technologies

MDPI

Article

Dual-Channel ADCMix–BiLSTM Model with Attention Mechanisms for Multi-Dimensional Sentiment Analysis of Danmu

Wenhao Ping [1], Zhihui Bai [1,*] and Yubo Tao [2]

[1] School of Information Engineering, Dalian University, Dalian 116622, China; gyou1t4@gmail.com
[2] State Key Laboratory of Computer-Aided Design & Computer Graphics, Zhejiang University, Hangzhou 310058, China; taoyubo@cad.zju.edu.cn
* Correspondence: baizhihui0810@126.com

Abstract

Sentiment analysis methods for interactive services such as Danmu in online videos are challenged by their colloquial style and diverse sentiment expressions. For instance, the existing methods cannot easily distinguish between similar sentiments. To address these limitations, this paper proposes a dual-channel model integrated with attention mechanisms for multi-dimensional sentiment analysis of Danmu. First, we replace word embeddings with character embeddings to better capture the colloquial nature of Danmu text. Second, the dual-channel multi-dimensional sentiment encoder extracts both the high-level semantic and raw contextual information. Channel I of the encoder learns the sentiment features from different perspectives through a **mix**ed model that combines the benefits of self-**A**ttention and **D**ilated **C**NN (ADCMix) and performs contextual modeling through bidirectional long short-term memory (BiLSTM) with attention mechanisms. Channel II mitigates potential biases and omissions in the sentiment features. The model combines the two channels to erase the fuzzy boundaries between similar sentiments. Third, a multi-dimensional sentiment decoder is designed to handle the diversity in sentiment expressions. The superior performance of the proposed model is experimentally demonstrated on two datasets. Our model outperformed the state-of-the-art methods on both datasets, with improvements of at least 2.05% in accuracy and 3.28% in F1-score.

Keywords: multi-dimensional sentiment analysis; Danmu; attention mechanism; bidirectional long short-term memory; dual-channel

Academic Editor: Juan Gabriel Avina-Cervantes

Received: 5 June 2025
Revised: 29 July 2025
Accepted: 1 August 2025
Published: 10 August 2025

Citation: Ping, W.; Bai, Z.; Tao, Y. Dual-Channel ADCMix–BiLSTM Model with Attention Mechanisms for Multi-Dimensional Sentiment Analysis of Danmu. *Technologies* **2025**, *13*, 353. https://doi.org/10.3390/technologies13080353

1. Introduction

As online video platforms proliferate, video viewers have increasingly expressed their opinions and insights through Danmu. Danmu synchronizes the user's comments with video playback time, scrolling them across the screen to display the user's interactions in real time, as shown in Figure 1. Danmu fosters a co-viewing experience among audiences, enhancing the enjoyment of the video content. As Danmu is closely synchronized with the timeline of the video content, it can accurately reflect the user's emotions and reactions at specific moments.

Sentiment analysis [1] is a core component of Danmu research because sentiment plays a crucial role in society and profoundly influences the decision-making, communication, and behavior styles of people. Danmu sentiment analysis aims to extract and analyze users'

sentiment states from Danmu texts. Multi-dimensional sentiment analysis can more accurately represent complex human emotional states such as happiness, anger, sadness, and fear than traditional binary (positive or negative) sentiment analysis. Multi-dimensional sentiment analysis categorizes sentiments in a fine-grained manner, allowing deep explorations of a user's emotional fluctuations and psychological states, thus providing a more comprehensive perspective on sentiment analysis than its binary counterpart. This also provides valuable emotional feedback for video creators and offers insights into public opinion trends, supporting applications in media optimization and social governance.

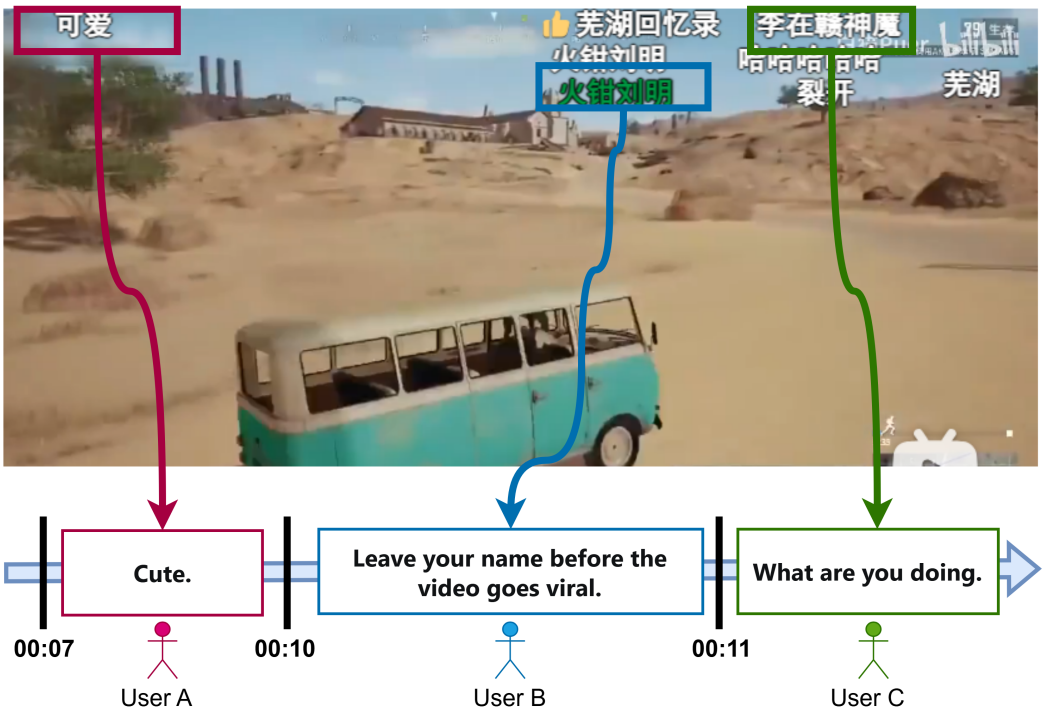


Figure 1. Screenshot of Danmu overlaid on video (Video source: https://www.bilibili.com/video/BV1Ap4y1i79S (accessed on 25 July 2025)). Viewers can express their emotions and interact through Danmu during video playback, enhancing the viewing experience.

In a typical current Danmu sentiment analysis, the text is represented with word embedding [2]. The sentiment features within the data are then analyzed and explored by leveraging the powerful feature learning capabilities of neural network models [3,4]. Nevertheless, multi-dimensional Danmu sentiment analysis faces several serious challenges. First, Danmu is typically colloquial and contains rich cultural references and internet slang, which complicates vocabulary processing and feature extraction. The second main problem is potential misjudgments caused by the indistinct boundaries between similar sentiments in Danmu. For example, "服了服了" (I am speechless) can express anger or joy, but the real sentiment in Danmu is joy. Third, the emotional expressions of users are frequently multifaceted and nuanced, conveyed through rhetorical devices such as antithesis and sarcasm. Therefore, the multi-dimensional sentiment characteristics of Danmu are difficult to capture.

To address the aforementioned challenges, this study proposes a dual-channel model integrated with attention mechanisms for multi-dimensional sentiment analysis of Danmu. Rather than extracting typical word-level features, the proposed ADCMix–BiLSTM method first extracts the char-level features using char embedding, which captures the inherent features of Danmu's colloquial style, rich cultural references, and widespread use of internet slang. In addition, the dual-channel multi-dimensional sentiment encoder enhances the extraction of multi-dimensional sentiment features. By separately extracting the high-

level semantic information and raw contextual information, the encoder erases the fuzzy boundaries between similar sentiments. In particular, Channel I learns the sentiment features from different perspectives through our proposed ADCMix, which effectively integrates self-attention and convolutional pathways for sentiment analysis. To accurately capture the high-level sentiment features, channel I then performs contextual modeling through its BiLSTM network with attention mechanisms. Meanwhile, channel II effectively mitigates bias and omission of sentiment features. To better integrate sentiment differences, we finally design a multi-dimensional sentiment decoder that effectively solves the diversity problem in sentiment expression. Experimental results demonstrate the higher performance of the proposed model than of several existing state-of-the-art (SOTA) methods.

The contributions of our paper are summarized below:

(1) This paper presents a dual-channel ADCMix–BiLSTM model integrated with attention mechanisms for multi-dimensional sentiment analysis of Danmu texts. Unlike traditional models with limited ability for processing short texts and informal language, the proposed architecture can effectively capture the complex, multi-dimensional sentiment information in Danmu.
(2) The ADCMix structure is designed to capture the high-level semantic information of multi-dimensional sentiment features in Danmu. The dilated CNN in the ADCMix component learns the sentiment features from multiple perspectives, and the self-attention mechanism with a convolutional module learns the sentiment features from different perspectives.
(3) The proposed model is validated through comprehensive experiments on two self-constructed datasets, *Romance of The Three Kingdoms* and *The Truman Show*.

2. Related Work

2.1. Danmu Sentiment Analysis

The continuous development of natural language processing (NLP) has enabled comprehensive studies on Danmu sentiment analysis, allowing the full exploitation of Danmu as a potentially valuable data resource.

For instance, Li et al. [3] proposed an innovative sentiment word embedding (EWE) method and developed a Deep Coupled Video and Danmu Neural Network (DCVDN) model that integrates video and Danmu for sentiment analysis. Although the DCVDN model yielded promising results in Danmu sentiment analysis, emerging Danmu contents are not always covered in the existing sentiment corpora; moreover, sentiment lexicons are not easily transformed into accurate and effective sentiment embedding. To better capture the colloquialism and diversity of Danmu, Wang et al. [4] proposed an enhanced sentiment analysis approach based on a BiLSTM model that classifies four sentiment dimensions (joy, anger, sadness, and happiness). However, the four-dimensional sentiment model cannot accurately capture and fully express the rich emotional nuances of Danmu sentiment, characterized by complex multi-dimensional features. Zhang and Ren [5] fine-tuned a sentiment classification model based on bidirectional encoder representation (BERT) to quantify the emotion polarity of Danmu. They also smoothed the emotion sequence utilizing techniques such as comprehensive weights and subsequently clustered the collected data using a shape-based distance (SBD)–K-shape method. Their method provides a reference for Danmu sentiment analysis using deep learning and transfer learning. Although the BERT-based method can attenuate the effects of irregular noise and temporal phase shifts, it cannot precisely capture multi-dimensional sentiments. In addition, smoothing can obscure subtle emotional fluctuations in the emotion sequences.

Research on Danmu sentiment analysis has notably advanced, providing increasingly accurate recognition and extraction of sentiment information. Nevertheless, the existing

sentiment dimensions do not encompass the intricacy and heterogeneity of Danmu sentiments. The major remaining challenges are the limitations of emotion corpora, difficulties in converting emotion lexicons into effective emotion embedding, and insufficient multi-dimensional capture of sentiments. When emotions expressed in Danmu are complex and intertwined, the existing methods cannot easily and comprehensively capture the full sentiment spectrum. To address these limitations, this paper introduces a seven-dimensional sentiment analysis approach that aims to comprehensively capture the intricate Danmu sentiment. Specifically, the conventional word embedding is substituted with a char embedding, enabling more accurate identification and extraction of the inherent multi-dimensional sentiment features in Danmu. Furthermore, the proposed ADCMix–BiLSTM model captures the multi-dimensional sentiment features of Danmu from different perspectives, enabling a comprehensive and accurate analysis of sentiment alterations.

As Danmu sentiment analysis remains a relatively underexplored area, we draw on insights from related short-text sentiment analysis, including studies on tweets [6], online comments [7], and Weibo posts [8]. These texts, like Danmu, are typically informal, contain slang, and are characterized by short, fragmented linguistic forms. Therefore, the following section reviews research on sentiment analysis of such short texts.

2.2. Short-Text Sentiment Analysis

Short-text sentiment analysis, defining the task of extracting and analyzing the sentiment states of users from textual data, plays a crucial role in NLP applications such as healthcare [9], public opinion analysis [10], and e-commerce [11]. Short-text sentiment analysis methods are currently divided into three main categories: machine learning-based approaches, deep learning-based approaches, and methods based on pretrained large language models.

Machine learning-based approaches: Sentiment features in text are commonly identified using traditional machine learning algorithms such as support vector machine (SVM) and naive Bayes (NB). Singh et al. [12] found that NB and logistic regression outperform deep learning methods in sentiment analysis of online video comments. Suresh et al. [13] combined an enhanced vector space model and a hybrid SVM classifier into a hybrid model that provides high performance but relies heavily on high-quality sentiment lexicons and well-designed feature engineering. However, as traditional machine learning methods require handcrafted feature extraction, they cannot easily handle grammatical complexity and contextual dependencies. The intricate semantic expressions and diverse sentiment contents in Danmu are especially challenging for these methods.

Deep learning-based approaches: Models for deep learning-based sentiment analysis have markedly progressed and are classified into two categories: methods using recurrent neural network (RNN) variants and methods combining hybrid neural network models. RNNs effectively capture contextual information and handle complex syntactic structures. Their improved variants, such as BiLSTM and Bidirectional Gated Recurrent Unit (BiGRU) [14–16], can better learn textual contextual features than traditional RNNs and have increased the effectiveness of sentiment analysis. Other researchers have combined different neural networks [17–19] for sentiment analysis. In particular, the local feature-extraction capabilities of convolutional neural networks (CNNs) and the global contextual dependency modeling abilities of RNNs [20] can be co-leveraged to enhance sentiment classification performance. For example, Gan et al. [21] proposed a scalable multi-channel dilated CNN–BiLSTM model for sentiment analysis. By virtue of the dilated CNN, which effectively extracts multi-scale features with different expansion rates, this model flexibly extracts the original and advanced contextual information through a multi-channel structure. Hybrid models combining the local feature-extraction capability of CNNs with the

sequential-dependency capturing ability of RNNs provide deep feature learning of both the temporal and semantic structures of the text. However, the difficulty of integrating the strengths of both models can limit the performance of hybrid models on specific tasks.

Pre-trained large language model-based approaches: Pretrained large-scale language models, such as BERT and its subsequent versions (e.g., Ernie–Tiny [22], RoBERTa-wwm-ext [23], and BGE-large-en [24]), acquire rich linguistic representations through unsupervised pre-training. As these models effectively capture the contextual information and semantic relationships, they have become essential tools in the NLP field. For example, Kumar et al. [25] proposed combining BERT, which leverages transfer learning for semantic understanding, with CNN for local feature extraction, achieving improved sentiment classification performance. Meanwhile, the Ens-RF-BERT model [26] proposed in recent research combines the strengths of Random Forest and BERT in an ensemble voting framework to outperform traditional classifiers.

Although deep learning models and pretrained large-scale language models have performed well in certain domains, they cannot comprehensively and accurately extract multi-dimensional sentiment features from Danmu. Replacing traditional word-level features with char embedding, we extract char-level features and capture the high-level sentiment features of Danmu texts using our ADCMix–BiLSTM model. Combining the powerful language-understanding capabilities of pretrained large language models with the contextual dependency handling capabilities of deep learning models, the ADCMix–BiLSTM model enhances the capture of different semantic layers and sentiment details in sentiment analysis.

3. Materials and Methods

The proposed dual-channel ADCMix–BiLSTM model integrated with attention mechanisms is composed of four main modules: char embedding, a dual-channel multi-dimensional sentiment encoder, a multi-dimensional sentiment decoder, and a multi-dimensional sentiment classifier (see Figure 2).

To clarify the main contribution of each module, we summarize them as follows:

- **Char Embedding**: Utilizes the BGE-large-en model to encode Danmu at the character level, capturing fine-grained semantics and handling the flexibility of Chinese character composition.
- **Dual-Channel Multi-Dimensional Sentiment Encoder**:
  - **Channel I**: Extracts multi-scale sentiment features using the proposed ADCMix and models long-range dependencies via BiLSTM with attention.
  - **Channel II**: Preserves original contextual information through a parallel BiLSTM with attention path.
  - **Integration**: Fuses both channels to enhance feature richness and multi-dimensional sentiment representation.
- **Multi-Dimensional Sentiment Decoder**: Combines a global attention module and an MLP to decode subtle sentiment variations and refine sentiment representations.
- **Multi-Dimensional Sentiment Classifier**: Maps the final feature representation to a sentiment dimension using a fully connected layer and SoftMax activation.

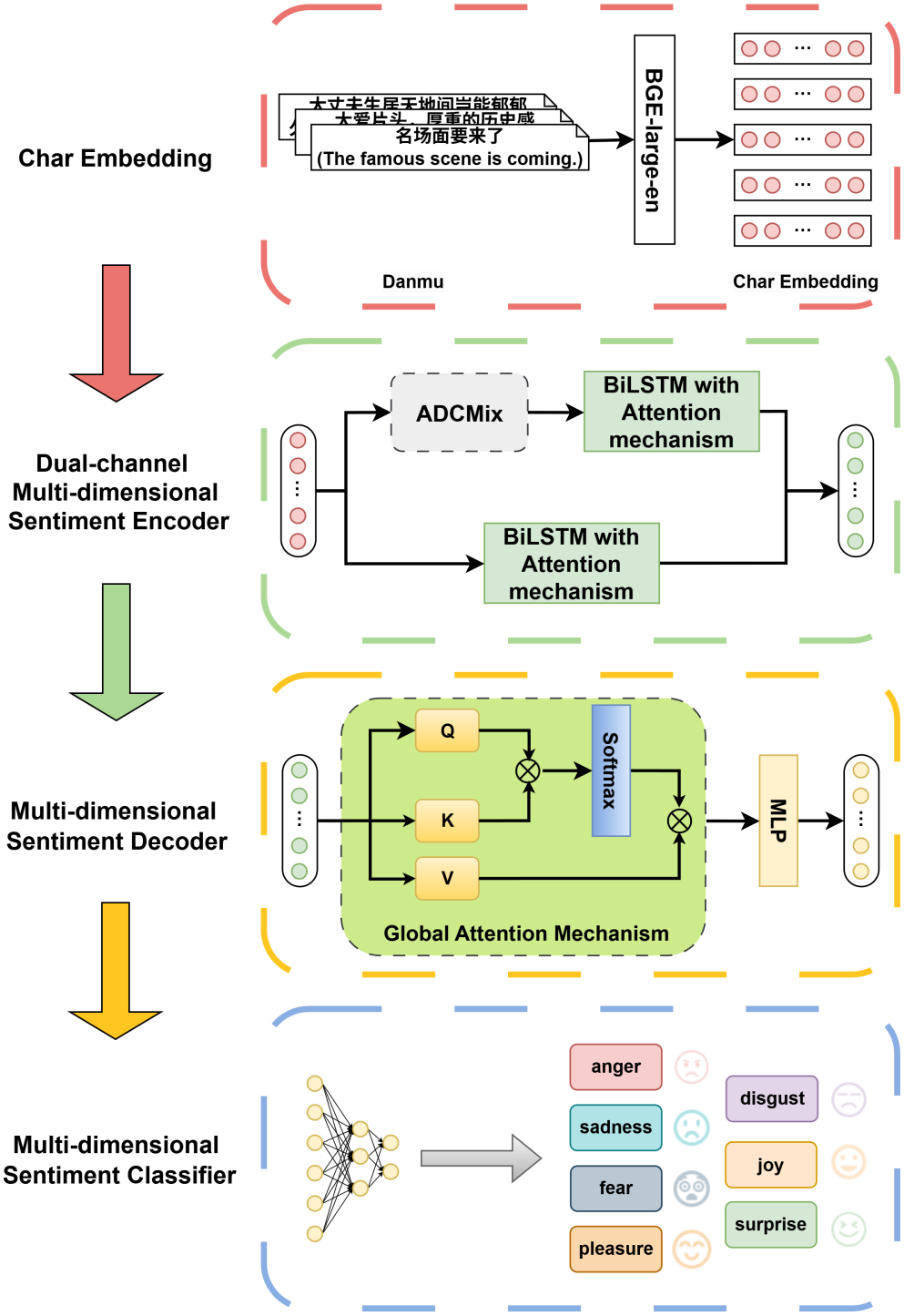


Figure 2. The four main modules of the proposed model: char embedding, dual-channel multi-dimensional sentiment encoder, multi-dimensional sentiment decoder, and a multi-dimensional sentiment classifier. The symbol ⊗ represents element-wise multiplication.

3.1. Char Embedding

Traditional word embeddings struggle to capture the complex and dynamic semantics of Danmu, which are often concise and filled with internet slang. An additional complication is the flexible structure of Chinese words. Different Chinese characters combine into various words, sometimes with substantial semantic differences. To address these challenges, this study extracts the char-level features from Danmu using BGE-large-en.

A sentence S is represented as a sequence of n characters. The BGE-large-en model maps each character in a sentence to a d-dimensional vector. Let $\mathbf{c_t} \in \mathbb{R}^d$ be the char embedding corresponding to the tth character. The char embedding matrix $\mathbf{T}$ of sentence S is then expressed as:

$$\mathbf{T} = \left[\mathbf{c_1}, \mathbf{c_2}, \ldots, \mathbf{c_t}, \ldots, \mathbf{c_n}\right] \in \mathbb{R}^{d \times n},\ t = 1, 2, \ldots, n. \tag{1}$$

In our experiments, we use the BGE-large-en model with embedding dimension $d = 768$. We did not perform additional fine-tuning on the BGE encoder but rely on its pretrained capability to handle mixed-language and informal text. This character-level embedding has proven effective in capturing the semantic richness of Chinese Danmu, as demonstrated by our improved sentiment analysis results.

3.2. Dual-Channel Multi-Dimensional Sentiment Encoder

As the sentiment information in Danmu is complex and the sentiment dependencies can vary between different words, the boundaries between similar sentiments are blurred. The dual-channel ADCMix–BiLSTM extracts both high-level semantic and raw contextual information from Danmu and integrates them to enhance multi-dimensional sentiment representation. The dual-channel structure consists of three main parts: channel I, channel II, and dual-channel integration.

3.2.1. Channel I

Channel I comprises two parts: ADCMix and BiLSTM with attention mechanisms. For each char embedding matrix $\mathbf{T}$, the multi-scale sentiment features are initially captured by ADCMix. The context dependency is then captured by BiLSTM with attention mechanisms. This process enhances the accuracy of the multi-dimensional Danmu sentiment features.

Previous works have shown that ACMix [27] mitigates the shortcomings of CNNs by strategically combining and leveraging the strengths of the convolutional and self-attention mechanisms. Nevertheless, the convolution operation in ACMix cannot fully capture the multi-dimensional sentiment features of nuanced sentiments or similar sentiments with indistinct boundaries. In such cases, the extracted sentiment can be biased toward a particular interpretation. Furthermore, as ACMix relies on convolution for local feature extraction, it cannot capture the long-distance dependencies in certain contexts, limiting the overall performance of the model. In contrast, the dilated CNN in our proposed ADCMix captures multi-dimensional sentiment features from different perspectives. These features are effectively integrated and fused through the combined self-attention mechanism and convolution.

ADCMix integrates the convolutional network with a self-attention mechanism. In this architecture, both modules form a single projection with the same dilated convolution operation. The resulting intermediate features are reused in distinct aggregation operations. The two stages of ADCMix are illustrated in Figure 3.

The first stage of ADCMix projects the features into a deeper space using the dilated CNN in the feature learning module. From each char embedding matrix $\mathbf{T}$, the high-level features at varying scales are extracted through dilated convolutions with different dilation rates ($r1$, $r2$, $r3$). The training process, stabilized and accelerated by batch normalization, is followed by a Rectified Linear Unit (ReLU) activation function. The convolution outputs of the three dilation rates are combined into an intermediate feature $\mathbf{I} = \{\mathbf{X}^{r1}, \mathbf{X}^{r2}, \mathbf{X}^{r3}\}$, where $\mathbf{X}^r$ represents the feature matrix of the dilated convolution output with dilation rate r:

$$\mathbf{X}^r = ReLU(\mathrm{conv}(\mathbf{W}_{ck}, \mathbf{T}, r)), \tag{2}$$

where $\mathbf{W}_{ck}$ is the weight of kernel in dilated convolution, and $ReLU(\cdot)$ denotes the ReLU activation function:

$$ReLU(x) = \max(0, x). \tag{3}$$

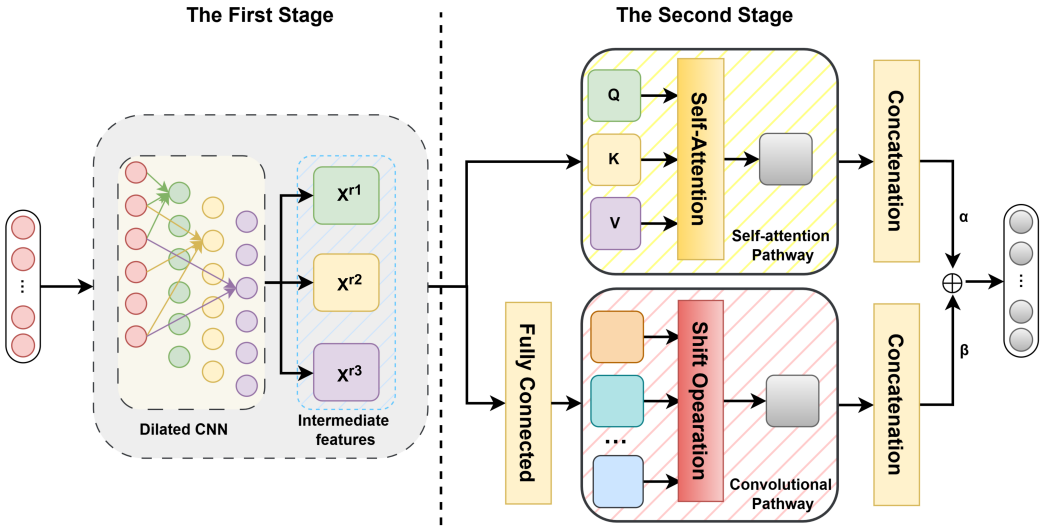


Figure 3. Structure of ADCMix. During the first stage, the dilated CNN projects the input features into a deeper space. During the second stage, the intermediate features are used in two pathways. The $\oplus$ symbol denotes a vector concatenation operation.

The second stage of ADCMix is the feature aggregation process, which follows distinct paradigms to ensure effective aggregation. The self-attention pathway processes the intermediate features through the traditional multi-head self-attention structure. The outputs of all attention heads are concatenated and linearly transformed to produce the final aggregated feature representation. This process is formulated as:

$$\mathbf{F}_{att} = \text{concat}(\mathbf{head}_1, \mathbf{head}_2, \ldots, \mathbf{head}_h) \cdot \mathbf{W}^O, \tag{4}$$

where $\mathbf{F}_{att}$ is the final output of self-attention pathway, $\mathbf{head}_i$ is the output of the ith head of attention, and $\mathbf{W}^O$ is the learned output weight matrix.

$$\mathbf{head}_i = \mathbf{A}_i(\mathbf{q}, \mathbf{k}) \cdot \mathbf{v}, \tag{5}$$

where $\mathbf{q}$, $\mathbf{k}$, $\mathbf{v}$ corresponding to $\mathbf{X}^{r1}$, $\mathbf{X}^{r2}$, and $\mathbf{X}^{r3}$ denote queries, keys, and values, respectively. Furthermore, $\mathbf{A}(\mathbf{q}, \mathbf{k})$ are the calculated attention weights based on the queries and keys.

$$\mathbf{A}(\mathbf{q}, \mathbf{k}) = \text{SoftMax}\left(\frac{\mathbf{q} \cdot \mathbf{k}^T}{\sqrt{d}} + \text{conv}(\mathbf{PE})\right), \tag{6}$$

where d is the dimension of $\mathbf{q}$, and $\mathbf{PE}$ is the relative position encoding, inspired by Swin-Transformer [28]. However, unlike Swin-Transformer, this $\mathbf{PE}$ is generated assuming a linear spacing between -1.0 and 1.0. It is followed by a convolution operation to enrich the positional information.

The convolutional pathway performs shifting and aggregation operations on the intermediate features. The features are first fused through a fully connected (FC) layer and then fed to a convolutional layer in a format mimicking the traditional convolutional methods, which collect information from local receptive fields:

$$\mathbf{F}_{conv} = \text{concat}(Con(\mathbf{X}^{r1}), Con(\mathbf{X}^{r2}), Con(\mathbf{X}^{r3})), \tag{7}$$

$$Con(\mathbf{X}^r) = \text{conv}(\text{FC}(\mathbf{X}^r)), \tag{8}$$

where $\mathbf{F}_{conv}$ is the final output of the convolutional pathway, and Con represents the shifting and aggregation operation.

Finally, the outputs of the two pathways are summed. Their weights are controlled by two learnable scalars, α and β. They are initialized as 0.5 and updated via backpropagation during training without additional constraints. They control the relative contribution of the attention and convolutional pathways in the final output fusion:

$$\mathbf{F}_{out} = \alpha \mathbf{F}_{att} + \beta \mathbf{F}_{conv}. \tag{9}$$

The ADCMix operation is followed by the BiLSTM with attention mechanisms to capture the multi-dimensional sentiment features from different perspectives while enhancing the feature richness and accuracy through feature integration. The main structure of the BiLSTM with attention mechanisms is shown in Figure 4:

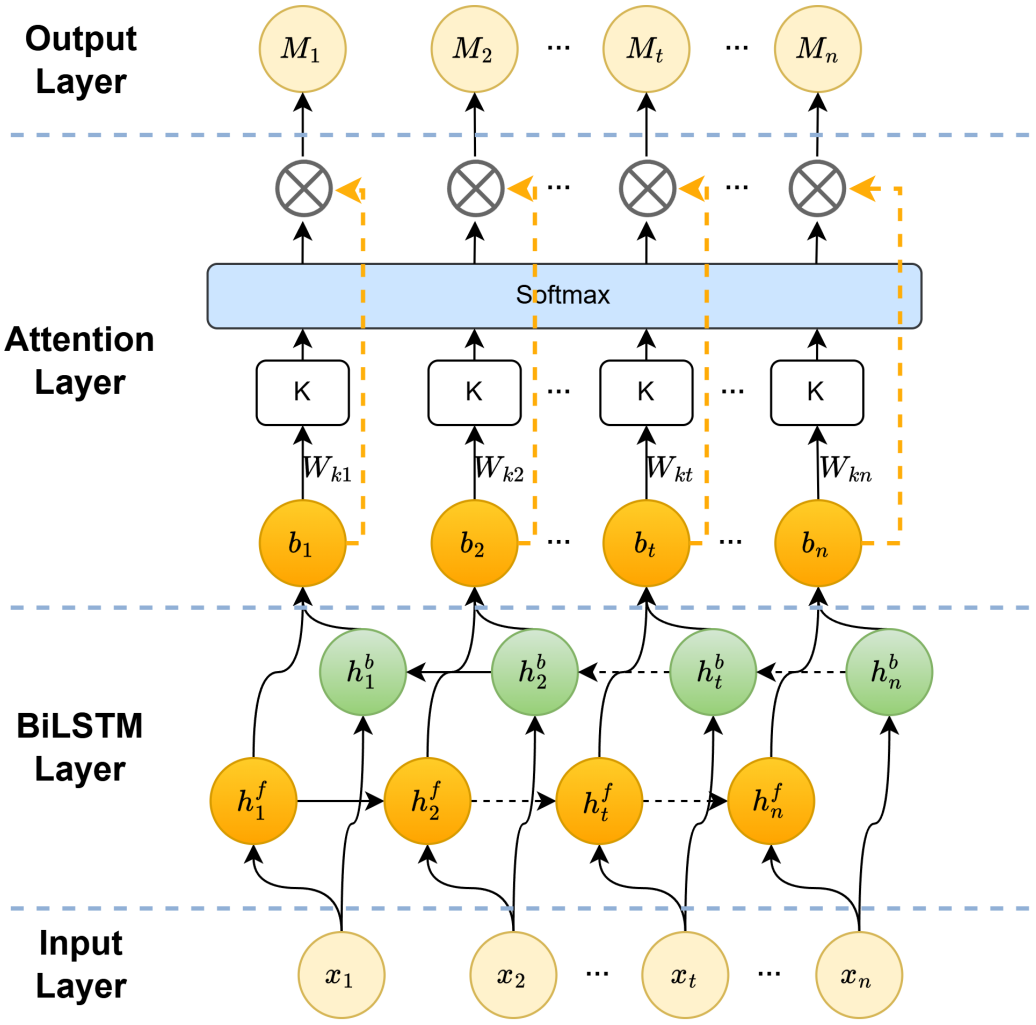


Figure 4. Structure of the BiLSTM with attention mechanisms.

The BiLSTM captures the enriched context information fused from two LSTM layers: the forward LSTM and the backward LSTM. The BiLSTM **B** output is given by

$$\mathbf{B} = [\mathbf{h}^f, \mathbf{h}^b], \tag{10}$$

$$\mathbf{h}^f = \overrightarrow{\mathbf{LSTM}}(\mathbf{F}_{out}), \tag{11}$$

$$\mathbf{h}^b = \overleftarrow{\mathbf{LSTM}}(\mathbf{F}_{out}), \tag{12}$$

where $\mathbf{h}^f$ and $\mathbf{h}^b$ are the outputs of the forward and backward LSTMs, respectively.

The attention mechanism assigns different weights to the output of the BiLSTM, enriching the model's capacity for sentiment representation. The output becomes the encoded features as detailed below:

$$\mathbf{K} = \tanh\left(\mathbf{W}_k \mathbf{B}^T\right), \tag{13}$$

$$\mathbf{A} = \text{softmax}(\mathbf{W}_a \mathbf{K}), \tag{14}$$

$$\mathbf{M}^I = \mathbf{A}\mathbf{B}. \tag{15}$$

Here, $\mathbf{B}^T$ is the transpose matrix of $\mathbf{B}$, $\mathbf{K}$ is the key matrix, $\mathbf{A}$ is the matrix of attention weights, and $\mathbf{M}^I$ denotes the output of the attention layer. $\mathbf{W}_k$ and $\mathbf{W}_a$ are learnable weight matrices used in the computations of key and attention-weight matrices, respectively.

3.2.2. Channel II

Channel II is designed to capture the original contextual information and hence enrich the sentiment information in the output. This channel processes the raw char embedding using a BiLSTM with attention mechanisms, structured identically to that of channel I. This design ensures that the model fully retains the original semantic information while extracting the complex sentiment features, providing a comprehensive and accurate basis for multi-dimensional sentiment analysis.

3.2.3. Dual-Channel Integration

Channels I and II are dedicated to the capture of high-level multi-dimensional sentiment features and the original context dependency, respectively. The two channels are combined for comprehensive and accurate extraction of the multi-dimensional sentiment features in Danmu. This dual-channel integration improves the accuracy and expressiveness of the sentiment analysis. Denoting $\mathbf{M}^I$ and $\mathbf{M}^{II}$ as the outputs of channels I and II, respectively, the output $\widehat{\mathbf{M}}$ of the dual-channel multi-dimensional sentiment encoder is described as follows:

$$\widehat{\mathbf{M}} = \mathbf{M}^I + \mathbf{M}^{II}. \tag{16}$$

3.3. Multi-Dimensional Sentiment Decoder

To bridge the gap between complex multi-dimensional sentiment encoding and accurate sentiment classification, we introduce a multi-dimensional sentiment decoder. Positioned between the encoder and classifier, the decoder aims to refine and project the high-level features into a sentiment space where subtle distinctions and compound sentiments become more separable. This is especially important for Danmu, where sentiment boundaries are often blurred. Our decoder, consisting of two main modules—a global attention module and a multi-layer perceptron (MLP) module—distinguishes subtle emotional variations and improves the comprehensiveness of the sentiment information analysis.

The global attention mechanism assigns global weights and decodes the sentiment features generated by the multi-dimensional sentiment encoder. By assigning different weights to each position, this mechanism captures the global dependencies of sentiment information across the entire sequence. The global attention mechanism enables a more comprehensive analysis of Danmu than local convolution operations. The mechanism is formulated as:

$$\tilde{\mathbf{C}} = \text{conv}\left(\mathbf{W}^C, \widehat{\mathbf{M}}\right), \tag{17}$$

$$\tilde{\mathbf{Q}} = \tilde{\mathbf{C}}\mathbf{W}^{Q}, \tag{18}$$

$$\tilde{\mathbf{K}} = \tilde{\mathbf{C}}\mathbf{W}^{K}, \tag{19}$$

$$\tilde{\mathbf{V}} = \tilde{\mathbf{C}}\mathbf{W}^{V}, \tag{20}$$

$$\tilde{\mathbf{M}} = \text{softmax}\left(\frac{\tilde{\mathbf{Q}}\tilde{\mathbf{K}}^{T}}{\sqrt{d}}\right)\tilde{\mathbf{V}}, \tag{21}$$

where $\widehat{\mathbf{M}}$ and $\tilde{\mathbf{M}}$ denote the input of the multi-dimensional sentiment decoder and the output of the global attention mechanism, respectively, $\mathbf{W}^{C}$ is the weight matrix of the convolution operation, d is the dimension of $\tilde{\mathbf{C}}$, and the other parameters are those of the attention mechanism.

The features are integrated in the MLP module, which compresses the sentiment features generated by the encoder into a more discriminative representation to enhance their correlation and complementarity and improve the accuracy of sentiment classification. Besides enhancing the model's capacity to represent complex sentiment features, the MLP module improves the overall classification performance. The detailed process is given below:

$$\mathbf{M}^{decoder} = \text{MLP}(\tilde{\mathbf{M}}). \tag{22}$$

3.4. Multi-Dimensional Sentiment Classifier

This module first converts the extracted sentiment features into a probability distribution of sentiment dimensions, providing corresponding assessments of the sentiment intensities in different categories. In practice, the multi-dimensional sentiment is evaluated through an FC layer and a SoftMax activation layer, which outputs the probability distribution of sentiment labels:

$$\mathbf{P} = \text{SoftMax}\left(\mathbf{M}^{decoder}\mathbf{W}_1 + \mathbf{B}_1\right), \tag{23}$$

where $\mathbf{M}^{decoder}$ refers to the output of the MLP module, and $\mathbf{W}_1$ and $\mathbf{B}_1$ are the trainable weight and bias parameters of the FC layer, respectively.

4. Results and Discussion

The performance of the proposed dual-channel ADCMix–BiLSTM model integrated with attention mechanisms was compared with that of SOTA models. The factors influencing the model's performance were then evaluated. This section describes the implementation details of the model and the performance evaluation experiments.

4.1. Experimental Setup

Evaluation criteria To evaluate the model performance in multi-dimensional Danmu sentiment analysis, we selected the accuracy (ACC) and F1-score. Accuracy reflects the overall correctness of predictions, while F1-score considers both precision and recall, making it more suitable for imbalanced and multi-class sentiment classification. Together, these metrics provide a balanced assessment of the model's effectiveness.

Datasets As publicly available datasets related to Danmu are limited, we constructed two new datasets based on *Romance of The Three Kingdoms* and *The Truman Show*. A substantial volume of Danmu was collected using the official application programming interface of the Internet video platform Bilibili (https://www.bilibili.com/ (accessed on 25 July 2025)). The Danmu texts from *Romance of The Three Kingdoms* reflect traditional Chinese cultural expressions, while those from *The Truman Show* are rich in contemporary colloquial and internet language. The combination of the two datasets captures both the cultural depth and the informal diversity of Danmu, providing a robust and representative foundation

for evaluating our model. Each dataset contains individual Danmu comments, annotated with one of seven dimensions of sentiment (joy, anger, sadness, fear, surprise, disgust, and pleasure). These categories are derived from the emotion wheel theory [29], which conceptualizes human emotions as a set of core affective states widely adopted in psychology and affective computing. All Danmu texts underwent preprocessing, including meaningless-symbol removal and noise filtering, to ensure data quality and consistency. For annotation, we invited annotators from diverse professional backgrounds (students, teachers, factory workers, and train attendants). All annotators first watched the original videos to understand the full context, and then assigned sentiment labels to each Danmu comment accordingly. Each Danmu was independently annotated by at least two individuals. The final sentiment labels were reviewed and validated by a panel of experts with backgrounds in affective computing. In cases where disagreements arose among expert reviewers, the final decision was determined by majority voting. This annotation process ensures the reliability and consistency of sentiment labels. The datasets were randomly split into three subsets: 70% for training, 10% for validation, and 20% for testing. Basic statistics of the datasets are provided in Table 1. To facilitate external validation and further research, we have publicly released the annotated datasets, which are available on Baidu AI Studio (https://aistudio.baidu.com/datasetdetail/329261 (accessed on 25 July 2025)).

Table 1. Basic statistics of the datasets.

Romance of the Three Kingdoms						
Number of Danmu:						**6736**
Joy	**Anger**	**Sadness**	**Fear**	**Surprise**	**Disgust**	**Pleasure**
831	1024	1460	42	391	446	2541
The Truman Show						
Number of Danmu:						**8858**
Joy	**Anger**	**Sadness**	**Fear**	**Surprise**	**Disgust**	**Pleasure**
1191	1072	2432	186	482	719	2776

Training details The network was trained using the Adam optimizer [30] with an initial learning rate of 0.00004 and a total of 10 training epochs. To mitigate overfitting and encourage convergence, we adopted a cosine annealing learning rate schedule [31] with a minimum learning rate of 0.000001 and a cycle length of 10 iterations. We employed cross-entropy as the primary loss function due to its effectiveness in multi-class classification tasks. The batch size was set to 32, and dropout was applied at multiple layers to reduce overfitting. The hyperparameters of the model, including the number of channels, kernel size, dilation rates, dropout rates, and attention head dimensions, were selected based on a combination of grid search and empirical performance on the validation set. The detailed architectural hyperparameters are listed in Table 2.

Table 2. Hyperparameters of the proposed ADCMix–BiLSTM model.

ADCMix					
Dilated CNN					
Out_Channels	**Kernel**	**Dilation**	**Padding Formula**		**Activation**
128	3	(1, 2, 3)	$\lfloor \frac{\text{Kernel}-1}{2} \rfloor \cdot \text{Dilation}$		ReLU
Self-Attention Pathway			*Convolutional Pathway*		
Embed_Dim	**Heads**	**Head_Dim**	**Out_Channels**	**Kernel**	**Groups**
128	4	32	128	1	32
BiLSTM with Attention Mechanism					
BiLSTM			*Attention Mechanism*		
Num_Layers	**Dropout**	**Hidden_Size**	**Activation**	**Dropout**	
2	0.5	256	tanh	0.3	

4.2. Comparisons of the ADCMix–BiLSTM and SOTA Models

Our model was pitted against the EWE–LSTM [3], BiLSTM [4], and multi-channel dilated CNN–BiLSTM with attention mechanism [21] deep learning models, which process text at word-level granularity. Word embedding was achieved using the Skip-gram model within Word2Vec [32]. To further validate the superiority of our model, we also pitted it against the pretrained models BERT–Kshape [5], Ernie–Tiny [22], Chinese-BERT-wwm [33], Chinese-ELECTRA-base [34], and RoBERTa-wwm-ext [23], which process text at char-level granularity. To ensure fairness, 5-fold cross-validation is also employed. Table 3 compares the results of all models after training and testing on the same samples. In addition, a significance test is performed on the results. Figure 5 presents the *p*-values from the paired t-test. The code of our model is available online (https://github.com/Gyou1t/ADCMix-BiLSTM (accessed on 25 July 2025)).

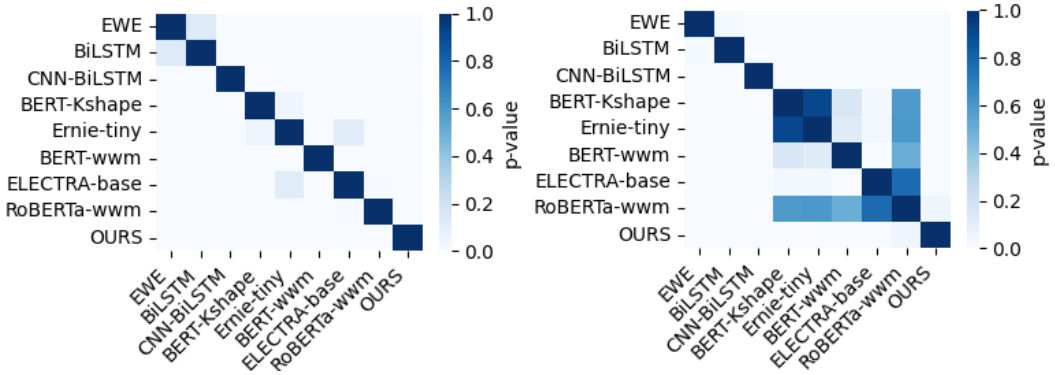
Table 3. Performance comparison on *Romance of The Three Kingdoms* (*The Three Kingdoms* for short) and *The Truman Show*.The data format is x^{+a}_{-b}, where x stands for the average of the results, a is the difference between the maximum and the average, and b represents the difference between the minimum and the average.

	Method	***The Three Kingdoms***		***The Truman Show***	
		ACC (%)	**F1-Score (%)**	**ACC (%)**	**F1-Score (%)**
deep learning	EWE-LSTM	$50.74^{-1.82}_{+0.26}$	$49.05^{-0.06}_{+9.52}$	$55.98^{-1.06}_{+0.13}$	$53.13^{-1.10}_{+2.08}$
	BiLSTM	$60.33^{-0.32}_{+0.43}$	$56.57^{-0.45}_{+0.31}$	$59.41^{-0.48}_{+0.40}$	$55.03^{-0.51}_{+0.41}$
	Dilated CNN-BiLSTM	$70.07^{-0.75}_{+0.65}$	$69.09^{-0.19}_{+10.11}$	$71.11^{-1.11}_{+0.89}$	$70.84^{-1.95}_{+0.82}$
pretrained	BERT-Kshape	$92.48^{-0.63}_{+0.42}$	$92.00^{-0.60}_{+0.47}$	$90.11^{-0.61}_{+0.34}$	$89.78^{-0.76}_{+0.21}$
	Ernie-tiny	$92.68^{-0.48}_{+0.47}$	$92.48^{-0.48}_{+0.46}$	$90.92^{-0.62}_{+0.28}$	$89.47^{-0.46}_{+0.42}$
	Chinese-BERT-wwm	$89.42^{-0.47}_{+0.38}$	$89.06^{-0.51}_{+0.31}$	$90.04^{-0.44}_{+0.26}$	$89.80^{-0.31}_{+0.33}$
	Chinese-ELECTRA-base	$93.29^{-0.49}_{+0.21}$	$93.06^{-0.53}_{+0.18}$	$89.98^{-0.48}_{+0.22}$	$89.04^{-0.52}_{+0.19}$
	RoBERTa-wwm-ext	$94.57^{-0.47}_{+0.33}$	$93.34^{-0.22}_{+0.35}$	$86.77^{-1.57}_{+0.73}$	$86.63^{-1.51}_{+1.25}$
	OURS	$\mathbf{96.69}^{-0.31}_{+0.64}$	$\mathbf{96.69}^{-0.31}_{+0.64}$	$\mathbf{94.58}^{-0.27}_{+0.26}$	$\mathbf{94.58}^{-0.29}_{+0.26}$

Note: Bold values indicate the best performance in each metric.

As indicated in Table 3, our method outperformed the current deep learning-based approaches for sentiment analysis. Traditional word embedding models rely strongly on pre-defined corpora, limiting their capability for handling variations across different domains or linguistic contexts. Therefore, traditional word embedding might not fully

capture the colloquial nature of Danmu. Our method substantially outperformed the EWE–LSTM, mainly because EWE relies on sentiment dictionaries that do not cover newly emerging Danmu terms. Moreover, the accurate conversion of sentiment dictionary entries into sentiment features is an unresolved problem in multi-dimensional sentiment analysis. Our method also significantly outperformed BiLSTM, primarily because the integration and attention mechanisms improve the integration of CNN and LSTM, thereby enhancing the sentiment analysis capability. Eventually, our method outperformed the multi-channel dilated CNN–BiLSTM model, demonstrating that ADCMix effectively captures the high-level semantic information of the multi-dimensional sentiment features in Danmu.



(**a**) *Romance of The Three Kingdoms* (**b**) *The Truman Show*

Figure 5. Visual representation of *p*-values of t-test results for two datasets.

Furthermore, the proposed model outperformed the pretrained models BERT–Kshape, Ernie–Tiny, Chinese–BERT-wwm, Chinese–ELECTRA-base, and RoBERTa-wwm-ext on both datasets. Compared to the SOTA methods, our method improved the ACC and F1-score by at least 2.05% and at least 3.28%, respectively, on the *Romance of the Three Kingdoms* dataset, and by at least 3.39% and 4.49%, respectively, on the *The Truman Show* dataset. The factor largely depends on the understanding abilities of pretrained models, which are insufficient for handling the informal language of Danmu. Moreover, pretrained models are unlikely to accurately capture the complex and subtle sentiment features across the different dimensions of multi-dimensional sentiment analysis.

In summary, the proposed model outperformed the existing methods on both datasets. As the ADCMix module effectively captures the high-level semantic information of multi-dimensional sentiment features in Danmu, our method can fully explore the multi-dimensional sentiment information embedded in the data. BiLSTM, integrated with the local attention mechanism, well captures the contextual dependencies. In addition, the multi-dimensional sentiment features are supplemented with the original contextual features, mitigating potential biases and omissions in the sentiment representation. The integration of global attention and MLP further enhances the accuracy of multi-dimensional sentiment feature extraction from Danmu.

4.3. Analysis of the Factors Influencing Model Performance

Table 3 demonstrates the superior performance of our model. The effectiveness of each component in our model was evaluated through a series of ablation studies.

4.3.1. Embedding and Channel Analysis

The selected embedding methods and channels of the proposed architecture largely determine the feature representation quality and overall performance of our model. The following analysis determines the impacts of different embedding methods and channels of the proposed architecture on the model's performance.

Analysis of different embedding methods: The effects of the word-level and char-level embedding methods on the model's performance were explored through ablation experiments. As shown in Table 4, char-level embedding generally yielded better results than word-level embedding. This result is likely explained by the colloquial style with rich cultural references and internet slang in Danmu texts. As word-level embedding cannot easily capture new or unconventional words, it often fails to accurately represent the sentiment features in Danmu. These challenges were more effectively handled by char-level embedding.

Table 4. Analysis of different embedding methods.

Embedding Method	*Romance of The Three Kingdoms*		*The Truman Show*	
	ACC (%)	**F1-Score (%)**	**ACC (%)**	**F1-Score (%)**
word-level	51.72	46.85	54.40	50.39
char-level	**96.62**	**96.62**	**94.31**	**94.29**

Note: Bold values indicate the best performance in each metric.

Analysis of different channel of the proposed architecture: Next, the impacts of single-channel and dual-channel configurations on the model's performance were investigated through ablation experiments. As shown in Table 5, the combined channels I and II generally yielded better results than the single-channel models. The high-level semantic information extracted by channel I was supplemented with the raw contextual information extracted by channel II, mitigating potential biases in sentiment features.

Table 5. Analysis of different channels of the proposed architecture.

Architecture	*Romance of The Three Kingdoms*		*The Truman Show*	
	ACC (%)	**F1-Score (%)**	**ACC (%)**	**F1-Score (%)**
Single Channel I	88.54	88.22	86.59	86.53
Single Channel II	81.41	81.19	77.70	77.23
Dual-channel	**96.62**	**96.62**	**94.31**	**94.29**

Note: Bold values indicate the best performance in each metric.

4.3.2. Parameter Analysis in ADCMix

ADCMix is the key component of the proposed model, designed to capture sentiment features from different perspectives. The hyperparameters of ADCMix are provided in Table 1. This subsection analyses the impact of the key ADCMix parameters on the model's performance.

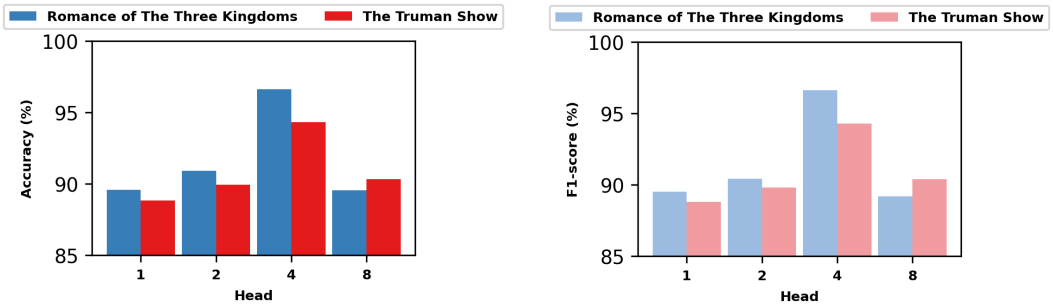
Analysis of dilation rate in the dilated CNN: The dilation rate of the dilated CNN is a critical parameter of the proposed model. If the receptive field is excessively small, the model might capture insufficient contextual information. Conversely, if the receptive field is excessively large, the model might overlook important details. The impacts of different dilation rate combinations on the model performance are detailed in Table 6. Clearly, increasing the dilation rate enhanced the capture of multi-scale features by the model. Specifically, changing the dilation rate from (1, 1, 1) to (1, 2, 3) improved the accuracies and F1-score to their highest values on both datasets. This result demonstrates that dilated CNNs more effectively extract the precise multi-dimensional sentiment features than traditional CNNs. However, further increasing the dilation rate slightly declined the model performance, possibly because excessively high dilution rates induce the grid effect, leading to biases in high-level features. They also increase the emotional distance between the sentiment characters, causing overall deviations in the sentiment analysis.

Table 6. Analysis of dilated rate of dilated CNN.

Dilated Rate of Channel	*Romance of The Three Kingdoms*		*The Truman Show*	
	ACC (%)	**F1-Score (%)**	**ACC (%)**	**F1-Score (%)**
(1,1,1)	96.02	96.02	92.02	92.74
(1,2,3)	**96.62**	**96.62**	**94.31**	**94.29**
(2,2,2)	95.31	95.31	93.77	93.74
(2,3,4)	95.67	95.65	93.00	93.00
(3,3,3)	95.37	95.33	91.96	91.96
(4,5,6)	95.31	95.32	92.28	92.26

Note: Bold values indicate the best performance in each metric.

Analysis of head number of the multi-head attention mechanism: The accuracy of the model also depends on the number of attention heads in the multi-head attention mechanism. Here, the accuracy of the proposed dual-channel ADCMix–BiLSTM model integrated with attention mechanisms was evaluated on two Danmu datasets with varying numbers of attention heads (1, 2, 4, and 8). The experimental results are shown in Figure 6. Four attention heads maximized the model performance on both datasets, with an accuracy and F1-score of 96.62% and 96.62%, respectively, on the *Romance of The Three Kingdoms* dataset and 94.31% and 94.29%, respectively, on the *The Truman Show* dataset. The performance improved as the number of heads increased from 1 to 4, reflecting the increasing number of diverse features extracted from different subspaces. However, increasing the number of heads to 8 reduced the performance, likely because the sparsity of attention scores adversely affected the stability of training.



(**a**) Accuracy for different attention heads (**b**) F1-score for different attention heads

Figure 6. Analysis of the head number of the multi-head attention mechanism.

Analysis of combining the outputs of two pathways: Finally, we explored the effects of different combinations of dilated CNN and self-attention mechanisms on the model performance. The experimental results of the various combination methods are summarized in Table 7. The performances of single-pathway models, Swin-T for self-attention and Conv–Swin-T (replacing window attention with dilated convolutions) for convolution, are also shown. Combining the dilated CNN and self-attention modules generally achieved better results than the single-pathway models. Moreover, fixing the ratio between convolution and self-attention tended to lower the performance. In contrast, the flexible use of learned parameters enables adaptive adjustment of the strengths of convolution and self-attention paths based on the filter's position throughout the network, thereby improving the model's performance.

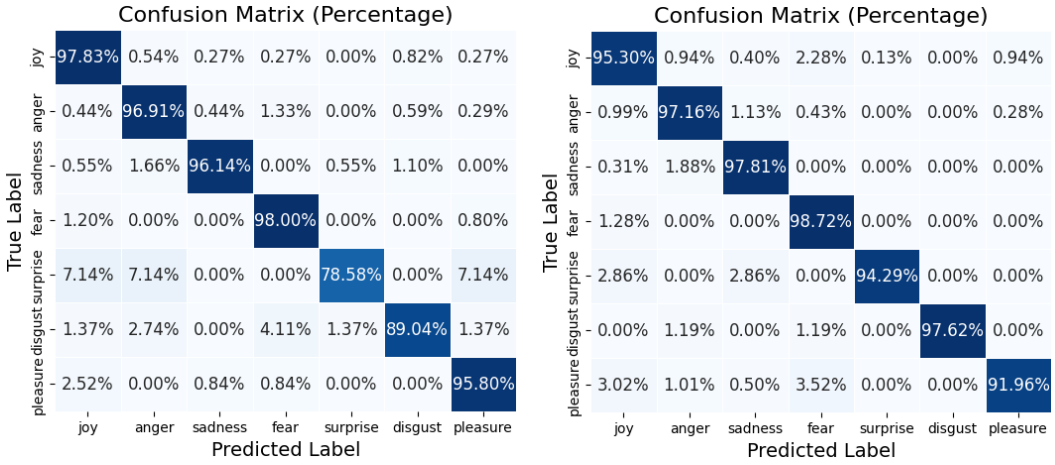
Table 7. Ablation study on combining the output of two pathways. The ADCMix final output is computed as $\mathbf{F}_{out} = \alpha\mathbf{F}_{att} + \beta\mathbf{F}_{conv}$.

Method	α	β	*Romance of The Three Kingdoms*		*The Truman Show*	
			ACC (%)	F1-Score (%)	ACC (%)	F1-Score (%)
Swin-T	1	-	95.13	95.12	90.88	90.87
Conv-Swin-T	-	1	93.17	93.20	90.70	90.70
Swin-ADCMix-T	1	1	96.56	96.55	91.96	91.96
	α	1	95.79	95.78	92.87	92.86
	α	1-α	95.61	95.61	92.51	92.50
	α	β	**96.62**	**96.62**	**94.31**	**94.29**

Note: Bold values indicate the best performance in each metric.

4.4. Analysis of Misclassifications

Figure 7 presents the confusion matrix results for both datasets. From the matrices, we observe that most categories are classified with high accuracy, particularly the core emotions such as joy and anger. However, several categories show notable confusion. On the *Romance of The Three Kingdoms* dataset, a common misclassification occurs between fear and anger. This suggests that when emotional expressions are implicit or nuanced, the model struggles to capture the intended sentiment, often mistaking reflective sadness for lightheartedness or tension for hostility. Similarly, on the *The Truman Show* dataset, we notice that surprise is frequently confused with joy, likely due to overlapping emotional cues such as excitement or unexpected positivity.



(**a**) *Romance of The Three Kingdoms* (**b**) *The Truman Show*

Figure 7. Visual representation of confusion matrix results.

Building upon this observation, we randomly sampled several misclassified cases from the test set for qualitative error analysis. It was found that when a text is relatively long and contains key emotional cues in the latter part or involves a sentiment shift, the model's prediction accuracy significantly drops. For example: "曹老板老了。跟第一集的差距太大了，节目组很用心，在我们来不及的发现曹老板变老了。害。" (Boss Cao has gotten old. The contrast with the first season is huge. The production team put in a lot of effort, and we didn't even notice when Boss Cao started aging. Sigh.). This Danmu expresses a sense of nostalgia and sadness overall and was manually labeled as sorrow. However, the model incorrectly classified it as joy. One possible reason is the presence of seemingly positive phrases such as "the production team put in a lot of effort" which

may have misled the model. Additionally, the ending word "sigh" is a common colloquial expression that conveys subtle emotional weight, which the model likely failed to capture.

5. Conclusions

We proposed a novel dual-channel ADCMix–BiLSTM architecture with attention mechanisms for multi-dimensional sentiment analysis of Danmu. Unlike most existing approaches that rely on word-level embeddings or single-channel architectures, our method employs character-level features to better capture the informal, abbreviated, and evolving nature of Danmu language. To address the challenges posed by sentiment ambiguity and diversity, our dual-channel encoder explicitly separates semantic abstraction (via ADCMix–BiLSTM in Channel I) and raw contextual features (in Channel II), enabling complementary learning and mitigating feature bias or omission. In addition, the multi-dimensional sentiment decoder further enhances the model's capacity to distinguish subtle emotional cues, which are often intertwined in Danmu expressions. Extensive experiments on two benchmark datasets demonstrate that our approach consistently outperforms several state-of-the-art (SOTA) baselines in capturing complex and nuanced emotional content. This confirms the effectiveness of dual-channel sentiment representation and the utility of multi-dimensional decoding in this unique domain.

Compared with related studies, which often simplify Danmu sentiment as binary or uni-dimensional classification, our method provides a finer-grained and culturally aware sentiment profiling, offering both academic and practical value. The insights gained can support not only content creators in refining their materials but also assist sociocultural analysis and public opinion monitoring by reflecting group value orientations in real time. The proposed method comprehensively analyses the sentiments of video viewers, highlighting its practical value. Video creators can use the sentiment feedback in Danmu to optimize the contents of their creations, whereas multi-dimensional Danmu sentiment analysis can effectively uncover the value orientations of social groups. It is valuable not only for understanding cultural dynamics but also for public opinion monitoring. Such insights can assist government authorities in identifying emotional trends and reasonably guiding online discourse.

Nevertheless, this study has several limitations. First, it does not incorporate sentiment cues embedded in the original video content, which may provide essential context for interpreting irony, sarcasm, or implicit sentiments. Second, all experiments are conducted solely on Chinese Danmu data, leaving the cross-lingual generalizability of the model untested. Third, given the fast-changing nature of online expressions, the current model may face difficulties in adapting to new or evolving linguistic patterns. To address these issues, future work will explore cross-modal fusion approaches to integrate visual and textual modalities for more accurate sentiment interpretation. Additionally, we plan to examine the model's adaptability in multilingual settings to evaluate its cross-lingual performance. Finally, we aim to incorporate adaptive strategies such as continual fine-tuning to maintain model robustness over time.

Author Contributions: W.P. contributed to the conceptualization, coding/experimentation, and manuscript writing. Z.B. was involved in the conceptualization, methodology, and writing—review and editing. Y.T. contributed to the writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Open Project Program of the State Key Laboratory of CAD&CG (Grant No. A2312), Zhejiang University.

Data Availability Statement: The data that support this study are available in Baidu AI Studio at https://aistudio.baidu.com/datasetdetail/329261 (accessed on 25 July 2025).

Acknowledgments: The authors would like to thank all those who contributed to this work.

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

References

1. Bai, Q.; Wei, K.; Zhou, J.; Xiong, C.; Wu, Y.; Lin, X.; He, L. Entity-level sentiment prediction in Danmaku video interaction. *J. Supercomput.* **2021**, *77*, 9474–9493. [CrossRef]
2. Al-Saqqa, S.; Awajan, A. The use of word2vec model in sentiment analysis: A survey. In Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control, Cairo, Egypt, 14–16 December 2019; pp. 39–43. [CrossRef]
3. Li, C.; Wang, J.; Wang, H.; Zhao, M.; Li, W.; Deng, X. Visual-texual emotion analysis with deep coupled video and danmu neural networks. *IEEE Trans. Multimed.* **2019**, *22*, 1634–1646. [CrossRef]
4. Wang, S.; Chen, Y.; Ming, H.; Huang, H.; Mi, L.; Shi, Z. Improved danmaku emotion analysis and its application based on bi-LSTM model. *IEEE Access* **2020**, *8*, 114123–114134. [CrossRef]
5. Zhang, R.; Ren, C. Sentiment time series clustering of Danmu videos based on BERT fine-tuning and SBD-K-shape. *Electron. Libr.* **2024**, *42*, 553–575. [CrossRef]
6. Sherin, A.; SelvakumariJeya, I.J.; Deepa, S. Enhanced Aquila optimizer combined ensemble Bi-LSTM-GRU with fuzzy emotion extractor for tweet sentiment analysis and classification. *IEEE Access* **2024**, *12*, 141932–141951. [CrossRef]
7. Zhu, Y.; Zhou, R.; Chen, G.; Zhang, B. Enhancing sentiment analysis of online comments: A novel approach integrating topic modeling and deep learning. *PeerJ Comput. Sci.* **2024**, *10*, e2542. [CrossRef] [PubMed]
8. Chen, X.; Shen, Z.; Guan, T.; Tao, Y.; Kang, Y.; Zhang, Y. Analyzing patient experience on weibo: Machine learning approach to topic modeling and sentiment analysis. *JMIR Med. Inform.* **2024**, *12*, e59249. [CrossRef] [PubMed]
9. Edara, D.C.; Vanukuri, L.P.; Sistla, V.; Kolli, V.K.K. Sentiment analysis and text categorization of cancer medical records with LSTM. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 5309–5325. [CrossRef]
10. Yenkikar, A.; Babu, C.N. AirBERT: A fine-tuned language representation model for airlines tweet sentiment analysis. *Intell. Decis. Technol.* **2023**, *17*, 435–455. [CrossRef]
11. Karabila, I.; Darraz, N.; EL-Ansari, A.; Alami, N.; EL Mallahi, M. BERT-enhanced sentiment analysis for personalized e-commerce recommendations. *Multimed. Tools Appl.* **2024**, *83*, 56463–56488. [CrossRef]
12. Singh, T.D.; Singh, T.J.; Shadang, M.; Thokchom, S. Review comments of manipuri online video: Good, bad or ugly. In Proceedings of the International Conference on Computing and Communication Systems: I3CS 2020, NEHU, Shillong, India, 28–30 April 2021; pp. 45–53. [CrossRef]
13. Suresh Kumar, K.; Radha Mani, A.; Ananth Kumar, T.; Jalili, A.; Gheisari, M.; Malik, Y.; Chen, H.C.; Jahangir Moshayedi, A. Sentiment Analysis of Short Texts Using SVMs and VSMs-Based Multiclass Semantic Classification. *Appl. Artif. Intell.* **2024**, *38*, 2321555. [CrossRef]
14. Hameed, Z.; Garcia-Zapirain, B. Sentiment classification using a single-layered BiLSTM model. *IEEE Access* **2020**, *8*, 73992–74001. [CrossRef]
15. Liu, K.; Feng, Y.; Zhang, L.; Wang, R.; Wang, W.; Yuan, X.; Cui, X.; Li, X.; Li, H. An effective personality-based model for short text sentiment classification using BiLSTM and self-attention. *Electronics* **2023**, *12*, 3274. [CrossRef]
16. Xiaoyan, L.; Raga, R.C. BiLSTM model with attention mechanism for sentiment classification on Chinese mixed text comments. *IEEE Access* **2023**, *11*, 26199–26210. [CrossRef]
17. Gao, Z.; Li, Z.; Luo, J.; Li, X. Short text aspect-based sentiment analysis based on CNN+ BiGRU. *Appl. Sci.* **2022**, *12*, 2707. [CrossRef]
18. Albahli, S.; Nawaz, M. TSM-CV: Twitter Sentiment Analysis for COVID-19 Vaccines Using Deep Learning. *Electronics* **2023**, *12*, 3372. [CrossRef]
19. Shan, Y. Social Network Text Sentiment Analysis Method Based on CNN-BiGRU in Big Data Environment. *Mob. Inf. Syst.* **2023**, *2023*, 8920094. [CrossRef]
20. Aslan, S. A deep learning-based sentiment analysis approach (MF-CNN-BILSTM) and topic modeling of tweets related to the Ukraine–Russia conflict. *Appl. Soft Comput.* **2023**, *143*, 110404. [CrossRef]
21. Gan, C.; Feng, Q.; Zhang, Z. Scalable multi-channel dilated CNN–BiLSTM model with attention mechanism for Chinese textual sentiment analysis. *Future Gener. Comput. Syst.* **2021**, *118*, 297–309. [CrossRef]
22. Su, W.; Chen, X.; Feng, S.; Liu, J.; Liu, W.; Sun, Y.; Tian, H.; Wu, H.; Wang, H. Ernie-tiny: A progressive distillation framework for pretrained transformer compression. *arXiv* **2021**, arXiv:2106.02241. [CrossRef]
23. Xu, Z. RoBERTa-WWM-EXT fine-tuning for Chinese text classification. *arXiv* **2021**, arXiv:2103.00492. [CrossRef]
24. Xiao, S.; Liu, Z.; Zhang, P.; Muennighoff, N. C-Pack: Packaged Resources To Advance General Chinese Embedding. *arXiv* **2023**, arXiv:2309.07597. [CrossRef]

25. Kumar, G.; Agrawal, R.; Sharma, K.; Gundalwar, P.R.; Agrawal, P.; Tomar, M.; Salagrama, S. Combining BERT and CNN for Sentiment Analysis A Case Study on COVID-19. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*. [CrossRef]
26. Jlifi, B.; Abidi, C.; Duvallet, C. Beyond the use of a novel Ensemble based Random Forest-BERT Model (Ens-RF-BERT) for the Sentiment Analysis of the hashtag COVID19 tweets. *Soc. Netw. Anal. Min.* **2024**, *14*, 88. [CrossRef]
27. Pan, X.; Ge, C.; Lu, R.; Song, S.; Chen, G.; Huang, Z.; Huang, G. On the integration of self-attention and convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 815–825. [CrossRef]
28. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022. [CrossRef]
29. Plutchik, R. A general psychoevolutionary theory of emotion. In *Theories of Emotion*; Elsevier: Amsterdam, The Netherlands, 1980. [CrossRef]
30. Zou, F.; Shen, L.; Jie, Z.; Zhang, W.; Liu, W. A Sufficient Condition for Convergences of Adam and RMSProp. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 11119–11127. [CrossRef]
31. Vrbančič, G.; Podgorelec, V. Efficient ensemble for image-based identification of Pneumonia utilizing deep CNN and SGD with warm restarts. *Expert Syst. Appl.* **2022**, *187*, 115834. [CrossRef]
32. Church, K.W. Word2Vec. *Nat. Lang. Eng.* **2017**, *23*, 155–162. [CrossRef]
33. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Wang, S.; Hu, G. Revisiting Pre-Trained Models for Chinese Natural Language Processing. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020, Online, 16–20 November 2020; pp. 657–668. [CrossRef]
34. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z. Pre-training with whole word masking for chinese bert. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 3504–3514. [CrossRef]