

Fake News Detection with Integration of Embedded Text Cues and Image Features

Deepak Mangal
Dept. of CEA, GLA University,
Mathura, India
deepak.mangal@gla.ac.in

Dilip Kumar Sharma
Dept. of CEA, GLA University,
Mathura, India
dilip.sharma@gla.ac.in

Abstract— A novel approach using Convolution neural Network (CNN) and Long short-term memory (LSTM) has been proposed to find the reliability of the news. In this research, image visual feature with embedded text feature and headline texts have been considered to find the comprehensive results. First the semantic information from the images have been captured as text (news tag line) and this tag has been compared to the original headline text. Individually image and text both are insufficient to find the semantic knowledge of publish news. So, the cosine similarity index (CSI) has been used to predict the reliability of the news. The threshold of CSI has been constrained greater than 0.62 for the news real. A repository has been created named as "imaged fake news". In this repository 1000 images have been considered with the headline texts, where 367 news were fake and 633 news were real. The accuracy of the proposed method is 91.07%. The result implies that the novel methodology is better than the state-of-the-art method.

Keywords— LSTM, Word2Vec, Cosine similarity, OCR, SoftMax

I. INTRODUCTION

Fake News is a yellow journalism spread over the world to gain individual profit, political benefit, acquire information of group of age group and increase number of viewers for media. As we know users do not have time to read or listen the complete information, they require summarized information. So, such people those are creating fake news using this habit of people and post such one-line text which is not related with real news or actual text. They have created ques to user sometimes positive or negative to take benefits. For example, AAJ TAK published a news on their channel that Indian army has been attacked on Pakistan post and destroy two post of Pakistan pimpal and minal, but one post is in India. This was creating an issue that how does the user identify such news as fake or real because an authorized media publishes it. Today, social media and social networking sites are mostly used to search, connect and communicate with family and friends. So, people are using this to post their happening, events or many information related or unrelated to it. Now, it is very useful to know anyone or to get latest news in all fields. Number of case studies have been done on the use of social media sites, the latest one published in [1]. It shows how use of social media has been increased. Now journalists are using social media as an important and usual tool to publish anything or latest news [2,3,5]. Social media is a fastest medium to transfer information whether bad or good in crisis [4]. However, the spreading power of social media generates problems many times. For that an individual, a group of people, a company or a political party pay cost because

unauthenticated news is being spreading [2,28]. Such situations have been born a word fake news. So, a fabricated news or article which has been intentionally written to influence people or group of people for any type of benefit. Fake news is coined by US president Donald trump in 2016 in US election whenever, various information which are spread not in favour of him.



Fig. 1. Fake News Story

To understand fake news, there is an example at the time of US election is that one of the supporters Bob of Donald trump posted a Facebook post of Donald Trump with pop Francis Shocks with text that pops endorses Donald trump for president. This news is highly spread by WTOE 5News without finding the fact. So, Donald Trump got gain in their campaign in US election 2016 with given fig 1[6]. However, facts were different, but people unknowingly forwarded this in a long chain. Why Social Media is important today? Fig- 2 shows importance of social media in today's life. People are easily communicating and propagate information by using various mode of social media without worrying about the impact of it. On social media people just see information available in the form of images, text or both and forward it. Some people fabricate content just to gain popularity among friends, group of people and others. These are the medium shown in fig-2 that reveal the percentage of people are using different mediums to get or post information.

The key features which give us motivation are as follows:

1. No specific definition is available for fake or real news. So, no one can understand the reality of text.

One specific way is required in research in future for fake news detection.

2. To understand available state of art methods and their working mechanism, a comparison is needed.
3. Identify various features for detection of social news. So future research work can be carried forward for smart mechanism.

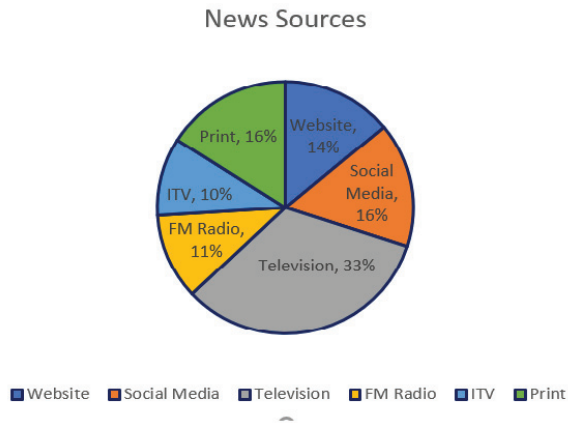


Fig. 2. News Sources in India

Our key contribution in this area is extracted text and objects available in the image and then check similarity between them to find the fraud in this information.

The next sections of this paper has been well organized in such a way: In Section II, it illustrates Literature survey in terms of fake news detection on image and fake news detection on text, section III shows the exact detail of proposed architecture, while section IV shows experimental results and comparison study with state-of-the-art approaches. Also, some further modifications that can be proposed and done so as to increase the range of the currently built work has been further discussed in section V future work and conclusion.

II. LITERATURE SURVEY

From Last 25 years in India or all over world number of cases have been detected as fraud disclosure. There are various types of fraud such as unrealistic information with images, fake tweet, wrong scientific research data and many more. So, we can consider unrealistic information with images as sub field of fake news. Many Researchers published research papers on this issue with various solutions: -

Before detecting fake news and unrealistic information, the researcher used credibility of user, information and follower's credibility and all [7, 8,9,10].

In the current scenario, fake text with addition of positive and negative sentiments is easily forwarded and spread over twitter, facebook and such social media, so this one is a big challenge for researchers to find the solution. Hence, we can have an emerging field to find an automatic solution to detect fake news. Few researchers have worked on finding solution for automatic fake news detection [11]. Images and text are the main focus in fake news detection in recent research. Many algorithms have been used to extract image features, which are used to train machine as a model to detect such images. The traditional approach are text based approaches which finds the pattern of text and matches it with already

saved pattern of fake news. These approaches come under the linguistic approach [12].

A. Fake News Detection on Images

First, The most recent work on images in the elf of fake news detection is given as follows: The state of art methods in detection of fake news on images are very few. In 2017, Jin et al.[13] found that each real image or fake image having different features and pattern. They proposed that visual feature is very important in any news and its verification. They used visual content and found hidden patterns in image distribution. They also used statistical features of images and formalized mathematical statics of the same. Apart from this they combined both features for fake news authentication. In 2016, Elkasrawi et. al [14] worked with reverse images search of google with combination of K-means clustering algorithm to find images across time to find changes in images in comparison with web published image. They applied edge detection, changes in color and event date or publishing date to find changes and use publishing date as first view to detect event is fake or real. In 2015, Pasquini et. al. [15] had been published research work in which they extracted visual and semantic features of images applied on news articles on similar subject to verify the consistency. In 2014, Jin, Cao, Jiang, and Zhang proposed three-layer architecture [14]. Layers were content based with name event layer, sub event and message layer. They created credibility-based network to calculate news credibility which was very helpful to detect fake images as fake news. Apart from all in 2011, Rashed, Renzel, and Klamma [16], developed a multimedia system to detect fake multimedia. They have used various modules such as trust management, feature extraction for identification. So, there was many research areas open for researchers such as detect tempered images, multimedia forensics, recovery algorithm etc. From 2014 to 2017 many researchers worked for the same [17, 18, 19].

B. Fake News Detection on Text

There are number of researchers who worked for fake text detection as fake news. In 2018 Alrubaian, Al-Qurishi, Hassan, and Alamri [20] found an issue of text credibility at twitter. To solve this, they worked and proposed an automated classification system with four building blocks. In this they considered reputation-based technique with classifier engine. They also used user experiences and rank based algorithm. In 2016 Qin, Wurzer, Lavrenko, and Tang [10] proposed a Novelty and pseudo feedback-based features to early detection of rumors. They have used different sentiment of eight types, hash-tags, URLs and punctuation emotions. In 2015 Vosoughi [21] used linguistic features of text for fake news detection. He has also considered temporal propagation and twitter dataset for result analysis. In 2016 Bouazizi and Ohtsuki [22] solve sarcasm issue by using pattern-based approach. In last two years, many researchers proposed many content-based machines learning classification algorithms to detect news on social media is fake or real.

III. PROPOSED ARCHITECTURE

Mostly research work from year 2014 to year 2019 focused on Image features or text features to identify fake news. However, individually these feature extractions hve not been efficient to achieve high accuracy fake news detection algo- rithm. So, in this proposed research work, a fusion of text features and image text features are used to get

an efficient algorithm to evaluate information is real or fake. The proposed model consists four processing modules. First module extract text from image and extract text features. Second module extract the features of image by visual geometry group (VGG) net [23]. VGG is a deep convolution neural network (CNN) model. Third module LSTM which takes input from first and second module generate more semantic text with visual image feature. The fourth module use the output of third one and match with headline of news by using cosine similarity and generate cosine index of both word vector.

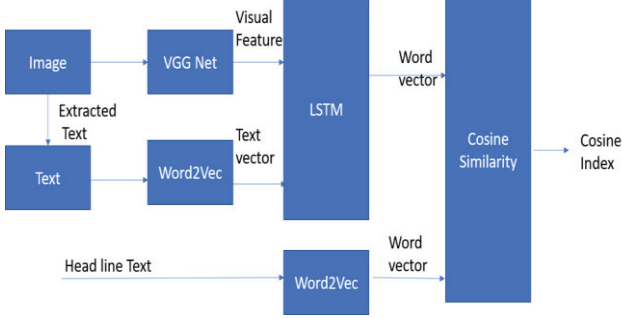


Fig. 3. Proposed Architecture

A. Visual Image feature

To get visual feature we have used a pre-trained model VGG Net [23] in proposed model. ImageNet [24] dataset has been used to trained this model. It is 3 x 3 convolution network. For down sampling Max polling method is used. Finally, total of 4096 nodes are used in two fully connected layers. Then 1000 classes have been classified by SoftMax classifier. So, for each input image say NIM, it computed a vector say v of 4096- dimensional features after finetuning of VGG Net.

B. Extracted Text Features

To extract text features an approach proposed in [25]. It has been represented in vector form by using bag of words algorithm Word2Vec [26]. This algorithm finds words which are very much close to spotted word in a vector space. Each which is extracted or spotted has been represented with 300-dimension vector say F_i^V . So, total feature extracted is 300 x Number of words. There is no direct relation with image of the selected word. However, different weight is assigned, to find correlation between image and word. To calculate the weight for each word. Equation 1 shows the calculation for weight for each word.

$$W_i = \exp (F_v^T \times UM \times F_i^V) \quad (1)$$

where exponent normalize all words in SoftMax and UM is bilinear matrix. And by sum polling method weights and word are combined as by using equation (2)

$$W_{\text{sum}} = \sum (W_i \times F_i^V) \quad (2)$$

C. LSTM based information extraction

Input gate, output gate and forget gate are three types of gates in LSTM, these gates are useful to control memory cell. The LSTM is using visual cues with extracted text vector as input to find more appropriated information including semantic data. For this LSTM taking two input, first input is x and second is m_{t-1} . x is visual feature input and output of first input of next is m_{t-1} . In our approach is using extracted text from image as a new input say E vector for LSTM to generate more precise or meaningful

information. The mathematical organization of our approach for memory cell is given as follows:

$$ig_t = \sigma(W_{igx}X_t + W_{igm}m_{t-1} + W_{ige}E) \quad (3)$$

$$fg_t = \sigma(W_{fgx}X_t + W_{fgm}m_{t-1} + W_{fge}E) \quad (4)$$

$$rg_t = \sigma(W_{ogx}X_t + W_{ogm}m_{t-1} + W_{oge}E) \quad (5)$$

$$Cg_t = fg_t \otimes Cg_{t-1} + ig_t \otimes h(W_{oge}X_t + W_{gm}m_t + W_{cge}E) \quad (6)$$

$$m_t = O_{gt} \otimes Cg_t \quad (7)$$

$$P_{t+1} = \text{SoftMax}(m_t) \quad (8)$$

Where E is holding extracted text feature.

To generate non-linearity sigmoid activation function $\Omega(\cdot)$, $h(\cdot)$ is hyperbolic tangent activation function, W are trained parameter matrix. At the end SoftMax function is used to calculate probability of each word on the basis of previous input. So, LSTM model has been trained by using visual feature and extracted text feature vector and generate word vector as output B . After all processing, the output categorizes a real news or fake news on the basis of cosine similarity. The cosine similarity unit is using vector B and vector A as shown in fig.3. This module calculates the cosine similarity index as follows:

The cosine similarity between two vectors $A = \{a_1, a_2, \dots, a_N\}$ and $B = \{b_1, b_2, \dots, b_N\}$ is defined as:

$$\cos(A, B) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^N A_i B_i}{\sqrt{\left(\sum_{i=1}^N A_i^2\right) \left(\sum_{i=1}^N B_i^2\right)}}$$

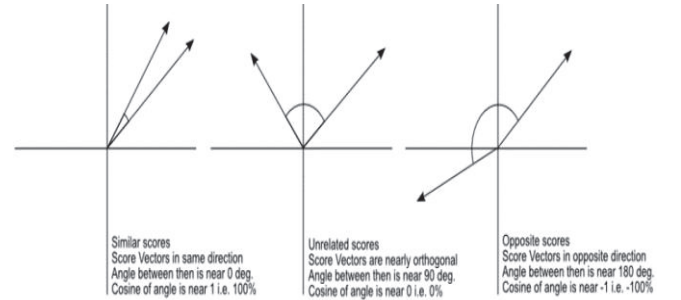


Fig. 4. Cosine Similarity Graph

IV. DISCUSSION AND EXPERIMENTAL RESULTS

In this section the proposed model will be compared with traditional state of art methods. There are limited data set available which are having fake images with text. In 2016, PHEME data set introduced by Zubiaga et. al [11]. The data set consist of total 5808 records in which 1972 records are fake and rest are real. Afterwards in 2018, Boididou[27] collected 193 real and 218 fake news. So, no specific data set are available for training and testing. We collected 1000 images from Google, Kaggle and onion for fake or real images with text. There is various news which are published in local area, national area or international area. We will test our proposed algorithm performance and comparison on it.

A. Evaluation Method

To evaluate the performance of proposed algorithm, the collected data set is divided in two parts in the ratio of 8:2. In which 80% data will be used to train the machine and 20% for testing. The data set consist URL of images, head line text and binary class column having values 0 or 1, as fake or real respectively.

Algorithm 1: Proposed Algorithm

Step1: Extract image from given URL from dataset
Step2: Extract text from image by using OCR
Step3: Collect visual feature of image by using VGG.
Step4: Collect text feature by using Word2Vec.
Step5: The LSTM using both extracted feature of step 3 and step 4 and generate a semantic word vector.
Step6: The headline from data set passed as input to Word2Vec model and generate another word vector.
Step7: Both vectors are input to soft cosine similarity model to generate cosine similarity.

IF cosine similarity is greater 0.62
THEN
Predict news → Real News
ELSE
Predict news → Fake News

B. Result

The proposed model has been tested on rest data of dataset and accuracy has been measured by calculating as follows:

To know how accurate your model is the precision has been calculated as shown in equation (9)

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (9)$$

And to know how many cases have been true positive cases captured by your model, recall calculate as illustrated in equation(10)

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (10)$$

To measure model accuracy F1 score is used as shown in equation (11)

$$F - measure = 2 * \left(\frac{precision * recall}{precision + recall} \right) \quad (11)$$

TABLE I. PERFORMANCE ANALYSIS OF CLASSES

Class	Accuracy	Precision	Recall	F-measure
I-Class	80%	58%	40%	50%
T-Class	83%	61%	72%	66%
TI-Class	91%	90.9%	91.3%	91.09%

I-Class: Similarity at the visual features level. Image visual feature vector collected from VGG Net and compare with headline text vector. So, 80% accuracy has been achieved.

T-Class: Similarity at the Extracted features level. Extracted Text feature vector collected from Word2Vec model and compare with headline text vector. So, 83% accuracy has been achieved.

TI-Class: Similarity at the Extracted features level combines with Image features level. Extracted Text feature vector collected and Visual Image feature vector both are combined by using LSTM and compare with headline text vector. This is proposed method in this paper. And proposed method achieved 91% accuracy.

The observational outcome in fig. 5 and Table 1 says that in all aspect our proposed methodology depicted better result in our data set. It shows that alone visual features, extracted text feature are not sufficient to check the news on social media is Fake or Real.

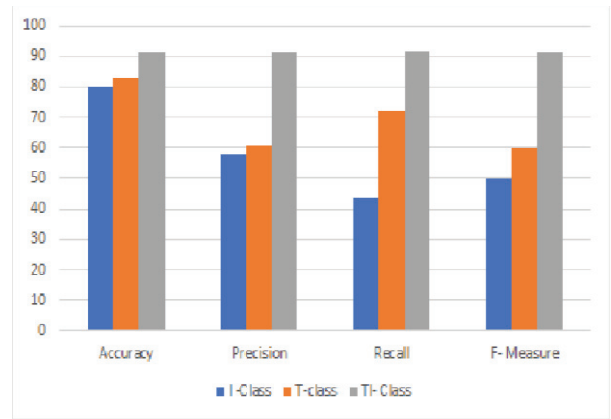


Fig. 5. Comparison Graph

V. CONCLUSION AND FUTURE WORK

In this research paper, It states that we achieve accuracy of 91.07% in detection of fake news. First time visual features and embedded text features on image has been involved to get meaningful information by using proposed model. The involvement of deep learning approach and long short term memory generate meaningful word vector. Then headline pass to Word2Vec model which provides another word vector. the cosine similarity has been used to get CSI between these two vectors for more accurate results .In future, other similarity model may be applied. Text available in image if it is semantically Irrelevant has not been considered.

REFERENCES

- [1] Ott, Myle, et al. "Finding deceptive opinion spam by any stretch of the imagination." Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1. Association for Computational Linguistics, 2011.
- [2] Xie, Sihong, et al. "Review spam detection via temporal pattern discovery." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012.
- [3] Shu, Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD Explorations Newsletter 19.1 22-36, 2017.
- [4] Wagner, Elliott O., and Michael Franke. "BRIAN SKYRMS Signals: Evolution, Learning, and Information." 883-887. 2013.
- [5] Garimella, Kiran, et al. "Balancing opposing views to reduce controversy." arXiv preprint arXiv:1611.00172 (2016):
- [6] Akoglu, Leman, Mary McGlohon, and Christos Faloutsos. "Oddball: Spotting anomalies in weighted graphs." Pacific-Asia Conference on

Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, 2010.

- [7] AlRubaiian, Majed, et al. "A multistage credibility analysis model for microblogs." 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2015.
- [8] Kang, Byungkyu, John O'Donovan, and Tobias Höllerer. "Modelling topic specific credibility on twitter." Proceedings of the 2012 ACM international conference on Intelligent User Interfaces. 2012.
- [9] Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. "Information credibility on twitter." Proceedings of the 20th international conference on World wide web. 2011.
- [10] Qin, Yumeng, et al. "Spotting rumors via novelty detection." arXiv preprint arXiv:1611.06322 (2016).
- [11] Zubiaga, Arkaitz, et al. "Detection and resolution of rumours in social media: A survey." ACM Computing Surveys (CSUR) 51.2 (2018): 1-36.
- [12] Tarmizi, Fatin Amanina Ahmad, et al. "Online news veracity assessment using emotional weight." Proceedings of the 2019 2nd International Conference on Information Science and Systems. 2019.
- [13] Jin, Zhiwei, et al. "News credibility evaluation on microblog with a hierarchical propagation model." 2014 IEEE International Conference on Data Mining. IEEE, 2014.
- [14] Elkasrawi, Sarah, et al. "What you see is what you get? Automatic Image Verification for Online News Content." 2016 12th IAPR Workshop on Document Analysis Systems (DAS). IEEE, 2016.
- [15] Pasquini, Cecilia, et al. "Towards the verification of image integrity in online news." 2015 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). IEEE, 2015.
- [16] Rashed, Khaled AN, Dominik Renzel, and Ralf Klammer. "Trust-aware media quality profiles in fake multimedia detection." 2011 Workshop on Multimedia on the Web. IEEE, 2011.
- [17] Zampoglou, M., Papadopoulos, S. and Kompatsiaris, Y., "Large-scale evaluation of splicing localization algorithms for web images", *Multimed Tools Appl* 76, 4801–4834. <https://doi.org/10.1007/s11042-016-3795-2>, 2017.
- [18] Ferreira, Anselmo, et al. "Behavior knowledge space-based fusion for copy-move forgery detection." *IEEE Transactions on Image Processing* 25.10 (2016): 4729-4742.
- [19] Dadkhah, Sajjad, et al. "An effective SVD-based image tampering detection and self-recovery using active watermarking." *Signal Processing: Image Communication* 29.10 (2014): 1197-1210.
- [20] Alrubaiian, Majed, et al. "A credibility analysis system for assessing information on twitter." *IEEE Transactions on Dependable and Secure Computing* 15.4: 661-674, 2016.
- [21] Vosoughi, Soroush, "Automatic detection and verification of rumors on Twitter", Diss. Massachusetts Institute of Technology, 2015.
- [22] M. Bouazizi and T. Otsuki Ohtsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," in *IEEE Access*, vol. 4, pp. 5477-5488, 2016.
- [23] Simonyan K, Zisserman A, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv: 1409.1556, 2014
- [24] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S et al, "Imagenet large scale visual recognition challenge", *Int J Comput Vis* 115(3):211–252, 2015
- [25] Gupta N, Jalal AS (2018) A robust model for salient text detection in natural scene images using MSER feature detector and Grabcut. *Multimedia Tools Appl* 78(8):10821–10835
- [26] Mikolov T, Chen K, Corrado G, Dean J, "Efficient estimation of word representations in vector space", arXiv preprint arXiv:1301.3781, 2013
- [27] Boididou, Christina, et al. "Detection and visualization of misleading content on Twitter." *International Journal of Multimedia Information Retrieval* 7.1: 71-86. 2018.
- [28] S. Sharma and D. K. Sharma, "Fake News Detection: A long way to go", In *Proceedings of IEEE Sponsored 2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, GLA University, Mathura, UP, India. Nov 21-22, 2019