

文章编号:1005-4642(2025)06-0033-06

中文文本分类在物理实验教学评价中的应用 ——以物理实验课后练习题文字回答的自动评分为例

杨泽斌^a, 洪玲儿^a, 胡亦菲^a, 吴先球^{a, b}

(华南师范大学 a. 物理学院; b. 物理学国家级实验教学示范中心(华南师范大学), 广东 广州 510006)

摘 要:以物理实验课后练习题文字回答的自动评分为例,探讨了基于中文文本分类技术在大学物理实验教学评价中的应用. 依据 SOLO 分类理论制定了大学物理实验课后练习题文字回答的分类规则,在比较不同分类器性能指标后选择出最优分类器,实现对学生文字回答的自动分类和评价. 结果表明,利用中文文本分类技术可实现学生物理实验课后练习题的文本自动分类,能够帮助教师对学生进行高效评价与分类统计,为教师教学提供有效的改进依据.

关键词:中文文本分类; SOLO 分类理论; 物理实验; 教学评价

中图分类号:G642.423

文献标识码:A

DOI:10.19655/j.cnki.1005-4642.2025.06.005

大学物理实验教学评价是课程教学中的重要环节,起到评价学生学习情况、检验教师实验课程教学效果与教学目标的作用^[1]. 近年来,不同学者对大学物理实验教学评价体系进行了探索,取得了丰富的研究成果^[2-4]. 随着“互联网+教育”和智慧课堂的发展,传统的评价方式较难满足以学生为中心的评价要求^[5],多维评价模型的实施相比于单一的考核方式操作程度更加复杂,加上学生人数的增多,这无疑增加了教师的工作量;对学生的评价依赖于教师的专业判断和经验,这容易受到教师主观因素的影响,如个人偏好、情绪状态等,从而影响评价的可靠性. 随着数字化时代的到来,运用机器学习、人工智能等新兴技术辅助大学物理实验的测量与评价是发展趋势^[6],许多专家学者将人工智能、大数据应用于教育中,为大学物理实验教学评价提供了新的发展方向. 如何捕捉和利用物理实验教学过程中学生产生的大量数据来对学生进行评价是目前该方向的研究重点. 中文文本分类技术是基于机器学习和自然语言处理的技术,其核心目的是将大量的文本数据按照一定的规则或标准进行分类. 该技术在信息检索、情感分析等领域具有广泛应用^[7],而在教

学评价中,中文文本分类技术凭借其强大的数据挖掘和分析功能可以更高效、准确地处理在物理实验教学过程中产生的文本信息,这也让其应用于物理实验教学评价中成为可能.

本研究以物理实验课后练习题文字回答的自动评分为例,先依据 SOLO 分类理论制定出学生文字回答的分类规则,再根据分类规则对学生的大学物理实验课后练习题的文字回答进行中文文本分类,通过对比不同分类器的性能指标,选择最优分类器对样本进行自动分类,并利用数据挖掘和统计分析提取出学生文字回答中的关键词汇、语句结构等重要特征,进而判断学生对实验原理是否掌握、实验步骤的理解是否正确等. 通过中文文本分类技术对学生物理实验课后练习题的文字回答进行自动化处理和分析,获取学生学习情况,发现学生在学习过程中存在的普遍问题和不足,为教师实验教学提供有效、客观的改进依据.

1 分类原理

1.1 SOLO 分类理论

实现中文文本分类需制定好文本分类规则,再根据分类规则对未知文本数据进行分类,从而

收稿日期:2024-09-26; **修改日期:**2024-12-10

基金项目:2024 年教育部高等教育司实验教学和教学实验室建设研究项目(教高司函[2024]1 号);华南师范大学学生课外科研项目(No. 24WDGA01);广东省高等教育教学改革项目(粤教高函[2023]4 号)

作者简介:杨泽斌(2001—),男,广东广州人,华南师范大学物理学院 2023 级硕士研究生,研究方向为信息技术在物理实验中的应用. E-mail:2023022075@m.scnu.edu.cn

通信作者:吴先球(1968—),男,广东平远人,华南师范大学物理学院教授,博士,研究方向为信息技术在物理实验中的应用、核磁共振技术及其应用. E-mail:xqw@foxmail.com

得到分类结果. 在本研究中, 大学物理实验课后练习题题目内容为“如何根据布儒斯特角测量样品折射率? 请写出具体的实验步骤”. 学生在进行文字回答时, 表现出学习者在解题过程中的思维能力. 而 SOLO 分类理论^[8]是在皮亚杰发展阶段学说基础上提出的学习质量评价方法, 该理论指出: 一个人的总体认知结构是不可测的, 而当一个人回答某个问题时所表现出来的结构却是可测的, 即“可观察的学习成果”, 通过学生可视化解题思维结构可直观地反映出学生达到教学目标的程度^[9], 以实现对学生的评价, 因此本文将该理论作为文本分类规则制定的理论基础.

SOLO 分类理论将学生学习成果结构从低到高分别为前结构、单点结构、多点结构、关联结构和抽象扩展结构共 5 个层次, 各结构层次及其具体表征如表 1 所示.

表 1 SOLO 分类理论的层次及其具体表征

水平层次	行为表现
前结构	学生尝试运用无关条件或错误条件来解决问题.
单点结构	学生只运用单一条件来解决问题.
多点结构	学生只能孤立地运用多个条件来解决问题, 不能对条件之间的关系进行关联和整合.
关联结构	学生可以将所有有关的条件联系在一起, 并对其间关系进行进一步综合.
抽象扩展结构	学生不仅能运用所有条件, 还能对问题进行抽象概括, 使其适用于其他新的问题情景.

参照 SOLO 分类理论, 结合学生具体的解题情景与回答表现, 对于理论中的“抽象扩展结构”要求较高, 依据其余 4 个水平来划分, 对学生解题的思维结构层次进行分类. 表 2 为学生文字回答布儒斯特角问题所反映的思维结构水平层次及其具体行为表现.

表 2 学生解题思维结构水平层次及行为表现

水平层次	行为表现	分类类别
前结构	学生运用了其他的实验原理或无原理.	1
单点结构	学生仅能写出布儒斯特角的基本原理或某个实验步骤.	2
多点结构	学生能写出布儒斯特角的原理与基本实验步骤, 但均有缺漏.	3
关联结构	学生能写出布儒斯特角的原理, 并根据原理设计出完整的完整实验步骤.	4

根据表 2 将学生的思维结构水平层次对应到分类类别, 制定本研究的文本分类规则, 再根据分类规则对学生的文字回答进行分类.

1.2 中文文本分类技术

随着信息技术与计算机技术的发展, 中文文本分类技术^[10]也较为成熟, 其分类过程主要分为 4 个阶段: 文本处理阶段、训练阶段、分类阶段以及分类器评估阶段.

文本处理阶段是进行中文文本分类的首要阶段, 该阶段分为文本预处理和特征处理, 在文本预处理中, 输入的中文文本先后经过文本分词、去停用词、文本表示的步骤转化为计算机可运算的数据格式, 得到原始的文本特征集. 在特征处理中, 选择合适的特征选择函数对文本预处理所得到的原始特征集进行筛选, 得到最终的特征集合. 本研究选择 TF - IDF 权重计算函数作为特征选择函数, TF 为词频, IDF 为逆文档频率指数, 该函数公式为

$$TF_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$IDF_t = \lg \frac{N}{DF_t}$$

$$W_{ij} = tf_{ij} \times IDF_t$$

其中, $n_{i,j}$ 代表某个词出现的次数, $\sum_k n_{k,j}$ 代表总词数, N 指所有文档总数, DF_t 表示含有特征词的文档数目.

训练阶段主要是对文本预处理阶段的最终特征集合进行机器学习, 而文本分类阶段是使用训练得到的分类器对未知文本进行分类. 在这 2 个阶段中, 关键在于选取合适的机器学习分类算法, 常见的机器学习分类算法^[11]有 SVM 算法、KNN 算法、逻辑回归算法和随机森林算法. 最后在分类器评估阶段中将训练得到不同类型的机器学习模型进行对比, 选择出最优分类算法.

1.3 结果评估指标

如何确定哪个文本分类器更佳? 这需要借助文本分类器的性能指标来判断. 文本分类器的性能评价指标主要有召回率、 $F1$ 值、精准率以及准确率等, 本研究主要关注分类器对学生文字回答样本的整体分类效果, 而准确率是分类器预测正确的结果占总样本的百分比, 突出模型整体的评价效果, 因此选择准确率作为性能评价指标之一.

在机器学习应用于教育评估的研究中, 往往

还会关注机器和人分配分数的一致性^[12],这种一致性表征了机器评分系统与人类评分者之间在评分结果上的相似程度,是判断机器评分系统在教學评估中是否可靠、替代人类评分是否有效的重要依据。

评分者之间的一致性通常用科恩 Kappa 值来进行量化。根据文献^[13],科恩 Kappa 值范围与一致性对应如表 3 所示。

表 3 科恩 Kappa 值范围与一致性对应表

Kappa 值	一致性
≤ 0.40	较差一致性
0.41~0.60	一般一致性
0.61~0.80	良好一致性
0.81~1.00	接近完美一致性

在本研究中用各分类器输出的结果与专家对数据集的分类结果来计算科恩 Kappa 值,并将科恩 Kappa 值作为分类器性能的另一个评价指标,若文本分类器与人类评分者之间的一致性在良好及以上则为研究可接受的范围。另外,为了更加直观对比不同模型的分类表现,绘制出各模型对应的 ROC 曲线,用比对图来对各个模型分类效果进行评估,并计算出各 ROC 曲线下面积 S_{AUC} 作为评价指标, S_{AUC} 取值范围在 0.5~1.0 之间,值越高则表示该训练模型的预测能力越好^[14]。

2 技术实现

2.1 数据介绍

本次研究的中文文本分类数据来自于大学物理实验课后练习题的学生文字回答,通过下发纸质问卷并由学生填写后收回,借助扫描软件对学生的纸质问卷进行扫描与文字识别处理,将数据录入到 Excel 中再进行后续分析。所收集到学生文字回答样本总共 372 份,学生来自华南师范大学物理学院各年级本科生。本研究主要是建立分类系统来更普遍地对学生的回答进行自动分类,因此对收集到的数据不作区分。

收集完毕之后,还需按照一定标准对数据进行分类。在本研究中,由 2 位大学物理实验教学经验丰富的教师作为评分者,根据上述标准对收集的数据进行分类评分,直至评分者之间的科恩 Kappa 值大于 0.8,以获得更高的一致性。本研究

数据中评分者之间的科恩 Kappa 值为 0.898,由表 3 可知 2 位评分者之间具有接近完美的一致性。

2.2 实现流程

为实现中文文本分类技术,本研究所使用的编程语言为 Python 3.8,开发环境为 PyCharm。在已配置好的开发环境中编写程序,程序基本流程如图 1 所示。

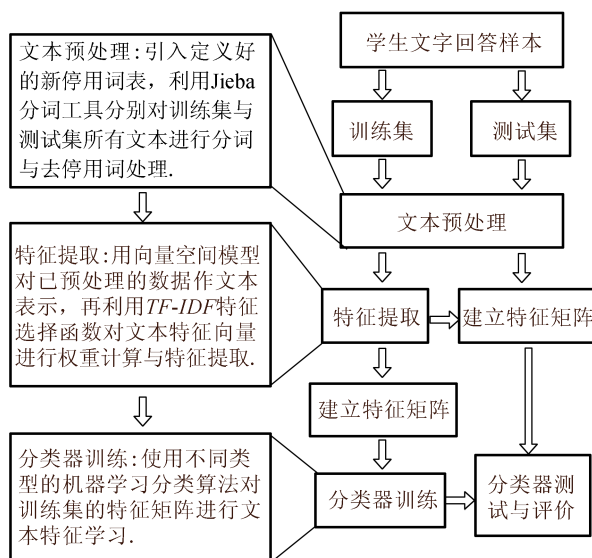


图 1 程序基本流程

首先,将学生文字回答样本进行分割,随机分为训练集与测试集。其中,训练集占比 90%,测试集占比 10%。在文本预处理部分,利用 Jieba 分词工具分别对训练集与测试集中的所有文本进行分词处理。在学生文字回答中有些词如“折射率”“平行光”等可以很明显地表现出类别特征,而像“的”“了”“地”等词,虽然出现的频率很高,但没有实际意义,这类词则为停用词,对于这类词语应将其删除。根据哈尔滨工业大学中文停用词表,再通过分析学生文字回答内容,记录出可能是停用词的词语并在原词表上进行补充,形成新的停用词表,并引入新词表作去停用词处理。

在特征提取部分,本研究使用了向量空间模型将学生的文字回答表示为特征空间的高维特征向量,对于学生的文字回答经过文本预处理后得到中文词语的集合 C ,根据向量空间模型^[15]可以表示为: $C = \{t_1, w_1; t_2, w_2, \dots, t_n, w_n\}$,其中 t_i 代表文字回答中的第 i 个特征词, w_i 代表特征词 t_i 的权重大小,从而实现数据格式的转化。再利用

TF-IDF 特征选择函数对待处理的文本特征向量进行权重计算和特征提取,并建立输出特征矩阵。

在分类器训练部分,分别使用 SVM 算法、KNN 算法、逻辑回归算法和随机森林算法对训练集的特征矩阵进行文本特征学习。在分类器训练时,运用了 K 折交叉验证以防止模型出现过拟合,并使用了参数网格搜索方法为各模型寻找更优参数,确保得到各模型下最优的分类器。

2.3 模型选择

在分类器测试与评价部分,用测试集对每个训练好的分类器进行测试与评价,得到各文本分类器的性能指标,如表 4 所示。

表 4 各模型性能指标

模型算法	准确率	科恩 Kappa 值
SVM 算法	0.765	0.589
KNN 算法	0.689	0.469
逻辑回归算法	0.758	0.576
随机森林算法	0.791	0.618

对照表 3,可以看出随机森林算法分类器的科恩 Kappa 值达到了良好的一致性,而 SVM 算法与逻辑回归算法的科恩 Kappa 值虽然在一般一致性的范围,但均接近 0.6,可认为除了 KNN 算法外,分类器与人类评分者之间的一致性在可接受范围。在 4 种分类器中,随机森林算法的准确率最高,但仅接近 0.8,这也说明了本研究所得到的分类器与人类评分者在进行同一特定分类任

务时还存在一定的差距。

下面绘制出 4 个模型 ROC 曲线比较图,并计算出曲线下面积 S_{AUC} ,如图 2 所示。

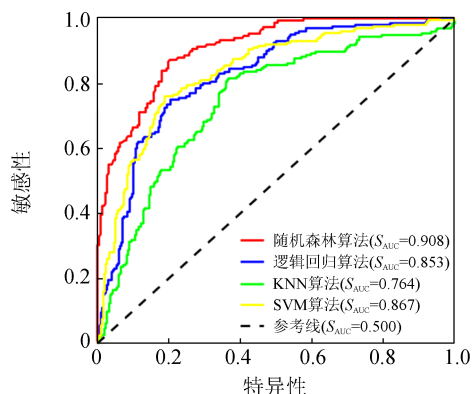


图 2 4 个模型的 ROC 曲线比较

由图 2 可以看出,随机森林算法下的 S_{AUC} 值最高。综上,对于本研究中样本的分类,随机森林算法的整体性能更好。因此,选择该模型下的最优分类器对大学物理实验课后练习题的文字回答进行自动分类。

3 研究结果

3.1 实验数据与分析

通过中文文本分类技术对学生文字回答的分类,经过特征提取部分得到学生文字回答中有效特征数共 72 个,再利用最优分类器对样本进行自动分类,之后对不同类别下的学生文字回答中出现的特征词组进行统计处理,得到如表 5 所示各类别的特征词组统计表。

表 5 各类别的特征词组统计表

类别	水平层次	特征个数	占比	主要特征词
1	前结构	10	13.8%	样品、布儒斯(特角)、折射率、偏振光、最小、偏向角、顶角、分光计、角度、入射光
2	单点结构	23	31.9%	样品、布儒斯(特角)、折射率、正切值、偏振光、入射角、角度、光强、消光、自然光
3	多点结构	48	66.7%	样品、布儒斯(特角)、折射率、正切值、偏振光、入射角、(起)检偏器、角度、消光、调节
4	关联结构	55	76.4%	样品、布儒斯(特角)、折射率、正切值、偏振光、入射角、(起)检偏器、角度、光强、转动

观察表 5 数据,可以发现中文文本分类技术对不同类别下学生文字回答提取的特征数会有所不同,所提取的特征个数会随着学生结构水平层次的增加而增加,下面对各层次中的特征词语进行统计分析:

1) 处于前结构水平层次的学生文字回答中,

主要特征词里面出现“样品”“折射率”“布儒斯(特角)”等重要特征词,但学生却处在较低结构水平,可以看出此水平学生可能对测布儒斯特角的原理不太熟悉,另外还有“最小”“偏向角”“顶角”等特征词,可以推测部分学生可能运用了最小偏向角测样品折射率实验的实验原理回答该问题。

2)在单点结构水平层次的学生除了回答出“样品”“折射率”“布儒斯(特角)”等重要特征词,还出现有关测布儒斯特角原理的“消光”“偏振光”等其他特征词,说明处于单点结构水平的学生对实验原理的理解比前结构水平层次的学生更好,且两层次中均没有出现与实验步骤有关的特征词组,可以得出前2个水平层次的学生对实验步骤掌握较差。

3)处于多点结构水平层次的学生文字回答中,既出现有关测布儒斯特角原理的重要特征词,又有如“(起)检偏器”“调节”等与实验步骤有关的特征词,且多点结构层次水平的特征个数是单点结构水平的2倍多,表明了处于该结构水平的学生对测布儒斯特角的原理与实验步骤均具有一定程度的掌握。

4)关联结构层次水平的学生文字回答中,无论是特征个数还是主要特征词,基本与多点结构层次水平相当,后经过比较2个结构水平中其他特征词的异同,发现关联结构层次水平存在更多关于实验步骤方面的特征词,即处于关联结构层次水平的学生在实验步骤上掌握可能更加扎实。

3.2 研究结论

1)实现学生物理实验课后练习题的文本自动分类可行。在本次研究中,收集学生大学物理实验课后练习题的文字回答作为样本数据,通过模型训练与评价选择出最优分类器,使用该分类器对学生文字回答进行自动分类,由分类器自动分类结果与专家对数据集的分类结果所计算出的科恩 Kappa 值为 0.618,说明分类器与人类评分者之间具有良好一致性,证明了利用中文文本分类技术对学生物理实验课后练习题的文字回答进行分类的可行性。

2)使用机器学习能够帮助教师对学生进行高效评价与分类统计。借助中文文本分类技术对学生文字回答实现自动分类,教师仅需将学生回答导入分类器,即可输出学生所处的结构水平层次,得知学生所达到教学目标的程度,实现对学生的教学评价。这样既可有效减轻教师的工作量,提高评价效率;又能降低教师个人主观因素影响,从而确保评价一致性和客观性。

3)中文文本分类技术可以有效获取学生实验的掌握情况。利用中文文本分类技术强大的数据挖掘功能对不同类别下的学生文字回答中出现的

特征词组进行统计处理,提取得到的特征数与主要特征词可以有效反映出学生解题思维的各结构水平层次。教师可查找学生文字回答中的主要特征词,通过对比特征数量与观察特征词语,判断学生的实验原理是否掌握、实验步骤是否正确,亦能通过对处于各结构水平层次学生回答中的主要特征词进行分析与统计,及时发现不同类别下学生所存在的问题与不足,从而更好地指导教学,对教学策略进行及时、有效的改进。

4 应用与展望

本研究尝试将中文文本分类技术应用在物理实验课后练习题文字回答中,为教师改进实验教学提供了可靠的依据。而该技术在其他场景的应用还可以继续作进一步的探索,如:实验课前预习是学生学习物理实验知识和技能的重要一环,而学生在预习过程中往往会存在有关实验原理、实验步骤、数据分析等方面的疑问,收集学生在预习过程中的笔记、问题记录或者思考等文本数据,利用中文文本分类技术对数据进行分类,并进一步挖掘数据中的有效特征词,以便更深入地理解学生的困惑来源。统计出学生在实验前存在的问题可以帮助教师更好地了解学生的学习状况和困惑点,从而提供更有针对性的实验指导。又如:学生在完成实验后,需撰写实验报告,报告内容主要包括实验目的、步骤、数据记录、数据分析和结论等,这也是在物理实验教学过程中学生产生的文本数据,利用中文文本分类技术对学生的实验报告进行自动评估与分析,评价报告的内容完整性、数据准确性、结论合理性,并提供详细的反馈和改进建议。自动化的报告评价一方面减少了教师的工作量,提高了批改效率,另一方面学生在提交报告后也可以获得及时的评价反馈。

5 结束语

随着机器学习与人工智能的发展,运用新兴技术辅助大学物理实验教学评价势在必行。本文将中文文本分类技术应用在大学物理实验教学评价中,对物理实验课后练习题文字回答的自动评分进行了研究与讨论。基于 SOLO 分类理论构建了学生回答的分类规则,并根据分类规则实现了对文字回答的文本分类,对比不同分类器的性能指标选择出最优分类器,验证了学生文字回答自

动分类的可行性. 利用数据挖掘功能提取了不同水平层次学生回答文本的关键特征, 进而判断出学生的实验掌握情况, 为教师提供客观可靠的评价依据, 助力教学的改进. 随着人工智能与教育深度融合, 中文文本分类技术有望在物理实验教学评价中发挥更大作用, 推动教学评价向智能化、精准化发展.

参考文献:

- [1] 教育部高等学校大学物理课程教学指导委员会. 理工科类大学物理课程教学基本要求、理工科类大学物理实验课程教学基本要求[M]. 北京: 高等教育出版社, 2023.
- [2] 刘艳, 王长昊. OBE 理念下大学物理实验课程教学模式的探索与实践[J]. 大学物理实验, 2023, 36(6): 136-138.
- [3] 王晓萌, 张亚萍, 凌翠翠, 等. 基于“三位一体”理念的大学物理实验多维度协同育人体系的探索与实践[J]. 物理实验, 2024, 44(4): 31-36.
- [4] 刘佳琪, 李春密, 冉露, 等. 大学生物理实验素养评价体系的构建[J]. 物理实验, 2020, 40(7): 19-24.
- [5] 王琦, 邱红, 谢铁文, 等. 基于大学物理实验混合式教育模式与评价体系研究[J]. 物理与工程, 2021, 31(S1): 19-27.
- [6] 李佳隆. 大学物理实验课程期末考试的实践研究[D]. 上海: 华东师范大学, 2017.
- [7] 景丽, 何婷婷. 基于改进 TF-IDF 和 ABLCNN 的中文文本分类模型[J]. 计算机科学, 2021, 48(S2): 170-175.
- [8] BIGGS J B, COLLIS K F. 学习质量评价: SOLO 分类理论[M]. 高凌飏, 张洪岩, 译. 北京: 人民教育出版社, 2010: 24-32.
- [9] 陈玉洁, 黄致新, 王琴. 面向深度学习的大学物理学习评价: 以大一学生解决某一力学问题为例[J]. 大学物理, 2020, 39(3): 63-68.
- [10] 张兵磊. 基于机器学习的中文新闻短文本分类研究[D]. 郑州: 华北水利水电大学, 2022.
- [11] 周文安. 机器学习[M]. 北京: 北京邮电大学出版社, 2021: 12-14.
- [12] ZHAI X, YIN Y, PELLEGRINO J W, et al. Applying machine learning in science assessment: A systematic review [J]. Studies in Science Education, 2020, 56(1): 111-151.
- [13] LANDIS J R, KOCH G G. The measurement of observer agreement for categorical data [J]. Biometrics, 1977, 33(1): 159-174.
- [14] 张晓龙, 江川, 骆名剑. ROC 分析技术在机器学习中的应用[J]. 计算机工程与应用, 2007, 43(4): 243-248.
- [15] YU H G. English semantic recognition based on abstract meaning representation and vector space model [J]. Journal of Physics: Conference Series, 2018, 1087(5): 052017-052052.

Teaching evaluation of physics experiment based on Chinese text classification

YANG Zebin^a, HONG Linger^a, HU Yifei^a, WU Xianqiu^{a,b}

(a. School of Physics; b. National Experimental Teaching Demonstration Center of Physics (South China Normal University), South China Normal University, Guangzhou 510006, China)

Abstract: Taking the automatic grading of text answers of physics experiment exercises as an example, the application of Chinese text classification technology was discussed in the evaluation of college physics experiment teaching. Based on the SOLO classification theory, the classification rules of the text answers to the after-class exercises in college physics experiments were formulated. After comparing the performance indicators of different classifiers, the optimal classifier was selected, and then the automatic classification and evaluation of students' text answers could be realized. The results showed that it was feasible to realize the automatic text classification of students' physics exercises after class by using Chinese text classification technology, which could help teachers to evaluate and classify students efficiently, and provided effective basis for teachers' teaching improvement.

Key words: Chinese text classification; SOLO classification theory; physical experiment; teaching evaluation

[编辑: 郭 伟]