



(12) 发明专利申请

(10) 申请公布号 CN 113987153 A

(43) 申请公布日 2022. 01. 28

(21) 申请号 202111306304.7

(22) 申请日 2021.11.05

(71) 申请人 京东科技信息技术有限公司

地址 100176 北京市北京经济技术开发区

科创十一街18号院2号楼6层601

(72) 发明人 宋阳 陈蒙

(74) 专利代理机构 北京清亦华知识产权代理事

务所(普通合伙) 11201

代理人 单冠飞

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 16/33 (2019.01)

G06F 40/279 (2020.01)

G06Q 30/06 (2012.01)

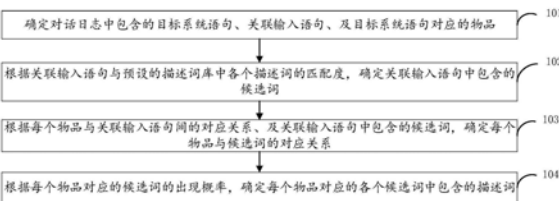
权利要求书3页 说明书13页 附图3页

(54) 发明名称

物品描述词的确定方法、装置、计算机设备及存储介质

(57) 摘要

本公开提出一种物品描述词的确定方法、装置、计算机设备及存储介质,涉及计算机技术领域。该方法包括:确定对话日志中包含的目标系统语句、关联输入语句、及所述目标系统语句对应的物品;根据所述关联输入语句与预设的描述词库中各个描述词的匹配度,确定所述关联输入语句中包含的候选词;根据每个物品与关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系;根据每个所述物品对应的候选词的出现概率,确定每个所述物品对应的各个候选词中包含的描述词。由此,通过对对话日志自动进行提取处理,即可生成对应的描述词,描述词的生成效率更高,所得描述词也更加准确和可靠。



1. 一种物品描述词的确定方法,其特征在于,包括:

确定对话日志中包含的目标系统语句、关联输入语句、及所述目标系统语句对应的物品;

根据所述关联输入语句与预设的描述词库中各个描述词的匹配度,确定所述关联输入语句中包含的候选词;

根据每个物品与所述关联输入语句间的对应关系、及所述关联输入语句中包含的候选词,确定每个所述物品与候选词的对应关系;

根据每个所述物品对应的候选词的出现概率,确定每个所述物品对应的各个候选词中包含的描述词。

2. 如权利要求1所述的方法,其特征在于,所述确定对话日志中包含的目标系统语句、关联输入语句、及所述目标系统语句对应的物品,包括:

将所述对话日志中用于指向物品所在页面的系统语句,确定为目标系统语句;

根据所述目标系统语句指向的物品所在页面的描述信息,确定所述物品;

将接收时刻位于所述目标系统语句之前的预设数量的输入语句,确定为所述关联输入语句。

3. 如权利要求1所述的方法,其特征在于,所述根据所述关联输入语句与预设的描述词库中各个描述词的匹配度,确定所述关联输入语句中包含的候选词,包括:

将所述关联输入语句进行切词,以确定所述关联输入语句中包含的各个分词;

在任一分词与各个所述描述词间的匹配度均小于第一阈值的情况下,确定所述任一分词为候选词。

4. 如权利要求1所述的方法,其特征在于,所述根据每个所述物品对应的候选词的出现概率,确定每个所述物品对应的各个候选词中包含的描述词,包括:

确定每个候选词在所述对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率;

根据所述第一出现概率和/或所述第二出现概率,确定每个所述物品对应的各个候选词中包含的描述词。

5. 如权利要求1所述的方法,其特征在于,所述根据每个所述物品对应的候选词的出现概率,确定每个所述物品对应的各个候选词中包含的描述词,包括:

确定每个候选词在所述对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率;

根据各个所述第一出现概率,确定每个所述候选词在所述对话日志中的第一高斯分布值;

根据各个所述第二出现概率,确定每个所述物品对应的候选词在关联输入语句中的第二高斯分布值;

根据所述第一出现概率、第二出现概率、第一高斯分布值及第二高斯分布值,确定每个所述物品对应的各个候选词中包含的描述词。

6. 如权利要求5所述的方法,其特征在于,所述根据所述第一出现概率、第二出现概率、第一高斯分布值及第二高斯分布值,确定每个所述物品对应的各个候选词中包含的描述词,包括:

响应于任一物品对应的任一候选词对应的第二出现概率大于第一出现概率、且第二高斯分布值与第一高斯分布值间的比值大于第二阈值,确定所述任一候选词为所述任一物品的描述词。

7.如权利要求1-6任一所述的方法,其特征在于,还包括:

确定每个所述物品的描述词在所述关联输入语句中的第二出现概率;

根据所述第二出现概率,确定每个所述物品的描述词的权重值。

8.如权利要求7所述的方法,其特征在于,在所述确定每个所述物品的描述词的权重值之后,还包括:

接收查询语句;

基于描述词库对所述查询语句进行处理,以确定所述查询语句中包含的目标描述词;

根据预设的描述词与物品的关联关系,确定所述目标描述词关联的候选物品;

根据所述目标描述词在每个所述候选物品描述词中的权重,确定目标物品;

返回所述目标物品。

9.一种物品描述词的确定装置,其特征在于,包括:

第一确定模块,用于确定对话日志中包含的目标系统语句、关联输入语句、及所述目标系统语句对应的物品;

第二确定模块,用于根据所述关联输入语句与预设的描述词库中各个描述词的匹配度,确定所述关联输入语句中包含的候选词;

第三确定模块,用于根据每个物品与所述关联输入语句间的对应关系、及所述关联输入语句中包含的候选词,确定每个所述物品与候选词的对应关系;

第四确定模块,用于根据每个所述物品对应的候选词的出现概率,确定每个所述物品对应的各个候选词中包含的描述词。

10.如权利要求9所述的装置,其特征在于,所述第一确定模块,具体用于:

将所述对话日志中用于指向物品所在页面的系统语句,确定为目标系统语句;

根据所述目标系统语句指向的物品所在页面的描述信息,确定所述物品;

将接收时刻位于所述目标系统语句之前的预设数量的输入语句,确定为所述关联输入语句。

11.如权利要求9所述的装置,其特征在于,所述第二确定模块,具体用于:

将所述关联输入语句进行切词,以确定所述关联输入语句中包含的各个分词;

在任一分词与各个所述描述词间的匹配度均小于第一阈值的情况下,确定所述任一分词为候选词。

12.如权利要求9所述的装置,其特征在于,所述第四确定模块,具体用于:

确定每个候选词在所述对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率;

根据所述第一出现概率和/或所述第二出现概率,确定每个所述物品对应的各个候选词中包含的描述词。

13.如权利要求9所述的装置,其特征在于,所述第四确定模块,包括:

第一确定单元,用于确定每个候选词在所述对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率;

第二确定单元,用于根据各个所述第一出现概率,确定每个所述候选词在所述对话日志中的第一高斯分布值;

第三确定单元,用于根据各个所述第二出现概率,确定每个所述物品对应的候选词在关联输入语句中的第二高斯分布值;

第四确定单元,用于根据所述第一出现概率、第二出现概率、第一高斯分布值及第二高斯分布值,确定每个所述物品对应的各个候选词中包含的描述词。

14.如权利要求13所述的装置,其特征在于,所述第四确定单元,具体用于:

响应于任一物品对应的任一候选词对应的第二出现概率大于第一出现概率、且第二高斯分布值与第一高斯分布值间的比值大于第二阈值,确定所述任一候选词为所述任一物品的描述词。

15.如权利要求9-14任一所述的装置,其特征在于,所述第一确定模块,还用于:

确定每个所述物品的描述词在所述关联输入语句中的第二出现概率;

根据所述第二出现概率,确定每个所述物品的描述词的权重值。

16.如权利要求15所述的装置,其特征在于,还包括:

接收模块,用于接收查询语句;

处理模块,用于基于描述词库对所述查询语句进行处理,以确定所述查询语句中包含的目标描述词;

第五确定模块,用于根据预设的描述词与物品的关联关系,确定所述目标描述词关联的候选物品;

第六确定模块,用于根据所述目标描述词在每个所述候选物品描述词中的权重,确定目标物品;

返回模块,用于返回所述目标物品。

17.一种计算机设备,其特征在于,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程序时,实现如权利要求1-8中任一所述的物品描述词的确定方法。

18.一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1-8中任一所述的物品描述词的确定方法。

19.一种计算机程序产品,其特征在于,包括计算机程序,所述计算机程序在被处理器执行时实现如权利要求1-8中任一项所述的物品描述词的确定方法。

物品描述词的确定方法、装置、计算机设备及存储介质

技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及一种物品描述词的确定方法、装置、计算机设备及存储介质。

背景技术

[0002] 随着互联网技术的快速发展,各种电商平台提供了多种多样的线上商品交易渠道,极大地方便了人们的工作和生活。在电商咨询对话系统中,可以根据用户输入的语句,通过查找各个物品对应的描述词,从而确定出用户想要了解的物品。

[0003] 相关技术中,通常需要人工先确定出各个物品对应的描述词,商品数量众多,需要耗费大量的人工成本,同时,人工总结的效率也不高。

[0004] 由此,如何快速准确的确定出物品描述词,显得至关重要。

发明内容

[0005] 本公开旨在至少在一定程度上解决相关技术中的技术问题之一。

[0006] 本公开第一方面实施例提出了一种物品描述词的确定方法,包括:

[0007] 确定对话日志中包含的目标系统语句、关联输入语句、及所述目标系统语句对应的物品;

[0008] 根据所述关联输入语句与预设的描述词库中各个描述词的匹配度,确定所述关联输入语句中包含的候选词;

[0009] 根据每个物品与所述关联输入语句间的对应关系、及所述关联输入语句中包含的候选词,确定每个所述物品与候选词的对应关系;

[0010] 根据每个所述物品对应的候选词的出现概率,确定每个所述物品对应的各个候选词中包含的描述词。

[0011] 本公开第二方面实施例提出了一种物品描述词的确定装置,包括:

[0012] 第一确定模块,用于确定对话日志中包含的目标系统语句、关联输入语句、及所述目标系统语句对应的物品;

[0013] 第二确定模块,用于根据所述关联输入语句与预设的描述词库中各个描述词的匹配度,确定所述关联输入语句中包含的候选词;

[0014] 第三确定模块,用于根据每个物品与所述关联输入语句间的对应关系、及所述关联输入语句中包含的候选词,确定每个所述物品与候选词的对应关系;

[0015] 第四确定模块,用于根据每个所述物品对应的候选词的出现概率,确定每个所述物品对应的各个候选词中包含的描述词。

[0016] 本公开第三方面实施例提出了一种计算机设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时,实现如本公开第一方面实施例提出的物品描述词的确定方法。

[0017] 本公开第四方面实施例提出了一种非临时性计算机可读存储介质,存储有计算机

程序,所述计算机程序被处理器执行时实现如本公开第一方面实施例提出的物品描述词的确定方法。

[0018] 本公开第五方面实施例提出了一种计算机程序产品,当所述计算机程序产品中的指令处理器执行时,执行本公开第一方面实施例提出的物品描述词的确定方法。

[0019] 本公开提供的物品描述词的确定方法、装置、计算机设备及存储介质,可以先确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品,之后根据关联输入语句与预设的描述词库中各个描述词的匹配度,确定关联输入语句中包含的候选词,再根据每个物品与所述关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系,之后即可根据每个物品对应的候选词的出现概率,确定每个物品对应的各个候选词中包含的描述词。由此,通过对对话日志自动进行提取处理,即可生成对应的描述词,描述词的生成效率更高,所得描述词也更加准确和可靠。

[0020] 本公开附加的方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本公开的实践了解到。

附图说明

[0021] 本公开上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0022] 图1为本公开一实施例所提供的物品描述词的确定方法的流程示意图;

[0023] 图2为本公开另一实施例所提供的物品描述词的确定方法的流程示意图;

[0024] 图3为本公开另一实施例所提供的物品描述词的确定方法的流程示意图;

[0025] 图4为本公开另一实施例所提供的物品描述词的确定方法的流程示意图;

[0026] 图5为本公开一实施例所提供的物品描述词的确定装置的结构示意图;

[0027] 图6示出了适于用来实现本公开实施方式的示例性计算机设备的框图。

具体实施方式

[0028] 下面详细描述本公开的实施例,所述实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,旨在用于解释本公开,而不能理解为对本公开的限制。

[0029] 下面参考附图描述本公开实施例的物品描述词的确定方法、装置、计算机设备和存储介质。

[0030] 图1为本公开实施例所提供的物品描述词的确定方法的流程示意图。

[0031] 本公开实施例以该物品描述词的确定方法被配置于物品描述词的确定装置中来举例说明,该物品描述词的确定装置可以应用于任一计算机设备中,以使该计算机设备可以执行物品描述词的确定功能。

[0032] 其中,计算机设备可以为个人电脑(Personal Computer,简称PC)、云端设备、移动设备等,移动设备例如可以为手机、平板电脑、个人数字助理、穿戴式设备、车载设备等具有各种操作系统、触摸屏和/或显示屏的硬件设备。

[0033] 为方便说明,物品描述词的确定装置,可以简记为“确定装置”。

[0034] 如图1所示,该物品描述词的确定方法可以包括以下步骤:

[0035] 步骤101,确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品。

[0036] 其中,对话日志中可以包含多个语句,比如可以为系统语句、输入语句等等,本公开对此不做限定。

[0037] 比如,对话日志中包含的语句可以为:推荐一款老年机;大概价位是多少;2000左右吧;请问有心仪的品牌吗;没有。好的请稍等;请点击下方链接查看推荐物品详情:XX手机¥1899…等等,本公开对此不做限定。

[0038] 可选的,可以将对话日志中用于指向物品所在页面的系统语句,确定为目标系统语句。

[0039] 比如,上述示例中的“请点击下方链接查看推荐物品详情:XX手机¥1899…”,即为可以指向物品所在页面的系统语句,从而可以将其作为目标系统语句。

[0040] 可选的,可以根据目标系统语句指向的物品所在页面的描述信息,确定目标系统语句对应的物品。

[0041] 比如,物品所在页面的描述信息为:XX品牌Y系列9代手机、一亿像素、支持快充、¥1899等等,则根据该描述信息,可以确定目标系统语句对应的物品为:XX手机Y系列9代手机。

[0042] 可选的,可以将接收时刻位于目标系统语句之前的预设数量的输入语句,确定为关联输入语句。

[0043] 其中,预设数量可以为提前设定好的数值,比如可以为3、5等等,本公开对此不做限定。

[0044] 比如,预设数量为3。则可以将上述示例中的输入语句“推荐一款老年机”、“2000左右吧”、“没有”,确定为关联输入语句。

[0045] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中对话日志中包含的语句、目标系统语句、关联输入语句、以及目标系统语句对应的物品等的限定。

[0046] 可以理解的是,同一时刻,在对话日志中可能存在多个目标系统语句,则可以分别确定各个目标系统语句、各个目标系统语句对应的物品以及关联输入语句等等,本公开对此不做限定。

[0047] 可以理解的是,对话日志中的内容较为全面,也更加贴近用户的表达与感受,从而本公开实施例中,通过对对话日志进行挖掘与处理,可使得确定出的内容也更加准确,更加符合用户需求。

[0048] 步骤102,根据关联输入语句与预设的描述词库中各个描述词的匹配度,确定关联输入语句中包含的候选词。

[0049] 其中,关联输入语句中包含的候选词,可以为一个,或者也可以为多个等等,本公开对此不做限定。

[0050] 可以理解的是,可以提前对大量物品的详情页面的内容进行解析处理,将处理后得到的结果作为预设的描述词库中的描述词。

[0051] 另外,可以先对关联输入语句进行解析,以得到其包含的分词,还可以将其中与物品描述无关的分词进行剔除,之后再将剩余分词与预设的描述词库中的描述词进行比对,从而确定关联输入语句中包含的候选词。

[0052] 比如说,通过对关联输入语句进行解析,确定出的分词分别为:你好、风扇、立柜式、大功率、可移动。其中,“你好”与物品描述不相关,则可以先将其剔除,之后将剩余的分词“风扇、立柜式、大功率、可移动”与预设的描述词库中的各个描述词分别进行匹配。若预设的描述词库中的描述词不包含“可移动”,即匹配度为零,则可以确定“可移动”即为关联输入语句中包含的候选词。

[0053] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中关联输入语句及其包含的候选词等的限定。

[0054] 本公开实施例中,可以先将关联输入语句进行解析,之后再与预设的描述词库中各个描述词进行匹配,以确定出对应的匹配度,之后根据各个匹配度,可以确定出对应的候选词,从而提高了确定的候选词的准确性和可靠性。

[0055] 步骤103,根据每个物品与关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系。

[0056] 可以理解的是,对话日志中的关联输入语句可能为一个,或者也可能为多个;不同的关联输入语句可能均对应于同一物品,或者也可能对应于不同物品等等;各个关联输入语句中包含的候选词可能相同,或者也可能不相同等等,本公开对此不做限定。

[0057] 比如,关联输入语句1对应于物品1、关联输入语句2对应于物品2、关联输入语句3对应于物品1,而关联输入语句1的候选词为A,关联输入语句2的候选词为B,关联输入语句3的候选词为C,则可以确定物品1与候选词A、候选词C相对应,物品2与候选词B对应。

[0058] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中物品、关联输入语句以及候选词间的对应关系等的限定。

[0059] 步骤104,根据每个物品对应的候选词的出现概率,确定每个物品对应的各个候选词中包含的描述词。

[0060] 其中,描述词,可以为可对物品进行描述的任意的词。比如可以为:老人用、冬天穿、会滑动、手感好等等,本公开对此不做限定。

[0061] 可选的,可以将物品对应的出现概率最大的候选词,确定为该物品对应的描述词。或者,也可以将物品对应的出现概率大于阈值的各候选词,确定为该物品对应的描述词。或者,还可以按照物品对应的出现概率的大小对各个候选词进行排序,可将前预设数量的候选词,确定为该物品对应的描述词。

[0062] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中确定每个物品对应的各个候选词中包含的描述词的方式等的限定。

[0063] 可以理解的是,本公开实施例中,通过对对话日志自动进行挖掘处理,即可确定出对应的描述词,从而可使得确定出的描述词,更加贴合用户的表达,更加准确,从而可以更符合用户需求。

[0064] 本公开实施例,可以先确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品,之后根据关联输入语句与预设的描述词库中各个描述词的匹配度,确定关联输入语句中包含的候选词,再根据每个物品与所述关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系,之后即可根据每个物品对应的候选词的出现概率,确定每个物品对应的各个候选词中包含的描述词。由此,通过对对话日志自动进行提取处理,即可生成对应的描述词,描述词的生成效率更高,所得

描述词也更加准确和可靠。

[0065] 图2为本公开一实施例所提供的物品描述词的确定方法的流程示意图,如图2所示,该物品描述词的确定方法可以包括以下步骤:

[0066] 步骤201,确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品。

[0067] 步骤202,将关联输入语句进行切词,以确定关联输入语句中包含的各个分词。

[0068] 比如说,关联输入语句为“推荐一款老年机”、“2000左右吧”、“粗糙一点的,不容易掉”,通过对关联输入语句进行切词处理,可以确定其中包含的分词为:老年机、2000左右、粗糙、不容易掉等等,本公开对此不做限定。

[0069] 可以理解的是,本公开实施例中,可以采用任何可取的方式对关联输入语句进行切词处理,本公开对此不做限定。

[0070] 步骤203,在任一分词与各个描述词间的匹配度均小于第一阈值的情况下,确定任一分词为候选词。

[0071] 其中,第一阈值可以为提前设定好的阈值,比如可以为0.15、0.2等等,本公开对此不做限定。

[0072] 比如说,第一阈值为0.25。任一分词A与各个描述词间的匹配度均小于第一阈值0.25,则可以确定该任一分词A为候选词等等,本公开对此不做限定。

[0073] 可以理解的是,本公开实施例中,可以采用任何可取的方式确定任一分词与各个描述词间的匹配度,本公开对此不做限定。

[0074] 步骤204,根据每个物品与关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系。

[0075] 步骤205,确定每个候选词在对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率。

[0076] 比如说,对话日志中共有100个分词,候选词B出现的次数为20,则该候选词B对应的第一出现概率可以为0.2;候选词B在物品1关联的输入语句中出现的次数为2,物品1关联的输入语句中共有5个分词,则该候选词B对应的第二概率为0.4。

[0077] 需要说明的是,上述示例只是示意性说明,不能作为对本公开实施例中第一出现概率、第二出现概率等的限定。

[0078] 步骤206,根据第一出现概率和/或第二出现概率,确定每个物品对应的各个候选词中包含的描述词。

[0079] 可选的,可以将第一出现概率和第二出现概率相加,之后将所得结果最大的候选词,确定为与该候选词对应的物品包含的描述词。或者,也可以将第二出现概率大于第一出现概率的候选词,确定为与该候选词对应的物品包含的描述词。

[0080] 或者,也可以将第二出现概率最大的候选词,确定为与该候选词对应的物品包含的描述词。或者,还可以将第一出现概率最小的候选词,确定为与该候选词对应的物品包含的描述词。

[0081] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中确定每个物品对应的各个候选词中包含的描述词的方式等的限定。

[0082] 可选的,在实际实现过程中,在确定出每个物品的描述词之后,还可以先确定每个

物品的描述词在关联输入语句中的第二出现概率,之后根据第二出现概率,确定每个物品的描述词的权重值。

[0083] 可以理解的是,描述词可以为候选词中的一部分,从而可以根据确定候选词在每个物品关联的输入语句中的第二出现概率,直接确定出对应的描述词的第二出现概率。

[0084] 可以理解的是,第二出现概率越大,描述词的权重值也越大;相应的,第二出现概率越小,描述词的权重值也越小。

[0085] 从而,本公开实施例中,在确定出描述词之后,还可以根据描述词对应的第二出现概率,确定每个物品的描述词的权重值。从而确定装置在进行物品推荐时,可以充分考虑到各个描述词对应的权重,再基于各权重进行物品推荐,可使得推荐的物品更加可靠、准确,符合用户需求

[0086] 本公开实施例,可以先确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品,之后可以将关联输入语句进行切词,以确定关联输入语句中包含的各个分词,在任一分词与各个描述词间的匹配度均小于第一阈值的情况下,确定任一分词为候选词,之后可以根据每个物品与关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系,之后可以先确定每个候选词在对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率,再根据第一出现概率和/或第二出现概率,确定每个物品对应的各个候选词中包含的描述词。由此,通过对对话日志自动进行提取处理,即可生成对应的描述词,描述词的生成效率更高,所得描述词也更加准确和可靠。

[0087] 图3为本公开一实施例所提供的物品描述词的确定方法的流程示意图,如图3所示,该物品描述词的确定方法可以包括以下步骤:

[0088] 步骤301,确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品。

[0089] 步骤302,根据关联输入语句与预设的描述词库中各个描述词的匹配度,确定关联输入语句中包含的候选词。

[0090] 步骤303,根据每个物品与关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系。

[0091] 步骤304,确定每个候选词在对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率。

[0092] 需要说明的是,步骤301至步骤304的具体内容及实现方式,可以参照本公开其他各实施例的说明,此处不再赘述。

[0093] 步骤305,根据各个第一出现概率,确定每个候选词在对话日志中的第一高斯分布值。

[0094] 比如,候选词 ω 在对话日志中的第二出现概率为 ω_{total} ,对话日志中所有候选词的出现概率的均值为 μ_{total} 、方差为 σ_{total} ,则可以根据如下公式(1)所示的形式,确定该候选词 ω 在对话日志中的第一高斯分布值:

$$[0095] \quad f_{total}(\omega) = \frac{1}{\sqrt{2\pi}\sigma_{total}} \exp\left(-\frac{(\omega_{total}-\mu_{total})^2}{2\sigma_{total}^2}\right) \quad (1)$$

[0096] 步骤306,根据各个第二出现概率,确定每个物品对应的候选词在关联输入语句中

的第二高斯分布值。

[0097] 比如,候选词 ω 在关联输入语句中的出现概率为 ω_{skuid} ,当前物品所有候选词出现概率的均值为 μ_{skuid} 、方差为 σ_{skuid} ,则可以根据如下公式(2)所示的形式,确定该候选词 ω 在对关联输入语句中的第二高斯分布值:

$$[0098] \quad f_{skuid}(\omega) = \frac{1}{\sqrt{2\pi}\sigma_{skuid}} \exp\left(-\frac{(\omega_{skuid}-\mu_{skuid})^2}{2\sigma_{skuid}^2}\right) \quad (2)$$

[0099] 步骤307,根据第一出现概率、第二出现概率、第一高斯分布值及第二高斯分布值,确定每个物品对应的各个候选词中包含的描述词。

[0100] 其中,可以将第二高斯分布与第一高斯分布间的比值记为:

$$[0101] \quad \text{score}(\omega) = f_{skuid}(\omega) / f_{total}(\omega) \quad (3)$$

[0102] 可选的,可以先确定出第二出现概率大于第一出现概率的各个候选词,再确定出第二出现概率大于第一出现概率的各个候选词各自对应的 $\text{score}(\omega)$ 值,将 $\text{score}(\omega)$ 值按照大小顺序进行排列,将前预设数量的候选词,确定为该物品对应的描述词。

[0103] 可选的,也可以在任一物品对应的任一候选词对应的第二出现概率大于第一出现概率、且第二高斯分布值与第一高斯分布值间的比值大于第二阈值的情况下,确定任一候选词为任一物品的描述词。

[0104] 其中,第二阈值,可以为提前设定的,比如可以1.5、1.7等等,本公开对此不做限定。

[0105] 比如,第二阈值为1.3,任一物品A对应的任一候选词a对应的第二出现概率为0.8,第一出现概率为0.4,第二概率大于第一概率,且第二高斯分布值与第一高斯分布值间的比值大于第二阈值1.3,则可以确定该任一候选词a为任一物品A的描述词。

[0106] 可选的,也可以在任一物品对应的任一候选词对应的第二出现概率大于第一出现概率的情况下,将最大的 $\text{score}(\omega)$ 对应的候选词,确定为该物品对应的描述词。

[0107] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中,确定每个物品对应的各个候选词中包含的描述词的方式等的限定。

[0108] 本公开实施例,可以先确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品,之后可以根据关联输入语句与预设的描述词库中各个描述词的匹配度,确定关联输入语句中包含的候选词,再根据每个物品与关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系,之后可以确定每个候选词在对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率,并确定每个候选词在对话日志中的第一高斯分布值以及每个物品对应的候选词在关联输入语句中的第二高斯分布值,之后可以根据第一出现概率、第二出现概率、第一高斯分布值及第二高斯分布值,确定每个物品对应的各个候选词中包含的描述词。由此,通过对对话日志自动进行提取处理,即可生成对应的描述词,描述词的生成效率更高,所得描述词也更加准确和可靠。

[0109] 图4为本公开一实施例所提供的物品描述词的确定方法的流程示意图,如图4所示,该物品描述词的确定方法可以包括以下步骤:

[0110] 步骤401,接收查询语句。

[0111] 可以理解的是,确定装置在检测到用户触发了“发送”控件之后,即可确定接收查

询语句;或者在检测到用户输入了相关内容之后,即可确定接收查询语句等等,本公开对此不做限定。

[0112] 步骤402,基于描述词库对查询语句进行处理,以确定查询语句中包含的目标描述词。

[0113] 其中,描述词库中,可以包含有描述词,描述词可以为描述物品相关信息的任意词语,比如可以为:适用年龄、季节、手感、物品名称、型号等等,本公开对此不做限定。

[0114] 其中,可以先对查询语句进行切词处理,以得到该查询语句对应的分词,之后可以将该分词与描述词库中的描述词进行匹配,满足匹配条件的描述词即为查询语句中包含的目标描述词。

[0115] 比如,将匹配度最高的描述词作为查询语句中包含的目标描述词;或者,也可以将匹配度大于阈值的各描述词作为查询语句中包含的目标描述词等等,本公开对此不做限定。

[0116] 需要说明的是,本公开实施例中,可以采用任何可取的方式确定分词与描述词库中的描述词的匹配度,本公开对此不做限定。

[0117] 步骤403,根据预设的描述词与物品的关联关系,确定目标描述词关联的候选物品。

[0118] 比如,目标描述词为“老年手机、1000元”,之后可以在预设的描述词与物品的关联关系中,遍历查找与“老年手机”、“1000元”关联的物品。比如说,与“老年手机”、“1000元”关联的物品为:手机1、手机2,则可以确定该目标描述词关联的候选物品为:手机1、手机2。

[0119] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中目标描述词以及关联的候选物品等的限定。

[0120] 步骤404,根据目标描述词在每个候选物品描述词中的权重,确定目标物品。

[0121] 比如说,目标描述词“老年手机”在“手机1”中的权重为0.5,“1000元”在“手机1”中的权重为0.2,二者的权重之和为0.7;而目标描述词“老年手机”在“手机2”候选物品2中的权重为0.3,在“手机2”中的权重为0.1,二者的权重之和为0.4,则可以确定“手机1”为目标物品。

[0122] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中确定目标物品的方式等的限定。

[0123] 本公开实施例中,在确定目标物品时,充分考虑到了各个目标描述词在候选物品描述词中的权重,从而基于各权重,所确定的目标物品更加准确、可靠,也更加符合用户需求。

[0124] 步骤405,返回目标物品。

[0125] 可以理解的是,确定装置在确定出目标物品之后,可以将目标物品发送给用户,以使用户可以及时得到反馈,从而可以给予用户良好的体验。

[0126] 本公开实施例,可以先接收查询语句,之后基于描述词库对查询语句进行处理,以确定查询语句中包含的目标描述词,之后可以根据预设的描述词与物品的关联关系,确定目标描述词关联的候选物品,再根据目标描述词在每个候选物品描述词中的权重,确定目标物品,之后可以返回目标物品。由此,在推荐物品时,充分考虑到了目标描述词对应的权重,从而可使得推荐的物品准确性更高,更符合用户需求,可以给予用户良好的使用感。

- [0127] 为了实现上述实施例,本公开还提出一种物品描述词的确定装置。
- [0128] 图5为本公开实施例所提供的物品描述词的确定装置的结构示意图。
- [0129] 如图5所示,该物品描述词的确定装置100可以包括:第一确定模块110、第二确定模块120、第三确定模块130以及第四确定模块140。
- [0130] 其中,第一确定模块110,用于确定对话日志中包含的目标系统语句、关联输入语句、及所述目标系统语句对应的物品。
- [0131] 第二确定模块120,用于根据所述关联输入语句与预设的描述词库中各个描述词的匹配度,确定所述关联输入语句中包含的候选词;
- [0132] 第三确定模块130,用于根据每个物品与所述关联输入语句间的对应关系、及所述关联输入语句中包含的候选词,确定每个所述物品与候选词的对应关系;
- [0133] 第四确定模块140,用于根据每个所述物品对应的候选词的出现概率,确定每个所述物品对应的各个候选词中包含的描述词。
- [0134] 可选的,所述第一确定模块110,具体用于:
- [0135] 将所述对话日志中用于指向物品所在页面的系统语句,确定为目标系统语句;
- [0136] 根据所述目标系统语句指向的物品所在页面的描述信息,确定所述物品;
- [0137] 将接收时刻位于所述目标系统语句之前的预设数量的输入语句,确定为所述关联输入语句。
- [0138] 可选的,所述第二确定模块120,具体用于:
- [0139] 将所述关联输入语句进行切词,以确定所述关联输入语句中包含的各个分词;
- [0140] 在任一分词与各个所述描述词间的匹配度均小于第一阈值的情况下,确定所述任一分词为候选词。
- [0141] 可选的,所述第四确定模块140,具体用于:
- [0142] 确定每个候选词在所述对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率;
- [0143] 根据所述第一出现概率和/或所述第二出现概率,确定每个所述物品对应的各个候选词中包含的描述词。
- [0144] 可选的,所述第四确定模块140,包括:
- [0145] 第一确定单元,用于确定每个候选词在所述对话日志中的第一出现概率、及在每个物品关联的输入语句中的第二出现概率;
- [0146] 第二确定单元,用于根据各个所述第一出现概率,确定每个所述候选词在所述对话日志中的第一高斯分布值;
- [0147] 第三确定单元,用于根据各个所述第二出现概率,确定每个所述物品对应的候选词在关联输入语句中的第二高斯分布值;
- [0148] 第四确定单元,用于根据所述第一出现概率、第二出现概率、第一高斯分布值及第二高斯分布值,确定每个所述物品对应的各个候选词中包含的描述词。
- [0149] 可选的,所述第四确定单元,具体用于:
- [0150] 响应于任一物品对应的任一候选词对应的第二出现概率大于第一出现概率、且第二高斯分布值与第一高斯分布值间的比值大于第二阈值,确定所述任一候选词为所述任一物品的描述词。

- [0151] 可选的,所述第一确定模块110,还用于:
- [0152] 确定每个所述物品的描述词在所述关联输入语句中的第二出现概率;
- [0153] 根据所述第二出现概率,确定每个所述物品的描述词的权重值。
- [0154] 可选的,上述装置,还包括:
- [0155] 接收模块,用于接收查询语句。
- [0156] 处理模块,用于基于描述词库对所述查询语句进行处理,以确定所述查询语句中包含的目标描述词。
- [0157] 第五确定模块,用于根据预设的描述词与物品的关联关系,确定所述目标描述词关联的候选物品。
- [0158] 第六确定模块,用于根据所述目标描述词在每个所述候选物品描述词中的权重,确定目标物品。
- [0159] 返回模块,用于返回所述目标物品。
- [0160] 本公开实施例中的上述各模块的功能及具体实现原理,可参照上述各方法实施例,此处不再赘述。
- [0161] 本公开实施例的物品描述词的确定装置,可以先确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品,之后根据关联输入语句与预设的描述词库中各个描述词的匹配度,确定关联输入语句中包含的候选词,再根据每个物品与所述关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系,之后即可根据每个物品对应的候选词的出现概率,确定每个物品对应的各个候选词中包含的描述词。由此,通过对对话日志自动进行提取处理,即可生成对应的描述词,描述词的生成效率更高,所得描述词也更加准确和可靠。
- [0162] 为了实现上述实施例,本公开还提出一种计算机设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,处理器执行程序时,实现如本公开前述实施例提出的物品描述词的确定方法。
- [0163] 为了实现上述实施例,本公开还提出一种非临时性计算机可读存储介质,存储有计算机程序,计算机程序被处理器执行时实现如本公开前述实施例提出的物品描述词的确定方法。
- [0164] 为了实现上述实施例,本公开还提出一种计算机程序产品,当计算机程序产品中的指令处理器执行时,执行如本公开前述实施例提出的物品描述词的确定方法。
- [0165] 图6示出了适于用来实现本公开实施方式的示例性计算机设备的框图。图6显示的计算机设备12仅仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。
- [0166] 如图6所示,计算机设备12以通用计算设备的形式表现。计算机设备12的组件可以包括但不限于:一个或者多个处理器或者处理单元16,系统存储器28,连接不同系统组件(包括系统存储器28和处理单元16)的总线18。
- [0167] 总线18表示几类总线结构中的一种或多种,包括存储器总线或者存储器控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构(Industry Standard Architecture;以下简称:ISA)总线,微通道体系结构(Micro Channel Architecture;以下简称:MAC)总线,增强型ISA总线、视频电子标准协会(Video Electronics Standards

Association;以下简称:VESA) 局域总线以及外围组件互连 (Peripheral Component Interconnection;以下简称:PCI) 总线。

[0168] 计算机设备12典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算机设备12访问的可用介质,包括易失性和非易失性介质,可移动的和不可移动的介质。

[0169] 存储器28可以包括易失性存储器形式的计算机系统可读介质,例如随机存取存储器(Random Access Memory;以下简称:RAM) 30和/或高速缓存存储器32。计算机设备12可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统34可以用于读写不可移动的、非易失性磁介质(图6未显示,通常称为“硬盘驱动器”)。尽管图6中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如:光盘只读存储器(Compact Disc Read Only Memory;以下简称:CD-ROM)、数字多功能只读光盘(Digital Video Disc Read Only Memory;以下简称:DVD-ROM) 或其它光介质) 读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线18相连。存储器28可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本公开各实施例的功能。

[0170] 具有一组(至少一个)程序模块42的程序/实用工具40,可以存储在例如存储器28中,这样的程序模块42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。程序模块42通常执行本公开所描述的实施例中的功能和/或方法。

[0171] 计算机设备12也可以与一个或多个外部设备14(例如键盘、指向设备、显示器24等)通信,还可与一个或者多个使得用户能与该计算机设备12交互的设备通信,和/或与使得该计算机设备12能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口22进行。并且,计算机设备12还可以通过网络适配器20与一个或者多个网络(例如局域网(Local Area Network;以下简称:LAN),广域网(Wide Area Network;以下简称:WAN) 和/或公共网络,例如因特网)通信。如图所示,网络适配器20通过总线18与计算机设备12的其它模块通信。应当明白,尽管图中未示出,可以结合计算机设备12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、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0172] 处理单元16通过运行存储在系统存储器28中的程序,从而执行各种功能应用以及数据处理,例如实现前述实施例中提及的方法。

[0173] 本公开的技术方案,可以先确定对话日志中包含的目标系统语句、关联输入语句、及目标系统语句对应的物品,之后根据关联输入语句与预设的描述词库中各个描述词的匹配度,确定关联输入语句中包含的候选词,再根据每个物品与所述关联输入语句间的对应关系、及关联输入语句中包含的候选词,确定每个物品与候选词的对应关系,之后即可根据每个物品对应的候选词的出现概率,确定每个物品对应的各个候选词中包含的描述词。由此,通过对对话日志自动进行提取处理,即可生成对应的描述词,描述词的生成效率更高,所得描述词也更加准确和可靠。

[0174] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特

点包含于本公开的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不必针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0175] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本公开的描述中,“多个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。

[0176] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现定制逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分,并且本公开的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能,这应被本公开的实施例所属技术领域的技术人员所理解。

[0177] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤,例如,可以被认为用于实现逻辑功能的可执行指令的定序列表,可以具体实现在任何计算机可读介质中,以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用,或结合这些指令执行系统、装置或设备而使用。就本说明书而言,“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。计算机可读介质的更具体的示例(非穷尽性列表)包括以下:具有一个或多个布线的电连接部(电子装置),便携式计算机盘盒(磁装置),随机存取存储器(RAM),只读存储器(ROM),可擦除可编程只读存储器(EEPROM或闪存存储器),光纤装置,以及便携式光盘只读存储器(CDROM)。另外,计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质,因为可以例如通过对纸或其他介质进行光学扫描,接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序,然后将其存储在计算机存储器中。

[0178] 应当理解,本公开的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。如,如果用硬件来实现和在另一实施方式中一样,可用本领域公知的下列技术中的任一项或它们的组合来实现:具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路,具有合适的组合逻辑门电路的专用集成电路,可编程门阵列(PGA),现场可编程门阵列(FPGA)等。

[0179] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,该程序在执行时,包括方法实施例的步骤之一或其组合。

[0180] 此外,在本公开各个实施例中的各功能单元可以集成在一个处理模块中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机

可读取存储介质中。

[0181] 上述提到的存储介质可以是只读存储器,磁盘或光盘等。尽管上面已经示出和描述了本公开的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本公开的限制,本领域的普通技术人员在本公开的范围内可以对上述实施例进行变化、修改、替换和变型。

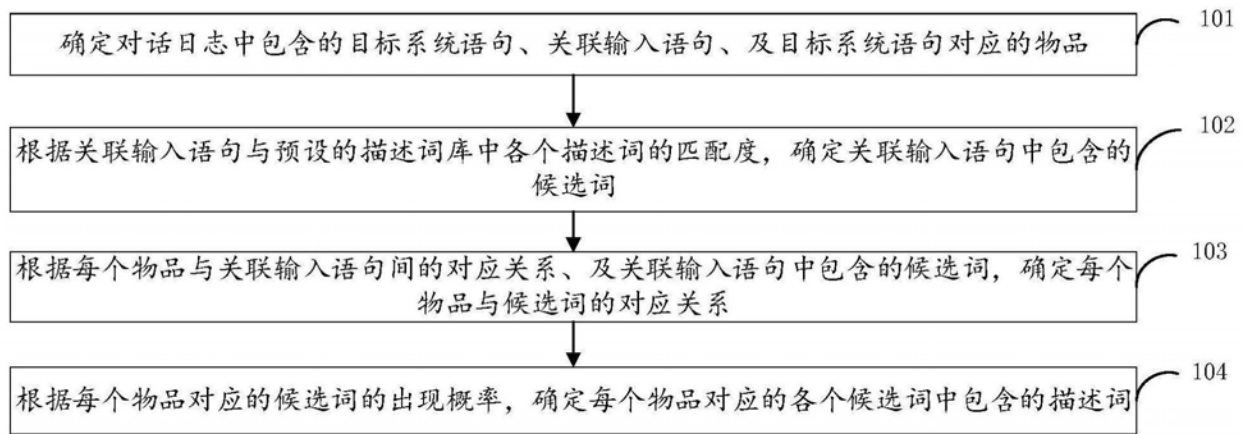


图1

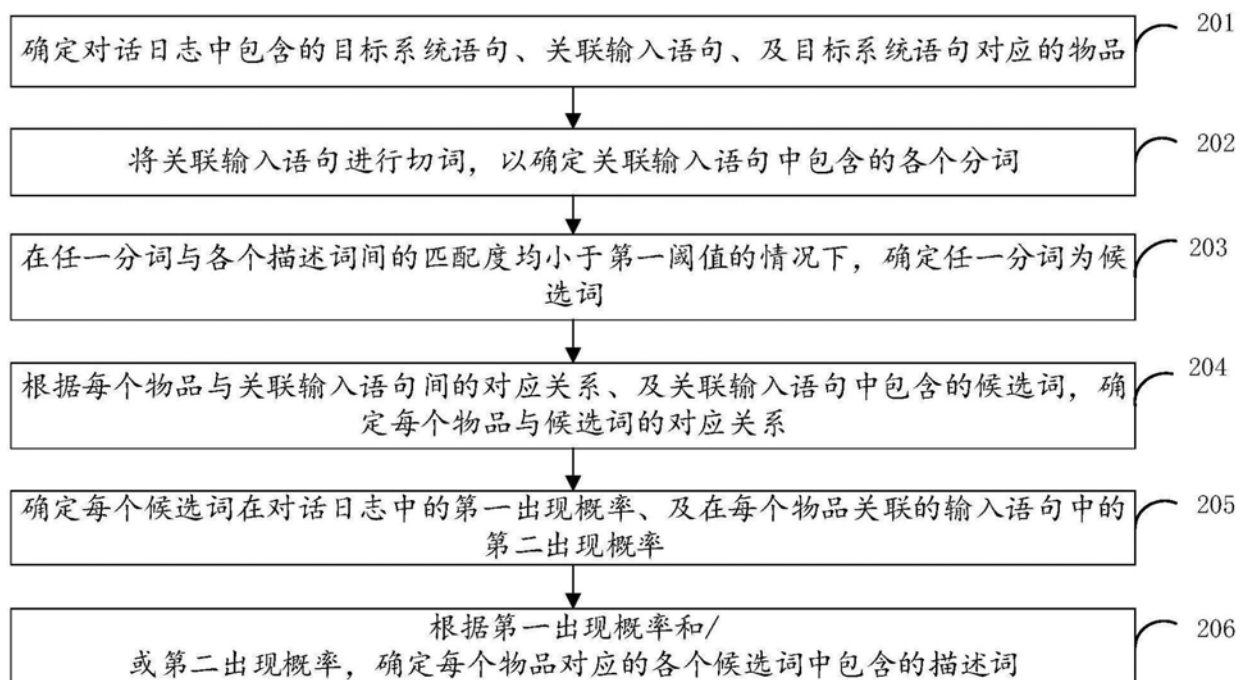


图2

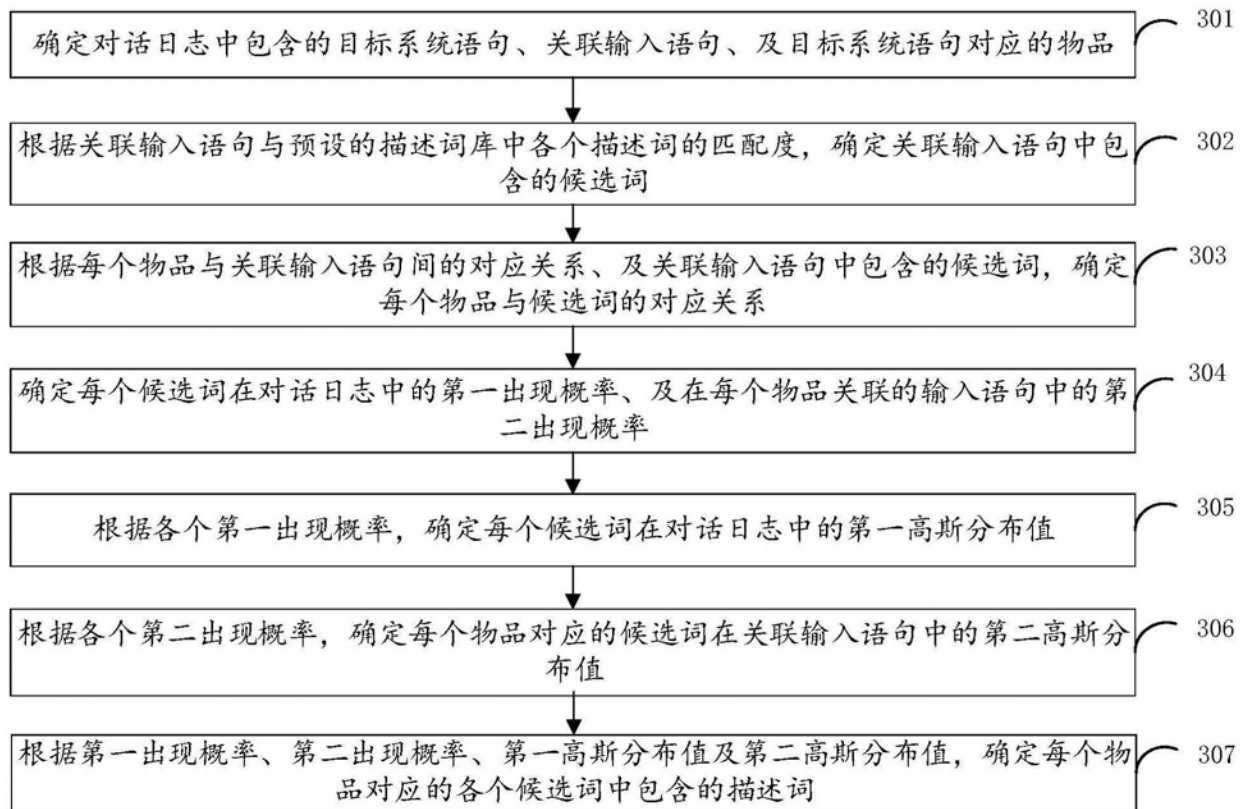


图3

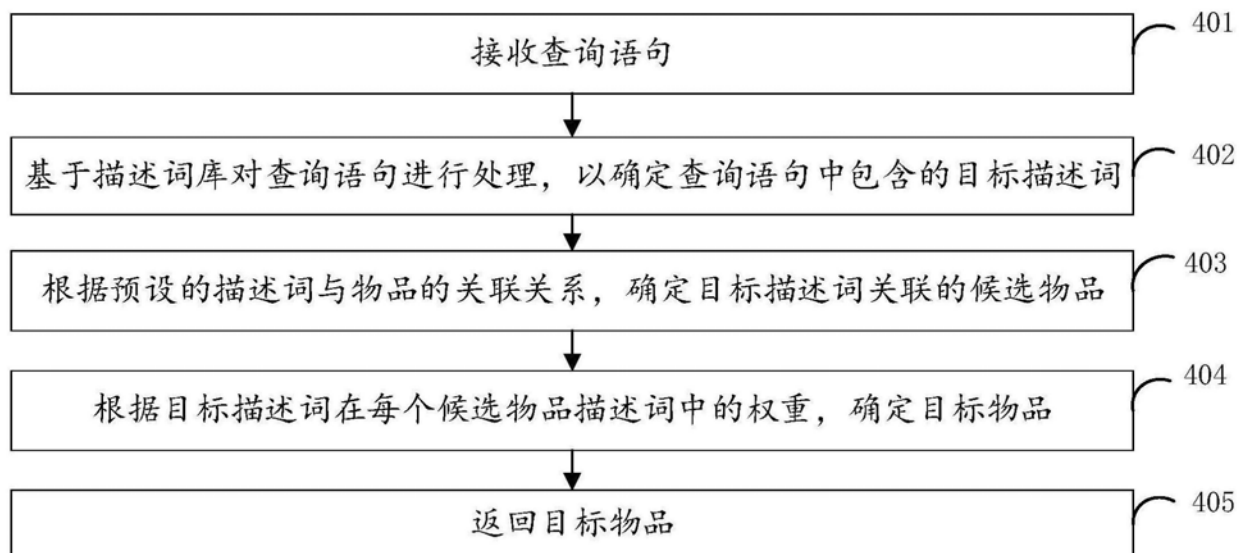


图4

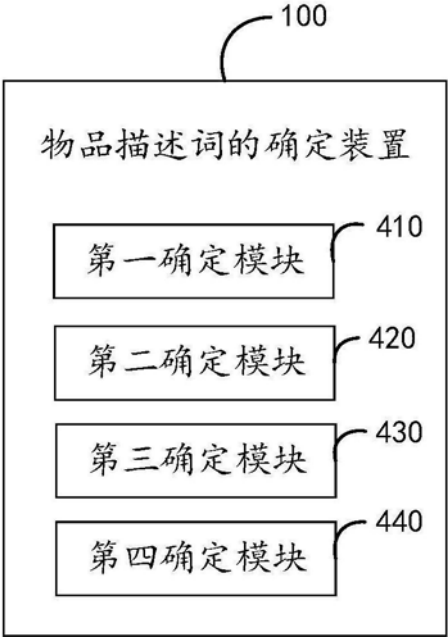


图5

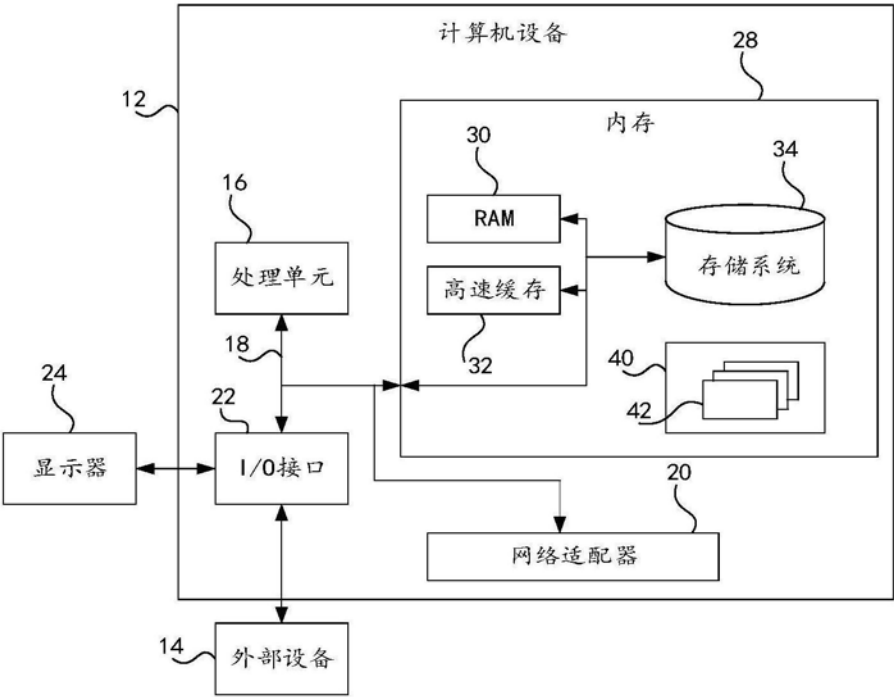


图6