The final publication is available at

https://doi.org/10.1007/978-3-030-58323-1_3

Additional Information

Multimodal Fake News Detection with Textual, Visual and Semantic Information

Anastasia Giachanou[1][0000−0002−7601−8667], Guobiao Zhang[2,1], Paolo Rosso[1]

[1] Universitat Politècnica de València, Valencia, Spain
[2] Wuhan University, Wuhan, China
{angia9@, prosso@dsic.}upv.es, zgb0537@whu.edu.cn

Abstract. Recent years have seen a rapid growth in the number of fake news that are posted online. Fake news detection is very challenging since they are usually created to contain a mixture of false and real information and images that have been manipulated that confuses the readers. In this paper, we propose a multimodal system with the aim to differentiate between fake and real posts. Our system is based on a neural network and combines textual, visual and semantic information. The textual information is extracted from the content of the post, the visual one from the image that is associated with the post and the semantic refers to the similarity between the image and the text of the post. We conduct our experiments on three standard real world collections and we show the importance of those features on detecting fake news.

Keywords: multimodal fake news detection, visual features, textual features, image-text similarity

1 Introduction

Recent years have seen a rapid growth in the amount of fake news that are published online. Although fake news is not a new phenomenon, the rise of social media has offered an easy platform for their fast propagation. A large amount of invalid claims, rumours and clickbaits are posted every day online with the aim to deceive people and to influence their opinions on different topics. For example, the outcome of Brexit[3] has been into question because of the amount of fake news that were posted before the referendum.

Fake news detection is not a trivial task since the content and the images are manipulated in many different ways which makes the development of an effective system difficult. Several researchers have tried to address the problem of fake news detection. Early works focused on using textual information extracted from the text of the document, such as statistical text features [2] and emotional information [6, 9]. Apart from the content, researchers have also explored the role of users [16, 8] and the credibility of the source where the post is published [14].

---

[3] https://www.theguardian.com/world/2017/nov/14/how-400-russia-run-fake-accounts-posted-bogus-brexit-tweets

Although content information is very important for the detection of fake news, it is not sufficient alone. Online articles and posts usually contain images that provide useful information for a classification system. Some researchers have proposed multimodal approaches for the detection of fake news [22, 12]. The majority of those systems combine textual and visual information to address the problem. However, in addition to the visual information, the similarity between the image and the text is very important since it is possible that in some fake news the image to be contradictory to the content. Although text-image similarity can be an additional useful information, it still remains under-explored.

In this paper we propose a system that uses multimodal information to differentiate between fake and real news. To this end, we combine textual, visual and semantic information. Our main motivation is that information that comes from different sources complement each other in detecting fake news. In addition, some of the fake news contain manipulated images that do not correspond to the post's content. Therefore, we also incorporate semantic information that refers to the similarity between the text and the image. Our experimental results on three different collections show that combining textual, visual and semantic information can lead to an effective fake news detection.

2   Related Work

Early attempts on fake news detection were based on textual information. Castillo et al. [2] explored the effectiveness of various statistical text features, such as count of word and punctuation, whereas Rashkin et al. [15] incorporated various linguistic features extracted with the LIWC dictionary [20] into a Long Short Term Memory (LSTM) network to detect credible posts.

Some researchers explored the role of emotions on the area of fake news. Vosoughi et al. [21] investigated true and false rumours on Twitter and found that false rumours triggered fear, disgust and surprise in their replies, whereas the true rumours triggered joy, sadness, trust and anticipation. Giachanou et al. [9] proposed an LSTM-based neural network that leveraged emotions from text to address credibility detection, whereas Ghanem et al. [6] explored the impact of emotions regarding the detection of the different types of fake news.

Visual information complements the textual one and improve the effectiveness of systems on fake news detection. Wang et al. [22] proposed the Event Adversarial Neural Networks (EANN) model that consists of the textual component represented by word embeddings and the visual that was extracted using the VGG-19 model pre-trained on ImageNet. Khattar et al. [12] proposed the Multimodal Variational Autoencoder (MVAE) model based on bi-directional LSTMs and VGG-19 for the text and image representation respectively. Zlatkova et al. [24] explored the effectiveness of text-image similarity in addition to other visual information but on the task of claim factuality prediction with respect to an image.

Different to the previous work, not only we explore the effectiveness of a wider range of visual features on fake news detection but also of the similarity between
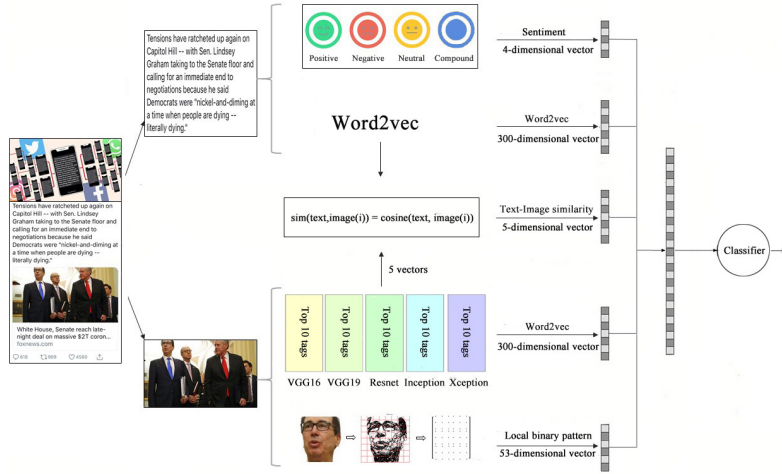
Fig. 1. Architecture of the multimodal fake news detection model.

the image and the text. Our visual features include image tags generated using five different models as well as LBP, whereas the similarity is calculated using the embeddings of the post's text and the image tags.

3   Multimodal Fake News Detection

In this section, we present our multimodal system that aims to differentiate between fake and real news. Our system is based on a neural network and combines the following three different types of information: textual, visual and semantic. The architecture of our system is depicted in Figure 1.

For the textual information, we combine word embeddings and sentiment. To extract the word embeddings, we use the public pre-trained words and phrase vectors GoogleNews-vectors-negative300 that contains 300-dimensional vectors for 3 million words and phrases.

In addition to the word embeddings, we also estimate the sentiment expressed in the posts. Sentiment analysis has in general attracted a lot of research attention and aims to annotate a text regarding to its polarity. Sentimental information has been shown to be useful for fake news detection as well as in other classification tasks [7]. To extract the sentiment score from the documents, we use the Valence Aware Dictionary for sEntiment Reasoning (VADER). VADER [4] is a sophisticated tool that in addition to the terms' sentiment, takes also into account factors such as negation and emoticon usage and predicts the normalized valence of positive or negative sentiment of the given text.

For the visual component, we combine image tags and Local Binary Patterns (LBP) [13]. The visual information can be very useful in case there are different patterns used in fake and real news or there are images that have been

Table 1. Label statistics of the collections.

Collection	Real		Fake	
	training	test	training	test
MediaEval	4,997	1,202	6,742	2,483
PolitiFact	1,067	266	1,712	428
GossipCop	16,495	4,124	16,495	4,124

manipulated. To extract the image tags we use pre-trained CNN-based models. These models are the VGG16, VGG19 [18], Resnet [10], Inception [19], and Xception [3]. The models are pre-trained on the visual dataset ImageNet that contains over 14 million hand-annotated images [5]. We extract the top ten image tags using the pre-trained models, so every image has in total 50 tags. Then for each tag, we use the word2vec embeddings to estimate the 300-dimension vector by averaging the embeddings.

In addition to the image tags, we also explore the effectiveness of LBP. LBP is a very efficient texture operator which labels the pixels of an image by putting a threshold on the neighborhood of each pixel and considers the result as a binary number. LBP has been proved to be very effective in many visual tasks such as face recognition [11]. Similar to previous studies that have used LBP for other tasks such as multimodal sentiment analysis [23], we reduce the original 256-dimensional LBP feature histogram to a 53-dimensional vector.

Finally, the semantic information refers to the text-image similarity. Estimating this similarity is very important since it is possible that fake news contain images that are not relevant to the text. To calculate the similarity we calculate the cosine similarity between the word embeddings of the text and the embeddings of the image tags extracted from the visual feature extraction. This feature provides a 5-dimensional vector, where each vector refers to one image tag model (e.g., VGG16) and is calculated based on the average similarity between the word embeddings of the text and the embeddings of the image tags.

4   Experimental Setup

In this section we describe the collections and the experimental settings used to run our experiments.

4.1   Collections

For our experiments we use MediaEval [1] and FakeNewsNet [17] that, to the best of our knowledge, are the only standard collections that contain tweets in English and that can be used for multimodal fake news detection. Table 1 shows the statistics regarding the labels of the collections.

- *MediaEval*: This collection was released as a part of the *Verifying Multimedia Use at MediaEval* challenge [1]. The aim of the task was to detect fake

Table 2. Neural network parameters.

layers	neurons	learning rate	dropout	activation	optimiser	epochs
4	1000, 500, 300, 100	0.001	0.6	sigmoid	adam	50

multimedia content on social media. The collection consists of tweets and each tweet is provided with textual content, image/video and social context information. After removing the tweets that did not have an image, we managed to have a training set of 11,739 tweets of which 4,997 are real and 6,742 are fake. Our test set contains 3,685 tweets of which 1,202 are real and 2,483 are fake.

- *PolitiFact*: This collection is based on tweets that have been collected with regards to the posts that are published in PolitiFact[4] and is part of the FakeNews-Net collection [17]. PolitiFact is a website that experts in journalism annotate news articles and political claims as fake or real. To create the FakeNewsNet collection Shu et al. used the headlines of those posts as queries to collect relevant tweets. We used the tweet_ids provided as part of the FakeNewsNet and the Twitter API to collect the tweets (text and image) that were available. In total, we managed to collect 2,140 fake and 1,333 real tweets posts.

- *GossipCop*: Similar to PolitiFact, this collection is based on the FakeNews-Net collection [17]. This collection is based on tweets that were collected using the headlines of articles that were posted and annotated in GossipCop[5]. FakeNewsNet contains 5,323 fake and 16,817 real news posted in GossipCop. Due to the imbalance between the classes, we decided to use under-sampling and we randomly selected 5,323 real news posts. We used the tweet_ids and the Twitter API to collect the tweets (text and image) that were still available. In total, we managed to collect 20,619 tweets for each class.

4.2   Experimental Settings

For our experiments on the PolitiFact and the GossipCop collections, we use 20% of our corpus of tweets for test and 80% for training. For the MediaEval, we use the sets as provided in the original collection, that refer to 23% for test and the rest for training. We initialize our embedding layer with the pretrained GoogleNews-vectors-negative300 words and phrase vectors. It is worth to mention that at the beginning of our experiments, we tested also other classifiers including Support Vector Machines and Random Forest. The overall results showed that the neural network performed better for this particular task. Table 2 shows the parameters for the neural network. We have experimented with other hyperparameters, such as different hidden layer number, hidden units, learning rate and dropout. The dropout is applied to each layer. We used the same parameters for all the three different collections.

---

[4] https://www.politifact.com/
[5] https://www.gossipcop.com/

We used keras to build the neural network and the VGG16, VGG19, Resnet, Inception and Xception. Finally, opencv and scikit-image libraries were used to extract the LBP features[6].

5 Results

Table 3 shows the performance results of the experiments on PolitiFact, MediaEval and GossipCop with regards to F1-metric. First, we evaluate the system when one type of information is used. From the results, we observe that in this case the word embeddings achieve the best performance in all the three collections compared to the other types of information. This is expected given that word embeddings are usually a strong indicator in many text classification tasks. In addition, we notice that in MediaEval, text-image similarity manages to achieve a high performance as well. With regards to the visual information, we observe that image tags perform better than LBP on all the collections. This can be due to the fact that image tags represent a larger vector compared to LBP.

Table 3. Performance results of the different combinations of information and the different collections on the fake news detection task. The best result for every collection is emphasized in bold.

	PolitiFact	MediaEval	GossipCop
embeddings	0.911	0.885	0.815
sentiment	0.474	0.352	0.562
tags	0.718	0.615	0.623
LBP	0.474	0.520	0.551
similarity	0.474	0.875	0.538
text-tags	0.924	0.637	0.825
text-LBP	0.909	0.896	0.814
text-tags-similarity	0.920	0.636	0.827
text-LBP-similarity	0.910	**0.908**	0.816
text-tags-LBP-similarity	**0.925**	0.622	**0.829**

Next, we explore the effectiveness of our system when the text (embeddings+sentiment) is combined with the visual information. We observe that on PolitiFact and GossipCop, the combination of text and image tags (*text-tags*) performs better than the combination of text and LBP, whereas in case of MediaEval, the text-LBP achieves a higher performance compared to text-tags. We believe that the poor performance of the image tags on MediaEval has to do with the images of the collection that refer to natural disasters and tend to be more complex than the images on PolitiFact and GossipCop.

Finally, we incorporate the text-image similarity into the system to evaluate its impact when it is combined with the rest of the information. From the results,

---

[6] https://www.pyimagesearch.com/2015/12/07/local-binary-patterns-with-python-opencv/

we observe that incorporating text-image similarity improves the performance. With regards to GossipCop, the *text-tags-LBP-similarity* combination improves the performance by 1.72% and 1.84% compared to *word embeddings* and to *text-LBP* respectively. Similar, in case of PolitiFact, the *text-tags-LBP-similarity* combination achieves a 1.51% increase compared to word embeddings.

Finally, regarding MediaEval we observe that the best performance is achieved by the *text-LBP-similarity* combination, whereas the *text-tags-LBP-similarity* combination is not very effective. When with *text-LBP-similarity* the system combines the text, LBP and text-image similarity it manages to outperform word embeddings by 2.53%.

6   Conclusions and Future Work

In this paper, we proposed a multimodal system to address the problem of fake news detection. The proposed system is based on a neural network and combines textual, visual and semantic information. The textual information was based on the word embeddings and the sentiment expressed in the post, the visual information was based on image tags and LBP, whereas the semantic one referred to the text-image similarity. The experimental results showed that combining textual, visual and text-image similarity information is very useful for the task of fake news detection. Finally, our results showed that different visual information is effective for the different collections.

In future, we plan to investigate more visual features extracted from images such as the color histogram. In addition, we plan to explore the effectiveness of the multimodal information on fake news detection across different languages.

References

1. Boididou, C., Andreadou, K., Papadopoulos, S., Dang-Nguyen, D.T., Boato, G., Riegler, M., Kompatsiaris, Y.: Verifying Multimedia Use at MediaEval 2015. In: MediaEval 2015 Workshop. pp. 235–237 (2015)
2. Castillo, C., Mendoza, M., Poblete, B.: Information Credibility on Twitter. In: WWW '11. pp. 675–684 (2011)
3. Chollet, F.: Xception: Deep learning with Depthwise Separable Convolutions. In: CVPR '17. pp. 1251–1258 (2017)
4. Davidson, T., Warmsley, D., Macy, M., Weber, I.: Automated Hate Speech Detection and the Problem of Offensive Language. In: ICWSM '17 (2017)

5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A Large-scale Hierarchical Image Database. In: CVPR '09. pp. 248–255 (2009)
6. Ghanem, B., Rosso, P., Rangel, F.: An Emotional Analysis of False Information in Social Media and News Articles. ACM Transactions on Internet Technology (TOIT). 20(2), 1–18 (2020)
7. Giachanou, A., Gonzalo, J., Mele, I., Crestani, F.: Sentiment propagation for predicting reputation polarity. In: ECIR '17. pp. 226–238 (2017)
8. Giachanou, A., Ríssola, E., Ghanem, B., Crestani, F., Rosso, P.: The Role of Personality and Linguistic Patterns in Discriminating between Fake News Spreaders. In: NLDB '20 (2020)
9. Giachanou, A., Rosso, P., Crestani, F.: Leveraging Emotional Signals for Credibility Detection. In: SIGIR '19. pp. 877–880 (2019)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: CVPR '16. pp. 770–778 (2016)
11. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Local Binary Patterns and its Application to Facial Image Analysis: A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C 41(6), 765–781 (2011)
12. Khattar, D., Goud, J.S., Gupta, M., Varma, V.: MVAE: Multimodal Variational Autoencoder for Fake News Detection. In: WWW '19. pp. 2915–2921 (2019)
13. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
14. Popat, K., Mukherjee, S., Yates, A., Weikum, G.: DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning. In: EMNLP '18. pp. 22–32 (2018)
15. Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y.: Truth of varying shades: Analyzing Language in Fake News and Political Fact-Checking. In: EMNLP '17. pp. 2931–2937 (2017)
16. Shu, K., Wang, S., Liu, H.: Understanding User Profiles on Social Media for Fake News Detection. In: MIPR '18. pp. 430–435 (2018)
17. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: FakeNewsNet: A Data Repository with News Content, Social Context and Spatialtemporal Information for Studying Fake News on Social Media. arXiv:1809.01286 (2018)
18. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-scale Image Recognition. arXiv:1409.1556 (2014)
19. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: CVPR '16. pp. 2818–2826 (2016)
20. Tausczik, Y.R., Pennebaker, J.W.: The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology 29(1), 24–54 (2010)
21. Vosoughi, S., Roy, D., Aral, S.: The Spread of True and False News Online. Science 359(6380), 1146–1151 (2018)
22. Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., Gao, J.: EANN: Event Adversarial Neural Networks for Multi-modal Fake News Detection. In: KDD'18. pp. 849–857 (2018)
23. Zhao, Z., Zhu, H., Xue, Z., Liu, Z., Tian, J., Chua, M.C.H., Liu, M.: An Image-Text Consistency Driven Multimodal Sentiment Analysis Approach for Social Media. Information Processing & Management 56(6) (2019)
24. Zlatkova, D., Nakov, P., Koychev, I.: Fact-checking meets Fauxtography: Verifying Claims about Images. In: EMNLP-IJCNLP '19. pp. 2099–2108 (2019)