

基于规则和案例推理集成的 刑事案件量刑决策支持研究

余贵清 著

- ★ **构建**基于本体的刑事案例库。
- ★ **探索**基于遗传算法的刑事案例属性优化与检索策略。
- ★ **研究**刑事案件推理集成决策模型。
- ★ **设计**刑事案件量刑决策辅助系统原型。

 中国工信出版集团

 电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

基于规则和案例推理集成的刑事 案件量刑决策支持研究

余贵清 著

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

本书构建了基于本体的审判案例自动抽取标注模型；针对审判案例特点，提出了支持多属性、多案例、多要素的 GA-KNN 案例检索方法；利用决策树、神经网络等数据挖掘算法构建了规则推理和案例推理集成的决策支持机制。对以半结构化或非结构化数据作为决策基础数据，以规则推理和案例推理同时支持的决策工作具有重要借鉴意义。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目（CIP）数据

基于规则和案例推理集成的刑事案件量刑决策支持研究 / 余贵清著. —北京：电子工业出版社，2016.3

ISBN 978-7-121-28142-6

I. ①基… II. ①余… III. ①刑事犯罪—量刑—研究—中国 IV. ①D924.134

中国版本图书馆 CIP 数据核字（2016）第 026454 号

责任编辑：秦绪军

特约编辑：田学清 赵海红

印 刷：北京季蜂印刷有限公司

装 订：北京季蜂印刷有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱

邮编：100036

开 本：720×1000 1/16

印张：16

字数：258 千字

彩插：1

版 次：2016 年 3 月第 1 版

印 次：2016 年 3 月第 1 次印刷

定 价：45.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zltts@phei.com.cn，盗版侵权举报请发邮件到 dbqq@phei.com.cn。

服务热线：（010）88258888。

前 言

随着国内经济发展、社会进步,国内司法领域审判量刑决策的科学性、规范性成为社会关注的焦点,法规与案例结合支持法官量刑决策成为司法审判领域和管理理论界的热点问题,迫切需要进行针对性、前瞻性、应用性强的研究。为了给法官决策提供有效的支持,本书创建了基于规则和案例推理集成的辅助量刑决策机制,并运用本体、遗传算法等理论与方法,结合案例进行了严谨的论证。本书研究主要构建了基于本体的刑事案例表示方法,提出和论证了对于多属性、多案例、多要素背景下,基于 GA-KNN 的案例属性优化与检索针对性方案,揭示了基于规则与案例推理集成的量刑决策机制。为了系统论证集成案例推理的可靠性,以 11 000 个故意伤害罪的半结构化案例文档集为初始样本,通过本体技术进行决策信息规则化、结构化,采用 GA-KNN 算法进行属性优化与案例检索,采用决策树和神经网络模型构建推理集成决策模型,通过与实际量刑决策结果比较论证了该方法对支持量刑决策有重大作用。

本书主要研究内容和结论归纳如下:

本书针对目前法官量刑决策判决书文档集的非结构化、半结构化特征,首先,通过分析刑事案件判决书文档集结构特征,从本体理论规范知识表达出发,建立了刑事案件本体,并以罪名为单位,对刑事案件相关的概念、类和属性进行了规范化表达;其次,以案例文档集和刑事案件本体为基础,通过信息抽取

和标注技术,建立了半结构化文本决策数据的自动抽取模板,实现了刑事案件实体信息的自动化描述性抽取;再次,通过数据仓库技术,对刑事案件决策信息进行逻辑设计,将描述性实体信息规则化,并形成规则化案例库;最后,对刑事案件本体设计和信息抽取方法进行了验证。研究发现,以本体为核心,从模型、实验和评价三个角度阐述的从半结构化数据向结构化案例库转化的机制,是利用信息技术手段实现更多决策信息可利用的重要途径,是科学决策的重要基础。该研究结论确立了蕴含在半结构化文档中决策信息支撑决策的途径,突破了从半结构化数据中获取决策信息的瓶颈,实现了本体理论在领域决策中的应用。

本书针对司法领域的多案例、多属性、多要素特征及推理机制依赖的案例库必须满足快速检索、及时支持法官决策的要求,确立了刑事案例属性优化与敏捷检索这一关键技术问题。首先,通过刑事案例的特点分析确立了刑事案例推理精确性目标与案例检索环节方法选择与路径;其次,通过遗传算法、改进的 AHP 方法对案例属性进行优化效果比较,形成适合刑事案件量刑决策的属性优化方法,分析比较了不同方法检索结果异同;最后,论证了多案例、多属性、多要素司法类案例应用 GA-KNN 方法组合完整检索方法的可行路径与检索精度。研究发现,基于权重优化的 GA-KNN 检索方法可以提高多案例、多属性、多要素司法类案例的检索精确度及检索效率。该研究拓宽了管理科学与信息技术交叉研究的范围,实现了决策过程依赖特定决策基础信息的最优集合,使决策推理可以在有限的时间内完成。

本书针对我国成文法依法规量刑和结合案例辅助约束法官自由裁量权的需要,重点研究刑事案件量刑决策中规则推理和案例推理集成决策机制问题。首先,分析了以案例库为基础的数据挖掘技术,根据刑事案件信息离散性和决策过程的非线性特点,提出并论证了决策树模型和神经网络模型的综合应用模式,

使法规与案例在决策推理中进行结合；其次，利用实验方法，对两种推理集成的模型进行训练，直至满足决策要求，将推理模型按案件类别（罪名）进行确定，形成各类型案件对应的决策模型；最后，通过具体使用该推理集成决策模型，对其应用效果进行了验证。研究发现，利用数据挖掘方法对领域知识进行深度挖掘，可以找到决策中潜在的关键信息，如量刑决策特征中的暴力方式细节与决策结果的关系、量刑结果类型与各犯罪特征集的对应等。该研究结论中两种推理的集成充实了管理决策方法，同时，拓宽了领域知识的挖掘和利用机制。

本书针对通过信息决策系统自动支持法官量刑的需要，确定了如何将决策模型在实际量刑决策中得以应用的前瞻性研究问题。首先，从上述研究成果出发，从系统需求与功能模块划分的角度对系统进行了分析与阐述，提出了量刑决策辅助系统原型；其次，基于角色和目标案例，进行了系统架构设计，研究了三个关键模块的设计与实现细节；最后，对该系统原型进行了初步实现，并展示了相关界面，验证了该原型的正确性。研究发现，建立量刑决策辅助系统需要有强大的数据库、知识库做支撑，将案件的案情结构化、法官化是其中的重要环节。该研究结论是决策理论、信息技术与领域知识等多学科研究成果综合应用的结果，拓宽了多学科交叉应用研究的领域。

关键词：量刑决策；刑事案件本体；案例属性；案例推理；推理集成

目 录

第 1 章 绪 论	1
1.1 研究背景及问题提出	1
1.2 研究意义	3
1.2.1 理论意义	4
1.2.2 实践意义	6
1.3 本书中相关概念界定	7
1.3.1 半结构化文本	7
1.3.2 领域本体	7
1.3.3 信息抽取	8
1.3.4 语义标注	9
1.3.5 刑罚裁量	10
1.3.6 案例推理	10
1.3.7 决策集成	11
1.4 研究对象与研究内容	11
1.4.1 研究对象	11
1.4.2 研究内容	11
1.5 研究方法与技术路线	13
1.5.1 研究方法	13
1.5.2 技术路线	14

1.6 章节安排	16
1.7 研究的创新之处	18
第2章 国内外研究综述	20
2.1 法律知识的表示与抽取	20
2.1.1 法律知识的表示	21
2.1.2 法律知识的抽取	28
2.2 法律案例的属性优化与案例检索方法	38
2.2.1 案例属性优化	38
2.2.2 案例检索方法	41
2.3 法律知识的推理	42
2.3.1 法律知识的推理逻辑	42
2.3.2 法律知识的推理方法	44
2.4 文献述评	50
2.4.1 研究问题的价值	50
2.4.2 研究面临的问题	52
第3章 基于本体的刑事案例库构建研究	56
3.1 构建刑事案例库的条件分析	57
3.1.1 刑事案件文书特点	57
3.1.2 刑事案例现状分析	59
3.1.3 刑事案件本体构建难度	60
3.2 刑事案件本体化框架设计	62
3.2.1 本体建模基本模型	62
3.2.2 刑事案件本体构建	64
3.2.3 刑事案件信息抽取	74
3.2.4 刑事案例库设计	81
3.3 刑事案件本体构建实现与信息抽取评价	85
3.3.1 刑事案件本体构建实现	85

3.3.2 刑事案件信息抽取评价	88
3.4 本章小结	91
第 4 章 基于遗传算法的刑事案例属性优化与检索策略研究	92
4.1 基于遗传算法对刑事案例属性优化的研究	93
4.1.1 刑事案例属性的特征及属性优化的意义	93
4.1.2 两种优化方法对刑事案例属性优化的针对性分析	94
4.1.3 针对刑事案例属性优化的两种优化方法对比分析	97
4.2 基于 GA-KNN 算法对刑事案例检索策略研究	108
4.2.1 案例相似性的 KNN 算法	108
4.2.2 案例相似性的 GA-KNN 算法	112
4.2.3 案例相似性的 KNN 方法与 GA-KNN 算法的比较	113
4.3 本章小结	116
第 5 章 刑事案件推理集成决策模型研究	117
5.1 基于本体的法律推理体系	117
5.2 刑事案件推理决策模型中的相关算法	119
5.2.1 特征选择	119
5.2.2 决策树	123
5.2.3 神经网络	128
5.3 刑事案件推理集成决策模型研究	130
5.3.1 推理集成决策模型建立意义	130
5.3.2 推理集成决策模型设计	132
5.3.3 案例特征提取策略	137
5.3.4 模型训练阶段的方案设计	139
5.3.5 模型使用阶段的方案设计	140
5.4 刑事案件推理集成模型实现	141
5.4.1 实验平台	141
5.4.2 实验数据及预处理	141

5.4.3 判决类型推理模型实验	146
5.4.4 判决刑期推理模型实验	150
5.5 刑事案件推理集成模型评价	151
5.6 本章小结	152
第 6 章 刑事案件量刑决策辅助系统原型设计	154
6.1 需求分析	155
6.1.1 系统需求概述	155
6.1.2 系统功能需求	155
6.2 系统总体设计	156
6.2.1 系统工作流程	157
6.2.2 系统架构设计	159
6.3 功能模块实现	160
6.3.1 案例库模块实现	160
6.3.2 决策推理模块实现	161
6.3.3 案例推荐模块实现	164
6.4 系统原型界面设计	165
6.5 本章小结	168
结论与展望	170
附录 A 刑事案件本体类与属性的详细设计	173
附录 B 刑事案件本体 OWL 文件	180
附录 C 故意伤害罪类刑事案件本体	215
附录 D 刑事案件本体 XML 抽取模板	219
附录 E 实验数据	230
后 记	233
参考文献	235

绪 论

1.1 研究背景及问题提出

伴随着我国经济飞速发展和人民生活水平不断提高，人们对社会公平正义的呼声越来越高，依法治国成为我国政治发展的必然趋势。2010年，我国的社会主义法律体系初步建成，明确要求法官遵循法律原则和法律规定的情况下，准确掌握案件事实，慎重把握裁判尺度，妥善处理矛盾纠纷，实现社会公平正义。

量刑规范化和案例指导制度等成为法律理论与实务界关注的重点，迫切需要通过信息技术手段解决量刑不统一的问题，为法院科学决策、科学发展进一步提供理论技术支持。因为我国法律体系建立时间不长，存在着两个天生的不足之处：一是我国唯一的量刑主体——法官队伍中，存在着大量理论功底缺乏和审判经验不足等问题，表现为从事刑事审判的法官不能很好地驾驭审判，

对案情不能形成较全面的自由心证,对如何量刑不能做出正确判断。二是《刑法》规定宽泛不严谨,使法官自由裁量权过大,成为司法不统一的主要因素。世界量刑体制大体有立法模式、司法模式、行政模式等。我国目前采用的是司法模式,也就是法律上对犯罪行为规定了相对确定的幅度刑,法官对具体案件宣判的是确定刑(定期刑)。我国的刑事立法规定的每个罪,应受的处罚要么过于宽泛,要么过于笼统。例如,《刑法》第234条规定,犯故意伤害罪“致人死亡或者以特别残忍手段致人重伤造成严重残疾的,处十年以上有期徒刑、无期徒刑或者死刑”,到底什么情况判多少年、什么情况判死刑由法官作判。为解决上述司法实践中的缺陷,近几年,最高法院连续发布量刑规范化办法,并着手建立案例指导制度,两个重要课题都离不开信息技术的支持,需要将管理决策理论、司法审判领域知识和信息技术相结合做前瞻性研究。

人民法院信息化的不断发展,为建立基于案例指导的司法审判制度提供了海量的决策数据,为科学决策提供了数据基础。我国法院信息化发展大都经历了十几年的快速发展,不仅有支持法官办案的审判业务系统,同时,也初步探索建立了支持决策的系统,并积累了数以千万计的电子化判决书。这些判决书蕴含着案件的案情、判决依据、判决结果和审理理念,它具有数据量大和半结构化两个特点。由于判决书量大,具有大数据体量巨大、数据类型繁多、价值密度低和处理速度快的特征,不可能采用人工方法,将其转成支持量刑决策的信息;由于其半结构化的特点,经过恰当的方式进行结构分析和表达,可以将其转化为结构化的决策信息。

现代决策理论的发展,使决策理论吸收了行为科学、系统理论、信息技术及运筹学等新兴学科的理论和方法,产生了决策技术论分支。决策技术论认为在决策过程中要对各种各样的决策进行深入细致的分析,借助一定的科学方法和手段,发展相应的决策技术,可以减少决策主体有限理性对决策结果的影响。

支持大数据的数据库、云计算、可扩充存储系统和挖掘技术的不断发展,使得将有价值的决策信息从半结构化的案例文本中进行挖掘,供决策中使用成为可能。数据库和数据挖掘的产生和发展,使得当今的计算机网络应用体系从业务管理逐步跃升到决策支持。同时,在“大数据”时代,要从海量、纷繁复杂的数据中快速找到决策者关注的信息,必须借助功能强大、智能化的信息检索工具。将系统蕴含的基础数据转化为决策知识,也为大规模数据统计分析处理提供了一套完整的解决方案。

面对社会对法律需求的机遇和法官努力实现司法统一的内在动力,从最高法院到地方法院都着手研究,借助网络与信息技术,就基于规则和案例支持的新的量刑决策模式进行前瞻性研究,创新提出量刑决策新模式。从而,提高司法审判的科学性和量刑的均衡统一,使审判决策结果朝着“同案同判”的方向发展。为此,从决策科学的角度入手,探索信息条件下的决策方法论,支持在审判领域创建规则和案例推理集成的决策模式,以建立一套相对科学的、处理复杂决策问题行之有效的支持动态决策的方法,这是本书研究的出发点和落脚点。

1.2 研究意义

在司法审判领域,我国采用的是成文法,坚持“有法可依”的原则,刑事案件审判的基本原则是以事实为根据,以法律为准绳,进行量刑决策。根据案件性质的不同,我国《刑法》将刑事案件分为400多个罪名,每个罪名又根据案件事实的区别,给出一定的量刑幅度,具体刑期就由法官进行自由裁量。

法官自由裁量的依据有三个，一是依据“重现的”案件事实。刑事案件审理都是事后的，依据的案件事实是通过各种审理的必要程序进行“重现的”，其带有明显的与真实事实的差异。同时，实践表明，即使相同罪名的案件，其情节也千差万别，各不相同，如何找出各类案件的共同特征，理清案情的关键要素，用以表示相同或相似案件的特征是从整体研究审判决策方法的前提。二是依据相关法律。我国法律体系初步建立，加上社会环境日新月异，法律条文的宽泛难以适应新的社会矛盾和审判工作的需求。以故意伤害罪在《刑法》第234条中的表述为例，“故意伤害他人身体的处三年以下有期徒刑、拘役或者管制。犯前款罪，致人重伤的，处三年以上十年以下有期徒刑；致人死亡或者以特别残忍手段致人重伤造成严重残疾的，处十年以上有期徒刑、无期徒刑或者死刑。”从中可见，无论从法的规定，还是量刑区间，都存在很大的弹性，需要裁判主体通过智慧去合理裁判。三是依据法官的智慧。法官的智慧和经验是隐性知识，来源于对大量相同或相似案件的审理，以及对法律法规的反复解释，是一种实践和参悟的结果。不能否认有许多优秀和成熟的法官具有这样的能力，但绝大多数法官并不具备这样的能力和素质，需要借助成熟的决策方法的指引。

以大量的案件数据为依据，通过科学的数据抽取方法，挖掘案件量刑的规则，建立案例库，形成辅助审判决策的信息平台，无论是在决策理论研究，还是决策方法实践研究，都具有重要的意义。

1.2.1 理论意义

决策技术论，即对各种各样的决策进行深入细致的分析，提出有针对性的方法，发展相应的决策技术。决策理论学派所主张的决策，是在解决复杂决策问题时，要求尽可能量化，建立数学模式，从而求得满意结果。即使结果满意，

计算工作量也是很大的,离开了电子计算机有时是很难办到的。计算机科学的兴起是决策理论学派得以存在和发展的基础。本书正是从帮助决策者处理信息、解决关键技术、建立决策模型并提供可计算机系统化的决策方法的研究,属于决策理论学派范畴。

目前,不少学者在不同领域研究规则推理,也有不少应用成果。案例推理由于其案例表达的困难和长期以来信息挖掘的瓶颈,使其处在研究较多,但专业领域应用却不是很深入的境地。将规则和案件推理相结合,应用到司法审判领域就更少见。这就需要管理科学与工程领域的决策管理研究者,对决策活动中的这些信息从挖掘技术、智能集成和辅助决策等方面进行研究,为决策者提供一套处理复杂决策问题的理论方法和智能决策途径与手段,形成信息条件下的决策模式,用以指导决策活动^[1]。

(1) 决策信息的表达方式。以本体理论为基础,依据刑事案件自身特征,探索领域知识本体化表示方法,为使半结构化或非结构化的决策信息规则化和利用化成为可能,拓宽了决策信息的规则化表达方式。

(2) 信息技术为决策服务。以数据挖掘技术为基础,对刑事法律法规和刑事案例进行本体化标注和实体信息的抽取研究,并以此形成支持决策的案例库,使决策活动直接依赖大数据提高决策水平,给管理科学的发展增加了新的内容。

(3) 推理集成促进决策质量。探索在专业领域,进行规则推理和案例推理的集成,形成从不同角度保证科学决策,提出新的决策路径。信息理论与决策理论的结合,会在不同领域的决策中起到示范效应,为其他相关决策研究提供新视角和方法。

(4) 管理科学与相关领域的交叉研究。将管理科学中知识表示、数据挖掘和决策方法与刑事案件量刑中的案件要素、法律法规推理和参阅案例研究进行了很好的结合,对服务专业领域决策具有一定的借鉴意义。

1.2.2 实践意义

法官作为量刑决策的主体,面临着来自知识、能力、领导和社会的各种压力,法官群体迫切需要用技术方法帮助其解决量刑决策问题,而且希望这种决策支持是自动的。通过数据挖掘和推理技术,在刑事案件量刑过程中具体解决拥有大量数据而有用信息却很贫乏的这种不正常的现象,无疑具有很强的现实意义。

(1) 确定描述刑事案件知识的方法。分析大量刑事案件,从刑事案件的组成细化和归纳出其构成,为法官提供一个分析案情的路线图,是绝大多数法官,特别是年轻法官迫切需要的。结合案情与裁判结果,努力借助大数据分析的思想,将法官从深陷于因果分析的劳累中解放出来,利用相关关系进行逻辑推理与判断,可以提高审判效率和审判质量,减轻审判的压力。

(2) 支持量刑决策的推理集成决策方法。法规与案例的数字化,使法官更容易得到法规的支持和案例的帮助。根据目标案件的案情,自动确定案件的刑罚类型,通过刑罚类型选择案例支持,进而得到目标案件的量刑结果和支持量刑的参考案例,使判决结果趋于均衡化。

(3) 提高量刑决策的能力。基于文本挖掘技术的发展,使得对法律法规和蕴含法官量刑智慧的判决书得以有效的信息挖掘,有效地提高了法官智慧的积累速度,扩大了法官经验传承的范围。从知识管理角度讲,提高了决策者内在的决策水平,突破了支持决策的知识瓶颈。

总之,研究案件规则化和推理集成,从而支持量刑决策,是对决策理论的丰富,也是决策理论在司法审判领域的创新应用,具有重要的理论价值与实践意义。

1.3 本书中相关概念界定

1.3.1 半结构化文本

半结构化文本是指满足以下条件的文本:文本由具有独立语义的内容模块组成,各模块可以由若干名词组成;每个独立模块间由非标点分隔,如回车;文本各模块内容可以归纳为一个名词集合,所有同类文本对应的集合是一个确定的有限集;每个模块内部语义的描述不一样,如简历和公文。法院的判决书是典型的半结构化文本。

1.3.2 领域本体

本体是用来描述某一领域甚至更广范围内的概念定义及概念之间的联系,使得这些概念和联系在共享的范围内有明确唯一的定义,达成一种共识,方便人机交流。领域本体是指在通用本体的基础上,以一个特定领域为描述对象的本体,提供该特定领域的概念定义和概念之间的关系、主要理论和基本原理,以及领域中发生的活动等。法律领域的本体可以分为法律条例本体和法律案例本体,刑事案件本体包括刑事法规本体和刑事案例本体。本书着重研究在对新案件进行量刑时,可以根据法规本体检索出相似案例进行刑罚类型的判定,通过案例本体进行刑期的判定。刑事案件本体属于法律领域本体的概念范畴。