

Dear Editors and Reviewers:

Thanks very much for taking the time to review this manuscript. We appreciate all your comments and suggestions. We provide a detailed response to all of the previous reviews, as well as an updated manuscript highlighting in **BLUE**.

Thanks again.

1. RESPONSE TO REVIEWER #1

Comment 1: Please thoroughly proofread the paper and make any necessary corrections to grammar errors. Ensure that sentence structures and other language conventions are properly used.

For example, in Abstract, ... the TCAMixer achieved over 85% to 98.7% performance of large pre-trained models, but only "own" -> "owns" 1/3000 to 1/2000 size of those models on most of test data sets.

Response: We deeply regret any deficiencies in the English expression used in our manuscript. As such, we have taken great care to meticulously review and revise the manuscript for any grammatical, syntactical, or spelling errors, in order to ensure that the manuscript is presented in the best possible light. Furthermore, we have proactively sought out the assistance of highly qualified native speakers to help improve the quality of the manuscript, both in terms of its scientific content and its linguistic accuracy. Our commitment to producing a manuscript of the highest quality remains unwavering.

2. RESPONSE TO REVIEWER #3

We are extremely grateful for the time and effort that you have dedicated to evaluating our work and providing valuable feedback. We will take your comments and suggestions into careful consideration as we continue to refine and strengthen our manuscript. Once again, we appreciate your support and wish you all the best in your personal and professional endeavors.

3. RESPONSE TO REVIEWER #4

We are extremely grateful for the time and effort that you have dedicated to evaluating our work and providing valuable feedback. We will take your comments and suggestions into careful consideration as we continue to refine and strengthen our manuscript. Once again, we appreciate your support and wish you all the best in your personal and professional endeavors.

4. RESPONSE TO REVIEWER #5

Comment 1: Its sections should be addressed at the end of the introduction.

Response: Thanks for the advice. We have addressed sections at the end

of the Introduction. Details are as follows:

This paper is organized into six sections. Section 1 provides a brief overview of the research background and related methods. In section 2, the shortcomings of existing research are identified, and the focus of this paper is introduced. Section 3 proposes the Triple Concepts Attention. Section 4 introduces the TCAMixer model based on TCA. Section 5 presents the experimental results and provides analysis. Finally, in section 6, the entire paper is summarized

Comment 2: The specifications of all software and hardware systems used must be reported.

Response: Thanks for the advice. We have supplemented the experimental hardware and software system descriptions prior to Section 5.1. Details are as follows:

The program was written using PyTorch 1.7 framework, and the effectiveness of the method was validated on a computer with an AMD R7 4800H CPU, an RTX3060 GPU, and 16GB of RAM.

During the experiment, the model's learning rate was carefully set to $5e-4$ in order to strike a balance between achieving good performance and avoiding overfitting. Additionally, weight decay was set to $1e-3$ to regularize the model and reduce the risk of overfitting. The batch size was chosen to be 100, which allowed for efficient processing of large amounts

of data during each iteration. Finally, the number of epochs was set to 1000, ensuring that the model had sufficient time to converge and produce accurate results.

Comment 3: A native language editing should be done and its certificate provided.

Response: We deeply regret any deficiencies in the English expression used in our manuscript. As such, we have taken great care to meticulously review and revise the manuscript for any grammatical, syntactical, or spelling errors, in order to ensure that the manuscript is presented in the best possible light. Furthermore, we have proactively sought out the assistance of highly qualified native speakers to help improve the quality of the manuscript, both in terms of its scientific content and its linguistic accuracy. Our commitment to producing a manuscript of the highest quality remains unwavering.

Comment 4: Abbreviations in the abstract must be complete! A complete list of abbreviations, variables, and control parameters along with their values should be presented and described in separate tables in an appendix.

Response: Thanks for the advice. We have modified the **abstract** and listed the abbreviations, variables and control parameters in **Appendix A**

Table 9.**Abstract:**

Large-scale model sizes and expensive computing costs cause the challenge of deploying and applying large pre-trained models. Hence, this paper presents a novel Triple Concepts Attention Mechanism and a lightweight TCAMixer model for edge devices to classify texts. Furthermore, the TCAMixer abstracts textual concepts in a human way, which is unmatched by other counterparts such as pNLP-Mixer (a projection-based MLP-Mixer model for Nature Language Processing) and HyperMixer (a hyper network using dynamic token-mixing layers). Experimental results on several public datasets demonstrate that the TCAMixer outperforms the counterparts by a significant margin, for example, achieving 3% higher accuracy with a smaller model size of 0.177M. Additionally, the TCAMixer achieves a performance of 85% to 98.7% compared to that of large pre-trained models but only occupies 1/3000 to 1/2000 of their size on most test datasets.

Nomenclature	
1D	One Dimensional
CNN	Convolutional Neural Network
MLP	Multi-layer Perceptron
NLP	Nature Language Processing
CV	Computer Vison
TCA	Triple Concepts Attention
LDA	Latent Dirichlet Allocation
FLOPs	Floating Point Operations
SOTA	state-of-the-art

Appendix A

Table 9

The description of notations

Notations	Description	Notations	Description
\mathbf{X}	token matrix	\mathbf{U}_{tm}	outputs of the token-mixing layers
m	number of tokens	\mathbf{U}_{SA}	outputs of the self-attention layers
d	embedding dimensions	g	number of groups
\mathbf{X}_w	Word-level Concepts	k	width of the convolution
σ	normalization	e	number of paddings
p	sampling probability	s	number of stride
N	number of layers	r	dilation rate
\mathbf{X}_t	Topic-level Concepts	\mathbf{X}_d	Document-level Concepts

Comment 5: The final outputs should be compared and verified with at least 3 to 5 recent similar studies.

Response: Your suggestion is very pertinent. At present, there is a dearth of related research in this field. To the best of our knowledge, **MobileBERT**, **pNLP-Mixer**, and **HyperMixer** are the only studies that have been conducted. **In section 5.3**, we have conducted a comprehensive comparison of these three methods and the results are tabulated below for clarity and reference.

Table 8

The accuracy and number of parameters of different models on each public data sets. To reduce the size of the layout, some dataset names are abbreviated, such as "AGN." means AGNews, "Ama." means Amazon, "DBp." means DBpedia and "Hyp." means Hyperpartisan. Top-3 scores are colored by red, gold and green.

Model	Accuracy									Param.
	AGN.	Ama.	DBp.	Hyp.	IMDb	Yelp	SST	CoLA	QQP	
RoBERTa	/	/	/	87.40	95.30	/	96.70	67.80	90.20	125
XLNet	95.55	/	99.40	/	96.21	98.63	94.40	69.00	90.40	240
Bert Large	/	97.37	99.36	/	95.49	/	93.70	/	/	340
UDA	/	96.50	98.91	/	95.80	97.95	/	/	/	340
Bert-ITPT	95.20	/	99.32	/	/	/	/	/	/	340
gMLP	/	/	/	/	/	/	94.80	/	/	365
Longformer	/	/	/	94.80	95.70	/	/	/	/	149
MobileBERT	/	/	/	/	/	/	92.80	51.10	70.50	25.3
pNLP-Mixer	91.03	93.50	98.40	89.20	82.90	91.70	80.90	69.94	84.90	5.3
HyperMixer	/	/	/	/	/	/	80.70	/	83.70	12.5
TCAMixer	91.30	93.30	98.10	90.70	85.20	94.00	82.10	72.30	84.60	0.177

Comment 6: A detailed paragraph of future work should be added to the conclusion.

Response: Thanks for your suggestion, we have added more details for

future work at the end of conclusion.

However, this is not the performance limit of the TCAMixer model. In the next step, optimization will be carried out specifically for the TCAMixer model, applying advanced training techniques (such as pre-training) to make its performance not inferior to current large pre-trained models.

Thanks for all the valuable reviews which help us improve the quality of our manuscript and give us more insights into the research.

Best regards,

Xiaoyan Liu, Huanling Tang, Jie Zhao, Quansheng Dou, Mingyu Lu

2023.3.24