中文错误语法校对综述

林语盈[1)]

[1)](计算机科学与技术专业 南开大学计算机学院, 天津市中国 300071)

摘 要 本文对中文错误语法校对近年来的研究成果进行汇总和概述。首先进行问题定义。接下来讨论了中文 GEC 任务的主要数据集以及数据增广方法、模型算法分类与介绍、评价指标。因任务的相关性，也提及适用于其他语言和多语言的算法模型。最后提出该领域未来的发展方向。

关键词 语法错误校对；语言生成；序列标注；机器翻译；自然语言处理

中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.01.2022.00001

A Survey of Chinese Grammatical Error Correction

Yuying Lin[1)]

[1)](Department of Computer science and technology,Nankai University, Tianjin 300071, China)

Abstract This paper summarizes the current baselines and research results of Chinese Grammatical Error Correction in recent years. First, We give the defination the problem. Next, We introduce the datasets of Chinese GEC tasks as well as data augmentation methods, classification of models, and evaluation indicators. Due to the relevance of tasks, multilingual models are also mentioned. Finally, we discuss prospective directions for future GEC researches.

Key words grammatical error correction; language generation; sequence labeling; machine translation; natural language processing

1 引言

Grammatical Error Correction(GEC), 错误语法校对，有巨大的需求和发展潜力，包括汉语学习者汉语作文自动订正、汉语母语者文章和公文自动校对、以及 NLP 文本预处理的需求等。近年来，随着数据和算力的发展，GEC 成为 NLP 一个受到关注的问题。

本文主要介绍中文错误语法校对，同时因任务的相关性，也提及适用于其他语言和多语言的算法模型。对 GEC 任务的主要数据集以及数据增广方法、模型算法分类与介绍、评价指标作简要概述。文章的第二章为问题定义，第三章为主要方法分类，分为模型、数据、评价指标三个部分，第四章为当前仍存在的挑战与未来展望。

收稿日期：2022-04-19；修改日期：2022-04-19 林语盈（**通信作者**），女，2000 年生，南开大学计算机科学与技术专业 2020 级学生，主要研究领域为中文错误语法校对 E-mail: 2012174@mail.nankai.edu.cn.
第 1 作者手机号码：13612191673，E-mail: 2012174@mail.nankai.edu.cn

2 问题定义

句子错误类型可能有很多，按照 CTC2021 比赛的分类，如图1所示，按难度依次递进，可分为拼写错误、语法错误、语病错误。这里给出了四种语法错误的例子，如图2所示，包括遗漏单词（M）、冗余单词（R）、单词选择错误（S）和词序错误（W）。

错误大类	错误小类
拼写错误	别字
	别词
语法错误	缺失
	冗余
	乱序
语病错误	语义重复
	句式杂糅

图 1 中文语法错误分类（CTC2021）

Error Type	Original Sentence	Correct Sentence
M	每个城市的超市能看到这些食品。	每个城市的超市**都**能看到这些食品。
R	我和妈妈**是**不像别的母女。	我和妈妈不像别的母女。
S	最重要的是**做**孩子想学的环境。	最重要的是**创造**孩子想学的环境。
W	“静音环境”**是对人体应该**有危害的。	“静音环境”**应该是对人体**有危害的。

图 2 四种语法错误例，遗漏单词（M）、冗余单词（R）、单词选择错误（S）和词序错误（W）

问题的输入输出可简要概括如下:

- 输入：中文句子;
- 输出：句子是否正确，若有错误，输出错误类型、位置、改正方案。

3 主要方法分类

目前，人们对 GEC 问题的研究主要包括三个方面，即模型算法、数据与评价指标。

3.1 模型算法

当前主要的模型可分为基于统计机器学习的方法、基于端对端神经网络机器翻译的方法、基于序列标注的方法、基于语言模型的方法与混合方法。其中，基于神经网络机器翻译的方法与基于序列标注的方法获得了更多的关注，也取得了更好的效果，这也是本文的重点。

3.1.1 基于统计机器学习的方法

统计机器学习 在使用统计机器学习方法之前，语法纠正仅仅使用简单的基于规则或者分类的方法，这类方法需要专家去手工设计既定的规则，并且需要大量的语言学信息，在实际应用中往往会受到专家知识和预料资源的限制，尤其是对于那些仅被少数人使用的语言。于此同时，几乎不存在针对所有语言都适用的规则，上下文和语言的不同都会限制这类方法等效果。在这种情况下，统计机器学习的方法逐渐受到研究者的重视。

在语法纠正领域，统计机器学习方法主要使用的是统计机器翻译（Statistical Machine Translation, SMT)。对于源语言给定的句子 x，它的目标是寻找目标语言对应的翻译 y，也就是寻找 y 使得 $p(y|x)$ 的概率最大。概率分布 $p(y|x)$ 就是统计机器翻译模型需要估计的。常见的统计机器翻译模型有基于噪音通道（Noisy Channel）和基于对数线性模型（Log-Linear Model）。基于噪音通道的翻译模型通过贝叶斯公式，将求解 $\underset{y}{argmax}, p(y|x)$ 转化为求解 $\underset{y}{argmax}, p(x|y)p(y)$，引入了一个反转翻译模型 $p(x|y)$ 和一个语言模型 $p(y)$，同时将这两个模型有效地结合起来。并且，训练语言模型本身不需要平行语料，有丰富的单语语料资源。基于对数线性模型（Log-Linear Model）的翻译模型基于最大熵原理，通过定义特征函数，可以有效地捕捉源句子中的先验知识和语义信息。

模型与方法的发展与现状 在使用统计机器翻译进行语法纠错的发展过程中,[1] 使用在中国学习者错误语料库 (Chinese Learner Error Corpus) 中发现的大量名词错误示例来指导工程训练集的创建,表明统计机器翻译范式的应用可以捕获为母语人士设计的广泛使用的校对工具无法很好解决的错误。[2] 研究了针对语法错误纠正任务的标准指标 M2 指标的参数调整，发现具有特定任务参数调整的基于短语的基本 SMT 设置会有明显的效果提升。

神经网络联合模型 (neural network joint model) 是 SMT 中经常使用的模型,[3] 使用特定于某个语言的学习器文本调整神经网络联合模型，并将其集成到基于统计机器翻译的语法错误纠正系统中，在中文、俄语和西班牙语作家撰写的英语文本的语法纠正上取得了不错的结果[4] 实现了基于短语统计机器翻译的无监督语法纠错（GEC)。[5] 指出基于短语的统计机器翻译系统受到离散词表示、线性映射和缺乏全局上下文的限制。因此通过使用神经网络全局词典模型和神经网络联合模型使用单词的连续空间表示以及学习非线性映射来更好地泛化。

结果重排序是对 SMT 方法的一个有效改进,[6] 针对 SMT 系统的 1-best 结果欠佳的问题，提出了一种语法错误纠正的重新排序方法用于对 SMT 的 N-best 重新评分并对结果重新排序。类似的,[7] 利用全局上下文信息构建了一个分类器，可用于选择适当的编辑或重新排列 N-best。

3.1.2 基于神经机器翻译的方法

神经机器翻译 因问题的相似性，基于端对端的深度网络的神经机器翻译的模型被广泛用于 GEC 任务。与统计机器学习的方法不同，此类方法无需构建句子的特征，而是直接采用正确句-错误句的平行数据对作为模型的输入，采用编码器-解码器的结构。具体来说，编码器或解码器的选择包括 CNN、RNN 或注意力机制。大规模的预训练模型也被广泛尝试。编码器首先将源语句 $x = \left\{x_1, x_2, \cdots, x_{T_x}\right\}$ 编码为向量 v。然后将该向量被传递给解码器，用于生成改正的结果 $y, p(y) = \prod_{t=1}^{T} p\left(y_t \mid y_{1,\cdots,t-1}, v\right)$ 在每一个时间步，目标词都会用向量和之前生成的词进行预测。

模型与方法的发展与现状 基于 **RNN** 的模型。Yuan 和 Briscoe[8] 于 2016 年首次在 GEC 中应用基于 RNN 的 NMT 模型。然后，Xie 等人[9] 为 GEC 应用了第一个字符级神经编解码模型，以更好地解

决 OOV（词汇表外）问题。接下来，结合单词级模型和字符级模型，[10] 提出了一种基于 NMT 的嵌套注意混合 GEC 模型。该模型的设计更多地考虑了 GEC 的特定特性。2017 年，Sakaguchi 等人创新性地将强化学习应用于 GEC，以更好地纳入任务特定指标，并克服传统 MLE 导致的暴露偏差。2018 年，[11] 提出了另一项使用基于 RNN 的 NMT 模型的新工作，该模型旨在提高源句子的流利度和可靠性，以纠正语法错误，称为流利度提升学习。这项工作为使用基于 NMT 的方法解决 GEC 提供了一个新的视角。

基于 **CNN** 的模型。[12] 在 2018 年成为第一个基于 CNN 的 NMT 的 GEC 系统。在此之后，2019 年 [13] 提出了另一项工作，通过卷积编码器-解码器模型将跨句上下文集成到 GEC 中。考虑到语境，该模型更能纠正某些语境相关的错误，如动词时态错误和冠词错误。与基于 RNN 的模型相比，基于 CNN 的模型的优势在于，CNN 在捕捉本地上下文方面更有效，从而纠正更广泛的语法错误。多层体系结构可以捕获长期的上下文信息。此外，无论输入长度如何，只对输入执行恒定数量的非线性计算，而在 RNN 中，非线性的数量为 O（n），从而减少了远距离单词的影响。

基于 **Transformer** 的模型。随着 Transformer[14] 的提出，许多基于 NMT 的 GEC 模型用 Transformer 取代了传统的基于 RNN 的编码器-解码器，并取得了有希望的结果。2019 年，[15] 提出基于 Transformer 模型的 NMT 系统，并通过整合几个增强功能对其进行了改进：将丢弃层应用于整个源词和目标词，加权目标子词，平均模型检查点，并反复使用训练过的模型来纠正中间翻译，在 BEA 2019 比赛中取得了 64.55 的 F0.5 score。[16] 基于之前使用有限状态传感器和强神经语言模型的工作，提出了一个用于少量数据的方法。2019 年晚些时候，复制机制 [17] 被创新地应用于 GEC 任务，允许模型将未更改的源单词直接复制到目标端，并实现了最先进的性能。与 RNN 和 CNN 相比，Transformer 通过注意在句子中建立远程依赖的能力更强，并且它能够实现更高效的并行计算。此外，[18] 提出了浅攻击解码 (SAD)，以提高 transformer 在线即时语法纠错 (GEC) 的计算性能。

大规模**预训练**模型。随着数据和算力的不断丰富，大规模预训练模型被用于 GEC 任务。2019 年，[19] 首先将大规模的预训练模型与 transformer 用于 GEC 任务。2020 年，[20] 将预训练的模型与编解码模型结合并运用到汉语语法纠错任务中。[21] 将预先训练好的掩码语言模型（MLM），整合到用于语法错误纠正（GEC）的编码器-解码器（EncDec）模型中。首先使用给定的 GEC 语料库对模型参数进行微调，然后将微调后的模型参数输出作为 GEC 模型中的附加特征。该模型在 BEA-2019 和 CoNLL-2014 中取得了当时最优的效果。[22] 研究了预训练的语言编码器（ELMo、BERT 和 RoBERTa）在遇到自然语法错误时的处理，收集非母语人士的真实语法错误，并在干净的文本数据上进行对抗性攻击来模拟这些错误，并使用这种方法来促进下游任务并调试模型。

2021 年左右，一些基于 NMT 的方法被扩展到**多语言环境**。使用元学习，以更好地处理除英语以外的一些语言的数据稀缺问题。多语言双向自回归 transformer（BART）[23] 在机器翻译任务中表现出了有效性，在其他语言上，包括德语、捷克语和俄语的 GEC 任务中也受到了关注。[24] 使用 BART 作为 GEC 通用预训练编码器-解码器模型，消除了耗时的预训练，单语言和多语言的 BART 模型在 GEC 中取得了很高的性能。[25] 研究了 GEC 的跨语言迁移学习方法，来自其他语言的迁移学习可以提高 GEC 的准确性，并且，接近源语言对纠正某些类型错误的准确性有重大影响。[26] 提出了一种语言无关的方法来生成大量数据，并使用大规模的多语言模型 (最多 11B 参数)，对特定语言数据集进行微调，在四种语言 (英语、捷克语、德语和俄语) 上的 GEC 基准测试结果超过了以前的最先进水平。针对小语种少量的数据，[27] 研究了如何在低数据场景中最好地利用 GEC 可用的数据源。[28] 提出了适用于语法错误纠正（GEC）任务的好的多语言模型，应用现有的预先训练的多语言模型来纠正多种语言中的语法错误，找到了一个模型，可以纠正七种不同的语言，是目前已知的最好的多语言 GEC 模型。

生成性对抗网络（Generative adj-arial Networks，简称 GANs）通过学习直接最小化人工生成文本和合成文本之间的差异，成功地生成了跨越许多不同任务的真实文本。[29] 使用生成器-鉴别器框架提出了一种 GEC 的对抗式学习方法。生成器是一个 transformer 模型，经过训练，可以在语法错误的句子中生成语法正确的句子。鉴别器是一个句子对分类模型，经过训练可以根据语法纠正的质量来判断给定的一对语法错误的正确句子。[30] 提出通过不断

识别模型的弱点来生成更有意义和价值的训练示例，并逐步将生成的对抗性示例添加到训练集中来增强模型，这样提高了 GEC 模型的泛化能力和鲁棒性。

最近，研究方向转向一些针对 GEC 任务的特定优化方法和其他的需求场景。[31] 提出了一种束搜索方法，以在局部序列转换任务中获得不同的输出, 不重写文本中的所有标记，而只重写需要进行不同更正的部分, 根据预测从源语句复制的概率调整 beam 中的搜索标记, 在不损失 GEC 任务准确性的情况下，可以产生更多不同的修正。[32] 针对文档级语法错误纠正任务进行改进，并提出了基于上下文感知方法的文档级 GEC 系统，采用了三步训练策略，从句子级和文档级数据中获取信息。在文档级别通用测试集上实现了最先进的性能。[33] 提出了用于多假设条件下 GEC 质量估计的神经网络 VERNet (Neural Verification Network)。VERNet 用推理图建立假设之间的交互，并通过两种注意机制传播 GEC 证据来验证生成的假设的质量。[34] 研究全词掩蔽对中文语境理解的影响，引入了两个与语法错误纠正相关的探测任务，并将预训练的模型以掩蔽语言建模的方式修改或插入标记，分别训练了三个具有标准字符级掩蔽（CLM）、WWM 和 CLM 与 WWM 组合的中文 BERT 模型。[35] 提出了一种结合语法分析树进行语法错误纠正（GEC）的方法, 为 GEC 提供了一个统一的 seq2seq 解析集成方法，该方法使用依赖树和选区解析树，以及语法图对原始模型进行优化。[36] 提出了一种用于 GEC 的类型驱动多轮校正方法。从每个训练实例中额外构造多个训练实例，每个训练实例都涉及到特定类型错误的纠正。然后使用这些额外构造的训练实例和原始实例依次训练模型。

3.1.3 基于序列标注的方法

序列标注 基于 NMT 的方法以自回归模式捕获输出令牌之间的依赖关系，并获得最佳结果。然而，顺序解码在推断阶段执行缓慢，时间复杂度为 O（n），其中 n 表示目标句子的长度。研究人员重新考虑了 GEC 任务的特点，即输入和输出文本之间的高度重叠。一般来说，源句子和目标句子之间的转换可以通过一些编辑操作来实现。因此，研究人员提出了基于序列标注的方法。基于 NMT 的方法相比，只需少量编辑，该方法的推理速度显著提高。

基于序列标注的方法放弃了直接预测目标句子 y 的想法，并基于源句子 x 预测一系列编辑操作 $e = e1e2...eT x$，这样在编辑序列 e 变换后，x 可以被转换为目标句子 y。一般来说，在训练阶段，该方法首先根据 x 和 y 获得 e 序列。然后，通过使用 x 作为模型输入，e 作为模型输出，训练模型学习预测编辑。在推理阶段，模型首先根据输入 x 预测编辑序列 e，然后将编辑 e 应用于 x 以获得最终目标序列 y。

模型与方法的发展与现状 研究人员提出了几种获得编辑操作的方法。PIE[37] 定义了四种编辑操作：复制（C）、删除（D）、用 n-gram 追加（A）、用 n-gram 替换（R）以及各种词汇转换。编辑序列是通过计算源语句和目标语句之间的 Levenshtein 距离得到的。LaserTagger[38] 中的编辑操作包括两部分，一个基本操作和一个附加短语 P 操作。基本操作是 KEEP 或 DELETE，它指示输出是否保留该单词。附加短语 P 操作强制 P 添加到相应单词之前，其中 P 属于短语词汇 V，这是通过组合优化获得的。他们使用贪婪算法得到最终的编辑序列。GECToR[39] 通过手动定义编辑操作进行编辑，编辑操作包括常见语法错误，如拼写、名词数量、主谓一致性和动词形式。它的标记词汇表包括 4971 个基本转换（保留、删除、1167 个附加操作和 3802 个替换操作）和 29 个 g 转换。编辑序列也是通过最小化 Levenshtein 距离得到的, 最好的单模型/整体 GEC 标记器在 CONLL-2014（测试）上达到 65.3/66.5 的 $F_{0.5}$，在 BEA-2019（测试）上达到 72.4/73.6 的 $F_{0.5}$。在之前工作的基础上，Seq2Edit[40] 提出了基于跨度的编辑，而不是令牌级别的编辑，这使得表示更容易学习。基于编辑的方法对改进模型几乎没有作用，通常使用 transformer 模型体系结构。在 BERT 的基础上，PIE 和 GECToR 分别使用全连接层和 softmax 层来获得标签预测结果。LaserTagger 在 BERT 模型中添加了浅层解码器，并探讨了前馈解码器和自回归 ACM 事务对智能系统和技术的各自影响。[41] 提出了尾对尾（TtT）非自回归序列预测。采用一个 bert 初始化的转换器编码器作为主干模型来进行信息建模和传输，将条件随机场（CRF）层堆叠在上尾端，通过对令牌相关性建模来进行非自回归序列预测，并将焦点损失惩罚策略集成到损失函数中以解决模型标签分类的失衡问题。[42] 研究了对 GEC 序列标记体系结构的改进，在大型模型中对最近最先进的基于 transformer 的编码器进行集成，通过跨级别编辑的分数来整合模型，并可以适应不同的模

型架构和词汇表大小。针对中文语料，在 CTC2021 的 CGEC 比赛中，采用序列标注的模型 S&A 取得了目前最好的效果。

3.1.4 基于语言模型的方法

此外，还有基于语言模型（LM）方法,[43] 利用预训练语言模型（LM）来定义语言评判器, 根据现实的非语法/语法对，以进一步使用 Break-It-Fix-It（BIFI）框架在学习在无标记数据的情况下进行错误语法检测。如果 LM 赋予句子比其局部扰动更高的概率，该模型会判断句子是否符合语法。

3.1.5 基于混合模型的方法

[44] 提出了一种自动组合系统的黑箱方法，自动检测每个错误类型的系统分数与多个系统的组合, 在直接优化 F 分数的同时提高精确度和召回率, 并分析与改进了 BERT 与拼写检查器等方法。[45] 提出了一种基于混合模型的中文 GEC 方法，它由一个基于 NMT 的模型、一个序列编辑模型和一个拼写检查器组成，在中文 GEC 任务中实现了较好的表现，而且不依赖于数据扩充或 GEC 特定的架构更改。[46] 将任务分为两个子任务：错误跨度检测（ESD）和错误跨度校正（ESC），来提高语法错误纠正（GEC）的效率。ESD 使用序列标记模型识别语法错误的文本范围，ESC 利用 seq2seq 模型将带有注释错误的句子作为输入，输出更正文本。

3.2 数据与数据增广

目前，中文 GEC 任务的主要数据集与其规模细节如表1所列。

语法纠错 (GEC) 缺乏足够的并行数据。关于 GEC 的研究提出了几种生成伪数据的方法，伪数据包括语法对和人工生成的非语法句子。目前，生成伪数据的主流方法包括反向翻译 (BT)、基于百科编辑记录的方法、基于噪声与规则的生成方法、基于机器翻译不同句子对的方法等。根据模型的结构不同, GEC 对生成的数据的特点也存在不同倾向。

2019 年,[51] 提出了 Singsound GEC system 错误数据生成方法，使用了基于维基百科编辑记录的数据，并引入更复杂的错误生成组件，提出了单词树的方法。2020 年,[52] 设计了一种自调整方法，通过利用现有模型的预测一致性对这些数据集进行去噪，优于强去噪的基线方法。[53] 提出将噪声应用于句子的潜在表示。通过编辑语法句子的潜在表征，生成具有各种错误类型的合成样本, 提高语法纠错模型的性能和鲁棒性。[54] 提出了基于机器翻译好-坏句子对的数据合成方法，基于一对不同质量（即差和好）的机器翻译模型（例如，中文到英文)。[55] 对 GEC 任务中两种生成合成平行数据对的方法进行比较分析，即使用基于反向拼写检查的混淆集和使用模式与位置的混淆集，并发现使用模式与位置的方法比基于反向拼写检查的方法表现出更强的使用价值。[56] 从单语数据中提取与学习者句子相似的句子并且通过考虑学习者经常犯的错误类型来生成真实的伪错误。2021 年,[57] 比较了三种不同架构的反向翻译 (BT) 模型 (Transformer、CNN 和 LSTM) 产生的伪数据训练 GEC 模型的校正趋势, 研究了使用不同 BT 模型产生的伪数据组合时的修正趋势, 并发现其提高或插值了每种错误类型的性能。[58] 使用来自自动标注工具 (如 ERRANT) 的错误类型标记来指导合成数据的生成, 带有与给定开发集匹配的错误标记频率分布。[27] 研究了在没有大量高质量训练数据的情况下，语言语法纠错的数据策略。证明了生成 GEC 人工训练数据的方法可以受益于包含形态错误, 以及从维基百科的修订历史和语言学习网站 Lang8 收集的嘈杂的纠错数据是有价值的数据来源，证明了在有噪声的数据源上预先训练的 GEC 系统可以使用少量高质量的、人类标注的数据进行有效的微调。2022 年,[59] 提出了两种数据合成方法，可以控制合成数据的错误率和错误类型比率。第一种方法是以固定的概率损坏单语语料库中的每个单词，包括替换、插入和删除。另一种方法是训练错误生成模型，并进一步过滤模型的解码结果。

3.3 评价指标

由于语法错误纠正任务的主观性和复杂性，如何对其结果进行合理的评估一直是热门的问题。通常来说, 常用的评价方法都需要有参考的正确答案, 通过将假设与参考结果进行比较来计算。由于传统的精度、召回率和 F 分数可能会产生误导，因此在语法错误纠正这一领域有一些特定的评估指标，包括 M^2[60]、LEU[61] 和 ERRANT[62]。当有多个参考答案可用时，M^2 和 ERRANT 选择最大化度量分数的一个，而 GLEU 随机选择一个参考并计算 500 个分数的平均值作为最终分数。此外,[63] 提出了基于范例的 GEC（EB-GEC），它向语言学习者展示范例，作为纠正结果的基础，提高了 GEC 模型的可解释性。

表 1 现有 CGEC 数据集

适用任务	数据集名称	数据来源	数据形式	错误句子数量	正确句子数量	出现字数	句子平均长度
GEC	NLPCC2018[1]	Lang-8.com 中真实外国汉语学习者的错句与改正句	1 错误句，0-6 改正句	717241	1097703	606982	18.73
	CGED2020[2]	真实外语学习者的错句与改正句	1 错误句，1 改正句	1129	1129	1683	42.7
	CTC2021[3]	由网络中爬取并生成的错句	1 错误句，1 改正句	217634	217634	224307	52.94
	YACLC[47]	汉语学习者错句与专家的多维标注数据	1 错误句，多个改偏标注	9000	约 90000		
CEC	Sighan13[48]	中文含错别字的句子与改正句	1 错误句，1 改正句	1700	1700		
	Sighan14[49]	中文含错别字的句子与改正句	1 错误句，1 改正句	4499	4499		
	Sighan15[50]	中文含错别字的句子与改正句	1 错误句，1 改正句	3439	3439		

3.4 M^2

MaxMatch（M^2）是当前错误语法纠正中最常用的评价指标。M^2 依赖于正确答案中的错误编码注释，它提取系统假设的短语级编辑与正确编辑集合的最大重叠，根据 F_β 指标对编辑进行评估。假设句子 i 的正确编辑集合为 g_i，系统假设的编辑集合是 e_i，召回率 R，准确率 P，以及 F_β 的定义如下：

$$R = \frac{\sum_{i=1}^{n} |g_i \cap e_i|}{\sum_{i=1}^{n} |g_i|}$$

$$P = \frac{\sum_{i=1}^{n} |g_i \cap e_i|}{\sum_{i=1}^{n} |e_i|}$$

$$F_\beta = \frac{(1+\beta^2) \times R \times P}{R + \beta^2 \times P}$$

一般来说系统的准确率更为重要，通常会使用 $F_{0.5}$ 来双倍地关注准确率。

3.5 GLEU

M^2 指标的不足之一是需要给出明确的错误注释。受机器翻译的启发，错误语法纠正任务可以被视为序列到序列的重写，由此可提出广义语言评估理解度量 (Generalized Language Evaluation Understanding metric, GLEU)。GLEU 计算 n-gram 假设答案相对参考答案的加权精度，它奖励正确更改的 n-gram，同时惩罚出现在源句中但不在参考答案中的 n-gram。在实验中，GLEU 与人类评价的相关度最强。它的计算公式为：

$$GLEU(S, \hat{Y}, R) = BP \cdot \exp\left(\frac{1}{N}\sum_{k=1}^{N} \log p_k\right)$$

其中 S、$\hat{Y}$、R 分别为源句子集合、假设答案集合、参考答案集合，p_k 是 n-gram 精度，惩罚系数 BP 的公式为：

$$BP = \begin{cases} 1, & \text{if } l_h > l_r \\ \exp\left(1 - l_r/l_h\right), & \text{if } l_h \leqslant l_r \end{cases}$$

其中 l_r 是参考答案的 token 数目，l_h 是所有假设答案的 token 数目。

3.6 ERRANT

ERRANT 是 M^2 指标的改进版本，它不要求提供正确的编辑以及错误类型，而是在答案和预测两方面都采取了自动提取编辑以及自动判断错误类型的方便措施。自动提取的好处是减小了标注员的负担，缺点是有一定误差。在自动提取编辑上，ERRANT 采用了以词性和词干作为考量的 Damerau-Levenshtein 算法。在判断错误类型上，ERRANT 采用了一系列针对词性和词干编写的规则。它将错误分为 25 个类别，是第一个能够评估不同错误类型性能的指标，更有利于错误语法纠正系统的开发。

4 未来方向

从目前仍存在的问题与挑战出发，GEC 未来的发展方向有以下几点。中文平行语料的进一步丰富。

- **数据**。对于语言生成模型，无论从数据集的数量还是质量都有很大发展空间。此外针对不同的应用场景，所需的数据类型亦有差异，如何生成特定任务的数据有待研究。如何更好的利用高质量的数据，也是一个待解决的问题。
- **基于序列标注的方法**。基于序列标注的方法仍处于发展阶段，其在模型结果上与基于 NMT 的方法相当，但拥有更快的时间复杂度，显示出很好的发展前景。目前，针对中文的标注空间以及不同粒度的设计仍有待丰富和细化。
- **更好评价指标**。针对不同的数据集，目前仍缺少统一的凭借方法。此外，针对 GEC 任务，现有的评估指标还无法很好的判断语法校正的语法正确性与流利性，原意的重视度以及生成句的多样性，这同样有待研究。

参考文献

[1] BROCKETT C, DOLAN B, GAMON M. Correcting esl errors using phrasal smt techniques[C]//21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL, Sydney, Australia. [S.l.: s.n.], 2006.

[2] JUNCZYS-DOWMUNT M, GRUNDKIEWICZ R. Phrase-based machine translation is state-of-the-art for automatic grammatical error correction[C]//2016 Conference on Empirical Methods in Natural Language Processing. [S.l.]: Association for Computational Linguistics, 2016: 1546-1556.

[3] CHOLLAMPATT S, HOANG D T, NG H T. Adapting grammatical error correction based on the native language of writers with neural network joint models[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. [S.l.: s.n.], 2016: 1901-1911.

[4] KATSUMATA S, KOMACHI M. (almost) unsupervised grammatical error correction using synthetic comparable corpus[C/OL]//Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. Florence, Italy: Association for Computational Linguistics, 2019: 134-138. https://aclanthology.org/W19-4413. DOI: 10.18653/v1/W19-4413.

[5] CHOLLAMPATT S, TAGHIPOUR K, NG H T. Neural network translation models for grammatical error correction[J]. arXiv preprint arXiv:1606.00189, 2016.

[6] MIZUMOTO T, MATSUMOTO Y. Discriminative reranking for grammatical error correction with statistical machine translation[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. [S.l.: s.n.], 2016: 1133-1138.

[7] HOANG D T, CHOLLAMPATT S, NG H T. Exploiting n-best hypotheses to improve an smt approach to grammatical error correction [J]. arXiv preprint arXiv:1606.00210, 2016.

[8] TOUTANOVA K, RUMSHISKY A, ZETTLEMOYER L, et al. Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies [C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. [S.l.: s.n.], 2021.

[9] XIE Z, AVATI A, ARIVAZHAGAN N, et al. Neural language correction with character-based attention[J]. arXiv preprint arXiv:1603.09727, 2016.

[10] JI J, WANG Q, TOUTANOVA K, et al. A nested attention neural hybrid model for grammatical error correction[J]. arXiv preprint arXiv:1707.02026, 2017.

[11] GE T, WEI F, ZHOU M. Fluency boost learning and inference for neural grammatical error correction[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). [S.l.: s.n.], 2018: 1055-1065.

[12] CHOLLAMPATT S, NG H T. A multilayer convolutional encoder-decoder neural network for grammatical error correction[C]//Proceedings of the AAAI Conference on Artificial Intelligence: volume 32. [S.l.: s.n.], 2018.

[13] CHOLLAMPATT S, WANG W, NG H T. Cross-sentence grammatical error correction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. [S.l.: s.n.], 2019: 435-445.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[15] NÁPLAVA J, STRAKA M. CUNI system for the building educational applications 2019 shared task: Grammatical error correction[C/OL]//Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. Florence, Italy: Association for Computational Linguistics, 2019: 183-190. https://aclanthology.org/W19-4419. DOI: 10.18653/v1/W19-4419.

[16] STAHLBERG F, BYRNE B. The CUED's grammatical error correction systems for BEA-2019[C/OL]//Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. Florence, Italy: Association for Computational Linguistics, 2019: 168-175. https://aclanthology.org/W19-4417. DOI: 10.18653/v1/W19-4417.

[17] ZHAO W, WANG L, SHEN K, et al. Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data[J]. arXiv preprint arXiv:1903.00138, 2019.

[18] SUN X, GE T, WEI F, et al. Instantaneous grammatical error correction with shallow aggressive decoding[C]//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). Online: Association for Computational Linguistics, 2021: 5937-5947.

[19] ALIKANIOTIS D, RAHEJA V. The unreasonable effectiveness of transformer language models in grammatical error correction[C/OL]//Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. Florence, Italy: Association for Computational Linguistics, 2019: 127-133. https://aclanthology.org/W19-4412. DOI: 10.18653/v1/W19-4412.

[20] WANG H, KUROSAWA M, KATSUMATA S, et al. Chinese grammatical correction using BERT-based pre-trained model[C/OL]//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. Suzhou, China: Association for Computational Linguistics, 2020: 163-168. https://aclanthology.org/2020.aacl-main.20.

[21] KANEKO M, MITA M, KIYONO S, et al. Encoder-decoder models can benefit from pre-trained masked language models in grammatical error correction[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: [s.n.], 2020: 4248-4254.

[22] YIN F, LONG Q, MENG T, et al. On the robustness of language encoders against grammatical errors[C]//Proceedings of the 58th An-

nual Meeting of the Association for Computational Linguistics. Online: [s.n.], 2020: 3386-3403.

[23] LEWIS M, LIU Y, GOYAL N, et al. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[J]. arXiv preprint arXiv:1910.13461, 2019.

[24] KATSUMATA S, KOMACHI M. Stronger baselines for grammatical error correction using a pretrained encoder-decoder model[C/OL]// 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. Suzhou, China: Association for Computational Linguistics, 2020: 827-832. https://aclanthology.org/2020.aacl-main.83.

[25] YAMASHITA I, KATSUMATA S, KANEKO M, et al. Cross-lingual transfer learning for grammatical error correction[C/OL]//Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020: 4704-4715. https://aclanthology.org/2020.coling-main.415. DOI: 10.18653/v1/2020.coling-main.415.

[26] ROTHE S, MALLINSON J, MALMI E, et al. A simple recipe for multilingual grammatical error correction[C]//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). [S.l.: s.n.], 2021: 702-707.

[27] FLACHS S, STAHLBERG F, KUMAR S. Data strategies for low-resource grammatical error correction[C/OL]//Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications. Online: Association for Computational Linguistics, 2021: 117-122. https://aclanthology.org/2021.bea-1.12.

[28] PAJAK K, PAJAK D. Multilingual fine-tuning for grammatical error correction[J]. Expert Systems with Applications, 2022:116948.

[29] RAHEJA V, ALIKANIOTIS D. Adversarial Grammatical Error Correction[C/OL]//Findings of the Association for Computational Linguistics: EMNLP 2020. Online: Association for Computational Linguistics, 2020: 3075-3087. https://aclanthology.org/2020.findings-emnlp.275. DOI: 10.18653/v1/2020.findings-emnlp.275.

[30] WANG L, ZHENG X. Improving grammatical error correction models with purpose-built adversarial examples[C/OL]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, 2020: 2858-2869. https://aclanthology.org/2020.emnlp-main.228. DOI: 10.18653/v1/2020.emnlp-main.228.

[31] HOTATE K, KANEKO M, KOMACHI M. Generating diverse corrections with local beam search for grammatical error correction[C/OL]// Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020: 2132-2137. https://aclanthology.org/2020.coling-main.193. DOI: 10.18653/v1/2020.coling-main.193.

[32] YUAN Z, BRYANT C. Document-level grammatical error correction [C]//Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications. Online: Association for Computational Linguistics, 2021: 75-84.

[33] LIU Z, YI X, SUN M, et al. Neural quality estimation with multiple hypotheses for grammatical error correction[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online: [s.n.], 2021: 5441-5452.

[34] DAI Y, LI L, ZHOU C, et al. " is whole word masking always better for chinese bert?": Probing on chinese grammatical error correction[J]. arXiv preprint arXiv:2203.00286, 2022.

[35] LI Z, PARNOW K, ZHAO H. Incorporating rich syntax information in grammatical error correction[J]. Information Processing & Management, 2022, 59(3):102891.

[36] LAI S, ZHOU Q, ZENG J, et al. Type-driven multi-turn corrections for grammatical error correction[J]. arXiv preprint arXiv:2203.09136, 2022.

[37] AWASTHI A, SARAWAGI S, GOYAL R, et al. Parallel iterative edit models for local sequence transduction[J]. arXiv preprint arXiv:1910.02893, 2019.

[38] MALMI E, KRAUSE S, ROTHE S, et al. Encode, tag, realize: High-precision text editing[J]. arXiv preprint arXiv:1909.01187, 2019.

[39] OMELIANCHUK K, ATRASEVYCH V, CHERNODUB A, et al. Gector–grammatical error correction: tag, not rewrite[J]. arXiv preprint arXiv:2005.12592, 2020.

[40] STAHLBERG F, KUMAR S. Seq2edits: Sequence transduction using span-level edit operations[J]. arXiv preprint arXiv:2009.11136, 2020.

[41] LI P, SHI S. Tail-to-tail non-autoregressive sequence prediction for Chinese grammatical error correction[C]//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). Online: Association for Computational Linguistics, 2021: 4973-4984.

[42] TARNAVSKYI M, CHERNODUB A, OMELIANCHUK K. Ensembling and knowledge distilling of large sequence taggers for grammatical error correction[J]. arXiv preprint arXiv:2203.13064, 2022.

[43] YASUNAGA M, LESKOVEC J, LIANG P. Lm-critic: Language models for unsupervised grammatical error correction[C]//Empirical Methods in Natural Language Processing (EMNLP). [S.l.: s.n.], 2021.

[44] KANTOR Y, KATZ Y, CHOSHEN L, et al. Learning to combine grammatical error corrections[C/OL]//Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. Florence, Italy: Association for Computational Linguistics, 2019: 139-148. https://aclanthology.org/W19-4414. DOI: 10.18653/v1/W19-4414.

[45] HINSON C, HUANG H H, CHEN H H. Heterogeneous recycle generation for Chinese grammatical error correction[C/OL]//Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020: 2191-2201. https://aclanthology.org/2020.coling-m

ain.199. DOI: 10.18653/v1/2020.coling-main.199.

[46] CHEN M, GE T, ZHANG X, et al. Improving the efficiency of grammatical error correction with erroneous span detection and correction [C/OL]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, 2020: 7162-7169. https://aclanthology.org/2020.emnlp-main.581. DOI: 10.18653/v1/2020.emnlp-main.581.

[47] WANG Y, KONG C, YANG L, et al. Yaclc: A chinese learner corpus with multidimensional annotation[J]. arXiv preprint arXiv:2112.15043, 2021.

[48] WU S H, LIU C L, LEE L H. Chinese spelling check evaluation at sighan bake-off 2013.[C]//SIGHAN@ IJCNLP. [S.l.]: Citeseer, 2013: 35-42.

[49] YU L C, LEE L H, TSENG Y H, et al. Overview of sighan 2014 bake-off for chinese spelling check[C]//Proceedings of The Third CIPS-SIGHAN Joint Conference on Chinese Language Processing. [S.l.: s.n.], 2014: 126-132.

[50] TSENG Y H, LEE L H, CHANG L P, et al. Introduction to sighan 2015 bake-off for chinese spelling check[C]//Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing. [S.l.: s.n.], 2015: 32-37.

[51] XU S, ZHANG J, CHEN J, et al. Erroneous data generation for grammatical error correction[C/OL]//Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. Florence, Italy: Association for Computational Linguistics, 2019: 149-158. https://aclanthology.org/W19-4415. DOI: 10.18653/v1/W19-4415.

[52] MITA M, KIYONO S, KANEKO M, et al. A self-refinement strategy for noise reduction in grammatical error correction[C/OL]//Findings of the Association for Computational Linguistics: EMNLP 2020. Online: Association for Computational Linguistics, 2020: 267-280. https://aclanthology.org/2020.findings-emnlp.26. DOI: 10.18653/v1/2020.findings-emnlp.26.

[53] WAN Z, WAN X, WANG W. Improving grammatical error correction with data augmentation by editing latent representation[C/OL]// Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020: 2202-2212. https://aclanthology.org/2020.coling-main.200. DOI: 10.18653/v1/2020.coling-main.200.

[54] ZHOU W, GE T, MU C, et al. Improving grammatical error correction with machine translation pairs[C/OL]//Findings of the Association for Computational Linguistics: EMNLP 2020. Online: Association for Computational Linguistics, 2020: 318-328. https://aclanthology.org/2020.findings-emnlp.30. DOI: 10.18653/v1/2020.findings-emnlp.30.

[55] WHITE M, ROZOVSKAYA A. A comparative study of synthetic data generation methods for grammatical error correction[C/OL]// Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications. Seattle, WA, USA → Online: Association for Computational Linguistics, 2020: 198-208. https://aclanthology.org/2020.bea-1.21. DOI: 10.18653/v1/2020.bea-1.21.

[56] TAKAHASHI Y, KATSUMATA S, KOMACHI M. Grammatical error correction using pseudo learner corpus considering learner's error tendency[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. Online: [s.n.], 2020: 27-32.

[57] KOYAMA A, HOTATE K, KANEKO M, et al. Comparison of grammatical error correction using back-translation models[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop. Online: [s.n.], 2021: 126-135.

[58] STAHLBERG F, KUMAR S. Synthetic data generation for grammatical error correction with tagged corruption models[C]//Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications. [S.l.: s.n.], 2021: 37-47.

[59] YANG L, WANG C, CHEN Y, et al. Controllable data synthesis method for grammatical error correction[J]. Frontiers of Computer Science, 2022, 16(4):1-10.

[60] DAHLMEIER D, NG H T. Better evaluation for grammatical error correction[C]//Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. [S.l.: s.n.], 2012: 568-572.

[61] NAPOLES C, SAKAGUCHI K, POST M, et al. Ground truth for grammatical error correction metrics[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). [S.l.: s.n.], 2015: 588-593.

[62] BRYANT C, FELICE M, BRISCOE E. Automatic annotation and evaluation of error types for grammatical error correction[C]//[S.l.]: Association for Computational Linguistics, 2017.

[63] KANEKO M, TAKASE S, NIWA A, et al. Interpretability for language learners using example-based grammatical error correction[J]. arXiv preprint arXiv:2203.07085, 2022.

[64] CHOLLAMPATT S, WANG W, NG H T. Cross-sentence grammatical error correction[C/OL]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 435-445. https://aclanthology.org/P19-1042. DOI: 10.18653/v1/P19-1042.

[65] OMELIANCHUK K, ATRASEVYCH V, CHERNODUB A, et al. GECToR – grammatical error correction: Tag, not rewrite[C/OL]// Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications. Seattle, WA, USA → Online: Association for Computational Linguistics, 2020: 163-170. https://aclanthology.org/2020.bea-1.16. DOI: 10.18653/v1/2020.bea-1.16.

[66] GRUNDKIEWICZ R, JUNCZYS-DOWMUNT M. Near human-level performance in grammatical error correction with hybrid machine translation[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. [S.l.: s.n.], 2018: 284-290.

[67] ROZOVSKAYA A, ROTH D. Grammatical error correction: Machine translation and classifiers[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). [S.l.: s.n.], 2016: 2205-2215.

Yuying Lin, 2020 undergraduate majoring in computer science and technology, Nankai University. Her research interests include GEC