

# 项目二：基于大模型的合作审阅报告

3220104041 华东东

2025 年 6 月 7 日

## 1 贡献总结

本项目致力于利用大语言模型 (LLMs) 解决合同条款审阅中的复杂知识密集型任务。通过引入检索增强生成 (RAG) 机制并结合自我一致性 (Self-Consistency) 策略, 显著提升了识别合同中风险条款的准确性和效率。核心贡献在于构建了一个能够有效利用专家知识库和风险审查清单的智能审阅系统, 使得模型在理解特定领域合同条款的含义及潜在风险方面表现更优, 尤其在 F1 分数上取得了显著提升, 证明了该方法在合同审阅自动化方面的潜力。

## 2 技术方案

### 2.1 问题定义与数据集

本项目的目标是根据给定的待审阅合同、风险审查清单和专家知识库, 利用大语言模型筛选出含有风险的合同条款, 并高效准确地识别这些风险条款。衡量指标为 F1 分数。

项目使用以下数据集:

- `contract.csv`: 待审阅合同, 已分割好的合同条款, 包含一级标题和具体条款内容.
- `checklist.csv`: 风险审查清单, 包含检查点 (Risk\_feature) 和常见的风险表现形式 (Identified\_items) .
- `clause_review_pair.csv`: 专家知识库, 包含检查点 (checkpoint)、案例条款 (clause) 和专家对条款的评价 (review) .
- `GT_contract_risk.csv`: Ground Truth, 风险条款参考结果, 其中“A”表示条款有风险, “B”为无风险.

### 2.2 RAG 检索增强生成

为了使大语言模型能够更好地处理特定领域 (合同审阅) 的复杂任务, 本项目采用了 RAG 机制。RAG 方案的核心流程如下:

1. **数据加载与嵌入**: 使用 `CSVLoader` 工具加载 `clause_review_pair.csv` 作为专家知识库。文本数据通过 `HuggingFaceEmbeddings` 的 `llmware/industry-bert-contracts-v0.1` 模型进行嵌入, 生成向量表示.

2. **向量存储与检索**：将嵌入后的知识库存储在 FAISS 向量数据库中。对于 `contract.csv` 中的每一个待审阅条款，使用其内容作为查询，从 FAISS 向量数据库中检索最相关的 K 个（本项目中设置为  $k=2$ ）专家评审案例。
3. **提示词构建**：将检索到的相关专家评审案例（包括检查点、识别项、案例条款及评审意见）与待审阅条款一同构建成结构化的提示词，输入给大语言模型。本项目使用的提示词模板为 `rag3_prompt_template`（详见附录 A）。
4. **大语言模型推理**：使用 ChatTongyi 的 `qwen-plus` 模型进行推理。模型根据提供的检查点、识别项、相关评审案例和待审阅条款，判断该条款是否与检查点相矛盾或相符，并给出解释和风险评估（A：有风险，B：无风险）。

## 2.3 自我一致性 (Self-Consistency)

为了替代 CoT (Chain-of-Thought) 中的贪婪解码并提高预测的鲁棒性，本项目引入了自我一致性策略。具体做法是对每个待审阅条款，重复多次（本项目中设置为 3 次）调用大语言模型进行推理，收集每次的评估结果（A 或 B）。最终的预测结果通过多数投票决定：如果“A”出现的次数多于“B”，则最终评估为“A”；否则为“B”。

## 2.4 评估方法

模型性能通过 `evaluator.py` 脚本进行评估，主要关注以下指标：

- **精确率 (Precision)**: 真正例 / (真正例 + 假正例)，即模型识别为风险条款中，实际为风险条款的比例。
- **召回率 (Recall)**: 真正例 / (真正例 + 假负例)，即所有实际风险条款中，模型正确识别的比例。
- **F1-分数 (F1-Score)**: 精确率和召回率的调和平均值，综合衡量模型的性能。 
$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

其中，“A”被定义为正例（风险条款），“B”为负例（无风险条款）。

# 3 实验分析

## 3.1 实验设置

- **测试集**： `contract.csv` 和 `GT_contract_risk.csv`。测试集中包含 75 条条款，其中 13 条为风险条款，62 条为非风险条款。
- **知识库**： `checklist.csv` 和 `clause_review_pair.csv`。
- **模型**：大语言模型为 `qwen-plus`，嵌入模型为 `llmware/industry-bert-contracts-v0.1`。

## 3.2 实验结果

表 1: 不同方案的性能对比

指标	Standard	RAG	RAG + SC
Precision	0.5714	0.6667	0.7778
Recall	0.3077	0.6154	0.5385
F1-Score	0.4000	0.6400	0.6364

## 3.3 结果分析

从表 1 可以看出：

- **Standard 对比 RAG**：从 Standard 方案（不使用 RAG）到 RAG 方案，F1-Score 从 0.4000 显著提升至 0.6400。这表明引入外部知识库对大语言模型处理合同审阅这类专业性任务的有效性。RAG 能够为模型提供更具体的背景信息和专家意见，帮助模型更准确地判断条款的风险。
- **RAG 对比 RAG + SC**：RAG + SC（RAG 结合自我一致性）方案的 F1-Score (0.6364) 与纯 RAG 方案 (0.6400) 相比略有下降，但 Precision 有明显提升（从 0.6667 提升至 0.7778），而 Recall 略有下降（从 0.6154 降至 0.5385）。这说明自我一致性策略可能使模型在做出“风险”判断时更加保守，从而减少了假阳性（提高了 Precision），但可能也因此错过了一些真正的风险条款（降低了 Recall）。在某些应用场景下，高 Precision 可能比高 Recall 更为重要，以减少误报。

### 3.3.1 案例分析

以“项目经理信息”检查点为例：

- **原始 Standard 方案的输出：**

```
specific condition: ```Article 6.1 The project manager for this project is
[Project Manager Name], but no ID number is provided, and the
authorization scope is described in a very general way without clear
boundaries.```
```

Condition situation: notfound

Explanation: The checkpoint "Project Manager Information" does not provide enough detail to determine if the specific condition either contradicts or entails it. The specific condition describes the lack of an ID number and vague authorization scope, but there is no explicit mention in the checkpoint about these requirements. Therefore, we cannot definitively conclude if there is a contradiction or entailment.

Assessment: B

在 Standard 方案中，由于没有检索到与“项目经理信息”检查点直接相关的详细要求，模型无法判断该条款是否存在风险，因此评估为“B”（无风险或未发现）。

- RAG 方案的输出：

Summary: The exemplars demonstrate that the checkpoint "Project Manager Information" is met when the project manager's name, ID number, and a clearly defined authorization scope are explicitly stated in the contract.

Condition situation: contradict

Explanation: The specific condition contradicts the checkpoint because it lacks an ID number and provides a vague description of the authorization scope, which does not meet the clarity required by the checkpoint.

Assessment: A

在 RAG 方案中，模型能够从知识库中检索到关于“项目经理信息”检查点的详细示例（例如要求明确的姓名、ID 号和授权范围），从而明确判断当前条款（缺少 ID 和授权范围模糊）与检查点相矛盾，评估为“A”（有风险）。这突出 RAG 在补充上下文信息和提高模型判断能力方面的优势。

### 3.3.2 局限性分析

尽管 RAG 带来了显著的性能提升，但仍存在一些局限性：

- **嵌入模型准确性要求高：** RAG 的效果很大程度上依赖于嵌入模型将查询与知识库中的相关案例进行精确匹配的能力。如果检索到的案例不够相关，即使知识库中存在正确信息，模型也可能无法做出准确判断。例如，对于条款“Lessee(Party B): Li Si ID Card No.: 123412199405064321”，期望检索到关于“Missing Subject Information”的识别项（例如“Party B (as a company) does not provide business license information”），但实际可能检索到关于“Property lacks ownership certificate”的案例，导致误判。
- **条款上下文信息缺失：** 当前的 RAG 方案主要针对单个合同条款进行审阅。然而，合同条款之间往往存在上下文关联，单个条款的风险可能需要结合前后条款或整个合同的语境才能准确判断。例如，某个条款看似无风险，但结合其他条款可能构成整体性风险。当前的方案可能无法完全捕捉到这种跨条款的复杂关系。

未来的工作可以探索更先进的检索策略，例如基于合同结构或语义关联的检索，以及如何将更广阔的上下文信息有效地融入到提示词中，以进一步提升模型的审阅能力。

## A 使用的 Prompt Template

本项目中使用的三个 Prompt Template 定义在 `rag.py` 文件中。其中 `rag3_prompt_template` 是在最终实验中使用的 RAG 提示词模板。

## A.1 Standard Prompt Template

```
1 standard_prompt_template = """You are a construction contract review AI assistant.
2 You will be provided with checkpoint, specific condition delimited by ```.
3 Your task is to perform the following actions:
4 1 - check whether the specific condition contradicts with or entails with the checkpoint.
5 2 - if there are not enough information to determine whether the specific condition contradicts
   with or entails with the checkpoint, simply output notfound.
6 3 - give explanation for your response.
7 4 - give your final assessment.
8 Use the following format:
9 Condition situation: <contradict or entail or not found>
10 Explanation: <explanation>
11 Assessment: <If there is a risk, answer A; if there is no risk or not found, answer B. Please
   answer ONLY with A or B>
12
13 checkpoint: ```{checkpoint}```
14 specific condition: ```{specific_condition}```
15 """
```

Listing 1: standard\_prompt\_template

## A.2 RAG Prompt Template (rag\_prompt\_template)

```
1 rag_prompt_template = """You are a construction contract review AI assistant.
2 You will be provided with checkpoint, specific condition, identified items, and relevant review
   delimited by ```.
3 Your task is to perform the following actions:
4 1 - summarize what have you learned from the exemplars first concisely, no need to repeat the
   new question in your answer.
5 2 - check whether the specific condition contradicts with or entails with the checkpoint.
6 3 - if there are not enough information to determine whether the specific condition contradicts
   with or entails with the checkpoint, simply output notfound.
7 4 - give explanation for your response.
8 5 - give your final assessment.
9 Use the following format:
10 Summary: <Summary of the exemplars>
11 Condition situation: <contradict or entail or not found>
12 Explanation: <Explanation>
13 Assessment: <If there is a risk, answer A; if there is no risk or not found, answer B. Please
   answer ONLY with A or B>
14
15 checkpoint: ```{checkpoint}```
16 identified items: ```{identified_items}```
17 {review_cases}
18 specific condition: ```{specific_condition}```
19 """
```

Listing 2: rag\_prompt\_template

### A.3 RAG Prompt Template (rag2\_prompt\_template)

```
1 rag2_prompt_template = """You are a construction contract review AI assistant.
2 You will be provided with checkpoint, specific condition, identified items, and relevant review
   delimited by ``` .
3 Your task is to perform the following actions:
4 1 - summarize what have you learned from the exemplars first concisely, no need to repeat the
   new question in your answer.
5 2 - check whether the specific condition contradicts with or entails with the checkpoint.
6 3 - if there are not enough information to determine whether the specificcondition contradicts
   with or entails with the checkpoint, simply output notfound.
7 4 - give explanation for your response.
8 5 - give your final assessment.
9 Use the following format:
10 Summary: <Summary of the exemplars>
11 Condition situation: <contradict or entail or not found>
12 Explanation: <Explanation>
13 Assessment: <If there is a risk, answer A; if there is no risk or not found, answer B. Please
   answer ONLY with A or B>
14
15 checkpoint: ```{checkpoint}```
16 {review_cases}
17 specific condition: ```{specific_condition}```
18 """
```

Listing 3: rag2\_prompt\_template

### A.4 RAG Prompt Template (rag3\_prompt\_template)

```
1 rag3_prompt_template = """You are a construction contract review AI assistant.
2 You will be provided with checkpoint, specific condition, identified items, and relevant review
   delimited by ``` .
3 Your task is to perform the following actions:
4 1 - summarize what have you learned from the exemplars first concisely, no need to repeat the
   new question in your answer.
5 2 - check whether the specific condition contradicts with or entails with the checkpoint.
6 3 - if there are not enough information to determine whether the specificcondition contradicts
   with or entails with the checkpoint, simply output notfound.
7 4 - give explanation for your response.
8 5 - give your final assessment.
9 Use the following format:
10 Summary: <Summary of the exemplars>
11 Condition situation: <contradict or entail or not found>
12 Explanation: <Explanation>
13 Assessment: <If there is a risk, answer A; if there is no risk or not found, answer B. Please
   answer ONLY with A or B>
14
15 {review_cases}
16 specific condition: ```{specific_condition}```
```

## Listing 4: rag3\_prompt\_template