

(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
2020 年 2 月 6 日 (06.02.2020)



(10) 国际公布号
WO 2020/024448 A1

- (51) 国际专利分类号:
G06Q 10/06 (2012.01) **G06K 9/62** (2006.01)
- (21) 国际申请号: PCT/CN2018/111124
- (22) 国际申请日: 2018 年 10 月 21 日 (21.10.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201810866432.9 2018年8月1日 (01.08.2018) CN
- (71) 申请人: 平安科技(深圳)有限公司 (PING AN TECHNOLOGY(SHENZHEN)CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
- (72) 发明人: 金戈(JIN, Ge); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 徐亮(XU, Liang); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 肖京(XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
- (74) 代理人: 北京中强智尚知识产权代理有限公司 (BEIJING STRONG WISDOM INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市朝阳区北苑东路 19 号院 3 号楼 2501 室, Beijing 100070 (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: GROUP PERFORMANCE GRADE IDENTIFICATION METHOD, DEVICE, STORAGE MEDIUM, AND COMPUTER APPARATUS

(54) 发明名称: 人群绩效等级识别方法、装置、存储介质及计算机设备

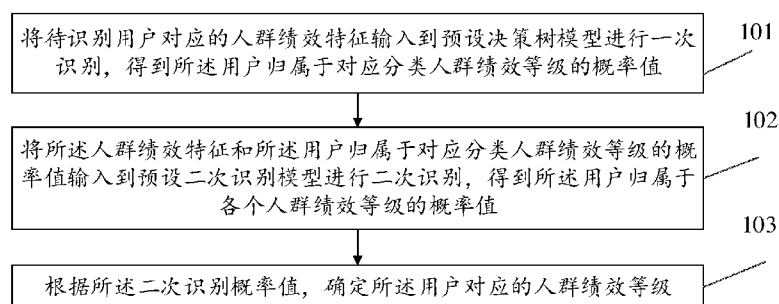


图 1

- 101 Input a group performance feature corresponding to a user awaiting identification into a pre-determined decision tree model and perform primary identification, and obtain a probability that the user belongs to a corresponding group performance grade
- 102 Input the group performance feature and the probability that the user belongs to the corresponding group performance grade into a pre-determined secondary identification model and perform secondary identification to obtain respective probability of the user belonging to each group performance grade
- 103 Determine a group performance grade corresponding to the user according to the probabilities of the secondary identification

(57) Abstract: A group performance grade identification method, a device, a storage medium, and a computer apparatus, pertaining to the technical field of artificial intelligence, and primarily aiming to reduce the impact of data distribution imbalances, reduce identification errors of a pre-determined decision tree model, and improve recall for a small target group while ensuring accurate identification. The method comprises: inputting a group performance feature corresponding to a user awaiting identification into a pre-determined decision tree model and performing primary identification, and obtaining a probability that the user belongs to a corresponding group

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)。

根据细则4.17的声明:

- 关于申请人有权申请并被授予专利 (细则4.17(ii))
- 发明人资格 (细则4.17(iv))

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

performance grade (101); inputting the group performance feature and the probability that the user belongs to a corresponding group performance grade into a pre-determined secondary identification model and performing secondary identification to obtain respective probabilities of the user belonging to each group performance grade (102); and determining a group performance grade corresponding to the user according to the probabilities of the individual group performance grades (103).

(57) 摘要: 一种人群绩效等级识别方法、装置、存储介质及计算机设备, 涉及人工智能技术领域, 主要目的在于能够在一定程度上克服数据分布平衡的影响, 减少预设决策树模型识别错误的情况, 且能够在保证识别精确度的前提下, 提升占比较低的目标人群召回率。所述方法包括: 将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别, 得到所述用户归属于对应分类人群绩效等级的概率值 (101); 将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别, 得到所述用户归属于各个人群绩效等级的概率值 (102); 根据所述各个人群绩效等级的概率值, 确定所述用户对应的人群绩效等级 (103)。

人群绩效等级识别方法、装置、存储介质及计算机设备

本申请要求与2018年8月1日提交中国专利局、申请号为2018108664329、申请名称为“人群绩效等级识别方法、装置、存储介质及计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在申请中。

技术领域

本申请涉及人工智能技术领域，尤其是涉及一种人群绩效等级识别方法、装置、存储介质及计算机设备。

背景技术

近年来，很多行业开始重视人群绩效，尤其是开始重视识别人群绩效等级，通过识别出目标人群并对目标人群进行奖励，能够大大促进企业整体效益的提升。

目前，在识别人群绩效等级时，通常仅通过决策树模型识别人群绩效等级，即将决策树模型对用户识别得到的人群绩效等级，确定为用户的人群绩效等级。然而，由于人群数据量庞大且各个人群绩效等级的人群数据分布不平衡，决策树模型可能无法取得理想拟合效果，通常占比较低的人群通常为目标人群。例如，高绩效等级的人群数据通常会远小于普通绩效等级的人群数据，一般高绩效等级的人群占比通常为20%，普通绩效等级的人群占比通常为80%，高绩效等级的人群通常为目标人群。若仅通过决策树模型识别人群绩效等级，会发生决策树模型识别错误的情况，造成识别出来的占比较低的目标人群的数量较少，从而造成占比较低的目标人群召回率较低。

发明内容

本申请提供了一种人群绩效等级识别方法、装置、存储介质及计算机设备，主要在于能够在一定程度上克服数据分布平衡的影响，减少预设决策树模型识别错误的情况，且能够在保证识别精确度的前提下，提升占比较低的目标人群召回率。

根据本申请的第一个方面，提供一种人群绩效等级识别方法，包括：

将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；

将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；

根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

根据本申请的第二个方面，提供一种人群绩效等级识别装置，包括：

识别单元，用于将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；

所述识别单元，还用于将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；

确定单元，用于根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

根据本申请的第三个方面，提供一种计算机非易失性可读存储介质，其上存储有计算机可读指令，该计算机可读指令被处理器执行时实现以下步骤：

将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；

将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；

根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

根据本申请的第四个方面，提供一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现以下步骤：

将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；

将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；

根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

本申请提供的一种人群绩效等级识别方法、装置、存储介质及计算机设备，与目前仅通过决策树模型识别人群绩效等级相比，本申请能够将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值。与此同时，能够将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；并能够根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。从而能

够实现在预设决策树模型的识别结果上，利用预设二次识别模型对所述用户进行二次识别，能够在一定程度上克服数据分布平衡的影响，减少预设决策树模型识别错误的情况，且能够在保证识别精确度的前提下，提升占比较低的目标人群召回率。

附图说明

此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

图 1 示出了本申请实施例提供的一种人群绩效等级识别方法流程图；

图 2 示出了本申请实施例提供的另一种人群绩效等级识别方法流程图；

图 3 示出了本申请实施例提供的一种人群绩效等级识别装置的结构示意图；

图 4 示出了本申请实施例提供的另一种人群绩效等级识别装置的结构示意图；

图 5 示出了本申请实施例提供的一种计算机设备的实体结构示意图。

具体实施方式

下文中将参考附图并结合实施例来详细说明本申请。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。

如背景技术，目前，在识别人群绩效等级时，通常仅通过决策树模型识别人群绩效等级，即将决策树模型对用户识别得到的人群绩效等级，确定为用户的人群绩效等级。然而，由于人群数据量庞大且各个人群绩效等级的人群数据分布不平衡，决策树模型可能无法取得理想拟合效果，通常占比较低的人群通常为目标人群。若仅通过决策树模型识别人群绩效等级，会发生决策树模型识别错误的情况，造成识别出来的占比较低的目标人群的数量较少，从而造成占比较低的目标人群召回率较低。

为了解决上述问题，本申请实施例提供了一种人群绩效等级识别方法，如图 1 所示，所述方法包括：

101、将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值。

其中，所述人群绩效特征可以包括但不限于：月均学习课程数、工作地址经纬度、单月内互联网交易产品数。所述人群绩效等级可以包括：高绩效等级和普遍绩效等级。所述预设决策树模型可以为根据样本人群绩效特征和样本人群绩效等级建立的，所述预设决策树模型可以保存有人群绩效特征和人群绩效等级之间的对应规则。所述用户对应的分类人群绩效等级可以为所述预设决策树模型对所述用户进行分类得到的人群绩效等级，具体地

可以为根据输入的人群绩效特征分类得到的，即每个人群绩效特征都可以作为识别所述用户类别的节点。在人群绩效特征输入到预设决策树模型后，预设决策树模型会根据人群绩效特征对所述用户进行分类，最后得到所述用户的人群绩效等级。

例如，月均学习课程数为 50、工作地址经纬度为（123.435，41.819）、单月内互联网交易产品数为 10，所述预设决策树模型会先根据月均学习课程数 50 对用户进行分类，然后根据工作地址经纬度（123.435，41.819）对用户进行分类；根据单月内互联网交易产品数 10 对用户进行分类，最后可以得到所述用户对应的分类人群绩效等级，并计算出所述用户归属于对应分类人群绩效等级的概率值。

102、将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值。

例如，若用户对应的人群绩效特征为：月均学习课程数为 50、工作地址经纬度为（123.435，41.819）、单月内互联网交易产品数为 10；所述预设决策树模型识别所述用户对应的分类人群绩效等级为高绩效等级，且计算所述用户归属于高绩效等级的概率值为 0.7，则可以将月均学习课程数 50、工作地址经纬度（123.435，41.819）、单月内互联网交易产品数 10、归属于高绩效等级的概率值 0.7 作为特征输入到预设逻辑回归模型对所述用户进行二次识别，得到所述用户归属于高绩效等级的概率值以及所述用户归属于普通绩效等级。

103、根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

需要说明的是，可以将最高概率值对应的人群绩效等级，确定为所述用户对应的人群绩效等级。例如，用户归属于高绩效等级的概率值为所有概率值中最高的，则可以将高绩效等级确定为所述用户对应的人群绩效等级，即确定用户为高绩效人群。

本申请实施例提供的一种人群绩效等级识别方法，与目前仅通过决策树模型识别人群绩效等级相比，本申请实施例能够将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值。与此同时，能够将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；并能够根据所述概率值，确定所述用户对应的人群绩效等级，从而能够实现在预设决策树模型的识别结果上，利用预设二次识别模型对所述用户进行二次识别，能够在一定程度上克服数据分布平衡的影响，减少预设决策树模型识别错误的情况，且能够在保证识别精确度的前提下，提升占比较低的目标人群召回率。

进一步的，为了更好的说明上述人群绩效等级识别的过程，作为对上述实施例的细化和扩展，本申请实施例提供了另一种人群绩效等级识别方法，如图 2 所示，所述方法包括：

201、获取待识别用户对应的人群绩效数据，并从所述人群绩效数据中提取出所述用户对应的人群绩效特征。

其中，所述人群绩效数据可以为人工上传的，也可以为从企业的绩效管控系统采集的。在本申请实施例中，可以通过特征关键字匹配的方式从人群绩效数据中提取出所述用户对应的人群绩效特征。具体地，可以将人群绩效特征关键字与所述人群绩效数据进行匹配，实现从所述人群绩效数据中提取出所述用户对应的人群绩效特征。

例如，若人群绩效特征关键字为“月均学习课程”，则可以从所述人群绩效数据中提取出所述用户对应的月均学习课程数，具体可以为 80；若人群绩效特征关键字为“工作地址经纬度”，则可以从所述人群绩效数据中提取出所述用户对应的工作地址经纬度，具体可以为 (123.436, 41.819)。若人群绩效特征关键字为“单月内互联网交易产品”，则可以从所述人群绩效数据中提取出所述用户对应的单月内互联网交易产品数，具体可以为 20。

对于本申请实施例，为了建立所述预设决策树模型，所述步骤 201 之前所述方法还可以包括：获取多个样本用户对应的样本人群绩效特征和样本人群绩效等级；根据所述样本人群绩效特征和所述样本人群绩效等级，建立所述预设决策树模型。具体地，可以通过决策树算法对所述样本人群绩效特征和所述样本人群绩效等级进行训练，得到所述预设决策树模型。所述预设决策树模型保存有人群绩效特征和人群绩效等级之间的对应规则。决策树算法可以是一种逼近离散函数值的方法。它是一种典型的分类方法，首先对数据进行处理，利用归纳算法生成可读的规则和决策树，然后使用决策对新数据进行分析。本质上决策树是通过一系列规则对数据进行分类的过程。

此外，为了建立所述二次识别模型，所述方法还可以包括：获取所述预设决策树模型计算的所述多个样本用户归属于对应分类人群绩效等级的概率值；根据所述样本人群绩效特征和所述各个样本人群绩效等级的概率值，建立所述预设二次识别模型。具体地，可以通过将所述样本人群绩效特征输入到所述预设决策树模型进行识别，得到所述多个样本用户归属于各个人群绩效等级的概率值。

在具体应用场景中，若所述预设二次识别模型为预设逻辑回归模型，则可以利用预设逻辑回归算法对所述样本人群绩效特征和所述多个样本用户归属于各个人群绩效等级的概率值进行训练，建立所述预设逻辑回归模型。预设逻辑回归算法又称为逻辑回归分析算

法，是分类和预测算法中的一种。可以通过历史数据的表现对未来结果发生的概率进行预测。在本申请实施例中，逻辑回归算法对所述各个样本用户对应的样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值进行训练时，可以将样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值作为自变量，把样本人群绩效等级作为因变量训练所述预设逻辑回归模型。

或者若所述预设二次识别模型为预设神经网络模型，则利用预设神经网络算法对所述样本人群绩效特征和所述多个样本用户归属于各个人群绩效等级的概率值进行训练，建立所述预设神经网络模型。所述预设神经网络算法可以为卷积神经网络算法、递归神经网络算法等，所述预设神经网络模型可以包括输入层、隐藏层、输出层。输入层可以对应于输入的样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值组成的训练集，所述隐藏层可以包括特征提取层和计算层，所述特征提取层可以用于根据所述训练集提取样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值，所述计算层可以用于计算各个人群绩效等级的概率值计算；所述输出层用于将计算的各个人群绩效等级的概率值进行输出。

202、将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值。

其中，得到所述用户归属于对应分类人群绩效等级的概率值的过程在步骤 101 已详细阐述，本申请实施例在此不进行赘述。

203、将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值。

其中，所述预设二次识别模型可以为预设逻辑回归模型或者预设神经网络模型。在所述预设二次识别模型中所述用户归属于对应分类人群绩效等级的概率值可以为与所述人群绩效特征类似的单独特征。即可以将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值作为共同特征输入到预设二次识别模型对所述用户进行二次识别。具体地，当所述预设二次识别模型为预设逻辑回归模型时，可以通过如下公式计算得到所述用户归属于各个人群绩效等级的概率值：

$$P(Y=1|x)=\pi(x)=\frac{1}{1+e^{-g(x)}}$$

其中， $x=(x_1, x_2, \dots, x_p)$ ， $x_1, x_2, \dots, x_{p-1}$ 可以为人群绩效特征、 x_p 可以为所述用户归属于对应分类人群绩效等级的概率值（作为一个单独特征）。因此，所述预设二次识

别模型可以以所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值作为输入特征或者训练集进行计算，得到所述用户归属于各个人群绩效等级的概率值。

例如，可以将月均学习课程数 80、工作地址经纬度（143.435，51.819）、单月内互联网交易产品数 30、归属于高绩效等级的概率值 0.6 作为共同特征输入到预设逻辑回归模型对所述用户进行二次识别，因此，可以得到所述预设逻辑回归模型输出的：所述用户归属于高绩效等级的概率值为 0.83，所述用户归属于普通绩效等级为 0.17。

204、对所述各个人群绩效等级的概率值进行排序，并将最高概率值对应的人群绩效等级，确定为所述用户对应的人群绩效等级。

例如，若所述预设二次识别模型计算用户归属于高绩效等级的概率值为 0.56，用户归属于普通绩效等级的概率值为 0.46，用户归属于高绩效等级的概率值最高，则可以将高绩效等级确定为所述用户对应的人群绩效等级，即确定用户为高绩效人群。若所述预设二次识别模型计算用户归属于高绩效等级的概率值为 0.32，计算用户归属于普通绩效等级的概率值为 0.68，用户归属于普通绩效等级的概率值最高，则可以将普通绩效等级确定为所述用户对应的人群绩效等级，即确定用户为普通绩效人群。

本申请实施例提供的另一种人群绩效等级识别方法，与目前仅通过决策树模型识别人群绩效等级相比，本申请实施例能够将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值。与此同时，能够将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；并能够根据所述概率值，确定所述用户对应的人群绩效等级。从而能够实现在预设决策树模型的识别结果上，利用预设二次识别模型对所述用户进行二次识别，能够在一定程度上克服数据分布平衡的影响，减少预设决策树模型识别错误的情况，且能够在保证识别精确度的前提下，提升占比较低的目标人群召回率。

进一步地，作为图 1 的具体实现，本申请实施例提供了一种人群绩效等级识别装置，如图 3 所示，所述装置包括：识别单元 31 和确定单元 32。

所述识别单元 31，可以用于将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值。所述识别单元 31 是本装置中将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值的主要功能模块。

所述识别单元 31，还可以用于将所述人群绩效特征和所述用户归属于对应分类人群绩

效等级的概率值输入到预设二次识别模型进行二次识别,得到所述用户归属于各个人群绩效等级的概率值;所述识别单元 31 还是本装置中将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别,得到所述用户归属于各个人群绩效等级的概率值的主要功能模块,也是核心模块。

所述确定单元 32,可以用于根据所述各个人群绩效等级的概率值,确定所述用户对应的人群绩效等级。所述确定单元 32 是本装置中根据所述各个人群绩效等级的概率值,确定所述用户对应的人群绩效等级的主要功能模块。

对于本申请实施例,为了确定为所述用户对应的人群绩效等级,所述确定单元 31,具体可以用于对所述各个人群绩效等级的概率值进行排序;将最高概率值对应的人群绩效等级,确定为所述用户对应的人群绩效等级。

对于本申请实施例,为了建立所述预设决策树模型和所述预设二次识别模型,所述装置还包括:获取单元 33 和建立单元 34,如图 4 所示。

所述获取单元 33,可以用于获取多个样本用户对应的样本人群绩效特征和样本人群绩效等级。所述获取单元 34 是本装置中获取多个样本用户对应的样本人群绩效特征和样本人群绩效等级的主要功能模块。

所述建立单元 34,可以用于根据所述样本人群绩效特征和所述样本人群绩效等级,建立所述预设决策树模型。所述建立单元 34 是本装置中根据所述样本人群绩效特征和所述样本人群绩效等级,建立所述预设决策树模型的主要功能模块。

所述获取单元 33,还可以用于获取所述预设决策树模型计算的所述多个样本用户归属于对应分类人群绩效等级的概率值。所述获取单元 34 还是本装置中获取所述预设决策树模型计算的所述多个样本用户归属于对应分类人群绩效等级的概率值的主要功能模块。

所述建立单元 34,还可以用于根据所述样本人群绩效特征和所述各个人样本群绩效等级的概率值,建立所述预设二次识别模型。所述建立单元 34 还是本装置中根据所述样本人群绩效特征和所述各个人样本群绩效等级的概率值,建立所述预设二次识别模型的主要功能模块。

在具体应用场景中,所述建立单元 34,具体可以用于当所述预设二次识别模型为预设逻辑回归模型时,利用预设逻辑回归算法对所述样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值进行训练,建立所述预设逻辑回归模型。

所述建立单元 34,具体还可以用于当所述预设二次识别模型为预设神经网络模型时,利用预设神经网络算法对所述样本人群绩效特征和所述多个样本用户归属于对应分类人

群绩效等级的概率值进行训练，建立所述预设神经网络模型。

所述获取单元 33，具体可以用于将所述样本人群绩效特征输入到所述预设决策树模型进行识别，得到所述用户归属于对应分类人群绩效等级的概率值。

对于本申请实施例，为了获取待识别用户对应的人群绩效特征，所述装置还可以包括：提取单元 35。

所述获取单元 33，还可以用于获取待识别用户对应的人群绩效数据。所述获取单元 33 还是本装置中获取待识别用户对应的人群绩效数据的主要功能模块。

所述提取单元 35，可以用于从所述人群绩效数据中提取出所述用户对应的人群绩效特征。所述提取单元 35 是本装置中从所述人群绩效数据中提取出所述用户对应的人群绩效特征的主要功能模块。

需要说明的是，本申请实施例提供的一种人群绩效等级识别装置所涉及各功能模块的其他相应描述，可以参考图 1 所示方法的对应描述，在此不再赘述。

基于上述如图 1 所示方法，相应的，本申请实施例还提供了一种计算机非易失性可读存储介质，其上存储有计算机可读指令，该计算机可读指令被处理器执行时实现以下步骤：将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；根据所述概率值，确定所述用户对应的人群绩效等级。

基于上述如图 1 所示方法和如图 3 所示人群绩效等级识别装置的实施例，本申请实施例还提供了一种计算机设备的实体结构图，如图 5 所示，该计算机设备包括：处理器 41、存储器 42、及存储在存储器 42 上并可在处理器上运行的计算机可读指令，其中存储器 42 和处理器 41 均设置在总线 43 上所述处理器 41 执行所述计算机可读指令时实现以下步骤：将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；根据所述概率值，确定所述用户对应的人群绩效等级。该计算机设备还包括：总线 43，被配置为耦接处理器 41 及存储器 42。

通过本申请的技术方案，能够将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值。与此同时，能够将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次

识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；并能够根据所述概率值，确定所述用户对应的人群绩效等级。从而能够实现在预设决策树模型的识别结果上，利用预设二次识别模型对所述用户进行二次识别，能够在一定程度上克服数据分布平衡的影响，减少预设决策树模型识别错误的情况，且能够在保证识别精确度的前提下，提升占比较低的目标人群召回率。

显然，本领域的技术人员应该明白，上述的本申请的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的计算机可读指令代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，并且在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本申请不限制于任何特定的硬件和软件结合。

以上所述仅为本申请的优选实施例而已，并不用于限制本申请，对于本领域的技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包括在本申请的保护范围之内。

权 利 要 求

1、一种人群绩效等级识别方法，其特征在于，包括：

将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述

5 用户归属于对应分类人群绩效等级的概率值；

将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；

根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

2、根据权利要求1所述的方法，其特征在于，所述根据所述各个人群绩效等级的概
10 率值，确定所述用户对应的人群绩效等级，包括：

对所述各个人群绩效等级的概率值进行排序；

将最高概率值对应的人群绩效等级，确定为所述用户对应的人群绩效等级。

3、根据权利要求1所述的方法，其特征在于，所述将待识别用户对应的人群绩效特
征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概
15 率值之前，所述方法还包括：

获取多个样本用户对应的样本人群绩效特征和样本人群绩效等级；

根据所述样本人群绩效特征和所述样本人群绩效等级，建立所述预设决策树模型，
并获取所述预设决策树模型计算的所述多个样本用户归属于对应分类人群绩效等级的概
率值；

20 根据所述样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概
率值，建立所述预设二次识别模型。

4、根据权利要求3所述的方法，其特征在于，所述预设二次识别模型为预设逻辑回
归模型，所述根据所述样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等
级的概率值，建立所述预设二次识别模型，包括：

25 利用预设逻辑回归算法对所述样本人群绩效特征和所述多个样本用户归属于对应分
类人群绩效等级的概率值进行训练，建立所述预设逻辑回归模型。

5、根据权利要求3所述的方法，其特征在于，所述预设二次识别模型为预设神经网络
模型，所述根据所述样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等
级的概率值，建立所述预设二次识别模型，包括：

30 利用预设神经网络算法对所述样本人群绩效特征和所述多个样本用户归属于对应分
类人群绩效等级的概率值进行训练，建立所述预设神经网络模型。

6、根据权利要求3所述的方法，其特征在于，所述获取所述预设决策树模型计算的所述多个样本用户归属于对应分类人群绩效等级的概率值，包括：

将所述样本人群绩效特征输入到所述预设决策树模型进行识别，得到所述用户归属于对应分类人群绩效等级的概率值。

5 7、根据权利要求1所述的方法，其特征在于，所述将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值之前，所述方法还包括：

获取待识别用户对应的人群绩效数据；

从所述人群绩效数据中提取出所述用户对应的人群绩效特征。

10 8、一种人群绩效等级识别装置，其特征在于，包括：

识别单元，用于将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；

所述识别单元，还用于将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；

15 确定单元，用于根据所述归属于各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

9、根据权利要求8所述的装置，其特征在于，所述确定单元，具体用于对所述各个人群绩效等级的概率值进行排序；将最高概率值对应的人群绩效等级，确定为所述用户对应的人群绩效等级。

20 10、根据权利要求8所述的装置，其特征在于，所述装置还包括：

所述获取单元，用于获取多个样本用户对应的样本人群绩效特征和样本人群绩效等级；

所述建立单元，用于根据所述样本人群绩效特征和所述样本人群绩效等级，建立所述预设决策树模型；

25 所述获取单元，还用于获取所述预设决策树模型计算的所述多个样本用户归属于对应分类人群绩效等级的概率值；

所述建立单元，还用于根据所述样本人群绩效特征和所述各个人样本群绩效等级的概率值，建立所述预设二次识别模型。

30 11、根据权利要求10所述的装置，其特征在于，所述建立单元，具体用于当所述预设二次识别模型为预设逻辑回归模型时，利用预设逻辑回归算法对所述样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值进行训练，建立所述预设逻辑

回归模型。

12、根据权利要求 10 所述的装置，其特征在于，所述建立单元，具体用于当所述预设二次识别模型为预设神经网络模型时，利用预设神经网络算法对所述样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值进行训练，建立所述预设神经网络模型。

13、根据权利要求 10 所述的装置，其特征在于，所述获取单元，具体用于将所述样本人群绩效特征输入到所述预设决策树模型进行识别，得到所述用户归属于对应分类人群绩效等级的概率值。

14、根据权利要求 8 所述的装置，其特征在于，所述装置还包括：
获取单元，用于获取待识别用户对应的人群绩效数据；
提取单元，用于从所述人群绩效数据中提取出所述用户对应的人群绩效特征。

15、一种计算机非易失性可读存储介质，其上存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现人群绩效等级识别方法，包括：

将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；

将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；

根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

16、根据权利要求 15 所述的计算机非易失性可读存储介质，其特征在于，所述计算机可读指令被处理器执行时实现所述根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级，包括：

对所述各个人群绩效等级的概率值进行排序；

将最高概率值对应的人群绩效等级，确定为所述用户对应的人群绩效等级。

17、根据权利要求 15 所述的计算机非易失性可读存储介质，其特征在于，所述计算机可读指令被处理器执行时实现所述将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值之前，所述方法还包括：

获取多个样本用户对应的样本人群绩效特征和样本人群绩效等级；

根据所述样本人群绩效特征和所述样本人群绩效等级，建立所述预设决策树模型，并获取所述预设决策树模型计算的所述多个样本用户归属于对应分类人群绩效等级的概率值；

根据所述样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值，建立所述预设二次识别模型。

18、一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现人群绩效等级识别方法，包括：

5 别方法，包括：

将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值；

将所述人群绩效特征和所述用户归属于对应分类人群绩效等级的概率值输入到预设二次识别模型进行二次识别，得到所述用户归属于各个人群绩效等级的概率值；

10 根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级。

19、根据权利要求 18 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时实现所述根据所述各个人群绩效等级的概率值，确定所述用户对应的人群绩效等级，包括：

对所述各个人群绩效等级的概率值进行排序；

15 将最高概率值对应的人群绩效等级，确定为所述用户对应的人群绩效等级。

20、根据权利要求 18 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时实现所述将待识别用户对应的人群绩效特征输入到预设决策树模型进行一次识别，得到所述用户归属于对应分类人群绩效等级的概率值之前，所述方法还包括：

获取多个样本用户对应的样本人群绩效特征和样本人群绩效等级；

20 根据所述样本人群绩效特征和所述样本人群绩效等级，建立所述预设决策树模型，并获取所述预设决策树模型计算的所述多个样本用户归属于对应分类人群绩效等级的概率值；

根据所述样本人群绩效特征和所述多个样本用户归属于对应分类人群绩效等级的概率值，建立所述预设二次识别模型。

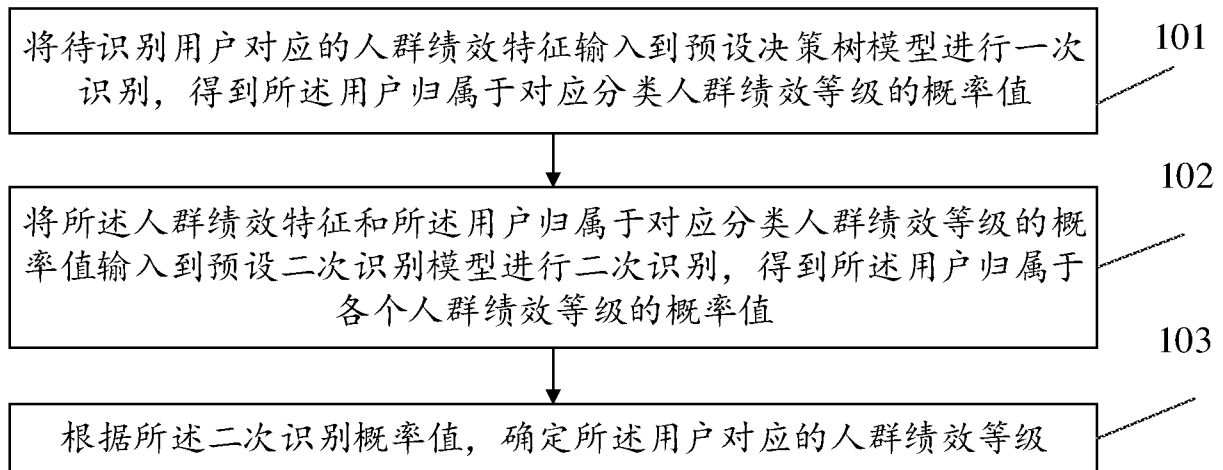


图 1

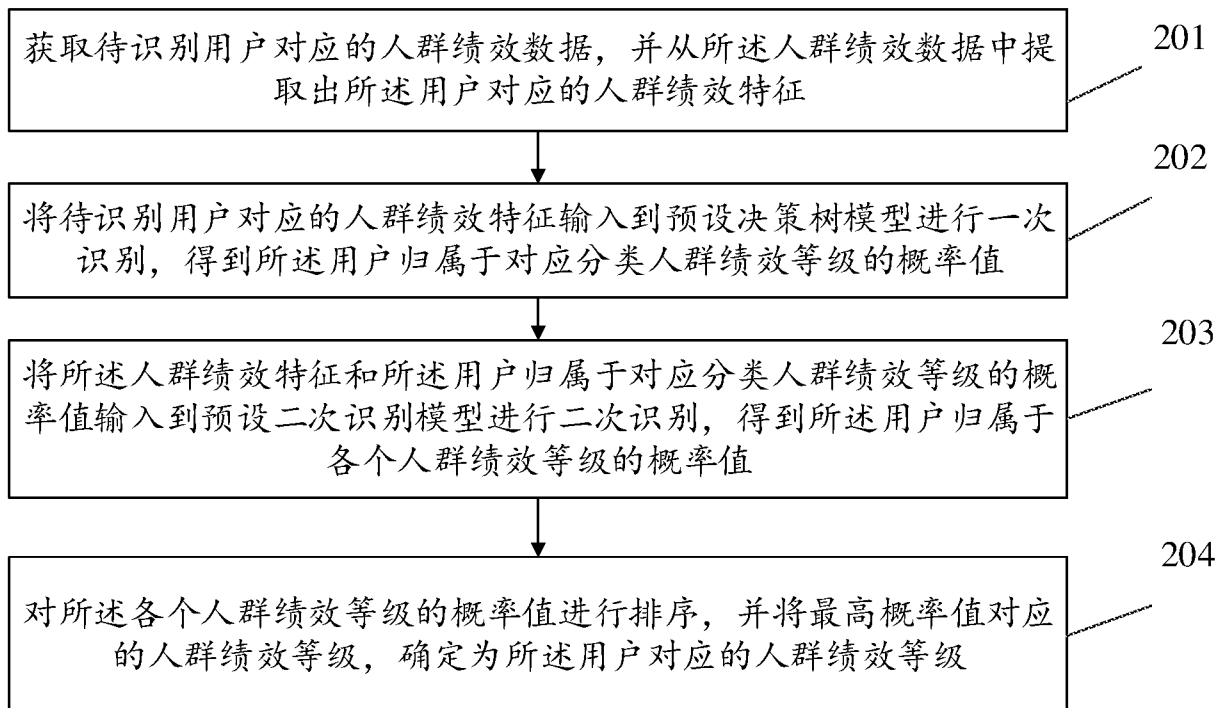


图 2

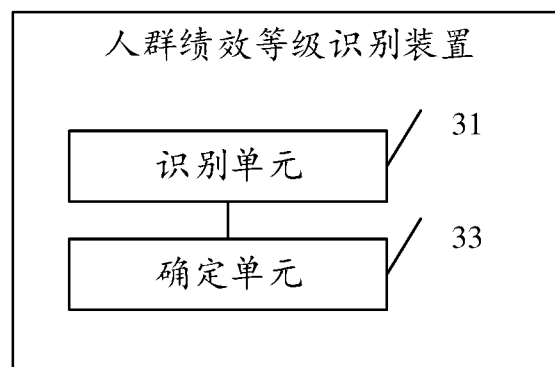


图 3

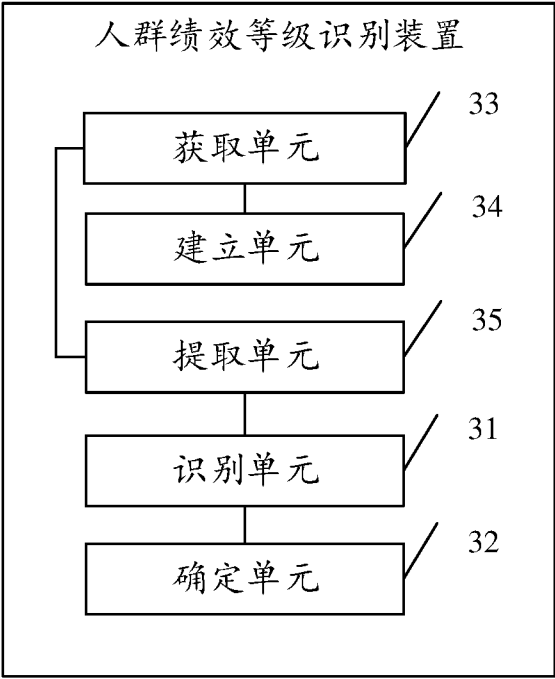


图 4

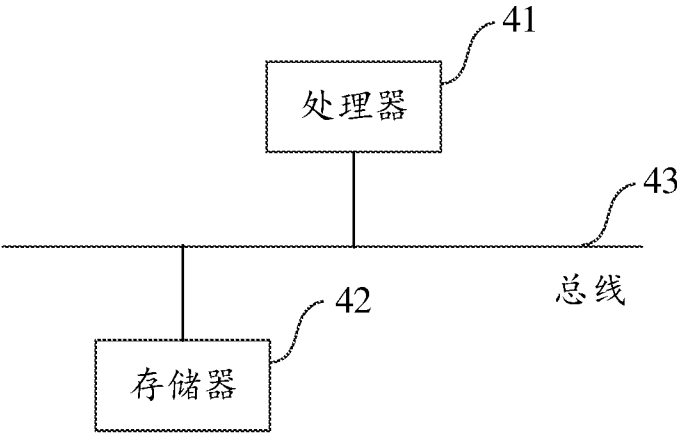


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/111124

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/06(2012.01)i; G06K 9/62(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q; G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, EPODOC, WPI, IEEE: 人群, 客户, 一次, 二次, 再次, 识别, 决策树, 逻辑回归, 神经网络, human, customer, one, two, identifying, model, decision tree, logistic regression, neural network

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108073883 A (SHENZHEN INTELLIFUSION TECHNOLOGY CO., LTD.) 25 May 2018 (2018-05-25) description, paragraphs [0032]-[0082], and figure 1	1-20
A	CN 108256691 A (CHENGDU ZHIBAO BIG DATA TECHNOLOGY CO., LTD.) 06 July 2018 (2018-07-06) entire document	1-20
A	CN 105225135 A (GUANGZHOU HUADUO NETWORK TECHNOLOGY CO., LTD.) 06 January 2016 (2016-01-06) entire document	1-20
A	CN 104123395 A (BEIJING CYCLE CENTURY DIGITAL TECHNOLOGY CO., LTD.) 29 October 2014 (2014-10-29) entire document	1-20

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

10 April 2019

Date of mailing of the international search report

28 April 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/111124

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	108073883	A	25 May 2018	None	
CN	108256691	A	06 July 2018	None	
CN	105225135	A	06 January 2016	None	
CN	104123395	A	29 October 2014	None	

国际检索报告

国际申请号

PCT/CN2018/111124

A. 主题的分类 G06Q 10/06 (2012.01)i; G06K 9/62 (2006.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																	
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) G06Q; G06K 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNPAT, CNKI, EPDOC, WPI, IEEE:人群, 客户, 一次, 二次, 再次, 识别, 决策树, 逻辑回归, 神经网络, human, customer, one, two, identifying, model, decision tree, logistic regression, neural network																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 108073883 A (深圳云天励飞技术有限公司) 2018年 5月 25日 (2018 - 05 - 25) 说明书第[0032]-[0082]段, 附图1</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108256691 A (成都智宝大数据科技有限公司) 2018年 7月 6日 (2018 - 07 - 06) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 105225135 A (广州华多网络科技有限公司) 2016年 1月 6日 (2016 - 01 - 06) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 104123395 A (北京赛科世纪数码科技有限公司) 2014年 10月 29日 (2014 - 10 - 29) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 108073883 A (深圳云天励飞技术有限公司) 2018年 5月 25日 (2018 - 05 - 25) 说明书第[0032]-[0082]段, 附图1	1-20	A	CN 108256691 A (成都智宝大数据科技有限公司) 2018年 7月 6日 (2018 - 07 - 06) 全文	1-20	A	CN 105225135 A (广州华多网络科技有限公司) 2016年 1月 6日 (2016 - 01 - 06) 全文	1-20	A	CN 104123395 A (北京赛科世纪数码科技有限公司) 2014年 10月 29日 (2014 - 10 - 29) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 108073883 A (深圳云天励飞技术有限公司) 2018年 5月 25日 (2018 - 05 - 25) 说明书第[0032]-[0082]段, 附图1	1-20															
A	CN 108256691 A (成都智宝大数据科技有限公司) 2018年 7月 6日 (2018 - 07 - 06) 全文	1-20															
A	CN 105225135 A (广州华多网络科技有限公司) 2016年 1月 6日 (2016 - 01 - 06) 全文	1-20															
A	CN 104123395 A (北京赛科世纪数码科技有限公司) 2014年 10月 29日 (2014 - 10 - 29) 全文	1-20															
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																	
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																	
国际检索实际完成的日期 2019年 4月 10日		国际检索报告邮寄日期 2019年 4月 28日															
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 白桦 电话号码 86-(10)-53961425															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/111124

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	108073883	A	2018年 5月 25日	无	
CN	108256691	A	2018年 7月 6日	无	
CN	105225135	A	2016年 1月 6日	无	
CN	104123395	A	2014年 10月 29日	无	