

中文语法纠错的多轮解码方法研究

王晓盈, 穆玲玲*, 许鸿飞

郑州大学计算机与人工智能学院, 河南省 郑州市, 450001

xywangnlp@qq.com; iellmu@zzu.edu.cn; hfxunlp@foxmail.com

摘要

在语法纠错 (Grammatical Error Correction, GEC) 任务上, 序列到序列 (Sequence-to-sequence, seq2seq) 模型与序列到编辑 (Sequence-to-edit, seq2edit) 模型相比可以取得相当或更好的性能。序列到编辑模型通常通过多次迭代解码, 而序列到序列模型则以从左到右的方式一次性解码, 不考虑后续的词。通过在序列到序列模型中应用多轮解码 (Multi-Turn Decoding, MTD) 来迭代改进前一轮的修正结果, 可能会进一步提升性能。然而, 多轮解码会增加推理的计算成本, 且前一轮修正中的删除或替换操作可能会导致原始输入中有用的源语句信息丢失。本文提出了一种早停机制来提高效率。同时, 为解决源语句信息丢失问题, 本文将原始输入与上一轮的修正结果合并为一个序列。在NLPCC2018测试集、FCGEC验证集和NaCGEC测试集的实验结果表明, 本文方法可在BART基线上能带来一致且显著的性能提升, F0.5值分别提高了+2.06, +2.31和+3.45, 分别取得了47.34, 54.58和62.09的F0.5值。

关键词: 语法纠错; 多轮解码; 序列到序列

Multi-Turn Decoding for Chinese Grammatical Error Correction

Xiaoying Wang, Lingling Mu*, Hongfei Xu

School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou Henan, 450001

xywangnlp@qq.com; iellmu@zzu.edu.cn; hfxunlp@foxmail.com

Abstract

Sequence-to-sequence (seq2seq) models achieve comparable or better Grammatical Error Correction (GEC) performance compared to sequence-to-edit (seq2edit) models. Seq2edit models normally iteratively refine the correction result, while seq2seq models decode only once in a left-to-right manner unaware of subsequent tokens. Iteratively refining the correction results from the previous round with seq2seq models via Multi-Turn Decoding (MTD) may lead to better performance. However, MTD increases the inference costs. Deleting or replacing corrections in previous rounds may lose useful information in the source input. This paper presents an early-stop mechanism to alleviate the efficiency issue. To address the source information loss issue, this paper proposes to merge the source input with the previous round correction result into one

*通讯作者

©2024 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

sequence. Experiment results on the NLPCC2018 test set, FCGEC development set and NaCGEC test set show that our approach can lead to consistent and significant improvements over the strong BART baselines (+2.06, +2.31 and +3.45 F0.5 scores respectively), and obtaining F0.5 scores of 47.34, 54.58 and 62.09.

Keywords: Multi-Turn Decoding, Grammatical Error Correction, Sequence-to-sequence

1 引言

语法纠错 (Grammar Error Correction, GEC) 旨在纠正给定句子中的语法错误[Ng et al., 2013, 2014]。目前, 主流的语法纠错方法分为序列到编辑和序列到序列两种。序列到编辑方法将语法纠错任务视为序列标注任务, 通过预测每个词语的编辑标签 (如保留、删除、插入、替换等) 并迭代处理, 直至所有词语均被标记为保留[Malmi et al., 2019, Stahlberg and Kumar, 2020, Omelanchuk et al., 2020, Yuan et al., 2021]。这种方法往往需要多次校验以达到准确修正。相对地, 序列到序列方法将语法纠错任务视为机器翻译任务, 即从有语法错误的句子转换到语法正确的句子[Zhao et al., 2019, Kiyono et al., 2019, Wang et al., 2021, Li et al., 2022a, Fang et al., 2023]。序列到序列模型对输入句子进行编码, 并自回归解码出修正后的句子来实现这一转换[Omelanchuk et al., 2020, Liu et al., 2021a]。为提升性能, 现有方法通常采用预训练模型来提高纠错性能, 例如序列到编辑方法使用BERT [Devlin et al., 2019]和XLNet [Yang et al., 2019], 序列到序列方法使用BART [Lewis et al., 2020]和T5 [Raffel et al., 2020]。

序列到序列模型在未使用特定语言编辑操作的情况下, 表现出与序列到编辑方法相匹配或更优的性能。然而, 当前的序列到序列语法纠错研究通常仅执行一次解码, 忽略了后续词语的影响。多轮解码 (Multi-Turn Decoding, MTD) 策略可以通过迭代优化提升性能。训练多轮解码模型以生成给定错误解码结果的参考结果, 可能通过自我修正机制增强学习效果。

多轮解码不直接采用纠错模型单轮的解码结果作为最终的纠错结果, 而是通过将上一轮的修正结果作为新一轮的输入, 要求模型产生新的纠错结果, 并迭代修正, 直至连续两轮的解码结果相同。尽管这种方法提升了性能, 但增加了推理的计算成本。为解决效率问题, 本文提出一个与序列到序列模型协同工作的早停机制。该机制以句子结束标记 (<eos>) 的隐藏表示作为输入, 并在以下两种情况下终止多轮解码: 1) 下一轮的修正结果与当前修正结果一致, 或2) 与当前修正结果相比, 下一轮的修正结果与参考结果的编辑距离更大。

此外, 在多轮解码过程中, 若模型在前几轮中删除或替换了某些词语, 那么后续的纠正将遗失这些词语的信息, 即使它们可能为后续纠正提供有价值的信息。为解决信息丢失问题, 本文提出同时将源语句与上一轮的修正结果同时输入纠错模型, 直接将两个序列拼接[Correia and Martins, 2019]输入预训练模型会使长度翻倍, 且不利于模型对比分析两个序列中对应位置的词语。双编码器结构[Denkowski, 2015]会增加模型的复杂度, 不利于对预训练模型语言能力的充分利用。本文提出基于序列对比的源信息融合方法, 利用两个序列相近的特性, 对比两个序列中的相同片段和差异片段后再依序拼接, 避免相同片段的重复输入。然后利用编辑标记或位置编号对混合序列中每个词语在两个序列中的角色予以区分。

本文的主要贡献如下:

- 本文针对现有的序列到序列纠错方法仅进行一次从左到右的解码, 在解码过程中不能考虑后续词语的问题, 提出通过多轮解码来重复检查纠正现有的纠错结果, 改进序列到序列语法纠错方法。
- 为提升多轮解码的效率, 本文提出早停机制, 在下一轮可能不会产生更好的纠错结果时, 提前结束多轮解码。
- 本文提出源信息融合方法来解决多轮解码中删除和替换词语导致的信息丢失问题, 并根据两个序列相近的特性, 提出基于对比的序列混合方法, 避免序列拼接导致的效率问题。
- 在NLPCC2018测试集、FCGEC验证集和NaCGEC测试集上, 本文的方法相对于强大的BART基线带来了+2.06, +2.31和+3.45的F0.5值提升, 分别取得了47.34, 54.58和62.09的F0.5值, 证明了本文方法的有效性。

算法 1 多轮解码与早停机制。输入：输入纠正句子 x ，中文语法纠错模型 M ，提前停止分类器 C_e ，最大解码轮数 n ，提前停止阈值 τ ；输出：更正后的句子 y 。

```

1:  $\hat{x}_0, h_{d,0}^{<eos>} = M(x)$ ;
2:  $p_e = C_e(h_{d,0}^{<eos>})$ ;
3: if  $p_e > \tau$  then
4:    $y = \hat{x}_0$ 
5: else
6:   for  $t = 1$  to  $n$  do
7:      $\hat{x}_t, h_{d,t}^{<eos>} = M(x, \hat{x}_{t-1})$ ;
8:      $p_e = C_e(h_{d,t}^{<eos>})$ ;
9:      $y = \hat{x}_t$ 
10:    if  $\hat{x}_{t-1} == \hat{x}_t$  or  $p_e > \tau$  then
11:      break;
12:    end if
13:  end for
14: end if
15: return  $y$ 

```

2 基于序列到序列的中文语法纠错

序列到序列模型 M 接收待纠正的输入序列 x ，并输出修正后的序列 \hat{x} 。该模型由编码器和解码器两部分构成。

编码器接收输入序列 x ，并计算上下文隐状态向量 h_e ：

$$h_e = \text{encoder}(x) \quad (1)$$

解码器根据编码器的隐状态 h_e 和已生成的解码序列 $\hat{x}^{<k>}$ 来生成当前的隐状态 h_d^k ：

$$h_d^k = \text{decoder}(h_e, \hat{x}^{<k>}) \quad (2)$$

在解码过程中， \hat{x}^k 代表序列中的第 k 个标记， \hat{x}^0 对应于序列的开始标记 $\langle \text{sos} \rangle$ ，而 $\hat{x}^{<k>}$ 则表示从 \hat{x}^0 到 \hat{x}^{k-1} 的标记序列。解码器中的分类器利用隐状态 h_d^k 来预测词汇表中每个词语的概率，并选择概率最高的词语作为下一个解码步骤的 \hat{x}^k 。

$$\hat{x}^k = \text{classifier}(h_d^k) \quad (3)$$

解码器将持续执行此过程，直至分类器在接收到隐状态 $h_d^{<eos>}$ 时输出序列结束标记 $\langle \text{eos} \rangle$ 。

通过重构损坏文本进行预训练，大规模语料库的知识可被编码进模型参数。损坏文本是指从原始文本中选择部分词进行mask或换序等操作后的句子。在语法纠错任务上对预训练模型进行微调（如BART和T5）可进一步提升性能[Sun et al., 2021, Rothe et al., 2021]。

3 本文方法

3.1 结合早停机制的多轮解码

在语法纠错任务中，序列到序列语法纠错模型 M 输入可能包含语法错误的句子 x ，并输出修正后的句子 \hat{x} 。本文不直接采用 \hat{x} 作为最终输出，而是通过将上一轮的修正结果 \hat{x}_{t-1} 作为新一轮的输入，要求模型将其修正为 \hat{x}_t ，以此迭代地进行修正，直至 $\hat{x}_t = \hat{x}_{t-1}$ 。迭代的终止条件是连续两次解码产生相同的序列。尽管这种方法提升了性能，但也增加了推理的计算成本。为解决效率问题，本文训练了一个与序列到序列模型协同工作的早停机制。

早停机制包括一个轻量级的逻辑回归分类器 C_e ，用于预测提前停止多轮解码的概率 p_e 。 C_e 由权重向量 w_e 和偏置 b_e 组成。在解码 \hat{x}_{t-1} 时，本文使用特殊句子结束标记 $\langle \text{eos} \rangle$ 的隐状态表示 $h_{d,t-1}^{<eos>}$ 来计算早停的概率：

原始文本 (x):	我们明天从石家庄出发去北京中科院。
前一轮解码结果 (\hat{x}_{t-1}):	我们明天从石家庄出发去河北中国科学院。
混合序列:	我们明天从石家庄出发去北京河北中国科学院。
编辑标记:	e e e e e e e e e e e d d i i e i e i e e
位置编码:	1 2 3 4 5 6 7 8 9 10 11 12 13 0 0 14 0 15 0 16 17 1 2 3 4 5 6 7 8 9 10 11 0 0 12 13 14 15 16 17 18 19

图 1: 基于序列对比的源信息融合。对于编辑标记, “e”、“d”和“i”分别表示相等、删除和插入。“0”表示位置编码中的填充。

$$p_e = \sigma(h_{d,t-1}^{<eos>} \bullet w_e + b_e) \quad (4)$$

其中“ \bullet ”和“ σ ”分别表示向量点积和sigmoid激活函数。

本文使用 p_e 和早停标签 y_e 之间的二元交叉熵 (Binary Cross Entropy, BCE) 损失来训练早停分类器:

$$l_e = \text{BCE}(p_e, y_e) \quad (5)$$

在多层解码训练中, 本文首先解码得到本轮的解码结果 \hat{x}_t , 随后根据前一轮的解码结果 \hat{x}_{t-1} 、本轮的解码结果 \hat{x}_t 和参考答案 r 来标记早停分类器的分类标签 y_e , 如果: 1) 本轮的解码结果 \hat{x}_t 与前一轮的解码结果 \hat{x}_{t-1} 相同, 或2) 参考答案 r 与本轮解码结果 \hat{x}_t 之间的编辑距离大于与前一轮解码结果 \hat{x}_{t-1} 之间的编辑距离, 则分类标签 y_e 为真。引入第二个编辑距离条件旨在确保多层解码不会降低性能。

训练损失是原始序列到序列生成损失 $l_{seq2seq}$ 和编辑损失 l_e 的加权组合:

$$l = l_{seq2seq} + \lambda * l_e \quad (6)$$

本文在推理阶段采用算法 1。设定最大解码轮数 n 为3, 若 $\hat{x}_t = \hat{x}_{t-1}$ 或 $p_e > \tau$, 则触发早停。参数 λ 和 τ 分别默认设为1和0.5。将 λ 设为1意味着在训练过程中, 纠正任务和早停分类器被赋予同等重要性。阈值 $\tau = 0.5$ 表示当停止解码的概率超过0.5时提前终止多层解码。对于二分类任务, $\tau = 0.5$ 是一个合理的选择。最大解码轮数 n 通过在验证集上测试得到, 结果表明超过3轮后不会再提升性能。尽管进一步调整 λ 和 τ 可能会提高性能, 但这不是本文关注的重点, 因此本文未对此进行深入研究。

3.2 基于序列对比的源语句信息融合

在多层解码过程中, 若模型在前几轮中删除或替换了某些词语, 那么这些原始词语将无法用于后续轮次的纠正, 即便它们可能为后续纠正提供有价值的信息。

以图 1中的示例为例, 模型应将句子:

“我们明天从石家庄出发去北京中科院。”

纠正为:

“我们明天从石家庄出发去河北中国科学院。”

在第一轮中, 模型将“中科院”修正为“中国科学院”, 同时将“北京”错误地修改为“河北”。在第二轮中, 如果没有原始语句中的词语“北京”, 模型就无法确定词语“中科院”准确的地域信息, 可能会导致错误的纠错结果。“北京”一词的存在可以帮助模型确定原始语句中“中科院”一词的地域信息。

因此, 在所有纠正轮次中保留完整的源语句信息可能会有利于多层解码的性能。然而, 预训练的序列到序列模型通常不具备处理源语句 x 和前一轮纠正结果 \hat{x}_{t-1} 的双编码器结构。将源语

句 x 和前一轮的解码结果 \hat{x}_{t-1} 直接拼接作为编码器输入会生成一个长且包含冗余信息的序列，且未改变的词语在拼接序列中会出现在两个不同的位置，不利于模型对比分析两个序列中对应位置的词语。为了有效地使用单个编码器处理源语句 x 和前一轮的解码结果 \hat{x}_{t-1} ，本文提出了一种将两者合并为单一序列的方法，如图1所示。

具体来说，本文首先比较源语句 x 和前一轮的解码结果 \hat{x}_{t-1} ，识别出它们相同的和不同的部分。然后根据它们在原序列中的顺序将它们合并成一个序列。合并后的序列包含了未改变的词语、插入的词语和删除的词语，且保留它们的原始顺序。替换操作可以被视为插入和删除的组合。

本文通过编辑标记或特定的位置编码来区分合并序列中各词语的角色。编辑标记使用“e”（相等）、“d”（删除）和“i”（插入）来标识合并序列中词语的状态，分别对应于同时出现在源语句 x 和前一轮的解码结果 \hat{x}_{t-1} 中的词语、仅在源语句 x 中出现的词语及新添加到上一轮解码结果 \hat{x}_{t-1} 中的词语。

本文引入一个特征向量层以学习编辑标记的表示，并在计算编码器层之前将这些编辑特征向量添加到序列到序列模型的词向量中。对于位置编码，本文在合并序列中使用了两个位置编码：源位置编码表示字符在源语句 x 中的位置，解码结果的位置编码表示其在上一轮解码结果 \hat{x}_{t-1} 中的位置。如果词语没有在某个序列中出现，则其位置编码为0。为了减少使用合并序列作为输入时新位置向量与预训练模型位置向量之间的差异，新的位置向量是基于预训练位置向量初始化的。然而，因为在使用合并序列作为输入时，位置向量会被累加两次：一次用于源位置，一次用于解码结果的位置，本文将预训练位置向量的权重减半。

4 实验

4.1 设置

为验证多轮解码方法的有效性，本文严格遵循Zhang et al. [2022a]的数据处理流程。本文使用NLPCC2018官方提供的Lang-8语料集、HSK汉语考试语料集和FCGEC训练集的组合作为训练集。NLPCC2018的训练集是从lang8收集处理得到的，是汉语的母语使用者对汉语学习者作文的修改，处理后包含1092285条数据，测试集来源于北京大学中文系构建的北京大学汉语学习者语料库，该语料库包含外国大学生的作文，旨在促进国际教育和汉语作为中介语言的研究，包含2000条数据。HSK动态作文语料库是外国汉语学习者参加等汉语水平考试等的作文答卷，处理后包含95320条数据。使用MuCGEC-dev[Zhang et al., 2022a]作为验证集，同时在NLPCC2018测试集、FCGEC的验证集[Xu et al., 2022]和NaCGEC测试集[Ma et al., 2022]上进行测试。

本文使用large设置的Transformer和预训练BART模型*作为基线，模型的训练和解码训练依据Yang and Quan [2024]的实验设定，批量大小使用1024，训练数据的最大句长使用128，最大训练周期数分别是20和10，解码的柱搜索大小是10。

为了与现有研究保持一致，对于NLPCC-2018测试数据集，本文采用官方的PKUNLP†工具对模型输出结果进行分词，并使用官方的MaxMatch (M2) ‡工具计算精度（Precision, P）、召回率（Recall, R）和F0.5值。对于FCGEC验证集和NaCGEC测试集，本文使用ChERRANT进行基于字符的评估，不受分词工具对结果的影响。

4.2 主要结果

根据消融实验的结果，多轮解码的训练仅采用单轮解码的结果，推理采用多轮解码和早停机制（详见第4.3节），同时使用编辑标记和位置编码来融合源语句信息（第4.4节）。在NLPCC2018测试集的结果如表1所示。FCGEC验证集和NaCGEC测试集的结果如表2所示。

表1和2表明：1) 尽管LLaMa2-7B和BaiChuan-7B大语言模型微调后的性能也较强，但其表现仍落后于经过微调的BART模型。2) 多轮解码显著且一致地提升了所有基线模型的性能，相对于Transformer和BART，F0.5值分别提高了+2.78和+2.06（NLPCC2018测试集）、+6.78和+2.30（FCGEC验证集）和+4.99和+3.45（NaCGEC测试集）。3) F0.5值超过

<http://lang-8.com/>

*<https://huggingface.co/fnlp/bart-large-chinese>

†<https://github.com/biug/pkunlp>

‡<https://github.com/nusnlp/m2scorer>

模型	P	R	F0.5
大模型（零样本）			
LLaMa2-7B	11.79	11.46	11.72
BaiChuan-7B	20.87	23.28	21.31
大模型（微调）			
LLaMa2-7B	45.85	27.44	40.43
BaiChuan-7B	51.69	27.92	44.17
序列到编辑			
BERT-base-Chinese[Devlin et al., 2019]	41.38	24.55	36.39
HRG [Hinson et al., 2020]	36.79	27.82	34.56
SG-GEC [Wu and Wu, 2022]	50.56	25.24	42.11
序列到序列			
AliGM[Zhao et al., 2018]	41.00	13.75	29.36
YouDao[Fu et al., 2018]	35.24	18.64	29.91
BLCU[Li et al., 2019]	47.63	12.56	30.57
Qiu and Qu [2019]	36.88	18.94	31.01
MaskGEC[Zhao and Wang, 2020a]	44.36	22.18	36.97
GPT2-Chinese [Du, 2019]	41.94	36.13	40.63
TEA[王辰成 et al., 2020]	39.43	22.8	34.41
WCDA[Tang et al., 2021]	47.29	23.89	39.49
Copy[Zhao et al., 2019]	51.25	32.55	45.97
SynGEC[Zhang et al., 2022b]	49.96	33.04	45.32
TemplateGEC [Li et al., 2023]	54.50	27.40	45.50
Alirector [Yang and Quan, 2024]	51.76	33.49	46.67
本文			
Transformer	42.37	23.49	36.50
+ 多轮解码	46.64	24.08	39.28
BART	50.63	31.83	45.28
+ 多轮解码	52.56	33.89	47.34

表 1: 在NLPC2018测试集上的结果

Method	FCGEC验证集			NaCGEC测试集		
	P	R	F0.5	P	R	F0.5
Transformer	47.83	22.99	39.33	59.67	28.69	49.07
+ 多轮解码	58.67	24.76	46.06	65.61	31.73	54.06
BART	56.26	40.71	52.27	65.85	40.79	58.64
+ 多轮解码	59.21	41.57	54.58	68.82	44.64	62.09

表 2: 在FCGEC验证集和NaCGEC测试集上的结果

设置	FCGEC验证集			NaCGEC测试集			Speed
	P	R	F0.5	P	R	F0.5	
BART	56.26	40.71	52.27	65.85	40.79	58.64	1.00x
训练							
单次	59.21	41.57	54.58	68.82	44.64	62.09	0.83x
多轮	57.86	42.45	53.94	67.51	43.97	60.98	
解码							
策略模型[2018]	58.07	40.09	53.29	67.22	43.52	60.62	0.27x
无早停机制							
$n = 1$	58.47	40.91	53.85	67.66	43.42	60.86	0.46x
$n = 2$	58.13	41.09	53.68	67.24	43.48	60.62	0.41x
$n = 3$	58.18	41.21	53.75	67.27	43.68	60.71	0.38x
早停机制, $n = 3$	59.21	41.57	54.58	68.82	44.64	62.09	0.83x

表 3: 各种多轮解码训练和推理设置的结果。速度是在NaCGEC测试集上的推理速度。

了最新的Alirector模型，提高了+0.67。尽管本文仅在广泛使用的Transformer和BART基线上应用了本文的方法，但本文认为本文的方法也可以用于更先进的基线模型，并带来更好的性能。

4.3 多轮解码训练和推理的消融实验

除了训练模型生成给定输入源语句 x 的参考答案 r 之外，多轮解码训练还可以进行单轮解码，如Omelianchuk et al. [2020]所述，或者进行几轮解码直至满足推理终止条件。本文探讨了在推理时采用多轮解码并带有早停机制的多轮解码训练中，单次解码与多轮解码的效果。其中，“多轮解码”是指解码方法，模型训练需要使用之前的解码结果，此结果可由单次解码产生，也可由多轮解码产生。

对于单次多轮解码训练，本文使用模型将输入源语句 x 解码为单次解码结果 \hat{x}_0 ，并训练模型在给定源语句 x 和单次解码结果 \hat{x}_0 的情况下生成参考答案 r ：

$$M(x, \hat{x}_0) \rightarrow r \quad (7)$$

对于多遍多轮解码训练，本文以源语句 x 作为初始解码结果 \hat{x}_{-1} 开始，迭代地将前一轮解码结果 \hat{x}_{i-1} 解码为本轮解码结果 \hat{x}_i ，重复数轮直至满足终止条件，并训练模型在给定源语句 x 和本轮解码结果 \hat{x}_i 的情况下生成参考答案 r ：

$$M(x, \hat{x}_i) \rightarrow r \quad (8)$$

本文还探讨了在多轮解码推理中最大解码轮数，以及是否使用早停机制的影响，采用单次多轮解码进行训练。此外，本文对比了本文提出的早停机制与Geng et al. [2018]提出的策略模型。Geng et al. [2018]采用强化学习方法，通过评估连续两次解码遍之间的差异来确定解码轮数，并针对机器翻译任务使用BLEU作为奖励优化策略模型。在本文的语法纠错任务实验中，奖励基于F0.5得分而非BLEU，结果如表3所示。其中，“训练”模块中的多轮是指模型在训练时使用带有早停机制的多轮解码算法产生训练所需的上一轮解码结果，具体的轮数受早停机制影响是变化的。

表3表明：1) 对于多轮解码训练，两种设置均能带来性能提升，但单次解码设置在训练效率更高的情况下，相较于多轮解码，在FCGEC验证集和NaCGEC测试集上都取得了更高的F0.5值（+0.64和+1.11）。2) 不同的最大解码轮数取得了相近的F0.5值，更大的最大解码轮数会导致推理速度下降，但早停机制可以有效缓解这一问题，并取得了最佳的性能，3) 基于策略网络的多轮解码在两个共享任务上同样带来了F0.5的改进，但本文简单的早停方法比策略网络[Geng et al., 2018]更有效率，并在FCGEC验证集和NaCGEC测试集上都取得了更高的F0.5值（+1.29，+1.47）。

方法	FCGEC验证集			NaCGEC测试集		
	P	R	F0.5	P	R	F0.5
BART	56.26	40.71	52.27	65.85	40.79	58.64
None	56.72	43.11	53.35	65.64	46.13	60.52
Concat	57.32	42.15	53.47	66.90	44.11	60.63
Edit	57.59	42.97	53.92	67.02	44.34	60.80
Pos	57.86	42.81	54.06	67.41	44.00	60.93
Pos+edit	59.21	41.57	54.58	68.82	44.64	62.09

表 4: 不同源语句信息融合方法的结果

现有最好的机器翻译多轮解码研究[Geng et al., 2018]采用了复杂的强化学习方法来决定所需的解码轮数。然而，强化学习训练可能不稳定，导致性能波动。本文的监督方法直接基于解码序列的向量表示训练简单的二分类器，本文认为本文的早停方法易于实现，并且在实践中证明了有效性。

4.4 源语句信息融合的影响分析

本文评估了不同源语句信息融合方法对BART的影响，包括：1) 在多轮解码推理中仅使用前一轮的解码结果 \hat{x}_{t-1} 而非同时使用前一轮的解码结果 \hat{x}_{t-1} 和源语句 x (“None”)，2) 序列拼接 (“Concat”)，3) 编辑标记 (“Edit”)，4) 位置编码 (“Pos”)，以及5) 编辑标记和位置编码的组合 (“Pos+Edit”)。实验结果如表4所示。

表4表明：1) 未融合源语句信息的普通多轮解码 (“None”) 在FCGEC验证集和NaCGEC测试集上分别带来了+1.08和+1.88的F0.5值提升，证明了多轮解码的有效性，2) 通过序列拼接 (“Concat”) 提供源语句信息较没有源语句信息 (“None”) 可以在FCGEC验证集和NaCGEC测试集上分别带来了+0.12和+0.11的F0.5值提升，证明了源语句信息的正面影响，3) 位置编码 (“Pos”) 和编辑标签 (“Edit”) 都显著优于序列拼接，表明基于序列对比的源语句信息融合方法的重要性，其中位置编码的性能略优于编辑标签，这可能是由于预训练位置向量的初始化，4) 位置编码与编辑标记的组合 (“Pos+Edit”) 带来了最佳性能。

5 相关研究

用于语法纠错的序列到编辑方法 可将语法纠错任务视为序列标注任务，通过迭代预测句子中每个位置的编辑标记。受GECToR[Omelianchuk et al., 2020]启发，Liang et al. [2020]将中文语法纠错任务视作序列标注任务。Zhang et al. [2022a]直接采用GECToR，并利用预训练模型进行增强。TtT[Li and Shi, 2021]提出了一种非自回归语法纠错方法，结合了BERT[Devlin et al., 2019]编码器和CRF编码器[Lafferty et al., 2001]。Li et al. [2022b]则提出一个序列到编辑模型，用于中文语法纠错任务，在每个字符前预测一个动作，动作由跳过、复制和生成三类构成，然后将这些动作与基本序列到序列框架融合，提供最终预测。Tan et al. [2023]提出了一种基于纠错类型的校正方法，在每个训练实例中，额外构建多个涉及特定类型错误训练实例，然后使用这些额外构造的训练实例和原始的训练实例依次训练模型。序列到编辑模型的一个局限性在于其对编辑操作手动构建的依赖，这降低了模型的可迁移性并可能影响输出的流畅性[Li et al., 2023]。上述方法均使用了多轮解码。

用于语法纠错的序列到序列方法 可将语法纠错视为序列生成任务[Fang et al., 2023, Li et al., 2022a, Liu et al., 2021b, Wang et al., 2021]。早期研究主要采用基于CNN[Ren et al., 2018]或RNN[Zhou et al., 2018]的模型来解决语法纠错任务。随着transformer[Vaswani et al., 2017]在机器翻译任务中取得巨大成功，它也被应用于语法纠错任务。HRG[Malmi et al., 2019]结合了拼写检查器、NMT-base模型和序列编辑模型，但其拼写检查器基于语言模型，未能充分利用上下文。Kaneko et al. [2020]首先对MLM模型进行微调，然后将微调后的模型输出作为额外特征用于语法纠错模型，以最大化MLM模型的优势。Zhao and Wang [2020b]提出了数据增强方法MaskGEC，该方法在训练过程中动态地对输入句子添加随机噪声。Tang et al. [2021]提出了使用数据增强的方法来扩展训练数据，以提升模型性能。Zhou et al. [2020]通过编

辑式投票机制集成了序列到序列模型和序列到编辑模型。Zhang et al. [2022a]提出了基于序列到序列架构的中文BART模型，适用于中文语法纠错任务。Fei et al. [2023], Zhang et al. [2022b]通过添加额外的语法知识或利用特定语言知识来提高纠错性能。Sun and Wang [2022]提出通过权衡精确度-召回率来提高纠错性能。Zhang et al. [2023]提出通过重排序来选择正确率更高的纠错结果，从而提高纠错性能。

多轮解码在NMT中的应用 多轮解码已被用于改进神经机器翻译（NMT）[Xia et al., 2017, Mahmood et al., 2017, Zhang et al., 2018, Geng et al., 2018, Liu et al., 2019]。自动后编辑（APE）也可视为多轮解码的一种特例[Correia and Martins, 2019, Pal et al., 2020, Bhattacharyya et al., 2022, Jung et al., 2023]。这些研究均强调了源信息融合对性能提升的重要性，但它们在源输入和已解码序列上采用了双编码器结构，这主要是因为它们采用不同的语言，这两种语言在拼写和结构上的差异。然而，在语法纠错任务中，源句子和解码序列通常较为接近。本文首次：1) 研究了中文语法纠错任务的序列到序列模型多轮解码，2) 提出了一种简单有效的早停机制来解决效率问题，3) 提出了一种将源语句和解码序列合并的方法，以适应只有一个编码器的预训练序列到序列模型。

6 总结

本文针对现有的序列到序列纠错方法仅进行一次从左到右的解码，在解码过程中不能考虑后续词语的问题，提出通过多轮解码来重复检查纠正现有的纠错结果，改进序列到序列语法纠错方法。为提升多轮解码的效率，本文提出早停机制，在下一轮可能不会产生更好的纠错结果时，提前结束多轮解码。本文提出源信息融合方法来解决多轮解码中删除和替换词语导致的信息丢失问题，并根据两个序列相近的特性，提出基于对比的序列混合方法，避免序列拼接导致的效率问题。

在NLPCC2018测试集、FCGEC验证集和NaCGEC测试集上的实验表明，本文的方法相对于强大的BART-large基线带来了+2.06, +2.31和+3.45的F0.5值提升，分别取得了47.34, 54.58和62.09的F0.5值，证明了本文方法的有效性。

致谢

感谢审稿人帮助改善论文的建议。本文受国家自然科学基金青年项目（62306284），中国博士后科学基金面上项目（2023M743189），河南省自然科学基金青年项目（232300421386）资助。

References

- Pushpak Bhattacharyya, Rajen Chatterjee, Markus Freitag, Diptesh Kanojia, Matteo Negri, and Marco Turchi. Findings of the wmt 2022 shared task on automatic post-editing. In *Proceedings of the Seventh Conference on Machine Translation*, pages 109–117, Abu Dhabi, December 2022. Association for Computational Linguistics. URL <https://aclanthology.org/2022.wmt-1.5>.
- Gonçalo M. Correia and André F. T. Martins. A simple and effective approach to automatic post-editing with transfer learning. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3050–3056, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1292. URL <https://aclanthology.org/P19-1292>.
- Michael J. Denkowski. Machine translation for human translators. 2015. URL <https://api.semanticscholar.org/CorpusID:8279024>.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL <https://aclanthology.org/N19-1423>.

- Zeyao Du. Gpt2-chinese: Tools for training gpt2 model in chinese language. <https://github.com/Morizeyao/GPT2-Chinese>, 2019.
- Tao Fang, Jinpeng Hu, Derek F. Wong, Xiang Wan, Lidia S. Chao, and Tsung-Hui Chang. Improving grammatical error correction with multimodal feature integration. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9328–9344, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-acl.594. URL <https://aclanthology.org/2023.findings-acl.594>.
- Yuejiao Fei, Leyang Cui, Sen Yang, Wai Lam, Zhenzhong Lan, and Shuming Shi. Enhancing grammatical error correction systems with explanations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7489–7501, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.413. URL <https://aclanthology.org/2023.acl-long.413>.
- Kai Fu, Jin Huang, and Yitao Duan. Youdao’s winning solution to the nlpcc-2018 task 2 challenge: A neural machine translation approach to chinese grammatical error correction. In *Natural Language Processing and Chinese Computing*, 2018. URL <https://api.semanticscholar.org/CorpusID:52007443>.
- Xinwei Geng, Xiaocheng Feng, Bing Qin, and Ting Liu. Adaptive multi-pass decoder for neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 523–532, Brussels, Belgium, October–November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1048. URL <https://aclanthology.org/D18-1048>.
- Charles Hinson, Hen-Hsen Huang, and Hsin-Hsi Chen. Heterogeneous recycle generation for Chinese grammatical error correction. In Donia Scott, Nuria Bel, and Chengqing Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2191–2201, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics. doi: 10.18653/v1/2020.coling-main.199. URL <https://aclanthology.org/2020.coling-main.199>.
- Baikjin Jung, Myungji Lee, Jong-Hyeok Lee, and Yunsu Kim. Bring more attention to syntactic symmetry for automatic postediting of high-quality machine translations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1433–1441, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-short.122. URL <https://aclanthology.org/2023.acl-short.122>.
- Masahiro Kaneko, Masato Mita, Shun Kiyono, Jun Suzuki, and Kentaro Inui. Encoder-decoder models can benefit from pre-trained masked language models in grammatical error correction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4248–4254, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.391. URL <https://aclanthology.org/2020.acl-main.391>.
- Shun Kiyono, Jun Suzuki, Masato Mita, Tomoya Mizumoto, and Kentaro Inui. An empirical study of incorporating pseudo data into grammatical error correction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1236–1242, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1119. URL <https://aclanthology.org/D19-1119>.
- John D. Lafferty, Andrew McCallum, and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *International Conference on Machine Learning*, 2001. URL <https://api.semanticscholar.org/CorpusID:219683473>.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-

- training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.703. URL <https://aclanthology.org/2020.acl-main.703>.
- Bei Li, Quan Du, Tao Zhou, Yi Jing, Shuhan Zhou, Xin Zeng, Tong Xiao, JingBo Zhu, Xuebo Liu, and Min Zhang. ODE transformer: An ordinary differential equation-inspired model for sequence generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8335–8351, Dublin, Ireland, May 2022a. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.571. URL <https://aclanthology.org/2022.acl-long.571>.
- Jiquan Li, Junliang Guo, Yongxin Zhu, Xin Sheng, Deqiang Jiang, Bo Ren, and Linli Xu. Sequence-to-action: Grammatical error correction with action guided sequence generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10):10974–10982, Jun. 2022b. doi: 10.1609/aaai.v36i10.21345. URL <https://ojs.aaai.org/index.php/AAAI/article/view/21345>.
- Piji Li and Shuming Shi. Tail-to-tail non-autoregressive sequence prediction for Chinese grammatical error correction. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4973–4984, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.385. URL <https://aclanthology.org/2021.acl-long.385>.
- Si Li, Jianbo Zhao, Guirong Shi, Yuanpeng Tan, Huifang Xu, Guang Chen, Haibo Lan, and Zhiqing Lin. Chinese grammatical error correction based on convolutional sequence to sequence model. volume 7, pages 72905–72913, 2019. doi: 10.1109/ACCESS.2019.2917631.
- Yinghao Li, Xuebo Liu, Shuo Wang, Peiyuan Gong, Derek F. Wong, Yang Gao, Heyan Huang, and Min Zhang. TemplateGEC: Improving grammatical error correction with detection template. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6878–6892, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.380. URL <https://aclanthology.org/2023.acl-long.380>.
- Deng Liang, Chen Zheng, Lei Guo, Xin Cui, Xiuzhang Xiong, Hengqiao Rong, and Jinpeng Dong. BERT enhanced neural machine translation and sequence tagging model for Chinese grammatical error diagnosis. In Erhong YANG, Endong XUN, Baolin ZHANG, and Gaoqi RAO, editors, *Proceedings of the 6th Workshop on Natural Language Processing Techniques for Educational Applications*, pages 57–66, Suzhou, China, December 2020. Association for Computational Linguistics. URL <https://aclanthology.org/2020.nlp tea-1.8>.
- Cao Liu, Shizhu He, Kang Liu, and Jun Zhao. Vocabulary pyramid network: Multi-pass encoding and decoding with multi-level vocabularies for response generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3774–3783, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1367. URL <https://aclanthology.org/P19-1367>.
- Xuebo Liu, Longyue Wang, Derek F. Wong, Liang Ding, Lidia S. Chao, and Zhaopeng Tu. Understanding and improving encoder layer fusion in sequence-to-sequence learning. 2021a. URL <https://openreview.net/forum?id=n1HD8M6WGn>.
- Xuebo Liu, Longyue Wang, Derek F. Wong, Liang Ding, Lidia S. Chao, and Zhaopeng Tu. Understanding and improving encoder layer fusion in sequence-to-sequence learning. 2021b. URL <https://openreview.net/forum?id=n1HD8M6WGn>.
- Shirong Ma, Yinghui Li, Rongyi Sun, Qingyu Zhou, Shulin Huang, Ding Zhang, Li Yangning, Ruiyang Liu, Zhongli Li, Yunbo Cao, Haitao Zheng, and Ying Shen. Linguistic rules-based

- corpus generation for native Chinese grammatical error correction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 576–589, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-emnlp.40. URL <https://aclanthology.org/2022.findings-emnlp.40>.
- Rehan Mahmood, Zulin Wang, and Qin Huang. Multi-pass decoding for the robust transmission of deep-space images. In *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pages 1–5, 2017. doi: 10.1109/VTCspring.2017.8108437.
- Eric Malmi, Sebastian Krause, Sascha Rothe, Daniil Mirylenka, and Aliaksei Severyn. Encode, tag, realize: High-precision text editing. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5054–5065, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1510. URL <https://aclanthology.org/D19-1510>.
- Hwee Tou Ng, Siew Mei Wu, Yuanbin Wu, Christian Hadiwinoto, and Joel Tetreault. The CoNLL-2013 shared task on grammatical error correction. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–12, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL <https://aclanthology.org/W13-3601>.
- Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. The CoNLL-2014 shared task on grammatical error correction. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–14, Baltimore, Maryland, June 2014. Association for Computational Linguistics. doi: 10.3115/v1/W14-1701. URL <https://aclanthology.org/W14-1701>.
- Kostiantyn Omelianchuk, Vitaliy Atrasevych, Artem Chernodub, and Oleksandr Skurzhanyski. 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 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.bea-1.16. URL <https://aclanthology.org/2020.bea-1.16>.
- Santanu Pal, Hongfei Xu, Nico Herbig, Sudip Kumar Naskar, Antonio Krüger, and Josef van Genabith. The transference architecture for automatic post-editing. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5963–5974, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics. doi: 10.18653/v1/2020.coling-main.524. URL <https://aclanthology.org/2020.coling-main.524>.
- Zhaoquan Qiu and Youli Qu. A two-stage model for chinese grammatical error correction. *IEEE Access*, 7:146772–146777, 2019. doi: 10.1109/ACCESS.2019.2940607.
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1), jan 2020. ISSN 1532-4435.
- Hongkai Ren, Liner Yang, and Endong Xun. A sequence to sequence learning for chinese grammatical error correction. In Min Zhang, Vincent Ng, Dongyan Zhao, Sujian Li, and Hongying Zan, editors, *Natural Language Processing and Chinese Computing*, pages 401–410, Cham, 2018. Springer International Publishing.
- Sascha Rothe, Jonathan Mallinson, Eric Malmi, Sebastian Krause, and Aliaksei Severyn. A simple recipe for multilingual grammatical error correction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 702–707, Online,

- August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-short.89. URL <https://aclanthology.org/2021.acl-short.89>.
- Felix Stahlberg and Shankar Kumar. Seq2Edits: Sequence transduction using span-level edit operations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5147–5159, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.418. URL <https://aclanthology.org/2020.emnlp-main.418>.
- Xin Sun and Houfeng Wang. Adjusting the precision-recall trade-off with align-and-predict decoding for grammatical error correction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 686–693, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-short.77. URL <https://aclanthology.org/2022.acl-short.77>.
- Xin Sun, Tao Ge, Furu Wei, and Houfeng Wang. Instantaneous grammatical error correction with shallow aggressive decoding. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5937–5947, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.462. URL <https://aclanthology.org/2021.acl-long.462>.
- Minghuan Tan, Min Yang, and Ruifeng Xu. Focal training and tagger decouple for grammatical error correction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5978–5985, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-acl.370. URL <https://aclanthology.org/2023.findings-acl.370>.
- Zecheng Tang, Yixin Ji, Yibo Zhao, and Junhui Li. 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 Sheng Li, Maosong Sun, Yang Liu, Hua Wu, Kang Liu, Wanxiang Che, Shizhu He, and Gaoqi Rao, editors, *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 813–824, Huhhot, China, August 2021. Chinese Information Processing Society of China. URL <https://aclanthology.org/2021.ccl-1.73>.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.
- Yu Wang, Yuelin Wang, Kai Dang, Jie Liu, and Zhuo Liu. A comprehensive survey of grammatical error correction. *ACM Trans. Intell. Syst. Technol.*, 12(5), dec 2021. ISSN 2157-6904. doi: 10.1145/3474840. URL <https://doi.org/10.1145/3474840>.
- Xiuyu Wu and Yunfang Wu. From spelling to grammar: A new framework for Chinese grammatical error correction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 889–902, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-emnlp.63. URL <https://aclanthology.org/2022.findings-emnlp.63>.
- Yingce Xia, Fei Tian, Lijun Wu, Jianxin Lin, Tao Qin, Nenghai Yu, and Tie-Yan Liu. Deliberation networks: Sequence generation beyond one-pass decoding. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/c6036a69be21cb660499b75718a3ef24-Paper.pdf.

- Lvxiaowei Xu, Jianwang Wu, Jiawei Peng, Jiayu Fu, and Ming Cai. FCGEC: Fine-grained corpus for Chinese grammatical error correction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 1900–1918, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-emnlp.137. URL <https://aclanthology.org/2022.findings-emnlp.137>.
- Haihui Yang and Xiaojun Quan. Alirector: Alignment-enhanced chinese grammatical error corrector. *ArXiv*, abs/2402.04601, 2024. URL <https://arxiv.org/abs/2402.04601>.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/dc6a7e655d7e5840e66733e9ee67cc69-Paper.pdf.
- Zheng Yuan, Shiva Taslimipoor, Christopher Davis, and Christopher Bryant. Multi-class grammatical error detection for correction: A tale of two systems. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8722–8736, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.687. URL <https://aclanthology.org/2021.emnlp-main.687>.
- Xiangwen Zhang, Jinsong Su, Yue Qin, Yang Liu, Rongrong Ji, and Hongji Wang. Asynchronous bidirectional decoding for neural machine translation. AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018. ISBN 978-1-57735-800-8.
- Ying Zhang, Hidetaka Kamigaito, and Manabu Okumura. Bidirectional transformer reranker for grammatical error correction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 3801–3825, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-acl.234. URL <https://aclanthology.org/2023.findings-acl.234>.
- Yue Zhang, Zhenghua Li, Zuyi Bao, Jiacheng Li, Bo Zhang, Chen Li, Fei Huang, and Min Zhang. MuCGEC: a multi-reference multi-source evaluation dataset for Chinese grammatical error correction. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *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 2022a. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.227. URL <https://aclanthology.org/2022.naacl-main.227>.
- Yue Zhang, Bo Zhang, Zhenghua Li, Zuyi Bao, Chen Li, and Min Zhang. SynGEC: Syntax-enhanced grammatical error correction with a tailored GEC-oriented parser. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 2518–2531, Abu Dhabi, United Arab Emirates, December 2022b. Association for Computational Linguistics. doi: 10.18653/v1/2022.emnlp-main.162. URL <https://aclanthology.org/2022.emnlp-main.162>.
- Wei Zhao, Liang Wang, Kewei Shen, Ruoyu Jia, and Jingming Liu. Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 156–165, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1014. URL <https://aclanthology.org/N19-1014>.
- Yuanyuan Zhao, Nan Jiang, Weiwei Sun, and Xiaojun Wan. Overview of the NLPCC 2018 shared task: Grammatical error correction. In Min Zhang, Vincent Ng, Dongyan Zhao, Sujian Li, and

- Hongying Zan, editors, *Natural Language Processing and Chinese Computing - 7th CCF International Conference, NLPCC 2018, Hohhot, China, August 26-30, 2018, Proceedings, Part II*, volume 11109 of *Lecture Notes in Computer Science*, pages 439–445. Springer, 2018. doi: 10.1007/978-3-319-99501-4_41. URL https://doi.org/10.1007/978-3-319-99501-4_41.
- Zewei Zhao and Houfeng Wang. Maskgec: Improving neural grammatical error correction via dynamic masking. In *AAAI Conference on Artificial Intelligence*, 2020a. URL <https://api.semanticscholar.org/CorpusID:213955573>.
- Zewei Zhao and Houfeng Wang. Maskgec: Improving neural grammatical error correction via dynamic masking. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):1226–1233, Apr. 2020b. doi: 10.1609/aaai.v34i01.5476. URL <https://ojs.aaai.org/index.php/AAAI/article/view/5476>.
- Junpei Zhou, Chen Li, Hengyou Liu, Zuyi Bao, Guangwei Xu, and Linlin Li. *Chinese Grammatical Error Correction Using Statistical and Neural Models: 7th CCF International Conference, NLPCC 2018, Hohhot, China, August 26–30, 2018, Proceedings, Part II*, pages 117–128. 08 2018. ISBN 978-3-319-99500-7. doi: 10.1007/978-3-319-99501-4_10.
- Wangchunshu Zhou, Tao Ge, Chang Mu, Ke Xu, Furu Wei, and Ming Zhou. Improving grammatical error correction with machine translation pairs. In Trevor Cohn, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 318–328, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.30. URL <https://aclanthology.org/2020.findings-emnlp.30>.
- 王辰成, 杨麟儿, 王莹莹, 杜永萍, and 杨尔弘. 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, 2020. URL <http://jcip.cipsc.org.cn/CN/Y2020/V34/I6/106>.