

基于改进 BERT-BiGRU 模型的文本情感分类研究

李芸^{1,2},潘雅丽¹,肖冬¹

(1. 杭州电子科技大学电子信息学院,浙江杭州310018;2. 浙江省装备电子研究重点实验室,浙江杭州310018)

摘要:针对目前网络评论文本情感分类准确性不高的问题,提出一种基于BERT和双向门控循环单元(BiGRU)的改进模型,使用能够表征文本丰富语义特征的BERT模型进行词向量表示,结合能够长期保留文本上下文关联信息的BiGRU神经网络提高模型的分类效果,并在此基础上引入注意力机制,突出文本中更能表达分类结果的情感词权重,提高情感分类的准确率。将上述模型分别在Acclmdb_v1和酒店评论两个公开数据集上进行测试,实验结果表明,该模型在中、英文文本情感分类任务中都获得了良好的性能。

关键词:文本情感分类;BERT;BiGRU;注意力机制

中图分类号: TP391.1

文献标志码: A

DOI: 10.16157/j.issn.0258-7998.223108

中文引用格式:李芸,潘雅丽,肖冬. 基于改进 BERT-BiGRU 模型的文本情感分类研究[J]. 电子技术应用,2023,49(2):9-14.

英文引用格式:Li Yun, Pan Yali, Xiao Dong. Research on text emotion classification based on improved BERT-BiGRU model [J]. Application of Electronic Technique, 2023, 49(2): 9-14.

Research on text emotion classification based on improved BERT-BiGRU model

Li Yun^{1,2}, Pan Yali¹, Xiao Dong¹

(1.School of Electronics Information, Hangzhou Dianzi University, Hangzhou 310018, China;

2.Zhejiang Provincial Key Laboratory of Equipment Electronics, Hangzhou 310018, China)

Abstract: Aiming at the problem that the accuracy of text emotion classification of online comment is not high, an improved model based on BERT and bidirectional gated recurrent unit (BiGRU) is proposed. The word vector representation is carried out by using the BERT model which can represent the rich semantic features of the text. The classification effect of the model is improved by combining the BiGRU neural network which can retain the text context related information for a long time. On this basis, the attention mechanism is introduced, to highlight the weight of emotional words which can better express the classification results in the text, and improve the accuracy of emotional classification. The above model was tested on Acclmdb_v1 data set and hotel reviews data set, which are public data set. The experimental results show that the model achieves good performance in both Chinese and English text emotion classification tasks.

Key words: text emotion classification; BERT; BiGRU; attention mechanism

0 引言

文本情感分类^[1]是自然语言处理(Natural Language Processing, NLP)的一个重要任务。随着互联网的飞速发展,人们不仅能够从网上获取信息,还能通过各种平台随心所欲地发表包含自己主观情绪的观点和看法,例如网络购物、微博评论等。对含有丰富情感信息的文本数据进行分析具有非常重要的现实意义,它被广泛应用于产品推荐、搜索排名等领域。但由于一些平台字符长度有限,而且用户表达较随意,存在用词不规范和词语拼写错误等问题,传统的情感分类方法在许多方面表现效果差强人意,实现高效准确的自动文本情感分类方法

是本文的主要研究内容。

1 相关研究

基于深度学习的文本情感分类方法能够联系上下文信息并捕捉到文本的顺序结构,在文本信息处理、分类任务中获得了较好的效果。Kim等人^[2]最先提出了使用卷积神经网络(Convolutional Neural Networks, CNN)进行文本分类,CNN的卷积层和池化层可以提取文本中的高层次特征,减少训练参数。Johnson等人^[3]在CNN的基础上提出了深层金字塔卷积神经网络(Deep Pyramid Convolutional Neural Networks, DPCNN),DPCNN引入残差结构,增加多尺度信息,通过不断加深

网络有效获取文本的长期关联,却没有增加太多计算成本。

虽然卷积神经网络在情感分类任务的研究中取得了不错的效果,但是它不能很好地处理序列信息,而循环神经网络(Recurrent Neural Networks, RNN)能解决这个问题。Socher 等人^[4]将 RNN 应用于文本情感分类,RNN 充分考虑了句中的结构信息,但在处理长序列数据时表现不佳。Schmidhuber^[5]基于 RNN 提出了长短时记忆(Long Short-Term Memory, LSTM)网络,它能够有效解决 RNN 的梯度消失和梯度爆炸的问题。Graves 等人^[6]首次将双向 LSTM 网络用于文本分类,与单向 LSTM 模型相比,双向 LSTM 可以更好地捕获文本的双向语义。Cho 等人^[7]提出了门控循环单元(Gated Recurrent Unit, GRU),GRU 相较于 LSTM 模型结构更加简单,参数更少,并且能长期保存隐藏层序列信息。Bahdanau 等人^[8]第一次将注意力机制应用到文本分类任务中,它对输入文本分配不同的关注,目的是协助分类模型捕捉重要信息。

在词向量表示方面,传统的文本情感分类方法大多数用词袋(bag-of-words)模型^[9]来表示,然后通过标注情感词典的方法进行分类或者通过机器学习的方法分类,这些方法往往忽略了文本的顺序结构特征和上下文信息,很难学习到准确的语义特征。近年来词嵌入^[10]模型逐渐被大家广泛应用,谷歌公司提出了词向量模型 Word2Vec^[11](Word to Vector),它将稀疏的词向量映射为稠密的词向量,按照表现形式的不同,将 Word2Vec 模型分为了连续词袋模型 CBOW (Continuous Bag-Of-Words Model) 和 Skip-Gram (Continuous Skip-gram Model) 模型。Pennington 等人^[12]提出了 Glove (Global Vectors) 模型来解决 Word2Vec 不能捕捉文本全局信息的问题,但是 Glove 模型没有考虑单词含义在不同语境下的含义不同。Devlin 等人提出了 BERT^[13](Bidirectional Encoder Representations from Transformer) 模型,它使用多层双向 Transformer 编码器,可以充分挖掘并提取上下文的情感信息,BERT 模型提出时刷新了 NLP 任务之前的 11 项最优性能记录。

因此,本文拟采用 BERT 预训练模型,结合双向 GRU 神经网络,对网络文本数据进行情感分类。

2 模型结构

2.1 BERT 模型

BERT 是 2018 年谷歌 AI 研究院提出的一种 NLP 预训练模型,它由多层双向 Transformer 编码器构成,其模型结构如图 1 所示。BERT 模型进行训练有两个步骤:预训练和微调。首先随机初始化模型参数,然后进行网络训练,不断调整模型的参数以达到最好的训练效果。将训练结果最好的模型参数记录下来,用自己的数据对

参数进行微调以适应数据。首先对模型进行预训练产生模型参数,之后用这些参数对下游任务进行处理。在每个输入文本前插入一个[CLS]符号作最终的文本语义表示,文本中插入的[SEP]表示特殊分隔符。

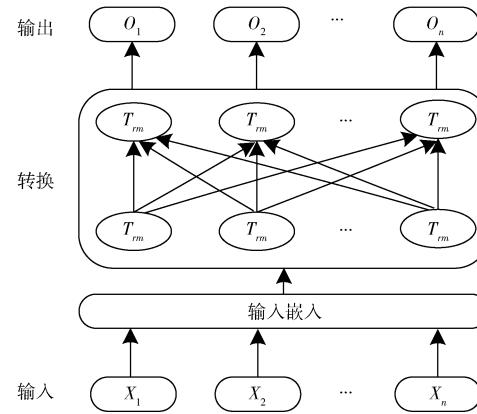


图 1 BERT 模型图

为了增强对文本语义的表达能力,BERT 模型提出了两个预训练任务,一个是遮蔽语言建模(Masked Language Model, MLM),对句中的部分词汇(token)进行随机遮蔽,然后预测被遮蔽的 token;另一个是下一句预测(Next Sentence Prediction, NSP),让模型理解句子与句子之间的关系,从而更好地完成文本情感分类任务。

2.2 BiGRU 模型结构

GRU 是一种由 LSTM 改进的深度网络模型,它将 LSTM 的遗忘门与输出门合并,减少了参数,提高了效率。GRU 模型结构如图 2 所示,参数说明如表 1 所示。GRU 模型只有两个控制门:更新门(update gate)和重置门(reset gate),更新门用于判断前一时刻的隐藏层的状态信息是否对当前层造成影响,重置门可以删除前一时刻中的无效信息,两个门能够长期保存隐藏层的状态信息,具有更好的捕捉长距离依赖关系能力。

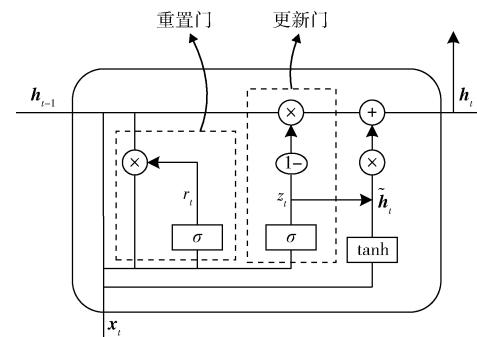


图 2 GRU 结构图

GRU 只能处理单向时间序列,而文本中上下文信息关联较大,因此本文采用双向 GRU(BiGRU)网络,即叠加两个 GRU 网络来保留一个词的前后文联系。

表1 参数说明表

参数	说明
x_t	输入数据
h_t	GRU单元输出
\tilde{h}_t	候选隐状态
r_t	重置门
z_t	更新门

2.3 BERT-BiGRU 模型

本文提出一种以 BERT 作为词向量模型,用双向 GRU 网络进行高效的特征提取的 BERT-BiGRU 文本情感分类模型,结构如图 3 所示。

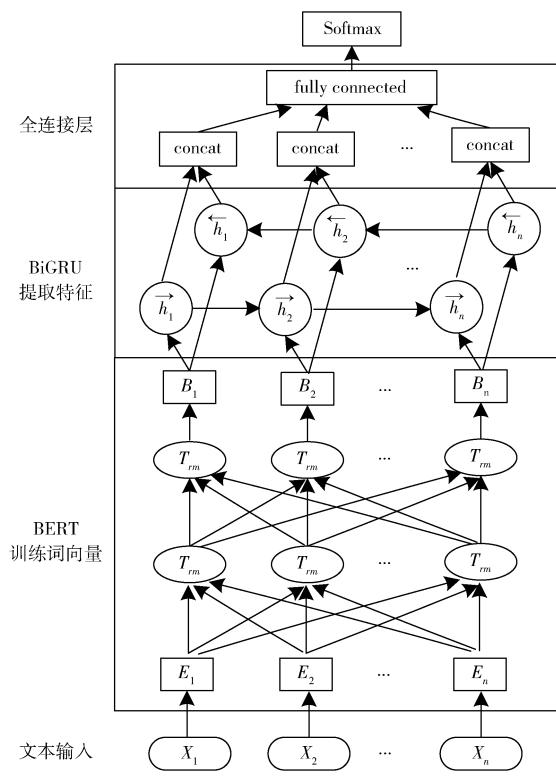


图 3 BBG 文本情感分类算法结构图

将经过 BERT 模型得到的文本词向量数据 X_w 输入到 GRU 网络中,通过 r_t 和 z_t 共同控制从隐藏状态 h_{t-1} 到 h_t 的计算。各门单元计算如下:

$$r_t = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r h_{t-1}) \quad (1)$$

$$z_t = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z h_{t-1}) \quad (2)$$

$$\tilde{h}_t = \tanh(\mathbf{W} \mathbf{x}_t + \mathbf{U}(r_t \cdot h_{t-1})) \quad (3)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (4)$$

其中, \mathbf{W} 和 \mathbf{U} 为权重矩阵, σ 为 sigmoid 激活函数。

根据式(3)、式(4),由重置门、更新门和隐含状态可以计算当前 GRU 输出。通过设置 r_t 的值来对上一时刻隐藏状态信息进行取舍,当 r_t 接近于 0 时,舍弃上一时刻候选隐藏状态,当 r_t 接近于 1 时保留,因此重置门有助于

获取文本序列中的短期关联信息。当 $z_t=1$ 时, $h_t = h_{t-1}$, 即更新门有助于获取长期关联信息。

在循环神经网络的训练中,状态的传递是单向的,即从前往后传输,但文本语义特征不仅取决于之前的状态,还与之后的状态有关,因此需要采用双向 GRU 网络来解决此问题,它由上下两个 GRU 叠加在一起组成,输出由两个 GRU 的状态共同决定,即:

$$h_t^{(i)} = [\overrightarrow{h_t^{(i)}}, \overleftarrow{h_t^{(i)}}] \quad (5)$$

其中, $h_t^{(i)}$ 表示第 i 个文本的 BiGRU 信息, $\overrightarrow{h_t^{(i)}}$ 和 $\overleftarrow{h_t^{(i)}}$ 分别表示第 i 个文本的前向 GRU 信息和后向 GRU 信息。BiGRU 网络不仅可以增加网络的深度,提高模型的训练效率和性能,更有助于扩展前向和后向文本情感信息,从整体上进行文本的综合特征提取。

将通过 BiGRU 网络获得的文本特征向量输入到全连接层获得更加精确的情感特征,同时引入 dropout 机制来防止模型出现过拟合现象以进一步增加模型分类的有效性,最后将提取到的文本特征整合后输入到输出层的 Softmax 函数中获得情感分类结果。

2.4 BERT-BiGRU 模型改进

文本情感分类中加入注意力机制(Attention model)是为了让文本中重点词语对分类结果产生更大的影响。通过计算概率分布,选择对当前目标任务更关键的信息,对深度学习模型进行优化。下面将注意力机制与 BERT-BiGRU 网络相结合,用注意力机制代替上一小节 BERT-BiGRU 模型中的全连接层,通过加权得到最终的文本信息。注意力机制的输入为上一层 BiGRU 网络的输出向量,权重系数计算如下:

$$\text{score}(\mathbf{x}_i, \mathbf{x}_j) = V_a^T \tanh(w_a[\mathbf{x}_i \oplus \mathbf{x}_j]) \quad (6)$$

$$\alpha_{ij} = \frac{e^{\text{score}(\mathbf{x}_i, \mathbf{x}_j)}}{\sum_j e^{\text{score}(\mathbf{x}_i, \mathbf{x}_j)}} \quad (7)$$

其中, V_a^T 是模型训练得到参数向量的转置,将各单词按不同权重加权,最后将提取到的文本特征整合后输入到输出层的 Softmax 函数中获得情感分类结果,如式(8)所示。

$$P_y^{(i)} = j! \mathbf{x}^{(i)}; \theta = \frac{e^{\theta^T x^{(i)}}}{\sum_k e^{\theta^T x^{(i)}}} \quad (8)$$

其中, θ 代表参数向量, k 表示类别数。

改进的 BERT-BiGRU 模型引入注意力机制,通过在关键信息上分配足够关注来突出局部重要信息,从而提高隐藏层特征提取质量,增加情感分类的准确性。

改进算法流程如图 4 所示。

(1) 对输入的待分类文本进行数据清洗,将换行符等特殊字符清洗掉、过滤掉非 ASCII 字符等。

(2) 将处理后的文本数据输入到 BERT 预训练语言

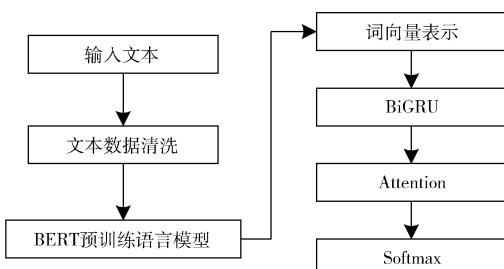


图4 改进算法流程

模型中,生成文本的词向量,将词向量拼接成词向量矩阵作为文本的向量表示。

(3)将词向量表示输入到BiGRU网络中进行特征提取后得到特征向量。

(4)通过注意力机制对文本特征向量分配不同的权值。

(5)将最终得到的文本特征向量输入到Softmax函数中得到文本的最终情感分类结果。

3 实验与分析

3.1 实验配置

本文实验环境配置如表2所示。

表2 实验环境配置

实验环境	配置
操作系统	64位Windows10
处理器	Intel Core i7-8550U
内存/GB	12
编程语言	Python
深度学习框架	TensorFlow

本文选用AclImdb_v1和酒店评论两个公开数据集进行验证。其中,AclImdb_v1是英文数据集,共有42 455条数据;酒店评论是中文数据集,共有10 000条数据。将数据集输入到模型进行训练之前,首先对文本数据进行清洗,将换行符等特殊字符清洗掉、过滤掉非ASCII字符,之后按照8:2的比例随机将数据集划分为训练集和测试集,具体划分情况如表3所示。实验模型参数设计如表4所示。

表3 数据集划分情况

数据集	划分	正向	负向
AclImdb_v1数据集	训练集	17 064	16 900
	测试集	4 266	4 225
酒店评论数据集	训练集	5 600	2 400
	测试集	1 400	600

本文采用以下评价指标:准确率 A_{cc} (accuracy)、精确率 P (precision)、召回率 R (recall)和 F_1 测度(F1-Score)。定义如下:

表4 模型参数设置

参数名称	参数值
词向量维度	500
epochs	7
batch_size	8
学习率	2×10^{-5}
dropout	0.5
损失函数	Cross-Entropy
优化器	SGD

$$A_{cc} = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

$$F_1 = \frac{2PR}{P + R} \quad (12)$$

其中,TP表示模型预测正确的正样本数量,TN表示模型预测正确的负样本数量,FP表示模型预测错误的负样本数量,FN表示模型预测错误的正样本数量, F_1 是精确率和召回率的加权平均。

3.2 实验结果与分析

首先验证BERT词向量模型的效果,在AclImdb_v1数据集上对BERT、Word2Vec和Glove 3种模型进行对比。将情感文本输入到这几个词向量模型中得到含有文本特征的词向量,然后通过最大池化层和输出层得到最终的情感分类结果。实验评价指标选取准确率和 F_1 值,实验结果如表5所示。

表5 不同词向量模型实验结果对比 (%)

词向量模型	准确率	F1-Score
Word2Vec	85.24	85.11
Glove	88.02	88.12
BERT	90.32	90.41

从准确率来看,BERT明显优于Word2Vec和Glove,达到了90.32%,比Glove提高了2.3%,比Word2Vec提高了5.08%; F_1 值方面,BERT也远优于其他两种词向量模型,因此本文采用BERT词向量模型。

接下来验证本文设计模型的有效性,将之与文献[14]中几种常用模型分别在AclImdb_v1和酒店评论数据集上进行对比,实验结果如表6所示。各模型依次为:

(1)BERT模型,直接利用BERT模型生成含有文本语义特征的句向量进行情感分类任务;

(2)BERT-CNN模型,通过BERT模型获得词向量,采用CNN网络进行训练;

(3)BERT-RNN 模型,通过 BERT 模型获得词向量,采用 RNN 网络进行训练;

(4)BERT-BiGRU 模型,在 BERT 模型的基础上采用 BiGRU 网络进行训练;

(5)BERT-BiGRU-attention 模型,改进的 BERT-BiGRU 模型,引入注意力机制。

表 6 实验结果 (%)

数据集	模型	准确率	精确率	召回率	F1-Score
Acldb	BERT	90.32	90.41	90.42	90.41
	BERT-CNN	91.18	91.23	91.08	91.15
	BERT-RNN	91.96	91.85	91.26	91.55
	BERT-BiGRU	93.52	92.57	92.40	92.48
	BERT-BiGRU-attention	94.87	94.53	94.56	94.54
酒店评论	BERT	90.11	90.13	90.21	90.17
	BERT-CNN	90.26	89.01	88.53	88.77
	BERT-RNN	91.06	90.41	90.47	90.44
	BERT-BiGRU	92.42	90.57	90.70	90.63
	BERT-BiGRU-attention	93.02	92.53	92.61	92.57

准确率方面,在 Acldb_v1 数据集上,BERT-BiGRU 达到了 93.52%,相较于 BERT-CNN 和 BERT-RNN 皆有提升,BERT-BiGRU-attention 模型准确率进一步提高了 1.35%;在酒店评论数据集上,BERT-BiGRU 准确率依然达到了 92.42%,相较于单纯使用 BERT 模型提高了 2.31%,BERT-BiGRU-attention 进一步提高到 93.02%。

F_1 值能够比较全面地反映精确率和召回率,在两个数据集上,本文设计的 BERT-BiGRU 和 BERT-BiGRU-attention 模型 F_1 值性能较基本 BERT 模型都有一定程度的提高。在较小的酒店评论数据集上,性能提升不如 Acldb_v1 数据集,是因为小数据情况下,最大池化提取特征具有单一性问题,导致 BERT-CNN 分类效果不尽人意。BERT-RNN 分类效果优于 BERT-CNN,因为 RNN 能够解决文本序列问题。BERT-BiGRU 效果最好,它能保留文本中词与词之间的上下文关联信息,提升训练效果。注意力池化层根据文本词汇的重要程度分配不同的权值,对最终情感极性产生不同程度的影响,能够缓解最大池化信息丢失和提取特征单一的缺陷,进一步提高模型的分类效果。

4 结论

本文基于 BERT 预训练语言模型,结合 BiGRU 神经网络和注意力机制,设计了一种网络文本情感分类模型 BERT-BiGRU-attention。该模型使用 BERT 生成词向量输入 BiGRU 网络,通过注意力机制对 BiGRU 网络提取

到的文本特征向量进行加权,最终得到文本情感分类结果。BiGRU 网络能够更好地捕获文本及其上下文的总体特征,达到高效训练的结果;注意力机制可以提高文本中更能表达情感的单词的权值,进一步提高分类的准确率。实验结果表明,和其他模型相比,BERT-BiGRU-attention 在中英文数据集上都获得了很好的分类效果,明显提升了情感分类性能。

参考文献

- [1] ABE H, TSUMOTO S. Text categorization with considering temporal patterns of term usages[C]//IEEE International Conference on Data Mining Workshops. IEEE Computer Society, 2010:1743–1748.
- [2] KIM Y. Convolutional neural networks for sentence classification [J]. arXiv preprint arXiv: 1408.5882, 2014.
- [3] JOHNSON R, TONG Z. Deep pyramid convolutional neural networks for text categorization[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017: 562–570.
- [4] SOCHER R, LIN C, MANNING C, et al. Parsing natural scenes and natural language with recursive neural networks[C]//Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011: 129–136.
- [5] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735–1780.
- [6] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM networks[C]// IEEE International Joint Conference on Neural Networks. IEEE, 2005.
- [7] CHO K, MERRIENBOER B V, GULCEHRE C, et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation [C]// Conference on Empirical Methods in Natural Language Processing, 2014:1–15.
- [8] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]// International Conference on Learning Representations, 2015:1309–1324.
- [9] WAN C, WANG Y, LIU Y, et al. Composite feature extraction and selection for text classification [J]. IEEE Access, 2019,7: 35208–35219.
- [10] HAGGERTY R A, PURVIS J E. Natural language processing: put your model where your mouth is [J]. Mo-

lecular Systems Biology, 2017, 13(12):958–969.

[11] GUPTA P, PAGLIARDINI M, JAGGI M. Better word embeddings by disentangling contextual n-Gram information [J]. arXiv:1904.05033,2019.

[12] PENNINGTON J, SOCHER R, MANNING C. Glove: global vectors for word representation[C]//Conference on Empirical Methods in Natural Language Processing, 2014:1532–1543.

[13] DEVLIN J, CHANG M W, LEE K, et al. BERT: pretraining of deep bidirectional transformers for language understanding [J]. Computation and Language, 2018, 23(2):

3–19.

[14] XU G X, MENG Y T, QIU X Y, et al. Sentiment analysis of comment texts based on BiLSTM [J]. IEEE Access, 2019, 7: 51522–51532.

(收稿日期:2022-06-21)

作者简介:

李芸(1977-),通信作者,女,博士,高级实验师,主要研究方向:信息处理、无线通信,E-mail:
liyunr@163.com。

潘雅丽(1997-),女,硕士,主要研究方向:信息处理。

肖冬(1999-),男,硕士研究生,主要研究方向:信息处理。



扫码下载电子文档

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所