

基于用户行为及语义相关实时更新用户兴趣的推荐系统^{*}

全海金^{1,2} 邱玉辉¹ 李 瑞¹

(西南师范大学计算机与信息科学学院¹ 数学与财经学院² 重庆400715)

摘 要 在现有的推荐系统中,其用户兴趣模型都能够有效地表达出用户的兴趣,但在用户兴趣发生变化时却不能很好地调整用户兴趣模型,不能及时适应用户兴趣的动态变化。本文提出一种基于语义相关实时更新用户兴趣模型的推荐系统。该系统能够及时响应用户兴趣变化,从而改善了以往推荐系统对用户兴趣更新不及时所导致的推荐结果不够全面、准确的问题。实验表明该系统能够准确表达用户兴趣,特别是在用户兴趣发生变化时比以往系统具有更高的准确性。

关键词 语义相关,用户行为,推荐系统,用户兴趣模型

The Recommender System with Real-time Updated User Interests Based on User Behaviors and Similar Semantic

QUAN Hai-Jin^{1,2} QIU Yu-Hui¹ LI Rui¹

(School of Computer and Information Science, Southwest China Normal University, Chongqing 400715)¹

(School of Mathematics&Finance, Southwest China Normal University, Chongqing 400715)²

Abstract The user interests models in the existing recommender systems can express the user's interests effectively, but they can't modify the user interests models while the user interests change, and cannot adapt the changes in time. This flaw of the existing recommender systems results in the suggestions to users aren't exact. So we propose a recommender system based on user behaviors and similar semantic. The system can adapt the changes of user interests quickly, and is more exact than the existing recommender systems.

Keywords Similar semantic, User behavior, Recommender system, User interests model

1 引言

目前,随着 Internet 的迅速增长,信息也正在呈爆炸式增长,WWW 已经发展成为包含多种信息资源、站点遍布全球的巨大信息服务网络。然而,在这浩瀚的信息中如何快速地查找有用的信息已成为困扰用户的一大难题。

现阶段所采用的信息检索技术一般都基于布尔表达式或向量空间模型(Vector Space Model),它们假设术语间相互独立,通过逻辑表达式或向量间的内积反映用户查询和文档的相似度。当中存在着这样的问题:1)假设术语间相互独立,并对于用户的查询项进行精确匹配,只能反映用户所需要检索内容的某一方面,无法保证语义概念上的匹配,从而导致检索结果不够全面;2)对用户兴趣更新不够及时,导致检索结果往往为用户以前兴趣,导致检索结果不准确。

本文提出一种基于语义相关实时更新^[1]用户兴趣模型的推荐系统。本推荐系统隐式跟踪^[2]用户行为,获取用户感兴趣文档的特征词及其权重,然后采用贝叶斯分类方法对用户兴趣进行分类,形成多个个性化向量,并且基于贝叶斯原理计算每一个特征词与用户兴趣中每个个性化向量条件概率,将每个特征词划分到与其最为相关的个性化向量中。如果该个性化向量中无该特征词,则为其新建一个特征项,若有则采用一个带有衰减因子的公式对其更新。在推荐过程中采用贝叶斯网络模型从语义相关的角度计算文档与用户兴趣相关度。本推荐系统能够从语义角度上细化用户兴趣,预测用户兴趣模

型和文档间的概念语义相似度,能够对用户兴趣变化作出及时响应,从而改善了用户兴趣更新不及时所导致的推荐结果不够全面、准确的问题。

本文第2节对基于用户行为过程的用户兴趣模型做了整体描述;第3节给出了一个使用该模型得到的一个系统;第4节分析了实验该系统所得的实验结果;最后进行总结并说明今后要做的工作。

2 基于语义相关实时更新的用户兴趣模型

现有的用户兴趣模型更新方式如下:提取用户浏览过的文档的特征术语,假设术语间相互独立,然后对该术语在用户兴趣模型中的权重进行累加。当用户兴趣发生改变时,由于以前的兴趣具有很高的权重,因此在接下来的一段时间里即使用户兴趣改变了,但优先推荐的却是用户以前感兴趣的内容。这显然是不合理的。这里我们提出一种基于语义相关实时更新的用户兴趣模型,该模型能够快速适应用户兴趣变化,及时给出更为准确的推荐结果。

在基于语义相关实时更新的用户兴趣模型中,将用户的兴趣细分为两种:持久兴趣和暂时兴趣。前者是用户长时间所具有的兴趣,它随时间的推移而发生改变,而暂时兴趣的产生常常是根据用户当前的一些操作得来的。在构造用户兴趣 $C = \{C_1, C_2, \dots, C_I\}$ 时会同时考虑用户的持久兴趣 C^{per} 和暂时兴趣 C^{today} , C^{per} 表示的是根据用户在 N 天内的行为所获得的用户兴趣。在表述 C^{per} 前,我们先做如下定义:

^{*} 本文研究得到重庆市自然科学基金资助。项目编号: CSTC2004BB2086。

用户在第 j 天所浏览的网页数为: $S_j (j=0, 1, 2, \dots, N)$ 其中 $j=0$ 表示是今天。

为了更为详细地研究用户一天中的行为, 我们把用户一天中的行为分成相对独立的操作集。操作集的划分标准为: 一系列相对连续的操作作为一个操作集。比如在一个小时内连续的操作作为一个操作集, 间隔了一段时间后的为下一个操作集。当前操作集定义为今天的第 cur 个操作集, 当前操作集的前一个操作集为第 $before$ 个操作集, 也就是存在以下关系:

$$Cur = before + 1$$

在构造 C^{today} 前先定义在第 $r (r=1, 2, \dots, before)$ 个操作集中的第 $hp^{(r)} (hp=1, 2, \dots, S_0^{(r)})$ 个用户浏览网页的特征向量 $w^{hp^{(r)}}$ 如下:

$$w^{hp^{(r)}} = (w_{t_1}^{hp^{(r)}}, w_{t_2}^{hp^{(r)}}, \dots, w_{t_m}^{hp^{(r)}})$$

其中 m 为第 $hp^{(r)}$ 个网页中的术语的个数, $t_k (k=1, 2, \dots, m)$ 为某一个具体的术语。根据术语文件频度和反频度^[9]定义 $w_{t_k}^{hp^{(r)}}$ 如下:

$$w_{t_k}^{hp^{(r)}} = const^{hp^{(r)}}.$$

$$\frac{(0.5 + 0.5 \frac{tf(t_k, hp^{(r)})}{tf_{\max}(hp^{(r)})}) (\log \frac{n}{df(t_k)})}{\sqrt{\sum_{i=1}^m ((0.5 + 0.5 \frac{tf(t_i, hp^{(r)})}{tf_{\max}(hp^{(r)})})^2 (\log \frac{n}{df(t_i)})^2)}$$

其中 $tf(t_k, hp^{(r)})$ 为术语 t_k 在 $hp^{(r)}$ 中出现的频率, $tf_{\max}(hp^{(r)})$ 为 $hp^{(r)}$ 中出现频率最高的术语的频率, n 是全部文档数, $df(t_k)$ 是包含术语 t_k 的文档数, $const^{hp^{(r)}}$ 是用来区分用户是否对网页感兴趣的常数, 其定义如下:

$$const^{hp^{(r)}} = \begin{cases} 1; time \geq Th \\ 0; time < Th \end{cases}$$

其中 $time$ 为用户阅读网页 $hp^{(r)}$ 的时间, Th 为阈值。当阅读时间超过阈值 Th 时认为用户对该网页有兴趣, 反之则认为用户无兴趣, 将该网页过滤掉。

第 r 个操作集的用户兴趣定义为:

$$C^{(r)} = (C_1^{(r)}, C_2^{(r)}, \dots, C_l^{(r)})$$

$C_i^{(r)}$ 对于每一个 $C_i^{(r)} (i=1, \dots, l)$ 当 $P(C_i/t_k^{(r)}) > th2$, $th2$ 为一阈值时 $C_i^{(r)} = C_i^{(r)}$

t_k 对类 C_i 的类条件概率是 $P(t_k^{(r)}/C_i)$, 则根据贝叶斯定理^[8], 可得到 C_i 的后验概率 $P(C_i/t_k^{(r)})$:

$$P(C_i/t_k^{(r)}) = \frac{P(t_k^{(r)}/C_i)P(C_i)}{P(t_k^{(r)})} = \frac{P(t_k^{(r)}/C_i)P(C_i)}{\sum_{i=1}^l P(t_k^{(r)}/C_i)P(C_i)} \quad (1)$$

每一个 $C_i^{(r)}$ 定义如下:

$$\begin{aligned} C_i^{(r)} &= \frac{1}{S_0^{(r)}} \sum_{hp=1}^{S_0^{(r)}} w_{t_k}^{hp^{(r)}} \\ &= \frac{1}{S_0^{(r)}} \sum_{hp=1}^{S_0^{(r)}} const^{hp^{(r)}} \\ &= \frac{(0.5 + 0.5 \frac{tf(t_k, hp^{(r)})}{tf_{\max}(hp^{(r)})}) (\log \frac{n}{df(t_k)})}{\sqrt{\sum_{i=1}^m ((0.5 + 0.5 \frac{tf(t_i, hp^{(r)})}{tf_{\max}(hp^{(r)})})^2 (\log \frac{n}{df(t_i)})^2}} \end{aligned}$$

其中 $S_0^{(r)}$ 为今天的第 r 个操作集中所浏览的网页数。到当前操作集前的用户所有兴趣 $C^{(br)}$ 定义如下:

$$C^{(br)} = (C_1^{(br)}, C_2^{(br)}, \dots, C_l^{(br)})$$

对于每一个 $C_i^{(br)} (i=1, \dots, l)$ 定义如下:

$$C_i^{(br)} = \sum_{r=1}^{before} C_i^{(r)}$$

类似地, 我们定义当前操作集中的网页 $hp^{(cur)} (hp=1, 2, \dots, S_0^{(cur)})$ 的特征向量 $w^{hp^{(cur)}}$ 为:

$$w^{hp^{(cur)}} = (w_{t_1}^{hp^{(cur)}}, w_{t_2}^{hp^{(cur)}}, \dots, w_{t_m}^{hp^{(cur)}})$$

其中 m 为第 $hp^{(cur)}$ 个网页中的术语的个数, $t_k (k=1, 2, \dots, m)$ 为某一个具体的术语。所使用的词频规则 $w_{t_k}^{hp^{(cur)}}$ 定义如下:

$$w_{t_k}^{hp^{(cur)}} = const^{hp^{(cur)}}.$$

$$\frac{(0.5 + 0.5 \frac{tf(t_k, hp^{(cur)})}{tf_{\max}(hp^{(cur)})}) (\log \frac{n}{df(t_k)})}{\sqrt{\sum_{i=1}^m ((0.5 + 0.5 \frac{tf(t_i, hp^{(cur)})}{tf_{\max}(hp^{(cur)})})^2 (\log \frac{n}{df(t_i)})^2}}$$

其中 $tf(t_k, hp^{(cur)})$ 为术语 t_k 在 $hp^{(cur)}$ 中出现的频率, $tf_{\max}(hp^{(cur)})$ 为 $hp^{(cur)}$ 中出现频率最高的术语的频率, n 是全部文档数, $df(t_k)$ 是包含术语 t_k 的文档数, $const^{hp^{(cur)}}$ 是用来区分用户是否对网页感兴趣的常数, 其定义如下:

$$const^{hp^{(cur)}} = \begin{cases} 1; time \geq Th \\ 0; time < Th \end{cases}$$

$time$ 为用户阅读网页 $hp^{(cur)}$ 的时间, Th 为阈值。当阅读时间超过阈值 Th 时认为用户对该网页有兴趣, 反之则认为用户无兴趣, 将该网页过滤掉。当前操作集 $C^{(cur)}$ 的用户兴趣定义为:

$$C^{(cur)} = (C_1^{(cur)}, C_2^{(cur)}, \dots, C_l^{(cur)})$$

$C_i^{(cur)} = C_i^{(cur)}$ (其中 $i=1, \dots, l, P(C_i/t_k) > th2$, $th2$ 为一阈值) $P(C_i^{(cur)}/t_k^{(cur)})$ 可由公式(1)计算得到。每一个 $C_i^{(cur)}$ 定义如下:

$$\begin{aligned} C_i^{(cur)} &= \frac{1}{S_0^{(cur)}} \sum_{hp=1}^{S_0^{(cur)}} w_{t_k}^{hp^{(cur)}} \\ &= \frac{1}{S_0^{(cur)}} \sum_{hp=1}^{S_0^{(cur)}} const^{hp^{(cur)}} \\ &= \frac{(0.5 + 0.5 \frac{tf(t_k, hp^{(cur)})}{tf_{\max}(hp^{(cur)})}) (\log \frac{n}{df(t_k)})}{\sqrt{\sum_{i=1}^m ((0.5 + 0.5 \frac{tf(t_i, hp^{(cur)})}{tf_{\max}(hp^{(cur)})})^2 (\log \frac{n}{df(t_i)})^2}} \end{aligned}$$

用 $C^{(br)}$ 和 $C^{(cur)}$ 构造 $C^{(today)}$ 如下:

$$C^{(today)} = xC^{(br)} + yC^{(cur)}$$

其中 x 和 y 满足 $x+y=1$, 为了强调当前操作集, 我们定义 y 比 x 大。

用户的持久兴趣 C^{per} 由用户近 N 天的兴趣组成。 C^{per} 定义如下:

$$C^{(per)} = (C_1^{(per)}, C_2^{(per)}, \dots, C_l^{(per)})$$

$C_i^{(per)}$ 对于每一个 $C_i^{(per)} (i=1, \dots, l)$ 当 $P(C_i/t_k^{(per)}) > th2$, $th2$ 为一阈值时 $C_i^{(per)} = C_i^{(per)}$ $P(C_i/t_k^{(per)})$ 可由公式(1)类似计算得到。

每一个 $C_i^{(per)}$ 定义如下:

$$C_i^{(per)} = \frac{1}{N} \sum_{j=1}^N \frac{1}{S_j} \sum_{hp=1}^{S_j} w_{t_k}^{hp^{(j)}} \cdot e^{-\frac{\log 2}{hl}(d-d_{t_k_init})}$$

其中 $e^{-\frac{\log 2}{hl}(d-d_{t_k_init})}$ 是一个遗忘因子, 因为用户的某一兴趣会随着时间的流失而衰减。 $d_{t_k_init}$ 是术语 t_k 首次出现在用户兴趣的日期, d 是 $d_{t_k_init}$ 后的某一个日期。 hl 是一个半衰期范围, 这里把它设为 10。也就是用户的一个兴趣十天会衰减一半。

由上所述, 用户兴趣可表示为:

$$C = aC^{per} + bC^{today} = aC^{per} + bx C^{(br)} + by C^{(cur)}$$

其中 a, b, x, y 满足 $a+b=1$ 和 $x+y=1$ 。

3 基于贝叶斯网络模型的相似度计算

贝叶斯网络模型是解决信息检索领域问题的有效手段,因为它可以表示术语间的条件概率和概念语义,并依此预测用户兴趣和文档间的相似度。与假设术语间相互独立的模型相比,它具有更高的查全率和查准率。

贝叶斯网络是一个带有概率注释的有向无环图,图中的点表示要解决的问题中的变量。这种概率图模型能表示变量之间的联合概率分布,分析变量之间的相互关系,可以实现预测、分类、聚类任务^[3]。在本系统中,主要是利用贝叶斯网络模型表示术语间的关系,预测用户兴趣模型与文档间的概念语义相似度,从而实现基于概念语义的网页推荐。

我们的贝叶斯网络模型^[4,5]分为三层,最上面一层的节点代表术语,表示用户兴趣项及其权重;第二层的节点也代表术语,它们与第一层中的术语相关,其值代表相关程度和权重;第三层节点代表文档,节点中的值表示文档与用户兴趣项的相关程度,如图1所示。

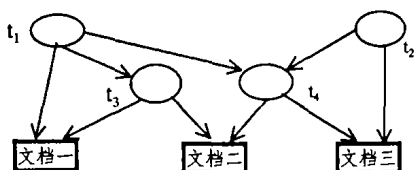


图1 贝叶斯网络模型结构图

最上层的节点及其权重从用户兴趣中可以直接获得,主要要解决的问题是如何得到第一层术语与第二层术语之间的关系,从而求出第二层节点的权重,进一步求出第一层与第三层间的相关度。

3.1 计算术语间的条件概率

我们认为经常出现在同一文件中的术语具有某种相关关系,如上下位关系,所属领域相同关系等,因此我们将其称为相关术语。这里我们运用数据挖掘中的相关规则发现方法挖掘术语间的相关关系。

为了求出第二层节点的权重,有一种方法是存储所有贝叶斯网络模型中由第一层节点到第二层节点的联合条件概率。显然这个数目是非常大的,要保存需要很多的空间,因此我们采用动态生成的方法来计算术语间的联合条件概率。

对每一个网页中的特征向量 $w^{A_p(r)} = (w_{t_1}^{A_p(r)}, w_{t_2}^{A_p(r)}, \dots, w_{t_m}^{A_p(r)})$ 进行统计可以得到网页中术语的文件频度及多个术语同时出现在同意文档中的文件频度。然后采用式(2)来计算术语间的条件概率(节点 A 的权重为通过所有与 A 相关的节点计算出来的权重及节点 A 的原始权重中的最大值):

$$P(T_j | T_{t_1}, \dots, T_{t_m}) = \max(w(T_j), P(T_j | T_{t_1}) * w(T_{t_1}), \dots, P(T_j | T_{t_m}) * w(T_{t_m})) \quad (2)$$

其中,术语 $T_{t_1}, \dots, T_{t_m}$ 是用户兴趣项,它们对应于贝叶斯网络模型中的第一层节点。 $P(T_j | T_{t_k})$ 表示规则 $T_{t_k} \rightarrow T_j$ 的置信度,即 $n(T_j, T_{t_k}) / n(T_{t_k})$, $n(T_j, T_{t_k})$ 是术语集 (T_j, T_{t_k}) 的文件频度, $n(T_{t_k})$ 是术语集 (T_{t_k}) 的文件频度, $w(T_{t_k})$ 表示术语 T_{t_k} 的权重,因为 $T_{t_1}, \dots, T_{t_m}$ 是用户兴趣项,所以 $P(T_j | T_{t_1}, \dots, T_{t_m})$ 就是术语 T_j 的权重。

3.2 用户兴趣模型与文档间概念语义相似度的预测

我们采用文[6]的计算方法,将用户兴趣模型与文档间概念语义相似度表示为用户个性化向量和文档向量间的向量内积。对于文档 D ,其特征向量为 $w^D = (w_{t_1}^D, w_{t_2}^D, \dots, w_{t_m}^D)$,用户兴趣 $C = \{C_1, C_2, \dots, C_l\}$ 中的每一个个性化向量为 $C_i = (w_{t_1}^{C_i},$

$w_{t_2}^{C_i}, \dots, w_{t_l}^{C_i}) (i=1, \dots, l)$ 。将 C_i 基于贝叶斯网络模型扩展后得到贝叶斯网络模型中的第二层节点,也就是基于贝叶斯网络模型的概念语义扩展后的新向量 $C_{i2} = (w_{t_1}^{C_{i2}}, w_{t_2}^{C_{i2}}, \dots, w_{t_l}^{C_{i2}})$ 。 w^D 与 C_{i2} 之间的相似程度可用向量之间的夹角来度量,夹角越小说明相似度越高,相似度的衡量值为向量之间夹角的余弦值,其计算公式如下:

$$\text{Sim}(w^D, C_i) = \text{Sim}(w^D, C_{i2}) = \frac{w^D \cdot C_{i2}}{\|w^D\| \cdot \|C_{i2}\|} \quad (3)$$

对于 $C = \{C_1, C_2, \dots, C_l\}$ 中的每一个个性化向量 $C_i (i=1, \dots, l)$ 都做类似操作,则文档与用户兴趣模型间相似度 $\text{Sim}(w^D, C_i)$ 为:

$$\text{Sim}(w^D, C) = \text{Max}(\text{Sim}(w^D, C_i), i=1, \dots, l)$$

4 实验

4.1 基于语义相关实时更新的用户兴趣模型的推荐系统

为了验证以上理论,我们构造了一个基于新闻浏览的推荐系统。在该系统中首先建立一个用户兴趣初始化模型,并为每一个注册用户分配一个初始的用户兴趣模型,然后实时监测用户的浏览行为,对用户阅读时间超过阈值的网页提取其特征术语,再使用贝叶斯分类器对特征术语分类,计算其权重,更新用户暂时兴趣、持久兴趣,最后得到当前用户兴趣。这样,当用户进行新闻浏览时,推荐系统先计算用户兴趣模型和用户未浏览过的新闻文档之间的相关性,从而预测出用户对未浏览新闻文档的感兴趣程度,然后再按照由高到低的顺序进行推荐。推荐系统整体结构如图2所示。

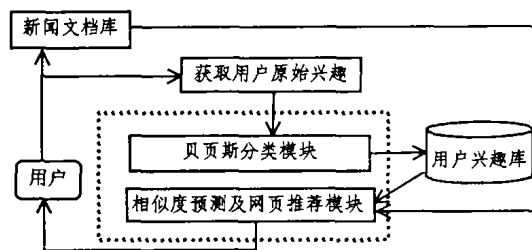
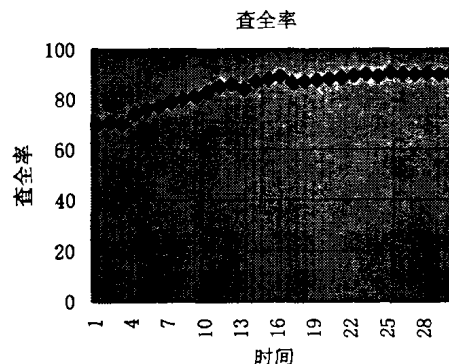


图2 系统结构图

4.2 实验结果

我们所选的实验数据为来自网站 www.sahu.net 的新闻频道中30天里共6235篇新闻页面。推荐系统采用 c# 实现,程序运行在网站的服务器端。该网站每天发布新闻200条左右,包括了社会、娱乐等各类新闻。在注册用户的浏览过程中,系统实时监测用户行为,获取用户兴趣,并根据用户兴趣给用户推荐新闻。当 $Th=0.317, \alpha=0.615, b=0.3825, x=0.151, y=0.849$ 时得到实验结果如下:



(下转第94页)

2.5%, NVOD(150,50)和SAM(150,25)的VCR操作失败率则要远远高于前两者,达到35%以上。并且,VCRSM是零时延的,不存在客户反悔的情况;而NVOD和SAM是基于批处理的,客户的平均等待时间为0.4分钟,客户反悔率达50%以上。

在实验中还发现,当有VCR操作时,服务器端的带宽随着 λ 近似线性增大。这是因为,实验中假设快进和快退操作播放速率恰好为客户端的带宽,因此对于快速流,服务器无法再进行合并,相当于单播流的性能。但如果客户端带宽大于快速播放速率,快速流之间也可合并,故当 λ 趋向于无穷大时,相应带宽不会线性增长。

结论 本文结合客户端主动缓存技术对基本流合并策略进行扩展,提出了一个支持VCR功能的多播调度算法VCRSM。VCRSM在客户端通过主动缓存机制一方面预取数据,同时提供局部VCR操作提供支持,并使用延迟请求为流合并和服务带宽优化创造条件。在服务器端,将流调度策略和VCR请求相结合,在提供实时响应的基础上通过优化流合并来提高流的共享度,减少服务器带宽消耗。

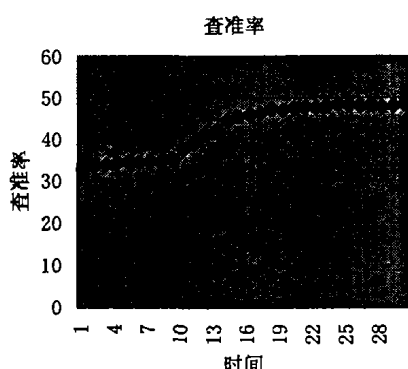
与已有VOD系统相比,VCRSM的主要优点是:真正支持实时交互式流媒体服务,对所有VCR功能提供即时服务,没有延时(初始延时和VCR操作延时);具有良好的可扩展性,实验表明,在一定磁盘空间支持下,VCRSM的服务器端带宽消耗受用户规模及VCR操作频度影响相对较小。

参考文献

- 1 Viswanathan S, Imielinski T. Pyramid Broadcasting for Video On Demand Service. In: Proc. SPIE Multimedia Computing and Networking Conf. (MMCN '95), 1995. 66~77
- 2 Hua K A, Shen S. Skyscraper broadcasting: A new broadcasting scheme for metropolitan video-on-demand systems. In: Proc. of ACM Sigcomm'97, 1997
- 3 Eager D L, Vernon M K. Dynamic Skyscraper Broadcasts for Video-on-Demand. In: Proc. 4th Intl. Workshop on Multimedia Information Systems, Sept. 1998

- 4 Dan A, Sitaram D, Shahabuddin P. Scheduling Policies for an On-Demand Video Server with Batching. In: Proc. of ACM Multimedia, Oct. 1994. 15~23
- 5 Gao Lixin, Towsley Don. Supplying Instantaneous Video-on-Demand Services Using Controlled Multicast. In: Proc. IEEE ICMCS'99, Florence, Italy, June 1999
- 6 Gao Lixin, Zhang Zhi-Li, Towsley D. Catching and Selective Catching: Efficient Latency Reduction Techniques for Delivering Continuous Multimedia Streams. In: Proc. of the Seventh ACM Intl. Conf. on Multimedia (Part 1), 1999. 203~206
- 7 Ramesh S, Rhee I, Guo K. Multicast with Cache (Mcache): An Adaptive Zero-Delay Video-on-Demand Service. In: Proc. of IEEE INFOCOM, 2001
- 8 Eager D, Vernon M, Zaborjan J. Minimizing Bandwidth Requirements for On-Demand Data Delivery. In: Proc. of the ninth annual ACM-SIAM symposium on Discrete algorithms, 1998. 11~20
- 9 Eager D, Vernon M, Zaborjan J. Optimal and Efficient Merging Schedules for Video-on-Demand Servers. In: Proc. 7th ACM Intl. Multimedia Conf. (ACM MULTIMEDIA '99), Nov. 1999. 199~202
- 10 Kwon J B, Yeom H Y. Providing VCR Functionality in Staggered Video Broadcasting. Trans. on Consumer Electronics (SCI), 2002, 48(1): 41~48
- 11 Fei Zongming, Kamel I, Mukherjee S, Ammar M H. Providing Interactive Functions for Staggered Multicast Near Video-On-Demand Systems. In: Proc. of the IEEE Intl. Conf. on Multimedia Computing and Systems, 1999
- 12 Almeroth K C, Ammar M H. On the Use of Multicast Delivery to Provide a Scalable and Interactive Video-on-Demand Service. IEEE Journal of Selected Areas in Communications, Aug. 1996, 14
- 13 Fei Zongming, Ammar M H, Kamel I, Mukherjee S. Providing Interactive Functions through Active Client-Buffer Management in Partitioned Video Multicast VOD Systems. In: Proc. of First Intl. Workshop on Networked Group Communications (NGC'99), Nov. 1999. 152~169
- 14 Biersack E W, Jean-Marie A, Nain P. Open-loop Video Distribution with Support of VCR Functionality. In: Proc. of the Performance 2002 Conf. Roma, Italy, 2002
- 15 Liao Wanjiun, Li V O K. The Split and Merge Protocol for Interactive Video-on-Demand. IEEE MultiMedia, 1997, 4(4): 51~62
- 16 Abram-Profeta E L, Shin K G. Providing Unrestricted VCR Functions in Multicast Video-on-Demand Servers. In: Proc. of IEEE Intl. Conf. on Multimedia Computing and Systems (ICMCS'98), Austin, Texas, 1998
- 17 The ns Manual. July 23 2003. <http://www.isi.edu/nsnam/ns/>

(上接第78页)



由此可见,该系统具有较高查全率及查准率,并能够较快地适应用户兴趣的变化。

结束语 本文提出了一种动态跟踪用户兴趣的推荐系统。该系统从语义角度上对用户兴趣细化分类,形成多个个性化向量,并从语义角度上计算了用户兴趣与文档间的相关度,从而能够实时检测用户兴趣及用户兴趣的变化。经实验证明,所构造的用户兴趣模型对用户作出的推荐较为全面和准确,而且在用户兴趣发生变化的情况下能快速、准确地对用户兴趣模型作出调整。我们要作的下一步工作是:进一步细化用户

行为对其兴趣模型的影响。比如鼠标滚动、页面保存、添加到收藏夹等等。

参考文献

- 1 Bonino D, Corno F. A Real-Time Evolutionary Algorithm for Web Prediction WI-2003. In: the 2003 IEEE/WIC Intl. Conf. on Web Intelligence, Halifax, Canada, Oct. 2003
- 2 Claypool M, Brown D. Inferring User Interest. IEEE Internet Computing, Nov./Dec. 2001
- 3 Callan J, Lu Z, Croft W. Searching distributed collections with inference networks. In: Proc. of ACM SIGIR Conf. Seattle, 1995. 21~28
- 4 Campos L M, Huete J F. Building Bayesian Network-Based Information Retrieval Systems. In: Proc. of the 11 Intl. Workshop on Database and Expert Systems Applications (DEXA'00)
- 5 Campos L M, Huete J F. Document Instantiation for Relevance Feedback in the Bayesian Network Retrieval Model. ACM SIGIR'01 Workshop on Mathematical/Formal Methods in IR, New Orleans, Louisiana, 2001
- 6 Rbeiso B, Muntz R. A belief network model for IR. In: Proc. of 19th ACM-SIGIR Conf., 1996
- 7 Kelly D, Teevan J. Implicit feedback for inferring user preference: a bibliography. SIGIR Forum, 2003, 37(2): 18~28
- 8 Han Jiawei, Kamber M. Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers, ISBN 1-55860-489-8, Aug. 2000. 550
- 9 Balabanovic M, Shoham Y, Yun Y. An Adaptive Agent for Automated Web Browsing. Journal of Visual Communication and Image Representation, 1995, 6(4)