Research on Cross-Cultural News Communication Modeling Method Driven by Multimodal Semantic Fusion

Di Xiao*

Renmin University of China, Beijing, China

*xd15232369916@163.com

Abstract. This paper proposes a cross-cultural news communication modeling method driven by multimodal semantic fusion. A cultural context-aware multimodal semantic fusion algorithm, CCAMF, is designed, and a cross-cultural news communication prediction model based on fusion communication dynamics is constructed. This method achieves deep collaborative expression of text and image information through three modules: hierarchical extraction of multimodal features, cultural context-aware semantic alignment, and residual-enhanced cross-modal fusion. It also combines cultural difference moderating factors and audience characteristics to jointly predict the scope, speed, and click-through rate. Experiments are conducted using both publicly available and self-built datasets for validation. Results show that the CCAMF semantic fusion accuracy reaches 89.7%, and the cross-cultural semantic alignment error is reduced to 0.078. The accuracy of the scope prediction reaches 88.2%, the MAE of the speed prediction is 0.082, and the F1 score of the click-through rate prediction reaches 87.6%. Ablation experiments further demonstrate that each core module has a significant effect on improving model performance. The research results show that this method can effectively improve the quality of multimodal semantic fusion and the accuracy of communication modeling in cross-cultural news scenarios, and has good stability, generalization, and application value.

Keywords: Multimodal Semantic Fusion; Cross-Cultural News Communication; Cultural Context Perception; Communication Modeling; Deep Learning.

1 Introduction

With the rapid iteration of globalized information technology, cross-cultural news communication has become a core carrier of information interaction. Multimodal data such as text, images, and audio, due to their intuitiveness and completeness in information transmission, have gradually become the mainstream forms of cross-cultural news communication. Currently, the combination of multimodal semantic fusion technology and communication modeling methods provides a technical possibility for the precise control of cross-cultural news communication, but existing research still faces significant technical bottlenecks. At the level of multimodal semantic fusion, traditional methods

S. Shamsudin et al. (eds.), *Proceedings of the 4th International Conference on Language and Cultural Communication (ICLCC 2026)*, Advances in Social Science, Education and Humanities Research 1025,
https://doi.org/10.2991/978-2-38476-597-3_48

mostly adopt a shallow feature splicing mode, failing to achieve deep semantic fusion and lacking a context adaptation mechanism for cross-cultural scenarios, resulting in semantic bias in multimodal features. At the level of communication modeling, existing models often neglect the deep correlation between multimodal semantic features and cultural difference elements, leading to low prediction accuracy of communication scope, speed, and audience feedback, making it difficult to meet the application needs of actual cross-cultural news communication [1]. This paper focuses on the aforementioned core issues, designs an original Cultural Context-Aware Multimodal Semantic Fusion Algorithm (CCAMF), and constructs a cross-cultural news communication model driven by multimodal semantic fusion. The effectiveness and superiority of the method are verified through systematic experimental simulation [2]. The research results can effectively improve the accuracy and efficiency of cross-cultural news communication modeling, perfect the technical system of multimodal semantic fusion and communication modeling, and provide technical support for the precise and intelligent development of cross-cultural news communication [3]-[4]. Existing research on multimodal semantic fusion mainly focuses on single-cultural scenarios, lacking cross-cultural adaptability; cross-cultural communication modeling often relies on single text features, failing to fully utilize the complementary advantages of multimodal information. Deep integration research between the two remains lacking. This research has clear technological innovation and engineering application value.

2 Design of Cultural Context-Aware Multimodal Semantic Fusion Algorithm (CCAMF) and Cross-Cultural News Communication Modeling

2.1 Algorithm Overall Architecture

The CCAMF algorithm adopts a three-layer progressive architecture, including three core modules: multimodal feature layered extraction, cultural context-aware semantic alignment, and residual-enhanced cross-modal fusion, forming a complete cross-cultural semantic fusion process [5]. First, in the feature extraction stage, differentiated extraction strategies are designed for text and image dual-modal data. The text modality is based on an improved XLM-R model with the introduction of a cultural keyword attention mechanism to enhance semantic information and culturally loaded word feature expression. The image modality is based on the EfficientNet-B0 model with the addition of a cultural visual feature enhancement module to extract cultural symbols, scene semantics, and emotional tendency features. Secondly, in the semantic alignment stage, a cross-cultural semantic knowledge base, semantic alignment matrix, and cultural adaptation factor are used to dynamically correct semantic deviations in different cultural backgrounds, improving the cultural adaptability and consistency of feature expression [6]. Finally, in the fusion stage, residual connections and adaptive attention mechanisms are introduced to dynamically allocate fusion weights for each modality while preserving key information of the original modality, enhancing feature complementarity and fusion depth. Ultimately, a unified and efficient multimodal semantic feature vector is output, effectively compensating for the shortcomings of traditional fusion methods in feature loss, insufficient alignment, and weak cultural adaptability.

2.2 Propagation Model Construction

The multimodal semantic feature vector output by the CCAMF algorithm is defined as $X \in \mathbb{R}^d$, where d is the feature dimension. Combining the theory of communication dynamics, a cross-cultural news communication dynamics model is constructed as shown in formula (1):

$$\frac{dN(t)}{dt} = rN(t)\left(1 - \frac{N(t)}{K}\right) \cdot \alpha(X) \cdot \beta(C) \tag{1}$$

In the formula: $N(t)$ is the number of the audience for the news at time t; r is the basic propagation rate; K is the upper limit of the audience capacity; $\alpha(X)$ is the multimodal semantic feature influence factor, which is calculated from the feature vector output by the CCAMF algorithm and reflects the promoting effect of semantic features on propagation; $\beta(C)$ is the cultural difference moderating factor, where C is the cross-cultural difference coefficient, and the value range is [0,1]. The larger the difference, the smaller the value, and vice versa. The calculation of the multimodal semantic feature influence factor $\alpha(X)$ is shown in formula (2):

$$\alpha(X) = \frac{1}{1+e^{w^T X+b}} \tag{2}$$

In the formula: $w \in \mathbb{R}^d$ is the feature weight vector; b is the bias term, which is obtained through iterative optimization of the training data to ensure that the value range of $\alpha(X)$ is [0,1]. The calculation of the cultural difference moderating factor $\beta(C)$ is shown in formula (3):

$$\beta(C) = 1 - C \cdot \exp(-\lambda t) \tag{3}$$

In the formula: λ is the cultural adaptation rate coefficient, reflecting the speed at which the audience adapts to cross-cultural news, with a value range of (0,1), obtained by calibration from experimental data.

A sub-model for predicting communication effects is constructed to predict the audience click-through rate P_{click}, as shown in formula (4):

$$P_{\text{click}} = \sigma(X^T W_1 + U^T W_2 + \theta) \tag{4}$$

In the formula: $\sigma(\cdot)$ is the Sigmoid activation function; U is the audience feature vector; $W_1 \in \mathbb{R}^d, W_2 \in \mathbb{R}^m$(m is the audience feature dimension) are weight matrices; θ is the bias term.

3 Experimental Simulation and Result Analysis

3.1 Experimental Environment and Dataset

The experimental platform was configured with an Intel Core i9-12900K processor, an NVIDIA RTX 3090 graphics card (24GB VRAM), 64GB DDR5 RAM, and a 1TB SSD. The software environment included Ubuntu 20.04 LTS, Python 3.8, PyTorch 1.12.0, and Pandas, NumPy, Matplotlib, and Seaborn. The datasets combined publicly available datasets with self-built datasets. The 20min-XD French-German bilingual news dataset contained 15,000 text-image news articles, while the self-built Chinese-English-Japanese cross-cultural news dataset contained 8,000 articles, supplemented with audio modalities [7]. Both classes of data were partitioned into training, validation, and test sets in a 7:1.5:1.5 ratio. Preprocessing included text denoising, word segmentation, normalization, image resizing, and cultural semantic alignment.

3.2 Experimental Design

The experiment consisted of a comparative experiment and an ablation experiment. Strictly unified experimental parameters were maintained throughout to ensure scientific fairness. Specific parameters included an initial learning rate of 0.001, a decay coefficient of 0.95, a convergence threshold of 10^{-5}, weight coefficients of 0.6 and 0.4, 100 iterations, and a batch size of 32. Three comparative experiments were set up, all using the propagation model presented in this paper, with only the multimodal fusion algorithm replaced: a traditional early feature concatenation method, a mainstream Transformer-based algorithm, and the proposed CCAMF algorithm [8]. The superiority of the CCAMF algorithm was verified by comparing semantic fusion, propagation modeling performance, and algorithm efficiency. Evaluation metrics covered two categories: fusion performance was assessed using semantic fusion accuracy and cross-cultural semantic alignment error; propagation modeling was assessed using propagation range prediction accuracy, propagation speed prediction error, and click-through rate F1 score. Efficiency was also evaluated using the average processing time per news item; higher positive metrics and lower negative metrics indicated better performance.

3.3 Experimental Results and Analysis

Performance Analysis of Multimodal Semantic Fusion.

The performance comparison results of multimodal semantic fusion are shown in Figure 1. Figure 1(a) shows the comparison of semantic fusion accuracy of different algorithms, and Figure 1(b) shows the comparison of cross-cultural semantic alignment error [9]. As shown in Figure 1(a), the semantic fusion accuracy of the experimental group (CCAMF algorithm) reached 89.7%, which is 12.3 percentage points higher than control group 1 (early splicing method, 77.4%) and 6.3 percentage points higher than control group 2 (Transformer-based fusion, 83.4%). As shown in Figure 1(b), the cross-cultural semantic alignment error of the experimental group was 0.078, which is 18.8% lower than control group 1 (0.096) and 12.4% lower than control group 2 (0.089).

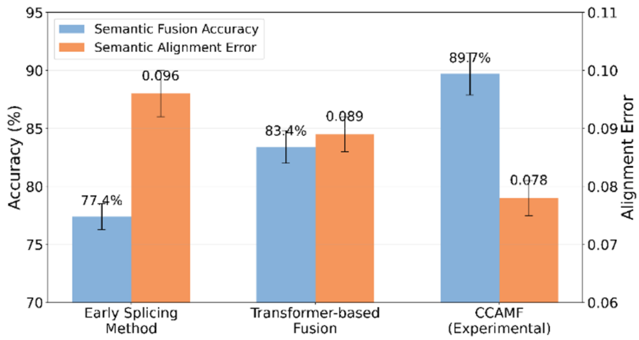

Fig. 1. Performance Comparison of Multimodal Semantic Fusion.

Performance Analysis of Cross-Cultural News Communication Modeling.

Figure 2 shows the performance comparison results of cross-cultural news communication modeling, illustrating the differences in accuracy of prediction of dissemination range, MAE of dissemination speed, and F1 score of audience click-through rate prediction [10]. As can be seen from Figure 2, the experimental group outperformed the two control groups in all modeling indicators: the accuracy of dissemination range prediction reached 88.2%, an increase of 10.5 percentage points compared to control group 1 (77.7%) and 5.8 percentage points compared to control group 2 (82.4%); the MAE of dissemination speed prediction was 0.082, a decrease of 12.8% compared to control group 1 (0.094) and 9.9% compared to control group 2 (0.091); the F1 score of audience click-through rate prediction reached 87.6%, an increase of 7.2 percentage points compared to control group 1 (80.4%) and 4.7 percentage points compared to control group 2 (82.9%).

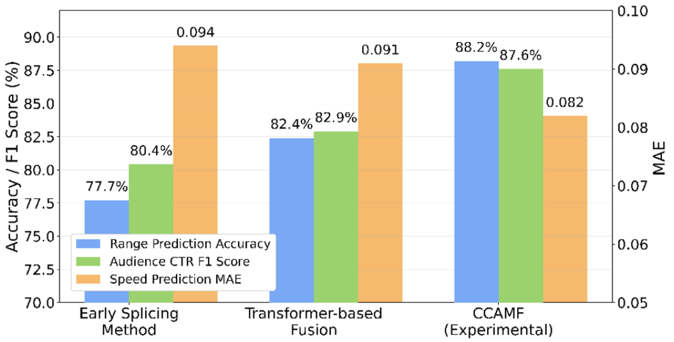

Fig. 2. Performance Comparison of Cross-Cultural News Communication Modeling.

4 Conclusion

This paper addresses the problem of multimodal semantic fusion and propagation prediction in cross-cultural news communication, proposing the CCAMF algorithm and its driven propagation modeling method. The results show that this method achieves superior performance in both semantic fusion and propagation prediction tasks: semantic fusion accuracy reaches 89.7%, cross-cultural semantic alignment error is 0.078, propagation range prediction accuracy reaches 88.2%, propagation speed prediction MAE is reduced to 0.082, and click-through rate prediction F1 score reaches 87.6%, outperforming earlier splicing methods and Transformer-based methods overall. Ablation experiments also verify the necessity of the semantic alignment module, residual fusion module, and cultural adaptation factor. This indicates that the proposed method can effectively solve problems such as large semantic bias, insufficient fusion, and inaccurate propagation prediction in cross-cultural scenarios.

References

1. Li, S., Yao, T., Li, S., Yan, L.: Semantic-enhanced multimodal fusion network for fake news detection. International Journal of Intelligent Systems 37(12), 12235–12251 (2022) DOI:10.1002/int.23084
2. Sittar, A., Mladenić, D., Grobelnik, M.: News dissemination: a semantic approach to barrier classification. Journal of Intelligent Information Systems 63(2), 535–565 (2025) DOI:10.1007/s10844-024-00894-5
3. Zhang, Y., Liu, Y., Guo, Z.: Optimising news dissemination pathways in the media convergence era: an interactive digital media technology approach. International Journal of Information and Communication Technology 26(29), 110–126 (2025) DOI:10.1504/IJICT.2025.147882
4. Zhu, P., Hua, J., Tang, K., Tian, J., Xu, J., Cui, X.: Multimodal fake news detection through intra-modality feature aggregation and inter-modality semantic fusion. Complex & Intelligent Systems 10(4), 5851–5863 (2024) DOI:10.1007/s40747-024-01473-5
5. Abbas, F., Taeihagh, A.: A multi-level fusion-based framework for multimodal fake news classification using semantic feature extraction. International Journal of Machine Learning and Cybernetics 16, 6531–6560 (2025) DOI:10.1007/s13042-025-02633-w
6. Zhou, D., Ouyang, Q., Lin, N., Zhou, Y., Yang, A.: GS2F: Multimodal fake news detection utilizing graph structure and guided semantic fusion. ACM Transactions on Asian and Low-Resource Language Information Processing 24(2), 1–22 (2025) DOI:10.1145/3708536
7. Hangloo, S., Arora, B.: Combating multimodal fake news on social media: methods, datasets, and future perspective. Multimedia Systems 28(6), 2391–2422 (2022) DOI:10.1007/s00530-022-00966-y
8. Uppada, S.K., Ashwin, B.S., Sivaselvan, B.: A novel evolutionary approach-based multimodal model to detect fake news in OSNs using text and metadata. The Journal of Supercomputing 80(2), 1522–1553 (2024) DOI:10.1007/s11227-023-05531-6
9. Zhang, L., Zhang, X., Zhou, Z., Zhang, X., Wang, S., Yu, P.S., Li, C.: Early detection of multimodal fake news via reinforced propagation path generation. IEEE Transactions on Knowledge and Data Engineering 37(2), 613–625 (2025) DOI:10.1109/TKDE.2024.3496701
10. Astuti, W.T., Mulyadi, A.: Selected Political Hoax Cases on Social Media: Multimodal in Forensic Linguistics. Policy & Governance Review 9(1), 44–63 (2025) DOI:10.30589/pgr.v9i1.1190