结合社会网络分析和多维特征 聚类的学习小组划分方法*

张红宇1,王坚强1,马 华2

(1. 中南大学商学院, 长沙 410083; 2. 湖南涉外经济学院 信息科学与工程学院, 长沙 410205)

摘 要:通过引入社会网络分析方法,建立了综合学习能力、学习状态和协作能力多维指标的学习者特征模型,应用模糊相似和模糊等价演算,以多层次的模糊动态聚类方法识别候选学习领袖,通过融合多类型学习者实现学习小组的迭代划分。实验分析表明,该方法可提高虚拟学习社区中学习小组划分结果的准确度,具有良好的自适应性。

关键词:虚拟学习社区;学习小组;划分;社会网络分析;特征聚类

中图分类号: TP391; C420 文献标志码: A 文章编号: 1001-3695(2013)03-0732-04 doi:10.3969/j.issn.1001-3695.2013.03.022

Grouping approach of learning team based on social network analysis and multidimensional feature clustering

ZHANG Hong-yu¹, WANG Jian-qiang¹, MA Hua²

(1. Business School of Central South University, Changsha 410083, China; 2. School of Information Science & Engineering, Hunan International Economics University, Changsha 410205, China)

Abstract: This paper presented a learner feature model considering learning capacity, learning status and cooperation ability by introducing social network analysis. Employing the fuzzy similarity and fuzzy equivalent operations, it put forward a grouping approach of learning team, in which learning leader candidate were identified by using dynamic multi-level fuzzy clustering method, and iterative grouping could be achieved compromising learners of different levels. Finally, the experiment proves that this approach can enhance accuracy of learning team grouping with self-adaptability.

Key words: virtual learning community; learning team; grouping; social network analysis; feature clustering

0 引言

在虚拟学习社区中构建合适的学习小组,有利于实现协作学习、交互学习,对于改善学习过程、提高学习效果具有重要意义^[1]。由教师指派或学习者自行选择组成的学习小组,通常难以保障建立良好的协作关系^[2]。

现有的文献主要从学习伙伴的角度入手开展了一系列研究。Greer等人^[3]介绍了 Helpdesk 系统中对等计算环境下基于面向主题的论坛讨论形式发现合适助学者的方法。Chen 等人^[2]针对基于问题学习(problem-based learning, PBL)环境通过挖掘学习交互社会网络,提出一种推荐准确的学习伙伴的方法。程向荣等人^[4]应用模糊隶属度建模学习者的动态协作信息,将其作为选择伙伴的主要依据。以上研究的重点是通过考察学习进度、认知能力和协作能力,为学习者个体寻找与之相匹配的协作学习对象。然而,学习小组的构建往往比学习伙伴的寻找更加复杂,其关注的重点是通过协作实现学习者群体在知识水平、学习积极性、能力水平上的提升^[5]。

因此,如何选定一组有效的特征指标,将学习者划分为不同层次的学习类型,进而组建融合多层次学习者的学习小组,

是当前虚拟学习社区的研究热点。Chen 等人^[6]以社会计量法(sociometry)衡量学生的社会声望,并对混合学习成绩、社会声望等约束下的学习小组应用遗传算法进行优化。Wang 等人^[7]在问卷调查基础上依据思维模式对学习者进行分类,并采用爬山算法和模拟退火算法实现各类别全覆盖条件下的分组优化。这些方法受学习者主观判断限制,降低了分组结果的准确度,并且缺乏自适应性。

近年来,相关学者以博客^[8]、论坛^[9]、QQ 群^[10]的社会网络分析为例,研究了虚拟学习社群中网络学习成效的评价方法,以及改善学习过程的指导建议。社会计算理论的引入为虚拟学习社区中学习小组的研究提供了新的思路。基于现有研究,本文围绕学习能力、学习状态和协作能力建立学习者特征模型,结合社会网络分析(social network analysis, SNA)方法给出了学习小组模型,该模型以虚拟学习社区中社会网络结构的量化分析为依据,探讨了点度中心度、接近中间度和间距中间度与学习者的学习状态和协作能力的关联与关系,通过融合学习者的学习能力数据,实现虚拟学习社区中学习者的特征化模糊聚类,并在其基础上进行学习小组的迭代划分。实验分析表明,本文方法可显著提高分组的准确性和自适应能力,从而确

收稿日期: 2012-07-11; **修回日期**: 2012-08-17 **基金项目**: 国家自然科学基金资助项目(70771115); 国家教育部人文社会科学研究青年基金资助项目(11YJCZH227); 湖南省教育科学"十二五"规划课题(XJK011QXJ002)

作者简介: 张红宇(1979-), 女, 安徽亳州人, 副教授, 博士, CCF 会员, 主要研究方向为信息管理和现代教育技术(hyzhang@ csu. edu. cn); 王坚强, 男, 教授, 主要研究方向为模糊决策; 马华, 男, 副教授, 主要研究方向为可信软件和分布式计算.

保通过协作学习实现虚拟学习社区中所有学习者在知识水平、学习积极性和协作能力上的整体提升。

1 虚拟学习社区中的学习者特征模型

学习者特征是对虚拟学习社区中学习者模型^[11]的关键描述。特征模型的完整性、准确性直接关系到虚拟学习社区能否为学习者提供与其相适应的学习策略、内容和资源,以及能否有效提高协作学习效果等。因此准确定义学习者特征模型,是建立虚拟学习社区中学习小组的重要前提之一。

通常,现有的学习者特征模型在建模过程中,不仅关注学习者知识,也会引入非智力因素,如学习风格、自主行为能力、能力倾向、情感态度的性格表现、情绪性格表现等[12]。现实的课程教学实践也是如此,教师将考试成绩与日常教学过程中学生的出勤、课堂表现,以及学生对教师课程教学的贡献等因素相结合,对学习者给出综合评价。因此,本文综合考虑智力和非智力因素,基于社会网络分析方法定义了虚拟学习社区下学习者的特征模型,该模型包含学习能力、学习状态和协作能力三维空间,如图1所示。



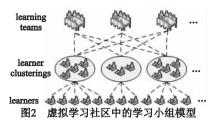
图1 学习者特征模型的三维空间

- 1)学习能力(learning capacity) 反映学习者的学习历史情况、知识状态、认知能力等智力性因素。本文以优秀度衡量学习者的学习能力。虚拟学习社区中学习者的相关课程期评成绩、本课程的起始单元测验成绩等可作为评价参考。
- 2)学习状态(learning status) 表征学习者的学习兴趣度、振奋度等。学习过程中兴趣度、振奋度越高,学习活动越积极,收到较好效果的可能性越大。本文以活跃度衡量学习者的学习状态。虚拟学习社区中,依据 SNA 方法将学习者参加学习讨论、交流活动的发帖、回帖次数作为评价依据。
- 3)协作能力(cooperation ability) 反映学习者在协作学习中体现的合作、沟通、协调能力。协作能力较强的学习者,在团队中具有更强的凝聚力,是潜在的学习领袖(即学习小组组长)。本文以学习影响力衡量学习者的协作能力。在虚拟学习社区中,依据 SNA 方法利用意见领袖评价指标衡量学习者的学习影响力。

虚拟学习社区中,学习领袖是学习小组中的核心人物,是协作学习的实际组织和实施者,对于小组学习的成功具有重要意义。如果基于现有的社会网络分析方法,将意见领袖直接视为学习领袖,将存在一定缺陷。由于在虚拟学习社区中可能存在大量灌水数据和低价值数据,会严重影响 SNA 结果。因此,本文综合学习者特征模型的三个度量指标,即优秀度、活跃度、影响力,将它们作为学习领袖的识别标准,从智力和非智力两个角度人手,以提高识别学习领袖的准确性。

2 虚拟学习社区中的学习小组模型

现有研究表明,在学习小组中融合不同类型学习者有利于 提高学习者群体的学习效果^[7]。依据第1章定义的学习者特 征模型,给出虚拟学习社区中学习小组模型,如图2所示。



该模型涉及学习者、学习者聚类和学习小组三个概念。

- 1)学习者(learner) 虚拟学习社区中进行课程学习的人。 本文为学习者特征模型定义了三个度量指标,即优秀度、活跃 度、影响力。
- 2)学习者聚类(learner clustering) 通过学习者特征模型 对学习者进行模糊综合评价,并对学习者进行聚类分析。将学 习能力、个性特征和协作能力相似的学习者识别出来,尤其是 那些具有最高模糊综合评价值的学习者(称为候选学习领 袖),为后续分组作准备。
- 3)学习小组(learning team) 按照一定的分组策略和算法,从各个聚类中抽取一定比例的学习者,组建学习小组。每个学习小组中应尽可能多地覆盖不同的学习者聚类,且至少应该包含一个学习领袖。

3 学习者特征的社会网络分析

为实现对学习者的聚类分析,先要对学习者特征的度量指标,即优秀度、活跃度、影响力,进行数据提取。其中,优秀度可以直接综合虚拟学习社区中的已有数据加权平均后得到,而活跃度、影响力指标则需应用 SNA 方法获取。

3.1 基于 SNA 的活跃度评价方法

在虚拟学习社区中,如果学习者在论坛和讨论区中发帖、回帖的次数较多,可认为其兴趣度、振奋度较高,表明其学习状态是活跃的,在社会网络分析中,可使用点度(degree centrality)来刻画。一个学习者的点度中心性为点出度和点入度的综合[13]。绝对点出度为该学习者发帖的数目,即

outDegree_i =
$$\sum_{i=1}^{n} Z_{ij}$$

绝对点入度为该学习者回帖的数目,即

$$inDegree_i = \sum_{i=1}^{n} Z_{ji}$$

学习者 i 的点度中心性计算为

$$C_t = \sum_{i=1}^{n} (Z_{ij} + Z_{ji}) / \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ij}$$

通过对虚拟学习社区中的点出度和点入度分析,可了解学习者参与交流和互动的程度。若某个学习者的出度较高,则他可能是积极主动型的学习者、学习气氛的营造者。入度较高的学习者在学习讨论中备受大家的关注,更有可能得到大家的欢迎和认可。若有的学习者出度和入度都很高,则他们可能既积极主动学习,又受到大家的关注,他们对虚拟学习社区的贡献可能较大;若有的学习者出度高而入度较低,则表明他虽积极关注其他学习者,但自己却很少被关注;若有的学习者出度较低但入度很高,则表明他虽很少主动发言,但却能吸引其他学习者的关注。

由此,第 i 个学习者的活跃度综合评价值为

$$A_i = \mathrm{outDegree}_i \times w^o + \mathrm{inDegree}_i \times w^i \tag{1}$$

其中: w° 和 w^{i} 分别为点出度和点入度的权重, w° + w^{i} = 1,通

常 $w^{\circ} > w^{i}$ 。为便于计算,对活跃度按下式进行归一化操作:

$$A_i' = \frac{A_i - A_{\min}}{A_{\max} - A_{\min}} \tag{2}$$

3.2 基于 SNA 的影响力评价方法

运用接近中心度和间距中心度综合评价学习者影响力。

1)接近中心度(closeness centrality)分析

接近中心度是依据社会网络图中学习者间的紧密性或距离进行测量得到的中心度,用于度量网络中学习者的重要性。所测得的距离越短,说明接近中心度越高,与其他行动者越密切^[14]。一个点的接近中心度是该点与网络中所有其他点的捷径距离之和^[15],用公式表示为

$$C_c^{-1} = \sum_{j=1}^{n} d(n_i, n_j)$$

其相对紧密中心度 C_c 满足:

$$C_{c}' = (n-1)C_{c}$$
 (3)

该指标反映了网络中一个行动者在多大程度上不受其他 人影响的程度。社会网络中,一个非核心成员必须通过其他人 才能传播信息,容易受制于其他节点。当他与其他人越接近, 则在传播信息的过程中越不依赖其他人。因此,在虚拟学习社 区中,一个学习者与其他学习者间的距离越短,则接近中心度 越高,在该学习社区中越重要,其影响力越大。

2)间距中心度(betweenness centrality)分析

间距中心度是指社会网络图中某一节点与其他各节点间 的间隔程度^[15]。如果一个学习者处于许多网络路径上,可认 为其处于重要地位。间距中心度由式(4)计算得到。

$$C_B = \sum_{i=1}^{n} g_{jk}(n_i) / g_{jk}$$
 (4)

间距中心度用于刻画一个行动者在多大程度上居于其他两个行动者之间,反映的是其控制网络中其他行动者的能力。在虚拟学习社区中,如果一个学习者具有较高的间距中心度,则可发挥沟通其他学习者的桥梁作用,说明其他学习者间的交流对该学习者的依赖性较大,该学习者对其他学习者很重要。如果一个学习者的间距中心度为0,则该学习者不能控制任何学习者;如果一个学习者的间距中心度为1,则该学习者控制其他学习者的能力强,影响力大。

3)影响力的综合评价与归一化

将第 i 个学习者的影响力综合评价值定义为

$$I_i = C_c' \times w^c + C_B \times w^b \tag{5}$$

其中: w^c 和 w^b 分别为接近中心度和间距中心度的权重, w^c + w^b = 1。为便于计算,对影响力进行归一化操作:

$$I_i' = \frac{I_i - I_{\min}}{I_{\max} - I_{\min}} \tag{6}$$

4 学习者的多维特征聚类

对于学习者优秀度的评价,可参考虚拟学习社区中其他相关课程的期评成绩。如果是首修课程,则将该课程的首个单元测验成绩作为评价参考,使用 E_i 表示第 i 个学习者的优秀度。由此,学习者的多维特征聚类过程如下:

a)结合活跃度、影响力指标,构建由n个学习者组成的学习者特征矩阵如下:

$$L = \begin{bmatrix} L_0 \\ L_2 \\ \vdots \\ L_n \end{bmatrix} = \begin{bmatrix} L_{01} & L_{02} & L_{03} \\ L_{21} & L_{22} & L_{23} \\ \vdots & \vdots & \vdots \\ L_{n1} & L_{n2} & L_{n3} \end{bmatrix} = \begin{bmatrix} E_0 & A_0 & I_0 \\ E_2 & A_2 & I_2 \\ \vdots & \vdots & \vdots \\ E_n & A_n & I_n \end{bmatrix}$$

其中: L_0 代表学习领袖期望指标,与该期望值具有较高相似度的学习者才具备作为学习领袖的基本素质。

b)应用绝对值减数法按式(7)建立模糊相似关系矩阵 R。

$$r_{ij} = \begin{cases} 1 & i = j \\ 1 - c \sum_{k=1}^{m} |L_{ik} - L_{jk}| & i \neq j \end{cases}$$
 (7)

其中:c = 0.1.m = 3。

c)因 R 可能不满足传递性,并非模糊等价关系,需使用平方法^[16]对 R 进行变换运算,即 $R \circ R \to R^*$, $R^* \circ R^* \to R^*$,…,直至 $R^m \circ R^m \to R^2m = R^m \circ R^m$ 即为模糊等价矩阵 R^* 。其中, R 的自乘采用 $M(\Lambda, V)$ 模型计算。

d)设置不同的 λ 值,由 λ 截矩阵 [16] 得到模糊动态聚类结果。设聚类个数为 n^c ,当 n^c > 3 时认为基本可满足聚类需求。若聚类数过少,将不利于识别不同层次的学习者,此时可降低期望指标 L_0 值。 L_0 所属的聚类被称为候选学习领袖类。考虑到学习小组的实际协作可行性,每个小组的人数不宜过多或过少,以 3 ~ 6 人为宜,通常候选学习领袖类中学习者人数约为n/5 人较为合适,通过调节 λ 值可满足该需求。

5 学习小组的迭代划分方法

以一个由 5 名学习者组成的学习小组 T_i 为例,定义该学习小组特征值为

$$F_i = \max(S_{i1}, S_{i2}, \dots, S_{im})$$

其中: $S_{11} \sim S_{15}$ 分别代表各个学习者的特征值与学习领袖期望指标 L_0 的相似度。具有最大相似度值的学习者作为该组的学习领袖,即学习小组长。

分组的目标是使每个学习小组的特征值尽可能大,即使得 所有学习小组中的最小特征值最大化,一个有 *m* 个小组的分 组结果 sol 应满足以下条件:

$$sol = \operatorname{argmax}(\min(F_1, F_2, \dots, F_m))$$
 (8)

为找出合适的小组划分结果 sol,将多维特征聚类得到的 n^c 个学习者聚类表示为 $C = \{c_0, c_1, c_2, \cdots, c_{nc-1}\}$, c_o 为候选学习领袖类, $|c_i|$ 表示聚类 c_i 中的学习者个数。由此,依据学习小组模型,本文给出了一种学习小组的迭代划分方法:

- a) 令待生成的学习小组编号 i=1, num = $\lfloor c_o \rfloor$ 。
- b) 如果 $|c_o| \neq 0$,从中抽取一个与 L_o 具有最大相似度的学习者 l_i ,将其作为学习小组 T_i 的学习领袖,从 c_o 中删除学习者 l_i : 否则,转步骤 d)。
- c)按顺序依次检查聚类 c_1 , c_2 , …, c_{nc-1} , 如果某个聚类非空,则从中抽取一个学习者加入 T_i , 并将其从所属聚类中删除,令 i=i+1,继续步骤 b);如果 c_1 , c_2 , …, c_{nc-1} 均为空,分组完成。
- d) 如果 $c_1, c_2, \cdots, c_{n^c-1}$ 中有任一个不为空,则令 i 复位, i=1
- e) 依次检查 $c_1, c_2, \cdots, c_{ne-1}$, 如果某个聚类非空,则从中抽取一个学习者加入 T_i , 并将其从所属聚类中删除,令 i = (i+1)% num,继续步骤 e); 如果 $c_1, c_2, \cdots, c_{ne-1}$ 均为空,则分组完成。

经以上迭代方法得到的分组方案可满足所有学习小组中最小特征值的最大化要求,该分组方法可有效避免优秀生或落后生扎堆的情况发生,能够为学习小组整体学习效果的提升奠定坚实基础。

6 实验及分析

基于 E-learning 2.0 环境,本研究小组建设了经管类信息系统课程的虚拟学习社区平台[17],该社区遵循 IEEE 1484 标准和 SCORM 模型,利用国际主流的开源 Moodle 作为核心框架,以 Web 2.0 技术为基础,可营造自由、开放、共享、协作的学习环境,支持教师、学习者间以合作方式展开教学活动。以《计算机网络》课程为例,本文选取该虚拟学习社区中某班 20名学习者(用代号1、2、3、…、20表示)为研究对象,采集 2011年9月至12月间的学习交流数据,导入到 UCINET工具进行社会网络分析。UCINET(University of California at Irvine Network)是一种功能强大的社会网络分析软件[15]。导入 UCINET后的发帖、回帖交互数据如图 3 所示。

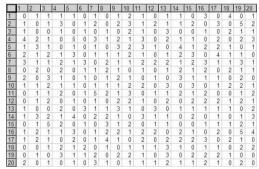


图3 虚拟学习社区中的学习者交互数据

根据图 3,使用 UCINET 软件可得该虚拟学习社区的社群 关系图,如图 4 所示。

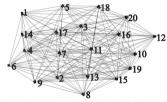


图4 虚拟学习社区社群图

由图 4 可知,该虚拟学习社区的学习者之间交互密切,大家参与学习讨论的积极性普遍较高,尤其是学习者 2、4、5、7、16、19等。

由式(1)计算学习者的活跃度,其中,令 $w^{o}=0.6$ 、 $w^{i}=0.4$;由式(5)计算学习者的影响力,其中,令 $w^{c}=0.5$ 、 $w^{b}=0.5$ 。根据学习者的学习成绩计算得到优秀度,从而得到如表1所示的学习者特征数据。

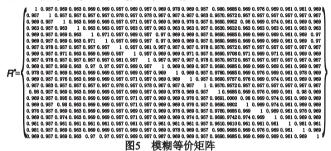
表1 学习者特征数据

代号	优秀度	活跃度	影响力	代号	优秀度	活跃度	影响力
1	0.73	0.2778	0.0939	11	0.37	0.6111	1.0000
2	0.31	0.6111	0.7686	12	0.75	0.2223	0.0000
3	0.69	0.3056	1.0000	13	0.78	0.5834	0.7662
4	0.43	0.7778	0.6464	14	0.31	0.6111	0.8168
5	0.42	0.6112	0.5324	15	0.31	0.5278	0.6538
6	0.7	0.3334	0.2598	16	0.95	0.7778	0.7081
7	0.53	0.8056	1.0000	17	0.6	0.5834	1.0000
8	0.71	0.4167	0.1598	18	0.69	0.2500	0.5941
9	0.57	0.4723	0.5460	19	0.68	0.7222	0.6538
10	0.68	0.6667	0.8168	20	0.53	0.4445	0.4880

由表1可知,根据学习交流的讨论参与情况,学习者7积极参与讨论,发帖和回帖较多,其活跃度值最高。根据传统的SNA方法。一般会得出该学习者学习积极的结论。但经过深入了解该学习者发帖和回帖情况后发现,该生发帖和回帖所传

递的学习信息非常有限,帖子内容与学习关系并不密切,多为 "水帖",他的帖子也因此常常引起同学"围观"。实际上,其学 习并不理想,如表1所示。

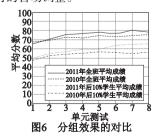
本文综合三维学习者特征数据对学习者进行聚类,将可消除灌水影响,识别出真正的学习领袖,提高分组准确度,从而显著改善分组质量。实验中定义了学习领袖期望指标 $L_0=\{0.80,0.70,0.70\}$,由此计算得到模糊等价矩阵 R^* 如图 5 所示。当 $\lambda=0.97$ 时,可分为 7 类,即 $c_o=\{l_{10},\ l_{13},\ l_{16},\ l_{19}\}$ 、 $c_1=\{l_2,\ l_7,\ l_{11},\ l_{14},\ l_{15},\ l_{17}\}$ 、 $c_2=\{l_1,\ l_6,\ l_8,\ l_{12}\}$ 、 $c_3=\{l_3\}$ 、 $c_4=\{l_4,\ l_5\}$ 、 $c_5=\{l_9,\ l_{20}\}$ 、 $c_6=\{l_{18}\}$ 。



因此,根据分组优化算法得到以下四个学习小组: $T_1 = \{l_{10}, l_2, l_{15}, l_8, l_5\}$ 、 $T_2 = \{l_{13}, l_7, l_{17}, l_{12}, l_9\}$ 、 $T_3 = \{l_{16}, l_{11}, l_1, l_3, l_{20}\}$ 、 $T_4 = \{l_{19}, l_{14}, l_6, l_4, l_{18}\}$ 。

由分组结果可知,具有较高优秀度、活跃度和影响力的学习领袖被均匀地划分配到不同的学习小组中,在实际教学实践过程中,他们确实能够较好地完成小组内协作学习的各项任务,对实现小组内所有学习者在知识水平、学习积极性、协作能力上的整体提升起到了积极作用。以 2011 年和 2010 年同一课程不同班级的学习者在同一虚拟学习社区环境下的学习效果进行分析,2010 年的学习小组划分采取学生自愿组合方式构建,2011 年的学习小组划分采用本文方法组建。通过统计两个班级全班学生的单元测试平均成绩和成绩排名在 10%的学生的单元测试平均成绩,得到的数据对比如图 6 所示。由于自愿组合学习小组可能导致学习尖子或学习困难生过于集中,不利于提高整体学习效果。而采用本文的分组方法,班级平均成绩有了明显提升,尤其对于基础较差的学生,在良好的小组学习氛围促进下,学习兴趣得到增加,学习质量得到改善。

此外,由于聚类和分组依据来源于虚拟学习社区中的论坛数据,本文方法可较客观地反映学习者的综合情况,避免了在问卷调查中因主观因素影响造成的偏差。同时,本文的多维特征聚类方法伸缩性好,可以适应不同规模学习者群体的分组需求,也可以动态支持如学习者退出或加入新的虚拟学习社区等意外情况出现时的自动调整。



7 结束语

学习小组的构建,有利于协作学习、交互学习活动的开展,构建合适的学习小组,对于提高学习效果具有重要意义。本文引入社会网络分析方法建立了综合学习能力、(下转第741页)

领域的且为网络评论,观点表达的结构复杂且用词变化大,所以情感词典的命中率往往不会特别高,导致分类精度也不会特别高。而唐慧丰等人^[9]提出的利用有监督机器学习模型 SVM 进行情感分类的方法与本文提出的第二种方法类似,由于其数据主要为单一产品领域,其分类精度能达到 90% 以上;而另一方面,对于本文所研究的跨领域评论文本在用有监督机器学习方法进行训练时比单一领域需要更大的语料库,但是由于本文的训练集较小,也同样导致了分类精度不高。

3 结束语

本文在对情感分类常用算法进行综述及确定跨领域评论 观点表达的特殊性的基础上提出了基于词典资源和有监督机 器学习这两种方法来对跨领域中文评论句的情感进行分类。 基于词典资源的方法为利用知网评价词词典和《同义词词 林》,并结合转折词和否定词确定观点表达的极性,同时还根 据语料的特征确定了褒义优先的原则,对知网评价词词典进行 了简单的修正并进行了实验以证明情感词典在情感分类中的 重要作用。基于有监督机器学习方法则采用所有词、四种词性 (名词、动词、形容词和副词)和补充词性这三种文本表示方 法,TFI-DF 和布尔函数这两种权重计算函数,IG 和 CHI 这两 种文本特征选择方法以及 SVM 和 NB 这两种文本分类模型进 行对比实验。从实验结果可以看到,补充词性的文本表示方法 起到了缩小数据集的作用,并且其性能与采用所有词进行特征 表示相当,同时布尔函数适合作为跨领域评论文本分类的权重 计算函数。在文本特征选择上,CHI 比 IG 在跨领域数据上相 对更有优势,而 SVM 比 NB 在跨领域评论句观点表达的情感 分类上具有明显的优势。最后对两种方法进行了比较。

参考文献:

[1] HATZIVASSILOGLOU V, McKEOWN K R. Predicting the semantic

(上接第735页)学习状态和协作能力多维指标的学习者特征模型,并以优秀度、活跃度和影响力对其进行量化分析,应用模糊相似关系演算和模糊等价演算,对学习者进行多层次的模糊动态聚类,识别候选学习领袖,并在其基础上通过融合多类型学习者,以迭代方式完成学习小组的划分。实验分析表明,本文方法可显著提高虚拟学习社区中学习小组划分的准确性和自适应性,有利于所有学习者在知识水平、学习积极性、协作能力上的整体提升。

参考文献:

- [1] 袁磊,黄道鸣. CSCL 环境中的社会交互[J]. 现代教育技术, 2004,14(4):46-49.
- [2] CHEN C, CHANG C. Mining learning social networks for cooperative learning with appropriate learning partners in a problem-based learning environment [J]. Interactive Learning Environments, 2012, 20(1): 1-28.
- [3] GREER J, McCALLA G, COOKE J, et al. The intelligent helpdesk: supporting peer-help in a university course [EB/OL]. [2012-07-03]. http://julita.usask.ca/Texte/Jim-html/I-Help.htm.
- [4] 程向荣,周竹荣,邓小清. 计算机支持的协作学习的伙伴模型 [J]. 计算机应用,2007,27(7):1763-1766.
- [5] 程艳, 许维胜, 何一文. 虚拟学习社区的绩效评估模型[J]. 计算机工程与应用,2011,47(4):17-21.
- [6] CHEN Rong-chang, CHEN Shih-ying, FAN Jyun-you, et al. Grouping partners for cooperative learning using genetic algorithm and so-

- orientation of adjectives [C]//Proc of the 8th Conference of European Chapter of the ACL. 1997:174-181.
- [2] TURNEY P D. Mining the Web for synonyms; PMI-IR versus LSA on TOEFL[C]//Proc of the 12th European Conference on Machine Learning, 2001;491-502.
- [3] TURNEY P D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews [C]//Proc of the 40th Annual Meeting Association for Computational Linguistics. 2002;417-424.
- [4] HU Ming-qing, LIU Bing. Mining and summarizing customer reviews [C]//Proc of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2004;168-177.
- [5] 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算 [J]. 中文信息学报, 2006, 20(1):14-20.
- [6] 刘群,李素建.基于《知网》的词汇语义相似度的计算[C]//第三届汉语词汇语义学研讨会.2002.
- [7] 孟凡博,蔡莲红,陈斌,等. 文本褒贬倾向判定系统的研究[J]. 小型微型计算机系统,2008,7(7):1458-1461.
- [8] PANG Bo, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques [C]//Proc of ACL Conference on EMNLP. 2002:79-86.
- [9] 唐慧丰,谭松波,程学旗.基于监督学习的中文情感分类技术比较研究[J].中文信息学报,2007,21(6):55-94.
- [10] HASSAN A, RADEV D. Identifying text polarity using random walks [C]//Proc of the 48th Annual Meeting of the ACL. 2010;395-403.
- [11] 《同义词词林》扩展版[EB/OL]. http://www.ir-lab.org/.
- [12] COAE[EB/OL]. http://www.ir-china.org.cn/information.html/.
- [13] LIBSVM[EB/OL]. http://www.csie.ntu.edu.tw/~cjlin/libsvm/.
- [14] KEERTHI S S, LIN C J. Asymptotic behaviors of supportvector machines with Gaussian kernel [J]. Neural Computation, 2003, 15 (7):1667-1689.
 - cial network analysis [J]. Procedia Engineering, 2012, 29 (1):
- [7] WANG Dai-yi, LIU Yen-chun, SUN Chuen-tsai. A grouping system used to form teams full of thinking styles for highly debating [C]// Proc of the 10th WSEAS International Conference on Systems. 2006: 721-726.
- [8] 黄伟. 社会网络分析视角下的虚拟学习社群研究[J]. 电化教育研究,2011(12):53-58.
- [9] 赖文华,叶新东.虚拟学习社区中知识共享的社会网络分析 [J].现代教育技术,2010,20(10):97-101.
- [10] 张豪锋,李瑞萍,李名. QQ 虚拟学习社群的社会网络分析[J]. 现代教育技术,2009,19(12):80-84.
- [11] 张舸,周东岱,葛情情. 自适应学习系统中学习者特征模型及建模方法述评[J]. 现代教育技术,2012,22(5):77-82.
- [12] 王万森, 龚文. E-Learning 中情绪认知个性化学生模型的研究 [J]. 计算机应用研究,2011,28(11):4174-4177.
- [13] 陈淑洁, 叶新东, 邹文才. 社会网络分析在网络课程评价中的应用研究[J]. 现代教育技术,2009,19(3): 26-30.
- [14] 林聚任. 社会网络分析:理论、方法与应用[M]. 北京: 北京师范 大学出版社,2009.
- [15] 刘军. 整体网分析讲义——UCINET 软件实用指南[M]. 上海: 北京师范大学出版社,2009.
- [16] 李士勇. 工程模糊数学及应用[M]. 哈尔滨: 哈尔滨工业大学出版社.2004.
- [17] 张红宇, 王坚强, 高阳. E-learning 2.0 环境下高校教学资源的构建及应用[J]. 计算机教育,2012(14):52-55,59.