

(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 2020/034750 A1

(43) 国际公布日
2020 年 2 月 20 日 (20.02.2020)

- (51) 国际专利分类号:
G06K 9/62 (2006.01) **G06N 20/00** (2019.01)
- (21) 国际申请号: PCT/CN2019/092199
- (22) 国际申请日: 2019 年 6 月 21 日 (21.06.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810924268.2 2018 年 8 月 14 日 (14.08.2018) 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: 郑毅 (ZHENG, Yi); 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)。 张鹏 (ZHANG, Peng); 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)。 潘健民 (PAN, Jianmin); 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: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (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,

(54) Title: CLASSIFICATION MODEL GENERATION METHOD AND DEVICE, AND DATA IDENTIFICATION METHOD AND DEVICE

(54) 发明名称: 一种分类模型生成方法及装置、一种数据识别方法及装置

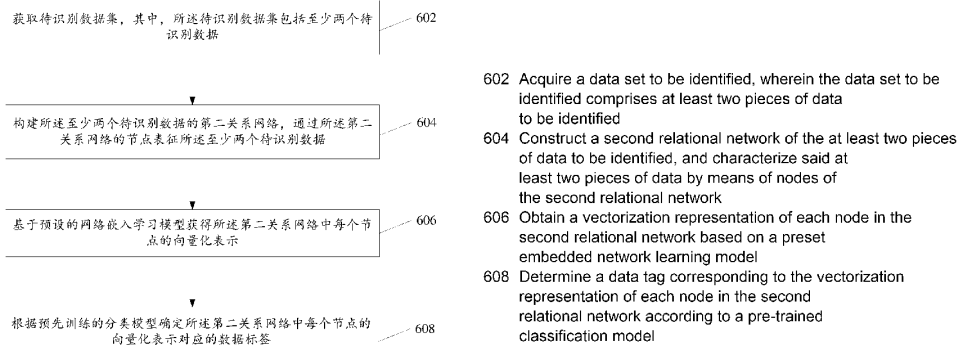


图 6

(57) Abstract: Provided are a classification model generation method and device, and a data identification method and device, wherein the data identification method comprises: acquiring a data set to be identified; constructing a second relational network of at least two pieces of data to be identified, and characterizing said at least two pieces of data by means of nodes of the second relational network; obtaining a vectorization representation of each node in the second relational network based on a preset embedded network learning model; and determining a data tag corresponding to the vectorization representation of each node in the second relational network according to a pre-trained classification model.

(57) 摘要: 本申请提供一种分类模型生成方法及装置、一种数据识别方法及装置, 其中, 所述数据识别方法包括获取待识别数据集; 构建所述至少两个待识别数据的第二关系网络, 通过所述第二关系网络的节点表征所述至少两个待识别数据; 基于预设的网络嵌入学习模型获得所述第二关系网络中每个节点的向量化表示; 根据预先训练的分类模型确定所述第二关系网络中每个节点的向量化表示对应的数据标签。

[见续页]

WO 2020/034750 A1

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,
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))。