

协作学习中基于协同过滤的学习资源推荐研究

叶树鑫¹, 何聚厚^{1,2}

(1. 陕西师范大学 计算机科学学院, 陕西 西安 710062;

2. 陕西师范大学 现代教学技术教育部重点实验室, 陕西 西安 710062)

摘要:符合学习者特征的学习资源对于提高协作学习效率具有重要的影响。但是传统的学习资源推荐,没有充分考虑学习者、学习资源的特征和高效的推荐算法。针对上述问题,提出了基于协同过滤的学习资源推荐算法,根据学习者学习特征、学习资源特征和学习者对学习资源历史评价信息,采用协同过滤推荐算法,实现学习资源推荐。首先,通过学习者特征和学习资源的评分,寻找相似学习者并计算学习资源预测评分,然后根据该评分值和学习资源与学习者匹配度推荐学习资源,从而为学习者推荐符合自己兴趣爱好最合适的学习资源。实验结果表明该算法在个性化学习资源推荐的准确性上优于传统算法。

关键词:协同过滤算法;学习资源推荐;协作学习

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2014)10-0063-04

doi:10.3969/j.issn.1673-629X.2014.10.015

Research on Learning Material Recommendation Based on Collaborative Filtering Algorithm in Cooperative Learning

YE Shu-xin¹, HE Ju-hou^{1,2}

(1. School of Computer Science, Shaanxi Normal University, Xi'an 710062, China;

2. Key Laboratory of Modern Teaching Technology of Ministry of Education, Shaanxi Normal University, Xi'an 710062, China)

Abstract: The appropriate learning material is very important to improve learners' learning efficiency in cooperative learning environment. However, traditional recommendation of learning material doesn't consider the learner features, learning materials features and recommendation algorithm enough. To solve the problems, propose a personalized learning materials recommendation algorithm based on collaborative filtering, which takes the learners' learning features, the features of learning materials and the historical assessment information of learners to learning material into consideration, using the collaborative recommendation algorithm to realize the learning material recommending. First, through the score of learner feature and learning material, search the similar learners and compute the learning material prediction score. Then, based on predicting ratings and relationship between learner and learning materials, produce the final recommending learning materials. Experimental results show that the proposed algorithm outperforms the other recommendation ones in recommending accuracy.

Key words: collaborative filtering algorithm; learning materials recommendation; collaborative learning

0 引言

协作学习中,为学习者推荐符合自己需求的学习资源,对于提高学习者的学习效率具有重要影响^[1-4]。在一些常见的学习管理系统中,如 Blackboard、WebCT、Moodle、Sakai 等,提供给学习者的都是千篇一律的学习资源,学习者是存在个性化差异的,具有不同

的学习风格、认知水平、学习动机、社会背景等^[1]。为此,L Sabine Graf 在 2007 年提出了采用 Felder-Silverman 学习风格度量表来实现个性化学习资源推荐^[2]。Chih-Ming Chen 在 2007 年提出了采用遗传算法,根据学习者认知能力提供相应难度级别的学习对象^[3]。Yao Jung Yang 等在 2009 提出了采用蚁群算法和 Kolb

收稿日期:2013-12-13

修回日期:2014-03-18

网络出版时间:2014-07-28

基金项目:国家自然科学基金资助项目(31271110);中央高校基本科研业务费专项资金项目(GK201002028, GK201101001)

作者简介:叶树鑫(1988-),男,硕士研究生,研究方向为计算机支持的协作学习;何聚厚,博士,副教授,通讯作者,研究方向为技术增强学习、知识工程与智能系统。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140728.1230.053.html>

学习风格模型的 AACs 学习资源推荐系统^[4],该系统依据学习者特征,为学习者提供个性化学习资源。学习资源推荐需要根据学习者的学习风格和兴趣进行推荐^[5-6]。鉴于上述学习资源推荐中对学习者属性的不完整性,文中结合学习者学习风格和认知能力,同时将学习资源进行分类,引入基于内存的协同过滤推荐算法,为协作学习中的学习者提供个性化学习资源推荐。

1 相关工作

协同过滤推荐是当今应用最成功、最广泛的推荐算法^[7-11]。基于协同过滤的个性化学习资源推荐,在基于内存的协同过滤算法基础上,根据学习者的学习风格和认知能力,为学习者推荐与自己认知能力相应难度级别和与自己学习风格相吻合的学习资源。

1.1 学习特征和学习资源特征

在学习资源推荐中,学习者学习特征分为学习风格和认知能力,学习资源特征包括资源类型和难度系数^[4]。

学习风格采用 Kolb 学习风格度量表^[12-13],学习风格维度如图 1 所示。

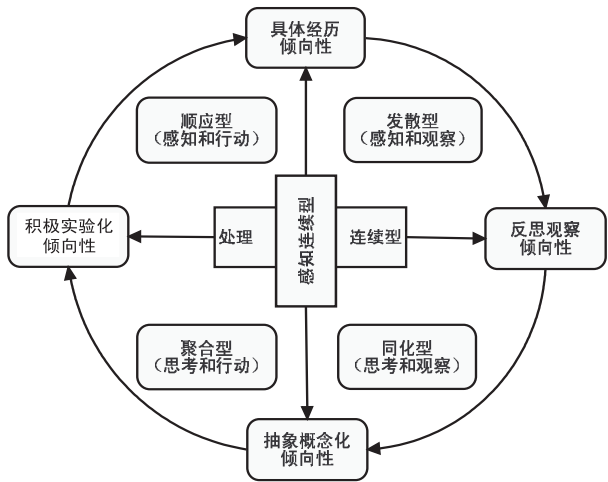


图 1 Kolb 学习风格维度

Kolb 学习风格维度包括具体经验(CE)、主动实践(AE)、抽象概念(AC)、反思观察(RQ)。根据四个维度,Kolb 学习风格分为四种类型:顺应型(Accommodating)、聚合型(Converging)、同化型(Assimilating)和发散型(Diverging)^[4]。

顺应型学习风格的学习者更喜欢学习资源类型是经验分享和实际练习,如例子、教程等^[4]。

聚合型学习风格的学习者更喜欢文本形式的学习资源,如 Word、pdf 等^[4]。

同化型学习风格的学习者更喜欢音频、视频等类型的学习资源^[4]。

发散型学习风格的学习者更喜欢表格、图片等类型的学习资源^[4]。

根据布鲁姆认知理论^[14],在个性化学习资源推荐中,学习者认知能力分为识记、领会、应用、分析(包括分析、综合和评价)四个层次。

学习资源列表根据学习者学习风格所喜欢的学习资源类别进行分类,分为 XML、文本、视频和图表^[4]。学习资源难度系数根据学习者的认知能力进行划分,学习资源难度系数分为初始资源、介绍资源、拓展资源和专业资源^[4]。

学习者特征与学习资源特征对应关系如表 1 所示。

表 1 学习者与学习资源类别

学习者		学习资源	
学习风格	认知能力	资源类型	难度系数
顺应型	识记	XML(Web, SCROM)	初始资源
聚合型	领会	文本(Word, PPT)	介绍资源
发散型	应用	图表(图片等)	拓展资源
同化型	分析	视频(音频等)	专业资源

1.2 学习资源推荐问题描述

定义 1:学习者特征集合 L 定义为:

$$L = \{lc, ls\}$$
 (1)

其中, $lc \in [0, 1]$ 表示学习者的认知能力,分为识记、领会、应用和分析级别四种类型; $ls \in [0, 1]$ 表示学习者的学习风格类型,包括顺应型、聚合型、同化型和发散型。

定义 2:学习者集合 S 定义为:

$$S = \{s_i \mid s_i = \langle lc_i, ls_i \rangle, i = 1, 2, \dots, N\}$$
 (2)

其中, lc_i 和 ls_i 分别是学习者 s_i 的特征值。

定义 3:学习资源特征集合 Materials 定义为:

$$\text{Materials} = \{md, mt\}$$
 (3)

其中, $md \in [0, 1]$ 表示学习资源的难度系数,学习资料按照难度分为四种,初始资料、介绍资料、拓展资料和专业资料^[4]; $mt \in [0, 1]$ 表示学习资源的类型,分为 XML、文本类、图表类和音频视频类四种^[4]。

定义 4:学习资源集合 M 定义为:

$$M = \{m_i \mid m_i = \langle md_i, mt_i \rangle, i = 1, 2, \dots, m\}$$
 (4)

其中, md_i 和 mt_i 分别代表学习资源 m_i 的特征值。

定义 5:学习资源的评分矩阵 A 定义如下:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}$$
 (5)

其中, n 表示学习者个数; m 表示学习资源个数; a_{ij} 表示第 i 个学习者对第 j 个学习资源的评分。一般 $a_{ij} \in [0, 5]$ 且 a_{ij} 为整数。该值表示学习者对该学习资源的喜欢程度。

定义 6:数据标准化方法为:

平移标准差变换:

$$u'_{ik} = \frac{u_{ik} - \bar{u}_k}{s_k} (i = 1, 2, \dots, n; k = 1, 2, \dots, m) \quad (6)$$

$$\text{其中, } \bar{u} = \frac{1}{n} \sum_{i=1}^n u_{ik}; s_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (u_{ik} - \bar{u})^2}。$$

平移极差变换:

$$u''_{ik} = \frac{u'_{ik} - \min_{1 \leq i \leq n} \{u'_{ik}\}}{\max_{1 \leq i \leq n} \{u'_{ik}\} - \min_{1 \leq i \leq n} \{u'_{ik}\}} \quad (7)$$

显然有 $0 \leq u''_{ik} \leq 1$, 数据标准化消除了量纲的影响。数据标准化采用平移标准差变换和平移极差变换, 将数据压缩到区间 $[0, 1]$ 上。

2 基于协同过滤的个性化学习资源推荐

第一步:数据初始化和预处理。

(1)数据初始化。

获取学习者特征集合 S (定义 2)、学习资源特征集合 M (定义 4)、学习资源的评分矩阵 A (定义 5)。构建学习者属性集合 G , 其中学习者属性包括学习者特征和学习资源评分, $G = (g_{ij})$ 为一个 $n \times (m+2)$ 矩阵, 其中 n 表示学习者个数, $m+2$ 表示学习者属性包括 m 个学习资源和两个学习者特征, 如表 2 所示。

表 2 学习者属性矩阵

学习者	学习资源评分				学习者特征	
	资源 1	资源 2	...	资源 M	lc	ls
学习者 1	3	5	...		0.6	0.2
学习者 2	1	4	...		0.8	0.5
...
学习者 N	2	3			0.7	0.9

(2)数据预处理。

对学习资源属性矩阵 $G = (g_{ij})$, 进行数据标准化处理, 使得 $g_{ij} \in [0, 1]$, 标准化方法如定义 6。

第二步:寻找最相似的学习者。

通过学习资料评分和学习者特征即学习者属性矩阵 G , 计算学习者之间的相似度。

学习者相似度计算方法:设学习者 s_a 和 s_b 的相似度为 $\text{sim}(s_a, s_b)$, 相似度计算方法采用常用的皮尔逊相关系数法, 计算公式如下:

$$\text{sim}(s_a, s_b) = \frac{\sum_{i \in U_{ab}} |g_{a,i} - \bar{g}_a| \cdot |g_{b,i} - \bar{g}_b|}{\sqrt{\sum_{i \in U_{ab}} (g_{a,i} - \bar{g}_a)^2} \cdot \sqrt{\sum_{i \in U_{ab}} (g_{b,i} - \bar{g}_b)^2}} \quad (8)$$

其中, 集合 U_{ab} 是学习者 s_a 和 s_b 共有的属性集合; $g_{a,i}$ 和 $g_{b,i}$ 分别表示学习者 s_a 和 s_b 的第 i 个属性值; \bar{g}_a

和 \bar{g}_b 分别表示学习者 s_a 和 s_b 所有属性的平均值。

第三步:计算学习资源的预测评分。

根据目标学习者的相似学习者, 计算目标学习者未评分的学习资源的预测评分。

(1)通过公式(8)计算目标用户与其他用户相似度值, 选取与目标学习者 s_T 相似度值最大的 k 个学习者, 构建集合 $NL = \{s_1, s_2, \dots, s_k\}$ 。

(2)基于集合 NL 对目标学习者 s_T 未评分的学习资源 m_T 预测评分值 p , 计算公式如下。

$$p = \bar{a}_{s_T} + \frac{\sum_{s \in NL} \text{sim}(s_T, s) \cdot (a_{s, m_T} - \bar{a}_s)}{\sum_{s \in NL} \text{sim}(s_T, s)} \quad (9)$$

其中, $\text{sim}(s_T, s)$ 表示目标学习者 s_T 与最相似学习者 s 的相似度; a_{s, m_T} 表示学习者 s 对学习资源 m_T 的评分; \bar{a}_{s_T} 和 \bar{a}_s 分别表示学习者 s_T 和 s 对学习资源评分的平均值。

第四步:产生推荐结果。

目标学习者对未评分学习资源的预测评分后, 基于预测评分和目标学习者同未评分学习资源关联度计算学习资源推荐度, 同时产生 TOP- N 推荐集。

(1)目标学习者 s_T 同未评分学习资源 m_T 的特征匹配度值包括学习者认知能力和学习资源难度的关联度 dsm 、学习者爱好类型与学习资源类型的关联度 rsm 。计算公式如下。

$$\text{dsm} = e^{-\left(\frac{\text{md}_T - \text{lc}_T}{\text{lc}_T/3}\right)^2} \quad (10)$$

$$\text{rsm} = e^{-\left(\frac{\text{mt}_T - \text{ls}_T}{\text{ls}_T/3}\right)^2} \quad (11)$$

其中, $\text{dsm} \in [0, 1]$; $\text{rsm} \in [0, 1]$, 值越大关联度越高。

(2)将未评分学习资源的预测评分 p 和学习者与学习资源关联度 dsm 和 rsm 构成推荐集合 $Q = \langle p, \text{dsm}, \text{rsm} \rangle$, 采用定义(8)的标准化方法将预测值标准化。基于 Q 计算该学习资源的推荐度。推荐度计算公式如下:

$$h = Q * w \quad (12)$$

其中, w 为各属性的权重值; Q 为未评分资源的推荐集合, 选取 Top- N 个学习资源为推荐学习资源。

3 实验及结果分析

3.1 参数设置

实验采用 Matlab7.0 实现该算法。学习者特征 L 、学习资源特征 Materials 、学习资源评分 A 的量化都采用产生随机数的方法生成数据。该推荐算法中, 学习资源数目 $m = 100$, 分别选取协作学习小组中的人数 $n = \{3, 4, 5, 6, 7, 8\}$, 计算在不同的小组人数下, 产生的

推荐集。

3.2 度量标准

评价推荐系统推荐质量的度量标准主要包括统计精度度量方法和决策支持精度度量方法两类^[1]。统计精度度量方法中的平均绝对偏差 (Mean Absolute Error, MAE) 易于理解,可以直观地对推荐质量进行度量,是最常用的一种对推荐质量度量的方法。在个性化推荐算法研究中,希望推荐的结果不仅是根据基本协同过滤算法中的预测评分值,而且还根据学习者个性化特征产生的个性化推荐结果。个性化推荐结果的度量标准,采用推荐学习资源的预测评分和学习资源与学习者属性匹配度的平均绝对偏差值之和 PMAE 作为度量标准,PMAE 越小,推荐精度越高,越符合学习者个性化需求。

设目标学习者 s ,产生的推荐结果为 m ,目标学习者对 m 的评分值为 p ,协同过滤算法计算出的预测评分为 p' ,则 PMAE 定义为:

$$PMAE = |p - p'| + |lc - md| + |ls - mt| \tag{13}$$

其中, $lc, ls \in s, md, mt \in m, |lc - md|, |ls - mt|$ 代表推荐学习资源同学习者的特征匹配度。

3.3 性能分析

为了验证该算法对学习者的推荐学习资源的准确性,使用三组数据分别为随机推荐法 (R)、基本协同过滤算法 (CF) 和个性化协同推荐算法 (PCLCF) 进行对比试验。

在实验中,协作小组人数从 3 个增加到 8 个,每次增加 1 个,查看小组中每个学习者推荐精度的平均值对比结果,如表 3 所示。

表 3 三种推荐算法的实验结果

小组人数	个性化推荐算法	传统协同过滤推荐算法	随机推荐算法
3	1.345	1.595	1.962
4	0.797	1.379	2.067
5	1.210	1.264	1.795
6	1.144	1.337	2.308
7	1.076	1.290	2.326
8	1.174	1.409	2.088

个性化学习资源推荐同其他推荐算法的推荐精度比较如图 2 所示。

由于推荐精度 PMAE 值越小,推荐精度越高,由图 2 可见,个性化学习资源推荐比传统推荐算法和随机推荐算法的个性化精度高。这是因为,个性化学习资源推荐,在使用传统推荐算法的基础上,考虑到了学习资源特征是否符合学习者的个性化需求,根据相似学习者的特征和学习资源的特征,为学习者提供更符合学习者特征的学习资源。

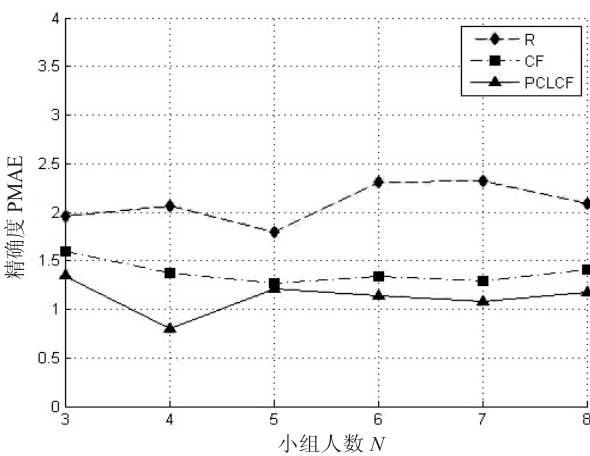


图 2 三种算法的 PMAE 值对比

4 结束语

采用基于内存的协同过滤推荐算法,通过学习者特征、学习资源特征和学习资源评分相结合,为学习者推荐更符合学习者学习风格的个性化学习资源,推荐的性能通过个性化推荐精度来判断,该值越小,推荐越准确。推荐精度的计算,采用传统推荐算法的推荐精度与个性化精度之和。结果表明,PCLCF 算法更好地根据学习者的特性,推荐个性化学习资源。此外,从教育而言,PCLCF 能够根据学习者认知能力、兴趣爱好进行推荐,为学习者推荐更适合自己的学习资源,而不是一味地根据学习资源的评分,这样能更好地提高学习者效率。

参考文献:

[1] 姜 强,赵 蔚,杜 欣,等. 基于用户模型的个性化本体学习资源推荐研究[J]. 中国电化教育,2010(5):106-111.

[2] Graf S. Adaptivity in learning management systems focusing on learning style[D]. Vienna:University of Vienna,2007.

[3] Chen Chih-Ming. Intelligent web-based learning system with personalized learning path guidance[J]. Computer & Education,2008,51(2):787-814.

[4] Yang Y J,Wu Chuni. An attribute-based ant colony system for adaptive learning object recommendation[J]. Expert Systems with Applications,2009,36(2):3034-3047.

[5] Klasnja-Milicevic A,Vesin B,Ivanovic M,et al. E-learning personalization based on hybrid recommendation strategy and learning style identification[J]. Computer & Education,2011,56(3):885-899.

[6] Marsh H W,Cooper T L. Prior subject interest,students' evaluations,and instructional effectiveness[J]. Multivariate Behavioral Research,1981,16(1):88-104.

[7] 李 聪,梁昌勇,马 丽,等. 基于领域最近邻的协同过滤

从表 2 中可以看出:考虑函数 f_1 , 三种 PSO 算法的收敛次数都同为 20 次, LDWPSO 算法的平均迭代次数是 486, 压缩因子 PSO 算法的平均迭代次数为 451, 而飞行时间单调递减的 PSO 算法的平均迭代次数是 249, 明显比前两种粒子群优化算法迭代的次数少。对于函数 f_2 , LDWPSO 算法的平均迭代次数是 64, 压缩因子 PSO 算法是 29, 而飞行时间单调递减的 PSO 算法是 31, 跟压缩因子的差不多。对于函数 f_3 , 三种粒子群算法的平均迭代次数分别为 394、318、198, 飞行时间单调递减的 PSO 算法的平均迭代次数较少。对于 f_4 函数来说, 三种算法的平均迭代次数分别为 168、177、144, 飞行时间单调递减的 PSO 算法的平均迭代次数依旧最少。所以, 通过以上分析可以看出, 与带压缩因子的粒子群算法和 LDWPSO 算法相比, 飞行时间单调递减的 PSO 算法的效果明显较好。综述, 飞行时间单调递减的 PSO 算法减少了振荡现象的产生, 改善了算法的寻优能力。

4 结束语

文中把飞行时间单调递减的思路引入到粒子群算法中, 主要依据这样的理论: 在算法的起始阶段, 粒子离最优位置较远, 所需飞行的时间要偏长些, 而到了后期, 粒子离最优位置较近, 所以所需的时间又偏短些, 如果依旧保持原来的状态继续朝前飞行, 则就越过最优位置的可能性, 进而就会产生振荡现象。引入之后便能比较合理地控制不同位置时飞行所需时间, 从而使算法的效率更高。算法测试结果表明了飞行时间单调递减的粒子群算法减少了振荡现象, 提高了算法的效率。

参考文献:

- [1] Kennedy J, Eberhart R C, Shi Y. Swarm intelligence [M]. San Francisco: Morgan Kaufmann, 2001.
- [2] Engelbrecht A P. Fundamentals of computational swarm intelligence [M]. New York: Wiley, 2009.
- [3] Das Sharma K, Chatterjee A, Rakshit A. A hybrid approach for design of stable adaptive fuzzy contrllers employing Lyapunov theory and particle swarm optimization [J]. IEEE Transactions on Fuzzy Systems, 2009, 17(2): 329–342.
- [4] Sun Jun, Fang Wei, Xu Wenbo. A quantum-behaved particle swarm optimization with diversity-guided mutation for the design of two-dimensional IIR digital filters [J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2010, 57(2): 141–145.
- [5] Gao Hao, Xu Wenbo, Sun Jun, et al. Multilevel thresholding for image segmentation through an improved quantum-behaved particle swarm algorithm [J]. IEEE Transactions on Instrumentation and Measurement, 2010, 59(4): 934–946.
- [6] Voumvoulakis E M, Hatziaargyriou N D. A particle swarm optimization method for power system dynamic security control [J]. IEEE Transactions on Power Systems, 2010, 25(2): 1032–1041.
- [7] Li Changhe, Yang Shengxiang. An adaptive learning particle swarm optimizer for function optimization [C]//Proc of IEEE congress on evolutionary computation. Trondheim: IEEE, 2009: 381–388.
- [8] Shi Yuhui, Eberhart R. A modified particle swarm optimizer [C]//Proceedings of the IEEE international conference on evolutionary computation. Anchorage, AK: IEEE, 1998: 69–73.
- [9] 曾建潮, 介倩, 崔志华. 微粒群算法 [M]. 北京: 科学出版社, 2004.
- [10] Eberhart R C, Shi Yuhui. Particle swarm optimization: developments, applications and resources [C]//Proceedings of the 2001 congress on evolutionary computation. Seoul: IEEE, 2001: 81–86.
- [11] 杨亚平, 谭瑛, 曾建潮. 二次微粒群算法及其参数选择 [J]. 计算机仿真, 2007, 24(1): 174–177.
- [12] Angeline P J. Using selection to improve particle swarm optimization [C]//Proceedings of the 1999 congress on evolutionary computation. Anchorage: IEEE, 1999: 84–89.
- [13] 孙书芳, 张家海, 徐林. 一种新型自适应 PSO 算法 [J]. 科教文汇, 2007(14): 196–197.
- [14] 许永峰, 张书玲. 带组织的粒子群优化算法—OPSO [J]. 计算机应用与软件, 2008, 25(2): 234–236.
- [15] New York ACM Press, 2001: 285–295.
- [16] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法 [J]. 软件学报, 2003, 14(9): 1621–1628.
- [17] Kolb D A. Learning style inventory technical manual [M]. Boston: McBer & Co, 1974.
- [18] Larkin-Hein T, Bundy D D. Research on learning style: application in the physics and engineering classroom [J]. IEEE Transactions on Education, 2001, 44(3): 276–281.
- [19] 布卢姆. 教育目标分类学: 第一分册 (认知领域) [M]. 上海: 华东师范大学出版社, 1986.

(上接第 66 页)

- [1] 推荐算法 [J]. 计算机研究与发展, 2008, 45(9): 1532–1538.
- [2] 黄创光, 印鉴, 汪静, 等. 不确定近邻的协同过滤推荐算法 [J]. 计算机学报, 2010, 33(8): 1369–1377.
- [3] 吴发青, 贺棵, 夏薇薇, 等. 一种基于用户兴趣局部相似性的推荐算法 [J]. 计算机应用, 2008, 28(8): 1981–1985.
- [4] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms [C]//Proceedings of the 10th international conference on World Wide Web.

协作学习中基于协同过滤的学习资源推荐研究

作者:

作者单位:


刊名:

英文刊名:

年, 卷(期):

叶树鑫, 何聚厚, YE Shu-xin, HE Ju-hou

叶树鑫, YE Shu-xin(陕西师范大学 计算机科学学院, 陕西 西安, 710062), 何聚厚, HE Ju-hou(陕西师范大学 计算机科学学院, 陕西 西安 710062; 陕西师范大学 现代教学技术教育部重点实验室, 陕西 西安 710062)

计算机技术与发展

Computer Technology and Development

2014(10)

本文链接: http://d.wanfangdata.com.cn/Periodical_wjz201410015.aspx