

(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 2 日 (02.07.2020)



(10) 国际公布号
WO 2020/134011 A1

- (51) 国际专利分类号:
G06Q 10/04 (2012.01)
- (21) 国际申请号: PCT/CN2019/094743
- (22) 国际申请日: 2019 年 7 月 4 日 (04.07.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201811647015.1 2018年12月29日 (29.12.2018) CN
- (71) 申请人: 北京字节跳动网络技术有限公司 (**BEIJING BYTEDANCE NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市石景山区实兴大街30号院3号楼2层B-0035房间, Beijing 100041 (CN)。
- (72) 发明人: 李亦钺 (**LI, Yitan**); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。余林韵 (**YU, Linyun**); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。陈嘉闽 (**CHEN, Jiamin**); 中国北京市海淀区知春路63号中

国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。黄训蓬 (**HUANG, Xunpeng**); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。李磊 (**LI, Lei**); 中国北京市海淀区知春路63号中国卫星通信大厦今日头条小邮局, Beijing 100190 (CN)。

(74) 代理人: 北京市立方律师事务所 (**LIFANG & PARTNERS**); 中国北京市东城区东四十条甲22号南新仓商务大厦A座1105室, Beijing 100088 (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 APPARATUS FOR DETERMINING DISPLAY INFORMATION COMBINATION, STORAGE MEDIUM, AND ELECTRONIC DEVICE

(54) 发明名称: 展示信息组合确定方法、装置、存储介质及电子设备

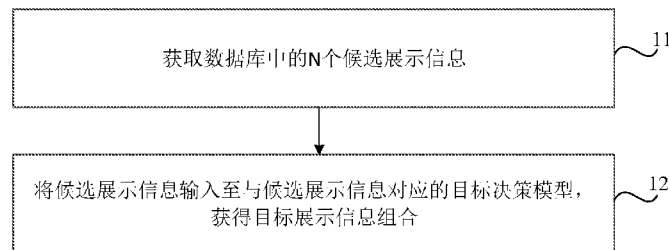


图 1

- 11 Obtain N pieces of candidate display information in a database
12 Input the candidate display information to a target decision model corresponding to the candidate display information to obtain a target display information combination

(57) **Abstract:** A method and apparatus for determining a display information combination, a storage medium, and an electronic device. The method comprises: obtaining N pieces of candidate display information in a database (11), the N pieces of candidate display information being all or some of display information in the database, and N being a positive integer greater than 1; and inputting the candidate display information to a target decision model corresponding to the candidate display information to obtain a target display information combination (12), the target display information combination comprising M pieces of the candidate display information, the target decision model being constructed by using a reinforcement learning network model, and the target decision model being used for determining, according to the input display information, a display information combination that can result in a maximum return value, wherein $1 \leq M \leq N$. In this way, a target display information combination can be automatically generated by means of a reinforcement learning method, thereby saving labor, and the target display information combination can result in a maximum return value, thereby maximizing user benefits.

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)。

根据细则4.17的声明:

- 发明人资格(细则4.17(iv))

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种展示信息组合确定方法、装置、存储介质及电子设备。所述方法包括: 获取数据库中的N个候选展示信息(11), 所述N个候选展示信息为数据库中所有展示信息的全部或部分, 其中, N为大于1的正整数; 将所述候选展示信息输入至与所述候选展示信息对应的目标决策模型, 获得目标展示信息组合(12), 所述目标展示信息组合包括M个所述候选展示信息, 所述目标决策模型是利用强化学习网络模型构建的, 且所述目标决策模型用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合, 其中, $1 \leq M < N$ 。这样, 利用强化学习方法, 可自动生成目标展示信息组合, 节省人力, 还可以使目标展示信息组合能够获得最大的回报值, 使用户利益最大化。

展示信息组合确定方法、装置、存储介质及电子设备

相关申请的交叉引用

5 本申请要求 2018 年 12 月 29 日在中国知识产权局提交的中国专利申请 No. 201811647015.1 的优先权，通过引用将该中国专利申请公开的全部内容并入本文。

技术领域

10 本公开涉及计算机技术领域，具体地，涉及一种展示信息组合确定方法、装置、存储介质及电子设备。

背景技术

信息流（Feeds）是社交媒体移动应用中最重要创新点之一，即当用户
15 加载显示界面时为用户呈现一系列信息。该一系列信息就是一系列展示信息，展示信息可以在展位进行展示，从而获得一定的回报值。展位可以为用于对展示信息进行展示的虚拟或实际的载体。在数据库中存在多种可供展示的展示信息，而在实际的展示中，由于每一次可用于展示的展位数量有限，这些可供展示的展示信息无法全部展示，只能从中选择一部分进行展示，这一部
20 分展示信息可视为展示信息组合。现有技术中，对于展示信息组合的确定需要人工收集数据后进行人为选择，这样，一方面需要投入大量的人力、物力、时间等，效率很低，另一方面，由于人工选择的方式受限于个人经验等因素，会使确定出的展示信息组合得到的回报值无法估量，易出现回报值过低的情况。

25

发明内容

本公开的目的是提供一种展示信息组合确定方法、装置、存储介质及电子设备，以自动生成展示信息组合。

为了实现上述目的，根据本公开的第一方面，提供一种展示信息组合确定方法，所述方法包括：

获取数据库中的 N 个候选展示信息，所述 N 个候选展示信息为数据库中所有展示信息的全部或部分，其中， N 为大于 1 的正整数；

5 将所述候选展示信息输入至与所述候选展示信息对应的目标决策模型，获得目标展示信息组合，所述目标展示信息组合包括 M 个所述候选展示信息，所述目标决策模型是利用强化学习网络模型构建的，且所述目标决策模型用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合，其中， $1 \leq M < N$ 。

10 可选地，与所述候选展示信息对应的所述目标决策模型通过以下方式获得：

获取历史展示信息组合以及所述历史展示信息组合对应的实际回报值，其中，所述历史展示信息组合包括 M 个所述候选展示信息；

15 根据所述 N 个候选展示信息、所述历史展示信息组合以及所述历史展示信息组合对应的实际回报值，对强化学习网络模型进行训练，以获得所述目标决策模型。

可选地，所述根据所述 N 个候选展示信息、所述历史展示信息组合以及所述历史展示信息组合对应的实际回报值对强化学习网络模型进行训练以获得所述目标决策模型包括：

20 初始化所述强化学习网络模型，并根据所述 N 个候选展示信息和所述强化学习网络模型生成第一展示信息组合，所述第一展示信息组合包括 M 个所述候选展示信息；

25 确定与所述第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为所述第一展示信息组合对应的实际回报值；

将所述第一展示信息组合对应的实际回报值输入至所述强化学习网络模型中，以调整所述第一展示信息组合中各候选展示信息在所述强化学习网络模型中的权重参数；

重复执行所述根据所述 N 个候选展示信息和所述强化学习网络模型生成

第一展示信息组合，以及所述确定与所述第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为所述第一展示信息组合对应的实际回报值，以及所述将所述第一展示信息组合对应的实际回报值输入至所述强化学习网络模型中，以调整所述第一展示信息组合中候选展示信息在所述强化学习网络模型中的权重参数的步骤，直至所述重复执行的次数达到预设次数；

将当前的所述强化学习网络模型确定为所述目标决策模型。

可选地，所述获取数据库中的 N 个候选展示信息包括：

获取所述数据库中的第一展示信息，所述第一展示信息为已被单独展示过的展示信息；

获取所述第一展示信息被单独展示时对应的实际回报值；

至少将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为所述候选展示信息。

可选地，所述至少将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为所述候选展示信息包括：

将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为所述候选展示信息，以及，将所述数据库中除所述第一展示信息之外的其他展示信息确定为所述候选展示信息。

可选地，所述方法还包括：

记录所述目标展示信息组合对应的实际回报值；

将所述实际回报值输入至所述目标决策模型，以更新所述目标决策模型中与所述目标展示信息组合中各候选展示信息对应的权重参数。

可选地，候选展示信息在目标决策模型中的权重参数与目标决策模型生成展示信息组合时选择该候选展示信息的几率相关，并且，展示信息组合对应的实际回报值越高，则在更新所述目标决策模型时，该展示信息组合中的候选展示信息在所述目标决策模型中的权重参数的增大程度越高。

根据本公开的第二方面，提供一种展示信息组合确定装置，所述装置包括：

获取模块，用于获取数据库中的 N 个候选展示信息，所述 N 个候选展示

信息为数据库中所有展示信息的全部或部分，其中， N 为大于 1 的正整数；

确定模块，用于将所述候选展示信息输入至与所述候选展示信息对应的目标决策模型，获得目标展示信息组合，所述目标展示信息组合包括 M 个所述候选展示信息，所述目标决策模型是利用强化学习网络模型构建的，且所述目标决策模型用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合，其中， $1 \leq M < N$ 。

可选地，所述确定模块包括：

第一获取子模块，用于获取历史展示信息组合以及所述历史展示信息组合对应的实际回报值，其中，所述历史展示信息组合包括 M 个所述候选展示信息；

训练子模块，用于根据所述 N 个候选展示信息、所述历史展示信息组合以及所述历史展示信息组合对应的实际回报值，对强化学习网络模型进行训练，以获得所述目标决策模型。

可选地，所述训练子模块被配置成：初始化所述强化学习网络模型，并根据所述 N 个候选展示信息和所述强化学习网络模型生成第一展示信息组合，所述第一展示信息组合包括 M 个所述候选展示信息；确定与所述第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为所述第一展示信息组合对应的实际回报值；将所述第一展示信息组合对应的实际回报值输入至所述强化学习网络模型中，以调整所述第一展示信息组合中各候选展示信息在所述强化学习网络模型中的权重参数；触发所述训练子模块重复执行所述根据所述 N 个候选展示信息和所述强化学习网络模型生成第一展示信息组合，以及所述确定与所述第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为所述第一展示信息组合对应的实际回报值，以及所述将所述第一展示信息组合对应的实际回报值输入至所述强化学习网络模型中，以调整所述第一展示信息组合中候选展示信息在所述强化学习网络模型中的权重参数，直至所述重复执行的次数达到预设次数；将当前的所述强化学习网络模型确定为所述目标决策模型。

可选地，所述获取模块包括：

第二获取子模块，用于获取所述数据库中的第一展示信息，所述第一展示信息为已被单独展示过的展示信息；

第三获取子模块，用于获取所述第一展示信息被单独展示时对应的实际回报值；

5 确定子模块，用于至少将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为所述候选展示信息。

可选地，所述确定子模块被配置成：将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为所述候选展示信息，以及，将所述数据库中除所述第一展示信息之外的其他展示信息确定为所述候选展
10 示信息。

可选地，所述装置还包括：

记录模块，用于记录所述目标展示信息组合对应的实际回报值；

更新模块，用于将所述实际回报值输入至所述目标决策模型，以更新所述目标决策模型中与所述目标展示信息组合中各候选展示信息对应的权重参
15 数。

可选地，候选展示信息在目标决策模型中的权重参数与目标决策模型生成展示信息组合时选择该候选展示信息的几率相关，并且，展示信息组合对应的实际回报值越高，则在更新所述目标决策模型时，该展示信息组合中的候选展示信息在所述目标决策模型中的权重参数的增大程度越高。

20 根据本公开的第三方面，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本公开第一方面所述的方法。

根据本公开的第四方面，提供一种电子设备，其包括：

存储器，其上存储有计算机程序；

25 处理器，用于执行所述存储器中的所述计算机程序，以实现本公开第一方面所述的方法。

通过上述技术方案，获取数据库中的 N 个候选展示信息，将候选展示信息输入至与该候选展示信息对应的目标决策模型，获得目标展示信息组合。其中，目标决策模型是利用强化学习网络模型构建的，且目标决策模型用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合。这样，利

用强化学习方法，可自动生成目标展示信息组合，节省人力，另外，还可以使目标展示信息组合能够获得最大的回报值，使用户利益最大化。

附图说明

5 附图是用来帮助对本公开的进一步理解，并且构成说明书的一部分，与下面的具体描述一起用于解释本公开，但并不构成对本公开的限制。在附图中：

图 1 是根据本公开的一种实施方式提供的展示信息组合确定方法的流程图；

10 图 2 是根据本公开提供的展示信息组合确定方法中，获取数据库中的 N 个候选展示信息的步骤的一种示例性实现方式的流程图；

图 3 是根据本公开提供的展示信息组合确定方法中，获得与候选展示信息对应的目标决策模型的方式的一种示例性实现方式的流程图；

图 4 是根据本公开的一种实施方式提供的展示信息组合确定装置的框图；

15 图 5 是根据示例性实施例的电子设备的框图。

具体实施方式

以下结合附图对本公开的具体实施方式进行详细说明。应当理解的是，此处所描述的具体实施方式仅用于说明和解释本公开，并不用于限制本公开。

20 需要说明的是，本公开的说明书和权利要求书以及上述附图中的术语“第一”等是用于区别类似的对象，而不必理解为描述特定的顺序或先后次序。

图 1 是根据本公开的一种实施方式提供的展示信息组合确定方法的流程图。如图 1 所示，该方法可以包括以下步骤。

在步骤 11 中，获取数据库中的 N 个候选展示信息。

25 其中，N 为大于 1 的正整数。示例地，N 个候选展示信息可以为数据库中所有的展示信息。再例如，N 个候选展示信息可以为数据库中所有的展示信息的一部分。

在一种可能的实施方式中，步骤 11 中获取数据库中的 N 个候选展示信息可以直接获取数据库中的所有展示信息作为 N 个候选展示信息，这里 N 即为

数据库中所有展示信息的个数。

在另一种可能的实施方式中，步骤 11 中获取数据库中的 N 个候选展示信息可以从数据库中的所有展示信息中随机选取 N 个展示信息作为候选展示信息，这里 N 小于数据库中所有展示信息的个数。

5 在另一种可能的实施方式中，如图 2 所示，步骤 11 可以包括以下步骤：
在步骤 21 中，获取数据库中的第一展示信息。

其中，第一展示信息为已被单独展示过的展示信息。

在步骤 22 中，获取第一展示信息被单独展示时对应的实际回报值。

10 针对已被单独展示过的第一展示信息，能够获取到第一展示信息被单独展示时各自对应的实际回报值。

在步骤 23 中，至少将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为候选展示信息。

在一种可能的实施例中，可以将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为候选展示信息。

15 采用上述方式，可以对数据库中的展示信息进行筛选以得到候选展示信息，即将单独展示时能够得到高回报值的展示信息作为候选展示信息，从而使候选展示信息本身在获得高回报值方面就具备一定的优势，从而对后续确定的目标展示信息组合获得高回报值具有积极效果。

20 在另一种可能的实施例中，除了将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为候选展示信息外，还可以将数据库中除第一展示信息之外的其他展示信息确定为候选展示信息。

也就是说，候选展示信息除了包括单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息外，还包括数据库中未被展示过的展示信息。

25 采用上述方式，一方面，可以将单独展示时能够得到高回报值的展示信息作为候选展示信息，从而使候选展示信息本身在获得高回报值方面具备一定的优势，从而对后续确定的目标展示信息组合获得高回报值具有积极效果。另一方面，也可以减小未被单独展示过的展示信息在候选展示信息选取中的劣势，避免出现未被单独展示过的展示信息总是不被选取为候选展示信息的

情况。

现在回到图 1，在步骤 12 中，将候选展示信息输入至与该候选展示信息对应的目标决策模型，获得目标展示信息组合。

其中，目标展示信息组合可以包括 M 个候选展示信息，且 $1 \leq M < N$ 。目标决策模型是利用强化学习网络模型构建的，且目标决策模型用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合。

下面将对强化学习进行简单介绍。强化学习网络模型以“试错”的方式进行学习，通过与环境进行交互获得的奖赏来指导行为。强化学习的目标是使强化学习网络模型获得最大的奖赏，强化学习不同于监督学习，主要表现在强化信号上。强化学习中由环境提供的强化信号是对产生动作的好坏作一种评价，而不是告诉强化学习网络模型如何去产生正确的动作。由于外部环境提供的信息很少，强化学习网络模型需要靠自身的经历进行学习。其中，强化学习网络模型获取外部环境的当前状态信息，对环境采取试探行为，并获取环境反馈的对此动作的评价和新的环境状态。如果强化学习网络模型的某动作导致环境正的奖赏，那么强化学习网络模型以后产生这个动作的趋势便会加强；反之，强化学习网络模型产生这个动作的趋势将减弱。通过这样的方式，强化学习网络模型通过逐步学习优化模型，可以使其选择的行为能够获得环境最大的奖赏。示例地，强化学习网络模型可以为多臂老虎机模型。

下面对本公开利用强化学习的原理进行简要说明。在本公开中，最终的目标是使展示信息组合能够尽可能获得高回报值。因此，强化学习网络模型所产生的动作就是生成展示信息组合，而动作导致的奖赏则通过生成的展示信息组合的实际回报值判定。例如，强化学习网络模型此次生成的展示信息组合所得到的实际回报值相较于上次生成的展示信息组合所得到的实际回报值有所增加，则可认为此次生成的展示信息组合导致环境正的奖赏，那么强化学习网络模型后续生成此次对应的展示信息组合的趋势将会增强。示例地，强化学习网络模型增强产生某一动作的趋势的具体的操作可以为增大此次的展示信息组合中各个展示信息在强化学习网络模型中的权重参数。其中，展示信息在强化学习网络模型中的权重参数可以与强化学习网络模型生成展示信息组合时选择该展示信息的几率相关。例如，某展示信息在强化学习网络

模型中的权重参数越大，则该展示信息在强化学习网络模型生成展示信息组合时被选择的几率越高。通过这样的方式，在一次又一次生成展示信息组合并收集相应的实际回报值后，逐步优化强化学习网络模型，可以使强化学习网络模型所生成的展示信息组合能够获得最大回报值。

5 在一种可能的实施方式中，可以首先初始化一强化网络学习模型，并直接将该初始化的强化网络学习模型作为与 N 个候选展示信息对应的目标决策模型。在这种实施方式中，目标决策模型内各个候选展示信息的权重参数是随机的。

10 在另一种可能的实施方式中，在获得与候选展示信息对应的目标决策模型时，可以首先确定是否存在历史展示信息组合。历史展示信息组合可以包括 M 个候选展示信息，且 $1 \leq M < N$ 。也就是说，历史展示信息组合需满足两个条件，第一个条件是组合内的展示信息均为候选展示信息，第二个条件是组合内的展示信息个数与欲构建的目标决策模型所要生成的目标展示信息组合内展示信息个数相同，也就是均为 M 个。若存在历史展示信息组合，则说明在构建目标决策模型之前，已经有相应的展示信息组合被展示，因此可以将与这些历史展示信息组合相关的数据收集起来，用以辅助获得目标决策模型。因此，在这种实施方式中，如图 3 所示，与候选展示信息对应的目标决策模型可以通过以下步骤获得。

15

20 在步骤 31 中，获取历史展示信息组合以及历史展示信息组合对应的实际回报值。

在步骤 32 中，根据 N 个候选展示信息、历史展示信息组合以及历史展示信息组合对应的实际回报值，对强化学习网络模型进行训练，以获得目标决策模型。

在一种可能的实施例中，步骤 32 可以包括以下步骤：

25 在第一步骤中，初始化强化学习网络模型；

在第二步骤中，根据 N 个候选展示信息和强化学习网络模型生成第一展示信息组合；

在第三步骤中，确定与第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为第一展示信息组合对应的实

实际回报值；

在第四步骤中，将第一展示信息组合对应的实际回报值输入至强化学习网络模型中，以调整第一展示信息组合中各候选展示信息在强化学习网络模型中的权重参数；

- 5 重复执行上述第二步骤、以及第三步骤、以及第四步骤，直至重复执行的次数达到预设次数；

将当前的强化学习网络模型确定为目标决策模型。

- 10 其中，第一展示信息组合包括 M 个候选展示信息。第一展示信息组合中每一个候选展示信息在强化学习网络模型中的权重参数可以与强化学习网络模型生成第一展示信息组合时选择该候选展示信息的几率相关，并且，第一展示信息组合对应的实际回报值越高，则在更新强化学习网络模型时，该第一展示信息组合中的候选展示信息在强化学习网络模型中的权重参数的增大程度越高。

- 15 首先，初始化强化学习网络模型。此时，当前的强化学习网络模型内各个候选展示信息的权重参数是随机的。之后，当前的强化学习网络模型可以根据 N 个候选展示信息生成第一展示信息组合。

- 20 在生成第一展示信息组合后，可以确定与第一展示信息组合相同的历史展示信息组合，若存在这样的历史展示信息组合，则将与该历史展示信息组合对应的实际回报值确定为第一展示信息组合对应的实际回报值。从而，可以认为该与第一展示信息组合对应的实际回报值就是第一展示信息组合实际展示所对应的回报值，因此可以将该第一展示信息组合对应的实际回报值输入至强化学习网络模型中，以调整第一展示信息组合中各候选展示信息在强化学习网络模型中的权重参数。而若不存在与第一展示信息组合相同的历史展示信息组合，则说明对于当前的第一展示信息组合没有可以参考的实际数据，因此可以忽略此次生成的第一展示信息组合而重新生成新的第一展示信息组合。
- 25

其中，若此次生成的第一展示信息组合对应的实际回报值有所增加，则增大该第一展示信息组合中各个候选展示信息在当前强化学习网络模型的权重参数。初始情况下，由于第一展示信息组合对应的实际回报值没有比较的

对象，因此可以预先设置一个比较阈值，将第一展示信息组合对应的实际回报值与该比较阈值进行比较，以判断此次第一展示信息组合对应的实际回报值的增加或减小情况。示例地，该比较阈值可以设置为 0。

另外，在调整强化学习网络模型的权重参数的过程中，还可以根据此次生成的第一展示信息组合对应的实际回报值相较于上次生成的第一展示信息组合对应的实际回报值的变化程度确定对此次的第一展示信息组合中各候选展示信息的权重参数的调整程度。例如，若此次生成的第一展示信息组合对应的实际回报值相较于上次生成的第一展示信息组合对应的实际回报值的增加程度较大，则在此次调整强化学习网络模型的权重参数时，可以将此次生成的第一展示信息组合对应的各个候选展示信息在强化学习网络模型中的权重参数以较大的增加幅度调整。

之后，可以重复执行上述第二步骤、第三步骤以及第四步骤，也就是生成新的第一展示信息组合对强化学习网络模型内的相应权重参数进行调整，直至重复执行的次数达到预设次数。这里，重复执行的次数可以从第一次重复执行开始计算。示例地，若在初始情况下，第二步骤、第三步骤以及第四步骤被执行过一次后，开始重复执行第二步骤、第三步骤以及第四步骤，那么在此次重复执行后，重复执行次数为 1。预设次数可以人为进行设定，例如可以根据历史展示信息组合的数量确定，示例地，预设次数可以大于或者等于历史展示信息组合的数量，以尽可能多地应用历史展示信息组合。在重复执行的次数达到预设次数后，则可以将当前的强化学习网络模型确定为目标决策模型。

需要说明的是，上述构建的目标决策模型是针对上述 N 个候选展示信息所生成的，并且该目标决策模型用于针对这 N 个候选展示信息生成含 M 个候选展示信息的展示信息组合。若候选展示信息的内容或者数量存在变化，或者目标决策模型需要生成的展示信息组合所包含的展示信息的数量存在变化，则均需要重新收集相关的数据重新构建目标决策模型。

通过上述方式，可以利用已被展示过的历史展示信息组合以及历史展示信息组合对应的实际回报值辅助构建目标决策模型，使构建的目标决策模型在初次应用时即可具备一定的优势，能够得到较高的回报值。

在另一种可能的实施例中，步骤 32 可以包括以下步骤：

在第一步骤中，初始化强化学习网络模型；

在第二步骤中，根据 N 个候选展示信息和强化学习网络模型生成第一展示信息组合；

5 在第三步骤中，确定与第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为第一展示信息组合对应的实际回报值；

10 在第四步骤中，将第一展示信息组合对应的实际回报值输入至强化学习网络模型中，以调整第一展示信息组合中各候选展示信息在强化学习网络模型中的权重参数；

重复执行上述第二步骤、以及第三步骤、以及第四步骤，直至所有的历史展示信息组合的实际回报值均被输入至强化学习网络模型以调整强化学习网络模型中候选展示信息组合的权重参数；

将当前的强化学习网络模型确定为目标决策模型。

15 所有的历史展示信息组合的实际回报值均被输入至强化学习网络模型以调整强化学习网络模型中候选展示信息组合的权重参数，是指强化学习网络模型在各次所生成的第一展示信息组合已经覆盖了全部的历史展示信息组合，也就是说，当前收集的历史展示信息组合以及历史展示信息组合对应的实际回报值均被使用过，且用于调整强化学习网络模型内各个候选展示信息的权重参数，已经没有其他数据可用，因此，可以不必再对强化学习网络模型继续进行训练。另外，关于上述实施例中的其他步骤，详细的执行方式在上文中均有描述，此处不赘述。

25 需要说明的是，上述构建的目标决策模型是针对上述 N 个候选展示信息所生成的，并且该目标决策模型用于针对这 N 个候选展示信息生成含 M 个候选展示信息的展示信息组合。若候选展示信息的内容或者数量存在变化，或者目标决策模型需要生成的展示信息组合所包含的展示信息的数量存在变化，则均需要重新收集相关的数据重新构建目标决策模型。

通过上述方案，获取数据库中的 N 个候选展示信息，将候选展示信息输入至与该候选展示信息对应的目标决策模型，获得目标展示信息组合。其中，

目标决策模型是利用强化学习网络模型构建的，且目标决策模型用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合。这样，利用强化学习方法，可自动生成目标展示信息组合，节省人力，另外，还可以使目标展示信息组合能够获得最大的回报值，使用户利益最大化。

5 在一种可能的实施方式中，本公开提供的方法还可以包括以下步骤：

记录目标展示信息组合对应的实际回报值；

将实际回报值输入至目标决策模型，以更新目标决策模型中与目标展示信息组合中各候选展示信息对应的权重参数。

10 其中，候选展示信息在目标决策模型中的权重参数与目标决策模型生成目标展示信息组合时选择该候选展示信息的几率相关，并且，目标展示信息组合对应的实际回报率越高，则在更新目标决策模型时，该目标展示信息组合中的候选展示信息在目标决策模型中的权重参数的增大程度越高。

另外，上述权重参数的更新与上文原理相同，且在上文中已有相关说明，此处不赘述。

15 通过上述方案，可通过收集目标展示信息组合对应的实际回报值，用以调整目标决策模型中与目标展示信息组合中各候选展示信息对应的权重参数，以优化目标决策模型，从而使目标决策模型的效果更加出色，使后续生成的目标展示信息组合能够获得更大的回报值。

20 需要说明的是，利用强化学习方法构建模型的方法与流程均为本领域技术人员公知，为理解方便在上文中对于其中的一种可能的情况进行了简要说明，但是本公开中构建模型的方式并不局限于此，对于其他的实现方式此处不赘述。

图 4 是根据本公开的一种实施方式提供的展示信息组合确定装置的框图。如图 4 所示，展示信息组合确定装置 40 可以包括：

25 获取模块 41，用于获取数据库中的 N 个候选展示信息，该 N 个候选展示信息为数据库中所有展示信息的全部或部分，其中，N 为大于 1 的正整数；

确定模块 42，用于将候选展示信息输入至与该候选展示信息对应的目标决策模型，获得目标展示信息组合，该目标展示信息组合包括 M 个候选展示信息，该目标决策模型是利用强化学习网络模型构建的，且该目标决策模型

用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合，其中， $1 \leq M < N$ 。

可选地，确定模块 42 可以包括：

5 第一获取子模块，用于获取历史展示信息组合以及该历史展示信息组合对应的实际回报值，其中，该历史展示信息组合包括 M 个候选展示信息；

训练子模块，用于根据该 N 个候选展示信息、该历史展示信息组合以及该历史展示信息组合对应的实际回报值，对强化学习网络模型进行训练，以获得该目标决策模型。

10 可选地，该训练子模块被配置成：初始化该强化学习网络模型，并根据该 N 个候选展示信息和该强化学习网络模型生成第一展示信息组合，该第一展示信息组合包括 M 个候选展示信息；确定与该第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为该第一展示信息组合对应的实际回报值；将该第一展示信息组合对应的实际回报值输入至该强化学习网络模型中，以调整该第一展示信息组合中各候选展示信息在该强化学习网络模型中的权重参数；触发该训练子模块重复执行根据该 N 个候选展示信息和该强化学习网络模型生成第一展示信息组合，确定与该第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为该第一展示信息组合对应的实际回报值，以及将该第一展示信息组合对应的实际回报值输入至该强化学习网络模型中，以调整该第一展示信息组合中候选展示信息在该强化学习网络模型中的权重参数，直至重复执行的次数达到预设次数；将当前的该强化学习网络模型确定为该目标决策模型。

可选地，获取模块 41 可以包括：

25 第二获取子模块，用于获取该数据库中的第一展示信息，该第一展示信息为已被单独展示过的展示信息；

第三获取子模块，用于获取该第一展示信息被单独展示时对应的实际回报值；

确定子模块，用于至少将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为该候选展示信息。

可选地，该确定子模块被配置成：将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为该候选展示信息，以及，将该数据库中除该第一展示信息之外的其他展示信息确定为该候选展示信息。

可选地，展示信息组合确定装置 40 还可以包括：

5 记录模块，用于记录该目标展示信息组合对应的实际回报值；

更新模块，用于将该实际回报值输入至该目标决策模型，以更新该目标决策模型中与该目标展示信息组合中各候选展示信息对应的权重参数。

10 可选地，候选展示信息在目标决策模型中的权重参数与目标决策模型生成展示信息组合时选择该候选展示信息的几率相关，并且，展示信息组合对应的实际回报值越高，则在更新该目标决策模型时，该展示信息组合中的候选展示信息在该目标决策模型中的权重参数的增大程度越高。

关于上述实施例中的装置，其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述，此处将不做详细阐述说明。

15 图 5 是根据一示例性实施例示出的一种电子设备的框图。例如，电子设备 1900 可以被提供为一服务器。参照图 5，电子设备 1900 包括处理器 1922，其数量可以为一个或多个，以及存储器 1932，用于存储可由处理器 1922 执行的计算机程序。存储器 1932 中存储的计算机程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外，处理器 1922 可以被配置为执行该计算机程序，以执行上述的展示信息组合确定方法。

20 另外，电子设备 1900 还可以包括电源组件 1926 和通信组件 1950，该电源组件 1926 可以被配置为执行电子设备 1900 的电源管理，该通信组件 1950 可以被配置为实现电子设备 1900 的通信，例如，有线或无线通信。此外，该电子设备 1900 还可以包括输入/输出 (I/O) 接口 1958。电子设备 1900 可以操作基于存储在存储器 1932 的操作系统，例如 Windows Server™, Mac OS X™, 25 Unix™, Linux™ 等等。

在另一示例性实施例中，还提供了一种包括程序指令的计算机可读存储介质，该程序指令被处理器执行时实现上述的展示信息组合确定方法。例如，该计算机可读存储介质可以为上述包括程序指令的存储器 1932，上述程序指令可由电子设备 1900 的处理器 1922 执行以完成上述的展示信息组合确定方

法。

以上结合附图详细描述了本公开的示例性实施方式，但是，本公开并不限于上述实施方式中的具体细节，在本公开的技术构思范围内，可以对本公开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保护范围。

- 5 另外需要说明的是，在上述具体实施方式中所描述的各个具体技术特征，在不矛盾的情况下，可以通过任何合适的方式进行组合。为了避免不必要的重复，本公开对各种可能的组合方式不再另行说明。

此外，本公开的各种不同的实施方式之间也可以进行任意组合，只要其不违背本公开的构思，就应被视为在本公开的范围內。

权 利 要 求 书

1、一种展示信息组合确定方法，包括：

获取数据库中的 N 个候选展示信息，所述 N 个候选展示信息为数据库中所有展示信息的全部或部分，其中， N 为大于 1 的正整数；

5 将所述候选展示信息输入至与所述候选展示信息对应的目标决策模型，获得目标展示信息组合，所述目标展示信息组合包括 M 个所述候选展示信息，所述目标决策模型是利用强化学习网络模型构建的，且所述目标决策模型用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合，其中， $1 \leq M < N$ 。

10 2、根据权利要求 1 所述的展示信息组合确定方法，其中，与所述候选展示信息对应的所述目标决策模型通过以下方式获得：

获取历史展示信息组合以及所述历史展示信息组合对应的实际回报值，其中，所述历史展示信息组合包括 M 个所述候选展示信息；

15 根据所述 N 个候选展示信息、所述历史展示信息组合以及所述历史展示信息组合对应的实际回报值，对强化学习网络模型进行训练，以获得所述目标决策模型。

20 3、根据权利要求 2 所述的展示信息组合确定方法，其中，所述根据所述 N 个候选展示信息、所述历史展示信息组合以及所述历史展示信息组合对应的实际回报值对强化学习网络模型进行训练以获得所述目标决策模型包括：

初始化所述强化学习网络模型，并根据所述 N 个候选展示信息和所述强化学习网络模型生成第一展示信息组合，所述第一展示信息组合包括 M 个所述候选展示信息；

25 确定与所述第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为所述第一展示信息组合对应的实际回报值；

将所述第一展示信息组合对应的实际回报值输入至所述强化学习网络模

型中，以调整所述第一展示信息组合中各候选展示信息在所述强化学习网络模型中的权重参数；

重复执行所述根据所述 N 个候选展示信息和所述强化学习网络模型生成第一展示信息组合，以及所述确定与所述第一展示信息组合相同的历史展示信息组合，并将该历史展示信息组合对应的实际回报值确定为所述第一展示信息组合对应的实际回报值，以及所述将所述第一展示信息组合对应的实际回报值输入至所述强化学习网络模型中，以调整所述第一展示信息组合中候选展示信息在所述强化学习网络模型中的权重参数的步骤，直至所述重复执行的次数达到预设次数；

将当前的所述强化学习网络模型确定为所述目标决策模型。

4、根据权利要求 1 所述的展示信息组合确定方法，其中，所述获取数据库中的 N 个候选展示信息包括：

获取所述数据库中的第一展示信息，所述第一展示信息为已被单独展示过的展示信息；

获取所述第一展示信息被单独展示时对应的实际回报值；

至少将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为所述候选展示信息。

5、根据权利要求 4 所述的展示信息组合确定方法，其中，所述至少将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为所述候选展示信息包括：

将单独展示时对应的实际回报值大于或者等于预设回报值阈值的第一展示信息确定为所述候选展示信息，以及，将所述数据库中除所述第一展示信息之外的其他展示信息确定为所述候选展示信息。

6、根据权利要求 1 所述的展示信息组合确定方法，其中，所述方法还包括：

记录所述目标展示信息组合对应的实际回报值；

将所述实际回报值输入至所述目标决策模型，以更新所述目标决策模型中与所述目标展示信息组合中各候选展示信息对应的权重参数。

7、根据权利要求 3 或 6 所述的展示信息组合确定方法，其中，候选展示信息在目标决策模型中的权重参数与目标决策模型生成展示信息组合时选择该候选展示信息的几率相关，并且，展示信息组合对应的实际回报值越高，则在更新所述目标决策模型时，该展示信息组合中的候选展示信息在所述目标决策模型中的权重参数的增大程度越高。

8、一种展示信息组合确定装置，包括：

获取模块，用于获取数据库中的 N 个候选展示信息，所述 N 个候选展示信息为数据库中所有展示信息的全部或部分，其中， N 为大于 1 的正整数；

确定模块，用于将所述候选展示信息输入至与所述候选展示信息对应的目标决策模型，获得目标展示信息组合，所述目标展示信息组合包括 M 个所述候选展示信息，所述目标决策模型是利用强化学习网络模型构建的，且所述目标决策模型用于根据输入的展示信息确定出能够获得最大回报值的展示信息组合，其中， $1 \leq M < N$ 。

9、一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现权利要求 1-7 中任一项所述的方法。

10、一种电子设备，包括：

存储器，其上存储有计算机程序；

处理器，用于执行所述存储器中的所述计算机程序，以实现权利要求 1-7 中任一项所述的方法。

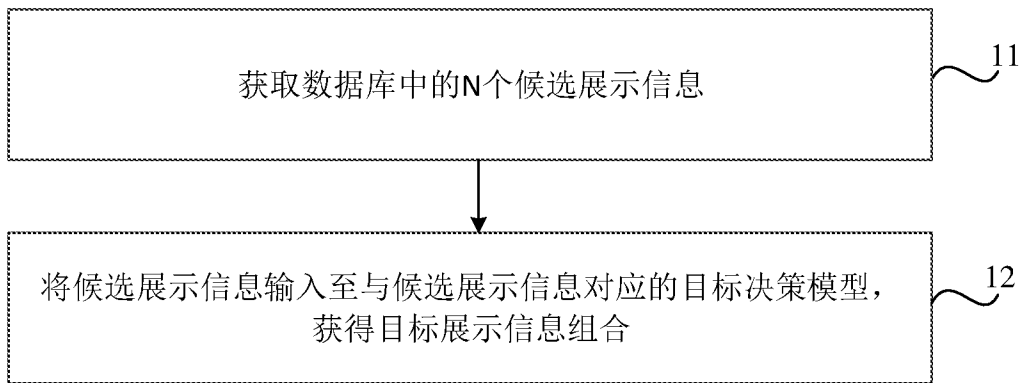


图 1

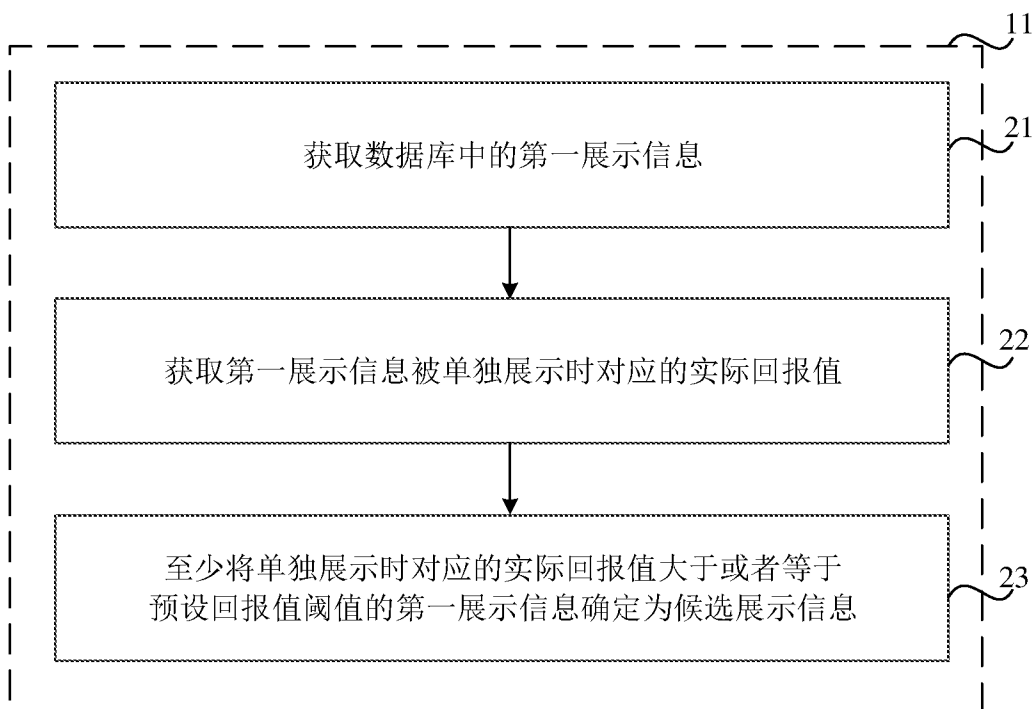


图 2

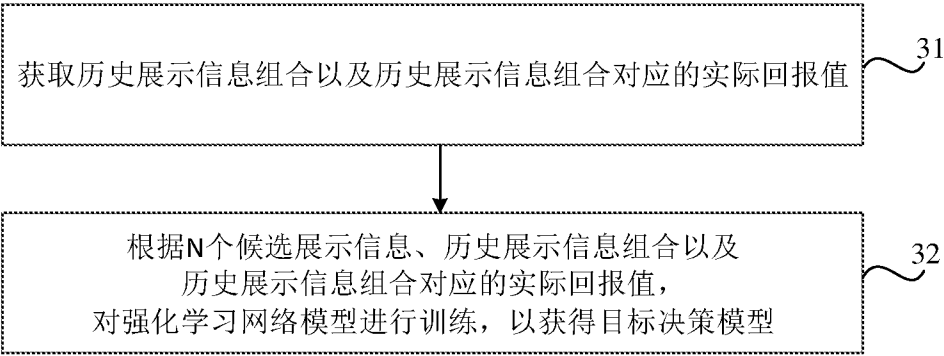


图 3

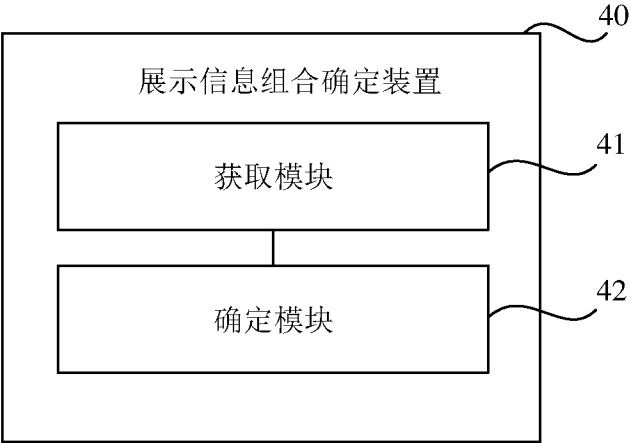


图 4

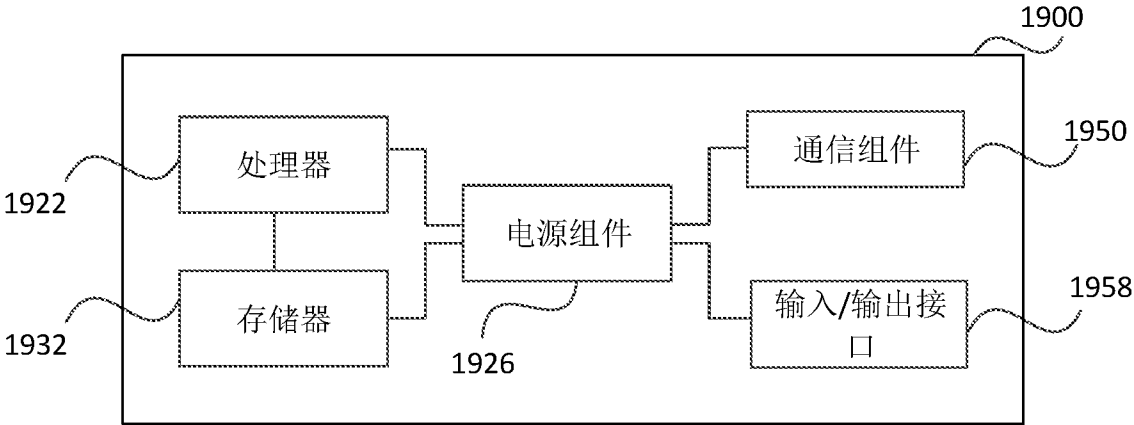


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/094743

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/04(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)

CNPAT, WPI, EPODOC, CNKI, IEEE: 信息, 展示, 数据库, 模型, 学习, 训练, 回报, 反馈, 调整, 优化, information, lay out, database, model, learn, train, return, feedback, adjust, optimization

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107798608 A (SHENZHEN NICEFILM TECHNOLOGY CO., LTD.) 13 March 2018 (2018-03-13) claims 1-10	1-10
A	CN 109062919 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 21 December 2018 (2018-12-21) entire document	1-10
A	CN 107908735 A (BEIJING SCIENCE AND TECHNOLOGY CO., THREE FAST ONLINE) 13 April 2018 (2018-04-13) entire document	1-10
A	CN 108108821 A (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP., LTD.) 01 June 2018 (2018-06-01) entire document	1-10
A	CN 108345419 A (HUAWEI TECHNOLOGIES CO., LTD.) 31 July 2018 (2018-07-31) entire document	1-10
A	WO 03065170 A2 (MANUGISTICS ATLANTA, INC.) 07 August 2003 (2003-08-07) entire document	1-10



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

21 September 2019

Date of mailing of the international search report

08 October 2019

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

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/094743

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107798608	A	13 March 2018	None			
CN	109062919	A	21 December 2018	None			
CN	107908735	A	13 April 2018	None			
CN	108108821	A	01 June 2018	WO	2019128435	A1	04 July 2019
				EP	3506184	A1	03 July 2019
				US	2019205701	A1	04 July 2019
				HK	1251328	A0	25 January 2019
				IN	201814041215	A	05 July 2019
CN	108345419	A	31 July 2018	WO	2018137420	A1	02 August 2018
WO	03065170	A2	07 August 2003	US	2003220773	A1	27 November 2003
				AU	2003207784	A1	02 September 2003
				EP	1479020	A2	24 November 2004

国际检索报告

国际申请号

PCT/CN2019/094743

A. 主题的分类

G06Q 10/04 (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))

CNPAT, WPI, EPDOC, CNKI, IEEE: 信息, 展示, 数据库, 模型, 学习, 训练, 回报, 反馈, 调整, 优化, information, lay out, database, model, learn, train, return, feedback, adjust, optimization

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
X	CN 107798608 A (深圳市耐飞科技有限公司) 2018年 3月 13日 (2018 - 03 - 13) 权利要求1-10	1-10
A	CN 109062919 A (腾讯科技深圳有限公司) 2018年 12月 21日 (2018 - 12 - 21) 全文	1-10
A	CN 107908735 A (北京三快在线科技有限公司) 2018年 4月 13日 (2018 - 04 - 13) 全文	1-10
A	CN 108108821 A (广东欧珀移动通信有限公司) 2018年 6月 1日 (2018 - 06 - 01) 全文	1-10
A	CN 108345419 A (华为技术有限公司) 2018年 7月 31日 (2018 - 07 - 31) 全文	1-10
A	WO 03065170 A2 (MANUGISTICS ATLANTA, INC.) 2003年 8月 7日 (2003 - 08 - 07) 全文	1-10



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

2019年 9月 21日

国际检索报告邮寄日期

2019年 10月 8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

胡妮

电话号码 86-(10)-53961429

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/094743

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	107798608	A	2018年 3月 13日	无			
CN	109062919	A	2018年 12月 21日	无			
CN	107908735	A	2018年 4月 13日	无			
CN	108108821	A	2018年 6月 1日	WO	2019128435	A1	2019年 7月 4日
				EP	3506184	A1	2019年 7月 3日
				US	2019205701	A1	2019年 7月 4日
				HK	1251328	A0	2019年 1月 25日
				IN	201814041215	A	2019年 7月 5日
CN	108345419	A	2018年 7月 31日	WO	2018137420	A1	2018年 8月 2日
WO	03065170	A2	2003年 8月 7日	US	2003220773	A1	2003年 11月 27日
				AU	2003207784	A1	2003年 9月 2日
				EP	1479020	A2	2004年 11月 24日