文章编号：1003-0077(2013)05-0100-07

适用于中国外语学习者的英文作文全自动集成评分算法

李 霞[1,2]，刘建达[2]

(1. 广东外语外贸大学 思科信息学院，广东 广州 510006；
2. 广东外语外贸大学 外国语言学及应用语言学研究中心，广东 广州 510420)

摘 要：中国英语学习者人数众多，迫切需要针对中国学生特点的、有效适用于大规模英文作文数据的全自动评分算法，以解决中国现有英语教学和大规模英语考试中英文作文批改量大和难度大的瓶颈问题。该文提出了一种能够有效识别中国英语学习者写作特点并能自动识别特征维数的特征选择方法，并在此基础上提出了适用于不平衡分布数据的集成分类评分算法。对来自中国英语学习者语料库中大学英语四、六级不同主题下的 1 115 篇作文的分类结果显示，该文提出的算法比传统的分类评分算法在类内及类间平均分类准确度、召回率及 F 度量值上均有较大幅度的提升。

关键词：作文自动评分；不平衡数据分类；多项式朴素贝叶斯

中图分类号：TP391 **文献标识码**：A

Ensemble Learning Based Essay Automated Scoring Algorithm for Chinese English Learners

LI Xia[1,2], LIU Jianda[2]

(1. School of Informatics, Guangdong University of Foreign Studies, Guangzhou, Guangdong 510006, China;
2. National Key Research Center for Linguistics & Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, Guangdong 510420, China)

Abstract: Nowaday, there are a large number of Chinese English learners in China, the substantial quantities and great difficulties in English writing assessment is now the bottleneck problem in English teaching and testing. So the effective automatic essay scoring algorithms are in great need of in China. In this paper, we first propose a feature selection method which can extract Chinese learners' writing characters effectively and automatically. And then we continue propose a resemble learning based essay automatic scoring algorithm for unbalanced essays data. The classification results on 1 115 university CET4 and CET6 essays from CLEC shows that our algorithm has dramatically promotion in precision, recall, and F-meature value compared with classification for balanced data.

Key words: essay automatic scoring; unbalanced data classification; multinorminal naïve bayes

1 引言

作文自动评分是指通过计算机技术对作文进行评价和分数预估的过程[1-3]。随着国内英语认证考试参加人数的逐年上涨，英语考试作文评分的工作量也逐年大幅上升。在二语习得(Second Language Acquisition)方面的许多研究表明[4-7]，随着写作任务的加重，作文的计算机自动评分成为一个必然的趋势，它可以消除传统人工评分过程中由于人工阅卷员之间的地域性、经验性、语言能力、评卷的严厉度等方面的差异而导致的评分结果的一致性、准确性、客观性和可靠性的降低。同时使用计算机对作文进行自动评分具有即时性、客观性、经济性和公平

收稿日期：2013-06-15 **定稿日期**：2013-08-03

基金项目：国家自然科学基金资助项目(61070061)；教育部人文社科重点研究基地资助项目(11JJD740012)；广东省高层次人才资助项目(粤教师函[2010]79 号)；广东外语外贸大学研究生科研创新资助项目

作者简介：李霞(1976—)，女，博士研究生，副教授，主要研究方向为自然语言处理和语言测试；刘建达(1967—)，男，教授，主要研究方向为语言测试，语用学。

性等优点。

目前国外对以英语为母语的英文作文自动评分技术的研究相对较为成熟，几个作文自动评分系统已被用于英语母语写作评分中，比较有代表性的系统包括：PEG(Project Essay Grade)[8-9]、IEA(Intelligent Essay Assessor)[9-10]、E-rater(Electronic Essay Rater)[9-11]和 IntelliMetric[3,9]等，这些系统分别从内容或形式上提取特征，并利用机器学习中的分类或回归技术来实现作文的自动评分，并取得了较好的效果。然而，受中国文化和汉语思维习惯的较大影响，中国学生在英语写作的语言特征如词汇、句法和篇章结构等层次上与国外以英语为母语的学生所写的英文作文差异较大[12-14]。例如，中国学生更容易出现高频率的词汇、短语搭配、介词使用、句法等方面的错误，更多使用话语虚词(如 well、now、anyway、however 等)、连接词语和形容词等语言特征，而以英语为母语的学生则更注重句型的变化性与灵活性等方面，如作文长度更长，更多使用各种从句等。已有自动评分技术对语言质量的分析主要考虑作文中的句法多样性等母语学生的写作特点，而忽略了非母语学生写作中特有的语言特征，这使得已有的作文自动评分系统无法很好的适用于中国学生的英语作文自动评分中[15]。

由于考生的作文分数普遍位于中等水平位置，即分数高和分数低的作文相对较少，因此，作文评分数据具有不平衡数据分布的特点，这使得传统的分类算法在对作文进行分类评分时效果不佳，通常大类样本数据(在本文中为中等水平作文)分类效果要好于小类样本数据(在本文中为高水平和低水平作文)。为此，本文首先依据中国学生的写作特点，在提出基于高频相邻搭配词组特征选择方法的基础上，利用不充分抽样 bagging 算法对大类数据进行多次随机抽样，并对多次分类结果进行组合，最终结果为各分类结果的投票得分。对中国英语学习者语料[16]大学英语四、六级不同主题作文下的 1 115 篇英文作文的评分结果表明，本文提出的算法能够较好的提取反应中国学生写作特点的特征，并有效适用于不平衡数据的分类，在类内和类间的正确率、召回率和 F 度量值上均有较大幅度的提升。

2　特征提取

传统的特征选择方法通常以分好的单个词为单位，依据所提取的特征对文本构建向量来进行分类处理。以单个词为单位的特征选择方法在作文自动评分中会导致对文字完全相同但顺序打乱前后的两篇作文评为相同的分数，这是因为单个词特征提取时没有考虑到词与词之间的前后顺序关系。依据中国学生英文写作的特点，如对介词和连词等掌握相对不是很好、习惯使用短语词组等特点，一方面避免出现词序打乱后的作文被误评，同时又能充分体现中国学生英文写作的特点，本文提出基于高频相邻搭配词组特征选择方法，该方法既考虑到词的前后顺序关系，同时也符合中国学生在英文写作中习惯使用短语词组的特点。

在提取特征时，本文没有过滤通常意义上的停用词，这是因为在一般的文本分类中诸如 of，to，that 等词由于具有较低的分类贡献度而通常被做为停用词过滤掉。但在英文作文中，介词、连词等的正确使用往往是衡量一个学生英文写作水平的一个重要方面，同时它也是很多短语词组的搭配词，为此本文在提取作文特征时，并没过滤任何停用词，详细特征选择算法描述如下。

输入：n 篇已知分数档的英文作文；

输出：有效词组特征；

1. 对英文作文依据空格分词得到词列表$\{t_1, t_2, \cdots, t_n\}$；

2. 依据从左至右的顺序提取相邻二元搭配词组，得到词组列表 $t_i t_j (i=0,\cdots,n;\ j=0,\cdots,n,\ i\neq j)$；

3. 计算每个相邻二元词组的信息增益值；

4. 对所有相邻二元词组的信息增益值排序；

5. 对横坐标为特征维数，纵坐标为相应二元词组特征的信息增益值为点对画散点图，计算急剧变化的点所对应的维数 k，k 为该训练集的有效特征维数；

6. 输出前 k 个相邻二元词组特征作为该训练集的有效特征。

文本分类领域常用的特征选择方法有文档频率(DF)方法、信息增益(IG)方法、互信息方法(MI)等[17]，本文采用应用广泛的信息增益(IG)方法来提取取作文特征，本文所采用的信息增益的计算公式描述如式(1)所示：

$$\begin{aligned} IG(t) = &-\sum_{i=1}^{k} p(c_i)\log(p(c_i)) + \\ &p(t)\sum_{i=1}^{k} p(c_i \mid t)\log p(c_i \mid t) + \\ &p(\bar{t})\sum_{i=1}^{k} p(c_i \mid \bar{t})\log p(c_i \mid \bar{t}) \end{aligned} \quad (1)$$

其中：t 表示某个相邻二元词组特征，c_i 为第 i 个分数档的作文集合，$\bar{t}$ 表示特征 t 没有出现时的情况，$p(c_i)$表示 c_i 分数档在整个训练语料中出现的概率，$p(t)$表示整个训练语料中包含相邻二元词组特征 t 的作文文档的概率，$p(c_i|t)$表示包含相邻二元词组特征 t 且属于 c_i 分数档的作文文档的条件概率，$p(\bar{t})$表示整个训练语料中不包含相邻二元词组特征 t 的作文文档的概率，$p(c_i|\bar{t})$表示不包含相邻二元词组特征 t 且属于 c_i 分数档类别的作文文档的条件概率，k 表示分数档的个数。针对本文的作文自动评分问题，我们把 c_i 设为 4 个分数档，分别是：4～6 分 1 档、7～9 分档、10～12 分 1 档、13～15 分 1 档，k 取 4。

为了验证我们所理解的中国学生英文写作的特点，我们利用信息增益特征选择方法分别对来自中国英语学习者语料库中大学英语四级作文中主题为 Global Shortage of Fresh Water 的 290 篇作文以及大学英语六级作文中主题为 Haste Makes Waste 下的 344 篇作文，分别以单词为单位和以相邻二元词组为单位进行了特征提取，提取结果如表 1 和表 2 所示。表 1 结果表明，相邻二元词组特征选择方法所提取的特征能较好的反应中国学生的写作特点，例如，较多使用固定搭配、中国式英语、主动句等。如在大学英语四级库中主题为 Global Shortage of Fresh Water 下 290 篇作文所提取的相邻二元词组特征中有较多的固定词组搭配以及反应中国学生写作特点的词汇，如 is_important(非常重要…)，use_them(使用他们…)，under_the(在…下)，are_also(也是…)，very_shortage(非常缺乏…)，must_be(必须)，at_present(目前…)，in_recent(最近…)，already_used(已经使用…)等。在大学英语六级库中主题为 Haste Makes Waste 的 344 篇文章中所提取的相邻二元词组特征中也同样反应了中国学生的写作特点，例如，try_to，it_must，easy_to，much_time，is_easy，makes_us，everyday_life，i_can，all_kinds 等。从这些结果中可以看出中国学生习惯使用短语词组和习惯使用主动句等写作习惯。通过提取这些特征，能够较好地提升中国英语学习者英语作文的分类评分效果。

表 1 主题为 Global Shortage of Fresh Water 共 290 篇大学英语四级作文中所提取的特征对比

前 40 个单词特征	前 40 个词组特征
people's//industry//third//average//needs//everyone//out//recent//actions//already//under//second//future//present//done//should//percent//on//also//searching//incease//abundant//researches//reasonable//earth//but//mustn't//realized//usually//ever//granted//amount//please//out//which//cause//obtain//solved//hand//problem//	is_important//use_them//under_the//are_also//very_shortage//that_can//the_average//must_be//do_our//the_serious//unfit_for//people's_daily//done_to//should_also//be_done//at_present//people_to//second_we//in_recent//we_can//global_shortage//all_the//need_of//people_not//but_in//know_the//so_on//will_not//this_condition//is_limit//countries_have//nowadays_many//already_used//for_ever//we_also//by_controlling//large_of//can_put//are_increasing//present_the//

表 2 主题为 Haste Makes Waste 共 344 篇大学英语六级作文中所提取的特征对比

前 40 个单词特征	前 40 个词组特征
waste//easy//last//going//meaning//they//some//brother//to//sit//classroom//decided//important//usually//understood//failure//chinese//impossible//successfully//who//finds//later//aspects//refused//angry//sentence//quality//invited//goods//cartons//play//birthday//intelligence//kinds//director//answered//dollar// continues//amount//client//	makes_waste//for_my//try_to//it_must//therefore_we//easy_to//much_time//you_did//say_haste//the_knowledge//is_easy//makes_us//things_quickly//everyday_life//the_classroom//however_we//is_that//the_matter//and_then//but_not//that_how//something_done//my_brother//brother_was//failure_if//do_must//it_tells//is_so//we_need//a_hurry//to_cut//save_time//to_achieve//example_one//to_work//he_finds//school_in//sometimes_haste//knowledge_of//do_what//

3　英文作文的表示

采用文本分类中广泛使用的向量空间模型(Vector Space Model)[18]来表示作文，每篇作文对应于一个空间向量，其格式为 $V(d_j)=(\langle t_1,w_1\rangle,\cdots,\langle t_i,w_i\rangle,\cdots,\langle t_m,w_m\rangle)(i=1,2,\cdots,m)$，这里的 $t_i(i=1,2,\cdots,m)$为训练作文数据中选出的 m 个相邻二元词组特征，$w_i(i=1,2,\cdots,m)$为每篇作文 $d_j(j=1,2,\cdots,n)$的第 i 个相邻二元词组特征所对应的权重值，权重值的计算方法主要包括：词频方法(TF)、逆文档频率方法(IDF)、词频—逆文档频率方法(TF-IDF)[19]，在中国学生英文作文数据上的分类评分结果显示，逆文档频率相对其他两个权重公式具有较优结果，其计算公式如式(2)所示：

$$w(t_i,d_j)=\log_{10}\left(\frac{N}{1+df(t_i)}\right) \tag{2}$$

其中，$w(t_i,d_j)$为特征词组 t_i 在作文 d_j 中的权重，$tf(t_i,d_j)$为特征词组 t_i 在作文 d_j 中出现的次数，N 为训练作文文档的总数，$df(t_i)$为作文训练集中包含特征词组 t_i 的作文文档个数。在本文算法中，逆文档频率 IDF 的计算效果最好。作文向量之间的相似度采用余弦相似度来计算，计算公式描述如式(3)所示：

$$sim(d_i,d_j)=\frac{\sum_{k=1}^{m}w_{ik}\times w_{jk}}{\sqrt{\left(\sum_{k=1}^{m}w_{ik}^2\right)\times\left(\sum_{k=1}^{m}w_{jk}^2\right)}} \quad (i,j=1,2,\cdots,n) \tag{3}$$

其中 $sim(d_i,d_j)$表示第 i 篇作文和第 j 篇作文之间的相似度，而 $w_{ik}(k=1,2,\cdots,m)$表示第 i 篇作文的第 k 个词的权重值，$w_{jk}(k=1,2,\cdots,m)$表示第 j 篇作文的第 k 个词的权重值，m 为表示整个作文数据中所有作文采用的相邻二元词组特征总数。

4　基于随机抽样和算法组合的不平衡作文数据分类评分算法

4.1　基于多项式模型的朴素贝叶斯分类算法

基于多项式模型的朴素贝叶斯分类算法[20]在信息检索领域被广泛使用，并且被国外的 BETSY 等评分系统[21-22]所使用，并取得了较好的效果，该算法的详细描述如下。

假定 d 表示用空间向量表示的一篇英文作文，$d=\{w_1,w_2,\ldots,w_m\}$，其中 m 是所选出的相邻二元词组特征的个数，w_i 表示第 i 个相邻二元词组特征在作文文档 d 中的权重值。对测试集中的某个文档 d，预测它属于某个分数档 $c_i(i\geqslant 1,i\leqslant k)$的概率为 $P(c_i\mid d)$，其中 k 为类别个数，表示不同的分数档。依据贝叶斯定理，$P(c_i\mid d)=\frac{P(c_i)P(d\mid c_i)}{P(d)}$，在基于多项式模型的朴素贝叶斯分类算法中，每篇作文被看成是所有所提取的相邻二元词组特征的一个抽样，对于给定的一篇作文，其得到某个分数的概率为包含在该作文中属性的概率乘积，$P(d\mid c_i)$的计算公式如式(4)所示：

$$P(d\mid c_i)=\prod_{k=1}^{V}\frac{P(w_k\mid c_i)^{N_k}}{N_k!} \tag{4}$$

其中 V 为所有二元词组特征的总数，N_k 表示第 k 个相邻二元词组特征 w_k 在作文 d 中出现的次数，$P(w_j\mid c_i)$表示相邻二元词组特征 w_j 在分数档 c_i 的作文中出现的概率，其计算公式如式(5)所示：

$$P(w_j\mid c_i)=\frac{N_{ji}+1}{N_{c_i}+\mid V\mid} \tag{5}$$

其中 N_{ji} 为第 j 个相邻二元词组特征在分数档 c_i 的作文文档中出现的次数，N_{c_i} 表示分数档 c_i 的所有作文文档中二元相邻词组的总数。

4.2　基于随机抽样和算法组合的不平衡作文数据分类评分算法

由于作文数据具有分布不平衡的特点，简单使用传统适用于分布均匀的分类算法将不能有效适用于作文的自动分类。以大学英语四级主题为“Global Shortage of Fresh Water”290 篇英文作文为例，分别对其进行基于信息增益的特征选择，特征维数为 100，使用多项式朴素贝叶斯进行评分结果如表 3 所示。在该结果中，2 分档和 5 分档的召回率和 F 度量值均为 0.4 和 0.5 左右，低于大类数据 3 分档和 4 分档的 R 值和 F 值近 30%，这说明传统基于均匀分布样本的分类算法不能很好的应用于作文的自动评分上。

表 3　多项式朴素贝叶斯分类算法在 Global Shortage of Fresh Water 主题 290 篇作文上的分类结果

分数档	作文个数	P	R	F
2 分档	20	0.692	0.45	0.545
3 分档	127	0.772	0.827	0.798
4 分档	122	0.764	0.795	0.779
5 分档	21	0.643	0.429	0.514

为了改变作文特征受大类作文数据的影响，本文提出了一种基于多次随机抽样及算法组合的不平衡数据分类评分方法，所提出的算法通过多次不充分抽样来平衡数据样本，并对多次评分结果进行投票获得最终评分结果，具体算法描述如下：

输入：待预测作文 x，不充分抽样次数 m；

输出：预测类别；

1. for(int $i=1$; $i\leqslant m$; $i++$)

2. {

3. 利用 Bagging 方法随机从大类数据中抽样得到与小类数据大小相同的样本；

4. 将小类作文数据与抽样得到的大类作文数据合并作为训练集；

5. 使用基于多项式模型的朴素贝叶斯分类算法对训练集进行分类，得到分类结果 y_i；

6. }

7. 利用投票策略，得到作文 x 的最终评分类别为 y'，$y'=\arg\max_{v}\sum_{1\leqslant i\leqslant m} I(v=y_i)$；其中，$v$ 是不同作文分数档，y_i 是某一个随机不充分抽样的分类结果，I(.)为指示函数，若其参数为真返回 1，否则返回 0；

8. 返回预测作文 x 的预测类别 y'。

5 实验结果

采用由桂诗春和杨慧中老师主编的中国学习者英语语料库（Chinese Learner English Corpus，CLEC）[16]作为测试数据，该语料库包含了大学英语四级和大学英语六级等不同级别考试的作文，并对所有作文进行了手工错误标注和分数归类。考虑到实际计算机作文评分时，是不包含有错误标注信息的，我们对所测试的不同主题的作文的错误标注信息进行了清除，使其尽量保持原始作文状态。

为了较为全面地测试本文的评分算法，分别选取了 CLEC 语料中大学英语四级作文库（ST3 子库）和大学英语六级作文库（ST4 子库）中来自四个主题的共计 1 115 篇英文作文进行评分测试。其中大学英语四级作文选取了 Global Shortage of Fresh Water 主题作文 290 篇和 Getting to Know the World Outside the Campus 主题作文 202 篇。大学六级作文选取了 Haste Makes Waste 主题作文 341 篇和 My View on Job-Hopping 主题作文 282 篇。按照大学英语四、六级的评分标准，进行评分时先把作文划分成 5 个分数等级，这 5 个登记分别是 2 分档，5 分档，8 分档，11 分档，14 分档。在本研究中，由于所有作文语料没有 2 分档的作文，为此本文将评分范围划分成了 4 个分数段：5 分档、8 分档、11 分档和 14 分档，并将其当作类标号，所测试的 1 115 篇作文按照分数档划分的详细信息如表 4 所示。

表 4 1 115 篇作文按分数段划分分布表

作文		4～6 分档	7～9 分档	10～12 分档	13～15 分档
四级作文	Global Shortage of fresh Water (290 篇)	20	127	122	21
	Getting to Know the World Outside the Campus (202 篇)	19	108	58	17
六级作文	Haste Makes Waste (341 篇)	51	210	80	无
	My View on Job-Hopping (282 篇)	62	175	45	无

算法评估指标采用分类准确率 P、召回率 R 和 F 度量值来进行评价，对某个分数或分数档类别 c_i，该类别样本分类的正确率 $Accuracy$、准确率 P、召回率 R、F 度量值的定义如下：

$$P=\frac{N_{c_i\to c_i}}{N_{c_i\to c_i}+N_{c_j\to c_i}}, \tag{6}$$

$$R=\frac{N_{c_i\to c_i}}{N_{c_i\to c_i}+N_{c_i\to c_j}}, \tag{7}$$

$$F=2\cdot\frac{P\cdot R}{P+R},(i\neq j) \tag{8}$$

$$Accuracy=\frac{\text{正确分类的作文个数}}{\text{总的作文个数}} \tag{9}$$

其中，$N_{c_i\to c_i}$ 表示类别 c_i 中正确分类的作文个数，$N_{c_j\to c_i}$ 表示属于类别 c_j 且别分类为类别 c_i 的作文个数，$N_{c_i\to c_j}$ 表示属于类别 c_i 且别分类为类别 c_j 的作文个数，其中 c_i 和 c_j 表示不同的分数或分数段类别。

整个实验是在一台配置 32 位 Win7 操作系统，安装内存为 2G，处理器为 Intel Celeron G530 2.4GHz的台式机器上进行，程序用VC++ 6.0实现，所有实验结果均采用十则交叉验证后得到。

表 5 为本文算法与传统多项式朴素贝叶斯分类

算法在大学英语四、六级不同不平衡作文数据中的分类结果对比，为了让结果具有可对比性，所有数据均为提取特征维数为 100 时的结果。从该结果可以看出，本文算法在特征维数为 100 的基础上，不仅平均分类评分精度、召回率及 F 度量值有所提高，其中 F 度量值均提高 10%以上，且大类作文数据和小类作文数据都取得了较平均和较好的 P、R 和 F 值，这证明本文的算法是有效可行的。

表 5　本文算法与多项式朴素贝叶斯算法在四、六级作文上的分类结果对比

作文题目	分数档	作文篇数	多项式朴素贝叶斯			本文算法					
						单词特征			相邻词组特征		
			P	R	F	P	R	F	P	R	F
Global Shortage of Fresh Water(四级作文)	2 分档	20	0.692	0.45	0.545	0.824	0.7	0.757	0.87	1	0.93
	3 分档	127	0.772	0.827	0.798	0.781	0.676	0.725	0.892	0.892	0.892
	4 分档	122	0.764	0.795	0.779	0.741	0.896	0.811	0.881	0.771	0.822
	5 分档	21	0.643	0.429	0.514	0.789	0.714	0.75	0.75	0.857	0.8
	平均值		**0.718**	**0.625**	**0.659**	**0.783**	**0.747**	**0.761**	**0.848**	**0.88**	**0.861**
Getting to Know the World Outside the Campus(四级作文)	2 分档	19	0.786	0.647	0.71	0.833	0.882	0.857	0.75	0.882	0.811
	3 分档	108	0.782	0.898	0.836	0.818	0.9	0.857	0.842	0.8	0.821
	4 分档	58	0.694	0.586	0.636	0.947	0.857	0.9	0.818	0.857	0.837
	5 分档	17	0.733	0.579	0.647	0.889	0.842	0.865	0.938	0.789	0.857
	平均值		**0.749**	**0.678**	**0.707**	**0.872**	**0.870**	**0.869**	**0.837**	**0.832**	**0.831**
Haste Makes Waste (六级作文)	2 分档	51	0.795	0.608	0.689	0.897	0.686	0.778	0.84	0.824	0.832
	3 分档	210	0.826	0.881	0.853	0.783	0.857	0.818	0.875	0.778	0.824
	4 分档	80	0.654	0.638	0.646	0.855	0.937	0.894	0.789	0.889	0.836
	平均值		**0.758**	**0.709**	**0.729**	**0.845**	**0.827**	**0.83**	**0.835**	**0.83**	**0.83**
My View on Job-Hopping(六级作文)	2 分档	62	0.756	0.5	0.602	0.855	0.855	0.855	0.941	0.774	0.85
	3 分档	175	0.805	0.966	0.878	0.855	0.929	0.89	0.831	0.914	0.871
	4 分档	45	0.613	0.422	0.5	0.897	0.778	0.833	0.796	0.867	0.83
	平均值		**0.725**	**0.629**	**0.66**	**0.869**	**0.854**	**0.859**	**0.856**	**0.852**	**0.85**

表 6 显示了多项式朴素贝叶斯分类算法和本文算法在自动计算最佳特征维数时所得到分类评分结果的对比，其中本文算法中全部采用基于二元词组特征。从实验结果可以看出，本文所提出的算法无论是在 P 值、召回率或是 F 度量值上都是最优的，并且相比传统多项式朴素贝叶斯分类算法在单词特征上的分类结果具有较大幅度的提升。

表 6　不同算法在 1 119 篇作文上的分类结果对比

作文主题	自动提取特征维数	多项式模型朴素贝叶斯分类算法			本文算法		
		P	R	F	P	R	F
Global Shortage of Fresh Water	400	0.855	0.845	0.837	0.989	0.988	0.988
Getting to Know the World Outside the Campus	360	0.785	0.777	0.762	0.988	0.987	0.987
Haste Makes Waste	700	0.877	0.865	0.856	1	1	1
My View on Job-Hopping	550	0.843	0.794	0.738	0.986	0.985	0.985
平均值		**0.84**	**0.82**	**0.798**	**0.991**	**0.99**	**0.99**

6 结论

利用计算机实现作文自动评分是自然语言处理领域一个比较崭新的研究方向，它拥有广阔的应用前景。本文结合中国学生受汉语影响以及所特有的写作特点，如介词掌握不好、短语搭配掌握不好等特征，提出了适用于中国英语学习者以及不平衡分布作文数据的集成分类评分算法，通过在CLEC语料库中大学英语四级和六级一共1 115篇作文中的分类评分结果显示，所提出的算法相比传统面向分布均匀数据的分类方法具有较高的准确率，能够有效应用于中国英语学习者的作文自动评分中。另外，由于本文的实验数据仅限于大学四、六级作文数据，并且每篇主题作文均不超过400篇，样本量还是比较小，在接下来的工作中，将继续探讨高考英文作文的分类评分处理以及大样本作文数据下的评分效果分析。

参考文献

[1] Sherm is, M. D., J. Burstein. Automated Essay Scoring: Cross-disciplinary Perspective. Computational Linguistics[J]. 2004,30(2):245-246.

[2] Rudner, Lawrence, Phill Gagne. An overview of three approaches to scoring written essays by computer. Practical Assessment [J], Research & Evaluation, 2001, 7(26).

[3] S Valenti, F Neri, A Cucchiarelli. An Overview of Current Research on Automated Essay Grading [J]. Journal of Information Technology Education, 2003, 2 (1):319-330.

[4] Hamp-Lyons L. On Second Language Writing[M]. Lawrence Erlbaum Associates,2001.

[5] Kukich K. Beyond Automated EssayScoring[C]//Proceedings of the debate on automated essay grading. IEEE Intelligent systems,2004:22-31.

[6] Hamp-Lyons L. Fourth Generation Writing Assessment[M]. Lawrence Erlbaum Associates,2001.

[7] Weigle S C. Assessing writing[M]. Cambridge University Press,2002.

[8] Shermis M, Mzumara H R, Olson J, et al. On-line grading of student essays: PEG goes on the world wide web[J]. Assessment & evaluation in higher education, 2001, 26(3).

[9] 梁茂成，文秋芳. 国外作文自动评分系统评述及启示[J]. 外语化教学，2007，17(2)：18-24.

[10] Dikli S. Automated Essay Scoring[J]. Turkish Online Journal of Distance Education, 2006,7(1).

[11] Yigal Attali, Jill Burstein. Automated Essay Scoring With E-rater v. 2. 0[M]. Princeton,2005.

[12] 方清. 中西方思维模式的不同及其对中国学生英语作文的影响[D]. 中山大学，2003.

[13] 马广惠. 中美大学生英语作文语言特征的对比分析. 外语教学与研究. 2002,34(5):345-380.

[14] 葛诗利，陈潇潇. 大学英语作文自动评分研究中的问题及对策[J]. 山东外语教学，2009，3：21-26.

[15] Jill Burstein, Martin Chodorow. Automated Essay Scoring for Nonnative English Speakers [C]//Proceedings of a Symposium on Computer Mediated Language Assessment and Evaluation in Natural Language Processing. 1999: 68-75.

[16] 桂诗春，杨惠中. 中国学习者英语语料库[M]. 上海外语教育出版社，2003.

[17] Yang Yiming. A comparison study on feature selection in text categorization [C]//Proceedings of the Fourteenth International Conference on Machine Learning (ICML 1997), Nashville, Tennessee, USA, July 8-12, 1997: 412-420.

[18] Salton G, Wong A, Yang C S. A vector space model for automatic indexing [J]. Communications of ACM, 1975, 18(11):613-620.

[19] G Salton, C Buckley. Term-weighting approaches in automatic text retrieval. Information Processing and Management [J]. 1998, 24 (5):513-523.

[20] Andrew Mccallum, Kamal Nigam: A Comparison of Event Models for Naive Bayes Text Classification. In: AAAI-98 Workshop on 'Learning for Text Categorization', 1998.

[21] Rudner L M, Liang T. Automated essay scoring using Bayes' Theorem [J]. The Journal of Technology, Learning and Assessment, 2002: (2).

[22] Larkey L, Croft W B. A Text Categorization Approach to Automated Essay Scoring[C]//Proceedings of Shermis M. D. and Burstein J. (eds.), Automated Essay Scoring: A Cross-Disciplinary Perspective, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 2003: 55-70.