文章编号：1003-0077(2023)09-0092-06

QV-Electra：引入 Query-Value 注意力机制的预训练文本分类模型

邵党国[1,2]，孔宪媛[1]，相 艳[1,2]，安 青[1]，黄 琨[1]，郭军军[1,2]

(1. 昆明理工大学 信息工程与自动化学院，云南 昆明 650500；
2. 昆明理工大学 云南省人工智能重点实验室，云南 昆明 650500)

摘 要：预训练语言模型的作用是在大规模无监督语料上基于特定预训练任务获取语义表征能力，故在下游任务中仅需少量语料微调模型且效果较传统机器学习模型(如 CNN、RNN、LSTM 等)更优。常见的预训练语言模型如 BERT、Electra、GPT 等均是基于传统 Attention 机制搭建。研究表明，引入 Query-Value 计算的 QV-Attention 机制效果较 Attention 机制有所提升。该文模型 QV-Electra 将 QV-Attention 引入预训练模型 Electra，该模型在保留 Electra 预训练模型参数的同时仅通过添加 0.1%参数获得性能提升。实验结果表明，QV-Electra 模型在同等时间的情况下，相较于传统模型以及同等参数规模预训练模型能取得更好的分类效果。

关键词：Electra 预训练模型；Attention 机制；QV-Attention 机制；文本分类

中图分类号：TP391 **文献标识码**：A

QV-Electra: A Pre-trained Text Classification Model with Query-Value Attention Mechanism

SHAO Dangguo[1,2], KONG Xianyuan[1], XIANG Yan[1,2], AN Qing[1], HUANG Kun[1], GUO Junjun[1,2]

(1.Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China;
2.Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)

Abstract: The pretrained language model outperforms the preceding deep learning fameworks such as CNN, RNN, LSTM, etc. Current pre-training language models, such as BERT, Electra, GPT, etc., are all built based on the ATTENTION mechanism. Studies have shown that the QV-attention mechanism with Query-value calculation is more effective. This paper establishes the QV-Electra by introducing QV-Attention into the pre-training model Electra. It can improve the performance by adding only 0.1% parameter while retaining the parameters of the pre-training model Electra. The experimental results show that the QV-Electra model achieves better classification performance than the CNN, RNN, LSTM and the pre-training models (BERT, Electra) of same parameter scale.

Keywords: Electra pre-training model; Attention mechanism; QV-Attention mechanism; text classification

0 引言

文本分类是自然语言处理的基础任务。随着 Web 2.0 时代的到来，互联网文本数量爆发式增长。对文本准确的分类有着巨大的经济价值和社会意义，如对商品评论正确分类可应用于商品推荐和商品改进等方面，而对文本质量的分类则可用于筛选低俗文本，有助于健康互联网环境的构建。

文本分类按照模型发展时间先后可以分为两个阶段，分别为基于词典统计的方法，基于机器学习的方法[1]。①基于词典统计方法的模型初步工作为手

收稿日期：2021-01-06 **定稿日期**：2021-08-11

基金项目：国家自然科学基金(62266025)；云南省基础研究专项面上项目(202001AT070047)

动构建情感词典。情感词典包含情感词、否定词、程度词等。该种方法首先需要对情感词设置合适的权重，并通过对文本包含情感词的权重进行计算来获得文本的最终得分。但该类方法存在灵活性低、分类效果差[2]的问题。②基于机器学习的方法随模型规模增大和特征提取能力增强而逐渐演化出深度学习。传统的机器学习方法包括决策树、k-means、SVM 等。该类方法在训练时学习数据特征，构建分类模型。但该类方法存在特征稀疏、特征提取困难、维度爆炸等缺点。

目前文本分类的主流方法是基于深度学习的方法。而基于 Transformer[3]的预训练模型因其效果好和仅需少量下游任务语料的特点而备受学界和业界关注。

随着深度学习的发展，词嵌入被正式引入自然语言处理任务中。2003 年 Bengio 提出 NMLM[4]模型，随后一系列词嵌入方法（如 Word2Vec[5]、Glove、FastTest 等）被提出。上述方法虽然可为文本提供数值化的表示方法，但无法解决词的唯一数值表示与一词多义之间的矛盾。为解决该问题 Elmo[6]被提出，它采用双向长短时记忆网络（Long Short-Term Memory，LSTM）预训练，使得词向量可结合上下文来获取嵌入表示。2018 年以后，基于注意力机制[3]的 Transformer 的预训练模型层出不穷，并且不断刷新自然语言处理各项任务的基线。

GPT[7]打开了无监督预训练任务与有监督微调任务结合的序幕，使得预训练好的模型能在下游任务上直接微调。BERT[8]首次将双向 Transformer 用于语言模型，使得该模型在语义提取上较 GPT 更加强大。Electra[9]因为其采用的 Generator-discriminator 预训练任务单词利用率为 100%，使其在文本分类任务中相对于 BERT 能以更小的模型达到同样甚至更优的分类效果。

现有 Transformer 预训练模型均是基于注意力机制，而传统的注意力机制仅将 Query-Key 做运算而疏忽了 Query-Value 之间的相互作用。Wu[10]等人提出了一种引入 Query-Value 交互的注意力机制 QV-Attention。且通过实验表明，QV-Attention 在文本分类和命名体识别效果优于传统注意力机制。

本文的主要贡献如下：

（1）将 QV-Attention 引入开源的预训练模型得到本文模型 QV-Electra；

（2）在 3 个中文数据集上实验，表明本文模型仅通过添加 0.1%参数，分类效果优于多项分类任务基线模型。

1　Electra 预训练模型介绍

预训练模型是基于特定的预训练任务，在大规模语料上训练深层 Transformer 网络结构从而得到一组模型参数。将预训练好的模型参数应用于后续具体的自然语言处理任务上微调模型参数，这些特定任务通常被称做“下游任务”。

掩码语言模型（Masked Language Model，MLM）是一种常见的预训练任务。多个预训练模型（如 BERT、ALBert、ROBERT、ENRIE）均采用了该预训练任务。MLM 预训练任务类似于完形填空，即在预训练阶段随机掩盖输入文本的部分词语，在输出层获得被掩盖位置词语的概率分布，进而通过最大化似然概率来优化模型参数。

如 BERT 预训练模型，随机选择输入文本序列中的 15%的词以供后续替换，但并非将这些词全都替换为[MASK]。其中，10%的词被替换为随机词，10%的词不做改变，另外 80%的词被替换为[MASK][8]。上述引入随机词的操作可以理解为通过引入随机噪声来提升模型的鲁棒性。

Electra 预训练模型提出一种新的预训练任务和框架，将生成式的 MLM 预训练任务改成判别式的替换词语检测（Replaced Token Detection，RTD）预训练任务，判断当前 Token 是否被语言模型替换过。

Electra 按照模型可分为生成器和判别器。生成器的作用与 MLM 相同，预测被替换为[MASK]的词生成完整的文本，通常使用小型 MLM 预训练模型如 BERT-small 等，而判别器则是预测经过生成器补全的文本中哪些 Token 是被生成器生成的。这有点类似对抗学习，但与对抗学习不同的是，判别器的输出结果没有参与生成器的梯度更新。（具体训练任务见图 1）。

由于 Electra 预训练 Token 均参与预训练任务，而如 BERT 仅 15%不到的 Token 被替换为[Mask]。如文献[9]所述 Electra-small 的参数量仅为 BERT-base 的 1/10，但分类效果优于 BERT-base，故选择 Electra 预训练模型来完成分类任务是可取的，特别是二分类任务。

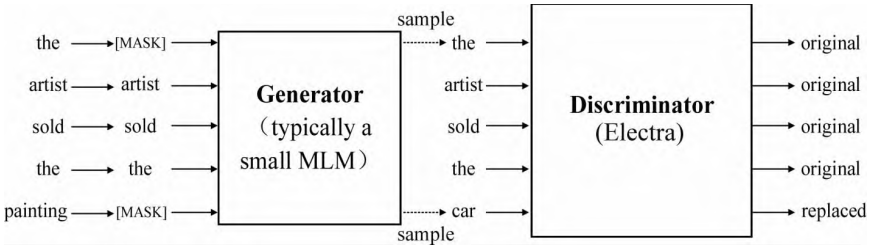

图 1 Electra 预训练任务图

本文模型为研究 QV-Electra 的中文文本分类任务，故使用哈尔滨工业大学于 2020 年 10 月开源的 Electra-180 GB-base[11] 的中文预训练模型，该模型在 180G 中文语料上预训练。Electra-180G-base 模型生成器参数量大约为 1 000 万，而判别器参数量大约为 1 亿。

2 QV-Attention 机制

注意力机制在各种 Transformer 预训练模型中扮演至关重要的作用。Attention 可以表述为一个三元函数，通过计算 Query 和 Key 之间的注意力权重加权到 Value 中将 Key，Query，Value 映射到输出。Attention 机制如式(1)所示。

$$O=f(\boldsymbol{Q},\boldsymbol{K})\boldsymbol{V}=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V} \tag{1}$$

与 Query，Key 交互类似，Query 和 Value 之间也存在内在的相关性，结合 Query 和 Value 交互可以通过根据 Query 和 Value 的特征学习定制来增强输出。但是现有的注意力机制忽略了 Query-Value 的交互作用，可能不是最优的。如文献[10]中，注意力网络用来学习新闻文本的表示，用于新闻推荐，Query 是用户嵌入表示，Value 是该用户点击新闻中单词的表示。在这种情况下，可以使用 Query 和 Value 之间的交互来根据用户特征调整单词的表示，以学习更好的个性化新闻表示。

Wu 等[10] 提出一种新型的注意力机制 QV-Attention。通过引入 Query 和 Value 之间的交互来增强传统 Attention 机制的语义表征能力，并根据原文实验可知 QV-Attention 相对于原始 Attention 在命名实体识别和文本分类任务上性能均有提升。

QV-Attention 相对于传统 Attention 机制在英文 AG 数据集和 Amazon 文本分类任务上准确率提升了 0.3%和 0.8%。为进一步适应分类任务且减少参数量，将原文 QV-Attention 计算公式稍作改动得到本文 QV-Attention 计算公式，如式(2)～式(6)所示。

$$\hat{O}=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)g(\boldsymbol{Q},\boldsymbol{V}) \tag{2}$$

$$Cg(\boldsymbol{Q},\boldsymbol{V})=(1-\beta)\,\hat{\boldsymbol{O}}*\boldsymbol{W}\boldsymbol{V}+\beta\boldsymbol{V} \tag{3}$$

$$\hat{O}=\mathrm{Softmax}\left(\frac{\boldsymbol{V}\boldsymbol{Q}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{Q} \tag{4}$$

$$\beta=\sigma(\boldsymbol{u}^{\mathrm{T}}\mathrm{Concat}(\hat{\boldsymbol{O}}*\boldsymbol{W}\boldsymbol{V};\boldsymbol{V})[:,0,:,:]) \tag{5}$$

$$\boldsymbol{W}\boldsymbol{V}=W^{\mathrm{T}}\boldsymbol{V}+\boldsymbol{b}_v \tag{6}$$

与传统的 Attention 机制相比，QV-Attention 将原始注意力机制公式中的 $\boldsymbol{V}$ 改变为 $g(\boldsymbol{Q},\boldsymbol{V})$，从而引入 Query-Value 交互计算。式(4)模仿注意力机制，$\hat{\boldsymbol{O}}$中的每个向量都是所有 Query 向量的加权和。式(6)是将 Value 向量做特征抽取。式(5)计算聚合参数 β，$*$ 为元素乘积，Concat 为向量拼接，而 $[:,0,:,:]$表示将拼接向量按照索引把第一个 token 向量取出，σ 表示 Sigmoid 激活函数。最终得到 Query-Value 交互计算式(3)，相对于原始 Attention 注意力机制，QV-Attention 注意力机制仅新添加三个学习参数：式(5)中的 u 和式(6)中的 W，$\boldsymbol{b}_v$。

3 QV-Electra 模型

如前所述，Electra 模型可分为生成器和判别器两部分。而下游分类任务仅需使用判别器，故本文仅对哈尔滨工业大学开源官方预训练模型 Electra-108G-base/discriminator[11] 做修改，而忽略 Electra-180G-base 的 Generator 模型。

基于 Transformer 的各种预训练模型可以视为深度 Attention 层堆叠的网络模型。具体可见图 2，图中将 Electra-180G-base/discriminaton 中 Attention 层修改为 QV-Attention 层。

观察图 2 可知，原始预训练模型 Electra-base 网络主要结构为堆叠的 12 层 Transformer 块，而每个 Transformer 的语义提取能力来源于 Attention 层，所以 Attention 机制在预训练模型中起着至关

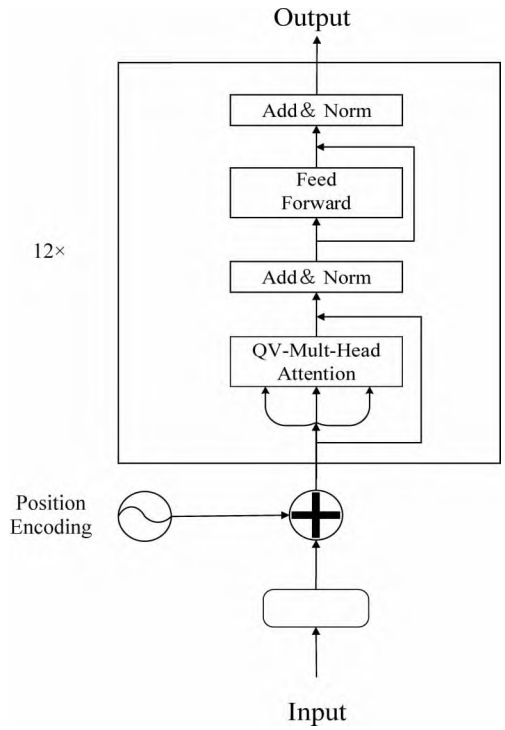

图 2　QV-Electra 文本分类模型图

重要的作用。因此注意力机制的改善将有利于提高预训练模型的性能，而 QV-Attention 机制在语义表征上理论上强于 Attention 机制。

注意观察式(3)和式(5)，β 由式(5)计算得来，调节着 Query-Value 交互计算得到的值与原始 Value 输入值的聚合比率。当 $\beta=0$ 时，QV-Attention 层中 $g(\boldsymbol{Q},\boldsymbol{V})$完全使用 Query-Value 交互计算得到的值。而当 $\beta=1$ 时，$g(\boldsymbol{Q},\boldsymbol{V})$则完全使用原始输入值 Value，那么此时 QV-Attention 退化为传统的 Attention 机制。β 值由网络模型学习得到，由万能逼近定理可知当某分类任务无须使用 Query-Value 交互的注意力机制，则网络将会把 β 调整至 0。而此时 QV-Attention 机制将会退化为传统 Attention 机制，而 QV-Electra 模型也会退化为原始 Electra 模型。从理论上来说，本文模型 QV-Electra 的性能下界为原始的 Electra，这是本文的理论基础。而从下文实验分析可知，QV-Electra 的分类效果优于多项中文文本分类的基线模型，这进一步表明了本文模型的可行性。

上述理论保证在同样大小预训练语料上将 Electra 所有 Attention 层改为 QV-Attention 层所得的预训练模型，理论上其特征提取能力强于原始 Electra 预训练模型。

但重新预训练一个模型成本高昂，QV-Attention 机制相对于传统 Attention 机制仅增加少许参数。故本文提出一种新的基于 QV-Attention 的 QV-Electra 预训练文本分类模型，该模型与 Electra/discriminator模型保持一致的模型结构，只是将每个 Transformer 块中的 Attention 层替换为 QV-Attention 层。并读取 Electra/discriminator 预训练参数，将每层 QV-Attention 新加入参数采用高斯初始化并在下游任务中以较大的学习率来学习。

4　实验结果及分析

4.1　实验环境

本文实验环境及其机器配置如表 1 所示。

表 1　配置表

实验环境	环境配置
操作系统	Windows 10
GPU	2080Super
内存	8 GB
编程语言	Python 3.7
配置库	Transformers 3.5.1
深度学习框架	Pytorch 1.5

4.2　数据集和评价指标

为验证本模型的可行性，本文采用携程酒店评论数据集 ChnSentiCrop，微博评论数据集 NLPCC2014，搜狗新闻数据集 SogouNews。由于数据集没有明确划分训练集和测试集，本文对上述数据集均采用 10 折交叉验证。其中携程酒店评论数据集、新浪微博评论数据集分类任务为二分类：正向和负向。而搜狗新闻数据集任务为四分类任务：文化、财经、军事、运动。具体数据信息如表 2 所示。

表 2　数据集介绍

数据集	大小	正类	负类	平均长度
ChnSentiCorp	6 000	3 000	3000	90
NLPCC2014	12 500	6 250	6250	40

数据集	大小	文化	财经	军事	运动	平均长度
SogouNews	8 000	2 000	2 000	2 000	2 000	534

本文采用四个评价指标：准确率(Accuracy)、精确率(Precision)、召回率(Recall)，具体计算

如式(7)～式(10)所示。

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}\times 100\% \quad (7)$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}}\times 100\% \quad (8)$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}}\times 100\% \quad (9)$$

$$F_1=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}\times 100\% \quad (10)$$

TP 为预测为正向实际上预测正确的样本数目，TN 为预测为负向实际上预测正确的样本数目，FP 为预测为正向实际上预测错误的样本数目，FN 为预测为负向实际上预测错误的样本数目。

4.3 实验超参数设置

本文模型参数可以分为两部分：① 原始 Electra 读取参数；② QV-Attention 新加入参数。本文实验将上述两部分参数设置不同的学习率：①0.000 01；②0.000 05。而其他对比预训练模型学习率均设置为 0.000 01。迭代次数(Epoch)设置为 20，批次数(batchSize)设置为 16。

4.4 实验结果分析

为检验模型的效果，本文将与多个深度学习方法进行对比，基准模型分别包括：卷积神经网络(CNN)[12-13]，循环神经网络(RNN)[14]，双向长短时记忆网络(BiLSTM)[15]，预训练模型 BERT-base[16]，预训练模型 Electra-180G/discriminator[11]。表 3 列出 6 种模型在上述三个数据集上的实验结果。

表 3 模型结果对比 (单位：%)

模型	数据集	Accuracy	Precision	Recall	F_1
CNN	ChnSentiCorp	87.50	87.57	87.50	87.53
	NLPCC2014	77.54	77.26	77.26	77.26
	SogouNews	84.32	84.63	84.13	84.25
RNN	ChnSentiCorp	87.80	87.91	87.80	87.85
	NLPCC2014	76.91	76.87	76.90	76.89
	SogouNews	84.62	84.50	84.14	84.18
BiLSTM	ChnSentiCrop	89.90	89.96	89.90	89.93
	NLPCC2014	78.67	78.63	78.63	78.63
	SogouNews	86.92	87.04	86.36	85.88
BERT-base	ChnSentiCorp	93.79	94.27	94.16	92.47
	NLPCC2014	80.28	80.20	80.29	81.24
	SogouNews	92.70	92.67	92.67	92.67
Electra-discriminator	ChnSentiCorp	95.27	96.20	95.36	94.67
	NLPCC2014	81.54	80.54	82.56	81.55
	SogouNews	93.50	93.72	93.42	93.54
QV-Electra	ChnSentiCorp	95.85	96.34	95.87	95.89
	NLPCC2014	83.53	81.41	84.21	82.32
	SogouNews	94.31	94.42	94.34	94.28

通过对比实验发现，本文模型在上述三个数据集上 4 个评价指标均相对于上述 5 种基线方法有所提升。ChnSentiCorcuracp 数据集上 Accuracy 相对于 CNN，RNN，BiLSTM，BERT-base，Electra 分别提高了 8.35%，8.05%，5.95%，2.06%，0.58%。而在 NLPCC2014 微博数据集上 Accuracy 相对于上述 5 个基线模型分别提高了 5.99%，6.62%，4.86%，3.25%，1.99%。而在文本长度较长的 SogouNews 新闻数据集上 Accuracy 相对于上述 5 个基线模型分别提高了 9.99%，9.69%，7.39%，1.61%，0.81% 。

比较 CNN、RNN、BiLSTM 和 BERT、Electra 效果发现，基于 Transformer 的预训练模型效果优于使用 word_embedding 后接初始化参数模型(RNN、RNN、BiLSTM)的效果，这是因为 Transformer 预训练模型强大的深层 Attention 网络结构和在大规模语料上预训练而获取的语义表征能力。

而对比 QV-Electra 和 BERT，Electra 模型的实验效果显示 QV-Electra 的分类效果较 BERT 和 Electra 有略微提升，这表明引入 QV-Attention 的预训练 Transformer 模型语义表征能力强于基于 Attention 的 Transformer 的预训练模型，即便 QV-Electra 并非重新预训练而来。

为验证本文模型性能提升不是因为在QV-Electra 新加入参数以较大学习率导致，本文在搜狗新闻数据集上对本文模型 QV-Electra 所有参数学习率均设置为 0.000 01，得到四项指标分别为：94.31%，94.42%，94.34%，94.28%。与实验结果表 3 对比，发现在相同的学习率情况下，QV-Electra 分类效果均优于上述多项基线模型，这表明本文模型 QV-Electra 本身的性能优于上述基线模型。而对 QV-Electra 新添加参数适当增大学习率可以进一步提升模型效果，这表明对新加入参数以较大学习率微调是正确的。

上述对比实验表明，本文模型 QV-Electra 在 3 个数据集上(囊括两个短文本数据集和一个长文本数据集)均优于非 Transforemer 结构的神经网络模型和现有同等参数规模的预训练模型，且在长短文本上提升效果幅度类似。这表明本文模型应用于中

文文本分类任务是可行的。

5　结论

本文在官方开源模型 Electra 的基础上，将 Attention 机制修改为 QV-Attention 机制，在 0.1% 参数提升（新加入参数规模约为 10 万，官方模型参数规模为 1 亿）的基础上获得性能提升。

该改进几乎没有额外的算力及时间消耗，仅需与原始训练模型 Electra 一样在下游任务中微调网络参数。这在预训练模型规模爆发式增长（如 GPT 1，2，3 三代参数分别为 1 亿，15 亿，1500 亿）的如今为算力所限的个人或机构提供一种新的思路。

本文模型在短文本评论数据集和长文本新闻数据集上实验，结果表明，本文模型相较于多项分类基线模型效果更优。这说明本文模型在中文分类任务上是有效且有意义的。后续工作中，我们考虑将本文模型应用于其他语言的文本分类任务以及其他自然语言处理任务中去。

参考文献

[1] 蒋盛益，黄卫坚，蔡茂丽，等.面向微博的社会情绪词典构建及情绪分析方法研究[J].中文信息学报，2015，29(06)：166-171.

[2] 钱凯雨，郭立鹏.融入习语信息的网络评论情感分析研究[J].小型微型计算机系统，2017，38(06)：1273-1277

[3] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[C]//Proceedings of the Advances in Neural Information Processing systems. 2017：5998-6008.

[4] BENGIO Y，DUCHARME R，VINCENT P，et al. A neural probabilistic language model[J]. The Journal of Machine Learning Research，2003，3：1137-1155.

[5] MIKOLOV T，CHEN K，CORRADO G，et al. Efficient estimation of word representations in vector space [J]. arXivpre-printarXiv：1307.3781，2013.

[6] PETERS M E，NEUMANN M，IYYER M，et al.Deep contextualized word representations[J].Association for Computational Linguistics，2018.1：2227-2237.

[7] RADFORD A，NARASIMHAN K，SALIMANS T，et al. Improving language understanding by generative pretraining[J].arXiv preprint arXiv：1301.3781，2018.

[8] DEVLIN J，CHANG M W，LEE K，et al.BERT：Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 29th Conference of the North American Chapter of the Association of Computational Linguistics，2018：4171-4186.

[9] CLARK K，LUONG M T，LE Q V，et al. Electra：Pre-training text encoders as discriminators rather than generators[J]. arXiv preprint arXiv：2003.10555，2020.

[10] WU C H，WU F ZH，QI T，et al.Improving attention mechanism with query-value interaction[J].arXiv e-prints：arXiv：2010.03766，2020.

[11] CUI Y，CHE W，LIU T，et al. Revisiting pre-trained models for Chinese natural language processing[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing，2020.

[12] GONG Y，ZHANG Q. Hashtag recommendation using attention-based convolutional neural network [C]//Proceedings of the IJCAI，2016：2782-2788.

[13] KIM S M，HOVY E. Extracting opinions，opinion holders，and topics expressed in online news media text[C]//Proceedings of the Workshop on Sentiment and Subjectivity in Text，2006：1-8.

[14] MIKOLOV T，KARAFIáT M，BURGET L，et al. Recurrent neural network based language model[C]//Proceedings of the 11th Annual Conference of the International Speech Communication Association，2010.

[15] SHARFUDDIN A A，TIHAMI M N，ISLAM M S. A deep recurrent neural network with BiLSTM model for sentiment classification[C]//Proceedings of the International Conference on Bangla Speech and Language Processing. IEEE，2018：1-4.

[16] CUI Y，CHE W，LIU T，et al. Pre-training with whole word masking for Chinese bert[J]. arXiv preprint arXiv：1906.08101，2019.

邵党国（1979—），博士研究生，副教授，主要研究领域为数据挖掘、文本处理。
Email：huntersdg@126.com

相艳（1979—），通信作者，博士研究生，副教授，主要研究领域为文本挖掘、情感分析。
E-mail：50691012@qq.com

孔宪媛（1995—），硕士研究生，主要研究领域为文本生成、文本分类。
Email：kongxianyuan05@163.com