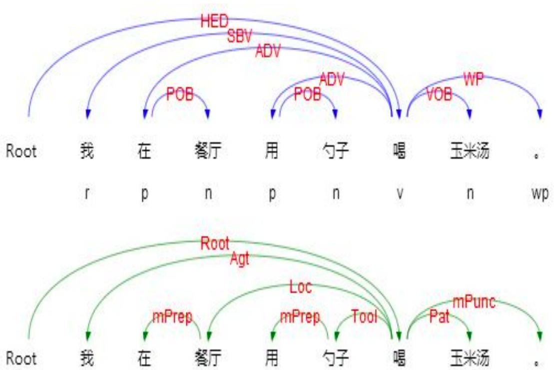
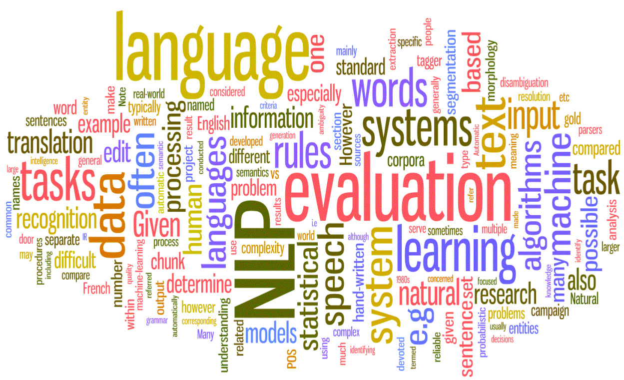
面向法律智能的自然语言处理

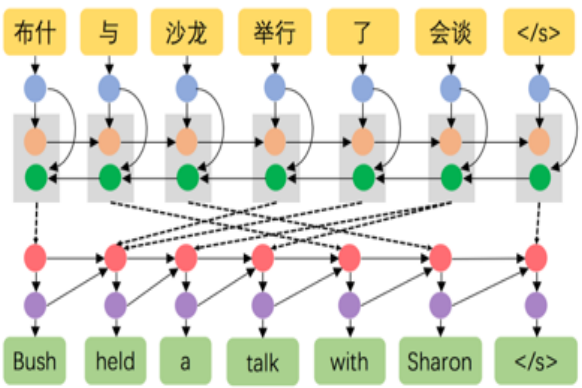
清华大学自然语言处理实验室

刘知远 涂存超

自然语言处理（NLP）任务







分词、词性标注、命名实体识别、指代消解、句法分析、语法分析…

关键词抽取、文本分类、情感分析与观点挖掘、信息抽取…

机器翻译、自动摘要、问答系统、对话系统…

NLP能为法律做什么？

- 自然语言是法律的载体
  - 高质量且形式丰富的文本数据
  - 基于法律文本的多样需求





案例文书	法律法规	司法解释	工商信息	法学文献
近5000万	1000+		千万商标、专利	数百种法学刊物

NLP在法律领域的应用前景

智能案例检索

判决预测

文书自动生成

法律智能推荐

法律文本翻译

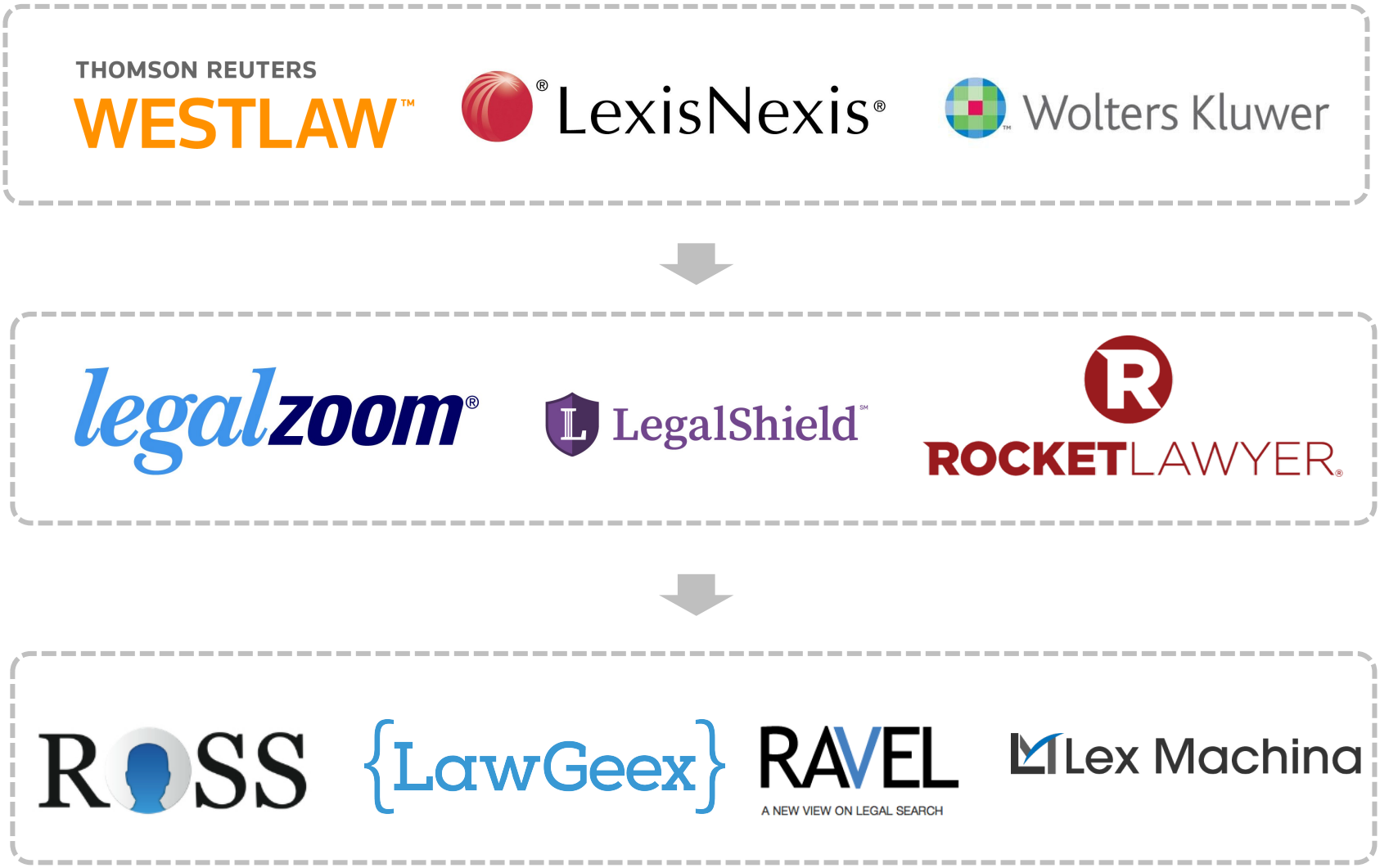
法律智能问答

风险提示

法律文本挖掘

合规审查

NLP在法律领域的应用前景



NLP在法律领域的应用挑战

多模态法律文本数据

针对法律文本的信息抽取、知识图谱构建

基于语义理解的法律检索

法律文本深度分析

多模态异构信息融合

智能推理决策

可解释性

面向案例文书的判决预测研究

- 刑事案件、民事案件
- 法条、罪名、刑期、赔偿金额等

面向案例文书的判决预测

研究问题 根据案件的案情描述，预测最终的判决结果

被告人黎某在本市白云区钟落潭镇五龙岗罗贤六巷16号401房梁某的租住处，因琐事与梁发生争吵，遂用枕头捂住梁的口、鼻部，用保温杯砸打梁的头部，将梁打晕，并持菜刀砍梁的颈、面部等处数刀，致梁受伤倒地，后拿了梁的三星牌3588型无线移动电话机1台等物品逃离现场。经鉴定，梁全身的损伤符合外力作用所致，其损伤致头皮疤痕累计长度大于20cm，面部单个疤痕长度大于6cm，颈部疤痕累计长度大于16cm，损伤属轻伤一级；上述无线移动电话机价值人民币493元。

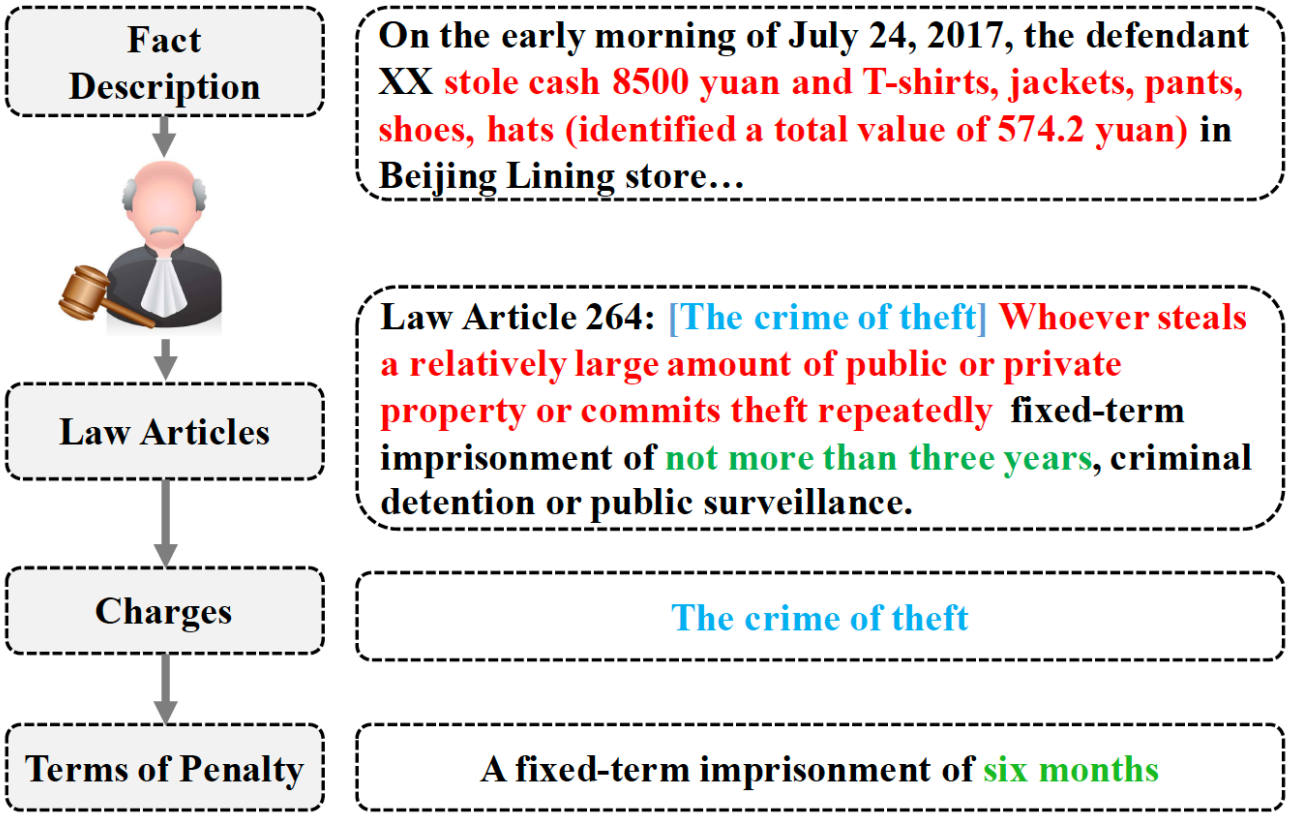
相关法条预测：
中华人民共和国刑法 第23条
中华人民共和国刑法 第55条
中华人民共和国刑法 第56条
中华人民共和国刑法 第232条

罪名预测：
故意杀人罪

量刑预测：
有期徒刑十年
剥夺政治权利三年

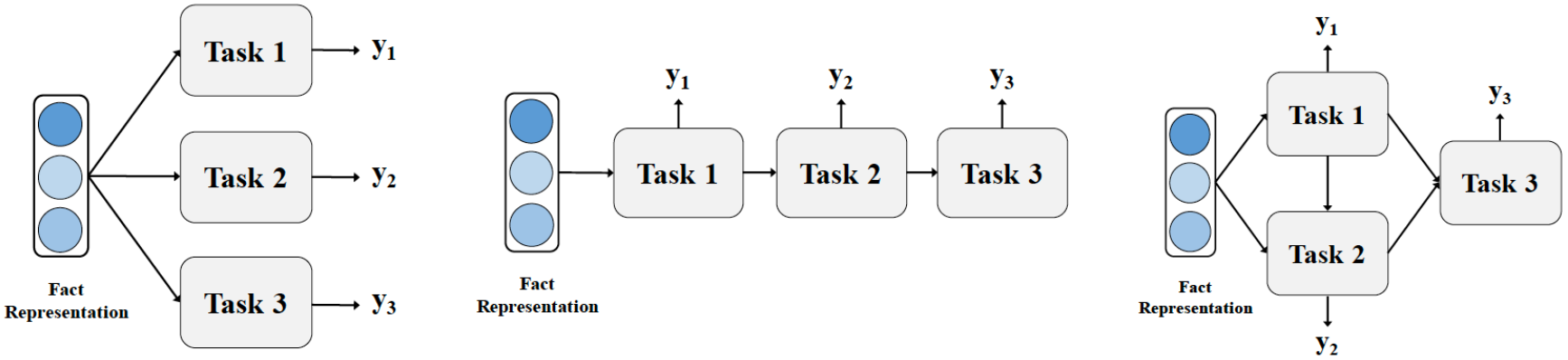
基于拓扑结构预测的判决预测

- 法官的判案逻辑：子任务之间的依赖关系



基于拓扑结构预测的判决预测

- 子任务之间依赖关系的拓扑结构



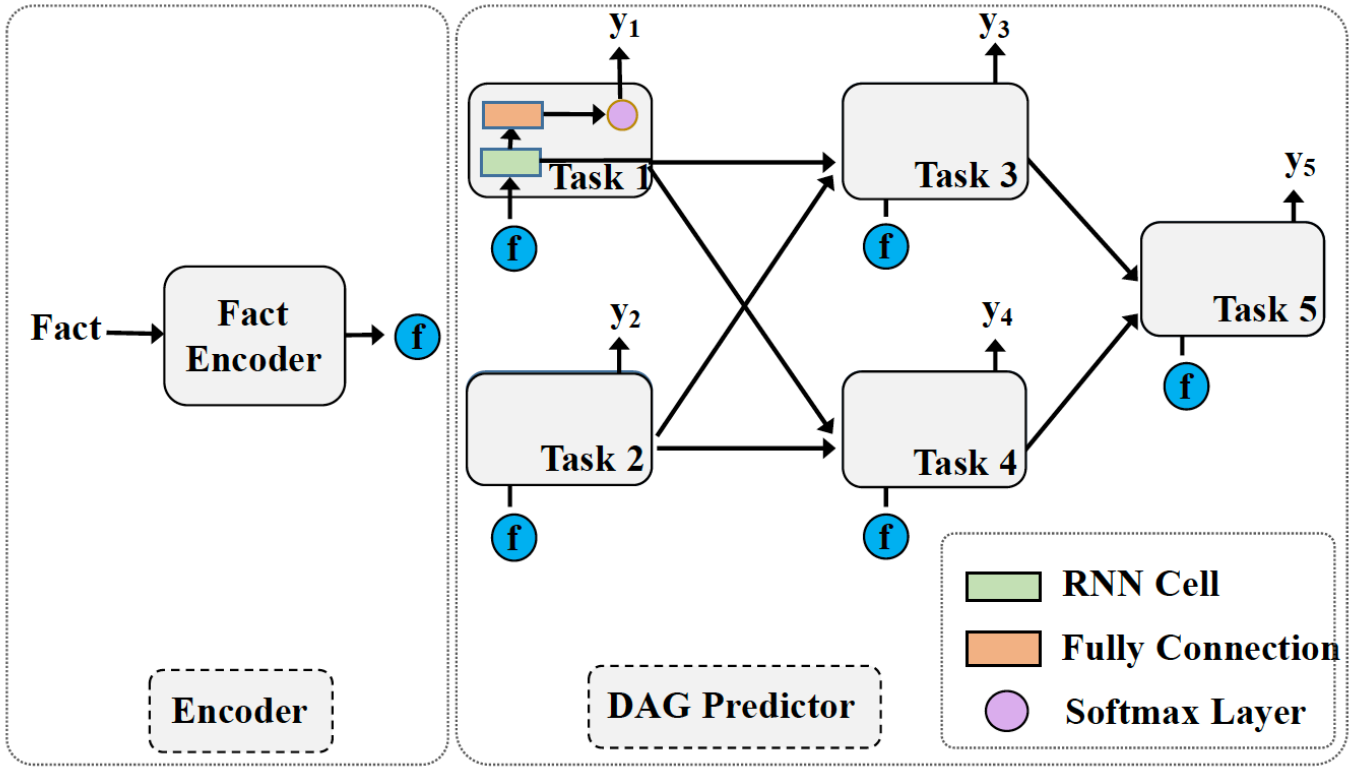
(a) Typical MTL form of DAG (b) Sequential form of DAG (c) General form of DAG

有向无环图（DAG）

Haoxi Zhong, Guo Zhipeng, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, Maosong Sun. **Legal Judgment Prediction via Topological Learning**. EMNLP 2018.

基于拓扑结构预测的判决预测

- TopJudge: 基于拓扑预测的多任务判决预测



Haoxi Zhong, Guo Zhipeng, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, Maosong Sun. **Legal Judgment Prediction via Topological Learning**. EMNLP 2018.

基于拓扑结构预测的判决预测

- 实验数据集

Datasets	CJO	PKU	CAIL
Cases	1,007,744	175,744	113,536
Law Articles	98	68	105
Charges	99	64	122
Term of Penalty	11	11	11

- 基准方法
  - TFIDF+SVM
  - CNN
  - H-LSTM
  - Fact-Law Attention Model
  - Pipeline Model
  - Multitask Learning

基于拓扑结构预测的判决预测

	Tasks	Law Articles				Charges				The Term of Penalty			
	Metrics	Acc.	MP	MR	F_1	Acc.	MP	MR	F_1	Acc.	MP	MR	F_1
Single	TFIDF+SVM	82.4	45.5	26.7	30.2	82.2	47.4	27.9	31.3	48.5	36.0	16.7	16.5
	CNN	92.5	46.9	38.4	40.0	92.3	41.2	32.3	33.7	57.4	35.6	22.2	22.7
	HLSTM	91.4	38.6	37.3	36.9	91.8	37.8	36.0	35.2	56.1	22.5	25.0	23.3
Multi	Fact-Law Att.	93.5	50.9	45.6	45.9	93.4	47.2	41.4	41.5	56.3	31.3	26.4	26.7
	PM	93.7	51.9	44.1	44.9	93.6	45.5	39.1	39.3	58.2	38.2	24.9	26.8
	CNN-MTL	94.3	53.0	46.0	46.9	94.1	48.5	41.7	42.5	58.7	39.9	28.8	29.4
	HLSTM-MTL	92.4	45.5	41.4	41.0	92.3	41.9	36.6	35.9	54.9	30.6	26.6	26.4
Ours	**TOPJUDGE**	**94.4**	**53.9**	**47.3**	**48.2**	**94.9**	**53.9**	**48.2**	**49.1**	**58.8**	**40.2**	**32.9**	**32.8**

Table 2: Judgment prediction results on CJO.

	Tasks	Law Articles				Charges				The Term of Penalty			
	Metrics	Acc.	MP	MR	F_1	Acc.	MP	MR	F_1	Acc.	MP	MR	F_1
Single	TFIDF+SVM	80.9	51.3	32.6	36.4	81.0	53.4	35.4	38.7	45.3	30.4	17.4	17.2
	CNN	93.1	64.3	52.6	54.3	93.3	61.9	49.3	51.1	57.6	24.1	23.1	23.3
	HLSTM	91.7	54.4	53.4	50.9	91.9	52.5	48.9	47.3	54.3	20.6	21.7	19.0
Multi	Fact-Law Att.	93.9	68.1	63.4	63.5	94.2	65.8	58.5	58.7	55.7	27.7	27.4	26.5
	PM	94.4	69.6	61.0	62.2	94.3	65.1	56.2	57.2	58.2	36.2	26.4	27.1
	CNN-MTL	95.0	73.8	64.9	66.0	95.0	70.7	60.6	61.7	**58.4**	36.0	28.7	28.9
	HLSTM-MTL	93.9	71.2	64.6	65.1	93.8	67.8	60.0	60.7	55.4	31.3	26.2	25.7
Ours	**TOPJUDGE**	**95.4**	**76.4**	**67.6**	**68.4**	**95.6**	**75.9**	**69.6**	**70.9**	57.8	**38.9**	**32.1**	**31.8**

Table 3: Judgment prediction results on PKU.

Haoxi Zhong, Guo Zhipeng, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, Maosong Sun. **Legal Judgment Prediction via Topological Learning**. EMNLP 2018.

基于拓扑结构预测的判决预测

- 对比实验

Tasks	t_1		t_2		t_3	
Metrics	Acc.	F_1	Acc.	F_1	Acc.	F_1
TOPJUDGE	**95.4**	**68.4**	**95.6**	**70.9**	**57.8**	**31.8**
- $t_3 \lhd t_1$	95.2	67.7	95.4	70.3	57.4	31.2
- $t_2 \lhd t_1$	94.8	64.7	94.9	60.2	57.0	31.6
ϕ	94.7	64.4	94.9	60.1	57.8	27.6

t1: 法条预测 t2: 罪名预测 t3: 刑期预测

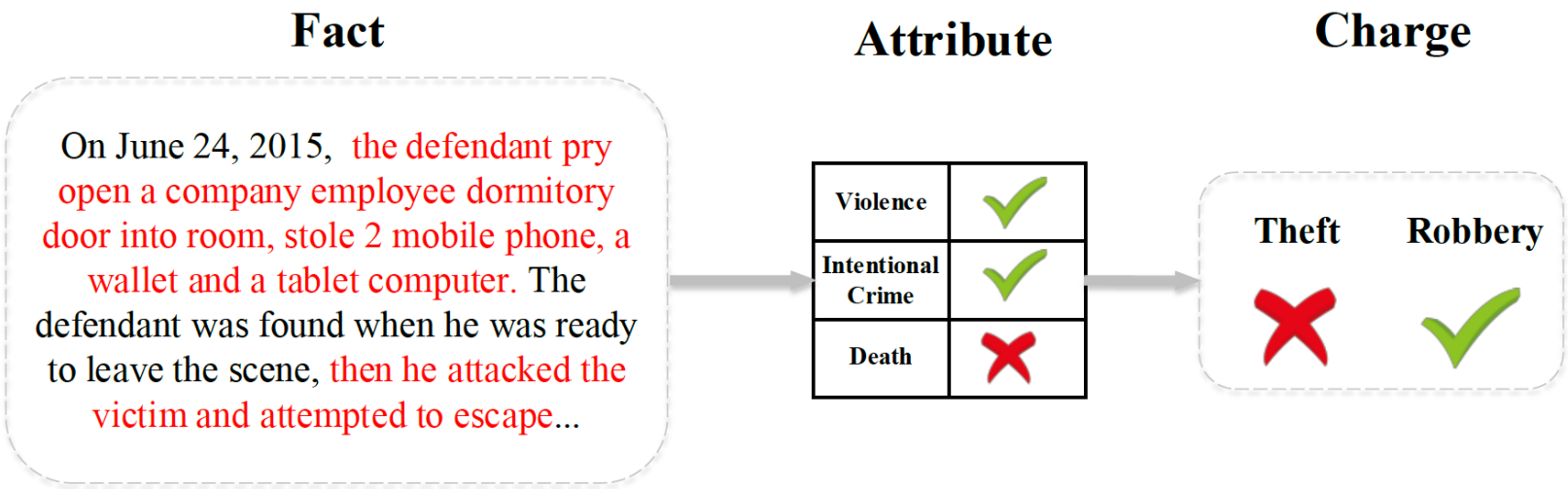
Haoxi Zhong, Guo Zhipeng, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, Maosong Sun. **Legal Judgment Prediction via Topological Learning**. EMNLP 2018.

引入区分性属性的罪名预测

- 低频罪名
  - 倒卖文物罪、扰乱法庭秩序罪、逃税罪
  - 10个高频罪名覆盖78%以上案例，50个低频罪名覆盖不到0.5%
- 混淆罪名
  - 盗窃罪、抢夺罪、抢劫罪
  - 贪污罪和挪用公款罪

引入区分性属性的罪名预测

- 引入显式的属性
  - 对低频罪名进行基于属性的判断
  - 对混淆罪名进行区分



引入区分性属性的罪名预测

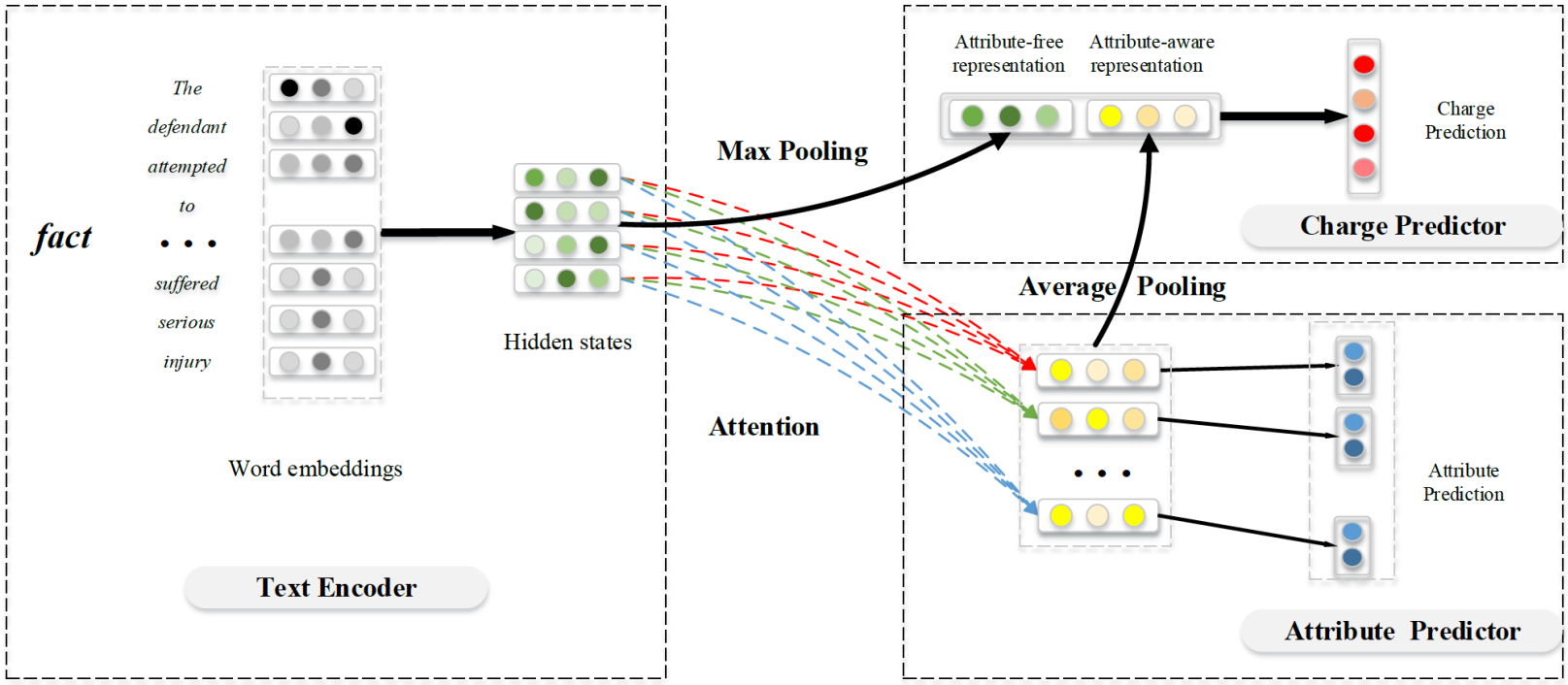
- 引入10个有区分性的罪名属性
  - 买卖行为、死亡、暴力行为、公共场所、营利目的等

Attributes	Description
Profit Purpose	Whether the criminal commits a crime on the purpose of getting profit.
Buying and Selling	Whether the criminal has buying or selling behavior during process of commission of crime.
Death	Whether death is caused to another person by the criminal.
Violence	Whether the criminal has the act of violence.
State Organ	Whether the case or the charge involves State organ or any functionary of a State organ.
Public Place	Whether the criminal commits a crime in a public place.
Illegal Possession	Whether the criminal commits a crime for the purpose of illegal possession.
Physical Injury	Whether a physical injury is caused to another person by the criminal.
Intentional Crime	Whether the criminal commits an intentional crime.
Production	Whether the criminal commits a crime during the production.

Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, Maosong Sun. Few-Shot Charge Prediction with Discriminative Legal Attributes. COLING 2018.

引入区分性属性的罪名预测

- 基于属性的罪名预测模型：
  - 多任务学习
  - 注意力机制



Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, Maosong Sun. Few-Shot Charge Prediction with Discriminative Legal Attributes. COLING 2018.

引入区分性属性的罪名预测

- 数据集
  - 40万刑事案件
  - 10个代表性属性
  - 149罪名

Datasets	Criminal-S	Criminal-M	Criminal-L
train	61,589	153,521	306,900
test	7,702	19,189	38,368
valid	7,755	19,250	38,429

Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, Maosong Sun. Few-Shot Charge Prediction with Discriminative Legal Attributes. COLING 2018.

引入区分性属性的罪名预测

- 罪名预测实验结果

Datasets	Criminal-S				Criminal-M				Criminal-L			
Metrics	Acc.	MP	MR	F1	Acc.	MP	MR	F1	Acc.	MP	MR	F1
TFIDF+SVM	85.8	49.7	41.9	43.5	89.6	58.8	5 0.1	52.1	91.8	67.5	54.1	57.5
CNN	91.9	50.5	44.9	46.1	93.5	57.6	48.1	50.5	93.9	66.0	50.3	54.7
CNN-200	92.6	51.1	46.3	47.3	92.8	56.2	50.0	50.8	94.1	61.9	50.0	53.1
LSTM	93.5	59.4	58.6	57.3	94.7	65.8	63.0	62.6	95.5	69.8	67.0	66.8
LSTM-200	92.7	60.0	58.4	57.0	94.4	66.5	62.4	62.7	95.1	72.8	66.7	67.9
Fact-Law Att.	92.8	57.0	53.9	53.4	94.7	66.7	60.4	61.8	95.7	73.3	67.1	68.6
Our Model	**93.4**	**66.7**	**69.2**	**64.9**	**94.4**	**68.3**	**69.2**	**67.1**	**95.8**	**75.8**	**73.7**	**73.1**

Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, Maosong Sun. Few-Shot Charge Prediction with Discriminative Legal Attributes. COLING 2018.

引入区分性属性的罪名预测

- 不同频率罪名的预测结果

Charge Type	Low frequency	Medium frequency	High frequency
Charge Number	49	51	49
LSTM-200	32.6	55.0	83.3
Our Model	**49.7 (↑ 17.1%)**	**60.0 (↑ 5.0%)**	**85.2 (↑ 1.9%)**

- 对比实验

Datasets	Criminal-S				Criminal-M				Criminal-L			
Metrics	Acc.	MP	MR	F1	Acc.	MP	MR	F1	Acc.	MP	MR	F1
Our model	93.4	**66.7**	**69.2**	**64.9**	94.4	68.3	**69.2**	**67.1**	**95.8**	**75.8**	**73.7**	**73.1**
w/o attention	**93.5**	63.4	60.1	60.0	94.7	**68.8**	58.2	60.9	94.9	70.9	54.4	58.6
w/o concatenation	**93.5**	59.3	59.0	57.2	**95.0**	64.6	62.4	62.5	95.7	69.4	64.5	65.4

Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, Maosong Sun. Few-Shot Charge Prediction with Discriminative Legal Attributes. COLING 2018.

小结

- 法律智能技术有着丰富的研究和应用前景
  - 信息检索、辅助判决、信息抽取、自动摘要、合规检查、风险预警、法治研究、…
- 法律智能特点：高度专业，富含知识
- 法律智能尚处起步充满挑战
  - 数据驱动（文书数据）+法律知识（法律人）
  - 应用场景

Q&A

liuzy@tsinghua.edu.cn
tucunchao@gmail.com