Computer Science and Application 计算机科学与应用, 2023, 13(3), 390-398
Published Online March 2023 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2023.133037

Hans汉斯

引入反馈机制的中文文本校对技术研究

杜晓童，李　崭，付萍萍，刘彦君

中国电子科技集团公司第十研究所第四事业部，四川　成都

收稿日期：2023年2月16日；录用日期：2023年3月16日；发布日期：2023年3月24日

摘　要

中文文本校对技术已取得了很大进展，然而目前很多技术研究依赖于深度学习，随着语言模型越来越复杂，训练成本迅速增加，导致落地应用较为困难。针对上述问题，本文提出了一种迭代式无监督文本自动校对技术，可同时纠正多字、少字、字序颠倒以及错别字等文本错误，并设计了反馈机制，可对校对错误结果进行反馈与实时修正。模型使用交叉位置融合算法定位错词索引，针对检测到的错词位置，采用并行多通道候选词构建策略得到候选词序列，并基于得分修正算法计算最优候选词。该方法在公开数据集SIGAHN和自构建数据集上进行了测试实验，纠正准确率和精度分别提升了6.39%和5.17%，高于Transformer等深度学习模型，且训练成本低，可作为文本自动校对技术普及应用的参考方案。

关键词

文本校对，反馈机制，位置融合，候选词策略，多类文本错误

Research on Chinese Proofreading Technology with Feed-Back Mechanism

Xiaotong Du, Zhan Li, Pingping Fu, Yanjun Liu

The Fourth Division, The 10th Research Institute of China Electronics Technology Group Corporation, Chengdu Sichuan

Received: Feb. 16th, 2023; accepted: Mar. 16th, 2023; published: Mar. 24th, 2023

Abstract

Chinese proofreading technology has made great progress. However, at present, many technical studies rely on deep learning. As the language model becomes more and more complex, the training cost increases rapidly, resulting in difficulties in landing applications. In view of the above problems, this paper proposes an iterative unsupervised text automatic proofreading technology, which can correct text errors such as multi word, few word, reversed word order and wrong type

文章引用: 杜晓童, 李崭, 付萍萍, 刘彦君. 引入反馈机制的中文文本校对技术研究[J]. 计算机科学与应用, 2023, 13(3): 390-398. DOI: 10.12677/csa.2023.133037

words at the same time, and designs a feedback mechanism to feed back and correct the proofreading error results in real time. The model uses the cross position fusion algorithm to locate the wrong word index. For the detected wrong word position, it uses the parallel multi-channel candidate word construction strategy to get the candidate word sequence, and calculates the optimal candidate word based on the score correction algorithm. The method has been tested on the public data set sigahn and the self built data set, and the correction accuracy and precision have been improved by 6.39% and 5.17% respectively, which are higher than the transformer deep learning model, and the training cost is low. It can be used as a reference scheme for the popularization and application of automatic text proofreading technology.

Keywords

Text Proofreading, Feedback Mechanism, Position Fusion, Candidate Strategy, Multiple Text Errors

1. 引言

随着信息技术的发展，人们使用电子文本信息进行数据分析的场合越来越多，对于文本内容的准确性要求越来越高，文本自动校对技术应运而生。该技术已广泛应用于各种领域，如电子法律文书[1]、语音识别后的文本[2]以及媒体新闻言论[3]等内容的纠错。文本自动校对技术逐渐成为自然语言处理领域的重要研究方向之一。

中文文本错误类型主要包含以下六种：1) 发音错误，即同音或近音字错误；2) 形似错误，指字形相似的字；3) 字序颠倒；4) 多字少字；5) 语法错误，如主谓搭配不当等；6) 知识错误，如“四川省省会绵阳”。前四种错误类型可以归属为文本字词级错误，后两种则需要对语句结构进行理解分析。目前文本校对技术大多针对前两种错误类型进行研究。

中文文本校对技术的研究[4]始于 20 世纪 90 年代，目前已有很多有效的方法被提出，且大多遵循错误检测－错误纠正的模型框架。主流文本自动校对模型共有两种，即传统语言模型[5]和深度学习语言模型[6]。前者提供了一种规则式检错、纠错方法，不需要大量标注数据从而易于实现落地，但模型过于稀疏且泛化能力差。深度学习模型可以实现长距离依赖，从而避免由数据稀疏导致的 OOV (out of vocabulary，词典溢出)问题，它使用一个端到端的网络结构依次进行检错和纠错。然而该类模型的可解释性较差，且训练强度依赖于硬件环境和标注数据，训练时间较长。

中文具有较高的复杂性[7]，普通话包含上万个常用字符，不同上下文中汉字读音以及词语使用也会发生变化，因此中文文本自动校对技术的研究是一项复杂的任务。同时在实际应用中，纠错技术常作为文本预处理中的一个环节，则其时效性也极为重要。综上所述，在标注资源、硬件条件受限的情况下，如何提高文本校对的准确率、召回率，缩短模型训练时间，扩展文本校对类型成为了当下急需解决的问题。

2. 技术现状

基于传统语言模型的中文文本自动校对技术的相关研究已有不少，石敏等人[8]提出了一种基于决策

列表的方法，使用同音词训练语言模型，然后从列表中选取正确的同音词，但该方法只能纠正同音词错误，对于其它类型错误无能为力；Yu 等人[9]使用双向字符级 N-gram 语言模型对句子中每个字符进行打分，得分低的视为错误位置；赵海等人[10]使用混淆集替换的方式做模糊匹配，并通过求解最短路径进行字词错误的校对。王琼等人[11]提出一种基于改进的 N-gram 模型和专业术语查错知识库的查错算法，增强模型对散串错误的查错率。王浩畅等人[12]基于语料库和 CRF 算法建立了 N-Gram 模型进行语法错误检测。以上这些基于规则词典、语言模型的校对方法可以很好地满足工程化时效性要求，但是准确率和召回率有待提高，而且纠错类型一般只为错别字、近音字等，并不适用于多字、少字和字序颠倒等错误。

随着深度学习的发展，越来越多的有监督语言模型被应用于文本自动校对技术。郝亚男等人[6]提出了一种基于神经网络和注意机制的中文文本校对技术，利用双向门控循环神经网络进行特征提取，可以纠正文本的语义错误；龚永罡[13]团队提出了一种基于 Seq2Seq 和 Bi-LSTM 结合的深度学习模型，可以有效地处理文本错误以及语义错误。然而 RNN 提取的是文本的序列化分布式特征，因此在长距离序列中存在信息丢失问题。龚永罡等人[14]又提出了 Transformer 模型，沿用经典的 Encoder-Decoder 结构，使用 Self-Attention 机制，提高了校对性能。百度 Ernie-CSC 文本纠错模型[15]改进了 Transformer 框架，提出了 MLM-phonetics 算法，使用海量标注数据(约 10 TB)进行预训练，又在超十万级的数据上进行迁移学习，可以很好地纠正广域文本字词级错误。基于深度学习的方法虽然在学术上取得了很大进展，但是受时间和资源的限制，很难应用落地。除此之外，端到端的模型无法在训练后进行人工干预，导致不能对错纠、误纠等结果进行反馈与修正。

针对以上问题，本文从工程应用角度出发，提出了引入反馈机制的中文文本校对技术。实验结果表明该方法提高了文本纠正准确率与精确率，兼容多种文本错误类型，大大缩短了模型训练时间，这充分证明了该方法的有效性。下面将详细介绍本文提出的技术方法。

3. 中文文本校对技术

完整的文本自动校对技术流程见图 1，首先对输入文本进行错误检测，使用交叉位置融合算法定位错词索引，然后将错误位置信息输入到基于多通道候选词构建策略以及反馈机制的文本错误纠正模型中，便得到了纠正后的输出文本。本文提出的方法加入了迭代思想，即对输入文本进行循环检错与纠错，并用纠正文本作为下一次的输入，直到纠正文本与输入文本相同，或循环次数大于 5 则停止。

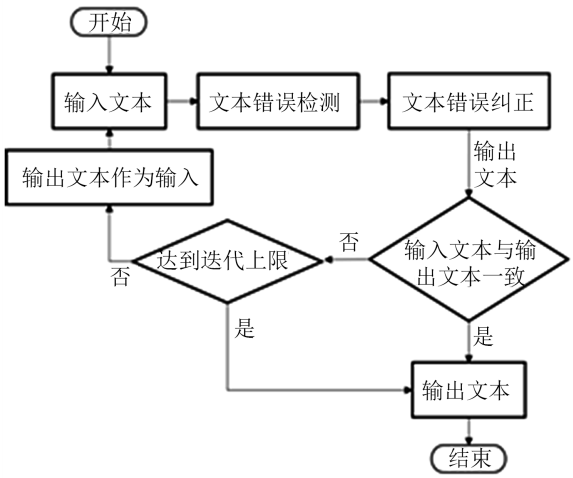

Figure 1. Automatic text proofreading process
图 1. 文本自动校对过程

下面将详细介绍技术流程中的文本错误检测、文本错误纠正两个算法步骤，以及隐含的校对错误反馈机制和中文知识库构建方法。

3.1. 文本错误检测

错误检测模型首先对输入文本进行归一化操作，即全角转半角，英文标点转为中文标点等，然后将归一化后的文本分别输入到知识库和语言模型两种检错算法中，最后将两种算法得到的错词索引位置进行交叉融合。

3.1.1. 错误索引定位

基于知识库的检错：对输入文本进行分词操作，得到分词集，对于集合中每一个长度大于 1 的中文分词，若它不是停用词，也不在常用词库中，则视其为疑似错词，记录该词的起止位置。

基于语言模型的检错：本文对原始 Berkeley 语言模型进行了优化：计算得到输入文本的 2-gram 和 3-gram 得分向量，分别记为 S_2 和 S_3，向量长度都为 N (表示输入文本的总字符数)，进而由式(1)得到整个句子的得分向量 $Score_N$。经实验发现，3-gram 得分较 2-gram 对句子错误检测具有更重要的作用，因此本文在融合 N-gram 得分时，为 3-gram 添加了权重。

$$Score_N = \left(S_2 + S_3 \times 2\right)/2 \tag{1}$$

通过平均绝对离差法(MAD)从 $Score_N$ 中计算出疑似错字的位置：如式(2)所示，可得到绝对离差向量 Mar_N，其中 S' 表示 $Score_N$ 的中位数，进而可以得到 Mar_N 的中位数 M'。若 M' 为 0，则表示没有检测到错误，否则继续计算 $ErrScore_N$。计算过程如式(3)所示，其中 RATIO 为可调节参数，本文根据实验值取 ratio = 0.6745。

$$Mar_i = \left|Score_i - S'\right|, i \in [1, N] \tag{2}$$

$$ErrScore_i = Mar_i \times \frac{RATIO}{M'}, i \in [1, N] \tag{3}$$

通过计算出的 $ErrScore_N$ 与另一个可调节参数 THRESHOLD 进行对比，判断出错误位置，判断方法如式(4)所示，其中 THESHOLD 实验取值为 2.4。最后将检测出错误的文本索引位置 $i\{i \in [1, N]\}$ 记录下来，过滤掉非中文字符以及常用词典检错中已经记录过的位置，便得到了基于语言模型的检错结果。

$$Location = \begin{cases} \text{无错}, & ErrScore_i \le \text{THRESHOLD} \\ \text{有错}, & ErrScore_i > \text{THRESHOLD} \end{cases} \tag{4}$$

3.1.2. 交叉位置融合

将基于知识库和模型分别得到的检错信息进行融合，由知识库得到的错误信息是错词的起止位置，而由模型得到的是错字位置，将索引位置连续的错词、错字连接起来，然后对连接结果进行分词操作，便得到了疑似错词候选集。

最后，使用专用词典和用户自定义错词库(知识库信息见 3.4 节)对候选集进行过滤：若疑似错词及其上下文文本构成了专用词，或者在不满足不转换规则的情况下构成了自定义错误词，则移除该疑似错词，最后得到的便是需要进行后续文本纠错处理的一系列疑似错词。

3.2. 文本错误纠正

错误纠正模型将对上阶段得到的所有疑似错词进行逐一纠正，纠正过程分为两步，第一步是使用多通道候选词构建策略得到候选词集，第二步是通过得分修正算法计算最优候选词。

3.2.1. 多通道候选词构建策略

针对四种错误类型，本文设计了不同候选词构建策略，如图 2 所示。

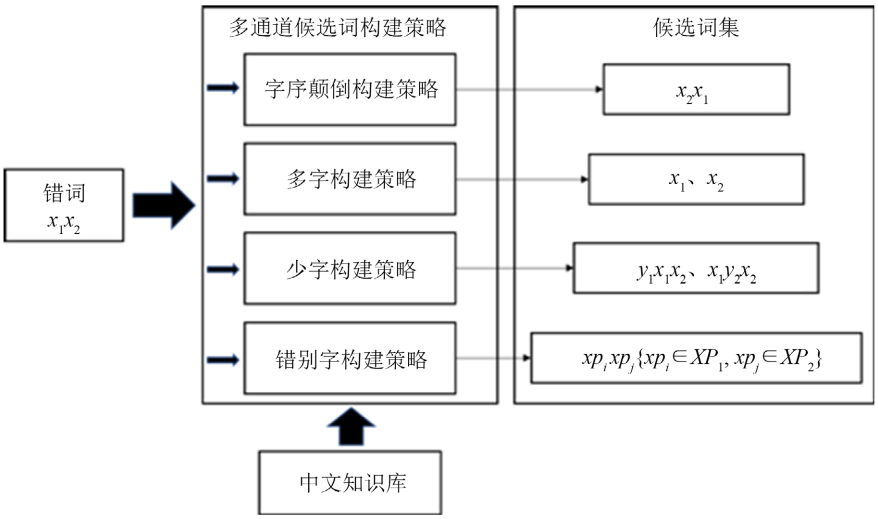

Figure 2. Multi-channel candidate construction strategy
图 2. 多通道候选词构建策略

字序颠倒：只针对疑似错词长度大于 1 的情况。

多字：只考虑多一个字的情况，若错词长度为 1，则候选词仅为空字符串。

少字：只考虑少一个字的情况，对于错词中的每一个字，查找常用词典中包含该字的二字常用词，然后用这些词替换这个字。图中例子即为使用常用词 y_1x_1 和 x_1y_2 替换原本的 x_1。

错别字：包含拼音错误候选词集和字形错误候选词集。对于疑似错词中的每一个字，先获取它的所有相同及相似读音，然后得到这些读音的所有字的集合，便是该字的拼音字集。设 x_1 的拼音字集为 XP_1，x_2 的拼音字集是 XP_2，则该错词的拼音错误候选词集为 $xp_ixp_j\left\{xp_i\in XP_1, xp_j\in XP_2\right\}$。字形错误候选词集构建方法与之类似，使用形近字进行替换组合即可。需要注意的是，若组合后的词与原错词的编辑距离大于 1，则该组合词必须在常用词典中才能加入候选词序列。这不仅可以降低误纠率，还能大大减少候选词的数量，提高纠正效率。

通过以上构建策略，取字序颠倒、多字、少字、错别字候选集的并集便得到了一个疑似错词的全部候选词集。

3.2.2. 计算最优候选词

本文通过使用语言模型计算句子困惑度来评价句子合理性。困惑度(*PPL*)计算过程如式(5)所示，其中 *S* 表示 Sentence，*N* 是句子长度，$P(w_i)$ 第 *i* 个词的概率。*PPL* 越小，$P(w_i)$ 越大，则该 Sentence 出现的概率就越大，句子越具合理性。

$$PPL(S)=P(w_1w_2\cdots w_N)^{-\frac{1}{N}} \tag{5}$$

使用语言模型计算原句评分记为 OriginS，然后用每一个候选词替换疑似错词并同样地计算句子得分，得分越小句子越合理。如果是少字错误，则需要对得分进行修正，即将计算出的句子得分乘以修正系数，本文设为 6。设 $Candi_1$ 是所有与疑似错词编辑距离为 1，且得分最低(记为 $MinS_1$)的候选词，$Candi_2$ 是编辑距离大于 1 的得分最低(记为 $MinS_2$)的候选词。式(6)得到的 SourceS 是与候选词进行评分比较的对

象，其中 MODIFY 是可调节参数，它控制候选词优于疑似错词的程度，值越小则候选词对句子困惑度降低的程度就需越大，本文对该参数进行调参实验的结果见第三章第三节。最优候选词的选择如式(7)所示，若两个最低候选词得分都比 SourceS 大，则视疑似错词无错，不予纠正。

$$\text{SourceS} = \text{OriginS} \times \text{MODIFY} \tag{6}$$

$$\text{最佳候选词} = \begin{cases} \text{原疑似错词}, \ \text{SourceS} \le MinS_1 \text{ 且 } \text{SourceS} \le MinS_2 \\ Candi_2, \ \text{SourceS} > MinS_1 \text{ 且 } MinS_1 \times \text{MODIFY} > MinS_2 \\ Candi_2, \ \text{其它情况} \end{cases} \tag{7}$$

对检错阶段给出的一系列疑似错词进行逐一纠正，然后使用最优候选词替换原错词，便得到了模型纠正文本。在经过模型校对之后还需要使用知识库进行纠正，在不满足所有的不转换规则条件下，将自定义错词修改为自定义正确词，便得到了输入文本的最终纠正文本。

3.3. 校对错误反馈机制

文本校对的结果存在三种错误情况，即误纠(正确词被纠错)、错纠(错误词被纠错)以及少纠(错误词未被纠出)，针对以上错误，本文设计了专用词典和自定义错词库两种表结构，可提供给用户自行维护，进行校对结果错误反馈。基于知识库的反馈机制如图 3 所示。自定义错词库的不转换规则指的是，这个错词在某些特定上下文关系中，不需要进行纠正。

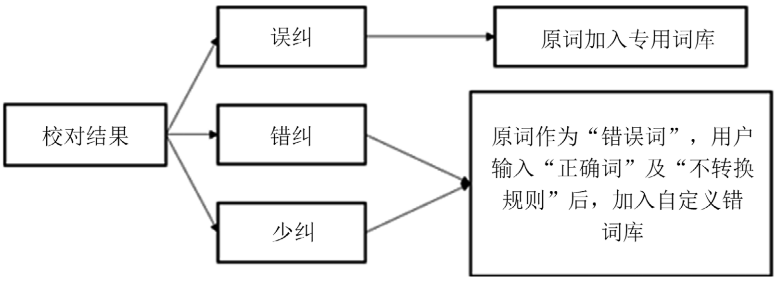

Figure 3. Feedback mechanism of proofreading errors based on knowledge base
图 3. 基于知识库的校对错误反馈机制

3.4. 知识库构建

本文构建的知识库由 8 个文本校对相关词典组成，分别是常用词典(超 34 万个词)、停用词典(近 1400 个词)、字音词典(8500 个汉字及其读音)、同音字词典(400 个拼音及其同音字)、近音字词典(256 组相似读音)、形似字词典(370 组相似字)、专用词典和用户自定义错词库。专用词典和用户自定义错词库初始化时不包含任何数据，是用户使用该文本校对方法过程中自行添加的。

4. 实验

4.1. 数据集

本文使用了全网新闻数据(SogouCA)、搜狐新闻数据(SogouCS)和人民日报 2014 语料库等共计 9GB 的开源文本作为语言模型的训练数据。本文改进的语言模型为无监督学习模型，在 CPU 环境下完成训练耗费近 5 个小时，这远远低于同数据量下的 ErnieCSC 训练时间，CPU 耗时 20 天，GPU 耗时 2 天。

错别字校对评估使用的数据集是 SIGHAN7CSC [16]，它是一个中文文本纠错竞赛提供的官方数据集，含有错误例句 1000 条，平均每句包含 70 个字，每句字词级错误率为 2%。

此外，由于之前鲜有多字、少字和字序颠倒这三种文本错误的研究，缺乏公开数据集，所以为了验证本文模型在这三种错误上的纠正准确率，依据搜狐新闻自构建了专项测试集，测试集详情见表 1。

Table 1. Three special test data sets
表 1. 三种专项测试数据集

数据集名称	样例数	字/句	错字/句
多字测试数据集	500	50	2
少字测试数据集	500	45	2
字序颠倒测试数据集	500	47	3

4.2. 评估指标

评估指标 LA (位置准确度)、CA (纠正准确度)以及 CP (纠正精度)的计算方法见公式(8)、(9)、(10)。其中 T 表示包含错误的句子总数，P 表示模型给出了纠正意见的句子总数。检错正确指的是检测出的错误位置必须与实际错误位置完全一致，少一处或多一处都不算检错正确。纠错正确指的是不包含任何错纠、少纠、误纠结果。

$$\text{LA} = \text{检错正确的句子数}/T \tag{8}$$

$$\text{CA} = \text{纠错正确的句子数}/T \tag{9}$$

$$\text{CP} = \text{纠错正确的句子数}/P \tag{10}$$

4.3. 实验结果与分析

4.3.1. 模型调参实验结果

前文介绍了文本纠错过程中有一个可调节参数 MODIFY，它控制候选词优于疑似错词的程度。对其进行调整，在 SIGHAN7CSC 数据集上得到的实验结果如图 4 所示。可以明显看出，MODIFY 值与纠正精度，即 CP 值呈负相关。这是因为该值是模型修改句子的阈值，所以它越小则模型对检错位置进行修改的概率就越小，这会导致最终进行纠正的句子总数变少，则纠正精度会增大，但由于纠错正确的句子数也在减少，因此 LA 和 CA 大大降低。平衡三种评估指标可以得到 MODIFY 的最佳值为 0.35。其它的可调节参数，如 THRESHOLD 等，由于篇幅原因不进行实验结果的展示，只给出最佳实验值。

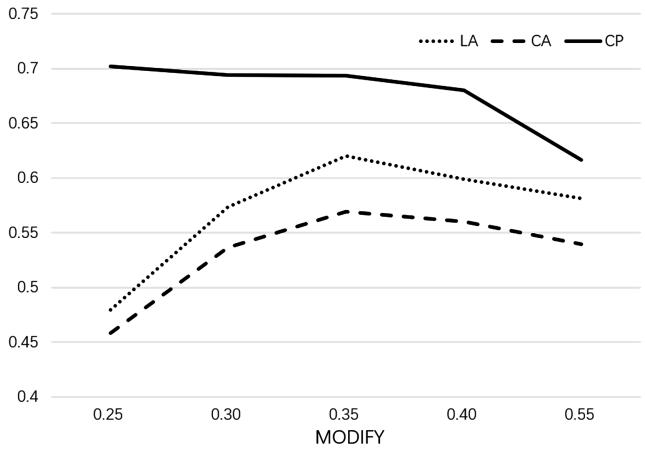

Figure 4. Model MODIFY parameter adjustment result diagram
图 4. 模型 MODIFY 调参结果图

4.3.2. 与其它模型结果对比

SIGHAN7CSC 上的实验结果如表 2 所示，包含了传统语言模型、深度学习模型以及官方参赛队伍的结果。可以明显看出，本文提出的模型在 LA、CA 以及 CP 评估指标中均取得了最优的实验测试结果，分别提升了 6.67%、6.39%以及 5.17%，这充分证明了本文提出方法的优越性。在实际应用过程中，通过校对错误反馈机制，准确率和精确度还会继续上升。

Table 2. Comparison with other models
表 2. 与其它模型对比结果

结果来源	LA	CA	CP
N-gram + 反转索引	0.6630	0.6250	0.7030
TOP1 参赛队伍	0.5590	0.5160	0.6158
TOP5 参赛队伍平均	0.5214	0.4862	0.6280
Ernie-CSC	0.62	0.573	0.665
本文方法	**0.7299**	**0.6889**	**0.7547**

4.3.3. 专项测试实验结果

本文提出的中文文本校对方法不仅适用于上述字形字音错误，还可以纠正字序颠倒、多字以及少字的情况。针对自构建的三种专项测试集，实验结果及分析如下。

表 3 给出了针对多字错误的专项测试结果，可以明显看出 LA 与 CA 是一致的，也就是说错词位置检测正确的句子全部都被正确纠正了，但是存在一些未被检测出的错误。此外 CP 值，即纠正精度很高，剩下不足 7%的未被正确纠正的句子多为人名误报。

Table 3. Experimental results of multi-word test
表 3. 多字测试实验结果

LA	CA	CP
0.5194	0.5194	0.9362

表 4 给出了针对少字错误的专项测试结果。由于同一个字在常用词库中可能包含多个意义相似的词，如算法把“反腐败争”修改为了“反腐败斗争”，然而正确答案是“反腐败争斗”。这会导致被正确检错的位置无法被正确纠正，即 CA 会低于 LA，同时纠正精度 CP 也会降低。

Table 4. Experimental results of missing-word test
表 4. 少字测试实验结果

LA	CA	CP
0.4521	0.4171	0.8800

表 5 给出了针对字序颠倒的专项测试结果，LA 和 CA 都达到了 80%以上，CP 达到了 97%，这充分说明了文本提出的字序颠倒错误纠正策略的有效性与准确性。

Table 5. Test results of word order reversal
表 5. 字序颠倒测试实验结果

LA	CA	CP
0.8256	0.8231	0.9757

5. 结束语

本文提出的引入反馈机制的中文文本自动校对技术在公共数据集和自构建数据集上均取得了最优的实验结果，这充分说明了该方法的可用价值与研究意义。除此之外，本文在实际应用基础上设计实现了校对结果错误反馈机制，使得自动校对方法流程更具容错性和交互性，使得文本自动校对技术作为数据预处理的一环更具有工程应用能力与价值。

本文的研究成果对于有相关需求的人员是一个完整而详细的参考，后续还有很多可以继续研究和优化的方向，例如降低人名、地名误报，提高错误位置检测准确率，以及词与词的颠倒问题纠正等。

参考文献

[1] 刘明洁, 梁毅, 艾中良, 贾高峰. 面向法律文书的中文文本校对方法研究[J]. 计算机工程与应用, 2020, 56(24): 274-278.

[2] 陈楠, 曹雪虹, 焦良葆, 等. 面向电力巡检语音指令识别后的文本纠错算法[J]. 计算机与数字工程, 2022, 50(1): 116-123, 134.

[3] 张鑫. 面向社会媒体的中文文本校对方法研究与实现[D]: [硕士学位论文]. 哈尔滨: 黑龙江大学, 2016.

[4] Chang, C.H. (1994) A Pilot Study on Automatic Chinese Spelling Error Correction. *Communication of COLIPS*, **4**, 143-149.

[5] 王福钊, 周雁. 基于匹配算法的藏文文本词语校对研究[J]. 计算机与数字工程, 2021, 49(7): 1433-1436.

[6] 郝亚男, 乔钢柱, 谭瑛. 基于神经网络与注意力机制的中文文本校对方法[J]. 计算机系统应用, 2019, 28(10): 190-195.

[7] Zhang, J. and Zhang, X. (2020) Comparison of Chinese Character Correct and Error Classifier for Overseas Students Based on Handwriting Motion Characteristics. *Journal of Physics*: *Conference Series*, **1646**, Article ID: 012064. https://doi.org/10.1088/1742-6596/1646/1/012064

[8] 石敏, 高尚. 基于决策列表的中文同音词自动识别与校对[J]. 电子设计工程, 2015(9): 39-41.

[9] Yu, J. and Li, Z. (2014) Chinese Spelling Error Detection and Correction Based on Language Model, Pronunciation, and Shape. *Proceedings of the 3rd CIPS-SIGHAN Joint Conference on Chinese Language Processing*, Wuhan, 20-21 October 2014, 220-223. https://doi.org/10.3115/v1/W14-6835

[10] Zhao, H., Cai, D., Xin, Y., Wang, Y. and Jia, Z. (2017) A Hybrid Model for Chinese Spelling Check. ACM Transactions on Asian and Low-Resource Language Information Processing, **16**, 1-22. https://doi.org/10.1145/3047405

[11] 王琼, 旷文珍, 许丽. 基于改进的 N-gram 模型和知识库的文本查错算法[J]. 计算机应用与软件, 2021, 38(10): 310-315, 320.

[12] 王浩畅, 周锦程. 中文语法自动纠错系统的研究与实现[J]. 企业科技与发展, 2020(2): 81-84, 87.

[13] 龚永罡, 吴萌, 廉小亲, 裴晨晨. 基于 Seq2Seq 与 Bi-LSTM 的中文文本自动校对模型[J]. 电子技术应用, 2020, 46(3): 42-46.

[14] 龚永罡, 裴晨晨, 廉小亲, 王嘉欣. 基于 Transformer 模型的中文文本自动校对研究[J]. 电子技术应用, 2020, 46(1): 30-33, 38.

[15] Zhang, R., Pang, C., Zhang, C., *et al*. (2021) Correcting Chinese Spelling Errors with Phonetic Pre-Training. *Findings of the Association for Computational Linguistics*: *ACL-IJCNLP* 2021, Online, 1-6 August 2021, 2250-2261. https://doi.org/10.18653/v1/2021.findings-acl.198

[16] Wu, S.-H., Liu, C.-L. and Lee, L.-H. (2013) Chinese Spelling Check Evaluation at SIGHAN Bake-off 2013. *Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing*, Nagoya, 14-18 October 2013, 35-42.