

文章编号: 2095-2163(2021)06-0178-03

中图分类号: TP301.6

文献标志码: A

# 司法案例建模方法综述

郭小丁, 张宏莉

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

**摘要:** 近年来,科技部和司法部全面贯彻习近平的创新驱动发展战略,推进“数字法治、智慧司法”科技创新工程实施,司法案例建模技术的研究是其中重要项目。司法案例建模是指司法案例数字化,将司法案例表示为计算机能处理的数据结构。司法案例建模是上层法律服务的基础,为其提供数据支撑。目前,司法案例建模方法主要围绕特征模型和矩阵分解模型展开,本文分别从上述两个角度介绍当前案例建模方法,对国内外的研究进行综述,为科研人员提供指导和参考。

**关键词:** 司法案例建模; 特征模型; 矩阵分解

## Summary of judicial case modeling methods

GUO Xiaoding, ZHANG Hongli

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

**[Abstract]** In recent years, the Ministry of Science and Technology and the Ministry of Justice comprehensively implement Xi Jinping's innovation-driven development strategy, promote the "digital rule of law, smart justice" project. Judicial case modeling technology is one of the important projects. Judicial case modeling refers to the digitization of judicial cases. It represents judicial cases as structures that can be processed by computers. Judicial case modeling is the foundation of legal services and provides data support for them. At present, modeling methods of judicial cases mainly focus on feature models and matrix decomposition models. This paper introduces current modeling methods from the above two perspectives and summarizes present research to provide guidance and reference for researchers.

**[Key words]** Judicial case modeling; Feature models; Matrix decomposition

### 0 引言

司法案例建模方法旨在将司法案例表示为案例相关预测算法能处理的数据结构,该数据结构保留原始数据中的有用信息,去除冗余信息,减少其对后续预测算法的干扰。司法案例建模方法是上层法律服务的基础,为其提供数据支撑。简洁有效的案例描述能有效地提高预测算法的准确率,降低其计算复杂度。

传统司法案例建模方法主要基于特征模型和矩阵分解。特征模型利用特征词典和特征抽取算法将司法案件表示为二维矩阵,如图1所示。然而,当特征和案例规模较大时,特征模型极易发生数据稀疏和维度爆炸。针对上述问题,基于矩阵分解的司法案例建模方法将特征模型得出的原始矩阵进行分解,并利用得出的矩阵表征案例。当矩阵分解算法与共现矩阵、主题模型相结合时,得出的司法案例表示具有一定的语义信息,如词汇间相似性、主题与特征间关联性。

基于矩阵分解的司法案例建模方法降低矩阵维度,避免维度爆炸及数据稀疏,有利于提高后续预测算法的准确率。但这类方法基于特征模型,存在特征模型天然缺陷,无法从多个层面描述案例,难以捕捉不同案例模块间的关联信息。上述因素不利于提高后续预测算法的准确率。

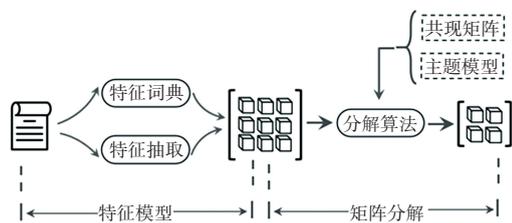


图1 基于特征模型以及矩阵分解的司法案例建模方法

Fig. 1 Judicial case modeling method based on feature models and matrix decomposition

### 1 基于特征模型的司法案例建模方法

针对司法案例建模问题,早期的研究主要围绕特征模型展开,大体分为基于案例推理、本体模型、分类算法的司法案例建模方法。根据大量的法律专

基金项目: 国家重点研发项目(2018YFC0830900)。

作者简介: 郭小丁(1991-),女,博士研究生,主要研究方向:大数据、人工智能、自然语言;张宏莉(1973-),女,博士,教授,博士生导师,主要研究方向:大数据、网络安全、人工智能等。

收稿日期: 2020-12-21

家知识和人工标注工作,上述方法利用特征抽取算法从原始案例数据中获取案例特征,结合案例推理、本体模型与分类算法,依据现有特征预测出新的案例特征,得出基于特征标签的司法案例模型。

### 1.1 基于案例推理的司法案例建模方法

针对司法案例建模问题,最早期的研究是利用大量的人工标注为案例及案例要素分配索引,进而从结构合理的案例文本中提取案例事实,并将其用于案例摘要分类和案例推理系统中。随着数据挖掘技术的发展,研究者提出一种基于信息抽取的司法案例建模方法。结合特定领域的知识,该方法抽取案例中的事件经过,利用特征抽取技术捕捉个体行为中的动作,并对其进行识别。

Zeleznikow 一种新的用于信息检索的司法案例建模方法<sup>[1]</sup>。该方法扩展传统案例因素和组件,将每一个因素分解为更加细致的案例要素,同时将所有要素分为主观要素、客观要素和天然要素,针对每一个司法案例利用特征挖掘技术匹配案例特征,进而得出案例表征。实验结果表明,针对信息检索问题,该方法具有更高的准确率和满意度,主要因素在于层次特征的使用。

长期以来,基于法律案例的推理(Legal Case-Based Reasoning,LCBR)一直是 AI 和法律感兴趣的话题,各种各样的方法已经发展。LCBR 的一项重要工作始于 HYPO,HYPO 根据法律规范中的论点与反论点在每个维度对司法案例进行建模,提出一种基于案件要素的建模方法。Kevin 出一系列利用特征模型刻画司法案例的方法,如 CATO 和 IBP,并将其用于法律案例推理应用中,根据大量调研和实践,上述方法构建案例框架,从司法案例中抽取原告、被告、指控罪名、主观以及客观态度等案例要素,为每一个案例要素设置索引,进而完成案例表征<sup>[2]</sup>。

### 1.2 基于本体模型的司法案例建模方法

Breuker 提出一种基于法律知识及用例分析的司法案例建模方法<sup>[3]</sup>。该方法根据专业人士从法律法规中获取的案例要素信息,利用特征挖掘算法从真实案例中抽取案例要素,根据大量调研,构建结构化论点模型,生成案例本体,进而建立案例要素与案例本体之间的映射,得出司法案例表征。Ashley 提出司法案例中责任与因果关系的发现方法,在充分考虑当事人责任归属的前提下,该方法构建专用的本体模型,并用其表征司法案例<sup>[4]</sup>。该模型涵盖法律领域中常见的责任与因果关系,如职位责任、雇佣关系、竞争关系等。根据已有责任划分和

因果关系,该模型利用推理规则,自动获取新的因果链以及责任归属。基于本体的司法案例责任与因果关系发现方法需要大量的概念知识库,该知识库中包含法律规范、法律概念和术语以及人们日常生活中常识性词汇表达等。

在此基础上,研究者提出一种基于 Web 本体语言的司法案例建模方法,以 CATO 和 LCBR 为基础,构建司法案例本体,确定本体中的关键案例要素,利用网络中案例将本体实例化,进而实现不同案例间的组件共享<sup>[5-7]</sup>。基于 OWL 的司法案例建模方法在案例信息检索和案例推理方面具有广泛的应用。

### 1.3 基于规则推理的司法案例建模方法

Lagos 以司法案例中事件为单位,提出一种基于事件抽取的司法案例建模方法<sup>[5]</sup>。该方法从原始案例数据中抽取出关键实体,如案件中当事人、当事人参与的事件以及当事人接触的其他人物等信息,利用规则推理技术抽取出实体之间的关系,进而得出实体与关系构成的事件,完成司法案例表征;Henry 提出一种基于规则推理和启发式要素搜索的司法案例建模方法,将规则推理嵌入案例建模过程中,半自动化抽取案例特征,在一定程度上减少了建模过程中的人工工作<sup>[6]</sup>。针对不同类型的司法案例,该方法依据大量的历史案例数据库,将最佳历史案例作为指导性案例,根据指导性案例中的法律推理规则及决策结果,提出不同的案例要素推理规则与裁决方法,并将其用于后续司法案例的建模过程中。

### 1.4 方法简析

基于特征模型的司法案例建模方法将司法案例表示为二维矩阵,此类方法主要存在以下不足:

(1)需要大量的法律专业知识和人工标注工作。特征模型构建特征词典,并利用特征挖掘算法抽取案例特征,进而将其表示为二维矩阵;

(2)对数据库中词汇和语法信息的依赖性强。当新增案例或者案例的特征词典发生变化时,特征模型易发生较强波动,影响后续预测算法的准确率和稳定性;

(3)极易发生维度爆炸和数据稀疏。当司法案例数量规模较大时,特征模型中易出现大量冗余信息,进而导致特征词典命中率降低,发生数据稀疏和维度爆炸;

(4)无法从多个层面描述案例,捕捉不同案例模块间的关联信息。

## 2 基于矩阵分解的司法案例建模方法

当司法案例规模和案例特征数量较大时,基于

特征模型的司法案例建模方法极易发生数据稀疏和维度爆炸。为了解决上述问题,研究者提出一系列基于矩阵分解的司法案例建模方法。

## 2.1 方法综述

矩阵分解算法将特征模型得出的代表司法案例的原始矩阵分解为维度更低的矩阵,去除了原始矩阵中冗余信息,有效地提高后续预测算法的计算速度和准确率。在以往研究中,矩阵分解算法常与共现矩阵、潜在语义分析相结合(Latent Semantic Analysis, LSA),共同完成司法案例建模。

当司法案例规模和特征数量较大时,特征模型极易出现数据稀疏和维度爆炸。为了解决上述问题,矩阵分解算法被广泛用于主成分抽取和矩阵降维中。矩阵的奇异值分解理论是矩阵理论的重要组成部分,是研究最小两乘积、广义逆矩阵等问题的有力工具。目前存在多种矩阵分解算法,如奇异值分解(Singular Value Decomposition, SVD)、非负矩阵分解(Non-negative Matrix Factorization, NMF)。矩阵分解算法旨在将代表司法案例的原始矩阵映射到新的向量空间中。该向量空间由一组新的基构成。原始矩阵集合在新向量空间中的投影更加离散、具有区分性。矩阵分解算法去除原始矩阵中大量无用信息,保留有效信息,有助于提高后续预测算法的计算速度和准确率。

Gui总结了矩阵分解在文本建模方面的应用,其提出的基于奇异值分解的矩阵分解算法可被用于司法案例建模<sup>[7]</sup>。该方法建立特征词词典,利用特征挖掘技术从原始数据中抽取特征词汇,将司法案例表示为二维原始矩阵。在此基础上,该方法利用基于奇异值分解的矩阵分解算法,将原始矩阵分解为维度更小的矩阵,并用其表征司法案例。实验结果表明,对于司法案例文本类别预测,相较于特征模型,基于奇异值分解的矩阵分解算法更有利于提高预测算法准确率和计算速度。主要原因在于矩阵分解算法去除了原始矩阵中大量无用信息。

Yin提出一种基于奇异值分解的司法案例建模方法,并将其用于法律咨询系统中<sup>[8]</sup>。该方法利用空间向量模型作为表示方法,借助潜在语义分析理论,通过奇异值分解的降维方法构建一个低维的语义空间,并在语义空间上实现了咨询问题与解答语句之间的相似度计算。针对司法案例中语句,该方法分别使用分词、去停用词、互信息扩展以及潜在语义分析,将咨询语句映射至主题语义空间中,并利用SVD分解得出其低维向量表示。

Symeonidis提出一种基于聚合矩阵分解的司法案例建模方法<sup>[9]</sup>。该方法在奇异值分解的基础上,根据文档的主题和可读性,利用余弦相似度或欧氏距离对司法案例进行建模。该方法使用特征模型将司法案例表示为二维矩阵,利用SVD算法降低原始矩阵的维度,结合当前案例和相似案例的向量表示,获得当前司法案例的数字化表征。实验结果表明,相对于矩阵分解算法,聚合矩阵分解更有利于提高后续预测算法的准确率。

## 2.2 方法简析

基于矩阵分解的司法案例建模方法在特征模型的基础上,将得出的代表司法案例的原始矩阵分解为维度更小的矩阵。矩阵分解算法极大地降低原始矩阵维度,减少后续预测算法的计算复杂度。矩阵分解算法去除了原始矩阵中冗余信息,提高后续预测算法准确率。然而,基于矩阵分解的司法案例建模方法仍然存在特征模型天然缺陷,不能从多个层面描述案例,捕捉不同案例模块之间的关联信息。

## 3 结束语

本文针对司法案例建模问题,给出近年来的相关研究成果,将司法案例建模方法划分为两类,即基于特征模型的方法和矩阵分解模型的方法,分析了二者的优缺点,为研究人员提供参考和指导。

## 参考文献

- [1] Zeleznikow, John. Knowledge Representation for the Intelligent Legal Case Retrieval[J]. Lecture Notes in Computer Science, 2005, 3681:339-345.
- [2] Kevin D. Ashley. Modelling Legal Argument: Reasoning with Cases and Hypotheticals[M]. Bradford Books/MIT Press, 1990.
- [3] Breuker J, Hoekstra R. Direct: Ontology based discovery of responsibility and causality in legal case descriptions[J]. Phycologia, 2004:59-68.
- [4] Ashley, K.: What a Legal CBR Ontology Should Provide. In: Proceedings of the 22<sup>nd</sup> Int'l FLAIRS Conf. Case-Based Reasoning Track, Sanibel Island, FL (May 2009)
- [5] Lagos N, Segond F, Castellani S, et al. Event Extraction for Legal Case Building and Reasoning[J]. Ifip Advances in Information & Communication Technology, 2010, 340:92-101.
- [6] Henry P, Adam W, Trevor B C, et al. A formalization of argumentation schemes for legal case-based reasoning in ASPIC+ [J]. Journal of Logic & Computation (5): ext010.
- [7] Gui L. Application of Matrix Singular Value Decomposition (SVD) to the Research of Text Categorization on Style[J]. Computer Engineering, 2002, 28(12):17-10.
- [8] Yin Z F, Wang X, Cai W Z, et al. Law Consultation System Based on Lucene and LSA [J]. Computer Systems & Applications, 2014.
- [9] P. Symeonidis, I. Kehayov, Yannis Manolopoulos. Text Classification by Aggregation of SVD [M]// Advances in Databases and Information Systems.