法学理论，实证研究，机器学习与法官决策

摘要

此研究认为，法学理论可以通过实证研究融入到数据驱动的机器学习当中，从而辅助法官纠正决策偏误。因此，本文选取了司法判决中的同案同判理论，通过实证研究方法，利用文本挖掘技术，探究中国司法判决是否实现了同案同判。研究发现司法判决中，案情越相似，说理越相似，该结果能够稳定出现；但是说理越相似，刑期越相似却在改变基准案件时出现问题。这说明法官的决策存在问题。为了解决这个问题，此研究基于实证研究此研究提出一种理论驱动的机器学习样本筛选方法，对比筛选前后的四种机器学习模型效果，发现法学理论能够通过实证研究提高机器学习模型的表现，同时其预测结果纠正了不符合同案同判理论的案件结果，纠正法官的决策偏误。

关键词：同案同判　实证研究　机器学习　计算法学　法律决策科学

绪论

理想的司法裁量追求同案同判，以体现法律的一般性，但实践中受多种客观因素影响，这一目标并不总能实现，故同案同判转而成为司法要求。近年来，国内积极推进“智慧司法”建设，最高人民法院与国务院相继发布文件，强调运用大数据和人工智能技术优化司法服务，提升审判智能化水平。[①]司法建设的其中一个重要的目标就是实现同案同案。[②]量刑预测模型的构建成为司法智能化辅助审判的重要发展环节，有助于提高司法公正与透明。本文以抢劫犯罪为研究样本，通过实证研究揭示了国内抢劫罪司法整体上实现说理和案情之间的同案同判，但部分案件在说理和刑期上尚未达到同案同判，因此量刑工作成为当前的重点。此研究从法律理论上出发，发现了一种能提高训练效率的筛选案件的方法。通过构建可解释性较强的刑事犯罪刑期预测模型，以辅助司法审判活动量刑准确性和智能化。

一、文献回顾

（一）同案同判理论

作为法律人的共同追求，无论是中国法学家还是海外法学家都非常关注同案同判的实现。法哲学家哈特认为，正义的潜在含义就在于在多种多样的正义要求中，个体和个体之间在同一个正义的理念下是被相似对待的，而这进一步就能够推理出同案同判的原则（也就是说，平等是同案同判的“渊源”）。同案同判的原则内涵是丰富的，我们要注意同的是什么，更要注意他包含了不同案不同判的内涵。当然法律并没有这样的规定，这是一个“一般的规范”，或者可以认为是正确适用法律的结果而已。[③]早在 1958 年，哈特在讨论法律和道德具有某种联系时就在其论文中指出，“同案同判”这个原则是法律需要去执行的“正义”而不是法律的正义本身，或者说，这是法律正确被适用的应有结果。因此法律中存在内容为同案同判的一般规范，他们保证法律被适用在他应当适用的地方，或者至少有助于减少不平等的发生。[④]具体而言，相似的说理逻辑、判决方式甚至道德判断在相互相似的案件中应该保持一致性，如此才能符合“形式正义”的要求。形式正义包括其中一个内容正是法律体系整体的协调一致，他要求类似案件类似判决，这也是保护性的体现——为了保护期待利益，相似的案件要具有相似的判决。[⑤]支撑同案同判理论成立的理由从古至今发展出了丰富的论证，包括但不限于“法律

[①] 最高人民法院关于加快建设智慧法院的意见
[②] 最高人民法院关于案例指导工作的规定
[③] H. L. A. Hart, The Concept of Law, Oxford: Clarendon Press, 1996, pp. 155-161.
[④] H. L. A. Hart, Positivism and the Separation of Law and Morals. Harvard Law Review, Vol. 71, No. 4, pp. 593-629 (1958).
[⑤] David Lyons. Formal Justice and Judicial Precedent, Vanderbilt Law Review, Vol. 38, No. 3, pp. 495-519 (1985).

规则秩序解释的统一”、“防止武断”、“刑罚的可预见性”、“法律的稳定性要求”、“判决的公开和权威保障”以及“可期利益的保护”等等。但是要注意的是，Andrei Marmor 指出这个问题（同案同判理论）不是一个简单的正确或错误的问题，他没有一个标准的答案。如果为了这些理由需要不同判，那么违背这个原则也是被允许的。也就是说，存在一种原则可以凌驾于“同案同判”之上。[①]尽管同案同判具有一些缺陷，但是，同案同判可以在大多数情况下阻止不好的结果发生，而在一个需要不同案同判的情况下，则需要对他进行评判，同案同判的原则就是对相同案件不同对待的一种警告。因此，尽管同案同判是一种“弱主张”，但是他被普遍接受。在实践中，国际上也普遍接受了同案同判理念。例如，对难民领域的司法决策不统一甚至引发了 1990 年代国际难民司法判决协会（International Association of Refugee Law Judges）的建立。[②]

同案同判在中国被认为来自于“法律人人平等”的宪法原则推导而来，[③]也曾一度引发了大讨论，尤其是其定义和存在的必要性。以周少华为代表的认为同案同判是一个“虚构的法制神话”，并且给出了同案同判是“相同的案件相同的判决”这样的定义，主张同案同判是从适用出发，却追求结果的实质公正的悖论。“同案同判”仅仅形式化地表达了刑法的平等适用原则，却无法说明“同判” 本身在实质意义上的正当性。[④]孙海波紧接着对其进行了回应，认为同案本质是类案类判，也不代表差异化判决不可以被容忍。[⑤]之后孙海波更从提高司法裁判的可预测性和确定性、保护人们的信赖利益、限制法官的自由裁量权等后果性三方面论证了同案同判的正当性，这也和先前张琪的研究相互呼应。[⑥]这场论战至少达成了一个共识，以周少华教授为主的认为同案同判是虚假神话的学者向同案同判妥协，并指出在刑事案件中同案同判当然有其价值。[⑦]雷磊在这场论战之后发表了一篇总结性的文章，认为“同案同判”指的就是“类似案件类似处理”，“同案”的确切所指是“同类案件”。[⑧]也就是“同类案件”要使用“同类判决”。因此本研究以雷磊教授对其的定义为基准出发，同时也是这场论战最后达成的共识，即同案

---

① 这一系列的同案同判理论上的支撑和同案同判是一个“弱主张”的讨论参见 Kenneth I. Winston. On Treating Like Cases Alike, The Independent Review, Vol. 4, No. 1, pp. 107-118 (1999) ; Andrei Marmor. Should Like Cases Be Treated Alike, Legal Theory, Vol. 11, Iss. 1, pp. 27-38 (2005) ; 陈景辉：《同案同判: 法律义务还是道德要求》，载《中国法学》2013 年第 03 期，第 46-61 页。

② Hugo Storey. Consistency in Refugee Decision-Making: A Judical Perspective, Refugee Survey Quarterly, Vol. 32, Iss. 4, pp. 112-125 (2013).

③ 白建军：《同案同判的宪政意义及其实证研究》，载《中国法学》2003 年第 03 期，第 131-140 页。

④ 周少华：《同案同判：一个虚构的法治神话》，载《法学》2015 年第 11 期，第 131-140 页；周少华：《刑事案件“同案同判”的理性审视》，载《法商研究》2020 年第 37 卷第 03 期，第 3-15 页。

⑤ 孙海波：《“同案同判”：并非虚构的法治神话》，载《法学家》2019 年第 05 期，第 141-157、195-196 页。

⑥ 孙海波：《类似案件应类似审判吗？》，载《法制与社会发展》2019 年第 25 卷第 03 期；张琪：《论类似案件应当类似审判》，载《环球法律评论》2014 年第 36 卷第 01 期，第 21-34 页。

⑦ 周少华：《刑事案件的差异化判决及其合理性》，载《中国法学》2019 年第 04 期，第 145-164 页。

⑧ 雷磊：《如何理解“同案同判”？——误解及其澄清》，载《政法论丛》2020 年第 05 期，第 28-38 页。

同判内涵为类案类判，是法治所应当追求的目标。但是为了操作化还要进行进一步区分。所谓同案，就是相似的案件，具有相似的情节、行为等等，这点很好理解。但是要注意这里的变量特别多，尤其是被告的情节方面的案情，通过列举变量赋值的方式是不可能穷尽的。更加棘手的是“同类判决”指的是什么呢？从 Christopher Sherri 和 Avani Mehta Sood 对裁判说理的描述和讨论中可以一探究竟。作为早已关注司法的同案同判的问题的美国，他们指出：一份不能同案同判（或者译为，不连贯）的判决结果（inconsistent verdict）要面临合理的怀疑（reasonable doubts），也意味着文书需要公开，在监督之下以防止他是错判（wrongful verdict）。在这个过程中，一份精细的裁判文书（special verdict）需要对怀疑的问题进行详细说理，从而论证其合理性。[①]这意味着，同案同判有这样的内涵：第一，同案之间要具有相同的判决结果，由此在逻辑上可以衍生出类案也应当具有类判结果。第二，判决应当具有充分说理，尤其是对合理怀疑（或者说是争议焦点）的充分论证。由此衍生出相似的判决结果应当匹配相似的说理，如果某案件具有转化型抢劫之要件，然而 A 法官以转化型抢劫判处 n 年有期徒刑，B 法官则说理论证此行为是盗窃+故意伤害两罪并罚判处相同强度的 n 年有期徒刑，显然本研究不能接受这属于“同案同判”。更具体来说，同案同判要实现的是法律适用、解释、量刑范围、过程的统一，或则说是一种“动静态结合”的统一。[②]第三，由第二点论证和第一点论证结合可以得到，在类似案件下，法官应当具有相同的说理（或者译为，理由和逻辑，reasons and logic）来推理出相似的结论，如此才能实现正义。[③]于是本研究得到了一个更加完整的理论，即“同案同判链条”。也就是案情——>说理——>刑期。这一链条重点关注的不仅是结果，更是说理过程，是“动”和“静”的同判的结合。

早在 2011 年，中国司法实践的“案例指导制度”就充分体现了同案同判原则。这是一种制度化实践，不但可以更好的落实“同案同判”的目标，而且还会进一步强化它本身所拥有的重要性，使之成为不能被轻易抛弃的部分。[④]另外，近几年来人工智能和大数据引发了司法审判的极大兴趣，而要在法院系统推行这种审判辅助技术，一个重要的前提就是承认同案同判对于裁判的基础性意义。2017 年的《最高人民法院关于加快建设智慧法院的意见》更主张要建设智慧法院，运用机器学习、深度学习技术“满足办案人员对法律、案例、专业知识的精准化

[①] Christopher Sherrin. Inconsistent Verdicts and the Possibility of Innocence: A Comment on R v RV, Wrongful Conviction Law Review, Vol. 2, No. 1, pp. 78-81 (2021) ; Avani Mehta Sood. What’s So Special About General Verdicts? Questioning the Preferred Verdict Format in American Criminal Jury Trials, Theoretical Inquiries in Law, Vol. 22, No. 2, pp. 55-84 (2021).
[②] 刘树德：《刑事司法语境下的“同案同判”》，载《中国法学》2011 年第 01 期，第 68-76 页。
[③] Andrei Marmor. Should Like Cases Be Treated Alike. Legal Theory, Vol. 11, Iss. 1, pp. 27-38 (2005).
[④] 陈景辉：《同案同判：法律义务还是道德要求》，载《中国法学》2013 年第 03 期，第 46-61 页。

需求，促进法官类案同判和量刑规范化”。学者对这一建设目标积极响应，开始思考如何运用人工智能帮助司法的同案同判实现。左卫民指出，类案类判可以为疑难案件提供新的解决途径，也能够统一司法裁判尺度，避免司法裁判不公，而引入人工智能技术有利于这个目标的实现。[①]2020 年中国实践中出现的类案检索制度，提出了了同案同判不能仅仅关注结果的价值面向。《指导意见》指出：“类案检索说明或者报告应当客观、全面、准确，包括检索主体、时间、平台、方法、结果，类案裁判要点以及待决案件争议焦点等内容，并对是否参照或者参考类案等结果运用情况予以分析说明。”这表明类案检索仍然需要关注判决结果是如何说理的。《意见》中表明法官需要关注焦点问题是如何解决的，类案是如何参考的等等问题，实际就是“说理”类似的要求。因此，无论在理论还是中国的实践中，我们都认可，同案同判应当包括说理上的同案同判，而不能只关注“结果正义”。因此可以自信地说，无论在学界还是实务界，司法领域的同案同判在中国已然成为共识。回到本研究要讨论的问题，本研究将先前的理论整合，结合实践中的《指导意见》，本研究认为同案同判是司法的追求，尽管同判不代表正义，但他被认为是正义的结果。**所以，司法的整体样态应该是：大多数判决是符合同案同判的三环节链条的，只有部分离群案件不符合这个规律。对这种离群案件完全可以单独提取查看不合理性。如果中国的司法判决符合这个规律，那么本研究就有信心说，中国的司法判决基本实现了同案同判这个一般规范目标。**

（二）先前的实证研究与讨论

中国的同案同判大论战中也存在着对实然状态的大讨论。如孙海波认为有学者因为看到了司法实践中频发的 “同案不同判”而认为这一理论不正确是不可取的。[②]但是没有一个学者可以很好地论证实然样态的同案同判是什么样的，大多数学者仅仅以个别案件的差异判决就认为同案同判在实务中并不存在。当然有不少研究对这个问题作出了实证检验，只不过局限于静态的，也就是“结果”的相同。其中白建军以实证研究指出中国司法实践呈现这样的状态：“尚有相当范围内的法律适用没有实现同案同判、等量等罚而是同案异判、等量异罚。”[③]白建军、吴雨豪更基于同案同判提出了集体裁量模型来实现这一目标，希望通过裁量模型稳定法官的量刑差异。[④]文书的说理应当是司法裁判中不可忽略的重要环节，说理不是判决结果的附庸（当然一些法律怀疑论者很可能不赞同这种观点，这涉及到法哲学的争论，

[①] 左卫民：《如何通过人工智能实现类案类判》，载《中国法律评论》2018 年第 02 期，第 26-32 页。
[②] 孙海波：《“同案同判”：并非虚构的法治神话》，载《法学家》2019 年第 05 期，第 141-157、195-196 页。
[③] 白建军：《同案同判的宪政意义及其实证研究》，载《中国法学》2003 年第 03 期，第 131-140 页。
[④] 白建军：《基于法官集体经验的量刑预测研究》，载《法学研究》2016 年第 38 卷第 06 期，第 140-154 页；吴雨豪：《量刑自由裁量权的边界：集体经验、个体决策与偏差识别》，载《法学研究》2021 年第 43 卷第 06 期，第 109-129 页。

本文不拟讨论），说理是法律和大众的沟通，在裁判中具有独立的地位，在研究中不可忽视。[①]然而有关同案同判说理方面研究仍然较少见，其中可能的原因在于文本处理的繁杂和量化的难度较高。幸运的是随着计算机技术的发展，NLP 技术可以为这方面的研究提供相当趁手的工具。一般而言，裁判的逻辑是：认定事实（本院查明）——>裁判说理（本院认为）——>定罪量刑（判决如下）。目前学者做的主要工作是认定事实——>定罪量刑这个有欠缺的链条的研究工作，缺乏完整逻辑链条（说理部分）的实证研究检验。因此本研究引入 NLP 技术来发展原有的同案同判实证研究的逻辑链。

（三）此前的预测工作

此前具有大量的量刑预测工作，但是法学工作者和计算机工作者各自为营。一些学者倾向于使用数据驱动的思维方式，而忽略了量刑要件的提取。他们通常依赖于语义分析，构建法条和案情文本的对应关系。[②]还有一种方法是利用 N-Gram 自动提取，缺点和前述的语义分析一致。[③]这些方法都容易产生伪相关，而且算法的可解释性太弱。部分学者尽管是理论驱动，使用了要件提取的办法，但是提取出现了严重的错误，比如把“首要分子”当作“主犯”。[④]另外，使用要件提取的方法进行分析的研究使用的罪名刑量跨度太小。[⑤]另外，此前的研究没有充分考虑同案同判的问题，也就是输出的结果并没有进行同案同判的检验。另外，许多数据驱动的方法可能存在算法歧视，因此必须利用可解释性更强的算法，才有可能投入实践。此外，浙江大学的团队基于同案同判理论制作了一个学习模型，他的工作原理是寻找类案然后通过类案的刑期推导出本案的刑期。但是这种方法不能排除“案件本身质量奇差”的情况。[⑥]**因此，本研究在数据提取上采用更加严谨的办法，同时筛选量刑不符合同案同判的案件进行剔除来进行训练，保证训练结果能够符合同案同判的规律，从而辅助法官量刑。**

研究设计与方法

（一）研究数据

[①] 周芳芳：《我国刑事判决说理的场域视角研究》，2018 年吉林大学博士论文，第 14-16 页。

[②] Haoxi Zhong, Guo Zhipeng, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, Maosong Sun. Legal Judgment Prediction via Topological Learning; Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, Maosong Sun. Few-Shot Charge Prediction with Discriminative Legal Attributes. COLING 2018.

[③] 舒洪水：《司法大数据文本挖掘与量刑预测模型的研究》，载《法学》2020 年第 464 卷第 7 期。

[④] 王芳，张蓝天，郭雷：《非线性递推辨识理论在量刑数据分析中的应用》，载《中国科学:信息科学》2022 年第 10 期，第 1837-1852 页。

[⑤] 陈鸿旭，陈铁今，王皓，田维，胡兵，王竹：《基于概率图模型的危险驾驶罪刑期预测》，载《四川大学学报(自然科学版)》2022 年第 06 期，第 13-18 页。

[⑥] Siying Zhou, Yifei Liu, Yiquan Wu, Kun Kuang, Chunyan Zheng, and Fei Wu. Similar Case Based Prison Term Prediction. CICAI 2022.

本文的数据来自 18 到 21 年的抢劫裁判文书。研究随机抽取了 5081 份文书，数据集在分析前已经过预处理，包括去除异常值和缺失值，并进行了标准化处理。同时，还删除了部分包含错别字的裁判文书。筛选的标准是：

第一，均为抢劫罪。如果犯多罪（数罪并罚），则不纳入本研究中，因为这会给刑期带来麻烦。如此选择的原因在于：

1、抢劫罪是比较高发的犯罪，样本量较大，随机抽样得到的结果不容易出现偏差。此外，相比盗窃罪和危险驾驶罪等很少使用简易程序，文书的内容较为丰富。而且，抢劫罪的刑期横跨了拘役到死刑，刑期跨度较大。

2、在刑法中，除了所有罪共有的刑事责任、犯罪形态、共犯、认罪认罚与否等等影响裁判的因素，抢劫罪还单独具有相当多的情况，如：抢劫具有 8 种情节加重犯，包括致人重伤死亡、持凶器等等；抢劫还具有转化型抢劫这个特殊形态，即我国刑法 269 条规定的“犯盗窃、诈骗、抢夺罪，为窝藏赃物、抗拒抓捕或者毁灭罪证而当场使用暴力或者以暴力相威胁的”，将被法律拟制技术转化为抢劫罪；在抢夺罪中，还出现了“飞车抢夺”为抢劫的特殊规定；另外，抢劫罪是复行为犯，包括“暴力或以暴力胁迫”和“为获取财物“两个行为，因此，其产生了非常多的共犯与未遂问题，例如：行为人暴力攻击他人后再起意取走财物，是抢劫还是故意伤害加盗窃、行为人暴力抢劫他人打晕被害人后，第三人加入一起拿走财物，如何认定共犯内容的问题等等。

3、抢劫罪处于财产犯罪中，其很容易与其它财产犯罪混淆，如敲诈勒索就极其容易与抢劫混淆。以收集的裁判文书“陈某勇抢劫敲诈勒索案”为例，法官以“采取暴力胁迫手段”和“采取恐吓手段”进行说理，从而区分其多个行为中哪些属于抢劫，哪些属于敲诈勒索。综上，抢劫罪为法官的裁判文书说理内容提供了丰富的素材，因此选择抢劫罪作为本研究的素材，可以让说理的相关数据具有更强的多样性，更好地反映问题。

第二，均为一审案件。理由在于：

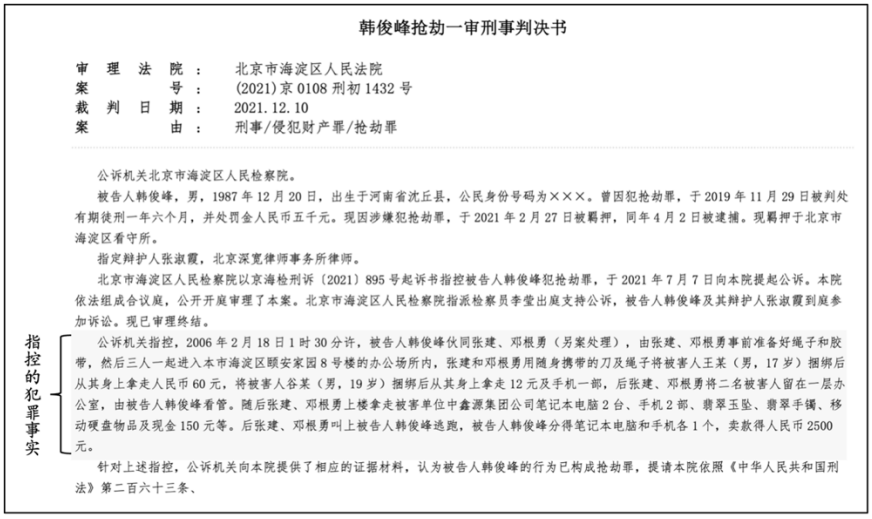
1、二审可能对一审的判决结果和说理均进行修改，形成不同的说理，为了防止这种情况导致结果的不稳定，本研究完全选用一审案件。二审对一审进行颠覆性的判决改变是否常见可以从发改率中一探究竟。《中国法律年鉴》的数据表明我国的改判率长期处于 10%—15%左右，[①]在最新的年鉴数据中，刑事整体上诉率（以结案的二审案件除去结案的一审案件）仅

[①] 《中国法律年鉴》1997—2007，转引自易延友：《我国刑事审级制度的建构与反思》，载《法学研究》2009 年第 31 卷第 03 期，第 59-76 页。

仅有 11.97%，[①]在这些案件中还要包括说理没有太大变化的情况，因此即便有部分一审被修正，也不会严重影响样本总体结构。

2、如前所述，二审经常出现维持原判的情况，为了提取的便利，本研究均适用一审案件，方便刑期的提取。

裁判文书一般具有一定的格式，大多数格式为“机关指控事实——本院查明事实——本院认为——判决如下”，其中，‘本院查明“部分为法院依据证据认定的事实，因为法庭的证据制度（如不合法证据制度），这部分事实与机关指控的事实并不一定一致，因此本院查明的案情才是判决的依据。“本院认为”部分为法院根据案情说理的部分，再通过说理得到最终的判决。要注意的是，部分文书没有本院查明部分，而是直接认定指控事实和本院查明事实一致，这类文书本研究直接录入机关指控部分。下面三张图展示了文书的切割方法：



韩俊峰抢劫一审刑事判决书

审　理　法　院：　北京市海淀区人民法院
案　　号：　(2021)京0108刑初1432号
裁　判　日　期：　2021.12.10
案　　由：　刑事/侵犯财产罪/抢劫罪

公诉机关北京市海淀区人民检察院。

被告人韩俊峰，男，1987年12月20日，出生于河南省沈丘县，公民身份号码为×××。曾因犯抢劫罪，于2019年11月29日被判处有期徒刑一年六个月，并处罚金人民币五千元。现因涉嫌犯抢劫罪，于2021年2月27日被羁押，同年4月2日被逮捕。现羁押于北京市海淀区看守所。

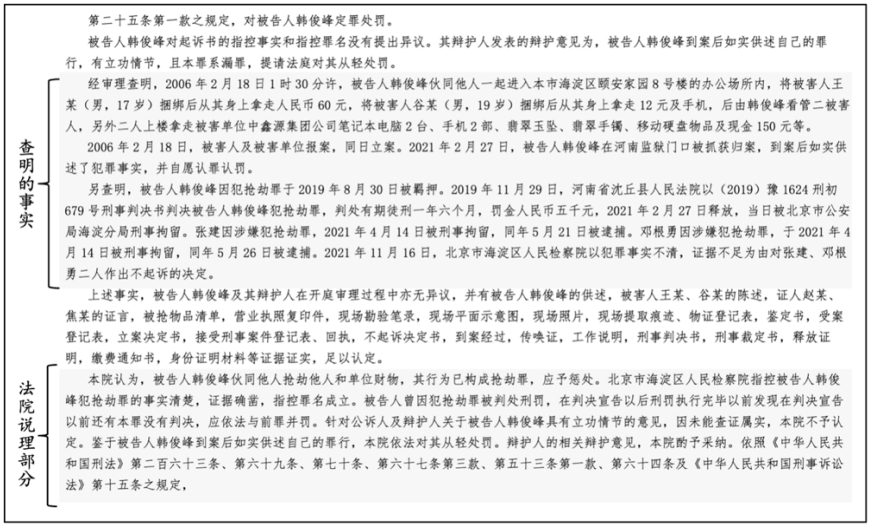
指定辩护人张淑霞，北京深宽律师事务所律师。

北京市海淀区人民检察院以京海检刑诉〔2021〕895号起诉书指控被告人韩俊峰犯抢劫罪，于2021年7月7日向本院提起公诉。本院依法组成合议庭，公开开庭审理了本案。北京市海淀区人民检察院指派检察员李莹出庭支持公诉，被告人韩俊峰及其辩护人张淑霞到庭参加诉讼。现已审理终结。

指控的犯罪事实

公诉机关指控，2006年2月18日1时30分许，被告人韩俊峰伙同张建、邓根勇（另案处理），由张建、邓根勇事前准备好绳子和胶带，然后三人一起进入本市海淀区颐安家园8号楼的办公场所内，张建和邓根勇用随身携带的刀及绳子将被害人王某（男，17岁）捆绑后从其身上拿走人民币60元，将被害人谷某（男，19岁）捆绑后从其身上拿走12元及手机一部，后张建、邓根勇将二名被害人留在一层办公室，由被告人韩俊峰看管。随后张建、邓根勇上楼拿走被害单位中鑫源集团公司笔记本电脑2台、手机2部、翡翠玉坠、翡翠手镯、移动硬盘物品及现金150元等。后张建、邓根勇叫上被告人韩俊峰逃跑，被告人韩俊峰分得笔记本电脑和手机各1个，卖款得人民币2500元。

针对上述指控，公诉机关向本院提供了相应的证据材料，认为被告人韩俊峰的行为已构成抢劫罪，提请本院依照《中华人民共和国刑法》第二百六十三条、



第二十五条第一款之规定，对被告人韩俊峰定罪处罚。

被告人韩俊峰对起诉书的指控事实和指控罪名没有提出异议。其辩护人发表的辩护意见为，被告人韩俊峰到案后如实供述自己的罪行，有立功情节，且本罪系漏罪，提请法庭对其从轻处罚。

查明的事实

经审理查明，2006年2月18日1时30分许，被告人韩俊峰伙同他人一起进入本市海淀区颐安家园8号楼的办公场所内，将被害人王某（男，17岁）捆绑后从其身上拿走人民币60元，将被害人谷某（男，19岁）捆绑后从其身上拿走12元及手机，后由韩俊峰看管二被害人，另外二人上楼拿走被害单位中鑫源集团公司笔记本电脑2台、手机2部、翡翠玉坠、翡翠手镯、移动硬盘物品及现金150元等。

2006年2月18日，被害人及被害单位报案，同日立案。2021年2月27日，被告人韩俊峰在河南监狱门口被抓获归案，到案后如实供述了犯罪事实，并自愿认罪认罚。

另查明，被告人韩俊峰因犯抢劫罪于2019年8月30日被羁押。2019年11月29日，河南省沈丘县人民法院以（2019）豫1624刑初679号刑事判决书判决被告人韩俊峰犯抢劫罪，判处有期徒刑一年六个月，罚金人民币五千元，2021年2月27日释放，当日被北京市公安局海淀分局刑事拘留。张建因涉嫌犯抢劫罪，2021年4月14日被刑事拘留，同年5月21日被逮捕。邓根勇因涉嫌犯抢劫罪，于2021年4月14日被刑事拘留，同年5月26日被逮捕。2021年11月16日，北京市海淀区人民检察院以犯罪事实不清，证据不足为由对张建、邓根勇二人作出不起诉的决定。

上述事实，被告人韩俊峰及其辩护人在开庭审理过程中亦无异议，并有被告人韩俊峰的供述，被害人王某、谷某的陈述，证人赵某、焦某的证言，被抢物品清单，营业执照复印件，现场勘验笔录，现场平面示意图，现场照片，现场提取痕迹、物证登记表，鉴定书，受案登记表，立案决定书，接受刑事案件登记表、回执，不起诉决定书，到案经过，传唤证，工作说明，刑事判决书，刑事裁定书，释放证明，缴费通知书，身份证明材料等证据证实，足以认定。

法院说理部分

本院认为，被告人韩俊峰伙同他人抢劫他人和单位财物，其行为已构成抢劫罪，应予惩处。北京市海淀区人民检察院指控被告人韩俊峰犯抢劫罪的事实清楚，证据确凿，指控罪名成立。被告人曾因犯抢劫罪被判处刑罚，在判决宣告以后刑罚执行完毕以前发现在判决宣告以前还有本罪没有判决，应依法与前罪并罚。针对公诉人及辩护人关于被告人韩俊峰具有立功情节的意见，因未能查证属实，本院不予认定。鉴于被告人韩俊峰到案后如实供述自己的罪行，本院依法对其从轻处罚。辩护人的相关辩护意见，本院酌予采纳。依照《中华人民共和国刑法》第二百六十三条、第六十九条、第七十条、第六十七条第三款、第五十三条第一款、第六十四条及《中华人民共和国刑事诉讼法》第十五条之规定，

[①] 《中国法律年鉴》2021

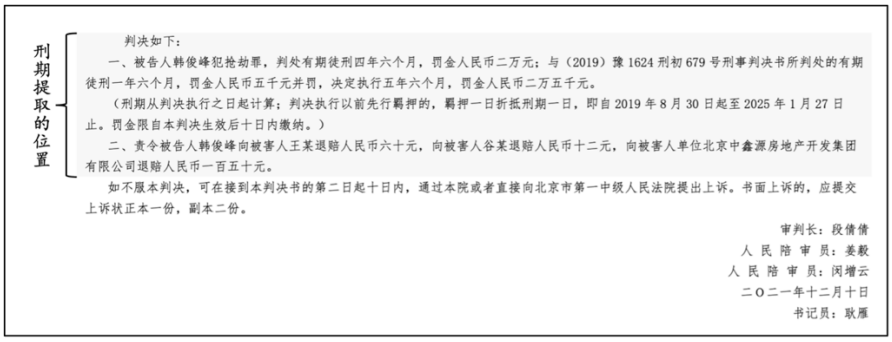


图 1.1 – 1.3

基于这样的规律，研究把这些文书处理成这样的样式：每个案件存储了案情、说理、刑期和一系列影响量刑的变量。案情部分和说理部分是非结构化文本数据，研究提取了裁判文书的“本院查明”部分作为案情，提取了“本院认为”作为说理。刑期部分，研究仅仅提取了主刑，并全部转化为月份，如 2 年 4 个月被转化为 28，尽管附加刑在刑期中也占有一定的比例，也有学者给出了相应的公式，[①]但是这些公式主观性太强，学者之间的公式也不相同，用于转化并不恰当，即便转化，也会面临众多质疑，无疑给模型的建立带来了很大的麻烦。一般而言附加刑强度都远远不如主刑，忽略主刑仅仅是一种数据降纬的结果，信息损失并不大，因此本研究只采用主刑。对于无期徒行和死刑，则以最高有期徒刑仅能判决 15 年的数据录入。影响量刑的要件部分，本研究通过团队中接受过法律教育的成员查询中国的法律文件来列举要件，然后从裁判文书中通过一定的逻辑提取，设置为哑变量。[②]法律人讲究“法言法语”，在文本处理的过程中，自然语言遇到的很大问题在于自然语言的多样性、意思表达的多样性、情感内涵的多样性、句子结构的多样性等等。但是法言法语提供了很好的平台，他的表达和措辞较为稳定，也几乎不出现一词多义情况。例如，如果法官在说理过程中想要表述被告的主观恶性较大，只会用专业的“主观恶性”这种词汇（除非他是不合格的法官），这让刑事司法判决书的措辞呈现出高度一致化。即便是案情描述，法官们也遵循几近相同的格式和措辞，如持刀这个事实，在文本中基本使用的是“携带”、“持刀”等词汇。因此只要文本中出现了某个关键词就可以表示这个要件。而一些特殊的要件，比如未成年人犯罪和对象是未成年人，仅仅使用未成年人是不能区分的，而需要再额外提取一个法条，来确定未成年人这个

[①] 白建军：《刑罚轻重的量化分析》，载《中国社会科学》2001 年第 06 期，第 114-125、206 页；白建军：《犯罪轻重的量化分析》，载《中国社会科学》2003 年第 06 期，第 123-133、208 页；白建军：《量刑基准实证研究》，载《法学研究》2008 年第 01 期，第 97-105 页；胡昌明：《社会结构因素对量刑影响的实证分析——以盗窃罪为例的案件社会学研究》，载《法律适用》2011 年第 03 期，第 54-59 页。

[②] 使用的法律文件包括：《中华人民共和国刑法及其修正案（十一）》、《最高人民法院最高人民检察院关于常见犯罪的量刑指导意见（试行）》（法发〔2021〕26 号）、《最高人民法院最高人民检察院关于常见犯罪的量刑指导意见（试行）》（法发〔2017〕7 号）。其中 21 年的指导意见对 17 年关于抢劫的改动并不太大，研究者进行了综合考量决定以 17 为标准。另外，尽管标准存在，但是可能标准出现前法官已经具有了这种集体经验，因此用 21 年作为辅助并无大碍。

词出现是作为被告还是被害。此外老年人、残疾人也是相同的逻辑。本研究提取的变量如下表所示：

表 1：变量提取列表

Types	Variables
犯罪方法 (Crime method)	入户、冒充军警、持枪、持刀/携带凶器、转化 (Entering the house, Pretending to be a military policeman, Carrying a gun, Carrying a knife/carrying a murder weapon, and Converting robbery)
犯罪地点 (Crime place)	交通工具、金融机构 (Rob in Transportation or Financial institutions)
犯罪次数和记录 (Times and prior crime)	累犯、前科、多次犯罪 (Recidivism, Ex-offenders, Repeat offenders)
犯罪形态 (Crime patterns)	预备、未遂、中止 (Preparation, Attempt, Abort)
犯罪地位 (Crime status)	主犯、从犯、首要分子 (Principal offender, Accessory, Ringleader)
被告人因素 (Defendants' factors)	自首、立功、坦白、自愿认罪、认罪认罚、精神疾病、未成年、赔偿 (Voluntary surrender, Meritorious service, Confession, Voluntary admission of guilt, Admission of guilt and acceptance of punishment, Mental illness, Juvenile, Compensation)
被害人因素 (Victim's factors)	谅解、和解、残疾、老年、孕妇、未成年 (Understanding or Reconciliation with defendants, Disability, Old Age, Pregnant, Minor)
伤情 (Injury)	轻微伤、轻伤、重伤、死亡 (Leading to Slight injuries, Minor injuries, Serious injuries or Death)
其他 (Other factors)	黑恶势力、数额巨大、灾害、军用物资、抢险救灾救济物资、罚金 (Underworld forces, Huge amounts, During disasters, Military supplies, Rob Emergency and Disaster relief materials, Fines)

（二）研究方法

本文实证部分对于文本数据，利用 rwmd 距离来计算相似度。对于刑期数据，使用欧几里得距离计算相似度。本研究使用 R 语言的 text2vec 包中已经训练好的 word2vec 和 glove 来进行词向量生成，保证近义词在文本相似度计算中被考虑，然后计算文本之间的向量距离。rwmd 方法擅长对长文本通过计算相似度然后分类，他的缺点也很明显，即无法处理否定词。但是这个缺点却很适合裁判文书数据的计算，因为裁判文书几乎不会出现否定词（法官不会说，

某人并没有主观恶性较大，而是仅仅列举肯定要件）。然后本研究选择一个案件作为基准案件，计算每个案件和这个基准案件的案情、说理和刑期相似度，再利用这些数据构建一元线性回归模型，查看其是否出现“案情越相似，说理越相似，刑期越相似”的规律。为了保证实证结论的可靠，本研究利用余弦相似度也进行了相同的操作；本研究也使用了更加传统的变量提取办法来对这个实证分析的结论再次进行验证；同时本研究通过随机抽样的方式来变换基准案件从而得到更加稳健的结果。

对于预测部分，由于该问题是预测模型，本研究考虑了在预测算法中表现较好的 lasso 模型、岭模型与线性模型对问题进行预测，通过 sklearn 库分别构建了模型，按照比例将数据集随机划分为测试集与训练集对模型进行训练和拟合，最终通过均方误差函数对训练结果进行评估。对于数据集部分本文采纳了两种数据集进行训练，然后对得到的结果进行对比。

三、研究结果

（一）司法判决实现同案同判了吗？

1. RWMD 方法和其回归诊断

表 2: OLS of RWMD

	No. 8	No. 1	No. 3923	No. 5071	No. 4806
C-R Cof.	0.15203	0.225497	0.116908	0.052300	0.16826
p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
R-P Cof.	-5.0235	-3.3419	-1.0013	0.1867	-1.948
p-value	< 0.01	< 0.01	< 0.01	0.3982	< 0.01

上表展示了 OLS 回归的结果。C-R 是案情-说理（case-reason），R-P 是说理-刑期（reason-prison term）。本研究随机抽取了五个案件作为基准，然后发现大多数案件能够实现同案同判，但是其中 5071 号案件存在一定的问题，即说理越相似，刑期却不会越相似。下列八张图展示了回归诊断结果。诊断结果表明此模型的变量基本呈现均匀分布，没有明显上升或者下降倾向，这表明其不存在共线性问题；样本的正态性良好；样本间不存在严重的内相关问题；没有点的库克距离长于 0.5，因此没有显著离群点。[①]

[①] Ford, C. 2020. Understanding Robust Standard Errors. *UVA Library StatLab*. https://library.virginia.edu/data/articles/understanding-robust-standard-errors/ (accessed February 1, 2023).

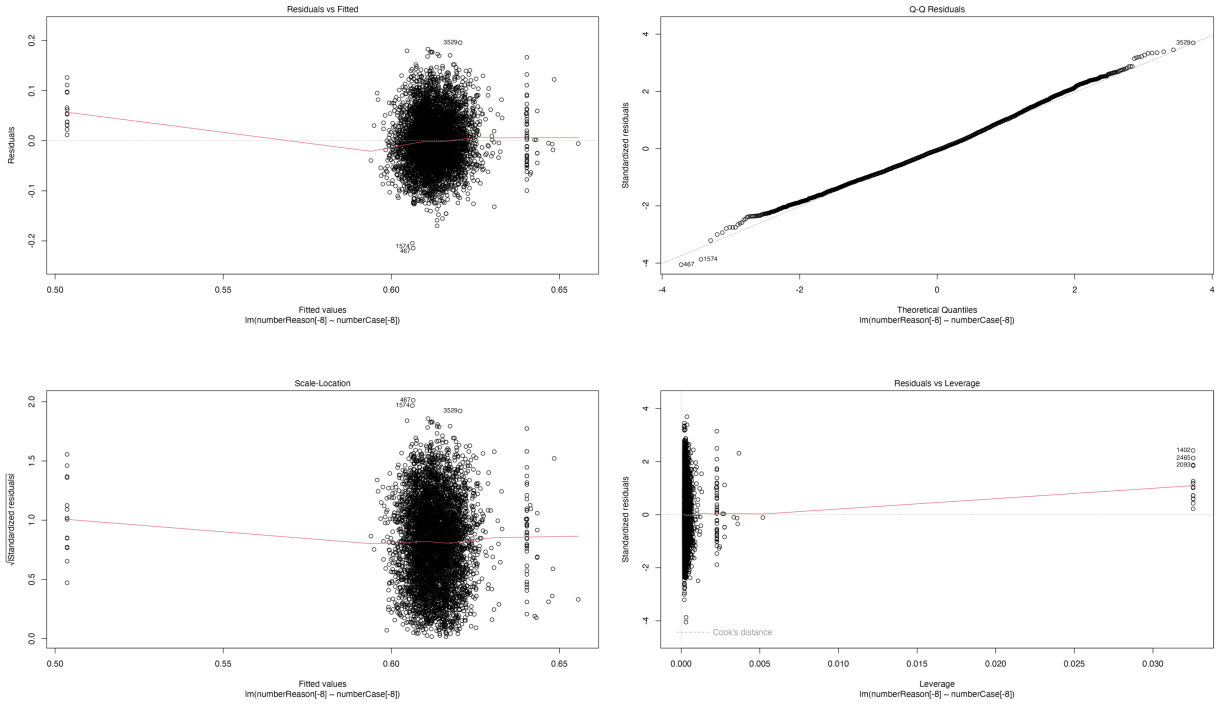

Residuals vs Fitted
Residuals
Fitted values
lm(numberReason[-8] ~ numberCase[-8])
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(numberReason[-8] ~ numberCase[-8])
Scale-Location
√|Standardized residuals|
Fitted values
lm(numberReason[-8] ~ numberCase[-8])
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(numberReason[-8] ~ numberCase[-8])

图 2.1 – 2.4

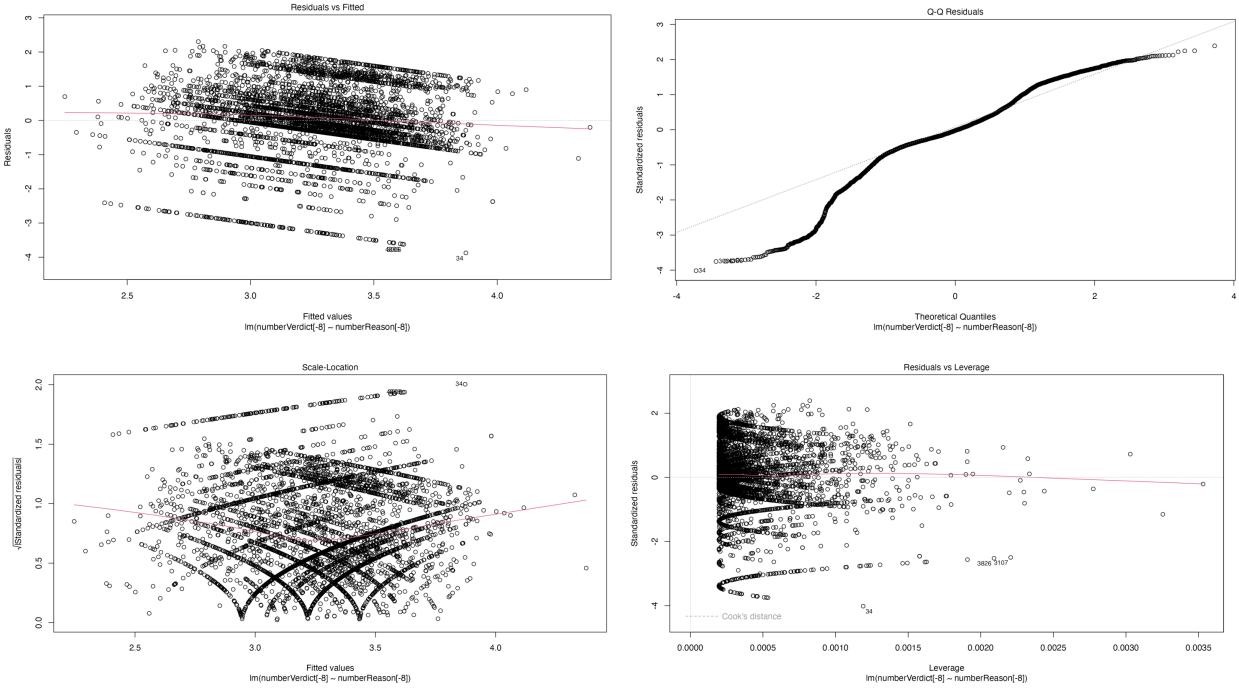

Residuals vs Fitted
Residuals
Fitted values
lm(numberVerdict[-8] ~ numberReason[-8])
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(numberVerdict[-8] ~ numberReason[-8])
Scale-Location
√|Standardized residuals|
Fitted values
lm(numberVerdict[-8] ~ numberReason[-8])
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(numberVerdict[-8] ~ numberReason[-8])

图 3.1 – 3.4

2. 稳健性检验

一般而言，法学领域的实证研究很少关注稳健性，大多数研究也没有回归诊断。但是这一研究则非常需要，其核心原因在于选择的基准案件的不可靠性和不同文本相似度计算方法产生的噪音不同。因此需要在这里补充一个稳健性检验结果。

稳健性检验有很多种方法，这里因为相似度计算问题，本研究使用替换基准的办法来检验稳健性。具体言之，本研究通常希望作为基准的案件是一个没有任何特别情况，非常标准，没有说理缺陷的抢劫罪模型。有几种方案可以选择：第一，选择最高院的指导性案例作为基准。但是最高院的指导性案例通常是为了解决某一特别法律问题诞生的，尽管权威，但是不符合“没有特殊情况”的要求。第二，作者自己选择一个认为适合作为基准的案件，然后公布这个案件的文书内容。然而这种方法并不可靠，因为选择的案件带有极大的个人主观性。公布的案件内容说理部分后，在读者之间也可能产生很大的争议。第三，使用不同案件的基准进行重复实验，得到相同结果则表明模型具有稳健性。这是本文采取的方法。最为可靠的方法是，对每一个案件都进行操作，但是这样对计算机的负担过大，还会造成时间和资源的浪费，因此随机抽样思想在这里是比较可靠的。即随机抽取三个案件，如果都没有出现了不显著结果，就可以认为这个结果可靠，因为刚好抽取到三个案件都幸运地能够让结果显著的可能性是极小的。本文前面已经使用了这种方法，报告了随机抽取的五个案件。

另一种可靠的方法是，改变变量。文本的相似度有多重计算方法，余弦相似度（Cosine Similarity）的基本原理是计算两个向量间夹角的余弦值大小，余弦值范围为[-1, 1]，余弦值越大，两个向量之间越相似。这样的方法常常被用在文本之间的相似度计算中。DTM 方法可以为长文本根据词频赋值，从而形成一个矩阵。通过 DTM 方法为裁判文书的文本进行分词并赋值，然后选取一个文本作为基准，以文本矩阵赋值的结果代入公式代码之中，计算所有文本和它的余弦相似度（当然和自己的相似度当然为 1）。通过这种方法，计算机就可以对成百上千的文本之间的相似度进行量化，从而汇报相似度结果。下面展示了余弦相似度的结果：

表 3: OLS of Cosine Similarity

	No. 1017	No. 4775	No. 2177	No. 5026	No. 1533
C-R Cof.	0.081784	0.045205	0.323915	0.212049	0.226299
p-value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
R-P Cof.	-1.0440	-0.1145	-0.6881	-0.5782	-0.6151
p-value	< 0.01	0.588	< 0.01	< 0.01	< 0.01

五次抽样中，只有案情—说理这一组相关性的链条是稳定显著的（每次的模型系数和整体 p 值都小于 0.01），另一模型无法出现较为稳健的结果，而这样就有理由认为其他模型的显著是因为抽样导致的随机结果，不够可靠。但是本研究发现，大多数情况下又能够出现负相关，因此结论值得商榷。本研究决定采用更多的方法进行检验。

另一种方法是相关相似度。在法律文本中，研究者经常做的操作是，把各种制度转化为一个列表，然后各种法律文本各自带有一个列表，整合起来变成一个矩阵，然后计算相关系数表示相似度。[①]文本项矩阵具有类似的结构，也可以进行相同的操作。直接使用 R 的 cor 函数计算相关系数即可。这个研究中相关系数的范围是 0 到 1（其实，相关系数的真正范围也是是-1 到 1，但是文本之间很难出现负数，除非一个文本有的词项另一个文本刚好没有，并且他没有的词项另一个文本刚好有）。下面展示的是同一个案件不同方法的对比，本研究发现方法之间具有稳健性。

表 4: OLS of different Algorithm

	RWMD	Cosine	Correlation
C-R Cof.	0.225497	0.249576	0.288393
p-value	< 0.01	< 0.01	< 0.01
R-P Cof.	-3.3419	-0.98190	-0.98056
p-value	< 0.01	< 0.01	< 0.01

3. 更准确的方法

本研究提取的变量已经在上表列出。为了保证探究的变量值存在法律文本内的要件，本研究用呀变量数据进行回归。本研究进行了全样本回归，输出不能够出现显著负相关的案件。然后本研究进行了输出，查看不同案同判的案件有什么，分析其特点。在所有的 5087 个案件中，本研究一共输出了 647 个案件不符合同案同判的规律。余弦相似度的计算要求不能全为 0，因此本研究添加了一列全部为 1 的数据进行运算。本研究仍然发现有很多案件不能符合这个规律。然后本研究希望观察这些案件的特点，于是本研究按照五年为重刑的标准，分别计算不同案同判的案件和同案同判的案件中非重刑的比例。发现分别为 18.08%和 72.899%。也

[①] David S. Law, Mila Versteeg. The Declining Influence of the United States Constitution, New York University Law Review, Vol. 87, No. 3 (2012); Anu Bradford, Yun-chien Chang, Adam Chilton, Nuno Garoupa. Do Legal Origins Predict Legal Substance?, Journal of Law and Economics, Vol. 64, No. 2, pp. 207-231 (2021). Yun-chien Chang, Property Law: Comparative, Empirical, Economic Analyses, Cambridge: Cambridge University Press, 2023, pp. 39-45.

就是说在不同案同判的案件中，大量的案件是重用了刑罚的。这一发现也符合直觉——法官在低刑期时表现更为谨慎，高刑期则表现更加肆意，裁量的空间更大。后续将更具体论述。

4. 对实证的总结

一个奇怪的现象是，为什么案件相似性和推理相似性之间存在显著相关性，而量刑与说理变量之间的相关性却不显著？本研究如何解释这种现象？我想用我的数据框架来讨论它。首先让我解释为什么相似的推理和案件相似性没有带来量刑的相似性。我认为可能有一个原因，那就是法外因素比本研究的预期发挥了更重要的作用。这导致预测裁决时的偏见。我之前提到过该因素，并且已经被学者证实，例如性别、种族，甚至法官的心情都会影响案件结果。但由于作为受过专业训练的法官或律师，他们会通过论理来隐藏量刑的真正原因。因此，论理可能是司法透明度和可信度的保证。[①]John 还指出，法官在审判过程中，部分会从理由推至结果，而有些法官则从结果进行倒推。[②]这种差异可能影响判决决策，进而出现该研究结果。此外，有些因素会在审判前后发生作用，但法官本人不会在判决中写下所有理由，例如，Björn Dressel 和 Tomoo Inoue 发现社交网络，无论是法庭外的还是法庭内的，都可能对判决发生作用，但没有人把它们写入判决，当然他们不必把它们全部写下来。[③]这个问题在中国尤为突出，在审判过程中法官往往掺杂道德因素。例如，一些法官可能从道德角度考虑判决。为了解决这个问题，陈和程建议法院应将这些要素公开。[④]笔者想补充的是，中国各省对法官的量刑决策有指导，这同样可能导致该研究结论。

因此，进一步的研究可以关注中国不同省份的情况。所有这些因素也导致了模型的 R2 较低。但是为什么案件相似性和推理相似性之间存在相关性呢？由于中国国家司法考试的专业训练，法官能够识别法律要求并正确应用法律，导致案件相似性和推理相似性之间的相关性。因此，法官可以为类似案件提供类似的推理，但量刑变化的具体原因却是个人和无法衡量的。例如，一些法官倾向于支持更严厉的判决，而另一些法官则更倾向于宽松的判决。更重要的是，中国国家司法考试是一个更多关注如何为案件说理的考试，通常忽略了如何正确地给出量刑。因此，中国的法官没有接受过量刑方面的训练。我相信许多其他国家也是如此。

---

① John Zhuang Liu, Xueyao Li. Legal Techniques for Rationalizing Biased Judicial Decisions: Evidence from Experiments with Real Judge. Journal of Empirical Legal Studies, Vol. 16, Iss. 3, pp. 630–670 (2019).

② John Zhuang Liu. Does Reason Writing Reduce Decision Bias? Experimental Evidence from Judges in China. The Journal of Legal Studies, Vol. 47, No. 1, pp. 83-118 (2018).

③ Björn Dressel, Tomoo Inoue. Informal networks and judicial decisions: Insights from the Supreme Court of the Philippines, 1986 2015. International Political Science Review, Vol. 39, No. 05, pp. 616-633 (2018).

④ Liang Chen, Jinhua Cheng. How to Import Moral Judgment into Judicial Adjudication. Exploration and Free Views, No. 08, pp. 59 72, 178 (2023).

加上前面讨论的原因，这导致了本文所获得的结果。因此，进一步的研究可以进行，以了解法庭外因素对裁决的确切影响程度，而不仅仅是调查哪个因素对裁决有影响。[①]

笔者想进一步说明的是，论理和判决之间不稳定的相关性。本研究可以看到，无论我使用文本分析还是传统方法，总是存在一个不稳定的显著线性模型。这可以解释为，如果本研究选择一个高质量的基准，例如我的样本中的第 1 个案例，其判决是正确的，那么本研究可以得到一个好的模型。如果基准案例本身就是有偏见的，那么模型的结果就不会很好。因此，如果法官的自由裁量被错误使用或没有任何限制，那么它将破坏类似案件类似处理的目标。另外，更严重的犯罪行为，会更容易出现错误判决，这可能的原因是，法官在低刑期时表现更为谨慎，高刑期表现更加肆意。同时，简单的比例量刑在基准刑高的时候允许的肆意性就变高了。从以上讨论中，本研究对中国判决的建议是更多地关注法官对高刑期案件的自由裁量权之限制，这也是中国最高法院通过一系列政策试图达到的效果。

（二）机器学习结果对比

1. 筛选操作

对于筛选部分，本研究在 RStudio 中进行运算，得出案由、说理、刑期三者之间相似度的相互关系。运用循环语句，本研究筛选出了大量的不同案同判的案件，然后进行回归。为了提高算法可解释性，本研究使用对案件本身的要件进行拆解之后的哑变量数据集代表案件相似度，然后再利用余弦距离计算案件之间的相似度，再进行回归。接着把每一个案件都各自作为基准案件重复同样的操作，如果出现某个案件作为基准时，案件的相似度和刑期的相似度无法出现显著负相关关系，那么这个案件应当被剔除。然后就形成了一组“同案同判”的数据集。本研究将在 RStudio 中表现一般，可能存在不符合同案同判的 600 个左右的样本去除后重新进行训练。逻辑在于，如果某个案件作为回归不能够显著负相关，那么这个案件就很可能不符合同案同判。本研究循环了 5078 次，删除输出的结果。

2. 构建模型和指标说明

本研究把判决书以 docx 格式存储，本研究首先通过 pandas 将判决书的文字部分写入到 csv 文件中的一列中。之后通过 openpyxl 库进行文书处理，本研究将判决书中公诉机关的主张和法院判决分割开，再从中确定案件的案由和说理，提高对量刑情节识别的准确性。本研究使用了 Python 和 RStudio 语言和 Pandas、Sklearn 等库进行数据处理，同时，为了计算余弦

[①] This question has been researched for years and has a lot of findings. Some early researches could be seen. See Reskin, Barbara, Visher, Christy. The Impact of Evidence and Extra-Legal Factors in Jurors' Decisions. Law & Society Review, Vol. 20, No. 3, pp. 423 438 (1986) ; Marilyn Chandler Ford. The Role of Extralegal Factors in Jury Verdicts, Vol. 11, No. 01, pp. 16-39 (1986).

值，本研究设置了一个虚拟变量，该变量的值恒为 1。本研究分别进行了三轮训练，然后对比了三次训练的结果。模型的参数设定如下：对于线性回归模型，本研究把测试集与验证集切割为比例三比七，其随机数种子为 40；对于岭回归。本研究仍然把测试集与验证集切割为比例三比七，随机数种子 48，系数 alpha 值为 1。对于 Lasso 回归。本研究仍然把测试集与验证集切割为比例三比七，随机数种子 4，系数 alpha 值为 0.1。对于多层感知机。本研究把测试集与验证集切割为比例二比八，随机数种子 4，设置一个隐藏层，包含 100 个神经元，允许迭代 2 万次,。对模型的评价上，本研究选用其得分和均方误差进行评价。其中，均方误差越小，模型表现越好，得分则通过 R-square 来评价。拟合程度越好，R-square 越接近 1。

3. 对比结果

如下表展示了 baseline（不筛选）的模型和 aft-filter（理论筛选后的模型）的 MSE 和得分情况表：

表 5: 前后模型比较

模型名称	Baseline R2	Baseline MSE	Aft-Filter R2	Aft-Filter MSE
Linear Reg	0.548/0.542	674.6032	0.680/0.784	245.6639
Ridge Reg	0.553/0.532	721.6177	0.701/0.687	257.6409
Lasso Reg	0.531/0.542	745.7592	0.682/0.682	264.0610
MLP Reg	0.764/0.504	747.2403	0.870/0.701	252.2937

四种模型的得分情况在引入法学理论后具有明显的提升，其 MSE 也显著减地减小了。尤其是三种线性回归模型得分情况较好。这一定程度上表明了利用理论帮助筛选数据的有效性。

4. 纠正效果法官决策效果

本研究在 R 语言中用 sample 函数随机抽取了三个不同案同判的案件，分别是第 1224，4166，3770 号案件，然后在网址中键入这些案件的情况，再计算是否能出现负相关。1224 号案件中，提取刑期为 180 个月，但是案件中没有任何重伤结果，因此明显存在偏重的情况。本研究把数据代入到模型中，除了多层感知机在这种高刑期表现不好（虽然很多案件是高刑期的），其他模型输出的结果均为 94 到 96 个月，然后本研究回代到代码中，出现了显著的负相关。之后的每个案件都进行相同的操作，得到下列结果：

表 6: 纠正效果表

	No. 1224	No. 4166	No. 3770	No. 1173	No. 2788
法官决策	180	37	36	171	180
模型结果	96	59	92	84	104
是否同案同判?	是	是	否	是	是

在 1224 号案件中，被告人仅仅是持刀入户抢劫，并且是初犯，也没有造成人员的伤亡，数额也并不巨大，只是存在法定的结果加重犯情节，罪不至无期徒刑，但是法官却为其期判处了无期徒刑，量刑明显偏高。在纠正后，量刑调整为 96 个月。4166 号案件中，被告人结伙持刀殴打被害人，法官为其量刑 37 个月，明显偏低，在其他没有持刀也没有进行殴打的案件中，被告是 36 个月的有期徒刑（比如不属于不符合同案同判的 4162 号案件），更何况本案中被本研究提取的是主犯，法官在说理中指出需要加重刑罚，却判处和一般抢劫相仿的刑期，不合理，纠正后为 59 个月的刑期。3770 号案件案情为入户盗窃后转化型抢劫，并且持刀，同时法官存在要件认定的一个小错误，即该案件根据司法解释应为既遂，而不是未遂。因此应当判处更高的刑期，而不是短短的 36 个月。这个案件没有显著负相关很可能是说理本身存在一些瑕疵，但是他能够出现负相关。1173 号案件为持刀抢劫，数额巨大，同时及时缴纳了罚金，除此之外没有其他法定的量刑情节，然而量刑居然达到了 171 个月，明显偏高，纠正后为 84 个月。2788 号案件的量刑为 180 个月，本研究检验发现是提取逻辑让这个案件存在一定的误差，实际提取为 48 个月。此案存在入户抢劫，被认定为未遂，但是实际上本案已经既遂。本研究修正为既遂，也是 105 个月的刑期。因此无论是 180 个月还是 48 个月，都不符合该案刑期。

结论

研究结果表明：第一，通过同案同判法学理论指导，用实证研究的方法融入机器学习模型，可以提高模型的预测表现。第二，模型在法学理论的指导下，可以一定程度上纠正法官的量刑偏误。这意味着，法学理论、实证研究、机器学习的关系是：法学理论可以通过实证研究融入到数据驱动的机器学习当中，从而辅助法官纠正决策偏误。法学理论在未来和计算机科学、机器学习技术的结合，可以通过这条道路进行。另外，本研究还提醒了计算机科学家，在数据端的筛选和质量的提高，对于预测结果来说是至关重要的。一味提升预测的准度并不能很好地实现“公平正义”的目标，因此理论驱动是必要的。本研究对法学的未来发展进行了探讨。主张法学的未来应当和其他学科积极交叉。