

Article

Implicit Emotion Analysis Based on Improved Supervised Contrastive Learning

Linghui Kong ^{1,2} , Wenzhong Yang ^{2,*} and Fuyuan Wei ^{1,2}¹ School of Software, Xinjiang University, Urumqi 830049, China² Xinjiang Key Laboratory of Multilingual Information Technology, Xinjiang University, Urumqi 830046, China

* Correspondence: ywz_xy@163.com

Abstract: Implicit sentiment analysis (ISA) can be distinguished from traditional text sentiment analysis by the fact that it does not rely on emotional words as emotional clues, and the expression is usually vaguer. Identifying implicit emotions is more difficult, as it requires a deeper understanding of the context, even when emotional words are absent. Researchers have focused on context feature modeling and developing sophisticated feature extraction mechanisms instead of starting from the emotional perspective. Enhancing the difference in emotional features of text samples is an intuitive method to address this challenge. We proposed a supervised contrastive learning (SCL) method during training that enables the model to conduct contrastive learning based on emotion labels even while training on weak emotion features. SCL training can strengthen the average embedding distance between texts with different emotion labels and enhance implicit emotion discrimination. Moreover, research indicates that contextual information can improve implicit emotion classification ability. Therefore, we applied a straightforward context feature fusion method (bi-affine) over a more complicated context feature modeling approach. To evaluate the effectiveness of our proposed method, we conducted experiments on the SMP2019-ECISA (Chinese implicit sentiment analysis) dataset. The results show a 2.13% enhancement in the F1 value compared to the BERT baseline, proving the effectiveness of our methods.

Keywords: implicit sentiment; sentiment analysis; supervised contrastive learning



Citation: Kong, L.; Yang, W.; Wei, F. Implicit Emotion Analysis Based on Improved Supervised Contrastive Learning. *Electronics* **2023**, *12*, 2172. <https://doi.org/10.3390/electronics12102172>

Academic Editor: Arkaitz Zubiaga

Received: 12 March 2023

Revised: 26 April 2023

Accepted: 6 May 2023

Published: 10 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Implicit emotion expression is defined as “language fragments (sentences, clauses, or phrases) that express subjective emotions but do not contain explicit emotion words” [1]. People tend to use subtle expressions, such as irony, metaphors, and humor, to convey their sentiments without using explicit emotion words, e.g., “When we arrived at the North Pole Village, it was already midnight. Although I was fatigued, I could not help but look up at the sky and keeps looking...”. The objective description is used without specific emotional words to express the speaker’s positive feelings of being attracted by the scenery of the North Pole Village, even forgetting the fatigue of the journey.

According to statistics in the Chinese sentiment dataset, about 15% to 20% of sentences utilize objective or rhetorical statements to convey emotional information implicitly [1,2]. For English sentiment analysis datasets, such as SemEval-2014 Task 4 [3] for Aspect-Based Sentiment Analysis (ABSA) in restaurant and laptop reviews, implicit emotions account for approximately 27.47% and 30.09%, respectively [4]. These findings suggest that implicit emotional expression is widely present across different cultures. Recent research has mainly focused on the explicit sentiment analysis field and has yielded many practical applications [5,6]. However, research on implicit emotions is just beginning. Further research on implicit emotions will continue to promote the development of sentiment analysis tasks.

Since implicit sentiment does not include emotional words as sentiment features, traditional methods for sentiment analysis relying on sentiment dictionary-based regularization techniques or statistical machine learning based on sentiment features are not

applicable. Recent research on implicit sentiment has mainly focused on deep learning methods. Researchers have proposed two approaches to address the problem of missing emotional words as sentiment features. The first approach involves designing sophisticated feature extraction mechanisms, such as orthogonal attention and multi-pooling operations, that capture more semantic features, even when emotion words are absent [7,8]. The second approach involves recognizing implicit emotions with contextual assistance. For instance, researchers have encoded contextual information into heterogeneous graphs consisting of words and sentences using graph neural network models to incorporate contextual information features [9,10]. However, these studies have rarely dealt with implicit sentiment problem from an emotional perspective. Augmenting differences in emotional features in text samples is an intuitive method to address this problem.

To address the problem mentioned above, we propose an implicit sentiment analysis method based on supervised contrastive learning (SCL) [4]. Contrastive learning imitates the learning strategy of human beings. When we classify a sample, we tend to compare it with similar samples to find commonalities and other dissimilar classes to find their differences, which correlates with the process of human generalization.

Supervised contrastive learning aligns the representation of feature embeddings based on a sample's emotion label. Features with the same label are grouped together, and those with different label samples are separated in vector space. Accordingly, we add orthogonal constraints in the supervised contrastive learning process, which can increase the distance of different sample clusters. This approach can reduce the fuzziness of feature embeddings in the vector space, particularly for tasks such as implicit sentiment analysis, for which label-based discrimination is not so clear.

In addition, research shows that contextual features can improve the performance of implicit emotion analysis. We use a simple feature fusion method called the bi-affine [11,12] module to leverage the context features. Compared to graph-based methods [13,14], this method can reduce the workload of complicated context feature modeling.

The main contributions of this paper are summarized as follows:

- We propose an improved supervised contrastive learning method. During training, samples with the same emotional label are aggregated, and samples with different emotional labels are separated with orthogonal constraints;
- Considering the role of context in implicit sentiment analysis, we propose a bi-affine feature fusion module that does not rely on complicated syntactic structures or context feature modeling;
- The experimental results show that the proposed model can achieve significant improvements over state-of-the-art baseline models.

2. Related Work

Existing research has mainly focused on explicit sentiments, and only limited work has been conducted on implicit sentiment analysis. In this section, we analyze relevant research on implicit sentiment analysis, then introduce the supervised contrastive learning method and its application in natural language processing, especially in text sentiment analysis. Eventually, this paper's research motivation is expounded based on previous research.

2.1. Implicit Emotion Analysis

Implicit sentiment analysis started much later than traditional text sentiment analysis tasks. Early researchers observed that certain emotions could only be deduced within the context of comments [15]. Subsequently, researchers attempted to find implicit emotions in events. Deng and Wiebe [16] analyzed implicit sentiments via the detection of explicit sentiment expression clues and the inference of events. Choi et al. [17] used +/- Effect-WordNet lexicon to identify event-related implicit emotions by assuming that emotional expression is usually associated with states and events that have positive/negative/ineffective effects on entities. In recent years, independent emotion analysis tasks have been

studied. Liao et al. [1] constructed a small-scale factual implicit emotional corpus and proposed a multilevel semantic fusion model; they studied the identification of factual implicit emotions in sentences and proposed a multi-level semantic fusion method that integrates emotional target representation, structural embedding representation, and semantic contextual representation. Researchers now use deep learning methods more to solve this problem. Wei et al. [8] proposed a BiLSTM model with multipolar orthogonal attention for implicit emotion analysis. Unlike traditional single-attention models, multipolar attention can identify differences between words and emotional orientations. Zuo et al. [9] obtained features of implicit emotional sentences and contexts through GCN and proposed a context-specific heterogeneous graph convolutional neural network (CsHGCN) to address the lack of emotional words. Zhou et al. [18] inferred the emotional polarity of sentences by utilizing emotion perception events. They represented events as a combination of their event type and event triplet. Based on this representation, they proposed a model with a hierarchical tensor composition mechanism to detect emotions in text.

2.2. Supervised Contrastive Learning

Contrastive learning is mainly applied in computer vision; according to the choice of contrast objectives, such methods can be divided into self-supervised and supervised learning methods. The main task of contrastive learning is to learn similar/dissimilar representations from a dataset composed of pairs of similar/dissimilar data.

Self-supervised contrastive learning is a variant of contrastive learning that makes use of both labeled and unlabeled data to improve performance on tasks. One of the key advantages of self-supervised contrastive learning is that it can be used to overcome the challenge of limited labeled data. This is particularly relevant for tasks for which labeled data are scarce or expensive to obtain. By leveraging unlabeled data to augment the labeled data, self-supervised contrastive learning can improve model performance without requiring a large amount of labeled data. Inspired by the performance achieved by discriminative approaches based on latent space comparison learning, Chen et al. [19] proposed a simple visual representation contrastive learning framework (SimCLR). The performance of SimCLR on 10 out of 12 datasets is comparable to or better than the supervised baseline. In the same year, the semisupervised learning framework SimCLRv2 [20] greatly improved the result of self-supervised contrastive learning (SimCLR). SimCLRv2 only uses 10% of the labeled data to achieve a better effect than supervised learning.

The supervised contrastive learning approach involves training a model to differentiate between similar and dissimilar examples in a supervised manner through the use of a contrastive loss function. The advantage of supervised contrastive learning is that it can improve the ability of a model to discriminate between different classes or categories of data. Khosla et al. [21] proposed a supervised comparison method with two forms. The self-supervised comparison method is extended to supervised learning to create orthogonal pairs between instances with the same class labels and combine their representations. SOTA results were obtained with ResNet-50 and the ResNet-200 backbone on ImageNet. This method led to better classification and prediction performance in NLP tasks such as sentiment analysis, document classification, and machine translation. Gunel et al. [4] used supervised contrastive learning loss instead of cross-entropy loss to fine-tune pretrained language models. The fine-tuned language models achieved significant improvements in evaluation metrics such as GLUE with few shots. One specific implementation of this is expressed by the following formula:

$$\mathcal{L}_{SCL} = \sum_{i=1}^N -\frac{1}{N_{y_i}-1} \sum_{j=1}^N \mathbf{1}_{i \neq j} \mathbf{1}_{y_i=y_j} \log \frac{\exp(\Phi(x_i) \cdot \Phi(x_j) / \tau)}{\sum_{k=1}^N \mathbf{1}_{i \neq k} \exp(\Phi(x_i) \cdot \Phi(x_k) / \tau)}$$

where N is the batch size of training examples, the sample denoted as $\{x_i, y_i\}_{i=1, \dots, N}$, N_{y_i} is the total number of examples in the batch that have the same label as y_i , $\Phi(x_i) \cdot \Phi(x_j)$

represents a comparison of the similarity of hidden states of samples x_i and x_j , and τ is the temperature hyperparameter.

Li et al. [22] found that previous ABSA research has rarely focused on implicit sentiment. Using supervised contrastive pretraining to learn implicit sentiment knowledge models from large-scale sentiment-annotated corpora, they achieved state-of-the-art performance and were able to effectively learn implicit sentiment.

The abovementioned studies demonstrate that implicit emotions can be better represented by supervised contrastive learning, thereby improving the accuracy of implicit emotion analysis [22]. However, the cost of retraining on a large-scale sentiment corpus is relatively high, although using cross entropy and fine-tuning pretrained language models can also achieve good results [4]. It has also been found that the bi-affine module [12] can also play a role in feature fusion. Therefore, we used supervised contrastive learning to fine-tune the pretrained model and bi-affine for contextual feature fusion in a study of Chinese implicit emotion analysis.

3. Methodology

Implicit sentiment analysis tasks are similar to explicit sentiment tasks in predicting the implicit emotional polarity of a given target sentence as positive, negative, or neutral. The structure of our model is shown in Figure 1, and the general processing flow is described as follows.

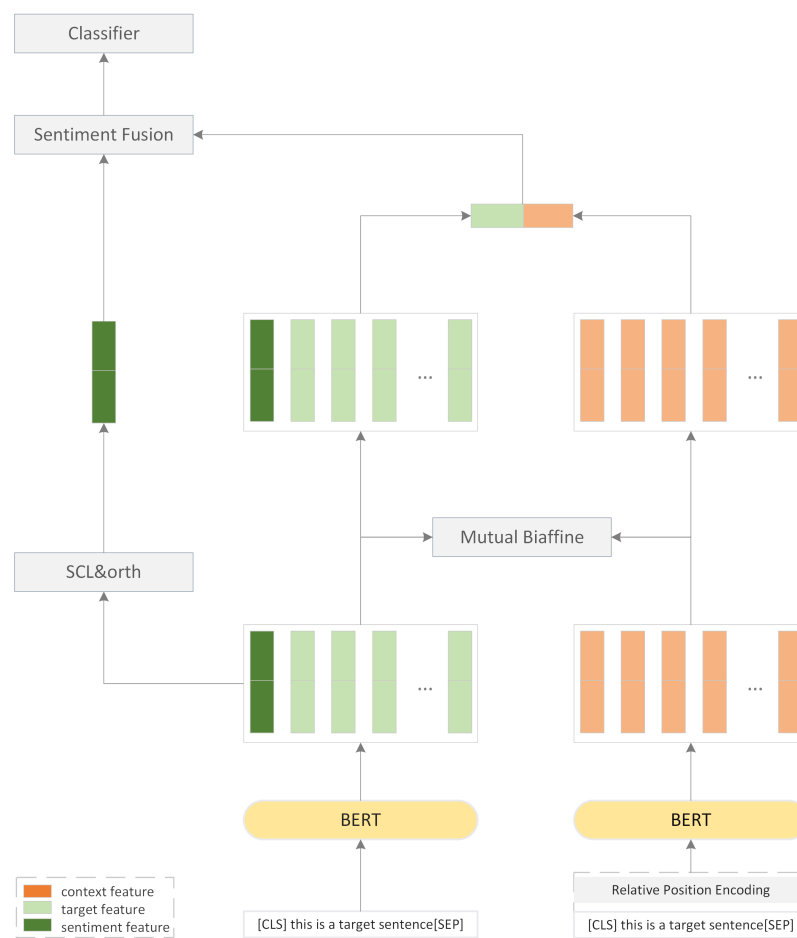


Figure 1. The architecture of our model.

1. Input the target sentence and context sentence into BERT to obtain the corresponding text hidden state;

2. Supervised contrastive learning can cluster emotional features based on emotion labels to obtain good emotional representations, while adding orthogonal constraints to different class features enhances feature discrimination;
3. The Bi-Affine module is used to fuse the target feature and context features. The aggregated feature of the entire context is obtained through pooling and summation operations;
4. The aggregated feature obtained in the previous step is concatenated with the emotion feature learned by supervised contrastive learning, forming the final feature representation;
5. Finally, a fully connected layer and softmax layer are used to calculate the probability of target emotion classification.

3.1. Model Inputs

The initial word node features of the model come from the BERT encoder, and the position information comes from positional embedding. Batch B includes n sentences $\{s_1, s_2, \dots, s_n\}$. The label of the i -th sentence (s_i) is denoted as y_i . Each sentence (s_i) includes its implicit sentiment sentence (s_i^t) and its context sentence (s_i^c). To control the length of context, we limit it to no more than three sentences before and after the target sentence (s_i^c).

Context-Relative Position Coding (CPOS): Location information helps to locate context-related target sentences, especially when there are multiple context sentences between a target sentence. Therefore, we explicitly encode the relative position of the context relative to the target sentence.

BERT Encoder: We used the pretrained BERT baseline model as the encoder to obtain the word representation. Specifically, we constructed the input as "[CLS]+sentence+[SEP]" and input the context and target sentence into BERT. For the context sentence, we added contextual relative position encoding (CPOS) before adding it to BERT. The hidden state representation (h_i^t) of the target sentence (s_i^t) and the hidden state representation (h_i^c) of the context (s_i^c) are obtained.

3.2. Improved SCL

Supervised Contrastive Learning: SCL encourages models to learn label-based feature representations. Here, we used supervised contrastive learning to encourage the sample to learn the implicit emotion sample representation of the same labels to improve familiarity.

First, we extracted the emotional representation ($s_i = W_s \bar{h}_i$) from the sentence representation. W_s can be seen as a trainable sentence emotion perceptron. $\bar{h}_i$ is the hidden representation of the target sentence s_c . Specifically, for (s_i, y_i) in batch B , we define the supervised contrastive loss as follows:

$$L_B^{sup} = \sum_{i \in B} -\frac{1}{N_{y_i} - 1} \sum_{y_i = y_j, i \neq j} \log \frac{\exp(s_i \cdot s_j / \tau)}{\sum_{b \in B, b \neq i} \exp(s_i \cdot s_b / \tau)}$$

where N_{y_i} represents the number of samples with label y_i in the same batch (B). We use $s_i \cdot s_j$ as a measure of similarity. The temperature (τ) is used to control the aggregation rate of sample points.

Add orthogonal constraints: In order to separate the emotional characteristics of different emotions, an orthogonal constraint is added to the emotional feature vectors of different labels in the same batch (B). We developed the following process:

$$L_B^{orth} = \sum \left| \frac{s_i \cdot s_j}{|s_i| \cdot |s_j|} \right| \quad (i, j \in [0, 1, 2], i \neq j)$$

3.3. Bi-affine Attention

We used the bi-affine attention module to effectively fuse the features of the target sentence and the context sentence; the specific process is expressed as follows:

$$h_i^{c'} = \text{softmax}\left(h_i^c W_c (h_i^t)^T\right) h_i^t$$

$$h_i^{t'} = \text{softmax}\left(h_i^t W_t (h_i^c)^T\right) h_i^c$$

where W_c and W_t are trainable weights.

After pooling the target sentence and contextual features transformed by bi-affine, they are combined and normalized to obtain the aggregated feature (r_i) for the entire context.

$$r_i = \text{Norm}(\text{pooling}(h_i^{c'}) + \text{pooling}(h_i^{t'}))$$

The pooling method uses average pooling to reduce the dimensions of features, and the norm is the standardization layer.

3.4. Emotional Feature Fusion

Finally, we concatenated the context-aggregated features obtained from the bi-affine module with the emotion features learned through supervised contrastive learning to obtain a final feature representation for the implicit emotion analysis task.

$$f = \tanh(W_c[s_i || r_i] + b_c)$$

where W_c and b_c are learnable parameters.

3.5. Sentiment Classification

Then, the obtained feature (f) is sent to a linear layer, followed by the softmax function, to generate an emotion probability distribution (p). L_c is the standard cross-entropy loss used to evaluate the accuracy of predictions.

$$p(a) = \text{softmax}(W_p f + b_p)$$

$$L_c = - \sum_{c \in C} \log p(a)$$

where W_p and b_p are trainable weights and biases, respectively, and $C = 0, 1, 2$ is the set of different emotional polarity samples in the same batch (B).

3.6. Loss Function

Our training goal is to minimize the following objective function:

$$L_t = L_c + \lambda_1 L_B^{\text{sup}} + \lambda_2 L_B^{\text{orth}}$$

where λ_1 and λ_2 are regularization coefficients, L_c is the standard cross-entropy loss for sentiment polarity classification, L_B^{sup} represents supervised contrastive loss, and L_B^{orth} is the added orthogonal constraint.

4. Experiments

4.1. Dataset

We conducted experiments using the Chinese implicit sentiment dataset SMP-ECISA 2019 https://conference.cipsc.org.cn/smp2019/smp_ecisa_SMP.html (accessed on 11 March 2023) (the Eighth National Conference on Social Media Processing of the Chinese implicit sentiment assessment). The dataset primarily comprises Weibo posts (Chinese social media platform), covering various topics, such as the Spring Festival Gala, tourism, haze, etc. A large-scale

sentiment lexicon was used to filter the dataset, ensuring that no text samples contained explicit sentiment words. The data are published in the form of chapters of cut sentences, which retains the complete contextual content information. Such data samples are categorized into three types: positive, negative, and neutral. The specific statistics are shown in Table 1. The data are published in XML format. An instance of a text sample is presented below.

```
<Doc ID="5">
<SentenceID="1">because you are old lady</Sentence>
<SentenceID="2" label="1">Finished watching, full of memories many
elements of that era</Sentence>
</Doc>
```

Table 1. Statistics of the SMP_ECISA2019 dataset.

Dataset	Positive	Negative	Neutral	Total
Train	3828	3957	6989	14,774
Validation	1232	1358	2553	5143
Test	919	979	1902	3800

4.2. Implementation Details

We implemented our model with the pretrained bert-base-uncased <https://huggingface.co/bert-base-uncased> (accessed on 11 March 2023) in pytorch transformers <https://huggingface.co/docs/transformers/> (accessed on 11 March 2023). The supervised contrastive loss uses the online version <https://github.com/HobbitLong/SupContrast> (accessed on 11 March 2023). The bug of only one positive or negative class was fixed in the batch, resulting in no other samples for comparison error. Then, we added a data normalization operation to maintain numerical stability.

The Adam optimizer uses a learning rate of 1×10^{-5} , with a batch size of 16 and dropout set to 0.1 during training. To identify the optimal hyperparameters, we conducted a grid search on λ and supervised contrast learning temperature (τ), with possible values of $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. In all experimental setups, models that achieved the best test accuracy used $\tau = 0.3$, $\lambda_1 = 0.3$, and $\lambda_2 = 0.5$.

4.3. Baseline

We selected some models that performed well in implicit emotion analysis tasks as the baseline. The description and specific details are as follows.

MPOA [8]: This model is based on the multihead attention model of BILSTM. Adopting orthogonal and score attention constraints encourages multiple-head attention dispersed to different parts of the text.

MPOAC [7]: This model proposes a multipolarity orthogonal attention mechanism to embed implicit sentiment sentences, establishing a multipolarity attention layer that integrates contextual information in model the context.

HAN [23]: This method adopts a hierarchical attention mechanism at the word and sentence levels so that the model can provide extra attention to the ability of sentences and words of different importance in the text. The method is divided into two parallel models, and the target sentence is spliced into the final representation through two layers of attention mechanisms.

CsHGCN [9]: This model proposes a context-specific heterogeneous graph convolutional network framework that can combine all context representations. It has a full context that comprehensively reflects the information of documents.

BERT [24]: This is the vanilla BERT model that uses the representation of CLS%for predictions.

BERT-SPC: This models composes the input sequence format as “[CLS] context sentence [SEP] target sentence” to the BERT model; the length of the context is limited to three sentences, using the [CLS] token for prediction.

4.4. Results and Analysis

We employed macro-P, macro-R, and macro-F as the performance metrics to assess our implicit sentiment analysis model. The main experimental results are shown in Table 2. Employing an attention module for LSTM(MPOA, MPOAC) leads to a 2–3% increase in F1 score compared to the vanilla LSTM model. For the method based on modeling a heterogeneous structure, the graph neural-network-based approach (CsHGCN) yields a substantial increase of 5.39% in the F1 score compared to the Bi-GRU method (HAN). This suggests that graph neural networks have better feature aggregation capabilities for heterogeneous data structures. BERT-based methods (BERT and BERT-SPC) outperform the other two methods based on LSTM attention (MPOA and MPOAC) and heterogeneous structure modeling methods (HAN and CSHGCN). In contrast to BERT, BERT-SPC shows a decline of 0.13% in the F1 score, implying that feeding context directly into the BERT model does not enhance its ability to detect implicit emotions. One plausible reason for this is that implicit emotions are highly context-sensitive and can be susceptible to more noise. Our proposed approach increases the f1 score from 80.47% to 82.6% compared to the vanilla BERT. This suggests that our proposed method, which combines supervised contrastive learning and bi-affine contextual feature fusion, can enhance the discriminative power of implicit emotions.

Table 2. Experimental result of the comparison model.

Model	macro-P (%)	macro-R (%)	macro-F1 (%)
LSTM	68.93	70.92	69.91
MPOA	-	-	73.3
MPOAC	-	-	74.6
HAN	72.69	71.91	72.3
CsHGCN	79.46	76	77.69
BERT	79.69	81.27	80.47
BERT-SPC	79.49	80.81	80.14
Proposed model	82.36	82.84	82.6

4.5. Ablation Study

To further explore the role of each module in the our model, a series of ablation studies was conducted. The experimental results are shown in Table 3.

Table 3. Experimental result of the ablation study.

Model	macro-P (%)	macro-R (%)	macro-F1 (%)
ours	82.36	82.84	82.6
w/o scl&orth	80.03	81.22	80.62
w/o orth	82.11	81.76	81.93
w/o biaffine	81.75	82.23	81.99

w/o scl&orth indicates that we removed both supervised contrastive learning and orthogonal restrictions, resulting in a 1.98% decrease in the F1 value. This is much more than the 0.61% decrease observed when removing the bi-affine module (w/o biaffine), indicating that supervised contrastive learning plays a more important role in our model.

When only removing orthogonal restrictions (w/o orth), we can see that the F1 value drops by 0.67%, which is comparable to the results of our double-context fusion model (w/o biaffine). This demonstrates the effectiveness of adding orthogonal restrictions.

4.6. The Impact of Contrast Feature

To examine the impact of contrast features in supervised contrast learning on experimental results, we employed three different pooling strategies to select contrast features—[MEAN], [MAX], and [CLS]—corresponding to mean pooling, max pooling, and using the [CLS] token as the whole implicit sentiment representation, respectively. As shown in Table 4, the [MEAN] and [CLS] pooling methods outperform [MAX] pooling, as they improve the final F1 score by 0.81% and 0.93%, respectively. This result suggests that using max-pooling as a comparative feature may result in losing more sample features, leading to poor performance. The difference in performance between [CLS] and [MEAN] is insignificant. The [CLS] method achieves a higher F1 score by 0.12%, implying that satisfactory results can be attained solely by using [CLS] without any extra pooling operations.

Table 4. The impact of contrast features.

Label	macro-P (%)	macro-R (%)	macro-F1 (%)
[Mean]	82.52	81.95	82.23
[MAX]	82.17	81.17	81.67
[CLS]	82.36	82.84	82.6

4.7. Difference of Label Distribution

To explore the performance of the model under different labels, we conducted the corresponding experiments. The results are shown in Table 5. The neutral tag sample performs the best in terms of accuracy, outperforming the positive and negative tags by 6.66% and 7.71% in the F1 score, respectively. This is due to the fact that the neutral samples constitute more than twice the number of positive or negative samples and therefore receive sufficient training. The F1 value of negative samples is 0.82% lower than that of positive samples, and its recognition accuracy is also lower. This may be because in implicit emotions, the use of metaphors and irony make it more difficult to distinguish negative emotional polarity.

Table 5. Results of different label distributions.

Label	Precision (%)	Recall (%)	F1 (%)
Neutral	85.46	87.02	86.23
Positive	79.39	79.76	79.57
Negative	78.11	79.32	78.71

4.8. Comparison of Runtime to the BERT Baseline

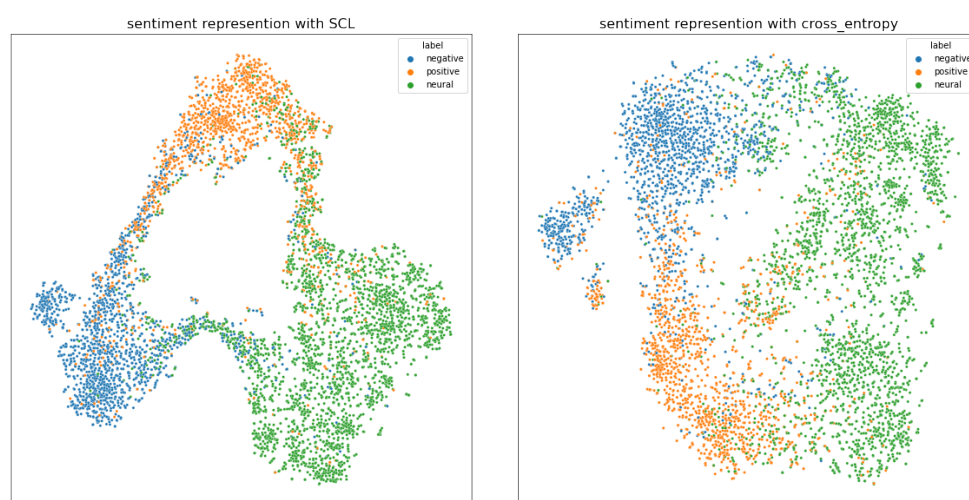
To ensure accurate comparisons, we maintained a consistent hardware and software environment throughout the experiments and utilized the same contextual conditions for the original input. Table 6 presents a comparison between our model and the baseline BERT in terms of the average runtime over 10 epochs on the SMP2019 dataset. Our supervised contrastive learning method did not involve additional parameters, resulting in our model having a similar number of parameters to BERT (around 100 million). During training, BERT takes an average of 6.81 minutes to complete one batch, while our model takes 12.85 minutes to complete one epoch. This difference is due to the fact that we input BERT in two segments to obtain the features of the context and target sentence separately, with the main time spent on BERT. In the prediction stage, the difference between our model and BERT is only 1.1 minutes because we do not need to calculate the supervised contrastive loss during the prediction phase. Therefore, the time difference is relatively small.

Table 6. The average runtime per epoch.

Model	Parameters	Train (min)	Predict (min)
Bert	102,269,955	6.81	3.1
Ours	104,630,786	12.85	4.2

4.9. Visualization of Hidden Representations

In order to understand the impact of supervised contrastive learning on implicit sentiment analysis, we used t-SNE [25] to perform dimensionality reduction and visualize the hidden state of the sentiment representation. As shown in Figure 2, the hidden state corresponding to the [CLS] token trained by SCL can learn feature representations by sentiment label more closely.

**Figure 2.** Visualization of the hidden sentiment representation.

5. Conclusions

In this paper, we investigated the implicit sentiment analysis task based on the SCL method. We argue that the relatively low accuracy of implicit emotion classification is due to weak emotional features caused by missing emotional words. Prior research has not fully utilized the available sentiment information. Hence, we employed supervised contrastive learning to minimize the embedding distance between sample features based on labels to help the model find a better sentiment representation. Our experiments demonstrate that our model outperforms the baseline, indicating the effectiveness of our approach. However, the improvement of the context fusion module on experimental results is limited. Our future research will focus on exploring more effective methods utilizing context.

In the future, our aim is to conduct more fine-grained research on implicit sentiment analysis. Employing prompt learning enables the alignment of pretrained tasks and downstream tasks. By setting templates, we can identify absent sentiment words in implicit sentiment sentences and align them with explicit emotional analysis tasks.

Author Contributions: Data curation, L.K.; funding acquisition, W.Y.; investigation, L.K. and F.W.; methodology, L.K.; project administration, L.K.; software, L.K.; supervision, L.K. and F.W.; validation, L.K. and F.W.; writing—original draft, L.K.; writing—review and editing, L.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of China grant number 202204120017, the Autonomous Region Science and Technology Program grant number 2022B01008-2, and Autonomous Region Science and Technology Program grant number 2020A02001-1.

Data Availability Statement: The data used in this study are public datasets, which can be obtained from the following link: https://conference.cipsc.org.cn/smp2019/smp_ecisa_SMP.html. (accessed on 11 March 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liao, J.; Wang, S.; Li, D. Identification of Fact-Implied Implicit Sentiment Based on Multi-Level Semantic Fused Representation. *Knowl.-Based Syst.* **2019**, *165*, 197–207. [CrossRef]
2. Jian, L.; Yang, L.; Suge, W. The Constitution of a Fine-Grained Opinion Annotated Corpus on Weibo. In *Proceedings of the Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*; Sun, M., Huang, X., Lin, H., Liu, Z., Liu, Y., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; pp. 227–240. [CrossRef]
3. Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, 23–24 August 2014; Association for Computational Linguistics: Dublin, Ireland, 2014; pp. 27–35. [CrossRef]
4. Gunel, B.; Du, J.; Conneau, A.; Stoyanov, V. Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning. *arXiv* **2021**, arXiv:2011.01403.
5. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
6. Liu, B. Sentiment Analysis and Opinion Mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*, 1–167.
7. Wang, H.; Hou, M.; Li, F.; Zhang, Y. Chinese Implicit Sentiment Analysis Based on Hierarchical Knowledge Enhancement and Multi-Pooling. *IEEE Access* **2020**, *8*, 126051–126065. [CrossRef]
8. Wei, J.; Liao, J.; Yang, Z.; Wang, S.; Zhao, Q. BiLSTM with Multi-Polarity Orthogonal Attention for Implicit Sentiment Analysis. *Neurocomputing* **2020**, *383*, 165–173. [CrossRef]
9. Zuo, E.; Zhao, H.; Chen, B.; Chen, Q. Context-Specific Heterogeneous Graph Convolutional Network for Implicit Sentiment Analysis. *IEEE Access* **2020**, *8*, 37967–37975. [CrossRef]
10. Yang, S.; Xing, L.; Li, Y.; Chang, Z. Implicit Sentiment Analysis Based on Graph Attention Neural Network. *Eng. Rep.* **2022**, *4*, e12452. [CrossRef]
11. Li, R.; Chen, H.; Feng, F.; Ma, Z.; Wang, X.; Hovy, E. Dual Graph Convolutional Networks for Aspect-based Sentiment Analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Virtual Event, 1–6 August 2021; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 6319–6329. [CrossRef]
12. Xu, M.; Wang, D.; Feng, S.; Yang, Z.; Zhang, Y. KC-ISA: An Implicit Sentiment Analysis Model Combining Knowledge Enhancement and Context Features. In *Proceedings of the 29th International Conference on Computational Linguistics*, Gyeongju, Republic of Korea, 12–17 October 2022.
13. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2017**, arXiv:1609.02907.
14. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, arXiv:1710.10903.
15. Toprak, C.; Jakob, N.; Gurevych, I. Sentence and Expression Level Annotation of Opinions in User-Generated Discourse. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, 13 July 2010; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 575–584.
16. Deng, L.; Wiebe, J. Sentiment Propagation via Implicature Constraints. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg, Sweden, 26–30 April 2014; pp. 377–385. [CrossRef]
17. Choi, Y.; Wiebe, J. +/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1181–1191. [CrossRef]
18. Zhou, D.; Wang, J.; Zhang, L.; He, Y. Implicit Sentiment Analysis with Event-centered Text Representation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic, 7–11 November 2021; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 6884–6893. [CrossRef]
19. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G.E. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the International Conference on Machine Learning*, Baltimore, MD, USA, 17–23 July 2022.
20. Chen, T.; Kornblith, S.; Swersky, K.; Norouzi, M.; Hinton, G. Big Self-Supervised Models Are Strong Semi-Supervised Learners. *arXiv* **2006**, arXiv:2006.10029.
21. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised Contrastive Learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 18661–18673.
22. Li, Z.; Zou, Y.; Zhang, C.; Zhang, Q.; Wei, Z. Learning Implicit Sentiment in Aspect-based Sentiment Analysis with Supervised Contrastive Pre-Training. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic, 7–11 November 2021; pp. 246–256. [CrossRef]

23. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical Attention Networks for Document Classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 7 June 2016; pp. 1480–1489. [[CrossRef](#)]
24. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805.
25. Van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.