**东北师范大学研究生课程说明文档**

## **项目题目 [增量提示和验证的层次事件模式归纳](https://arxiv.org/abs/2307.01972" \t "https://www.bing.com/_blank)**

**课程名称 数据挖掘**

**姓 名 王玉卿 学　　号 2024103474**

**专 业 人工智能 年 级 2024级**

**院 、 所 信息科学与技术学院 年 月 日 2025年6月15日**

**研究生课程项目评价标准**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **指 标** | **评价内容** | **评价等级（分值）** | | | | **得分** |
| **A** | **B** | **C** | **D** |
| 选 题 | 选题是否新颖；是否有意义；是否与本门课程相关。 | 20-16 | 15-11 | 10-6 | 5-0 |  |
| 论 证 | 思路是否清晰；逻辑是否严密；结构是否严谨；研究方法是否得当；论证是否充分。 | 20-16 | 15-11 | 10-6 | 5-0 |  |
| 文 献 | 文献资料是否翔实；是否具有代表性。 | 20-16 | 15-11 | 10-6 | 5-0 |  |
| 规 范 | 文字表达是否准确、流畅；体例是否规范；是否符合学术道德规范。 | 20-16 | 15-11 | 10-6 | 5-0 |  |
| 能 力 | 是否运用了本门课程的有关理论知识；是否体现了科学研究能力。 | 20-16 | 15-11 | 10-6 | 5-0 |  |
| 评阅教师签名：  年 月 日 | | 总分： | | | | |

**东北师范大学研究生院**

# 一、代码实现目标：

1.核心任务是给定场景名称（如 “疾病爆发”“政治腐败”），利用 LLM生成包含时间关系和层次关系的事件模式图，具体包括：

事件骨架构建：生成场景的主要事件链。

事件扩展：为每个事件补充前置、后置及子事件。

关系验证：验证事件间的时间顺序和层次包含关系。

2.关键能力

支持开放域场景，无需依赖特定领域的信息抽取系统。通过模块化提示模板扩展不同类型的事件关系（如新增关系时仅需修改提示模板）。结合检索增强（Retrieval-Augmented）提升事件描述的泛化性，避免生成过于具体的实例。

# 二、实现的主要思路

1. 增量提示与验证流程

将复杂事件图的构建分解为三个阶段，每个阶段通过预设的提示模板与 LLM 交互：

阶段 1：事件骨架构建  
 输入场景名称或章节结构，提示 LLM 生成主要事件列表（如 “疾病爆发” 的主要事件包括 “病原体出现”“病例增加”“诊断”）。

阶段 2：事件扩展与验证  
 对每个已生成的事件，通过提示查询其前置事件（What happened before...）、后置事件（What happens after...）、子步骤（What are the steps in...）等，扩展事件图。新增事件需通过重复性检测（字符串 / 嵌入相似度）、特异性检测（排除包含具体人名、时间的实例）和章节归属检测。

阶段 3：事件关系验证  
基于 Allen 区间代数，通过三个问题验证事件对的时间和层次关系：

“Does event A start before event B?”

“Does event A end before event B?”

“Is the duration of event A longer than event B?”  
根据回答概率判断关系类型（如 “先于”“包含”），并通过贪心算法去除时间环，确保图的一致性。

2. 检索增强提示（Retrieval-Augmented Prompting）

利用 Wikipedia 类别页和新闻文档构建场景相关的语料库，通过 TCT-ColBERT 检索模型获取相关段落。

将检索到的段落作为提示的上下文，引导 LLM 生成更具泛化性的事件描述，避免依赖单一实例。

3. 模型与工具集成

LLM 接口：调用 OpenAI API生成事件和关系。

向量检索：使用 Pyserini 实现的 TCT-ColBERT 进行文档检索。

相似度计算：基于 SBERT 计算事件描述的余弦相似度，Jaro-Winkler 计算字符串相似度，用于重复性检测。