

(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 2022/016556 A1

(43) 国际公布日
2022 年 1 月 27 日 (27.01.2022)

- (51) 国际专利分类号:
G06F 16/9535 (2019.01) **G06N 3/04** (2006.01)
- (21) 国际申请号: PCT/CN2020/104653
- (22) 国际申请日: 2020 年 7 月 24 日 (24.07.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (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: 程朋祥(CHENG, Pengxiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 董振华(DONG, Zhenhua); 中国广东省深圳市龙岗区坂田华为总部办公楼,

Guangdong 518129 (CN)。 何秀强(HE, Xiuqiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 张小莲(ZHANG, Xiaolian); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 殷实(YIN, Shi); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 胡粤麟(HU, Yuelin); 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: 深圳市深佳知识产权代理事务所(普通合伙)(SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦B座18C2、18D、18E、18E2, Guangdong 518001 (CN)。

(54) Title: NEURAL NETWORK DISTILLATION METHOD AND APPARATUS

(54) 发明名称: 一种神经网络蒸馏方法以及装置

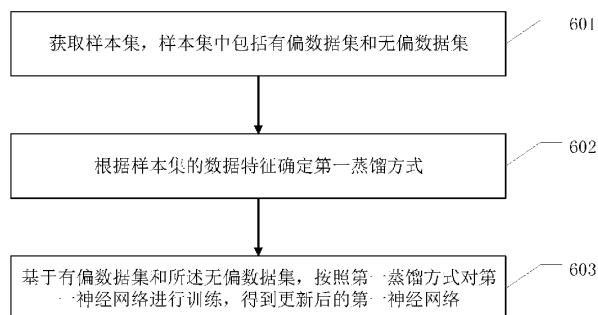


图 6

601 ACQUIRE A SAMPLE SET, THE SAMPLE SET COMPRISING A BIASED DATASET AND AN UNBIASED DATASET
602 DETERMINE A FIRST DISTILLATION MEANS ACCORDING TO A DATA FEATURE OF THE SAMPLE SET
603 TRAIN A FIRST NEURAL NETWORK ACCORDING TO THE FIRST DISTILLATION MEANS ON THE BASIS OF THE BIASED DATASET AND THE UNBIASED DATASET SO AS TO OBTAIN AN UPDATED FIRST NEURAL NETWORK

(57) Abstract: A neural network distillation method and apparatus in the field of artificial intelligence, which are used for providing a neural network having a lower output bias, and improving the output precision of the neural network, and can select a suitable distillation means according to different scenarios, and have a strong generalization capability. The method comprises: acquiring a sample set, the sample set comprising a biased dataset and an unbiased dataset (601), the biased dataset comprising a biased sample, and the unbiased dataset comprising an unbiased sample; determining a first distillation means according to a data feature of the sample set (602), training a teacher model by using the unbiased dataset in the first distillation means, and training a student model by using the biased dataset; and training a first neural network according to the first distillation means on the basis of the biased dataset and the unbiased dataset

[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))。

so as to obtain an updated first neural network (603).

(57) 摘要: 人工智能领域的一种神经网络蒸馏方法以及装置, 用于提供输出偏置更低的神经网络, 提高神经网络的输出精度, 且可以根据不同的场景选择合适的蒸馏方式, 泛化能力强。该方法包括: 获取样本集, 该样本集包括有偏数据集和无偏数据集 (601), 有偏数据集中包括有偏置的样本, 无偏数据集中包括无偏置的样本; 根据样本集的数据特征确定第一蒸馏方式 (602), 在第一蒸馏方式中使用无偏数据集训练老师模型, 使用有偏数据集训练学生模型; 基于有偏数据集和无偏数据集, 按照第一蒸馏方式对第一神经网络进行训练, 得到更新后的第一神经网络 (603)。