

CCKS 2019面向金融领域的事件主体抽取比赛task4:基于BERT的多模型融合的事件主体抽取模型

李振¹, 刘恒¹, 赵兴莹¹, 李毓瑞¹, 秦培歌¹

¹ 民生科技有限责任公司, 北京 101300, 中国
lizhen, liuheng, zhaoxingying, liyurui, qinpeige@mskj.com

摘要. 本文报告了本团队在CCKS 2019面向金融领域的事件主体抽取任务的提交。本文遵循从简单到复杂的思想设计多个学习器, 并基于集成学习的思想将多个学习器进行融合, 其中每个学习器采用编码-解码的思路进行设计。最终, 模型在A榜最高成绩达到0.92620, 排名第11位, 在B榜最高成绩达到0.82816, 排名第17位。

Keywords: 模型融合, 编码-解码, BERT

1 引言

事件抽取是从描述事件信息的文本中抽取出用户感兴趣的事件信息并以结构化的形式呈现出来。它不仅是构建知识图谱的关键任务, 也是自然语言处理的难点和热点问题之一, 在自动文摘、自动问答、信息检索等领域有着广泛的应用。本文所研究的面向金融领域的事件主体抽取任务, 是事件抽取任务中的限定域事件抽取。任务的目标是从一段金融事件文本中抽取出发生特定事件类型的主体, 主体范围限定为: 公司和机构, 事件类型共有21种, 包括交易违规、投诉维权、涉嫌欺诈和财务造假等。

限定域事件抽取是在进行抽取之前, 预先定义好目标事件的类型。限定域事件抽取是信息抽取和知识图谱构建的重要环节之一, 受到了学术界和产业界的广泛关注, 研究者针对该任务进行了大量的研究。目前, 对于限定域事件抽取主要有两种方法: 基于模式匹配的方法和基于机器学习的方法。基于模式匹配的方法是在一些模式的指导下进行某种类型的事件的识别和抽取。由于基于模式匹配的方法依赖于文本的具体形式(语言、领域和文档格式等), 且制定的模式很难覆盖所有的事件类型, 当语料发生变化时, 需要重新获取模式, 因此基于机器学习的事件主体抽取成为主流方法。基于机器学习的事件抽取方法中最具有代表性的是基于有监督学习的事件抽取, 即将事件抽取建模成一个多分类问题, 提取特征向量后再使用有监督的分类器进行事件抽取。本文采用基于有监督学习的方法从金融事件文本中抽取事件主体。

模型的整体设计思路是设计从简单到复杂的多个学习器, 基于集成学习的思想将多个学习器结合起来, 从而获得更好的性能。首先, 以BERT+FC作为基准模型, 添加一层self-attention作为编码器, 构建BERT+self-attention+FC模型。

其次，在上述模型中添加两个1维卷积层，另外为了避免过拟合，增加了残差连接，构建 BERT+Cov.+ self-attention+FC模型。通过对上述模型进行融合来得到最终的模型，这样既避免了直接建模的复杂性，又能提升整体系统的性能。经过模型的训练和调参，单模的F1值最高达到0.8949，进行模型融合之后，F1值提升到了0.9262。

综上所述，本文的贡献总结如下：

本文采用从简单到复杂的思路设计多个模型，并进行融合以提升模型效果；

本文采用编码-解码的设计思路来构建单个模型，从而使每个模型独立而结构类似，可以很好的适应模型融合。

本文其余内容组织如下：第二章介绍方法描述，第三章展示实验结果，第四章总结本文。

2 系统描述

2.1 设计思路

从简单到复杂设计整体系统。法国科学家笛卡尔从逻辑学、几何学和代数学中发现的4条规则之一就是：思想必须从简单到复杂。这条规则不仅适用于以上学科，也适用于本文的方法设计。因此，本文依循简单到复杂的原则来设计解决事件主体抽取的方法。事件主体抽取的任务是非平凡的，面向金融领域的限定域事件主体抽取亦然。根据奥卡姆剃刀原理，若能用简单的方式解决复杂的问题，就没有必要去设计复杂的方案。

单个模型的设计思路是编码+解码。事件识别的复杂性在于对事件类型和事件主体的判断。事件类型多种多样，相应的事件主体也各不相同。因此，本文采用编码-解码的思路来设计深度学习模型，即由编码器自行学习复杂的规则，由解码器解析学到的规则。

简单和复杂模型的融合。模型融合是解决复杂建模问题的有效方法，本文采用集成学习的思路来融合简单和复杂的模型。集成方法来源于下述观察：如果模型之间近似相互独立，则多个模型联合的性能要优于单个模型的。如果一个分类器以55%的概率可以给出正确的结果，这样的分类器只能说是中等水平，但是如果拥有100个这样的分类器，则大多数分类器的结果都正确的概率可以上升到82%。通过集成学习的方式将简单和复杂的模型结果进行融合，可以显著提升最终效果。

2.2 系统设计

Word2Vec和Glove等词向量模型的出现使预训练技术得到了一定的发展，但其优势并不突出。直到BERT (Pre-training of Deep Bidirectional Transformers for Language Understanding) 出现，它进一步增强了词向量模型的泛化能力，充分描述字符级、词级、句子级甚至句间关系特征，刷新了11项NLP任务的最优结果。因此，本文在设计模型时均采用了BERT作为模型的最底层。

遵循从简单到复杂的原则，本文首先设计了一个简单模型作为基准模型。模型的结构是BERT+FC。

BERT+FC模型

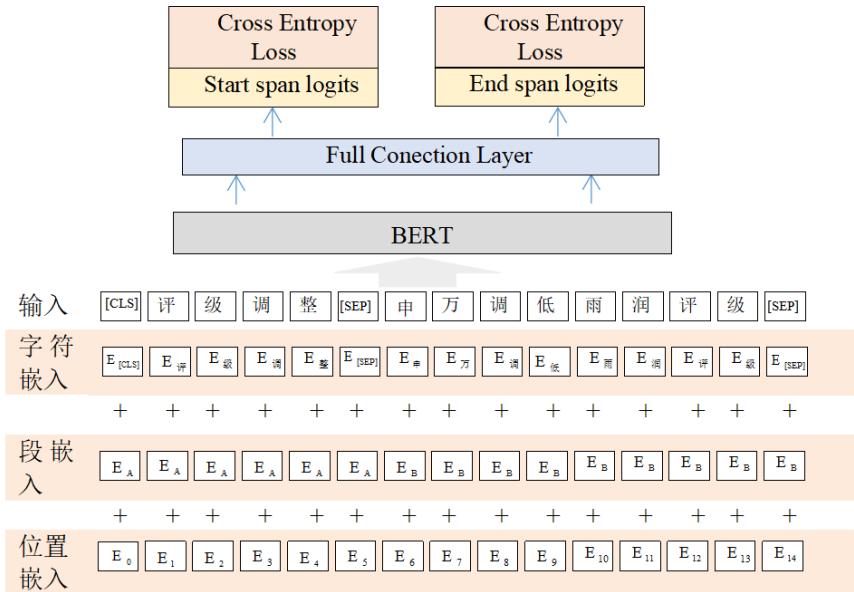


图 1. BERT+FC 模型

BERT+FC模型的结构如图1所示。由BERT作为编码器，一个全连接层作为解码器。每个输入序列由事件类型和事件描述组成，每个输出对应的事件主体是事件描述中的连续子串。模型的输入是通过对相应词块的词块嵌入、段嵌入和位置嵌入求和来构造的。图1的下半部分直观展示了构建模型输入的过程。具体如下：

- 字符嵌入：使用Google开源的预训练BERT中文模型提供的字典进行字嵌入，该字典总共包含21128个字符。其中，没有包含在字典中的字符用[UNK]替代。句子序列由事件类型和事件描述组成，并由特殊字符[SEP]分割，且序列的第一个字符始终是特殊嵌入字符[CLS]；
- 位置嵌入：可以支持的序列长度最多为512个词块；
- 段落嵌入：将事件类型和事件描述拼接成一个序列，事件类型（用A表示）嵌入到序列中，以 E_A 表示；事件描述（用B表示）嵌入到序列，以 E_B 表示。

构建好输入序列后，输入到BERT的预训练网络中，经BERT编码，序列中的每个字被编码成了768维的向量输出。经过BERT网络的输出，输入到一个全

连接层，如图1所示，全连接层的参数个数是输入序列长度的两倍，将全连接层的输出分成两个向量，分别与表示开始位置和结束位置的向量计算交叉熵。

模型中获取事件主体的过程如图2所示。图中 $start_logits$ 和 end_logits 分别对应每个字作为主体的开始字符与结束字符的得分，得分越高，表明越有可能是要提取的事件主体在事件描述中的开始位置和结束位置。 $index$ 是描述文本中每个字符的编号，而 $start_index$ 是排名前20的 $start_logits$ 相对应的编号，同理， end_index 是排名前20的 end_logits 相对应的编号。

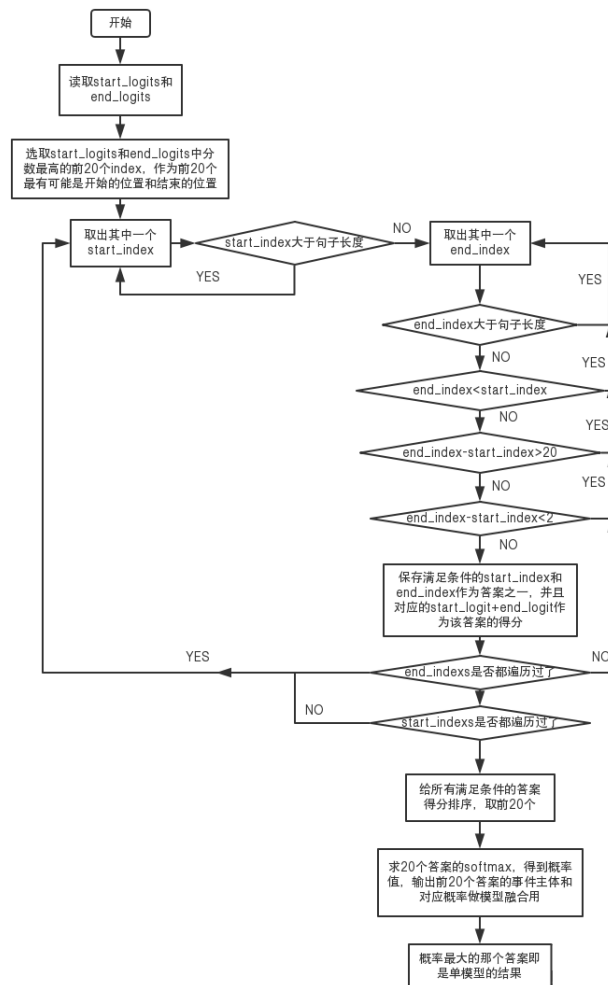


图 2. 获取事件主体的过程

BERT+self-attention+FC模型

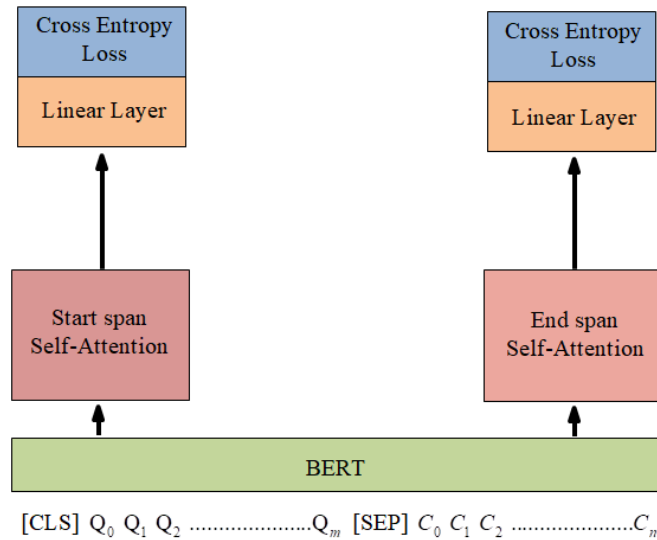


图 3. BERT+self-attention+FC 模型

在BERT+FC模型的基础上，本文设计了BERT+self-attention+FC模型，结构如图3所示，模型由BERT和self-attention作为编码器，全连接层作为解码器。模型输入与图1模型的输入相同，经BERT编码后的输出向量，分别输入到计算开始位置的self-attention层和计算结束位置的self-attention层中，再分别输入到线性连接层，然后分别与表示开始位置和结束位置的向量计算交叉熵作为损失，最后取两者损失和作为最终的损失。

BERT+Cov.+ self-attention+FC模型

在BERT+self-attention+FC的基础上，本文设计了BERT+Cov.+ self-attention+FC模型，模型输入和图1中模型输入的过程相同。如图4，该网络是在BERT+self-attention+FC的基础上添加了两个1维卷积层，使其可以获取上下文的一些信息，丰富网络学习到的参数。另外，为了避免过拟合，添加了一个残差层，将Bert的输出和卷积后的结果相加，然后分别输入到计算开始位置的self-attention层和计算结束位置的self-attention层中。其后的网络结构和BERT+self-attention+FC的结构相同。

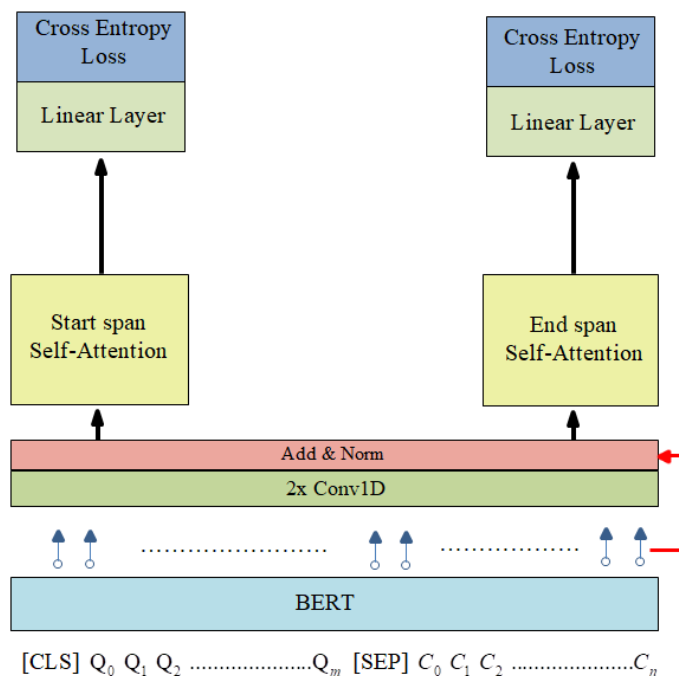


图 4. BERT+Cov.+ self-attention+FC 模型

2.3 模型融合

模型融合通过把从不同角度对问题域建立的模型进行融合来得到最终的模型，避免了直接建模的复杂性。本文设计的单个模型均是输出一个主体，而符合给定事件类型的主体往往不止一个，因此，本文采用的融合方式是融合多个模型的最优答案作为单个样本的最终预测结果。

每个模型对某个样本的预测都可能存在错误，而有些错误是显而易见的，而且存在固定的模式。因此在融合多个模型的多个主体的过程中，针对错误答案做了一定的过滤处理，主要方法有：

(1) 词性标注；调用预训练好的词性标注模型，对每个模型输出结果进行了词性标注。若词性是名词类，则保留此结果；否则，将其过滤掉，进而取其模型的第二第三预测值。

(2) 停用词过滤；将主体中不可能出现的符号作为停用词，如标点符号，特殊动词，特殊名词等，若预测结果出现停用词，则将其过滤。

3 实验结果

为了测试本文所提出的模型在事件主体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，分别采用上述三种模型以及融合之后的模型进行测试。

在调参过程中，测试集在三个模型上达到的F1-score最高分和模型融合后的F1-score最高分如表1所示：

表 1. 不同模型以及模型融合在测试集上的F1-score

模型名称	F1-score
BERT+FC	0.8900
BERT+self-attention+FC	0.8913
BERT+Conv.+self-attention+FC	0.8949
模型融合	0.9262

模型提升效果如图5所示，以BERT+FC作为基准，可以看到，改进后的单个模型的F1-score都有一定的提高。添加self-attention层后的BERT+self-attention+FC模型可以学习到句子内部长距离依赖关系，忽略词之间的距离直接计算依赖关系，能够学习一个句子的内部结构，从而使得F1-score得到提高。BERT+Conv.+self-attention+FC模型在上一个模型的基础上增加了两个1维卷积层，对句子空间相邻位置的关系进行学习，进一步提升了单模型的F1-score。但是单个模型的提升效果仍然有限，经过差异较大的模型融合之后，F1-score提高了3%。

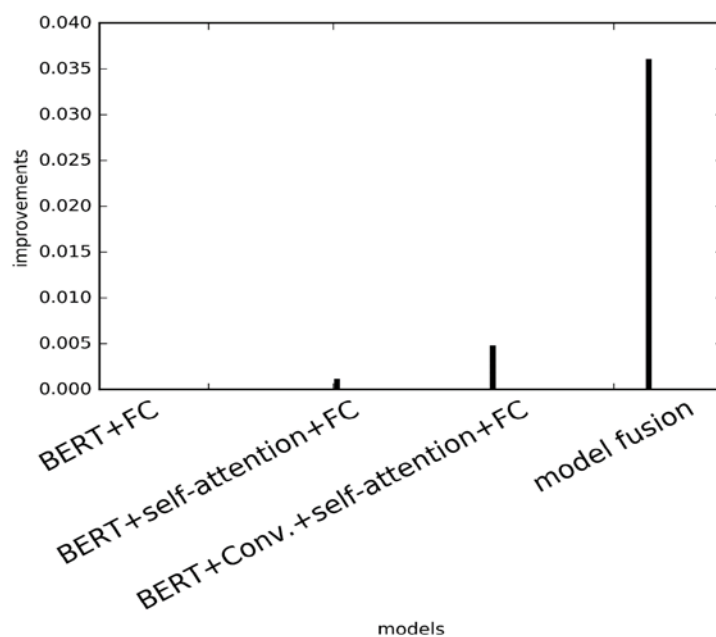


图 5. 模型提升效果

4 总结展望

针对比赛任务，本文提出基于BERT的多模型融合的事件主体抽取模型，实现了从金融领域事件文本中抽取特定事件类型的主体。本文遵循从简单到复杂的设计思路设计多个学习器，并基于集成学习的思想将多个学习器进行融合，最终，融合模型的F1值在最终数据集上达到0.82816，排名17位。

在未来，我们将探索更多关于事件主体抽取的方法，包括使用更多的外部知识和更深层的网络来提高模型性能，我们会在更多具有结构差异的模型上做集成尝试，以寻求性能更好的模型融合方法。

References

1. Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.
2. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[J]. 2017.
3. Cho K, Van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014.
4. Zhou Z H. Ensemble Methods - Foundations and Algorithms[M]. Taylor & Francis, 2012.
5. Rajpurkar P, Zhang J, Lopyrev K, et al. Squad: 100,000+ questions for machine comprehension of text[J]. arXiv preprint arXiv: 1606.05250, 2016.
6. Radford A, Narasimhan K, Salimans T, et al. Improving language understanding with unsupervised learning[J]. Technical report, OpenAI, 2018.