(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 7 月 23 日 (23.07.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/147259 A1

(51) 国际专利分类号:
G06F 16/951 (2019.01) *G06K 9/62* (2006.01)
G06F 16/953 (2019.01)

(21) 国际申请号: PCT/CN2019/091529

(22) 国际申请日: 2019 年 6 月 17 日 (17.06.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910041270.X 2019年1月16日 (16.01.2019) 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: **杨晟 (YANG, Sheng)**; 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)。 **陈爽 (CHEN, Shuang)**; 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)。 **陈源 (CHEN, Yuan)**; 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: **深圳中一联合知识产权代理有限公司(SHENZHEN ZHONGYI UNION INTELLECTUAL PROPERTY AGENCY CO., LTD.)**; 中国广东省深

(54) **Title:** USER PORTAIT METHOD AND APPARATUS, READABLE STORAGE MEDIUM, AND TERMINAL DEVICE

(54) **发明名称:** 一种用户画像方法、装置、可读存储介质及终端设备

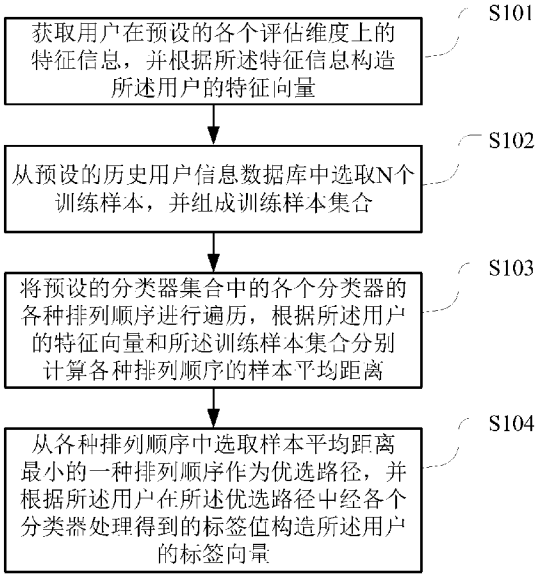


图 1

S101 Acquire feature information of a user in each evaluation dimension and, on the basis of the feature information, construct an eigenvector of the user

S102 Select N training samples from a preset historical user information database, and form a training sample set

S103 Traverse the various sorting orders of the classifiers in a preset classifier set and, on the basis of the eigenvector of the user and the training sample set, respectively calculate the average distances of the samples of the various sorting orders

S104 Select the sorting order having the shortest average distance of the samples from the various sorting orders as a preferred path and, on the basis of a label value of the user obtained by being processed by each classifier in the preferred path, construct a label vector of the user

(57) **Abstract:** The present application relates to the technical field of computers, and specifically relates to a user portrait method and apparatus, a computer readable storage medium, and a terminal device. The method comprises: acquiring feature information of a user in each evaluation dimension and, on the basis of the feature information, constructing an eigenvector of the user; selecting N training samples from a preset historical user information database, and forming a training sample set; traversing the various sorting orders of the classifiers in a preset classifier set and, on the basis of the eigenvector of the user and the training sample set, respectively

WO 2020/147259 A1

[见续页]

圳市福田区园岭街道深南中路1014号报春大厦9楼（5号信箱）, Guangdong 518028 (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))。

calculating the average distances of the samples of the various sorting orders; selecting the sorting order having the shortest average distance of the samples from the various sorting orders as a preferred path and, on the basis of a label value of the user obtained by being processed by each classifier in the preferred path, constructing a label vector of the user. The result of a preceding classifier is involved in the processing of a later classifier, greatly increasing the user portrait accuracy.

(57) 摘要：本申请属于计算机技术领域，尤其涉及一种用户画像方法、装置、计算机可读存储介质及终端设备。所述方法获取用户在各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量；从预设的历史用户信息数据库中选取N个训练样本，并组成训练样本集合；将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离；从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量。在前的分类器的结果会参与到在后的分类器的处理之中，大大提升了用户画像的准确率。

一种用户画像方法、装置、可读存储介质及终端设备

本申请要求于 2019 年 1 月 16 日提交中国专利局、申请号为 201910041270.X、发明名称为“一种用户画像方法、装置、可读存储介质及终端设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请属于计算机技术领域，尤其涉及一种用户画像方法、装置、计算机可读存储介质及终端设备。

背景技术

随着互联网的普及与发展，越来越多的数据可以被各个互联网运营商收集起来。例如，对于电商网站，可以获得用户的购买记录、浏览记录等信息；对于搜索引擎，可以获得用户的搜索记录、点击记录等信息。为了更好的利用这样的信息，以提供更为高效优质的服务，用户画像这一技术得到了普遍重视。用户画像是根据用户社会属性、生活习惯和消费行为等信息而抽象出的一个标签化的用户模型。

但是，目前的用户画像一般都是针对单一的标签进行的，然而，在很多实际场景中，一个用户可能同时具有多个不同维度的标签。针对这种多标签的场景，现有技术一般会将各个不同的标签维度独立进行处理，分别在各个标签维度上进行用户画像，这种方式未能充分考虑到各个标签维度之间的关联性，往往导致用户画像的准确率较低。

技术问题

有鉴于此，本申请实施例提供了一种用户画像方法、装置、计算机可读存储介质及终端设备，以解决现有技术一般会将各个不同的标签维度独立进行处理，导致用户画像的准确率较低的问题。

技术解决方案

本申请实施例的第一方面提供了一种用户画像方法，可以包括：

获取用户在预设的各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量；

从预设的历史用户信息数据库中选取 N 个训练样本，并组成训练样本集合，N 为正整数；

将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离，所述分类器集合中包括两个以上的不同分类器，每个分类器用于确定一个标签维度上的标签值；

从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据

所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量。

本申请实施例的第二方面提供了一种用户画像装置，可以包括用于实现上述用户画像方法的步骤的模块。

本申请实施例的第三方面提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机可读指令，所述计算机可读指令被处理器执行时实现上述用户画像方法的步骤。

本申请实施例的第四方面提供了一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述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上述用户画像方法的步骤。

有益效果

通过本申请，排列顺序在前的分类器对用户的特征向量的处理结果会传递到排列顺序在后的分类器中，并参与到排列顺序在后的分类器的处理过程之中，从而将各个分类器有机的耦合在一起，充分利用了这些分类器之间的关联性，大大提升了用户画像的准确率。

附图说明

图 1 为本申请实施例中一种用户画像方法的一个实施例流程图；

图 2 为各个分类器之间的网状拓扑结构的示意图；

图 3 为在网状拓扑结构中的优选路径的示意图；

图 4 为根据用户的特征向量和训练样本集合分别计算各种排列顺序的样本平均距离的示意流程图；

图 5 为构造出的分类器序列的示意图；

图 6 为示例中 4 个分类器之间的网状拓扑结构的示意图；

图 7 为选取出的优选路径的示意图；

图 8 为本申请实施例中一种用户画像装置的一个实施例结构图；

图 9 为本申请实施例中一种终端设备的示意框图。

本发明的实施方式

请参阅图 1，本申请实施例中一种用户画像方法的一个实施例可以包括：

步骤 S101、获取用户在预设的各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量。

具体的评估维度可以根据实际情况进行设置，例如，若要判断用户对某些产品的购买意向，则所需的评估维度包括但不限于性别信息、年龄信息、受教育程度信息、收入水平信息、工作种类信息等等。

由于上述各个评估维度上的特征信息并不是数值化的信息，不利于进行分析计算，因此，需要首先对各个维度的信息进行数值化处理。

例如，对于性别信息这一评估维度而言，若性别为男性，则可将该评估维度的取值数值化为 1，若性别为女性，则可将该评估维度的取值数值化为 0；对于年龄这一评估维度而言，若年龄为 20 岁以下，则可将该评估维度的取值数值化为 0，若年龄为 20 岁至 40 岁，则可将该评估维度的取值数值化为 1，若年龄为 40 岁至 60 岁，则可将该评估维度的取值数值化为 2，若年龄为 60 岁以上，则可将该评估维度的取值数值化为 3；对于受教育程度这一评估维度而言，若受教育程度为高中及高中以下，则可将该评估维度的取值数值化为 0，若受教育程度为本科，则可将该评估维度的取值数值化为 1，若受教育程度为硕士研究生，则可将该评估维度的取值数值化为 2，若受教育程度为博士研究生及以上，则可将该评估维度的取值数值化为 3；其它各个评估维度的数值化处理过程与之类似，此处不再赘述。

需要注意的是，以上的数值化处理过程只是一种可能的数值化处理方式，还可以根据实际情况进行其它类似的数值化处理，此处不对其做具体限制。

在进行完数值化处理之后，则可将各个评估维度上的特征信息构造为如下所示的特征向量：

$$TgtEgVec=(TgtEgVal_1,TgtEgVal_2,...,TgtEgVal_{en},...,TgtEgVal_{EN})$$

其中，$TgtEgVec$ 为所述用户的特征向量，$TgtEgVal_{en}$ 为所述用户的特征向量在第 en 个评估维度上的特征信息，1≤en≤EN，EN 为评估维度的总数。

步骤 S102、从预设的历史用户信息数据库中选取 N 个训练样本，并组成训练样本集合。

N 为正整数，其具体取值可以根据实际情况进行设置，一般地，为了保证训练结果的准确性，训练样本的数目应尽量的大，例如，可以将其设置为 10000、20000、50000 或者其它取值等等。

所述训练样本集合中的每个训练样本均包括一个历史用户的特征向量和标签向量，即：

$$Sample_n=(EigenVec_n,LabelVec_n)$$

其中，$Sample_n$ 为所述训练样本集合中的第 n 个训练样本，1≤n≤N，$EigenVec_n$ 为第 n 个训练样本的特征向量，且：

$$EigenVec_n=(EigenVal_{n,1},EigenVal_{n,2},...,EigenVal_{n,en},...,EigenVal_{n,EN}),$$

$EigenVal_{n,en}$ 为第 n 个训练样本的特征向量在第 en 个评估维度上的特征信息，1≤en≤EN，EN 为评估维度的总数，$LabelVec_n$ 为第 n 个训练样本的标签向量，且：

$LabelVec_n=(LabelVal_{n,1},LabelVal_{n,2},...,LabelVal_{n,ln},...,LabelVal_{n,LN})$，

$LabelVal_{n,ln}$ 为第 n 个训练样本的标签向量在第 ln 个标签维度上的标签值，1≤ln≤LN，LN 为标签维度的总数。

为了保证训练结果的准确性，所述训练样本集合在任一标签维度上的标签值均应是均衡分布的，将第 ln 个标签维度上的标签值的总数，也即第 ln 个标签维度上的标签值的可能取值的个数记为 LVN_{ln}，每个标签值的序号记为 lvn_{ln}，$1\le lvn_{ln}\le LVN_{ln}$，第 lvn_{ln} 个标签值在所述训练样本集合中出现的次数记为 $ValCount_{lvn_{ln}}$，且 $\sum_{lvn_{ln}=1}^{LVN_{ln}} ValCount_{lvn_{ln}}=N$，则应满足下列条件：

$$\frac{\sum_{lvn_{ln}=1}^{LVN_{ln}}(ValCount_{lvn_{ln}}-\frac{N}{LVN_{ln}})^2}{LVN_{ln}}<Threshold$$

其中，$Threshold$ 为预设的均衡度阈值，其具体取值可以根据实际情况进行设置，例如，可以将其设置为 50、100、200 或者其它取值等等。

步骤 S103、将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离。

所述分类器集合中包括两个以上的不同分类器，每个分类器用于确定一个标签维度上的标签值。这些分类器可以是基于决策树、朴素贝叶斯、支持向量机、神经网络、遗传算法中的任意一种算法，各个分类器可以基于相同的算法，也可以基于不同的算法。

各个分类器之间可以建立如图 2 所示的网状拓扑结构，图 2 中给出了 LN=2，LN=3，LN=4，LN=5 这四种情况下各个分类器之间的网状拓扑结构，其中，圆圈表示分类器，箭头表示分类器的传递方向，任意两个分类器之间均有双向的路径连接，其它情况可以依此类推。

对于由 LN 个分类器所组成的网状拓扑结构，依次遍历各个分类器节点且每个分类器节点仅遍历一次的路径共有 LN 的阶乘，即 LN！=1×2×3×…×(LN-1)×LN，，本实施例中即是要通过对所有可能的路径进行遍历选择一条如图 3 所示的优选路径。

如图 4 所示，步骤 S103 具体可以包括以下过程：

步骤 S1031、对所述用户的特征向量进行扩展。

例如，可以根据下式对所述用户的特征向量进行扩展：

$ExTgtEgVec_{CandiSq,ln}=(ExTgtEgVec_{CandiSq,ln-1},SelTgtLbVal_{CandiSq,ln-1})$

其中，$ExTgtEgVec_{CandiSq,ln}$ 为所述用户在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，1≤CandiSq≤LN！，LN！为 LN 的阶乘，特殊地，

设置 $ExTgtEgVec_{CandiSq,0}=TgtEgVec$，$TgtEgVec$ 为所述用户的特征向量，$SelTgtLbVal_{CandiSq,ln}$ 为所述用户与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelTgtLbVal_{CandiSq,0}=\varnothing$，即将其设为空，由此可知，$ExTgtEgVec_{CandiSq,1}=TgtEgVec$。

步骤 S1032、对所述训练样本集合的各个训练样本的特征向量进行扩展。

例如，可以根据下式对所述训练样本集合的各个训练样本的特征向量进行扩展：

$$ExEigenVec_{CandiSq,ln,n}=(ExEigenVec_{CandiSq,ln-1,n},SelLabelVal_{CandiSq,ln-1,n})$$

其中，$ExEigenVec_{CandiSq,ln,n}$ 为第 n 个训练样本在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，特殊地，设置 $ExEigenVec_{CandiSq,0,n}=EigenVec_n$，$SelLabelVal_{CandiSq,ln,n}$ 为第 n 个训练样本与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelLabelVal_{CandiSq,0,n}=\varnothing$，即将其设为空，由此可知，$ExEigenVec_{CandiSq,1,n}=EigenVec_n$。

步骤 S1033、根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离。

首先，根据下式计算第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离：

$$Dis(CandiSq,ln,clvn)=\frac{\sum_{SpSeq=1}^{CValCount(CandiSq,ln,clvn)}\left(\sum_{i=1}^{EN+ln-1}(ExTgtElm(CandiSq,ln,i)-ExSpElm(CandiSq,ln,clvn,SpSeq,i))^2\right)}{CValCount(CandiSq,ln,clvn)}$$

其中，$ExTgtElm(CandiSq,ln,i)$ 为所述用户扩展得到的特征向量的第 i 个元素，$ExSpElm(CandiSq,ln,clvn,SpSeq,i)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的第 $SpSeq$ 个训练样本扩展得到的特征向量的第 i 个元素，$1\le clvn\le CLVN(CandiSq,ln)$，$CLVN(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的标签值的总数，$1\le SpSeq\le CValCount(CandiSq,ln,clvn)$，$CValCount(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值在所述训练样本集合中出现的次数，$Dis(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离；

然后，根据下式分别构造第 CandiSq 种排列顺序中的各个分类器的样本平均距离序列：

$$DisSq(CandiSq,ln)=(Dis(CandiSq,ln,1),Dis(CandiSq,ln,2),...,$$
$$Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，$DisSq(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离序列；

接着，根据下式分别计算第 CandiSq 种排列顺序中的各个分类器的样本平均距离：

$$MinDis(CandiSq,ln)=MIN(DisSq(CandiSq,ln))$$
$$=MIN(Dis(CandiSq,ln,1),Dis(t,sn,2),...,Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，MIN 为求最小值函数，$MinDis(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离；

最后，根据下式计算第 CandiSq 种排列顺序的样本平均距离：

$$TotalDis_{CandiSq}=\frac{\sum_{ln=1}^{LN}MinDis(CandiSq,ln)}{LN}$$

其中，$TotalDis_{CandiSq}$ 为第 CandiSq 种排列顺序的样本平均距离。

步骤 S104、从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量。

将各种排列顺序的样本平均距离构造为如下所示的序列：

$$TotalDisSq=(TotalDis_1,TotalDis_2,...,TotalDis_{CandiSq},...,TotalDis_{LN!})$$

再从中选取样本平均距离最小的一种排列顺序作为优选路径，即：

$$SelPath=\arg\min(TotalDisSq)$$
$$=\arg\max(TotalDis_1,TotalDis_2,...,TotalDis_{CandiSq},...,TotalDis_{LN!})$$

其中，$SelPath$ 即为选取出的优选路径的序号。

经过上述的过程，确定了分类器所组成的网状拓扑结构的优选路径，按照该路径中各个分类器的顺序依次连接，即可构造出如图 5 所示的分类器序列。

最后，根据下式构造所述用户的标签向量：

$$TgtLbVec=(TgtLbVal_1,TgtLbVal_2,...,TgtLbVal_{ln},...,TgtLbVal_{LN})$$

其中，$TgtLbVal_{ln}$ 为所述用户在所述优选路径中经第 ln 个分类器处理得到的标签值，$TgtLbVec$ 为所述用户的标签向量。

现对以上整个过程举例具体说明如下：

若采集的用户数据分别为性别信息、年龄信息、受教育程度信息、收入水平信息、工作种类信息，现需要根据这些信息分别评估该用户对产品 A、产品 B、产品 C、产品 D 这四类产品的购买意向，即 4 个标签值。

预先设置 4 个分类器，其中，分类器 A 用于对用户对产品 A 的购买意向进行评估，分类器 B 用于对用户对产品 A 的购买意向进行评估，分类器 C 用于对用户对产品 C 的购买意向进行评估，分类器 D 用于对用户对产品 D 的购买意向进行评估，4 个分类器构成图 6 所示的网状拓扑结构。

其中的可选路径共有 24 种，分别为：

分类器 A→分类器 B→分类器 C →分类器 D

分类器 A→分类器 B→分类器 D →分类器 C

分类器 A→分类器 C→分类器 B →分类器 D

分类器 A→分类器 C→分类器 D →分类器 B

分类器 A→分类器 D→分类器 B →分类器 C

分类器 A→分类器 D→分类器 C →分类器 B

……

……

按照前述方式针对该用户选取出其优选路径，假设最终选取出的优选路径如 7 所示，则最终构造出的分类器序列即为：

分类器 C→分类器 B→分类器 A→分类器 D。

使用该分类器序列进行用户的多标签画像，首先，将待评估的用户的性别信息、年龄信息、受教育程度信息、收入水平信息、工作种类信息这 5 个维度上的特征信息输入到分类器 C 中，得到该用户对产品 C 的购买意向，然后，将该用户的性别信息、年龄信息、受教育程度信息、收入水平信息、工作种类信息以及对产品 C 的购买意向这 6 个维度上的特征信息输入到分类器 B 中，得到该用户对产品 B 的购买意向，再将该用户的性别信息、年龄信息、受教育程度信息、收入水平信息、工作种类信息以及对产品 C 的购买意向、对产品 B 的购买意向这 7 个维度上的特征信息输入到分类器 A 中，得到该用户对产品 A 的购买意向，最后，将该用户的性别信息、年龄信息、受教育程度信息、收入水平信息、工作种类信息以及对产品 C 的购买意向、对产品 B 的购买意向、对产品 A 的购买意向这 8 个维度上的特征信息输入到分类器 D 中，得到该用户对产品 D 的购买意向。

综上所述，通过本申请，排列顺序在前的分类器对用户的特征向量的处理结果会传递到排列顺序在后的分类器中，并参与到排列顺序在后的分类器的处理过程之中，从而将各个分类器有机的耦合在一起，充分利用了这些分类器之间的关联性，大大提升了用户画像的准确率。

对应于上文实施例所述的一种用户画像方法，图 8 示出了本申请实施例提供的一种用户画像装置的一个实施例结构图。

本实施例中，一种用户画像装置可以包括：

特征向量构造模块 801，用于获取用户在预设的各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量；

训练样本选取模块 802，用于从预设的历史用户信息数据库中选取 N 个训练样本，并组成训练样本集合，N 为正整数，所述训练样本集合中的每个训练样本均包括一个

历史用户的特征向量和标签向量。

遍历计算模块 803，用于将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离，所述分类器集合中包括两个以上的不同分类器，每个分类器用于确定一个标签维度上的标签值；

标签向量构造模块 804，用于从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量。

进一步地，所述遍历计算模块可以包括：

第一扩展单元，用于对所述用户的特征向量进行扩展；

第二扩展单元，用于对所述训练样本集合的各个训练样本的特征向量进行扩展；

样本平均距离计算单元，用于根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离。

进一步地，所述样本平均距离计算单元可以包括：

第一子单元，用于计算第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离；

第二子单元，用于分别构造第 CandiSq 种排列顺序中的各个分类器的样本平均距离序列；

第三子单元，用于分别计算第 CandiSq 种排列顺序中的各个分类器的样本平均距离；

第四子单元，用于计算第 CandiSq 种排列顺序的样本平均距离。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的装置，模块和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

图 9 示出了本申请实施例提供的一种终端设备的示意框图，为了便于说明，仅示出了与本申请实施例相关的部分。

在本实施例中，所述终端设备 9 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。该终端设备 9 可包括：处理器 90、存储器 91 以及存储在所述存储器 91 中并可在所述处理器 90 上运行的计算机可读指令 92，例如执行上述的用户画像方法的计算机可读指令。所述处理器 90 执行所述计算机可读指令 92 时实现上述各个用户画像方法实施例中的步骤。

在本申请各个实施例中的各功能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，可以存储在一个计算机可读存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干计算机可读指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁碟或者光盘等各种可以存储计算机可读指令的介质。

权利要求书

1. 一种用户画像方法，其特征在于，包括：

获取用户在预设的各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量；

从预设的历史用户信息数据库中选取 N 个训练样本，并组成训练样本集合，N 为正整数；

将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离，所述分类器集合中包括两个以上的不同分类器，每个分类器用于确定一个标签维度上的标签值；

从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量。

2. 根据权利要求 1 所述的用户画像方法，其特征在于，所述训练样本集合中的每个训练样本均包括一个历史用户的特征向量和标签向量，即：

$$Sample_n=(EigenVec_n,LabelVec_n)$$

其中，$Sample_n$ 为所述训练样本集合中的第 n 个训练样本，1≤n≤N，$EigenVec_n$ 为第 n 个训练样本的特征向量，且：

$$EigenVec_n=(EigenVal_{n,1},EigenVal_{n,2},...,EigenVal_{n,en},...,EigenVal_{n,EN}),$$

$EigenVal_{n,en}$ 为第 n 个训练样本的特征向量在第 en 个评估维度上的特征信息，1≤en≤EN，EN 为评估维度的总数，$LabelVec_n$ 为第 n 个训练样本的标签向量，且：

$$LabelVec_n=(LabelVal_{n,1},LabelVal_{n,2},...,LabelVal_{n,ln},...,LabelVal_{n,LN}),$$

$LabelVal_{n,ln}$ 为第 n 个训练样本的标签向量在第 ln 个标签维度上的标签值，1≤ln≤LN，LN 为标签维度的总数。

3. 根据权利要求 2 所述的用户画像方法，其特征在于，所述分别计算各种排列顺序的样本平均距离包括：

根据下式对所述用户的特征向量进行扩展：

$$ExTgtEgVec_{CandiSq,ln}=(ExTgtEgVec_{CandiSq,ln-1},SelTgtLbVal_{CandiSq,ln-1})$$

其中，$ExTgtEgVec_{CandiSq,ln}$ 为所述用户在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，1≤CandiSq≤LN！，LN！为 LN 的阶乘，特殊地，设置 $ExTgtEgVec_{CandiSq,0}=TgtEgVec$，$TgtEgVec$ 为所述用户的特征向量，$SelTgtLbVal_{CandiSq,ln}$ 为所述用户与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelTgtLbVal_{CandiSq,0}=\varnothing$；

根据下式对所述训练样本集合的各个训练样本的特征向量进行扩展：

$$ExEigenVec_{CandiSq,ln,n}=(ExEigenVec_{CandiSq,ln-1,n},SelLabelVal_{CandiSq,ln-1,n})$$

其中，$ExEigenVec_{CandiSq,ln,n}$ 为第 n 个训练样本在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，特殊地，设置 $ExEigenVec_{CandiSq,0,n}=EigenVec_n$，$SelLabelVal_{CandiSq,ln,n}$ 为第 n 个训练样本与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelLabelVal_{CandiSq,0,n}=\varnothing$；

根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离。

4. 根据权利要求 3 所述的用户画像方法，其特征在于，所述根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离包括：

根据下式计算第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离：

$$Dis(CandiSq,ln,clvn)=\frac{\sum_{SpSeq=1}^{CValCount(CandiSq,ln,clvn)}\left(\sum_{i=1}^{EN+ln-1}(ExTgtElm(CandiSq,ln,i)-ExSpElm(CandiSq,ln,clvn,SpSeq,i))^2\right)}{CValCount(CandiSq,ln,clvn)}$$

其中，$ExTgtElm(CandiSq,ln,i)$ 为所述用户扩展得到的特征向量的第 i 个元素，$ExSpElm(CandiSq,ln,clvn,SpSeq,i)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的第 $SpSeq$ 个训练样本扩展得到的特征向量的第 i 个元素，$1\le clvn\le CLVN(CandiSq,ln)$，$CLVN(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的标签值的总数，$1\le SpSeq\le CValCount(CandiSq,ln,clvn)$，$CValCount(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值在所述训练样本集合中出现的次数，$Dis(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离；

根据下式分别构造第 CandiSq 种排列顺序中的各个分类器的样本平均距离序列：

$$DisSq(CandiSq,ln)=(Dis(CandiSq,ln,1),Dis(CandiSq,ln,2),...,$$
$$Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，$DisSq(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离序列；

根据下式分别计算第 CandiSq 种排列顺序中的各个分类器的样本平均距离：

$$MinDis(CandiSq,ln)=MIN(DisSq(CandiSq,ln))$$
$$=MIN(Dis(CandiSq,ln,1),Dis(t,sn,2),...,Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，MIN 为求最小值函数，$MinDis(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离；

根据下式计算第 CandiSq 种排列顺序的样本平均距离：

$$TotalDis_{CandiSq}=\frac{\sum_{ln=1}^{LN}MinDis(CandiSq,ln)}{LN}$$

其中，$TotalDis_{CandiSq}$ 为第 CandiSq 种排列顺序的样本平均距离。

5. 根据权利要求 1 至 4 中任一项所述的用户画像方法，其特征在于，所述训练样本集合满足下列条件：

$$\frac{\sum_{lvn_{ln}=1}^{LVN_{ln}}(ValCount_{lvn_{ln}}-\frac{N}{LVN_{ln}})^2}{LVN_{ln}}<Threshold$$

其中，LVN_{ln} 为第 ln 个标签维度上的标签值的总数，每个标签值的序号记为 lvn_{ln}，$1\le lvn_{ln}\le LVN_{ln}$，$ValCount_{lvn_{ln}}$ 为第 lvn_{ln} 个标签值在所述训练样本集合中出现的次数，且 $\sum_{lvn_{ln}=1}^{LVN_{ln}}ValCount_{lvn_{ln}}=N$，$Threshold$ 为预设的均衡度阈值。

6. 一种用户画像装置，其特征在于，包括：

特征向量构造模块，用于获取用户在预设的各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量；

训练样本选取模块，用于从预设的历史用户信息数据库中选取 N 个训练样本，并组成训练样本集合，N 为正整数；

遍历计算模块，用于将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离，所述分类器集合中包括两个以上的不同分类器，每个分类器用于确定一个标签维度上的标签值；

标签向量构造模块，用于从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量。

7. 根据权利要求 6 所述的用户画像装置，其特征在于，所述训练样本集合中的每个训练样本均包括一个历史用户的特征向量和标签向量，即：

$$Sample_n=(EigenVec_n,LabelVec_n)$$

其中，$Sample_n$ 为所述训练样本集合中的第 n 个训练样本，1≤n≤N，$EigenVec_n$ 为第 n 个训练样本的特征向量，且：

$$EigenVec_n=(EigenVal_{n,1},EigenVal_{n,2},...,EigenVal_{n,en},...,EigenVal_{n,EN})$$，

$EigenVal_{n,en}$ 为第 n 个训练样本的特征向量在第 en 个评估维度上的特征信息，1≤en≤EN，EN 为评估维度的总数，$LabelVec_n$ 为第 n 个训练样本的标签向量，且：

$$LabelVec_n=(LabelVal_{n,1},LabelVal_{n,2},...,LabelVal_{n,ln},...,LabelVal_{n,LN})$$，

$LabelVal_{n,ln}$ 为第 n 个训练样本的标签向量在第 ln 个标签维度上的标签值，1≤ln≤LN，LN 为标签维度的总数。

8. 根据权利要求 7 所述的用户画像装置，其特征在于，所述遍历计算模块包括：

第一扩展单元，用于根据下式对所述用户的特征向量进行扩展：

$$ExTgtEgVec_{CandiSq,ln}=(ExTgtEgVec_{CandiSq,ln-1},SelTgtLbVal_{CandiSq,ln-1})$$

其中，$ExTgtEgVec_{CandiSq,ln}$ 为所述用户在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，1≤CandiSq≤LN！，LN！为 LN 的阶乘，特殊地，设置 $ExTgtEgVec_{CandiSq,0}=TgtEgVec$，$TgtEgVec$ 为所述用户的特征向量，$SelTgtLbVal_{CandiSq,ln}$ 为所述用户与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelTgtLbVal_{CandiSq,0}=\varnothing$；

第二扩展单元，用于根据下式对所述训练样本集合的各个训练样本的特征向量进行扩展：

$$ExEigenVec_{CandiSq,ln,n}=(ExEigenVec_{CandiSq,ln-1,n},SelLabelVal_{CandiSq,ln-1,n})$$

其中，$ExEigenVec_{CandiSq,ln,n}$ 为第 n 个训练样本在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，特殊地，设置 $ExEigenVec_{CandiSq,0,n}=EigenVec_n$，$SelLabelVal_{CandiSq,ln,n}$ 为第 n 个训练样本与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelLabelVal_{CandiSq,0,n}=\varnothing$；

样本平均距离计算单元，用于根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离。

9. 根据权利要求 8 所述的用户画像装置，其特征在于，所述样本平均距离计算单元包括：

第一子单元，用于根据下式计算第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离：

$$Dis(CandiSq,ln,clvn)=\frac{\sum_{SpSeq=1}^{CValCount(CandiSq,ln,clvn)}\left(\sum_{i=1}^{EN+ln-1}(ExTgtElm(CandiSq,ln,i)-ExSpElm(CandiSq,ln,clvn,SpSeq,i))^2\right)}{CValCount(CandiSq,ln,clvn)}$$

其中，$ExTgtElm(CandiSq,ln,i)$ 为所述用户扩展得到的特征向量的第 i 个元素，$ExSpElm(CandiSq,ln,clvn,SpSeq,i)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的第 $SpSeq$ 个训练样本扩展得到的特征向量的第 i 个元素，$1\le clvn\le CLVN(CandiSq,ln)$，$CLVN(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的标签值的总数，$1\le SpSeq\le CValCount(CandiSq,ln,clvn)$，$CValCount(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值在所述训练样本集合中出现的次数，$Dis(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中

的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离；

第二子单元，用于根据下式分别构造第 CandiSq 种排列顺序中的各个分类器的样本平均距离序列：

$$DisSq(CandiSq,ln)=(Dis(CandiSq,ln,1),Dis(CandiSq,ln,2),...,$$
$$Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，$DisSq(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离序列；

第三子单元，用于根据下式分别计算第 CandiSq 种排列顺序中的各个分类器的样本平均距离：

$$MinDis(CandiSq,ln)=MIN(DisSq(CandiSq,ln))$$
$$=MIN(Dis(CandiSq,ln,1),Dis(t,sn,2),...,Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，MIN 为求最小值函数，$MinDis(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离；

第四子单元，用于根据下式计算第 CandiSq 种排列顺序的样本平均距离：

$$TotalDis_{CandiSq}=\frac{\sum_{ln=1}^{LN}MinDis(CandiSq,ln)}{LN}$$

其中，$TotalDis_{CandiSq}$ 为第 CandiSq 种排列顺序的样本平均距离。

10. 根据权利要求 6 至 9 中任一项所述的用户画像装置，其特征在于，所述训练样本集合满足下列条件：

$$\frac{\sum_{lvn_{ln}=1}^{LVN_{ln}}(ValCount_{lvn_{ln}}-\frac{N}{LVN_{ln}})^2}{LVN_{ln}}<Threshold$$

其中，LVN_{ln} 为第 ln 个标签维度上的标签值的总数，每个标签值的序号记为 lvn_{ln}，$1\le lvn_{ln}\le LVN_{ln}$，$ValCount_{lvn_{ln}}$ 为第 lvn_{ln} 个标签值在所述训练样本集合中出现的次数，且 $\sum_{lvn_{ln}=1}^{LVN_{ln}}ValCount_{lvn_{ln}}=N$，$Threshold$ 为预设的均衡度阈值。

11. 一种计算机可读存储介质，所述计算机可读存储介质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如下步骤：

获取用户在预设的各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量；

从预设的历史用户信息数据库中选取 N 个训练样本，并组成训练样本集合，N 为正整数；

将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离，所述分类器集

合中包括两个以上的不同分类器，每个分类器用于确定一个标签维度上的标签值；

从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量。

12. 根据权利要求 11 所述的计算机可读存储介质，其特征在于，所述训练样本集合中的每个训练样本均包括一个历史用户的特征向量和标签向量，即：

$$Sample_n=(EigenVec_n,LabelVec_n)$$

其中，$Sample_n$ 为所述训练样本集合中的第 n 个训练样本，1≤n≤N，$EigenVec_n$ 为第 n 个训练样本的特征向量，且：

$$EigenVec_n=(EigenVal_{n,1},EigenVal_{n,2},...,EigenVal_{n,en},...,EigenVal_{n,EN}),$$

$EigenVal_{n,en}$ 为第 n 个训练样本的特征向量在第 en 个评估维度上的特征信息，1≤en≤EN，EN 为评估维度的总数，$LabelVec_n$ 为第 n 个训练样本的标签向量，且：

$$LabelVec_n=(LabelVal_{n,1},LabelVal_{n,2},...,LabelVal_{n,ln},...,LabelVal_{n,LN}),$$

$LabelVal_{n,ln}$ 为第 n 个训练样本的标签向量在第 ln 个标签维度上的标签值，1≤ln≤LN，LN 为标签维度的总数。

13. 根据权利要求 12 所述的计算机可读存储介质，其特征在于，所述分别计算各种排列顺序的样本平均距离包括：

根据下式对所述用户的特征向量进行扩展：

$$ExTgtEgVec_{CandiSq,ln}=(ExTgtEgVec_{CandiSq,ln-1},SelTgtLbVal_{CandiSq,ln-1})$$

其中，$ExTgtEgVec_{CandiSq,ln}$ 为所述用户在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，1≤CandiSq≤LN！，LN！为 LN 的阶乘，特殊地，设置 $ExTgtEgVec_{CandiSq,0}=TgtEgVec$，$TgtEgVec$ 为所述用户的特征向量，$SelTgtLbVal_{CandiSq,ln}$ 为所述用户与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelTgtLbVal_{CandiSq,0}=\varnothing$；

根据下式对所述训练样本集合的各个训练样本的特征向量进行扩展：

$$ExEigenVec_{CandiSq,ln,n}=(ExEigenVec_{CandiSq,ln-1,n},SelLabelVal_{CandiSq,ln-1,n})$$

其中，$ExEigenVec_{CandiSq,ln,n}$ 为第 n 个训练样本在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，特殊地，设置 $ExEigenVec_{CandiSq,0,n}=EigenVec_n$，$SelLabelVal_{CandiSq,ln,n}$ 为第 n 个训练样本与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelLabelVal_{CandiSq,0,n}=\varnothing$；

根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离。

14. 根据权利要求 13 所述的计算机可读存储介质，其特征在于，所述根据所述用

户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离包括：

根据下式计算第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离：

$$Dis(CandiSq,ln,clvn)=\frac{\sum_{SpSeq=1}^{CValCount(CandiSq,ln,clvn)}(\sum_{i=1}^{EN+ln-1}(ExTgtElm(CandiSq,ln,i)-ExSpElm(CandiSq,ln,clvn,SpSeq,i))^2)}{CValCount(CandiSq,ln,clvn)}$$

其中，$ExTgtElm(CandiSq,ln,i)$ 为所述用户扩展得到的特征向量的第 i 个元素，$ExSpElm(CandiSq,ln,clvn,SpSeq,i)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的第 $SpSeq$ 个训练样本扩展得到的特征向量的第 i 个元素，$1\le clvn\le CLVN(CandiSq,ln)$，$CLVN(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的标签值的总数，$1\le SpSeq\le CValCount(CandiSq,ln,clvn)$，$CValCount(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值在所述训练样本集合中出现的次数，$Dis(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离；

根据下式分别构造第 CandiSq 种排列顺序中的各个分类器的样本平均距离序列：

$$DisSq(CandiSq,ln)=(Dis(CandiSq,ln,1),Dis(CandiSq,ln,2),...,Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，$DisSq(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离序列；

根据下式分别计算第 CandiSq 种排列顺序中的各个分类器的样本平均距离：

$$MinDis(CandiSq,ln)=MIN(DisSq(CandiSq,ln))$$
$$=MIN(Dis(CandiSq,ln,1),Dis(t,sn,2),...,Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，MIN 为求最小值函数，$MinDis(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离；

根据下式计算第 CandiSq 种排列顺序的样本平均距离：

$$TotalDis_{CandiSq}=\frac{\sum_{ln=1}^{LN}MinDis(CandiSq,ln)}{LN}$$

其中，$TotalDis_{CandiSq}$ 为第 CandiSq 种排列顺序的样本平均距离。

15. 根据权利要求 11 至 14 中任一项所述的计算机可读存储介质，其特征在于，所述训练样本集合满足下列条件：

$$\frac{\sum_{lvn_{ln}=1}^{LVN_{ln}}(ValCount_{lvn_{ln}}-\frac{N}{LVN_{ln}})^2}{LVN_{ln}}<Threshold$$

其中，LVN_{ln}为第 ln 个标签维度上的标签值的总数，每个标签值的序号记为lvn_{ln}，$1 \le lvn_{ln} \le LVN_{ln}$，$ValCount_{lvn_{ln}}$为第$lvn_{ln}$个标签值在所述训练样本集合中出现的次数，且$\sum_{lvn_{ln}=1}^{LVN_{ln}} ValCount_{lvn_{ln}} = N$，$Threshold$为预设的均衡度阈值。

16. 一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下步骤：

获取用户在预设的各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量；

从预设的历史用户信息数据库中选取 N 个训练样本，并组成训练样本集合，N 为正整数；

将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离，所述分类器集合中包括两个以上的不同分类器，每个分类器用于确定一个标签维度上的标签值；

从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量。

17. 根据权利要求 16 所述的终端设备，其特征在于，所述训练样本集合中的每个训练样本均包括一个历史用户的特征向量和标签向量，即：

$$Sample_n = (EigenVec_n, LabelVec_n)$$

其中，$Sample_n$为所述训练样本集合中的第 n 个训练样本，1≤n≤N，$EigenVec_n$为第 n 个训练样本的特征向量，且：

$$EigenVec_n = (EigenVal_{n,1}, EigenVal_{n,2}, ..., EigenVal_{n,en}, ..., EigenVal_{n,EN})$$，

$EigenVal_{n,en}$为第 n 个训练样本的特征向量在第 en 个评估维度上的特征信息，1≤en≤EN，EN 为评估维度的总数，$LabelVec_n$为第 n 个训练样本的标签向量，且：

$$LabelVec_n = (LabelVal_{n,1}, LabelVal_{n,2}, ..., LabelVal_{n,ln}, ..., LabelVal_{n,LN})$$，

$LabelVal_{n,ln}$为第 n 个训练样本的标签向量在第 ln 个标签维度上的标签值，1≤ln≤LN，LN 为标签维度的总数。

18. 根据权利要求 17 所述的终端设备，其特征在于，所述分别计算各种排列顺序的样本平均距离包括：

根据下式对所述用户的特征向量进行扩展：

$$ExTgtEgVec_{CandiSq,ln} = (ExTgtEgVec_{CandiSq,ln-1}, SelTgtLbVal_{CandiSq,ln-1})$$

其中，$ExTgtEgVec_{CandiSq,ln}$为所述用户在第 CandiSq 种排列顺序中的第 ln 个分类器

的处理过程中扩展得到的特征向量，1≤CandiSq≤LN！，LN！为 LN 的阶乘，特殊地，设置 $ExTgtEgVec_{CandiSq,0}=TgtEgVec$，$TgtEgVec$ 为所述用户的特征向量，$SelTgtLbVal_{CandiSq,ln}$ 为所述用户与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelTgtLbVal_{CandiSq,0}=\varnothing$；

根据下式对所述训练样本集合的各个训练样本的特征向量进行扩展：

$$ExEigenVec_{CandiSq,ln,n}=(ExEigenVec_{CandiSq,ln-1,n},SelLabelVal_{CandiSq,ln-1,n})$$

其中，$ExEigenVec_{CandiSq,ln,n}$ 为第 n 个训练样本在第 CandiSq 种排列顺序中的第 ln 个分类器的处理过程中扩展得到的特征向量，特殊地，设置 $ExEigenVec_{CandiSq,0,n}=EigenVec_n$，$SelLabelVal_{CandiSq,ln,n}$ 为第 n 个训练样本与第 CandiSq 种排列顺序中的第 ln 个分类器对应的标签值，特殊地，设置 $SelLabelVal_{CandiSq,0,n}=\varnothing$；

根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离。

19. 根据权利要求 18 所述的终端设备，其特征在于，所述根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离包括：

根据下式计算第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离：

$$Dis(CandiSq,ln,clvn)=\frac{\sum_{SpSeq=1}^{CValCount(CandiSq,ln,clvn)}\left(\sum_{i=1}^{EN+\ln-1}(ExTgtElm(CandiSq,ln,i)-ExSpElm(CandiSq,ln,clvn,SpSeq,i))^2\right)}{CValCount(CandiSq,ln,clvn)}$$

其中，$ExTgtElm(CandiSq,ln,i)$ 为所述用户扩展得到的特征向量的第 i 个元素，$ExSpElm(CandiSq,ln,clvn,SpSeq,i)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的第 $SpSeq$ 个训练样本扩展得到的特征向量的第 i 个元素，$1\le clvn\le CLVN(CandiSq,ln)$，$CLVN(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的标签值的总数，$1\le SpSeq\le CValCount(CandiSq,ln,clvn)$，$CValCount(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值在所述训练样本集合中出现的次数，$Dis(CandiSq,ln,clvn)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的第 $clvn$ 个标签值的样本平均距离；

根据下式分别构造第 CandiSq 种排列顺序中的各个分类器的样本平均距离序列：

$$DisSq(CandiSq,ln)=(Dis(CandiSq,ln,1),Dis(CandiSq,ln,2),...,$$
$$Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，$DisSq(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离序列；

根据下式分别计算第 CandiSq 种排列顺序中的各个分类器的样本平均距离：

$$MinDis(CandiSq,ln)=MIN(DisSq(CandiSq,ln))$$
$$=MIN(Dis(CandiSq,ln,1),Dis(t,sn,2),...,Dis(CandiSq,ln,clvn),...,Dis(CandiSq,ln,CLVN(CandiSq,ln)))$$

其中，MIN 为求最小值函数，$MinDis(CandiSq,ln)$ 为第 CandiSq 种排列顺序中的第 ln 个分类器的样本平均距离；

根据下式计算第 CandiSq 种排列顺序的样本平均距离：

$$TotalDis_{CandiSq}=\frac{\sum_{ln=1}^{LN}MinDis(CandiSq,ln)}{LN}$$

其中，$TotalDis_{CandiSq}$ 为第 CandiSq 种排列顺序的样本平均距离。

20. 根据权利要求 16 至 19 中任一项所述的终端设备，其特征在于，所述训练样本集合满足下列条件：

$$\frac{\sum_{lvn_{ln}=1}^{LVN_{ln}}(ValCount_{lvn_{ln}}-\frac{N}{LVN_{ln}})^2}{LVN_{ln}}<Threshold$$

其中，LVN_{ln} 为第 ln 个标签维度上的标签值的总数，每个标签值的序号记为 lvn_{ln}，$1\le lvn_{ln}\le LVN_{ln}$，$ValCount_{lvn_{ln}}$ 为第 lvn_{ln} 个标签值在所述训练样本集合中出现的次数，且 $\sum_{lvn_{ln}=1}^{LVN_{ln}}ValCount_{lvn_{ln}}=N$，$Threshold$ 为预设的均衡度阈值。

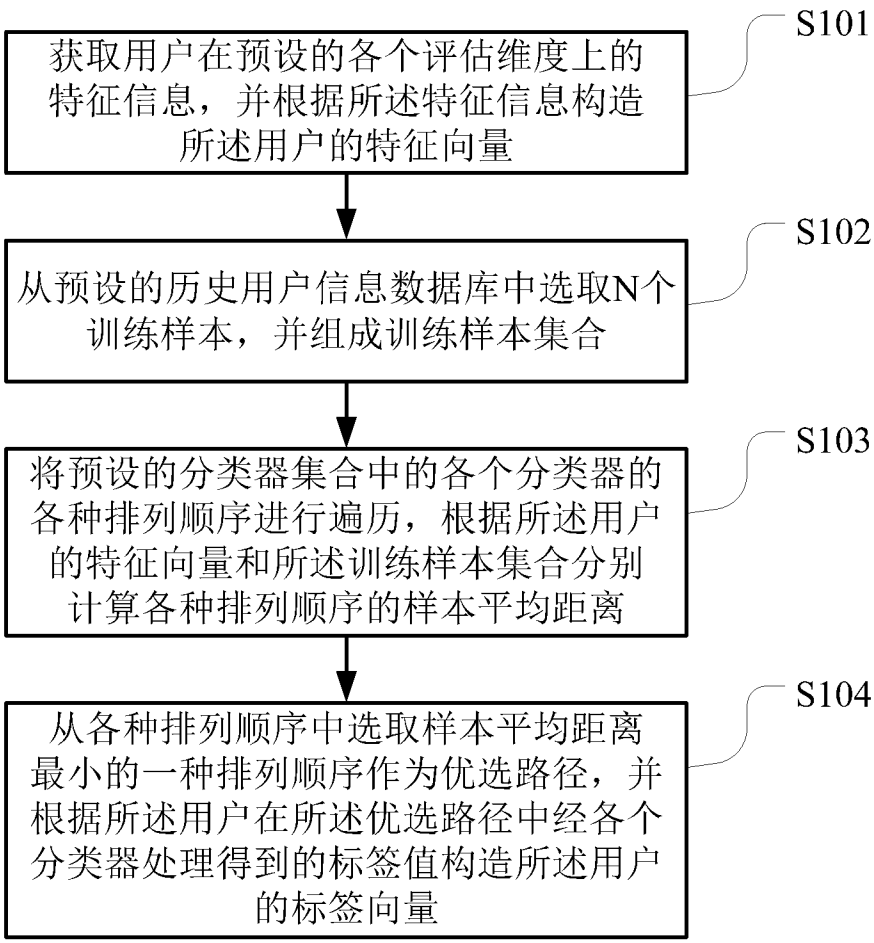
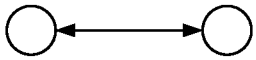
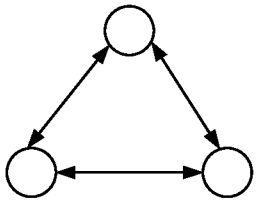

S101
获取用户在预设的各个评估维度上的特征信息，并根据所述特征信息构造所述用户的特征向量
S102
从预设的历史用户信息数据库中选取N个训练样本，并组成训练样本集合
S103
将预设的分类器集合中的各个分类器的各种排列顺序进行遍历，根据所述用户的特征向量和所述训练样本集合分别计算各种排列顺序的样本平均距离
S104
从各种排列顺序中选取样本平均距离最小的一种排列顺序作为优选路径，并根据所述用户在所述优选路径中经各个分类器处理得到的标签值构造所述用户的标签向量

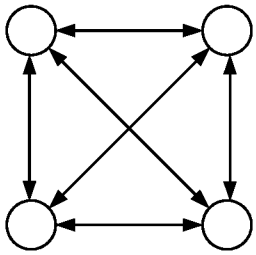
图 1



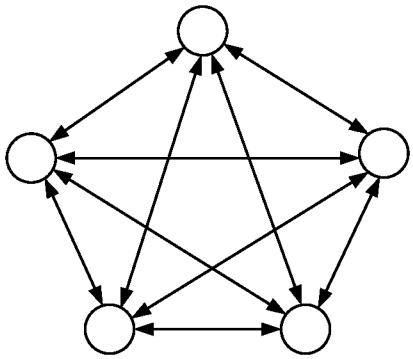
LN=2



LN=3



LN=4



LN=5

.....

.....

图 2

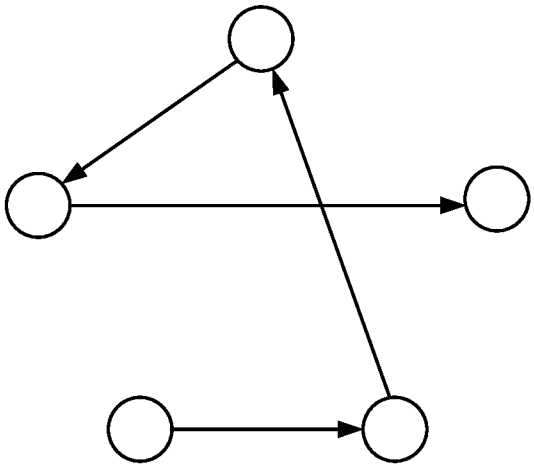


图 3

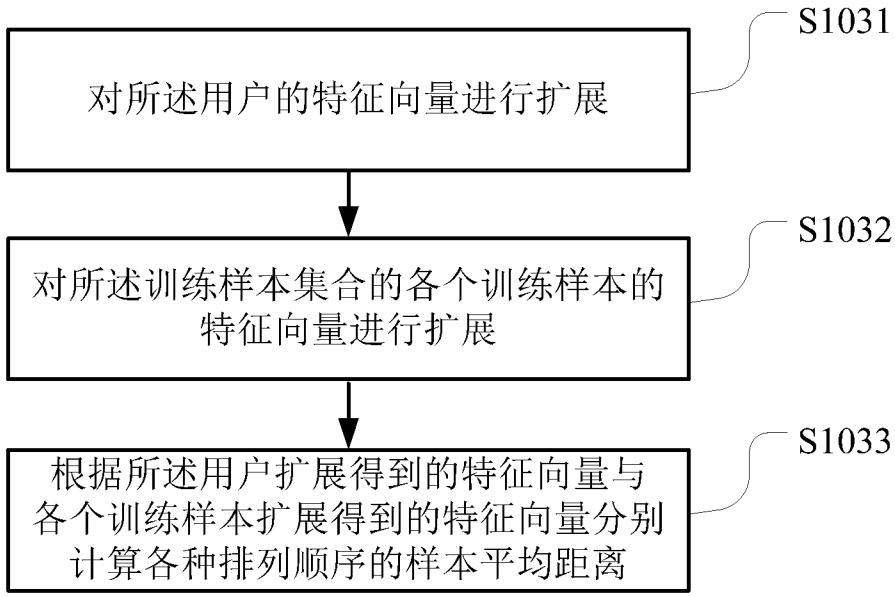

S1031
对所述用户的特征向量进行扩展
S1032
对所述训练样本集合的各个训练样本的特征向量进行扩展
S1033
根据所述用户扩展得到的特征向量与各个训练样本扩展得到的特征向量分别计算各种排列顺序的样本平均距离

图 4

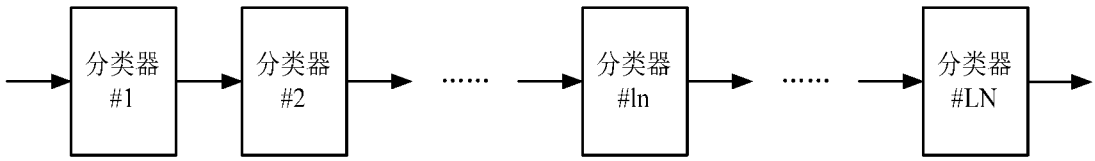

分类器 #1
分类器 #2
.....
分类器 #ln
.....
分类器 #LN

图 5

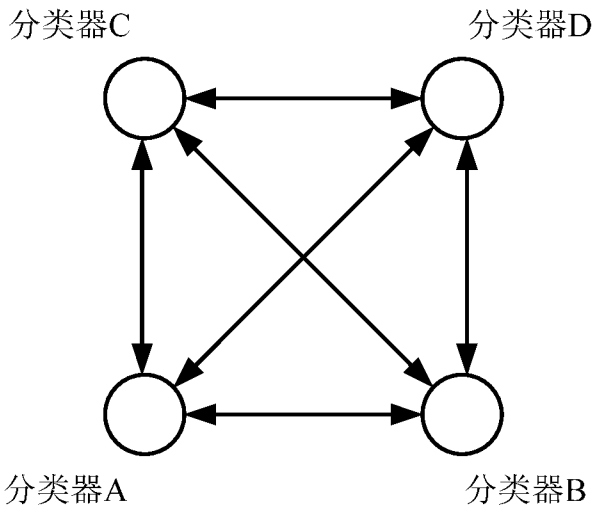

分类器C
分类器D
分类器A
分类器B

图 6

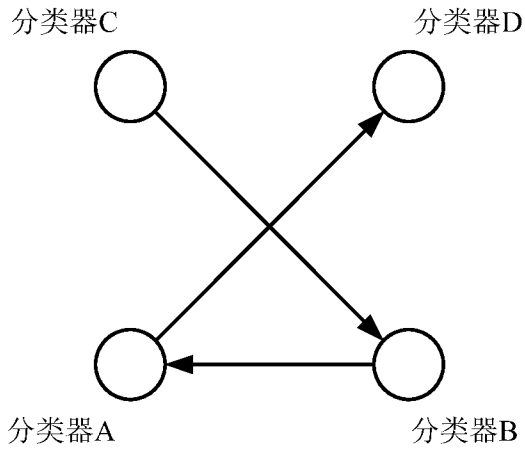

分类器C
分类器D
分类器A
分类器B

图 7

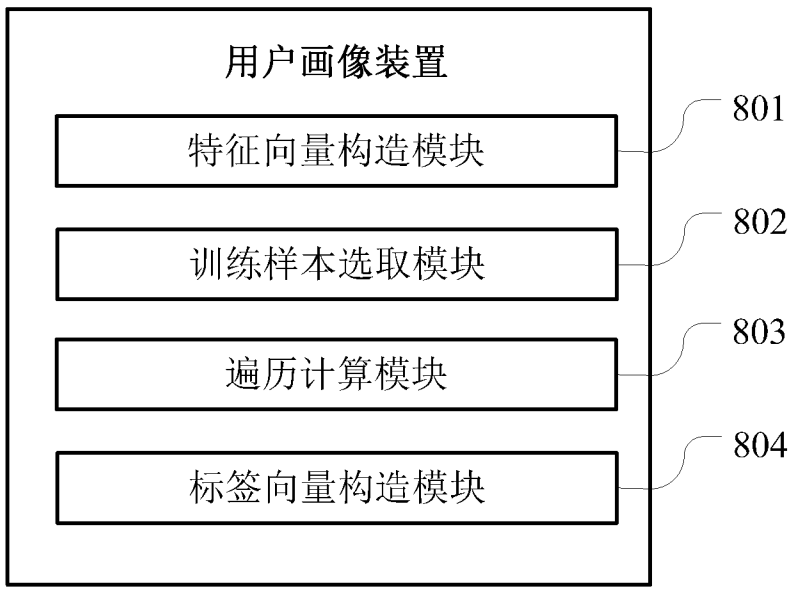

用户画像装置
特征向量构造模块
训练样本选取模块
遍历计算模块
标签向量构造模块
801
802
803
804

图 8

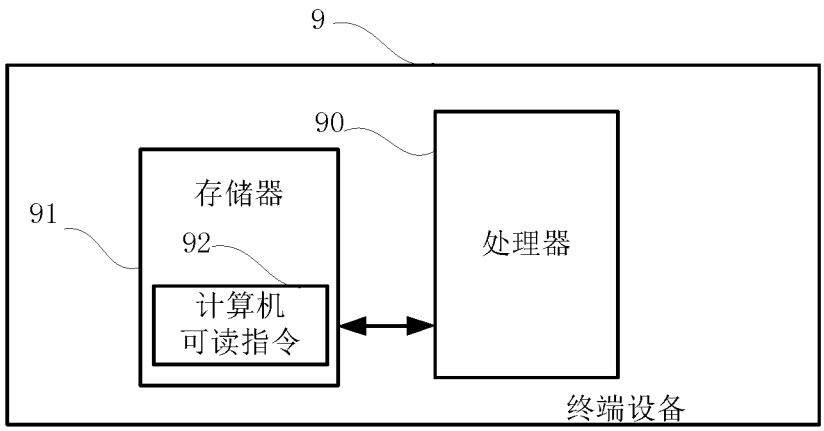

9
90
存储器
91
92
计算机
可读指令
处理器
终端设备

图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/091529

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/951(2019.01)i; G06F 16/953(2019.01)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)

G06F,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, WPI, EPODOC, IEEE: 平安, 杨晟, 用户, 画像, 标签, 分类器, 训练, 遍历, 最优, 路径, 拓扑, 特征, 向量, 维度, 多维, 高维, 顺序, 排序, user, image, label, classifier, train+, traversal, optimal, path, topology, characteristic, vector, dimension, multi+, many, high, order, sequence

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109885745 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 14 June 2019 (2019-06-14) entire document	1-20
Y	CN 108229590 A (ALIBABA GROUP HOLDING LIMITED) 29 June 2018 (2018-06-29) description, paragraphs [0023]-[0054], and figures 1-3	1-20
Y	CN 104615730 A (INSPUR GROUP LTD.) 13 May 2015 (2015-05-13) description, paragraphs [0042]-[0046]	1-20
A	CN 109002490 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 14 December 2018 (2018-12-14) entire document	1-20
A	CN 107633036 A (GUANGZHOU WANGWANG INFORMATION TECHNOLOGY CO., LTD.) 26 January 2018 (2018-01-26) entire document	1-20
A	CN 108629633 A (INSPUR SOFTWARE CO., LTD.) 09 October 2018 (2018-10-09) 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	Date of mailing of the international search report
09 July 2019	**25 July 2019**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 106779929 A (BEIJING KNOWNSEC INFORMATION TECHNOLOGY CO., LTD.) 31 May 2017 (2017-05-31) entire document	1-20
A	KR 2018131221 A (SKTE SK PLANET CO., LTD.) 10 December 2018 (2018-12-10) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/091529

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 109885745 A	14 June 2019	None	
CN 108229590 A	29 June 2018	None	
CN 104615730 A	13 May 2015	CN 104615730 B	27 October 2017
CN 109002490 A	14 December 2018	None	
CN 107633036 A	26 January 2018	None	
CN 108629633 A	09 October 2018	None	
CN 106779929 A	31 May 2017	None	
KR 2018131221 A	10 December 2018	None	

国际检索报告

国际申请号
PCT/CN2019/091529

A. 主题的分类

G06F 16/951(2019.01)i; G06F 16/953(2019.01)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)

G06F,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, WPI, EPODOC, IEEE:平安，杨晟，用户，画像，标签，分类器，训练，遍历，最优，路径，拓扑，特征，向量，维度，多维，高维，顺序，排序，user,image, label,classifier,train+,traversal,optimal,path,topology,characteristic,vector,dimension,multi+,many,high,order,sequence

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 109885745 A (平安科技深圳有限公司) 2019年 6月 14日 (2019-06-14) 全文	1-20
Y	CN 108229590 A (阿里巴巴集团控股有限公司) 2018年 6月 29日 (2018-06-29) 说明书第[0023]-[0054]段，图1-3	1-20
Y	CN 104615730 A (浪潮集团有限公司) 2015年 5月 13日 (2015-05-13) 说明书第[0042]-[0046]段	1-20
A	CN 109002490 A (腾讯科技深圳有限公司) 2018年 12月 14日 (2018-12-14) 全文	1-20
A	CN 107633036 A (广州汪汪信息技术有限公司) 2018年 1月 26日 (2018-01-26) 全文	1-20
A	CN 108629633 A (浪潮软件股份有限公司) 2018年 10月 9日 (2018-10-09) 全文	1-20
A	CN 106779929 A (北京知道创宇信息技术有限公司) 2017年 5月 31日 (2017-05-31) 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	国际检索报告邮寄日期
2019年 7月 9日	2019年 7月 25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 姜丹 电话号码 86-010-53961823

国际检索报告

国际申请号
PCT/CN2019/091529

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
A	KR 2018131221 A (SKTE SK PLANET CO., LTD.) 2018年 12月 10日 (2018-12-10) 全文	1-20

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/091529

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 109885745 A	2019年 6月 14日	无	
CN 108229590 A	2018年 6月 29日	无	
CN 104615730 A	2015年 5月 13日	CN 104615730 B	2017年 10月 27日
CN 109002490 A	2018年 12月 14日	无	
CN 107633036 A	2018年 1月 26日	无	
CN 108629633 A	2018年 10月 9日	无	
CN 106779929 A	2017年 5月 31日	无	
KR 2018131221 A	2018年 12月 10日	无	