



## 融入学习者模型在线学习资源协同过滤推荐方法

刘芳, 田枫, 李欣, 林琳

引用本文:

刘芳, 田枫, 李欣, 等. 融入学习者模型在线学习资源协同过滤推荐方法[J]. 智能系统学报, 2021, 16(6): 1117–1125.

LIU Fang, TIAN Feng, LI Xin, et al. A collaborative filtering recommendation method for online learning resources incorporating the learner model[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(6): 1117–1125.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202009005>

## 您可能感兴趣的其他文章

### 融合用户特征优化聚类的协同过滤算法

Collaborative filtering algorithm combining user features and preferences in optimized clustering

智能系统学报. 2020, 15(6): 1091–1096 <https://dx.doi.org/10.11992/tis.201710024>

### 基于知识图谱和用户长短期偏好的个性化景点推荐

Personalized attraction recommendation based on the knowledge graph and users' long-term and short-term preferences

智能系统学报. 2020, 15(5): 990–997 <https://dx.doi.org/10.11992/tis.201904064>

### 融合协同过滤与用户偏好的旅游组推荐方法

A tourist group recommendation method combining collaborative filtering and user preferences

智能系统学报. 2018, 13(6): 999–1005 <https://dx.doi.org/10.11992/tis.201802011>

### 面对智能导诊的个性化推荐算法

A personalized recommendation algorithm for intelligent guidance

智能系统学报. 2018, 13(3): 352–358 <https://dx.doi.org/10.11992/tis.201711036>

### 个性化信息推荐方法研究

Research on the recommendation method of personalized information

智能系统学报. 2018, 13(2): 189–195 <https://dx.doi.org/10.11992/tis.201701002>

### 在线学习的大规模网络流量分类研究

Large-scale network traffic classification based on online learning

智能系统学报. 2016, 11(3): 318–327 <https://dx.doi.org/10.3969/j.issn.1673-4785.201603033>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202009005

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.tp.20210920.0227.002.html>

# 融入学习者模型在线学习资源协同过滤推荐方法

刘芳<sup>1</sup>, 田枫<sup>1</sup>, 李欣<sup>2</sup>, 林琳<sup>1</sup>

(1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318; 2. 讷河市第一中学, 黑龙江 讷河 161300)

**摘要:** 在线教育存在“信息迷航”问题, 而传统的信息推荐方法往往忽视教育的主体——学习者的特征。本文依据教育教学理论, 根据在线教育平台中的学习者相关数据, 研究构建了适用于在线学习资源个性化推荐的学习者模型。以协同过滤推荐方法为切入点, 融合学习者模型中的静态特征和动态特征对协同过滤方法进行改进, 建立融入学习者模型的在线学习资源协同过滤推荐方法。以 2020 年 3~7 月时间段的东北石油大学“C 程序设计”课程学生的真实学习数据和行为数据为数据集, 对本文提出的方法进行验证和对比, 最后证明本文提出的方法在性能上均优于对比方法。

**关键词:** 学习者模型; 在线学习资源; 协同过滤; 个性化学习; 学习资源推荐; 学习风格特征; 认知水平特征; 兴趣偏好特征

**中图分类号:** TP391; G434    **文献标志码:** A    **文章编号:** 1673-4785(2021)06-1117-09

**中文引用格式:** 刘芳, 田枫, 李欣, 等. 融入学习者模型在线学习资源协同过滤推荐方法 [J]. 智能系统学报, 2021, 16(6): 1117-1125.

**英文引用格式:** LIU Fang, TIAN Feng, LI Xin, et al. A collaborative filtering recommendation method for online learning resources incorporating the learner model[J]. CAAI transactions on intelligent systems, 2021, 16(6): 1117-1125.

## A collaborative filtering recommendation method for online learning resources incorporating the learner model

LIU Fang<sup>1</sup>, TIAN Feng<sup>1</sup>, LI Xin<sup>2</sup>, LIN Lin<sup>1</sup>

(1. School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China; 2. Nehe No. 1 Middle School, Nehe 161300, China)

**Abstract:** Online education exhibits the problem of “information loss”. At the same time, traditional information recommendation methods often ignore the characteristics of learners, i.e., the main body of education. Based on the theory of education and teaching as well as the relevant data of learners on the online education platform, this paper constructs a learner model suitable for personalized recommendations for online learning resources. Based on the collaborative filtering recommendation method, the static and dynamic features of the learner model are integrated, with the aim to improve the collaborative filtering method, thereby establishing a collaborative filtering recommendation method for online learning resources incorporating the learner model. The real learning and behavior records of students taking the C programming course in the Northeast Petroleum University starting from March 2020 to July 2020 were selected as the dataset to conduct experiments and evaluations on the proposed research method. The comparative test shows that the performance of the proposed method is better than that of the comparative method.

**Keywords:** learner models; online learning resources; collaborative filtering; personalized learning; learning resources recommendation; learning style characteristics; cognitive level characteristics; interest preference characteristics

收稿日期: 2020-09-07. 网络出版日期: 2021-09-23.

**基金项目:** 国家自然科学基金项目 (61502094); 黑龙江省教育科学规划重点课题 (GJB1421113); 黑龙江省高等教育教学改革研究项目 (SJGY20190098); 东北石油大学引导性创新基金项目 (2020YDL-11); 东北石油大学优秀中青年科研创新团队项目 (KYCXTD201903); 东北石油大学研究生教育创新工程项目 (JYCX\_11\_2020).

通信作者: 刘芳. E-mail: [lfliufang1983@126.com](mailto:lfliufang1983@126.com).

近年来, 随着互联网技术高速发展, 在线教育平台的使用越来越被学习者所接受, 尤其新型冠状病毒爆发之后, 线上学习是在不能正常进行线下学习的情况下优选的学习模式。数据表明, 截至 2020 年 12 月, 我国在线教育用户规模达 3.42 亿, 占网民整体的 34.6%。该项数据在 2020 年

3月时达到高峰,为4.23亿,占比数据为46.8%<sup>[1]</sup>。面对巨大的在线学习需求,在线教育机构提供免费在线课程,共享在线教学资源,在线教育行业呈现爆发式增长趋势。然而随着在线学习平台的广泛使用,在线学习资源的数量也急剧增长,在数量庞大的在线学习资源中,学习者很难快速定位自己需要的资源,导致“信息迷航”。

信息推荐是解决用户从海量对象中迅速有效地筛选出符合自己偏好特征的对象的方法<sup>[2]</sup>。目前,信息推荐技术被广泛应用于诸多领域。将信息推荐技术应用到在线教育,实现在线教育过程中学习资源的个性化推荐是解决“信息迷航”问题的一个有效途径。协同过滤算法是目前信息推荐技术的主流算法,该算法的相关研究工作大多集中在相似度计算和用户对资源评分的优化上<sup>[3]</sup>。在相似度计算方面,Wang等<sup>[4]</sup>提出了基于Kullback-Leibler散度的项目相似度计算方法来改进相似度计算,提高推荐准确率;Jiang等<sup>[5]</sup>提出了基于Quasi-norm的用户相似度计算方法,来提高推荐精度;Mu等<sup>[6]</sup>提出了一种改进的Pearson相关系数的方法改进相似度计算从而改善推荐结果。他们都是通过提高用户相似性的度量来提高信息推荐的质量。但是协同过滤在应用到学习资源个性化推荐方面,仅靠提高相似度很难提高推荐质量,将学习者这一学习资源推荐的主体与信息资源推荐的用户相对比,学习者的学习风格、偏好、背景、认知水平等个性化特征更为复杂,学习者的复杂特征对在线学习资源推荐质量影响较大,国内外诸多研究人员对学习者的展开研究。Wang等<sup>[7]</sup>设计了自适应的推荐模型,该模型研究了学习者的兴趣偏好特征,并以学习者和学习资源的相关性为依据,挖掘基于本体的学习资源的语义关系;Segal等<sup>[8]</sup>将社会选择特征融入传统的协同过滤算法,提出了Edu Rank算法,该算法可适应个性化教学;Zhang等<sup>[9]</sup>提出基于本体的语义关系模型,并将协同过滤算法与改模型融合;Aleksandra等<sup>[10]</sup>提出采用聚类方法的学习者学习风格模型,并改进矩阵分解方法进行推荐。

学习者是在线学习的参与者,学习资源推荐的个性化程度,直接决定着学习者的学习效果,因此在推荐时不仅要考虑相似度的计算方法的改进,更要研究学习者的个性化特征。个性化的学习者模型的构建是学者们研究的重点<sup>[11-13]</sup>,但大多研究都较集中在学习者行为数据分析、学习资源特征分析、语义特征分析等单一方面,缺乏对学习者的整体特征的综合分析和学习者多维度特征

的个性化研究。本文以在线学习平台中的学习者真实数据为依托,以学习者为中心,提出融合学习者多维度模型的在线学习资源协同过滤推荐方法,优化在线教学学习效果,进一步提高在线教育的个性化程度。

## 1 多维度学习者模型构建

学习者是在线学习的主体,具有静态和动态的个性化特征,学习者模型用于描述学习者特征,该模型的构建是提高在线学习资源推荐性能,优化推荐精度,实现个性化推荐的核心。在构建学习者模型时首选要确定学习者的个性化特征,本文依据CELT-11学习者信息模型规范<sup>[14]</sup>,以学习风格理论<sup>[15]</sup>和教育目标分类理论<sup>[16]</sup>为指导,将学习者特征分为静态特征和动态特征两个部分,其中,静态特征包括学习者基本信息、学习风格和静态兴趣偏好,动态特征包括认知水平和动态兴趣偏好。静态特征是学习者的初始特征,在整个学习过程中不能随着学习深入而发生变化,不能表示学习者的个性化特征程度,但作为基本特征可以解决初始用户在推荐过程中存在的冷启动问题。动态特征是指随着学习行为的发生,学习者的一些隐含特征逐渐显现,如学习认知状态和对某些资源的学习评价等都会随着时间发生变化,因此动态特征是构建学习者模型的重点。学习者的静态和动态数据通过采集层进行数据采集,在数据层完成信息归类,数据分析层将归类好的信息进行进一步数据挖掘,为表示层的学习风格、认知水平、静态和动态兴趣偏好特征提供数据基础。学习者模型构建过程如图1所示。

### 1.1 数据采集

学习者的初始静态数据和动态行为数据是构建学习者模型的数据基础。通过学习者注册时所填写的问卷、量表等获取学习者模型的基本信息、学习风格以及静态的兴趣偏好等特征信息;通过调取学习平台的章节知识点测评数据和各类学习者行为数据获取学习者的认知水平和动态兴趣偏好特征。通过数据采集层实现基础数据的获取与收集,为下一步的归类分析挖掘以及特征表示做基础。

### 1.2 学习风格的特征表示

学习风格的概念是1954年由美国赛伦首次提出的,它是反应学习者生理、心理等需要的概念,学习风格的研究为学习者模型的个性化要求提供了依据<sup>[17]</sup>。以Felder-Silverman风格模型为基

础,以所罗门学习风格量表(index of learning style questionnaire, ILSQ)<sup>[18]</sup>为手段,将学习者的学习风格从感知、输入、处理和理解4个维度进行量

化,在数据采集层每一个学习者都要填写学习风格调查量表,将获取的ILSQ量表结果送入数据层和分析层,在表示层构建学习风格特征。

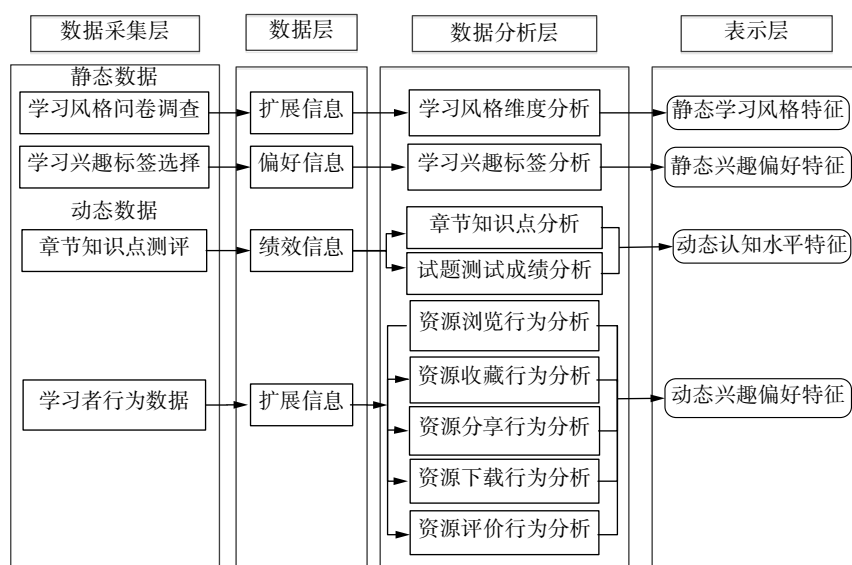


图1 学习者模型构建过程

Fig. 1 Building process of learner model

学习风格特征量化的具体流程如下:

1)以四元组 $\langle L_i, V_i \rangle (i=\{1,2,3,4\})$ 的形式表示学习风格量化结果, $L_i$ 表示ILSQ的4个维度; $V_i$ 表示在 $L_i$ 维度下的学习风格倾向的量化数值,其形式化定义为

$$L_S = \{(\langle L_1, V_1 \rangle, \langle L_2, V_2 \rangle, \langle L_3, V_3 \rangle, \langle L_4, V_4 \rangle) | V_i \in [-1, 1]\}$$

2)学习者填写ILSQ量表时,共44道题,每题包含两个选项A和B,答题结果的值定义为 $P_j$ ,其中 $j$ 表示题号;

3)根据 $P_j$ 的结果筛选处理,分类累加,最后的累加结果用 $a$ 和 $b$ 表示;

4)对 $a$ 和 $b$ 值的大小进行判断,如果 $a > b$ ,则 $V_i = (a-b)a$ ;如果 $a < b$ ,则 $V_i = (b-a)a$ ;

5)学习风格特征的测试结果四元组 $L_S$ 则为学习者的学习风格特征量化结果。

### 1.3 兴趣偏好的特征表示

学习者的兴趣偏好特征分为静态兴趣偏好特征和动态兴趣偏好特征。将数据集中的学习资源通过部分人工标注,再将剩余的资源通过相似度计算、最近邻排序等方法,实现自动标注,最后通过人工查询相关反馈机制进行校核,保证学习资源特征表示的准确性。以学习资源特征构成的规范化标签集合为选项,构建静态兴趣偏好问卷,在数据采集层每一个学习者都要填写静态兴趣偏好问卷,将获取的结果送入数据层和分析层,在表示层构建静态兴趣偏好特征。

学习者的学习过程是一个动态变化的过程,学习过程中各种操作都会产生相应的行为信息,该信息反映了当前学习者的兴趣偏好,本文将随着时间变化产生的兴趣偏好称为动态兴趣偏好<sup>[19]</sup>,具体的量化过程如下:

#### 1)学习者行为分类及权重计算

学习者行为主要分为5类,即浏览行为、收藏行为、分享行为、下载行为以及评价行为,不同的行为所代表的学习者的隐含偏好程度是不同的<sup>[20]</sup>,这里引入权重来表示不同的学习行为的贡献程度。权重的确定方法有很多种,专家评测或者经验主义权重具有一定的主观性,本文采用熵权法确定权重<sup>[21]</sup>学习者行为分类、权重分布及本文最后采用的权重数值,如表1所示,其中 $w_i$ 表示第 $i$ 个行为所占的权重分配。

表1 学习者行为分类及权重分布

Table 1 Learner behavior classification and weight distribution

行为分类	权重分配	数值
浏览行为	$w_1$	0.176
收藏行为	$w_2$	0.193
分享行为	$w_3$	0.195
下载行为	$w_4$	0.212
评价行为	$w_5$	0.224

#### 2)学习者-学习资源评分矩阵构建

依据学习行为及其所占的权重分配构建学习

者-资源评分矩阵  $\mathbf{P}_{m \times n}$ , 该矩阵可作为学习者对学习资源的评价依据,  $\mathbf{P}_{m \times n}$  为

$$\mathbf{P}_{m \times n} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix}$$

$\mathbf{P}_{m \times n}$  矩阵中的每个值都表示学习者  $u_m$  对资源  $i_n$  的行为权重, 如果  $s_{ui}=0$ , 那么说明学习者  $u_j$  并未对  $i_k$  产生任何行为, 如果矩阵元素全为 0, 则说明学习者  $u_j$  并没有开始学习。

### 3) 学习资源-学习标签矩阵构建

为了建立学习者与学习资源标签的直接关联关系, 首先构建学习资源标签矩阵来表征学习资源的特征:

$$\mathbf{Q}_{n \times l} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1l} \\ r_{21} & r_{22} & \cdots & r_{2l} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nl} \end{bmatrix}$$

$\mathbf{Q}_{n \times l}$  矩阵中的元素  $r_{jk}$  表示资源  $i_j$  是否拥有标签  $t_k$ ,  $r_{jk}=1$  表示标签  $t_k$  标注了资源  $i_j$ ;  $r_{jk}=0$  表示未被标注, 因此矩阵  $\mathbf{Q}_{n \times l}$  是一个由 0 和 1 构成的矩阵。

依据学习者-学习资源评分矩阵  $\mathbf{P}_{m \times n}$  和学习者-学习资源标签矩阵  $\mathbf{Q}_{n \times l}$  构建学习者-标签矩阵  $\mathbf{T}_{m \times l}$ :

$$\mathbf{T}_{m \times l} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1l} \\ g_{21} & g_{22} & \cdots & g_{2l} \\ \vdots & \vdots & & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{ml} \end{bmatrix}$$

式中:  $g_{jk} = \sum_{c=1}^n s_{jc} r_{ck}$ ,  $\mathbf{T}_{m \times l}$  矩阵中的元素  $g_{jk}$  表示在学习资源标签  $t_k$  上不断累积的学习者  $u_j$  的行为权重,  $\mathbf{T}_{m \times l}$  用来表示学习者产生行为之后的动态兴趣偏好矩阵。

### 4) 学习者动态兴趣偏好行为特征表示

学习者的不同行为操作在学习资源上累积可以用动态兴趣偏好矩阵  $\mathbf{T}_{m \times l}$  来表示, 学习者对资源的偏好程度体现着学习者之间的差异, 这一差异在表征学习者的行为特征属性时是一个递增函数, 其计算公式为

$$F_{u_k}^{\text{op}} = \exp\left(\frac{g_{uk}}{v_{u_k}} - \lambda\right), \quad 1 \leq k \leq l$$

式中:  $g_{uk} (1 \leq k \leq l)$  是学习者兴趣偏好在资源上的累加值, 是学习者  $u$  在关联资源的标签  $t_k$  上不断累加的行为之和;  $v$  是学习者平均兴趣偏好值;  $\lambda$  是学习者行为累加和的最小值, 用来消除不同学习者间兴趣偏好偏差。

### 5) 时间因素调整动态兴趣偏好特征的偏移

学习者的兴趣偏好特征会随着学习的深入产生偏移, 动态兴趣偏好特征的调整包括各类行为的特征表示和时间因素, 行为特征采用上述的递增函数, 而时间因素表征了学习者基于时间参数的特征, 采用时间衰减函数来进行计算<sup>[22]</sup>。学习者的动态兴趣偏好特征时间因素的计算公式为

$$F_{u_k}^{\text{time}} = \theta + (1 - \theta) \exp[-(t_{\text{now}} - t_{u_k})], \quad 1 \leq k \leq l$$

式中:  $t_{\text{now}}$  为当前时间;  $t_{u_k}$  表示学习者  $u$  被标签  $t_k$  标注的时间集合里的最近值; 超参数  $\theta \in [0, 1]$  可以影响时间因素对动态兴趣特征的计算, 二者表现为负相关。

将行为特征和时间权重特征进行综合, 得到学习者的动态兴趣偏好特征, 即

$$F_{u_k} = F_{u_k}^{\text{op}} F_{u_k}^{\text{time}}, \quad 1 \leq k \leq l$$

## 1.4 认知水平的特征表示

学习者的认知水平特征描述的是学习者在学习某个知识点之后, 对该知识点对应的试题进行测试, 获取的对该知识点的掌握程度。

以“布鲁姆教育目标分类理论”为依据, 知识点对应的学习资源的学习目标被分为 6 个等级 (如图 2), 这 6 个等级代表着不同学习者对核心知识点掌握程度, 即认知水平。在学习过程中, 采集层的章节知识测试数据代表了绩效信息, 通过分析章节知识点和试题测试成绩, 获取认识水平特征, 由于该项指标分为 6 个等级, 不同的学习者会有不同的整体认知水平, 同一个学习者不同时期对于不同的知识点也会有不同的水平状态, 因此认知水平体现着学习者的个性化特征。学习者的认知水平特征表达式为

$$K_L = \{(k_1, l_1), (k_2, l_2), \dots, (k_i, l_i), \dots, (k_n, l_n)\}$$

式中:  $k_i$  表示第  $i$  个知识点;  $l_i$  表示对第  $i$  个知识点的掌握程度, 即认知水平,  $n$  是学已学过的知识点数量。

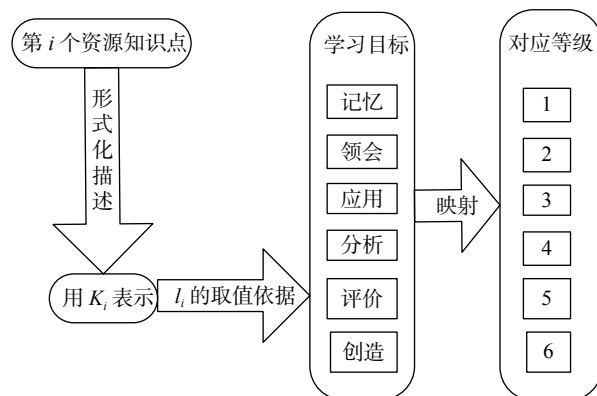


图 2 学习资源知识点掌握程度的表示方法

Fig. 2 Representation method of learning resources knowledge points master degree

## 2 融合学习者模型的协同过滤改进

协同过滤是信息推荐技术中经典的推荐方法<sup>[23-26]</sup>,本文采用协同过滤作为在线学习资源推荐的基本算法,实现基本的推荐,在基本推荐的基础上融合学习者模型的多维度特征,进行精准推荐。

### 2.1 基于协同过滤的在线学习资源推荐

经典协同过滤技术的推荐过程分成3个部分:1)收集学习者对学习资源的评分,构建学习者-学习资源评分矩阵;2)学习者-学习资源评分矩阵隐含着学习者对学习资源的兴趣偏好,因此可以通过相似度计算求出与被推荐的学习者具有相似兴趣偏好的学习者集合,构成 $K$ 近邻学习者集合;3)计算 $K$ 近邻学习者集合中每个学习者对学习资源的评分,产生被推荐学习者对学习资源的预测评分,按照评分进行排序,产生学习资源推荐集合,在该集合中筛选出没有被被推荐学习者学习过的Top-N个资源,这Top-N个资源就是最后的推荐结果。

### 2.2 融合学习者静态特征的协同过滤推荐改进

冷启动问题是推荐系统的共性问题,它指的是在面对刚刚进入推荐系统的新学习者时,由于行为数据较少,因此系统无法获取初始学习者的隐含偏好信息,本文构建的学习者模型的静态特征可以较好地解决冷启动问题。在数据采集层通过问卷、量表等方式获取学习者的学习风格以及静态的兴趣偏好等特征信息,通过加权融合计算学习者静态综合特征相似度,按相似度排序构成 $K$ 近邻学习者集合,根据 $K$ 近邻学习者的学习资源列表完成初始学习者的推荐。

#### 1) 静态兴趣偏好特征相似度计算

学习者 $u_n$ 的初始兴趣偏好标签个数为 $q$ ,学习者 $u_m$ 的初始兴趣偏好标签个数为 $p$ ,学习者 $u_n$ 和 $u_m$ 之间含相同标签数目为 $k$ ,相似度计算公式为

$$\text{sim}(u_m, u_n)_{\text{interest}} = \frac{k}{q + p - k}$$

#### 2) 学习风格特征相似度计算

学习风格包括4个维度,不同维度的分数值表示为 $S = \{s_1, s_2, s_3, s_4\}$ ,将 $S$ 值标准化之后,利用欧几里得距离公式计算学习风格距离,其计算公式为

$$d(S_m, S_n)_{\text{style}} = \sqrt{\sum_{k=1}^4 (S_m^k - S_n^k)^2}$$

$u_m$ 和 $u_n$ 的学习风格相似度计算公式为

$$\text{sim}(u_m, u_n)_{\text{style}} = \frac{1}{1 + d(S_m, S_n)}$$

#### 3) 静态综合特征相似度计算

将学习风格特征相似度和静态兴趣偏好特征相似度加权融合形成学习者静态综合特征相似度,超参数 $\alpha \in [0, 1]$ 可以影响学习风格和静态兴趣偏好的影响程度, $\alpha$ 的具体数值通过实验统计数据的经验获得,静态综合相似度计算公式为

$$\text{sim}(u_m, u_n)_{\text{static}} = \alpha \text{sim}(u_m, u_n)_{\text{interest}} + (1 - \alpha) \text{sim}(u_m, u_n)_{\text{style}}$$

### 2.3 融合学习者动态特征的协同过滤推荐改进

为丰富推荐结果的多样性,提高推荐结果的准确性,在融合静态特征推荐的基础上,引入学习资源-学习标签矩阵,结合基础协同过滤算法中用到的学习者-学习资源评分矩阵,构建学习者-学习标签矩阵,更新动态兴趣偏好特征。通过分析章节知识点和试题测试成绩,获取认识水平特征,通过计算融合动态兴趣偏好特征和认知水平特征的相似度,得到融合学习者动态特征的 $K$ 近邻学习者集合,由 $K$ 近邻集合得到推荐资源列表的过程与基于协同过滤的方法一致。

#### 1) 融合行为特征和时间权重特征的兴趣偏好相似度计算

以学习者对学习资源的评分来描述学习者的兴趣偏好特征是不全面不准确的,在构建学习者模型动态兴趣偏好特征时,融入各类学习行为,再对这些行为加权计算,构建融合行为加权的学习者-学习资源评分矩阵,从而构建学习者-学习标签矩阵,再融合时间因素,表示学习者动态兴趣偏好特征随着学习行为的持续和时间的深入产生的偏移问题。

融合行为特征和时间特征的兴趣偏好特征向量表示为 $F_u = \{F_{ut_1}, F_{ut_2}, \dots, F_{ut_j}\}$ ,学习者 $u_m$ 和 $u_n$ 之间的相似性可以通过皮尔逊相关系数进行计算,即

$$\text{sim}(u_m, u_n)_{\text{interest\_update}} = \frac{\sum_{t_i \in T_{mn}} (F_{m,t_i} - \bar{F}_m) \times (F_{n,t_i} - \bar{F}_n)}{\sqrt{\sum_{t_i \in T_{mn}} (F_{m,t_i} - \bar{F}_m)^2} \times \sqrt{\sum_{t_i \in T_{mn}} (F_{n,t_i} - \bar{F}_n)^2}}$$

式中: $T_{mn}$ 由学习者 $u_m$ 和 $u_n$ 的兴趣偏好标签的交集构成的标签集合; $F_{m,t_i}$ 和 $F_{n,t_i}$ 分别表示学习者 $u_m$ 和 $u_n$ 对标签 $t_i$ 的兴趣特征值; $\bar{F}_m$ 和 $\bar{F}_n$ 分别表示学习者 $u_m$ 和 $u_n$ 对集合中所有标签的平均兴趣值。计算学习者的兴趣特征值相似度,并按相似度的高低进行排序,构建出与目标学习者兴趣特征最为相似的近邻学习者集合 $U = \{u_1, u_2, \dots, u_m, \dots,$

$u_k$ }, 这里  $k$  为超参数, 具体数值通过经验或实验验证给出。

## 2) 融合认知水平特征的相似度计算

将学习者  $u_m$  在已学习过的知识点上的认知水平的集合表示为  $L(u_m) = \{(k_1 u_m, h_1 u_m), (k_2 u_m, h_2 u_m), \dots, (k_j u_m, h_j u_m)\}$ 。其中,  $k_j u_m$  表示学习者  $u_m$  掌握的第  $j$  个知识点;  $h_j u_m$  表示对第  $j$  个知识点的学习者  $u_m$  的掌握程度, 即认知水平。使用余弦相似度计算公式计算学习者的认知水平相似度, 构建出与目标学习者认知水平特征最为相似的近邻学习者集合, 学习者  $u_m$  和  $u_n$  的认知水平相似度计算公式为

$$\text{sim}(u_m, u_n)_{\text{level}} = \frac{u_m \cdot u_n}{|u_m| \cdot |u_n|} \frac{\sum_k h_{j, u_m} h_{j, u_n}}{\sqrt{\sum_k h_{j, u_m}^2} \sqrt{\sum_k h_{j, u_n}^2}}$$

## 3) 学习者动态综合特征相似度计算

将 2 种动态特征相似度加权, 计算出学习者的动态综合特征相似度。设置参数  $\beta$  调整融合比例, 具体计算公式为

$$\text{sim}(u_m, u_n)_{\text{static}} = \beta \text{sim}(u_m, u_n)_{\text{interest\_update}} + (1 - \beta) \text{sim}(u_m, u_n)_{\text{style\_update}}$$

# 3 实验结果与分析

## 3.1 实验数据集

以超星为在线学习资源推荐研究依托平台, 《C 程序设计》课程的学习资料为学习资源, 东北石油大学《C 程序设计》学习者 2020 年 3 月到 2020 年 7 月时间段的学生的真实学习行为为数据开展实验, 数据集中主要有 3 类文件: 1) 学习者特征中的静态数据文件, 包括学生基本信息、学习风格信息和学习兴趣信息; 2) 学习者特征中的动态数据文件, 包括知识点测评信息, 学习资源浏览、收藏、下载、评价和分享行为数据; 3) 带有标签信息的学习资源数据。原始数据经过数据预处理后, 共计数据 20 547 条, 学生 849 人, 学习资源 19 876 个, 其中静态数据 8 567 条, 动态数据 11 980 条。将数据集按照 4:1 的比例分为训练集和测试集进行模型的训练和测试。

## 3.2 评价标准

准确率、召回率、 $F_1$  是常用的用于评价推荐性能的评价标准。准确率是系统推荐给学习者的资源与学习者在测试集上感兴趣的资源的交集和系统推荐给学习者的资源的比率, 即

$$P = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|}$$

召回率是系统推荐给学习者的资源和学习者

在测试集上感兴趣的资源的交集与学习者在测试集上感兴趣的资源的比率, 即

$$R = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|}$$

式中:  $R(u)$  表示推荐产生的学习资源;  $T(u)$  表示学习者在测试集上关注的学习资源。随着学习资源推荐个数的增多, 准确率会有所下降, 但是召回率有所上升。对测试集所有学习者的上述度量求均值计算平均准确率 AP 和平均召回率 AR, 引入  $F_1$  值度量整体推荐方法的性能,  $F_1$  值越大, 表示该推荐方法性能越好,  $F_1$  值计算公式为

$$F_1 = 2 \times AP \times AR / (AP + AR)$$

## 3.3 实验结果与分析

本文构建的学习者模型包括静态特征和动态特征, 其中, 静态特征包括静态兴趣偏好特征和静态学习风格特征; 动态特征包括动态兴趣偏好特征和动态认知水平特征。从验证融合学习者特征算法有效性角度出发, 首先基于学习者对学习资源的评分矩阵, 实现了基于经典协同过滤的推荐。在此基础上融合各项动静态特征, 本文的经典协同过滤算法采用文献 [2] 中的通过构建“用户-项目”评分矩阵计算用户相似度, 匹配近邻用户进行推荐的方法。

### 1) 融合学习者静态特征实验分析

基于协同过滤方法, 融合学习者静态特征, 在学习资源推荐个数为 5、10、15、20、25、30、35 时的准确率、召回率和  $F_1$  值比较如图 3 所示。通过实验结果可知, 对比于只依靠学习者对学习资源评分矩阵的经典协同过滤推荐, 融合了学习者模型的单项的静态特征会提高整体推荐的性能, 但是单项实验并不会得出哪个特征对推荐结果的影响更大, 而且多项特征的融合效果也不会通过单项实验得到, 因此多项特征的融合参数如何选择也是要解决的问题。

实验中涉及学习资源推荐个数  $k$  的实验参数和静态兴趣偏好特征与静态学习风格特征相融合的权重系数  $\alpha$ 。实验过程中, 先定义推荐的学习资源个数为 5 个, 再对融合参数  $\alpha$  取值从 0.1~1 的推荐结果计算准确率和召回率的  $F_1$  值, 如图 4 所示。当  $\alpha$  取值为 0.6 时  $F_1$  值最高 (见图 4), 因此在后续测试结果中令  $\alpha=0.6$ , 再通过实验测试对比分析融合静态兴趣偏好特征和静态学习风格特征在不同学习资源推荐个数情况下的准确率和召回率。

### 2) 融合学习者动态特征实验分析

学习者动态特征融合包括学习者动态变化的认知水平和学习者对学习资源持续性的学习体现出的动态兴趣偏好特征。通过多次实验对比分析

准确率、召回率和  $F_1$  值,将时间参数  $\theta$ 、动态特征融合参数  $\beta$  调到最优值,最终在  $\theta=0.2$ ,  $\beta=0.7$  时,推荐结果最准确。

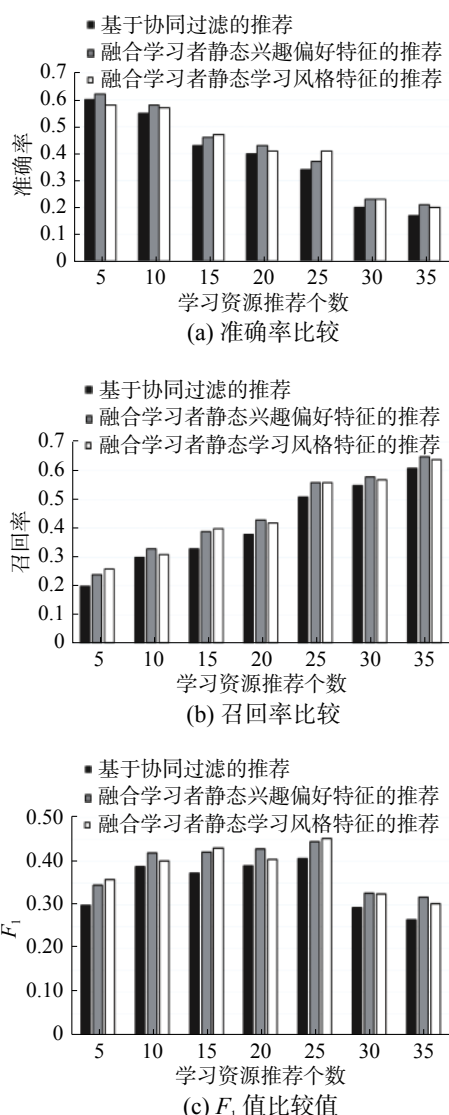


图3 融合学习者各项静态特征的推荐性能

Fig. 3 Recommended performance of integrating learners' various static characteristics

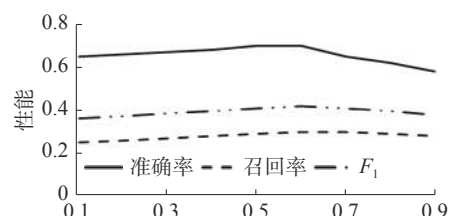


图4 静态特征融合参数选择比较

Fig. 4 Static feature fusion parameter selection comparison

### 3) 综合对比分析

综合对比分析基于协同过滤的推荐、融合学习者静态特征的推荐和融合学习者动态特征的推荐,从准确率、召回率和  $F_1$  值 3 个角度进行分析,实验数据如图 5~7 所示。

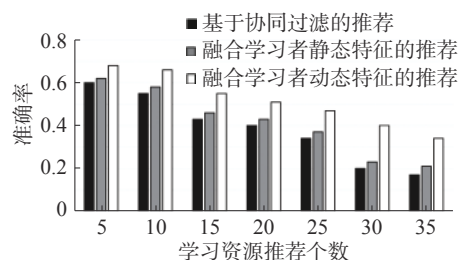


图5 不同推荐方法的准确率比较

Fig. 5 Accuracy comparison of different recommended methods

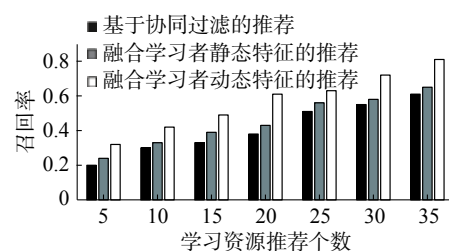


图6 不同推荐方法的召回率比较

Fig. 6 Recall rate comparison of different recommended methods

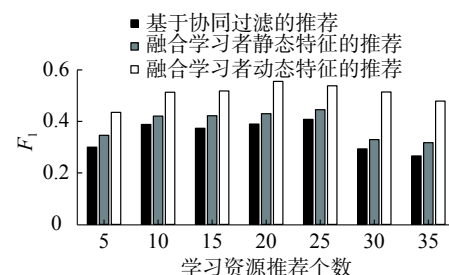


图7 不同推荐方法  $F_1$  值比较

Fig. 7  $F_1$  value comparison of different recommended methods

通过综合对比分析,得出结论:动态特征方面,融合行为特征和时间特征的动态兴趣偏好特征对最后推荐结果影响相对较大;静态特征方面,学习者的兴趣偏好特征要比学习者学习风格特征影响大。整体上,融合学习者动态特征的推荐性能优于融合学习者静态特征的推荐和基于协同过滤的推荐。

## 4 结束语

目前在线教育学习平台中存在海量学习资源,然而提供的服务个性化程度却不高,针对在线学习过程中的“信息迷航”问题,本文以在线学习平台中的学习资源数据和学习者数据为采集层的基础数据,通过数据分析和挖掘,构建了多维度的个性化学习者模型。该学习者模型包括学习者静态特征和学习者动态特征,静态特征包括学

习风格特征和静态兴趣偏好特征,动态特征包括认知水平特征和动态兴趣偏好特征。采用协同过滤作为在线学习资源的基础方法,将学习者静态特征和动态特征分别融入协同过滤的推荐方法中,通过实验得到的数据证实,本文构建的学习者模型,以及基于该模型构建的学习资源推荐方法提高了在线学习资源协同过滤推荐的性能。该方法对于满足个性化学习的需求、提高在线学习的学习效果具有重要意义。

## 参考文献:

- [1] 中国互联网络信息中心 (CNNIC). 第 45 次中国互联网络发展状况统计报告 [R]. 北京: 中国互联网络信息中心 (CNNIC), 2020: 4.
- [2] AL-SHAMRI M Y H. Power coefficient as a similarity measure for memory-based collaborative recommender systems[J]. Expert systems with applications, 2014, 41(13): 5680–5688.
- [3] NAJAFABADI M K, MOHAMED A, ONN C W. Impact of time and item influencer in collaborative filtering recommendations using graph-based model[J]. Information processing & management, 2019, 56(3): 526–540.
- [4] WANG Yong, DENG Jiangzhou, GAO J, et al. A hybrid user similarity model for collaborative filtering[J]. Information sciences, 2017, 418–419: 102–118.
- [5] JIANG Shan, FANG S C, AN Qi, et al. A sub-one quasi-norm-based similarity measure for collaborative filtering in recommender systems[J]. Information sciences, 2019, 487: 142–155.
- [6] MU Yi, XIAO Nianhao, TANG Ruichun, et al. An efficient similarity measure for collaborative filtering[J]. Procedia computer science, 2019, 147: 416–421.
- [7] WANG T I, TSAI K H, LEE M C, et al. Personalized learning objects recommendation based on the semantic-aware discovery and the learner preference pattern[J]. Educational technology & society, 2007, 10(3): 84–105.
- [8] SEGAL A, KATZIR Z, GAL Y, et al. EduRank: a collaborative filtering approach to personalization in E-learning[C]//Proceedings of the 7th International Conference on Educational Data Mining. London, UK, 2014: 68–74.
- [9] ZHANG Fuzheng, YUAN N J, LIAN Defu, et al. Collaborative knowledge base embedding for recommender systems[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 353–362.
- [10] KLAŠNJA-MILIĆEVIĆ A, IVANOVIĆ M, VESIN B, et al. Enhancing e-learning systems with personalized recommendation based on collaborative tagging techniques[J]. Applied intelligence, 2018, 48(6): 1519–1535.
- [11] BAKER R S J D, CORBETT A T, ALEVEN V. More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing[C]//Proceedings of the 9th International Conference on Intelligent Tutoring Systems. Montreal, Canada, 2008.
- [12] SALEHI M. Application of implicit and explicit attribute based collaborative filtering and BIDE for learning resource recommendation[J]. Data & knowledge engineering, 2013, 87: 130–145.
- [13] KURILOVAS E, SERIKOVIENE S, VUORIKARI R. Expert centred vs learner centred approach for evaluating quality and reusability of learning objects[J]. Computers in human behavior, 2014, 30: 526–534.
- [14] 现代远程教育技术标准化委员会. CELTS-11, 学习者模型规范 [S]. 现代远程教育技术标准化委员会, 2000: 11.
- [15] COSTA R D, SOUZA G F, VALENTIM R A M, et al. The theory of learning styles applied to distance learning[J]. Cognitive systems research, 2020, 64: 134–145.
- [16] ARIEVITCH I M. Reprint of: the vision of Developmental Teaching and Learning and Bloom's Taxonomy of educational objectives[J]. Learning, culture and social interaction, 2020, 27: 100473.
- [17] ZLATKOVIC D, DENIC N, PETROVIC M, et al. Analysis of adaptive e-learning systems with adjustment of Felder-Silverman model in a Moodle DLS[J]. Computer applications in engineering education, 2020, 28(4): 803–813.
- [18] DASCALU M I, BODEA C N, MOLDOVEANU A, et al. A recommender agent based on learning styles for better virtual collaborative learning experiences[J]. Computers in human behavior, 2015, 45: 243–253.
- [19] GONZÁLEZ G, LÓPEZ B, DE LA ROSA J L. A multi-agent smart user model for cross-domain recommender systems[C]//Proceedings of Beyond Personalization 2005: The Next Stage of Recommender Systems Research, International Conference on Intelligent User Interfaces IUI 2005. San Diego, USA, 2005.
- [20] 谢修娟, 陈永, 李香菊, 等. 融入信任的变权重相似度模型在线学习协同推荐算法 [J]. 小型微型计算机系统, 2018, 39(3): 525–528.
- XIE Xiujuan, CHEN Yong, LI Xiangju, et al. Collaborative recommendation algorithm of online learning based on trust-combined similarity model with variable

- weight[J]. Journal of Chinese computer systems, 2018, 39(3): 525–528.
- [21] 陈秀明, 刘业政. 多粒度犹豫模糊语言环境下未知权重的多属性群推荐方法 [J]. 控制与决策, 2016, 31(9): 1631–1637.
- CHEN Xiuming, LIU Yezheng. Method of group recommender systems with unknown attribute weights in a multi-granular hesitant fuzzy linguistic term environment[J]. Control and decision, 2016, 31(9): 1631–1637.
- [22] KOREN Y. Collaborative filtering with temporal dynamics[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, 2009: 89–97.
- [23] 孙歆, 王永固, 邱飞岳. 基于协同过滤技术的在线学习资源个性化推荐系统研究 [J]. 中国远程教育, 2012(8): 78–82.
- [24] 郑洁, 钱育蓉, 杨兴耀, 等. 基于信任和项目偏好的协同过滤算法 [J]. 计算机应用, 2016, 36(10): 2784–2788, 2798.
- ZHENG Jie, QIAN Yurong, YANG Xingyao, et al. Collaborative filtering algorithm based on trust and item preference[J]. Journal of computer applications, 2016, 36(10): 2784–2788, 2798.
- [25] 丁永刚, 张馨, 桑秋侠, 等. 融合学习者社交网络的协同过滤学习资源推荐 [J]. 现代教育技术, 2016, 26(2): 108–114.
- DING Yonggang, ZHANG Xin, SANG Qiuxia, et al. The collaborative filtering recommendation of learning resources combined with learners' social network[J]. Modern educational technology, 2016, 26(2): 108–114.
- [26] 刘忠宝, 宋文爱, 孔祥艳, 等. 云环境下学习者建模与学习资源推荐方法研究 [J]. 电化教育研究, 2017, 38(7): 58–63.
- LIU Zhongbao, SONG Wenai, KONG Xiangyan, et al. Research on learner modeling and learning resources recommendation in cloud environment[J]. E-education research, 2017, 38(7): 58–63.

### 作者简介:



刘芳, 副教授, 博士, 主要研究方向为智慧教育、多媒体与现代教育技术、计算机视觉、智能数据分析处理。获黑龙江省科技进步二等奖 1 项、大庆市科技进步二等奖 1 项。主持和参与国家自然科学基金项目、黑龙江省自然科学基金项目 6 项。发表学术论文 21 篇。



田枫, 教授, 博士, 主要研究方向为计算机视觉、智能数据分析处理。主持和参与国家自然科学基金项目、国家科技重大专项项目 8 项。获发明专利授权 16 项。发表学术论文 31 篇。



李欣, 助教, 主要研究方向为智慧教育。