

## 基于审判逻辑的裁判文书生成方法

廖进超, 杨卫哲, 秦永彬, 黄瑞章, 陈艳平, 周裕林

### 引用本文

廖进超, 杨卫哲, 秦永彬, 黄瑞章, 陈艳平, 周裕林. [基于审判逻辑的裁判文书生成方法](#)[J]. 计算机科学, 2025, 52(11): 223-229.

LIAO Jinchao, YANG Weizhe, QIN Yongbin, HUANG Ruizhang, CHEN Yanping, ZHOU Yulin. [Method for Generating Judgment Documents Based on Trial Logic](#) [J]. Computer Science, 2025, 52(11): 223-229.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

### Similar articles recommended (Please use Firefox or IE to view the article)

#### [Instruct-Malware:基于控制流图的大型语言模型恶意软件分析](#)

Instruct-Malware:Control Flow Graph Based Large Language Model Analysis of Malware  
计算机科学, 2025, 52(11): 40-48. <https://doi.org/10.11896/jsjcx.241100118>

#### [DF-RAG:基于查询重写和知识选择的检索增强生成方法](#)

DF-RAG:A Retrieval-augmented Generation Method Based on Query Rewriting and Knowledge Selection  
计算机科学, 2025, 52(11): 30-39. <https://doi.org/10.11896/jsjcx.241000117>

#### [基于大语言模型增强的零样本知识抽取方法](#)

Zero-shot Knowledge Extraction Method Based on Large Language Model Enhanced  
计算机科学, 2025, 52(11): 22-29. <https://doi.org/10.11896/jsjcx.241000049>

#### [基于组合上下文提示的大型语言模型领域知识问答研究](#)

Research on Domain Knowledge Question Answering via Large Language Models with Compositional Context Prompting  
计算机科学, 2025, 52(11): 13-21. <https://doi.org/10.11896/jsjcx.241200198>

#### [大语言模型与谣言:生成与检测的综述](#)

Large Language Models and Rumors:A Survey on Generation and Detection  
计算机科学, 2025, 52(11): 1-12. <https://doi.org/10.11896/jsjcx.250700034>

# 基于审判逻辑的裁判文书生成方法

廖进超 杨卫哲 秦永彬 黄瑞章 陈艳平 周裕林

公共大数据国家重点实验室(贵州大学) 贵阳 550025

贵州大学计算机科学与技术学院 贵阳 550025

(1762897859@qq.com)

**摘要** 裁判文书自动生成是智慧法院建设中的关键任务之一,旨在提高司法效率与文书质量。由于大模型在司法审判认知上存在盲区,难以理解审理机制和文书规范,导致生成文书在逻辑一致性和结构合理性上存在不足。针对以上问题,提出了一种基于审判逻辑的裁判文书生成方法,利用大语言模型模拟审判推理过程,分阶段生成裁判文书。首先,使用法律要素填充预设模板以描述“基本案情”;其次,对事实与证据进行分析对齐得到“审理事实”;最后,结合知识库检索相关法条生成“法院判决”,并进行拼接生成完整的文书。实验结果表明,相较于基线模型,所提方法在真实案件卷宗数据上的 F1 值,在 ROUGE-1, ROUGE-2 和 ROUGE-L 方面分别提升了 6.03, 6.56 和 7.98 个百分点,验证了所提方法的有效性。

**关键词:** 大语言模型;裁判文书生成;知识库;审判逻辑;智慧法院

**中图分类号** TP391.1

## Method for Generating Judgment Documents Based on Trial Logic

LIAO Jinchao, YANG Weizhe, QIN Yongbin, HUANG Ruizhang, CHEN Yanping and ZHOU Yulin

State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China

School of Computer Science and Technology, Guizhou University, Guiyang 550025, China

**Abstract** The automatic generation of judicial documents is one of the key tasks in the construction of smart courts, aiming to enhance judicial efficiency and document quality. However, due to the blind spots of large models in judicial cognition, they struggle to understand the trial mechanism and document norms, resulting in deficiencies in the logical consistency and structural rationality of the generated documents. To address these issues, this paper proposes a method for generating judicial documents based on trial logic, which utilizes large language models to simulate the trial reasoning process and generate documents in stages. Firstly, legal elements are used to fill in the preset template to describe the “basic case facts”. Secondly, the facts and evidence are analyzed and aligned to obtain the “trial facts”. Finally, relevant legal provisions are retrieved from the knowledge base to generate the “court judgment”, and the complete document is assembled. Experimental results show that, compared with the baseline model on real case file data, the proposed method has improved the F1 values of ROUGE-1, ROUGE-2, and ROUGE-L by 6.03, 6.56, and 7.98 percentage points respectively, verifying the effectiveness of the proposed method.

**Keywords** Large language model, Judgment document generation, Knowledge base, Trial logic, Smart court

法律文书自动生成旨在利用人工智能技术,基于案件信息、法律条文等相关信息,自动生成各类法律文书,是实现智慧法院建设的重要技术支撑<sup>[1]</sup>。随着人民法院受理案件数量不断激增,2024 年各类受理案件数量已达 4 600 余万件<sup>[2]</sup>,“案多人少”的矛盾愈加突出,法官在繁重的工作压力下难以兼顾文书质量与效率。智能化文书撰写凭借其快速且精确的文书生成能力,为解决这一问题提供了有效路径,对推进审判

体系现代化和司法智能化具有重要意义。其中,裁判文书作为诉讼活动结果的载体,清晰记录了案件审理与审判的详细经过,是法律文书自动生成的重要应用实例。

裁判文书生成方法多种多样,目前基于编码器-解码器框架<sup>[3]</sup>的深度学习模型已成为主流的技术路线。这类模型将生成任务建模为序列转换问题,在多种文本生成任务中都展现了卓越的性能。然而,这些模型在特定任务中通常依赖大量

到稿日期:2025-05-14 返修日期:2025-06-27

基金项目:国家重点研发计划(2023YFC3304500);贵州省科学技术基金重点资助项目(黔科合重大专项字[2024]003);贵州省研究生科研基金(2024YJSKYJJ041)

This work was supported by the National Key Research and Development Program of China(2023YFC3304500), Key Funded Projects of the Science and Technology Foundation Program of Guizhou Province(Qian-Kehe Major Special Project No. [2024] 003) and Department of Education of Guizhou Province(2024YJSKYJJ041).

通信作者:秦永彬(ybqin@gzu.edu.cn)

的标注数据和计算资源,且在处理复杂语言结构和长距离依赖时仍存在不足。为解决这些问题,预训练技术应运而生。通过预训练语言模型,生成文本的质量和效率得到显著提升。随着模型规模的扩大,大语言模型(Large Language Model, LLM)逐渐崛起。以 ChatGPT<sup>[4]</sup>为例,它们通过更多参数和更复杂的结构,在海量数据上进行训练,不仅能生成流畅文本,还能完成问答、翻译等任务,展现了强大的通用性。

尽管如此,大模型在面对裁判文书生成任务时仍存在明显局限与挑战。由于其未能捕捉审判过程中的复杂机制,且缺乏对文本逻辑结构的理解,因此生成文本质量不佳。表 1 列出了 GPT 可能会生成的内容,从中可以看出该文本不仅缺乏基本的文书规范,且判决结果缺乏相关司法解释的条文支撑(划线处),这显然违背了审判常识。

表 1 GPT 生成示例(部分有删减)

Table 1 GPT generation example

参考文本	GPT 生成文本
<p>【…】本院认为,被告人醉酒驾驶机动车在道路上行驶,其行为已构成危险驾驶罪。公诉机关指控的罪名成立,本院予以支持。鉴于其到案后如实供述犯罪事实,系坦白,且签署了《认罪认罚具结书》自愿认罪认罚,可以从轻处罚。依照《中华人民共和国刑法》第一百三十三条之一第一款第(二)项、第六十七条第三款、第七十三条第一款、第三款之规定,判决如下:被告人犯危险驾驶罪,判处有期徒刑三个月缓刑四个月,并处罚金人民币一万元。</p>	<p>【…】被告人醉酒驾驶机动车在道路上行驶,其行为已构成危险驾驶罪。依照相关规定,应当依法追究其刑事责任。被告人到案后能如实供述自己的犯罪事实,自愿认罪认罚,可以从轻处罚。综合考虑本案的具体情节,决定被告人判处以下刑罚:被告人犯危险驾驶罪,判处有期徒刑五个月,并处罚金人民币六千元。</p>

裁判文书本质上是案件审判逻辑的书面表达。在司法审判中,法官的审理程序<sup>[5]</sup>主要概括为以下 3 个步骤:1)收集材料,核对案件信息;2)组织调查,查明案件事实;3)依据相关司法解释,作出定罪量刑。而裁判文书各部分内容应与其相互对应。因此,将审判逻辑嵌入裁判文书中,有助于提升文书的逻辑性与准确性。

基于以上分析,本文提出了一种基于审判逻辑的裁判文书生成方法,将文书结构划分为“基本案情”“审理事实”和“法院判决”3 个部分,通过模拟审判流程将任务分解为多阶段子任务。针对各部分文本的特性,设计不同的生成策略,最终拼接生成结构完整、逻辑严谨的裁判文书。其中,由于大模型缺乏对司法领域专业知识的学习,在“法律适用”部分生成的内容往往准确性不足,因此本文引入 RAG(Retrieval-Augmented Generation)<sup>[6]</sup>策略,通过检索外部法条知识库来提高文书的法律适用准确性。经真实法院案件数据验证,所提方法在裁判文书生成任务中优于基线方法。

## 1 相关工作

在早期研究中,文本生成任务主要依赖规则驱动的方法和统计模型来实现,具有代表性的技术包括经典的 N-gram<sup>[7]</sup>模型、TF-IDF<sup>[8]</sup>以及基于短语结构语法<sup>[9]</sup>(Phrase Structure Grammar, PSG)等生成方法。这些方法虽然在特定任务中取得了一定成效,但存在泛化能力差、数据稀疏性强及语义理解不足等问题。这促使研究者转向更先进的生成方法,以突破传统技术的瓶颈。

随着深度学习技术的蓬勃发展,各种深度学习模型被广泛应用,如变分自编码器<sup>[10]</sup>、扩散模型<sup>[11]</sup>等。其中最成功的便是基于编码器-解码器的架构,该框架将输入序列编码为固定长度的语义表示,再解码生成目标序列,为文本生成任务提供了统一的解决方案。Wang 等<sup>[12]</sup>根据用户意图规划子主题,使用改进的循环神经网络(Recurrent Neural Network, RNN)逐行生成诗歌。Rahman 等<sup>[13]</sup>在长短期记忆网络(Long Short-Term Memory, LSTM)中引入了多层注意力机制,用于从长文本中抽取摘要。Yang 等<sup>[14]</sup>通过引入外部常识性知识,帮助模型生成更新颖和多样化的论文。以上方法在文本生成任务中取得了显著的成果。然而,这些模型在实际应用中的性能提升依赖于大规模标注数据和昂贵的计算资源。此外,在处理复杂句法结构和长距离语义关联时,模型的表现仍不尽如人意。

近几年,预训练语言模型逐渐成为文本生成领域的新范式。通过大规模文本语料的预训练,使这类模型能够有效捕捉深层次的语言规律和丰富的语义信息。基于此,众多学者针对不同应用场景对其进行创新改进。Pei 等<sup>[15]</sup>通过在预训练语言模型中引入知识编码机制,特别选取司法领域的专业知识进行特征增强,有效解决了司法文本摘要中罪行不匹配的关键问题;Alokla 等<sup>[16]</sup>则通过对 Bart 模型进行针对性微调,成功实现了从源代码到伪代码的自动生成;Abadi 等<sup>[17]</sup>在 T5 模型上进行三阶段微调,以适应波斯语新闻文本的特点,随后引入强化学习进一步优化模型性能。使用预训练语言模型不仅减小了对数据的依赖,还显著提升了模型在下游任务中的表现。

随着任务复杂度的提升,研究者们开始关注规模更大、能力更强的大语言模型。近期研究揭示了大语言模型在文本生成中的巨大潜力。例如,Ali 等<sup>[18]</sup>使用 ChatGPT 通过有限的临床输入来生成临床信件并对其进行评估,结果显示 ChatGPT 在生成高质量临床信件方面具有潜力;Wang 等<sup>[19]</sup>利用大语言模型将内容大纲转化为具体的角色行为,实现作者、角色与玩家共同创作动态叙事;Li 等<sup>[20]</sup>将检测到的解剖结构转化为文本提示,通过大语言模型生成结构化的胸片报告。得益于对海量通用知识的学习,大语言模型凭借其强大的语言理解与生成能力,在多样化任务中展现了卓越的通用性。然而,在面对裁判文书生成等垂直领域任务时,大语言模型的表现却受到明显制约。这种局限性主要体现在以下两个方面:

- 1)大语言模型在司法审判认知上存在盲区,难以准确理解案件审理的具体环节和运作机制;
- 2)由于缺乏对法律领域专业知识的深入理解,大语言模型往往难以准确把握法律文书的组织方式和表达规范。

## 2 本文方法

### 2.1 任务定义

本文任务输入为案件的卷宗材料  $D = \{d_i | i = 1, 2, \dots, n\}$ ,其中  $d_i$  代表了案件的不同组成部分(如起诉书、证据清单等),共同构成案件的完整记录。大模型需要依据  $D$  中的信息自动生成符合法律要求的裁判文书。

2.2 方法框架

本文方法框架如图1所示。首先,通过正则与大语言模型从给定的案件材料D中抽取关键法律要素以筛除冗余信息,得到的部分关键要素会填充到预定义模板中,生成结构化的基本案情。随后,大模型根据提取的关键信息,结合证据

材料、庭审记录及各方陈述进行合法性和关联性分析,整理形成完整的事实认定。法院判决部分则依据上一步得到的审理事实,检索外部法条知识库中的适用法条,并将检索到的法律法规融入到最终的法律判决。最终,拼接各阶段子文本输出一篇完整的裁判文书。

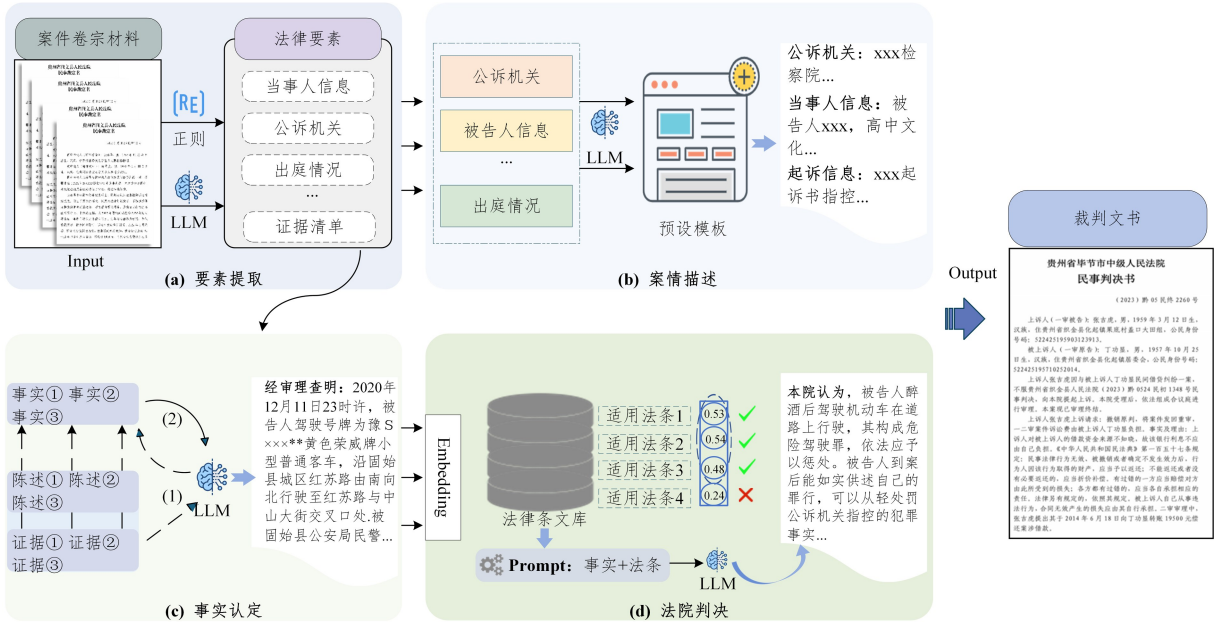


图1 基于审判逻辑的裁判文书生成框架示意

Fig. 1 Framework of judgment document generation based on trial logic

2.3 基于规则与大语言模型的法律要素抽取器

司法审判中,从非结构化的案件材料中精准提取案件核心要素能为后续步骤提供基础支持。由于案件材料具有明确的结构特点,各要素以特定字段开头,因此大部分核心要素可通过规则高效提取。所用的部分正则表达式如下:

Regex-1  $r$ “公诉机关(.+?)[,|\s]”

其中,(.+?)表示非贪婪匹配任意字符(除换行外),圆括号形成捕获组,[,|\s]表示匹配句点、或空白字符\s的字符集。整体匹配模式为:从“公诉机关”开始,截取到第一个出现的句点或空白字符前的内容。

对于非结构化要素,因其以自由文本形式存在,难以通过规则直接提取,需结合大语言模型进行语义理解和深度解析,以补充规则方法的不足,实现全面准确的要素抽取。所采用的部分提示词如下:

Prompt-1 任务:从起诉书中提取出指控事实,按原文输出。

起诉书:{固检一部刑诉[2020]xxx号...}

抽取的法律要素最终会存放到一个 json 字典中,以便后续处理和生成。

2.4 基于模板的基本案情生成

基本案情是裁判文书的基础部分,用于交代案件的基本背景。其文本具有固定性和规范性,易于标准化处理。本文依据裁判文书的编纂规范,总结出基本案情的通用格式如下:

模板:  
公诉机关:[公诉机关名称]  
被告人:[被告人的基本信息]

案件背景:[案件背景、涉嫌罪名以及事件发生时间]

起诉信息:[起诉书编号、起诉时间和指控罪名]

审理情况:[包括出庭情况与案件审理过程]

通过大语言模型将法律要素  $E = \{e_1, e_2, \dots, e_n\}$  填充到对应的固定字段  $F = \{f_1, f_2, \dots, f_n\}$  中,生成标准化的基本案情文本  $T$ ,即:

$$T = \bigcup_{i=1}^n LLM(e_i \rightarrow f_i) \quad (1)$$

2.5 证据驱动的审理事实生成

审理事实是裁判文书中对案件事实进行查明和描述的部分,直接影响法律适用和裁判结果。其文本逻辑严密,同时需要基于证据逐一核实并形成完整认定。本文设计的提示词如下:

Prompt-2 任务:基于提供的指控事实,从证据材料中选取相关证据进行推理分析,最终输出与证据直接相关且逻辑一致的查明事实。

指控事实 1:{血样乙醇含量 210.1mg/100ml...}  
证据材料:{血液酒精含量鉴定报告...}

注意:如果与证据相互印证则直接输出;如果无证据支持与证据矛盾,则输出该事实与证据不符或无证据证明。不要生成其他内容。

Prompt-3 任务:根据提供的起诉书与查明事实,整合生成一段审理事实段落,不要分点。

起诉书:{恒检一部刑诉[2020]xxx号...}  
查明事实:{被告人将一块眉毛肉盗走;被告将两块精瘦肉盗走...}  
证据:{证人证言;视频资料...}

请按照以下要求组织你的回答:

- 1)“经审理查明”,作为段落开头。
- 2)明确指出事件发生的时间和地点。
- 3)尽量不要更改查明事实。
- 4)按时间顺序描述各方的行为过程。
- 5)“上述事实由 xxx 证据证明,足以认定”作为结尾。

生成过程分为两个阶段:事实验证和事实整合。第一阶段结合证据验证主张逐一核实每一条指控事实,第二阶段将各查明事实进行整合,生成完整段落。模型以时间、地点、行为顺序为框架生成连贯内容。

## 2.6 结合法律知识库的法院判决生成

法院判决是基于审理事实结合适用法律作出明确的判决结果。然而,大语言模型因缺乏精准的法律知识,常出现法律适用错误和事实与结论脱节的问题。为此,本文构建了一个外部的法律法规知识库,使大模型能够结合案件事实准确引用法律条文,生成逻辑严密且符合法律规范的法院判决。

本文采用北京智源研究院发布的通用语义向量模型 BGE(BAAI General Embedding)<sup>[21]</sup>。该模型基于对比学习框架,在大规模语料上进行预训练,能够将文本语义高效地映射为低维稠密向量空间中的分布式表示。BGE 支持中英双语文本嵌入,在语义相似度计算和信息检索任务中表现出色。

本文通过构建包含案件事实与适用法条专用数据集对 BGE 模型进行微调,以学习案件事实与法律条文之间的语义关联。数据为三元组形式,即  $D = \{(q_i, p_i^+, \{n_{i,j}^-\}_{j=1}^m)\}_{i=1}^N$ , 其中,  $q_i$  表示第  $i$  个案件事实描述,  $p_i^+$  表示与  $q_i$  相关的法律条文,  $\{n_{i,j}^-\}_{j=1}^m$  表示与  $q_i$  不相关的  $m$  条法律条文。微调后的 BGE 模型将法律法规知识库中的每一条法律条文及案件事实分别转化为高维向量表示:

$$\mathbf{V}_{F_i} = \text{BGE}(F_i), \mathbf{V}_{L_j} = \text{BGE}(L_j) \quad (2)$$

其中,  $\mathbf{V}_{F_i}$  和  $\mathbf{V}_{L_j}$  分别是案件事实  $F_i$  和法律条文  $L_j$  的向量表示。通过余弦相似度来计算案件事实向量与法律条文向量之间的相关性。

$$\text{sim}(F_i, L_j) = \frac{\mathbf{V}_{F_i} \cdot \mathbf{V}_{L_j}}{\|\mathbf{V}_{F_i}\| \|\mathbf{V}_{L_j}\|} \quad (3)$$

根据得分筛选出与案件事实最相关的  $k$  条法律条文,与案件事实共同输入大模型以生成最终的判决结果。所使用的提示词如下:

Prompt-4 任务:法院判决文本中记录了法院认定的事实、公诉机关的指控与法院态度、适用的法律依据和判决结果。现在请你根据提供的审理事实与参考法条生成对应的法院判决文本,格式参考法院判决示例。

法院判决示例 1: {本院认为...}

法院判决示例 2: {本院认为...}

法院判决示例 3: {本院认为...}

审理事实: {经审理查明...}

参考法条: {《刑法》第一百四十一条...}

## 3 实验与结果分析

本文实验中各部分文本均以接口调用的方式使用 OpenAI 的大语言模型 GPT-4o-mini 生成。该模型擅长处理长

文本,支持灵活的提示设计,适用于裁判文书自动化生成研究。

### 3.1 数据集

本文使用的数据集来源于贵州省高级人民法院提供的真实刑事案件卷宗,涵盖盗窃、抢劫、故意杀人等多种案件类型。该数据集共包含 1916 条案件记录,每条数据包括证据材料、裁判文书、庭审笔录等完整内容。如图 2 所示,裁判文书文本长度分布广泛,最长为 11533 字,最短为 757 字,其中大部分长度均超过 3000 字。

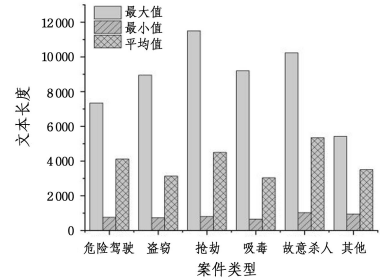


图 2 裁判文书文本长度统计

Fig. 2 Text length statistics of judgment documents

为确保模型训练与评估的科学性,将数据集按照 80%, 10%, 10% 的比例划分为训练集、验证集和测试集。

### 3.2 评价方法

本文采用 ROUGE<sup>[22]</sup> 指标中的 F1 值作为核心评价方式,以综合衡量生成文书在词汇级别 (ROUGE-1)、短语级别 (ROUGE-2) 及全局结构 (ROUGE-L) 上的生成质量。ROUGE-N 通过统计生成文本与参考文本之间重叠的  $N$  元词组比例来衡量匹配程度,其定义为:

$$\text{ROUGE-N} = \frac{C_{\text{overlap}}}{C_{\text{ref}}} \quad (4)$$

其中,  $C_{\text{overlap}}$  表示生成文本与参考文本之间重叠的  $N$ -gram 数量,  $C_{\text{ref}}$  表示参考文本中的  $N$ -gram 总数, ROUGE-1 和 ROUGE-2 分别对应 Unigram (单个词) 和 Bigram (两个词的组合)。对于全局结构的匹配, ROUGE-L 基于最长公共子序列 (LCS) 进行计算,其定义为:

$$\text{ROUGE-L} = \frac{\text{LCS}(S_{\text{gen}}, S_{\text{ref}})}{|S_{\text{ref}}|} \quad (5)$$

其中,  $\text{LCS}(S_{\text{gen}}, S_{\text{ref}})$  表示生成文本  $S_{\text{gen}}$  与参考文本  $S_{\text{ref}}$  之间的最长公共子序列长度,  $|S_{\text{ref}}|$  表示参考文本的长度。F1 值作为精确率  $P$  和召回率  $R$  的调和平均数,用于综合评估生成文本的准确性和完整性,其表达式为:

$$F1 = \frac{2PR}{P+R} \quad (6)$$

鉴于自动化指标无法评估逻辑严谨性和内容合理性,本文进一步引入了人工评测,由领域专家从准确性、逻辑性、流畅性、完整性、可读性 5 个维度对生成文书进行综合评价。

1) 确保法律条款引用和事实陈述无误,信息准确可靠。

2) 评估论述过程是否符合逻辑,论证是否严密。

3) 审查语言表达是否自然流畅,句式结构是否得当。

4) 确认文书涵盖所有必要的法律要素和事实背景,结构完整。

5) 评估内容划分是否合理且层次分明,便于快速浏览和理解。

### 3.3 实验结果及分析

#### 3.3.1 方法对比

本文以 Few-shot<sup>[23]</sup> 提词法(Few-shot Learning)构建基线模型(Baseline)。此外,还选取其他常见的生成模型进行对比分析。表2列出了所提方法与其他对比模型在数据集上的性能对比结果。结果显示,本文方法在所有指标上均取得了最优性能。尤其是,与Baseline相比,在F-1,F-2和F-L指标上分别提升了6.03个百分点、6.56个百分点和7.98个百分点,验证了本文方法的有效性。

表2 本文方法和对比模型的 ROUGE 指标 F1 值

Table 2 ROUGE F1 scores of the proposed method and the comparison models

模型	F-1	F-2	F-L
TextRank <sup>[24]</sup>	61.58	34.64	29.06
TF-IDF	62.33	26.98	13.11
Seq2Seq+Atten <sup>[25]</sup>	54.53	30.19	18.78
Bart <sup>[26]</sup>	73.46	49.56	40.92
CPT <sup>[27]</sup>	73.61	54.13	38.56
Baseline	71.89	50.94	41.96
Ours	<b>77.92</b>	<b>57.50</b>	<b>49.94</b>

在对比方法中,例如 TextRank 等传统技术仅从原文本中抽取信息,难以捕捉词汇间的语义关系及长程依赖。尽管 Seq2Seq 模型相比传统方法有所改进,但其上下文捕捉能力仍然有限,且训练和优化成本较高,因此性能反而不如部分传统方法。此外,即使在 Bart 等预训练模型上进行微调,模型也仅能掌握裁判文书的基本规范和语言风格,而无法深入理解案件信息之间的因果关系及法律推理的复杂性。相比之下,本文通过引入审判逻辑,显著增强了模型对案件事实、法律条文及推理过程的理解能力,同时提供了更精确的结构信息和因果逻辑信息,从而提升了生成文书的质量。

#### 3.3.2 模块验证

本文采用了一种任务分解+模块化生成的框架,针对裁判文书每个部分的特点设计了不同的生成方法。为验证每个模块的有效性,分别将基线方法与所提方法对裁判文书各部分的生成效果进行比较。表3列出了本文方法和基线模型的 ROUGE 指标 F1 值。

表3 本文方法和基线模型的 ROUGE 指标 F1 值

Table 3 ROUGE F1 scores of the proposed method and the baseline model

文本	方法	F-1	F-2	F-3
基本案情	Baseline	74.13	64.70	63.48
	Ours	<b>85.31</b>	<b>77.59</b>	<b>77.18</b>
审理事实	Baseline	65.78	39.37	35.09
	Ours	<b>67.81</b>	<b>43.73</b>	<b>38.05</b>
法院判决	Baseline	60.38	40.71	33.27
	Ours	<b>61.56</b>	<b>42.25</b>	<b>35.99</b>

#### 3.3.3 泛化性研究

本文还在智谱 AI 开源的 ChatGLM3-6B<sup>[28]</sup> 模型和 Meta 提供的 LLaMa3-8B-Instruct<sup>[29]</sup> 模型上进行了实验,以验证所提方法在裁判文书生成任务上的泛化性。实验结果如表4所列,可以看出在不同模型上,所提方法的 F1 值相较于基线模

型均有所提升,证明引入审判逻辑在文书生成性能上产生了积极的效果。

表4 两个模型下本文方法和基线模型的 ROUGE 指标 F1 值

Table 4 ROUGE F1 scores of the proposed method and the baseline model under two models

模型	方法	F-1	F-2	F-L
ChatGLM3-6B	Baseline	69.08	39.83	28.40
	Ours	<b>73.04</b>	<b>45.12</b>	<b>34.91</b>
LLaMa3-8B-Instruct	Baseline	74.60	47.23	35.56
	Ours	<b>75.67</b>	<b>52.66</b>	<b>39.95</b>

#### 3.3.4 人工评测

为全面评估生成的裁判文书是否符合人类语言表达习惯,从生成的文书中随机抽取 100 个样本,对其进行匿名化处理后再进行人工评测。评估团队由 5 名丰富经验的律师和 8 名法学研究生组成,他们从准确性、逻辑性、流畅性、完整性和可读性 5 个维度对每篇文书逐一进行审阅。评估人员从 5 个维度上对本文模型、Baseline 以及 T5<sup>[30]</sup> 模型生成的文书进行评分。单项满分为 10 分,其中 9~10 分表示优秀,6~8 分表示良好,3~5 分表示合格,0~2 分表示不合格。样本的最终得分分为各人员评分的简单平均值。

实验结果如图3所示,可以看出,本文模型在各个方面均超越了对比模型,该模型在生成高质量裁判文书方面具有显著优势。

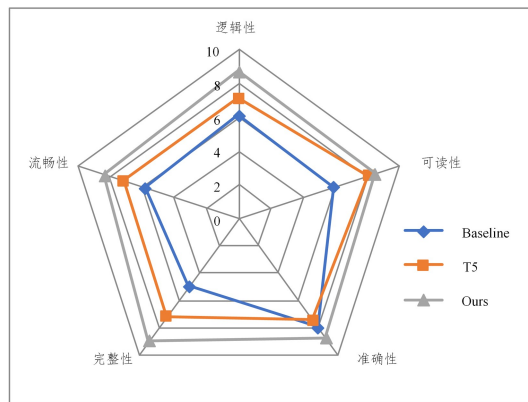


图3 人工评测结果

Fig. 3 Manual evaluation results

#### 3.3.5 案例研究

为了更直观地感受本文方法的文本生成效果,表5中列出了本文方法生成文书与标准文书示例的对比。可以看到,生成文书在结构、内容和语言表达上与标准文书高度一致,能够有效模拟真实文书的风格和格式。具体来说,生成文书能够准确、详细地记录事故发生的时间、地点、经过及被告人的行为等关键信息。此外,文书的整体结构清晰、逻辑严谨,按照案件事实、审理过程、判决结果等顺序进行叙述。

然而,从表5中下划线标识的文本可以看出,生成文本会遗漏一些案件信息,例如生成样例中辩护人的出庭情况以及辩护情况均未提及,这种遗漏主要是大模型的注意力分散所导致的。此外,在法律知识库中检索得到的结果不够精准或

全面,这是因为简单的匹配算法难以完全捕捉法条与案件事实之间的深层次关系。在未来的研究中,将从以下两个方面进行改进:1)引入后处理校验机制以筛查模型遗漏信息;

2)结合规则推理或图神经网络等方法,以提升法条检索的准确性。综上所述,生成文书已经具备较高的基础质量,但在细节和深度上仍有提升空间。

表5 裁判文书生成结果对比示例(部分有删减)

Table 5 Example of comparison of results generated by judgment documents

基本案情	审理事实	法院判决	基本案情
生成文书	<p>【…】公诉机关贵州省麻江县人民检察院以麻检刑诉 xxx 号起诉书指控被告人陈某犯过失致人死亡罪,向本院提起公诉。本院适用简易程序,实行独任审判,公开开庭审理了本案。贵州省麻江县人民检察院指派检察员出庭支持公诉,被告人陈某到庭参加诉讼。本案现已审理终结。</p>	<p>【…】经鉴定,被告人陈某的血样中乙醇含量为 127.76 mg/100 ml。事发后,被告人陈某明知他人拨打 110 报警电话仍在现场等候民警处理。案发后,被告人陈某已赔偿被害人家属,并取得谅解。可以认定被告人陈某的行为已构成过失致人死亡罪【…】</p>	<p>【…】依据《中华人民共和国刑法》第二百三十三条、第六十七条、第七十三条、第一百三十三条条之规定,判决如下:被告人陈某犯过失致人死亡罪,判处有期徒刑一年零八个月。【…】</p>
参考文书	<p>贵州省麻江县人民检察院以麻检刑诉 xxx 号【…】起诉书指控被告人陈某犯过失致人死亡罪,于 2021 年 1 月 4 日向本院提起公诉并建议本院适用简易程序审理,经审查本院决定适用简易程序,实行独任审判,公开开庭审理了本案。贵州省麻江县人民检察院指派检察员张某某出庭支持公诉,被告人陈某及其辩护人到庭参加诉讼。现已审理终结。</p>	<p>【…】经鉴定,被告人陈某的血样中乙醇含量为 127.76mg/100ml。事发后,被告人陈某明知他人拨打 110 报警电话仍在现场等候民警处理。案发后,被告人陈某已赔偿被害人家属,并取得谅解。据此认为被告人陈某的行为已构成过失致人死亡罪。【…】</p>	<p>【…】依照《中华人民共和国刑法》第二百三十三条、第六十七条第一款、第七十二条第一款、第七十三条第二款、第三款及《中华人民共和国民事诉讼法》第二百九十条之规定,判决如下:被告人陈某犯过失致人死亡罪,判处有期徒刑一年零两个月,缓刑二年。【…】</p>

**结束语** 本文提出了一种基于审判逻辑的裁判文书生成方法,该方法通过模块化设计将基本案情、审理事实和判决生成有机结合,构建了端到端的裁判文书生成系统;通过模拟审判人员的思维过程,依托大语言模型与知识库检索技术,确保生成文书的逻辑性与法律规范性。实验结果表明,本文方法在裁判文书生成任务中表现优异,在文本的准确性、逻辑性和流畅性等方面展现出显著优势。

在未来的工作中,将继续探索更高效的生成方法和技术路径。例如,通过增强模型对复杂案件的推理能力,改进细节生成机制以减少信息遗漏,以及探索更精准的法条引用方法,进一步提升生成文书的质量与适用性。本文作为结合审判逻辑设计裁判文书生成模型的初步尝试,旨在为智能司法辅助提供新的解决思路。希望这项工作不仅能够为本生成领域带来新的启示,还能在法律咨询、智能审判等实际应用场景中发挥更大的价值,推动智能司法技术的进一步发展与落地。

## 参考文献

- [1] WU J. Legal document system reform under the background of intelligent justice [J]. South China Sea Jurisprudence, 2020, 4(3): 1-5.
- [2] Supreme People's Court. Supreme People's Court Re-releases Key Data on Judicial Adjudication Work for 2024 [EB/OL]. (2024-01-26) [2025-03-01]. <https://www.chinacourt.org/article/detail/2025/01/id/8686406.shtml>.
- [3] SUTSKEVER I, VINYALS O, LE Q V. Sequence to Sequence Learning with Neural Networks [C]// Proceedings of the 28th International Conference on Neural Information Processing Systems(NIPS'14). MIT Press, 2014: 3104-3112.
- [4] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.
- [5] DUAN M Q. Basic Trial Procedure for First-Instance Ordinary Criminal Cases [EB/OL]. (2015-11-02) [2025-03-01]. <https://www.66law.cn/domainblog/74744.aspx>.

- [6] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks[J]. Advances in Neural Information Processing Systems, 2020, 33: 9459-9474.
- [7] DE NOVAIS E M, DIAS T T, PARABONI I. Improved Text Generation Using N-gram Statistics [C]// 12th Ibero-American Conference on Artificial Intelligence (IBERAMIA). Springer-Verlag, 2010: 316-325.
- [8] CHRISTIAN H, AGUS M P, SUHARTONO D. Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF) [J]. ComTech: Computer, Mathematics & Engineering Applications, 2016, 7(4): 285-294.
- [9] PROUDIAN D, POLLARD C. Parsing Head-Driven Phrase Structure Grammar [C]// Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics. 1985: 167-171.
- [10] HU H J, LIAO M F, MAO W M, et al. Variational Auto-encoder for Text Generation [C]// Proceedings of the 5th IEEE International Conference on Information Technology and Mechatronics Engineering(ITOEC). 2020: 595-598.
- [11] LIN Z H, GONG Y Y, SHEN Y L, et al. Text Generation with Diffusion Language Models: A Pre-training Approach with Continuous Paragraph Denoise [C]// International Conference on Machine Learning. 2023: 21051-21064.
- [12] WANG Z, HE W, WU H, et al. Chinese Poetry Generation with Planning Based Neural Network [C]// Proceedings of the 26th International Conference on Computational Linguistics (COLING). 2016: 1051-1060.
- [13] RAHMAN M M, SIDDIQUI F H. Multi-layered attentional peephole convolutional LSTM for abstractive text summarization[J]. Etri Journal, 2021, 43(2): 288-298.
- [14] YANG P C, LI L, LUO F L, et al. Enhancing Topic-to-Essay Generation with External Commonsense Knowledge [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019: 2002-2012.

- [15] PEI B S, L X, HU K Q, et al. Generation of judicial text abstracts based on knowledge enhancement pre-training model [J]. *Science Technology and Engineering*, 2024, 24(20): 8587-8597.
- [16] ALOKLA A, GAD W, NAZIH W, et al. Pseudocode generation from source code using the bart model [J]. *Mathematics*, 2022, 10(21): 3967.
- [17] ABADI V N M, GHASEMIAN F. Enhancing Persian text summarization through a three-phase fine-tuning and reinforcement learning approach with the mT5 transformer model [J]. *Scientific Reports*, 2025, 15(1): 80.
- [18] ALI S R, DOBBS T D, HUTCHINGS H A, et al. Using ChatGPT to write patient clinic letters [J]. *The Lancet Digital Health*, 2023, 5(4): e179-e181.
- [19] WANG Y, ZHOU Q, LEDO D. StoryVerse: Towards Co-authoring Dynamic Plot with LLM-based Character Simulation via Narrative Planning [C] // *Proceedings of the 19th International Conference on the Foundations of Digital Games*. 2024: 1-4.
- [20] LI H Z, WANG H Y, SUN X, et al. Prompt-guided Generation of Structured Chest X-ray Report Using a Pre-trained LLM [C] // *Proceedings of the IEEE International Conference on Multimedia and Expo(ICME)*. 2024: 1-6.
- [21] XIAO S T, LIU Z, ZHANG P T, et al. C-Pack: Packed Resources for General Chinese Embeddings [C] // *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2024: 641-649.
- [22] LIN C. Rouge: A Package for Automatic Evaluation of Summaries [C] // *Proceedings of the Workshop on Text Summarization Branches Out*. 2004: 74-81.
- [23] MIN S, LYU X, HOLTZMAN A, et al. Rethinking the Role of Demonstrations: What Makes In-Context Learning Work? [C] // *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 2022: 11048-11064.
- [24] MIHALCEA R, TARAU P. TextRank: Bringing Order into Text [C] // *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. 2004: 404-411.
- [25] SHI Y S, MENG J, WANG J. Seq2Seq Model with RNN Attention for Abstractive Summarization [C] // *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*. 2019: 348-353.
- [26] LEWIS M, LIU Y, GOYAL N, et al. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension [C] // *58th Annual Meeting of the Association for Computational Linguistics*. 2020: 7871-7880.
- [27] SHAO Y F, GENG Z C, LIU Y T, et al. Cpt: A pre-trained unbalanced transformer for both chinese language understanding and generation [J]. *Science China Information Sciences*, 2024, 67(5): 152102.
- [28] DU Z X, QIAN Y J, LIU X, et al. GLM: General Language Model Pre-training with Autoregressive Blank Infilling [C] // *60th Annual Meeting of the Association for Computational Linguistics*. 2022: 320-335.
- [29] GRATTAFIORI A, DUBEY A, JAUHRI A, et al. The llama 3 herd of models [J]. *arXiv:2407.21783*, 2024.
- [30] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [J]. *Journal of Machine Learning Research*, 2020, 21(1): 5485-5551.



**LIAO Jinchao**, born in 1999, postgraduate. His main research interest is natural language processing.



**QIN Yongbin**, born in 1980, Ph.D, professor. His main research interests include big data governance and application, and multi-source data fusion.

(责任编辑:柯颖)