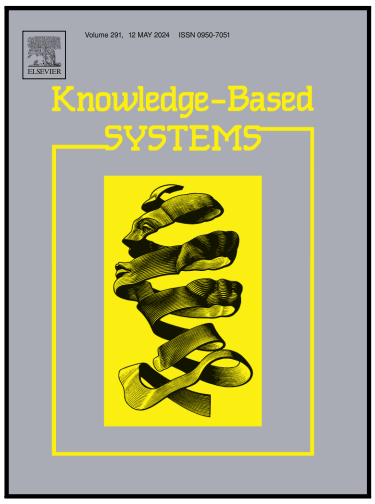


Personalized Multimodal Sentiment Analysis under Uncertain Modalities Missing via Pretraining and Online Learning

Hongxiang Sun, Zhizhong Liu, Dianhui Chu, Quan Z. Sheng, Zhaowei Liu, Jian Yu

PII: S0950-7051(25)01328-0
DOI: https://doi.org/10.1016/j.knosys.2025.114287
Reference: KNOSYS 114287

To appear in: *Knowledge-Based Systems*

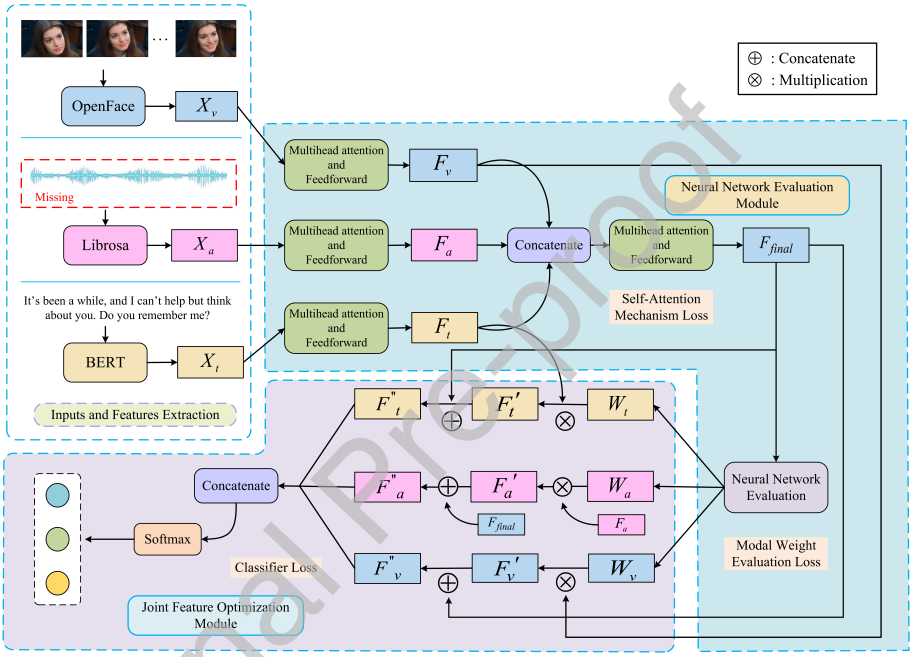
Received date: 29 April 2025
Revised date: 15 July 2025
Accepted date: 14 August 2025

Please cite this article as: Hongxiang Sun, Zhizhong Liu, Dianhui Chu, Quan Z. Sheng, Zhaowei Liu, Jian Yu, Personalized Multimodal Sentiment Analysis under Uncertain Modalities Missing via Pretraining and Online Learning, *Knowledge-Based Systems* (2025), doi: https://doi.org/10.1016/j.knosys.2025.114287

Graphical Abstract

Personalized Multimodal Sentiment Analysis under Uncertain Modalities Missing via Pretraining and Online Learning

Hongxiang Sun, Zhizhong Liu, Dianhui Chu, Quan Z. Sheng, Zhaowei Liu, Jian Yu



Highlights

Personalized Multimodal Sentiment Analysis under Uncertain Modalities Missing via Pretraining and Online Learning

Hongxiang Sun, Zhizhong Liu, Dianhui Chu, Quan Z. Sheng, Zhaowei Liu, Jian Yu

- To develop an effective MSA model for personalized users, we propose a personalized MSA model under uncertain modalities missing via pretraining and online learning (named PMSAPO), which is first pretrained with some public datasets, and then it autonomously adapts to personalized users through some online learning strategies. To the best of our knowledge, this is the first work that propose to develop personalized MSA model under uncertain modalities missing.

- To effectively to handle the uncertain missing modalities in MSA for personalized users, We propose a method of reusing the fused modalities to complete the missing modalities. In this method, a neural network evaluation module is proposed to assign different weights to each modality in the joint feature (fused modalities), and then the joint feature is fused with each weighted modality, so as to optimize each modality by reusing the fused modalities, which has excellent flexibility and generalization ability.

- We propose a set of online learning strategies to enable the pretrained MSA model to autonomously adapt to personalized users. Specifically, the adaptive learning rate adjustment strategy can adjust the learning rate according to the amount of data of personalized users. The online meta-learning strategy can obtain more adaptive personalized user parameters by optimizing the meta-task loss. The adaptive weights assigning strategy endows different weights to data samples of personalized users, thus to avoid catastrophic forgetting problems.

- We conduct extensive experiments to verify the performance of our proposed model based on three public benchmark datasets (IEMOCAP, MELD and CMU-MOSI). Experimental results prove that PMSAPO completely outperforms other 12 baseline models. Compared to the second-best performing model (SMCMSA), on the IEMOCAP dataset,

our proposed model PMSAPO has an average increase of 2.64% in M-F1 and 2.55% in ACC. On the MELD dataset, PMSAPO increase M-F1 by 3.19% on average, and improves ACC by 2.30% on average. On the CMU-MOSI dataset, PMSAPO increase M-F1 by 1.72% on average, and improves ACC by 2.02% on average.

Personalized Multimodal Sentiment Analysis under Uncertain Modalities Missing via Pretraining and Online Learning

Hongxiang Sun[a], Zhizhong Liu[a,*], Dianhui Chu[b,c], Quan Z. Sheng[d], Zhaowei Liu[a], Jian Yu[e]

[a] *The School of Computer and Control Engineering, Yantai University, Yantai, 264005, China*
[b] *College of Computer Science and Technology, Harbin Institute of Technology (Weihai), Weihai, 264209, China*
[c] *Shandong Key Laboratory of Digital Service Computing and Systems, Yantai, Yantai, 264005, China*
[d] *School of Computing, Macquarie University, Sydney, NSW, 2109, Australia*
[e] *Department of Computer Science, Auckland University of Technology, Auckland, 1142, New Zealand*

Abstract

Currently, multimodal sentiment analysis (MSA) for personalized users under uncertain modalities missing has become a new challenging problem. To address this issue, we propose a two-step idea. First, we propose an effective MSA model under uncertain modalities missing and train it with some public datasets, thus to enable the model to possess better preliminary MSA ability. Then, we make the pretrained model to continuously learn user's personalized characteristics with online learning methods, thereby enable the model grow into a robust model for personalized MSA. Based on this idea, we propose a Personalized MSA model under uncertain modalities missing via Pretraining and Online Learning (termed as PMSAPO). For Personalized MSA under uncertain modalities missing, PMSAPO firstly generates the fused modality and allocate weights for each modality with a Fully Connected Neural

*Corresponding author. Email: zhizhongliu@ytu.edu.cn
Email addresses: 2363453828@s.ytu.edu.cn (Hongxiang Sun), zhizhongliu@ytu.edu.cn (Zhizhong Liu), cdh@hitwh.edu.cn (Dianhui Chu), michael.sheng@mq.edu.au (Quan Z. Sheng), lzw@ytu.edu.cn (Zhaowei Liu), jian.yu@aut.ac.nz (Jian Yu)

Network Evaluation Module. Then, PMSAPO completes the final sentiment classification based on the fusion modality with a Joint feature optimization module. For the pretrained PMSAPO, we make it autonomously learn the personalized users via our proposed online learning techniques, including an online meta-learning method, a learning rate adaptive adjustment strategy, and a dynamic weight assignment strategy for sample data. Finally, based on three public benchmark datasets (IEMOCAP, MELD and CMU-MOSI), we conduct extensive experiments and prove that PMSAPO completely outperforms the Twelve state-of-the-art baseline models. (Code is available at https://github.com/SHX-AI/PMSAPO.)

Keywords: Multimodal sentiment analysis, Uncertain modalities missing, Pretraining, Online learning

1. Introduction

In people's daily lives and work, sentiment plays a crucial role in various human cognitive processes, such as learning, memory, and decision-making [1]. Accurate sentiment analysis has the greatest importance in various applications, including smart education [2], healthcare [3], intelligent recommendation [4], and human-computer interaction [5]. In the past few years, sentiment analysis technology has become a prominent focus in the artificial intelligence field [6]. In the early stages of research, sentiment analysis has been conducted based on single modal data (e.g., text data) with machine learning techniques, such as sentiment analysis based on personal blogs [7], sentiment analysis based on product reviews [8], sentiment analysis based on movie critiques [9], and so on.

Recently, with the rapid development of intelligent terminals and Internet technology, data on social platforms has transitioned from single modality to multiple modalities (text, visual, and audio) [10]. Actually, multimodal data can capture nuanced sentimental variations across multiple dimensions and reveal affective cues that are often imperceptible in single modality. Sentiment analysis based on multimodal data can yield more accurate and comprehensive insights. Therefore, Multimodal Sentiment Analysis (MSA) has emerged as a hot topic in the fields of artificial intelligence and affective computing.

MSA aims to recognize users' sentiment based on multimodal data (text, visual, and audio) [11]. In the past few years, some effective MSA models

have been developed with deep learning techniques [12], such as MSA based on recurrent neural networks (RNN) [13, 14], MSA based on graph convolutional neural networks (GCNN) [15, 16], MSA based on Transformer [17, 18], MSA based on contrastive learning models [19, 20], and so on. Undoubtedly, existing MSA research has achieved remarkable results and promoted the development of sentiment analysis technology significantly [18].

However, existing research only focuses on developing models for MSA, without considering how to tackle the challenges for personalized MSA. The first challenge is that the multimodal data of a personalized user is relatively limited, making it difficult to train an effective MSA model. Although MSA models can be trained with some public datasets (e.g., MELD [21]), however, models trained with public datasets do not have perfect performance for personalized users, because individual users' demographics often differ from those of the training datasets. It is worth mentioning that, we have conducted some experiments to verify the above phenomenon. Experimental results show that the performance of MSA models trained with public dataset almost halved when they are tested with personalized users' data. Therefore, how to develop an efficient MSA model for personalized users become a critical issue.

The second challenge is that, uncertain modalities missing occurs frequently due to some uncontrollable factors [12] in personalized MSA. For example, as illustrated in Fig. 1, when the camera device is occluded, the visual modality cannot be captured. When the users keep silent, the audio modality cannot be obtained. Recently, to solve this problem, some effective MSA models considering uncertain modalities missing have been proposed [12, 18]. However, existing works [12] always complete the uncertain missing modalities with a pre-trained model that trained with the full multimodal data. However, actually, it is difficult to obtain substantial high quality full multimodal data for training the pre-trained model in practical applications, due to the dynamism of the environment, users' privacy protection, or devices' abnormality. Moreover, the pre-trained model will be invalid when it is used for personalized users. Therefore, how to tackle the problem of uncertain modalities missing in personalized MSA is still a challenging problem. In all, although existing research on MSA has achieved some wonderful results, but there are still some challenges to be addressed for personalized MSA:

- Existing research on MSA does not consider how to develop effective MSA model for personalized users. Moreover, existing MSA models

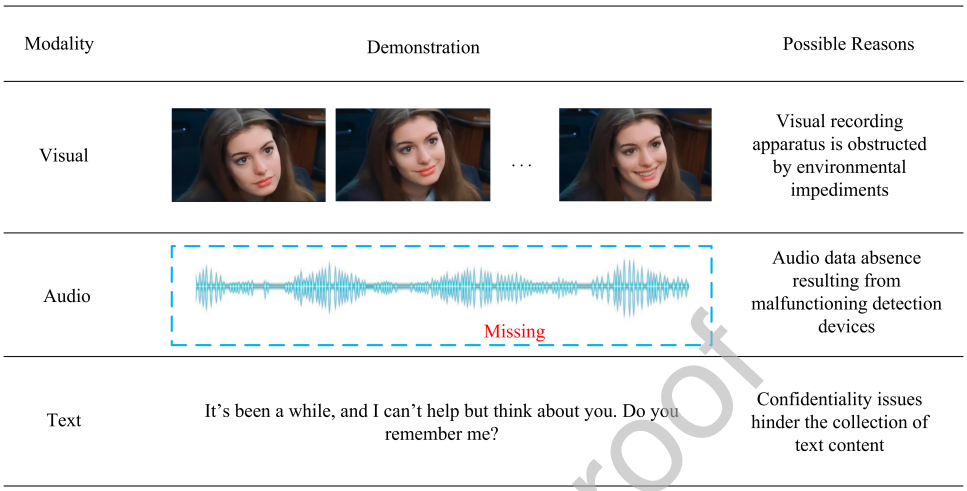


Fig. 1. Example of uncertain modalities missing in MSA

trained with public datasets do not have perfect performances for personalized users (which has been proved by some experiments). Therefore, how to develop an effective MSA model that can continuously learning personalized users' characteristics is a new challenging problem.

- For MSA under uncertain modalities missing, existing MSA models always adopt a pre-trained model (which is trained with the full modalities) to complete the missing modalities. However, in practical applications, it is difficult to collect high quality full multimodal data for getting the pre-trained model. Moreover, when data changes, the pre-trained model embedded in the MSA models will become invalid. Therefore, how to effectively tackle the problem of uncertain modalities missing in personalized MSA is still a challenging problem.

To tackle the above issues, this work proposes a Personalized MSA model under uncertain modalities missing via Pretraining and Online Learning (termed as PMSAPO). PMSAPO is first trained with some public datasets, thus to possess good preliminary MSA abilities for personalized users. Then, the pretrained model PMSAPO is enabled to continuously learn the personalized users with our proposed online learning techniques, thereby growing

into a robust model for personalized MSA. For MSA under uncertain modalities missing, PMSAPO first encodes the uncertain modalities missing (text, visual and audio) with multi-head attention, and then concatenates the embeddings of the three modalities. Next, the concatenated embeddings are input into the self-attention module to realize the internal interaction and generate the joint feature of the three modalities. Then, PMSAPO estimates the quality weight of each modality in the joint feature with a fully connected neural network, and produces the weighted modalities through weighting each modality with its quality weight. After that, the joint feature and each weighted modality are fused to generate the optimized fusion feature of each modality. Finally, the three optimized modalities are concatenated and the final sentiment analysis is conducted with the softmax function.

To enable the pretrained model PMSAPO to own self-learning ability when applied for personalized users, we proposed the following online learning techniques: firstly, we adopt an online meta-learning mode for pre-trained PMSAPO, which applying an inner and outer loop structure to enable PMSAPO to quickly learn features of personalized users. Secondly, we adjust the learning rate of PMSAPO adaptively with the increase of data of the personalized user. Thirdly, to address the catastrophic forgetting problem in online learning, we propose to assign dynamic weights for the sample data of the personalized user, which can guide the PMSAPO to focus on new samples while avoiding forgetting old ones. The main contributions of our work are summarized as follows:

- To tackle the issue of personalized MSA, we propose a personalized MSA model under uncertain modalities missing via pretraining and online learning (named PMSAPO), which is first trained with some public datasets, and then it autonomously adapts to personalized users through some online learning techniques. To the best of our knowledge, this is the first work that propose to develop personalized MSA model under uncertain modalities missing.

- To effectively handle the uncertain missing modalities in MSA for personalized users, we propose to complete the missing modalities with a method of reusing the fused modalities. In this method, a neural network evaluation module is proposed to assign different weights to each modality in the joint feature (fused modalities), and then the joint feature is fused with each weighted modality, so as to optimize each

modality by reusing the fused modalities, which has excellent flexibility and generalization ability.

- To enable the pretrained MSA model to autonomously adapt to the personalized users, we propose a set of online learning techniques. Specifically, the adaptive learning rate adjustment strategy can adjust the learning rate according to the amount of data of personalized users. The online meta-learning strategy can obtain more adaptive personalized user parameters by optimizing the meta-task loss. The adaptive weights assigning strategy endows different weights to data samples of personalized users, thus to avoid catastrophic forgetting problems.

The structure of this work is outlined as follows: Section 2 provides an overview of existing related works. Section 3 presents our pre-trained MSA model. Section 4 presents online learning strategy. Section 5 details our experiments and results analysis. Finally, Section 6 summarizes our contributions and outlines future research directions.

2. Related Work

We delve into representative studies that address the MSA problem under uncertain modalities missing.

2.1. MSA under Uncertain Modalities Missing

Recently, several outstanding MSA models have been proposed to solve the problem of uncertain modalities missing. These models can be broadly categorized into two main types: generative learning based models and joint learning based models. In the following, we review the representative works in each of these categories.

Generative learning based models: which propose to generate new data by analyzing the distribution of the available modalities. Zhou et al. [22] proposed a feature to enhance generator that is used to produce for missing modal correlation characteristics of the generator based on neural network end-to-end feature enhancements and depth. Zhang et al. [23] regard the learning of multiview latent representations as a degenerate process, and successfully achieve consistency and complementarity across different views. Pham et al. [24] the multimodal network (MCTN) cycle translation, through the loop between modal translation, learn to deal with missing modal robust coalition said. Research [25] based on the end-to-end translation put

forward by the modal characteristics of the fusion model (TransM), through the cycle of conversion between modal multimodal interaction. Liu et al. [18] proposed the MTMSA model to transform visual and audio modalities into text modalities to improve the quality of modalities to solve the problem of lack of modal uncertainty and improve the accuracy of sentiment analysis.

Joint learning methods: which propose to learn the joint representation by exploring the interactions between different modalities [26]. Work [27] enhances the accuracy of sentiment analysis through text-based reconstruction of missing information and guided fusion. Sun [12] introduced a strategy to mitigate the impact of missing modalities by incorporating analogous content from other datasets. Work [28] proposed a unified self-distillation framework (UMDF) for MSA that handles uncertain modalities missing by leveraging bidirectional knowledge transfer and multi-grained crossmodal interactions to enhance the robustness and performance of the model. Work [27] enhances the accuracy of sentiment analysis through text-based feature extraction, reconstruction of missing information, and guided fusion.

Although existing works for MSA under uncertain modalities missing are wonderful, the generalization ability of these models is weak, and these models do not have self-learning ability. Therefore, they cannot address the problem of personalized MSA under uncertain modalities missing.

3. MSA Model under Uncertain Modalities Missing with Pretraining (PMSAP)

In this section, we first present the problem studied in our work. Then, we introduce the details of our proposed model PMSAP. Finally, we introduce the pretraining process of our proposed model.

3.1. Problem Statement

Assume that the multimodal data for sentiment analysis contains three modalities: $P = [X_v, X_a, X_t]$, where X_v, X_a and X_t denote the visual, audio and text modality, respectively. Without loss of generality, we denote the missing modality by X_M^m, where $M \in \{v, a, t\}$. Table 1 outlines several possible scenarios of uncertain modalities missing. The first challenge addressed in this work is to develop a robust sentiment analysis model for the data set P while can handle uncertain modalities missing effectively. For convenience, in the following sections, we use $\{X_v^m, X_a, X_t\}$ to represent the multimodal data of an individual user with uncertain modalities missing.

Table 1: The seven possible scenarios of uncertain modalities missing.

Missing scenario	**Missing modality**	**Multimodal data**
No missing	/	$P = [X_v, X_a, X_t]$
Single missing	Visual	$P = [X_v^m, X_a, X_t]$
Single missing	Text	$P = [X_v, X_a, X_t^m]$
Single missing	Audio	$P = [X_v, X_a^m, X_t]$
Multiple missing	Visual & Text	$P = [X_v^m, X_a, X_t^m]$
Multiple missing	Audio & Text	$P = [X_v, X_a^m, X_t^m]$
Multiple missing	Visual & Audio	$P = [X_v^m, X_a^m, X_t]$

3.2. The structure of PMSAP

The structure of PMSAP is depicted as Fig2, which contains two modules, which are Fully Connected Neural Network Evaluation Module and Joint feature optimization module. Firstly, PMSAP generates the fused modality and allocate weights for each modality with the first module. Then, PMSAP completes the final sentiment classification based on the fused modality through the second module. Specifically, PMSAP deals with the uncertain modalities missing as follows:

PMSAP first encodes the uncertain missing modalities (text, visual and audio) with multi-head attention, and then it concatenates the embeddings of the three modalities to form the fused modality. Next, it inputs the fused modality into a self-attention module to realize the internal interaction and generate the joint feature of the three modalities. Then, PMSAP estimates the quality weight of each modality in the joint feature with a fully connected neural network, and produces the weighted modalities through weighting each modality with its quality weight. After that, the joint feature and each weighted modality are fused to generate the optimized fusion feature of each modality. Finally, the three optimized modalities are concatenated and the final sentiment analysis results are obtained with the softmax function. The details of the two modules of PMSAP are described as follows.

A. Fully Connected Neural Network Evaluation Module

In the Fully Connected Neural Network Evaluation module, we first use the multi-head attention mechanism to realize the coding of the three modalities, and then splice fusion and self-attention interaction to generate the joint

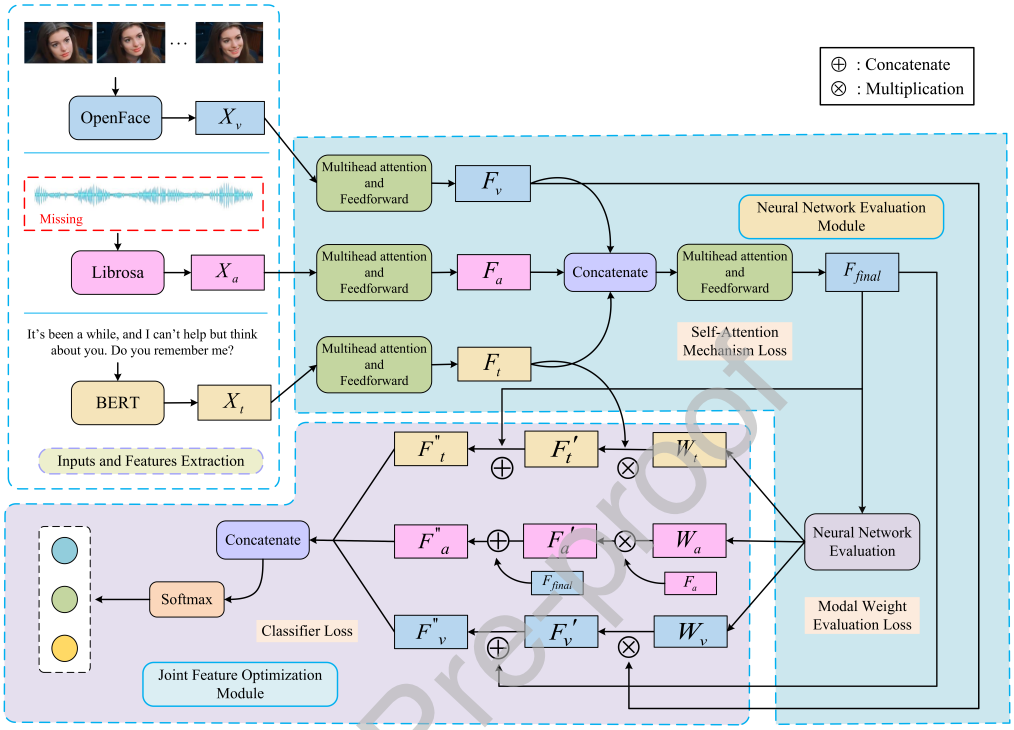


Fig. 2. The structure of PMSAP.

features. Then, we dynamically compute weights for each modality feature according to its quality with the linear transformation of the fully connected layer, the nonlinear transformation of the *tanh* activation function and the $softmax$ weight normalization of the joint features. By this way, PMSAP can adaptively adjust the weight of each modality according to the actual situation of the input data, so that PMSAP can still maintain high sentiment analysis accuracy when dealing with uncertain modalities missing.

With the multi-head attention mechanism and feed-forward neural network, the FCNN deeply mines the complex interaction in the multi-modal fusion features, and assigns weights to each modality based on the joint features, thus to improve the accuracy and robustness of the model. Next, we describe the implementation steps of the module in detail (which is illustrated as Fig. 3).

Firstly, we utilize the multi-head attention mechanism to encode the text, visual and audio modality, separately. Thus, to obtain the feature represen-

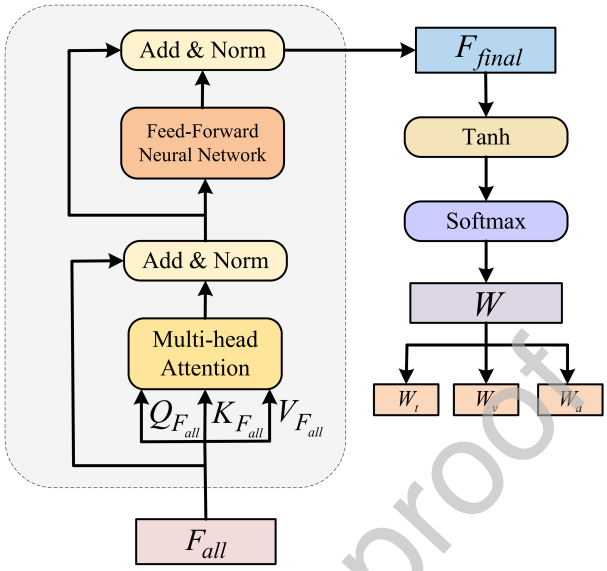


Fig. 3. The specific process of Fully Connected Neural Network Evaluation Module

tation of each modality. For the input features X_i of each modality i, the process of the multi-head attention mechanism can be expressed as Eq.(1):

$$\mathbb{Z}_i = MultiHead(X_i, X_i, X_i) \tag{1}$$

where $MultiHead$ represents the multi-head attention mechanism.

To further capture the nonlinear relationships within each modality, we use a Feedforward Neural Network (FFN) for deep feature extraction based on the output of the multi-head attention function. This process can be described as Eqs. (2):

$$F_i = FFN(Z_i) \tag{2}$$

After this step, we can obtain the feature representations of three modalities, which are denoted as F_t, F_a and F_v, respectively. Then, we concatenate the features of each modality to get a fusion feature matrix:

$$F_{all} = Concat(F_t, F_\alpha, F_\nu) \tag{3}$$

After obtaining the fused features, we apply the multi-head attention mechanism and feed-forward neural network to process the fused features, thus to realize deeper interaction between modalities and the joint features.

The process is described as Eqs.(4):

$$\mathrm{Z} = MultiHead(F_{all}, F_{all}, F_{all}) \tag{4}$$

$$F_{final} = FFN(Z) \tag{5}$$

Then, the concatenated features are transformed into a weight matrix through a linear transformation via a fully connected layer, enabling the model to learn the correlations between different modality features. Subsequently, the weight matrix undergoes a non-linear transformation with a $Tanh$ activation function, which allows the model to better capture the complex relationships among the features. Finally, the weights are normalized with the softmax function, thereby getting the weights according to the quality of each modality. This process can be described as Eqs. (6):

$$W = Soft\max(Tanh(F_{final}W_1 + b_1)) \tag{6}$$

where W_1 and b_1 are trainable parameters in the neural network. The output W is a modal weight matrix, which represents the quality of each modality, and each column of the matrix represents the corresponding weight of each modality. The $Tanh$ activation function (hyperbolic tangent function) is used to enhance the nonlinear expression ability of the neural network and limit the output value range between -1 and 1, so as to better capture the complex relationship in the data.

B. Joint Feature Optimization Module

Existing MSA models considering uncertain modalities missing always fulfill sentiment analysis based on the fusion of the three modalities, but without utilizing the fused modality to complete the uncertain missing modalities. Actually, the fusion of the three modalities usually contains richer emotional information than a single modality under uncertain modalities missing. Moreover, the fused modality absorb the sentiment information of each modality through the fusion interaction. Therefore, the fused modality is worth exploring and utilizing for completing the missing modalities. Thus, we propose a new method to better complete the uncertain missing modalities, that is, after fusing the three modalities, we further fuse the fused modality with the original three modalities. By re-utilizing the fused modality, it not only can complete the missing modalities, but also can enhance the quality of each modality.

In the Joint Feature Optimization Module, we first multiply the features of each modality by the corresponding weights evaluated by the fully connected neural network to obtain the weighted features of each modality, this can enable the better modality to play a greater role in features' fusion. The process for getting the weighted features can be described as follows:

$$F_t' = F_t \times W_t F_a' = F_a \times W_{\mathrm{a}} F_\nu' = F_\nu \times W_\nu \tag{7}$$

where F_t, F_a and F_v are the original features of each modality. W_t, W_a and W_v are obtained by separating from each column in the matrix W and represent the corresponding weights of each modality.

Then, the joint features are fused with the weighted feature of each modality, and the modality optimization is realized by deep interaction through the fully connected layer, which can be presented as follows:

$$\begin{aligned} F_t'' &= Concat(F_t', F_{final}) \\ F_a'' &= Concat(F_a', F_{final}) \\ F_\nu'' &= Concat(F_\nu', F_{final}) \end{aligned} \tag{8}$$

Next, the optimized modalities were fused to further deepen the interaction of each modality information with Eqs. (9):

$$F_{final}'' = Concat(F_t'', F_a'', F_\nu'') \tag{9}$$

Finally, the F_{final}'' is passed to the softmax function, and the final sentiment classification $\hat{y}$ can be obtained with Eqs.
(10).

$$\hat{y} = \mathrm{Softmax}(F_{final}'') \tag{10}$$

C. Training Objective

For model PMSAP, the total loss is composed of the loss of multiple modules, including loss of the attention mechanism, loss of weight evaluation for modality, loss of classification and loss of L2 Regularization. Each loss function will be introduced as follows.

(1) Loss of the attention mechanism ($\boldsymbol{L}_{attention}$): Within each modality, we use the multi-head attention mechanism to capture the context information within the modality. The mechanism of different modal characteristics of the query, the key and the value operation are helpful for generating

better semantic representation features. The loss of the attention mechanism is calculated with the cross entropy, which is defined as Eqs. (11):

$$L_{attention} = \sum_{h=1}^{H} L_{head_h} \tag{11}$$

where H represents the number of heads of multi-head attention, L_{head_h} represents the loss of the $h-th$ attention head.

(2) Loss of weight evaluation for modality ($\boldsymbol{L}_{weight}$): To allocate weights for multimodal fusion, we proposed the fully connected neural network evaluation module. The loss of this module is defined as Eqs. (12):

$$L_{weight} = -\sum_{h=1}^{H} p_i \log(\hat{p}_i) \tag{12}$$

where p_i means the actual weight distribution, $\hat{p}_i$ denotes the modal weight predicted by the module. By optimizing this loss, we are able to obtain the optimal weights for each modality for features fusion.

(3) Loss of L2 Regularization ($\boldsymbol{L}_{regularization}$): To prevent overfitting of the model during training, the L2 regularization is adopted. L2 regularization works by adding a penalty term of sum of squares to all the trainable parameters of the model. Loss of L2 Regularization is defined as Eqs. (13):

$$\boldsymbol{L}_{regularization} = \sum_{j} \|\boldsymbol{\theta}_j\|_2^2 \tag{13}$$

where θ_j denotes the j-th trainable parameter in the model.

(4) Loss of Classification ($\boldsymbol{L}_{cls}$): The F''_{final} is fed into a fully connected network with softmax activation function for the sentiment classification, and generating prediction score $\hat{y}$, which is obtained with Eq. (14):

$$\hat{y} = \text{softmax}\left(W_c F''_{final} + b_c\right) \tag{14}$$

where W_c and b_c are the learned weights and biases, respectively. In this work, we employ the standard cross-entropy loss for classification, so, the loss function of classification is defined as Eq. (15):

$$\boldsymbol{L}_{cls} = -\frac{1}{N}\sum_{n=1}^{N} y_n \log \hat{y_n} \tag{15}$$

where N indicates the number of samples, y_n means the true label of the n^{th} sample, and $\hat{y}_n$ denotes the predicted label.

(5) The total loss function ($\boldsymbol{L}_{total}$): the total loss function of the model is the weighted sum of the losses of modules, which is defined as Eqs. (16):

$$\boldsymbol{L}_{total} = \boldsymbol{L}_{attention} + \boldsymbol{L}_{cls} + \boldsymbol{L}_{weight} + \boldsymbol{\lambda} \cdot \boldsymbol{L}_{regularization} \tag{16}$$

where λ is the weight hyperparameter of the regularization term. In our experiments, we tune the loss of each part and finally obtain the best performance. Moreover, the whole calculation process of the PMSAP model is described as Algorithm 1.

D. Pretraining with public datasets

For MSA for a personalized user, the amount of multimodal data about the personalized user is relatively small, making it difficult to train a better MSA model for the personalized user. Inspired by the idea of pretraining of Large Models, we propose to pretrain our proposed model (introduced in Section 3.2.) with some public datasets (e.g.), thus to enable the model to learn rich and generalizable feature representations, and thus to have preliminary MSA ability for personalized MSA [29] (which is named as PMSAP). This not only can reduce the cost for collecting large-scale multimodal data about the personalized user, but also can provide a model with good initial ability for personalized MSA. The training process for our proposed model is described as follows:

4. Personalized MSA via Online Learning

Although model PMSAP has preliminary MSA ability, but its performance is not perfect in personalized MSA, which has been proved with some experiments (which will be presented in the experiment Section). Therefore, it needs to enable PMSAP to autonomously learn the characteristics of the personalized users. To solve this problem, we propose to adjust the parameters of pretrained model PMSAP through a series of online learning techniques, so as to get an effective model for personalized MSA (named as PMSAPO). In the following sections, we will introduce the online learning strategies in details.

A. Adaptive adjustment of the learning rate

Firstly, with the continuous use of PMSAP for personalized users, the learning rate of PMSAP is gradually reduced, so that the model can quickly adapt

Algorithm 1: : Personalized Multimodal Sentiment Analysis under Uncertain Modalities Missing

Input: The complete multimodal data: $[X_v, X_t, X_a]$, where $X_m \in \mathbb{R}^{l_m \times d_m}, m \in \{v, t, a\}$;
Multimodal data with uncertain modalities missing : $[X_v^M, X_t, X_a]$, where $X_m^M \in \mathbb{R}^{l_m \times d_m}, m \in \{v, t, a\}$, M denotes missing modality.
Output: The predicted sentiment category $\hat{y}$

Phase I. In the Fully Connected Neural Network Evaluation Module
Encoder: Produce the encoded representation F_v,F_t,F_a of each modality according to Eqs. (1)-(3)
$F_v \leftarrow \text{encoder}\,(X_v^M, X_v^M, X_v^M)$
$F_t \leftarrow \text{encoder}\,(X_t, X_t, X_t)$
$F_a \leftarrow \text{encoder}\,(X_a, X_a, X_a)$
Concatenate the encoded representations of all modalities.
$F_{all} \leftarrow [F_v||F_a||F_t]$
Apply multi-head attention and feed-forward networks to the concatenated representation.
$F_{all} \leftarrow \text{multihead\_attention}\,(F_{all}, F_{all}, F_{all})$
$F_{final} \leftarrow \text{ff}\,(F_{all})$
Compute the attention weights for each modality.
$W_m \leftarrow \text{Dense}(F_{all}, \text{units} = 3, \text{activation} = \tanh)$
$W_m \leftarrow \text{softmax}(W_m, \text{axis} = 2)$
$W_v \leftarrow W_m[:, : l_v, 0 : 1]$
$W_a \leftarrow W_m[:, l_v : l_v + l_a, 1 : 2]$
$W_t \leftarrow W_m[:, l_v + l_a :, 2 : 3]$
Phase II. In the Joint feature optimization module
Weight the encoded representations of each modality.
$F_v' \leftarrow F_v \times W_v$
$F_a' \leftarrow F_a \times W_a$
$F_t' \leftarrow F_t \times W_t$
Concatenate the encoded representations and the flag data.
$F_v'' \leftarrow \text{concat}\,([F_v', F_{final}])$
$F_a'' \leftarrow \text{concat}\,([F_a', F_{final}])$
$F_t'' \leftarrow \text{concat}\,([F_t', F_{final}])$
Concatenate the projected representations.
$F_{final}'' \leftarrow [F_v''||F_a''||F_t'']$
Phase III. Predict the Sentiment Category
$\hat{y} \leftarrow \text{softmax}(F_{final}'')$
Return $\hat{y}$
End

to personalized users when the amount of user data is small. This strategy can enable the model to learn more detailed characteristics of the personalized user when the amount of user data is large. Moreover, during the application process, the model calculates the loss with the current batch of data, and dynamically adjusts the learning rate based on changes in the loss, so as to avoid getting stuck in local optimal due to continuously decreasing learning rates. By combining the loss changes, this strategy enables more precise control of the learning rate at different stages of user usage, thereby improving the model's optimization effectiveness and performance. Overall, this strategy ensures that the learning rate decreases adaptively with the application of PMSAP, allowing the model to automatically adapt to the personalized user.

Specifically, the initial learning rate is set to α_0, which decreases as the number of learning steps increases, the formula is defined as Eqs. (17):

$$\alpha = \alpha_0 \beta^{\frac{t}{T}} \tag{17}$$

where β is the decay factor of the learning rate (0¡β¡1), t indicates the current number of steps, and T means the preset step interval. This formula ensures that the learning rate gradually decreases over time, so that when the amount of user data is limited, a larger learning rate is employed to enable the model to rapidly adapt to new users. As the amount of user data grows, the learning rate adaptively decreases, allowing the model to capture more nuanced and detailed characteristics of the user.

The adjustment of the combined loss is shown below: through the evaluation of the loss, if the loss of the current batch L_{post} is significantly better than the loss of the previous batch L_{pre}(i.e., $L_{\text{post}} < L_{\text{pre}} \times (1-\theta)$, where θ is a preset performance improvement threshold.), the model will keep the current learning rate and continue to optimize. If the current loss L_{post} does not show significant improvement compared to the previous loss L_{pre}(i.e., $L_{\text{post}} >= L_{\text{pre}} \times (1-\theta)$), the learning rate is appropriately increased to enable larger exploration. The formula for increasing the learning rate is defined as Eqs. (18):

$$\alpha = \alpha \times \eta \tag{18}$$

where η denotes the increasing factor of the learning rate.

Moreover, if the loss does not improve significantly for N consecutive steps (where N is a predefined patience value), the learning rate is halved to allow the model to reintroduce some exploration capability, thus to prevent

the model from getting stuck in local optima. The formula for this operation is defined as Eqs. (19):

$$\alpha = \alpha \times 0.5 \tag{19}$$

Regarding the gradient accumulation aspect, considering the real-time requirements of online learning, the model employs a strategy that combines gradient accumulation with step size updating Specifically, after every N batches, the accumulated gradients are used to update the model's parameters. This process is described as formula (20):

$$\theta_{t+1} = \theta_t - \frac{\alpha}{N}\sum_{i=1}^{N} \nabla_\theta L_i \tag{20}$$

where θ means the model's parameter, L_i is the loss function of the i batch data. Through this gradient accumulation strategy, the fluctuation of each parameter can be effectively reduced and the stability of online learning can be improved.

B. Online meta-learning method

To further enhance the generalization and adaptability of model PMSAP and to ensure more effective continuous learning and updating, we propose an online meta-learning method. Meta-Learning is a machine Learning method with the main idea as "Learning to Learn". Meta-learning tries to get the optimal parameter initialization point by allowing the model to learn from multiple tasks, so that the model can quickly converge on new tasks, so as to quickly adapt to new tasks and improve the generalization ability and learning efficiency of the model. However, meta-learning cannot conduct online learning. For this issue, we improve meta-learning through the following operations: treating different batches of personalized user data as multiple tasks in meta-learning, treating each application of PMSAP as a new task, and adjusting the related structure to make meta-learning to realize online learning. Through the online meta-learning strategy, PMSAP can learn to learn and quickly adapt to the personalized users.

In this work, the online meta-learning process is divided into an inner loop and an outer loop. In the inner loop, parameters are updated based on the loss for each batch of data, followed by another loss calculation to obtain the relevant gradients, known as meta-gradients. However, the parameters are not updated immediately after obtaining the meta-gradients; instead, these gradients are accumulated over multiple batches. When the

accumulated gradients reach a preset number, the outer loop is triggered, where the global parameters are updated with the meta-optimizer based on the accumulated meta-gradients. After the updating, the accumulated gradients are reseted to zero for the next accumulation cycle. The inner loop focuses on the model's rapid adaptation to each small batch of samples, while the outer loop is responsible for optimizing the model's overall performance across multiple tasks. This nested structure of inner and outer loops enables the model to learn quickly in complex tasks while maintaining good generalization capabilities.

Specifically, we assume that the input data for the model PMSAP is the multimodal dataset $(x_i^m, y_i)_{i-1}^N$. Here, x_i^m represents the m modality input of the i sample, y_i denotes its corresponding sentimental label, m indicates the number of modalities, and N is the total number of samples.

The meta-loss is obtained through the standard forward propagation process of the model. The loss function is defined as Eq. (21):

$$L_{task} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) \tag{21}$$

where $\hat{y}_i$ represents the final sentiment classification generated.

Let assume that the parameters at step t is θ_t, and the current meta-loss for data is L_{task}^t. Then, the meta-gradient can be calculated with Eq. (22):

$$\mathbf{g_t} = \nabla_{\theta_t} \mathbf{L}_{task}^t \tag{22}$$

The procedure for aggregating meta-gradients across multiple batches is described as Eq. (23).

$$G_{meta} = \sum_{t=1}^{T} \mathbf{g}_t \tag{23}$$

The global parameters θ are subsequently updated by incorporating the accumulated gradients with Eq. (24):

$$\theta = \theta - \beta G_{meta} \tag{24}$$

where β is the meta-learning rate.

C. Assigning weights to different samples

In this work, to avoid the catastrophic forgetting problem in online learning,

we propose the strategy to assigning weights to different samples. That is, we assign a weight s_i to each sample data of the personalized users, which is initially set as 1 and then gradually increases it over time, which is described as Eq. (25)

$$s_i = 1 + 0.01 \cdot \frac{t - t_0}{T} \tag{25}$$

where t_0 means the time step at which the model starts online learning, and T denotes the total training time step.

The loss function based on the sample weight is defined as Eq. (26)

$$\boldsymbol{L}_{weighted} = \sum_{i=1}^{N} \boldsymbol{s}_i \cdot \boldsymbol{L}_{task}^{i} \tag{26}$$

According to our proposed strategy, the new and old samples will have different weights to represent the different attention of the new and old samples, and the weights are not too different to better balance the new and old samples. Finally, the weights of the old and new samples were fused during the parameters' updating, so as to better balance the contributions of the old and new samples. Therefore, this strategy allows model PMSAP to pay more attention to the current samples without forgetting the previous samples too much.

The combination of pretraining and online learning strategy ensures that the model can acquire general knowledge from large-scale public datasets while making fine-grained adjustments of the personalized users, thus to ultimately achieve superior performance in personalized multimodal sentiment analysis.

5. Experiments

To assess the effectiveness of PMSAPO, we performed comprehensive experiments on three widely recognized public datasets: IEMOCAP[30], MELD[21] and CMU-MOSI[31]. In the subsequent sections, we first introduce these datasets and the data preprocessing procedures. We then describe the experimental setup and the baseline models used for comparison. Next, we introduce the experiments on proving the performance of PMSAPO in dealing with uncertain modalities missing. Then, we describe the experiments on proving the performances of PMSAPO in dealing with personalized MSA. Finally, the ablation experiments are presented and analyzed.

5.1. Benchmark Datasets

In our experiments, three MSA benchmark datasets (IEMOCAP, MELD and CMU-MOSI) are used to verify the performance of our proposed model PMSAPO. These three datasets and their features and processing process are explained in detail as follows.

IEMOCAP[30]: IEMOCAP is a recorded video dataset consisting of 5 binary dialogue sessions, each of which has about 30 videos and contains at least 24 utterances. Each sample in this dataset is labeled with emotion labels such as neutral, frustration, anger, sadness, happiness, excitement, surprise, fear, and disappointment. In our experiments, a binary classification experiment is conducted in IEMOCAP to map sentiment labels into positive and negative. Specifically, negative sentiment labels include frustration, anger, sadness, fear, and disappointment, while positive sentiment labels include happiness and excitement.

MELD[21]: MELD contains 13,708 utterances from 1,433 conversations from the Friends TV series. Each conversation was annotated with sentiment labels such as happy, angry, sad, neutral, surprise, disgust, and fear. In our experiments, three classification experiments are conducted in MELD, so the sentiment labels are mapped as negative, neutral and positive in our experiments. Specifically, negative sentiment labels include anger, sadness, disgust, and fear, while positive sentiment labels include happiness and surprise.

CMU-MOSI[31]: The Carnegie Mellon University Multimodal Opinion Sentiment and Intensity (CMU-MOSI) dataset consists of 2,199 monologue video segments extracted from 93 YouTube movie reviews. Each segment is annotated with a sentiment score ranging from -3 to 3, capturing a fine-grained spectrum of sentimental expressions. In our experiments, we conducted binary classification on CMU-MOSI, specifically focusing on negative and positive labels. Table 2 presents a detailed breakdown of the class distributions for three datasets.

5.2. Preprocessing of Benchmark Datasets

We adopt the method used in [32] to conduct multi-modal features extraction of the three benchmark datasets IEMOCAP, MELD and CMU-MOSI. Next, we introduce the feature extraction methods of each modality in detail.

Visual Modality Feature Extraction: We employed the OpenFace2.0 toolkit [33] to extract a comprehensive set of 709-dimensional facial features.

Table 2: Distribution of different sentimental label counts in the IEMOCAP, MELD and CMU-MOSI datasets

Category	Partition	Neg.	Neu.	Pos.
Two-classes (IEMOCAP)	Train Set	3219	-	1299
	Test Set	859	-	337
Three-class(MELD)	Train Set	2945	4708	2333
	Test Set	833	1256	521
Two-classes(CMU-MOSI)	Train Set	914	-	996
	Test Set	108	-	42

Note: "Partition" represents Dataset Partition, "Neg." represents negative labels, "Neu." represents neutral labels, "Pos." represents positive labels. "-" indicates that the corresponding category is not applicable in the two-classes.

These features encompass temporal information, confidence metrics, recognition flags, ocular movements, head orientation, and facial dynamics, providing a rich representation of facial characteristics.

Text Modality Feature Extraction: To leverage the power of transfer learning, we utilized a pre-trained BERT model [34] to generate 768-dimensional textual embeddings. These embeddings capture rich semantic and contextual information, enabling effective representation of textual content.

Audio Modality Feature Extraction: The audio data preprocessing included mono-mixing and resampling to 16 kHz. For each 512-sample frame, we extracted a 33-dimensional feature vector, which combines the zero-crossing rate, Mel-frequency Cepstral Coefficients (MFCC), and Constant Q Transform (CQT) features. This extraction process was facilitated using the Librosa library [35], ensuring high-quality feature representation.

To ensure a fair performance comparison, we preprocessed the data provided by [32] with the same preprocessing steps for all baseline models and our proposed model. This standardized approach helps in maintaining consistency and reliability in models' evaluation.

5.3. Experiment Settings

Our experiments were carried out on a high-performance computing system with the following configuration: Windows 10 operating system, Intel®

Table 3: Parameter settings of PMSAPO

Description	Symbol	Value
Epoch number	e	15
Missing rate	η	[0.1-0.5]
Hidden size	d	300
Batch size	b	32
Maximum video length	n_v	100
Maximum audio length	n_a	150
Maximum text length	n_t	25
Loss weights	λ	0.1

Core™ i9-10900K CPU, NVIDIA RTX 3090 GPU, and 96GB of RAM. The model was implemented with TensorFlow version 1.14.0 and Python 3.7. The hyperparameters used in our experiments are summarized in Table 3.

To evaluate the performance of our proposed model, two widely-recognized evaluation metrics are adopted, which are accuracy (Acc) and macro-F1 score (M-F1), and are defined as Eq. (27) and Eq. (28):

$$\text{Acc} = \frac{N_{true}}{N} \tag{27}$$

$$\text{M-F1} = \frac{2PR}{P+R}. \tag{28}$$

where N_{true} represents the number of correctly classified samples, N denotes the total sample size, P signifies the precision and R indicates the recall value.

5.4. Baseline Models

To illustrate the effectiveness of PMSAPO, we selected twelve state-of-the-art models as baseline models, which are introduced as follows:

- AE [36]: A robust framework for analyzing both linear and non-linear self-encoding architectures, aimed at optimizing the consistency between input and output in neural networks.

- CRA [37]: An advanced technique for reconstructing missing modalities using cascaded residual autoencoders, which enhance input approximation through residual connections.
- MCTN [24]: A novel method that employs inter-modal translation to strengthen cross-modal interactions and joint relationship learning.
- TransM [25]: An innovative end-to-end translation-based model for multimodal feature fusion, facilitating dynamic interactions between modalities through cyclic translation mechanisms.
- MMIN [38]: A highly refined feature reconstruction model is proposed to address the challenge of missing modalities. This model deploys cascaded residual auto-encoders in conjunction with bidirectional imagination modules to carry out cross-modal transformations.
- ICDN [17]: An integrated approach that combines consistency and difference networks, leveraging cross-modality Transformers to map information effectively across different modalities.
- MRAN [39]: A state-of-the-art model that employs multimodal and missing index embeddings to guide feature reconstruction, aligning audio-visual features with textual data to address modality absence.
- TATE_C [32]: An advanced tag-assisted Transformer encoder designed to manage uncertain modalities, integrating pre-trained models to enhance joint representation learning.
- MTMSA [18]: A modality translation method that converts visual and audio into text modality, effectively managing missing modalities and capturing deep cross-modal interactions.
- TATE_J [40]: An improved version of TATE_C, introducing modality-specific weighting schemes to maximize the utilization of each modality's unique characteristics.
- TgRN [27]: A text-guided method that enhances the accuracy of sentiment analysis through text-based feature extraction, reconstruction of missing information, and guided fusion.

- SMCMSA [12]: In scenarios with missing modalities, this model selects similar samples from a pre-constructed comprehensive modality sample database to impute the missing data.

Table 4: Performances of models when missing a single modality on the IEMOCAP dataset (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
IEMOCAP	AE	76.15	82.09	75.24	80.26	75.02	78.01	73.92	77.43	70.19	76.01	67.27	76.43
	CRA	77.05	82.13	75.95	80.97	75.13	78.09	74.02	78.11	70.69	76.12	67.75	76.49
	MCTN	78.57	82.27	77.74	81.02	75.37	78.27	74.69	78.52	71.75	76.29	68.17	76.63
	TransM	79.57	82.64	78.03	81.86	76.33	80.43	75.83	78.64	72.01	77.27	68.57	76.65
	ICDN	77.37	82.81	76.46	81.34	74.13	80.56	65.00	78.04	73.26	75.17	60.50	73.35
	MRAN	81.21	85.98	81.06	84.88	80.61	84.38	79.99	83.51	78.63	82.90	75.82	81.33
	MMIN	80.83	83.43	78.85	82.58	77.09	81.27	76.63	80.43	72.81	78.43	70.58	77.45
	TATE_C	81.15	85.39	79.99	85.09	79.10	84.07	78.45	83.25	76.74	82.75	74.43	82.43
	MTMSA	81.36	86.14	81.81	85.24	81.47	84.46	80.20	84.28	79.53	82.94	75.84	82.55
	TATE_J	81.76	86.46	81.25	85.24	79.57	84.80	78.06	83.76	77.84	82.97	75.76	82.51
	TgRN	82.53	86.38	81.87	84.98	80.33	83.49	79.21	83.54	77.47	82.97	76.21	82.17
	SMCMSA	84.17	87.84	82.10	85.64	81.92	85.05	80.21	84.80	79.99	83.78	77.64	83.36
	Ours	**85.21**	**89.16**	**84.72**	**88.51**	**83.69**	**87.45**	**82.97**	**87.31**	**82.47**	**86.33**	**80.09**	**84.90**

Table 5: Performances of models when missing a single modality on the MELD dataset (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
MELD	AE	56.27	61.75	55.62	60.47	54.96	59.77	54.57	59.42	52.99	57.76	51.39	56.61
	CRA	58.51	63.63	56.61	61.68	55.01	61.17	53.92	59.74	52.06	58.20	52.11	55.84
	MCTN	57.93	62.44	56.31	61.82	56.29	61.33	55.45	61.26	54.76	59.48	53.96	57.41
	TransM	57.18	61.33	56.66	61.27	54.42	60.54	54.17	59.21	54.07	58.96	53.98	57.64
	ICDN	56.01	62.29	53.88	60.69	55.89	61.97	55.48	61.26	54.33	59.57	53.41	58.54
	MRAN	58.26	63.24	56.23	61.64	55.26	61.51	53.47	59.82	53.84	59.08	53.06	57.86
	MMIN	58.46	63.14	56.89	61.65	55.13	61.19	53.85	60.68	52.07	59.62	49.25	56.97
	TATE_C	59.09	64.31	58.43	63.61	57.96	62.80	56.52	61.67	55.29	59.39	55.21	59.71
	MTMSA	58.47	64.19	57.29	62.81	57.15	62.50	55.12	62.11	54.08	60.88	53.95	60.58
	TATE_J	64.11	66.53	61.96	65.31	61.49	65.01	60.77	63.80	60.28	63.52	59.18	62.61
	TgRN	63.87	67.97	62.87	67.33	62.57	66.33	60.48	63.57	61.95	64.33	58.67	61.84
	SMCMSA	63.78	68.79	62.94	67.81	62.49	67.00	62.06	66.15	61.28	64.51	60.11	64.06
	Ours	**66.78**	**70.63**	**66.38**	**69.50**	**65.68**	**68.75**	**64.83**	**67.97**	**63.16**	**66.84**	**61.76**	**65.36**

5.5. Performance of PMSAPO in Dealing with Uncertain Modalities Missing with 2-classification and 3-classification

To assess the efficacy of the PMSAPO model in resolving the MSA problem under uncertain modalities missing in less-classification, we conduct experiments by implementing 2-classification (negative and positive) on the

Table 6: Performances of models when missing a single modality on the CMU-MOSI dataset (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
CMU-MOSI	AE	74.65	83.59	73.74	78.76	72.25	76.53	71.47	75.99	70.96	74.51	65.77	72.93
	CRA	75.51	83.63	74.56	82.47	73.63	76.59	72.57	75.61	71.16	74.63	66.25	71.99
	MCTN	77.07	83.66	76.21	82.52	73.79	76.44	73.19	75.92	72.15	74.73	66.79	73.86
	TransM	81.17	84.41	76.57	83.67	74.39	81.69	74.03	79.81	70.51	78.77	67.67	75.91
	ICDN	75.87	84.31	74.69	82.14	72.36	82.07	71.51	79.45	71.26	75.67	69.00	72.78
	MRAN	82.75	85.87	82.52	85.38	79.11	81.07	78.92	80.01	77.33	79.40	74.32	78.38
	MMIN	82.33	84.36	77.53	84.18	78.59	82.66	77.31	79.39	74.13	78.38	72.58	75.59
	TATE_C	82.99	85.87	80.49	83.68	80.60	83.57	79.95	82.57	78.24	80.51	76.93	79.39
	MTMSA	84.97	86.33	81.48	83.66	81.56	83.44	80.69	82.66	78.24	80.47	75.95	80.05
	TATE_J	84.33	86.09	82.73	84.90	81.67	83.88	78.36	81.90	76.52	81.09	75.82	79.09
	TgRN	84.78	86.36	82.79	84.77	81.33	83.33	79.66	82.78	77.01	80.98	76.32	79.98
	SMCMSA	85.37	86.96	83.09	85.63	82.41	84.41	81.74	83.84	79.47	82.78	78.67	80.69
	Ours	**86.48**	**88.51**	**85.15**	**88.07**	**83.97**	**86.68**	**83.69**	**85.06**	**81.98**	**83.86**	**80.17**	**82.83**

IEMOCAP and CMU-MOSI datasets, and 3-classification (negative, neutral and positive) on dataset MELD. In this experiment, the online learning strategies are not executed, so, we use PMSAP to denote our proposed model. This experiment consists of two parts, the first part considered the case of single modality missing, and the second part considered the case of multiple modalities missing. Experimental results of the 12 baseline models on dataset IEMOCAP are picked from work [12], experimental results of the 12 baseline models on dataset MELD are obtained by reproducing all models. In the following sections, we will introduce the two experimental segments in detail.

Experiment on single modality missing. In this experiment, the modality missing rate is set to 0, 0.1, 0.2, 0.3, 0.4 and 0.5, respectively. Experimental results are presented in Table 4, Table 5 and Table 6.

From Table 4 we can find that, on dataset IEMOCAP, our proposed PMSAP model outperforms the other 12 baseline models in the two evaluation metrics (ACC and M-F1) on all missing rates (0, 0.1, 0.2, 0.3, 0.4 and 0.5). Especially, compared with other baseline models, when the missing rate is 0.2, our proposed model PMSAP increase the metric M-F1 from 1.77% to 8.67%, and improve the values of ACC from 2.40% to 9.44%.

From Table 5 we can find that, on dataset MELD, for all the missing rates (0, 0.1, 0.2, 0.3, 0.4 and 0.5), PMSAP outperforms the other 12 baseline models in terms of the two evaluation metrics ACC and M-F1. Especially, compared with other baseline models, when the missing rate is 0.5, our pro-

Table 7: Performance of models when missing multiple modalities on the IEMOCAP dataset (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
IEMOCAP	AE	76.15	82.09	75.07	79.84	74.20	76.91	71.55	76.07	69.73	75.16	67.15	75.22
	CRA	77.05	82.13	75.21	79.95	74.22	77.03	71.86	76.41	70.13	75.29	67.31	75.42
	MCTN	78.57	82.27	76.83	80.56	74.77	77.89	72.27	77.03	71.02	75.84	67.51	75.88
	TransM	79.57	82.64	77.21	81.13	75.87	79.01	72.36	78.15	71.38	76.88	68.02	76.04
	ICDN	77.37	82.81	72.56	79.25	71.73	78.99	69.94	77.17	69.59	74.65	68.98	73.26
	MRAN	81.21	85.98	80.22	85.07	79.86	83.60	79.14	82.89	75.80	81.25	68.61	78.30
	MMIN	80.83	83.43	78.02	82.23	76.38	79.53	73.05	79.02	71.22	77.27	69.39	77.01
	TATE_C	81.15	85.39	78.37	83.63	77.55	82.33	76.14	82.21	74.09	81.94	72.49	80.57
	MTMSA	81.36	86.14	80.28	85.17	80.39	84.12	79.30	83.85	76.07	83.07	74.80	82.03
	TATE_J	81.76	86.46	80.67	85.30	79.12	83.77	78.99	84.64	78.44	82.75	76.97	82.25
	TgRN	82.53	86.38	80.33	83.87	78.97	81.66	78.09	81.03	77.02	80.33	75.33	79.68
	SMCMSA	84.17	87.84	81.55	85.81	80.55	84.38	78.29	82.93	74.47	80.49	77.35	82.85
	Ours	**85.21**	**89.16**	**84.03**	**88.26**	**83.47**	**87.33**	**82.19**	**87.12**	**81.88**	**85.74**	**78.11**	**84.08**

posed model PMSAP increase the metric M-F1 from 1.65% to 12.51%, and improve the values of ACC from 1.30% to 9.52%.

In addition, From Table 6 we can find that on the CMU-MOSI dataset, for all the missing rates (0, 0.1, 0.2, 0.3, 0.4 and 0.5), PMSAP outperforms the other 12 baseline models in terms of the two evaluation metrics ACC and M-F1. Therefore, based on the experimental results in Table 4, Table 5 and Table 6 can be concluded that our proposed model PMSAP is effective in solving the problem of MSA when a single modality missing. Compared to the second-best performing model (SMCMSA), on the IEMOCAP dataset, our proposed model PMSAP has an average increase of 2.19% in M-F1 and 2.20% in ACC. On the MELD dataset, PMSAP increase M-F1 by 2.66% on average, and improves ACC by 1.79% on average. On the CMU-MOSI dataset, PMSAP increase M-F1 by 1.78% on average, and improves ACC by 1.79% on average.

Experiment on multiple modalities missing. In this experiment, multiple modalities uncertain missing are generated by randomly setting modality missing rates (0, 0.1, 0.2, 0.3, 0.4 and 0.5) for three modalities in datasets IEMOCAP and MELD. Experimental results are presented in Table 7, Table 8 and Table 9.

From Table 7 we can find that, on dataset IEMOCAP, for all the missing rates (0, 0.1, 0.2, 0.3, 0.4 and 0.5), our proposed model PMSAP obtains the best results of ACC and M-F1 among the 12 baseline models. And from Table 8 and Table 9 we can find that, on the MELD dataset and CMU-MOSI

Table 8: Performance of models when missing multiple modalities on the MELD dataset (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
MELD	AE	56.27	61.75	55.00	60.36	54.78	59.51	54.18	58.48	52.74	57.37	50.64	55.34
	CRA	58.51	63.63	55.63	61.13	54.35	59.96	53.47	58.77	51.91	57.18	51.07	55.56
	MCTN	57.93	62.44	56.29	61.63	56.23	60.91	55.28	59.29	54.68	58.85	52.19	55.38
	TransM	57.18	61.33	56.25	61.13	56.07	60.36	54.09	58.44	53.32	57.19	53.10	56.77
	ICDN	56.01	62.29	53.85	60.43	54.76	59.61	53.36	59.02	52.44	57.71	51.48	57.23
	MRAN	58.26	63.24	56.14	61.63	55.15	60.67	53.07	59.11	53.76	58.91	51.63	56.34
	MMIN	58.46	63.14	56.49	61.52	54.23	60.94	53.27	60.02	51.54	59.53	48.76	56.58
	TATE_C	59.09	64.31	58.13	63.14	57.89	62.19	55.91	61.54	54.62	58.51	53.87	57.53
	MTMSA	58.47	64.19	57.14	62.31	56.32	62.38	54.60	61.51	53.54	60.85	52.59	59.63
	TATE_J	64.11	66.53	61.03	65.28	59.53	64.96	58.29	63.52	57.13	62.19	56.22	60.90
	TgRN	63.87	67.97	61.67	65.02	60.33	64.33	58.87	62.17	57.23	61.58	56.33	60.18
	SMCMSA	63.78	68.79	62.29	66.31	61.46	65.63	60.44	64.68	57.37	62.74	56.91	62.18
	Ours	**66.78**	**70.63**	**66.21**	**69.33**	**64.74**	**68.44**	**64.02**	**67.22**	**62.05**	**66.47**	**60.76**	**65.08**

dataset, for all the missing rates (0, 0.1, 0.2, 0.3, 0.4 and 0.5), our proposed PMSAP model outperforms the other 12 baseline models on two evaluation metrics (ACC and M-F1).

Compared to the second-best performing model (SMCMSA), on dataset IEMOCAP, our proposed model PMSAP improves the value of M-F1 by 3.09% on average, and increase the value of ACC by 2.90% on average. On the MELD dataset, our proposed model PMSAP improves M-F1 by 3.72% on average, and enhances ACC by 2.81% on average. On the CMU-MOSI dataset, our proposed model PMSAP improves M-F1 by 1.65% on average, and enhances ACC by 2.25% on average.

Based on the above experimental results, it can be concluded that the overall performance of our proposed model PMSAP is better than other baseline models when solving MSA under uncertain modalities missing.

Theoretical Analysis. From Table 4, Table 5, Table 6, Table 7, Table 8 and Table 9 we can find that, the MCTN and TransM models have better performance than models AE and CRA. This can prove that the recurrent translation mechanism adopted in model MCTN and TransM can extract and integrate information from different modalities more effectively than the autoencoder mechanism used in the models AE and CRA.

Compared with models MTMSA and TgRN, our proposed model PMSAP has acheieved more excellent results. This is because that both the MTMSA and TgRN models take the text modality as the dominant modality. MTMSA translates all the modalities into the text modality to address

Table 9: Performance of models when missing multiple modalities on the CMU-MOSI dataset (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
CMU-MOSI	AE	74.65	83.59	75.87	80.84	71.70	76.41	70.15	75.57	68.23	73.66	65.65	70.72
	CRA	75.51	83.63	74.19	81.45	73.72	76.53	72.36	74.91	71.63	74.19	66.03	70.92
	MCTN	77.07	83.66	76.17	81.86	73.27	76.39	73.17	74.53	72.12	74.34	66.01	72.38
	TransM	81.17	84.41	75.71	82.63	73.37	80.51	73.16	79.65	70.08	78.38	67.52	75.54
	ICDN	75.87	84.31	73.06	80.75	72.23	80.49	71.44	78.67	71.09	75.15	68.48	72.76
	MRAN	82.75	85.87	81.72	84.57	78.36	81.00	78.64	80.01	76.30	79.35	70.11	76.80
	MMIN	82.33	84.36	77.52	83.73	77.88	81.03	74.55	79.02	72.72	77.77	70.89	74.51
	TATE_C	82.99	85.87	80.17	83.13	79.05	82.83	77.64	81.71	75.59	79.44	73.99	77.07
	MTMSA	84.97	86.33	80.29	83.46	81.36	83.42	79.39	82.31	76.32	80.31	74.88	78.31
	TATE_J	84.33	86.09	82.68	84.31	79.73	82.31	78.21	81.66	76.21	80.13	75.44	78.88
	TgRN	84.78	86.36	82.33	84.07	80.11	82.51	77.77	80.11	76.12	80.02	75.80	78.51
	SMCMSA	85.37	86.96	82.91	85.19	82.32	84.38	80.86	83.48	78.10	80.72	77.03	79.05
	Ours	**86.48**	**88.51**	**84.20**	**87.63**	**83.59**	**86.41**	**82.83**	**84.77**	**80.88**	**83.38**	**78.53**	**82.59**

the issue of uncertain modalities missing, while TgRN uses the text modality to guide the feature learning and reconstruction of other modalities to solve the problem of uncertain modalities missing. The two models primarily rely on the text modality, therefore, once the text modality is missing, the ability of these models in handling the missing modalities will be significantly reduced, thereby severely affecting the performance of the two models. In contrast, our proposed PMSAP relies on fused features rather than a single modality. Consequently, PMSAP exhibits better robustness, adaptability, and generalization ability, thus to acheive the most superior performance.

Compared with model SMCMSA, our proposed model PMSAP also shows more excellent results. This is because that SMCMSA tries to complete the missing modalities by searching for similar samples. However, there are two issues with this approach. Firstly, the quality of the similar samples found is unstable. Secondly, as the degree of modality missing increases, it becomes extremely difficult for SMCMSA to find highly-matching similar samples. In contrast, our proposed PMSAP model does not need to introduce additional samples. It only needs the fused features within the current sample, thus avoiding the aforementioned problems.

From Table 4, Table 5, Table 6, Table 7, Table 8 and Table 9, it can be observed that as the missing rate increases, all indicators of PMSAP decline. This is because a higher missing rate implies fewer available features. Consequently, there are fewer reference features for generating new features through interaction. Moreover, the quality of the fused modality is closely

related to the features of the available modalities. As the number of available features decreases, the quality of the fused modality also deteriorates.

5.6. Performance of PMSAPO in Dealing with Uncertain Modalities Missing with 4-classification and 7-classification

Table 10: Distribution of multi-class sentimental label counts in the IEMOCAP dataset

Category	Par.	Hap.	Ang.	Sad	Neu.	Fru.	Exc.	Sur.
Four-class	Train	367	655	661	1016	-	-	-
	Test	228	448	423	692	-	-	-
Seven-class	Train	354	670	636	1013	1109	608	57
	Test	241	433	448	695	740	433	50

Note: "Par." represents Dataset Partition, "Train" represents Train Set, "Test" represents Test Set. "Hap.", "Ang.", "Sad", "Neu.", "Fru.", "Exc." and "Sur." represent Happy, Angry, Sad, Neutral, Frustration, Excited and Surprised labels, respectively. "-" indicates that the corresponding category is not applicable in the Four-class.

To further verify the performance of PMSAPO in dealing with uncertain modalities missing, we conduct two kinds of experiments with multi-classification (4-classification and 7-classification). In the first experiment, we conduct 4-classification (happy, angry, sad and neutral) test on dataset IEMOCAP. In the second experiment, we conduct 7-classification (Happy, angry, sad, neutral, frustrated, excited and surprised) test on dataset IEMOCAP. The distribution of sentiment labels in dataset IEMOCAP is shown in Table 10. In this experiment, the online learning strategies are not used, so, we use PMSAP to denote our proposed model. Moreover, models TATE_J, MRAN, ICDN, MTMSA and SMCMSA are selected as baseline models.

Experimental results for the 4-classification sentiment classification are illustrated in Fig. 4. Fig. 5 shows results of the 7-classification sentiment classification. The vertical axes in Fig. 4 and Fig. 5 represent the values of evaluation metrics (M-F1 or Acc). Here, "4 M-F1" and "4 Acc" denote the M-F1 and Acc values of each model under various missing rates in the 4-classification sentiment classification experiment, while "7 M-F1" and "7 Acc" represent the M-F1 and Acc values of each model under various missing

rates in the 7-classification sentiment classification experiment. The horizontal axes in Fig. 4 and Fig. 5 stand for different missing rates (0, 0.1, 0.2, 0.3, 0.4, 0.5). Moreover, the results of TATE_J, MRAN, ICDN, MTMSA and SMCMSA are selected from work [12].

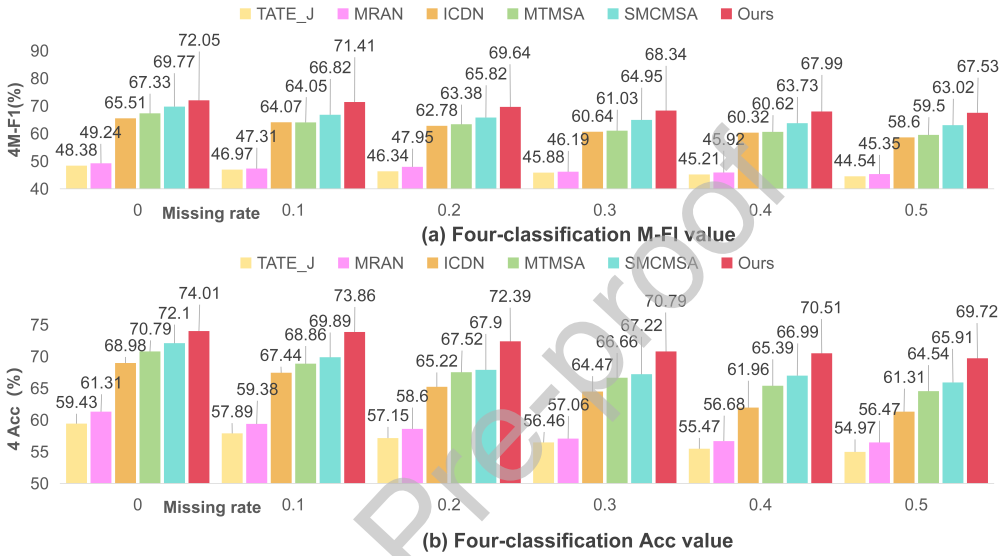


Fig. 4. Experimental results of different models on four-classification

From Fig. 4(a) and 4(b) we can find that, the performances of our proposed model PMSAP are the best among all the baseline models on all the missing rates. Among all the models, the performances of model SMCMSA are the second best. Compared with model SMCMSA, for all the missing rates, the M-F1 and ACC of our proposed model PMSAP are 3.81% and 3.55% larger than those of model SMCMSA averagely.

From Fig. 5(a) and 5(b) we find that, our proposed model outperforms all the baseline models for all the baseline models. Among all the models, the performances of model SMCMSA are the second best. Compared with model SMCMSA, for all the missing rates, the M-F1 and ACC of our proposed model PMSAP are 2.84% and 2.53% larger than those of model SMCMSA averagely.

Experimental results in Fig. 4 and Fig. 5 can prove that, for the 4-classification and 7-classification sentiment analysis, our proposed model PMSAP consistently outperforms the other five baseline models in terms of F1

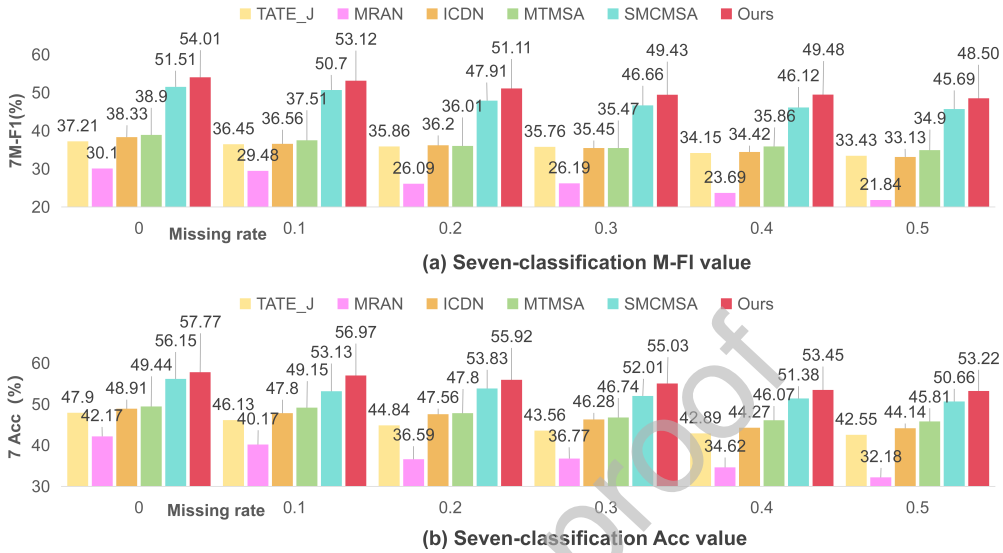


Fig. 5. Experimental results of different models on seven-classification

and ACC on all the missing rates. In summary, according to the above experimental results, we can come to a conclusion that our proposed model PMSAP has the best performance in multi-classification sentiment analysis under uncertain modalities missing.

5.7. Performance of PMSAPO in Personalized MSA

To verify the performance of PMSAPO for personalized MSA under uncertain modalities missing, we first perform pretraining with one dataset to obtain the optimal parameters, and then perform online learning with another dataset to adjust the optimal parameters and test the dataset to obtain the final result. In this work, we conduct three experiments.

The first experiment aims to prove the effectiveness of our proposed online learning strategies. In this experiment, we conduct 3-classification (angry, sad and happy) and 5-classification (happy, angry, sad, neutral and surprised) based on datasets IEMOCAP and MELD, respectively. The detailed distribution of labels of IEMOCAP and MELD is shown in Table 11 and Table 12. To fully test the performance of PMSAPO, we first take MELD as the pretraining dataset and take IEMOCAP as the testing dataset. Then, we take IEMOCAP as the pretraining dataset and use MELD as the testing dataset.

Table 11: Distribution of Three-class and Five-class sentimental label counts in the IEMOCAP dataset

Category	Par.	Hap.	Ang.	Sad	Neu.	Sur.
Three-class	Train	661	367	655	-	-
	Test	228	448	423	-	-
Five-class	Train	354	670	636	1013	57
	Test	241	433	448	695	50

Note: "Par." represents Dataset Partition, "Train" represents Train Set, "Test" represents Test Set. "Hap.", "Ang.", "Sad", "Neu." and "Sur." represent Happy, Angry, Sad, Neutral and Surprised labels, respectively. "-" indicates that the corresponding category is not applicable in the Three-class.

The second experiment aims to verify the performance of PMSAPO in Multi-Classification via comparing with the baseline models. In this experiment, we also conduct 3-classification (angry, sad and happy) and 5-classification (happy, angry, sad, neutral and surprised) based on datasets IEMOCAP and MELD, respectively. The detailed distribution of labels of IEMOCAP and MELD is also shown in Table 11 and Table 12. To fully test the performance of PMSAPO in Multi-Classification, we first take MELD as the pretraining dataset and take IEMOCAP as the testing dataset. Then, we take IEMOCAP as the pretraining dataset and use MELD as the testing dataset.

The third experiment aims to verify the performance of PMSAPO in Less Classification via comparing with the baseline models. In this experiment, we conduct 2-classification (negative and positive) based on IEMOCAP, MELD and CMU-MOSI datasets. The detailed distribution of labels of IEMOCAP, MELD and CMU-MOSI is shown in Table 13. To fully test the performance of PMSAPO in Less Classification, we first take MELD as the pretraining dataset and then take CMU-MOSI as the testing dataset. Next, we take IEMOCAP as the pretraining dataset and then use CMU-MOSI as the testing dataset. Finally, we take CMU-MOSI as the pretraining dataset and use IEMOCAP and MELD as the testing dataset, respectively.

Experiment I. To prove the effectiveness of the online learning strategies, we execute model PMSAP that without the online learning strategies, and model PMSAPO that have the online learning strategies. Experimen-

Table 12: Distribution of Three-class and Five-class sentimental label counts in the MELD dataset

Category	Par.	Hap.	Ang.	Sad	Neu.	Sur.
Three-class	Train	1109	683	1742	-	-
	Test	345	208	402	-	-
Five-class	Train	1742	1109	683	4708	1205
	Test	402	345	208	1256	281

Note: "Par." represents Dataset Partition, "Train" represents Train Set, "Test" represents Test Set. "Hap.", "Ang.", "Sad", "Neu." and "Sur." represent Happy, Angry, Sad, Neutral and Surprised labels, respectively. "-" indicates that the corresponding category is not applicable in the Three-class.

tal results are shown in Table 14 and Table 15, where (MELD, IEMOCAP) means that MELD is the pretraining dataset and IEMOCAP is the testing dataset, (IEMOCAP, MELD) indicates that IEMOCAP is the pretraining dataset and MELD is the testing dataset. From Table 14 and Table 15 we can find that, the performances of model PMSAP (without the online learning strategies) drops by about half on both 3-classification and 5-classification under various missing rates. These experimental results can prove that our proposed online learning strategy has good effectiveness, and can enable the pretrained model PMSAP to have excellent self-learning ability, thus can effectively address the problem of personalized MSA.

Experiment II. To verify the performance of our proposed model PMSAPO in Multi-Classification, in this experiment, we select TATE_C, TATE_J, MMIN, MTMSA, TgRN and SMCMSA as the baseline models, and we integrate our proposed online learning strategies into these baseline models, thus to enable them to have the online learning ability. Experimental results of 3-classification are described in Table 16, experimental results of 5-classification are shown in Table 17, where (MELD, IEMOCAP) means that MELD is the pretraining dataset and IEMOCAP is the testing dataset, (IEMOCAP, MELD) indicates that IEMOCAP is the pretraining dataset and MELD is the testing dataset.

From Table 16 we can find that, compared with the second-best model SMCMSA, when the IEMOCAP dataset is used as the testing dataset, our proposed model PMSAPO has an average increase of 6.63% in terms of M-F1

Table 13: Distribution of Two-class sentimental label counts in the IEMOCAP, MELD and CMU-MOSI datasets

Category	Partition	Negative	Positive
IEMOCAP	Train Set	3219	1299
	Test Set	859	337
MELD	Train Set	2945	2333
	Test Set	833	521
CMU-MOSI	Train Set	914	996
	Test Set	108	42

Note: "Partition" represents Dataset Partition.

Table 14: Experimental results of three-classification (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
(MELD, IEMO.)	PMSAP	42.99	50.32	41.78	49.56	41.07	48.59	40.98	48.12	40.66	47.95	40.13	47.66
	Ours	**81.24**	**83.48**	**80.39**	**83.23**	**80.68**	**83.12**	**80.43**	**82.58**	**78.20**	**81.11**	**76.36**	**78.65**
(IEMO., MELD)	PMSAP	39.64	53.74	38.12	52.01	37.71	51.88	37.24	51.70	37.01	51.42	35.46	51.09
	Ours	**65.63**	**67.89**	**64.56**	**67.63**	**64.94**	**66.14**	**61.87**	**65.01**	**58.60**	**63.72**	**55.23**	**60.54**

and an average increase of 5.86% in terms of ACC. When the MELD dataset is used as the testing dataset, our proposed model PMSAPO increases M-F1 by 6.59% on average, and increases ACC by 6.23% on average.

From Table 17 we can find that, compared to the second-best model SMCMSA, our proposed model PMSAPO demonstrates significant improvements in both M-F1 and ACC. Specifically, when the IEMOCAP dataset is used as the testing dataset, our proposed model PMSAPO has an average increase of 6.37% in M-F1 and an average increase of 6.74% in ACC, and when the MELD dataset is used as the testing dataset, PMSAPO increases M-F1 by 4.54% on average, and increases ACC is by 4.14% on average. Based on the above experimental results, it can be concluded that our proposed model PMSAPO demonstrates significantly better performance in multi-classification in personalized MSA under uncertain modalities missing.

Experiment III. To verify the performance of our proposed model PMSAPO in Less Classification, in this experiment, we select TATE_C, TATE_J, MMIN, MTMSA, TgRN and SMCMSA as the baseline models, and we integrate our proposed online learning strategies into these baseline models,

Table 15: Experimental results of five-classification (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
(MELD, IEMO.)	PMSAP	35.57	45.41	35.02	44.88	34.73	43.75	34.21	43.56	33.87	43.29	33.41	42.74
	Ours	**66.81**	**71.18**	**65.18**	**71.36**	**64.49**	**69.40**	**62.38**	**69.34**	**61.77**	**67.17**	**60.88**	**67.60**
(IEMO., MELD)	PMSAP	32.90	43.35	32.70	42.40	31.85	42.15	31.65	42.05	31.50	41.75	31.35	41.70
	Ours	**59.98**	**67.64**	**59.02**	**66.27**	**58.26**	**64.72**	**56.34**	**63.15**	**55.35**	**61.37**	**53.92**	**58.93**

Table 16: Experimental results of three-classification (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
(MELD, IEMO.)	MMIN	74.45	76.88	72.74	76.61	67.38	72.88	65.75	71.19	63.91	68.06	63.48	67.05
	TATE_C	74.33	78.41	72.86	77.27	68.22	73.62	67.34	71.81	66.04	69.82	64.08	67.46
	MTMSA	73.72	77.58	71.83	75.95	69.61	77.01	69.09	73.83	68.60	72.11	65.66	68.91
	TATE_J	74.71	79.45	73.07	76.92	70.84	76.24	67.79	75.47	69.86	73.11	67.23	71.56
	TgRN	74.19	79.33	73.56	77.33	71.89	75.91	69.56	73.89	68.33	72.19	67.91	71.48
	SMCMSA	74.73	79.20	74.14	78.05	73.75	77.64	73.05	75.25	71.45	74.00	70.42	72.85
	Ours	**81.24**	**83.48**	**80.39**	**83.23**	**80.68**	**83.12**	**80.43**	**82.58**	**78.20**	**81.11**	**76.36**	**78.65**
(IEMO., MELD)	MMIN	55.29	58.95	54.29	57.85	53.22	55.69	51.70	54.85	47.69	54.22	47.65	53.43
	TATE_C	56.86	59.83	56.87	59.34	54.19	57.22	51.85	53.99	49.48	53.36	47.84	52.23
	MTMSA	58.38	60.82	58.01	58.69	55.50	57.96	53.38	56.25	49.63	55.16	49.43	53.83
	TATE_J	58.75	61.35	59.66	60.90	56.69	59.21	54.15	57.38	49.18	52.95	49.10	53.23
	TgRN	59.01	62.33	58.67	61.33	56.01	58.97	54.33	56.67	52.19	54.43	49.01	53.33
	SMCMSA	59.36	62.97	59.17	61.82	56.63	59.12	54.76	57.84	52.07	56.35	49.29	55.48
	Ours	**65.63**	**67.89**	**64.56**	**67.63**	**64.94**	**66.14**	**61.87**	**65.01**	**58.60**	**63.72**	**55.23**	**60.54**

thus to enable them to have the online learning ability. Experimental results are shown in Table 18 and Table 19. (CMU., IEMO.) and (CMU., MELD) show the result that CMU-MOSI dataset is used as the pretraining dataset, IEMOCAP and MELD datasets are used as the testing datasets, respectively. (IEMO., CMU.) and (MELD, CMU.) show the result that IEMOCAP and MELD datasets are used as the pretraining datasets, CMU-MOSI dataset is used as the testing dataset, respectively.

From Table 18 we can find that, compared with the second-best model SMCMSA, when the IEMOCAP dataset is used as the testing dataset, our proposed model PMSAPO has an average increase of 1.91% in terms of M-F1 and an average increase of 2.21% in terms of ACC. When the MELD dataset is used as the testing dataset, our proposed model PMSAPO increases M-F1 by 4.84% on average, and increases ACC by 4.29% on average.

From Table 19 we can find that, compared to the second-best model SMCMSA, our proposed model PMSAPO demonstrates significant improvements in both M-F1 and ACC. Specifically, when the IEMOCAP dataset is

Table 17: Experimental results of five-classification (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
(MELD, IEMO.)	MMIN	57.04	61.95	54.40	58.80	54.08	59.13	53.20	58.09	51.39	56.82	49.77	54.14
	TATE_C	57.65	64.27	56.88	63.04	57.93	62.89	56.01	59.87	54.06	57.95	51.28	54.62
	MTMSA	57.73	63.49	57.71	61.73	56.90	60.82	56.24	61.18	54.06	57.98	51.73	57.24
	TATE_J	58.30	64.85	59.63	65.47	58.34	62.77	56.95	62.85	56.24	61.94	53.92	57.99
	TgRN	59.33	65.98	57.24	63.33	56.77	62.09	56.14	61.63	55.90	60.06	52.82	58.13
	SMCMSA	59.51	65.66	58.68	65.37	57.92	63.49	57.33	62.54	55.54	62.32	54.32	60.20
	Ours	**66.81**	**71.18**	**65.18**	**71.36**	**64.49**	**69.40**	**62.38**	**69.34**	**61.77**	**67.17**	**60.88**	**67.60**
(IEMO., MELD)	MMIN	54.26	59.12	52.22	57.31	50.70	56.12	46.97	51.79	43.44	48.66	44.24	49.70
	TATE_C	54.76	57.66	50.92	56.74	49.17	55.25	45.68	51.42	44.95	49.90	37.62	44.01
	MTMSA	55.80	59.96	52.89	58.99	52.22	59.25	51.15	56.16	46.12	51.76	44.40	50.24
	TATE_J	55.36	61.73	54.91	60.50	53.62	57.61	52.81	57.42	50.16	55.27	47.15	52.06
	TgRN	55.33	61.40	53.70	59.43	52.16	58.24	50.33	56.12	50.01	55.79	48.33	53.15
	SMCMSA	55.51	61.99	54.28	62.05	52.92	62.26	53.41	60.24	50.83	56.96	49.29	53.75
	Ours	**59.98**	**67.64**	**59.02**	**66.27**	**58.26**	**64.72**	**56.34**	**63.15**	**55.35**	**61.37**	**53.92**	**58.93**

used as the pretraining dataset, our proposed model PMSAPO has an average increase of 2.47% in M-F1 and an average increase of 3.15% in ACC, and when the MELD dataset is used as the pretraining dataset, PMSAPO increases M-F1 by 2.54% on average, and increases ACC is by 3.09% on average. Based on the above three experiments, it can be concluded that our proposed model PMSAPO has the best performance in the personalization MSA under uncertain modalities missing.

5.8. Significance test

To verify the reliability of the performance improvement of our proposed PMSAPO model, we conducted a significance test using ANOVA for each missing rate based on the experimental results of personalized MSA under uncertain modalities missing conditions. The experiments were performed using the MELD dataset as the pretraining dataset and the IEMOCAP dataset as the testing dataset for 3-classification. Moreover, to further enhance the credibility of our experimental results, we exclusively selected the baseline models from 2024 and 2025 (MTMSA, TgRN, SMCMSA) for calculations. The significance test results, including F-values and P-values for both M-F1 and ACC, are reported in Table 21.

From Table 21 we can find that, for both M-F1 and ACC, the F-values are consistently high (ranging from 18.76 to 25.34), and the P-values are all less than 0.001. This indicates that the performance differences between the PMSAPO model and the baseline models are statistically significant, and the

Table 18: Experimental results of two-classification with the CMU-MOSI dataset used as the pretraining dataset (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
(CMU., IEMO.)	MMIN	81.33	84.19	79.01	83.21	78.56	82.33	77.48	81.48	73.50	80.73	71.39	78.33
	TATE_C	82.26	86.50	81.12	85.33	80.10	84.84	79.27	83.59	77.15	82.63	75.27	82.50
	MTMSA	82.27	87.15	82.01	86.23	81.65	85.22	80.71	84.79	80.01	83.23	76.33	82.98
	TATE_J	82.27	87.32	82.01	86.39	80.12	84.28	79.42	84.33	78.31	83.34	76.34	82.83
	TgRN	83.33	87.07	82.13	85.24	81.33	84.17	79.27	83.01	78.41	83.33	77.17	82.88
	SMCMSA	85.01	87.93	83.33	86.54	82.21	85.65	81.31	84.80	80.21	84.01	78.88	83.79
	Ours	**85.89**	**89.33**	**84.88**	**88.79**	**84.01**	**87.98**	**83.33**	**87.54**	**82.96**	**86.67**	**81.31**	**85.67**
(CMU., MELD)	MMIN	70.33	72.27	69.52	71.63	68.25	71.01	67.67	70.56	66.33	68.56	65.32	67.36
	TATE_C	70.37	72.63	69.31	71.83	67.99	70.01	67.33	69.38	66.39	68.93	65.67	67.76
	MTMSA	71.33	74.28	70.01	73.56	69.76	71.98	67.56	69.77	67.33	69.41	65.34	67.98
	TATE_J	71.73	74.91	70.43	72.74	69.96	71.75	68.38	70.72	67.16	69.93	66.83	68.88
	TgRN	71.61	74.33	70.65	73.64	68.73	72.85	67.30	71.17	66.88	70.75	66.01	68.98
	SMCMSA	72.33	75.42	71.21	74.07	70.09	74.41	69.30	73.33	69.16	72.98	68.58	71.41
	Ours	**76.51**	**79.33**	**76.01**	**78.67**	**75.74**	**78.35**	**74.88**	**77.59**	**74.01**	**77.29**	**72.53**	**76.12**

improvements are not due to random chance. These results further validate the effectiveness of the PMSAPO model in handling missing data across various missing rates. In conclusion, through the significance test, it can be confirmed that our proposed PMSAPO model significantly outperforms other baseline models at all missing rates.

5.9. Analysis of the Dynamic Changes in Training Loss

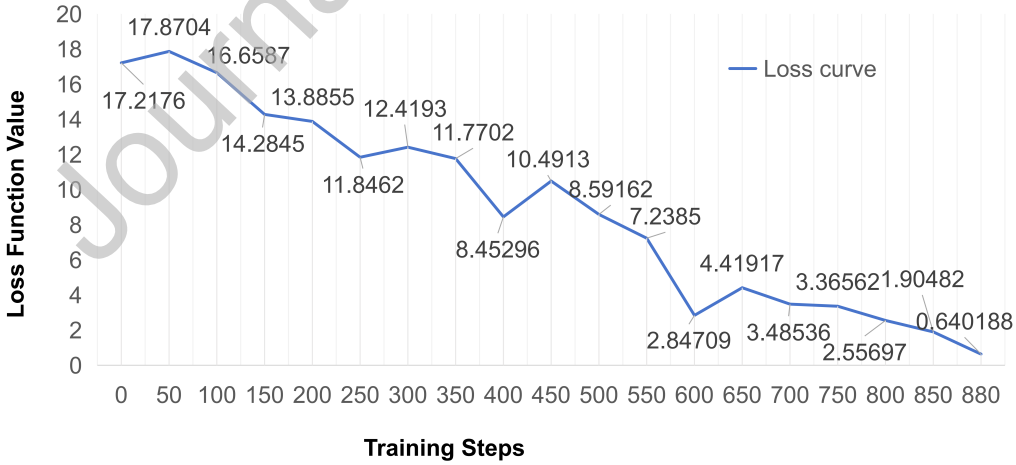


Fig. 6. The change of the loss values in training of PMSAPO

Table 19: Experimental results of two-classification with the CMU-MOSI dataset used as the testing dataset (the best results are bolded)

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
(IEMO., CMU.)	MMIN	83.11	85.21	80.73	84.56	79.93	82.33	78.64	81.21	76.21	79.33	74.56	77.21
	TATE_C	83.69	86.10	81.56	84.33	80.98	83.78	80.10	83.21	79.32	81.33	76.11	80.57
	MTMSA	85.18	87.10	82.51	84.74	81.15	83.33	81.01	83.10	79.38	81.51	76.67	81.22
	TATE_J	84.77	86.89	83.17	85.21	82.38	84.69	80.33	82.18	77.77	80.89	76.24	80.01
	TgRN	85.01	86.98	84.17	85.33	82.52	84.66	80.55	83.01	78.67	81.55	77.51	80.69
	SMCMSA	85.98	87.33	84.01	86.12	83.19	85.33	82.28	84.21	80.56	83.57	79.39	82.83
	Ours	**87.78**	**89.41**	**85.65**	**88.98**	**85.47**	**88.33**	**84.69**	**87.98**	**83.98**	**87.06**	**82.67**	**86.50**
(MELD, CMU.)	MMIN	83.41	85.38	80.83	84.67	79.89	82.15	78.51	81.69	76.77	79.41	74.33	77.56
	TATE_C	83.97	86.51	81.77	84.69	81.01	83.89	79.89	82.89	79.18	80.97	76.21	80.41
	MTMSA	85.18	87.21	82.41	84.69	81.33	83.56	81.01	83.12	79.83	81.91	76.56	80.51
	TATE_J	84.91	87.28	83.04	85.57	82.45	84.69	80.47	83.33	78.01	81.12	76.63	80.57
	TgRN	85.12	87.12	83.91	85.17	82.39	84.79	80.63	83.12	78.33	81.63	77.15	80.88
	SMCMSA	85.79	87.29	84.33	86.39	82.57	85.01	82.12	84.63	81.33	83.33	79.89	82.72
	Ours	**87.56**	**89.33**	**86.64**	**88.79**	**85.83**	**88.48**	**84.79**	**87.45**	**83.89**	**87.17**	**82.55**	**86.66**

Table 20: The results of ablation experiment in the PMSAP model, where the best results are bolded.

Modules	0		0.1		0.2		0.3		0.4		0.5	
	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
PMSAP-JF	84.33	87.71	82.60	87.33	81.35	86.87	81.06	85.17	79.56	84.12	78.03	82.77
PMSAP-JN	84.67	88.32	82.13	86.66	81.96	86.02	81.33	85.88	80.02	85.27	78.99	83.60
PMSAP	**85.21**	**89.16**	**84.72**	**88.51**	**83.69**	**87.45**	**82.97**	**87.31**	**82.47**	**86.33**	**80.09**	**84.90**

In the training process of our proposed PMSAPO model, the value of the loss function is a critical factor in guiding the model to learn the target features. Therefore, we conduct experiment to describe the dynamic change curve of the loss function with the number of training steps in personalized MSA, where the MELD dataset is used for pretraining and the CMU-MOSI dataset is used for testing. Experimental results are presented in Fig. 6.

From Fig. 6 we can find that, the loss curve generally exhibits a downward trend. The initial loss value is high, with a value 17.2176, this is because that parameters of PMSAPO are learned from the MELD dataset may not adapt to the CMU-MOSI dataset. As the training progresses, the loss values gradually decrease and finally reaches 0.640188 at step 880, these experimental results indicate that PMSAPO can converge to a stable and optimal state.

Detailed Analysis. Despite the use of multiple loss terms (e.g., loss of attention mechanism, loss of classification, loss of modality weight evaluation, and L2 regularization loss), the loss values of PMSAPO decrease

Table 21: The results of Significance test.

Evaluate	0		0.1		0.2		0.3		0.4		0.5	
metrics	F	P	F	P	F	P	F	P	F	P	F	P
M-F1	25.34	0.00012	22.56	0.00021	24.78	0.00015	23.45	0.00018	20.89	0.00031	19.76	0.00042
ACC	24.12	0.00016	21.34	0.00025	23.56	0.00017	22.45	0.00019	19.89	0.00033	18.76	0.00045

smoothly and consistently throughout the training process, which can prove that PMSAPO can effectively handle these multiple loss terms. The absence of sudden peaks or valleys in the loss curve suggests that the PMSAPO does not experience issues such as vanishing or exploding gradients. The gradual and steady decrease in the loss values indicate that the gradients are well-behaved, and PMSAPO can learn stably.

Moreover, it is inevitable that there are some fluctuations in the loss values at certain training steps, such as 250 to 300 steps, 400 to 450 steps, and 600 to 650 steps. These fluctuations can be attributed to the following reasons: (1) Learning rate adjustments: During these steps, the learning rate may have been adjusted, causing temporary fluctuations in the loss values. (2) Variations in data batches: The distribution of data in each batch may differ, leading to fluctuations in the loss values. However, these fluctuations do not affect the overall performance of model PMSAPO.

In conclusion, the smooth and consistent decrease in the loss values throughout the training process and the final convergence of PMSAPO can demonstrate the robustness and effectiveness of our proposed model PMSAPO. Moreover, the consistent reduction in loss values can prove that PMSAPO can learn the target features and perform sentiment classification with high accuracy.

5.10. Ablation Experiment

The aims of this ablation experiment including to verify the effectiveness of different modules of model PMSAP, to verify that pretraining can improve the performance of PMSAPO model personalized MSA, to prove the usefulness of each online learning technique, and to verify the fact that MSA models trained with public datasets do not have perfect performances for personalized MSA. For the module ablation experiments, we performed two-classification (positive and negative) experiments based on dataset IEMOCAP. For the pretrained ablation experiments, we performed three-classification (angry, sad, and happy) experiments based on dataset IEMOCAP. For the online learning strategy ablation experiments, we perform three-classification

(angry, sad, and happy) experiments based on dataset IEMOCAP, with pretrained model PMSAP on dataset MELD. For the Personalized MSA experiments, we performed three-classification (angry, sad, and happy) and five-classification (angry, sad, happy, neutral, and surprised)) experiments based on IEMOCAP and MELD datasets.

Module ablation experiment. In this experiment, different variants are generated by removing certain modules from PMSAP, and the effectiveness of different modules in PMSAP is verified by testing the performance of variants. The variants are generated as follows: (1) Variant PMSAP-JF is generated by removing the joint feature optimization module from PMSAP. (2) Variant PMSAP-JN is produced by removing the fully connected neural network evaluation module from model PMSAP .

Experimental results are presented in Table 20. From Table 20 we can find that, for PMSAP-JF, compared with PMSAP, the values of M-F1 and ACC are decreased under all the missing rates. Especially, when the missing rate is 0.3, the performance of PMSAP-JF decreases by 1.91% in terms of M-F1 and decreases 2.14% in terms of ACC. The above experimental results can prove that the joint feature optimization module in the PMSAP model is effective.

For PMSAP-JN, compared with PMSAP, it is evident that the M-F1 and ACC values decline under all the missing rates. When the miss rate is 0.1, PMSAP-JN decreases 2.59% in terms of M-F1 and decreases 1.85% in terms of ACC. When the missing rate is 0.3, the reduction of M-F1 of PMSAP-JN is 1.91%, and the ACC is reduced by 1.43%. These results can verify that the fully connected neural network evaluation module can improve the performance of PMSAP.

Pretrained ablation experiment. In this experiment, we train and test on the IEMOCAP dataset, that is, without pretraining, and pretrain on the MELD dataset and then train and test on the IEMOCAP dataset on the basis of pretraining, that is, with pretraining. To ensure that the experimental results are only affected by the pretraining, we update the online learning method for both. At the same time, to make this experiment more convincing, in addition to using the model PMSAPO our proposed, we also use the baseline models SMCMSA, TATE_J, MTMSA, TATE_C, MMIN to conduct experiments. Experimental results are presented in Table 22, where PMSAPO, SMCMSA, TATE_J, MTMSA, TATE_C, MMIN with pretraining, PMSAPO2, SMCMSA2, TATE_J2, MTMSA2, TATE_C2, MMIN2 without pretraining. Bold indicates better model results with the same missing rate

for the same model.

From Table 22, we can find that under the same model and same missing rate, the performance of the case with pretraining is better than that of the case without pretraining, which can prove that pretraining on the public dataset and then training and testing on personalized users can solve the PMSA problem more effectively, and is more in line with the actual application scenario.

Table 22: The results of pretraining ablation experiments.

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
IEMOCAP	MMIN2	72.45	75.11	69.48	73.43	66.05	71.11	64.82	70.10	63.22	66.16	61.45	64.23
	MMIN	**74.45**	**76.88**	**72.74**	**76.61**	**67.38**	**72.88**	**65.75**	**71.19**	**63.91**	**68.06**	**63.48**	**67.05**
	TATE_C2	73.12	77.03	70.05	75.67	66.22	71.24	65.36	69.09	63.26	67.76	62.63	65.88
	TATE_C	**74.33**	**78.41**	**72.86**	**77.27**	**68.22**	**73.62**	**67.34**	**71.81**	**66.04**	**69.82**	**64.08**	**67.46**
	MTMSA2	72.21	76.55	70.05	73.22	68.21	75.03	67.49	70.22	66.49	69.54	63.51	66.03
	MTMSA	**73.72**	**77.58**	**71.83**	**75.95**	**69.61**	**77.01**	**69.09**	**73.83**	**68.60**	**72.11**	**65.66**	**68.91**
	TATE_J2	73.88	78.34	71.11	74.61	70.38	73.55	66.21	72.18	68.01	71.36	66.11	69.14
	TATE_J	**74.71**	**79.45**	**73.07**	**76.92**	**70.84**	**76.24**	**67.79**	**75.47**	**69.86**	**73.11**	**67.23**	**71.56**
	SMCMSA2	72.20	77.55	71.73	76.88	71.13	74.96	70.19	73.34	69.18	72.61	68.13	70.13
	SMCMSA	**74.73**	**79.20**	**74.14**	**78.05**	**73.75**	**77.64**	**73.05**	**75.25**	**71.45**	**74.00**	**70.42**	**72.85**
	PMSAPO2	80.33	82.11	78.90	81.74	78.13	81.17	77.30	80.43	76.14	79.23	74.64	76.39
	PMSAPO	**81.24**	**83.48**	**80.39**	**83.23**	**80.68**	**83.12**	**80.43**	**82.58**	**78.20**	**81.11**	**76.36**	**78.65**

Table 23: The results of personalized module ablation, where the best results are bolded.

Modules	0		0.1		0.2		0.3		0.4		0.5	
	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
PMSAPO-A	78.56	81.68	77.50	81.38	77.31	80.57	76.61	78.88	75.92	77.21	73.01	75.43
PMSAPO-M	77.35	81.91	77.01	80.48	76.63	79.91	74.62	78.69	74.01	77.52	73.46	75.21
PMSAPO-U	80.41	82.22	79.63	82.69	78.97	82.00	78.66	81.89	77.31	80.36	75.59	77.12
PMSAPO-O	74.42	79.78	74.01	78.33	73.96	77.64	72.39	76.50	71.55	74.01	70.52	73.18
PMSAPO	**81.24**	**83.48**	**80.39**	**83.23**	**80.68**	**83.12**	**80.43**	**82.58**	**78.20**	**81.11**	**76.36**	**78.65**

Ablation experiment on online learning strategies. In this experiment, we take dataset MELD as the pretraining dataset to obtain the pretrained model PMSAP, and we take dataset IEMOCAP as the testing dataset. Different model variants are generated by removing certain online learning strategy from model PMSAPO, and the effectiveness of certain online learning strategy is verified by comparing the performances of model variants with performances of model PMSAPO. The model variants are generated as follows: (1) PMSAPO-A is generated by removing the adaptive learning rate adjustment strategy from PMSAPO. (2) PMSAPO-M is produced by removing the online meta-learning strategy from model PMSAPO.

(3) PMSAPO-U is generated by removing the strategy of assigning weights to different samples. (4) PMSAP is the pretrained model without the all online learning strategies, which implementing the online learning capabilities through fine-tuning operations.

Experimental results are shown in Table 23. From Table 23 we can find that, for PMSAPO-A, compared with PMSAPO, the values of M-F1 and ACC decline for all the missing rates. Especially, when the missing rate is 0.3, the M-F1 of PMSAPO-A decreases by 3.82%, and the ACC of PMSAPO-A decreases by 3.70%. The above experimental results can prove that the adaptive learning rate adjustment strategy is effective to enable the pretrained model PMSAP to own online learning ability.

For PMSAPO-M and PMSAPO-U, compared with PMSAPO, it can be found that the values of M-F1 and ACC are decreased under all the missing rates. These experimental results can verify that the online meta-learning strategy and the strategy of assigning weights to different samples can contribute significantly to enable the pretrained model PMSAP to own online learning ability.

For PMSAPO-O, in comparison with model PMSAPO, we observe a reduction in the values of M-F1 and ACC under all the missing rates. When the miss rate is 0.4, the M-F1 value of PMSAPO-O is decreased by 6.65%. When the missing rate is set to 0.5, the ACC value of PMSAPO-O model is reduced by 5.47%. Experimental results in Table 23 can prove that our proposed online learning strategies are effective.

Personalized MSA experiments. This experiment aims to verify the fact that MSA models that trained with public datasets will not have perfect performances for personalized MSA. In these experiments, four representative MSA models are selected, which are TATE_C [32], MTMSA [18], TATE_J [40], SMCMSA [12]. Two public datasets (IEMOCAP [30] and MELD [21]) are adopted. In the first experiment, IEMOCAP is used as the training dataset, and MELD is regarded as the dataset of personalized users. In the second experiment, MELD was used as the training dataset, and IEMOCAP was used as the personalized users' data. Experimental results are presented in Table 24 and Table 25, where the normal numbers in the two tables indicate models' performance when they are trained and tested with the same datasets, and the bold numbers denote models' performance when they are trained and tested with different datasets.

Based on the experimental results we can draw a conclusion that, the performance of MSA models trained with public dataset almost halved when

they are tested with personalized users. Therefore, the establishment of an efficient MSA model for personalized MSA has transformed into a crucial problem that needs to be addressed urgently.

Table 24: Three-class sentiment classification results. Non-bold data reflects standard operation; bold data reflects personalized user usage scenarios.

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
IEMOCAP	TATE_C	73.28	77.09	71.53	76.67	67.17	73.11	65.78	70.07	64.35	68.13	63.75	65.59
	TATE_C	**41.60**	**49.05**	**41.37**	**48.72**	**41.05**	**48.70**	**40.71**	**48.48**	**40.21**	**47.79**	**39.08**	**46.78**
	MTMSA	71.82	76.40	69.93	74.14	68.31	74.75	67.55	72.23	66.61	69.73	64.21	66.74
	MTMSA	**41.47**	**49.14**	**41.35**	**48.88**	**40.96**	**48.77**	**40.58**	**48.53**	**40.08**	**47.56**	**39.55**	**47.78**
	TATE_J	72.72	77.19	71.35	75.24	69.03	74.07	66.52	73.39	68.23	70.76	64.97	70.20
	TATE_J	**41.62**	**49.23**	**41.52**	**48.97**	**41.12**	**48.89**	**40.69**	**48.66**	**40.15**	**47.95**	**39.77**	**47.39**
	SMCMSA	72.92	77.28	72.06	76.15	72.12	75.91	71.15	73.17	69.19	71.65	67.89	71.13
	SMCMSA	**41.77**	**49.39**	**41.56**	**49.01**	**41.27**	**48.96**	**40.76**	**48.77**	**40.39**	**48.07**	**39.96**	**47.51**
MELD	TATE_C	55.23	57.93	55.33	57.08	52.59	54.82	50.13	52.54	47.49	51.28	46.03	49.97
	TATE_C	**32.14**	**46.02**	**30.36**	**44.77**	**30.03**	**44.59**	**29.56**	**44.16**	**29.19**	**43.67**	**28.47**	**43.04**
	MTMSA	56.48	58.83	56.43	57.57	53.60	55.94	51.48	54.17	48.24	53.26	46.99	51.93
	MTMSA	**32.07**	**45.94**	**30.29**	**44.61**	**29.74**	**44.44**	**29.67**	**44.28**	**29.42**	**43.79**	**28.11**	**42.83**
	TATE_J	57.12	59.18	58.21	58.79	54.86	56.77	52.07	55.48	47.46	51.14	46.66	50.97
	TATE_J	**32.46**	**46.49**	**30.87**	**44.99**	**30.59**	**44.27**	**30.12**	**44.26**	**29.82**	**43.97**	**28.49**	**43.46**
	SMCMSA	57.28	61.34	57.36	62.18	54.91	56.86	53.22	56.03	51.60	54.18	49.57	53.49
	SMCMSA	**32.69**	**46.78**	**31.11**	**45.27**	**30.77**	**44.76**	**30.28**	**44.49**	**29.86**	**44.24**	**28.53**	**43.79**

5.11. Practical Implications

In real-world application, on different devices, as long as the hardware environment is the same as ours and the data pre-processing methods are identical, highly accurate sentiment analysis results can be obtained. Moreover, the time required for a single user test is only 0.008563 s, which can fully meet users' needs. The memory required for a single test is 29.12 MB, and the model size is 281 KB, which can completely satisfy the requirements of practical deployment.

Moreover, in real-world application, we have integrated our PMSAPO model with a service recommendation system to detect users' sentimental states and provide tailored services to regulate their sentiments, enhance their well-being, prevent extreme behaviors, and mitigate the risk of depression, which is recognized as the fourth leading global disease. For instance, when the PMSAPO model detects signs of negative sentiments in a user, the service recommendation system suggests appropriate mental health services, such as counseling or relaxation exercises, to alleviate the user's sentimental distress. Empirical results from real-world application demonstrate that our

Table 25: Five-class sentiment classification results. Non-bold data reflects standard operation; bold data reflects personalized user usage scenarios.

Datasets	Models	0		0.1		0.2		0.3		0.4		0.5	
		M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC	M-F1	ACC
IEMOCAP	TATE_C	55.65	61.82	54.97	60.62	55.49	61.19	53.29	58.22	51.89	55.78	49.11	53.47
	TATE_C	**29.94**	**38.82**	**29.38**	**38.79**	**29.02**	**38.54**	**28.51**	**38.07**	**28.35**	**37.71**	**27.77**	**37.16**
	MTMSA	55.46	61.24	55.36	59.58	54.28	59.49	53.71	59.11	51.89	57.09	49.67	54.73
	MTMSA	**29.97**	**38.89**	**30.21**	**38.66**	**29.33**	**38.51**	**28.65**	**37.87**	**28.64**	**37.66**	**28.29**	**37.42**
	TATE_J	55.22	62.67	57.53	62.94	55.62	60.97	54.78	60.77	53.71	59.68	52.49	56.19
	TATE_J	**29.81**	**38.77**	**30.10**	**38.61**	**29.22**	**38.58**	**28.71**	**37.74**	**28.57**	**38.05**	**28.16**	**37.36**
	SMCMSA	57.35	63.84	57.45	63.03	55.48	61.14	54.62	60.64	53.58	59.53	52.53	58.31
	SMCMSA	**30.24**	**38.98**	**29.29**	**38.81**	**29.10**	**38.66**	**28.64**	**37.97**	**28.40**	**37.74**	**27.83**	**37.22**
MELD	TATE_C	52.46	55.86	48.78	54.67	46.91	52.83	44.05	48.56	42.81	48.29	35.46	42.49
	TATE_C	**26.93**	**37.47**	**26.76**	**36.44**	**25.93**	**36.27**	**25.70**	**36.13**	**25.59**	**35.87**	**25.42**	**35.81**
	MTMSA	53.27	57.89	50.54	57.19	49.66	56.83	48.42	53.31	43.51	50.47	41.99	48.75
	MTMSA	**26.51**	**37.92**	**26.35**	**37.04**	**25.92**	**36.80**	**25.74**	**36.72**	**25.63**	**36.41**	**25.45**	**36.37**
	TATE_J	52.82	59.91	52.38	58.25	51.55	55.69	51.47	55.35	48.93	53.72	45.81	50.69
	TATE_J	**26.65**	**37.56**	**26.43**	**36.66**	**25.54**	**36.44**	**25.32**	**36.33**	**25.21**	**36.02**	**25.05**	**35.96**
	SMCMSA	53.64	60.28	52.21	60.09	51.38	59.52	51.23	58.72	49.74	55.58	48.31	53.37
	SMCMSA	**26.77**	**37.71**	**26.59**	**36.79**	**25.73**	**36.58**	**25.57**	**36.46**	**25.42**	**36.16**	**25.26**	**36.13**

PMSAPO model achieves high accuracy in sentiment analysis under uncertain modalities missing conditions and delivers timely and effective service recommendations, thereby underscoring its practical utility and adaptability in related domains.

However, we have identified several challenges that impact the performance of the PMSAPO model in real-world scenarios. The first challenge is the discrepancy between users' expressed multimodal signals and their true sentimental states. For example, a user may laugh heartily to mask feelings of embarrassment, while their true sentimental state is neutral. In such cases, the PMSAPO model may erroneously classify the user's sentiment as happiness. The second challenge arises when users remain taciturn and exhibit consistent facial expressions, making it difficult for the model to accurately infer their sentimental states. To address these issues, we plan to introduce eye-movement and electroencephalogram (EEG) modalities in future work, as these physiological signals are less susceptible to conscious control and exhibit significant variations in response to sentimental changes.

In addition to experiments on personalized multimodal sentiment analysis under uncertain modalities missing, we also conducted experiments on personalized multimodal intention recognition under uncertain modalities missing. The results indicate that our model outperforms state-of-the-art models, demonstrating its superior generalization capabilities in other personalized

multimodal classification tasks under uncertain modalities missing scenarios. This further highlights the robustness and versatility of our PMSAPO model across diverse application domains.

6. Conclusion

Currently, Personalized Multimodal Sentiment Analysis (MSA) under uncertain modalities missing has become a new challenging problem. Although some effective models have been proposed for MSA under uncertain modalities missing, these models still have some serious deficiencies. Firstly, when dealing with uncertain modal missing, existing models' flexibility and generalization ability are relatively weak. Secondly, existing models perform poorly in personalized MSA under uncertain modalities missing. To tackle the above issues, we propose a personalized MSA model under uncertain modalities missing with pretraining and online learning (named PMSAPO). PMSAPO is pretrained with public datasets, and then it autonomously adapts to the personalized users with our proposed online learning strategies. Moreover, we propose a joint feature optimization and a fully connected neural network evaluation method to complete the missing modalities, which are simple and efficient. Extensive experiments have been conducted based on three public benchmark datasets (IEMOCAP, MELD and CMU-MOSI), and have proved that our proposed PMSAPO outperforms 12 baseline models and is effective for Personalized Multimodal Sentiment Analysis under uncertain modalities missing.

Compared with existing models, our proposed model PMSAPO has significantly improved its sentiment analysis ability. However, there are still the following deficiencies: (1) The performance of the PMSAPO model may be significantly degraded if the quality of the fused modalities is low, i.e., the initial available non-missing modalities are of low quality. This low quality can have a substantial impact on the effectiveness of modality completion. To solve this problem, in the future, we plan to enhance the quality of each modality before performing modality fusion. (2) In real-world applications, there are situations where all three modalities are missing simultaneously. If all three modalities are missing, our proposed model PMSAPO (along with all relevant models) will completely fail. To address this, we consider introducing two other modalities in the field of sentiment analysis-electroencephalogram (EEG) and eye-movement modalities, and combine them with the current three modalities to form a five-modality system. This can alleviate the oc-

currence of such extreme situations, and the five-modality system can handle more modality-missing scenarios in real-world applications.

CRediT authorship contribution statement

Hongxiang Sun: Methodology, Experiments Validation, Writing - Original Draft, Visualization. **Zhizhong Liu**: Conceptualization, Methodology, Writing - Original Draft, Investigation, Resources. **Dianhui Chu**: Conceptualization, Review, Editing, Investigation. **Quan Z. Sheng**: Formal analysis, Validation, Visualization. **Zhaowei Liu**: Experiments Validation, Editing. **Jian Yu**: Experiments Validation, Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 62273290, 61872126), the Special Funding Program of Shandong Taishan Scholars Project, Australian Research Council (ARC) Future Fellowship FT140101247 and Discovery Project DP200102298, Henan Province Science and Technology Research Project (No. 252102210138).

References

[1] C. M. Tyng, H. U. Amin, M. N. Saad, A. S. Malik, The influences of emotion on learning and memory, Frontiers in psychology 8 (2017) 235933.

[2] A. Ortigosa, J. M. Martín, R. M. Carro, Sentiment analysis in facebook and its application to e-learning, Computers in human behavior 31 (2014) 527–541.

[3] S. L. Angie Nguyen, Robert Pellerin, B. Lekens, Managing demand volatility of pharmaceutical products in times of disruption

through news sentiment analysis, International Journal of Production Research 61 (2023) 2829–2840. URL: https://doi.org/10.1080/00207543.2022.2070044. doi:10.1080/00207543.2022.2070044.

[4] S. Verma, Sentiment analysis of public services for smart society: Literature review and future research directions, Government Information Quarterly 39 (2022) 101708. URL: https://www.sciencedirect.com/science/article/pii/S0740624X22000417. doi:https://doi.org/10.1016/j.giq.2022.101708.

[5] P. Mahendhiran, K. Subramanian, Clsa-capsnet: Dependency based concept level sentiment analysis for text, Journal of Intelligent & Fuzzy Systems (2022) 1–17.

[6] A. Alslaity, R. Orji, Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions, Behaviour & Information Technology 43 (2024) 139–164.

[7] B. Erkantarci, G. Bakal, An empirical study of sentiment analysis utilizing machine learning and deep learning algorithms, Journal of Computational Social Science 7 (2024) 241–257.

[8] S.-J. Wu, R.-D. Chiang, H.-C. Chang, Applying sentiment analysis in social web for smart decision support marketing, Journal of Ambient Intelligence and Humanized Computing 15 (2024) 1927–1936.

[9] D. Xu, Z. Tian, R. Lai, X. Kong, Z. Tan, W. Shi, Deep learning based emotion analysis of microblog texts, Information Fusion 64 (2020) 1–11.

[10] K. Chan, Future Communication Technology and Engineering: Proceedings of the 2014 International Conference on Future Communication Technology and Engineering (FCTE 2014), Shenzhen, China, 16-17 November 2014, CRC Press, 2015.

[11] B. Yang, B. Shao, L. Wu, X. Lin, Multimodal sentiment analysis with unidirectional modality translation, Neurocomputing 467 (2022) 130–137.

[12] Y. Sun, Z. Liu, Q. Z. Sheng, D. Chu, J. Yu, H. Sun, Similar modality completion-based multimodal sentiment analysis under uncertain missing modalities, Information Fusion 110 (2024) 102454.

[13] Z. Quan, T. Sun, M. Su, J. Wei, Multimodal sentiment analysis based on cross-modal attention and gated cyclic hierarchical fusion networks, Computational Intelligence and Neuroscience 2022 (2022).

[14] C. Cai, Y. He, L. Sun, Z. Lian, B. Liu, J. Tao, M. Xu, K. Wang, Multimodal sentiment analysis based on recurrent neural network and multimodal attention, in: Proceedings of the 2nd on multimodal sentiment analysis challenge, 2021, pp. 61–67.

[15] Y. Fu, S. Okada, L. Wang, L. Guo, Y. Song, J. Liu, J. Dang, Context-and knowledge-aware graph convolutional network for multimodal emotion recognition, IEEE MultiMedia 29 (2022) 91–100.

[16] Y. Shou, T. Meng, W. Ai, S. Yang, K. Li, Conversational emotion recognition studies based on graph convolutional neural networks and a dependent syntactic analysis, Neurocomputing 501 (2022) 629–639.

[17] Q. Zhang, L. Shi, P. Liu, Z. Zhu, L. Xu, Icdn: integrating consistency and difference networks by transformer for multimodal sentiment analysis, Applied Intelligence (2022) 1–14.

[18] Z. Liu, B. Zhou, D. Chu, Y. Sun, L. Meng, Modality translation-based multimodal sentiment analysis under uncertain missing modalities, Information Fusion 101 (2024) 101973. URL: https://www.sciencedirect.com/science/article/pii/S1566253523002890. doi:https://doi.org/10.1016/j.inffus.2023.101973.

[19] S. Mai, Y. Zeng, S. Zheng, H. Hu, Hybrid contrastive learning of tri-modal representation for multimodal sentiment analysis, IEEE Transactions on Affective Computing (2022).

[20] C. Fan, K. Zhu, J. Tao, G. Yi, J. Xue, Z. Lv, Multi-level contrastive learning: Hierarchical alleviation of heterogeneity in multimodal sentiment analysis, IEEE Transactions on Affective Computing (2024).

[21] S. Poria, D. Hazarika, N. Majumder, G. Naik, E. Cambria, R. Mihalcea, Meld: A multimodal multi-party dataset for emotion recognition in conversations, arXiv preprint arXiv:1810.02508 (2018).

[22] T. Zhou, S. Canu, P. Vera, S. Ruan, Feature-enhanced generation and multi-modality fusion based deep neural network for brain tumor segmentation with missing mr modalities, Neurocomputing 466 (2021) 102–112.

[23] C. Zhang, Y. Cui, Z. Han, J. T. Zhou, H. Fu, Q. Hu, Deep partial multi-view learning, IEEE transactions on pattern analysis and machine intelligence (2020).

[24] H. Pham, P. P. Liang, T. Manzini, L.-P. Morency, B. Póczos, Found in translation: Learning robust joint representations by cyclic translations between modalities, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 6892–6899.

[25] Z. Wang, Z. Wan, X. Wan, Transmodality: An end2end fusion method with transformer for multimodal sentiment analysis, in: Proceedings of The Web Conference 2020, 2020, pp. 2514–2520.

[26] H. Akbari, L. Yuan, R. Qian, W.-H. Chuang, S.-F. Chang, Y. Cui, B. Gong, Vatt: Transformers for multimodal self-supervised learning from raw video, audio and text, Advances in Neural Information Processing Systems 34 (2021) 24206–24221.

[27] P. Shi, M. Hu, S. Nakagawa, X. Zheng, X. Shi, F. Ren, Text-guided reconstruction network for sentiment analysis with uncertain missing modalities, IEEE Transactions on Affective Computing (2025).

[28] M. Li, D. Yang, Y. Lei, S. Wang, S. Wang, L. Su, K. Yang, Y. Wang, M. Sun, L. Zhang, A unified self-distillation framework for multimodal sentiment analysis with uncertain missing modalities, in: Proceedings of the AAAI conference on artificial intelligence, volume 38, 2024, pp. 10074–10082.

[29] X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, Y. Yao, A. Zhang, L. Zhang, et al., Pre-trained models: Past, present and future, AI Open 2 (2021) 225–250.

[30] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, S. S. Narayanan, Iemocap: Interactive emotional dyadic motion capture database, Language resources and evaluation 42 (2008) 335–359.

[31] A. Zadeh, R. Zellers, E. Pincus, L.-P. Morency, Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages, IEEE Intelligent Systems 31 (2016) 82–88.

[32] J. Zeng, T. Liu, J. Zhou, Tag-assisted multimodal sentiment analysis under uncertain missing modalities, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 1545–1554.

[33] T. Baltrusaitis, A. Zadeh, Y. C. Lim, L.-P. Morency, Openface 2.0: Facial behavior analysis toolkit, in: 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), IEEE, 2018, pp. 59–66.

[34] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[35] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, O. Nieto, librosa: Audio and music signal analysis in python, in: Proceedings of the 14th python in science conference, volume 8, 2015, pp. 18–25.

[36] P. Baldi, Autoencoders, unsupervised learning, and deep architectures, in: Proceedings of ICML workshop on unsupervised and transfer learning, JMLR Workshop and Conference Proceedings, 2012, pp. 37–49.

[37] L. Tran, X. Liu, J. Zhou, R. Jin, Missing modalities imputation via cascaded residual autoencoder, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1405–1414.

[38] J. Zhao, R. Li, Q. Jin, Missing modality imagination network for emotion recognition with uncertain missing modalities, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 2608–2618.

[39] W. Luo, M. Xu, H. Lai, Multimodal reconstruct and align net for missing modality problem in sentiment analysis, in: MultiMedia Modeling: 29th International Conference, MMM 2023, Bergen, Norway, January 9–12, 2023, Proceedings, Part II, Springer, 2023, pp. 411–422.

[40] J. Zeng, J. Zhou, T. Liu, Robust multimodal sentiment analysis via tag encoding of uncertain missing modalities, IEEE Transactions on Multimedia (2022).