(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 16 日 (16.07.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/143377 A1

(51) 国际专利分类号:
G06Q 20/42 (2012.01)

(21) 国际申请号: PCT/CN2019/124199

(22) 国际申请日: 2019 年 12 月 10 日 (10.12.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910014638.3 2019年1月8日 (08.01.2019) CN

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

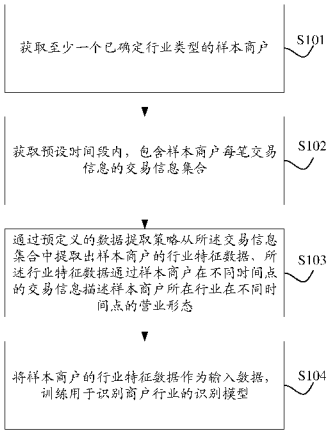
(72) 发明人: 王立 (**WANG, Li**); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

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

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, 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,

(54) **Title:** INDUSTRY RECOGNITION MODEL DETERMINATION METHOD AND APPARATUS

(54) 发明名称: 一种行业识别模型确定方法和装置



S101 Acquire at least one sample merchant, the industry type of which has been determined

S102 Acquire a transaction information set, containing information of each transaction of the sample merchant, within a pre-set time period

S103 Extract, by means of a pre-defined data extraction policy and from the transaction information set, industry feature data of the sample merchant, wherein the industry feature data describes, by means of transaction information of the sample merchant at different time points, business forms at the different time points of an industry to which the sample merchant belongs

S104 Take the industry feature data of the sample merchant as input data to train a recognition model for recognizing the industry of a merchant

图 1

(57) **Abstract:** Disclosed are an industry recognition model determination method and apparatus. The method comprises: firstly, acquiring a sample merchant, the industry type of which has been determined (S101); extracting, by means of a pre-defined data extraction policy and from a transaction information set, industry feature data for describing a business form of an industry to which the sample merchant belongs (S103); and training, by means of the industry feature data, a recognition model for recognizing the industry of a merchant (S104). Therefore, an industry to which a merchant belongs can be determined by using a recognition model and by means of recognizing industry features, and compared with summarizing word-bank-based screening policies, a merchant of a fake industry type cannot avoid the recognition of industry features in the method.

WO 2020/143377 A1

[见续页]

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))。

(57) 摘要：一种行业识别模型确定方法和装置。先获取已确定行业类型的样本商户（S101），通过预定义的数据提取策略从所述交易信息集合中提取出描述样本商户所在行业的营业形态的行业特征数据（S103），通过行业特征数据训练用于识别商户行业的识别模型（S104）。进而可利用识别模型通过识别行业特征确定商户的所在行业，和汇总基于词库的筛选策略相比，虚假行业类型商户难以规避上述方法对于行业特征的识别。

一种行业识别模型确定方法和装置

技术领域

[01] 本说明书涉及互联网技术领域，尤其涉及一种行业识别模型确定方法和装置。

背景技术

[02] 随着电子支付的发展，很多商户选择接入支付平台，交易时买家可通过扫码等方式进行电子支付。而对于不同行业的商家，支付平台可能会采用不同的接入规则。举例说明：商户是支付宝会员体系中重要的一部分，为了支持餐饮商户和支付宝的合作，支付宝规定，所有餐饮商户接入支付宝收银即可享受支付宝收款零费率。而与此同时，很多非餐饮商家为了利益，可能会伪造商户门头照、商户名称等申请参与零费率活动。针对这类风险，虽然已经有人工客服的排查，还是会漏过一批非餐饮商户。为了避免不符合条件的商户伪造信息。需要一种商户行业识别方法。

[03] 对于商户的行业识别只有基于关键词的识别策略，即维护不同行业对应的相关词库，对提交审核商户的商户名称，商品名称进行排查，但是这种方法比较容易被对方规避，同时还需要定期的对词库进行扩充。

发明内容

[04] 针对上述技术问题，本说明书实施例提供一种行业识别模型确定方法和装置，技术方案如下：

[05] 根据本说明书实施例的第一方面，提供一种行业识别模型确定方法，该方法包括：

[06] 获取至少一个已确定行业类型的样本商户；

[07] 获取预设时间段内，包含样本商户每笔交易信息的交易信息集合；

[08] 通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态；

[09] 将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型。

[10] 根据本说明书实施例的第二方面，提供一种基于行业识别模型的商户行业确定方

法，该方法包括:

[11] 获取预设时间段内，包含待检验商户每笔交易信息的交易信息集合;

[12] 通过预定义的数据提取策略从所述交易信息集合中提取出待检验商户的行业特征数据;

[13] 将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业。

[14] 根据本说明书实施例的第三方面，提供一种行业识别模型确定装置，该装置包括:

[15] 样本确定模块：用于获取至少一个已确定行业类型的样本商户;

[16] 信息获取模块：用于获取预设时间段内，包含样本商户每笔交易信息的交易信息集合;

[17] 数据提取模块：用于通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态;

[18] 模型训练模块：用于将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型。

[19] 根据本说明书实施例的第四方面，提供一种基于行业识别模型的商户行业确定装置，该装置包括:

[20] 信息获取模块：用于获取预设时间段内，包含待检验商户每笔交易信息的交易信息集合;

[21] 数据提取模块：用于通过预定义的数据提取策略从所述交易信息集合中提取出待检验商户的行业特征数据;

[22] 行业识别模块：用于将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业。

[23] 根据本说明书实施例的第五方面，提供一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其中，所述处理器执行所述程序时实现一种行业识别模型确定方法，该方法包括:

[24] 获取至少一个已确定行业类型的样本商户;

[25] 获取预设时间段内，包含样本商户每笔交易信息的交易信息集合；

[26] 通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态；

[27] 将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型。

[28] 根据本说明书实施例的第六方面，提供一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其中，所述处理器执行所述程序时实现一种基于行业识别模型的商户行业确定方法，该方法包括：

[29] 获取预设时间段内，包含待检验商户每笔交易信息的交易信息集合；

[30] 通过预定义的数据提取策略从所述交易信息集合中提取出待检验商户的行业特征数据；

[31] 将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业。

[32] 本说明书实施例所提供的技术方案，提供了一种行业识别模型确定方法，先获取已确定行业类型的样本商户，通过预定义的数据提取策略从所述交易信息集合中提取出描述样本商户所在行业的营业形态的行业特征数据，通过行业特征数据训练用于识别商户行业的识别模型。进而可利用识别模型通过识别行业特征确定商户的所在行业，和现有技术汇总基于词库的筛选策略相比，虚假行业类型商户难以规避对于本申请对于行业特征的识别。

[33] 应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，并不能限制本说明书实施例。

[34] 此外，本说明书实施例中的任一实施例并不需要达到上述的全部效果。

附图说明

[35] 为了更清楚地说明本说明书实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本说明书实施例中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据这些附图获得其他的附图。

[36] 图 1 是本说明书一示例性实施例示出的行业识别模型确定方法的一种示意图；

[37] 图 2 是本说明书一示例性实施例示出的交易数据的一种示意图;

[38] 图 3 是本说明书一示例性实施例示出的基于行业识别模型的商户行业识别方法的一种流程图;

[39] 图 4 是本说明书一示例性实施例示出的餐饮行业识别模型的确定方法的一种流程图;

[40] 图 5 是本说明书一示例性实施例示出的一种交易数据的资金热力示意图;

[41] 图 6 是本说明书一示例性实施例示出的模型网络构成示意图;

[42] 图 7 是本说明书一示例性实施例示出的行业识别模型确定装置的一种示意图;

[43] 图 8 是本说明书一示例性实施例示出的基于行业识别模型的商户行业识别装置的一种示意图;

[44] 图 9 是本说明书一示例性实施例示出的一种计算机设备的结构示意图。

具体实施方式

[45] 这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本说明书相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本说明书的一些方面相一致的装置和方法的例子。

[46] 在本说明书使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书。在本说明书和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[47] 应当理解，尽管在本说明书可能采用术语第一、第二、第三等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[48] 随着电子支付的发展，很多商户选择接入支付平台，交易时买家可通过扫码等方式进行电子支付。而对于不同行业的商家，支付平台可能会采用不同的接入规则。举例

说明：商户是支付宝会员体系中重要的一部分，为了支持餐饮商户和支付宝的合作，支付宝规定，所有餐饮商户接入支付宝收银即可享受支付宝收款零费率。而与此同时，很多非餐饮商家为了利益，可能会伪造商户门头照、商户名称等申请参与零费率活动。针对这类风险，虽然已经有人工客服的排查，还是会漏过一批非餐饮商户。为了避免不符合条件的商户伪造信息。需要一种商户行业识别方法。

[49] 对于商户的行业识别只有基于关键词的识别策略，即维护不同行业对应的相关词库，对提交审核商户的商户名称，商品名称进行排查，但是这种方法比较容易被对方规避，同时还需要定期的对词库进行扩充。

[50] 针对以上问题，本说明书实施例提供一种行业识别模型确定方法，以及一种用于执行该方法的行业识别模型确定装置。通过训练能够进行行业识别的模型来判断商户的行业。

[51] 下面对本实施例涉及的行业识别模型确定方法进行详细说明，参见图 1 所示，该方法可以包括以下步骤：

[52] S101，获取至少一个已确定行业类型的样本商户；

[53] 获取已确定行业类型的样本商户的方式有多种方式，可以根据应用场景自行设定。举例说明：若应用场景为支付宝电子支付平台接入的商户，商户行业类型为餐饮行业，则可从支付宝平台各个行业的口碑商户中获取对应类型的样本商户。

[54] 进一步地，可设定一种需要让模型重点识别的目标行业类型，如餐饮行业类型，则在获取样本商户时，可采集比例较高的行业类型为餐饮行业的商户作为正样本，采集比例较低的行业类型为非餐饮行业的商户作为负样本。以重点训练模型识别商户是否为餐饮行业的能力。

[55] S102，获取预设时间段内，包含样本商户每笔交易信息的交易信息集合；

[56] S103，通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态；

[57] 其中，数据提取策略需要预先根据行业类型进行规定，不同的行业类型对应着不同的数据提取策略。通过将样本商户的交易信息进行整合提取，获取不同行业对应的不同形式的行业特征数据。

[58] 其中，行业特征数据可包括样本商户在预定时间段内，每个单位时间的至少一种类别的交易信息。参见图 2，为其中一种提取出的行业特征数据的示意图，即获取不同商户单日内每个小时段的其中一种交易数据（如营业额数据），将提取结果作为营业特征数据。

[59] 举例说明，可为某一类型行业设定 5 种类别的交易数据作为营业特征数据，这 5 种类别的交易数据分别为该样本商户单个小时内收款的总金额、单个小时的交易次数、单个小时均笔交易金额、单个小时交易金额方差以及单个小时的交易人数，这五种类别的数据用于刻画餐饮商户的营业形态。当行业类型为餐饮行业时，通过上述行业类型数据可以看出，大部分的餐饮商户交易集中在早上晚三餐，且交易金额都比较少。提取上述交易数据作为行业特征数据能够看出餐饮行业的营业形态。

[60] 进一步地，通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据后，还可将同一样本商户任一类别的交易信息表达为交易图像，所述交易图像的横坐标用于表示交易的发生日期，所述交易图像的纵坐标用于表示每个发生日期内的不同发生时间点，所述交易图像中像素点的颜色用于表示所述类别的交易信息的数值大小。

[61] S104，将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型。

[62] 具体地，将样本商户的已确定的行业类型和对应的行业特征数据作为一组训练数据，训练用于识别商户行业的识别模型。

[63] 如上文所述，将同一样本商户任一类别的交易信息表达为交易图像后，还可将表达同一样本商户不同类型交易特征的交易图像数据作为 CNN 神经网络不同通道的输入数据，训练用于识别商户行业的识别模型。

[64] 其中，行业识别模型的训练方式具体过程可为：

[65] 将表达不同类型交易特征的交易图像数据分别处理为对应的一维向量序列；

[66] 将所述不同一维向量序列分别进行数据标准化处理，并将数据标准化处理后的不同一维向量序列整合重构为一个三维向量序列；

[67] 将所述三维向量序列作为不同通道的输入数据，输入 CNN 神经网络的卷积池化层，将反复卷积池化后的输出结果作为特征向量；

[68] 将特征向量展开为一维向量输入CNN神经网络全连接层，以完成行业识别模型对于所述样本商户的行业特征的学习。

[69] 基于上述训练出的行业识别模型，可进一步针对未知类型的商户进行行业识别方法，参见图3，包括以下步骤:

[70] S301，获取预设时间段内，包含待检验商户每笔交易信息的交易信息集合；具体实施方式参照步骤S102;

[71] S302，通过预定义的数据提取策略从所述交易信息集合中提取出待检验商户的行业特征数据；具体实施方式参照步骤S103;

[72] S303，将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业。

[73] 具体地，获取未知行业类型的商户的行业特征数据，将该行业特征数据输入已针对特定行业类型训练完成的行业识别模型，进而可判断出该未知商户是否属于该特定行业类型。

[74] 针对不同类型的商户行业，均可根据该行业的行业特征训练对应的行业识别模型，下面以餐饮行业类型举例，进行更具体的模型训练方法介绍，参考图4，包括如下步骤:

[75] S401，获取至少一个餐饮行业的样本商户；

[76] S402，将每个样本商户预定天数内，每小时内预定类型的交易数据合并成为天数*24*类型数量的交易图像;

[77] 具体地，将每个样本商户最近七天，每小时的5种交易数据合并成为7*24*5的资金热力图。

[78] 其中，这5种类别的交易数据分别为该样本商户单个小时内收款的总金额、单个小时的交易次数、单个小时均笔交易金额、单个小时交易金额方差以及单个小时的交易人数，这五种类别的数据用于刻画餐饮商户的营业形态。

[79] 将上述交易转化为交易图像数据，参见图5，为六个餐饮商户7天内交易金额特征对应的热力图，其中横纵坐标为交易发生的日期和具体的时间节点，图像中像素点的颜色深浅代表了交易金额的数值，具体对应情况见右侧的色条。

[80] 可以看出，大部分的餐饮商户交易集中在早上晚三餐，且交易金额都比较少。此即为餐饮商户的营业形态。

[81] 需要注意，7 天和 24 小时都是其中一种可以选择的时间划分方式，具体实施时可根据实际情况选择其他的时间划分方式。收款的总金额、交易次数、均笔交易金额、交易金额方差以及交易人数这 5 种类别的交易数据也可以根据实际情况增加，减少或变动为其他类型的数据。上述时间划分和类型选择方式均不对本实施例造成限制。

[82] S403，将 5 种不同资金热力图数据作为五个通道的输入数据，训练用于识别商户行业的识别模型。

[83] 具体地，该训练模型可仿照 VGG-Net 网络结构，采用[卷积-激活-池化]*3- Dropout-全连接-Dropout-分类的结构。

[84] 其中，在本实施例中，数据流的整体流程如下：

[85] a) 将餐饮商户七天内每小时内收款的总金额、每小时内的交易次数、每小时内均笔交易金额、每小时内交易金额方差以及每小时的内的交易人数作为模型的五个通道输入，五条序列都为 7*24 长的一维向量。

[86] b) 将五条序列信息处理成[-1,1]区间内数值，并重构成为 7*24*5 的三维向量。其中，金额序列中大于 1000 的金额设置为 1000，然后将(金额-500)/1000 作为输入序列。

[87] c) 参见图 6，采用 CNN 神经网络对输入序列反复卷积池化，得到包括 5 种交易数据类型的资金序列特征向量。

[88] d) 将资金序列特征向量展开成为一维向量输入全连接神经网络进行分类训练。

[89] 通过上述流程训练出针对餐饮行业的识别模型后，在后续应用阶段，仍使用上述 a，b 步骤清洗出 3 维交易数据，将清洗出的数据输入训练好的模型产出评价分数，并将评价分数不符合标准的高危商户返回人工客服进行复审。

[90] 可以看出，该基于交易热力图的深度神经网络行业识别模型，可以对符合餐饮交易形态的商户进行排查，即选择出交易集中在早中晚三餐、交易金额较小的商户。和基于词库的商户名称筛选策略相比，虚假餐饮商户难移规避策略，同时模型效果不会随着时间衰退，不用定期重训模型。

[91] 相应于上述方法实施例，本说明书实施例还提供一种行业识别模型确定方法装置，参见图 7 所示，所述装置可以包括：样本确定模块 710，信息获取模块 720，数据提取模块 730 和模型训练模块 740。

[92] 样本确定模块 710：用于获取至少一个已确定行业类型的样本商户；

[93] 信息获取模块 720：用于获取预设时间段内，包含样本商户每笔交易信息的交易信息集合；

[94] 数据提取模块 730：用于通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态；

[95] 模型训练模块 740：用于将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型。

[96] 相应于上述方法实施例，本说明书实施例还提供一种基于权利要求 10 行业识别模型的商户行业确定装置，参见图 8 所示，所述装置包括：信息获取模块 810，数据提取模块 820 和行业识别模块 830。

[97] 信息获取模块 810：用于获取预设时间段内，包含待检验商户每笔交易信息的交易信息集合；

[98] 数据提取模块 820：用于通过预定义的数据提取策略从所述交易信息集合中提取出待检验商户的行业特征数据；

[99] 行业识别模块 830：用于将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业。

[100] 本说明书实施例还提供一种计算机设备，其至少包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其中，处理器执行所述程序时实现前述行业识别模型确定方法，所述方法至少包括：

[101] 获取至少一个已确定行业类型的样本商户；

[102] 获取预设时间段内，包含样本商户每笔交易信息的交易信息集合；

[103] 通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态；

[104] 将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型。

[105] 本说明书实施例还提供一种计算机设备，其至少包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其中，处理器执行所述程序时实现一种基于前述行业识别模型的商户行业确定方法，所述方法包括：

[106] 获取预设时间段内，包含待检验商户每笔交易信息的交易信息集合；

[107] 通过预定义的数据提取策略从所述交易信息集合中提取出待检验商户的行业特征数据；

[108] 将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业。

[109] 图 9 示出了本说明书实施例所提供的一种更为具体的计算设备硬件结构示意图，该设备可以包括：处理器 1110、存储器 1120、输入/输出接口 1130、通信接口 1140 和总线 1150。其中处理器 1110、存储器 1120、输入/输出接口 1130 和通信接口 1140 通过总线 1150 实现彼此之间在设备内部的通信连接。

[110] 处理器 1110 可以采用通用的 CPU（Central Processing Unit，中央处理器）、微处理器、应用专用集成电路（Application Specific Integrated Circuit，ASIC）、或者一个或多个集成电路等方式实现，用于执行相关程序，以实现本说明书实施例所提供的技术方案。

[111] 存储器 1120 可以采用 ROM（Read Only Memory，只读存储器）、RAM（Random Access Memory，随机存取存储器）、静态存储设备，动态存储设备等形式实现。存储器 1120 可以存储操作系统和其他应用程序，在通过软件或者固件来实现本说明书实施例所提供的技术方案时，相关的程序代码保存在存储器 1120 中，并由处理器 1110 来调用执行。

[112] 输入/输出接口 1130 用于连接输入/输出模块，以实现信息输入及输出。输入输出/模块可以作为组件配置在设备中（图中未示出），也可以外接于设备以提供相应功能。其中输入设备可以包括键盘、鼠标、触摸屏、麦克风、各类传感器等，输出设备可以包括显示器、扬声器、振动器、指示灯等。

[113] 通信接口 1140 用于连接通信模块（图中未示出），以实现本设备与其他设备的通信交互。其中通信模块可以通过有线方式（例如 USB、网线等）实现通信，也可以通过无线方式（例如移动网络、WIFI、蓝牙等）实现通信。

[114] 总线 1150 包括一通路，在设备的各个组件（例如处理器 1110、存储器 1120、输入/输出接口 1130 和通信接口 1140）之间传输信息。

[115] 需要说明的是，尽管上述设备仅示出了处理器 1110、存储器 1120、输入/输出接口 1130、通信接口 1140 以及总线 1150，但是在具体实施过程中，该设备还可以包括实现

正常运行所必需的其他组件。此外，本领域的技术人员可以理解的是，上述设备中也可以仅包含实现本说明书实施例方案所必需的组件，而不必包含图中所示的全部组件。

[116] 本说明书实施例还提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现前述的行业识别模型确定方法，所述方法至少包括：

[117] 检测到调用图形编码页面后，启动当前设备的前置图像采集模块进行图像采集；

[118] 对采集到的图像进行图像识别，若所述采集到的图像中包含预置的扫码模块的特征，则计算所述扫码模块相对于图形编码当前展示位置的偏移距离；

[119] 根据所述偏移距离计算目标展示位置，将所述图形编码在所述目标展示位置进行展示。

[120] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存（PRAM）、静态随机存取存储器（SRAM）、动态随机存取存储器（DRAM）、其他类型的随机存取存储器（RAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、快闪记忆体或其他内存技术、只读光盘只读存储器（CD-ROM）、数字多功能光盘（DVD）或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体（transitory media），如调制的数据信号和载波。

[121] 对于装置实施例而言，由于其基本对应于方法实施例，所以相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本说明书方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

[122] 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即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的，其中所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，在实施本说明书实施例方案时

可以把各模块的功能在同一个或多个软件和/或硬件中实现。也可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

[123] 以上所述仅是本说明书实施例的具体实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本说明书实施例原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也应视为本说明书实施例的保护范围。

权利要求书

1、一种行业识别模型确定方法，所述方法包括:

获取至少一个已确定行业类型的样本商户;

获取预设时间段内，包含样本商户每笔交易信息的交易信息集合;

通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态;

将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型。

2、如权利要求 1 所述的方法，所述样本商户的行业特征数据，包括样本商户在预定时间段内，每个单位时间的至少一种类别的交易信息。

3、如权利要求 2 所述的方法，所述预定时间段内，每个单位时间内至少一种类别的交易信息，包括: 在预定时间段内，每个单位时间的收款总金额信息、交易次数信息、均笔交易金额信息、交易金额方差信息和交易人数信息。

4、如权利要求 1 所述的方法，所述通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据后，还包括:

将同一样本商户任一类别的交易信息表达为交易图像，所述交易图像的横坐标用于表示交易的发生日期，所述交易图像的纵坐标用于表示每个发生日期内的不同发生时间点，所述交易图像中像素点的颜色用于表示所述类别的交易信息的数值大小。

5、如权利要求 4 所述的方法，所述将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型，包括:

将表达同一样本商户不同类型交易特征的交易图像数据作为 CNN 神经网络不同通道的输入数据，训练用于识别商户行业的识别模型。

6、如权利要求 4 所述的方法，所述行业识别模型的训练方式，包括:

将表达不同类型交易特征的交易图像数据分别处理为对应的一维向量序列;

将所述不同一维向量序列分别进行数据标准化处理，并将数据标准化处理后的不同一维向量序列整合重构为一个三维向量序列;

将所述三维向量序列作为不同通道的输入数据，输入 CNN 神经网络的卷积池化层，将反复卷积池化后的输出结果作为特征向量;

将特征向量展开为一维向量输入 CNN 神经网络全连接层，以完成行业识别模型对于所述样本商户的行业特征的学习。

7、如权利要求 1 所述的方法，所述已确定行业类型的样本商户为已确定为餐饮行

业的样本商户，所述样本商户的营业形态为，样本商户发生交易的时间集中在单日早中晚单位时间内，且交易金额在预设阈值之下。

8、一种基于权利要求 1 行业识别模型的商户行业确定方法，所述方法包括:

获取预设时间段内，包含待检验商户每笔交易信息的交易信息集合;

通过预定义的数据提取策略从所述交易信息集合中提取出待检验商户的行业特征数据;

将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业。

9、如权利要求 8 所述的方法，所述将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业，包括:

将待检验商户的行业特征数据输入训练好的行业识别模型，获取输出的评价分数，并将评价分数低于预设阈值的商户确定为问题商户。

10、一种行业识别模型确定装置，所述装置包括:

样本确定模块: 用于获取至少一个已确定行业类型的样本商户;

信息获取模块: 用于获取预设时间段内，包含样本商户每笔交易信息的交易信息集合;

数据提取模块: 用于通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态;

模型训练模块: 用于将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型。

11、如权利要求 10 所述的装置，所述样本商户的行业特征数据，包括样本商户在预定时间段内，每个单位时间的至少一种类别的交易信息。

12、如权利要求 11 所述的装置，所述预定时间段内，每个单位时间内至少一种类别的交易信息，包括: 在预定时间段内，每个单位时间的收款总金额信息、交易次数信息、均笔交易金额信息、交易金额方差信息和交易人数信息。

13、如权利要求 10 所述的装置，所述通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据后，还包括:

将同一样本商户任一类别的交易信息表达为交易图像，所述交易图像的横坐标用于表示交易的发生日期，所述交易图像的纵坐标用于表示每个发生日期内的不同发生时间点，所述交易图像中像素点的颜色用于表示所述类别的交易信息的数值大小。

14、如权利要求 13 所述的装置，将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型，包括:

将表达同一样本商户不同类型交易特征的交易图像数据作为 CNN 神经网络不同通道的输入数据，训练用于识别商户行业的识别模型。

15、如权利要求 13 所述的装置，所述行业识别模型的训练方式，包括:

将表达不同类型交易特征的交易图像数据分别处理为对应的一维向量序列;

将所述不同一维向量序列分别进行数据标准化处理，并将数据标准化处理后的不同一维向量序列整合重构为一个三维向量序列;

将所述三维向量序列作为不同通道的输入数据，输入 CNN 神经网络的卷积池化层，将反复卷积池化后的输出结果作为特征向量;

将特征向量展开为一维向量输入 CNN 神经网络全连接层，以完成行业识别模型对于所述样本商户的行业特征的学习。

16、如权利要求 10 所述的装置，所述已确定行业类型的样本商户为已确定为餐饮行业的样本商户，所述样本商户的营业形态为，样本商户发生交易的时间集中在单日早中晚单位时间内，且交易金额在预设阈值之下。

17、一种基于权利要求 10 行业识别模型的商户行业确定装置，所述装置包括:

信息获取模块: 用于获取预设时间段内，包含待检验商户每笔交易信息的交易信息集合;

数据提取模块: 用于通过预定义的数据提取策略从所述交易信息集合中提取出待检验商户的行业特征数据;

行业识别模块: 用于将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业。

18、如权利要求 17 所述的装置，将待检验商户的行业特征数据输入训练好的行业识别模型，以确定待检验商户的所在行业，包括:

将待检验商户的行业特征数据输入训练好的行业识别模型，获取输出的评价分数，并将评价分数低于预设阈值的商户确定为问题商户。

19、一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其中，所述处理器执行所述程序时实现如权利要求 1 所述的方法。

20、一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，其中，所述处理器执行所述程序时实现如权利要求 8 所述的方法。

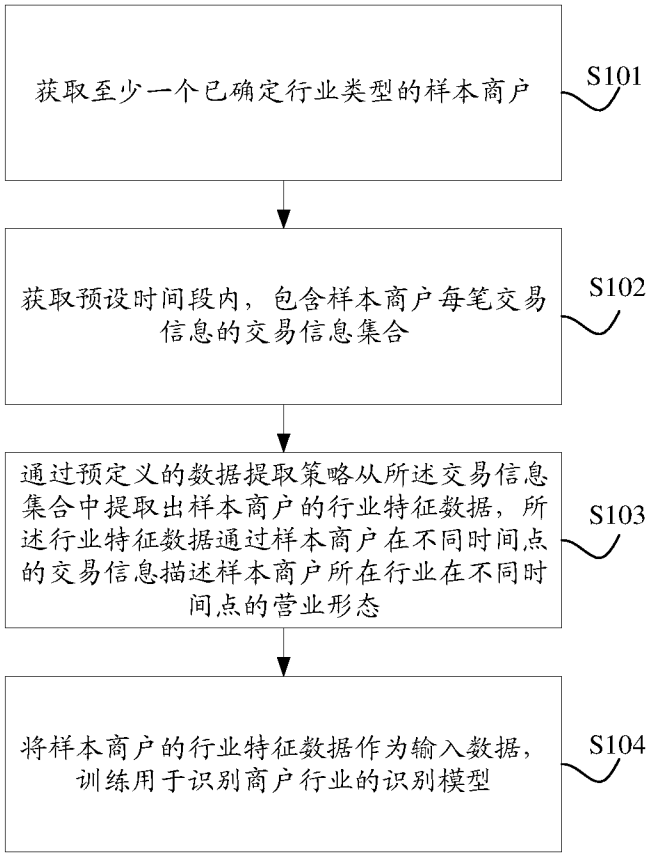

获取至少一个已确定行业类型的样本商户
S101
获取预设时间段内，包含样本商户每笔交易信息的交易信息集合
S102
通过预定义的数据提取策略从所述交易信息集合中提取出样本商户的行业特征数据，所述行业特征数据通过样本商户在不同时间点的交易信息描述样本商户所在行业在不同时间点的营业形态
S103
将样本商户的行业特征数据作为输入数据，训练用于识别商户行业的识别模型
S104

图 1

	1点	2点	3点	...	14点	..	24点
商户A	00	00	00		1350		9008
商户B	1200	800	00	...	0	...	1002
商户C	00	00	00	..	342	...	38
...				...			
商户N	100	40	00		259		102
...				...			

图 2

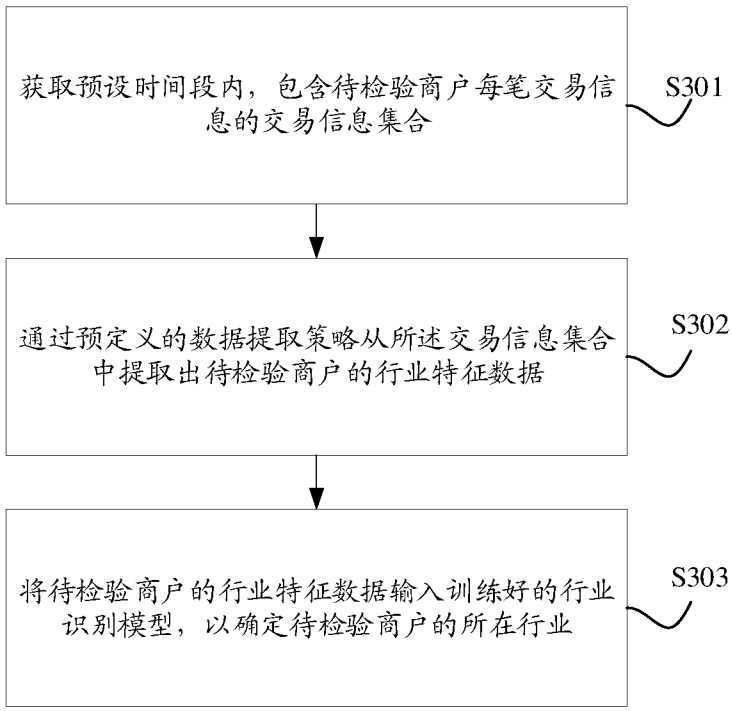


图 3

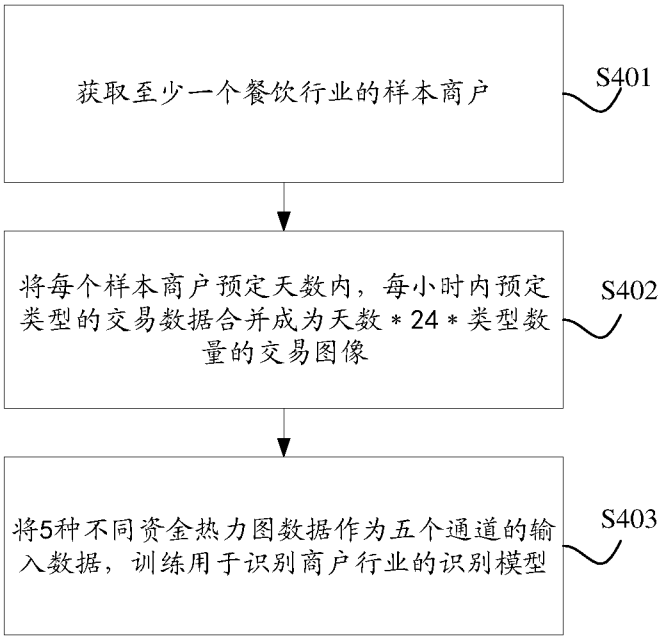

获取至少一个餐饮行业的样本商户
S401
将每个样本商户预定天数内，每小时内预定类型的交易数据合并成为天数＊24＊类型数量的交易图像
S402
将5种不同资金热力图数据作为五个通道的输入数据，训练用于识别商户行业的识别模型
S403

图 4

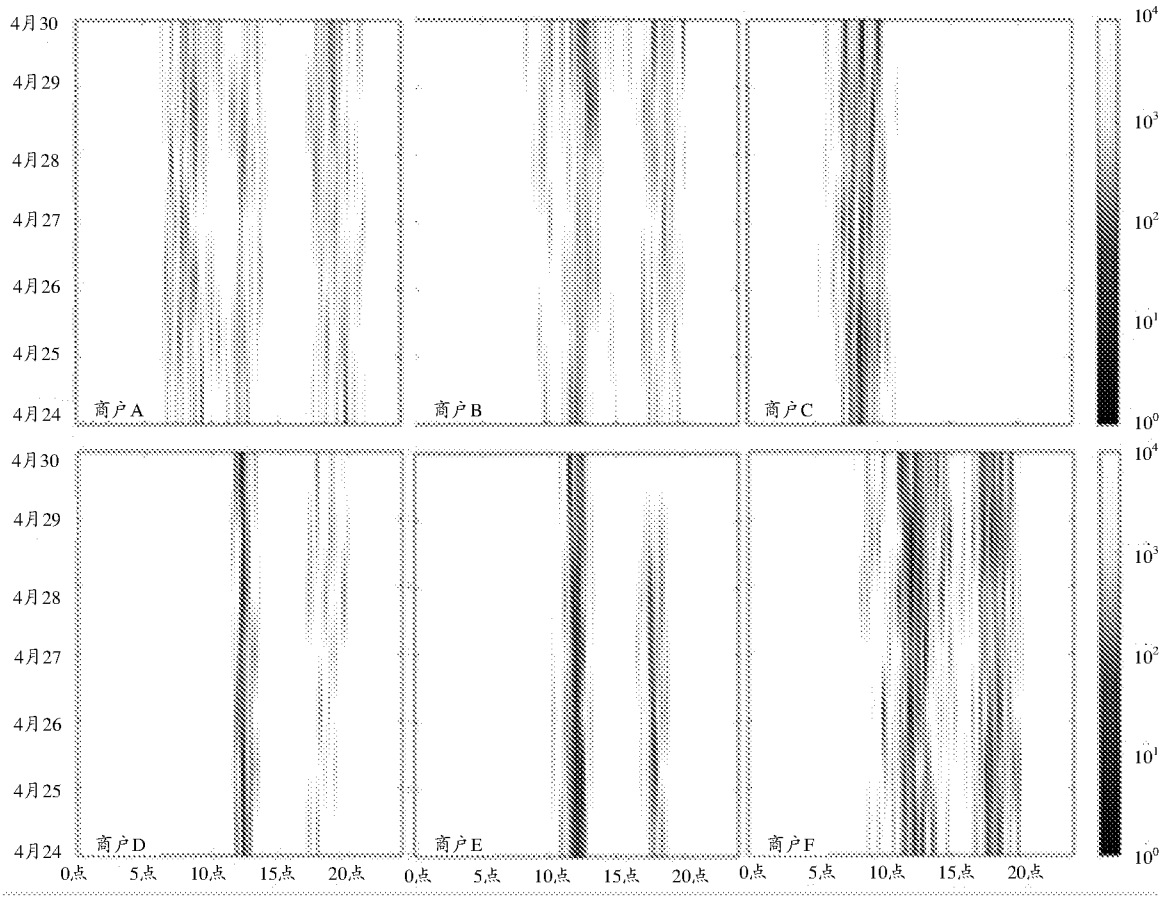

4月30
4月29
4月28
4月27
4月26
4月25
4月24
商户A
商户B
商户C
商户D
商户E
商户F
0点
5点
10点
15点
20点
10^4
10^3
10^2
10^1
10^0

图 5

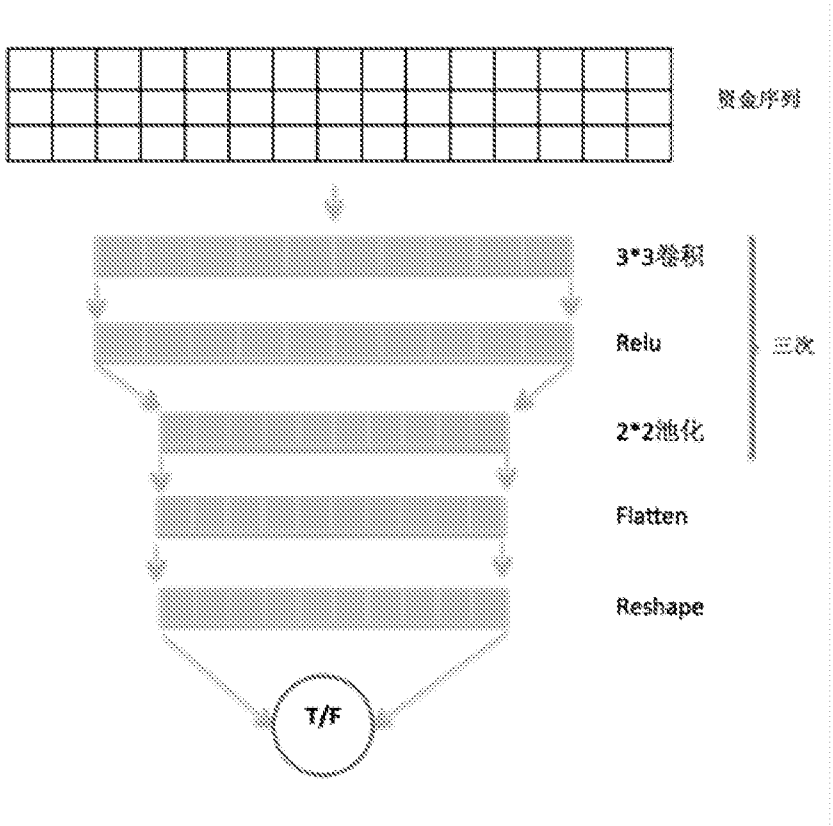

资金序列
3*3卷积
Relu
三次
2*2池化
Flatten
Reshape
T/F

图 6

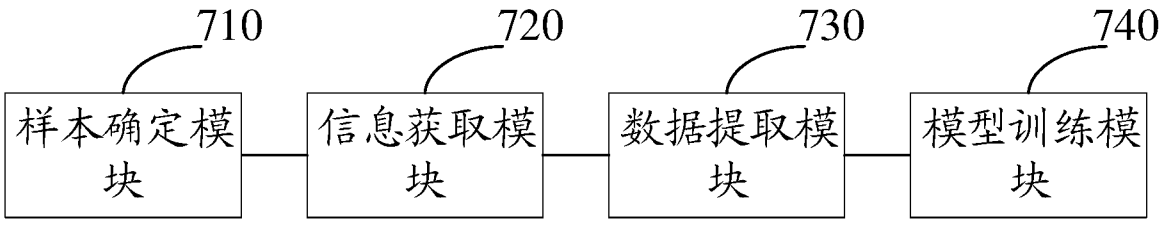

710
720
730
740
样本确定模块
信息获取模块
数据提取模块
模型训练模块

图 7

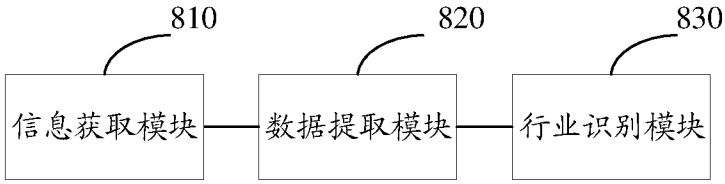

810
820
830
信息获取模块
数据提取模块
行业识别模块

图 8

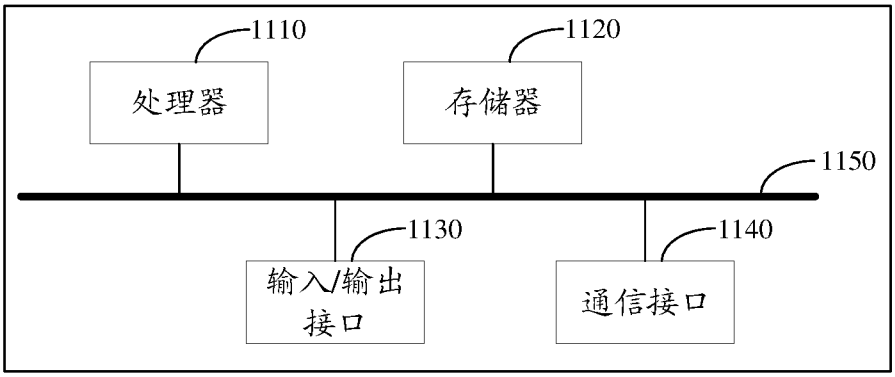

1110
1120
处理器
存储器
1150
1130
1140
输入/输出
接口
通信接口

图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/124199

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 20/42(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, 模型, 行业, 识别, 交易, 商户, 特征, model, industry, identif+, transact+, merchant, characteristic

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110009364 A (ALIBABA GROUP HOLDING LIMITED) 12 July 2019 (2019-07-12) claims 1-20	1-20
X	CN 109101989 A (ALIBABA GROUP HOLDING LIMITED) 28 December 2018 (2018-12-28) description, paragraphs [0035]-[0150]	1-20
A	CN 104881783 A (INSTITUTE OF INFORMATION ENGINEERING, CHINESE ACADEMY OF SCIENCES) 02 September 2015 (2015-09-02) entire document	1-20
A	CN 105184574 A (UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA) 23 December 2015 (2015-12-23) 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
22 February 2020	**09 March 2020**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/124199

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 110009364 A	12 July 2019	None	
CN 109101989 A	28 December 2018	None	
CN 104881783 A	02 September 2015	None	
CN 105184574 A	23 December 2015	CN 105184574 B	07 September 2018

国际检索报告

国际申请号
PCT/CN2019/124199

A. 主题的分类

G06Q 20/42(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，EPODOC，CNPAT，CNKI，模型，行业，识别，交易，商户，特征，model，industry，identif+，transact+，merchant，characteristic

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 110009364 A (阿里巴巴集团控股有限公司) 2019年 7月 12日 (2019-07-12) 权利要求1-20	1-20
X	CN 109101989 A (阿里巴巴集团控股有限公司) 2018年 12月 28日 (2018-12-28) 说明书第[0035]-[0150]段	1-20
A	CN 104881783 A (中国科学院信息工程研究所) 2015年 9月 2日 (2015-09-02) 全文	1-20
A	CN 105184574 A (电子科技大学) 2015年 12月 23日 (2015-12-23) 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
2020年 2月 22日	2020年 3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 焦永涵 电话号码 86-(10)-53961463

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/124199

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 110009364 A	2019年 7月 12日	无	
CN 109101989 A	2018年 12月 28日	无	
CN 104881783 A	2015年 9月 2日	无	
CN 105184574 A	2015年 12月 23日	CN 105184574 B	2018年 9月 7日