Fake News on Social Media: Current Research and Future Directions

Luciano Caroprese[1(✉)], Carmela Comito[2], and Ester Zumpano[3]

[1] University G. D'Annunzio, Chieti-Pescara, Chieti, Italy
luciano.caroprese@unich.it
[2] Institute for High Performance Computing and Networking (ICAR), Rende, Italy
carmela.comito@icar.cnr.it
[3] DIMES, University of Calabria, Rende, Italy
e.zumpano@dimes.unical.it

Abstract. The escalation of false information related to the massive use of social media has became a challenging problem and great is the effort of the research community in providing effective solutions to detecting it. Fake news are spreading since decades, but with the rise of social media the nature of misinformation has evolved from text based modality to visual modalities, such as images, audio and video. Therefore, the identification of media-rich fake news requires an approach that exploits and effectively combines the information acquired from different multimodal categories. Multimodality is a key approach to improve fake news detection, but effective solutions supporting it are still poorly explored. More specifically, many different works exist that investigate if a text, an image or a video is fake or not, but effective research on a real multimodal setting, 'fusing' the different modalities with their different structure and dimension is still an open problem. The paper is a focused survey concerning a very specific topic that is the use of Deep Learning methods (DL) for multimodal fake news detection on social media.

Keywords: Fake News Detection · Multimodal · Social Media

1 Introduction

The world is highly connected and ideas easily spread in it. Moreover, the easy access to social media platforms has greatly increased so that offering the possibility to produce and share information, ideas and emotions in different forms such as text, video, audio, images. The freedom to share and access content without cost and supervision has surely positive implications, but it has also let to the consequent spread of low quality news and false news, referred to as fake news. Inaccurate and fake information is often intentionally posted online by malicious users in order to manipulate public emotions, influence people thoughts and actions, damage a group or a community, generate confusion and gain profits by misinformation.

C. Comito and D. Talia (Eds.): PerSOM 2022, LNICST 494, pp. 57–64, 2023.
https://doi.org/10.1007/978-3-031-31469-8_4

Fake news are misleading and difficult to catch by humans but also by AI algorithms, as often false information combines both fake and real information. The propagation of false information through social media has negative effects in many different aspects of social life.

The widespread diffusion of fake news on social media is a challenging problem. The research community is devoting great attention to the topic, putting in place important efforts to provide effective fake news detection solutions.

Early works on the fake news detection topic just rely on textual content. Anyhow, even if it is undoubted the necessity of analyzing news content in order to obtain a good indicator for detecting misinformation, it is clear that the only analysis of content is not sufficient. Post and online articles contain not only textual information but also audio, images and video, and misinformation can, therefore, spread through different modalities. Many different sophisticated tools exist to produce fake images or fake videos so that attracting users' attention and thus being shared.

Revealing a fake image involves an accurate analysis of the features related to the image, its associated caption, and the relationship between the image and the caption. Revealing a fake video implies, among others, a detailed analysis of the features related to the images, the sounds and the narrative associated to the video.

Thus far, besides textual information it is important to exploit and correctly combine information acquired from images and audios in order to detect fake news. Multimodality is the real key point to properly address the misinformation detection challenge. However, results obtained by the research community are not yet very effective. More specifically, many different works exist that investigate if a text, an image or a video is fake or not, but effective research on a real multimodal setting, 'fusing' the different modalities with their different structure and dimension, also including the news propagation network, is still an open problem. This paper surveys the recent literature covering various aspects of multimodality, like the news propagation network, text, image, audio, and discusses the fusion strategies proposed in the literature to merge the different modalities for fake news detection.

The proposal investigates and discusses an extensive collection of papers published in the recent years with the purpose of highlighting how deep learning can help fighting fake news. The paper is a focused survey exploring the specific topic of multimodal fake news detection with the lens of deep learning techniques. In fact, even if there are several useful surveys on fake news detection [7,14,15], only a few of them focus on multimodal strategies and even a smaller number of them is restricted to the use of deep leaning methods [1,2] and none of them relies on this topic in the social media domain. Therefore, the final purpose of this survey is to undertake a complete analysis of multimodal fake news detection by considering only the recent advancements in artificial intelligence brought by deep neural networks based solutions.

The paper is structured as follows. Section 2 presents a comprehensive overview of the literature, including models, methodologies, modality, data for

fake news detection. Section 3 provides a discussion about the major challenges and opportunities, and traces future research directions. Section 3.2 concludes the survey.

2 Literature Review

In this section we review the works in the literature discussing models, methods and applications of deep learning techniques for multimodal fake news detection.

In particular, a detailed analysis of the selected papers is provided throughout the section. For each study in the literature, we extracted the most important features like the method implemented, the data type and size used, the evaluation methods adopted, the accuracy for each method, the results achieved.

[16] proposed SpotFake, a multimodal framework for fake news detection, exploiting both the textual and visual features of an article. Specifically, BERT is used to learn text features, while image features are learned from a CNN, VGG19 pre-trained on ImageNet dataset. All the experiments were performed on two publicly available datasets, MediaEval and Weibo A. Authors stated that the proposed model performs better than the current state-of-the-art.

[2] exploited that accepts an article's text and image as inputs. After that, a single vector is created by concatenating the outputs. Experimental validation has been carried out on the Fakeddit dataset, using both unimodal and multimodal solutions. According to experiments, the multimodal technique had an accuracy of 87% and produced the best outcomes.

[11] proposed the EM-FEND framework that is based on the extraction of visual entities to understand the semantics of images. To this purpose the authors considered a variety of data modality: text, OCR text, news-related high level semantics of images e.g., celebrities and landmarks, visual CNN features of the image, the embedded text in images. The different features are then concatenated by considering among others, correlations between text and images and inconsistency. Authors claimed that extensive experiments show their model outperforms the state of the art.

[20] proposed a Similarity-Aware FakE news detection method (SAFE) which exploit multimodal data, more exactly the textual and visual features from news. To this purpose, neural networks are used to extract the textual and visual features, also deriving a similarity among them. The aim of the approach is to classify a news by using either its text or images, or the mismatch between the text and images. Experiments have been performed on large-scale real-world data (PolitiFact, GossipCop), showing the effectiveness of the proposed method.

In the paper of [19] is proposed the Multimodal Consistency Neural Network (MCNN) tool, which is composed of five modules: the textual feature extraction that exploit BERT, the visual semantic feature extraction, the visual tampering feature extraction, the similarity measurement, and the multimodal fusion module. The visual tampering feature extraction focuses on physical levels feature extraction such as malicious image tampering and recompression by usiNg ResNet. The key aspect of the approach is THE similarity measurement module

that evaluates the correlation between the text information and the visual one. The different features are then fused by means of attention mechanisms. The framework has been evaluated over 4 Twitter datasets, MCG-FNeWS, PolitiFact, MediaEval, Yang dataset, showing promising results.

[17] proposed a multimodal fake news detection model exploiting text, comments and images and based on word embedding and convolutional neural network (VGG-19). Precisely, the model is composed of the following components: (1) input embedding layer to obtain word embedding and image embedding; (2) Cross-modal Attention Residual (CARN) layer to reinforce the target modality feature representation by selectively extracting information from another source modality; (3) self-attention residual network layer to capture the interactions between different sequence element pairs and transmit original textual information to MCN; (4) By simultaneously extracting textual feature representation from the original and fused textual data, the Multichannel Convolutional Neural Network (MCN) can reduce the impact of noisy information that may be produced by the crossmodal fusion component. (5) fake news detection module. Experiments have been performed on four real-world datasets: MediaEval, Weibo A, Weibo B. The model exceeds cutting-edge techniques, according to the results, and learns more discriminable feature representations.

[12] makes use of network, textual, and relaying elements like hashtags and URLs and categorizes articles by concatenating the embeddings of the features. Textual features are obtained by using word embedding to represent each word by a low dimensional vector and input this to an LSTM to find the contextual embedding of each tweet. As relaying features are considered five tweet-level features, including hashtag count, URL count, retweet count, mention count and favorite count. For what concerns network features, the framework constructs a network that captures the interactions between users and tweets, creating this way a directed graph of user mentions such that each tweet is connected to a user if their name is mentioned in the tweet text. Using this graph, authors created a one-hot vector of user mentions per tweet. The framework has been evaluated over two datasets, PHEME and Volkova dataset. Results shown that the approach is comparable with state-of-the-art performance.

[13] propose Shared Cross Attention Transformer Encoders (SCATE), a new idea that uses shared layers and cross-modal attention to encode both text and image information using deep convolutional neural networks and transformer-based techniques. Through attentional mechanisms, SCATE integrates the many modalities by focusing on the crucial components of each in relation to the others. A detailed experimental evaluation has been carried out over both Twitter and Weibo datasets like MediaEval, Weibo A, Weibo B.

[8] propose the Attention based Multimodal Factorized Bilinear (AMFB) framework to detect multimodal fake news. The framework has been designed with the intention to reveal the maximum correlation between visual and textual information. This framework has four different sub-modules: i) Attention Based Stacked Bidirectional Long Short Term Memory (ABSBiLSTM) for textual feature representation, ii) Attention Based Multilevel Convolutional Neural

Network-Recurrent Neural Network (ABM-CNN-RNN) for visual feature extraction, iii) multimodal Factorized Bilinear Pooling (MFB) attention mechanism for feature fusion and finally iv) Multi-Layer Perceptron (MLP) for the classification. Experimental results performed on two real-world dataset, MediaEval and Weibo A, shown the effectiveness of the approach.

[4] proposed the TRANSFAKE framework that considers different modalities like news content, comments and images for fake news detection. The textual features are extracted with BERT while for the visual one is used a Faster-RCNN model. TRANSFAKE fuses the different features with a Transformer-based model. It employs multiple tasks, i.e. rumor score prediction and event classification, as intermediate tasks for extracting useful hidden relationships across various modalities. These intermediate tasks promote each other and encourage TRANSFAKE making the right decision. Extensive experiments on three real-life datasets (PolitiFact, GossipCop, Weibo A) demonstrate that TRANSFAKE outperforms state-of-the-art methods.

[9] proposed the GCAN framework, Graph-aware Co-Attention Networks whose main aim is to enable explainable fake news detection on social media. After employing a dual co-attention approach to capture the correlations between user interaction/propagation and tweet's content, Lu et al. concatenate models of user interaction, word representations, and features related to the propagation. To learn the representation of retweet propagation based on user attributes, authors used convolutional and recurrent neural networks. In order to learn the graph-aware representation of user interactions, a graph convolution network is employed to model the potential interactions between users. Both the co-influence of the source tweet and user engagement, as well as the relationship between the source tweet and retweet propagation, can be learned using the dual co-attention mechanism. The binary forecast is produced using the learned embeddings. The framework has been evaluated on a real-world dataset, the Ma dataset [10]. The outcomes shown that the novel approach could be successfully applied for fake news detection by exploiting the propagation network.

[3] propose an interesting multimodal multi-image system in order to perform a binary classification of on line articles by combining textual, visual and semantic information; moreover, differently from other approaches, in the case of an article in which more that an image is present, it extracts and combines features extracted from all of them. BERT is used to obtain textual features, whereas to a obtain visual characteristics a VGG-19 model, an LSTM layer, and a mean pooling layer are employed.

In terms of semantic representation, it refers to the correspondence between text and image that is obtained applying the cosine similarity between the image tags embeddings and the title, this last is a type of information that is rarely considered in fake news detection. Experimentation is performed using the FakeNewsNet collection. More in details, from the GossipCop posts of such collection authors collects 2,745 fake news and 2,714 real news. The proposed multimodal multi-image system outperforms the BERT baseline by 4.19% and SpotFake by 5.39% and achieves an F1-score of 79.55%.

[5] propose a binary classification of fake news called DeepNet. DeepNet is modeled as a deep neural network that performs its task by considering not only the content of the news shared on social media but also exploits the relationship the user exhibits in the social network. The proposal is built considering the tensor factorization method, therefore a tensor is in charge of expressing the social context of news articles as combination of different information related to the news itself, the user, and group with whom the user interacts. DeepNet is structured as follows: it presents one embedding layer, three convolutional layers, one LSTM layer, seven dense layers, moreover it uses ReLU for activation and the softmax function to perform the binary classification. DeepNet is tested on the following datasets: Fakeddit and BuzzFeed (Kaggle, a). This last contains news from articles obtained related to U.S. election in a temporal interval of a week and labeled as either true news or fake news. The binary classification accuracy for the Fakeddit dataset is 86.4 %, and the accuracy for the BuzzFeed dataset is 95.2%.

[6] use four different modalities to perform binary classification of fake news over the Fakeddit dataset: the news' text content, related comments, photos, and any remaining metadata from other modalities. The proposed architecture allows to aggregate these modalities at different levels and considering different data fusion methods. The best result shows an accuracy of 95.5% and has been obtained by separately pre-training each modality, and then training only the fusion and classification layers on top.

[18] propose SERN, the Stance Extraction and Reasoning Network that allows to associate, given a post, its stances representations that are implied in the reply associated to the post itself. Text and images are considered into the proposal and a multimodal representation of these features is performed in order to binary classify fake news. The method works as follows: given a post containing multimodal news, an extractor first construct stances, i.e. post-reply pairs. Then, BERT is used to extract textual features and a pretrained ResNet-152 is used to retrieve visual features. Textual and visual features are therefore concatenated so that obtaining a multimodal feature representation. This last is then the input of a Multi-Layer Perceptron (MLP) that is in charge of performing a binary classification of the post. Experimentation demonstrates the proposal outperforms the state-of-the art baselines on two public datasets: PHEME dataset and a reduced version of the Fakeddit dataset created by the authors. Results show an accuracy of 96.63 % for Fakeddit and of 76.53% on PHEME.

3 Discussion

3.1 Major Challenges in Multimodal Fake News Detection

Fake news affect both online and offline social communities and different proposals exist in the recent literature investigating at different levels and with different strategies the problem. Multimodal approaches for fake news detection

have been proved to be a viable effective approach to address disinformation, however, many are still the challenges that remain to be addressed.

- **Datasets:** Different multimodal datasets exist, but they often are related to two or few modalities such as text and images. These datasets have generally small size, expose content in just one language, and often are imbalanced either in the fake or real news. An additional issue is that in order to cope with different styles and different topics, datasets from heterogeneous platforms should be available. Therefore, urgent is the need of real and complete multimodal datasets containing different modalities such as text, images, video audio, social content, temporal and network propagation features.
- **Finer classification:** Existing fake news detection models are mainly binary classifiers that determine whether a piece of news is false or not. This strategy is often not sufficient and a multi-class classification or even a regression task should be used. The final aim should allow to enable prioritized reasoning and consequent strategies in the presence of fake news detection.
- **Scalability:** Since deep neural networks are complex and costly to build, and as most existing multimodal models use multiple deep neural networks (one per modality), they are not scalable as the number of modalities grows. Furthermore, many existing models require extensive computing resources, including large amounts of memory storage and processing units. As a result, when developing new architectures, the scalability of proposed models should be considered.

3.2 Conclusion

The paper provided a rigorous and in-depth survey on a very specific topic related to the use of deep learning for multimodal fake news detection on social media. The paper analyzed a large number of deep learning approaches and provided, for each work surveyed, an analysis of the rational behind the approach, highlighting some relevant features such as the DL method used, the type of data analyzed, the datasets used, the fusion strategy adopted and the eventual domain-invariant features.

Acknowledgments. This work was partially supported by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU.

References

1. Alam, F., et al.: A survey on multimodal disinformation detection (2021). https://doi.org/10.48550/ARXIV.2103.12541, https://arxiv.org/abs/2103.12541
2. Alonso-Bartolome, S., Segura-Bedmar, I.: Multimodal fake news detection (2021). https://doi.org/10.48550/ARXIV.2112.04831
3. Giachanou, A., Zhang, G., Rosso, P.: Multimodal multi-image fake news detection. In: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), pp. 647–654 (2020). https://doi.org/10.1109/DSAA49011.2020.00091

4. Jing, Q., Yao, D., Fan, X., Wang, B., Tan, H., Bu, X., Bi, J.: TRANSFAKE: multi-task transformer for multimodal enhanced fake news detection. In: 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2021)
5. Kaliyar, R.K., Kumar, P., Kumar, M., Narkhede, M., Namboodiri, S., Mishra, S.: DeepNet: an efficient neural network for fake news detection using news-user engagements. In: 2020 5th International Conference on Computing, Communication and Security (ICCCS), pp. 1–6 (2020). https://doi.org/10.1109/ICCCS49678.2020.9277353
6. Kirchknopf, A., Slijepčević, D., Zeppelzauer, M.: Multimodal detection of information disorder from social media. In: 2021 International Conference on Content-Based Multimedia Indexing (CBMI), pp. 1–4 (2021). https://doi.org/10.1109/CBMI50038.2021.9461898
7. Kumar, S., Shah, N.: False information on web and social media: a survey (2018). https://doi.org/10.48550/ARXIV.1804.08559, https://arxiv.org/abs/1804.08559
8. Kumari, R., Ekbal, A.: AMFB: attention based multimodal factorized bilinear pooling for multimodal fake news detection. Expert Syst. Appl. **184**, 115412 (2021)
9. Lu, Y.J., Li, C.T.: GCAN: graph-aware co-attention networks for explainable fake news detection on social media. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 505–514 (2020)
10. Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F., Cha, M.: Detecting rumors from microblogs with recurrent neural networks, IJCAI 2016, pp. 3818–3824 (2016)
11. Qi, P., et al.: Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues, pp. 1212–1220 (2021)
12. Rezayi, S., Soleymani, S., Arabnia, H.R., Li, S.: Socially aware multimodal deep neural networks for fake news classification. In: 2021 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 253–259 (2021). https://doi.org/10.1109/MIPR51284.2021.00048
13. Sachan, T., Pinnaparaju, N., Gupta, M., Varma, V.: SCATE: shared cross attention transformer encoders for multimodal fake news detection. In: Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2021, pp. 399–406 (2021)
14. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: a data mining perspective. SIGKDD Explor. Newsl. **19**(1), 22–36 (2017)
15. da Silva, F.C.D., Vieira, R., Garcia, A.C.B.: Can machines learn to detect fake news? A survey focused on social media. In: HICSS (2019)
16. Singhal, S., Dhawan, M., Shah, R.R., Kumaraguru, P.: Inter-modality discordance for multimodal fake news detection. In: MMAsia 2021 (2021)
17. Song, C., Ning, N., Zhang, Y., Wu, B.: A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. Inf. Process. Manage. **58**(1), 102437 (2021)
18. Xie, J., Liu, S., Liu, R., Zhang, Y., Zhu, Y.: SERN: stance extraction and reasoning network for fake news detection. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2520–2524 (2021). https://doi.org/10.1109/ICASSP39728.2021.9414787
19. Xue, J., Wang, Y., Tian, Y., Li, Y., Shi, L., Wei, L.: Detecting fake news by exploring the consistency of multimodal data. Inf. Process. Manag. **58**(5), 102610 (2021)
20. Zhou, X., Wu, J., Zafarani, R.: SAFE: similarity-aware multi-modal fake news detection (2020). https://doi.org/10.48550/ARXIV.2003.04981, https://arxiv.org/abs/2003.04981