

2024 年(第 12 届)“泰迪杯”数据挖掘挑战赛——

C 题：竞赛论文的辅助自动评阅

一、问题背景

近年来我国各领域各层次学科竞赛百花齐放，层出不穷，学生参与度也越来越高。随着参赛队伍的增加，评阅论文的工作量急剧增加，这对评阅论文的人力要求也越来越大。因此引入机器辅助评阅成为竞赛主办方的现实需求。

在学术界，建立基于 AI 的学术论文自动评审模型已得到了许多研究者的关注。论文的自动评阅涉及多种传统的自然语言处理技术如文本分类、信息抽取、论辩挖掘等。近年来，随着深度学习和自然语言处理技术的不断发展，特别是以 GPT 为代表的大语言模型的出现，进一步促进了论文自动评阅技术的发展，使得利用 AI 进行文本的自动评阅变得越来越可行，逐步从实验室走向学校和更多组织机构，成为当前的技术热点。但是在特定领域实现论文自动评阅仍然存在很多挑战，需要利用预训练的大语言模型适配具体的应用场景来解决问题。

二、解决问题

1、构造论文质量特征

每个指标的分数范围为 0-10 分。

(1) 论文的完整性评价

对照赛题，比对竞赛论文中相关问题的章节或段落，对论文的完整性进行评价。评估竞赛论文是否能完整解答赛题，并给出评价论文完整性的技术手段和评分标准。

(2) 论文有无实质性工作

对照赛题评阅要点，查找竞赛论文中相关问题的章节或段落，考察论文是否就赛题问题做出了相关的研究。需给出相关的技术方法和评价标准。

(3) 摘要质量

摘要与内容的一致性评价。评价摘要是否如实反映正文的中心思想，即衡量内容摘要与正文的相关性、一致性。需给出摘要质量评价指标及其依据。

(4) 写作水平评价

评价文字流畅性、写作规范（图、表、摘要）性和论文逻辑性。在传统论文评分（essay scoring）技术基础上，从文本通顺、立意分析、篇章结构、论证挖掘等维度进行探索，挖掘文本蕴含的论点论据、论证关系、结构信息，结合论证挖掘角度评估论文一致性、逻辑性，综合给出论文写作水平的评分。

2、竞赛论文辅助评分

根据上面构造的各项评分指标建立论文的整体评分模型，根据提供的论文集，按照十分制给出每篇论文的综合评分，将结果保存到 result.xlsx 文件中。综合评分结果要求满足如下限制条件：

8-10 分的不超过 3%；

6-7 分的不少于 10%，6-10 分不超过 15%；

4-5 分不少于 20%，4-10 分不超过 35%；

其他的为 0-3 分。

一般而言，在综合评分中论文的完整性和写作水平的分数占比之和不超过 40%。

注 1 若使用预训练的大语言模型完成赛题任务，需要给出实现过程，如提问时使用的提示词及如何利用提问结果。

注 2 自 2022 年底 ChatGPT 发布以来，大语言模型的能力突飞猛进，可考虑将大语言模型技术应用于本次竞赛。一方面可考虑使用 ChatGPT、讯飞星火、文心一言、智谱清言等国内外大模型接口，基于大语言模型设计算法和构建合适的提示词等，辅助完成本赛题的任务。另一方面也可考虑微调训练开源大语言模型，例如 ChatGLM、Qwen、Baichuan 等系列开源大语言模型，设计训练任务，让知识赋能大语言模型以更好地解决问题。

三、附件说明

附件 1 为竞赛论文集，附件 2 为赛题和参考评阅标准，附件 3 为 result.xlsx 的结果模板。

表 1 result.xlsx 样例

论文编号	完整性	实质性	摘要	写作水平	综合评分
C001