

(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 3 月 5 日 (05.03.2020)



(10) 国际公布号
WO 2020/042580 A1

(51) 国际专利分类号:
G06Q 10/06 (2012.01)

(21) 国际申请号: PCT/CN2019/077227

(22) 国际申请日: 2019 年 3 月 6 日 (06.03.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810982127.6 2018 年 8 月 27 日 (27.08.2018) CN

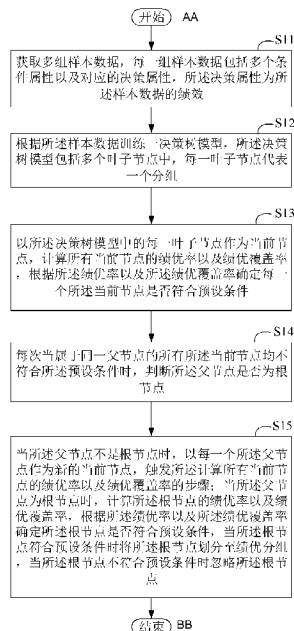
(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 邓悦(DENG, Yue); 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)。 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: 深圳市赛恩倍吉知识产权代理有限公司(SHENZHEN SCIENBIZIP INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深

(54) Title: PERSONNEL GROUPING METHOD AND DEVICE, ELECTRONIC DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 人员分组方法、装置、电子装置及存储介质



S11 Obtain a plurality of sets of sample data, each set of sample data comprising a plurality of condition attributes and corresponding decision attributes, and the decision attributes being performance of the sample data

S12 Train a decision tree model, in a plurality of leaf nodes comprised in the decision tree model, each leaf node representing a group

S13 Taking each leaf node of the decision tree model as the current node, calculate the excellent performance rate and the excellent performance coverage rate of all the current nodes, and determine whether each of the current nodes meets a preset condition according to the excellent performance rate and the excellent performance coverage rate

S14 Each time all the current nodes belonging to the same parent node do not meet the preset condition, determine whether the parent node is a root node

S15 When the parent node is not the root node, taking each of the parent nodes as the new current node, trigger the step of calculating the excellent performance rate and the excellent performance coverage rate of all the current nodes; when the parent node is the root node, calculate the excellent performance ratio and the excellent performance coverage rate of the root node, determine whether the root node meets the preset condition according to the excellent performance ratio and the excellent performance coverage rate, classify the root node into excellent performance groups when the root node meets the preset condition, and ignore the root node when the root node does not meet the preset condition

AA Start

BB End

图 1

(57) Abstract: A personnel grouping method and device, an electronic device, and a storage medium. The method comprises: obtaining a plurality of sets of sample data; training a decision tree model; taking each leaf node of the decision tree model as the current node, calculating the excellent performance rate and the excellent performance coverage rate of all the current nodes, and determining whether each of the current nodes meets a preset condition; when all the current nodes belonging to the same parent node do not meet the preset condition, determining whether the parent node is a root node; when the parent node is the root node, calculating the excellent

圳市龙华新区龙观东路83号荣群大厦9楼, Guangdong 518109 (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, 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))。

performance ratio and the excellent performance coverage rate of the root node, determining whether the root node meets the preset condition according to the excellent performance ratio and the excellent performance coverage rate, classifying the root node into excellent performance groups when the root node meets the preset condition, and ignoring the root node when the root node does not meet the preset condition. According to the method, distinguishing features of excellent performers are identified by means of condition attributes of an excellent performance group, so that a user portrait is constructed.

(57) 摘要: 一种人员分组方法、装置、电子装置及存储介质, 所述方法包括: 获取多组样本数据; 训练决策树模型; 以决策树模型每一叶子节点作为当前节点, 计算所有当前节点的绩优率及绩优覆盖率, 确定每一当前节点是否符合预设条件; 当属于同一父节点的所有当前节点均不符合所述预设条件时, 判断所述父节点是否为根节点; 当所述父节点为根节点时, 计算所述根节点的绩优率以及绩优覆盖率, 根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件, 当所述根节点符合预设条件时将所述根节点划分至绩优分组, 当所述根节点不符合预设条件时忽略所述根节点。该方法通过绩优分组的条件属性识别绩优人员的显著特征, 从而构建用户画像。

人员分组方法、装置、电子装置及存储介质

本申请要求于 2018 年 08 月 27 日提交中国专利局，申请号为 201810982127.6 发明名称为“人员分组方法、装置、电子装置及计算机可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及分组方法，具体涉及一种人员分组方法、人员分组装置、电子装置以及存储介质。

背景技术

企业通常拥有海量的人员基础数据和行为数据。企业通常期望搭建用户画像，对绩优人员的显著特征进行识别。现有技术中，通常采用无监督学习的聚类算法对无序的对象进行分群、归类，已达到更好地理解研究对象的目的。然而，利用该类方法建立用户画像存在缺陷，如，分群目标不明确，分群效果不可控，分群结果不可解释，且各分群的重要特征无法识别等。

发明内容

鉴于以上内容，有必要提出一种人员分组方法、人员分组装置、电子装置以及存储介质，能够解决以上问题。

本申请一较佳实施方式提供一种人员分组方法，包括：获取多组样本数据，每一组样本数据包括多个条件属性以及对应的决策属性，所述决策属性为所述样本数据的绩效；根据所述样本数据训练一决策树模型，其中，所述决策树模型包括多个叶子节点，每一叶子节点代表一个分组；以所述决策树模型中的每一叶子节点作为当前节点，计算所有当前节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件；当属于同一父节点的所有所述当前节点均不符合所述预设条件时，判断所述父节点是否为根节点；当所述父节点不是根节点时，以每一个所述父节点作为新的当前节点，触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤；当所述父节点为根节点时，计算所述根节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件，当所述根节点符合预设条件时将所述根节点划分至绩优分组，当所述根节点不符合预设条件时忽略所述根节点；针对每一个所述当前节点，执行以下步骤：将所述当前节点作为目标节点，当所述目标节点符合所述预设条件时，将所述目标节点划分至所述绩优分组；当所述目标节点不符合所述预设条件时，判断与所述目标节点属于同一父节点的其它同层节点是否符合所述预设条件；当属于同一父节点的至少一其它同层节点符合所述预设条件时，将所述目标节点划分至绩差分组；当属于同一父节点的所有其它同层节点均不符合所述预设条件时，忽略所述目标节点。

本申请一较佳实施方式还提供一种人员分组装置，包括：获取模块，用于获取多组样本数据，每一组样本数据包括多个条件属性以及对应的决策属性，所述决策属性为所述样本数据的绩效；训练模块，用于根据所述样本数据训练一决策树模型，其中，所述决策树模型包括多个叶子节点，每一叶子节点代表一个分组；执行模块，用于计算所有当前节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件；当属于同一父节点的所有所述当前节点均不符合所述预设条件时，所述执行模块判断所述父节点是否为根节点；当所述父节点不是根节点时，所述执行模块以每一个所述父节点作为新的当前节点，触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤；当所述父节点为根节点时，所述执行模块计算所述根节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件，当所述根节点符合预设条件时将所述根节点划分至绩优分组，当所述根节点不符合预设条件时忽略所述根节点；针对每一个所述当前节点，所述执行模块用于：将所述当前节点作为目标节点，当所述目标节点符合所述预设条件时，将所述目标节点划分至所述绩优分组；当所述目标节点不符合所述预设条件时，判断与所述目标节点属于同一父节点的其它同层节点是否符合所述预设条件；当属于同一父节点的至少一其它同层节点符合所述预设条件时，将所述目标节点划分至绩差分组；当属于同一父节点的所有其它同层节点均不符合所述预设条件时，忽略所述目标节点。

本申请一较佳实施方式还提供一种电子装置，包括处理器和存储器，所述存储器中存储有至少一个计算机可读指令，所述处理器用于执行所述计算机可读指令以实现如前所述的人员分组方法。

本申请一较佳实施方式还提供一种非易失性可读存储介质，所述非易失性可读存储介质上存储至少一个计算机可读指令，所述计算机可读指令被处理器执行时实现如前所述的人员分组方法。

本申请实施例在决策树建模中加入绩差人员的信息，使得分组目标更明确，且数据的执行效果更好，保证最终获得的绩优分组为绩优人数比例达到一定程度的群体；而且，可通过绩优分组的条件属性识别绩优人员的显著特征，从而构建用户画像。

附图说明

图 1 是本申请一较佳实施例提供的人员分组方法的流程图。

图 2 是图 1 的人员分组方法所使训练的决策树的示意图。

图 3 是本申请一较佳实施例提供的人员分组装置的结构示意图。

图 4 为本申请一较佳实施例提供的电子装置的结构示意图。

具体实施方式

图 1 是本申请一较佳实施例提供的人员分组方法的流程图。所述人员分组方法应用于一电子装置 1 中。根据不同需求，所述人员分组方法的步骤顺序可以改变，某些步骤可以省略或合并。所述人员分组方法包括以下步骤：

步骤 S11，获取多组样本数据，每一组样本数据包括多个条件属性以及对

应的决策属性，所述决策属性为所述样本数据的绩效。

其中，所述样本数据需同时包括绩效较优的人员(即，绩优人员)和绩效较差的人员(即，绩差人员)的数据，所述样本数据可存储于所述电子装置中。在另一实施方式中，所述电子装置还可以通过接入网络连接之一外部样本库，进而来获取所述样本库存储的样本数据。在其它实施方式中，所述电子装置还可以通过大数据方式来收集并建立所述样本库。

在本实施方式中，所述条件属性可为行为轨迹(如，出差情况)、app 活跃情况、业务扩展情况、消费情况、兴趣爱好、参加培训情况、考勤率等。以所述条件属性包括出差情况、业务扩展情况、参加培训情况以及考勤率为例进行说明，所述样本数据如表 1 所示。

表 1 样本数据

条件属性				决策属性
出差情况(C)	业务扩展情况(Y)	参加培训情况(P)	考勤率(K)	绩效
每月出差 0 次	每月增加 0 件业务	每月 0 次	<80%	差
每月出差 0 次	每月增加 0 件业务	每月 0 次	≥80%	优
每月出差 0 次	每月增加 1 件业务	每月 1 次	≥80%	优
每月出差 0 次	每月增加 ≥2 件业务	每月 0 次	≥80%	优
每月出差 0 次	每月增加 0 件业务	每月 ≥2 次	≥80%	优
每月出差 1 次	每月增加 0 件业务	每月 0 次	<80%	差
每月出差 1 次	每月增加 1 件业务	每月 0 次	<80%	优
每月出差 1 次	每月增加 0 件业务	每月 1 次	<80%	差
每月出差 1 次	每月增加 1 件业务	每月 0 次	≥80%	优
每月出差 ≥2 次	每月增加 1 件业务	每月 0 次	<80%	差
每月出差 ≥2 次	每月增加 1 件业务	每月 1 次	<80%	优
每月出差 ≥2 次	每月增加 1 件业务	每月 1 次	≥80%	优
每月出差 ≥2 次	每月增加 0 件业务	每月 ≥2 次	<80%	优
每月出差 ≥2 次	每月增加 0 件业务	每月 1 次	≥80%	优
每月出差 ≥2 次	每月增加 0 件业务	每月 1 次	≥80%	优
每月出差 ≥2 次	每月增加 0 件业务	每月 0 次	<80%	差

步骤 S12, 根据所述样本数据训练一决策树模型, 其中, 所述决策树模型包括多个叶子节点中, 每一叶子节点代表一个分组。

其中, 决策树算法属于监督学习分类算法, 而决策树模型代表对象属性与对象值之间的一种映射关系。其中, 所述决策树模型具有 N 层(N 为自然数, $N > 2$), 所述电子装置根据所述条件属性设置所述决策树模型中的每个节点(包括根节点、内部节点和叶子节点)的值。所述决策树模型的第一层为根节点, 所述决策树模型的第二层代表以第一级条件属性对所述根节点进行分割得到的多个节点, 所述决策树模型的第三层代表以第二级条件属性对上一层节点进行分割得到的多个节点, 等等。如图 2 所示, 所述决策树模型包括四层, 即, $N=4$ 。

在本实施方式中, 训练所述决策树模型具体包括:

步骤 S121, 以所述样本数据作为训练集计算每一条件属性的信息增益(information gain)。

步骤 S122, 选择信息增益最大的条件属性作为所述决策树模型的根节点以分割所述样本数据以得到下一层节点; 其中, 某一条件属性的信息增益越大, 说明选择所述条件属性对分类提供的信息越多, 越有利于确定性, 越有利于将所述样本数据进行分类。如图 2 所示, 若计算信息增益最大的条件属性为“出差情况”, 则选择“出差情况”条件属性作为根节点以分割所述样本数据。

步骤 S123, 以每一节点包含的样本数据作为训练集重新计算每一条件属性的信息增益。

步骤 S124, 选择信息增益最大的条件属性分割所述节点以得到下一层节点。如图 2 所示, 对于包含每月出差情况为 0 次($C=0$)的样本的节点, 若计算信息增益最大的条件属性为“业务拓展情况”, 则根据“业务拓展情况”条件属性对所述节点进行分割而获得下一层节点。对于包含每月出差情况为 ≥ 2 次($C \geq 2$)的样本的节点, 若计算信息增益最大的条件属性为“参加培训情况”, 则根据“参加培训情况”条件属性对所述节点进行分割而获得下一层节点。实际训练过程中, 对位于同一层的多个节点进行分割所采用的条件属性通常不同。每一次分割后形成的节点比上一层节点的数据纯度更高。

步骤 S125, 递归执行步骤 S123 以及 S124, 直至分割停止。此时, 每一叶节点代表一个分组, 每一分组包含的绩优人员数量与绩差人员数量之间具有固定的比值, 而多个分组包含的绩优人员数量与绩差人员数量之间的比值可以互不相同。如, 叶子节点“分组 1”的所述比值可为 1:8。当然, 某一叶节点包含的可均为绩优人员, 也可以均为绩差人员。其中, 每一组样本数据只能被划分至其中一个叶子节点中, 即, 每一组样本数据不能同属于两个及两个以上的叶子节点。

在本实例中, 在所述决策树的层数达到一预设层数(如: 4 层)时停止下一次分割。在另一实施例, 在所述决策树的当前层的每一节点包含的样本数量小于一预设数量时停止下一次分割。

其中，每一条件属性的信息增益 $Gain(S, A)$ 可通过如下方式计算：

$$Gain(S, A) = Entropy(S) - Entropy(S, A)$$

其中， $Gain(S, A)$ 表示 A 条件属性在训练集 S 上的信息增益， $Entropy(S)$ 表示训练集 S 的信息熵， $Entropy(S, A)$ 表示 A 属性的信息熵。

例如，如表 1 所示，在步骤 S121 中，所述训练集共有样本数据 16 组，绩优的有 11 组，绩差的有 5 组。以“出差情况”条件属性的信息增益计算为例进行说明，其它条件属性的信息增益计算过程同理。对于 $C=0$ 的情况，共有样本数据 5 组，绩优的有 4 组，绩差的有 1 组；对于 $C=1$ 的情况，共有样本数据 4 组，绩优的有 2 组，绩差的有 2 组；对于 $C \geq 2$ 的情况，共有样本数据 7 组，绩优的有 5 组，绩差的有 2 组。因此，“出差情况”条件属性的信息增益的计算方式如下：

$$\begin{aligned} Entropy(S) &= -\frac{11}{16} \log_2 \frac{11}{16} - \frac{5}{16} \log_2 \frac{5}{16} \\ Entropy(A) &= \frac{4}{16} \times (-\frac{4}{5} \log_2 \frac{4}{5} - \frac{1}{5} \log_2 \frac{1}{5}) + \frac{4}{16} \times (-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}) \\ &\quad + \frac{7}{16} \times (-\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7}) \\ Gain(S, A) &= -\frac{11}{16} \log_2 \frac{11}{16} - \frac{5}{16} \log_2 \frac{5}{16} - \frac{4}{16} \times (-\frac{4}{5} \log_2 \frac{4}{5} - \frac{1}{5} \log_2 \frac{1}{5}) \\ &\quad - \frac{4}{16} \times (-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}) - \frac{7}{16} \times (-\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7}) \end{aligned}$$

步骤 S13，以所述决策树模型中的每一叶子节点作为当前节点，计算所有当前节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件。

针对每一个所述当前节点，执行以下步骤：

将所述当前节点作为目标节点，当所述目标节点符合所述预设条件时，将所述目标节点划分至所述绩优分组；

当所述目标节点不符合所述预设条件时，判断与所述目标节点属于同一父节点的其它同层节点是否符合所述预设条件；

当属于同一父节点的至少一其它同层节点符合所述预设条件时，将所述目标节点划分至绩差分组；

当属于同一父节点的所有其它同层节点均不符合所述预设条件时，忽略所述目标节点。即，既不将所述目标节点划分至所述绩优分组，也不将所述目标节点划分至所述绩差分组。

其中，每一节点的绩优率为所述节点的绩优人员数量与所述节点的待分组人员的总数量之间的比值，所述节点的绩优率可通过所述节点包含的绩优人员数量与绩差人员数量之间的比值进行计算。每一节点的绩优覆盖率为所述节点的绩优人员数量与整个决策树模型的绩优人员的总数量之间的比值。当所述节点的绩优率以及绩优覆盖率分别达到一绩优率阈值与一绩优覆盖率阈值时，确定所述节点符合所述预设条件。反之，当所述节点的绩优率未达

到所述绩优率阈值，或是，所述节点的绩优覆盖率未达到所述绩优覆盖率阈值，又或是，所述节点的绩优率以及绩优覆盖率均未达到所述绩优率阈值与所述绩优覆盖率阈值时，确定所述节点不符合所述预设条件。

例如，设置所述绩优率阈值为 40%，所述绩优覆盖率阈值为 20%。若计算出节点“分组 8”的绩优率为 80%，绩优覆盖率为 20%，则节点“分组 8”符合预设条件，被划分至绩优分组。对于节点“分组 7”，若绩优覆盖率为 30%，绩优覆盖率为 25%，则节点“分组 7”不符合预设条件，但是由于与所述节点“分组 7”属于同一父节点的同层节点“分组 8”预设条件，因此节点“分组 7”被划分至绩差分组。反之，若节点“分组 7”与节点“分组 8”均不符合预设条件，则忽略节点“分组 7”以及节点“分组 8”。

在本实施方式中，所述绩优率阈值与所述绩优覆盖率阈值预先存储于所述电子装置中。所述绩优率阈值与所述绩优覆盖率阈值可根据实际需要进行调整。

步骤 S14，每次当属于同一父节点的所有所述当前节点均不符合所述预设条件时，判断所述父节点是否为根节点。

步骤 S15，当所述父节点不是根节点时，以每一个所述父节点作为新的当前节点，触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤；当所述父节点为根节点时，计算所述根节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件，当所述根节点符合预设条件时将所述根节点划分至绩优分组，当所述根节点不符合预设条件时忽略所述根节点。

根据上述例子，若节点“分组 7”与节点“分组 8”均不符合预设条件，则需要考虑父节点是否为根节点。由于节点“分组 7”与节点“分组 8”的父节点并非根节点，则继续判断所述父节点的所述绩优率以及所述绩优覆盖率是否符合预设条件。可以理解，当所有的内部节点和叶子节点均不符合预设条件时，才对所述根节点进行判断。

其中，通过调整所述绩优率阈值与所述绩优覆盖率阈值，可控制所述绩优分组与所述绩差分组的大小和质量。

其中，最终得到的绩优分组可包括具有不同特征的绩优人群。如，若“分组 4”被划分至所述绩优分组，则该部分绩优人群涉及的条件属性包括：出差情况=0 次、业务拓展情况 ≥ 1 次，考勤率 ≥ 80 。也就是说，该部分绩优人群的显著特征为：不出差、业务能力强、出勤情况较好。

若“分组 8”被划分至所述绩优分组，则该部分绩优人群涉及的条件属性包括：出差情况为 ≥ 2 次、参加培训情况 ≥ 1 次，考勤率 ≥ 80 。也就是说，该部分绩优人群的显著特征为：经常出差、乐于学习、出勤情况较好。

图 3 为本申请一较佳实施方式提供的人员分组装置 300 的结构示意图。在一些实施例中，所述人员分组装置 300 运行于电子装置中。所述人员分组装置 300 可以包括多个由程序代码段所组成的功能模块。所述人员分组装置

300 的各个程序段的程序代码可以存储于电子装置的存储器中，并由所述至少一个处理器所执行，以实现人员分组功能。

本实施例中，所述人员分组装置 300 根据其所执行的功能，可以被划分为多个功能模块。如图 3 所示，所述人员分组装置 300 包括：获取模块 301、训练模块 302、执行模块 303。本申请所称的模块是指一种能够被至少一个处理器所执行并且能够完成固定功能的一系列计算机可读指令段，其存储在存储器中。在本实施例中，关于各模块的功能将在后续的实施例中详述。

所述获取模块 301 用于获取多组样本数据，每一组样本数据包括多个条件属性以及对应的决策属性，所述决策属性为所述样本数据的绩效。

其中，所述样本数据需同时包括绩效较优的人员(即，绩优人员)和绩效较差的人员(即，绩差人员)的数据，所述样本数据可存储于所述电子装置中。在另一实施方式中，所述电子装置还可以通过接入网络连接之一外部样本库，进而来获取所述样本库存储的样本数据。在其它实施方式中，所述电子装置还可以通过大数据方式来收集并建立所述样本库。

在本实施方式中，所述条件属性可为行为轨迹(如，出差情况)、app 活跃情况、业务扩展情况、消费情况、兴趣爱好、参加培训情况、考勤率等。以所述条件属性包括出差情况、业务扩展情况、参加培训情况以及考勤率为例进行说明，所述样本数据如上表 1 所示。

所述训练模块 302 用于根据所述样本数据训练一决策树模型，其中，所述决策树模型包括多个叶子节点，每一叶子节点代表一个分组。

其中，决策树算法属于监督学习分类算法，而决策树模型代表对象属性与对象值之间的一种映射关系。其中，所述决策树模型具有 N 层(N 为自然数， $N > 2$)，所述训练模块 302 根据所述条件属性设置所述决策树模型中的每个节点(包括根节点、内部节点和叶子节点)的值。所述决策树模型的第一层为根节点，所述决策树模型的第二层代表以第一级条件属性对所述根节点进行分割得到的多个节点，所述决策树模型的第三层代表以第二级条件属性对上一层节点进行分割得到的多个节点，等等。如图 2 所示，所述决策树模型包括四层，即， $N=4$ 。

在本实施方式中，所述训练模块 302 以所述样本数据作为训练集计算每一条件属性的信息增益(information gain)，并选择信息增益最大的条件属性作为所述决策树模型的根节点以分割所述样本数据以得到下一层节点。其中，某一条件属性的信息增益越大，说明选择所述条件属性对分类提供的信息越多，越有利于确定性，越有利于将所述样本数据进行分类。如图 2 所示，若计算信息增益最大的条件属性为“出差情况”，则选择“出差情况”条件属性作为根节点以分割所述样本数据。

所述训练模块 302 进一步以每一节点包含的样本数据作为训练集重新计算每一条件属性的信息增益，并选择信息增益最大的条件属性分割所述节点以得到下一层节点。如图 2 所示，对于包含每月出差情况为 0 次($C=0$)的样本的节点，若计算信息增益最大的条件属性为“业务拓展情况”，则根据“业务拓展情况”条件属性对所述节点进行分割而获得下一层节点。对于包含每月出差

情况为 ≥ 2 次($C \geq 2$)的样本的节点, 若计算信息增益最大的条件属性为“参加培训情况”, 则根据“参加培训情况”条件属性对所述节点进行分割而获得下一层节点。实际训练过程中, 对位于同一层的多个节点进行分割所采用的条件属性通常不同。每一次分割后形成的节点比上一层节点的数据纯度更高。

所述训练模块 302 进一步递归执行每一条件属性的信息增益的计算步骤以及选择信息增益最大的条件属性分割所述节点以得到下一层节点的步骤, 直至分割停止。此时, 每一叶节点代表一个分组, 每一分组包含的绩优人员数量与绩差人员数量之间具有固定的比值, 而多个分组包含的绩优人员数量与绩差人员数量之间的比值可以互不相同。如, 叶子节点“分组 1”的所述比值可为 1:8。当然, 某一叶节点包含的可均为绩优人员, 也可以均为绩差人员。其中, 每一组样本数据只能被划分至其中一个叶子节点中, 即, 每一组样本数据不能同属于两个及两个以上的叶子节点。

在本实例中, 在所述决策树的层数达到一预设层数(如: 4 层)时停止下一次分割。在另一实施例中, 在所述决策树的当前层的每一节点包含的样本数量小于一预设数量时停止下一次分割。

其中, 每一条件属性的信息增益 $Gain(S, A)$ 可通过如下方式计算:

$$Gain(S, A) = Entropy(S) - Entropy(S, A)$$

其中, $Gain(S, A)$ 表示 A 条件属性在训练集 S 上的信息增益, $Entropy(S)$ 表示训练集 S 的信息熵, $Entropy(S, A)$ 表示 A 属性的信息熵。

例如, 如表 1 所示, 在所述训练模块 302 选择根节点的过程中, 所述训练集共有样本数据 16 组, 绩优的有 11 组, 绩差的有 5 组。以“出差情况”条件属性的信息增益计算为例进行说明, 其它条件属性的信息增益计算过程同理。对于 $C=0$ 的情况, 共有样本数据 5 组, 绩优的有 4 组, 绩差的有 1 组; 对于 $C=1$ 的情况, 共有样本数据 4 组, 绩优的有 2 组, 绩差的有 2 组; 对于 $C \geq 2$ 的情况, 共有样本数据 7 组, 绩优的有 5 组, 绩差的有 2 组。因此, “出差情况”条件属性的信息增益的计算方式如下:

$$\begin{aligned} Entropy(S) &= -\frac{11}{16} \log_2 \frac{11}{16} - \frac{5}{16} \log_2 \frac{5}{16} \\ Entropy(A) &= \frac{4}{16} \times (-\frac{4}{5} \log_2 \frac{4}{5} - \frac{1}{5} \log_2 \frac{1}{5}) + \frac{4}{16} \times (-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}) \\ &\quad + \frac{7}{16} \times (-\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7}) \\ Gain(S, A) &= -\frac{11}{16} \log_2 \frac{11}{16} - \frac{5}{16} \log_2 \frac{5}{16} - \frac{4}{16} \times (-\frac{4}{5} \log_2 \frac{4}{5} - \frac{1}{5} \log_2 \frac{1}{5}) \\ &\quad - \frac{4}{16} \times (-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}) - \frac{7}{16} \times (-\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7}) \end{aligned}$$

所述执行模块 303 用于以所述决策树模型中的每一叶子节点作为当前节点, 计算所有当前节点的绩优率以及绩优覆盖率, 根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件。

针对每一个所述当前节点, 所述执行模块 303 用于:

将所述当前节点作为目标节点，当所述目标节点符合所述预设条件时，将所述目标节点划分至所述绩优分组；

当所述目标节点不符合所述预设条件时，判断与所述目标节点属于同一父节点的其它同层节点是否符合所述预设条件；

当属于同一父节点的至少一其它同层节点符合所述预设条件时，将所述目标节点划分至绩差分组；

当属于同一父节点的所有其它同层节点均不符合所述预设条件时，忽略所述目标节点。即，既不将所述目标节点划分至所述绩优分组，也不将所述目标节点划分至所述绩差分组。

其中，每一节点的绩优率为所述节点的绩优人员数量与所述节点的待分组人员的总数量之间的比值，所述节点的绩优率可通过所述节点包含的绩优人员数量与绩差人员数量之间的比值进行计算。每一节点的绩优覆盖率为所述节点的绩优人员数量与整个决策树模型的绩优人员的总数量之间的比值。当所述节点的绩优率以及绩优覆盖率分别达到一绩优率阈值与一绩优覆盖率阈值时，确定所述节点符合所述预设条件。反之，当所述节点的绩优率未达到所述绩优率阈值，或是，所述节点的绩优覆盖率未达到所述绩优覆盖率阈值，又或是，所述节点的绩优率以及绩优覆盖率均未达到所述绩优率阈值与所述绩优覆盖率阈值时，确定所述节点不符合所述预设条件。

例如，设置所述绩优率阈值为 40%，所述绩优覆盖率阈值为 20%。若计算出节点“分组 8”的绩优率为 80%，绩优覆盖率为 20%，则节点“分组 8”符合预设条件，被划分至绩优分组。对于节点“分组 7”，若绩优覆盖率为 30%，绩优覆盖率为 25%，则节点“分组 7”不符合预设条件，但是由于与所述节点“分组 7”属于同一父节点的同层节点“分组 8”符合预设条件，因此节点“分组 7”被划分至绩差分组。反之，若节点“分组 7”与节点“分组 8”均不符合预设条件，则忽略节点“分组 7”以及节点“分组 8”。

在本实施方式中，所述绩优率阈值与所述绩优覆盖率阈值预先存储于所述电子装置中。所述绩优率阈值与所述绩优覆盖率阈值可根据实际需要进行调整。

所述执行模块 303 还用于每次当属于同一父节点的所有所述当前节点均不符合所述预设条件时，判断所述父节点是否为根节点。

当所述父节点不是根节点时，所述执行模块 303 以每一个所述父节点作为新的当前节点，触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤；当所述父节点为根节点时，计算所述根节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件，当所述根节点符合预设条件时将所述根节点划分至绩优分组，当所述根节点不符合预设条件时忽略所述根节点。

根据上述例子，若节点“分组 7”与节点“分组 8”均不符合预设条件，则需要考虑父节点是否为根节点。由于节点“分组 7”与节点“分组 8”的父节点并非根节点，则继续判断所述父节点的所述绩优率以及所述绩优覆盖率是否符合预设条件。可以理解，当所有的内部节点和叶子节点均不符合预设条件时，

才对所述根节点进行判断。

其中，通过调整所述绩优率阈值与所述绩优覆盖率阈值，可控制所述绩优分组与所述绩差分组的大小和质量。

其中，最终得到的绩优分组可包括具有不同特征的绩优人群。如，若“分组 4”被划分至所述绩优分组，则该部分绩优人群涉及的条件属性包括：出差情况=0 次、业务拓展情况 ≥ 1 次，考勤率 ≥ 80 。也就是说，该部分绩优人群的显著特征为：不出差、业务能力强、出勤情况较好。

若“分组 8”被划分至所述绩优分组，则该部分绩优人群涉及的条件属性包括：出差情况为 ≥ 2 次、参加培训情况 ≥ 1 次，考勤率 ≥ 80 。也就是说，该部分绩优人群的显著特征为：经常出差、乐于学习、出勤情况较好。

如图 4 所示，图 4 是本申请一较佳实施方式中实现所述人员分组方法的电子装置 1 的结构示意图。所述电子装置 1 包括存储器 101、处理器 102 以及存储于所述存储器 101 中并可在所述处理器 102 上运行的计算机可读指令 103，例如人员分组程序。

所述处理器 102 执行所述计算机可读指令 103 时实现上述实施例中人员分组方法的步骤：

步骤 S11，获取多组样本数据，每一组样本数据包括多个条件属性以及对应的决策属性，所述决策属性为所述样本数据的绩效。

其中，所述样本数据需同时包括绩效较优的人员(即，绩优人员)和绩效较差的人员(即，绩差人员)的数据，所述样本数据可存储于所述电子装置中。在另一实施方式中，所述电子装置还可以通过接入网络连接之一外部样本库，进而来获取所述样本库存储的样本数据。在其它实施方式中，所述电子装置还可以通过大数据方式来收集并建立所述样本库。

在本实施方式中，所述条件属性可为行为轨迹(如，出差情况)、app 活跃情况、业务扩展情况、消费情况、兴趣爱好、参加培训情况、考勤率等。以所述条件属性包括出差情况、业务扩展情况、参加培训情况以及考勤率为例进行说明，所述样本数据如上表 1 所示。

步骤 S12，根据所述样本数据训练一决策树模型，其中，所述决策树模型包括多个叶子节点，每一叶子节点代表一个分组。

其中，决策树算法属于监督学习分类算法，而决策树模型代表对象属性与对象值之间的一种映射关系。其中，所述决策树模型具有 N 层(N 为自然数， $N > 2$)，所述电子装置根据所述条件属性设置所述决策树模型中的每个节点(包括根节点、内部节点和叶子节点)的值。所述决策树模型的第一层为根节点，所述决策树模型的第二层代表以第一级条件属性对所述根节点进行分割得到的多个节点，所述决策树模型的第三层代表以第二级条件属性对上一层节点进行分割得到的多个节点，等等。如图 2 所示，所述决策树模型包括四层，即， $N=4$ 。

在本实施方式中，训练所述决策树模型具体包括：

步骤 S121，以所述样本数据作为训练集计算每一条件属性的信息增益

(information gain)。

步骤 S122, 选择信息增益最大的条件属性作为所述决策树模型的根节点以分割所述样本数据以得到下一层节点; 其中, 某一条件属性的信息增益越大, 说明选择所述条件属性对分类提供的信息越多, 越有利于确定性, 越有利于将所述样本数据进行分类。如图 2 所示, 若计算信息增益最大的条件属性为“出差情况”, 则选择“出差情况”条件属性作为根节点以分割所述样本数据。

步骤 S123, 以每一节点包含的样本数据作为训练集重新计算每一条件属性的信息增益。

步骤 S124, 选择信息增益最大的条件属性分割所述节点以得到下一层节点。如图 2 所示, 对于包含每月出差情况为 0 次($C=0$)的样本的节点, 若计算信息增益最大的条件属性为“业务拓展情况”, 则根据“业务拓展情况”条件属性对所述节点进行分割而获得下一层节点。对于包含每月出差情况为 ≥ 2 次($C \geq 2$)的样本的节点, 若计算信息增益最大的条件属性为“参加培训情况”, 则根据“参加培训情况”条件属性对所述节点进行分割而获得下一层节点。实际训练过程中, 对位于同一层的多个节点进行分割所采用的条件属性通常不同。每一次分割后形成的节点比上一层节点的数据纯度更高。

步骤 S125, 递归执行步骤 S123 以及 S124, 直至分割停止。此时, 每一叶节点代表一个分组, 每一分组包含的绩优人员数量与绩差人员数量之间具有固定的比值, 而多个分组包含的绩优人员数量与绩差人员数量之间的比值可以互不相同。如, 叶子节点“分组 1”的所述比值可为 1:8。当然, 某一叶节点包含的可均为绩优人员, 也可以均为绩差人员。其中, 每一组样本数据只能被划分至其中一个叶子节点中, 即, 每一组样本数据不能同属于两个及两个以上的叶子节点。

在本实例中, 在所述决策树的层数达到一预设层数(如: 4 层)时停止下一次分割。在另一实施例, 在所述决策树的当前层的每一节点包含的样本数量小于一预设数量时停止下一次分割。

其中, 每一条件属性的信息增益 $Gain(S, A)$ 可通过如下方式计算:

$$Gain(S, A) = Entropy(S) - Entropy(S, A)$$

其中, $Gain(S, A)$ 表示 A 条件属性在训练集 S 上的信息增益, $Entropy(S)$ 表示训练集 S 的信息熵, $Entropy(S, A)$ 表示 A 属性的信息熵。

例如, 如表 1 所示, 在步骤 S121 中, 所述训练集共有样本数据 16 组, 绩优的有 11 组, 绩差的有 5 组。以“出差情况”条件属性的信息增益计算为例进行说明, 其它条件属性的信息增益计算过程同理。对于 $C=0$ 的情况, 共有样本数据 5 组, 绩优的有 4 组, 绩差的有 1 组; 对于 $C=1$ 的情况, 共有样本数据 4 组, 绩优的有 2 组, 绩差的有 2 组; 对于 $C \geq 2$ 的情况, 共有样本数据 7 组, 绩优的有 5 组, 绩差的有 2 组。因此, “出差情况”条件属性的信息增益的计算方式如下:

$$Entropy(S) = -\frac{11}{16} \log_2 \frac{11}{16} - \frac{5}{16} \log_2 \frac{5}{16}$$

$$Entropy(A) = \frac{4}{16} \times (-\frac{4}{5} \log_2 \frac{4}{5} - \frac{1}{5} \log_2 \frac{1}{5}) + \frac{4}{16} \times (-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}) \\ + \frac{7}{16} \times (-\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7})$$

$$Gain(S, A) = -\frac{11}{16} \log_2 \frac{11}{16} - \frac{5}{16} \log_2 \frac{5}{16} - \frac{4}{16} \times (-\frac{4}{5} \log_2 \frac{4}{5} - \frac{1}{5} \log_2 \frac{1}{5}) \\ - \frac{4}{16} \times (-\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}) - \frac{7}{16} \times (-\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7})$$

步骤 S13, 以所述决策树模型中的每一叶子节点作为当前节点, 计算所有当前节点的绩优率以及绩优覆盖率, 根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件。

针对每一个所述当前节点, 执行以下步骤:

将所述当前节点作为目标节点, 当所述目标节点符合所述预设条件时, 将所述目标节点划分至所述绩优分组;

当所述目标节点不符合所述预设条件时, 判断与所述目标节点属于同一父节点的其它同层节点是否符合所述预设条件;

当属于同一父节点的至少一其它同层节点符合所述预设条件时, 将所述目标节点划分至绩差分组;

当属于同一父节点的所有其它同层节点均不符合所述预设条件时, 忽略所述目标节点。即, 既不将所述目标节点划分至所述绩优分组, 也不将所述目标节点划分至所述绩差分组。

其中, 每一节点的绩优率为所述节点的绩优人员数量与所述节点的待分组人员的总数量之间的比值, 所述节点的绩优率可通过所述节点包含的绩优人员数量与绩差人员数量之间的比值进行计算。每一节点的绩优覆盖率为所述节点的绩优人员数量与整个决策树模型的绩优人员的总数量之间的比值。当所述节点的绩优率以及绩优覆盖率分别达到一绩优率阈值与一绩优覆盖率阈值时, 确定所述节点符合所述预设条件。反之, 当所述节点的绩优率未达到所述绩优率阈值, 或是, 所述节点的绩优覆盖率未达到所述绩优覆盖率阈值, 又或是, 所述节点的绩优率以及绩优覆盖率均未达到所述绩优率阈值与所述绩优覆盖率阈值时, 确定所述节点不符合所述预设条件。

例如, 设置所述绩优率阈值为 40%, 所述绩优覆盖率阈值为 20%。若计算出节点“分组 8”的绩优率为 80%, 绩优覆盖率为 20%, 则节点“分组 8”符合预设条件, 被划分至绩优分组。对于节点“分组 7”, 若绩优覆盖率为 30%, 绩优覆盖率为 25%, 则节点“分组 7”不符合预设条件, 但是由于与所述节点“分组 7”属于同一父节点的同层节点“分组 8”符合预设条件, 因此节点“分组 7”被划分至绩差分组。反之, 若节点“分组 7”与节点“分组 8”均不符合预设条件, 则忽略节点“分组 7”以及节点“分组 8”。

在本实施方式中, 所述绩优率阈值与所述绩优覆盖率阈值预先存储于所述电子装置中。所述绩优率阈值与所述绩优覆盖率阈值可根据实际需要进行

调整。

步骤 S14, 每次当属于同一父节点的所有所述当前节点均不符合所述预设条件时, 判断所述父节点是否为根节点。

步骤 S15, 当所述父节点不是根节点时, 以每一个所述父节点作为新的当前节点, 触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤; 当所述父节点为根节点时, 计算所述根节点的绩优率以及绩优覆盖率, 根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件, 当所述根节点符合预设条件时将所述根节点划分至绩优分组, 当所述根节点不符合预设条件时忽略所述根节点。

根据上述例子, 若节点“分组 7”与节点“分组 8”均不符合预设条件, 则需要考虑父节点是否为根节点。由于节点“分组 7”与节点“分组 8”的父节点并非根节点, 则继续判断所述父节点的所述绩优率以及所述绩优覆盖率是否符合预设条件。可以理解, 当所有的内部节点和叶子节点均不符合预设条件时, 才对所述根节点进行判断。

其中, 通过调整所述绩优率阈值与所述绩优覆盖率阈值, 可控制所述绩优分组与所述绩差分组的大小和质量。

其中, 最终得到的绩优分组可包括具有不同特征的绩优人群。如, 若“分组 4”被划分至所述绩优分组, 则该部分绩优人群涉及的条件属性包括: 出差情况=0 次、业务拓展情况 ≥ 1 次, 考勤率 ≥ 80 。也就是说, 该部分绩优人群的显著特征为: 不出差、业务能力强、出勤情况较好。

若“分组 8”被划分至所述绩优分组, 则该部分绩优人群涉及的条件属性包括: 出差情况为 ≥ 2 次、参加培训情况 ≥ 1 次, 考勤率 ≥ 80 。也就是说, 该部分绩优人群的显著特征为: 经常出差、乐于学习、出勤情况较好。

或者, 所述处理器 102 执行所述计算机可读指令 103 时实现上述人员分组装置实施例中各模块/单元的功能, 例如图 3 中的单元 301-303。

本申请实施例在决策树建模中加入绩差人员的信息, 使得分组目标更明确, 且数据的执行效果更好, 保证最终获得的绩优分组为绩优人数比例达到一定程度的群体; 而且, 可通过绩优分组的条件属性识别绩优人员的显著特征, 从而构建用户画像。

示例性的, 所述计算机可读指令 103 可以被分割成一个或多个模块/单元, 所述一个或者多个模块/单元被存储在所述存储器 101 中, 并由所述处理器 102 执行, 以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机可读指令段, 该指令段用于描述所述计算机可读指令 103 在所述电子装置 1 中的执行过程。例如, 所述计算机可读指令 103 可以被分割成图 3 中的获取模块 301、训练模块 302、执行模块 303。

所述电子装置 1 可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。本领域技术人员可以理解, 所述示意图仅仅是电子装置 1 的示例, 并不构成对电子装置 1 的限定, 可以包括比图示更多或更少的部件, 或者组合某些部件, 或者不同的部件, 例如所述电子装置 1 还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 102 可以是中央处理单元(Central Processing Unit, CPU), 还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器 30 也可以是任何常规的处理器等, 所述处理器 102 是所述电子装置 1 的控制中心, 利用各种接口和线路连接整个电子装置 1 的各个部分。

所述存储器 101 可用于存储所述计算机可读指令 103 和/或模块/单元, 所述处理器 102 通过运行或执行存储在所述存储器 101 内的计算机可读指令和/或模块/单元, 以及调用存储在存储器 101 内的数据, 实现所述电子装置 1 的各种功能。所述存储器 101 可主要包括存储程序区和存储数据区, 其中, 存储程序区可存储操作系统、至少一个功能所需的应用程序 (比如声音播放功能、图像播放功能等) 等; 存储数据区可存储根据电子装置 1 的使用所创建的数据 (比如音频数据、电话本等) 等。此外, 存储器 101 可以包括高速随机存取存储器, 还可以包括非易失性存储器, 例如硬盘、内存、插接式硬盘, 智能存储卡 (Smart Media Card, SMC), 安全数字 (Secure Digital, SD) 卡, 闪存卡 (Flash Card)、至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

所述电子装置 1 集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时, 可以存储在一个非易失性可读取存储介质中。基于这样的理解, 本申请实现上述实施例方法中的全部或部分流程, 也可以通过计算机可读指令来指令相关的硬件来完成, 所述的计算机可读指令可存储于一非易失性可读取存储介质中, 该计算机可读指令在被处理器执行时, 可实现上述各个方法实施例的步骤。其中, 所述计算机可读指令包括计算机可读指令代码, 所述计算机可读指令代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述非易失性可读介质可以包括: 能够携带所述计算机可读指令代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、电载波信号、电信号以及软件分发介质等。需要说明的是, 所述非易失性可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减, 例如在某些司法管辖区, 根据立法和专利实践, 非易失性可读介质不包括电载波信号和电信信号。

在本申请所提供的几个实施例中, 应该理解到, 所揭露的电子装置和方法, 可以通过其它的方式实现。例如, 以上所描述的电子装置实施例仅仅是示意性的, 例如, 所述单元的划分, 仅仅为一种逻辑功能划分, 实际实现时可以有另外的划分方式。

另外, 在本申请各个实施例中的各功能单元可以集成在相同处理单元中, 也可以是各个单元单独物理存在, 也可以两个或两个以上单元集成在相同单元中。上述集成的单元既可以采用硬件的形式实现, 也可以采用硬件加软件

功能模块的形式实现。

对于本领域技术人员而言,显然本申请不限于上述示范性实施例的细节,而且在不背离本申请的精神或基本特征的情况下,能够以其他的具体形式实现本申请。因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本申请的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本申请内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。电子装置权利要求中陈述的多个单元或电子装置也可以由同一个单元或电子装置通过软件或者硬件来实现。第一,第二等词语用来表示名称,而并不表示任何特定的顺序。

最后应说明的是,以上实施例仅用以说明本申请的技术方案而非限制,尽管参照较佳实施例对本申请进行了详细说明,本领域的普通技术人员应当理解,可以对本申请的技术方案进行修改或等同替换,而不脱离本申请技术方案的精神和范围。

权 利 要 求 书

1. 一种人员分组方法，其特征在于，包括：

获取多组样本数据，每一组样本数据包括多个条件属性以及对应的决策属性，所述决策属性为所述样本数据的绩效；

根据所述样本数据训练一决策树模型，其中，所述决策树模型包括多个叶子节点，每一叶子节点代表一个分组；

以所述决策树模型中的每一叶子节点作为当前节点，计算所有当前节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件；

当属于同一父节点的所有所述当前节点均不符合所述预设条件时，判断所述父节点是否为根节点；

当所述父节点不是根节点时，以每一个所述父节点作为新的当前节点，触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤；

当所述父节点为根节点时，计算所述根节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件，当所述根节点符合预设条件时将所述根节点划分至绩优分组，当所述根节点不符合预设条件时忽略所述根节点；

针对每一个所述当前节点，执行以下步骤：

将所述当前节点作为目标节点，当所述目标节点符合所述预设条件时，将所述目标节点划分至所述绩优分组；

当所述目标节点不符合所述预设条件时，判断与所述目标节点属于同一父节点的其它同层节点是否符合所述预设条件；

当属于同一父节点的至少一其它同层节点符合所述预设条件时，将所述目标节点划分至绩差分组；

当属于同一父节点的所有其它同层节点均不符合所述预设条件时，忽略所述目标节点。

2. 如权利要求 1 所述的人员分组方法，其特征在于，所述根据所述样本数据训练一决策树模型包括：

以所述样本数据作为训练集计算每一条件属性的信息增益；

选择信息增益最大的条件属性作为所述决策树模型的根节点以分割所述样本数据以得到下一层节点；

以每一节点包含的样本数据作为训练集重新计算每一条件属性的信息增益；

选择信息增益最大的条件属性分割所述节点以得到下一层节点；以及

递归执行重新计算每一条件属性的信息增益的步骤以及选择信息增益最大的条件属性分割所述节点的步骤，直至分割停止。

3. 如权利要求 2 所述的人员分组方法, 其特征在于, 在所述决策树的层数达到一预设层数时停止下一次分割。

4. 如权利要求 2 所述的人员分组方法, 其特征在于, 在所述决策树的当前层的每一节点包含的样本数量小于一预设数量时停止下一次分割。

5. 如权利要求 1 所述的人员分组方法, 其特征在于, 每一节点的绩优率为所述节点的绩优人员数量与所述节点的待分组人员的总数量之间的比值, 每一节点的绩优覆盖率为所述节点的绩优人员数量与整个决策树模型的绩优人员的总数量之间的比值。

6. 如权利要求 5 所述的人员分组方法, 其特征在于, 所述根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件包括:

当所述当前节点的绩优率以及绩优覆盖率分别达到一绩优率阈值与一绩优覆盖率阈值时, 确定所述当前节点符合所述预设条件; 以及

当所述当前节点的绩优率未达到所述绩优率阈值, 或所述当前节点的绩优覆盖率未达到所述绩优覆盖率阈值, 或所述当前节点的绩优率以及绩优覆盖率均未达到所述绩优率阈值与所述绩优覆盖率阈值时, 确定所述当前节点不符合所述预设条件。

7. 如权利要求 1 所述的人员分组方法, 其特征在于, 所述样本数据同时包括绩优人员和绩差人员的数据。

8. 一种人员分组装置, 其特征在于, 包括:

获取模块, 用于获取多组样本数据, 每一组样本数据包括多个条件属性以及对应的决策属性, 所述决策属性为所述样本数据的绩效;

训练模块, 用于根据所述样本数据训练一决策树模型, 其中, 所述决策树模型包括多个叶子节点, 每一叶子节点代表一个分组;

执行模块, 用于计算所有当前节点的绩优率以及绩优覆盖率, 根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件;

当属于同一父节点的所有所述当前节点均不符合所述预设条件时, 所述执行模块判断所述父节点是否为根节点;

当所述父节点不是根节点时, 所述执行模块以每一个所述父节点作为新的当前节点, 触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤;

当所述父节点为根节点时, 所述执行模块计算所述根节点的绩优率以及绩优覆盖率, 根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件, 当所述根节点符合预设条件时将所述根节点划分至绩优分组, 当所述根节点不符合预设条件时忽略所述根节点;

针对每一个所述当前节点, 所述执行模块用于:

将所述当前节点作为目标节点, 当所述目标节点符合所述预设条件时, 将所述目标节点划分至所述绩优分组;

当所述目标节点不符合所述预设条件时, 判断与所述目标节点属于同一父节点的其它同层节点是否符合所述预设条件;

当属于同一父节点的至少一其它同层节点符合所述预设条件时, 将所述目标节点划分至绩差分组;

当属于同一父节点的所有其它同层节点均不符合所述预设条件时，忽略所述目标节点。

9. 一种电子装置，其特征在于，包括处理器和存储器，所述存储器中存储至少一个计算机可读指令，所述处理器执行所述计算机可读指令以实现以下步骤：

获取多组样本数据，每一组样本数据包括多个条件属性以及对应的决策属性，所述决策属性为所述样本数据的绩效；

根据所述样本数据训练一决策树模型，其中，所述决策树模型包括多个叶子节点，每一叶子节点代表一个分组；

以所述决策树模型中的每一叶子节点作为当前节点，计算所有当前节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件；

当属于同一父节点的所有所述当前节点均不符合所述预设条件时，判断所述父节点是否为根节点；

当所述父节点不是根节点时，以每一个所述父节点作为新的当前节点，触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤；

当所述父节点为根节点时，计算所述根节点的绩优率以及绩优覆盖率，根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件，当所述根节点符合预设条件时将所述根节点划分至绩优分组，当所述根节点不符合预设条件时忽略所述根节点；

针对每一个所述当前节点，执行以下步骤：

将所述当前节点作为目标节点，当所述目标节点符合所述预设条件时，将所述目标节点划分至所述绩优分组；

当所述目标节点不符合所述预设条件时，判断与 said 目标节点属于同一父节点的其它同层节点是否符合所述预设条件；

当属于同一父节点的至少一其它同层节点符合所述预设条件时，将所述目标节点划分至绩差分组；

当属于同一父节点的所有其它同层节点均不符合所述预设条件时，忽略所述目标节点。

10. 如权利要求 9 所述的电子装置，其特征在于，所述处理器在根据所述样本数据训练一决策树模型时，执行所述计算机可读指令以实现以下步骤：

以所述样本数据作为训练集计算每一条件属性的信息增益；

选择信息增益最大的条件属性作为所述决策树模型的根节点以分割所述样本数据以得到下一层节点；

以每一节点包含的样本数据作为训练集重新计算每一条件属性的信息增益；

选择信息增益最大的条件属性分割所述节点以得到下一层节点；以及

递归执行重新计算每一条件属性的信息增益的步骤以及选择信息增益最大的条件属性分割所述节点的步骤，直至分割停止。

11. 如权利要求 10 所述的电子装置, 其特征在于, 所述处理器执行所述计算机可读指令时还用以实现以下步骤:

在所述决策树的层数达到一预设层数时停止下一次分割。

12. 如权利要求 10 所述的电子装置, 其特征在于, 所述处理器执行所述计算机可读指令时还用以实现以下步骤:

在所述决策树的当前层的每一节点包含的样本数量小于一预设数量时停止下一次分割。

13. 如权利要求 9 所述的电子装置, 其特征在于, 每一节点的绩优率为所述节点的绩优人员数量与所述节点的待分组人员的总数量之间的比值, 每一节点的绩优覆盖率为所述节点的绩优人员数量与整个决策树模型的绩优人员的总数量之间的比值。

14. 如权利要求 13 所述的电子装置, 其特征在于, 所述根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件时, 所述处理器执行所述计算机可读指令以实现以下步骤:

当所述当前节点的绩优率以及绩优覆盖率分别达到一绩优率阈值与一绩优覆盖率阈值时, 确定所述当前节点符合所述预设条件; 以及

当所述当前节点的绩优率未达到所述绩优率阈值, 或所述当前节点的绩优覆盖率未达到所述绩优覆盖率阈值, 或所述当前节点的绩优率以及绩优覆盖率均未达到所述绩优率阈值与所述绩优覆盖率阈值时, 确定所述当前节点不符合所述预设条件。

15. 一种非易失性可读存储介质, 其特征在于, 所述非易失性可读存储介质上存储至少一个计算机可读指令, 所述计算机可读指令被处理器执行时实现以下步骤:

获取多组样本数据, 每一组样本数据包括多个条件属性以及对应的决策属性, 所述决策属性为所述样本数据的绩效;

根据所述样本数据训练一决策树模型, 其中, 所述决策树模型包括多个叶子节点, 每一叶子节点代表一个分组;

以所述决策树模型中的每一叶子节点作为当前节点, 计算所有当前节点的绩优率以及绩优覆盖率, 根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件;

当属于同一父节点的所有所述当前节点均不符合所述预设条件时, 判断所述父节点是否为根节点;

当所述父节点不是根节点时, 以每一个所述父节点作为新的当前节点, 触发所述计算所有当前节点的绩优率以及绩优覆盖率的步骤;

当所述父节点为根节点时, 计算所述根节点的绩优率以及绩优覆盖率, 根据所述绩优率以及所述绩优覆盖率确定所述根节点是否符合预设条件, 当所述根节点符合预设条件时将所述根节点划分至绩优分组, 当所述根节点不符合预设条件时忽略所述根节点;

针对每一个所述当前节点, 执行以下步骤:

将所述当前节点作为目标节点, 当所述目标节点符合所述预设条件时,

将所述目标节点划分至所述绩优分组；

当所述目标节点不符合所述预设条件时，判断与所述目标节点属于同一父节点的其它同层节点是否符合所述预设条件；

当属于同一父节点的至少一其它同层节点符合所述预设条件时，将所述目标节点划分至绩差分组；

当属于同一父节点的所有其它同层节点均不符合所述预设条件时，忽略所述目标节点。

16. 如权利要求 15 所述的存储介质，其特征在于，在根据所述样本数据训练一决策树模型时，所述计算机可读指令被所述处理器执行以实现以下步骤：

以所述样本数据作为训练集计算每一条件属性的信息增益；

选择信息增益最大的条件属性作为所述决策树模型的根节点以分割所述样本数据以得到下一层节点；

以每一节点包含的样本数据作为训练集重新计算每一条件属性的信息增益；

选择信息增益最大的条件属性分割所述节点以得到下一层节点；以及

递归执行重新计算每一条件属性的信息增益的步骤以及选择信息增益最大的条件属性分割所述节点的步骤，直至分割停止。

17. 如权利要求 16 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还用以实现以下步骤：

在所述决策树的层数达到一预设层数时停止下一次分割。

18. 如权利要求 16 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还用以实现以下步骤：

在所述决策树的当前层的每一节点包含的样本数量小于一预设数量时停止下一次分割。

19. 如权利要求 15 所述的存储介质，其特征在于，每一节点的绩优率为所述节点的绩优人员数量与所述节点的待分组人员的总数量之间的比值，每一节点的绩优覆盖率为所述节点的绩优人员数量与整个决策树模型的绩优人员的总数量之间的比值。

20. 如权利要求 19 所述的存储介质，其特征在于，所述根据所述绩优率以及所述绩优覆盖率确定每一个所述当前节点是否符合预设条件时，所述计算机可读指令被所述处理器执行以实现以下步骤：

当所述当前节点的绩优率以及绩优覆盖率分别达到一绩优率阈值与一绩优覆盖率阈值时，确定所述当前节点符合所述预设条件；以及

当所述当前节点的绩优率未达到所述绩优率阈值，或所述当前节点的绩优覆盖率未达到所述绩优覆盖率阈值，或所述当前节点的绩优率以及绩优覆盖率均未达到所述绩优率阈值与所述绩优覆盖率阈值时，确定所述当前节点不符合所述预设条件。

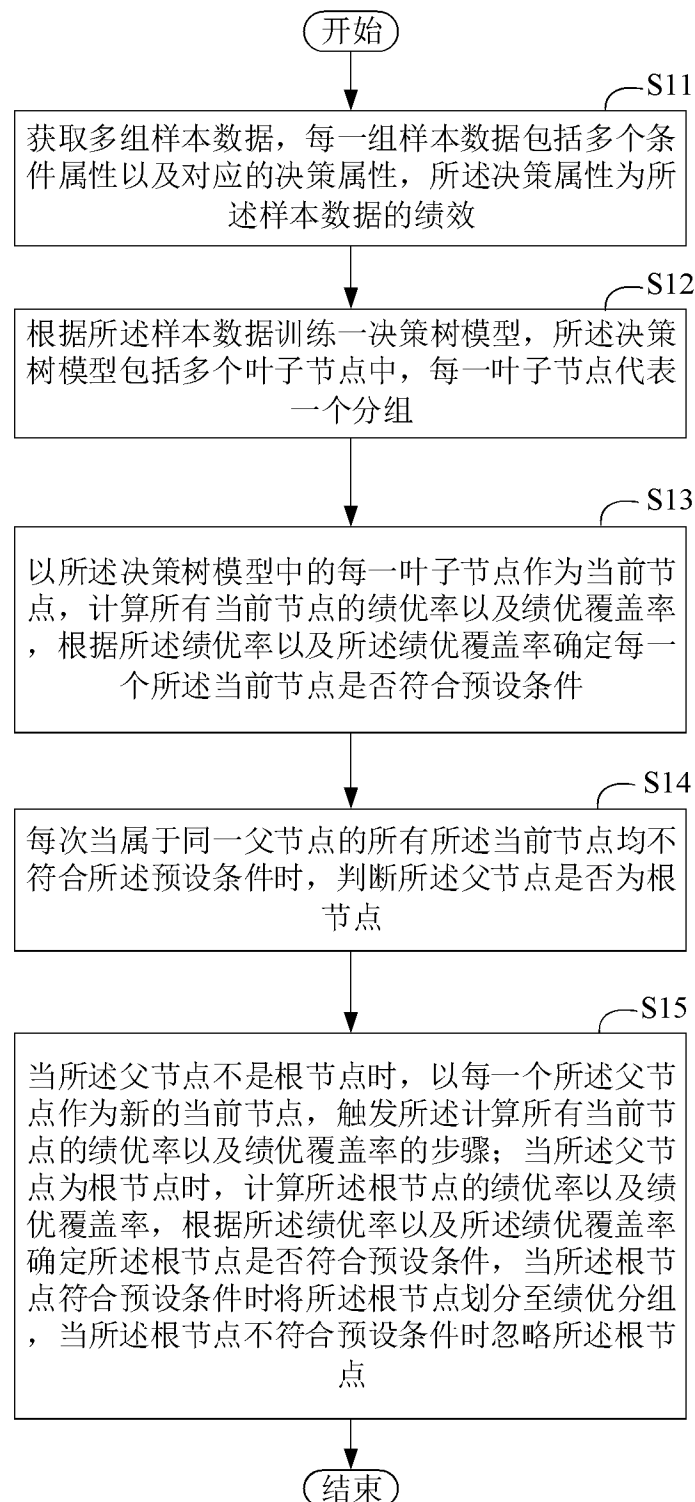


图 1

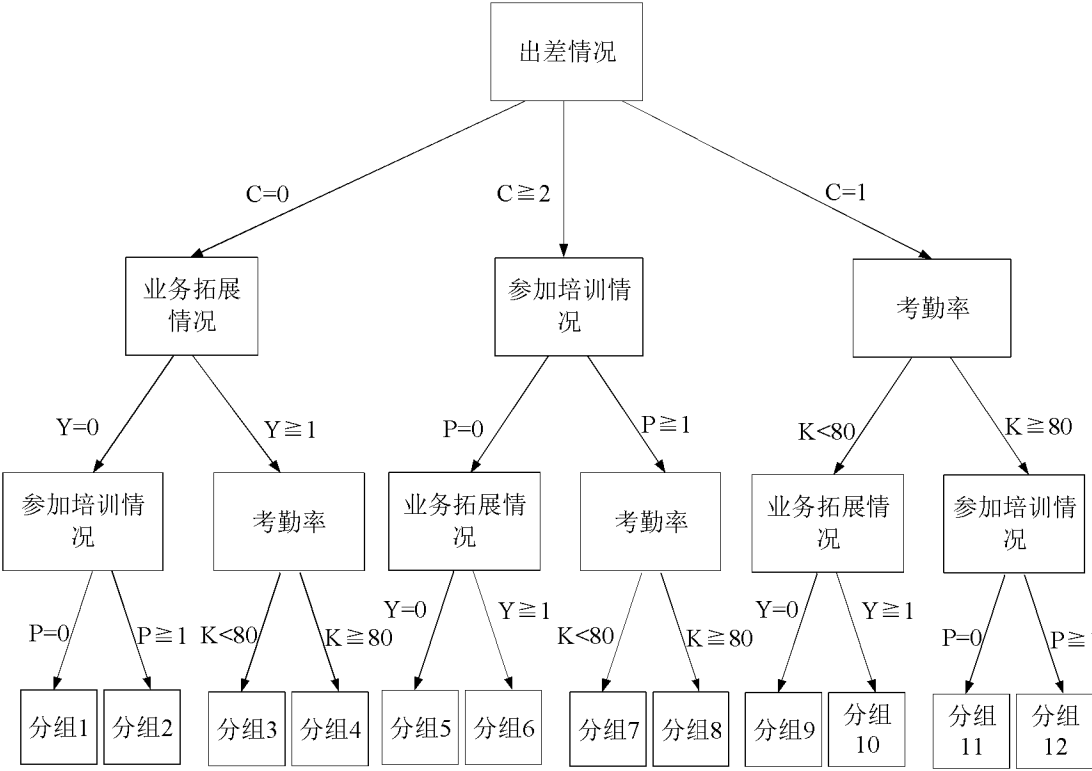


图 2

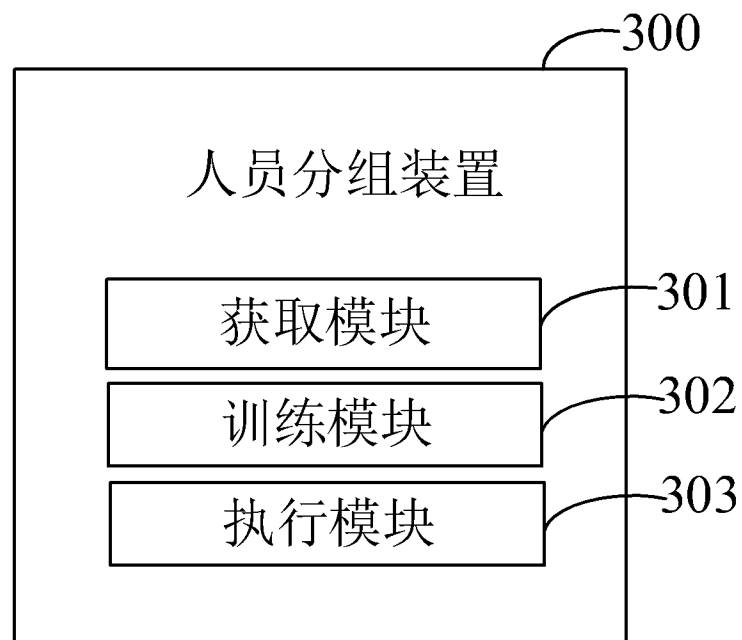


图 3

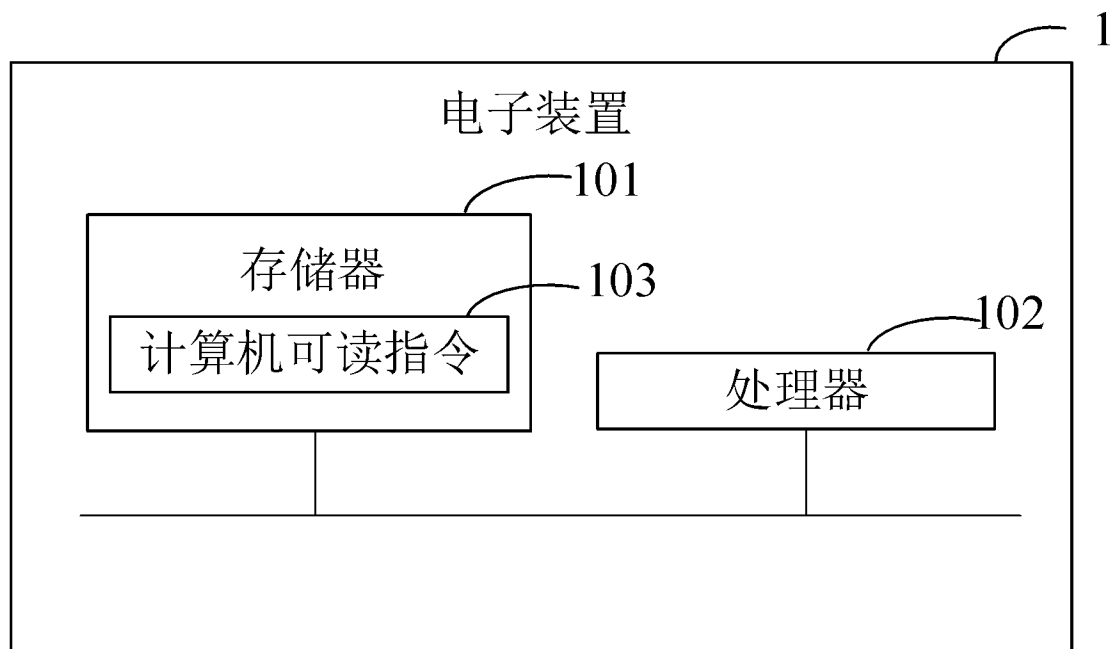


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/077227

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/06(2012.01)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

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, WPI, EPODOC, CNKI, IEEE: 叶子, 绩优, 根节点, 决策树, 绩差, 绩效, 分组, 人员, 样本, leaf, excellent, root, decision, tree, performance, group, personnel, template

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109214671 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 15 January 2019 (2019-01-15) claims 1-10	1-20
A	CN 105894177 A (STATE GRID CORPORATION OF CHINA ET AL.) 24 August 2016 (2016-08-24) description, paragraphs [0041]-[0097]	1-20
A	CN 102214213 A (INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 12 October 2011 (2011-10-12) entire document	1-20
A	CN 105354588 A (BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS) 24 February 2016 (2016-02-24) 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

25 May 2019

Date of mailing of the international search report

05 June 2019

Name and mailing address of the ISA/CN

**China National Intellectual Property Administration (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/CN2019/077227

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109214671	A	15 January 2019	None	
CN	105894177	A	24 August 2016	None	
CN	102214213	A	12 October 2011	None	
CN	105354588	A	24 February 2016	None	

国际检索报告

国际申请号

PCT/CN2019/077227

A. 主题的分类

G06Q 10/06 (2012.01) 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

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, WPI, EPODOC, CNKI, IEEE: 叶子, 绩优, 根节点, 决策树, 绩差, 绩效, 分组, 人员, 样本, leaf, excellent, root, decision, tree, performance, group, personnel, template

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 109214671 A (平安科技深圳有限公司) 2019年 1月 15日 (2019 - 01 - 15) 权利要求1-10	1-20
A	CN 105894177 A (国家电网公司 等) 2016年 8月 24日 (2016 - 08 - 24) 说明书第0041-0097段	1-20
A	CN 102214213 A (中国科学院计算技术研究所) 2011年 10月 12日 (2011 - 10 - 12) 全文	1-20
A	CN 105354588 A (北京邮电大学) 2016年 2月 24日 (2016 - 02 - 24) 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 5月 25日

国际检索报告邮寄日期

2019年 6月 5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

授权官员

王芳

电话号码 86-(10)-53961369

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/077227

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	109214671	A	2019年 1月 15日	无	
CN	105894177	A	2016年 8月 24日	无	
CN	102214213	A	2011年 10月 12日	无	
CN	105354588	A	2016年 2月 24日	无	