

事件抽取与事件图谱构建

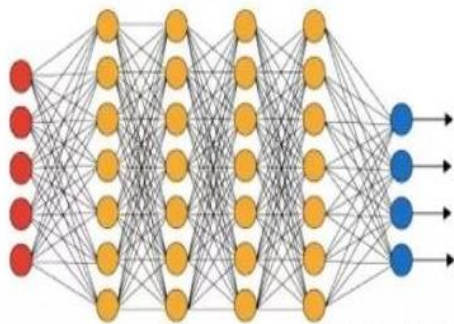
陈玉博

中国科学院自动化研究所
模式识别国家重点实验室



知识图谱

知识图谱是下一代人工智能的基础设施，
是实现**可解释人工智能**的重要手段。



深度学习



知识图谱

政府/企业	项目名称	启动时间	金额
美国国防部	Machines Common Sense (机器常识)	2018	3034万美元
	Automated Knowledge Acquisition (自动知识获取)	2019	2410万美元
	Knowledge-directed Artificial Intelligence Reasoning Over Schemas (知识指导的模式推理)	2019	1548万美元
欧盟地平线2020计划	Knowledge in the Making in the European Society (社会知识)	2017	---
Google	Knowledge Graph / Vault (知识图谱)	2012	---
百度	百度知心	2015	---
阿里巴巴	藏经阁计划	2018	---

事件图谱的意义：通用领域

丰富现有的知识图谱

名称	创建时间	数据来源	数据规模
OpenCyc	1984	专家知识	23万实体, 无事件
WordNet	1985	专家知识	15万实体, 无事件
YAGO	2007	WordNet + Wikipedia	459万实体, 无事件
DBpedia	2007	Wikipedia + 专家知识	1694万实体, 8万事件
Freebase	2008	Wikipedia + 领域知识 + 群体智慧	5872万实体, 2万事件

支撑其它信息获取引擎 语义搜索



事件监控



事件图谱构建的关键技术：事件抽取

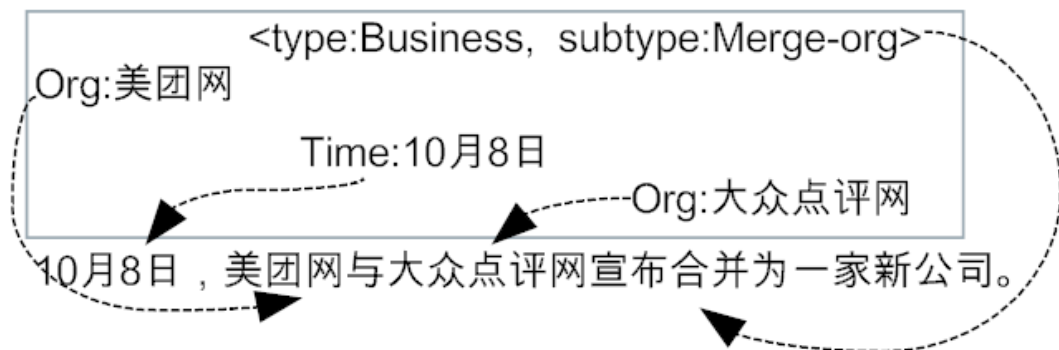
- 事件抽取

从自然语言文本中抽取用户感兴趣的事件信息并以结构化的形式呈现出来，如什么人/组织，什么时间，在什么地方，做了什么事

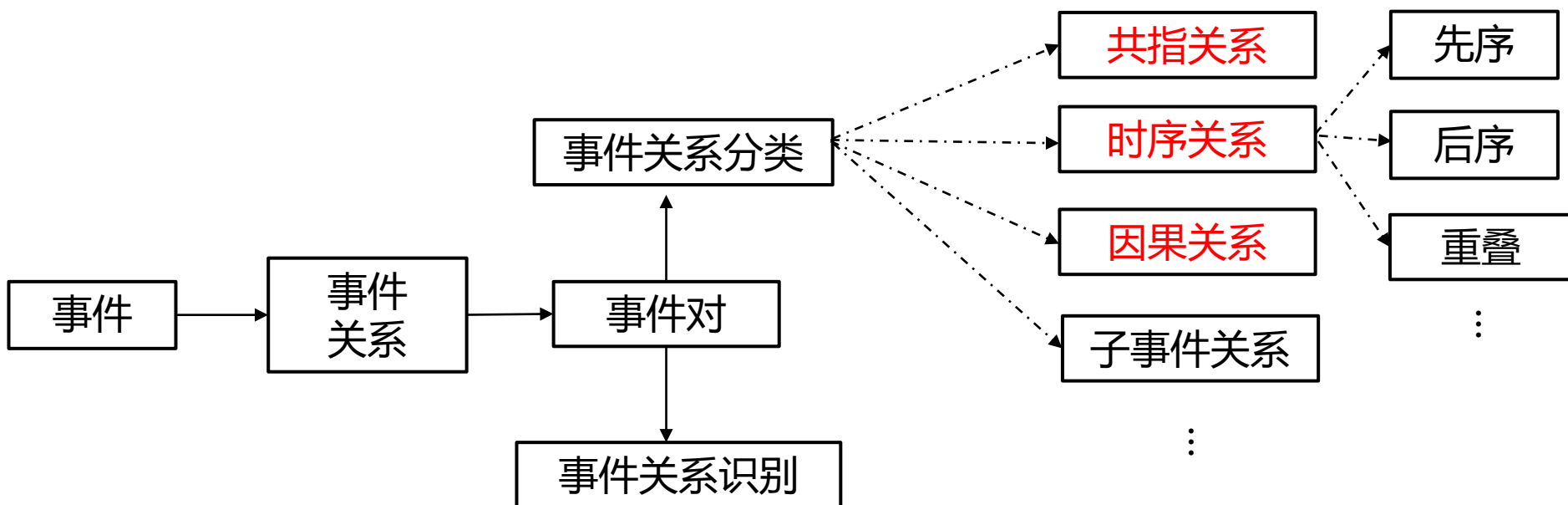
- 事件抽取相关的任务

事件发现 (Event Detection)：从文本中发现事件触发词 (Event Trigger)，

事件元素抽取 (Argument Extraction)：从文本中识别事件元素 (Event Argument) 并判断元素扮演的角色 (Argument role)



事件图谱构建的关键技术：事件关系抽取



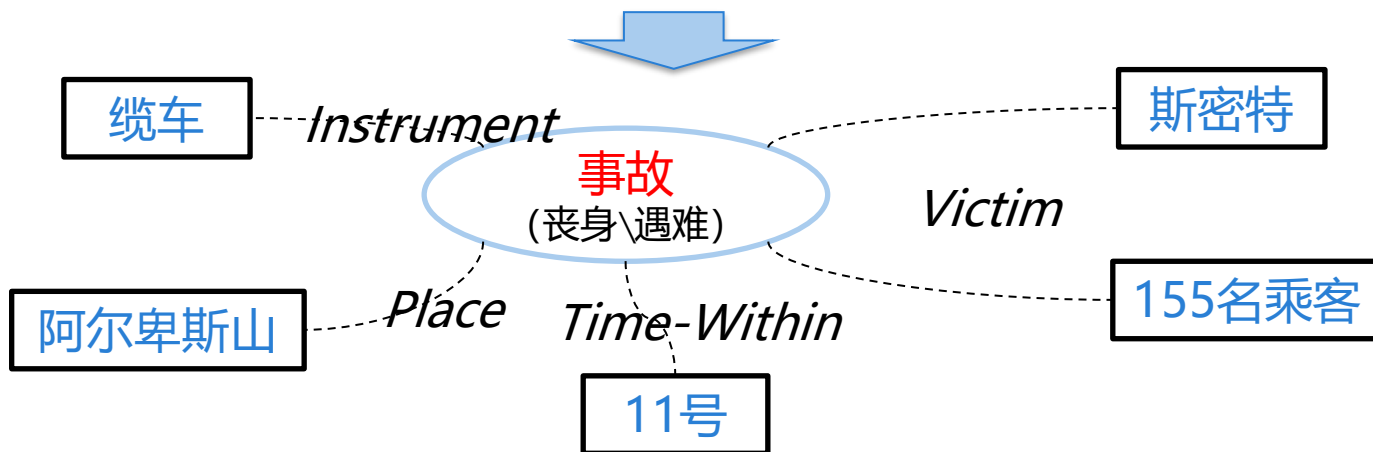
共指关系

- 数据来源：ACE-Chinese

S1:根据奥地利救灾组织的统计,在阿尔卑斯山登山缆车失火惨剧中丧生的155名乘客中包括有1999年世界女子花式滑雪冠军施密特。

S2:调查单位仍然无法断定事故发生的原因。但是指出,乘客的滑雪服装和设备都是易燃材料。

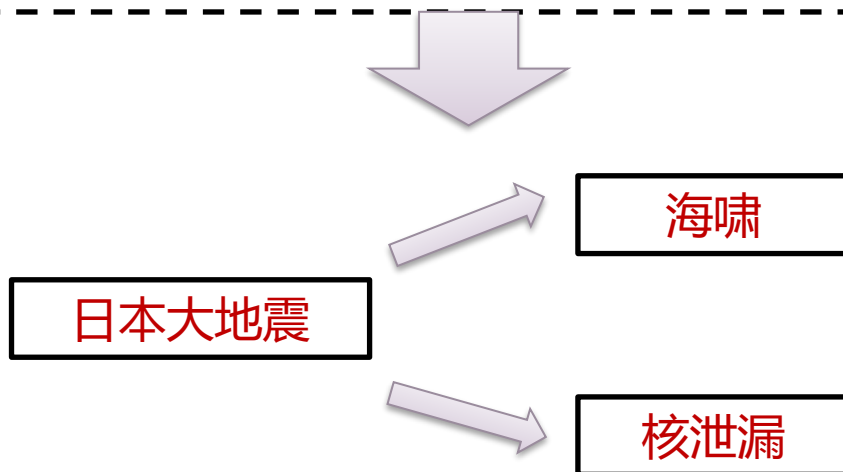
S3:奥地利一处滑雪胜地的登山缆车11号在隧道发生缆车失火惨剧。事发后有18名乘客及时逃脱存活,有155人不幸遇难。



因果关系

- 数据来源：百度百科

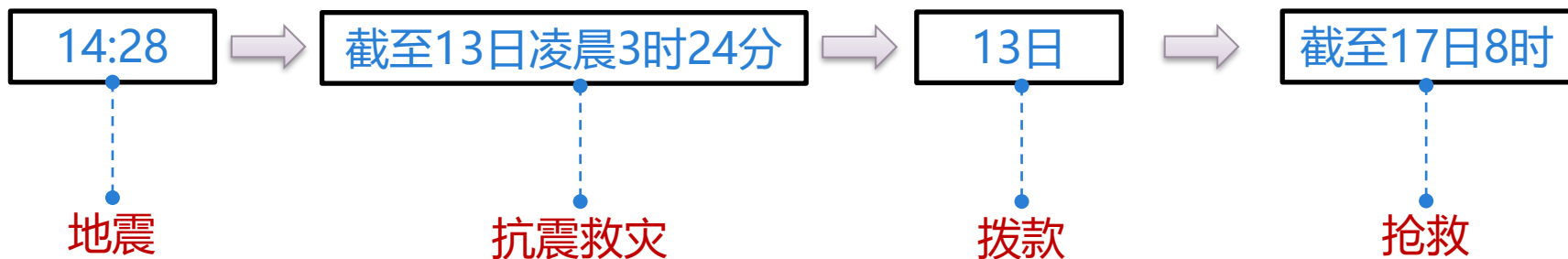
2011年3月11日13时46分，日本发生里氏9.0级大地震，地震随即引发了海啸，后来更使福岛核电站发生核泄漏危机。地震震中位于宫城县以东太平洋海域，震源深度20公里，东京有强烈震感。



时序关系

• 数据来源：百度百科

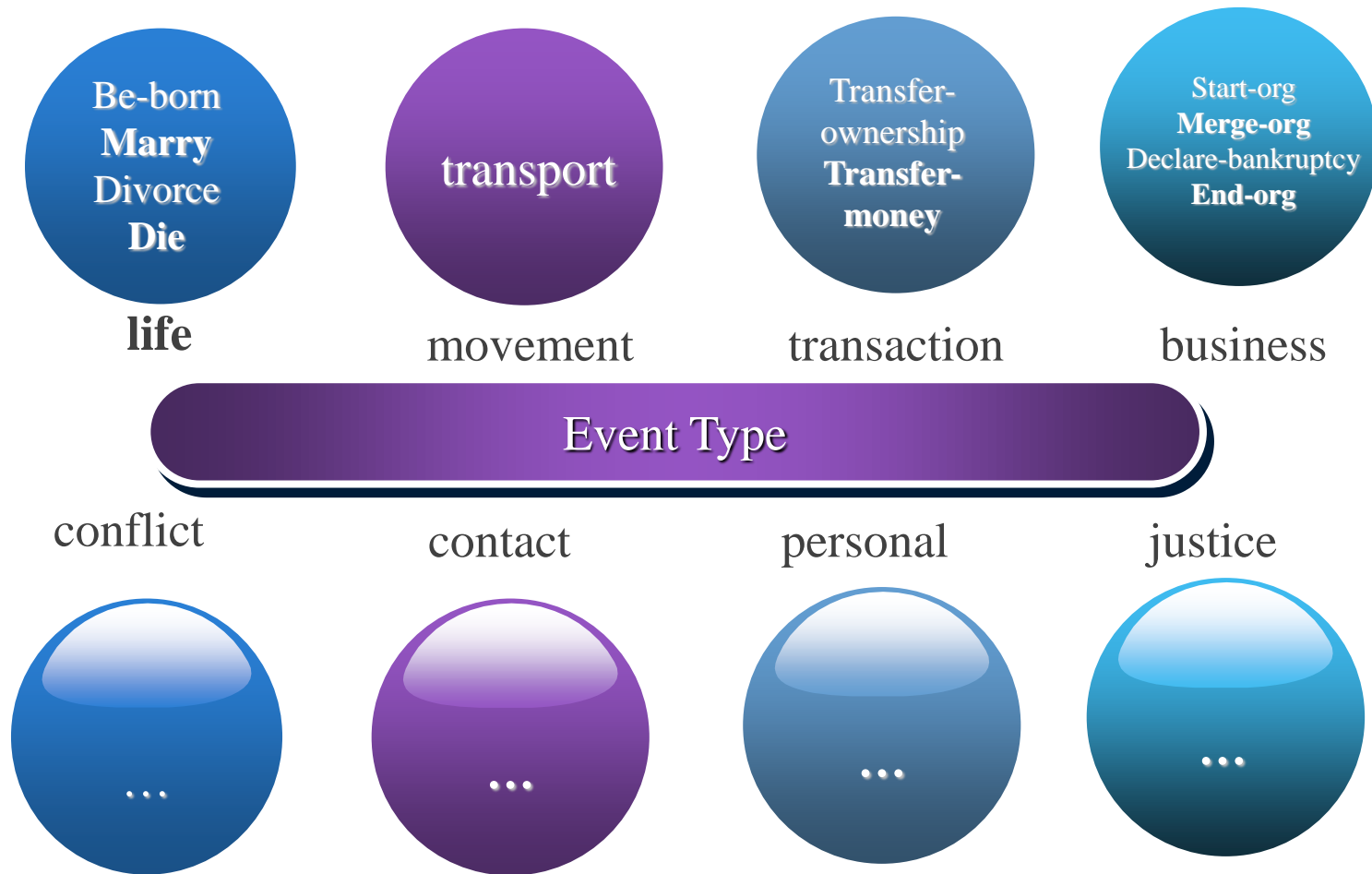
5月12日14:28，四川汶川发生7.8级地震。截至13日凌晨3时24分，武警部队已出动13000余名官兵急赴灾区抗震救灾。13日，为帮助地震灾区开展紧急救灾工作，中央财政紧急下拨款地震救灾资金8.6亿元。截至17日8时，参加救援的民兵预备役部队人员已从灾区废墟中一共抢救出1352名幸存者，发现和掩埋遇难者遗体2756人，救治伤员7600多人。



相关评测及语料

	MUC	TDT	ACE(KBP)	ECB+
Name	Message Understanding Conference	Topic Detection and Tracking	Automatic Content Extraction	EventCorefBank
Organizer	DARPA	DARPA	NIST	UA
Time	1987-1997	1998-2004	ACE: 2000-2008 KBP: 2014-2021	2014
Task	Event Extraction	Event Extraction	Event Extraction & Coreference Relation	Event Extraction & Coreference Relation

ACE事件类型示例



事件抽取

2005

基于特征的方法

代表性方法：
Ralph Grishman
et al. 2005、Ahn.
2006、Liao et
al. 2010、Hong
2011 et al.等

2013

基于结构的方法

代表性方法：
McClosky et al.
2011、Li et
al. 2013、Li et
al. 2014等

2015

基于神经网络的方法

代表性方法：
Chen et al. 2015、
Nguyen et al. 2016、
Yang et al. 2019、
Wang et al. 2019、
Liu et al. 2020等

Ralph Grishman, et al. "NYU's English ACE 2005 system description" . In JACE 2005

David Ahn, et al. "The stages of event extraction" . In ACL 2006

Shasha Liao, et al. "Using document level cross-event inference to improve event extraction". In ACL 2010

Yu Hong, et al. "Using cross-entity inference to improve event extraction" . In ACL 2011

David McClosky , et al. "Event extraction as dependency parsing" . In ACL 2011

Qi Li, et al. "Joint event extraction via structured prediction with global features" . In ACL 2013

Yubo Chen , et al. " Event Extraction via Dynamic Multi-Pooling Convolutional Neural Network" . In ACL 2015

Nguyen , et al. " Joint Event Extraction via Recurrent Neural Networks" . In NAACL 2016

Hongyu Lin , et al. " Nugget Proposal Networks for Chinese Event Detection" . In ACL 2018

Sen Yang , et al. " Exploring pre-trained language models for event extraction and generation " . In ACL 2019

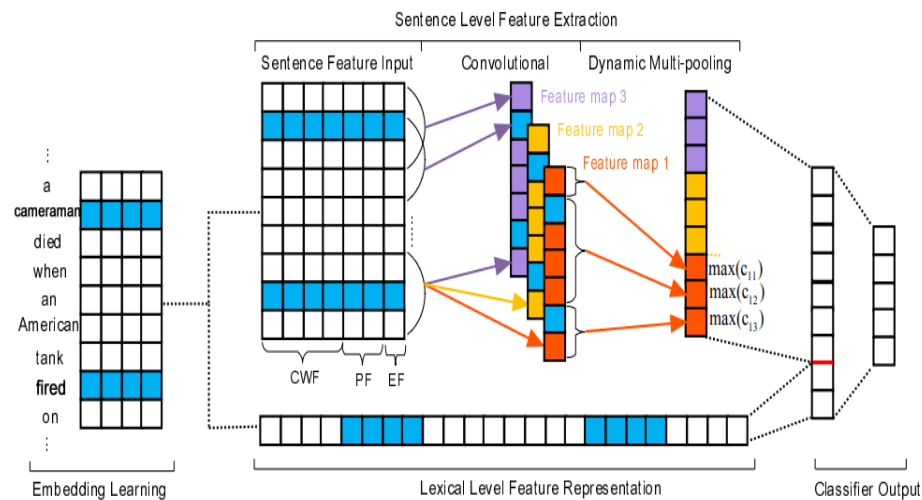
Xiao Zhi , et al. " HMEAE: Hierarchical Modular Event Argument Extraction" . In EMNLP 2019

Jian Liu , et al. " Event Extraction as Machine Reading Comprehension" . In EMNLP 2020

神经网络事件抽取

不同网络结构的事件抽取模型

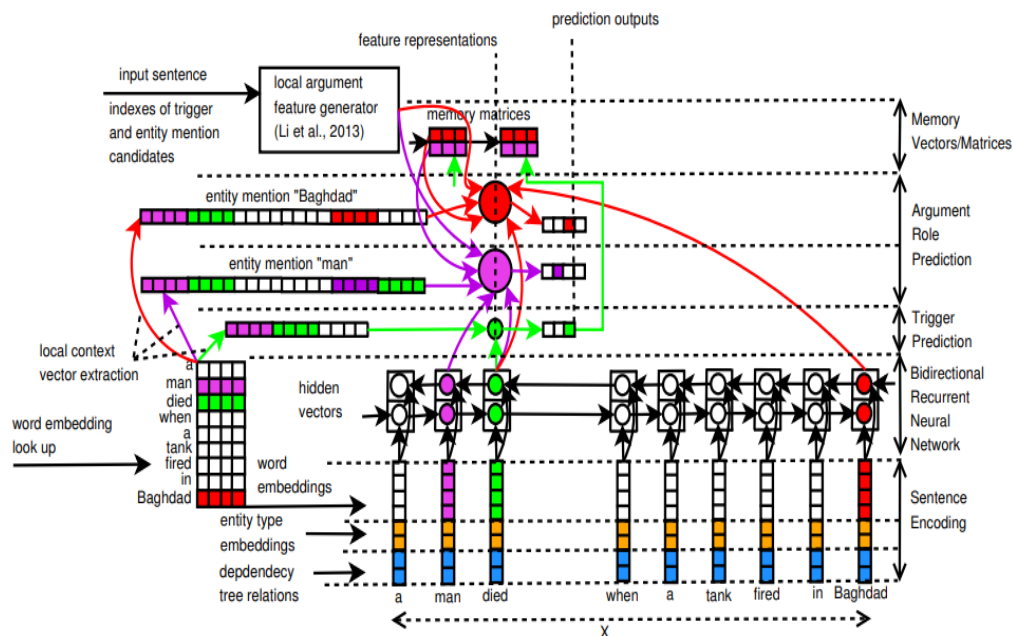
- **Convolutional Neural Networks**
- Recurrent Neural Networks
- Graph Neural Networks
- Transformers



神经网络事件抽取

不同网络结构的事件抽取模型

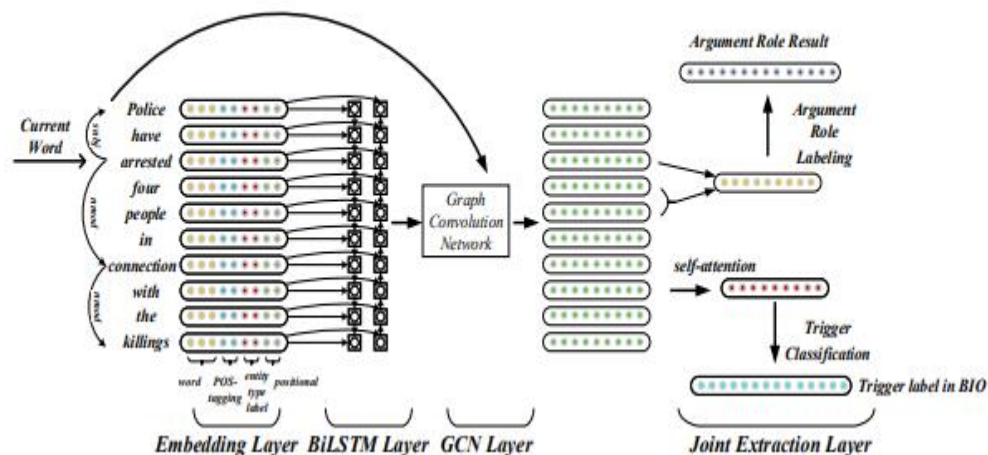
- Convolutional Neural Networks
- **Recurrent Neural Networks**
- Graph Neural Networks
- Transformers



神经网络事件抽取

不同网络结构的事件抽取模型

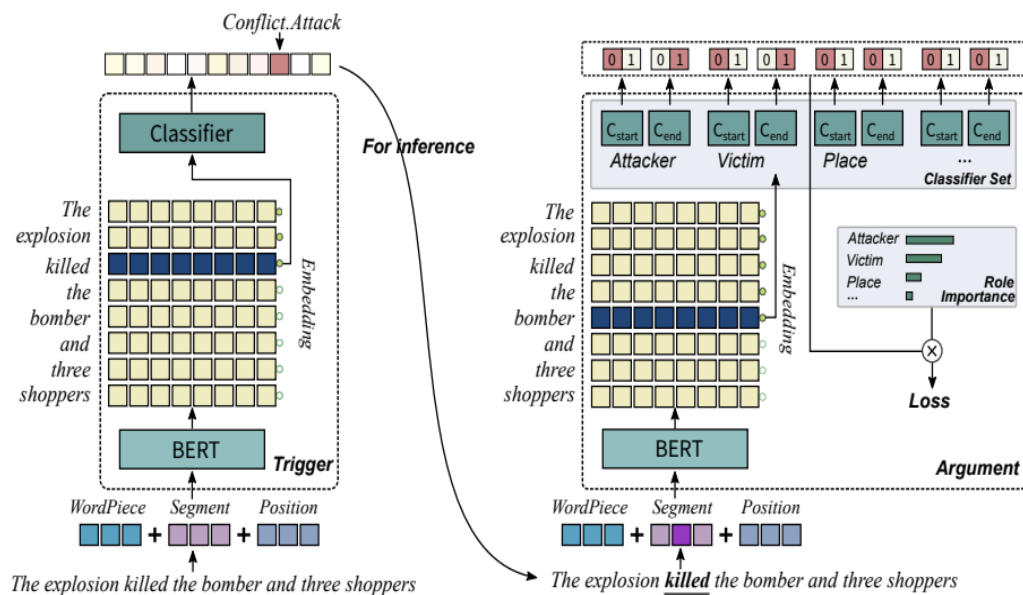
- Convolutional Neural Networks
- Recurrent Neural Networks
- **Graph Neural Networks**
- Transformers



神经网络事件抽取

不同网络结构的事件抽取模型

- Convolutional Neural Networks
- Recurrent Neural Networks
- Graph Neural Networks
- **Transformers**



神经网络事件抽取

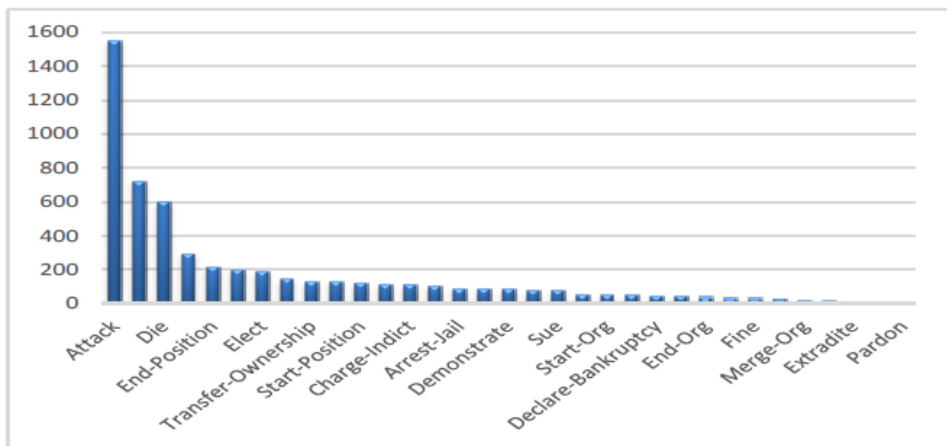
- 神经网络的事件抽取模型大力推动了事件抽取的研究
 - 在ACE 2005英文数据集上，事件检测性能
 - 2013年：67.5^[1]
 - 2020年：74.9^[2]
 - 在ACE 2005英文数据集上，元素抽取性能
 - 2013年：52.7^[1]
 - 2020年：70.1^[2]
- 问题：依赖大量人工标注的训练数据，仅利用文本从句子中抽取固定类别的事件知识。

[1] Li et al. Joint event extraction via structured prediction with global features. In ACL 2013

[2] Liu et al. Event Extraction as Machine Reading Comprehension. In EMNLP 2020

数据层面

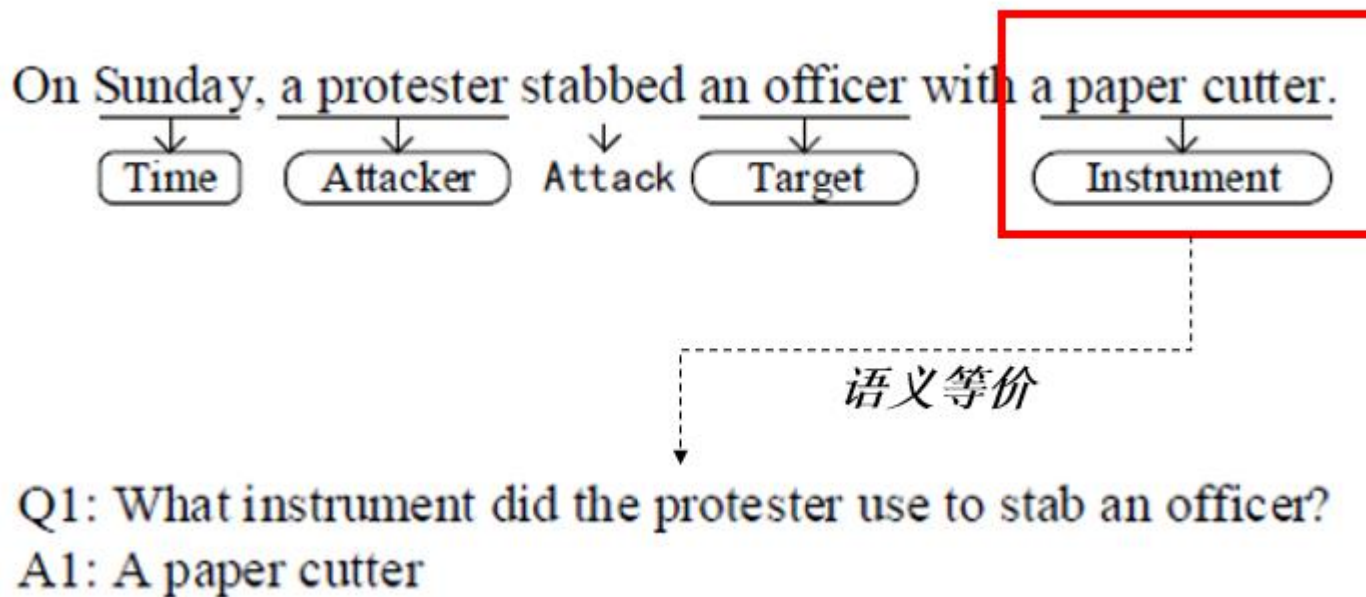
ACE2005 33类事件 599篇标注文档



事件类型	标注个数
机构合并（Merge-Org）	14
人事提名（Nominate）	12
司法引渡（Extradite）	7
无罪释放（Acquit）	6
假释出狱（Release-Parole）	2

基于阅读理解范式的事件抽取

- 动机
 - 传统方法将事件抽取建模为分类/序列标注的任务
 - 在样本较多的常见事件类型上表现较好
 - **在样本较少的罕见事件类型上表现较差**



基于阅读理解范式的事件抽取

- 挑战
 - 如何生成事件框架相关的问题?
 - 问句主题生成
 - 问句内容生成

问句主题

What is the instrument that a protester use to stab an officer?

问句内容

基于阅读理解范式的事件抽取

- 挑战
 - 如何生成事件框架相关的问题？
 - 问句主题生成

类别	语义角色	主题生成模板
时间	Time	When
地点	Place	Where
人物	Victim, Attacker, ...	Who is the #role
通用语义角色	Instrument, Target, ...	What is the #role

问句主题

What is the instrument that a protester use to stab an officer?

问句内容

基于阅读理解范式的事件抽取

- 挑战
 - 如何生成事件框架相关的问题？
 - 问句内容生成
 - 单语翻译（风格转换）任务
 - 无监督机器翻译

On Sunday, a protester stabbed an officer with a paper cutter

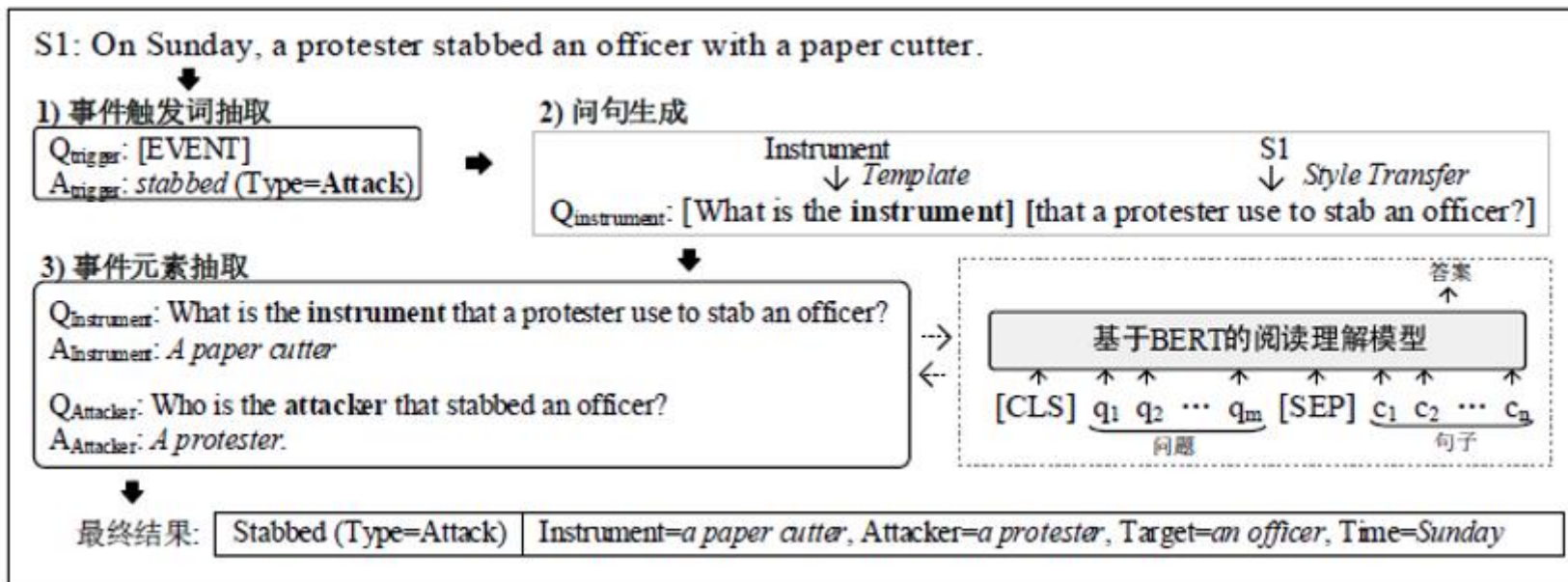
单语风格转换

What is the instrument **that a protester use to stab an officer?**

问句内容

基于阅读理解范式的事件抽取

- 模型框架



基于阅读理解范式的事件抽取

实验结果

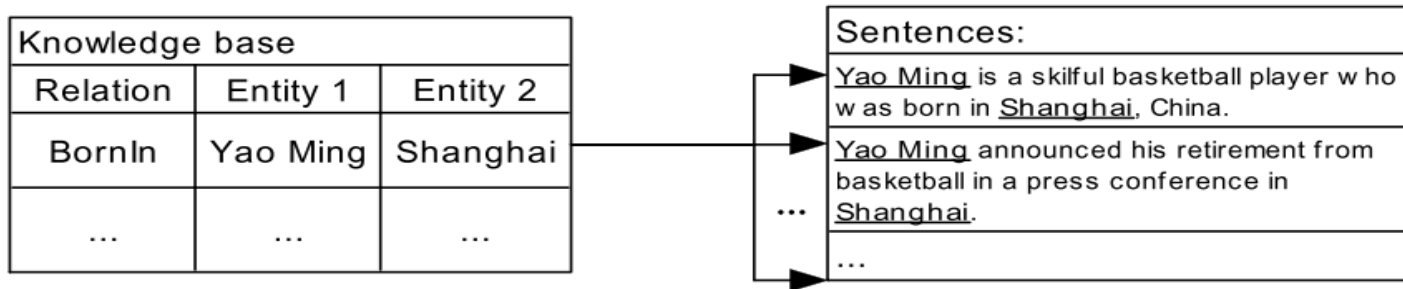
METHOD	TRIGGER EX.			ARGUMENT EX.			ARGUMENT EX.(O)		
	P	R	F1	P	R	F1	P	R	F1
JointBeam (Li et al., 2013)	73.7	62.3	67.5	64.7	44.4	52.7	-	-	-
DMCNN (Chen et al., 2015)	75.6	63.6	69.1	62.2	46.9	53.5	59.0 [†]	54.8 [†]	56.8 [†]
JRNN (Nguyen et al., 2016)	66.0	73.0	69.3	54.2	56.7	55.4	57.5 [†]	58.2 [†]	57.9 [†]
dbRNN (Sha et al., 2018)	74.1	69.8	71.9	66.2	52.8	58.7	58.4 [†]	64.2 [†]	61.2 [†]
JMEE (Liu et al., 2018b)	76.1	71.3	73.7	66.8	54.9	60.3	59.8 [†]	64.2 [†]	62.0 [†]
BERTEE	74.8 [†]	73.9 [†]	74.3 [†]	70.5 [†]	52.2 [†]	60.6 [†]	66.8 [†]	62.6 [†]	64.7 [†]
RCEE_ER (ours)	75.6	74.2	74.9*	63.0	64.2	63.6*	71.2	69.1	70.1*
RCEE_ER w/o DA (ours)	-	-	-	61.8	63.6	62.7	69.6	68.4	69.0

METHOD	1%	5%	10%	20%
DMCNN	-	8.7	16.6	23.7
dbRNN	-	8.1	17.2	24.1
BERTEE	2.20	10.5	19.3	28.6
RCEE	38.8	51.3	55.7	59.4
RCEE w/o DA	2.00	23.8	35.2	49.2
RCEE_ER	49.8	59.9	65.1	67.6
RCEE_ER w/o DA	2.20	26.5	37.8	54.1

DATASET	MODEL	P	R	F1
ACE2005	RCEE	25.5	26.0	25.8
	RCEE_ER	38.2	35.8	37.0
FrameNet	RCEE	18.2	15.3	16.6

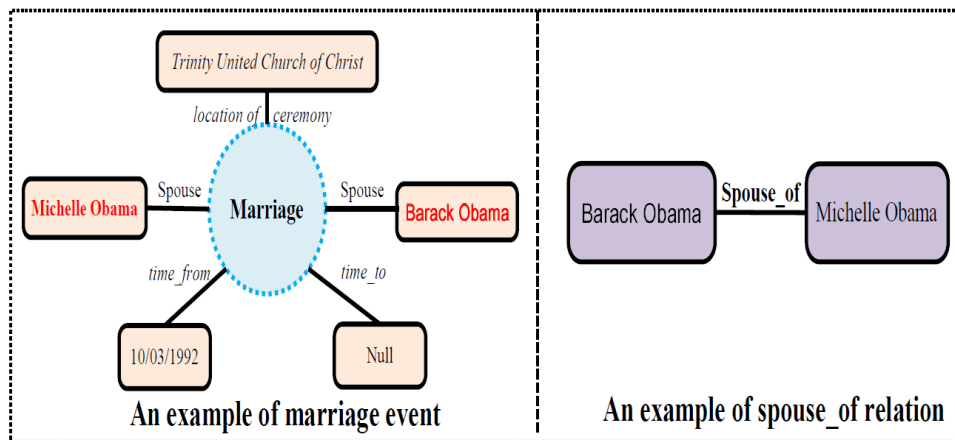
基于世界知识和语言学知识的数据自动生成

- 动机
 - 运用结构化的知识库自动生成语料
- 远距离监督的方法在关系抽取中取得成功



基于世界知识和语言学知识的数据自动生成

- 挑战1: 现有事件知识库中缺少触发词信息



- 关系知识库: (实体1, 关系, 实体2)
 - 可以利用Michelle Obama 和 Barack Obama回标
- 事件知识库: (事件实例, 事件类型; 角色1, 事件元素1; 角色2, 事件元素2 ... 角色N, 事件元素N)
 - 上图的例子中事件实例是Michelle Obama和Barack Obama结婚的事件, 知识库中用虚拟节点m.02nqglv表示。
 - 无法利用m.02nqglv和Barack Obama直接回标。

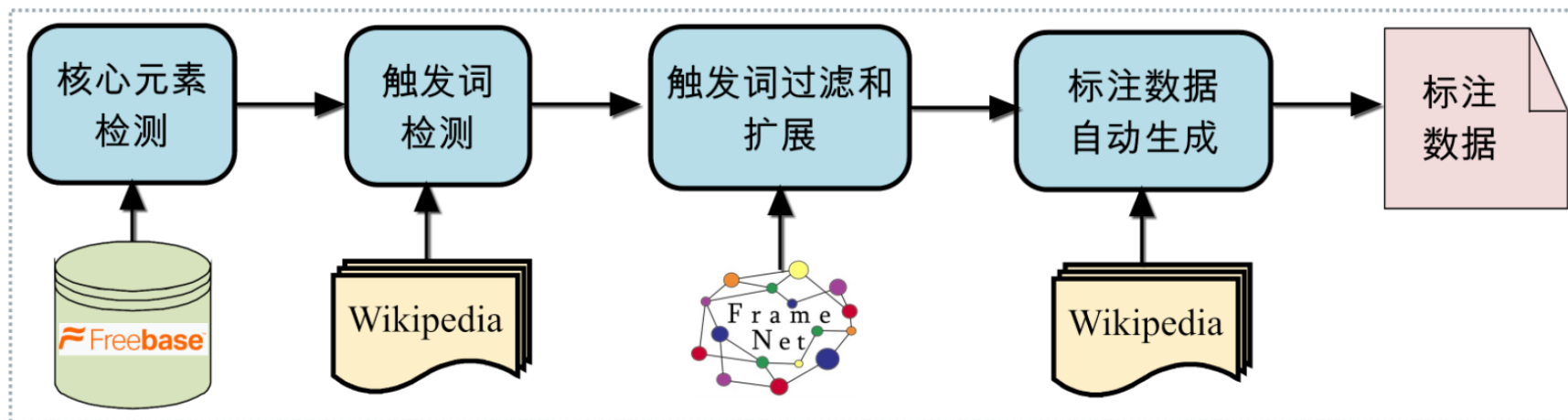
基于世界知识和语言学知识的数据自动生成

- 挑战2：事件元素并不都是必须出现
 - 知识库中只有60%的事件实例包含所有的事件元素
- 挑战3：一个事件的多个元素可以出现在多个句子中
 - 只有0.02%的事件实例能在一句话中找到所有的事件元素

事件类型	EI#	A#	S#
education.education	530,538	8	0
film.film_crew_gig	252,948	3	8
people.marriage	152,276	5	0
...
military.military_service	27,933	6	0
olympics.olympic_medal_honor	20,790	5	4
sum of the selected 21 events	3,870,492	100	798

基于世界知识和语言学知识的数据自动生成

- 方法
 - 利用世界知识和语言学知识
 - 自动生成大规模事件语料



基于世界知识和语言学知识的数据自动生成

- 自动生成的数据

Event Type	Freebase Size	Sentences (KA)	Sentences (KA+T)	Examples of argument roles sorted by KR	Examples of triggers
people.marriage	152,276	56,837	26,349	spouse, spouse, from, to, location	marriage, marry, wed, wedding, couple,..., wife
music.group_membership	239,813	90,617	20,742	group, member, start, role, end	musician, singer, sing, sang, sung, concert,..., play
education.education	530,538	26,966	11,849	student, institution, degree,..., minor	educate, education, graduate, learn, study,..., student
organization.leadership	43,610	5,429	3,416	organization, person, title,..., to	CEO, charge, administer, govern, rule, boss,..., chair
olympics.olympic_medal_honor	20,790	4,056	2,605	medalist, olympics, event,..., country	win, winner, tie, victor, gold, silver,..., bronze
...
sum of 21 selected events	3,870,492	421,602	72,611	argument1, argument2 ,..., argumentN	trigger1, trigger2, trigger3, ... , triggerN

- 当仅利用两个核心元素回标时，生成421,602个标注数据，但是这个数据中没有标注触发词信息
- 当同时利用核心元素和事件触发词回标时，生成72,611个标注数据
- 与ACE人工标注的将近6,000个的标注数据相比，我们提出的方法能自动生成大规模训练数据

基于世界知识和语言学知识的数据自动生成

- 标注数据的人工评价

##001 He is the uncle of [Amal Clooney], [wife] of the actor [George Clooney].

Trigger: wife Event Type: Marriage MannalAnotate[Y/N]:

Argument: Amal Clooney Role:Spouse MannalAnotate[Y/N]:

Argument: George Clooney Role:Spouse MannalAnotate[Y/N]:

##002 She was [married] to the cinematographer [Theo Nischwitz] and was sometimes credited as [Gertrud Hinz-Nischwitz].

Trigger: married Event Type: Marriage MannalAnotate[Y/N]:

Argument: Theo Nischwitz Role:Spouse MannalAnotate[Y/N]:

Argument: Gertrud Hinz-Nischwitz Role:Spouse MannalAnotate[Y/N]:

- 随机地从标注数据中选择500个样例，重复三次以平均正确率作为人工评价的结果
- 评判中每个句子都由三个标注者评价，最终投票决定

阶段	平均正确率 (%)
触发词标注	88.9
元素标注	85.4

基于世界知识和语言学知识的数据自动生成

- 标注数据的自动评价
 - 数据：ACE, ED only, ACE+ ED
 - 评价指标：同ACE一样
 - 实验结果：

Methods	Trigger Identification(%)			Trigger Identification + Classification(%)			Argument Identification(%)			Argument Role(%)		
	P	R	F	P	R	F	P	R	F	P	R	F
Li's structure trained with ACE	76.9	65.0	70.4	73.7	62.3	67.5	69.8	47.9	56.8	64.7	44.4	52.7
Chen's DMCNN trained with ACE	80.4	67.7	73.5	75.6	63.6	69.1	68.8	51.9	59.1	62.2	46.9	53.5
Nguyen's JRNN trained with ACE	68.5	75.7	71.9	66.0	73.0	69.3	61.4	64.2	62.8	54.2	56.7	55.4
DMCNN trained with ED Only	77.6	67.7	72.3	72.9	63.7	68.0	64.9	51.7	57.6	58.7	46.7	52.0
DMCNN trained with ACE+ED	79.7	69.6	74.3	75.7	66.0	70.5	71.4	56.9	63.3	62.8	50.1	55.7

ACE+ED取得到最好的效果，证明自动生成的数据可以有效扩展人工标注的数据

Shulin Liu, et al. " Leveraging FrameNet to Improve Automatic Event Detection" . In ACL 2016

Jian Liu, et al. " Event Detection via Gated Multilingual Attention Mechanism" . In AAAI 2018

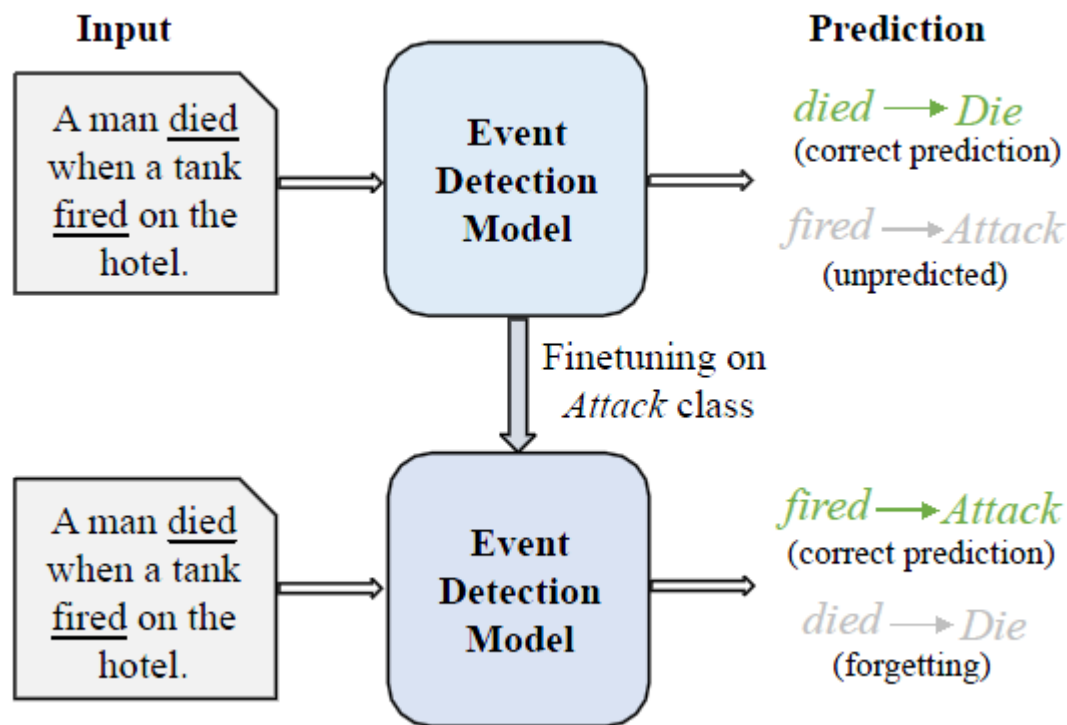
Jian Liu, et al. " Event Neural Cross-Lingual Event Detection with Minimal Parallel Resources" . In EMNLP 2019

Yang Zhou, et al. " What the role is vs. What plays the role:Semi-supervised Event Argument Extraction via Dual Question Answering" . In AAAI 2021

类别层面

- 动机

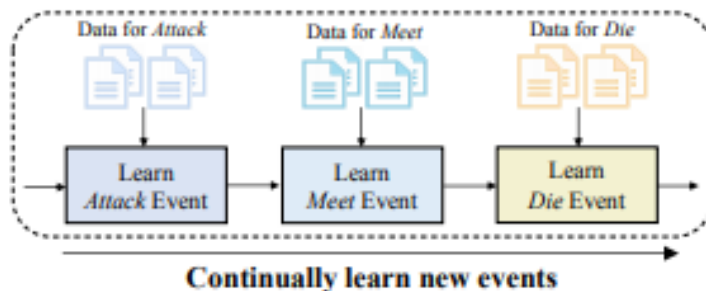
- 传统事件抽取方法都是在预定义类别的数据集上训练，事件类型固定
- 原始数据可能无法全部保存，全部类别重新训练时间和空间成本过高
- 直接在新类别上微调原有模型，会有遗忘灾难问题



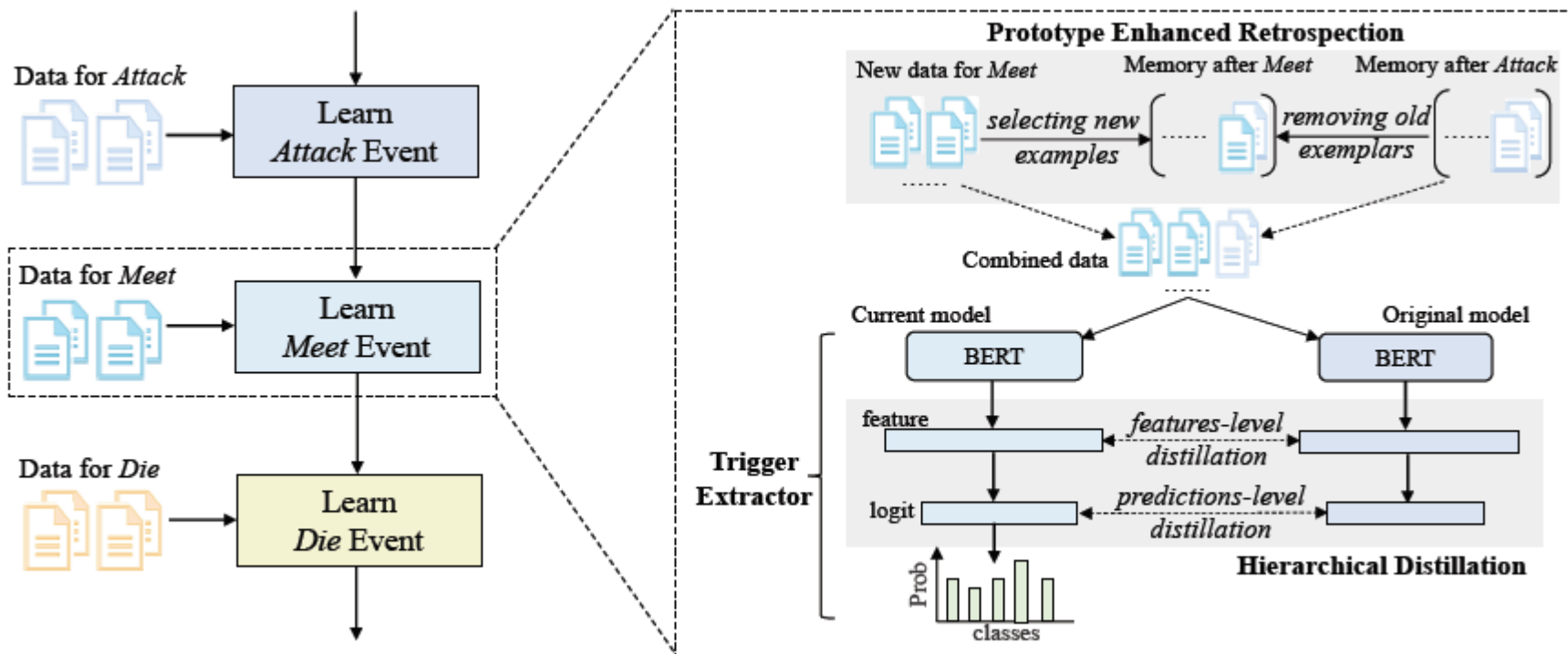
基于知识巩固网络的增量事件识别

挑战

- 如何保存有代表性的样本
- 新旧类别数据量不一致，会有偏置问题

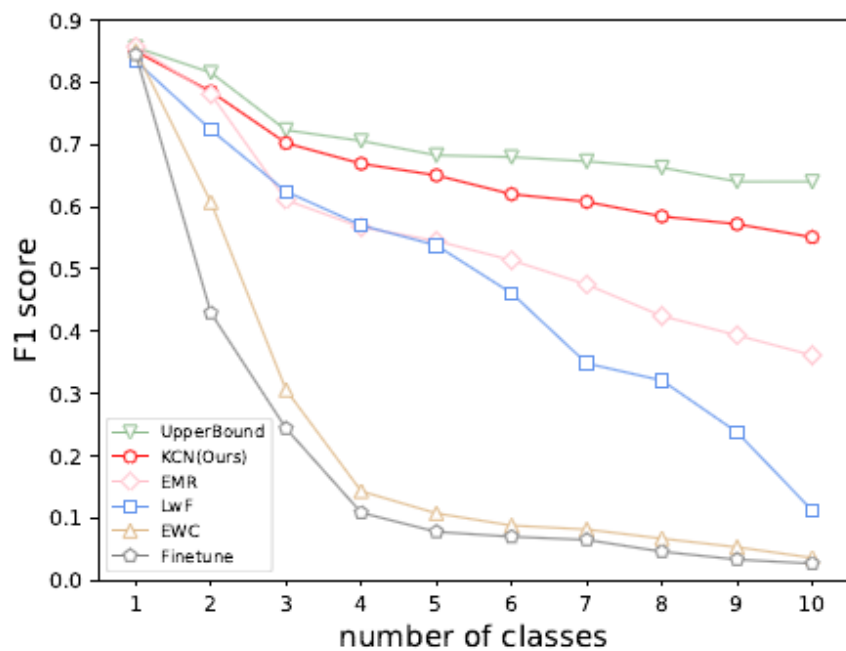


方法

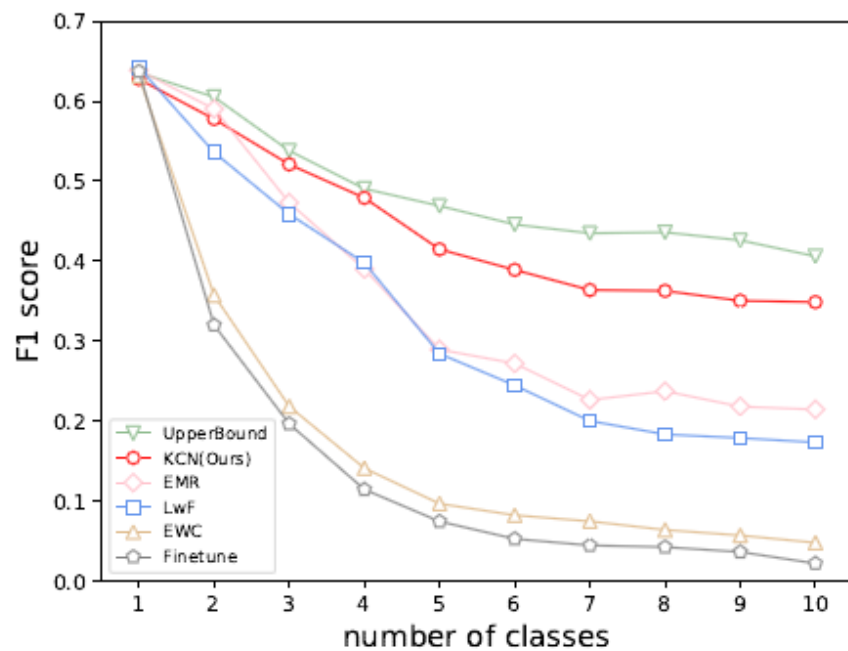


基于知识巩固网络的增量事件识别

- 实验



(a) ACE benchmark



(b) TAC KBP benchmark

基于知识巩固网络的增量事件识别

- 实验

Method	ACE		TAC KBP	
	Avg	Whole	Avg	Whole
Finetune	19.5	2.7	15.4	2.2
EWC	23.4	3.6	17.7	4.8
LwF	47.7	11.2	33.1	17.4
EMR	55.3	36.1	35.5	21.5
KCN(Ours)	65.9	55.1	44.3	34.9

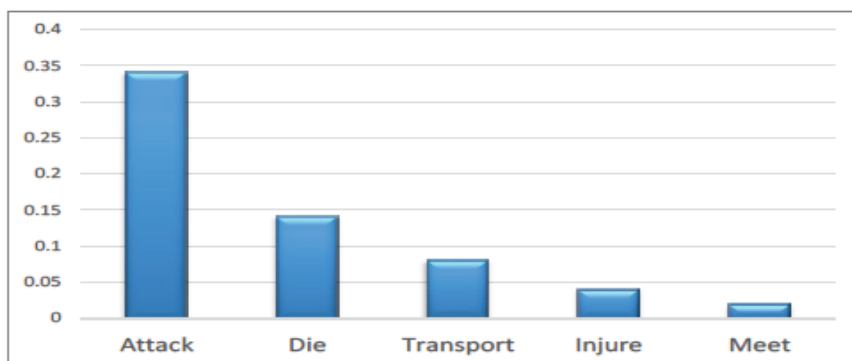
在观察数据集上的平均F1值
在全部测试集上的最终F1值

Models	ACE		TAC KBP	
	Avg	Whole	Avg	Whole
KCN	65.9	55.1	44.3	34.9
w/o PS	62.1	51.6	41.8	32.2
w/o FD	64.1	53.2	42.4	32.7
w/o PD	59.7	50.6	38.2	27.4
w/o HD	58.2	44.3	36.8	24.0
w/o PS and HD	55.3	36.1	35.5	21.5

消融实验

文本层面

- 动机
 - 仅利用句子文本，单独抽取多个事件，很难准确抽取全部事件
 - 一句话的多个事件之间具有依存关系



Die



S1: In Baghdad, a cameraman **died** when an American tank **fired** on the Palestine hotel.

Attack or End-position?

S2: The project leader was **fired** for the **bankruptcy** of the subsidiary company.



Bankruptcy

多事件协同抽取

- 动机
 - 仅利用句子文本，单独抽取多个事件，很难准确抽取全部事件
 - 篇章级信息能帮助句子级信息消除歧义

篇章级信息更重要的情况:

哈哈下班了



Transport

S3: He **left** the company.



End-position

He planned to go shopping before he went home, because he got off work early today

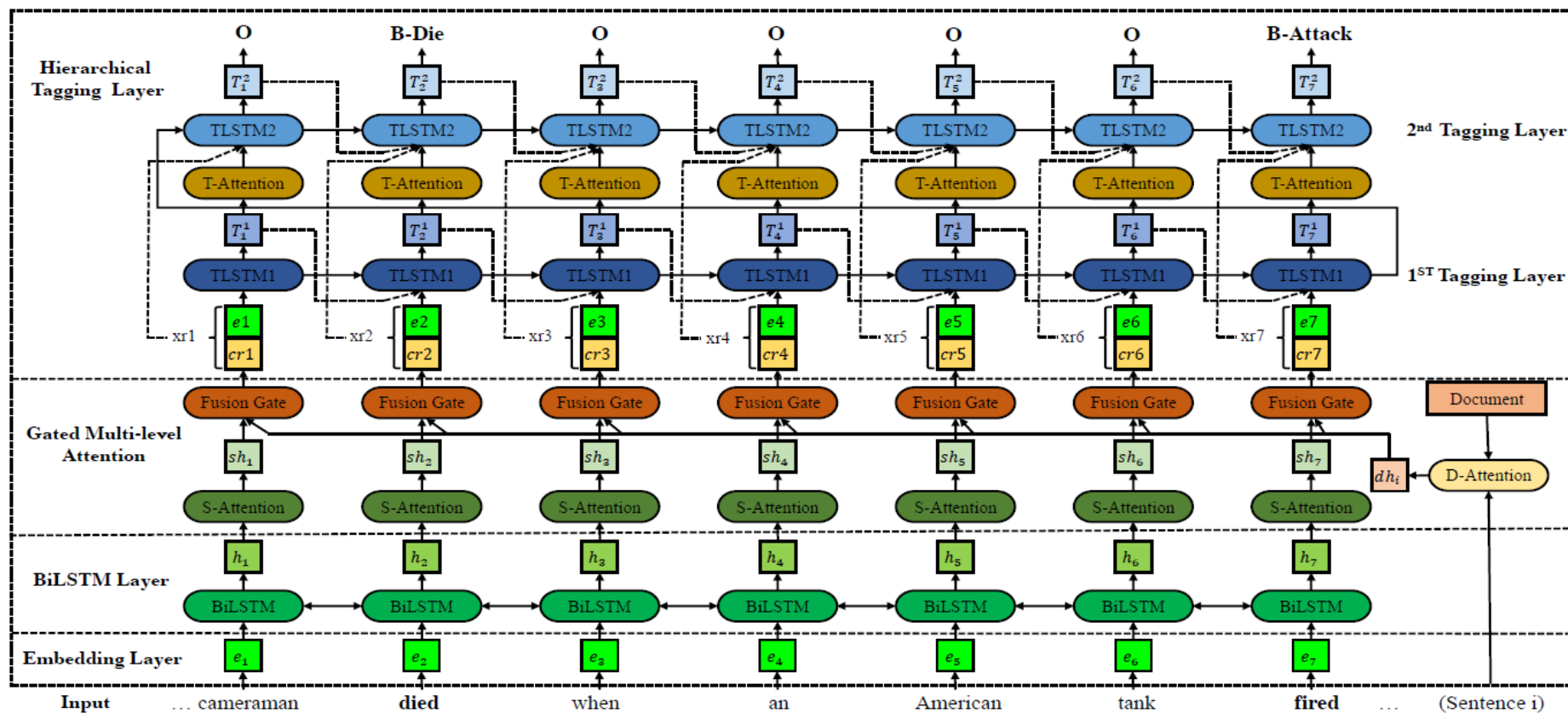
They held a party for his retirement.

句子级信息更重要的情况:

S1: In Baghdad, a cameraman **died** when an American tank **fired** on the Palestine hotel.

多事件协同抽取

- 基于层次偏置标注网络和多粒度门控关注机制的多事件协同抽取



多事件协同抽取

实验结果

Methods	P	R	F_1
Li's MaxEnt (2013)	74.5	59.1	65.9
Liao's CrossEvent (2010)	68.7	68.9	68.8
Hong's CrossEntity (2011)	72.9	64.3	68.3
Chen's DMCNN (2015)	75.6	63.6	69.1
Chen's DMCNN+ † (2017)	75.7	66.0	70.5
Liu's FrameNet † (2016a)	77.6	65.2	70.7
Liu's ANN-Aug † (2017)	76.8	67.5	71.9
Li's Structure (2013)	73.7	62.3	67.5
Yang's JointEE (2016)	75.1	63.3	68.7
Nguyen's JRNN (2016)	66.0	73.0	69.3
Liu's PSL (2016b)	75.3	64.4	69.4
Ours HBTNGMA	77.9	69.1	73.3

Method	1/1	1/N	all
LSTM+Softmax	74.7	44.6	66.8
LSTM+CRF	75.1	49.5	68.5
LSTM+TLSTM	76.8	51.2	70.2
LSTM+HTLSTM	77.9	57.3	72.4
LSTM+HTLSTM+Bias	78.4	59.5	73.3

Table 2: Performance of different ED systems. 1/1 means one sentence that only has one event and 1/N means that one sentence has multiple events.

Method	P	R	F_1
Word Only	70.1	63.4	66.6
Word+SA	75.6	68.2	71.7
Word+DA	73.1	65.8	69.3
Word+Average MA	76.5	68.7	72.4
Word+Gated MA	77.9	69.1	73.3

Table 3: Performance of gated multi-level attention.

篇章级金融事件抽取

- 动机：仅利用句子信息很难抽取全事件信息

证券代码：600747股票简称：大连控股编号：临2017-04大连大福控股股份有限公司关于大股东股份冻结的公告本公司董事会及全体董事保证本公告内容不存在任何虚假记载、误导性陈述或者重大遗漏，并对其内容的真实性、准确性和完整性承担个别及连带责任。

公司于近日收到通知，公司第一大股东长富瑞华持有的上市公司520,000股被大连市中级人民法院于2017年5月5日冻结。冻结期限为3年。自转为正式冻结之日起计算。

本次轮候冻结包括孳息（包括派发的送股、转增股及现金红利），其效力从登记在前的冻结证券解除冻结且本次轮候冻结部分或全部生效之日起产生。

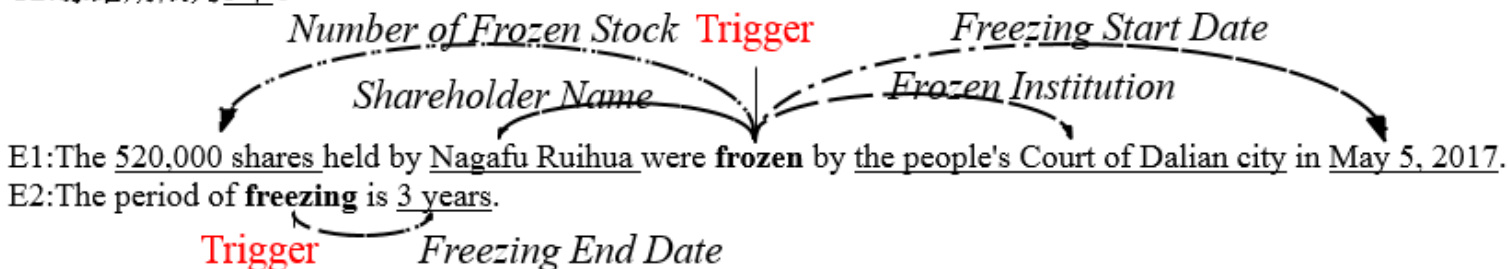
此次轮候冻结股数占公司总股本的35.51%，截止本公告日，长富瑞华持有本公司股份520,000,000股，占公司总股本35.51%；此次股份冻结后累计股份冻结的数量520,000,000股，占公司总股本的35.51%，经公司向大股东长富瑞华了解，此次股份冻结事项不会对公司的控制权造成影响，也不影响公司正常经营，长富瑞华将与相关方积极协商妥善处理解决相关事宜，公司将密切关注该事项的进展并及时履行信息披露义务，特此公告。

大连大福控股股份有限公司董事会二〇一七年一月十三日2



C1:长富瑞华持有的上市公司520,000股被大连市中级人民法院于2017年5月5日冻结。

C2:冻结期限为3年。

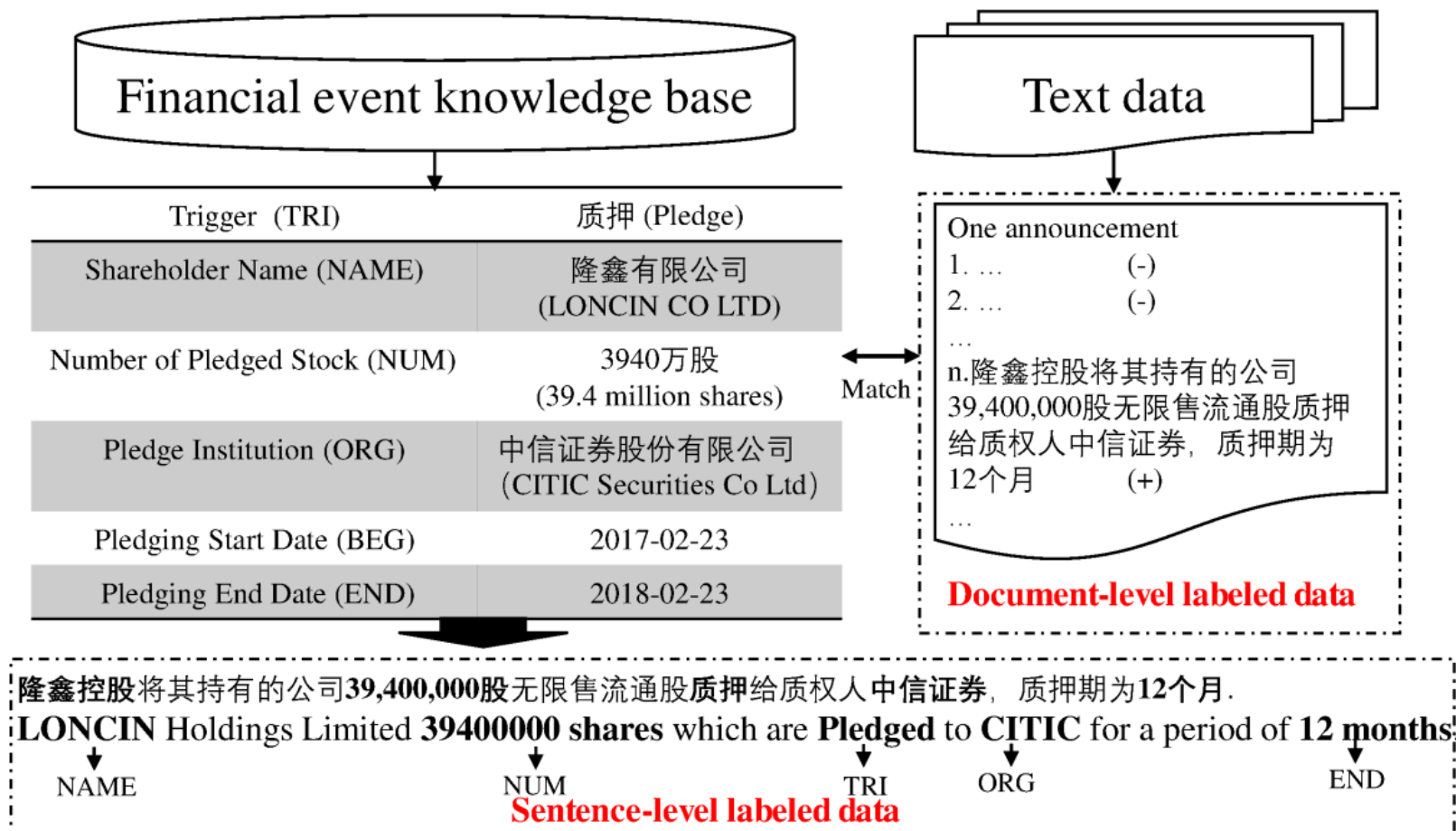


篇章级金融事件抽取

- 挑战
 - 标注数据的缺失
 - 现有的事件抽取系统性能都依赖于人工标注数据
 - 人工标注数据耗时费力，成本高昂，金融领域缺乏大规模高质量的标注数据
 - 篇章级事件抽取
 - 目前大多数的事件抽取系统都是针对一个句子进行抽取
 - 由多个句子描述一个事件，一个事件的多个元素分布在不同的句子中

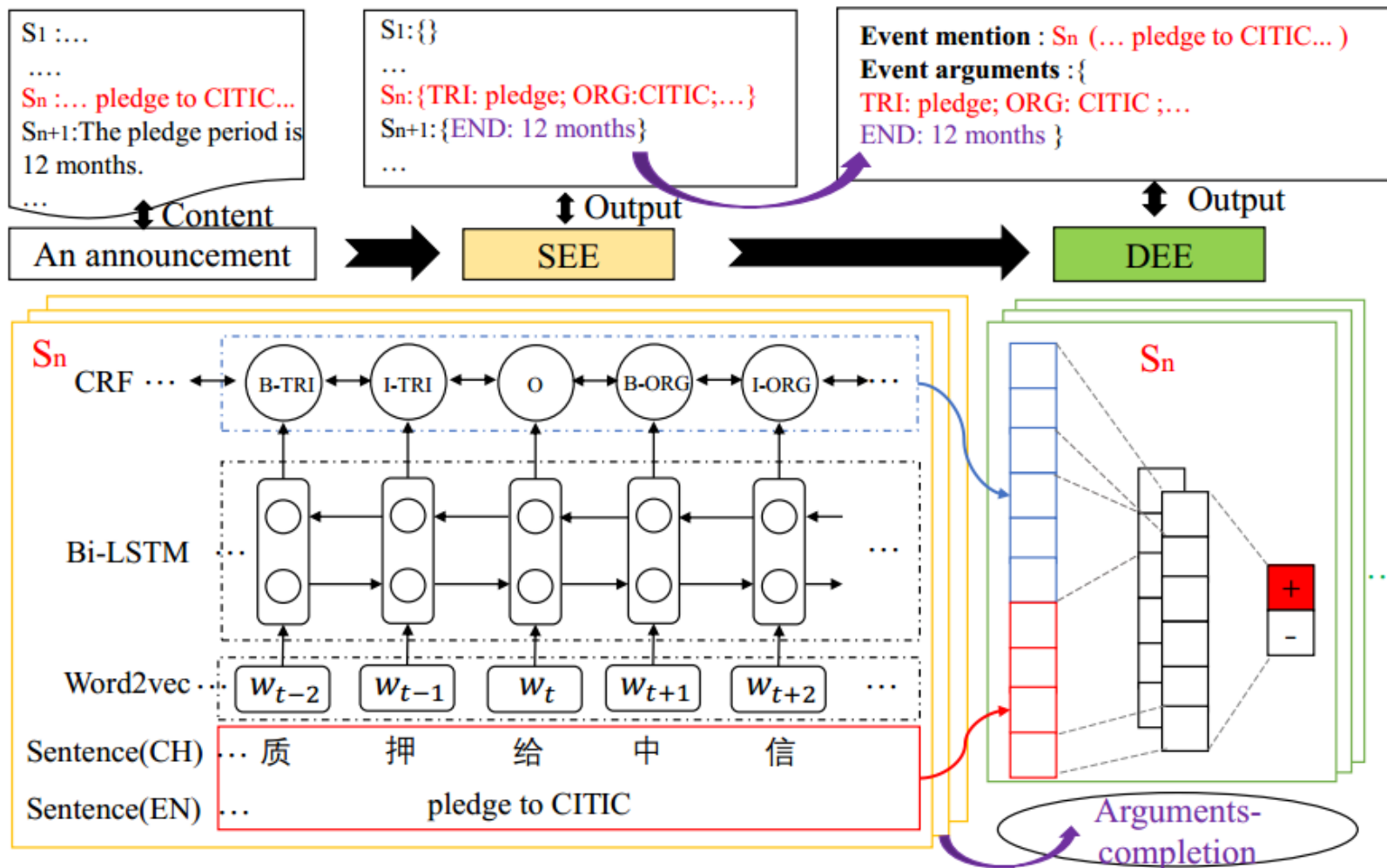
篇章级金融事件抽取

- 自动生成标注数据



篇章级金融事件抽取

- 基于序列标注+元素补齐策略的篇章级事件抽取



篇章级金融事件抽取

- 实验结果

- 数据集

Dataset	NO.BUL	NO.POS	NO.NEG
<i>EF</i>	526	544	2960
<i>EP</i>	752	775	6392
<i>EB</i>	1178	1192	11590
<i>EI</i>	520	533	11994
<i>Total</i>	2976	3044	32936

- 系统性能

Stage	SEE			DEE		
Type	$P(\%)$	$R(\%)$	$F_1(\%)$	$P(\%)$	$R(\%)$	$F_1(\%)$
<i>EF</i>	90.00	90.41	90.21	80.70	63.40	71.01
<i>EP</i>	93.31	94.36	93.84	80.36	65.91	72.30
<i>EB</i>	92.79	93.80	93.29	88.79	82.02	85.26
<i>EI</i>	88.76	91.88	90.25	80.77	45.93	58.56

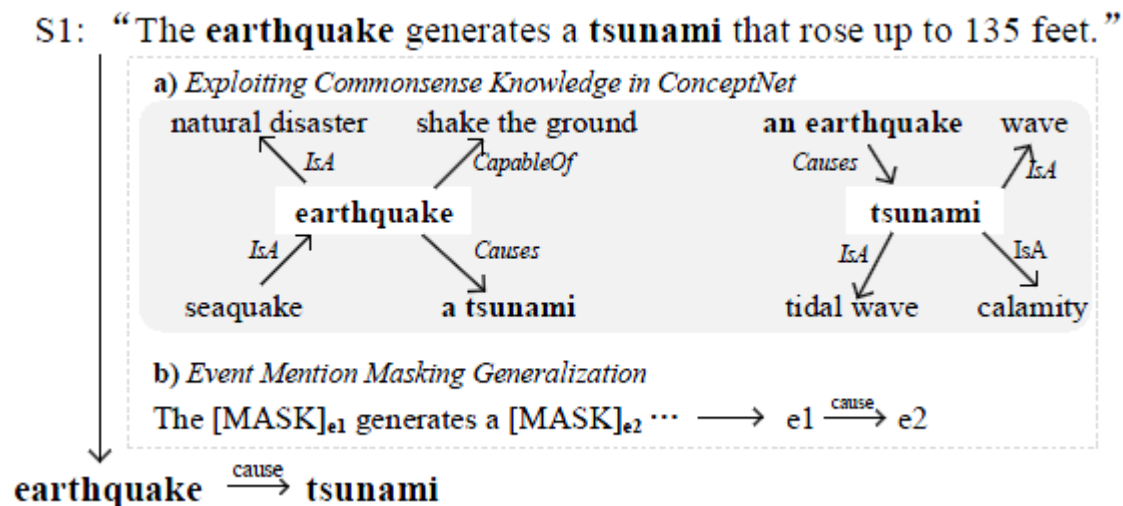
知识层面

- 动机

- 仅利用文本信息很难完成事件因果关系推断

- 知识图谱中存有相关的背景知识

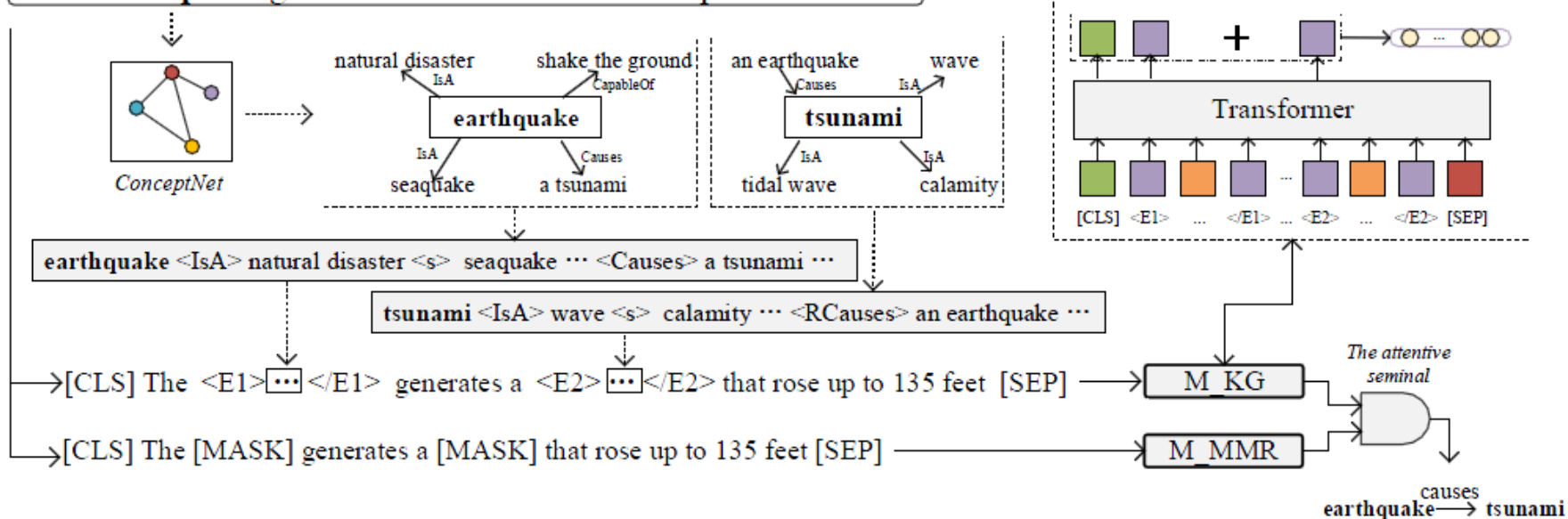
- 因果语义表达规律+背景知识能有效提升因果推断能力



知识增强的事件因果关系识别

• 方法

S1: “The **earthquake** generates a **tsunami** that rose up to 135 feet.”



M_KG: the knowledge aware reasoner

M_MMR: the mention masking reasoner

知识增强的事件因果关系识别

实验

– 数据集: EventStoryLine、Causal-TimeBank、EventCausality

	METHOD	PRE.	REC.	F1
CV	Rule-based [Mirza, 2014b]	36.8	12.3	18.4
	Data-driven [Mirza, 2014a]	67.3	22.6	33.9
	BERT	30.3	41.1	34.9
	M _{KG} (Ours)	38.7	44.4	41.3
	M _{MMR} (Ours)	31.1	51.9	38.8
	M _{FULL} (Ours)	36.6	55.6	44.1*
TE	RB [Mirza and Tonelli, 2016]	91.7	42.3	57.9
	ML [Mirza and Tonelli, 2016]	42.9	11.5	18.2
	HB [Mirza and Tonelli, 2016]	73.7	53.8	62.2
	BERT	56.6	42.3	48.4
	M _{KG} (Ours)	50.0	57.7	53.5
	M _{MMR} (Ours)	52.8	73.1	61.2
	M _{FULL} (Ours)	61.3	73.1	66.7*

Causal-TimeBank

METHODS	PRE.	REC.	F1
OP [Caselli and Vossen, 2017]	22.5	98.6	36.6
LSTM [Cheng and Miyao, 2017]	34.0	41.5	37.4
Seq [Choubey and Huang, 2017]	32.7	44.9	37.8
LR+ [Gao <i>et al.</i> , 2019]	37.0	45.2	40.7
LIP [Gao <i>et al.</i> , 2019]	38.8	52.4	44.6
BERT	37.9	38.5	38.2
M _{KG} (Ours)	44.5	39.3	41.8
M _{MMR} (Ours)	37.6	52.6	43.9
M _{FULL} (Ours)	41.9	62.5	50.1*

EventStoryLine

METHODS	PRE.	REC.	F1
PMI [Do <i>et al.</i> , 2011]	26.6	20.8	23.3
ECD PMI [Do <i>et al.</i> , 2011]	40.9	23.5	29.9
CEA [Do <i>et al.</i> , 2011]	62.2	28.0	38.6
BERT	16.8	30.7	21.7
M _{BERT} (Ours)	17.2	68.2	27.5
M _{MMR} (Ours)	20.7	77.3	32.6
M _{FULL} (Ours)	34.1	68.2	45.4*

EventCausality

金融领域的应用：输入为公司公告文本

请在下方输入你的文字，或点击样例尝试。

右键可取消/隐藏实体。点击名称可以跳转到图谱页面。

证券代码：300295 证券简称：三五网公告编号：2016-019 江苏三五网络股份有限公司关于持股5%以上股东减持计划的公告 本公司及董事会全体成员保证信息披露的内容真实、准确、完整，没有虚假记载、误导性陈述或重大遗漏。特别提示：本公司股东邢炜先生计划在2016年6月10日后的12个月内，以大宗交易或集中竞价方式减持不超过本公司股份989万股（占本公司总股本比例5.15%）。其中，通过集中竞价减持股份的总数连续三个月不能超过公司总股本的1%。江苏三五网络股份有限公司（以下简称：“三五网”或“公司”）于2016年5月10日收到公司股东邢炜先生《关于持股5%以上股东股份减持计划告知函》，现将有关情况公告如下：一、股东情况1、股东姓名：邢炜2、截至本公告日，邢炜先生直接持有公司19,780,240股，占公司总股本的10.30%；上述股票将在2016年6月10日解除限售（董事离职后根据有关规定锁定）。二、本次减持计划的主要内容1、减持目的：个人资金安排。2、减持期间：自2016年6月10日起十二个月内。3、拟减持的数量：不超过989万股，不超过公司总股本的5.15%。若计划减持期间有送股、资本公积金转增股本等股份变动事项，上述股份数量做相应调整。4、减持方式：包括但不限于集中竞价交易、大宗交易等。5、减持价格：根据减持时市场价格确定。三、与股份相关股东承诺履行情况：邢炜先生在公司首次公开发行及上市时承诺为：“自公司股票上市之日起三十六个月内，不转让或者委托他人管理本次发行前已直接或间接持有的发行人股份，也不由发行人回购该部分股份。”邢炜先生在担任公司董事监事、高管时承诺：“在其任职期间每年转让直接或间接持有的公司股份不超过其所持有公司股份总数的百分之二十五；离职后半年内，不转让其直接或间接持有的公司股份。”2015年3月12日，邢炜作为原共同实际控制人，与其他三位共同实际控制人共同承诺：“在限售股解禁后12个月内，四名共同实际控制人及其关联人李东、沈丽除出售及赠予给员工持股计划的股份外，实际减持其直接持有股份数不超过占公司总股本的4.65%；”截至本公告日，邢炜先生已正常完毕履行了上述承诺，不存在违反上述承诺的行为。四、其他事项1、在按照该计划减持股份期间，邢炜先生将严格遵守《深圳证券交易所创业板股票上市规则》、《深圳证券交易所创业板上市公司规范运作指引》等有关法律法规及公司规章制度。2、邢炜先生已不是公司共同实际控制人，本减持计划不会对公司治理结构、股权结构及持续经营产生重大影响，公司基本面未发生重大变化。敬请广大投资者理性投资。本次减持计划如完全实施后，邢炜先生仍是公司持股5%以上股东。五、备查文件1、邢炜先生的《关于持股5%以上股东股份减持计划告知函》。特此公告！江苏三五网络股份有限公司董事会2016年5月11日

识别

样例1

金融领域的应用：逐句分析结果

概览	文本
全部分析结果	返回
🔍 增减持事件 1	
⚙️ 事件触发词 15	
📈 增减持股数 3	
▶ 计划起始日 3	
■ 计划截止日 2	
📊 增减持目的 1	
👤 股东名称 8	
	<p>证券代码：300295证券简称：三六五网公告编号：2016-019江苏三六五网络股份有限公司关于持股5%以上股东减持计划的公告本公司及董事会全体成员保证信息披露的内容真实、准确、完整，没有虚假记载、误导性陈述或重大遗漏。 特别提示：本公司股东邢炜先生计划在2016年6月10日后的12个月内,以大宗交易或集中竞价方式减持不超过本公司股份989万股（占本公司总股本比例5.15%）。 其中，通过集中竞价减持股份的总数连续三个月不能超过公司总股本的1%。江苏三六五网络股份有限公司（以下简称：“三六五网”或“公司”）于2016年5月10日收到公司股东邢炜先生《关于持股5%以上股东股份减持计划告知函》，现将有关情况公告如下：一、股东情况1、股东姓名：邢炜 2、截至本公告日，邢炜先生直接持有公司19,780,240股,占公司总股本的10.30%；上述股票将在2016年6月10日解除限售(董事离职后根据有关规定锁定)。二、本次减持计划的主要内容1、减持目的：个人资金安排。2、减持期间：自2016年6月10日起十二个月内。3、拟减持的数量：不超过989万股，不超过公司总股本的5.15%。若计划减持期间有送股、资本公积金转增股本等股份变动事项，上述股份数量做相应调整。4、减持方式：包括但不限于集中竞价交易、大宗交易等。5、减持价格：根据减持时市场价格确定。三、与股份相关股东承诺履行情况：邢炜先生在公司首次公开发行及上市时承诺为：“自</p>

金融领域的应用：篇章分析结果（单句）

概览	文本
<div>全部分析结果</div> <div>增减持事件1</div>	<div>返回</div>
<div>事件:1</div>	
<div>事件触发词1</div>	
<div>增减持股数1</div>	
<div>计划起始日1</div>	
<div>计划截止日1</div>	
<div>增减持目的0</div>	
<div>股东名称1</div>	
	<p>证券代码：300295证券简称：三六五网公告编号：2016-019江苏三六五网络股份有限公司关于持股5%以上股东减持计划的公告本公司及董事会全体成员保证信息披露的内容真实、准确、完整，没有虚假记载、误导性陈述或重大遗漏。特别提示：本公司股东邢炜先生计划在2016年6月10日后的12个月内,以大宗交易或集中竞价方式减持不超过本公司股份989万股（占本公司总股本比例5.15%）。其中，通过集中竞价减持股份的总数连续三个月不能超过公司总股本的1%。江苏三六五网络股份有限公司（以下简称：“三六五网”或“公司”）于2016年5月10日收到公司股东邢炜先生《关于持股5%以上股东股份减持计划告知函》，现将有关情况公告如下：一、股东情况1、股东姓名：邢炜2、截至本公告日，邢炜先生直接持有公司19,780,240股,占公司总股本的10.30%；上述股票将在2016年6月10日解除限售(董事离职后根据有关规定锁定)。二、本次减持计划的主要内容1、减持目的：个人资金安排。2、减持期间：自2016年6月10日起十二个月内。3、拟减持的数量：不超过989万股，不超过公司总股本的5.15%。若计划减持期间有送股、资本公积金转增股本等股份变动事项，上述股份数量做相应调整。4、减持方式：包括但不限于集中竞价交易、大宗交易等。5、减持价格：根据减持时市场价格确定。三、与股份相关股东承诺履行情况：邢炜先生在公司首次公开发行及上市时承诺为：“自</p>

金融领域的应用：篇章分析结果（多句）

概览

[全部分析结果](#)

🔍 冻结事件	1
⚙️ 事件触发词	1
📈 冻结股数	1
▶ 冻结起始日期	1
■ 冻结截止日期	1
🏛️ 执行冻结机构	1
👤 被冻结股东名称	1

文本

[返回](#)

长久物流：关于控股股东股份冻结的公告，“原始公告”：证券代码：603569证券简称：长久物流公告编号：2016-【002】北京长久物流股份有限公司关于控股股东股份冻结的公告本公司董事会及全体董事保证本公告内容不存在任何虚假记载、误导性陈述或者重大遗漏，并对其内容的真实性、准确性和完整性承担个别及连带责任。北京长久物流股份有限公司（以下简称“长久物流”或“公司”）于近日收到上海证券交易所通知，公司控股股东吉林省长久实业集团有限公司（以下简称“长久集团”）持有公司的部分股份被冻结，具体情况如下：一、股份冻结的具体情况因长久集团与中国光大银行股份有限公司长春分行（以下简称“光大长春”）及第三人大理鸿迪汽车销售服务有限公司（以下简称“大理鸿迪”）合同纠纷一案，光大长春向吉林省长春市中级人民法院（以下简称“长春中院”）申请冻结长久集团在银行的存款人民币4,307万元或查封、扣押其相应价值的财产。长春中院根据其作出的（2016）吉-01民初字第682-1号《民事裁定书》和《协助执行通知书》，对长久集团持有的公司4,307万股股份（限售流通股）进行司法冻结。此次冻结包括孳息（包括派发的送股、转增股利及现金股利），冻结期限为两年，从2016年8月12日至2018年8月11日止。截止本公告出具日，长久集团持有公司30,463.618万股限售股，占公司总股本的76.16%。本次冻结股份后，长久集团所持公司股份冻结数量变为4,307万股，占其所持股份的14.48%，占公司总股本的14.48%。

金融领域的应用

数据统计

公司

23637

股票代码
证券名称
工商登记号
法人代表
注册资本

.....

事件

269100

股权冻结
股权质押
关联交易
担保
增减持计划

.....

重要高管

188723

董事长
总经理
监事长
前十大股东
法人代表

.....

学术评测

• 面向金融领域的事件抽取系列评测

- 2019年: https://www.biendata.com/competition/ccks_2019_4/
- 2020年: https://www.biendata.xyz/competition/ccks_2020_4_1/

CCKS & 蚂蚁金服 & 中科院自动化所 • ¥15,000 • 488 支队伍 • 1176 名参赛者

CCKS 2019 面向金融领域的事件主体抽取

组队截止时间 2019-07-27
结束时间 2019-08-02

开始时间 2019-05-01

主页 > 比赛 > CCKS 2019 面向金融领域的事件主体抽取

任务描述

“事件识别”是舆情监控领域和金融领域的重要任务之一，“事件”在金融领域是投资分析，资产管理的重要决策参考。“事件识别”的复杂性在于事件类型和事件主体的判断，比如“公司A产品出现添加剂，其下属子公司B和公司C遭到了调查”，对于“产品出现问题”事件类型，该句中事件主体是“公司A”，而不是“公司B”或“公司C”，我们称发生特定事件类型的主体成为事件主体，本任务中事件主体范围限定为：公司和机构。事件类型范围确定为：产品出现问题、高管减持、违法违规...

本次评测任务的主要目标是从真实的新闻语料中，抽取特定事件类型的主体。即给定一段文本T，和文本所属的事件类型S，从文本T中抽取确定事件类型S的事件主体。

CCKS & 蚂蚁金服 & 中科院自动化所 • ¥20,000 • 614 支单人队伍 • 32 支多人队伍 • 702 名参赛者

CCKS 2020 : 面向金融领域的篇章级事件主体与要素抽取 (一) 事件主体抽取

组队截止时间 2020-09-27
结束时间 2020-09-30

开始时间 2020-03-20

主页 > 比赛 > CCKS 2020 : 面向金融领域的篇章级事件主体与要素抽取 (一) 事件主体抽取

2020/05/15 通知

根据最新安排，比赛时间调整如下：
评测任务发布：3月20日
报名时间：3月20日—9月20日
训练及验证数据发布：3月20日
验证集提交开放：3月20日 23:59:59 - 9月27日 23:59:58 (北京时间)
测试数据发布：9月27日 23:59 (北京时间)
测试集提交开放：9月27日 23:59:59 - 9月30日 23:59:59 (北京时间)
评测论文提交：10月25日
CCKS2020会议日期：11月12日—15日

任务描述

“事件抽取”是舆情监控领域和金融领域的重要任务之一，“事件”在金融领域是投资分析，资产管理的重要决策参考；事件也是知识图谱的重要组成部分，事件抽取是进行图谱推理、事件分析的必要过程。“事件抽取”的挑战体现在文本的复杂和任务的复杂。文本的

总结

- 事件知识不可或缺
 - 企业信息监控
 - 风险信用控制
 - 智能投顾
- 通用领域的事件抽取很难
 - 大规模、高质量的训练数据
 - 鲁棒的特征表示
- 限定域的事件抽取有可能取得不错的性能
 - 文本类型受限
 - 语言表示规律性较强，知识密集
- 未来工作
 - 复杂场景下的事件抽取：篇章级事件抽取、增量事件抽取
 - 人机协同式事件关系抽取



中国科学院自动化研究所
INSTITUTE OF AUTOMATION
CHINESE ACADEMY OF SCIENCES

感谢各位专家!