

移动网络服务中基于认知心理学的 用户偏好提取方法

王立才, 孟祥武, 张玉洁

(1. 北京邮电大学计算机学院, 北京 100876; 2. 北京邮电大学智能通信软件与多媒体北京市重点实验室, 北京 100876)

摘 要: 迅速增长的移动网络服务给人们带来沉重的移动信息负担. 移动用户偏好提取方法是缓解“移动信息过载”问题的有效手段. 受加工水平模型和分布式认知理论的启发, 提出一种基于认知心理学的移动用户偏好提取方法. 在移动用户偏好信息结构建模的基础上, 引入服务加工水平认知、有效上下文认知的概念, 并计算其对用户偏好提取的影响, 然后分别提取基于服务加工水平认知和基于有效上下文认知的用户偏好, 最终提取综合的用户偏好. 实验结果表明, 该方法能有效提高移动用户偏好提取精确度, 为用户提供满足个性化需求的移动网络服务.

关键词: 移动网络服务; 用户偏好; 加工水平; 分布式认知; 上下文

中图分类号: TP301 **文献标识码:** A **文章编号:** 0372-2112 (2011) 11-2547-07

A Cognitive Psychology-Based Approach to User Preferences Elicitation for Mobile Network Services

WANG Li-cai, MENG Xiang-wu, ZHANG Yu-jie

(1. School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,

Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: The rapid growth of the number of mobile services provides enormous potential for mobile users in different contexts to find mobile information of interest. Mobile user preferences elicitation has been used as a valid means to ease the “mobile information overload” problem. Inspired by the model of level of processing and the theory of distributed cognition, a cognitive psychology-based approach to user preferences elicitation for mobile network services is proposed. It uses a six-tuple to describe the data structure of user preferences information, analyzes the level of processing of services for users to elicit context-free user preferences, then identifies valid types of contexts as well as their influences on user preferences, and finally elicits converged user preferences. Experimental comparisons of this approach against some baseline methods with a synthetic data set have been conducted, showing improvements in performance.

Key words: mobile network services; user preferences; level of processing; distributed cognition; context

1 引言

随着下一代网络技术的飞速发展, 移动通信网络在与计算机网络逐渐融合的过程中, 对互联网信息服务进行延伸, 为用户提供了比传统电信业务更加丰富多彩(在信息内容、价格、功能特性、服务质量等方面)的移动网络服务. 与此同时, 由于智能移动设备日益普及, 信息资源的获取和推送可以发生在“任何时间、任何地点、以任何方式”, 为用户提供无处不在的移动网络服务已经成为可能. 然而, 服务类型和信息内容的日益增长将逐

渐超出人们所能接受的范围, 加之移动设备的界面显示、终端处理、输入输出等能力有限, 为移动用户带来沉重的“移动信息过载”问题, 导致移动网络资源利用率和用户体验受到严重影响.

个性化移动网络服务研究, 即不断发现和满足移动用户对移动网络服务的个性化需求及其变化, 准确提供和推荐其真正感兴趣的移动服务及其信息内容, 成为学术界和工业界近年来的研究热点. 由于移动用户相比传统互联网用户面临更加复杂、融合、协作的移动网络环境和泛在的移动网络服务提供环境, 只有在充分、准确

理解移动用户在上下文(如时间、位置等)影响下对各种移动网络服务的偏好后,才能实现“服务按需提供、内容按需推荐”的目标.因此,移动用户偏好提取方法成为个性化移动网络服务研究的首要问题.

获取移动用户需求并提供个性化移动网络服务,其实是一种辅助人们进行决策的过程;而认知能力是指人脑加工、储存和提取信息的能力,与用户需求获取和个性化决策存在着密切的联系.本文面向“移动信息过载”问题,提出一种基于认知心理学的用户偏好提取方法(Cognitive Psychology-based Approach to User Preference Elicitation, CPUPE),将加工水平模型和分布式认知理论引入移动用户偏好提取过程,以期提取更加全面、精确的用户偏好.其中,引入加工水平模型的目的在于,突破传统基于“记忆两过程理论”划分“长期偏好”和“短期偏好”的模式,强调了语义层次认知的重要性及其对移动用户偏好的影响;分布式认知理论则强调系统对各种环境因素的认知差异以及这些差异对系统的不同影响程度,而不将各种上下文因素对移动用户偏好的影响程度等权重看待.

2 相关工作

20世纪90年代中期至今,推荐系统^[1]作为一种缓解“信息过载”问题的重要手段,发现用户最感兴趣的项目(如商品、新闻信息^[2]等)并推荐给用户,已广泛应用于电子商务、信息检索等领域.通用推荐系统流程^[1]由用户偏好提取和推荐生成两部分组成.目前,用户偏好技术主要用于计算不同用户对不同项目的选择顺序关系^[3],大都采用“用户-项目”二维定量评分矩阵,而偏好提取过程则基于机器学习与数据挖掘技术^[3](如最近邻算法、神经网络、聚类、决策树、关联规则和贝叶斯网络等).此外,还有一些模型采用定性分析方法提取用户偏好^[4],从逻辑推理或偏序模型角度,提取用户对项目或其属性特征的二元偏序关系.

在泛在的移动网络服务提供环境下,移动用户偏好受上下文因素的影响更加显著^[5].出于“用户在不同上下文实例条件下的需求可能不尽相同”^[6]的考虑,有些研究人员开始将上下文信息引入传统推荐系统和用户偏好提取技术.例如,Adomavicius等^[7]较早提出“上下文感知推荐系统”概念;文献^[8]面向移动应用领域,提出一种利用严格偏序偏好模型定性挖掘上下文用户偏好的方法;文献^[9]基于本体方法获取移动用户偏好并生成上下文感知的移动服务推荐.

此外,也有研究人员结合认知心理学研究用户偏好提取技术.例如,有些研究人员利用“记忆两过程理论”将移动用户偏好分为长期偏好和短期偏好进行建模、获取与修正^[10];文献^[11]考虑到情绪因素在人类认

知决策过程中的重要性,将其引入协同过滤推荐系统;文献^[12]则借鉴“需要层次理论”的思想,提出一种适应用户需求进化的个性化信息服务模型.鉴于认知心理学在人机交互、用户建模、推荐系统等领域的重要性,研究如何从认知心理学的角度提取移动用户偏好,将是个性化移动网络服务研究的重要挑战之一.

3 基于认知心理学的移动用户偏好提取模型视图

3.1 加工水平模型与分布式认知理论

Craik等^[13]在20世纪70年代提出与“记忆两过程理论”相对立的“加工水平模型”,认为没有必要将记忆分为长期记忆和短期记忆,而强调用信息加工深度来解释记忆,即人们对识记材料保持时间的长短和记忆得清楚与否,决定于对材料的加工水平:如果对输入信息的加工只是在粗浅的感觉水平上,那么其痕迹保持的时间便是短暂的;如果对信息的加工包含有语义的性质,那么其痕迹的保持便会是持久的;加工愈深,所需的时间也愈多,以语义水平的加工需时最多,再认知成绩也愈好.Craik等进行了很多实验,表明从信息加工角度用加工水平模型取代“记忆两过程理论”的有效性.

传统认知心理学一直忽略工具、环境等因素的重要地位.20世纪90年代,Hutchins等^[14]提出“分布式认知”的概念,其核心思想是:认知活动不仅存在于人的大脑,也存在于外界的环境、媒介、文化、社会和时间等参与认知活动的全部要素,是认知过程中内外表征的统一.分布式认知理论不仅关注认知在个体的内分布,更强调认知在各种外界要素(如环境、媒介、文化、社会和时间等)的分布,从而阐述了对各种外界要素的分布式、差异化认知在整体认知活动中的重要性,被当作人机交互的新理论基础.

3.2 受加工水平模型与分布式认知理论启发的移动用户偏好提取模型视图

从第2节相关工作中,我们可以看出:(1)现有移动用户偏好提取技术主要基于传统的机器学习、数据挖掘技术,很少考虑用户对移动网络服务的加工水平对用户信息需求的影响,即忽略了不同用户对不同信息内容的认知水平与用户偏好之间的内在关联;(2)有些移动用户偏好提取技术也融入上下文因素,但是它们仅考虑到同种上下文类型(如时间)的不同上下文实例(如上午、下午等)对用户偏好的影响,而忽略了不同种上下文类型对个性化移动网络服务中用户需求分布式、差异化影响,即没有对各种有效上下文及其影响移动用户偏好的优先级顺序进行分布式检测和量化.

针对上述问题,本文受“加工水平模型”和“分布式

认知理论”的启发,提出从研究层次化移动用户需求以及有效上下文信息的分布式认知、差异化计算的角度,逐步精确地提取移动用户偏好:(1)移动用户信息需求与用户对移动网络服务的“加工水平”联系紧密:用户对服务的加工水平越深,获得的信息量就越多,用户偏好越高,再认知成绩也愈好;以语义水平的加工需时最多,如果用户对服务的认知深度达到语义层次,反映用户对该服务的信息需求越强烈,则加强用户偏好;(2)用户偏好不仅依赖于用户对移动网络服务的内在加工水平认知,也在一定程度上受到外界上下文环境的影响;对上下文因素进行分布式认知,分析各种有效上下文类型及其具体实例对用户偏好的差异化影响程度,从而提取基于有效上下文认知的用户偏好;最终提取综合的用户偏好.因此,本文提出如图 1 所示的用户偏好提取模型视图:以用户、移动网络服务、上下文以及用户历史行为为输入,分别挖掘用户的服务加工水平认知和有效上下文认知,然后综合考虑两种因素提取更为精确的用户偏好.

文类型的数目, Z_k 代表该上下文类型的具体实例数.例如,若“时间”上下文 $C_1 = \{\text{上午, 中午, 下午, 晚上}\}$, 则 $c_{11} = \text{上午}, Z_1 = 4$.

定义 5 (移动用户历史行为信息集合) $UBH = \{ubh_{ij} | i \in [1, N], j \in [1, M], ubh_{ij} \in B_1 \times \dots \times B_w\}$, 定义为移动用户历史行为信息集合, 通过“用户-服务”矩阵来描述用户与移动网络服务交互行为; 其中 ubh_{ij} 为单个或者多个非负实数表示的行为变量(如服务使用时长, 浏览/下载行为等)组成的行为向量.

定义 6 (移动用户历史行为上下文信息集合) $UBHC = \{ubhc_{ijr} | i \in [1, N], j \in [1, M], ubhc_{ijr} \in B_1 \times \dots \times B_w, c_r \in C, r \in [1, \prod_{k=1}^L Z_k]\}$, 定义为含上下文信息的移动用户历史行为信息集合, 通过“用户-上下文-服务”多维度向量来描述不同用户在不同上下文环境下使用移动网络服务的情况; 其中, $c_r \in C$ 为多维度组合上下文实例, 即上下文实例 c_{k_q} 的笛卡尔乘积, $ubhc_{ijr}$ 的取值空间与 ubh_{ij} 一致.

定义 7 (移动用户偏好集合) $UP = \{up_{ijr} | i \in [1, N], j \in [1, M], c_r \in C_r \subset C, up_{ijr} \in [a, b]$ 且 a, b 为实数}, 定义为用户在不同上下文环境下对不同服务的偏好集合, up_{ijr} 值越大表示偏好越高; 其中 c_r 表示组合有效上下文实例.

5 基于认知心理学的用户偏好提取方法

在移动用户偏好信息结构建模的基础上, 本节首先从研究层次化移动用户偏好的角度, 提取基于服务加工水平认知的用户偏好; 然后从移动用户偏好受各种上下文因素的分布式、差异化影响程度的角度, 提取基于有效上下文认知的用户偏好; 最后提取综合的基于认知心理学的用户偏好.

5.1 基于服务加工水平认知的用户偏好提取

基于加工水平模型和 3.2 节模型所强调的初次认知、再认知成绩以及语义加工水平, 定义相关概念集, 并设计基于服务加工水平认知的用户偏好提取公式.

定义 8 (服务加工水平认知概念集) LP 表示用户对服务的加工水平认知深度矩阵(如, LP_{ij} 表示 u_i 对 s_j 的加工水平认知深度), SC 表示达到稳定认知(初次认知及再认知)所需要的认知次数矩阵; td 表示单次服务认知处理时间矩阵(用户历史行为变量的直接函数), to 表示两次服务认知的时间间隔距离矩阵, $LP_{threshold}$ 表示用户对服务的语义水平认知阈值, UP_{LP} 为基于服务加工水平认知的用户偏好矩阵.

从加工水平模型的角度, 用户对服务的加工水平深度与达到稳定认知的前 SC_{ij} 次对服务的加工水平相关, 而且服务认知处理时间越长, 服务认知间隔距离越

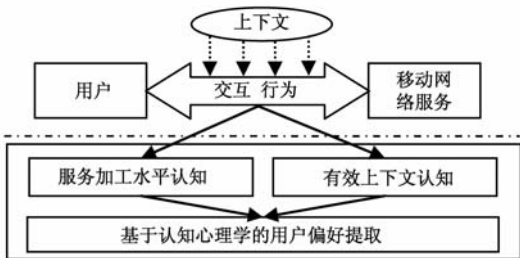


图1 移动网络服务中基于认知心理学的用户偏好提取模型视图

4 移动用户偏好信息结构建模

本节提出一种移动用户偏好信息结构建模方法, 将相关数据的具体定义描述如下:

定义 1 (移动用户偏好信息结构六元组) $UPM = \{U, S, C, UBH, UBHC, UP\}$ 定义为“移动用户偏好信息结构”六元组.

定义 2 (移动用户集合) $U = \{u_i | i \in [1, N]\}$, 定义为移动用户的集合, 其中 N 代表用户数目; 任意用户 $u_i = \{u_{io} | o \in [1, X]\}$ 定义为用户属性的集合, X 代表用户属性类型数目.

定义 3 (移动网络服务集合) $S = \{s_j | j \in [1, M]\}$, 定义为移动网络服务的集合, 其中 M 代表服务数目; 任意服务 $s_j = \{s_{jp} | p \in [1, Y]\}$ 定义为服务属性的集合, 其中 Y 代表服务属性类型的数目.

定义 4 (多维度上下文集合) $C = C_1 \times C_2 \times \dots \times C_L$, 定义为多维度上下文集合, 表示为单维度上下文类型 C_k 的笛卡尔乘积, 而 $C_k = \{c_{kq} | q \in [1, Z_k]\}$ 定义为单维度上下文具体实例集合. 其中 $k \in [1, L], L$ 代表上下

短,用户对服务的加工水平认知深度越高.因此,服务加工水平认知深度是用户对服务达到稳定认知之前的服务认知处理时间与服务认知间隔距离的正反比例函数.定义服务加工水平认知深度如式(1):

$$LP_{ij} = \frac{\sum_{u=1}^{SC_i} (td_{ij})_u}{\sum_{u=1}^{SC_i-1} (to_{ij})_u} \quad (1)$$

公式 1(计算服务语义水平认知阈值) 为了判断用户对服务的加工水平认知深度是否达到语义水平,需要计算每种服务的语义水平认知阈值 $LP_{threshold}$. 因为缺少经验数据,所以将所有已使用该服务的用户的服务加工水平认知深度(需要去除 $LP_{ij} = 0$ 的噪音数据)作为数据源,使用最大类间方差法(OSTU 方法)^[15]来计算 $LP_{threshold}$ (式(2)),使得语义水平认知深度与非语义水平认知深度的类间分离性最好.

$$(LP_{threshold})_j = h_{Ostu} (LP_{ij}) \Big|_{LP_{ij} \neq 0} \quad (2)$$

公式 2(计算基于服务加工水平认知的用户偏好) UP_{LP} 是加工水平深度的函数(式(3)). u_i 对 s_j 的加工水平认知深度越高,获取的信息量越多,用户偏好值越高;如果加工水平达到语义层次,则加强用户偏好值.

$$(UP_{LP})_{ij} = \begin{cases} \left(\frac{\rho \times LP_{ij} - \min_{1 \leq i \leq M} \{LP_{ij}\}}{\rho \times \max_{1 \leq j \leq M} \{LP_{ij}\} - \min_{1 \leq j \leq M} \{LP_{ij}\}} \right) \times (b - a) + a, & \text{if } LP_{ij} \geq (LP_{threshold})_j \\ \left(\frac{LP_{ij} - \min_{1 \leq j \leq M} \{LP_{ij}\}}{\max_{1 \leq j \leq M} \{LP_{ij}\} - \min_{1 \leq j \leq M} \{LP_{ij}\}} \right) \times (b - a) + a, & \text{if } 0 < LP_{ij} < (LP_{threshold})_j \\ 0, & \text{if } LP_{ij} = 0 \end{cases} \quad (3)$$

其中,用户偏好的输出采用区间量化方法,取值为 $[a, b]$ 上的实数. $\rho > 1$ 表示语义水平认知对用户偏好的影响因子.如果 u_i 对 s_j 的加工认知水平不小于语义水平认知阈值,认为 u_i 对 s_j 的偏好水平达到语义水平,则引入语义水平认知影响因子 ρ ,以加强语义水平认知对用户偏好的影响;否则,不引入 ρ .

5.2 基于有效上下文认知的用户偏好提取

基于分布式认知理论和 3.2 节模型,本小节首先定义有效上下文认知概念集,并对有效上下文进行分布式认知检测和分布式影响因子量化;然后,以单维度移动用户历史行为上下文数据为数据源,提取单维度上下文用户偏好;最后,结合分布式影响因子,提取基于有效上下文认知的多维度上下文用户偏好.

定义 9(有效上下文认知概念集) $ubhc_{ijk_q}$ 表示单维度移动用户历史行为上下文,有效上下文集合 $(C_{valid})_{ij}$

表示对用户偏好有显著影响的上下文类型的集合, $D_{threshold}$ 表示有效上下文检验阈值,影响因子 η_{ijk} 表示 u_i 对 s_j 的偏好受某种有效上下文类型 C_k 的影响程度, $(up_c)_{ijk_q}$ 表示 u_i 在单维度上下文实例 c_k 影响下对 s_j 的偏好, $(UP_C)_{ij}$ 表示基于有效上下文认知的用户偏好,即 u_i 在组合有效上下文实例 c_r 影响下对 s_j 的偏好.

在不考虑其他维度上下文的情况下,将 u_i 对 s_j 的移动用户历史行为上下文按照某种单维度上下文类型的某个具体实例 c_k 求平均值,定义为单维度用户历史行为上下文 $ubhc_{ijk_q}$ (式(4)).其中, d_{ij} 为包含 c_k 的用户历史行为上下文的数目.

$$ubhc_{ijk_q} = \frac{1}{d_{ij}} \sum_{r=1}^{d_{ij}} ubhc_{ijr} \mid c_r = c_1 \cdots c_{k-1} c_k c_{k+1} \cdots c_L \quad (4)$$

定义 10(有效上下文认知检测) 有效性上下文认知检测(式(5))对可能影响用户偏好提取的各种上下文因素进行分布式认知检测.以某种单维度上下文 C_k 作用下的用户历史行为上下文为数据源,计算 u_i 在该单维度上下文作用下对 s_j 的用户行为变量方差 D_{ijk} ,从而分析用户偏好受该上下文类型影响程度的波动特征.如果 D_{ijk} 不小于阈值 $D_{threshold}$,则用户 u_i 对服务 s_j 的偏好是受 C_k 影响的;否则,用户偏好不受 C_k 的显著影响.

$$(C_{valid})_{ij} = \left\{ C_k \mid D_{ijk} \geq D_{threshold}, \text{ 其中} \right. \\ D_{ijk} = \sum_{q=1}^{Z_k} \left[\left((abhc_{ijk_q}) - E_{ijk_q} \right)^2 \right] \times p_{kq}, \\ E_{ijk_q} = \sum_{q=1}^{Z_k} \left[(abhc_{ijk_q}) \times p_{kq} \right] \quad (5)$$

其中, $D_{threshold}$ (常量)定义为有效上下文检验阈值,有效上下文即大于给定有效阈值的上下文; E_{ijk} 表示 u_i 在 C_k 作用下对 s_j 的用户行为期望值; p_{kq} 表示 u_i 与 s_j 的交互行为处于上下文具体实例 c_k 的概率(相比于 C_k 的其他具体实例).

基于 $(C_{valid})_{ij}$ 和 D_{ijk} ,定义 u_i 对 s_j 的偏好受单维度有效上下文类型 C_k 影响的影响因子 η_{ijk} (式(6)).其中, $L' \leq L$ 表示有效上下文类型的数量.

$$\eta_{ijk} = \begin{cases} \frac{D_{ijk}}{\sum_{k=1}^{L'} D_{ijk}}, & \text{if } C_k \in (C_{valid})_{ij} \\ 0, & \text{if } C_k \notin (C_{valid})_{ij} \end{cases} \quad (6)$$

公式 3(计算单维度上下文用户偏好) 以单维度用户历史行为上下文 $ubhc_{ijk_q}$ (需去除 $ubhc_{ijk_q} = 0$ 的噪音数据)为数据源,通过 K-Means 方法^[3]计算单维度上下文用户偏好 $(up_c)_{ijk_q}$ (式(7)),其取值区间为 $[a, b]$.

$$(up_c)_{ijk_q} = f_{KMeans} (ubhc_{ijk_q}) \mid ubhc_{ijk_q} \neq 0 \quad (7)$$

公式 4(计算基于有效上下文认知的用户偏好) 在上下文分布式影响因子 η_{ijk} 作用下,根据单维度上下文用户偏好计算多维度有效上下文用户偏好(式(8)).

$$(UP_C)_{ij\bar{v}} = \sum_{k=1}^L (\eta_{ijk} \times (up_c)_{ijk_v}) \quad (8)$$

5.3 基于加权融合的用户偏好提取

本小节综合考虑基于服务加工水平认知的用户偏好和基于有效上下文认知的用户偏好,加权融合提取基于认知心理学的用户偏好(式(9)).其中, $\alpha \in [0, 1]$ 用于衡量用户偏好受多维度上下文因素影响的程度,其值可以由系统自适应地调整(即自适应选择能够使用户偏好提取精确度达到最高的权值),也可以由用户指定(例如,在实际应用中提供相应用户接口,由用户自己选择不同上下文因素对其个性化服务的影响).本文采用前者.

公式 5(计算融合的用户偏好)

$$up_{ij\bar{v}} = (1 - \alpha) \times (UP_{LP})_{ij} + \alpha \times (UP_C)_{ij\bar{v}} \quad (9)$$

6 实验及分析

实验软硬件环境为: Intel 双核 CPU 2.8GHz, 2GB 内存, Windows XP 操作系统, 开发语言环境 JDK1.5.0_04, Eclipse3.2, 数据库 MySQL5.0.

6.1 模拟数据集 MobileServices

目前,移动网络服务领域没有公开可用的包含移动用户历史行为上下文信息的数据集.本文在调研分析移动用户行为研究报告和抓取中国移动网上应用商城(<http://www.mmarket.com/>)提供的移动网络服务的基础上,搭建了一个面向个性化移动网络服务的测试平台,通过设计合理的上下文规则、移动用户历史行为上下文规则约束,构造了一个模拟数据集 MobileServices. 该数据集的 Schema 如下:

(1)移动用户数据集:500 位,用户属性包含用户标识、年龄、性别、职业、消费水平等;

(2)移动网络服务数据集:100 种,服务属性包含服务标识、服务价格、QoS、服务类型等;

(3)上下文数据集:5 种,即时间、位置、使用设备、活动状况、周围人员;其中,时间分为:上午、中午、下午和晚上,并区分工作日和周末;位置分为:办公室、教室、餐厅、家里、超市、交通场所、体育场所、娱乐场所;使用设备分为:PDA、智能手机、笔记本电脑;活动状况分为:工作、学习、会议、饮食、购物、步行、等待、乘坐交通工具、运动、娱乐、睡眠;周围人员分为:家人、朋友、同事、陌生人、独处.

(4)移动用户历史行为上下文数据集:163402 条用户历史行为上下文记录(按 30 天统计,用户行为变量以服务使用时长作为主要考量因素);

(5)移动用户行为数据集:用户-服务行为矩阵(500 * 100),由用户已使用服务的行为变量值与用户未使用服务的行为变量值(即 0)组成.

数据集分为两部分:前 24 天数据(80%)作为训练集,用于提取用户偏好;后 6 天数据(20%)作为测试集,用于评测用户偏好提取的精确度.

6.2 评价标准

评价标准采用 $P@R$ ^[11],即根据所提取用户偏好的大小排序为用户推荐 Top-R 种服务,并将其与测试集中用户常使用的 Top-R 种服务比较,来计算精确度.

$P@R =$

$$\frac{\text{Top-R 推荐服务集所包含测试集 Top-R 种服务的数目}}{R}$$

6.3 实验步骤

实验分为两个阶段:

(1)**Phase1**:基于服务加工水平认知的用户偏好提取(LPUPE)方法的有效性验证实验,基准(Baseline)方法为随机服务选择方式.步骤为:

(a)以训练集数据为输入,使用 LPUPE 方法提取用户偏好,并为用户推荐用户偏好值最高的 R 种服务;同时以训练集数据为输入,获取用户已使用的服务列表,然后为用户随机选择 R 种服务(Baseline 方法).

(b)将测试集中用户使用的服务按照服务使用时长总和排序,取 Top-R 种服务作为用户常使用的服务;然后利用 $P@R$ 计算上述两种方法的服务推荐精确度,并对比分析实验结果.

(2)**Phase2**:基于有效上下文认知的用户偏好提取(DCUPE)方法、基于认知心理学的用户偏好提取(CPUPE)方法的有效性验证实验.步骤为:

(a)以训练集数据为输入,分别利用 MCUPE、DCUPE 和 CPUPE 方法提取上下文相关的用户偏好,然后分别为用户在各种组合上下文实例下推荐用户偏好值最高的 R 种服务.其中,MCUPE(Multi-dimensional Contextual User Preferences Elicitation)方法按照式(7)计算各种单维度上下文用户偏好之后,将用户偏好受各类上下文类型的影响看作等权重(即各上下文类型的影响因子 η_{ijk} 相同),然后再按照式(8)提取多维度上下文用户偏好.

(b)以测试集为输入,将用户在各种组合上下文实例下使用的服务按照服务使用时长总和排序,取 Top-R 种服务作为用户在该组合上下文实例下常使用的服务;然后,利用 $P@R$ 计算 MCUPE、DCUPE、CPUPE 在所有组合上下文实例下的平均服务推荐精确度,并对比分析实验结果.

6.4 实验结果及分析

图 2 展示了 LPUPE 和 Baseline 方法的平均精确度

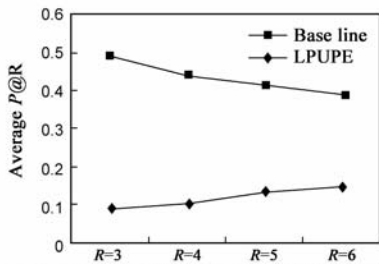


图2 LPUPE方法的有效性验证实验结果

$P@R(R=3,4,5,6)$ 的对比实验结果.其中,稳定认知次数 $SC=3$ (实验筛选值),语义水平认知影响因子 $\rho=1.5(>1)$.

可以看出,以 $P@R$ 为评价标准,LPUPE 的整体精度优于没有提供个性化服务的随机服务选择方式,这说明将加工水平模型引入用户偏好提取方法是有意义的.但是,随着 R 增大,LPUPE 方法的平均精确度将逐渐缓慢下降,而 Baseline 方法的平均精确度缓慢上升.针对这一问题,我们通过增大服务推荐数目 R 的值展开进一步实验发现,当 R 增大为 21 时,LPUPE 的 $P@R$ 值由缓慢下降开始转为缓慢上升;而 Baseline 方法则一直保持缓慢上升.这说明 LPUPE 方法尤其适用于移动环境下设备输出能力有限(R 值也因此受限)的情形.

图 3 展示了 MCUPE、DCUPE 和 CPUPE 三种方法的 $P@R$ 对比实验结果.其中, $SC=3$, $\rho=1.5$ 保持不变; $D_{threshold}=1.2$, $\alpha=0.8$ (经多次实验筛选确定).

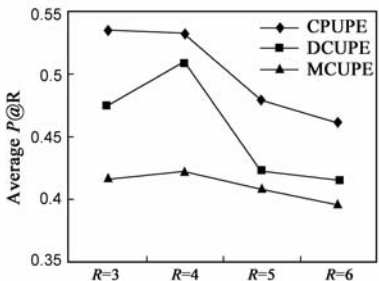


图3 MCUPE、DCUPE和CPUPE对比实验结果

可以看出,以 $P@R$ 为评价标准,DCUPE 方法优于 MCUPE 方法,这说明将各种上下文类型及其实例对移动用户偏好的影响程度区别对待并量化,有利于提高移动用户偏好提取精确度.从直观意义上来讲,某用户对某种移动服务的偏好可能受位置和活动内容影响较大,而受时间、设备和周围人员上下文的影响较小,或许反之;即不能将所有上下文类型看作有效上下文,更不能将它们等权重考量.这与本文的实验结果是一致的.这是因为,进行分布式有效上下文认知检测,有利于去除噪音上下文因素.

从图 3 还可以看出,CPUPE 方法总体上优于 DCUPE 方法,即融合了加工水平模型和分布式认知理论的用户偏好提取精确度在一定程度上优于基于分布式认知

理论的用户偏好提取精确度.这说明,将 DCUPE 方法应用于组合上下文具体实例时,存在一定程度的不确定性,即有些用户偏好较少或者不受上下文因素影响,也有些用户偏好受上下文因素的影响是随机的,需要通过其他方法(如 LPUPE)来修正.

此外,对比图 2、图 3 发现,虽然 DCUPE 与 LPUPE 方法的 $P@R$ 没有直接可比性(因为基于 DCUPE 的推荐结果受到上下文条件的约束),但在数值上 DCUPE 却并不比 LPUPE 低.这表明移动用户的移动网络服务偏好在一定程度上确实受到上下文条件的影响.例如,某用户可能“很喜欢”某游戏类移动服务,但用户在“办公室”且“工作”的组合上下文实例下对该类服务的用户偏好值可能会很低.

7 结束语

移动用户偏好提取方法是缓解“移动信息过载”问题的有效手段,具有重要的研究意义和实用价值.本文在用户偏好信息结构建模基础上,提出一种移动网络服务中基于认知心理学的用户偏好提取方法,从研究层次化移动用户需求、有效上下文信息的分布式认知与差异化计算的角度,分析各种因素对偏好提取过程的影响,并逐步提取精确的用户偏好.仿真实验结果表明,该方法能够有效挖掘层次化移动用户需求以及上下文关联的移动用户行为数据,发现不同上下文类型及其实例对移动用户偏好的差异化影响程度,并降低噪音上下文因素带来的消极影响,从而提高了个性化移动网络服务精确度.下一步主要工作包括^[16]:检测不同上下文类型之间存在的依赖关系及其对移动用户偏好的影响,计算基于内容的移动用户偏好(即通过分析上下文关联的移动用户历史行为提取用户对移动网络服务属性特征的偏好),研究如何缓解多维度上下文用户偏好稀疏性问题和基于社会化网络分析的上下文推荐问题.

参考文献

- [1] 许海玲,吴潇,李晓东,阎保平.互联网推荐系统比较研究[J].软件学报,2009,20(2):350-362.
Xu Hailing, Wu Xiao, Li Xiaodong, Yan Baoping. Comparison study of Internet recommendation system[J]. Journal of Software, 2009, 20(2): 350-362. (in Chinese)
- [2] 吴永辉,王晓龙,丁宇新等.基于主题的自适应、在线网络热点发现方法及新闻推荐系统[J].电子学报,2010,38(11):2620-2624.
Wu Yonghui, Wang Xiaolong, Ding Yuxin, et al. Adaptive online web topic detection method for web news recommendation system[J]. Acta Electronica Sinica, 2010, 38(11): 2620-2624. (in Chinese)

- [3] Gemmis De M, et al. Preference learning in recommender systems[A]. In ECML/PKDD-09 Workshop on Preference Learning[C]. Bled, Slovenia: Springer-Verlag, 2009. 41 – 55.
- [4] Jung SY, Hong JH, Kim TS. A statistical model for user preference[J]. IEEE Trans on Knowledge and Data Engineering, 2005, 17(6): 834 – 843.
- [5] Bettini C, et al. A survey of context modelling and reasoning techniques[J]. Pervasive and Mobile Computing, 2009, 6(2): 161 – 180.
- [6] Stefanidis K, Pitoura E. Fast contextual preference scoring of database tuples[A]. Proc EDBT' 08[C]. German: Springer-Verlag, 2008. 344 – 355.
- [7] Adomavicius G, Tuzhilin A. Context-aware recommender systems[A]. Recommender Systems Handbook[M]. Springer Press, 2011. 217 – 253.
- [8] Jembere E, Adigun MO, Xulu SS. Mining context-based user preferences for m-services applications[A]. Proc IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)[C]. Washington, DC: IEEE Computer Society, 2007. 757 – 763.
- [9] Jason J J. Contextualized mobile recommendation service based on interactive social network discovered from mobile users[J]. Expert Systems with Applications, 2009, 36(9): 11950 – 11956.
- [10] Kuo MH, Chen LC, et al. Building and evaluating a location-based service recommendation system with a preference adjustment mechanism[J]. Expert Systems with Applications, 2009, 36(2): 3543 – 3554.
- [11] Wang LC, Meng XW, Zhang YJ and Shi YC. New approaches to mood-based hybrid collaborative filtering[A]. Proc ACM CAMRa' 10[C]. New York: ACM Press, 2010. 28 – 33.
- [12] 谢海涛, 孟祥武. 适应用户需求进化的个性化信息服务模型[J]. 电子学报, 2011, 39(3): 643 – 648.
Xie Haitao, Meng Xiangwu. A personalized information service model adapting to user requirement evolution[J]. Acta Electronica Sinica, 2011, 39(3): 643 – 648. (in Chinese)
- [13] Craik F, et al. Levels of processing: Past, present, and future? [J]. Memory, 2002, 10(5/6): 305 – 318.
- [14] Hollan J, Hutchins E, Kirsh D. Distributed cognition: toward a new foundation for human-computer interaction research[J]. ACM Trans on Computer-Human Interaction, 2000, 7(20): 174 – 196.
- [15] 许向阳, 宋恩民, 金良海. Otsu 准则的阈值性质分析[J]. 电子学报, 2009, 37(12): 2716 – 2719.
Xu Xiangyang, Song Enmin, Jin Lianghai. Characteristic analysis of threshold based on otsu criterion[J]. Acta Electronica Sinica, 2009, 37(12): 2716 – 2719. (in Chinese)
- [16] Wang LC. Understanding and using contextual information in recommender systems[A]. Proc ACM SIGIR' 11[C]. New York: ACM Press, 2011. 1329 – 1330.

作者简介



王立才 男, 1984 年生于山东阳谷, 北京邮电大学博士研究生. 研究方向为推荐系统、智能信息处理.

E-mail: wiizane@gmail.com



孟祥武 男, 1966 年生于山东招远, 北京邮电大学计算机学院教授, 博士生导师. 研究方向为通信软件, 网络服务, 人工智能.

张玉洁 女, 1969 年生于天津杨柳青, 北京邮电大学讲师. 研究方向为智能信息处理, 通信软件, 网络服务.