

CCL2022-CLTC 赛道四：“啊对对对”队技术评测报告

叶劲亨¹, 李映辉¹, 马仕镭¹, 郑海涛^{1,2,*}

¹清华大学深圳国际研究生院 深圳 518055

²鹏城实验室 深圳 518055

{yejh22, liyinghu20, msr21}@mails.tsinghua.edu.cn

zheng.haitao@sz.tsinghua.edu.cn

摘要

本论文阐述了在CCL2022汉语学习者文本纠错评测中赛道四（CCL2022-CLTC Track4）——多参考多来源汉语学习者文本纠错（Multi-reference Multi-source Chinese Learner Text Correction, 简称MMCLTC）（王莹莹et al., 2022）中“啊对对对”队伍提交的参赛系统。在模型方面，我们分别训练Seq2Edit-based和Seq2Seq-based模型，并使用了模型集成技术。在数据方面，我们使用了基于字词混淆集的数据增强策略，有效提高了Seq2Edit-based模型的纠错性能。另外，现存的中文语法纠错数据集存在大量的单源句多目标样本，我们发现使用单源句多目标样本进行训练会稀释模型的预测概率质量，从而降低模型的纠正性能。因此，我们提出了一个简单又有效的数据清洗策略，过滤后的训练集样本数仅占原始数据集的50~60%，在加速训练的同时还能显著提升模型性能。

关键词： 汉语学习者文本纠错评测；多参考多来源汉语学习者文本纠错；数据清洗

CCL2022-CLTC Track 4: Technical Report of “啊对对对” Team

Jingheng Ye¹, Yinghui Li¹, Shirong Ma¹, Hai-Tao Zheng^{1,2,*}

Tsinghua Shenzhen International Graduate School, Tsinghua University

Peng Cheng Laboratory

{yejh22, liyinghu20, msr21}@mails.tsinghua.edu.cn

zheng.haitao@sz.tsinghua.edu.cn

Abstract

This paper explains the system submitted by the team “啊对对对” in CCL2022-CLTC Track4: Multi-reference Multi-source Chinese Learner Text Correction (MMCLTC) (Wang et al., 2022). In terms of models, we trained Seq2Edit-based and Seq2Seq-based models respectively, both of which can be integrated into ensemble model. In terms of data, we introduce data augmentation based on character and word-based confusion set, which effectively improves the performance of Seq2Edit-based model. In addition, there are a large number of single-source multi-target samples in the existing Chinese grammatical error correction dataset, and we found that these single-source multi-target samples will dilute the prediction probability mass of models, thus reducing the performance of the model. Therefore, we propose a simple and effective data cleaning strategy. The filtered training set samples only account for 50~60% of the original dataset, which can significantly improve the model performance while accelerating training.

指导老师，通讯作者：zheng.haitao@sz.tsinghua.edu.cn

Source	我能胜任这此职务 I am competent for this the position.
Ref. 1	我能胜任这此职务。 I am competent for this position.
Ref. 2	我能胜任这此职务。 I am competent for the position.

Table 1: 具有两个参考目标句的语法纠错样例。纠错部位已用红色标记。

Keywords: Chinese Learner Text Correction , Multi-reference Multi-source Chinese Learner Text Correction , Data cleaning

1 引言

MMCLTC任务旨在令模型检测并修改汉语学习者多领域文本中的语法错误，评估系统将基于给定的多条参考答案对模型预测进行评价。该任务使用的评估集为多源多参考中文语法纠错评测数据集（Multi-Reference Multi-Source Evaluation Dataset for Chinese Grammatical Error Correction, 以下简称MuCGEC）(Zhang et al., 2022), MuCGEC评估集具有参考答案答案多样、来源多样和标注质量高等优点，能够更真实地衡量中文语法纠错（Chinese Grammatical Error Correction, CGEC）模型的语法纠正性能，推动了CGEC领域的发展。

我们首先使用基于字词混淆集的数据增强技术(Tang et al., 2021; Zhang et al., 2021)构造合成的并行数据集用于预训练模型。另外，我们提出了一种简单但有效的数据清洗策略，并将其应用在现存的CGEC训练集。我们分别训练了Seq2Seq-based和Seq2Edit-based模型，由于这两种异构模型在纠错结果上能充分互补，合理地将这两种模型进行集成能显著提升系统的纠错性能。

我们观察到现存的CGEC数据集中同一个源句可能对应多个不同的纠正目标句，使用这些单源句多目标样本进行训练会稀释模型的预测概率质量，造成模型性能下降。因此，对于训练集中大量的单源句多目标样本，我们基于编辑距离选择与源句最相似的参考目标加入到单目标数据集中，该数据集中每个源句正好对应一个参考目标句。使用单目标训练集进行训练，相比全数据集训练，不仅训练时间大大缩短（单目标训练集的样本数量仅占全数据集的50-60%），而且能提升模型的纠错性能大约1~2 F0.5分数。不过，我们的模型在面对长编辑时纠正正确率依旧有待提高。我们训练的单个Seq2Edit-based模型最终在MuCGEC测试集中取得了45.81 F0.5分数，Seq2Seq-based模型取得了44.66 F0.5分数，最终提交的集成模型5×Seq2Edit+5×Seq2Seq取得了51.15 F0.5分数，取得了CCL2022-CLTC Track4比赛的top1成绩。

2 背景

Dataset	#sent	#err. sent (perc.)	chars/sent	edits/ref	refs/sent
Original NLPCC18	2000	1983 (99.2%)	29.7	2.0	1.1
MuCGEC (NLPCC18)	1996 (4)	1904 (95.4%)	29.7	2.5	2.5
MuCGEC (CGED)	3125 (12)	2988 (95.6%)	44.8	4.0	2.3
MuCGEC (Lang8)	1942 (58)	1652 (85.1%)	37.5	2.8	2.1
MuCGEC	7063 (74)	6544 (92.7%)	38.5	3.2	2.3

Table 2: MuCGEC评估集数据统计量，从左到右，表格的列依次表示句子数量、错误句子数量（占比）、每个句子的平均字符数、每个参考的平均编辑数以及每个句子的平均参考数量(Zhang et al., 2022)。

CCL2022-CLTC Track4比赛旨在对其中存在的拼写、词法、语法等各类错误进行自动纠正。给定可能包含语法错误的句子，要求模型输出正确的纠正目标句，如表1。由于MuCGEC评估集包含多个参考目标句，因此在对模型预测进行评估时需要将模型预测的假设句和每个参考目标句使用字级别编辑抽取工具ERRANT-ZH (Zhang et al., 2022)逐一计算F0.5分数，最终选择最高的F0.5分数作为该源句的纠正分数。

本次比赛所使用的评估集MuCGEC主要来自汉语学习者，这是一个多来源的中文语法纠错评估集，评估样本分别采样自以下数据集：NLPCC18测试集（来自于NLPCC18-shared Task2评测任务）、CGED测试集（来自于CGED18&20评测任务）以及中文Lang8训练集（来自于NLPCC18-shared Task2评测任务）(Zhang et al., 2022)。数据的整体统计见表2。

在模型层面上，早期的中文语法纠错使用基于规则的方法对中文句子进行语法错误诊断或纠错(Yu and Li, 2014)。较近的研究将CGEC视作序列到序列生成（即机器翻译）任务，并使用Seq2Seq-based模型实现CGEC任务(Yuan and Briscoe, 2016; Duan et al., 2021)。预训练语言模型的出现显著提升了深度模型在自然语言处理领域中各种下游任务的性能，CGEC任务也不例外，Zhang等人(2021)使用BERT预训练模型分别进行拼写检查和语法纠错。Zhang等人(2022)使用BART以序列到序列生成的方式实现CGEC，并且对比了不同预训练序列到编辑模型（BERT、RoBERTa、MacBERT和StructBERT）的纠错性能。

在数据层面上，现存的CGEC数据集包含的样本数量较少且包含噪声较多，较大程度上限制了模型的纠错性能。为此，大量的工作(王辰成et al., 2020; Zhao and Wang, 2020; Tang et al., 2021; Zhang et al., 2021)通过生成更接近原始样本分布的合成并行数据集，以预训练的方式提升模型纠错性能。Xie等人(2018)首次提出使用基于反向翻译的数据增强技术来生成并行语料库，Wang等人(2020)通过对单语语料的腐化从而生成更多的纠错数据，进一步提高模型的性能。Zhao等(2020)则提出一种动态随机掩盖的方式，通过掩盖并替换部分数据达到扩充语料集的目的。Tang等人(2021)提出融合字词粒度噪声的数据增强方法，从而构建出大规模且质量较高的并行数据集。遵循以往的工作，我们同样也构造了基于字词混淆集的合成并行数据集，用于对模型的预训练。

预训练之后，还需要使用CGEC标注数据集对模型进行微调。现存的语法错误标注数据集存在大量的单源句多目标样本，现存的工作(Zhang et al., 2022; Katsumata and Komachi, 2020)几乎都使用全数据集微调模型，我们发现这些多目标样本不仅增加了训练成本，还会降低模型的纠错性能。

3 系统

我们使用的纠错系统由Seq2Seq-based和Seq2Edit-based集成。在本节的剩余部分，我们依次介绍Seq2Edit-based和Seq2Seq-based模型的工作方式。

3.1 Seq2Edit-based模型

我们使用GECToR(Omelianchuk et al., 2020)作为Seq2Edit-based模型⁰。GECToR是一个简单又高效的语法纠错序列标注模型，并行地对句子中所有标记进行解码，推理速度相比Seq2Seq能提高十倍左右。另外，为了解决GECToR的纠正标签为单个字符的缺陷，可以采取迭代解码策略，对GECToR模型的推理结果进行多次回送输入。GECToR模型训练完毕后，可以通过设置\$KEEP标签的概率偏置 η ，调整模型在推理时的精度和召回率，以取得最佳的F0.5分数。

3.2 Seq2Seq-based模型

我们使用Chinese-BART(Shao et al., 2021)作为我们的Seq2Seq-based模型，因为原始Chinese-BART的词表缺少一些常见的中文标点/字符，我们为Chinese-BART的词表添加了原本不存在的常见中文标点/字符，并且使用近义标记初始化新加入的词嵌入向量，添加的部分标记如表3。

⁰我们没有对GECToR模型进行额外改进，因此不在此做过多赘述。

新添加标记	近义标记
“ (中文左双引号)	” (英文双引号)
” (中文右双引号)	” (英文双引号)
...	。
—	-
‘ (中文右单引号)	’ (英文单引号)
’ (中文右单引号)	’ (英文单引号)
‘	’
-	-
鯖	鱼
—	-

Table 3: Chinese-BART扩展标记集，使用近义标记的词嵌入向量进行初始化。

4 训练数据集

4.1 合成数据集的构建

我们分别从字和词粒度对文本进行加噪处理，以生成并行语料库。具体地说，我们选用开源的社区问答语料库¹作为种子语料库，该语料库已经经过过滤处理，只保留高质量问题和回复，并且涉及的社区问答话题广泛，共包含410万条语料。合成数据集的构建需要分别用到字和词粒度的两种混淆集，表4展示了混淆集示例。

- 字混淆集²：包含读音、字形或意义相近的汉字，我们过滤掉其中的低频字。
- 词混淆集³：包含容易被误用的词语。

我们首先对语料进行简单的过滤，移除不正确使用标点符号以及包含推广、超链接以及特殊字符的语料，并且将语料进行分句，因为MuCGEC是句子级中文语法纠错评测集。然后，我们分别将得到的句子进行分字或使用分词工具jieba⁴进行分词，为后续的字词粒度数据增强做准备。对于句子中的每个字或词，我们按一定比例使用替换、删除、添加和乱序等操作对句子进行损坏处理，作为CGEC任务的源句。关于损坏处理的具体策略，我们参考了文献(Tang et al., 2021; Zhang et al., 2021)。

混淆集类型	原字/词	混淆字/词
字混淆集	一	倚、三、甚、哀、饴、不、昇、以、义、已、亦
词混淆集	人民	民、国民、公民、平民、黎民、庶、庶民、老百姓、苍生

Table 4: 混淆集示例

4.2 单目标训练集

模型经过预训练后，在汉语学习者文本域上的语法纠错性能依旧较低，还需要使用更符合汉语学习者语法错误分布的标注训练集进行微调。由于本次比赛不限制使用开源数据集，我们使用的并行数据集列举如下：

- Lang8⁵：来源于在线语言学习平台Lang8，第二语言学习者在平台上提出包含语法纠错的句子，由网友进行纠正；

¹https://github.com/brightmart/nlp_chinese_corpus

²https://github.com/HillZhang1999/MuCGEC/blob/main/scorers/ChERRANT/data/confusion_dict.txt

³<https://github.com/HillZhang1999/MuCGEC/blob/main/scorers/ChERRANT/data/cilin.txt>

⁴<https://github.com/foxsjy/jieba>

⁵<http://tcci.ccf.org.cn/conference/2018/taskdata.php>

- HSK⁶: 来源于汉语水平考试 (Hanyu Shuiping Kaoshi) 的作文;
- CGED⁷: 来源于中文语法错误诊断任务(Rao et al., 2018; Rao et al., 2020);
- YACLC-minimal⁸: 是一个多维度中文语法纠错数据集(Wang et al., 2021), 包含最小改动版本 (YACLC-minimal) 和流畅性改动版本 (YACLC-fuency)。我们只使用最小改动版本的开发集部分。

我们观察到现存的CGEC数据集中同一个源句可能对应多个参考目标句, 当模型同时使用不同的参考目标进行训练时, 不同的纠正标签会稀释模型的预测概率质量, 令模型倾向于预测\$KEEP\$标签, 从而降低模型纠正性能。为了消除因多目标训练造成的概率质量稀释问题, 我们对现有的训练集进行简单但十分有效的数据清洗策略, 不仅只需要使用相当于原始数据集样本数量一半的训练样本, 还能显著提高模型的纠错性能。

具体地说, 给定一个源句 \mathcal{S} , 该源句对应参考目标句集合 $\{\mathcal{T}\}_{k=1}^K$, 我们逐一计算每个目标句 \mathcal{T}_k 与源句的莱文斯坦比 r_k :

$$r_k = \frac{sum - ldist(\mathcal{S}, \mathcal{T}_k)}{sum}$$

其中 sum 指源句和目标句的长度总和, $ldist(\cdot, \cdot)$ 表示两个序列之间的类编辑距离, 不同于编辑距离, 类编辑距离的删除、插入的距离为1, 但是替换的距离为2。

我们选择莱文斯坦比最大的(源句, 目标句)对加入到单目标数据集中。最终得到的单目标数据集的统计量如表5, 其中样本数量仅占全数据集的54%, 因此能显著降低模型训练成本。

数据集名称	样本数量
全数据集	1,325,133 (源句可能重复)
单目标数据集	716,683 (源句不重复)

Table 5: 训练集统计量

5 实验

5.1 训练设置

项目主要使用PyTorch库搭建, 我们收集了来自Lang8、HSK、CGED和YACLC的数据作为训练数据, 并且严格按照官方的指示过滤了与MuCGEC评估集中重复的训练样本⁹。模型学习率设置为 $1e-5$, Seq2Seq的批大小设置为32, 梯度积累步数设置为1, Seq2Edit的批大小设置为32, 梯度积累步数设置为4, 根据官方提供的评测脚本选择在开发集上表现最优的模型进行测试。

对于Seq2Edit-based模型, 我们首先在合成数据集上进行预训练, 然后使用4.2节中介绍的单目标数据集进行微调。出于简化, 我们没有进行多阶段的微调。

对于Seq2Seq-based模型, 我们没有在合成数据集上进行预训练, 仅在单目标数据集上进行微调。我们尝试过加入预训练, 但是对模型性能没有显著提升, 因为BART的噪声自编码预训练任务和基于规则的数据增强技术类似, 因此BART模型难以进一步从数据增强中受益。我们对Seq2Seq-based模型进行了两阶段的微调, 其中第一阶段使用由Lang8、HSK、CGED和YACLC组成的单目标数据集进行微调, 第二阶段我们剔除了包含噪声较多的Lang8数据集, 使用由HSK、CGED和YACLC组成的数据集进行微调, 并且只训练模型的softmax层。

对于模型集成, 遵循Zhang等人(2022), 我们使用了简单的编辑投票机制, 选择大多数模型同意的编辑作为最终保留的编辑。

⁶<http://hsk.blcu.edu.cn>

⁷https://github.com/blcuicall/cged_datasets

⁸<https://github.com/blcuicall/CCL2022-CLTC/blob/main/datasets/track3/README.md>

⁹我们先将句子转为简体, 然后过滤掉出现在MuCGEC评估集中的训练样本

5.2 开发集结果

对于Seq2Edit-based模型，我们使用StructBERT预训练模型初始化模型权重，训练阶段分为两个步骤：预训练和微调。对于Seq2Seq-based模型，我们使用Chinese-BART预训练模型初始化模型权重，进行两阶段微调。对于集成模型，我们更换不同的随机种子训练多个Seq2Edit-based和Seq2Seq-based模型，并分别选择表现最好的模型用于集成。表6展示了模型在MuCGEC开发集上的结果。

对于Seq2Edit-based模型，添加预训练可以显著提高模型的精确度，并且保持召回率基本保持，由于F0.5分数更看重精确度，所以F0.5分数显著提升。使用我们提出的数据清洗策略得到的单目标数据集进行微调，不仅加速模型训练，还可以使模型的精确度和召回率同时提高，最终的F0.5分数达到了45.68。对于Seq2Seq-based模型，使用单目标数据集微调模型同样能提高模型的纠错F0.5分数，不过提升幅度不如Seq2Edit-based，可能的原因是Seq2Edit-based模型的样本预处理使得每个标记只能对应单标记标签，因此对数据质量要求更高。

对于集成模型，我们分别选择了1+1、3+3和5+5的模型组合。随着集成模型数量的增加，纠错系统的性能逐渐增加，这说明了Seq2Edit-based模型和Seq2Seq-based模型可以充分地互补，Seq2Edit-based模型精确度较高，而Seq2Seq-based模型召回率较高，合理地集成异构的语法纠错模型可使得纠错系统性能显著提高。

模型	训练策略	开发集结果		
		精确度	召回率	$F_{0.5}$
<i>Seq2Edit</i>	全数据集微调	47.90	24.88	40.42
<i>Seq2Edit</i>	预训练+全数据集微调	54.67	24.00	43.54
<i>Seq2Edit</i>	预训练+单目标数据集微调	58.24	24.53	45.68
<i>Seq2Seq</i>	全数据集微调	51.28	29.27	44.57
<i>Seq2Seq</i>	单目标数据集微调	53.97	27.86	45.45
<i>Seq2Seq</i>	单目标数据集微调+第二阶段微调	52.43	31.02	46.07
$1 \times Seq2Edit + 1 \times Seq2Seq$	-	75.13	19.30	47.59
$3 \times Seq2Edit + 3 \times Seq2Seq$	-	74.68	21.07	49.50
$5 \times Seq2Edit + 5 \times Seq2Seq$	-	73.38	22.70	50.73

Table 6: MuCGEC开发集上模型结果

5.3 测试集结果

由于比赛成绩仅由B榜决定，而且每个队伍只能在B榜上提交三次纠正结果，因此我们在B榜提交了三种不同的集成模型，最终以 $5 \times Seq2Edit + 5 \times Seq2Seq$ 的51.15 F0.5分数领先。比赛结束后取消了提交限制，因此我们重新提交了单模型的预测结果，表7报告了在MuCGEC测试集上的模型结果。从表7可以看出我们的模型在MuCGEC测试集上的结果和开发集是类似的，这说明了我们提出的数据清洗策略的有效性。

6 总结

在本次CCL2022-CLTC Track4比赛中，我们观察到现存的中文语法纠错数据集中一条源句可能对应多个参考目标句，这些单源句多目标样本可能会稀释模型的预测概率质量，导致模型纠错性能下降。因此，我们提出了简单又有效的方法来构建样本量接近原始数据集一半的单目标数据集。我们在Seq2Edit-based和Seq2Seq-based模型上的实验结果表明，使用单目标数据集训练不仅使训练时间大大缩短，而且能显著提升模型性能。另外，我们还构建基于字词混淆集的语法纠错合成数据集用于模型预训练，使Seq2Edit-based模型的纠错性能显著提高。最终，我们“啊对对对”队伍的纠错系统在测试集上的总得分为51.15 F0.5，位列第一。

然而，我们提出的单目标数据集过滤了较多的训练样本，存在样本利用不充分问题。未来，我们将研究如何在提升模型性能的前提下充分利用数据集中的单源句多目标样本，探索多目标设置下的语法纠错。

模型	训练策略	开发集结果		
		精确度	召回率	$F_{0.5}$
<i>Seq2Edit</i>	全数据集微调	47.59	26.72	41.16
<i>Seq2Edit</i>	预训练+全数据集微调	53.40	25.89	44.04
<i>Seq2Edit</i>	预训练+单目标数据集微调	53.70	28.85	45.81
<i>Seq2Seq</i>	全数据集微调	47.76	29.10	42.33
<i>Seq2Seq</i>	单目标数据集微调	49.45	27.33	42.56
<i>Seq2Seq</i>	单目标数据集微调+第二阶段微调	50.68	30.27	44.66
$5 \times Seq2Edit + 5 \times Seq2Seq$	-	72.71	23.4	51.15

Table 7: MuCGEC测试集上模型结果

参考文献

- Jianyong Duan, Yang Yuan, and Hao Wang. 2021. Chinese spelling correction method based on transformer local information and syntax enhancement architecture. *Beijing Da Xue Xue Bao*, 57(1):61–67.
- Satoru Katsumata and Mamoru Komachi. 2020. Stronger baselines for grammatical error correction using a pretrained encoder-decoder model. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 827–832, Suzhou, China, December. Association for Computational Linguistics.
- Kostiantyn Omelianchuk, Vitaliy Atrasevych, Artem Chernodub, and Oleksandr Skurzshanskyi. 2020. GECToR – grammatical error correction: Tag, not rewrite. In *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 163–170, Seattle, WA, USA → Online, July. Association for Computational Linguistics.
- Gaoqi Rao, Qi Gong, Baolin Zhang, and Endong Xun. 2018. Overview of NLPTEA-2018 share task Chinese grammatical error diagnosis. In *Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications*, pages 42–51, Melbourne, Australia, July. Association for Computational Linguistics.
- Gaoqi Rao, Erhong Yang, and Baolin Zhang. 2020. Overview of NLPTEA-2020 shared task for Chinese grammatical error diagnosis. In *Proceedings of the 6th Workshop on Natural Language Processing Techniques for Educational Applications*, pages 25–35, Suzhou, China, December. Association for Computational Linguistics.
- Yunfan Shao, Zhichao Geng, Yitao Liu, Junqi Dai, Fei Yang, Li Zhe, Hujun Bao, and Xipeng Qiu. 2021. CPT: A pre-trained unbalanced transformer for both chinese language understanding and generation. *CoRR*, abs/2109.05729.
- Zecheng Tang, Yixin Ji, Yibo Zhao, and Junhui Li. 2021. 基于字词粒度噪声数据增强的中文语法纠错(chinese grammatical error correction enhanced by data augmentation from word and character levels). In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 813–824.
- Yingying Wang, Cunliang Kong, Liner Yang, Yijun Wang, Xiaorong Lu, Renfen Hu, Shan He, Zhenghao Liu, Yun Chen, Erhong Yang, and Maosong Sun. 2021. YACL: A chinese learner corpus with multidimensional annotation. *CoRR*, abs/2112.15043.
- Yingying Wang, Cunliang Kong, Xin Liu, Xuezhi Fang, Yue Zhang, Nianning Liang, Tianshuo Zhou, Tianxin Liao, Liner Yang, Zhenghua Li, Gaoqi Rao, Zhenghao Liu, Chen Li, Erhong Yang, Min Zhang, and Maosong Sun. 2022. Overview of cltc 2022 shared task : Chinese learner text correction.
- Ziang Xie, Guillaume Genthial, Stanley Xie, Andrew Ng, and Dan Jurafsky. 2018. Noising and denoising natural language: Diverse backtranslation for grammar correction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 619–628, New Orleans, Louisiana, June. Association for Computational Linguistics.

- Junjie Yu and Zhenghua Li. 2014. Chinese spelling error detection and correction based on language model, pronunciation, and shape. In *Proceedings of The Third CIPS-SIGHAN Joint Conference on Chinese Language Processing*, pages 220–223, Wuhan, China, October. Association for Computational Linguistics.
- Zheng Yuan and Ted Briscoe. 2016. Grammatical error correction using neural machine translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 380–386, San Diego, California, June. Association for Computational Linguistics.
- Yue Zhang, Zuyi Bao, Bo Zhang, Chen Li, Jiacheng Li, and Zhenghua Li. 2021. Technical report of suda-alibaba team on ctc-2021. Technical report.
- Yue Zhang, Zhenghua Li, Zuyi Bao, Jiacheng Li, Bo Zhang, Chen Li, Fei Huang, and Min Zhang. 2022. MuCGEC: a multi-reference multi-source evaluation dataset for Chinese grammatical error correction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3118–3130, Seattle, United States, July. Association for Computational Linguistics.
- Zewei Zhao and Houfeng Wang. 2020. Maskgrec: Improving neural grammatical error correction via dynamic masking. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1226–1233.
- 王莹莹, 孔存良, 刘鑫, 方雪至, 章岳, 梁念宁, 周天硕, 廖田昕, 杨麟儿, 李正华, 饶高琦, 刘正皓, 李辰, 杨尔弘, 张民, and 孙茂松. 2022. Cltc 2022: 汉语学习者文本纠错技术评测及研究综述.
- 王辰成, 杨麟儿, 王莹莹, 杜永萍, and 杨尔弘. 2020. 基于transformer 增强架构的中文语法纠错方法. *中文信息学报*, 34(6):106–114.