

# (12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局



(10) 国际公布号

WO 2020/181910 A1

(43) 国际公布日

2020 年 9 月 17 日 (17.09.2020)

(51) 国际专利分类号:

G06Q 10/06 (2012.01)

G06Q 30/06 (2012.01)

G06Q 30/00 (2012.01)

(72) 发明人: 印晓华(YIN, Xiaohua); 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(21) 国际申请号:

PCT/CN2020/070672

(22) 国际申请日:

2020 年 1 月 7 日 (07.01.2020)

(25) 申请语言:

中文

(26) 公布语言:

中文

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街9号嘉华大厦B座409, Beijing 100085 (CN)。

(30) 优先权:

201910192832.0

2019年3月14日 (14.03.2019) 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,

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层847号邮箱, Grand Cayman (KY)。

(54) Title: METHOD AND DEVICE FOR IDENTIFYING A RISK MERCHANT

(54) 发明名称: 识别风险商家的方法及装置

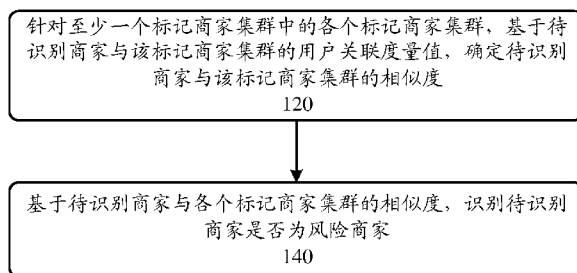


图 1

120 For each marked merchant cluster in at least one marked merchant cluster, on the basis of a user correlation degree value of a merchant to be identified and the marked merchant cluster, determine a similarity between the merchant to be identified and the marked merchant cluster  
140 On the basis of the similarity between the merchant to be identified and each marked merchant cluster, identify whether the merchant to be identified is a risk merchant

(57) Abstract: A method and device for identifying a risk merchant. The method comprises: for each marked merchant cluster in at least one marked merchant cluster, on the basis of a user correlation degree value of a merchant to be identified and the marked merchant cluster, determining a similarity between the merchant to be identified and the marked merchant cluster (120), each marked merchant in each marked merchant cluster being marked as a risk merchant of a corresponding risk category; and on the basis of the similarity between the merchant to be identified and each marked merchant cluster, identifying whether the merchant to be identified is a risk merchant (140).

(57) 摘要: 一种识别风险商家的方法及装置。该方法包括: 针对至少一个标记商家集群中的各个标记商家集群, 基于待识别商家与该标记商家集群的用户关联度量值, 确定所述待识别商家与该标记商家集群的相似度 (120), 所述各个标记商家集群中的各个标记商家被标记为相应风险类别的风险商家; 以及基于所述待识别商家与所述各个标记商家集群的相似度, 识别待识别商家是否为风险商家 (140)。

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,  
UZ, VC, VN, WS, 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))。

## 识别风险商家的方法及装置

### 技术领域

[01] 本公开涉及计算机技术领域，具体地，涉及识别风险商家的方法及装置。

### 背景技术

5 [02] 随着电商技术的发展，越来越来的交易通过互联网进行。在互联网交易过程中，反风险团队一般会对可疑交易行为进行监控，以及时识别出有风险的行为。

[03] 与此同时，可疑分子也会竭尽手段规避反风险团队的监控。因此，监控风险行为的任务量变得越来越庞大。此外，互联网交易快速便捷的特点也使得需要监控的交易行为的数量越来越大。因此，现有技术亟需能够协助反风险团队快速有效识别风险交易的技术。

10

### 发明内容

[04] 鉴于上述，本公开提供了一种识别风险商家的方法及装置，利用该方法和装置，通过基于被标记为风险商家的标记商家集群与待识别商家的相似度，来识别待识别商家中的风险商家，不仅能够准确识别风险商家，而且能够降低识别过程的计算量，提高识别效率。

15

[05] 根据本公开的一个方面，提供了一种识别风险商家的方法，包括：针对至少一个标记商家集群中的各个标记商家集群，基于待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度，所述各个标记商家集群中的各个标记商家被标记为风险商家；以及基于所述待识别商家与所述各个标记商家集群的相似度，识别所述待识别商家是否为风险商家。

20

[06] 可选的，在一个示例中，所述待识别商家与所述标记商家集群的用户关联度量值可以包括所述待识别商家与所述标记商家集群的用户关联系数和所述待识别商家与所述标记商家集群的相同用户数量。其中，所述用户关联系数可以包括：在所述待识别商家的用户中，与所述标记商家集群的相同用户所占的比例；和/或所述待识别商家与所述标记商家集群的用户特征相似度。

25

[07] 可选的，在一个示例中，所述各个标记商家集群具有至少一个代表标记商家，基

于至少一个待识别商家中的待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度包括：基于针对该标记商家集群的所述待识别商家的向量表示和该标记商家集群的至少一个代表标记商家的向量表示，确定所述待识别商家与该至少一个代表标记商家的相似度；以及基于所述待识别商家与该至少一个代表标记商家的相似度，确定所述待识别商家与该标记商家集群的相似度。其中，针对该标记商家集群的商家的向量表示是基于所述商家与该标记商家集群的用户关联系数和所述商家与该标记商家集群的相同用户数量的向量表示。

[08] 可选的，在一个示例中，在基于所述待识别商家的向量表示和该标记商家集群的至少一个代表标记商家的向量表示，确定所述待识别商家与该至少一个代表标记商家的相似度之前，基于待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度还包括：将所述向量表示中的相同用户数量维度和用户关联系数维度处理为具有相同的取值范围。

[09] 可选的，在一个示例中，所述各个标记商家集群的至少一个代表标记商家是基于该标记商家集群中的各个标记商家的用户数量确定的。

[10] 可选的，在一个示例中，所述待识别商家与所述各个标记商家集群的至少一个代表标记商家的相似度用以下中的任一者来表征：欧氏距离、曼哈顿距离和夹角余弦距离。

[11] 可选的，在一个示例中，基于所述待识别商家与所述各个标记商家集群的相似度，识别所述至少一个待识别商家中的风险商家包括：针对所述各个标记商家集群，将至少两个待识别商家与该标记商家集群的相似度中的大于预定阈值的相似度值最大的前预定个数的待识别商家，识别为与该标记商家集群对应类别的风险商家。

[12] 根据本公开的另一方面，还提供一种识别风险商家的装置，包括：相似度确定单元，被配置为针对至少一个标记商家集群中的各个标记商家集群，基于待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度，所述各个标记商家集群中的各个标记商家被标记为相应风险类别的风险商家；以及风险商家识别单元，被配置为基于所述待识别商家与所述各个标记商家集群的相似度，识别所述待识别商家是否为风险商家。

[13] 可选的，在一个示例中，所述待识别商家与所述标记商家集群的用户关联度量值包括所述待识别商家与所述标记商家集群的用户关联系数和所述待识别商家与所述标记商家集群的相同用户数量。其中，所述用户关联系数可以包括：在所述待识别商家的

用户中,与所述标记商家集群的相同用户所占的比例;和/或所述待识别商家与所述标记商家集群的用户特征相似度。

[14] 可选的,在一个示例中,所述各个标记商家集群具有至少一个代表标记商家,所述相似度确定单元包括:第一相似度确定模块,被配置为基于针对该标记商家集群的所述待识别商家的向量表示和该标记商家集群的至少一个代表标记商家的向量表示,确定所述待识别商家与该至少一个代表标记商家的相似度;以及第二相似度确定模块,被配置为基于所述待识别商家与该至少一个代表标记商家的相似度,确定所述待识别商家与该标记商家集群的相似度。其中,针对该标记商家集群的商家的向量表示是基于所述商家与该标记商家集群的用户关联系数和所述商家与该标记商家集群的相同用户数量的向量表示。

[15] 可选的,在一个示例中,所述相似度确定单元还包括:归一化处理模块,被配置为在基于待识别商家与该标记商家集群的用户关联度量值,确定所述待识别商家与该标记商家集群的相似度之前,将所述向量表示中的相同用户数量维度和用户关联系数维度处理为具有相同的取值范围。

[16] 可选的,在一个示例中,所述风险商家识别单元被配置为:针对所述各个标记商家集群,将至少两个待识别商家与该标记商家集群的相似度中的大于预定阈值的相似度值最大的前预定个数的待识别商家,识别为与该标记商家集群对应类别的风险商家。

[17] 根据本公开的另一方面,还提供一种计算设备,包括:至少一个处理器;以及存储器,所述存储器存储指令,当所述指令被所述至少一个处理器执行时,使得所述至少一个处理器执行如上所述的识别风险商家的方法。

[18] 根据本公开的另一方面,还提供一种非暂时性机器可读存储介质,其存储有可执行指令,所述指令当被执行时使得所述机器执行如上所述的识别风险商家的方法。

[19] 利用本公开的识别风险商家的方法和装置,通过基于待识别商家与风险的标记商家集群的用户关联度量值来计算待识别商家与标记商家集群的相似度,进而基于该相似度识别待识别商家中的风险商家,从而能够综合地基于标记商家集群中的多个标记商家的特征来进行识别,由此不仅能准确识别出风险商家,而且在识别过程中的运算量较小,因而还能够提高识别效率。

[20] 利用本公开的识别风险商家的方法和装置,基于在待识别商家与标记商家集群的用户关联系数和待识别商家与标记商家集群的相同用户数量,来计算待识别商家与标记

商家集群的相似度，从而能够基于待识别商家与标记商家集群的相对关联属性和绝对关联属性来确定待识别商家与标记商家集群的相似度，以提高风险识别的准确性。

[21] 利用本公开的识别风险商家的方法和装置，以代表标记商家来代表各个标记商家集群，从而基于待识别商家和各个标记商家集群的代表标记商家的向量表示来确定待识别商家与各个标记商家集群的相似度，能够降低风险商家识别过程的复杂度，进一步提高识别效率。

[22] 利用本公开的识别风险商家的方法和装置，在确定代表标记商家与待识别商家间的相似度之前，将向量表示中的相同用户数量维度和用户关联系数维度处理为具有相同的取值范围，从而避免因为各维度的取值范围差异导致忽略某一维度的特征，由此能够进一步提高识别风险商家的准确性。

[23] 利用本公开的识别风险商家的方法和装置，基于各个标记商家集群中的各个标记商家的用户数量来确定各个标记商家集群的代表标记商家，从而能够根据实际情况选取能够代表该标记商家集群的代表标记商家，以提高识别过程的灵活性。

## 附图说明

[24] 通过参照下面的附图，可以实现对于本公开内容的本质和优点的进一步理解。在附图中，类似组件或特征可以具有相同的附图标记。附图是用来提供对本发明实施例的进一步理解，并且构成说明书的一部分，与下面的具体实施方式一起用于解释本公开的实施例，但并不构成对本公开的实施例的限制。在附图中：

[25] 图 1 是根据本公开的一个实施例的识别风险商家的方法的流程图；

[26] 图 2 是根据本公开的一个实施例的识别风险商家的方法中的相似度确定过程的一个示例的流程图；

[27] 图 3 是根据本公开的一个实施例的识别风险商家的方法中的风险商家识别过程的一个示例的流程图；

[28] 图 4 是根据本公开的一个实施例的识别风险商家的装置的结构框图；

[29] 图 5 是根据本公开的一个实施例的识别风险商家的装置中的相似度确定单元的一个示例的结构框图；

[30] 图 6 是根据本公开的另一实施例的用于实现用于识别用户意图的方法的计算设备

的结构框图。

## 具体实施方式

[31] 以下将参考示例实施方式讨论本文描述的主题。应该理解，讨论这些实施方式只是为了使得本领域技术人员能够更好地理解从而实现本文描述的主题，并非是对权利要求书中所阐述的保护范围、适用性或者示例的限制。可以在不脱离本公开内容的保护范围的情况下，对所讨论的元素的功能和排列进行改变。各个示例可以根据需要，省略、替代或者添加各种过程或组件。另外，相对一些示例所描述的特征在其它例子中也可以进行组合。

[32] 如本文中使用的，术语“包括”及其变型表示开放的术语，含义是“包括但不限于”。术语“基于”表示“至少部分地基于”。术语“一个实施例”和“一实施例”表示“至少一个实施例”。术语“另一个实施例”表示“至少一个其他实施例”。术语“第一”、“第二”等可以指代不同的或相同的对象。下面可以包括其他的定义，无论是明确的还是隐含的。除非上下文中明确地指明，否则一个术语的定义在整个说明书中是一致的。

[33] 现在结合附图来描述本公开的识别风险商家的方法及装置。

[34] 图 1 是根据本公开的一个实施例的识别风险商家的方法的流程图。

[35] 如图 1 所示，在块 120，针对至少一个标记商家集群中的各个标记商家集群，基于待识别商家与该标记商家集群的用户关联度量值，确定待识别商家与该标记商家集群的相似度，各个标记商家集群中的各个标记商家被标记为相应风险类别的风险商家。风险商家是指有进行异常交易风险（例如洗钱犯罪交易行为等违法交易行为）的商家。

[36] 可以由人工对所收集的风险商家进行标注，并将被标注为相同类别的风险商家聚合为同一个标记商家集群。还可以利用经过训练的分类模型将已知的风险商家集群为各个标记商家集群。分类模型可以利用所收集的商家的用户数据来训练。各个标记商家集群对应的风险类别例如可以是毒品风险、走私风险、赌博风险、欺诈风险等类别。在获得各个标记商家集群之后，可以合并该标记商家集群中的各个标记商家的用户群体以作为该标记商家集群的用户群体。由此，能够将各个标记商家的用户特征综合为对应的标记商家集群的用户特征。各个商家的用户可以基于用户与商家之间的资金往来关系来确定。

[37] 合并同类别的风险商家以获得标记商家集群时，可以通过确定待识别商家与各个标记商家集群的相似度来识别待识别商家中的风险商家，而不需要一一计算待识别商家与每个已知的标记商家的相似度，从而大幅降低了运算量，提高了运算效率。此外，基于标记商家集群中的各个标记商家的用户群体合并而生成的用户群体来计算相似度，能够丰富各个标记商家集群的特征，从而能够提高识别风险商家的准确性。

[38] 在一个示例中，待识别商家与标记商家集群的用户关联度量值可以包括待识别商家与标记商家集群的用户关联系数和待识别商家与标记商家集群的相同用户数量。其中，用户关联系数可以是在待识别商家的用户中，与标记商家集群的相同用户所占的比例。即，用户关联系数可以利用如一数学式一来计算。

10 [39] 数学式一：

$$M(A, G) = \frac{|A \cap G|}{|A|}$$

[40] 在数学式一中，A 表示待识别商家的用户数量，G 表示标记商家集群的用户数量，M(A, G) 表示待识别商家与标记商家集群的用户关联系数， $A \cap G$  表示待识别商家与标记商家集群的相同用户数量。对于数学式一，可以定为当 A 的值为 0 时，M(A, G) 的值为 0。

15 [41] 用户关联系数还可以包括待识别商家与标记商家集群的用户特征相似度。在一个示例中，可以基于待识别商家的用户数据，利用经过训练的分类模型来确定待识别商家与各个标记商家集群的用户特征相似度，进而基于所确定的用户特征相似度来确定待识别商家与各个标记商家集群的相似度。在另一示例中，还可以利用特征提取模型（例如逻辑回归模型等）来从待识别商家和标记商家集群的用户数据（例如用户基本信息、用户行为数据等）中提取用户特征，进而基于所提取的待识别商家的用户特征和各个标记商家集群的用户特征，来确定待识别商家与各个标记商家集群的用户特征相似度。

[42] 在该示例中，用户关联系数可以看作待识别商家与标记商家集群之间的相对用户关联度量值，而相同用户数量可以看作待识别商家与标记商家集群的绝对用户关联度量值。因此，该示例可以基于待识别商家与标记商家集群的相对关联属性和绝对关联属性来确定相似度，从而使所计算的相似度更加准确。

25 [43] 在确定待识别商家与各个标记商家集群的相似度之后，在块 140，基于待识别商家与各个标记商家集群的相似度，识别待识别商家是否为风险商家。

[44] 在一个示例中，可以将待识别商家与各个标记商家集群的相似度中的最大值确定

为该待识别商家的风险系数。然后，可以基于待识别商家的风险系数和预定风险阈值，来确定该待识别商家是否为风险商家。例如，可以设置为在待识别商家的风险系数高于某一风险阈值时，确定其为风险商家。

[45] 在一个示例中，各个标记商家集群可以有至少一个代表标记商家。代表标记商家可以基于标记商家集群中的各个标记商家的用户数量来确定。例如，可以将各个标记商家集群中的标记商家按照用户数量进行排序。代表标记商家例如可以是排序结果中序数为中位数所对应的标记商家，还可以是提排序结果中居于中间位置的两个以上标记商家。此外，还可以是排序中间隔预定序数的两个以上标记商家。例如，如果某一标记商家集群中的标记商家按用户数量排序后的序数为 1 至 300，可以每隔 50 个序数从第 1 个至第 300 个标记商家中选取代表标记商家。当各个标记商家集群具有代表标记商家时，可以通过确定待识别商家与代表标记商家的相似度，来确定待识别商家与对应标记商家集群的相似度。

[46] 图 2 是根据本公开的一个实施例的识别风险商家的方法中的基于代表标记商家来确定相似度的示例的流程图。

[47] 如图 2 所示，在块 202，基于针对该标记商家集群的待识别商家的向量表示和该标记商家集群的至少一个代表标记商家的向量表示，确定待识别商家与该至少一个代表标记商家的相似度。针对某一标记商家集群的商家的向量表示是基于该商家与该标记商家集群的用户关联系数和相同用户数量而建立的。

[48] 待识别商家与代表标记商家的相似度可以用欧氏距离、曼哈顿距离、夹角余弦距离等中的任一者来表征。以下以利用欧氏距离的情形为例，来说明计算待识别商家与代表标记商家的相似度确定过程的示例。在下述示例中，待识别商家或代表标记商家与标记商家集群的用户关联系数为在该商家的用户中与该标记商家集群的相同用户所占的比例。

[49] 针对某一标记商家集群，待识别商家的向量表示可以是 $(M(A, G), |A \cap G|)$ ，该标记商家集群的代表标记商家的向量表示可以是 $(M(G_p, G), |G_p \cap G|)$ 。其中， $G_p$ 表示代表标记商家的用户数量。由于代表标记商家的用户全部来自于该标记商家集群，因而代表标记商家与该标记商家集群的用户关联系数 $M(G_p, G)$ 的值为 1，代表标记商家与该标记商家集群的相同用户数量 $|G_p \cap G|$ 为 $G_p$ 。待识别商家与代表标记商家之间的相似度可以利用如下数学式二来确定。

[50] 数学式二：

$$D(A, G_p) = \sqrt{(M(A, G) - 1)^2 + (|A \cap G| - |G_p|)^2}$$

[51] 在数学式二中,  $G_p$ 表示代表标记商家的用户数量,  $D(A, G_p)$ 为待识别商家与代表标记商家之间的欧氏距离。

[52] 在块 204, 基于待识别商家与至少一个代表标记商家的相似度, 确定待识别商家与该标记商家集群的相似度。

5 [53] 通常向量表示中的用户关联系数的取值较小 (例如, 取值范围为[0,1])。当待识别商家与某一标记商家集群的相同用户数量较大时, 由于在向量表示中相同用户数量维度的取值远大于用户关联系数, 因而用户关联系数维度可能会被相同用户数量维度覆盖。这会导致所确定的相似度不够准确。

[54] 因而, 在执行块 202 的操作之前, 可以将向量表示中的相同用户数量维度和用户  
10 关联系数维度处理为具有相同的取值范围。然后基于处理后的向量表示执行块 202 至块 204 的操作。

[55] 例如, 可以对向量表示中的相同用户数量维度进行归一化处理。可以利用如下数学式三对相同用户数量维度进行归一化处理。

[56] 数学式三:

$$f(x) = \frac{\max(x, |G_p|)}{|G_p|}$$

15 [57] 归一化处理后的待识别商家的向量表示为 $(M(A, G), \frac{\max(|A \cap G|, |G_p|)}{|G_p|})$ , 代表标记商家的向量表示为(1,1)。待识别商家与代表标记商家之间的相似度可以修改为如下数学式四。

[58] 数学式四:

$$D(A, G_p) = \sqrt{(M(A, G) - 1)^2 + (\frac{\max(|A \cap G|, |G_p|)}{|G_p|} - 1)^2}$$

[59] 数学式四中的 $D(A, G_p)$ 取值范围为 $[0, \sqrt{2}]$ 。还可以进一步基于数学式四对相似度的计算式进行改造。例如可以基于如下数学式五来计算最终的相似度。

20 [60] 数学式五:

$$S(A, G_p) = 1 - \frac{D(A, G_p)}{\sqrt{2}}$$

[61] 在数学式五中,  $S(A, G_p)$  为最终得出的待识别商家与代表标记商家的相似度, 其取值范围为[0,1]。

[62] 以上虽然示出了利用数学式三进行归一化处理的情形, 但应当理解的是, 还可以采用其它方式来对相同用户数量维度进行归一化处理。例如, 可以将各个向量表示中的相同用户数量的值除以相同用户数量维度的最大值, 以进行归一化处理。

[63] 在确定出待识别商家与代表标记商家之间的相似度之后, 可以确定待识别商家与标记商家集群之间的相似度。例如, 如果标记商家集群具有一个代表标记商家, 则可以将待识别商家与该代表标记商家的相似度确定为待识别商家与该标记商家集群的相似度。在一个示例中, 标记商家集群可以具有两个以上代表标记商家。此时, 可以对待识别商家与各个代表标记商家的相似度取平均值, 从而得出待识别商家与该标记商家集群的相似度。此外, 当代表标记商家是基于标记商家集群中的各个标记商家的用户数量排序而选取的多个代表标记商家时, 还可以基于各个代表标记商家的用户数量为其赋予不同的权重, 从而对待识别商家与各个代表标记商家的相似度取加权平均值或加权求和, 以得到待识别商家与该标记商家集群的相似度。

[64] 在确定出待识别商家与各个标记商家集群的相似度时, 可以针对各个标记商家集群, 将待识别商家与该标记商家集群的相似度中的大于预定阈值的相似度值最大的前预定个数的待识别商家, 识别为与该标记商家集群对应类别的风险商家。图 3 是根据本公开的一个实施例的识别风险商家的方法中的风险商家识别过程的一个示例的流程图。

[65] 如图 3 所示, 在块 302, 针对各个标记商家集群, 对至少两个待识别商家与该标记商家集群的相似度进行排序。

[66] 然后, 在块 304, 将排序结果中大于预定阈值的相似度值最大的前预定个数的待识别商家识别为该标记商家集群所对应的类别的风险商家。

[67] 此外, 还可以将待识别商家与各个标记商家集群的相似度中的最大值确定为该待识别商家的风险系数。然后将待识别商家的犯罪风险系数中的值最大的前预定个数的待识别商家确定为风险商家。或者, 将待识别商家的犯罪风险系数中的大于预定阈值的值最大的前预定个数的待识别商家确定为风险商家。

[68] 在确定出风险商家之后, 还可以进一步对所确定出的风险商家进行验证征处理, 以排除其中风险较低的商家。例如, 可以将所确定出的风险商家的数据发送至反风险监控团队, 由团队中的专业人员进一步对这些商家进行分析, 以最终确定风险商家。

[69] 图 4 是根据本公开的一个实施例的风险商家识别装置 400 的结构框图。如图 4 所示，风险商家识别装置 400 包括相似度确定单元 410 和风险商家识别单元 420。

[70] 相似度确定单元 410 被配置为针对至少一个标记商家集群中的各个标记商家集群，基于待识别商家与该标记商家集群的用户关联度量值，确定待识别商家与该标记商家集群的相似度，各个标记商家集群中的各个标记商家被标记为相应风险类别的风险商家。风险商家识别单元 420 被配置为基于待识别商家与各个标记商家集群的相似度，识别至少一个待识别商家中的风险商家。

[71] 在一个示例中，风险商家识别单元 420 可以被配置为针对各个标记商家集群，将至少两个待识别商家与该标记商家集群的相似度中的大于预定阈值的相似度值最大的前预定个数的待识别商家，识别为与该标记商家集群对应类别的风险商家。

[72] 在一个示例中，待识别商家与标记商家集群的用户关联度量值可以包括待识别商家与标记商家集群的用户关联系数和待识别商家与标记商家集群的相同用户数量。其中，用户关联系数可以为在待识别商家的用户中，与标记商家集群的相同用户所占的比例。用户关联系数还可以是待识别商家的用户与标记商家集群的用户的特征相似度。

[73] 图 5 是图 4 所示的风险商家识别装置 400 中的相似度确定单元 410 的一个示例的结构框图。在该示例中，各个标记商家集群可以具有至少一个代表标记商家。如图 5 所示，相似度确定单元 410 可以包括归一化处理模块 411、第一相似度确定模块 412 和第二相似度确定模块 413。其中，在另一示例中，可以不包括归一化处理模块 411。以下为了说明上的方便性，首先对第一相似度确定模块 412 和第二相似度确定模块 413 进行说明。

[74] 第一相似度确定模块 412 被配置为基于针对该标记商家集群的待识别商家的向量表示和该标记商家集群的至少一个代表标记商家的向量表示，确定待识别商家与该至少一个代表标记商家的相似度。第二相似度确定模块 413 被配置为基于待识别商家与该至少一个代表标记商家的相似度，确定待识别商家与该标记商家集群的相似度。其中，针对该标记商家集群的商家的向量表示是基于商家与该标记商家集群的用户关联系数和所述商家与该标记商家集群的相同用户数量的向量表示。

[75] 为了避免向量表示中不同维度的取值范围差异过大而导致所确定的相似度不准确，可以利用归一化处理模块 411 在基于至少一个待识别商家中的待识别商家与该标记商家集群的用户关联度量值，确定待识别商家与该标记商家集群的相似度之前，将向量表示中

的相同用户数量维度和用户关联系数维度处理为具有相同的取值范围。关于归一化处理的示例已在上述内容中进行了说明，此处不再赘述。

[76] 以上参照图 1-5 对识别风险商家的方法和装置进行了说明。需要说明的是，在以上对识别风险商家的方法的说明中提及的细节同样适用于识别风险商家的装置。本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见。

[77] 本公开的识别风险商家的装置可以采用硬件实现，也可以采用软件或者硬件和软件的组合来实现。以软件实现为例，作为一个逻辑意义上的装置，是通过其所在设备的处理器将非易失性存储器中对应的计算机程序指令读取到内存中运行形成的。在本公开中，识别终端设备上显示的应用程序控件的装置例如可以利用计算设备实现。

[78] 图 6 是根据本公开的另一实施例的识别风险商家的方法的计算设备 600 的结构框图。如图 6 所示，计算设备 600 可以包括至少一个处理器 610、存储器 620、内存 630、通信接口 640 以及内部总线 650，该至少一个处理器 610 执行在计算机可读存储介质（即，存储器 620）中存储或编码的至少一个计算机可读指令（即，上述以软件形式实现的元素）。

[79] 在一个实施例中，在存储器 620 中存储计算机可执行指令，其当执行时使得至少一个处理器 610：针对至少一个标记商家集群中的各个标记商家集群，基于至少一个待识别商家中的待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度，所述各个标记商家集群中的各个标记商家被标记为相应风险类别的风险商家；以及基于所述待识别商家与所述各个标记商家集群的相似度，识别所述至少一个待识别商家中的风险商家。

[80] 应该理解，在存储器 620 中存储的计算机可执行指令当执行时使得至少一个处理器 610 进行本公开的各个实施例中以上结合图 1-5 描述的各种操作和功能。

[81] 根据一个实施例，提供了一种例如非暂时性机器可读介质的程序产品。非暂时性机器可读介质可以具有指令（即，上述以软件形式实现的元素），该指令当被机器执行时，使得机器执行本公开的各个实施例中以上结合图 1-5 描述的各种操作和功能。

[82] 具体地，可以提供配有可读存储介质的系统或者装置，在该可读存储介质上存储着实现上述实施例中任一实施例的功能的软件程序代码，且使该系统或者装置的计算机或处理器读出并执行存储在该可读存储介质中的指令。

[83] 在这种情况下，从可读介质读取的程序代码本身可实现上述实施例中任何一项实

施例的功能，因此机器可读代码和存储机器可读代码的可读存储介质构成了本发明的一部分。

[84] 可读存储介质的实施例包括软盘、硬盘、磁光盘、光盘(如 CD-ROM、CD-R、CD-RW、DVD-ROM、DVD-RAM、DVD-RW、DVD-RW)、磁带、非易失性存储卡和 ROM。

5 可选择地，可以由通信网络从服务器计算机上或云上下载程序代码。

[85] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是  
10 可以的或者可能是有利的。

[86] 在整个本说明书中使用的术语“示例性”意味着“用作示例、实例或例示”，并不意味着比其它实施例“优选”或“具有优势”。出于提供对所描述技术的理解的目的，具体实施方式包括具体细节。然而，可以在没有这些具体细节的情况下实施这些技术。在一些实例中，为了避免对所描述的实施例的概念造成难以理解，公知的结构和装置以  
15 框图形式示出。

[87] 以上结合附图详细描述了本公开的实施例的可选实施方式，但是，本公开的实施例并不限于上述实施方式中的具体细节，在本公开的实施例的技术构思范围内，可以对本公开的实施例的技术方案进行多种简单变型，这些简单变型均属于本公开的实施例的保护范围。

20 [88] 本公开内容的上述描述被提供来使得本领域任何普通技术人员能够实现或者使用本公开内容。对于本领域普通技术人员来说，对本公开内容进行的各种修改是显而易见的，并且，也可以在不脱离本公开内容的保护范围的情况下，将本文所定义的一般性原理应用于其它变型。因此，本公开内容并不限于本文所描述的示例和设计，而是与符合本文公开的原理和新颖性特征的最广范围相一致。

### 权利要求书

1、一种识别风险商家的方法，包括：

针对至少一个标记商家集群中的各个标记商家集群，基于待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度，所述各个标记商家集群中的各个标记商家被标记为相应风险类别的风险商家；以及

基于所述待识别商家与所述各个标记商家集群的相似度，识别所述待识别商家是否为风险商家。

2、如权利要求 1 所述的方法，其中，所述待识别商家与所述标记商家集群的用户关联度量值包括所述待识别商家与所述标记商家集群的用户关联系数和所述待识别商家与所述标记商家集群的相同用户数量，

其中，所述用户关联系数包括：

在所述待识别商家的用户中，与所述标记商家集群的相同用户所占的比例；和/或所述待识别商家与所述标记商家集群的用户特征相似度。

3、如权利要求 2 所述的方法，其中，所述各个标记商家集群具有至少一个代表标记商家，基于待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度包括：

基于针对该标记商家集群的所述待识别商家的向量表示和该标记商家集群的至少一个代表标记商家的向量表示，确定所述待识别商家与该至少一个代表标记商家的相似度；以及

基于所述待识别商家与该至少一个代表标记商家的相似度，确定所述待识别商家与该标记商家集群的相似度，

其中，针对该标记商家集群的商家的向量表示是基于所述商家与该标记商家集群的用户关联系数和所述商家与该标记商家集群的相同用户数量的向量表示。

4、如权利要求 3 所述的方法，其中，在基于所述待识别商家的向量表示和该标记商家集群的至少一个代表标记商家的向量表示，确定所述待识别商家与该至少一个代表标记商家的相似度之前，基于待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度还包括：

将所述向量表示中的相同用户数量维度和用户关联系数维度处理为具有相同的取值范围。

5、如权利要求 3 所述的方法，其中，所述各个标记商家集群的至少一个代表标记商家是基于该标记商家集群中的各个标记商家的用户数量确定的。

6、如权利要求 3 所述的方法，其中，所述待识别商家与所述各个标记商家集群的至少一个代表标记商家的相似度用以下中的任一者来表征：欧氏距离、曼哈顿距离和夹角余弦距离。

5 7、如权利要求 1 所述的方法，其中，基于所述待识别商家与所述各个标记商家集群的相似度，识别所述待识别商家中的风险商家包括：

针对所述各个标记商家集群，将所述至少两个待识别商家与该标记商家集群的相似度中的大于预定阈值的相似度值最大的前预定个数的待识别商家，识别为与该标记商家集群对应类别的风险商家。

8、一种识别风险商家的装置，包括：

10 相似度确定单元，被配置为针对至少一个标记商家集群中的各个标记商家集群，基于待识别商家中的待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度，所述各个标记商家集群中的各个标记商家被标记为相应风险类别的风险商家；以及

15 风险商家识别单元，被配置为基于所述待识别商家与所述各个标记商家集群的相似度，识别所述待识别商家是否为风险商家。

9、如权利要求 8 所述的装置，其中，所述待识别商家与所述标记商家集群的用户关联度量值包括所述待识别商家与所述标记商家集群的用户关联系数和所述待识别商家与所述标记商家集群的相同用户数量，

其中，所述用户关联系数包括：

20 在所述待识别商家的用户中，与所述标记商家集群的相同用户所占的比例；和/或所述待识别商家与所述标记商家集群的用户特征相似度。

10、如权利要求 8 或 9 所述的装置，其中，所述各个标记商家集群具有至少一个代表标记商家，所述相似度确定单元包括：

25 第一相似度确定模块，被配置为基于针对该标记商家集群的所述待识别商家的向量表示和该标记商家集群的至少一个代表标记商家的向量表示，确定所述待识别商家与该至少一个代表标记商家的相似度；以及

第二相似度确定模块，被配置为基于所述待识别商家与该至少一个代表标记商家的相似度，确定所述待识别商家与该标记商家集群的相似度，

30 其中，针对该标记商家集群的商家的向量表示是基于所述商家与该标记商家集群的用户关联系数和所述商家与该标记商家集群的相同用户数量的向量表示。

11、如权利要求 10 所述的装置，其中，所述相似度确定单元还包括：

归一化处理模块，被配置为在基于待识别商家与该标记商家集群的用户关联度量值，确定所述待识别商家与该标记商家集群的相似度之前，将所述向量表示中的相同用户数量维度和用户关联系数维度处理为具有相同的取值范围。

12、如权利要求 8 所述的装置，其中，所述风险商家识别单元被配置为：

- 5       针对所述各个标记商家集群，将至少两个待识别商家与该标记商家集群的相似度中的大于预定阈值的相似度值最大的前预定个数的待识别商家，识别为与该标记商家集群对应类别的风险商家。

13、一种计算设备，包括：

至少一个处理器；以及

- 10       存储器，所述存储器存储指令，当所述指令被所述至少一个处理器执行时，使得所述至少一个处理器执行如权利要求 1 到 7 中任一所述的方法。

14、一种非暂时性机器可读存储介质，其存储有可执行指令，所述指令当被执行时使得所述机器执行如权利要求 1 到 7 中任一所述的方法。

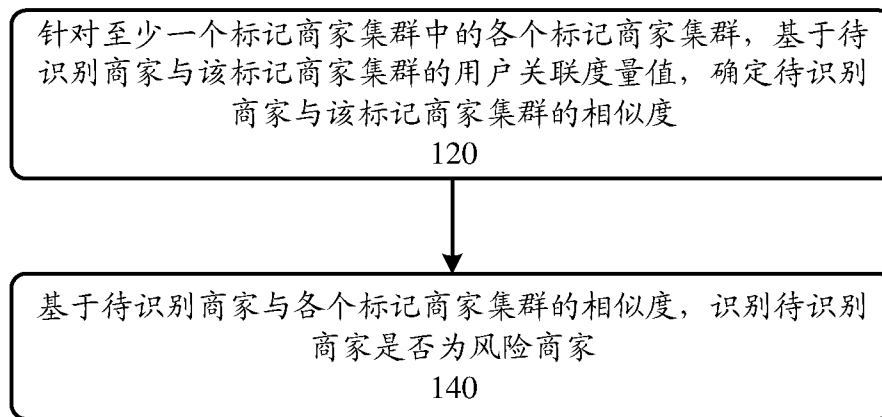


图 1

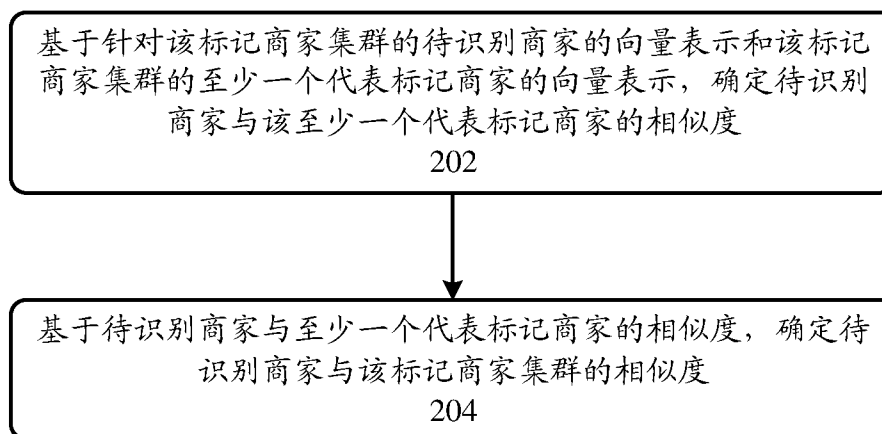


图 2

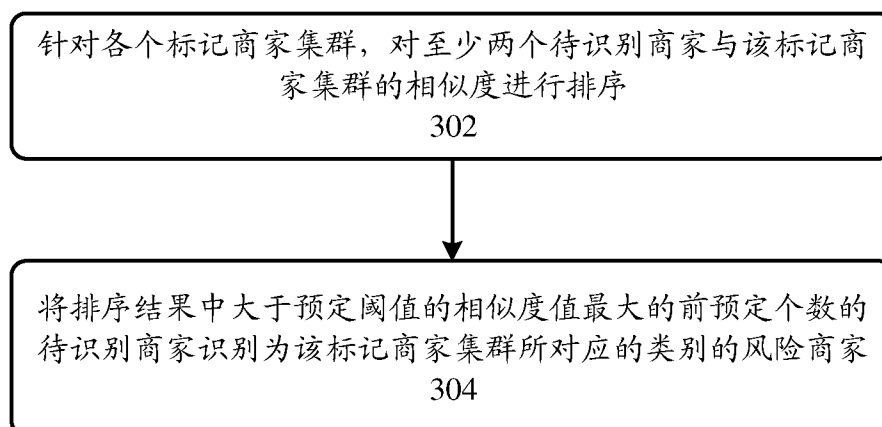


图 3



图 4

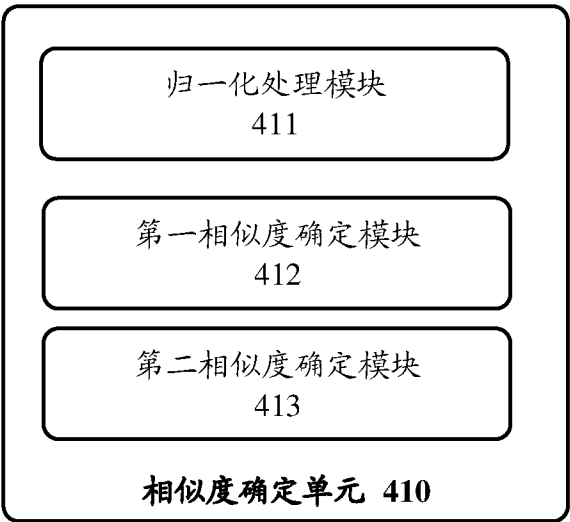


图 5

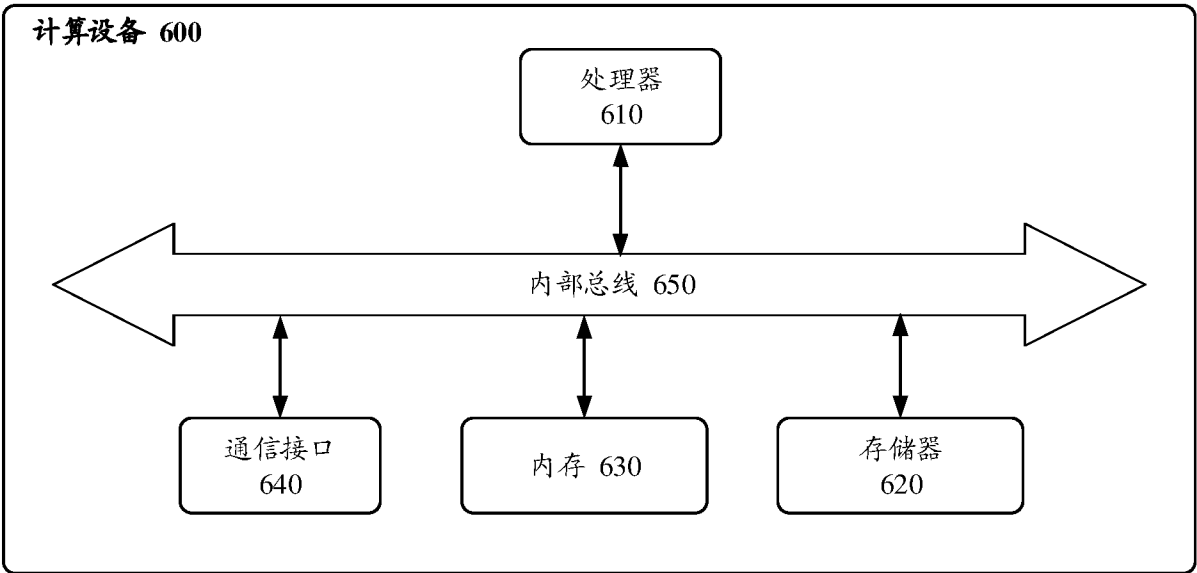


图 6

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/070672

**A. CLASSIFICATION OF SUBJECT MATTER**

G06Q 10/06(2012.01)i; G06Q 30/00(2012.01)i; G06Q 30/06(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)

WPI, EPODOC, CNPAT, CNKI, IEEE: 识别, 风险, 商家, 集群, 关联, 相似度, 值, 向量, 模型, identify, risk, merchant, cluster, correlation, degree, value, similarity, vector, model

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                  | Relevant to claim No. |
|-----------|-----------------------------------------------------------------------------------------------------|-----------------------|
| PX        | CN 110033170 A (ALIBABA GROUP HOLDING LIMITED) 19 July 2019 (2019-07-19)<br>claims 1-14             | 1-14                  |
| X         | CN 109461068 A (SHENZHEN ONECONNECT TECHNOLOGY CO., LTD.) 12 March 2019 (2019-03-12)<br>claims 1-10 | 1-14                  |
| A         | CN 109360089 A (SICHUAN UNIVERSITY) 19 February 2019 (2019-02-19)<br>entire document                | 1-14                  |
| A         | CN 105159879 A (BEIJING INSTITUTE OF TECHNOLOGY) 16 December 2015 (2015-12-16)<br>entire document   | 1-14                  |
| A         | US 2018025428 A1 (XEROX CORPORATION) 25 January 2018 (2018-01-25)<br>entire document                | 1-14                  |



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

18 March 2020

Date of mailing of the international search report

09 April 2020

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/CN2020/070672**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|--------------------------------------|
| CN                                        | 110033170  | A  | 19 July 2019                         | None                    |                                      |
| CN                                        | 109461068  | A  | 12 March 2019                        | None                    |                                      |
| CN                                        | 109360089  | A  | 19 February 2019                     | None                    |                                      |
| CN                                        | 105159879  | A  | 16 December 2015                     | None                    |                                      |
| US                                        | 2018025428 | A1 | 25 January 2018                      | None                    |                                      |

## 国际检索报告

国际申请号

PCT/CN2020/070672

## A. 主题的分类

G06Q 10/06(2012.01)i; G06Q 30/00(2012.01)i; G06Q 30/06(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))

WPI, EPDOC, CNPAT, CNKI, IEEE: 识别, 风险, 商家, 集群, 关联, 相似度, 值, 向量, 模型, identify, risk, merchant, cluster, correlation, degree, value, similarity, vector, model

## 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 110033170 A (阿里巴巴集团控股有限公司) 2019年 7月 19日 (2019 - 07 - 19)<br>权利要求1-14  | 1-14    |
| X    | CN 109461068 A (深圳壹账通智能科技有限公司) 2019年 3月 12日 (2019 - 03 - 12)<br>权利要求1-10 | 1-14    |
| A    | CN 109360089 A (四川大学) 2019年 2月 19日 (2019 - 02 - 19)<br>全文                | 1-14    |
| A    | CN 105159879 A (北京理工大学) 2015年 12月 16日 (2015 - 12 - 16)<br>全文             | 1-14    |
| A    | US 2018025428 A1 (XEROX CORPORATION) 2018年 1月 25日 (2018 - 01 - 25)<br>全文 | 1-14    |

☐ 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

“&amp;” 同族专利的文件

国际检索实际完成的日期

2020年 3月 18日

国际检索报告邮寄日期

2020年 4月 9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

张涛

电话号码 86-(10)-53961356

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2020/070672

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|----------------|
| CN          | 110033170  | A  | 2019年 7月 19日   | 无    |                |
| CN          | 109461068  | A  | 2019年 3月 12日   | 无    |                |
| CN          | 109360089  | A  | 2019年 2月 19日   | 无    |                |
| CN          | 105159879  | A  | 2015年 12月 16日  | 无    |                |
| US          | 2018025428 | A1 | 2018年 1月 25日   | 无    |                |