A Triplet Network Approach for Chinese Confusing Text Classification

Rui Xu[1], Cheng Zeng[1,2*], Yu Jin Liu[1], Peng He[1,3], Min Chen[1]
[1]School of Computer Science and Information Enginerring, Hubei University, Wuhan, China
[2]School of Artificial Intelligence, Hubei University, Wuhan, China
[3]School of Cyber Science and Technology, Hubei University, Wuhan, China

***Abstract*—The pre-trained model in the Chinese text classification task has made significant progress. However, there is a lot of semantically ambiguous and confusing text in the Chinese text, which has a negative impact on the classification model. A triplet network approach for Chinese confusing text classification is proposed to address this problem. This method improves the traditional triplet network's way of randomly constructing sample combinations, compares the feature similarity between the screened confusing text, straightforward text, and ordinary text, and improves the clustering effect of Chinese text features. At the same time, embedding text and label are jointly learned in the same latent space to learn the similarities and differences between texts of the same category and texts of confused categories. Experiments on multiple Chinese text classification datasets demonstrate the negative impact of confusing text on model accuracy and verify the method's effectiveness in this paper.**

***Keywords*—text classification; triplet network; confusing text**

I. Introduction

As a semantic language, Chinese differs from a morphological language (e.g., English). Changing a small amount of characters will lead to a massive difference in Chinese semantics. This special nature makes the Chinese language highly complex and diverse, often resulting in confusing text that is difficult to learn. The personality characteristics of Chinese confusing text are not evident, usually contain more confusing words, and the feature similarity with the confused category is high, so it is difficult to distinguish accurately. As shown in Fig. 1, the most confusing text is concentrated around the classification decision line, and the model cannot effectively identify the ground-truth class.

The confusing text belongs to the research category of the hard sample, and a series of research on the hard sample is currently carried out in the field of computer vision[1-2]. Wang et al. [3] used data augmentation technology to balance the data volume of the hard sample and easy sample, which improved the model performance to a certain extent. Shrivastava et al. [4] proposed a hard example mining algorithm to add the screened hard sample to a new training batch for training, but this method is generally used for mining hard negative examples. Compared with the processing of the dataset, Jiang et al. [5] changed the model structure. They used the supervised contrastive learning mode to optimize the loss of hard negative examples and increase the distance between different classes. In fact, the essence of processing confusing text in the hard sample is to reduce the distance of the same category and expand the distance of the confusing category. In pedestrian re-identification, Cai et al. [6] used the characteristics of the triplet network to effectively draw similar samples closer and push away dissimilar samples.

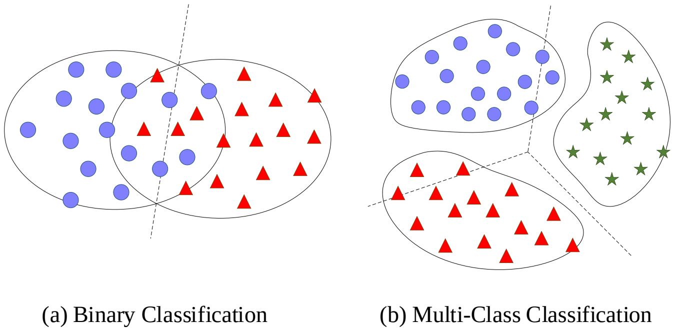

(a) Binary Classification (b) Multi-Class Classification

Fig. 1. Confusing text

However, using the traditional triplet network to construct sample combinations randomly, there are still noticeable differences between the feature vectors of the same class in Chinese texts. So, we propose based on a triplet network for confusing text in Chinese text classification. This method suppresses confusing text from deviating from the gold label by learning text feature information and label features in the same latent space. Moreover, improve the way of randomly constructing sample combinations in the traditional triplet network, let the confusing text select the positive and negative texts from the straightforward texts to form a triplet sample combination. On this basis, add a negative example selected from ordinary text to further improve the model's generalization ability and pass the obtained sample combination to the triplet network for targeted training. Let the model dig deep into the feature coding differences between the confusing text and the straightforward text in Chinese predictions to improve the classification effect of the model.

II. Related Work

A. Chinese Confusing Text

Chinese confusing text restricts the performance of existing models in text classification tasks to a certain extent, and the confusing text is shown in Table 1. In Chinese text classification, the confusing text tends to be confused due to vague opinions or more confusing words. A thorny issue is how to effectively avoid the negative impact of confusing text on the model.

DOI reference number: 10.18293/SEKE23-037

TABLE I. CONFUSING TEXT EXAMPLE

Confusing text	True label	Wrong label
科研人员发现癌症预警和诊断方法	Health	Technology
哇！这一次成功真的不容易	Happiness	Surprise
有点慢，其他的还好	Positive	Negative

Prabhakar et al. [7] combined the attention mechanism and the Focal Loss function, and some text category confusion has been improved to a certain extent. Xu et al. [8] used a graph neural network combined with an attention mechanism to learn the feature differences between confusing legal texts. It has a good effect on legal texts but ignores the relevance between legal entries of the same type. Therefore, how to further cluster texts of the same class while expanding the distance of confusing class is an essential idea for studying confusing text. Surprisingly, in pedestrian re-identification, the triplet network has outstanding performance in dealing with pedestrians' overlapping and confusing problems, which can solve this problem very well. Therefore, we uses the triplet network to train confusing text and optimizes and improves it according to the characteristics of Chinese texts.

B. *Triplet Network*

The triplet network is developed from the Siamese network. Chopra et al. [9] proposed the Siamese network to solve the problem in the field of face recognition that the model cannot accurately identify personality characteristics due to similar face structures. The Siamese network uses two neural networks with the same structure and shared weight and inputs a positive face picture and a negative face picture into the model. After calculation, the feature similarity of the two samples can be obtained, which can be effectively increased through training—the distance between different classes.

However, the Siamese network is more sensitive to anchor samples, and the error in distinguishing different individuals in the same group category is relatively large. In this regard, Hoffer proposed that the triplet network [10] uses anchor samples, positive and negative examples to form a training group and uses the neural network model shared by three weight to extract input features for triplet loss calculation, which effectively solves the problem in the Siamese network. For problems with poor individual recognition ability in the same class, the triplet network performs better than the Siamese network on multiple tasks. Especially for pedestrian re-identification [11], the triplet network can track the trajectories of overlapping pedestrians very well. The triplet network can also be used to compare the feature information of confusing text with other texts to mine rich semantic information. Additionally, when Chen et al. [12] added a negative sample to the traditional triplet training criteria when constructing triplet combinations, they found that it could better reduce intra-class differences and increase inter-class differences. Inspired by this work, we improve the traditional method of randomly selecting samples by triplets. We select positive examples, negative examples, and confusing texts from straightforward texts to form a triplet sample group. At the same time, adding a negative example selected from ordinary texts can Improve the model's generalization ability. Therefore, while measuring different categories of texts, Our method focuses on the samples selected from straightforward texts to dig out the standard text features of easy and hard samples.

III. APPROACH

Fig. 2 shows our triplet network's main idea for Chinese confusing text. The main steps include 1) fused label embedding in standard classification, 2) screening confusing text and straightforward text, and 3) training confusing text based on the triplet network.

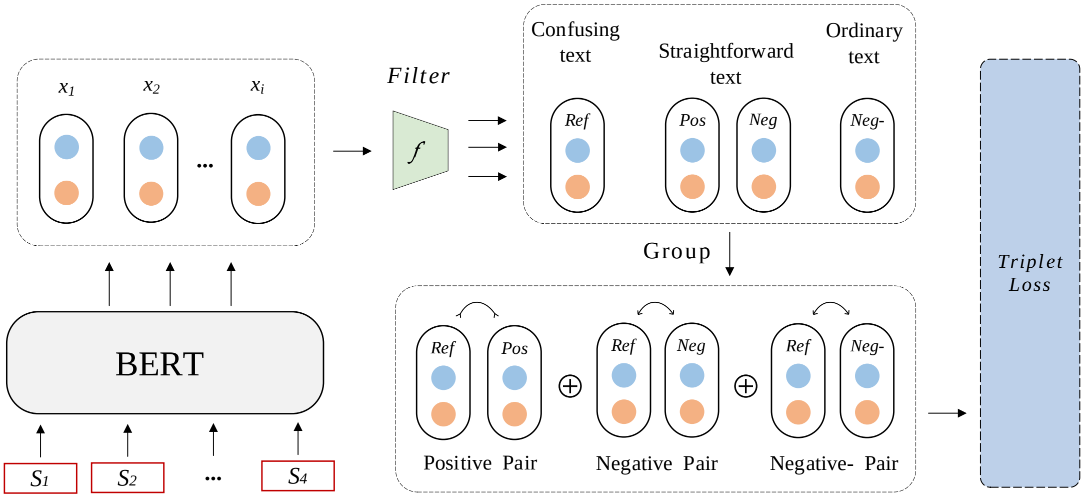


Fig. 2. The framework of our method

A. *Standard Classification*

Consider a text classification task with K classes. In the text representation stage, the label embedding work is fused [13], and the text and label features are learned in the same latent space, which can inhibit the confusing text from deviating from the golden label. The input text can be expressed as:

$$s_i =< [CLS], L_1, L_2, ..., L_K, [SEP], e_1, e_2, ..., e_t, [SEP] > \quad (1)$$

Corresponding author: Cheng Zeng (zc@hubu.edu.cn)

As shown in Fig. 3, use the encoder to obtain the feature representation h_i of the text s_i and applies the softmax function to calculate the predicted function calculates the predicted probability $\hat{y}_i$ of each category, and $\sum_{j=1}^{K} \hat{y}_i^{\,j} = 1$ formalized by:

$$x_i = \boldsymbol{W} h_i + b \tag{2}$$

$$\hat{y}_i = \mathrm{softmax}(x_i) \tag{3}$$

Where $\boldsymbol{W}$ and b are the learnable weight matrix and bias respectively. The maximum value in the probability $\hat{y}_i$ is the category prediction value of the current input text s_i . In classification tasks, we usually use cross-entropy (CE) loss. For a dataset $\{x_i, y_i\}_{i=1,\dots,N}$ containing N samples, define the following cross-entropy loss:

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} y_i^j \log(x_i^j) \tag{4}$$

Although CE loss works well in most cases, since the input sample label is represented as a one-hot vector. CE loss is not sensitive to obfuscated text, so it cannot effectively deal with it in the dataset.

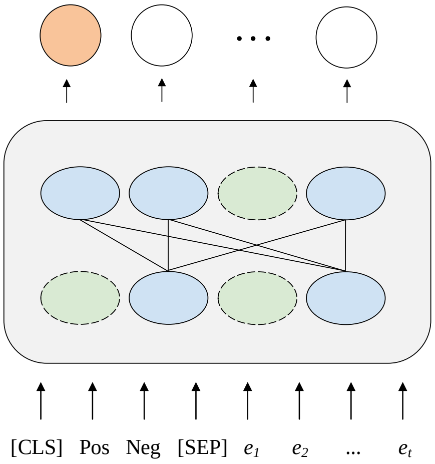


Fig. 3. Classifier

B. Screening Confusing Text and Straightforward Text

The input text s_i is subjected to standard classification to obtain the probability predictions of different classes. Then a filter function is designed to filter confusing text and straightforward text. For Chinese confusing text, as it is difficult for the classifier to learn the personality characteristics of the correct class from the current text, the similarity to the feature vector of the confused category is high, and the prediction scores of the two classes are very close. The straightforward text can quickly converge during model training, and the error between the predicted and ground-truth class is small. This kind of text is also defined as a easy sample, and ordinary text is other texts. From the perspective of the loss function, the loss of easily confusing text is relatively large during training, and the loss of straightforward text is relatively small. To this end, we introduces a screening strategy for these two texts.

The text whose difference between the first two categories of predicted probabilities is within threshold λ will be judged as confusing text. At this time, the text representation of this text is screened out, and the formalization is as follows:

$$\lambda \ge \left|\max(\hat{y}_i) - \max\left(\hat{y}_i - \max(\hat{y}_i)\right)\right| \tag{5}$$

The screening rules for straightforward text are as follows:

$$\max(\hat{y}_i) \ge \sum\nolimits_{j=1,\, j\ne\max(\hat{y}_i)}^{K} \hat{y}_i^{\,j} \tag{6}$$

Straightforward text is essential in shrinking confusing text to the correct class, more effectively comparing the feature similarity of two type of text, and improving the model's ability to classify confusing text.

C. Training Strategy based on Triplet Network

For the Chinese confusing text training strategy, first randomly select positive and negative samples from straightforward texts, construct a triplet sample group by combining the two texts with the confusing text, and add a negative sample randomly selected from ordinary text. The triplet loss function is obtained through the triplet network to make the training target closer to the distance between the anchor sample and the positive example while keeping the anchor sample away from the negative example. The modified triplet loss is designed as follows:

$$\begin{aligned} L_{tl} = &\sum_{i,j,k}^{N} [d(x_i, x_j) \text{ - } d(x_i, x_k) + \delta_1]_+ \\ &+ (1-\beta)\sum_{i,j,v}^{N} [d(x_i, x_j) \text{ - } d(x_i, x_v) + \delta_2]_+ \end{aligned} \tag{7}$$

where $[z]_+ = \max(z, 0)$, x_i is confusing text, x_j and x_k are straightforward text, and x_v is ordinary text.

The first item is called the strong push to build from the straightforward texts, and the second is the weak push to add the selection and construction from the ordinary texts to play a balancing role. In order to allow the model to dig out the features in the straightforward text deeply, set the weight of the first item to be greater than the second item, set to 0.2. The threshold δ_1 and δ_2 is a margin that is enforced between positive and negative pairs. $d(x_i, x_j)$ represents the feature similarity distance between samples. We select Euclidean distance as the distance measurement method, and the sample feature vector is mapped to the Euclidean distance space to achieve the goal of triplet learning.

$$L = \alpha L_{tl} + (1-\alpha) L_{ce} \tag{8}$$

In the training, we designs and uses an optimized objective function based on the triplet loss function to ensure that the ordinary text is not affected while training the confusing text.

The objective function is shown in Eq.8. α is a hyperparameter Used to adjust the weight of both.

IV. Experiments and Analysis

A. Datasets

We conduct experiments on Chinese text classification benchmark datasets with various granularities. 1) *nlpcc2014* is derived from the emotion recognition of Weibo comments in the NLPCC2014 task. It is a dataset with seven classes of emotion classification tasks. 2) *waimai_10k* is a two-category sentiment classification task dataset for takeaway meal evaluation. 3) *THUCNews* is based on the historical data of Sina News RSS subscription channels. After data cleaning, reintegration, and division into finance, stocks, science, society, politics, and entertainment, a total of six classes of news subject classification datasets. 4) *SHNews* uses the open-source Sohu news dataset for data cleaning to remove some missing label data in the data. The dataset has 12 categories, including entertainment, finance, real estate, tourism, technology, sports, health, education, automobiles, news, culture, woman. In the data preprocessing stage, text data is normalized by removing username cards, special character fragments, and improving the standardization of the data. The experimental data was partitioned into training, testing, and validation sets at a ratio of 8:1:1. See Table 2 for detailed statistics on datasets.

TABLE II. Dataset Statistics

Dataset	Classes	Type	Train	Dev.	Test	Length
nlpcc2014	7	Sentiment	13324	2855	2855	50
waimai_10k	2	Sentiment	8391	1798	1798	40
THUCNews	6	Topic	48000	6000	6000	20
SHNews	12	Topic	22699	5755	5764	20

B. Implementation Details

The experimental parameters in this paper mainly include classifier model parameters, confusing filter hyperparameters, and triplet loss parameters. The model is trained using the Adam gradient descent algorithm, the batch number is 128, the maximum number of rounds is set to 20 rounds, and the initial learning rate is set to 2E-5. According to the specific characteristics of different datasets, the parameter sensitivity analysis of the confusing category threshold $\lambda \in \{0.02 \sim 0.2\}$ of the confusing filter is carried out. The optimal parameter is finally selected as the experimental parameter.

C. Baseline Methods

In order to evaluate the classification effect of the proposed method on confusing text, we uses BERT [14] as the benchmark encoder. It selects methods that perform better in hard sample and confusing text to conduct comparative experiments.

- **EDA**: Through data enhancement on Chinese confusing text, balance the amount of data to improve the learning ability of the model for confusing text.
- **Focal-Loss**: Build a hard sample loss function to alleviate the problem that a small amount of confusing text and a large amount of ordinary text contribute differently to classifier learning [15].
- **H-SCL**: Supervised Contrastive Learning for Chinese Confusing Text, Better Performance than Random Sampling for Unsupervised Contrastive Learning.
- **TN**: Targeted training on obfuscated text using traditional triplet network methods.

D. Analysis of Experimental Results

The Accuracy of different methods on the four datasets are shown in Table 3. It can be seen from the experimental results that the model processed for confusing text is generally better than the benchmark model BERT, and the accuracy rate has been improved to a certain extent. It is highly beneficial to optimize the model's training method and process the text data while retaining the benchmark encoder. This can mitigate the adverse effects of confusing text. Among them, the method proposed in this paper performs better overall on the four Chinese text classification datasets than the other listed methods. The accuracy rates on *waimai_10k*, *nlpcc2014*, *THUCNews*, and *SHNews* have increased by 1.45 %, 1.18%, 0.95%, and 0.96%, respectively. Compared with traditional triplet network methods, the proposed method has improved the accuracy rates by 0.49%, 0.14%, 0.40%, and 0.26% respectively. Although the results are not particularly impressive, they provide some insights and ideas for studying the issues with Chinese confusing text. At the same time, if the label embedding technology (LE) is removed, the classification performance of the method in this paper will decline to a certain extent, proving the proposed method's rationality and effectiveness. To demonstrate the effectiveness of our method in addressing the issue of text obfuscation and check its compatibility with various classification models, we conducted relevant experiments using other pre-trained models like RoBERTa[16] and XLNET[17], which are indicated in Table 4. The experimental results show that the proposed method can still significantly enhance the classification performance of the replaced BERT when other classification models are used. Further clustering of confusing text in the data can improve the classification performance of the model. The experimental results confirm the necessity of handling text obfuscation issues in Chinese text classification research.

TABLE III. Experimental Results

Method	waimai_10k	nlpcc2014	THUCNews	SHNews
BERT	90.04	64.57	93.57	86.55
(+) EDA	90.71	64.74	93.03	86.21
(+) Focal-Loss	90.82	65.53	94.25	87.00
(+) OHEM	91.10	65.92	94.32	87.18
(+) H-SCL	90.88	65.49	94.42	86.95
(+) TN	91.00	65.81	94.12	87.25
(+) Ours w/o LE	91.32	**65.95**	93.87	87.35
(+) Ours	**91.49**	65.75	**94.52**	**87.51**

TABLE IV. Performance of our method on other models

Method	waimai_10k	nlpcc2014	THUCNews	SHNews
RoBERTa	90.88	64.71	94.02	86.66
(+) Ours	**90.99**	**65.28**	**94.18**	**87.65**
XLNet	91.27	64.46	93.62	86.39
(+) Ours	**91.43**	**64.53**	**93.68**	**87.00**

E. *Hyperparameter Influence*

The experiment was designed to explore the impact of the confusion text filter threshold parameter λ on the model performance. The filter threshold determines the level of tolerance for confused text, and the optimal parameter for the filter threshold varies across different datasets due to differences in text characteristics, quality and length. This paper aims to explore the range of data confusion ratios for which the proposed method enhances the performance of the model to a significant extent. The condition for such an analysis is that all other hyperparameters are set to their respective optimal values. The results of our experiments are depicted in Fig. 4. The selection of filter thresholds has a significant impact on the improvement of the model's performance. Our experiments show that the optimal range for the filter threshold selection is between 0.03 and 0.1, where the overall improvement in the model's performance is most significant.

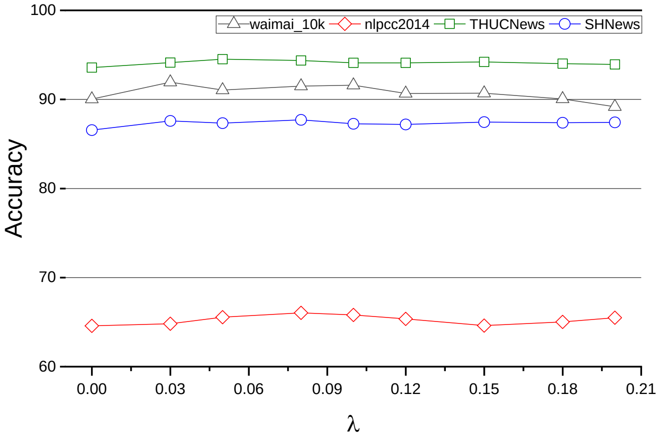


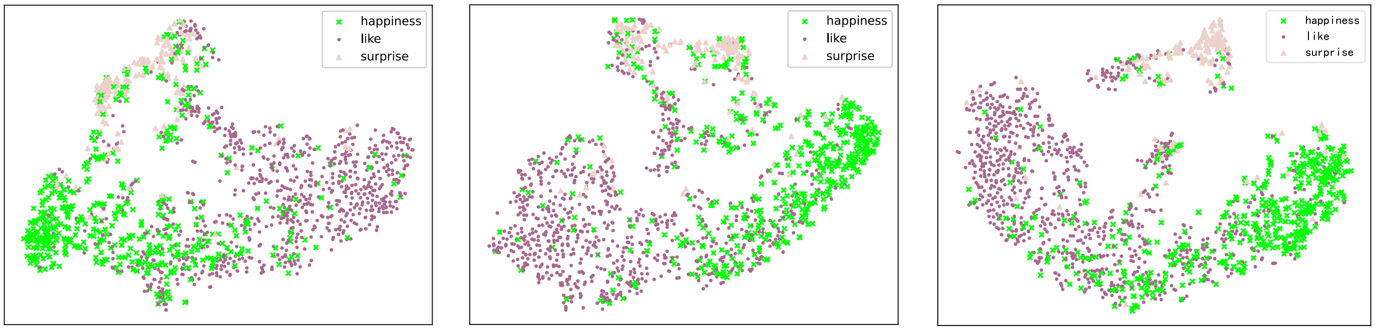
Fig. 4. Accuracy under different confusion ratios

In order to further illustrate the performance improvement of the triplet loss function proposed in this paper compared with the traditional triplet loss function, we selected *happiness*, *like*, and *surprise* of data from the *nlpcc2014* test dataset and used the t-SNE (t-distributed Stochastic Neighbor Embedding) visually display the text features, as shown in Fig. 5.

Fig. 5(a) shows the two-dimensional space representation of the feature vectors of the three categories after the benchmark model BERT is trained. The distribution of the test set data in the embedding space is relatively scattered, the problem of text confusion is prominent, and the model's classification performance is restricted. Fig. 5(b) uses the traditional triple loss to build a model for training. Compared with Fig. 5(a), there is a significant difference in the distance between different categories, but the problem of interleaved stacking of *happiness* and *like* categories is prominent, and the distance between classes is relatively scattered. The use of triplet loss cannot eliminate the wrong movement of samples but can only constrain and suppress this negative trend, so there are still many outlier samples interlaced with each other in Fig. 5(b) and Fig. 5(c). Fig. 5(c) is the optimized model using the method of this paper. It is obvious that the distance between the same categories has been shortened, and the confusion problem has been significantly improved. The outlier text has also been reduced, which shows that the confusing text has been corrected. Strengthen the model's ability to distinguish sentiment data.

V. Conclusion

In this paper, we propose a text classification method based on triplet network to investigate the impact of confusing text on Chinese text classification tasks. Our proposed method offers a simple yet effective way to enhance the overall performance of the model without making any structural changes. The experimental results presented in this paper confirm the effectiveness and rationality of our approach. However, there are also some deficiencies, such as the requirement to fine-tune the filter threshold parameters for different datasets, which reduces flexibility. There is a need to propose targeted optimization methods for other areas of natural language understanding. In our subsequent work, we will endeavor to produce adaptive weights by taking into account both the overall quality and average distance of the dataset, to further optimize the detection of confusing text in various domains.



(a) BERT

(b) BERT+TN

(c) BERT+**Ours**

Fig. 5. The tSNE plots of the learned representations on the *nlpcc2014* dataset

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No.62102136) , the Key R & D projects in Hubei Province (No.2021BAA188, 2021BAA184, 2022BAA044) , the Science and Technology Innovation Program of Hubei Province (No.2020AEA008).

REFERENCES

[1] Zhu C, Hu Z, Dong H, et al. Construct informative triplet with two-stage hard-sample generation[J]. Neurocomputing, 2022, 498: 59-74.

[2] Shao X, Wei J, Guo D, et al. Pedestrian detection algorithm based on improved faster rcnn[C]//2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2021, 5: 1368-1372.

[3] Wang X, Shrivastava A, Gupta A. A-fast-rcnn: Hard positive generation via adversary for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2606-2615.

[4] Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 761-769.

[5] Jiang R, Nguyen T, Ishwar P, et al. Supervised Contrastive Learning with Hard Negative Samples[J]. arXiv preprint arXiv:2209.00078, 2022.

[6] Cai X, Liu L, Zhu L, et al. Dual-modality hard mining triplet-center loss for visible infrared person re-identification[J]. Knowledge-Based Systems, 2021, 215: 106772.

[7] Prabhakar S K, Rajaguru H, Won D O. Performance Analysis of Hybrid Deep Learning Models with Attention Mechanism Positioning and Focal Loss for Text Classification[J]. Scientific Programming, 2021, 2021: 1-12.

[8] Xu N, Wang P, Chen L, et al. Distinguish confusing law articles for legal judgment prediction[J]. arXiv preprint arXiv:2004.02557, 2020.

[9] Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 1: 539-546.

[10] Hoffer E, Ailon N. Deep metric learning using triplet network[C]//Similarity-Based Pattern Recognition: Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, October 12-14, 2015. Proceedings 3. Springer International Publishing, 2015: 84-92.

[11] Zeng K, Ning M, Wang Y, et al. Hierarchical clustering with hard-batch triplet loss for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 13657-13665.

[12] Chen W, Chen X, Zhang J, et al. Beyond triplet loss: a deep quadruplet network for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 403-412.

[13] Xiong Y, Feng Y, Wu H, et al. Fusing label embedding into bert: An efficient improvement for text classification[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021: 1743-1750.

[14] Cui Y, Che W, Liu T, et al. Revisiting pre-trained models for Chinese natural language processing[J]. arXiv preprint arXiv:2004.13922, 2020.

[15] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[16] Liu Y, Ott M, Goyal N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv preprint arXiv:1907.11692, 2019.

[17] Yang Z, Dai Z, Yang Y, et al. Xlnet: Generalized autoregressive pretraining for language understanding[J]. Advances in neural information processing systems, 2019, 32.