



HAL
open science

Employing Contrastive Strategies for Multi-label Textual Emotion Recognition

Yangyang Zhou, Xin Kang, Fuji Ren

► **To cite this version:**

Yangyang Zhou, Xin Kang, Fuji Ren. Employing Contrastive Strategies for Multi-label Textual Emotion Recognition. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.299-310, 10.1007/978-3-031-03948-5_24 . hal-04178710

HAL Id: hal-04178710

<https://inria.hal.science/hal-04178710v1>

Submitted on 8 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Employing Contrastive Strategies for Multi-label Textual Emotion Recognition

Yangyang Zhou¹[0000-0001-7476-0908], Xin Kang²[0000-0001-6024-3598], and
Fuji Ren³[0000-0003-4860-9184]

¹ Tokushima University, Tokushima, Japan
dango_agnes@foxmail.com

² kang-xin@is.tokushima-u.ac.jp

³ ren@is.tokushima-u.ac.jp

Abstract. Textual emotion recognition is an important part of the human-computer interaction field. Current methods of textual emotion recognition mainly use large-scale pre-trained models fine-tuning. However, these methods are not accurate enough in the semantic representation of sentences. Contrastive learning has been shown to optimize the representation of vectors in the feature space. Therefore, we introduce the contrastive strategies to the textual emotion recognition task. We propose two approaches: using self-supervised contrastive learning before fine-tuning the pre-trained model, and using contrastive training on the same inputs during fine-tuning. We experiment on two multi-label emotion classification datasets: Ren-CECps and NLPCC2018. The experimental results demonstrate that the latter contrastive approach effectively improves the accuracy of emotion recognition.

Keywords: Textual emotion recognition · Multi-label classification · Large-scale pre-trained models · Contrastive learning · Fine-tuning.

1 Introduction

Emotion recognition is the process by which a machine identifies human emotion. The response of the machine based on the user’s emotional state can enhance the user’s experience. Emotion recognition can be applied in many realistic scenarios. For example, obtaining user satisfaction with a product can help a sales platform develop a better sales strategy [25]. Machines usually recognize the emotion by acquiring emotion-induced physiological signals or behavioral information. In this paper, we focus on emotion recognition from the text.

According to the level of the content, emotion analysis of the text can be divided into word level, sentence level, chapter level and target level [32]. Building an emotion dictionary is a common method for word-level emotion analysis [14]. The information of each word in a sentence/chapter is pooled or concatenated to represent the sentence-level/chapter-level emotion as a whole [21]. The target-level target refers to the emotion of an entity or an attribute [16]. Regardless of the level, words that contain emotional tendencies are the most critical factors in the emotion analysis process [22].

Treating pooled or concatenated word vectors as sentence vectors will inevitably lose information. To solve this problem, some language models tend to design semantic tasks to train feature representations of whole sentences [20]. Large-scale pre-trained language models, such as bidirectional encoder representations from transformers (BERT), predict whether the sentences are contextually relevant at the first token position [7]. Semantic text similarity measures the meaning similarity of sentences. In this kind of task, the state-of-the-art results are almost obtained by large-scale pre-trained language models [4].

In large-scale pre-trained language models, the representation of word vectors is correlated with word frequency, resulting in uneven distribution of word vectors in the feature space [12]. The idea of contrastive learning is to represent the features uniformly in the feature space [29]. Therefore, Gao et al. introduce contrastive learning into the training process of large-scale pre-trained language models, which in turn achieves uniform distribution of sentence vectors in the feature space [8].

This paper is devoted to investigating how contrastive strategies affect the accuracy of multi-label emotion recognition from text. Using unsupervised contrastive learning to adjust parameters before fine-tuning a large-scale pre-trained language model has no significant benefit for emotion recognition. In contrast, adding a supervised contrastive learning approach to fine-tune the large-scale pre-trained language model is beneficial for accuracy improvement. We experiment on two multi-label textual emotion classification datasets based on our methods.

The main contribution of this paper is to introduce contrastive strategies in the training phase of multi-label emotion recognition from text to further improve the recognition accuracy based on large-scale and training language models. The remainder of this paper is organized as follows. Section 2 presents a review of the related work. Section 3 gives the details of the mechanism of the proposed methods. The experimental setup is presented in Section 4. The system performance analysis is presented in Section 5. Finally, Section 6 presents the conclusion and future work of this study.

2 Related Work

Textual emotion recognition has recently become a hot topic due to its commercial and academic potential[6]. In tasks such as product comments analysis, sentiment analysis is generally performed with positive and negative [19]. We classify emotion more fine-grained into multiple categories. Some scholars regard text analysis as a single-label classification task [3], which is different from the complex emotional states in real-world conditions. In this article, we regard textual emotion analysis as a multi-label classification task, with the goal of recognizing all possible emotions in the textual expression [17].

Approaches for emotion recognition from the text can be classified into 4 categories: rule-based approaches, classical learning-based approaches, deep learning approaches, and hybrid approaches [1]. The emotion rules are usually extracted

by statistics or linguistic concepts. These rules rely heavily on preprocessing steps including tokenization, lemmatization, POS tagging, and stop words removal. Udochukwu et al. proposed a rule-based emotion recognition model called Ortony, Clore, and Collins model [28]. The results show that the rule-based approach is very sensitive to text quality. Support vector machine is a common classification algorithm used in classical learning-based approaches. This algorithm relies on extracting and selecting features with the most information gain and then outputting the optimal hyperplane. Experiments by Anusha et al. show that the important part of the sentence is essential for improving the results [2]. The convolutional neural network, recurrent neural network and transformer are the commonly used frameworks for deep learning approaches. With a large amount of data information and automatic feature engineering, deep learning approaches achieve better results. Hybrid approaches inherit the advantages and disadvantages of the above methods.

Most of the datasets involve a specific domain rather than an open domain, so there is also some work to make up for the limitations of the datasets by introducing background knowledge or prior information [24]. Recently, language models pre-trained on large-scale datasets have been shown to be effective in capturing contextual information and benefiting various downstream tasks through fine-tuning [9]. Language models have been classified into autoregressive and autoencoder language models based on whether there is information leakage. Autoregressive language models, represented by Generative Pre-Training 2 [23], predict words using frameworks such as unidirectional transformers. The autoencoder language model, represented by BERT [7], uses a framework such as bidirectional transformers to incorporate contextual information into the model. Since different mask strategies for autoencoder language models during training and testing lead to errors, we try to reduce this error by the contrastive strategy. In this paper, we introduce two different contrastive strategies based on the autoencoder language model BERT.

At first, contrastive learning used a self-supervised approach to avoid the cost of annotating large-scale datasets in the field of computer vision [10]. Contrastive learning constructs positive and negative samples by means of data augmentation and adjusts the distribution of features in space. Later, contrastive learning was also gradually used in the field of natural language processing. The most commonly used methods of text data augmentation are replacement, insertion, deletion and swap [26]. Gao et al. used the property of dropout to implicitly construct positive samples and achieved state-of-the-art results on several semantic text similarity tasks [8]. Liang et al. added a regularization strategy to improve the robustness of the model based on the inconsistency of dropout output [15]. Inspired by the above work, we use dropout and construct two large-scale pre-trained models based on fine-tuning to explore the influence of contrastive strategies on emotion recognition.

3 Model Architecture

In this paper, we explore how the contrastive strategy can be introduced into the textual emotion recognition model to achieve better results. The process of textual emotion recognition is described as follows. Given a sentence s , we use an encoder to map s to a representation s' . Then, we use a classifier to transform s' into the emotion e .

3.1 Dropout and Contrastive Strategy

Contrastive learning adjusts the representation of features by decreasing the distance between similar samples and increasing the distance between different samples. Similar samples are generally transformed from the original sample, and we use the dropout method to transform in this article.

Dropout is a method commonly used to prevent overfitting of neural networks [27]. By dropping some neurons randomly during the training process, the model can be prevented from over-relying on some neurons and dependencies. Since dropout has randomness, the model will produce similar but not the same output each time with the same input. According to the property of dropout, it is easy to construct two similar samples.

3.2 Encoder and Classifier

We construct a framework consisting of an encoder and a classifier.

For encoding, we use a large-scale pre-trained model, load the weights, and fine-tune them. In this paper, we use BERT as an encoder. To be unified with BERT, each sentence is prefixed with a $[cls]$ token at first before it is fed into the model. BERT is designed with two pre-training tasks. The text involves the masked language modeling task, while the $[cls]$ token involves the next sentence prediction task. Thus, the first token of the output contains the sentence semantics to some extent. We use this token for classification.

We take the fully connected network with bias as a classifier. Since the first token does not contain the sequence axis, we can directly map it to the emotion state e .

$$e = (e_1, e_2, \dots, e_n) \quad (1)$$

in which n is the number of emotion categories. When the output of emotion e_i exceeds the threshold, we believe that sentence s expresses the emotion e_i .

We introduce the contrastive strategy into the encoding and classification parts separately to get two different recognition models.

3.3 Model with Contrastive Strategy added to the Encoding part

The semantics represented by sentence vectors is the most critical part of emotion recognition. Different sentence vectors represented by $[cls]$ token in BERT have been confirmed to be very similar and do not represent the semantics well

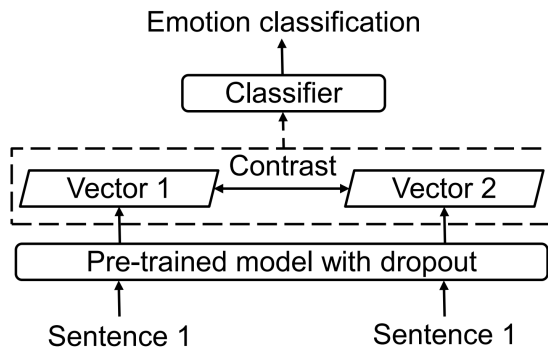


Fig. 1. Overall framework of the model with contrastive strategy added to the encoding part.

[12]. We try to enhance the semantic representation of sentence vectors with a contrastive strategy.

According to the description above, contrastive learning is a self-supervised training process, the same as what large-scale pre-trained models do. We continue to train the weights of the model with similar samples constructed by dropout on the basis of BERT. We expect that this weight for fine-tuning will improve the accuracy of emotion recognition.

As shown in Fig. 1, we fine-tune the encoder before classification. The encoder predicts two similar representations Vector1 and Vector2 by dropout from the same input Sentence 1. Using unsupervised contrastive learning, the parameters of the encoder can be fine-tuned so that the distribution of the encoded results in the feature space is more favorable for emotion classification. Based on this, we add the classifier at the top for supervised training.

3.4 Model with Contrastive Strategy added to the Classification part

In addition to self-supervised training, we can use the contrastive strategy for supervised training. Dropout only works at training time, i.e., the model is not consistent at training and prediction. To reduce the gap of the model between training and prediction, we close the distance between the outputs of the model after randomly dropping neurons.

Specifically, we take input during training twice through the model and add a Kullback-Leibler (KL) divergence between the two outputs. KL divergence is a metric used to estimate the distance between distributions [11]. When calculating the loss function, adding KL divergence ensures that the model outputs are closer after each random drop of neurons to reduce the effect of dropout on the results.

As shown in Figure 2, we supervised train the classifier and the encoder at the same time. Due to dropout, the same input Sentence 1 has similar results Vector 3 and Vector 4 when prediction. For the stability of results, we introduce

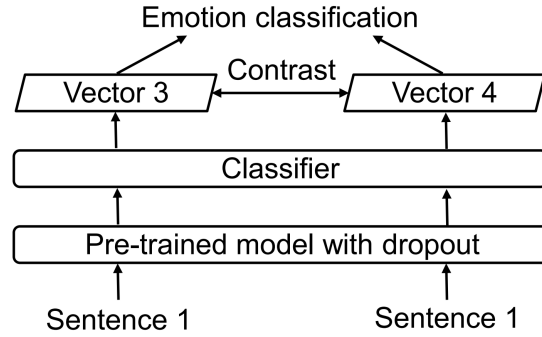


Fig. 2. Overall framework of the model with contrastive strategy added to the classification part.

the KL loss function during training to make Vector 3 and Vector 4 as close as possible.

4 Experiments Setup

4.1 Datasets

To verify that the contrastive strategy is effective in the emotion recognition from text, we select two multi-label emotion classification datasets as follows.

The Ren-CECps dataset selects Chinese blog texts and annotates 8 kinds of emotions: joy, hate, love, sorrow, anxiety, surprise, anger and expect. Emotions are in range $[0, 1]$. We labeled data points with emotion scores over 0 as 1 and others as 0 during preprocessing. We labeled the data points with all 0 emotion scores as neutral. The dataset can be regarded as a multi-label classification task where the average number of emotions expressed per sentence is about 1.45. This dataset includes a total of 27,091 sentences of training data and 7,681 sentences of test data. The details of this dataset can be found in the article [13].

The NLPC2018 dataset contains 5 kinds of emotions: happiness, sadness, anger, fear and surprise. As with the Ren-CECps dataset, we added the neutral to the labels. Compared to the Ren-CECps dataset, the average number of emotions expressed per sentence in the NLPC2018 dataset is less (about 1.15). This dataset includes a total of 4,611 sentences of training data and 955 sentences of test data. The details of this dataset can be found in the article [30].

4.2 Evaluation Metrics

Unlike the evaluation metrics for single-label classification tasks, the commonly used accuracy does not reflect the effectiveness of multi-label classification models well. To compare with other public models, we use the following evaluation metrics.

The F1 score is a combination of precision and recall. Micro F1 score is calculated overall for all categories, with errors when it is used for imbalanced datasets. Macro F1 score is calculated separately for different categories with the same weights and is strongly influenced by extreme precision and recall. Average precision (AP) is the average score of accuracy for recall from 0 to 1. For the metrics of micro-F1, macro-F1 and AP, higher values indicate better performance. Coverage error (CE) indicates the average number of labels that can cover the ground truth labels. Ranking loss (RL) is the proportion of misclassifications after weighting by the number of relevant labels. For the metrics of CE and RL, lower values indicate better performance.

4.3 Model for Comparison

We choose the model with the best-published results on both datasets to compare with our proposed models. The multi-label emotion detection architecture from sentence (MEDA-FS) [5] is a multichannel hierarchical model. MEDA-FS captures the underlying emotion-specified features through a multi-channel feature extractor and then predicts emotion sequences through an emotion correlation learner.

Commonly used baseline models for multi-label classification such as binary relevance and backpropagation for multi-label learning have been compared in the article of the MEDA-FS and will not be repeated in this paper.

To better control the variables, in addition to the MEDA-FS model that uses the pre-trained model as a feature extractor, we also compare fine-tuned models that do not use the contrastive strategy.

4.4 Experimental Details

We used the hugging face open source large-scale pre-training model BERT (chinese-bert-wwm-ext) as our backbone [31]. For the model with contrastive strategy added to the encoding part, 0.1M news titles are used for contrastive learning to adjust the semantic expressiveness of sentence embedding. For the model with contrastive strategy added to the classification part, we add KL loss between the replicated data points in a batch.

We used AdamW [18] optimizer to train the model. The learning rate of the classifier is 1e-2 and the learning rate of the backbone is 1e-5. When predicting, if the output result of an emotion is greater than the boundary threshold, we assume that the emotion is expressed.

5 Experimental Results and Discussion

5.1 Experimental Results

Table 1 shows the results of the Ren-CECps dataset. MEDA-FS uses the feature extraction approach for downstream tasks. BERT FT refers to fine-tuning

Table 1. The results of the Ren-CECps dataset. For the metrics of Micro F1, Macro F1 and AP, higher values indicate better performance. For the metrics of CE and RL, lower values indicate better performance.

	Micro F1	Macro F1	AP	CE	RL
MEDA-FS	60.76	48.31	76.51	2.2226	0.1062
BERT FT	65.42	53.57	81.99	1.9046	0.0871
CE model	61.78	45.80	79.58	2.0187	0.1008
CC model	66.16	53.75	82.00	1.9021	0.0868

approach with BERT on the dataset. The BERT FT improves from 60.76/48.31 to 65.42/53.57 in f1 scores compared to MEDA-FS. CE model means that the weights are adjusted with the self-supervised contrastive learning approach before fine-tuning. This method makes the evaluation metrics worse, especially the macro f1 score of 45.80 which is even lower than that of the MEDA-FS. CC model refers to the addition of supervised contrastive training to fine-tuning. This method improves the f1 score from 65.42/53.57 to 66.16/53.75 compared to the BERT FT. CC model also performs better in AP, CE and RL metrics.

Table 2. The results of the NLPCC2018 dataset. For the metrics of Micro F1, Macro F1 and AP, higher values indicate better performance. For the metrics of CE and RL, lower values indicate better performance.

	Micro F1	Macro F1	AP	CE	RL
MEDA-FS	63.02	49.42	77.12	1.7288	0.1681
BERT FT	59.05	49.94	85.13	1.2883	0.1114
CE model	60.67	43.40	84.74	1.2958	0.1137
CC model	58.68	50.57	85.09	1.2775	0.1090

Table 2 shows the results of the NLPCC2018 dataset. The results are similar to those in Table 1. In terms of evaluation metrics, the fine-tuning approach outperforms the feature extraction approach to a certain extent. BERT FT achieves scores of 85.13 on the AP metrics, which is the best results in the comparison. Compared to BERT FT, the macro f1 score of the CE model decrease from 49.94 to 43.40. Compared to BERT FT, the macro f1 scores of the CC model improve to 50.57. CC model also achieves scores of 1.2775 and 0.1090 on the CE and RL metrics. It proves that the model with a contrastive strategy added to the classification part is more effective in the textual emotion recognition task than to the encoding part.

5.2 Discussion

Compared to feature extraction, fine-tuning has better performance in downstream tasks. The large-scale pre-trained model utilizes a large corpus for self-supervised learning, and this corpus may deviate a little from the corpus used for the downstream task. The fine-tuning approach adaptively adjusts the model weights to make it more accurate in feature representation for downstream tasks. This is reflected in both Table 1 and Table 2. However, fine-tuning requires supervised learning, so the computational overhead is greater.

The self-supervised contrastive learning is adopted before fine-tuning, but the results become worse. We intended to represent the sentence semantics that was slighted in the pre-training task better by contrastive learning. However, this approach was not effective enough in the textual emotion classification task. The CE model does not perform well in all metrics for both datasets.

We originally thought the possible reason was that emotion classification did not depend entirely on sentence semantics. However, after fine-tuning by replacing the *[cls]* token with pooled word vectors, the results slightly decreased. This suggests that sentence semantics plays a bigger role in emotion classification than word embeddings. Therefore, the reason for the worse results may be that the semantics compared to contrastive learning is not exactly related to emotion. That is, a better comparison scheme may improve the results.

Another possible reason is the inconsistent training goals of the model at different stages. The CE model focuses on sentence similarity rather than classification when encoding. Compared to the CE model, the CC model has better performance because it adds the comparison of hidden states similarity to the classification part.

We tried fine-tuning the pre-contrastive training with data from the unlabeled training set. We combined the textual parts of Ren-CECs and NLPCC2018 (excluding labels) for unsupervised training. The results were worse than the classification results obtained with the additionally found title data. This indicates that the self-supervised learning corpus affects the training results. A more general, larger corpus facilitates parameter fine-tuning during unsupervised learning.

The contrastive training on the same sample representation is adopted during fine-tuning and the results are improved. This suggests that the contrastive strategy used in models with dropout is beneficial to the textual emotion classification task.

We experimented with the effect of different drop rates on the results. In this kind of task, the classification results get worse as the drop rate increases. We believe that dropout is a regularization method, so the larger the drop rate, the stronger the constraint on the model and the more unstable results predicted by the model. Under the contrastive strategy, stability training becomes more difficult as the drop rate increases. If the amount of data is large enough, a larger drop rate may have a better performance. However, in our experiment, the best result is achieved by the BERT model with a default rate of 0.1.

We also tried both contrastive learnings before and during fine-tuning. The results are better than the CE model but worse than the CC model. We believe that the two contrastive strategies are independent and do not affect each other. Overall, the strategy of contrastive training during fine-tuning is a better choice for the textual emotion classification task.

6 Conclusions

In this article, we introduce the contrastive strategies for multi-label textual emotion classification tasks. Based on the large-scale pre-trained model BERT, we propose two approaches: using self-supervised contrastive learning before fine-tuning the model, and using contrastive training on the same inputs during fine-tuning. We experiment with the effectiveness of the strategies on two multi-label emotion classification datasets: Ren-CECps and NLPCC2018. The experimental results demonstrate that using the contrastive strategy in the classification part is more effective in improving the accuracy of emotion recognition than using the contrastive strategy in the encoding part.

We use the `[cls]` token of the pre-trained model for classification in the experiments. In the future, we will investigate the relationship between word embeddings and `[cls]` token to minimize the loss of emotional information contained in sentences during training.

Acknowledgements

This research has been supported by JSPS KAKENHI Grant Number 19K20345 and Grant Number 19H04215.

References

1. Alswaidan, N., Menai, M.E.B.: A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge & Information Systems* **62**(8) (2020)
2. Anusha, V., Sandhya, B.: A learning based emotion classifier with semantic text processing. In: *Advances in intelligent informatics*, pp. 371–382. Springer (2015)
3. Bandhakavi, A., Wiratunga, N., Padmanabhan, D., Massie, S.: Lexicon based feature extraction for emotion text classification. *Pattern recognition letters* **93**, 133–142 (2017)
4. Chandrasekaran, D., Mago, V.: Evolution of semantic similarity—a survey. *ACM Computing Surveys (CSUR)* **54**(2), 1–37 (2021)
5. Deng, J., Ren, F.: Multi-label emotion detection via emotion-specified feature extraction and emotion correlation learning. *IEEE Transactions on Affective Computing* (2020)
6. Deng, J., Ren, F.: A survey of textual emotion recognition and its challenges. *IEEE Transactions on Affective Computing* (2021)
7. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018)

8. Gao, T., Yao, X., Chen, D.: Simcse: Simple contrastive learning of sentence embeddings. arXiv preprint arXiv:2104.08821 (2021)
9. Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Zhang, L., Han, W., Huang, M., et al.: Pre-trained models: Past, present and future. *AI Open* (2021)
10. Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F.: A survey on contrastive self-supervised learning. *Technologies* **9**(1), 2 (2021)
11. Kullback, S., Leibler, R.A.: On information and sufficiency. *The annals of mathematical statistics* **22**(1), 79–86 (1951)
12. Li, B., Zhou, H., He, J., Wang, M., Yang, Y., Li, L.: On the sentence embeddings from pre-trained language models. arXiv preprint arXiv:2011.05864 (2020)
13. Li, J., Ren, F.: Creating a chinese emotion lexicon based on corpus ren-cepceps. In: 2011 IEEE International Conference on Cloud Computing and Intelligence Systems. pp. 80–84. IEEE (2011)
14. Li, J., Xu, Y., Xiong, H., Wang, Y.: Chinese text emotion classification based on emotion dictionary. In: 2010 IEEE 2nd Symposium on Web Society. pp. 170–174. IEEE (2010)
15. Liang, X., Wu, L., Li, J., Wang, Y., Meng, Q., Qin, T., Chen, W., Zhang, M., Liu, T.Y.: R-drop: Regularized dropout for neural networks. arXiv preprint arXiv:2106.14448 (2021)
16. Liu, B.: Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* **5**(1), 1–167 (2012)
17. Liu, S.M., Chen, J.H.: A multi-label classification based approach for sentiment classification. *Expert Systems with Applications* **42**(3), 1083–1093 (2015)
18. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
19. Mai, L., Le, B.: Joint sentence and aspect-level sentiment analysis of product comments. *Annals of Operations research* **300**(2), 493–513 (2021)
20. Majumder, G., Pakray, P., Gelbukh, A., Pinto, D.: Semantic textual similarity methods, tools, and applications: A survey. *Computación y Sistemas* **20**(4), 647–665 (2016)
21. Majumder, N., Poria, S., Gelbukh, A., Cambria, E.: Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems* **32**(2), 74–79 (2017)
22. Quan, C., Ren, F.: An exploration of features for recognizing word emotion. In: Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). pp. 922–930 (2010)
23. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al.: Language models are unsupervised multitask learners. *OpenAI blog* **1**(8), 9 (2019)
24. Ren, F., Deng, J.: Background knowledge based multi-stream neural network for text classification. *Applied Sciences* **8**(12), 2472 (2018)
25. Satrio, D., Priyanto, S.H., Nugraha, A.K.: Viral marketing for cultural product: The role of emotion and cultural awareness to influence purchasing intention. *Montenegrin Journal of Economics* **16**(2), 77–91 (2020)
26. Shorten, C., Khoshgoftaar, T.M., Furht, B.: Text data augmentation for deep learning. *Journal of big Data* **8**(1), 1–34 (2021)
27. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15**(1), 1929–1958 (2014)
28. Udochukwu, O., He, Y.: A rule-based approach to implicit emotion detection in text. In: International Conference on Applications of Natural Language to Information Systems. pp. 197–203. Springer (2015)

29. Wang, T., Isola, P.: Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In: International Conference on Machine Learning. pp. 9929–9939. PMLR (2020)
30. Wang, Z., Li, S., Wu, F., Sun, Q., Zhou, G.: Overview of nlpcc 2018 shared task 1: Emotion detection in code-switching text. In: CCF International Conference on Natural Language Processing and Chinese Computing. pp. 429–433. Springer (2018)
31. Wolf, T., Chaumond, J., Debut, L., Sanh, V., Delangue, C., Moi, A., Cistac, P., Funtowicz, M., Davison, J., Shleifer, S., et al.: Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. pp. 38–45 (2020)
32. Zhang, L., Wang, S., Liu, B.: Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **8**(4), e1253 (2018)