



(12) 发明专利

(10) 授权公告号 CN 112818230 B

(45) 授权公告日 2023. 10. 20

(21) 申请号 202110129782.9

G06F 18/2431 (2023.01)

(22) 申请日 2021.01.29

G06F 18/22 (2023.01)

(65) 同一申请的已公布的文献号

G06N 3/0464 (2023.01)

申请公布号 CN 112818230 A

G06N 20/00 (2019.01)

(43) 申请公布日 2021.05.18

(56) 对比文件

(73) 专利权人 北京百度网讯科技有限公司

CN 108121770 A, 2018.06.05

地址 100085 北京市海淀区上地十街10号

CN 111428138 A, 2020.07.17

百度大厦2层

US 10529000 B1, 2020.01.07

(72) 发明人 刘欢 张鑫睿 陈坤斌 刘准

US 2019042585 A1, 2019.02.07

何伯磊 和为

CN 102880687 A, 2013.01.16

CN 102754094 A, 2012.10.24

(74) 专利代理机构 北京猷德知识产权代理有限公司

US 10783167 B1, 2020.09.22

公司 16084

审查员 耿翠萍

专利代理师 范继晨

(51) Int. Cl.

G06F 16/9535 (2019.01)

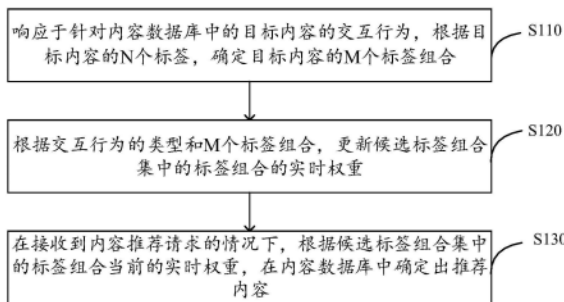
权利要求书3页 说明书12页 附图3页

(54) 发明名称

内容推荐方法、装置、电子设备和存储介质

(57) 摘要

本公开提供了内容推荐方法、装置、电子设备和存储介质,涉及自然语言处理领域,尤其涉及智能推荐领域。具体实施方案为:响应于针对内容数据库中的目标内容的交互行为,根据目标内容的N个标签,确定目标内容的M个标签组合;根据交互行为的类型和M个标签组合,更新候选标签组合集中的标签组合的实时权重;在接收到内容推荐请求的情况下,根据候选标签组合集中的标签组合当前的实时权重,在内容数据库中确定出推荐内容。根据本公开实施例,能够更准确、更贴近用户偏好地确定推荐内容。



1. 一种内容推荐方法,包括:

响应于针对内容数据库中的目标内容的交互行为,根据所述目标内容的N个标签,确定所述目标内容的M个标签组合;其中,N为大于等于2的整数,M为正整数;

根据所述交互行为的类型和所述M个标签组合,更新候选标签组合集中的标签组合的实时权重;

在接收到内容推荐请求的情况下,根据所述候选标签组合集中的标签组合当前的实时权重,在所述内容数据库中确定出推荐内容;

其中,所述根据所述交互行为的类型和所述M个标签组合,更新候选标签组合集中的标签组合的实时权重,包括:

在所述候选标签组合集不包含所述M个标签组合中的第j个标签组合的情况下,若所述候选标签组合集中的标签组合的数量小于第一预设阈值,则在所述候选标签组合集中添加所述第j个标签组合,并根据所述交互行为的类型,确定所述第j个标签组合的实时权重;其中,j为小于等于M的正整数。

2. 根据权利要求1所述的方法,其中,所述根据所述交互行为的类型和所述M个标签组合,更新候选标签组合集中的标签组合的实时权重,包括:

在所述候选标签组合集包含所述M个标签组合中的第i个标签组合的情况下,根据所述交互行为的类型,更新所述第i个标签组合的实时权重;其中,i为小于等于M的正整数。

3. 根据权利要求1-2中任一项所述的方法,还包括:

对在预定时段内触发的针对所述内容数据库中的各个内容的交互行为进行统计,得到K个标签组合的统计权重;其中,K为正整数;

根据所述K个标签组合的统计权重,在所述K个标签组合中选取L个标签组合;其中,L为小于等于K的正整数;

基于所述L个标签组合得到所述候选标签组合集。

4. 根据权利要求1-2中任一项所述的方法,所述根据所述候选标签组合集中的标签组合的实时权重,在所述内容数据库中确定出推荐内容,包括:

根据所述候选标签组合集中的各个标签组合的实时权重,从所述候选标签组合集中确定出感兴趣标签组合;

将所述内容数据库中与所述感兴趣标签组合对应的内容,确定为所述推荐内容。

5. 根据权利要求4所述的方法,其中,所述将所述内容数据库中与所述感兴趣标签组合对应的内容,确定为所述推荐内容,包括:

在所述感兴趣标签组合为高频标签组合的情况下,根据预先确定的高频标签组合与内容队列之间的映射关系,确定与所述感兴趣标签组合对应的内容队列;

在所述内容数据库中选取所述内容队列中的内容,并将选取的所述内容确定为所述推荐内容。

6. 根据权利要求5所述的方法,还包括:

确定与至少一个用户标识分别对应的至少一个候选标签组合集;

根据预先配置的至少一个标签组合中的每个标签组合在所述至少一个候选标签组合集中的出现次数,从所述至少一个标签组合中确定出所述至少一个高频标签组合。

7. 根据权利要求5所述的方法,还包括:

针对所述至少一个高频标签组合中的每个高频标签组合,在与所述高频标签组合对应的内容队列所包含的内容的数量小于第二预设阈值的情况下,在所述内容数据库中增加与所述高频标签组合对应的内容。

8. 一种内容推荐装置,包括:

交互响应模块,用于响应于针对内容数据库中的目标内容的交互行为,根据所述目标内容的N个标签,确定所述目标内容的M个标签组合;其中,N为大于等于2的整数,M为正整数;

权重更新模块,用于根据所述交互行为的类型和所述M个标签组合,更新候选标签组合集中的标签组合的实时权重;

推荐模块,用于在接收到内容推荐请求的情况下,根据所述候选标签组合集中的标签组合当前的实时权重,在所述内容数据库中确定出推荐内容;

其中,所述权重更新模块,包括:

第二更新单元,用于在所述候选标签组合集不包含所述M个标签组合中的第j个标签组合的情况下,若所述候选标签组合集中的标签组合的数量小于第一预设阈值,则在所述候选标签组合集中添加所述第j个标签组合,并根据所述交互行为的类型,确定所述第j个标签组合的实时权重;其中,j为小于等于M的正整数。

9. 根据权利要求8所述的装置,其中,所述权重更新模块包括:

第一更新单元,用于在所述候选标签组合集包含所述M个标签组合中的第i个标签组合的情况下,根据所述交互行为的类型,更新所述第i个标签组合的实时权重;其中,i为小于等于M的正整数。

10. 根据权利要求8-9中任一项所述的装置,还包括:

权重确定模块,用于对在预定时段内触发的针对所述内容数据库中的各个内容的交互行为进行统计,得到K个标签组合的统计权重;其中,K为正整数;

组合选取模块,用于根据所述K个标签组合的统计权重,在所述K个标签组合中选取L个标签组合;其中,L为小于等于K的正整数;

第一集合确定模块,用于基于所述L个标签组合得到所述候选标签组合集。

11. 根据权利要求8-10中任一项所述的装置,所述推荐模块包括:

组合筛选单元,用于根据所述候选标签组合集中的各个标签组合的实时权重,从所述候选标签组合集中确定出感兴趣标签组合;

内容确定单元,用于将所述内容数据库中与所述感兴趣标签组合对应的内容,确定为所述推荐内容。

12. 根据权利要求11所述的装置,其中,所述内容确定单元用于:

在所述感兴趣标签组合为高频标签组合的情况下,根据预先确定的高频标签组合与内容队列之间的映射关系,确定与所述感兴趣标签组合对应的内容队列;

在所述内容数据库中选取所述内容队列中的内容,并将选取的所述内容确定为所述推荐内容。

13. 根据权利要求12所述的装置,还包括:

第二集合确定模块,用于确定与至少一个用户标识分别对应的至少一个候选标签组合集;

高频确定模块,用于根据预先配置的至少一个标签组合中的每个标签组合在所述至少一个候选标签组合集中的出现次数,从所述至少一个标签组合中确定出所述至少一个高频标签组合。

14. 根据权利要求13所述的装置,还包括:

内容增加模块,用于针对所述至少一个高频标签组合中的每个高频标签组合,在与所述高频标签组合对应的内容队列所包含的内容的数量小于第二预设阈值的情况下,在所述内容数据库中增加与所述高频标签组合对应的内容。

15. 一种电子设备,其特征在于,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-7中任一项所述的方法。

16. 一种存储有计算机指令的非瞬时计算机可读存储介质,其特征在于,所述计算机指令用于使计算机执行权利要求1-7中任一项所述的方法。

内容推荐方法、装置、电子设备和存储介质

技术领域

[0001] 本公开涉及自然语言处理领域,尤其涉及智能推荐领域。

背景技术

[0002] 随着互联网发展以及内容的迅速膨胀,越来越多的场景中需要为用户推荐相关内容。例如,随着企业的发展,企业内部沉淀的文档知识越来越多,为了让知识在企业内部高效流动,需要构建知识推荐系统,实现知识主动找人。如何更准确地为用户推荐相关内容,是智能推荐领域的热点问题。

发明内容

[0003] 本公开提供了一种内容推荐方法、装置、电子设备和存储介质。

[0004] 根据本公开的一方面,提供了一种内容推荐方法,包括:

[0005] 响应于针对内容数据库中的目标内容的交互行为,根据目标内容的N个标签,确定目标内容的M个标签组合;其中,N为大于等于2的整数,M为正整数;

[0006] 根据交互行为的类型和M个标签组合,更新候选标签组合集中的标签组合的实时权重;

[0007] 在接收到内容推荐请求的情况下,根据候选标签组合集中的标签组合当前的实时权重,在内容数据库中确定出推荐内容。

[0008] 根据本公开的另一方面,提供了一种内容推荐装置,包括:

[0009] 交互响应模块,用于响应于针对内容数据库中的目标内容的交互行为,根据目标内容的N个标签,确定目标内容的M个标签组合;其中,N为大于等于2的整数,M为正整数;

[0010] 权重更新模块,用于根据交互行为的类型和M个标签组合,更新候选标签组合集中的标签组合的实时权重;

[0011] 推荐模块,用于在接收到内容推荐请求的情况下,根据候选标签组合集中的标签组合当前的实时权重,在内容数据库中确定出推荐内容。

[0012] 根据本公开的另一方面,提供了一种电子设备,包括:

[0013] 至少一个处理器;以及

[0014] 与该至少一个处理器通信连接的存储器;其中,

[0015] 该存储器存储有可被该至少一个处理器执行的指令,该指令被该至少一个处理器执行,以使该至少一个处理器能够执行本公开任一实施例中的方法。

[0016] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,该计算机指令用于使计算机执行本公开任一实施例中的方法。

[0017] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现本公开任一实施例中的方法。

[0018] 根据本公开的技术,当用户对内容数据库中的目标内容做出交互行为时,先确定出目标内容的标签组合,再根据交互行为的类型和目标内容的标签组合更新候选标签组合

集中的标签组合的实时权重。由于用于确定推荐内容的该权重是以标签组合为单位更新的，标签组合能够更准确表征用户偏好，并且，该权重是基于交互行为的类型实时更新的，因此，基于该权重能够更准确、更贴近用户偏好地确定推荐内容。

[0019] 应当理解，本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征，也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0020] 附图用于更好地理解本方案，不构成对本公开的限定。其中：

[0021] 图1是本公开一个实施例提供的内容推荐方法的示意图；

[0022] 图2是本公开另一个实施例提供的内容推荐方法的示意图；

[0023] 图3是本公开实施例中内容推荐系统的一个应用示例的示意图；

[0024] 图4是本公开一个实施例提供的内容推荐装置的示意图；

[0025] 图5是本公开另一个实施例提供的内容推荐装置的示意图；

[0026] 图6是用来实现本公开实施例的内容推荐方法的电子设备的框图。

具体实施方式

[0027] 以下结合附图对本公开的示范性实施例做出说明，其中包括本公开实施例的各种细节以助于理解，应当将它们认为仅仅是示范性的。因此，本领域普通技术人员应当认识到，可以对这里描述的实施例做出各种改变和修改，而不会背离本公开的范围和精神。同样，为了清楚和简明，以下的描述中省略了对公知功能和结构的描述。

[0028] 在智能推荐领域，为了直观体现推荐内容与用户需求之间的匹配度，推荐一般应具有可解释性。为了实现推荐的可解释性，一般基于标签进行内容推荐。

[0029] 一种示例性的方式是，在一个标签体系中为用户和内容分别确定单个标签，对用户的单个标签以及内容的单个标签进行匹配，根据匹配结果进行推荐。如此，用户和推荐内容之间相关性较高。但随着内容的膨胀以及知识边界的扩展，基于多维度的知识类型、主题等信息维护标签体系的成本较高，为用户和内容确定准确的单个标签也存在较高难度。

[0030] 另一种示例性的实施方式是，维护与多个维度分别对应的多个扁平简单的标签体系，为用户和内容分别确定多个维度上的标签。基于用户的每个标签分别确定相匹配的内容。例如，用户的标签包括A、B和C，则为用户推荐标签为A的内容、标签为B的内容以及标签为C的内容。如此，虽然易于维护、扩展标签体系以及为用户和内容确定标签，但导致了推荐内容与用户之间的相关性的漂移。

[0031] 本公开实施例提供的方案，主要用于解决上述问题中的至少一个。

[0032] 图1示出了本公开一个实施例提供的内容推荐方法的示意图。如图1所示，该方法包括：

[0033] 步骤S110，响应于针对内容数据库中的目标内容的交互行为，根据目标内容的N个标签，确定目标内容的M个标签组合；其中，N为大于等于2的整数，M为正整数；

[0034] 步骤S120，根据交互行为的类型和M个标签组合，更新候选标签组合集中的标签组合的实时权重；

[0035] 步骤S130，在接收到内容推荐请求的情况下，根据候选标签组合集中的标签组合

当前的实时权重,在内容数据库中确定出推荐内容。

[0036] 示例性地,内容推荐方法可以由电子设备执行,其中,电子设备例如是内容推荐系统的服务器或用户设备等。内容推荐系统可以集成在信息分享App(Application,应用程序)、社交App、企业办公App等之中。

[0037] 本公开实施例中,内容可以指信息的各种载体,例如知识文档、文章、广告等。内容数据库可以是至少一个内容的集合,例如内容推荐系统中的各内容的集合。目标内容可以是内容数据库中的任意一个内容。目标内容也可以是内容数据库中的某个特定的内容,例如在特定时段内发布的内容或某种类型的内容等。

[0038] 示例性地,电子设备可以基于多个维度的标签体系,确定内容数据库中的各内容的标签。确定内容的标签的过程也可以称作为内容打标签。

[0039] 在一些实施方式中,标签体系可以是预先配置的标签集合。相应的,针对内容数据库中的每个内容,电子设备可以在多个维度对应的多个标签集合中,分别选取标签,作为该内容的标签。

[0040] 可选地,电子设备可以通过在内容数据库的各内容中提取关键词、识别实体或接收人工归纳的标签集合等方式,构建标签体系。

[0041] 可选地,电子设备可以利用分类器、相似度预测模型等深度学习模型为内容打标签。例如,将内容输入深度学习模型,由深度学习模型输出相应的标签。电子设备也可以基于预设规则为内容打标签。

[0042] 以下以企业办公App的内容推荐系统为例,提供多种构建标签体系和为内容打标签的示例性方式:

[0043] 示例一:

[0044] 本示例中,多个维度包括关键实体维度。其中,关键实体可以指内容涉及的实体。例如,关键实体可以包括内容涉及的部门、项目、技术专有名词、工具软件等。

[0045] 构建关键实体维度的标签体系的过程可以包括:利用无监督关键词提取算法和NER(Named Entity Recognition,命名实体识别)模型,从内容数据库中挖掘企业内常见的关键词,作为标签体系中的标签。和/或,获取企业办公App中已有的结构化信息,例如ERP(Enterprise Resource Planning,企业资源计划)中的部门、项目、工具软件等,作为标签体系中的标签。利用上述方式构建的关键实体维度的标签体系中包含的标签的数量较多,例如是1.8万、3万等。

[0046] 为内容打上关键实体维度的标签的过程可以包括:对标签体系中所有标签进行词语切分,得到每个标签对应的切词,并建立标签和切词之间的倒排索引,即索引为切词,查找对象为标签的索引体系。对内容的标题和正文进行词语切分,得到内容的切词。基于内容的切词以及上述倒排索引,在标签体系中查找至少一个标签,作为候选标签。再利用相似度预测模型例如基于SimNet网络结构的预测模型,确定内容与各候选标签之间的相似度。将相似度超过预设阈值的标签确定为内容对应的关键实体维度上的标签,或者将相似度最大的标签确定为内容对应的关键实体维度上的标签。

[0047] 示例二:

[0048] 本示例中,多个维度包括主题维度。例如,主题可以包括机器学习、深度学习、软件测试、企业文化等。

[0049] 构建主题维度的标签体系的过程可以包括：人工根据企业内部知识分布对涉及的知识进行梳理、归纳，得到主题维度上的多个标签。电子设备接收输入的多个标签，基于输入的多个标签得到主题维度的标签体系。主题维度的标签体系中包含的标签的数量相对较少，例如是50、83等。

[0050] 为内容打上主题维度的标签的过程可以包括：利用多分类模型，基于主题维度的标签体系以及输入的内容，得到内容对应的主题维度上的标签。示例性地，多分类模型可以是利用样本数据对，对预训练模型例如Ernie2.0模型进行微调(fine-tune)得到的。其中，样本数据对可以是基于预设规则得到的内容和对标签的数据对。其中，预设规则例如包括：包含关键词“模型”、“神经网络”或“卷积”等关键词的内容的标签为深度学习，存储于企业资料库的内容的标签为工作资料等。

[0051] 示例三：

[0052] 本示例中，多个维度包括倾向性维度。不同的内容包括文章、知识文档等有不同形态、深度，适于不同的人群阅读。例如，热门新闻适合所有人群阅读，而软件测试基础知识则适合测试方向的初级学者或对测试方向感兴趣的人们阅读。倾向性维度的标签可以指与内容所适宜的受众类型相关的标签，例如热门新闻、领域动态、公司动态、公司制度与福利、技术入门知识、技术深度知识、产品入门知识、产品深度知识、通用能力等。

[0053] 构建倾向性维度的标签体系的过程可以包括：人工根据企业内部知识分布对涉及的知识进行梳理、归纳，得到倾向性维度上的多个标签。电子设备接收输入的多个标签，基于输入的多个标签得到倾向性维度的标签体系。倾向性维度的标签体系中包含的标签的数量相对较少，例如是7、9等。

[0054] 为内容打上倾向性维度的标签的过程可以包括：利用多分类模型，基于倾向性维度的标签体系以及输入的内容，得到内容对应的倾向性维度上的标签。示例性地，多分类模型可以是利用样本数据对，对预训练模型例如Ernie2.0模型进行微调(fine-tune)得到的。其中，样本数据对可以是基于人工标注得到。

[0055] 从上述示例性方式可以看到，内容数据库中的内容，可以具有多个标签。一种示例中，根据上述步骤S110，在检测到针对内容数据库中的目标内容的交互行为时，可以对目标内容的N个标签进行排列组合，得到目标内容的M个标签组合。这里，每个标签组合可以包括至少两个标签，以使标签组合与内容的相关性高于单个标签与内容的相关性。例如，某个文章的标签包括A1、B6和C2，则可以得到以下4个标签组合：包括A1和B6的第一标签组合、包括A1和C2的第二标签组合、包括B6和C2的第三标签组合、以及包括A1、B6和C2的第三标签组合。

[0056] 示例性地，交互行为可以包括用户对内容的浏览、收藏、取消收藏、反馈是否喜欢等，也可以包括向用户展现、推荐了某个内容后用户没有浏览等。根据上述步骤S120，以触发针对目标内容的交互行为的用户为第一用户为例，可以基于交互行为的类型以及目标内容的M个标签组合，针对第一用户，更新候选标签组合集中各候选标签组合的实时权重。其中，候选标签组合集可以包括多个标签组合，这些标签组合可以称为候选标签组合。

[0057] 在一些示例性的实施方式中，各用户的候选标签组合集是相同的。示例性地，可以对各个维度的标签体系中的标签进行排列组合，得到多个候选标签组合。例如，构建关键实体维度、主题维度和倾向性维度的标签体系，各标签体系中包含的标签的数量分别为

18000、83和9。通过排列组合得到 $18000 \times 83 \times 9 = 13446000$ 个候选标签组合,基于这些候选标签组合得到各用户的候选标签组合集。在这些实施方式中,针对每个用户,候选标签组合集中的各标签组合的实时权重可以不同。例如,对于候选标签组合集中的包括标签A1和B6的第一标签组合,用户1的实时权重为50,用户2的实时权重为60。在交互行为触发时,电子设备根据当前使用内容推荐系统的用户的ID (Identifier,标识) 确定当前待更新的实时权重,并对该实时权重进行更新。这里,实时权重用于表征标签组合和用户的相关性的大小,其数值可以在更新过程中累加,没有上限。

[0058] 在一些示例性的实施方式中,各用户的候选标签组合集是不同的。例如,针对不同用户的相关信息,例如职业、阅读习惯等,预先为用户配置专有的候选标签组合集。

[0059] 在这些实施方式中,各用户可以有自己的候选标签组合集。各候选标签组合集中的各标签组合分别对应一个实时权重。例如,用户1的候选标签组合集中存在包括标签A1和B6的第一标签组合,但不存在包括B6和C2的第三标签组合。用户2的候选标签组合集中则同时存在第一标签组合和第三标签组合。用户1的候选标签组合集中的第三标签组合的实时权重为50,用户2的候选标签组合集中的第三标签组合的实时权重为60。电子设备可以记录每个候选标签组合集所对应的用户ID。在交互行为触发时,电子设备根据当前使用内容推荐系统的用户的ID从多个用户ID对应的多个候选标签组合集中确定待更新的候选标签组合集,对该候选标签组合集中的候选标签组合的实时权重进行更新。

[0060] 实际应用中,每次更新实时权重时,可以根据交互行为的类型,对实时权重进行不同幅度的增加或减小。例如,浏览对应的增加幅度为1,收藏对应的增加幅度为1.5;取消收藏对应的增加幅度为-1即减小幅度为1;反馈不喜欢的增加幅度为-4;推荐了目标内容但用户没有浏览目标内容对应的增加幅度为 $X = (-1/120 + \text{pos}/(120 \times 15))$ 。其中,每次为用户推荐内容时,分为120页显示内容的标识,每页显示15个内容的标识(用户可通过点击内容的标识浏览内容),pos表示目标内容在每页中的位置序号,取值为1至15中的一个。根据上述公式,目标内容在页面中的显示位置越后,则实时权重需要减小的幅度越小。例如,当序号为15,即目标内容在页面中的最后一个显示位置时,即使用户不浏览目标内容,也不会对目标内容的标签组合所对应的实时权重产生影响。

[0061] 本公开实施例中,候选标签组合集中的标签组合的实时权重用于在内容数据库中确定出推荐内容。例如,在上述步骤S130中,可以根据候选标签组合集中所有标签组合的实时权重,确定出实时权重最大的Y个标签组合,从内容数据库中确定出与Y个标签组合中的任一个对应的内容,作为推荐内容,为用户展现推荐内容。

[0062] 需要说明的是,每次更新实时权重后,不一定会确定推荐内容。确定推荐内容的步骤可以在接收到内容推荐请求的情况下执行。其中,内容推荐请求可以基于用户操作生成,例如用户在App中切换到推荐页面时,生成内容推荐请求。内容推荐请求也可以自动生成,例如,每隔一段时间,即基于预定频率,生成内容推荐请求,以在App中显示弹窗,弹窗中显示推荐内容。

[0063] 可见,根据本公开实施例提供的内容推荐方法,当用户对内容数据库中的目标内容做出交互行为时,先确定出目标内容的标签组合,根据交互行为的类型和目标内容的标签组合更新候选标签组合集中的标签组合的实时权重。由于用于确定推荐内容的该权重是以标签组合为单位更新的,标签组合能够更准确表征用户偏好,并且,该权重是基于交互行

为的类型实时更新的,因此,基于该权重能够更准确、更贴近用户偏好地确定推荐内容,提升了推荐内容的相关性。

[0064] 示例性地,如果候选标签组合集一定包含M个标签组合中的所有标签组合,则可以对M个标签组合中的所有标签组合的实时权重进行更新。例如,在候选标签组合集是对多个标签体系中的各标签进行排列组合得到的情况下,可以直接对候选标签组合集中的上述M个标签组合的实时权重进行更新。

[0065] 如果候选标签组合集不一定包含M个标签组合中的所有标签组合,则可以根据如下示例性实施方式实现上述步骤S120:

[0066] 根据交互行为的类型和M个标签组合,更新候选标签组合集中的标签组合的实时权重,可以包括:

[0067] 在候选标签组合集包含M个标签组合中的第i个标签组合的情况下,根据交互行为的类型,更新第i个标签组合的实时权重其中,i为小于等于M的正整数。

[0068] 可选地,根据交互行为的类型和M个标签组合,更新候选标签组合集中的标签组合的实时权重,可以包括:在候选标签组合集不包含上述第i个标签组合的情况下,可以不对第i个标签组合进行处理。

[0069] 例如,目标内容的M个标签组合包含由标签A1和B6组成的第一标签组合,如果待更新的候选标签组合集中也包含第一标签组合,则更新第一表情组合的实时权重。如果待更新的候选标签组合集中不包含第一标签组合,无需针对第i个标签组合更新实时权重。

[0070] 如此,可以通过控制候选标签组合集中标签组合的数量,限制实时在线更新时电子设备的运算量,减少电子设备的运算压力,提高更新效率。

[0071] 可选地,根据交互行为的类型和M个标签组合,更新候选标签组合集中的标签组合的实时权重,可以包括:

[0072] 在候选标签组合集不包含M个标签组合中的第j个标签组合的情况下,若候选标签组合集中的标签组合的数量小于第一预设阈值,则在候选标签组合集中添加第j个标签组合,并根据交互行为的类型,确定第j个标签组合的实时权重;其中,j为小于等于M的正整数。

[0073] 例如,目标内容的M个标签组合包含由标签A1和B6组成的第一标签组合,如果待更新的候选标签组合集中也包含第一标签组合,则更新第一表情组合的实时权重。如果待更新的候选标签组合集中不包含第一标签组合,但此时待更新的候选标签组合集中标签组合的数量小于第一预设阈值,则可以在候选标签组合集中增加第一标签组合,并确定第一表情组合的实时权重,作为初始的实时权重。

[0074] 如此,可以将候选标签组合集中标签组合的数量限制在第一预设阈值内。在该数量小于第一预设阈值的情况下,可以更新实时权重。因此,在限制实时在线更新时电子设备的运算量的同时,可以尽可能少地忽略与用户相关的标签组合,提高推荐的准确性。

[0075] 示例性地,如图2所示,内容推荐方法还可以包括:

[0076] 步骤S210,对在预定时段内触发的针对内容数据库中的各个内容的交互行为进行统计,得到K个标签组合的统计权重;其中,K为正整数;

[0077] 步骤S220,根据K个标签组合的统计权重,在K个标签组合中选取L个标签组合;其中,L为小于等于K的正整数;

[0078] 步骤S230,基于L个标签组合得到候选标签组合集。

[0079] 可选地,上述步骤S210至S230,可以在上述步骤S110至步骤S130之前执行,也可以在上述步骤S110至步骤S130之后执行。

[0080] 具体而言,在一些示例性的实施方式中,可以预先确定出候选标签组合集,再基于交互行为更新候选标签组合集中各标签组合的实时权重,以及针对内容推荐请求确定推荐内容。在一些示例性的实施方式中,也可以在实时更新实时权重以及确定推荐内容的过程中,定期重新确定候选标签组合集。例如每天0点进行离线计算,对之前一个月内用户的所有交互行为进行统计,得到各标签组合的统计权重,重新确定一候选标签组合集。

[0081] 示例性地,在得到新的候选标签组合集时,候选标签组合集中各标签组合的实时权重,即初始的实时权重,可以是预设值,例如0或1等。或者,对于候选标签组合集中的每个标签组合,如果执行步骤S210前该标签组合也在候选标签组合集中,则该标签组合的实时权重可以沿用重新确定候选标签组合集前的实时权重;如果执行步骤S210前该标签组合不在候选标签组合集中,则该标签组合的实时权重的初始值可以是预设值。

[0082] 实际应用中,可以针对某个用户,例如第一用户,对第一用户在预定时段内触发的针对各内容的交互行为进行统计,以得到第一用户的候选标签组合集。在上述步骤S210中,可以针对K个标签组合中的每个标签组合,在预定时段内涉及交互行为的各个内容中,确定该标签组合所对应的内容。再根据该内容对应的在预定时段内的交互行为的类型和次数,确定该标签组合的统计权重。或者,遍历预定时段内的每次交互行为,基于交互行为的类型,对K个标签组合中的与交互行为相关的标签组合的统计权重进行累加,得到最终的统计权重。

[0083] 可选地,K个标签组合可以是预先配置的。例如,K个标签组合是对多个标签体系中的各标签进行排列组合得到的。

[0084] 可选地,L可以是预先配置的,例如,预先确定在上述步骤S220中要选取的标签组合的数量,根据该数量选取标签组合,例如选取统计权重最大的L个标签组合。实际应用中,如果根据候选标签组合集中所有标签组合的实时权重,确定出实时权重最大的Y个标签组合,基于Y个标签组合确定推荐内容,则L可以跟Y相关。例如 $L=2Y$ 。

[0085] 根据图2所示的实施例,可以基于预定时段内的交互行为准确地确定候选标签组合集。例如在用户兴趣转移时,可以及时捕捉到与用户的新兴趣相关性高的候选标签组合集。在通过控制候选标签组合集中标签组合的数量,限制实时在线更新时电子设备的运算量的同时,可以确保被忽略更新的标签组合与用户的相关性低,提高推荐内容的准确性。

[0086] 在一种示例性的实施方式中,上述步骤S130中,根据候选标签组合集中的标签组合的实时权重,在内容数据库中确定出推荐内容,包括:

[0087] 根据候选标签组合集中的各个标签组合的实时权重,从候选标签组合集中确定出感兴趣标签组合;

[0088] 将内容数据库中与感兴趣标签组合对应的内容,确定为推荐内容。

[0089] 示例性地,可以将实时权重最大的Y个标签组合确定为感兴趣标签组合。进一步地,Y可以等于 $0.5L$ 。

[0090] 示例性地,可以针对某个用户,例如第一用户,根据第一用户的候选标签组合集中的各个标签组合的实时权重,从候选标签组合集中确定出感兴趣标签组合;并将与感兴趣

标签组合对应的内容,确定为第一用户的推荐内容。

[0091] 根据上述实施方式,可以准确确定用户的感兴趣标签组合。基于感兴趣标签组合确定对应的推荐内容。由于基于感兴趣标签组合进行内容推荐,标签组合能够更准确表征用户偏好,因此,能够更准确、更贴近用户偏好地确定推荐内容,提升了推荐内容的相关性。

[0092] 在一些实施方式中,可以实时根据感兴趣标签组合,从内容数据库中搜索对应的内容。在一些实施方式中,还可以预先确定某些标签组合,例如较多用户感兴趣的标签组合,记为高频标签组合,预先筛选与高频标签组合对应的至少一个内容,形成内容队列。在接收到内容推荐请求时,基于高频标签组合对应的内容队列快速确定推荐内容。具体的,将内容数据库中 with 感兴趣标签组合对应的内容,确定为推荐内容,包括:

[0093] 在感兴趣标签组合为高频标签组合的情况下,根据预先确定的高频标签组合与内容队列之间的映射关系,确定与感兴趣标签组合对应的内容队列;

[0094] 在内容数据库中选择内容队列中的内容,并将选取的内容确定为推荐内容。

[0095] 示例性地,可以采用key-value存储方式存储与至少一个高频标签组合分别对应的至少一个内容队列。

[0096] 根据该实施方式,预先设置高频标签组合的内容队列,在接收到内容推荐请求时,可以利用高频标签组合与内容队列之间的映射关系快速地从内容数据库中选择到推荐内容,提高推荐效率。

[0097] 示例性地,针对某个标签组合,例如感兴趣标签组合或者高频标签组合,确定其对应的内容的过程可以包括:

[0098] 根据内容数据库中各内容的至少两个标签,确定各内容的至少一个标签组合;

[0099] 根据各内容的至少一个标签组合、各内容的热度、各内容的时效性中的至少之一,计算各内容与待匹配的标签组合例如感兴趣标签组合或高频标签组合之间的相似度;例如先根据各内容的至少一个标签组合,筛选出包含待匹配的标签组合的内容,再根据筛选出的内容的标签、热度和时效性,利用相似度预测模型计算内容与待匹配的标签组合的相似度;

[0100] 将相似度符合预设条件的内容,例如相似度大于一定阈值的内容或者相似度最大的P个内容,确定为与待匹配的标签组合对应的内容。其中,P为预先设置的正整数。

[0101] 如此,可以针对各标签组合,获得更准确的与标签组合对应的内容。

[0102] 本公开实施例还提供一种确定高频标签组合的示例性的实施方式。具体的,内容推荐方法还可以包括:

[0103] 确定与至少一个用户标识分别对应的至少一个候选标签组合集;

[0104] 根据预先配置的至少一个标签组合中的每个标签组合在至少一个候选标签组合集中的出现次数,从至少一个标签组合中确定出至少一个高频标签组合。

[0105] 也就是说,不同用户对应于不同的候选标签组合集。离线统计各用户的候选标签组合集,如果一个标签组合在多个候选标签组合集中出现,则该标签组合与较多用户的相关性较高,可以将其确定为高频标签组合。

[0106] 示例性地,可以将出现次数大于Q的标签组合确定为高频标签组合,也就是说,如果一个标签组合与至少Q个用户具有较高的相关性,则可以将该标签组合确定为高频标签组合。其中,Q可以根据电子设备的承载能力设置的阈值。

[0107] 根据上述实施方式,可以准确确定高频标签组合。从而提高预先确定的内容队列的利用率,减少实时确定标签组合对应的内容的需求和次数,降低电子设备的处理压力,提高推荐效率。

[0108] 根据本公开实施例,还可以帮助内容数据库的管理人员增补内容。具体的,内容推荐方法还可以包括:

[0109] 针对至少一个高频标签组合中的每个高频标签组合,在与高频标签组合对应的内容队列所包含的内容的数量小于第二预设阈值的情况下,在内容数据库中增加与高频标签组合对应的内容。

[0110] 根据上述步骤,对于大多数用户感兴趣的标签组合,如果内容数据库中较少相关的内容,则可以及时进行增补,促进内容推荐系统更好的运转。用于企业知识管理系统,可以主动发现企业内部对于知识的需求是否得到满足,引入更多员工感兴趣的知识。

[0111] 图3示出了内容推荐系统的一个应用示例的示意图。该应用示例以企业知识推荐为例,其中,内容数据库为知识数据库。内容为知识,或者称为文档知识、知识文档。如图3所示,用户可以在推荐用户界面310对知识进行点击、收藏、取消收藏等交互行为。实时交互行为触发用户兴趣标签组合画像320的更新。这里,用户兴趣标签组合画像可以是用户的候选标签组合集中各标签组合的实时权重。并且该实时交互行为触发行为日志生成和落盘,对应的行为日志将存储于行为日志数据库360中。利用行为日志数据库360中记录的用户的各种交互行为的类型和数量,可以对用户兴趣标签组合画像进行离线定期矫正,即更新用户的候选标签组合集。基于用户兴趣标签组合画像320,在接收到内容推荐请求时,可以确定兴趣标签组合对330。兴趣标签组合对也可以称为感兴趣标签组合,其相关知识可以作为推荐内容。

[0112] 此外,可以定期进行相关知识召回340。例如当高频标签组合的相关知识较少,即高频标签组合中存在低知识覆盖率标签组合350时,可以补充低覆盖知识。根据补充的文档知识370,完成相关知识召回340。基于召回的相关知识离线更新兴趣标签组合对所对应的的相关知识。

[0113] 可见,根据本公开实施例提供的内容推荐方法,当用户对内容数据库中的目标内容做出交互行为时,先确定出目标内容的标签组合,根据交互行为的类型和目标内容的标签组合更新候选标签组合集中的标签组合的实时权重。由于用于确定推荐内容的该权重是以标签组合为单位更新的,标签组合能够更准确表征用户偏好,并且,该权重是基于交互行为的类型实时更新的,因此,基于该权重能够更准确、更贴近用户偏好地确定推荐内容,提升了推荐内容的相关性。

[0114] 作为对上述各方法的实现,本公开还提供一种内容推荐装置。图4示出了本公开一个实施例提供的内容推荐装置的示意图。如图4所示,该装置400包括:

[0115] 交互响应模块410,用于响应于针对内容数据库中的目标内容的交互行为,根据目标内容的N个标签,确定目标内容的M个标签组合;其中,N为大于等于2的整数,M为正整数;

[0116] 权重更新模块420,用于根据交互行为的类型和M个标签组合,更新候选标签组合集中的标签组合的实时权重;

[0117] 推荐模块430,用于在接收到内容推荐请求的情况下,根据候选标签组合集中的标签组合当前的实时权重,在内容数据库中确定出推荐内容。

[0118] 如图5所示,在一种实施方式中,权重更新模块420包括:

[0119] 第一更新单元421,用于在候选标签组合集包含M个标签组合中的第i个标签组合的情况下,根据交互行为的类型,更新第i个标签组合的实时权重。

[0120] 在一种实施方式中,权重更新模块420,包括:

[0121] 第二更新单元422,用于在候选标签组合集不包含M个标签组合中的第j个标签组合的情况下,若候选标签组合集中的标签组合的数量小于第一预设阈值,则在候选标签组合集中添加第j个标签组合,并根据交互行为的类型,确定第j个标签组合的实时权重。

[0122] 如图5所示,在一种实施方式中,的装置还包括:

[0123] 权重确定模块440,用于对在预定时段内触发的针对内容数据库中的各个内容的交互行为进行统计,得到K个标签组合的统计权重;其中,K为正整数;

[0124] 组合选取模块450,用于根据K个标签组合的统计权重,在K个标签组合中选取L个标签组合;其中,L为小于等于K的正整数;

[0125] 第一集合确定模块460,用于基于L个标签组合得到候选标签组合集。

[0126] 如图5所示,在一种实施方式中,推荐模块430包括:

[0127] 组合筛选单元431,用于根据候选标签组合集中的各个标签组合的实时权重,从候选标签组合集中确定出感兴趣标签组合;

[0128] 内容确定单元432,用于将内容数据库中与感兴趣标签组合对应的内容,确定为推荐内容。

[0129] 在一种实施方式中,内容确定单元432用于:

[0130] 在感兴趣标签组合为高频标签组合的情况下,根据预先确定的高频标签组合与内容队列之间的映射关系,确定与感兴趣标签组合对应的内容队列;

[0131] 在内容数据库中选取内容队列中的内容,并将选取的内容确定为推荐内容。

[0132] 如图5所示,在一种实施方式中,还包括:

[0133] 第二集合确定模块470,用于确定与至少一个用户标识分别对应的至少一个候选标签组合集;

[0134] 高频确定模块480,用于根据预先配置的至少一个标签组合中的每个标签组合在至少一个候选标签组合集中的出现次数,从至少一个标签组合中确定出至少一个高频标签组合。

[0135] 在一种实施方式中,还包括:

[0136] 内容增加模块490,用于针对至少一个高频标签组合中的每个高频标签组合,在与高频标签组合对应的内容队列所包含的内容的数量小于第二预设阈值的情况下,在内容数据库中增加与高频标签组合对应的内容。

[0137] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0138] 图6示出了可以用来实施本公开的实施例的示例电子设备600的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限

制本文中描述的和/或要求的本公开的实现。

[0139] 如图6所示,电子设备600包括计算单元601,其可以根据存储在只读存储器 (ROM) 602中的计算机程序或者从存储单元608加载到随机访问存储器 (RAM) 603中的计算机程序来执行各种适当的动作和处理。在RAM 603中,还可存储电子设备600操作所需的各种程序和数据。计算单元601、ROM 602以及RAM 603通过总线604彼此相连。输入输出 (I/O) 接口605也连接至总线604。

[0140] 电子设备600中的多个部件连接至I/O接口605,包括:输入单元606,例如键盘、鼠标等;输出单元607,例如各种类型的显示器、扬声器等;存储单元608,例如磁盘、光盘等;以及通信单元609,例如网卡、调制解调器、无线通信收发机等。通信单元609允许电子设备600通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0141] 计算单元601可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元601的一些示例包括但不限于中央处理单元 (CPU)、图形处理单元 (GPU)、各种专用的人工智能 (AI) 计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器 (DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元601执行上文所描述的各个方法和处理,例如内容推荐方法。例如,在一些实施例中,内容推荐方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元608。在一些实施例中,计算机程序的部分或者全部可以经由ROM 602和/或通信单元609而被载入和/或安装到电子设备600上。当计算机程序加载到RAM 603并由计算单元601执行时,可以执行上文描述的内容推荐方法的一个或多个步骤。备选地,在其他实施例中,计算单元601可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行内容推荐方法。

[0142] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列 (FPGA)、专用集成电路 (ASIC)、专用标准产品 (ASSP)、芯片上系统的系统 (SOC)、负载可编程逻辑设备 (CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0143] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0144] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、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 (EPROM)

或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0145] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入、或者触觉输入来接收来自用户的输入。

[0146] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)和互联网。

[0147] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。

[0148] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0149] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

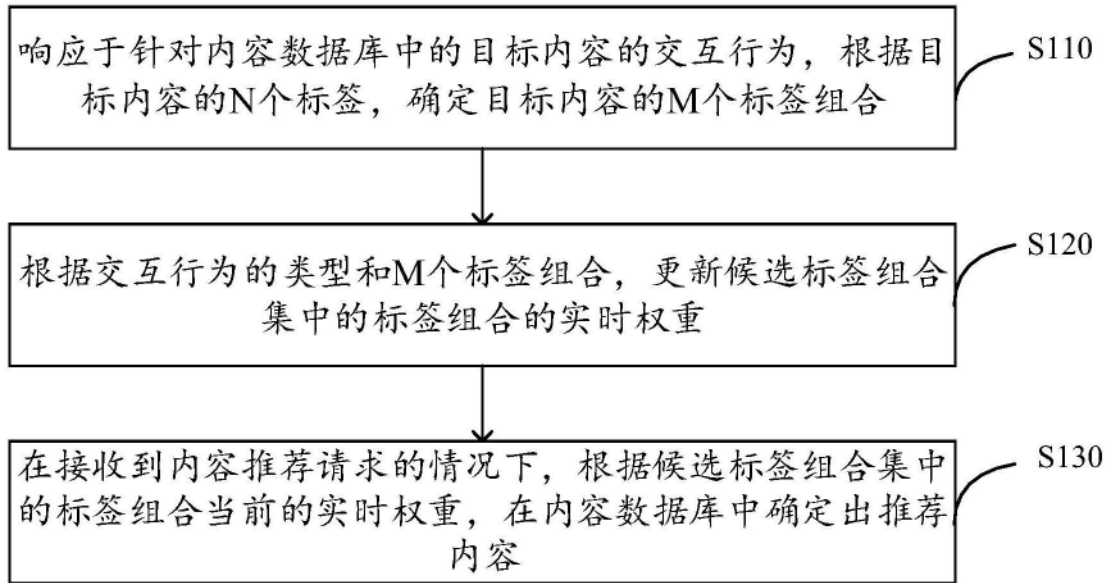


图1

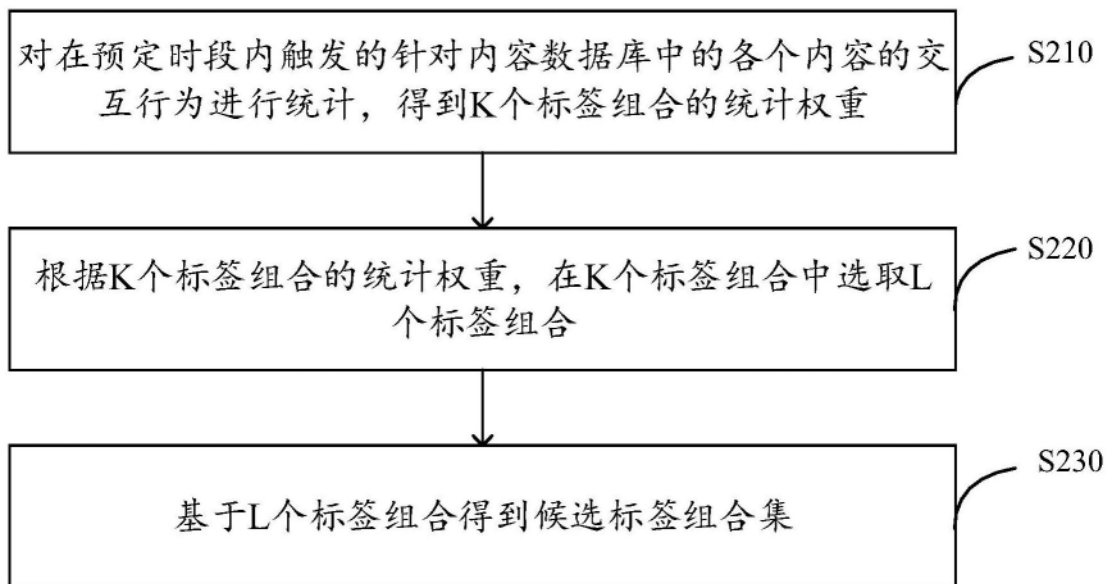


图2

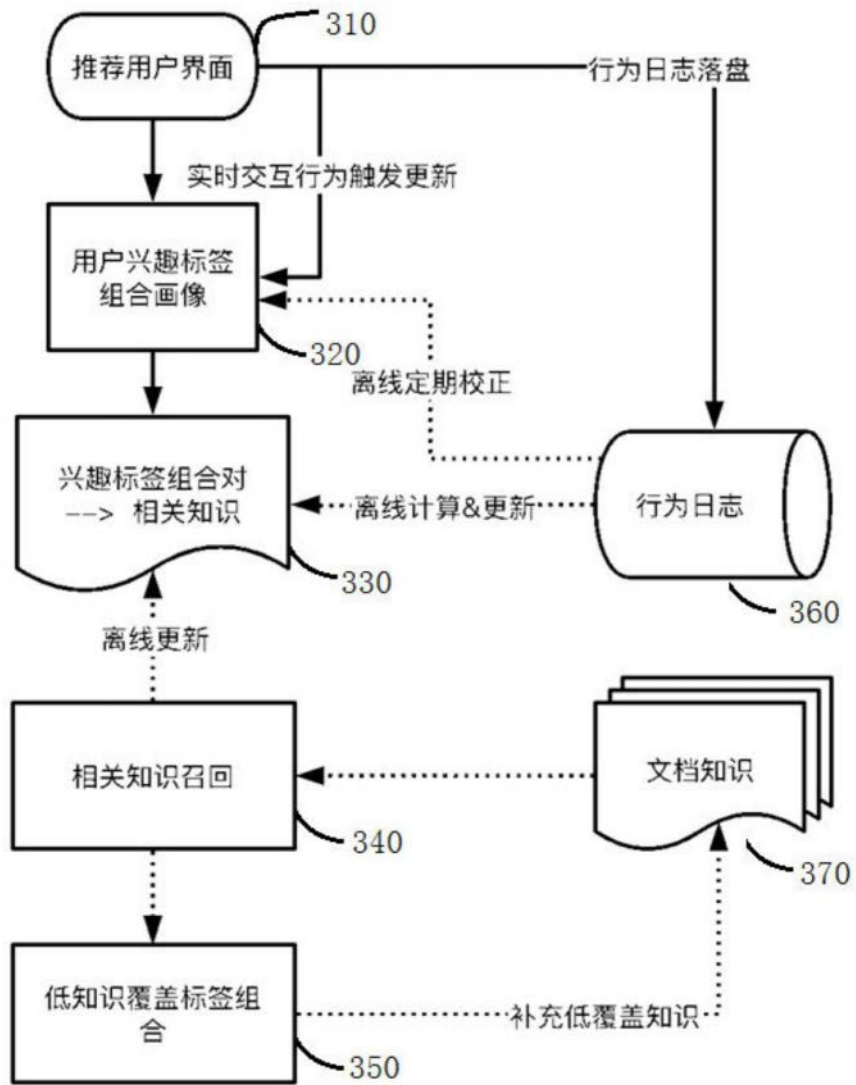


图3

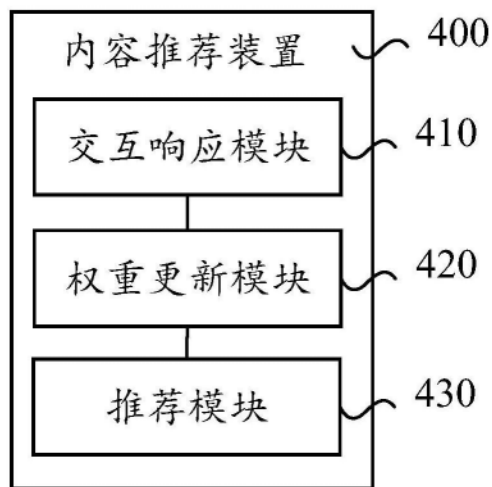


图4

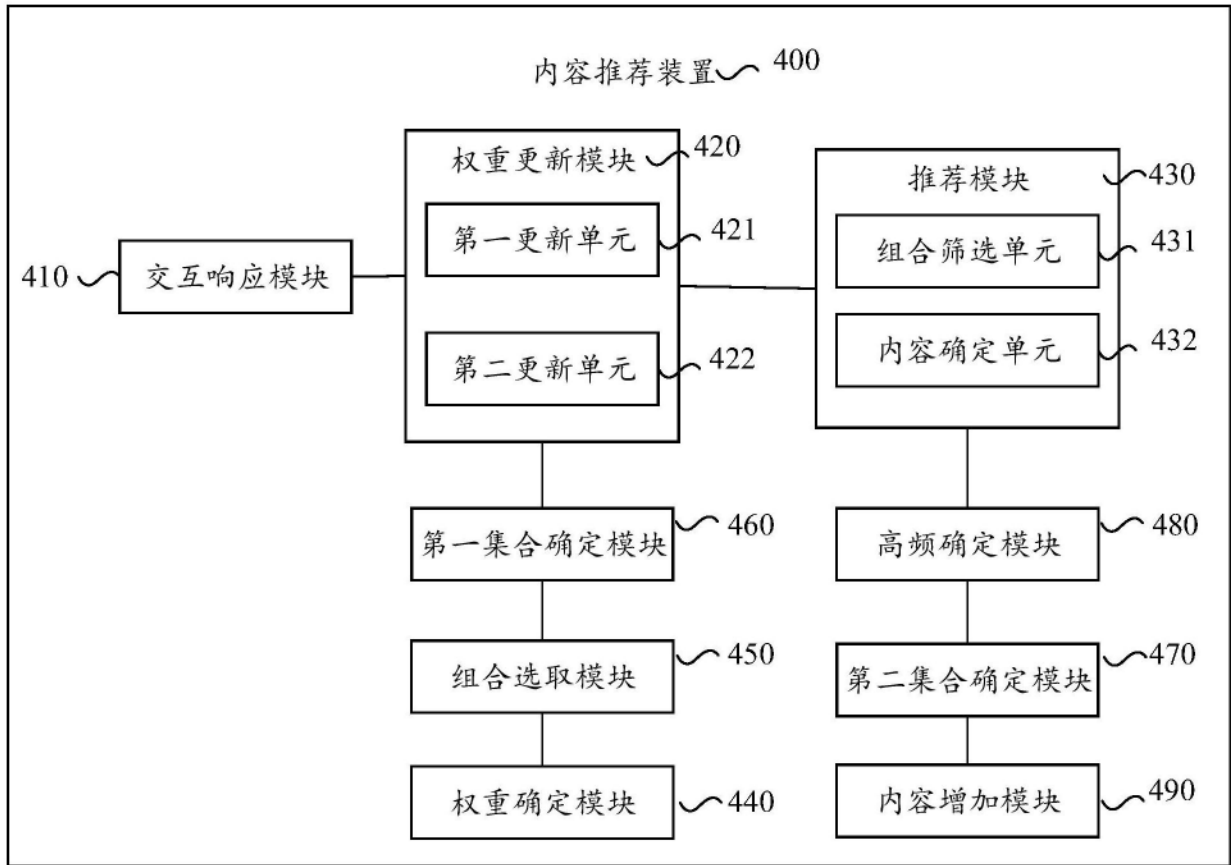


图5

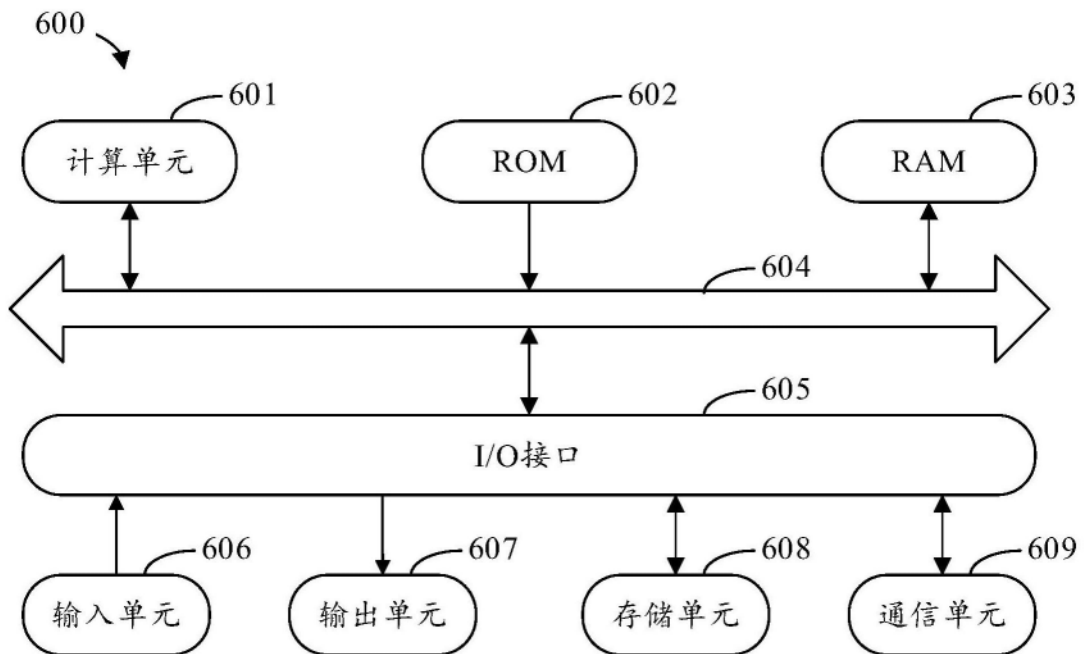


图6