

基于大语言模型, 法律引导的的案例融合方法, 用于 司法判决预测

作者名¹⁾ 作者名^{2),3)} 作者名³⁾(* 字体为 3 号仿宋 * 作者)*

¹⁾(单位全名部门(系) 全名, 市(或直辖市) 国家名邮政编码)* 字体为 6 号宋体 * 单位

²⁾(单位全名部门(系) 全名, 市(或直辖市) 国家名邮政编码)* 中英文单位名称、作者姓名须一致 *

³⁾(单位全名部门(系) 全名, 市(或直辖市) 国家名邮政编码)

论文定稿后, 作者署名、单位无特殊情况不能变更。若变更, 须提交签章申请, 国家为中国可以不写, 省会城市不写省名, 其他国家必须写国家名。

摘 要 * 中文摘要内容置于此处(英文摘要中要有这些内容), 字体为小 5 号宋体。摘要贡献部分, 要有数据支持, 不要出现“... 大大提高”、“... 显著改善”等描述, 正确的描述是“比...提高 X%”、“在...上改善 X%”。* 摘要

关键词 * 关键词(中文关键字与英文关键字对应且一致, 应有 5-7 个关键词); 关键词; 关键词; 关键词 *

中图法分类号 TP DOI 号: * 投稿时不提供 DOI 号

Title * (中英文题目一致) 字体为 4 号 Times New Roman, 加粗 * Title

NAME Name-Name¹⁾ NAME Name²⁾ NAME Name-Name³⁾ * 字体为 5 号 Times new Roman*Name

¹⁾(Department of ****, University, City ZipCode, China) * 字体为 6 号 Times new Roman* Depart.Correspond

²⁾(Department of ****, University, City ZipCode)* 中国不写国家名 *

³⁾(Department of ****, University, City ZipCode, country)* 外国写国家名 *

Abstract (500 英文单词, 内容包含中文摘要的内容). 字体为 Times new Roman, 字号 5 号 * Do not modify the amount of space before and after the artworks. One- or two-column format artworks are preferred. and Tables, create a new break line and paste the resized artworks where desired. Do not modify the amount of space before and after the artworks. One- or two-column format artworks are preferred. All Schemes, Equations, Figures, and Tables should be mentioned in the text consecutively and numbered with Arabic numerals, and appear below where they are mentioned for the first time in the main text. To insert Schemes, Equations, Figures, and Tables, create a new break line and paste the resized artworks where desired. Do not modify the amount of space before and after the artworks. One- or two-column format artworks are preferred.Do not modify the amount of space before and after the artworks. One- or two-column format artworks are preferred. and Tables, create a new break line and paste the resized artworks where desired.

*收稿日期: - - ; 最终修改稿收到日期: - - * 投稿时不填写此项 * 本课题得到……基金中文完整名称(No. 项目号)、……基金中文完整名称(No. 项目号)、……基金中文完整名称(No. 项目号)资助。作者名 1(通信作者), 性别, xxxx 年生, 学位(或目前学历), 职称, 是/否计算机学会(CCF)会员(提供会员号), 主要研究领域为 ****、****.E-mail: *****. 作者名 2 (通信作者), 性别, xxxx 年生, 学位(或目前学历), 职称, 是/否计算机学会(CCF)会员(提供会员号), 主要研究领域为 ****、****.E-mail: *****. 作者名 3 (通信作者), 性别, xxxx 年生, 学位(或目前学历), 职称, 是/否计算机学会(CCF)会员(提供会员号), 主要研究领域为 ****、****.E-mail: *****.(给出的电子邮件地址应不会因出国、毕业、更换工作单位等原因而变动。请给出所有作者的电子邮件) 第 1 作者手机号码(投稿时必须提供, 以便紧急联系, 发表时会删除): ……., E-mail: ……* 此部分 6 号宋体 *

Keywords 中文关键字与英文关键字对应且一致, 不要用英文缩写); key word; key word; key word* * 字体为 5 号 Times new Roman * Key words

1 方法

1.1 总体流程

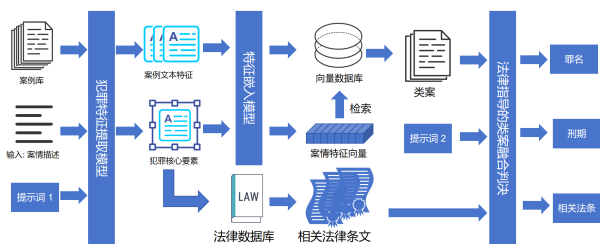


Fig. 1 基于大语言模型, 法律引导的案例融合方法, 用于司法判决预测流程

本研究提出的司法判决预测方法其核心在于构建一个能够有效整合案件事实、法律法规及类案判例的智能推理框架。整体流程如图??所示。首先, 系统接收待判决案件的详细事实描述 (C), 并利用预训练的大语言模型进行初步语义分析和关键信息抽取。该模型通过自然语言理解能力, 从复杂的案情描述中识别并初步推断出罪名类别、犯罪构成要件 (包括主体、主观、客体、客观) 及证据特征等核心法律要素, 形成结构化的犯罪特征表示 (F)。其次, 基于抽取的犯罪特征 F , 系统并行地从两个专门构建的法律知识库中检索相关信息。一方面, 从权威的法律条文数据库中检索出与案件特征 F 高度相关的法律法规条款集合 (L)。另一方面, 从海量的历史案例数据库中智能检索出与当前案件在罪名构成、事实情节和证据方面最为相似的判例集合 (S)。最后, 将原始案件事实描述 C 、初步提取的犯罪特征 F 、检索到的相关法律条文 L 以及筛选出的相似判例 S 共同作为上下文信息, 输入至大语言模型 (LLM), 由其进行综合分析推理, 输出最终的判决结果 (J)。该模型作为核心推理引擎, 通过整合这些多源异构信息, 综合考量法律原则、司法解释及类案判例的指导作用, 有效弥补了 LLM 在法律专业知识和复杂逻辑推理方面的固有局限性, 旨在显著增强判决预

测的专业性、准确性和可解释性。

1.2 判决核心要素提取

在大陆法系的刑法理论中, 一项行为要被判定为犯罪, 其客观事实必须符合刑法分则具体条文所规定的全部构成要件。此外, 法官在进行司法判决时, 还需要根据检查机关提供的证据, 犯罪嫌疑人的行为动机, 造成的事实性后果综合给判决。然而, 传统方法可能“未能识别法律要素间的逻辑依赖”, 导致判决预测缺少法律推理的前提, 最终造成判决结果的不可信甚至是判决错误。为了解决该问题, 本研究提出的“判决核心要素提取”, 通过首先明确识别出这些要素, 为后续的逻辑关联分析和推理奠定基础。根据中国刑法理论, 判决核心要素 F 由犯罪核心要素和证据核心要素。其中犯罪要素包括犯罪主体, 犯罪客体, 犯罪主观方面, 犯罪客观方面四个方面; 证据核心要素由证据和证明力组成, 证据是检查机关提供的直接证据, 证明力是证据的配合, 可信程度。该证据特征由 LLM 通过特别定制的提示词 1 从原始输入中提取。

提示词 1:

你是一名法律案例分析员, 负责根据输入的法律案件信息提取关键特征, 并生成结构化输出。你需要根据输入的和原始事实描述, 提取法律裁判的核心特征并按照示例格式输出结果, 字段命名和结构需与示例一致。要求如下: 1. 必含字段: 罪名类别、犯罪构成、证据特征 2. 数值单位统一: 刑期以月为单位 (如“9 个月”), 罚金保留原数值 (如“1000 元”)。4. 语言与格式: 使用中文, 严格遵循示例的示例结构, 禁止添加额外字段。

输入参考示例: 原始事实描述: 公诉机关指控: 2016 年 3 月 28 日 20 时许, 被告人颜某在本市洪山区马湖新村足球场马路边捡拾到被害人谢某的 VIVOX5 手机一部, 并在同年 3 月 28 日 21 时起, 分多次通过支付宝小额免密支付功能, 秘密盗走被害

人谢某支付宝内人民币 3723 元。案发后, 被告人颜某家属已赔偿被害人全部损失, 并取得谅解。公诉机关认为被告人颜某具有退赃、取得谅解、自愿认罪等处罚情节, 建议判处被告人颜某一年以下 $\times\times$ 、 $\times\times$ 或者 $\times\times$, 并处罚金你应该输出回答:

[犯罪构成]

主体 = "自然人、成年、无特殊身份"

主观 = "直接故意 (非法占有他人财物)"

客体 = "公私财产所有权"

客观 = "秘密窃取他人支付宝资金 (3723 元)"

[证据特征]

核心证据 = ["支付宝交易记录", "被害人陈述", "退赃凭证"]

证明力 = "电子数据与被害人证言相互印证"

通过明确提取犯罪核心要素和证据核心要素, 不仅优化了案件结构化理解的基础, 也为后续的法律条文检索、司法三段论构建及判决文本生成提供了更为精准和逻辑严谨的输入, 从而提升整个智能判决系统的透明度、准确性和与人类司法实践的一致性。

1.3 相关法律条文检索

为了解决其在法律知识方面的不足和潜在的“幻觉”问题, 相关法律条文检索通过引入真实的法律条文为大语言模型提供准确、权威的法律规范依据, 确保模型在进行判决预测时, 能够明确犯罪的构成要件、法律定义以及量刑法定的边界。首先构建一个包含各类法律法规的法律条文数据库 (DB_{law})。数据库中的每条法律条文均通过文本嵌入模型转换为高维向量表示, 并存入向量数据库以支持高效检索。当处理新的案件时, 从案件事实描述 C 中提取的犯罪特征 F (或其与法律相关的部分, 如初步认定的罪名、关键情节等) 同样被转换为查询向量 $embed(F)$ 。随后, 系统采用近似最近邻 (ANN) 搜索算法, 通过计算查询向量与数据库中法律条文向量之间的相似度 (Sim, 通常采用内积作为度量标准), 检索出与案件最为相关的 k 条法律条文, 形成集合

L 。该过程可表示为:

$$L = (l_1, l_2, \dots, l_k) = \text{Topk}(\text{Sim}(\text{embed}(F), DB_{law}))$$

其中, Topk 表示取相似度最高的 k 个结果。

1.4 相似案例检索

为了解决仅凭法条和案例描述难以覆盖所有复杂情况的问题, 并促进“同案同判”, 本研究提出相似案例检索, 在从历史判例中寻找与当前待审案件在核心特征上相似的案例, 为大语言模型提供司法实践层面的参考。相似案例检索可以让 LLM 理解法律条文在具体情境下的应用方式, 学习既往判决中蕴含的裁判经验和量刑酌情考量。

首先, 构建一个结构化的案例数据库。该数据库中的每个案例都包含详细的字段, 如“罪名类别”、“犯罪构成”, 量刑情节”、“证据特征”、“法律适用”、“裁判逻辑”和“判决结果”的案例。对于案例中的关键文本字段, 特别是“罪名类别”、“犯罪构成”和“证据特征”, 采用文本嵌入模型将其转换为向量表示, 并构建相应的向量索引。需要检索相似案例时, 针对当前案件提取的犯罪特征 F 中的对应字段 (令检索字段集合为 $R = \{\text{罪名, 构成, 证据}\}$, 即 F_{field} (其中 $field \in R$), 分别将其通过相同的文本嵌入模型转换为查询向量 $embed(F_{field})$ 。接着, 对每一个查询向量, 在案例数据库对应字段的向量索引中, 利用近似最近邻搜索算法和相似度计算, 各自独立检索出 m 个最相似的案例。为了得到与当前案件整体最为匹配的案例, 需要对上述各字段检索出的候选案例进行综合评估。对于每一个候选案例 s_j , 计算其与当前案件在 R 中所有字段上的平均相似度

$$\text{AvgSim}(s_j) = \frac{1}{|R|} \sum_{i \in R} \text{Sim}(\text{embed}(F_i), \text{embed}(s_j, i))$$

其中, $\text{embed}(s_j, i)$ 表示候选案例 s_j 在特定字段 i 上的向量表示。最后, 选取平均相似度最高的 n 个案例 ($n < m$) 作为最终的相似案例集合 S

$$S = (s_1, s_2, \dots, s_n) = \text{Topn}(\text{AvgSim}(s_j))$$

通过引入与当前案件高度相似的历史判例,为大语言模型提供了宝贵的经验性知识。这不仅有助于模型更准确地把握特定罪名的构成要件和量刑尺度,还能使其判决建议更贴近司法实践,增强判决结果的合理性和可接受性。

1.5 法律指导的类案融合判决

为了整合前述模块获取的全部信息,有效融合多源异构信息,并进行复杂法律推理的问题,法律指导的类案融合判决将原始案件事实描述 C 、大语言模型从 C 中提取的结构化犯罪特征 F 、从法律条文数据库中检索到的相关法律条文集合 L ,以及从案例数据库中检索到的相似案例集合 S ,共同组织成一个全面的上下文提示词²。

提示词 2:

你是一名法官,负责根据案情描述,参考初步推断,相关法律条文和相似案例,对该案情进行判决,最后你需要按照示例输出的结构的要求生成结构化输出.要求如下:

1. 输出必须严格遵守输出示例的格式要求
2. 数值单位: 判决的刑期结果单位是“月”,也就是“term_of_imprisonment”的“imprisonment”字段.罚金“criminal’s”单位为“元”,“relevant_articles”使用数字字符串,例如“234”,而不是中文表示的数字.
3. 请注意,一个案情描述可能会判决多个罪名,有多个相关法条.因此“accusation”和“relevant_articles”字段的值是数组,即使只有一个罪名和相关法条,类型也是数组,只不过数组只有一个元素.
4. 在进行结果判决时,具体的刑期你需要参考相似案例的判刑情况和法律条文的具体定义.
5. 语言与格式: 使用中文,严格遵循示例的示例输出的结构,禁止添加额外字段.

输入示例:

案情描述: %C

初步推断:

%F

相关法律条文:

%L

相似案例:

%S

输出示例:

```
accusation = ["故意伤害"]
[punish_of_money]
punish_of_money = 0
[term_of_imprisonment]
imprisonment = 6
death_penalty = false
life_imprisonment = false
relevant_articles = ["234"]
```

这个提示被输入到预训练的大语言模型(LLM)中。大语言模型利用其强大的自然语言理解、知识整合和逻辑推理能力,对这些输入信息进行深度分析和融合。模型在推理过程中,会考量法律条文 L 的规定(作为法律依据),并参考相似案例 S 中的裁判思路和判决结果(作为实践经验)。最终,大语言模型生成结构化的判决结果 J ,其内容通常包括建议的罪名、刑期、是否适用死刑或无期徒刑、相关的法律条文编号以及可能的罚金等。该过程可以概念化地表示为:

$$J = LLM(C|F, S, L)$$

其中, $C|F, S, L$ 表示以 F, S, L 为条件上下文信息,结合原始描述 C 进行判决。法律指导的类案融合判决通过整合多源数据和大语言模型的综合推理,实现了对案件事实、法律规范和司法判例的有效融合。它不仅提升了判决预测的准确性和专业性,还通过结合明确的法律条文和相似案例,增强了判决结果的可解释性和说服力。

1.5.1 参考文献

这是参考文献示例。参考文献应遵循 GB/T 7714-2015 标准。引用文献 1?, 文献 2?, 文献 3-5 ???。

致 谢 *致谢内容.*致谢

参 考 文 献

ボハンデ. 過去及び現在に於ける英国と会[J]. 日本時報, 1928, 17: 5-9.

陈晋鏊, 张惠民, 朱士兴, 等. 蓟县震旦亚界研究[M]//中国地质科学院天津地质矿产研究所. 中国震旦亚界. 天津: 天津科学技术出版社, 1980: 56-114.

BRAVO H, OLAVARRIA J. Comparative study of visual inter and intrahemispheric cortico-cortical connections in five

native chilean rodents[J]. Anatomy and embryology, 1990, 181(1): 67-73.

牛志明, 斯温兰德, 雷光春. 综合湿地管理国际研讨会论文集[C]. 北京: 海洋出版社, 2013.

袁训来, 陈哲, 肖书海. 蓝田生物群: 一个认识多细胞生物起源和早期演化的新窗口 – 篇一[J]. 科学通报, 2012, 57(34): 3219.

附录 X.

* 附录内容置于此处, 字体为小 5 号宋体. 附录内容包括: 详细的定理证明、公式推导、原始数据等 *



First A. Author * 计算机学报第 1 作者提供照片电子图片, 尺寸为 1 寸。英文作者介绍内容包括: 出生年, 学位 (或目前学历), 职称, 主要研究领域 (与中文作者介绍中的研究方向一致)。* * 字体为小 5 号 Times New Roman*



Second B. Author * 英文作者介绍内容包括: 出生年, 学位 (或目前学历), 职称, 主要研究领域 (与中文作者介绍中的研究方向一致)。* * 字体为小 5 号 Times New Roman*

Background

* 论文背景介绍为英文, 字体为小 5 号 Times New Roman 体 *

论文后面为 400 单词左右的英文背景介绍。介绍的内容包括:

本文研究的问题属于哪一个领域的什么问题。该类问题目前国际上解决到什么程度。

本文将问题解决到什么程度。

课题所属的项目。

项目的意义。

本研究群体以往在这个方向上的研究成果。

本文的成果是解决大课题中的哪一部分, 如果涉及 863\973 以及其项目、基金、研究计划, 注意这些项目的英文名称应书写正确。