

面向移动数字图书馆的情境敏感型知识推荐研究

胡慕海¹ 蔡淑琴² 谭婷婷²

(武汉纺织大学管理学院 武汉 430074)¹ (华中科技大学管理学院 武汉 430074)²

摘 要 随着移动通信的发展,数字图书馆的各种知识服务更多地在移动终端上实施,但现有的移动数字图书馆对于用户环境、场景等“情境”的感知能力不足,缺乏为用户提供适应当前情境的个性化知识推荐服务。因此提出基于信息熵度量读者的情境敏感性。基于情境敏感度对读者进行相似性比较,并利用这种比较扩展了协同过滤算法。试验结果表明,该方法为移动数字图书馆具备情境敏感的能力,提高知识推荐的准确性提供了一种新的途径,有利于给读者提供既满足情境又符合其个性化需求的知识资源。

关键词 移动数字图书馆,协同过滤,情境,知识推荐,信息熵

中图法分类号 TP391 **文献标识码** A

Mobile Digital Library-oriented Context Sensitive Knowledge Recommendation

HU Mu-hai¹ CAI Shu-qin² TAN Ting-ting²

(Management School, Wuhan Textile University, Wuhan 430074, China)¹

(Management Department, Huazhong University of Science and Technology, Wuhan 430074, China)²

Abstract With the development of mobile communication, the knowledge service of digital library is implemented on mobile terminals, but the existing digital libraries lack of capacity to aware the context such as reader's environment and scenario, so can not offer personalization knowledge recommendation service, which also adapt to the current context and the reader. So a method to measure reader's context sensitivity based on entropy was proposed, and the measure was used to compare two different users based on what the collaborative filtering can be improved. The experiment results show that the improved algorithm can offer mobile digital library a new approach to increase the knowledge recommendation accuracy and its context-aware ability, then provide reader personalized knowledge resource which also adapt to his current context.

Keywords Mobile digital library, Collaborative filtering, Context, Knowledge recommendation, Entropy

1 引言

数字图书馆是用数字技术处理和存储各种图文并茂的文献的图书馆,实质上是一种多媒体制作的分布式信息系统,把各种不同载体、不同地理位置的知识资源用数字技术存储,以跨越区域、面向对象的网络查询和传播的一个大型知识管理系统^[1]。当前,互联网上的信息、知识爆炸式增长,而同时信息、知识过载和组织形式的无序性使得用户难以准确定位自己所需的内容。数字图书馆则能够有序、智能地组织各种知识,因此成为管理网络知识资源的一种很好的方法,能高效、高质量地满足读者的需求。

随着移动通信的发展,移动信息终端已经成为大多数人日常生活中不可缺少的一部分,诸如手机、PDA等已经成为在所有办公和生活用品中最广泛被人们随身携带的工具,移动通信网络覆盖了广大人群所能到达的大多数领域。各种资讯想要更深刻地“嵌入”到人们工作、生活的每个领域,移动终端是已知的最佳载体。移动数字图书馆就是将数字图书馆的

应用领域拓展到移动通信领域的一种新型知识管理和服务平台,它的主要目标是使图书馆的用户能够通过无线通讯网来访问数字图书馆,享受数字图书馆的各项知识服务^[2]。

随着 Web2.0 的发展,诸如 Wiki 等由网络用户共同创造、维护的知识、信息体系产生了越来越多的资源,新的词汇、概念、标签、多媒体等“微内容”层出不穷,因此数字图书馆也面临着如何更有效地管理这些资源,更高效地为读者提供知识服务的问题。近年来,个性化信息推荐在数字图书馆领域得到越来越多的关注^[1]。数字图书馆环境下的个性化推荐技术就是采集描述读者兴趣、爱好等的关联信息,形成抽象描述读者偏好的知识元,支持各种推荐模式的推理运作,使得数字图书馆能主动提供满足读者个性化需求的知识资源。读者无需亲临图书馆,就能够随时随地远程查询、浏览、下载和借阅数字图书馆的电子资源。

移动数字图书馆的主要用户是拥有移动手持设备的读者,读者所处的环境、场景是动态变化的,而研究表明这种动态特性对读者潜在的需求会产生不同程度的影响^[3]。比如读

到稿日期:2010-08-30 返修日期:2010-11-23 本文受国家自然科学基金项目(71071066),国家自然科学基金重点项目(70731001)资助。

胡慕海(1976—),男,博士,讲师,主要研究方向为信息管理、移动商务、电子商务, E-mail: humuhai@163.com; 蔡淑琴(1955—),女,教授,博士生导师,主要研究方向为信息管理、决策支持理论与方法; 谭婷婷(1983—),女,博士生,主要研究方向为众包。

者当前处于工作环境中,可能更多地需要与工作相关的科技类电子期刊;而在周末的时候,读者可能对娱乐、体育等休闲类电子杂志更感兴趣,又比如它和传统数字图书馆不同,读者的浏览设备主要是手机、PDA 等手持设备,在这些设备上,对于资源阅览的支持千差万别,设备终端的不同会影响推荐给读者的知识内容的展示方式、组织方式。研究者提出“情境”(Context)的概念来描述这些动态变化的影响因素,并指出如果缺乏对上述情境因素的感知和应用,会对个性化推荐的服务能力产生消极影响^[4]。但目前移动数字图书馆的知识推荐服务对情境因素的研究和应用尚不够深入,使得当前的移动数字图书馆缺乏有效的情境感知能力,服务的质量无法保证。

鉴于以上情况,本文基于信息熵提出了一种情境敏感性的度量方式,并在协同过滤中应用这种情境敏感度扩展了寻找当前读者的相似者的方式,最后通过试验验证了算法的有效性。基于本文研究结果,移动数字图书馆可以建立情境敏感的个性化知识推荐模式,通过对情境的感知和应用,为读者提供更符合当前情境和其兴趣、爱好的知识资源。

2 相关研究

2.1 移动数字图书馆研究

目前数字图书馆已经在海外蓬勃发展起来,比较有影响力的数字图书馆项目^[5]包括 VD17, Medoc, Bibliotheca Universalis, California. Mary^[6]和 Lippincot^[7]对数字图书馆的起源、趋势和研究现状、未来研究趋势作了综述,认为移动数字图书馆向移动平台迁移是必然的,移动数字图书馆将有良好的发展前景。Eunice 和 Tedjasaputra^[8]设计、开发和评估了应用于教育系统的移动 ICT 数字图书馆平台,分析和探讨了资源的组织和检索问题。Shukla 和 Gugnani^[5]将目前的数字图书馆研究分为用户为中心、内容为中心和系统架构为中心 3 种研究类型,比较了它们在资源组织过程中的差异。近年来,国内也开始重视移动数字图书馆的建设,如中国高等教育文献保障系统和中科院华建集团合作开发了移动数字图书馆系统^[2]、清华大学中心图书馆的移动数字图书馆系统^[5]。张成昱等^[9]和丰江帆等^[10]探讨了移动数字图书馆构建中的一般性问题,解决对策、原则。孙一钢和王安生^[11]对移动数字图书馆一般的体系结构和表现形式进行了研究。总体来讲,目前国内外研究中更多的研究侧重在移动数字图书馆设计、开发和运行过程中的一般性问题探讨,重视资源的组织、搜索,但从知识推荐的角度分析移动数字图书馆的功能和实现途径的研究较少,同时对情境因素影响的度量、整合和应用研究也有待进一步深入。

2.2 协同过滤

协同过滤^[4]是应用最为广泛的一种知识推荐方式。在日常生活中,人们往往会根据亲朋好友的推荐来做出一些选择,协同过滤就是将这一思想运用到网络信息服务中,基于其他用户对某一信息的评价来向当前用户推荐资源。即找到与当前用户 u 相似(如偏好、兴趣、习惯等相似)的其他用户 r ,计算对象 r 对资源项目 i 的效用值 $R_{r,i}$,利用 $R_{r,i}$ 对所有 i 进行排序或者加权等操作,找到最适合 u 的 i ^[4]。协同过滤的一个优点是推荐行为是主动产生的,即用户所获得的推荐是系统从用户购买或浏览等行为中隐式获得的,不需要用户主动去查找适合自己兴趣的推荐对象。另一个优点是对推荐对象没

有特殊要求,不需要了解推荐对象自身的特征信息,因此能够处理非结构化的复杂对象。同时,由于是基于用户间的关联性进行推理,因此与社会网络研究有交叉点,有丰富的研究基础和广阔的前景,故在数字图书馆的知识推荐中受到重视,有着较多应用。如何定义用户相似性以及如何选择用户近邻集,是协同过滤算法研究的重点^[4]。采用的思路可以分为启发式(如基于余弦相似度、评价相关性等启发算子)和基于模型的方式(如 SVM、神经网络、贝叶斯网络等)^[4]。现有的用户间相似度度量主要基于用户的评价相关性或用户自身统计学特征(年纪、职业、性别等)的相似性,一些研究也考虑了情境因素^[12,13],但出发点都是将情境作为用户自身特征属性的扩展,然后基于扩展的用户属性集进行用户间的比较。与上述思路不同,本文针对特定类型的项目,定义了用户的情境敏感度,即用户兴趣随特定情境改变的确定程度。这种情境敏感度作为一种新的描述用户特征的方式,能够实现用户间的比较,并基于这种视角扩展传统的协同过滤。

3 情景敏感性度量

读者在不同情境下对移动数字图书馆推荐的相同的知识项目(以下简称项目)很可能产生不同的感受,即对项目的兴趣程度随着情境的变化而发生变化。本文基于信息熵^[14]可以表征读者兴趣随情境发生改变这一事件发生的确定程度,由此提出情境敏感度的概念。情境敏感度的获取分 3 个步骤:推荐服务的情境化、情境化推荐服务评价以及基于信息熵的情境敏感度计算,分别介绍如下。

3.1 推荐服务的情景化

情景是任何能够将实体特征化的信息,实体指人、地点或者与用户、系统交互关联的任何对象^[3]。情景具有关联的特征属性集合,每个属性具有可选的一组值,基于具体的环境或场景能够确定具体的值^[4]。不同特征属性值的集合区分了不同的情景,比如 Week 和 Location 定义为电子杂志推荐服务的两个情景属性,前者的属性状态集为 {weekend, weekday},后者属性状态集为 {home, office}。考虑到读者一般在周末和在家里喜欢浏览体育、娱乐等休闲资讯,在工作日和办公室则需要浏览财经等工作相关的资讯,可以将触发体育、娱乐类电子杂志推荐的情景定义为 $C_{sport} = \{\text{weekend, home}\}$,触财经类电子杂志推荐的情景定义为 $C_{finance} = \{\text{weekday, office}\}$ 。

基于上述分析,假定情景限定为读者接受推荐服务和进行反馈评价时的情景,形式化 $C^S = \{C_1^S, C_2^S, \dots, C_i^S, \dots, C_J^S\}$, $i \in \{1, 2, 3, \dots, J\}$ 。其中 S 是所推荐项目的领域类型,如体育类、娱乐类或财经类, C^S 是 S 的情景属性集。 $C_i^S = \{c_{i1}, c_{i2}, c_{i3}, \dots, c_{im}\}$ 是 C^S 第 i 个情景属性的值域; C_{ij} 是第 i 个情景属性的第 j 个属性的值,该情景属性共有 n 个离散的属性值。

3.2 情境化推荐服务评价

在服务过程中,用户所有显性和隐性反馈都能够反映用户的意图,通过分析这些反馈可以推断读者的潜在需求。假定读者反馈是积极的,意味着对服务的评价会提高。如果读者反应消极,则该服务的评价会降低。将反馈信息和用户评价定量化,可以通过对用户访问路径、用户姿态、点击网页的次数、驻留网页的时间等反馈信息进行数据挖掘和融合,建立反馈 fbc 到评价 $rating$ 的映射: $f: fbc \rightarrow rating$ 。本文研究的 $rating$ 只包括两种状态值:积极、消极,由此对于读者 U_{id} 而

言,在第 i 个情景属性 C_i 的第 j 个状态 C_{ij} 下,对 S 的评价定义为该 C_{ij} 中读者积极反馈的次数与该 C_S 下读者所有积极反馈的次数的比:

$$R_{U_{id},C_{ij}}^S = \frac{\text{Count}\{u_r | u_r = "1", \text{Context} = C_{ij}\}}{\text{Count}\{u_r | u_r = "1"\}} \quad (1)$$

式中, C_i 下所有对 S_{ID} 的评价形成集合 $R_{U_{id},G}^S = \{R_{U_{id},G1}^S, R_{U_{id},G2}^S, \dots, R_{U_{id},Gj}^S, \dots, R_{U_{id},Gn}^S\}$, $\sum_{j=1}^n R_{U_{id},Gj}^S = 10 \leq R_{U_{id},Gj}^S \leq 1$, $c_{ij_n} \in c_i, i \in \{1, 2, 3, \dots, J\}, j_i \in \{1, 2, 3, \dots, n\}$ 。

表1是从某读者 u 的日志挖掘出的反馈信息,显示了在特定的 $C_{S_{ID}}$ 下 u 对基于情景属性 Week 所推送的体育类杂志的反应。当 u 浏览杂志的时间超过一定阈值,可认为 u 对所推荐的项目是满意的,读者反馈 u_r 为“1”,否则 u_r 为“0”。将根据式(1):

$$R_{u, \text{weekday}}^{\text{Sport}} = \frac{\text{Count}\{u_r | u_r = "1", \text{Context} = \text{Week.weekday}\}}{\text{Count}\{u_r | u_r = "1"\}} = 1/3$$

$$R_{u, \text{weekend}}^{\text{Sport}} = \frac{\text{Count}\{u_r | u_r = "1", \text{Context} = \text{Week.weekend}\}}{\text{Count}\{u_r | u_r = "1"\}} = 2/3$$

可见,当其他情景属性的状态不发生变化时, u 在周末比在平日更容易对图书馆推荐的体育类电子杂志产生积极响应。

表1 读者 u 的反馈信息

LogID	推荐杂志	情境: Week	用户反馈
001	Sport	weekday	0
002	Sport	weekday	1
003	Sport	weekday	1
004	Sport	weekend	1
005	Sport	weekend	1
006	Sport	weekend	1
007	Sport	weekend	1
008	Sport	weekend	0

3.3 读者的情境敏感度计算

当 $R_{U_{id},G}^S$ 的各元素值彼此接近,如 $|R_{U_{id},G}^{\text{Sport}, \text{weekday}} - R_{U_{id},G}^{\text{Sport}, \text{weekend}}| \leq 3(3 \rightarrow 0^+)$ 表明读者不会因是工作日还是周末,对体育资讯的感兴趣程度产生较大差异,可以理解为读者 U_{id} 接受服务 S 时,不存在对其情景属性 C_i 的偏好;反之,如果 $R_{U_{id},G}^S$ 的各个元素存在较大差异,特别是存在突出的值,则表明用户对 C_i 是存在明显敏感的。对情景的敏感可以理解为一个随机事件,发生与否具有不确定性。香农提出熵可以作为一个随机事件不确定性的量度^[13],由此本文引入信息熵度量特定情境下读者对特定服务产生兴趣的确定程度。即给定 $R_{u,c_{ij}}^s$ 是当前情境属性 c_{ij} 下读者 u 对服务 s 的评价, $I_{c_i} = -\sum_{j=1}^k R_{u,c_{ij}}^s \log_2 R_{u,c_{ij}}^s$ 是对应 C_i 的熵,对于 I_{c_i} , $\exists \text{Min}(I_{c_i}) = 0$, $\exists \text{Max}(I_{c_i}) = -\sum_{j=1}^n \frac{1}{\text{Count}(r_{w}=1)} \log_2 (\frac{1}{\text{Count}(r_{w}=1)})$, 则情境敏感度定义为:

$$P_{c_i} = \left| \frac{\text{Max}(I_{c_i}) - I_{c_i}}{\text{Max}(I_{c_i}) - \text{Min}(I_{c_i})} \right| \quad (2)$$

式中, I_{c_i} 越小,说明 u 在体验服务 S 时,随着情境 c_i 变化,其满意程度更容易发生变化,这里称为对情境 c_i 越敏感; I_{c_i} 越大,表明 u 对 c_i 越不敏感。当 I_{c_i} 达到最大值 $\text{Max}(I_{c_i})$ 时,表明 u 对 c_i 完全不敏感。 P_{c_i} 可视为 I_{c_i} 的规范化,反映出 u 对

s 的兴趣随 c_i 发生改变这一现象发生的确定程度,随着用户对 c_i 情景的敏感程度的增加而递增。最终 u 的情境敏感性描述模型为 $P_c^u = (P_{c_1}^u, P_{c_2}^u, \dots, P_{c_n}^u)$ 。

4 基于情境敏感性的移动数字图书馆协同过滤算法

移动数字图书馆的推荐问题可描述为:给定 $U = (U_1, U_2, \dots, U_i)$, $I = (I_1, I_2, \dots, I_j)$, $C = (C_1, C_2, \dots, C_n)$ 分别是描述读者、电子资源和情境的非空的属性集,则当前情境 c 、目标读者 u 以及资源 i 可视为“推荐空间” $S: U_1 \times U_2 \times \dots \times U_i \times I_1 \times I_2 \times \dots \times I_j \times C_1 \times C_2 \times \dots \times C_n$ 中的一个点 (u, c, i) , f 是资源评价函数,实现 (u, c, i) 到评价空间 R 的评价值 $R_{u,c,i}$ 的映射 $f: S \rightarrow R, R_{u,c,i}$ 是一定范围内全序的非负实数, R 是 $R_{u,c,i}$ 的集合。推荐要研究的问题就是找到使 $R_{u,c,i}$ 最大的那些资源 i' , 当读者处于当前情境 c 下时,将 $I(i' \in I)$ 提供给 u , 即 $\forall u \in U, \forall c \in C, i' = \arg\max_{(i \in I)} f(i, c, u)$ 。

4.1 基于情境敏感性的用户相似度量

基于传统协同过滤的思想,设读者 r 和读者 u 共同评分过的项目集合用 I_{ru} 表示,则 r 和 u 之间的相似性 $\text{sim}(r, u)_1$ 可以通过皮尔森相关系数来度量。皮尔森相关系数用于衡量两个变量之间的线性关系:

$$\text{sim}(r, u)_1 = \frac{\sum_{i \in I_{ru}} (R_{r,i} - \bar{R}_r)(R_{u,i} - \bar{R}_u)}{\sqrt{\sum_{i \in I_{ru}} (R_{r,i} - \bar{R}_r)^2} \cdot \sqrt{\sum_{i \in I_{ru}} (R_{u,i} - \bar{R}_u)^2}}$$

式中, $R_{r,i}$ 表示 r 对项目 i 的评分, $R_{u,i}$ 表示 u 对项目 i 的评分, $\bar{R}_r$ 表示 r 对项目评分的平均值, $\bar{R}_u$ 表示 u 对项目评分的平均值。

读者对特定资源类型的情境敏感性可以作为一种新的表征读者特征的因素,从情境对兴趣爱好影响程度的视角表示读者间的差异。对于读者 u 而言,给定资源项目的特征属性集 $I = (I_1, I_2, \dots, I_j)$, 每个属性 I_i 对应推荐项目 i 的领域类型,关联的情境属性集为 $C^I = (C_1, C_2, \dots, C_i, \dots, C_n)$ 。 $P_c^u = (P_{c_1}^u, P_{c_2}^u, \dots, P_{c_n}^u)$ 描述其对 I_i 的兴趣相对于关联情境 C^I 的敏感度的集合,则基于情境敏感度的 u, r 相似性比较,可通过皮尔森相关性系数计算如下:

$$\text{Sim}(r, u)_I = (1/j) \sum_{I_S \in I} \left[\sum_{C_i \in C_w} (p_{c_i}^r - \bar{p}_{c_i}^r)(p_{c_i}^u - \bar{p}_{c_i}^u) \right] / \left[\sum_{C_i \in C_w} (p_{c_i}^r - \bar{p}_{c_i}^r)^2 \sum_{C_i \in C_w} (p_{c_i}^u - \bar{p}_{c_i}^u)^2 \right]^{1/2}$$

式中, $C_w (C_w \subseteq C)$ 表示读者 r, u 共同给出评价时关联的情境属性集。 $p_{c_i}^r, p_{c_i}^u$ 分别表示 r, u 对 c_i 的敏感度, $\bar{p}_{c_i}^r, \bar{p}_{c_i}^u$ 分别表示 r, u 对 C^I 所有情境属性的敏感度均值。

用户 u, r 间总体相似度为:

$$\text{Sim}(r, u) = \alpha \text{Sim}(r, u)_1 + (1 - \alpha) \text{Sim}(r, u)_2 \quad (3)$$

式中, α 是权重因子,用来调节情境敏感性在读者相似度量中的重要程度, $0 \leq \alpha \leq 0$ 。

4.2 移动数字图书馆推荐项目的评价预测

一旦基于相似性测量确定出 N 个最相似的读者,这 N 个读者就能形成目标读者 u 的近邻集 NBS , 下一步就是查看相似读者的评分,然后通过计算得到预测。皮尔森相关系数属于启发式推荐算法^[4], 这种类型推荐算法对未知资源的评价预测一般基于早期用户评价过的其他项目,也就是说用户 u 对项目 i 评价的预测值 $R_{u,i}$ 常常是其他用户 r (如近邻集

NBS中的 N 个用户 $r, r \in NBS$ 对 i 的评价的聚合 $R_{u,i} = \text{aggr}(R_{r,u})$ 。常用的评价聚合函数如下^[15]:

$$\begin{aligned} (A) R_{u,i} &= (1/N) \sum_{r \in NBS} R_{r,i} \\ (B) R_{u,i} &= (1/k) \sum_{r \in NBS} \{ \text{sim}(r,u) \times R_{r,i} \} \\ (C) R_{u,i} &= \bar{R}_u + k \sum_{r \in NBS} \text{sim}(r,u) \times (R_{r,i} - \bar{R}_r) \end{aligned}$$

总体来说,函数(C)的预测评价精确度更高^[15],因此本文采用该函数计算聚合评价价值 $R_{u,i}$ 。正态化因子 $k = 1/(\sum_{r \in NBS} \text{sim}(r,u))$,预测评价函数为:

$$R_{u,i} = \bar{R}_u + \frac{\sum_{r \in NBS} \text{sim}(r,u) \times (R_{r,i} - \bar{R}_r)}{\sum_{r \in NBS} \text{sim}(r,u)} \quad (4)$$

式中, $\bar{R}_u$ 表示 u 对所有项目评分的平均值, $\bar{R}_r$ 表示相似读者 r 对所有项目评分的平均值, $\text{sim}(r,u)$ 是根据式(3)得到的 r 和 u 的相似度。根据对项目 i 的预测评分,选取其中的 $top-N$ 的项目推荐给读者 u ,完成推荐。

5 试验及分析

5.1 数据准备

假定移动数字图书馆推荐的是电影知识,试验在 MovieLens(<http://movielens.umn.edu>)的电影评分数据集上展开。采用的数据文件包含了943位用户对1682部电影的10万次评价。数据类型包括用户对服务商提供的电影推荐资讯的反馈评分(1~5级)、影片类型、用户性别、职业和年纪、用户对项目打分时的时间戳 TimeStamp、用户的邮政区号和国籍。由于时间戳反映某电影被评价时的 UTC 时刻,结合用户的邮政区号和国籍,可确定用户当地的时区划分,进而挖掘出用户提供评价反馈的当地时刻,将这些时刻数据离散化,可确定出服务关联的两种情境属性:当日时间段(上午、下午、晚上),周(工作日、周末)。构建了情境属性集,选择电影类型为喜剧。

5.2 度量

假定喜剧类型项目总数为 K , N 名用户对其的总评价数为 M ,稀疏度定义为 $\text{Sparsity} = M/(NK)$ 。

平均绝对偏差(MAE)方法易于理解,可以直观地度量推荐质量,是最常用的一种评估推荐质量的指标。试验采用 MAE 预测用户评分与实际用户评分之间的偏好, $\text{MAE} = (\sum_{i=1}^N |R_{u,i} - \hat{R}_{u,i}|)/N$ 。其中, $R_{u,i}$ 表示用户 u 对项目 i 的实际评分, $\hat{R}_{u,i}$ 表示根据推荐算法预测的用户 u 对项目 i 的评分, N 表示预测评分的个数。MAE 越小,推荐准确度越高。

5.3 试验过程及结果比较

基于稀疏度考察,选择出500名用户, $\text{Sparsity} = 21.5\%$,保证了可用数据集的密度。500名用户分为两组:选择300名为训练集(训练集大小依次变化为100,200,300),剩余200名用户作为测试集。根据上述式(1)~式(4),预测用户对给定的未知喜剧类电影资讯项目的评分,通过 MAE 评估预测准确度。试验调节参数为权重因子 α (参见式(3))和用户近邻数 k (参见式(4))。

试验1设置 $k = 25$,考察本算法推荐质量和情境影响程度间的关联关系。从图1看出,权重因子从0~1以0.2的间隔取值,当权重因子 $\alpha = 0.6$ 时,本算法的准确度达到最佳,因此在试验2中设置前置参数 $\alpha = 0.6$,比较本算法和以皮尔森

相关系数作为用户相似度量依据的传统协同过滤算法(即 $\alpha = 1$ 时)的推荐准确度,纵坐标基于3个训练集对应的 MAE 的均值得到。图2表明,本算法在不同的近邻数下均能表现出比不考虑情境影响的传统协同过滤算法更优的推荐质量, $k = 25$ 时推荐质量最高。

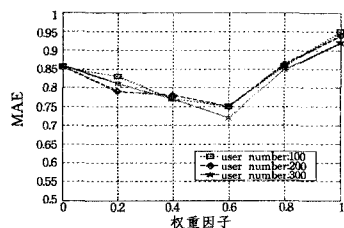


图1 不同权重因子下的 MAE 比较

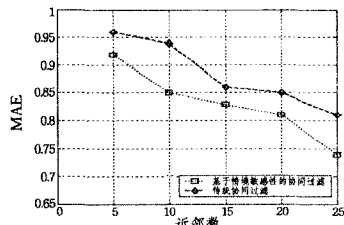


图2 不同近邻数下的 MAE 比较

由此可知,通过选择合适的权重因子 α ,本文提出的基于情境敏感性的协同推荐算法能更好地保证推荐质量,并通过权重因子的调节,使得用户相似度量的方式具有灵活性,以适应动态变化的情境对当前服务的影响程度的变化。总体来讲,试验验证本文提出的新的扩展用户间相似度量的方式,相比未扩展的传统用户相似度量方式,更有利于推荐准确性的提升,因此,以本文提出的情境敏感度为基础,整合应用到一些更复杂的、以用户相似度量为核心的、以协同过滤为基本思路的推荐算法中,以利于移动数字图书馆系统灵活整合、应用情境信息,提供给用户既适合其情境又满足其个性化需求的资讯服务。

结束语 随着移动通信的发展,数字图书馆的各种服务更多地移动终端上实施,但现有的移动数字图书馆对于读者环境、场景、移动终端内存、操作平台等“情境”因素的感知能力不足。而在动态变化的移动环境中,这些情境因素很可能影响着读者潜在的信息需求,导致移动数字图书馆不能为读者提供适应当前情境的个性化推荐服务。因此,本文基于信息熵度量了读者的这种情境敏感性,并在协同过滤中应用情境敏感度改进寻找当前读者的近邻读者的算子,试验验证了算法的有效性。基于本文的研究结果,移动数字图书馆可以建立情境敏感的资讯推荐模式,通过对情境的感知和应用,可为读者提供更符合当前环境、场景和读者兴趣、爱好的资讯服务。

读者的情境敏感度是基于读者在不同情境下的反馈信息推理得到的。但如果反馈信息很少,即数据稀疏性问题严重时,就无法有效地评估用户的情境敏感性。因此在移动数字图书馆为读者提供初始服务时,如何在反馈信息稀疏甚至没有的情况下评估读者的情境敏感性,需要进一步研究。同时,权重因子 α 决定了情境敏感性对推荐准确度的影响程度,是否可以基于反馈学习的方式确定 α 的值也值得研究。本算法

(下转第105页)

然后根据给定的 Ad-hoc 网络 5 个可信度影响因子的测试集数据,使用不确定模拟技术计算出 $E[\bar{b}_i](i=1,2,3,4)$, 得到该 Ad-hoc 网络的可信度评判结果为 (0.1531, 0.578, 0.2688, 0)。4 个数中 0.578 最大,可见该 Ad-hoc 网络的可信度等级为 II 级,与实际情况基本一致,但结果以数据的形式量化表示,更加真实、可靠。

结束语 本文提出了一种基于不确定理论的 Ad-hoc 网络可信度评判方法,首先采用不确定变量表示各个评判因子的权重系数,增强了各因素权重的合理性;然后通过单因素评价模型对各因素进行评价;最后利用不确定综合评判模型确定整个 Ad-hoc 网络的可信度标准等级,提高了评判的准确性。在今后的工作中,关于 Ad-hoc 网络可信性的定义、信任度影响因子的确定还需要进一步的研究和探讨,在实际应用中不确定变量的选取和权重系数的确定方法也都需要进一步完善,从而提出更加科学的模型和评价方法。

参考文献

- [1] Xu Li, Lin Zhi-wei, Ayong Y E. Analysis and counter measure of selfish node problem in mobile Ad hoc network[C]//Proc of the 10th Computer Supported Cooperative Work in Design International Conference, 2006; 1-4
- [2] Li Jing-tao, Jing Yi-nan, Xiao Xiao-chun, et al. A trust model based on similarity-weighted recommendation for P2P environments [J]. Journal of Sothware, 2007, 18(1): 157-167
- [3] Blaze M, Lacy F J. Decentralized trust management[C]//Proceedings of the 17th Symposium on Security and Privacy. IEEE Computer Society Press, 1996; 164-173
- [4] Blaze M, Feigenbaum J, Keromitis D A. Keynote: trust management for public-key infrastructure[C]//1998 security protocols International Workshop. Springer-Verlag, 1999; 59-63
- [5] Josang A. A logic for uncertain probabilities[J]. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 2001, 9(3): 279-211
- [6] Josang A, Kaapskog S J. A metric for trusted systems[M]. Global IT Security. Wien: Austrian Computer Society, 1998; 541-549
- [7] 唐文, 陈钟. 基于模糊集合理论的主观信任管理模型研究[J]. 软件学报, 2003, 14(8): 1401-1408
- [8] 张艳群, 张辰. 基于模糊理论的信任度评估模型[J]. 计算机工程与设计, 2007, 28(3): 532-533
- [9] 王小斌, 矫立峰, 孙延明, 等. 基于模糊模拟的主观信任评价模型研究[J]. 计算机工程与应用, 2008, 44(16): 102-104
- [10] 陈涛, 周学广, 张焕国. 基于模糊理论的 Ad-hoc 网络可信性研究[J]. 计算机工程与设计, 2010, 31(5): 958-960
- [11] Liu B. Theory and Practice of Uncertain Programming(2nd ed) [M]. Berlin: Springer-Verlag, 2009
- [12] 曾文才. 模糊综合评判失效的分析与对策[J]. 系统工程理论方法应用, 1995, 4(2): 53-59
- [13] 宋世德, 周静芋, 袁志发. 解决模糊综合评判中失效与失效问题的方法[J]. 西北农业大学报, 1996, 24(01): 75-78
- [14] 林衍, 顾恒岳, 盛湘渝, 等. 模糊综合评判误判原因的探讨[J]. 系统工程理论方法应用, 1997, 6(2): 67-70
- [15] 王道勇. 模糊综合评判的失效与消除[J]. 系统工程理论方法应用, 1998, 7(2): 6
- [16] 付强, 王立, 宋艳纷. 对模糊综合评判决策模型的认识和改进[J]. 农机化研究, 2002(2): 27-29
- [17] 谢华, 朱学峰. 模糊综合评判中几种数学模型的比较[J]. 广东自动化与信息工程, 2001, 22(4): 5-7
- [18] 张兴芳, 管恩瑞, 孟广武, 等. 区间值模糊综合评判及其应用[J]. 系统工程理论方法应用, 2001(12): 80-85

(上接第 95 页)

在大数量集下的运算效率和推荐质量也需要进一步评估。

参考文献

- [1] 王代琳, 刘亚秋, 王真谛. 基于平均差异度的数字图书馆个性化推荐算法研究[J]. 图书情报工作, 2009, 53(11): 119-122
- [2] 丁会平. 移动数字图书馆内容管理系统的设计与实现[D]. 南京: 南京理工大学, 2004
- [3] Prahalad C K. Beyond CRM; C. K. Prahalad Predicts Customer-Context Is the Next Big Thing[M]. New York: AMACOM, 2004
- [4] Adomavicius G, Tuzhilin A. Toward the Next Generation of Recommendation Systems: A Survey of the State-of-the-Art and Possible Extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749
- [5] Shukla V N, Gujani K K A V. Digital Library: Language Centered Research, Test Beds and Applications[EB/OL]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.787&rep=rep1&type=pdf>, 2010-08-27
- [6] Mary A K. Book Review: Open Access to Knowledge and Information; Scholarly Literature and Digital Library Initiatives-the South Asian Scenario[J]. Library Management, 2008, 29(8): 811-812
- [7] Lippincott K A. Mobile future for academic libraries [J]. Reference Services Review, 2010, 38(2): 205-213
- [8] Eunice S A, Tedjasaputra A D. Exploring potentials and challenges of mobile ICT for learning in Finland and Indonesia[J]. International Journal of Mobile Learning and Organisation, 2008, 2(2): 103-118
- [9] 张成昱, 方玮, 周虹. 关于移动数字图书馆建设的几点思考[J]. 图书馆建设, 2009, 9(1): 68-71
- [10] 丰江帆, 朱冠宇. 基于 TD-SCDMA 的移动数字图书馆研究[J]. 图书馆学研究, 2009(12): 28-31
- [11] 孙一钢, 王安生. 移动数字图书馆集成服务的体系结构与表现形式[J]. 国家图书馆学报, 2008(3): 51-54
- [12] Sankaranarayanan R, Adomavicius G. Incorporating contextual information in recommendation systems using a multidimensional approach[J]. ACM Transactions on Information Systems, 2005, 11(23): 103-145
- [13] Park H S, Yoo J O, et al. A context-aware music recommendation system using fuzzy bayesian networks with utility theory [C]//International Conference on Fuzzy Systems and Knowledge Discovery. 2006; 970-979
- [14] 傅祖芸. 信息论——基础理论与应用[M]. 北京: 电子工业出版社, 2001
- [15] Delgado J, Ishii N. Memory-based Weighted-majority Prediction for Recommender Systems[C]//Proc. ACM SIGIR '99 Workshop Recommender Systems: Algorithms and Evaluation. 1999; 671-674