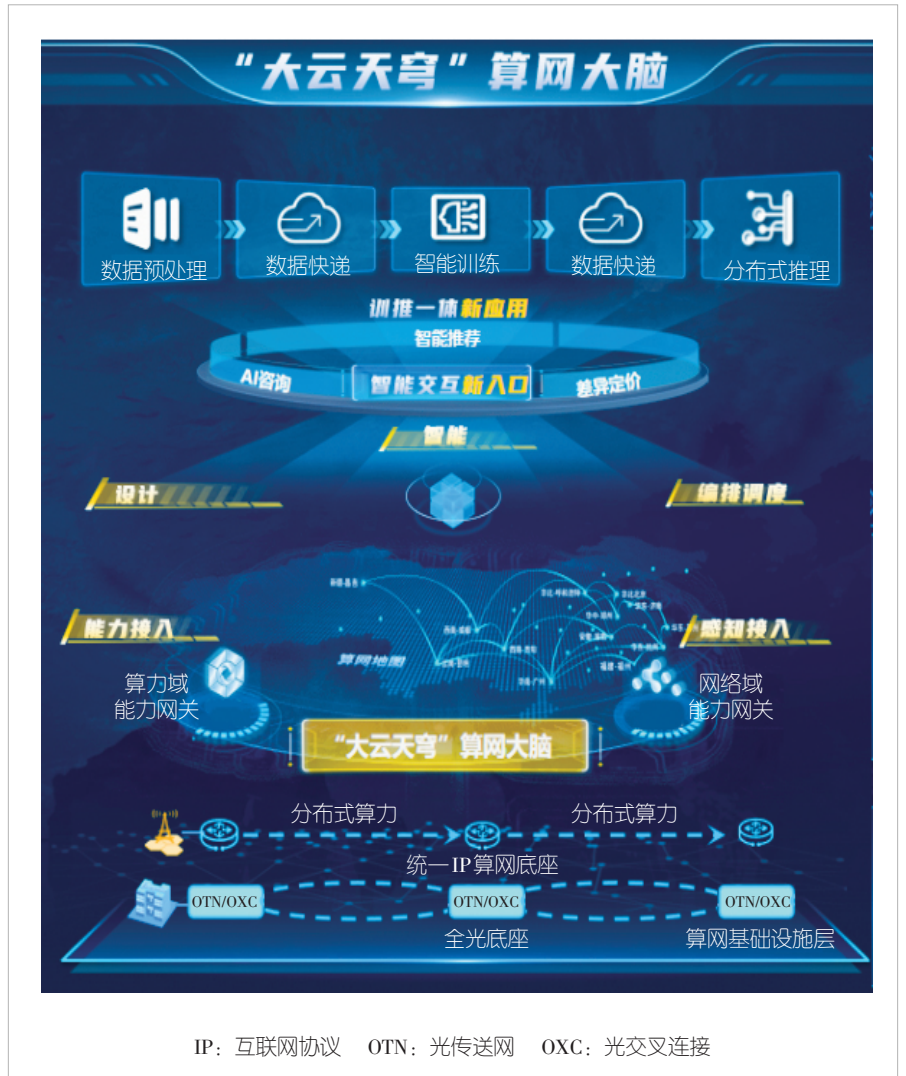


图2 “大云天穹”算网大脑体系架构示意图

源与软件架构先进性的综合评估体系，融合机器学习算法，构建了一种创新的算力资源管理与优化平台，并开展了模拟实验验证。该验证实验平台由算力度量模型、资源优化算法及动态调整机制3部分组成，依托算网大脑的智能调度能力，部分解决了传统算力度量方式单一、资源调度效率偏低及能源效率不高等问题。验证结果表明，所提出的“三元组”度量优化方法在提升资源利用率、降低运营成本、减少环境影响等方面具有显著效果，有助于数据中心与云计算平台实现资源的精细化管理。

3.1 算法与计算模型设计

1) 数据预处理与需求预测：收集并预处理算力系统相关数据，包括任务需求、资源状态、网络状况等，通过时序预测模型对算力需求进行预测。

2) 算法模型：构建基于机器学习的模型（如随机森林或神经网络），以各项度量指标为输入，输出综合评估分数。训练采用历史数据集进行交叉验证，优化模型参数，确保预测准确性。

3) 动态调整机制：集成在线学习模块，根据新数据实时调整模型权重，以适应硬件与软件架构的变化。权重因子可依据实时监控数据与反馈循环动态调整。例如，当系统检测到能源消耗过高时，可自动增加 w_2 的权重，以优化能源效率。

4) 参数赋权的动态监测与反馈机制：随着任务性质、资源状况及外部条件（如能源价格波动）的变化， w_1 、 w_2 、 w_3 的最优组合需进行动态调整。例如，在算力中心投产初期，系统可能更注重效率与时间（即较高的 w_1 和 w_3 ）；随着能耗数据的积累与节能目标的凸显， w_2 的比重可能上升。通过实时监控算力网络状态与任务执行情况，并结合反馈循环与机器学习机制，权重策略可不断优化，实现最佳资源分配，形成闭环优化，持续提升资源调度效果。

3.2 系统设计及体系模块架构

系统核心功能模块如下：

- 1) 用户界面：提供交互界面，直观展示算力资源的综合评分，并支持定制化查询与排序功能。
- 2) 数据输入/输出模块：支持用户手动或自动上传硬件配置、能耗及性能数据。

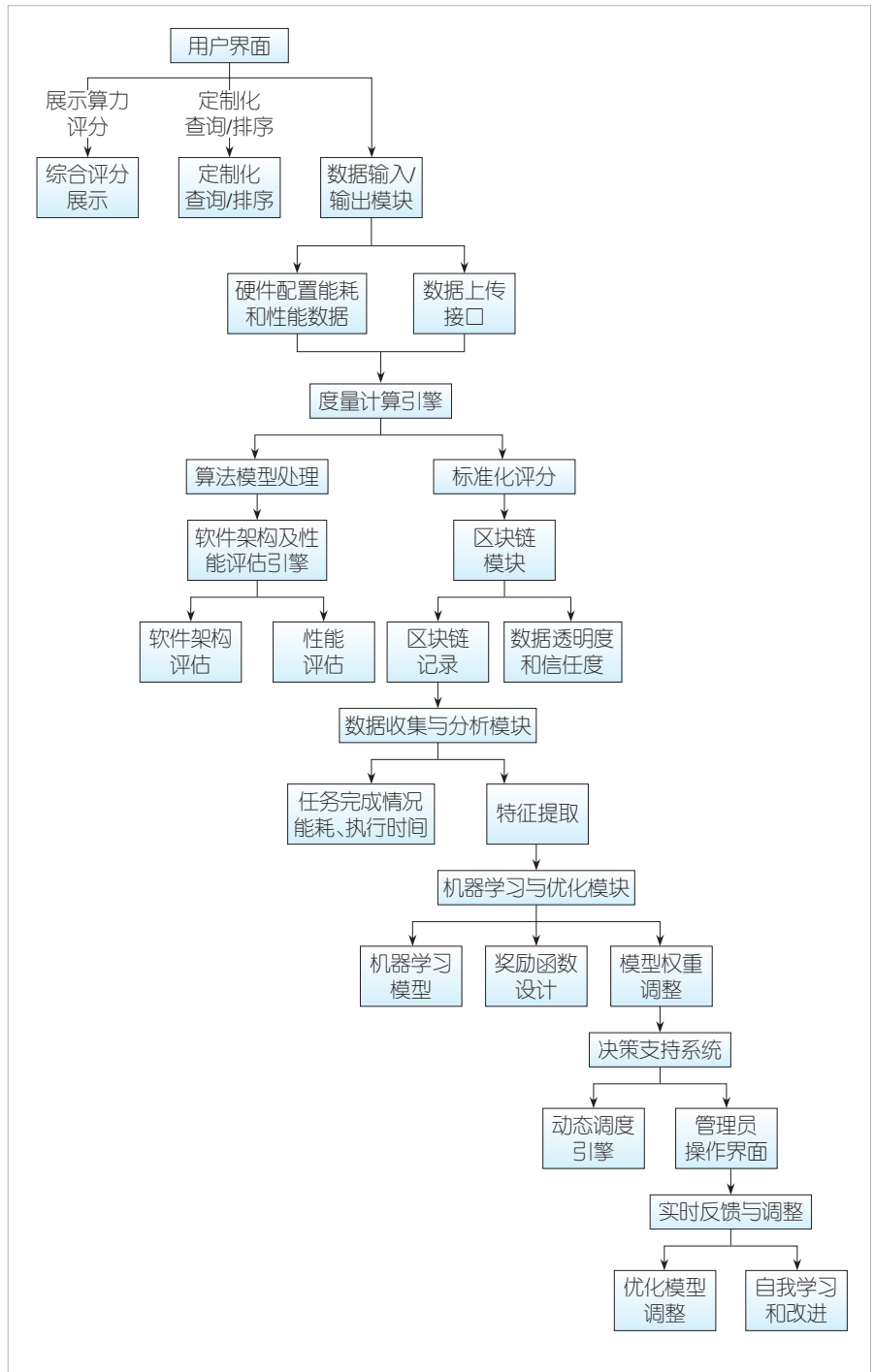


图3 系统核心功能模块及相互关系示意图

3) 度量计算引擎：内置基于信息-能源-时间三元组的算法模型，接收输入数据后，依据统一算力标识符进行处理，输出标准化的算力评分。

4) 软件架构及性能评估引擎：集成代码分析工具，支持计算引擎与业务算法的智能解耦，自动评估软件架构的先进性与性能。

5) 区块链模块：采用区块链技术（私有链）实现数据存证，记录硬件状态与调度决策，通过智能合约自动执行权重更新。

6) 数据收集与分析模块：实时监控并收集算力资源的运行数据，从中提取关键特征。

7) 机器学习与优化模块：设计并训练机器学习模型，实现权重的自动调整与多目标奖励函数的优化。

8) 决策支持系统：基于当前权重与实时数据，动态调整算力资源分配策略，并提供管理员操作界面。

9) 实时反馈与调整：形成闭环控制，根据任务执行后的反馈持续优化模型，实现自我学习与迭代改进。

系统具体功能模块及其相互关系示意如图3所示。

3.3 仿真实验验证

3.3.1 实验目的与评价指标

本文验证了基于“信息-能源-时间”三元组的归一化评分模型在异构资源池调度中的理论有效性，分析了不同权重配置策略下的硬件选择差异，并对比了软件生态存在差距场景下的模型表现。实验通过测算综合得分 S ，考察不同策略下首选硬件的变化以及模型的区分能力。

3.3.2 实验设计与实验场景

1) 实验采用包含两款 GPU 设备的异构资源池：NVIDIA A10 (FP16 125 TFLOPS、150W TDP、能效比 0.83 TFLOPS/W) 与华为昇腾 910B (FP16 320 TFLOPS、310W TDP、能效比 1.03 TFLOPS/W)。两款设备性能接近但架构不同，用于验证三元组模型在不同维度权重下的选择逻辑。

2) FP16 算力、功耗及能效比数据来源于公开参数。

3) 执行时间定义为完成单次 epoch 训练所需的时间

(s)。基于本文定义的基准测试任务 (ResNet-50 模型、ImageNet 数据集、批次大小 256、FP16 混合精度训练)，NVIDIA A10 的执行时间为 320 s/epoch，华为昇腾 910B 的执行时间为 380 s/epoch。

3.3.3 实验权重配置

为展示三元组模型各维度的主导作用，实验采用特征化的权重配置，具体如表3所示。

3.3.4 实验结果

仿真实验结果如表4、表5所示，由此可得出以下结论：

1) 不同权重策略对应不同的 GPU 选择，这表明模型能够有效区分各维度的贡献。当硬件在各维度上各有优势时，模型支持基于权重配置实现差异化选择。

2) 三元组模型适用于同类型硬件资源池，而对于不同硬件类型之间的比较，需考虑软件适配度差异。

3) 在软件栈统一的场景下，差异化选择难度较大，模型的核心价值在于正确识别综合最优设备。

4) S 值是“硬件 × 任务”的联合属性。进行算力评估时，必须明确所针对的任务类型；在调度决策中，应优先考虑任务与硬件的匹配度，而非单纯比较不同硬件的 S 值。

表3 特征化的权重配置

策略	权重配置(w_1, w_2, w_3)	适用场景	配置意图
算力优先	(0.8, 0.1, 0.1)	大规模 AI 训练	算力维度主导选择
能效优先	(0.1, 0.8, 0.1)	绿色数据中心	能效维度主导选择
时间优先	(0.1, 0.1, 0.8)	实时推理服务	时间维度主导选择
均衡策略	(0.34, 0.33, 0.33)	通用计算场景	各维度均衡评估

AI: 人工智能 CPU: 中央处理器 GPU: 图形处理器

表4 GPU各维度归一化得分

GPU 型号	I_{norm} (算力)	η_{norm} (能效)	T_{norm} (时间)	维度优势
NVIDIA A10	0.390 6	0.805 8	1.000 0	时间维度领先
华为昇腾 910B	1.000 0	1.000 0	0.842 1	算力、能效维度领先

CPU: 中央处理器 GPU: 图形处理器

表5 GPU4种策略综合得分对比

策略	权重(w_1, w_2, w_3)	首选 GPU	综合得分 S	A10 得分	昇腾得分	选择依据
算力优先	(0.8, 0.1, 0.1)	昇腾 910B	0.984 2	0.492 6	0.984 2	$I_{norm}=1.0$, 算力维度满分占优
能效优先	(0.1, 0.8, 0.1)	昇腾 910B	0.984 2	0.783 8	0.984 2	$\eta_{norm}=1.0$, 能效维度满分占优
时间优先	(0.1, 0.1, 0.8)	NVIDIA A10	0.919 6	0.919 6	0.873 6	$T_{norm}=1.0$, 时间维度满分占优
均衡策略	(0.34, 0.33, 0.33)	昇腾 910B	0.947 9	0.728 6	0.947 9	综合各维度评估, 算力+能效优势明显

GPU: 图形处理器

4 结束语

本文提出一种基于“信息-能源-时间”三元组的算力资源综合度量与优化方法。通过构建以最优硬件为基准的归一化评分模型，实现算力资源的标准化管理与场景化优化配置。该模型首次将三要素统一为可量化的算力评分，引入可解释的权重分配机制，支持算力优先、能效优先、时间优先等场景化策略切换，从而更精细地匹配任务需求与算力资源，同时兼顾能耗效率。

当前工作的主要局限在于：实验仅基于单一数据中心，尚未验证跨域调度效果；权重设定仍依赖人工经验初值。

下一步研究计划包括：1) 构建跨域试验环境，验证模型在“东数西算”跨枢纽场景下的效果；2) 引入强化学习自动搜索最优权重组合，减少人工干预；3) 与行业联盟合作，推动算力网络标准化认证，加快速度量模型与资源管理方案的标准化进程，促进“信息-能源-时间”三元组指标成为算力交易的行业标准。

综上所述，本文构建了一种可随技术进步与市场变化持续优化的综合算力资源度量与分配方法，验证了所提评估框架的有效性与竞争力。通过标准化度量、算法优化、系统设计与自动化赋能，为“东数西算”工程提供了一种可参考的技术方案，能够满足远期算力网络调度需求，具备一定的理论价值与工程落地可行性。

参考文献

- [1] 苗青, 归律, 张开飞, 等. 算力网络与“东数西算”工程实践 [M]. 北京: 人民邮电出版社, 2023
- [2] Radovanovic A, Koningstein R, Schneider L, et al. Carbon-aware computing framework [PP/OL]. arXiv(2021-06-11)[2026-02-18]. <https://arxiv.org/abs/2106.11750>
- [3] Microsoft. Resource central: AI-driven datacenter efficiency [R]. Microsoft Research, 2022
- [4] 阿里云. 绿色计算白皮书: 面向碳中和的云资源调度 [R]. 2023
- [5] 华为. 全栈AI能效优化技术白皮书 [R]. 2022
- [6] 中国移动通信集团有限公司. 算力网络白皮书 [R]. 2021
- [7] 工业和信息化部电子第五研究所. 异构融合计算技术白皮书(2023年)[R]. 2023

作者简介



刘超清，中国移动通信集团设计院有限公司高级工程师；主要研究方向为算力网络、业务网、AI领域关键技术等。



周丹媛，中国移动通信集团有限公司项目经理；主要研究方向为云计算和人工智能技术、高性能网络等。



缪政，中移（苏州）软件技术有限公司计划部工程师；主要研究方向为分布式存储、混合云、绿色低碳与成本优化等。



丁亦志，中国移动通信集团设计院有限公司正高级工程师；主要研究方向为业务网、算力网络、信息能源融合关键技术等。



武振宇，中国移动通信集团设计院有限公司高级工程师；主要研究方向为算力网络、智算中心等。