

自动作文评分在大学英语教学中的应用与前景

○白 云

〔摘要〕 自动作文评分研究在我国尚处于起步阶段。本文基于在大学英语教学中对该系统的试用,提出应将计算机评分与教师讲评相结合,将限时作文与课后作文相结合,将结果写作法与过程写作法相结合的使用策略,并指出了注重其针对性反馈、通用性和人机界面划分的发展方向。

〔关键词〕 自动作文评分;大学英语;写作教学

〔作者简介〕 白云(1973-),女,硕士,山东大学外国语学院讲师,研究方向:英语教学、英语语言学。(山东济南 250001)

〔课题来源〕 本文为2011年山东大学实验室软件建设项目“大学英语写作评分标准的制定与应用”的阶段性能成果。

一、国内外研究现状

所谓自动作文评分(Automated Essay Scoring,简称为AES)就是利用计算机技术对作文进行评估与记分。同教师的人工评阅相比较,计算机自动作文评分有着极大的优势。首先,它可以使评分过程更公平、更高效。电脑不会疲劳、厌烦、恼怒或者疏忽,它会比人更加细致地检查、分析每一篇作文,并且不带任何荒诞的判断、错误的信念或者价值偏见。其次,它可以极大地降低教师的阅卷工作量,教师可以把充足的时间和精力放在教学组织和讲解上。最后,也是最重要的,就是学生可以写出更多的文章,写作能力会提高得更快。尤其是在大学英语教学改革之后,英语课程实行分级、分课型教学改革,每位读写课教师面对数百名学生,如何使学生获得更多的写作机会,并得到有指导性的反馈,已成为写作教学的燃眉之急。

国外该方向的研究至今已历时40余年,并于1999年进入实际应用阶段。一个合理的作文自动评分系统能够充分利用统计技术、自然语言处理技术、信息检索技术等,从作文文本中挖掘能够直接反映文章质量的文本特征项作为变量,以有效地预测作文质量。目前国外最具代表性的三种作文自动评分系统是PEG(Project Essay Grader)、IEA(Intelligent Essay Assessor)和E-rater(Electronic Essay Rater)。自动作文评分领域的开拓者是Page,他在1966年开发了Project Essay Grader(PEG)系统,该系统基于统计技术中的浅层文本特征分析技术,目的就是为了使大规模作文评分更加实际而高效。IEA是上世纪90年代末基于统计技术中的潜在语义分析技术开发的。潜在语义分析的定义为“一个单词用法的统计模型,该模型允许对片断文本包含的信息之间的语义相似性进行比

趣,体现高校自身的办学特色,改变高校课程设置千校一面的局面。

(三)学术创新环节

面授结束后,教师除了要引领学员围绕教学要求进一步自学,巩固所学知识外,还要引领学员学习远程辅助教学平台所提供的深度学习资料,围绕本学科的最新研究前沿、围绕本土文化、围绕自己的本职工作开展学术研究,完成课程论文和毕业论文,为学员进一步深造和从事本学科的研究打下良好的基础。

四、强化四种能力

通过以上三个教学环节的学习,我们设想强化学员的四种能力,来实现“培养服务区域经济和地方文化教育所需要的应用型人才”的培养目标。一是强化学员阅读文言文的能力,为学员继承优秀的文化

遗产、提高自身人文素养打下良好的基础,实现课程的近期培养目标。二是强化学员驾驭中小学文言文教学的能力,提高学员从事中小学文言文教学的能力与水平,实现课程的远期培养目标。三是强化学员围绕本职工作、尤其是围绕中小学文言文教学工作开展学术研究的能力,实现课程的创新培养目标。四是强化学员围绕本土文化开展学术研究的能力,实现课程的特色培养目标。

经过几年来的努力,《古代汉语专题》课的教学内容得以优化,教学形式与方法得以改进,初步彰显出自我的特色。学员的自主学习能力和应用知识能力、自主创新能力普遍得到提高,课程改革初见成效。

责任编辑:王峰

较”(Foltz,1996 见 Dikli 2006:52)。而 E-rater 于 1997 年面世,它结合了自然语言处理与统计技术,目的是基于作文思想组织、句子结构和文章内容给出整体评分。它于 1999 年开始用于 GMAT 考试作文评分,2005 年开始用于托福考试作文评分。试卷的最终得分由 E-rater 和一名评卷员决定,如果 E-rater 跟评卷员的评分差异超过 1 分,就需要第二名评卷员来处理。

国内自动作文评分研究尚处于起步阶段,偏重于介绍性、探讨性。研究表明,英语作为母语的作文评分与英语作为外语的作文评分,尤其是与我国大学生的作文评分,存在着较大的差异,其中主要的差异就在于词汇、句法和篇章结构等方面。对中国学生英语作文的自动评分,最早是由梁茂成做出了有益的尝试。他以提取浅层文本特征为主,结合针对内容的潜在语义分析,进行线性回归,得到了与人工评分较高的相关度,并于 2005 年获得专利。但由于以浅层特征的统计分析为主,以及样本数量、范围的局限性,其评分模型与实用系统尚有一定的距离。梁茂成主持的大规模考试英语作文自动评分系统的研制也已于 2009 年初步完成,处于国内领先水平。另一位研究自动作文评分的是李亚男,其研究方向是汉语作为第二语言测试的作文自动评分,以中国少数民族汉语水平考试作文为样本。曹亦薇和杨晨则首先使用潜在语义分析技术对汉语作文进行了自动评分的探索研究。

二、自动作文评分在大学英语教学中的应用

自动作文评分的应用主要是在两个方面,一是用于大规模考试,作文是各类语言考试中必不可少又很有效的题型,而人工对作文的评分是最费人力、信度相对较低的;另一方面则是英语教学,学生要想提高写作水平只能通过多写和有针对性的反馈,但在实际教学中,老师没有足够的时间和精力来仔细批改每一篇作文并给出反馈,因此,自动作文评分就具有很高的实用性。

鉴于此,笔者在实际教学中试用了一个自动作文评分系统,时间为一学期,共六个班近三百名理工科的大二学生参与了此次尝试。该系统是由浙江大学外语学院与杭州某网络科技公司联合开发的冰果英语智能作文评阅系统。在每次布置作文时,笔者以教师身份登陆该系统,在自己的账户里设定英语作文题目以及要求。每篇设定好的英语作文题目都有一个作文序列号。之后学生可以登录学生账号,通过老师指定的序列号找到要写的英语作文题目,直接在电脑上输入文章后,电脑就会按照设定,给出相应的分数和简

单的评语。作文提交时限之后教师就可以批量下载学生的作文及系统给出的成绩、评语。

通过一学期的试用,结合课下对学生的访谈,笔者认为在当前大学英语教学中应用自动作文评分系统需注意以下几个使用策略。

(一)应将计算机评分与教师讲评相结合

从本质上看,自动作文评分是计算机根据一定数量的已经评分的作文的某些数据特征将一篇未评分的作文和分数量表上的某个数值建立起一个对应关系。因此,目前的评分系统在技术层面上绝没有达到“欣赏和评鉴”文章,例如计算机无法发现幽默的表达及独特的推理风格。这就需要将计算机评分与教师讲评相结合。教师可借助计算机系统的评价,从中选出较好和较差的作文进行讲评,也可引导学生更好地进行构思、组织文章内容,还可对作文中出现的诸如文体、句式、词汇等方面的问题进行总结讲解,有针对性地各个击破。事实上,自动评分系统的使用本身就给教师节省了宝贵的时间和精力来进行学生作文的分析和讲评,从而更高效、更有针对性地提高学生的写作水平。而如果只是单纯依赖评分系统,老师不进行任何讲评,则会引发学生的不安和烦躁,以致限制学生写作能力的发展。

(二)将限时作文与课后作文相结合

目前很多非英语专业大学生练习英语写作的目的是在各项考试中取得高分,即他们的写作动机有很强的应试性。要进行应试的写作训练,教师最好采用评分系统中的限时作文模式。教师可当堂布置作文题目,并限定提交的最后期限为三十或四十分钟之后。这样,学生不仅要练习如何谋篇布局、遣词造句来表达自己的意思,还要练习如何更好地安排时间,在规定时间内完成作文。这种模式要求学生的电脑可以连接外网,在基于网络的自主学习逐步普及的今天,这具有越来越高的可操作性。其它的写作题目可以布置作业的形式,交给学生课下完成,在规定的期限内提交。

(三)将结果写作法与过程写作法相结合

传统的结果教学法把写作视为一种结果,要求学生一次写成作文,教师批阅后给出分数和评语。对学生而言,交上了作文就感觉“一劳永逸”,发下作文时,只关心分数的高低,对教师的评语和修正往往不认真阅读,导致同样的错误在以后的作文中重复出现。而过程教学法要求学生多次修改同一篇作文,想写什么就写什么,比较重视思想内容的挖掘和表达。它注重反馈和评改,强调多次修改,认为修改是写作过程中

极为重要的一个环节,学生应通过一次次修改来提高自己的认知能力、思维能力和写作能力。国内已有大量对过程写作法的介绍与研究(如邓鹏鸣等,2003;张莹,2006),但其在实际教学中应用遇到的最大问题就是教师数量相对较少,根本没有时间和精力评阅如此大量的作文。而自动作文评分系统的出现恰恰突破了这个瓶颈,把教师从繁重的评分和纠错中解脱出来,使教师注重于帮助学生拓宽思路,充实文章内容,鼓励学生自由发挥。学生可以借助系统的反馈完成从初稿到终稿的修改,在不断的修改过程中领略写作的实质和乐趣。

三、自动作文评分在教学中的发展方向

目前,大学英语写作教学仍然停留在老师讲解写作技巧,学生背作文模板,写作训练严重匮乏的阶段。随着计算机的普及和网络自主学习的推广,自动作文评分和反馈必将起到越来越大的作用。要使作文评分和反馈系统对日常教学更有帮助,在今后的研究中应注重以下三个方面。

(一)有针对性的反馈

从教学的角度来看,有针对性、个性化反馈的重要性要超过准确的作文评分。如果只给学生的作文打一个分数,哪怕这个分数再准确,也是远远不够的。对于学生而言,他需要知道一篇文章存在什么问题,哪些地方需要改正或规避;对于教师而言,他需要知道哪些语言点学生已经基本掌握,哪些语言点还需要深入地讲解。这些目标只有通过计算机作文评估中给出的尽可能详尽的反馈才有可能达到。而现在系统给出的评语中,只有语言表达不够完善、单词拼写有错误等,但具体哪句话或哪个单词就不得而知了。

(二)通用性

通用性是指评分模型实现一次训练多次使用,适用于非特定题目的作文评阅。目前自动作文评分研究,都是注重自动评分研究的第一个目标,即大规模考试评分,采用一个题目训练一个评分模型,强调的是评分的准确性。大规模考试中一个题目的作文会有几十万上百万篇,训练一个准确的评分模型可以解决很大的问题。但是对于第二个目标,即作文写作教学中的反馈工具,相关的研究还比较薄弱。在日常教学中,每个老师都会有不同的题目,而且对一位老师来说,每次写作的学生数量都不会很大。不可能每次都训练一个计算机作文评分模型。所以从教学的实用角度来讲,非特定题目的通用计算机评分才是解决问题的最好方法。

既然是非特定题目的通用的计算机评分,就要把

语言和内容分别处理。在收集训练作文时要覆盖尽可能多的作文题目,在人工评分时要偏重于语言的使用,在构建评分模型时剔除与题目有关的内容。由此构建通用的关于语言使用的自动评分模型,再辅之以不经训练的内容评价法,如文本聚类或关键词法,就可以用于大学英语的写作教学中。

(三)人机界面的划分

所谓人机界面的划分,就是必须决定哪些问题交给计算机去解决,哪些问题必须要有人的参与,从而达到以最少量的人工参与达到系统效率的最大化。从现有的技术来看,作文评阅完全由机器来自动完成是不现实的。因此,研究人机结合的作文评阅方法,设计一种人和机器最佳结合的评阅流程,使人和机器最大限度地相互取长补短,应该是自动作文评分的发展方向。比如,在作文内容的评价方面,通过同一题目的文本聚类,可以发现某一篇或数篇作文与大多数作文用词不同,这些少量的疑似跑题作文需要提交给教师,人工判断是否确实跑题。

自动作文评分完全符合当今的《大学英语课程教学要求》,顺应了“以现代信息技术,特别是网络技术为支撑,使英语教学不受时间和地点的限制,朝着个性化学习、自主式学习方向发展”的潮流。虽然尚不完善,但这一先进技术已经进入外语教学领域。我们要探索性地应用它,扬长避短,使其发挥更大的作用。

参考文献:

- [1] Shermis, M.D. & Burstein, J. Automated Essay Scoring: A Cross Disciplinary Perspective [M]. Mahwah, NJ: Lawrence Erlbaum Associates, 2003.
- [2] Streeter, L., Pstoka, J., Laham, D. & MacCuish, D. The credible grading machine: Automated essay scoring in the 21st century [EB/OL]. <http://www.k-a-t.com/papers/CredGrading2002>.
- [3] Dik, I. S. Automated Essay Scoring [J]. Turkish Online Journal of Distance Education, 2006, (1).
- [4] 梁茂成. 中国学生英语作文自动评分模型的构建 [D]. 南京大学博士学位论文, 2005.
- [5] 李亚男. 汉语作为第二语言测试的作文自动评分研究 [D]. 北京语言大学博士论文, 2006.
- [6] 曹亦薇, 杨晨. 使用潜在语义分析的汉语作文自动评分研究 [J]. 考试研究, 2007, (1).
- [7] 邓鹏鸣等. 过程写作法的系统研究及其对大学英语写作教学改革的启示 [J]. 外语教学, 2003, (6).
- [8] 张莹. 成品写作法向过程写作法的嬗变 [J]. 外语研究, 2006, (6).

责任编辑: 方钢山