

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号
WO 2024/140355 A1

(43) 国际公布日
2024 年 7 月 4 日 (04.07.2024)

(51) 国际专利分类号:
G06Q 10/0631 (2023.01)

(21) 国际申请号: PCT/CN2023/140084

(22) 国际申请日: 2023 年 12 月 20 日 (20.12.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202211676872.0 2022 年 12 月 26 日 (26.12.2022) CN

(71) 申请人: 杭州阿里云计算技术
有限公司(HANGZHOU ALICLOUD FEITIAN INFOR-
MATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国
浙江省杭州市余杭区五常街道文一西路 969 号
3 幢 5 层 553 室, Zhejiang 310023 (CN)。

(72) 发明人: 程明 (CHENG, Ming); 中国浙江省杭州
市西湖区三墩镇灯彩街 1008 号云谷园区,
Zhejiang 310030 (CN)。 闫月君 (YAN, Yuejun); 中
国北京市朝阳区望京东园四区 6 号楼金辉大
厦, Beijing 100102 (CN)。 毛宏举 (MAO, Hongju);

中国浙江省杭州市西湖区三墩镇灯彩街 1008
号云谷园区, Zhejiang 310030 (CN)。 刘文字 (LIU,
Wenyu); 中国北京市朝阳区望京东园四区 6 号
楼金辉大厦, Beijing 100102 (CN)。

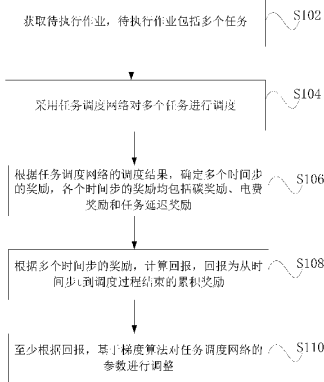
(74) 代理人: 北京同钧律师事务所 (BEIJING TONGJUN
LAW FIRM); 中国北京市海淀区西直门北大街 32
号院 1 号楼 7 层 808 室, Beijing 100082 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家
保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG,
BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU,
CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ,
IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ,
LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN,
MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA,
PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,
SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ,
UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ,

(54) Title: TASK SCHEDULING NETWORK TRAINING METHOD BASED ON DEEP REINFORCEMENT LEARNING

(54) 发明名称: 基于深度强化学习的任务调度网络训练方法



(图 1)

- S102 Acquire a job to be executed, wherein said job comprises a plurality of tasks
S104 Schedule the plurality of tasks by using a task scheduling network
S106 Determine rewards of a plurality of time steps according to a scheduling result of the task scheduling network, wherein rewards of each time step comprise carbon rewards, electricity bill rewards and task delay rewards
S108 Calculate a return according to the rewards of the plurality of time steps, wherein the return is a cumulative reward from a time step t to the end of a scheduling process
S110 At least according to the return, adjust parameters of the task scheduling network on the basis of a gradient algorithm

(57) Abstract: Disclosed in the embodiments of the present invention is a task scheduling network training method based on deep reinforcement learning. The task scheduling network training method based on deep reinforcement learning comprises: acquiring a job to be executed; scheduling a plurality of tasks by using a task scheduling network; determining rewards of a plurality of time steps according to a scheduling result of the task scheduling network, wherein rewards of each time step comprise carbon rewards, electricity bill rewards and task delay rewards; calculating a return G_t according to the rewards of the plurality of time steps, wherein the return G_t is a cumulative reward from a time step t to the end of a scheduling process; and at least according to the return G_t , adjusting parameters of the task scheduling network on the basis of a gradient algorithm. The present invention is conducive to realizing low-carbon and economical operation of a data center, and solves the technical problem, in the relevant art, of it being impossible for a task scheduling algorithm to realize the low-carbon and economical operation of the data center on the basis of ensuring the timeliness of task processing, and the problem, in the relevant art, of non-convergence or slow convergence of scheduling algorithms.

[见续页]

NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚
(AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE,
BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,
HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO,
PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN,
TD, TG)。

根据细则4.17的声明:
— 关于申请人有权要求在先申请的优先权(细则
4.17(iii))

本国际公布:
— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本发明实施例公开了一种基于深度强化学习的任务调度网络训练方法, 基于深度强化学习的任务调度网络训练方法, 包括: 获取待执行作业; 采用任务调度网络对多个任务进行调度; 根据任务调度网络的调度结果, 确定多个时间步的奖励, 各个时间步的奖励包括碳奖励、电费奖励和任务延迟奖励; 根据多个时间步的奖励, 计算回报 G_t , 回报 G_t 为从时间步 t 到调度过程结束的累积奖励; 至少根据回报 G_t , 基于梯度算法对任务调度网络的参数进行调整。通过本发明, 有利于实现数据中心低碳、经济运行, 解决相关技术中的任务调度算法无法在保证任务处理及时性的基础上实现数据中心低碳经济运行的技术问题, 以及相关技术中调度算法不收敛或收敛慢的问题。