



(12)发明专利申请

(10)申请公布号 CN 106952130 A

(43)申请公布日 2017.07.14

(21)申请号 201710110168.1

(22)申请日 2017.02.27

(71)申请人 华南理工大学

地址 510640 广东省广州市天河区五山路  
381号

(72)发明人 周智恒 劳志辉 俞政 黄俊楚  
代雨琨 李立军

(74)专利代理机构 广州市华学知识产权代理有  
限公司 44245

代理人 罗观祥

(51)Int.Cl.

G06Q 30/06(2012.01)

G06F 17/30(2006.01)

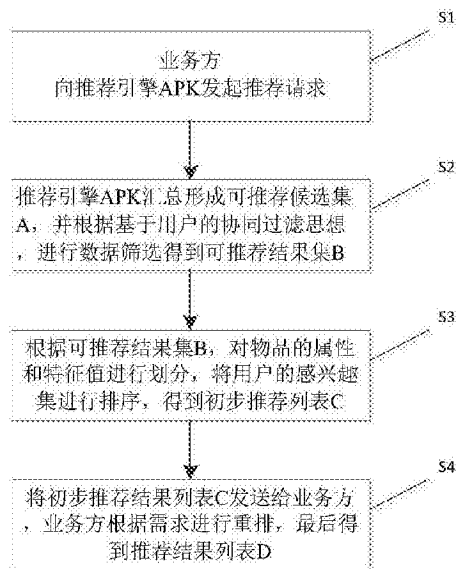
权利要求书2页 说明书5页 附图1页

(54)发明名称

基于协同过滤的通用物品推荐方法

(57)摘要

本发明公开了一种基于协同过滤的通用物品推荐方法,该方法利用基于用户的协同过滤方法,通过欧几里得距离计算用户之间的相似度,得到相似用户集并进一步得出不同用户的候选推荐集;之后再通过对用户初始感兴趣物品属性的分类和特征赋值,计算出候选推荐集的推荐分数,从而得到一种适用性和鲁棒性较强的推荐结果,帮助用户更方便的获取感兴趣的内容。相比于传统的基于人口统计学和基于内容的推荐方法相比,本发明更为注重了用户之间的个体差异性,通过分析用户的历史行为数据,挖掘用户的兴趣进行推荐;因此,推荐结果因人而异,更多的考虑了用户的个人喜好,也可以针对不同的推荐场景定制不同的方法参数。



1. 一种基于协同过滤的通用物品推荐方法,其特征在于,所述通用物品推荐方法包括下列步骤:

S1、业务方对推荐物品及用户的属性及特征值进行初始化定义,将用户的行为分类并设置初始搜索引擎的条件,完成初始化设定和推荐引擎APK接入后,向推荐引擎APK发起推荐请求,并将初始化的数据集发送给推荐引擎APK;

S2、推荐引擎APK首先根据业务方的初始搜索引擎,将满足搜索条件的数据,汇总形成可推荐候选集A,对可推荐候选集A通过相似度判定得到相似用户集,并根据基于用户的协同过滤思想,进行数据筛选得到可推荐结果集B;

S3、将可推荐结果集B根据物品的属性,对属性的特征值进行划分,根据用户的感兴趣集中各个特征值所占的比例,得到可推荐结果集B的用户中每个属性特征值所占的权重,并根据各个最显著的特征集去得到用户感知最敏感的属性,根据感知最敏感的属性的不同参考权重,将用户的感兴趣集进行排序,得到初步推荐列表C;

S4、将初步推荐结果列表C发送给业务方,业务方根据需求进行重排,最后得到推荐结果列表D。

2. 根据权利要求1所述的基于协同过滤的通用物品推荐方法,其特征在于,所述步骤S2的具体过程如下:

S201、将用户的行为分为 $T_1 \sim T_K$ 共K类,并对这K类行为分别进行权重赋值 $w_1 \sim w_K$ ,根据不同的用户行为区分为正面、负面以及高、中、低六个维度,赋值向量 $w$ 的取值为 $w = [-2, -1, 0, 1, 2, 3]$ ;

S202、获取用户对物品的行为操作累加值得到用户对物品的喜好度 $H = \sum w$ ;

S203、通过不同用户对各个物品的喜好度 $H$ ,利用欧几里得距离 $d(x, y) = \sqrt{\sum H_x - \sum H_y}$ ,计算得到用户之间的相似度: $sim(x, y) = \frac{1}{1 + d(x, y)}$ ,

当两个用户之间的相似度 $sim(x, y) > k$ 时,其中 $k$ 由业务方决定,即认为两者相似从而得到相似用户集,并根据基于用户的协同过滤思想,得到各个用户的可推荐结果集B。

3. 根据权利要求1所述的基于协同过滤的通用物品推荐方法,其特征在于,所述步骤S3的具体过程如下:

S301、根据可推荐结果集B,对物品的属性和特征值进行划分,设物品的属性向量为:

$$S = [S_1, S_2, \dots, S_w],$$

属性 $S_i$ 的特征值向量为:

$$V_i = [v_1, v_2, \dots, v_i];$$

S302、通过各个属性的区分度向量和候选推荐集构造属性特征矩阵;

对用户A而言,感兴趣物品集 $Q_A$ 中某一物品属性 $S_i$ 的特征值 $v_k$ 在 $V_i$ 所占的比例为 $q_{v_k}^i$ ,则在候选推荐集 $Q_T$ 中,设该物品属性的特征值 $v_k$ 所占的权重:

$$q = q_{v_k}^i$$

另,当 $q_{v_k}^i > q_{v_k}^j$  ( $i, k$ 取任意可取值),则认为属性 $S_x$ 的区分度最强;各个属性的区分度取:

$$P_w = \max q_v^w$$

所以,当给出 $Q_A$ 后,可得索性区分度向量:

$$V = [p_1, p_2, \dots, p_w]$$

得到 $Q_T$ 后,可得物品的属性特征矩阵:

$$A_{Q_T} = \begin{bmatrix} x_1^1, x_1^2, \dots, x_1^r \\ x_2^1, x_2^2, \dots, x_2^r \\ \dots \\ x_w^1, x_w^2, \dots, x_w^r \end{bmatrix};$$

S303、根据属性特征矩阵和属性区分度向量,可以得到候选推荐集 $Q_T$ 的各物品推荐分数向量:

$$F = V \times A_{Q_T};$$

S304、根据得到的推荐分数向量对可推荐结果集B的排序,确定初步推荐结果列表C。

4. 根据权利要求1所述的基于协同过滤的通用物品推荐方法,其特征在于,

所述步骤S2中可推荐候选集A的人数如果不满足最低推荐人数要求,则向业务方请求扩大搜索条件。

5. 根据权利要求1所述的基于协同过滤的通用物品推荐方法,其特征在于,所述用户的属性及特征值的定义的方式为【属性-值】键值对。

6. 根据权利要求1所述的基于协同过滤的通用物品推荐方法,其特征在于,用户的初始属性由用户在注册时填写,推荐引擎可以根据用户填写的属性值,为初始用户进行特征划分。

7. 根据权利要求2所述的基于协同过滤的通用物品推荐方法,其特征在于,当用户对物品的喜好度 $H > 3$ 时,则认为用户对该物品感兴趣。

8. 根据权利要求1至7任一所述的基于协同过滤的通用物品推荐方法,其特征在于,

所述通用物品包括:书籍、音乐、电影,不同的通用物品的区别在于,喜好度的衡量标准以及物品的属性特征差异可根据不同的推荐场景由业务方调整参数。

## 基于协同过滤的通用物品推荐方法

### 技术领域

[0001] 本发明涉及信息推荐技术领域,具体涉及一种基于协同过滤的通用物品推荐方法。

### 背景技术

[0002] 如今已经进入了一个数据爆炸的时代,随着Web 2.0的发展,Web已经变成数据分享的平台,那么,如何让人们在海量的数据中想要找到他们需要的信息将变得越来越难。

[0003] 在这样的情形下,搜索引擎(Google,Bing,百度等等)为大家快速找到目标信息的最好途径。在用户对自己需求相对明确的时候,用搜索引擎很方便的通过关键字搜索很快的找到自己需要的信息。但搜索引擎并不能完全满足用户对信息发现的需求,那是因为很多情况下,用户其实并不明确自己的需要,或者他们的需求很难用简单的关键字来表述。又或者他们需要更加符合他们个人口味和喜好的结果,因此出现了推荐系统,与搜索引擎对应,大家也习惯称它为推荐引擎。

[0004] 随着推荐引擎的出现,用户获取信息的方式从简单的目标明确的数据的搜索转换到更高级更符合人们使用习惯的信息发现。

[0005] 如今,随着推荐技术的不断发展,推荐引擎已经在电子商务(E-commerce,例如Amazon,当当网)和一些基于social的社会化站点(包括音乐,电影和图书分享,例如豆瓣,Mtime等)都取得很大的成功。这也进一步的说明了,Web2.0环境下,在面对海量的数据,用户需要这种更加智能的,更加了解他们需求,口味和喜好的信息发现机制。协同过滤推荐算法是诞生最早,并且较为著名的推荐算法。主要的功能是预测和推荐。算法通过对用户历史行为数据的挖掘发现用户的偏好,基于不同的偏好对用户进行群组划分并推荐品味相似的商品。协同过滤推荐算法分为两类,分别是基于用户的协同过滤算法(user-based collaborative filtering),和基于物品的协同过滤算法(item-based collaborative filtering)。简单的说就是:人以类聚,物以群分。

### 发明内容

[0006] 本发明的目的是为了解决现有技术中的上述缺陷,提供一种基于协同过滤的通用物品推荐方法。

[0007] 本发明的目的可以通过采取如下技术方案达到:

[0008] 一种基于协同过滤的通用物品推荐方法,所述通用物品推荐方法包括下列步骤:

[0009] S1、业务方对推荐物品及用户的属性及特征值进行初始化定义,将用户的行为分类并设置初始搜索引擎的条件,完成初始化设定和推荐引擎APK接入后,向推荐引擎APK发起推荐请求,并将初始化的数据集发送给推荐引擎APK;

[0010] S2、推荐引擎APK首先根据业务方的初始搜索引擎,将满足搜索条件的数据,汇总形成可推荐候选集A,对可推荐候选集A通过相似度判定得到相似用户集,并根据基于用户的协同过滤思想,进行数据筛选得到可推荐结果集B;

[0011] S3、将可推荐结果集B根据物品的属性,对属性的特征值进行划分,根据用户的感兴趣集中各个特征值所占的比例,得到可推荐结果集B的用户中每个属性特征值所占的权重,并根据各个最显著的特征集去得到用户感知最敏感的属性,根据感知最敏感的属性的不同参考权重,将用户的感兴趣集进行排序,得到初步推荐列表C;

[0012] S4、将初步推荐结果列表C发送给业务方,业务方根据需求进行重排,最后得到推荐结果列表D。

[0013] 进一步地,所述步骤S2的具体过程如下:

[0014] S201、将用户的行为分为 $T_1 \sim T_k$ 共K类,并对这K类行为分别进行权重赋值 $w_1 \sim w_k$ ,根据不同的用户行为区分为正面、负面以及高、中、低六个维度,赋值向量 $w$ 的取值为 $w = [-2, -1, 0, 1, 2, 3]$ ;

[0015] S202、获取用户对物品的行为操作累加值得到用户对物品的喜好度 $H = \sum w$ ;当 $H > 3$ 时,则认为用户对该物品感兴趣;

[0016] S203、通过不同用户对各个物品的喜好度 $H$ ,利用欧几里得距离

$$d(x, y) = \sqrt{\sum H_x - \sum H_y}, \text{ 计算得到用户之间的相似度: } sim(x, y) = \frac{1}{1 + d(x, y)},$$

[0017] 当两个用户之间的相似度 $sim(x, y) > k$ 时,其中 $k$ 由业务方决定,即认为两者相似从而得到相似用户集,并根据基于用户的协同过滤思想,得到各个用户的可推荐结果集B。

[0018] 进一步地,所述步骤S3的具体过程如下:

[0019] S301、根据可推荐结果集B,对物品的属性和特征值进行划分,设物品的属性向量为:

$$S = [S_1, S_2, \dots, S_w],$$

[0021] 属性 $S_i$ 的特征值向量为:

$$V_i = [v_1, v_2, \dots, v_j];$$

[0023] S302、通过各个属性的区分度向量和候选推荐集构造属性特征矩阵;

[0024] 对用户A而言,感兴趣物品集 $Q_A$ 中某一物品属性 $S_i$ 的特征值 $v_k$ 在 $V_i$ 所占的比例为 $q_{v_k}^i$ ,则在候选推荐集 $Q_T$ 中,设该物品属性的特征值 $v_k$ 所占的权重:

$$q = q_{v_k}^i$$

[0026] 另,当 $q_{v_i}^x > q_{v_i}^k$  ( $i, k$ 取任意可取值),则认为属性 $S_x$ 的区分度最强;各个属性的区分度取:

$$P_w = \max q_{v_i}^w$$

[0028] 所以,当给出 $Q_A$ 后,可得索性区分度向量:

$$V = [p_1, p_2, \dots, p_w]$$

[0030] 得到 $Q_T$ 后,可得物品的属性特征矩阵:

$$[0031] \quad A_{Q_r} = \begin{bmatrix} x_1^1, x_1^2, \dots, x_1^T \\ x_2^1, x_2^2, \dots, x_2^T \\ \dots \\ x_w^1, x_w^2, \dots, x_w^T \end{bmatrix};$$

[0032] S303、根据属性特征矩阵和属性区分度向量,可以得到候选推荐集 $Q_T$ 的各物品推荐分数向量:

$$[0033] \quad F = V \times A_{Q_r};$$

[0034] S304、根据得到的推荐分数向量对可推荐结果集B的排序,确定初步推荐结果列表C。

[0035] 进一步地,所述步骤S2中可推荐候选集A的人数如果不满足最低推荐人数要求,则向业务方请求扩大搜索条件。

[0036] 进一步地,所述用户的属性及特征值的定义的方式为【属性-值】键值对。

[0037] 进一步地,用户的初始属性由用户在注册时填写,推荐引擎可以根据用户填写的属性值,为初始用户进行特征划分。

[0038] 本发明相对于现有技术具有如下的优点及效果:

[0039] 相比于传统的基于人口统计学和基于内容的推荐方法相比,本发明借鉴协同过滤思想和基于内容的推荐思想,更为注重了用户之间的个体差异性,通过分析用户的历史行为数据,挖掘用户的兴趣进行推荐;因此,推荐结果因人而异,更多的考虑了用户的个人喜好,也可以针对不同的推荐场景定制不同的方法参数。

## 附图说明

[0040] 图1是本发明中公开的基于协同过滤的通用物品推荐方法的流程图。

## 具体实施方式

[0041] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作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。

[0042] 实施例

[0043] 如附图1所示,本实施例公开了一种基于协调过滤的通用物品推荐方法,该通用物品推荐方法借鉴协同过滤思想和基于内容的推荐思想,设计一套通用物品的推荐引擎。其中,通用物品包括但不限于:书籍、音乐、电影,不同的通用物品的区别在于,喜好度的衡量标准以及物品的属性特征差异可根据不同的推荐场景由业务方调整参数。

[0044] 该通过物品推荐方法具体包括下列步骤:

[0045] S1、业务方对推荐物品及用户的属性及特征值进行初始化定义,定义的方式为【属性-值】键值对,例如【身高-170CM】;将用户的行为分类并设置初始搜索引擎的条件,完成初始化设定和推荐引擎APK接入后,向推荐引擎APK发起推荐请求,并将初始化的数据集发送

给推荐引擎APK。

[0046] 具体应用中,用户的初始属性由用户在注册时填写,推荐引擎可以根据用户填写的属性值,为初始用户进行特征划分。

[0047] S2、推荐引擎APK首先根据业务方的初始搜索引擎,将满足搜索条件的数据,汇总形成可推荐候选集A,对可推荐候选集A通过相似度判定得到相似用户集,并根据基于用户的协同过滤思想,进行数据筛选得到可推荐结果集B;如果可推荐候选集A的人数不满足最低推荐人数要求,则向业务方请求扩大搜索条件。

[0048] 其中,所述进行数据筛选得到可推荐结果集B的具体过程如下:

[0049] S201、将用户的行为分为 $T_1 \sim T_k$ 共K类,并对这K类行为分别进行权重赋值 $w_1 \sim w_k$ ,根据不同的用户行为区分为正面、负面以及高、中、低六个维度,赋值向量 $w$ 的取值为 $w = [-2, -1, 0, 1, 2, 3]$ ;

[0050] 具体应用中,根据用户在线上的行为数据,定义不同行为的特征。

[0051] S202、获取用户对物品的行为操作累加值得到用户对物品的喜好度 $H = \sum w$ ,当 $H > 3$ 时,则认为用户对该物品感兴趣;

[0052] S203、通过不同用户对各个物品的喜好度 $H$ ,利用欧几里得距离

$$d(x, y) = \sqrt{\sum H_x - \sum H_y}$$

[0053] 计算得到用户之间的相似度:  $sim(x, y) = \frac{1}{1 + d(x, y)}$

[0054] 当两个用户之间的相似度 $sim(x, y) > k$ 时,其中 $k$ 由业务方决定,即认为两者相似从而得到相似用户集,并根据基于用户的协同过滤思想,得到各个用户的可推荐结果集B。

[0055] S3、将可推荐结果集B根据物品的属性,对属性的特征值进行划分,根据用户的感兴趣集中各个特征值所占的比例,得到可推荐结果集B的用户中每个属性特征值所占的权重,并根据各个最显著的特征集去得到用户感知最敏感的属性,根据感知最敏感的属性不同参考权重,将用户的感兴趣集进行排序,得到初步推荐列表C。

[0056] 该步骤具体过程如下:

[0057] S301、根据可推荐结果集B,对物品的属性和特征值进行划分,设物品的属性向量为:

$$S = [S_1, S_2, \dots, S_w]$$

[0059] 属性 $S_i$ 的特征值向量为:

$$V_i = [v_1, v_2, \dots, v_j] ;$$

[0061] S302、通过各个属性的区分度向量和候选推荐集构造属性特征矩阵;

[0062] 对用户A而言,候选推荐集 $Q_T$ 中某一物品属性 $S_i$ 的特征值 $v_k$ 在 $V_i$ 所占的比例为 $q_{v_k}^i$ ,则在候选推荐集 $Q_T$ 中,设该物品属性的特征值 $v_k$ 所占的权重:

$$q = q_{v_k}^i$$

[0064] 另,当 $q_{v_k}^x > q_{v_k}^i$  ( $i, k$ 取任意可取值)

[0065] 则认为属性 $S_x$ 的区分度最强;各个属性的区分度取:

$$P_w = \max q_{v_k}^w$$

[0067] 所以,当给出 $Q_A$ 后,可得索性区分度向量:

$$[0068] \quad V = [p_1, p_2, \dots, p_w]$$

[0069] 得到 $Q_T$ 后,可得物品的属性特征矩阵:

$$[0070] \quad A_{Q_T} = \begin{bmatrix} x_1^1, x_1^2, \dots, x_1^r \\ x_2^1, x_2^2, \dots, x_2^r \\ \dots \\ x_w^1, x_w^2, \dots, x_w^r \end{bmatrix};$$

[0071] S303、根据属性特征矩阵和属性区分度向量,可以得到候选推荐集 $Q_T$ 的各物品推荐分数向量:

$$[0072] \quad F = V \times A_{Q_T};$$

[0073] S304、根据得到的推荐分数向量对可推荐结果集B的排序,确定初步推荐结果列表C。

[0074] S4、将初步推荐结果列表C发送给业务方,业务方根据需求进行适当重排,最后得到推荐结果列表D,即为最终结果。

[0075] 综上所述,本发明借鉴协同过滤思想和基于内容的推荐思想,注重用户之间的个体差异性,通过分析用户的历史行为数据,挖掘用户的兴趣进行推荐;因此,推荐结果因人而异,更多的考虑了用户的个人喜好,也可以针对不同的推荐场景定制不同的方法参数。

[0076] 上述实施例为本发明较佳的实施方式,但本发明的实施方式并不受上述实施例的限制,其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化,均应为等效的置换方式,都包含在本发明的保护范围之内。



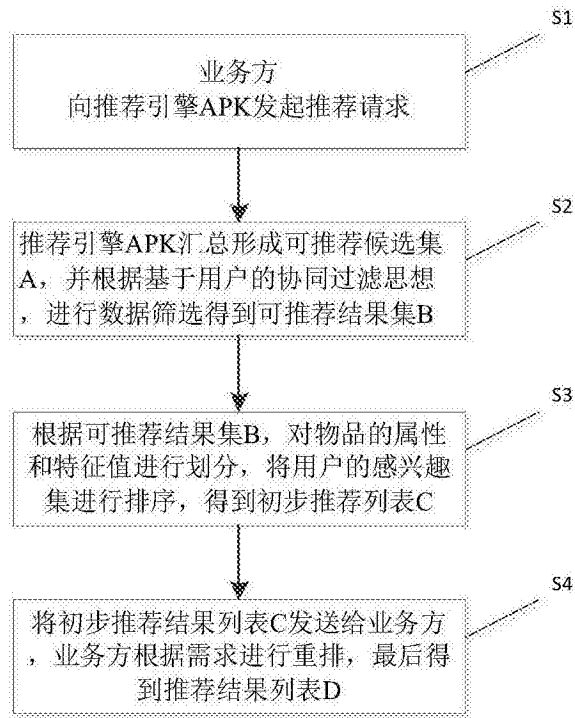


图1