



ALICE:一种面向中文科技文本分析的预训练语言表征模型

王英杰, 谢 彬, 李宁波

(中国电子科技集团公司第三十二研究所, 上海 201808)

摘 要:深度学习模型应用于自然语言处理任务时依赖大型、高质量的人工标注数据集。为降低深度学习模型对大型数据集的依赖,提出一种基于 BERT 的中文科技自然语言处理预训练模型 ALICE。通过对遮罩语言模型进行改进并将其与命名实体级遮罩相结合,改善基础模型在下游任务中的表现,使其学习到的语言表征更贴合中文的语言特性。实验结果表明,与 BERT 模型相比,ALICE 模型对于中文科技文本的分类准确率和命名实体识别的 F1 值分别提高 1.2% 和 0.8%。

关键词:预训练模型;迁移学习;BERT 模型;文本分类;命名实体识别;自然语言推断

开放科学(资源服务)标志码(OSID):



中文引用格式:王英杰,谢彬,李宁波. ALICE:一种面向中文科技文本分析的预训练语言表征模型[J]. 计算机工程,2020,46(2):48-52,58.

英文引用格式:WANG Yingjie, XIE Bin, LI Ningbo. ALICE: a pre-trained language representation model for Chinese technological text analysis[J]. Computer Engineering, 2020, 46(2): 48-52, 58.

ALICE: A Pre-trained Language Representation Model for Chinese Technological Text Analysis

WANG Yingjie, XIE Bin, LI Ningbo

(The 32nd Research Institute of China Electronics Technology Group Corporation, Shanghai 201808, China)

[Abstract] The deep model of natural language processing rely on huge, high-quality and human-annotated dataset. In order to alleviate such dependency, this paper proposes a BERT-based natural language processing pre-trained model for Chinese technological text named ALICE. Improve Masked Language Model (MLM) and combine it with entity-level mask to boost the base model's performance on downstream tasks, and let the learned representations fit Chinese trait much better. Experimental results show that, compared with the BERT model, ALICE model improves the classification accuracy of Chinese technological texts and the F1 value of named entity recognition by 1.2% and 0.8%, respectively.

[Key words] pre-trained model; transfer learning; BERT model; text classification; named entity recognition; natural language inference

DOI:10.19678/j.issn.1000-3428.0055246

0 概述

近年来,网络已经成为科研人员获得信息的重要渠道。由于网络中科技文献的数量飞速增长,因此亟需一种能自动分析大量科技文献并抽取有价值信息的自然语言处理工具。随着深度学习的快速发展,其在自然语言处理领域得到广泛应用^[1],该领域的相关研究也取得了重大进展。然而,深度学习本身的局限性使得模型在训练时需要大量人工标注的数据集^[2]。对一般领域而言,通过一些众包平台可

以得到大量带标签的数据集,且获取方式比较简单,但是对于科技、医疗、军事等特殊领域,由于数据可能涉及不同层面的专业知识,标注数据需要专家的参与,因此数据集的获取难度较大且成本较高。

以 BERT^[3]为代表的预训练方法降低了模型对大型数据集的依赖性,其在多种自然语言处理任务中的表现均超越了 ELMo^[4]和 GPT^[5]方法。由于这些模型都是针对通用语料进行训练,因此在一些通用领域的自然语言处理任务中可以获得较好的效果。但是科技、医疗等专业领域存在大量专有名词,

基金项目:国家部委基金。

作者简介:王英杰(1994—),男,硕士研究生,主研方向为自然语言处理;谢彬,研究员;李宁波,高级工程师。

收稿日期:2019-06-18 **修回日期:**2019-08-06 **E-mail:**wangyj0233@163.com

这些专有名词的词义与通用领域大不相同,模型对某些词的理解可能会出现偏差,进而影响模型决策的准确性^[6-7]。

本文提出一种面向中文科技文本分析的预训练模型 ALICE。在 BERT 模型的基础上对大量科技语料进行预训练,利用迁移学习得到的语言表征改善模型在小数据集上的自然语言处理性能。

1 相关工作

1.1 预训练模型

针对语言表征预训练的研究工作已经进行了数十年,从传统密集分布的非神经词嵌入模型^[8]到基于神经网络的 Word2Vec^[9]和 GloVe^[10],这些工作为自然语言理解任务提供了一种初始化权值的方法,对模型的表现有明显的提升作用^[11]。Word2Vec 通过对前后 n 个词的依赖关系建模,学习词毗邻出现的概率,GloVe 使用词共现矩阵学习更广泛的共现概率。CoVe^[12]利用神经翻译的编码器在词嵌入中添加含有上下文背景的特征向量,使模型学习上下文背景化的语义。

ELMo^[4]未采用这种传统的学习静态词向量的方法,其在大量语料中训练一个双向神经语言模型(bidirectional Neural Language Model, biNLM),通过该模型从语料集中抽取词汇在不同语境下的含义。OpenAI GPT^[5]使用多层 Transformer^[13]解码器加强模型对上下文背景知识的感知能力。BERT 使用多层 Transformer 编码器学习词汇前后的语义依赖关系,并通过遮罩语言模型(Mask Language Model, MLM)解决 BERT 的输入在多层 Transformer 结构中的“镜像问题”。文献[14]提出知识融合与对话语言模型的概念,针对中文通用领域的自然语言处理任务对 BERT 进行优化,得到 ERNIE 模型。

1.2 迁移学习

在自然语言处理领域,数据量最多的是无标签数据,在大量无标签的数据集上训练模型可以提高有监督任务的表现^[15]。Skip-Thought^[16]使用迁移学习方法通过上下文语句的向量预测输出内容,CoVe^[12]从机器翻译任务中抽取知识并将其成功迁移至其他自然语言理解任务中。

深度学习模型需要数量庞大的人工标注数据^[17],很多领域由于缺乏人工标注的数据资源,模型的表现受到限制。对于小数据集的自然语言处理任务,文献[18]提出一种高效的迁移学习方法,以缓解数据不足的问题。对于一些特殊的领域,如生物医疗,研究者直接从大量的生物医疗语料中抽取信息,并将其用于其他相关的自然语言处理任务中,以提高模型的表现^[19-20]。即使在拥有大量有标签数据的情况下,也可以利用从大量语料中迁移学习得到的知识训练词表征,提升模型的表现^[9]。

1.3 BERT 预训练方法

BERT 通过对语言模型进行无监督预训练,抽取出大量的语义知识,将这些知识迁移到其他自然语言处理任务中,不仅可以提高模型的表现,而且对于小数据集的任务有较好的效果。

BERT 采用双向 Transformer 结构,通过放缩点积注意力与多头注意力直接获取语言单位的双向语义关系,其信息获取的准确性优于传统的循环神经网络(Recurrent Neural Network, RNN)模型和正反向 RNN 网络直接拼接的双向 RNN 模型。此外, BERT 将输入词的 15% 按一定规则进行遮罩(Mask),通过训练语言模型进行预测,从而抵消“镜像问题”的影响。然而,语言模型无法学习到句子间的关系,为了解决这一问题, BERT 引入下句话预测(Next Sentence Predicts, NSP)机制。在预训练后, BERT 无需引入复杂的模型结构,只需对不同的自然语言处理任务进行微调并学习少量新参数,即可通过归纳式迁移学习将模型用于源域与目标域均不相同的新任务中。需要说明的是, BERT 模型对 100 多种语言进行了预训练,而本文主要针对中文进行自然语言处理,因此,若没有特殊说明,下文出现的 BERT 均指针对中文语料训练的 BERT-Chinese。

2 ALICE 文本分析模型

本文 ALICE 文本分析模型主要从词语遮罩语言模型、命名实体级遮罩、Transformer 编码器、预处理方法和预训练语料 5 个方面进行介绍。

2.1 词语遮罩语言模型

英文中单词(word)是可以表达具体意思的最小单位,与英文不同,在中文里,由两个或者两个以上的字(character)组成的词语(phrase)才是表达具体意思的最小单位。BERT 对 CJK Unicode 范围内的每个字添加空格,并且使用 WordPiece 对句子进行分字(tokenize),将句子拆成不连续的字后作为 Transformer 的输入。这种方法虽然提高了模型对不同语言的适应能力和处理能力,但是削弱了对于中文这种以词语为最小语义单位的语言的语义表达。针对上述问题,本文对 BERT 的中文输入方式进行改进,将整个词语作为输入,并以词语为最小语言单位,选取 15% 的词进行类似于 BERT 中遮罩、替换和保持不变的操作,如图 1 所示。经过这种词语遮罩预测任务的训练,模型可以学习到更丰富的语义关系。

例句	傍	晚	小	街	路	面	上	沁	出	微	雨	后	的	湿	润
BERT	M	晚	小	街	路	面	上	M	出	微	雨	M	的	湿	润
ERNIE	M	M	小	街	路	面	上	M	M	微	雨	M	的	湿	润
ALICE	M		小街	路面	上			M		微雨		M	的	湿润	

图 1 BERT、ERNIE、ALICE 输入方式的比较

Fig. 1 Comparison of the input modes of BERT, ERNIE and ALICE

由于中文的字和词语数量庞大,为了降低训练时的内存开销,需要对词汇表进行筛选,选取53 494个常用词作为 ALICE 的词汇表,对于一些词汇表中没有出现的词语(OOV 问题),需对其进行分字处理。

2.2 命名实体级遮罩

ERNIE^[14]模型中引入了命名实体级遮罩的概念,对 BERT 模型的遮罩方式进行修改,ERNIE 以命名实体为单位,选择连续的输入字进行遮罩,其遮罩形式如图 1 所示,这种连续的遮罩方式使模型的自然语言处理效果更好。

在 ALICE 模型中,由于词汇表只覆盖一部分常用的中文词语,一些命名实体、新词等并没有出现在词汇表中。因此,本文借鉴 ERNIE 的方法,使用命名实体级别的遮罩方式来弥补分字处理的不足,让模型学习到的语言表征包含更丰富的语义关系。引入命名实体级遮罩后,ALICE 与 ERNIE 输入方式的对比如图 2 所示。

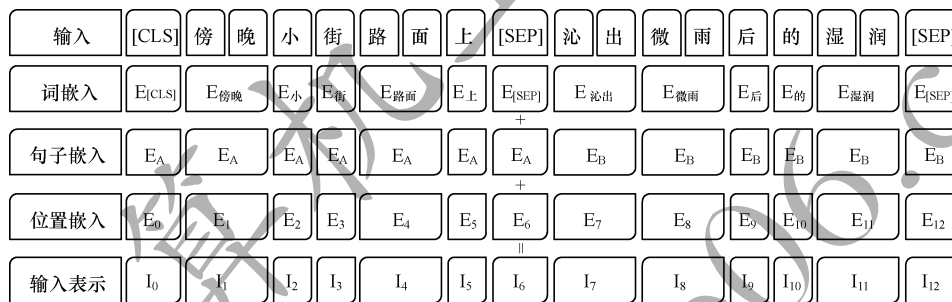


图 3 Transformer 输入方式
Fig.3 Transformer input mode

2.4 预处理方法

对于需要进行预训练的句子,用[CLS]表示句子的开端,以[SEP]充当逗号对句子进行分割。如前文所述,本文未使用 BERT 的 WordPiece 分词方法,而是使用 pkuseg^[21]进行分词操作,然后在每个词的周围添加空格。pkuseg 分词工具与 THULAC、jieba 相比拥有更高的 F1 值,在多种数据集上的平均 F1 值分别比 THULAC、jieba 高出 3.21% 和 9.68%。

2.5 预训练语料

科技预训练语料集包括约 80 万篇文章,其中 60 万篇来自百度百科、微信公众号以及各领域期刊论文,其余的 20 万篇来自各大门户网站的相关文章。计算机、生物科技、军事科技、空间科技的文章分别占 50%、20%、20% 和 10%。语料集少部分期刊论文为全文,大部分仅包含摘要部分。为了防止广告、推荐阅读列表、网站链接等与文章无关的噪音文字干扰模型的训练,需要对语料进行彻底的清洗,以保证其质量。此外,预训练语料还包括 BERT 所使用的维基百科中的 100 多万篇中文文章。表 1 统计了该语料的详细信息,需要注意的是,由于预处理时对新词、连词进行了分字处理,在统计词数时需将这些特殊词丢弃,因此统计数值与实际数值略有偏差。

例句	三	体	是	中	国	作	家	刘	慈	欣	创	作	的	科	幻	小	说
ERNIE	三	体	是	M	M	作	家	M	M	M	创	作	的	M	M	小	说
ALICE	三	体	是	M		作	家	M	M	M	创	作	的	M		小	说

图 2 ALICE 和 ERNIE 的输入方式对比

Fig.2 Comparison of the input modes between ERNIE and ALICE

2.3 Transformer 编码器

ALICE 与 BERT 的结构类似,都使用多层 Transformer 编码器,通过多头自注意力层捕捉句中每个词的上下文信息。对于一个给定字,Transformer 将该字在词汇表中的位置、分句的位置以及与其他字的相对位置相加,以获得该字的输入表示。对 Transformer 的输入方式进行修改后,ALICE 模型的输入如图 3 所示。其中,E 代表嵌入,I 代表输入。

表 1 预训练语料详细信息

Table 1 Details of pre-trained corpus

语料集	文章数/10 ⁶	字数/10 ⁶	词数/10 ⁶
中文维基百科	1.0	417	136
科技语料集	0.8	3 390	1 200
总计	1.8	3 807	1 336

3 实验结果与分析

为了方便比较,ALICE 与 BERT-Chinese 使用同样的模型结构,其中包含 12 个编码器层、768 个隐层单元和 12 个注意力头。本文主要在 3 个自然语言处理任务中,比较本文方法与 BERT、ERNIE 方法的性能。需要说明的是,除文本分类任务外,其他任务都采用 BERT^[3]中的微调方法。

3.1 中文自然语言处理任务

本文的中文自然语言处理任务主要包括文本分类、命名实体识别和自然语言推断。

3.1.1 文本分类

本文参照 DocBERT^[22]的方法对模型进行微调。在最后一层的隐层状态中引入一个响应输入标示为 [CLS] 的全连接层,通过新的 softmax 分类器对模型

进行端到端优化,参数 $W \in \mathbb{R}^{H \times K}$, H 表示隐层状态向量的维度, K 表示类别数。由于本文的文本分类属于多分类任务,因此需要最小化交叉熵。

本文使用的文本分类数据集共包含 8 000 个人工标注的文本,其中生物科技、空间科技、军事科技、高新科技相关文章各 1 500 篇,其他 6 个领域的文章共 2 000 篇。将其中 6 000 篇文章作为训练集,1 000 篇作为开发集,剩下的 1 000 篇作为测试集。

3.1.2 命名实体识别

命名实体识别数据集包括 5 000 条人工标记数据,其采用 BIO 标注策略,实体设计的范围包括生物、空间、军事、高新技术 4 个科技领域的人名、地名、组织名和其他专有名词。在该数据集中,将 4 000 条数据作为训练集,其余数据平均分配给开发集和测试集。

3.1.3 自然语言推断

自然语言推断任务的目的是判断两句话之间是否存在蕴含、中立和矛盾关系。为了证明 ALICE 在通用领域和大数据集上的自然语言推断性能,本文在 XNLI^[23] 的中文数据集上进行实验,详细信息如表 2 所示。

表 2 3 种任务数据集的数据量

Table 2 Data size of the 3 types of task datasets

任务名称	训练集	开发集	测试集
文本分类	6 000	1 000	1 000
命名实体识别	4 000	500	500
自然语言推断	392 703	2 491	5 011

3.2 SOTA 模型

为了证明经过微调的 ALICE 模型性能优于其他具有复杂结构的模型,本文分别选取在上述中文自然语言处理任务中取得 SOTA (State Of The Art) 的模型进行比较。例如,文本分类任务的 SOTA 模型选取 RMDL^[24],命名实体识别任务选取 MSRA 中文数据集的 Lattice LSTM 模型^[25],自然语言推断选取 XNLI 多语言数据集上的 LASER 模型^[26]。由于 RMDL 与 Lattice LSTM 需要词嵌入层来初始化模型权值,因此需在科技语料中训练 200 维的 GloVe 词向量作为 RMDL 的词嵌入矩阵。对于 Lattice LSTM,本文采用文献[25]提供的 50 维 Chinese Giga-Word 和 Word2Vec 词嵌入矩阵。

经测试可知,RMDL 在公开文本分类数据集 THUCNews 中的准确率为 92.3%,Lattice LSTM 在 MSRA 命名实体识别数据集集中的 F1 值为 93.18%。THUCNews 数据集包含 14 个类别,每个类别约 63 086 篇新闻,总共 88 万篇新闻,本文从 10 个类别中随机抽取 10 万篇作为训练集,3 000 篇作为开发集,5 000 篇作为测试集,具体的数据集信息如表 3 所示。

表 3 THUCNews 与 MSRA 数据集的数据量

Table 3 Data size of THUCNews and MSRA datasets

数据集	训练集	开发集	测试集
THUCNews	100 000	3 000	5 000
MSRA	20 865	2 319	4 637

3.3 结果分析

表 4 给出 ALICE 在 3 个中文自然语言处理任务中的性能对比,可以看出,ALICE 在 3 个任务上的性能均优于 BERT 模型和针对不同任务的 SOTA 模型,但是在自然语言推断任务中,ERNIE 的性能更好。

表 4 在 3 个中文自然语言处理任务中的性能对比

Table 4 Comparison of performance in 3 Chinese NLP tasks

模型	文本分类		命名实体		自然语言	
	准确率/%		识别的 F1 值/%		推断准确率/%	
	开发集	测试集	开发集	测试集	开发集	测试集
SOTA	91.7	90.3	92.3	91.8	—	71.4
BERT	92.5	90.8	94.7	93.4	78.4	77.0
ERNIE	93.0	91.5	95.0	93.9	79.8	78.6
ALICE	93.4	92.0	94.9	94.2	79.5	77.9

在公开的通用数据集上,RMDL 与 Lattice LSTM 模型(即文本分类任务和命名实体识别任务的 SOTA 模型)分别取得了 92.3% 与 93.18% 的优秀表现^[25],但是在科技领域的任务中,其性能指标分别降低了 2% 与 1.38%,其主要原因是本文构建的科技自然语言数据集的数据量不到上述通用领域数据集的 1/5,针对这种小数据集任务,传统的深度学习模型无法发挥其优势。由于 BERT 模型需要在大量的通用语料上进行预训练,可以通过迁移学习获得的知识解决小数据集数据量不足的问题,因此其文本分类的准确率和命名实体识别的 F1 值比 SOTA 模型分别提升了 0.5% 和 1.6%。

与 BERT 模型相比,ALICE 模型通过词语遮罩语言模型与命名实体遮罩两处改进以及从科技语料预训练中学习到的专业知识,其科技文本分类准确率和命名实体识别的 F1 分别提高 1.2% 和 0.8%。此外,自然语言推断任务的准确率提高了 0.9%。

ERNIE 针对中文进行优化,通过知识融合与对话语言模型学习更丰富的语义关系。与 ERNIE 模型相比,针对科技领域进行预训练的 ALICE 模型的科技文本分类准确率与命名实体识别任务的 F1 值分别提高 0.5% 和 0.3%。在自然语言推断任务中,ALICE 模型的准确率略低于 ERNIE 模型。

3.4 消融实验

为了证明本文改进的词遮罩语言模型能够理解更丰富的语义关系,本文从语料集中随机选取 10% 的数据进行消融实验,结果如表 5 所示。

表5 不同遮罩策略的效果对比

Table 5 Comparison of effects of different mask strategies

遮罩策略	文本分类准确率/%		自然语言推断准确率/%	
	开发集	测试集	开发集	测试集
MLM	90.8	89.9	93.3	92.1
PMLM	91.0	90.3	93.2	92.6
PMLM + NERM	91.3	90.5	93.9	93.0

由表5可知,在中文的自然语言处理任务中,ALICE的词语遮罩策略PMLM的效果优于BERT的遮罩语言策略MLM,添加ERNIE的命名实体遮罩策略NERM后效果更佳。此外,本文在10倍数据量的预训练语料上进行实验后发现,模型的文本分类准确率和命名实体识别F1值分别提高1.5%和1.2%。

4 结束语

本文提出一种针对科技领域文本处理的预训练模型ALICE。在大量科技文本上进行无监督预训练,通过改进的遮罩语言模型,使其更适用于中文的自然语言处理任务。实验结果表明,该模型在科技领域的文本分类、命名实体识别任务中的性能均优于BERT模型。下一步将通过动态遮罩对ALICE模型进行预训练,提高模型对预训练语料的利用率。

参考文献

- [1] XI Xuefeng, ZHOU Guodong. A survey on deep learning for natural language processing [J]. Acta Automatica Sinica, 2016, 42(10): 1445-1465. (in Chinese)
奚雪峰, 周国栋. 面向自然语言处理的深度学习研究[J]. 自动化学报, 2016, 42(10): 1445-1465.
- [2] LI Liuyu, DU Qiu. The ability and inability of deep learning [J]. International Financing, 2018(8): 25-27. (in Chinese)
李留宇, 杜秋. 深度学习的能与不能[J]. 国际融资, 2018(8): 25-27.
- [3] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1810.04805.pdf>.
- [4] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1802.05365.pdf>.
- [5] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [EB/OL]. [2019-06-01]. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf.
- [6] ZHANG Yukun, LIU Maofu, HU Huijun. Chinese medical entity classification and relationship extraction based on joint neural network model [J]. Computer Engineering and Science, 2019, 41(6): 1110-1118. (in Chinese)
- [7] LI Jianlong, WANG Panqing, HAN Qiyu. Military named entity recognition based on bidirectional LSTM [J]. Computer Engineering and Science, 2019, 41(4): 713-718. (in Chinese)
李健龙, 王盼卿, 韩琪羽. 基于双向LSTM的军事命名实体识别[J]. 计算机工程与科学, 2019, 41(4): 713-718.
- [8] BROWN P F. Class-based n-gram models of natural language [J]. Computational Linguistics, 1992, 18(4): 467-479.
- [9] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2013: 3111-3119.
- [10] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation [EB/OL]. [2019-06-01]. <https://nlp.stanford.edu/pubs/glove.pdf>.
- [11] TURIAN J P, RATINOV L A, BENGIO Y. Word representations: a simple and general method for semi-supervised learning [C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. New York, USA: ACM Press, 2010: 384-394.
- [12] MCCANN B, BRADBURY J, XIONG C, et al. Learned in translation: contextualized word vectors [C]//Proceedings of Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 6294-6305.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 5998-6008.
- [14] SUN Yu, WANG Shuohuan, LI Yukun, et al. ERNIE: enhanced representation through knowledge integration [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1904.09223v1.pdf>.
- [15] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12(8): 2493-2537.
- [16] KIROS R, ZHU Y, SALAKHUTDINOV R R, et al. Skip-thought vectors [C]//Proceedings of Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2015: 3294-3302.
- [17] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [18] HOWARD J, RUDER S. Universal language model fine-tuning for text classification [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1801.06146.pdf>.
- [19] ZHU H, PASCHALIDIS I C, TAHMASEBI A. Clinical concept extraction with contextual word embedding [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1810.10566.pdf>.

(上接第 52 页)

- [20] LEE J, YOON W, KIM S, et al. Biobert: pre-trained biomedical language representation model for biomedical text mining [EB/OL]. [2019-06-01]. <https://arxiv.org/ftp/arxiv/papers/1901/1901.08746.pdf>.
- [21] SUN Xu, WANG Houfeng, LI Wenjie. Fast online training with frequency-adaptive learning rates for Chinese word segmentation and new word detection[C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2012:253-262.
- [22] ADHIKARI A, RAM A, TANG R, et al. DocBERT: BERT for document classification [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1904.08398v1.pdf>.
- [23] CONNEAU A, LAMPLE G, RINOTT R, et al. XNLI: evaluating cross-lingual sentence representations[EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1809.05053.pdf>.
- [24] KOWSARI K, HEIDARYSAFA M, BROWN D E, et al. RMDL: random multimodel deep learning for classification[C]//Proceedings of the 2nd International Conference on Information System and Data Mining. New York, USA: ACM Press, 2018:19-28.
- [25] ZHANG Yue, YANG Jie. Chinese NER using lattice LSTM [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1805.02023.pdf>.
- [26] ARTETXE M, SCHWENK H. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1812.10464.pdf>.