

普适环境下的群体用户建模机制研究

武瑞娟 周兴社 王海鹏 王柱

(西北工业大学计算机学院 西安 710072)

摘要 普适环境下经常会出现多个用户一起活动的情形,群体用户建模技术可以有效支持针对群体用户的个性化服务。对普适环境下的群体建模需求进行了分析,并讨论了个体兴趣模型及群体兴趣模型的定义和形式化描述。重点给出了一种基于痛苦避免、兴趣满足和公平三者均衡下的兴趣融合算法,最后在一种应用场景下给出实验,并对实验结果进行了分析和讨论。

关键词 普适计算,群体建模,兴趣融合,个性化服务

中图分类号 TP391 **文献标识码** A

Research on Group Modeling in the Pervasive Environment

WU Rui-juan ZHOU Xing-she WANG Hai-peng WANG Zhu

(Department of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract People usually tend to take activities in groups; the group modeling technology can support personality services towards group people effectively. This paper analysed the requirement of group modeling in the pervasive environment, then discussed the definition and formalization of the individual and group preference model. Emphatically, this paper presented an algorithm which is based on the least misery, the pleasure satisfying, and the fairness. To proof our algorithm, we designed an experiment, and discussed the result of it.

Keywords Pervasive computing, Group modeling, Interest merging algorithm, Personalized service

1 引言

在普适环境中,经常会出现多个用户一起活动的情形。如在智能博物馆多个游客组成一个团体,一起参观游览。文献[1]中提到,博物馆的游客在游览方式上,只有5%的游客是单独来参观,而大约20%的游客是和朋友一起来,30%的游客同家人一起来,45%的游客以雇有导游的团体形式游览。因此,研究普适环境下群体用户的特征,以及如何对群体用户进行建模以支持个性化服务,具有重要意义。

已有的关于群体建模的研究多集中于群体推荐系统。MusicFX系统对异构群体进行背景音乐的推荐^[2];INTRUGUE对群体进行旅游景点的推荐,该项目提出一种群体划分思想^[3],将异构群体划分为若干同构子群,得到子群的推荐列表后再进行融合。这些典型的群体推荐系统或者通过简单计算群体成员的总体评分值,或者根据群体成员的参数选择和待推荐项目之间的关联度,直接给出推荐项目列表,没有对群体兴趣的建模做详细研究。

而在普适环境下,环境上下文是一个重要因素。如在博物馆环境中,对展品的推荐要考虑的不仅仅是用户需求和展品之间的关联,还要考虑包括诸如展品前的人数、展品开发时

间等信息,以及游客游览时间约束等信息。因此,对群体兴趣进行建模,从而得到群体整体的兴趣,而不是仅依靠群体各成员的参数选择,可以更好地进行推荐或路线规划等个性化服务。

本文首先对普适环境下的群体用户建模需求进行分析,并阐述了个体兴趣模型及群体兴趣模型。然后重点提出一种基于均衡策略的群体兴趣融合算法,最后通过实验对此算法进行验证和评价。

2 群体用户建模

在现代化信息服务环境下,用户的信息需求日趋多元化和个性化。随着社会信息化进程的加速,个性化的生产和服务开始流行。普适个性化服务以人为中心,通过收集和分析用户信息来学习用户的兴趣和行为,为不同用户提供不同的服务,以满足不同的需求。

当前的个性化服务多集中于针对个体的服务。在普适环境中,很多情况下人们是以群体为单位进行活动,如在智能博物馆中,游客经常以团体形式进行游览。群体用户建模技术通过分析获取群体兴趣,可以为针对群体的个性化服务提供数据支持。

到稿日期:2008-06-24 本文受国家863高技术研究发展计划基金项目(2006AA01Z198),教育部博士点基金新教师课题(20070699014),西北工业大学科技创新基金项目(2006CR13),西北工业大学研究生创业种子基金(200858)资助。

武瑞娟(1984-),女,硕士研究生,主要研究方向为普适计算、用户建模,E-mail: wuruijuan@gmail.com;周兴社(1955-),男,教授,博士生导师,主要研究方向为嵌入式计算、分布式计算、传感器网络;王海鹏(1975-),男,博士,讲师,主要研究方向为普适计算、资源管理;王柱(1983-),男,硕士研究生,主要研究方向为普适计算、个性化服务。

2.1 群体建模需求分析

用户模型领域的研究自 20 世纪 70 年代末期开始后,已有许多应用领域的研究人员开始进行了相关研究,并取得了一定的成果^[4]。在用户建模技术领域里,一些典型的用户建模服务系统,如 BGP-MS^[4],Personis^[5]等,是以个体用户为目标构建用户模型。近年来在群体模型方面的研究多集中于群体推荐领域,如 INTRIGUE^[3]为群体推荐旅游路线,TRAVEL DECISION FORUM^[6]为群体推荐休假地点。这些系统没有把群体兴趣单独列出来考虑,而是根据群体成员的参数选择和待推荐项目之间的关联度,直接给出推荐项目列表。

对普适环境下群体的兴趣模型进行研究,可以有效地对群体兴趣、服务内容、环境上下文信息进行综合考虑,从而更好地为群体提供个性化服务,提高群体的整体满意度。

2.2 兴趣模型

兴趣模型是进行群体建模研究的基础,以下从模型的定义和形式化描述两方面对兴趣模型进行具体的阐述。

2.2.1 模型定义

群体兴趣是群体个性化服务中重要的数据支撑,群体建模研究如何对群体中成员的个体兴趣进行融合进而获得群体的整体兴趣。定义 1 和定义 2 说明了个体兴趣模型和群体兴趣模型的含义。

定义 1(个体兴趣模型, Individual Preference, IP) 描述群体中某个成员的个人兴趣,是存储个体兴趣的知识源。兴趣可以抽象为加权关键词列表。

定义 2(群体兴趣模型, Group Preference, GP) 描述一个群体的整体兴趣所在,即群体成员一定程度上可接受的兴趣关键词,可以抽象为一个兴趣关键词列表。

群体建模系统模型见图 1,其中 IP 指个体兴趣模型,GP 指群体兴趣模型,决策器是决定如何由个体兴趣融合出群体兴趣的组件。决策器的工作机制是本文重要研究的对象,关于它更详细的描述见第 3 节。

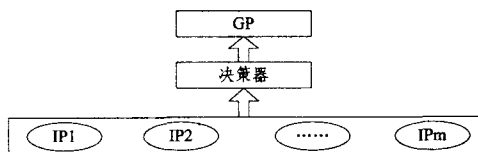


图 1 群体建模系统模型

2.2.2 模型形式化描述

对个体兴趣模型和群体兴趣模型进行形式化描述,是做下一步群体兴趣融合的基础。

在个体兴趣模型的表示上,本文采用加权关键词向量来表示,即某个兴趣关键词及其权重表示一个兴趣,多个兴趣组成一个向量代表个体兴趣。这样可以用一个二元组的集合表示个体兴趣,即

$$IP = \{(p_1, w_1), (p_2, w_2), \dots, (p_n, w_n)\} \quad (1)$$

p_1 到 p_n 代表 n 个兴趣关键词; w_1 到 w_n 是针对不同兴趣关键词的权重,值越大表示兴趣度越大。 p_i 在不同的应用中有不同的取值,例如在博物馆应用中,可以指代“陶瓷类型”、“书画类型”、“秦汉时期展品”等关键词。

在某个特定应用中,可以把 p_1 到 p_n 的取值和顺序都设置为固定,那么就可以这样简化兴趣模型:

$$IP = (w_1, w_2, \dots, w_n) \quad (2)$$

将个体兴趣模型简化为一个 n 维的向量。在此基础上,在两个层面上表示群体兴趣。第一层面表示群体一般兴趣,即用向量集合表示群体中所有成员的兴趣。

$$GP1 = \{(w_{11}, w_{12}, \dots, w_{1n}), (w_{21}, w_{22}, \dots, w_{2n}), \dots, (w_{m1}, w_{m2}, \dots, w_{mn})\} \quad (3)$$

w_{ij} 指代群体中成员 i 对兴趣关键词 p_j 的权重。

第二层面表示群体的整体兴趣,用一个兴趣关键词列表来表示。

$$GP2 = (P_1, P_2, \dots, P_k) \quad (4)$$

式(4)用 P_1 到 P_k 表示 k 个群体整体兴趣,即群体成员一定程度上可接受的兴趣关键词。 P_k 从式(1)中的 p_1 到 p_n 中取值,不同之处在于式(1)中的 p_1 到 p_n 值和顺序固定,而式(4)中的 P_1 到 P_k 顺序并不固定,关于其次序的排列,体现了一种群体意识,具体到实际应用中即成员更愿意接受的观看顺序, P_1 到 P_k 的次序会影响成员的满意度,这将在下面通过实验说明。

3 群体用户兴趣融合机制

3.1 群体描述

普适环境中的用户群体是可以分类的,不同类型的用户群体有着不同的目标。且由于不同群体在成员组成及成员关系上的不同,在个体兴趣向群体兴趣的融合过程中,对所取融合机制的选择应该有不同的侧重。本节所阐述的兴趣融合机制针对普通群体,我们假设此群体有如下特征:

- 1) 群体为固定群体,即群体的人数及组成无动态变化;
- 2) 成员之间平等,无优先级。

3.2 基于均衡思想的群体兴趣融合算法

在给出群体一般兴趣模型 GP1 的基础上,分析此兴趣融合算法得出的群体整体兴趣 GP2。GP1 如表 1 所列。

表 1 群体一般兴趣

	P_1	P_2	P_3	P_4	P_5	P_6	P_7
User 1	1.0	2.0	5.3	4.5	4.8	5.2	3.8
User 2	2.0	5.2	4.2	4.6	4.8	5.3	3.7
User 3	5.4	4.9	4.1	4.6	3.5	1.0	4.1

假设兴趣关键词固定为 7 个,群体中有 3 个成员。关键词的权重分为 5 个级别:1 表示特别不喜欢,2 表示不喜欢,3 表示感觉一般,4 表示喜欢,5 表示特别喜欢。

针对上小节所述目标群体,先进行了一个问卷调查,目的在于研究人们倾向于选择哪种策略进行兴趣融合。该调查以表 1 为基础,以博物馆为应用背景,让实验者为这 3 个人选出几个兴趣关键词组成群体兴趣,作为群体观看展品的基础。

实验结果显示,在 18 个实验者中,有 15 个人使用了痛苦避免思想^[7],他(她)们都选择了首先排除 P_1, P_2 等含低权值的兴趣。不同之处在于排列顺序上,部分人用了权重和从大到小排列,部分人用了从小到大排列。痛苦避免思想的核心在于,尽量使群体中的任何成员都不遭受痛苦,一般的做法是或者将含低权值的兴趣关键词排除,或者按关键词之上的最小权值排序,例如在表 1 中,7 个兴趣关键词之上的最小权值依次是:1.0, 2.0, 4.1, 4.5, 3.5, 1.0, 3.7, 按从大到小排序,则是 $P_4, P_3, P_7, P_5, P_2, P_1, P_6$ 。两个人使用了公平策略^[7],也可称为轮流投票策略,他们认为需要保证每个人最好有一个

自己最喜欢的,因此分别取3个成员中权值最大的关键词,即 P_3, P_6, P_1 。另外有1个人认为应使用重要人物策略^[7],即以群体中某个用户为中心决定群体的选择。重要任务策略适用于有特殊人物或重要人物的群体。

以上调查中实验者虽然是作为局外人对群体进行推荐,但实验者的选择体现了实验者更能接受何种策略。综上可看出,痛苦避免的思想得到了绝大多数人的接纳,因此在设计兴趣融合算法时,应该将这种思想作为一个重要因素加以考虑。但本文认为,不能将该思想作为唯一的决定因素,因为首先,若一个异构群体成员数量较多,兴趣冲突表现得可能更明显,若划定一个阈值,将含低权值的兴趣全部排除在外,可能会排除大多数兴趣关键词,最坏情况下可能所有关键词都因含低权值而被排除;其次,可能存在这样的情形,某个关键词大多数成员的权重都很高,而只有少数人的权重低,如表1中的 P_6 ,而痛苦避免的思想可能会过滤掉这些含较多高权值的关键词。

综合考虑上述内容,本文提出一种基于均衡思想的兴趣融合算法。该算法在以下3点间寻找一种均衡,由群体成员的个体兴趣融合得出群体整体兴趣。

- 1)痛苦避免:尽量不选择含低权值的兴趣关键词;
- 2)兴趣满足:对于某些关键词,如果多数人的权值较高,可考虑取这个关键词;
- 3)公平:对2)造成的不良后果的补救。如果选择了一个A特别不感兴趣的关键词,应再加一个A很感兴趣的关键词以降低他的不适度/痛苦(Misery)。

算法描述如下:

输入:群体一般兴趣,用向量集合 $GP1$ 表示,具体见3.1中的公式(3)。

输出:群体整体兴趣 $GP2$,见公式(4)。

过程描述:

$Sum[] = 0, number[] = 0, P'[k] = 0, k = 1$

//将三个数组初始化为0,其中 Sum 用来存各关键词的权值

//和, $number$ 用来存各关键词之上权值大于5.0的用户的

//个数, P' 用来存无低权值的关键词下标

1)for $j = 1 : n$ {

$misery = 0;$

 for $i = 1 : m$ {

$Sum[j] += w_{ij};$

 if ($w_{ij} < 3.0$) $misery = 1;$

 //将含权值低于阈值3.0的关键词排除

 if ($w_{ij} > 5.0$) $number[j] += 1;$

 //记录权值大于5.0的用户的个数

 }

 if ($misery = 0$) $P'[k++] = j;$

 //关键词之上的所有权值都大于3.0,则加入 P'

2)取 P' 中权值和最大所对应的两项(通过比较 $Sum[P'[k]]$ 的值) $P'[x]$ 和 $P'[y]$,将 $P_{P'[x]}$ 和 $P_{P'[y]}$ 加入 $GP2$ 。

3)取 $number[]$ 中最大值,若 $k > m/2$,将 P_k 加入 $GP2$ 。

 //值大于 $m/2$ 表示有超过一半的人的兴趣权值大于5.0

4)for $i = 1 : m$

 if ($w_{ik} < 3.0$)

 {

 for $j = 1 : n$

 取 w_{ij} 中的最大值所对应的 j ;

 将 P_j 加入 $GP2$;

 }

以表1所描述的群体一般兴趣 $GP1$ 为基础,下面给出一个算法具体实现的例子。假设 $GP2$ 的长度为4,即群体整体兴趣中含4个兴趣关键词。

1)痛苦避免思想,选一个所有成员都不反感的关键词;

 排除含 $weight < 3.0$ 的关键词,剩下 $P_3(13.6), P_4(13.7), P_5(13.1), P_7(11.6)$,括号中为各关键词的群体权值之和,取最高的两项 P_4 和 P_3 。

 结果: P_4, P_3

2)兴趣满足,选多数人权值较高的关键词;

 取含 $w_{ij} > 5.0$ 的关键词,其中 P_6 含两个大于5.0的权值,所以取 P_6 。

 结果: P_6

3)如果2)的结果中含 $w_{ij} < 3.0$ 的,针对这些人,选择一个他们兴趣向量中权值最高的关键词;

P_6 中 $w_{36} = 1.0$,取 User3 最大权值关键词 P_1 。

 结果: P_1

 综上,得出群体整体兴趣 $GP2 = (P_4, P_3, P_6, P_1)$ 。

4 实验结果分析与讨论

本次实验以表1为基础,以智能博物馆为应用背景,目的在于调查统计群体游览过程中各成员对上文提出的兴趣融合算法的满意度,以及群体更愿意接受的兴趣排列顺序。

针对上文提出的兴趣融合算法得出的群体整体兴趣 $GP2$,本实验对18个实验者的满意度进行了调查。这18人分为6组,每组3人,分别代指 $User1, User2, User3$,即3个一起去博物馆看展品的朋友。假设 P_1 到 P_7 分别指7种类型或年代的展品,让3人依次观看 P_4, P_3, P_6, P_1 所属类型年代的展品,最后每个人给出一个满意度的评分,取值介于0到1之间。

实验者的评分结果见图2,其中评分在0.7到0.8之间的约有50%,0.8到0.9之间的约有40%,小于0.7的有10%。

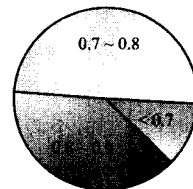


图2 评分分布图

由此可以看出,绝大多数人对兴趣融合得出的结果是较为满意的。少数不满意的实验者给出的原因是,观看特别不喜欢的东西会大大降低自己的满意度。

经过分析实验者的评分结果和所述原因,得出如下结果:

① 绝大多数人希望群体整体兴趣中能够包含他们权值高的(大于5.0)兴趣关键词。

② 评分大于0.8的实验者给出的原因基本都是:因为有最喜欢看到,所以即使有个别特别不感兴趣的,也还可以接受。说明高权值兴趣关键词可以在一定程度上降低低权值关键词带来的不适度。

③ 在评分介于 0.7 到 0.8 的实验者中,半数人认为如果给出的群体整体兴趣中虽不含权值最低的,但也不含权值最高的,不会提升他们的满意度。这说明了权值介于中间的关键词并不能有效提高用户满意度,提高用户满意度还需要依靠用户最感兴趣的东西。

④ 群体整体兴趣,即 GP2 中关键词的排列顺序会影响满意度。约 35% 的人希望按由低到高的兴趣观看,35% 的人希望按由高到低的兴趣观看,20% 的人希望以起伏式如由高到低再到高的兴趣观看,剩下的人持无所谓态度。但是在对于最不感兴趣物品的态度上,大多数人认为最好不要将其放在最开始的位置。

结束语 个性化的主动式信息服务已经成为人们日益关注的一个热点,群体兴趣是进行群体个性化服务的重要数据支撑。本文对个体兴趣模型和群体兴趣模型进行了分析与定义,提出了一种由群体成员的个体兴趣得到群体整体兴趣的融合算法,并通过实验证明了此算法的合理性和有效性。

参考文献

[1] Petrelli, et al. A user centered approach to user modeling // Proceedings of the 7th Int. Conference on User Modeling (UM99).

Springer Wien, New York, 1999: 255-264

[2] McCarthy J F, Anagnost T D. MusicFX: An Arbiter of Group Preferences for Computer Supported Collaborative Workouts [C] // Proceedings of the 1998 Conference on Computer-Supported Cooperative Work, 1998: 363-372

[3] Ardissono L, Goy A, Petrone G, et al. INTRIGUE: Personalized recommendation of tourist attractions for desktop and handset devices [J]. Applied Artificial Intelligence, 2003, 17: 687-714

[4] Kobsa A. Generic User Modeling Systems [J]. User Modeling and User-Adapted Interaction Journal, 2001, 11(1/2): 49-63

[5] Kay J, Kummerfeld B, Lauder P. Personix: A server for user models // AH'02: Proceedings of Adaptive Hypermedia and Adaptive Web-based Systems. London, UK: Springer-Verlag, 2002

[6] Jameson A. More than the sum of its members: challenges for group recommender systems // AVI '04: Proceedings of the Working Conference on Advanced Visual Interfaces. Gallipoli, Italy: ACM Press, 2004

[7] Masthoff J. Group modeling: Selecting a sequence of television items to suit a group of viewers [J]. User Modeling and User-Adapted Interaction, 2004, 14(1): 37-85

(上接第 224 页)

0.4, 最小可信度为 0.3, 图 2 所示为算法的 CPU 时间随库中节点元素的变化情况。其中 □ 为 Bordat 算法, ○ 为 Chein 算法, △ 为批量式建造算法, ◇ 为增量式建造算法。

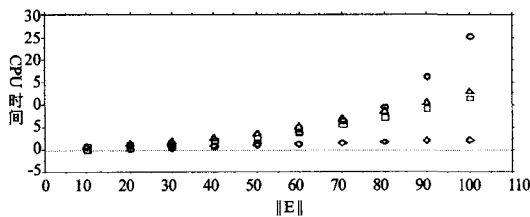


图 2 四种算法的 CPU 时间比较

采用文献[9]所提供的 UCI 知识库验证本算法, UCI 知识库是一个开放的平台, 包括有 17 个大的集合, 13000 个数据, 43 个属性, 791 个规则, 对其做预处理以使它的格式适用于本文对知识的要求。结果表明, 采用基于广义概念格的规则生成, 从规则集中得到了 298 个假设, 由于原有的算法不对规则知识进行处理, 因此这一点上没有可比性。

结束语 本文在已经提出的概念格生成算法的基础上, 给出了广义概念格结构的增量式生成算法。由于增加了支持度和实时的可信度信息, 大大提高了知识发现的效率。同时通过构建树结构, 缩小产生子格节点的范围, 这样就无需对每个格节点都测试其成为产生子的可能性。这种算法在简单修改后均可以有效适用于通常意义下的概念格的生成。

参考文献

[1] Cole R, Eklund P. Scalability in formal concept analysis [J]. Computational Intelligence, 1999, 15 (1): 11-27

[2] Martin P, Eklund P W. Knowledge Retrieval and the World Wide Web [J]. IEEE Intelligent Systems, 2000, 15(3): 18-25

[3] Godin R, Mineau G, Missaoui R, et al. Applying concept formation methods to software reuse [J]. International Journal of Knowledge Engineering and Software Engineering, 1995, 5(1): 119-142

[4] Kent R E, Neuss C. Creating a Web Analysis and Visualization Environment [J]. Computer Networks and ISDN Systems, 1995, 28(1/2): 109-117

[5] Corbett D, Burrow A L. Knowledge reuse in SEED exploiting conceptual graphs [C] // International Conference on Conceptual Graphs. Sydney, 1996: 56-60

[6] Krajić, Stanislav. A Generalized Concept Lattice [J]. Logic Journal of the IGPL, 2005, 13(5): 543-550

[7] Godin R, Missaoui R, Alaoui H. Incremental concept formation algorithms based on Galois (concept) lattices [J]. Computational Intelligence, 1995, 11(2): 246-267

[8] Bordat J-P, Berry A, Sigayret A. A local approach to concept generation [J]. Source Annals of Mathematics and Artificial Intelligence archive, 2007, 49 (1): 117-136

[9] Newman D J, Hettich S, Blake C L, et al. UCI Repository of Machine Learning Databases [DB/OL]. Available at: <http://www.ics.uci.edu/~mllearn/MLRepository.html>

[10] 胡可云, 陆玉昌, 石统一. 粗糙集理论及其应用进展 [J]. 清华大学学报: 自然科学版, 2001, 41(1): 64-68

[11] 谢志鹏, 刘宗田. 概念格与关联规则发现 [J]. 计算机研究与发展, 2000, 37(2): 1415-1421

[12] 王志海, 胡可云, 胡学钢, 等. 概念格上规则提取的一般算法与渐进式算法 [J]. 计算机学报, 1999, 22(1): 66-70