International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 3132-3140

https://doi.org/10.70517/ijhsa464259

Deep learning-integrated human-machine communication cross-cultural empathy expression and adaptive interaction dual-modal construction

Xinruo Zhang[1,*]
[1] School of Literature and Media, Xi'an Institute of Translation, Xi'an, Shaanxi, 710015, China
Corresponding authors: (e-mail: zhangxinruo21@163.com).

Abstract Aiming at the problem of semantic fragmentation of multimodal data and insufficient dynamic modeling of emotion communication in cross-cultural human-computer communication, this paper proposes an improved TD-SIR emotion communication model that integrates multimodal alignment and self-attention mechanism. Contrastive learning technique is adopted to realize cross-modal semantic alignment, and Transformer-based self-attention network is designed to realize multimodal emotion inference at character level. Using three-degree influence theory to model the emotion propagation model and optimize the propagation threshold parameters of the TD-SIR model. Based on 128,000 cross-cultural multimodal data from Sina Weibo, the effectiveness of the improved TD-SIR model is verified. Compared with the TD-SIR model, in the initial propagation stage, the improved TD-SIR model is closer to the real data and has a better fitting effect. Setting different experimental parameters, the improved TD-SIR model achieves the highest accuracy of 92.48% when the propagation probability threshold is 0.28 and the forgetting probability threshold is 0.035. Under this experimental parameter, the model proposed in this paper better simulates the sentiment evolution trend of public opinion events and performs better than ESIS and EC models.

Index Terms multimodal alignment, self-attention mechanism, Transformer, TD-SIR, emotion propagation modeling

I. Introduction

Today's international situation is full of many unstable and uncertain factors, despite the complexity and change of the international situation, the world's multipolarity and globalization is still an irreversible trend of the times, and peace, development, stability, communication, and cooperation are the common vision of the peoples of the world [1]-[4]. The global outbreak and spread of new crown epidemics have made the peoples of the world realize the necessity of solidarity and common response to global public health events. The progress of human civilization and the realization of global well-being require the rational cognition and common emotion formed by the peoples of the world, and culture is the link to achieve such rational consensus and emotional empathy [5], [6]. Culture, as a flexible force with emotion and temperature, can construct a stable empathic structure that promotes the pro-social behavior of the peoples of the world in a state of empathy based on the common human emotions especially in the face of global public disasters [7]-[9].

Empathy is a concept active in different fields such as aesthetics, psychology, philosophy and communication. It emphasizes the homogeneous interpretation and emotional resonance among different cultural subjects, which is conducive to solving the communication dilemma of “speaking to the void” in global communication, and at the same time, the combination of empathy and cross-cultural communication activities is cross-cultural empathic communication [10]-[12]. However, the loss of cultural expression in cross-cultural communication leads to communication misunderstanding, of which the most significant misunderstanding is caused by unimodal interaction, with the loss of cultural expression exceeding 50%, and affects the communication effect of human-computer interaction [13]-[15]. In contrast, the accuracy of cultural expression performed by multimodal fusion was improved [16]. Therefore, it is considered to apply multimodal fusion techniques to cross-cultural empathy in order to explore its communication and interaction phenomena.

In this paper, we first propose to use multimodal alignment to solve the semantic consistency problem of different modal data, and design a multimodal sentiment inference method based on the attention mechanism. Based on the principle of three degrees of influence, the TD-SIR emotion propagation model is constructed, which is improved by combining multimodal alignment with self-attention mechanism. Three typical cross-cultural events were selected as sample categories, and the improved TD-SIR model was used to analyze the dynamic process of emotion communication. With the help of RMSE, the fitting effects before and after the model improvement are analyzed to

examine the effectiveness of the improvement scheme in this paper. The optimal experimental parameters are determined, and the superior performance of this paper's model is proved through control experiments with ESIS and EIC models.

II. Design of a model for affective communication that combines multimodal alignment with the enhancement of self-attention mechanisms

Existing studies have mostly focused on single-modal sentiment analysis, but the synergistic mechanism of linguistic symbols, visual elements and acoustic features in cross-cultural scenarios has not been fully explored. In addition, the traditional sentiment communication model has limitations in dynamic evolution and cultural heterogeneity adaptation, which makes it difficult to capture the semantic consistency of multimodal data and the time-sensitive features of sentiment communication. In this paper, we propose an improved TD-SIR model based on multimodal alignment and self-attention mechanism, which provides theoretical support and technical paths for cross-cultural empathic expression and adaptive decision-making in human-computer communication by constructing a framework for cross-cultural multimodal emotion communication.

II. A.Multimodal alignment based on contrast learning

Contrastive learning-based multimodal alignment technique is a rapidly developing approach in the field of multimodal machine learning in recent years, which aims at solving the problem of semantic consistency between different modal data (e.g., image, text, audio, etc.). Under the contrast learning framework, the model receives paired multimodal inputs, and during training, the algorithm actively seeks out those samples that are truly paired and maximizes their similarity scores in the joint embedding space while minimizing the similarity between mismatched samples. In this way, the model can learn to map different modal data with the same semantics onto the same or highly correlated vector representations, thus achieving alignment.

The structure of the multimodal model based on contrastive learning alignment is shown in Fig. 1. The CLIP model uses contrastive learning techniques to realize deep semantic alignment between large-scale image and textual data, so that the distance between the textual descriptions with the same semantic concept and the encoding results of the image samples is as close as possible, and the encoding distance of the mismatched sample pairs is pulled away as far as possible, which effectively breaks down the boundaries between linguistic and visual information. It is the first to kick off the multimodal macromodeling. The ALBEF model, on the other hand, adds a form of momentum updating to the encoding module constrained by contrast learning on the basis of the CLIP model, which makes the contrast learning between pairs of modal features more efficient and further enhances the representation of inter-modal coherence and consistency.The BLIP model, in addition to using the underlying image-text contrast loss for the spatial alignment between the visual features and the linguistic representations, which strengthens the In addition to using the basic image-text contrast loss for spatial alignment between visual features and linguistic representations to strengthen the model's ability to understand cross-modal information associations, the BLIP model further uses the difficult-to-negative samples in contrast learning to construct an image-text matching loss, which is used to identify the correctness of the input image-text pairings with the help of contrast learning to refine the process of multimodal alignment. Transformer architecture, with the help of comparative learning, ensures that positively matched multimodal pairs are close in the feature space while negative sample pairs remain separated, successfully realizing the improvement of multimodal data comprehension under the unified architecture.

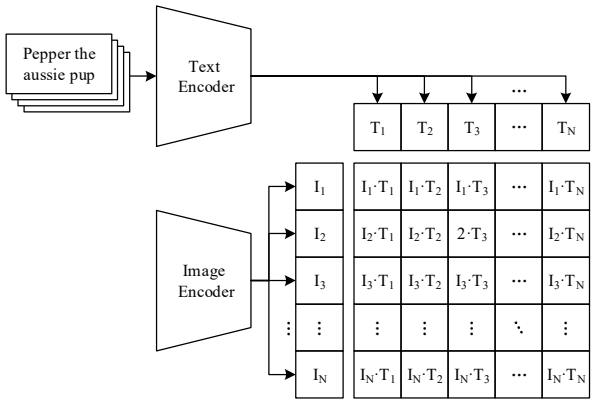


Figure 1: Multimodal model based on contrastive learning alignment

Multimodal alignment based on contrast learning has a strong multimodal alignment capability, which helps the model extract consistent semantic information from different data modalities by maximizing the similarity between positive sample pairs as well as minimizing the difference between negative sample pairs. And since contrast learning focuses on the structural construction of the underlying semantic space rather than the exact matching of specific instances, the model is better able to adapt to unseen data combinations and realize a wider range of migration applications. However, this approach relies heavily on large-scale aligned datasets, and when large-scale labeled data is lacking, contrastive learning may have difficulty in adequately mining deep abstract associations between modalities. Moreover, contrast learning usually involves a large amount of negative sample sampling and computation, especially in the large-scale pre-training phase, which also places high demands on hardware resources and training time.

II. B.Self-attention based modeling framework

In this paper, the following inference strategies are considered: on a psychological level, first, characters' emotions are continuously changing, so the emotional context of a character plays an important role in inferring the emotion of his or her current state. Second, characters' emotions are transmitted and influenced by each other, e.g., both happy and sad moods may be rapidly contagious, while conversing with an angry person may cause fearful emotions. Finally, a character's a priori knowledge, such as personality and past history, can portray the character better, as different characters have different emotions for the same event. Therefore, in order to perform character-level multimodal affective reasoning, this paper proposes a model based on an attention mechanism to adequately model a character's affective context, affective propagation between characters, and a priori knowledge about a character's personality.

The attention method used in this paper is the deflated dot product attention used in Transformer:

$$Att\left(Q,K,V\right)=soft\max\left(\frac{QK^{T}}{\sqrt{d}}\right)V \tag{1}$$

where Q, K, and V are the query vector, key vector, and value vector, respectively, and d is the dimension of the query vector and key vector. Suppose a video $V=\left(\{P_i\}_{i=1}^{M},\{S_j\}_{j=1}^{N}\right)$ has N clips and M characters, the task of this paper is to recognize the emotion category $E_{m,n}$ of the target character P_m in the target emotion clip S_n.

Multimodal embedding vectors. In this paper, three encoders E_a, E_t, E_v are used to encode the multimodal input features $\left\{\left(a_{i,j},t_{i,j},v_{i,j}\right)\right\}_{i=1,j=1}^{M,N}$ to obtain the multimodal embedding vector with uniform dimension $\left\{\left(f_{i,j}^{(a)},f_{i,j}^{(t)},f_{i,j}^{(v)}\right)\right\}_{i=1,j=1}^{M,N}$, and $f_{i,j}^{(k)}\in\square^{256}$, where $k\ in\{a,t,v\}$ represents sound, text and visual modalities.

Character embedding vector: in this paper, a two-layer perceptual machine E_p is used to encode the character traits p_i of a person P_i to obtain the character embedding vector $f_i^{(p)}$. Subsequently, it is used as a global knowledge to augment the multimodal embedding vectors, yielding

$$\hat{f}_{i,j}^{(k)}=\left[f_{i,j}^{(k)};f_i^{(p)}\right] \tag{2}$$

where $k\in\{a,t,v\}$ represents sound, text and visual modalities.

Attention mechanism at the modal level. Prior to fusion of the multimodal embedding vectors, character-guided temporal and inter-character attentional enhancement was performed at the modal level. First, at each semantic segment S_j, a self-attention operation was performed on the inter-character

$$h_{i,j}^{(k),1}=\hat{f}_{i,j}^{(k)}+Att\left(\hat{f}_{i,j}^{(k)},\hat{f}_{:,j}^{(k)},\hat{f}_{:,j}^{(k)}\right) \tag{3}$$

In this way, relationships and emotion propagation between pairs of characters are implicitly modeled, guided by the character traits $f^{(p)}$. Next, capturing the contextual emotional associations of each character P_i in its time, i.e., performing self-attention between contextual semantic fragments, yields

$$h_{i,j}^{(k),2}=h_{i,j}^{(k),1}+Att\left(h_{i,j}^{(k),1},h_{i,:}^{(k),1},h_{i,:}^{(k),1}\right) \tag{4}$$

If the modality k is missing for the character P_i on the semantic segment S_j, then $\hat{f}_{i,j}^{(k)}$ is set to 0 in the attention mechanism at the modal level.

Multimodal feature fusion. In this paper, we adopt an early fusion approach for multimodal feature fusion and connect the character embedding vectors to the multimodal features again after fusion, so as to directly utilize the character's personality traits at a higher level, and thus obtain the multimodal fusion features on each semantic segment:

$$h_{i,j} = \left[h_{i,j}^{(a),2}; h_{i,j}^{(t),2}; h_{i,j}^{(v),2}; f_i^{(p)} \right] \tag{5}$$

Character-level attention mechanism. For the target emotion segment S_j, the character-level self-attention is used to further capture its high-level inter-character emotion propagation and enhance the character-level multimodal representation. Thus, the final enhanced multimodal representation is:

$$\hat{h}_{i,j} = h_{i,j} + Att\left(h_{i,j}, h_{:,j}, h_{:,j}\right) \tag{6}$$

Ultimately, the augmented multimodal fusion features $\hat{h}_{m,n}$ about the target person P_m under the target emotion moment S_n are fed into a three-layer perceptual machine for emotion classification.

II. C.Affective communication model based on TD-SIR

The theory of three degrees of influence shows how people interact with each other in social networks, and it applies not only to the behavioral communication of users in social networks, but also to the emotional communication in social networks, which is based on the principle of three degrees of influence. In addition, social networks are in a state of dynamic change, with the continuous change of the network structure, the network will continue to evolve, more than three degrees of nodes in the social network connection is extremely unstable. The principle of three degrees of influence can more accurately describe the propagation of emotions in social networks, compared with the traditional view that people are only influenced by their own one degree nodes (i.e., direct friends), the principle of three degrees of influence can more realistically portray the real world.

The propagation of emotion in TD-SIR affects at most three degrees of nodes, and does not affect nodes outside the three degrees. Moreover, the degree of emotion propagation within three degrees is not fixed, the emotion in the process of propagation will continue to decay and will continue to reduce the intensity of the emotion propagation, with the emotion propagation chain continues to lengthen, the infectious force of the emotion, the credibility of the emotion will continue to decrease, this feature is characterized as the inherent attenuation. And the propagation of the emotion will decay as the one degree node to the second degree node, and the propagation of the emotion will decay again when it spreads from the second degree node to the third degree node, where the decay coefficient of the propagation of the emotion as the one degree node to the second degree node is α, which is referred to as the one degree decay coefficient, and the propagation of the emotion as the two degree node to the three degree node is β, referred to as the second degree decay coefficient, and $0 \le \alpha, \beta \le 1$.

The TD-SIR algorithm consists of the following three parts:

Part 1: Calculate the first degree node set attribute (Neighbor1), second degree node set attribute (Neighbor2) and third degree node set attribute (Neighbor3) of each node in the given social network graph and save the calculation results so that the above three attributes of each node can be directly obtained by the function in the subsequent steps, without the need to again re-do the traversal process on the social network graph. The first-degree node set attribute of each node in the social network graph can be computed directly using the graph's breadth-first search algorithm. Calculating the second-degree node set attributes of each node needs to be based on the first-degree node set attributes. For example, to find the second degree node set attribute of node V, use the breadth-first search algorithm to find the concatenation set of the first degree node set of each element J in the first degree node set of node V, and do the difference between this result and the Neighbor1 set of node V and node V to get the Neighbor2 attribute of node V. . Calculating the third degree node set property for each node requires the second degree node set property as a basis. For example, the method for finding the attribute of the three-dimensional node set of node V is: For each element K in the second-degree node set of node V, the breadth-first search algorithm is used to find the union of the first-degree node set of each element K. And by taking the difference between this result and the Neighbor1 set, Neighbor2 set of node V and node V, the Neighbor3 attribute of node V can be obtained.

Part II: TD-SIR's Emotion Infection Process In each iteration, a susceptible state node (S) that comes into contact with an infected state node (I) has a probability of becoming an infected state node.TD-SIR incorporates the idea of the principle of the three-degree of influence into this process. Using the node's three-degree attribute on top of the basic SIR model, the probability of emotional infection when a susceptible node encounters an infected node is not directly equal to the emotional contagion rate as in the SIR model, but rather is equal to the average of the emotional contagion rate of a one-degree node, the emotional contagion rate of a second-degree node multiplied

by a one-degree decay coefficient, and the emotional contagion rate of a third-degree node multiplied by a two-degree decay coefficient. TD-SIR can achieve a more accurate portrayal of the emotion propagation process than the basic SIR model.

Part III: Emotional healing process of TD-SIR, in each iteration, an infected node will be cured with probability μ and become an immune state node, the healing rate μ takes a constant value ranging from 0 to 1. The cure rate is typically set to the average network degree, and nodes are irreversibly exited from the propagation process when they are cured.

The second and third parts are repeated in an iterative process until there are no more nodes with infected states in the network graph.

The TD-SIR algorithm ends when there are no more nodes with infected state in the network graph.

III. Analysis of the effectiveness of the TD-SIR model enhanced by applying multimodal fusion techniques

The specific algorithmic implementation of the improved TD-SIR model proposed in this paper, which combines multimodal alignment with the enhancement of self-attention mechanism, is carried out on a server loaded with Ubuntu 18.04.1 system. In this experiment, we used Python crawler program to obtain data with the help of API interface function provided by Sina Weibo, and selected three sample categories with significant cultural differences, namely, public health events, holiday custom events and social movement events. Multimodal posts containing text, images, video and voice were crawled with event keywords during the period of 2024, and a total of 128,000 valid samples were obtained, covering user-generated content from 12 countries, including China, the United States, Japan, Germany and so on.

III. A. Dynamic Processes of Emotional Communication

Communication events often last for a long time, but their communication process is usually characterized by a long-tailed distribution, i.e., most of the discussions are completed within a relatively short period of time after the event, while a few of the discussions will stretch over weeks or even months to form a long tail. In this paper, in the process of acquiring data, according to the hour as the granularity of searching according to the event keywords, each hour can obtain about 1,000 related microblogging samples, and set the truncation when the number of samples collected in the hour is less than 50 (i.e., the collection rate is less than 5%), that is to say, the main body of the event tends to end the dissemination of the event, defining the event began to disseminate to this truncated point in time as the dissemination of the event time. Formally, the event i propagation time T_i is defined as:

$$T_i = \tau_{sample<50} - \tau_{start} \tag{7}$$

The dissemination time of the three types of events in the dataset is shown in Figure 2. From the figure, we can see that the dissemination of microblog message subjects of public health events and social movement events generally continues for a longer period of time, with a mean value of 157h and 158h, respectively, which tends to continue with the occurrence, discussion, and fermentation of the event, which also indicates that public health and social movement types of events are more likely to trigger a lasting discussion, and microblog dissemination of festive customs is more time-sensitive, and their attenuation is also faster, with a mean value of around 108h.

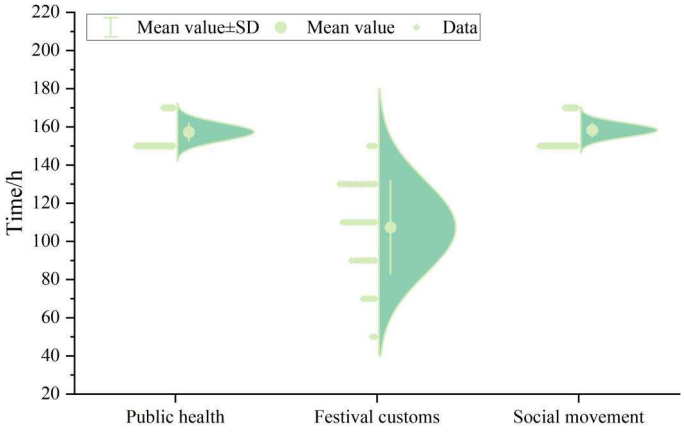


Figure 2: The propagation time of the three types of events

III. B. Assessment of the effectiveness of the fit

Setting the initial moment t of microblog posting, the fitting experimental results of the TD-SIR model and the improved TD-SIR model are obtained as shown in Fig. 3, in which the vertical axis represents the amount of microblog retweets I(t), and the horizontal axis represents the time t ($0 \le t \le 48h$). The microblog text grows faster in the initial stage of the forwarding volume, and the microblog information spreads slower in the later stage, and the propagation rate tends to 0, which means the end of propagation. In the initial propagation stage, the improved TD-SIR model is closer to the real data, while in the later stage with the increase of iteration number, the TD-SIR model is closer to the real data. Considering the timeliness of emotion propagation, the highest impact of emotion propagation is the initial propagation in the early stage, so the improved TD-SIR model fits better than the TD-SIR model.

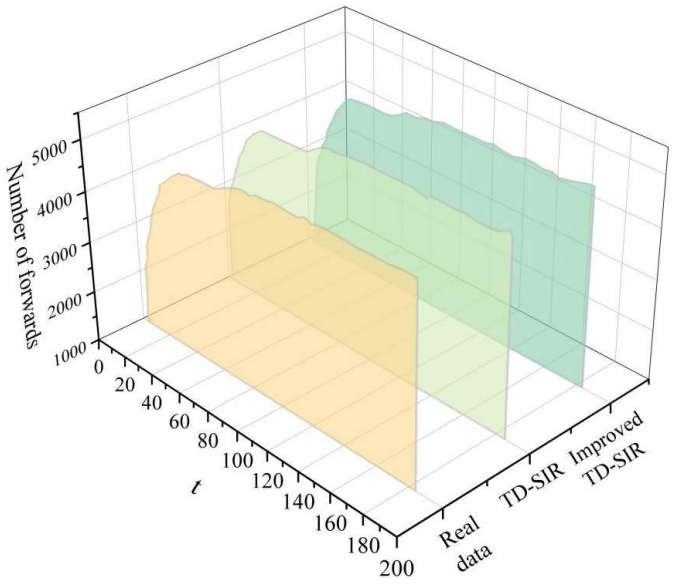


Figure 3: Fitting the experimental results

The root mean square error (RMSE) was used to compare and analyze the results of the two model fits, and a comparison of the root mean square error of the model fits is shown in Figure 4. It can be clearly seen that the RMSE of the improved TD-SIR model is smaller than the RMSE of the TD-SIR model, and it is known that the smaller the value of the root mean square difference, the better the fitting effect, the average RMSE of the TD-SIR model is -23.47, and the average RMSE of the improved TD-SIR model is -10.94, so the fitting effect of the improved TD-SIR model is more superior for the dissemination of emotional information.

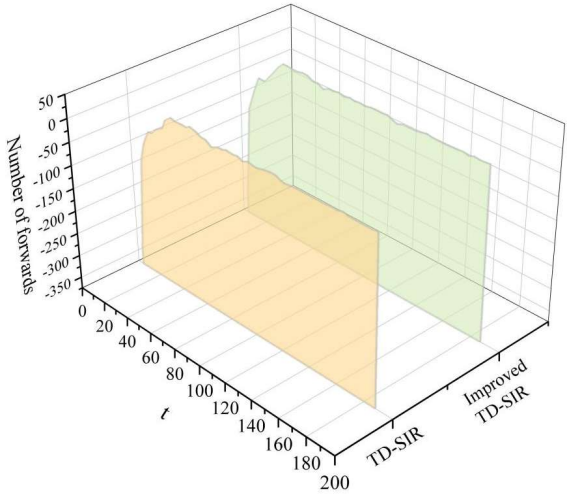


Figure 4: Comparison of root mean square differences in model fitting

III. C. Comparative Experimental Analysis

III. C. 1) Determination of experimental parameters

During the experiment, $\alpha = 0.5$ is set, and two experimental parameters exist: propagation probability threshold λ_a and forgetting probability threshold λ_b. The propagation probability threshold λ_a indicates that the propagation probability of uninformed users $\alpha_1(i)$ or $\alpha_2(i)$ is greater than λ_a, i.e., the uninformed S is transformed into a negative sentiment N or a positive sentiment P, and the forgetting probability threshold λ_b indicates that the probability of microblogging users forgetting is greater than λ_b. This experiment adopts the control parameter method, first fix one of the parameters, then debug the other parameter, set λ_a between the interval $[0.01, 0.03]$, set λ_b between the interval $[0.2, 0.3]$, and finally calculate the model's Accuracy Accuracy, as shown in Equation (8).

$$Accuracy = \frac{1}{T}\sum_{t=0}^{T}\left(1-\left|y_o(t)-y_p(t)\right|\right) \tag{8}$$

where T is the total time required for the entire dissemination process, $y_o(t)$ denotes the percentage of people actually participating in the opinion discussion at moment t, where $y_o(t) = \frac{N_o(t)+p_o(t)+R(t)}{M}$; and $y_p(t)$ denotes the percentage of people participating in the opinion discussion as predicted by the model at moment t, where $y_p(t) = \frac{N_p(t)+p_p(t)+R_p(t)}{M}$.

Meanwhile, in order to reduce the experimental error, this paper repeats the realization of each experiment for 100 times, and sums and averages the accuracies obtained from 100 calculations, and the accuracies under different propagation probability thresholds and forgetting probability thresholds are shown in Fig. 5. When the propagation probability threshold is 0.28 and the forgetting probability threshold is 0.035, the improved TD-SIR model achieves the highest accuracy rate of 92.48%, i.e., the model performance is optimal under this parameter.

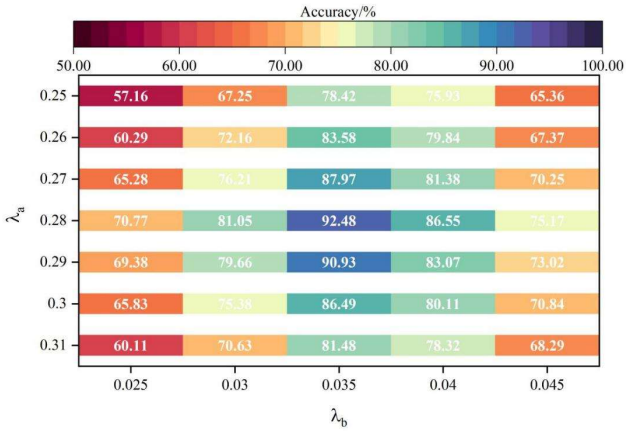


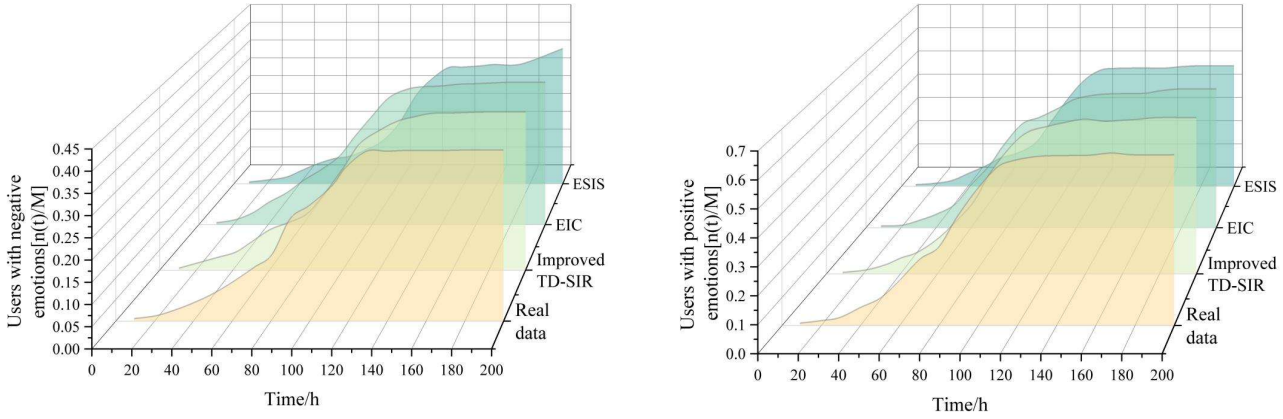
Figure 5: Accuracy rate under different experimental parameters

III. C. 2) Comparison experiments

In order to measure the model performance objectively, the ESIS model, EIC model and the improved TD-SIR model are selected for comparison experiments. Since these models have different underlying conditions, they need to be adjusted to the same baseline. The ESIS model categorizes emotions into fine-grained classes, taking the proportion of a particular emotion forwarded among users as a weight. Based on this, we categorize emotions into two types in the ESBS model, where unemotional and happy are considered positive, and anger, sadness, fear, and disgust are all considered negative. In the EIC model, the weights of the edges indicate the degree of influence among users, where the emotion values of neutral and positive are considered positive, and negative are considered negative.

The model parameters are set to α=0.5, λa=0.28, and λb=0.035, and the results of the fitting experiments of the model proposed in this paper with the ESIS and EIC models and the actual negative and positive sentiment evolution are shown in Fig. 6 (a~b), respectively. It can be seen that in the initial stage, the negative sentiment users and positive sentiment users grow slowly, and the growth of the propagation trend begins to accelerate as the event receives attention, but the number of negative sentiment users in the initial stage is less than the number of

positive sentiment users, and thus the opinion event as a whole is dominated by the propagation of positive sentiment, and as the time continues to grow, the trend of the propagation of positive sentiment and the propagation of negative sentiment gradually grows at a slower rate, indicating that the public opinion event gradually becomes more and more concerned about the evolution of negative sentiment. As time goes by, the trend of positive and negative emotions gradually decreases at a slower rate, indicating that microblog users gradually lose interest in opinion events and slowly fail to spread to other uninformed people. The model proposed in this paper better simulates the trend of sentiment evolution of public opinion events. Although the ESIS and EC models have the same development trend as the real data in the initial stage, the propagation trend of the EC model is higher than that of the real data, while the propagation trend of the ESIS model is obviously lower than that of the real data, and the gap with the real data is getting bigger and bigger in the later stage.



(a) Negative emotion

(b) Positive mood

Figure 6: Comparison of the fitting experiment results

IV. Conclusion

In this paper, we design a sentiment propagation model of TD-SIR enhanced by multimodal fusion technology, and experimentally explore its performance for cross-cultural category data.

The dynamic analysis of sentiment propagation using the improved TD-SIR model reveals that the propagation of microblog message subjects of public health events and social movement events generally has a longer duration, with mean values of 157h and 158h, respectively, while the microblog propagation of holiday and customary categories is more time-sensitive, and its decay is faster, with a mean value of around 108h.

Compared with the TD-SIR model, in the initial propagation stage, the improved TD-SIR model is closer to the real data and has a better fitting effect. Setting different experimental parameters, the improved TD-SIR model achieves the highest accuracy of 92.48% when the propagation probability threshold is 0.28 and the forgetting probability threshold is 0.035. Under the experimental parameters, the model proposed in this paper better simulates the sentiment evolution trend of public opinion events. Although the initial stage of the ESIS and EC models have the same development trend as that of the real data, the propagation trend of the EC model is higher than that of the actual data, and the propagation trend of the ESIS model is obviously lower than that of the actual data, and the gap between the later stage and that of the real data is getting bigger and bigger.

Funding

This research was supported by the Research on the high-quality development of the gymnasium (Project Number: 20240745).

References

[1] Chitondo, L., Chanda, C. T., Mwila, M. G., & Madoda, D. (2024). The paradox of a world aspiring for peace amidst pervasive conflicts. International Journal of Research and Innovation in Social Science, 8(2), 2471-2484.

[2] Abdenur, A. E. (2019). UN peacekeeping in a multipolar world order: Norms, role expectations, and leadership. United Nations operations in a changing global order, 45-65.

[3] Salvia, A. L., Leal Filho, W., Brandli, L. L., & Griebeler, J. S. (2019). Assessing research trends related to Sustainable Development Goals: Local and global issues. Journal of cleaner production, 208, 841-849.

[4] Li, B., Ma, X., Yu, Y., Wang, G., Zhuang, N., Liu, H., ... & Fu, W. (2020). Intensifying International Exchanges and Cooperation. Tutorial for Outline of the Healthy China 2030 Plan, 225-234.

[5] Sarmast, B. (2024). The Integration of Emotion and Rationality and Its Application in Explaining Social Consensus. Journal of Sociology of Lifestyle (SLS), 9(4), 81-103.
[6] Jami, P. Y., Walker, D. I., & Mansouri, B. (2024). Interaction of empathy and culture: a review. Current Psychology, 43(4), 2965-2980.
[7] Park, S., & Yu, J. (2017). The effects of multicultural experience on empathy in adolescents: Focused on mediating effect of multicultural acceptance and cultural empathy. Journal of Digital Convergence, 15(4), 499-510.
[8] Butovskaya, M. L., Burkova, V. N., Randall, A. K., Donato, S., Fedenok, J. N., Hocker, L., ... & Zinurova, R. I. (2021). Cross-cultural perspectives on the role of empathy during COVID-19's first wave. Sustainability, 13(13), 7431.
[9] Martí-Vilar, M., Serrano-Pastor, L., & Sala, F. G. (2019). Emotional, cultural and cognitive variables of prosocial behaviour. Current Psychology, 38, 912-919.
[10] Zeng, Z. (2022). Cross-cultural empathic communication of national image: Analysis of international communication strategies and effects of the Beijing Winter Olympic Games. Asian Social Science, 18(11), 1-4.
[11] Zhang, Y. S. D., & Noels, K. A. (2024). Cultural empathy in intercultural interactions: the development and validation of the intercultural empathy index. Journal of Multilingual and Multicultural Development, 45(10), 4572-4590.
[12] Shen, J., Na, L., & Qian, L. (2023). Empathy and cross-cultural communication: atypical experience of The Song of New China. International Communication of Chinese Culture, 10(1), 31-39.
[13] Abu Hatab, W., & Al-Badawi, M. (2020). Cross-cultural pragmatic failure in Jordanian media discourse. Jordan J. Mod. Lang. Lit, 12(3), 347-358.
[14] Zhai, C., & Wibowo, S. (2022). A systematic review on cross-culture, humor and empathy dimensions in conversational chatbots: the case of second language acquisition. Heliyon, 8(12).
[15] Oster, U. (2019). Cross-cultural semantic and pragmatic profiling of emotion words. Regulation and expression of anger in Spanish and German. Current approaches to metaphor analysis in discourse, 35-56.
[16] Parra, F., Scherer, S., Benezeth, Y., Tsvetanova, P., & Tereno, S. (2019). Development and Cross-Cultural Evaluation of a Scoring Algorithm for the Biometric Attachment Test: Overcoming the Challenges of Multimodal Fusion with "Small Data". IEEE Transactions on Affective Computing, 13(1), 211-225.