**CCKS 2020 技术评测任务书**

# 面向金融领域的小样本跨类迁移事件抽取

# 任务描述

在金融领域，事件抽取是一项十分重要的任务，也是自然语言处理领域一项比较复杂的任务，而小样本下的事件抽取模型在落地应用中也极为需要。本任务需要从金融领域新闻资讯句子中，抽取事件知识（包括事件类型、触发词和事件元素），并将大样本下训练的模型跨类迁移到小样本的其他事件类型上。其中，事件类型分为两类，**初始事件**类型限定为：质押、股份股权转让、投资、起诉和高管减持，**需要迁移**的事件类型为：收购、担保、中标、签署合同和判决，每个事件类型都有其对应的事件框架，需要抽取出每个事件对应的事件元素[[1]](#footnote-1)。即给出一段句子级新闻资讯文本，针对该文本需要判断其所属的事件类型，抽取该事件的各个事件元素。

**示例数据：**

**输入：**

刚刚，A公司发布情报通告，称已于2019年10月28日向广州知识产权法院就B公司涉嫌滥用市场支配地位等相关事宜提起诉讼，并于2019年11月4日得到受理。

**输出：**

* 事件类型：起诉
* 触发词：诉讼
* 原告（公司）：A公司
* 被告（公司）：B公司
* 起诉日期：2019年10月28日

# 数据描述

本次数据来自真实金融新闻，由专业人员标注，训练集、验证集及测试集的说明如下：

**训练集：**

**训练集分为两个文件给出：**

1. **包含初始事件类型的训练集，数据规模为7000左右；**
2. **包含需要迁移的五个事件类型的训练集，其中收购、判决两个事件训练数量约各200条，其余三个事件约各100条。**

训练集每行为一个JSON字符串，形式如下：

{

 "id": "id字符串",

"content": "文本内容"

 "events": [

 {

 "type": "事件类型",

 "mentions": [

 {

 "word":"事件元素",

 "span": [起始位置, 终止位置], //两个元素，分别为起始终止位置

 "role": "trigger"

 },

 {

 "word":"事件元素",

 "span":[起始位置, 终止位置], //两个元素，分别为起始和终止位置

 "role": "sub",

 },

 ......

 ]

 },

 {

 "type": "事件类型",

 ......

 }

 ]

}

**验证集：**

**验证集分为两个文件给出：**

1. **与初始事件类型对应的验证集，数据规模为20万左右；**
2. **收购、判决两个事件类型对应的验证集，数据规模为10万左右。**

第二个文件格式与训练集相同。其余两个文件每行为一个JSON字符串，形式如下：

{

 "id": "id字符串",

"content": "文本内容"

}

选手需要在第一个文件预测与训练集对应的5个事件，在第三个文件预测收购、判决两个事件。

**测试集：**

**测试集将给出五个需要转移的事件类型的测试集，数据规模为35万左右。**

测试集每行为一个JSON字符串，形式如下：

{

 "id": "id字符串",

"content": "文本内容"

}

选手需要在第三个文件预测收购、判决及新增的三个事件。文件每行为一个JSON字符串，形式如下：

{

 "id": "id字符串",

"content": "文本内容"

}

# 评价指标

本任务采用调整后的精准率（Precision,P）、调整后的召回率（Recall， R）、F1值（F1-measure,F1）来评估事件抽取的效果：

$$精准率（precision）=\frac{\sum\_{抽取的事件i}^{m}\sum\_{抽取的事件i中的元素j}^{k}\frac{x}{k}}{抽取的事件总数量m}$$

$$召回率（recall）=\frac{\sum\_{抽取的事件i}^{m}\sum\_{抽取事件i中的元素j}^{k}\frac{x}{k}}{标注的事件总数量n}$$

$$F1=\frac{2\*precision\*recall}{(precision+recall)}$$

其中k 为标注数据中事件i的元素个数， x = 1 if 事件i的事件类型与标注一致且元素j与标注一致 else 0, n 为标注的事件总数量。最终榜单的评分会综合验证集和测试集排名。

**评价指标伪代码**

# 任务提交

本次任务将采取刷榜的方式。验证集发布后，允许参赛队伍多次向平台提交结果，将所有预测结果写入一个文件，文件命名为“valid\_result.txt”，格式与结果文件result.txt相同，排名随时更新。参赛队伍可在评测集发布之前随时上传验证集的计算结果（每日最多可上传3次），管理系统会及时更新各队伍的最新排名情况；

测试集发布后，允许参赛队伍多次提交结果文件，结果文件提交格式如下。

选手将结果保存为result.txt，以utf-8编码格式保存，每行为一个JSON字符串，形式为

{

 "id": "id字符串",

 "events": [

 {

 "type": "事件类型",

 "mentions": [

 {

 "word":"事件元素",

 "span": [起始位置, 终止位置], //两个元素，分别为起始和终止位置

 "role": "trigger"

 },

 {

 "word":"事件元素",

 "span":[起始位置, 终止位置], //两个元素，分别为起始和终止位置

 "role": "sub",

 },

 ......

 ]

 },

 {

 "type": "事件类型"

 ......

 }

 ]

}

最终提交文件要求：

每一个参赛队需提交的材料如下：

1. 事件抽取运行结果文件
2. 代码及说明
3. 方法描述文档

代码及其文档需打包成一个文件（tar，zip，gzip，rar等均可），用code.xxx命名，要求提交所有的程序代码及相关的配置说明，确保程序能够正确运行，且所得结果与result.txt相符。

评测平台：本次评测将依托biendata平台（https://biendata.com/）展开，请有意向的参赛队伍关注平台上的竞赛列表。

# 时间安排

• 任务征集截止：3月1日

• 任务准备时间：3月1日—3月20日

• 评测任务发布：3月20日

• 报名时间：3月20日—7月10日

• 训练集数据发布：3月20日

* 验证集数据发布：4月5日

•   验证集提交开放：4月5日23:59 - 7月5日23:59 （北京时间）

•     测试数据发布：7月13日23:59（北京时间）

•     测试集提交开放： 7月13日23:59 - 7月20日23:59（北京时间）

•     评测论文提交：8月5日

•     CCKS会议日期：8月15日—8月18日

# 规则

1. 所有参赛选手都必须在biendata平台管理系统中注册，本次比赛的参赛对象仅限全日制在校大学生（本科、硕士、博士均可）和企业员工；

2. 参赛选手需确保注册时提交信息准确有效，所有的比赛资格及奖金支付均以提交信息为准；

3. 参赛选手在管理系统中组队，参赛队伍成员数量不得超过5个，报名截止日期之后不允许更改队员名单；

4. 每支队伍需指定一名队长，队伍名称不超过15个字符，队伍名的设定不得违反中国法律法规或公序良俗词汇，否则组织者有可能会解散队伍；

5. 每名选手只能参加一支队伍，一旦发现某选手以注册多个账号的方式参加多支队伍，将取消相关队伍的参赛资格；

6. 允许使用开源代码或工具，但不允许使用任何未公开发布或需要授权的代码或工具；

7. 除主办方提供的数据集外，参赛选手允许使用预训练数据（如词向量、字向量等），但是不能用额外的标注数据。

8. 参赛队伍可在参赛期间随时上传验证集的预结果，一天不能超过3次 ，管理系统会实时更新各队伍的最新排名情况。

# 任务组织者及联系人

组织者：招商银行 人工智能实验室

联系人：徐洁馨 jiexinx@cmbchina.com

 吴垠 xyionwu@cmbchina.com

 盛剑 jsheng@cmbchina.com

届时会开放微信群作为任务讨论平台。

# 奖项安排（包含大会专委会奖金及招商银行提供奖金）

第一名：10000

第二名：10000

第三名：5000

第四名：5000

1. 第七名：1500

创新技术奖（两名）：5000

优胜者有机会进入2021年招商银行IT团队校园招聘终面。

# 附录

## 事件抽取相关概念

* 1. 事件（Event）

事件是发生在某个特定的时间点或时间段、某个特定的地域范围内，由一个或者多个角色参与的一个或者多个动作组成的事情或者状态的改变。一种事件由一种事件框架定义。

* 1. 事件框架（Event Schema）

一个事件框架定义了一类事件，包括事件中含有的触发词、事件元素和元素角色。

* 1. 事件类型（Event Type）

本任务第一阶段需抽取的事件类型包含以下五个：质押、股份股权转让、投资、起诉和高管减持。**一个事件类型在一句话中可能会重复出现多次，如“A公司转让给B公司，C公司转让股权给D公司”中股权转让事件出现了两次，此时应该分别抽取。**

* 1. 触发词（Event Trigger）

触发词是事件描述中最能代表事件发生的词汇，决定事件类别的重要特征，一般是动词或者名词。有些情况下，触发词会出现两次，或触发词分散在几处，如： “*4月20日，翰宇药业实控人曾少贵、曾少强完成了一笔协议转让，2人合计向共赢基金转让5569.60万股，占总股本6%，交易总价款为5.73亿元。*”中，触发词“转让”出现了两次；“*某公司转让了%6的股权*”中，触发词“转让股权”分隔为两个词。这种情况下，应抽取第一句中的第二个“转让”和第二句中的“转让”为触发词。

一般触发词应是能单独表达或代表对应事件类型的词语。

* 1. 事件元素（Event Arguments）

事件元素是事件的重要信息，或者说是实体描述（entity mention），主要由实体、属性值等表达完整语义的细粒度单位组成，每个事件类型的事件元素由事件框架确定。

当同一词语出现在句中多次，均需抽取，用逗号分隔。

* 1. 元素角色（Argument Role）

元素角色指的是事件元素在事件中扮演的角色，是事件元素与事件的语义关系，每个事件类型的元素角色由事件框架确定。

## **金融事件类型解释**

* 1. 质押

债务人或第三方将其动产或权利移交债权人占有，将该动产或权利作为债权的担保。债务人或第三方为出质人，债权人为质权人，移交的动产或权利为质押物。动产质押物包含出质人合法占有的一切动产，权利质押物包含汇票、支票、本票、债权、存款单、仓单、提单、股份、股票、商标专用权、专利权、著作权等。

质押有区别于抵押、质权。

* 1. 股份股权转让

公司股东依法将自己的股东权益有偿转让给他人，股东（转让方）让他人（受让方）双方当事人意思表示一致而发生的股份股权转移，转让形式包含持份转让、股份转让、书面股权转让、非书面股权转让、即时股权转让、预约股权转让、公司参与的股权转让、公司非参与的股权转让、有偿股权转让、无偿股权转让等。

如发生A公司收购B公司，在此事件类型中，A为受转让公司，B为标的物公司。

* 1. 投资

投资，指国家或企业以及个人，为了特定目的，与对方签订协议，促进社会发展，实现互惠互利，输送资金的过程。又是特定[经济主体](https://baike.baidu.com/item/%E7%BB%8F%E6%B5%8E%E4%B8%BB%E4%BD%93/11059502)为了在未来可[预见](https://baike.baidu.com/item/%E9%A2%84%E8%A7%81/5738463)的时期内获得[收益](https://baike.baidu.com/item/%E6%94%B6%E7%9B%8A/638677)或是[资金](https://baike.baidu.com/item/%E8%B5%84%E9%87%91/1671461)增值，在一定时期内向一定领域投放足够数额的资金或实物的货币等价物的经济行为。可分为[实物投资](https://baike.baidu.com/item/%E5%AE%9E%E7%89%A9%E6%8A%95%E8%B5%84/7588488)、资本投资和[证券投资](https://baike.baidu.com/item/%E8%AF%81%E5%88%B8%E6%8A%95%E8%B5%84/2378)等。前者是以[货币](https://baike.baidu.com/item/%E8%B4%A7%E5%B8%81/85299)投入企业，通过生产经营活动取得一定[利润](https://baike.baidu.com/item/%E5%88%A9%E6%B6%A6/2714509)，后者是以货币购买企业发行的股票和公司债券，间接参与企业的利润分配。

* 1. 起诉

起诉是依法向法院提出诉讼，请求法院对特定案件进行审判的行为。刑事案件的起诉，在于请求法院通过审判对被告人定罪判刑；刑事附带民事诉讼的提起，在于请求法院通过审判追究被告人的经济赔偿责任。起诉必须是发起诉讼权的公民或法人依法向有管辖权的法院进行。在中国，刑事案件的起诉，主要是指人民检察院提起公诉，只有对告诉才处理和其他不重要进行侦查的轻微刑事案件，才由被害人或其法定代理人提起自诉；刑事附带民事诉讼的提起，可以是被害人及其法定代理人，也可以是人民检察院和受侵害的法人单位。

* 1. 高管减持

高管减持是指上市公司的高管在二级市场（二级市场：指在证券发行后各种证券在不同的投资者之间买卖流通所形成的市场）卖出自己公司股票的行为。创业板高管减持的方式主要有在职抛售和辞职套现两种方式。《公司法》对高管减持进行了限制，要求上市公司高管在股票上市交易之日起一年内不得转让其所持本公司的股份，且任职期内每年转让的股份不得超过持有股份总数的25％，并在辞职半年内不得转让其所持股份。在解禁期满后，如果公司没有对股票出售的特殊限制，按照《公司法》的规定，高管每年减持的股票数量不得超过持有总量的25％。

* 1. 收购

收购是指一个公司通过[产权交易](https://baike.baidu.com/item/%E4%BA%A7%E6%9D%83%E4%BA%A4%E6%98%93/2369989)取得其他公司一定程度的控制权，以实现一定经济目标的经济行为。收购有区别于回购。

* 1. 判决

司法机关对审理结束的案件作出裁决。

* 1. 担保

当事人根据法律规定或双方约定，为促使债务人履行债务，向债权人订立担保合同。担保方式有保证、抵押、质押、留置、定金。

* 1. 中标

招标人向经过评选的投标人发出中标通知书，并在规定时间内与之订立书面合同。

* 1. 签署合同

指合作双方经过商定后在文件、条约、凭证等书面文件上签字或盖章等形式确立合作关系。

## 事件框架定义

* 1. 质押

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub-org | 质押公司：发起质押的**公司** |
| sub-per | 质押人：发起质押的**个人** |
| obj-org | 质权公司：接受质押的**公司** |
| obj-per | 质权人：接受质押的**个人** |
| collateral | 质押物（标的，可为公司股票、大额存单等等） |
| date | 质押日期（**显示**表达的日期，如“2019年9月24日”） |
| money | 质押**金额** |
| number | 质押**数量** |
| proportion | 质押**比例** |

* 1. 股份股权转让

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub-org | 股份股权转让公司：发起股份股权转让的**公司** |
| sub-per | 股份股权转让人：发起股份股权转让的**个人** |
| obj-org | 受转让公司：接受股份股权转让的**公司** |
| obj-per | 受转让人：接受股份股权转让的**个人** |
| collateral | 股份股权转让物（标的，可为公司股票、股权等等） |
| date | 转让日期（**显示**表达的日期，如“2019年9月24日”） |
| money | 转让交易**金额** |
| number | 转让**数量** |
| proportion | 转让**比例** |
| Target-company | 标的公司：指收购行为中的收购对象公司 |

* 1. 起诉

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub-per | 原告（**个人**） |
| sub-org | 原告（**公司**） |
| obj-per | 被告（**个人**） |
| obj-org | 被告（**公司**） |
| date | 起诉日期 |

* 1. 投资

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub | 发起投资的**组织**或**单位**（投资方） |
| obj | 被投资的**组织**或**单位**（被投资方） |
| money | 投资**金额** |
| date | 日期 |

* 1. 减持

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub | 减持方（通常为人，少数情况为组织） |
| title | 减持方的职务 |
| date | 日期 |
| share-per | 减持的股份占**个人**股份百分比 |
| share-org | 减持的股份占**公司**股份的百分比 |
| org | 减持方所在组织或单位 |

* 1. 收购

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub-org | 收购公司：发起收购的**公司** |
| sub-per | 收购方：发起收购的个人 |
| obj-org | 被收购公司：接受收购的**公司** |
| way | 收购方式（现金收购、股份股权收购、承担债务式收购等等） |
| date | 收购日期（**显示**表达的日期，如“2019年9月24日”） |
| money | 收购**金额** |
| number | 收购股份**数量** |
| proportion | 收购股份**比例** |

* 1. 担保

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub-org | 担保公司：发起担保的**公司** |
| sub-per | 担保人：发起担保的**个人** |
| obj-org | 受担保公司：接受担保的**公司** |
| way | 担保方式（保证、抵押、质押、留置、定金等等） |
| amount | 担保金额 |
| date | 担保日期（**显示**表达的日期，如“2019年9月24日”） |

* 1. 中标

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub | 中标方（**公司**） |
| obj | 招标方（**单位或组织**） |
| amount | 中标金额 |
| date | 中标日期 |

* 1. 签署合同

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| sub-org | 发起合同签署的**组织**或**单位**（甲方） |
| sub-per | 发起合同签署的自然人（甲方） |
| obj-org | 接受合同签署的**组织**或**单位**（乙方） |
| obj-per | 接受合同签署的**自然人**（乙方） |
| amount | 成交额 |
| date | 日期 |

* 1. 判决

|  |  |
| --- | --- |
| **标签** | **说明** |
| trigger | 触发词 |
| institution | 裁判单位（**法院**） |
| sub-per | 原告**（个人）** |
| sub-org | 原告**（公司）** |
| obj-per | 被告（**个人**） |
| obj-org | 被告（**公司**） |
| date | 判决日期 |
| money | 判决金额 |

1. 事件抽取相关概念及解释说明详见附录。 [↑](#footnote-ref-1)