

(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 2020/078098 A1

(43) 国际公布日
2020 年 4 月 23 日 (23.04.2020)

(51) 国际专利分类号:
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2019/101335

(22) 国际申请日: 2019 年 8 月 19 日 (19.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811210339.9 2018 年 10 月 17 日 (17.10.2018) 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: 陈超超 (CHEN, Chaochao); 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); 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,

(54) Title: GRADIENT BOOSTING DECISION TREE-BASED METHOD AND DEVICE FOR MODEL TRAINING

(54) 发明名称: 一种基于梯度提升决策树的模型训练方法及装置

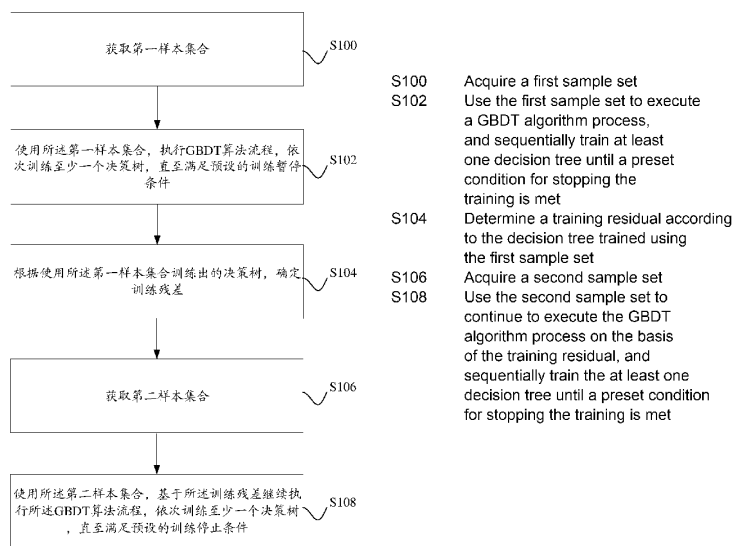


图 1

(57) Abstract: Disclosed are a gradient boosting decision tree (GBDT)-based method and device for training a model, the method comprising: dividing a GBDT algorithm process into two phases; in the former phase, acquiring a labelled sample from a data region of a service scenario which is similar to a target service scenario, sequentially training a plurality of decision trees, and determining a training residual generated after undergoing the former phase training; in the latter phase, acquiring a labelled sample from the data region of the target service scenario, and on the basis of the training residual, continuing to train the plurality of decision trees. Finally,

[见续页]

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, 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))。

the model applied to the target service scenario is actually obtained by integrating the decision trees trained in the former phase and the decision trees trained in the latter phase.

(57) 摘要: 公开了一种基于梯度提升决策树的模型训练方法及装置。将一个GBDT算法流程划分为两个阶段, 在前一阶段, 从与目标业务场景相近的业务场景的数据域获取已标注样本依次训练若干决策树, 并确定经过前一阶段训练后产生的训练残差; 在后一阶段, 从目标业务场景的数据域获取已标注样本, 并基于所述训练残差, 继续训练若干决策树。最终, 应用于目标业务场景的模型实际上是由前一阶段训练出的决策树与后一阶段训练出的决策树集成得到的。