

一种基于语义网的个性化学习资源推荐算法

刘志勇^{1,2}, 刘磊¹, 刘萍萍¹, 杨帆³, 贾冰³

(1. 吉林大学 计算机科学与技术学院, 长春 130012; 2. 东北师范大学 软件学院, 长春 130024; 3. 东北师范大学 理想信息技术研究院, 长春 130024)

摘要:在 E-learning 环境中, 为了满足用户对学习资源的个性化需求, 提出了一种基于语义网技术的学习资源个性化推荐算法。首先根据用户评价和浏览行为得到用户感兴趣的学习资源集合与核心概念集合, 然后根据领域本体中概念间的关系分别计算不同用户评价的学习资源集合间的语义相似度和核心概念集合间的语义相似度, 最后根据得出的两个相似度值共同决定用户兴趣偏好的相似性, 找到具有相似兴趣的最近邻居, 从而实现学习资源的协同推荐。此外, 在学习资源管理上引入了学习对象概念, 降低了相似度计算的复杂度。并将该算法应用到了基于语义网的个性化学习资源推荐系统中, 实验表明, 该算法有效改善了学习资源推荐效果, 特别是对于新加入的资源和新注册用户效果显著。

关键词: 计算机应用; 学习资源; 语义相似度; 用户兴趣; 语义网

中图分类号: TP301.2 **文献标识码:** A **文章编号:** 1671-5497(2009)Sup. 2-0391-05

Learning resource personalizing recommendation algorithm based on semantic Web

LIU Zhi-yong^{1,2}, LIU Lei¹, LIU Ping-ping¹, YANG Fan³, JIA Bing³

(1. School of Computer Science and Technology, Jilin University, Changchun 130012, China; 2. School of Software, Northeast Normal University, Changchun 130024, China; 3. School of Ideal Information Technology Research, Northeast Normal University, Changchun 130024, China)

Abstract: In E-learning, in order to satisfy users' personalized demands for learning resources, in this paper we present a learning resources personalized recommendation algorithm based on semantic web technology. First of all, based on users' evaluations and browsing behaviors of the learning resources, we define the learning resource set and core concept set for a user; secondly, based on the relations of concepts in domain ontology, we obtain the user's preference similarity by computing semantic similarity between the different users' learning resource sets and core concept sets respectively. At last, we recommend collaboratively the learning resources according to the user's preference similarity. In addition, the learning objects for management of learning resources can reduce the complexity of similarity computation. We applied this algorithm to the personalized learning resources recommendation system, the experiments show that the algorithm improved the effectiveness of the recommended learning resources, especially for new resources and new registered users.

收稿日期: 2009-05-10.

基金项目: 国家自然科学基金项目(60873044); 东北师范大学自然科学基金青年基金项目(20061007).

作者简介: 刘志勇(1978-), 男, 博士研究生. 研究方向: 数字化学习, 知识工程. E-mail: liuzy452@nenu.edu.cn

通信作者: 刘萍萍(1979-), 女, 博士研究生. 研究方向: 人工智能, 思维计算. E-mail: liupp@jlu.edu.cn

Key words: computer application; learning resource; semantic similarity; user preference; semantic Web

随着 E-learning^[1] 和网络技术的发展,出现了越来越多的学习资源站点。但用户在海量的学习资源^[2]中找到需要的学习资源十分困难。越来越多的学者致力于个性化推荐算法的研究^[2-3]。

目前主要有协同推荐算法^[4]、基于内容的推荐算法^[5-6]和基于规则的推荐算法。尽管协同推荐算法能够挖掘用户的潜在兴趣并提供新的学习资源,但也存在一定的缺点,比如用户评价矩阵稀疏问题、新用户没有评价任何资源以及新资源没有任何用户评价等问题。基于内容的推荐算法也存在着挖掘信息不充分、仅能处理文本类型文件和缺乏用户反馈等缺点。基于规则的推荐算法,随着资源不断的增长,规则的制定和维护变得十分困难。

本文提出一种基于语义网技术的学习资源个性化推荐算法^[7]。首先根据用户对学习资源的评价和浏览行为得到用户感兴趣的学习资源集合与核心概念集合,然后根据领域本体中概念间的关系分别计算不同用户评价的学习资源集合间的语义相似度和核心概念集合间的语义相似度,最后根据得出的两个相似度值共同决定用户兴趣偏好的相似性,找到具有相似兴趣的最近邻居,从而实现学习资源的协同推荐。实验结果表明,该算法有效改善了学习资源推荐的有效性,特别是对于新加入的资源和新注册用户效果显著。

1 基于语义的学习资源组织

基本的思路是利用领域内意义明确的本体组织大量无序异构的学习资源,明确体现学习资源的语义关系。学习资源是在概念空间中(本体)被组织而不是页面空间,使得学习资源能够以一种更加灵活的操作方式被组织和重用,提供智能化推荐服务。学习资源组织模型如图 1 所示。

2 用户兴趣相似度算法

我们认为:不同用户评价了不同的学习资源和核心概念,根据领域知识,由于学习资源之间和核心概念之间都存在一定的相似性,所以用户之间也存在相似兴趣。对于不同的学习资源与核心概念,相同的用户评分对用户的相似度的贡献是不同的。用户评分的学习资源对用户相似度的贡献度由用户的评分和学习资源的相似度共同确

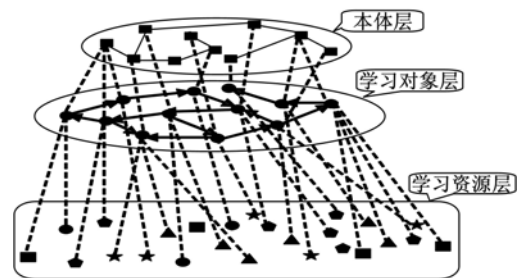


图 1 学习资源组织模型

Fig. 1 Learning resources organization model

定。用户评分的核心概念对用户相似度的贡献度由用户的评分和核心概念的相似度共同确定。因此用户的相似性通过用户评价学习资源的集合相似性与用户评价核心概念的集合相似度共同决定。

用户兴趣可以通过用户的浏览行为和对资源以及概念的评价获得。我们定义了两个集合:核心概念集合和学习资源集合,核心概念集合中的元素是用户评价过的核心概念,学习资源集合中的元素是用户评价过的学习资源。基于领域本体中概念之间的关系,分别计算出不同用户对应的两个集合的相似性。然后通过该相似性找到用户的最近邻居。最后将最近邻居感兴趣的资源排序产生推荐列表。用户兴趣的相似度由两个集合的相似度共同决定。两个集合中数据的获取方法见 2.1。

2.1 获取用户反馈

用户的反馈包括核心概念和学习资源的评分。获取途径有以下 5 种方式。

(1)用户在注册时可以粗略的给出几个核心概念来描述他们的兴趣,然后基于这几个概念发现更多关联的扩展概念。

(2)部分用户可以通过提交自己感兴趣的学习资源实例,比如教师。

(3)跟踪用户的查询来找到用户的核心概念。用户提出的查询请求通常能直接反映用户的兴趣,因此可以把频繁使用的查询关键词提取出来作为用户的核心关键词。

(4)跟踪用户的浏览行为找到用户感兴趣的学习资源。

(5)用户直接对资源 and 核心概念(资源标签)进行评价。

2.2 选择学习对象

选择一个或多个主题相关的学习对象,选取的原则是要有超过一定阈值的学习资源数量,结构如图 2 所示。

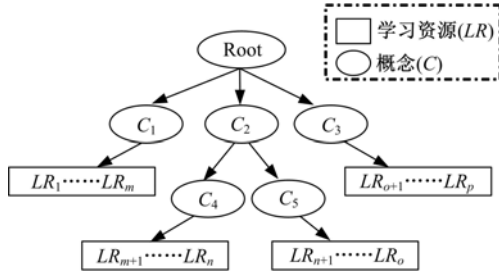


图 2 学习资源组织结构图

Fig. 2 Learning resources organization chart

2.3 学习资源语义相似度算法

定义 1 LR_i 和 LR_j 为任意学习资源节点, $LRSS(LR_i, LR_j)$ 为 LR_i 和 LR_j 的语义相似度, $0 \leq LRSS \leq 1$ 。当 LR_i 和 LR_j 为同一节点时, $LRSS=1$, 当 LR_i 和 LR_j 没有任何关系时, $LRSS=0$ 。 $depth(n)$ 是从根节点开始到节点 n 的路径的长度, 集合 $\{CA_1, CA_2, \dots, CA_n\}$ 为 LR_i 和 LR_j 的共同祖先节点集合, 若 $depth(CA_i)$ 为最大值, 则 $CA(LR_i, LR_j) = CA_i$ 。

$LRSS$ 可由公式(1)计算得到:

$$LRSS(LR_i, LR_j) = \frac{2 \times depth(CA(LR_i, LR_j))}{depth(LR_i) + depth(LR_j)} \quad (1)$$

2.4 核心概念语义相似度算法

定义 2 KC_i 和 KC_j 为任意概念节点, $KCSS(KC_i, KC_j)$ 为 KC_i 和 KC_j 的语义相似度。 $KCSS=1$ 。当 KC_i 和 KC_j 为同一节点时, $KCSS=1$, 当 KC_i 和 KC_j 没有任何关系时, $KCSS=0$ 。

$KCSS$ 可由公式(2)计算得到:

$$KCSS(KC_i, KC_j) = \frac{2 \times depth(CA(KC_i, KC_j))}{depth(KC_i) + depth(KC_j)} \quad (2)$$

2.5 学习资源得分的贡献度算法

定义 3 $Diff(PLR_{m,i}, PLR_{n,j})$ 为用户对学习资源的评价的差异度, $PLR_{m,i}$ 是用户 m 对学习资源 i 的评价, $PLR_{n,j}$ 是用户 n 对学习资源 j 的评价。为用户评分的最大值。

则 $Diff(PLR_{m,i}, PLR_{n,j})$ 可由公式(3)计算得到:

$$Diff(PLR_{m,i}, PLR_{n,j}) = \frac{\alpha - |PLR_{m,i}, PLR_{n,j}|}{\alpha} \quad (3)$$

定义 4 $LRSim(LR_i, LR_j, PLR_{m,i}, PLR_{n,j})$ 为根据用户评价的学习资源的得分得出的用户相似性。

根据公式(1)和公式(3), $LRSim$ 可由公式(4)计算得到:

$$LRSim(LR_i, LR_j, PLR_{m,i}, PLR_{n,j}) = LRSS(LR_{m,i}, LR_{n,j}) \times Diff(PLR_{m,i}, PLR_{n,j}) \quad (4)$$

2.6 核心概念得分的贡献度算法

同用户评价学习资源算法相同。

定义 5 $KCSim(KC_i, KC_j, PCK_{m,i}, PKC_{n,j})$ 为根据用户评价的概念的得分得出的用户相似性。 $PCK_{m,i}$ 是用户 m 对概念 i 的评价, $PKC_{n,j}$ 是用户 n 对概念 j 的评价。

$KCSim$ 可由公式(5)计算得到:

$$KCSim(KC_i, KC_j, PCK_{m,i}, PKC_{n,j}) = LCSS(KC_{m,i}, KC_{n,j}) \times Diff(PKC_{m,i}, PKC_{n,j}) \quad (5)$$

2.7 集合相似度算法

有了评分的学习资源之间相似度的计算方式, 我们就可以根据它计算用户的相似性。集合之间相似度的计算需要考虑相似的对称性性质和集合元素个数对相似计算结果的影响。

参考[8]中关于集合相似性算法:

1) 首先计算两个用户评分学习资源集合的所有学习资源两两之间的相似度;

2) 从所有的相似度值中选择最大的一个, 将这个相似度值对应的两个学习资源对应起来;

3) 从所有的相似度值中删去那些已经建立对应关系的学习资源的相似度值;

4) 重复上述第 2) 步和第 3) 步, 直到所有的相似度值都被删除;

5) 没有建立起对应关系的学习资源与空元素对应;

6) 求平均相似度(累加和除以对应学习资源个数), 返回。

上面的算法首先按照相似度建立起两个集合中学习资源的一一对应关系, 然后计算用户评分学习资源集合的相似度: 用户评分学习资源集合的相似度等于其元素对的相似度的加权平均。因为评分学习资源集合的元素之间都是平等的, 所

以所有的权值取成相同,那么学习资源集合的相似度等于其学习资源对的相似度的算术平均。

同理,我们可以得出用户评价的核心概念集合的相似度。

则用户间的相似度即为核心概念集合的相似度与学习资源集合相似度的算数平均。

定义 6 $userSim(setLRSim, setKCSim)$ 为用户兴趣相似度,其中 $setLRSim$ 为用户评价的学习资源集合相似度, $setKCSim$ 为用户评价的核心概念集合的相似度。

$userSim$ 可由公式(6)计算得到:

$$userSim(setLRSim, setKCSim) = (setLRSim + setKCSim)/2 \quad (6)$$

2.8 推荐资源

根据以上算法得到的相似度,把相似度超过一定阈值的用户认为是有相似兴趣的用户,然后将相似用户感兴趣的资源按照评分排序,选择评分超过一定阈值的学习资源推荐给用户。

3 实验结果分析

为了验证我们算法的有效性,我们将该算法应用到了基于语义网的个性化学习资源推荐系统中。

3.1 系统结构

系统结构如图 3 所示。系统有两个模块组成:个性化推荐模块和学习资源模块。个性化推荐模块是系统的核心,系统根据其中存储的用户兴趣进行资源的推荐。学习资源模块存储了领域本体、学习对象和学习资源,学习对象与学习资源的存储和管理建立在本体知识分类基础上。

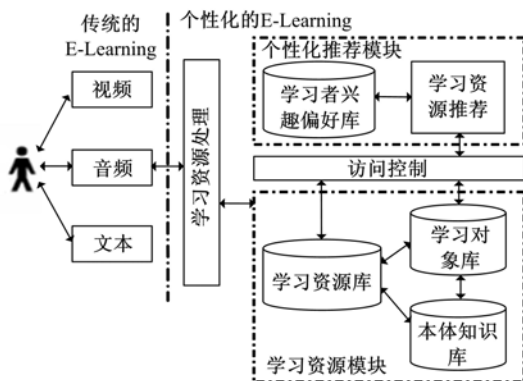


图 3 系统结构图

Fig.3 System architecture

3.2 例子

以体育比赛视频为例,知识结构如图 4 所示。

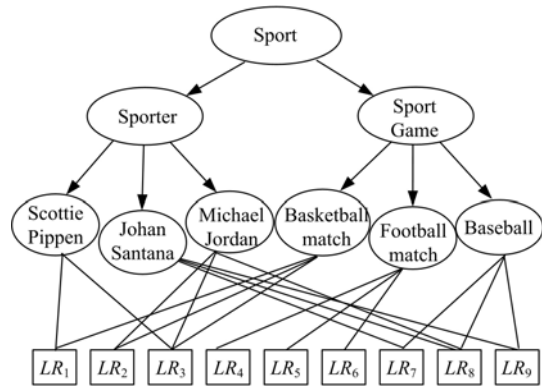


图 4 知识结构图

Fig.4 Knowledge architecture

定义 7 图中节点 LR1...LR9 为体育比赛视频片断,即学习资源。其他节点表示体育比赛涉及的概念,有向线表示概念间的关系,线段表示概念和实例关系。

定义 8 用户对概念的评价表示为 $\langle Concept_i, Score_i \rangle$ 形式,其中 $Concept_i$ 为树中的任意概念。

定义 9 用户对学习资源的评价表示为 $\langle LR_i, Score_i \rangle$ 形式,其中 LR_i 为树中的任意概念。

表 1 用户评价

Table 1 User's Evaluations

UserName	LR Evaluation Set	Concept Evaluation Set
User1	$\langle LR1, 80 \rangle \langle LR6, 70 \rangle$	$\langle Michael\ Jordan, 80 \rangle$ $\langle Football\ match, 80 \rangle$
User2	$\langle LR1, 90 \rangle$	$\langle Michael\ Jordan, 85 \rangle$
User3	$\langle LR1, 90 \rangle$	$\langle Michael\ Jordan, 90 \rangle$
User4	$\langle LR6, 80 \rangle$	$\langle Football\ match, 90 \rangle$
User5	$\langle LR6, 90 \rangle$	$\langle Football\ match, 90 \rangle$
User6	$\langle LR2, 90 \rangle \langle LR3, 85 \rangle$ $\langle LR9, 80 \rangle$	$\langle Basketball\ match, 95 \rangle$ $\langle Baseball\ match, 85 \rangle$
User7	$\langle LR3, 95 \rangle$	$\langle Basketball\ match, 90 \rangle$
User8	$\langle LR3, 60 \rangle \langle LR6, 90 \rangle$	$\langle Football\ match, 95 \rangle$
User9		$\langle Basketball\ match, 95 \rangle$

如表 1 所示,实验结果分析如下:

(1) User1 和 User2 的相似度为 0.925, User1 和 User3 的相似度为 0.9,可以看出尽管 User2 和 User3 对学习资源 LR1 的评价是相同的,由于对概念 Michael Jordan 的评价不同,所以对用户相似性的贡献是不同的。

(2) User1 和 User4 的相似度为 0.9, User1

和 User5 的相似度为 0.85,可以看出尽管 User4 和 User5 对概念 Football match 的评价是相同的,由于对学习资源 LR6 的评价不同,所以对用户相似性的贡献是不同的。

通过(1)(2)分析得出:用户的相似度由概念集合相似度和学习资源集合相似度共同确定。

(3)User6 和 User7 的相似度为 0.85,User6 和 User8 的相似度为 0.375,则 User6 和 User7 为具有相似兴趣的用户。按照资源评分顺序,依次推荐学习资源 LR2, LR9。LR3 为用户评价过的学习资源不做推荐。

通过(3)分析得出:根据用户的兴趣相似度可以推荐出用户潜在的感兴趣的资源。

(4) User9 和 User6 的相似度为 0.95,依次推荐学习资源 LR2, LR3, LR9。

通过(4)分析得出:User9 仅对一个概念做出了评价,可以视为新注册的用户,根据我们的算法仍然可以有效的推荐感兴趣的资源。

对于特殊情况,用户没有对任何资源和概念做出评价,则根据用户浏览的资源和时间长度作为用户感兴趣的资源和资源的评价。

4 结束语

本文将语义网技术应用到了学习资源个性化推荐中,利用领域本体中概念间的关系计算不同用户的兴趣相似度,根据最近邻居的兴趣实现学习资源的协同推荐。相对于传统推荐算法该算法有效提高了学习资源推荐的有效性和精确性。总体来说,该方法的优点如下:

(1)使用具有明确语义的领域本体组织大量无序异构的学习资源,学习资源模型分为三层:本体层(概念层)、学习对象层和学习资源层,三层结构大大的降低了用户兴趣偏好语义相似度的算法复杂度。

(2)通过判断学习资源间的语义相似度解决了新资源没有任何用户评价的而无法产生推荐的问题。

(3)通过跟踪用户行为解决了新用户没有评价任何资源而无法协同推荐的问题。

(4)综合考虑了用户对学习资源的评价和核心概念的评价,二者共同决定用户兴趣偏好的相似性。

参考文献:

- [1] Karl M Kapp. Winning E-Learning Proposals: The Art of Development and Delivery[M]. Ross, J Publishing, Incorporated. ISBN:1932159045,2003.
- [2] Romero P C, Ventura P S. Educational data mining: A survey from 1995 to 2005[J]. Expert Systems with Applications: An International Journal, 2007,33(1):135-146.
- [3] Perugini Saverio, Marcos Andre Goncalves, Edward A Fox. Recommender systems research: a connection-centric survey[J]. Journal of Intelligent Information Systems, 2004,23(2):107-143.
- [4] Magdalini Eirinaki, Michalis Vazirgiannis, Dimitris Kapogiannis. Web path recommendations based on page ranking and Markov models[C]// Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management (WIDM'05). Bremen, Germany, 2005.
- [5] 张丙奇. 基于领域知识的个性化推荐算法研究[J]. 计算机工程,2005,31(21):7-9.
Zhang Bing-qi. A collaborative filtering recommendation algorithm based on domain knowledge[J]. Computer Engineering,2005, 31(21):7-9.
- [6] Pluis Martinez, Luis G Perez, Manuel Barranco. A multigranular linguistic content-based recommendation model: Research Articles [C] // International Journal of Intelligent Systems, 2007, 22 (5): 419-434.
- [7] Lee W N, Shah N, Sundlass K, et al. Comparison of ontology-based semantic-similarity measures [C] // AMIA Annual Symposium Proceedings: 384-388. 2008
- [8] 谢圣献,谢光. 语义检索在电子商务中的应用研究[J]. 微计算机信息,2008,24(12):135-137.
Xie Sheng-xian, Xie Guang. Study on applying semantic retrieval to E-commerce[J]. Control and Automation, 2008,24(12):135-137.