



(12)发明专利

(10)授权公告号 CN 107590232 B

(45)授权公告日 2019.12.06

(21)申请号 201710799698.1

(22)申请日 2017.09.07

(65)同一申请的已公布的文献号
申请公布号 CN 107590232 A

(43)申请公布日 2018.01.16

(73)专利权人 北京师范大学
地址 100875 北京市海淀区新街口外大街
19号

(72)发明人 余胜泉 卢宇 庄梓君 李葆萍

(74)专利代理机构 北京科迪生专利代理有限责
任公司 11251

代理人 安丽

(51)Int.Cl.
G06F 16/9535(2019.01)
H04L 29/06(2006.01)

(56)对比文件

CN 104050258 A,2014.09.17,
CN 106708938 A,2017.05.24,
CN 104331459 A,2015.02.04,
CN 106815297 A,2017.06.09,
CN 102135999 A,2011.07.27,
US 2009288150 A1,2009.11.19,
程罡 等.《基于真实用户评论信息构建移动
学习资源的评价模型》.《现代远程教育》
.2014,(第1期),第43-53页.

审查员 侯鹏

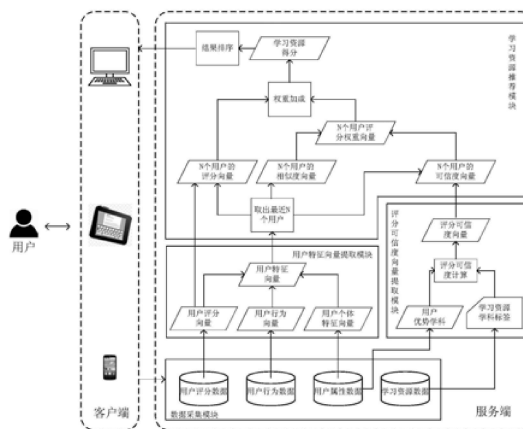
权利要求书3页 说明书6页 附图2页

(54)发明名称

一种基于网络学习环境的资源推荐系统及方法

(57)摘要

本发明涉及一种基于网络学习环境的资源推荐系统及方法,包括:数据采集模块、用户多维度特征向量和评分向量提取模块、评分可信度向量提取模块、学习资源推荐模块。本发明利用协同过滤的思想,筛选出与目标用户相似的用户群体,结合相似用户群与目标用户的相似度和用户评分的可信度进行学习资源的评分与推荐,使学习资源的评分具有用户针对性和评分客观性,从而为用户推荐个性化和高质量的学习资源。此发明可应用于在线网络学习平台,为用户提供个性化和可靠的学习资源推荐服务,从而提高用户对推荐的学习资源的满意度。



CN 107590232 B

1. 一种基于网络学习环境的资源推荐系统,其特征在於:所述网络学习环境是指线上的网络学习平台,所述资源是提供用户学习的各类资料,所述系统包括数据采集模块、用户多维度特征向量和评分向量提取模块、评分可信度向量提取模块、学习资源推荐模块,其中:

数据采集模块:采集的数据有两类,第一类是与用户相关的数据,其中包括用户属性数据、用户行为数据和用户评分数据;所述用户属性数据是指用以描述用户属性的信息,用户属性数据包括用户ID、姓名、密码、性别、年级、优势学科,所述用户行为数据是指用户对学习资源的交互数据,用户行为数据包括点赞、收藏、浏览时长、点击次数;所述用户评分数据是指用户对学习资源的评价分数;第二类是与学习资源相关的数据,包括学习资源和学习资源的学科标签,所述学习资源的学科标签是指学习资源所属的学科范畴;采集的用户属性数据、用户行为数据和用户评分数据提供给用户多维度特征向量和评分向量提取模块调用;采集的用户属性数据和学习资源数据提供给评分可信度向量提取模块调用;

用户多维度特征向量和评分向量提取模块:对于任意用户 i ,基于数据采集模块中的用户属性数据、用户行为数据和用户评分数据,形成用户个性化特征向量、用户行为向量和用户评分向量,进而综合形成用户多维度特征向量,从用户属性数据中选取性别、年级数据,将这些数据进行量化后形成用户个人特征向量 P_i ;用户行为数据中部分数据属于非数值型数据,包括点赞、收藏,需要先进行量化,而另外一些数据属于数值型数据,包括浏览时长、点击次数,无需量化,最终得到用户行为向量 B_i ;用户评分数据属于数值型数据,无需量化,直接形成用户评分向量 V_i ,将这三个特征向量综合形成用户的多维度向量 $T_i = [P_i, B_i, V_i]$,包括用户个人特征向量 P_i 、用户行为向量 B_i 和用户评分向量 V_i ,提取所有用户多维度特征向量和用户评分向量发送至学习资源推荐模块;

评分可信度向量提取模块:基于数据采集模块中得到的用户属性数据和学习资源数据分别筛选出用户的优势学科数据和学习资源的学科标签数据,利用用户的优势学科数据和学习资源的学科标签数据,赋予用户在学习资源上的评分可信度,所述评分可信度是指用户对学习资源的评分是否客观、可信,即评分是否能反映出学习资源的真实情况;如果一个用户的优势学科和某学习资源的学科标签是一致的,说明用户对该学习资源比较了解,则评分也较能真实反映学习资源的优劣情况,此时将该用户的评分可信度值设为系统预设的可信度值 c ,其中 $c > 1$;若用户的优势学科和学习资源学科标签是不一致的,该用户的评分可信度则设为默认值 1 ;基于所有用户在各个学习资源上的评分可信度,得到评分可信度向量 Φ ,将 Φ 发送至学习资源推荐模块;

学习资源推荐模块:基于用户多维度特征向量和评分向量提取模块得到用户多维度特征向量,利用目标用户 i 的多维度特征向量 T_i 和用户 j 的多维度特征向量 T_j ,计算目标用户 i 与用户 j 的欧氏距离 D_{ij} ;然后利用高斯函数作为反转函数将目标用户 i 与用户 j 之间的欧氏距离 D_{ij} 转变为目标用户 i 与用户 j 之间的相似度 S_{ij} ,通过计算得到目标用户 i 和其余所有用户的相似度,从中筛选出相似度最高的 N 个用户,根据这 N 个用户与目标用户 i 的相似度的大小降序排列,从而形成这 N 个用户的相似度向量;根据评分可信度向量提取模块中得到的评分可信度向量,筛选出这 N 个用户对某一学习资源的评分可信度,根据上述相似度大小排序的用户序列,将这 N 个用户的评分可信度按照该用户序列进行排列,从而得到这 N 个用户的可信度向量,所述的用户序列是指根据相似度大小降序排列的用户顺序;同理,根据用户多

维度特征向量和评分向量提取模块中得到的用户评分向量,筛选出这N个用户在该学习资源上的评分,根据上述的用户序列将N个用户的评分进行排列,从而得到这N个用户的评分向量;利用这N个用户的相似度向量和评分可信度向量,计算得到评分权重向量;最后,通过这N个用户的评分权重向量和其评分向量进行权重加成,得到该学习资源对于目标用户i的评分分数;对每个学习资源进行评分,获得每个学习资源对目标用户i的评分分数,然后根据分数的高低进行降序排列,最后将排好序的资源集推荐给目标用户i;

所述学习资源推荐模块中,利用高斯函数作为反转函数将目标用户i与用户j之间的欧氏距离 D_{ij} ,转变为目标用户i与用户j之间的相似度 S_{ij} ,具体实现如下:

$$S_{ij} = f(D_{ij}; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(D_{ij} - \mu)^2}{2\sigma^2}\right)$$

D_{ij} 为用户i与用户j的综合特征向量间的欧氏距离, σ 和 μ 为决定高斯函数形状的常数;

所述学习资源推荐模块中,根据N个用户的相似度向量和评分可信度向量,计算得到评分权重向量,具体实现如下:

分别取出这N个用户的相似度和评分可信度,其中用户j的评分权重为该用户与目标用户i的相似度 S_{ij} 和该用户在某学习资源上的评分可信度 θ_j 的乘积,即:

$$W_j = S_{ij} * \theta_j$$

W_j 为用户j在此资源上的评分权重,基于这N个用户每个用户的评分权重,根据这N个用户的用户序列进行排序,从而得到评分权重向量。

2.一种基于网络学习环境的资源推荐方法,其特征在于:步骤如下:

步骤(1),数据采集,通过用户在学校的个人信息和学业信息,管理员利用这些信息导入用户ID、姓名、密码、性别、年级、优势学科,形成用户属性数据;用户每次登录使用平台时,记录用户对学习资源的交互行为,包括点赞、收藏、浏览时长、点击次数,形成用户行为数据;另外收集用户对学习资源的评分数据,形成用户评分数据;通过管理员用户上传的学习资源和标记的学习资源学科标签数据,形成学习资源数据,其中,用户属性数据、用户行为数据和用户评分数据供给用户多维度特征向量和评分向量提取模块调用;用户属性数据和学习资源数据供给评分可信度提取模块使用;

步骤(2),基于数据采集模块中的用户属性数据、用户行为数据和用户评分数据形成的用户个性化特征向量、用户行为向量和用户评分向量综合形成用户多维度特征向量,从用户属性数据中选取性别、年级数据,将这些数据进行量化后形成用户个人特征向量;用户行为数据中部分数据属于非数值型数据,包括点赞、收藏,需要先进行量化,而另外一些数据属于数值型数据,包括浏览时长、点击次数,无需量化,最终得到用户行为向量;用户评分数据属于数值型数据,无需量化,直接形成用户评分向量,通过得到的三个向量综合形成用户多维度特征向量,提取用户多维度特征向量和用户评分向量发送至学习资源推荐模块;

步骤(3),判断用户对学习资源的评分是否可信,得到评分可信度,基于各用户在各个学习资源上的评分可信度,形成评分可信度向量;基于步骤(1)得到的用户属性数据和学习资源数据分别筛选出用户的优势学科数据和学习资源的学科标签数据,利用用户的优势学科数据和学习资源的学科标签数据,赋予用户在学习资源上的评分可信度,所述评分可信度是指用户对学习资源的评分是否客观、可信,即评分是否能反映出学习资源的真实情况,

如果用户的优势学科和某学习资源的学科标签是一致的,则用户的评分可信度值设为预设的可信度值 c ,其中 $c>1$;若用户的优势学科和学习资源学科标签是不一致的,则用户的评分可信度则设为默认值1;基于各用户在各个学习资源上的评分可信度,得到评分可信度向量,将其发送至学习资源推荐模块;

步骤(4),基于步骤(2)得到的用户多维度特征向量,利用目标用户的多维度特征向量和其余用户的多维度特征向量,计算目标用户与其余用户的欧氏距离;然后利用高斯函数作为反转函数将目标用户与其余用户之间的欧氏距离转变为目标用户与其余用户之间的相似度;然后从中筛选出相似度最高的 N 个用户,根据这 N 个用户与目标用户的相似度的大小降序排列,得到这 N 个用户的相似度向量;根据步骤(3)得到的评分可信度向量,筛选出这 N 个用户对某一学习资源的评分可信度,根据相似度大小排序的用户序列,将这 N 个用户的评分可信度按照该用户序列进行排列,从而得到这 N 个用户的可信度向量,所述的用户序列是指根据相似度大小降序排列后的用户顺序;同理,根据步骤(2)得到的用户评分向量,筛选出这 N 个用户在该学习资源上的评分,根据用户序列将 N 个用户的评分进行排列,从而得到这 N 个用户的评分向量;利用这 N 个用户的相似度向量和评分可信度向量,计算得到评分权重向量;最后,通过这 N 个用户的评分权重向量和其评分向量进行权重加成,得到该学习资源对于目标用户的评分分数;对每个学习资源进行评分,获得每个学习资源对目标用户的评分分数,然后根据分数的高低进行降序排列,最后将排好序的资源集推荐给目标用户。

一种基于网络学习环境的资源推荐系统及方法

技术领域

[0001] 本发明涉及一种基于网络学习环境的资源推荐系统及方法,属于推荐系统领域,应用于网络学习环境中。

背景技术

[0002] 在信息过载的时代,推荐系统能够帮助用户找到符合其个性化需求的内容,因而在电子商务、社交网站中得到了广泛的应用。现有的推荐系统大部分是借助与目标用户具有相似特征、共同经验的其他用户的信息数据来为该用户推荐其感兴趣的内容。其中相似特征中的特征通常指的是用户的个人特征,共同经验中的经验通常指的是用户对项目的评分或行为。在学习资源推荐系统中,用户并不单单只想获得感兴趣的学习资源,他们同时渴望获得高质量的学习资源,以助于提高学习的效果。然而,由于部分用户(特别是非内行的用户)对资源的评分具有较大的主观性,其评分不能真实、客观的反映学习资源,如果只根据用户间的相似特征、共同经验数据来推荐学习资源,而不考虑学习资源的可靠性,就会导致得到的推荐集并不可靠。

[0003] 在此背景下,本发明首先基于协同过滤算法思想,找到与目标用户具有高度相似特征与经验的其他用户的评分数据,然后结合用户的评分可信度,进行学习资源的评分,最终为用户推荐具有针对性和可靠性的学习资源。

发明内容

[0004] 本发明要解决的问题是:克服现有推荐结果集质量不高的问题,提供一种基于网络学习环境的资源推荐系统及方法,达到为用户推荐具有针对性和高质量的学习资源的目的。

[0005] 本发明解决其问题所采用的方案是:一种基于网络学习环境的资源推荐系统,包括数据采集模块、用户多维度特征向量和评分向量提取模块、评分可信度向量提取模块、学习资源推荐模块,其中:

[0006] 数据采集模块:采集的数据有两类,第一类是与用户相关的数据,其中包括用户属性数据、用户行为数据和用户评分数据;所述用户属性数据是指用以描述用户属性的信息,用户属性数据包括用户ID、姓名、密码、性别、年级、优势学科,所述用户行为数据是指用户对学习资源的交互数据,用户行为数据包括点赞、收藏、浏览时长、点击次数;所述用户评分数据是指用户对学习资源的评价分数;第二类是与学习资源相关的数据,包括学习资源和学习资源的学科标签,所述学习资源的学科标签是指学习资源所属的学科范畴;采集的用户属性数据、用户行为数据和用户评分数据提供给用户多维度特征向量和评分向量提取模块调用;采集的用户属性数据和学习资源数据提供给评分可信度向量提取模块调用;

[0007] 用户多维度特征向量和评分向量提取模块:对于任意用户*i*,基于数据采集模块中的用户属性数据、用户行为数据和用户评分数据,形成用户个性化特征向量、用户行为向量和用户评分向量,进而综合形成用户多维度特征向量。从用户属性数据中选取性别、年级数

据,将这些数据进行量化后形成用户个人特征向量 P_i ;用户行为数据中部分数据属于非数值型数据,包括点赞、收藏,需要先进行量化,而另外一些数据属于数值型数据,包括浏览时长、点击次数,无需量化,最终得到用户行为向量 B_i ;用户评分数据属于数值型数据,无需量化,直接形成用户评分向量 V_i ,将这三个特征向量综合形成用户的多维度向量 $T_i = [P_i, B_i, V_i]$,包括用户个人特征向量 P_i 、用户行为向量 B_i 和用户评分向量 V_i 。提取所有用户多维度特征向量和用户评分向量发送至学习资源推荐模块;

[0008] 评分可信度向量提取模块:基于数据采集模块中得到的用户属性数据和学习资源数据分别筛选出用户的优势学科数据和学习资源的学科标签数据,利用用户的优势学科数据和学习资源的学科标签数据,赋予用户在学习资源上的评分可信度,所述评分可信度是指用户对学习资源的评分是否客观、可信,即评分是否能反映出学习资源的真实情况;如果一个用户的优势学科和某学习资源的学科标签是一致的,说明用户对该学习资源比较了解,则评分也较能真实反映学习资源的优劣情况,此时将该用户的评分可信度值设为系统预设的可信度值 c ,其中 $c > 1$;若用户的优势学科和学习资源学科标签是不一致的,该用户的评分可信度则设为默认值1;基于所有用户在各个学习资源上的评分可信度,得到评分可信度向量 θ ,将 θ 发送至学习资源推荐模块;

[0009] 学习资源推荐模块:基于用户多维度特征向量和评分向量提取模块得到用户多维度特征向量,利用目标用户 i 的多维度特征向量 T_i 和用户 j 的多维度特征向量 T_j ,计算目标用户 i 与用户 j 的欧氏距离 D_{ij} ;然后利用高斯函数作为反转函数将目标用户 i 与用户 j 之间的欧氏距离 D_{ij} 转变为目标用户 i 与用户 j 之间的相似度 S_{ij} ;通过计算得到目标用户 i 和其余所有用户的相似度,从中筛选出相似度最高的 N 个用户,根据这 N 个用户与目标用户 i 的相似度的大小降序排列,从而形成这 N 个用户的相似度向量;根据评分可信度向量提取模块中得到的评分可信度向量,筛选出这 N 个用户对某一学习资源的评分可信度,根据上述相似度大小排序的用户序列,将这 N 个用户的评分可信度按照该用户序列进行排列,从而得到这 N 个用户的可信度向量,所述的用户序列是指根据相似度大小降序排列的用户顺序;同理,根据用户多维度特征向量和评分向量提取模块中得到的用户评分向量,筛选出这 N 个用户在该学习资源上的评分,根据上述的用户序列将 N 个用户的评分进行排列,从而得到这 N 个用户的评分向量;利用这 N 个用户的相似度向量和评分可信度向量,计算得到评分权重向量;最后,通过这 N 个用户的评分权重向量和其评分向量进行权重加成,得到该学习资源对于目标用户 i 的评分分数;对每个学习资源进行评分,获得每个学习资源对目标用户 i 的评分分数,然后根据分数的高低进行降序排列,最后将排好序的资源集推荐给目标用户 i 。

[0010] 所述学习资源推荐模块中,利用高斯函数作为反转函数将目标用户 i 与用户 j 之间的欧氏距离 D_{ij} ,转变为目标用户 i 与用户 j 之间的相似度 S_{ij} ,具体实现如下:

$$[0011] \quad S_{ij} = f(D_{ij}; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(D_{ij} - \mu)^2}{2\sigma^2}\right)$$

[0012] D_{ij} 为用户 i 与用户 j 的综合特征向量间的欧氏距离, σ 和 μ 为决定高斯函数形状的常数。

[0013] 所述学习资源推荐模块中,根据 N 个用户的相似度向量和评分可信度向量,计算得到评分权重向量,具体实现如下:

[0014] 分别取出这N个用户的相似度和评分可信度,其中用户j的评分权重为该用户与目标用户i的相似度 S_{ij} 和该用户在某学习资源上的评分可信度 θ_j 的乘积,即:

$$[0015] \quad W_j = S_{ij} * \theta_j$$

[0016] W_j 为用户j在此资源上的评分权重。基于这N个用户每个用户的评分权重,根据这N个用户的用户序列进行排序,从而得到评分权重向量。

[0017] 一种基于网络学习环境的资源推荐方法,步骤如下:

[0018] 步骤(1),数据采集,通过用户在学校的信息和学业信息,管理员利用这些信息导入用户ID、姓名、密码、性别、年级、优势学科,形成用户属性数据;用户每次登录使用平台时,记录用户对学习资源的交互行为,包括点赞、收藏、浏览时长、点击次数,形成用户行为数据;另外收集用户对学习资源的评分数据,形成用户评分数据;通过管理员用户上传的学习资源和标记的学习资源学科标签数据,形成学习资源数据,其中,用户属性数据、用户行为数据和用户评分数据供给用户多维度特征向量和评分向量提取模块调用;用户属性数据和学习资源数据供给评分可信度提取模块使用;

[0019] 步骤(2),基于数据采集模块中的用户属性数据、用户行为数据和用户评分数据形成的用户个性化特征向量、用户行为向量和用户评分向量综合形成用户多维度特征向量。从用户属性数据中选取性别、年级数据,将这些数据进行量化后形成用户个人特征向量;用户行为数据中部分数据属于非数值型数据,包括点赞、收藏,需要先进行量化,而另外一些数据属于数值型数据,包括浏览时长、点击次数,无需量化,最终得到用户行为向量;用户评分数据属于数值型数据,无需量化,直接形成用户评分向量,通过得到的三个向量综合形成用户多维度特征向量。提取用户多维度特征向量和用户评分向量发送至学习资源推荐模块;

[0020] 步骤(3),判断用户对学习资源的评分是否可信,得到评分可信度,基于各用户在各个学习资源上的评分可信度,形成评分可信度向量;基于步骤(1)得到的用户属性数据和学习资源数据分别筛选出用户的优势学科数据和学习资源的学科标签数据,利用用户的优势学科数据和学习资源的学科标签数据,赋予用户在学习资源上的评分可信度,所述评分可信度是指用户对学习资源的评分是否客观、可信,即评分是否能反映出学习资源的真实情况。如果用户的优势学科和某学习资源的学科标签是一致的,那么用户的评分可信度值设为系统预设的可信度值 c ,其中 $c > 1$;若用户的优势学科和学习资源学科标签是不一致的,那么用户的评分可信度则设为默认值1;基于各用户在各个学习资源上的评分可信度,得到评分可信度向量,将其发送至学习资源推荐模块;

[0021] 步骤(4),基于步骤(2)得到的用户多维度特征向量,利用目标用户的多维度特征向量和其余用户的多维度特征向量,计算目标用户与其余用户的欧氏距离;然后利用高斯函数作为反转函数将目标用户与其余用户之间的欧氏距离转变为目标用户与其余用户之间的相似度;然后从中筛选出相似度最高的N个用户,根据这N个用户与目标用户的相似度的大小降序排列,得到这N个用户的相似度向量;根据步骤(3)得到的评分可信度向量,筛选出这N个用户对某一学习资源的评分可信度,根据相似度大小排序的用户序列,将这N个用户的评分可信度按照该用户序列进行排列,从而得到这N个用户的可信度向量,所述的用户序列是指根据相似度大小降序排列后的用户顺序;同理,根据步骤(2)得到的用户评分向量,筛选出这N个用户在该学习资源上的评分,根据用户序列将N个用户的评分进行排列,从

而得到这N个用户的评分向量;利用这N个用户的相似度向量和评分可信度向量,计算得到评分权重向量;最后,通过这N个用户的评分权重向量和其评分向量进行权重加成,得到该学习资源对于目标用户的评分分数;对每个学习资源进行评分,获得每个学习资源对目标用户的评分分数,然后根据分数的高低进行降序排列,最后将排好序的资源集推荐给目标用户。

[0022] 本发明与现有方法相比的有益效果为:

[0023] (1) 本发明使推荐的学习资源在符合用户个性化需求的同时,也提高了推荐的学习资源的质量,从而提高学习资源对用户的有效性,最终达到提高用户学习效果的目的。

[0024] (2) 本发明利用大规模数据集形成用户多维度特征向量和用户评分可信度向量,结合用户相似度和评分可信度建立评分机制,增加了评分所考虑的数据维度与信息量,同时在算法与工程上易于实现与拓展。

附图说明

[0025] 图1为本发明一种基于网络学习环境的资源推荐系统的结构图;

[0026] 图2为本发明中的用户行为数据储存结构;

[0027] 图3为本发明中基于用户特征计算相似度流程;

[0028] 图4为本发明中学习资源的评分计算流程。

具体实施方式

[0029] 下面结合附图及具体实施方式详细介绍本发明。

[0030] 如图1所示,本发明为一种基于网络学习环境的资源推荐系统,包括数据采集模块、用户多维度特征向量和评分向量提取模块、评分可信度向量提取模块、学习资源推荐模块。

[0031] 数据采集模块具体实现如下:

[0032] (1) 通过用户在学校的个人信息和学业信息,管理员利用这些信息导入用户ID、姓名、密码(密码是平台的初始密码,学生使用初始密码登录后可以进行修改)、性别、年级、优势学科,存储在用户属性数据库;

[0033] (2) 用户每次登录使用平台时,记录用户对学习资源的交互行为,包括点赞、收藏、浏览时长、点击次数,将这些数据存储在用户行为数据库;另外收集用户对学习资源的评分数据,存储在用户评分数据库;

[0034] (3) 管理员用户登录并上传学习资源和标记学习资源的学科标签时,将这些数据存储在学习资源数据库。

[0035] 用户多维度特征向量和评分向量提取模块具体实现如下:

[0036] (1) 提取用户相关数据形成用户特征向量。从用户属性数据中提取性别、年级数据,将这些数据量化后形成用户个人特征向量。例如用户i的年级是初二,其性别是女,那么用户i的个人特征向量 $P_i = [8, 1]$,其中年级编码和性别编码由系统预先设定;用户行为数据中的点赞、收藏等非数值型数据转换为数值型,而浏览时长、点击次数等数值型数据就无需量化,最终形成一个矩阵向量如图2所示,每一行对应的是该用户对每个学习资源的行为,而列分别对应的是该用户对具体某个学习资源的各个行为:是否点赞,是否收藏,浏览

时长和点击次数。例如用户*i*对学习资源A有收藏且有点赞,浏览时长为40分钟,点击次数为6,而对学习资源B有收藏但没点赞,浏览时长为8分钟,点击次数为2,那么该用户的行为向量 $T_i = \begin{bmatrix} 1 & 1 & 40 & 6 \\ 0 & 1 & 8 & 2 \end{bmatrix}$;用户评分数据无需量化,例如某用户对学习资源A的评分是4分,对学习资源B的评分是2分,那么该用户的评分向量是 $V_i = [4, 2]$ 。

[0037] (2) 将用户各个特征向量形成多维度特征向量 $T_i = [P_i, B_i, G_i]$ 用以表征一个用户,其中多维度向量包括用户个体特征向量 P_i ,用户行为向量 B_i 和用户评分向量 V_i 。

[0038] (3) 提取所有用户的多维度特征向量和评分向量发送至学习资源推荐模块。

[0039] 评分可信度向量提取模块具体实现如下:

[0040] 如果用户的优势学科和学习资源学科标签是一致的,如用户*i*的优势学科为英语,该学习资源的学科标签也为英语,那么该用户的对该学习资源的可信度为系统预设的可信度值 c ,其中 $c > 1$;若用户的优势学科和学习资源学科标签是不一致的,如用户*i*的优势学科为数学,该学习资源的学科标签为英语,那么该用户的评分可信度则设为系统默认值1。基于所有用户的评分可信度数值,将其作为元素得到评分可信度向量;

[0041] 学习资源推荐模块,其具体实现如下:

[0042] (1) 通过多维度特征向量得到相似度向量,实现流程如图3所示,具体而言,基于用户多维度特征向量和评分向量提取模块产生的目标用户*i*的多维度特征 T_i ,计算 T_i 与系统中具有相似个体属性的其他用户的多维度特征 T_j 的欧式距离 D_{ij} 。为了进一步得到相似度的数值,需要利用反转函数将目标用户*i*与用户*j*之间的欧氏距离 D_{ij} 转化为相似度 S_{ij} 。此模块中使用高斯函数作为反转函数,将目标用户*i*与用户*j*之间的欧氏距离,转变为目标用户*i*与用户*j*之间的相似度。高斯函数的数学表达如下:

$$[0043] \quad S_{ij} = f(D_{ij}; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(D_{ij} - \mu)^2}{2\sigma^2}\right)$$

[0044] 其中的 S_{ij} 为目标用户*i*与用户*j*之间的相似度, D_{ij} 为目标用户*i*与用户*j*的综合特征向量间的欧氏距离, σ 和 μ 为决定高斯函数形状的常数。通常情况下,可以取 $\mu = 0$; $\sigma = 1$ 。

[0045] 计算目标用户和其余所有用户间的相似度,将其作为元素得到相似度向量。

[0046] (2) 获得最近*N*个用户的相似度向量和评分可信度向量和评分向量。根据计算得到的目标用户*i*与其余用户的相似度,筛选出与目标用户*i*相似度最高的前*N*个用户,根据这*N*个用户的相似度大小进行降序排列,从而形成这*N*个用户的相似度向量。根据评分可信度向量提取模块中得到的评分可信度向量,筛选出这*N*个用户对某一学习资源的评分可信度,根据上述按相似度大小排序的用户序列,将这*N*个用户的评分可信度按照该用户序列进行排列,从而得到这*N*个用户的可信度向量,所述的用户序列是指根据相似度大小降序排列的用户顺序;同理,根据用户多维度特征向量和评分向量提取模块中得到的用户评分向量,筛选出这*N*个用户在该学习资源上的评分,根据上述的用户序列将*N*个用户的评分进行排列,从而得到这*N*个用户的评分向量。

[0047] (3) 利用*N*个用户的相似度向量、评分可信度向量和评分向量进行学习资源评分计算,计算流程如图4所示。首先利用筛选出来的*N*个用户的相似度向量和评分可信度向量计算得到评分权重向量,具体过程为:分别取出这*N*个用户的相似度和评分可信度,计算每个

用户的评分权重(每个用户的评分权重为其相似度与评分可信度的乘积),根据这N个用户的相似度大小排列的用户序列,对其评分权重按该用户序列排序,从而得到评分权重向量。例如对于目标用户i,系统根据其他用户与目标用户i的相似度找到了前5个与目标用户相似度最高的用户(即N=5),其相似度向量是[1,0.99,0.99,0.84,0.82];且这5个用户在此学习资源上的评分可信度向量为[1,1,1,1.5,1.5],那么这5个用户的评分权重分别为1,0.99,0.99,1.26,1.23,得到评分权重向量是[1,0.99,0.99,1.26,1.23]。然后将计算得到的评分权重向量和这N个用户的评分向量进行权重加成,获得学习资源对于用户i的评分 \hat{V}_i ,具体的计算方法如下:

$$[0048] \quad \hat{V}_i = \frac{\sum_{j \in N} W_j V_j}{\sum_{j \in N} W_j};$$

[0049] 例如根据上述得到的5个用户的评分权重向量后,即[1,0.99,0.99,1.26,1.23],利用其评分权重向量和评分向量,假设评分向量为[5,5,5,4,3],最终得到该学习资源对于用户i的评分 $\hat{V} = 4.32$ 。

[0050] (4)对每个学习资源进行评分,获得每个学习资源对目标用户i的评分分数,然后根据分数的高低进行降序排列,最后将排好序的资源集推荐给目标用户i。

[0051] 本发明未详细阐述的部分属于本领域公知技术。

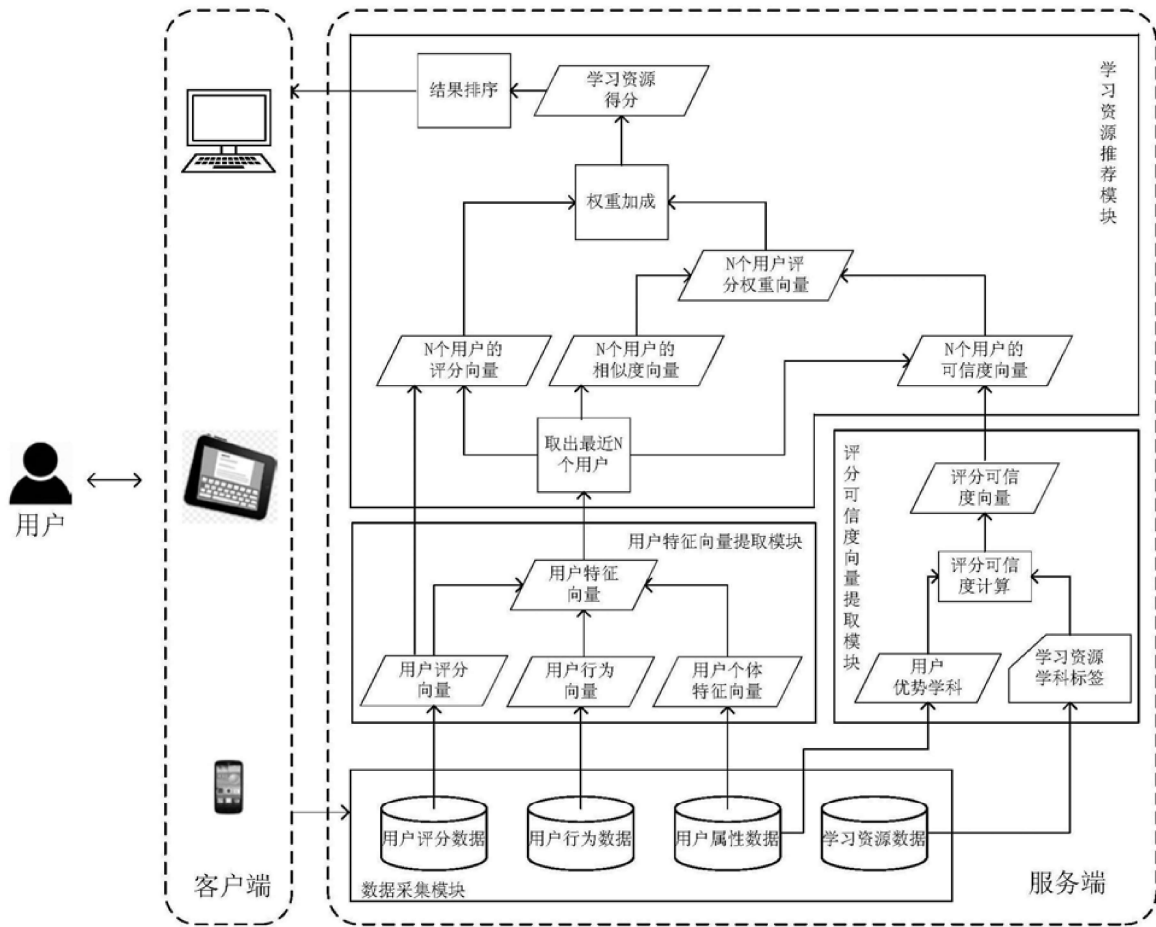


图1

行为 \ 学习资源	收藏	点赞	浏览时长	点击次数
学习资源1	a_1	b_1	c_1	d_1
学习资源2	a_2	b_2	c_2	d_2
...
学习资源n	a_n	b_n	c_n	d_n

图2

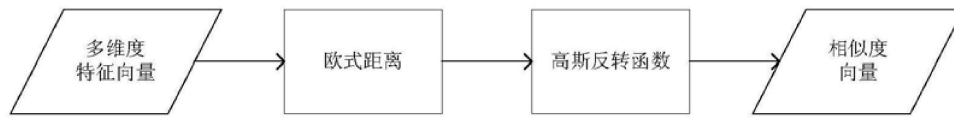


图3

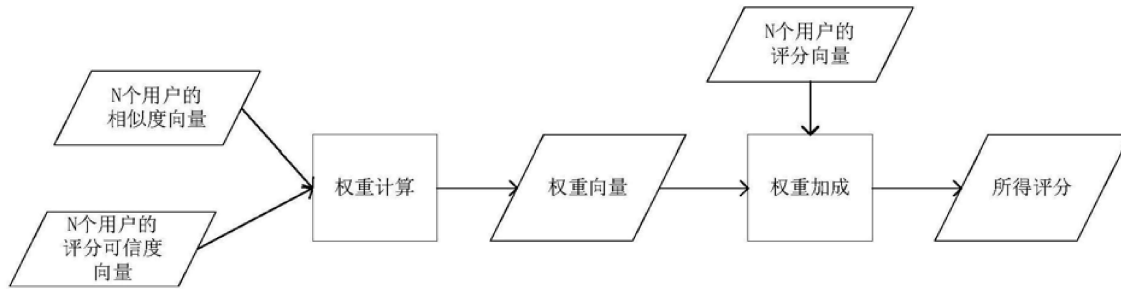


图4