

基于多视图群集认知优化的情景感知推荐方法

高全力¹⁾, 高岭^{1),2)}, 曹瑞²⁾, 杨建锋²⁾

¹⁾(西安工程大学计算机科学学院 西安 710048)

²⁾(西北大学信息科学与技术学院 西安 710127)

摘要 现有的情景感知推荐方法, 仅从单视图角度获取用户偏好, 并且未考虑在社会化网络环境下用户认知心理与其偏好行为间的内在联系, 导致了推荐结果无法有效的覆盖用户潜在的偏好特征, 难以建立准确的用户偏好模型, 影响了推荐质量. 受分布式认知模型及多视图学习理论启发, 提出了一种基于多视图群集认知优化的情景感知推荐方法. 首先基于用户的真实社交行为(通过记录、位置记录等)与历史偏好行为记录(包含上下文的评分、浏览等行为), 建立用户的社交关系网络, 并从中提取用户的认知行为特征, 获取个体认知行为. 然后, 基于社交关系网络与个体认知行为, 从社会化认知的角度获取用户认知行为间的关联关系, 建立了包含认知依赖、认知风险、群集认知等因素的认知关系网络模型. 并从社会关系本身属性出发, 提出了一种包含关系传递、共点增强、约束传递等性质的增强策略, 以提高认知数据密度. 最后, 从用户属性、认知关系、项目认知价值等视图出发, 通过特征融合、投影变换并引入 Fisher 判别准则, 获取用户的认知偏好, 并建立起基于多视图优化的偏好获取模型. 在大规模真实数据集上的实验结果显示, 与现有方法 CABT、CABT、WSDM 等相比, 推荐准确度平均提升 15.76%, 多样性提高近 41.03%, 表明所提出方法能够取得更好的推荐质量, 并且有更大的概率覆盖用户的潜在偏好特征.

关键词 多视图学习; 认知行为; 群集认知; 线性判别分析; 上下文感知

中图法分类号 TP391

Context-aware Recommendation via Multi-view Cluster Cognition Optimization

GAO Quan-Li¹⁾, GAO Ling^{1),2)}, CAO Rui²⁾, YANG Jian-Feng²⁾

¹⁾(Department of Computer Science, Xi'an Polytechnic University, Xi'an 710048)

²⁾(Department of Information Science and Technology, Northwest University, Xi'an 710127)

Abstract To solve the problem that the existing methods acquire users' preference only from single view, which have not consider the internal relation between users' cognitive psychology and their preference behavior under the social network environment, this may lead to the incapable of recommend results to cover users' latent preference features, and the difficult for the existing methods to build the accuracy users' preference model, the two insufficient lead to the low predictive quality of the recommender system. Inspired by distributed cognitive model and multi-view learning theory, a context-aware recommendation method based on multi-view cluster cognitive optimization model is proposed in this paper. In the beginning, the proposed method establish the users' social relation network based on the user's real social behavior(call records and location records) and historical preference behavior records(include rate,browse under context environment), and it extract the features of users' cognitive behavior from their historical preference records to acquire the individual cognitive behavior. This method measures the incidence relation among users' cognitive behavior from the view of social cognition, which based on the social network and the individual cognitive behavior, and it build the cognitive relation model which includes reliant cognition, cognition risk and cluster cognition, and

本课题得到国家自然科学基金(No.61373176, No.61572401, No.61672426); 陕西省工业公关项目(2014K05-42); 陕西省教育厅科研基金(No.11JK1059); 陕西省教育厅 2013 年科学研究计划(No.2013JK1178)资助. 高全力, 男, 1988 年生, 博士, 讲师, 主要研究领域为智能推荐系统, E-mail: gaoquanli@nwu.edu.cn. 高岭, 男, 1964 年生, 博士, 教授, 计算机学会(CCF)高级会员 (E200005713S), 主要研究领域为计算机网络、服务计算, E-mail: gl@nwu.edu.cn. 曹瑞, 男, 1991 年生, 硕士研究生, 主要研究领域为智能推荐系统, E-mail: caorui@stumail.nwu.edu.cn; 杨建锋, 男, 1981 年生, 博士研究生, 工程师, 主要研究领域为智能推荐系统, E-mail: yangjianfeng@nwu.edu.cn.

so on. Furthermore, the authors proposed an enhancement strategy that includes the properties of relational transfer, covariance enhancement, constraint transfer and so on, which based on the view of social relations, its purpose is to improve the data density of users cognitive which can solve the sparsity problem of the recommender systems. Then the method uses feature fusion, projection transformation and Fisher discriminant criterion to acquire multi-view optimal projection direction from the views of users' properties, cognitive relations and items' cognitive values. Finally, the proposed method use the optimal projection direction to compute the cognitive preference, and basing on which the method get the users' preference model based on the multi view optimization of the above multidimensional data, and the recommendation list is generated based on the users' preference model. Experiments which include two main parts(the parameter test of the proposed method and the compare with existing methods) are carried out based on a real large scale data set(which is divided into different test set and training set by randomly hiding the data), in addition we adopt both P@R and diversity as the evaluation metrics, which are two popular evaluation metrics for preference prediction models at present, to evaluate the predicting accuracy for missing ratings of different recommendation models. The experimental results show that the proposed algorithm can achieve better prediction accuracy and diversity about 20.19% and 44.89 respectively , which are compare with the existing methods CABT(context aware recommend method)、CACF(the modified collaborative filtering method which is the most widely used recommendation algorithm) and WSDM(the context aware method which consider social relations data) . It has been demonstrated that the proposed method can achieve better recommendation quality, and has bigger probability to cover users' latent preference.

Key words multi-view learning ; cognitive behavior; cluster cognition; linear discriminant analysis
context-aware

1 引言

大数据背景下, 由于推荐系统可以主动的为用户推送符合其偏好的信息资源, 能够缓解“信息过载”问题对于内容服务的影响, 受到了学术界与工业界的广泛关注^[1]. 传统的推荐系统多利用历史偏好记录构建用户偏好模型, 并据此为用户生成推荐服务. 例如, 高明等^[2]提出了一种采用隐语义主题模型的喜好微博获取方法, 能够实时、连续的为用户推荐符合其偏好的微博内容; Baltrunas 等^[3]提出了一种考虑项目划分与聚合距离的改进协同过滤方法, 能够有效的提高服务推送的准确率; Wu Hui 等^[4]提出了一种基于用户间信任度阈值, 来改进传统协同过滤推荐系统中最近邻的获取方法.

上述推荐系统仅关注于“用户-项目”间的二维关系(用户和项目), 多以用户之间的历史偏好数据为核心, 通过度量用户偏好行为间的相互影响关系, 建立用户的偏好预测模型. 由于用户偏好的产生受多种内在与外界因素的影响, 用户的职业、年龄、所处上下文环境等都会影响用户的偏好决策^[5]. 并且对于同一用户而言, 在不同上下文环境影响下, 其偏好会产生相应的波动^[5], 以位置上下文的影响为例, 当位置属性为“家”时, 用户会倾向于“电视节目”, 当位置属性为“公共交通”时, 则会更倾向于“音乐”、“新闻资讯”等. 而上述推荐

系统由于无法辨别用户的位置等上下文环境间的差异, 影响了推荐结果的准确度, 难以取得高质量的推荐效果.

随着“普适计算”等计算方法的提出, 系统能够主动发现并处理用户状态、所处位置等上下文信息为用户提供服务, 上下文情景感知推荐系统逐渐成为了新的研究热点. 例如, Alhamid^[5]等采用情景信息融合偏好获取流程, 能够获取用户的隐含偏好; Huang 等^[6]提出了一种 Pairwise 学习方法融合用户与项目的上下文信息; Hong 等^[7]基于角色的信任网络能够建立起准确的用户偏好模型; Feng 等^[8]提出了 Gray 预测模型, 能够缓解内容推荐所面临的数据稀疏性问题.

这些方法多是从用户或者项目等角度结合上下文环境获取用户偏好, 也即是从单视图的角度生成推荐结果(即从单个视图下对建立用户偏好模型, 并据此生成推荐结果). 然而在实际的应用中, 推荐结果的产生需要平衡用户、内容提供商、服务提供商等各方面的利益; 并且由于推荐系统普遍面临的数据稀疏性问题, 基于单视图的偏好获取方法, 无法更有效的覆盖用户的潜在偏好特征和隐藏的兴趣点, 导致了偏好预测的适用性低、准确度差等问题. 并且由于用户的偏好行为是在其认知心理的支配下形成的, 是认知心理在行为层面的直观反

映, 只对用户的偏好行为进行分析处理, 而不考虑其认知心理及相关的认知影响因素, 也在一定程度上影响了偏好获取的准确性。

为解决上述问题, 本文提出了一种融合认知心理与多视图学习的偏好预测方法。多视图指的是对于同一事物可以从多种不同的途径或不同的角度对其进行描述, 此处的多种描述即构成事物的多个视图。而单视图仅是上述多种表述中的一种描述。所提出方法的主要流程为: 基于社交关系网络与个体认知行为, 从社会化认知的角度获取用户认知行为间的关联关系, 建立了包含认知依赖、认知风险、群集认知等因素的认知关系网络模型, 并从用户属性、认知关系、项目认知价值等视图出发, 通过特征融合、投影变换并引入 Fisher 判别准则, 建立基于多视图优化的认知偏好模型, 并基于此为用户生成推荐结果。最后以大规模真实数据为实验对象, 验证了该方法的有效性。即方法能够提高推荐质量, 有助于缓解大数据环境下的“信息过载”问题。

本文的主要贡献如下:

(1) 提出了一种基于位置交互、通话交互等真实社交关系的社交关系网络建立方法, 并建立起基于关系传递、共点增强等修正策略的优化模型;

(2) 将个体认知行为、认知依赖、认知风险、群集认知等认知领域概念引入偏好获取流程, 并建立起基于认知心理与社交关系的认知关系网络模型;

(3) 提出了一种采用多视图学习融合个体认知行为、集群认知、项目认知价值等多维数据的偏好获取模型, 并在大规模真实数据集上对比验证。

2 相关工作

多视图学习主要适用于样本空间可以采用多视图描述的学习问题, 研究起源于 Yarowsky^[9]用多视图解决图像样本特征分类问题, 其定义比较宽泛, 一般只要满足“学习的数据可以用多视图描述”, 即可称为多视图学习问题。其主要是依据无监督数据的视图一致性, 也即是分类器在一无监督数据不同视图下的分类结果应基本保持一致。目前是样本分类、模型优化与半监督学习等领域的热门研究方向, 例如, 将多视图学习用户复杂标签样本分类^[10]、手写数字识别^[11]、人脸识别^[12]等。在推荐领域, 推荐的过程可以看作是样本分类的过程, 即是以用户偏好为分类特征, 将相符的项目分类到相应的用户偏好集合中。并且, 相关的样本数据可以采用多视图来描述: 可以从用户、项目、内容提供商等视图对相关数据进行描述, 也即是推荐的产生可以转换为多视图优化问题。例如, Chechev 等^[13]提出了一种基于多视图的内容推荐策略, 能够为用户在 Twitter 上推荐感兴趣的人; Lu 等^[14]提出了采用多视图学习融合评分信息、社交信息、项目信息, 提高了用户偏好模型的准确性。

另一方面, 现有的用户偏好建模方法, 多从行为层面挖掘用户偏好行为间的关联关系、用户历史偏好对当前偏好的影响程度, 进而预测用户的偏好趋势。而用户的偏好行为是在其认知心理的支配下形成的, 基于认知心理建立用户的偏好模型, 将能够更深入的挖掘出影响用户偏好的主要因素。因此, 近年来开始有研究者将认知心理引入偏好获取过程, 例如, 将认知水平、认知有用性等认知领域概念融入偏好获取流程^[15]; 将记忆两过程理论与认知心理融合, 建立认知心理对于用户偏好的影响关系模型, 进而为用户生成推荐^[16]; 采用心理学模型进行多媒体的内容推荐^[17]等。

基于上述分析, 多视图学习理论及认知心理学均为推荐系统领域的主流研究方法, 并能有效解决现有推荐系统存在的缺陷和不足, 开始受到学术界与工业界的广泛关注, 并逐渐的成为该领域最新的发展方向。而上述方法, 未能同时考虑多视图学习理论与社会化网络环境下的认知行为, 影响了推荐质量。基于线性判别分析算法^[10]以及认知心理^[16], 本文基于认知心理与社交关系获取群集认知行为, 并通过将带有 Fisher 准则的线性判别分析应用到了推荐领域, 建立起用户的认知偏好模型, 并据此为用户生成推荐服务。

3 基于多视图群集认知优化的情景感知推荐方法

本文推荐问题的定义为: 给定用户集合 U , 项目集合 I , 以及包含上下文的历史偏好记录 H , 预测结果 R , 则

$$U \times I \times H \rightarrow R$$

其中, R 采用最优分类准则获取。

本文所提出推荐方法的核心是获取用户的认知偏好, 通过认知偏好来为用户生成推荐结果。而情景感知主要体现在相比于传统的推荐系统考虑了上下文信息。本文的上下文信息包括用户的位置、时间等内容, 基于此获取认知偏好, 并生成推荐结果等。

方法的核心流程为: 首先用户的历史行为记录, 建立起一种有向有权的社交关系网络模型; 其次基于社交关系网络模型与个体认知行为建立起多种认知关系交互影响的认知关系网络模型; 然后从上述两模型中获取到各视图下的特征数据, 并基于 FLDA 进行特征融合、投影变化, 获取最优的认知偏好; 最后根据认知偏好生成推荐结果。整体总结流程如图 1 所示。

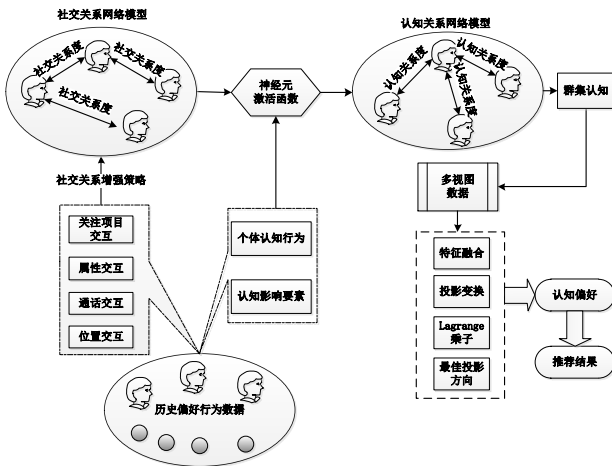


图1 基于群集认知优化的偏好获取总体结构

3.1 认知关系模型的建立

本节基于个体认知行为、认知交互影响关系、社交关系网络等建立起用户间认知关系网络模型,具体过程及定义如下。

定义 1. 个体认知行为,指的是用户个体在内在与外在上下文因素影响下,并受制于个人认知能力所产生的一种对外界事物的认知过程。其中,认知的过程表示用户对于各种上下文信息加工处理的过程;认知能力指的是人脑加工、存储与处理相关信息的能力,影响用户的需求与偏好选择^[14]。由于认知水平表示用户在外在上下文环境与内在因素影响下对外界(一般指商品、项目等)的接受与认可程度。那么对个体认知行为的度量就可以转换为对认知水平的度量,影响认知水平的主要因素包括:(1)所处上下文环境,主要包括周围环境、时间状态、天气状态等;(2)历史经验,主要指用户对于与当前项目同类的其他项目的历史行为所引起的喜好倾向,也可表述为对某类项目有无先入性认知行为;(3)项目属性,主要有内容信息、规格信息、价值信息等;(4)用户属性,主要有教育信息、年龄信息、职业信息等。上述因素对用户认知水平的影响可分为正相关因素与负相关因素,正相关因素指的是能够增强用户对当前项目喜好程度的因素,例如,用户对于相关项目有较多的喜好行为记录,或当前用户的社交群集中多数用户的历史经验一致等。负相关因素指的是对当前用户偏好有消极影响的因素,例如,历史喜好记录波动较小等。显然认知水平的度量是上述两因素的正反比例函数,具体如下:

$$Cl_a = \frac{\sum_{m \in pc} (Pc_a)_m}{\beta + \sum_{n \in nc} (Nc_a)_n} \quad (1)$$

其中, Cl_a 表示用户 U_j 对于项目 S_s 的认知水平,且 $Cl_{js} \in CD$, Pc_a 与 Nc_a 分别表示其正、负相关因素, pc 与 nc 分别表示正、负相关因素集,其获取方式有两种,一是由多数用户的历史经验;二是根据与

具体应用环境相对应的构造规则来生成。 β 为引入的修正因子,取值如下:

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

定义 2. 认知关系网络模型,是基于用户的社交关系网络与个体认知行为,建立用户间认知层面的交互关系模型,也即是从社会化角度来度量用户间的认知关系、用户群集的认知状态、认知交互影响关系等。主要过程包括社交关系网络的建立,并基于此建立认知关系网络。

其中,社交关系网络指的是基于真实用户间的社交关系,所形成的涉及用户情感、社会背景、思想观念等社会因素的人际关系模型。目前,社交关系网络的建立方法主要使用平均权重、专家经验^[18]等主观方法,并且传统的社会网络构建方法把社会网络看作无向有权图^[19],而从个人情感的角度而言,社会网络关系应是一种有向有权图,用户间的社会关系认知并非等值的。例如,用户 A 认为用户 B 为“知心好友”,而用户 B 可能认为用户 A 只是其“普通好友”等,上述因素在一定程度上影响了模型建立的客观性与准确性^[20]。

本文利用用户间真实的交互关系提出了一种有向有权的社交关系网络,用有向有权图 $G=(V,E)$ 表示,如图 2(图 1 中此部分为方便表示,合并了双向箭头),其中 V 指代用户集合, $V=\{U_i\}, i=1,2,3,\dots,n$, U_i 为任一用户, $E=\{S_j | j \in V\}$ 为社交关系集合,在有向有权图 G 中, S 以有值邻接矩阵的形式存储用户间的双向社交关系, $S=\{S_{U_i \rightarrow U_j} | U_i, U_j \in V\}$, $S_{U_i \rightarrow U_j}$ 表示用户 U_i 对 U_j 的社交关系度值,其定义如下:

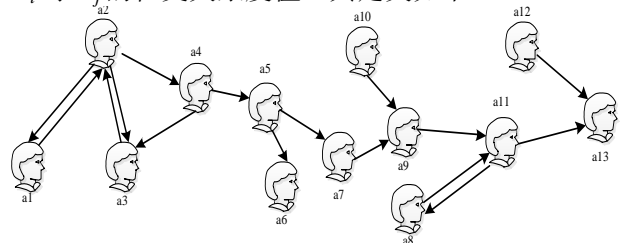


图2 社交关系网络示例

定义 3. 社交关系度,表示社交关系网络中的任两用户间社交关系度量值。其具有不对称性(方向性和差异性)、动态性等特征,它是不同用户间各种交互行为的内在体现。本文提出了一种基于交互因子的度量方法,交互因子包含通话交互、位置交互、关注项目交互、基本属性交互等。

(1)属性交互:指的是用户职业、年龄等属性信息所反映出的社交关系,例如:年龄差距较小、相同职业等特点的用户间有更大的概率具有较大的社交关系度;

(2)通话交互:指的是用户间通话频次、时间等;

(3)位置交互：指的是用户在相同时间段内，处于同一区域的频次与时长；

(4)关注项目交互：指的是用户间的历史关注项目集合的重合程度，反映的是用户是否对同一项目或同类项目有无相同的喜好行为。

其中，属性交互反映的是用户的自然属性对社交关系的影响；通话交互与位置交互是用户间社交关系度的行为层面的直观体现；关注项目交互则是指由用户偏好行为对社交关系度的影响。四种交互因子涵盖了用户社交行为的主要影响因素，那么基于交互因子的社交关系度的度量方法可以表示为

$$S_{U_i \rightarrow U_j} = \alpha_1 \cdot B_{U_i \rightarrow U_j} + \alpha_2 \cdot T_{U_i \rightarrow U_j} + \alpha_3 \cdot L_{U_i \rightarrow U_j} + \alpha_4 \cdot K_{U_i \rightarrow U_j} \quad (3)$$

$$B_{U_i \rightarrow U_j} = \frac{B_{ij}}{B_i} \quad (4)$$

$$T_{U_i \rightarrow U_j} = \sqrt{\frac{|T_t(U_i, U_j) - \overline{T_t(U_i)}|}{T_t(U_i, U_j) + \overline{T_t(U_i)}} \times \frac{|T_p(U_i, U_j) - \overline{T_p(U_i)}|}{T_p(U_i, U_j) + \overline{T_p(U_i)}}} \quad (5)$$

$$L_{U_i \rightarrow U_j} = \frac{|(L_t(U_i, U_j) - \overline{L_t(U_i)})(L_p(U_i, U_j) - \overline{L_p(U_i)})|}{L_t(U_i) + L_p(U_i)} \quad (6)$$

$$K_{U_i \rightarrow U_j} = \frac{K_{ij}}{K_i} \quad (7)$$

由于公式中参数较多，所以用列表的方式来表述各参数的具体含义，具体如表1所示。

表1 参数含义表

参数名	含义
$\alpha_1, \alpha_2, \alpha_3, \alpha_4$	修正参数
$B_{U_i \rightarrow U_j}$	U_i 与 U_j 间基本属性交互值
$T_{U_i \rightarrow U_j}$	U_i 与 U_j 间通话交互值
$L_{U_i \rightarrow U_j}$	U_i 与 U_j 间位置交互值
$K_{U_i \rightarrow U_j}$	U_i 与 U_j 间关注项目交互值
B_{ij}	用户 U_i 与 U_j 间基本属性重合(包含相同与相近)个数
$T_t(U_i, U_j)$	U_i 与 U_j 间通话时长
$\overline{T_t(U_i)}$	用户 U_i 与其他用户的平均通话时长
$T_p(U_i, U_j)$	U_i 与 U_j 间通话频次
$\overline{T_p(U_i)}$	用户 U_i 与其他用户的平均通话频次
$L_t(U_i, U_j)$	用户 U_i 与 U_j 处于同一位置的时长
$L_p(U_i, U_j)$	用户 U_i 与 U_j 处于同一位置的频次
$\overline{L_t(U_i)}$	用户 U_i 与其他相关用户处于同一地点的平均时长
$\overline{L_p(U_i)}$	用户 U_i 与其他相关用户处于同一地点的平均频次
K_{ij}	用户 U_i 与 U_j 间共同行为项目数
K_i	用户 U_i 的历史行为项目数

在四种交互因子中， $T_{U_i \rightarrow U_j}$ 与 $L_{U_i \rightarrow U_j}$ 由于考虑了当前用户与对应的其他用户间交互关系，其值具有双向性，即 $T_{U_i \rightarrow U_j} \neq T_{U_j \rightarrow U_i}$ ， $L_{U_i \rightarrow U_j} \neq L_{U_j \rightarrow U_i}$ 。

公式(3)中，参数估计方法如下：

任意用户 U_i 与 U_n 间的社交关系度获取方法可由下式表示

$$S_{U_i \rightarrow U_n} = \sum_{j=1}^4 C_{ij} \cdot a_j \quad (8)$$

其中， C_{ij} 表示上述四种交互因子($j=1,2,3,4$)，对于任意用户而言，存在社交关系度的用户数要大于参数的数量，则上述方程为超定方程，可由最小二乘法求解最优参数取值，引入残项平方和函数，如下

$$S'(\alpha) = \sum_{n \in RU_i} \left\| \sum_{j=1}^4 (\alpha_j \cdot C_{ij} - S_{U_i \rightarrow U_n}) \right\|^2 \quad (9)$$

其中， RU_i 表示与 U_i 存在社交关系的用户集合。

假设 $\alpha = \hat{\alpha}$ 时， $S'(\alpha)$ 取得最小值，则有

$$\hat{\alpha} = \arg \min(S'(\alpha)) \quad (10)$$

对 $S'(\alpha)$ 进行微分求最值

$$\hat{\alpha} = (C^T C)^{-1} C^T \cdot S_{U_i \rightarrow U_n} \quad (11)$$

在真实的社交环境下，受现有的数据采集方式及用户自身交往范围的影响，社交关系数据往往达不到所需要的数据密度（指的是非零数据所占的比例），这也是现有的推荐系统所普遍面临的问题之一：数据稀疏性。据此，本文提出一种社交关系增强模型，从社会关系本身属性出发，提出了一种包含关系传递、共点增强、约束传递等性质的增强策略，优化通过公式(3)获取的社交关系度，以提高社交关系度的准确度与数据密度。具体如图3-图5所示(图中虚线表示进行修正的社交关系度)。具体定义如下：

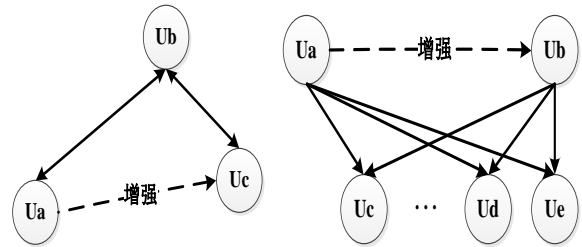


图3 关系传递示例

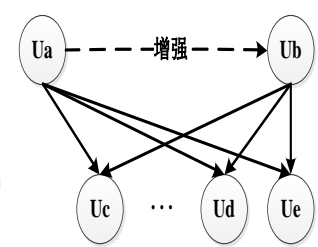


图4 共点增强示例

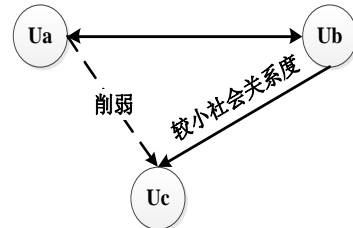


图5 约束传递示例

(1)关系传递,如图3,若 U_a 与 U_b 、 U_c 间皆有较大(大于与其他用户社交关系度均值)双向社交关系度值,则增大 $S_{U_a \rightarrow U_c}$.由于增强模型需要考虑 $S_{U_a \rightarrow U_c}$ 的初始值,其初始值越大可增强幅度越小;同时还增强与 $S_{U_a \rightarrow U_b}$ 、 $S_{U_b \rightarrow U_c}$ 的值为正相关关系,因此采用如下方式来表示:

$$S_{U_a \rightarrow U_c} = S_{U_a \rightarrow U_c} \cdot \left(1 + \sqrt{\frac{S_{U_a \rightarrow U_b} + S_{U_b \rightarrow U_c}}{S_{U_a \rightarrow U_c} + S_{U_b \rightarrow U_c} + S_{U_a \rightarrow U_c}}}\right) \quad (12)$$

(2)共点增强,如图4,若 U_a 与 U_b 间有较多的具有较大(大于社交关系值均值)社交关系值的共同用户,则增强 $S_{U_a \rightarrow U_b}$.由此可知,共点增强由以上述用户数决定,本文提出了一种考虑其他用户间存在较大社交关系值用户数的增强策略,具体如下:

$$S_{U_a \rightarrow U_b} = S_{U_a \rightarrow U_b} \cdot \left(1 + \frac{x - x^*}{x}\right) \quad (13)$$

其中, x 为 U_a 与 U_b 间具有较大社交关系度值的共同用户数, x^* 为其他用户间平均的共同用户数.

(3)约束传递,如图5,若 $S_{U_a \rightarrow U_b}$ 、 $S_{U_b \rightarrow U_a}$ 皆有较大社交关系度值, $S_{U_b \rightarrow U_c}$ 较小,则削弱 $S_{U_a \rightarrow U_c}$.约束传递是一种削弱策略,一方面其受 $S_{U_b \rightarrow U_a}$ 与 $S_{U_b \rightarrow U_c}$ 间的差异程度的影响,差异越大说明 U_a 与 U_b 间社交关系越强,而 U_a 与 U_c 间越弱,所以需要削弱的幅度也就越大;另一方面需要考虑 U_b 与 U_a 间社交关系度的差异程度,差异越大则单向社交关系越明显,其间影响程度越小.据此,约束传递的度量方式如下:

$$S_{U_a \rightarrow U_c} = S_{U_a \rightarrow U_c} \cdot \left(1 - \frac{(S_{U_b \rightarrow U_a} - S_{U_b \rightarrow U_c}) - |S_{U_b \rightarrow U_a} - S_{U_a \rightarrow U_b}|}{S_{U_b \rightarrow U_a}}\right) \quad (14)$$

那么基于所建立的社交关系网络模型,本文通过融合个体认知行为、用户间认知交互关系,建立起度量用户间认知交互关系的认知关系模型.

目前对于用户间的认知关系,存在两个结论^[21]

- (1)每个认知节点均受其他认知节点的影响;
- (2)认知影响应排除认知节点对自身的影响.

基于此,本文提出了一种从认知层面度量用户间关联关系的认知关系网络模型,该模型在形式上类似单层神经网络,以个体认知、社会关系度为输入,以用户间认知交互关系为输出,关联关系为“神经元激活函数”,如下

$$f(cl, S)_{U_a \rightarrow U_b} = \frac{e^{cl} - e^{-cl} + e^S - e^{-S}}{e^{cl} + e^{-cl} + e^S + e^{-S}} \quad (15)$$

其中, $f(cl, S)_{U_a \rightarrow U_b}$ 为认知关系度.认知关系网络模型图形上近似于立体双曲面正切函数,功能是模拟人的真实认知过程,其值表示在认知网络中认知节点间的初始值.

根据结论1,在认知关系网络中,用户认知节点间存在交互影响关系,本文基于认知心理学与社

交关系相关理论,提出了一种包含认知依赖、认知风险、群集认知等性质的认知关系网络模型.相关定义如下:

定义5. 认知依赖,指的是在认知层面个体用户对其他用户的依赖性,表现为个体用户的认知行为受与其关联用户的影响关系,是一种认知社会性与动态性的体现.其值由与当前用户最关联的 n 个用户中(即认知关系度最大的 n 个用户),认知关系度波动的用户数及波动量的共同影响.影响关系表示如下

$$R_a = 1 + \frac{\text{Card}(GU) - k}{n} \cdot \frac{\sum_{i \in GU} (f'(cl, S)_{U_a \rightarrow U_i} - f(cl, S)_{U_a \rightarrow U_i})}{\sum_{i \in GU} f(cl, S)_{U_a \rightarrow U_i}} \quad (16)$$

其中, R_a 为在认知关系网络模型中用户 U_a 的认知依赖系数, k 为认知依赖修正启动值, $k \in N^+$, $k \leq \text{Card}(GU)$,即是只有当与 U_a 相关的用户数达到一定程度时才会启动修正策略. GU 为与 U_a 关联的认知产生波动的用户集合, $\text{Card}(GU)$ 表示 GU 中用户数量, $f'(cl, S)_{U_a \rightarrow U_i}$ 表示受用户 U_a 与 U_i 间在各种认知影响因素作用后的认知关系度.

定义6. 认知风险,指的是认知关系度较高的用户间,由于认知关系的非对称性以及用户的社会性,导致的认知不确定现象,表现为决策心理上的不确定感^[14].例如,用户 U_a 与 U_b 间有较高的双向认知关系度(大于各自均值),而 U_a 与其他用户的认知关系度普遍较小, U_b 则反之,在这种情况下, U_b 对于 U_a 而言就存在某种程度上的认知风险.本文提出一种认知风险判别策略,包含两种判别条件,如下

$$\begin{cases} \frac{\sum_{i \in HU_b} f(cl, S)_{U_b \rightarrow U_i}}{\text{Card}(HU_b)} > \frac{\sum_{j \in HU_a} f(cl, S)_{U_a \rightarrow U_j}}{\text{Card}(HU_a)} \\ \frac{\text{Card}(HU_a)}{\text{Card}(HU_b)} < \rho \end{cases} \quad (17)$$

其中, HU_a 与 HU_b 分别表示与用户 U_a 、 U_b 认知关系度较高的用户集合, $\text{Card}(HU_a)$ 与 $\text{Card}(HU_b)$ 分别表示 HU_a 与 HU_b 中用户数量, ρ 为修正参数.

两个条件同时满足时, U_a 对于 U_b 即存在认知风险,需要减弱 U_a 对于 U_b 的认知关系度.本文基于两用户的社交关系差异提出了一种认知补偿策略,如下所示:

$$\begin{aligned}
f(cl, S)_{U_a \rightarrow U_b} &= f(cl, S)_{U_a \rightarrow U_b} \\
&\frac{\sum_{i \in HU_b} f(cl, S)_{U_b \rightarrow U_i}}{Card(HU_b)} \cdot \frac{\sum_{j \in HU_a} f(cl, S)_{U_a \rightarrow U_j}}{Card(HU_a)} \\
&\cdot \left(1 - \frac{Card(HU_b)}{Card(HU_b) - Card(HU_a)}\right) \\
&= f(cl, S)_{U_a \rightarrow U_b} \\
&\frac{Card(HU_a) \cdot \sum_{i \in HU_b} f(cl, S)_{U_b \rightarrow U_i} - Card(HU_b) \cdot \sum_{j \in HU_a} f(cl, S)_{U_a \rightarrow U_j}}{Card(HU_b)^2 \cdot Card(HU_a) - Card(HU_a)^2 \cdot Card(HU_b)} \\
&\cdot \left(1 - \frac{Card(HU_b)}{Card(HU_b) - Card(HU_a)}\right) \quad (18)
\end{aligned}$$

定义 7. 认知群集, 指的是在认知关系网络中, 由关联用户组成的用户群集间认知影响关系, 即是从社交群集的角度度量整体的认知状态, 是认知及用户行为社会性的体现. 若任意用户 U_b 为用户 U_a 的关联用户, 则需满足如下条件

$$\begin{cases} f(cl, S)_{U_a \rightarrow U_b} > 0, f(cl, S)_{U_a \rightarrow U_b} > 0 \\ f(cl, S)_{U_a \rightarrow U_b} \geq \overline{f(cl, S)_{U_a}} \end{cases} \quad (19)$$

其中, $\overline{f(cl, S)_{U_a}}$ 表示用户 U_a 与其他用户间的平均社交认知度. 判断条件也即是关联用户间应具有双向非 0 值社交关系度, 且具有较大的社交关系度值. 那么满足上述条件的用户群集就构成了用户 U_a 的认知群集. 本文提出了一种采用群集认知度来表示认知集群的认知状态的度量方法, 如下

$$\begin{aligned}
Q_{U_a} &= \frac{|Card(HU_a) - Card(HU_{oa})|}{Card(HU)} \\
&\times \left(\frac{\sum_{j \in HU_a} f(cl, S)_{U_a \rightarrow U_j}}{Card(HU_a)} - \frac{\sum_{k \in HU_a} \sum_{m \in HU_k} f(cl, S)_{U_k \rightarrow U_m}}{\sum_{k \in HU_a} Card(HU_k)} \right) \quad (20)
\end{aligned}$$

式中, $\overline{Card(HU_{oa})}$ 表示 U_a 的关联用户所包含的平均关联用户数, HU_a 表示用户 U_a 的认知群体. Q_{U_a} 是在认知关系网络的基础上, 基于认知关系度描述用户间群体性的认知状态, 即是认知社会性的特征数据.

3.2 项目认知价值

定义 9. 项目认知价值, 指的是项目的价值认知程度, 是用户能否从认知层面接受该项目的影响要素. 以商品为例, 用户对该商品的认知价值与其实际价值的差距直接影响该项目的价格需求性, 用户会更倾向于接受认知价值与实际价值大致相符的商品. 影响项目认知价值的主要因素包括: 项目自身属性、用户的认知水平、该项目的偏好用户(由于用户与项目间在社交网络下的关联关系仍是采用评分值度量的偏好关系, 因此其计算方法仍是基于用户的历史偏好数据)与当前用户的社交关系度等, 计算方法如下

$$IR_{I \rightarrow U_a} = \eta \cdot \frac{Pc_i}{Nc_i + Pc_i} + cl_a \cdot \frac{\sum_{i \in IU} (f(cl, S)_{U_i \rightarrow U_a} - \overline{f(cl, S)_{U_a}})}{Card(IU)}$$

其中, IU 表示偏好当前项目的用户集合, η 为修正参数, Pc_i 与 Nc_i 为项目属性中的正、负影响因素, 分别表示项目属性中, 增强/削弱用户对该项目认可程度的属性因素, 获取方法有项目属性与用户喜好项目中属性差异、专家经验等.

3.3 基于FLDA的认知偏好优化

由于推荐结果是基于用户的偏好生成的, 由于认知偏好直接影响了用户的偏好行为, 因此本文方法主要通过获取用户的认知偏好来生成推荐.

个体认知是从用户的视图下度量其自身的认知能力; 社交认知度、群集认知描述的是用户间认知影响关系视图下获取的认知社会性特征数据; 项目认知价值是从项目的视图下获取的认知吸引程度. 仅从数据特征层面难以为上述视图数据建立统一的分类或判别准则, 难以获取用户的认知偏好^[22]. 据此本文提出了一种基于 FLDA 的认知偏好优化方法, 其中优化指的是采用 FLDA 对认知偏好的获取方法进行优化, 并且采用认知偏好生成推荐, 是对传统推荐算法的一种优化. 算法流程如图 6 所示.

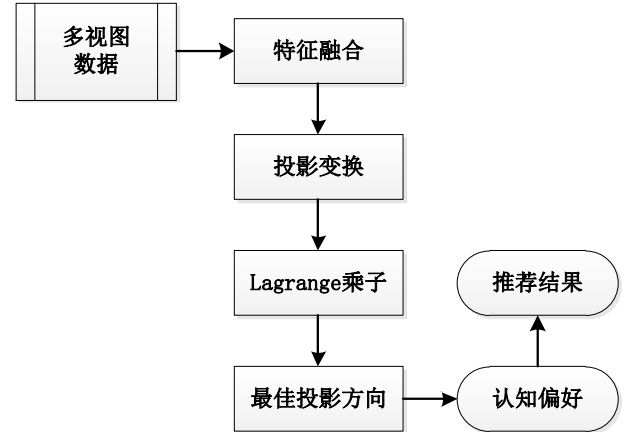


图 6 基于 FLDA 的认知偏好优化流程

基于上述数据, 建立多视图样本数据, 包含用户属性矩阵 $U = \{x_1^{(1)}, x_2^{(1)}, \dots, x_{m_1}^{(1)}\}$ 、项目属性矩阵 $I = \{x_1^{(2)}, x_2^{(2)}, \dots, x_{m_2}^{(2)}\}$, 项目认知价值 $IR = \{x_1^{(3)}, x_2^{(3)}, \dots, x_{m_3}^{(3)}\}$ 、个体认知行为 $UR = \{x_1^{(4)}, x_2^{(4)}, \dots, x_{m_4}^{(4)}\}$ 、群集认知矩阵 $UQ = \{x_1^{(5)}, x_2^{(5)}, \dots, x_{m_5}^{(5)}\}$, m_1 到 m_5 表示各样本中样本数量.

首先依据样本均值的定义获取各样本的均值, 采用 A_j 来表示上述各样本均值, 具体的

$$A_j = \frac{1}{m_j} \sum_{i \in X_{(j)}} x_i^{(j)} \quad j \in c \quad (22)$$

其中, 用户与项目属性的样本矩阵示例为: 用户的平均年龄、项目的平均价格等, c 表示总的样本类别数. 各样本的类内离散度矩阵可表示为

$$S_{k_j} = \sum_{i=1}^{m_j} (x_i^{(j)} - A_j)(x_i^{(j)} - A_j)^T, j \in c \quad (23)$$

总的类内离散度矩阵为

$$S_k = \sum_{k=1}^c \sum_{i=1}^{m_j} (x_i^{(j)} - A_j)(x_i^{(j)} - A_j)^T \quad (24)$$

类间离散度矩阵为

$$S_t = \sum_{k=1}^c m_k (A_1 - A_2)(A_1 - A_2)^T \quad (25)$$

其中矩阵 $(A_1 - A_2)(A_1 - A_2)^T$ 在数学表达上为协方差矩阵, 描述了每条偏好数据与总体数据间的关联关系. 其中, 此矩阵对角线上的元素表示了相应的偏好样本与总体间的样本方差, 非对角线上的元素表示了其总体矩阵的协方差. 那么 A_j 即描述了样本总体内各类间的离散冗余程度, S_t 即表示了各类间(符合用户认知偏好的项目集合)的离散冗余程度.

在获取到两类样本的类间离散度与类内离散度后, 为了提高推荐结果的准确度, 即分类规则的精确度, 需要尽可能的降低分类后的各用户认知偏好项目间的耦合程度, 增加其集合内部的聚合程度. 换言之, 需要获取分类后属于同一类的项目, 其间差异最小; 属于不同类的项目, 其间差异最大. 即是找到满足 A_j 尽可能小, 而 S_t 尽可能大的分类准则, 才能使得分类后的样本是满足特定用户认知需求的项目集合. 为此, 将原样本数据看作相应维度的向量, 并以任一向量 C 为方向进行投影变换, 转换规则为

$$r_i^{(j)} = C^T x_i^{(j)} \quad (26)$$

变换后的两类类样本均值为

$$\begin{aligned} A_j' &= \frac{1}{m_j} \sum_{i \in X_{(j)}, j=1,2} r_i^{(j)} = \frac{1}{m_j} \sum_{i \in X_{(j)}, j \in c} C^T x_i^{(j)} \\ &= C^T \left(\frac{1}{m_j} \sum_{i \in X_{(j)}, j \in c} x_i^{(j)} \right) \\ &= C^T A_j \end{aligned} \quad (27)$$

变换后的类内离散度为

$$\begin{aligned} S_{k_j}' &= \sum_{i=1}^{m_j} (r_i^{(j)} - A_j')(r_i^{(j)} - A_j')^T = \sum_{i=1}^{m_j} (C^T x_i^{(j)} - C^T A_j)^2 \\ &= C^T \left[\sum_{i=1}^{m_j} (x_i^{(j)} - A_j)(x_i^{(j)} - A_j)^T \right] C \\ &= C^T S_{k_j} C, j \in c \end{aligned} \quad (28)$$

变换后的类间离散度为

$$\begin{aligned} S_t' &= (C^T A_1 - C^T A_2)(C^T A_1 - C^T A_2)^T \\ &= C^T (A_1 - A_2)(A_1 - A_2)^T C \\ &= C^T S_t C \end{aligned} \quad (29)$$

对于变换后的认知偏好项目集合的数据要求仍然是, 需要尽可能的降低分类后的各集合间的耦合程度, 增加集合内项目间的聚合程度. 为此引入了 Fisher 判别准则, 如下

$$J_{Fisher} = \frac{(A_1' - A_2')^2}{S_{k_1}' + S_{k_2}'} \quad (30)$$

其核心思想为计算能够使 J_{Fisher} 取得最大值的投影方向, 将前面所述投影变换代入可以得出

$$J_{Fisher}(C) = \frac{C^T S_t C}{C^T S_k C} \quad (31)$$

我们采用 Lagrange 乘子法求解其最值, 令分母为非零常数 b , 那么 Lagrange 函数定义如下

$$L(C, \delta) = C^T S_t C - \delta(C^T S_k C - b) \quad (32)$$

对 C 求偏导数可以得出

$$\frac{\partial L(C, \delta)}{\partial C} = 2S_t C - 2\delta S_k C \quad (33)$$

令偏导数为 0, 即是

$$S_t C' = \delta S_k C' \quad (34)$$

两边同乘 S_k^{-1} , 即可变形为

$$S_k^{-1} S_t C' = \delta C' \quad (35)$$

原问题也就转换为了求解矩阵 $S_k^{-1} S_t$ 的特征值问题, 其最大特征值所对应的特征向量也即是所求的最佳投影方向, 由此也可推出具有最佳认知偏好的投影方向.

获取认知偏好是生成推荐项目的前提和核心步骤, 那么基于认知偏好为用户生成推荐结果的过程如下: 依据公式(35)获取最大特征值对应的特征向量, 即最佳认知偏好的投影方向. 在此方向上能够获得各样本间最大的类间离散度、最小的类内离散度, 即分类的效果最优. 即能够将符合特定用户认知偏好的项目依据用户的偏好特征进行分类, 并将分类结果推送给用户.

4 实验步骤及结果分析

本节通过实验来验证如下问题: (1)所建立认知关系网络模型的有效性; (2)项目属性与用户认知关系度对项目认知价值的影响程度; (3)与现有方法相比, 所提出方法能能否挖掘出用户隐藏的偏好特征, 以及能否取得更好的推荐质量.

4.1 实验数据集

目前尚无公开的同时包含位置信息、通话信息、用户历史偏好行为记录等要素的数据集, 本文采用麻省理工大学 Reality Mining^[20]提供个标准数

据集与 *BookCrossing* 数据集^[23]，并以用户属性最近邻为关联准则，形成同时包含上述要素的模拟数据集 *RBookCrossing*。其中，*Reality Mining* 所提供的标准数据集中包含麻省理工大学志愿者们的移动轨迹、相遇数据和通话记录等，时间跨度为九个月。*BookCrossing* 数据集由 Cai-Nicolas Ziegler 从 *Book-Crossing* 图书社区获取的 278858 个用户对 271379 本图书的行为信息，并包含用户与图书的属性信息。

针对所采用的数据集，经过反复检验测试，正、负相关因素的判定规则^[14]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。

4.2 实验结果度量标准

本文使用 $P@R$ 度量用户偏好获取的准确性， $P@R$ 依据所获取的待预测偏好用户的偏好值对项目进行排序，并向其推荐最符合偏好的 $Top-R$ 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。

$$P@R = \frac{Top-R \text{ 推荐项目集所包含测试集 } Top-R \text{ 种项目的数量}}{R} \quad (36)$$

评价推荐系统的另一个重要指标是多样性，它指的是推荐结果中项目间的不相似性。其值越大，表明推荐结果覆盖大多数用户兴趣点的概率也就越大。假设 $s(i, j) \in [0, 1]$ 定义了项目 I_i 和 I_j 之间的相似度， $|G|$ 表示推荐列表的长度，则推荐列表 G 的多样性定义如下：

$$Diversity = 1 - \frac{\sum_{I_i, I_j \in G} s(i, j)}{\frac{1}{2}|G|(|G|-1)} \quad (37)$$

$$S_{item}(i, j) = \frac{\sum_{a \in U_{ij}} (P_{ai} - \bar{P}_a)(P_{aj} - \bar{P}_a)}{\sqrt{\sum_{a \in U_{ij}} (P_{ai} - \bar{P}_a)^2 \cdot \sum_{a \in U_{ij}} (P_{aj} - \bar{P}_a)^2}} \quad (38)$$

其中， P_{ai} 表示用户 U_a 对项目 i 的历史偏好， \bar{P}_a 表示用户的 U_a 的历史平均偏好。

4.3 实验设计及结论

实验一：认知关系网络模型有效性检验

本次实验目的在于检验所提出的认知关系网络模型的有效性，认知关系网络模型主要是用于高效、准确的度量用户认知节点间的关联关系，为进一步的应用研究提供数据支撑与决策依据。通用的检测指标为：整网密度 $NOD(NOD \in [0, 1])$ 、度数中

心势 $DCP(DCP \in [0, 1])$ ^[20]，具体定义如下：

NOD 指的是认知关系网络模型节点间的紧密程度， NOD 越大表示认知关系网络模型越发达。若认知关系网络中节点总量为 M ，认知关系总量为 CM ，则

$$NOD = \frac{CM}{M(M-1)} \quad (39)$$

在所提出的认知关系网络模型中，由于为有向有权图，认知关系度只有大于 0 才认为其对应节点间存在认知关系。

DCP 指的是认知关系网络中认知节点间差异程度， DCP 越大表示网络分布越不均匀，网络健壮性越差。度量方法如下

$$DCP = \frac{\sum_{i=1}^M D_{RDMax} - D_{RD}(i)}{M-2} \quad (40)$$

$$D_{RD}(i) = \frac{d_{out}(i) + d_{in}(i)}{2(M-1)} \quad (41)$$

其中 i 表示不同的认知节点， $D_{RD}(i)$ 认知节点的相对度数中间度， D_{RDMax} 则为所有相对度数中间度的最大值， $d_{in}(i)$ 、 $d_{out}(i)$ 分别指代认知节点 i 的初度与入度。

本文选取具有代表性的两种社会关系认知模型 $HGSM$ ^[24] 与 AM ^[25]。其中， $HGSM$ 通过挖掘用户历史位置记录建立用户间的社交关系网络； AM 通过建立基于用户通话的关联策略，度量用户间社交关系。对比实验结果如图 7 与图 8 所示。

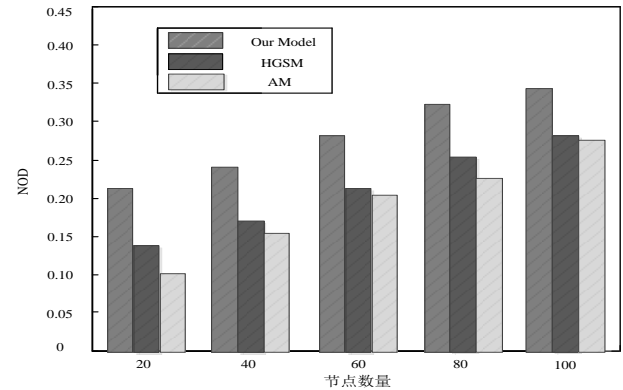


图 7 NOD 对比实验结果

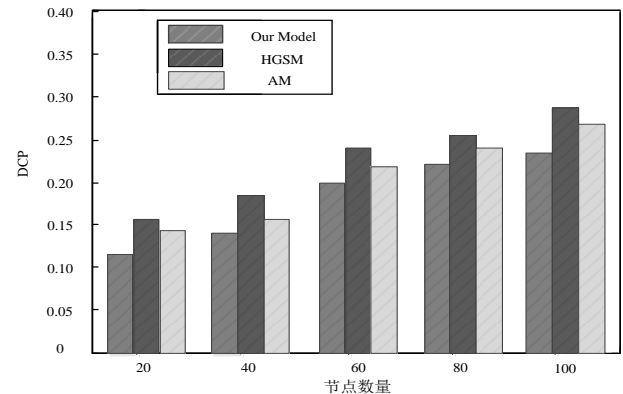


图 8 DCP 对比实验结果

从实验结果图 7 可以看出, 随着认知节点数量的增加, 三种算法的 *NOD* 都随之增加, 表示网络模型中社会关系随着认知节点数的增加而增多, 而本文提出的认知关系网络模型的增长幅度明显大于两个对比模型, 原因在于所提出的认知关系网络模型是基于个体认知行为与社交关系网络进一步融合, 能够更好的挖掘用户间社会性关联关系, 也使得认知模型能够度量用户认知节点间认知关系的变化趋势. 并且, 由于本文定义了包含认知依赖、认知风险等认知关系增强策略, 将认知关系间隐藏的传递性与约束性反馈于认知关联间, 有助于建立起更广泛的认知关系, 这一结论可以从实验结果图 8 上得到验证, 所提出的认知关系模型 *DCP* 增长幅度缓慢, 优于两种对比模型.

实验二: 参数 η 影响实验

在 *RBookCrossing* 数据集上, 为保证实验结果的客观性, 经过反复实验对比我们分别选取 20%、30%、50%、60%、70% 五种比例数据作为训练集, 其余作为测试集进行实验. 在四种数据模式下, 作者对三个参数都进行了实验检验, 由于检验思路和方法基本一致, 在本节以项目认知价值修正参数 η 作为代表性实验进行展示. 其中, 经过反复测试, 选取认知依赖参数 $k=12$, 认知风险系数 $\rho=0.4$. η 主要作用是控制项目属性的正、负影响因素与相关用户认知关系度在项目认知价值中的影响程度. 本次实验中, 我们基于项目认知价值获取用户对当前项目的喜好程度, 并与数据集中真实的记录进行对比, 用预测结果与实际值间的偏差作为度量参数取值的评测标准. 为增强实验结果的客观性, 我们对于 *RBookCrossing* 数据集采用 *ABO* 算法进行隐藏处理, 即是每次运算测试集数据的隐藏都随机进行, 每组数据进行运算时测试数据集都在变化, 在这种情况下, 运行实验的结果会有一定的浮动空间. 因此我们在每个节点将算法运行 5 次, 取其均值作为当前的预测精度值. 经过反复实验对比我们选取出 4 组有代表性实验结果如图 9-图 12 所示.

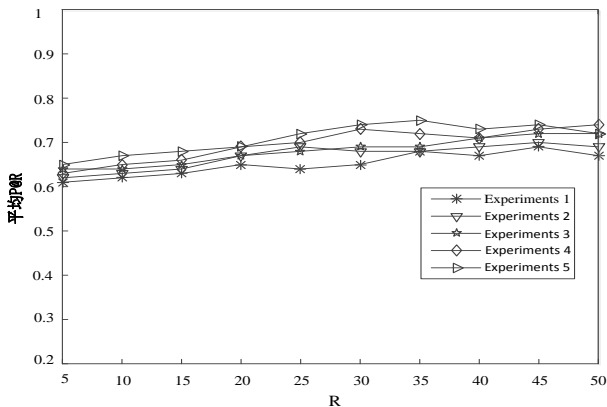


图 9 $\eta=0.28$ 时实验结果

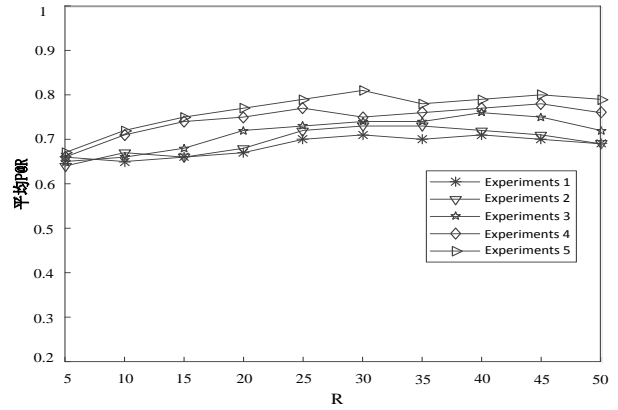


图 10 $\eta=0.43$ 时实验结果

从实验结果可以得出, 在不同的项目认知参数 η 取值下, 随着训练集比例的增加, 预测的准确度也随之增加, 在 70% 的 *RBookCrossing* 数据作为训练集时能够取得最优的推荐准确度. 分别对比四个结果图中, 相同数据集比例下 η 对于推荐准确度的影响可以看出, 随着 η 取值的增加, *P@R* 整体上呈现递增趋势, 并随着 η 的持续增加又呈现递减趋势, 并在 $\eta=0.43$ 时取得最优值, 这说明基于用户属性的正、负影响因素对项目认知价值的影响, 要小于与当前项目相关用户的认知关系. 也即说明了, 在群集认知网络模型中, 由外界(主要指用户认知喜好

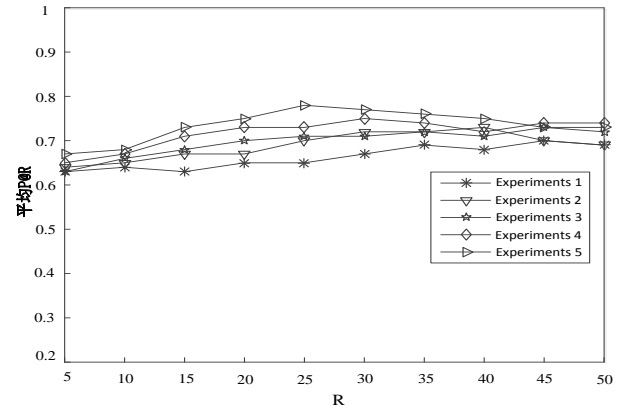


图 11 $\eta=0.56$ 时实验结果

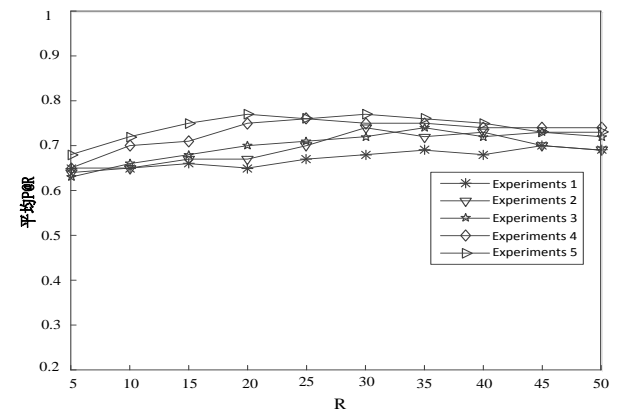


图 12 $\eta=0.74$ 时实验结果

行为)的交互作用所导致的对项目认知层面价值的影响, 要大于项目自身属性对其的影响, 是用户的

社会性在项目认知层面的体现。

实验三：与现有方法对比实验

经过反复试验对比，我们选取 40% 与 70% 两种训练集比例进行实验，将本文提出的算法，与经典的协同过滤算法 CACF^[26]、基于上下文情景感知的推荐算法 CABT^[27] 以及融入社会关系的矩阵分解算法 WSDM^[28] 进行实验对比。其中，CACF 算法采用余弦相似与多维距离，并结合独立准则函数对协同过滤算法进行优化。由于 CACF 算法仅考虑“用户-项目”间的相互影响，因此本次试验中其仅采用相同数据集中的用户对图书的评分记录。CABT 最新的融合上下文信息的情景感知推荐方法，对用户的购买行为(本次实验中以对图书的评分值代替)及相关上下文环境进行分析，抽取其中行为信任关系并生成基于内容的推荐。WSDM 把社会化网络信息与矩阵分解进行结合，提高了推荐系统的准确性和可扩展性。

(1) 准确度对比

准确度是度量推荐系统的最常用度量方法，是推荐质量的最直观体现。本次实验依然采用 ABO 算法对数据集进行随机隐藏，并在每个节点运行 5 次，以均值作为当前值，选取有代表性的 40% 与 70% 两种训练集比例下，四种算法的对比结果如图 13-图 14 所示。

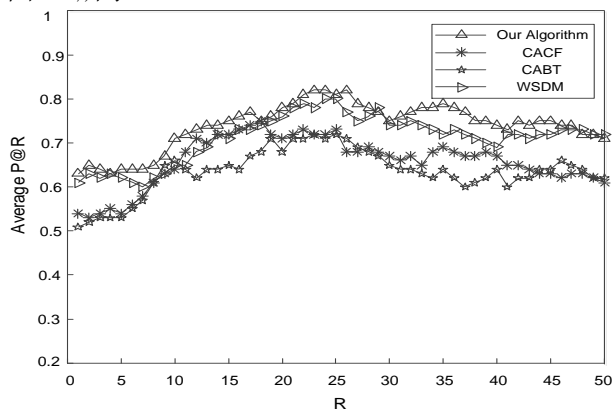


图 13 40%训练集对比实验结果

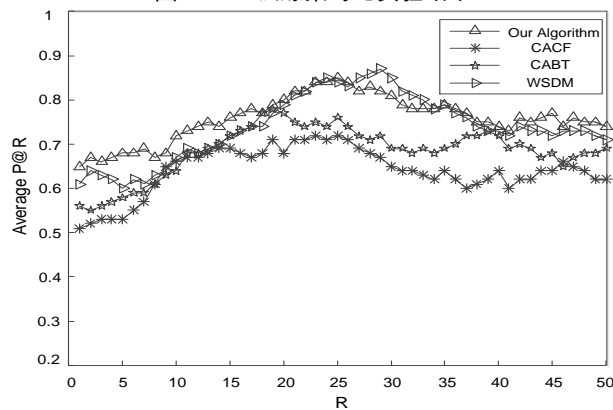


图 14 70%训练集对比实验结果

从实验结果可以看出，以 P@R 为评价标准在不同的数据集模式下，随着训练集比例的增加，四种算法的准确度都呈现出增加的趋势，说明提高训练集比例有助于取得更好的推荐效果。本文算法、CABT、CACF、WSDM 的最优 P@R 值分别为 0.83、0.74、0.71、0.86，其中 WSDM 的最优值要优于本文算法。而整体上本文算法相比于 CABT、CACF、WSDM，平均 P@R 分别提升了 21.46%、18.92%、6.89%，也即是验证了本文算法的有效性，并且优于现有方法，具体分析如下：

本文算法、CABT 以及 WSDM 要优于传统的协同过滤算法，这说明在获取用户偏好时，考虑社会化关系等上下文信息相比于只考虑用户历史评分记录的推荐算法，能够取得更加准确的推荐结果；并且本文算法在一定程度上要优于其他三种算法，这说明基于群集认知行为的偏好获取方法，由于更多的考虑了用户的社会性特征与认知心理特征，能够取得更好的推荐效果。

(2) 多样性对比

多样性表示推荐列表中项目间差异程度，其差异性越大，覆盖用户潜在兴趣点的概率也就越大。与准确度的区别体现在准确度只是从与历史偏好的匹配程度的角度去度量推荐效果，而多样性则是从推荐结果本身来度量，相比准确度具有更好的客观性。依据多样性评价标准，经过反复对比，我们选取出 40%、70% 四种训练集比例，四种算法的多样性对比如表 2 所示。

从表 2 可以看出，在不同数据集比例下，本文算法相比 CACF、CABT、WSDM 多样性分别提高 60.28%、37.46%、25.36% (40% 作为训练集)，56.31%、31.69%、24.49% (70% 作为训练集)。由此可见，从用户与项目属性、用户间认知交互关系等视图下，获取用户的偏好，相比于现有方法能够取得更好的推荐多样性。也即是，本文算法能够有效的避免由于仅采用历史偏好数据，所导致的用户偏好模型无法覆盖用户潜在的兴趣点。

从上述实验可以看出，多样性与准确度存在一定程度的不一致现象，若仅考虑准确度会导致所建立的偏好模型，无法全面的覆盖用户潜在的兴趣特征，即是具有较高准确度的推荐方法不一定能够满足用户的实际需求。实验结果表明本文算法的有效性，并且能够在取得更好推荐准确的与多样性，同时也证明了从认知层面获取用户偏好优于从行为层面。

表2 多样性对比实验结果

	Our Method		CACF		CABT		WSDM	
	40%	70%	40%	70%	40%	70%	40%	70%
5	0.62	0.66	0.38	0.37	0.48	0.48	0.44	0.45
10	0.63	0.68	0.39	0.39	0.48	0.50	0.46	0.47
15	0.69	0.68	0.42	0.42	0.49	0.51	0.49	0.50
20	0.67	0.69	0.43	0.44	0.49	0.52	0.47	0.48
25	0.68	0.71	0.42	0.47	0.50	0.52	0.49	0.51
30	0.69	0.72	0.43	0.46	0.51	0.54	0.52	0.53
35	0.71	0.71	0.44	0.47	0.51	0.54	0.54	0.56
40	0.72	0.69	0.44	0.48	0.53	0.55	0.58	0.58
45	0.73	0.69	0.46	0.47	0.51	0.56	0.57	0.58
50	0.72	0.71	0.47	0.47	0.49	0.55	0.56	0.58

5 结束语

现有的基于用户历史偏好记录的推荐方法, 由于数据稀疏性等问题, 导致所建立的偏好模型无法覆盖用户潜在的偏好特征, 影响了推荐的准确度与多样性. 本文将认知心理学与社会关系模型引入偏好获取领域, 提出了一种基于多视图群集认知优化的偏好获取方法, 从社会化认知的角度获取用户认知行为间的关联关系, 基于社交关系网络与个体认知行为, 建立了包含认知依赖、认知风险、群集认知等因素的认知关系网络模型, 并从用户属性、认知关系、项目认知价值等多视图出发, 通过特征融合、投影变换并引入 Fisher 判别准则, 建立起基于多视图优化的偏好获取模型. 通过与现有方法对比实验表明, 推荐准确度平均提升了 20.19%, 多样性提高近 44.89%, 验证了所提出方法的能够显著地提高推荐准确度与推荐质量, 并且有更大的概率覆盖用户的潜在偏好特征. 未来的研究方向, 将会将会致力于解决多视图学习在推荐系统领域的应用中所存在的小样本及秩限制问题, 以期进一步的提高推荐质量.

参考文献

- [1] Bagci H, Karagoz P. Context-aware location recommendation by using a random walk-based approach. *Knowledge and Information Systems*, 2016, 44(2):241-260.
- [2] Gao Ming, Jin Che-Qing, Qian Wei-Ning, et al. Real-time and personalized recommendation on microblogging systems. *Chinese Journal of Computers*, 2014, 37(4):963-975 (in Chinese)
(高明, 金澈清, 钱卫宁, 等. 面向微博系统的实时个性化推荐. *计算机学报*, 2014, 37(4):963-975)
- [3] Baltrunas L, Ricci F. Experimental evaluation of context-dependent collaborative filtering using item splitting[J]. *User Modeling and User-Adapted Interaction*, 2014, 24(1-2):7-34
- [4] Hui WU, BIAN Yi-Jie, Zhao Z, et al. Collaborative filtering algorithm based on trust. *Computer Systems & Applications*, 2014, 7(6):131-135
- [5] Pessemier T D, Dooms S, Martens L. Context-aware recommendations through context and activity recognition in a mobile environment. *Multimedia Tools & Applications*, 2014, 72(3):2925-2948
- [6] Alhamid M F, Rawashdeh M, Dong H, et al. Exploring latent preferences for context-aware personalized recommendation systems. *IEEE Transactions on Human-Machine Systems*, 2016, 46(4):1-9
- [7] Huang Z, Shijia E, Zhang J, et al. Pairwise learning to recommend with both users' and items' contextual information. *IET Communications*, 2016, 10(16): 2084-2090
- [8] Hong L, Zou L, Zeng C, et al. Context-aware recommendation using role-based trust network. *Acm Transactions on Knowledge Discovery from Data*, 2015, 10(2):1-25
- [9] YaroWsky D. Unsupervised word sense disambiguation rivaling supervised methods//*Proceedings of the 3th annual meeting on Association for Computational Linguistics*. Uppsala, Sweden, 1995: 189-196
- [10] Nguyen C, Wang X, Liu J, et al. Labeling complicated objects: multi-view multi-instance multi-label learning. *National Conference on Artificial Intelligence*. 2014,(3):2013-2019
- [11] Yang M, Sun S. Multi-view uncorrelated linear discriminant analysis with applications to handwritten digit recognition// *Proceedings of the International Joint Conference on Neural Networks*. Beijing, China, 2014:4175-4181
- [12] Zhu Z, Luo P, Wang X, et al. Multi-View Perceptron: a deep model for learning face identity and view representations. *Neural Information Processing Systems*, 2014, 13(4):217-225
- [13] Chechev M, Georgiev P. A multi-view content-based user recommendation scheme for following users in twitter. *social informatics*, 2012, 12(3): 434-447.
- [14] Lu H, Chen C, Kong M, et al. Social recommendation via multi-view user preference learning. *Neurocomputing*, 2016, 216(1): 61-71.
- [15] Gao Quan-Li, Gao Ling, Yang Jian-Feng, et al. A preference elicitation method based on users' cognitive behavior for context-aware recommender system. *Chinese Journal of Computers*, 2015, 38(9): 1767-1776 (in Chinese)
(高全力, 高岭, 杨建锋, 等. 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. *计算机学报*, 2015, 38(9):1767-1776)
- [16] 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)

- [17] Wonik P, Sanggil K, Youngkuk K. A personalized multimedia contents recommendation using a psychological model. *Computer Science and Information Systems*,2012,9(1):1-21
- [18] Zhang Hui-Qi, Dantu R. Predicting social ties in mobile phone networks//*Proceedings of the IEEE International Conference on Intelligence and Security Information(ISI)*.Vancouver, Canada, 2010: 25-30
- [19] Xie Xiao-Qin, Han Shuai, Chen Min,et al. Research on team formation method in social network, *Chinese Journal of Computers*, 2016,39(87):1-16(in Chinese)
(谢晓芹, 韩帅, 陈敏, 等. 基于社会网络的团队生成方法研究, *计算机学报*, 2017,40(3):712-728)
- [20] An Jian, Gui Xiao-Lin, Zhang Wen-Dong, et al.Social relation cognitive model of mobile nodes in the internet of things,*Chinese Journal of Computers*,2012,35(6):1164-1174(in Chinese)
(安健, 桂小林, 张文东, 等. 物联网移动感知中的社会关系认知模型, *计算机学报*,2012,35(6):1164-1174)
- [21] Chen Zhuang, Montazemi Ali R.The methodology of mining cognitive maps based on data resources,*Chines Journal of Computers*,2007,30(8):1446-1454(in Chinese)
(陈庄, 阿里蒙特瑟密. 基于数据资源的认知图挖掘方法, *计算机学报*, 2007,30(8):1446-1454)
- [22] Fischhoff B. Cognitive processes in stated preference methods. *Handbook of Environmental Economics*, 2006,2(5): 937-968
- [23] Zielger 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
- [24] Li Quan-Nan, Zheng Yu, Xie Xing, et al. Mining user similarity based on location history//*Proceedings of the 16th ACM International Conference on Advances in Geographic Information Systems*. New York, USA,2008:34-44
- [25] Zhang Hui-Qi, Dantu R. Predicting social ties in mobile phone networks//*Proceedings of the IEEE International Conference on Intelligence and Security Information*. Vancouver, Canada, 2010:25-30
- [26] Wiranto, Edi Winarko, Sri H, et al. Improving the prediction accuracy of multicriteria collaborative filtering by combination algorithms. *Proceedings of the International Journal of Advanced Computer Science and Application*, 2014,5(4): 51- 58
- [27] Michele G, Umberto P, Alexander T. The effect of context-aware recommendations on customer purchasing behavior and trust// *Proceedings of the Fifth ACM Conference on Recommender Systems*. Foster City,USA,2013:85-92
H Ma, D Zhou, C Liu, et al Recommender systems with social regularization// *Proceedings of the Forth International conference on Web Search & Web data mining*. Banff, Canada, 2011 :287-296



GAO Quan-Li, born in 1988, Ph. D, lecturer. His research interests focus on recommender systems.

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

CAO Rui, born in 1991, Master candidate. His research interests are in data mining and recommender system.

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

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 preference, 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 problem.

In this paper, we study the problems which belongs to the field of recommender systems, and it also related to intelligent computing. We proposed a multi-view cluster cognition optimization model based context-aware method is proposed. This method measure the incidence relation among users' cognitive behavior from the view of social cognition based on social networks and individual cognitive behavior, it build the cognitive relation model which includes reliant cognition, cognition risk, cluster cognition. Then the method use feature fusion, projection transformation and Fisher discriminant criterion to build multi-view optimization preference elicit model from the views of users' properties, cognitive relation, items' cognitive value The experimental results shows 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.61373146, No.61572401; Shanxi Industrial Tacking Foundation under Grant No.2014k05-42; Shanxi Provincial Natural Science Foundation of China under Grant No.2012JQ8047; Foundation of Shanxi Educational Committee under Grant No.11JK1059; Foundation of 2013 Science Research of Shanxi Educational Committee under Grant No.2013JK1178.