(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
2019 年 7 月 25 日 (25.07.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/141125 A1

(51) 国际专利分类号:
G06Q 40/02 (2012.01)

(21) 国际申请号: PCT/CN2019/071088

(22) 国际申请日: 2019 年 1 月 10 日 (10.01.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810050531.X 2018年1月18日 (18.01.2018) 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: 程微宏 (**CHENG, Weihong**); 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)。

赵星 (**ZHAO, Xing**); 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)。

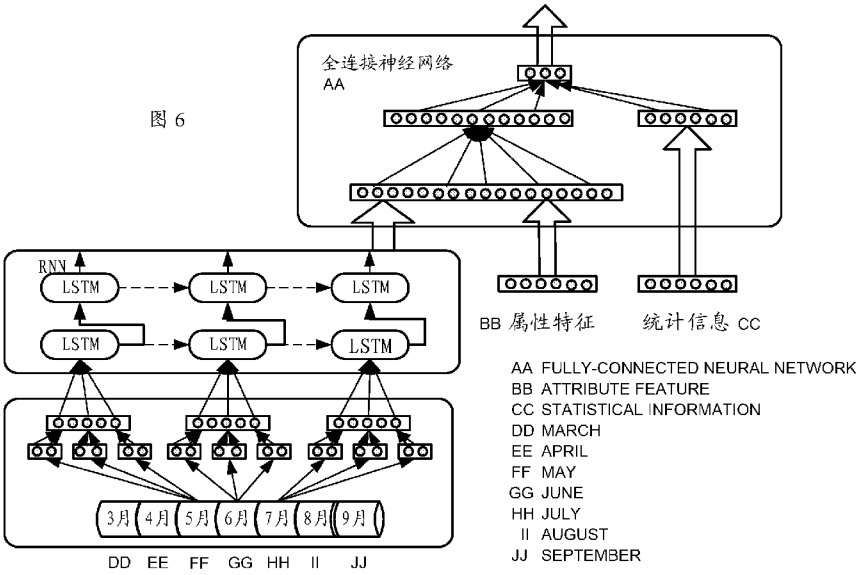
(74) 代理人: 北京博思佳知识产权代理有限公司 (**BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION**); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (CN)。

(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, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** METHOD AND DEVICE FOR ASSESSING FINANCIAL DEFAULT RISK

(54) 发明名称: 评估金融违约风险的方法及装置

图 6



(57) **Abstract:** A method and device for assessing financial default risk of a user. The method comprises: first obtaining a plurality of short-period fine-grained data sets comprising various amount information and merchant information related to the borrowing behaviors of the user, processing the data sets by using a recurrent neural network to obtain the intermediate processing result; obtaining coarse-grained statistical information of the user in the long period to serve as the breadth feature, and further processing the statistical information and the obtained intermediate processing result by using a fully-connected deep neural network, so that final financial risk assessment is carried out. The device corresponds to the method. In this case, the financial default risk is better accessed by combining the depth features and the breadth features.

WO 2019/141125 A1

[见续页]

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：一种评估用户的金融违约风险的方法和装置。方法包括，首先获取若干个短周期的细粒度数据集，该细粒度数据集包括用户的各种与借贷行为相关的金额信息和商户信息，利用循环神经网络处理这些数据集，得到中间处理结果。另一方面，还获取用户在长周期的粗粒度统计信息作为广度特征，利用全连接的深度神经网络进一步处理这些统计信息和以上得到的中间处理结果，从而进行最终的金融风险评估。所述装置与上述方法对应。如此，通过深度特征和广度特征结合的方式，更好地评估金融违约风险。

评估金融违约风险的方法及装置

技术领域

[01]本说明书一个或多个实施例涉及计算机技术领域，尤其涉及评估金融违约风险的方法及装置。

背景技术

[02]计算机和网络技术的发展使得互联网已经渗透进人们生活的方方面面，人们越来越多地利用互联网进行各种各样的操作，包括金融相关操作，例如网络购物，电子支付，电子转账，在线理财，在线借贷等。在用户的诸多网络金融操作中，有一些操作行为有一定的金融风险，例如请求先享后付类服务，采用花呗、白条等透支服务，申请借贷等。这就需要对用户的金融违约风险预先进行评估和判断。尽管已经存在诸如芝麻信用这样的信用评估系统，但是仅仅根据信用分判断金融违约风险，维度比较单一，粒度不够细致。

[03]因此，需要更有效的方式，对用户进行全面分析，从而评估其金融违约风险。

发明内容

[04]本说明书一个或多个实施例描述了一种方法和装置，可以全面地对用户的金融违约风险进行分析和评估。

[05]根据第一方面，提供了一种评估金融违约风险的方法，包括:

[06]获取数据集序列，所述数据集序列包括对应于连续的多个第一时间段、且按照时间顺序排列的多个数据集，所述多个数据集中的各个数据集包括，对应的第一时间段中，用户的与借贷行为相关的金额信息和商户信息;

[07]利用循环神经网络处理所述数据集序列，获得第一输出结果;

[08]获取与所述用户在第二时间段内的借贷行为相关的统计信息，所述第二时间段大于所述第一时间段;

[09]利用全连接神经网络处理所述第一输出结果和所述统计信息，获得第二输出结果;

[10]根据所述第二输出结果评估所述用户的金融违约风险。

[11]在一种可能的设计中，所述商户信息包括，与所述用户的借贷行为相关的商户 ID。

[12]进一步地，在一个实施例中，与所述用户的借贷行为相关的商户 ID 包括以下中的一项或多项，用户申请借款的商户 ID，用户借款成功的商户 ID，用户借款失败的商户 ID，用户履约的商户 ID，用户违约的商户 ID。

[13]在一个实施例中，所述方法还包括：对所述商户信息进行预处理，所述预处理包括：通过词嵌入模型将所述商户 ID 转化为 ID 向量。

[14]根据一个实施例，上述预处理还包括：在同一数据集中，对所述用户的同类型借贷行为所针对的商户 ID 的 ID 向量求平均，获得各类型对应的均值向量；将所述均值向量包含在预处理的商户信息中。

[15]在一个实施例中，上述预处理还包括：对所述用户的不同类型的借贷行为中各个类型对应的所述均值向量求和，获得商户向量和；将所述商户向量和包含在预处理的商户信息中。

[16] 根据一个实施例，所述多个数据集中的各个数据集还包括，用户在对应的第一时间段内的信用信息。

[17]在一个实施例中，所述方法还包括，获取所述用户的属性特征信息，所述利用全连接神经网络处理所述第一输出结果和所述统计信息，获得第二输出结果，包括：将所述第一输出结果、所述统计信息和所述属性特征信息输入所述全连接神经网络，获得第二输出结果。

[18]进一步地，在一个实施例中，所述全连接神经网络为多层全连接神经网络，其中将所述第一输出结果、所述统计信息和所述属性特征信息输入所述全连接神经网络包括：将所述第一输出结果和所述属性特征信息输入所述多层全连接神经网络的第一层，将所述统计信息输入所述全连接神经网络的最后一层。

[19]根据一种可能的设计，所述循环神经网络包括多层叠加的长短期记忆模型 LSTM。

[20]根据第二方面，提供了一种评估金融违约风险的装置，包括：

[21]数据集获取单元，配置为获取数据集序列，包括对应于连续的多个第一时间段、且按时间顺序排列的多个数据集，所述多个数据集中的各个数据集包括，对应的第一时间段中，用户的与借贷行为相关的金额信息和商户信息；

[22]第一处理单元，配置为利用循环神经网络处理所述数据集序列，获得第一输出结果；

[23]统计信息获取单元，配置为获取与所述用户在第二时间段内的借贷行为相关的统计信

息，所述第二时间段大于所述第一时间段；

[24]第二处理单元，配置为利用全连接神经网络处理所述第一输出结果和所述统计信息，获得第二输出结果；

[25]评估单元，配置为根据所述第二输出结果评估所述用户的金融违约风险。

[26]根据第三方面，提供了一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行第一方面的方法。

[27]根据第四方面，提供了一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现第一方面的方法。

[28]通过本说明书实施例提供的方法及装置，采用了深度模型和广度模型相结合的架构分析用户的金融违约风险。具体地，采用循环神经网络作为深度模型，对用户的多个短周期中细粒度的数据集进行深度分析，反映用户的借贷状态的变化。广度模型基于用户长周期的统计特征，获取用户更为宏观的统计信息，与深度模型的分析结果进行全连接，进而全面分析用户的金融违约风险。

附图说明

[29]为了更清楚地说明本发明实施例的技术方案，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发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其它的附图。

[30]图 1 为本说明书披露的一个实施例的实施场景示意图；

[31]图 2 示出根据一个实施例的评估金融违约风险的方法的流程图；

[32]图 3 示出根据一个实施例的循环神经网络的处理示意图；

[33]图 4 示出一个实施例中全连接神经网络的处理示意图；

[34]图 5 示出另一实施例中全连接神经网络的处理示意图；

[35]图 6 示出根据一个实施例的网络模型结构示意图；

[36]图 7 示出根据一个实施例的评估装置的示意框图。

具体实施方式

[37]下面结合附图，对本说明书提供的方案进行描述。

[38]图 1 为本说明书披露的一个实施例的实施场景示意图。如图 1 所示，用户有可能通过网络，进行多种与借贷相关的行为，例如向各种商户申请借款，还款，申请延期还款等。相应地，商户对应的服务器可以记录用户的与借贷相关的行为记录。例如，用户通过支付宝下的“蚂蚁借呗”申请了一笔借款，然后每月还款，这时可以认为商户即为支付宝平台，那么在支付宝平台服务器中会记录用户的该借贷行为。用户还通过某电子借贷平台申请了一笔借款，那么该借贷平台即为对应商户，并且该借贷平台的服务器会记录用户的借贷行为。可以理解，取决于用户进行借贷的商户，记录用户借贷行为的服务器可以是集中的服务器，也可以分布式服务器，还可以是互相完全独立的多个服务器，在此不做限定。

[39]为了对用户的金融违约风险进行评估，各个借贷平台会将用户的借贷行为数据经过整理之后提供给计算平台，由计算平台进行综合分析，确定出用户的违约风险。在说明书披露的实施例中，计算平台获取各个借贷商户所提供的用户行为相关数据之后，采用深度模型和广度模型相结合的构架，利用机器学习和循环神经网络，对这些数据进行全面分析，从而对金融风险进行评估。上述计算平台可以是任何具有计算、处理能力的装置、设备和系统，例如可以是服务器，它既可以作为独立的计算平台，也可以集成到记录用户借贷行为的某些服务器中。更具体地，计算平台一方面获取若干个短周期（例如 1 个月）的细粒度数据集，该细粒度数据集包括用户的各种与借贷行为相关的金额信息和商户信息，利用循环神经网络处理这些数据集，得到中间处理结果。另一方面，计算平台还获取用户在长周期的粗粒度统计信息作为广度特征，利用全连接的深度神经网络进一步处理这些统计信息和以上得到的中间处理结果，从而进行最终的金融风险评估。下面描述计算平台评估金融违约风险的具体过程。

[40]图 2 示出根据一个实施例的评估金融违约风险的方法的流程图。该方法流程的执行主体可以是任何具有计算、处理能力的装置、设备和系统，例如图 1 中的计算平台。如图 2 所示，在该实施例中，评估金融违约风险的方法包括以下步骤：步骤 21，获取数据集序列，包括按时间顺序排列的多个数据集，各数据集分别对应预设的第一时间段；所述多个数据集中的各个数据集包括，用户的与借贷行为相关的金额信息和商户信息；步骤 22，利用循环神经网络处理所述数据集序列，获得第一输出结果；步骤 23，获取与所述用户在第二时间段内的借贷行为相关的统计信息，所述第二时间段大于所述第一时间段；步骤 24，利用

全连接神经网络处理所述第一输出结果和所述统计信息，获得第二输出结果；以及步骤 25，根据所述第二输出结果评估所述用户的金融违约风险。下面具体描述以上各个步骤的执行过程。

[41]首先，在步骤 21，获取数据集序列，其包括按时间顺序排列的多个数据集，各数据集分别对应预设的第一时间段；其中各个数据集包括，对应的第一时间段中，用户的与借贷行为相关的金额信息和商户信息。

[42]在一个实施例中，上述第一时间段为相对较短的时间段，例如一周，半个月，一个月等。典型地，上述第一时间段为一个月。为了描述简单，在本文中又将该第一时间段称为短周期。针对这样的短周期，可以获取每个短周期内与用户的借贷行为有关的细粒度的信息，将这样的信息整理为一个数据集。具体地，针对每个短周期的细粒度数据集可以包括，对应短周期内与用户的借贷行为有关的金额信息和商户信息。

[43]在一个实施例中，上述金额信息可以包括以下中的一项或多项：对应短周期内的新增借款金额、还款金额、订单金额、应还金额、逾期金额、总结欠金额等。例如，在 3 月份，用户在 A 借贷平台借款 1 万元，约定每月还款 2500 元，分四期完成还款；在 4 月份，该用户向 A 借贷平台还款 2000 元，同时还向 B 借贷平台借款 5000 元，约定每月还款 1000 元，分五期完成还款。那么，3 月份对应的金额信息具体可以包括：新增借款金额 1 万元，订单金额 1 万，总结欠 1 万元；4 月份对应的金额信息具体可以包括：新增借款金额 5000 元，还款金额 2000 元，订单金额 15000 元（1 万元+5000 元），应还金额 2500 元，逾期金额 500 元，总结欠金额 13000 元（15000-2000 元）。金额信息还可以包含更多项的与借贷金额相关的数据。

[44]此外，短周期对应的数据集还包括，对应时间段中用户的与借贷行为相关的商户信息。在一个实施例中，上述商户信息包括，与用户的借贷行为相关的商户 ID。更具体地，与所述用户的借贷行为相关的商户 ID 可以包括以下中的一项或多项：用户申请借款的商户 ID，用户借款成功的商户 ID，用户借款失败的商户 ID，用户履约的商户 ID，用户违约的商户 ID 等。例如，在一个例子中，用户在 3 月份共向 4 个借贷平台 A，B，C 和 D 申请了借款，其中 A，B，C 平台同意了该用户的借款请求并放款，而 D 平台拒绝了用户的借款请求；在 4 月份，该用户又向 E 平台申请借款并成功，同时，按照约定向 A，B 平台还款，而对于已经借款的 C 平台，未能按时还款。那么针对该用户，其 3 月份数据集中的商户信息如下面的表 1A 所示，4 月份的数据集中的商户信息如表 1B 所示。

表 1A

申请借款的商户 ID	借款成功的商户 ID	借款失败的商户 ID	用户履约的商户 ID	用户违约的商户 ID
A, B, C, D	A, B, C	D		

表 1B

申请借款的商户 ID	借款成功的商户 ID	借款失败的商户 ID	用户履约的商户 ID	用户违约的商户 ID
E	E		A, B	C

[45]在一个实施例中，为了便于机器模型更好地处理商户信息，对获得的商户信息进行预处理。在一个实施例中，上述预处理包括，通过词嵌入模型（embedding 模型）将以上的商户 ID 转化为向量的形式，此处称为 ID 向量。

[46]可以理解，词嵌入模型是自然语言处理 NLP 中用到的一种模型，用于将单个词转换为一个向量。在最简单的模型中，为每个单词构造一组特征作为其对应向量。更进一步地，为了体现单词之间的关系，例如类别关系，从属关系，可以采用各种方式训练语言模型，优化向量表达。例如，word2vec 的工具中包含了多种词嵌入的方法，能够快速得到单词的向量表达，并且向量表达能够体现单词之间的类比关系。例如，单词“北京”与单词“中国”的对应向量之间的关系，与单词“巴黎”与单词“法国”的对应向量之间的关系相一致，如此，通过词向量体现对应的单词之间的类别和类比关系。还存在一些其他的词嵌入算法。

[47]在商户信息的预处理中，可以根据需要，选取适当的词嵌入模型，将商户 ID 转换为对应的 ID 向量。如此，得到预处理的商户信息，在其中用不同的 ID 向量表示不同商户。例如，对于表 1A 和表 1B 所示的商户信息，通过词嵌入预处理，可以得到预处理后的商户信息，如表 2A 和表 2B 所示。

表 2A

申请借款的商户 ID	借款成功的商户 ID	借款失败的商户 ID	用户履约的商户 ID	用户违约的商户 ID
$\overrightarrow{V_A}$, $\overrightarrow{V_B}$, $\overrightarrow{V_C}$, $\overrightarrow{V_D}$	$\overrightarrow{V_A}$, $\overrightarrow{V_B}$, $\overrightarrow{V_C}$	$\overrightarrow{V_D}$		

表 2B

申请借款的商户 ID	借款成功的商户 ID	借款失败的商户 ID	用户履约的商户 ID	用户违约的商户 ID
$\overrightarrow{V_E}$	$\overrightarrow{V_E}$		$\overrightarrow{V_A}$, $\overrightarrow{V_B}$	$\overrightarrow{V_C}$

[48]在表 2A 和 2B 中，$\overrightarrow{V_A}$，$\overrightarrow{V_B}$，$\overrightarrow{V_C}$，$\overrightarrow{V_D}$ 和 $\overrightarrow{V_E}$ 分别表示商户 A，B，C，D，E 对应的 ID 向量。如此，用向量的方式表征各个商户的商户 ID。

[49]在一个实施例中，对以上 ID 向量进行进一步处理。具体地，在同一数据集中，对所述用户的同类型借贷行为所针对的商户 ID 的 ID 向量求平均，获得各类型对应的均值向量，作为预处理的商户信息。例如，在某个数据集中，商户信息包括，用户向 3 个商户申请借款的行为，那么可以计算这 3 个商户的 ID 向量的平均值，作为申请借款这一类型对应的均值向量。表 3A 示出对表 2A 进行进一步处理，获取的各类型下的均值向量。

表 3A

申请借款的商户 ID	借款成功的商户 ID	借款失败的商户 ID	用户履约的商户 ID	用户违约的商户 ID
$\overrightarrow{V_1}$	$\overrightarrow{V_2}$	$\overrightarrow{V_D}$		

[50]在表 3A 中，$\overrightarrow{V_1}$ 为 ID 向量 $\overrightarrow{V_A}$，$\overrightarrow{V_B}$，$\overrightarrow{V_C}$，$\overrightarrow{V_D}$ 的平均值，$\overrightarrow{V_2}$ 为 ID 向量 $\overrightarrow{V_A}$，$\overrightarrow{V_B}$，$\overrightarrow{V_C}$ 的平均值。如此，仅用一个向量表征一个类别下的商户 ID 信息。

[51]在一个实施例中，在商户信息预处理过程中，对以上的均值向量进行进一步处理。具体地，在同一数据集中，对用户的不同类型的借贷行为中各个类型对应的均值向量求和，获得商户向量和。如前所述，经过向量均值处理，已经将用户的一个类型的借贷行为对应的商户 ID 处理为一个均值向量。进一步地，可以对不同类型借贷行为各自对应的均值向量进行求和，来获得商户向量和，作为预处理的商户信息。在一个例子中，将不同类型借贷行为中各个类型对应的均值向量简单求和，从而获得商户向量和。例如，对于表 3A 所示的商户信息，可以简单地将不同类型下的均值向量，即申请借款的商户 ID 的均值向量$\overrightarrow{V_1}$，借款成功的商户 ID 的均值向量$\overrightarrow{V_2}$，以及借款失败的商户 ID 的均值向量$\overrightarrow{V_D}$，进行求和$\vec{V}=\overrightarrow{V_1}+\overrightarrow{V_2}+\overrightarrow{V_D}$，获得商户向量和$\vec{V}$。在另一个例子中，为各个类型的借贷行为赋予一定权重，对各个类型的借贷行为对应的均值向量进行加权求和。仍以表 3A 为例，假定申请借款对应的权重因子为 w1，借款成功对应的权重因子为 w2，借款失败对应的权重因子为 w3，那么在该例子中，商户向量和可以计算为$\vec{V}=w_1*\overrightarrow{V_1}+w_2*\overrightarrow{V_2}+w_3*\overrightarrow{V_D}$。通过这样的方式，用一个向量，即商户向量和，来综合地表征一个短周期内与用户的借贷行为相关的商户信息。如前所述，在词嵌入过程中，优选地使得转化的 ID 向量能够体现对应商户 ID 的类别关系。因此，对于商户 ID 的 ID 向量进行均值计算，以及进一步的求和计算，最终的结果能够从总体上反映，用户借贷行为相关的商户的平均特征和综合特征。

[52]以上描述了对数据集中商户信息中的商户 ID 的处理。

[53]在一个实施例中，商户信息还可以包括，用户的各类型借贷行为相关的商户数目。例如，基于表 1A 所示的 3 月份中各类型借贷行为对应的商户 ID，可以确定出各借贷行为对应的商户数目，比如申请借款的商户数为 4，借款成功的商户数为 3，借款失败的商户数为 1。这些有关商户数目的信息也可以包含在商户信息中。

[54]除了以上的金额信息和商户信息，对应于短周期的数据集还可以包含更多与借贷行为相关的信息。例如，在一个实施例中，数据集中还可以包括与借贷行为相关的订单数目信息，比如借款订单数目，还款订单数据，履约订单数目等。

[55]在一个实施例中，各个数据集还包括，用户在对应的时间段内的信用信息。该信用信息例如包括，芝麻信用分，第三方征信机构提供的信用评分，以及借贷商户可能调用的其他信用相关特征，例如交易活跃度、支付完成率等。

[56]在更多实施例中，各个数据集还可以包括更多种细粒度的与用户借贷行为相关的数据

信息。

[57]对应于多个短周期的数据集可以构成一个具有先后时序的数据集序列。如前所述，在一个实施例中，上述数据集中的商户信息可以经过预处理，转化为不同层级的向量形式。这样的数据集序列可以输入到循环神经网络中进行处理。也就是，在步骤 S22，利用循环神经网络处理多个数据集构成的数据集序列，获得中间结果，或称为第一输出结果。

[58]可以理解，循环神经网络（RNN, Recurrent Neural Networks）是一种时间递归神经网络，可用于处理序列数据。在 RNN 中，一个序列当前的输出与其前面的输出相关联。具体的，RNN 会对前面的信息进行记忆并应用于当前输出的计算中，即隐藏层之间的节点是有连接的，并且隐藏层的输入不仅包括输入层的输出还包括上一时刻隐藏层的输出。也就是说，第 t 次的隐含层状态可以表示为：

$$St=f(U*Xt+W*St-1)$$

[59]其中，Xt 为第 t 次输入层的状态，St-1 为第 t-1 次隐含层状态，f 为计算函数，W,U 为权重。如此，RNN 将之前的状态循环回当前输入，考虑了历史输入的影响，因而适合于具有时序的数据序列。

[60]在一个实施例中，在步骤 22，采用多层 RNN 来处理上述数据集序列。

[61]更进一步地，在一个实施例中，在 RNN 架构下，采用长短期记忆模型（LSTM，Long Short Term Memory）进行处理。

[62]如前所述，在 RNN 中当前隐含层状态依赖于之前的状态输出，因此在处理长期记忆的情况下，需要将当前的隐含态的计算与前 n 次的计算关联，即 St = f(U*Xt + W1*St-1 + W2*St-2 + … + Wn*St-n)。随着 n 的增大，计算量呈指数式增长，导致模型训练的时间大幅增加。为此，提出 LSTM 模型来解决长期依赖的问题。

[63]在 LSTM 模型中，通过设置让信息选择性通过的“遗忘门”来丢弃某些不再需要的信息，如此对输入的不必要的干扰信息进行判断和屏蔽，从而更好地对长期数据进行分析处理。

[64]在一个实施例中，采用多层叠加的 LSTM 模型来处理多个数据集构成的数据集序列。

[65]图 3 示出根据一个实施例的循环神经网络的处理示意图。在图 3 的例子中，短周期采取一个月，从而可以获取每个月对应的数据集，各个数据集中包括当月用户的与借贷行为有关的信息，包括金额信息，商户信息，订单信息等。这样的连续 6 个月的数据集构成一

个数据集序列。对于这样的数据集序列，可选地，在一个例子中，在输入层对各个数据集中的商户信息进行预处理。预处理可以包括如上所述的 ID 向量化，求均值向量，求商户向量和中的至少一部分，还可以包括更多的预处理方式。图中条形框中的每个圆圈示意性地对应数据集中的一项特征。经过预处理的数据集序列被输入到循环神经网络 RNN 进行处理。在图 3 的例子中，循环神经网络包括多层叠加的LSTM模型，图 3 示例中每一纵列的LSTM块对应于一个月的数据集。经过多层RNN的处理，可以获得中间处理结果。

[66]除了如上所述获得各个短周期内的细粒度借贷行为数据之外，另一方面，在步骤 23，还获取与用户在第二时间段内的借贷行为相关的统计信息，该第二时间段大于步骤 21 中的第一时间段。如果将第一时间段称为短周期的话，在本文中对应地将第二时间段称为长周期。一般地，长周期可以是短周期的若干倍数。例如，在短周期典型地为一个月的情况下，长周期可以是例如 3 个月，6 个月，或 12 个月。相应地，在步骤 23 中，获取用户在 3/6/12 个月中与借贷行为相关的统计信息。这样的统计信息可以包括，借款总金额，借款总次数，还款总金额等。相比于步骤 21 中获得的数据集中的各种具体信息，步骤 23 中的统计信息是更加粗粒度的综合信息。

[67]在此基础上，在步骤 24，将步骤 22 获得的、循环神经网络输出的中间结果，和步骤 23 获得的长周期内粗粒度的统计信息，共同输入到全连接神经网络进行处理。

[68]可以理解，在全连接神经网络中，每一个节点都与上一层的所有节点相连，用于把前边提取到的所有特征综合起来。相应地，在步骤 24，利用全连接神经网络，对输入的统计信息和中间结果进行全面关联分析，从而提供全连接输出，即第二输出。

[69]基于这样的全连接输出，在步骤 25，可以确定用户的金融违约风险。在一个实施例中，全连接神经网络对输入数据进行分析处理后，可以输出该用户未来的金融违约概率。于是可以基于该金融违约概率确定用户的金融违约风险。在另一实施例中，全连接神经网络可以输出用户的各类型行为的风险指数，例如借贷风险指数，透支风险指数，购买先享后付类服务的风险指数等等。在步骤 25，可以根据需要，确定用户的各类型行为的风险指数以及综合金融违约风险。在又一实施例中，全连接神经网络可以输出针对不同类型商户的风险指数。相应地，在步骤 25，可以根据需要，确定针对不同商户的金融违约风险，以及综合金融违约风险。可以理解，全连接神经网络的输出结果的具体形式，取决于对全连接神经网络，以及循环神经网络的训练过程。

[70]在一个实施例中，为了更全面地对用户进行评估，还获取上述用户的属性特征，例如

年龄、性别、地域、职业、学历等等。将用户的属性特征，和上述的中间结果和统计信息一起，共同输入到全连接神经网络进行进一步处理。

[71]图 4 示出一个实施例中全连接神经网络的处理示意图。在图 4 的示意图中，全连接神经网络的输入包含 3 个部分，第一部分为图 3 的循环神经网络所输出的中间结果，第二部分为用户长周期的统计信息，第三部分为用户的属性特征。在图 4 的例子中，全连接神经网络是 2 层全连接网络，所有数据输入到第一层，第一层的处理结果输入到第二层，第二层的每一个节点都与第一层的所有节点相连。第二层输出的结果可以用于评估用户的金融违约风险。

[72]图 5 示出另一实施例中全连接神经网络的处理示意图。在图 5 的示意图中，全连接神经网络为多层全连接神经网络，其中，图 3 的循环神经网络输出的中间结果与用户的属性特征一起，输入到多层全连接神经网络的第一层，而用户的长周期的统计信息输入到该网络的最后一层。这是因为，用户的属性特征相对比较泛化，需要与别的特征结合做全连接。因此，将用户的属性特征与中间结果一起输入到网络的第一层。而用户的统计信息具有独立而明确的特征，直接将其输入到最后一层，也就是最外层，更方便全连接神经网络的训练和调试。

[73]图 6 示出根据一个实施例的网络模型结构示意图。可以理解，图 6 将图 3 的循环神经网络与图 5 的全连接神经网络结合在一起，示出了图 2 方法中数据的总体处理结构。通过图 6 可以看到，在该实施例的方法中，采用了深度模型和广度模型相结合的架构。具体地，采用循环神经网络作为深度模型，对用户的多个短周期中细粒度的数据集进行深度分析，充分利用用户借贷行为中的时序特征，反映用户的借贷状态的变化。广度模型基于用户长周期的统计特征，获取用户更为宏观的统计信息，与深度模型的分析结果进行全连接，进而全面分析用户的金融违约风险。并且，在深度模型分析中，输入的数据集中包含商户信息的特征，从而在金融风险分析中引入了商户的风控能力和客群特征，可以增加模型对风险的区分度。

[74]根据另一方面的实施例，还提供一种评估金融违约风险的装置。图 7 示出根据一个实施例的评估装置的示意框图。如图 7 所示，用于评估金额违约风险的装置 700 包括：数据集获取单元 71，配置为获取数据集序列，其包括按时间顺序排列的多个数据集，各数据集分别对应预设的第一时间段；所述多个数据集中的各个数据集包括用户的与借贷行为相关的金额信息和商户信息；第一处理单元 72，配置为利用循环神经网络处理所述数据集序列，

获得第一输出结果；统计信息获取单元 73，配置为获取与所述用户在第二时间段内的借贷行为相关的统计信息，所述第二时间段大于所述第一时间段；第二处理单元 74，配置为利用全连接神经网络处理所述第一输出结果和所述统计信息，获得第二输出结果；评估单元 75，配置为根据所述第二输出结果评估所述用户的金融违约风险。

[75]在一个实施例中，所述商户信息包括，与所述用户的借贷行为相关的商户 ID。

[76]进一步地，在一个例子中，与所述用户的借贷行为相关的商户 ID 包括以下中的一项或多项：用户申请借款的商户 ID，用户借款成功的商户 ID，用户借款失败的商户 ID，用户履约的商户 ID，用户违约的商户 ID。

[77]在一个实施例中，装置 700 还包括预处理单元 711，配置为对所述商户信息进行预处理，所述预处理包括：通过词嵌入模型将所述商户 ID 转化为 ID 向量。

[78]进一步地，在一个实施例中，所述预处理单元 711 还配置为：在同一数据集中，对所述用户的同类型借贷行为所针对的商户 ID 的 ID 向量求平均，获得各类型对应的均值向量；将所述均值向量包含在预处理的商户信息中。

[79]更进一步地，在一个实施例中，所述预处理单元 711 还配置为：对所述用户的不同类型的借贷行为中各个类型对应的所述均值向量求和，获得商户向量和；将所述商户向量和包含在预处理的商户信息中。

[80]根据一种实施方式，在一个实施例中，商户信息还可以包括，用户的各类型借贷行为相关的商户数目。

[81]除了以上的金额信息和商户信息，在一个实施例中，以上数据集还可以包含与借贷行为相关的订单数目信息。在另一实施例中，以上数据集还包括用户在对应的第一时间段内的信用信息。

[82]根据一个实施例，上述装置 700 还包括属性特征获取单元（未示出），配置为获取所述用户的属性特征信息。相应地，第二处理单元 74 配置为：将所述第一输出结果、所述统计信息和所述属性特征信息输入所述全连接神经网络，获得第二输出结果。

[83]进一步地，在一个实施例中，所述第二处理单元 74 还配置为：将所述第一输出结果和所述属性特征信息输入所述多层全连接神经网络的第一层，将所述统计信息输入所述全连接神经网络的最后一层。

[84]根据一个实施例，所述循环神经网络包括多层叠加的长短期记忆模型 LSTM。

[85]可以看到，以上的装置 700 采用了深度模型和广度模型相结合的方式，评估用户的金融违约风险。具体地，采用循环神经网络作为深度模型，对用户的多个短周期中细粒度的数据集进行深度分析。广度模型基于用户长周期的统计特征，获取用户更为宏观的统计信息，与深度模型的分析结果进行全连接，进而全面分析用户的金融违约风险。并且，在深度模型分析中，输入的数据集中包含商户信息的特征，从而在金融风险分析中引入了商户的风控能力和客群特征，可以增加模型对风险的区分度。

[86]根据另一方面的实施例，还提供一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行结合图 2 所描述的方法。

[87]根据再一方面的实施例，还提供一种计算设备，包括存储器和处理器，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现结合图 2 所述的方法。

[88]本领域技术人员应该可以意识到，在上述一个或多个示例中，本发明所描述的功能可以用硬件、软件、固件或它们的任意组合来实现。当使用软件实现时，可以将这些功能存储在计算机可读介质中或者作为计算机可读介质上的一个或多个指令或代码进行传输。

[89]以上所述的具体实施方式，对本发明的目的、技术方案和有益效果进行了进一步详细说明，所应理解的是，以上所述仅为本发明的具体实施方式而已，并不用于限定本发明的保护范围，凡在本发明的技术方案的基础之上，所做的任何修改、等同替换、改进等，均应包括在本发明的保护范围之内。

权利要求书

1. 一种评估金融违约风险的方法，包括:

获取数据集序列，所述数据集序列包括按照时间顺序排列的多个数据集，各数据集分别对应预设的第一时间段；所述多个数据集中的各个数据集包括用户的与借贷行为相关的金额信息和商户信息；

利用循环神经网络处理所述数据集序列，获得第一输出结果；

获取与所述用户在第二时间段内的借贷行为相关的统计信息，所述第二时间段大于所述第一时间段；

利用全连接神经网络处理所述第一输出结果和所述统计信息，获得第二输出结果；

根据所述第二输出结果评估所述用户的金融违约风险。

2. 根据权利要求 1 所述的方法，其中所述商户信息包括，与所述用户的借贷行为相关的商户 ID。

3. 根据权利要求 2 所述的方法，其中与所述用户的借贷行为相关的商户 ID 包括以下中的一项或多项，用户申请借款的商户 ID，用户借款成功的商户 ID，用户借款失败的商户 ID，用户履约的商户 ID，用户违约的商户 ID。

4. 根据权利要求 2 所述的方法，还包括：对所述商户信息进行预处理，所述预处理包括：通过词嵌入模型将所述商户 ID 转化为 ID 向量。

5. 根据权利要求 4 所述的方法，其中所述预处理还包括:

在同一数据集中，对所述用户的同类型借贷行为所针对的商户 ID 的 ID 向量求平均，获得各类型对应的均值向量；

将所述均值向量包含在预处理的商户信息中。

6. 根据权利要求 5 所述的方法，其中所述预处理还包括:

对所述用户的不同类型的借贷行为中各个类型对应的所述均值向量求和，获得商户向量和；

将所述商户向量和包含在预处理的商户信息中。

7. 根据权利要求 1 所述的方法，其中所述多个数据集中的各个数据集还包括，用户在对应的第一时间段内的信用信息。

8. 根据权利要求 1 所述的方法，还包括，获取所述用户的属性特征信息，所述利用全连接神经网络处理所述第一输出结果和所述统计信息，获得第二输出结果，包括：将所述第一输出结果、所述统计信息和所述属性特征信息输入所述全连接神经网络，获得第二输出结果。

9. 根据权利要求 8 所述的方法，其中所述全连接神经网络为多层全连接神经网络，其中将所述第一输出结果、所述统计信息和所述属性特征信息输入所述全连接神经网络包括：将所述第一输出结果和所述属性特征信息输入所述多层全连接神经网络的第一层，将所述统计信息输入所述全连接神经网络的最后一层。

10. 根据权利要求 1-9 中任一项所述的方法，其中所述循环神经网络包括多层叠加的长短期记忆模型 LSTM。

11. 一种评估金融违约风险的装置，包括：

数据集获取单元，配置为获取数据集序列，所述数据集序列包括按照时间顺序排列的多个数据集，各数据集分别对应预设的第一时间段，所述多个数据集中的各个数据集包括，用户的与借贷行为相关的金额信息和商户信息；

第一处理单元，配置为利用循环神经网络处理所述数据集序列，获得第一输出结果；

统计信息获取单元，配置为获取与所述用户在第二时间段内的借贷行为相关的统计信息，所述第二时间段大于所述第一时间段；

第二处理单元，配置为利用全连接神经网络处理所述第一输出结果和所述统计信息，获得第二输出结果；

评估单元，配置为根据所述第二输出结果评估所述用户的金融违约风险。

12. 根据权利要求 11 所述的装置，其中所述商户信息包括，与所述用户的借贷行为相关的商户 ID。

13. 根据权利要求 12 所述的装置，其中与所述用户的借贷行为相关的商户 ID 包括以下中的一项或多项，用户申请借款的商户 ID，用户借款成功的商户 ID，用户借款失败的商户 ID，用户履约的商户 ID，用户违约的商户 ID。

14. 根据权利要求 12 所述的装置，还包括预处理单元，配置为对所述商户信息进行预处理，所述预处理包括：通过词嵌入模型将所述商户 ID 转化为 ID 向量。

15. 根据权利要求 14 所述的装置，其中所述预处理单元还配置为：

在同一数据集中，对所述用户的同类型借贷行为所针对的商户 ID 的 ID 向量求平均，获得各类型对应的均值向量；

将所述均值向量包含在预处理的商户信息中。

16. 根据权利要求 15 所述的装置，其中所述预处理单元还配置为：

对所述用户的不同类型的借贷行为中各个类型对应的所述均值向量求和，获得商户向量和；

将所述商户向量和包含在预处理的商户信息中。

17. 根据权利要求 11 所述的装置，其中所述多个数据集中的各个数据集还包括，用户在对应的第一时间段内的信用信息。

18. 根据权利要求 11 所述的装置，还包括，属性特征获取单元，配置为获取所述用户的属性特征信息，所述第二处理单元配置为：将所述第一输出结果、所述统计信息和所述属性特征信息输入所述全连接神经网络，获得第二输出结果。

19. 根据权利要求 18 所述的装置，其中所述第二处理单元配置为：将所述第一输出结果和所述属性特征信息输入所述多层全连接神经网络的第一层，将所述统计信息输入所述全连接神经网络的最后一层。

20. 根据权利要求 11-19 中任一项所述的装置，其中所述循环神经网络包括多层叠加的长短期记忆模型 LSTM。

21.一种计算机可读存储介质，其上存储有计算机程序，当所述计算机程序在计算机中执行时，令计算机执行权利要求 1-10 中任一项的所述的方法。

22. 一种计算设备，包括存储器和处理器，其特征在于，所述存储器中存储有可执行代码，所述处理器执行所述可执行代码时，实现权利要求 1-10 中任一项所述的方法。

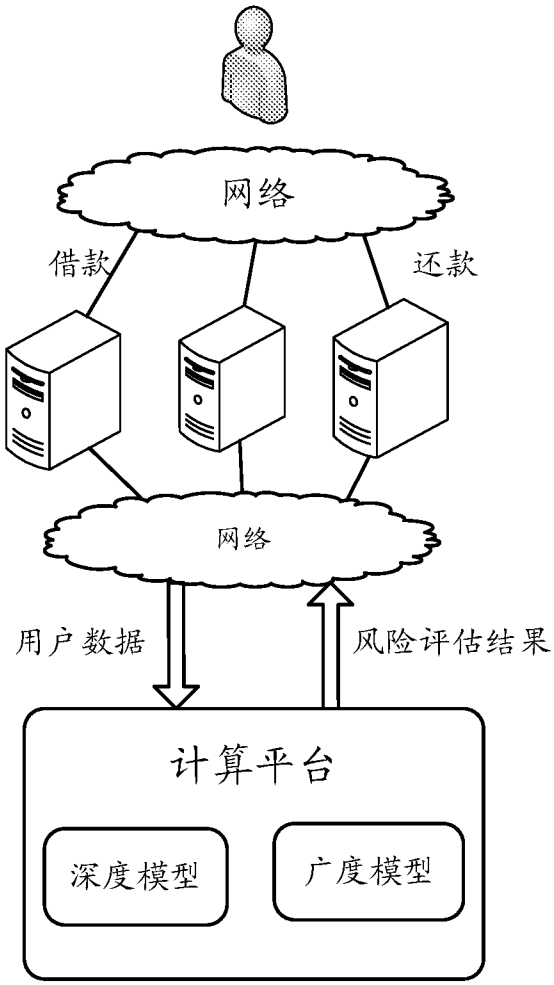


图 1

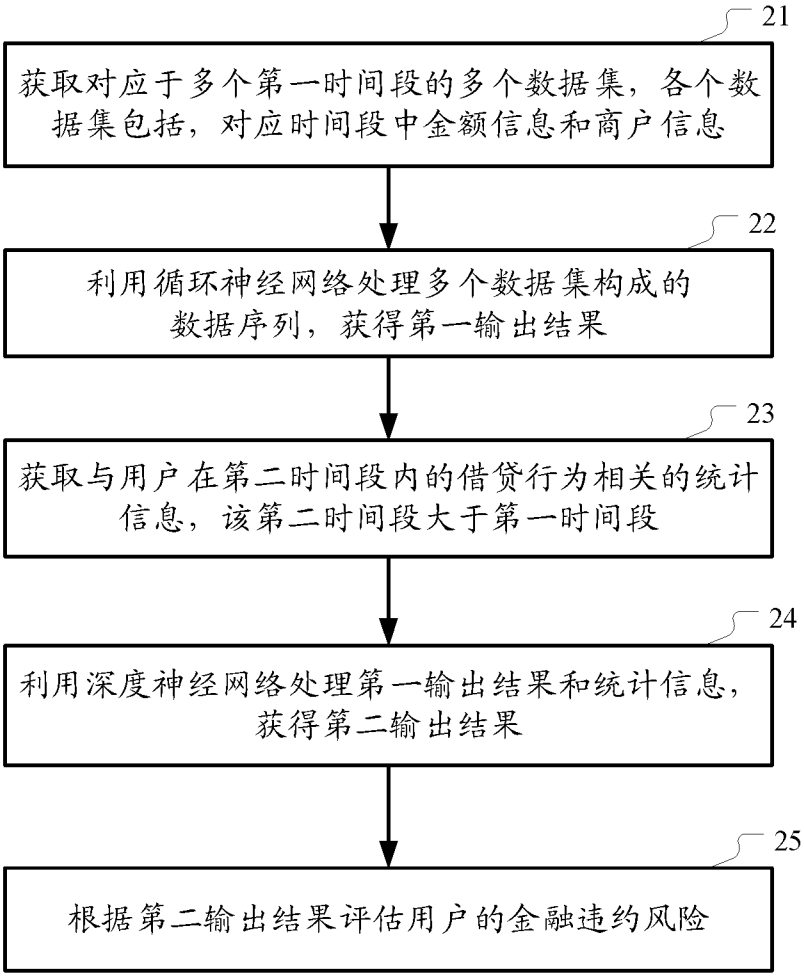

21
获取对应于多个第一时间段的多个数据集，各个数据集包括，对应时间段中金额信息和商户信息
22
利用循环神经网络处理多个数据集构成的数据序列，获得第一输出结果
23
获取与用户在第二时间段内的借贷行为相关的统计信息，该第二时间段大于第一时间段
24
利用深度神经网络处理第一输出结果和统计信息，获得第二输出结果
25
根据第二输出结果评估用户的金融违约风险

图 2

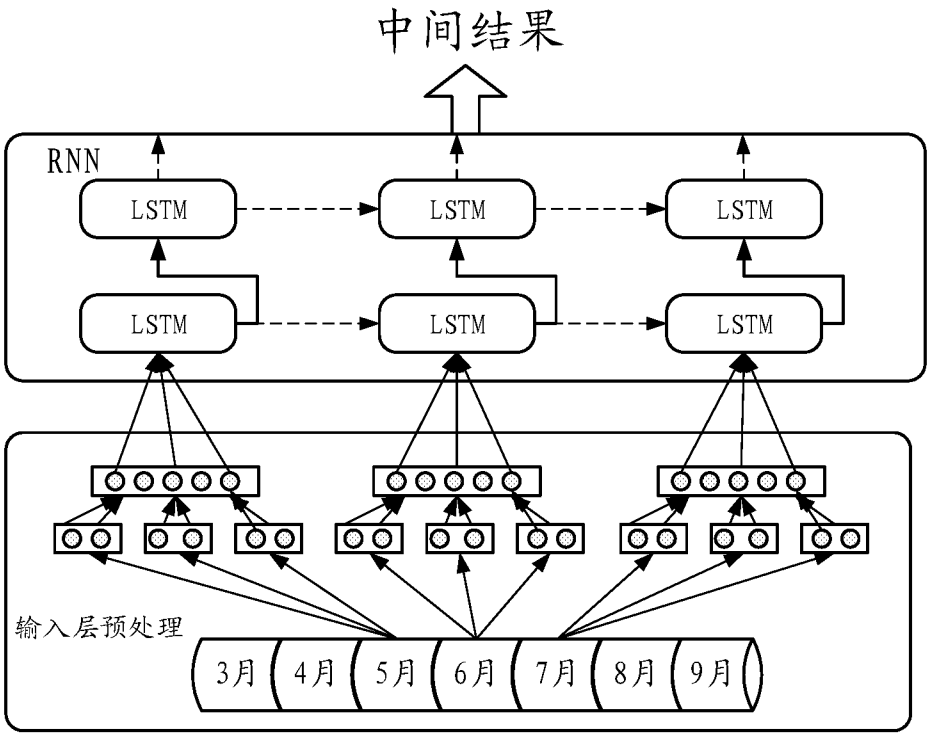

中间结果
RNN
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
输入层预处理
3月
4月
5月
6月
7月
8月
9月

图 3

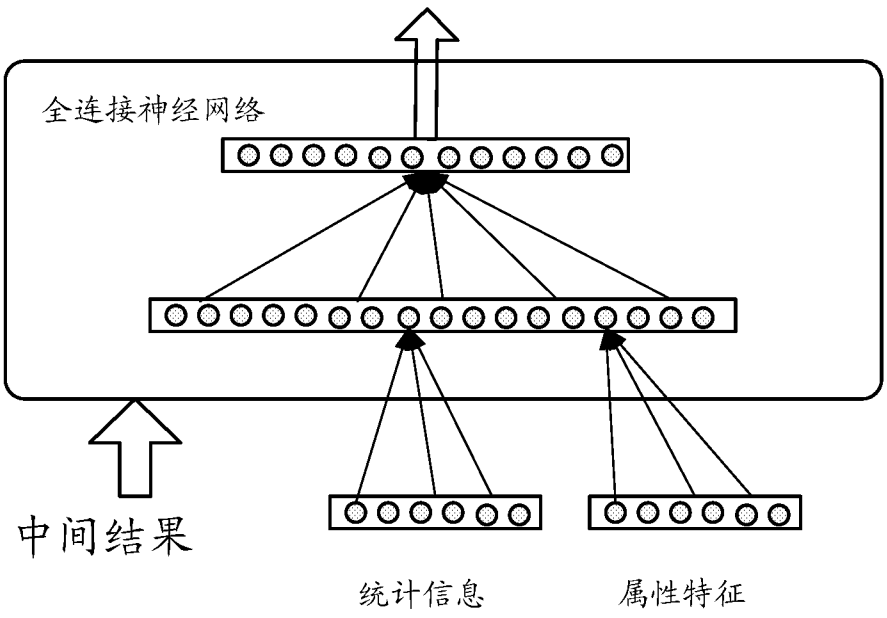

全连接神经网络
中间结果
统计信息
属性特征

图 4

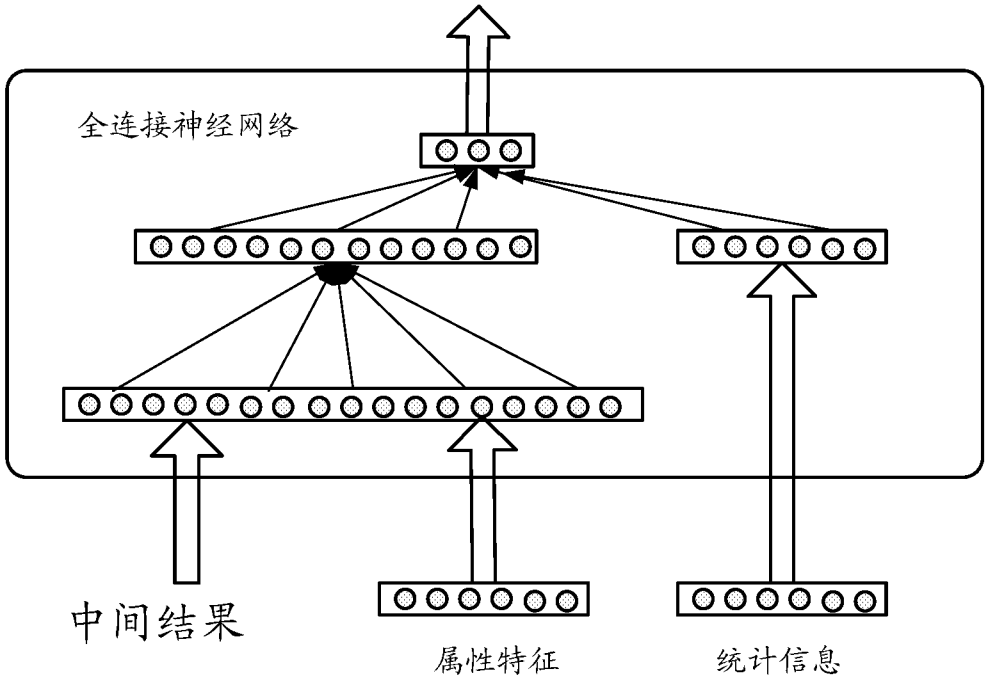

全连接神经网络
中间结果
属性特征
统计信息

图 5

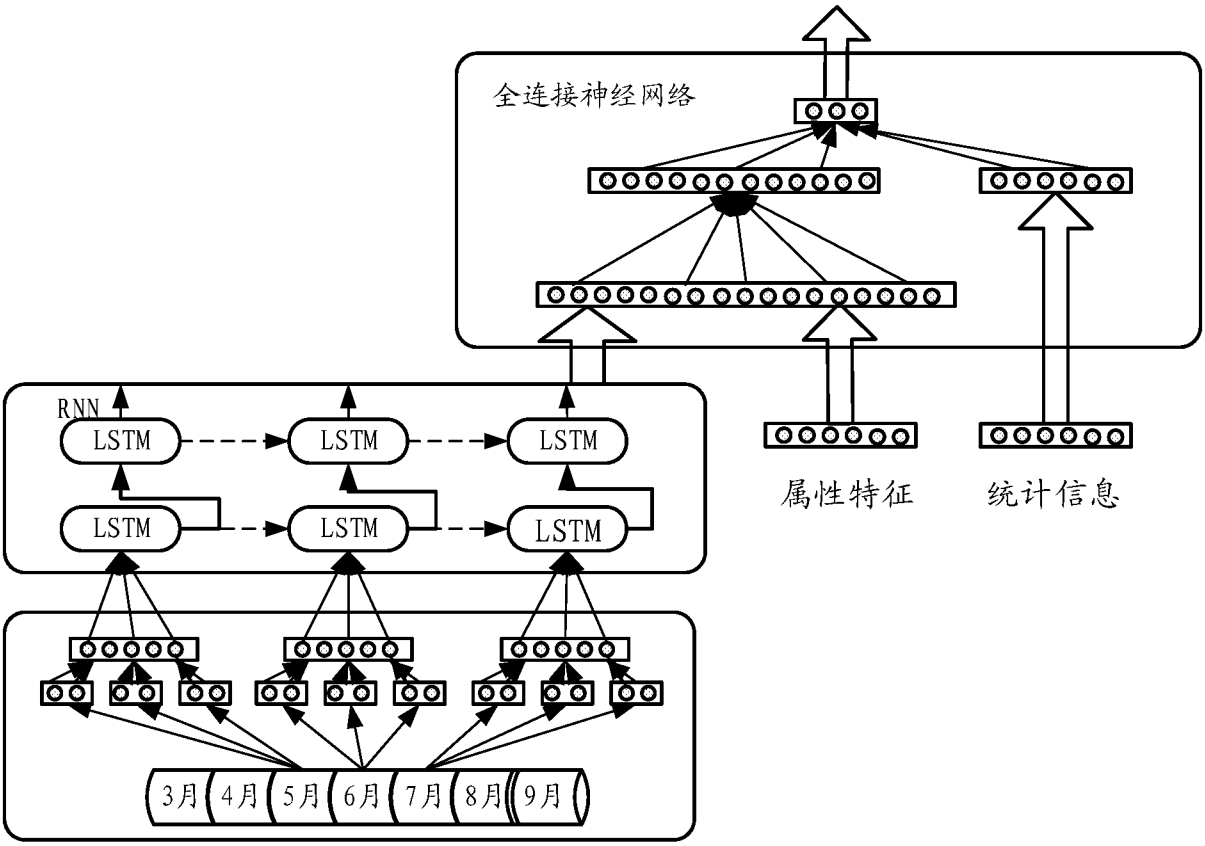

全连接神经网络
RNN
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
3月
4月
5月
6月
7月
8月
9月
属性特征
统计信息

图 6

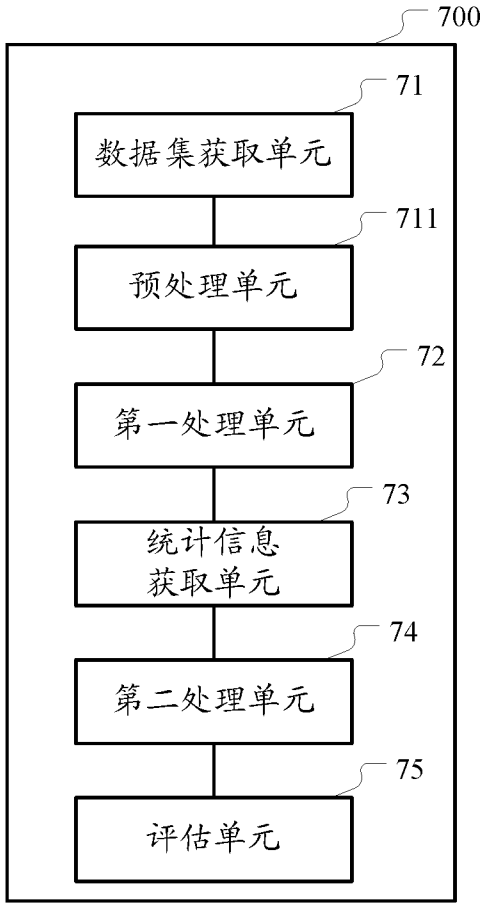

700
71
数据集获取单元
711
预处理单元
72
第一处理单元
73
统计信息
获取单元
74
第二处理单元
75
评估单元

图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/071088

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 40/02(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)

CNKI, CNPAT, WPI, EPODOC: 评估, 金融, 违约, 风险, 借贷, 短期, 长期, 神经网络, evaluate, insurance, contract, breach, risk, neural, network, short, long, period, borrow

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 108416663 A (ALIBABA GROUP HOLDING LIMITED) 17 August 2018 (2018-08-17) claims 1-22	1-22
A	CN 107424070 A (GUANGZHOU HUIRONGYI INTERNET FINANCE INFORMATION SERVICE CO., LTD.) 01 December 2017 (2017-12-01) claims 1-10	1-22
A	CN 107153977 A (ALIBABA GROUP HOLDING LIMITED) 12 September 2017 (2017-09-12) entire document	1-22
A	CN 107316198 A (ALIBABA GROUP HOLDING LIMITED) 03 November 2017 (2017-11-03) entire document	1-22

☐ 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 March 2019	**12 April 2019**

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

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/071088

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 108416663 A	17 August 2018	None	
CN 107424070 A	01 December 2017	None	
CN 107153977 A	12 September 2017	None	
CN 107316198 A	03 November 2017	None	

国际检索报告

国际申请号
PCT/CN2019/071088

A. 主题的分类

G06Q 40/02(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））

CNKI, CNPAT, WPI, EPODOC: 评估， 金融， 违约， 风险， 借贷， 短期， 长期， 神经网络， evaluate, insurance, contract, breach, risk, neural, network, short, long, period, borrow

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 108416663 A (阿里巴巴集团控股有限公司) 2018年 8月 17日 (2018-08-17) 权利要求1-22	1-22
A	CN 107424070 A (广州汇融易互联网金融信息服务有限公司) 2017年 12月 1日 (2017-12-01) 权利要求1-10	1-22
A	CN 107153977 A (阿里巴巴集团控股有限公司) 2017年 9月 12日 (2017-09-12) 全文	1-22
A	CN 107316198 A (阿里巴巴集团控股有限公司) 2017年 11月 3日 (2017-11-03) 全文	1-22

☐ 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 3月 22日	2019年 4月 12日

ISA/CN的名称和邮寄地址	受权官员
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	邹予婷
传真号 (86-10)62019451	电话号码 86-(10)-53961440

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/071088

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 108416663 A	2018年 8月 17日	无	
CN 107424070 A	2017年 12月 1日	无	
CN 107153977 A	2017年 9月 12日	无	
CN 107316198 A	2017年 11月 3日	无	