



中国科学院大学
University of Chinese Academy of Sciences

神经事件抽取综述

自然语言处理

樊润泽

中科院计算技术研究所

目录

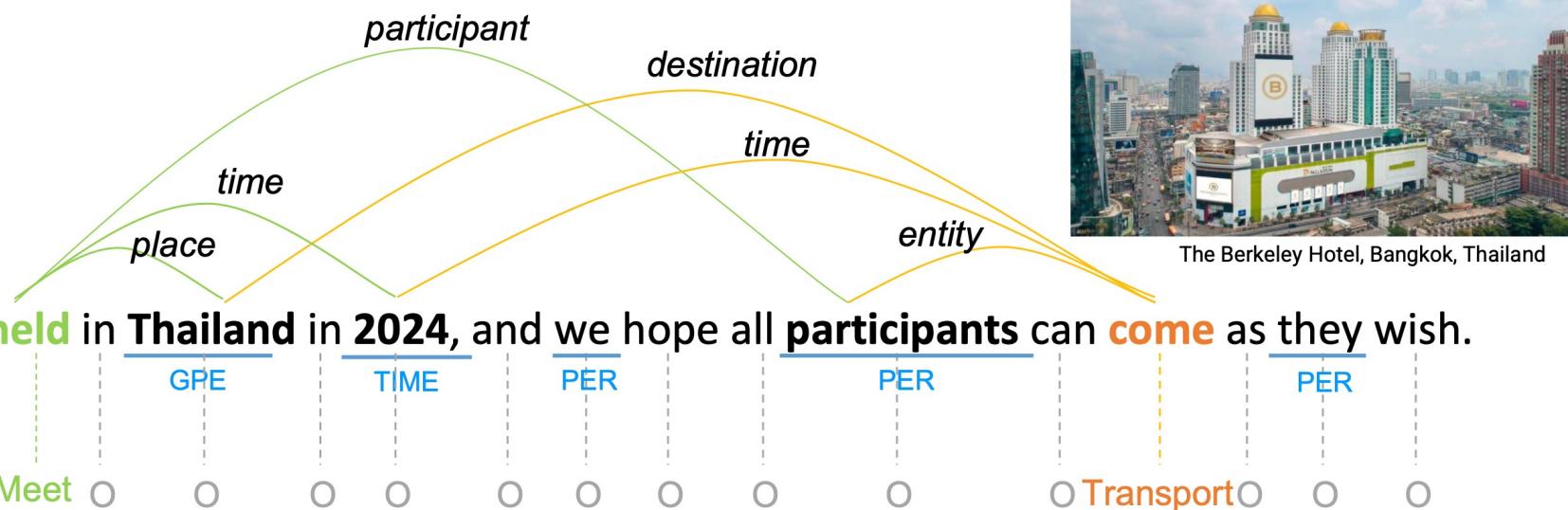
CONTENTS

- 01 任务定义
- 02 数据集与评价方法
- 03 主要方法
- 04 总结与未来展望



什么是事件抽取？

- 事件是指发生在某个特定时间，某个特定地点，由一个或多个角色参与的某个事情的发生。
- 事件可以被描述为状态的改变。



The Berkeley Hotel, Bangkok, Thailand

ACL will be **held** in **Thailand** in **2024**, and we hope all **participants** can **come** as they wish.

(Li and Ji, ACL 2021 Tutorial)

任务定义

■ 输入：

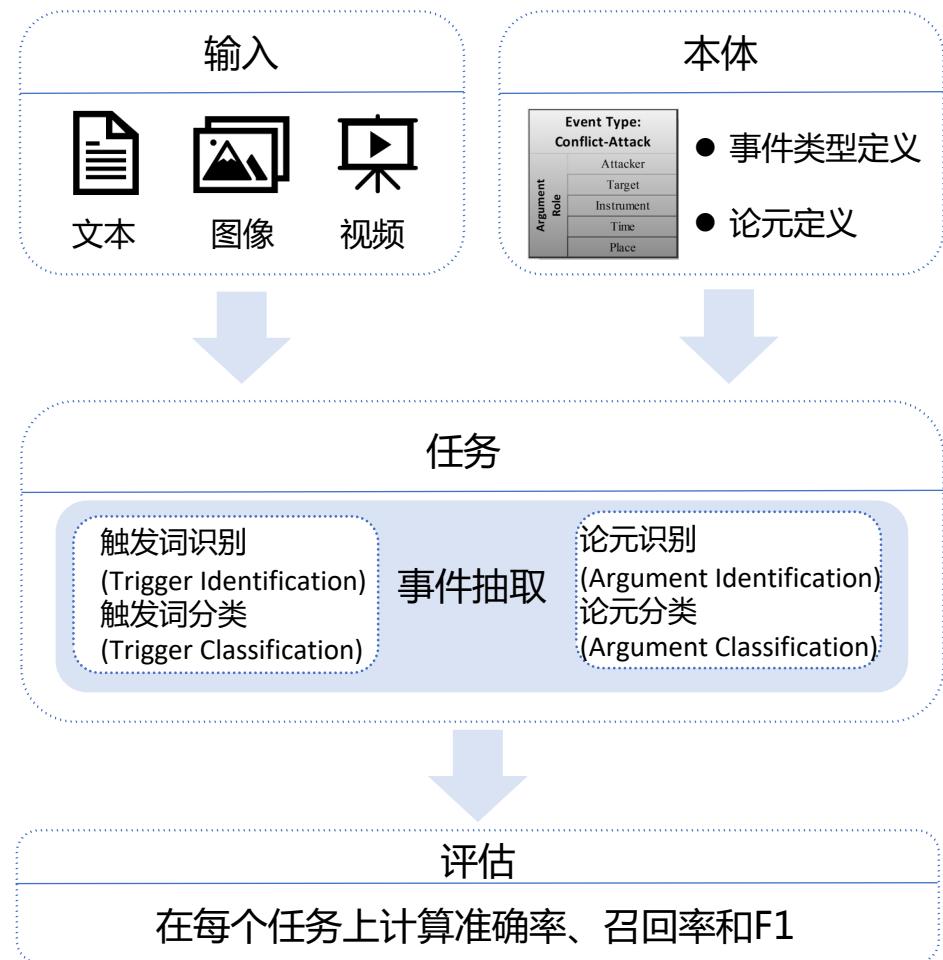
- 文本或图片或视频
- 本体：事件类型及其论元定义

■ 抽取：

- 事件 (Events)
 - 触发词识别 (Trigger Identification)
 - 触发词分类 (Trigger Classification)
- 论元 (Arguments)
 - 论元识别 (Argument Identification)
 - 论元分类 (Argument Classification)

■ 评估：

- 准确率(Precision)、召回率(Recall)、F1





任务分类





数据集

数据集	文档数	句子数	事件类型数量	语言	事件范围
ACE2005	599	18117	33	英语/汉语/阿拉伯语	句子级
TAC KBP 2015	360	12976	38	英语	句子级
TAC KBP 2016	505	9042	18	英语/汉语/西班牙语	句子级
TAC KBP 2017	500	8951	18	英语/汉语/西班牙语	句子级
MAVEN	4480	118732	168	英语	句子级
FewFC	-	8982	10	汉语	句子级
DuEE	-	17000	65	汉语	句子级
Google	11909	-	30	英语	篇章级
Twitter	1000	-	20	英语	篇章级
ChFinAnn(Doc2EDAG)	32040	-	5	汉语	篇章级

评价方法

- 触发词识别 (Trigger Identification): 触发词预测正确表示触发词字符串、触发词开始的索引和触发词结束的索引相等。

$$P_{TI} = \frac{\sum I(\hat{T} = T \wedge \hat{T}_s = T_s \wedge \hat{T}_e = T_e)}{N_{\hat{T}}} \quad (1)$$

$$R_{TI} = \frac{\sum I(\hat{T} = T \wedge \hat{T}_s = T_s \wedge \hat{T}_e = T_e)}{N_T} \quad (2)$$

$$F1_{TI} = \frac{2 \times P_{TI} \times R_{TI}}{P_{TI} + R_{TI}} \quad (3)$$

其中 \hat{T} 表示模型预测的触发词, T 表示数据集中标注的正确触发词。

- 论元识别 (Argument Identification): 论元识别正确是指论元字符串、论元开始的索引、论元结束的索引和论元所属事件的类别均相等。

$$P_{AI} = \frac{\sum I(\hat{A} = A \wedge \hat{A}_s = A_s \wedge \hat{A}_e = A_e \wedge \hat{T}_t = T_t)}{N_{\hat{A}}} \quad (7)$$

$$R_{AI} = \frac{\sum I(\hat{A} = A \wedge \hat{A}_s = A_s \wedge \hat{A}_e = A_e \wedge \hat{T}_t = T_t)}{N_A} \quad (8)$$

$$F1_{AI} = \frac{2 \times P_{AI} \times R_{AI}}{P_{AI} + R_{AI}} \quad (9)$$

其中 A 表示标注数据中正确的论元短语, \hat{A} 表示模型预测的论元短语, T_t 是指标注数据中正确的此论元 A 所属事件的类别, \hat{T}_t 是模型预测的此论元 A 所属事件的类别。

- 触发词分类 (Trigger Classification): 触发词分类正确表示在触发词识别正确的基础上, 其事件类型也预测正确, 即触发词字符串、触发词开始的索引、触发词结束的索引和触发词类型相等。

$$P_{TC} = \frac{\sum I(\hat{T} = T \wedge \hat{T}_s = T_s \wedge \hat{T}_e = T_e \wedge \hat{T}_t = T_t)}{N_{\hat{T}}} \quad (4)$$

$$R_{TC} = \frac{\sum I(\hat{T} = T \wedge \hat{T}_s = T_s \wedge \hat{T}_e = T_e \wedge \hat{T}_t = T_t)}{N_T} \quad (5)$$

$$F1_{TC} = \frac{2 \times P_{TC} \times R_{TC}}{P_{TC} + R_{TC}} \quad (6)$$

- 论元分类 (Argument Classification): 论元分类正确是指在论元识别正确的基础上, 论元在事件中充当的论元角色也分类正确。

$$P_{AC} = \frac{\sum I(\hat{A} = A \wedge \hat{A}_s = A_s \wedge \hat{A}_e = A_e \wedge \hat{T}_t = T_t \wedge \hat{A}_t = A_t)}{N_{\hat{A}}} \quad (10)$$

$$R_{AC} = \frac{\sum I(\hat{A} = A \wedge \hat{A}_s = A_s \wedge \hat{A}_e = A_e \wedge \hat{T}_t = T_t \wedge \hat{A}_t = A_t)}{N_A} \quad (11)$$

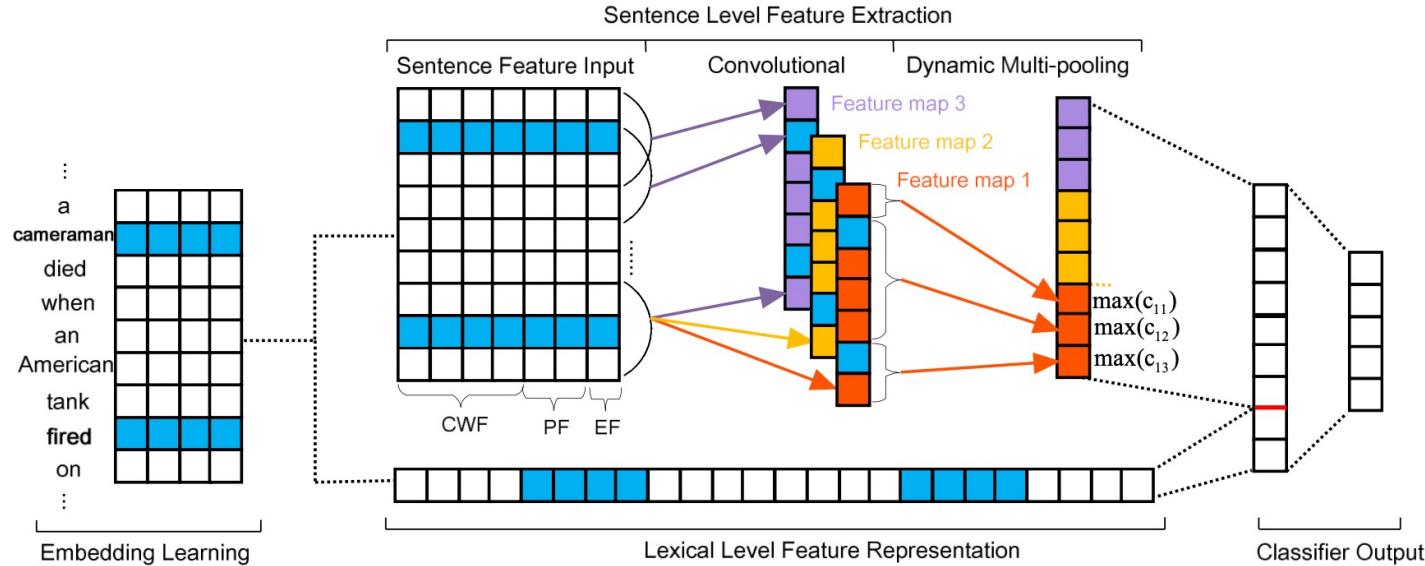
$$F1_{AC} = \frac{2 \times P_{AC} \times R_{AC}}{P_{AC} + R_{AC}} \quad (12)$$

其中 A_t 表示标注数据中正确的论元类型, \hat{A}_t 表示模型预测的论元类型。



卷积神经网络 (CNN)

(Chen et al., ACL 2015)



特点：词向量表示、词级别特征学习、句级别特征学习、事件触发词与论元分类。

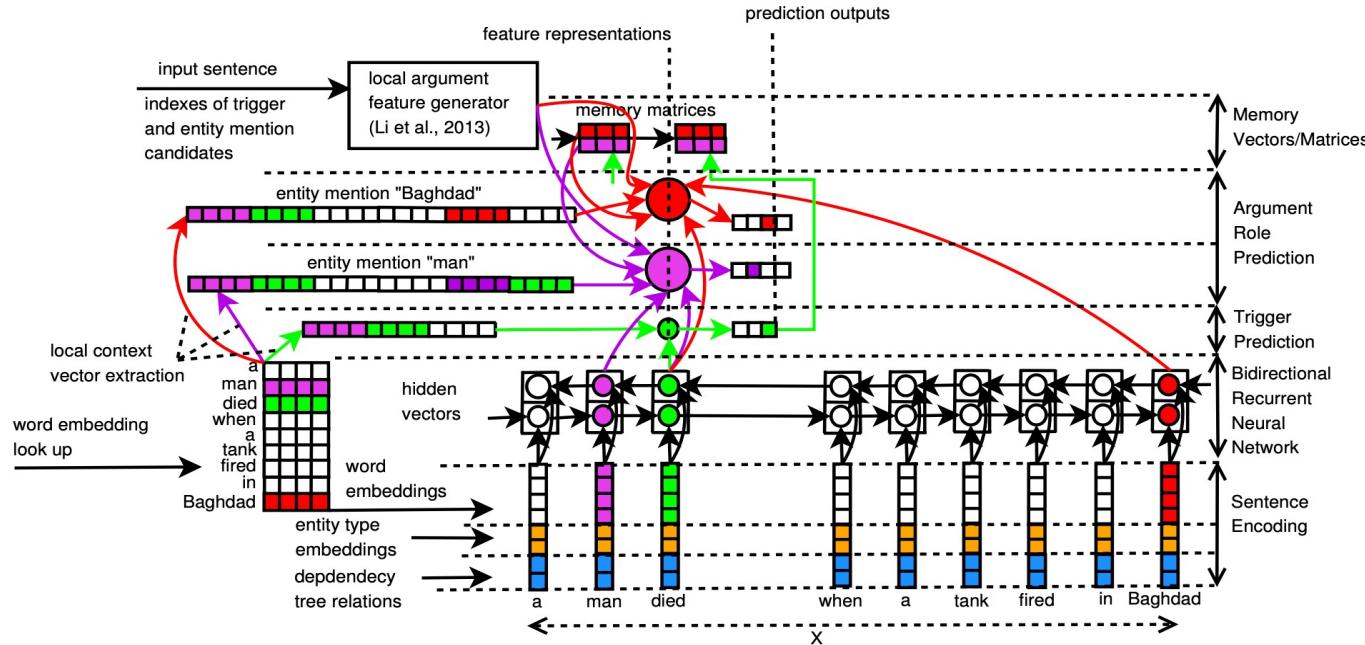
动态多池化技术针对触发词和论元的候选位置对句子分段，每一段做最大池化。

缺点：采用Pipeline的方法，会导致误差传播问题。

采用CNN模型，只能捕捉局部特征。

循环神经网络 (RNN)

(Nguyen et al., NAACL 2016)



特点：编码部分由词向量、依存关系、实体类别组成。

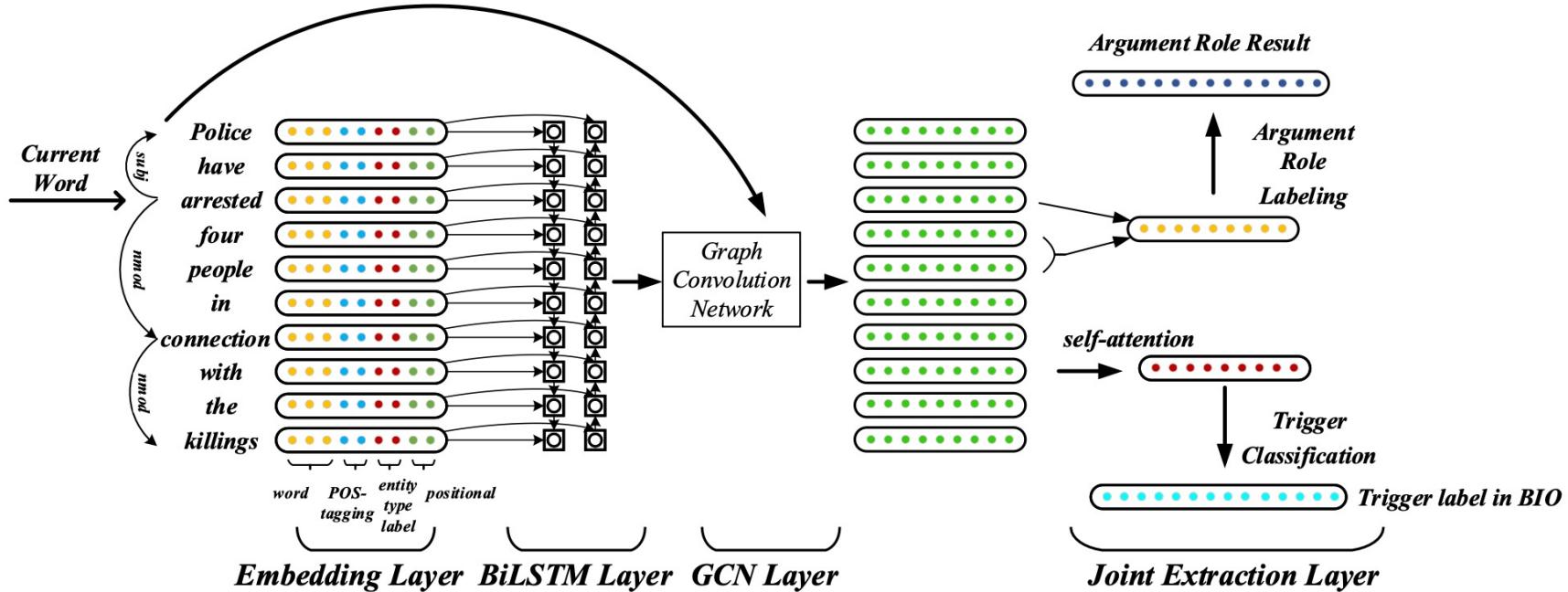
预测部分使用RNN联合预测事件触发词和论元。

缺点：实体类别、依存关系需要提前给定，若采用其他算法得到依然存在错误传播。

RNN存在长距离依赖问题，若触发词与论元距离较远则无法捕捉。



图卷积网络 (GCN) 与注意力机制



特点：采用双向LSTM学习句子特征，采用GCN和自注意力机制学习全局依赖信息。

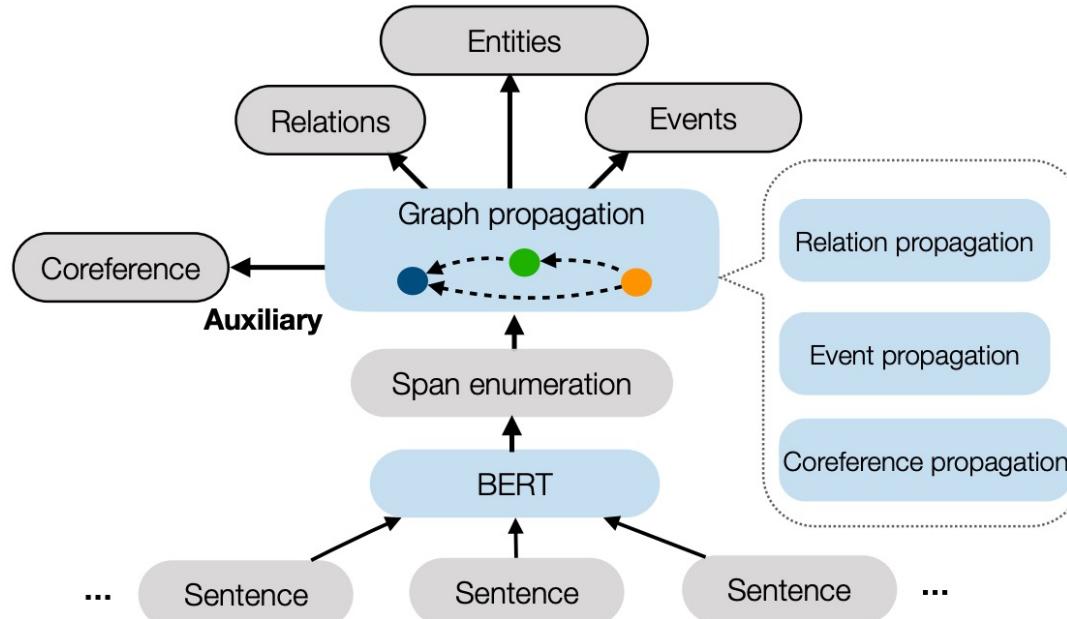
可以同时抽取一个句子中的多个事件及其论元。

缺点：实体类别、词性标注需要提前给定，若采用其他算法得到依然存在错误传播。



多任务学习 (Multi-task Learning)

(Wadden et al., EMNLP 2019)



特点：采用多任务学习同时解决命名实体识别、事件抽取和关系抽取任务，缓解了错误传播问题。

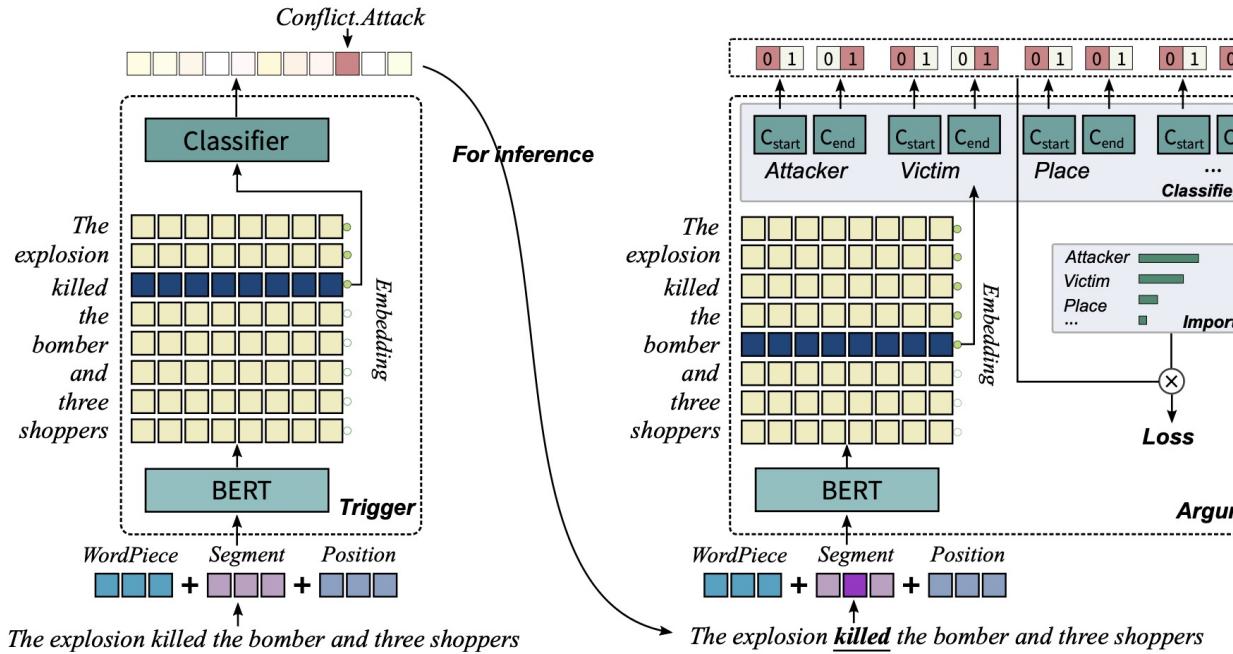
引入了Span表示学习。

缺点：实验证明在事件抽取任务中，图传播技术未起到作用。



预训练-微调范式

(Yang et al., ACL 2019)



特点：采用BERT动态词向量表示。

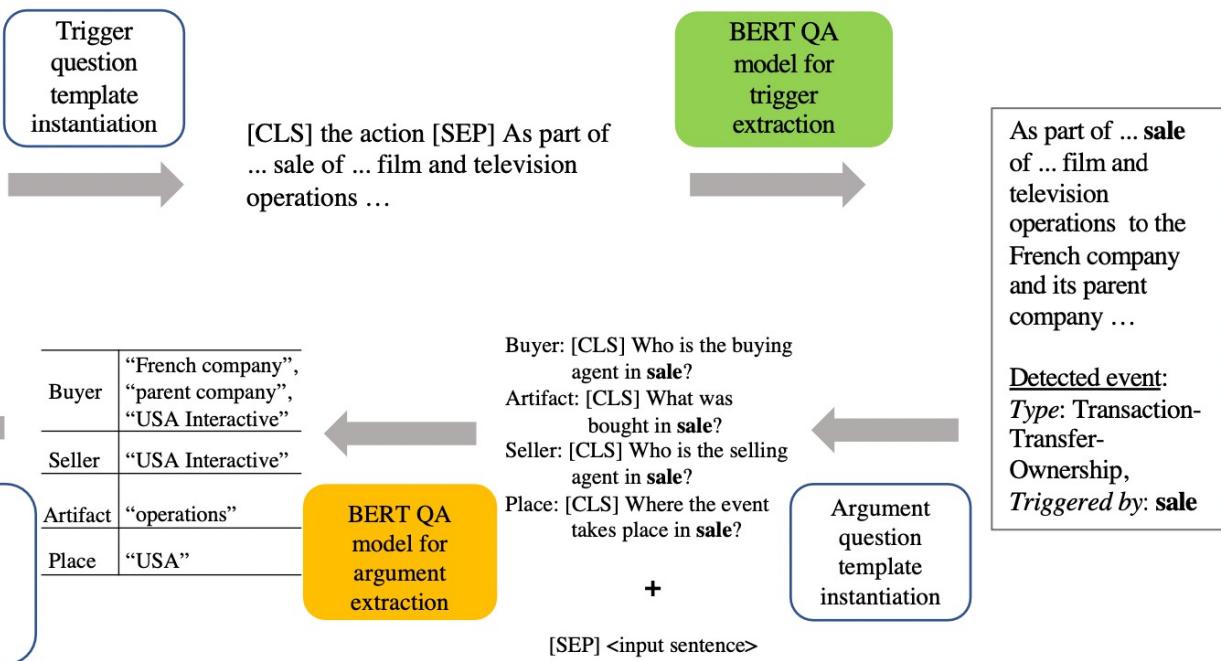
通过MSAK机制，使用BERT预测来进行数据增强。



QA范式

Input sentence:

As part of the 11-billion-dollar sale of USA Interactive's film and television operations ...



特点：将事件抽取任务转换为QA任务，从而避免了前置任务(如NER)所导致的错误传播问题。

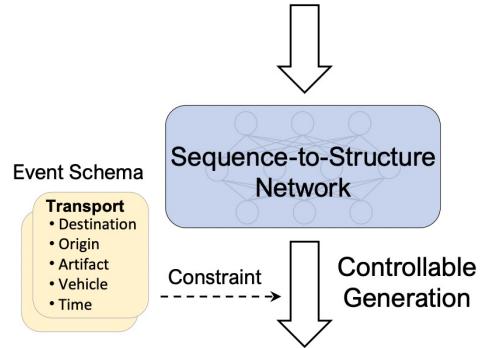
缺点：需要手动设计好问题模版。



生成式模型

(Lu et al., ACL 2021)

The man returned to Los Angeles from Mexico following his capture Tuesday by bounty hunters.

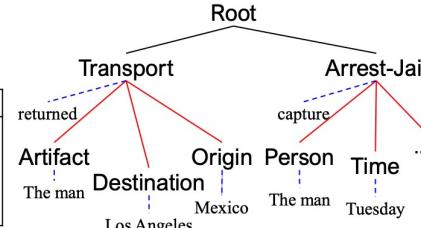


The man returned to Los Angeles from Mexico following his capture Tuesday by bounty hunters.

Event Type	Transport	Event Type	Arrest-Jail
Trigger	returned	Trigger	capture
Artifact	The man	Person	The man
Destination	Los Angeles	Time	Tuesday
Origin	Mexico	Agent	bounty hunters

(a) Record format.

Event Type	Transport	Event Type	Arrest-Jail
Trigger	returned	Trigger	capture
Artifact	The man	Person	The man
Destination	Los Angeles	Time	Tuesday
Origin	Mexico	Agent	bounty hunters



(b) Tree format.

(Transport returned
(Artifact The man)
(Destination Los Angeles)
(Origin Mexico))
(Arrest-Jail capture
(Person The man)
(Time Tuesday)
(Agent bounty hunters))

(c) Linearized format.

特点：将事件抽取任务转换为结构生成式任务。



提示学习 (Prompt)

(Liu et al., 2021)

让各个任务向预训练语言模型靠拢。

Paradigm	Engineering	Task Relation
a. Fully Supervised Learning (Non-Neural Network)	Features (e.g. word identity, part-of-speech, sentence length)	 CLS LM TAG GEN
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	 CLS LM TAG GEN
c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	 CLS LM TAG GEN
d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	 CLS LM TAG GEN



提示学习 (Prompt)

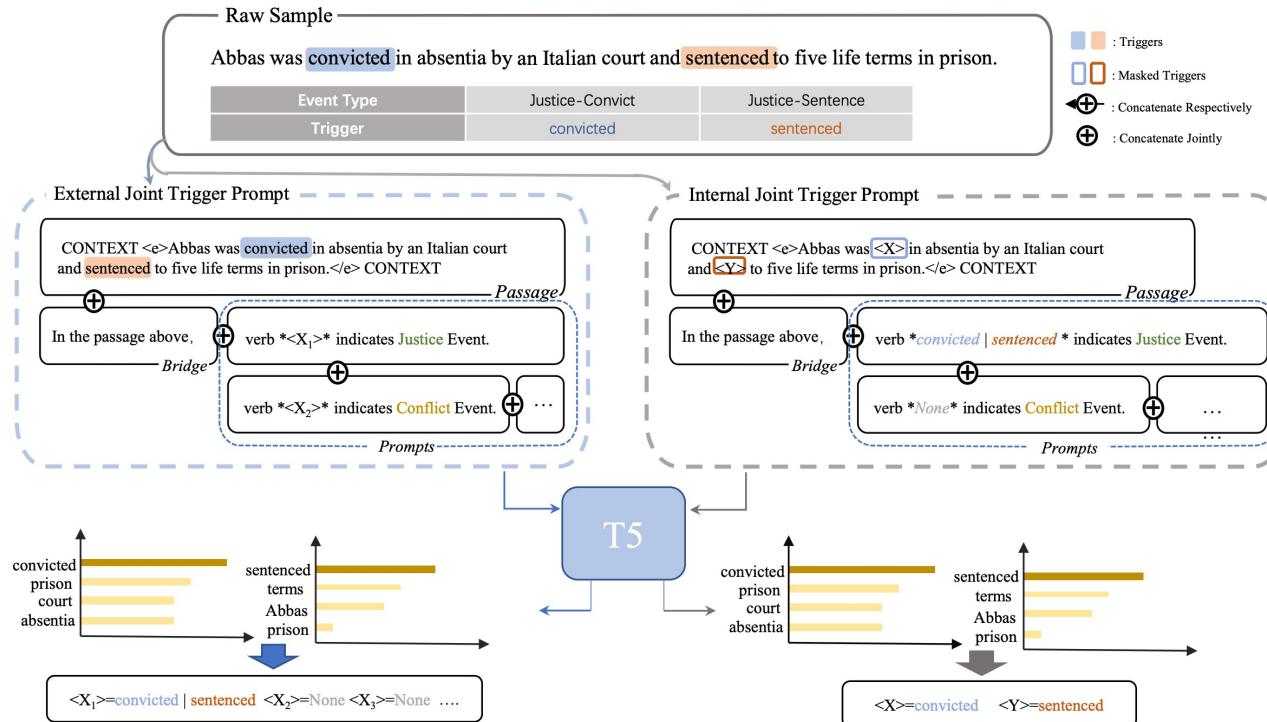
(Liu et al., 2021)

Type	Task	Input ([x])	Template	Answer ([z])
Text CLS	Sentiment	I love this movie.	[X] The movie is [Z].	great fantastic ...
	Topics	He prompted the LM.	[X] The text is about [Z].	sports science ...
	Intention	What is taxi fare to Denver?	[X] The question is about [Z].	quantity city ...
Text-span CLS	Aspect Sentiment	Poor service but good food.	[X] What about service? [Z].	Bad Terrible ...
	NLI	[X1]: An old man with ...		Yes
		[X2]: A man walks ...	[X1]? [Z], [X2]	No ...
Tagging	NER	[X1]: Mike went to Paris.		organization
		[X2]: Paris	[X1] [X2] is a [Z] entity.	location ...
Text Generation	Summarization	Las Vegas police ...	[X] TL;DR: [Z]	The victim ... A woman
	Translation			
		Je vous aime.	French: [X] English: [Z]	I love you. I fancy you. ...



提示学习 (Prompt)

(Lin et al., 2021)

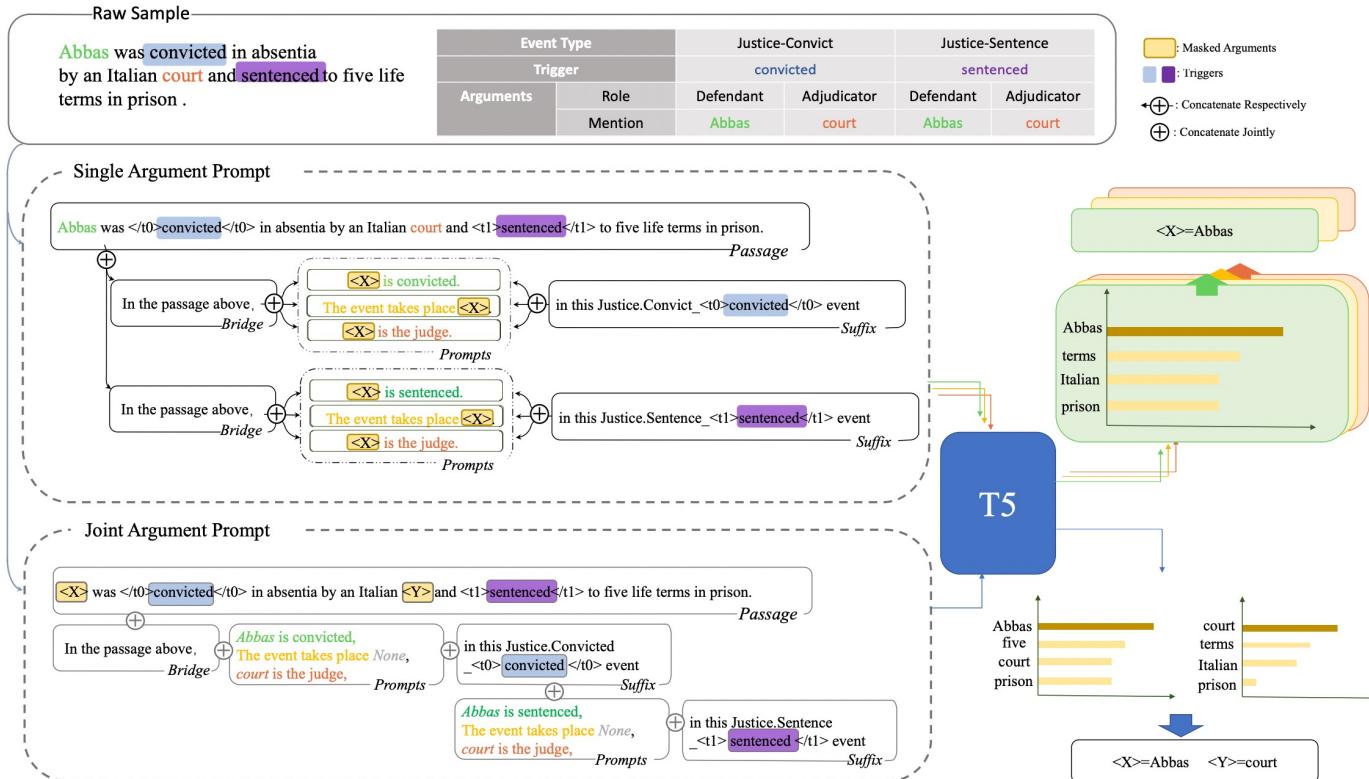


特点：为每一个事件类型设计一个抽取触发词的Prompt模版，让预训练语言模型去填词。
缺点：需要手动设计模版，此外并不知道什么模版才是最好的。



提示学习 (Prompt)

(Lin et al., 2021)



特点：为每一个事件类型设计一组抽取论元的Prompt模版，让预训练语言模型去填词。
缺点：需要手动设计模版，此外并不知道什么模版才是最好的。



总结

- 针对长距离依赖问题，采用图神经网络与注意力机制来解决。
- 针对错误传播问题，采用多任务学习、QA范式、生成式范式、Prompt范式来解决。
- 深度学习刚出现时，事件抽取任务通常被看作序列标注问题或分类问题来解决。
- 随着预训练语言模型的发展，事件抽取任务被看作QA问题、生成问题等。



未来展望

- 针对Prompt范式，目前已出现自动学习Prompt模版、连续型向量模版等，可以尝试将其用于事件抽取任务。
- 预训练语言模型蕴含着强大的能力，但由于预训练任务与事件抽取任务的不一致性，目前事件抽取任务并未很好的利用预训练模型。应从提示学习和设计事件抽取的专用预训练模型来入手，使其更好的利用大模型。
- 篇章级事件抽取的研究
- 多语言事件抽取的研究
- 多模态事件抽取的研究
- 开放域事件抽取的研究



中国科学院大学
University of Chinese Academy of Sciences

请批评指正

樊润泽

2021.12.30