

学习资源精准推荐模型及应用研究

王晓东,时俊雅,李淳,吴慧萍

(河南师范大学 计算机与信息工程学院,河南 新乡 453007)

摘要:根据属性特征推荐资源,由于存在冷启动和稀疏性问题,限制了在线学习资源推荐的性能.基于知识表示和协同过滤,将学习者的学习水平和学习风格等特征融入推荐过程,进行协同过滤个性化推荐,提出了一种学习资源精准推荐模型,构建了学习者和学习资源知识表示模型;通过实验表明知识表示-协同过滤相结合的推荐算法在个性化推荐和推荐准确度方面优于传统的 CF 算法.

关键词:知识表示;学习资源推荐;协同过滤;在线学习

中图分类号:TP391

文献标志码:A

推荐系统在解决“信息过载的问题”上表现优异.传统的推荐技术,包括协同过滤(Collaborative Filtering, CF)和基于内容的推荐(Content-Based, CB),已在多个领域得到应用,如:淘宝为用户推荐商品、新闻推荐、亚马逊图书推荐、Netflix 电影推荐和 Coursera 课程推荐等^[1-2].在 e-learning 环境中,学习者具有不同属性特征(包括动态特征和静态特征),如学习动机、学习水平、学习风格与偏好等,这些学习者特征都会影响学习者的学习.一般的推荐技术由于冷启动和稀疏性问题,限制了在线学习推荐系统的性能^[3-6],即当存在新的未进行评价的学习资源或者没有评论过任何项目的新用户时,会导致无法进行精准推荐,随着系统规模的扩大,学习者对学习资源的评分数据就越来越庞大,获取这些数据会变得很稀疏^[7].因此,提高个性化学习资源推荐精度及提升性能是目前亟待解决的问题.

针对上述问题,本工作基于知识表示(Knowledge Representation, KR)和 CF,研究学习资源推荐算法.为提高推荐个性化和准确性,利用知识表示将学习者学习水平、学习风格和学习偏好等属性特征融入推荐过程中,利用知识之间的语义关系缓解冷启动问题;通过协同过滤算法,解决稀疏性问题.

1 模型构建与精准推荐算法

将学习者在学习过程中的属性特征和学习资源特点作为设计学习者 KR 和学习资源 KR 模型的依据.在强调学习者自主学习的基础上,基于 CF 和 KR,构建学习资源精准推荐模型,如图 1 所示.该模型中主要包含 4 个部分:学习者 KR 模型、学习资源 KR 模型、CF 推荐引擎和个性化学习资源推荐.本文将详细解释推荐模型是如何工作的.

(1)学习者 KR 模型:该模型主要包括学习者的基本信息、偏好信息和学习者属性特征信息.学习者的基本信息可通过显式和隐式获取;学习者 KR 模型中存储学习者属性特征,包括学习水平、学习动机、学习风格等信息.学习者 KR 模型根据学习者的偏好和属性特征对学习者的个人资料进行个性化分析,而 CF 推荐引擎利用学习者和学习资源 KR 信息一起进行评分预测,为学习者进行个性化推荐.

(2)学习对象模型:该部分存储了学习资源的信息,包括文本、图像、动画、音频或视频等多种格式.

(3)数据预处理:将数据预处理组件中的学习者和学习资源数据准备好并预处理成推荐引擎可以识别的正确格式.

收稿日期:2018-03-06;**修回日期:**2018-11-22.

基金项目:横向研究项目(5201119160004);河南师范大学研究生科研创新项目(YL201623).

作者简介(通信作者):王晓东(1963-),男,河南永城人,河南师范大学教授,博士,博士生导师,研究方向为本体工程、教育大数据, E-mail: wxd@htu.cn.

(4)推荐引擎:一旦数据预处理成功,推荐引擎就根据学习者和学习资源 KR 来计算目标学习者评分的相似度和对目标学习者的评分预测.最后,推荐引擎为目标学习者生成个性化推荐列表.

1.1 学习者和学习资源 KR 模型

KR 是学科知识与教学规律的组织方式,其实质是知识的符号化形式,主要是为了方便计算机对知识进行存储和处理.目前,更多的 KR 技术包括谓词逻辑表示法、Web 表示法、产生式规则、语义网络和框架表示法等^[8].

1.1.1 学习者 KR 模型

由于个性化学习资源推荐需要了解学习者信息,根据学习者的学习目标、学习需求、教学内容、学习问题与环境、偏好等提供学习资源,因此需要建立学习者模型.在教育领域对学习资源知识

概念进行检索,基于 KR 的学习者模型可以根据概念之间的联系,将与学习检索关系密切的一系列概念反馈给学习者,有利于学习者发现学习目标、兴趣领域与学习兴趣相联系的新概念,并且通过对概念的属性和关系的描述更好的表达学生的偏好,从而对其进行个性化的推荐.学习者模型 U_R 的形式化定义如下: $U_R = \{U_b, U_o, U_n, U_h, U_p, U_v\}$,其中用户的个人信息 $U_b = \{ID, Name, Age, Sex, Edu, Tel\}$,包括用户名、姓名、年龄、教育背景、性别、电话等个人基本信息.而学习者 KR 偏好本体为: $U_o = \{C, R_c, R_n, F_c, A, I\}$ 其中, C 表示本体中学习者偏好概念; R_c 描述了学习者偏好本体中概念间的分类关系; R_n 描述了学习者偏好本体中的非分类关系(属性、同位等); F_c 表示函数; A 表示公理; I 表示实例. U_n 为学习风格信息,学习风格分为 4 个维度,主要分为活跃型/沉思型、感悟型/直觉型、视觉型/言语型、综合型/序列型,即学习者的学习风格是由 4 个维度排列组合构成;本文中不同的学习风格表示为: $U_n = \{\text{活跃型/沉思型、感悟型/直觉型、视觉型/言语型、综合型/序列型}\} = \{1, 2, 3, 4\}$. U_h 为学习水平信息,为了获取学习者的学习水平,随机设置一个包含 10 个问题进行学习水平测试,通过测试情况为学习者进行学习水平的分配,其中{初级=0~3、中级=4~6、高级=7~10};对于不同的学习水平用 U_h 来表示, $U_h = \{\text{初级、中级、高级}\} = \{1, 2, 3\}$. U_p 为用户偏好信息,描述用户对界面、字体、语言、资源内容类型和资源媒体类型等的偏好. U_v 为访问日志信息,主要记录学习者登录学习系统的时间、访问的学习资源、开始访问学习资源的时间、访问完成的时间和退出系统的时间,可以帮助学习者判断哪个时间段的学习效率高.其中,学习者 KR 模型如图 2 所示.

构建学习者 KR 模型需要囊括学习者的个人数据信息、学习水平、学习风格信息和访问日志信息等,并通过显式和隐式两种方法获取学习者的资料信息.而且,将学习者资料信息如学习者偏好、个人基本信息(姓名、性别、年龄等)、学习风格和学习水平等属性特征存储在学习者 KR 模型中.一旦获取学习者的数据信息,构建的学习者模型就会自动更新,根据学习者的偏好、学习风格和学习水平等特点,形成个性化的学习者知识,CF 推荐引擎根据学习者 KR 信息和学习资源 KR 信息来进行评分预测,将预测评分高的学习对象推荐给目标学习者,进而对目标学习者进行个性化的推荐.当新的学习者进入该模型时,会根据学习者注册信息进行语义搜索,分析学习者的属性特征,确定学习者的偏好,并和学习者 KR 模型进行匹配,从而缓解冷启动问题.

1.1.2 学习资源 KR 模型

在线教育环境中,学习资源的结构存在复杂性和多样性,其表现形式包括课件、案例、文献资料、资源目录索引、网络课程、试题、试卷、作业、文本、动画、论坛答疑、教案和媒体等多种形式;为了更好地建设学习资源,促进各级各类学习资源之间的数据共享,提高学习资源检索的效率与准确度,根据学习资源对象的特点

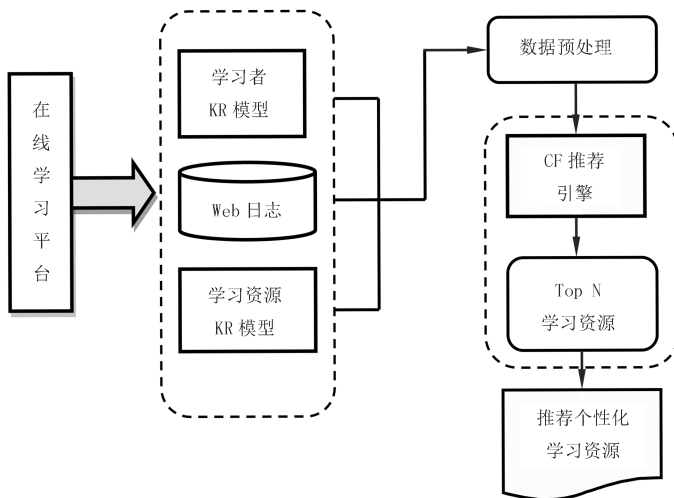


图 1 学习资源精准推荐模型

Fig.1 The accurate recommendation model of learning resources

和资源存储的无序性、非结构性,将学习资源划分为以下几类:文本类资料、媒体类资料.其中文本类资料包括课件、案例、试卷和课本资料等;媒体资料包括图形/图像资料、动画资料、音频资料、视频资料.为了能够对学习资源进行结构化重组和描述,本文采用基于元数据(Meta Data)的 KR 本体构建方法(Metadata-based knowledge representation Ontology Buidling, 简称为 MOB 方法)来定义领域知识,这种方法具有规范性、合理性和可扩展性等特点.

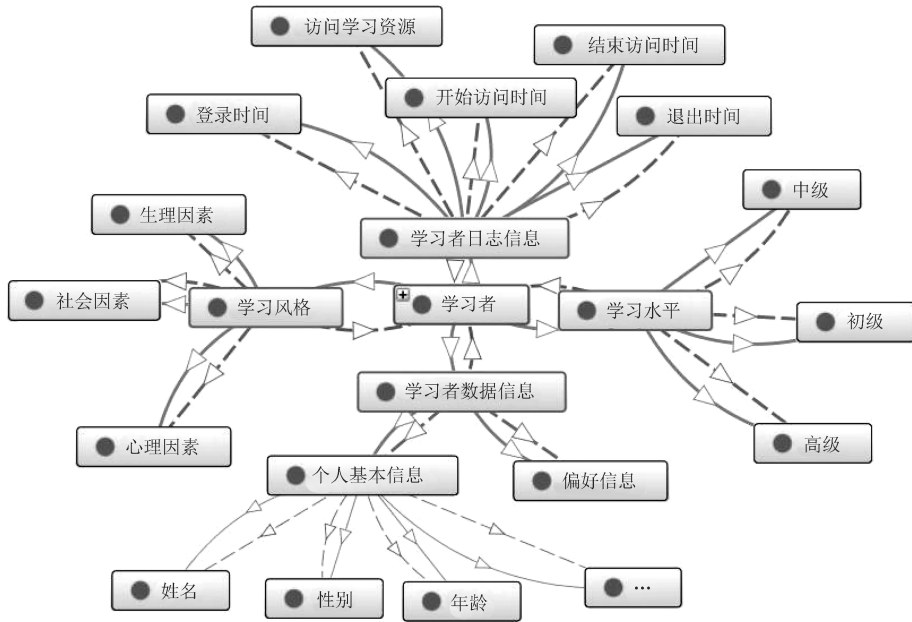


图2 学习者 KR 模型

Fig.2 The KR model of learner

首先,构建学习资源 KR 模型能够实现基于知识点的学习资源的管理和检索,在构建学习资源描述本体时利用计算学科知识点的本体进行表示,通过关联知识点属性对知识点本体进行映射;其次,从元数据中提取术语列表:媒体资料类、视频/音频类、文本类、课件类、网络课程类、标题、资源位置、姓名、联系方式和关联关系等;最后,定义类和确立类的层次,根据分类资源的层级来定义类,学习对象为父类(LearningObject)包含课件类、文本类和媒体资料类等这些子类,媒体资料类为文本类、视/音频类、图形/图画类等子类;建立类之间的层次关系主要利用“part-of”,“kind-of”“instance-of”和“attribute-of”等.学习资源 KR 模型中类的层级结构图如图 3 所示.

构建学习资源 KR 模型利用本体描述语言 OWL 和 protégé 工具,用来表示学习者和学习者之间、学习者和学习资源之间、学习资源中知识点之间的概念和实体之间的关系.所有的学习资源根据各知识点的属性特征进行分类,并且在学习实践课堂中进行举例和小测试,这些小测试和例子与学习目标息息相关.CF 推荐引擎利用学习者和学习资源之间的语义关系对学习者的评分预测,同时预测目标学习者的相似性.然后,根据在线学习平台提供的数据集构建学习者和学习资源 KR 模型,将他们与 Web 日志数据一起预处理成 CF

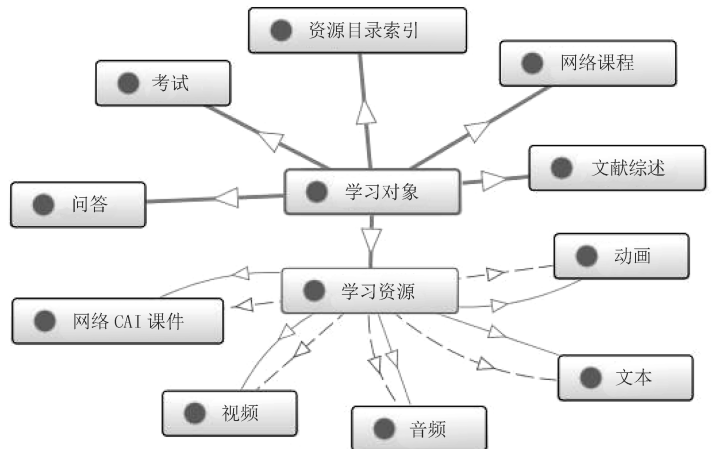


图3 学习资源 KR 模型

Fig.3 The KR model of learning resources

推荐引擎所需要的格式。

以数据结构课程为例进行应用,主要从以下几个方面入手:

- (1) 数据结构课程的关键知识点;
- (2) 与知识点相关的内容,例如学习目标、学习的重点和难点;
- (3) 前后知识点之间的关系.课程的学习往往具有前后性,在学习某一个知识点之前必须要先学会另外的知识点;

(4) 举例解释概念,用一些例子或者课前导入来说明概念,帮助对概念的理解.为了能够更准确地获取资源的完整性,通过使用约束条件和类之间的一些特殊关系来建立了类的层次结构,规范的提取类与类之间的层次关系.以数据结构为例构建的部分类层次结构如图 4 所示。

1.2 CF 相似度计算和预测

CF 推荐系统的推荐列表是这样生成的,首先寻找几个目标学习者的最近邻居,然后提取最近邻居对学习资源的评分数据,最后再根据这些评分数据预测目标学习者的学习资源的评分,从而生成推荐列表.对算法来说数据集是极其重要的,将数据集定义为 $Data = (U, I, R)$, 其中 $U = \{u_1, u_2, u_3, \dots, u_m\}$ 是基本的学习者集合, $|U| = m$; $I = \{i_1, i_2, i_3, \dots, i_n\}$ 是学习资源

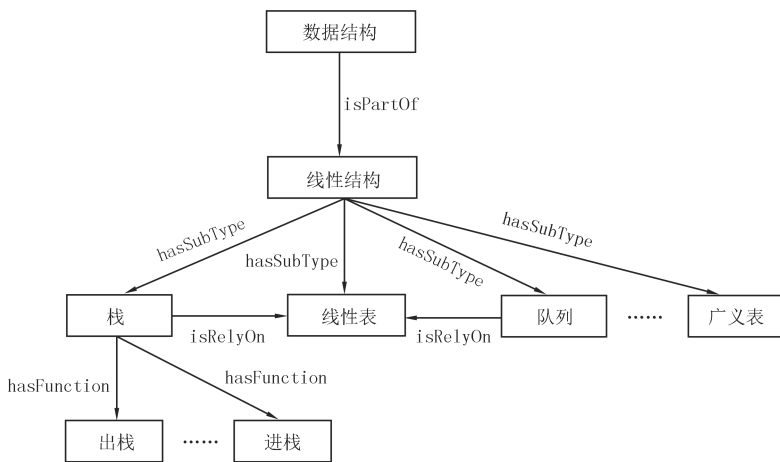


图 4 《数据结构》部分类层次结构图
Fig.4 The hierarchy structure diagram of partial classes in "Data Structure"

集合, $|I| = n$. $m * n$ 阶矩阵 R 是学习者对各学习资源的评分矩阵,其中的元素 r_{ij} 表示 U 中第 i 个用户对 I 中第 j 个学习资源的评分。

1.2.1 相似度计算

CF 推荐的关键是准确定位目标学习者的最近邻居,而确定最近邻居的根本是计算学习者之间的相似度,主要有下面 3 种常用的计算方法。

A) Pearson 相关相似度^[9]: Pearson 与修正的余弦相似度不同的是分母为用户的共同评分项目,如下所示,

$$S(u, v) = \frac{\sum_{a \in P_{uv}} (R_{u,a} - \bar{R}_u) (R_{v,a} - \bar{R}_v)}{\sqrt{\sum_{a \in P_{uv}} (R_{u,a} - \bar{R}_u)^2} \sqrt{\sum_{a \in P_{uv}} (R_{v,a} - \bar{R}_v)^2}} \tag{1}$$

B) 余弦相似性度 (Cosine)^[10]:

$$S(u, v) = \frac{u * v}{\|u\| * \|v\|} \tag{2}$$

$S(u, v)$ 的值越接近 1 说明用户 u 和用户 v 的相似性越高。

C) 修正的余弦相似度 (Adjusted Cosine)^[11]: 鉴于传统的余弦相似度没有考虑用户的评分偏好,也就是说有的用户喜欢评分高的有的喜欢评分低的,比如两个学生对两门课程物理和数学的评分分别为 5、4 和 3、2,如果按照传统的余弦相似度计算发现两个用户的相似度很低,不过其实这两个学生的偏好是一致的,也就是说相对于数学课两个学生更喜欢物理.如(3)式所示是两个用户 u 和 v 的修正余弦相似度。

$$S(u, v) = \frac{\sum_{\alpha \in P_{uv}} (R_{u,\alpha} - \bar{R}_u)(R_{v,\alpha} - \bar{R}_v)}{\sqrt{\sum_{\alpha \in P_u} (R_{u,\alpha} - \bar{R}_u)^2} \sqrt{\sum_{\alpha \in P_v} (R_{v,\alpha} - \bar{R}_v)^2}}, \quad (3)$$

其中, P_{uv} 表示评分资源项目集合, P_u 和 P_v 分别表示用户 u 和 v 评分过的资源项目, $R_{u,\alpha}$ 和 $R_{v,\alpha}$ 表示对资源项目 α 的评分, \bar{R} 表示平均评分.

本文需要考虑本体领域知识和用户对学习资源的评分,因此,(4)式使用修正的余弦相似性进行计算.学习资源对象 i 和 j 的 KR 相似性计算公式 $S(i, j)$:

$$S(i, j) = \frac{\sum (R_{l,i} - \bar{R}_l)(R_{l,j} - \bar{R}_l)}{\sqrt{\sum (R_{l,i} - \bar{R}_l)^2} \sqrt{\sum (R_{l,j} - \bar{R}_l)^2}}, \quad (4)$$

其中, $R_{l,i}$ 是指学习者 l 对学习资源对象 i 的评分, $R_{l,j}$ 是指学习者 l 对学习资源对象 j 的评分; \bar{R}_l 表示的是对学习资源对象 l 提供的所有评分的平均值.

最后计算目标学习者对学习资源对象的预测评分.从(4)式中获得 N 个最相似的学习资源对象.计算预测评分的具体如(5)式所示:

$$P_{l,i} = \frac{\sum_{t \in N} (S(i, t) \times R_{l,t})}{\sum_{t \in N} \| S(i, t) \|}, \quad (5)$$

其中, N 是与学习资源对象 i 相似的学习资源对象集合, $R_{l,t}$ 是学习者 l 对学习资源对象 t 的评分.

1.3 推荐算法

根据目标学习者对学习资源的预测评分、学习者知识表示和学习资源知识表示模型,利用 CF 推荐引擎生成学习资源对象的 top- N 推荐列表.本文用 L 代表所有的学习者集合,即 $L = \{l_1, l_2, l_3, \dots, l_n\}$;用 I 表示所有可能推荐的学习资源,即 $I = \{i_1, i_2, i_3, \dots, i_n\}$;用 k 代表学习者和学习资源领域知识,即 $k = \{k_1, k_2, k_3, \dots, k_q\}$;用 R 表示学习者对学习资源的评分情况,并将评分范围定义为 $R = \{1, 2, 3, 4, 5\}$.算法 1 描述如下.

算法 1:生成 Top- N 学习资源推荐列表

输入:

学习资源对象集合

$I = \{i_1, i_2, i_3, \dots, i_n\}$

KR

$k = \{\text{学习者}, \text{学习资源}\}$

学习者 R 的评分值 $R = \{1, 2, 3, 4, 5\}$

输出:预测评分和 top N 学习资源推荐列表

方法:

步骤 1 for each

$i \in I, j \in I, n \in N$, do

步骤 2 利用(4)式计算相似度 $S(i, j)$

End for each

步骤 3 利用(5)式计算预测评分 $P_{l,i}$

步骤 4 将预测评分最高的前 N 学习资源对象为目标学习者 l_i 生成 top- N 学习资源推荐列表.

2 模型应用与实验分析

2.1 模型应用

本文将模型应用在自主研发的“移动自主学堂”在线学习平台中,主要采用的是“以学习者为中心”的教

学模式.为了满足学习者的个性化学习需求,从学习者和学习资源两个方面入手,提出学习者 KR 模型和学习资源 KR 模型.该模型通过分析学生在平台中的学习行为,根据学习者学习水平、学习风格和学习偏好等属性特征综合考虑个性化资源推荐模型.模型考虑学习者在实际学习应用中的学习特点与学习偏好而推出最优的学习资源.综合了学习者的基本信息、学习者的属性特点、学习者的兴趣爱好和用户的评分,使学习资源的推荐更加的系统全面.其中,考虑学习者的学习风格信息因素的推荐,能使学习者获得适合当时学习情境的推荐资源;考虑学习者属性特点因素的推荐,能使系统推荐与学习者的学习层次水平、学历背景等特点相匹配的学习资源;考虑学习者偏好因素的推荐,能帮助促进学习者对推荐资源的学习兴趣,提高学习行为的持续性;考虑学习目标因素的推荐,更容易使学习者高效地、有针对性地学习;考虑学习评价因素的推荐,能够维持推荐资源的时效性,使学习者接触到最新的、最精准的学习资源.个性化学习资源推荐模型综合以上属性特征来系统地进行个性化的推荐,节省了学习者选择资源的时间和精力,并且提高学习资源的使用率和学习者的学习成绩.

2.2 实验数据集

实验数据来源是“移动自主学堂”^[12]提供的 meclass 数据集,该数据集包含 30 000 条评分记录,产生于 200 名学习者在 2 个月内对 650 个学习资源的评价,评分数值的范围 1~5(1-非常不相关,5-非常相关),0 表示学习者未做出任何评价.实验将收集的数据按 1:4 的比例分为两部分,一部分作为训练集,另一部分作为测试集,构造推荐模型.具体的数据如表 1 所示.

表 1 数据集

Tab.1 Dataset

学习者数量	学习资源数量	评分数量
200	650	30 000

2.3 算法评价标准

本文使用平均绝对误差(MAE)这一评价指标对提出算法的准确度进行评估,MAE 通过准确的预测用户的评分来评估推荐系统的性能^[13].MAE 通过对比用户对项目预测评分与用户对项目实际评分的偏差,通过本文提出的算法计算不同邻居社区的 MAE 的计算公式为:

$$E' = \frac{1}{m} \sum_{i=1}^m |P_i - R_i|, \quad (6)$$

其中: E' 为平均绝对误差, P_i 为用户对项目的预测评分; R_i 为用户对项目的实际评分; m 为预测评分的次数.从(6)式中可以看出 MAE 的值越小,则算法的准确性越高.

2.4 实验结果与分析

将 KR-CF 算法与传统的推荐算法 Cosine-CF, Pearson-CF 进行比较,比较结果如图 5 所示.

从图 5 中可知,KR-CF 算法的 MAE 值在不同的邻居数目下都明显低于传统的算法,随着邻居数目的增加、数据的稀疏性降低,即算法的推荐精度随着 MAE 值的降低而增加,最后趋于稳定.KR-CF 算法在预测评分时,综合考虑了学习者的学习水平、学习风格和学习偏好改进用户相似度,这样突出了学习者之间的语义关系对学习资源推荐的重要性,提高学习者最近邻的识别精度,使推荐更加合理.实验表明 KR-CF 算法比传统的 CF 算法的推荐性能高.

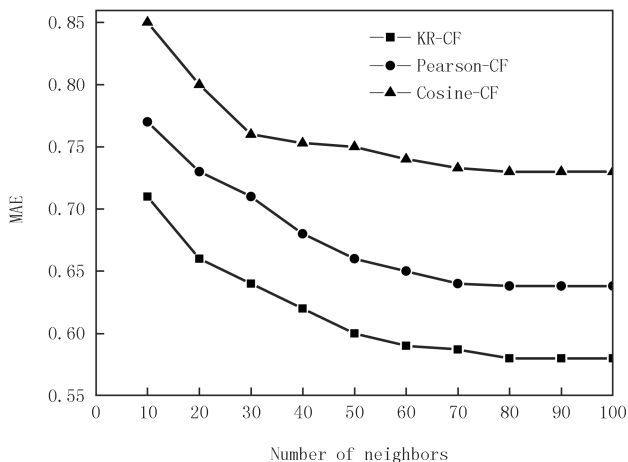


图 5 3 种算法在不同邻居数目下的 MAE 值对比图

Fig.5 The MAE values of three algorithms in the case of different neighbor numbers

3 结束语

本文研究的重点和创新点在于提出了一种基于 KR 和 CF 的学习资源精准推荐模型,利用 KR 构建学习者和学习资源模型,并且融合了学习者的属性特征,更好地实现了学习资源的个性化和精准推荐.实验结果表明,本文提出的 KR-CF 的推荐算法较之 CF 算法更具优势.此外,提出的推荐模型在缺乏足够学习资源评分的情况下,将用户-资源-本体 KR 三元关系进行分析,缓解数据稀疏问题和初始阶段的冷启动问题.后续工作中,将对学习者偏好的动态变化和学习者在推荐过程中的学习行为做进一步研究.

参 考 文 献

- [1] Erdt M, Fernandez A, Rensing C. Evaluating Recommender Systems for Technology Enhanced Learning: A Quantitative Survey[J]. IEEE Transactions on Learning Technologies, 2015, 8(4): 326-344.
- [2] 叶海智, 刘骏飞. 基于用户评论的协同过滤推荐算法[J]. 河南师范大学学报(自然科学版), 2017, 45(1): 79-84.
- [3] Schafer J B, Dan F, Herlocker J, et al. Collaborative Filtering Recommender Systems[M]. Berlin: Springer, 2007: 46-45.
- [4] Najafabadi M K, Mahrin M N. A systematic literature review on the state of research and practice of collaborative filtering technique and implicit feedback[J]. Artificial Intelligence Review, 2015, 45(2): 1-35.
- [5] Le H S. HU-FCF++: A novel hybrid method for the new user cold-start problem in recommender systems[J]. Engineering Applications of Artificial Intelligence, 2015, 41: 207-222.
- [6] Barjasteh I, Forsati R, Ross D, et al. Cold-Start Recommendation with Provable Guarantees: A Decoupled Approach[J]. IEEE Transactions on Knowledge & Data Engineering, 2016, 28(6): 1462-1474.
- [7] Konstan J A, Walker J D, Brooks D C, et al. Teaching recommender systems at large scale: evaluation and lessons learned from a hybrid MOOC[J]. ACM Transactions on Computer-Human Interaction(TOCHI), 2015, 22(2): 10.
- [8] 王晓东. 基于 Ontology 知识库系统建模与应用研究[D]. 上海: 华东师范大学, 2003.
- [9] Ricci F. Introduction to Recommender Systems Handbook[M]. Berlin: Springer, 2011: 1-4.
- [10] 贺桂和. 基于用户偏好挖掘的电子商务协同过滤推荐算法研究[J]. 情报科学, 2013(12): 38-42.
- [11] Tarus J, Niu Z, Khadidja B. E-Learning Recommender System Based on Collaborative Filtering and Ontology[J]. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 2017, 11(2): 225-230.
- [12] 王瑞, 李永波, 王晓东, 等. 移动自主学习堂及其应用[J]. 河南师范大学学报(自然科学版), 2014, 42(6): 162-166.
- [13] Bobadilla J. Recommender systems survey[J]. Knowledge-Based Systems, 2013, 46(1): 109-132.

Accurate recommendation model and application of learning resources

Wang Xiaodong, Shi Junya, Li Chun, Wu Huiping

(College of Computer and Information Engineering, Henan Normal University, Xixiang 453007, China)

Abstract: Due to the existence of cold start and sparsity problems, the performance of online learning resource recommendation based on attribute characteristics is limited. This paper is based on knowledge representation and collaborative filtering. We integrate learners' learning levels and learning styles into the recommendation process and perform personalized recommendation based on collaborative filtering. Through the works of this paper, an accurate learning resource recommendation model is proposed, a knowledge representation model for learners and learning resources is built. Experiments show that the recommendation algorithm combined with knowledge representation and collaborative filtering outperforms the traditional CF algorithm in terms of personalized recommendation and recommendation accuracy.

Keywords: knowledge representation; learning resource recommendation; collaborative filtering; online learning

[责任编辑 陈留院]