# 从计算语言学的角度分析香港的法律判决

Sankalok Sen[♣◇1]♣香港大学计算机科学系◇
*ssen2001@connect.hku.hk*

arXiv:2305.02558v1 [cs.CL] 4 May 2023

**摘要**

使用计算语言学分析和提取法律判决中的有用信息，是信息检索领域最早提出的问题之一。目前已有多家商业供应商能够自动完成此类任务。然而，在分析香港法律体系所作出的判决书时，高昂的费用和可用资源匮乏构成了关键瓶颈。本文尝试通过提供多种基于统计学、机器学习、深度学习和零样本学习的方法，有效分析香港法院体系的法律判决，以弥合这一差距。提出的方法包括：(1) 引文网络图生成，(2) PageRank算法，(3) 关键词分析与摘要生成，(4) 情感极性，以及 (5) 段落分类，以便从单份判决以及多份判决中提取关键见解。这将使香港判决分析的整体流程不再那么繁琐，并实现更高程度的自动化，从而通过快速推理迅速提取见解。我们还利用大语言模型对结果进行基准测试，并借助 HuggingFace 生态系统进行了稳健分析。

**关键词**。 法律判决、香港法律制度、自然语言处理、引文网络图、知识表示、关键词提取、摘要生成、情感极性检测、逐段语义分析

## 1 引言

在以下各节中，我们将简要介绍香港法律体系的历史、本论文从学术角度而言的重要性、本论文所考虑的工程方案，最后说明本论文试图实现的总体目标。

[1]本研究是在作者于2022年7月至2023年5月期间在香港大学担任研究助理职位时完成的。

## 1.1 香港法律制度简史

中华人民共和国香港特别行政区（HKSAR）的司法机构负责处理该地区的司法需求和要求。根据基本法 [1] 的规定，司法机构独立于立法机构和行政机构的影响。基本法于1990年4月4日经全国人民代表大会批准，并于英国移交该地区后，即1997年7月1日生效。基本法取代了由1917年[2]的香港英皇制诰和香港王室训令组成的殖民地规则。总体而言，在基本法保护下作出判决的香港法院大致分为表格1所列的8种不同类型。

| **法院** |
| --- |
| 1. 终审法院 |
| 2. 高等法院上诉法庭 |
| 3. 竞争事务审裁处 |
| 4. 区域法院 |
| 5. 家事法庭 |
| 6. 土地审裁处 |
| 7. 其他：裁判法院、劳资审裁处、小额钱债审裁处，淫亵物品审裁处，死因裁判官法院 |

表格 1：香港法院的类别

## 1.2 背景与动机

表1中提及的每个法院每周都会作出多项判决。为促进法律研究并开展学术教学，香港的法学院需要不断更新数据库，记录这些判决在案件类型、重要性和措辞方面的差异。这是一项极其繁琐的手动工作，因此有多家商业第三方公司提供计算自动化解决方案 [3]。然而，这类解决方案费用高昂，而且通常不适用于香港的判决。

因此，本文提出一种解决方案，将自然语言处理中的多种计算技术结合起来。这样便能更轻松地从无监督学习和监督学习两种视角，有效分析较新的法律判决。所采用的方法包括：(1) 基于引文网络图的知识生成，(2) PageRank算法，(3) 关键词分析与摘要生成，(4)

情感极性，以及(5) 段落分类，从而能够从单个判决以及一组判决中提取关键见解。

## 1.3 工程选择

近年来，随着性能更强大的计算处理器问世，许多论文开始运用深度学习模型的理论。深度学习的理念源自人脑的结构，而人脑由神经元构成。正如人脑会从新经验中学习，深度学习模型也被认为能够学习并模仿人类神经元。具体到自然语言处理领域，自 2014 年 Bahdanau 等人提出注意力机制以来，相关领域取得了巨大进展 [4] ；随后，Vaswani 等人提出了 Transformer架构 2017 [5],，这两者都运用了深度学习模型。这些模型在文本摘要、生成、情感分析和问答任务等自然语言理解任务方面取得了显著进展。

然而，将这些模型应用于特定的社会科学领域以加深对该领域的理解时，它们往往会泛化，因为这类模型通常是规模庞大的预训练模型，且训练所用的数据集极为通用，与社会科学领域的性质并不十分契合。因此，本文选择采用更通用的概率机器学习模型。本文提出的三个模型均基于这一理论。相比之下，深度学习被视为机器学习中更具生态位特征的子集，其模型似乎能够模仿人类神经元。在尝试解决社会科学问题时，研究者往往会选择更通用的机器学习模型，而非特定的深度学习模型，[6], [7], [8], [9], [10], [11]便是如此。

因此，本文通过将所提出的方法与深度学习模型（特别是大型预训练语言模型）的结果进行比较来评估，并基于这项分析，从精确率和准确率的角度得出哪些模型适用于任务的哪些部分。

## 1.4 目标

考虑到这些因素，本文提出了以下目标:

- 为每项包含引文的判决构建基于引文网络模型的知识图谱，以找出过去25年中更重要的引文，以及香港回归以来各项判决之间的因果联系。

- 实现 Google 的 PageRank 算法 [12] 并将其应用于已构建的引文网络模型，以找出数据中排名最高的引文

可用于为每份判决提供总体评分。

- 关键词分析算法（TextRank [13], YAKE [14],RAKE [15], KeyATM [16], LDA [17]）的实现，用于提取关键词和短语，并有效总结每份判决，同时使用BERT（基础版）进行基准测试。

- 实施情感分析（VADER [16]），以按段落提取判决中的情感分布。

- 实现针对每项判决的段落级分类器，使用朴素贝叶斯、支持向量机（SVM）、伯努利受限玻尔兹曼机（Bernoulli RBM）、随机梯度下降（SGD）、多层感知机（MLP）等算法，以及经过多体裁自然语言推理训练的 BART，并采用少样本学习方法进行基准测试，以提升各段落的语义提取质量。

# 2 文献综述

在接下来的章节中，我们将全面回顾过去二十年来运用机器智能开展法律判决分析的相关研究。

[19]介绍了一个意大利语和德语法律文本的平行翻译语料库，用于测试不同的翻译方法，例如对齐架构。 [20] 使用了一个由英国上议院判决编纂而成的语料库，用于创建与法律判决相关的有效摘要技术。 [21] 尝试采用一种问答方法，创建类似人类的工具，并用其参加美国举行的律师资格考试测试；他们将律师资格考试备考材料作为法律语料库。 [22] 考察了竞争者在日本法律文档判例信息抽取竞赛中提出的各种方法。 [23]介绍了一个数据集，用于在巴西法律文本上测试命名实体识别任务。[24]介绍了一个包含中国最高人民法院公布的超过 280 万起刑事案件的数据集，尝试使用机器学习和深度学习技术进行判决预测，并提供两种技术之间的基准比较。 [25] 提供了一个包含不同法律合同的数据集，并尝试将其总结为非正式英语，以便人类理解。[26] 介绍了一个根据《联合国人权公约》判决编纂而成的数据集，并尝试使用不同的神经网络技术预测结果。 [27]介绍了一个包含欧盟 5.7 万份判决的大型数据集，并尝试采用一种稳健的多标签分类方法。 [28] 介绍了一个用于中文司法阅读的数据集，其中包含 1 万份判决和 5 万个问题。

并将方法（如 BERT 和 TF-iDF）与人工标注的基准进行比较，回答问题并测试这些方法。

[29] 随机选取了中国法院在线发布的 5 万份中国判决书，并采用单层和多层分类以及双向 GRU 架构，对刑事案件中的罪名预测任务进行建模。 [30] 同样处理了罪名预测任务，但还在其深度神经架构之上使用了位置嵌入、词性标签、二元语法模型和 WordNet，从而获得了更高的预测准确率。[31]使用带注意力机制的编码器，为法律阅读理解文本准确预测判决。 [32] 采用马尔可夫网络模型，预测与离婚案件相关的判决。

[33] 尝试分析刑事案件，使用带注意力机制的标签条件序列到序列模型来处理法院观点生成任务。[34]使用基于注意力的编码器架构以及配备指针生成器的创新型反事实解码器架构，处理同一问题。 [35], [36], [37] 尝试使用不同的命名实体识别技术提取法律实体。 [38], [39]尝试使用时间推理等技术从法律文本中提取事件。 [40] 利用段落和引文信息，在法律判决上尝试了信息检索技术。[41] 结合摘要生成和词汇匹配技术，构建了用于信息检索的预训练短语评分模型。 [42]结合基于规则和统计的方法，首先创建了一个用于法律判决的自动摘要工具。 [43] 尝试使用 LDA 算法对法律文书进行摘要。 [44]有效运用领域知识方法进行法律文本摘要。

# 3 编程语言与工具

本项目利用香港大学计算机科学系 GPU 集群的资源，借助现代高性能处理器来取得成果。用于编写模型和算法的主要编程语言是 Python。

选择 Python 的理由是，它能够轻松处理大数据，并在数据集上轻松训练语言模型。脚本在 Jupyter Notebook 中编写，并在 GPU 集群上运行。选择使用 Jupyter Notebook，是因为它提供了便于交互式模拟的界面。例如，对于给定的数据集，可以轻松地以图形、表格和图表的形式查看不同类型的输出。GPU（图形处理器）是一种能够加快数值计算的特殊电路，因此使用香港大学的 GPU 集群来加快模型的整体计算。

# 4 数据集

法律自然语言处理数据集是自然语言处理实验室、香港大学计算机科学系（HKUNLP）另一项进行中项目的一部分，该数据集存放在项目主管的院系服务器上。每份判决均被提取并存储为一个 JSON 文件，其结构如表格 2 所示。

| **键 值** | **配对值** |
|---|---|
| 判决 | $[data_i, [type \quad = \quad \{other, para, heading, quote\}]] \quad \forall i$ (列表 of 列表) |
| ref | $[ref_i] \ \forall i$ (参考文献列表) |
| date | $[date_i] \ \forall i$ (所有聆讯日期) |
| 当事方 | $[party_i] \ \forall i$ (争议当事方) |
| 法官席前 | $[coram_i] \ \forall i$ （法官席前） |
| 代理 | $[rep_i] \ \forall i$ (所有各方的代理) |
| 案号 | [案号](提取的案号) |

表格 2：每份包含一个判决的 JSON 文档结构

在预处理为 JSON 文件之前，该文件是从香港法律信息研究所（HKLII） [45]下载的。预处理使用语义解析和正则表达式完成。本小节正式介绍法律自然语言处理数据集。从数学上看，数据集 $\mathcal{D}$ 由判决书 $|\mathcal{J}|$组成，可描述如下：

$$\begin{aligned} \forall c = \{c_1, c_2, ..., c_M\} \in \mathcal{C} \ (\text{Courts of Law}) \\ \exists j = \{j_{c,1}, j_{c,2}, ..., j_{c,N}\} \in |\mathcal{J}|_{C,N} \ (\#\{\text{Judgments per Court}\}) \\ \text{such that} \sum_{c=1}^{c=M} \sum_{j=1}^{j=N} J_{c_m, j_n} = |\mathcal{J}| = \mathcal{D} \end{aligned} \tag{Equation 1}$$

该数据集由约 115,000 份双语判决书组成，其中约 80,000 份为英文判决书，其余约 35,000 份为繁体中文判决书。如表格 1 所述，法院类别可划分为 28 个半宽泛实体，而该数据集中约有 18 个半宽泛实体存在相关案件。图 1 展示了概括这一分布情况的图。

共识别出73种不同类型的案件，总数为 112 [46], 种；每种案件均可通过独有的识别代码进行描述。数据集中的词语总数约为2.51亿，平均每份判决包含3000个词。

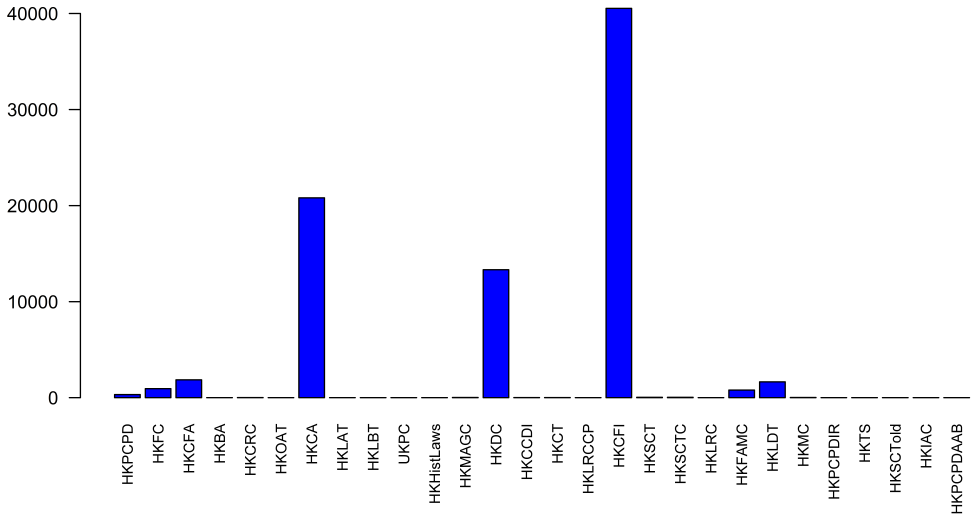



图1：案件数量与香港法院

# 5 方法

## 5.1 基于引文网络的知识图谱与 PageRan k

每项判决都有特定数量的引文；为将日期范围集中在特定的 25 年期间，已剔除移交前的案件。引文是根据其引用格式，使用多种句法方法提取的，包括暴力破解法和正则表达式，最终生成键值对字典数据类型。键值字典是指，对于每个唯一的键（判决引文编号），都存在一定数量的引文，并以类似列表的形式呈现，这些引文称为值。

清理预处理后的图，以应对不同引文格式造成的结构差异后，我们实现了PageRank算法。PageRank算法是对“Google算法” [12] 的重新实现——这是Google 搜索引擎用于根据特定搜索内容对网页进行排名的官方算法。我们将其应用于图引文网络，以找出香港回归后历史上最重要且对当今影响最大的案件。

然后，我们使用以下规则为引文网络图着色。各种颜色的含义如下：红色 [正在查看的主要引领案件], 蓝色 [引用红色案件], 绿色[引用蓝色案件], 黄色 [引用绿色案件], 紫色 [引用黄色案件], 粉色 [引用不同层级的混合案件]。粉色案件十分重要，因为它有助于我们追踪“过渡型”案件，即这些案件不仅引用了引领父案件，还引用了该父案件的其他子案件，这表明随着时间的推移，在对这类相似案件作出特定判决时，意识形态发生了转变，最终形成了一项判决

它们不仅从父案件中汲取灵感，也从引用了该父案件的不同子案件中汲取灵感。

如下所示的算法1展示了状态空间图的创建过程。

**算法1** 状态空间图生成

**要求：** DICT()类型的案件列表 $\mathbb{C} = \{c_1, c_2, ..., c_N\}$
**确保：** 状态空间图 $\mathbb{G} =$ DICT()
  **对于** $c_i$ 在 $\mathbb{C}$**中执行**
    **如果** $c_i$[案件编号] **不在** $\mathbb{G}$.KEYS() **则**
      $\mathbb{G}[c_i$[案件编号]] = []
    **结束如果**
    **对于** $r_i$ **在** $c_i$[参考文献列表] **执行**
      **如果** $r_i$ **不在** $\mathbb{G}$.键()**则**
        $\mathbb{G}[r_i] = [c_i$[案件编号]]
      **else**
        $\mathbb{G}[r_i]$.追加($c_i$[案件编号])
      **结束条件判断**
    **结束循环**
  **结束循环**
**要求：** $\mathbb{K} = \mathbb{G}$.键()
  **对于** $k$ **中的** $\mathbb{K}$ **，执行**
    $\mathbb{G}[k] =$ 列表(集合($\mathbb{G}[k]$))
  **结束循环**

该状态空间图由键值对构成，我们利用这些键值对生成一个包含 12068 个节点和 12663 条边的有向无环图。我们报告的图密度为 8.695648829995403e-05。我们使用阻尼因子的默认值 0.85 应用 PageRank算法，该值即为阻尼参数。

与网页类似，每个判决案件在互联网（或者在本例中，即香港法律制度）的万维网中都被表示为一个图。正如同名论文所述，PageRank 本身的灵感来源于学术引文分析，因此我们得出结论，将其应用于判决引文分析能够取得良好的结果，因为两者所要解决的任务具有相似的性质。

之后，我们实施算法2，以便根据特定的案例编号生成一个有色引文网络图，该图如引用网络着色中所述。

使用通过算法2得出的这一图，我们将每个节点的位置0、1、2、3和4分别与红色、蓝色、绿色、黄色和紫色关联。如果一个节点关联了多个位置，意味着它被交叉引用，因此将其标为粉色。

**算法2**引用网络着色

**要求：** 案例编号、子图边 $=[]$，树结构 $=$ DICT()，树结构[0] $=$ [案例编号-编号], 树结构[1] $=[]$，树结构[2] $=[]$，树结构[3] $=[]$，状态空间图 $\mathbb{G}$

**确保：** $\mathbb{L}$ =LIST(SET(LIST($\mathbb{G}$.前驱(案例编号))))

**对于** $l$ **，在** $\mathbb{L}$**中执行**

子图边.APPEND(($l$，案例编号))

**如果** $l$ **不在** tree-structure.键()**则**

树结构[$l$] $=[l]$

**else**

树结构[$l$].APPEND($l$)

**结束如果**

**结束循环**

**确保：** 深度 $=2$，循环 $=1$

**当** $\mathbb{L}$ ！$=[]$ **或** 深度 $>3$**时，执行**

**for** $l$ **in** $\mathbb{L}$ **do**

前驱-l = LIST(SET(LIST($\mathbb{G}$.前驱($l$))))

$\mathbb{L}_{\text{update}} = []$

**对于** $i$ 中的前驱-l **执行**

子图边.APPEND(($i$, $l$))

**如果** 深度**不在** tree-structure.KEYS() **则**

树结构[深度] $=[i]$

**else**

树结构[深度].APPEND($i$)

**结束 if**

**结束 for**

**结束 for**

$\mathbb{L}_{\text{update}}$.APPEND($i$)

深度 $=$ 深度 $+1$

循环 $=$ 循环 $+1$

**结束 while**

**确保：** 新图 $\mathbb{G}_{\text{new}}$

$\mathbb{G}_{\text{new}}$.ADD-EDGES-FROM(子图边)

## 5.2 关键词分析与摘要生成

我们对从 PageRank 中提取的一些排名靠前的案件执行以下可视化算法，然后对这些算法进行基准比较，以评估它们在香港终审法院（HKCFA）判决子集上的结果相似度。

### 5.2.1 TextRank

TextRank论文 [13] 提出了一种创新方法，通过将文本转换为图结构，并采用类似于以下所示 PageRank 的投票和推荐方法，来提取关键词并生成文本摘要

，使用相同的评分指标。

给定一个文档 $\mathcal{D}$，执行分词，然后根据分词后的文本构建一个图。接着，使用 PageRank 评分机制对该图进行排序，其中，对于构建的图中的一个顶点 $V_j$ ， $In(V_j)$ 是指向前驱顶点的顶点集合，而 $Out(V_j)$ 是指向后继顶点的顶点集合。顶点的得分定义为：

$$S(V_i) = (1-d) + d \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j)$$

参数 $d$ 是阻尼因子，默认设为 0。85，这一建议出自同名的 PageRank 论文。

### 5.2.2 快速自动关键词提取（RAKE）

RAKE 论文 [15] 提出了一种关键词提取方法，通过使用标点符号和停用词对文档进行分区，构建词级共现矩阵，并使用计算出的词语得分提取得分最高的词语。

初始候选关键词被选定为出现在停用词或短语分隔符之间的短语，随后通过关键词的共现关系进行图构建。根据构建的图计算单词 $w$ 的度数为 $deg(w)$，并将词频计算为 $freq(w)$，最后计算比率为 $^{deg}{}_{(w)\,freq(w)}$。对于最初选定的每个候选短语，其评分通过对该短语中每个关键词的单独得分求和来计算：

$$S(\text{cand-phr}) = \sum_{w \in \text{cand-phr}} S(w)$$

最后，它会考虑多次出现的停用词是否应纳入候选关键词短语，并对这些短语进行最终评分，再按评分降序排列。

### 5.2.3 YetAnotherKeyword Extractor（YAKE）

YAKE 论文 [14] 介绍了另一种近期的关键词提取器。它受 RAKE 启发，采用额外的统计估计量作为提取特征，包括词语位置、词频、术语与上下文的相关性，以及术语在不同句子结构中的表现。随后对术语进行评分、去重，最后重新排序。

候选关键词的评分如下：

$$S(kw) = \frac{\prod_{t \in kw} S(t)}{KF(kw) \times (1 + \sum_{t \in kw} S(t))}$$

$kw$ 表示用于评分的一个或多个 $n$-gram关键词， $S(t)$是构成候选关键词的 $n$-gram概率评分，而 $KF(kw)$是论文中所述的候选关键词的整体关键词频率。

### 5.2.4 潜在狄利克雷分配（LDA）

LDA 论文 [18] 介绍了如何从文本语料库中生成主题。根据该论文，数学描述如下。

对于给定的 $N$ 个文档语料库，每个文档长度为 $n_i$，选择 $\theta_i \sim Dir(\alpha)$，并为 $i = \{1, ..., N\}$设置稀疏参数 $\alpha < 1$。要提取 $K$ 个主题，选择 $\phi_k \sim Dir(\beta)$，并为 $k = \{1, ..., K\}$设置稀疏参数 $\beta < 1$。对于文档中的每个词语位置，以及文档长度 $i \in \{1, ..., M\}$，分别计算主题$t_{i,\ j} \sim Multinomial(\theta_i)$和词语 $w_{i,\ j} \sim Multinomial(\phi_{t_{i,\ j}})$。

## 5.3 情感分析

我们使用《情感推理效价感知词典》（VADER） [16] 论文，提取判决中每个段落的语言情感，从而深入了解法官作出判决时的情感变化。根据人工核查，大多数案件会在判决开头或结尾的段落中描述原告或被告一方诉求的受理或驳回情况。最后，我们通过绘制段落层面的情感变化分布图来进行可视化。

局限性：这种方法凸显了文档中缺乏对同情性或其他原始情绪变化进行区分的问题。可以通过文本分类为段落添加标注器来解决这一问题，从而辨别各段落之间的差异。例如，某个段落可能写道“让我们向受害者家属表示同情.....”，并被标记为正向语言情感。然而，在整个案件的语境中，对结果的分析可能存在歧义。不过，如果标注器能将其描述为案件事实描述/相关内容的一部分，而非判决法官的意见/裁定，就能极大提升法律研究人员/学者对语料库文本的理解。为应对这一挑战，我们将在以下小节提出多种逐段文本分类方法，并提出一种用于分析结果的基准测试方法，最后介绍一种创新的、基于零样本学习的快速推理方法。这种方法利用了近期大语言模型的兴起。这些模型通常基于海量数据进行训练，且由于能够更好地理解文本数据中的语言子结构，因此比训练所得的模型更有助于提升结果的泛化能力。

## 5.4 逐段文本分类

在下一节中，我们将针对逐段判决文本分类任务，实现多种机器学习和深度学习算法。仔细检查各份判决后，我们对其中50份文档的子集进行了分类，并将其段落分为4种类型:

- **关于：** 法官对案件的事实描述。
- **裁定：** 描述经中性化处理的法官裁定或意见。
- **准许：** 描述法官支持原告或被告的意见。
- **驳回：** 描述法官反对原告或被告的意见。

### 5.4.1 朴素贝叶斯

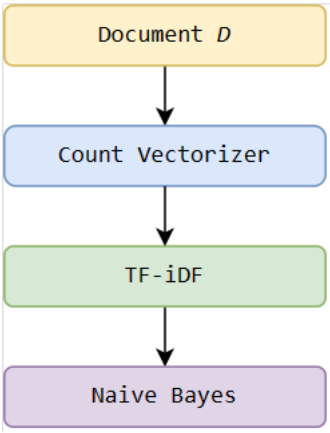



图 2：朴素贝叶斯模型

上图展示了已实现的朴素贝叶斯模型。计数向量化器有助于将文本文档语料库转换为词元计数矩阵。对于词语/术语 $t$、单个文档$d$、文档簇 $D$、文档总数 $|D| = N$，以及术语 $t$ 出现于其中的文档数量 $|d \in D$ ，词频–逆文档频率（TF-iDF） $t \in d|$定义为:

$$
\begin{aligned}
TF(t,d) &= \frac{freq(t,d)}{\sum_{t' \in d} freq(t',d)} \\
iDF(t,D) &= \log \frac{N}{1 + |d \in D : t \in d|} \\
TF\text{-}iDF(t,d,D) &= TF(t,d) \cdot iDF(t,D)
\end{aligned}
$$

为执行朴素贝叶斯分类，我们希望预测句子（$x_1$，$x_2$，..... $x_n$）的类别 $\hat{y}$：

$$\mathbb{P}(y|x_1, x_2, ..., x_n) \propto \mathbb{P}(y) \prod_{i=1}^{n} \mathbb{P}(x_i|y)$$

$$\hat{y} = \arg\max_i \mathbb{P}(y) \prod_{i=1}^{n} \mathbb{P}(x_i|y)$$

对于先验概率，我们考虑三种类型的朴素贝叶斯子模型：(1) 伯努利，(2) 多项式，以及 (3) 补集（适用于稀疏数据集的多项式）。

### 5.4.2 线性支持向量机

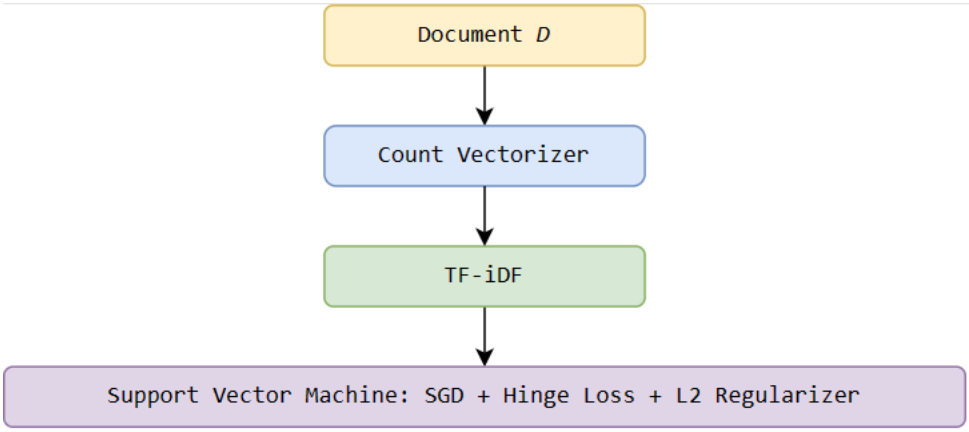



图 3：支持向量机

上图展示了已实现的支持向量机。对于给定的一组刑罚和类别（$\vec{X}, Y$），我们希望根据各自的类别对数据进行线性分离。为完成这项任务，我们使用随机梯度下降、合页损失和 $L_2$ 正则化项。将数据分成簇的超平面可定义为$w^T \cdot \vec{X} - b = 0$。给定参数 $\lambda > 0$，该优化问题可表述为：

$$\min \left( \lambda ||w||_2^2 + \frac{1}{n} \sum_{i=1}^{n} \max(0, 1 - y_i(w^T \cdot x_i - b)) \right)$$

求解这一问题可将数据分类到所需类别，并有助于评估基于拟合后的支持向量机模型的数据集。

### 5.4.3 逻辑回归

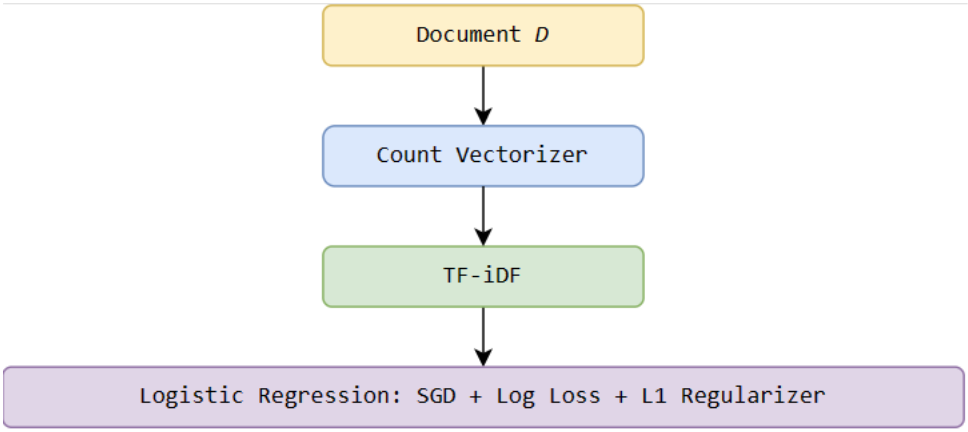



图4：逻辑回归

上图展示了所实现的逻辑回归模型。我们将先验概率计算如下：

$$p_X(\vec{x_i}) = \mathbb{P}(\vec{x_i}) = \frac{1}{1 + e^{-\beta \vec{x_i}}}$$

我们希望最小化对数似然估计，并预测类别如下：

$$C = \min \sum_{i=1}^{n} y_i \log(p_X(\vec{x_i})) + (1 - y_i) \log(1 - p_X(\vec{x_i}))$$

### 5.4.4 伯努利受限玻尔兹曼机

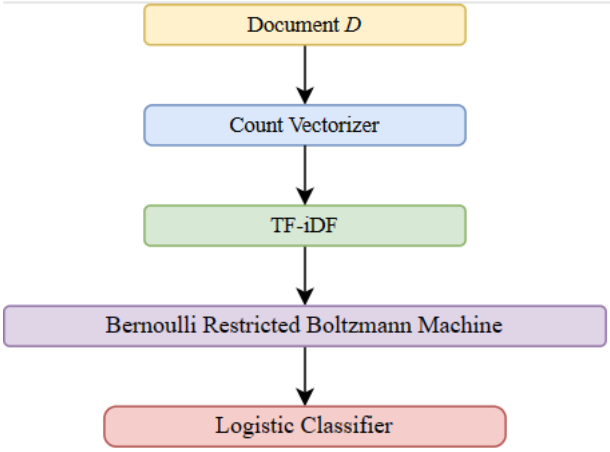



图5：受限玻尔兹曼机

上图展示了使用逻辑分类器实现的受限玻尔兹曼机。

它由可见的 $v$ （偏移量为 $a$）和隐藏的 $h$ （偏移量为 $b$）二值单元组成，是一种随机生成式神经网络和权重矩阵 $w_{i,\ j} \in W$。分类模型的计算如下：

$$
\begin{aligned}
E(v,h) &= -a^T v - b^T h - v^T W h \\
\mathbb{P}(v,h) &= \frac{e^{-E(v,h)}}{\sum e^{-E(v,h)}} \\
\mathbb{P}(h_j = 1|v) &= \sigma\left(b_j + \sum_i w_{i,j} v_i\right) \\
\mathbb{P}(v_i = 1|h) &= \sigma\left(a_i + \sum_j w_{i,j} h_j\right)
\end{aligned}
$$

对于要进行分类的输入刑罚，我们希望最大化以下内容：

$$
\arg\max_W \prod_{x \in \vec{X}} \mathbb{P}(x)
$$

### 5.4.5 带嵌入的基础线性模型

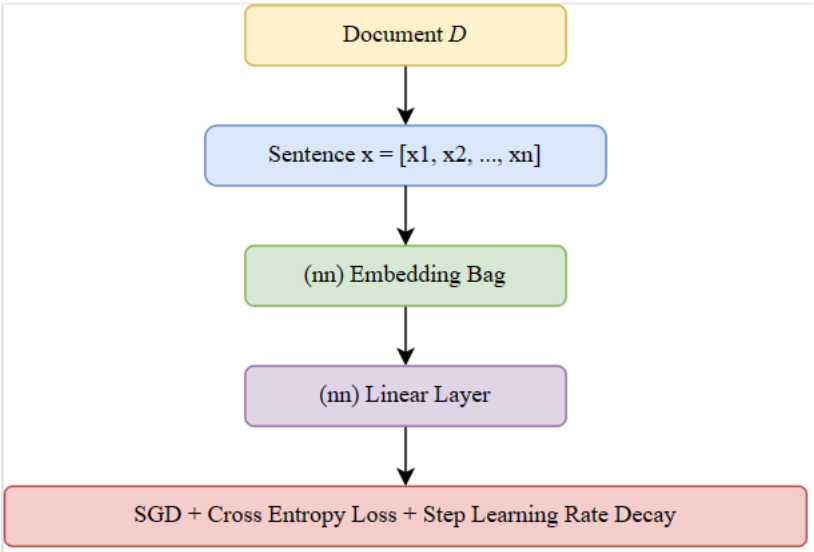



图 6：带嵌入的线性模型

对于基础深度学习模型，我们构建了一个由嵌入和线性层组成的简单模型。我们使用交叉熵损失、阶梯式学习率衰减和随机梯度下降算法进行优化。模型如图6所示。

### 5.4.6 编码器（长短期记忆网络） + 解码器（长短期记忆网络 +自注意力）

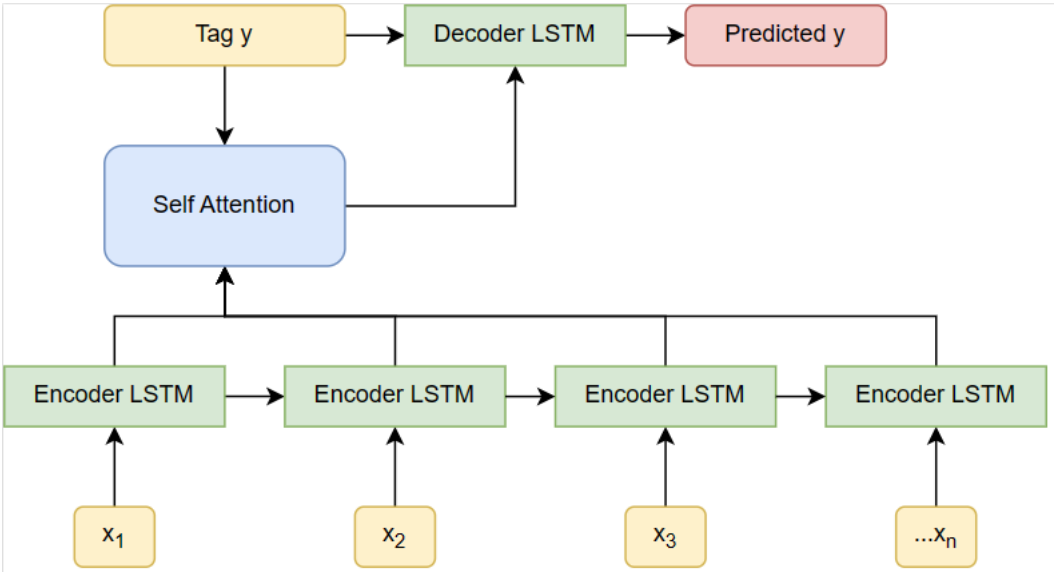



图7：编码器（长短期记忆网络） + 解码器（长短期记忆网络 + 自注意力）

对于改进后的深度学习模型，我们构建了一个基于长短期记忆网络的编码器，以及一个基于长短期记忆网络并采用自注意力机制的解码器。模型如图7所示。

我们针对键（$K$）、值（$V$）和查询（$Q$）对计算自注意力，如下所示：

$$\text{SelfAttention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{dim(K)}}\right)V$$

### 5.4.7 用于段落分类的 BART Large MultiNLI

最后，我们将结果与由 Facebook AI Research 创建的预训练 BART Large MultiNLI 大型语言模型进行比较，并展示其在快速进行判决段落级分类方面的实用性。

# 6 结果

在以下小节中，我们将介绍实验结果与研究发现。首先，我们陈述 PageRank 算法的结果。然后，我们展示已实施的机器学习、深度学习和零样本学习推理的结果，并比较不同模型的异同。最后，我们呈现从 PageRank 算法中提取的部分排名靠前案件的引文网络图、关键词分析及摘要方法的结果。

## 6.1 PageRank 结果

以下表格 3 所示的前 20 个案件，是使用默认参数通过 PageR- ank 算法提取的，并列出了相应的评分。

| 案件编号 | PageRank 得分 |
|---|---|
| 7 香港终审法院报告 187 | 0.00676 |
| CACV 284/2017 | 0.00232 |
| CACV 54/2018 | 0.00221 |
| CACV 219/2018 | 0.00199 |
| 2 HKLRD 1121 | 0.00162 |
| 1 香港法律报告 69 | 0.00162 |
| 1 香港法律报告 1 | 0.00136 |
| 3 香港法律报告 691 | 0.00115 |
| 10 香港终审法院裁判汇编 676 | 0.00108 |
| 1 HKC 261 | 0.00106 |
| 2 香港法律报告 437 | 0.00101 |
| CACC 338/2007 | 0.00098 |
| HCAL 106/2017 | 0.00096 |
| 5 香港法律报告 1 | 0.00095 |
| 5 HKCFAR 356 | 0.00084 |
| 2 香港法律报告 12 | 0.00071 |
| 2 香港法律报告 1 | 0.00070 |
| CACV 65/2014 | 0.00068 |
| FACV 2008年第16号 | 0.00066 |

表格 3：PageRank（判决前20名）

## 6.2 关键词分析结果

本节概述关键词提取模型的结果。我们假设，如果某个模型与其他模型具有较高的相关性和评分，或者相较于其他模型能够提取出质量更高、数量更多的关键词，那么我们就可以选择该模型。我们实现了 5 种模型：TextRank、RAKE、YAKE、LDA（生成单个主题）和 KeyBERT，其中 KeyBERT 作为基准模型，因为它是大型语言模型。我们在表4和表5中总结了结果。

| 指标/CN | HKCCDI HK | FAMC HK | MAGC HKSC | TC HKOAT | |
|---|---|---|---|---|---|
| **案件数量** | **8** | **749** | **20** | **37** | **1** |
| TextRank-RAKE | 0.0902 | 0.1118 | 0.0924 | 0.1067 | 0.1714 |
| TextRank-YAKE | 0.1360 | 0.1900 | 0.1258 | 0.1448 | 0.2553 |
| TextRank-LDA | 0.0645 | 0.1189 | 0.0610 | 0.0773 | 0.1162 |
| TextRank-KeyBERT | 0.0160 | 0.0279 | 0.0209 | 0.0216 | 0.0232 |
| RAKE-YAKE | 0.1000 | 0.0968 | 0.0787 | 0.1120 | 0.1509 |
| RAKE-LDA | 0.1250 | 0.0771 | 0.0582 | 0.0944 | 0.0869 |
| RAKE-KeyBERT | 0.0000 | 0.0160 | 0.0134 | 0.0186 | 0.0000 |
| YAKE-LDA | 0.3043 | 0.3843 | 0.3699 | 0.3516 | 0.2857 |
| YAKE-KeyBERT | 0.0000 | 0.0575 | 0.0840 | 0.0670 | 0.0454 |
| LDA-KeyBERT | 0.0000 | 0.0572 | 0.0621 | 0.0525 | 0.0909 |

表4：各法院关键词分析指标（第1部分）

| 指标/CN | 香港终审法院 | 香港区域法院 | HKCT | HKFC | HKMC |
|---|---|---|---|---|---|
| **案件数量** | **1544** | **8** | **12** | **872** | **18** |
| TextRank-RAKE | 0.1173 | 0.0902 | 0.1026 | 0.1127 | 0.0966 |
| TextRank-YAKE | 0.1899 | 0.1360 | 0.1804 | 0.1916 | 0.1289 |
| TextRank-LDA | 0.1452 | 0.0645 | 0.1252 | 0.1215 | 0.0627 |
| TextRank-KeyBERT | 0.0559 | 0.0160 | 0.0302 | 0.0294 | 0.0211 |
| RAKE-YAKE | 0.2005 | 0.1000 | 0.1107 | 0.1002 | 0.0806 |
| RAKE-LDA | 0.1448 | 0.1250 | 0.1068 | 0.0790 | 0.0605 |
| RAKE-KeyBERT | 0.0621 | 0.0000 | 0.0207 | 0.0165 | 0.0148 |
| YAKE-LDA | 0.4348 | 0.3043 | 0.3983 | 0.3858 | 0.3610 |
| YAKE-KeyBERT | 0.1280 | 0.0000 | 0.0920 | 0.0603 | 0.0813 |
| LDA-KeyBERT | 0.1488 | 0.0000 | 0.0820 | 0.0607 | 0.0523 |

表格 5：各法院的关键词分析指标（第 2 部分）

结论： 从展示计算所得关键词分析指标的两个表格中，我们得出结论：**YAKE-LDA** 之间的结果关键词共同出现比例最高。而 **KeyBERT** 自身的结果共同出现比例最低，这凸显出 KeyBERT 提取的关键词与传统模型提取的不同且独特。

因此，在考虑不同的关键词分析算法时，我们必须考虑当前的任务，并使用多种不同的算法，平均而言独立地可视化提取出的关键词，从判决中提取独特的见解。

## 6.3 摘要生成结果

在分析我们的摘要生成结果时，我们使用面向召回率的摘要评估指标（Recall-Oriented Understudy for Gisting Evaluation，简称 ROUGE）[47] 进行计算分析。我们将 Rouge-1、Rouge-2 和 Rouge-L 可视化，并列出它们的精确率、召回率和 F1 分数。我们将 TextRank 摘要生成与 Facebook 的 BART-Large-CNN 模型在摘要生成任务上的表现进行基准比较。

根据以下表格 6 汇总的统计数据，我们将其定义如下：

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

| **指标/CN** | **香港终审法院** | **HKCA** | **HKDC** | **HKFC** | **HKLDT** |
|---|---|---|---|---|---|
| **案件数量** | **25** | **25** | **25** | **25** | **25** |
| ROUGE-1 召回率 | 0.3242 | 0.3236 | 0.2783 | 0.2057 | **0.2233** |
| ROUGE-1 精确率 | 0.4489 | 0.5060 | 0.3958 | 0.4545 | **0.6285** |
| ROUGE-1 F1 分数 | 0.3270 | 0.3541 | 0.3019 | 0.2762 | **0.3297** |
| ROUGE-2 召回率 | 0.1719 | 0.1792 | 0.1236 | **0.0649** | 0.1069 |
| ROUGE-2 精确率 | 0.1999 | 0.2641 | 0.1670 | **0.1974** | 0.3578 |
| ROUGE-2 F1 值 | 0.1577 | 0.1811 | 0.1292 | **0.0950** | 0.1646 |
| ROUGE-L 召回率 | 0.2973 | **0.2945** | 0.2535 | 0.1822 | 0.1928 |
| ROUGE-L 精确率 | 0.4034 | **0.4512** | 0.3608 | 0.4068 | 0.5428 |
| ROUGE-L F1 | 0.2980 | **0.3192** | 0.2747 | 0.2458 | 0.2846 |

表格 6：所选法院的摘要生成指标

结论：我们计算了来自案件数量最多的 5 家法院的 25 个案件样本的 ROUGE 指标。关于摘要生成指标，我们得出以下结论：

- **HKLDT** 拥有最高的 ROUGE-1 指标 [单字]
- **香港家事法院** 具有最低的 ROUGE-2 指标 [二元词组]
- **香港上诉法院** 的ROUGE-L 指标 [最长公共子序列]

## 6.4 段落级判决分类结果

首先，我们在下表中总结了各模型的结果。我们选取了 1000 个段落样本，使用 4 个分类器进行 80-20 的训练集-测试集划分，类别如第 5.4 节所述。

| 模型/指标 | 准确率 | 精确率 | 召回率 | F1分数 |
|---|---|---|---|---|
| **玻尔兹曼机** | **0.55** | **0.30** | **0.55** | **0.39** |
| **伯努利朴素贝叶斯** | 0.73 | 0.67 | 0.73 | 0.69 |
| **多项式朴素贝叶斯** | 0.68 | 0.71 | 0.67 | 0.61 |
| **互补朴素贝叶斯** | 0.72 | 0.73 | 0.72 | 0.69 |
| **支持向量机** | 0.77 | 0.76 | 0.76 | 0.75 |
| **逻辑回归** | 0.78 | 0.78 | 0.78 | 0.77 |
| **多层感知机** | 0.74 | 0.74 | 0.74 | 0.73 |
| **嵌入+线性** | 0.62 | 0.61 | 0.62 | 0.61 |
| **E(长短期记忆网络)+D(长短期记忆网络+注意力)** | 0.98 | **0.97** | **0.94** | **0.95** |

表格 7：分类模型指标（测试集）

因此，我们看到伯努利受限玻尔兹曼机的表现最差，而所提出的编码器-解码器风格架构表现最佳。

局限性：然而，这些模型对测试集中的不同样本泛化能力不佳。因此，我们建议使用在多体裁自然语言推理数据集上训练的 *BART-Large*模型，以便更好地关注长段落中的语言子结构。对于较长的段落，我们的编码器-解码器架构表现不佳；当段落长度超过 *100* 个词元时，模型准确率会降至 *0.47*。由于预训练模型最多可处理 *512* 个词元，而我们的数据中段落长度均不超过 *473* 个词元，因此它具有更好的泛化能力。

结论：因此，我们得出结论：零样本学习能够以相对较快的推理时间，更真实地检验法律数据的文本分类，这得益于其在多个大型数据集上的预训练。BART 的架构[48]被描述为一种基于 Transformer 的编码器-解码器风格语言模型，包含双向编码器和自回归解码器。该模型使用经过任意加噪函数破坏的文本进行预训练，从而通过去噪学习重建原始文本。

下图展示了 BART 预测所完成的部分工作，并突出显示了经人工交叉核对后发现会使模型倾向于某一类别而非另一类别的文本。

PARAGRAPH: 14. In drug trafficking cases, the fact that all or part of the drug involve
d is for the defendant's self-consumption is recognized as a mitigating factor (see, fo
r example, R v Chan Mung-lung [1992] 2 HKCLR 127 and R v Chung Kam Fai [1993] HKC 42).
LABEL: ALLOWED
PARAGRAPH: 15. In R v Meah & Marlow (1991) 92 Cr App R 254, a case concerning drug traf
ficking by way of importation of drugs, Jupp J also made the following observations (at
256):
LABEL: RULING
PARAGRAPH: "Importing is a distinct offence from possessing. The penalties are diff
erent and in our view it is not right to say that this must be treated simply as a case
of possession. Nevertheless there must be a considerable reduction in sentence to refle
ct the fact that the drugs were for the appellant's own consumption."
LABEL: RULING

图 8：为什么 BART 的泛化能力更强？（第 1 部分）

PARAGRAPH: 22. With 15 previous convictions, the Appellant's criminal record is worse t
han those of other defendants in similar cases. It must also be borne in mind that the
Appellant re-offended when he was wanted. Having regard to self-consumption by the Appe
llant of most of the "ice" involved, the Judge reduced the starting point from 5 years
and 10 months to 5 years and 3 months. We agree that this 7-month discount (i.e. approx
imately 10%) is on the conservative side. Nevertheless, in our judgment, the Judge has
not erred in principle, nor is the final sentence of 3 years and 6 months on the first
charge manifestly excessive.
LABEL: ALLOWED
PARAGRAPH: 23. The ground of appeal advanced by the Appellant fails. We therefore dismi
ss his appeal against sentence.
LABEL: DISMISSAL

图 9：为什么 BART 的泛化能力更强？（第 2 部分）

## 6.5 案例分析 1：7 HKCFAR 187

对于以下案例，我们按段落展示关键词分析结果及后续的情感分布，并附上标签。

| 方法 | 结果 |
|---|---|
| TextRank 摘要 | 本上诉中的关键法律原则问题是，律政司司长是否热这里只有精品<br>在裁定潜在被递解人士的酷刑申诉时，依据<br>有权仅依赖联合国难民署未作解释的政策娱乐主管<br>拒绝向相关人士授予难民身份，但未能……<br>对该项主张作出任何评估。 |

表格 8：关键词分析与摘要：7 HKCFAR 187

| 方法 | 结果 |
|---|---|
| TextRank 关键词 | [’阿尼’, ’联合国难民署’, ’避难’, ’国家’, ’秘书处’, ’酷刑’, ’原因’, ’mr’, ’法律’, ’驱逐出境’, ’被驱逐出境者’, ’公约’, ’关注’, ’法律’, ’国家高’, “艺术”、“人员”、“相关考虑包括”、“斯里兰卡” ] |
| RAKE | [“焦虑的考虑” mr 普拉巴卡尔表示，” “十一月”, “基于”、“承认”、“可允许的途径”、“确定难民 身份”、“所主张的保护”、“九月”、“以不承担——的方式” 入境’, ‘接受联合国难民署’, ‘秘书仅遵循联合国难民署’, ‘遗漏’, ‘疑似’, ‘依据’, ‘难民陈述’, ‘是否遭受身体方面的痛苦—— 方面’] |
| YAKE | [‘秘书’, ‘联合国难民署’, ‘难民’, ‘公约’, ‘酷刑’, ‘答辩人—— ’, ‘人士’, ‘主任’, ‘香港’, ‘索赔’, ‘斯里’, ‘身份’, ‘有关’, ‘国家’, ‘命令’, ‘兰卡’, ‘条’, ‘原因’, ‘国家’] |
| LDA | [’秘书’, ’难民’, ’联合国难民署’, ’酷刑’, ’会’, ’答辩人’, ’person’, ’公约’, ’主张’, ’身份’] |

表格 9：关键词分析与摘要：7 HKCFAR 187（续）

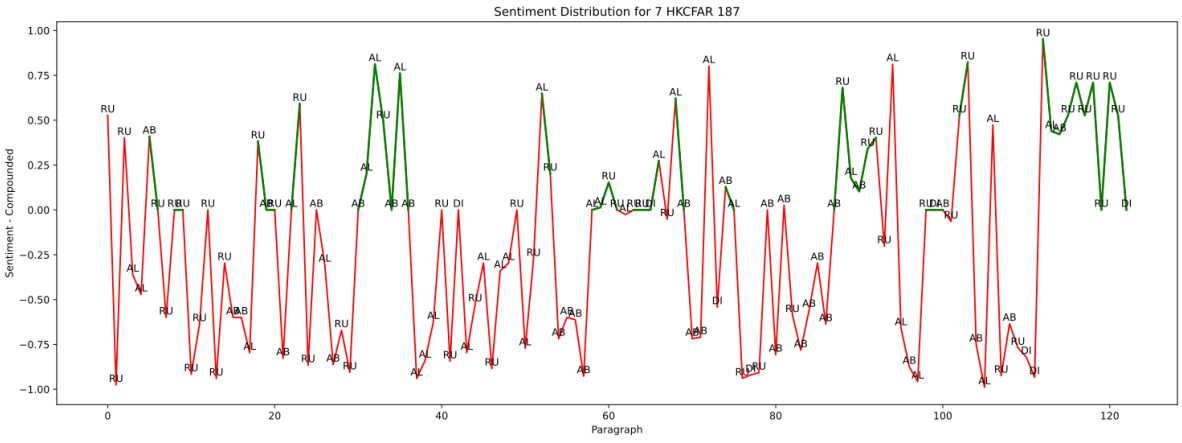



图 10：带标签的情感分布：7 HKCFAR 187

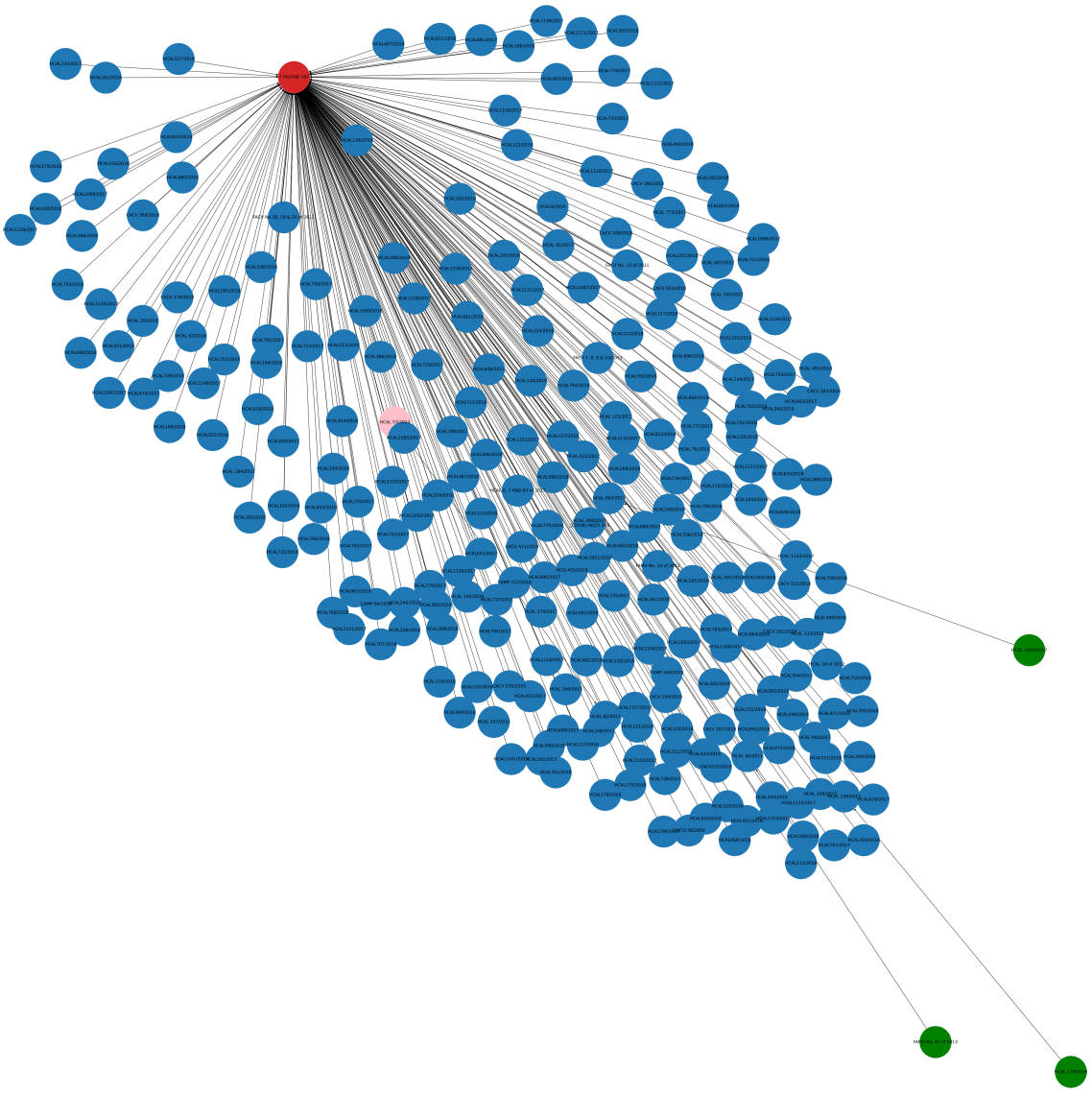


图 11：引文网络图：7 香港终审法院 187

## 6.6 案例分析 2：2 HKLRD 1121

对于以下案例，我们将按段落展示关键词分析结果及后续的情感分布，并附上相应的标注。

| 方法 | 结果 |
|---|---|
| TextRank 摘要 | 法官强调了贩运<br>危险药物罪行的严重性，并指出根据量刑<br>上诉法院所确立的指引，是……的起点。<br>贩运多达10克“冰毒”的刑罚为3至7年监禁。<br>ment. |
| TextRank 关键词 | [’判刑’, ’判决’, ’毒品’, ’ma’, ’控罪’, ’上诉’, ’Appel Chow Chun’, ’折扣’, ’案件’, ’罪行’, ’适当’, ’辩护’, ’考虑’, ’考量’, ’仅’, ’警方’, ’公平 香港法律报告’, ’贩运’, ’减刑’] |
| RAKE | [’法官考虑’, ’被告’, ’器具’, ’执行’, ’危险-<br>性毒品’, ’认罪’, ’宁路’, ’法院事实’, ’贩运’,<br>’警务人员’, ’上诉人强调’, ’折扣约为’,<br>’不足’, ’服刑月数’, ’已被判刑’, ’提出异议’, ’上-<br>接近’] |
| YAKE | [’上诉人’, ’法官’, ’法院’, ’毒品’, ’年’, ’月’, ’冰毒’, ’判<br>刑’, ’毒品’, ’被告’, ’贩运’, ’上诉’, ’克’, ’消费’, ’TMCC’, ’<br>要点’, ’起点’, ’管有’, ’折扣’, ’监<br>禁’] |
| LDA | [’上诉人’, ’月’, ’年’, ‘毒品’, ‘法官’, ’冰毒’, ‘刑罚’, ‘法院’, ‘被告’, ‘贩运’] |

表10：关键词分析与摘要：2 HKLRD 1121

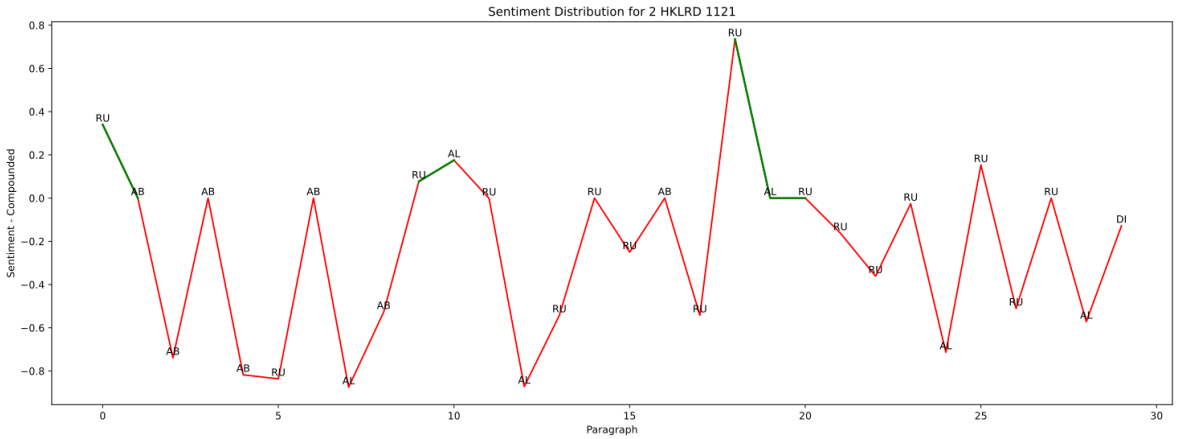



图12：带标签的情感分布：2 HKLRD 1121

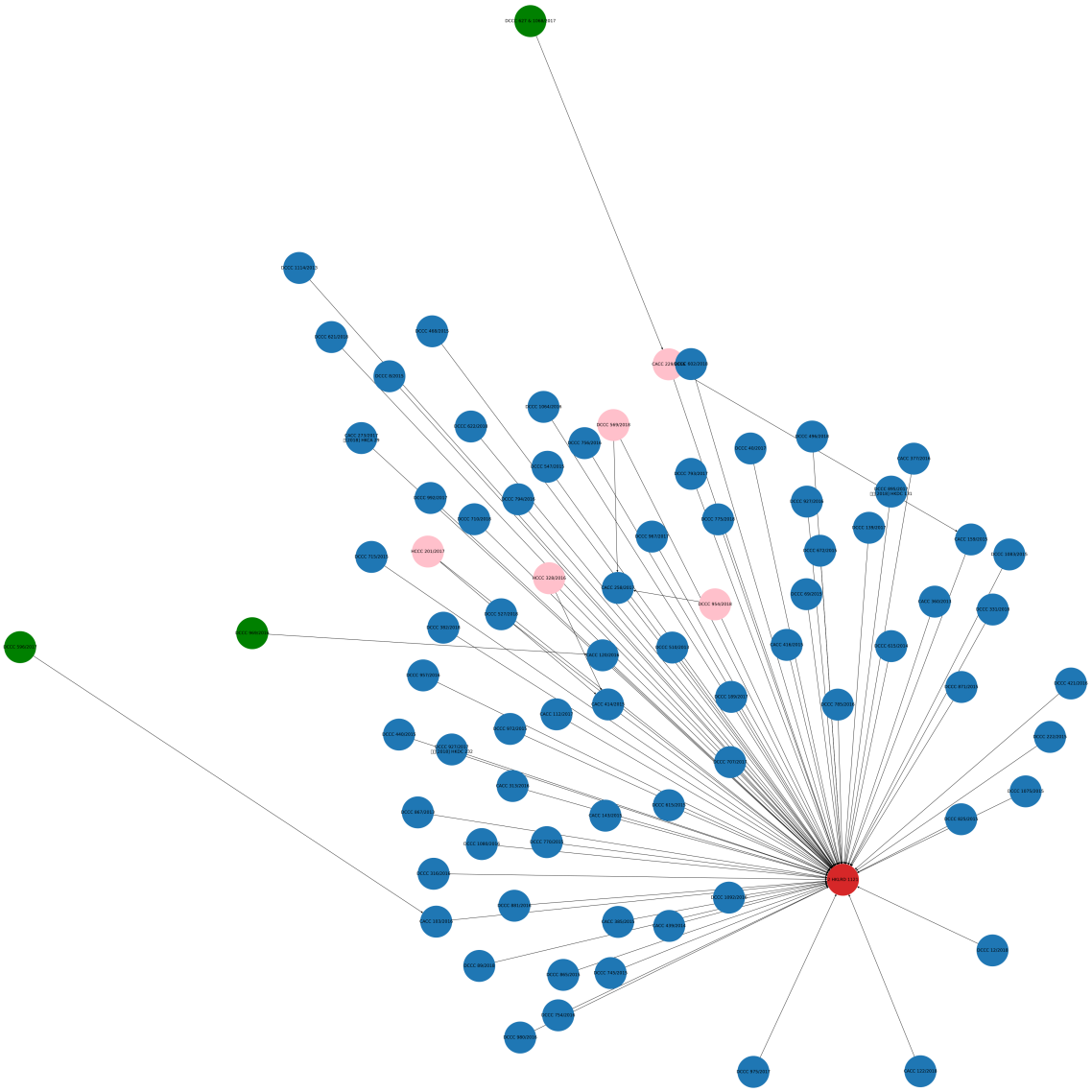
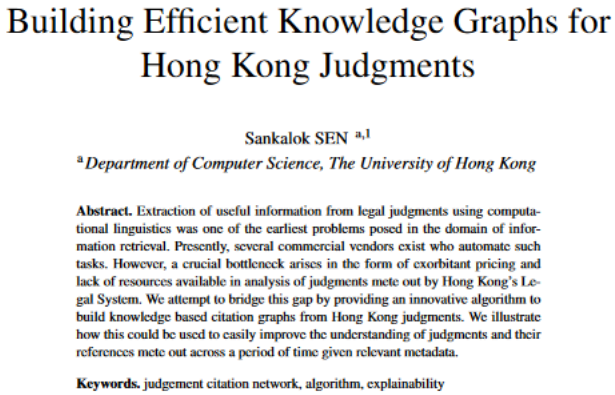


图 13：引文网络图：2 HKLRD 1121

# 7 结论与未来工作

本文从计算语言学的角度出发，针对香港法律体系作出的判决，对从法律判决中提取有用信息进行了广泛的实验与分析。我们采用了五折方法：（1）引文网络图生成；（2）PageRank算法；（3）关键词分析与摘要生成；（4）情感极性；以及（5）段落分类，以从单个判决以及一组判决中提取关键见解，从而自动提取有用见解，并缩短推理时间。我们还使用大语言模型对实验结果进行了基准测试。

未来工作包括对我们为本项目撰写的两篇补充论文进行微调，并计划投稿至即将举行的法律知识与信息系统国际会议（JURIX 2023）。我们计划在未来几周内将论文上传至 arXiv，并期待收到有关论文的反馈。



Building Efficient Knowledge Graphs for Hong Kong Judgments

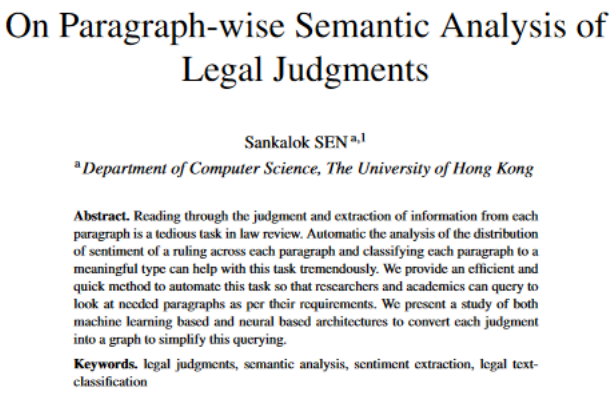
Sankalok SEN [a,1]

[a] *Department of Computer Science, The University of Hong Kong*

**Abstract.** Extraction of useful information from legal judgments using computational linguistics was one of the earliest problems posed in the domain of information retrieval. Presently, several commercial vendors exist who automate such tasks. However, a crucial bottleneck arises in the form of exorbitant pricing and lack of resources available in analysis of judgments mete out by Hong Kong's Legal System. We attempt to bridge this gap by providing an innovative algorithm to build knowledge based citation graphs from Hong Kong judgments. We illustrate how this could be used to easily improve the understanding of judgments and their references mete out across a period of time given relevant metadata.

**Keywords.** judgement citation network, algorithm, explainability

(a) 引文知识图谱



On Paragraph-wise Semantic Analysis of Legal Judgments

Sankalok SEN [a,1]

[a] *Department of Computer Science, The University of Hong Kong*

**Abstract.** Reading through the judgment and extraction of information from each paragraph is a tedious task in law review. Automatic the analysis of the distribution of sentiment of a ruling across each paragraph and classifying each paragraph to a meaningful type can help with this task tremendously. We provide an efficient and quick method to automate this task so that researchers and academics can query to look at needed paragraphs as per their requirements. We present a study of both machine learning based and neural based architectures to convert each judgment into a graph to simplify this querying.

**Keywords.** legal judgments, semantic analysis, sentiment extraction, legal text-classification

(b) 逐段语义分析

图 14：提交至 JURIX 2023 的论文摘要

# 8 致谢

我谨向我的项目导师孔令鹏博士致以诚挚的谢意，感谢他在本项目中给予我的持续支持，帮助我进一步探索如何弥合自然语言处理与社会科学领域之间的差距，从而深化我的研究兴趣。我还要感谢上海人工智能实验室的吴志勇博士，以及香港大学计算机科学系博士后研究员吴凯文博士，感谢他们提供的宝贵建议和指导，帮助我解决本项目旨在攻克的多个瓶颈问题。

# 9 参考文献

[1] S. H. C. Lo、K. K. Cheng 和 W. H. Chui，《香港法律制度》，剑桥大学出版社，2019】【。

[2]J. Chan 和 C. L. Lim，《从殖民地到特别行政区》，载于《香港宪法法》，第 2 版，Sweet Maxwell，2015 年。

[3] D. Locke 和 G. Zuccon，《判例检索：过去 30 年的成果、问题、方法与评估》，arXiv:2202.07209，2022 年。

[4] D. Bahdanau、K. Cho 和 Y. Bengio，《通过联合学习对齐与翻译实现神经机器翻译》，载于《国际学习表征会议》论文集（口头报告），2015 年。

[5] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、L. Kaiser 和 I. Polosukhin，《注意力机制就是你所需要的一切》，载于第 31 届神经信息处理系统会议论文集，2017 年。

[6]D. Nguyen、N. A. Smith、C. P. Rose。2011。《使用线性回归从文本预测作者年龄》。载于第五届 ACL-HLT文化遗产、社会科学与人文学科语言技术研讨会论文集，第115–123页。

[7]M. Sap、S. Swayamdipta、L. Vianna、X. Zhou、Y. Choi、N. A. Smith。2021。《带有态度的标注者：标注者信念与身份如何影响有害语言检测》。arXiv 预印本，arXiv:2111.07997。

[8] S. Gururangan、D. Card、S. K. Dreier、E. K. Gade、L. Z. Wang、Z. Wang、L. Zettlemoyer、N. A. Smith。2022。《什么样的语言才算高质量？衡量文本数据选择中的语言意识形态》。arXiv 预印本，arXiv:2201.10474。

[9]J. Gross、B. Acree、Y. Sim、N. A. Smith。2013年。检验 Etch-a-Sketch假设：通过候选人对关键词组的使用来衡量意识形态信号传递。发表于2013年芝加哥美国政治学协会年会。

[10] E. G. Altmann、J. B. Pierrehumbert、A. E. Motter。2011。生态位作为在线群组中词语命运的决定因素。PLOS ONE 6(5)。19

[11]A. Garimella, R. Mihalcea。2016。聚焦社交媒体中的性别差异。载于《社交媒体中人类观点、个性与情感计算建模研讨会论文集》，第1–10页。

[12]L. Page、S. Brin、R. Motwani 和 T. Winograd，《PageRank 引文排名：为万维网带来秩序》，发表于 WWW，1999。

[13] R. Mihalcea 和 P. Tarau。2004年。《TextRank：为文本带来秩序》。载于 2004年自然语言处理实证方法会议论文集，第 404–411 页，西班牙巴塞罗那。计算语言学协会。

[14] R. Campos、V. Mangaravite、A. Pasquali、A. Jatowt、A. Jorge、C. Nunes 和 A. Jatowt。2020。YAKE！使用多种局部特征从单篇文档中进行关键词提取。发表于《信息科学》期刊。Elsevier，第 509 卷，第 257–289 页。

[15]S. Rose、D. Engel、N. Cramer 和 W. Cowley。2010。《从单个文档中自动关键词提取》。载于《文本挖掘：应用与理论》（第 1–20 页）

[16] C.J. Hutto 和 Eric Gilbert。2014。VADER：一种用于社交媒体文本情感分析的简约规则模型。载于第八届国际网络日志与社交媒体会议（ICWSM）论文集。

[17]S. Eshima、K. Imai、T. Sasaki。2020年。关键词辅助主题模型。https://arxiv.org/abs/2004.05964。

[18]D. M. Blei、A. Y. Ng 和 M. I. Jordan。2003 年。潜在狄利克雷分配。载于《机器学习研究杂志》。

[19] J. Gamper。2000。意大利语/德语法律文本的平行语料库。载于 LREC 论文集。

[20]C. Grover、B. Hachey 和 I. Hughson。2004年。《HOLJ语料库：支持法律文本摘要》。载于 COLING 会议论文集。

[21]B. Fawei、A. Wyner 和 J. Pan。2016。通过美国全国律师资格考试：用于实验的首个语料库。载于 LREC 会议论文集。

[22]Y. Kano、M. Kim、M. Yoshioka、Y. Lu、J. Rabelo、N. Kiyota、R. Goebel 和 K. Satoh。2018 年。Coliee-2018：法律信息提取与蕴含竞赛评估。载于 JSAI 会议论文集。

[23] P. H. L. de Araujo、T. E. de Campos、R. R. R. de Oliveira、M. Stauffer、S. Couto 和 P. Bermejo。2018。Lener-br：一个用于巴西法律文本中命名实体识别的数据集。发表于 PROPOR 会议论文集。

[24]C. Xiao、H. Zhong、Z. Guo、C. Tu、Z. Liu、M. Sun、Y. Feng、X. Han、Z. Hu、H. Wang 和 J. Xu。2018。CAIL2018：用于判决预测的大规模法律数据集。

[25] L. Manor 和 J. J. Li。2019。合同的通俗英语摘要生成。发表于自然法律语言处理研讨会论文集。

[26]I. Chalkidis、I. Androutsopoulos 和 N. Aletras。2019。《英语神经网络法律判决预测》。载于 ACL 会议论文集。

[27]I. Chalkidis、M. Fergadiotis、P. Malakasiotis 和 I. Androutsopoulos。2019。《欧盟立法的大规模多标签文本分类》。载于 ACL 会议论文集。

[28]X. Duan、B. Wang、Z. Wang、W. Ma、Y. Cui、D. Wu、S. Wang、T. Liu、T. Huo 和 Z. Hu。2019。《CJRC：用于中文司法阅读理解的可靠人工标注基准数据集》。载于 CCL 会议论文集。

[29] B. Luo、Y. Feng、J. Xu、X. Zhang 和 D. Zhao。2017。《结合法律依据学习预测刑事案件的罪名》。载于 EMNLP 会议论文集。

[30]Y. Shen、J. Sun、X. Li、L. Zhang、Y. Li 和 X. Shen。2018。用于罪名预测的法律条文感知端到端记忆网络。发表于 CSAE 会议论文集。

[31]S. Long、C. Tu、Z. Liu 和 M. Sun。2018。通过法律阅读理解实现自动判决预测。https://arxiv.org/abs/1809.06537。

[32]J. Li、G. Zhang、H. Yan、L. Yu 和 T. Meng。2018。基于马尔可夫逻辑网络的离婚案件司法判决预测方法。发表于 IEEE SmartCloud 会议论文集。

[33]H. Ye、X. Jiang、Z. Luo 和 W. Chao。2018。刑事案件中可解释的罪名预测：学习根据事实描述生成法院观点。发表于 NAACL-HLT 会议论文集。

[34] Y. Wu、K. Kuang、Y. Zhang、X. Liu、C. Sun、J. Xiao1、Y. Zhuang、L. Si 和 F. Wu。2020。利用因果关系生成去偏见的法院观点。发表于 EMNLP 会议论文集。

[35]C. Cardellino、M. Teruel、L. A. Alemany 和 S. Villata。2017。结合本体、维基百科与课程学习的法律命名实体识别。载于 EACL 会议论文集。

[36]A. Elnaggar、R. Otto 和 F. Matthes。2018。《利用迁移学习对法律文书进行命名实体链接的深度学习》。载于 AICCC 会议论文集。

[37]E. Leitner、G. Rehm 和 J. Moreno-Schneider。2019。法律文书中的细粒度命名实体识别。载于 SEMANTiCS 会议论文集。

[38] N. Lagos、F. Segond、S. Castellani 和 J. O'Neill。2010。用于法律案件构建和推理的事件抽取。载于 IIP 会议论文集。

[39] M. Truyens 和 P. V. Eecke。2014。文本挖掘的法律方面。载于 LREC 会议论文集。

[40]K. Raghav、P. K. Reddy 和 V. B. Reddy。2016 年。利用段落级和引文信息分析相关法律判决的提取。在 ECAI 会议论文集。

[41]V. Tran、M. L. Nguyen 和 K. Satoh。2020。使用预训练短语评分模型构建基于词汇匹配与摘要生成的法律案例检索系统。https://arxiv.org/abs/2009.14083。

[42]C. Grover、B. Hachey、L. Hugson、C. Korycinski。2003。法律文书的自动摘要。载于 ICAIL 会议论文集。

[43] R. Kumar V 和 K. Raghuveer。2012。使用潜在狄利克雷分配进行法律文档摘要生成。载于 IJCST 会议论文集。

[44]R. S. Wagh 和 D. Anand。2020。通过利用领域知识增强法律文本分段训练数据的新方法。载于《2020年技术与应用会议论文集》。

[45]香港特别行政区，香港法律信息研究所（HKLII），https://www.hklii.hk/eng/。

[46]香港特别行政区法律参考系统，https://legalref.judiciary.hk/lrs/common/help/hlptopic.htm。

[47]C.-Y. Lin。2004年。《ROUGE：自动摘要评估工具包》。载于《文本摘要的拓展》，计算语言学协会。

[48]HuggingFace，https://huggingface.co/facebook/bart-large。