

基于模糊树的个性化电子学习推荐系统

姜书浩, 金 格

(天津商业大学信息工程学院, 天津 300400)

摘 要: 电子学习系统的快速发展为学习者在线学习提供了巨大的机会。然而, 在线学习系统中太多的学习活动使个体学习者很难找到合适自己的学习活动, 所以在线学习系统必须有能够提供个性化产品的推荐系统。本研究首先提出了一种模糊树状结构学习活动模型, 然后结合基于知识和协同过滤推荐算法的优点提出了基于混合学习活动推荐方法的模糊树匹配方法。

关键词: 电子学习; 模糊集; 推荐系统; 树匹配

中图分类号: TP3-0 **文献标识码:** A

A Fuzzy Tree Matching-Based Personalized E-Learning Recommender System

JIANG Shuhao, JIN Ge

(Dept. of Information Technology, Tianjin University of Commerce, Tianjin 300400, China)

Abstract: The rapid development of e-learning systems provides learners with vast opportunities to access online learning. However, too many learning activities are emerging in an e-learning system, making it difficult for individual learners to select proper activities, thus it is necessary for such recommender systems as to provide personalized recommendations from products in the e-learning system. This paper first proposes a fuzzy tree-structured learning activity model, a fuzzy tree matching-based hybrid learning activity recommendation approach is then developed, which takes advantage of both the knowledge-based and collaborative filtering-based recommendation approaches.

Keywords: E-learning; fuzzy set; recommender systems; tree matching

1 引言(Introduction)

计算机网络技术与多媒体技术的发展, 对传统的教育教学模式有了很大的冲击, 远程教育应运而生^[1]。而在信息过载问题越来越严重的大数据时代, 学习者很难在众多种类的学习活动中选择最适合他们学习活动, 因此电子学习系统必须能够自动生成个性化的建议来指导学习者的学习。目前推荐系统已经广泛应用于商业领域^[2], 在电子学习中应用很少, 因为学习活动不同于电子商务中的商业产品, 它具有很大的不确定性, 涉及特殊的推荐方法和相似性度量的需求。为了处理电子学习推荐系统的特殊要求, 本研究提出了一种基于模糊树匹配的混合推荐方法^[3]。用一种模糊树形结构数据模型来描述学习者档案和学习活动, 应用模糊集技术处理不确定问题, 相似性计算作为推荐方法的核心技术^[4]。该混合推荐方法结合基于知识(KB)和协同过滤(CF)的推荐方法的优点, 以及

考虑学习者之间的语义相似性和协同过滤相似性来生成推荐结果。

2 推荐方法(Recommendation approaches)

目前推荐技术已经备受关注, 并且许多推荐方法也已被提出。最常用的两个推荐方法是: CF(协同过滤)、基于内容的推荐(CB)。CF技术基于其他具有相似兴趣的人提出的建议帮助人们做出选择。它可以进一步分为基于用户和基于项目的CF推荐方法。CB技术给一个特定的用户基于他以前喜欢的东西推荐类似的物品。每个方法都有其局限性, 如CB的项目内容依赖问题和过度专门问题, CF的冷启动和稀疏性问题^[5]。为了获得更高的性能和避免传统推荐算法的缺点, 可以在一个混合方法中结合两个或者更多个推荐方法的优点。

3 电子学习推荐系统(E-Learning recommender systems)

推荐系统已经应用于电子学习领域, Zaiane提出了一种使

用数据挖掘技术的方法,利用关联规则挖掘建模用户行为和推荐的学习活动,并提出了个性化的电子学习材料推荐系统的框架^[6]。我国的个性化推荐技术在电子学习中的应用起步较晚,更多的是对于电子商务中推荐算法的改进,却并没有注意到学习资源与传统电子商务的区别。尽管有很多公司和高校已经开始探索个性化推荐技术在电子学习上的应用,但整体因缺乏国家推动或者别的原因,发展速度比较缓慢。取得效果比较明显的是上海交通大学开发的系统,该系统采用基于项目的协同过滤算法进行个性化推荐。北京大学也正在进行“共享北大”计划,也是典型的系统^[7]。

4 模糊树状结构数据模型和树匹配方法(Fuzzy tree structured data model and tree matching method)

4.1 模糊树形结构数据模型

定义1: 树形结构数据模型是一棵树,以下特性被添加到树节点。

(1)对于一组属性 $A=\{a_1, a_2, \dots, a_n\}$ 。 $a_i \in A$ 代表一个节点的语义含义的一个方面。相应的一组值域 $D=\{d_1, d_2, \dots, d_n\}$ 。对于属性 a_i ,取赋值函数 $a_i: V \rightarrow d_i$,这样每个节点可以为它的属性分配值。

(2)一组相似度量 $S=\{S_1, S_2, \dots, S_m\}$ 定义给节点属性从不同的角度来评估不同的节点之间的相似性。每个相似性测量如果被定义为一个函数 $S_i: \Delta \times \Delta \rightarrow \Delta[0, 1] \in 2D$ 根据特定的应用程序可以指定 Δ 。

(3)一个加权函数 $w: V \rightarrow [0, 1]$ 为每个节点分配一个权值来表示其对于兄弟姐妹的重要性程度。

定义2: 模糊树形结构的数据模型表示的是一个树结构数据,它的节点属性问题,节点之间相似度,或节点的权重都用模糊集表示。

在下面几节中,树木和节点使用以下符号表示。假设我们每棵树有一个编号。让 $t[i]$ 是树 T 的第 i 个节点。 $T[i]$ 是以 $t(i)$ 为根节点的子树, $F(i)$ 是从 $T[i]$ 删除 $t[i]$ 获得的无序森林。让 $t[i_1], t[i_2], \dots, t[i_{n_i}]$ 为 $t[i]$ 的孩子。

4.2 树形结构数据匹配方法

本节基于之前的研究总结树形结构数据匹配方法^[8-10]。构造最大概念相似性树映射来识别两棵树最相似的部分,当比较两棵树,一种是两棵树的权重都应该考虑,另一个是一个子树匹配目标树找出目标树是否包括子树,此时子树的权重应该被主要考虑。因此,树匹配分为对称匹配和不对称匹配。分别定义为 SCT_{sym} 和 SCT_{asym} 。

两棵树 $T1[i]$ 和 $T2[j]$ 比较,它们的概念相似度计算如公式(1)。

$$scT(T1[i], T2[j]) = \begin{cases} sc(t1[i], t2[j]), & F1[i]=\varphi, F2[j]=\varphi \\ \alpha \cdot sc(t1[i], t2[j]) + (1-\alpha) \sum_{t=1}^{n_i} wjt \cdot scT(T1[i], T2[t]), & F1[i]=\varphi, F2[j] \neq \varphi \\ \alpha \cdot sc(t1[i], t2[j]) + (1-\alpha) \sum_{t=1}^{n_j} wjt \cdot scT(T1[t], T2[j]), & F1[i] \neq \varphi, F2[j]=\varphi \\ \alpha \cdot sc(t1[i], t2[j]) + (1-\alpha) \cdot scF(F1[i], F2[j]), & F1[i] \neq \varphi, F2[j] \neq \varphi \end{cases} \quad (1)$$

正如对树结构数据定义的讨论,树节点概念之间的相似性度量 $SC()$ 如方程(1)所示。 wjt 和 wit 分别是 $t2[t]$ 和 $t1[it]$ 的归一化权重。 α 是父节点的影响因素。为了从两个森林的根中找到最对应的节点对,应用了maximum weighted bipartite matching(MWBM)。在计算两棵树之间概念相似度的过程中,记录了MWBM的结果。相应的两棵树的根是节点对。两个根的孩子相应的节点基于两根的孩子的MWBM被定义。

5 学习活动的模糊树状结构(Fuzzy tree structured learning activities)

模糊类别树和模糊类别相似

(1)模糊类别树: 电子学习系统中介绍了学习活动类别来区分学习活动。本系统中学习活动类别有六大类,每个大类别分为几个子类别如图1所示。

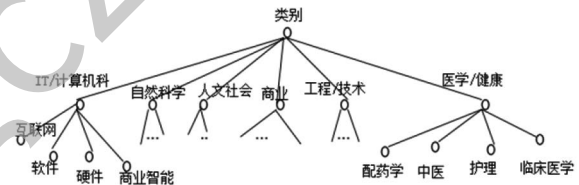


图1 学习活动类别树

Fig.1 Learning activity category tree

在真实的应用程序中,每个学习活动可能不同程度的属于几类,如图3所示,图中每个子类下面的数字表示学科对于子类隶属度。子目录和相应的隶属度由学习活动提供者插入到系统时提供。两个模糊分类树如图2所示。

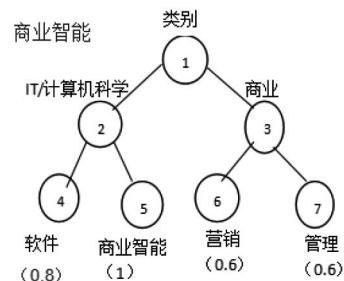


图2 商业智能学科的模糊类别树

Fig.2 Fuzzy category tree of the subject business intelligence

(2)模糊类别相似性: 模糊类别的树都是基于图1所示的学习活动类别树。学习活动类别树的编号用于表示树节点。让 $T1[i]$ 和 $T2[i]$ 分别代表两个学习活动 a_1 和 a_2 的模糊类别树。为

了评估两个模糊类别树的相似度,所有节点的值都必须考虑。

根据T1[i]和T2[i]的孩子是否为0,公式中给出了四种情况,a1和a2的模糊类别相似计算如公式2所示。

$$sfc(T1[i],T2[i]) = \begin{cases} vc(t1[i]) \wedge vc(t2[i]), & vc(F1[i])=0, vc(F2[i])=0 \\ vc(t1[i]) \wedge vc(T2[i]), & vc(F1[i])=0, v(F2[i]) \neq 0 \\ vc(T1[i]) \wedge vc(t2[i]), & vc(F1[i]) \neq 0, vc(F2[i])=0 \\ (ah-di-ah) \cdot (vc(T1[i]) \wedge vc(T2[i])) + (1-ah-di+ah) \cdot (\bigvee_{j=1}^{n_i} Sfc(T1[ij],T2[ij])), & vc(F1[i]) \neq 0, vc(F2[i]) \neq 0 \end{cases} \quad (2)$$

6 学习者档案模糊树形结构(Fuzzy tree structured learner profiles)

当一个学习者选择学习活动时,有许多影响学习者决定的因素,系统做推荐时这些因素都要被考虑到。每个方面又包括几个子方面,因此够构造了学习者档案模糊树如图3所示。

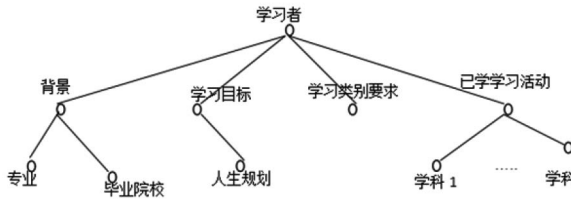


图3 学习者档案结构

Fig.3 Structure of a learner profile

我们的系统定义了一个语言学术语集 $R=\{\text{非常低要求(VLR)}, \text{低要求(LR)}, \text{中等要求(MR)}, \text{高要求(HR)}, \text{非常高要求(VHR)}\}$ 用于学习者表达他们的要求。运用模糊集技术处理这些术语。这些语言相关的模糊数详见表1。

表1 学习者要求模糊数表

Tab.1 Table of fuzzy numbers for learner requirements

VLR	LR	MR	HR	VHR
(0,0,0.25)	(0,0.25,0.5)	(0.25,0.5,0.75)	(0.5,0.75,1)	(0.75,1,1)

6.1 有关模糊要求分类树的相似性度量

(1)模糊要求类别相似:

Tr1和Tr2分别代表两个模糊要求类别树,Tr1和Tr2的模糊要求类别相似性计算如公式(3)。

$$sfc(Tr1[i],Tr2[i]) = \begin{cases} sv(vr(tr1[i]),vr(tr2[i])), & vr(Fr1[i])=0, vr(Fr2[i])=0 \\ sv(vr(tr1[i]),vr(Tr2[i])), & vr(Fr1[i])=0, vr(Fr2[i]) \neq 0 \\ sv(vr(Tr1[i]),vr(Tr2[i])), & vr(Fr1[i]) \neq 0, vr(Fr2[i])=0 \\ (ah-di-ah) \cdot sv(vr(Tr1[i]),vr(Tr2[i])) + (1-ah-di+ah) \cdot \sum_{j=1}^{n_i} wjsfc(Tr1[ij],Tr2[ij]), & vr(Fr1[i]) \neq 0, vr(Fr2[i]) \neq 0 \end{cases} \quad (3)$$

6.2 模糊类别相似度匹配

Tr是学习者的模糊要求类别树,Tc是学习活动的模糊类别树,Tr和Tc的模糊相似性匹配计算如公式(4)。

$$sfc(Tr[i],Tc[i]) = \begin{cases} sv(vr(tr[i]),vc(tc[i])), & vr(Fr[i])=0, vc(Fc[i])=0 \\ sv(vr(tr[i]),vc(Tc[i])), & vr(Fr[i])=0, vc(Fc[i]) \neq 0 \\ sv(vr(Tr[i]),vc(tc[i])), & vr(Fr[i]) \neq 0, vc(Fc[i])=0 \\ (a-h-di-a-h) \cdot sv(vr(Tr[i]),vc(Tc[i])) + (1-a-h-di+a-h) \cdot \sum_{j=1}^{n_i} w(tr[ij]) \cdot sfc(Tr[ij],Tc[ij]), & vr(Fr[i]) \neq 0, vc(Fc[i]) \neq 0 \end{cases} \quad (4)$$

7 基于混合推荐方法的模糊树匹配方法(Fuzzy tree matching based on hybrid recommendation approaches)

步骤1: 确定推荐方案

从两个方面确定与学习者ut有关的学习类别: ut和其他有相同学习目标的学者已经学过的科目; ut的模糊要求类别树Tfrc。gt是ut的学习目标。学习目标是gt的学习者构成一个组Ugt。每一个学习者 $U_i \in U_{gt}$ 的学习活动为 $\{ai,1,ai,2,\dots,ai,ni\}$,相应的模糊类别树为 $\{Ti,1,Ti,2,\dots,Ti,ni\}$ 。结合Ugt中所有用户的学习类别树,以gt为学习目标的学习类别树为Tgt。

$Tcr = combine(\{Tfrc, Tgt\})$ 。对于任何学习活动a的模糊类别树为Tca,如果 $sfc(Tca, Tcr) > 0$,则a为预选推荐的活动。

预选学习活动时教育约束条件也要考虑。学习者Ut的档案树为Tt,Tt中代表学习活动的子树计为Tt,l。学习活动的子树为 $\{at,1,at,2,\dots,at,nt\}$,对学习活动a的学习序列和先决约束条件分别进行验证。对于序列条件,如果 $(a-at,i) \in Sprior, 1 \leq i \leq nt$,则a不适合被推荐。对于前提条件,让学习活动的前提子树表示为Ta,p。通过这一步,选择一套推荐方案,对于每一个选择学习活动a,用以下措施来预测其评级。

步骤2: 计算学习活动a和学习者要求的匹配度

学习者Ut的模糊要求类别树为Treq,学习活动的模糊类别树为Tca。两者的匹配度计算如公式(5)。

$$sm(ut,a) = sfc(Treq, Tca). \quad (5)$$

步骤3: 计算用户之间的语义相似性

学过学习活动a的用户记为 $U_a = \{u1, u2, \dots, um\}$ 。对于每个用户 $ui \in U_a$ 的档案树为Ti。计算ut和ui之间的语义相似性如公式(6)。

$$Ssem(ut,ui) = SCTsym(Tt,Ti). \quad (6)$$

在计算 $Ssem(ut, ui)$ 时, 在 ut 和 ui 的档案树之间构造一个最大概念相似性树映射。匹配最相似的学习活动, 被匹配到的学习活动记在 Mt, i , 对于 $(p, q) \in Mt, i$, p 和 q 分别为 ut 和 ui 学过的学习活动。

步骤4: 计算用户之间的CF相似性

预定义学习活动的相似性阈值为 ast , 对于任何学习活动对 (p, q) , Tp 和 Tq 为相应的树, 如果 $Ssem(p, q) = scTsym(Tp, Tq)$ 小于 ast , 则 p 和 q 无关。 Mt, i 的子集 $M' t, i = \{(p, q) : (p, q) \in Mt, i, Ssem(p, q) > ast\}$, 基于 $M' t, i$, ut 和 ui 的CF相似性计算如公式(7)。

$$sCF(ut, ui) = \frac{\sum_{(p, q) \in M' t, i} r_{t, p} \times r_{t, q}}{\sqrt{\sum_{(p, q) \in M' t, i} r_{t, p}^2} \times \sqrt{\sum_{(p, q) \in M' t, i} r_{t, q}^2}} \quad (7)$$

步骤5: 选择前 n 个相同的用户

基于上两个步骤计算总相似性如公式(8)。

$$su(ut, ui) = \beta \times Ssem(ut, ui) + (1 - \beta) \times sCF(ut, ui) \quad (8)$$

$\beta \in [0, 1]$ 是一个在组合计算中指定相似权重的语义结合系数, ua 中的用户根据总相似性排序, 头 n 个最相似的用户被选为邻居用来预测评级。

步骤6: 计算预测评级

用户 ut 的学习活动 a 的预测评级计算如公式(9)。

$$prt, a = \theta \times sm(ut, a) \times rmax + (1 - \theta) \times \frac{\sum_{i=1}^N r_{i, a} \times su(ut, ui)}{\sum_{i=1}^N su(ut, ui)} \quad (9)$$

$\theta \in [0, 1]$, $rmax$ 是评级的最大值, 方程包括两部分。 $sm(ut, a) \times rmax$ 是基于预测评级的要求匹配, 如果目标学习活动完全匹配用户的要求, 则达到最高级。 $ri, aSu(ut, ui) / Su(ut, ui)$ 是传统的基于项目CF预测评级。 θ 是一个结合两个部分的参数。

步骤7: 生成的建议

计算 ut 所有供选择的学习活动的预测评级并排序。前 K 名适合推荐给用户。

8 结论(Conclusion)

本文概述了电子学习系统中基于混合推荐的模糊树匹配方法的发展。这个方法构建了学习活动和学子这档案的模糊树模型。运用模糊树相似性测量计算学习活动或学习者之间的相似性。在学习活动模糊树形结构模型中, 定义了模糊类别树指定每个学习活动大致属于的类别, 模糊类别相似性测

量来计算学习活动之间的语义相似性。通过分析处理学习序列和建模前提学习活动确定学习活动之间的优先级关系, 学习者通过模糊要求类别树表达他们的需求。推荐方法结合了CF和KB的推荐方法的优点。就寻找相似的学习者来说, 这套系统结合了语义相似和CF相似的优点。在计算CF相似时, 利用的是被匹配的学习活动的评级, 而不是用户使用的两个常见完的学习活动, 减轻稀疏的问题。

参考文献(References)

- [1] 赵蔚, 余延东, 张赛男. 基于数据挖掘的个性化E-Learning解决方案推荐系统研究[J]. 现代远距离教育, 2011(4): 60-64.
- [2] D. Wu, G. Zhang, J. Lu. A fuzzy preference tree-based recommender system for personalized business-to-business e-services[A]. 2015: 29-43.
- [3] J. Lu, Q. Shambour, Y. Xu, et al. A web-based personalized business partner recommendation system using fuzzy semantic techniques, Comput[A]. Intell, 2013: 37-69.
- [4] Z. Zhang, J. Lu. A hybrid fuzzy-based personalized recommender system for telecom products/services[A]. 2013: 117-129.
- [5] 陈洁敏, 汤勇, 李建国, 等. 个性化推荐算法研究[J]. 华南师范大学学报(自然科学版), 2014(5): 9-13.
- [6] O. R. Zaiane. Building a recommender agent for e-learning systems in Proc[A]. Int. Conf. Comput Educ. 2002: 55-59.
- [7] 王凯, 陈建. 支持个性化学习资源推荐的在线辅助学习系统的研究与设计[D]. 陕西师范大学, 2014: 5-6.
- [8] D. Wu, G. Zhang, J. Lu, W. A. Halang. A similarity measure on treestructured business data, in Proc[A]. 23rd Australas. Conf. Inf. Syst., 2012: 1-10.
- [9] D. Wu, G. Zhang, J. Lu. A fuzzy tree similarity measure and its application in telecom product recommendation in Proc[A]. IEEE Int. Conf. Syst. U. K., 2013: 3483-3488.
- [10] D. Wu, G. Zhang, and J. Lu. A fuzzy tree similarity based recommendation approach for telecom products in Proc[A]. Joint IFSA World Congr. NAFIPS Annu. Meet, Edmonton, Canada, 2013: 813-818.

作者简介:

姜书浩(1980-), 男, 硕士, 副教授. 研究领域: 电子商务.

金 格(1994-), 女, 本科生. 研究领域: 电子商务.