

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



案件文本分析

硕士研究生 董勃

2020年03月29日

- 背景介绍
- 基本概念
 - 案件文本介绍
 - 案件文本任务
- 算法原理
 - 算法框架
 - 技术技巧
- 应用总结
- 参考文献

- 预期收获
 - 1. 熟悉案件文本处理任务类型
 - 2. 了解法研杯（CAIL）竞赛
 - 3. 了解案件文本各任务的常用框架
 - 4. 了解任务的应用

- 概念与目的
 - 人工智能在法律中的应用，其目标是充分提升法治效能，将法律工作者从繁杂的工作中解放出来。
- 落地应用
 - 智慧法院、智慧检务建设。
- 竞赛
 - **CAIL2018**——中国法研杯司法人工智能挑战赛
 - **CAIL2019**
 - COLIEE——加拿大联邦法院/日本法律律师考试

- CAIL2018
 - **罪名预测**
 - 法律条文预测
 - 刑期预测

- CAIL2019
 - 阅读理解
 - 要素识别
 - **相似案例匹配**

• 裁决文书要素

a. 原被告信息

被告人黄利民，男，汉族，1974年12月20日出生于湖南省隆回县，初中文化，农民，户籍地隆回县××镇××村×组××号，租住地湖南省长沙市雨花区尊邸华庭×栋×单元×××房。2014年6月19日被逮捕。现在押。

被告人杨杰彪，男，汉族，1985年3月10日出生于广东省惠来县，小学文化，无业。

b. 案件事实描述

一、关于被告人黄利民贩卖、制造毒品事实

2012年9月，被告人黄利民得知同案被告人尹有缓（已判刑）会制造甲基苯丙胺片剂（俗称“麻古”），遂向尹有缓提议共同制造、贩卖甲基苯丙胺片剂牟利，尹有缓表示同意，黄利民又纠集侄子黄某某（同案被告人，犯罪时不满十八周岁，已判刑）参与。黄利民负责筹集资金，购买甲基苯丙胺（冰毒）、咖啡因等主要制毒原料和制毒工具并销售毒品，尹有缓负责购买制毒其他部分原料和制造毒品，黄某某为二人提供协助，三人共同在湖南省隆回县××镇黄利民家、隆

c. 依据法条

判处死刑，可不立即执行。依照《中华人民共和国刑法》第三百四十七条第一款、第二款第（一）项、第七款、第二十五条第一款、第二十六条第一款、第四款、第四十八条第一款、第五十七条第一款、第五十九条、第六十七条第三款和《中华人民共和国刑事诉讼法》第二百四十六条、第二百五十条、《最高人民法院关于适用〈中华人民共和国刑事诉讼法〉的解释》第三百五十二条的规定，判决

d. 判决罪名

e. 刑期

一、核准湖南省高级人民法院（2016）湘刑终440号刑事裁定中维持第一审以贩卖、制造毒品罪判处被告人黄利民死刑，剥夺政治权利终身，并处没收个人全部财产的部分。

T	基于原始案件文书，完成罪名、刑期、法条预测，QA
I	CAIL比赛数据集、中国裁判文书网文书
P	For { 1. 文书预处理 2. 预训练模型选择 3. fine-tune }
O	任务匹配的分类模型

P	构建匹配任务的数据集，找到恰当的模式
C	自行构建数据集，需自行构造结构化数据
D	数据不平衡，文书长度过长，多标签任务
L	水平不一

- 论文名称
 - Learning to Predict Charges for Criminal Cases with Legal Basis
 - 对刑事案件进行裁决预测
- 模型
 - Attention-based 神经网络模型
- 目标
 - 相关法条提取
 - 罪名判断

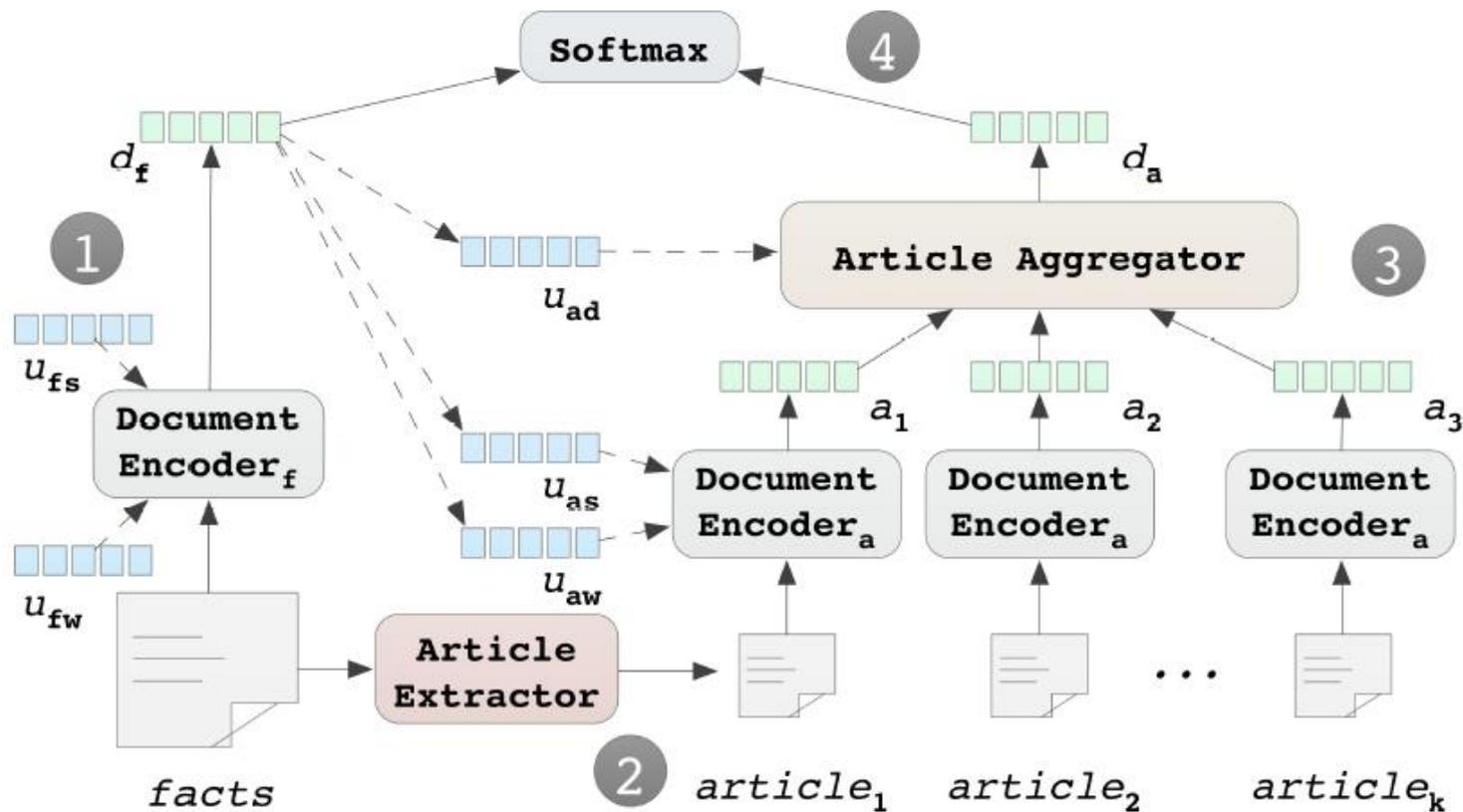
- 难点
 - 两个不同的文书差异细微
 - 一个案件可能涉及多个罪名
 - 需提供法条才更有说服力
- 任务说明

... **经审理查明**, 2011年10月6日凌晨, 被告人AA携带改锥、扳手、破坏钳、刀等物品到尉氏县张市镇尹庄村BB家门口盗窃农用车上的电瓶时被被害人BB发现, 在逃跑过程中AA为抗拒抓捕持刀将BB致伤。 ...

本院认为, 被告人AA在盗窃过程中携带凶器, 为抗拒抓捕而当场使用暴力致被害人BB轻微伤, 其行为已构成抢劫罪, ... 依照《**中华人民共和国刑法**》第二百六十三条、第二百六十九条、... 之规定, **判决如下**:

被告人AA犯**抢劫罪**, 判处有期徒刑三年, 并处罚金人民币一千元。 ...

- 模型架构



- 论文名称
 - Few-Shot Charge Prediction with Discriminative Legal Attributes
 - 根据刑事案件事实描述预测罪名

- Few-shot charges
 - 各类型案件数量极度不平衡
- Confusing charges
 - 存在许多易混淆的案件
- 解决方法
 - 引入法律属性 (legal attributes)
 - 提出多任务学习的框架, 以同时预测每个案例的属性和判决结果

- 十个区分属性

Attributes
Profit Purpose
Buying and Selling
Death
Violence
State Organ
Public Place
Illegal Possession
Physical Injury
Intentional Crime
Production

对于每个(charge, attribute)对，都会标记Yes、No、NA。过失杀人(manslaughter)罪名在故意犯罪(Intentional Crime)属性上标记为No，在死亡(Death)上为Yes，在国家机关(State Organ)上为NA。

故意伤害罪

滋事罪（打架斗殴）

暴力（Violence）✓

身体伤害（Physical Injury）✗

- 论文名称
 - CAIL2019-SCM: A Dataset of Similar Case Matching in Legal Domain
- 任务介绍：
 - 每份数据由(d,d1,d2)三篇文书组成，找到与d更相关的一篇文书。
- 目标
 - 提取相似案例
- 模型
 - Bert-base-chinese、Bi-GRU、Bi-LSTM融合模型。

- 文本统计类特征
 - 文本长度特征（文书分部分提取）
 - 词频统计（tf-idf）
- 业务理解抽象特征
 - 原告被告特征
 - 担保特征（一般，连带）
 - 利息特征（民事借贷相关）
 - 其余有业务含义特征

- 数据增广
 - 原因：现有数据集较小，需要更多数据集
 - 方式：原始数据 (A,B,C) , 0

序号	描述	增广产物
1	反对称增广	(A,C,B) , 1
2	自反性增广	(C,C,A) , 0
3	自反性+反对称	(C,A,C) , 1
4	启发式增广	(B,A,C) , 0
5	启发式增广+反对称	(B,C,A) , 1

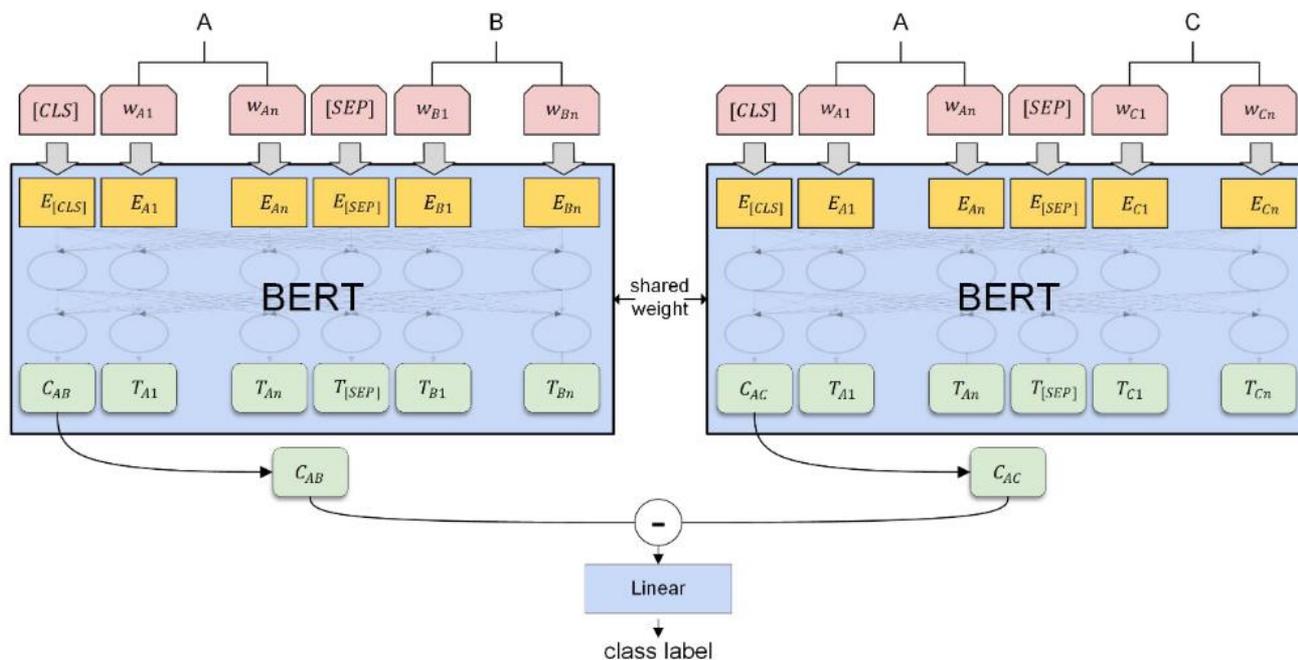
- 损失函数
 - 三元组 (A,B,C)
 - 二分类绝对相似模型
 - 三元组相对相似问题

$$L = \max((d(a, p) - d(a, n)) + \text{margin}, 0)$$

$$d(a, p) = \sqrt{\sum (a - p)^2}$$

$$d(a, n) = \sqrt{\sum (a - n)^2}$$

- 孪生网络



- 使用孪生网络结构，两个BERT共享权重，分别将AB、AC输入BERT，最后拼接一个线性层输出后进行二分类。

- 算法的应用领域
 - 文本分类问题
 - 问答-结构化数据获取, 辅助判案

- 未来的发展
 - Q&A领域
 - 案件文本结构化处理

- [1]Xiao, C., et al.CAIL2019-SCM: A Dataset of Similar Case Matching in Legal Domain. arXiv preprint arXiv:1911.08962, 2019.
- [2]Luo, B., et al. Learning to predict charges for criminal cases with legal basis. arXiv preprint arXiv:1707.09168, 2017.
- [3]Hu, Z., et al. Few-shot charge prediction with discriminative legal attributes. 2018. p. 487--498.



知人者智，自知者明。
胜人者有力，自胜者
强。知足者富。强行
者有志。不失其所者
久。死而不亡者，寿。

谢谢!

