# A Social Media Emotion Detection Model Based on Implicit Emotion Multi-Hop Memory and Knowledge Embedding

Meiling Liu, Xinlei Xi,Ming Li*
*School of Computer and Control Engineering, Northeast Forestry University, Harbin, China*
{mlliu, liming_nefu, xixl0119}@nefu.edu.cn

***Abstract*—Social media platforms are a primary source for public sentiment analysis. However, a significant portion of online expression is implicitly conveyed through metaphors or context, posing a challenge for traditional models. Existing approaches relying on surface-level semantics often fail to capture these cues due to a lack of background knowledge. To address this, we propose a Social Media Emotion Detection Model based on Implicit Emotion Multi-Hop Memory and Knowledge Embedding. Our model incorporates a Knowledge-Embedded Attention Mechanism to retrieve external commonsense knowledge, bridging the gap between literal text and implicit intent. Furthermore, we design an Implicit Emotion Multi-Hop Memory Network to capture long-range contextual dependencies and learn the prior distribution of emotional states. Experiments on the fine-grained GoEmotions dataset demonstrate that our method achieves state-of-the-art performance, yielding a Precision of 54.66%, Recall of 53.80%, and Macro-F1 score of 52.34%. These results significantly outperform strong baselines such as BERT (Macro-F1: 46%) and BiDLSTM (Macro-F1: 50.4%) in detecting implicit and complex emotions. This superior performance is attributed to bridging semantic gaps via knowledge embeddings and capturing long-range dependencies through the multi-hop memory network.**

***Index Terms*—Social Media; emotion Detection; Implicit Emotion; Knowledge Embedding; Multi-Hop Memory**

## I. Introduction

Emotion is an interdisciplinary field involving psychology and computer science. In psychology, emotions are expressed as mental states different from levels of meditative, emotion, behavioral reactions, and joy or unhappiness [1]. In computer science, emotions are identified via audio, video, or text. Textual emotion analysis is challenging because emotions are often derived implicitly from conceptual interactions rather than explicit emotion words. Emotional expression is an important way of communication in interpersonal relationships [2]. It can be expressed as positive, negative, or neutral emotions [3]. These encompass various states like happiness and sadness. In this way, emotions are expressed in various forms for communication, and rich sources of textual information are obtained from YouTube, social networking sites such as Twitter, Facebook, where people spend most of their time posting and expressing their emotions [4]. Analyzing these social media platforms is crucial for investigating social trends, tracking customer feedback, and shaping business strategies.

With the continuous development of deep learning methods, a novel deep architecture based on neural attention has been proposed, called Transformer [5]. Transformers are good at handling long-term dependencies in text. Since the introduction of the initial Transformer model, several advanced architectures have been proposed: BERT [6],XLNet [7], ROBERTa [8],ALBERT [9]. However, these models often struggle to discern subtle differences between fine-grained emotion categories. Moreover, they exhibit deficiencies in capturing probabilistic distribution patterns and contextually relevant information tied to specific emotions. To address these issues, our main contributions are:

1) We constructed the PeM-Bert pre-training model, integrating dictionary-based emotional information via emotion masking language modeling. It utilizes P-tuning to continuously optimize emotion templates, enhancing generalization for downstream tasks.
2) We introduced a knowledge-embedded attention mechanism to address the insufficient recognition of fine-grained emotion differences, integrating external emotional knowledge into PeM-Bert's representations.
3) We proposed an Implicit Emotion Multi-hop Memory Model, containing an Implicit Emotion Module to learn prior emotion distributions, and a Multi-hop Memory Module to capture specific contextual emotion features. Their collaborative learning resolves the lack of prior knowledge regarding sentence emotion distributions.
4) Extensive comparative and ablation experiments demonstrate our model's effectiveness and superior performance in emotion detection tasks.

## II. Related Work

In the past, most work in text emotion detection primarily focused on identifying the six basic emotions proposed by Paul Ekman [10] (sadness, happiness, disgust, anger, surprise, and fear) and the eight primary emotions proposed by Robert Plutchik [11] (joy vs. sadness, surprise vs. anticipation, anger vs. fear, and trust vs. disgust). However, real-life human emotional expression is far more complex and diverse, encompassing states like shame, guilt, and pride.

It is well-known that external knowledge sources can provide explicit knowledge for the task at hand, complementing the representations implicitly learned by deep learning models [12]. Two main approaches integrate knowledge into pre-trained models. First, these language models can be retrained

from scratch by modifying the raw input [13], through multi-task learning [14], or by augmenting word embeddings [15]–[17]. While effective, they require expensive retraining when adding new knowledge sources. The second approach enriches representations through early or late fusion. Early fusion integrates knowledge at the input stage. These knowledge sources are often linguistic in nature, such as auxiliary sentences [18], etc.

Researchers have analyzed emotion detection from different perspectives in recent years: Dibyendu et al. [19] started from the semantic aspect and proposed a sentence-level emotion detection technology by applying semantic rules, which also included negative words, and built an emotion detection model accordingly. Xiao Zhang et al. [20] solved the problem of various emotion detection in social media from the user level. They adopted a factor graph-based emotion recognition model, incorporated emotion label relevance, social relevance and temporal relevance into a general framework, and detected various emotions based on multi-label learning methods.

Table I lists public datasets for emotion detection. Among the commonly used public datasets, we use the GoEmotions [21] dataset, which is the largest artificially annotated corpus. Crowdsourced from 58k Reddit comments, it features 27 fine-grained emotion labels plus neutrality. Table II provides an illustrative example.

TABLE I
EMOTION DETECTION RESEARCH DATASET DESCRIPTION

| Name | Description | Labels |
|---|---|---|
| ISEAR | Obtained from cross-cultural studies across 37 countries | Joy, Fear, Anger, Guilt, Sadness, Shame, Disgust |
| SemEval | Based on Google News, headlines, major newspapers, Twitter, and dialogues | Joy, Sadness, Fear, Anger, Surprise, Disgust |
| Emotion-Stimulus | Contains 1,594 sentences with emotion labels, annotated from Frame Nets | Joy, Sadness, Fear, Anger, Surprise, Disgust |
| WASSA | Constructed from tweets | Joy, Sadness, Fear, and Anger |
| Emobank | News headlines, blogs, newspapers, essays, fiction, and letters | Joy, Fear, Sadness, Anger, Surprise, Disgust |
| GoEmotions | 58k curated comments extracted from Reddit | Amusement, Disgust, Remorse, Sadness, Anger, etc. |

## III. METHODOLOGY

We propose a unified framework for fine-grained emotion detection that integrates a Knowledge-Embedded Attention mechanism with an Implicit Emotion Multi-hop Memory network. The overall architecture is illustrated in Fig. 1.

TABLE II
ILLUSTRATIVE EXAMPLES OF THE GOEMOTIONS DATASET

| Text | Labels |
|---|---|
| "OMG, yep!!! That is the final answer. Thank you so much!" | gratitude, approval |
| "I'm not even sure what it is, why do people hate it" | confusion |
| "Guilty of doing this tbph" | remorse |
| "This caught me off guard for real. I'm actually off my bed laughing" | surprise, amusement |
| "I tried to send this to a friend but [NAME] knocked it away." | disappointment |
| "Almost everyone has MASSIVE problems with some of the decisions Nintendo makes." | neutral |
| "Does nobody notice that this is a doctored photo..??" | disapproval |

### A. Construction of the PeMBKEA model

To improve generalization, we construct the PeM-Bert pre-trained model, utilizing emotion masking language modeling (eMLM) to integrate dictionary-based emotional information. It uses P-tuning to optimize emotion templates for downstream tasks. To differentiate fine-grained emotion categories, we designed a knowledge-embedded attention mechanism. This mechanism transforms input text into emotion encodings using external dictionary knowledge and combines them with PeM-Bert's contextual representations to form an emotion-integrated representation.The model structure is shown in Fig. 2.

We adopt BERT as the pre-trained language model. An input token sequence $X_{1:n} = \{x_0, x_1, \cdot \cdot \cdot, x_n\}$ passes through the pre-trained layer $e \in M$, yielding embeddings $\{e(x_0), e(x_1), ..., e(x_n)\}$. Let unmasked context X and masked target y be organized by prompt function P to predict y, forming template T. Unlike traditional Prompts, P-tuning replaces discrete tokens with pseudo-[P], utilizing continuous vectors to construct T, as in formula (1):

$$\{h_0, \ldots, h_i, e(x), h_{i+1}, \ldots, h_m, e(y)\} \tag{1}$$

Here, $h_i$ $i(0 \leq i < m)$ is a trainable vector. Through the downstream loss function, continuous prompt $h_i$ $(0 \leq i < m)$ is progressively optimized, as shown in Equation (2):

$$\hat{h}_{0:m} = \arg_h \min(M(x, y)) \tag{2}$$

To avoid the complexities of an additional LSTM for token embeddings, we construct training masks based on the eMLM proportion, generating no new masks for validation/test sets. During training, only the embedding layer's gradients are updated. We employ a mask matrix (setting only template positions to 1) and a register-hook to exclusively retain and update the template gradients.

For masking, we employ our proposed Emotion Masked Language Modeling (eMLM), a variant of MLM. Unlike BERT's uniform 15% masking probability, eMLM assigns

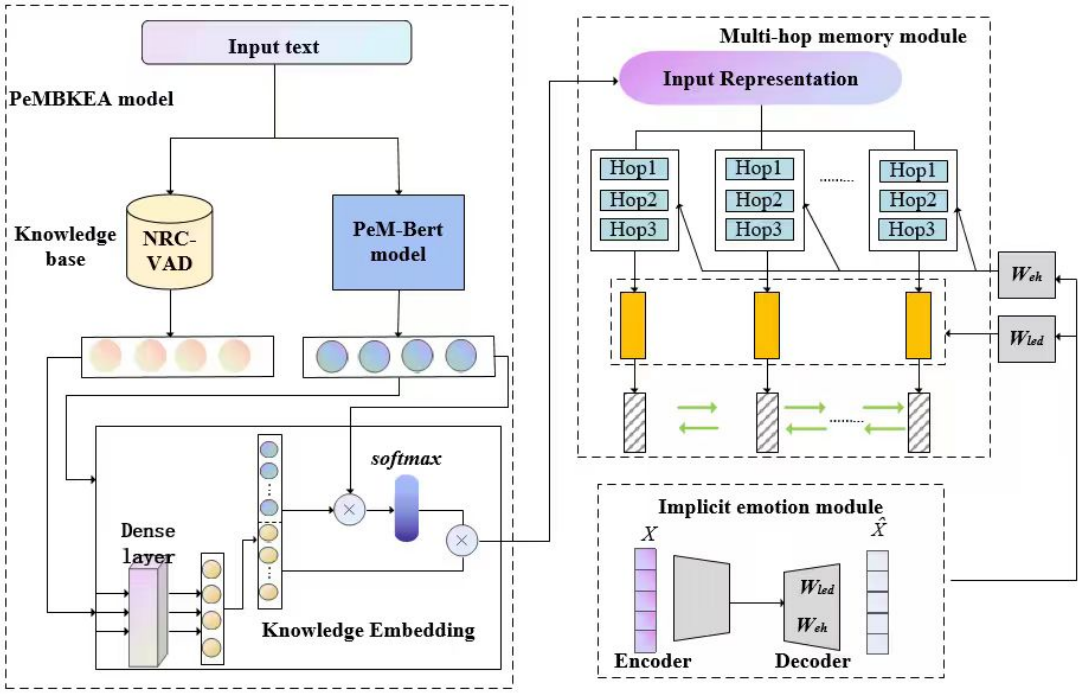


Fig. 1. The model structure of PeMBKEA-LEMJM

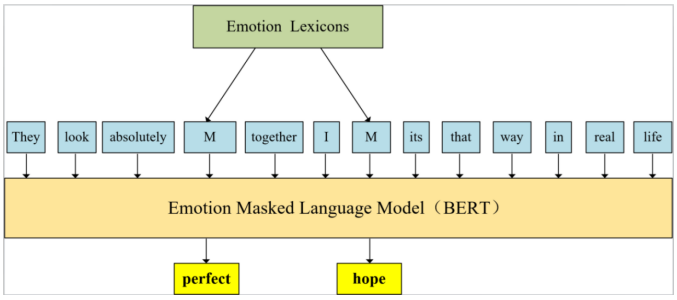


Fig. 2. Emotional mask Language modeling

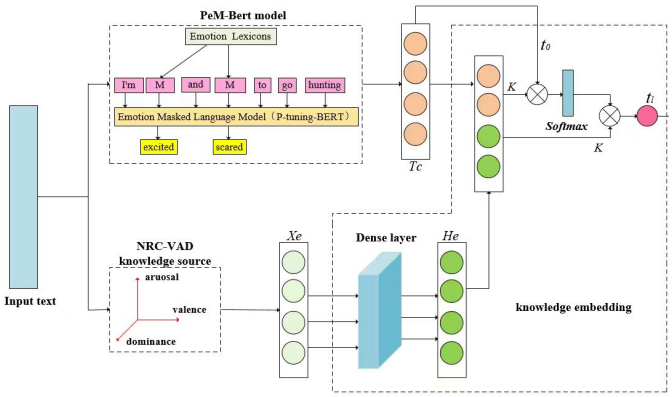


Fig. 3. The model structure of PeMBKEA

a higher probability to emotion-rich words selected from a dictionary. We use the EmoLex [22] dictionary, which enables eMLM to focus on emotion-rich vocabulary during the masking task and correspondingly increases their masking probability. To maintain a stable total probability sum, the masking probability for words expressing weaker emotions is reduced, as shown in Fig. 3.

Denoting this probability as hyperparameter k, the masking process is: given an input sentence, extract dictionary words denoted by E, and set their masking probability to $P(w_e) = k \nabla w_e \in E$. To ensure a total of 15% of words are blocked, formula (3) is used to reduce the masking probability of non-emotion-rich words:

$$P(w_n) = \frac{\max(|S| \cdot 0.15 - |E| \cdot k, 0)}{|S| - |E|}, \forall w_n \notin E \quad (3)$$

Where $|.|$ represents the size of the set.Thus, the PeM-Bert pre-training model, combining P-tuning with eMLM, is successfully constructed.

Pre-trained context representations enhance language understanding by encoding meaningful content and surrounding contexts. Thus, we denote the hidden states of the last layer from the PeM-Bert model as matrix $T_c = (t_0, t_1, ..., t_n)$, where $T_c \in \mathbb{R}^{N \times l}$. Here, $l_c$ is the output representation size, and the [CLS] token $t_0$ serves as the contextual representation of the entire input.

In the knowledge-embedded attention mechanism, the input X is transformed into a feature vector $X_e$ (padded to a fixed length of 512, the base model's default). This vector is projected via a dense layer to form the sentence-level emotion encoding $H_e$, where $H_\epsilon \in \mathbb{R}^{l_e \times l_c}$. Based on self-attention,

$H_e$ is concatenated with the context representation $T_c$ to form K, where $K \in \mathbb{R}^{(N\times l_e)\times l_c}$. Using $t_0$ as a Query, we obtain softmax-attention scores s to weight the matrix K. Including $H_e$ in K with $T_c$ retains both contextual information and the integrated emotional knowledge. The overall attention layer is defined by Equations (4), (5), and (6):

$$K = \text{concat}(T_c, H_e) \tag{4}$$

$$s = \text{softmax}(t_0^T \cdot K) \tag{5}$$

$$t_l = S^T \cdot K \tag{6}$$

Finally, the enhanced context representation $t_l$ serves as the input for the implicit emotion module.

### B. Construction of the LEMJM model

*1) Implicit Emotion Module :*

Given that it is not possible to precisely determine the distribution of emotions, we construct it as a set of implicit variables and use the Variational Autoencoder [23] (VAE) method to learn the latent multimodal distribution representation during the reconstruction process of the input features. The schematic diagram of the Implicit Emotion Module is shown in Fig. 4.

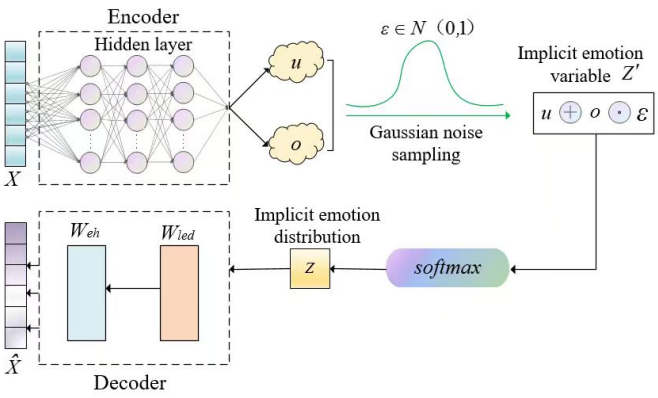


Fig. 4. The model structure of Implicit emotion

Its input X is directly derived from the enhanced context representation generated by the PeMBKEA model's attention mechanism.The encoder $h_e(\cdot)$, consisting of multiple non-linear hidden layers, transforms the input $X \in \mathbb{R}^L$ (where L represents the maximum sequence length) into prior parameters u and $\sigma$, as shown in equations (7) and (8):

$$u = h_{e,u}(X) \tag{7}$$

$$\log \sigma = h_{e,\sigma}(X) \tag{8}$$

Subsequently, we define an implicit emotion variable $Z' = u + \sigma \cdot \epsilon$, where $\epsilon$ is a Gaussian noise variable sampled $from\mathcal{N}(0,1)$, and $Z' \in \mathbb{R}^K$ (K represents the number of emotion labels). It is normalized into the implicit emotion distribution Z through the softmax function, as shown in equation (9):

$$Z = \text{softmax}(Z') \tag{9}$$

Accordingly, the implicit emotion distribution $p(\epsilon_k, |k = 1, \cdots, K)$ is reflected in Z. In the decoder part, variational inference approximates the posterior distribution of Z. First, Z is converted into emotion feature embeddings using a linear hidden layer, with the transformation formula shown in equation (10):

$$\mathbb{R}^{\text{led}} = h_{\text{led}}(Z; W_{\text{led}}) \tag{10}$$

Where $W_{led} \in \mathbb{R}^{K\cdot led}$ (with corresponding embedding dimension $E_{tot}$) is the learned embedding of the implicit emotion distribution, designed to guide the memory module in feature extraction. To learn a high-level representation of rich emotional features during reconstruction, another non-linear hidden layer $h_{eh}(\cdot)$ further decodes $R^{led}$, as shown in equation (11):

$$\mathbb{R}^{\text{eh}} = h_{\text{eh}}(R^{\text{led}}; W_{\text{eh}}) \tag{11}$$

Where $W_{eh} \in \mathbb{R}^{L\cdot E_{eh}}$ ($E_{eh}$ representing the embedding dimension) is the global embedding of emotion features, acting as external memory to assist the multi-hop memory module. Finally, the decoding formula is shown in equation (12):

$$\hat{X} = \text{softmax}(h_{\text{rec}}(R^{\text{eh}})) \tag{12}$$

Where $h_{rec}(\cdot)$ is the final hidden layer, and $\hat{X}$ is the reconstructed input feature.

*2) Multi-hop Memory Module :*

In a single sentence, emotion-related evidence is often dispersed and intermingled, complicating effective feature extraction. Therefore, we employs a multi-hop memory network [24] to design a multi-hop memory module specifically for emotion detection tasks to unearth the rich contextual information of the relevant emotions. The overall structure of the model is shown in Fig. 5.

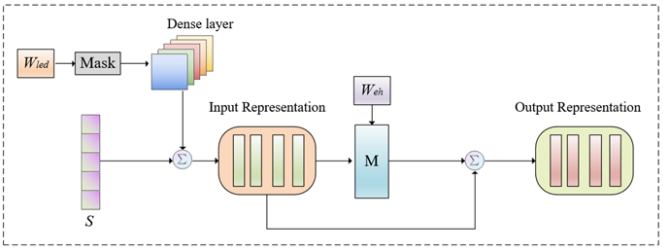


Fig. 5. The model structure of Multiple hop memory

In this module, masking is applied to $W_{led}$ to select the k-th emotion embedding, as shown in equation (13):

$$\text{Mask} = \underbrace{[\vec{1}; \cdots; \vec{0}]}_{K} \tag{13}$$

Where Mask contains K vectors; only the k-th vector is entirely ones, encouraging the module to focus on its corresponding emotion. Subsequently, $W_{led,k}$ is transformed into $R_k^{le}$ via a dense layer. Given an input text representation $S = \{e_1, \cdots, e_L\}$ (where $e_i \in \mathbb{R}^{E_e}$), an attention mechanism calculates the relevance between the text embedding and the implicit emotion representation $R_L^{le}$, as shown in equation (14):

$$\mathbf{a}_{\text{input}} = \text{softmax}(W_{\text{input}}[R_k^{\text{le}}; S]) \tag{14}$$

where $W_{input}$ is the parameter matrix, and [;] denotes vector concatenation. Based on weights $a_{input}$ and text representation S, the emotion-related input context representation $R^{input}$ is obtained:

$$\mathbf{R}^{\text{input}} = \sum \mathbf{a}_{\text{input}} \cdot S \tag{15}$$

The memory representation M is derived from the global emotion feature embedding $W_{ch}$ via a linear transformation:

$$\mathbf{M} = f_\beta(W_{\text{eh}}) \tag{16}$$

Attention is again used to compute the match between $R^{input}$ and M, yielding the current hop's output context representation, as shown in equations (17) and (18):

$$\mathbf{a}_{\text{output}} = \text{softmax}\left(o(W_{\text{output}}[\mathbf{R}^{\text{input}}; \mathbf{M}])\right) \tag{17}$$

$$\mathbf{R}^{\text{output}} = \sum \mathbf{a}_{\text{output}} \cdot \mathbf{R}^{\text{input}} \tag{18}$$

The output $R^{output}$ is then combined with S to form the emotion feature representation $R_k^m$, as shown in equation (19):

$$\mathbf{R}_k^m = \mathbf{R}^{\text{output}} + S \tag{19}$$

Thus, the current hop's output carries features related to the corresponding emotion. The model contains K private memory modules for K emotion categories, sharing the implicit emotion module, embeddings $W_{led}$, and $W_{ch}$. The multi-hop memory module uses multi-hop attention to iteratively explore features based on previous states. The final representation $R_p^m$ is obtained after multiple hops, as shown in equation (20):

$$\mathbf{R}_{p,k}^m = \text{Mem}_{p,k}(\mathbf{R}_{p-1,k}^m) \tag{20}$$

Finally, $R_{p,k}^m$ is concatenated with the implicit emotion distribution embedding $W_{led.k}$ to form the ultimate feature representation $R_{total,k}$ for emotion detection, as shown in equation (21):

$$\mathbf{R}_{\text{total},k} = \left[\mathbf{R}_{p,k}^m; W_{\text{led},k}\right] \tag{21}$$

## IV. Experiment

### A. Dataset and Evaluation Metrics

We utilize the GoEmotions [21] dataset, which is currently the largest manually annotated corpus. It features 27 fine-grained emotion labels plus neutrality. In this dataset, 83% of examples have one label, 15% have two, 2% have three, and 0.2% have four or more. We employ Precision, Recall, and Macro-F1 score as primary evaluation indicators.

### B. Lexicon Features

The external knowledge source we use is the NRC-VAD [24] dictionary data, which contains quantified assessments of the pleasure, arousal, and dominance of 20,000 words. Ratings range from 0 to 1 (negative to positive valence, calm to excited arousal, and submissive to dominant dominance). To convert input text X into these vectors, words are replaced by their corresponding rating values. Words not found in the dictionary are assigned a neutral score of 0.5.

### C. Experimental Results and Analysis

To evaluate the performance of the PeMBKEA-LEMJM model proposed in this paper, we not only compared and analyzed it with models such as BERT [6], BiLSTM [25], XLNet [7]. ALBERT [9], and ELECTRA [26], but also selected several comparable advanced research methods for comparative analysis. The experimental results are shown in Table III. Relevant methods include:

Seq2Emo [27]: This study proposes a sequence-to-emotion (Seq2Emo) method, which is similar to the classifier chain approach. It implicitly models emotion associations within a bidirectional decoder, utilizing an attention mechanism to focus on input words relevant to the current emotion.

BiDLSTM [28]: This research introduces an innovative method aimed at implementing fine-grained emotion classification on tweets using a bidirectional expanded long short-term memory (BiDLSTM) model with integrated attention. Incorporating bidirectional, expansion, and attention mechanisms into the traditional LSTM enhances its ability to handle complex data dependencies.

TABLE III
Experimental Results Comparison

| Model | Precision | Recall | Macro-F1 |
|---|---|---|---|
| BERT | 40.00 | 63.00 | 46.00 |
| BiLSTM | 56.50 | 39.30 | 43.90 |
| XLNet | 47.09 | 51.92 | 48.50 |
| ALBERT | 48.60 | 52.90 | 49.60 |
| ELECTRA | 47.40 | 50.40 | 47.50 |
| BiDLSTM | 49.50 | 53.50 | 50.40 |
| Seq2Emo | 48.86 | 53.11 | 50.17 |
| **PeMBKEA-LEMJM** | **54.66** | **53.80** | **52.34** |

Our model significantly improves Precision, Recall, and Macro-F1. Combining P-tuning and eMLM during pre-training compensates for limited generalization in social media contexts. Furthermore, the knowledge-embedded attention mechanism leverages external emotion data to address fine-grained

differences, while LEMJM captures prior distributions and related context, comprehensively boosting detection performance.

### *D. Ablation Experimental*

To demonstrate our model's effectiveness, we designed four ablation experiments on the GoEmotions dataset, comparing BERT, PeM-Bert, KEA-Bert, PeMBKEA, and PeMBKEA-LEMJM. The experimental results are shown in Table IV. PeM-Bert significantly outperforms BERT in Precision and Macro-F1 Score, proving that combining P-tuning and eMLM addresses pre-trained models' insufficient generalization. Similarly, KEA-Bert's outperformance confirms that enriching context with external emotion knowledge helps differentiate fine-grained emotions. PeMBKEA improves upon both PeM-Bert and KEA-Bert, indicating their combination successfully addresses both generalization and fine-grained differences simultaneously. Finally, applying the LEMJM method yields the best overall performance (PeMBKEA-LEMJM). This indicates that LEMJM captures subtle emotion-related contexts and long-distance dependencies, further enhancing emotion detection capabilities beyond general models.

TABLE IV
ABLATION STUDY RESULTS ON GOEMOTIONS DATASET

| Model | Precision | Recall | Macro-F1 |
|---|---|---|---|
| BERT [6] | 40.00 | 63.00 | 46.00 |
| PeM-Bert | 56.98 | 48.98 | 50.12 |
| KEA-Bert | 51.40 | 52.50 | 51.00 |
| PeMBKEA | 53.42 | 54.37 | 51.96 |
| **PeMBKEA-LEMJM** | **54.66** | **53.80** | **52.34** |

## V. CONCLUSION

We proposed PeMBKEA, integrating PeM-Bert (combining P-tuning and eMLM to optimize templates and resolve generalization limits) and a knowledge-embedded attention mechanism utilizing external knowledge for fine-grained emotion distinction. Building on this, we proposed LEMJM. Its Implicit Emotion Module learns prior distributions, while the Multi-hop Memory Module extracts contextual features, collaboratively mitigating prior knowledge deficits. Future work will encode multimodal information (emojis, images) to handle diverse comments, address interpersonal expression variability, and integrate knowledge graphs to explore relational impacts on emotion recognition.

## REFERENCES


[1] Lutz C A. Unnatural emotions: Everyday sentiments on a Micronesian atoll and their challenge to Western theory[M]. University of Chicago Press, 2011.

[2] Gross J J, Carstensen L L, Pasupathi M, et al. Emotion and aging: experience, expression, and control[J]. Psychology and aging, 1997, 12(4): 590.

[3] Kopelman S, Rosette A S, Thompson L. The three faces of Eve: Strategic displays of positive, negative, and neutral emotions in negotiations[J]. Organizational Behavior and Human Decision Processes, 2006, 99(1): 81-101.

[4] Fox, Andrew S., et al., eds. The nature of emotion: Fundamental questions. Oxford University Press, 2018.

[5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[6] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

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

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

[9] Lan Z, Chen M, Goodman S, et al. Albert: A lite bert for self-supervised learning of language representations[J]. arXiv preprint arXiv:1909.11942, 2019.

[10] Ekman P. Basic emotions. Handbook Cognit Emot. 1999;98(45-60):16.

[11] Plutchik R. A general psychoevolutionary theory of emotion. Amsterdam,Netherlands: Elsevier; 1980 (pp. 3–33).

[12] Roy A, Pan S. Incorporating extra knowledge to enhance word embedding[C]//Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. 2021: 4929-4935.

[13] Poerner N, Waltinger U, Schütze H. E-BERT: Efficient-Yet-Effective Entity Embeddings for BERT[C]//Findings of the Association for Computational Linguistics: EMNLP 2020. 2020: 803-818.

[14] Tian H, Gao C, Xiao X, et al. SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis[J]. arXiv preprint arXiv:2005.05635, 2020.

[15] Ke P, Ji H, Liu S, et al. SentiLARE: Sentiment-Aware Language Representation Learning with Linguistic Knowledge[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural .

[16] Language Processing (EMNLP). 2020: 6975-6988.Chen W, Su Y, Yan X, et al. KGPT: Knowledge-grounded pre-training for data-to-text generation[J]. arXiv preprint arXiv:2010.02307, 2020.

[17] Liu Y, Wan Y, He L, et al. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(7): 6418-6425.

[18] Wu Z, Ong D C. Context-guided bert for targeted aspect-based sentiment analysis[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(16): 14094-14102.

[19] Seal D, Roy U K, Basak R. Sentence-level emotion detection from text based on semantic rules[M]//Information and Communication Technology for Sustainable Development. Springer, Singapore, 2020: 423-430.

[20] Zhang X, Li W, Ying H, et al. Emotion detection in online social networks: a multilabel learning approach[J]. IEEE Internet of Things Journal, 2020, 7(9): 8133-8143.

[21] Demszky D, Movshovitz-Attias D, Ko J, et al. GoEmotions: A dataset of fine-grained emotions[J]. arXiv preprint arXiv:2005.00547, 2020.

[22] Mohammad S M, Turney P D. Crowdsourcing a word–emotion association lexicon[J]. Computational intelligence, 2013, 29(3): 436-465.

[23] Kingma D P, Welling M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.

[24] Sukhbaatar S, Weston J, Fergus R. End-to-end memory networks[J]. Advances in neural information processing systems, 2015, 28.

[25] Mohammad S. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words[C]//Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers). 2018: 174-184.

[26] Lin Z, Feng M, Santos C N, et al. A structured self-attentive sentence embedding[J]. arXiv preprint arXiv:1703.03130, 2017.

[27] Clark K, Luong M T, Le Q V, et al. Electra: Pre-training text encoders as discriminators rather than generators[J]. arXiv preprint arXiv:2003.10555, 2020.

[28] Huang C, Trabelsi A, Qin X, et al. Seq2emo: a sequence to multi-label emotion classification model[C]//Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies. 2021: 4717-4724.

[29] Schoene A M, Turner A P, Dethlefs N. Bidirectional Dilated LSTM with Attention for Fine-grained Emotion Classification in Tweets[J]. Affcon@ aaai, 2020, 2614: 100-1