

# 面向开放文本的逻辑推理知识抽取与事件影响推理探索



刘焕勇<sup>1</sup>, 李瑞<sup>1</sup>, 任红萍<sup>1</sup>, 陈贺<sup>1</sup>, 薛云志<sup>1,2</sup>, 张鹏<sup>3</sup>

<sup>1</sup>中国科学院软件研究所, 北京, 100190 <sup>2</sup>计算机科学国家重点实验室, 北京, 100190  
<sup>3</sup>数据地平线(广州)科技有限公司, 广州, 511458



## 1. 工作应用场景

·**真实事件**: 2018年06月04日, 崔永元爆料范冰冰疑似“阴阳合同”的事件, A股股市开盘后, 整个文化传媒板块整体陷入暴跌, 截至收盘, 文化传媒板块40家公司集体飘绿, 一天之内, 市值累计蒸发179.33亿元。**此类事件是否有规律可循? 可否利用推理逻辑加以模拟刻画?**

公司名称	最新价 (元)	涨跌幅 (%)	涨跌幅 (元)	总股本 (亿)	市值变化 (亿)	
华谊兄弟	5.12	-5.69	-0.57	-10.026	6.78	-3.86
华谊兄弟	7.36	8.18	-0.82	-10.026	27.75	-22.76
鲁信影视	14.85	16.5	-1.65	-10.006	4	-6.60
当代明诚	15.27	16.92	-1.65	-9.756	4.87	-8.04
誉辰文化	5.82	6.38	-0.56	-8.798	8.6	-8.82
敦煌种业	5.55	5.99	-0.44	-7.458	9.81	-8.32
傲文传媒	28.1	29.91	-1.81	-6.008	3.39	-6.14



### ·工作目标

1. 构建大规模逻辑推理知识库, 建模领域常识推理路径
2. 探索基于逻辑推理知识库的事件影响推理策略模拟

### ·应用场景

1. 常识搜索: 面向事件与实体的逻辑推理知识获取与搜索
2. 事件预警: 金融与情报领域事件影响驱动建模

## 2. 逻辑知识库的定义与构成

逻辑推理知识库, 指具有逻辑刻画能力、由具有逻辑性推理因子组成的一类常识知识库。形式化定义为:  $G(V, E)$ ,  $V$ 表示推理因子节点, 常以名词、名词性词组、谓词、谓词性词组或短语充当,  $E$ 表示推理因子之间有向的逻辑演化关系边

类别	产业链逻辑推理知识库	实体描述逻辑推理知识库	事件因果逻辑推理知识库
节点类型	行业、公司、产品等行实体	人物、地点、机构等实体; 实体属性特征	主语(高)形式的事件短语或短句
关系类型	<公司-主营-产品>, <产品-原料-产品>, <产品-下游-行业>等	<实体-描述-属性>, <描述-属性-实体>	<事件-因果-事件>
用途	实体间关联推理	实体关联属性推理	事件演化关系推理

表2.1-逻辑推理知识库构成



图2.1-产业链逻辑推理知识库示例



图2.2-事件因果逻辑推理知识库示例



图2.3-实体描述推理知识库示例

## 3. 基于序列标注方式的逻辑知识抽取

针对产业链逻辑知识, 通过收集及专家标注方式构建, 事件因果逻辑、实体描述逻辑抽取, 采用面向开放文本的序列标注方式抽取

数据类型	数据量	描述
训练集	8,349条	标注数据包括单实体-多个描述
验证集	1,296条	多个实体-多个描述句, 多个产品-下游行业-等
测试集	1,172条	实体-单句描述等多种类型
数据总量	10,817条	训练集+验证集+测试集 8:1:1

表3.1-实体描述逻辑抽取数据情况

数据类型	数据量	描述
训练集	7,000条	标注数据包括单触发词-无因果、
验证集	1,000条	单触发词-单因果、单触发词-多因果、
测试集	2,000条	多触发词-多因果、多触发词-因果等多种类型
数据总量	10,000条	训练集+验证集+测试集 7:2:1

表3.2-事件因果推理逻辑抽取数据情况



图3.1-实体描述推理逻辑抽取序列标注模型

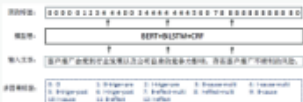


图3.2-事件因果推理逻辑抽取序列标注模型

评测项	Precision	Recall	F1
实体-Entity	97.62%	96.47%	97.04%
触发词-Trigger	96.51%	97.70%	96.59%
实体概念描述-Desc	88.42%	94.38%	91.30%
Token	93.68%	95.17%	94.90%
<实体-触发词, 实体-描述>	84.75%	86.29%	85.51%

表3.3-实体描述推理逻辑抽取效果

模型	Precision	Recall	F1
规则因果抽取模型	42.68%	100%	59.83%
单标体系+BERT+BLSTM+CRF	90.14%	87.23%	88.68%
多标体系+BERT+BLSTM+CRF	85.34%	82.76%	84.03%

表3.4-事件因果推理逻辑抽取效果

## 6. 逻辑推理知识库进一步的工作

1. 实时逻辑推理知识学习与检索系统 <https://xueji.datahorizon.cn>
2. 开放逻辑推理知识库 <http://www.openkg.cn/organization/datahorizon>



## 4. 基于逻辑推理知识库的知识实证与影响推理

### ·逻辑推理知识库的现有规模

数据来源	搜房网、东方财富网等行业网站、主流媒体网站、门户网站、百科类资讯、社区类网站共1375个, 合计文本49,019,726篇
产业链逻辑推理知识库	4大类实体, 8大类关系, 实体规模37,116个, 关系边311,781条。包括产品上下位、产品上游原料、下游产品、产品所属行业、公司所属板块、公司所属行业类等关系。包括产品、公司、行业、板块四大类实体。
实体描述逻辑推理知识库	包括8,904,569个实体描述逻辑对
事件因果逻辑推理知识库	包括12,904,569个事件因果逻辑对

### ·基于逻辑推理知识库的知识实证

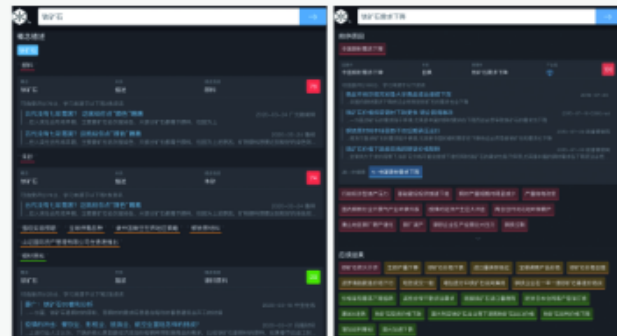


图4.1-实体描述推理逻辑知识学习示例

图4.2-因果事件推理逻辑知识学习示例

### ·逻辑推理知识库的事件影响推理



图4.3-智利地震芬兰港口罢工及铁矿石需求下降的影响推理

## 5. 总结

1. 开放文本为构建大规模推理逻辑知识库提供了丰富素材, 逻辑推理知识库, 可为实体关联查询、事件逻辑推理、领域事件推理提供支撑
2. 领域逻辑推理知识抽取的复用、逻辑推理知识可信度计算、推理逻辑的演算和量化是下一步的挑战

## 7. 联系我们

联系邮箱: [huanyong@iscas.ac.cn](mailto:huanyong@iscas.ac.cn)  
联系地址: 北京市海淀区中关村南四街4号  
项目主页: <https://liuhuanrong.github.io>  
工业应用: <https://www.datahorizon.cn>

