

上下文感知推荐系统中 基于用户认知行为的偏好获取方法

高全力 高 岭 杨建锋 王 海

(西北大学信息科学与技术学院 西安 710127)

摘 要 针对现有的偏好获取方法,因未考虑不同用户在各类型上下文环境中的认知行为与用户偏好间的内在联系所导致用户偏好预测准确度不高的问题,受分布式认知理论与记忆信息加工模型启发,提出了一种基于用户认知行为的上下文感知偏好获取方法。在多维上下文环境下,将认知水平、认知有用性、认知风险、有效认知行为等认知领域概念引入偏好获取过程,并分别给出其概念定义及计算方法,通过建立多种认知因素交互影响的偏好获取模型,分别提取在单维与多维上下文环境下的用户偏好。在大规模真实数据集上的实验结果表明,与经典的协同过滤算法及上下文感知算法相比,显著地提高了偏好获取的准确度。

关键词 认知行为;上下文感知;偏好获取;推荐系统

中图法分类号 TP391

DOI号 10.11897/SP.J.1016.2015.01767

A Preference Elicitation Method Based on Users' Cognitive Behavior for Context-Aware Recommender System

GAO Quan-Li GAO Ling YANG Jian-Feng WANG Hai

(School of Information Science and Technology, Northwest University, Xi'an 710127)

Abstract To solve the problem that the internal relation between user's cognitive behaviors and preferences have not been considered which lead to low predictive accuracy in existing recommender systems, inspired by the theory of distributed cognition and the processing model of memory, a cognitive behavior based approach is proposed. Firstly, the paper introduces many cognition concepts into the process of acquiring user preferences under multidimensional context environment, such as cognitive level, cognitive usefulness, cognitive risk, effective cognitive and so on. Secondly, this paper provides the definitions and calculation methods of those concepts. Finally, it elicits user preferences under unidimensional and multidimensional context environments by establishing the cognitive factors' mutual effect model. The authors present empirical experiments by using a real extensive dataset, experimental results show that the proposed algorithms can achieve better prediction accuracy compared with collaborative filtering and context-aware algorithm.

Keywords cognitive behavior; context-aware; preference elicitation; recommender system

收稿日期:2014-04-01;最终修改稿收到日期:2014-12-23。本课题得到国家自然科学基金(61373176)、教育部博士点基金(20116101110016)、陕西省自然科学基金(2012JQ8047)、陕西省教育厅科研基金(11JK1059)、陕西省重大科技创新专项资金(2012ZKC05-2)及陕西省教育厅 2013 年科学研究计划(2013JK1178)资助。高全力,男,1988 年生,博士研究生,主要研究方向为智能推荐系统。E-mail: gaoquanli@nwu.edu.cn。高 岭,男,1964 年生,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为计算机网络、服务计算。杨建锋,男,1981 年生,博士研究生,工程师,主要研究方向为智能推荐系统。王 海,男,1977 年生,博士,讲师,主要研究方向为机器学习、服务计算。

1 引 言

随着信息技术的发展,网络上的各种信息也随之急剧增长,而这种增长远远超出了个人或系统所能接受、处理和有效利用的范围,这就是“信息过载”问题.随着“普适计算”和“面向服务”等新技术的出现,“信息过载”的问题越来越严重,用户需要处理的信息数量更是呈现几何级增长,并且大量无关的冗余信息严重干扰了用户对相关有用信息的选择,若仅仅依靠人们自身去处理这些信息,所需的时间是不可接受的^[1].虽然搜索引擎、信息检索等技术的发展在一定程度上缓解了这个问题,但仍无法解决由教育背景、时间、情绪等方面的不同所带来的偏好差异^[2],这直接导致了网络资源的利用率低下与用户体验受到较大影响的问题.

在这种环境下,能够解决不同用户的个性化需求的推荐系统(Recommender System)成为学术界与工业界的热门研究领域.现有的推荐系统往往利用用户的历史行为信息及用户间或项目间的相似性关系来挖掘其潜在偏好,或仅仅聚焦于挖掘“用户-项目”(User-Item)间的二维关系模型,为独立的用户推荐最相关的资源,而不考虑用户当前的时间、位置、环境等上下文信息^[3].这些技术方案在实际应用中是不合理的,因为用户的偏好会随着时间、教育程度、职业等因素的变化而变化,并且用户所处的环境也会实时地影响用户的决定^[2].在这种情况下,上下文感知(Context-Aware)技术在推荐领域里的应用研究,成为学者们新的研究热点^[1-3].

多数的上下文感知推荐算法,通过将上下文信息融入到“用户-项目”的二维关系模型中,然后定量分析其对用户偏好的影响^[1-2,4],也即是在上下文环境下对用户的行为进行分析所得出的预测,而用户行为又受其认知心理的直接支配,其认知心理波动又在一定程度上受当前上下文环境的影响.于是我们从用户认知行为间的差异入手,通过研究不同用户在各类上下文环境下的认知行为与其偏好间的内在联系,我们提出了一种面向“信息过载”问题的基于认知行为的上下文偏好获取方法,将各种上下文信息及用户和项目的自有属性对用户认知行为的影响,抽象为认知水平、认知有用性、认知风险、有效认知行为等概念模型,并给出了其相互影响修正关系及全面的偏好获取流程,以获得更加准确的用户偏好.本文主要的贡献为:

(1)将认知水平、认知有用性、认知风险、有效认知行为等认知领域概念引入偏好获取过程,并给出其概念定义、算法模型及其间交互影响关系;

(2)提出了一种融合上下文信息及认知行为的单维与多维偏好提取算法,并在大规模真实数据集上进行了验证.

本文第2节给出相关研究工作;第3节详细描述本文所提出算法的实现方法;第4节给出实验设计及结论;第5节给出全文小结及以后研究方法.

2 相关工作

推荐系统的目的是通过对各种影响用户偏好的自身及外界因素进行分析建模,以获取用户潜在的兴趣内容并推荐给用户.传统的推荐技术主要包括基于内容的推荐算法(Content-based Recommender)^[4]、协同过滤算法(Collaborative Filtering)^[5]、混合推荐算法(Hybrid Recommender)^[6]等.这些算法都是将“用户-项目”历史评分记录作为数据源,通过数据挖掘与优化算法(神经网络、蚁群算法、决策树等)挖掘其最近邻域用户或最近邻域项目,并通过定量分析行为关系模型来预测用户的偏好关系.

随着“普适计算”概念的提出,系统能够自主发现并分析利用时间、周围环境、地理位置等上下文信息为用户提供服务,上下文感知计算理论也随之成为研究热点^[1-3,7].在推荐系统领域中,许多研究者也随即发现仅仅依靠“用户-项目”二元关系,在实际应用场景中其准确性与适用性都存在很大问题^[7].例如有些用户平时喜欢惊悚题材的电影,而在其压力较大或心情烦躁时,就更倾向于轻松幽默题材的电影;有的用户更倾向于在晚上而非中午被推荐一些餐厅、商场等.上下文感知推荐系统随即成为新的研究热点,其把与用户相关的上下文信息融入偏好获取算法,进一步提高了推荐适用性与精确度^[7].例如在广告推送应用中,通过将上下文信息与数学模型结合,能够线性增强用户体验与广告点击率^[8];在移动社交网络领域,将上下文相似度引入移动服务选择过程中,大幅提高了服务选择的准确性与可靠性^[9].

现有的上下文偏好获取算法大多是通过定量分析上下文环境下用户的行为信息来获取用户的偏好,而未考虑用户的行为是受其认知心理直接支配以及不同用户对于各类型的上下文信息的认知行为与用户偏好间的内在联系,导致了偏好预测准确性的不足^[3].所以近年来,有些研究者开始结合认知心

理学来研究用户偏好获取技术. 例如, 有些研究者通过引入认知心理学与信息加工水平模型来获取用户的偏好信息^[10], 或者将“记忆两过程”理论融合上下文信息对用户偏好进行建模分析^[11]. 鉴于用户认知心理学理论在教育教学、模式识别、用户建模、推荐系统中的广泛应用^[10-11], 如何从用户认知行为的角度出发, 并结合上下文信息来获取用户偏好, 将成为推荐系统领域的重要研究热点之一.

3 基于用户认知行为的上下文偏好获取

本节首先对上下文感知系统所涉及的相关信息结构化定义, 给出上下文感知系统的形式化描述, 然后从各种上下文信息及用户和项目的自有属性对用户偏好影响的角度出发, 将认知水平、认知风险、认知有用性等相关概念模型引入偏好获取过程, 以提取在单维与多维上下文环境下基于认知行为的用户偏好.

3.1 上下文感知推荐系统

目前“上下文”仍无标准的定义, 较多采用的是 Dey^[12] 给出的一个被认为是相对规范的定义: “上下文是环境本身及其所包含的实体通过隐式或显式的方式, 表达出用来描述其状态的任何信息. 其中, 实体可以是人、地理位置或与用户及应用程序间交互的相关客体(包括用户及应用程序本身)”. 环境中 k 种不同类型的上下文信息形式化定义如下:

$$C = (C_1, C_2, \dots, C_k),$$

其中, $C_i \subset C (i \in [1, k])$ 表示任一类型的单维度上下文信息, 其具体描述为 $C_i = \{c_{i_p} | p \in [1, q_i]\}$, q_i 则表示该维上下文类型的实例信息数目. 量化的上下文信息示例如下:

$C_i: Emotion$ (愉悦[1], 郁闷[2], 悲伤[3], 难过[4]),

$C_j: Location$ (办公场所[1], 移动环境[2], 家[3],

学校[4], 餐厅[5]).

那么在上下文环境下, 传统的“用户-项目”二维评分预测模型就扩展为包含上下文信息的多维预测模型 $R: User \times Context \times Item \rightarrow Rating$ ^[3], 其中 $User, Context, Item$ 的结构化描述示例如下:

$$User \subseteq UName \times Address \times Income \times Age;$$

$$Item \subseteq IName \times Type \times Price;$$

$$Context \subseteq Location \times Emotion \times Time \times$$

$$Environment.$$

基于上述描述, 我们提出了一种在上下文环境

下, 将用户的认知行为及上下文感知偏好获取系统进行建模分析的方法, 相关数据的定义描述如下.

定义 1. 完全项目集, 在上下文环境中, 项目的提供者包含不同项目与内容集合的网络服务商, 任一类别的项目实体 l 可描述为 $Ent_l = \{S_s | s \in [1, N]\}$, N 为项目数目. 完全项目集合仅描述与特定的上下文类型无关的初始项目集合, 即为各单类项目实体的最大并集, 结构化描述为 $S_{vr} = \{svr_n | n \in [1, m]\}$, 其中 $svr_n \subseteq \{Ent_1 \cup \dots \cup Ent_l\}$, $m = Card\{Ent_1 \cup \dots \cup Ent_l\}$ 为初始项目集合中的总项目数量.

定义 2. 多维上下文用户偏好, 对于任意用户 U_j , $Pre_{jis} = \{Pre_{jis} | j \in [1, U], i \in [1, k], s \in [1, N]\}$, 表示任一用户 U_j 在上下文环境 C_i 下对项目 S_s 的偏好, 其中 $S_s \in S_{vr}$, Pre_{jis} 值越大说明项目 S_s 符合用户 U_j 偏好的概率越大.

定义 3. 上下文用户偏好信息结构, $\forall U_j$, $UP_j = \{U_j, S_{vj}, C_j, PreS_j\}$ 描述上下文用户 U_j 的偏好信息结构, 其中 S_{vj} 表示用户 U_j 的初始完全项目集, 即为用户 U_j 可能接触到的各类项目的合集, C_j 为当前上下文信息, $PreS_j$ 为 U_j 的偏好项目集.

定义 4. 上下文感知推荐系统形式化定义, 若 $UD_{pre} = \{D_{j1}, \dots, D_{j(Ent_l)}\} (j \in [1, U], Ent_l \subseteq S_{vr})$ 为待预测偏好的目标维度空间, $D_{pre} = \{D_{i1}, \dots, D_{i(Ent_k)}\} (i \in [1, U], Ent_k \subseteq S_{vr})$ 为偏好结果维度空间, 且当前 $UD_{pre} \cap D_{pre} = \emptyset$, 效用函数 $F_{uty}(\cdot)$ 用于计算用户在高维度上下文条件下对项目的偏好 Pre_{jis} , 那么上下文感知推荐系统就是为 $d_{j1}, \dots, d_{j(Ent_l)}$ 找到偏好值最大的那些元组 $d_{i1}, \dots, d_{i(Ent_k)}$, 其形式化定义如下^[13]:

$$\forall (d_{j1}, \dots, d_{j(Ent_l)}) \subseteq UD_{pre},$$

$$(d_{i1}, \dots, d_{i(Ent_k)}) = \arg \max F_{uty}(UD_{pre}, D_{pre}).$$

3.2 认知行为相关理论的引入

在 20 世纪 90 年代, Cole 等人^[14] 认为人们的认知活动不仅存在于大脑, 还存在于各种物理环境、社会关系、时间因素等参与认知活动的一切要素中, 验证了人们的认知活动与外界上下文环境的关联性. Craik 等人^[15] 提出“信息加工水平模型”, 从信息加工水平的深度与时间的角度分析了上下文环境对再认知水平的影响, 给出了影响认知水平的外界因素. 王立才等人^[10] 将认知心理学融入移动用户偏好提取的过程中, 能够大幅提高推荐的精确度. 因此, 我们受“分布式认知理论”与“记忆信息加工模型”相关理论的启发, 将认知行为引入偏好获取过程, 将各种

上下文信息及用户和项目的自有属性对用户认知行为的影响,抽象为认知水平、认知有用性、认知风险、有效认知行为等概念模型,其相互影响关系模型视图如图 1,各概念模型具体描述如下。

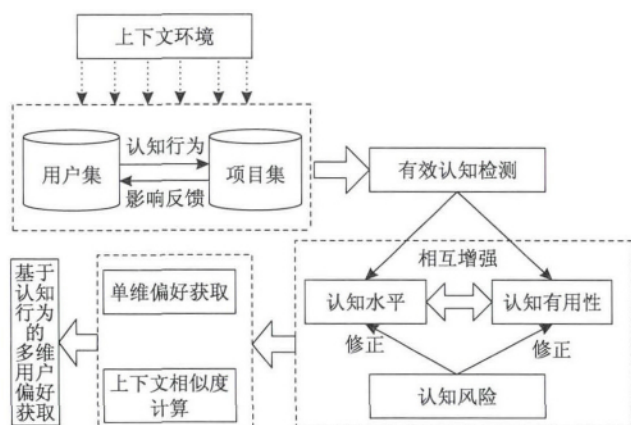


图 1 基于用户认知行为的上下文偏好获取模型视图

定义 5. 用户及认知,用户(User)在本文中指认知活动中特定的个体,由其自身各种固有属性构成,其描述为 $User = \{Name, Age, Income, Address\}$. 认知(Cognition)指通过心理活动(如形成概念、知觉、判断或想象)来获取信息,并接受各种外在信息的影响反馈;认知的过程即为用户对各种上下文信息的加工过程,认知能力指的是人脑加工、存储和提取信息的能力,与用户需求获取及个性化决策存在着密切的联系。

定义 6. 认知域,表示各种用户认知水平的集合,根据认知水平的五等级模型^[16],我们将认知水平划分为:(1)初级认知(Primary Cognition).利用陈述性知识来解决偏好选择问题;(2)条件再认知(Conditional Re-cognition).通过再认知历史熟悉的模式来确定偏好;(3)条件推理(Conditional Inference).根据历史熟悉模式来推理新模式的偏好;(4)语义认知(Semantic Cognition).对某类项目的认知达到语义级别;(5)条件构建(Conditional Construction).没有相关历史经验参考,自行构建缺失条件进行认知偏好分析.因此认知域可描述为 $CD = \{Pc, Cr, Ci, Sc, Cc\}$.引入认知域的目的在于给出认知水平的规范分级模型,本文不对其具体分级取值进行讨论分析。

定义 7. 认知水平,表示用户在各种外在环境及自身属性影响下对当前项目的接受程度.为度量用户的认知水平,本文总结出影响用户认知水平的各种影响因素,主要包括:(1)用户固有属性,包括年龄、性别、教育程度、职业等;(2)项目的固有属

性,包括价格、内容、展现形式等;(3)对当前项目的关注程度,受对其同类项目的浏览次数、浏览时间、浏览间隔影响;(4)历史经验影响,对该类项目的历史接触所带来的喜好倾向,即对该类项目是否有先入性认知(历史经验),一般由对其同类项目的量化评分值来衡量;(5)当前上下文环境,包括时间、位置、周围人员、使用设备等.因为不同的外界环境对于不同人的影响有很大的差异,其影响程度也难以进行具体的取值分析,所以本文从对认知水平影响类型的角度将上述因素分类为:①正相关因素(Positive Correlation),即能够增强用户对当前项目偏好程度的因素,例如对该类项目较大的自身需求,该类项目具有较好的评价信息等;②负相关因素(Negative Correlation),即能够降低用户对当前项目偏好程度的因素,例如嘈杂的外界环境、较差的历史经验等(其可由数据采集时人为判定或采用与具体应用环境相结合的构造规则进行判定).显然,认知水平判定是正、负相关因素的正反比例函数,其定义方式如下:

$$Cl_{js} = \frac{\sum_{m \in pc} (Pc_{js})_m}{\beta + \sum_{n \in nc} (Nc_{js})_n} \quad (1)$$

其中, Cl_{js} 表示用户 U_j 对于项目 S_s 的认知水平,且 $Cl_{js} \in CD$, Pc_{js} 与 Nc_{js} 分别表示其正、负相关因素,其为等值的量化单位,默认初始值为 1,即若 Pc_{js} 的数量为 3, Nc_{js} 为 5,则当前用户认知水平度量值为 $3/5$, pc 与 nc 分别表示正、负相关因素集, β 为引入的修正因子,取值如下:

$$\beta = \begin{cases} 1, & \text{如果 } \sum_{n \in nc} (Nc_{js})_n = 0 \\ 0, & \text{其他} \end{cases} \quad (2)$$

定义 8. 认知有用性,为了表示项目提供商所提供项目对用户有无吸引力,我们引入认知有用性的概念,其是决定用户是否接受所提供项目的关键性因素.由于缺少历史经验数据,所以将所有已使用该项目的用户认知水平作为数据源.为了判断在当前上下文环境 C_i 下用户 j 对项目 S_s 的认知有用性,我们采用平均偏差法(Mean Variation)^[17]来表示认知有用性的量化数值, US 表示对项目 S_s 有过行为记录的用户集,则认知平均偏差如下:

$$MV_{is} = \frac{\sum_{r \in US} \left| Cl_{rs} - \frac{\sum_{r \in US} Cl_{rs}}{\text{card}(US)} \right|}{\text{card}(US)} \quad (3)$$

其中 $\text{card}(US)$ 表示用户集 US 中用户的个数,则认

知有用性量化数值为

$$CU_{jis} = \frac{\sum_{r \in US} Cl_{rs}}{card(US)} + MV_{is} \quad (4)$$

其中 CU_{jis} ($j \in [1, U], i \in [1, k], s \in [1, N]$) 表示用户 U_j 在上下文环境 C_i 下对项目 S_s 的认知有用性度量值. 采用这种定义方式的原因为: (1) 对当前项目有过行为记录用户的平均认知水平, 其表示项目自身的吸引力; (2) 在当前上下文环境下, 不同用户间的认知水平偏差程度, 其值越大就表示当前上下文环境下, 外界影响因素越复杂, 用户的认知分布越广泛, 从整体而言项目满足用户需求的概率就越大, 即作为一种考虑当前上下文的对于平均认知水平的修正性因素.

定义 9. 认知风险, 指用户决策前无法预知其是否符合自身偏好, 即在决策中存在一种心理上的不确定感. 认知风险是影响用户决策的最大障碍, 会暗示用户推迟或取消本次选择. 其受用户先入性认知(历史经验)与项目本身属性的影响(价格、类型等), 主要包括用户对该项目的信息收集程度、项目的被选择次数、评价信息、内容风险(质量不能达到用户需求的风险, 用户本身对此类商品的功能需求越高, 内容风险越大)等. 我们设定认知风险临界阈值 Rf_{thd} , 其默认初始值为各影响因素的均值(根据具体应用而改变), 即若该项目被选择次数大于其同类项目的平均被选择次数或者商品评价高于评价均值等, 其认知风险就会降低, 反之就会增加. 用 Rfb_{si} 与 Rfl_{si} 分别表示项目 S_s 第 i 个影响因素大于等于、小于 Rf_{thd} , 其在第 i 个影响因素处的正、负风险取值 Rfp_{si} 与 Rfn_{si} 分别为

$$Rfp_{si} = \begin{cases} 1, & Rfb_{si} \\ 0, & Rfl_{si} \end{cases} \quad (5)$$

$$Rfn_{si} = \begin{cases} 0, & Rfb_{si} \\ 1, & Rfl_{si} \end{cases} \quad (6)$$

用户的认知风险不但与当前项目的固有属性相关也与其历史风险(以往的认知风险)相关, 用户对项目 S_s 同类项目历史风险越大, 本次认知存在风险的可能性就越高. 因此, 为描述用户 U_j 对于项目 S_s 的认知风险, 我们综合考虑各风险影响因素与历史风险, 采用动态风险下界的方法来描述, 如下:

$$Rf_{js}(t) = \left(\frac{Rf_{js}(t-1)}{Rf_{js}(t-1) + \varphi} \right) \cdot \left(\frac{\sum_{i=1}^{rs} (Rfp_{si})}{\sum_{i=1}^{rs} (Rfn_{si}) + \xi} \right) \quad (7)$$

其中, $Rf_{js}(t)$ 表示当前的认知风险, $t(t \in N)$ 为当前的认知风险次数, 其取值依赖于不同用户的历史行为记录数, 且 $Rf_{js}(0) = 1$, rs 为项目 S_s 的影响因素个数, φ 为波动参数, 其取值范围为 $(0, 1)$, 用以控制历史风险对于当前认知风险的影响幅度, ξ 为防止分母为 0 的修正参数, 默认初值如下:

$$\xi = \begin{cases} 1, & \text{如果 } \sum_{i=1}^{rs} (Rfn_{si}) = 0 \\ 0, & \text{其他} \end{cases} \quad (8)$$

定义 10. 有效认知行为, 为去除认知域中各种噪声数据, 防止情绪化的认知、突发性认知、偶然性认知错误^[15]等对于偏好获取的影响, 我们引入有效认知的概念. 为了判断认知的有效性, 对认知水平的波动幅度与认知风险进行评测, 用 VC_{js} 表示如下:

$$VC_{js} = \begin{cases} 1, & \text{如果 } Cl_{js} \geq vc_{thd} + Cl_{j(s-1)}, \\ & Rf_{js}(t) \geq (Rf_{js}(t-1) + Rf_{thd}) \\ 0, & \text{其他} \end{cases} \quad (9)$$

即若此次认知有较大的认知风险且认知水平波动幅度大于阈值, 则为无效认知, 即 $VC_{js} = 1$, 否则 $VC_{js} = 0$. 其中, vc_{thd} 认知水平波动阈值, 初始值为认知平均偏差 MV_{is} , Rf_{thd} 为认知风险参数.

3.3 基于用户认知行为的上下文偏好获取

本小节通过量化分析用户各认知行为间的约束关系及多维上下文间相似性, 来提取单维与多维上下文环境下的用户偏好.

公式 1. 单维上下文偏好提取算法 UDCP (UniDimensional Context-aware Prediction), 以单维上下文用户行为数据为数据源, 用户 U_j 在上下文类型 C_i 下对于项目 S_s 的偏好, 随着认知水平 Cl_{js} 与认知有用性 CU_{jis} 的增加而增加, 即用户对于该项目的信息收集程度、认可程度较高, 其符合用户偏好的概率就越大, 且这种偏好概率受认知风险 $Rf_{js}(t)$ 的约束修正. 提取单维上下文偏好如下:

$$Pre_{jis} = (1 - Rf_{js}(t)) \cdot (\lambda_1 \cdot CU_{jis} + \lambda_2 \cdot Cl_{js}) \quad (10)$$

其中 λ_1 与 λ_2 为认知有用性与认知水平的权重因子, 通过 λ_1 与 λ_2 的不同取值来调节认知有用性与认知水平的权重. 由式(10)不难发现, 当认知风险过高时, Pre_{jis} 值可能为负, 而这种情况并不符合客观实际, 因此, 我们引入修正算法来修正 Pre_{jis} 的取值, 如下:

$$Pre_{jis} = \begin{cases} Pre_{jis}, & \text{如果 } Pre_{jis} \geq 0 \\ 0, & \text{其他} \end{cases} \quad (11)$$

公式 2. 多维上下文相似性, 我们采用泊松相关系数^[3]来表示多维上下文间的相似性. 在任意两种不同类型的上下文环境 C_a 与 C_b 下, 设同时对项目 S_s 有过行为记录的用户集合为 US , 上下文 C_a 与 C_b 间相似性为

$$Sim(C_a, C_b) = \frac{\sum_{\substack{C_a, C_b \in C \\ j \in US}} (Pre_{jC_a} - \overline{Pre_j})(Pre_{jC_b} - \overline{Pre_j})}{\sum_{\substack{C_a, C_b \in C \\ j \in US}} \sqrt{(Pre_{jC_a} - \overline{Pre_j})^2 (Pre_{jC_b} - \overline{Pre_j})^2}} \quad (12)$$

其中, Pre_{jC_a} 与 Pre_{jC_b} 分别表示用户 U_j 在单维上下文环境 C_a 与 C_b 下对项目 S_s 的偏好值, $\overline{Pre_j}$ 表示用户 j 在不同上下文环境下的平均偏好值. 通过减去不同用户自身的平均偏好值来修正其偏好差异. 相似度值越大, 表示两个上下文环境越相似.

公式 3. 融合多维上下文的偏好提取算法 MDCP (MultiDimensional Context-aware Prediction), 根据式(12)计算出的多维上下文间相似度, 得出相似度最高的 k 个上下文集合 $CM (CM \subset C)$, 用户 U_j 在多维上下文环境下对项目 S_s 的偏好为

$$(Pre_{final})_{jC_s} = \left(1 + \frac{\sum_{C_k \in CM} |Pre_{jC_k} - \overline{Pre_{C_k}}|}{k} \right) \cdot Pre_{jC_s} \quad (13)$$

其中, $\overline{Pre_{C_k}}$ 为项目 S_s 在上下文环境 C_k 下对用户的平均偏好值, 在多维环境下的偏好值即是在各最相似的单维上下文环境下偏好偏差对其单维上下文偏好的影响.

4 实验步骤及结果分析

本节通过实验来验证如下问题: (1) 验证认知水平与认知有用性在不同规模数据集上的最佳权重取值及 3 种认知因素对于偏好获取的影响; (2) 所提出的 UDCP 与 MDCP 算法与其他偏好获取算法相比, 能否取得更好的推荐质量.

4.1 实验数据集

本文实验所采用的数据集为在真实数据集 BookCrossing^[18]的基础上, 通过加入合理的上下文生成规则, 构造了一个模拟真实数据集 MBookCrossing. BookCrossing 数据集是由 Cai-Nicolas Ziegler 使用爬虫程序从 Book-Crossing 图书社区上采集的 278858 个用户对 271379 本书进行的行为信息. 构造后的数据集主要包括 4 部分:

(1) MBX-Users, 包含用户的 ID、位置、年龄.

(2) MBX-Books, 包含图书的 ISBN、标题、作者、发行年代、出版社和缩写.

(3) MBX-Book-Ratings, 包含用户对图书的评分信息.

(4) MBX-Book-Contexts, 包含时间、位置、周围环境等上下文信息. 其中, 时间分为: 早上、上午、下午、晚上; 位置分为: 办公场所、家、学校、餐厅; 周围环境分为: 视觉环境、声音环境、空间环境; 状态信息分为: 乘坐交通工具、娱乐、工作、学习等.

针对当前的数据集环境, 我们经过反复筛选, 正、负相关因素的判定规则为: 以下因素若为“是”则为正相关因素, 否则为负相关因素. 发行年代为是否最近 5 年图书, 出版社是否为排名前 1/2, 教育程度是否为大学及以上, 是否为嘈杂的外界环境、用户评分是否大于均值. 有用性强的物品特征为: 用户评分大于评分均值; 对于该物品评分的用户的教育程度、年龄具有多样性特点; 用户对与当前物品同类的物品没有低于评分均值的评分. 认知风险大的特征为: 用户对与当前物品同类的物品存在低于评分均值的评分; 当前物品的平均评分小于其它物品的平均评分; 图书类型与教育程度差别过大 (儿童刊物与大学以上教育程度); 发行年份为十年之前.

4.2 实验结果度量标准

我们采用 $P@R^{[10]}$ 作为偏好获取准确性的度量标准, $P@R$ 即根据偏好获取方法所获取的当前用户偏好值的大小为其进行排序, 为其推荐 Top- R 个项目, 并将此 Top- R 个项目与测试集中用户常用的 Top- R 个项目进行对比, 以计算推荐准确率, 其值越大表明推荐的准确率越高.

$$P@R = \frac{\text{Top-}R \text{ 推荐项目集所包含测试集 Top-}R \text{ 种项目的数量}}{R}.$$

4.3 实验设计及结论

实验 1. 3 种认知因素影响分析.

本次实验是针对文中所提出的 3 种认知因素对于用户偏好的影响类型而进行的, 目的是确定 3 种认知因素对推荐性能的影响类型与程度. 经过反复测试与筛选, 我们选取以 60% 的 MBookCrossing 数据为训练集, 其余为测试集, 相似上下文个数 $k=6$, λ_1 与 λ_2 的取值空间为 $[0, 10]$, 进行本次实验. 经过反复测试, 确定实验方案为:

(1) 针对认知风险, 我们固定参数 λ_1 与 λ_2 的取值, 通过改变 φ 的值来度量其对 $P@R$ 的影响, 由于

推荐结果的获取依赖于 Pre_{jis} 的相对大小, λ_1 与 λ_2 的取值不会对本次实验结果产生影响, 因此我们选取 $\lambda_1 = \lambda_2 = 5$ 进行本次实验;

(2) 针对认知水平, 我们固定参数 φ 与 λ_1 的取值, 通过改变 λ_2 来度量其对 $P@R$ 的影响, 为了既能覆盖较大的认知风险与认知水平的权值比, 又要在两个参数的取值空间中, 选取 $\varphi = 0.5, \lambda_1 = 5$;

(3) 针对认知有用性, 选取参数 $\varphi = 0.5, \lambda_2 = 5$ (取值原因同(2)), 通过改变 λ_1 来度量其对 $P@R$ 的影响。

每个实验的纵坐标为选取在 $R = 4, 5, 6, 7, 8$ 上的 $P@R$ 均值, 经过反复实验, 针对当前数据集我们选取了几组有代表性的实验结果, 如图 2~图 4 所示。

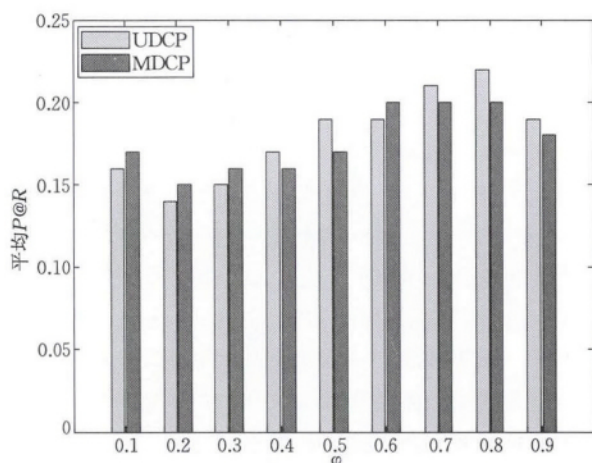


图 2 认知风险对偏好获取结果的影响

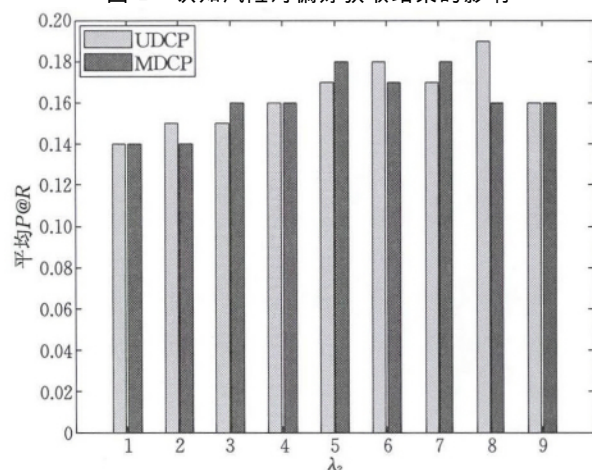


图 3 认知水平对偏好获取结果的影响

在实验中, 参数 $\varphi, \lambda_1, \lambda_2$ 取值范围的选取结合了 3 种认知因素在当前数据集环境下具体取值, 并使其能够较为广泛的覆盖 3 个参数的取值空间。从实验结果可以看出, 随着 3 个参数取值的增大, 平均 $P@R$ 的值在总体上呈现递增的趋势, 即说明了认

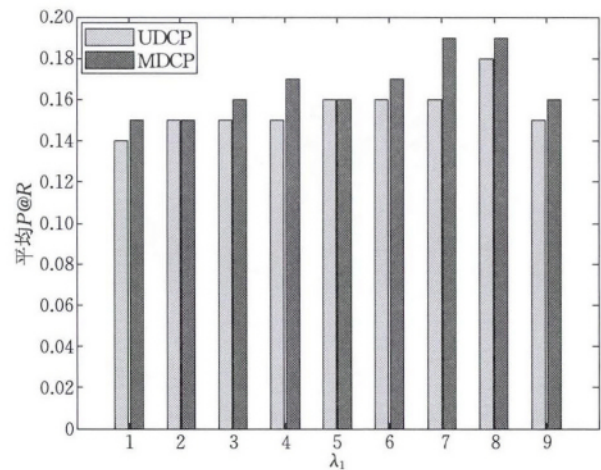


图 4 认知有用性对偏好获取结果的影响

知风险对于用户偏好获取为负相关关系, 认知有用性与认知水平相对用户偏好获取为正相关关系, 这也验证了式(10)假设的合理性。

实验 2. 数据集与参数影响实验。

针对当前数据集, 我们分别提取 20%、40%、60%、80% 的 MBookCrossing 数据集作为训练集, 并在 4 种数据集模式进行实验, 在 4 种模式下, 经过反复测试与筛选选取波动参数 $\varphi = 0.5$, 相似上下文个数 $k = 6$, 并针对单维上下文偏好提取算法 UDCP 与融合多维上下文的偏好提取算法 MDCP 分别对参数 λ_1, λ_2, R 对于推荐准确率的影响进行了反复实验, 并选取了具有代表性的实验结果, 如表 1~表 3 所示。

表 1 $\lambda_1 = 3, \lambda_2 = 5$ 时实验数据

不同数据规模下 算法对比		R				
		4	5	6	7	8
UDCP	20%	0.25	0.22	0.17	0.14	0.13
	40%	0.32	0.31	0.18	0.24	0.14
	60%	0.23	0.23	0.20	0.18	0.14
	80%	0.28	0.28	0.28	0.25	0.16
MDCP	20%	0.37	0.47	0.48	0.43	0.34
	40%	0.37	0.46	0.49	0.43	0.36
	60%	0.36	0.46	0.47	0.44	0.33
	80%	0.34	0.36	0.37	0.34	0.33

表 2 $\lambda_1 = 5, \lambda_2 = 1$ 时实验数据

不同数据规模下 算法对比		R				
		4	5	6	7	8
UDCP	20%	0.28	0.21	0.16	0.13	0.13
	40%	0.29	0.23	0.19	0.16	0.19
	60%	0.24	0.24	0.19	0.17	0.14
	80%	0.28	0.25	0.22	0.19	0.21
MDCP	20%	0.33	0.36	0.35	0.31	0.19
	40%	0.37	0.37	0.36	0.35	0.29
	60%	0.33	0.37	0.32	0.37	0.28
	80%	0.36	0.38	0.36	0.36	0.27

表 3 $\lambda_1=5, \lambda_2=3$ 时数据对比

不同数据规模下 算法对比		R				
		4	5	6	7	8
UDCP	20%	0.25	0.21	0.16	0.16	0.13
	40%	0.29	0.31	0.26	0.24	0.14
	60%	0.24	0.23	0.21	0.18	0.15
	80%	0.32	0.32	0.28	0.27	0.19
MDCP	20%	0.34	0.38	0.38	0.37	0.33
	40%	0.34	0.34	0.33	0.33	0.28
	60%	0.35	0.34	0.38	0.38	0.34
	80%	0.32	0.36	0.37	0.34	0.29

表 1~表 3 分别展示了 4 种数据集模式下, 参数 λ_1, λ_2 在不同 R 值处对算法 UDCP 与 MDCP 偏好提取精度的影响. 分别对比 3 个表中数据可以发现, λ_1 与 λ_2 保持其中一个权值不变而增加另一个, 大多都能够一定程度上提高推荐结果的准确率. 这说明在认知风险固定的情况下, 认知水平与认知有用性对于推荐结果的准确率是一种正相关的影响, 也即与实验 1 结论保持一致. 综合考虑实验数据可以发现, MDCP 的整体推荐准确率要优于 UDCP, 这说明考虑多维上下文因素的偏好获取算法要比单维偏好获取算法能够取得更好的推荐质量. 从实验结果可以看出, 以 $P@R$ 作为评价标准, 随着 R 的增加, 推荐准确率大多为先增后减, 当 $\lambda_1=3, \lambda_2=5, R=5$ 或 6 时两种算法的整体准确率较高.

实验 3. 与现有偏好获取算法对比实验.

经过反复实验对比, 我们取 $\lambda_1=3, \lambda_2=5$, 并选取 20% 与 60% 两种训练集比例进行实验, 将本文的 UDCP 及 MDCP 算法, 与经典的协同过滤算法 (CF)^[5] 及基于上下文情景感知的推荐算法 (IMLBS)^[4] 进行实验对比, 其中由于 CF 算法仅考虑“用户-项目”间的相互影响, 因此本次实验中其仅采用相同数据集中的 MBX-Book-Ratings 部分. 对比结果如图 5~图 6 所示.

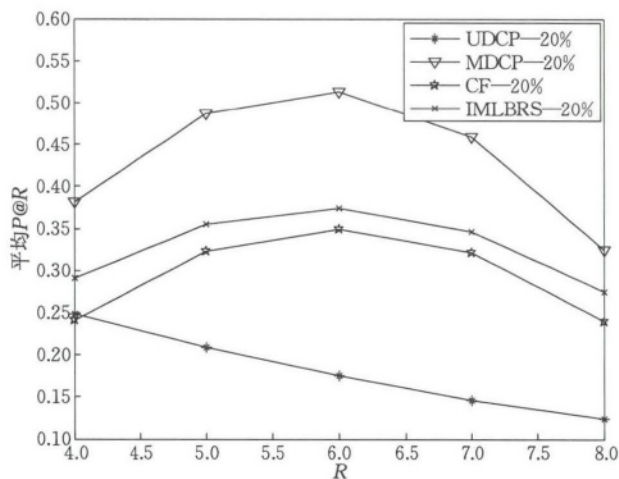


图 5 20% 作为训练集几种算法对比的结果

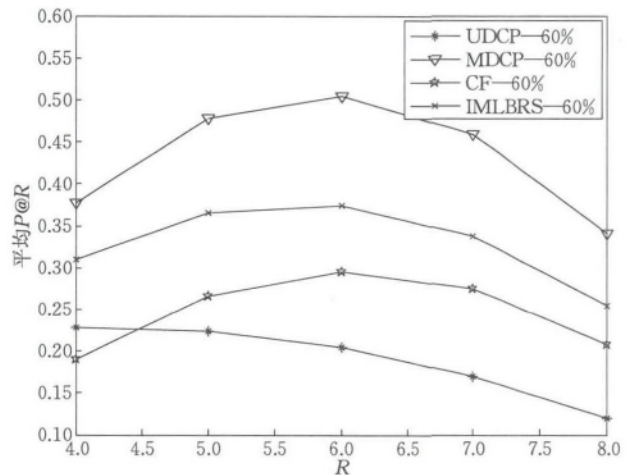


图 6 60% 作为训练集几种算法对比的结果

从图 5~图 6 可以看出, 以 $P@R$ 为评价标准, (1) IMLBS 算法及本文提出的算法整体上优于 CF 算法, 这说明在获取推荐结果时考虑上下文信息对用户偏好的影响比单纯的依靠“用户-评分”信息的协同过滤算法, 有助于提高用户偏好提取的准确度, 即是用户所做出的每次选择在较大程度上会受到当前上下文环境的影响; (2) MDCP 算法在一定程度上优于其他 3 种偏好获取算法, 这说明考虑多维上下文信息对用户认知行为的影响, 能够提高预测的准确度, 并且只有少部分用户的偏好不受上下文信息的影响; (3) 融合多维上下文的 MDCP 算法优于只考虑单维上下文的 UDCP 算法, 因为在不同的上下文环境下, 用户的偏好存在一定的差异性, 其偏好行为受多种上下文环境的交互影响, 并且由结果可知, 考虑多维上下文环境对用户的影响能够获取到更为精确的用户偏好.

5 结束语

推荐系统是缓解“信息过载”问题的有效手段, 具有重要的学术研究价值与工业研究价值. 针对现有的上下文感知推荐系统中, 未考虑不同用户对于各类型的上下文信息的认知行为与用户偏好间的内在联系及对用户偏好预测准确度不高的问题, 本文受“分布式认知理论”及“记忆信息加工模型”启发, 提出了基于用户认知行为的上下文偏好获取算法 UDCP 与 MDCP. 该方法通过分析各种上下文信息对用户认知行为的影响, 将认知水平、认知有用性、认知风险、有效认知行为等认知领域概念融入偏好获取流程, 并通过构建其相互间约束修正关系, 逐步提取更为精确的用户偏好. 在真实数据集上的仿真

实验结果表明,该方法能够有效的挖掘用户的认知行为与其偏好之间的关联关系,以 $P@R$ 为评价标准,偏好获取的精确度优于经典的协同过滤算法及上下文感知算法. 在以后的研究工作中,将重点研究以用户群体间认知行为的相互影响关系为基础,融合粒子群或遗传算法等优化算法的社会化认知偏好获取算法.

参 考 文 献

- [1] Kim J, Lee D, Chung K Y. Item recommendation based on context-aware model for personalized u-healthcare service. *Multimedia Tools and Applications*, 2013, 67(3): 1-18
- [2] Van Setten M, Pokraev S, Koolwaaij J. Context-aware recommendations in the mobile tourist application COMPASS // *Proceedings of the 3rd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*. Eindhoven, Netherlands, 2004: 235-244
- [3] Adomavicius G, Tuzhilin A. *Recommender Systems Handbook: Context-Aware Recommender Systems*. New York: Springer, 2011
- [4] Chen C M. An intelligent mobile location-aware book recommendation system that enhances problem-based learning in libraries. *Interactive Learning Environments*, 2013, 21(5): 469-495
- [5] Wei S, Ye N, Zhang S, et al. Collaborative filtering recommendation algorithm based on item clustering and global similarity // *Proceedings of the 5th International Conference on Business Intelligence and Financial Engineering (BIFE)*. Lanzhou, China, 2012: 69-72
- [6] Cantador I. An enhanced semantic layer for hybrid recommender systems. *Semantic Web: Ontology and Knowledge Base Enabled Tools Services and Applications*, 2013, 72(4): 235-238
- [7] Wang Li-Cai, Meng Xiang-Wu, Zhang Yu-Jie. Context-aware recommender systems. *Journal of Software*, 2012, 23(1): 1-20 (in Chinese)
(王立才, 孟祥武, 张玉洁. 上下文感知推荐系统. *软件学报*, 2012, 23(1): 1-20)
- [8] Tu Dan-Dan, Shu Cheng-Chun, Yu Hai-Yan. Using unified probabilistic matrix factorization for contextual advertisement recommendation. *Journal of Software*, 2013, 24(3): 454-464 (in Chinese)
(涂丹丹, 舒承椿, 余海燕. 基于联合概率矩阵分解的上下文广告推荐算法. *软件学报*, 2013, 24(3): 454-464)
- [9] Wang Yu-Xiang, Qiao Xiu-Quan, Li Xiao-Feng, et al. Research on context-awareness mobile SNS service selection mechanism. *Chinese Journal of Computers*, 2010, 33(11): 2126-2135 (in Chinese)
(王玉祥, 乔秀全, 李晓峰等. 上下文感知的移动社交网络服务选择机制研究. *计算机学报*, 2013, 33(11): 2126-2135)
- [10] Wang Li-Cai, Meng Xiang-Wu, Zhang Yu-Jie. A cognitive psychology-based approach to user preferences elicitation for mobile network services. *Acta Electronica Sinica*, 2011, 39(11): 2547-2553 (in Chinese)
(王立才, 孟祥武, 张玉洁. 移动网络服务中基于认知心理学的用户偏好提取方法. *电子学报*, 2011, 39(11): 2547-2553)
- [11] Kuo M H, Chen L C, Liang C W. Building and evaluating a location-based service recommendation system with a preference adjustment mechanism. *Expert Systems with Applications*, 2009, 36(2): 3543-3554
- [12] Dey A K. Understanding and using context. *Personal and Ubiquitous Computing*, 2001, 5(1): 4-7
- [13] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6): 734-749
- [14] Cole M, Engeström Y. *A Cultural-Historical Approach to Distributed Cognition*. Distributed Cognitions: Psychological and Educational Considerations. USA: Cambridge University Press, 1993
- [15] Craik F I M, Lockhart R S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11(6): 671-684
- [16] Kommers P A M, Jonassen D H, Mayes J T. *Cognitive tools for learning (Nato ASI Series (closed), Nato ASI Subseries F)*. Heidelberg: Springer Publishing Company, Incorporated, 2012
- [17] Forsyth P A, Kennedy J S, Tse S T, et al. Optimal trade execution: A mean quadratic variation approach. *Journal of Economic Dynamics and Control*, 2012, 36(12): 1971-1991
- [18] Ziegler C N, McNee S M, Konstan J A, et al. Improving recommendation lists through topic diversification // *Proceedings of the 14th International Conference on World Wide Web*. Chiba, Japan, 2005: 22-32



GAO Quan-Li, born in 1988, Ph.D. candidate. His research interests focus on recommender systems and data mining.

GAO Ling, born in 1964, Ph.D., professor. His current research mainly concentrates on services computing and computer network.

YANG Jian-Feng, born in 1981, Ph.D. candidate, engineer. His research interests focus on personalized recommendation.

WANG Hai, born in 1977, Ph.D., lecturer. His main research interests include machine learning and services computing.

Background

Traditional recommender systems have been extensively used in various applications to make recommendations based on users' history of preferences. There are many researchers focus on this field and figure out many solutions to get the items that fit users' personal habits and preferences, such as collaborative filtering, singular value decomposition, mathematic models. But the existing methods have been widely suffered from the sparse and cold-start problems, and in some applications failure to consider the users' current situations may result in considerable performance degradation in recommendation effectiveness because users may have different preferences for items in different problems.

In this paper, we study the problems which belong to the field of recommender systems, and it is also related to intelligent computing. We proposed a preference elicitation method based on users' cognitive behavior for context-aware recommender system to overcome those drawbacks that mentioned above. Firstly, this method analysis the influence the multidimensional context environment and users' and

items' objective attributes. Secondly, it puts forward many calculation methods of concept models such as cognitive level, cognitive usefulness, cognitive risk, effective cognitive and so on. Finally, we coalesce those concept models to the users' preference eliciting processing to get the users' personal preference items. The experimental results show that the methods that mentioned above can achieve better recommendation accuracy.

This work is supported by the National Natural Science Foundation of China under Grant No. 61373176; the Ph. D. Programs Foundation of Ministry of Education of China under Grant No. 20116101110016; the Shaanxi Provincial Natural Science Foundation of China under Grant No. 2012JQ8047; the Foundation of Shaanxi Educational Committee under Grant No. 11JK1059; the Foundation of Shaanxi Science Innovation under Grant No. 2012ZKC05-2; the Foundation of 2013 Science Research of Shaanxi Educational Committee under Grant No. 2013JK1178.