



面向法律文本的 自然语言分析与理解

冯岩松
北京大学

fengyansong@pku.edu.cn

SMP 2018 , 哈尔滨



北京大学



法律与人工智能

- 1980年代前
 - 逻辑推理与法律文本分析的结合[Layman Allen, 1957] [Buchanan and Headrick , 1970]
 - TAXMAN：将规则推理应用于公司税务文本分析 [L. Thorne McCarty, 1977]
 - 基于法律概念及关联的法律文本检索[Carole Hafner , 1978] [挪威计算机与法律研究中心，1987]

- 1980-90年代

80年代 AI+Law 蓬勃发展
但却迎来了专家系统的落寞

- ICAIL (自1987), IAAIL (自1991), JAIL (自1992), Jurix (自1988)



北京大学



早期工作

- 专家系统

- 支持严密逻辑推理，但基于大量专家手写规则
- 适用于限定领域，难以大规模扩展
- 规则集的覆盖与冲突
- 模糊信息处理

- 一些经验

- 机器并不能代替法律实务工作者
- 应成为信息化中的重要组成 (两高、律所.....)
- 以法律文书为主要研究对象



北京大学



围绕法律文书

- 从实用性出发
 - 服务司法实践的信息化
 - 面向法条、案情的分析与辅助决策
- 《纲要》：“**智能化**”
 - 智慧检务、智慧法院……
 - 以业务为核心，以文书为主线
- 引起工业界广泛关注
 - 从业务信息化，向“**智能化**”进发
 - 知名传统企业、明星AI企业、初创AI企业……
- 学术界新宠 —— **巨大挑战**
 - 应用场景丰富
 - 任务设定多样

以法律文书为主线

诉讼文书、非诉文书
裁判文书、仲裁文书
侦察文书、检察文书
公证文书、律师实务文书

……



北京大学



典型任务

- 信息获取与检索
 - 关键词抽取
 - 合同、专利、法条检索
 - 类案推送、方案推送
 - 智能问答
- 智能辅助决策
 - 辅助判案、辅助量刑、辅助民事调解
 - 文书生成
- 基础支持
 - 图像、语音、文字识别
 - 卷宗管理



北京大学



基于文本分析的法律智能

核心应用

公安

检察

法院

律师

用户

核心技术

法律知识
体系
构建

信息
抽取

语义
分析

文本
分类

文本
生成

信息
检索

文本、法
律推理

篇章、论
辩分析

.....

法律文本资源

法条、法
规

诉讼、非
诉文书

裁判、仲
裁文书

侦察相关
文书

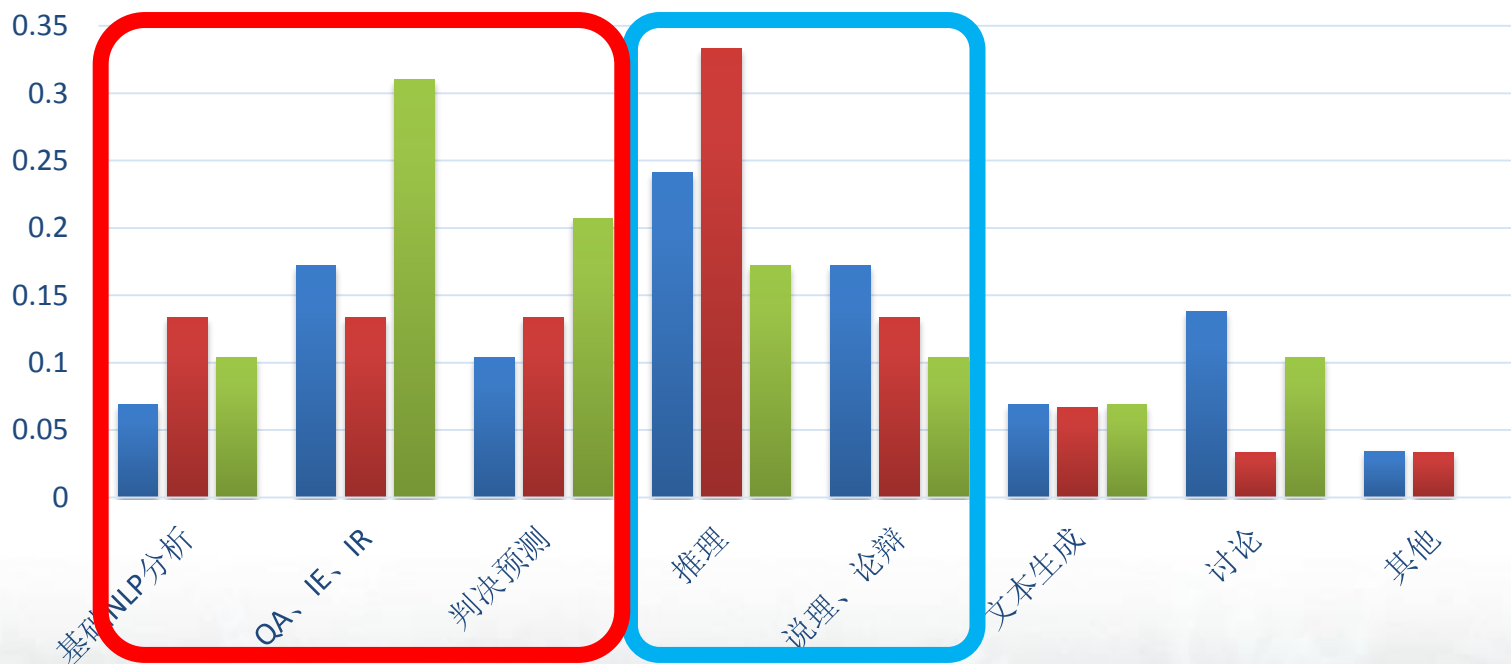
律师实务
文书

.....

相关进展

- ICAIL: International Conference on Artificial Intelligence
– 自1987年开始创办，每2年举办一次

ICAIL 文章领域分布



以法律文书的分析为主



北京大学



发展现状

- 一流院校开设的相关项目
 - Stanford CodeX
 - 北大法律人工智能研究中心
 - 清华法学院计算法学专业
- 初创企业
 - LawGeex, Premonition, Ross Intelligence, ...
 - 前9名创业公司融资约 \$29 million (2017)
- 技术层面
 - 民商领域为主
 - 帮助普通人理解专业用语、熟悉流程
 - 为法律或商务专业人士节约时间



北京大学

技术现状





仍存在巨大挑战

- 许多任务性能仍有待提高
 - 如何在NLP模型中应用法律知识
 - 基础标注资源
- 如何解释模型输出的结果？
 - 可解释性
 - A：盗窃罪，刑期5年；B：抢劫罪，刑期8年
 - 说理（论辩结构）
- 如何有理有据？
 - 专业术语与日常用语
 - 争议焦点与说理

法律领域自身的复杂性
现有AI、NLP技术的局限性

利用知识进行深度解析

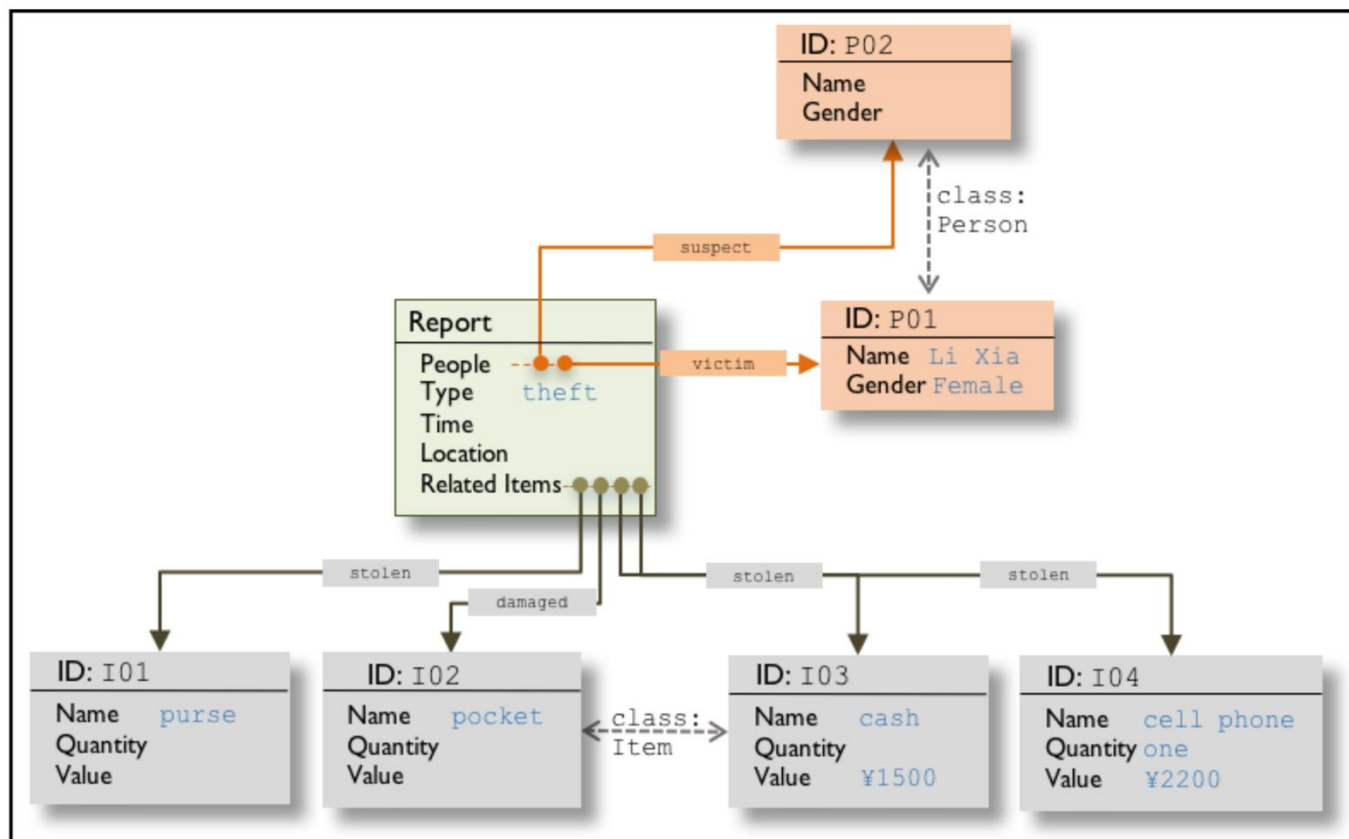
接李霞（女）报称其在梅西百货逛街时钱夹被盗。犯罪分子趁事主不备割开上衣口袋将钱夹偷走。被盗现金：1500元，手机一部价值2200元。

HED

Report from Li Xia (female) that her purse was stolen while shopping in Macy's. The thief cut a hole in her coat pocket and stole the purse while the victim was off guard. The lost items include ¥1500 cash, and one cell phone worth ¥2200.

Original Chinese:

接李霞(女)报称其在梅西百货逛街时钱夹被盗。犯罪份子趁事主不备割开上衣口袋将钱夹偷走。被盗现金：1500元，手机一部价值2200元。



将法律领域知识转化为语义分析的结构目标，边阅读边解析[ACL2018]

解释预测结果



or

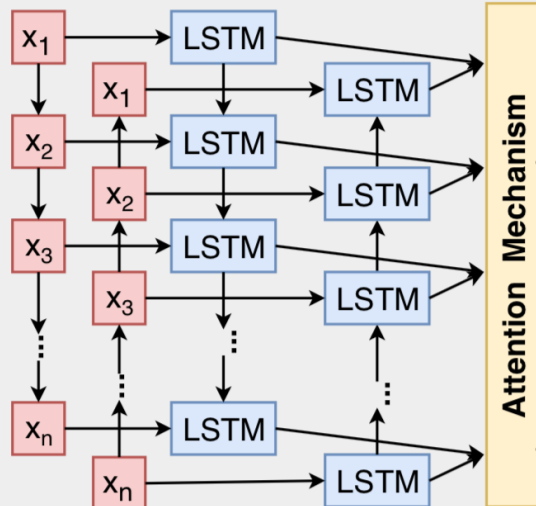


charge decision

Fact Description

... After hearing, our court identified that at 23:00 on July 10,2009, the defendant Chen together with other eight or nine young men stopped Lee who was riding a motorcycle on street near the road in Xinliao townXuwen County, after that the defendant Chen and the others beat Lee with steel pipe and knife. According to forensic identification, Lee suffered minorwound. ...

X



Attention Mechanism

<S>

v

charge

LSTM

LSTM

LSTM

LSTM

LSTM

LSTM

y1

y2

y3

y4

ym

Y

Court View

Our court hold that the defendant Chen ignored the state law and caused others minor wound with equipment together with others

本院认为，被告人吴春丽伙同他人以营利为目的，采用欺骗手段，拐卖妇女一人，本案被告人吴春丽作案时间为1989年，根据从旧兼从轻原则，其行为已触犯1979年《中华人民共和国刑法》之规定，构成拐卖人口罪。....。

解释预测结果-2

- 如何找出辅助决策的支持理由？
 - 没有标注数据？
 - 利用强化学习方法

CASE 1 [Official Embezzlement]_{charge}

... PP 利用其担任公司业务员的职务便利，从公司仓库提走多部手机，后将手机卖掉，货款挥霍...

... Using his position as a company salesman, PP took several phones from the company's warehouse, sold the phones, and squandered the money...

CASE 2 [Larceny]_{charge}

... PP₁ 趁 PP₂ 家中无人之机，进入到 PP₂ 家卧室内伺机盗窃。被 PP₂ 回家后发现，PP₁ 翻墙逃跑...

... When PP₂ was not at home, PP₁ went to PP₂'s bedroom to steal. When PP₂ came home, PP₁ fled the wall and ran...

CASE 3 [Negligently Causing Fire]_{charge}

... 在焚烧耕地上的杂草时，不慎引发山林火灾。案发后，PP 积极救火，主动向上级说明失火情况...

... When burning weeds on cultivated land, PP inadvertently ignited the mountain fire. Then, PP actively doused the fire and reported the fire situation...

CASE 4 [Arson]_{charge}

... PP₁ 因生意竞争与 PP₂ 产生积怨。PP₁ 酒后萌生放火烧 PP₂ 将机店的念头，进入 PP₂ 的店内将纸箱点燃...

... PP₁ hates PP₂ for his business competition. After drinking, PP₁ produced the idea of burning PP₂'s shop. PP₁ entered the shop and lights the carton...

CASE 5 [Negligent Homicide]_{charge}

... 在狩猎过程中，PP 因地滑摔跤，导致其所持鸟铳击发走火，将走在前面的 PP 打伤致死...

... In the process of hunting, PP fell down due to the slippery ground, leading to the shotgun fire, killing PP who was walking in front...

CASE 6 [Intentional Homicide]_{charge}

... PP₁ 从家中携带匕首出门寻找 PP₂ 进行报复，将 PP₂ 捅倒后，在颈部来回割，致 PP₂ 当场死亡...

... PP₁ took the dagger and looked for PP₂ for revenge. He stabbed PP₂ and then cut him back and forth on the neck, causing PP₂ to die on the spot...

标注数据有限

- 数据稀疏、标注数据有限怎么办？
 - 如何精简专家的投入？

| | 盈利目的 | 买卖过程 | 犯罪致死 | 暴力行为 | 国家机构 | 公共场合 | 非法占有 | 物理伤害 | 故意犯罪 | 生产过程 |
|--------------------|------|------|------|------|------|------|------|------|------|------|
| 过失损坏武器装备、军事设施、军事通信 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 污染环境 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 遗弃 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 非法处置查封、扣押、冻结的财产 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 故意毁坏财物 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 窝藏、转移、隐瞒毒品、毒赃 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 容留他人吸毒 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 拒不执行判决、裁定 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 扰乱无线电通讯管理秩序 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 重婚 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 伪证 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 传播性病 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 引诱、教唆、欺骗他人吸毒 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 诬告陷害 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 帮助毁灭、伪造证据 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 窝藏、包庇 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 妨害作证 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 破坏监管秩序 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 脱逃 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 |

05101520253035404550

Charge Label Id



从技术角度

- 篇章理解
 - 如何理解长文本
- 多源知识
 - 如何利用知识
- 可解释性算法
 - 如何解释模型输出
- 多模态数据整合
 - 图像、视频与文本的关联与融合
- 标注数据匮乏



北京大学



法律文本的深层理解

- 利用知识
 - 深层语义分析与世界知识的结合
 - 与法律条文的结合与判断
- 理解长文本、复杂文本
 - 篇章结构分析
 - 论辩结构与说理
- 细粒度标注数据
 - 需要专业背景
 - 强化学习、深度学习



北京大学



请各位老师指正
谢谢！



北京大学