文章编号:1006-2467(2016)09-1422-08 DOI: 10.16183/j.cnki.jsjtu.2016.09.013

考虑时序性和动态信任的工程经验知识推荐技术

黄 颖, 蒋祖华, 刘璞凌, 王海丽
(上海交通大学 机械与动力工程学院,上海 200240)

摘 要: 随着用户和词条数量的增长,"用户-词条"评分矩阵变得极其稀疏,导致基于相似度计算的推荐算法精度降低. 提出一种基于用户短期兴趣时序性变化的评分矩阵进行预填充. 由于 MediaWiki社区用户群之间的信任呈动态变化,定义了追随率反映同一时间窗内具有相似兴趣的用户对知识推荐的参考性. 最后设计实验,确定时间窗长度 T 的最优参数,通过比较 CFBDT(Collaborative filtering based on dynarnic trust)算法与 3 类现有算法的效果,验证其可行性.

关键词: 知识推荐; 工程经验知识; 动态信任; MediaWiki 平台
中图分类号: TP 391 **文献标志码**: A

Engineering Empirical Knowledge Recommendation Mechanism Considering Time Sequence and Dynamic Trust

HUANG Ying, *JIANG Zuhua*, *LIU Puling*, *WANG Haili*
(School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: Users' rating matrix of User-Item-Time model suffers from sparsity severely with the increase of users and items. A pre-filling algorithm to pre-process the rating matrix sequentially was developed to avoid the rapid accuracy loss of recommendation system. Moreover, loyalty L was defined to measure the probability of which a target user would be convinced by an experienced user, as trust between users varied dynamically according to their historical interactions. Finally, experiments were conducted to determine the proper length of time window and verify the effectiveness of the proposed algorithm compared with the three previous collaborative filtering recommendation algorithms.

Key words: knowledge recommendation; engineering empirical knowledge; dynamic trust; MediaWiki platform

工程领域设计作为一个复杂的知识密集型过程,蕴含着大量复杂的不同类型的知识,如公式类知识、规则和模糊规则类知识、隶属函数类知识,以及设计经验类知识等,其中在企业应用中反映最实用和有价值的是经验知识. 工程师需要先了解相关的设计知识和历史经验,才能快速地进行工艺设计或生产设计;同时,成熟的历史经验知识也是进行创新设计的依据和参考[1]. 然而,在 Web2.0 环境下,互联网成为全球最大的知识库,也引发了知识泛滥、知识迷航等问题. Wiki 技术作为典型的 Web 2.0 技术

收稿日期:2015-11-06
基金项目:国家自然科学基金项目(70971085,71271133),上海市教育委员会科研创新重点项目 (13ZZ012),上海市科委科技创新行动计划 (13111104500),上海汽车工业教育基金会项目资助
作者简介:黄 颖(1991-),女,湖南省郴州市人,硕士生,主要研究方向为知识管理.
蒋祖华(联系人),男,教授,博士生导师,电话(Tel.):021-34206819;E-mail: zhjiang@sjtu.edu.cn.

之一，支持具有相同兴趣，或相近专业和领域的用户群体，以开放协作的方式共同创造和积累知识. 因此，如何在海量信息中有效、准确地给用户推荐工程领域经验知识，是在较短时间内获得高质量设计效果的关键因素.

本文引入用户相似度概念，重新定义社交网络中相似度属性、相似度构成及其计算方法，设计一种改进的协同过滤推荐算法，提出基于社会网络动态信任的考虑时序性的工程经验知识协同过滤推荐技术，并给出推荐质量度量方法.

1　国内外研究现状

知识推送旨在依据用户的知识需求主动将合适的知识提供给用户. 当更多的用户加入 MediaWiki 平台时，用户间的交互行为和职能关系将发挥出更加重要的作用，因此社会网络被引入虚拟社区的知识推荐中.

YANG 等[2]考虑社交网络对用户兴趣的影响，采用分析了基于社交网络的矩阵重构方式和最近邻搜索这 2 种方式的协同过滤推荐效率，通过 K 均值聚类的方式，首先计算边的权重，对商品进行推荐. HOHFELD 等[3]通过交换评分向量，并更新用户邻居进行推荐. ESSLIMANI 等[4]针对购物网站这个数据量庞大的系统，分析用户浏览行为的相似性进行推荐. PIZZATO 等[5]研究了互为推荐方的一类推荐系统，即用户既是推荐主体又是推荐的客体，例如招聘网站和在线交友网站等，通过分析澳大利亚交友网站的大量数据集，考虑用户的交互图谱，对用户的个人信息进行分类取值. 文献[5-9]中考虑用户兴趣的时序性，较之用户-词条二维模型，能更好地察觉到用户兴趣随时间推移而产生的变化.

现有的考虑社会网络的基于内存的协同过滤推荐算法研究大多是建立二维矩阵模型，即将存储在网站中的每个用户信息和词条建立一个二维矩阵，对用户向量组或者词条向量组进行分类，使得类中的成员具有最大的相似度，最后通过同类成员信息向目标用户推荐词条. 这种基于二维矩阵模型的协同过滤推荐算法将一个用户不同时间产生的兴趣评分不加以区分地记录在一个向量中；另一方面，文献[6]中提出的用户-词条-时间三维模型，虽然考虑了用户兴趣的时序性，但倾向分析用户的长期兴趣和遗忘规律，较少考虑用户之间的互动. 综上，现有推荐算法未能充分利用虚拟社区中用户属性和互动信息分析用户的兴趣偏好和评分操作行为，推荐效率与准确性偏低.

2　基于动态信任的工程经验知识推荐框架

通过对国内外关于考虑社会网络的协同过滤方法已有研究成果的分析，结合工程领域设计经验知识的特点，主要实现以下几个目标：

（1）以 MediaWiki 为经验知识积累的载体，建立平台上用户档案和时序性的兴趣评分矩阵，构建修正的兴趣评分矩阵预填充算法，解决带时间窗的评分矩阵数据稀疏性关键问题；

（2）结合平台用户动态信任和预填充的评分矩阵，生成目标用户 V 短期兴趣相似性最高的 K 个最近邻居；

（3）根据目标用户的用户档案产生相应的推荐决策，依据用户的知识需求主动将合适的知识提供给用户，同时解决新用户冷启动问题.

模型建立分为 2 个阶段：① 根据滚动时间窗内用户兴趣评分矩阵，以及定义词条关系强度，设计修正的评分矩阵预填充算法；② 定义 MediaWiki 下用户间的动态信任元，完成模型建立. 最后，实验验证提出的知识推荐方法在实际应用中的有效性和可行性.

3　基于时间窗的 UIT 三维的知识推荐模型

近年来，国内外在研究协同过滤推荐技术上逐渐开始考虑用户兴趣的时序性，用以反映用户兴趣的变化规律. 文献[6]中将推荐模型定义为三元组形式：$R:\{U, I, T\}$. 其中：U 表示参与知识推荐的用户，I 表示词条集合，T 表示时间维. 区分用户在不同时刻的评分行为，能够更好地察觉到用户兴趣随时间推移产生的变化.

本文在现有三维模型研究[6]的基础上，对协同过滤推荐引擎中的时序性问题做了进一步研究. 考虑到工程师在 MediaWiki 平台获取工程经验知识是基于项目背景，词条在大多数情况是独立的，而且若用户对一些特定的经验知识主题确实具有长期兴趣，会反映到用户进行多次搜索、评价或者与其他用户讨论等操作中，因此，本文结合工程经验知识的特点，着重研究了用户的短期兴趣.

如图 1 所示，UIT 三维知识推荐模型定义滚动时间窗，记录用户在不同时间窗内的评分行为. 随着时间窗的滚动，时间窗内较早的评分记录被清除，最新的评价被包含进来.

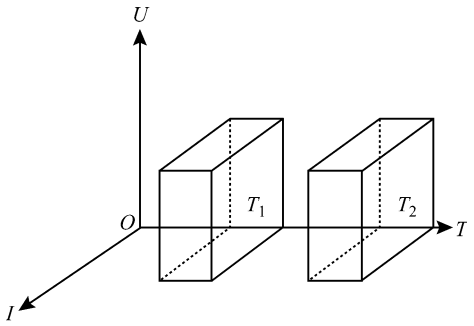

图 1 UIT 三维知识推荐模型

Fig. 1 UIT 3-D recommendation model

基于 UIT 三维知识推荐模型的算法主体分为 5 个阶段：

(1) 建立用户电子档案，包括用户的 ID，用户进行评分操作的词条，用户进行评分操作的具体时间，兴趣评分的 0-1 矩阵，用户部分注册信息等.

(2) 确定时间窗长度. 时间窗内的词条兴趣得分和用户的评分情况反映了用户在一定时间窗长度内的短期兴趣. 算法初期先参考部分文献，给时间窗长度定初值. 为了使算法达到较好的推荐效果，需要不断调整时间窗长度参数，选择多次实验中的最优参数.

(3) 建立时间窗长度内的 0-1 兴趣评分矩阵. 由于数据的稀疏性，需要进行评分矩阵预填充.

(4) 时间窗滚动，更新时间窗内的评分矩阵，直到时间窗达到最新时间.

(5) 进行推荐决策. 若目标用户 V 属于 MediaWiki平台的老用户群，则选择追随率最大的 K 个最近邻居，推送 K 个最近邻居预测兴趣评分最高的前 N_0 条词条. 本文设定 N_0 为 5. 若目标用户 V 是平台的新用户，但是注册信息表明了自己所处的职能部门，则推送与这个职能部门的 K 个用户的预测兴趣评分最高的前 N_0 条词条；否则，推荐所有用户群在当前时间窗被搜索最多的前 N_0 条词条.

知识推荐算法流程如图 2 所示.

本文提出的评分矩阵预填充算法在 1 个时间窗内预测用户对未评分的词条的兴趣，随着时间窗滚动，更新时间窗内的兴趣评分值，最后根据推荐决策为目标用户推荐最可能感兴趣的词条.

3.1 用户兴趣评分矩阵

$\boldsymbol{E}(U,I,T)$ 表示用户-词条-时间三维评分矩阵. 矩阵中每个元素 $e(U_i,I_m,T_n)$ 表示用户 U_i 在第 n 个时间窗对第 m 条词条的评分，用户感兴趣则元素赋值为 1；否则，赋值为 0. 当时间窗向左滚动 1 次后，矩阵的第1列清除，其余各列向左平移1列，第

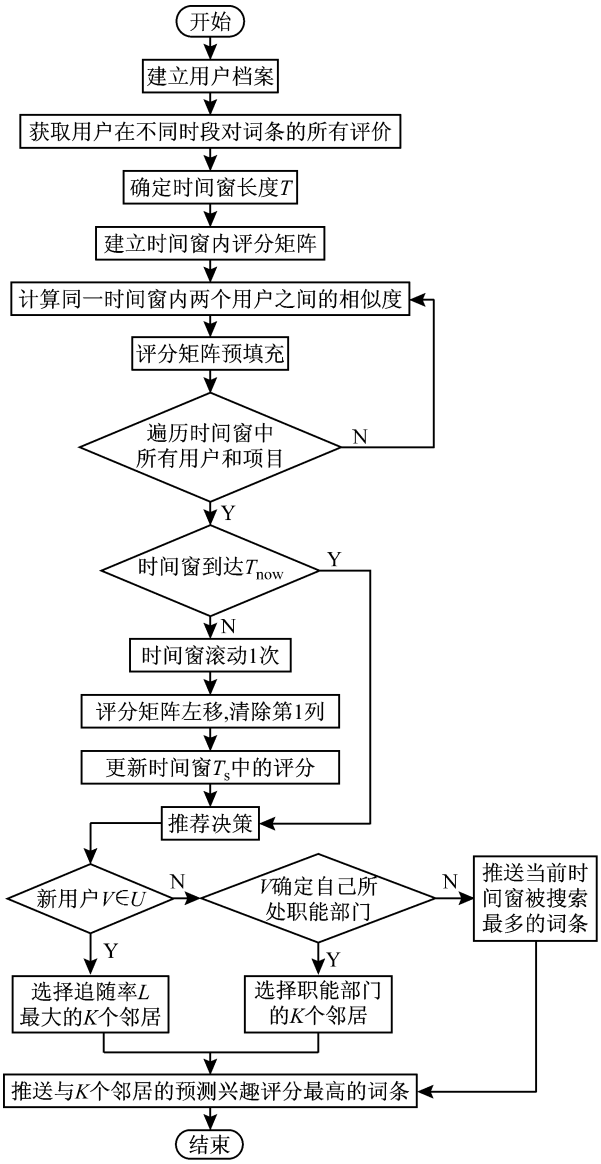

图 2 知识推荐算法流程图

Fig. 2 Flowchart of CFBDT knowledge recommendation algorithm

n 列是用户在新时间窗 T_n 内的兴趣评分，从而得到新的评分矩阵. 兴趣评分矩阵记录了用户对词条的评分，如用户 U_i 的兴趣评分矩阵：

$$\boldsymbol{E}(U_i,I,T)=\begin{matrix}\\ I_1\\ \vdots\\ I_m\end{matrix}\begin{matrix}T_1 & \cdots & T_n\\ \end{matrix}\begin{bmatrix}e_{11} & \cdots & e_{1n}\\ \vdots & & \vdots\\ e_{m1} & \cdots & e_{mn}\end{bmatrix} \tag{1}$$

3.2 考虑用户兴趣变化的评分矩阵预填充

协同过滤算法主要包括基于记忆算法、基于模型算法，以及各种混合算法. 基于记忆的协同过滤算法中，首先根据目标用户以往的行为记录，计算用户之间、词条之间的相似性，向目标用户推荐与其相似度大的现有用户所查询、评分的词条，或者推荐和该目标用户以前查询的词条相似度大的词条. 然而，实

际的 MediaWiki 系统数据量非常庞大，用户评分矩阵十分稀疏，利用传统的基于记忆推荐算法会增大推荐知识条目的误差. 本文考虑基于模型的协同推荐算法，根据用户的历史兴趣评分计算出用户评分行为模型，再根据模型对目标用户在下一个时间窗内的评分行为进行预测. 很多推荐算法都通过计算用户的全局相似度来寻找近邻进行推荐，但是用户的兴趣可能只是在某一方面相似. 因此 MIYAHARA 等[10]提出基于贝叶斯分类的推荐模型，把评分分类分别采用不同的相似用户进行推荐. HOFMANN 等[11]提出概率隐语意(PLSA)模型，引入隐含变量 $Z=\{z_1,z_2,\cdots,z_k\}$，认为用户选择了一个词条是由于某个隐含变量 Z 导致的，隐含变量可以理解成具有某种相同兴趣偏好的社区，也可以理解为某类相似词条所具有的共同属性，并用一个混合高斯分布表示用户对词条评分概率. 这些广泛使用的推荐算法大都是静态模型，只是单纯的整合用户历史数据，对用户在各个时段的评分不加以区别，并未考虑用户兴趣变化情况. 沈建等[6]首次建立用户-词条-时间的三维模型，考虑用户的长期兴趣随时间的变化.

本文基于用户-词条-时间的三维模型，通过引入用户评分时间窗，着重对用户短期兴趣进行建模. 设定时间窗的长度，在用户每次给一个词条新增评分时，往时间窗中添加该评分信息，并且去掉时间窗中对词条最早的评分，形成滚动的时间窗. 如图 3 所示，第 1 个时间窗中包含对词条 I_1、I_2、I_3 和 I_4 的评分；滚动到第 2 个时间窗时，则对词条 I_1 较早的评分 e_{I_1} 被消去；滚动到第 3 个时间窗时，则消去对词条 I_2 和 I_3 较早的评分 e_{I_2} 和 e_{I_3}. 通过消去用户较早评分信息的方法，可以准确地捕捉用户的短期兴趣，提高对用户的未来短期评分行为的预测. 同时，由于用户不可能在每个时间窗对每个词条都留下兴趣评分信息，所以需要对其中空缺的评分进行预填充.

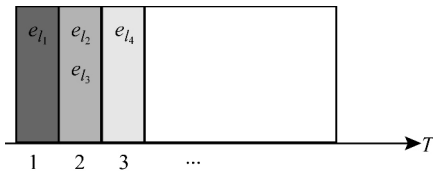

图 3　评分时间窗示意图

Fig. 3　Evaluation time-window

随着用户评分记录的增加，在时间窗内始终保存着用户近期评分词条的记录. 因为用户在某段连续时间内评分的词条有很大的相似性，所以该时间窗可以有效反映用户的短期兴趣. 考虑到工程师在 MediaWiki 平台获取工程经验知识是基于项目背景，词条在大多数情况是独立的，而且若用户对一些特定的经验知识主题确实具有长期兴趣，会反映到用户进行多次搜索、评价或者与其他用户讨论等操作中，因此不考虑用户的长期兴趣. 结合时间窗中的内容，预测用户 u 对词条 I_j 的评分预测公式为

$$e'_{t(U,I)}=\tilde{e}_u+\frac{\sum\limits_{i\in \mathrm{window}(u)}\mathrm{sim}(I_j,I_i)(e_{u,i}-\tilde{e}_i)}{\sum\limits_{i\in \mathrm{window}(u)}\mathrm{sim}(I_j,I_i)} \tag{2}$$

式中：$\tilde{e}_u$ 为该用户对所有评价过的词条给出的平均评分；$\tilde{e}_i$ 为词条 I_i 收到的来自所有用户的平均评分；$\mathrm{window}(u)$ 表示用户近期评分窗口中的所有词条；$\mathrm{sim}(I_j,I_i)$ 表示待预测的词条 I_j 和评分时间窗中词条 I_i 的相似度，

$$\mathrm{sim}(I_j,I_i)=\frac{\boldsymbol{E}_i\cdot\boldsymbol{E}_j}{\|\boldsymbol{E}_i\|\cdot\|\boldsymbol{E}_j\|} \tag{3}$$

式中，$\boldsymbol{E}_i$ 和 $\boldsymbol{E}_j$ 分别为词条 I_i 和 I_j 收到的来自所有用户的评分向量. 对于 $\mathrm{sim}(I_j,I_i)$，陈登科等[12]在评分预测模型上采用传统的评分余弦相似度进行计算. 这种融合算法是预测模型和基于词条相似性模型的迭代，虽然能提高准确率，但是运算时间太长，不适用于 MediaWiki 平台实时响应的场合.

在现有研究中，$\mathrm{sim}(I_j,I_i)$ 的计算将所有评分视为相互独立，相互的关系强度 s 都默认为 1. 然而，在实际过程中，计算词条两两的相似度时，考虑多个用户可能在评分时间窗中同时对一条词条进行了兴趣评分，被重合评分的词条之间的关系强度更大，换而言之，被评分的次数越少的词条之间的关系强度应该越小. 将在同一个时间窗对一个词条 I_i 进行了评分的用户集合记为 U_{I_i}，在同一个时间窗对一个词条 I_j 进行了评分的用户集合记为 U_{I_j}，则词条 I_i 和 I_j 之间的关系强度为

$$s=\frac{N(U_{I_i}\cap U_{I_j})}{N(U_{I_i}\cup U_{I_j})} \tag{4}$$

式中：$N(U_{I_i}\cap U_{I_j})$ 表示对词条 I_i 和词条 I_j 同时进行评分的用户人数；$N(U_{I_i}\cup U_{I_j})$ 表示参与对词条 I_i 或词条 I_j 评分的总的用户人数. 因此，将关系强度 s 作为修正因子，则词条 I_i 和 I_j 之间的相似度修正为

$$\mathrm{sim}'(I_j,I_i)=s\frac{\boldsymbol{E}_i\cdot\boldsymbol{E}_j}{\|\boldsymbol{E}_i\|\cdot\|\boldsymbol{E}_j\|} \tag{5}$$

将 $\mathrm{sim}'(I_j,I_i)$ 代入评分预测公式，可得

$$e'_{t(U,I)}=\tilde{e}_u+\frac{\sum\limits_{i\in \mathrm{window}(t)}\mathrm{sim}'(I_j,I_i)(e_{t(U,I)}-\tilde{e}_t)}{\sum\limits_{i\in \mathrm{window}(u)}\mathrm{sim}'(I_j,I_i)} \tag{6}$$

基于用户兴趣变化的评分矩阵预填充算法：

(1) 调用用户历史评分记录；

(2) 确定时间窗的大小；

(3) 从窗口获取近期评分词条，求出用户对所有当前时间窗内未评分词条的短期预测评分 $e'_{t(U,I)}$；

(4) 当用户进行新的评分时，记录用户评分信息，并更新窗口内容；

(5) 时间窗滚动，若已遍历历史评分记录中的所有时间窗则结束，否则转步骤(3).

3.3 用户相似度计算

用户相似度计算的目的是找到与目标用户 U_i 具有相同或相似爱好的其他用户 U_j，称为最近邻居. 通过预填充的用户-词条兴趣矩阵得到，用户相似度：

$$\mathrm{sim}(U_i,U_j)=\frac{\boldsymbol{E}_{U_i}\cdot\boldsymbol{E}_{U_j}}{\|\boldsymbol{E}_{U_i}\|\cdot\|\boldsymbol{E}_{U_j}\|}=\frac{\sum_{t=1}^{n}\sum_{I=1}^{m}(e'_{U_i(I,t)}e'_{U_j(I,t)})}{\sqrt{\sum_{t=1}^{n}\sum_{I=1}^{m}e^2_{U_i(I,t)}}\cdot\sqrt{\sum_{t=1}^{n}\sum_{I=1}^{m}e^2_{U_j(I,t)}}} \tag{7}$$

为了寻找目标用户的最近邻居集合，需要度量用户相似度的参考意义. 在传统网络中，由于用户兴趣偏好或需求的资源往往存在巨大的差异，这样找出来的最近邻居集合不够准确. MediaWiki 社区由一定数量的用户在平台上经过一段时间的交流、讨论，彼此拥有足够的信任. 因此，社区用户在做出某项决定之前往往会咨询社区内的其他用户，寻求推荐意见.

针对于虚拟社区中 MediaWiki 平台上用户之间存在的关系，动态信任是在信息咨询和反馈中建立的用户之间的信任程度，是一个随着时间和用户之间的交互而变化的动态矢量. 记 $D(u_i,u_j)$ 为社区用户 u_i 对社区用户 u_j 的动态信任程度，$0\leqslant D\leqslant 1$. 用户之间的动态信任是单向的. k 表示用户 u_i 向用户 u_j 发出的信息咨询的次数，$f_k(u_i,u_j)$ 表示用户 u_j 是否对用户 u_i 的第 k 次咨询进行反馈. 若用户 u_j 对 u_i 的信息咨询进行了回应，则 $f_k(u_i,u_j)$ 取值为 1，反之为 0. 考虑时间因素，用户之间最近的信息交互操作被赋予较大的权重，过去的信息交互则赋予较小的权重，记用户每次操作的权重系数为 $w^{-\tau(t_{\mathrm{now}}-t_k)}$. 根据上述分析，则动态信任为

$$D(u_i,u_j)=\frac{\sum_{k=1}^{n}w^{-\tau(t_{\mathrm{now}}-t_k)}f_k(u_i,u_j)}{\sum_{k=1}^{n}w^{-t(t_{\mathrm{now}}-t_k)}} \tag{8}$$

式中：t_{now} 表示当前时间；t_k 表示用户的信息咨询请求最后得到反馈的时间；τ 是一个修正量，本文取值 1/365，防止动态信任元下降过快.

结合基于社会网络分析的用户间的动态信任，则目标用户 V 对其最近邻居推荐的词条的追随率记为

$$L=D(U_i,U_j)\mathrm{sim}(U_i,U_j) \tag{9}$$

3.4 目标用户 V 的推荐决策

MediaWiki 的工程经验知识推送机制的内涵是：当目标用户 V 访问 MediaWiki，进行查询、询问其他用户、评分等操作，期望得到特定的知识，MediaWiki 将会为用户 V 推送词条列表 L. 若目标用户 V 至少对一条词条进行评分，即 $V\in U:\{u_1,u_2,\cdots,u_i\}$，则选择 L 值最大的 K 个最近邻居，根据这 K 个最近邻居，预测目标用户 V 在当前时间窗对所有词条的兴趣评分矩阵的元素，取 $\max\{e'_{t(U,I)},e_{t(U,I)}\}$ 对应行的词条，推荐给目标用户 V.

若目标用户 V 是新用户，很少进行词条查询、询问其他用户和评分操作，则遇到冷启动问题，即矩阵稀疏性问题的极端情况，也称为第 1 评价人问题或早期评价人问题. 因为协同过滤推荐算法对用户的分类主要依据目标用户与其他用户的比较，需要基于不断累积的用户评分. 如果一个新用户从未对系统中的词条进行评分，则系统无法获知新用户的兴趣点，也就无法为其进行推荐. 在协同过滤推荐系统刚投入运行时，每个用户在每个词条上都将面临冷启动问题.

近年来，出现了许多解决冷启动的方法. 一些推荐系统在用户注册时会询问一些问题，从而发掘用户兴趣. 一些学者提出通过挖掘可用的信息来减轻冷启动问题带来的影响，如隐语义模型[13]. 有些学者提出了混合算法，通过结合基于内容的推荐算法和协同过滤算法来应对冷启动问题[14]. 文献[15]中提出使用社会网络分析来解决冷启动问题的方法.

本文通过对最终的推荐决策分类来解决冷启动问题. 若目标用户 V 在注册时已经给自己所处的职能定位，则选择与用户具有相同职能的最近邻居，根据最近邻居的短期兴趣进行推荐. 若目标用户 V 没有提供任何参考信息，则推送现有用户群最感兴趣的 N_0 个知识条目，即当前时间窗下用户群搜索次数最多的 N_0 个知识条目.

4 实验与算法验证

受多媒体检索技术的数据特征模型的启发，本文针对文档类的经验知识的特征，将经验知识的结构分为 3 个层次，自顶向下依次为基本属性、语义特

征、数据内容. 前两者属于知识的外部特征的标签. 基本属性包含贡献者、经验知识 ID 号、名称以及创建修改时间. 语义特征主要是反映工程领域特定背景和特定工艺的关键字标签. 数据内容包含知识的内容特征.

图 4 所示是 MediaWiki 平台下的汽车研发经验知识示例，包含以上 3 种结构层次，即词条名称，词条和贡献者 ID 号、问题描述，解决方法等.

图 4　汽车研发经验知识示例

Fig. 4　An example of EEK in automotive R and D

4.1　数据预处理和评价指标

整理和筛选 2014 年汽车设计网论坛发布的经验知识和评价数据，建立用户-词条测试集 S_{test}，其中包含用户 105 个，词条数 97 条，打分次数 592 次. 由此计算得到数据稀疏度为 $1-592/(105\times 97)=0.942$，矩阵非常稀疏.

实验采用命中率和预测精确度作为评价标准. 命中率表示：用户在下一个时间窗中浏览的词条存在于当前推荐词条列表中，则记为当前时刻对该用户命中；若用户在下一个时间窗中无任何行为则不考虑其命中情况. 综合各个用户的命中情况，求出该时间窗下的命中率.

预测精确度是评价推荐系统好坏的一项重要指标，平均绝对误差(MAE)是一种常用的评价推荐系统精确度指标，是测试集中所有用户对资源打分的实际值与预测值的绝对值的平均. 对于在测试集 S_{test} 中的 N 组实际评分值-预测值对 $(e_{U,I},\hat{e}_{U,I})$，平均绝对误差的计算公式为

$$\text{MAE}=\frac{1}{N}\sum_{\langle U,I\rangle\in S_{\text{test}}}|e_{U,I}-\hat{e}_{U,I}| \qquad (10)$$

MAE 值越小说明推荐算法的预测精度越高.

4.2　算法参数实验

本文提出的模型及算法中，时间窗长度的选取至关重要. 时间窗内的词条兴趣得分和用户的评分情况反映了用户在一定时间窗长度内的短期兴趣.

由图 5 可见，随着时间窗长度增加，知识推荐的平均命中率增高，但是当时间窗长度数增加到 5 周，命中率会逐渐达到饱和，命中率提高的幅度非常有限，因此时间窗无需太长，本算法选择时间窗长度为 5 周.

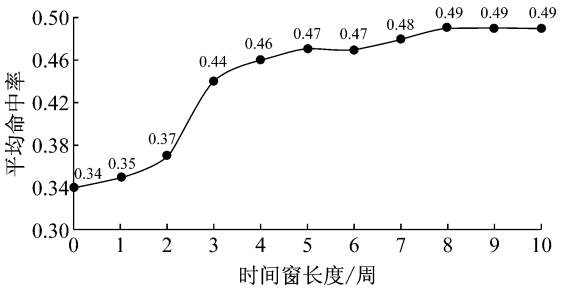

图 5　不同时间窗长度下的算法平均命中率

Fig. 5　Hitrate within value of varied time-window

4.3　修正的评分矩阵预填充算法与现有算法的效果比较

本文在 5 种最近邻居集合规模{2,5,10,15,20}上，分别对经典的协同过滤算法(CCF)[11]、基于词条评分相似性进行评分填充的协同过滤(CFBRF)算法[16]、以及彭石等[17]提出的基于加权 Jaccard 系数的词条综合相似度进行评分矩阵预填充(CFB-WJI)算法及本文算法(CFBDT)进行对比实验.

由图 6 可见，当最近邻居的规模取 2，并依此给用户 V 推荐工程经验知识时，本文和其他算法的 MAE 值均较大，CFBWJI 算法效果稍强，原因是在参考的最近邻居较少时，CFBWJI 算法通过修正后的词条综合相似度进行预评分，而其他 3 种算法受最近邻居的影响，评分矩阵极其稀疏，导致推荐的误差较大. 然而，由于本算法综合考虑了与最近邻居的兴趣相似度和用户之间积累的动态信任，并通过定义词条间的关系强度来修正评分矩阵预填充算法，所以，当最近邻居规模较大，如 $K=10$ 时，本算法的 MAE 值降低到 0.62，相比于 CFBWJI 算法降低了 22.5%，而传统的 CCF 算法的 MAE 值高达 0.85，CFBRF 算法的 MAE 值达到 0.78. 随着最近邻居的规模超过 10，各个算法的 MAE 值都有所降低，降低幅度逐渐减小. 当参考 20 位最近邻居时，本算法的 MAE 值小于 0.5，相比于 CFBWJI 算法降低了近 24%.

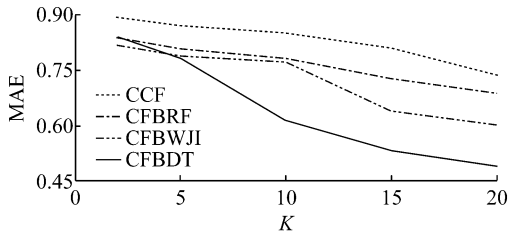

图 6 不同最近邻居规模下 4 种算法的比较

Fig. 6 MAE of the four algorithms in value of varied nearest neighbor

当用户浏览一条经验知识词条，并认为有必要得到知识推荐的建议时，可以启动本文研究的知识推荐方法，自动查找用户潜在最感兴趣的词条. 图 7 显示了对于 VTEC 可变气门正时及升程电子控制系统这一经验知识条目的推荐结果. 根据算法的设定和推荐决策，选取前 5 条知识条目推荐给目标用户 V，并分别显示了知识条目创建者的用户 ID 和词条名称. 至此，目标用户 V 可以选择推荐列表中的词条进行查阅学习.

图 7 对目标用户 V 的知识推荐结果

Fig. 7 An example of knowledge recommended for target user V

5 结 语

本文基于开放的 MediaWiki 平台提出了一个新的知识推荐方法. 结合 MediaWiki 平台的特征和隐性知识积累的形式，本文 CFBDT 方法的优势体现在：① 考虑目标用户对社交网络中其他用户的信任度，研究了针对 UIT 三维动态模型的工程经验知识的协同过滤推荐算法；② 针对工程经验知识的特点，考虑用户短期兴趣的时序性，提出一种基于用户兴趣时序性变化的评分矩阵预填充算法，首次考虑词条间的关系强度用以修正用户评分的预测值，解决用户-词条评分矩阵稀疏性问题；③ 在同一时间窗长度内，充分地考虑 MediaWiki 社区用户群之间的信任度的动态变化，通过定义追随率，反映具有相似兴趣的用户对知识推荐的参考性发生的变化，提高算法的精度.

用户的兴趣和需求随着与时间和空间相关的情境的改变而有所不同. 因此，如何结合社会网络分析

中用户之间的关系和用户所处的工作情境，更准确地预测用户的短期偏好信息是本文接下来的研究重点.

参考文献：

[1] 黄咏文. 工程领域设计经验知识的积累和重用方法研究[D]. 上海：上海交通大学机械与动力工程学院，2015.

[2] YANG Xiwang, GUO Yang, LIU Yong. A survey of collaborative filtering based social recommender systems [J]. **Computer Communications**, 2014, 41(5): 1-10.

[3] HOHFELD A, GRATZ P, BECK A, *et al*. Self-organizing collaborative filtering in global-scale massive multi-user virtual environments [C]// **Proceedings of the 2009 ACM Symposium on Applied Computing**. NY: ACM, 2009: 1719-1723.

[4] ESSLIMANI I, BRUN A, BOYER A. From social networks to behavioral networks in recommender systems [C]// **Proceedings of the International Conference on Advances in Social Network Analysis and Mining**. NY: ACM, 2009: 143-148.

[5] PIZZATO L A, REJ T, ChUNG T, *et al*. RECON: A reciprocal recommender for online dating[C]// **Proceeding of the 4th ACM Conference on Recommendation System**. NY: ACM, 2010.

[6] 沈键，杨煜普. 基于滚动时间窗的动态协同过滤推荐模型及算法[J]. 计算机科学，2013，40(2)：206-209.
SHEN Jian, YANG Yupu. Dynamic collaboration filtering model based on rolling time window and its algorithm [J]. **Computer Science**, 2013, 40(2): 206-209.

[7] LI Dan, CAO Peng, GUO Yucai, *et al*. Time weight update model based on the memory principle in collaborative filtering [J]. **Journal of Computers**, 2013, 8(11): 2763-2767.

[8] XIANG Liang, YANG Qing. Time-dependent models in collaborative filtering based recommender system [C]// **Proceedings of the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology**. UK: IET, 2009: 450-457.

[9] KOREN Y. Collaborative filtering with temporal dynamics[C]// **Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. NY: ACM, 2009: 447-456.

[10] MIYAHARA K, PAZZANI M J. Collaborative filtering with the simple Bayesian classifier[C]// **Proceedings of the 6th Pacific Rim International Conference on Artificial Intelligence**. Berlin: Springer, 2000: 679-689.

[11] HOFMANN T, PUZICHA J. Latent class models for collaborative filtering[C]// **proceeding of the 16th International Joint Conference in Artificial Intelligence**. California: IJCAI, 1999: 688-693.

[12] 陈登科，孔繁胜. 基于高斯 pLSA 模型与项目的协同过滤混合推荐[J]. 计算机工程与应用，2010，46(23)：209-211.
CHEN Dengke, KONG Fansheng. Hybrid Gaussian pLSA model and item based collaborative filtering recommendation[J]. **Computer Engineering and Applications**, 2010, 46(23): 209-211.

[13] ADOMAVOCIUS G, TUZHILIN A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions [J]. **IEEE Transactions on Knowledge and Data Engineering**, 2005, 17 (6): 734-749.

[14] LEUNG W K, CHAN C F, CHUNG F L. An empirical study of a cross level association rule mining approach to cold start recommendations [J]. **Knowledge Based Systems**, 2008, 21(7): 515-529.

[15] CASTILLEJO E, ALMEIDA A, DIEHO L I. Alleviating cold-user start problem with user's social network data in recommendation systems [EB/OL]. (2013-02-09)[2015-08-03]. http://morelab.deusto.es/publications/info/alleviating-cold-user-start-problem-with-users-social-network-data-in-recommendation-systems/.

[16] LEMIREL D, MACLACHLAN A. Slope one predictors for online rating-based collaborative filtering. Computer Science [C]// **Proceedings of 2005 SIAM International Conference on Data Mining (SDM'05)**. California: Newport Beach, 2005.

[17] 彭石，周志彬，王国军. 基于评分矩阵预填充的协同过滤算法[J]. 计算机工程，2013，39(1)：175-182.
PENG Shi, ZHOU Zhibin, WANG Guojun. Collaborative filtering algorithm based on rating matrix prefilling [J]. **Computer Engineering**, 2013, 39(1): 175-182.