

文章编号: 1003-0077(2018)01-0001-08

英语学习者书面语法错误自动检测研究综述

刘磊¹, 梁茂成²

(1. 燕山大学 外国语学院, 河北 秦皇岛 066004;

2. 北京外国语大学 中国外语与教育研究中心, 北京 100089)

摘要: 英语学习者书面语法错误检测和修改系统可为作文自动评分提供参数, 评测作文整体质量; 也可用于计算机辅助英语教学, 为学生提供书面纠错反馈, 促进其二语写作能力的发展。该文概述了近十年来自然语言处理技术在英语学习者语法错误自动检测研究中的应用, 首先介绍了基于大规模本族语和学习者语料库的三种数据驱动的系统设计方法, 然后讨论了语误检测系统的评测标准, 最后提出了提高现有系统准确率的一些建议。

关键词: 英语学习者书面语; 语法错误自动检测; 作文质量自动评测

中图分类号: TP391

文献标识码: A

A Survey of Automatic Grammar Error Detection for English Learners' Writing

LIU Lei¹, LIANG Maocheng²

(1. School of Foreign Studies, Yanshan University, Qinhuangdao, Hebei 066004, China;

2. National Research Centre for Foreign Language Education, Beijing Foreign
Studies University, Beijing 100089, China)

Abstract: Automatic detection and correction of grammatical errors in EFL learners' writing can help perform automatic evaluation of the essay quality and assist learners by providing written corrective feedback to facilitate autonomous learning. In this paper, we provide a survey of the application of NLP technology in automatic grammar error correction of English learners' writing in the last decade. Firstly, we introduce the large-scale native and learner English corpora publically available and three data-driven methods for the automatic grammatical correction system. Then the development of system evaluation methods is discussed. Lastly, we conclude with some suggestions for future directions.

Key words: EFL learners' writing; automated grammatical error correction; automated essay evaluation

0 引言

在英语教学和测试领域, 作文是检测英语学习者语言能力的重要指标。目前, 学习者作文通常依靠教师或评分员人工审阅, 这一过程需要耗费大量的人力和物力, 同时很难保证作文评测的信度和效度^[1]。

为了克服上述弊端, 国内外学者近年来开始借助自然语言处理技术, 利用计算机自动评测学习者的作文质量^[2-4]。其中, 语法错误的自动检测和修改

是作文自动评测的重要环节, 能为学习者提供书面纠错反馈, 并提高其自主学习意识^[5-6]。

自 20 世纪 80 年代起, 英语书面语法错误自动检测和修改研究 (Grammar error correction, GEC) 已经历了三个阶段的发展^[7]: 第一代 GEC 系统采用简单的字符串匹配和替换识别、修改错误, 如 Unix 系统上的 Writer's Workbench 工具, 第二代系统使用人工编纂的语法规则对文本进行句法分析, 通过编写错误模板匹配和纠正语法错误, 其代表是 IBM 公司开发的 Epistle 和 Critique 工具, 以及微软公司 Word 字处理软件中的语法检查模块, 随着大数据

收稿日期: 2017-05-24 定稿日期: 2017-06-08

基金项目: 河北省社会科学基金项目 (HB17YY046)

的涌现和计算机数据处理能力的提高;第三代系统采用数据驱动的方法,从大规模本族语或学习者语料库中提取词汇-句法特征,通过机器学习算法自动构建统计模型检测语法错误,如微软公司开发的 ESL Assistant 系统。

近十年来,GEC 领域出现了一些新的研究方法,这些方法以数据驱动的统计模型为主,主要涉及以下内容:①使用海量基于网络的英语本族语者语料库构建语言模型检测语法错误;②采用基于统计的机器翻译模型提高系统性能;③重视英语学习者语误类型与其母语/L1 的关系,在统计模型中加入学习者的 L1 信息提高语误检测的准确率;④针对不同错误类型选用不同检测方法,主流的统计模型与传统基于规则的方法相结合。

同时,为了更加客观地评测基于不同范式开发的语误检测系统性能,自 2011 年起,GEC 领域的学者先后组织了四次系统评测任务,并发布了训练和测试系统性能的学习者语料库,以保证参评团队在使用相同数据的前提下开发系统^[8-11]。这些系统评测任务部分解决了以往研究中数据不一致的问题,推动了 GEC 研究的进展。

本文首先介绍英语学习者和本族语者语料库的标注体系和规模与 GEC 研究的关系,接着分析基于统计和基于规则的研究方法,以及 GEC 系统的评测标准,最后阐述学习者语误检测的研究趋势,并提出进一步提高系统准确率的几点建议。

1 GEC 研究中语料库

1.1 学习者语料库

英语学习者语料库对 GEC 研究至关重要。研究者可借助人工标注和修改错误后的语料归纳语法错误类别;还可将其作为训练和测试语料,构建错误检测模型和评测系统性能。例如 Leacock 等^[12]根据剑桥学习者语料库(cambridge learner corpus, CLC)统计了英语学习者书面语中的常见错误类型及频率分布(图 1)。CLC 语料库规模较大,约 5 000 万词,包含错误类型、错误批改及学习者背景(如年龄、性别和国籍等)的标注,非常适合 GEC 系统的训练和评测^[13]。遗憾的是,由于版权原因,CLC 语料库无法公开,目前只有少数学者使用该语料库开展语误检测研究。为了改善这一状况,近年来,从事 GEC 研究的学者陆续建立和公布了一批可供研究

者免费使用的学习者语料库,如表 1 所示。

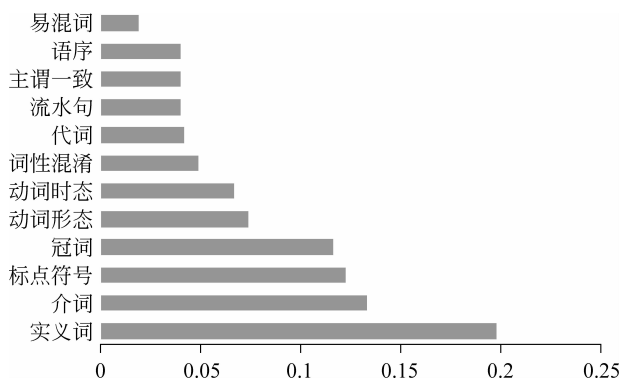


图 1 CLC 语料库中的免费语法错误类型分布

UICLE 语料库是首个公开的标注母语背景/L1 和错误类型的学习者语料库^[14]。Rozovskaya^[15]根据 UICLE 语料库中不同 L1 学习者的错误分布概率优化了朴素贝叶斯模型的参数,提高了错误识别和修改的准确率。FCE 语料库由剑桥 FCE 考试作文构成,包含多达 75 种错误类型^[16]。在实际应用中,研究者通常对其错误赋码进行归并,将其作为训练或测试语料开发 GEC 系统^[17]。

表 1 GEC 研究中的英语学习者语料库^①

语料库	文本/句子数	词数	错误种类	母语种类
UICLE	2 646 句	6 万词	7	9
FCE	1 244 篇	56 万词	75	16
Lang-8	99 074 篇	1 200 万词	—	50
NUCLE	1 400 篇	120 万词	28	1

近期研究表明,GEC 系统的性能与学习者语料库的规模关系密切:训练语料的规模越大,系统的准确率也越高^[18]。因此,为了弥补 UICLE 和 FCE 语料库规模较小的不足,近年来出现了一些大型的英语学习者语料库,如 Lang-8 语料库^[19]。Lang-8 语料库由来自 50 余个国家的英语学习者作文及其批改构成,这些作文均选自在线语言学习交流网站 lang-8.com。该网站用户使用所学外语写文章,交给以此外语为母语的其他用户批改。自发布以来,Lang-8 语料库为 GEC 系统的研制提供了新的语料和研究视角,推动了机器翻译模型在 GEC 研究中的应用^[20]。

① Lang-8 语料库只修改了作文中的语法错误,未标注错误类型。

NUCLE 语料库是 2013 和 2014 年两届 CoNLL 大会 GEC 系统评测任务的训练和测试语料^[21]。与上述语料库使用单一标注人员标注和修改语法错误不同,NUCLE 的测试语料由两名英语母语者共同标注。用多人标注的语料作为评测标准充分考虑了错误修改的多种可能性,可以更加客观地评价 GEC 系统的性能。同时,NUCLE 的开发团队还公布了评测 GEC 系统性能的工具 Max Match /M²。目前,这一工具已成为计算 GEC 系统准确率和召回率的标准,是与人工错误批改相关度较高的评测方法^[22]。

1.2 本族语者语料库

与英语学习者语料库相比,本族语者语料库更易于建立,因此其规模更大,种类也更多,如 BNC 语料库、Gigaword 语料库和 Wikipedia 语料库等^[12]。在 GEC 研究中,研究者通常对上述语料进行预处理,如分词、词性标注和句法标注等,以便提取语言特征构建统计模型。例如,Chodorow 等^[23]在检测英语学习者介词错误时选用了介词周围的名词短语、动词短语和 n 元序列等特征构建了自动分类模型。

在基于 n 元语法的 GEC 研究中,研究者需要从本族语者语料库提取 n 元序列的频率和概率,构建语言模型。为了解决数据稀疏问题,近年出现了一批大规模基于网络的 n 元序列语料库,如表 2 所示。

表 2 GEC 研究中的英语本族语者语料库

	Google ^[24]	Microsoft ^[25]	Common Crawl ^[26]
1 元组	0.14 亿	12 亿	26 亿
2 元组	3.15 亿	117 亿	153 亿
3 元组	9.77 亿	601 亿	619 亿
4 元组	13.14 亿	1 485 亿	1 568 亿
5 元组	11.8 亿 ^①	2 370 亿	2 637 亿

表 2 中的 n 元序列从网络获取,其规模较传统语料库增大不少。为了高效检索海量数据中的 n 元序列,Heafield^[27]以及 Pauls 和 Klein^[28]开发了专门针对大规模 n 元序列语料库的数据压缩和索引算法,极大地推进了 n 元语法模型在 GEC 研究中的应用。

2 GEC 研究中的常用方法

随着学习者和本族语者语料库的数量和规模持续扩大,GEC 领域出现了一些新的研究方法,这些方法以数据驱动的统计模型为主,具体可分为三类:(1) n 元语法模型;(2)自动分类模型;(3)机器翻译模型。在实际应用中,统计方法常与传统的基于规则的方法相结合,以充分发挥两者的优点,开发规则与统计的 GEC 系统。图 2 是 GEC 研究方法与本族语者和学习者语料库的关系图。

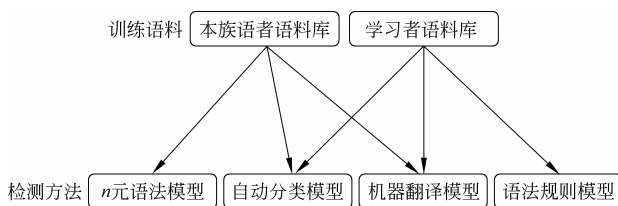


图 2 训练语料与检测方法间的关系

2.1 n 元语法模型

n 元语法模型的最初目的是利用统计学中的马尔科夫链预测句子的出现概率,其核心是根据句子中每个词 w_n 与其之前的语境 $(w_1, w_2, \dots, w_{n-1})$ 计算词序列概率。基于 n 元语法设计的自动语法错误修改系统使用语法正确的文本即本族语者语料库构建模型,如果输入文本的词序列概率过低,则将其判定为语法错误^[29]。Gamon 等^[30]利用该模型检测了英语学习者书面语中涉及冠词、介词、动词和形容词的八类语法错误,其检测方法包括两个步骤:(1)确定易混词/短语集合 C ;(2)依次计算包含词或短语 $c(c \in C)$ 的句子概率,并建议学习者按照概率最大的句子修改原文。例如,在检测例句“*I want to travel Disneyland in March.*”中的介词使用错误时,系统首先确定易混词集合 $C = \{on, from, for, of, about, to, at, in, with, by, \emptyset\}$, \emptyset 表示介词省略;替换集合中其他介词并计算句子概率后可知使用介词 *to* 时概率最大,因此建议将“*travel Disneyland*”修改为“*travel to Disneyland*”。

近年来,随着训练语料规模的扩大,基于 n 元语法模型开发的 GEC 系统也逐渐增多,在 CoNLL2013 和 CoNLL2014 两次系统评测任务中取得

① Google n -gram 语料库只收录了频率大于 40 的 n 元序列。因此,5 元组的数量少于 4 元组的数量。

了较好的效果^[31-32]。除统计方法外,上述 GEC 系统均使用了语法规则模板以提高系统性能。例如 Lee 和 Lee^[33]首先从 NUCLE 学习者语料库中提取易混词/短语列表,如{rise, raise}、{well, good}等,然后再计算句子概率,提升了 n 元语法模型修改实义词错误的准确率。

2.2 自动分类模型

在 GEC 研究中,自动分类模型常用于冠词、介词等虚词错误,以及动词时态和形态错误的自动检测和修改^[34]。利用自动分类模型设计自动语法检查系统时需要选择合适的语言特征进行训练,通过训练语料求得各语言特征的权重,预测句子中出现语法错误的概率。常见的语言特征有词形、词性、句法和语义信息等。自动分类模型包括多种算法,如多元逻辑回归、支持向量机和朴素贝叶斯算法等^[35-36]。例如,在检测介词错误时,朴素贝叶斯模型将介词周围的词汇和句法结构等作为特征向量 f 输入模型,输出结果为介词的类别 c ,如式(1)所示。

$$\begin{aligned}\hat{c} &= \operatorname{argmax}_c P(c | f) \\ &= \operatorname{argmax}_c P(f | c)P(c)\end{aligned}\quad (1)$$

早期基于朴素贝叶斯模型的介词修改系统使用英语本族语者语料库作为训练语料,从中提取介词周围的语言特征概率 $P(f | c)$ 和介词出现的先验概率 $P(c)$ 。Rozovskaya 和 Roth^[37]在统计 UICLE 语料库中的介词错误后发现,英语学习者介词类语法错误的分布与其母语关系密切。因此,研究者修改了式(1)所示的朴素贝叶斯模型,在计算先验概率时考虑了学习者母语背景对介词误用的影响,如式(2)、式(3)所示。

$$\begin{aligned}\hat{c} &= \operatorname{argmax}_c P(c | f) \\ &= \operatorname{argmax}_c P(f | c)P(c, s, L_1)\end{aligned}\quad (2)$$

$$P(c, s, L_1) = \frac{C_{L_1}(s, c)}{C_{L_1}(s)}\quad (3)$$

其中 s 为学习者使用的介词, c 是修改后的介词; C_{L_1} 函数通过学习者语料库中介词 s 和 c 的频率计算先验概率,提高了介词修改的准确率。

自动分类模型一般根据具体错误类型确定特征集合,从训练语料提取特征并通过机器学习算法获取特征参数,构建针对特定错误的语言模型。因此,与 n 元语法模型类似,基于自动分类模型设计的 GEC 系统通常由多个检测模块构成,各个模块按线性顺序修改文本中的语法错误。

2.3 机器翻译模型

基于不同语法错误检测模块开发的 GEC 系统如 n 元语法模型和自动分类模型无法处理例句“*Social network plays a role in providing and also filtering information.*”中的语法错误^[12]。该句中包含两处明显的语法错误:名词单复数错误(network→networks)和主谓不一致错误(plays→play)。如果 GEC 系统将主谓一致模块置于名词单复数模块之前,系统将无法得到正确的修改结果。为了检测包含多处语法错误的句子,有学者开始借鉴统计机器翻译模型开发 GEC 系统。

式(4)是统计机器翻译模型的表达式。

$$\begin{aligned}\hat{e} &= \operatorname{argmax}_e P(e | f) \\ &= \operatorname{argmax}_e P(f | e)P(e)\end{aligned}\quad (4)$$

其中, f 是源语, e 是目标语。该模型从双语平行语料库中自动抽取基于短语的双语词典,计算翻译模型参数 $P(f | e)$;从单语语料库中提取 n 元序列计算目标语 e 的概率 $P(e)$ 。

基于机器翻译模型设计的 GEC 系统将学习者的语言产出视作源语,将修改语法错误后的文本等同目标语,通过大规模经人工修改错误的学习者语料库(如 Lang-8 和 NUCLE 语料库)训练统计模型参数,大幅度提高了语法错误修改系统的准确率^[38]。Rozovskaya 和 Roth^[39]归纳了自动分类和机器翻译模型的优缺点,在训练语料、特征选取、错误类型和泛化能力等层面对比了两类系统的异同,如表 3 所示。

表 3 自动分类和机器翻译模型对比

	自动分类模型	机器翻译模型
训练语料标注	错误类型 错误修改	错误修改
特征选取	人工	自动
错误类型覆盖	低	高
复杂错误检测	否	是
泛化能力	强	弱

从表 3 可以看出,机器翻译模型无需标注训练语料中的语法错误类型,降低了学习者语料库建设的时间和人力成本。同时,机器翻译模型可自动选取特征训练模型,更加擅长处理复杂错误。自动分类模型根据具体错误类型人工选取词汇、短语和句法层面的语言特征,因此具有较强的泛化能力,能够

检测未出现在训练语料中的新样本所包含的语法错误。由于两种模型可以处理不同类型的语法错误,融合两种模型开发的混合模型准确率高于基于单一模型构建的 GEC 系统。

3 GEC 系统的评测

GEC 系统使用人工标注语法错误的学习者语料库作为评测标准,通过对比机器批改与人工批改的异同衡量系统的准确率和召回率。由于语法错误的标注和修改需要耗费大量的人力,为了提高标注效率,早期的 GEC 研究一般使用单人标注测试语料中的语法错误。但近期研究发现,由于本族语者对学习语法错误的标注和修改存在差异,为了更准确地测试 GEC 系统的性能,应尽可能使用多个标注人员标注测试语料。表 4 是学习者语法错误标注信度研究的测试结果,其中 Kappa 系数是测量标注人

表 4 语法错误标注的一致性^①

	文献[14]	文献[21]	文献[40]
语料库	UICLE	NUCLE	—
错误类型	冠词、介词、名词、动词等 7 类错误	词法、句法、搭配等 28 类错误	介词
Kappa 系数	0.40	0.48	0.63

员间标注一致性的统计值,Kappa 值越大表明一致性越高。

从该表可以看出,语法错误标注的一致性普遍较低,尤其当语法错误类别增加时,标注人员间的分歧也随之增大。虽然个别语法错误(如主谓不一致)的标注信度较高,涉及冠词和介词使用错误的标注存在较大差异。因此,为了更加客观地对比不同 GEC 系统的准确率和召回率,多数研究者主张使用多人标注的学习者语料库作为评测集^[41]。为了提高多人标注的效率,可以借助网络众包标注平台如 Amazon Mechanical Turk 和 CrowdFlower 加快标注速度^[42]。

在得到人工批改的标准集后,GEC 系统一般使用 M^2 算法对比系统修改与人工修改的异同,计算系统的准确率和召回率。以学习者原文“*There is no a doubt, tracking system has bought many benefits in this information age.*”为例,如果人工修改集合 $g = \{a \text{ doubt} \rightarrow \text{doubt}, \text{system} \rightarrow \text{systems}, \text{has} \rightarrow \text{have}\}$,系统修改集合 $e = \{a \text{ doubt} \rightarrow \text{doubt}\}$,则该系统的准确率为 $|g \cap e| / |e| = 100\%$;召回率为 $|g \cap e| / |g| = 33.3\%$ 。表 5 列举了 CoNLL-2014 GEC 评测任务中各类错误检测的最高召回率及各个系统在检测具体语法错误时所使用的检测方法和训练语料。

表 5 CoNLL-2014 GEC 系统评测结果

语法错误类型	召回率/%	方法	训练语料
动词时态/Vt	19.61	规则+ n 元语法模型	Google N-gram
情态动词/Vm	20.51	规则+ n 元语法模型	Google N-gram
动词缺失/V0	15.19	规则+机器翻译模型	Mircosoft N-gram;FCE
动词形式/Vform	18.99	规则+ n 元语法模型	Google N-gram
主谓一致/SVA	57.67	自动分类模型	Google N-gram
冠词/ArtOrDet	47.87	自动分类模型	Google N-gram
名词单复数/Nn	46.18	自动分类模型	Google N-gram
名词所有格/Npos	11.11	机器翻译模型	Lang-8;CommonCrawl
代词格/Pform	15.52	规则+机器翻译模型	Mircosoft N-gram;FCE
代词指代/Pref	9.35	规则+机器翻译模型	Mircosoft N-gram;FCE
介词/Prep	24.94	规则+机器翻译模型	Mircosoft N-gram;FCE
搭配/Wci	6.79	机器翻译模型	Lang-8;CommonCrawl

^① 作者未公开测试语料。

续表

语法错误类型	召回率/%	方法	训练语料
缩写/Wa	0.00	—	—
词性混用/Wform	36.57	机器翻译模型	Lang-8;CommonCrawl
措辞不当/Wtone	39.13	机器翻译模型	Lang-8;CommonCrawl
流水句/Srun	0.00	—	—
状语从句/Smold	0.00	—	—
并列句/Spar	50.00	规则+ n 元语法模型	—
不完整句/Sfrag	7.14	规则+自动分类模型	—
定语从句错误/Ssub	11.27	规则+机器翻译模型	Mircosoft N-gram;FCE
疑问句语序/WOinc	2.86	机器翻译模型	WMT2014
形容词/副词语序/WOadv	29.41	规则+机器翻译模型	Mircosoft N-gram;FCE
连词/Trans	11.97	规则+机器翻译模型	Mircosoft N-gram;FCE
拼写/标点/Mec	22.62	规则+ n 元语法模型	Google N-gram
多余/Rloc	11.41	规则+机器翻译模型	Mircosoft N-gram;FCE
意思模糊/Um	8.16	n 元语法+自动分类模型	Gigaword

4 总结和展望

由于 GEC 系统需要处理英语学习者书面语中的多种语法错误,目前系统的准确率和召回率仍然较低。例如在 CoNLL-2014 的 GEC 系统评测任务中,排名第一系统的准确率和召回率分别为 39.71% 和 30.10%。本文认为,GEC 系统的性能仍有提升空间。为进一步提高自动语法错误检测系统的性能,需要解决以下几个问题。

(1) 加强语法规则在统计模型中的应用

虽然统计模型以大规模真实语料为知识来源,避免了人工编写规则的繁琐,但这种方法对语言的层级结构考虑不够,无法检测涉及长距离语法关系的错误。因此,在设计 GEC 系统时应结合上述两种方法的优点:使用统计模型检测涉及冠词、介词和搭配类等与语境关系密切的语法错误;使用基于语法规则的错误模板处理句法类错误,如主谓不一致、连词缺失和流水句错误等。鉴于英语句法自动分析技术日益成熟,具有较高的准确率和健壮性,可以使用句法分析工具提取学习者文本中的主语-谓语、动词-宾语、助动词-动词关系和主句-状语从句等语法关系,为后续错误模板的编写和统计模型特征的提取奠定基础。例如识别主语-谓语关系可辅助编写主谓不一致错误模板;主句-状语从句关系可作为语

言特征之一构建动词时态错误检测模型。

以表 5 所示的主谓不一致错误为例,在 CoNLL-2014 参赛团队所提交的系统中,使用基于统计的自动分类模型检测该类错误的最高召回率为 57.67%。本文认为,可以通过编写基于语法规则的错误模板弥补统计模型的不足,进一步提高主谓不一致错误检测的准确率。在编写错误模板时,应充分考虑英语主语内部结构的复杂性,分析主语为并列结构、数量短语+名词、集体名词和从句时的情况。

(2) 优化统计模型中语言特征的选取

基于统计模型的 GEC 系统根据具体错误类型从训练语料提取语言特征,构建针对特定错误的语言模型。现有 GEC 研究通常在句子内部选取词或多词序列作为语言特征训练模型,无法识别和修改涉及语篇层面的语法错误。例如在检测动词时态错误时,除了考虑同一句子中动词周围的词和短语外,还应从该句前后更大的语境中提取以下语言特征:动词的语义类别、时间副词、从句类型及前后语境中动词的时态等。

(3) 重视学习者母语对语言产出的影响

语言迁移理论认为英语学习者的书面语法错误类型受其母语影响。基于学习者语料库的研究也发现,不同母语背景学习者的错误类型和分布概率存在差异。现有 GEC 系统大多选择语法正确的本族

语者语料库构建统计模型,未考虑学习者母语这一变量。随着大规模经错误标注的学习者语料库的增多,GEC 系统的设计应结合学习者和本族语者语料库,根据学习者的母语背景调整统计模型的参数。虽然文献[37]从 UICLE 学习者语料库提取了不同母语背景英语学习者的介词错误概率,并将这一信息加入朴素贝叶斯模型,提高了介词错误修改的准确率,但 UICLE 语料库规模较小,作者只研究了九类不同母语背景学习者的介词使用错误。随着学习者语料库规模的扩大(如 Lang-8 语料库),可以验证该方法是否适用于其他母语背景学习者,以及对动词、名词等实义词语法错误的检测效果。

参考文献

- [1] 梁茂成,文秋芳. 国外自动评分系统评述及启示[J]. 外语电化教学,2007(5): 18-24.
- [2] 梁茂成. 大规模考试英语作文自动评分系统的研制[M]. 北京: 高等教育出版社,2012.
- [3] 王耀华,李舟军,何跃鹰,等. 基于文本语义离散度的自动作文评分关键技术研究[J]. 中文信息学报,2016,30(6): 173-181.
- [4] Shermis M D, Burstein J, Bruskey S A. Introduction to automated essay evaluation[C]//Proceedings of Handbook of Automated Essay Evaluation: Current Applications and New Directions. New York: Routledge, 2013: 1-15.
- [5] Li J, Link S, Hegelheimer V. Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction[J]. Journal of Second Language Writing, 2015(27): 1-18.
- [6] 陈功. 学习者语法错误自动检查研究述评[J]. 语料库语言学, 2016, 3(1): 70-81.
- [7] Dale R. Checking in on grammar checking[J]. Natural Language Engineering, 2016, 22(3): 491-495.
- [8] Dale R, Kilgarriff A. Helping our own: The HOO 2011 pilot shared task[C]//Proceedings of the 13th European Workshop on Natural Language Generation. Nancy: ACL, 2011: 242-249.
- [9] Dale R, Anisimoff I, Narroay G. HOO 2012: A report on the preposition and determiner error correction shared task[C]//Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications. Montreal: ACL, 2012: 54-62.
- [10] Ng H T, Wu S M, Wu Y, et al. The CoNLL-2013 shared task on grammatical error correction[C]//Proceedings of the 17th CoNLL Shared Task. Sofia: ACL, 2013: 1-12.
- [11] Ng H T, Wu S M, Briscoe T, et al. The CoNLL-2014 shared task on grammatical error correction[C]//Proceedings of the 18th CoNLL Shared Task. Baltimore: ACL, 2014: 1-14.
- [12] Leacock C, Chodorow M, Tetreault J. Automated grammatical error detection for language learners, Second Edition[M]. San Rafael, CA: Morgan and Claypool, 2014.
- [13] Nicholls D. The Cambridge learner corpus - error coding and analysis for lexicography and ELT[C]//Proceedings of Corpus Linguistics 2003 Conference. Lancaster: UCREL, 2003: 572-581.
- [14] Rozovskaya A, Roth D. Annotating ESL errors: challenges and rewards[C]//Proceedings of the 5th Workshop on Innovative Use of NLP for Building Educational Applications. Stroudsburg, Los Angeles: ACL, 2010: 28-36.
- [15] Rozovskaya A. Automated methods for text correction[D]. Urbana and Champaign: University of Illinois at Urbana-Champaign, 2013: 62-86.
- [16] Yannakoudakis H, Briscoe T, Medlock B. A new dataset and method for automatically grading ESOL texts[C]//Proceedings of the 49th Annual Meeting of ACL. Portland: ACL, 2011: 180-189.
- [17] Yuang Z, Felice M. Constrained grammatical error correction using statistical machine translation[C]//Proceedings of the 17th CoNLL Shared Task. Sofia: ACL, 2013: 52-61.
- [18] Mizumoto T, Hayashibe Y, Komachi M, et al. The effect of learner corpus size in grammatical error correction of ESL writings[C]//Proceedings of COLING 2012 Posters. Mumbai: COLING 2012 Organizing Committee, 2012: 863-872.
- [19] Mizumoto T, Komachi M, Nagata M, et al. Mining revision log of language learning SNS for automated Japanese error correction of second language learners[C]//Proceedings of the 5th IJCNLP. Chiang Mai, Thailand: Asian Federation of NLP, 2011: 147-155.
- [20] Junczys-Dowmunt M, Grundkiewicz R. The AMU system in the CoNLL-2014 shared task: Grammatical error correction by data-intensive and feature-rich statistical machine translation[C]//Proceedings of the 18th Conference on CoNLL Shared Task. Baltimore: ACL, 2014: 25-33.
- [21] Dahlmeier D, Ng H T, Wu S M. Building a large annotated corpus of learner english: The NUS corpus of learner english[C]//Proceedings of the 8th Workshop on Innovative Use of NLP for Building Educational Applications. Atlanta: ACL, 2013: 22-31.
- [22] Grundkiewicz R, Junczys-Dowmunt M, Gillian E. Human evaluation of grammatical error correction systems[C]//Proceedings of EMNLP 2015. Lisbon:

- ACL, 2015; 461-470.
- [23] Chodorow M, Tetreault J, Han N R. Detection of grammatical errors involving prepositions[C]//Proceedings of the 4th ACL-SIGSEM Workshop on Prepositions. Prague; ACL, 2007; 25-30.
- [24] Brants T, Franz A. The Google Web 1T 5-gram corpus version 1.1[DB]. Philadelphia: Linguistic Data Consortium, 2006.
- [25] Wang K, Thrasher C, Viegas E, et al. An overview of Microsoft Web N-gram corpus and applications[C]//Proceedings of the NAACL HLT 2010 Demonstration Session. Los Angeles; ACL, 2010; 45-48.
- [26] Buck C, Heafield K, van Ooyen B. N-gram counts and language models from the common crawl[C]//Proceedings of LREC 2014. Reykjavik; ELRA, 2014; 3579-3584.
- [27] Heafield K. KenLM: Faster and smaller language model queries[C]//Proceedings of the 6th Workshop on SMT. Edinburgh; ACL, 2011; 187-197.
- [28] Pauls A, Klein D. Faster and smaller n-gram language models[C]//Proceedings of the 49th Annual Meeting of the ACL. Portland; ACL, 2011; 258-267.
- [29] 谭咏梅, 王晓辉, 杨一泉. 基于语料库的英语文章语法错误检查及纠正方法[J]. 北京邮电大学学报, 2016, 39(4): 92-97.
- [30] Gamon M, Gao J, Brockett C, et al. Using contextual speller techniques and language modeling for ESL error correction[C]//Proceedings of the 3rd IJCNLP. Hyderabad; Asian Federation of NLP, 2008; 449-456.
- [31] Boros T, Dumitrescu S D, Zafiu A, et al. RACAI GEC: A hybrid approach to grammatical error correction[C]//Proceedings of the 18th CoNLL Shared Task. Baltimore; ACL, 2014; 43-48.
- [32] Kao T H, Chang Y W, Chiu H W, et al. CoNLL-2013 shared task: grammatical error correction NTHU system description [C]//Proceedings of the 17th CoNLL Shared Task. Sofia; ACL, 2013; 20-25.
- [33] Lee K, Lee G G. POSTECH grammatical error correction system in the CoNLL-2014 shared task[C]//Proceedings of the 18th CoNLL Shared Task. Baltimore; ACL, 2014; 65-73.
- [34] Tajiri T, Komachi M, Matsumoto Y. Tense and aspect error correction for ESL learners using global context[C]//Proceedings of the 50th Annual Meeting of ACL. Jeju Island; ACL, 2012; 198-202.
- [35] Rozovskaya A, Roth D, Srikumar V. Correcting grammatical verb errors[C]//Proceedings of the 14th Conference of EACL. Gothenburg; ACL, 2014; 358-367.
- [36] 李霞, 刘建达. 适用于中国外语学习者的英文作文自动集成评分算法[J]. 中文信息学报, 2013, 27(5): 100-106.
- [37] Rozovskaya A, Roth D. Algorithm selection and model adaptation for ESL correction tasks[C]//Proceedings of ACL 2011. Portland; ACL, 2011; 924-933.
- [38] Junczys-dowmunt M, Grundkiewicz R. Phrase-based machine translation is state-of-the-art for automatic grammatical error correction [C]//Proceedings of EMNLP 2016. Austin; ACL, 2016; 1546-1556.
- [39] Rozovskaya A, Roth D. Grammatical error correction; Machine translation and classifiers [C]//Proceedings of ACL 2016. Berlin, ACL, 2016; 2205-2215.
- [40] Tetreault J, Chodorow M. Native judgments of non-native usage: Experiments in preposition error detection[C]//Proceedings of the Workshop on Human Judgements in Computational Linguistics. Manchester; COLING 2008 Organizing Committee, 2008; 24-32.
- [41] Sakaguchi K, Napoles C, Post M, et al. Reassessing the goals of grammatical error correction: Fluency instead of grammaticality[J]. Transactions of the Association for Computational Linguistics, 2016(4): 169-182.
- [42] Madnani N, Tetreault J, Chodorow M, et al. They can help: Using crowdsourcing to improve the evaluation of grammatical error detection systems[C]//Proceedings of ACL 2011. Portland; ACL, 2011; 508-513.



刘磊(1980—), 博士, 讲师, 主要研究领域为计算语言学、语料库语言学。
E-mail: liulei@ysu.edu.cn



梁茂成(1964—), 博士, 教授, 主要研究领域为计算语言学、语料库语言学。
E-mail: frankliang0086@163.com