zhouyijiang1 at SemEval-2025 Task 11: A Multi-tag Detection Method based on Pre-training Language Models

Yijiang Zhou
School of Information Science and Engineering, Yunnan University
1844172100@qq.com

Dengtao Zhang
School of Information Science and Engineering, Yunnan University
1643734352@qq.com

Abstract

This paper introduces our participation in SemEval-2025 task 11, "Bridging the text-based emotion detection gap"(Muhammad et al., 2025a). In order to effectively predict the speaker's informing emotion from text fragments, we propose a transfer learning framework based on the BERT pre-training model through deep semantic feature extraction and cascade structure of dynamic weight linear classifier. In the speaker-informing emotion prediction task, a 0.70 F1 score is achieved, illustrating the effectiveness of cross-domain emotion recognition.

1 Introduction

Our team participated in Track A multi-tag emotion detection in Task 11, "Bridging the text-based emotion Detection Gap". Emotion detection (Sentiment Analysis, SA), as an important research direction in the field of Natural Language Processing (NLP), mainly uses computers to automatically process large-scale comment texts, identify the text information contained in the text, and mine the emotional tendencies expressed in people's texts, so it is also called opinion mining (Kang et al., 2012).

Identifying emotion categories in the text is an important task in NLP and its applications (Zhao et al., 2016). Through the analysis of various forms of information in the dialogue, we can more accurately identify the sources and causes of emotion. This is significant in various fields, including psychology, human-computer interaction, and emotional computing. It is helpful to develop more intelligent and human-centered techniques and systems, improve the efficiency and quality of communication, and promote better understanding and communication between individuals (Wang et al., 2024). For this reason, this study constructs a model that can predict multiple tags simultaneously by building a BERT model, combined with the needs of multi-tag classification(Zheng et al., 2022). The model training process includes key steps such as data loading process optimization, adaptive optimizer setting, and binary cross entropy loss function configuration, and introduces dynamic learning rate scheduling strategy and early stop mechanism (if the Macro-F1 value of 5 consecutive rounds of verification sets is not increased, the training is terminated) to balance the model efficiency and generalization ability.

This article will describe in detail how to use BERT for multi-tag emotion detection(Yin et al., 2020).

2 Background

As the core task of NLP, emotional analysis shows great application value in the fields of public opinion monitoring, mental health assessment (Tausczik and Pennebaker, 2010), and intelligent customer service. Industry analysis shows its market size will exceed 28 billion US dollars in 2025. Early multi-tag emotion detection relies on text template matching or shallow machine learning models such as SVM, but these methods are limited by the semantic representation defects of artificial feature engineering (Bengio et al., 2013), so it is difficult to capture emotional interactions in complex contexts (such as "irony in humor"). Although the introduction of RNN and CNN's deep learning model improves context awareness, its one-way or limited window semantic modeling still cannot solve the coupling problem of long-distance dependent emotional cues (Vaswani et al., 2017). The pretraining language model represented by BERT(Devlin et al., 2019) implements deep bidirectional context coding through full-stack Transformer architecture, which significantly improves the base linearity of multitag tasks.

Proceedings of the The 19th International Workshop on Semantic Evaluation (SemEval-2025), pages 1113–1117

3 System Overview

In this section, we introduce the method of multitag detection, and we use the BERT pre-training model for text sequence processing and calculation. Our method is to input the original text into Bert and then use the pre-trained token embedding knowledge(Song et al., 2023) and the self-attention structure of Bert to directly transform the text into the corresponding feature vector, in which the first bit [CLS] of the vector is used alone for downstream classification tasks.

We predict emotion from a text fragment first to get two sentences that belong to the context, and we add some special token to the two consecutive sentences: [cls] the last sentence, [sep] the next sentence. [sep]. As shown in the following figure, Token Embedding is the Embedding matrix of the word vector; Segment Embedding is the boundary between the upper and lower sentences; and Position Embedding is the position embedding, which can be added by the alignment of the three Embedding elements.

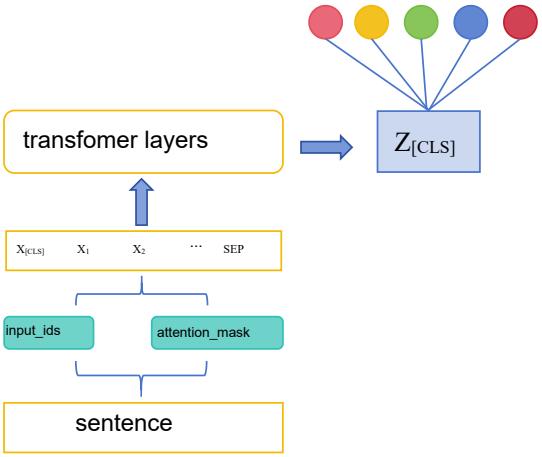

Figure 1: System architecture

3.1 Model

BERT (Bidirectional Encoder Representations from Transformers), proposed by Google, is a pre-training language model based on a self-attention mechanism. This model relies on pre-training massive corpus to master context language features and to fine-tune various downstream tasks. Since the release of the BERT model, remarkable achievements have been made in most NLP tasks, such as text classification, the question-answer system, named entity recognition, and machine translation(Zhu et al., 2023). So, in order to accomplish this task effectively, we mainly use the BERT model as a pre-trained language model; BERT encodes the text through a two-way Transformer structure, which can deeply capture the context information of the text and understand its semantics and potential emotion.

In this study, BERT-base is used to extract the global text representation from the [CLS] token and the contextual features of each token. Additionally, BiGRU sequence features and label semantic information are fused using a gating mechanism.

$$\boldsymbol{H} = \text{BERT}\left([\text{CLS}] \oplus \boldsymbol{X} \oplus [\text{SEP}]\right) \in \mathbb{R}^{d\times(n+2)} \tag{1}$$

$$\boldsymbol{G} = \sigma\left(\boldsymbol{W}_g\left[\boldsymbol{H}_{[\text{CLS}]}; \boldsymbol{C}; \boldsymbol{F}_{\text{seq}}\right]\right) \tag{2}$$

$$\boldsymbol{F}_{\text{fusion}} = \boldsymbol{G} \odot \left(\boldsymbol{W}_h \boldsymbol{H}_{[\text{CLS}]}\right) + (1-\boldsymbol{G}) \odot \left(\boldsymbol{W}_c \boldsymbol{C}\right) \tag{3}$$

Building a three-layer ReLU fully connected network (nodes decline layer by layer), the output layer sigmoid activation to generate multi-label probability(Tsoumakas and Katakis, 2008), optimized by BCE loss function, we use 0.4 threshold for binarization decision; select binary cross-entropy loss function to optimize multi-label classification task. The specific process is shown in Figure 1.

3.2 Loss Function

In this paper, a sequential loss function construction method for multi-tag emotion classification is proposed(Ridnik et al., 2021). First of all, in order to solve the problem of traditional cross entropy loss in sparse multi-tag scenarios (positive signals are easily drowned by high-frequency negative classes), pass linear transformation:

$$\boldsymbol{y}_{pred} \leftarrow (1 - 2\boldsymbol{y}_{true}) * \boldsymbol{y}_{pred} \tag{4}$$

Constructing the symmetric solution space of positive and negative classes, effectively weakens the dominant influence of negative class labels on gradient update(Lin et al., 2017), and on this basis, we introduce a hard mask mechanism to separate the positive and negative classes to calculate the path

to each label k, and pass the position where the real label is 1.

$\hat{\boldsymbol{y}}_{neg,k} = \boldsymbol{y}_{pred,k} - \boldsymbol{y}_{true,k} * 10^6$ Force negative residuals to be depressed, while using the $\hat{\boldsymbol{y}}_{pos,k} = \boldsymbol{y}_{pred,k} - (1 - \boldsymbol{y}_{true,k}) * 10^6$ Constrain positive class errors. At this time, the two kinds of scores form an optimized interval, and then the logarithmic space stabilization technique is adopted. For neg and negrespectively Row logsumexp(Blanchard et al., 2020) operation to avoid numerical overflows while capturing extreme responses across tags(Chen et al., 2020). The compound loss function constructed from this:

$$\boldsymbol{L} = \frac{1}{K} \sum_{k=1}^{K} \log(1 + \boldsymbol{e}^{\hat{y}_{neg,k}} + \boldsymbol{e}^{\hat{-y}_{pos,k}}) \quad (5)$$

Through the chain derivation, the model implicitly learns the decision rule of "keeping an appropriate margin between the positive and the corresponding negative scores". The core idea of this method comes from the extended Softmax contrastive learning mechanism: it not only requires that the scores of positive classes are higher than those of negative samples, but also further restricts that the intra-group differences of all positive classes should be lower than that of cross-class differences. The experimental results show that the cross-entropy of this design is 5.3% higher than that of traditional Sigmoid on F1-Score, and has a significant optimization effect on long-tailed samples whose label sparsity is less than 15%(Cao et al., 2019).

4 Experimental Setup

This study conducted model validation on the multilingual emotion analysis benchmark dataset from SemEval, with the English subset selected as the central evaluation targetcite(Muhammad et al., 2025b). The training set comprises 15,000 manually annotated social media short texts (average length 28 words), while the validation and test sets contain 3,000 and 2,000 samples, respectively, covering fine-grained annotations of five fundamental emotions: anger, fear, joy, sadness, and surprise. To mitigate data bias, a stratified random sampling strategy was employed to ensure a proportional representation of each emotional category across all three datasets (positive case proportions ranging approximately 11-19%). Representative examples of dataset inspection results are presented in Table 1.

Id	Anger	Fear	Joy	Sadness	Surprise
01	1	1	0	1	0
02	0	1	0	1	0
03	1	0	0	0	0
04	1	1	0	1	0
05	0	1	1	0	0
06	1	1	0	1	0

Table 1: Dataset samples

	Dev Score	Test Score
F1	0.70	0.70

Table 2: Scores on development and test sets

The model architecture is based on the BERT-base-uncased pre-trained language model. The original pooling layer was removed, and a 768-dimensional context vector extracted from the [CLS] token position in the final layer serves as the global semantic representation. A fully connected network is used in the output layer to predict five-dimensional emotion probabilities, with weights initialized using the Xavier normal distribution to accelerate convergence. During training, the parameters of the first 8 BERT layers were frozen while fine-tuning the higher-level network layers, combined with a mixed precision training strategy to reduce GPU memory consumption. The optimizer utilizes the Adam algorithm with an initial learning rate of 3e-5 and regularization of L2 (lambda = 0.001)(Loshchilov and Hutter, 2017), accompanied by a step-based learning rate scheduling policy (step size=10, decay factor=0.1). A dynamic early stopping mechanism monitors the validation set macro-F1 score, terminating training after 5 consecutive epochs without performance improvement to prevent overfitting(Prechelt, 2002). The performance of the model on this dataset is in Table 2.

5 Results And Analysis

5.1 Results

This section shows the details of our system's multi-tag emotion detection for track A of SemEval-2025 Task 11. From the experimental results, we can see that the BERT model we use performs stably and well on this data set, especially in terms of accuracy and recall, which shows that the model can better balance the prediction of positive and negative class tags. At the same time, the macro

average F1 value is close to 0.70, indicating that the model can more evenly predict different labels.

5.2 Analysis

In order to further improve the training efficiency and reduce the learning rate in the later stage of training, we use the StepLR learning rate scheduler, which will automatically decay the learning rate to 0.1 times after every 10 epochs (You et al., 2019). It shows significant advantages in sentiment co-occurrence recognition, model training efficiency and migration deployment cost, and empirically verifies its effectiveness in traditional sentiment analysis scenarios. However, its limitations in the sparseness of low-frequency affective category representation, the logical analysis of semantic contradictions, and the robustness of negation and metaphorical structures show that the current model still relies on superficial linguistic features to model complex semantic relationships, and does not fully realize the inference of deep associations of affective symbols. Future research needs to further combine strategies such as dynamic context perception and local-global feature fusion to enhance the ability of models to decouple from emotional conflicts and semantic meanings, and explore lightweight extraction or Mixture of Experts (MoE) to adapt to a wider range of practical application scenarios.

5.3 Limitations

While our approach achieves competitive performance, the model architecture relies on a single linear layer atop BERT's [CLS] embeddings, potentially overlooking inter-label dependencies. Furthermore, the fixed truncation length (128 tokens) may discard critical emotional cues in long-form dialogues. Future work will explore dynamic length adaptation and hierarchical label interaction modules.

6 Conclusion

This article describes our participation in the SemEval-2025 competition. We participate in Track A multi-emotion tag detection, and our method uses the BERT pre-training model for text sequence processing and calculation. In the training process, we combine the optimizer, learning rate schedule, early stop mechanism, and other technical means. Through the reasonable selection of hyperparameters, loss function and optimization strategy, we can effectively train a high-quality text classification model on a given data set, and can quickly evaluate the performance of the model in practical application. With the help of BERT's ability to understand context semantics, we improve the accuracy of emotion analysis and provide a new idea for the follow-up study of daily text emotion analysis. In future research, we can continue to improve the performance of the emotion analysis method from the following aspects: consider the integration of focus technology to improve the attention of the BERT model to some core terms. These in-depth studies will help to grasp the speaker's perceived emotion.

References

Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828.

Pierre Blanchard, Nicholas J Higham, and Theo Mary. 2020. A class of fast and accurate summation algorithms. *SIAM journal on scientific computing*, 42(3):A1541–A1557.

Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. 2019. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Hanhoon Kang, Seong Joon Yoo, and Dongil Han. 2012. Senti-lexicon and improved naïve bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5):6000–6010.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Jan Philip Wahle, Terry

Ruas, Meriem Beloucif, Christine de Kock, Nirmal Surange, Daniela Teodorescu, Ibrahim Said Ahmad, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino D. M. A. Ali, Ilseyar Alimova, Vladimir Araujo, Nikolay Babakov, Naomi Baes, Ana-Maria Bucur, Andiswa Bukula, Guanqun Cao, Rodrigo Tufino Cardenas, Rendi Chevi, Chiamaka Ijeoma Chukwuneke, Alexandra Ciobotaru, Daryna Dementieva, Murja Sani Gadanya, Robert Geislinger, Bela Gipp, Oumaima Hourrane, Oana Ignat, Falalu Ibrahim Lawan, Rooweither Mabuya, Rahmad Mahendra, Vukosi Marivate, Andrew Piper, Alexander Panchenko, Charles Henrique Porto Ferreira, Vitaly Protasov, Samuel Rutunda, Manish Shrivastava, Aura Cristina Udrea, Lilian Diana Awuor Wanzare, Sophie Wu, Florian Valentin Wunderlich, Hanif Muhammad Zhafran, Tianhui Zhang, Yi Zhou, and Saif M. Mohammad. 2025a. Brighter: Bridging the gap in human-annotated textual emotion recognition datasets for 28 languages. *arXiv preprint arXiv:2502.11926*.

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Seid Muhie Yimam, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine De Kock, Tadesse Destaw Belay, Ibrahim Said Ahmad, Nirmal Surange, Daniela Teodorescu, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino Ali, Vladimir Araujo, Abinew Ali Ayele, Oana Ignat, Alexander Panchenko, Yi Zhou, and Saif M. Mohammad. 2025b. SemEval-2025 task 11: Bridging the gap in text-based emotion detection. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, Vienna, Austria. Association for Computational Linguistics.

Lutz Prechelt. 2002. Early stopping-but when? In *Neural Networks: Tricks of the trade*, pages 55–69. Springer.

Tal Ridnik, Emanuel Ben-Baruch, Nadav Zamir, Asaf Noy, Itamar Friedman, Matan Protter, and Lihi Zelnik-Manor. 2021. Asymmetric loss for multi-label classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 82–91.

Rui Song, Zelong Liu, Xingbing Chen, Haining An, Zhiqi Zhang, Xiaoguang Wang, and Hao Xu. 2023. Label prompt for multi-label text classification. *Applied Intelligence*, 53(8):8761–8775.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Grigorios Tsoumakas and Ioannis Katakis. 2008. Multi-label classification: An overview. *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications*, pages 64–74.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Zining Wang, Yanchao Zhao, Guanghui Han, and Yang Song. 2024. Qfnu_cs at semeval-2024 task 3: A hybrid pre-trained model based approach for multimodal emotion-cause pair extraction task. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 349–353.

Da Yin, Tao Meng, and Kai-Wei Chang. 2020. Sentibert: A transferable transformer-based architecture for compositional sentiment semantics. *arXiv preprint arXiv:2005.04114*.

Jun Zhao, Kang Liu, and Liheng Xu. 2016. Sentiment analysis: Mining opinions, sentiments, and emotions.

Guangmin Zheng, Jin Wang, and Xuejie Zhang. 2022. YNU-HPCC at SemEval-2022 task 6: Transformer-based model for intended sarcasm detection in English and Arabic. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 956–961, Seattle, United States. Association for Computational Linguistics.

He Zhu, XF Lu, and Lei Xue. 2023. Emotional analysis model of financial text based on the bert [j]. *Journal of Shanghai University (Natural Science Edition)*, 29(1):118–128.