

手机广告推荐中的用户兴趣建模研究 *

顾其威, 郭 鹏, 潘 锋[†]
(深圳大学 管理学院, 广东 深圳 518060)

摘 要: 为了增强基于 WAP 网页的手机广告推荐中用户建模的准确性, 并对“非邀”式广告推荐中脱离用户兴趣试探性推荐进行修正, 针对手机广告推荐中手机屏幕小、用户注意力集中等特点, 根据用户对广告的访问历史和操作模式建立其广告兴趣模型和非兴趣模型, 同时分析用户网页访问模式探测其网页兴趣度, 在此基础上建立用户综合兴趣模型。分别采用基于网页兴趣模型、基于广告兴趣模型和基于用户综合兴趣模型进行广告推荐, 随着样本空间增大, 综合兴趣模型的查准率明显优于另两者。实验验证了用户综合兴趣模型在手机广告推荐中的有效性和优越性。

关键词: 电子商务; 手机广告; 综合兴趣; 个性化推荐

中图分类号: TP309; TP301.6 **文献标志码:** A **文章编号:** 1001-3695(2012)02-0579-03
doi:10.3969/j.issn.1001-3695.2012.02.047

Research on user interest modeling in mobile ad recommendation

GU Qi-wei, GUO Peng, PAN Feng[†]
(College of Management, Shenzhen University, Shenzhen Guangdong 518060, China)

Abstract: For enhancing the accuracy of mobile advertising recommendation on the basis of WAP Web page, and amending the exploratory recommended in the mobile advertising recommendation of which recommendation pattern is No Invitation, on account of mobile phone's small screen and user's great attention, builds advertising interest model and noninterest model according to the ads visiting history and operation mode. Simultaneously, analyzed the Web access mode to detect the user's Web interest degree, proposed a user comprehensive interest model which considering the user's Web interest and user's ads interest. Employed three recommended patterns to recommend the mobile advertising based on Web page interest model, advertising interest model and comprehensive interest model, with the sample space growing, the accuracy of comprehensive interest model was superior to other two models. Experiments prove that the effectiveness and superiority of comprehensive model in the mobile advertising recommendation.

Key words: E-commerce; mobile advertising; comprehensive interest; personalized recommendation

3G 时代的启动将数字营销纵深推向无线互联网媒介, 广告作为推动企业营销的工具, 广告主将要求其具有更高的精准性和有效性。在企业追求高精准、高效率、高附加值、低成本的营销战略下, 个性化广告推荐系统将在 3G 时代被委以更高的营销使命^[1]。

目前基于电脑的个性化推荐中, 数据获取具有较大的局限性, 无法将获取的信息与用户完全对应, 只有用户注册网站后才可以获得其准确的数据信息, 而电脑浏览更多的属非注册应用。手机和使用者是一一对应关系, 手机以其灵巧性、便携性和智能性已经成为主人形影不离的帮手, 手机浏览完全等同于注册用户应用, 获取的数据具有确切的个人特征。

因为手机终端的个人性^[2], 可以精确地统计广告的受众人数^[3], 广告主借助于精确统计的数据评价广告效果, 进一步审定广告投放策略, 并利用手机媒体的互动和实时特点, 按照需要及时变更广告形式和内容。基于这些技术优势, 从基础层面确立了手机广告的经济价值, 即节省广告主的费用投入, 增强广告的实际效果, 改变传统媒体广告浪费广告费的固有缺陷。手机广告与传统广告比较如表 1 所示。

表 1 手机广告与传统广告比较

媒体特性	印刷媒体	广播	电视	互联网	手机
到达率	低	中	高	中	高
个性化	低	低	低	中	高
随身性	中	低	低	低	高
实时性	低	高	高	高	高
信息容量	高	中	高	高	低
互动性	低	低	低	中	高
可靠性	中	中	中	高	高
专注性	低	低	中	中	高
可测量性	低	低	低	中	高
分众性	低	低	低	中	高

手机广告与传统广告相比, 在精准、实时、互动和高效性上具有巨大优势, 但因为手机自身硬件限制, 譬如屏幕尺寸小、其信息容量有限^[4,5]等, 因此手机广告推荐有其自身特点。

1 手机广告个性化推荐

个性化推荐是根据用户的兴趣特点和购买行为, 向用户推荐用户感兴趣的信息和商品。个性化推荐分受邀和非邀两类。受邀推荐指用户主动发出请求并提供线索以获取与之相关的

收稿日期: 2011-07-05; 修回日期: 2011-08-18 基金项目: 国家自然科学基金资助项目(71001072); 广东省自然科学基金资助项目(9451806001002694)

作者简介: 顾其威(1957-), 男, 上海人, 教授, 主要研究方向为电子商务、信息系统、个性化服务(gqw@szu.edu.cn); 郭鹏(1987-), 男, 河南新乡人, 硕士研究生, 主要研究方向为电子商务、个性化服务; 潘锋(1977-), 男(通信作者), 河南驻马店人, 博士研究生, 主要研究方向为数据挖掘、电子商务。

信息。非邀推荐指用户未发出显性请求而系统主动推荐其所需信息。手机广告可以是受邀推荐,比如基于 LBS 的手机广告搜索^[6,7],其推荐的内容属同质产品;本文主要研究基于 WAP 网页的非邀类手机广告推荐,其推荐的内容可以是同质产品,也可以是异质产品。这里同质产品推荐指推荐的内容与用户当前关注的内容属同类,异质产品推荐反之。目前的研究中较多对手机广告按照受邀推荐来研究。

非邀类广告推荐主要按照用户兴趣进行同质产品推荐,但考虑到广告主尽量扩大其广告影响力的需要,也可以偏离用户兴趣进行异质产品推荐,即对用户进行非兴趣试探性推荐。但非兴趣试探性推荐必须有所节制,以免破坏用户体验,引起用户反感,比如其推荐产品不能是用户明显不感兴趣产品。

手机广告推荐按照广告位进行。所谓广告位指能够显示手机广告的位置。广告位决定了所能显示的广告类型、广告媒体类型和广告的尺寸,同时通过广告位可以反馈用户的广告访问行为。广告主在投放广告时对广告进行了描述,包括广告主题、广告内容、广告费用、投放用户群、广告类型、广告媒体类型和广告尺寸等。广告类型、广告媒体类型和广告尺寸是为匹配广告位而提供的,与用户兴趣建模没有关系,因此称之为手机广告的自然属性。该属性在手机广告推荐时起决定性作用。广告费用和投放用户群是广告主体现的广告属性,与手机广告本身没有关系,因此称之为手机广告的社会属性。该属性在手机广告推荐时起重要作用。广告主决定了投放广告的人群,广告费用越高,被推荐概率越大。广告主题和广告内容是手机广告表达内容的直接体现,并据此对广告进行分类,因此称之为内容属性。该属性是用户广告兴趣建模的主体,它与用户兴趣进行匹配产生推荐结果。

由于手机屏幕小的缘故,手机广告推荐不可能像电脑互联网广告推荐一样,一次可以推荐数个广告产品,因为手机每屏只有一个广告位,所以其推荐只能是每次推荐一个广告产品。大部分手机广告在广告位不能显示全部内容,只能显示其 logo 或索引,用户有兴趣可点击查看内容,因此可以根据用户的行为来分析得出用户对广告的感兴趣程度。同样由于每一屏只有一个广告位,更易吸引用户注意力,用户操作率也更高,因此更方便探测用户的广告兴趣度。

2 用户兴趣建模

手机广告推荐中用户兴趣模型的构建需要考虑两方面内容:通过对用户浏览过的网页分析可以捕捉到用户的兴趣,建立相应的用户网页兴趣模型;通过对用户浏览过的广告进行分析,更能准确地捕捉用户兴趣,建立相应的用户广告兴趣模型。

2.1 用户网页兴趣建模

用户网页兴趣建模流程如图 1 所示。

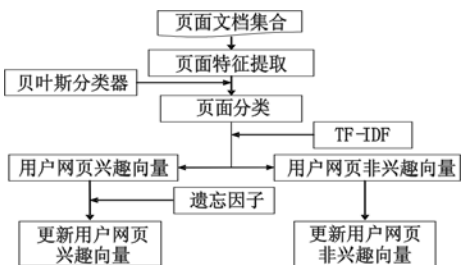


图1 用户网页兴趣建模流程

用户兴趣的获取有两种方式,即显式获取和隐式获取^[8]。显式获取主要是人口特征信息,如背景、性别、年龄、爱好等,这些由用户注册时提供。隐式获取是兴趣获取的主要形式,从用户浏览的内容分析中得出。

获取用户浏览过的网页及相应网页操作行为,去除噪声数据,如百度、Google 等搜索引擎主页。根据用户操作行为判断用户的网页兴趣度,如用户刚打开就关掉,以及浏览时间和网页内容量不符合的网页等可判断为用户不感兴趣。本文用向量空间模型(VSM)^[9]来描述网页和用户兴趣。

有效网页采用朴素贝叶斯^[10]分类得到用户兴趣向量 $PI = (w_1, w_2, w_3, \dots, w_n)$ 和用户非兴趣向量 $NPI = (Nw_1, Nw_2, Nw_3, \dots, Nw_n)$, w_i, Nw_i 表示特征项 $Word_i$ 的权重,产生用户兴趣模型和用户非兴趣模型。

对于一个待分类网页 d_k 来说,划分为感兴趣类或者非兴趣类取决于其概率的大小。在贝叶斯分类器中,取经验值后验概率阈值为 $p = 0.6$,但是某网页归属感兴趣的概率 p_{kl} 在 0.6 附近,则不能简单地判断为感兴趣类。为了解决这个问题,本文利用用户非兴趣模型来解决。计算 p_{kNI} ,若满足

$$|p_{kNI} - p|/p \leq \xi \tag{1}$$

$$(p_{kNI} - p_{kl}) \geq \theta \tag{2}$$

其中: p_{kl} 为网页归属于感兴趣类的概率; p_{kNI} 为网页归属于非兴趣类的概率, ξ 为大于0的小常量, θ 为大于0的小常量, ξ 取值为 0.0010.01, θ 取值为 0.050.15。满足式(1)说明 d_k 与 NPI 有很强的相似性,满足式(2)说明 d_k 与 NPI 的相似度大于 PI 的相似度。满足以上两式保证了 d_k 归属用户非兴趣向量集的正确性。

因为用户的兴趣是不断变化的,所以要定期更新用户网页兴趣模型和用户非兴趣模型。本文引入心理学里的遗忘因子^[11],即根据艾宾浩斯遗忘曲线,人接触新知识一周后记忆便开始衰退,因此引入遗忘参数 $e^{\log_2(\text{Now-Last})/T}$, T 代表生命周期为 7, Now 为系统现在时间, Last 为最近一次访问某类的时间。因此用户的兴趣向量为

$$PI = e^{\log_2(\text{Now-Last})/T} \times PI_{old} + Ctr \tag{3}$$

$$Ctr = (1/\cos(d_k, X_i))d_k \tag{4}$$

其中: Ctr 为网页对用户兴趣模型的贡献。

2.2 用户广告兴趣建模

用户广告兴趣模型建立是用户兴趣建模中的重要一环。通过分析用户对广告的浏览行为,建立用户广告兴趣模型和用户广告非兴趣模型。

2.2.1 手机广告的 VSM 描述

对于网页来说,基于 TF_IDF 的空间向量模型是最合适的表达方式,但是对于手机广告来说,TF_IDF 并不完全适用。因为手机广告有自己的自然属性、社会属性和内容属性。这些属性由广告商投放广告时自行进行描述。手机广告的内容属性可以参考网页兴趣建模。因此手机广告在 VSM 中可以表示为 $Ad_i = (N_i, S_i, C_i)$ 。其中: $N_i = (\text{广告类型, 广告媒体类型, 广告尺寸})$,与广告位属性匹配,返回值为布尔类型; $S_i = (\text{广告费用, 投放人群})$,第一维广告费用 $S_i(1)$ 返回值越大,该条广告被推荐概率越大;第二维投放人群 $S_i(2)$ 与显式获取的用户信息匹配,返回值为布尔类型; $C_i = (\omega_1, \omega_2, \dots, \omega_n), (\omega_1, \omega_2, \dots, \omega_n)$ 为表征广告的特征词,与用户的兴趣向量匹配。

2.2.2 用户广告行为分析

因为手机屏幕限制,手机广告分为完全显示和分级显示,且大部分广告采用分级显示。分级显示广告指在广告位只显示其品牌 logo,用户将鼠标放在 logo 上可以放大查看索引,点击 logo 或索引可以进入广告交互界面。完全显示广告如用户登录时的全屏广告或者页面切换之间的全屏广告等,对此类广告用户可选择播放和关闭播放。针对图片、视频广告,若用户允许图片显示 5 s,则认为用户对广告稍感兴趣;若用户允许图片显示 8 s,则认为用户很感兴趣;若用户选择关闭,则认为不感兴趣。具体如图 2 所示。

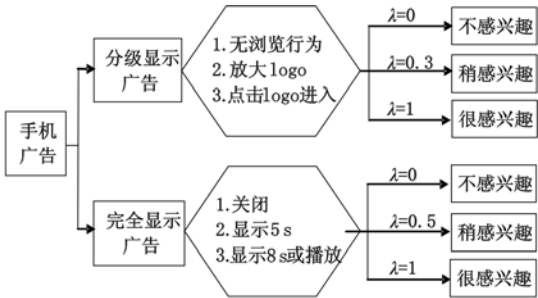


图2 用户手机广告兴趣

用户广告兴趣度 $DAI_i = \lambda C_i = \lambda(\omega_1, \omega_2, \dots, \omega_n)$ 。如图 2 所示,在分级显示广告中,若用户放大 logo,则该条广告的兴趣度 $DAI_i = 0.3(\varphi_1, \varphi_2, \dots, \varphi_n)$ 。当 $\lambda = 0$ 时,则该条广告属于用户广告非兴趣模型,其中 λ 为经验值。

2.2.3 用户广告兴趣模型

用户广告兴趣建模流程如图 3 所示。

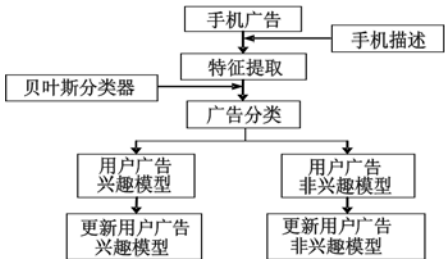


图3 用户广告兴趣建模流程

因为手机广告的特有社会属性,使其可脱离用户兴趣进行试探性推荐,但需在试探频度和试探内容两方面加以约束。

广告兴趣推荐和试探性推荐可按 $n:1$ 的比例进行,比如 $9:1$,即向用户推荐的 10 条广告中,只有一条是试探性推荐。用户广告非兴趣模型中数据大部分来自试探性推荐的结果。广告平台向用户试探性推荐了一条广告,若用户选择关闭,则认为用户对此广告不感兴趣;若用户点击查看了该广告,则认为对其感兴趣。但是根据用户的单次关闭操作就把广告归类于用户非兴趣模型而不再进行推荐,则较为武断。用户可能是因为心情或者时间关系没有浏览广告,因此对于用户不感兴趣的广告应该有逐步判断的过程。假设用户网页兴趣在用户综合兴趣模型中其权重为 α ,用户广告兴趣在用户综合兴趣模型中占 β , $\alpha + \beta = 1$ 。对于每条广告,其推荐权重与用户的浏览行为有十分重要的关系,若第一次推荐被关闭,则将其归类于非用户兴趣模型中,但是推荐权重为 $(1/2)\beta$,若第二次推荐被关闭,则其推荐权重变为 $(1/4)\beta$,其推荐权重按 $(1/2^n)\beta$ 变化, n 是该类广告被连续关闭的次数。在此过程中,若根据用户的网页兴趣发现用户对此类的主题很关注,则该类广告的推荐权重变为 α ,并且从用户广告非兴趣模型中删除,实现对用户广告

兴趣模型和广告非兴趣模型的更新。

广告主在把广告提交广告平台时已经进行了简单的分类,在此基础上运用朴素贝叶斯方法进行分类。手机广告的前两个属性不参与分类,第三维采用朴素贝叶斯分类器进行分类。在 VSM 中,待推荐手机广告首先经过自然属性的匹配,然后经过用户广告非兴趣模型过滤掉用户不感兴趣的广告,再经过社会属性的选择,最后经过用户兴趣模型选择出用户感兴趣的广告,产生最优手机广告推荐集。

2.3 用户综合兴趣模型

根据用户网页兴趣模型只能推荐与网页内容主题相关的广告产品,则扭曲了广告推荐的本质。若只通过用户广告兴趣模型进行广告推荐,虽然得到的用户兴趣比较准确,但是获取的用户兴趣不够完整。因此,将两者结合并辅助用户非兴趣模型^[12],对用户进行试探性推荐,这样可以得到更好的推荐效果,也能够取得较好的商业效果,更能体现广告推荐的真正内涵。

在个性化广告推荐中,与用户网页兴趣模型相比,用户广告兴趣模型更能直接体现用户对广告的兴趣,两者以不同的权重共同决定用户的综合兴趣。用户综合兴趣为

BI=α×PI+β×DAI*(5)*

其中: α 是用户网页兴趣在综合兴趣中的权重, β 是用户广告兴趣在综合兴趣中的权重。用户综合兴趣是因变量,两者满足 $\alpha + \beta = 1$ 。

3 实验验证

实验数据采用手机阅读杂志中两个频道,即财经和体育,包括页面 400 个左右。推荐系统采用开发的基于 Symbian 系统的 v1.0 手机客户端,每个页面只推荐一条广告。

a) 首先确定用户综合兴趣模型中的两个参数 α 和 β 。将获取的数据随机进行五次划分,得到五组数据子集(训练集和测试集),采用交叉验证的方法对 α 进行确定。在这里采用平均绝对误差(MAE)评价方法。

MAE=Σ*(p_r-r_i)/n**(6)*

其中: p_{r_i} 为预测评分; r_i 为实际评分; n 是测试集中目标用户真实评分项数。五次实验结果如图 4 所示。

由图 4 可知,当 α 取 0.3,用户评分的 MAE 值最小,此时 β 取 0.7,即用户综合兴趣模型为

BI=0.3PI+0.7DAI*(7)*

b) 为了验证用户综合兴趣模型的有效性,本文分别采用基于网页用户兴趣模型、广告兴趣模型和用户综合兴趣模型进行广告推荐。因为每个页面上仅推荐一条广告,因此只考察系统的查准率。查准率=正确推荐的广告数目/推荐的广告数目。三种方法结果如图 5 所示。

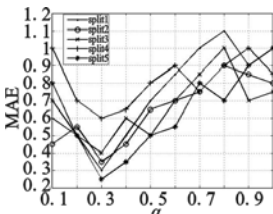


图4 五次交叉实验结果

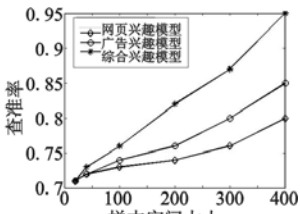


图5 查准率对比

从图 5 可看出,综合兴趣模型开始与其他两种模型的查准率一样,这是因为系统处于开始学习阶段,随着样本空间增大,综合兴趣模型的查准率明显优于其他两种模型,(下转第 585 页)

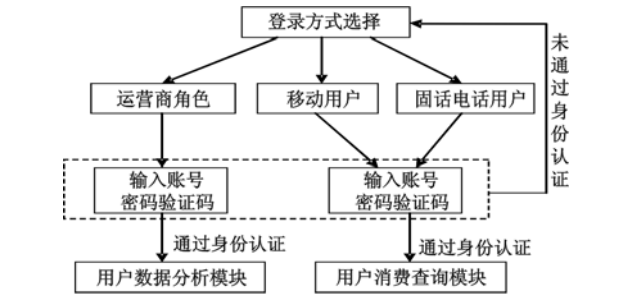


图4 登录模块流程

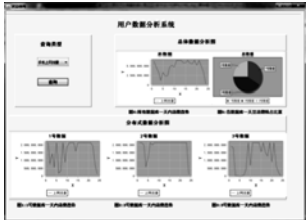


图5 用户数据分析系统

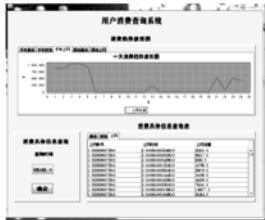


图6 用户消费查询系统

用户点击查询按钮,图 6 上面会显示该用户一天内的消费趋势波形图,下面的表格显示使用业务的具体信息情况;如果该用户在当天通话了,表格中会显示通话的所有记录,包括主叫方、被叫方、通话时间、通话时长;如果该用户发短信了,点击短信标签,会显示短信的所有记录,包括发送号码、接收号码、发送时间。

5 结束语

本文提出的基于大规模廉价计算平台的海量数据处理系统模型,主要实现以数据为中心的计算密集型的经济性超级计算平台,对大数据集进行处理、分析,向用户提供高效的服务。进一步需要考虑是在采用分布式存储的方式来存储数据时,如

(上接第 581 页)在样本达到 400 时,单独使用网页兴趣模型的准确率为 80%,单独使用广告兴趣模型的准确率为 85%,而综合兴趣模型的准确率达到 95%,体现了综合兴趣模型对用户兴趣的准确把握。

4 结束语

随着手机终端的普及和 3G 时代的到来,手机广告推荐在广告产业中将发挥重要的作用,用户兴趣建模在个性化推荐中具有决定性作用。本文通过对用户浏览过的网页分析获得用户网页兴趣,对用户浏览过的广告进行内容分析和浏览行为分析,以获得用户广告兴趣,并辅以用户网页非兴趣模型和用户广告非兴趣模型辅助建模,进而建立用户综合兴趣模型。通过实验证明,综合兴趣模型具有更好的精确度,但由于实验样本数据较少,模型的实际应用效果有待进一步检验。针对广告的试探性推荐,本文提出采用调整推荐频度系数的方法加以约束,但由于条件所限,其效果有待商业实践加以验证。下一步研究方向是将广告营销策略与模型结合进行个性化推荐。

参考文献:

[1] DROSSOS D A, FOUSKAS K G. Mobile advertising: product involvement and its effect on intention to purchase[C]//Proc of the 9th International Conference on Mobile Business(ICMB'2010). Athens, Greece; International Journal of Electronic Commerce, 2010:183-189.

何采用有效冗余存储的方式来保证存储数据的可靠性,以高可靠软件来弥补硬件的不可靠,同时又不会浪费存储资源。该平台适用于多种海量数据的分布式处理。进一步需要改进的是系统的编程模型应该让后台复杂的并行执行和任务调度对用户和编程人员透明。目前对系统的管理和监控方式是集中式的,为了避免出现系统瓶颈和单点失效问题,进一步研究的重点是使大量的数据计算节点协同工作,方便地进行分布式管理和监控,从而快速地发现和恢复系统故障。

参考文献:

[1] 陈康,郑纬民. 云计算:系统实例与研究现状[J]. 软件学报, 2009,20(5):1337-1348.
[2] 杨晓亮,张翀,汤大权,等. 基于对等计算的分布式时空查询处理系统的设计及应用研究[J]. 计算机应用研究, 2011,28(4):1342-1347.
[3] BARROSO L A, DEAN J, HOLZLE U. Web search for a planet; the Google cluster architecture[J]. IEEE Micro,2003,23(2):22-28.
[4] Apache hadoop[EB/OL]. (2010-05-16). <http://hadoop.apache.org/core/>.
[5] DEAN J, GHEMAWAT S. Distributed programming with mapreduce[M]. Sebastopol, CA: O'Reilly Media, Inc., 2007:371-384.
[6] CHANG F, DEAN J, GHEMAWAT S, et al. Bigtable: a distributed storage system for structured data[C]// Proc of the 7th USENIX Symposium on Operating Systems Design and Implementation. 2006: 205-218.
[7] 徐小龙,王汝传,姜波. 一种新的基于 P2P 的电信海量数据分析系统模式[J]. 南京邮电大学学报:自然科学版,2007,27(3):22-27.
[8] VmWare Hyperic HQ[EB/OL]. (2010-09-26). <http://www.hyperic.com/products/vmware-monitoring>.

[2] 王传三. 移动广告的个性化研究[D]. 合肥:中国科学技术大学, 2008.
[3] LEE C C, HSIEH M C. The influence of mobile self-efficacy on attitude towards mobile advertising[C]//Proc of International Conference on New Trends in Information and Service Science(NISS' 09). Beijing; IEEE Press,2009:1231-1236.
[4] 廖安舟,王纯. 广告系统的研究与设计[J]. 计算机系统应用, 2009,18(8):15-18.
[5] 陈明,孙丽丽. 移动电话内容服务系统的个性化推荐[J]. 计算机工程与应用,2007,43(4):4-7.
[6] 谭磊,刘钰峰,李仁发. 一种个性化移动搜索技术的研究[J]. 计算机应用研究,2010,27(9):3403-3405,3414.
[7] CHOI D Y. Personalized local internet in the location-based mobile Web search[J]. Decision Support Systems,2007,43(1):31-45.
[8] 相培峰,杨欣,史元春. 手持设备上基于隐式反馈的网页兴趣块提取[J]. 清华大学学报:自然科学版,2009,49(1):102-105.
[9] 邵秀丽,也聚科,田振雷,等. 用户个性化推荐系统的设计与实现[J]. 计算机工程与设计,2009,30(20):4681-4685.
[10] XU D J, LIAO S S, LI Qiu-dan. Combining empirical experimentation and modeling techniques: a design research approach for personalized mobile advertising applications[J]. Decision Support Systems,2008,44(3):710-724.
[11] 于洪,李转运. 基于遗忘曲线的协同过滤推荐算法[J]. 南京大学学报:自然科学版,2010,46(5):520-527.
[12] 秦光浩,张颖. 基于综合兴趣度的协同过滤推荐算法[J]. 计算机工程,2009,35(17):81-83.