EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

MFUIE: A Fake News Detection Model Based on Multimodal Features and User Information Enhancement

Xiulan Hao[1], Wenjing Xu[1], Xu Huang[2,*], Zhenzhen Sheng[1] and Huayun Yan[1]

[1] Zhejiang Province Key Laboratory of Smart Management & Application of Modern Agricultural Resources, School of Information Engineering, Huzhou University, Huzhou, Zhejiang 313000, China

[2] School of Electronic Information, Huzhou College, Huzhou, Zhejiang 313000, China

Abstract

INTRODUCTION: Deep learning algorithms have advantages in extracting key features for detecting fake news. However, the existing multi-modal fake news detection models only fuse the visual and textual features after the encoder, failing to effectively utilize the multi-modal contextual relationships and resulting in insufficient feature fusion. Moreover, most fake news detection algorithms focus on mining news content and overlook the users' preferences whether to spread fake news.
OBJECTIVES: The model uses the multi-modal context relationship when extracting model features, and combines with user features to assist in mining multi-modal information to improve the performance of fake news detection.
METHODS: A fake news detection model called MFUIE (Multimodal Feature and User Information Enhancement) is proposed, which utilizes multi-modal features and user information enhancement. Firstly, for news content, we utilize the pre-trained language model BERT to encode sentences. At the same time, we use the Swin Transformer model as the main framework and introduce textual features during the early visual feature encoding to enhance semantic interactions. Additionally, we employ InceptionNetV3 as the image pattern analyser. Secondly, for user's historical posts, we use the same model as the news text to encode them, and introduce GAT (Graph Attention Network) to enhance information interaction between post nodes, capturing user-specific features. Finally, we fuse the obtained user features with the multi-modal features and validate the performance of the model.
RESULTS: The proposed model's performance is compared with those of existing methods. MFUIE model achieves an accuracy of 0.926 and 0.935 on the Weibo dataset and Weibo-21 dataset, respectively. F1 on Weibo is 0.926, 0.017 greater than SOAT model BRM; while F1 on Weibo-21 is 0.935, 0.009 greater than that of BRM.
CONCLUSION: Experimental results demonstrate that MFUIE can improve the fake news recognition in some degree.

Keywords: Multimodal, user information, fake news detection, deep learning.

Received on 10 10 2024, accepted on 12 12 2024, published on 12 12 2024

doi:10.4108/eetsis.7517

*Corresponding author. Email: hx @zjhzu.edu.cn

1. Introduction

Researches have proved that socially sensitive topics and curiosity-driven questions contributes to the dissemination of fake news. These fake news articles gradually become shorter in length and possess strong emotional appeal. Additionally, fake news often comes together with low credibility images and highly sensational characteristics [1]. Existing cross-modal interaction methods, such as information concatenation [2], information contrast [3], and information enhancement [4], usually deploy after the encoder and fail to effectively utilize the multi-modal contextual relationships. In existing work, fake news detection mainly focuses on leveraging multi-modal information from news content for modelling, but overlooks the preferences of users whether to spread fake news.

To address these issues, a fake news detection model called MFUIE (Multimodal Feature and User Information Enhancement) is proposed, which is based on deep learning. Deep leaning is widely used in all walks of life. Medical image processing algorithms based on deep learning are used in assisted diagnosis, such as diabetic eye disease identification in [5,6], early detection of mild cognitive impairment [7–10], atherosclerosis detection in coronary CT angiography [11], lung image segmentation [12], etc. Some also attempted to identify antisocial behavior using it [13]. It is also used in other scientific research, such as online decision support systems [14], remote sensing image processing [15], natural language processing [16], etc.

Optimization is a critical component in deep learning [17], general algorithms for training neural networks include SGD and its variants, SGD with momentum and SGD with Nesterov momentum, adaptive gradient methods, such as AdaGrad, RMSProp and Adam. They mainly focus on resolving “local issues” of training, and the theoretical results can at most ensure convergence to local minima.

Global optimization is a subarea of optimization which aims to design and analyze algorithms that find globally optimal solutions. Evolutionary methods and particle swarm optimization are such global search algorithms for general non-convex problems.

Evolutionary methods find fit place in enhancing database privacy and utility [18,19], efficiently solving Sudoku puzzles [20], group insurance portfolio so that the total payout of the whole group can be maximized [21], multimodal optimization problems (MMOPs)[22, 23], adaptive resource allocation[24], coordinated charging scheduling of electric vehicles [25], effectively tackling the multi-solution traveling salesman problem [26], aerodynamic airfoil design real-world optimization problem[27].

Yang, J. Q. et al. [28] proposes a bi-directional feature fixation (BDFF) framework for Particle swarm optimization (PSO) and provides a novel idea to reduce the search space in large-scale feature selection.

B. B, Rani KS et al. [29] propose an optimized hashing-based heuristic search technique to solve the multidimensional constraint reachability queries.

The proposed MFUIE is based on multi-modal features and a user information enhancement module.

Firstly, for news content information, a text pre-training model is utilized to get textual features. During the visual feature encoding stage, textual features for semantic interaction are combined to enhance the visual feature representation. Specifically, at each stage of the network, semantic correlations between fine-grained image regions and textual words are learned by using attention mechanisms. This approach helps to extract useful multi-modal contextual information. Furthermore, to verify the authenticity of visual content in fake news, the credibility of images is considered.

Secondly, for the user's historical posts, the same method as the news text information is employed to encode them, and the information interaction between post nodes is enhanced by introducing GAT to capture user-specific features.

Finally, the resulted user features are fused with the multi-modal features of the content to obtain the final feature representation, which is then fed into the classifier. In previous fake news datasets, the user's historical post information is often not omitted. Therefore, we also crawl data to obtain historical posts related to the target news, constructing a dataset that includes both news content and user historical posts.

2. Literature Survey

Here, fake news detection can be divided into two classes: content based and user based, where content-based detection includes many methods focusing on the single modality content features for detection and many recent works utilizing multimodal content in news [30]. Our focus is the latter.

Combination of textual and visual modalities. This provides complementary information for fake news detection. Therefore, methods that leverage multi-modal information for fake news detection have received much attention. Existing approaches can be roughly categorized into three types: concatenation fusion, contrast between modalities, and enhancement of multi-modal information.

In Singhal et al. [31], a multi-modal fake news detection framework is proposed, where BERT is deployed to capture semantic and contextual meanings and generate textual feature vectors. VGG19 is used to extract visual features, which are then dimensionally reduced through fully connected layers. The textual and visual feature vectors are simply concatenated and fed into a classifier. In Wang et al. [2] and Khattar et al. [32], VGG extracts visual features, while Text-CNN or Bi-LSTM extracts textual features. Visual and textual features are concatenated to represent the news. In Qu et al. [33], a quantum multimodal fusion-based model is proposed for fake news detection (QMFND). QMFND integrates the extracted images and textual features, and passes them through a proposed quantum convolutional neural network (QCNN) to obtain discriminative results.

Concatenation fusion cannot utilize the corelation between visual and textual features.

In Zhou et al. [3], a representative work on detecting the consistency between news text and images is proposed. Text-CNN is employed to extract textual and visual features and their correlation is defined by cosine similarity. In Xue et al. [34], a multi-modal consistent neural network (MCNN) is proposed, which extracts textual features using BERT and temporal attributes using Bi-GRU. Image features are encoded by a residual neural network and tampering features are highlighted by ELA algorithm to detect fake news. The similarity measurement module calculates the similarity between text and images, and the multi-modal fusion module weights and fuses their features. Considering the similarity between text and images improve the fake news recognition to some degree.

In Jin et al. [35], LSTM is deployed to extract textual information, VGG to extract visual information, and attention mechanisms to enhance the information understanding between modalities. To more effectively fuse multi-modal features, in Wu et al. [36], the relationships between modalities are considered and a two-layer image-text co-attention network is designed. First, VGG-19 is employed to extract spatial domain features, CNN to extract frequency domain features, and BERT to extract textual features. Then, the co-attention model fuses the multi-modal features to get the final features. The attention mechanisms make the model weigh features in different modalities and further improve the accuracy of fake news detection.

Exploitation of user features. Fake news writers often mimic the writing style of real news when writing fake news. Therefore, verifying the authenticity by content of news is not reliable. The position and opinions of social media users can effectively help identify fake news more accurately. In Yang et al. [37], an SVM classifier with an RBF kernel function is trained to judge rumours on Sina Weibo based on user information, including identity, whether the user has a personal description, gender, age, and username type. In Liu et al. [38], contextual information from tweets is not used but only user context information. Instead of modelling the propagation patterns of news, it constructs the propagation paths of news. Each propagation path is a sequence composed of multiple user feature vectors. Both convolutional neural networks and recurrent neural networks take the propagation path as input, and their output features are used to determine the category of news. These works only consider user information and news content is not exploited.

In Yin et al. [39], CNN is used to extract user features, LSTM to extract textual features, and those features are concatenated to form a joint representation that includes user attributes and textual content. In Chen et al. [40], user multi-view learning and attention are used for rumour detection. It can better learn the representations of different opinions of users in the news propagation path and can fuse the learned representations through fusion mechanisms. In Jiang et al. [41], user attributes are utilized to discover potential user connections in the friendship network, and the news-user network is reconstructed to enhance the embeddings of news and users in the news propagation network, effectively identifying users who are prone to spreading fake news. These works only utilize user and textual information.

3. Model establishment

Figure 1 illustrates our proposed fake news detection model, MFUIE, which is based on multi-modal feature and user information enhancement. The design of this model includes five main components: textual semantic encoding, language-aware visual fusion encoding, image pattern encoding, user information enhancement module, and classification.

3.1. Text semantic encoding

Real sentences often exhibit complexity, such as polysemy, which requires considering the relationships between words to obtain accurate solutions. BERT has powerful contextual understanding and generalization capabilities [42]. By learning on a large-scale corpus, the model has a range of syntactic and commonsense knowledge. Therefore, we use the BERT-base-Chinese model, which is suitable for the Chinese context, for textual semantic encoding. As shown in Figure 2, the BERT-base-Chinese model first encodes the sentence and then trains. Output vector of the last hidden layer is extracted as the final textual semantic representation, as shown in Eq. (1):

$$F_t = \text{bert-base-chinese}(T), F_t \in R^{768} \quad (1)$$

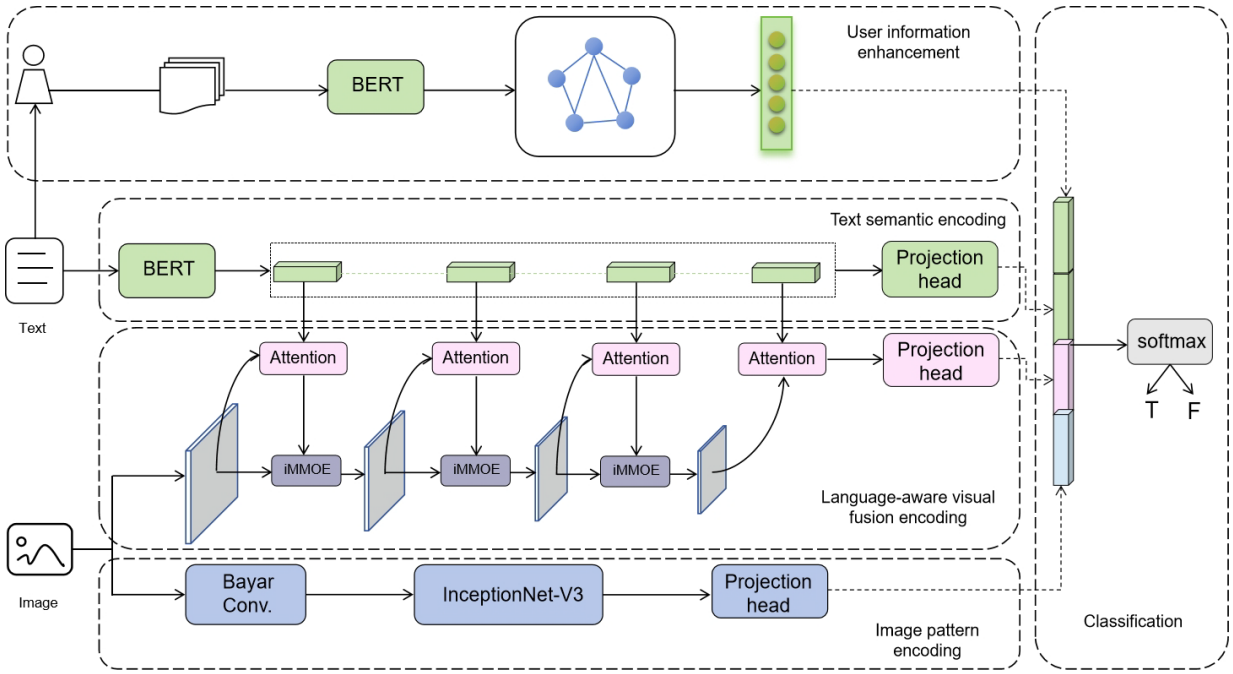


Figure 1. MFUIE model.

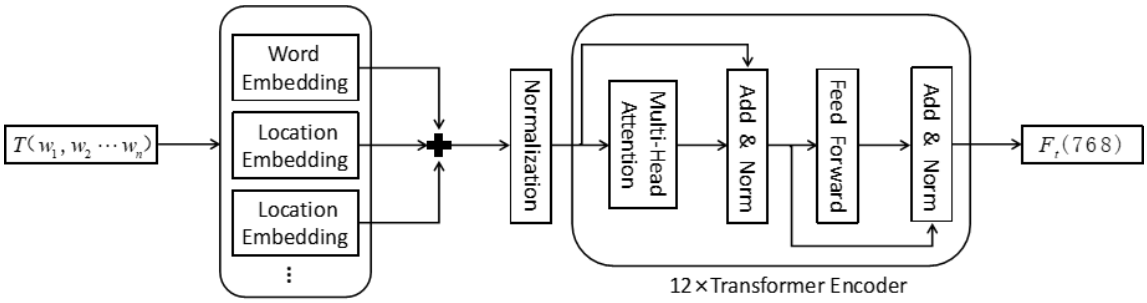


Figure 2. Process of textual semantic encoding.

3.2. Language-aware visual fusion encoding

Swin Transformer provides four stages of hierarchical feature representation. To make full use of visual features, a fusion scheme is proposed, which performs language-aware visual encoding in each Transformer layer at different stage. As shown in Figure 3, at each stage, the pretrained Swin-Base-Patch-Window7-224 model embeds textual semantic features in three steps.

First, the Transformer layer takes the features from the previous stage as input and generates rich visual feature representation $V_i \in R^{C_i \times H_i \times W_i}$.

Next, cross-modal fusion of visual features and textual semantic features is done by an attention mechanism. Attention mechanism is widely used in deep learning. Common attention mechanisms used in cross-modal fusion include co-attention [36], cross-attention [43, 44], etc.

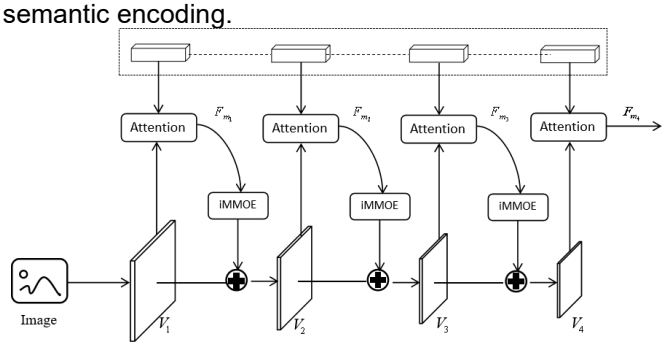


Figure 3. Language-aware visual fusion encoding.

In comparison, the single-head scaled dot-product attention mechanism [45,46] is computationally efficient as it only calculates the attention matrix within an input sequence and takes less memory space. As shown in Figure 4, input visual features V_i is used as queries, and the embedded textual semantic features F_t is used as keys and values to perform single-head scaled dot-product attention. Dot-product computation forms a set of features $G_i \in R^{C_i \times H_i \times W_i}$ with the

same spatial size as V_i. Then, element-wise multiplying G_i with V_i gets the multimodal representation $F_{m_i} \in R^{C_i \times H_i \times W_i}$. The attention mechanism is defined by equations (2), (3), (4), (5), (6):

$$Q = \text{flatten}(V_i \times W_q) \tag{2}$$

$$K = F_t \times W_k \tag{3}$$

$$V = F_t \times W_v \tag{4}$$

$$G_i = \text{unflatten}(\text{softmax}(\frac{Q \times K^T}{\sqrt{d}}) \times V) \tag{5}$$

$$F_{m_i} = V_i \bullet G_i \tag{6}$$

where $W_q \in R^{d_{V_i} \times d}$, $W_k \in R^{d_L \times d}$, $W_v \in R^{d_L \times d}$, and $\otimes$ is matrix multiplication, and $\bullet$ is element-wise multiplication.

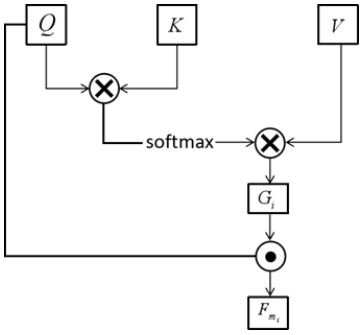


Figure 4. Framework of the attention mechanism [46].

Finally, to better utilize the multimodal context, the multimodal features are fused with visual features and used as the input of next stage. To avoid excessive use of multimodal information, the iMMOE module is introduced to control the flow of multimodal information. Based on the MMOE network, the iMMOE module divides multi-view single-modal representation and the fusion of cross-modal features into different subtasks. This allows to obtain the enhanced visual feature vector E_i, which serves as the input for the next stage of feature extraction and processing, as shown in equations (7) and (8):

$$X_i = F_{m_i} \oplus V_i \tag{7}$$

$$E_i = \sum_{i=1}^{n} (\text{gate}_i^k (\sum_{j=1}^{t} \text{mlp}_k (X_i)) \cdot \text{expert}_i (X_i) \tag{8}$$

where $\oplus$ is the concatenation operation, $Expert_i$ and $Gate_i^k$ are the outputs of the i-th expert and the i-th gate for task k, t is the number of labels, MLP_k is the label attention for task k, and $k \in [1,2]$ is the task assignment. Here, the extraction of single-modal information is task 1 and the fusion of multimodal features is task 2.

3.3. Image pattern encoding

Compared to images in genuine news, images in fake news often characterize by lower quality and periodic recompression or image tampering [47]. InceptionNetV3 model [48] is employed to extract image features, which outperforms traditional CNN models. The specific implementation is illustrated in Figure 5.

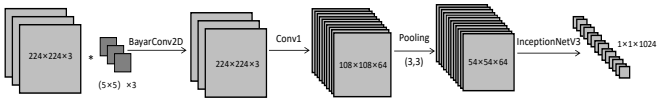


Figure 5. Image pattern encoding.

First, BayarConv [49] is applied to the image to enhance its detail features. Then, after processed by convolutional and pooling layers, the image is fed into InceptionNetV3 for feature extraction, ultimately obtaining the pattern features F_v of the image.

3.4. User information enhancement module

In Ahmad et al. [50], users' historical posts are utilized to model their personality, emotions, and perspectives. Similarly, as shown in Figure 6, textual features also are extracted from users' historical posts to encode their intrinsic preferences.

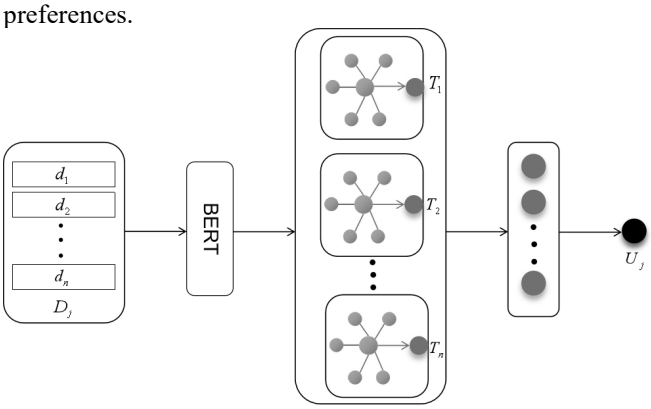


Figure 6. User information enrichment module.

A pretrained BERT-base-Chinese model effectively captures the semantic and contextual information in the posts. For each user's historical post document $D_j = \{ d_1, d_2 \ldots . d_n \}$, due to the length limitation of the BERT-base-Chinese model for input sequences, 50 historical posts couldn't be encoded as a complete sequence. Therefore, individually encoding each post is an alternate. For the 50 historical posts, each post is encoded into a feature vector. Since user posts are usually shorter than news texts, maximum input sequence length of the BERT-base-Chinese model is set to 16 tokens, which speeds up the encoding. After these operations, each sub-sentence is transformed into a fixed-length vector representation, and these feature vectors

are integrated to obtain the final semantic representation $D_j = \{T_1, T_2 \ldots\ldots T_n\}$. The formula is as follows:

$$T_i = BERT - base - Chinese(d_i), T_i \in R^{768} \quad (9)$$

After obtaining the independent vector representations for each sentence, to capture the deep structural information between sentences, the GAT was introduced to establish relationships between sub-sentences. Sub-sentence vector representations T_i is used as node features, where edges represent the relationships between sentences. For each node, the GAT model determines the relationships with its neighbouring nodes by calculating attention weights. Based on the resulting attention weights, the GAT model aggregates the weighted features of neighbouring nodes to obtain the contextual representation. The calculation is as follows:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{a}^T[\mathrm{W}\vec{T}_i \,\|\, \mathrm{W}\vec{T}_j]))}{\sum_{k\in N_i}\exp(\text{LeakyReLU}(\vec{a}^T[\mathrm{W}\vec{T}_i \,\|\, \mathrm{W}\vec{T}_k]))} \quad (10)$$

$$\vec{T}_i = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in N_i}\alpha_{ij}^k W^k \vec{T}_j\right) \quad (11)$$

where a is a mapping $\mathrm{R}^{F'} \times \mathrm{R}^{F'} \to \mathrm{R}$, $W \in \mathrm{R}^{F'\times F}$ is a weight matrix (shared by all $\vec{T}_i$), the weights are calculated by the attention module, which simulates the importance of each neighbouring node T_j to T_i. N is the number of nodes in the node set, and F is dimension of corresponding feature vector. Then, they are averaged to obtain the representation of the user's preferences.

Post feature needs to be further processed to make it compatible with multimodal features as user information. By averaging the feature vectors of all posts, a fixed-length vector is obtained to represent the overall preference of the user in these posts. The formula is as follows:

$$U_i = (T_1[i] + T_2[i] + \ldots + T_{50}[i]) / 50 \quad (12)$$

where $T_1[i]$ represents the value of the first post in the i-th dimension, and $T_2[i]$ represents the value of the second post in the i-th dimension.

3.5. Fusion and classification of multimodal and user information features

To maintain consistency between user information features and the resulting multimodal features, linear concatenation is used, as shown in the formula:

$$Z = [U \oplus F_t \oplus F_m \oplus F_v] \quad (13)$$

where U is the user information features, F_t, F_m, and F_v are the multimodal features extracted from the news content, Z is the total feature vector after concatenation, and symbol $\oplus$ denotes operation of column-wise vector concatenation.

Finally, the concatenated features are fed into a fully connected layer, and its output is passed through a softmax layer to obtain the distribution of classification labels, as shown in formula 14. Cross-entropy is used as the loss function for the model, as shown in formula 15:

$$p = softmax(W_C Z + b_C) \quad (14)$$

$$L = -\sum[y^f \log p^f + (1 - y^f)\log(1 - p^f)] \quad (15)$$

where W_c and b_c are the parameters of the model; y^f is the true label: 1 is fake news and 0 is genuine news; p^f is the probability that the sample is predicted as fake.

4. Experiments and Analysis

4.1. Experimental datasets

The proposed model is evaluated by training and testing on two publicly available Chinese datasets, Weibo [35] and Weibo-21 [51]. The datasets consist of news content and user ID information. To obtain more comprehensive historical post information, user historical posts are crawled based on their user IDs. The most recent 50 posts are crawled for each accessible account.

To apply text representation learning algorithms, data are pre-processed. Firstly, irrelevant characters, including many unrelated foreign language texts, traditional Chinese characters, and special characters are removed, as these data will affect the training and prediction process of the Chinese text model. Then, Jieba, a Chinese word segmentation tool is used to do word segmentation. Next, stop words and other common words are removed. Finally, the processed text is saved as a document $D_i = \{d_1, d_2 \ldots d_n\}$ by user, which d_i represents the ith clause in the document D, so that the file can be used directly in the subsequent experiments. The detailed are shown in Table 1 and Table 2.

Table 1. Distribution of Weibo Dataset

Dataset	training	testing	total
Fake news	2986	829	3815
Real news	3305	826	4131
total	6291	1655	7946

Table 2. Distribution of Weibo-21 Dataset

Dataset	training	testing	total
Fake news	2460	308	2768
Real news	2430	300	2730
total	4890	608	5498

4.2. Experiment setting

The model is implemented using PyTorch and trained using an NVIDIA Tesla V-100 GPU. It utilizes the Adam optimizer with default settings and a fixed learning rate of 0.0001. Due to the balanced ratio of real news to fake news in the training set (approximately 1:1), the threshold for the dataset was set to 0.5. The experiment is done in 100 epochs, with a text embedding dimension of 768 and a batch size of 16. The GAT has 3 layers and 4 multi-head attention. The model is evaluated by cross-validation, and the average accuracy, average precision, average recall, and average F1 score are used as evaluation metrics. They are calculated using equations (16), (17), (18), and (19), respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{19}$$

4.3. Comparisons among different models

Related models

The proposed model is compared with four models:

1.EANN [2]: It is an end-to-end framework based on event adversarial neural networks. It combines textual and visual information by simple concatenation. Its advantage is the introduction of an event discriminator to improve the detection performance. However, its simple concatenation to fuse textual and visual information may result in insufficient fusion.

2.SPOTFAKE [31]: This is a multimodal framework that utilizes BERT to extract textual features and VGG-19 to extract visual features. The advantage is that it exploits features from two different modalities and directly performs real vs. fake classification, reducing training and model size overheads. Its disadvantage is similar to EANN.

3.CAFE [47]: It is an end-to-end model that uses a common attention mechanism to fuse features from three modalities: text semantics, image spatial domain, and frequency domain. The advantage is that it takes image frequency domain features and improves the fusion effect of multimodal features using attention mechanisms. However, it does not further consider the contextual relationship between multimodal features.

4.BMR [53]: It combines cross-modal consistency learning, extracts feature from different modalities and different perspectives of news, and constructs a model with interpretability by bootstrapping multi-representation and optimizing multimodal feature learning. The advantage is the utilization of cross-modal consistency scores to weight a learnable representation, and improve the interpretability of the model. However, it does not utilize the contextual relationships between multimodal features.

Results and analysis

As shown in Table 3, the average performance of the five models is compared on the two datasets and visualized in Figure 7 and Figure 8. It can be found that MFUIE model outperformed the first three methods significantly in accuracy and slightly better than the state-of-the-art fourth model. On the Weibo dataset, MFUIE achieved an accuracy of 92.6%, and on the Weibo-21 dataset, MFUIE achieved an accuracy of 93.5%.

The result implies that MFUIE has improved in extracting and fusing multimodal information and user information features, enhancing its capability in clue mining.

4.4. Ablation experiments

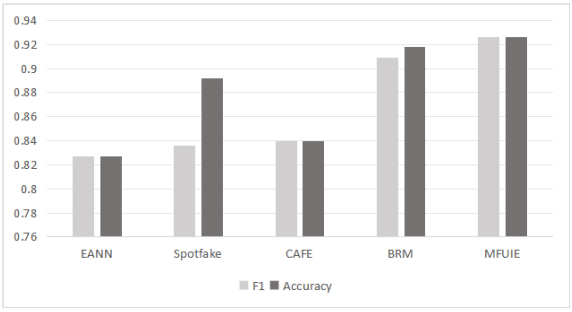
Related module

In the user information enhancement module, the GAT model is introduced to process the deep structure information between user historical posts. Therefore, the effectiveness of GAT is verified by using only the user 's historical posts for ablation experiments, and the experiment is set to 50 rounds. For all clause features after BERT processing, in method 1, GAT is used to enhance the features, then the average value of the features is taken, and finally input into the fully connected layer and the softmax layer for label classification. In method 2, the average value of all clause features obtained by the BERT model is input into the fully connected layer, and then the softmax layer is used for classification.

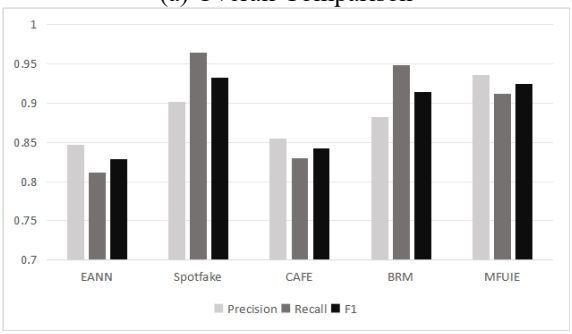
Table 3. Performance Comparison of Different Methods

Dataset	Approach	F1	Accuracy	Fake News			Real News		
				Precision	Recall	F1	Precision	Rcall	F1
Weibo	EANN	0.827	0.827	0.847	0.812	0.829	0.807	0.843	0.825
	Spotfake	0.836	0.892	0.902	**0.964**	**0.932**	0.847	0.656	0.739

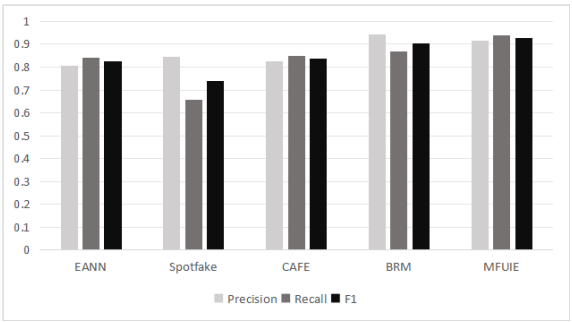
	CAFE	0.841	0.840	0.855	0.830	0.842	0.825	0.851	0.837
	BRM	0.909	0.918	0.882	0.948	0.914	**0.942**	0.870	0.904
	MFUIE	**0.927**	**0.926**	**0.936**	0.912	0.924	0.917	**0.940**	**0.929**
Weibo -21	EANN	0.869	0.870	0.902	0.825	0.862	0.841	0.912	0.875
	Spotfake	0.847	0.851	**0.953**	0.733	0.828	0.786	**0.964**	0.866
	CAFE	0.881	0.882	0.857	0.915	0.885	0.907	0.844	0.876
	BRM	0.926	0.929	0.908	**0.947**	0.927	**0.946**	0.906	0.925
	MFUIE	**0.935**	**0.935**	0.942	0.926	**0.934**	0.944	0.929	**0.936**



(a) Overall Comparison



(b) Comparison on Fake News



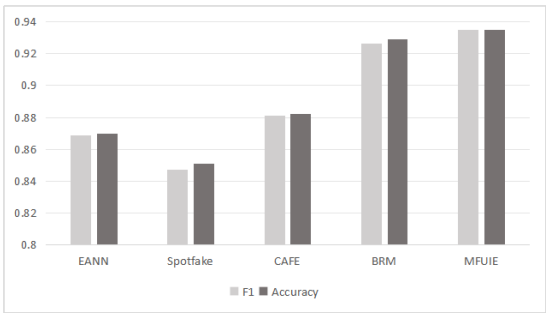
(c) Comparison on Real News

Figure 7. Performance Comparison of Different Models on the Weibo Dataset

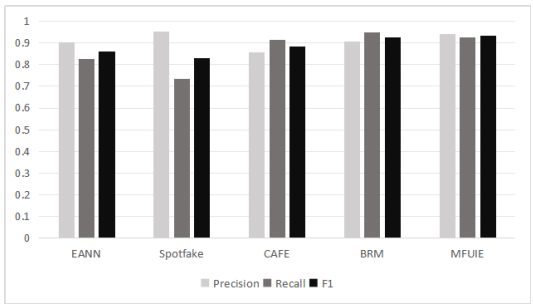
Analysis of effect

For the MFUIE model, the impact of introducing user information features on the classification performance is verified.

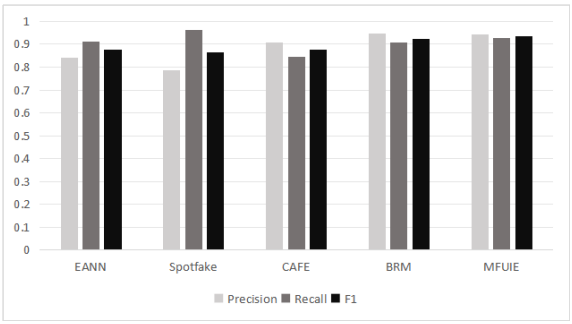
It can be seen from Figure 9, when epoch<30, the loss value of BERT fluctuates sharply; while in Figure 10, when epoch<25, this trend occurs. In the case of adding the GAT model, the loss decline trend is more stable than the case without GAT. This shows that the GAT model can help improve the convergence of the model and reduce the oscillation and instability at early stage in the training process. This shows that by introducing the GAT model, the model can converge at early stage with smaller epochs.



(a)Overall Performance Comparison of Models



(b)Performance Comparison of Models on Fake News



(c)Performance Comparison of Models on Real News

Figure 8. Performance Comparison of Various Models on the Weibo-21 Dataset

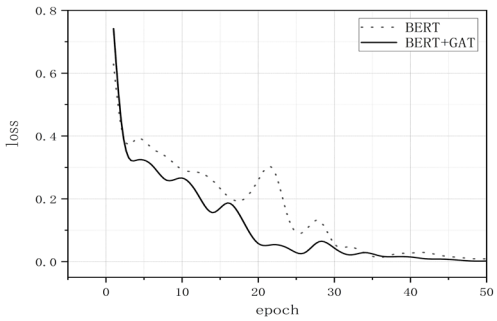


Figure 9. Loss Curve of the Two Methods Based on Weibo Dataset

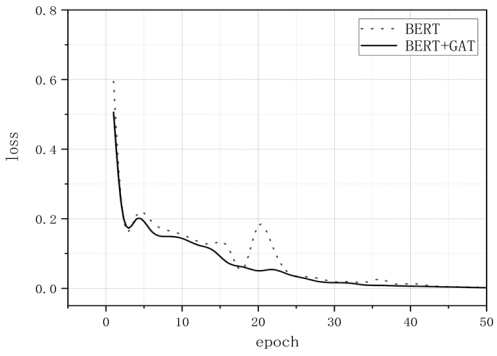


Figure 10. Loss Curve of the Two Methods Based on Weibo-21 Dataset

Figures 11 and 12 demonstrate the trend curves of the accuracy. It can be observed that in BERT with GAT model, the accuracy trend is more stable, the accuracy is always above 0.9 when epoch>10, with a faster convergence.

As shown in Table 4, the F1 of model with user information enhancement on Weibo is 0.926, greater than that of without user information 0.918. The F1 of model with user information enhancement on Weibo-21 is 0.935, greater than that of without user information 0.930.

The average performance of the models with and without user information enhancement module on the two datasets are compared and visualized in Figure 13 and Figure 14.

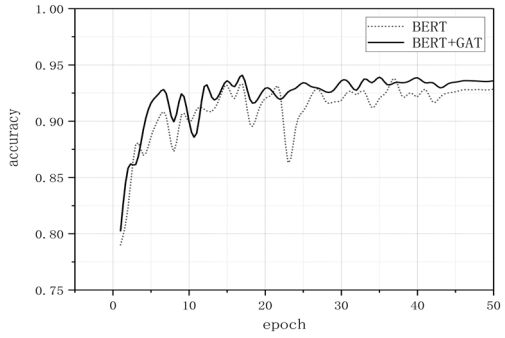


Figure 11. Accuracy Curve of the Two Methods on Weibo Dataset

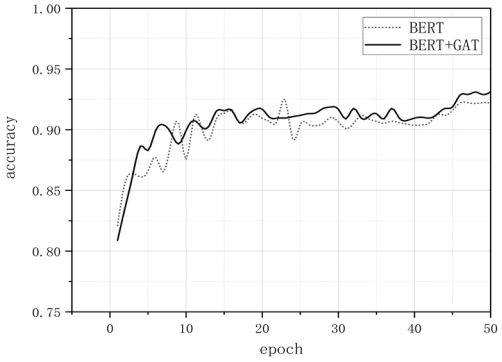
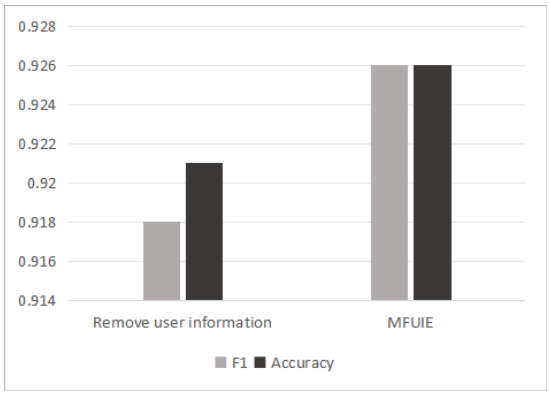


Figure 12. Accuracy Curve of the Two Methods on Weibo-21 Dataset

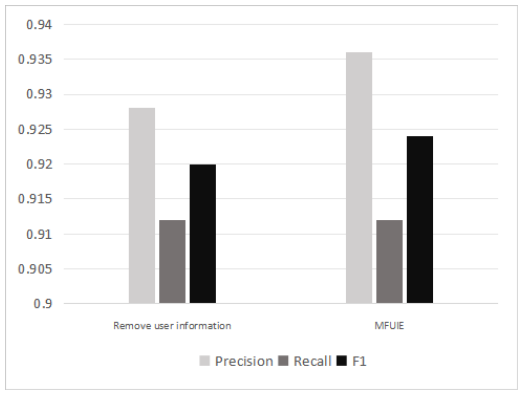
Table 4. Performance Comparison of Different Modules

Dataset	Approach	F1	Accuracy	Fake News			Real News		
				Precision	Recall	F1	Precision	Rcall	F1
Weibo	without user information	0.918	0.921	0.928	**0.912**	0.920	0.903	0.930	0.916
	MFUIE	**0.926**	**0.926**	**0.936**	**0.912**	**0.924**	**0.917**	**0.940**	**0.929**

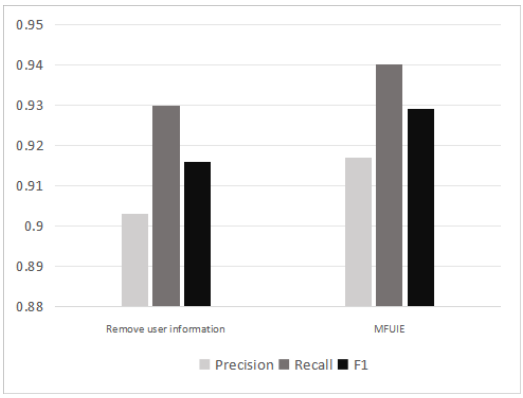
Weibo-21	without user information	0.930	0.932	0.937	0.922	0.929	0.936	0.925	0.930
	MFUIE	**0.935**	**0.935**	**0.942**	**0.926**	**0.934**	**0.944**	**0.929**	**0.936**



(a) Overall Performance Comparison

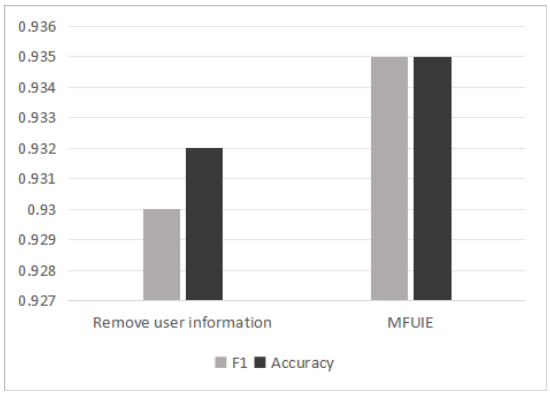


(b) Performance Comparison on Fake News

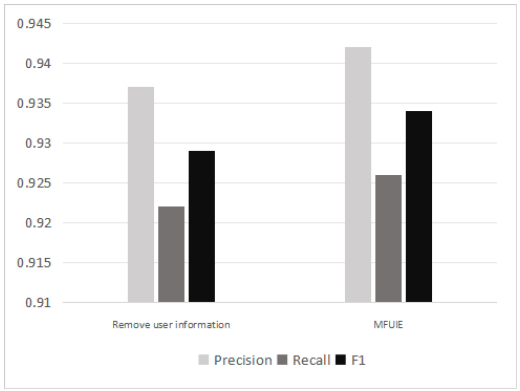


(c) Performance Comparison on Real News

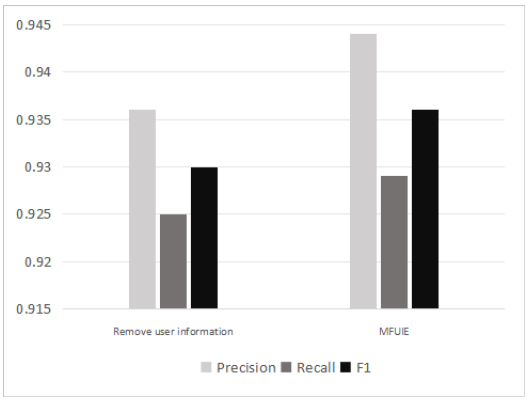
Figure 13. Performance Comparison of Models on Weibo Dataset



(a) Overall Performance Comparison



(b) Performance Comparison on Fake News



(c) Performance Comparison on Real News

Figure 14. Performance Comparison on Weibo-21 Dataset

It can be found that adding user information enhancement module was beneficial for the overall classification performance, especially in precision for real news.

On Weibo dataset, F1 of MFUIE increased by more than 0.8% compared to the model without the user information enhancement module, particularly with an improvement of over 1.4% in precision for real news.

On Weibo-21 dataset, MFUIE achieved an accuracy improvement of more than 0.3% compared to the model without the user information enhancement module, especially with a performance improvement of over 0.8% in precision for real news. This further demonstrates the importance of user information.

5. Conclusion and Future Work

Existing multimodal fake news detection models only fuse the visual and textual features after the encoder, failing to effectively utilize the multimodal context relationships, resulting in insufficient feature fusion. Moreover, most fake news detection algorithms focus on mining news content while neglecting user-related auxiliary information. MFUIE proposed integrates the multimodal content of news with user information features. MFUIE is compared with those of existing methods. MFUIE model achieves an accuracy of 0.926 and 0.935 on the Weibo dataset and Weibo-21 dataset, respectively. F1 on Weibo is 0.926, 0.017 greater than SOAT model BRM; while F1 on Weibo-21 is 0.935, 0.009 greater than that of BRM. These results demonstrate that MFUIE can improve the fake news recognition in some degree.

This year, Hu et al. [52] investigated if large LMs (Language Models) help in fake news detection and how to properly utilize their advantages for improving performance. Results show that the large LM (GPT-3.5) underperforms the task-specific small LM (BERT), but could provide informative rationales and complement small LMs in news understanding. Based on these, they design an adaptive rationale guidance network (ARG), in which SLMs selectively acquire insights on news analysis from the LLMs' rationales. But rationales generated by LLM may add noise to the detection, how to remove the noise is worth exploring.

Zhang et al. [30] use LLM to provide external knowledge, and AKA-Fake learns a compact knowledge subgraph to further improve the fake news detection in cross-modality news. To measure the importance of the concepts for the text, the Concept Kernel Attention Network (CKAN) is proposed in Li H et al. [54]. It uses the text-to-concept attention mechanism (TCAM) and the entity-to-concept attention mechanism (ECAM) to assign weights to concepts, which limit the importance of noisy concepts as well as contextually irrelevant concepts and assign more weights to concepts that are important for classification. We can borrow their idea to design appropriate attention mechanism to reduce the noise introduced by external knowledge.

In our work, Adam optimizer is employed. In Öztürk MM et al. [55], the results of four optimization methods including Neldermead, Genetic algorithm, Bayes, and Random search are evaluated for that validation. The best method is selected by comparing the prediction accuracies of the optimization methods. In future, we will evaluate different optimizers on small data set to decide the best optimizer for fake news detection. Also, we can combine global optimizer like evolutionary methods in the deep learning framework to find global maxima.

References

[1] Qi, Cao, & Sheng. (2021). Semantic-enhanced multimodal false news detection. C*omputer Research and Development*, 58 (7), 1456-1465.

[2] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., & Gao, J. (2018, July). Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 849-857).

[3] Zhou, X., Wu, J., & Zafarani, R. (2020). Safe: similarity-aware multi-modal fake news detection. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 354-367). Cham: Springer International Publishing.

[4] Zhang, H., Fang, Q., Qian, S., & Xu, C. (2019). Multi-modal knowledge-aware event memory network for social media rumor detection. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 1942-1951).

[5] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., Ma, J. & Wang, K. (2021) Image preprocessing in classification and identification of diabetic eye diseases. Data Science and Engineering 6(4): 455–471.

[6] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2021) Convolutional neural network for multi-class classification of diabetic eye disease. EAI Endorsed Transactions on Scalable Information Systems 9(4). doi: https://doi.org/10.4108/eai.16-12-2021.172436.

[7] Tawhid, M.N.A., Siuly, S., Wang, K. & Wang, H. (2023) Automatic and efficient framework for identifying multiple neurological disorders from eeg signals. IEEE Transactions on Technology and Society 4(1): 76–86. doi: https://doi.org/10.1109/TTS.2023.3239526.

[8] Alvi, A.M., Siuly, S., & Wang, H. (2023) A long short term memory based framework for early detection of mild cognitive impairment from eeg signals. IEEE Transactions on Emerging Topics in Computational Intelligence 7(2): 375–388. doi: https://doi.org/10.1109/TETCI.2022.3186180.

[9] Alvi, A.M., Siuly, S., Wang, H., Wang, K., & Whittaker, F. (2022) A deep learning based framework for diagnosis of mild cognitive impairment. Knowledge-Based Systems 248: 108815. doi: https://doi.org/10.1016/j.knosys.2022.108815.

[10] Tawhid, M.N.A., Siuly, S., Wang, H., Whittaker, F., Wang, K., & Zhang, Y. (2021) A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from eeg. Plos One 16(6): e0253094. doi: https://doi.org/10.1371/journal.pone.0253094.

[11] Laidi A, Ammar M, Daho MEH, Mahmoudi S. GAN Data Augmentation for Improved Automated Atherosclerosis Screening from Coronary CT

Angiography. EAI Endorsed Scalable Information System [Internet]. 2022 May 17 [cited 2024 Jun. 3];10(1):e4. Available from: https://publications.eai.eu/index.php/sis/article/view/1027

[12] Hao X, Zhang C, Xu S. Fast Lung Image Segmentation Using Lightweight VAEL-Unet. EAI Endorsed Scalable Information System [Internet]. 2024 Apr. 8 [cited 2024 Jul. 13];11(6). Available from: https://publications.eai.eu/index.php/sis/article/view/4788

[13] Singh, R., Subramani, S., Du, J., Zhang, Y., Wang, H., Miao, Y., & Ahmed, K. (2023) Antisocial behavior identification from twitter feeds using traditional machine learning algorithms and deep learning. EAI Endorsed Transactions on Scalable Information Systems 10(4). doi: https://doi.org/10.4108/eetsis.v10i3.3184.

[14] Du, J., Rong, J., Wang, H., & Zhang, Y. (2021). Neighbor-aware review helpfulness prediction. Decision Support Systems, 148, 113581.

[15] Xu, S., Song, Y., & Hao, X. (2022) A comparative study of shallow machine learning models and deep learning models for landslide susceptibility assessment based on imbalanced data. Forests 13(11). doi: https://doi.org/10.3390/f13111908 URL https://www.mdpi.com/1999-4907/13/11/1908.

[16] Cao, Q., Hao, X., Ren, H., Xu, W., Xu, S., & Asiedu, C.J. (2022) Graph attention network based detection of causality for textual emotion-cause pair. World Wide Web: 1–15 doi: https://doi.org/10.1007/s11280-022-01111-5, URLhttps://doi.org/10.1007/s11280-022-01111-5.

[17] Sun, R. Y. (2020). Optimization for deep learning: An overview. Journal of the Operations Research Society of China, 8(2), 249-294.

[18] Ge, Y. F., Wang, H., Bertino, E., Zhan, Z. H., Cao, J., Zhang, Y., & Zhang, J. (2023). Evolutionary dynamic database partitioning optimization for privacy and utility. IEEE Transactions on Dependable and Secure Computing, 21(4), 2296-2311. doi: https://doi.org/10.1109/TDSC.2023.3302284.

[19] Ge, Y. F., Yu, W. J., Cao, J., Wang, H., Zhan, Z. H., Zhang, Y., & Zhang, J. (2020). Distributed memetic algorithm for outsourced database fragmentation. IEEE Transactions on Cybernetics, 51(10), 4808-4821.

[20] Wang, C., Sun, B., Du, K. J., Li, J. Y., Zhan, Z. H., Jeon, S. W., ... & Zhang, J. (2024). A novel evolutionary algorithm with column and sub-block local search for sudoku puzzles. IEEE Transactions on Games, 16(1), 162-172. doi: https://doi.org/10.1109/TG.2023.3236490.

[21] Shi, W., Chen, W. N., Kwong, S., Zhang, J., Wang, H., Gu, T., ... & Zhang, J. (2021). A coevolutionary estimation of distribution algorithm for group insurance portfolio. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52(11), 6714-6728.

[22] Huang, T., Gong, Y. J., Chen, W. N., Wang, H., & Zhang, J. (2020). A probabilistic niching evolutionary computation framework based on binary space partitioning. IEEE Transactions on Cybernetics, 52(1), 51-64. doi: https://doi.org/10.1109/tcyb.2020.2972907

[23] Chen, Z.-G., Zhan, Z.-H., Wang, H., & Zhang, J. (2019). Distributed individuals for multiple peaks: a novel differential evolution for multimodal optimization problems. IEEE Transactions on Evolutionary Computation, 1–1. doi: https://doi.org/10.1109/tevc.2019.2944180

[24] Li, J. Y., Du, K. J., Zhan, Z. H., Wang, H., & Zhang, J. (2023). Distributed differential evolution with adaptive resource allocation. IEEE Transactions on Cybernetics, 53(5), 2791-2804. doi: https://doi.org/10.1109/TCYB.2022.3153964.

[25] Liu, W. L., Gong, Y. J., Chen, W. N., Liu, Z., Wang, H., & Zhang, J. (2019). Coordinated charging scheduling of electric vehicles: A mixed-variable differential evolution approach. IEEE Transactions on Intelligent Transportation Systems, 21(12), 5094-5109

[26] Huang, T., Gong, Y.-J., Kwong, S., Wang, H., & Zhang, J. (2019). A niching memetic algorithm for multi-solution traveling salesman problem. IEEE Transactions on Evolutionary Computation, 1–1. doi: https://doi.org/10.1109/tevc.2019.2936440

[27] Li, J. Y., Zhan, Z. H., Wang, H., & Zhang, J. (2020). Data-driven evolutionary algorithm with perturbation-based ensemble surrogates. IEEE Transactions on Cybernetics, 51(8), 3925-3937. doi: https://doi.org/10.1109/tcyb.2020.3008280

[28] Yang, J. Q., Yang, Q. T., Du, K. J., Chen, C. H., Wang, H., Jeon, S. W., ... & Zhan, Z. H. (2023). Bi-directional feature fixation-based particle swarm optimization for large-scale feature selection. IEEE Transactions on Big Data, 9(3), 1004-1017. doi: https://doi.org/10.1109/TBDATA.2022.3232761

[29] B. B, Rani KS, Neog A. Finding Multidimensional Constraint Reachable Paths for Attributed Graphs. EAI Endorsed Scalable Information System [Internet]. 2022 Aug. 22 [cited 2024 Jun. 3];10(1): e8. Available from: https://publications.eai.eu/index.php/sis/article/view/2581

[30] Zhang, L., Zhang, X., Zhou, Z., Huang, F., & Li, C. (2024). Reinforced Adaptive Knowledge Learning for Multimodal Fake News Detection. In Proceedings of the AAAI Conference on Artificial Intelligence 38(15): AAAI-24 Technical Tracks 15, 16777-16785

[31] Singhal, S., Shah, R. R., Chakraborty, T., Kumaraguru, P., & Satoh, S. I. (2019). Spotfake: A multi-modal framework for fake news detection. In *2019 IEEE fifth International Conference on Multimedia Big Data (BigMM)* (pp. 39-47). IEEE.

[32] Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). Mvae: Multimodal variational autoencoder for fake news detection. In The World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 2915–2921. https://doi.org/10.1145/3308558.3313552

[33] Qu, Z., Meng, Y., Muhammad, G., & Tiwari, P. (2024). Qmfnd: A quantum multimodal fusion-based fake news

detection model for social media. Information Fusion, 104, 102172.

[34] Xue, J., Wang, Y., Tian, Y., Li, Y., Shi, L., & Wei, L. (2021). Detecting fake news by exploring the consistency of multimodal data. *Information Processing & Management*, 58(5), 102610.

[35] Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In Proceedings of the 25th ACM International Conference on Multimedia (MM '17). Association for Computing Machinery, New York, NY, USA, 795–816. https://doi.org/10.1145/3123266.3123454.

[36] Wu, Y., Zhan, P., Zhang, Y., Wang, L., & Xu, Z. (2021). Multimodal fusion with co-attention networks for fake news detection. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021,*. 2560-2569.

[37] Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, 1-7.

[38] Liu, Y., & Wu, Y. F. (2018). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 354-361.

[39] Yin Pengbo, Pan Weimin, Peng Cheng, & Zhang Haijun. (2020). Research on early detection of Weibo rumours based on user feature analysis. *Intelligence Magazine*, 39 (7), 81-86.

[40] Chen, X., Zhou, F., Trajcevski, G., & Bonsangue, M. (2022). Multi-view learning with distinguishable feature fusion for rumor detection. *Knowledge-Based Systems*, 240, 108085.

[41] Jiang, S., Chen, X., Zhang, L., Chen, S., & Liu, H. (2019). User-characteristic enhanced model for fake news detection in social media. In *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part I 8*, 634-646.

[42] Jacob Devlin, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171-4186.

[43] Luo, G., Zhou, Y., Ji, R., Sun, X., Su, J., Lin, C. W., & Tian, Q. (2020). Cascade grouped attention network for referring expression segmentation. In *Proceedings of the 28th ACM International Conference on Multimedia* ,1274-1282.

[44] Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M.M. (2019). Fake News Detection on Social Media using Geometric Deep Learning. *ArXiv, abs/1902.06673*.

[45] Tuan, N. M. D., & Minh, P. Q. N. (2021). Multimodal fusion with BERT and attention mechanism for fake news detection. In *2021 RIVF International Conference on Computing and Communication Technologies (RIVF)* (pp. 1-6). IEEE.

[46] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

[47] Qi, P., Cao, J., Yang, T., Guo, J., & Li, J. (2019). Exploiting multi-domain visual information for fake news detection. In *2019 IEEE International Conference on Data Mining (ICDM)* (pp. 518-527). IEEE.

[48] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818-2826).

[49] Bayar, B., & Stamm, M. C. (2018). Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection. *IEEE Transactions on Information Forensics and Security*, 13(11), 2691-2706.

[50] Ahmad, N., & Siddique, J. (2017). Personality assessment using Twitter tweets. *Procedia computer science, 112*, 1964-1973.

[51] Nan, Q., Cao, J., Zhu, Y., Wang, Y., & Li, J. (2021). Mdfend: Multi-domain fake news detection. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 3343-3347).

[52] Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., & Qi, P. (2024). Bad actor, good advisor: Exploring the role of large language models in fake news detection. In Proceedings of the AAAI Conference on Artificial Intelligence 38 (20), 22105-22113

[53] Ying, Q., Hu, X., Zhou, Y., Qian, Z., Zeng, D., & Ge, S. (2023). Bootstrapping multi-view representations for fake news detection. In Proceedings of the AAAI conference on Artificial Intelligence,5384-5392.

[54] Li, H., Huang, G., Li, Y., Zhang, X., & Wang, Y. (2023). Sentence classification based on the concept kernel attention mechanism. EAI Endorsed Transactions on Scalable Information Systems, 10(1), e3-e3.

[55] Öztürk MM. Developing a hyperparameter optimization method for classification of code snippets and questions of stack overflow: HyperSCC. EAI Endorsed Scalable Information System [Internet]. 2022 May 27 [cited 2024 Jun. 3];10(1):e5. Available from: https://publications.eai.eu/index.php/sis/article/view/1267