(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
2019 年 1 月 10 日 (10.01.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/007417 A1

(51) 国际专利分类号:
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2018/094786

(22) 国际申请日: 2018 年 7 月 6 日 (06.07.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710552377.1 2017年7月7日 (07.07.2017) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层847号邮箱, Grand Cayman (KY)。

(72) 发明人; 及

(71) 申请人 (仅对*US*): 王力 (WANG, Li) [CN/CN]; 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 赵沛霖 (ZHAO, Peilin) [CN/CN]; 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 周俊 (ZHOU, Jun) [CN/CN]; 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 李小龙 (LI, Xiaolong) [US/CN]; 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲2号渔阳置业大厦B座605, Beijing 100029 (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, 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,

(54) **Title:** TRAINING SAMPLE GENERATION METHOD AND DEVICE BASED ON PRIVACY PROTECTION

(54) **发明名称**: 基于隐私保护的训练样本生成方法和装置

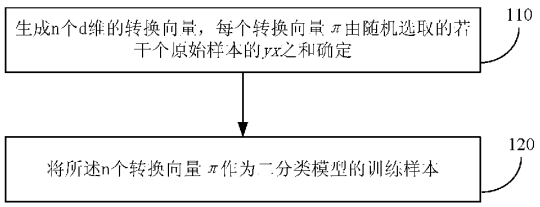

图 1

110 Generate *n* d-dimensional conversion vectors π, each of the conversion vectors π being determined according to a sum of *x* and *y* of a plurality of randomly selected raw samples

120 Take the *n* d-dimensional conversion vectors π as training samples for a binary classification model

(57) **Abstract:** Provided in the specification is a training sample generation method based on privacy protection. Mined raw data comprises *m* raw samples. Each raw sample comprises a d-dimensional raw vector *x* and an outputted flag value *y*, and each of *m* and *d* is a natural number. The method comprises: generating *n* d-dimensional conversion vectors π, each of the conversion vectors π being determined according to respective sums of *y* and *x* of a plurality of randomly selected raw samples; and taking the *n* d-dimensional conversion vectors π as training samples for a binary classification model.

(57) **摘要**: 本说明书提供一种基于隐私保护的训练样本生成方法，被挖掘的原始数据包括m个原始样本，每个原始样本包括d维的原始向量x和输出标记值y，m、d为自然数，所述方法包括：生成n个d维的转换向量π，每个转换向量π由随机选取的若干个原始样本的yx之和确定；将所述n个转换向量π作为二分类模型的训练样本。

WO 2019/007417 A1

[见续页]

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。