A Review of Image and Text Feature Extraction Methods in Fake News Detection Tasks

Feng Li[1,2] , Marshima Mohd Rosli[1*] , Yujiang Wang[1]

[1] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM), Shah Alam 40450, Malaysia
[2] College of Computer and Information Engineering, Hebei Finance University, Baoding 071000, China

Corresponding Author Email: marshima@fskm.uitm.edu.my

ABSTRACT

With the high-speed development of multimedia technologies, news content is very rich, including not only text but also image information. One of the most essential approaches for detecting fake news is through content analysis. Feature extraction and representation are the crucial steps of the task. How to accurately characterize news content is still a challenging problem. This article seeks to assist readers comprehend the various strategies connected with feature extraction and representation. Therefore, we scan various digital libraries to find all relevant papers published since 2010. This paper reviews methods that can extract and represent features from three perspectives: text, image, and multi-modal. In particular, we count the usage of these methods in various fake news detection tasks and detail the related theories. We hope that this review can promote the advancement of machine learning, neural networks, and other technologies so as to provide better services for fake news detection.

1. INTRODUCTION

Given the network's rapid expansion, especially the rise of 5G networks, the social media platforms have also accelerated the spread of all kinds of fake news. Fake news is acknowledged by all countries as one of the biggest threats to economic and social stability. Due to its negative impacts, fake news detection (FND) has become increasingly important, and this problem has attracted enormous public attention, as well as rising attention from the scholarly community. As far as we know, the existing review articles on FND are all about different detection methods, mainly discussing their basic theories, frameworks, and respective advantages and disadvantages. The academic literature on FND has reviewed and evaluated methods from four perspectives [1]. Previous studies have explored the existing methods and techniques for identifying and mitigating fake news [2], focusing on the advantages and limitations of each approach. Previous research has established the primary methods for detecting fake news that are now accessible, as well as how these methods can be used in various contexts [3]. But they all ignored the most important part of the FND, which is feature extraction and representation (FE&R).

Through the literature review of the past thirteen years (2010-2023), We can verify the authenticity of news in terms of knowledge, style, dissemination, and sources [1]. We can also test the authenticity of news from the perspectives of content, feedback, and intervention [2]. We can even carry out false news detection from the perspective of context and feature fusion [4]. Regardless of the method used, the main technologies used are traditional machine learning and deep learning. The so-called knowledge-based FND is essentially to carry out fact verification. Automatic news fact verification is made up of two primary components: fact extraction and fact checking. With the tremendous growth of science and technology, automatic fact-verification approaches have been created that heavily rely on SVM, Decision Trees, Machine Learning, and neural network technology [5]. Both the style-based FND method and the knowledge-based detection method analyze the content of news. The difference is that knowledge-based detection approaches primarily examine the veracity of news, whereas style-based methods assess the goal of news and decide whether it has a subjective intention to deceive readers. We think that certain malicious entities create fake news to manipulate people's trust, and these fake news articles have a distinct type of content. The style of news can be represented by machine learning features, which are composed of textual features [6] and visual features [7]. The FND method based on propagation first obtains the relevant information in the process of news communication and then constructs a tree or graph structure for binary classification tasks (true or fake). The entity that first posts news articles is represented by the root node of the tree structure, and users who forward articles are represented by other nodes [1]. A graph structure can more flexibly capture news dissemination data, and a heterogeneous graph structure has different types of nodes and edges [2].

Feature extraction and feature representation are closely related but distinct concepts in the fields of machine learning and data analysis. Feature extraction involves transforming raw data into a more compact and meaningful representation by extracting relevant information or patterns. This procedure seeks to minimize the dimension of the data, filter out noise, and capture the most relevant features. Feature extraction

techniques can be domain-specific or general-purpose and may include methods like edge detection, local feature extraction, or frequency analysis. On the other hand, feature representation refers to the encoding or representation of the extracted features in a manner appropriate for machine learning techniques or other data analysis techniques. It involves transforming the extracted features into a numeric or structured format that can be easily processed by algorithms. The type of feature representation used is determined by the nature of the data and the specific purpose. They are widely used in domains including machine learning, computer vision, and natural language processing. Table 1 shows the summary of feature extraction technologies in FND. There are numerous applications of FE&R, such as text classification [8], image recognition [9], news filtering [10], machine translation [11], and spam detection [12]. We categorize existing evaluated studies into two types: The first type is based on a single-modal approach, which only extracts and represents features of news text or image content [13]. The second type is a multi-modal approach based on mixed representations of news' text and image feature extraction [14].

FE&R is a fundamental problem in FND. Regardless of which detection method is summarized above, FE&R are always crucial steps. The sole distinction is that some concentrate on the textual or visual aspects of news information, while others consider both aspects simultaneously. In a follow-up study, Conroy et al. [6] suggested that textual aspects describe content style on four levels of language: lexicon, syntax, discourse, and semantics. They hold the opinion that malicious entities seek to provide fake information in a specific manner to captivate readers and gain their trust. Jin et al. [15] not only pointed out that images are extremely popular and have a significant impact on Weibo news propagation, but also claimed that image propagation patterns differ between fake news and real news events. Therefore, they characterize these patterns visually and statistically for FND. According to the study by Qian et al. [16], social media posts contain a large amount of multi-model features that can be utilized to determine whether the news is fake by merging textual and visual information.

Table 1. The summary of feature extraction in FND

Methods	Literatures	Feature Extraction		
		Text	Image	Text-Image
Knowledge	[5, 17, 18]	√		
Style	[6, 7, 15, 19, 20]	√	√	
Propagation	[21, 22]	√		
Source	[23]	√		
Content	[8, 9, 14]	√	√	√
Feedback	[24, 25]	√		
Intervention	[26, 27]	√		
Context	[28-31]	√	√	√
Feature fusion	[32-35]	√		√
Machine learning	[7, 30]	√	√	√
Deep Learning	[36-40]	√	√	√

This review seeks to aid knowledge of the various FE&R methodologies. Therefore, we search across multiple digital libraries to find all relevant papers published since 2010. We hope that this paper can promote the evolution of machine learning, neural networks, and other technologies so as to provide better services for FND.

The remainder of the paper is arranged as follows: Section 2 delves into the most prevalent textual feature representation methods, discussing their properties and some famous applications in the reference. Section 3 summarizes the approaches used to express visual features. Section 4 provides a detailed introduction to multi-modal feature representation methods. Finally, some important discussions and conclusions are proposed in Sections 5 and 6.

2. TEXT-BASED FEATURE REPRESENTATION

Textual feature representation refers to the methods and techniques used to extract meaningful information from text data and represent it in a format suitable for analysis or machine learning tasks (Figure 1).

2.1 Bag-of-Words (BoW)

It is one of the most common methods for the classification of text [41, 42]. This basic technique represents a document as a collection of words, disregarding the order or grammar. Every word in the document is considered a distinct feature, and the feature value is determined by how frequently each word appears. BoW is simple but doesn't capture the semantic relationships between words [43]. An elementary illustration of a text-based BoW model is provided, commencing with the following pair of basic text documents:

Document1: "human machine interface for computer applications"

Document2: "human machine system engineering testing of interface"

Create the following dictionary using the words found in the aforementioned two documents:

Vocabulary= {"human": 2, "machine": 2,"interface": 2, "for": 1, "computer": 1, "applications": 1, "system": 1, "engineering": 1,"testing": 1, "of": 1}.

Since there are ten words in the aforementioned dictionary, each with its own index, we may represent each sentence with a 10-dimensional vector. The number of times a word appears in the document can be indicated using the integer values 0-n, where n is a positive integer:

Document 1: [1,1,1,1,1,1,0,0,0,0]

Document 2: [1,1,1,0,0,0,1,1,1,1]

This vector represents the frequency with which each dictionary word occurs in the text, rather than the order in which they appear in the original text.

2.2 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF, a statistical measurement method, is used to evaluate the importance of words in documents in a set or corpus [1, 2, 16].

Each word is given a weight determined by its inverse frequency in the corpus and its frequency in the document. TF-IDF helps identify words that are relatively more important and distinctive in a document compared to others. Given a corpus D, a word w, and a single document $d \in D$, we compute:
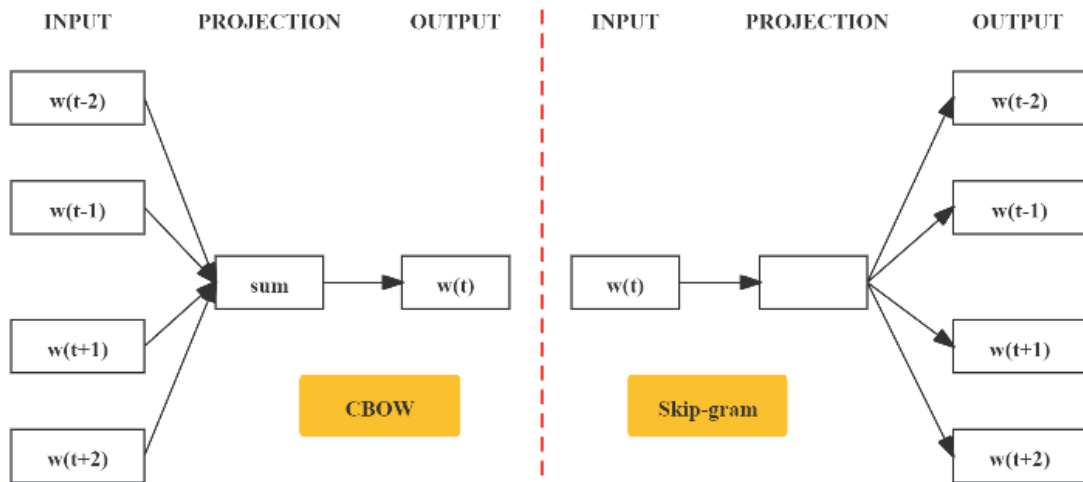
$$TF - IDF = TF \times IDF \qquad (1)$$
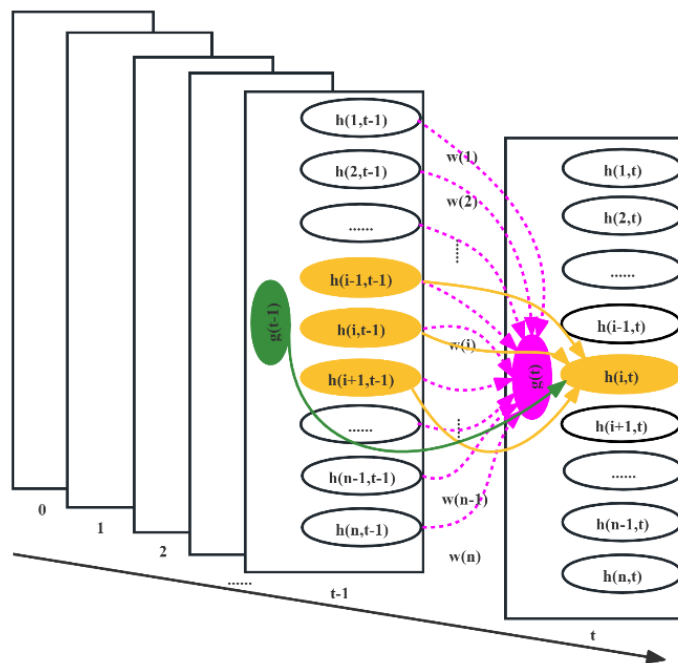
$$TF = \frac{f_{w,d}}{|d|} \qquad (2)$$
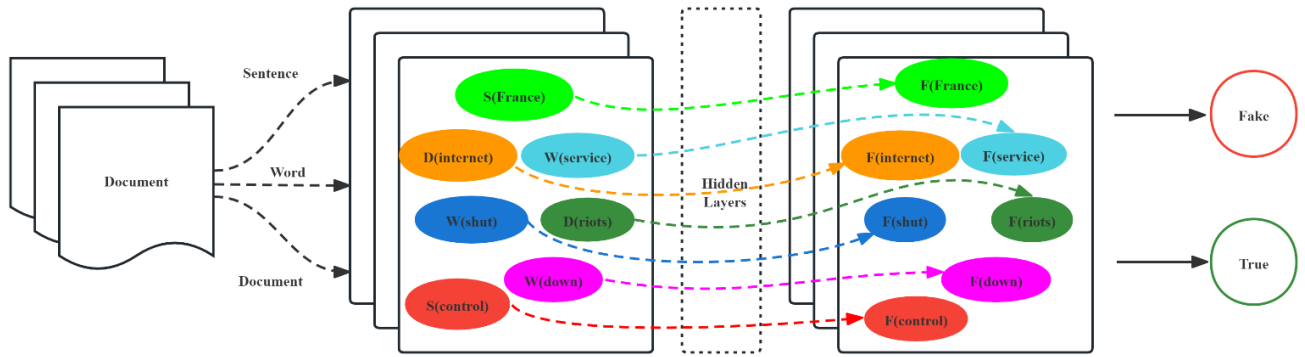
Figure 1. Schematic of text-based feature representation

Nodes begin with "D" are document nodes, with "S" are sentence nodes, with "W" are word nodes. F(x) means the representation of x.

Figure 2. Two models in Word2vec

Figure 3. The scheme of sentence-state LSTM

411

$$IDF = \log \frac{|D|}{f_{w,D} + 1} \qquad (3)$$

where, TF represents the frequency of a word appearing in a document, $f_{w,d}$ is equivalent to the quantity of times w appears in d, $|d|$ is the dimensions of the document, and IDF represents the inverse document frequency, which a word's frequency is inversely proportionate to its value, $|D|$ is the corpus's size, and $f_{w,D}$ is equivalent to the quantity of documents in D where w appears.

2.3 Word embeddings

In a continuous vector space, words with similar meanings are positioned closer to one another, and this is how word embeddings depict words. Often used word embedding methods include Word2Vec [44], GloVe [45], and FastText [46]. Many different natural language processing tasks make use of these embeddings, which capture the semantic and contextual links between words. For example, in Word2vec, the CBOW (Continuous Bag of Word) and skip-gram models were presented. The former predicts the current word based on the contextual word, while the latter predicts its contextual word based on the current word (Figure 2).

2.4 Deep learning architectures

Advanced deep learning models, such as RNN [47] and CNN [8], have been successfully applied to text representation tasks. These models can capture complex relationships and dependencies in text data, enabling more accurate FE&R. According to the study by Zhang et al. [48], a novel LSTM structure for text encoding was explored that uses a parallel state for each word and models the concealed conditions of every word at once during every iterative phase, as opposed to one word at a time (Figure 3). The Text Graph Convolutional Network (TextGCN) is suggested to use a single text graph

and a graph convolutional network to train word and text embeddings [49].

These technologies are used in various applications, including text classification [8], sentiment analysis, information retrieval, document summary, machine translation, and many more. The particular task at hand and the properties of the text data under analysis will determine which technique is best.

3. IMAGE-BASED FEATURE REPRESENTATION

Tian et al. [9, 50, 51] extracted relevant features from images to identify potential signs of fake news or incorrect information. Although most FND methods focus on textual analysis, incorporating visual information from images can provide additional insights and enhance the overall detection accuracy [52, 53].

3.1 Image feature extraction

Image feature extraction can be carried out in three aspects: color features [54, 55], texture features [49, 56], and shape features. Due to the widespread belief that textures are used for recognition and interpretation by the human visual system, we will focus on introducing texture feature extraction (TFE). As stated by the field of texture feature extraction, it is mainly separated into two categories: spectral texture feature extraction and spatial texture feature extraction. The former uses pixel statistics, while the latter extracts texture features by calculating the modified image's features [56]. Table 2 summarizes their respective advantages and disadvantages. Keyvanpour et al. [56] classified texture feature methods into three types: statistical, model-based, and filter-based. Before conducting qualitative evaluations and comparisons, they provided a comprehensive introduction to each representative method, as shown in Table 3.

Table 2. Contrast of texture feature extraction

TFE	Advantage	Disadvantage
Spatial texture	Interesting, easy to understand, able to extract from any shape without losing information.	Noise sensitivity prone to distortion
Spectral texture	Strong robustness, less computational work	Without semantic meaning, Need sufficient space

Table 3. Comparing and evaluating TFE qualitatively

Approach	Method	Accuracy	Generality	Noise Sensitivity	Precision	Time Complexity
Filter	Fourier Transforms	moderate	moderate	moderate	high	moderate
	Gabor Filters	higher	higher	low	higher	higher
	Laws Filter Masks	moderate	low	higher	low	high
	Wavelet Transforms	high	high	low	higher	low
Model	AR Models	low	lower	high	low	moderate
	Fractal Models	low	moderate	low	low	moderate
	MRFs	moderate	high	high	moderate	higher
Statistical	Auto-correlation	low	low	high	low	low
	GLCM	high	high	moderate	low	moderate
	GLRLM	low	low	high	low	low
	Histogram	lower	low	higher	lower	low

3.2 Image feature representation

According to the study by Tian [9], there are two important image feature representation methods: Image segmentation

[57] and interest point detectors. A key component of picture semantic understanding and a significant research area in computer vision is image segmentation technology. The process of splitting a picture into many sections with

comparable characteristics is known as image segmentation. Image segmentation is the mathematical technique of breaking an image up into separate sections. In terms of algorithm evolution, image segmentation technology can be roughly divided into three categories: graph-based, pixel clustering-based, and deep semantic-based.

First, the graph-based approach applies graph theory theories and techniques, the image is mapped into an undirected graph with weights $G = (V, E)$, where $V = \{v_1, \cdots, v_n\}$ is the set of vertices, and E is the set of edges. Every pixel in the image corresponds to a node in the graph, and every edge joins two adjacent pixels, with the weight of the edge $w(v_i, v_j)$ representing the non-negative similarity in grayscale, color, or texture between adjacent pixels. The principle of segmentation is to ensure that the partitioned subgraphs maintain maximum similarity internally and minimum similarity between subgraphs. Representative methods based on graph theory include NormalizedCut [58], GraphCut [59], and GrabCut [60]. The sketch map of the graph-based method is shown in Figure 4.

Second, the general steps of the machine learning clustering-based method can also be used to picture segmentation problems: 1) Get the rough clustering started. 2) Grouping pixels with comparable characteristics—like color, brightness, and texture—into a single super-pixel through iterative techniques. 3) Until convergence, iterate in order to achieve the final result for picture segmentation. Typical pixel clustering-based techniques include Spectral clustering [61], Meanshift [62], SLIC [63], etc.

Third, although clustering methods can segment images into super pixel blocks with uniform size and appropriate compactness, in practical scenarios, the structure of objects is relatively complex and there are significant internal differences. Only utilizing low-level content information such as pixel color, brightness, and texture is not sufficient to generate good segmentation results. As a result, more high-level information offered by images must be combined to aid in image segmentation, a process known as image semantic segmentation. The picture classification challenge has seen tremendous progress since the advent of deep learning technology, particularly in its capacity to represent advanced

semantics, which has essentially resolved the issue of missing semantic information in conventional image segmentation techniques. According to the study by Shelhamer et al. [64], proposed the FCN method was proposed, which created a framework for teaching whole convolutional networks end-to-end to solve picture segmentation issues, targeting input images of any size to achieve pixel classification. In order to overcome the lack of spatial location information in the final output layer of the convolution network, the rough segmentation result is converted to the dense segmentation result by bi-linear interpolation, up sampling, and combining the feature map of the middle layer output.

DeepLab-v1 [65] extends the FCN framework with fully connected CRFs. To produce the rough segmentation result, bi-linear interpolation is used to sample the FCN output result, and then to increase the model's ability to collect details, every pixel in the outcome is used as a node to build a CRF model. To gather more context information, the Receptive field in this model is enlarged using the Divided Convolution approach, which overcomes the resolution deterioration problem produced by recurrent maximum pooling and down sampling in CNN.

DeepLab-v2 [66] offered the ASPP (atmosphere spatial pyramid pooling) model, it gathers an image's context at several scales by parallelly scanning holes at various sampling rates on a particular input. Simultaneously, it replaces VGG16 with a deep residual network to improve the model's fitting capabilities.

DeepLab-v3 [67] proposed an effective bootstrapping method, which investigated the usage of hollow convolutions and modifies the ASPP module to better capture multi-scale contexts, obtaining excellent results in actual applications. Figure 5 shows the feature extraction performance of different versions of the DeepLab model.

It's important to note that image-based feature representation should be used in conjunction with other FND techniques, including text analysis and source credibility assessment, to achieve more robust results. Additionally, as deep learning models have been successful in image classification tasks, leveraging convolutional neural networks (CNN) can also be explored to automatically learn feature discrimination from images for FND.
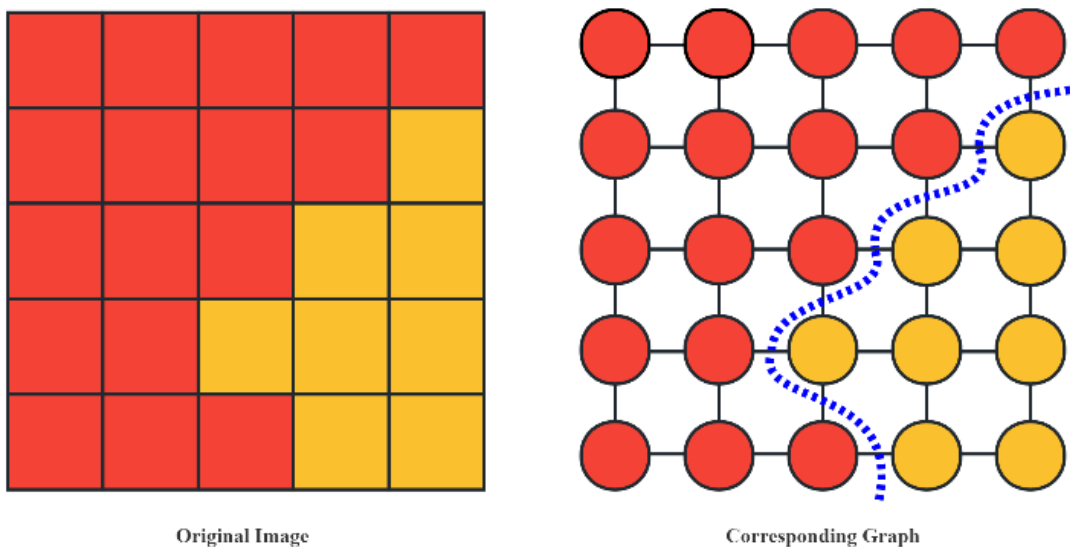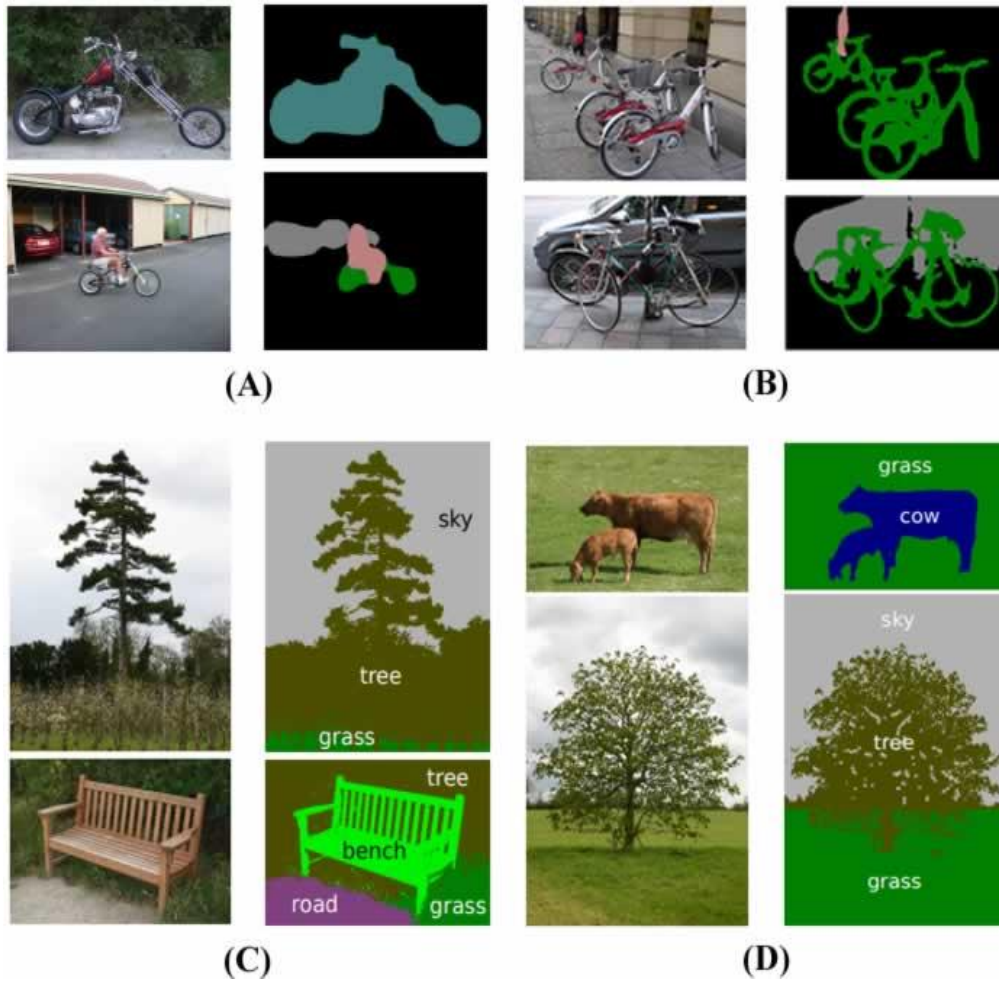

Original Image   Corresponding Graph

Figure 4. Sketch of the graph-based method

Figure 5. Comparison of four image feature representation methods

The subgraph A: The ASPP model can successfully capture objects as well as image context at multiple scales. The subgraph B: The effect of hard pictures bootstrap method in picture segmentation. The subgraph C: The effectiveness of a fully connected CRF framework in extracting classification features. The subgraph D: The training effect of model CRF on dataset MSRC-21.

4. MULTIMODAL-BASED FEATURE REPRESENTATION

Multimodal-based feature representation for FND includes combining data from several modalities—such as text, photos, and videos—to obtain a more thorough comprehension of the topic and raise FND's accuracy. In order to identify fake news, most studies are based on text content, and only a few studies are about the integration of news text, images, and videos. Recently, multi-modal FND has received considerable attention.

4.1 Text-image fusion

Combine textual features extracted from the news article with visual features extracted from associated images to create a joint representation. This can be achieved by employing techniques such as late fusion, early fusion, or hybrid fusion, which combine the features at different stages of the model architecture. Intended to automatically predict whether tweets containing multimedia content are fake or true. As part of the 2015 MediaEval benchmark, Boididou et al. [68] offer a summary of validating multimedia usage tasks. The automatic identification of content tampering and abuse on multimedia platforms is the task at hand. In the latest research, several statistical and visual features based on microblogs were

proposed by Jin et al. [15] to describe these patterns from the visual and statistical perspectives and then combine them with textual features to detect Fake news. Experiments have shown that these novel image features are very effective.

4.2 Visual-text alignment

Align the textual and visual content to identify correspondences between them. For example, the textual description in the news article can be matched with specific objects or scenes depicted in the associated images. Techniques like attention mechanisms or cross-modal embeddings can be utilized to establish meaningful associations between text and images. For the purpose of creating image descriptions, researchers suggested using a deep visual-semantic alignments (DVSA) method to utilize image datasets and their sentence descriptions to understand the modal correspondence between language and visual data [69].

4.3 Fusion of modal features

Extract modality features from text, images, videos, users, and other aspects and then fuse them at a later stage. For example, text-based features can be derived from the news article using techniques like TF-IDF, word embeddings, or

topic modeling, while image-based features can be obtained using the methods mentioned in the previous response. These features can be concatenated or combined using fusion techniques like stacking, averaging, or attention mechanisms. Jin et al. [70] proposes a two-level classification model to utilize the information that tweets on the same topic may have the same credibility. The Figure 6 shows a subcomponent of the proposed two-level classification model.

4.4 Multi-modal deep learning

Because deep neural networks have demonstrated exceptional performance in nonlinear representation learning [71-73], many multi-modal representation approaches employ deep schemes to learn representative features and achieve greater performance in FND [14, 74]. Deep neural networks have proven to be more capable than typical hand-crafted features at learning correct image and phrase representations. Use deep learning architectures like multi-modal convolutional neural networks or multi-modal recurrent neural networks to assess and model textual and visual input simultaneously. These architectures can include independent routes for each modality and train to aggregate multi-modal signals successfully for FND.

Although it exhibits deep neural networks' potential to bridge the "semantic gap" in multimodal data interpretation. However, due to variations in task settings, existing multimodal fusion models have different input feature sets and optimization assumptions than FND tasks. An end-to-end system known as an Event Adversarial Neural Network

(EANN) was proposed in literature [74]. EANN may produce event-invariant characteristics and hence aid in the detection of false information on recently received events. The three main parts of it are the event discriminator, the fake news detector, and the multi-modal feature extractor. (Figure 7).

Zhou et al. [75] proposed a similarity-aware Fake news detection technique (SAFE) that evaluates multimodal (text and visual) information in news items. In order to extract textual and visual aspects for news expression, neural networks are first utilized. Next, examine the relationship in different modalities between the features that were extracted. Finally, SAFE integrates and understands the representation of news text and visual information, as well as the links between them. The proposed method aids in determining the veracity of news based on its text, photographs, or "mismatches" (Figure 8).

4.5 Graph-based representation

Build a graph in which nodes represent distinct modalities (e.g., text, image) and edges reflect the relationships between them. Graph-based models like graph convolutional networks (GCN) or graph attention networks (GAT) can be used to transfer information and learn properties across modalities for enhanced false news identification. A multi-depth graph convolutional network framework (M-GCN) was proposed by Hu et al. [13]. It uses graph embedding to obtain each news node's representation and multi-depth GCN blocks to gather multi-scale information from neighbors, which are then fused via attention mechanisms (Figure 9).
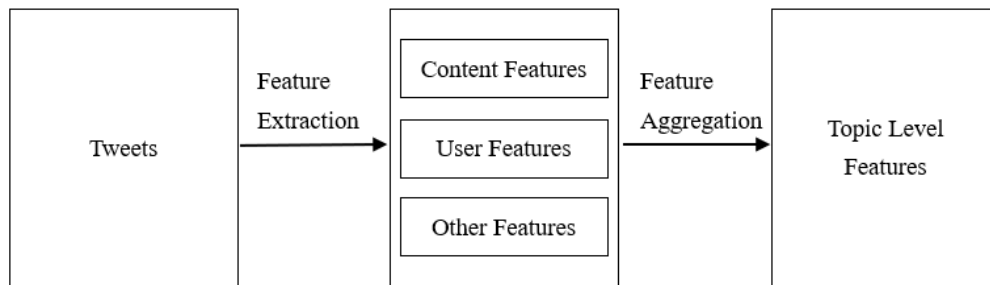

Figure 6. The features of two-level classification model
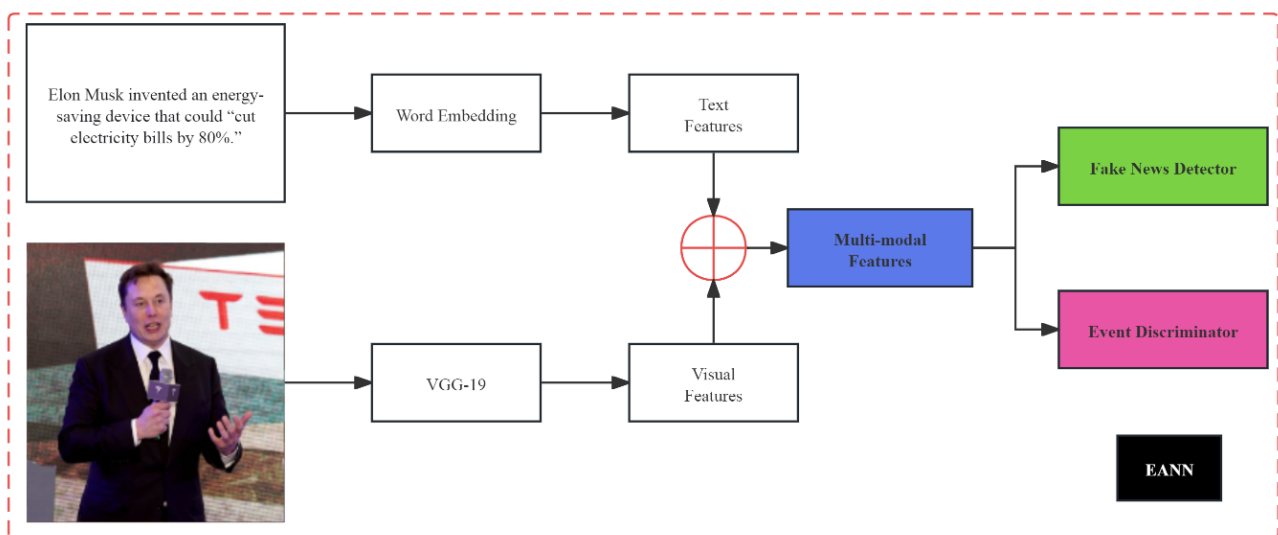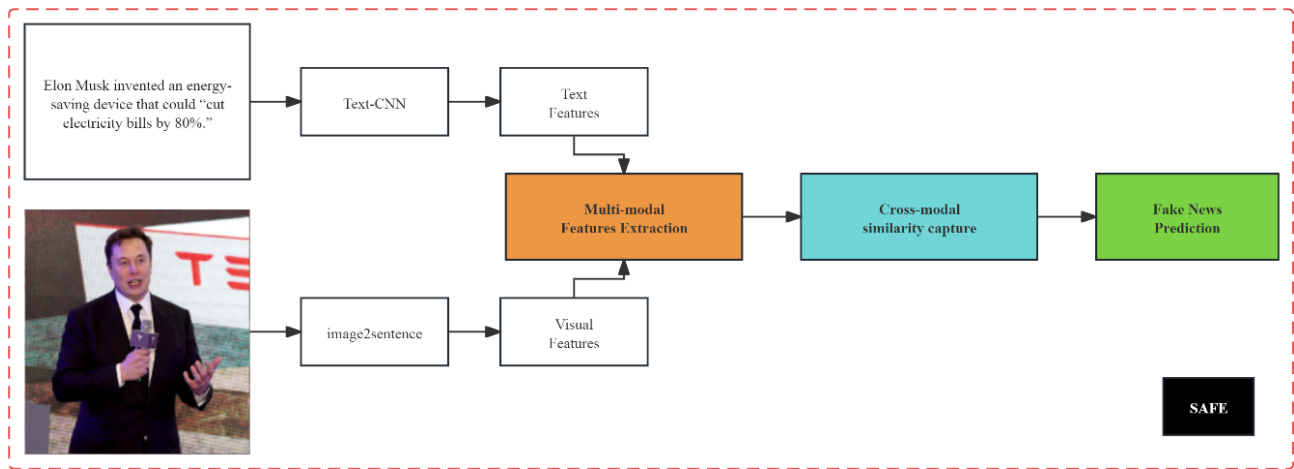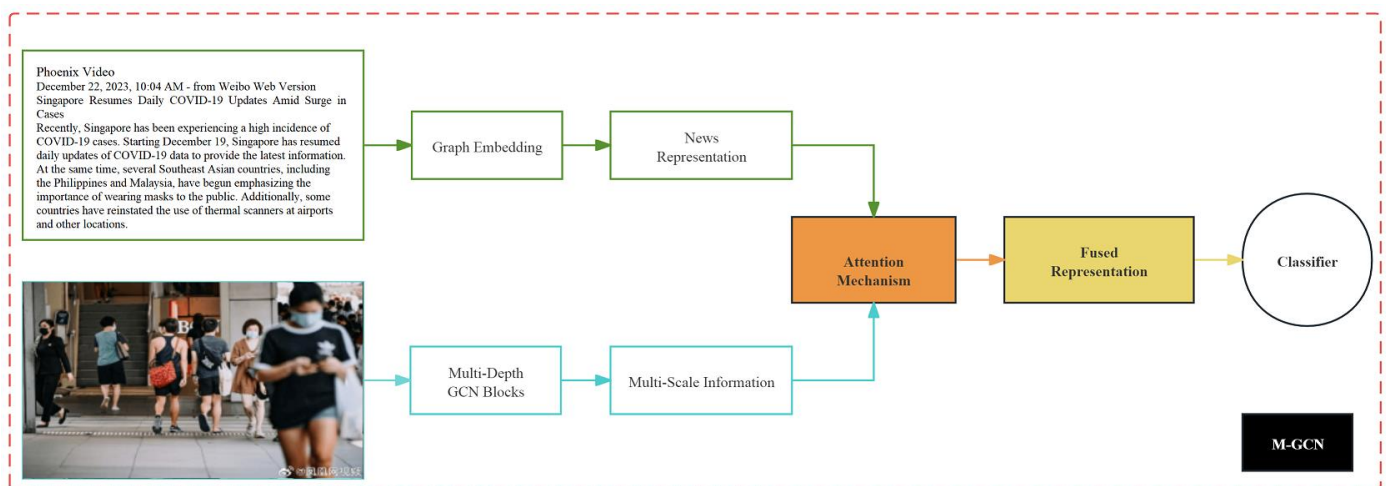

Figure 7. Basic architecture of EANN

Figure 8. Overview of the SAFE framework

Figure 9. A summary of the suggested model M-GCN

It's important to note that both deep neural networks EANN and SAFE, and graph neural network-based M-GCN, are based on multi-modal feature representation. These models require labeled multi-modal datasets for training and may involve additional computational complexity compared to single modal methods. However, by utilizing complementary information from multiple modalities, multimodal feature representation may improve the accuracy and robustness of FND systems.

5. DISCUSSION

Future studies on the topic of feature extraction and representation in FND are recommended. Here are some topics that researchers can further explore:

Techniques for Multi-modal Fusion: More research into improved fusion approaches for merging several senses may improve the efficacy of multimodal-based feature representation. Investigating novel fusion architectures, such as graph-based fusion or attention mechanisms, can improve the ability to effectively capture and combine information from various modalities.

Deep Learning Methodologies: Deep learning has demonstrated potential for a variety of natural language

processing and computer vision problems. Further research into deep learning architectures, such as graph neural networks (GNN), intended expressly for multi-modal FND, may offer more powerful models capable of capturing detailed correlations between text, images, and other modalities.

Explainability and Interpretability: Developing methods to provide explanations or interpretations for the extracted features can enhance the trustworthiness and transparency of FND systems. Techniques like attention mechanisms, knowledge graphs, or rule-based explanations can help understand which features contribute most to the decision-making process.

By addressing these research areas, we can enhance the effectiveness and practical applicability of FND systems.

6. CONCLUSION

In conclusion, feature extraction and representation play a crucial role in FND systems. By capturing relevant information from different modalities, such as text and images, the objective of these strategies is to improve the detection process's robustness and accuracy. For text-based feature representation, approaches like TF-IDF, word embeddings, and deep learning enable the extraction of semantic and

contextual information from news articles. These features help identify patterns, linguistic cues, and stylistic differences between fake and genuine news. In image-based feature representation, analyzing metadata, conducting reverse image searches, and examining visual similarity aid in detecting inconsistencies, manipulations, or reuse of images associated with fake news. Incorporating contextual analysis and social media insights further strengthens the assessment of image credibility. Multimodal-based feature representation takes advantage of integrating information from multiple modalities. By fusing textual and visual features, aligning modalities, or utilizing cross-modal embeddings, these approaches capture nuanced relationships and inconsistencies between different aspects of fake news content. It is important to note that no single feature representation method can guarantee perfect detection of fake news. Training data diversity and quality determine how well a feature representation technique works, the chosen classification algorithm, and the domain-specific characteristics of the fake news problem. Combining multiple feature extraction methods and employing advanced machine learning techniques, such as deep learning or graph-based models, can yield more accurate and reliable FND systems.

ACKNOWLEDGMENT

REFERENCES

[1] Zhou, X., Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5): 109. https://doi.org/10.1145/3395046

[2] Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Transactions on Intelligent Systems and Technology (TIST), 10(3): 21. https://doi.org/10.1145/3305260

[3] De Beer, D., Matthee, M. (2021). Approaches to identify fake news: A systematic literature review. Integrated Science in Digital Age 2020, pp. 13-22. https://doi.org/10.1007/978-3-030-49264-9_2

[4] Guo, B., Ding, Y., Yao, L., Liang, Y., Yu, Z. (2020). The future of false information detection on social media: New perspectives and trends. ACM Computing Surveys (CSUR), 53(4): 68. https://doi.org/10.1145/3393880

[5] Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E. (2015). A review of relational machine learning for knowledge graphs. Proceedings of the IEEE, 104(1): 11-33. https://doi.org/10.1109/JPROC.2015.2483592

[6] Conroy, N.K., Rubin, V.L., Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1): 1-4. https://doi.org/10.1002/pra2.2015.145052010082

[7] Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1): 22-36. https://doi.org/10.1145/3137597.3137600

[8] Yao, L., Mao, C., Luo, Y. (2019). Graph convolutional networks for text classification. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1): 7370-7377. https://doi.org/10.1609/aaai.v33i01.33017370

[9] Tian, D.P. (2013). A review on image feature extraction and representation techniques. International Journal of Multimedia and Ubiquitous Engineering, 8(4): 385-396.

[10] Silva, R.M., Santos, R. L.D., Almeida, T.A., Pardo, T.A.S. (2020). Towards automatically filtering fake news in Portuguese. Expert Systems with Applications, 146: 113199. https://doi.org/10.1016/j.eswa.2020.113199

[11] Wang, H., Wu, H., He, Z., Huang, L., Church, K.W. (2022). Progress in machine translation. Engineering, 18: 143-153. https://doi.org/10.1016/j.eng.2021.03.023

[12] Rao, S., Verma, A.K., Bhatia, T. (2021). A review on social spam detection: Challenges, open issues, and future directions. Expert Systems with Applications, 186: 115742. https://doi.org/10.1016/j.eswa.2021.115742

[13] Hu, G., Ding, Y., Qi, S., Wang, X., Liao, Q. (2019). Multi-depth graph convolutional networks for fake news detection. In Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, pp. 698-710. https://doi.org/10.1007/978-3-030-32233-5_54

[14] 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, California, USA, pp. 795-816. https://doi.org/10.1145/3123266.3123454

[15] Jin, Z., Cao, J., Zhang, Y., Zhou, J., Tian, Q. (2016). Novel visual and statistical image features for microblogs news verification. IEEE Transactions on Multimedia, 19(3): 598-608. https://doi.org/10.1109/TMM.2016.2617078

[16] Qian, S., Hu, J., Fang, Q., Xu, C. (2021). Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17(3): 98. https://doi.org/10.1145/3451215

[17] Lin, H., Yan, J., Qu, M., Ren, X. (2019). Learning dual retrieval module for semi-supervised relation extraction. In The World Wide Web Conference, San Francisco, USA, pp. 1073-1083. https://doi.org/10.1145/3308558.3313573

[18] Dong, X.L., Gabrilovich, E., Murphy, K., Dang, V., Horn, W., Lugaresi, C., Sun, S.H., Zhang, W. (2015). Knowledge-based trust: Estimating the trustworthiness of web sources. arXiv preprint arXiv:1502.03519. https://doi.org/10.48550/arXiv.1502.03519

[19] Zhou, X., Jain, A., Phoha, V.V., Zafarani, R. (2020). Fake news early detection: A theory-driven model. Digital Threats: Research and Practice, 1(2): 12. https://doi.org/10.1145/3377478

[20] Feng, S., Banerjee, R., Choi, Y. (2012). Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju, Republic of Korea, pp. 171-175.

[21] Ma, J., Gao, W., Wong, K.F. (2018). Rumor detection on twitter with tree-structured recursive neural networks. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, pp. 1980-1989. https://doi.org/10.18653/v1/P18-1184

[22] Zhou, X., Zafarani, R. (2019). Network-based fake news detection: A pattern-driven approach. ACM SIGKDD Explorations Newsletter, 21(2): 48-60. https://doi.org/10.1145/3373464.3373473

[23] Varol, O., Ferrara, E., Davis, C., Menczer, F., Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. Proceedings of the International AAAI Conference on Web and Social Media, 11(1): 280-289. https://doi.org/10.1609/icwsm.v11i1.14871

[24] Castillo, C., Mendoza, M., Poblete, B. (2011). Information credibility on twitter. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, pp. 675-684. https://doi.org/10.1145/1963405.1963500

[25] Kwon, S., Cha, M., Jung, K. (2017). Rumor detection over varying time windows. PloS one, 12(1): e0168344. https://doi.org/10.1371/journal.pone.0168344

[26] Nguyen, N.P., Yan, G., Thai, M.T., Eidenbenz, S. (2012). Containment of misinformation spread in online social networks. In Proceedings of the 4th Annual ACM Web Science Conference, Evanston, Illinois, pp. 213-222. https://doi.org/10.1145/2380718.2380746

[27] Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L., Zha, H.Y. (2017). Fake news mitigation via point process based intervention. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, pp. 1097-1106.

[28] Shu, K., Wang, S., Liu, H. (2018). Understanding user profiles on social media for fake news detection. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, pp. 430-435. https://doi.org/10.1109/MIPR.2018.00092

[29] Cao, J., Guo, J., Li, X., Jin, Z., Guo, H., Li, J. (2018). Automatic rumor detection on microblogs: A survey. arXiv preprint arXiv:1807.03505. https://doi.org/10.48550/arXiv.1807.03505

[30] Ma, J., Gao, W., Wei, Z., Lu, Y., Wong, K.F. (2015). Detect rumors using time series of social context information on microblogging websites. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, pp. 1751-1754. https://doi.org/10.1145/2806416.2806607

[31] Liu, Y., Xu, S. (2016). Detecting rumors through modeling information propagation networks in a social media environment. IEEE Transactions on Computational Social Systems, 3(2): 46-62. https://doi.org/10.1109/TCSS.2016.2612980

[32] Della Vedova, M.L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., De Alfaro, L. (2018). Automatic online fake news detection combining content and social signals. In 2018 22nd Conference of Open Innovations Association (FRUCT), Jyvaskyla, Finland, pp. 272-279. https://doi.org/10.23919/FRUCT.2018.8468301

[33] Tacchini, E., Ballarin, G., Della Vedova, M.L., Moret, S., De Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506. https://doi.org/10.48550/arXiv.1704.07506

[34] Volkova, S., Jang, J.Y. (2018). Misleading or falsification: Inferring deceptive strategies and types in online news and social media. In Companion Proceedings of the Web Conference 2018, Lyon, France, pp. 575-583. https://doi.org/10.1145/3184558.3188728

[35] Shu, K., Wang, S., Liu, H. (2019). Beyond news contents: The role of social context for fake news detection. Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, Melbourne, Australia, pp. 312-320. https://doi.org/10.1145/3289600.3290994

[36] Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F., Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks. In Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, USA, pp. 3818-3824.

[37] Yu, F., Liu, Q., Wu, S., Wang, L., Tan, T. (2019). Attention-based convolutional approach for misinformation identification from massive and noisy microblog posts. Computers & Security, 83: 106-121. https://doi.org/10.1016/j.cose.2019.02.003

[38] Guo, H., Cao, J., Zhang, Y., Guo, J., Li, J. (2018). Rumor detection with hierarchical social attention network. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, pp. 943-951. https://doi.org/10.1145/3269206.3271709

[39] Monti, F., Frasca, F., Eynard, D., Mannion, D., Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning. arXiv preprint arXiv:1902.06673. https://doi.org/10.48550/arXiv.1902.06673

[40] Dong, M., Zheng, B., Quoc Viet Hung, N., Su, H., Li, G. (2019). Multiple rumor source detection with graph convolutional networks. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, pp. 569-578. https://doi.org/10.1145/3357384.3357994

[41] Qader, W.A., Ameen, M.M., Ahmed, B.I. (2019). An overview of bag of words; importance, implementation, applications, and challenges. In 2019 International Engineering Conference (IEC), Erbil, Iraq, pp. 200-204. https://doi.org/10.1109/IEC47844.2019.8950616

[42] Jin, Z., Cao, J., Jiang, Y.G., Zhang, Y. (2014). News credibility evaluation on microblog with a hierarchical propagation model. In 2014 IEEE International Conference on Data Mining, Shenzhen, China, pp. 230-239. https://doi.org/10.1109/ICDM.2014.91

[43] Ali, N.M., Jun, S.W., Karis, M.S., Ghazaly, M.M., Aras, M.S.M. (2016). Object classification and recognition using Bag-of-Words (BoW) model. In 2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA), Melaka, Malaysia, pp. 216-220. https://doi.org/10.1109/CSPA.2016.7515834

[44] Hu, N., Sui, L. (2022). Chinese textual entailment recognition model based on contextual feature extraction. In Proceedings of the 2022 2nd International Conference on Control and Intelligent Robotics, pp. 838-841. https://doi.org/10.1145/3548608.3559318

[45] Pennington, J., Socher, R., Manning, C.D. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qata, pp. 1532-1543.

[46] Athiwaratkun, B., Wilson, A.G., Anandkumar, A. (2018). Probabilistic fasttext for multi-sense word embeddings. arXiv preprint arXiv:1806.02901. https://doi.org/10.48550/arXiv.1806.02901

[47] Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., Khudanpur, S. (2010). Recurrent neural network based language model. Interspeech, 2(3): 1045-1048.

[48] Zhang, Y., Liu, Q., Song, L. (2018). Sentence-state LSTM for text representation. arXiv preprint arXiv:1805.02474. https://doi.org/10.48550/arXiv.1805.02474

[49] Humeau-Heurtier, A. (2019). Texture feature extraction methods: A survey. IEEE Access, 7: 8975-9000. https://doi.org/10.1109/ACCESS.2018.2890743

[50] Morris, M.R., Counts, S., Roseway, A., Hoff, A., Schwarz, J. (2012). Tweeting is believing? Understanding microblog credibility perceptions. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, Seattle Washington, USA, pp. 441-450. https://doi.org/10.1145/2145204.2145274

[51] Chow, T.W., Rahman, M.K.M. (2007). A new image classification technique using tree-structured regional features. Neurocomputing, 70(4-6): 1040-1050. https://doi.org/10.1016/j.neucom.2006.01.033

[52] Gupta, M., Zhao, P., Han, J. (2012). Evaluating event credibility on twitter. In Proceedings of the 2012 SIAM International Conference on Data Mining, Anaheim, CA, USA, pp. 153-164. https://doi.org/10.1137/1.9781611972825.14

[53] Wu, K., Yang, S., Zhu, K.Q. (2015). False rumors detection on sina weibo by propagation structures. In 2015 IEEE 31st International Conference on Data Engineering, Seoul, Korea (South), pp. 651-662. https://doi.org/10.1109/ICDE.2015.7113322

[54] Zhang, D., Islam, M. M., Lu, G. (2012). A review on automatic image annotation techniques. Pattern Recognition, 45(1): 346-362. https://doi.org/10.1016/j.patcog.2011.05.013

[55] Kamarasan, M. (2016). Unified technique for colour image retrieval based on orthogonal polynomial model and multiresolution features. International Journal of Image and Data Fusion, 7(3): 243-263. https://doi.org/10.1080/19479832.2015.1046955

[56] Keyvanpour, M.R., Vahidian, S., Mirzakhani, Z. (2021). An analytical review of texture feature extraction approaches. International Journal of Computer Applications in Technology, 65(2): 118-133. https://doi.org/10.1504/IJCAT.2021.114990

[57] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(7): 3523-3542. https://doi.org/10.1109/TPAMI.2021.3059968

[58] Chew, S.E., Cahill, N.D. (2015). Semi-supervised normalized cuts for image segmentation. In 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, pp. 1716-1723. https://doi.org/10.1109/ICCV.2015.200

[59] Yi, F., Moon, I. (2012). Image segmentation: A survey of graph-cut methods. In 2012 International Conference on Systems and Informatics (ICSAI2012), Yantai, China, pp. 1936-1941. https://doi.org/10.1109/ICSAI.2012.6223428

[60] Wang, Z., Lv, Y., Wu, R., Zhang, Y. (2023). Review of GrabCut in image processing. Mathematics, 11(8): 1965. https://doi.org/10.3390/math11081965

[61] Jia, H., Ding, S., Xu, X., Nie, R. (2014). The latest research progress on spectral clustering. Neural Computing and Applications, 24: 1477-1486. https://doi.org/10.1007/s00521-013-1439-2

[62] Li, Y., Xu, J.W., Zhao, J.F., Zhao, Y.D., Li, X. (2015). An improved mean shift segmentation method of high-resolution remote sensing image based on LBP and canny features. Applied Mechanics and Materials, 713: 1589-1592. https://doi.org/10.4028/www.scientific.net/AMM.713-715.1589

[63] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11): 2274-2282. https://doi.org/10.1109/TPAMI.2012.120

[64] Shelhamer, E., Long, J., Darrell, T. (2016). Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4): 640-651. https://doi.org/10.1109/TPAMI.2016.2572683

[65] Krähenbühl, P., Koltun, V. (2011). Efficient inference in fully connected CRFs with Gaussian edge potentials. In Proceedings of the 24th International Conference on Neural Information Processing Systems, Granada, Spain, pp. 109-117.

[66] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4): 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

[67] Chen, L.C., Papandreou, G., Schroff, F., Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587. https://doi.org/10.48550/arXiv.1706.05587

[68] Boididou, C., Andreadou, K., Papadopoulos, S., Dang Nguyen, D.T., Boato, G., Riegler, M., Kompatsiaris, Y. (2015). Verifying multimedia use at mediaeval 2015. In MediaEval 2015, Wurzen, Germany.

[69] Karpathy, A., Li, F.F. (2016). Deep visual-semantic alignments for generating image descriptions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4): 664-676. https://doi.org/10.1109/TPAMI.2016.2598339

[70] Jin, Z., Cao, J., Zhang, Y., Zhang, Y. (2015). MCG-ICT at MediaEval 2015: Verifying multimedia use with a two-level classification model. MediaEval 2015 Workshop, Wurzen, Germany.

[71] Liu, S., Qian, S., Guan, Y., Zhan, J., Ying, L. (2020). Joint-modal distribution-based similarity hashing for large-scale unsupervised deep cross-modal retrieval. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, pp. 1379-1388. https://doi.org/10.1145/3397271.3401086

[72] Zhang, Y., Qian, S., Fang, Q. Xu, C. (2019). Multi-modal knowledge-aware hierarchical attention network for explainable medical question answering. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, pp. 1089-1097. https://doi.org/10.1145/3343031.3351033

[73] Zhao, L., Hu, Q., Wang, W. (2015). Heterogeneous feature selection with multi-modal deep neural networks

and sparse group lasso. IEEE Transactions on Multimedia, 17(11):1936-1948. https://doi.org/10.1109/TMM.2015.2477058

[74] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., Gao, J. (2018). 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, London, UK, pp. 849-857. https://doi.org/10.1145/3219819.3219903

[75] Zhou, X., Wu, J., Zafarani, R. (2020). SAFE: Similarity-aware multi-modal fake news detection. In Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, Singapore, pp. 354-367. https://doi.org/10.1007/978-3-030-47436-2_27