

(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
2018 年 9 月 20 日 (20.09.2018)

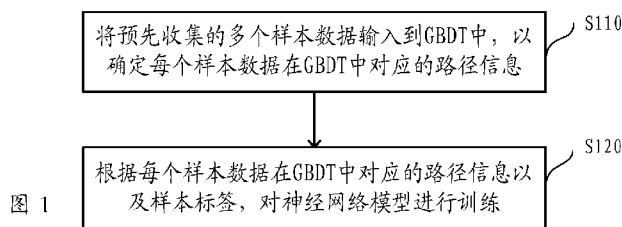


(10) 国际公布号
WO 2018/166457 A1

- (51) 国际专利分类号:
G06N 3/08 (2006.01) **G06K 9/62** (2006.01)
- (21) 国际申请号: PCT/CN2018/078906
- (22) 国际申请日: 2018 年 3 月 14 日 (14.03.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710153115.8 2017年3月15日 (15.03.2017) 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; 及
(71) 申请人 (仅对US): 李龙飞 (LI, Longfei) [CN/CN]; 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)。周俊 (ZHOU, Jun) [CN/CN]; 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)。李小龙 (LI, Xiaolong) [CN/CN]; 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: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (CN)。

(54) Title: NEURAL NETWORK MODEL TRAINING METHOD AND DEVICE, TRANSACTION BEHAVIOR RISK IDENTIFICATION METHOD AND DEVICE

(54) 发明名称: 神经网络模型训练、交易行为风险识别方法及装置



S110 INPUT A PLURALITY OF PIECES OF PRE-COLLECTED SAMPLE DATA INTO A GBDT, SO AS TO DETERMINE PATH INFORMATION IN THE GBDT CORRESPONDING TO EACH PIECE OF SAMPLE DATA
S120 ACCORDING TO THE PATH INFORMATION IN THE GBDT CORRESPONDING TO EACH PIECE OF SAMPLE DATA AND A SAMPLE LABEL, TRAIN A NEURAL NETWORK MODEL

(57) Abstract: A neural network model training method and device, and a transaction behavior risk identification method and device. The neural network model training method comprises: inputting a plurality of pieces of pre-collected sample data into a gradient boosting decision tree (GBDT), so as to determine path information in the GBDT corresponding to each piece of sample data (S110); and according to the path information in the GBDT corresponding to each piece of sample data and a sample label, training a neural network model (S120). The method firstly determines the path information according to the GBDT, and then trains the neural network models according to the path information and the sample label. It is known from features of the GBDT itself that a certain piece of path information generally comprises multi-dimensional information of the sample data. Thus, the invention can improve the efficiency of training the neural network model.

(57) 摘要: 一种神经网络模型训练、交易行为风险识别方法及装置。所述神经网络模型训练方法包括: 将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中, 以确定每个样本数据在 GBDT 中对应的路径信息 (S110); 根据每个样本数据在 GBDT 中对应的路径信息以及样本标签, 对神经网络模型进行训练 (S120)。所述方法首先根据 GBDT 来确定路径信息, 之后根据路径信息以及样本标签来训练神经网络模型, 而根据 GBDT 本身的特点可知, 其一条路径信息通常会包含样本数据中多个维度的信息, 由此, 可以提高神经网络模型训练的效率。

(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))。

神经网络模型训练、交易行为风险识别方法及装置

技术领域

本申请涉及计算机技术领域，尤其涉及一种神经网络模型训练、交易行为风险识别方法及装置。

背景技术

传统技术中，在搜集到样本数据之后，直接根据样本数据以及样本数据的样本标签，来训练神经网络模型。然而，上述搜集的样本数据通常会包括多个维度的信息，这会导致神经网络模型训练的效率比较低。

发明内容

本申请描述了一种神经网络模型训练、交易行为风险识别方法及装置，可以提高神经网络模型训练的效率。

第一方面，提供了一种神经网络模型训练方法，包括：

将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中，以确定每个样本数据在所述 GBDT 中对应的路径信息；所述每个样本数据具有对应的样本标签；

根据所述每个样本数据在所述 GBDT 中对应的路径信息以及样本标签，对神经网络模型进行训练。

第二方面，提供了一种交易行为风险识别方法，包括：

获取用户的交易行为数据；

将所述交易行为数据输入到梯度提升决策树 GBDT 中，以确定所述交易行为数据在所述 GBDT 中对应的路径信息；

将所述路径信息输入到神经网络模型中；

输出交易行为风险识别结果。

第三方面，提供了一种神经网络模型训练装置，包括：

确定单元，用于将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中，以确定每个样本数据在所述 GBDT 中对应的路径信息；所述每个样本数据具有对应的样本标签；

5 训练单元，用于根据所述确定单元确定的所述每个样本数据在所述 GBDT 中对应的路径信息以及样本标签，对神经网络模型进行训练。

第四方面，提供了一种交易行为风险识别装置，包括：

获取单元，用于获取用户的交易行为数据；

10 确定单元，用于将所述获取单元获取的所述交易行为数据输入到梯度提升决策树 GBDT 中，以确定所述交易行为数据在所述 GBDT 中对应的路径信息；

输入单元，用于将所述确定单元确定的所述路径信息输入到神经网络模型中；

输出单元，用于输出交易行为风险识别结果。

15 本申请提供的神经网络模型训练、交易行为风险识别方法及装置，将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中，以确定每个样本数据在 GBDT 中对应的路径信息。根据每个样本数据在 GBDT 中对应的路径信息以及样本标签，对神经网络模型进行训练。也即本申请首先根据 GBDT 来确定路径信息，之后根据路径信息以及样本标签来训练神经网络模型，而根据 GBDT 本身的特点可知，其一条路径信息通常会包含样本数据中多个维度的信息，由此，可以提高神经网络模型训练的效率。

20

附图说明

25 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。

图 1 为本申请一种实施例提供的神经网络模型训练方法流程图；

图 2 为本申请提供的决策树的示意图；

图 3 为本申请提供的训练 DNN 的过程示意图；

5 图 4 为本申请提供的交易行为风险识别方法示意图；

图 5 为本申请一种实施例提供的神经网络模型训练装置示意图；

图 6 为本申请另一种实施例提供的交易行为风险识别装置示意图。

具体实施方式

10 下面结合附图，对本申请的实施例进行描述。

本申请实施例提供的神经网络模型训练方法适用于对深度神经网络 (Deep Neural Network, DNN) 或者人工神经网络 (Artificial Neural Network, ANN) 等神经网络模型进行训练的场景。训练好的神经网络模型可以用于进行模式识别以及分类的场景，如，可以用于对交易行为进行风险识别。

15 图 1 为本申请一种实施例提供的神经网络模型训练方法流程图。所述方法的执行主体可以为具有处理能力的设备：服务器或者系统或者装置，如图 1 所示，所述方法具体包括：

步骤 110，将预先收集的多个样本数据输入到梯度提升决策树 (Gradient Boosting Decision Tree, GBDT) 中，以确定每个样本数据在 GBDT 中对应的
20 路径信息。

在执行步骤 110 之前，可以先训练好 GBDT 模型。具体的训练过程后续进行说明。

步骤 110 中，以训练的神经网络模型用于交易行为风险识别的场景为例来说，上述样本数据可以是指用户的交易行为数据。具体地，可以从支付
25 宝系统的后台数据库中搜集样本数据。此处，样本数据可以归属于如下五个类别的用户数据：1) 用户的历史行为信息。如，a，若干天（如，180 天）内

用户来电次数；b，最后一次登录城市；c，最后一次登录距今时间；d，若干天（如，90 天）内登录次数等。2）用户的交易信息。如，a，若干天（如，90 天）平均支付金额；b，若干天（如，180 天）内支付天数；c，若干天（如，180 天）内支付金额；d，最后一次支付距今时间等。3）用户的基本信息。如，

5 a，用户是否单身；b，用户是否装修；c，用户是否已婚；d，用户年龄；e，用户注册时长；f，用户教育水平等。4）用户的远程过程调用（Remote Procedure Call，RPC）行为信息。此处的 RPC 行为信息是指用户在使用客户端的时候，客户端与服务器之间的 RPC 调用。在一种实现方式中，可以搜集每个用户在最近一个给定时间窗口的这些操作。如，可以搜集用户近 2 天访问的 RPC 接

10 口的次数变量。5）用户的统一资源定位器（Uniform Resource Locator，URL）地址信息。

对上述收集的多个样本数据，如果某样本数据与当前用户不相关或者该样本数据能给用户带来负面影响的，则将该样本数据分类为正样本数据。如，某一交易行为由非用户本人操作的或者对用户的账户带来一定的损失且报案

15 的，则将该交易行为数据标记为正样本数据。否则，如果某样本数据为用户本人正常的交易行为数据，则将该样本数据标记为负样本数据。

需要说明的是，通常负样本数据比较容易搜集。如，可以很容易从支付宝系统的后台数据库中搜集到正常支付行为的数据。所以，样本数据集合中负样本数据会占绝大多数的比重，如，大于 99.999%。然而，当负样本数据的

20 比重比较高时，训练的神经网络模型往往会有偏差，如，只能识别安全的交易行为，而不能识别有风险的交易行为，这影响了交易行为风险识别的准确性。

为了能提升交易行为风险识别的准确性，可以对样本数据进行预处理。在一种实现方式中，可以对正样本数据进行升采样处理；和/或，对负样本数据

25 数据进行降采样处理。其中，对正样本数据进行升采样处理可以包括：通过复制等方式增加正样本数据的数量。对负样本数据进行降采样处理可以包括：

通过删除等方式减小负样本数据的数量。在一个例子中，可以将正样本数据与负样本数据的比例调整为 1:300。

还需要说明的是，对上述预处理后的样本数据，还可以为正、负样本数据添加对应的样本标签。具体地，为正样本数据添加正样本标签，为负样本数据添加负样本标签。

步骤 110 中，将预先收集的多个样本数据输入到 GBDT 中具体可以包括：针对每个样本数据，可以先根据该样本数据，确定多个特征对应的特征值。之后，将特征的特征值输入到 GBDT 的决策树中。

此处的特征可以归属于多个类别。在一种实现方式中，上述特征中的部分特征可以采用现有交易行为风险识别模型在线沉淀的模型变量，该模型变量归属于如下三个类别：1) 用户的历史行为信息。2) 用户的交易信息。3) 用户的基本信息。

然而，上述模型变量需要根据业务数据来确定，而业务数据通常来自不同业务部门，其采集和整理需要一定的时间，所以仅通过上述模型变量不能得到用户最新的状态，从而也不能对用户最新的交易行为进行风险识别。为解决该问题，本申请中增加了归属于用户的 RPC 行为信息的特征和归属于用户的 URL 地址信息的特征。

综上，本申请的特征可以为归属于如下五个类别的特征：1) 用户的历史行为信息。2) 用户的交易信息。3) 用户的基本信息。4) 用户的 RPC 行为信息。5) 用户的 URL 地址信息。其中，每个类别如上所述，在此不复赘述。

对上述设定的特征，在根据具体的样本数据，确定其对应的特征值之后，就可以将特征值输入到 GBDT 中。此处的 GBDT 可以由多棵决策树组成，每棵决策树包括多个节点，每个节点与一个特征相对应。以一棵决策树为例来说，该决策树可以如图 2 所示，图 2 中，节点 1、节点 2 和节点 3 分别与特征：“用户性别是否是男”、“用户年龄大于 20 岁”以及“交易金额是否超过 1000 元”相对应。在将特征的特征值输入决策树之后 就可以在决策树中确定出多条

路径信息。如，假设样本数据包含用户性别是男，用户年龄大于 20 岁，交易金额超过 1000 元时，确定出的路径信息可以如图 2 中的粗线所示。

作为示例性说明，图 2 中只是展示了一条路径信息，实际上样本数据输入 GBDT 时，可以确定出多条路径信息，本申请在此不复赘述。

5 需要说明的是，本申请中，在将该特征值输入到 GBDT 之前，还可以将该特征值表示为 one-hot 形式的特征向量。在还确定特征值对应的特征向量的情况下，上述将特征值输入到 GBDT 中可以替换为：将特征值对应的特征向量输入到决策树中，以确定相应的路径信息。其中，确定特征值的特征向量的过程可以举例如下：

10 以特征为“用户性别”为例来说，如果用户性别为男，也即特征的特征值为“男”，则该特征值对应的特征向量可以为： $[0\ 1]$ 。如果用户性别为女，也即特征的特征值为“女”，则该特征值对应的特征向量可以为： $[1\ 0]$ 。

再以特征为用户的 RPC 行为信息为例来说，其特征值对应的特征向量的确定可以通过如下两种方式来实现：第一种实现方式中，首先设定规则：出现
15 出现过则标识为 1，否则为 0。具体地，假设预设的 RPC 行为信息为：a, b 和 c。而某个样本数据包含用户两天内的 RPC 行为信息为：a, a 和 b，也即特征值为：a, a 和 b。则对应的特征向量可以为： $[1\ 1\ 0]$ 。在另一种实现方式中，可以设定规则：统计预设的 RPC 行为信息的频次，然后归一化。具体地，假设
20 预设的 RPC 行为信息为：a, b 和 c。而某个样本数据包含用户两天内的 RPC 行为信息为：a, a, b, b 和 c，也即特征值为：a, a, b, b 和 c。则对应的特征向量可以为：2, 2 和 1。因为需要归一化，所以最终的特征向量为： $[0.4\ 0.4\ 0.2]$ 。

需要说明的是，上述将特征值表示为特征向量属于传统常规技术，在此不复赘述。

25 需要说明的是，为了提升神经网络模型的准确性，本申请中设定了比较多的特征，从而会确定多个特征值。对于越来越多的特征值，其处理往往需

要花费很多的时间，受限于同时观察的特征值的个数，人很难对多个特征值之间的关系进行深入的分析，并手工生成新的特征值。而本申请通过将样本数据输入 GBDT，来得到路径信息，该路径信息由于包含了多个特征值。从而可以大大减小特征值的数量，由此可以显著地减少人工的工作。

5 步骤 120，根据每个样本数据对应的路径信息以及样本标签，对神经网络模型进行训练。

 此处的神经网络模型可以包括 DNN 或者 ANN 等。其中，DNN 最近几年发展迅速，相比传统使用的浅层模型（如，逻辑回归（Logistic Regression，LR），随机森林(Random forest，RF)），DNN 有着其特有的先进性：模型表
10 达能力强大，适合大数据和分布式训练。因此，本说明书中，以训练 DNN 为例进行说明。

 本申请中，DNN 的训练过程可以如图 3 所示，图 3 中，DNN 的输入层用于输入 GBDT 中的各条路径信息，而输出层即可输出第一预测结果。可以理解的是，针对每个样本数据，即在将该样本数据对应的路径信息输入到 DNN
15 之后，DNN 都会输出相应的第一预测结果。对样本集合中的多个样本数据，若第一预测结果与样本数据的样本标签相符合的概率达到预设阈值，此处的预设阈值可以根据经验值设定，则可以认为已经得到了优化的 DNN。

 可以理解的是，随着路径信息的个数的不同，图 3 中 DNN 的层数是可以改动的。

20 通过实验发明，本申请训练得到的神经网络模型会比其它模型（LR 或者 RF）的效果都好。同时特征处理的时间大大的减少了，整体建模流程变快了很多。

 以下对如何训练 GBDT 模型进行说明：

 在根据每个样本数据，确定多个特征对应的特征值之后，可以将多个特征对应的特征值输入 GBDT 的各个决策树中。之后将各个决策树的结论累加
25 起来以确定第二预测结果。可以理解的是 针对每个样本数据，GBDT 模型

都会输出相应的第二预测结果。对样本集合中的多个样本数据，若第二预测结果与样本数据的样本标签相符合的概率达到预设阈值，此处的预设阈值可以根据经验值设定，则可以认为已经得到了优化的 GBDT 模型。而若第二预测结果与样本数据的样本标签相符合的概率未达到预设阈值，则可以通过调整决策树的数目、决策树的深度以及正则化项（用于表示特征）来继续执行上述输入和输出的操作，直至达到预设阈值为止。

综上，本申请具有如下几方面的优点：

1) 由于本申请的特征包括了类别为用户 RPC 行为信息的特征，因此本申请训练的神经网络模型能够满足时效性要求，也即能够识别用户最新的交易行为。

2) 本申请训练的神经网络模型的准确性比传统的浅层模型高。

3) 通过将样本数据输入 GBDT，获得了路径信息。而一条路径信息由多个特征值组合而成，也即一条路径信息包含了样本数据的多个维度的信息，由此，可以极大地减小 DNN 输入层输入的数据量，从而可以提高神经网络模型训练的效率。

需要说明的是，在通过图 1 所示的各步骤训练得到神经网络模型之后，就可以将该神经网络模型部署到线上，并对用户的交易行为进行风险识别了。

图 4 为本申请提供的交易行为风险识别方法的过程示意图。如图 4 所示，该方法可以包括：

步骤 410，获取用户的交易行为数据。

此处的交易行为数据与上述样本数据的定义相同，在此不复赘述。

步骤 420，将交易行为数据输入到梯度提升决策树 GBDT 中，以确定交易行为数据在 GBDT 中对应的路径信息。

上述 GBDT 由多棵决策树组成，每棵决策树包括多个节点，每个节点与一个特征相对应。步骤 420 中将交易行为数据输入到梯度提升决策树 GBDT 中，以确定交易行为数据在 GBDT 中对应的路径信息的步骤具体可以包括：

根据交易行为数据，确定多个特征对应的特征值；根据特征值，在决策树中确定路径信息。其中，确定路径信息的过程可以参照图 2，在此不复赘述。

步骤 430，将路径信息输入到神经网络模型中。

即将步骤 420 中确定的路径信息输入 DNN 的输入层中。

5 步骤 440，输出交易行为风险识别结果。

具体地，由 DNN 的输出层输出交易行为风险识别结果。此处，如果识别结果为风险的交易行为，则可以发起报警。在支付场景下，若识别结果为风险的支付行为，则可以冻结该用户账户以防止财产流失。与上述神经网络模型训练方法对应地，本申请实施例还提供的一种神经网络模型训练装置，如

10 图 5 所示，该装置包括：

确定单元 501，用于将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中，以确定每个样本数据在 GBDT 中对应的路径信息。

此处，每个样本数据具有对应的样本标签。

15 训练单元 502，用于根据确定单元 501 确定的每个样本数据在 GBDT 中对应的路径信息以及样本标签，对神经网络模型进行训练。

可选地，GBDT 由多棵决策树组成，每棵决策树包括多个节点，每个节点与一个特征相对应。

确定单元 501 具体用于：

20 对多个样本数据中的每个样本数据，根据样本数据，确定多个特征对应的特征值。

此处，特征可以包括：用户的远程过程调用 RPC 行为信息和/或用户的统一资源定位器 URL 地址信息。

根据特征值，在决策树中确定路径信息。

25 可选地，样本标签可以包括：正样本标签和负样本标签。上述装置还可以包括：

处理单元 503，用于对样本标签为正样本标签的样本数据进行升采样处

理；和/或，

对样本标签为负样本标签的样本数据进行降采样处理。

本申请实施例装置的各功能模块的功能，可以通过上述方法实施例的各步骤来实现，因此，本申请提供的装置的具体工作过程，在此不复赘述。

5 本申请提供的神经网络模型训练装置，确定单元 501 将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中，以确定每个样本数据在 GBDT 中对应的路径信息。训练单元 502 根据每个样本数据在 GBDT 中对应的路径信息以及样本标签，对神经网络模型进行训练。由此，可以提高神经网络模型训练的效率。

10 与上述交易行为风险识别方法对应地，本申请实施例还提供的一种交易行为风险识别装置，如图 6 所示，该装置包括：

获取单元 601，用于获取用户的交易行为数据。

确定单元 602，用于将获取单元 601 获取的交易行为数据输入到梯度提升决策树 GBDT 中，以确定交易行为数据在 GBDT 中对应的路径信息。

15 输入单元 603，用于将确定单元 602 确定的路径信息输入到神经网络模型中。

输出单元 604，用于输出交易行为风险识别结果。

可选地，GBDT 由多棵决策树组成，每棵决策树包括多个节点，每个节点与一个特征相对应；

20 确定单元 602 具体用于：

根据交易行为数据，确定多个特征对应的特征值。

根据特征值，在决策树中确定路径信息。

其中，特征可以包括：用户的远程过程调用 RPC 行为信息和/或用户的统一资源定位器 URL 地址信息。

25 本申请实施例装置的各功能模块的功能，可以通过上述方法实施例的各步骤来实现，因此，本申请提供的装置的具体工作过程，在此不复赘述。

本申请提供的交易行为风险识别装置，可以提高交易行为风险识别的效率和准确性。

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。

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、一种神经网络模型训练方法，其特征在于，包括：

将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中，以确定每个样本数据在所述 GBDT 中对应的路径信息；所述每个样本数据具有对应的
5 样本标签；

根据所述每个样本数据在所述 GBDT 中对应的路径信息以及样本标签，对神经网络模型进行训练。

2、根据权利要求 1 所述的方法，其特征在于，所述 GBDT 由多棵决策树组成，每棵决策树包括多个节点，每个节点与一个特征相对应；

10 所述将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中，以确定每个样本数据在 GBDT 中对应的路径信息，包括：

对所述多个样本数据中的每个样本数据，根据所述样本数据，确定多个特征对应的特征值；

根据所述特征值，在所述决策树中确定所述路径信息。

15 3、根据权利要求 1 或 2 所述的方法，其特征在于，所述样本标签包括：正样本标签和负样本标签；

在所述将预先收集的多个样本数据输入到梯度提升决策树 GBDT 之前，还包括：

对样本标签为正样本标签的样本数据进行升采样处理；和/或，

20 对样本标签为负样本标签的样本数据进行降采样处理。

4、根据权利要求 2 所述的方法，其特征在于，所述特征包括：

用户的远程过程调用 RPC 行为信息和/或用户的统一资源定位器 URL 地址信息。

5、一种交易行为风险识别方法，其特征在于，包括：

25 获取用户的交易行为数据；

将所述交易行为数据输入到梯度提升决策树 GBDT 中，以确定所述交易

行为数据在所述 GBDT 中对应的路径信息；

将所述路径信息输入到神经网络模型中；

输出交易行为风险识别结果。

- 5 6、根据权利要求 5 所述的方法，其特征在于，所述 GBDT 由多棵决策树组成，每棵决策树包括多个节点，每个节点与一个特征相对应；

所述将所述交易行为数据输入到梯度提升决策树 GBDT 中，以确定所述交易行为数据在所述 GBDT 中对应的路径信息，包括：

根据所述交易行为数据，确定多个特征对应的特征值；

根据所述特征值，在所述决策树中确定所述路径信息。

- 10 7、根据权利要求 6 所述的方法，其特征在于，所述特征包括：

用户的远程过程调用 RPC 行为信息和/或用户的统一资源定位器 URL 地址信息。

8、一种神经网络模型训练装置，其特征在于，包括：

- 15 确定单元，用于将预先收集的多个样本数据输入到梯度提升决策树 GBDT 中，以确定每个样本数据在所述 GBDT 中对应的路径信息；所述每个样本数据具有对应的样本标签；

训练单元，用于根据所述确定单元确定的所述每个样本数据在所述 GBDT 中对应的路径信息以及样本标签，对神经网络模型进行训练。

- 20 9、根据权利要求 8 所述的装置，其特征在于，所述 GBDT 由多棵决策树组成，每棵决策树包括多个节点，每个节点与一个特征相对应；

所述确定单元具体用于：

对所述多个样本数据中的每个样本数据，根据所述样本数据，确定多个特征对应的特征值；

根据所述特征值，在所述决策树中确定所述路径信息。

- 25 10、根据权利要求 8 或 9 所述的装置，其特征在于，所述样本标签包括：正样本标签和负样本标签；所述装置还包括：

处理单元，用于对样本标签为正样本标签的样本数据进行升采样处理；
和/或，

对样本标签为负样本标签的样本数据进行降采样处理。

11、根据权利要求 9 所述的装置，其特征在于，所述特征包括：

5 用户的远程过程调用 RPC 行为信息和/或用户的统一资源定位器 URL 地址信息。

12、一种交易行为风险识别装置，其特征在于，包括：

获取单元，用于获取用户的交易行为数据；

10 确定单元，用于将所述获取单元获取的所述交易行为数据输入到梯度提升决策树 GBDT 中，以确定所述交易行为数据在所述 GBDT 中对应的路径信息；

输入单元，用于将所述确定单元确定的所述路径信息输入到神经网络模型中；

输出单元，用于输出交易行为风险识别结果。

15 13、根据权利要求 12 所述的装置，其特征在于，所述 GBDT 由多棵决策树组成，每棵决策树包括多个节点，每个节点与一个特征相对应；

所述确定单元具体用于：

根据所述交易行为数据，确定多个特征对应的特征值；

根据所述特征值，在所述决策树中确定所述路径信息。

20 14、根据权利要求 13 所述的装置，其特征在于，所述特征包括：

用户的远程过程调用 RPC 行为信息和/或用户的统一资源定位器 URL 地址信息。

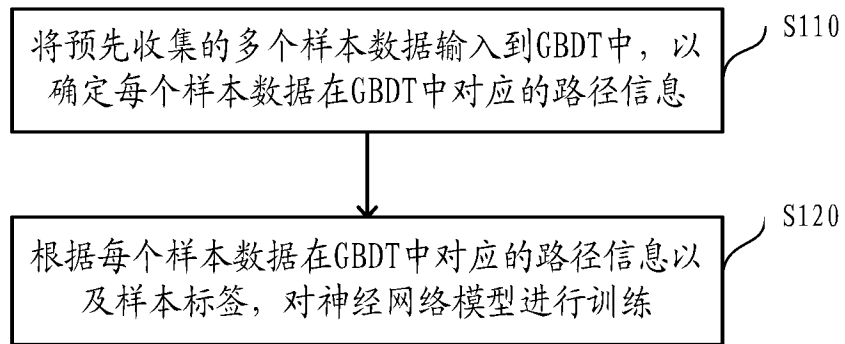


图 1

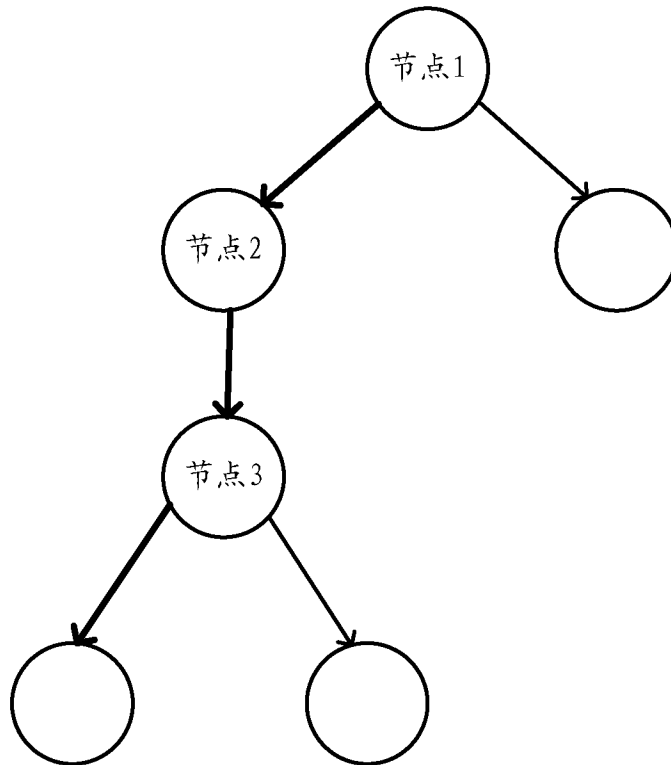


图 2

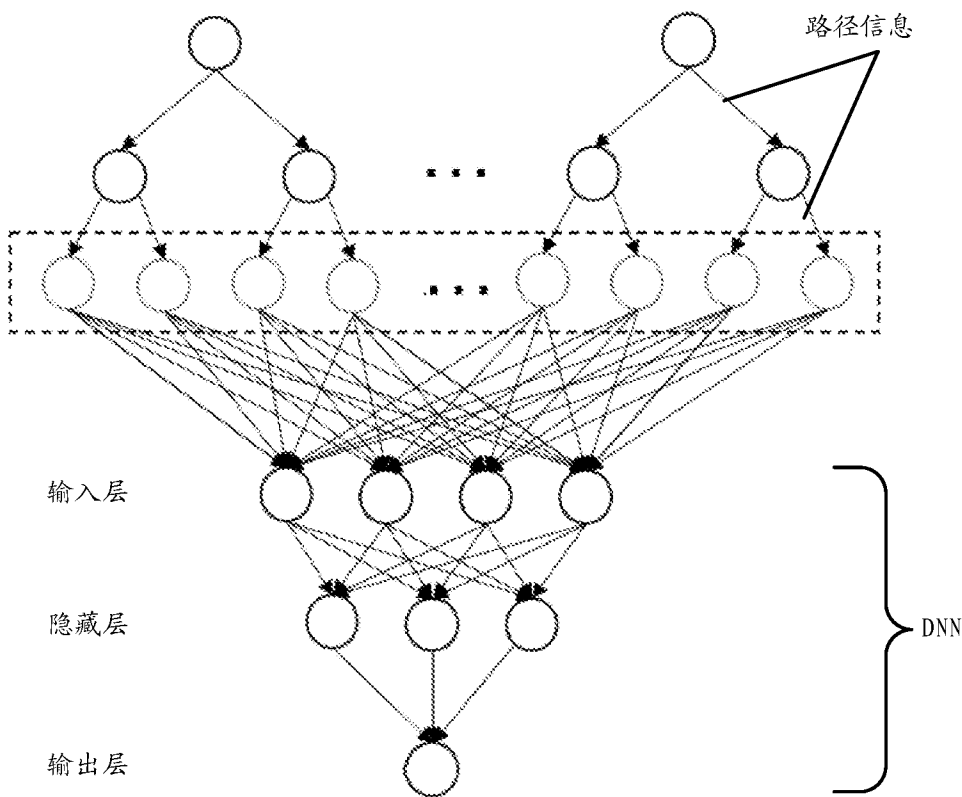


图 3

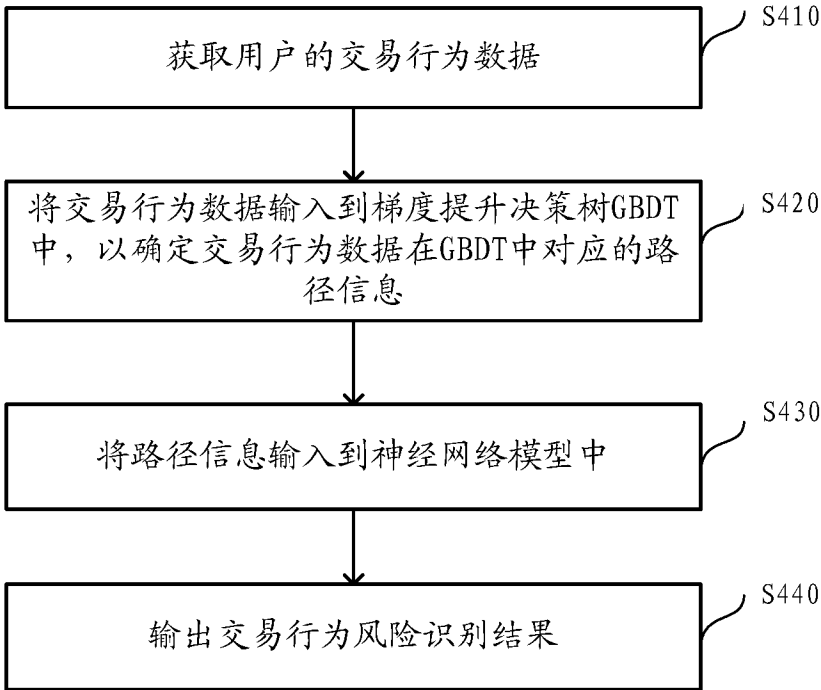


图 4

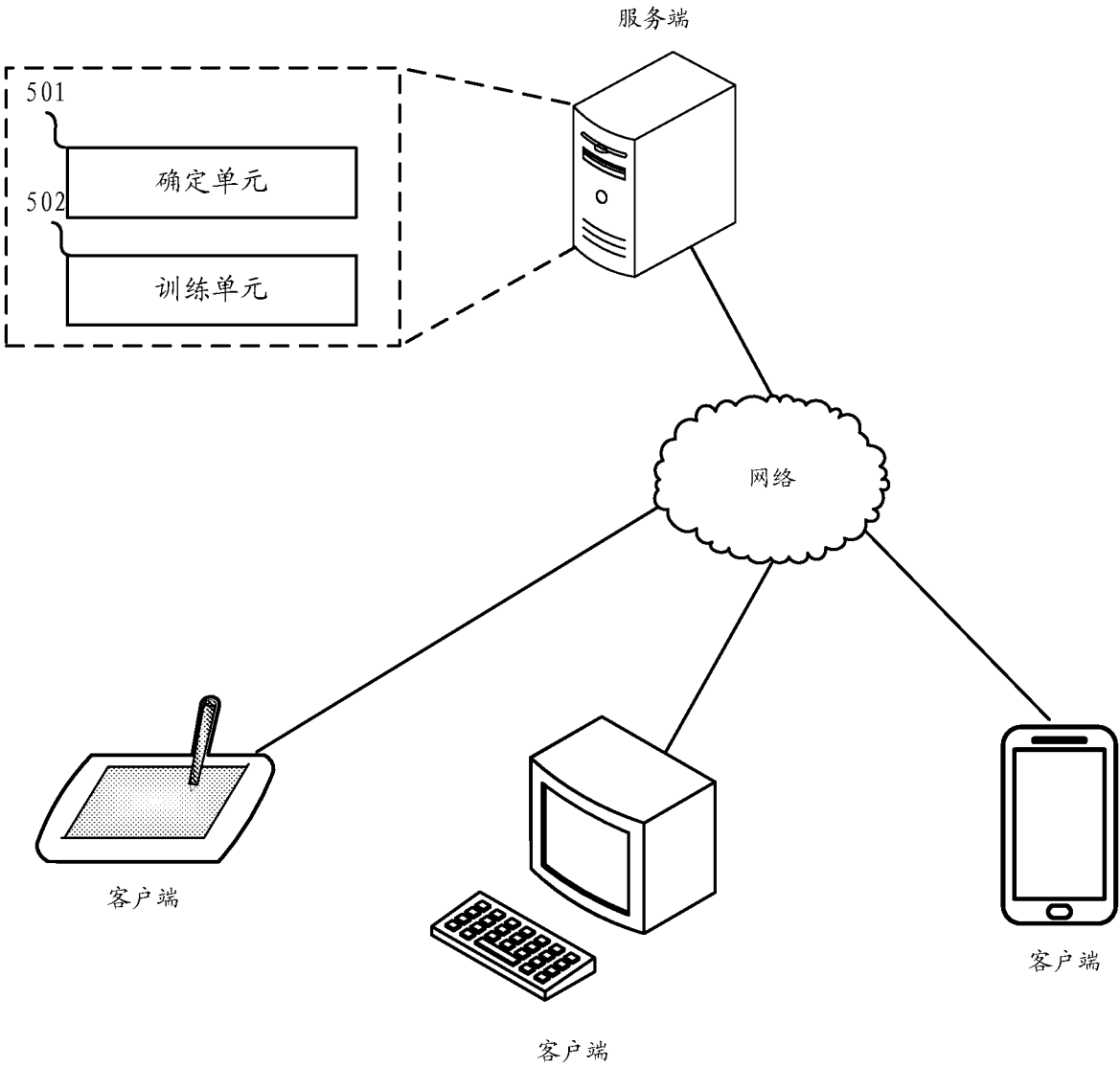


图 5

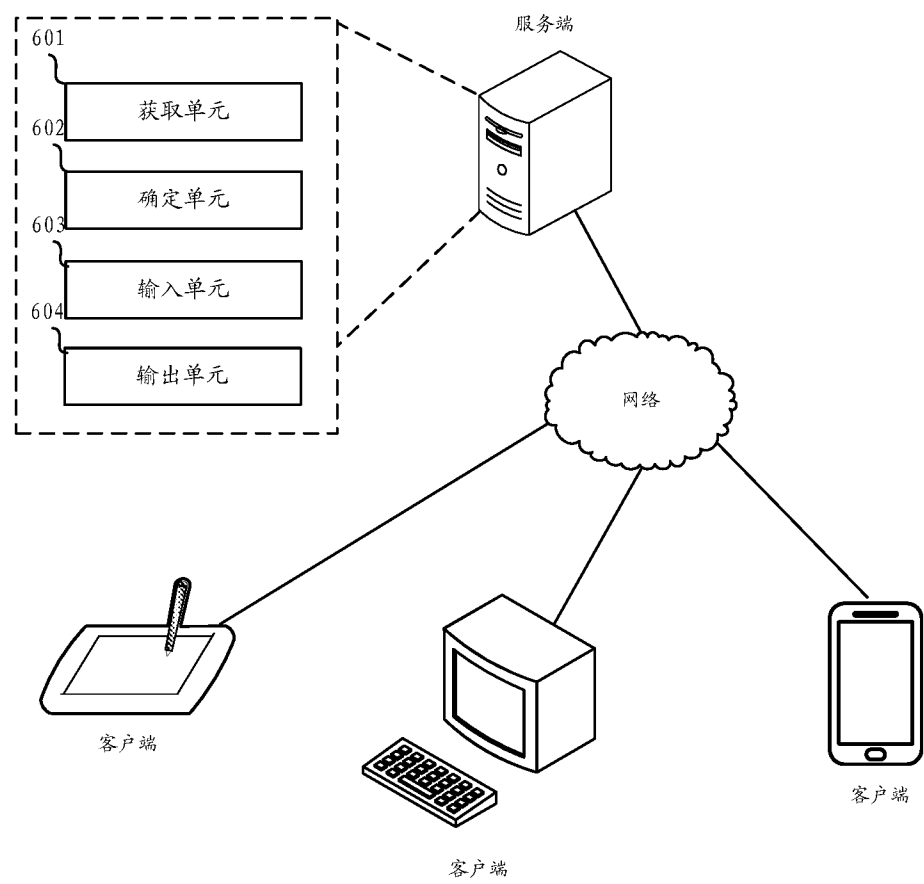


图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/078906

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/08 (2006.01) 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)

G06N, 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, WPI, EPODOC, CNKI: 神经网络, 模型, 训练, 样本, 梯度, 提升, 决策树, 分类树, 路径, 标签, 节点, 特征, 正, 负, 升采样, 降采样, 远程过程调用, RPC, 统一资源定位器, URL, 地址, neural network?, DNN, ANN, model?, train+, sample?, gradient, boost+, decision tree?, classification tree?, path?, label?, node?? characteristic?, feature?, positive, negative, upsampl+, downsampl+, remote procedure call, uniform resource locator?, address??

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 105844501 A (SHANGHAI YIBAO HEALTH MANAGEMENT CO., LTD.) 10 August 2016 (10.08.2016), description, paragraphs [0041]-[0069]	1-14
Y	CN 105279691 A (CHINA UNIONPAY CO., LTD.) 27 January 2016 (27.01.2016), description, paragraphs [0032]-[0065], and figures 1 and 2	1-14
A	CN 106296195 A (ALIBABA GROUP HOLDING LTD.) 04 January 2017 (04.01.2017), entire document	1-14
A	CN 102890803 A (ALIBABA GROUP HOLDING LTD.) 23 January 2013 (23.01.2013), entire document	1-14

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“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	“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
“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)	“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
“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	“&”document member of the same patent family

Date of the actual completion of the international search 01 June 2018	Date of mailing of the international search report 13 June 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer WEI, Jingyao Telephone No. (86-10) 53961340

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/078906

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105844501 A	10 August 2016	None	
CN 105279691 A	27 January 2016	None	
CN 106296195 A	04 January 2017	None	
CN 102890803 A	23 January 2013	HK 1176725 A1	26 August 2016

国际检索报告

国际申请号

PCT/CN2018/078906

A. 主题的分类 G06N 3/08(2006.01)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) G06N, 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, WPI, EPDOC, CNKI: 神经网络, 模型, 训练, 样本, 梯度, 提升, 决策树, 分类树, 路径, 标签, 节点, 特征, 正, 负, 升采样, 降采样, 远程过程调用, RPC, 统一资源定位器, URL, 地址, neural network?, DNN, ANN, model?, train+, sample?, gradient, boost+, decision tree?, classification tree?, path?, label?, node?, characteristic?, feature?, positive, negative, upsampl+, downsampl+, remote procedure call, uniform resource locator?, address??																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>CN 105844501 A (上海亿保健康管理有限公司) 2016年 8月 10日 (2016 - 08 - 10) 说明书第[0041]-[0069]段</td> <td>1-14</td> </tr> <tr> <td>Y</td> <td>CN 105279691 A (中国银联股份有限公司) 2016年 1月 27日 (2016 - 01 - 27) 说明书第[0032]-[0065]段, 图1-2</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 106296195 A (阿里巴巴集团控股有限公司) 2017年 1月 4日 (2017 - 01 - 04) 全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 102890803 A (阿里巴巴集团控股有限公司) 2013年 1月 23日 (2013 - 01 - 23) 全文</td> <td>1-14</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	Y	CN 105844501 A (上海亿保健康管理有限公司) 2016年 8月 10日 (2016 - 08 - 10) 说明书第[0041]-[0069]段	1-14	Y	CN 105279691 A (中国银联股份有限公司) 2016年 1月 27日 (2016 - 01 - 27) 说明书第[0032]-[0065]段, 图1-2	1-14	A	CN 106296195 A (阿里巴巴集团控股有限公司) 2017年 1月 4日 (2017 - 01 - 04) 全文	1-14	A	CN 102890803 A (阿里巴巴集团控股有限公司) 2013年 1月 23日 (2013 - 01 - 23) 全文	1-14
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
Y	CN 105844501 A (上海亿保健康管理有限公司) 2016年 8月 10日 (2016 - 08 - 10) 说明书第[0041]-[0069]段	1-14															
Y	CN 105279691 A (中国银联股份有限公司) 2016年 1月 27日 (2016 - 01 - 27) 说明书第[0032]-[0065]段, 图1-2	1-14															
A	CN 106296195 A (阿里巴巴集团控股有限公司) 2017年 1月 4日 (2017 - 01 - 04) 全文	1-14															
A	CN 102890803 A (阿里巴巴集团控股有限公司) 2013年 1月 23日 (2013 - 01 - 23) 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 “&” 同族专利的文件																	
国际检索实际完成的日期 2018年 6月 1日		国际检索报告邮寄日期 2018年 6月 13日															
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局(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 魏晶瑶 电话号码 86-(10)-53961340															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/078906

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	105844501	A	2016年 8月 10日	无	
CN	105279691	A	2016年 1月 27日	无	
CN	106296195	A	2017年 1月 4日	无	
CN	102890803	A	2013年 1月 23日	HK 1176725 A1	2016年 8月 26日