

裁判文书潜在结构对司法判决预测的影响

——以彩礼返还纠纷案件的判决书为例

杨棋^{1,2} 胡威^{3,4*}

1. 西华大学法学与社会学学院 四川成都 610039; 2. 中国政法大学民商经济法学院 北京 100091;
3. 中国科学院成都文献情报中心 四川成都 610299; 4. 中国人民大学统计学院 北京 100872

摘要: 构建司法判决预测模型, 有助于优化类案检索与类案推荐的效果, 从而提高审判质效, 支撑智慧法院的辅助决策支持系统建设。文章结合法律领域特点进行知识表征与推理, 从判决书中提取关键事实信息、地域信息和裁判理由的文本潜在结构, 构建法律主题与文本潜在结构之间的映射关系, 进一步地构建了司法判决预测模型。该模型可用于类案检索、类案推荐, 支撑智慧法院的具体功能设计。实证结果表明: 该模型在测试集上的 AUC 值达到 0.876, 比基线模型提升了 9.9%; 提高了判决的可解释性; 并揭示了彩礼返还案件的地域差异性。研究验证了模型的预测准确性及可解释性, 为类案检索与推荐提供了新的思路。

关键词: 司法判决预测; 类案检索与推荐; 智慧法院; 辅助决策支持系统; 文本主题模型

中图分类号: D926 **文献标志码:** A **文章编号:** 1672-8505(2025)02-0086-18

doi: 10.12189/j.issn.1672-8505.2025.02.009

The Influence of the Potential Structure of Judicial Documents on the Prediction of Judicial Judgments

—Taking the Judgments of Bride Price Refund Dispute as an Example

YANG Qi^{1,2} HU Wei^{3,4*}

1. School of Law and Sociology, Xihua University, Chengdu, Sichuan, 610039, China;
2. School of Civil, Commercial and Economic Law, China University of Political Science and Law, Beijing, 100091, China;
3. National Science Library (Chengdu), Chinese Academy of Sciences, Chengdu, Sichuan, 610299, China;
4. School of Statistics, Renmin University of China, Beijing, 100872, China

Abstract: This article constructs a judicial judgment prediction model that optimizes similar case retrieval and recommendation, improves the quality and efficiency of trials, and supports the design of an intelligent court system. It combines the characteristics of the legal field for knowledge representation and reasoning, incorporates specialized legal knowledge, and extracts objective factual information, geographical details, and the latent text structure from legal judgments. A mapping between legal topics and the latent text struc-

收稿日期: 2024-11-27

基金项目: 中国科学院特别研究助理资助项目(项目编号: E3Z0000804); 中国科学院成都文献情报中心创新基金青年项目“概率图模型视角下关键核心技术发展态势感知与预测”(项目编号: E4C0001302)。

第一作者: 杨棋, 女, 讲师, 博士, 主要从事民法总论、婚姻家庭法研究。

* 通信作者: 胡威, 男, 助理研究员, 博士, 主要从事文献计量学、复杂网络数据分析研究。

引用格式: 杨棋, 胡威. 裁判文书潜在结构对司法判决预测的影响——以彩礼返还纠纷案件的判决书为例[J]. 西华大学学报(哲学社会科学版), 2025, 44(2): 86-103.

ture is established. Based on these features, a judicial decision prediction model is constructed to conduct similar case retrieval and recommendation, further supporting the functional design of an intelligent court system. Experimental results show that the judicial decision prediction model achieves an AUC value of 0.876 on the test set, a 9.9% improvement over the baseline model. The proposed model enhances the interpretability of judgments, and significant regional variations are observed among the "bride price refund" cases. The experimental results confirm the predictive accuracy and interpretability of the proposed legal judgment prediction model. This research provides a new tool for similar case retrieval and recommendation.

Key words: legal judgment prediction; similar case search and recommendation; intelligent court; decision support system; text topic model

司法判决是法律适用的具体结果,代表了抽象法律规范对具体现实案件的妥当性处理。习近平总书记在中国共产党第二十次全国代表大会上的报告、《法治中国建设规划(2020—2025年)》等一系列重要文件均指出,“努力让人民群众在每一个司法案件中都感受到公平正义”^[1-3]。法律规范中的公平正义通过司法判决的内容呈现。司法判决作为纠纷处理的最终结果,其核心要义在于根据法律规范,公平、公开地保护各方当事人的权利义务。司法判决一般以裁判文书的形式呈现。2013年随着《最高人民法院裁判文书上网公布暂行办法》开始实施,中国裁判文书网开始正式使用。截至2021年,全国范围内的电子卷宗整合率超过95%^[4]。海量的裁判文书数据的电子化记录,使得司法判决的可计算化的研究成为可能^[5-7]。

近年来,最高人民法院指导建设了中国司法大数据服务网。2021年最高人民法院发布《人民法院信息化建设五年发展规划(2021—2025)》,提出“十四五”时期加强法院审判能力现代化,要求进一步加强构建司法数据平台,以信息化的方式促进司法现代化,全面加强智慧法院建设,以形成统一的管理系统、智能化辅助工具^[8],以裁判文书网为载体的司法数据构成了当代中国法律大数据的基本源泉^[5]。

司法判决预测是法律文书自然语言理解的一项重要任务^[9],包括刑期预测、案件分类等,后者也称司法判决分类。如何融合法律基础与司法解释、推理模式、特定民事领域知识等特征,从而构建预测精度高,结果可解释的判决预测模型,是现有研究面临的重要挑战^[10]。现有的机器学习、深度学习等方法已经在司法判决预测领域达到了较高的预测精度^[11-13],但是由于对裁判文书的潜在特征提取不足,使得模型依然存在可解释性不够的问题^[10],对于法律逻辑的理解尚不深入。模型的可解释性包括:预测结果会受到哪些因素的显著影响,各影响因素的影响程度,各影响因素在法律领域的实际含义。

针对上述问题,本文构建了一种司法判决预测模型。首先,本文根据法律用语辞典及该领域案件关键词,对裁判文书的内容进行关键事实和地域信息的识别。然后,使用文本主题模型对于裁判理由进行主题建模,由此提取潜在结构信息,构建法律主题与文本潜在结构之间的映射关系。基于此,本文依据双固定效应逻辑回归构建司法判决预测模型,分析司法判决的关键影响因素,以及探究省份间差异对判决的影响。最后,本文基于所构建的模型,提出支持智慧法院辅助决策支持系统的功能设计,从不同受众的角度提供辅助决策支持。

本文的主要贡献有以下三点。

其一,本文从法理基础出发,结合法律行业关键词进行裁判文书中关键的客观事实信息、地域信息和潜在结构信息的提取,构建了法律主题与文本潜在结构之间的映射关系,赋予了文本结构以法律解释含义,为司法判决预测模型奠定了数据基础。

其二,本文构建了司法判决预测模型,在保证预测准确度的情况下,提高了模型的可解释性。该模型将关键事实信息、地域信息、法律主题信息作为输入特征,以此探究了司法判决的影响因素,为理解司法判决和检验同案同判提供了新方法视角。

其三,本文根据所构建的司法判决预测模型的研究结果,进行智慧法院辅助决策支持系统的功能设

计。利用关键词匹配、分类模型集成方法、文本相似性指标、聚类相结合的方法,提出面向法院的案件分类管理与公平性检验功能,面向基层法官的案例推荐与辅助支持判决功能,和面向群众的以案释法功能。

本文安排如下:第一章介绍相关工作,第二章介绍变量定义和司法判决预测模型;第三章介绍数据分析结果,给出了法律主题的分类结果,以及验证了所提出的判决预测模型的效果;第四章是面向实际应用场景的辅助决策支持系统的功能设计;第五章是总结与展望;限于篇幅,本文实验数据的描述性分析与彩礼返还纠纷的法律基础见第六章附录部分。

一、相关工作

已有的裁判文书实证研究包括命名实体识别、司法判决预测模型、案件辅助判决与管理系统,这些研究为智慧法院辅助决策支持系统的建设提供了重要支持^{[14][15]}。

裁判文书中的命名实体识别^[16]指的是识别文本中有特定意义的实体,包括提取法条、证物等专有名词,已有的命名实体识别的方法可以分为两类^[17]。第一类是基于规则的方法^{[18][19]},通过建立词典通过正则表达式进行关键词匹配,这一类方法优点在于对于规范性较强的文档识别准确率较高,缺点在于需要构建针对特定案例的词典;第二类是基于机器学习或深度学习的方法^[20-22],如条件随机场、随机森林模型、循环神经网络、预训练语言模型等,这一类方法的优点是无须建立专用的词库,缺点是需要对识别结果进行人工标注。高效准确的实体识别方法为司法大数据的建模分析和应用提供了支持^[23]。

目前,利用人工智能算法在司法判决预测领域中的应用以刑事领域为主,如量刑预测、信用卡诈骗识别,而针对民事案件的较少。对于刑法领域的量刑预测问题,舒洪水^[23]、谭红叶等^[24]、白建军^[25]、文姬^[26]、王芳等^[27]探究了基于过往案例数据进行量刑预测的可能性,以及探究了适用的统计、数学模型。值得注意的是,对于民事领域而言,不同子领域的立法精度存在着显著差异,比如商事领域的规范往往具备更强的技术性,立法精度更高^{[28][29]},而婚姻家庭领域向来秉持着“宜粗不宜细”的立法精神,立法并未对于此类案件进行过于细致的规定。作为影响婚姻家庭的重要财产行为,目前立法上存在规范供给不足的问题。在婚姻家庭案件中,法官往往拥有更大的自由裁量权,同类案件的判决结果存在差异较大的情况。因此,婚姻家庭领域的司法判决预测亟待针对性的研究。

对于案件辅助判决与管理系统的建设研究,已有研究基于各类裁判文书的数量、案由、性质等内容特征进行了研究,如左卫民^[30]探究了诉前调解机制的可行性;朱青等^[31]研究了基于诉讼文本内容的判决预测;石勇等^[32]研究了案件的分类管理系统以支持检察官进行起诉罪名的决策。此外,在内容特征的基础上结合对裁判文书结构的分析,洪文兴等^[33]探究了司法案情知识图谱的构建,杨超群等^[34]提出了案件适用法条的推荐模型,梁柱等^[35]提出了高相关的裁判文书的推荐系统。各地法院借助互联网技术不断深化司法智能化改革,相继推出各项智慧司法举措,构建类案检索系统、司法问答系统,从而实现案件的辅助判决与管理。例如北京地区的“睿法官”、苏州法院的“智慧审判苏州模式”、浙江金华法院的“类案大数据分析平台”,均为司法信息化在我国审判实践中产生的成功经验^[36]。技术的不断发展使得对裁判文书的分析不断深化,也为司法辅助决策支持系统的建设创造了条件。

上述已有研究通过分析裁判文书的内容与结构,研究司法判决预测模型,构建案件辅助判决与管理系统。然而,已有研究对于裁判文书的特征提取和分析方面尚有不足,体现在地区差异和细粒度的法律主题这两个方面。

其一,由于我国幅员辽阔情况复杂,地区之间往往存在显著差异,体现在经济^{[37][38]}、金融^[39]、文化^{[40][41]}、环境^{[42][43]}等方面,在构建与解释判决预测模型时需考虑量化表达公平正义等法学基本问题^[44],比如企业知识产权纠纷问题中存在的司法地方保护主义^[45]。

其二,现有研究对于裁判文书结构的分析集中于裁判文书的整体结构,如案情描述、裁判理由、裁判结果、相关法条。但是,裁判文书的整体结构的粒度不足以刻画判决影响因素。根据最高法对于裁判文

书撰写的指导意见^[46]，裁判理由内容显示了法官对特定类型案件的个性化考量，其阐释裁断的形成过程蕴含了更为细粒度的法律主题。因此，现有研究的司法判决预测模型对于判决的法律逻辑的理解尚不充分。

以上两点使得司法判决预测模型的可解释能力存在不足。本文将省份及当地风俗差异特征融入模型，并构建细化的法律主题与文本潜在结构之间的映射关系，从而构建司法判决预测模型，以反映法官的真实考量因素的细节。

二、变量定义和司法判决预测模型

(一) 变量定义

前文中指出在婚姻家庭案件中，法官常有较大的自由裁量权，使得同类案件的判决结果可能差异显著。因此，本文聚焦婚姻家庭领域的司法判决预测研究。以彩礼返还案件为例，对相关变量进行定义。记因变量 Y 表示裁判结果：是否判决返还彩礼，由判决书的判决结果部分提取。对于第 i 份判决书，如判决为返还，则记 $Y_i = 1$ ；否则判决为不返还，记 $Y_i = 0$ 。值得注意的是，常见的作为彩礼的物品如“三金”，以及嫁妆如家具，这些物品由于来源不一、折旧等因素难以衡量其真实市场价格，因此本文未考虑彩礼返还的比例作为因变量。

本文考虑的用于构建司法判决预测模型的自变量包括三部分，分别是：关键事实信息，地域信息，和裁判理由的法律主题，如表 1 所示。自变量均由判决书文本提取得到。三部分自变量的提取过程具体解释如下。

表 1 变量的详细说明

变量类别	变量名称	变量类型	变量定义与取值范围
因变量	是否返还	0—1变量	判决返还，则为1；不返还则为0
关键事实信息	案件标的额	多分类变量	包括五类：未提及、0—10万、10—50万、50—100万、100—500万
	同居时间	0—1变量	同居时间较短；时间较长或未提及
	曾经怀孕	0—1变量	曾怀孕；未怀孕或未提及
	育有子女	0—1变量	有子女；无子女或未提及
	举办过婚礼	0—1变量	举办过婚礼；未举办或未提及
	办理结婚登记手续	0—1变量	未办理；已办理或未提及
地域信息	省份	多分类变量	河南、甘肃、安徽、江西、江苏、山东
	考虑当地风俗	0—1变量	办理过；未办理或未提及
裁判理由的法律主题	原告主张与原被告争议	连续型变量	裁判理由对此方面的重视程度，可根据实际数据得到的二级法律主题进行细分，取值范围0~1
	证据认定	连续型变量	同上
	影响返还的具体因素	连续型变量	同上
	调解情况	连续型变量	同上
	相关法律基础	连续型变量	同上
	法律程序表述	连续型变量	同上

注：对于分类变量，在变量定义与取值范围一列中，最后介绍的取值记为基准组，比如对于“办理结婚登记手续”而言，“已办理或未提及”为基准组

关键事实信息包括：案件标的额、同居时间、是否曾经怀孕、是否育有子女、是否举办过婚礼、是否办理结婚登记手续。裁判文书的结构和表达较规范，利用《法律术语辞典》，这类自变量通过关键词及正则表达式匹配的方式提取。值得注意的是，特征提取过程中发现法官常会将“同居时间较短”作为裁判理由，而“同居时间较长”却未在裁判理由之中出现，故进行了表中所述的变量定义方式。

地域信息包括：所属省份，和裁判理由中是否考虑了当地风俗，后者反映了法官对于当地特色的彩礼风俗的重视程度，其中“当地”指代当事人所在的乡镇、县、市等区域的模糊概念。以变量“考虑当地风

俗”为例,根据“当地风俗”“本地的风俗”“当地习惯”等关键词,并结合中文语料库的BERT预训练模型进行近义词自动化合并,由此通过构造正则表达式对裁判理由段落文本进行匹配提取特征。

裁判理由的法律主题信息反映了法官在作出判决时对各方面因素的看重程度。裁判理由通常包括以下六个一级法律主题:原告主张与原被告争议、证据认定、影响判决(返还)的具体因素、调解情况、相关法律基础、法律程序表述。根据最高法对于裁判文书撰写的指导意见^[46],和“彩礼返还”案件的领域特点,可以细化法律主题。本文利用主题模型提取裁判理由的潜在结构,然后将潜在结构对应的高频主题词分布与《法律术语词典》及领域关键词进行匹配,从而构成映射关系,得到一级、二级法律主题。主题模型的具体细节如下:

主题模型是一种常见的无监督式的文本挖掘方法,假设“文档”(本研究中,即判决书中的裁判理由)和“词语”之间存在一个隐藏层,称之为“主题”,这里即法律主题。主题建模是一种生成式的模型,对于每个文档,其中的词语通过“以一定概率选择了某个主题,并从这个主题中以一定概率选择某个词语”的过程生成得到^[47]。以第*i*份判决书为例,假设其由*K*个法律主题构成,则对于词汇*j*的生成概率的形式可以表达如下:

$$P(\text{词汇}_j|\text{判决书}_i) = \sum_{k=1, \dots, K} P(\text{法律主题}_k|\text{判决书}_i)P(\text{词汇}_j|\text{法律主题}_k)$$

其中词汇*j*表示第*j*个词汇,法律主题*k*表示第*k*个法律主题,判决书*i*表示第*i*份判决书。而在实际数据中,法律主题的个数未知,各法律主题有若干对应的术语及领域关键词。

这里本文考虑潜在狄利克雷分布模型^[48-50](latent Dirichlet allocation,LDA)。LDA通过基于狄利克雷先验的贝叶斯学习方法,得到各文本主题在某文档中的概率分布,和各词语在某文本主题中的概率分布。此外,可以通过困惑度值确定LDA模型的主题个数^[48]。困惑度值越低代表模型的拟合效果越好,一般来说,主题个数越高,困惑度值越低。裁判理由的文本结构如图1所示,其中文档层和词语层是可观测的,而主题层作为隐藏层是不可观测的,主题层的结构可根据实际计算得到的文本主题进行比表1更为细致的刻画和调整。通过计算主题词概率分布与法律术语词典及领域关键词间的匹配程度,从而按匹配程度从高到低进行排序,构成了法律主题与文本潜在结构(即文本主题)之间的映射关系。最后,通过各法律主题在裁判理由中的概率分布,可表示该判决对各方面因素的重视程度。

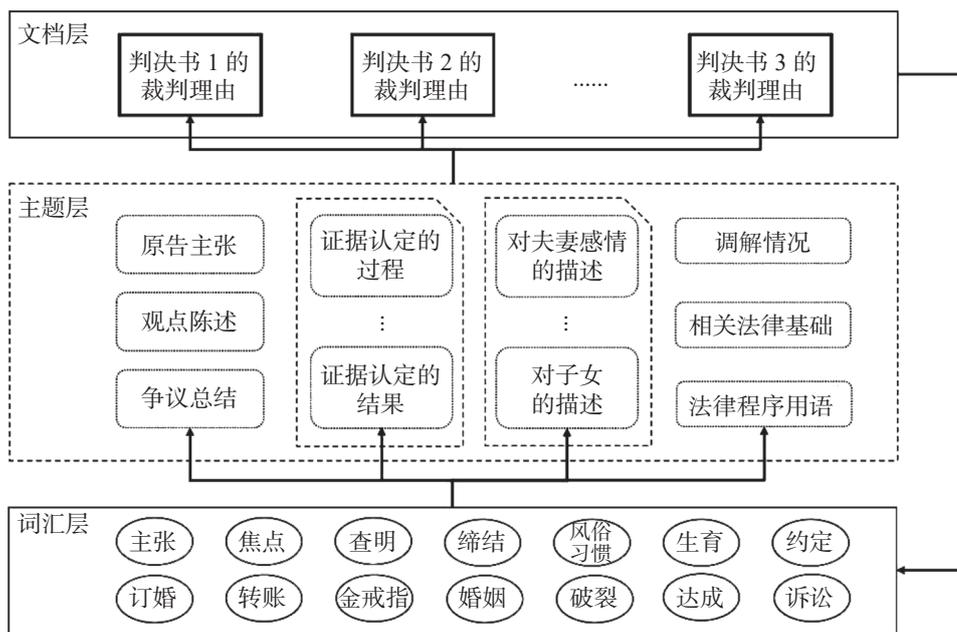


图1 裁判理由的主题结构图

(二) 司法判决预测模型

基于裁判文书的背景,和上述三部分自变量:关键事实信息,地域信息,和裁判理由的法律主题信息的分析,本文依据逻辑回归模型的形式并结合领域先验知识,首先构建不含交互项的司法判决预测模型形式如下:

$$P(Y_i = 1) = \text{Sigmoid}\{Year_i + Provence_i + Fact_i\beta + Costom_i\theta + Topic_i\gamma\}$$

其中 $P(Y_i = 1)$ 表示第 i 份判决书的裁判结果为“返还”的概率, $\text{Sigmoid}(x)$ 表示函数 $\exp(x)/(1+\exp(x))$, x 为标量, $Year_i$ 表示年份固定效应, $Provence_i$ 表示判决书 i 的省份固定效应, $Fact_i$ 表示第 i 份判决书所蕴含的关键事实信息, $Costom_i$ 表示第 i 份判决书是否考虑了当地风俗等地域信息, $Topic_i$ 表示第 i 份判决书的裁判理由在各法律主题上的分布权重得分,参数 β, θ, γ 是对应自变量的系数向量,分别表示关键事实信息、地域信息和裁判理由的法律主题信息对判决结果的直接影响程度。

在法律实证研究中,常利用逻辑回归模型等统计方法进行司法判决的影响因素分析和公正性检验^{[30][51][52]}。对于逻辑回归模型中的参数,一般使用极大似然算法进行估计。记不含交互项的司法判决预测模型形式为 $P(Y_i = 1) = \text{Sigmoid}\{\Theta_i\}$,参数的估计值为:

$$(\hat{\beta}, \hat{\theta}, \hat{\gamma}) = \text{argmax}_{\beta, \theta, \gamma} \prod_{i=1}^n \left\{ \frac{\exp(\Theta_i)}{1 + \exp(\Theta_i)} \right\}^{Y_i} \left\{ \frac{1}{1 + \exp(\Theta_i)} \right\}^{1-Y_i}$$

其中 n 表示样本量,右侧的连乘式被称为似然函数 $L(\beta, \theta, \gamma)$ 。然后,根据参数含义及估计结果可以度量其对于司法判决预测模型的影响。(1)系数的大小表示自变量的变化对于判决的影响程度。具体解释为:保持其他因素不变的情况下,自变量每上升一个单位时,返还的对数优势比率(Log-odds Ratio)上升的倍数,即判决返还与不返还的可能性的比值的对数 $\log[P(Y_i = 1)/P(Y_i = 0)]$ 。比如,以特征“考虑当地风俗”为例,若其对应的系数估计值 $\hat{\theta}$ 为0.5且通过了显著性检验,则解读为:在保持其他因素不变的情况下,该省份考虑当地风俗比不考虑的案件,判决返还的可能性增大 $\exp(0.5)/(1+\exp(0.5)) = 62.2\%$ 。(2)系数的显著性解读为,在设定的水平下该自变量对于是否返还具有显著的解释/预测力。

为了细化描述不同省份的判决影响因素的差异性,本文考虑在司法判决预测模型的设定中加入省份与其他自变量之间的交互项,模型形式如下:

$$P(Y_i = 1) = \text{Sigmoid}\{Year_i + Provence_i + Fact_i\beta + Costom_i\theta + Topic_i\gamma + Provence_i \times (Fact_i\beta_{pro_i} + Costom_i\theta_{pro_i} + Topic_i\gamma_{pro_i})\}$$

其中 $Provence_i \times Fact_i$ 表示省份与关键事实信息的交互项,参数 $\beta_{pro_i}, \theta_{pro_i}, \gamma_{pro_i}$ 表示相应省份的交互项所对应的系数向量,下标 pro_i 即表示对应的省份,使用极大似然算法得到各个系数的估计结果。加入交互项特征后,模型中的自变量较多,因此采用BIC准则进行变量选择,将BIC值最小的变量子集构成最终的司法判决预测模型,从而提高最终模型的泛化能力和可解释性。最终模型的估计结果中,若存在某省份对应的参数估计值 $\hat{\beta}_{pro_i}, \hat{\theta}_{pro_i}$ 、或 $\hat{\gamma}_{pro_i}$ 显著不为0,则表示该省份受对应特征的影响程度与基准省份存在显著差异。

三、数据分析与结果

(一) 数据集

本文使用判决书文本数据,数据来源于“威科先行—TM法律信息库”数据库。本文以“彩礼返还”为检索关键词,然后对案由、审判程序、法院级别、地区、文书类型进行选择。判决书是这类案件裁判文书的主要文书类型。本文选择彩礼返还民事案件的基层法院一审判决书,地区跨度上选择彩礼返还案件

频率最高的六个省份,分别为河南、安徽、甘肃、江西、江苏、山东。由以上步骤,本文共收集得到6096份裁判文书。这些判决书的时间跨度为2009年1月至2023年3月。每份判决书的核心要素包括:案例编号、案例名称、审理法院、所属省份、审判日期、原被告及其代理律师、审理经过、标的额、受理费、裁判理由、裁判结果、审理人员。

(二) 数据集的法律背景

彩礼是中国传统中重要的婚俗习惯之一,意为婚聘时男方以结婚为目的,对女方进行金钱或实物赠与。男女双方因未办理结婚登记、感情破裂办理离婚等原因,男方诉请女方返还彩礼,因此产生纠纷请求法院判决。彩礼返还的请求权基础在于“给付目的嗣后不存在”的不当得利^[53]。随着近年来彩礼数额的不断增加,某些地区甚至出现“天价彩礼”现象,使得彩礼返还纠纷的数量不断增加。与婚姻家庭领域的其他纠纷相比,彩礼返还案件具有更强的地方性和风俗性,甚至于“十里不同风,百里不同俗”,法官在判决案件时往往涉及更多个体因素的判断^[54]。

彩礼返还的司法流程一般如下:给付彩礼的男方提起诉讼后,法院先对双方进行调解,若调解成功将出具调解书,若调解不成将继续进行裁判。法院依据相关法律法规和具体案情中的各项考量因素,说明裁判理由,得出判决结果。最后,关于彩礼返还纠纷及其法律基础的详细解释可参见附录。

(三) 基线模型的实验结果

基线模型是根据计算到的关键事实信息、地域信息构建逻辑回归模型,进行司法判决预测,而未考虑裁判理由的法律主题信息以及省份效应的交互项。将是否返还作为因变量,对模型进行BIC变量选择,基线模型的结果如表2所示,本文采用测试集上的AUC值进行预测效果评价。其中BIC(贝叶斯信息准则)是一种用于模型选择的统计量,它通过惩罚模型复杂度来帮助选择最佳模型;AUC(Area Under the Curve,曲线下面积)是评估分类模型性能的一个指标,取值在0到1之间,表示模型在区分正负样本时的综合性能,值越接近1表示模型性能越优。由表2,可得判决预测的影响因素结果如下。

表2 基线模型结果

自变量	六个省份	甘肃省	河南省	江西省	安徽省	山东省	江苏省
考虑当地风俗	0.585***	0.878**	0.854***	###	0.916**	###	###
标的额0—10万	0.921***	1.427***	0.623***	1.247***	0.767***	1.349***	###
标的额10—50万	1.494***	2.016***	0.934***	1.934***	1.358***	1.780***	###
同居时间较短	0.821***	1.091***	0.573**	0.990***	1.059***	###	1.048***
曾经怀孕	###	###	###	###	###	###	1.650*
育有子女	-0.974***	###	-1.151***	-1.556***	-1.131***	-1.127**	-1.216***
未办理结婚登记	1.156***	###	1.183***	1.468***	0.895***	1.429***	2.487***
裁判理由的法律主题信息	否	否	否	否	否	否	否
省份固定效应	是	—	—	—	—	—	—
年份固定效应	是	是	是	是	是	是	是
样本量	6096	941	1728	889	987	745	806
测试集AUC值	0.797	0.785	0.744	0.835	0.761	0.818	0.846

注:是否举办过婚礼这一特征在各回归模型中系数均不显著,故省去这一行。系数上标*、**、***分别代表在0.05、0.01、0.001水平下显著不为0。

###表示该项在BIC模型选择的过程中没有被纳入最终模型。

在影响程度方面,对于甘肃、河南、安徽三省,当地风俗对于是否返还存在显著影响,考虑当地风俗的案件返还的可能性更高。从定量的角度,以河南省为例,保持其他条件不变的情况下,考虑当地风俗的

案件比不考虑当地风俗的案件判决未返还的可能性增大 $\exp(0.854)/(1+\exp(0.854))=70.1\%$ 。

从影响因素的显著性的角度,将六个省份样本的模型结果,与各省份内部样本的模型结果进行对比,可以看出,具有显著解释力的自变量的符号是相同的。比如,“同居时间较短”和“未办理结婚登记”自变量的回归系数均显著为正,这与法律基础及司法解释的要求相吻合。这体现出模型符合法律裁判的逻辑,具有可解释性,自变量对应的法律含义合理。

(四) 法律主题与文本潜在结构之间的映射关系

本文利用主题模型提取裁判理由的文本潜在结构,并与法律主题之间建立映射关系。设置主题个数在 1—30 之间并根据拟合的困惑度进行选择,归一化后的困惑度值如图 2 所示。由图 2 可知,当主题数目大于等于 14 时,困惑度值的曲线相对平稳。根据困惑度值、法律术语及领域关键词的匹配程度,最终选择主题个数为 19。

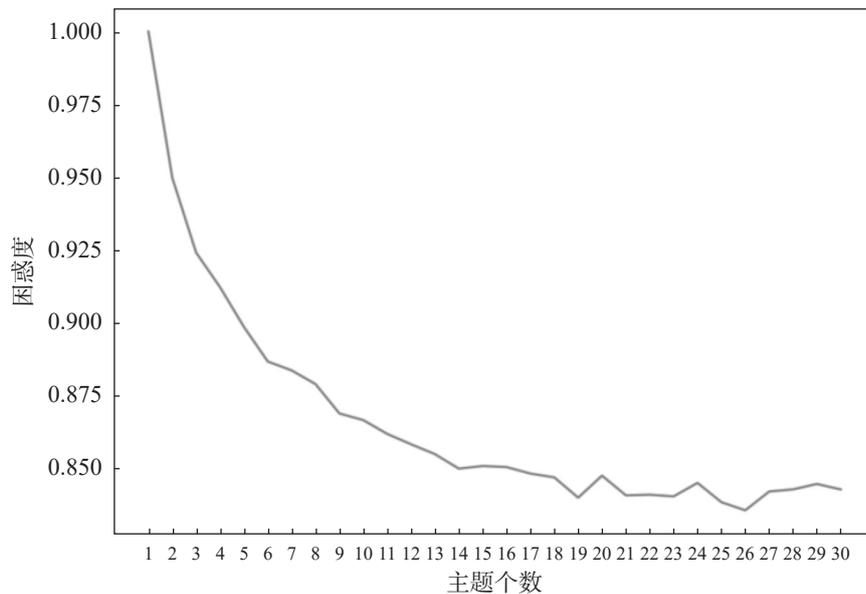


图 2 不同主题数目的困惑度值

一级法律主题包括主张与争议、证据认定、影响判决(返还)的具体因素、调解情况、相关法律基础、法律程序表述共六大类,如表 3 所示。各一级法律主题包含若干个细分的二级法律主题,表 3 给出了各二级法律主题在实际数据集上的前 10 个高频主题词,这一结果与根据经验的裁判理由主题结构相吻合。

表 3 一级、二级法律主题及其主题词

一级法律主题	二级法律主题	前 10 个高频主题词
主张与争议	原被告观点陈述	被告方,原告方,订婚,实际,现金,打发,收取,辩称,农村,解除婚约
	原被告争议点的总结	万元,争议,焦点,是否,借款,主张,支付,车辆,认定,房屋
	原告的主张	主张,提供,认可,物品,证明,购买,辩称,陪嫁,数额,采信
证据认定	已查明事实	年月日,转账,期间,微信,款项,支付,恋爱,主张,提交,认定
	证据认定的过程	证人,证言,陈述,证明,媒人,事实,认定,提供,原被告,认可
	原告的举证	证明,主张,提供,承担,事实,诉讼请求,责任,举证,依据,后果
影响返还的具体因素	原告生活困难情况	情形,查明,困难,以下,属于,导致,办理,婚前,人民法院,但确
	彩礼物品的认定	价值,金项链,戒指,购买,金戒指,黄金,一枚,一条,现金,项链
	彩礼法律性质分析	缔结,数额,目的,当地,婚姻,风俗习惯,财物,一方,时间,因素

续表 3

一级法律主题	二级法律主题	前 10 个高频主题词
影响返还的具体因素	结婚仪式与登记情况的考量	同居, 关系, 举行, 法律, 结婚仪式, 原被告, 结婚登记, 农村, 应予, 第一项
	原被告家庭情况	女方, 父母, 男女双方, 家庭, 接受, 男方, 主体, 涉及, 之间, 缔结
	对夫妻感情的描述	离婚, 原被告, 夫妻感情, 婚后, 婚前, 第三十二条, 破裂, 夫妻, 确已, 时间
	不认定为彩礼的婚前开销	属于, 费用, 赠与, 购买, 主张, 见面礼, 订婚, 花费, 支出, 范畴
	认定为彩礼的婚前开销	赠与, 行为, 目的, 条件, 成就, 缔结, 婚姻关系, 法律, 男女双方, 财产
	对子女的描述	子女, 抚养, 生育, 孩子, 抚养费, 女儿, 支付, 承担, 精神, 小孩
调解情况	调解情况	协议, 法律, 表示, 达成, 约定, 意思, 履行, 原被告, 调解, 退还
相关法律基础	支持返还的法律基础	第五条, 婚姻家庭, 民法典, 第一千, 四十二条, 数额, 第六十七条, 时间, 款元, 综合
	支持返还的法律解释	婚姻, 财物, 禁止, 索取, 第三条, 收取, 数额, 法律, 行为, 收受
法律程序表述	法律程序表述	诉讼, 一方, 到庭, 一百四十四, 视为, 权利, 解除, 放弃, 传唤, 参加

影响返还的具体因素中, 细分的二级法律主题包括: “结婚仪式与登记情况的考量”“对夫妻感情的描述”“原告生活困难情况”, 这体现了实验得到的裁判理由与法律基础及最高法出台的司法解释的吻合性。

裁判理由因特定领域而被表述得更为具体, 实验结果得到了比法律基础及司法解释更为细粒度的影响因素。对于共同生活时间较短、认定为彩礼的婚前开销情况的表述占比较高, 则往往也是支持返还的条件; 相对应地, 育有子女、不认定为彩礼的婚前开销的表述占比较高, 则往往是不支持返还的条件。

裁判理由体现了与法律基础并非直接相关的影响因素。由于部分案件的地域性更强, 因此裁判理由中会包括依据当地风俗习惯进行的彩礼法律性质分析, 当地风俗会影响婚前开销及实体物品的彩礼认定。

综上, 实验结果表明, 裁判理由所讨论的事实内容和法律主题, 提供了比法律基础和司法解释更为丰富的生活细节及地域性含义, 更细粒度地阐释了裁判理由的内涵。

(五) 司法判决预测模型的实验结果

对于司法判决预测模型, 该模型将裁判理由的法律主题信息自变量进行标准化处理, 并加入省份地域特征的交互项。经 BIC 准则变量选择后, 最终模型的结果如表 4 所示。接下来, 本文首先进行模型效果的比较, 然后从可解释司法判决预测的角度, 分别从系数含义解释、省份地区差异、司法公正性、当事人的细分角度, 结合对于表 2 的分析进行模型实证结果的分析解读。

表 4 司法判决预测模型的结果

自变量	六个省份	甘肃省	河南省	江西省	安徽省	山东省	江苏省
同居时间较短	0.620***	0.865***	0.790***	0.904**	0.935**	0.883**	###
未办理结婚登记	0.533***	###	0.602**	###	###	###	-1.813***
原被告观点陈述	0.421***	###	###	###	###	###	###
原被告争议点的总结	###	###	0.531***	###	###	-0.354***	###
原告的主张	0.572***	0.582***	0.673***	0.813*	0.624***	###	0.434**
已查明事实	###	###	###	###	###	###	###
证据认定的过程	0.251***	0.738***	0.327***	###	0.369**	###	###
原告的举证	###	###	0.199***	###	###	-0.517***	###
原告生活困难情况	0.267***	###	0.419***	###	###	###	###

续表 4

自变量	六个省份	甘肃省	河南省	江西省	安徽省	山东省	江苏省
彩礼物品的认定	0.421***	0.504***	0.382***	###	0.373**	###	0.460**
彩礼法律性质分析	0.494***	0.545***	0.460***	###	0.825***	-0.690***	0.714***
结婚仪式与登记情况的考量	0.570***	0.652***	0.535***	###	0.773***	###	###
原被告家庭情况	0.372***	###	0.624***	###	###	###	###
对夫妻感情的描述	###	###	###	-1.100***	###	-1.232***	###
不认定为彩礼的婚前开销	0.627***	0.738***	0.541***	0.629**	0.810***	###	###
认定为彩礼的婚前开销	0.581***	###	0.774***	###	0.750***	###	###
对子女的描述	-0.208***	###	###	-0.641***	###	-1.004***	-0.487***
调解情况	###	###	###	-0.514***	###	-0.896***	###
支持返还的法律基础	0.269***	###	0.317***	###	1.068***	###	###
支持返还的法律解释	0.762***	###	0.843***	###	1.072***	###	###
法律程序表述	0.432***	0.712***	0.401***	###	1.226***	###	0.332***
省份固定效应	是	—	—	—	—	—	—
年份固定效应	是	是	是	是	是	是	是
样本量	6096	941	1728	889	987	745	806
测试集AUC 值	0.876	0.836	0.850	0.896	0.877	0.907	0.863

注：为保证预测结果可比性，表4与表2中各模型AUC值之间的比较是基于相同的测试集。

从模型预测效果的角度，对于六个省份的样本，判决预测模型在测试集上的 AUC 值达到 0.876，比基线模型的 AUC 值 0.797 提升了 9.9%；对于各省份内部样本而言，判决预测模型的 AUC 值比基线模型提升了 2.0% 至 15.2%。这表明本文提出的融合法律文本潜在结构的司法判决预测模型的泛化预测能力更强，所得到的预测结果更为可信。

从模型系数含义解释的角度，随着法律主题信息的加入，地域信息和一些关键事实信息的回归系数变得不再显著。比如，对彩礼物品进行认定的过程需要依据当地风俗，例如安徽省淮北市的“红衣款”、甘肃省天水市的“追节礼”等，更多细节解读见附录（六）。因此，当考虑了“彩礼物品的认定”法律主题信息作为自变量时，“考虑当地风俗”自变量对于判决的影响便不再显著。以及“育有子女”这一关键的客观事实信息，与“对子女的描述”的法律主题含义相同。表 4 中有显著影响的特征与 2023 年 12 月 11 日最高人民法院举行的“推进移风易俗 治理高额彩礼”新闻发布会所公布的四个典型案例的细节高度吻合，包括本文所提取的关键事实信息，生活情况，彩礼物品及开销的举证认定等。这表明该模型分析得到的司法判决影响因素具有典型意义，提高了司法判决的可解释性。

从省份地区差异的角度，可以发现不同省份的裁判理由的法律主题对于是否返还的主要影响因素存在差异。以江苏省与河南省的对比为例，河南省在彩礼返还判决中更多侧重于风俗习惯，考量原被告争议以及具体开销的彩礼认定；而江苏省则重点考虑结婚登记等与法律基础及司法解释相匹配的特征，这一差异或与当地习俗及经济发展水平相关。

当前在省级层面的司法裁判结果未出现违反司法公正性的问题。尽管各省份判决的主要影响因素存在差异，但是 BIC 准则得到的最终模型不包含各省份与其他自变量的交互项，这表明交互项对于判决结果没有显著的解释力。因此，可认为不同省份的差异是当地经济文化等客观条件使然，实验结果不能

拒绝该原假设“省份差异对于判决结果有显著影响”。

从当事人的角度,该司法判决预测模型能够给出比基线模型更为细节切实的指导建议:建议男方保存易被法院认定为彩礼的财物支出相关证据(例如购买“三金”的发票),建议女方保存易被法院认定为赠予行为的相关证据(例如特定节日、特定数额的红包赠予),从而利于己方举证,保护自己的合法权益。

此外需要指出的是,在表2和表4中,尽管部分省份的自变量在BIC模型选择过程中未被纳入最终模型,但这并不影响模型的整体准确性。这是因为BIC模型选择的核心目标是通过惩罚模型复杂度来避免过拟合,从而提高模型的泛化能力。因此,未被纳入的自变量对模型的预测能力贡献较小,或其信息已被模型中的其他变量所涵盖。因此,即使这些变量未被包含在最终模型中,其对于模型在测试集上的AUC值影响不显著。

(六) 稳健性检验

本文进行了以下几个方面的稳健性检验。

关于交互项及自变量选择的问题,本文尝试将年份与其他自变量的交互项纳入判决预测模型,以及加入对应的基层法院名称作为固定效应项,均得到了与基线模型表2、司法判决预测模型表4的结论相一致的结果。

关于样本不平衡问题,本文计算了模型在不同阈值下的精准率(precision)和召回率(recall),未出现因样本不平衡导致的少数类样本分类不准确的现象。

关于主题个数选择问题,本文根据困惑度值图2进行主题个数的选择。若将主题个数减少为14个,则根据各主题下的高频词分析发现,存在一些重要主题如夫妻感情与育有子女的情况这两个法律主题无法区分,使得到的裁判理由的结构不够细致。限于篇幅,省略了具体数值结果。将主题个数增加为26个,则对于裁判理由的结构分解过于细致,导致文本主题之间的含义难以区分,难以准确映射到法律主题。

关于不同主题个数下司法判决预测模型的性能问题,相比较于司法判决预测模型的结果表4,主题个数为14或26时,具有显著的解释力的自变量均有减少,且AUC值也显著降低。

综上,这些实验结果表明本文所构建的司法判决预测模型是合适的,该模型对于文本主题个数的选择、法律主题与文本潜在结构之间映射关系的构建、判决影响因素的解释及判决结果预测是可靠的。

(七) 模型在不同领域案件中的通用性说明

对于本文提出的融合法律文本潜在结构的司法判决模型,其通用性说明如图3所示。该模型利用了关键事实信息、地域信息、裁判理由的法律主题信息这三类特征,其中关键事实信息依据《法律术语词典》进行自动化分词提取得到;地域信息则依据法院所属地区的关键词结合BERT模型进行近义词合并提取得到;裁判理由中的一级法律主题,以及除了“3)影响判决的具体因素(领域)”以外的二级法律主题,依据文本主题模型结果结合法律术语关键词提取得到。以上特征通过自动化提取纳入司法判决预测模型,能够适用于不同领域类型的案件,通过自动化交叉验证设置文本主题个数,可以保证优良的模型预测效果。

对于相关法条及司法解释的规定明确细致的领域案件,比如交通事故等领域,该司法判决预测模型无须人工介入,可以得到优良的预测效果及可解释性。对于秉持“宜粗不宜细”立法精神的部分领域案件,该司法判决预测模型仍具有优良的预测效果。而因为不同法官的裁判理由的个性化考量因素差异较大,因此需要对影响判决的具体因素之下的二级法律主题进行关键词核对及主题个数的微调,保证二级法律主题与文本潜在结构的准确映射关系,从而能够进一步提高可解释性。比如本文所探究的彩礼返还案件,以及商标权侵权、精神损害赔偿等案件,其影响判决的具体因素往往比法律基础和司法解释更为

细致,如图 3 所示。综上,通过结合自动化工具与领域知识,该司法判决预测模型在多种类型案件具有适用性,能够得到优良可靠的预测效果及可解释性。

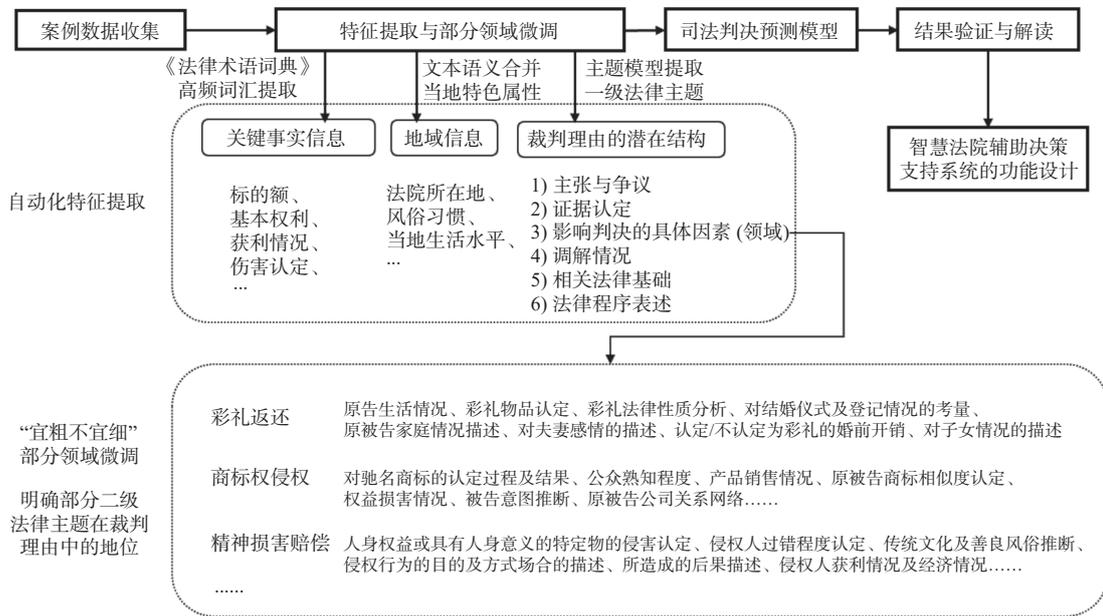


图 3 司法判决预测模型在不同领域案件的通用性说明

四、类案检索与推荐辅助决策支持系统

结合《最高人民法院关于统一法律适用加强类案检索的指导意见(试行)》《最高人民法院关于完善统一法律适用标准工作机制的意见》以及前文的研究,本文构建设计智慧法院辅助决策支持系统的若干功能,其整体设计如图 4 所示。该系统主要面向三类群体和场景,分别是:面向法院的类似案件分类管理与公平性检验功能、面向基层法官的类案推荐与辅助支持判决功能、面向群众的类案释法以提升法治素养功能。

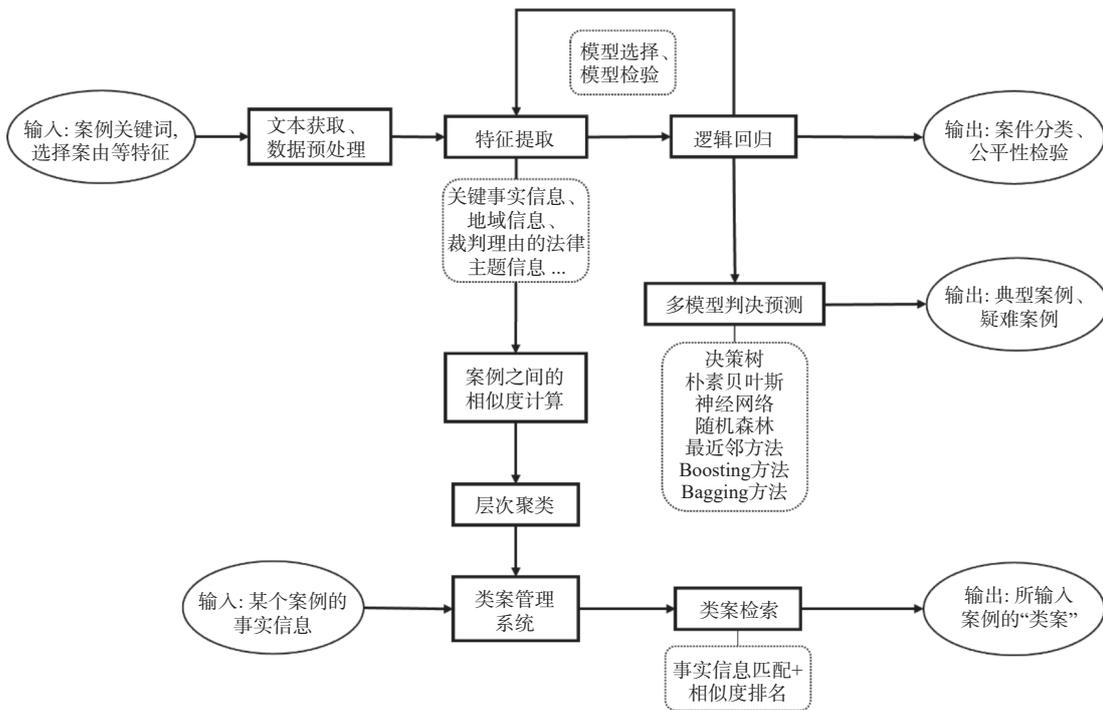


图 4 司法判决的辅助决策支持系统设计

面向法院的类似案件分类管理与公平性检验功能。首先根据选择的案件关键词,选择案由、审判程序等关键词进行类案检索,以保证该案件集合内部的案件之间具有可比性,由此得到输入的文本数据,进行类案分类管理。然后,对于输入的非结构化文本,经过预处理提取三类特征,根据特征可以对案件进行标记,便于按特定目的进行案件分类。构建逻辑回归模型后,通过比较参数估计的数值差异和显著性,得到不同地区(省、市、县)的案件判决的影响因素,和地区之间的差异对于案件判决的影响。由此结果结合经验,可以初步判断是否可认为满足司法公正性。

面向基层法官的类案推荐与辅助支持判决功能。将(1)中输出的逻辑回归模型的判决预测结果,与其他多种不同模型如决策树、朴素贝叶斯模型的判决预测结果做比较。对于某一案件,若多数个分类模型预测错误,则作为疑难案例;若多数个分类模型预测正确,则作为典型案例,这两类案件都可作为参考性案例,实现例以辅律^[55]。另外,对于(1)中预处理后的三类特征信息,通过层次聚类,使得该案件集合得以根据相似性归档。因此,对于输入的某个新案件的事实信息,可以根据关键的事实信息的匹配,和相似度排名,快速索引到对应的类案。最后,基层法官可依据此系统提供的:以往的典型案例、疑难案例、和新案件的类案,而作出判决。若新案件的判决结果与以往类案的判决结果差异较大,则提醒法官补充解释裁判原因,规范行使自由裁量权,由此为落实“同案同判”提供支持。

面向群众的类案释法以提升法治素养功能。群众可在此系统中输入所关注的关键事实信息,比如彩礼返还、未登记结婚、已举办结婚仪式等,选择相应的省份等地区,系统将检索并输出类案的裁判过程、裁判理由和裁判结果,群众可据此了解相似案件的审理判决情况。通过这样的类案式普法,可以进一步加深群众对于相关法律和公平正义的理解,有助于实现案结事了。

五、总结与展望

本文以彩礼返还纠纷案件的判决书为例,使用命名实体识别、文本主题模型的方法提取了关键事实信息、地域信息和裁判理由的文本潜在结构。随后,本文构建了裁判理由的法律主题与文本潜在结构的映射关系,提出了可解释的司法判决预测模型,探究了司法判决的影响因素,给出了该模型的通用性说明,为理解司法判决和检验同案同判提供了新的方法视角。最后,本文据此进行智慧法院辅助决策支持系统的功能设计,提出了面向法院的案件管理与公平性检验功能、面向基层法官的辅助支持判决功能、和面向群众的类案式普法宣传功能。为智慧法院建设,普及法律知识,法院审判资源优化配置而贡献力量。

在未来的研究中可以对本文的研究内容进行拓展。首先,本文所依据的裁判文书的样本量依然较小,使得本文从省份的视角而未能够从基层法院的视角分析差异性和判决的公平性。未来可以针对更为大规模的案件集合探究其判决的影响因素,和进行地区差异与司法公正性的检验。其次,本文在提取客观信息、地域特征、文本主题信息的过程属于无监督的学习,而结合有价值的人工标注的信息进行分析,是下一步研究的一个重要方向。最后,本文研究指出对于秉持“宜粗不宜细”立法精神的部分领域案件,可以基于大规模的历史案例语料库进行自监督预训练,并结合多层次结构的文本主题模型,得到细粒度的法律主题与文本潜在结构的准确映射关系。由此可以进一步地减少人工介入,从而增强模型在更多类型案件的通用性和自动化程度,这也是未来研究将探索的重要方向。

附录:实验数据的描述性分析、彩礼返还纠纷的法律基础

(一)司法判决文本内容的描述性分析

本文共收集到六个省份共计6096份判决书,各个省份的样本量如表5所示,其中932份判决书的判决结果为不返还,5161份的判决结果为返还。在裁判理由中,不同省份对当地风俗的考虑程度差异较大。在彩礼返还案件中,江苏省的判决对当地风俗的考虑比例最高,达该省案件总数的41.6%,而甘肃省

对当地风俗的考虑比例最低,为 19.0%。这体现出江苏省基层法院的法官在作出判决时,对于当地特殊彩礼习俗情况的考量较多,而甘肃省则更多考虑除当地风俗以外的信息,如证据认定、家庭生活情况等。根据最高法正式颁布的司法解释,判决书中将对关键的客观事实信息进行阐述,如:未办理结婚登记手续的应予以支持彩礼返还。

表 5 六个省份的彩礼返还案件的判决书数目和考虑当地风俗的比例

省份	河南	安徽	甘肃	江西	江苏	山东
样本量	1728	987	941	889	806	745
考虑当地风俗的比例	29.8%	31.8%	19.0%	30.5%	41.6%	30.3%

表 6 对于地域信息和关键事实信息的特征的描述性分析,可以看出可能影响彩礼返还的关键因素的分布情况。此外,这些解释特征之间的相关程度较低,各变量的方差膨胀因子均小于 5,可认为变量间不存在多重共线性问题。

表 6 地域信息和关键事实信息的描述性统计

考虑当地风俗	比例(%)	同居时间	比例(%)	曾经怀孕	比例(%)
是	30.2%	较短	41.4%	是	8.7%
未提及	69.8%	长或未提及	58.6%	否或未提及	91.3%
育有儿女	比例(%)	举办过婚礼	比例(%)	办理结婚登记	比例(%)
是	11.6%	是	3.3%	是或未提及	43.0%
否或未提及	88.4%	否或未提及	96.7%	否	57.0%
案件标的额	未提及	0—10万	10—50万	50—100万	100—500万
比例(%)	30.4%	39.0%	30.1%	0.2%	0.0%

注:标的额在 100—500 万的案件仅一个。

接下来,本文利用词频分析对分词预处理后的裁判理由文本数据,进行高频词汇的统计,绘制词云图如图 5 所示,图中字体的大小代表该关键词出现的频率的高低。从左侧支持返还的裁判理由高频词来看,主要聚焦于在法条引用和解释,结婚登记手续等方面。从右侧来看,裁判理由则主要聚焦于在证据认定,共同生活情况等方面。这反映出,对于支持返还或不返还的判决书,其裁判理由所表达的侧重点会有所不同,裁判理由会阐明其重点考量的因素从而给出判决结果。这表明裁判理由所蕴含的文本法律主题对判决结果有明显的解释力。



图 5 支持返还(左侧)和不予支持返还(右侧)的裁判理由词云图

(二) 司法判决的地区差异性

由于我国地区之间的经济、文化等方面的差异性客观存在,且彩礼返还案件具有强的地方性和风俗性,法官在判决案件时往往涉及更多个体因素判断,往往拥有更大的自由裁量权。因此,本文从省级层

面的数据探究两个问题:①省份之间在考虑当地风俗的比率上是否存在显著差异?②法官对于当地风俗的考量是否会对彩礼返还产生影响?

六个省份的判决书中裁判理由考虑当地风俗的比率如表5所示,对于上一自然段提到的问题(1),是两比率差的假设检验问题^[56],各省份的比率可看作某个两点分布 $b(1,r)$ 中的参数 r ,原假设 H_0 为:两个省份考虑当地风俗的比率相同。根据所获得的样本数据,可得每两个省份的假设检验的 p 值,如表7所示,从而对原假设进行判断。根据表2的结果,选取显著性水平为0.05,若 p 值小于0.05则拒绝原假设,表示在给定的显著性水平下,两个省份的比率的差异是显著的。反之,若 p 值大于等于0.05,则表示在给定的显著性水平下,不能认为两个省份的比率的差异是显著的。由此,根据样本数据可以得到,省份之间在考虑当地风俗的比率上存在显著差异,比如甘肃省判决书中的考虑当地风俗的比率显著低于其他五个省份。

表7 两两省份的比率差的假设检验的 p 值

省份\p值	甘肃省	河南省	江西省	安徽省	山东省	江苏省
甘肃省	—	<0.001	<0.001	<0.001	<0.001	<0.001
河南省	—	—	0.868	0.319	0.003	<0.001
江西省	—	—	—	0.483	0.014	<0.001
安徽省	—	—	—	—	0.062	0.001
山东省	—	—	—	—	—	0.190
江苏省	—	—	—	—	—	—

接下来探究问题(2),也即对于各省来说“是否考虑当地风俗”与彩礼是否返还的关系是否独立。根据所获得的样本数据,能够对每个省份的判决书的数目的 2×2 列联表如表8所示。通过卡方独立性检验,可以检验这两个变量是否独立,原假设 H_0 为:某省份“是否考虑风俗”与彩礼是否返还是独立的。在以省为单位的样本数据的计算中,得到六个省份对应的 p 值均小于0.001。根据样本数据的假设检验得到的结论为:对于各省份,地域特征“是否考虑当地风俗”会对“彩礼是否返还”产生影响,这表明法官做出裁判时对当地风俗的考量是一项重要影响因素。

表8 各省份的“是否考虑当地风俗”与彩礼是否返还的 2×2 列联表

甘肃省	返还	未返还	河南省	返还	未返还	江西省	返还	未返还
考虑	76	11	考虑	243	25	考虑	132	8
不考虑	711	146	不考虑	1106	250	不考虑	593	102
安徽省	返还	未返还	山东省	返还	未返还	江苏省	返还	未返还
考虑	160	15	考虑	133	16	考虑	186	12
不考虑	669	127	不考虑	441	96	不考虑	488	119

(三) 彩礼的概念

彩礼是中国传统中重要的婚俗习惯之一,是聘娶婚制度的产物^[57],其最早起源于西周时期“六礼”制度。《礼记·昏义》记载“六礼”为:“昏礼,纳采,问名,纳吉,纳征,请期,亲迎”^[58]。“纳征”又称“纳币”,意为男方向女方交送聘礼。经过不断的历史演变,成为如今民众生活中的“彩礼”。

社会学对于彩礼的解释理论可分为两种:婚姻补偿理论与婚姻资助理论^[59]。婚姻补偿理论认为:婚姻中的新娘是有价值的,当新娘从女方家嫁到男方家,女方劳动力与生育能力发生转移。由此,彩礼是对女方价值转移而支付给女方家庭的补偿^[60]。婚姻资助理论认为:彩礼是财产代际传承的方式,虽然彩礼先由男方家庭流转至女方家庭,但部分将以嫁妆的方式流入小家庭^[61]。随着男女平等观念的普及,女性

社会地位不断提升,彩礼逐渐摆脱交易性质,但作为婚俗习惯仍然广泛地存在于民众的婚嫁行为中。

(四) 法学角度看彩礼返还纠纷的种类

所谓彩礼返还,指男方将彩礼给予女方后,因各种纠纷的发生,男方主张女方返还已经给付的彩礼。纠纷发生的原因不同往往导致法院判决的结果不尽相同。实践中按照请求返还时男女双方是否结婚,可以分为两类:(1)未结婚的彩礼返还;(2)离婚彩礼返还。前者是指男方已经给予彩礼,但最终男女双方未办理结婚登记,男方请求法院判决女方返还彩礼。后者是指男女双方结婚后,因感情破裂办理离婚后,男方请求女方返还婚前给付的彩礼。

就目前的学术理论来看,一般认为彩礼的性质是:有解除条件的赠与^[62-64]。赠与的条件为登记结婚或者婚姻存续。当男方给付彩礼,随着赠与物交付,所有权发生转移。但由于该赠与附解除条件,当解除条件发生时,原本赠与的基础灭失,接受赠与的一方负返还义务。因此,彩礼返还请求权基础在于不当得利,且属于“给付目的嗣后不存在”的不当得利^[53]。一般实践中大部分纠纷属于彩礼交付后,当事人因各种原因未能办理结婚登记,男方赠与彩礼的目的不能达成,因此请求法院判决返还彩礼。但也有部分纠纷属于男女双方已经办理结婚登记,但离婚后诉请女方返还彩礼。

(五) 彩礼返还的规范基础

彩礼原本属于民众生活中的婚俗习惯,但近年来不断呈现高额化的特点,由彩礼导致的纠纷冲突也层出不穷。2019年中央1号文件指出:要持续推进农村移风易俗工作,治理天价彩礼等社会不良风气^[65]。2021年《国务院关于印发中国妇女发展纲要和中国儿童发展纲要的通知》再次重申抵制高价彩礼现象^[66]。

彩礼在实践中所引发的纠纷亟待法律进行回应,而我国对于彩礼问题的规制也经历了一个漫长的过程。革命时期,彩礼作为封建思想的余毒被禁止,1931年《中华苏维埃共和国婚姻条例》和1934年《中华苏维埃共和国婚姻法》都明令禁止^[67]。新中国成立后,1950年《婚姻法》以及相关的司法解释并未规定彩礼问题,但因彩礼产生的纠纷并未停止。为此,最高人民法院在1951年的一项《复批》大致表达了裁判的主要观点:区分赠与性质的聘礼与买卖性质的聘礼。若属于买卖婚姻性质的聘礼视情况予以没收,若属于赠与性质的聘礼不得没收,原则上也不得请求返还^[68]。此后无论是1980年《婚姻法》与2001年《婚姻法》修正案也并未对此问题进行正式的立法回应。直到2003年《关于适用〈中华人民共和国婚姻法〉若干问题的解释(二)》(以下简称《婚姻法司法解释(二)》)向社会公开征求意见时,“群众意见最多、最集中、分歧最大的,就是对彩礼问题到底应当如何规定”^[69]。

最终,2003年最高法正式颁布《婚姻法司法解释(二)》,其中第10条对此问题进行了正面回应。该条规定了三种人民法院判决支持返还的事由:未办理结婚登记;办理结婚登记但未共同生活;婚前给付导致给付人生活困难。后两种事由以离婚为前提条件。2017年修改《婚姻法司法解释(二)》保留了这一规定。2020年《民法典》正式通过,随之颁布的《最高人民法院关于适用〈中华人民共和国民法典〉婚姻家庭编的解释(一)》第5条仍然保留了原本《婚姻法司法解释(二)》的规定,成为实践中法院处理此类案件的主要法律依据。

(六) 案件中法官的其他细节考量因素

司法判决的第一要义是依法裁判。因此,正式的规范性文件是法官作出判决的核心依据。但上述司法解释只规定了哪些情况下法官可以支持返还,并未规定哪些属于彩礼的范围,应当返还多少以及如何返还等一系列问题。从实践来看,法院更多地地在具体的裁判中结合了当地的各类习惯风俗进行裁判,具体来看,可以分为以下种类。

第一,考量哪些财物属于法律规定的“彩礼”。实践中因风俗产生的各类婚嫁支出种类繁多复杂,除较为常见的“三金”“见面礼”“改口费”等常见的婚嫁支出,各地还存在各式各样的其他支出。例如安徽省淮北市的“红衣款”、甘肃省天水市的“追节礼”、江苏省宿迁市的“换鸡腿”。如何认定这些种类繁多

的婚嫁支出是否属于“彩礼”的范围,还需要法官在案件中根据当地的习惯具体判断。

第二,考量返还的范围和比例。对此,法院在实践中往往考量:当事人之间是否同居;同居时间的长短;女方是否有妊娠或流产;双方是否举办仪式婚礼;双方感情破裂的过错方;彩礼金额的数量;双方是否共同花销等等一系列因素。

第三,考量返还的责任主体。一般认为彩礼是男方家庭支付给女方的家庭的金钱或者财物,但是具体到实际案件,女方家庭谁作为返还责任主体法院的认定往往也不相同。例如安徽临泉县法院认为彩礼的给付离不开女方父母,因此女方父母应当承担返还责任。甘肃省榆中县法院认为爷爷奶奶代孙女收下彩礼也符合习俗,因此爷爷奶奶也承担返还责任。但江苏省邳州市法院认为只有实际收取彩礼的人才负返还义务,父母未接受彩礼因此不负返还责任。

注释:

- ① 参见安徽省阜阳市临泉县人民法院(2020)皖1221民初10789号民事判决书。
- ② 参见甘肃省兰州市榆中县(2021)甘0123民初1947号民事判决书。
- ③ 参见江苏省邳州市人民法院(2019)苏0382民初3763号民事判决书。

参考文献:

- [1] 新华社. 高举中国特色社会主义伟大旗帜为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告[EB/OL]. [2023-06-01]. https://www.gov.cn/xinwen/2022-10/25/content_5721685.htm.
- [2] 何莉. 努力让人民群众在每一个司法案件中感受到公平正义[EB/OL]. [2023-06-01]. <https://www.chinacourt.org/article/detail/2021/09/id/6254330.shtml>.
- [3] 新华网. 中共中央印发《法治中国建设规划(2020—2025年)》[EB/OL]. [2023-06-01]. <https://www.chinacourt.org/article/detail/2021/01/id/5710121.shtml>.
- [4] 李林,田禾. 中国法院信息化发展报告[M]. 北京:社会科学文献出版社,2017.
- [5] 左卫民,王婢媛. 基于裁判文书网的大数据法律研究:反思与前瞻[J]. 华东政法大学学报,2020,23(2):64-76.
- [6] 左卫民. 中国计算法学的未来:审思与前瞻[J]. 清华法学,2022,16(3):196-208.
- [7] 邓胜利,钱倩文,夏苏迪,等. 大数据的融合应用与拓展——首届“大数据管理与应用”暑期学校纪要[J]. 图书情报知识,2023,40(2):57-64.
- [8] 娄必县. 智慧法院建设让公平正义更高效更可观[EB/OL]. [2023-06-01]. <https://www.chinacourt.org/article/detail/2022/11/id/7013475.shtml>.
- [9] 安震威,来雨轩,冯岩松. 面向法律文书的自然语言理解[J]. 中文信息学报,2022,36(8):1-11.
- [10] 王婉臻,饶元,吴连伟,等. 基于人工智能的司法判决预测研究与进展[J]. 中文信息学报,2021,35(9):1-14.
- [11] 刘宗林,张梅山,甄冉冉,等. 融入罪名关键词的法律判决预测多任务学习模型[J]. 清华大学学报(自然科学版),2019,59(7):497-504.
- [12] JIANG X, YE H, LUO Z, et al. Interpretable rationale augmented charge prediction system[C]//Proceedings of the 27th international conference on computational linguistics: System demonstrations. 2018:146-151.
- [13] LI J J, ZHANG G Y, YU L X, et al. Research and design on cognitive computing framework for predicting judicial decisions[J]. *Journal of signal processing systems*, 2019, 91(10):1159-1167.
- [14] 左卫民. 关于法律人工智能在中国运用前景的若干思考[J]. 清华法学,2018,12(2):108-124.
- [15] 刘品新. 智慧司法的中国创新[J]. 国家检察官学院学报,2021,29(3):81-101.
- [16] 王燕玲. 论命名实体识别技术在司法大数据中的适用[J]. 政法论坛,2022,40(5):40-52.
- [17] 王颖洁,张程焯,白风波,等. 中文命名实体识别研究综述[J]. 计算机科学与探索,2023,17(2):324-341.
- [18] 余贵清,张永安. 审判案例自动抽取与标注模型研究[J]. 现代图书情报技术,2013(6):23-29.
- [19] 刘晨玥,李兵,吴卫星. 基于罪名相关成分标注的刑事裁判文书概要信息提取[J]. 山东科技大学学报(自然科学版),2018,37(4):92-101+124.
- [20] 黄茵,王宏宇,王晓光. 结合主动学习的条件随机场模型用于法律术语的自动识别[J]. 数据分析与知识发现,2019,3(6):66-74.
- [21] 王昊,林克柔,孟镇,等. 文本表示及其特征生成对法律判决书中多类型实体识别的影响分析[J]. 数据分析与知识发现,2021,5(7):10-25.
- [22] 王义真,沈雪莹,欧石燕. 民事裁判文书中论辩元素识别研究[J]. 数据分析与知识发现,2024(8):168-178.
- [23] 舒洪水. 司法大数据文本挖掘与量刑预测模型的研究[J]. 法学,2020(7):113-129.
- [24] 谭红叶,张博文,张虎,等. 面向法律文书的量刑预测方法研究[J]. 中文信息学报,2020,34(3):107-114.
- [25] 白建军. 基于法官集体经验的量刑预测研究[J]. 法学研究,2016,38(6):140-154.
- [26] 文姬. 醉酒型危险驾驶罪量刑影响因素实证研究[J]. 法学研究,2016,38(1):165-186.
- [27] 王芳,张蓝天,郭雷. 非线性递推理论在量刑数据分析中的应用[J]. 中国科学:信息科学,2022,52(10):1837-1852.

- [28] 中华人民共和国民法典[EB/OL]. [2023-06-01]. <http://www.npc.gov.cn/npc/c30834/202006/75ba6483b8344591abd07917e1d25cc8.shtml>.
- [29] 最高人民法院关于审理建设工程施工合同纠纷案件适用法律问题的解释(一)[EB/OL]. [2023-06-01]. <https://www.court.gov.cn/zixun-xiangqing-282111.html>.
- [30] 左卫民. 通过诉前调解控制“诉讼爆炸”——区域经验的实证研究[J]. 清华法学, 2020, 14(4): 89-106.
- [31] 朱青, 卫柯臻, 丁兰琳, 等. 基于文本挖掘和自动分类的法院裁判决策支持系统设计[J]. 中国管理科学, 2018, 26(1): 170-178.
- [32] 石勇, 安文录, 曲艺. 基于文本挖掘的检察起诉决策支持与案卷分类管理系统[J]. 管理评论, 2022, 34(6): 143-152.
- [33] 洪文兴, 胡志强, 翁洋, 等. 面向司法案件的案情知识图谱自动构建[J]. 中文信息学报, 2020, 34(1): 34-44.
- [34] 杨超群, 庞彦燕, 严若冰, 等. 基于迁移学习的适用法条推荐模型[J]. 四川大学学报(自然科学版), 2021, 58(2): 13-18.
- [35] 梁柱, 沈思, 叶文豪, 等. 基于结构内容特征的裁判文书自动推荐研究[J]. 情报学报, 2022, 41(2): 167-175.
- [36] 黄国栋. 比较法视野下智慧法院建设的中国经验、实践困境与路径优化[J]. 法律适用, 2023(3): 129-138.
- [37] 钞小静, 任保平. 中国经济增长质量的时序变化与地区差异分析[J]. 经济研究, 2011, 46(4): 26-40.
- [38] 郝颖, 辛清泉, 刘星. 地区差异、企业投资与经济增长质量[J]. 经济研究, 2014, 49(3): 101-114+189.
- [39] 李建军, 彭俞超, 马思超. 普惠金融与中国经济发展: 多维度内涵与实证分析[J]. 经济研究, 2020, 55(4): 37-52.
- [40] 唐跃, 王东浩, 陈暮紫, 等. 中国不良贷款回收率地区差异的原因分析[J]. 系统工程理论与实践, 2011, 31(3): 438-449.
- [41] 杨扬, 谢佳松, 林建浩, 等. 信任网络、资金流动与区域发展差距[J]. 系统工程理论与实践, 2020, 40(9): 2222-2235.
- [42] 赵子乐, 林建浩. 经济发展差距的文化假说: 从基因到语言[J]. 管理世界, 2017, 33(1): 65-77.
- [43] 高超, 黄玫立, 李坤望. 方言、移民史与区域间贸易[J]. 管理世界, 2019, 35(2): 43-57.
- [44] 王芳, 郭雷. 人机融合社会中的系统调控[J]. 系统工程理论与实践, 2020, 40(8): 1935-1944.
- [45] 龙小宁, 王俊. 中国司法地方保护主义: 基于知识产权案例的研究[J]. 中国经济问题, 2014(3): 3-18.
- [46] 最高人民法院关于加强和规范裁判文书释法说理的指导意见[EB/OL]. [2023-06-01]. <https://www.court.gov.cn/zixun-xiangqing-101552.html>.
- [47] 贾君君, 胡虞巧, 吴华清. 碳市场中文新闻主题建模: 热点与时空演变分析[J]. 计量经济学报, 2023, 3(2): 487-512.
- [48] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of machine learning research, 2003(3): 993-1022.
- [49] GRIFFITHS T L, STEYVERS M. Finding scientific topics[J]. Proceedings of the national academy of sciences of the United States of America, 2004, 101(suppl_1): 5228-5235.
- [50] 余乐安, 赵晨珊, 宋正阳. 融合网络新闻主题分布与主题分类下新闻情感的原油价格预测研究[J]. 计量经济学报, 2023, 3(2): 443-463.
- [51] 白建军. 论法的确定性与公正的可检验性[J]. 中国法学, 2008(2): 51-61.
- [52] 张妮, 蒲亦非. 计算法学导论[M]. 成都: 四川大学出版社, 2015.
- [53] 傅广宇. “中国民法典”与不当得利: 回顾与前瞻[J]. 华东政法大学学报, 2019, 22(1): 116-132.
- [54] 胡云红, 宋天一. 彩礼返还纠纷法律适用研究——以全国法院158份问卷调查和相关裁判文书为对象[J]. 中国政法大学学报, 2022, 92(6): 5-27.
- [55] 胡云腾, 于同志. 案例指导制度若干重大疑难争议问题研究[J]. 法学研究, 2008, 13(6): 3-24.
- [56] 吕晓玲, 黄丹阳. 数据科学统计基础[M]. 北京: 中国人民大学出版社, 2021.
- [57] 许莉. 婚姻家庭继承法学[M]. 北京: 北京大学出版社, 2019.
- [58] 彭林. 仪礼[M]. 北京: 中华书局, 2012.
- [59] 李永萍. 北方农村高额彩礼的动力机制——基于“婚姻市场”的实践分析[J]. 青年研究, 2018(2): 24-34+94-95.
- [60] 莫里斯·弗里德曼. 中国东南的宗族组织[M]. 刘晓春, 译. 上海: 上海人民出版社, 2000.
- [61] 阎云翔. 礼物的流动: 一个中国村庄中的互惠原则与社会网络[M]. 上海: 上海人民出版社, 2000.
- [62] 吴娟. 解除同居关系后彩礼返还的法律分析[J]. 东南大学学报(哲学社会科学版), 2017, 19(S1): 110-113.
- [63] 黄松有. 最高人民法院婚姻法解释(二)的理解与适用[M]. 北京: 人民法院出版社, 2004.
- [64] 史尚宽. 亲属法论[M]. 北京: 中国政法大学出版社, 2000.
- [65] 新华网. 中共中央 国务院关于坚持农业农村优先发展做好“三农”工作的若干意见[EB/OL]. [2023-06-01]. <https://china.huanqiu.com/article/9CaKmKif2A>.
- [66] 新华社. 国务院印发《中国妇女发展纲要(2021—2030年)》和《中国儿童发展纲要(2021—2030年)》[EB/OL]. [2023-06-01]. https://www.gov.cn/xinwen/2021-09/27/content_5639545.htm.
- [67] 中国社会科学院法学研究所, 韩延龙, 常兆儒. 中国新民主主义革命时期根据地法制文献选编(第4卷)[M]. 北京: 中国社会科学出版社, 1984.
- [68] 最高人民法院、司法部关于婚姻案件中聘金或聘礼处理原则问题的函(1951)(1951年8月10日法编字第9577号)[EB/OL]. [2023-06-01]. <http://fgcx.bjcourt.gov.cn:4601/law?fn=chl521s774.txt&dbt=chl>.
- [69] 最高人民法院民事审判第一庭. 最高人民法院婚姻法司法解释二的理解与适用[M]. 北京: 人民法院出版社, 2015.