

(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 2023/216902 A1

(43) 国际公布日
2023 年 11 月 16 日 (16.11.2023)

- (51) 国际专利分类号:
G06F 9/50 (2006.01) **G06N 3/08** (2023.01)
G06F 16/9535 (2019.01) **G06N 20/00** (2019.01)
G06N 3/04 (2023.01)
- (21) 国际申请号: PCT/CN2023/091189
- (22) 国际申请日: 2023 年 4 月 27 日 (27.04.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202210524865.2 2022年5月13日 (13.05.2022) 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)。脸萌有限公司 (LEMON INC.) [GB/GB]; 英国英属开曼群岛大开曼西湾路 802 号木槿道大楼邮政信箱 31119, Cayman Islands (GB)。

- (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)。杨鑫 (YANG, Xin); 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); 中国北京市朝阳区七圣中街 12 号院融中心 B1 小邮局, Beijing 100028 (CN)。吴迪 (WU, Di); 中国北京市朝阳区七圣中街 12 号院融中心 B1 小邮局, Beijing 100028 (CN)。

(54) Title: METHOD AND APPARATUS FOR MODEL PERFORMANCE EVALUATION, AND DEVICE AND MEDIUM

(54) 发明名称: 用于模型性能评估的方法、装置、设备和介质

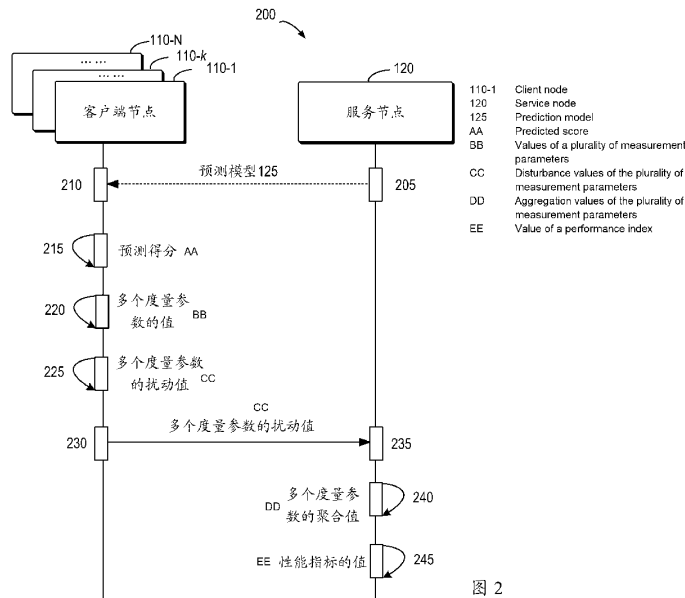


图 2

(57) Abstract: Provided in the embodiments of the present disclosure are a method and apparatus for model performance evaluation, and a device and a medium. The method comprises: at a client node, respectively applying a plurality of data samples to a prediction model, so as to obtain a plurality of predicted scores, which are output by the prediction model, wherein the plurality of predicted scores respectively indicate prediction probabilities of the plurality of data samples belonging to a first category or a second category; on the basis of a plurality of ground-truth labels and the plurality of predicted scores of the plurality of data samples, determining values

[见续页]

(74) 代理人:北京世辉律师事务所(SHIHUI PARTNERS);
中国北京市朝阳区建国门外大街2号北京银
泰中心C座42层, Beijing 100022 (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,
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))。

of a plurality of measurement parameters which are related to a predetermined performance index of the prediction model; applying disturbance to the values of the plurality of measurement parameters, so as to obtain disturbance values of the plurality of measurement parameters; and sending the disturbance values of the plurality of measurement parameters to a service node. Therefore, the aim of privacy protection of local label data of a client node is achieved while model performance evaluation is realized.

(57) 摘要: 根据本公开的实施例, 提供了用于模型性能评估的方法、装置、设备和介质。该方法包括: 在客户端节点处, 将多个数据样本分别应用到预测模型, 以得到预测模型输出的多个预测得分, 多个预测得分分别指示多个数据样本属于第一类别或第二类别的预测概率; 基于多个数据样本的多个真值标签和多个预测得分, 确定与预测模型的预定性能指标相关的多个度量参数的值; 对多个度量参数的值施加扰动, 得到多个度量参数的扰动值; 以及将多个度量参数的扰动值发送给服务节点。由此, 在实现模型性能评估的同时, 达到了对客户端节点本地标签数据的隐私保护目的。