

用于金融文本挖掘的多任务学习预训练金融语言模型

刘 壮¹ 刘 畅² Wayne Lin³ 赵 军⁴
¹(东北财经大学应用金融与行为科学学院 辽宁大连 116025)
²(中国石油物资采购中心 沈阳 110031)
³(南加州大学计算机学院 美国加利福尼亚州洛杉矶 90007)
⁴(IBM 研究院 北京 100101)
(liuzhuang@dufe.edu.cn)

Pretraining Financial Language Model with Multi-Task Learning for Financial Text Mining

Liu Zhuang¹, Liu Chang², Wayne Lin³, and Zhao Jun⁴
¹(School of Applied Finance and Behavioral Science, Dongbei University of Finance and Economics, Dalian, Liaoning 116025)
²(China Petroleum Materials Procurement Center, Shenyang 110031)
³(School of Computer Science, University of Southern California, Los Angeles, CA, USA 90007)
⁴(IBM Research, Beijing 100101)

Abstract Financial text mining is becoming increasingly important as the number of financial documents rapidly grows. With the progress in machine learning, extracting valuable information from financial literature has gained attention among researchers, and deep learning has boosted the development of effective financial text mining models. However, as deep learning models require a large amount of labeled training data, applying deep learning to financial text mining is often unsuccessful due to the lack of training data in financial fields. Recent researches on training contextualized language representation models on text corpora shed light on the possibility of leveraging a large number of unlabeled financial text corpora. We introduce F-BERT (BERT for financial text mining), which is a domain specific language representation model pre-trained on large-scale financial corpora. Based on the BERT architecture, F-BERT effectively transfers the knowledge from a large amount of financial texts to financial text mining models with minimal task-specific architecture modifications. The results show that our F-BERT outperforms most current state-of-the-art models, which demonstrates the effectiveness and robustness of the proposed F-BERT.

Key words BERT; financial text mining; multi-task learning; pre-training; transfer learning; fintech

摘 要 近年来,机器学习,尤其是深度神经网络方法促进了金融文本挖掘研究的发展,在金融科技(Fintech)领域起着越来越重要的作用.如何从金融大数据中提取有价值的信息已经成为学术界和工业界一个非常有挑战的研究.由于深度神经网络需要大量标注数据,但是金融领域缺乏训练数据,因此,将深度学习应用于金融大数据挖掘并没有取得很好效果.为了更好地解决该问题,利用自监督学习和多任

务学习等深度学习领域最新的思想和技术,提出了基于 BERT 模型架构的开源金融预训练语言模型 F-BERT. F-BERT 是一种针对特定领域(金融文本数据)的预训练语言模型,它通过在大型金融语料库上进行无监督训练得到.基于 BERT 架构, F-BERT 可以有效地自动将知识从金融大数据中提取出并记忆在模型中,而无需进行特定于金融任务的模型结构修改,从而可以直接将其应用于下游各种金融领域任务,包括股票涨跌预测、金融情绪分类、金融智能客服等.在金融关系提取、金融情绪分类、金融智能问答任务上的大量实验表明了 F-BERT 模型的有效性和鲁棒性.同时, F-BERT 在这 3 个有代表性的金融文本挖掘任务上均取得了很高的模型准确率,进一步验证了模型的性能.

关键词 BERT;金融文本挖掘;多任务学习;预训练;迁移学习;金融科技

中图法分类号 TP181

海量的互联网金融信息在金融市场中有着举足轻重的地位,对网络金融文本信息的挖掘工作具有很大的实际价值.随着大数据时代的到来,金融大数据挖掘已成为行业热点趋势,面向金融的机器学习技术吸引了越来越多的关注.面对每日产生的数量惊人的金融文本数据,如何从中提取有价值的信息已经成为学术界和工业界一个非常有挑战的研究.如果我们采取人工的方式来分析这些文本信息并从中获得可行的见解几乎是一项极其艰巨的任务.机器学习技术的进步使金融科技中的金融文本挖掘模型成为可能.但是,在金融文本挖掘任务中,构建有监督训练数据代价非常高昂,因为这需要使用财务领域的专家知识.由于可用于金融文本挖掘任务的有标签训练数据量很少,因此大多数金融文本挖掘模型无法直接利用深度学习技术.

在本文中,我们创新地提出了 F-BERT 模型,通过利用自监督学习和多任务学习的深度神经网络方法来解决该问题.当前,金融科技中的金融文本挖掘模型主要是采取基于深度学习(deep learning)的自然语言处理(natural language processing)技术.

目前,自然语言处理主要使用基于深度神经网络的技术,其发展主要有两大里程碑工作.首先是 2013 年提出并不断发展的以 Word2Vec^[1]为代表的词向量技术,例如 Word2Vec, GloVe^[2]等;第 2 个里程碑是在 2018 年底由谷歌提出的以 BERT(bidirectional encoder representations from transformers)^[3]为典型代表的预训练语言模型(pre-training language models)技术,例如 BERT, ELMO^[4]等.其中,诸如 Word2Vec, GloVe 之类的词编码是从无监督语料库中提取知识的一种方式,已成为自然语言处理的主要进步之一.但是,由于在金融领域中包含了很多专业术语,因此这些简单的词向量方法不够有效.另一方面,预先训练的语言模型技术,例如 BERT,

ALBERT^[5]等,采取在大规模无监督数据(例如维基百科数据等)上进行了预先训练,经过大规模语言模型预训练, BERT 获得了有效的上下文表示.与 Word2Vec 词向量相比, BERT 能够充分利用深度预训练模型的参数,可以更有效地学习上下文知识.但是,模型的预训练(例如 BERT)主要使用基于简单的预训练任务来对语言模型进行训练,从而使得模型具备掌握单词或句子共现的能力.实际上,除了共现之外,还有其他词汇、句法和语义信息需要在训练语料库中进行检查.尤其是对于金融文本数据,例如,股票、债券类型和金融机构名称之类的命名实体包含唯一的词汇信息.此外,诸如句子顺序和句子之间的接近度之类的信息也使模型能够学习语义感知表示.并且, BERT 的预训练数据来自 Books Corpus 和英文维基百科,这些语料不一定和目标任务的语言环境相近,如金融领域.因此,如果我们直接使用 BERT 进行金融文本挖掘,效果并不是很好.

为了在大型金融训练语料库中有效地捕获语言知识和语义信息,我们构建了涵盖更多知识的 4 个自监督学习的预训练任务,并通过对训练数据进行多任务学习来训练 F-BERT.具体而言,我们创新地提出了 F-BERT 模型,在该模型架构中,我们构建了 4 个自监督学习预训练任务,并同时与普通文本语料和金融领域文本语料进行了学习,从而使得 F-BERT 可以更好地捕获金融文本数据的语言知识和语义等信息.总结来说,本文的主要贡献有 5 个方面:

- 1) 利用深度学习领域的自监督学习和多任务学习技术,提出了一种基于 BERT 模型架构的大规模语料上训练的开源金融预训练模型 F-BERT.

- 2) 构建了 4 个自监督学习预训练任务,可以通过多任务自监督学习来进行同时预训练,通过该机制可以有效地捕获大规模预训练语料中的金融语言知识和语义等信息.

3) 分别在金融关系提取、金融情绪分类、金融智能问答任务上进行了金融任务实验,验证了 F-BERT 的有效性和鲁棒性.实验结果表明 F-BERT 模型在这 3 个有代表性的金融文本挖掘任务均取得了最佳准确性,优于所有当前其他模型性能;另外,针对金融命名实体识别、金融短文本分类这 2 个实际金融科技任务也进行了实验验证,所提出的 F-BERT 模型均取得了最佳准确性.

4) 采取了一种混合精度训练方法,并在 Horovod 框架上进行 F-BERT 的分布式训练,既使得整个训练过程具有稳定加速比,同时拥有较好的可扩展性.

5) 在 Github 上进行了 F-BERT 的开源,包括其模型架构、源代码、超参数、训练模型,以及用于微调的源代码.通过该开源代码,F-BERT 可用于各种其他下游金融文本挖掘任务,对特定金融文本分析任务进行最少的修改即可帮助实现最新的性能.

1 相关工作

1.1 预训练模型

预训练模型最早应用于机器图像和计算机视频领域.在 2012 年的大规模图像识别竞赛 ImageNet^[6]中,当时取得第一名成绩的 AlexNet^[7]横空出世,它采取了基于 CNN^[8]的卷积神经网络模型架构.此后,AlexNet 被广泛应用在众多的机器视觉任务中.虽然很多新模型并不是完全借鉴 AlexNet 模型架构从头开始训练,而是主要利用在 ImageNet 上训练得到的 AlexNet 模型的参数和神经网络架构,再进一步结合实际任务进行少量修改,然后在新的数据上训练和优化.实验结果表明,复用预训练模型可以显著增加目标任务的准确度,也大大缩短了新模型的训练时间.

预训练模型的一大优势是,可以利用在其他相似任务的大量数据上建立有效的模型再迁移到目标任务,从而解决了目标任务(例如股票涨跌预测)训练数据不足的问题.另外,从垂直领域文档(例如金融文本数据等)中提取有价值的信息正在变得越来越重要,深度学习促进了有效的垂直领域文本挖掘的研究发展.

1.2 预训练语言模型 BERT

Google 在 2018 年下半年发布了预训练语言模型 BERT^[3],在自然语言处理界具有开创时代的意义.接下来,BERT 模型在十多个自然语言理解任务中全部取得了最佳模型准确率,在多个公开自然语

言处理竞赛上取得大幅提升,甚至在常识推理、自动问答、情绪分析、关系抽取、命名实体识别等竞赛均取得了超过了人类准确率的成绩.同时,在 BERT 模型的源码发布之后,后续研究人员基于开源代码和预训练模型进行了各种自然语言处理任务,进一步大幅提升了各个 NLP 任务模型的成绩.例如,在 SQuAD^[9-12]竞赛排行榜上排名前 20 名的模型均选择使用 BERT 模型,且最好成绩超过人类水平;在 CoQA^[13-15]竞赛排行榜上,排名前 12 名的模型也全部基于 BERT 模型,排名第一的模型也同样取得了超过人类水平的成绩.

BERT 本质上也是一种多层 Transformer^[16]结构.它的输入是一段文本中每个单词的词向量(分词由 Word Piece 生成),输出是每个单词的 BERT 编码.单词的 BERT 编码表示一般包含了该单词及其上下文的信息.BERT 模型采用了 2 个预训练任务:双向语言模型和判断下一段文本.这 2 个任务均属于无监督学习,即只需要文本语料库,不需要任何人工标注数据.

1) 掩码机制

BERT 提出了掩码机制.BERT 在一段文本中随机挑选 15% 的单词,以掩码符号[MASK]代替.然后,利用多层 Transformer 机制预测这些位置的单词.由于输入中没有被掩去的单词的任何信息,这些位置上的 Transformer 输出可以用来预测对应的单词.因此 BERT 是一个双向语言模型.但是,由于原文本中并不包含掩码[MASK],从而使得预训练任务与真正目标任务会出现不一致的情况.为了更好地缓解这个问题,BERT 采取了在选取被掩单词后以 80% 的概率替换成[MASK],同时以 10% 的概率替换成一个随机单词,以 10% 的概率保持原单词的方法.最终实验也证明该方法可以有效地提升目标任务的准确性.

2) 判断下一段文本机制

BERT 的第 2 个预训练任务是二分类问题:给定 2 段文本 A, B ,判断 B 是否是原文中 A 的下一段文本.为了尽可能多地考虑上下文,文本 A 和 B 的长度总和最大为 512 个词.训练中,50% 的正例来自原文中紧挨着的 2 段文本,50% 的负例来自 2 段无关联的文本.由于 Transformer 结构只接收一段文本输入,BERT 将 A 和 B 拼接起来,并加上起始符号[CLS]和分隔符[SEP].为了使模型区分文本 A 和 B ,还加入了段编码,即给文本 A 和文本 B 中的单词分配不同的编码.设起始符[CLS]位置的

BERT 编码为 CLS,则模型预测文本 B 是文本 A 的下一段文本的概率.判断下一段文本的预训练任务属于分类问题,提高了预训练阶段与微调阶段任务的契合度.这也是 BERT 取得比其他预训练语言模型更优秀结果的原因之一.

3) BERT 预训练

BERT 的预训练数据来自公开语料库 Books Corpus(共 8 亿个词)和英文维基百科(共 25 亿个词)^[3]. BERT 公开的预训练模型有 BASE 与 LARGE 两种.

① BERT_{BASE}:12 层 Transformer,输入和输出维度为 768,注意力含 12 个 Head,共 1.1 亿个参数.

② BERT_{LARGE}:24 层 Transformer,输入和输出维度为 1 024,注意力含 24 个 Head,共 3.4 亿个参数.

其中,2 个模型均训练了 40 轮.BERT_{BASE} 在 4 个 Cloud TPU 上训练,BERT_{LARGE} 在 16 个 Cloud TPU 上训练,均花费了 4 天左右的时间.特别地,相比于 GPU 来说,针对深度学习训练 TPU 能够进行更好的硬件和算法的优化.在 4 个 GPU 上训练 BERT_{LARGE} 需要花费初略计算约 100 天时间.因此,BERT 语言模型被提出时是当时耗费计算资源最多且模型规模最大的自然语言预训练模型.

2 F-BERT 模型

如图 1 所示,我们利用深度学习领域的自监督学习和多任务学习技术,创新地提出了基于 BERT 模型架构的垂直领域预训练语言模型:金融预训练语言模型,我们将其命名为 F-BERT 模型.F-BERT 是同样采取预训练(pre-training)加微调(fine-tuning)的两阶段架构.其中,在模型预训练阶段,F-BERT 与传统的 BERT 预训练不同之处在于,与其使用少量的预训练目标进行训练,不如考虑同时引入多种预训练任务来尝试更好地帮助模型进行有效地学习训练.因为除了语言共现信息之外,在训练语料中实际上还涉及到多方面的语言知识,例如词法、句法以及更高层次的语义和语用等知识.同时,也还包括金融实体等词语概念的知识、文本语义相似度等语义知识.最重要的是,F-BERT 会通过多任务学习来不断地更新预训练模型.并且,与现有的预训练模型相比,F-BERT 同时在通用语料库和金融领域语料库上进行训练.在微调阶段,首先使用预训练的参数初始化 F-BERT,然后使用特定于任务的监督数据进行微调训练.

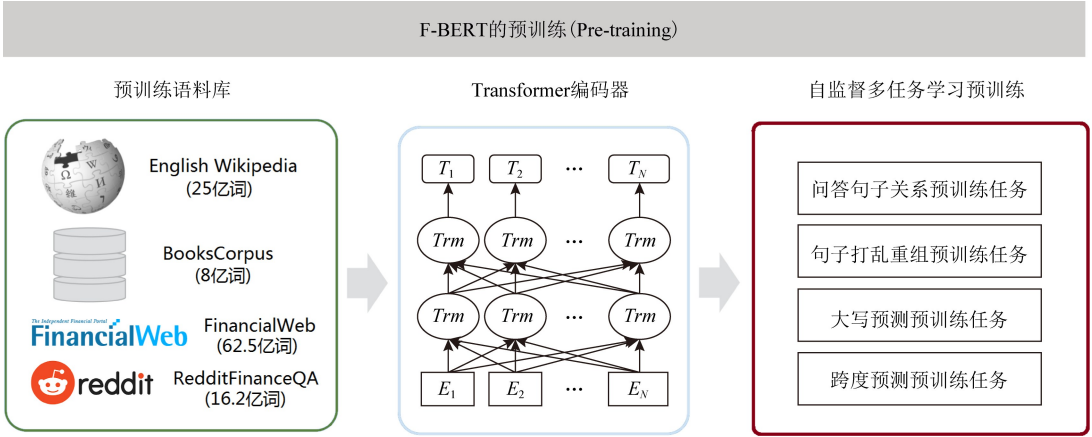


Fig. 1 An illustration of the architecture for F-BERT
图 1 F-BERT 模型的预训练架构

本节将对 F-BERT 模型架构进行详细描述.

2.1 编码器

在预训练的阶段,我们使用 Transformer 编码器,采取和 BERT 类似的编码方法,即位置编码(position embedding)、段编码(segment embdding)和令牌编码(token embedding),但除此之外我们额外设计了一个任务编码(task embedding).针对不同的预训练任务,我们使用任务编码来精细化地建模

不同类型的任务.对于 N 个任务,任务的 ID 范围就是 $1 \sim N$,每个任务 ID 都会被映射到不同的编码上,最终输入编码的输入为:输入编码=位置编码+段编码+令牌编码+任务编码.具体如图 2 所示.

接下来,针对输入编码计算注意力,我们使用 Transformer 的多头自注意力机制进行计算.它主要由查询 Query(使用 Q 来代替),键 Key(使用 K 来代替)和值 Value(使用 V 来代替)作为输入,然后

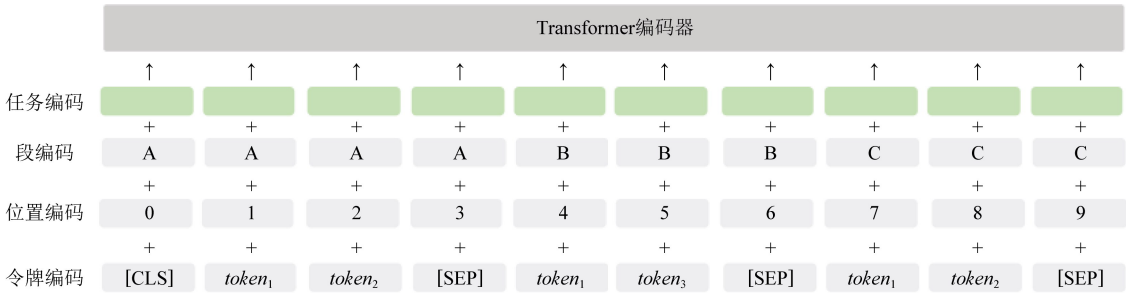


Fig. 2 Transformer encoder layer

图 2 Transformer 编码层

使用线性变换对 Q, K, V 进行投影,并且可以同步计算多次,即使用多头 h 来表示 h 个不同的线性变化.最后再将全部的注意力值结果进行拼接,从而完成一次多头注意力计算.对于单个 Q ,注意力功能的输出是 V 的加权组合.为了简化计算,在这里我们将自注意力取相同的 Q, K, V 值,并且注意力计算使用的是缩放点积.具体注意力函数为

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

其中, d 表示 Q, K, V 向量维度.

2.2 自监督多任务学习预训练任务

目前预训练阶段主要是利用句子或词的共现信号设计不同任务进行语言模型的预训练.例如,原始 BERT 模型构建了 2 个预训练(遮蔽语言模型任务和下一句预测任务);XLNet^[9]模型利用全排列的语言模型进行自回归方式的预训练.相比较之下,我们在 F-BERT 模型中堆叠了大量的预训练目标.就好像人类进行外语考试,在卷子上面有多种不同的题型,如果可以进行综合训练,那么对整体学习提升一定有很大帮助.具体来说,在预训练阶段, F-BERT 构造了 4 个自监督学习的预训练任务,并通过多任务学习方式从训练文本语料中学习不同层面的知识.如图 1 所示,这 4 个自监督的预训练任务分别为:跨度替换预测预训练任务、大写预测预训练任务、句子打乱重组预训练任务,以及问答句子关系预训练任务.

1) 跨度替换预测预训练

我们采取了一种分词级别的预训练任务,实现跨度替换预测预训练. BERT 模型的输入主要是以字为单位进行切分,这样就不能够更好地考虑到共现单词和共现词组间的关系,进一步导致不能充分学习其所包括对隐含先验知识.针对金融领域文本语料该问题就会更进一步被放大,从而降低

模型的学习效果.通过借鉴 Mandar 等人提出的 SpanBERT^[17]思想,我们对谷歌原生的 BERT 模型进行了 2 方面改进:①不再对单个随机词定义掩码,而是定义了一种更好的跨度掩码方案,即采取随机对一定窗口范围的邻接词来设计掩码;②不再对分词内单个词的表示过多依赖,而是定义一种分词边界表示来预测被添加掩码的分词内容,从而实现跨度替换预测的训练目标,最终我们可以更充分地学习到领域性更强的金融领域文本知识,提升 F-BERT 模型性能.

关于跨度替换预测预训练任务,具体来说,首先,我们迭代地采样文本序列,即针对每一个单词序列 $X = (x_1, x_2, \dots, x_n)$,通过定义一个掩码比例阈值(例如整个序列的 20%)来构造一个序列子集 Y .在 F-BERT 训练任务中,我们采取基于几何分布来随机选择一定长度的文本.由于几何分布(geometric distribution)是一种离散型概率分布,在它的第 m 次伯努利试验中,试验 k 次才得到第 1 次成功的机率,因此我们能够均匀地(随机地)得到分词起点并可以得到一个较短的序列.为了获得更佳的采样长度分布, F-BERT 模型中的几何概率分布超参数 $P = 0.18$,最大长度上限 $T = 16$ (超出部分按照丢弃处理),通过实验我们得到的最佳平均序列长度是 4.6.接下来,我们具体实现了该实现跨度替换预测预训练任务.在 F-BERT 训练过程中,我们将跨度定义为取跨度前后边界的 4 个词,如果这些词没有在跨度范围内,就使用这 4 个词的向量再加上跨度中被遮盖掉词的向量来预测原来的单词.具体实现就是使用一个 2 层的前馈神经网络,并使用层正则化,其中的激活函数我们使用 $ReLU$,从而将编码向量和位置向量拼接起来:

$$h = \text{LayerNorm}(\text{ReLU}(W_1 X)),$$
$$f = \text{LayerNorm}(\text{ReLU}(W_2 h)).$$

我们同样使用交叉熵作为损失函数,就是跨度替换预测预训练目标的损失来用于模型训练.

为了更直观地理解跨度替换预测训练,图3中给出一个具体例子,其中使用了跨度替换预测预训练任务.假设输入序列为“股价预测很困难”,以“很”字为例,我们首先随机掩盖一个连续的跨度(图中的 x_4 到 x_6),针对待预测字的位置信息和基于该跨度前后 S 个字的表示(这里 $S=2$)来进行最终预测,这个过程就是跨度替换预测训练任务,同理,最终的“测”和“困”也使用同样的跨度替换方法来进行预测.在跨度替换预测过程中,对每个文本序列相当于进行多次验码,因此任意字都会成为跨度替换预测连续内容,这样就实现了各个字的表示均会包含周围跨度的信息,从而将跨度前后字的知识体现到跨度中的字的语义信息中.针对抽取类任务中(例如金融文本边界检测,跨度和起止位置的知识对模型而已非常重要),我们通过对预训练语料的不断训练,最终就可以使得 F-BERT 模型在这样的抽取类任务中取得很好的效果.

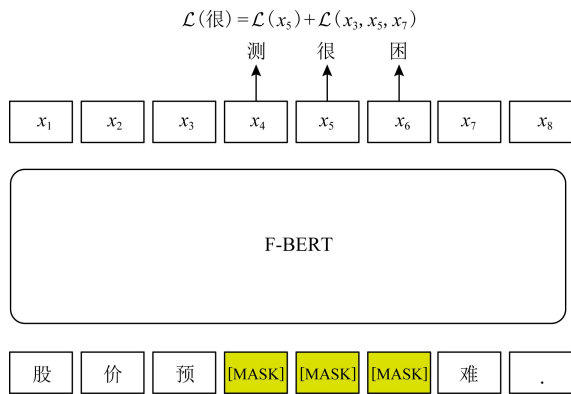


Fig.3 An example of span boundary prediction task
图3 跨度替换预测预训练示例

- 2) 大写预测预训练任务
- 大写预测预训练任务主要是由 F-BERT 模型预测一个字母是不是大写.这个对特定的任务例如命名实体识别比较有用,与句子中的其他单词相比,大写单词通常具有某些特定的语义价值,尤其对于金融文本处理,其中包含很多专业命名实体,因此该任务可以发挥很大作用.同时,针对中文的训练语料,我们还额外定义了一个预测某个词组是不是专有短语的缩写训练任务,这个针对中文的金融文本处理作用较大.
- 3) 句子打乱重组预训练任务
- 句子打乱重组预训练任务主要是进行顺序还

原,就是将输入文本序列随机打乱,然后通过该任务将其还原回来.受到 ALBERT 和 T5^[18] 预训练任务的启发,我们提出一种更优的重组任务.具体来说,首先把一篇文章随机分为 $n=1$ 到 k 份,对于每种分法都有 $n!$ 种组合,然后我们再让模型去预测这篇文章是第几种,本质相当于构造了一个多分类的问题,通过该任务就能够让模型学到句子之间的顺序关系.

- 4) 问答句子关系预训练任务
- 问答句子关系预训练任务主要是判断 2 个句子是否属于问答对.问答或者对话的数据对语义表示很重要,因为对于相同回答的提问一般都是具有类似语义的,通过修改段嵌入,使之能够适用多轮问答或多轮对话的形式.具体:定义 3 个句子的组合 [CLS] Sentence1 [CLS] Sentence2 [SEP] Sentence3 [SEP] 的格式,这种组合可以表示多轮问答和多轮对话,例如 QRQ, QRR.在这里, Q 表示“问题”文本, R 表示“答案”文本.

3 实验分析

本节我们首先给出预训练数据集的构建;然后在多个金融任务上将 F-BERT 与最新的预训练模型进行了性能比较;最后进行了详细的模型分析,包括预训练对模型性能的影响、少量预训练数据进行预训练分析和预训练语料库的大小的讨论等.

3.1 预训练数据集

- 为了更好地训练 F-BERT 模型,我们主要构建了基于通用领域和金融领域 2 部分的训练语料.其中,通用领域训练数据集与 BERT 模型训练相似,使用来自 Wikipedia 和 Book Corpus 的英语语料库.为了将 F-BERT 模型更好地应用于金融文本,我们还构建了金融领域训练数据集,通过在金融网站上爬取各种金融文本数据,包括金融新闻和对话等.如表 1 所示,我们具体构建了 4 个大小不同、领域不同的英语语料库,总计超过 40 GB 大小的文本(约 108.4 亿个词):
- 1) 通用英语训练语料库 Book Corpus 和英语 Wikipedia.即用于训练 BERT 的原始数据,共计 13 GB(约 33.1 亿词)文本.
- 2) Financial Web 金融文本数据集(共 23 GB,约 62.5 亿词).主要是从 Common Crawl News 数据集中进行收集的,其中包含 2012-01 至 2019-12 期间

的 1300 万条财经新闻以及爬取自 FINWEB 网站的金融文章.

3) Reddit Finance QA 金融问答文本数据集 (共 4 GB,约 12.8 亿词).主要从 Reddit 网站上收集有关金融问题然后构建一个问答训练语料.

表 1 中报告了 F-BERT 全部预训练数据的统计信息.我们已经建立并维护了一个开放的金融文本语料库并进行了开源处理,任何人都可以访问和利用该资料库资源.

Table 1 List of Pre-training Corpora Used for F-BERT
表 1 F-BERT 预训练文本语料

无监督语料库	大小/GB	词数量	领域
English Wikipedia	9	25.0 亿	通用
Books Corpus	4	8.1 亿	通用
Financial Web	23	62.5 亿	金融
Reddit Finance QA	4	12.8 亿	金融

3.2 分布式预训练

F-BERT 的预训练过程对整个计算力有着非常大的要求,我们主要采取基于 AI Framework On YARN 分布式框架来进行 F-BERT 预训练.该训练框架是使用 Horovod^[19]的分布式训练方案,整个架构基于 YARN 实现的 16 张 NVidia Tesla v100 GPU 卡进行训练,框架在调度上非常灵活,同时还具有调度作业高伸缩性和高容错力.Horovod 是由 Uber 发布的开源深度学习工具,它是一个分布式深度学习框架.Horovod 库为分布式训练提供了帮助,可以使分布式深度学习更易于使用且训练更高效,该库集成了百度 Ring Allreduce 和 Facebook 的一小时训练 ImageNet 论文的优点.Uber 官方发布的 Horovod 包括 TensorFlow 和 Horovod 两种分布式方案,根据其架构特点,GPU 的数量越多,Horovod 性能损失越小,要优于分布式 TensorFlow,甚至训练速度可以达到 TensorFlow 的 2 倍;同时,即使 GPU 卡数量达到几十张,仍可以保证稳定的加速比.我们使用混合精度的训练方式来训练 F-BERT 模型,以便可以更好地提高预训练效率.

目前的神经网络模型,它们的训练过程基本都使用双精度或单精度的数据类型.但由于显存大小的限制,如果神经网络模型很深且比较复杂时通常会使得处理的 batch 较小.如果调整 batch 的大小,会降低数据的处理效率,训练速度会受到影响,并且一旦 batch 很小的话,也会造成模型训练不稳定并影响模型整体性能.2018 年,百度在 ICLR 提出了一

种单精度和双精度的混合精度训练,该方法可以减少显存消耗并加速训练,通过在语言模型任务、语言识别任务和图像检测任务上进行了有效性验证,均表现非常高效.通常而言,Nvidia 的 Pascal 系列显卡或新一代 GPU 架构 Volta 系列在单精度和底精度计算方面都有非常好的支持,例如,NVIDIA Tesla P40 硬件支持 INT8 计算加速,NVIDIA Tesla V10 支持 FP16 计算加速,同时,就计算峰值而言,相比单精度浮点,低精度浮点计算的峰值要高得多.因此,本节提出的 F-BERT,在实验中采用了混合精度训练方式.具体而言,F-BERT 采用的是 FP16 和 FP32 的混合精度训练.其中,模型的计算过程使用 FP16 精度进行加速,而权重参数会以 FP32 数据格式进行存储、FP32 格式进行参数更新,从而使得模型兼具了 FP16 的速度和 FP32 的稳定性.因此,F-BERT 的训练过程既减少了缓存的开销也加速了训练,同时因为 FP16 的速度优势,使得 F-BERT 的整体预训练和推理都有了一定提升.另外,由于神经网络的训练梯度一般会超过 FP16 的范围,为了控制梯度使其在反向传播过程中在其表示范围内,在训练过程中,我们针对 F-BERT 特别使用了 Loss Scaling 策略控制来对损失的控制,对其梯度进行了放大的处理.

3.3 微调任务

微调(fine-tuning)是机器学习和深度学习中常用的一种调参方法.在实际中,考虑到数据不够大,通常很少且很难从头开始重新进行神经网络模型的训练.通常采取的办法是探索如何利用已有的神经网络,然后在其基础上结合具体任务特点进行进一步训练处理.通常,预训练语言模型 BERT 包括 2 个连续的步骤:一个是预训练阶段(pre-training),另外一个则是微调阶段(fine-tuning).它首先在预训练阶段对大型语料库进行无监督预训练,接下来在微调阶段再对下游的自然语言处理任务进行有监督微调.与 BERT 相似,从头开始在这些大型无监督语料上训练 F-BERT 模型,并将其微调/应用于各种下游有监督的金融文本处理任务.在这里,我们分别在金融关系提取、金融情绪分类、金融智能问答任务这 3 个有代表性的金融文本挖掘任务上进行了微调处理.

1) 金融文本边界检测

金融文本关系抽取常被称作金融文本边界检测,它是金融文本挖掘的基本任务.由于句子是自然语言的基本单位,检测句子的开头和结尾,或者句子边界检测是许多自然语言处理任务(例如词性标记、

信息提取等)非常基础的一步.在金融领域,招股说明书等文献中包含了投资方式等投资基金详细信息,金融句子边界检测任务,就是指如何将嘈杂的非结构化文本转换为半结构化文本,从而实现通过检测招股说明书的开始和结束边界来提取句子的细分.2019年,IJCAI19会议发布了FinNLP金融检测挑战赛数据集FinSBD Shared Task^[20-22].本节使用FinSBD Shared Task数据集作为金融句子边界检测任务的有监督微调数据.FinSBD-2019数据集包含已自动分段的金融文本,它的训练集和开发集中有953个不同的开始标记和207个不同的结束标记.

2) 金融情绪分析

金融情绪分析是金融科技中一项基本任务,也是金融文本分析最基本的任务之一.金融情绪分析主要是指根据给定金融文本数据,检测文本的目标情绪得分.本节使用2个金融情绪分析数据,分别是Financial PhraseBank和金融情绪分析挑战数据集(FiQA Task1)^[23].Financial PhraseBank数据集包含了4845个英语单词,这些文本是从LexisNexis数据库中发现的财经新闻中随机挑选的,然后由具有金融和商业背景的16个专家进行了标注处理.金融情绪分析挑战数据集(FiQA Task1)包括2种类型:金融新闻标题和金融微博,分别带有人工标注的目标实体和情感得分.其中,金融新闻标题数据集总共包含529个带标注的标题样本(训练集为436个样本,测试集为93个样本),而金融微博包含总共774个带标注的帖子样本(训练集为675个样本,测试集为99个样本).

3) 金融智能问答

金融智能问答是金融科技的一项具有挑战的任务,其目的是自动提供与给定短文本或文章相关的问题的答案.2018年,WWW18会议发布了2个金融文本处理数据集^[23-24],分别是金融情绪分析挑战数据集(FiQA Task1)、金融金融智能问答挑战数据集(FiQA Task2).这里使用FiQA Task2的数据集作为金融金融智能问答任务的有监督微调数据.FiQA Task2数据集是通过抓取Stack Exchange网站在2009—2017年间所有涉及“投资”主题的帖子内容人工构建而成的.例如,类似这样的问题“Why are big companies like Google or Apple not included in the Dow Jones Industrial Average (DJIA) index?”.最终FiQA Task2数据集包含了57640个答案集,其中包含了17110个“问题-答案”用于训练集和531个“问题-答案”用于测试集.

3.4 实验结果

1) 金融文本边界检测任务

在IJCAI19金融检测挑战赛数据集FinSBD Shared Task测试集上提交方法以进行评估,详细内容参见表2.在表2中我们可以看到,针对测试集上的准确率,所提出的F-BERT_{LARGE}和F-BERT_{BASE}优于模型发布时的其他方法,尤其取得最佳结果的F-BERT_{LARGE}模型获得了0.93的ES(end of sentence)得分,0.95的BS(beginning of sentence)得分和0.938的MEAN(平均值)分数,均为评估指标的最佳成绩.正如实验结果所看到,对通用领域语料库和金融领域语料库进行预训练的F-BERT是非常有效的,在金融句子边界检测任务上获得了显著的模型性能提升.

Table 2 Experimental Results on Test Set for the FinSBD Dataset

表 2 金融文本边界检测任务数据集 FinSBD 的实验结果

模型	ES	BS	MEAN
Rule-based ^[25]	0.80	0.86	0.830
BiLSTM-CRF ^[11]	0.83	0.88	0.855
Deep-Att ^[26]	0.83	0.91	0.875
BERT-S ^[27-28]	0.88	0.89	0.885
F-BERT _{BASE}	0.89	0.90	0.897
F-BERT _{LARGE}	0.93	0.95	0.938

2) 金融情绪分析任务

从表3~4我们可以看到F-BERT模型和其他有竞争力的方法在Financial PhraseBank和金融情感分析数据集(FiQA Task1)上的性能.

如表3~4所示,F-BERT模型(包括F-BERT_{LARGE}和F-BERT_{BASE})的模型准确率优于基线模型分别在PhraseBank情感分析数据集和FiQA Task1情感分析数据集上取得的成果.其中,在Financial PhraseBank^[29]上,最优模型F-BERT_{LARGE}获得了0.90的准确率和0.89的F1值.

情感分析数据集FiQA Task1主要包括2种类型的数据:金融新闻头条和金融微博.FiQA Task1具体有2个评估指标^[27]:均方误差(MSE)和R方(R²).在表4中,MSE(H)和MSE(P)分别表示金融微博的均方误差和金融新闻头条的均方误差,同样R²(H)和R²(P)分别表示为金融微博和金融新闻头条的R方值.通过表4可以清晰看到,在FiQA Task1上,最优模型F-BERT_{LARGE}获得了MSE(H)=

0.30 和 $R^2(H)=064$ 的值,以及 $MSE(P)=0.34$ 和 $R^2(P)=0.27$ 的值.

表 3~4 的实验结果表明在 Financial PhraseBank 和 FiQA Task1 上,F-BERT 的表现均明显优于提交时的所有其他方法,这证明了方法的有效性.考虑到使用众多语言功能的最新模型的复杂性,目前的实验结果令人鼓舞,这些结果突出了特定于金融领域的语料库预训练设计的重要性.

Table 3 Experimental Results on Test Set for the PhraseBank Financial Sentiment Analysis Dataset

表 3 金融情绪分析数据集 PhraseBank 的实验结果

模型	准确率	F1 值
CUKG ^[11]	0.13	0.46
ULMFit ^[11]	0.83	0.79
FB-SA ^[27-28]	0.86	0.84
F-BERT _{BASE}	0.88	0.86
F-BERT _{LARGE}	0.90	0.89

Table 4 Experimental Results on Test Set for the FiQA Task1 Financial Sentiment Analysis Dataset

表 4 金融情绪分析数据集 FiQA Task1 的实验结果

模型	$MSE(H)$	$R^2(H)$	$MSE(P)$	$R^2(P)$
IIT-Dehi ^[11]	0.20	0.18	0.10	0.08
Inf-UFG ^[11]	0.21	0.17	0.10	0.16
SC-V ^[27]	0.08	0.40		
RCNN ^[23]	0.09	0.41		
FB-SA ^[27-28]	0.07	0.55		
F-BERT _{BASE}	0.22	0.57	0.25	0.19
F-BERT _{LARGE}	0.30	0.64	0.34	0.27

3) 金融智能问答任务

表 5 展示了金融文本智能问答挑战数据赛 (FiQA Task2)上方法和其他基准模型的性能对比.为了更好地进行金融智能问答模型的性能比较,FiQA Task2 数据集主要定义了 2 个评估指标^[27]:归一化贴现累积收益($nDCG$)和平均倒数排名(MRR).从表 5 可以看到,我们的最佳模型 F-BERT_{LARGE}在测试集上可达到 $nDCG=0.60$ 和 $MRR=0.52$,同时 F-BERT_{BASE}达到了 $nDCG=0.51$ 和 $MRR=0.41$,其准确率也高于其他的基准模型.由于金融智能问答的数据一般都很小(只有几十或者上百个训练样本),由此可见,模型预训练模型在金融智能问答任务中可以起到非常重要的作用.

Table 5 Experimental Results on Test Set for the FiQA Task2 Question Answering Dataset

表 5 金融智能问答数据集 FiQA Task2 的实验结果

模型	$nDCG$	MRR
CUKG-TongJi ^[11]	0.17	0.10
eLabour ^[28]	0.31	0.19
F-BERT _{BASE}	0.51	0.41
F-BERT _{LARGE}	0.60	0.52

3.5 实验分析与讨论

为了详细分析架构内各组件对最终性能的影响,在本节中对架构进行了详细的模型分析,包括预训练对模型性能的影响、少量预训练数据进行预训练分析和预训练语料库大小的讨论.

1) 预训练对模型性能的影响

如表 6 所示,进一步测量了预训练对模型性能的影响.详细比较了 4 个模型:没有进一步的预训练,由 Vanilla BERT 和 Vanilla F-BERT 表示;对训练集进行进一步的预训练,使用 BERT-task 和 F-BERT-task 表示.在测试数据集上使用准确率、精确率和召回率的得分对相应模型进行评估.从表 6 可以明显看出,F-BERT 模型能带来可观的性能提升.尽管 BERT 任务在金融领域训练集上进行了进一步的预训练,但 Vanilla F-BERT 的性能优于 2 个基于 Vanilla BERT 的模型 Vanilla BERT 和 BERT-task,在准确率上,比它们分别提高 0.06 和 0.03.这表明 F-BERT 在预训练期间有效地利用了来自大量未标记金融文本的领域特定知识.

Table 6 Experimental Results on Test Set for the Financial Classification Dataset

表 6 金融文本分类数据集的消融实验

模型	准确率	精确率	召回率
Vanilla BERT	0.77	0.31	0.29
BERT-Task	0.80	0.47	0.41
Vanilla F-BERT	0.83	0.51	0.47
F-BERT-Task	0.85	0.59	0.61

2) 少量预训练数据预训练分析

预训练模型通常需要庞大的训练语料库来进行训练,但是,在金融领域的许多应用程序中,很难拥有大型的有标注训练语料库.因此,为进一步验证 F-BERT 的优势,本文进行了一个模型实验.实验使用小型语料库来分别对 BERT 和 F-BERT 进行预训练.具体地说,通过在整个金融训练数据集中随机选择

1/8 大小的文本数据作为训练语料库.然后,基于自己模拟构造的这个小学语料库,对所有模型进行预训练,并和 1) 中实验的相同任务上进行测试,实验结果详见表 7.

Table 7 The Performance of BERT and F-BERT on Three Financial Tasks When They Are Trained on a Small Corpus
表 7 小数据集预训练数据消融实验

模型	金融文本边界检测	金融情绪分析		金融智能问答	
	MEAN	准确率	F1 值	nDCG	MRR
BERT	0.83	0.82	0.81	0.37	0.27
F-BERT	0.85	0.83	0.82	0.45	0.39

从表 7 中的实验数据我们可以看到,与之前的实验相比,该实验呈现了相同的趋势,F-BERT 模型在所有任务中始终胜过 BERT.该实验结果进一步证实,当模型在不同大小的语料库上训练时,F-BERT 可提供稳定的增强.如表 7 所示,这些实验数

据也表明,F-BERT 模型可以在具体金融文本处理任务上提供更多帮助,例如金融机器阅读理解任务、金融情感分析任务和金融句子边界检测任务.总体而言,该实验模拟了在数据有限的情况下对金融文本进行预训练的情况,这是特定领域经常遇到的问题,因此证明了 F-BERT 有潜力处理特定领域的小训练数据集问题.

3) 预训练语料库的大小及训练时长变化分析
为了进一步评估针对不同大小的数据集 F-BERT 模型的表现,我们针对预训练语料库的规模大小进行了消融实验.通过在 3 个具体下游金融任务上的性能改进变化,进一步验证了在预训练阶段数据大小和多样性的重要性.表 8 详细地展示了不同大小的数据集上 F-BERT 模型的性能变化.通过该实验我们还观察到,随着训练时间的提升 F-BERT 并没有过分拟合数据,模型同样会从进一步训练中获益.

Table 8 The Performance of F-BERT on Different Traing Size and Steps of the Corpus
表 8 预训练语料库的大小及训练时长变化的分析实验

模型	训练集/GB	训练步数	金融文本边界检测	金融情绪分析		金融智能问答	
			MEAN	准确率	F1 值	nDCG	MRR
① F-BERT _{LARGE}	36	200 k	0.88	0.85	0.83	0.42	0.33
② 再增加数据集(Reddit Finance QA+ Financial Web)	40	200 k	0.89	0.86	0.83	0.49	0.38
③ 进一步训练	40	400 k	0.90	0.88	0.85	0.51	0.41
④ 再进一步训练	40	1 200 k	0.94	0.90	0.89	0.60	0.52

注:①选定训练数据,可以观察到 F-BERT 优于原始结果;②增加数据集;③增加训练步数;④进一步增加训练步数.

3.6 真实业务数据的实验

我们利用真实线上业务数据对所提出的 F-BERT 模型有效性进行了进一步验证.具体来说,我们从金融科技实际业务中抽象并人工构建了 2 个典型的金融领域数据集:包括金融命名实体识别任务数据集和金融短文本分类任务数据集,然后分别针对金融命名实体识别和金融短文本分类这 2 个实际金融任务进行了实验.

1) 金融命名实体识别
金融命名实体识别,是一种非常常见的金融业务,此任务主要是对给定的金融文本,利用训练模型可以自动完成有特定意义的金融实体的识别,包括识别出金融机构名、人名(例如高管、人名),以及其他专有实体名词等.该任务也是关系抽取和实体提取的基础任务,也多用于金融大数据和金融知识图谱相关任务中,是金融数据分析的一项非常基础的任务.在这里,我们对实际业务数据进行了简化和脱

敏处理,从真实数据中人工整理出一个金融命名实体识别数据集.该数据集共包含 53 000 条训练样本,测试数据集 45 000 条,训练数据样本 8 000 条.

2) 金融短文本分类
金融短文本分类是一种非常常见的金融业务,此任务主要是帮助判断该金融短文本属于哪一类别,从而完成金融文本内容对自动类别分类,最终对于不同的金融业务场景可以有不同的业务处理方案.在这里,我们对针对实际金融短文本分类进行了精简处理,基于真实的业务数据,我们构造了一个包含 6 600 条样本、6 个类别的金融短文本分类数据集.该数据集的训练集和测试集分别有 2 000 条和 2 850 条.

3) 实验结果
基于构建的金融命名实体识别数据集、金融短文本分类数据集,我们将提出的 F-BERT 模型分别与 BERT 模型和 RoBERTa-wwm-ext^[6] 模型、BERT-

wwm^[26]模型进行金融文本挖掘任务的实验比较.表 9 和表 10 给出了最终的实验对比结果.实验过程中,我们针对这 3 个基准模型(BERT, RoBERTa-wwm-ext, BERT-wwm)和 F-BERT 均设置成相同的学习率等超参数,最终在 F1 值上平均可以提升 3~4 个百分点的准确率.通过对真实业务数据的对比实验,包括金融命名实体识别任务、金融类短文本类型分类任务的基线测试,我们可以清楚看到, F-BERT 模型在相比于 BERT 等其他基线模型在金融文本发掘任务上有比较明显的提升.

Table 9 Experimental Results on Named Entity Recognition Task

表 9 金融命名实体识别实验结果

模型	F1 值
Vanilla BERT	0.89
RoBERTa-wwm-ext	0.92
BERT-wwm	0.91
F-BERT	0.94

Table 10 Experimental Results on SMS Text Classification Task

表 10 金融短文本分类实验结果

模型	F1 值
Vanilla BERT	0.85
RoBERTa-wwm-ext	0.88
BERT-wwm	0.87
F-BERT	0.91

4 总 结

在本文中,我们利用深度学习领域的自监督学习和多任务学习技术,创新地提出了一种基于 BERT 模型架构的大规模语料上训练的开源金融预训练模型 F-BERT.虽然 BERT 是为理解通用语言而构建的,但 F-BERT 通过多任务自监督学习进行同时预训练,有效地捕获了大规模预训练语料中的金融领域知识和语义信息.我们通过最小限度地修改特定于具体金融任务的模型结构, F-BERT 在多个金融文本数据挖掘任务(包括金融关系提取任务、金融情绪分类任务、金融智能问答任务)上的性能均优于当前的最新模型.同时,我们还从金融科技线上业务中构建了 2 类有代表性的金融大数据挖掘任务,即:金融命名实体识别、金融短文本分类,在这 2

个实际金融科技任务进行了实验, F-BERT 模型同样取得了最佳准确性.

将来的工作中,我们将进一步扩展 F-BERT 模型以进行消费者情绪分析和个性化理财智能投顾问问题的研究.

参 考 文 献

[1] Tomas M, Chen Kai, Greg C, et al. Efficient estimation of word representations in vector space [J]. arXiv preprint, arXiv:1301.3781, 2013

[2] Jeffrey P, Richard S, Christopher D M. GloVe: Global vectors for word representation [C] //Proc of the 2014 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1532-1543

[3] Jacob D, Chang Mingwei, Kenton L, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C] //Proc of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 4171-4186

[4] Matthew E P, Mark N, Mohit I, et al. Deep contextualized word representations [C] // Proc of the 2018 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2018: 2227-2237

[5] Lan Zhenzhong, Chen Mingda, Sebastian G, et al. ALBERT: A lite BERT for self-supervised learning of language representations [C] //Proc of the 8th Int Conf on Learning Representations. La Jolla, CA: ICLR, 2020: 1-17

[6] Deng Jia, Dong Wei, Richard S, et al. ImageNet: A large-scale hierarchical image database [C] //Proc of the 2009 IEEE Computer Society Conf on Computer Vision and Pattern. Piscataway, NJ: IEEE, 2009: 248-255

[7] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet classification with deep convolutional neural networks [C] // Proc of the 26th Annual Conf on Neural Information Processing Systems. San Diego, CA: NIPS, 2012: 1106-1114

[8] Huang HsinYuan, Zhu Chenguang, Shen Yelong, et al. FusionNet: Fusing via fully-aware attention with application to machine comprehension [C] //Proc of the 6th Int Conf on Learning Representations. La Jolla, CA: ICLR, 2018: 1-16

[9] Yang Zhilin, Dai Zihang, Yang Yiming, et al. Xlnet: Generalized autoregressive pre-training for language understanding [J]. arXiv preprint, arXiv:1906.08237, 2019

[10] Liu Zhuang, Huang Kaiyu, Huang Degen, et al. Dual head-wise coattention network for machine comprehension with multiple-choice questions [C] //Proc of the 29th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2020: 1015-1024

- [11] Macedo M, Siegfried H, Andre F, et al. Financial aspect and sentiment predictions with deep neural networks: An ensemble approach [C] //Proc of the 2018 World Wide Web Conf. New York: ACM, 2018: 1973-1977
- [12] Pranav R, Robin J, Percy L. Know what you don't know: Unanswerable questions for SQuAD [C] //Proc of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 784-789
- [13] Qiao Liang, Chen Ying, Cheng Zhazhan, et al. MANGO: A mask attention guided one-stage scene text spotter [C] //Proc of the 35th AAAI Conf on Artificial Intelligence and 33rd AAAI Conf on Innovative Applications of Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 2467-2476
- [14] Jia Qi, Huang Hongru, Kenny Q. DDRel: A new dataset for interpersonal relation classification in dyadic dialogues [C] //Proc of the 35th AAAI Conf on Artificial Intelligence and 33rd AAAI Conf on Innovative Applications of Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 13125-13133
- [15] Reddy S, Chen D, Manning C D. CoQA: A conversational question answering challenge [J]. Transactions of the Association for Computational Linguistics, 2019, 7(1): 249-266
- [16] Ashish V, Noam S, Niki P, et al. Attention is all you need [C] //Proc of the 30th Annual Conf on Neural Information Processing Systems. San Diego, CA: NIPS, 2017: 5998-6008
- [17] Mandar J, Chen Danqi, Liu Yinhan, et al. SpanBERT: Improving pre-training by representing and predicting spans [J]. Transactions of the Association for Computational Linguistics, 2020, 8: 64-77
- [18] Colin R, Noam S, Adam R, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [J]. Journal of Machine Learning Research, 2021, 21(140): 1-67
- [19] Alex S, Mike D B. Meet Horovod: Uber's open source distributed deep learning framework for TensorFlow [R/OL]. Uber, 2017 [2021-03-31]. <https://eng.uber.com/horovod/>
- [20] Du Jinhua, Huang Yan, Moilanen K. Sentence boundary detection through sequence labeling and BERT fine-tuning [C] //Proc of the 28th Int Joint Conf on Artificial Intelligence. California: IJCAI, 2019: 81-87
- [21] Tian Ke, Peng Zijun. Sentence boundary detection in noisy texts from financial documents using deep attention model [C] //Proc of the 28th Int Joint Conf on Artificial Intelligence. California: IJCAI, 2019: 88-92
- [22] Mehwish F, Mark C M, Christoph M. Machine learning vs. rule-based sentence boundary detection [C] //Proc of the 28th Int Joint Conf on Artificial Intelligence. California: IJCAI, 2019: 115-121
- [23] Piao Guangyuan, John G B. Financial aspect and sentiment predictions with deep neural networks: An ensemble approach [C] //Proc of the 2018 World Wide Web Conf. New York: ACM, 2018: 1973-1977
- [24] Steve Y, Jason R, Jacques M. Financial aspect-based sentiment analysis using deep representations [J]. arXiv preprint, arXiv:1808.07931, 2018
- [25] Oren M, Jacob G, Ido D. Context2vec: Learning generic context embedding with bidirectional LSTM [C] //Proc of the 20th SIGNLL Conf on Computational Natural Language Learning. Stroudsburg, PA: ACL, 2016: 51-61
- [26] Cui Yiming, Che Wanxiang, Liu Ting, et al. Pre-training with whole word masking for Chinese BERT [J]. arXiv preprint, arXiv:1906.08101, 2019
- [27] Liu Zhuang, Huang Degen, Huang Kaiyu. A pre-trained financial language representation model for financial mining [C] //Proc of the 29th Int Joint Conf on Artificial Intelligence. California: IJCAI, 2020: 4513-4519
- [28] Dogu R. FinBERT: Financial sentiment analysis with pre-trained language models [J]. arXiv preprint, arXiv:1908.10063, 2019
- [29] Pekka M, Ankur S, Pekka J K, et al. Good debt or bad debt: Detecting semantic orientations in economic texts [J]. Journal of the Association for Information Science and Technology, 2014, 65(4): 782-796



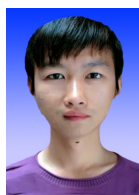
Liu Zhuang, born in 1982, PhD. His main research interests include natural language processing (e.g., dialogue generation, question answering, and sentiment analysis).
刘 壮, 1982 年生. 博士. 主要研究方向为自然语言处理(包括对话生成、自动问答、情感分析).



Liu Chang, born in 1982, Master. Her main research interests include information retrieval and data analytics.
刘 畅, 1982 年生. 硕士. 主要研究方向为信息检索和数据分析.



Wayne Lin, born in 1988, PhD candidate. His main research interests include natural language processing, multimodal learning, and recommendation systems.
Wayne Lin, 1988 年生. 博士研究生. 主要研究方向为自然语言处理、多模态学习、推荐系统.



Zhao Jun, born in 1986, PhD candidate. His main research interests include natural language processing and question answering.
赵 军, 1986 年生. 博士研究生. 主要研究方向为自然语言处理与自动问答.