

教学反思内容自动评估模型研究

邱欢堂¹, 何聚厚^{1,2}, 何秀青¹

(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 教师专业能力发展中心, 陕西 西安 710062)

摘要: 教学反思是教师专业能力发展的重要途径, 对反思内容进行自动评估是网络环境下教学反思系统亟待解决的关键问题。对网络环境下提交的反思文本进行中文分词、停用词过滤等预处理, 采用向量空间模型构建反思文本向量, 基于余弦理论计算反思文本与语料库文本的相似度。根据最大相似度语料库文本的等级及系统预设阈值, 实现反思内容的自动评估。实验结果表明, 自动评估结果和专家认定的评估结果相比, 正确率达到90%以上, 基本实现了反思内容的自动评估。

关键词: 教学反思; 自动评估; 向量空间模型; 文本相似度

中图分类号: TP39

文献标识码: A

文章编号: 1673-629X(2012)09-0173-04

Automatic Assessment Model for Content of Teaching Reflection

QIU Huan-tang¹, HE Ju-hou^{1,2}, HE Xiu-qing¹

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

2. Center for Teacher Professional Ability Development, Shaanxi Normal University, Xi'an 710062, China)

Abstract: Self-reflection is an effective way in teachers' professional development. How to evaluate the content of self-reflection automatically is a critical issue in the web-based software system. The content of self-reflection submitted under web environment is pre-processed such as Chinese word segmentation, stop words filtering and so on, build self-reflection text vector with vector space model, calculate the similarity between self-reflection text and corpus text based on cosine theory. According to the level of corpus text with maximum similarity and previous threshold of system, achieve the automatic evaluation of self-reflection content. The experiment shows that the correct rate of the automatic evaluation result is more than 90% compared with the expert evaluation result, implement the automatic assessment of self-reflection content well.

Key words: self-reflection; automatic evaluation; vector space model; text similarity

0 引言

在教学过程中, 教师对教学课堂的课后反思是促进教师专业能力发展的重要手段之一, 培养反思型教师也已经成为教师教育的重要目标之一^[1]。随着计算机网络技术的发展, 基于互联网通过远程指导和反思成为在职教师有效提高其专业能力的有效手段^[1-5]。CLANDRA等^[2]讨论了通过观看教学视频的方法, 帮助实训者进行教学反思, 解决教学中的相关问题, 如: 教学内容和教学活动的选择、学习的多样性和课堂互动的促进、学习资源的管理、教学策略的选择等。SO等^[3]进一步提出了基于Web的视频系统, 使实训者以

便捷、直观的方式通过网络观看视频进行教学反思。KONG等^[4]则开发了相应的视频系统, 实现基于网络观看视频并且根据系统向导进行教学反思。这些研究一般都是指导者根据实训人员的反思内容进行人工评估, 而当实训者数量比较多时, 指导者则无法有效实现一一指导。

国内学者关于如何有效利用信息化手段进行教学反思方面的研究, 更多的集中在如何利用Blog或BBS等方式撰写教学反思日志, 或通过这些方法进行反思交流、讨论等, 以达到教学反思的网络化^[5]。

文中提出的教学反思内容自动评估模型基于向量空间模型和中文分词技术, 采用自主学习构建评估内容语料库, 通过计算实训者的反思内容与语料库中已分级文本的相似度, 实现对反思内容的自动评估。

1 系统结构

教学反思自动评估模型, 通过基于Web的教学反思平台收集实训者的教学反思内容, 对学员的教学反

收稿日期: 2012-02-04; 修回日期: 2012-05-08

基金项目: 国家985优势学科"教师教育创新平台"项目(GJ9850104); 中央高校基本科研业务费专项资金(GK201002028)

作者简介: 邱欢堂(1981-), 男, 陕西西安人, 硕士研究生, 主要研究方向为网络安全、信息教育技术; 何聚厚, 副教授, 硕士生导师, 主要研究方向为信息安全、信息教育技术。

思内容进行自动评估后将结果反馈给用户,其系统结构如图 1 所示:

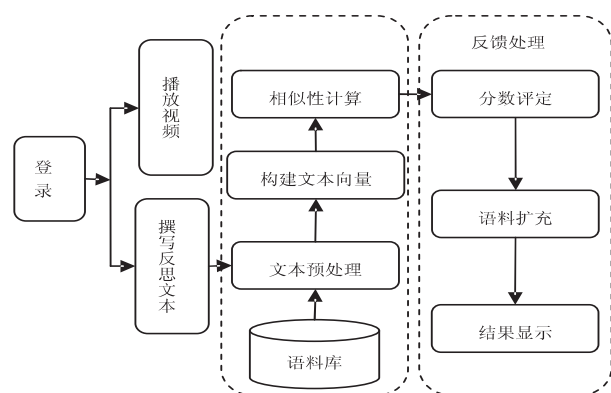


图 1 教学反思自动评估系统结构

(1) 登录。

该模块提供学员登录系统的界面,需要输入用户名和密码。

(2) 播放视频、撰写反思文本。

登录后,学员可以选择播放课堂录像并撰写反思内容。建议课堂录像采用三分屏录制,包含教师视角、学生视角及电子课件视角,学员可以从不同角度针对具体教学问题更好地进行教学反思。

(3) 评估计算。

实训学员提交教学反思内容后,通过文本预处理对反思文本及语料库中的文本进行分词、停用词过滤,根据结果中的词维和权重构建基于向量空间模型构建文本向量。采用余弦理论计算反思文本向量与语料库文本向量的最大相似度。

(4) 反馈处理。

将评估计算的最大相似度结果与系统预设阈值进行比较,若满足阈值则最大相似度语料文本对应的等级即为教学反思文本的等级,并将反思文本添加到语料库对应等级目录中,对语料库进行扩充。最大相似度不满足阈值的反思文本进行标记由人工进一步评定。最后显示系统评定结果。

2 文本处理

2.1 文本处理

由于待处理的反思文本内容为自然语言描述的纯文本,计算机无法直接理解其语义并进行相应处理,所以需要对其文本内容进行预处理。

2.1.1 中文分词

目前的中文分词算法主要分为三大类:基于词典的方法、基于统计的方法和基于规则的方法^[6,7]。由中国科学院计算技术研究所研制的 ICTCLAS(Institute of Computing Technology, Chinese Lexical Analysis System)在中文分词领域效果得到广泛认可。该系统支持用户

词典,支持 gb2312、GBK、UTF8 等多种编码格式。分词速度单机达到 500KB/s,分词精度 98.45%^[8]。

文中采用中科院汉语分词系统 ICTCLAS(C#版本),对反思文本内容及语料库文本内容进行分词处理,并根据停用词词典对分词结果进行过滤。将待处理的教学反思文本和语料文本切分为单词序列为后续计算提供基础。

2.1.2 停用词处理

待处理的反思文本通常包含对文本表达含义影响不大的词语或符号,如“的”、“啊”、“吗”等。它们本身并没有太多意义,虽然在语料库中出现频率较高,但对文本相似性度量贡献不大,称为停用词。通过构建的停用词表,对语料库及反思文本中所包含的停用词进行过滤以降低文本表示的向量维数从而减少计算量、提升效率和正确率。

2.2 文本相似度计算

将自然语言描述的反思文本表示为便于计算的形式化结构,根据所采用的算法计算文本之间的相似程度。

2.2.1 向量空间模型

给定的反思文本内容可以看成由许多词条组成。根据“朴素贝叶斯模型假设”:在文本向量中,各特征的出现是相互独立的^[9],即文本中各词条相互独立,其出现的概率与它所处的位置和代表的内容无关。因此采用向量空间模型^[10],可以将文本中的每一个词作为该文本向量的一个维,那么反思文本向量 V_q 可以表示为 $(w_1, w_2, \dots, w_i, \dots, w_n)$,其中 w_i 表示该向量的第 i 个词的权重, n 表示该向量空间的维数即文档集中总的词数。同理,语料库文本向量 V_d 也可表示。

根据向量空间模型理论,当两个文本内容相近时,其对应的向量之间的夹角会越小。相反,若内容相差较大是,向量之间的夹角就会越大。我们可以用向量之间的余弦值来度量文本之间的相似度。其相似度计算公式如下:

$$\text{Similarity}(q, d) = \cos(\theta) = \frac{V_q \cdot V_d}{|V_q| \times |V_d|} = \frac{\sum_{i=1}^n w_{i,q} \times w_{i,d}}{\sqrt{\sum_{i=1}^n w_{i,q}^2} \times \sqrt{\sum_{i=1}^n w_{i,d}^2}} \quad (1)$$

其中,分别表示反思文本向量和当前比较的语料文本向量。其中表示词 i 在反思文本 q 中的权重,表示词 i 在语料库文档 d 中的权重, n 表示向量空间的维数。由于各权重大于等于 0,所以公式 1 中文档之间的相似度在 $[0, 1]$ 之间,值越大表示文本之间越相似,其中 1 表示反思文本与语料库文本完全相似, 0 表示完全

不同。

2.2.2 TF-IDF

无论对语料文本或反思文本表示时,为了更好地体现词在文本相似性度量中的作用,设置词权重来描述词的贡献程度,提升文本之间度量的准确率。权重的计算采用 TF-IDF^[11,12] (Term Frequency - Inverse Document Frequency) 来进行表示。其主要思想是:如果某个词在文档中出现的次数越多就越重要,并且在其他文档中很少出现,则认为该词重要性就越高。

其中 TF (Term Frequency) 表示词频,即词 t 在文本 d 中出现的次数,记为: $tf(t, d)$ 。在计算中为了防止某些词频过大,取真实频率的平方根。

IDF (Inverse Document Frequency) 为逆文档频率,描述词在文档集中的普遍性。一般来说,文档集合中包含某个词的文档越多,说明它区分文本之间相似性的能力越低,权值就越小。其计算公式如下:

$$idf(t) = 1 + \log \frac{N}{n_t + 1} \quad (2)$$

其中 $idf(t)$ 表示词 t 的逆文档频率, N 表示文档集的文档总数, n_t 表示出现词 t 的文档数。

据 TF-IDF 权重计算方法,综合公式(1)、(2)则单词 t 的权重为:

$$w_{t,d} = tf(t, d) \times idf(t) = tf(t, d) \times (1 + \log \frac{N}{n_t + 1}) \quad (3)$$

2.2.3 相似度计算与评分

相似度计算是模型研究的核心,根据反思内容及语料库内容的分词结果构建文本向量,采用余弦定理计算反思文本向量与语料库中各文本向量的夹角余弦值作为文本之间的相似度。

其具体实现步骤如下:

- (1) 对反思内容及语料文本进行分词;
- (2) 根据停用词表对分词结果进行停用词过滤处理;
- (3) 计算分词总数作为向量空间的维数;
- (4) 统计词频,采用 TF-IDF 方法(见公式 3) 计算各词的权重;
- (5) 采用向量空间模型构建文本向量;
- (6) 根据余弦定理(见公式 1) 计算反思文本向量与语料库各文本向量之间的相似度。

通过相似度计算,可以获取反思内容与语料库各文本之间的相似度。在计算结果中取最大的文本相似度与预先设定的阈值进行比较,如果大于系统设定的阈值则将该最大相似度所对应的语料文本的分数作为反思文本的分数。如果低于阈值则对反思内容进行标记,交由人工进一步评定防止错误评判或漏判。

3 反思内容语料库

语料库是反思评估的依据,提供经人工或系统自动评级确定的典型反思文本。语料文本包括:反思题目、分数等级、反思内容、创建时间、评级专家等信息。为便于语料库信息的加工处理,语料文本采用 XML 格式化语言描述。

为使评估模型有较好的适应性,语料库采用自动构建方法。对满足相似度阈值的反思内容自动创建 XML 语料文件并添加到语料库对应类别目录中,实现语料库的自动扩充。

3.1 语料库构建

根据基于 Web 的语料数据收集系统,自动收集专家提交的典型反思文本及相关信息。提取收集的数据信息,自动保存为初始语料。语料库的构建流程包括:数据采集、信息提取、语料自动生成。其流程图如图 2 所示:

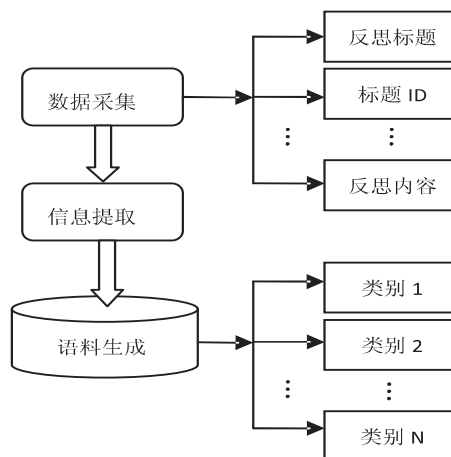


图2 语料库构建流程图

数据采集: 反思专家通过登录基于 Web 的语料管理系统,填写反思标题及标题 ID、典型反思内容、分数等信息,提供语料库的数据源。

信息提取: 服务器收到反思专家提交的数据,更加收集数据的 Web 页面中各文本组件读取反思内容、分数等级、标题 ID 等信息,实现语料数据的提取。

语料库生成: 语料库由 N 个问题类别组成,每个类别中包括若干已经确定评估等级的反思文本。根据反思标题 ID 将提取的反思信息划分到对应语料库类别中,在相应目录生成 XML 文件保存,实现语料文件的生成。

3.2 语料库的自动扩充

语料库构建完成后,可以根据反思文本的相似度结果对语料库自动扩充,增加同类语料文本规模以提高系统的自动评估效果,减少对反思内容误判和漏判的,如图 3 所示。

将反思内容计算的最大相似度与系统预设阈值进行比

较。若满足系统阈值,则该反思文本达到典型反思文本要求,反思内容通过基于 Web 的语料数据收集系统自动提交。服务器获取到数据信息后根据反思标题 ID 将反思内容添加到语库对应类别中,并以 XML 文件格式保存,实现语料库的动态扩充。

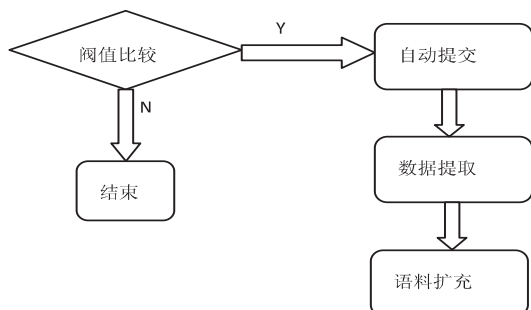


图 3 语料库的自动扩充

4 实验结果及分析

从上述研究过程可以看出,系统实现的基本原理是基于文本相似度计算的文本分类思想。为了对系统的评分结果进行分析,文中采用文本分类系统中常用的三种评价标准:准确率、召回率及 F1 值。公式定义^[13,14]如下:

准确率 (precision) = (分类正确的文本数) / (实际分类的文本数)

召回率 (recall) = (分类正确的文本数) / (应有分类正确的文本数), 其中 (p 为准确率, r 为召回率)

实验数据为从多位教学反思专家采集的典型反思内容语料库,其中针对某具体反思问题共有 60 分到 100 分 41 个等级的典型反思文本 3105 篇。测试集选取了 375 篇反思文本,分布在 81 分到 90 分 10 个级别中,各级别测试集的文档数目在 26 个到 51 个之间,语料集与测试集中文档数量比例在 2:1 左右。测试集文本通过评估系统进行评估测试,其查准率、查全率以及 F1 测试值分别如图 4、图 5、图 6 所示:

从上图实验结果可以看出,系统评估模型具有较好的准确率、召回率、F1 值。并将评分结果与经过专家认定的原评分结果对比,效果良好正确率达到 90% 以上。

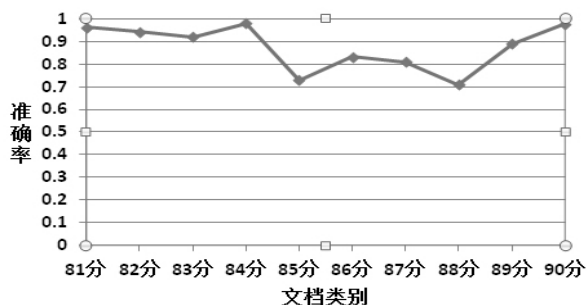


图 4 查准率

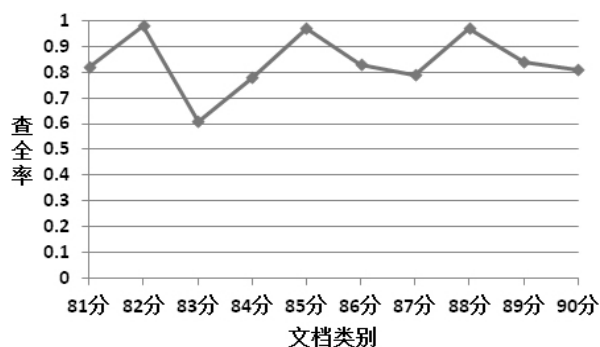
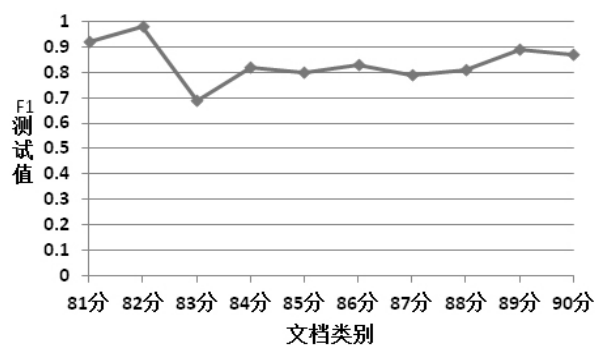


图 5 查全率

图 6 F₁ 值

5 结束语

文中探讨了网络环境下,教学反思自动评估模型的原理、设计及实现,并通过实验进行验证。采用向量空间模型、文本分词技术等,通过度量文本相似度的方法实现教学反思内容的自动评估,为教师的教学反思效果进行实时反馈。在测试实验中系统评分效果良好。在后续的研究中将文本相似度过度到基于本体的语义相似度计算方法上会更好提升评分结果的准确性。

参考文献:

- [1] 申继亮,刘加霞.论教师的教学反思[J].华东师范大学学报(教育科学版) 2004 22(3):44-49.
- [2] Calandra B, Dias L B, Lee J et al. Using video editing to cultivate novice teachers' practice [J]. Journal of Research on Technology in Education 2009 42(1):73-94.
- [3] So W W, Pow J W, Hung V H. The interactive use of a video database in teacher education: creating a knowledge base for teaching through a learning community [J]. Computers and Education 2009 53(3):775-786.
- [4] Kong S C, Shroff R H, Hung H K V. A web enabled video system for self reflection by student teachers using a guiding framework [J]. Australasian Journal of Educational Technology 2009 25(4):544-558.
- [5] 乔拥军.基于 Blog 的中小学教师教学反思的应用研究[D].兰州:西北师范大学 2009.

(下转第 180 页)

⑧ 报告生成。经过结果分析后,可以讲结果生成研究报告,以方便查阅、演示和存档。

4 Simulation 优化分析

在工程设计中,设计方案往往不是唯一的,从多个可行方案中寻找最优方案的过程,成为优化设计(Optimal Design)。机械优化设计是设计工作者一直追求的目标。从数学的观点看,工程中的优化问题就是求解极大值和极小值的问题,亦即极值问题。通过 Simulation 的优化分析,可以从众多的设计方案中搜索最佳解决方案,以最低的成本获得最好的效益,提高设计效率。

以零件轴为例,进行 Simulation 优化分析,方法如下:

单击【评估】下的【设计算例】,然后分别对【变量】、【约束】和【目标】进行设置。选择轴的直径为设计变量,选择节应力的大小为约束条件,最后选择质量减小为设计目标。设置完成之后,点击【运行】,就可以得到优化结果,结果如图 6 所示。

		当前	情形 2	情形 3	情形 4	情形 5
轴直径		20 mm	15 mm	20 mm	25 mm	30 mm
应力 3	>0 N/mm ²	251.11 N/mm ²	603.11 N/mm ²	251.11 N/mm ²	164.29 N/mm ²	111.43 N/mm ²
质量 3	Minimize	284.004 g	157.762 g	284.004 g	447.305 g	647.627 g

图 6 优化分析结果

5 结束语

SolidWorks Simulation 为了体现设计仿真一体化的解决方案,在无缝集成界面做了创造性的改变,将仿真界面、仿真流程无缝融入到 SolidWorks 的设计过程中。文中通过几个简单实例研究、分析了 SolidWorks Simulation 的一些基本功能及其使用方法。SolidWorks Simulation 的功能非常强大,工程应用中,用户可以使用 SolidWorks Simulation 软件对 SolidWorks 构建的零件和装配体进行高性能的应力分析和优化分析。

(上接第 176 页)

- [6] 许高建,胡学钢,王庆人.文本挖掘中的中文分词算法研究及实现[J].计算机技术与发展,2007,17(12):122-124.
- [7] 刘源,谭强.信息处理用现代汉语分词规范及自动分词方法[M].北京:清华大学出版社,1994.
- [8] 中国科学院计算技术研究所. ICTCLAS 汉语分词系统[EB/OL]. 2008. <http://www.ictclas.org>.
- [9] 薛德军,张钹,孙茂松.中文文本自动分类中的关键问题研究[D].北京:清华大学,2004.
- [10] Salton G, Wong A, Yang C S. A Vector Space Model for Automated Indexing[J]. Communications of the ACM, 1975, 18

参考文献:

- [1] 贾红雨,李成,杨洁,等.复合材料储能飞轮转子有限元分析[J].玻璃钢/复合材料,2009,207(4):35-38.
- [2] Ireland R C, Torres C R. Finite element modeling of a circumferential magnetizer[J]. Sensors and Actuators, 2006, 129: 197-202.
- [3] 殷波.玻璃钢信号灯架的有限元分析与应用[J].玻璃钢/复合材料,2009,207(4):74-76.
- [4] 黄旌,高涛. ANSYS 用于机翼有限元分析的建模研究[J]. 计算机辅助工程, 2006, 2(2): 1-4.
- [5] 许文彬,张华良.基于三角形连接的有限元网格划分[J]. 计算机技术与发展, 2011, 21(9): 59-62.
- [6] 陈荣征,李代平,黄健,等. EBE-PCG 算法在有限元并行计算中的应用研究[J]. 计算机技术与发展, 2008, 18(3): 232-235.
- [7] Mercan B, Schultz A E, Stolarski H K. Finite element modeling of prestressed concrete spandrel beams[J]. Engineering Structures, 2010, 32(9): 2804-2813.
- [8] 冯立浩.基于 SolidWorks Simulation 的随车起重机吊臂有限元分析[J]. 机械工程师, 2011(5): 51-54.
- [9] 祁辉,司建铭,曹世海,等.基于 SolidWorks Simulation

的管端定径机的研究与分析[J]. 机械, 2011(8): 44-48.

[10] SolidWorks 公司.

SolidWorks Simulation 高级教程[M]. 杭州新迪数字工程系统有限公司译. 北京:机械工业出版社, 2009.

- [11] Hawileh R A, Rahma A, Tabatabai H. Nonlinear finite element analysis and modeling of a precast hybrid beam-column connection subjected to cyclic loads[J]. Applied Mathematical Modeling, 2010, 34(9): 2562-2583.
- [12] 张向宇,熊计,郝铨,等.基于 Ansys 的立柱有限元分析与结构优化设计[J]. 机械科学与技术, 2008(12): 56-61.
- [13] 庞艳红,刘尧,吴先良.多小波在一维有限元方法中的应用[J]. 计算机技术与发展, 2008, 18(8): 223-225.

(1): 613-620.

- [11] Baeza-Yates R, Ribeiro-Neto B. Modern Information Retrieval[M]. New York: Addison Wesley, 1999.
- [12] Gospodnetic O, Hatcher E. Lucene in Action2[M]. Stamford: Manning Publications Co, 2010.
- [13] 黄丽琼,何中市,张杰慧.基于文本相似度的自动文摘评价方法[J]. 计算机应用研究, 2007, 24(8): 97-99.
- [14] 徐威,董渊,白若鹤,等.针对中文文本自动分类算法的评估体系[J]. 计算机科学, 2007, 34(8): 44-49.