



计算机科学

COMPUTER SCIENCE

基于大语言模型的刑事案件智能判决研究

丛颖男, 韩林睿, 马佳羽, 朱金清

引用本文

丛颖男, 韩林睿, 马佳羽, 朱金清. 基于大语言模型的刑事案件智能判决研究[J]. 计算机科学, 2025, 52(5): 248-259.

CONG Yingnan, HAN Linrui, MA Jiayu, ZHU Jinqing. [Research on Intelligent Judgment of Criminal Cases Based on Large Language Models](#) [J]. Computer Science, 2025, 52(5): 248-259.

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)

[基于RISC-V Matrix指令集扩展的LLM向量点积加速研究](#)

Research on LLM Vector Dot Product Acceleration Based on RISC-V Matrix Instruction Set Extension
计算机科学, 2025, 52(5): 83-90. <https://doi.org/10.11896/jsjcx.241200074>

[基于大语言模型自身的提示语公平性自动优化与评估](#)

Automatic Optimization and Evaluation of Prompt Fairness Based on Large Language Model Itself
计算机科学, 2025, 52(4): 240-248. <https://doi.org/10.11896/jsjcx.240900008>

[元宇宙中三维场景重建技术综述](#)

Survey on 3D Scene Reconstruction Techniques in Metaverse
计算机科学, 2025, 52(3): 17-32. <https://doi.org/10.11896/jsjcx.241000043>

[基于大小语言模型协同增强的中文电子病历依存句法分析](#)

Dependency Parsing for Chinese Electronic Medical Record Enhanced by Dual-scale Collaboration of Large and Small Language Models
计算机科学, 2025, 52(2): 253-260. <https://doi.org/10.11896/jsjcx.231200054>

[大语言模型驱动的多元关系知识图谱补全方法](#)

Large Language Model Driven Multi-relational Knowledge Graph Completion Method
计算机科学, 2025, 52(1): 94-101. <https://doi.org/10.11896/jsjcx.240600170>

基于大语言模型的刑事案件智能判决研究

丛颖男¹ 韩林睿^{2,3} 马佳羽⁴ 朱金清⁵

1 中国政法大学商学院 北京 100088

2 教育部哲学社会科学实验室——中国政法大学数据法治实验室 北京 100088

3 中国政法大学数据法治研究院 北京 100088

4 清华大学法学院 北京 100084

5 北京字节跳动网络技术有限公司 北京 100043

(cyn_2010@163.com)

摘要 刑事案件判决的智能化一直是数字法院建设中的研究热点。传统方法基于自然语言处理技术,由模型依据案件事实直接预测判决结果,但应对复杂刑事案件案情时,模型难以发现法律要件之间的逻辑依赖关系,也难以清晰表达法律推理过程。文中提出一种基于大语言模型的刑事案件智能判决方法,该方法以“标记案件语料—预训练大模型—强化判决逻辑”为思路,首先通过自动化标注与人工校正相结合的方式,标注案情中的主体、客体、主观要件和客观要件等法律要素,构建结构化的推理数据集;其次基于 GLM 预训练框架,选取 ChatGLM3-6b-32k 作为基座大语言模型进行增量预训练;最后采用 LoRA 参数高效微调策略与大模型检索增强技术对模型进行参数调优与法律知识扩展,实现判决逻辑的强化。实验结果表明,与 Qwen-7B-Chat 和 Baichuan2-7B-Chat 相比,ChatGLM3-6b-32k 模型在指令监督微调后性能更优。引入司法三段论显著增强了判决文本的逻辑性,使其更贴近人类法官的裁判说理。在罪名预测和刑期预测任务中,所提模型准确率相较于 MTL-Fusion, Lawformer 和 BERT 模型均有显著提升。此外,与基于欧美法律文本训练的 Legal-BERT 和 CaseLawBERT 相比,所提模型更适应中国刑事案件的判决逻辑,在处理长文本任务上展现出更强的能力。该研究不仅探索了大语言模型在刑事案件智能判决中的应用,还为司法领域大模型研究的范式提供了有益参考。

关键词: 数字法院;法律判决预测;司法三段论;大语言模型;参数高效微调

中图分类号 TP183

Research on Intelligent Judgment of Criminal Cases Based on Large Language Models

CONG Yingnan¹, HAN Linrui^{2,3}, MA Jiayu⁴ and ZHU Jinqing⁵

1 Business School, China University of Political Science and Law, Beijing 100088, China

2 Ministry of Education Laboratory of Philosophy and Social Sciences—The Data Law Lab, China University of Political Science and Law, Beijing 100088, China

3 The Institute for Data Law, China University of Political Science and Law, Beijing 100088, China

4 School of Law, Tsinghua University, Beijing 100084, China

5 Beijing Bytedance Network Technology Co. Ltd., Beijing 100043, China

Abstract The intelligentization of criminal case trials has been a hot research topic in the development of digital courts. In the conventional method based on natural language processing, the model directly predicts the final judgment based on the facts of the case. However, when dealing with complex criminal cases, the model may fail to identify the logical dependencies between legal elements and to clearly present the legal reasoning process. The intelligent criminal case trial method based on large language models proposed in this paper follows the approach of “annotating case corpus-pre-training large language model-reinforcing trial logic”. The first step is to annotate the legal elements of the case such as subjects, objects, subjective elements, and objective elements by combining automated annotating with manual correction and create a structured reasoning dataset. The second step is to

到稿日期:2024-11-18 返修日期:2025-03-06

基金项目:2025 年中国政法大学青年教师学术创新团队支持计划(25CXTD04);2022 年国家重点研发计划“社会治理与智慧社会科技支撑”重点专项(2022YFC3303000);教育部人文社会科学研究一般项目(22YJC190003)

This work was supported by the Program for Young Innovative Research Team in China University of Political Science and Law(25CXTD04), 2022 National Key R&D Program of China“Social Governance and Smart Society Technology Support” Key Special Project(2022YFC3303000) and General Project of Humanities and Social Sciences Research of the Ministry of Education(22YJC190003).

通信作者:朱金清(zhujinqing@bytedance.com)

use ChatGLM3-6b-32k as the foundational large language model for incremental pre-training based on the GLM pre-training framework. The last step is to fine-tune the parameter and increase legal knowledge using the LoRA parameter-efficient fine-tuning strategy and large language model retrieval enhancement technology, thereby reinforcing the trial logic. Experimental results indicate that, compared to Qwen-7B-Chat and Baichuan2-7B-Chat, the ChatGLM3-6b-32k model exhibits superior performance after supervised fine-tuning. The introduction of judicial syllogism significantly enhances the logicity of the judgment texts, making them closer to the reasoning of human judges. In the tasks of charge prediction and sentencing prediction, the model created using this method shows a significant improvement in accuracy compared to the MTL-Fusion, Lawformer, and BERT models. In addition, compared to Legal-BERT and CaseLawBERT, which are trained on European and American legal texts, the ChatGLM3-6b-32k model better suits the trial logic of Chinese criminal cases and demonstrates stronger capabilities in handling long texts. This paper not only explores the application of large language models in intelligent criminal case trials, but also provides valuable references for research on large language models in justice.

Keywords Digital court, Legal judgement prediction, Judicial syllogism, Large language model, Parameter-efficient fine-tuning

1 引言

智慧司法一直是人工智能应用的重要场景。从国务院2017年印发的《新一代人工智能发展规划》将智慧法庭建设作为“推进社会治理智能化”的重要举措,指出要“促进人工智能在证据收集、案例分析、法律文件阅读与分析中的应用,实现法院审判体系和审判能力智能化”,到最高人民法院2022年印发《最高人民法院关于规范和加强人工智能司法应用的意见》,为实现2025年“基本建成较为完备的司法人工智能技术应用体系”和2030年“建成具有规则引领和应用示范效应的司法人工智能技术应用和理论体系”的“两步走”目标专门制定了工作意见,我国始终高度重视推进人工智能技术与审判执行、诉讼服务、司法管理和服社会治等工作的深度融合。

法律判决预测(Legal Judgement Prediction, LJP)是利用人工智能技术和大数据分析对法律案件的可能判决结果进行预测的过程,数据集及子任务多集中于刑事案件领域(如罪名、刑期、罚金等)^[1]。在预测过程中,众多研究聚焦于根据给定的法律事实,通过分类、回归、神经网络等算法,预测形式相对确定的结果。然而,由于算法的“黑箱”属性,具体的预测过程并不能被很好地诠释,其法律信任程度存在高敏感性与不确定性。基于此,一部分学者探索文本生成任务^[2-3],尝试可视化LJP的推理过程,但受限于生成模型的能力,生成的文本并不能完整展现并支持法律判决的逻辑^[4]。本研究认为,这种局限性出现的原因有两点可能:一是缺乏高质量的数据,如同Silver所言,分开“噪声”和“信号”并不容易^[5];二是模型本身能力不足,数理统计学的方法并不能很好地解释法律领域的复杂逻辑,预训练模型也因领域知识匮乏而无法揭示垂直领域的深层次规律。

因此,本研究针对上述局限进行了完善,旨在提升刑事案件判决的智能化水平,进一步提高数字时代的办案质效,保障司法公正。在数据方面,考虑到传统LJP采用“事实-判决”的单向化数据,本研究参考Deng等基于司法三段论构建的SLJA(Syllogistic Legal Judgement Analysis)^[6]数据集,并以中国裁判文书网、北大法宝中公开的刑事案件裁判文书作为

补充,通过引导大语言模型(Large Language Model, LLM)按照“法条检索—事实映射—主观要素判断—客观要素生成”的路径标注裁判文书中的主体、客体、主观及客观要件等案件推理的关键要素,构建SLJ(Structured Legal Judgements)数据集。

在模型方面,本研究通过增量预训练与指令微调大语言模型的方式,将模型的通用能力扩展应用于法律垂直领域,同时以大模型检索增强的方式扩充法律法规知识库与司法裁判规则,以期缓解大模型的幻觉、遗忘问题,优化其迁移学习与领域适应的效果,进而更高效、准确地完成罪名与刑期的预测以及判决文本的生成,实现刑事案件的智能判决。

2 相关工作

LJP任务通常由经过多年法律学习或专业培训的法律从业者完成,包括分析案件事实,查找相关法律条文或案例,确定指控范围和刑期等步骤。若在纯人工的背景下,该任务的每一部分都十分耗时,且需要掌握极为专业的法律知识和丰富的从业经验,这给时间或能力有限的人带来了巨大负担。早期的LJP工作,由于数据规模相对较小且计算能力有限,主要依靠统计学方法分析和预测法官在非裁判文书信息下的投票^[7]。例如,Kort^[8]采用多因素复合分析法(如罪犯责任能力、犯罪性质等)预测美国最高法院是否会推翻州级法院的裁决,揭示法院在处理“律师辩护权”案件时的一致性和规律性;Nagel等^[9]探讨了如何应用相关性分析来预测判决结果;Lawlor^[10]提出“传统先例原则(Traditional Stare Decisis)”和“个人先例原则(Personal Stare Decisis)”,并通过构建逻辑公式来描述法院的决策行为;Marshall^[11]利用灾难模型(Catastrophic Model)模拟法律决策过程中的复杂动态。该时期,人工智能在法律推理领域的研究聚焦于法律专家系统,尝试结构化法律论证,将法律知识转化成计算机可以理解的形式^[12]。

然而,依赖于规则或统计学方法的LJP系统往往对噪声较为敏感。一方面,基于法律文本的规则制定、特征标注较为复杂,且法律事实的描述常包含冗余的表达、不相关的事实或不完整的数据;另一方面,LJP往往受到多种因素的影响,无

法通过简单的数学模型捕捉,噪声数据可能会扭曲学习过程,从而影响模型的性能^[1]。随着机器学习和文本挖掘技术的不断进步,基于文本分类和案例标注的特征提取技术成为研究的热点。多分类任务中,将案件事实作为输入,判决标签作为输出,推理过程通常依赖专家知识、简单的分类范式和基础的文本分析技术^[13]。有学者利用监督式算法自动学习中文法律文件的标记、案件分类及量刑^[14-16]。进一步地,为了更好地提升分类模型在法律领域的效果,研究者在模型训练中融合大量法律知识(如纳入案件事实和法律条款的相关性评估,手动推导法律属性)。例如,Liu等^[17]借助PAT树模型与HowNet研究了基于短语的中文刑事司法文书分类,在基于起诉类别的分类任务上取得了较以往研究^[18]更好的效果。犯罪风险、犯罪热点的预测借助支持向量机(Support Vector Machine, SVM),以稳定有效的方式对非线性关系进行建模^[19]。Sulea^[20]构建了一种结合多个SVM分类器输出的平均概率集成系统,模型以案情事实描述和时间跨度信息作为输入,能够输出判决结果、法律范围、估算判决日期等信息。Katz^[21]使用随机森林,从案情描述中提取有效特征对美国最高法院的判决结果进行预测。但这些方法尚不能挖掘深层的文本特征,且因人工设计的特性,需要大量的人力成本,无法深入应用到其他领域。

深度学习技术的不断发展,使得深度神经网络在自然语言处理中的应用持续深入,这些模型能捕捉更复杂的特征和模式,为LJP提供新的思路与方法。Zeleznikow等^[22]结合神经网络和基于规则的推理,提出Split-up系统预测澳大利亚家庭法领域财产纠纷的判决结果,其基于Toulmin论证结构的解释生成机制弥补了神经网络解释性缺乏的问题,展示了在家庭法领域开发法律专家系统的可行性。结合长短期记忆网络及其衍生方法,Huang等^[23]针对犯罪态势预测得到了较为理想的结果。Luo等^[24]在参照深度学习文本分类模型的基础上,提出基于注意力机制的刑事罪名预测模型,该模型能够更好地捕捉事实描述与法律适用之间的关系,从而提升预测的准确性。为了进一步克服该模型选用的小样本、易混淆罪名带来的弊端,Hu等^[25]引入了具有区分性的法律属性进行干预。Zhao等^[26]则更深入地利用法律构成要素,自动提取与案件事实相匹配的法律构成要素,在减少人工标注的任务量的同时更好地辅助法律判断。在法律人工智能的不断拓展下,针对LJP子任务(法条预测^[13]、类案预测^[27]、法律问答^[28]等)的研究逐渐成熟,多任务模型构建成为学术界研究的热点,利用多种子任务的共同点和差异,可以实现有用信息在不同任务间的转移,从而提升判决预测的效果。例如,Zhong等^[29]提出的TOPJUDGE框架,不仅考虑单个子任务(如法律适用、罪名预测、刑期预测等),还考虑这些子任务之间的拓扑依赖关系,以有向无环图(Directed Acyclic Graph, DAG)的方式将其形式化,以更好地体现司法逻辑。Yue等^[30]提出的基于情境感知的LJP框架(NeurJudge),利用中间子任务的结果区分描述不同的事实,并利用它们对其他子任务进行预测。除此以外,为了提高法律判决的可解释性,Ye等^[4]提出了带有罪名标签的Seq2Seq模型,以便在解码器中生成与标签条

件相关的法律说理,使其更具区分度和准确性。

随着语料库的不断扩充,预训练模型(Pre-trained Model, PLM)^[31-33]已被证明可以有效地从大规模未标记的语料库中捕获丰富的语言知识,获得了一定的泛化能力,并能较好地适应下游任务。受其启发,人们不断探索如何利用强大的PLM来促进“法律+人工智能”的发展,如使用欧盟和英国法律语料库进行微调的Legal-BERT^[34],在美国判例法文件和合同语料库上进行预训练的CaseLawBERT^[35],基于LongFormer、专门针对中国司法长文本文书进行预训练的Lawformer等^[36]。这些模型尽管都取得了一定的效果,但在与目前流行的LLM的对比中展现出的能力仍然有限。

如今,LLM的应用成为自然语言处理领域的研究热点,其在常识推理方面取得了令人满意的结果^[37]。除了大模型本身性能外,各种提示工程方法(Prompt Engineering)已经证明,LLM能够基于提示词进行正确的响应,进而取得更好的效果。例如,提示“一步一步思考”(think step by step)^[38],或是利用自助法迭代循环来引导自身,产生更复杂的推理方法等^[39]。然而,LLM在更专业的垂直领域上面临明显的困难。以法律领域为例,优化思路主要集中在:1)利用法律提示工程(Legal Prompt Engineering, LPE),以提示词引导、思维链增强(Chain-of-Thought, CoT)的方式提升其生成正确答案的能力^[40-41],例如Jiang等^[42]利用司法三段论强化LLM进行法律推理的逻辑;2)以法律领域数据微调LLM,使其具备一定的“法律素养”,例如Huang等^[43]基于法律数据集微调推出的Lawyer LLaMA模型,相比原始基座模型、中文LLaMA模型,均能给出更优质的法律回答。这些优化方式都取得了一定效果,但也面临一些难题:单独优化提示词方式中Prompt的独有属性意味着它并不能成为“通用设计”,而微调LLM以适应于一个新的领域可能会导致对其一般知识的灾难性遗忘^[44],训练数据集的形式和数量也是微调的难点^[45]。

因此,本研究认为将提示词优化与大模型微调相结合以“取其精华”,是实现刑事案件智能判决的可行路径。该方法能够通过平衡知识保留与适应能力,高效地利用有限的训练数据。此外,结合大模型微调与检索增强技术,可以突破单一技术的局限,为模型带来更高的灵活性和可扩展性。

3 基于大语言预训练模型的刑事案件智能判决方法

3.1 总体流程

基于大语言模型的刑事案件智能判决方法的总体思路为“标记案件语料—预训练大模型—强化判决逻辑”。首先,利用自动化标注与人工校正相结合的方法标注裁判文书案情中的主体、客体、主观及客观要件等案件推理的关键要素,形成指令微调数据集;其次,选取并部署ChatGLM3-6b-32k模型作为基座大语言模型,在多个开源的法律数据集上进行无监督的增量预训练;最后,采用低秩适应(Low-Rank Adaptation, LoRA)参数高效微调策略与大模型检索增强技术(Retrieval-Augmented Generation, RAG)对预训练模型进行指令微调与法律知识扩展,实现司法三段论对判决结果逻辑性与准确性的提升。整体流程如图1所示。

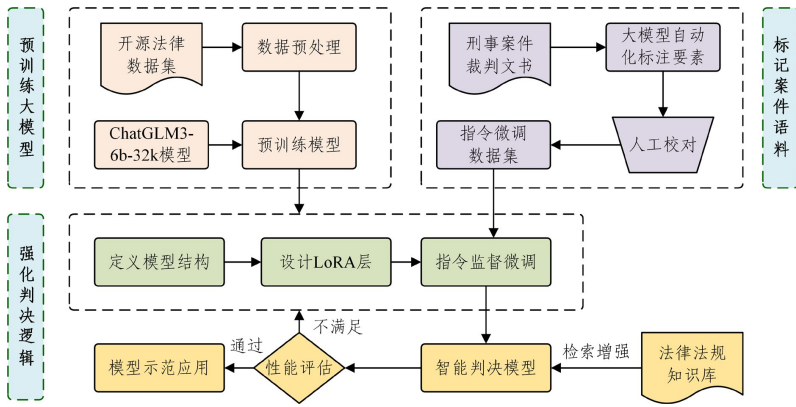


图1 刑事案件智能判决方法流程图

Fig. 1 Flowchart of the intelligent judgment method for criminal case

3.2 GLM 预训练框架

现有的主流预训练框架可分为3类:自回归模型、自编码模型与编码器-解码器模型。自回归模型以 GPT^[45] (Generative Pre-trained Transformer) 为代表,仅使用解码器,侧重于长文本生成任务,但其只具备单向注意力机制,无法捕捉上下文联系,在自然语言理解任务中表现不佳;自编码模型以 BERT^[46] (Bidirectional Encoder Representation from Transformers) 为代表,仅使用编码器,通过掩码语言模型产生双向语言表征,被广泛应用于自然语言理解任务。T5^[47] (Text-To-Text Transfer Transformer) 是典型的编码器-解码器模型,其核心思想是将所有 NLP 任务转化为文本到文本的任务。T5 综合采用编码器的双向注意力和解码器的单向注意力,并在二者之间进行注意力调整,但由于自编码和自回归模型目标本质上不同,简单的统一无法完全继承二者的优势。

上述3种主流的框架虽在长文本生成、自然语言理解等任务上各具优势,但均存在无法全面覆盖所有 NLP 任务类型、参数体量庞大等局限性。通用语言模型预训练框架^[48] (General Language Model, GLM) 通过引入自回归空白填充任务有效结合了各框架的优势,通过在训练中隐去连续的词语片段并重建,不仅继承了自回归模型的生成能力,还获得了自编码模型的理解能力。

3.2.1 GLM 预训练目标

GLM 的训练依赖于自回归式填空目标的优化。针对

输入文本 $\alpha = [\alpha_1, \dots, \alpha_n]$, 选择多个文本片段 $\beta = [\beta_1, \dots, \beta_n]$, 每个文本片段 β_i 匹配至 α 中的连续标记序列 $\beta = [\beta_{i1}, \dots, \beta_{in}]$ 。替换每个片段为单一的 [MASK] 标记, 构成受损文本 α_{corrupt} 。模型采用自回归方式预测受损文本中片段的缺失标记, 使得在此过程中模型能够访问到受损文本及之前的预测片段, 通过随机化片段顺序来捕获不同片段间的依赖性。

定义 Z_m 为所有长度为 m 的索引序列 $[1, 2, 3, \dots, m]$ 的集合, 并设 $\beta_{z_i} = [\beta_{z_{i1}}, \dots, \beta_{z_{i-1}}]$, 则预训练目标为:

$$\max_{\theta} \mathbb{E}_{z \sim Z_m} \left[\sum_{i=1}^m \log p_{\theta}(\beta_{z_i} | \alpha_{\text{corrupt}}, \beta_{z_{<i}}) \right] \quad (1)$$

在填空过程中, 标记的生成始终遵循从左至右的顺序, 生成片段 β_i 的概率分解为:

$$p_{\theta}(\beta_i | \alpha_{\text{corrupt}}, \beta_{z_{<i}}) = \prod_{j=1}^{l_i} p_{\theta}(\beta_{ij} | \alpha_{\text{corrupt}}, \beta_{z_{<i}}, \beta_{i<j}) \quad (2)$$

3.2.2 GLM 框架实现路径

GLM 框架下, 自回归填空首先将输入文本划分为两部分: 第一部分为未遮蔽(受损)的文本 α_{corrupt} ; 第二部分为遮蔽的片段。第一部分的标记间允许相互注意, 但不得注意第二部分标记。而第二部分标记可以注意第一部分及其自身的先行标记, 但禁止相互注意。为实现自回归生成, 每个片段均以 [START] 和 [END] 特殊标记填充, 以便输入和输出响应的生成。此方法使得模型可以自行学习双向编码器(针对第一部分)与单向解码器(针对第二部分)于一体化模型之中。GLM 自回归填空任务的实现如图 2 所示。

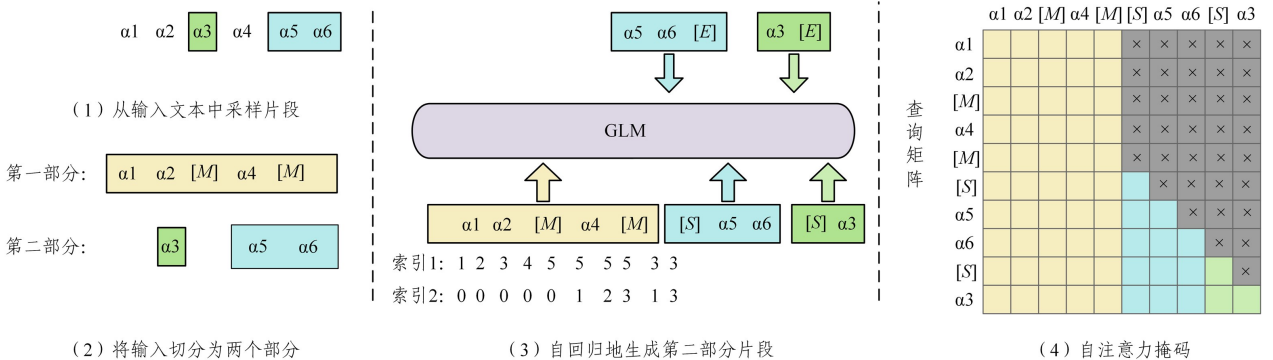


图2 自回归填空任务的实现路径

Fig. 2 Realization path of autoregressive blank filling tasks

3.3 司法三段论增强判决逻辑的实现

尽管基于 GLM 框架的大型语言预训练模型已广泛应用于各类自然语言处理任务,但在涉及刑事案件智能判决的复杂场景中,直接应用原始模型往往难以达到理想效果。为优化模型在特定法律领域的表现,本研究采取大模型增量预训练与有监督的指令微调策略,旨在确保模型在保持对广泛法律知识的全面理解与掌握的同时,能够按照司法三段论的方式对刑事案件进行精准分析,并生成具备高度参考价值的判决文本,从而提升判决的准确性和公正性。

司法三段论包括大前提(法律规范)、小前提(案件事实)和结论(判决结果),是一种基本的法律推理模式。本研究进一步应用了法官在刑事案件审理中使用的犯罪主体、犯罪客体、主观要件和客观要件四要素分析方法,并基于此设计了指令微调数据模板。通过 LoRA 微调策略,引导大语言模型高效地从复杂案情中抽取四要素关键信息,形成推理的小前提;同时,采用 RAG 技术帮助模型快速定位适用的法律条文,并将其作为推理的大前提;最终,结合大前提与小前提推导出符合逻辑的判决结果。算法伪代码如算法 1 所示。

算法 1 基于司法三段论的智能判决生成算法

输入:案件事实 case_facts,外部知识库 legal_database

输出:判决结果 judgment

1. /* 案件事实抽取 */
2. 调用函数 extract_key_elements(case_facts) 提取案件事实的四要素,将结果存储为小前提 small_premise。
3. /* 法律条文检索 */
4. 使用检索增强技术,调用函数 rag_search(legal_database, small_premise) 从法律数据库中定位与案件事实相关的法律条文,生成大前提 legal_provisions。
5. /* 推理生成判决结果 */
6. 初始化 judgment 设为 None。
7. 大语言模型遍历 legal_provisions 中的每一条法律条文 provision:
8. 调用函数 calculate_sentence(provision, small_premise), 根据案件情节和适用条文动态计算刑期。
9. 设置 judgment \gets {罪名:provision. 罪名, 刑期:刑期}。
10. 返回 judgment。

通过李某某醉酒驾驶机动车案件展示基于司法三段论的智能判决生成算法的流程。该案件的基本案情是:被告人李某某因醉酒驾驶机动车被执勤民警查获。经检测,其血液中酒精含量为 85 mg/100 ml。案发当日,李某某驾驶车辆在城市道路上行驶,未造成交通事故或人员伤亡,事后主动配合调查并如实供述违法事实。公安机关对其危险驾驶行为提起公诉。

该案件的事实包括主体“李某某”、客体“公共安全”、主观要件“明知醉酒”、客观要件“驾驶机动车,血液酒精含量 85 mg/100ml”。模型首先提取上述案件的事实四要素,构成推理的小前提。其次,利用 RAG 技术从外部知识库中定位相关法律条文,匹配到《中华人民共和国刑法》第一百三十三条

以及最高人民法院、最高人民检察院、公安部《关于办理醉酒驾驶机动车刑事案件适用法律若干问题的意见》,具体规定“血液酒精含量大于 80 mg/100 ml 构成危险驾驶罪”,形成推理的大前提。最后,大语言模型在匹配案件事实与条文适用条件后,结合自首等减轻情节,动态计算刑期,生成最终的智能判决结果:罪名为“危险驾驶罪”,刑期为拘役 3 个月,并处罚金人民币 2000 元。在整个案例判决生成过程中,大语言模型不仅展示了从法律条文解读到裁判要件的逻辑推理能力,而且凸显了大模型在法律领域应用的优势和潜力。这一过程体现了人工智能在辅助司法决策及提高判决效率和准确性方面的重要作用。

3.3.1 LoRA 大模型参数高效微调策略

参数高效微调技术(Parameter-Efficient Fine-Tuning, PEFT)可通过最小化微调过程中的参数数量和计算复杂度,既提升模型在新任务上的表现,又避免严重的灾难性遗忘,在增强模型性能的同时,显著降低训练时间和计算成本^[49]。

当前主流的 PEFT 技术包含 Adapter, Prefix Tuning, LoRA 三大类型。Adapter 技术^[50]通过在预训练模型的层间插入小型神经网络模块来实现微调;而 Prefix Tuning 技术^[51]则在模型的输入或隐层添加额外的可训练前缀标记;LoRA^[52]的核心思想在于通过对模型内部的权重矩阵进行低秩分解,来减少在微调过程中需要更新的参数量,从而提升计算效率和微调速度。

尽管 Adapter 与 Prefix Tuning 两种技术各有优势,但本研究选择 LoRA 技术进行大模型微调,考虑基于 GLM 预训练框架的大语言模型中的权重矩阵 $W \in \mathbb{R}^{m \times n}$,其中 m 和 n 分别代表输入和输出的维度。可将 W 近似为两个低秩矩阵乘积的形式来减少微调过程中需更新的参数量,即 W 被近似表示为 $W \approx A \times B$,其中 $A \in \mathbb{R}^{m \times r}$, $B \in \mathbb{R}^{r \times n}$,且 $r < \min(m, n)$ 表示这两个矩阵的秩相对较低。因此,通过更新较小的矩阵 A 和 B 而非原始矩阵 W ,LoRA 能够显著减少参数的数量,从而提升计算效率和微调速度^[53]。此外,由于只有少数参数被更新,大多数原始参数保持不变,因此模型在微调期间可保留绝大多数预训练及增量预训练期间获得的通用知识与法律领域知识,减小了过拟合的风险,可快速适用于罪名及刑期预测、判决文本生成等新任务。

3.3.2 大模型检索增强技术应用

大型语言模型的幻觉问题是指在处理输入并保持输出连贯性以及与现实世界事实一致时出现的偏差或错误^[54]。法律判决预测任务中,确保法律知识的准确性和时效性对于保证最终模型生成结果的公正性和增强其可参考性具有重要意义。若模型所援引的法律知识存在错误、欠缺或滞后,极易导致灾难性的后果。为解决大语言模型在法律专业知识领域的幻觉问题,本研究引入检索增强生成技术^[55]对微调后的 GLM 模型进行优化,流程如图 3 所示。

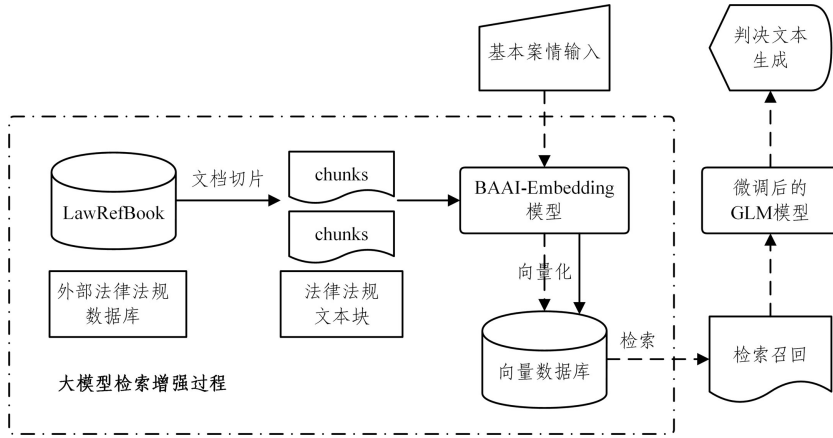


图3 RAG技术知识增强大语言模型流程图

Fig. 3 Flowchart of knowledge-enhanced large language model using RAG technology

外部知识库选取中国法律快查手册 LawRefBook, 首先, 对其进行文档载入与切片化操作, 形成法律法规文本块; 然后, 使用 BAAI/bge-large-zh-v1.5 模型对文本块进行向量化操作并存储至向量数据库 Chroma, 完成外部法律专业知识准备; 最后, 用户输入基本案情, 便可通过相似度搜索完成检索召回并传递至微调后的 GLM 模型, 实现检索增强生成。模型在生成判决文本前通过查询机制定位相关法律条文, 然后将法律知识作为上下文引入模型, 以辅助进行更准确的推理和判断。这不仅强化了模型对特定刑事罪名及其对应法律条文的理解, 也优化了案件判决结果的生成过程, 降低了模型基于不完全信息做出误判的风险。

4 实验与讨论

4.1 实验数据

实验以开源的 SLJA 数据集^[6]为基础, 结合 2021—2023 年中国裁判文书网公开的刑事案件判决书构建 SLJ 指令微调数据集, 共涵盖 10200 个案例, 覆盖 104 种刑事罪名和 117 条刑法法规, 对应的罪名与刑期统计如表 1 所列。

表1 数据集刑事案例的数据分布

Table 1 Distribution of criminal case data in the dataset

罪名	样本数量	占比/%	刑期区间	样本数量	占比/%
盗窃罪	2 927	28.70	≤1	2 571	25.21
故意伤害罪	1 604	15.73	1—3	4 847	47.52
合同诈骗罪	1 921	18.83	3—5	1 521	14.91
危险驾驶罪	853	8.36	>5	439	4.30
其他	2 895	28.38	—	822	8.06

罪名样本分布存在显著的不均衡性, 其中盗窃罪、故意伤害罪、合同诈骗罪等高频罪名的样本占比远高于其他罪名, 而低频罪名的样本数量相对不足, 限制了模型在预测任务上的泛化能力。同样地, 刑期分布呈现右偏特性, 大部分样本集中于 1~3 年的较短刑期区间, 较长刑期的样本比例较低。这种样本不均衡性可能导致模型在低频罪名预测中表现不佳, 并倾向于预测高频刑期, 忽略低频但具有重要意义的长刑期情境。为此, 本研究引入加权交

叉熵损失函数, 通过提高低频类别的权重, 增强模型对稀有类别的学习能力。

$$\mathcal{L}_{\text{weighted}} = -\frac{1}{n} \sum_{i=1}^N \sum_{j=1}^C \omega_j \cdot y_{i,j} \cdot \log(p_{i,j}) \quad (3)$$

其中, N 为样本数量, C 为类别总数, $y_{i,j}$ 为样本 i 在类别 j 中的真实标签, $p_{i,j}$ 为样本 i 属于类别 j 的预测概率, ω_j 为类别 j 的权重系数。

此外, 本研究选取 CAIL-Long 数据集^[56]、华律网法律咨询数据集^[57]、JEC-QA^[58]、裁判文书数据集作为中文无监督司法语料, 用于增量预训练大语言模型。

在数据预处理阶段, 本研究采用大语言预训练模型 Llama, 按照“法条检索—事实映射—主观要素判断—客观要素生成”的流程, 从原始裁判文书案件案情部分自动提取与标注关键法律推理要素, 包括犯罪主体、犯罪客体、犯罪主观要件及犯罪客观要件。完成自动化标注后, 采用法律专家人工校对核验的方式, 确保 SLJ 数据集中各数据项的完整性和准确性。

构建的 SLJ 指令微调数据集严格遵循“指令-输入-响应” (Instruct-Input-Output) 的规范化格式, 数据样例如表 2 所列。

表2 SLJ指令微调数据集样例

Table 2 Sample of SLJ instruction tuning dataset

Instruction	Input	Output
你是一位正在处理刑事案件案件的法官, 请依据基本案情, 采用三段论的推理方式得到该案件的判决结果, 判决结果应包含: 1. 罪名; 2. 刑期 两部分。	基本案情: 经审理查明: 明...	犯罪要件分析如下: (1) 犯罪主体: ...; (2) 犯罪客体: ...; (3) 犯罪主观要件: ...; (4) 犯罪客观要件: ...。 适用法条: 《中华人民共和国刑法》第 X 条【...罪名】 判决结果: 1) 罪名: ...; 2) 刑期: ...。

4.2 实验设置

为避免其他参数设置对模型的影响, 本实验均在同一条件下进行。实验采用的操作系统为 Ubuntu 18.04.6 LTS, 硬件配置为 Intel Xeon Gold 6330 处理器, 256 GB 运行内存, 40 TB 存储空间, 4 块 NVIDIA A40 48 GB 显卡, 实验过程中采用单机多卡并行计算的推理方式提升模型训练效率。

考虑到 SLJ 数据集中的参考文本长度, 序列分词后的最大长度限制为 1024, 批处理大小 (batch size) 设置为 4, 将在总共 100 000 步的训练中每批重复 3 次。AdamW 优化器采用 0.00005 的学习率 (Learning Rate), 并在 10 000 步的预热期后通过余弦调度策略调整学习率, 确保权重更新的平滑性。为了提高效率, 采用混合精度训练 (fp16)。LoRA 配置包括 16 层低秩更新和 0.1 的学习率乘数, 优化模型在保持原有知识的同时对新任务进行适应。为保证模型对数据的充分利用, 并确保评价结果的客观性, 实验按照 9:1 的比例将 SLJ 数据集随机划分为训练集和测试集, 验证集的比例设置为 10%。

4.3 性能评价指标

本研究选取 BLEU^[59] (Bilingual Evaluation Understudy) 和 ROUGE-L^[60] (Recall-Oriented Understudy for Gisting Evaluation) 两个指标对基座模型微调后的生成效率进行评价。BLEU 通过计算模型生成文本与参考文本的相似度来评估模型文本生成的准确率; ROUGE 则计算参考文本中的内容被模型生成的文本所涵盖的比率, 用于评估模型的召回率, 本研究采用 ROUGE-1, ROUGE-2, ROUGE-L 共同对大语言模型文本生成性能进行评估。

$$\left\{ \begin{aligned} BLEU_n(C, S) &= b(C, S) \exp\left(\sum_{n=1}^N \omega_n \lg CP_n(C, S)\right) \\ CP_n(C, S) &= \frac{\sum_i \sum_k \min(h_k(c_i), \max_{j \in m} h_k(s_{ij}))}{\sum_i \sum_k h_k(c_i)} \\ b(C, S) &= \begin{cases} 1, & \text{if } l_c > l_s \\ e^{1-l_c/l_s}, & \text{if } l_c \leq l_s \end{cases} \end{aligned} \right. \quad (4)$$

其中, $CP_n(C, S)$ 为匹配准确度, c_i 为生成文本, s_{ij} 为参考文本, m 为参考文本的个数, $h_k(c_i)$ 和 $h_k(s_{ij})$ 分别代表一个或多个连续的词在生成文本和参考文本中出现的次数; l_c 是生成文本总长度, l_s 是参考文本总长度; ω_n 为权重; $b(C, S)$ 是惩罚机制, 用于防止短句高分情形。

$$ROUGE-L = \frac{\sum_{s \in (\text{RefSum})} LCS(\text{output}, s)}{\sum_{s \in (\text{RefSum})} length(s)} \quad (5)$$

其中, $LCS(\text{output}, s)$ 是输出和评估文本之间最长公共子序列的长度, $length(s)$ 是评估文本的长度。

其次, 本研究中涉及的法律判决预测中的罪名、刑期预测两项子任务, 都属于多标签的多分类问题。为避免样本之间的不平衡性, 本研究采用精度 (Precision, P)、召回率 (Recall, R) 和 F1 值 3 项指标来衡量模型的预测效果。

$$P = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FP_i} \quad (6)$$

$$R = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FN_i} \quad (7)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (8)$$

其中, 下标 i 为分类任务中类别的种类; TP_i 为被正确地划分了类别 i 的样本个数; FP_i 为实际为其他类但被分类器划分为

i 类的样本数; FN_i 为实际为 i 类, 但是被分类器划分错误的样本数。

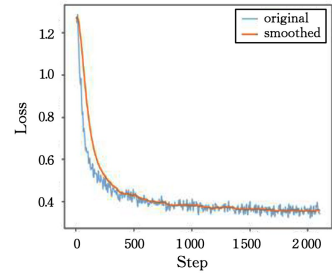
4.4 基线模型

为检验基座大语言模型的性能和本研究提出的刑事案件智能判决模型的有效性, 及其与基于域外语料训练的法律大模型能力相比的优劣, 进行了 3 组对比实验。1) 在 SLJ 数据集上对比不同大语言模型的指令微调效果。在对 Open LLM^[61] 以及 LawBench^[62] 等公开且认可度高的大模型评测榜单进行分析后, 选取 Qwen-7B-Chat^[63] 和 Baichuan2-7B-Chat^[64] 两种主流对话大语言模型作为对比模型。2) 对比本文模型与 3 种基线模型的效果差异, 基线模型的选取覆盖基于词嵌入的深度学习模型 MTL-Fusion 模型^[65]、中国司法长文本文书预训练模型 Lawformer^[36] 以及预训练模型 BERT^[66]。3) 对比本文模型与 LegalBERT^[34] 和 CaseLawBERT^[35] 两个基于域外法律语料微调得到的法律大模型。

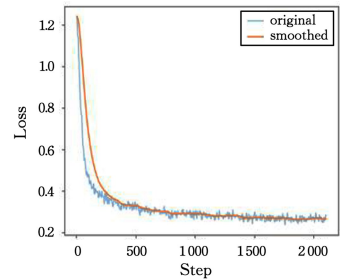
4.5 实验结果与分析

4.5.1 指令微调结果对比分析

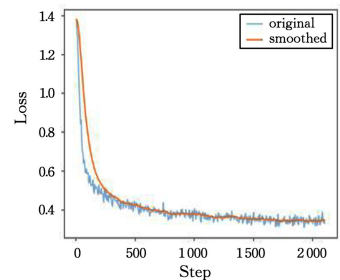
本研究对比了 ChatGLM3-6B-32k, Qwen-7B-Chat 和 Baichuan2-7B-Chat 这 3 个大语言模型在指令微调过程中的损失率的变化情况, 结果如图 4 所示。



(a) ChatGLM3-6B-32k



(b) Qwen-7B-Chat



(c) Baichuan2-7B-Chat

图 4 模型训练过程中损失率变化的对比图

Fig. 4 Comparison of loss rate changes during model training

图4可以看出,ChatGLM3-6B-32k与Qwen-7B-Chat的稳态表现及损失率下降速度接近,表明两者在泛化能力和新任务适应性上较为相似。相比之下,Baichuan2-7B-Chat虽在初期表现出较快的适应速度,但后期稳定性较差,出现一定程

度的过拟合。从损失率的稳态表现来看,ChatGLM3-6B-32k和Qwen-7B-Chat在训练后期展现出更强的稳定性和泛化能力。微调后的各模型在测试集上进行评估的结果如表3所列。

表3 不同大语言模型指令微调结果评估

Table 3 Evaluation of instruction tuning results for different large language model

模型	Evaluation of instruction tuning results for different large language model (%)							
	BLEU		ROUGE-1		ROUGE-2		ROUGE-L	
	原始	微调后	原始	微调后	原始	微调后	原始	微调后
ChatGLM3-6B-32k	5.18	24.58	27.41	50.24	11.54	39.99	17.11	45.16
Qwen-7B-Chat	3.25	20.52	21.53	49.11	10.27	39.51	13.66	43.15
Baichuan2-7B-Chat	1.35	7.76	5.67	11.41	0.92	7.25	4.21	10.57

ChatGLM3-6B-32k在BLEU指标上的优异表现,表明其在生成和评估语义相关性强的文本方面能力较强。刑事案件智能判决对特定法律术语的准确使用等精确性要求极高,因此ChatGLM3-6B-32k模型具有显著优势。

相比之下,Qwen-7B-Chat虽在各项指标上略低于ChatGLM3-6B-32k,但两者ROUGE-L得分接近,反映出它们在内容覆盖方面的表现相当。然而,它们在BLEU指标上的差距明显,表明Qwen-7B-Chat在生成与源文本字面相似内容的能力上仍有不足。Baichuan2-7B-Chat的整体性能明显低于前两者,尤其在BLEU指标上表现较弱,反映出其在理解和复现法律文本特定术语及结构方面存在明显短板。

4.5.2 判决预测结果对比分析

表4列出了本文提出的模型与基准模型在罪名及刑期预测任务上的表现。

表4 法律判决预测结果的对比

Table 4 Comparison of judicial judgment prediction results

模型	精度 P		召回率 R		F1 值	
	罪名	刑期	罪名	刑期	罪名	刑期
MTL-Fusion	0.6861	0.3512	0.6911	0.3567	0.6886	0.3539
Lawformer	0.6927	0.3581	0.7082	0.3629	0.7004	0.3605
BERT	0.7011	0.4311	0.7178	0.4308	0.7094	0.4309
Ours	0.7517	0.4712	0.7478	0.4671	0.7497	0.4691

Lawformer和MTL-Fusion模型性能较差,主要原因在于其在解码法律文本复杂结构和细粒度语义信息方面能力有限,特别是在捕捉与罪名和刑期密切相关的细节信息时表现不足。相比之下,BERT模型展现出一定的性能优势,这得益于其在预训练阶段积累的丰富语义理解能力,显著提升了预测效果,证明了预训练大语言模型在处理复杂结构和语义信息上的突出能力。

本文提出的模型在罪名和刑期预测的精确度上明显超越了基准模型,这得益于本文模型针对法律语言细节和裁判逻辑所设计的微调语料和策略。在召回率上,本文模型在两项任务中均取得了最高值,表明其更能准确地捕捉法律文本中的复杂语义关系,并深入理解案件事实与法律规定之间的联系。预训练大语言模型突出的语境理解能力,使其能够更准确地识别裁判文书中的关键信息,从而更全面地涵盖真实标签集中的所有标签。

此外,在F1值方面,本文提出的模型相比BERT模型,在罪名预测任务上提升了4%,在刑期预测任务上提升了

3%。这表明该模型不仅能够减少误判,还能够全面识别所有正确标签,在精确率与召回率之间找到平衡,确保所有相关标签都能被准确识别并覆盖。

4.5.3 法律大模型横向对比分析

本研究依据《法律大模型评估指标和评测办法(征求意见稿)》中所提出的法律大模型评估指标与方法,针对法律判决预测任务,选取语言理解(案件要素抽取)、内容生成(法律文书生成)、逻辑推理(司法决策推理)、逻辑推理(案情分析)为核心评估指标,基于司南OpenCompass大模型测评平台,在中文法律数据集上对Legal-BERT、CaseLawBERT以及本文模型进行评估,评分范围为1~10,结果如图5所示。

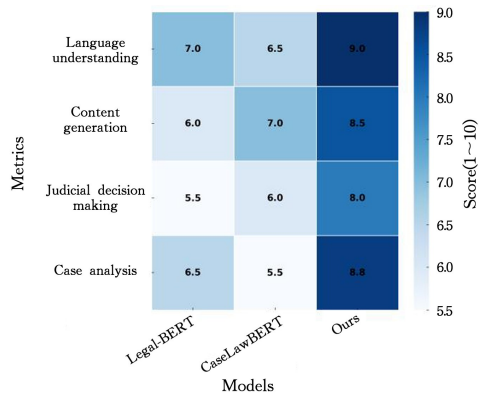


图5 法律大模型评估热力图

Fig. 5 Heatmap of legal large language models evaluation

本文所构建的模型在4项评估指标上均表现突出,尤其在语言理解和案情分析方面展现出对中文法律文本的深刻理解能力以及对复杂案情逻辑关系的高效解析能力。相比之下,Legal-BERT在语言理解和案情分析上表现尚可,但在内容生成和司法决策推理方面表现不佳。CaseLawBERT虽在内容生成上稍优于Legal-BERT,但整体性能仍有较大改进空间。主要原因在于Legal-BERT和CaseLawBERT基于欧美法律语料预训练,对中国司法系统特有的刑法体系和判决逻辑适应性不足,导致模型在罪名预测和刑期预测任务上的准确性受到限制。此外,中国刑事案件通常为长文本,而Legal-BERT和CaseLawBERT在长文本处理上能力有限,难以有效捕捉复杂案情中的细粒度信息。

4.6 具体案例分析

将测试集中的何某某销售假冒注册商标的商品罪一审刑

事判决书结构化数据作为对象进行分析,真实判决结果、原始模型生成结果以及微调后模型生成的结果如表 5 所列。

表 5 大语言模型微调与优化后刑事案件判决结果分析

Table 5 Analysis of criminal case trial results after fine-tuning and optimization of large language models

真实判决结果	<p>本院认为,被告人何某某销售明知是假冒注册商标的商品,销售货物金额数额巨大,其行为已构成销售假冒注册商标的商品罪。部分行为系因被告人意志以外的原因而使犯罪未得逞,系犯罪未遂,可比照既遂犯减轻处罚。公诉机关指控的事实清楚,罪名成立,适用法律正确,予以支持。被告人何某某具有自首情节,依法予以减轻处罚。被告人何某某在案发后已退出违法所得,可酌情从轻处罚。辩护人提出的应按照被告人之前所销售价格中的最低价来确定尚未销售商品部分的单价的观点不合理,不予采纳,但采纳其要求对被告人何竹娟减轻处罚的辩护意见。依照《中华人民共和国刑法》第二百一十四条、第二十三条、第六十七条第一款、第七十二条第一款、第三款、第七十三条第二款、第三款、第五十二条、第五十三条、第六十四条之规定,判决如下:</p> <p>一、被告人何某某犯销售假冒注册商标的商品罪,判处有期徒刑二年,缓刑二年六个月,并处罚金人民币十五万元。罚金限于判决生效后五日内缴纳。</p> <p>(缓刑考验期限,从判决确定之日起计算)。</p> <p>二、没收违法所得及被扣押的假冒 X 品牌内裤。</p>
原始模型生成的案件判决结果	<p>首先,根据基本案情,我们可以得出以下信息:</p> <p>被告人何某某在其经营的阿里巴巴店铺内销售涉嫌假冒 X 品牌注册商标的男士内裤,实际销售金额共计人民币 62506.17 元。</p> <p>案发后,金华市婺城区市场监管局在何某租用的 2 个车库内查获了尚未销售的涉嫌假冒 X 的内裤 50440 条,货物价值计人民币 260270.4 元。</p> <p>公诉机关认为被告人何竹娟的行为已构成犯罪,部分系犯罪未遂,系自首。</p> <p>根据以上信息,利用三段论的推理方式得到判决结果:</p> <p>罪名:被告人何某某的行为已构成销售假冒注册商标的商品罪。刑期:根据我国刑法的规定,对于自首的犯罪分子,可以从轻或者减轻处罚。因此,对于被告人何竹娟的刑期,应在法律规定的刑罚幅度内,根据案件的具体情节,从轻或者减轻处罚。</p> <p>综上,根据基本案情,被告人何某某的行为已构成销售假冒注册商标的商品罪,应判处有期徒刑五年,并处罚金。</p>
微调后模型生成的案件判决结果	<p>犯罪要件:</p> <p>犯罪主体:犯罪主体为被告人何某某,已达到法定年龄且为完全行为能力人。</p> <p>构成要件包括:主观方面的故意犯罪、客观方面的实施行为和犯罪结果。</p> <p>犯罪客体:犯罪客体为国家对商标的管理制度和他人注册商标的专用权。被告人何竹娟在明知所销售的是涉嫌假冒 X 品牌注册商标的男士内裤的情况下,仍在其经营的店铺内销售上述品牌的不同款式的男士内裤,侵犯了 X 品牌注册商标的专用权。</p> <p>犯罪主观要件:被告人明知所销售的男士内裤涉嫌假冒 X 品牌注册商标,仍销售并获取非法利益。犯罪意图为非法牟利。</p> <p>犯罪客观要件:犯罪时间:2016 年 7 月 13 日至 2016 年 8 月 9 日。犯罪地点:金华市婺城区被告人何竹娟经营的阿里巴巴网店“Y”店铺。犯罪行为:在明知所销售的是涉嫌假冒 X 品牌注册商标的男士内裤的情况下,销售上述品牌的不同款式的男士内裤,实际销售金额共计人民币 62506.17 元。犯罪结果:被查获尚未销售的涉嫌假冒 X 品牌的内裤 50440 条,货物价值计人民币 260270.4 元。</p> <p>适用法条:《中华人民共和国刑法》第二百一十四条【销售假冒注册商标的商品罪】,销售明知是假冒注册商标的商品,销售金额数额较大的,处三年以下有期徒刑或者拘役,并处或者单处罚金;销售金额数额巨大的,处三年以上七年以下有期徒刑,并处罚金。原因:被告人销售假冒 X 品牌注册商标的男士内裤,销售金额数额较大,符合该罪名的构成要件。但被告人何竹娟具有自首情节,依法予以减轻处罚。</p> <p>判决结果:罪名:销售假冒注册商标的商品罪</p> <p>刑期:三年以下有期徒刑或者拘役,并处或者单处罚金。</p>

模型微调前后的生成判决结果对比如表 6 所列。

表 6 大语言模型微调前后刑事案件判决效果的对比

Table 6 Comparison of criminal case trial effects before and after fine-tuning of large language models

效果测度	微调前	微调后
案件事实描述	提供了被告人何某某的犯罪行为的具体金额和查获的货物价值,以及公诉机关的认定观点。	深入抽取提炼了案件事实的法律要件,例如时间、地点、行为、结果等,更具有结构性。
定罪分析	成功认定罪名为“销售假冒注册商标的商品罪”,但并没有给出适用的法条及原因,推理逻辑性不强。	成功认定罪名为“销售假冒注册商标的商品罪”的基础上,给出推理及法律适用逻辑。
量刑分析	没有充分考虑自首情节,也没有明确说明“犯罪未遂”对于减轻、从轻处罚的作用,给出的五年刑期也不符合法院给出的标准判决。	充分考虑到了量刑因素,并正确返回了适用的法条,给出的量刑范围也更贴合法院最后的判决结果。

总的来说,微调后的回答在案件事实、定罪量刑、法律适用的处理上更为详细和合理,体现了对法律文档的理解的深入,更展现了其在司法决策领域的强大辅助能力。

微调后的模型所生成的案件判决结果能够更清晰地呈现法院在作出判决时的考虑要件,同时对于犯罪主体、犯罪

客体、犯罪的主观要件、犯罪的客观要件提取较为合理,应用法条、得出判决更为自然,改善了智能判决预测结果可解释性与可采信度低的问题。但对比法院裁判说理,在细节上仍略有不足,如对犯罪主体的构成要件、适用法条的召回捕捉不佳。而法院的裁判说理考虑了犯罪未遂、退出违法所得等因素,给出了具体的刑罚和罚金,以及缓刑的考验期限。微调后的模型所生成的案件判决结果在定罪方面符合实际判决,但较为笼统,没有给出具体的刑罚,只是指出了刑罚的范围;量刑方面考虑到了自首等减轻处罚的情节,但不够周延全面,对量刑的法定因素把握尚不深刻。微调后的模型虽在现阶段不能完全替代法官进行裁判,但综合上述回答效果,其能在犯罪构成提取、推理逻辑构建等方面有力辅助法官完成说理。

结束语 本研究针对罪名预测与刑期预测任务,提出了以“标记案件语料—预训练大模型—强化判决逻辑”为思路的刑事案件智能判决方法。在参考 SLJA 数据集的基础上,构建了更加均衡且丰富的 SLJ 刑事案件数据集。通过 LoRA 参数微调策略和 RAG 技术,高效地对增量预训练后的 ChatGLM3-6B-32k 模型进行了参数调优和法律知识扩充。与 Qwen-7B-Chat, Baichuan2-7B-Chat 等同级参数的大语言模型相比,微调后的模型性能提升更为出色。实验结果表明,得

益于对法律数据的深度解析与综合应用,本研究构建的模型在生成判决文本的语义推理方面更符合实际法官的裁判逻辑。在法律判决预测领域,该模型在罪名判定及刑期量定的精确度上相较于主流基线模型有显著提升。

数字时代,人工智能的迭代发展已被应用于产业升级和治理能力现代化的实现,我国也启动了智慧社会、智慧司法、法治信息化的建设。技术的引入使得法律判决过程更加透明,解释性更加合理,大大降低了法律的判决成本、执行成本;与此同时,智能判决在传统法律思维方式之外,确立了智慧司法、智慧治理的新理念,呈现出法律规则和技术规则相互补充、协同促进的局面^[67]。因此,判决智能化将是一个必然趋势,这是规范判决逻辑、巩固司法秩序、筑牢法律信任、实现“数字正义”的关键。

但不能否认的是,智慧司法也存在风险,智能的一个极端是技术理性的流失与法律逻辑的崩坏。智能判决的核心要义在于实现类案同判,但类案同判并不等同于个案公平。当前智能判决仍面临公众认可度低、算法经验归纳的非周延性及系统性偏差^[68]等问题。未来的研究工作将以“可视正义”为目标,一方面,通过探索司法判决数据集的多样性,拓宽刑事案件案件的覆盖范围,消除因语料规模不足导致的模型偏见;另一方面,应用知识图谱技术提升模型的法律推理能力,同时捕捉与可视化大模型中的关键决策信号与推理步骤,综合提升智能判决模型的可解释性。

大模型技术在司法领域的广泛应用将推动司法制度和理念的不断革新,相信在顺应法治信息化、智能化的趋势下,通过完善技术应用的保障机制,可以充分发挥大语言模型的优势,从而向“类案同判与个案公正相统一”的目标持续迈进。

参 考 文 献

- [1] CUI J, SHEN X, WEN S. A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges[J]. IEEE Access, 2023, 11: 102050-102071.
- [2] YUE L, LIU Q, WU H, et al. Circumstances enhanced Criminal Court View Generation[C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 1855-1859.
- [3] WU Y, KUANG K, ZHANG Y, et al. De-Biased Court's View Generation with Causality[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020: 763-780.
- [4] YE H, JIANG X, LUO Z, et al. Interpretable Charge Predictions for Criminal Cases: Learning to Generate Court Views from Fact Descriptions[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. 2018: 1854-1864.
- [5] SILVER N. The Signal and the Noise: Why So Many Predictions Fail-But Some Don't[M]. Penguin, 2012.
- [6] DENG W, PEI J, KONG K, et al. Syllogistic Reasoning for Legal Judgment Analysis[C]// Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023: 13997-14009.
- [7] ALETRAS N, TSARAPATSANIS D, PREOTIUC-PIETRO D, et al. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing Perspective[J]. PeerJ Computer Science, 2016, 2: e93.
- [8] KORT F. Predicting Supreme Court Decisions Mathematically: A Quantitative Analysis of the "Right to Counsel" Cases[J]. American Political Science Review, 1957, 51(1): 1-12.
- [9] NAGEL S. Applying Correlation Analysis to Case Prediction[J]. Texas Law Review, 1963, 42: 1006.
- [10] LAWLOR R. What Computers Can Do: Analysis and Prediction of Judicial Decisions[J]. American Bar Association Journal, 1963, 49(4): 337-344.
- [11] KEOWN R. Mathematical Models For Legal Prediction, 2 Computer L. J. 829(1980)[J]. UIC John Marshall Journal of Information Technology & Privacy Law, 1980, 2(1): 29.
- [12] SUSSKIND R. Expert Systems in Law: A Jurisprudential Approach to Artificial Intelligence and Legal Reasoning[J]. The Modern Law Review, 1986, 49(2): 168-194.
- [13] LIU C L, LIAO T M. Classifying Criminal Charges in Chinese for Web-Based Legal Services[C]// Asia-pacific Web Conference. 2005: 64-75.
- [14] BOELLA G, DI CARO L, HUMPHREYS L. Using Classification to Support Legal Knowledge Engineers in the Eunomos Legal Document Management System[C]// Fifth International Workshop on Juris-informatics. 2011.
- [15] GONÇALVES T, QUARESMA P. Evaluating Preprocessing Techniques in a Text Classification Problem[C]// Proceedings of the Conference of the Brazilian Computer Society. 2005.
- [16] LIN W, KUO T, CHANG T, et al. Exploiting Machine Learning Models for Chinese Legal Documents Labeling, Case Classification, and Sentencing Prediction[C]// Proceedings of ROCLING. 2012: 140.
- [17] LIU C, HSIEH C. Exploring Phrase-based Classification of Judicial Documents for Criminal Charges in Chinese[C]// Proceedings of the 16th International Conference on Foundations of Intelligent Systems. 2006: 681-690.
- [18] LIU C, CHANG C, HO J. Case Instance Generation and Refinement for Case-Based Criminal Summary Judgments in Chinese. [J]. Journal of Information Science and Engineering, 2004, 20(4): 783-800.
- [19] KIANMEHR K, ALHAJJ R. Crime Hot-Spots Prediction Using Support Vector Machine[C]// IEEE International Conference on Computer Systems and Applications. 2006: 952-959.
- [20] SULEA O, ZAMPIERI M, MALMASI S, et al. Exploring the Use of Text Classification in the Legal Domain[C]// Proceedings of the Second Workshop on Automated Semantic Analysis of Information in Legal Texts co-located with the 16th International Conference on Artificial Intelligence and Law. 2017: 2143.
- [21] KATZ D. Quantitative Legal Prediction-or-How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry[J]. Emory Law Journal, 2013, 62(4): 909.
- [22] ZELEZNIKOW J, STRANIERI A. The Split-up System: Integrating Neural Networks and Rule-Based Reasoning in the Le-

- gal Domain[C]// Proceedings of the 5th International Conference on Artificial Intelligence and Law. 1995;185-194.
- [23] HUANG N, HE J, SUN J, et al. Improved Lawformer-based Approach for Forecasting Crimes[J]. Journal of Beijing University of Technology, 2019, 45(8):742-748.
- [24] LUO B, FENG Y, XU J, et al. Learning to Predict Charges for Criminal Cases with Legal Basis[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017;2727-2736.
- [25] HU Z, LI X, TU C, et al. Few-Shot Charge Prediction with Discriminative Legal Attributes[C]// Proceedings of the 27th International Conference on Computational Linguistics. 2018: 487-498.
- [26] ZHAO J, GUAN Z, XU C, et al. Charge Prediction by Constitutive Elements Matching of Crimes[C]// International Joint Conferences on Artificial Intelligence Organization. 2022: 4517-4523.
- [27] CHEN Y, LIU Y, HO W. A text mining approach to assist the general public in the retrieval of legal documents[J]. Journal of the American Society for Information Science and Technology, 2013, 64(2):280-290.
- [28] KIM M, XU Y, GOEBEL R. Legal Question Answering Using Ranking SVM and Syntactic/Semantic Similarity [C] // New Frontiers in Artificial Intelligence. 2015;244-258.
- [29] ZHONG H, GUO Z, TU C, et al. Legal Judgment Prediction via Topological Learning[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 3540-3549.
- [30] YUE L, LIU Q, JIN B, et al. NeurJudge: A Circumstance-aware Neural Framework for Legal Judgment Prediction [C] // Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021:973-982.
- [31] PETERS M, NEUMANN M, IYYER M, et al. Deep Contextualized Word Representations[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. 2018;2227-2237.
- [32] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2019;4171-4186.
- [33] LIU Y, OTT M, GOYAL N, et al. A robustly optimized BERT pre-training approach with post-training [C]// China National Conference on Chinese Computational Linguistics. Cham: Springer International Publishing, 2021;471-484.
- [34] CHALKIDIS I, FERGADIOTIS M, MALAKASIOTIS P, et al. LEGAL-BERT: The Muppets straight out of Law School [C]// Findings of the Association for Computational Linguistics. 2020: 2898-2904.
- [35] ZHENG L, GUHA N, ANDERSON B, et al. When does pre-training help? assessing self-supervised learning for law and the CaseHOLD dataset of 53,000+ legal holdings [C] // Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law. 2021;159-168.
- [36] XIAO C, HU X, LIU Z, et al. Lawformer: A Pre-trained Language Model for Chinese Legal Long Documents [J]. AI Open, 2021, 2:79-84.
- [37] 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.
- [38] KOJIMA T, GU S, REID M, et al. Large Language Models are Zero-Shot Reasoners [J]. Advances in Neural Information Processing Systems, 2022, 35:22199-22213.
- [39] ZELIKMAN E, WU Y, MU J, et al. STaR: self-taught reasoner bootstrapping reasoning with reasoning [C]// Proceedings of the 36th International Conference on Neural Information Processing Systems. 2024;15476-15488.
- [40] TRAUTMANN D, PETROVA A, SCHILDER F. Legal Prompt Engineering for Multilingual Legal Judgement Prediction [J]. arXiv:2212.02199, 2022.
- [41] YU F, QUARTEY L, SCHILDER F. Legal Prompting: Teaching a Language Model to think like a Lawyer [J]. arXiv:2212.01326, 2022.
- [42] JIANG C, YANG X. Legal Syllogism Prompting: Teaching Large Language Models for Legal Judgment Prediction [C] // Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law. 2023;417-421.
- [43] HUANG Q, TAO M, ZHANG C, et al. Lawyer LLaMA Technical Report [J]. arXiv:2305.15062, 2023.
- [44] CHEN S, HOU Y, CUI Y, et al. Recall and Learn: Fine-tuning Deep Pretrained Language Models with Less Forgetting [C] // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020;7870-7881.
- [45] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving Language Understanding by Generative Pre-Training [EB/OL]. [2025-03-11]. <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf>.
- [46] DEEPA M. Bidirectional Encoder Representations from Transformers (BERT) Language Model for Sentiment Analysis Task [J]. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 2021, 12(7):1708-1721.
- [47] 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(140):1-67.
- [48] DU Z, QIAN Y, LIU X, et al. GLM: General Language Model Pretraining with Autoregressive Blank Infilling [C] // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022;320-335.
- [49] WEN S, QIAN L, HU H, et al. Review of Research Progress on Question-Answering Techniques Based on Large Language Models [J]. Data Analysis and Knowledge Discovery, 2024, 8(6):16-29.
- [50] HOULSBY N, GIURGIU A, JASTRZEBSKI S, et al. Parameter-Efficient Transfer Learning for NLP [C] // Proceedings of the

- 36th International Conference on Machine Learning. PMLR, 2019;2790-2799.
- [51] LI X L, LIANG P. Prefix-Tuning: Optimizing Continuous Prompts for Generation[J]. arXiv:2101.00190, 2021.
- [52] HU E J, SHEN Y, WALLIS P, et al. LoRA: Low-Rank Adaptation of Large Language Models[J]. arXiv:2106.09685, 2021.
- [53] ZHANG Q T, WANG Y C, WANG H X, et al. A Comprehensive Review of Large Language Model Fine-tuning[J]. Computer Engineering and Applications, 2024, 60(17): 17-33.
- [54] ZHAO Y, HE J W, ZHU S C, et al. Security of Large Language Models: Current Status and Challenges[J]. Computer Science, 2024, 51(1): 68-71.
- [55] 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.
- [56] XIAO C, ZHONG H, GUO Z, et al. CAIL2018: A Large-Scale Legal Dataset for Judgment Prediction[J]. arXiv:1807.02478, 2018.
- [57] CHENG F D. Legal Consultation Data and Corpus from China Law Network; Replication Data for Design and Research of Text Classification System[EB/OL]. [2025-03-11]. <https://doi.org/10.18170/DVN/OLO4G8>.
- [58] ZHONG H, XIAO C, TU C, et al. JEC-QA: a Legal-domain Question Answering Dataset[C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 9701-9708.
- [59] PAPINENI K, ROUKOS S, WARD T, et al. BLEU: a Method for Automatic Evaluation of Machine Translation[C] // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 2002: 311-318.
- [60] LIN C Y. ROUGE: A Package for Automatic Evaluation of Summaries[C] // Text Summarization Branches Out, Barcelona, Spain: Association for Computational Linguistics, 2004: 74-81.
- [61] XU L, LI A, ZHU L, et al. Superclue: A Comprehensive Chinese Large Language Model Benchmark [J]. arXiv: 2307.15020, 2023.
- [62] FEI Z, SHEN X, ZHU D, et al. LawBench: Benchmarking Legal Knowledge of Large Language Models[J]. arXiv: 2309.16289, 2023.
- [63] BAI J, BAI S, CHU Y, et al. Qwen Technical Report[J]. arXiv: 2309.16609, 2023.
- [64] YANG A, XIAO B, WANG B, et al. Baichuan 2: Open Large-scale Language Models[J]. arXiv: 2309.10305, 2023.
- [65] LIU Z L, ZHANG M S, ZHENG R R, et al. Multi-task Learning Model for Legal Judgment Predictions with Charge Keywords [J]. Journal of Tsinghua University (Science and Technology), 2019, 59(7): 497-504.
- [66] FAN A M, WANG Y C. Multi Task Intelligent Legal Judgment Method Based on BERT Model[J]. Microelectronics & Computer, 2022, 39(9): 107-114.
- [67] MA C S. Law Towards a Digital Society[M]. Beijing: Law Press China, 2021.
- [68] ZHENG Y. The Embedding of Artificial Intelligence in the Rule of Law: A Jurisprudential Reflection of Intelligent Justice[J]. Jianghai Academic Journal, 2024, (3): 172-180, 256.



CONG Yingnan, born in 1985, Ph.D, associate professor, Ph.D supervisor, is a member of CCF (No. J0079M). His main research interests include big data on business and law, artificial intelligence, blockchain, Fin-tech, Reg-tech and complex system.



ZHU Jinqing, born in 1984, postgraduate, engineer, is a member of CCF (No. D0034M). His main research interests include database systems, content data analysis, artificial intelligence, and knowledge graphs.

(责任编辑:何杨)