

(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 2021/196800 A1

(43) 国际公布日
2021 年 10 月 7 日 (07.10.2021)

(51) 国际专利分类号:
G06K 9/62 (2006.01) *G06N 20/20* (2019.01)
G06N 3/08 (2006.01)

(21) 国际申请号: PCT/CN2020/142061

(22) 国际申请日: 2020 年 12 月 31 日 (31.12.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010246686.8 2020 年 3 月 31 日 (31.03.2020) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 司晓云(SI, Xiaoyun); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 胡新宇(HU, Xinyu); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 薛莉(XUE, Li); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 张亮(ZHANG, Liang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 尘福兴(CHEN, Fuxing); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 北京三高永信知识产权代理有限公司(BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (CN)。

(54) Title: METHOD AND APPARATUS FOR OBTAINING ARTIFICIAL INTELLIGENCE MODEL, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 获取人工智能模型的方法、装置、设备及存储介质

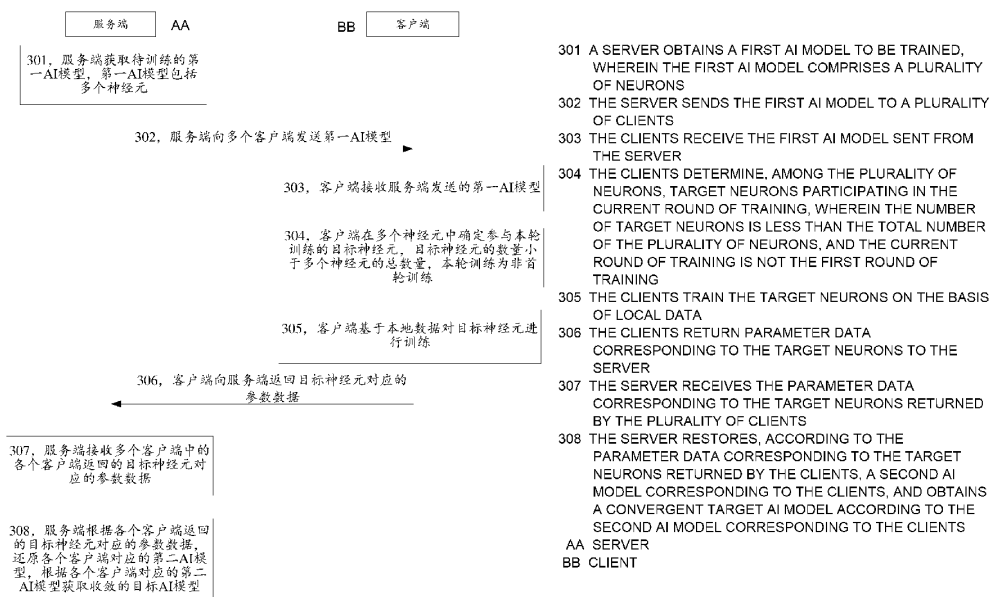


图 3

(57) Abstract: A method and apparatus for obtaining an artificial intelligence (AI) model, a device, and a storage medium. In an example of clients executing the method, clients receive a first AI model sent from a server (303), wherein the first AI model comprises a plurality of neurons; the clients determine, among the plurality of neurons, target neurons participating in the current round of training, wherein the current round of training is not the first round of training, and the number of target neurons is less than the total number of the plurality of neurons (304); the clients train the target neurons on the basis of local data (305); the clients return parameter data

[见续页]

(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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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))。

corresponding to the target neurons to the server (306), wherein the parameter data corresponding to the target neurons is used by the server to obtain a convergent target AI model. The target neurons are trained according to an active condition without training all the neurons in the model, thereby reducing power consumption and accelerating the training speed. Also, the clients merely transmit parameters corresponding to trained neurons to the server, thereby saving communication bandwidth.

(57) 摘要: 一种获取人工智能模型的方法、装置、设备及存储介质, 以客户端执行该方法为例, 客户端接收服务端发送的第一人工智能AI模型(303), 第一AI模型包括多个神经元; 客户端在多个神经元中确定参与本轮训练的目标神经元, 本轮训练为非首轮训练, 目标神经元的数量小于多个神经元的总数量(304); 客户端基于本地数据对目标神经元进行训练(305); 客户端向服务端返回目标神经元对应的参数数据(306), 目标神经元对应的参数数据用于服务端获取收敛的目标AI模型。由于无需对模型中的所有神经元进行训练, 而是按照活跃条件来训练目标神经元, 从而能够降低功耗, 加快训练速度。又由于客户端向服务端仅传输训练的神经元对应的参数, 因而可节省通信带宽。