



ARTICLE

Event-Aware Sarcasm Detection in Chinese Social Media Using Multi-Head Attention and Contrastive Learning

Kexuan Niu, Xiameng Si*, Xiaojie Qi and Haiyan Kang

College of Computer Science, Beijing Information Science and Technology University, Beijing, 100192, China

*Corresponding Author: Xiameng Si. Email: sixiameng@bistu.edu.cn

Received: 11 March 2025; Accepted: 17 July 2025; Published: 29 August 2025

ABSTRACT: Sarcasm detection is a complex and challenging task, particularly in the context of Chinese social media, where it exhibits strong contextual dependencies and cultural specificity. To address the limitations of existing methods in capturing the implicit semantics and contextual associations in sarcastic expressions, this paper proposes an event-aware model for Chinese sarcasm detection, leveraging a multi-head attention (MHA) mechanism and contrastive learning (CL) strategies. The proposed model employs a dual-path Bidirectional Encoder Representations from Transformers (BERT) encoder to process comment text and event context separately and integrates an MHA mechanism to facilitate deep interactions between the two, thereby capturing multidimensional semantic associations. Additionally, a CL strategy is introduced to enhance feature representation capabilities, further improving the model's performance in handling class imbalance and complex contextual scenarios. The model achieves state-of-the-art performance on the Chinese sarcasm dataset, with significant improvements in accuracy (79.55%), F1-score (84.22%), and an area under the curve (AUC, 84.35%).

KEYWORDS: Sarcasm detection; event-aware; multi-head attention; contrastive learning; NLP

1 Introduction

The rapid proliferation of social media platforms has significantly transformed the ways in which individuals communicate, with sarcasm emerging as a prevalent form of expression. Often subtle and indirect, sarcasm enables users to convey complex emotions and opinions by juxtaposing a literal statement with its intended, often contradictory, meaning. This phenomenon is particularly pronounced within digital discourse, where the informal nature of interactions lends itself to the frequent use of sarcastic remarks. However, despite its ubiquity, sarcasm presents substantial challenges for numerous tasks in the field of Natural Language Processing (NLP), particularly those aimed at sentiment analysis. The intricate linguistic features and contextual dependencies of Chinese sarcasm make its identification particularly challenging for NLP systems.

Sarcasm detection is inherently challenging due to its reliance on contextual and cultural nuances. Conventional methods often fail to model event-specific contexts [1], while cross-lingual approaches struggle with Chinese linguistic traits (e.g., slang, wordplay) and data scarcity [2]. These limitations hinder accurate interpretation of sarcastic intent in Chinese social media.

In recent years, the emergence of multimodal approaches has opened up new avenues for sarcasm detection research. Farabi et al., in their systematic review, emphasized the significance of incorporating multimodal data (such as speech, images, and videos) to enhance sarcasm detection capabilities [3]. Despite



these advancements, text-based sarcasm detection remains fraught with challenges, particularly as it has predominantly focused on English-language corpora, with relatively limited attention given to Chinese corpora. On one hand, sarcastic expressions frequently rely on implicit elements—such as irony, exaggeration, or metaphor—to convey their intended meaning, making it difficult for traditional single-modality sentiment analysis models to capture the nuanced semantics of such expressions [4]. On the other hand, interpreting sarcasm often hinges on understanding the specific event context, a factor that is especially crucial in Chinese social media datasets. For example, as illustrated in Fig. 1, a comment like “What an unforgettable moment!” could express an entirely different sentiment depending on the context (e.g., a thrilling sports victory or an unfortunate mishap). Conventional architectures exhibit limited capacity for contextual salience detection, thereby constraining their discriminative power in pragmatic interpretation tasks.

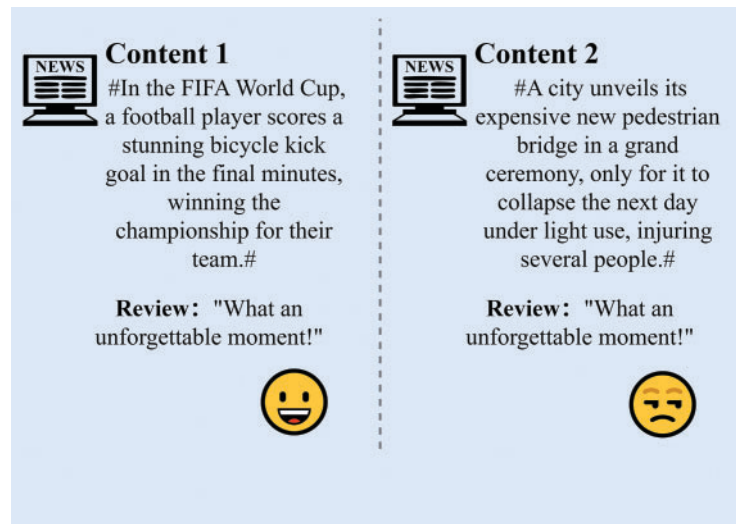


Figure 1: Context-dependent sarcasm in different scenarios

The ubiquity of sarcasm in social media, driven by its ability to convey contradictory meanings, presents both challenges and opportunities. Accurate sarcasm detection is critical for applications like public opinion analysis [5], yet existing systems struggle with contextual dependencies, especially on Chinese platforms [6]. In conclusion, the unique linguistic traits and the strong contextual dependence of sarcasm in Chinese social media present considerable challenges to traditional natural language processing (NLP) methods. Achieving accurate sarcasm detection in this domain is essential for advancing sentiment analysis and gaining a deeper understanding of public sentiment [7].

Despite notable advancements in sarcasm detection research, several critical limitations continue to hinder the development of more effective models:

- **Inadequate Contextual Modeling:** A majority of existing approaches treat sarcasm detection as a task limited to isolated sentence-level classification, failing to account for the complex semantic interrelations between sarcasm and the broader event context [8]. For instance, the same comment may convey contradictory sarcastic intents depending on the specific circumstances surrounding the event. This oversight in integrating event-specific context with textual content significantly hampers the model's overall performance and its ability to accurately interpret the intended sarcasm.
- **Data Scarcity:** While sarcasm detection studies have primarily concentrated on English-language corpora, the research in the Chinese context remains constrained by the availability of high-quality

datasets [9]. The linguistic characteristics of sarcasm in Chinese differ significantly from those in English, including variations in semantic structures, syntactic patterns, and rhetorical devices. Furthermore, the scarcity of sarcastic samples in Chinese datasets exacerbates the inherent difficulty of the detection task, limiting the depth and scope of research in this area [10].

- **Class Imbalance:** Sarcastic expressions are often underrepresented in real-world datasets, leading to a class imbalance that detrimentally affects model training. Despite efforts to alleviate this issue through techniques such as data sampling and reweighting, performance bottlenecks persist, particularly when dealing with large-scale datasets that exacerbate the imbalance problem.

The main contributions of this study are as follows: (1) An event-aware sarcasm detection framework is proposed, in which a dual-path BERT [11] architecture is utilized to separately encode comment text and event context, thereby enhancing the interpretation of context-dependent sarcasm. (2) A multi-head attention mechanism is incorporated to facilitate deeper feature interaction and to capture complex semantic relationships between comments and event context, thus improving the detection of subtle sarcastic expressions. (3) A hybrid training strategy integrating contrastive learning and focal loss is designed to enhance feature representation and address class imbalance, thereby improving model robustness by adaptively reweighting hard samples and optimizing cross-modal representation consistency. This approach is particularly crucial when dealing with limited or specialized datasets where robust feature learning and effective handling of imbalanced classes are paramount.

The structure of this paper is organized as follows: [Section 2](#) provides a comprehensive review of the current state of research in sarcasm detection. [Section 3](#) offers an in-depth discussion of the proposed model architecture, including data preprocessing, model design, and training methodologies. [Section 4](#) outlines the experimental setup and presents the results of the analysis. Finally, [Section 5](#) concludes with a summary of the study's contributions.

2 Related Work

Sarcasm is a complex linguistic phenomenon that conveys emotions and opinions through the contrast between literal meaning and actual intent. Against the backdrop of the widespread use of social media, sarcasm detection has gradually emerged as a critical research topic in the field of sentiment analysis. However, due to the often implicit nature of sarcastic expressions and their strong reliance on contextual information, sarcasm detection poses significant challenges both in theory and practice. Existing research primarily focuses on traditional machine learning, deep learning, and the integration of external knowledge. This paper reviews the progress in sarcasm detection research from the following four perspectives.

2.1 Based on Traditional Machine Learning

Early sarcasm detection methods primarily relied on traditional machine learning models, such as Support Vector Machines (SVM) and Logistic Regression (LR), which utilized manually crafted linguistic features for classification. Recent studies in Arabic sentiment analysis [12] further highlight the limitations of traditional classifiers in capturing nuanced cultural expressions, reinforcing the necessity of context-aware modeling even in non-English languages. González-Ibáñez et al. [13] employed an SVM model combined with bag-of-words and part-of-speech features to classify sarcastic expressions in a Twitter dataset, demonstrating superior performance relative to baseline models.

Traditional methods suffer from key limitations: handcrafted features fail to capture the nuanced semantics of sarcasm, and their dependence on domain-specific knowledge restricts generalizability. Moreover, these approaches overlook the essential interplay between text and context, which is critical for accurate

sarcasm detection. Their performance further deteriorates on Chinese corpora due to linguistic and cultural complexities, highlighting the need for more sophisticated and event-aware models.

2.2 Based on Deep Learning

The rapid development of deep learning technologies has provided new breakthroughs for sarcasm detection. Compared to traditional methods, deep learning can automatically learn features and capture the deep semantic information of text, thereby significantly improving the performance of sarcasm detection. Poria et al. [14] employed convolutional neural networks (CNNs) combined with sentiment, emotion, and personality features, which effectively enhanced the model's performance across multiple datasets. Ghosh et al. [15] proposed a neural network model that integrates context, user psychological states, and linguistic features, significantly improving sarcasm recognition. Furthermore, Ghosh et al. [16] combined CNNs, LSTMs, and deep neural networks (DNNs), achieving an F1 score of 0.92 across multiple datasets.

In addition to traditional neural networks, methods based on pre-trained language models have demonstrated excellent performance in sarcasm detection in recent years. Kalaivani et al. [17] utilized the BERT model to detect sarcasm in social media, and the experimental results showed that BERT exhibits strong robustness in handling informal expressions and conversational contexts. However, these methods typically focus on modeling single texts while neglecting the strong dependence of sarcastic expressions on external background knowledge, such as contextual information and event backgrounds. Effectively integrating external knowledge with text modeling remains a pressing challenge for further improving sarcasm detection performance.

2.3 Integrating External Knowledge

The formation and understanding of sarcasm often require extensive background knowledge and contextual information. In recent years, researchers have attempted to incorporate external knowledge into sarcasm detection models to address the limitations of single-text modeling. For instance, Wen et al. [18] proposed a sarcasm detection model (SAAG) that integrates sememe knowledge and auxiliary information, demonstrating excellent performance in complex Chinese sarcasm detection tasks. Peled et al. [19] developed a sarcasm explanation model (Sarcasm SIGN) based on monolingual machine translation, which enhances the understanding of sarcastic expressions by converting sarcastic text into non-sarcastic text aligned with its true meaning through machine translation techniques. Hazarika et al. [20] proposed the contextual sarcasm detector (CASCADE) model, which combines personalized features from users' historical comments with contextual information of discussion topics, achieving remarkable performance in sarcasm detection tasks.

In the field of multimodal sarcasm detection, Cai et al. [21] proposed a hierarchical fusion model that effectively models sarcastic expressions by integrating textual, visual, and image attribute information, significantly improving the accuracy of sarcasm detection. These studies indicate that incorporating external knowledge and multimodal information is a crucial direction for improving sarcasm detection performance.

Recent studies have attempted to enhance sarcasm detection performance by leveraging external knowledge augmentation and prompt learning. For instance, the Sarcasm Detection with Context and Common Sense (CCSD) model [22] generates auxiliary text through pre-trained commonsense knowledge bases and concatenates it with BERT-encoded features. However, its static knowledge struggles to adapt to dynamically evolving events. The Entity Knowledge-based Prompt Learning (EKPL) model [23] employs entity-driven prompt templates to inject domain-specific knowledge into pre-trained models, yet template-induced bias may constrain its generalizability to diverse linguistic expressions. The Sarcasm Detection with Topic and Incongruity (TISD) model [24] extracts topic-opinion chunks via syntactic parsing and models semantic contradictions, but exhibits limited capability in capturing cross-hierarchical semantic correlations

(e.g., hybrid expressions combining metaphors and irony). The Retrieval–Detection method for Verbal Irony (RDVI) model [25] innovatively introduces retrieval-augmented mechanisms to expand contextual information through similar document retrieval; however, retrieval noise and computational overhead may compromise practical utility.

While these methods demonstrate promising results in specific scenarios, they remain insufficient in modeling sarcastic expressions within Chinese social media contexts, which are highly event-dependent and semantically implicit. Crucially, the dynamic interaction mechanisms between user comments and event backgrounds have not been thoroughly explored.

2.4 Based on Attention Mechanisms

Attention mechanisms have been widely applied in natural language processing tasks in recent years. By dynamically calculating weight relationships between different elements, they can effectively capture key features within text. In sarcasm detection tasks, Diao et al. [26] proposed a model based on a multi-dimensional question-answering network, which leverages memory networks and attention mechanisms to model the relationship between context and target text, significantly improving detection performance. Kumar et al. [27] proposed a multi-head attention-based bidirectional LSTM model (MHA-BiLSTM), which effectively utilizes contextual information and handcrafted features, achieving excellent performance in sarcasm detection. Tay et al. [28] introduced a multi-dimensional intra-attention mechanism (MIARN), which explicitly models semantic contrast relationships to capture inconsistencies in sarcastic language, demonstrating outstanding performance in both detection accuracy and model interpretability. Lin et al. [29] proposed a sentence embedding model based on self-attention mechanisms, which enhances the expressive power of detection models by capturing multi-faceted semantic information of sentences.

In particular, multi-head attention mechanisms have demonstrated powerful modeling capabilities in handling long-range semantic dependencies and complex semantic expressions. Studies have shown that multi-head attention can effectively capture feature relationships between comment text and event context, offering a novel technical pathway for sarcasm detection.

3 Methodology

This section provides a detailed explanation of the event-aware Chinese sarcasm detection model based on the multi-head attention mechanism. The model aims to enhance the understanding and recognition of Chinese sarcastic expressions by incorporating event context information and optimizing the model structure. This section first introduces the construction and preprocessing of the dataset, followed by a detailed description of the overall model architecture and key components. Finally, the training and inference methods of the model are explained.

3.1 Data Preprocessing

This study utilizes the dataset provided by the Chinese Sarcasm Computation Evaluation Task (CSCET) for model training and evaluation. The dataset includes Weibo topics and related comments, annotated with sarcasm labels and sentiment polarity, providing high-quality corpus support for sarcasm detection tasks (detailed information about the dataset is provided in [Section 4.1](#)).

The event context in this study includes the Weibo topic title and topic content. The topic title concisely summarizes the event theme (e.g., “U.S. Presidential Election”), while the topic content describes event details (e.g., official reports, user discussions). This structured information enables the model to link comments with their contextual backgrounds.

During the data preprocessing stage, the following steps were performed to improve data quality and ensure the model's ability to fully understand the semantic relationships between comments and contextual information. First, the raw text underwent initial cleaning, removing null values and other noise to ensure the purity of the input data [30]. Second, to distinguish different text types in the input, a [REVIEW] token was added before the comment text, while [TITLE] and [CONTENT] tokens were added before the topic title and content, respectively, to help the model better understand the semantic relationships between comments and context. Finally, the pre-trained BERT-base-Chinese tokenizer [31] was used to tokenize and encode the comment and background text.

3.2 Model Architecture

The overall architecture of the proposed event-aware sarcasm detection model is illustrated in Fig. 2. The model consists of the following main modules: a dual-path BERT-based encoder, which extracts features separately from the comment text and event context; a multi-head attention layer, which models the interactions between the comment text and the event context; a feature fusion layer, which integrates interaction features of different granularities; a contrastive learning module, which enhances the discriminative power of sarcasm features through feature transformation and similarity computation. A classifier that determines the sarcasm of the comment based on the fused features. These modules work together to achieve deep semantic interactions and feature fusion between the comment text and event context.

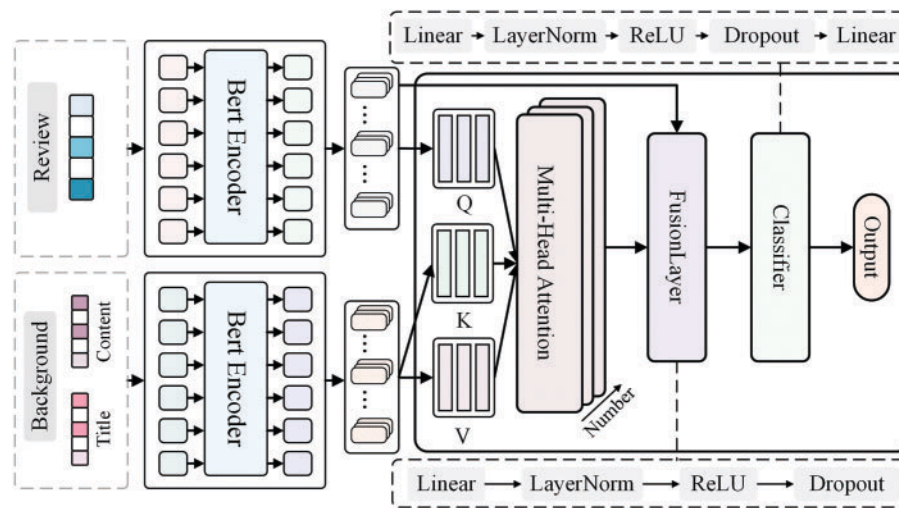


Figure 2: The architecture of the event-aware sarcasm detection model

3.2.1 Dual-Path BERT Encoder

To separately capture the semantic features of comment text and event context, the model employs two BERT encoders with shared parameters. BERT-base-chinese is selected as the base encoder to fully leverage its language knowledge acquired through pretraining on large-scale Chinese corpora. By employing a parameter-sharing mechanism, the model effectively handles two different types of text inputs while keeping the number of parameters manageable. The output of the BERT encoder provides contextual representations for each token, with a dimensionality of 768. These representations serve as the foundation for subsequent interaction and fusion layers.

BERT-base-chinese was adopted as the backbone encoder in this study owing to its demonstrated efficacy and extensive utilization in Chinese natural language processing. Pre-trained on large-scale Chinese corpora, it effectively captures the intricate linguistic, syntactic, and semantic properties unique to the Chinese language. In contrast to multilingual or English-centric models, BERT-base-chinese exhibits superior capability in addressing challenges such as word segmentation, idiomatic usage, and context-dependent semantics. Its established robustness and transferability across diverse Chinese NLP tasks render it a reliable foundation for sarcasm detection in Chinese social media. Moreover, the parameter-shared dual-path architecture facilitates computationally efficient feature distillation while maintaining generalization fidelity, as empirically demonstrated through systematic evaluations in [Section 4.3](#).

3.2.2 Multi-Head Attention Interaction Layer

To achieve deep semantic interactions between comment text and event context, the model incorporates an interaction layer based on an 8-head attention mechanism. In this layer, the representation of the comment text serves as the query (Q), while the representation of the event context serves as the key (K) and value (V). For the i -th attention head, the query, key, and value are denoted as Q_i , K_i , V_i , respectively. The formula for calculating the attention weights is:

$$\text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (1)$$

where d_k denotes the dimensionality of the query and key vectors (here, $d_k = 768$), and softmax normalizes the attention weights. By performing parallel computations across 8 attention heads, the model can capture semantic associations between the two types of text from multiple perspectives. This design enables the model to dynamically focus on relevant background information based on the content of the comment, thereby achieving a more accurate understanding of sarcastic intent. The output of this layer is then passed to the feature fusion layer for further processing.

Existing studies employing multi-head attention primarily focus on intra-textual semantic mining. In contrast, our model uniquely integrates event context as dynamic key-value pairs to establish cross-modal interactions between comment semantics and contextual triggers. Specifically, by utilizing comment representations as queries and event context as keys/values, the attention mechanism dynamically weights context-comment relevance across multiple semantic subspaces. This design addresses the implicit contextual dependencies in Chinese sarcasm, enabling more precise alignment between sarcastic expressions and their triggering contexts.

3.2.3 Feature Fusion Layer

After the attention interaction, the feature fusion layer integrates the semantic information of the comment text and the event context. This layer first concatenates the output of the attention interaction with the original comment representation, and then generates the fused feature representation through nonlinear transformations and layer normalization. Specifically, a fully connected layer is used to map the concatenated 1536-dimensional features (768×2) to 768 dimensions, followed by layer normalization, a ReLU activation function, and dropout (with a rate of 0.1). The use of layer normalization helps mitigate the issue of internal covariate shift in deep networks, while the dropout mechanism prevents overfitting by randomly deactivating some neurons. The fused feature representation is then used as input to the contrastive learning module.

3.2.4 Contrastive Learning Projection Head

To enhance the model's ability to represent sarcastic features, a contrastive learning mechanism is introduced. The projection head applies two layers of nonlinear transformation to map the 768-dimensional feature vector into 384-dimensional and 128-dimensional spaces sequentially. In this low-dimensional space, contrastive learning optimizes the feature representations by bringing the representations of the comment text and event context from the same sample closer together, while pushing apart the representations from different samples. This design helps the model learn more discriminative feature representations, improving its ability to identify sarcastic expressions. The optimized feature representations are then fed into the classifier for final prediction.

3.3 Loss Function Design

The training objective of the model consists of two parts: the classification loss L_{cls} and the contrastive loss L_{con} . To address the issue of class imbalance, the classification loss is constructed as a weighted combination of the standard cross-entropy (CE) loss and Focal Loss. First, the CE loss for supervised sarcasm classification is defined as:

$$L_{ce} = - \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (2)$$

where y_i is the true label of the i -th sample, and \hat{y}_i is the model's predicted probability for that label.

To further mitigate class imbalance, this paper introduces Focal Loss, which assigns higher weights to hard-to-classify samples, thereby enhancing the model's ability to recognize minority classes. The loss function is defined as follows:

$$L_{focal} = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (3)$$

here, p_t denotes the predicted probability for the true class, α is the class balancing factor, and γ is a hyperparameter that adjusts the focus on hard vs. easy samples.

The final classification loss, L_{cls} is then computed as a weighted sum of the CE loss and Focal Loss:

$$L_{cls} = (1 - \lambda_{focal})L_{ce} + \lambda_{focal}L_{focal} \quad (4)$$

where λ_{focal} is the weighting hyperparameter. Based on experimental results, α and γ are set to 1.55 and 2.4, respectively, while λ_{focal} is set to 0.66. This design improves the robustness of the model under class-imbalanced scenarios.

The contrastive loss optimizes feature learning by maximizing the mutual information between the comment text representation and the event context representation of the same sample. Let r_i and e_i denote the comment text representation and the event context representation of the i -th sample, respectively. The contrastive loss is defined as:

$$L_{con} = - \sum_{i=1}^N \log \frac{\exp(\text{sim}(r_i, e_i)/\tau)}{\sum_{j=1}^N \exp(\text{sim}(r_i, e_j)/\tau)} \quad (5)$$

here, $\text{sim}(r_i, e_j)$ computes the cosine similarity between vectors r_i and e_j , and τ (set to 0.07) is a temperature hyperparameter that scales the similarity scores.

$$L = L_{cls} + \lambda L_{con} \quad (6)$$

The contrastive loss weight $\lambda = 0.1$ balances the two objectives, determined through ablation studies (see [Section 4.3](#)).

3.4 Training Strategies

To effectively train the proposed model, the following strategies were adopted:

- **Batch size:** Considering the dataset size and GPU memory constraints, the batch size was set to 16, achieving a good balance between efficiency and performance.
- **Learning rate:** The initial learning rate was set to 2×10^{-5} , and a linear decay strategy was applied during training. This helps the model perform more refined optimization in the later stages of training.
- **Regularization:** The AdamW optimizer with built-in weight decay was used to implement L2 regularization, reducing the risk of overfitting. Additionally, a dropout probability $p = 0.1$ was applied after the fully connected and attention layers.
- **Early stopping:** To prevent overfitting, an early stopping mechanism was implemented. Training was automatically terminated if the F1 score on the validation set did not improve for 5 consecutive epochs.
- **Gradient clipping:** To avoid gradient explosion, gradient clipping was applied after each backpropagation step. The gradient norm was constrained within 1, ensuring stable updates.
- **Learning rate scheduling:** A linear learning rate scheduler was employed, which linearly increased the learning rate during the first 10% of the training steps and then linearly decreased it over the remaining steps. This facilitated rapid convergence in the early stages and stable optimization in the later stages.
- **Optimizer:** The AdamW optimizer was utilized to update the model parameters. The parameter update process follows the standard AdamW optimization formula, which consists of several steps.

First, the gradient of the loss function $L(\theta)$ with respect to the model parameters θ at iteration step t is computed as:

$$g_t = \nabla_{\theta} L(\theta_t) \quad (7)$$

The first-order moment estimate v_t and the second-order moment estimate v_t are updated as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (8)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (9)$$

Bias-corrected first-order and second-order moment estimates are then computed:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (10)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (11)$$

Finally, the model parameters are updated using the following equation:

$$\theta_t = \theta_{t-1} - \eta \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda \theta_{t-1} \right) \quad (12)$$

where η is the learning rate, λ represents the weight decay coefficient, β_1 and β_2 are the decay rates for the first- and second-order moments, respectively, and ϵ is a small smoothing term to prevent division by zero.

$$\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1e - 8, \lambda = 0.01 \quad (13)$$

Using the aforementioned strategies, the model was trained for 5 epochs on the training set. On the validation set, the optimal F1 score was achieved at the 3rd epoch. Therefore, the model parameters from this epoch were selected for evaluation on the test set.

3.5 Model Training Algorithm

To provide a holistic and procedural view of our methodology, this section synthesizes the preceding architectural, loss-related, and optimization components into a unified computational workflow, formally specified as Algorithm 1. The algorithm delineates the step-by-step process, commencing with dual-path encoding of text and context, proceeding through cross-modal interaction via multi-head attention, and calculating a hybrid loss that synergizes Focal Loss with a contrastive objective. The workflow culminates in parameter updates driven by the AdamW optimizer. This formal specification serves to enhance the methodological rigor and ensure the reproducibility of our proposed event-aware model.

Algorithm 1: Optimization algorithm for the event-aware sarcasm detection model

Require:

- Training dataset $D = (c_i, e_i, y_i) | i = 1, \dots, N$.
- Model parameters to be optimized, θ .
- Training hyperparameters: $\{\eta, B, E_{max}, G_{norm}\}$.
- Loss function hyperparameters: $\{\lambda, \tau, \lambda_{focal}, \alpha, \gamma\}$.

Ensure:

- The optimized model parameters, θ^* .

Procedure:

1. Initialize model parameters θ , AdamW optimizer (as per Eqs. (7)–(13)), and learning rate scheduler.
 2. **for** epoch = 1 **to** E_{max} **do**:
 3. **for** each batch $\{(c_b, e_b, y_b)\} \subset D$ **do**:
 - //Step 1: Dual-Path Parallel Feature Encoding
 - 4. $H_c \leftarrow \text{BERT_Encoder}(c_b; \theta)$ //Extract contextual representations for the comment:
 $H_c \in \mathbb{R}^{B \times L \times d}$
 - 5. $H_e \leftarrow \text{BERT_Encoder}(e_b; \theta)$ //Extract contextual representations for the event context:
 $H_e \in \mathbb{R}^{B \times L \times d}$
 - //Step 2: Cross-Modal Interaction via Multi-Head Attention
 - 6. $h_{c,cls} \leftarrow H_c[:, 0]$ //Extract the [CLS] representation for the comment
 - 7. $H_{inter} \leftarrow \text{MultiHeadAttn}(Q = H_c, K = H_e, V = H_e)$ //Perform deep cross-modal interaction (Eq. (1))
 - 8. $h_{inter,cls} \leftarrow H_{inter}[:, 0]$ //Extract the [CLS] representation after interaction
 - //Step 3: Feature Fusion and Classification
 - 9. $h_{fused} \leftarrow \text{FusionLayer}(\text{Concat}(h_{c,cls}, h_{inter,cls}))$
 - 10. $\hat{y} \leftarrow \text{Classifier}(h_{fused})$ //Predict classification logits: $\hat{y} \in \mathbb{R}^{B \times \text{NumClasses}}$
 - //Step 4: Projection for Contrastive Learning
 - 11. $h_{e,cls} \leftarrow H_e[:, 0]$ //Extract the [CLS] representation for the event context
 - 12. $z_c \leftarrow \text{ProjectionHead}(h_{c,cls})$ //Project comment features to the contrastive space:
 $z_c \in \mathbb{R}^{B \times d_{proj}}$
 - 13. $z_e \leftarrow \text{ProjectionHead}(h_{e,cls})$ //Project context features to the contrastive space:
 $z_e \in \mathbb{R}^{B \times d_{proj}}$
-

(Continued)

Algorithm 1 (continued)

```

//Step 5: Hybrid Loss Computation
14.  $L_{cls} \leftarrow (1 - \lambda_{focal})L_{CE}(\hat{y}, y_b) + \lambda_{focal} \cdot L_{Focal}(\hat{y}, y_b; \alpha, \gamma)$  //Classification Loss
(Eqs. (2)–(4))
15.  $L_{con} \leftarrow \text{ContrastiveLoss}(\{z_c, z_e\}; \tau)$  //Contrastive Loss (Eq. (5))
16.  $L_{total} \leftarrow L_{cls} + \lambda \cdot L_{con}$  //Weighted Total Loss (Eq. (6))
//Step 6: Gradient Backpropagation and Parameter Update
17.  $g \leftarrow \nabla_{\theta} L_{total}$  //Compute gradients of the total loss w.r.t.  $\theta$  (Eq. (7))
18.  $g_{clipped} \leftarrow \text{ClipGradientNorm}(g, \max\_norm = G_{norm})$  //Apply gradient clipping
19.  $\theta \leftarrow \text{AdamW\_Update}(\theta, g_{clipped}, \eta)$  //Update parameters via AdamW (Eqs. (8)–(12))
20. end for
21. end for
22. return The optimized model parameters,  $\theta^*$ 

```

4 Experiments and Results

To comprehensively evaluate the performance of the proposed event-aware Chinese sarcasm detection model based on the multi-head attention mechanism, a series of experiments were conducted on the dataset provided by the Chinese Sarcasm Computation Evaluation Task. The effectiveness of the model was validated through rigorous experimental setups and evaluation metrics, employing both quantitative and qualitative analyses. Experimental results demonstrate that the proposed method achieves significant improvements across multiple evaluation metrics, showcasing excellent sarcasm detection capabilities.

4.1 Datasets

To comprehensively evaluate the effectiveness and generalizability of our proposed model, this study utilizes two benchmark datasets. The primary dataset is the Chinese Sarcasm Computation Evaluation Task (CSCET) [32]. It comprises Sina Weibo comments from various domains, including politics, sports, entertainment, and economics, collected between October 2020 and May 2021. The dataset includes 50 Weibo topics and their associated comments, totaling 30,327 samples, of which approximately 65% (19,553) are sarcastic and 35% (10,774) are non-sarcastic. During data collection, topics with high controversy and sarcasm potential were carefully selected to ensure diverse sarcastic expressions within the samples. The sample distribution across topics is shown in Fig. 3.

For further validation, we also employ the Guanchazhe Chinese Sarcasm Dataset (GuanSarcasm) [18]. This dataset, collected from the Guanchazhe news website, contains 4972 user comments manually annotated for sarcasm, drawn from 720 news articles published between April and May 2019. Each comment is accompanied by contextual information such as the news title and article content. The dataset's diversity of topics and rich contextual cues makes it well-suited for evaluating the robustness of sarcasm detection models.

Dataset Example

To intuitively demonstrate the characteristics and annotation quality of the dataset, Table 1 presents a representative sample example. This sample is taken from the topic “U.S. Presidential Election,” which involves interactions between Biden and Trump. It includes the event context (topic content and topic title) and related comments. The example in Table 1 illustrates the sarcasm judgments of comments and their dependence on context.

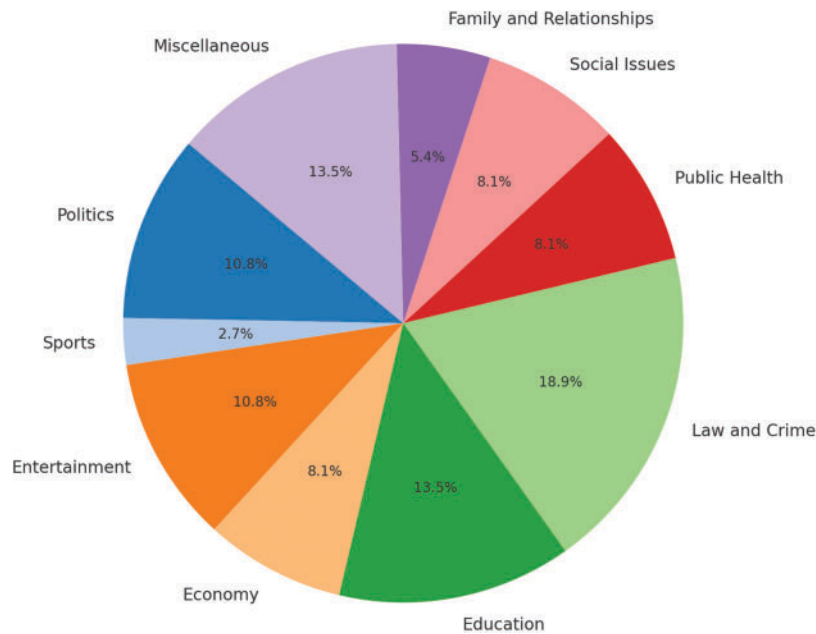


Figure 3: Distribution of samples across domains

Table 1: Example of a dataset sample

topicContent	topicTitle	Review	isSarcasm
On December 4th, local time, Biden stated during an interview that he believes Trump should attend his own inauguration. According to CNN, this was the first time Biden had participated in an interview alongside Vice President-elect Harris since the election. In addition, Biden mentioned, "This is his decision, it doesn't affect me, but I think it's for the good of the country."	"Biden Says He Hopes Trump Will Attend the Inauguration for the Good of the Country"	"Trump: Young man, you don't play by the rules."	1
		"Biden: I've extended the invitation, they don't have to come, just sending the RSVP is enough."	1
		"After all, it's still uncertain who the president is."	0

As shown in [Table 1](#), this example demonstrates the following characteristics of the dataset: First, each sample contains rich contextual information (topic content and topic title), which provides essential contextual support for understanding and judging the sarcasm in comments. Second, the comment text exhibits diverse language styles and expression forms, including explicit sarcasm (e.g., "Trump: Young man, you don't play by the rules.") as well as subtle satire or irony. Furthermore, the dataset annotations are detailed and capable of accurately distinguishing between sarcastic and non-sarcastic expressions, providing a solid data foundation for training effective models.

4.2 Experimental Setup

In this study, the dataset was divided using stratified sampling by topic to ensure consistent distribution across the training, validation, and test sets for each topic while maintaining a balanced ratio of sarcastic to non-sarcastic samples. Specifically, the comment texts and their corresponding event context information were first merged based on their topic identifiers (topicId) to construct the complete dataset. Subsequently, all distinct topics in the dataset were extracted, and the samples within each topic were stratified and divided into training, validation, and test sets in an 8:1:1 ratio. To ensure balanced sample distribution, stratified sampling was applied within each subset to maintain the original sarcastic and non-sarcastic sample ratio. This division strategy effectively evaluates the model's generalization ability across different topics while avoiding evaluation bias caused by inconsistent data distribution.

During the data preprocessing stage, systematic normalization was performed on the raw text, including the removal of irrelevant noise (e.g., extra spaces and punctuation), the addition of standardized symbols for special markers, and serialization processing. These steps not only improved data quality but also ensured efficient processing of textual inputs by the model.

During model training, the robust AdamW optimizer was used, and hyperparameter configurations were dynamically adjusted based on validation set performance to optimize the model's performance. To prevent overfitting during training, early stopping and weight decay techniques were employed. The final hyperparameter configurations, including learning rate and batch size, were determined through multiple experimental iterations.

4.3 Experimental Results and Analysis

This section provides an in-depth analysis of the experimental outcomes from three perspectives: (1) a comparative evaluation of the proposed model against a range of baseline and state-of-the-art methods; (2) an investigation of the training dynamics to assess performance trends, convergence behavior, and model stability; and (3) an ablation study to quantify the individual contributions of key architectural components. This comprehensive evaluation framework substantiates the effectiveness of the proposed approach for Chinese sarcasm detection and elucidates the roles of critical model design elements.

4.3.1 Comparison of Model Performance

Table 2 presents a comprehensive comparison of the proposed model against a range of baseline and state-of-the-art methods across two benchmark datasets: the Chinese Sarcasm Computation Evaluation Task (CSCET) and the GuanSarcasm dataset. The results unequivocally demonstrate that our model consistently surpasses all competing approaches on both datasets, as evidenced by superior performance in accuracy, precision, recall, and F1-score.

Table 2: Performance comparison of models on the test set

Model	Accuracy		Precision		Recall		F1	
	CSCET	GuanSarcasm	CSCET	GuanSarcasm	CSCET	GuanSarcasm	CSCET	GuanSarcasm
BERT Baseline	75.79%	75.57%	79.82%	81.40%	83.62%	66.29%	81.67%	73.07%
TextCNN	72.95%	75.19%	72.82%	75.09%	92.65%	75.38%	81.54%	75.24%
BiLSTM	71.61%	71.40%	72.37%	73.44%	90.57%	67.05%	80.45%	70.10%
BERT + Attention	75.43%	72.54%	75.53%	71.79%	91.58%	74.24%	82.79%	73.00%
CNN-Transformer	76.35%	73.67%	82.04%	75.30%	81.09%	70.45%	81.56%	72.80%
Our Model	79.55%	75.24%	83.91%	74.56%	84.54%	79.92%	84.22%	77.15%

On the CSCET dataset, our model achieves an accuracy of 79.55% and an F1-score of 84.22%, representing substantial improvements of 3.76 and 2.56 percentage points over the BERT Baseline, respectively. This performance underscores the efficacy of our event-aware architecture, which synergistically integrates comment text with event context through a multi-head attention mechanism. Notably, models that rely exclusively on comment text, such as traditional deep learning architectures and standard BERT variants, exhibit diminished capacity to capture the nuanced contextual dependencies intrinsic to sarcastic expressions.

The evaluation on the GuanSarcasm dataset further substantiates the robustness and generalizability of our approach. Our model attains the highest F1-score of 77.15%, outperforming all baselines, including the CNN-Transformer hybrid architecture. This result underscores the adaptability of our method to diverse domains and comment styles, affirming its applicability beyond a single data source.

In comparison to conventional models such as TextCNN and BiLSTM, which tend to achieve elevated recall at the expense of precision and overall F1-score, our model maintains a judicious balance between these metrics. This balance enables more reliable discrimination between sarcastic and non-sarcastic instances, mitigating the prevalence of false positives that often afflict traditional approaches.

The inclusion of the CNN-Transformer hybrid model in the comparative analysis reveals inherent limitations of static feature fusion approaches. While the CNN-Transformer leverages convolutional operations for local feature extraction and self-attention for global sequence modeling, it fundamentally relies on static feature concatenation. In contrast, the multi-head attention mechanism employed in this study enables dynamic alignment between comment semantics and evolving contextual information, thereby facilitating more precise modeling of context-dependent sarcastic expressions. Experimental results demonstrate that the proposed model achieves statistically significant improvements over this strong baseline on both datasets, underscoring its effectiveness in capturing intricate semantic interactions and adapting to complex real-world scenarios.

To comprehensively evaluate the effectiveness of the proposed model, [Table 3](#) presents comparative results with state-of-the-art approaches: the CCSD model, EKPL model, TISD model, and RDVI model. The experimental results demonstrate that our model achieves superior performance across key metrics, including accuracy and F1-score, thereby validating its effectiveness.

Table 3: Performance comparison with state-of-the-art models

Model	Accuracy	F1
CCSD	71.94%	71.23%
EKPL	73.77%	74.04%
TISD	73.18%	72.73%
RDVI	79.54%	79.41%
Our Model	79.55%	84.22%

The CCSD model excels in integrating contextual and commonsense features, particularly for sarcastic expressions requiring commonsense reasoning. However, it exhibits limited capability in dynamic semantic modeling of event backgrounds, hindering its adaptation to complex social media contexts.

- The EKPL model demonstrates strong adaptability in low-resource scenarios through entity knowledge and prompt learning. Nevertheless, its heavy reliance on predefined prompt templates constrains scalability when processing large-scale datasets.

- The TISD model advances intra- and inter-sentential context modeling by capturing topic information and contextual incongruity. However, its focus on noun-centric opinion chunks restricts its ability to explore broader semantic units and deep semantic relationships.
- The RDVI model leverages document retrieval and prompt learning to enhance semantic understanding, showing advantages in detecting sarcasm requiring extensive domain knowledge. Yet, its performance heavily depends on the quality of retrieved documents, and potential noise from irrelevant or redundant snippets may degrade detection accuracy.

Overall, our model outperforms all baseline approaches across evaluation metrics, confirming its effectiveness and superiority in complex sarcasm detection tasks. This result further highlights our model's capability in multimodal feature integration and deep semantic pattern extraction.

4.3.2 Model Training Process

To comprehensively elucidate the performance evolution during model training, the trajectories of key evaluation metrics across training epochs are depicted in Fig. 4.

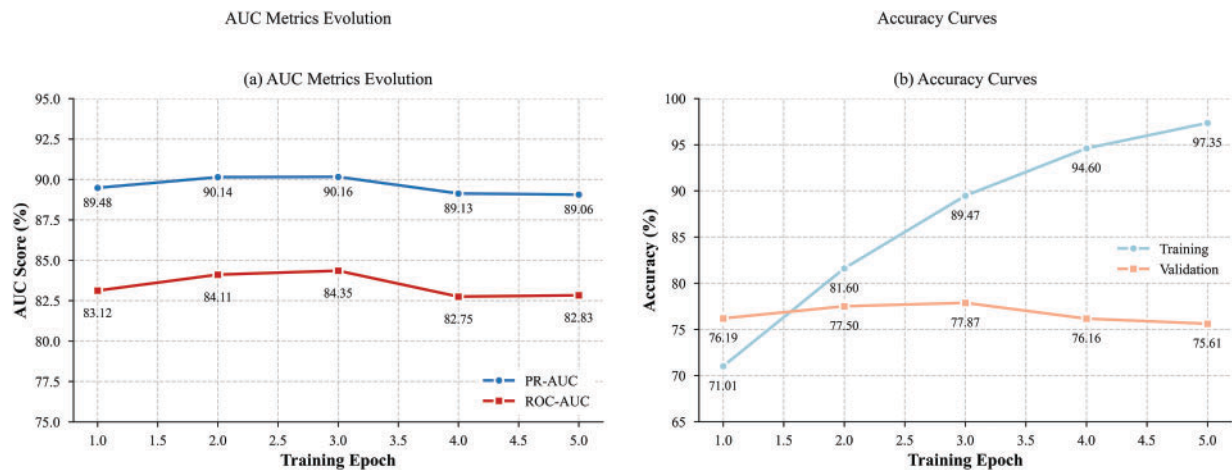


Figure 4: Performance metrics evolution during model training

Fig. 4a illustrates the trends of PR-AUC and ROC-AUC on the validation set. It can be observed that the model achieves peak values of PR-AUC (0.9016) and ROC-AUC (0.8435) as early as the second epoch, after which these metrics exhibit minor fluctuations but remain generally stable, indicating the model's robust capacity for capturing salient feature representations. Notably, PR-AUC consistently surpasses ROC-AUC, further substantiating the model's sensitivity and discriminative power for minority classes (i.e., sarcastic samples) under class-imbalanced conditions.

Fig. 4b further elucidates the dynamic changes in accuracy for both the training and validation sets. The training accuracy increases steadily from 71.01% to 97.35%, whereas the validation accuracy remains within the range of 75.61% to 77.87%. The pronounced gap between the two indicates a certain degree of overfitting. To mitigate this issue, Dropout regularization (dropout rate = 0.15), L2 weight decay (weight decay = 0.01), and the Focal Loss function were incorporated into the model architecture to enhance the model's focus on hard-to-classify samples and improve generalization capability. Experimental results demonstrate that, following these enhancements, the model exhibits improved F1 scores and PR-AUC on the validation set, with particularly notable gains in the identification of minority class samples. Nevertheless,

the performance gap between the training and validation sets persists. Future work will focus on exploring more effective regularization mechanisms and architectural optimization strategies to further enhance the model's robustness and generalization in real-world applications.

4.3.3 Ablation Experiments

To rigorously assess the contributions of individual architectural components and the qualitative impact of contextual information, a comprehensive suite of ablation experiments was conducted. The evaluation encompassed five distinct configurations: (1) the complete model, (2) removal of the multi-head attention mechanism, (3) exclusion of event context, (4) omission of contrastive learning, and (5) introduction of noisy event context, wherein the background information was randomly substituted with unrelated content to simulate real-world scenarios of context corruption or ambiguity. The results are detailed in [Table 4](#).

Table 4: Results of the ablation experiments

Model	Accuracy	Precision	Recall	F1
Our Model	79.55%	83.91%	84.54%	84.22%
Without Multi-head Attention	75.83%	77.66%	87.78%	82.41%
Without Event Context	76.19%	79.85%	84.38%	82.05%
Without Contrastive Learning	78.34%	84.84%	80.88%	82.81%
Noisy Event Context	72.87%	78.10%	80.57%	79.32%

The architectural design was validated through a comprehensive suite of ablation experiments. The findings empirically justify the proposed methodology by dissecting the contribution of each key component:

- **Multi-head Attention:** The exclusion of the multi-head attention mechanism precipitates a substantive decline in performance, with the F1-score dropping from 84.22% to 82.41%. This degradation extends beyond a mere numerical reduction, underscoring the mechanism's critical function in facilitating dynamic semantic alignment. Without it, the model would be constrained to a static fusion of features, failing to judiciously weigh the relevance between specific tokens in the comment and salient cues within the event context. This finding confirms that multi-head attention is the linchpin for modeling the intricate, non-linear interdependencies that are crucial for extracting nuanced sarcastic features.
- **Event Context:** The complete removal of event context validates the central premise of our research—that sarcasm in social media is profoundly context-dependent. The resulting drop in F1-score to 82.05% demonstrates that, in isolation, the comment text is often semantically ambiguous. This experiment provides unequivocal evidence that integrating event-specific background is not merely an auxiliary enhancement but a fundamental necessity for resolving such ambiguities and achieving accurate interpretation.
- **Contrastive Learning:** Ablating the contrastive learning module (F1-score: 82.81%) highlights its efficacy in sculpting a more discriminative feature space. By refining feature representations, it sharpens the model's ability to resolve ambiguous cases with high lexical overlap, thereby enhancing overall robustness.
- **Noisy Event Context:** Most revealingly, introducing noisy context triggers the most severe performance degradation (F1-score: 79.32%), a result worse than excluding context entirely. This demonstrates that the model relies on the fidelity and relevance of context, not its mere presence. The penalty for incongruent

information confirms the model is truly event-aware, actively forging semantic links rather than using superficial cues.

In summary, these findings collectively substantiate our architectural choices. The synergy between multi-head attention, event context, and contrastive learning confirms the methodological soundness of our event-aware framework.

4.3.4 Analysis of the Comment Attention Module

Fig. 5 presents the attention distribution learned by the multi-head attention mechanism within the comment “专家真懂, 就是没用” (The experts really get it; they’re just useless). The tokens “没” (méi, meaning “not”) and “用” (yòng, meaning “useful”) demonstrate the strongest mutual attention, highlighting the model’s emphasis on the core negative sentiment that conveys sarcasm. Additionally, the token “就是” receives moderate attention, functioning as a semantic connector that reinforces the contrast and irony within the sentence. In comparison, tokens such as “真” (zhēn, meaning “really”) and “懂” (dǒng, meaning “understand”) exhibit weaker attention, indicating their relatively minor role in the overall sarcastic expression. This attention pattern suggests that the model is capable of identifying and prioritizing key elements central to the sarcastic intent, even in the absence of explicit contextual information. The visualization further demonstrates the effectiveness of the multi-head attention mechanism in capturing subtle semantic cues and internal contrasts that are characteristic of sarcastic language. Such interpretability not only enhances our understanding of the model’s decision-making process but also provides evidence that the model can focus on linguistically meaningful features, thereby improving the reliability of sarcasm detection in complex social media discourse.

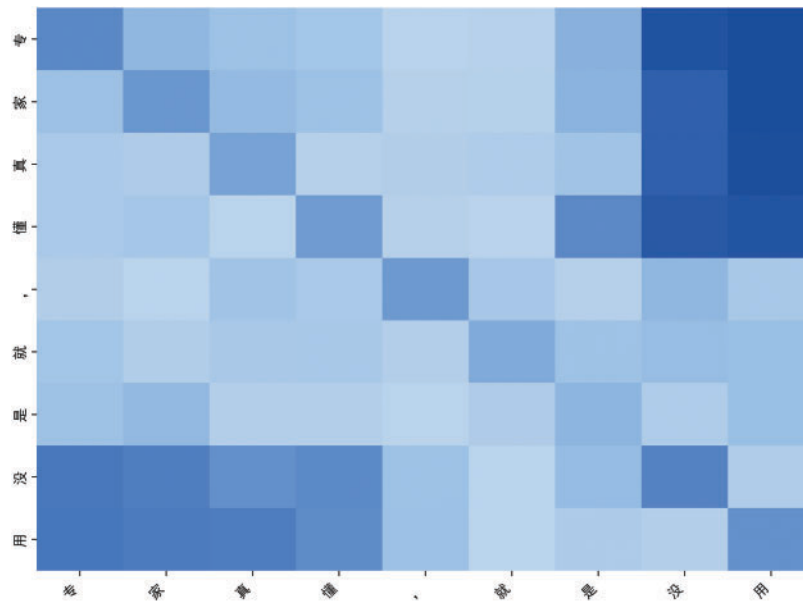


Figure 5: Attention heatmap of sarcastic comment

4.3.5 Qualitative Analysis of Misclassified Cases

Analysis of misclassified samples reveals key limitations in the model’s handling of event-context dependencies in Chinese social media. For instance, the comment “大家别骂了, 留学生都原谅学校了” (Everyone, stop the criticism; the international students have already forgiven the school) was misclassified

as non-sarcastic (confidence: 0.59). This exemplifies implicit sarcasm conveyed through ostensibly positive sentiment, where the underlying critique requires integration of contextual information about the international students' incident. The model failed to capture how this comment ironically critiques the institutional response, revealing limitations in contextual reasoning. Conversely, “不在北京的表示雨我无瓜” (Those not in Beijing state it's 'irrelevant to me') was incorrectly classified as sarcastic (confidence: 0.82), likely due to the model's limited understanding of internet slang. The phrase “雨我无瓜” (yǔ wǒ wú guā) is a form of creative wordplay, phonetically similar to the standard expression “与我无关” (yǔ wǒ wú guān), which translates to “it has nothing to do with me.” This type of slang is used for humorous effect and does not inherently convey sarcasm. These cases demonstrate that, despite leveraging multi-head attention for event-context integration, the model still struggles with nuanced expressions and culturally embedded references. Future work should focus on incorporating external knowledge and improving the modeling of internet vernacular to enhance contextual understanding in Chinese sarcasm detection.

5 Conclusion

This paper proposes an event-aware model for Chinese sarcasm detection, leveraging multi-head attention and contrastive learning to integrate comment text and event context. The model significantly enhances contextual modeling and semantic interaction, outperforming existing methods across various evaluation metrics (e.g., accuracy, precision, recall, F1-score, and AUC), thus demonstrating the critical role of event context in sarcasm detection. In low-resource scenarios, it exhibits strong generalization and robustness, highlighting the effectiveness of contrastive learning in feature discrimination. The model's key strengths lie in its multi-head attention mechanism, which captures deep semantic associations, and its event context modeling approach, which improves understanding of context-dependent sarcastic expressions. Ablation experiments further validate the importance of these mechanisms, while the contrastive learning strategy maximizes mutual information to refine feature representation.

Acknowledgement: The authors are grateful to all the editors and potential reviewers for their time and effort in evaluating this work and thank all the colleagues and collaborators who have contributed to this research.

Funding Statement: This work was granted by Qin Xin Talents Cultivation Program (No. QXTCP C202115), Beijing Information Science & Technology University; the Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing Fund (No. GJJ-23), National Social Science Foundation, China (No. 21BTQ079).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Kexuan Niu and Xiameng Si; methodology, Kexuan Niu; validation, Kexuan Niu, Xiameng Si, and Xiaojie Qi; formal analysis, Kexuan Niu; investigation, Kexuan Niu; resources, Kexuan Niu; data curation, Kexuan Niu and Xiaojie Qi; writing—original draft preparation, Kexuan Niu; writing—review and editing, Kexuan Niu and Xiameng Si; visualization, Kexuan Niu; supervision, Xiameng Si and Haiyan Kang; project administration, Xiameng Si; funding acquisition, Xiameng Si and Haiyan Kang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Joshi A, Bhattacharyya P, Carman MJ. Automatic sarcasm detection: a survey. *ACM Comput Surv.* 2017;50(5):1–22. doi:10.1145/3124420.

2. Chen W, Lin F, Li G, Liu B. A survey of automatic sarcasm detection: fundamental theories, formulation, datasets, detection methods, and opportunities. *Neurocomputing*. 2024;578(3):127428. doi:10.1016/j.neucom.2024.127428.
3. Farabi S, Ranasinghe T, Kanojia D, Kong Y, Zampieri M. A survey of multimodal sarcasm detection. In: Larson K, editor. *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*; 2024 Aug 3–9; Jeju, Republic of Korea. doi:10.24963/ijcai.2024/887.
4. Aboobaker J, Ilavarasan E. A survey on sarcasm detection and challenges. In: *Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*; 2020 Mar 6–7; Coimbatore, India. doi:10.1109/ICACCS48705.2020.9074163.
5. Bouazizi M, Ohtsuki T. Opinion mining in Twitter: how to make use of sarcasm to enhance sentiment analysis. In: *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*; 2015 Aug 25–28; Paris, France. doi:10.1145/2808797.2809350.
6. Kumar A, Garg G. Systematic literature review on context-based sentiment analysis in social multimedia. *Multimed Tools Appl*. 2020;79(21):15349–80. doi:10.1007/s11042-019-7346-5.
7. Fan X, Liu J, Liu J, Tuerxun P, Deng W, Li W. Identifying hate speech through syntax dependency graph convolution and sentiment knowledge transfer. *IEEE Access*. 2024;12:2730–41. doi:10.1109/ACCESS.2023.3347591.
8. Liang B, Lin Z, Qin B, Xu R. Topic-oriented sarcasm detection: new task, new dataset and new method. In: *Proceedings of the 21st Chinese National Conference on Computational Linguistics*; 2022 Oct 14–16; Nanchang, China.
9. Gong X, Zhao Q, Zhang J, Mao R, Xu R. The design and construction of a Chinese sarcasm dataset. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*; 2020 May 11–16; Marseille, France.
10. Liu P, Chen W, Ou G, Wang T, Yang D, Lei K. Sarcasm detection in social media based on imbalanced classification. In: Li F, Li G, Hwang S, Yao B, Zhang Z, editors. *Web-age information management*. Berlin/Heidelberg, Germany: Springer; 2014. p. 459–71. doi:10.1007/978-3-319-08010-9_49.
11. Devlin J, Chang MW, Lee K, Toutanova K, Burstein J, Doran C, et al. 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)*; 2019 Jun 2–7; Minneapolis, MN, USA. doi:10.18653/v1/N19-1423.
12. Musleh D, Alkhwaja I, Alkhwaja A, Alghamdi M, Abahussain H, Alfawaz F, et al. Arabic sentiment analysis of YouTube comments: NLP-based machine learning approaches for content evaluation. *Big Data Cogn Comput*. 2023;7(3):127. doi:10.3390/bdcc7030127.
13. González-Ibáñez R, Muresan S, Wacholder N, Lin D, Matsumoto Y, Mihalcea R. Identifying sarcasm in Twitter: a closer look. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*; 2011 Jun 19–24; Portland, OR, USA.
14. Poria S, Cambria E, Hazarika D, Vij P, Matsumoto Y, Prasad R. A deeper look into sarcastic tweets using deep convolutional neural networks. In: *Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics*; 2016 Dec 11–16; Osaka, Japan.
15. Ghosh A, Veale T, Palmer M, Hwa R, Riedel S. Magnets for sarcasm: making sarcasm detection timely, contextual and very personal. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*; 2017 Sep 9–11; Copenhagen, Denmark. doi:10.18653/v1/D17-1050.
16. Ghosh A, Veale T, Balahur A, van der Goot E, Vossen P, Montoyo A. Fracking sarcasm using neural network. In: *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*; 2016 Jun 16; San Diego, CA, USA. doi:10.18653/v1/W16-0425.
17. Kalaivani A, Thenmozhi D, Klebanov BB, Shutova E, Lichtenstein P, Muresan S, et al. Sarcasm identification and detection in conversation context using BERT. In: *Proceedings of the Second Workshop on Figurative Language Processing*; 2020 Jul 9; Online. doi:10.18653/v1/2020.figlang-1.10.
18. Wen Z, Gui L, Wang Q, Guo M, Yu X, Du J, et al. Sememe knowledge and auxiliary information enhanced approach for sarcasm detection. *Inf Process Manag*. 2022;59(3):102883. doi:10.1016/j.ipm.2022.102883.

19. Peled L, Reichart R, Barzilay R, Kan MY. Sarcasm SIGN: interpreting sarcasm with sentiment based monolingual machine translation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics; 2017 Jul 30–Aug 4; Vancouver, BC, Canada. doi:10.18653/v1/P17-1155.
20. Hazarika D, Poria S, Gorantla S, Cambria E, Zimmermann R, Mihalcea R, et al. CASCADE: contextual sarcasm detection in online discussion forums. In: Proceedings of the 27th International Conference on Computational Linguistics; 2018 Aug 21–24; Santa Fe, NM, USA.
21. Cai Y, Cai H, Wan X, Korhonen A, Traum D, Màrquez L. Multi-modal sarcasm detection in Twitter with hierarchical fusion model. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; 2019 Jul 28–Aug 2; Florence, Italy. doi:10.18653/v1/P19-1239.
22. Liu Q, Li B, Huang Z. CCSD: topic-oriented sarcasm detection. *Comput Sci*. 2024;51(9):310–8. doi:10.11896/jsj.kx.230600217.
23. Zhou Y, Zhang S, Wang C, Wang Y, Wang X, Li K. Topic-oriented sarcasm detection via entity knowledge-based prompt learning. *Comput Sci Inf Syst*. 2025;22(1):33–57. doi:10.2298/CSIS240320067Z.
24. Duan Y, Zhang S, Qian L, Wen H, Ding Y, Ge C. A sarcasm detection model incorporating topic information and contextual incongruity information. *Comput Eng Appl*. 2024;1–9. (In Chinese).
25. Wen Z, Wang R, Chen S, Wang Q, Ding K, Liang B, et al. RDVI: a retrieval-detection framework for verbal irony detection. *Electronics*. 2023;12(12):2673. doi:10.3390/electronics12122673.
26. Diao Y, Lin H, Yang L, Fan X, Chu Y, Xu K, et al. A multi-dimension question answering network for sarcasm detection. *IEEE Access*. 2020;8:135152–61. doi:10.1109/ACCESS.2020.2967095.
27. Kumar A, Narapareddy VT, Aditya Srikanth V, Malapati A, Neti LBM. Sarcasm detection using multi-head attention based bidirectional LSTM. *IEEE Access*. 2020;8:6388–97. doi:10.1109/ACCESS.2019.2963630.
28. Tay Y, Luu AT, Hui SC, Su J, Gurevych I, Miyao Y. Reasoning with sarcasm by reading in-between. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics; 2018 Jul 15–20; Melbourne, Australia. doi:10.18653/v1/P18-1093.
29. Lin Z, Feng M, dos Santos CN, Yu M, Xiang B, Zhou B, et al. A structured self-attentive sentence embedding. *arXiv:1703.03130*. 2017.
30. Babanejad N, Agrawal A, An A, Papagelis M, Jurafsky D, Chai J, et al. A comprehensive analysis of preprocessing for word representation learning in affective tasks. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020 Jul 5–10; Online. doi:10.18653/v1/2020.acl-main.514.
31. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4–9; Long Beach, CA, USA.
32. pjzj220113. Chinese sarcasm calculation [Internet]. GitHub; 2021 [cited 2024 Mar 8]. Available from: <https://github.com/pjzj220113/chinese-sarcasm-calculation>.