小样本多任务联合训练的事件抽取模型

任君翔，钟曦
万达信息股份有限公司，上海，201112
renjunxiang1215@sina.com

摘要：本文的目标，是解决金融领域的小样本跨类迁移事件抽取问题。本文基于 RoBERTa 作为预训练模型，采用文本分类、阅读理解相结合的方案，通过共享语言模型、联合训练子任务的方式，兼顾运算效率和模型精度，从文本中准确的抽取事件类型、触发词和事件元素，达到良好的事件抽取效果。

关键词：文本分类，阅读理解，联合训练

1 引言

在金融领域，小样本下的事件抽取模型是一项比较复杂的任务，同时在落地应用中也极为重要。本次比赛的任务是从初始的“质押”、“股份股权转让”、“投资”、“起诉”和“减持”五类事件，迁移到“收购”、“担保”、“中标”、“签署合同”和“判决”五类事件。针对句子级新闻资讯文本，判断其所属的事件类型（以下简称 type），抽取该事件的触发词（以下简称 tirgger）和各个事件元素（以下简称 mention）。

事件抽取从早期基于模式匹配[1]的方法发展到基于机器学习[2]的方法，如今基于深度学习[3]的方法已经被广泛应用于事件抽取中。Zhang 等提出了一种基于生成对抗模仿学习的实体-事件联合抽取框架[4]，利用逆向强化学习动态机制直接评估实体和事件抽取中实例的正确和错误标签，以提升模型的鲁棒性。Wang 等首先通过在改进的命名实体识别方法（NER）上引入 BERT 预训练模型增强汉字间的语义表示，然后利用基于 CNN 和 LSTM 多神经网络融合的事件主题抽取方法，提升模型的特征学习能力并增强事件主题识别的准确率，F1 score 达到 86.99%[5]。Sha 等提出了基于桥依赖 RNN 和论元张量交互的事件抽取联合模型[6]，在 bilstm 单元的基础上增加了单词之间的依赖桥信息，然后在每两个候选论元上建立一个张量层，来获取论元之间的信息交互。利用 Tree-LSTM 和 Bi-GRU 循环神经网络对候选事件句子进行编码，以获取丰富的语法信息和候选事件句子的文本特征。基于句子嵌入，候选事件类型可以直接被识别，而不是先通过识别触发词的事件类型然后使用字典匹配方法间接识别，该模型[7]提升了事件类型识别的性能。基于知识库的树状结构短期记忆网络（Tree-LSTM）框架[8]，其依赖结构可捕捉广泛的上下文，通过实体链接从外部本体获得实体属性（类型和类型描述），更好地对上下文信息和外部背景知识进行编码。Li 等基于 FrameNet 语料库构造了一个细粒度的事件层级框架[9]，在此基础上提出了基于 MLN 的联合预测模型来更有效地从文本中抽取事件。

本次比赛，我们在 BERT 预训练模型下游任务[10] 的基础上，将事件抽取拆解为文本分类（Classify，以下简称 CLF）、抽取式阅读理解（Machine Reading Comprehension，以下简称 MRC）和多分类的抽取式阅读理解（Machine Reading Comprehension + Classify，以下简称 MRC-C），通过共享预训练模型、联合训练三个子任务的方式实现事件抽取。

2 数据

本次比赛主要围绕句子级新闻资讯文本进行分类和信息抽取，我们对句子的长度、事件的类型等进行数据分析，以便于后续的模型设计和调优。

通过数据分析可以发现，99%的文本长度低于 250（表 1），大多为 100 以内的短文本（图1和图2）。因此在训练过程中可以将文本长度限制在250以内，从而扩大训练的batchsize。

表 1 文本长度分位数

分位数	训练集 base	训练集 trans
50	74	83
90	130	134
95	153.45	156
99	218	197.24
99.5	251.45	209.62
99.9	370.959	298.706
max	443	320

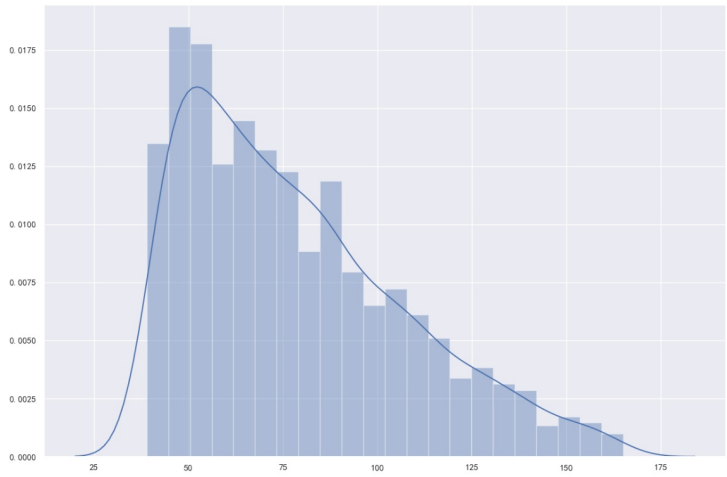

图 1 训练集 base 文本长度分布

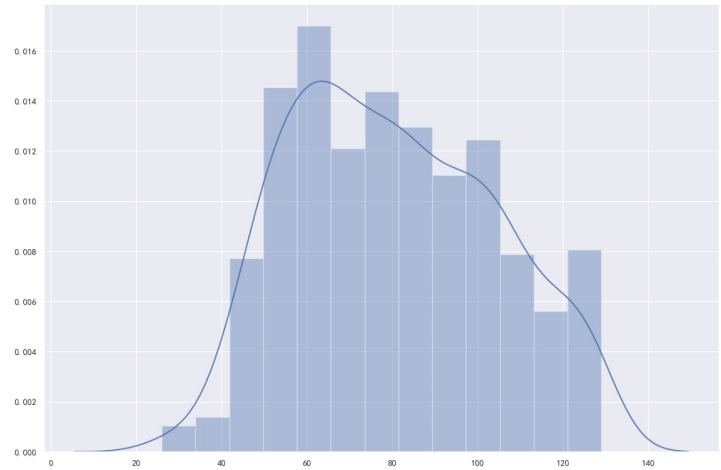

图 2 训练集 trans 文本长度分布

通过对文本事件的频数统计和分布来看（表 2），数据中存在不少同一类型事件出现多

次的情况，在设计模型时需要考虑同一类事件的拆分，后续我们通过识别 trigger 这个模型以及事件要素组合后处理的方式来解决这个问题。初始的五类事件和迁移的五类事件整体分布比较均匀，不需要额外的采样工作。

表 2 事件类型分布

事件类型	重复	去重
股份股权转让	1581	1063
投资	1083	795
质押	815	725
减持	670	491
起诉	533	477
判决	200	198
中标	200	179
收购	200	163
担保	163	153
签署合同	132	127

通过对文本事件频数进行分析可以发现（表 3），训练集 base 中一个文本包含多种事件的比例不低，因此需要通过每个事件二分类的方式有效识别是否包含此类事件。训练集 trans 没有这种情况，说明需要迁移的几个事件差别明显，不容易出现互相包含的情况。

表 3 事件频数

事件类型数	训练集 base	训练集 trans
0	4	0
1	1927	820
2	779	0
3	22	0

3 模型

多任务联合训练的优点主要体现在以下方面：小样本的情况下，少量的噪声很容易导致模型有偏和过拟合，多任务在一定程度上实现了数据扩充，增加了训练样本的大小，提高模型的泛化能力；多任务有助于提高模型捕捉各项任务特征的能力，学习到的信息会更加通用，从而提高模型的泛化能力和可迁移性。

因此，我们对事件抽取的各个要素进行拆解，拆分为事件类型、事件触发词和事件要素。通过文本数据的初步分析以及事件要素抽取文献的研究，我们认为需要先确定事件类型和触发词，再识别要素，然后采用联合训练的方式微调预训练模型。

3.1 事件分类及触发词识别

事件的界定由 type 和 trigger 共同决定，因此我们最初设计了两种方案：

1.先通过实体识别（以下简称 NER）的方式，经过双指针，利用 sigmoid 计算文本序列首尾指针概率，从而抽取 trigger；再通过观点型阅读理解（以下简称 OQMRC）的方式，将 trigger 作为 query，对[CLS]位置的输出进行分类，利用 sigmoid 计算事件概率，从而识别 type；

2. 先通过 CLF 的方式，对文本[CLS]位置的输出进行分类，利用 sigmoid 计算事件概率，从而识别 type；再通过 MRC 的方式，将 type 作为 query，经过双指针，利用 sigmoid 计算文本序列首尾指针概率，从而抽取 tirgger；

尽管方案一先识别 trigger 的方式避免了人为构造 query 和更贴近人类认知过程，但考虑到验证集过于庞大，OQMRC 会增加很多无意义的计算量，因此我们采用第二种方案（图 3）。

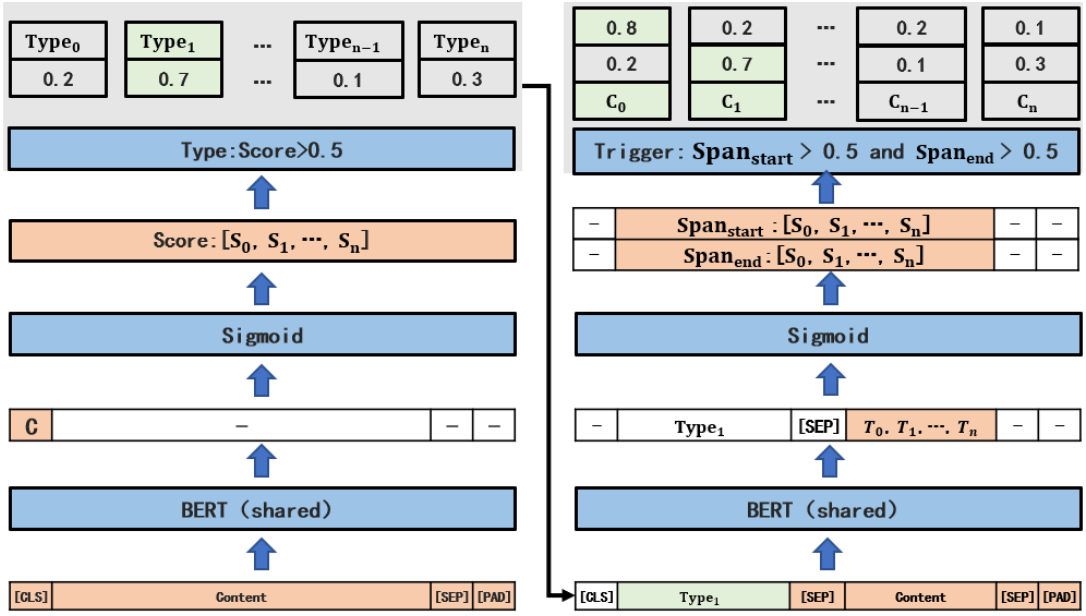

图 3 事件分类及触发词识别

在复赛更换事件类型后，我们发现采用初始预训练重新微调的分数要高于经过初赛微调后的预训练做迁移学习。原因可能是不同的事件描述差异较大，事件类型和触发词密切相关，不同事件对应不同的触发词，导致数据集差异过大。因此，这两个任务我们认为重新做微调会更合适。

3.2 要素识别

抽取 mention 我们最初设计了两种方案：

1. 通过 MRC 的方式，将 type＋trigger＋role 作为 query，利用 sigmoid 计算文本序列首尾指针概率，从而抽取 mention；

2. 通过 MRC-C 的方式，将 type＋trigger 作为 query，计算文本序列首尾指针 role 概率，当首尾最大概率对应的 role 一致时，抽取 mention；

我们发现，方案一在推断过程中会出现抽出其他 role 的 mention。通过查看 task 层的输出也可以发现，query 细微的差异并不足以让输出有足够差异，我们猜测可能因为 query 的构建仅有 role 不同，类似“收购公司”、“被收购公司”这种细微的差异并没有被模型学到，模型更关注 type＋trigger。

因此，我们将 role 的识别转移到 task 层，将方案一的双指针转为 role 的多分类双指针，从而实现方案二 type＋trigger 作为 query 一次性识别出全部的 word＋role（图 4），同时还避免了训练和推断匹配大量 role 需要的计算次数。

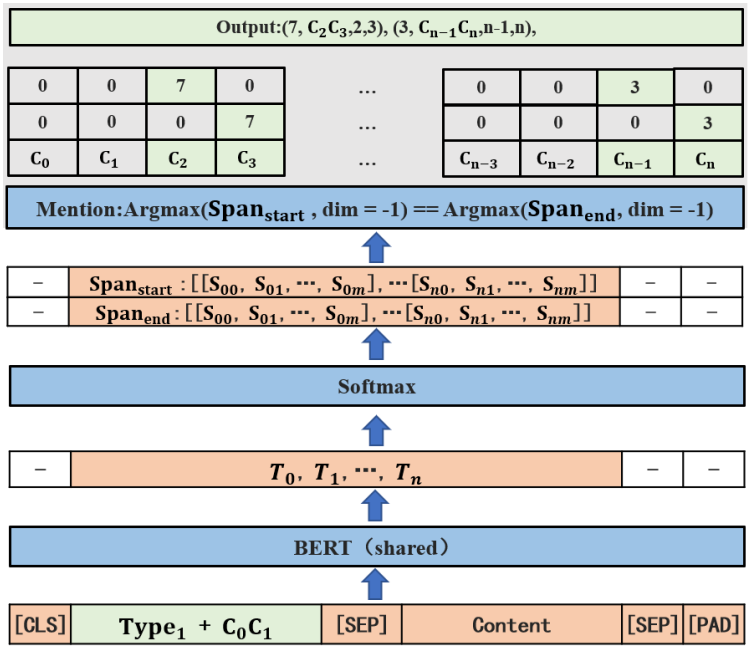

图 4 要素识别

由于 role 在初始五类标签和需要迁移的五类标签存在交集，对比“type＋role”和“role”的分类，我们发现同类型 role 合并后的分数会更高，意味着不同事件、相同 role 的文本描述比较接近，该任务做迁移学习会有一定效果。考虑到联合训练前两个任务不适合做迁移学习，因此针对新的事件类型，我们将训练集 base 中和训练集 trans 相同 role 的数据合并训练。

3.3 模型训练及融合

我们最终采用将事件抽取拆解为文本分类、抽取式阅读理解和多属性的抽取式阅读理解，通过共享预训练模型、联合训练三个子任务的方式实现事件抽取（图 5），使得模型更加合理和高效。完整流程如下：

1. BERT 作为编码器，content 经 BERT 编码后输出为向量；
2. 选择 content 的[CLS]位置，，经过全连接层，得到事件向量；
3. 通过 sigmoid 转为 0-1 的概率，选择概率 p>0.5 的位置索引，从而识别 type；
4. BERT 作为编码器，将 type 作为 query，拼接 content，通过 MRC 的方式，type+content 经 BERT 编码后输出为向量；
5. 抽取序列中 content 的位置向量，经过两个全连接层，输出首尾指针向量；
6. 利用 sigmoid 将 content 序列向量转为 0-1 的概率表示，计算文本序列首尾指针概率，当首尾指针概率 p 均大于 0.5，锁定 type 对应的 trigger；
7. BERT 作为编码器，将 type+trigger 作为 query，拼接 content，通过 MRC-C 的方式，type+trigger+content 经 BERT 编码后输出为向量；
8. 抽取序列中 content 的位置向量，经过两个全连接层，输出首尾指针向量；
9. 计算 content 序列向量中每个字的最大概率 role，当首尾指针的 role 相同，锁定 type+trigger 对应的 mention；

模型输出以 mention 为单位，形式为（type，trigger，role，word，span）的五元组，方便后续模型融合时进行加权投票以及后处理优化。融合的时候通过模型的 F1 进行加权投票，超过阈值的 mention 被保留。如果一条文本的在各个模型的输出完全一致，我们将该条文本和预测结果保存为伪标签，作为半监督学习的伪标签训练集再次训练。

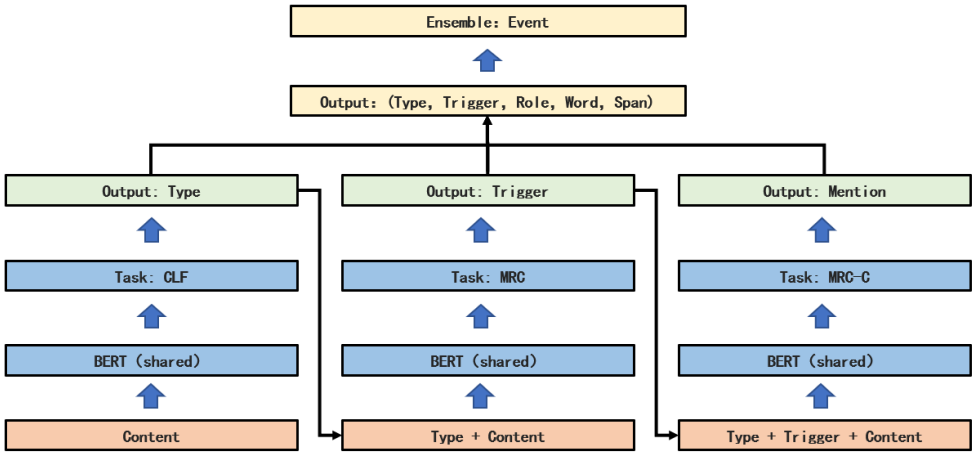

图 5 模型整体结构

训练过程，我们最终采用 RoBERTa-wwm-ext-large[11]作为预训练模型。为了让三个任务的损失在同一个量级，我们均采用 Focal loss 作为损失函数，Focal loss 参数为默认值 gamma=2、 alpha=1，其他超参分别为 optimizer=Adam、lr=1e-5、batchsize=2。考虑到文本长度在 250 的情况下，batchsize 较小模型不易收敛至最优，我们采用梯度累积的方法来扩大 batchsize 至 8。

4 实验结果

为了测试本文所提出的模型在事件抽取方面的效果差异，在使用相同测试样本数据集的基础上，分别采用上述方案以及融合之后的模型进行测试，模型效率（表 4）及得分（表 5）如下。可以发现，预训练采用 large 的效果要优于 base，任务层采用 CLS+MRC+MRC-C 要优于 CLS+MRC+MRC，模型融合较单模可以提升约 4 个百分点。

表 4 模型效率（2080ti）

模型	训练（约 3000 条）	推断（约 33000 条）
BERT-base	0.5 分钟	3 分钟
BERT-large	2 分钟	10 分钟

表 5 模型得分

预训练	模型方案	复赛线上得分
RoBERTa-wwm-ext	CLS+MRC+MRC	0.749
RoBERTa-wwm-ext	CLS+MRC+MRC-C	0.795
RoBERTa-wwm-ext-large	CLS+MRC+MRC-C	0.809
RoBERTa-wwm-ext-large(ensemble)	CLS+MRC+MRC-C	0.849

5 结论

针对比赛任务，本文提出了基于“文本分类（CLF）”+“抽取式阅读理解（MRC）”+“多分类的抽取式阅读理解（MRC-C）”的联合训练方案，实现了从金融领域小样本事件文本中

抽取事件的需求。该方案的三个算法可以独立拆解使用，也可以根据业务场景自由组合，具有较高的灵活性、模型性能和可迁移性，能够胜多种事件抽取场景的集成和部署。最终，融合模型的 F1 值在测试集上达到 0.84943，排名第四。

金融领域的小样本跨类迁移事件抽取，是 NLP 领域一个非常重要的课题，结合多种任务实现联合训练，兼顾效率和精度是业界的发展方向。我们将尝试更多的方案，不断提高模型的整体性能。

参考文献

1. Ruihong Huang, Ellen Riloff. Bootstrapped training of event extraction classifiers. Conference: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (2012)
2. Peifeng Li, Qiaoming Zhu and Guodong Zhou: Joint Modeling of Argument Identification and Role Determination in Chinese Event Extraction with Discourse-Level Information. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (2014)
3. Hongyu Lin1, Yaojie Lu, Xianpei Han, Le Sun, Nugget Proposal Networks for Chinese Event Detection. arXiv:1805.00249v1 (2018)
4. Tongtao Zhang, Heng Ji and Avirup: Joint Entity and Event Extraction with Generative Adversarial Imitation Learning. Data Intelligence (2019)
5. Zhu Wang, Zhiming Liu, Lingyun Luo, et al: A Multi-Neural Network Fusion based Method for Financial Event Subject Extraction. Computers and Software Engineering (2020)
6. Lei Sha, Feng Qian, Baobao Chang and Zhifang Sui, Jointly Extracting Event Triggers and Arguments by Dependency-Bridge RNN and Tensor-Based Argument Interaction. Association for the Advancement of Artificial Intelligence (2018)
7. WenTao Yu, MianZhu Yi, XiaoHui Huang: Make It Directly: Event Extraction Based on Tree-LSTM and Bi-GRU. IEEE Access (2020)
8. Diya Li, Lifu Huang, Heng Ji and Jiawei Han: Biomedical Event Extraction Based on Knowledge-driven Tree-LSTM. Proceedings of NAACL-HLT (2019)
9. Wei Li, DeZhi Cheng, Lei He, et al: Joint Event Extraction Based on Hierarchical Event Schemas from FrameNet. IEEE Access (2019)
10. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL .(2019)
11. Cui, Yiming and Che, Wanxiang and Liu, Ting and Qin, Bing and Wang, Shijin and Hu, Guoping: Revisiting Pre-Trained Models for Chinese Natural Language Processing. Findings of EMNLP .(2020)