

CCL24-Eval任务6系统报告：中小学作文修辞识别与理解

赵亮
洪恩完美未来科技有限公司
zhaoliang802652@ihuman.com
余浩
洪恩完美未来科技有限公司
yuhao602137@pwr.com

王伟轩
洪恩完美未来科技有限公司
wuweixuan801519@ihuman.com
鲁文斌
洪恩完美未来科技有限公司
luwenbin@ihuman.com

摘要

本技术报告是对2024CCL评测任务(中小学作文修辞识别与理解评测)的一种解决方案。在中小学生的学习过程中，修辞手法不仅是阅读理解和写作技巧的核心组成部分，同时也是塑造优秀文学作品的不可或缺的元素。识别并理解学生作文中的修辞使用，可以帮助学生提高作文表达能力，指导学生更高质量的叙述和描写。对修辞的识别目前属于自然理解领域比较困难的任務，因为需要用到人类领域的大量先验知识，而且很多时候不同的修辞之间的边界还是模糊的。我们通过lora技术直接微调基于qwen-chat-7B的大语言预训练模型，来进行修辞类别的识别。

我们的主要创新技术点为：基于相同的输入输出数据来构造多条训练数据提升算法表现；分级分层来进行修辞的判断，先进行大的修辞类别判断，再把大的修辞类别做为输入对修辞的子类别进行判断；针对修辞成分抽取的任务，直接输出对应的结果文本，再对应回原文本进行位置检索，而不是直接输出索引下标。

关键词： 大语言预训练模型；修辞判断；文本分类

System Report for CCL24-Eval Task 6: Identification and Understanding of Rhetoric in Primary and Secondary School Essays

Liang Zhao
ihuman
zhaoliang802652@ihuman.com
Hao Yu
ihuman
yuhao602137@pwr.com

WeiXuan Wu
ihuman
wuweixuan801519@ihuman.com
WenBin Lu
ihuman
luwenbin@ihuman.com

Abstract

This technical report presents a solution to the 2024 CCL evaluation task (Rhetorical Device Recognition and Understanding in Primary and Secondary School Essays). In the learning process of primary and secondary school students, rhetorical devices are not only core components of reading comprehension and writing skills but also indispensable elements in shaping excellent literary works. Recognizing and understanding the use of rhetoric in student essays can help students improve their writing skills and guide them to produce higher-quality narratives and descriptions. Identifying rhetorical devices is currently a challenging task in the field of natural understanding because it requires extensive human domain knowledge, and often the boundaries between different rhetorical devices are blurred. We utilize LoRA technology to fine-tune

a large pre-trained language model based on Qwen-Chat-7B to recognize categories of rhetorical devices.

Our main innovative technical points are: constructing multiple training data instances based on the same input-output data to improve algorithm performance; using a hierarchical approach to judge rhetorical devices by first determining broad categories of rhetoric and then using these broad categories as input to determine subcategories of rhetoric; and for the task of extracting rhetorical components, directly outputting the corresponding result text and then mapping it back to the original text for position retrieval, rather than directly outputting index positions.

Keywords: Large Pre-trained Language Model , Rhetorical Judgment , Text Classification

1 引言

1.1 引言

本报告主要是提出一种帮助中小学教师进行修辞识别的大模型自动识别算法。熟练使用修辞是中小学生学习写作的一项重要而基本的要求，所以能正确高效的对中小学生学习写作的修辞手法进行准确的识别，可以为语文老师提供十分重要的帮助。本报告做的三个任务分别是修辞形式分类，修辞内容分类，和修辞成分抽取。

我们首先会进行修辞的大类别识别，主要是中小学常见的四个修辞类别：比喻，拟人，排比，夸张。然后我们会对修辞形式类型进行细粒度的识别，比喻包括：明喻、暗喻、借喻；比拟包括：名词、动词、形容词、副词；夸张包括：直接夸张、间接夸张、融合夸张；排比包括：句子排比、成分排比。再然后我们会对修辞内容类型进行细粒度的识别，比喻包括：实在物、动作、抽象概念；比拟包括：拟人、拟物；夸张包括：扩大夸张、缩小夸张、超前夸张；排比包括：并列、承接、递进。最后，进行修辞成分的抽取，修辞成分由连接词、连接对象、连接内容三个部分组成。我们通过利用大模型算法对这三个任务进行训练，得到了最终适合的结果。

1.2 主要困难

比喻和比拟的近似。很多时候比喻和比拟之间的界限是十分模糊的。比如说图1中的多个例子能准确区分是比喻还是比拟实际上是比较困难的。

夸张的判断。很多时候一个句子是否是夸张需要大量结合具体知识甚至很多时候需要结合具体说话场景，夸张的判断对语境的理解和要求是比较高的，如图2这里例举了一些比较难判断的具体例子。

修辞成分的抽取。修辞成分给的输出直接是对应的索引数值，直接从输入的文本，转成输出的索引是非常抽象的。但是如果输出的是文本内容，则即使只差一个字符也无法对应到原来的文本内容中去，所以修辞抽取这个任务的最终输出的方式也是一个比较困难的问题。

1.3 数据集分析

全部数据集包括训练集、验证集、测试集，每个集合条数的分布可以参照表1。训练集和测试集中大的修辞类别和不同的子类别所占的比例可以参照表2和表3。可以看出不同修辞类别尤其是子类别之间的数量相差是十分巨大的，尤其是“排比”和“夸张”两个类别中很多子类别的数目是十分的稀疏的，这种分布特性也非常增大了分类的难度。

为了进行较为充分的实验比对，本文主要实验数据为线下测试集的统计，线上效果最终为任务1: 53.77，任务2: 60.15，任务3: 68.26。

2 技术方案

我们的整体技术方案是使用lora(Hu et al. [2021])方法对qwen-chat-7B(Bai et al. [2023])大语言模型进行针对性的微调，lora方法已经集成到peft库中(Han et al. [2024])。我们做3个任务使用的是同一个语言大模型，实际上包括了4个子任务：修辞大类别分类，修辞形式分类，修辞内容分类，以及修辞成分抽取。

```

{
  "id": 1229,
  "conversations": [
    {
      "from": "user",
      "value": "给你一个句子，‘‘她个子高挑，是班上最高的女生，却又很瘦，远看就像一根纤细的竹竿。’’已知使用了‘‘比拟’’的修辞手法，请你对这个句子进行判断修辞具体的‘‘内容类型’’",
    },
    {
      "from": "assistant",
      "value": "拟物"
    }
  ]
},
{
  "id": 1265,
  "conversations": [
    {
      "from": "user",
      "value": "请找到下面句子中全部的修辞手法，‘‘奋斗在一线的，大多是青年，你们总要接过先辈们手中的火炬，继承发展先辈们的精神与品格。’’",
    },
    {
      "from": "assistant",
      "value": "[ 比喻 ]"
    }
  ]
}
}

```

Figure 1: 相近的比喻和比拟的示例

```

{
  "id": 1360,
  "conversations": [
    {
      "from": "user",
      "value": "请找到下面句子中全部的修辞手法，‘‘外祖父是方圆几十里最出名的人物，二胡声总能把鸟儿引来与他一同伴奏。’’",
    },
    {
      "from": "assistant",
      "value": "[ 夸张 ]"
    }
  ]
},
{
  "id": 3424,
  "conversations": [
    {
      "from": "user",
      "value": "给你一个句子，请你找到句子中所有的修辞手法。给你的句子是‘‘我怕得是欢喜的回到家后，人都不在，好像只要我回家慢一点，就可以让家中有人等待着我。’’",
    },
    {
      "from": "assistant",
      "value": "[ 夸张 ]"
    }
  ]
}
}

```

Figure 2: 比较难判断的夸张的示例

数据集	数目
训练集	约700
验证集	约300
盲测集	约5000

Table 1: 数据集统计

修辞形式	数目
明喻	193
暗喻	129
借喻	76
名词	12
动词	131
形容词	20
副词	7
直接夸张	56
间接夸张	28
融合夸张	24
成分排比	67
句子排比	48

Table 2: 修辞形式的统计分布

修辞内容	数目
实在物	233
动作	33
抽象概念	131
拟人	145
拟物	19
扩大夸张	98
缩小夸张	9
并列	84
承接	5
递进	26

Table 3: 修辞内容的统计分布

2.1 预处理阶段

针对“修辞形式”分类和“修辞内容”分类，首先我们把数据拆成两层，第一层针对大的修辞进行分类，第二层则针对两个子任务进行分类。在对第二层进行分类的时候，把第一层分类的输出直接当作输入，举个例子，输入是：修辞的“形式类型”指的是对一个修辞手法进行更细粒度的修辞分类，给你一个句子，已知使用了“比拟”的修辞手法，请你基于这个修辞手法，对这个句子进行判断修辞具体的“形式类型”，给你的句子是它枝干虬劲有力，高大挺拔，宛如一位镇守在诸陪葬坑前的披坚执锐的秦朝战士。对应的输出为：“名词”。

“修辞成分抽取”则把索引转化成对应的文本进行生成来增加预测的准确率和稳定性，举个例子，输入是：一个句子是“白衣天使把我们带进了真正的春天。”，请判断具体的“修辞成分”，分别是“描写对象、连接词、描写内容”。对应的输出为：[None, None, “真正的春天”]。

2.2 一级修辞大类别的分类

直接利用lora的技术finetune语句的输出进行分类，不再使用额外的分类网络，从而使学习更加集中。先构建一个prompt，在相应位置上填充对应的句子，输出对应的修辞大类别的序列，可以为零个、一个、或者多个，如果为多个大分类则进行全排列有多个相同训练权重的对应输出。

为了增强算法的稳定性，同时进一步优化算法的泛化能力，采用多种prompt开头的方法对其进行针对性的优化，即一个相同的输入数据使用多种prompt来构建多个输入数据来增强模型对应的综合能力。原始输入为“请找到下面句子中全部的修辞手法，火红的夕阳下，那只金色的翅膀直直地伸向天空，犹如一块金色的墓碑，这是老狼紫岚的墓。”，改变后的输入为一个集合{“请找到下面句子中全部的修辞手法，火红的夕阳下，那只金色的翅膀直直地伸向天空，犹如一块金色的墓碑，这是老狼紫岚的墓。”，“给你一个句子，请你找到句子中所有的修辞手法。给你的句子是火红的夕阳下，那只金色的翅膀直直地伸向天空，犹如一块金色的墓碑，这是老狼紫岚的墓。”，……}，然后对输出结果进行判断。做预测的时候则随机选择一个prompt进行输出。

2.3 基于一级修辞分类的二级修辞子类别的分类

和一级修辞大类别的分类一样，直接利用lora的技术finetune语句的输出进行分类，不再使

用额外的分类网络。同样如同“预处理阶段”所说的那样，把第一层分类的输出直接当作输入来辅助二级修辞子类别的分类。

为了增强算法的稳定性，也同样采用多种prompt开头的方法对其进行针对性的优化。并且两个任务“修辞形式”和“修辞内容”，都会把第一阶段的一级修辞大类别输出的结果当作前置条件做为输入，当在训练阶段是直接把真实的一级修辞大类别做为输入这样会使得相对的训练误差降低，然后在预测阶段由于没有准确的一级于此分类就把第一阶段的预测结果来当作输出。通过不同参数和prompt的组合在验证集上的反复调试，我们最终得到了相对最优的训练结果。

2.4 修辞成分抽取的文本和索引之间的转化

修辞成分抽取的任务，输入部分还是会针对同一个输入数据利用多个prompt来构造多条有差异性的输入数据，输出部分我们把索引在原文中进行了定位得到了原本的输出内容。然后训练数据变成了文本输入和文本输出，从而降低了训练的难度。值得注意的是，在预测阶段，预测输出的文本可能与原文不完全一致，我们采用了子串对比编辑距离的方式，只要输出文本和输入文本的子串之间的编辑距离小于一定的值，则认为完全一致，直接把相应的子串当作结果。

最后我们再根据确定的子串在原文中的索引，来得到最终的输出结果。在实验中我们同样尝试了不同的最大编辑距离阈值，最终我们选择了2做为最大可接受的编辑距离。具体的结果可以参见实验篇章。

3 实验结果

针对第二章的技术方案我们在官方提供的数据集上进行了充分的实验，具体的实验结果可以参照表4和表5。

如表4所示，在修辞成分抽取的任务中，我们也实验了多个不同最大编辑距离的阈值来进行比较。最终我们确定了把最大编辑距离设置为2，可以得到相对最优的输出结果。

如表5所示，最优模型为复合了多个句子做为输入数据进行训练得到的，我们针对线上数据集的预测也和线下的预测方式和相关参数是保持相应一致的。最终我们确立每组输入数据配合3个promot进行组合，可以训练得到相对最优模型。

最大编辑距离的阈值 \ 输入的prompt数目	1	2	3	4	5	6
1	0.705	0.711	0.713	0.720	0.709	0.711
2	0.741	0.747	0.753	0.749	0.748	0.752
3	0.732	0.731	0.736	0.729	0.719	0.721
4	0.730	0.729	0.741	0.719	0.731	0.732

Table 4: 修辞成分抽取实验结果统计

任务 \ 输入的prompt数目	1	2	3	4	5	6
task1	0.559	0.562	0.590	0.588	0.581	0.583
task2	0.551	0.558	0.571	0.572	0.566	0.569

Table 5: 修辞形式分类和修辞内容分类实验结果统计

4 总结

我们通过实验证明了针对中小学作文修辞识别和理解任务的几个方法的有效性：对单条输入数据使用不同prompt扩充为多条训练数据来构造训练集；分级分层来进行修辞判断的方法，即先进行大的修辞类别判断，再把大的修辞类别做为输入对修辞的子类别进行判断，来优化算法输出；针对修辞成分抽取的任务，直接输出对应的结果文本，再对应回原文本进行位置检索，同时使用控制最大编辑距离的方法来检查位置的合理性。

在中小学作文修辞识别和理解这个任务中后续还有很大的算法效果提升空间，希望本报告的方法能帮助更多相关研究人员进行相关算法的优化和改进。

References

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023.

Zeyu Han, Chao Gao, Jinyang Liu, Sai Qian Zhang, et al. Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv preprint arXiv:2403.14608*, 2024.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.