

任务：面向金融领域的篇章级事件抽取和事件因果关系抽取

1. 任务描述

事件抽取是舆情监控和金融领域的重要任务之一。“金融事件”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。文本的复杂体现在事件抽取的输入文本可能是句子、段落或者篇章，不定长度的文本使得限制文本长度的模型无法使用；任务的复杂体现在事件识别的任务包括：事件类型识别，事件要素抽取，事件关系抽取等等。本评测任务是解决篇章级事件元素抽取和事件因果关系抽取这两个核心的知识抽取问题。

本次评测任务的文本语料来自于互联网上的公开新闻、报告。在篇章级事件元素抽取任务中，给定篇章级长文本和事件类型，从篇章级文本中识别事件的元素。在事件关系抽取任务中，给定一段描述因果或影响关系的文本，从文本中抽取原因事件的表示和结果事件的表示，其中事件的表示包括事件类型和事件的三个要素：影响地域、产品、行业。本次评测包括两个子任务：篇章级事件元素抽取和事件因果关系抽取。具体内容如下：

子任务一：篇章级事件元素抽取

事件用事件类型和事件要素来表示，该任务旨在从给定的长文本中抽取事件的 13 个要素，即给定文本 T 抽取 T 中的 13 个事件要素。同时，该任务提供文本所属的事件类型作为输入，选手可参考使用。每个事件要素的定义和描述如下表所示：

事件要素名称	事件要素取值描述
涉案平台	涉案交易发生的平台。注意如果有多个平台需要都识别出来
案发城市	城市信息，能抽取到的最细粒度，如“xx省xx市xx区”
嫌疑人	嫌疑人姓名、账号、代称、账号昵称、账号号码等唯一描述性的称谓。注意如果有多个不同的称谓需要都识别出来
交易号	加密后的交易单号
订单号	加密后的订单号
手机号	加密后的手机号
案发时间	文本中表述最完整的时间
身份证号	加密后的身份证号
支付渠道	可枚举的支付工具。注意如果有多个渠道需要都识别出来
受害人	受害人姓名、账号、代称、账号昵称、账号号码等唯一描述性的称谓。注意如果有多个不通的称谓都需要识别出来
银行卡号	加密后的银行卡号。注意如果有多个银行卡需要都识别出来
资损金额	包含金额和单位。注意如果有多次金额需要都识别出来
受害人身份	文本中描述的受害人的家庭或者社会身份。如“老人”、“未成年人”、“我妈”，“我老婆”

输入：一段文本 T 和额外的三级事件类型 level1,level2,level3

输出：表格中描述的事件要素

示例：

样例

输入： `{“text_id”:“123456”, “text”: “王某新认识一个网友小刘,聊了一会之后看了一下情况就很自来熟一样,过了几天小刘说肚子疼上医院要软件打五百块钱给她,没过多久脚又被行李箱砸伤又问王某要了八百, 之后联系不上”, “level1”:“ 欺诈风险”, “level2”:“ 冒充身份类”, “level3”:“ 好友/领导/婚恋交友”}`

输出： `{“text_id”:“123456”, “attribute”:[{“type”:“支付渠道”, “entity”:“软件”, “start”:47, “end”:48}, {“type”:“资损金额”, “entity”:“五百块钱”, “start”:50, “end”:53}, {“type”:“资损金额”, “entity”:“八百”, “start”:75, “end”:76}, {“type”:“嫌疑人”, “entity”:“小刘”, “start”:9, “end”:10}, {“type”:“受害人”, “entity”:“王某”, “start”:0, “end”:1}]}`

子任务二：事件关系抽取

该任务旨在从文本中抽取具有因果关系两个事件：原因事件和结果事件，并通过事件的四个要素来表示原因事件和结果事件。即给定文本 T，抽取 T 中所有的原因事件 M 和结果事件 N。即分别抽取 M 和 N 的事件类型和 3 个事件要素，如下表所示：

事件要素	事件要素描述
事件类型	原因事件有 40 个事件类型，结果事件有 19 个事件类型
影响地域	事件影响的地域，文本中完整的地域描述
产品	事件影响的产品。产品与行业的区别是产品的范围小，通常只是一类具体的产品
行业	事件影响的行业。行业与产品的区别是行业的范围大，通常是多类产品组成一个行业

注：每个事件要素的取值可能有多个，需要全部抽取出来并用列表来表示

输入： 一段文本 T

输出： 原因事件和结果事件的类型及其事件要素

示例：

样例

输入： `{“text_id”:“123456”, “text”: “货运行业盈利水平大幅度缩水是卡车需求不旺的主要原因”}`

输出： `{“text_id”:“123456”, [{“reason_type ”:“产品利润下降”, “reason_region”:[], “reason_product”:[], “reason_industry”:[“货运行业”], “result_type”:“需求减少”, “result_region”:[], “result_product”:[“卡车”], “result_industry”:[]}]}`

2. 数据描述

本次数据主要来自金融领域的公开新闻、报道，样本包含正样本和负样本，训练集、验证集及测试集的说明如下：

任务一：

训练集&验证集：

在训练及验证数据发布阶段，我们会发布 5.5 千条左右的文本及其所标注的三个事件类型以及每个文本标注的事件要素，和 1 千条左右的验证文本及其所标注的三个事件类型。训练集以 json 格式，包含“text”、“level1”、“level2”、“level3”分别表示文本和三个事件类型，包含“attributes”表示训练集中事件的要素，注意同一个要素可能有多个取值，需要都识别出来。验证集以 json 格式，“text”、“level1”、“level2”、“level3”分别表示文本和三个事件类型。

测试集：

在测试数据发布阶段，我们将会再发布 1 千条左右的文本数据集以及一定数量的干扰数据，不含标注结果，作为测试。以 json 格式，“text”、“level1”、“level2”、“level3”分别表示文本和三个事件类型。

任务二：

训练集&验证集：

在训练及验证数据发布阶段，我们会发布 8 千条左右的文本及其所标注原因事件和结果

事件的事件类型和三个事件要素，以 json 格式，包含“text”、“reason_type”、“reason_region”、“reason_product”、“reason_industry”、“result_type”、“result_region”、“result_product”、“result_industry”。注意，对于一段文本一个原因事件和结果事件组合作为一个唯一结果单元，一段文本可能抽取多个事件类型组合，三个事件要素的取值可能有多，需要都识别出来。验证集包含 1 千条左右的验证文本，以 json 格式，包含“text”

测试集：

在测试数据发布阶段，我们将会再发布 1 千条左右的文本数据集，不含标注结果，作为测试，以 json 格式，包含“text”

说明：以上数据来源于互联网公开信息样本，仅用于本次比赛交流目的，不具有统计意义

3. 评价指标

任务 1 评价指标：

本次任务采用精确率（Precision, P）、召回率（Recall, R）、F1 值（F1-measure, F1）来评估篇章事件要素的识别效果。采用微平均计算 F 值即所有样本一起计算 P 和 R。

$$\text{事件要素精确率} = \frac{\text{识别篇章事件元素和标注相同的数量}}{\text{识别出篇章事件元素的总数量}}$$

$$\text{事件要素召回率} = \frac{\text{识别篇章事件元素和标注相同的数量}}{\text{标注的篇章事件元素总数量}}$$

$$\text{事件要素F1值} = \frac{2 * \text{事件主体精确率} * \text{事件主体召回率}}{\text{事件主体精确率} + \text{事件主体召回率}}$$

任务 2 评价指标：

本次任务采用精确率（Precision, P）、召回率（Recall, R）、F1 值（F1-measure, F1）来评估事件要素的识别效果。采用微平均计算 F 值即所有样本的要素一起计算 P 和 R。原因事件和结果事件两个事件类型及其某一要素为一个统计项，如果只识别原因和结果事件类型没有正确识别事件要素不算正确。同一原因事件结果事件类型的组合最多有一组结果。

给定一个文档，假设答案包含两个事件实例：

A1: 原因事件类型 1，结果事件类型 2，原因事件要素一取值 1 和 6，原因事件要素二取值 2，结果事件要素一取值 3，结果事件要素二取值 4

A2: 原因事件类型 3，结果事件类型 4，原因事件要素一取值 1，原因事件要素二取值 6，结果事件要素二取值 7，结果事件要素三取值 8

假设选手给出两个预测结果：

P1: 原因事件类型 1，结果事件类型 2，原因事件要素一取值 1，原因事件要素二取值 2，原因事件要素三取值 3，结果事件要素二取值 4

P2: 原因事件类型 3，结果事件类型 4，原因事件要素一取值 1，结果事件要素二取值 2

P3: 原因事件类型 1，结果事件类型 4，原因事件要素一取值 9，结果事件要素二取值 10

评测时，会以标准答案中原因事件类型和结果事件类型的组合去预测结果中查找对应的事件要素，如果预测结果中的原因结果事件类型组合不在标准答案中，则认为该结果所有要素全部错误。如果找到事件类型组合依次判断 6 个事件要素（原因事件三个要素，结果事件

三个要素)匹配的数量。例如上例中 A1 和 P1 的原因和结果事件类型相同,相同要素包括:原因事件要素一取值 1,原因事件要素二取值 2,结果事件要素二取值 4。A2 和 P2 原因和结果事件类型相同,相同的要素包括:原因要素一取值 1。P3 的原因和结果事件类型组合不和标准答案中任何一组相同,认为结果全部错误。可以计算 Precision 和 Recall,如下

$$\text{Precision} = (3+1)/(4+2+2), \quad \text{Recall} = (3+1)/(5+4)$$

$$\text{事件要素精确率} = \frac{\text{识别出因果事件的要素和标注相同的数量}}{\text{识别出的因果事件要素总数量}}$$

$$\text{事件要素召回率} = \frac{\text{识别出因果事件的要素和标注相同的数量}}{\text{标注的因果事件要素总数量}}$$

$$\text{事件要素F1值} = \frac{2 * \text{事件要素精确率} * \text{事件要素召回率}}{\text{事件要素精确率} + \text{事件要素召回率}}$$

4. 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,文件命名为“参赛队名称_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。

结果文件格式说明:

选手将结果保存为 result.txt,以 utf-8 编码格式保存。对于任务一:使用 attribute 中每项的 type、entity 和 text_id 对应的标准答案计算评测指标, start 和 end 截取的字符串如果和 entity 不一致当做 entity 识别错误。

```
{“text_id”:“123456”, “attribute”:[{“type”:“支付渠道”, “entity ”:“软件”, “start”:47,“end”:48},{“type”:“资损金额”,“entity”:“五百块钱”, “start”:50,“end”:53},{“type”:“资损金额”,“entity”:“八百”, “start”:75,“end”:76},{“type”:“嫌疑人”,“entity”:“小刘”, “start”:9,“end”:10},{“type”:“受害人”,“entity”:“王某”, “start”:0,“end”:1}]}
```

对于任务二:和 text_id 对应的标准答案计算评测指标

```
{“text_id”:“123456”, [ {“reason_type ”:“产品利润下降”, “reason_region”:[], “reason_product”:[], “reason_industry”:[“货运行业”], “result_type”:“需求减少”, “result_region”:[], “result_product”:[“卡车”], “result_industry”:[ ]}] }
```

最终提交文件要求:

每一个参赛队需提交的材料如下:

1. 事件主体运行结果文件
2. 代码及说明
3. 方法描述文档(非评测论文,评测论文撰写要求见 CCKS 2021 官网)

以上三个文件需在任务提交截止日期前发送至邮箱:

子任务 1: 发送至 ccks2021taskFEE_t1@126.com。邮件的标题为:“CCKS-评测任务 FEE 子任务 1-最终提交文件-参赛队名称”,例如“CCKS-评测任务 FEE-子任务 1-最终提交文件-火箭队”。

子任务 2: 发送至 ccks2021taskFEE_t2@126.com。邮件的标题为:“CCKS-评测任务 FEE-子任务 2-最终提交文件-参赛队名称”,例如“CCKS-评测任务 FEE-子任务 2-最

终提交文件-火箭队”。

结果文件用 `result.txt` 命名，采用 `utf-8` 格式存储，文件格式需要与训练数据中的标注结果文件完全一样，结果可用 `json` 解析。

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。

方法描述文档用 `SM.pdf` 命名，包含算法描述及参数设置，需用 `pdf` 格式存储，页数不超过 10 页。

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。

5. 时间安排

- 任务准备时间：3 月 20 日—4 月 1 日
- 评测任务发布：4 月 1 日
- 报名时间：4 月 1 日—7 月 15 日
- 训练及验证数据发布：4 月 20 日
- 验证集提交开放：4 月 20 日 23:59 - 7 月 17 日 23:59（北京时间）
- 测试数据发布：7 月 17 日 23:59（北京时间）
- 测试集提交开放：7 月 17 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(评测报告及颁奖)：8 月 18 日—21 日

6. 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测结果，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。

7. 奖励设置

第一名：人民币 5000 元

第二名：人民币 3000 元

第三名：人民币 2000 元

创新奖：人民币 5000 元

备注：优胜者有机会进入 2022 蚂蚁集团校园招聘终面或者实习生绿色通道。

8. 任务组织者和联系人

任务组织者：

王太峰（蚂蚁集团）

王峰（蚂蚁集团）

张可尊（蚂蚁集团）

陈玉博（中科院自动化所）

刘康（中科院自动化所）

赵军（中科院自动化所）

任务联系人：

王峰：zifan.wf@antgroup.com

陈玉博：yubo.chen@nlpr.ia.ac.cn

9. 任务交流平台

任务交流平台采用钉钉群的方式。所有报名参赛的参赛队要求有一名成员加入到该讨论群中，后续所有有关本任务评测及数据的相关说明和通知将只在讨论群中发布和交流，不再另行发邮件或更新官网通知，请所有参赛队务必在报名成功之后加入该讨论群。钉钉群加入二维码如下，在钉钉上扫一扫添加：



子任务 1 钉钉群二维码



子任务 2 钉钉群二维码

加入讨论组时请务必注明所属参赛队队名及所属单位，加入讨论组有问题的参赛队请联系：**ccks2021taskFEE_t1@126.com** 或 **ccks2021taskFEE_t2@126.com**。