

基于相似学习者发现的资源推荐系统

王志梅^{1,2}, 杨帆¹

(1. 上海交通大学 计算机科学与工程学系, 上海 200030; 2. 温州职业技术学院 计算机系, 浙江 温州 325035)

摘 要: 为了解决给规模大、分布广的远程学习者提供个性化的学习资源推荐的这一难点问题, 提出了一种新颖的基于多代理的远程协作学习资源推荐机制, 引入学习状态评估向量对复杂的学习行为进行建模, 并基于组代理之间的交互来发现具有相似学习状态的学习者, 同时构建了一个供社区学习者共享的即时交互和学习资源推荐平台, 为社区学习者的学习资源共享提供了有针对性的推荐. 实验证明, 具有较高的相似学生发现准确率和社区构建效率, 同时通过社区学生的资源推荐和共享, 提高了学习者的学习效率和成绩.

关键词: 远程教育; 资源过滤; 推荐系统; 学习社区

中图分类号: TP 391.7

文献标识码: A

文章编号: 1008-973X(2006)10-1688-04

Resource recommendation system based on similar learners exploitation

WANG Zhi-mei^{1,2}, YANG Fan¹

(1. Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, China;

2. Department of Computer, Wenzhou Vocational and Technical College, Wenzhou 325035, China)

Abstract: A novel E-learning resource collaborative recommendation mechanism based on multi agent was proposed to solve the big challenge to provide personalized learning resource to large scale and distributed E-learners. A learning status evaluation vector was introduced to model the E-learners' behaviors, and help to find and reorganize the learners' share of similar learning status into smaller communities based on the interaction between group agents. Further more a collaborative communication and learning resource recommendation platform was developed to enable the learner to share personalized recommended resources. Experimental results showed that this algorithm has higher discovery accuracy of similar E-learner and construction efficiency of community. Based on the learning resource recommendation and sharing among E-learners in the common community, this system is proved to enhance the learning effect and scores.

Key words: E-learning resource filtering recommendation system learning communities

E-Learning 在线课程的设置通常涉及大量的具有不同专业背景、不同学习偏好及不同需要的学生, 这些远程学习者通常很难判断哪些学习资源最适合他们的需要及阅读. 因此, 迫切需要构建一个能进行个性化学习资源推荐的机制^[1].

个性化推荐方法最早应用于电子商务领域^[2-4], 针对特定用户提供一系列的推荐产品以帮助顾客找到自己想要购买的产品^[5-6]. 一些文献表明, 已经有

很多的研究都尝试将推荐系统应用到 E-Learning 领域中^[7-8]. 然而, 这些系统需要老师或者专家对教学资源进行特征描述和重新组织, 而这些工作往往要耗费大量的时间和精力. 同时, 这些系统主要都是从资源本身出发, 而忽略了学习者之间的关联. 通过对真实学习者的调查, 发现在学习过程中, 那些具有相似学习偏好及学习水平的学生通常都对学习资源具有相似的需求. 因此, 如果能够提出一种有效的方

法以帮助相似的学习者共享有用的学习资源, 将会大大减少老师的工作量, 同时, 利用学习者之间的互助性, 也可以较好地提高其学习效率.

本文提出了一种新颖的 E-Learning 协作资源推荐算法, 通过发现具有相似学习状态的学习者之间的关联, 构建了一个供社区学习者共享及交互的资源推荐平台, 为社区学习者的学习资源共享提供了有针对性的推荐. 该系统为每个学习者建立一个学习者代理(learn agent, LA), 负责收集学习者的学习数据, 并构建学习评估矩阵. 在整个学习过程中, 每个学习者代理必须向一个组代理注册并动态地发送其学习状态, 而组代理则基于社区自组织算法, 动态地发现具有相似学习状况和兴趣的学习者并重组到同一社区. 同社区的学习者可以在本个性化推荐平台中进行即时交流, 并推荐感兴趣的资源, 最终使同一社区的学习者能通过互助获得有帮助的学习资源.

1 系统描述

本系统主要由 3 部分构成: 学习者档案构建及维护模块、相似学习者关联发现模块、学习资源推荐模块, 系统框架如图 1 所示. 其中每一部分都与 E-Learning 的界面、学习日志数据库及基于概念图组织的学习资源库相关联.

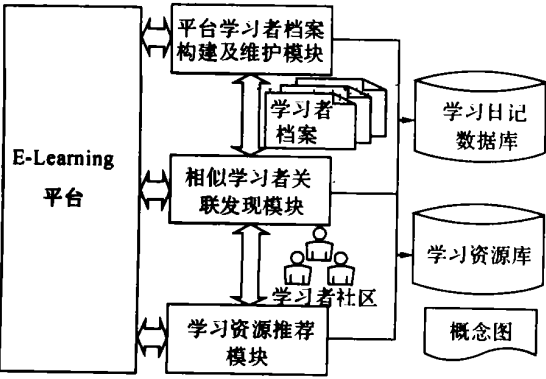


图 1 E-Learning 资源推荐系统框架

Fig. 1 Framework of E-Learning resource recommendation system

1.1 学习者档案构建及维护模块

推荐系统开发技术的实现在很大程度上要依赖所使用的信息类型^[9]. 在学习过程中, 学习者要通过 E-Learning 平台在线浏览课程、提交问题及作业、完成测试等. 所有的这些行为都代表了学习者的学习状况及需求. 本模块的目的就是获取每一个已经在该 E-Learning 平台注册的学习者的学习信息, 即

从学习者个人的历史日志数据中获取学习数据, 并为其构建和维护一个可重用的学习者档案.

为了实现对学习数据进行监控和分析, 笔者为每个学习者构建了一个 LA, 其主要负责:

1) 监控学习行为

LA 将监控其代表的学习者的学习行为, 如学习路径、访问频率、提交的问题及查询的资源等, 并负责从各种数据资源(web 服务、课件学习平台、问答系统、在线作业系统及讨论组等)中收集这些历史学习行为, 并对噪音数据进行清理.

2) 学习者档案的产生及维护

一个学习者档案包括两个部分: ① 学习资源列表, 包括已学过的学习资料、总结的学习经验和对某些问题的解答; ② 学习状态评估向量, 令 $C = \{c_1, \dots, c_2, \dots, c_n\}$ 为与某一特殊学习领域相关的学习概念集合, 那么学习者 a 的学习状况评估向量可定义为 $V = \{(c_1, s_1), (c_2, s_2), \dots, (c_n, s_n)\}$, 其中 $c_i \in C$ 是一个学习概念, s_i 是学习者 a 对概念 c_i 的学习掌握状况的评估值. 这里对学习掌握状况的评估值主要依赖于显式的和隐式的两种方式: 显式的方式可以通过学习者对某个学习资源的评分或是一次测试的分数来进行定量的表示; 隐式的方式则需要从学习行为数据中分析得出, 如访问频率、学习的持续时间等. E-Learning 的日志数据中包含了很多被喻为离线学习数据的详细资料, 这有助于理解学习者的学习行为, 获得全面的学习状况评估.

1.2 相似学习者关联发现模块

“学习者社区”在这里被定义为具有相似学习状况并能提供互助的、学习资源的学习者组成的一个团体. 本模块负责为管理学习者社区生成若干个组代理(group agent, GA). 一个组代理 GA 负责管理学习者、与其他组代理交互并发现相似用户、交换相似用户以构建社区. 这种基于组代理的社区结构可以通过 LA 与 GA 之间的映射 $m: L \rightarrow G$ 来描述, 这里 $m(l) = g$ 表示学习者 l 是组 g 的一个成员. g 所管理的所有学习者代理可以用式(1)来表示:

$$A_g = \{l \in L \mid m(l) = g\}.$$
 (1)

在初始化阶段, 每个学习者代理 a 都将随机地选择一个 GA g 并且报告它当前的 V_a . 组代理 g 将建立成员组 A_g 并基于每个成员的 V 构建组状态的学习评估矩阵 M_g . 在学习过程中, 如果监控学习者已经学习了一个概念, 学习者代理 a 将发送一组信息 (a, c_i^a, s_i^a, V_a) 给组代理 g , 这里 a 代表 LA, (c_i^a, s_i^a) 代表 <概念-状况> 二元组, V_a 是更新的学习状况评估向量.

GA g 接收到这个询问后将向它的成员发送消息并随机地转发给其他 m 个的组代理 g' , 其中 $m < \text{topSearch}$, 这里, 参数 topSearch 表示搜索宽度, 以减少通信量降低通信阻塞的概率. 接收到这个消息后, 每一个学习者 l 将检查它是否学习过这个概念, 如果 l 学习过这个概念则比较学习者 l 和 a 对概念 c_i 的评估值, 如果其评估值之差低于某阈值 α , 即 $|S_i^a - S_i^l| < \alpha$, 则计算 V_a 与 V_l 之间皮尔森相关系数, 如式(2)所示:

$$\text{si } m_{a,l} = \frac{\sum_i (s_i^a - s^a)(s_i^l - s^l)}{\sqrt{\sum_i (s_i^a - s^a)^2 \sum_j (s_j^l - s^l)^2}} \cdot x \cdot (2)$$

如果相关系数 $\text{si } m_{a,j}$ 高于某阈值 β , 那么意味着学习者代理 a 与 l 具有相似的学习状态, 然后学习者代理 l 将反馈 $(l, \text{si } m_{a,l})$ 给组代理 g . 如果 l 收到的这个消息是由组代理 g' 转发的, 则 l 会将消息反馈给组代理 g' 而由 g' 返回给组代理 g .

一段等待时间 T 后, 组代理 g 对反馈回来的信息按 $\text{si } m_{a,j}$ 进行升序排列, 并为匹配用户 l 选择一个 $\text{si } m_{a,j}$ 最高的作为本次匹配的最相似学习者; 然后组代理 g 将检查学习者代理 a 和 l 是否在同一个组代理中, 如果不在, 那么组代理 g 和 g' 将采用类似公式(2)的方法计算每一个学习者代理的交互相似系数, 然后将该系数小的学习者代理移出到另外的组中, 社区自组织算法描述如下:

- 1) 创建学习者代理;
- 2) 创建组代理;
- 3) 对每一个学习者进行初始化;

- 4) 随机选择一个组代理 GA 进行注册;
- 5) 向组代理发布其学习评估向量 V_a ;
- 6) 增加 a 到组代理 $g: A_g = A_g \cup a$;
- 7) 更新学习状态矩阵 $M_g: M_g = M_g \cup V_a$;
- 8) 结束初始化;
- 9) 对每一个学习者进行轮询, 发送个性化推荐请求;
- 10) 如果某学习者学完了一个学习对象;
- 11) 其学习者代理向组代理发布最新的学习状态评估值, 并由组代理负责寻找最新的具有相似学习状态的学习者: send a message to g and call: $g \cdot \text{ReceiveEnquire}(a, c_i^a, s_i^a, V_a)$;
- 12) 对相似学习者按相似度进行排序;
- 13) 选择具有最高相似度的学习者;
- 14) 对匹配学习者进行判断, 如果不在同一个组中, 将具有小的交互相似系数的 LA 移动到另外的组中 if $a \in A_g, l \in A_{g'}, g \neq g'$ then Exchange LA (a, g, l, g')
- 15) 结束.

1.3 学习资源推荐模块

为了协助相似学习者进行交流、共享和推荐学习资源, 本模块基于 JADE 平台开发了一个基于 Agent 的学习资源推荐平台, 如图 2 所示. 本平台左边框包括 3 个主要面板: “学习资源列表”、“推荐资源列表”、“邻居列表”. 单击面板内每一个对象将激活对应的功能, 并显示在右边框中: “学习资源列表”

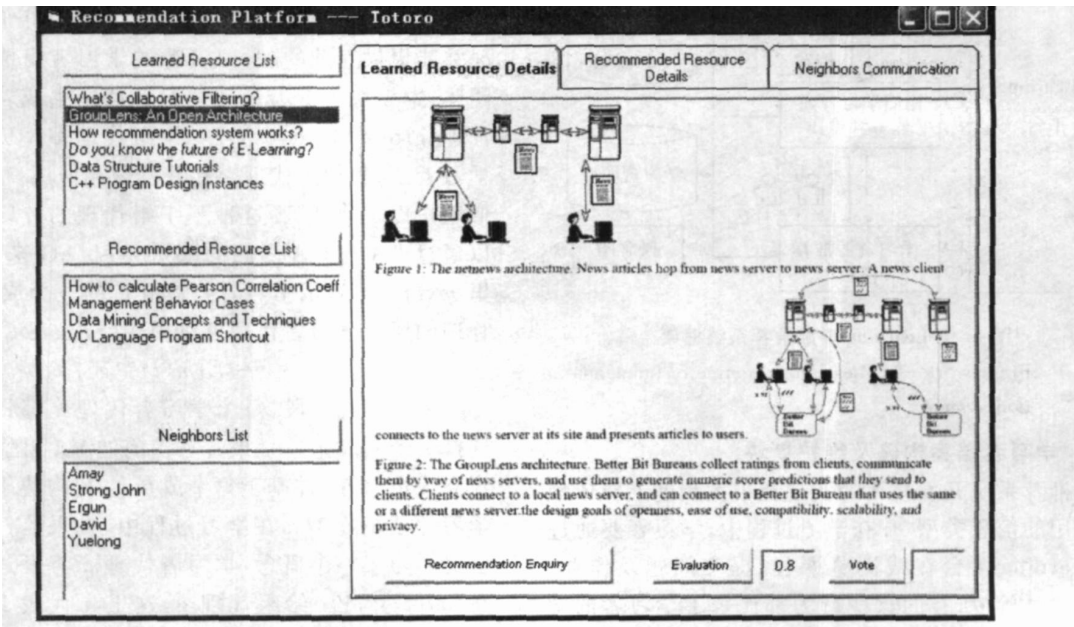


图 2 推荐系统平台

Fig. 2 Recommendation platform

主要给出了学习者已经学习过的资源列表, 点击某一个资源名称, 可以激活右边框中的“已学资源详情”面板并显示该资源的详细内容, 学习者可以阅读该内容并对该内容进行评估, 同时还可以对自己感兴趣的资源进行推荐. 例如, 一个学习者选择了“GroupLens :An Open Architecture”条目, 其在右边框中相关的面板“学习资源详细情况”将被激活并显示出详细的内容. 如果这个学习者认为该资源值得推荐, 则可以点击“Recommend”按钮进行推荐. 一旦该学习者选择了推荐该资源, 其学习者代理 l 将捕获该请求并递交给其组代理 g , 而后 g 将该资源传递给同组的其他分布的邻居学习者. 学习者代理 l 的每一个相似学习者都将查看到该资源推荐, 并可以选择将其添加到“推荐资源列表”, 学习者可以通过点击该列表中的文件名称来查看其详细内容, 对于自己感兴趣的推荐资源, 它可以给定自己的评估值并保存到已读资源中. 特别地, “邻居列表”列出了 l 的所有相似学习者, 并通过“通信模块”和社区中的所有学习者进行交互. 推荐过程对所有的学习者是透明的.

2 实验分析

从上海交通大学网络教育学院选择了 200 个自愿者, 以“大学英语”这门课程作为测试课程, 并设计了一系列限期完成的测试题目. 这些志愿者被分成 A 和 B 两组, 每组 100 人. A 组的学生采用普通的学习方式, 并在基于 E-Learning 平台上进行测试; B 组的学生按要求使用在线推荐系统. 也就是说, 在学习过程中, 系统为 B 组中的每个学生产生了一个 LA 以及几个 GA 来管理这个学习者社区. 每个 LA 监控着相关概念的测试成绩, 而 GA 执行社区组织算法并组织具有相似状况的学习者进入同一个社区, 同时引导 B 组学生去学习与测试相关的资源, 写读书笔记, 收集某些特殊问题的解答等, 还鼓励他们在测试过程中使用该推荐系统平台进行学习资源的推荐. 每个学生都能收到、读到及评估由其他学生推荐的学习资源.

根据试验的方案, 笔者认为当学生完成一个测试以后, 将发生 <概念, 学习状态> 请求. 对每个请求, 将检查当运行接收询问算法后交换是否发生并计算每个请求(学生)的平均交换率作为社区组织效率. 图 3 给出了每个学生平均发出 100 次请求后的交换频率曲线, 描述了社区交换频率 f 随平均每个学生的请求次数 n 的变化曲线. f 值越大表示社区

交换次数越多, 社区越不稳定; f 值越小则表示社区越趋于动态稳定. 由图 3 可见, 在初始化阶段当学生还没有组织好的时候, 交换频率很高. 这是因为在初始状态 j 时, 组代理 GA 常常不能在它自己的组内找到匹配的学生而要转发请求给其他的 GA. 大概在平均发出 23 次请求后, 系统组织具有相似学习状态的学习者的能力明显增强, 而且这些学习者都被组织到了相同的社区, 达到了一个稳定的状态, 因此其社区交换效率逐渐趋于一个较小的值, 说明社区构建已经基本完成. 为了证明本系统的效率, 在学生完成了所有的测试以后进行了一个评估, 两组学生分别完成了同一套试题, A 组和 B 组学生的平均成绩比较数据如图 4 所示. 从图可以看出, 分数分布在区间 $[B, A +]$ 的人数中, B 组学生比 A 组学生多, 而分数分布在区间 $[D, B -]$ 的人数中, 却刚好相反. 同时还在 B 组的 100 个自愿者中作了一项调查, 其中 54% 的学生认为该系统是有帮助且有效的.

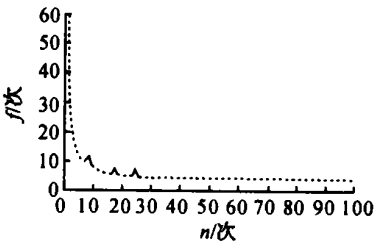


图 3 社区组织效率曲线图
Fig. 3 Community organizing efficiency

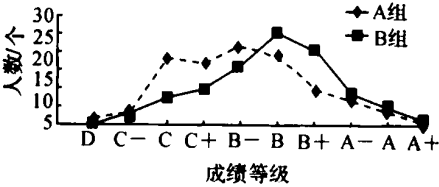


图 4 两组自愿者的成绩比较
Fig. 4 Achievement comparison between two volunteer groups

3 结 语

本文提出了一种新颖的基于学习状态匹配的相似学习者关联发现算法, 并为形成的自组织学习者社区构建了一个协作交互及资源推荐系统. 该系统对解决 E-Learning 领域所面临的学习者分散学习状态难获取、学习者相对孤立学习积极性低、学习资源统一缺乏针对性等瓶颈问题, 提供了一个有效的解决方案.

表 4 工况 9 预测情况数据表
Tab. 4 Forecast of condition 9

<i>g</i>	<i>C</i>	$\delta/\%$	<i>n</i>	<i>t/s</i>
8×10^{-5}	150	1.6	49	$\ll 1$
2×10^{-5}	800	1.9	49	$\ll 1$
0.03	200	6.7	48	$\ll 1$

本的分布和所包含信息是否充分有关. 在选择训练样本时应尽量使样本含有充分的信息.

通过以上参数的选择过程可知, 参数 *g* 和 *C* 对模型的性能都具有很大的影响, 其影响只在一定区域内比较明显, 这一点为寻优操作带来方便, 使寻优范围大大减小, 与寻优算法结合时效率会更高. 在选择参数时可以先对其中一个参数进行选择, 选择好后再对另一个进行选择. 用支持向量机建立的 NO_x 排放模型可以快速、准确地做出预测, 核函数参数的选择和罚因子的选择对模型的性能有很大的影响, 通过调整参数可以获得性能好的模型. 虽然支持向量机可从理论上保证模型的泛化能力, 但这个保证是比较宽松的, 模型的预测能力还与样本的数量和分布有一定关系, 从预测精度考虑选择训练样本时应该尽量选择分布比较均匀、数量尽量多的样本.

3 结 语

本文采用支持向量机方法建立了 NO_x 排放量预测模型. 通过将热态实炉试验数据作为训练样本、检测样本, 对模型进行了训练和检验, 并通过参数选择使模型达到了较好的效果, 可以快速、准确地对 NO_x 排放量进行预测. 与其他建模方法相比, 支持

向量机从理论上保证了模型的泛化能力, 具有更加可靠的推广性, 计算速度更快, 更适合于在线工作, 结合全局寻优算法, 可以在燃用任何煤种条件下寻找出最优的操作参数, 实现在线指导优化运行.

参考文献(References):

[1] 周昊, 朱洪波, 曾庭华, 等. 大型四角切圆燃烧锅炉 NO_x 排放特性的神经网络模型[J]. 中国电机工程学报, 2002, 22(1): 33-37.
ZHOU Hao, ZHU Hong bo, ZENG Hua ting, et al. An artificial neural network model on NO_x emission property of a high capacity tangentially firing boiler [J]. **Proceeding of the CSEE**, 2002, 22(1): 33-37.

[2] 王培红, 李磊磊, 陈强, 等. 电站锅炉 NO_x 排放与效率的响应特性模型[J]. 动力工程, 2004, 24(2): 254-259.
WANG Pei hong, LI Lei lei, CHEN Qiang, et al. Response characteristics model of NO_x emission and efficiency for power station boiler [J]. **Power Engineering**, 2004, 24(2): 254-259.

[3] CRISTIANINI N, SHAWE-TAYLOR J. 支持向量机导论[M]. 李国正, 王猛, 增华军, 译. 北京: 电子工业出版社, 2004.

[4] Vapnik. 统计学习理论[M]. 张学工, 译. 北京: 电子工业出版社, 2004.

[5] KEERTHI S S, LIN C J. Asymptotic behaviors of support vector machines with Gaussian kernel [J]. **Neural Computation**, 2003, 15(12): 1667-1689.

[6] LIN H T, LIN C J. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO type methods [EB/OL]. [2003-03-15]. <http://www.csie.ntu.edu.tw/~cjlin/papers.html>.

(上接第 1691 页)

参考文献(References):

[1] AAI ANE O R. Web usage mining for better web based learning environment [C] // **Proceedings of Computers and Advanced Technology in Education**. Alberta: IAST-ED, 2001: 60-64.

[2] LEE L, LU T C. Intelligent agent based systems for personalized recommendations in Internet commerce [J]. **Expert Systems with Applications**, 2002(22): 275-284.

[3] RESNICK P, VARIAN H R. Recommender systems [J]. **Communications of the ACM**, 1997(40): 56-58.

[4] BALABANOVIC M, SHOHAM Y. Fab: Content-based, collaborative recommendation [J]. **Communications of the ACM**, 1997, 40(3): 66-72.

[5] CHEUNG K W, KWOK J T, LAW M H, et al. Mining customer product ratings for personalized marketing [J]. **Decision Support Systems**, 2003(35): 231-243.

[6] SCHAFER J B, KONSTAN J A, RIEDL J. E-commerce recommendation applications [J]. **Data Mining and Knowledge Discovery**, 2001(5): 115-153.

[7] SHEN L P, SHEN R M. Learning content recommendation service based on simple sequencing specification [C] // **Proceedings of Advanced in Web-based Learning (ICWL 2004)**. Heidelberg: Springer, 2004: 363-370.

[8] LUO S, SHA S, SHEN D, et al. Conceptual network based courseware navigation and web presentation mechanisms [C] // **Proceedings of Advanced in Web-Based Learning (ICWL 2002)**. Heidelberg: Springer, 2002: 81-91.

[9] YANG J D, LEE D G. Incorporating concept-based match into fuzzy production rules [J]. **Information Sciences**, 1998(104): 213-239.