

CCL2022-CLTC赛道二：基于序列到编辑的多阶段中文语法纠错

穆玉芝	陈志刚	徐雪帆	周正茂	韩伟	张健
达观数据	达观数据	达观数据	达观数据	达观数据	达观数据
muyuzhi@	chenzhigang@	xuxuefan@	zhouzhengmao@	hanwei@	zhangjian@
datagrand.com	datagrand.com	datagrand.com	datagrand.com	datagrand.com	datagrand.com

摘要

中文语法纠错任务旨在对文本中存在的拼写、语法等错误进行自动检测和纠正，是自然语言处理领域一项重要的任务。本文基于序列到编辑的模型融合了局部最优路径和困惑度等方法，从而使得模型能够更好地融合局部和全局的特征信息，提升模型的纠错能力。我们还针对数据中包含的错误类型利用两阶段的拼写纠错和语法纠错模型以融合不同模型的纠错能力。同时，针对训练数据中存在一些错误数据而且有效数据较少的情况，我们采用基于规则的数据增强技术生成了一定数量的训练数据，并进一步提升了模型的纠错效果。

关键词： 序列标注；数据增强；困惑度；语法纠错

CCL2022-CLTC Track 2: Multi-Stage Chinese Grammatical Error Correction Based on Seq2Edit Model

Yuzhi Mu	Zhigang Chen	Xuefan Xu	Zhengmao Zhou	Wei Han	Jian Zhang
DataGrand	DataGrand	DataGrand	DataGrand	DataGrand	DataGrand
muyuzhi@	chenzhigang@	xuxuefan@	zhouzhengmao@	hanwei@	zhangjian@
datagrand.com	datagrand.com	datagrand.com	datagrand.com	datagrand.com	datagrand.com

Abstract

The Chinese Grammatical Error Correction task aims to automatically detect and correct the grammatical errors in the text. It's an important task in the field of natural language processing. This paper is based on the Seq2Edit model and integrates with local optimal path and perplexity to merge local move edits feature and global semantic information. We also design specific modules targeting at particular errors, including spelling and grammatical, etc. Since there are some erroneous data in the training data and there are few valid data, we use the rule-based data augmentation to generate amount of training data, and further improve the error correction effect of the model.

Keywords: Sequence Labeling , Data Augment , Grammar Correction , Perplexity

1 引言

随着汉语在世界范围内的流行，越来越多的外国人开始学习汉语；如何快速地对汉语学习者的写作内容进行检错和纠正具有实际的应用价值。本赛道重点关注汉语学习者的书写内容，并对其中包含的拼写、语法和标点符号等错误进行自动检测和修改，从而获得符合原意的正确句子。

为了实现语法错误的检测和纠正，我们基于Seq2Edit模型并结合局部最优路径和模型集成等方法获得模型的最优输出¹。首先我们对数据集进行分析和统计以了解数据中各错误类型的占比，同时对杂乱数据进行清洗，以降低错误的特征映射对模型拟合的影响；然后我们针对不同类型的错误构建拼写纠错，序列标注等模型以融合不同模型的优势，增强系统的纠错能力。下面，我们将分别从数据和模型的表现等方面来说明本系统的效果和能力。

2 背景

本评测报告参与的是CCL 2022的中文语法纠错任务 [Wang et al., 2022]，本赛道主要考察系统对汉语学习者写作内容的语法纠错能力，本次比赛数据中的错误类型包括用词错误、缺失、乱序、冗余以及符号等错误。由表1所示的数据样例可知，输入是可能包含错误的句子，输出是纠正后的句子，但是最终提交的是纠正后句子的编辑操作，包括错误的起止位置，错误类型和修改答案。对于本赛道的语法纠错任务必须正确检测到错误位置并完成纠正才算完整的纠错，其中任意环节出了差错均会影响到最后的效果。该赛道采用多种评价指标来更加合理地评估系统的纠错能力，包括假阳性、侦测层、识别层、定位层和修正层，并以各指标的综合得分评价系统的性能。

输入句子	输出句子	输出编辑
因为我的皮肤和嘴唇容易粗糙了。	因为我的皮肤和嘴唇在冬天容易变得粗糙。	10,10,M,在冬天 12,12,M,变得 14,14,R
还有是一个很好的机会家人在一起	还有是家人在一起一个很好的机会	4,15, W
在日本吸烟者越来越少，可是一部分的人还很爱吸烟。我觉得如果勉强禁止吸烟的话，对吸烟者不公平。所以现在我们应该做的是考虑吸烟的好处和坏处，承认两方面的意见。	在日本吸烟者越来越少，可是一部分人还很爱吸烟。我觉得如果强迫禁止吸烟的话，对吸烟者不公平。所以现在我们应该做的是考虑吸烟的好处和坏处，考虑两方面的意见。	17,17,R 30,31,S,强迫 69,70,S,考虑

Table 1: 数据错误类型示例

2.1 数据分析

本次比赛中，官方提供了CGED的历年比赛数据(41,239条)和Lang8数据(1212,457条)供参赛队伍训练模型，同时提供了3767条评测数据用来验证模型的效果和性能。因评测数据未开放标注结果，因此在构建系统解决语法错误的过程中，我们尝试对CGED-21中的验证集进行分析，总结数据集中的错误分布以及测验数据的质量。CGED-21验证集中的错误分布情况如图1所示，由此我们可以看出数据集中占绝大多数的均为用词错误，其次为缺失错误，而乱序错误的占比最少。

在数据测验的过程中，我们发现数据存在以下问题：

- (1) 源句子与目标句子完全不相关；
- (2) 目标句子是对源句子的批注；
- (3) 源句子中存在错误编辑距离较大的情况；
- (4) 数据集中的用词错误不仅包括形似音近和近义字词的错误，还包括反义词等词义混淆类的错误；
- (5) 数据集中末尾处存在多字的缺失错误。

¹https://github.com/Anil999999/MS_Seq2Edit

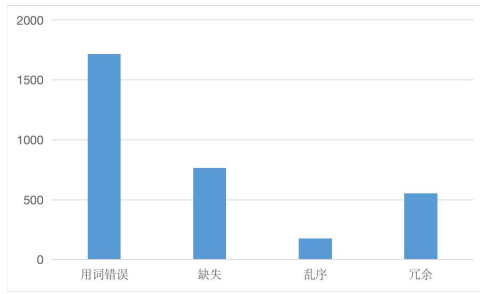


Figure 1: 验证集错误占比统计

对于数据中存在的问题，我们也在表2中进行了展示。由表中的数据可知，不管是采用生成模型还是序列标注模型，其中的一些错误数据均会导致模型产生一些错误的特征映射，从而影响模型的性能。为了提高训练数据的质量，我们利用编辑距离、字数差异和纠正的字数等多维度来衡量句子的质量，从而去除质量较差的数据，以降低错误数据对模型拟合和特征映射的影响。

原始句子	目标句子	问题
拔火罐是以罐为工具，利用燃火、抽气等方法产生负压，使之吸附于体表，以达到行气活血、消肿止痛、祛风散寒、血压、等等。	太极是一种以缓慢深呼吸进行的一项有氧运动，不但对身体好、能减少压力、增强免疫系统、降低血压，也会恢复体力的平衡。	不相关
	针灸是另一种中医的方法，用特别的针刺入皮肤下的穴位点进行治疗的一种方法，它是一个平衡能量或生命力的流动技术，也会止痛。	
	中医是一个中国传统的医学，到了现代还广泛应用，但没有古代时那么普及。	
他以他的诚实为臭名昭著。	他以他的诚实而闻名。臭名昭著是贬义词，不能形容诚实。	增加批注
那些孩子们的衣服很便宜但是很干净。	不需要修改。	批注
比如我父母、兄弟姐妹、朋友、老师等。在这二十四年的人生之中，我遇到的人们像星星一样多，	在这二十四年的人生之中，我遇到的人们像星星一样多，比如我父母、兄弟姐妹、朋友、老师等。	长距离的乱序
从小的孺学灌输使我对于夫妻、兄弟等之间各关系起着很深刻的行为规范。	从小的孺学灌输使我对于处理夫妻、兄弟等之间各关系起着很深刻的行为规范的作用。	句尾多字缺失

Table 2: 数据集中错误数据示例

2.2 相关工作

中文文本的语法纠错任务是当前自然语言处理领域中一个比较重要的下游任务；随着机器学习和深度学习的发展，中文文本纠错也逐渐从基于规则和模板的纠错发展为基于机器学习和深度学习的文本纠错模型并在一些应用场景下取得了不错的效果。

中文文本纠错任务按照错误类型一般分为拼写纠错和语法纠错。其中拼写纠错任务包括形似音近错误，并且该任务主要应用在OCR和ASR等场景。近期拼写纠错的研究集中于深度学习领域，2019年Dingmin Wang和Yi Tay等 [Wang et al., 2019]提出了基于使用混淆集和门控机制相结合的中文纠错模型，该模型在解码的过程中，每一时间步同时利用门控单元的输出和解码生成的向量空间的概率输出来解决当前位置是否需要纠错；并且当模型需要纠错时，也仅仅从当前词的混淆集中选择，而不是从向量空间中选择一词，该方法有效地降低了模型的误纠率。2021年Shaohua Zhang和Hang Li [Zhang et al., 2020]等人提出了利用GRU作为检

错模块来对输入文本的错误位置进行检测，然后融合字向量的特征信息和检错模块的错误信息作为Bert模型的输入特征来进行纠错。之后，Shulin Liu和Tao Yang等 [Liu et al., 2021]提出了PLOME模型，利用GRU将拼音和笔画信息引入到模型中，极大提升了模型的拼写纠错能力。同时，云从科技 [Zhao et al., 2021]利用BART预训练模型应用于语音识别场景下的拼写纠错任务，而且该模型不仅能对拼写错误进行纠错，还能对一些常识错误，语法错误甚至一些推理的错误进行纠正。接着，平安科技 [Guo et al., 2021]提出了基于全局注意力解码的拼写纠错模型，其中利用基于confusion set的替换策略来拉近Bert和CSC两个任务的距离，同时利用全局注意力解码的方法以学习到全局的上下文标注信息以避免噪声带来的影响。2022年Chenxi Zhu和Ziqiang Ying等 [Zhu et al., 2022]提出了针对拼写任务的多任务方法，将检错任务和纠错任务合并在一起，利用检错模型的输出特征表示来辅助模型纠错。

而中文语法纠错任务的实现方法主要集中于基于seq2seq的生成模型和基于GECToR的序列标注模型。2020年，Masahiro Kaneko和Masato Mita等 [Kaneko et al., 2020]提出了将Bert等预训练模型合并到Encoder-Decoder模型中，利用Bert输出的特征信息辅助Encoder-Decoder模型进行语法纠错。与此同时，Omelianchuk等人提出了序列到编辑操作的模型，将文本纠错任务看做序列标注，在保证模型精度的情况下大大缩减了预测的时间。因为基于seq2seq的模型是基于生成的方式根据上下文特征生成符合原始句子含义的正确句子，因为seq2seq模型每次仅能生成一个字符，无法进行并行推理，而基于seq2edit的模型不仅推理速度大幅提升，而且模型的性能也没有下降，因此基于seq2edit的中文纠错方法也出现了许多更进一步的研究方案。2021年，Piji Li和Shuming Shi [Li and Shi, 2021]提出了基于BERT初始化的Transformer编码器作为骨架模型并通过条件随机场层进行非自回归的序列预测模型。之后，Shaopeng Lai和Qingyu Zhou等 [Lai et al., 2022]提出了通过顺序学习数据集中的错误类型来帮助模型更好地文本的中间特征表示，而且先解决某种错误类型可以让模型记忆一些特征，来解决后面的复杂度更高的错误类型。

中文文本纠错任务除了模型结构和性能上的优化外，还有一些通过模型集成 [Tarnavskiy et al., 2022]，重排序 [Wang et al., 2021]等方法来提升模型纠错能力的研究和方法。

3 系统

针对本赛道的纠错任务类型，我们尝试采用不同的模型有针对性地解决数据中的错误。首先，我们利用开源的cbert模型¹和基于多任务的MDCSpell模型来解决句子中的拼写错误；然后利用拼写纠错后的结果进行语法纠错，这里主要利用基于序列到编辑的GECToR模型解决剩余的语法错误。在解码过程中，我们还引入了规则和局部最优路径等策略来获得全局最优的输出结果。

3.1 拼写纠错模型

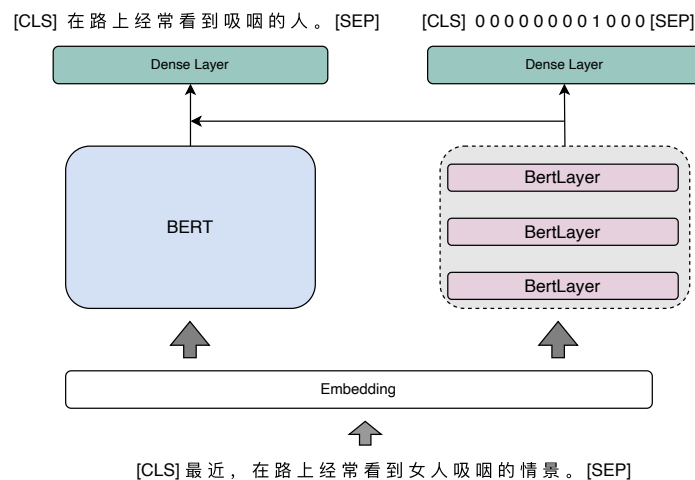


Figure 2: 拼写纠错模型结构图

¹<https://github.com/xueyouluo/speller-bertconfusion>

对于拼写纠错任务，我们首先尝试了开源预训练模型cbert，不过在验证过程中发现cbert模型虽然能够纠正多数的错误，但是精确率和召回率均不够高。因此我们收集了不同来源的混淆集词典，并基于混淆集和微信新闻语料生成了大量针对拼写错误的数据库，该任务中我们参考MDCSpell模型来实现拼写纠错任务；MDCSpell模型检测模块的输入是获得Word Embedding之后去掉CLS和SEP两个位置的特征信息，但是我们在对模型进行复现和验证的过程中认为检测模块可以直接使用Bert输出的词嵌入特征，不需要舍弃CLS和SEP，这样可以保证两个模块的输出维度相同，更方便特征信息的融合。我们的拼写模型如图2所示，其主要的计算公式如下：

$$H^c = \{C_{h1}, C_{h2}, C_{h3}, \dots, C_{hn}\}$$

$$H^d = \{D_{h1}, D_{h2}, D_{h3}, \dots, D_{hn}\}$$

$$H = H^c + H^d$$

$$P(y_i|X) = softmax(Wh_i)$$

通过以上步骤，将检错模块的错误特征信息融合到纠错模块中，提高模型的纠错能力。同时在解码阶段通过设置阈值等方式减少模型的误召回，尽可能保证纠错的高准确性。

3.2 语法纠错模型

除了拼写错误外，数据中还包括其他的一些错误信息，因此我们利用基于序列标注的Seq2Edit模型来解决剩余的语法纠错任务。Seq2Edit是目前最优的语法纠错模型，它通过预训练模型获得输入句子的语义编码特征，然后通过全连接层预测句子中对应的每个句子的编辑标签，该模型对应的解码空间为插入、删除、替换和保持等编辑操作。除了纠错模块外，该模型还引入了检错模块，以便利用多任务的方式来增强模型的纠错效果。该语法纠错的模型结构和解码流程如下图所示。

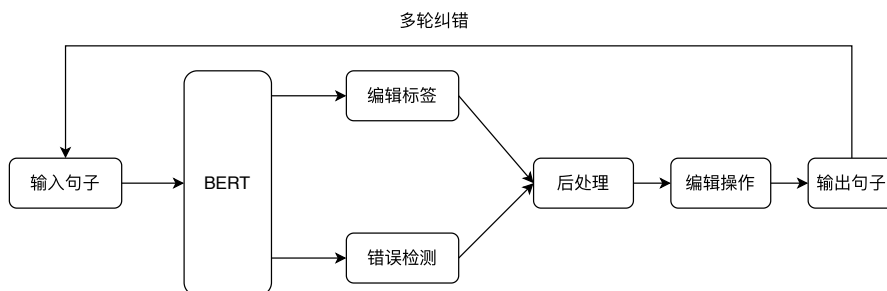


Figure 3: Seq2Edit纠错流程图

在实验过程中，我们发现原始Seq2Edit模型对于乱序错误的纠错能力较差，其在于模型的训练过程和推理过程存在暴露偏差，因此对于连续字符的缺失错误，模型可能多个轮次也无法纠正。因此原始模型将乱序错误看作冗余和缺失两种错误的集合会导致模型对删除操作的置信度较高，同时也因为缺少上下文特征信息不能预测出所有缺失的字符。

输入	输出
小间是个珍贵的东西在我们的生活中	在我们的生活中
我们要去寺庙拜一拜去外婆家以后。	我们要去寺庙拜一拜。

Table 3: Seq2Edit模型的乱序输出示例表

由上表的数据可知，输入句子是乱序错误，但在推理过程中模型对一侧短语片段进行删除操作之后，由于缺少了大量的上下文信息从而不能对另一侧的插入操作进行补齐。因此，我们引入移动(move)操作来解决模型中的乱序问题。通过对数据集乱序错误编辑距离统计，我们保留了最多12个位置的移动距离，最终我们的纠错标签体系如表4所示。

标签	描述	备注
KEEP	保持	保持当前字符不变
DELETE	删除	删除当前位置的字符
APPEND	插入	在当前位置的后面插入一个字符
REPLACE	替换	将当前字符替换为其他字符
MOVE	移动	将当前字符移动到其他位置，只有存在局部路径和为0的情况下才接受该操作

Table 4: Seq2Edit模型的标签体系

模型	输入句子	目标句子	输出结果	第一轮编辑操作
Seq2Edit	他的一生经历的确不凡。	他一生的经历的确不凡。	他的一生经历的确不凡。	1-2,DELETE 3,3,APPEND_的
Seq2Edit	但是当你结了婚，有了孩子以后，你就要对其他人的生命负责。	但是当你结了婚，有了孩子以后，你就要对其他人的生命负责。	但是当你结了婚，有了孩子以后，你就要对其他人的生命负责。	18,19,DELETE 19,20,DELETE 27,27,APPEND_负 (第二轮delete ₀)
Seq2Edit_Move	他的一生经历的确不凡。	他一生的经历的确不凡。	他一生的经历的确不凡。	1,2,MOVE_2 2,3,MOVE_-1 3,4,MOVE_-1
Seq2Edit_Move	但是当你结了婚，有了孩子以后，你就要对其他人的生命负责。	但是当你结了婚，有了孩子以后，你就要对其他人的生命负责。	但是当你结了婚，有了孩子以后，你就要对其他人的生命负责。	18-20; MOVE_6(2) 20-26;MOVE_-2(6)

Table 5: 模型引入编辑操作前后对比表

模型引入移动操作的输出对比结果如表5所示，由表中的数据可知，Seq2Edit对于短距离的乱序错误能够正确地调整字符的顺序，但是错误片段较长且错误位于句尾，因此该模块并没有输出正确的结果。但是将Move操作加入标签体系之后，模型仅通过一个轮次就将错误纠正。所以移动操作能够有效地解决长距离的乱序错误，并降低模型对删除操作的倾向性。

3.3 局部路径解码

在解码的过程中，我们发现在乱序移动操作的局部路径中，每个字符对应的最大概率并不能很好地解决乱序问题。但是对模型的输出概率进行分析，发现在局部区域内能够找到一条偏移距离和为0且概率相对最大的路径，能够解决乱序的问题。例如表6所示的部分数据中可以找到一条[3, 3, -2, -2, -2]路径，使得局部路径和为0且能正确解决乱序问题。因此在解码的过程中，对于每一轮次输出的编辑操作，我们都进行一个后处理用来解决移动操作等问题。

字符	局部概率
两	[MOVE_2, MOVE_-1, MOVE_4, MOVE_3, MOVE_-2]
周	[MOVE_2, MOVE_3, MOVE_4, MOVE_5, MOVE_-1]
短	[MOVE_-2, KEEP, MOVE_1, MOVE_-3, DELETE]
短	[MOVE_-2, MOVE_-3, KEEP, MOVE_-1, MOVE_1]
的	[MOVE_-2, KEEP, MOVE_-3, MOVE_-1, MOVE_-5]

Table 6: 局部概率输出结果

3.4 模型集成

不同的模型能够学习到不同的语义特征等信息，因此将多个差异较大而性能略差的模型正

确的组合能够极大地提升模型的性能。在实验过程中，我们发现按照一定的权重组合策略将多个模型的输出结果进行加权平均后作为最终的概率值，可以得到最优的结果。因此我们对语法纠错模型进行了模型的集成。但是在对多个模型的集成结果进行评测的过程中，我们发现有些相对正确的模型纠错输出在集成的过程中反而被加权平均了，因此为了能够取得组合之后的最优结果，我们还利用困惑度来对不同的纠错结果进行句子流畅度评估，选择困惑度最低的纠错句子为最优解。

3.5 数据增强和其他后处理策略

在实验分析中，我们发现模型对多字词的缺失和句子不同位置的错误的纠错能力不同，并且当前的数据集未完全覆盖绝大多数的错误，存在OOV的问题。因此我们尝试通过不同的策略生成数据，用来解决因OOV导致的错误无法纠正的问题。本次比赛中，拼写纠错和语法纠错两个任务都用到了数据增强技术，且均使用微信公众号语料¹作为种子数据。对于拼写纠错任务，我们基于字和词两个维度进行数据扩充，并维护了生僻词表、形近混淆集和音近混淆集以保证生成的均为常用的字词。

针对拼写纠错任务，数据增强的过程如下所述。

- (1) 对句子进行预处理并确定能够进行替换操作的字符位置；
- (2) 进行第一次采样确定句子进行替换操作的字符个数；
- (3) 迭代替换，每次确定替换的位置，替换类型和替换的字符。

而针对语法纠错任务，其过程与拼写任务的生成过程类似，主要是替换操作包括四大类；但是不同错误的生成比例不同，语法纠错的比例是W:M:R:S5分别为5%:10%:15%:70%，拼写纠错生成数据的错误比例均为40%的形似和音近错误，20%的词错误除了数据之外，我们还采用了成语纠错解决数据中出现的成语俗语，并且通过不同的策略来干预模型的解码过程，例如对常用词，连词的修改操作的接受概率值相对较高，插入操作的接受修改的概率值对其他操作来说相对较低，同时利用困惑度、规则以及召回度和精确度之间的平衡策略来提高模型的性能。

4 实验

4.1 实验描述

1. 拼写纠错模型

拼写纠错模型的训练数据使用的是微信语料生成的200万条数据和赛道二通过混淆集清洗后的数据。在训练过程中，该模型使用的预训练模型为Bert，检错模块为3层的transformer blocks，lr为2e-5，epoch为10。

2. 语法纠错模型

语法纠错模型的核心代码参考MuCGEC²，该模块的预训练模型为Roberta和StructBert，训练数据是按3.4增强后的微信数据以及按照2.1的数据处理方法对数据清洗之后的数据，同时为了解决模型FPR较高的问题，我们保留了数据中的正确数据，以期模型通过学习正确句子的字词顺序特征降低模型的误召回。

4.2 实验结果

在本次比赛过程中，我们首先对比了基于Seq2Seq的模型和基于Seq2Edit模型的基准模型效果，然后选择了Seq2Edit模型作为本次比赛的基本框架。由该赛道的评测指标可知，本次比赛不仅考察模型的纠错能力，还考察模型正确区分句子对错的能力；因此我们训练了不同性能的多个模型并通过模型集成和困惑度来选择最优结果；不同模型的模型效果对比如下表³所示。由表中数据可知，基于Seq2Seq的模型的检错能力较好，但同时模型引入了较高的误召回，从而使得假阳性的指标偏高；而基于Seq2Edit的方法更能够权衡精确率和召回率，使得模型在评测数据上取得更好的结果。同时通过对比可知，利用拼写纠错模型预先纠正用词错误，然后再对其他错误进行纠错，能够提升模型的效果。

¹https://github.com/nonamestreet/weixin_public_corpus

²<https://github.com/HillZhang1999/MuCGEC>

³DA:数据增强;PPL:困惑度;Ens:模型集成

模型	COM	FPR	DET	IDE	POS	COR
Seq2Seq	30.37	76.4	84.6	55.45	32.44	25.39
Seq2Edit	38.82	21.82	82.23	51.13	27.11	16.66
Seq2Edit+CSC	41.95	23.3	83.85	53.97	31.65	21.64
Seq2Edit+CSC+Move	43.33	24.77	81.4	55.91	37.35	27.41
Seq2Edit+CSC+Move+DA	44.41	22.27	78.96	55.56	37.55	27.83
Seq2Edit+CSC+Move+DA+PPL+Ens	46.58	19.76	81.25	56.68	28.98	29.19

Table 7: 不同模型效果对比表

4.3 错误分析

在本次比赛中，我们利用不同模型的优势来解决数据中存在的不同错误。虽然通过拼写纠错、移动操作和局部路径解码等方法解决了大多数的错误问题。但是，本系统仍然存在一些问题。首先本系统虽然加入了移动编辑操作来解决乱序的问题，但是只保留的出现频率较高的12个字符之内的移动操作，长距离的乱序仍然无法解决。本系统的一些样例输出结果如表8所示，我们还看出对于一些连词的冗余和缺失的纠正，虽然使得句子更顺畅，但是与标注数据不同，反而引入了错误。

错误类型	输入句子	目标句子	模型输出
S	但是我妈妈扔然勿视流行音乐。	但是我妈妈仍然无视流行音乐。	但是我妈妈仍然无视流行音乐。
W	比如我父母、兄弟姐妹、朋友、老师等。在这二十四年的人生之中，我遇到的人们像星星一样多，	比如我父母、兄弟姐妹、朋友、老师等。在这二十四年的人生之中，我遇到的人们像星星一样多，	比如我父母、姐、老师等。在这二十四年的人生之中，我遇到的人们像星星一样多，
M&R	其实面对挫折愈多，人的解决困难的能力也相应提高。	其实面对的挫折愈多，人的解决困难的能力也会相应提高。	其实面对的挫折愈多，人解决困难的能力也会相应提高。

Table 8: 模型样式输出结果

5 总结

本次比赛过程中，我们针对汉语学习者的写作数据设计了不同的方法以解决尽可能多的错误，在此过程中我们利用拼写模型、基于序列标注的seq2edit模型和模型集成以及一些额外的策略来对输入句子中的错误数据进行纠正并在修正层等指标下取得了较好的效果。但是我们的工作仍然存在一些问题，一方面是当前的模型复杂度较高，另一方面是系统的假阳性指标仍然较高。未来，我们的工作主要从降低模型复杂度、通过流畅度等方法降低FPR指标的同时保证模型的纠错能力等方向来进一步提升中文语法纠错的能力。

参考文献

- Guo, Z., Ni, Y., Wang, K., Zhu, W., and Xie, G. (2021). Global attention decoder for chinese spelling error correction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1419–1428.
- Kaneko, M., Mita, M., Kiyono, S., Suzuki, J., and Inui, K. (2020). Encoder-decoder models can benefit from pre-trained masked language models in grammatical error correction. *arXiv preprint arXiv:2005.00987*.
- Lai, S., Zhou, Q., Zeng, J., Li, Z., Li, C., Cao, Y., and Su, J. (2022). Type-driven multi-turn corrections for grammatical error correction. *arXiv preprint arXiv:2203.09136*.

- Li, P. and Shi, S. (2021). Tail-to-tail non-autoregressive sequence prediction for chinese grammatical error correction. *arXiv preprint arXiv:2106.01609*.
- Liu, S., Yang, T., Yue, T., Zhang, F., and Wang, D. (2021). Plome: Pre-training with misspelled knowledge for chinese spelling 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 1: Long Papers)*, pages 2991–3000.
- Tarnavskiy, M., Chernodub, A., and Omelianchuk, K. (2022). Ensembling and knowledge distilling of large sequence taggers for grammatical error correction. *arXiv preprint arXiv:2203.13064*.
- Wang, B., Che, W., Wu, D., Wang, S., Hu, G., and Liu, T. (2021). Dynamic connected networks for chinese spelling check. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2437–2446.
- Wang, D., Tay, Y., and Zhong, L. (2019). Confusionset-guided pointer networks for chinese spelling check. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5780–5785.
- Wang, Y., Kong, C., Liu, X., Fang, X., Zhang, Y., Liang, N., Zhou, T., Liao, T., Yang, L., Li, Z., Rao, G., Liu, Z., Li, C., Yang, E., Zhang, M., and Sun, M. (2022). Overview of cltc 2022 shared task : Chinese learner text correction.
- Zhang, S., Huang, H., Liu, J., and Li, H. (2020). Spelling error correction with soft-masked bert. *arXiv preprint arXiv:2005.07421*.
- Zhao, Y., Yang, X., Wang, J., Gao, Y., Yan, C., and Zhou, Y. (2021). Bart based semantic correction for mandarin automatic speech recognition system. *arXiv preprint arXiv:2104.05507*.
- Zhu, C., Ying, Z., Zhang, B., and Mao, F. (2022). Mdcspell: A multi-task detector-corrector framework for chinese spelling correction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1244–1253.