

(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 8 月 27 日 (27.08.2020)



(10) 国际公布号
WO 2020/168839 A1

(51) 国际专利分类号:
G06F 16/33 (2019.01)

(21) 国际申请号: PCT/CN2020/070014

(22) 国际申请日: 2020 年 1 月 2 日 (02.01.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910130377.1 2019年2月21日 (21.02.2019) CN

(71) 申请人: 北京京东尚科信息技术有限公司 (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD) [CN/CN]; 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。北京京东世纪贸易有限公司 (BEIJING JINGDONG CENTURY TRADING CO., LTD.) [CN/

CN]; 中国北京市大兴区北京经济技术开发区科创十一街 18 号 C 座 2 层 201 室, Beijing 100176 (CN)。

(72) 发明人: 胡忆桐 (HU, Yitong); 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。高云 (GAO, Yun); 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。王娜 (WANG, Na); 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。左丽丽 (ZUO, Lili); 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。张亚红 (ZHANG, Yahong); 中国北京市海淀区知春路 76 号 8 层, Beijing 100086 (CN)。

(74) 代理人: 上海弼兴律师事务所 (SHANGHAI BESHINING LAW OFFICE); 中国上海市小木桥路 681 号外经大厦 21 楼, Shanghai 200032 (CN)。

(54) Title: ITEM RECALL METHOD AND SYSTEM, ELECTRONIC DEVICE AND READABLE STORAGE MEDIUM

(54) 发明名称: 物品召回方法、系统、电子设备及可读存储介质

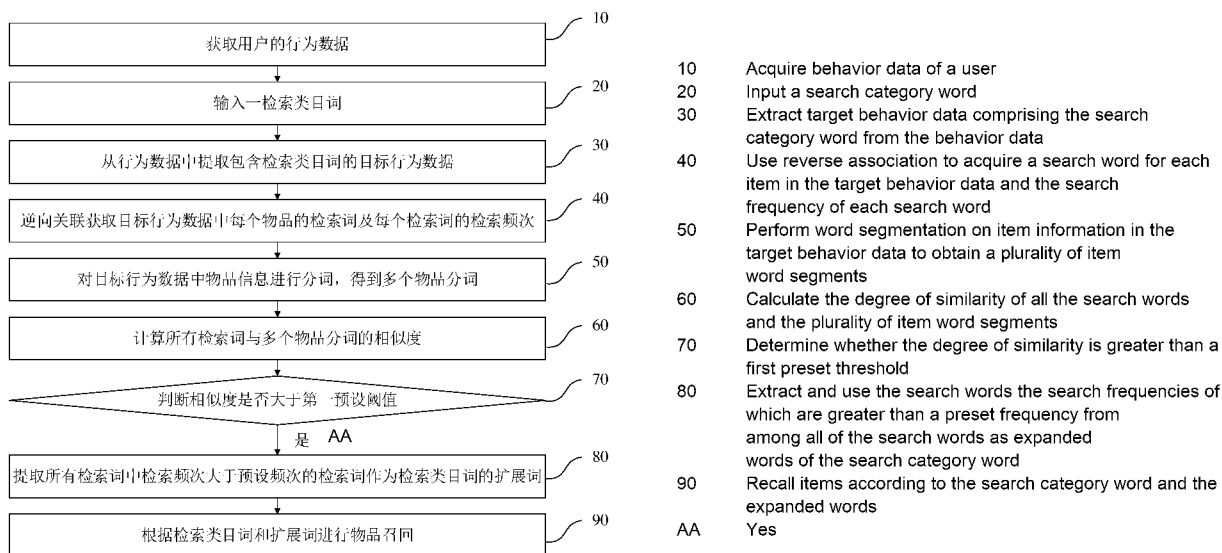


图 1

(57) Abstract: An item recall method and system, an electronic device and a readable storage medium, the item recall method comprising: acquiring behavior data of a user, the behavior data comprising clicked and/or purchased items and item information of each item (10); inputting a search category word (20); extracting target behavior data comprising the search category word from the behavior data (30); using reverse association to acquire a search word for each item in the target behavior data and the search frequency of each search word (40); performing word segmentation on the item information in the target behavior data to obtain a plurality of item word segments (50); calculating the degree of similarity of all the search words and the plurality of item word segments (60); determining whether the degree of similarity is greater than a first preset threshold (70); if so, extracting and using the search words the search frequencies of which are greater than a preset frequency from among all of the search words as expanded words of the search category

(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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

根据细则4.17的声明:

- 关于申请人有权申请并被授予专利(细则4.17(ii))
- 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))。

word (80); and recalling items according to the search category word and the expanded words (90). The described method utilizes acquired behavior data to extract expanded words of an item category word so as to improve the coverage of item recall.

(57) 摘要: 一种物品召回方法、系统、电子设备及可读存储介质, 所述物品召回方法包括: 获取用户的行为数据, 行为数据包括点击和/或购买的物品及每个物品的物品信息(10); 输入一检索类目词(20); 从行为数据中提取包含检索类目词的目标行为数据(30); 逆向关联获取目标行为数据中每个物品的检索词及每个检索词的检索频次(40); 对目标行为数据中物品信息进行分词, 得到多个物品分词(50); 计算所有检索词与多个物品分词的相似度(60); 判断相似度是否大于第一预设阈值(70), 若是, 则提取所有检索词中检索频次大于预设频次的检索词作为检索类目词的扩展词(80); 根据检索类目词和扩展词进行物品召回(90)。该方法利用获取的行为数据提取得到物品类目词的扩展词, 提高了召回物品的覆盖率。

物品召回方法、系统、电子设备及可读存储介质

本申请要求申请日为 2019/2/21 的中国专利申请 2019101303771 的优先权。本申请引用上述中国专利申请的全文。

技术领域

本发明属于大数据处理领域，特别涉及一种物品召回方法、系统、电子设备及可读存储介质。

背景技术

在电商领域中，经常出现需要基于物品类目去做物品召回的类似场景，比如基于类目（例如“连衣裙”）去召回一批符合要求的物品的类似场景，为了后续筛选或者排序算法可以有更大的发挥空间，我们需要在不失准确的基础上，召回尽可能多的物品。现有技术会用到文字匹配的相关方法，但对于同一物品，可能存在描述上的差异，不一定会包含该类目词，导致无法被召回。

发明内容

本发明要解决的技术问题是为了克服现有技术中由于描述差异导致物品召回效果差的缺陷，提供一种物品召回方法、系统、电子设备及可读存储介质。

本发明是通过下述技术方案来解决上述技术问题：

一种物品召回方法，所述物品召回方法包括：

获取用户的行为数据，所述行为数据包括点击和/或购买的物品及每个物品的物品信息；

输入一检索类目词；

从所述行为数据中提取包含所述检索类目词的目标行为数据；

逆向关联获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次；

对所述目标行为数据中物品信息进行分词，得到多个物品分词；

计算所有检索词与所述多个物品分词的相似度；

判断所述相似度是否大于第一预设阈值，若是，则提取所述所有检索词中检索频次大于预设频次的检索词作为所述检索类目词的扩展词；

根据所述检索类目词和所述扩展词进行物品召回。

较佳地，所述逆向关联获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次的步骤具体包括：

根据所述行为数据生成关系列表，所述关系列表用于表征检索词与点击和/或购买的物品的对应关系；

基于所述关系列表获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次。

较佳地，所述计算所有检索词与所述多个物品分词的相似度的步骤具体包括：

对所有的检索词进行去重生成检索词集合；

对所述多个物品分词进行去重生成物品分词集合；

基于 Jaccard 算法（用于比较样本集之间的相似性与差异性的算法）计算所述检索词集合和所述物品分词集合的相似度。

较佳地，所述根据所述检索类目词和所述扩展词进行物品召回物品的步骤之前，所述物品召回方法还包括：

构建语料库，所述语料库包括物品推荐文章和物品基本信息；

基于 word2vec（一种用于产生词向量的模型）对所述语料库进行训练，得到一词向量库，所述词向量库存储有多个物品特征词及与每个物品特征词对应的特征向量；

基于相似度算法计算所述检索类目词与每个物品特征词的相似度；

提取与所述检索类目词的相似度大于第二预设阈值的物品特征词加入所述扩展词。

较佳地，所述根据所述检索类目词和所述扩展词进行物品召回物品的步骤之前，所述物品召回方法还包括：

预设一物品类目列表，所述物品类目列表包括多个物品类目及每个物品类目下所有物品的属性；

判断所述扩展词是否包含在所述物品类目列表中，若是，则滤除包含在所述类目列表的扩展词；

所述根据所述检索类目词和所述扩展词进行物品召回的步骤具体包括：

根据滤除后的扩展词和所述检索类目词进行物品召回。

一种电子设备，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，所述处理器执行所述计算机程序时实现上述的物品召回方法。

一种可读存储介质，其上存储有计算机程序，所述程序被处理器执行时实现上述的物品召回方法的步骤。

一种物品召回系统，所述物品召回系统包括行为数据获取模块、输入模块、数据提

取模块、检索词获取模块、分词模块、第一计算模块、第一判断模块、扩展词生成模块和召回模块；

所述行为数据获取模块用于获取用户的行为数据，所述行为数据包括点击和/或购买的物品及每个物品的物品信息；

所述输入模块用于输入一检索类目词；

所述数据提取模块用于从所述行为数据中提取包含所述检索类目词的目标行为数据；

所述检索词获取模块用于逆向关联获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次；

所述分词模块用于对所述目标行为数据中物品信息进行分词，得到多个物品分词；

所述第一计算模块用于计算所有检索词与所述多个物品分词的相似度；

所述第一判断模块用于判断所述相似度是否大于第一预设阈值，若是，则调用所述扩展词生成模块；

所述扩展词生成模块用于提取所述所有检索词中检索频次大于预设频次的检索词作为所述检索类目词的扩展词；

所述召回模块用于根据所述检索类目词和所述扩展词进行物品召回。

较佳地，所述检索词获取模块用于根据所述行为数据生成关系列表，所述关系列表用于表征检索词与点击和/或购买的物品的对应关系，并基于所述关系列表获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次。

较佳地，所述第一计算模块包括集合生成单元和计算单元；

所述集合生成单元用于对所有的检索词进行去重生成检索词集合，还用于对所述多个物品分词进行去重生成物品分词集合；

所述计算单元用于基于 Jaccard 算法计算所述检索词集合和所述物品分词集合的相似度。

较佳地，所述物品召回系统还包括语料库构建模块、词向量库生成模块和第二计算模块；

所述语料库构建模块用于构建语料库，所述语料库包括物品推荐文章和物品基本信息；

所述词向量库生成模块用于基于 word2vec 对所述语料库进行训练并得到一词向量库，所述词向量库存储有多个物品特征词及与每个物品特征词对应的特征向量；

所述第二计算模块用于基于相似度算法计算所述检索类目词与每个物品特征词的相似度；

所述扩展词生成模块还用于提取与所述检索类目词的相似度大于第二预设阈值的物品特征词加入所述扩展词。

较佳地，所述物品召回系统还包括第二判断模块、滤除模块和一物品类目列表，所述物品类目列表包括多个物品类目及每个物品类目下所有物品的属性；

所述第二判断模块用于判断所述扩展词是否包含在所述物品类目列表中，若是，则调用所述滤除模块；

所述滤除模块用于滤除包含在所述类目列表的扩展词；

所述召回模块用于根据滤除后的扩展词和所述检索类目词进行物品召回。

本发明的积极进步效果在于：本发明在使用类目词进行物品召回时，利用获取的行为数据提取得到物品类目词的扩展词，在确保召回精准度的同时，提高了召回物品的覆盖率。

附图说明

图1为本发明实施例1的物品召回方法的流程图。

图2为本发明实施例1的物品召回方法中步骤60的流程图。

图3为本发明实施例2的物品召回方法的流程图。

图4为本发明实施例3的电子设备的结构示意图。

图5为本发明实施例5的物品召回系统的模块示意图。

图6为本发明实施例6的物品召回系统的模块示意图。

具体实施方式

下面通过实施例的方式进一步说明本发明，但并不因此将本发明限制在所述的实施例范围之中。

实施例1

一种物品召回方法，如图1所示，所述物品召回方法包括：

步骤10、获取用户的行为数据；行为数据包括点击和/或购买的物品及每个物品的物品信息；

步骤20、输入一检索类目词；

步骤30、从行为数据中提取包含检索类目词的目标行为数据；

步骤40、逆向关联获取目标行为数据中每个物品的检索词及每个检索词的检索频次；

步骤50、对目标行为数据中物品信息进行分词，得到多个物品分词；

步骤 60、计算所有检索词与多个物品分词的相似度；

步骤 70、判断相似度是否大于第一预设阈值，若是，则执行步骤 80；若否，则表示检索词均不能作为所述检索类目词的扩展词；

步骤 80、提取所有检索词中检索频次大于预设频次的检索词作为检索类目词的扩展词；

步骤 90、根据检索类目词和扩展词进行物品召回。

其中，步骤 40 具体包括：

根据所述行为数据生成关系列表，所述关系列表用于表征检索词与点击和/或购买的物品的对应关系，并基于所述关系列表获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次。

参见图 2，步骤 60 具体包括：

步骤 601、对所有的检索词进行去重生成检索词集合；

步骤 602、对多个物品分词进行去重生成物品分词集合；

步骤 603、基于 Jaccard 算法计算检索词集合和物品分词集合的相似度。

本实施例中，在使用类目词进行物品召回时，利用获取的行为数据提取得到物品类目词的扩展词，在确保召回精准度的同时，提高了召回物品的覆盖率。

实施例 2

本实施例的物品召回方法是在实施例 1 的基础上进一步改进，如图 3 所示，步骤 90 之前，所述物品召回方法还包括：

步骤 81、构建语料库，语料库包括物品推荐文章和物品基本信息；

步骤 82、基于 word2vec 对语料库进行训练，得到一词向量库；词向量库存储有多个物品特征词及与每个物品特征词对应的特征向量；

步骤 83、基于相似度算法计算检索类目词与每个物品特征词的相似度；

步骤 84、提取与检索类目词的相似度大于第二预设阈值的物品特征词加入扩展词。

本实施例中，考虑到是对物品类目的扩展，基于上述方法得到的扩展词很容易扩展到邻近类目或者上级类目或者其他现有的类目，比如，“牛肉”的扩展词中有“鸡肉”，香蕉的扩展词中有“菠萝”，“电脑桌”的扩展词中有“桌子”，如果需要召回现有的类目词对应的物品，只需选取对应的现有的类目词进行物品召回即可，在本申请中，可以通过现有类目表去过滤来排除掉这些扩展词，参见图 3，步骤 90 之前，所述物品召回方法还包括：

步骤 85、预设一物品类目列表，物品类目列表包括多个物品类目及每个物品类目下

所有物品的属性；

步骤 86、判断扩展词是否包含在物品类目列表中，若是，则执行步骤 87；

步骤 87、滤除包含在类目列表的扩展词；

进一步的，步骤 90 具体包括：

根据滤除后的扩展词和检索类目词进行物品召回。

本实施例中，利用各个平台的物品推荐文章、物品名称等作为语料，训练 word2vec 模型，并基于训练的模型产出特征词及对应的词向量，再基于相似度计算得到相似度最高的若干词作为检索类目词的扩展词，与上述利用行为数据进程检索类目词的扩展相辅相成，进一步增加召回的覆盖率。

实施例 3

一种电子设备，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，所述处理器执行所述计算机程序时实现实施例 1 或 2 中任意一个实施例所述的物品召回方法。

图 4 为本实施例提供的一种电子设备的结构示意图。图 4 示出了适于用来实现本发明实施方式的示例性电子设备 90 的框图。图 4 显示的电子设备 90 仅仅是一个示例，不应对本发明实施例的功能和使用范围带来任何限制。

如图 4 所示，电子设备 90 可以以通用计算设备的形式表现，例如其可以为服务器设备。电子设备 90 的组件可以包括但不限于：至少一个处理器 91、至少一个存储器 92、连接不同系统组件（包括存储器 92 和处理器 91）的总线 93。

总线 93 包括数据总线、地址总线和控制总线。

存储器 92 可以包括易失性存储器，例如随机存取存储器（RAM）921 和/或高速缓存存储器 922，还可以进一步包括只读存储器（ROM）923。

存储器 92 还可以包括具有一组（至少一个）程序模块 924 的程序工具 925，这样的程序模块 924 包括但不限于：操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。

处理器 91 通过运行存储在存储器 92 中的计算机程序，从而执行各种功能应用以及数据处理。

电子设备 90 也可以与一个或多个外部设备 94（例如键盘、指向设备等）通信。这种通信可以通过输入/输出（I/O）接口 95 进行。并且，电子设备 90 还可以通过网络适配器 96 与一个或者多个网络（例如局域网（LAN），广域网（WAN）和/或公共网络，例如因特网）通信。网络适配器 96 通过总线 93 与电子设备 90 的其它模块通信。应当明白，尽

管图中未示出，可以结合电子设备 90 使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理器、外部磁盘驱动阵列、RAID（磁盘阵列）系统、磁带驱动器以及数据备份存储系统等。

应当注意，尽管在上文详细描述中提及了电子设备的若干单元/模块或子单元/模块，但是这种划分仅仅是示例性的并非强制性的。实际上，根据本申请的实施方式，上文描述的两个或更多单元/模块的特征和功能可以在一个单元/模块中具体化。反之，上文描述的一个单元/模块的特征和功能可以进一步划分为由多个单元/模块来具体化。

实施例 4

一种计算机可读存储介质，其上存储有计算机程序，所述程序被处理器执行时实现实施例 1 或 2 中任意一个实施例所述的物品召回方法的步骤。

其中，可读存储介质可以采用的更具体可以包括但不限于：便携式盘、硬盘、随机存取存储器、只读存储器、可擦拭可编程只读存储器、光存储器件、磁存储器件或上述的任意合适的组合。

在可能的实施方式中，本发明还可以实现为一种程序产品的形式，其包括程序代码，当所述程序产品在终端设备上运行时，所述程序代码用于使所述终端设备执行实现实施例 1 或 2 中任意一个实施例所述的物品召回方法的步骤。

其中，可以以一种或多种程序设计语言的任意组合来编写用于执行本发明的程序代码，所述程序代码可以完全地在用户设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户设备上部分在远程设备上执行或完全在远程设备上执行。

实施例 5

一种物品召回系统，如图 5 所示，所述物品召回系统包括行为数据获取模块 11、输入模块 12、数据提取模块 13、检索词获取模块 14、分词模块 15、第一计算模块 16、第一判断模块 17、扩展词生成模块 18 和召回模块 19；

所述行为数据获取模块 11 用于获取用户的行为数据，所述行为数据包括点击和/或购买的物品及每个物品的物品信息；

所述输入模块 12 用于输入一检索类目词；

所述数据提取模块 13 用于从所述行为数据中提取包含所述检索类目词的目标行为数据；

所述检索词获取模块 14 用于逆向关联获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次；具体的，所述检索词获取模块 14 用于根据所述行为数据生成关系列表，所述关系列表用于表征检索词与点击和/或购买的物品的对应关系，并基于所

述关系列表获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次。

所述分词模块 15 用于对所述目标行为数据中物品信息进行分词,得到多个物品分词;

所述第一计算模块 16 用于计算所有检索词与所述多个物品分词的相似度;

其中,所述第一计算模块 16 包括集合生成单元 161 和计算单元 162;

所述集合生成单元 161 用于对所有的检索词进行去重生成检索词集合,还用于对所述多个物品分词进行去重生成物品分词集合;

所述计算单元 162 用于基于 Jaccard 算法计算所述检索词集合和所述物品分词集合的相似度。

所述第一判断模块 17 用于判断所述相似度是否大于第一预设阈值,若是,则调用所述扩展词生成模块 18;

所述扩展词生成模块 18 用于提取所述所有检索词中检索频次大于预设频次的检索词作为所述检索类目词的扩展词;

所述召回模块 19 用于根据所述检索类目词和所述扩展词进行物品召回。

本实施例中,在使用类目词进行物品召回时,利用获取的行为数据提取得到物品类目词的扩展词,在确保召回精准度的同时,提高了召回物品的覆盖率。

实施例 6

本实施例的物品召回系统是在实施例 5 的基础上进一步改进,如图 6 所示,所述物品召回系统还包括语料库构建模块 21、词向量库生成模块 22 和第二计算模块 23;

所述语料库构建模块 21 用于构建语料库,所述语料库包括物品推荐文章和物品基本信息;

所述词向量库生成模块 22 用于基于 word2vec 对所述语料库进行训练并得到一词向量库,所述词向量库存储有多个物品特征词及与每个物品特征词对应的特征向量;

所述第二计算模块 23 用于基于相似度算法计算所述检索类目词与每个物品特征词的相似度;

所述扩展词生成模块 18 还用于提取与所述检索类目词的相似度大于第二预设阈值的物品特征词加入所述扩展词。

本实施例中,考虑到是对物品类目的扩展,基于上述方法得到的扩展词很容易扩展到邻近类目或者上级类目或者其他现有的类目,比如,“牛肉”的扩展词中有“鸡肉”,香蕉的扩展词中有“菠萝”,“电脑桌”的扩展词中有“桌子”,如果需要召回现有的类目词对应的物品,只需选取对应的现有的类目词进行物品召回即可,在本申请中,可以通过现有类目表去过滤来排除掉这些扩展词,参见图 6,所述物品召回系统还包括第二判断模

块 25、滤除模块 26 和一物品类目列表 24，所述物品类目列表 24 包括多个物品类目及每个物品类目下所有物品的属性；

所述第二判断模块 25 用于判断所述扩展词是否包含在所述物品类目列表 24 中，若是，则调用所述滤除模块 26；

所述滤除模块 26 用于滤除包含在所述类目列表的扩展词；

所述召回模块 19 用于根据滤除后的扩展词和所述检索类目词进行物品召回。

本实施例中，利用各个平台的物品推荐文章、物品名称等作为语料，训练 word2vec 模型，并基于训练的模型产出特征词及对应的词向量，再基于相似度计算得到相似度最高的若干词作为检索类目词的扩展词，与上述利用行为数据进程检索类目词的扩展相辅相成，进一步增加召回的覆盖率。

虽然以上描述了本发明的具体实施方式，但是本领域的技术人员应当理解，这些仅是举例说明，在不背离本发明的原理和实质的前提下，可以对这些实施方式做出多种变更或修改。因此，本发明的保护范围由所附权利要求书限定。

权利要求

- 1.一种物品召回方法，其特征在于，所述物品召回方法包括：
获取用户的行为数据，所述行为数据包括点击和/或购买的物品及每个物品的物品信息；
输入一检索类目词；
从所述行为数据中提取包含所述检索类目词的目标行为数据；
逆向关联获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次；
对所述目标行为数据中物品信息进行分词，得到多个物品分词；
计算所有检索词与所述多个物品分词的相似度；
判断所述相似度是否大于第一预设阈值，若是，则提取所述所有检索词中检索频次大于预设频次的检索词作为所述检索类目词的扩展词；
根据所述检索类目词和所述扩展词进行物品召回。
- 2.如权利要求1所述的物品召回方法，其特征在于，所述逆向关联获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次的步骤具体包括：
根据所述行为数据生成关系列表，所述关系列表用于表征检索词与点击和/或购买的物品的对应关系；
基于所述关系列表获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次。
- 3.如权利要求1-2中至少一项所述的物品召回方法，其特征在于，所述计算所有检索词与所述多个物品分词的相似度的步骤具体包括：
对所有的检索词进行去重生成检索词集合；
对所述多个物品分词进行去重生成物品分词集合；
基于 Jaccard 算法计算所述检索词集合和所述物品分词集合的相似度。
- 4.如权利要求1-3中至少一项所述的物品召回方法，其特征在于，所述根据所述检索类目词和所述扩展词进行物品召回物品的步骤之前，所述物品召回方法还包括：
构建语料库，所述语料库包括物品推荐文章和物品基本信息；
基于 word2vec 对所述语料库进行训练，得到一词向量库，所述词向量库存储有多个物品特征词及与每个物品特征词对应的特征向量；
基于相似度算法计算所述检索类目词与每个物品特征词的相似度；
提取与所述检索类目词的相似度大于第二预设阈值的物品特征词加入所述扩展词。

5.如权利要求4所述的物品召回方法,其特征在于,所述根据所述检索类目词和所述扩展词进行物品召回物品的步骤之前,所述物品召回方法还包括:

预设一物品类目列表,所述物品类目列表包括多个物品类目及每个物品类目下所有物品的属性;

判断所述扩展词是否包含在所述物品类目列表中,若是,则滤除包含在所述类目列表的扩展词;

所述根据所述检索类目词和所述扩展词进行物品召回的步骤具体包括:

根据滤除后的扩展词和所述检索类目词进行物品召回。

6.一种电子设备,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,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至5任一项所述的物品召回方法。

7.一种可读存储介质,其上存储有计算机程序,其特征在于,所述程序被处理器执行时实现权利要求1至5任一项所述的物品召回方法的步骤。

8.一种物品召回系统,其特征在于,所述物品召回系统包括行为数据获取模块、输入模块、数据提取模块、检索词获取模块、分词模块、第一计算模块、第一判断模块、扩展词生成模块和召回模块;

所述行为数据获取模块用于获取用户的行为数据,所述行为数据包括点击和/或购买的物品及每个物品的物品信息;

所述输入模块用于输入一检索类目词;

所述数据提取模块用于从所述行为数据中提取包含所述检索类目词的目标行为数据;

所述检索词获取模块用于逆向关联获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次;

所述分词模块用于对所述目标行为数据中物品信息进行分词,得到多个物品分词;

所述第一计算模块用于计算所有检索词与所述多个物品分词的相似度;

所述第一判断模块用于判断所述相似度是否大于第一预设阈值,若是,则调用所述扩展词生成模块;

所述扩展词生成模块用于提取所述所有检索词中检索频次大于预设频次的检索词作为所述检索类目词的扩展词;

所述召回模块用于根据所述检索类目词和所述扩展词进行物品召回。

9.如权利要求8所述的物品召回系统,其特征在于,所述检索词获取模块用于根据所述行为数据生成关系列表,所述关系列表用于表征检索词与点击和/或购买的物品的对应

关系，并基于所述关系列表获取所述目标行为数据中每个物品的检索词及每个检索词的检索频次。

10.如权利要求 8-9 中至少一项所述的物品召回系统，其特征在于，所述第一计算模块包括集合生成单元和计算单元；

所述集合生成单元用于对所有的检索词进行去重生成检索词集合，还用于对所述多个物品分词进行去重生成物品分词集合；

所述计算单元用于基于 Jaccard 算法计算所述检索词集合和所述物品分词集合的相似度。

11.如权利要求 8-10 中至少一项所述的物品召回系统，其特征在于，所述物品召回系统还包括语料库构建模块、词向量库生成模块和第二计算模块；

所述语料库构建模块用于构建语料库，所述语料库包括物品推荐文章和物品基本信息；

所述词向量库生成模块用于基于 word2vec 对所述语料库进行训练并得到一词向量库，所述词向量库存储有多个物品特征词及与每个物品特征词对应的特征向量；

所述第二计算模块用于基于相似度算法计算所述检索类目词与每个物品特征词的相似度；

所述扩展词生成模块还用于提取与所述检索类目词的相似度大于第二预设阈值的物品特征词加入所述扩展词。

12.如权利要求 11 所述的物品召回系统，其特征在于，所述物品召回系统还包括第二判断模块、滤除模块和一物品类目列表，所述物品类目列表包括多个物品类目及每个物品类目下所有物品的属性；

所述第二判断模块用于判断所述扩展词是否包含在所述物品类目列表中，若是，则调用所述滤除模块；

所述滤除模块用于滤除包含在所述类目列表的扩展词；

所述召回模块用于根据滤除后的扩展词和所述检索类目词进行物品召回。

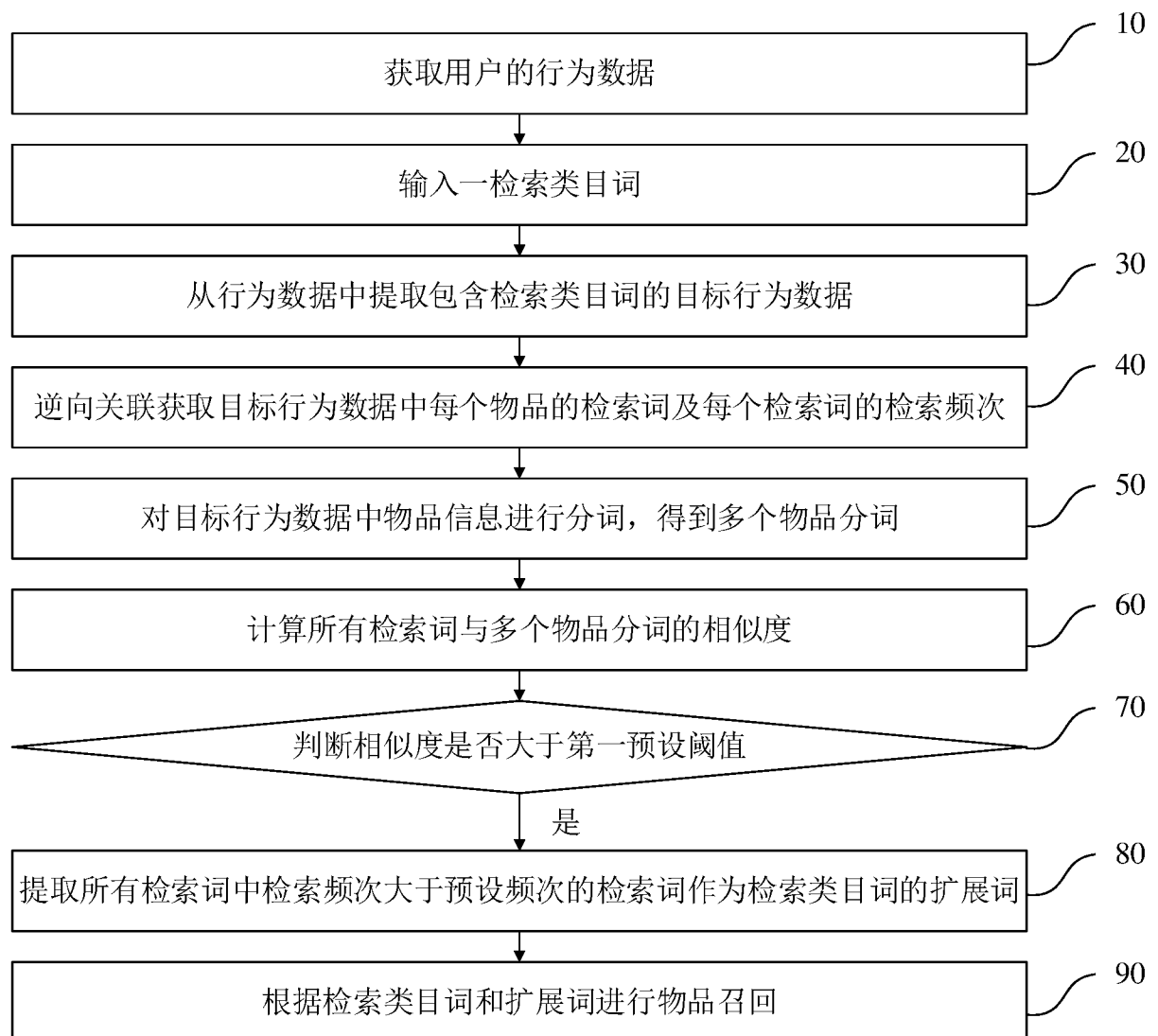


图 1

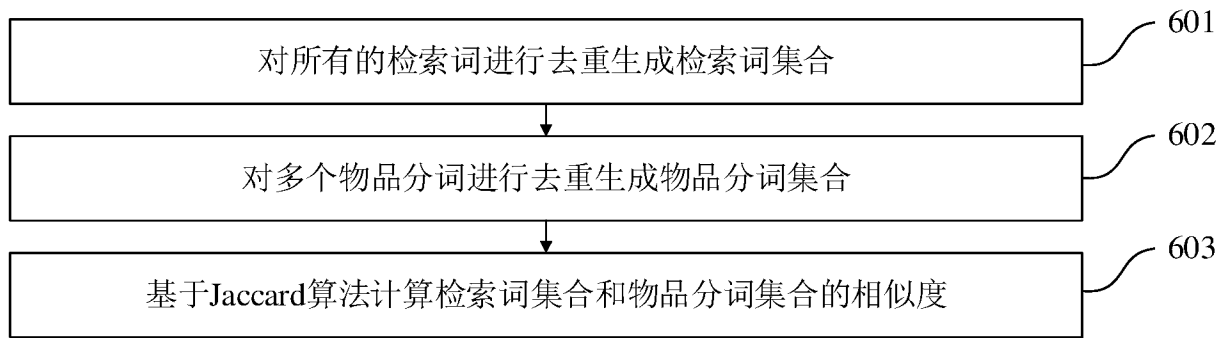


图 2

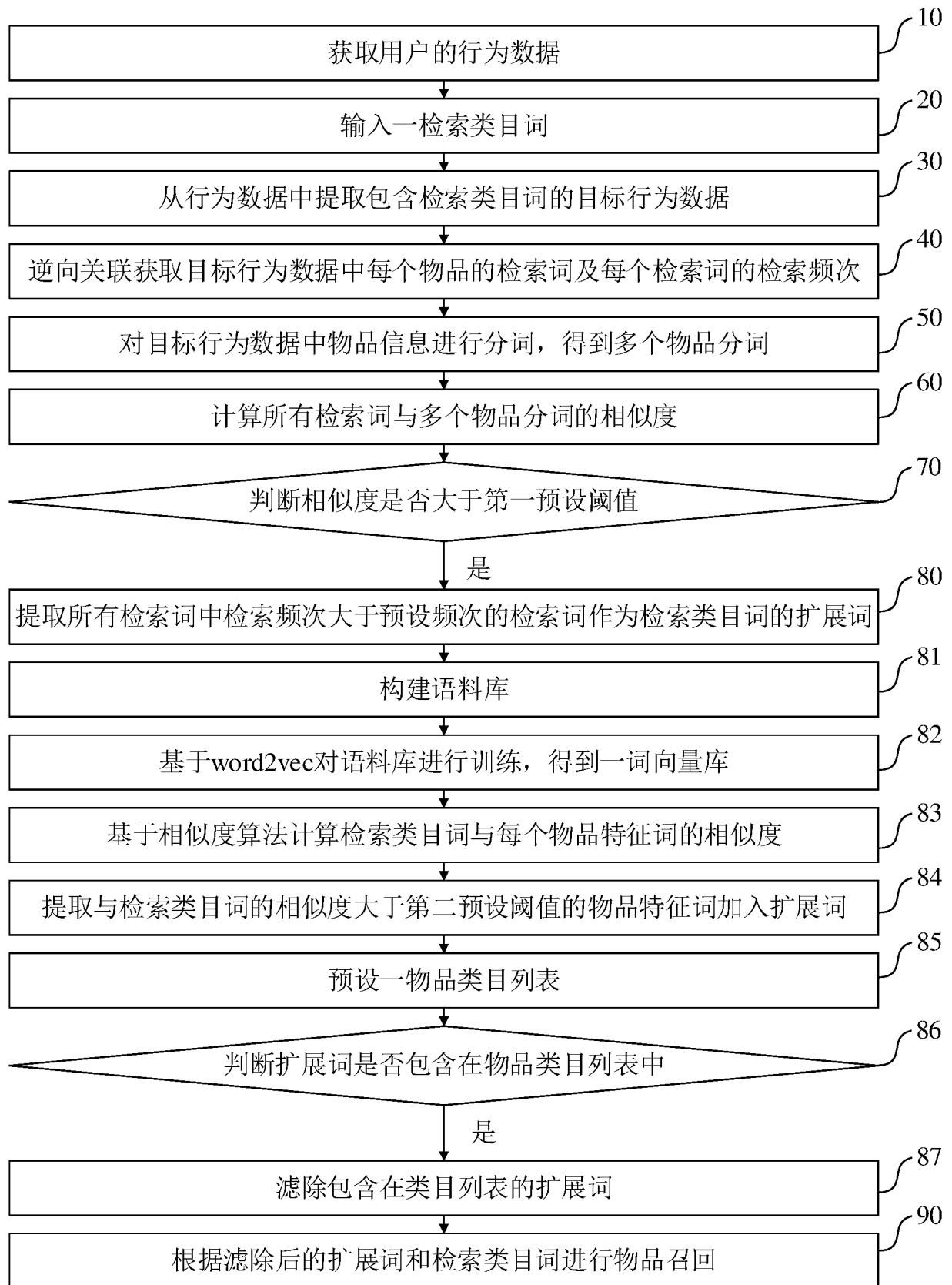


图 3

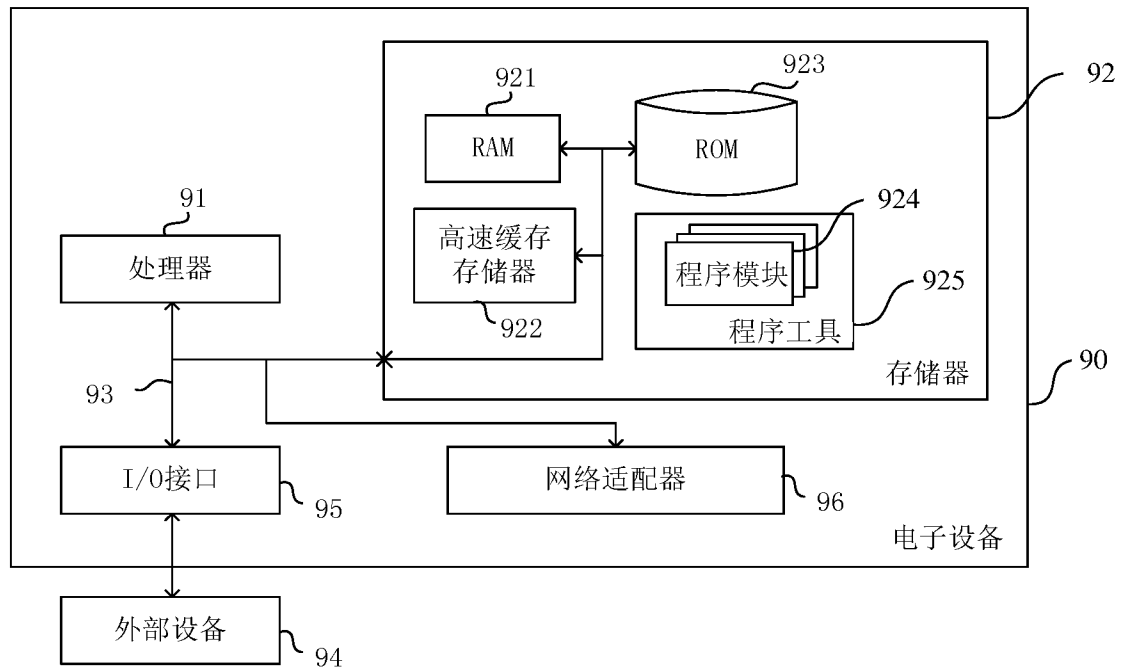


图 4

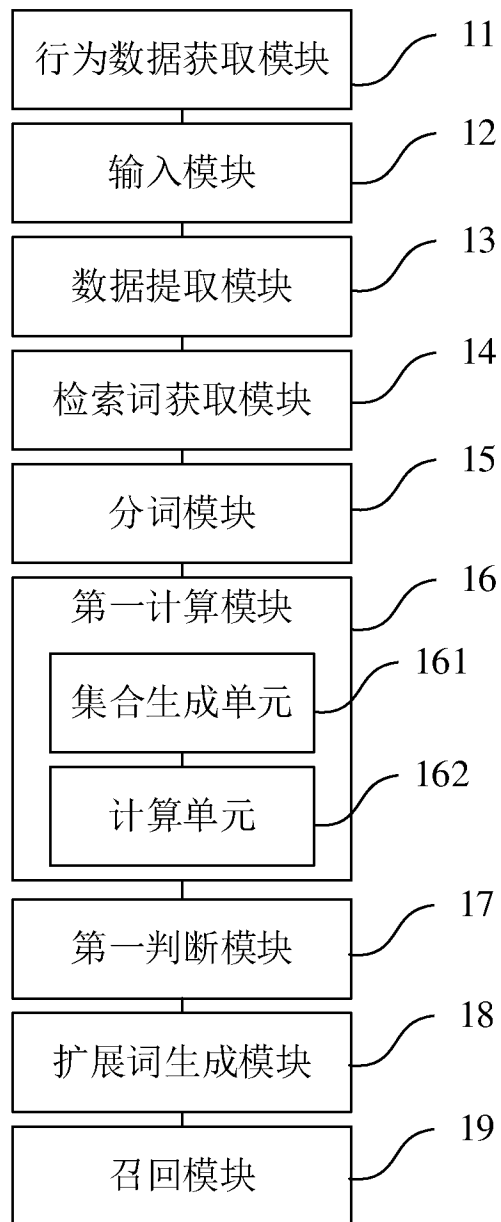


图 5

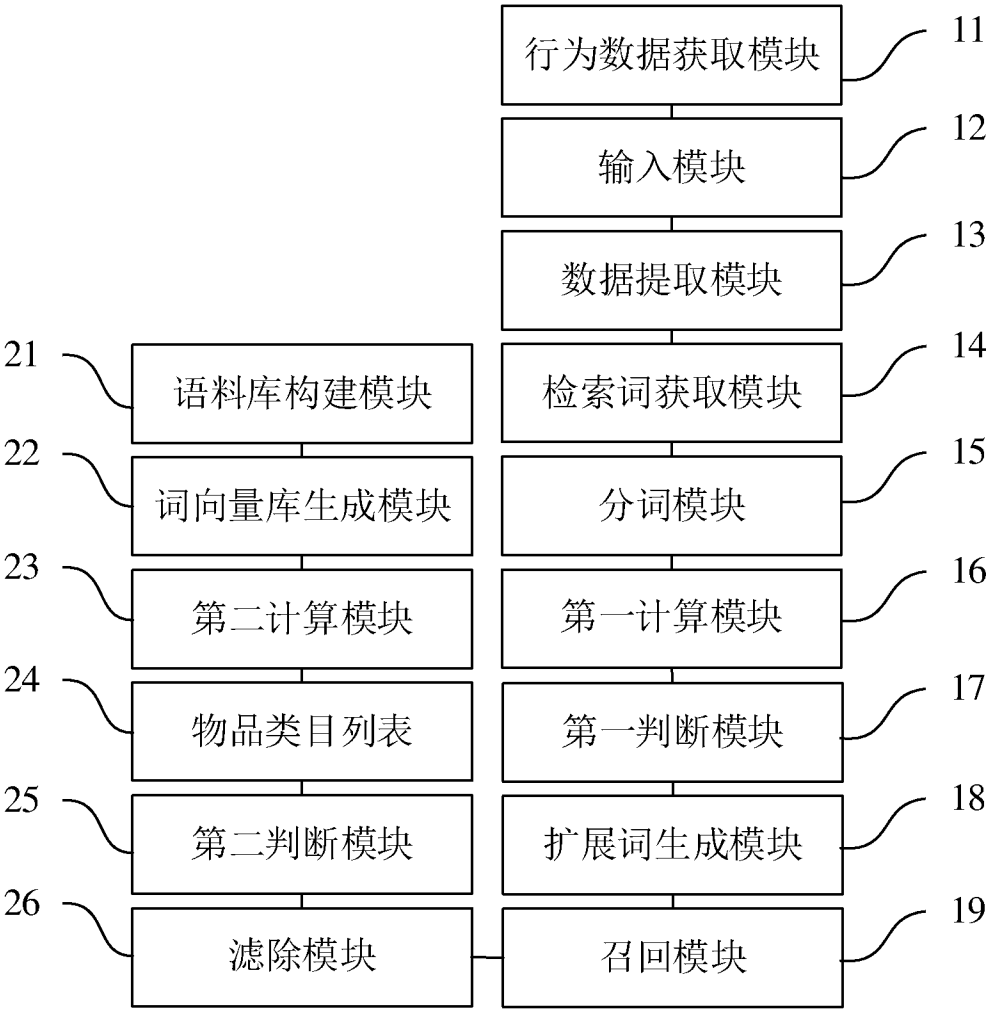


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/070014

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/33(2019.01)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)

G06F

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, GOOGLE, IEEE: 检索, 召回, 扩展, 关键词, 类目, 分词, 历史, 频次, 相似, search+, correspond+, keyword?, segment+, similar, train+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 106874492 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD. et al.) 20 June 2017 (2017-06-20) description, paragraphs [0018]-[0040], and figures 2-4	1-3, 6-10
Y	CN 106874492 A (BEIJING JINGDONG SHANGKE INFORMATION TECHNOLOGY CO., LTD. et al.) 20 June 2017 (2017-06-20) description, paragraphs [0018]-[0040], and figures 2-4	4, 5, 11, 12
Y	CN 108664637 A (WELLONG ETOWN INTERNATIONAL LOGISTICS CO., LTD.) 16 October 2018 (2018-10-16) description, paragraphs [0026]-[0051], and figures 1 and 2	4, 5, 11, 12
A	CN 104636334 A (ALIBABA GROUP HOLDING LIMITED) 20 May 2015 (2015-05-20) entire document	1-12
A	US 2008162118 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 03 July 2008 (2008-07-03) entire document	1-12



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

16 March 2020

Date of mailing of the international search report

26 March 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/070014

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106874492	A	20 June 2017	None			
CN	108664637	A	16 October 2018	None			
CN	104636334	A	20 May 2015	None			
US	2008162118	A1	03 July 2008	JP	2008151926	A	03 July 2008

国际检索报告

国际申请号

PCT/CN2020/070014

A. 主题的分类

G06F 16/33 (2019.01) 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)

G06F

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, EPDOC, GOOGLE, IEEE: 检索, 召回, 扩展, 关键词, 类目, 分词, 历史, 频次, 相似, search+, correspond+, keyword?, segment+, similar, train+

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 106874492 A (北京京东尚科信息技术有限公司 等) 2017年 6月 20日 (2017 - 06 - 20) 说明书第[0018]-[0040]段, 附图2-4	1-3、6-10
Y	CN 106874492 A (北京京东尚科信息技术有限公司 等) 2017年 6月 20日 (2017 - 06 - 20) 说明书第[0018]-[0040]段, 附图2-4	4-5、11-12
Y	CN 108664637 A (惠龙易通国际物流股份有限公司) 2018年 10月 16日 (2018 - 10 - 16) 说明书第[0026]-[0051]段, 附图1-2	4-5、11-12
A	CN 104636334 A (阿里巴巴集团控股有限公司) 2015年 5月 20日 (2015 - 05 - 20) 全文	1-12
A	US 2008162118 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2008年 7月 3日 (2008 - 07 - 03) 全文	1-12

☐ 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

2020年 3月 16日

国际检索报告邮寄日期

2020年 3月 26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

授权官员

吴瑶

电话号码 86-10-53961339

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/070014

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	106874492	A	2017年 6月 20日	无		
CN	108664637	A	2018年 10月 16日	无		
CN	104636334	A	2015年 5月 20日	无		
US	2008162118	A1	2008年 7月 3日	JP	2008151926	A 2008年 7月 3日