题目：**基于大语言模型数据增强的文本评测研究**

论文提纲

1 **引言（**本章节将详细阐述自然语言处理领域中文本评分任务的重要性及其面临的挑战。特别指出标注数据不足对模型性能的影响，并介绍数据增强技术在解决这一问题中的潜在价值。

* 1. **研究背景及意义**

在自然语言处理（NLP）领域，尤其是在文本评分任务中，数据集的质量和规模对模型性能有着决定性的影响。然而，高质量的标注数据往往难以获得，因为它们需要大量的人力和时间来创建和验证。此外，对于特定领域，如罕见语言或专业主题，高质量的标注数据可能微乎及微。而在其中，数据增强技术可以通过生成额外的训练样本来扩大和多样化数据集，从而提高模型的泛化能力和鲁棒性。在文本评分领域，这意味着能够更准确地评估文本的质量，无论是在教育、出版还是内容审核等领域。本研究旨在探讨如何通过数据增强技术，特别是利用大语言模型和文本蒸馏方法，来提高文本评分模型的性能。本研究提出了一种基于大语言模型的数据增强稳步评测研究，探究在数据紧缺的情况下，借由大语言模型进行数据增强和数据蒸馏方法，是否能够得出用于训练的较好质量的数据集，以提升文本分类模型的性能。

* 1. **国内外研究现状分析**

在自然语言处理领域，数据增强技术已经成为提升模型性能的重要手段，特别是对于文本分类任务。当前对于文本数据不足想出来的数据增强方法有很多，如Wei等人[1]发明的EDA方法通过同义词替换，随机插入，随机删除，随机交换，回译，文本重组的方式对数据进行增强，从而实现数据的多样化。Qian等人[2]发明的AEDA则通过更简单地插入标点符号的方式来进行数据增强，更加简化了对文本的处理，实现处理的轻量化。还有研究者通过相似度的计算来进行未标注文本的分类，变向实现从未标注数据到标注数据的转化，从而实现数据增强(如Su等人[5])。

还有一些研究者，为了适应不同领域作文所带来的作文分布差异，通过优化模型的方式构建了一些鲁棒性较好的跨领域作文自动评分模型。

随着近些年transfomer的发展，大语言模型如GPT等被证实在数据增强上具有较高的可靠性，因此也出现了其他数据增强的方法，比如运用大语言模型进行数据增强。

然而，在现有的方法中，传统的方法会使得模型学到的特征比较单一，使得训练得出的模型灵活性较低。跨领域的作文评分模型能够通过模型较好的鲁棒性实现较为不错的评分效果，但是建模却较为复杂，且难以适应不同评分标准的作文数据集。而现代运用大语言模型生成数据的方法优化大多在于修改大语言模型的提示词（即运用提示词工程）以得到更高质量的文本（如Chen等人[4]的研究）。但由于大语言模型生成的随机性，即使是通过缜密的提示词工程得到的文本依然无法保证所有文本的高质量性。

总的来说，现有方法面临着三重挑战：传统文本增强易导致特征单一化学习；跨领域模型构建存在复杂度与适配性矛盾；大语言模型生成数据的质量不可控性。这些瓶颈制约了自动评分系统的实用化进程，亟待建立有效的生成数据质量控制机制。

* 1. **本文研究内容**

本研究提出基于大语言模型及数据蒸馏的双阶段增强框架（LLM-DA），通过生成-蒸馏协同机制提升文本评分模型性能。研究内容包含三个核心模块：

大语言模型数据增强模块：通过当前给定的作文评分标准，构建提示词模板，运用大语言模型生成大量对应评分的数据。

数据过滤模块：设计基于词汇复杂及多样性指数评估，文章句型运用与主题连贯性评估，以及基于规则语法检测的三通道过滤机制，对进入的数据集进行过滤。将数据集进行分类，分离出符合要求的高质量数据集和不符合要求的低质量数据集。

数据自适应循环改写模块：记录下不符合要求的低质量数据集，根据其未达标的指标构建不同的大语言提示词模版，重新利用大语言模型进行修改，直至符合数据过滤模块的标准，得到质量良好的数据集。

总体来说，本研究的贡献在于：1. 提出了双阶段增强框架（LLM-DA），解决了大语言模型因为生成的随机性导致数据质量参差不齐的情况。2. 将数据增强运用于AES模型构建当中，并通过蒸馏得到的高质量数据集得出了评分能力良好的模型。3. 通过将模型进行对比分析，证明了此方法优于传统数据增强，及未经过数据蒸馏训练得出的模型。

* 1. **论文结构**

本研究方法的模型如下所示：

相关研究（本章节将概述一些面对数据不足的当今研究方法。包括当前数据增强技术的发展状况以及本研究旨在解决的具体问题：如何通过数据增强技术提升文本评分模型在数据稀缺情况下的性能。）

任务定义（本章节将介绍aes评分任务）

研究方法介绍（本章节将详细介绍本研究采用的数据增强技术，包括大语言模型的文本生成方法和文本蒸馏技术的原理及其工作流程）

模型构建与实验设计（本章节将描述基准模型的构建过程，包括多种不同条件下的模型：未经过数据增强的模型、经过大语言模型数据增强但没有经过蒸馏的模型、经过大语言模型数据增强且数据经过蒸馏的模型，运用其他数据增强方法（如eda）得出的模型，以及通过其他非此主题文章训练得出的模型。详细说明实验的设计，包括数据集的选择、大模型提示词的更改，模型的选择和验证过程，以及性能评估的标准。）

1. **相关技术介绍**
   1. **机器学习技术**
   2. **深度学习技术**
   3. **大语言模型技术**
2. **基于大语言模型的数据增强**
   1. **研究动机**
   2. **数据集的选择**
   3. **大模型提示词设计**
   4. **数据增强方法**
3. **实验结果与分析**
   1. **数据集**
   2. **基准模型**
   3. **评价指标**
   4. **结果分析与讨论**

**5 总结**

结果分析与讨论（本章节将展示实验结果，对比不同模型的性能，并分析数据增强与文本蒸馏技术对模型性能的影响。通过对比分析，展示本数据增强技术在经过数据蒸馏之后如何提升模型在数据稀缺情况下的性能）

局限性探讨（本章节将讨论模型可能会存在的潜在问题，以及其对应的可能解决方法，表面后续研究努力的方向）