(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

(43) 国际公布日
2025 年 9 月 18 日 (18.09.2025)

WIPO | PCT

(10) 国际公布号
WO 2025/189915 A1

(51) 国际专利分类号:
G06N 20/00 (2019.01)

(21) 国际申请号: PCT/CN2024/143782

(22) 国际申请日: 2024 年 12 月 30 日 (30.12.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202410299301.2 2024年3月14日 (14.03.2024) CN
202410798950.7 2024年6月18日 (18.06.2024) CN

(71) 申请人: 华为云计算技术有限公司 (**HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.**) [CN/CN]; 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心 550025 (CN)。

(72) 发明人: 李亿 (**LI, Yi**); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心 550025 (CN)。 练韵文 (**LIAN, Yunwen**); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心 550025 (CN)。 王志超 (**WANG, Zhichao**); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心 550025 (CN)。 董谷音 (**DONG, Guyin**); 中国贵州省贵阳市贵安新区黔中大道交兴功路华为云数据中心 550025 (CN)。

(74) 代理人: 北京中博世达专利商标代理有限公司 (**BEIJING ZBSD PATENT & TRADEMARK AGENT LTD.**); 中国北京市海淀区交大东路31号11号楼8层 100044 (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,

(54) **Title:** MODEL TRAINING METHOD AND APPARATUS

(54) 发明名称: 一种模型训练方法及装置

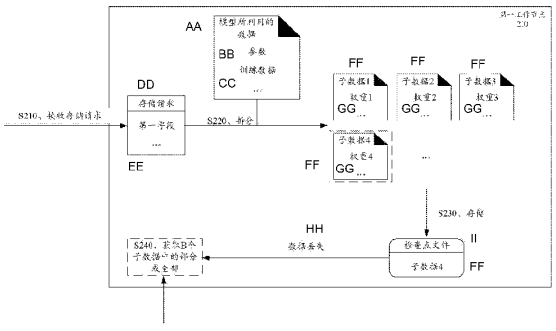

图 2

210 First working node
S210 Receive a storage request
S220 Split
S230 Store
S240 Acquire some or all of the B pieces of sub-data
AA Data used by a model
BB Parameters
CC Training data
DD Storage request
EE First field
FF Sub-data
GG Weight
HH Data loss
II Check point file

(57) **Abstract:** The present application relates to the technical field of cloud computing. Disclosed are a model training method and apparatus. In the model training method, in the process of a first working node and at least one second working node collaboratively training a model, check point files stored by the first working node are only part of complete data used by the model, and the complete data used by the model can be obtained by means of integrating check point files respectively stored by the first working node and the at least one second working node. If data loss occurs during the training of the model by the first working mode, the first working node acquires some or all of the checkpoint files, which are stored in the second working node and include B pieces of sub-data. Therefore, the first working node only stores part of the complete data used by the model, and the volume of stored data is reduced, thereby reducing the time consumed for storage and improving the efficiency of model training.

WO 2025/189915 A1

[见续页]

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, 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)。

本国际公布:
— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 公开了一种模型训练方法及装置，涉及云计算技术领域。该模型训练方法包括：在第一工作节点和至少一个第二工作节点协同对模型进行训练的过程中，第一工作节点保存的检查点文件仅为完整的模型所利用的数据中的部分数据，且，第一工作节点和至少一个第二工作节点分别保存的检查点文件综合后可得到完整的模型所利用的数据。若第一工作节点在对模型进行训练过程中数据丢失，则第一工作节点获取第二工作节点存储的包含B个子数据的检查点文件中的部分或全部。进而，第一工作节点仅保存完整的模型所利用的数据中的部分数据，减小了保存的数据量，从而减少了保存所需耗时，提高了对模型进行训练的效率。