(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
2022 年 1 月 13 日 (13.01.2022)

WIPO | PCT

(10) 国际公布号
WO 2022/007651 A1

(51) 国际专利分类号:
G06F 21/62 (2013.01) *G06N 20/00* (2019.01)
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2021/102488

(22) 国际申请日: 2021 年 6 月 25 日 (25.06.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010640985.X 2020年7月6日 (06.07.2020) CN

(71) 申请人: 北京字节跳动网络技术有限公司 (**BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市石景山区实兴大街30号院3号楼2层B-0035房间, Beijing 100041 (CN)。 字节跳动有限公司 (**BYTEDANCE INC.**) [US/US]; 美国加利福尼亚州洛杉矶西杰弗逊大道12655号六层第137室, California 90066 (US)。

(72) 发明人: 孙建凯 (**SUN, Jiankai**); 美国加利福尼亚州洛杉矶西杰弗逊大道12655号六层第137室, California 90066 (US)。 高伟豪 (**GAO, Weihao**); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。 张宏毅 (**ZHANG, Hongyi**); 美国加利福尼亚州洛杉矶西杰弗逊大道12655号六层第137室, California 90066 (US)。 王崇 (**WANG, Chong**); 美国加利福尼亚州洛杉矶西杰弗逊大道12655号六层第137室, California 90066 (US)。 解浚源 (**XIE, Junyuan**); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。 吴良超 (**WU, Liangchao**); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。 刘小兵 (**LIU, Xiaobing**); 美国加利福尼亚州洛杉矶西杰弗逊大道12655号六层第137室, California 90066 (US)。

(74) 代理人: 北京天达共和律师事务所 (**EAST & CONCORD PARTNERS**); 中国北京市朝阳

(54) **Title:** DATA PROTECTION METHOD AND APPARATUS, AND SERVER AND MEDIUM

(54) **发明名称**: 数据保护方法、装置、服务器和介质

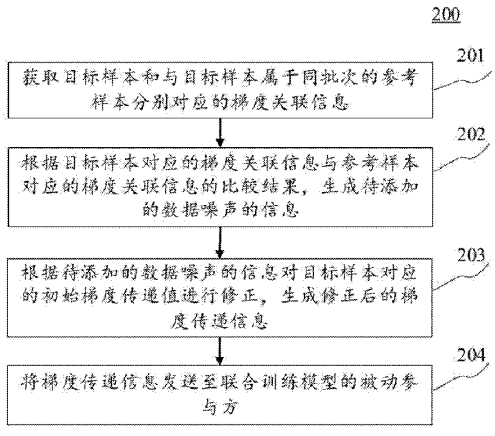

图 2

201 ACQUIRE GRADIENT ASSOCIATED INFORMATION, WHICH RESPECTIVELY CORRESPONDS TO A TARGET SAMPLE AND A REFERENCE SAMPLE THAT BELONGS TO THE SAME BATCH AS THE TARGET SAMPLE
202 ACCORDING TO A COMPARISON RESULT OF THE GRADIENT ASSOCIATED INFORMATION CORRESPONDING TO THE TARGET SAMPLE AND THE GRADIENT ASSOCIATED INFORMATION CORRESPONDING TO THE REFERENCE SAMPLE, GENERATE INFORMATION OF DATA NOISE TO BE ADDED
203 ACCORDING TO THE INFORMATION OF SAID DATA NOISE, CORRECT AN INITIAL GRADIENT TRANSFER VALUE CORRESPONDING TO THE TARGET SAMPLE, SO AS TO GENERATE CORRECTED GRADIENT TRANSFER INFORMATION
204 SEND THE GRADIENT TRANSFER INFORMATION TO A PASSIVE PARTY OF A JOINT TRAINING MODEL

(57) **Abstract:** Disclosed are a data protection method and apparatus, and a server and a medium. A particular embodiment of the method comprises: acquiring gradient associated information, which respectively corresponds to a target sample that belongs to a binary classification sample set with unbalanced distribution and a reference sample that belongs to the same batch as the target sample; according to a comparison result of the gradient associated information corresponding to the target sample and the gradient associated information corresponding to the reference sample, generating information of data noise to be added; according to the information of said data noise, correcting an initial gradient transfer value corresponding to the target sample, such that corrected gradient transfer information corresponding to samples in the sample set that belong to different types is consistent; and sending the gradient transfer information to a passive party of a joint training model. By means of the embodiment, there is no significant difference between corrected gradient transfer information corresponding to positive and negative samples, thereby effectively protecting the security of data.

WO 2022/007651 A1

[见续页]

区东三环北路8号亮马河大厦2座19层, Beijing 100004 (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, 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))。

(57) 摘要：本申请实施例公开了数据保护方法、装置、服务器和介质。该方法的一具体实施方式包括：获取属于分布不均衡的二分类样本集的目标样本和与该目标样本属于同批次的参考样本分别对应的梯度关联信息；根据该目标样本对应的梯度关联信息与该参考样本对应的梯度关联信息的比较结果，生成待添加的数据噪声的信息；根据该待添加的数据噪声的信息对该目标样本对应的初始梯度传递值进行修正，以使样本集中属于不同类别的样本所对应的修正后的梯度传递信息一致；将该梯度传递信息发送至联合训练模型的被动参与方。该实施方式使得正负样本对应的修正后的梯度传递信息没有明显差异，从而有力地保护了数据安全。