

基于线性衰减的用户兴趣建模^①

蒋 翀 (湖南女子职业大学 现代教育技术中心 湖南 长沙 410004)

费洪晓 (中南大学 软件学院 湖南 长沙 410079)

摘要: 针对电子商务参与者兴趣变化异常频繁的特点,提出了基于线性衰减的用户兴趣模型。根据初始注册信息和浏览行为构建了链式用户偏好表,所有用户的偏好链表组成用户兴趣模型。兴趣模型中的用户偏好值在固定时间间隔 t 进行线性衰减直至变为 0 而被淘汰。若在衰减过程中产生了用户对某项目新的评分,该项目兴趣模型中的偏好值将被更新为此评分。实验证明,基于线性衰减的用户兴趣建模方法是有效的。

关键词: 线性衰减; 用户偏好; 兴趣模型

Automated Testing of Firewall Using TTPlatform

JIANG Chong¹, FEI Hong-Xiao²

(1. Modern Education Technology Center, Hunan Woman's Vocational University, Changsha 410004, China; 2. School of Software, Central South University of China, Changsha 410079, China)

Abstract: For the extremely frequent changes in e-commerce users' interest, a linear-attenuation based user profile is proposed. The registered information and browsing behaviors are used to set up the initial chain user preference list. The User Profile is made up of all users' chain preference list. User ratings decrease at a fixed time interval until the rating has been eliminated to zero. User preference value is probably updated by a new rating in the course of attenuation. The efficiency of user profile is proved by the experiment.

Keywords: linear-attenuation; user preference; user profile

1 引言

用户兴趣模型是个性化推荐系统的基础和核心^[1],直接关系到个性化推荐系统的推荐质量。为了准确的创建用户兴趣模型,就需要从用户感兴趣的信息中提取兴趣特征,获取用户偏好信息,应用建模技术建立用户兴趣模型进行兴趣记录和管理。用户兴趣建模主要包括模型的建立和更新两个过程。

在建立模型时,需要根据推荐资源的类别选取一个合适的表示方式。由于个性化推荐系统是以网络平台为依托,所以用户兴趣模型并不是一般性的描述,而是一种面向算法的、具有特定数据结构的和形式化的描述^[2]。每个人的兴趣喜好不是一成不变的,所以

用户兴趣模型也必须进行更新,这样才能持续地为用户提供高质量的推荐。模型的更新分为显式更新和隐式更新两种方式^[3]。显式更新主要是指系统要求用户对推荐的项目进行反馈和评价,这是最有效也是最直接的方式,但会影响用户的正常浏览行为,为用户带来额外的负担,隐式更新只要是通过跟踪用户浏览行为和挖掘 Web 日志信息来实现^[4]。

本文在研究基于向量空间模型的用户兴趣建模的基础上,结合电子商务参与者兴趣偏好变化频繁的特点,提出了基于线性衰减的用户兴趣建模。该模型采用一个 0.5~5 的数值表示用户偏好,表示用户对每个项目的喜好程度,偏好值根据时间因素自动进行线

^① 基金项目:湖南省科技计划基金(2006JT1040)

收稿时间:2009-09-25;收到修改稿时间:2009-11-11

性衰减,以反映用户兴趣的变化。通过个性化推荐系统的对比实验证明,该模型能在一定程度上提高协作过滤推荐质量。

2 基于线性衰减的用户兴趣模型

现有的推荐算法存在一个问题:注重用户或资源间的相似性,忽略了用户兴趣的动态变化^[5,6],从而导致系统在很大程度上偏离了用户的需求。如果能够在兴趣模型上体现用户偏好的动态变化,就能够有效地弥补推荐算法的不足。用户对某项目的喜好程度一方面可通过用户直接给出的数值反映,也就是对该项目的评分,这类似于问卷调查,本文称之为显式反馈,可获得显式评分;另一方面,用户的一些浏览行为,例如在某页面的停留时间、鼠标点击次数和鼠标滚动时间,均能反映用户兴趣,并可进一步转换为一个表征喜好程度的数值,本文称之为隐式反馈,可获得隐式评分,该过程可通过嵌入了脚本语言的浏览器实现。

考虑到用户兴趣动态变化和时间空间开销,本文引入了基于线性衰减的用户兴趣建模,并从用户及用户偏好表示、兴趣模型设计和线性衰减算法三个方面对其进行分析说明。

2.1 用户及用户偏好表示

在基于线性衰减的用户兴趣模型中,用户对资源项目的评分仅代表评价当时的兴趣度,随着时间的推移,用户对该资源项目的评分将规律性地自动衰减,当项目评分衰减到0时,该资源项目将被兴趣模型所淘汰。评分衰减按照线性规律进行,如图1所示。

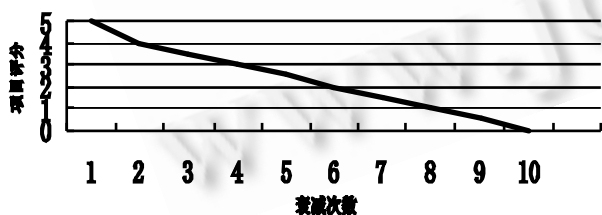


图1 线性衰减图

上图表示的是一个初始评分为5的项目经过9次兴趣衰减后评分变为0的过程,衰减方式是在第一次时间间隔t用户评分递减1,以后每隔t间隔用户评分将递减0.5,直到评分为0而被淘汰。如果在此过程中,用户对该资源项目有新的浏览、评价或购买记录,评分将以显式或隐式评分为准。为了避免兴趣衰减影

响用户正常浏览,整个过程将离线进行。

每位用户*i*都拥有一个用户-项目向量 V_i , $V_i=(R_{t_{i,1}}, R_{t_{i,2}}, \dots, R_{t_{i,j}}, \dots, R_{t_{i,t}})$,其中 $R_{t_{i,j}}$ 表示用户*i*对项目*j*的评分。由于 V_i 的维度随着用户偏好项目的增加而不断增加,同时兴趣衰减又会减少兴趣模型中的项目数,所以 V_i 的维度将一直处于频繁变动过程中。由于链表是动态分配存储空间,不会产生溢出,在任何位置上进行插入和删除只需要修改指针,这些特点非常适合 V_i 的数据存储,所以本文采用链式结构表示 V_i ,每一维分量都用链表中的一个节点来表示,具体结构如图2所示。



图2 用户偏好表示

用户ID是每个用户的唯一标识, $i_{i,1}$ 表示用户*i*的第一个兴趣项目,为保证唯一性和便于计算,采用国际通用的13位商品标准码表示, $R_{t_{i,1}}$ 表示用户*i*对第一个项目的评分。链表按照 $i_{i,1}$ 升序排列,当插入新项目节点时,需要寻找合适插入位置,时间开销为 $O(n)$,但可以为下一步的推荐算法节约一定的时间开销。

2.2 用户兴趣模型设计

用户兴趣模型主要是管理推荐算法所需要使用的反映用户兴趣的源数据,包括用户、项目、用户偏好以及相互关系的表示和更新。用户兴趣模型以若干个用户偏好链表作为数据源,链表的数目由当前用户数决定,每个用户的偏好由属于该用户的偏好链表表示。图3表示用户兴趣模型的逻辑结构,每位用户所有偏好的表示是采用图2所示的链表方式。

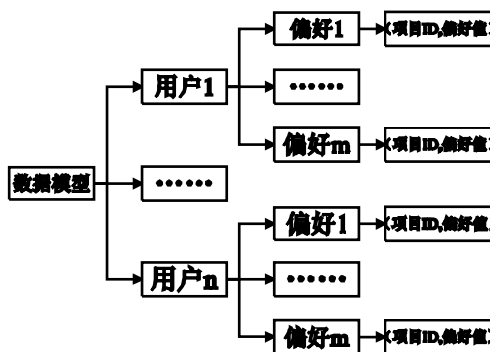


图3 用户兴趣模型结构

图3中的(项目, 偏好值)键值对依赖于用户的某一偏好而存在, 单个偏好依赖于某个具体的用户, 而 n 个相同结构的用户共同组成了用户兴趣模型。在用户兴趣衰减过程中, (项目, 偏好值)键值对中的“偏好值”会根据衰减规律进行更新。在用户偏好的频繁变化过程中, 兴趣模型会自动删除“偏好值”衰减至 0 的单个项目, 也会根据新获取的显式或隐式评分数据添加新项目和更新原有项目。

2.3 线性衰减算法

用户兴趣模型清晰的层次结构最终是为了用户偏好的线性衰减, 实现适用于电子商务参与者用户兴趣模型的更新。用户兴趣偏好的线性衰减就是用户偏好链表中单个偏好键值对(项目, 偏好值)中“偏好值”的规律递减, 具体算法描述如下:

输入: 用户兴趣模型中所有用户的原偏好链表 Preference[]

输出: 更新后的所有用户的偏好链表 Preference[]

算法步骤:

第一步 统计用户兴趣模型 DataModel 中所有用户数目为 n

第二步 创建临时用户链表 tempPreference[], 用户存储当前正在操作的用户偏好链表

第三步 对用户偏好链表进行更新

```
for every user i in DataModel do
tempPreference=useri.Preference[];
for every Preference[j] in tempPreference[]
//若偏好值已经为 0,该用户偏好将被删除
if Preference[j].value==0
delete Preference[j]
//若存在新的评分值,则更新为新偏好值
else if newPreference[j] exist
Preference[j]=newPreference[j]
//否则的话,将偏好值减少 0.5,进行衰减
Else
Preference[j]=Preference[j]-0.5
end for
```

//将修改完成的临时偏好列表赋给用户列表,完成单个用户的偏好列表更新

```
Preference=tempPreference[]
```

```
end for
```

第四步 算法结束

考虑到断电或意外关机等原因导致系统中断运行的特殊情况, 算法加入了临时变量 tempPreference[], 直接复制当前用户偏好列表

到该临时变量, 并进行更新操作, 有可能是兴趣衰减、删除或更新某个具体偏好。在这个过程中, 原用户偏好列表暂时保持不变。当操作完毕后, 直接用临时变量更新原用户列表, 避免数据不一致的情况。

3 用户兴趣模型的建立和更新

3.1 建立

值得注意的是, 用户的个性化信息需求与通过搜索引擎进行的临时的和随机的查询有着根本的区别。用户基于一些概念中的项目关键字进行的站外(谷歌、百度和雅虎等)或站内(亚马逊、当当和淘宝等)随机查询主要存在的问题是基于关键字的需求模式不能全面反映用户兴趣^[7], 没有用户兴趣保存和维护的功能; 同时, 简单的关键字匹配, 往往输出大量相关项目, 真正符合用户兴趣的不多, 用户不得不花费大量时间去处理无关项目。由此看来, 通过建立有效的用户兴趣模型, 帮助用户准确快速的找到符合自身兴趣喜好的资源是非常必要的。

本文在建立初始用户兴趣模型时, 具体算法描述如下:

输入: 结构化用户注册信息和用户浏览行为

输出: 初始用户兴趣模型

算法步骤:

第一步 从用户注册信息和反馈表单中提取兴趣项目获得对应评分 ERT

第二步 生成显式兴趣项目向量 (l_1, l_2, \dots, l_n) 及对应的评分

第三步 根据用户浏览行为获取用户兴趣项目获得隐式评分 IRT

第四步 生成隐式兴趣项目向量 (l_1, l_2, \dots, l_m) 及对应的评分

第五步 合并第二步和第四步生成的兴趣向量得到合并后向量 (l_1, l_2, \dots, l_k) 及对应的评分

第六步 对向量 (l_1, l_2, \dots, l_k) 进行升序排列

第七步 按照排序后向量的各分量及对应评分生成用户偏好链表得到初始用户兴趣模型

第八步 算法结束

从用户注册信息中获取的用户兴趣项目的数量一般是十分有限的, 更多的是用户对某一项目类别的兴趣, 算法将从中提取关注度高的具体项目, 并推测用户评分为“3”; 从用户浏览行为获取的偏好项目相对具体, 并可进一步获得对应项目的隐式评分。这两种途径获取的项目及评分共同构建了用户最初的兴趣模型。

3.2 更新

用户兴趣模型更新主要是依据线性衰减、显式反馈和隐式反馈来进行。线性衰减表示偏好值有规律逐步减少的过程，是对用户自然遗忘行为的一种模拟。显式和隐式反馈是获取用户偏好值的两种不同方式，得到的是某一时刻用户对某项目的偏好值。更新具体算法描述如下：

输入：显式反馈、隐式评分和当前用户兴趣模型

输出：更新后用户兴趣模型

算法步骤：

第一步 对每个显式反馈的项目(l_i, R_{t_i})

$newPreference[j] = R_{t_i}$

第二步 将用户浏览行为转化为对特定项目隐式评分(l_j, R_{t_j})

$newPreference[j] = R_{t_j}$

第三步 按照固定时间间隔 t 进行兴趣衰减，算法描述见 2.3

第四步 算法结束

为了节约系统开销，保证正常的网络浏览速度，用户显式反馈的提交和用户行为的记录是在线进行，用户兴趣模型中项目的匹配、更新和兴趣衰减等操作都是离线进行。

4 实验结果与分析

为了验证基于线性衰减用户兴趣模型的有效性，本文选取传统的基于协作过滤的推荐系统为平台，分别采用基于传统向量空间模型和基于线性衰减的用户兴趣模型进行实验。本实验采用 GroupLens 研究小组在其站点提供的免费数据集，以绝对平均误差(Mean Absolute Error, MAE)为评估标准。

实验采用基于项目的协作过滤，最近邻体积为 30，实验结果如图 4 所示。

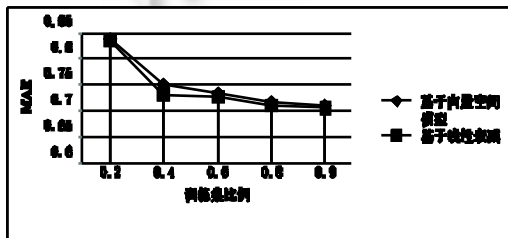


图 4 与传统协作过滤系统推荐质量比较测试结果

从上述实验结果可以得出，随着训练集比例增大，

绝对平均误差越来越小。当该比例从 0.2 增至 0.4 时，MAE 减少幅度最大，之后趋于平缓，在达到 0.8 之后，MAE 变化不再明显。在训练集比例为 0.4、0.6 和 0.8 时，红点所代表的 MAE 值明显比蓝点所代表的 MAE 值要小，差距在 0.005–0.018 之间。实验证明，当选取最近邻体积为 30 的协作过滤推荐中，当训练集比例为 0.2 和 0.9 时，本文提出的建模方式优势并不明显，当训练集比例为 0.4、0.6 和 0.8 时，该建模方式具有一定的优势，特别是达到 0.8 时，得到的 MAE 最小。

本实验在一定程度上证明了基于线性衰减的用户兴趣模型在基于项目的协作过滤系统中要优于基于传统向量空间模型的用户兴趣模型。

5 结语

本文立足于电子商务领域，根据参与者的兴趣特点，提出基于线性衰减的用户兴趣模型，使用(项目, 偏好值)键值对表示用户对单个项目的偏好，若干个该键值对组成的链表表示单个用户的所有偏好，所有用户的偏好链表组成用户兴趣模型。同时，实现所有偏好值的规律递减，模拟自然遗忘的过程，达到用户兴趣动态更新的目的。

参考文献

- 1 应晓敏,刘明,窦文华.一种面向个性化服务的无需反例集的用户建模方式.计算机工程与科学, 2003,25(6):39–42.
- 2 Zhang Y, Koren J. Efficient Bayesian Hierarchical User Modeling for recommender systems. Proc. of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, 2007. 47–54.
- 3 Gadanho SC, Lhuillier N. Addressing Uncertainty in Implicit Preferences. Proc. of the ACM Conference on Recommender Systems, 2007. 97–104.
- 4 Geyer W, Dugan C, Millen DR, Muller M, Freyne J. Recommending topics for self-descriptions in online user profiles. Proc. of the ACM Conference on Recommender Systems, 2008. 59–66.
- 5 邢春晓,高凤荣,战思南等.适应用户兴趣变化的协同过滤推荐算法.计算机研究与发展, 2007,44(2):296–301.
- 6 费洪晓,蒋翀,徐丽娟.基于树状向量空间模型的用户兴趣建模.计算机技术与发展, 2009,19(5):79–81.
- 7 吴晶,张品,罗辛,等.门户个性化兴趣获取与迁移模式发现.计算机研究与发展, 2007,44(8):1284–1292.