

(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 2019/223384 A1

(43) 国际公布日
2019 年 11 月 28 日 (28.11.2019)

(51) 国际专利分类号:
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2019/076264

(22) 国际申请日: 2019 年 2 月 27 日 (27.02.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810488062.X 2018 年 5 月 21 日 (21.05.2018) 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: 方文静 (FANG, Wenjing); 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)。高利翠 (GAO, Licui); 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,

(54) Title: FEATURE INTERPRETATION METHOD AND DEVICE FOR GBDT MODEL

(54) 发明名称: GBDT 模型的特征解释方法和装置

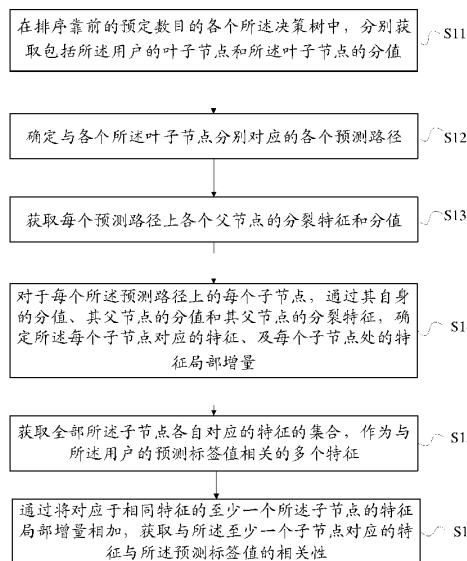


图 1

(57) Abstract: A method and device for acquiring a feature interpretation of a predicted tag value of a user. The method is performed after a tag value of a user is predicted by a GBDT model, and comprises: acquiring, from a predetermined number of top-ranked decision trees, leaf nodes comprising the user and a score of each of the leaf nodes (S11); determining a prediction path corresponding to each of the leaf nodes (S12); acquiring a splitting feature and a score of each parent node on each prediction path (S13); for each child node on each prediction path, determining, according to a score thereof, the score of the parent node thereof, and the splitting feature of the

[见续页]

WO 2019/223384 A1

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, 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))。

parent node thereof, a feature corresponding to the child node and a local feature increment at the child node (S14); acquiring a set of features corresponding to all of the child nodes, and using same as multiple features related to the predicted tag value of the user (S15); and acquiring, by adding the local feature increment of at least one child node corresponding to a same feature, a correlation between the feature corresponding to the at least one child node and the predicted tag value (S16).

(57) 摘要: 一种获取对用户的预测标签值的特征解释的方法和装置, 所述方法在通过GBDT模型预测用户的标签值之后执行, 包括: 在排序靠前的预定数目的各个决策树中, 分别获取包括所述用户的叶子节点和所述叶子节点的分值(S11); 确定与各个所述叶子节点分别对应的各个预测路径(S12); 获取每个预测路径上各个父节点的分裂特征和分值(S13); 对于每个预测路径上的每个子节点, 通过其自身的分值、其父节点的分值和其父节点的分裂特征, 确定所述每个子节点对应的特征、及每个子节点处的特征局部增量(S14); 获取全部所述子节点各自对应的特征的集合, 作为与用户的预测标签值相关的多个特征(S15); 以及通过将对应于相同特征的至少一个子节点的特征局部增量相加, 获取与至少一个子节点对应的特征与所述预测标签值的相关性(S16)。