



基于预训练模型的金融事件分析及应用

PLM based Financial Event Analysis and its Applications

马永亮 澜舟科技搜索与金融NLP技术负责人

DataFunSummit2022 智慧金融在线峰会

1

金融事件分析的
主要任务

2

金融事件分析
技术

3

总结和展望

1

金融事件分析的 主要任务

任务定义

非结构化数据智能解析

新闻资讯 公司公告
研究报告 政策文件



事件语义理解

事件检测识别
事件要素抽取 事件关系抽取



事件图谱分析

事件链分析 事件预测

核心技术

金融事件体系：按照公司、行业、宏观、股票、债券、基金等不同主体，建立不同的事件体系

- 场景化：细分场景事件体系支持有针对性的分析
- 完整性：完整事件类型和要素支持场景的深度分析
- 可扩展：未知事件类型发现和聚类泛化，事件体系的无监督归纳学习

金融事件图谱：抽取事件的类型、主体、要素和关系，结合事件主体的属性和关系，以及场景化的事件知识，实现事件图谱推理分析

- 事件检测和要素抽取：事件类型+事件触发词+要素
- 事件关系：远程监督+无监督学习
- 事件表示：结合事件上下文和关系的事件表示学习
- 事件预测：基于事件链模型的事件预测

分析目标

什么企业

发生了什么事

事件的要素

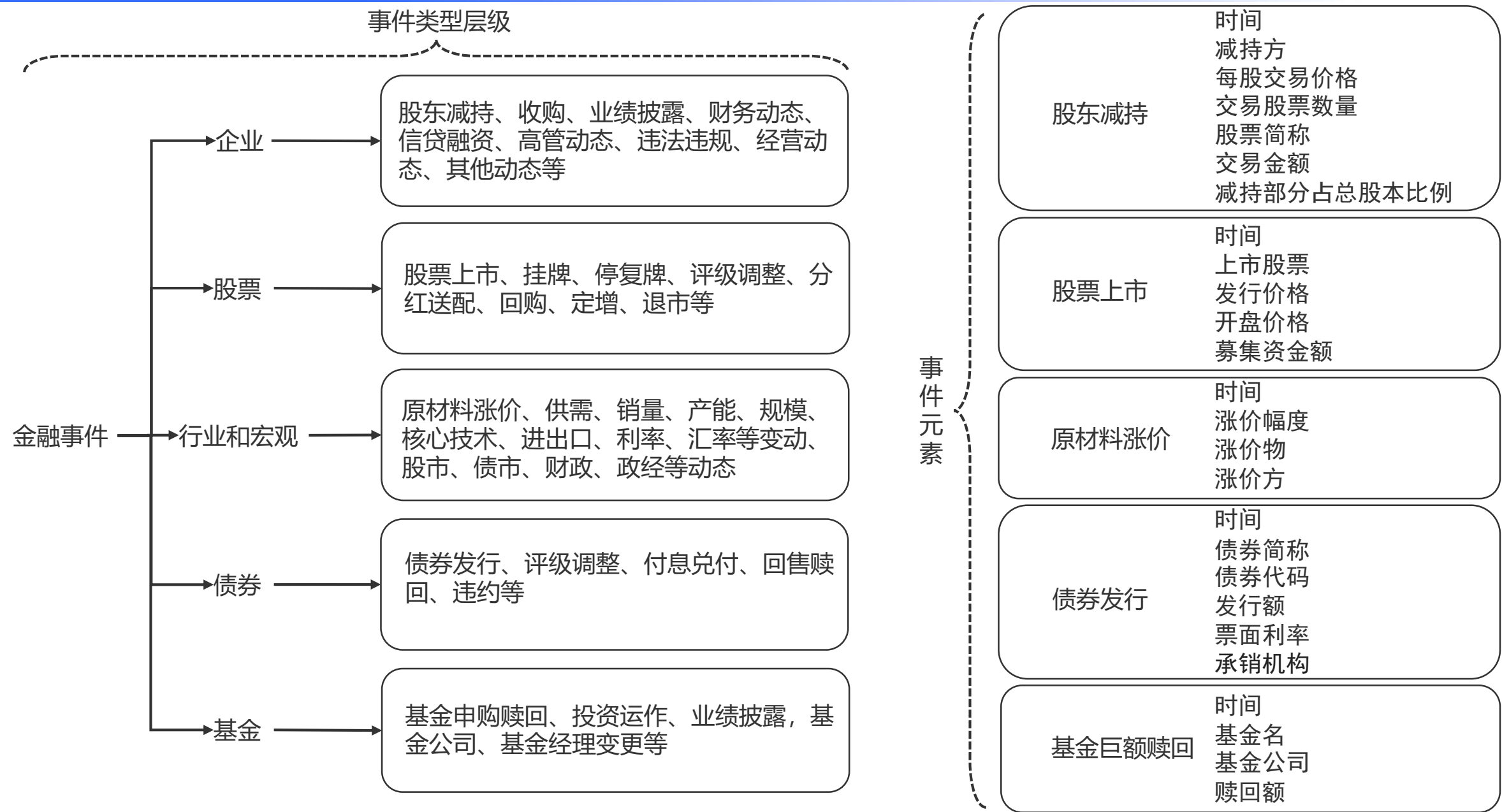
事件的情感

事件的关系

事件的影响

为什么发生

将导致什么事发生

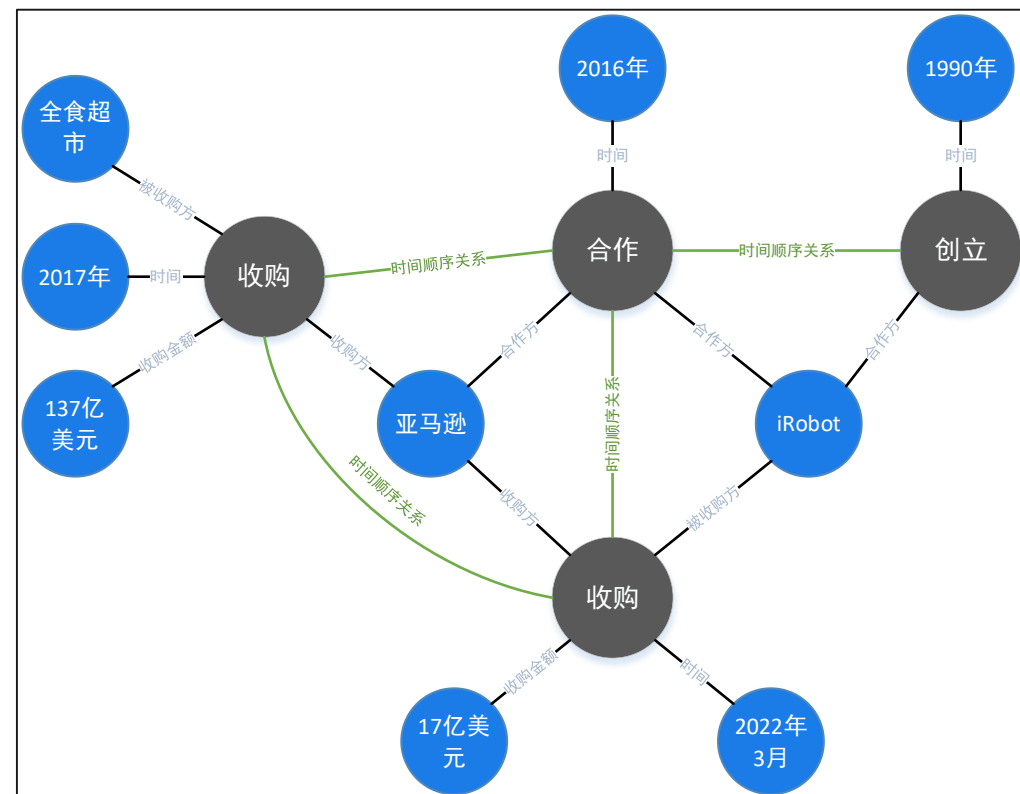


事件图谱是将事件和事件元素(实体)作为节点，将节点用表示他们之间关系的边连接起来而成的图。

- 节点：事件，事件元素(实体)
- 边：事件-事件的关系，事件-实体的关系，实体-实体的关系

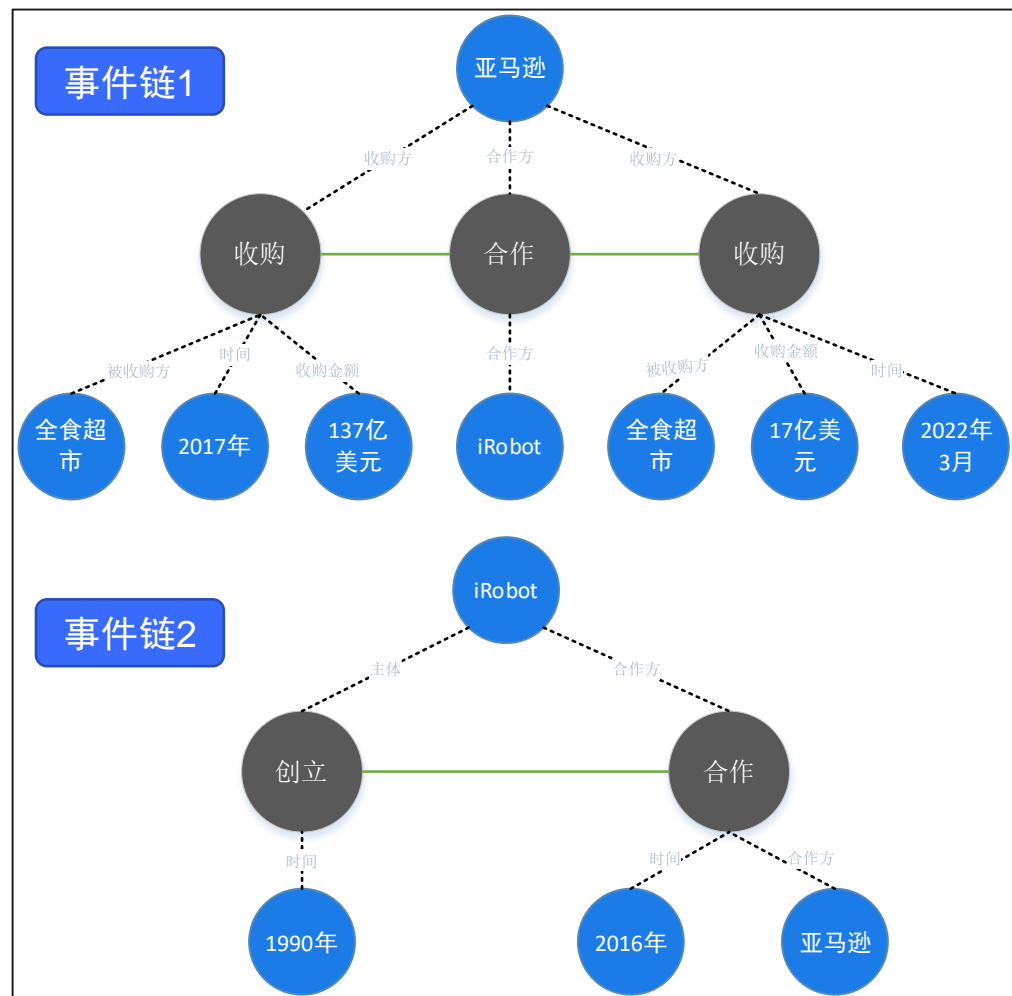
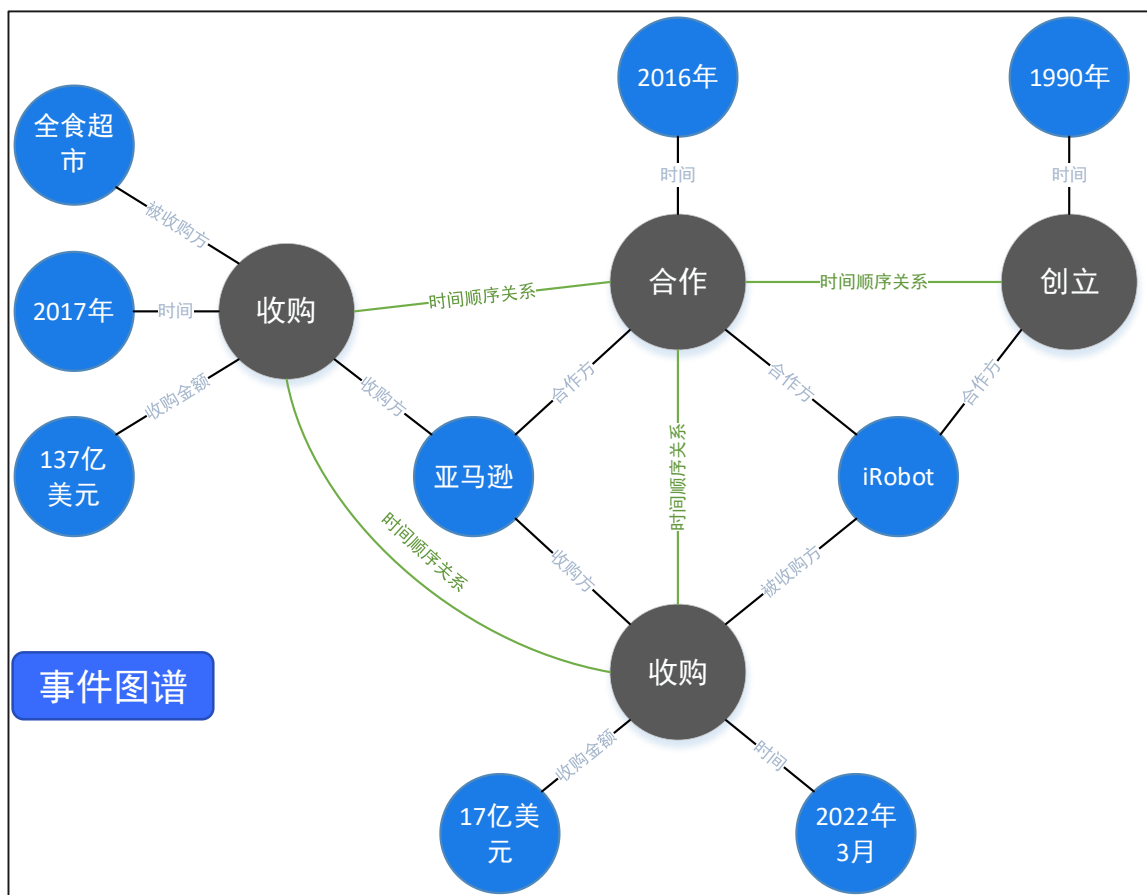
事件图谱的应用：

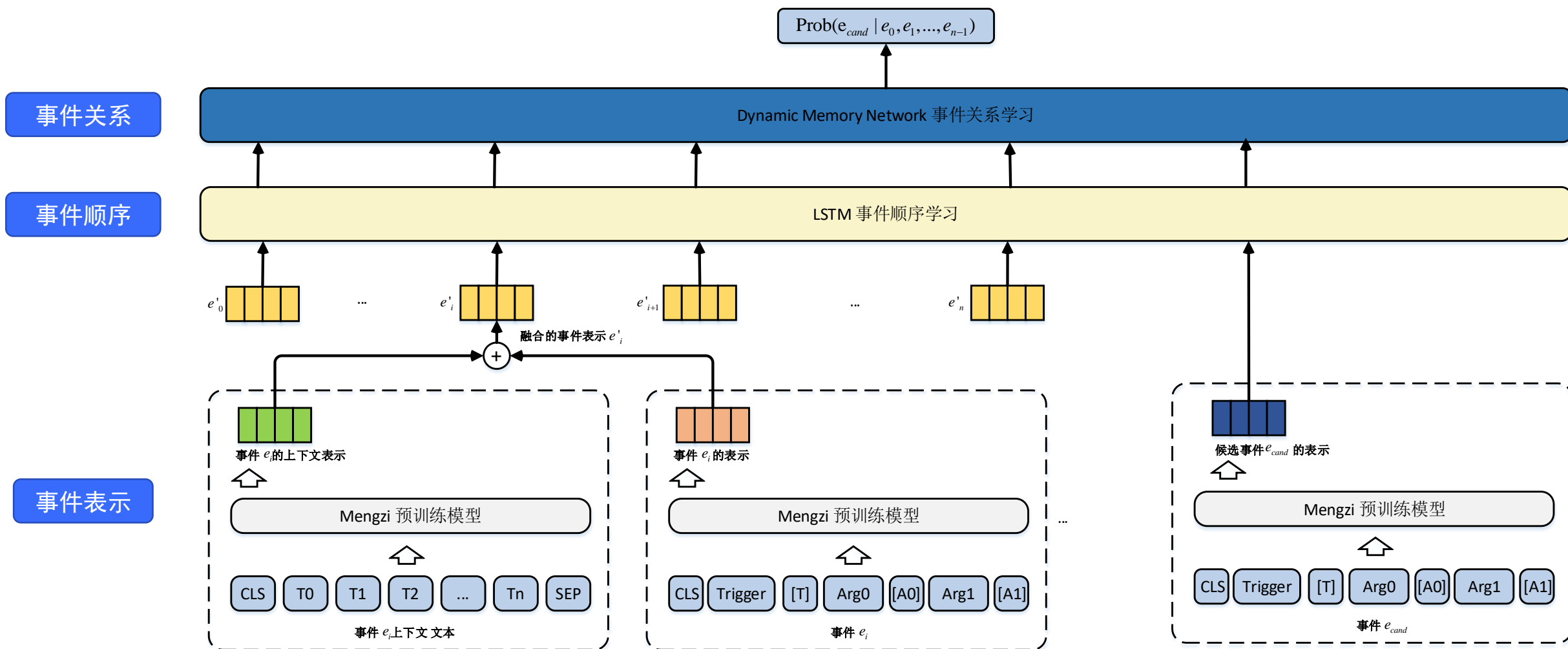
- 搜索
- 问答
- 推荐
- 事件文本生成
- 事件图谱预测
- 量化投资分析
- 风险事件监控
- 知识管理



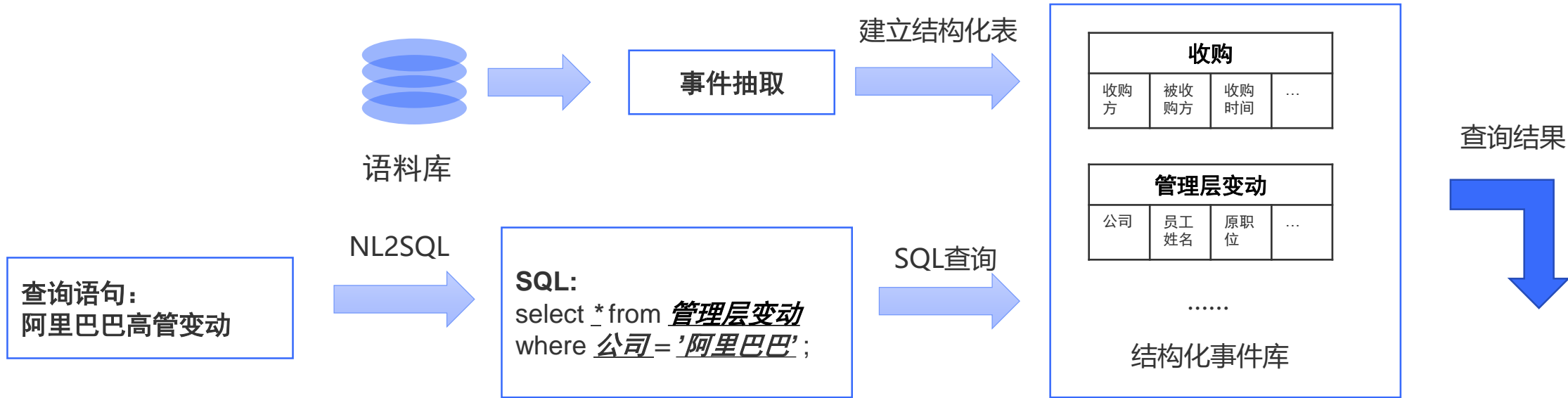
事件链是一个事件的偏序集合，并且其中的每个事件有一个**主要参与者**

- 事件图谱的复杂性：关系种类多-事件间，实体间，事件和实体间，关系涉及的节点类型多-实体，事件。
- 事件链是一种特殊的事件图谱，事件链上的事件表示、事件链建模和预测为图谱的应用提供了一种更有效的形式。





基于孟子预训练的事件链模型

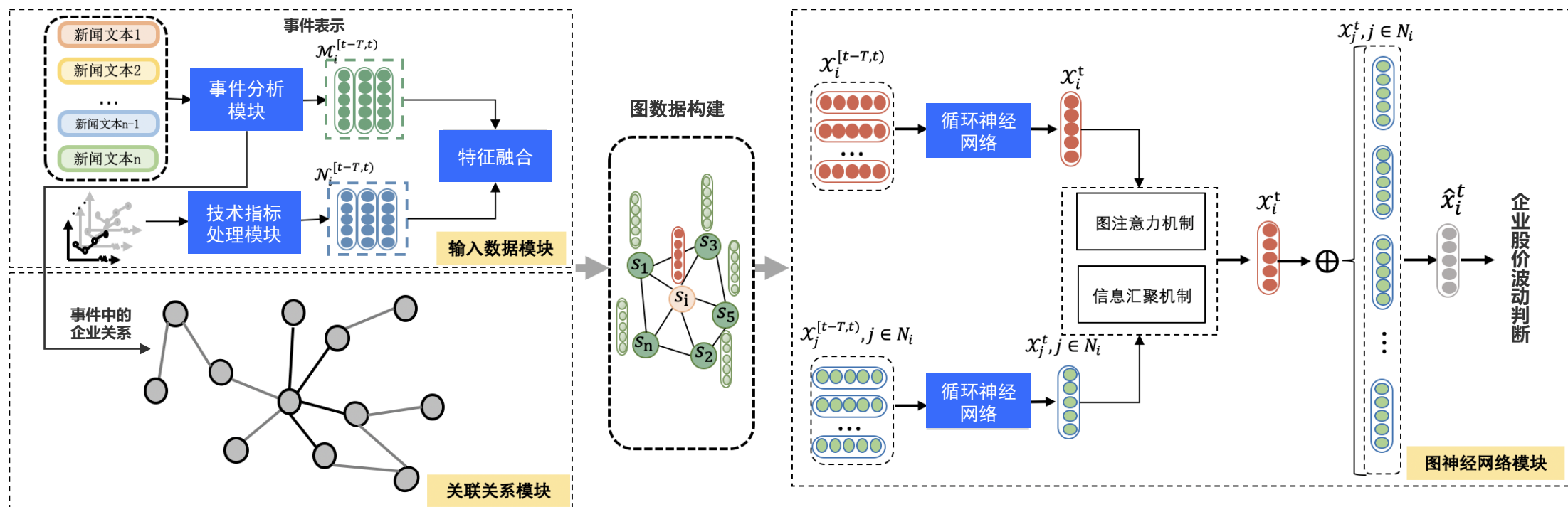


1. 支持多种事件类型、全要素的自然语言搜索, 如近期的公司收购事件、某公司历史上的高管变动等等
2. 结构化数据搜索较快且较为精确

公司	员工姓名	原职位	新职位	时间
阿里巴巴	利蕴莲	希慎兴业集团有限公司执行主席	独立董事	2022-08-04	...
阿里巴巴	吴港平	安永会计师事务所中国前主席	独立董事	2022-08-04	...
阿里巴巴	华先胜	阿里巴巴集团副总裁、达摩院城市大脑实验室负责人	离职	2022-05-22	...
阿里巴巴	肖利华	阿里巴巴集团副总裁、阿里云研究院院长	离职	2022-05-22	...
...

基于事件分析特征的图神经网络模型

- 抽取媒体新闻事件类型和表示，与数值型金融时序数据融合，作为图神经网络模型输入
- 抽取企业在事件中的关系和企业图谱关系（产业链等），利用图神经网络模型捕捉企业之间的动量溢出影响



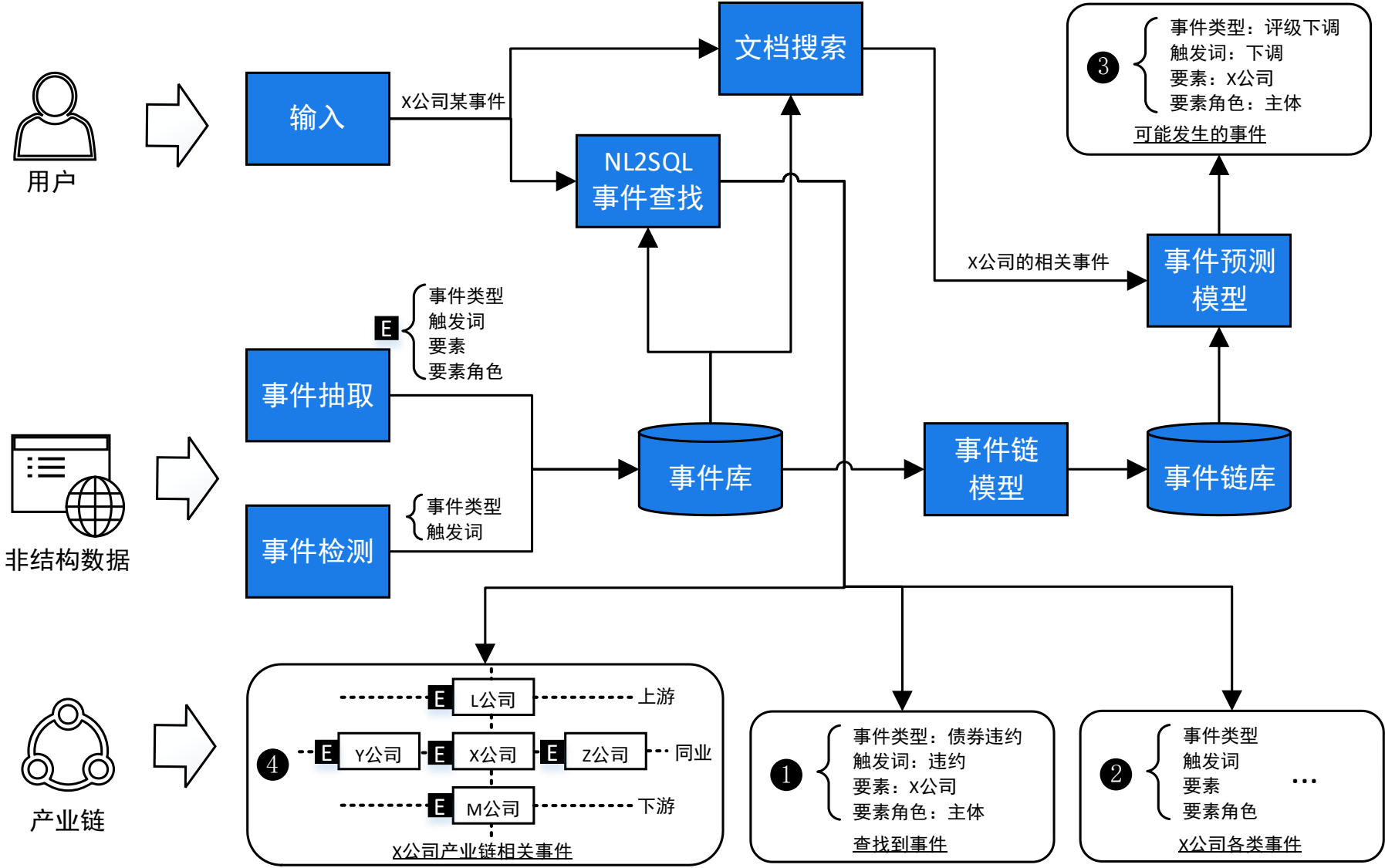
基于事件分析特征的图神经网络模型

事件分析框架落地4大能力：

- 搜索某个事件
- 搜索某主体的各类事件
- 结合产业链搜索某公司
同行业、上下游行业相
关公司的各类事件
- 根据某公司的搜索结果
预测未来可能发生的事
件

技术优势：

- 基于预训练模型，有迁
移能力，zero-shot/few-
shot能力，冷启动快
- 体系化方法，能力全面，
支持丰富场景



2

金融事件分析 技术

任务定义

- 事件检测 (Event Detection, ED) 的目标是从文本中检测发生的事件并对它们进行分类。通常的做法需要抽取事件触发词，再对触发词进行分类，从而得到事件分类。例如，

“公司副董事长、副总经理黄世霖因个人事业考虑将辞去公司副董事长、董事、董事会战略委员会委员和副总经理职务。”



事件检测 结果	事件类型：企业-高管变动
	事件触发词：“辞去”

挑战

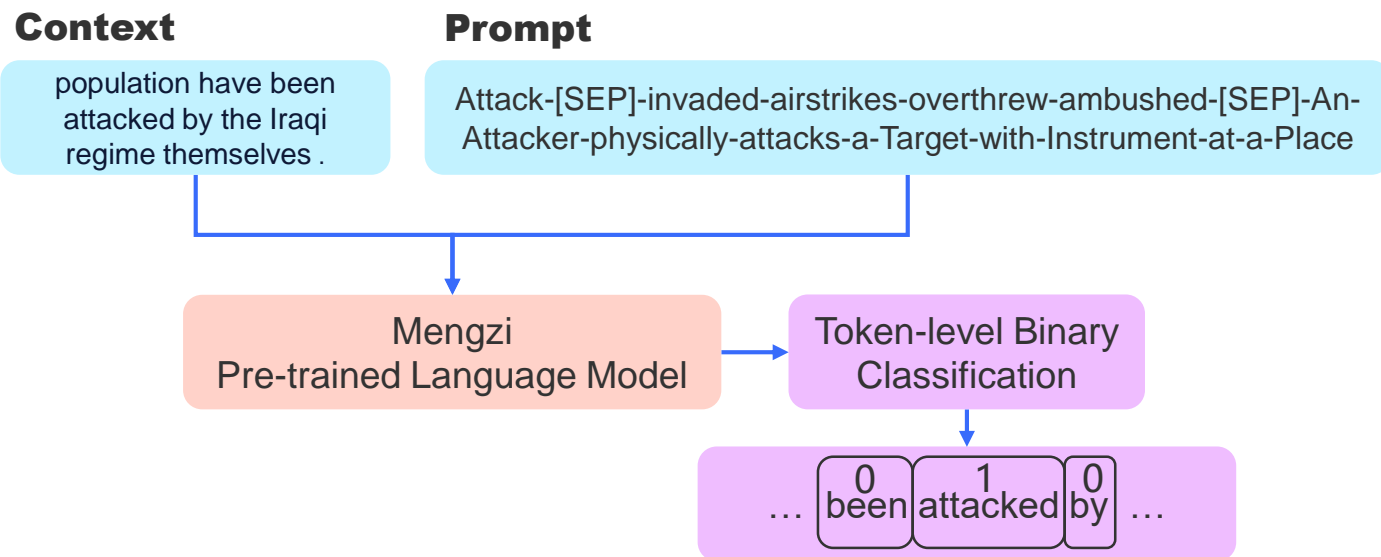
- 新的事件类型层出不穷，在没有标注数据或者只有少量标注数据的情况下进行事件检测

方法

- 基于提示词和预训练模型的事件检测

触发词检测

- Prompt提示由四部分信息组成：事件类型，事件触发词，事件定义和事件要素的角色
- 特定于事件的提示既可以是事件的详细定义，也可以是一些事件相关的关键词，两种都能提供事件知识
- 模型的输出是BIO标签，触发词的位置可以通过转换BIO标签得到



基于预训练的Prompt-based事件检测模型+触发词检测

Model	Full	5-way 0-shot	5-way 1-shot	5-way 5-shot	5-way 10-shot
ONEIE	74.7	-	4.2	38.5	53.3
DEGREE	72.2	46.8	53.3	55.5	61.2
Ours	78.8	48.6	56.3	64.0	63.3

ACE05数据集上的zero/few-shot事件检测结果，评价指标为Micro-F1 (%)

任务定义

- 事件抽取 (Event Extraction, EE) 的目标除了简单地识别出事件类型和触发词之外, 还需要提取出事件中具有特定角色的实体-事件元素, 例如

“格力电器9月9日宣布回购完成, 累计成交金额为150亿元, 其中最高成交价为56.11元/股, 最低成交价为40.21元/股”



事件抽取 结果	事件类型: 企业行为-股份回购
	事件触发词: “回购”
	时间: “9月9日”
	交易金额: “150亿元”
	最高成交价: 56.11元/股
	最低成交价: 40.21元/股

挑战

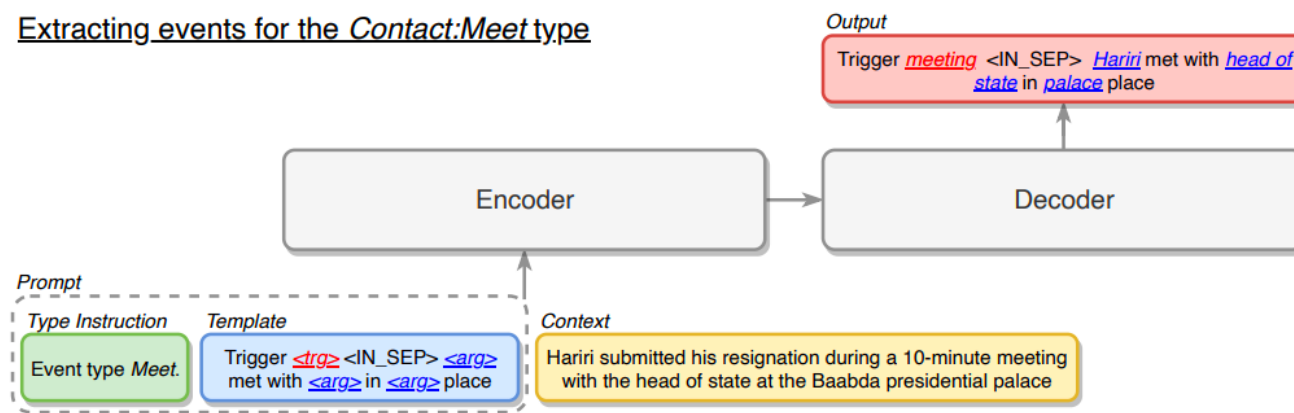
- 事件抽取需要在句子级进行序列标注, 要素的标注经常需要领域专家来完成, 成本较高, 难以扩展到数量较多的事件类型上

方法

- 基于生成式的方法在标注数据更少的场景下也能达到较为不错的模型性能, 具有更高的数据利用率和灵活性, 更容易扩展到新的事件类型

基于模板的生成式事件抽取 (Generative Template-based Event Extraction, GTEE)

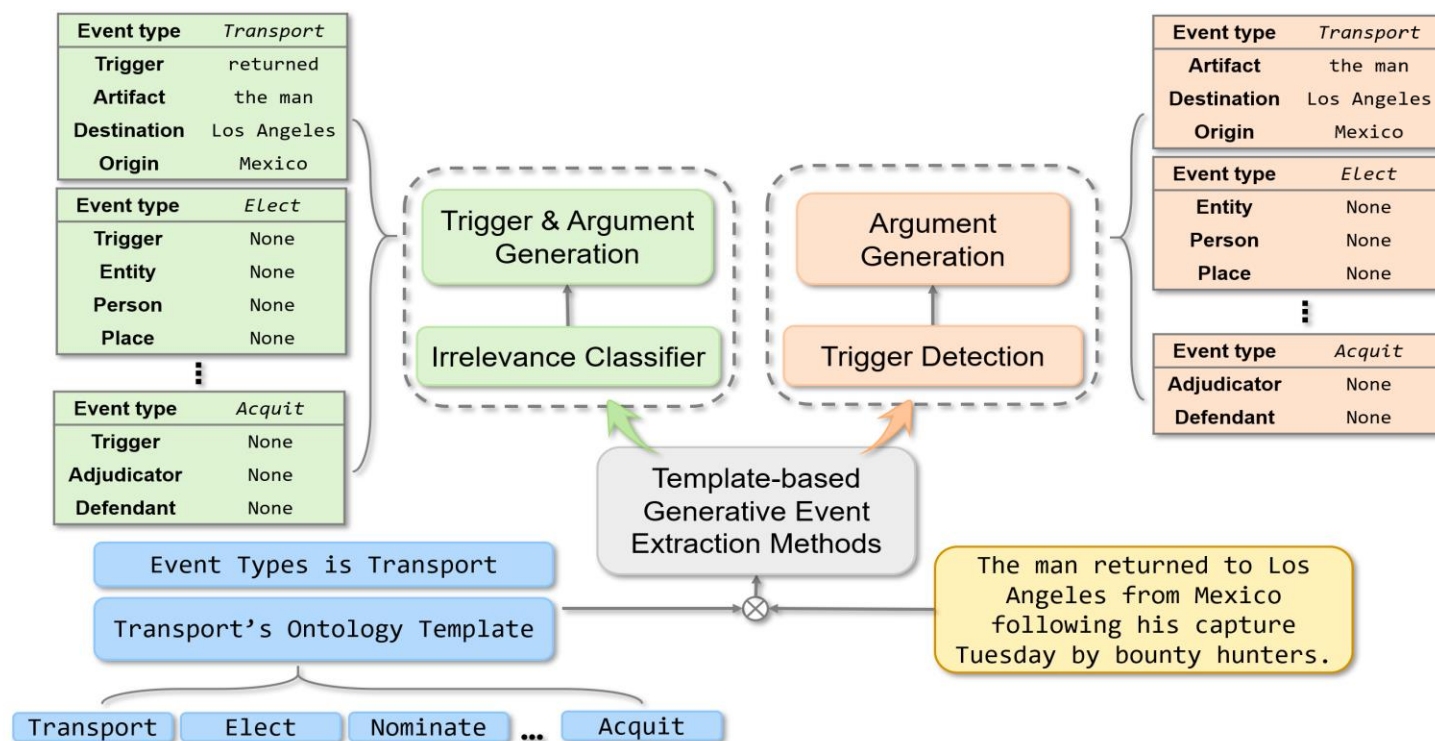
- GTEE将事件抽取任务转换为一个带有模板的端到端的条件生成任务，该任务允许模型以类似完型填空的方式提取事件要素
- GTEE使用提示调优 (Prompt tuning) 的训练方式，以适应少样本 (Few-shot) ，甚至零样本 (Zero-shot) 的场景。提示 (Prompt) 包含了事件类型指示和特定事件的模板
- GTEE以端到端方式同时学习触发词和事件要素，这鼓励模型更好地利用它们之间的共享知识和依赖关系



基于模板的生成式事件抽取模型 GTEE

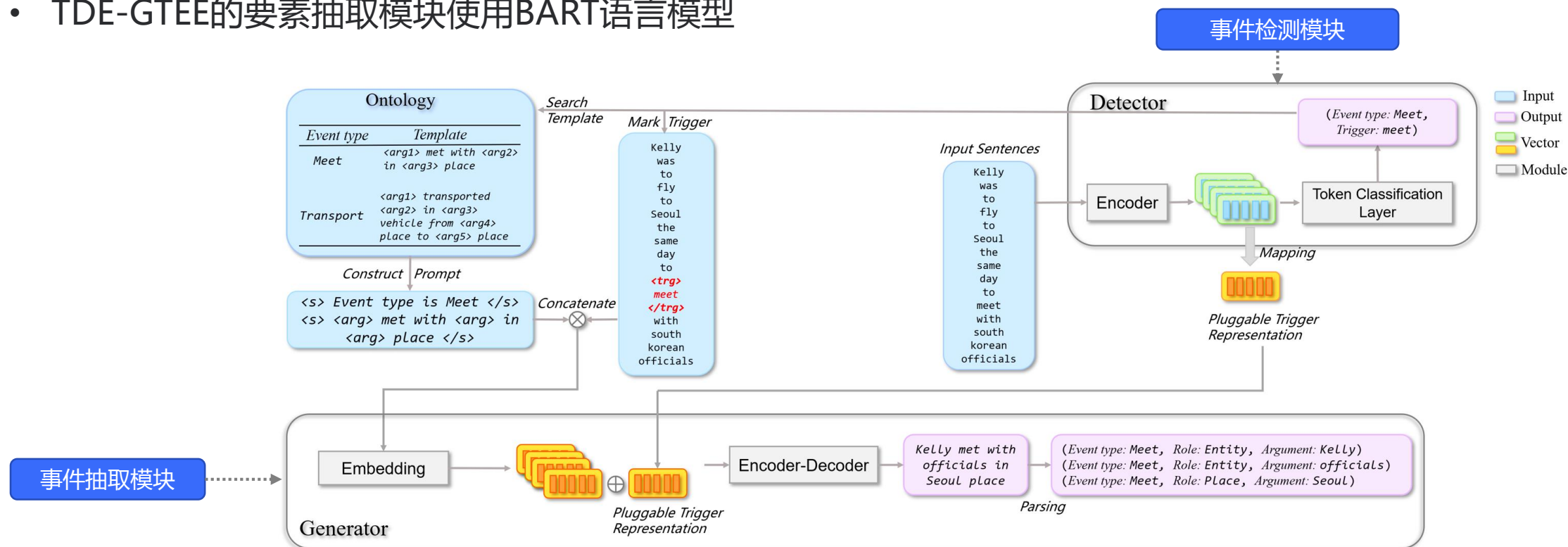
Generative Event Extraction的不足

- 通过模板建立的触发词和事件元素之间的联系较弱
- 推理时需要轮询所有的事件类型，增加大量的推理时间
- 事件检测和事件要素抽取模块采用Pipeline组织，存在错误传播



基于触发词检测增强的生成式事件抽取 (Trigger Detection Enhanced-GTEE, TDE-GTEE)

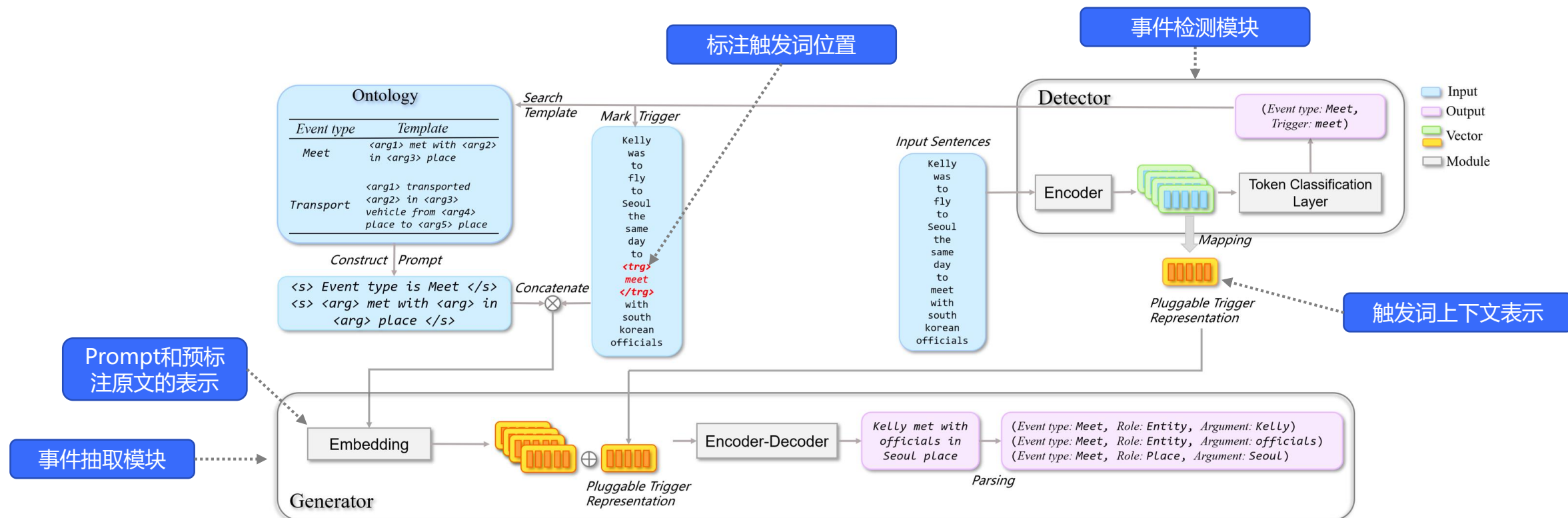
- TDE-GTEE将事件检测与基于生成的事件抽取相结合，通过增强触发词的抽取效果来增强事件要素的抽取效果
- TDE-GTEE的事件检测模块使用BERT语言模型，并将触发词检测视为词级的多分类任务，针对每个词同时给出“是否为触发词”以及“事件类型”信息
- TDE-GTEE的要素抽取模块使用BART语言模型



基于触发词检测增强的生成式事件抽取模型

基于触发词检测增强的生成式事件抽取 (TDE-GTEE)

- 事件检测模型提供触发词的位置和事件类型
- 辅助过滤无关事件类型
- 为了增强触发词和论元之间的联系，在原输入中重新显示标注触发词位置，并将触发词上下文表示作为可插拔的知识加入生成模型



基于触发词检测增强的生成式事件抽取模型

基于触发词检测增强的生成式事件抽取 (TDE-GTEE)

- 为了克服Pipeline的误差传播，事件检测模块和事件要素抽取模块组成一个端到端的模型，并联合训练
- 在公开数据集ACE和ERE上，达到SOTA水平
- ACE上的消融实验显示了两种提示的有效性

Model	ACE05-E		ACE05-E+		ERE-EN	
	Trg-C	Arg-C	Trg-C	Arg-C	Trg-C	Arg-C
<i>classification-based</i>						
ONEIE (Lin et al., 2020)	74.7	56.8	72.8	54.8	59.1	50.5
BERT_QA (Du and Cardie, 2020)	72.3	53.3	-	-	-	-
MQAEE (Li et al., 2020)	71.7	53.4	-	-	-	-
<i>generation-based</i>						
TEXT2EVENT (Lu et al., 2021)	69.2	49.8	71.8	54.4	59.4	48.3
BART-GEN (Li et al., 2021b)	71.1	53.7	-	-	-	-
DEGREE-E2E (Hsu et al., 2021)	70.9	54.4	72.7	55.0	57.1	49.6
GTEE-DYNPREF (Liu et al., 2022)	72.6	55.8	74.3	54.7	66.9	55.1
Our method	78.8	57.4	77.1	57.7	65.1	52.7

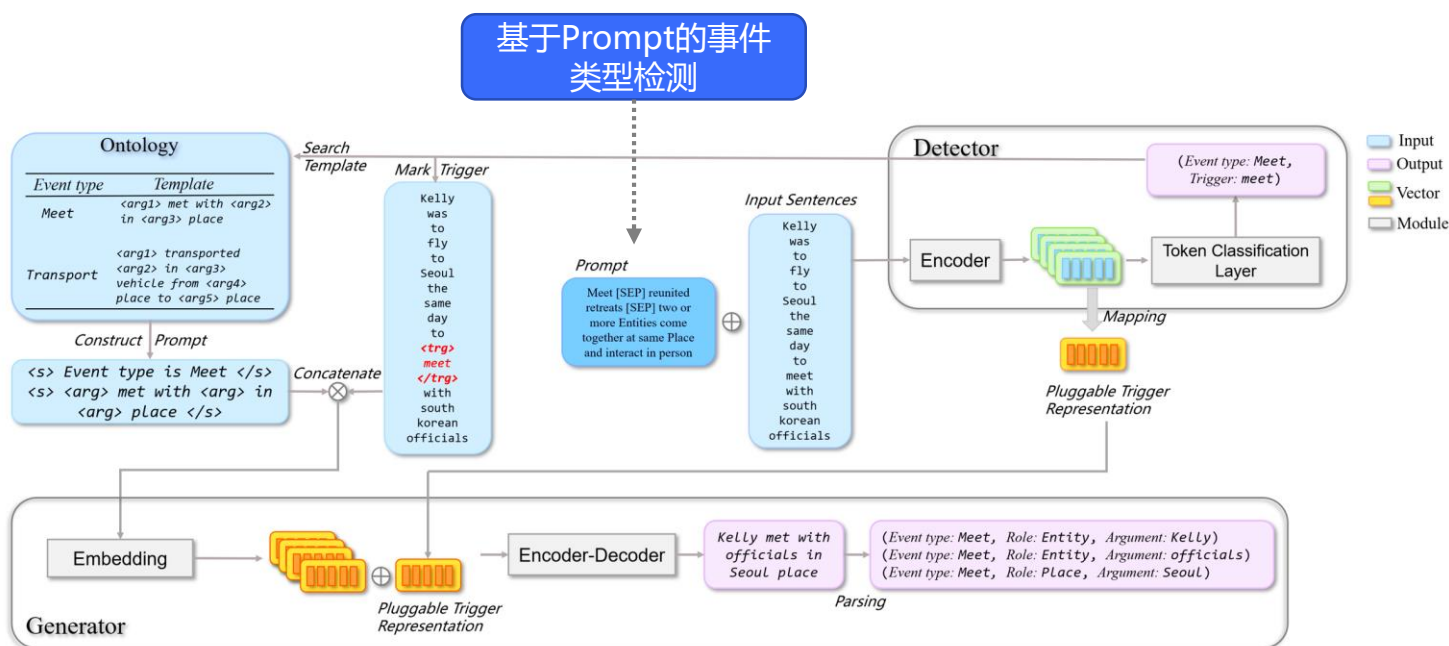
Results on ACE05-E, ACE05-E+ and ERE-EN for event extraction. For each column, we bold the highest score.

Insert Setting	Arg-I			Arg-C		
	P	R	F1	P	R	F1
full model	59.7	59.9	59.8	57.6	57.8	57.7
w/o TM	57.9	59.1	58.5	55.9	57.1	56.5
w/o PTR	58.2	59.3	58.7	56.3	57.3	56.8
w/o PTR and TM	56.8	59.1	57.9	54.8	57.0	55.9

Ablation study for the pluggable components on ACE05-E+

基于触发词检测增强的生成式事件抽取 (TDE-GTEE)

- 为了更好地适应小样本环境，事件检测模块使用基于Prompt的方式替换原先的多分类方式
- 在ACE05上与其他具有Few/Zero-shot事件检测模型对比



Model	Full	5-way 0-shot	5-way 1-shot	5-way 5-shot	5-way 10-shot
ONEIE	74.7/ 56.8	-	4.2/ 1.5	38.5/ 22.8	53.3/ 34.9
DEGREE	72.2/ 54.4	46.8/ 25.1	53.3/ 31.6	55.5/ 36.1	61.2/ 42.0
GTEE	72.6/ 55.8	48.6/ 35.2	52.9/ 40.2	52.4/ 41.4	56.7/ 45.2
TDE-GTEE	78.8/ 57.4	48.6/ 37.2	56.3/ 38.6	64.0/ 44.5	63.3/ 45.6

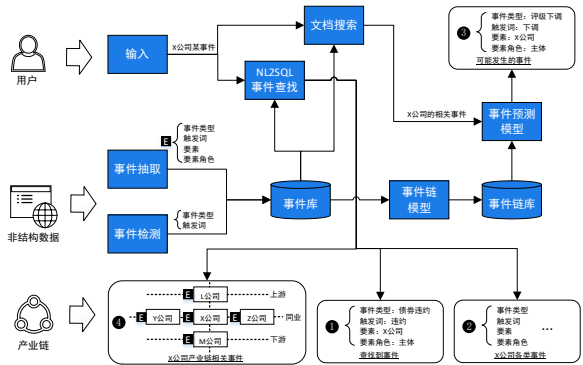
在数据集ACE05上，有监督, Zero-shot, Few-shot的对比，指标为Trigger-Classification和Argument-Classification的F1-score(%)

基于触发词检测增强的生成式事件抽取模型+基于Prompt的事件类型检测

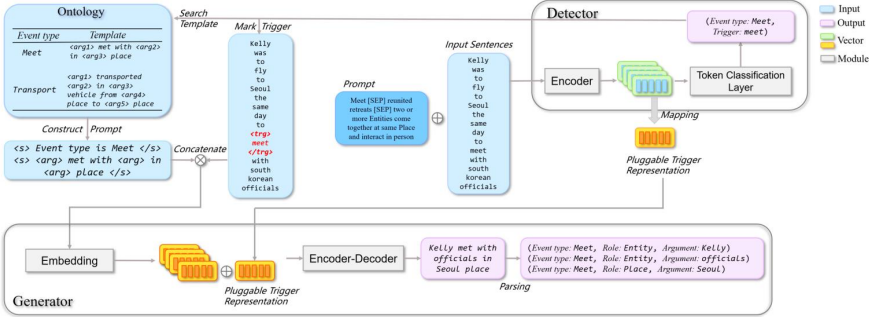
3

总结和展望

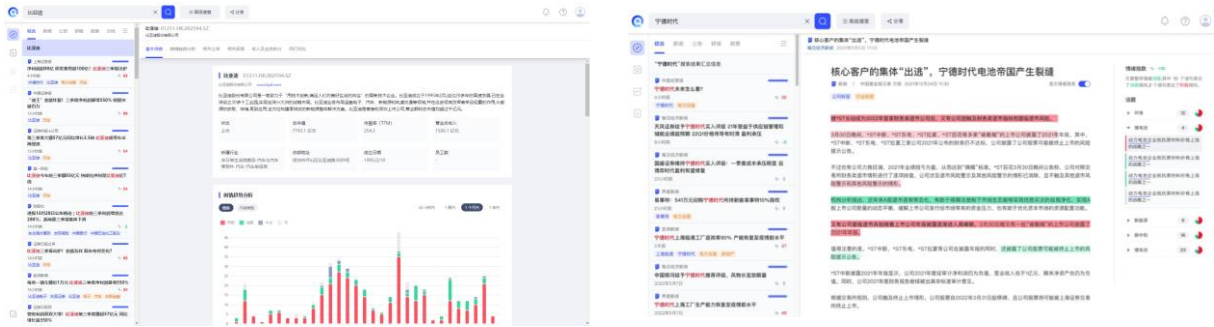
金融事件分析任务和技术：一个完善的事件分析框架
将为金融领域的信息理解提供强有力的技术支持



澜舟科技在金融事件分析领域中提出的TDE-GTEE等模型的效果和零样本/少样本学习能力达到了SOTA水平



澜舟科技持续推进金融事件分析和相关技术的落地
和产品化



Thank you!

北京澜舟科技有限公司

bd@langboat.com



关注我们



合作咨询

访问澜舟科技官网 <https://langboat.com> 了解更多澜舟科技产品服务