



(12) 发明专利

(10) 授权公告号 CN 110866181 B

(45) 授权公告日 2022.04.22

(21) 申请号 201910970985.3
 (22) 申请日 2019.10.12
 (65) 同一申请的已公布的文献号
 申请公布号 CN 110866181 A
 (43) 申请公布日 2020.03.06
 (73) 专利权人 平安国际智慧城市科技股份有限公司
 地址 518000 广东省深圳市前海深港合作区妈湾兴海大道3048号前海自贸大厦1-34层
 (72) 发明人 陆园丽 余玉霞
 (74) 专利代理机构 深圳市沃德知识产权代理事务所(普通合伙) 44347
 代理人 高杰 于志光
 (51) Int. Cl.
 G06F 16/9535 (2019.01)

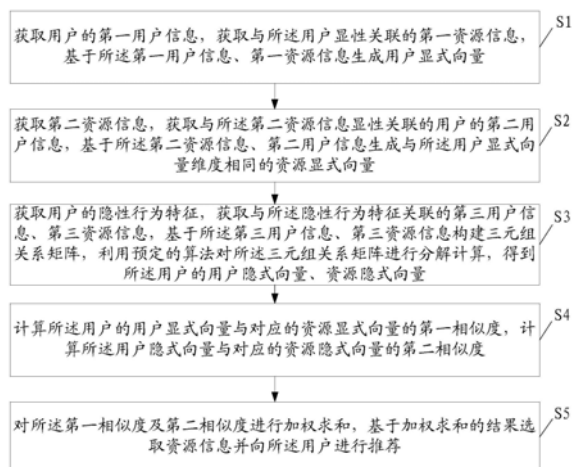
(56) 对比文件
 CN 106815297 A, 2017.06.09
 CN 109241405 A, 2019.01.18
 CN 105589916 A, 2016.05.18
 CN 110188208 A, 2019.08.30
 US 2019114687 A1, 2019.04.18
 CN 104462385 A, 2015.03.25
 CN 109947983 A, 2019.06.28
 陈登科.“基于潜在向量模型与项目的协同过滤混合推荐”.《中国优秀硕士学位论文全文数据库(信息科技辑)》.2010,(第2010年08期),
 Jichun Zhao等.“A recommendation algorithm based on collaborative filtering technology in distance learning”.《2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)》.2017,

审查员 姚晓斌

权利要求书3页 说明书10页 附图2页

(54) 发明名称
 资源推荐的方法、装置及存储介质

(57) 摘要
 本发明涉及一种数据分析技术,揭露了一种资源推荐的方法、装置及存储介质,该方法包括:获取用户的第一用户信息,获取与用户显性关联的第一资源信息,生成用户显式向量;获取第二资源信息,获取与第二资源信息显性关联的用户的第二用户信息,生成资源显式向量;获取用户的隐性行为特征,获取与隐性行为特征关联的第三用户信息、第三资源信息,构建三元组关系矩阵,对三元组关系矩阵进行分解计算,得到用户的用户隐式向量、资源隐式向量;计算用户的用户显式向量与对应的资源显式向量的第一相似度,计算用户隐式向量与对应的资源隐式向量的第二相似度;进行加权求和,基于加权求和的结果选取资源信息并推荐。本发明能够提高资源推荐的准确性。



CN 110866181 B

1. 一种资源推荐的方法,其特征在于,所述资源推荐的方法包括:

获取用户的第一用户信息,获取与所述用户显性关联的第一资源信息,基于所述第一用户信息、第一资源信息生成用户显式向量;

获取第二资源信息,获取与所述第二资源信息显性关联的用户的第二用户信息,基于所述第二资源信息、第二用户信息生成与所述用户显式向量维度相同的资源显式向量;

获取用户的隐性行为特征,获取与所述隐性行为特征关联的第三用户信息、第三资源信息,基于所述第三用户信息、第三资源信息构建三元组关系矩阵,利用预定的算法对所述三元组关系矩阵进行分解计算,得到所述用户的用户隐式向量、资源隐式向量;

计算所述用户的用户显式向量与对应的资源显式向量的第一相似度,计算所述用户隐式向量与对应的资源隐式向量的第二相似度;

对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐。

2. 根据权利要求1所述的资源推荐的方法,其特征在于,所述第一用户信息包括用户的基本信息及行为信息,所述第一资源信息中包括具有相同或不同业务属性的各资源信息,所述基于所述第一用户信息、第一资源信息生成用户显式向量的步骤,具体包括:

获取预先定义的多维向量,所述多维向量包括基础信息及业务属性;

基于所述第一用户信息的基本信息为所述多维向量中的基础信息赋值;

基于所述第一资源信息中的各资源信息,以及所述第一用户信息的基本信息及行为信息,为所述多维向量中的业务属性赋值;

以赋值后的多维向量作为所述用户显式向量。

3. 根据权利要求2所述的资源推荐的方法,其特征在于,所述行为信息包括显性行为特征及隐性行为特征,所述基于所述第一资源信息中的各资源信息,以及所述第一用户信息的基本信息及行为信息,为所述多维向量中的业务属性赋值的步骤,具体包括:

获取所述用户对所述第一资源信息中的各资源信息操作时产生的显性行为特征及时间信息,基于所述显性行为特征及时间信息计算所述用户对相应的资源信息的偏好程度,以该偏好程度作为对应的业务属性的值;或者

基于所述基本信息对用户进行分组,通过群组内关联规则预测每个分组内用户对所述第一资源信息中的各资源信息的偏好程度,以该偏好程度作为对应的业务属性的值。

4. 根据权利要求1至3任一项所述的资源推荐的方法,其特征在于,所述对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐的步骤,具体包括:

分别将所述第一相似度及第二相似度归一化,获取预定的权值,基于归一化后的第一相似度、归一化后的第二相似度及权值进行加权求和,得到总相似度;

获取各资源信息的上架时间及热度,基于所述总相似度、各资源信息的上架时间及热度选取多个资源信息并向所述用户进行推荐。

5. 根据权利要求4所述的资源推荐的方法,其特征在于,所述基于所述总相似度、各资源信息的上架时间及热度选取多个资源信息并向所述用户进行推荐的步骤,具体包括:

基于所述总相似度、各资源信息的上架时间及热度计算各资源信息的优先级,根据所述各资源信息的优先级选取多个资源信息并向所述用户进行推荐。

6. 一种资源推荐的装置,其特征在於,所述资源推荐的装置包括存储器及与所述存储器连接的处理器,所述存储器中存储有可在所述处理器上运行的处理系统,所述处理系统被所述处理器执行时实现如下步骤:

获取用户的第一用户信息,获取与所述用户显性关联的第一资源信息,基于所述第一用户信息、第一资源信息生成用户显式向量;

获取第二资源信息,获取与所述第二资源信息显性关联的用户的第二用户信息,基于所述第二资源信息、第二用户信息生成与所述用户显式向量维度相同的资源显式向量;

获取用户的隐性行为特征,获取与所述隐性行为特征关联的第三用户信息、第三资源信息,基于所述第三用户信息、第三资源信息构建三元组关系矩阵,利用预定的算法对所述三元组关系矩阵进行分解计算,得到所述用户的用户隐式向量、资源隐式向量;

计算所述用户的用户显式向量与对应的资源显式向量的第一相似度,计算所述用户隐式向量与对应的资源隐式向量的第二相似度;

对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐。

7. 根据权利要求6所述的资源推荐的装置,其特征在於,所述第一用户信息包括用户的基本信息及行为信息,所述第一资源信息中包括具有相同或不同业务属性的各资源信息,所述基于所述第一用户信息、第一资源信息生成用户显式向量的步骤,具体包括:

获取预先定义的多维向量,所述多维向量包括基础信息及业务属性;

基于所述第一用户信息的基本信息为所述多维向量中的基础信息赋值;

基于所述第一资源信息中的各资源信息,以及所述第一用户信息的基本信息及行为信息,为所述多维向量中的业务属性赋值;

以赋值后的多维向量作为所述用户显式向量。

8. 根据权利要求7所述的资源推荐的装置,其特征在於,所述行为信息包括显性行为特征及隐性行为特征,所述基于所述第一资源信息中的各资源信息,以及所述第一用户信息的基本信息及行为信息,为所述多维向量中的业务属性赋值的步骤,具体包括:

获取所述用户对所述第一资源信息中的各资源信息操作时产生的显性行为特征及时间信息,基于所述显性行为特征及时间信息计算所述用户对相应的资源信息的偏好程度,以该偏好程度作为对应的业务属性的值;或者

基于所述基本信息对用户进行分组,通过群组内关联规则预测每个分组内用户对所述第一资源信息中的各资源信息的偏好程度,以该偏好程度作为对应的业务属性的值。

9. 根据权利要求6至8任一项所述的资源推荐的装置,其特征在於,所述对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐的步骤,具体包括:

分别将所述第一相似度及第二相似度归一化,获取预定的权值,基于归一化后的第一相似度、归一化后的第二相似度及权值进行加权求和,得到总相似度;

获取各资源信息的上架时间及热度,基于所述总相似度、各资源信息的上架时间及热度选取多个资源信息并向所述用户进行推荐。

10. 一种计算机可读存储介质,其特征在於,所述计算机可读存储介质上存储有处理系统,所述处理系统被处理器执行时实现如权利要求1至5中任一项所述的资源推荐的方法的

步骤。

资源推荐的方法、装置及存储介质

技术领域

[0001] 本发明涉及数据分析技术领域,尤其涉及一种资源推荐的方法、装置及存储介质。

背景技术

[0002] 用户画像及资源画像是提高推荐系统准确性的重要方式,全面精准的标签可以充分体现用户特征及资源特征,依据画像形成的特征可为用户生成个性化资源池,从而实现千人千面的效果,提高推荐的精准度,同时,提升用户满意度。

[0003] 目前,画像在推荐系统应用中,采用两种方法进行资源的筛选及并形成预测:一种是采用显性特征(例如,内容和/或属性等相似的特征)进行资源筛选,这种方法一般需要进行大量的特征工程找到合适的特征组合,特征组合的效果在一定程度上决定着最终筛选及预测效果的好坏,准确性有待提高;另一种是为采用机器学习算法计算隐式特征(例如,内容和/或属性等不相似,但存在一定关联的特征)进行资源筛选,在数据量很大的情况下,这种方法在一定程度上能缓解数据稀疏性,但是存在资源更新慢、结果的可解释性低的特点,准确性也有待提高。

发明内容

[0004] 本发明的目的在于提供一种资源推荐的方法、装置及存储介质,旨在提高资源推荐的准确性。

[0005] 为实现上述目的,本发明提供一种资源推荐的方法,所述资源推荐的方法包括:

[0006] 获取用户的第一用户信息,获取与所述用户显性关联的第一资源信息,基于所述第一用户信息、第一资源信息生成用户显式向量;

[0007] 获取第二资源信息,获取与所述第二资源信息显性关联的用户的第二用户信息,基于所述第二资源信息、第二用户信息生成与所述用户显式向量维度相同的资源显式向量;

[0008] 获取用户的隐性行为特征,获取与所述隐性行为特征关联的第三用户信息、第三资源信息,基于所述第三用户信息、第三资源信息构建三元组关系矩阵,利用预定的算法对所述三元组关系矩阵进行分解计算,得到所述用户的用户隐式向量、资源隐式向量;

[0009] 计算所述用户的用户显式向量与对应的资源显式向量的第一相似度,计算所述用户隐式向量与对应的资源隐式向量的第二相似度;

[0010] 对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐。

[0011] 优选地,所述第一用户信息包括用户的基本信息及行为信息,所述第一资源信息中包括具有相同或不同业务属性的各资源信息,所述基于所述第一用户信息、第一资源信息生成用户显式向量的步骤,具体包括:

[0012] 获取预先定义的多维向量,所述多维向量包括基础信息及业务属性;

[0013] 基于所述基本信息为所述多维向量中的基础信息赋值,基于所述第一资源信息中

的各资源信息、基本信息及行为信息为所述多维向量中的业务属性赋值,以赋值后的多维向量作为所述用户显式向量。

[0014] 优选地,所述行为信息包括显性行为特征及隐性行为特征,所述基于所述第一资源信息中的各资源信息、基本信息及行为信息为所述多维向量中的业务属性赋值的步骤,具体包括:

[0015] 获取所述用户对所述第一资源信息中的各资源信息操作时产生的显性行为特征及时间信息,基于所述显性行为特征及时间信息计算所述用户对相应的资源信息的偏好程度,以该偏好程度作为对应的业务属性的值;或者

[0016] 基于所述基本信息对用户进行分组,通过群组内关联规则预测每个分组内用户对所述第一资源信息中的各资源信息的偏好程度,以该偏好程度作为对应的业务属性的值。

[0017] 优选地,所述对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐的步骤,具体包括:

[0018] 分别将所述第一相似度及第二相似度归一化,获取预定的权值,基于归一化后的第一相似度、归一化后的第二相似度及权值进行加权求和,得到总相似度;

[0019] 获取各资源信息的上架时间及热度,基于所述总相似度、各资源信息的上架时间及热度选取多个资源信息并向所述用户进行推荐。

[0020] 优选地,所述基于所述总相似度、各资源信息的上架时间及热度选取多个资源信息并向所述用户进行推荐的步骤,具体包括:

[0021] 基于所述总相似度、各资源信息的上架时间及热度计算各资源信息的优先级,根据所述各资源信息的优先级选取多个资源信息并向所述用户进行推荐。

[0022] 为实现上述目的,本发明还提供一种资源推荐的装置,所述资源推荐的装置包括存储器及与所述存储器连接的处理器,所述存储器中存储有可在所述处理器上运行的处理系统,所述处理系统被所述处理器执行时实现如下步骤:

[0023] 获取用户的第一用户信息,获取与所述用户显性关联的第一资源信息,基于所述第一用户信息、第一资源信息生成用户显式向量;

[0024] 获取第二资源信息,获取与所述第二资源信息显性关联的用户的第二用户信息,基于所述第二资源信息、第二用户信息生成与所述用户显式向量维度相同的资源显式向量;

[0025] 获取用户的隐性行为特征,获取与所述隐性行为特征关联的第三用户信息、第三资源信息,基于所述第三用户信息、第三资源信息构建三元组关系矩阵,利用预定的算法对所述三元组关系矩阵进行分解计算,得到所述用户的用户隐式向量、资源隐式向量;

[0026] 计算所述用户的用户显式向量与对应的资源显式向量的第一相似度,计算所述用户隐式向量与对应的资源隐式向量的第二相似度;

[0027] 对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐。

[0028] 优选地,所述第一用户信息包括用户的基本信息及行为信息,所述第一资源信息中包括具有相同或不同业务属性的各资源信息,所述基于所述第一用户信息、第一资源信息生成用户显式向量的步骤,具体包括:

[0029] 获取预先定义的多维向量,所述多维向量包括基础信息及业务属性;

[0030] 基于所述基本信息为所述多维向量中的基础信息赋值,基于所述第一资源信息中的各资源信息、基本信息及行为信息为所述多维向量中的业务属性赋值,以赋值后的多维向量作为所述用户显式向量。

[0031] 优选地,所述行为信息包括显性行为特征及隐性行为特征,所述基于所述第一资源信息中的各资源信息、基本信息及行为信息为所述多维向量中的业务属性赋值的步骤,具体包括:

[0032] 获取所述用户对所述第一资源信息中的各资源信息操作时产生的显性行为特征及时间信息,基于所述显性行为特征及时间信息计算所述用户对相应的资源信息的偏好程度,以该偏好程度作为对应的业务属性的值;或者

[0033] 基于所述基本信息对用户进行分组,通过群组内关联规则预测每个分组内用户对所述第一资源信息中的各资源信息的偏好程度,以该偏好程度作为对应的业务属性的值。

[0034] 优选地,所述对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐的步骤,具体包括:

[0035] 分别将所述第一相似度及第二相似度归一化,获取预定的权值,基于归一化后的第一相似度、归一化后的第二相似度及权值进行加权求和,得到总相似度;

[0036] 获取各资源信息的上架时间及热度,基于所述总相似度、各资源信息的上架时间及热度选取多个资源信息并向所述用户进行推荐。

[0037] 本发明还提供一种计算机可读存储介质,所述计算机可读存储介质上存储有处理系统,所述处理系统被处理器执行时实现上述的资源推荐的方法的步骤。

[0038] 本发明的有益效果是:本发明首先生成用户显式向量、资源显式向量,然后基于用户的隐性行为特征构建三元组关系矩阵,分解该三元组关系矩阵得到用户隐式向量、资源隐式向量,最后计算用户的用户显式向量与对应的资源显式向量的第一相似度,计算用户隐式向量与对应的资源隐式向量的第二相似度,对第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向用户进行推荐,本发明在用户显性特征及资源显性特征的基础上,融合了用户隐性特征及资源隐性特征,对现有的推荐算法进行修正,能够提高资源推荐准确度,同时增强系统的可解释性。

附图说明

[0039] 图1为本发明资源推荐的装置一实施例的硬件架构的示意图;

[0040] 图2为图1中处理系统一实施例的程序模块图;

[0041] 图3为本发明资源推荐的方法一实施例的流程示意图。

具体实施方式

[0042] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0043] 需要说明的是,在本发明中涉及“第一”、“第二”等的描述仅用于描述目的,而不能理解为指示或暗示其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。另外，各个实施例之间的技术方案可以相互结合，但是必须是以本领域普通技术人员能够实现为基础，当技术方案的结合出现相互矛盾或无法实现时应当认为这种技术方案的结合不存在，也不在本发明要求的保护范围之内。

[0044] 参阅图1所示，是本发明资源推荐的装置一实施例的硬件架构的示意图。资源推荐的装置1是一种能够按照事先设定或者存储的指令，自动进行数值计算和/或信息处理的设备。所述资源推荐的装置1可以是计算机、也可以是单个网络服务器、多个网络服务器组成的服务器组或者基于云计算的由大量主机或者网络服务器构成的云，其中云计算是分布式计算的一种，由一群松散耦合的计算机集组成的一个超级虚拟计算机。

[0045] 在本实施例中，资源推荐的装置1可包括，但不限于，可通过系统总线相互通信连接的存储器11、处理器12、网络接口13，存储器11存储有可在处理器12上运行的处理系统。需要指出的是，图1仅示出了具有组件11—13的资源推荐的装置1，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。

[0046] 其中，存储器11包括内存及至少一种类型的可读存储介质。内存为资源推荐的装置1的运行提供缓存；可读存储介质可为如闪存、硬盘、多媒体卡、卡型存储器（例如，SD或DX存储器等）、随机访问存储器（RAM）、静态随机访问存储器（SRAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、可编程只读存储器（PROM）、磁性存储器、磁盘、光盘等的非易失性存储介质。在一些实施例中，可读存储介质可以是资源推荐的装置1的内部存储单元，例如该资源推荐的装置1的硬盘；在另一些实施例中，该非易失性存储介质也可以是资源推荐的装置1的外部存储设备，例如资源推荐的装置1上配备的插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）等。本实施例中，存储器11的可读存储介质通常用于存储安装于资源推荐的装置1的操作系统和各类应用软件，例如存储本发明一实施例中的处理系统10的程序代码等。此外，存储器11还可以用于暂时地存储已经输出或者将要输出的各类数据。

[0047] 所述处理器12在一些实施例中可以是中央处理器（Central Processing Unit, CPU）、控制器、微控制器、微处理器、或其他数据处理芯片，用于运行所述存储器11中存储的程序代码或者处理数据，例如运行处理系统10等。

[0048] 所述网络接口13可包括标准的无线网络接口、有线网络接口，该网络接口13通常用于在所述资源推荐的装置1与其他电子设备之间建立通信连接。

[0049] 所述处理系统10存储在存储器11中，包括至少一个存储在存储器11中的计算机可读指令，该至少一个计算机可读指令可被处理器12执行，以实现本申请各实施例的方法；以及，该至少一个计算机可读指令依据其各部分所实现的功能不同，可被划为不同的逻辑模块。

[0050] 在一实施例中，上述处理系统10被所述处理器12执行时实现如下步骤：

[0051] 获取用户的第一用户信息，获取与所述用户显性关联的第一资源信息，基于所述第一用户信息、第一资源信息生成用户显式向量；

[0052] 获取第二资源信息，获取与所述第二资源信息显性关联的用户的第二用户信息，基于所述第二资源信息、第二用户信息生成与所述用户显式向量维度相同的资源显式向量；

[0053] 获取用户的隐性行为特征,获取与所述隐性行为特征关联的第三用户信息、第三资源信息,基于所述第三用户信息、第三资源信息构建三元组关系矩阵,利用预定的算法对所述三元组关系矩阵进行分解计算,得到所述用户的用户隐式向量、资源隐式向量;

[0054] 计算所述用户的用户显式向量与对应的资源显式向量的第一相似度,计算所述用户隐式向量与对应的资源隐式向量的第二相似度;

[0055] 对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐。

[0056] 优选地,所述第一用户信息包括用户的基本信息及行为信息,所述第一资源信息中包括具有相同或不同业务属性的各资源信息,所述基于所述第一用户信息、第一资源信息生成用户显式向量的步骤,具体包括:

[0057] 获取预先定义的多维向量,所述多维向量包括基础信息及业务属性;

[0058] 基于所述基本信息为所述多维向量中的基础信息赋值,基于所述第一资源信息中的各资源信息、基本信息及行为信息为所述多维向量中的业务属性赋值,以赋值后的多维向量作为所述用户显式向量。

[0059] 优选地,所述行为信息包括显性行为特征及隐性行为特征,所述基于所述第一资源信息中的各资源信息、基本信息及行为信息为所述多维向量中的业务属性赋值的步骤,具体包括:

[0060] 获取所述用户对所述第一资源信息中的各资源信息操作时产生的显性行为特征及时间信息,基于所述显性行为特征及时间信息计算所述用户对相应的资源信息的偏好程度,以该偏好程度作为对应的业务属性的值;或者

[0061] 基于所述基本信息对用户进行分组,通过群组内关联规则预测每个分组内用户对所述第一资源信息中的各资源信息的偏好程度,以该偏好程度作为对应的业务属性的值。

[0062] 优选地,所述对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐的步骤,具体包括:

[0063] 分别将所述第一相似度及第二相似度归一化,获取预定的权值,基于归一化后的第一相似度、归一化后的第二相似度及权值进行加权求和,得到总相似度;

[0064] 获取各资源信息的上架时间及热度,基于所述总相似度、各资源信息的上架时间及热度选取多个资源信息并向所述用户进行推荐。

[0065] 参照图2所示,为图1中处理系统10的程序模块图。所述处理系统10被分割为多个模块,该多个模块被存储于存储器12中,并由处理器13执行,以完成本发明。本发明所称的模块是指能够完成特定功能的一系列计算机程序指令段。

[0066] 所述处理系统10可以被分割为:第一生成模块101、第二生成模块102、分解模块103、计算模块104及推荐模块105。

[0067] 所述第一生成模块101,用于获取用户的第一用户信息,获取与所述用户显性关联的第一资源信息,基于所述第一用户信息、第一资源信息生成用户显式向量;

[0068] 所述第二生成模块102,用于获取第二资源信息,获取与所述第二资源信息显性关联的用户的第二用户信息,基于所述第二资源信息、第二用户信息生成与所述用户显式向量维度相同的资源显式向量;

[0069] 所述分解模块103,用于获取用户的隐性行为特征,获取与所述隐性行为特征关联

的第三用户信息、第三资源信息,基于所述第三用户信息、第三资源信息构建三元组关系矩阵,利用预定的算法对所述三元组关系矩阵进行分解计算,得到所述用户的用户隐式向量、资源隐式向量;

[0070] 所述计算模块104,用于计算所述用户的用户显式向量与对应的资源显式向量的第一相似度,计算所述用户隐式向量与对应的资源隐式向量的第二相似度;

[0071] 所述推荐模块105,用于对所述第一相似度及第二相似度进行加权求和,基于加权求和的结果选取资源信息并向所述用户进行推荐。

[0072] 具体原理请参照下述图3关于该方法的流程图的介绍。

[0073] 如图3所示,图3为本发明资源推荐的方法一实施例的流程示意图,资源推荐的装置1的处理器13执行存储器12中存储的处理系统10时实现该方法的如下步骤:

[0074] 步骤S1,获取用户的第一用户信息,获取与所述用户显性关联的第一资源信息,基于所述第一用户信息、第一资源信息生成用户显式向量;

[0075] 其中,第一用户信息包括用户的基本信息及行为信息,基本信息包括性别、年龄、消费能力、工作信息等,行为信息为用户在浏览或操作资源时的行为操作信息,可以从日志中获取,包括显性行为特征及隐性行为特征。显性行为特征能够直接反应用户对资源的喜好,显性行为特征例如为收藏、点赞、分享等,隐性行为特征不能够直接反应用户对资源的喜好,隐性行为特征例如为资源页面浏览时间、搜索关键字、评论、点击、鼠标滑动等。

[0076] 第一资源信息为网络上的资源信息,包括具有相同或不同业务属性的各资源信息,按照业务属性区分,例如资源信息可以是产品信息、销售信息、培训信息、人工智能信息等。

[0077] 其中,从用户的角度来看,用户在浏览或操作资源信息时的行为信息为显性行为特征,则该用户与该资源信息显性关联。

[0078] 其中,基于第一用户信息、第一资源信息生成用户显式向量的步骤,具体包括:

[0079] 预先定义多维的向量 $(a_1, a_2, \dots, a_j, b_1, b_2, \dots, b_k)$,其中,多维的向量以配置文件的形式支持可扩展、可配置操作。 a_1, a_2, \dots, a_j 为用户的各基础信息(包括性别、年龄、消费能力、工作信息等),其值为0或1,基于基本信息为多维向量中的基础信息赋值,对于离散变量可以直接获取对应的值,对于连续变量的值采用最小熵分箱方法进行离散化得到对应的值。例如,对于性别,性别为男对应的值为0,性别为女对应的值为1;对于年龄,20岁以上含20岁对应的值为0,20岁以下对应的值为1;对于工作信息,是作家对应的值为0,不是作家对应的值为1。

[0080] b_1, b_2, \dots, b_k 为与第一用户信息显性关联的第一资源信息的各资源信息的业务属性。对于第一资源信息的各资源信息的业务属性,可以通过以下方法确定:预先构建符合业务发展目标的业务属性标签结构,然后,对第一资源信息的各资源信息的文本信息进行抽取,对文本信息后续处理可采用现有技术,即分词、数据清洗、LDA主题提取、向量化、基于向量进行业务属性相似计算,当文本信息与相应的业务属性标签的相似度超过阈值时,则该第一资源信息的各资源信息的业务属性为该业务属性标签所指的属性。

[0081] 各业务属性的值可以通过以下预定的方式中的任一种获得:

[0082] 第一种方式为,获取所述用户对所述第一资源信息中的各资源信息操作时产生的显性行为特征及时间信息,基于所述显性行为特征及时间信息计算用户对相应的资源信息

的偏好程度,以该偏好程度作为对应的业务属性的值,即通过用户的显性行为特征及时间因子计算用户对各业务属性的资源信息的偏好程度,作为各业务属性的值:对于某一业务属性的资源信息,用户执行相应的行为是与时间密切相关的。从行为信息中获取用户的显性行为特征(例如,点赞、收藏等),采用如下公式计算用户的偏好程度 b :

$$[0083] \quad b = \frac{\alpha e^{\beta}}{(t+t_{\gamma})^c};$$

[0084] 其中, t 为该用户对该资源信息执行显性行为特征距离当前的天数, α 、 β 、 c 、 t_{γ} 均为常量参数, $\alpha > 0$ 、 $\beta > 0$ 、 $c > 0$ 、 α 、 β 、 c 、 t_{γ} 默认值分别为1、0.42、0.025、0.0025,当然,也可以依据该业务属性的数据变化,生成相应的 α 、 β 、 c 、 t_{γ} 的值。由于用户可能在不同时间对同一资源信息进行浏览或操作,因此可按照用户、业务属性进行偏好程度汇总,例如,用户在某一时间段与同一资源信息产生交互行为,可以取该时间段的最大偏好程度作为 b 的值,最后,将各业务属性对应的偏好程度分别对应作为各业务属性 b_1, b_2, \dots, b_k 的值。

[0085] 本实施例中提出融合时间因子计算用户的偏好程度,其通过时间因素的关联有利于提高资源推荐的准确性。

[0086] 第二种方式为基于所述基本信息对用户进行分组,通过群组内关联规则预测用户对各业务属性的资源信息的偏好程度,作为各业务属性的值:对用户进行分组,群组的定义可按照用户的数据规模判断,例如,如果用户的数据规模较小,则所有用户为整个群组,如果用户的数据规模较大,可以基于用户的基本信息,如按照地区或者行业进行分组,每个用户有对应的分组。在spark平台中,在每个组内利用关联规则(FP-Growth)算法对用户偏好的资源信息进行预测。具体地,构建各分组记为 $G = \{g_1, g_2, g_3, \dots, g_n\}$,其中 n 为组数。获取用户的显性行为特征,用户 u_i 的显性行为特征 v_i 记为 $\{u_i, v_i\}$,每个用户 u_i 有对应的分组 g_1 ,生成对应的关系 $R = \{r_1, r_2, \dots, r_m\}$, $r_i = \{g_i, v_i\}$,其中 m 为用户数,从而构造每个分组对应的资源频繁项。基于资源频繁项生成各群组偏好资源列表及推荐分值,进而依据关系 R 、各群组偏好资源列表及推荐分值获得(用户,偏好资源对应的业务属性的推荐分值),以推荐分值作为对应的业务属性 b_1, b_2, \dots, b_k 的值。

[0087] 本实施例中提出的群组内关联规则,一方面有助于解决关联规则算法计算资源消耗过大的问题,一方面有助于增强用户的群组效应,有利于提高资源推荐的准确性。

[0088] 第三种方式为融合上述的第一种方式及第二种方式,即将第一种方式中的业务属性的值 b_1, b_2, \dots, b_k 与第二种方式中的业务属性的值 b_1, b_2, \dots, b_k 进行对应后,进行加权求和,各权值可以预先确定。其中,第一种方式中各业务属性的值对应的权值均相同,例如均为0.55,第二种方式中各业务属性的值对应的权值均相同,例如均为0.45,加权求和后得到最终的业务属性 b_1, b_2, \dots, b_k 的值。

[0089] 上述的多维的向量 $(a_1, a_2, \dots, a_j, b_1, b_2, \dots, b_k)$ 的各值确定后,即生成了用户显式向量。

[0090] 步骤S2,获取第二资源信息,获取与所述第二资源信息显性关联的用户的第二用户信息,基于所述第二资源信息、第二用户信息生成与所述用户显式向量维度相同的资源显式向量;

[0091] 其中,第二资源信息也为网络上的资源信息,包括具有相同或不同业务属性的各资源信息。第二用户信息也包括用户的基本信息及行为信息。

[0092] 其中,从资源的角度来看,用户在浏览或操作资源信息时的行为信息为显性行为特征,则该资源信息显性与该用户显性关联,获取该用户的用户信息,所有显性关联的用户的用户信息构成第二用户信息。

[0093] 其中,基于所述第二资源信息、第二用户信息生成与所述用户显式向量维度相同的资源显式向量,具体包括:

[0094] 预先定义多维的向量 $(A_1, A_2, \dots, A_j, B_1, B_2, \dots, B_k)$,其中,多维的向量以配置文件的形式支持可扩展、可配置操作,其维度与上述的用户显式向量的维度相同。 B_1, B_2, \dots, B_k 为第二资源信息中的各资源信息的业务属性,通过以下方法确定:预先构建符合业务发展目标的业务属性标签结构,然后,对第二资源信息中的各资源信息的文本信息进行抽取,对文本信息后续处理可采用现有技术,即分词、数据清洗、LDA主题提取、向量化、基于向量进行业务属性相似计算,当文本信息与相应的业务属性标签的相似度超过阈值时,则该第二资源信息中的各资源信息的业务属性为该业务属性标签所指的属性,其值为与相应的业务属性标签的相似度。

[0095] A_1, A_2, \dots, A_j 为与第二资源信息显性关联的用户的各基础信息。第二资源信息显性关联的用户可能具有多种属性标签(例如男性、高消费群体、研发工程师等),这就使得用户的基础信息比较分散。一种业务属性的资源信息可能会被不同用户浏览及操作,但该种业务属性的资源信息实际上可能并不适用于所有的这些用户,故本实施例通过聚类的方式选取对应属性标签的用户,并基于这些属性标签的用户的基础信息得到 A_1, A_2, \dots, A_j 的值。

[0096] 本实施例聚类可采用kmeans算法:对用户进行分组(可以基于用户的基本信息,如按照地区或者行业进行分组),基于对用户组分析可以获得预设的中心点数量,设为k,对有历史行为信息记录的用户涉及的基础信息聚类,可获得基础信息的聚类中心,得到用户基础信息与聚类中心的关系[基础信息列表,聚类中心]。聚类之后,通过用户历史行为信息可获得关系[资源信息,基础信息列表],将这两种关系进行融合,可获得关系[资源信息,聚类中心],由于一个资源信息会对应多个聚类中心,因此依据属性类别用户总量取N个聚类中心的值,可默认取前3个聚类中心的值,则最终按照用户量占比作为权重进行加权求和,通过加权求和结果得到最终的 A_1, A_2, \dots, A_j 的值。

[0097] 本实施例对用户显式特征与资源显式特征进行向量化,得到用户显式向量与资源显式向量,避免了特征组合及相关大量的特征预处理的过程,能够降低计算的复杂度。

[0098] 步骤S3,获取所述用户的隐性行为特征,获取与所述隐性行为特征关联的第三用户信息、第三资源信息,基于所述第三用户信息、第三资源信息构建三元组关系矩阵,利用预定的算法对所述三元组关系矩阵进行分解计算,得到所述用户的用户隐式向量、资源隐式向量;

[0099] 本实施例中,可从日志中获取用户的隐性行为特征,获取与隐性行为特征关联的第三用户信息、第三资源信息,其中,第三用户信息也包括用户的基本信息,第三资源信息也为网络上的资源信息,包括具有相同或不同业务属性的各资源信息。基于第三用户信息、第三资源信息构建三元组关系矩阵 $R[user, product, rating]$,在该三元组关系矩阵中,包含m个user和n个product,user表示用户,product表示资源,rating表示评分(即偏好程度),本实施例将隐性行为特征对应的rating评分统一定义为1。

[0100] 在实际应用中,由于n和m的数量都十分巨大,因此三元组关系矩阵R的规模很大。

这时,传统的矩阵分解方法对于这么大的数据量难以处理;再者一个用户也不可能给所有资源product进行评分,因此,三元组关系矩阵R是个稀疏矩阵,有很多的缺失项。

[0101] 本实施例基于spark平台,利用预定的算法(交替最小二乘ALS)计算用户隐式向量及资源隐式向量,从而得到用户隐式向量及资源隐式向量。其中,由于三元组关系矩阵R为 $m*n$ 的矩阵,可以看做由 $m*k$ 和 $k*n$ 两个矩阵相乘得到的,其中 $k \ll m, n$, k 的典型取值一般是20~200,因此得到下式:

$$[0102] \quad R_{m*n} = u_{m*k} \times p_{k*n};$$

[0103] 上面的公式中, u_{m*k} 表示用户对隐性行为特征的偏好程度, p_{k*n} 表示资源包含隐性行为特征的程度,通过上式可计算出 u_{m*k} 、 p_{k*n} , 将计算得到的 u_{m*k} 作为用户隐式向量, 将 p_{k*n} 作为资源隐式向量。

[0104] 步骤S4, 计算所述用户的用户显式向量与对应的资源显式向量的第一相似度, 计算所述用户隐式向量与对应的资源隐式向量的第二相似度;

[0105] 由于直接计算用户显式向量与对应的资源显式向量之间的相似度存在计算量大且计算复杂度高的问题, 因此, 采用局部敏感哈希算法(Locality-Sensitive Hashing, LSH) 计算相似度。首先, 将用户显式向量与对应的资源显式向量进行hash映射处理, 指定特征列及唯一标识列, 将特征作为算法的输入, 指定算法输出列, 然后, 对于映射后的向量采用近似相似度连接的方法计算向量间的欧式距离作为相似度值:

[0106] 计算用户显式向量与资源显式向量之间的欧式距离, 在实现过程中采用上述描述的局部敏感哈希算法LSH获取第一相似度 $sim_{explicit_init}$;

[0107] 计算用户隐式向量与资源隐式向量之间的欧式距离, 在实现过程中采用上述描述局部敏感哈希算法LSH获取第二相似度 $sim_{implicit_init}$ 。

[0108] 步骤S5, 对所述第一相似度及第二相似度进行加权求和, 基于加权求和的结果选取资源并向所述用户进行推荐。

[0109] 分别将第一相似度 $sim_{explicit_init}$ 、第二相似度 $sim_{implicit_init}$ 归一化到(0,1)区间, 得到 $sim_{explicit}$ 、 $sim_{implicit}$, 对两组相似度值进行加权求和得到总相似度:

$$[0110] \quad Sim = \alpha * sim_{explicit} + \beta * sim_{implicit},$$

[0111] 其中, 权值 α 、 β 主要采用两种方式进行设定, 一种方式为专家评分, 设定 α 、 β 的固定值。另一种方式为通过线性回归进行确定, 通过对用户随机采样的方式, 将采样用户作为体验官, 让其对提供的资源进行相似度打分, 将打分结果作为训练数据, 生成 α 、 β 的值。

[0112] 最后, 依据加权求和得到的总相似度Sim、资源信息的上架时间及热度, 选取topN

个资源信息计算各资源信息的优先级, 优先级排序采用 $\frac{\lg(view) * j + Sim}{(\frac{age}{2} + 1)^i}$, view表示热度(默

认取前一天热度), age表示资源信息的上架时间距离当前的天数, 常量参数 i 、 j 均默认取1。最后, 可按照排序后的该topN个资源信息向用户推送。

[0113] 通过上述的描述可以得出, 本实施例在用户显性特征及资源显性特征的基础上, 融合了用户隐性特征及资源隐性特征, 对现有的推荐算法进行修正, 能够提高资源推荐准确度, 同时增强系统的可解释性。

[0114] 此外, 本发明实施例还提出一种计算机可读存储介质, 所述计算机可读存储介质

可以是硬盘、多媒体卡、SD卡、闪存卡、SMC、只读存储器 (ROM)、可擦除可编程只读存储器 (EPROM)、便携式紧致盘只读存储器 (CD-ROM)、USB存储器等等中的任意一种或者几种的任意组合。所述计算机可读存储介质中包括处理系统,该处理系统被处理器执行时实现的功能,请参照上述关于图3的介绍,在此不再赘述。

[0115] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0116] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、装置、物品或者方法不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、装置、物品或者方法所固有的要素。

[0117] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本发明各个实施例所述的方法。

[0118] 以上仅为本发明的优选实施例,并非因此限制本发明的专利范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

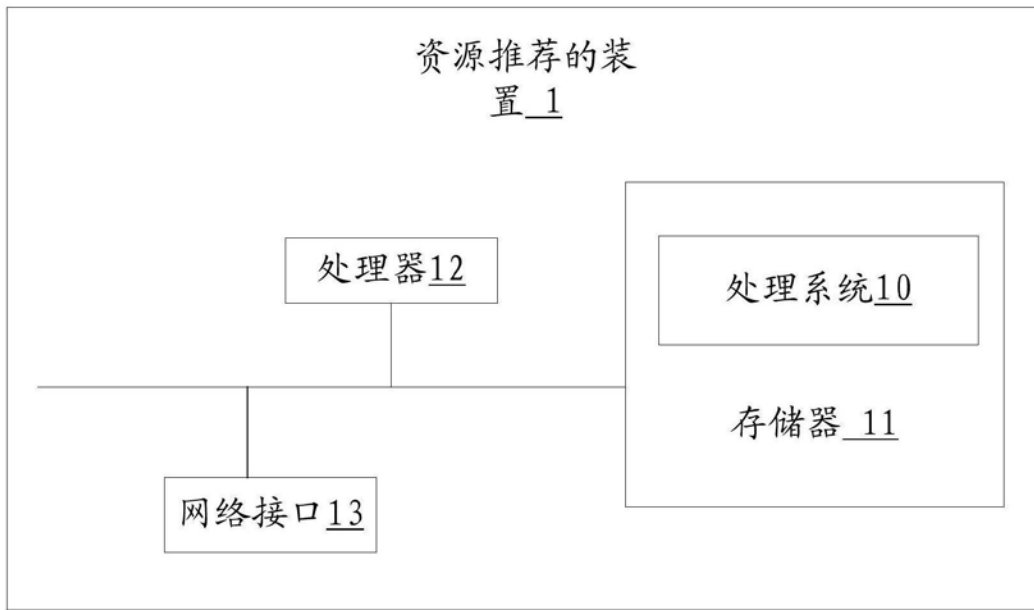


图1



图2

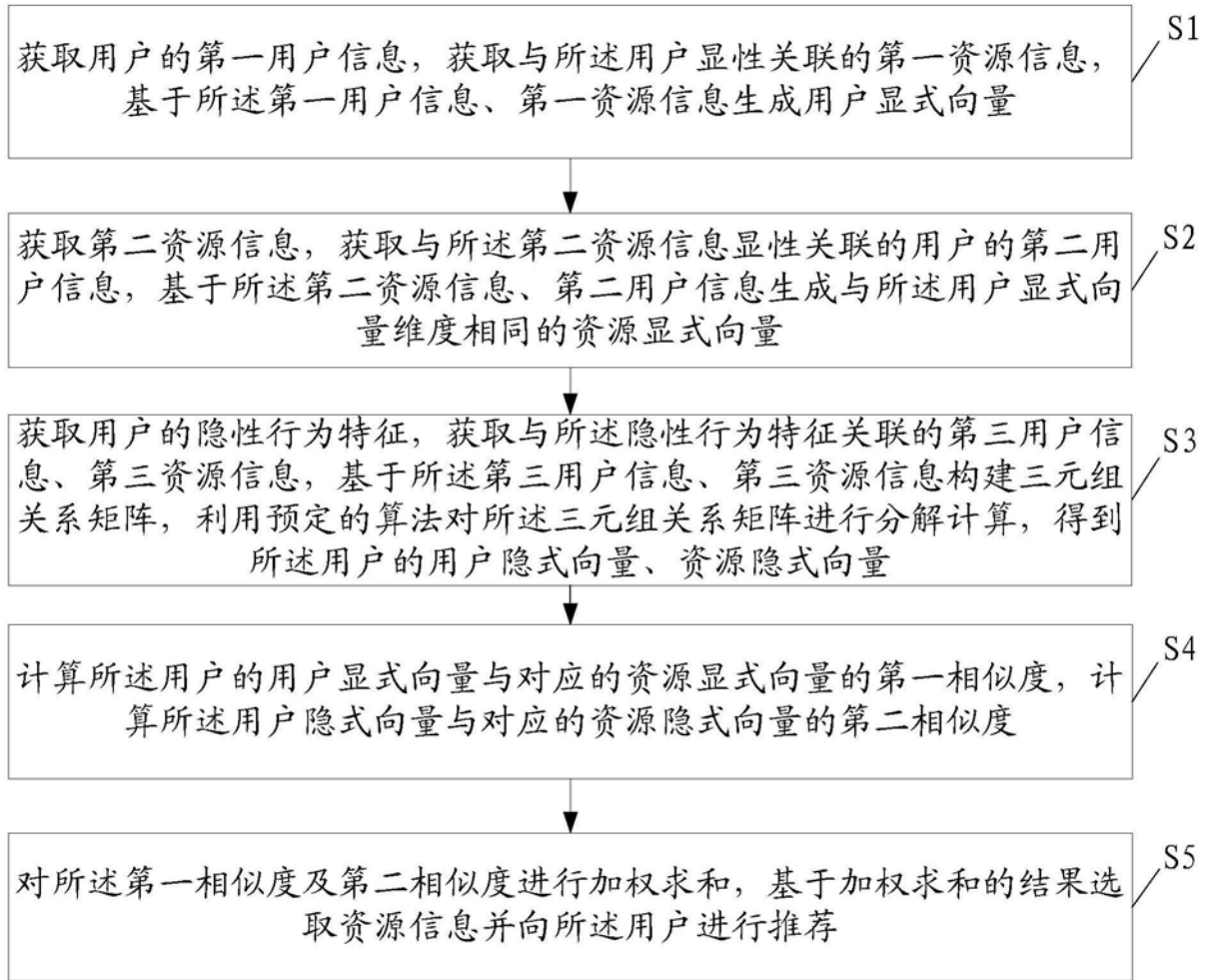


图3