NewsBag: A Multimodal Benchmark Dataset for Fake News Detection

Sarthak Jindal,[1] **Raghav Sood,**[1] **Richa Singh,**[2] **Mayank Vatsa,**[2] **Tanmoy Chakraborty**[1]
[1]IIIT-Delhi, India, [2]IIT Jodhpur, India
{sarthak15169, raghav16259, tanmoy}@iiitd.ac.in, {richa, mvatsa}@iitj.ac.in

Abstract

The spread of fake news poses a critical problem in today's world, where most individuals consume information from online platforms. Fake news detection is an arduous task, marred by the lack of a robust ground truth database for training classification models. Fake News articles manipulate multimedia content (text and images) to disseminate false information. Existing fake news datasets are either small in size or predominantly contain unimodal data. We propose two novel benchmark multimodal datasets, consisting of text and images, to enhance the quality of fake news detection. The first dataset includes manually collected real and fake news data from multiple online sources. In the second dataset, we study the effect of data augmentation by using a Bag of Words approach to increase the quantity of fake news data. Our datasets are significantly larger in size in comparison to the existing datasets. We conducted extensive experiments by training state of the art unimodal and multimodal fake news detection algorithms on our dataset and comparing it with the results on existing datasets, showing the effectiveness of our proposed datasets. The experimental results show that data augmentation to increase the quantity of fake news does not hamper the accuracy of fake news detection. The results also conclude that the utilization of multimodal data for fake news detection substantially outperforms the unimodal algorithms.

Introduction

News consumption by people has increasingly grown over the years. The primary reason is the ease of accessibility of news. With the help of social networking sites such as Facebook and Twitter, people not only share existing news, but also "create news" and then share it (Chen, Conroy, and Rubin 2015). Moreover, the era of content driven websites is becoming increasingly visible. For example, there are many existing popular news websites, and many more smaller websites come up every day. These websites contain news articles written by mostly paid content writers. Even though it is good that news is so easily accessible, these days, both with respect to consumption and production, it poses a serious challenge in the form of fake news (Jin et al. 2017). Fake news is any news written with the purpose of deception or providing misinformation to the reader (Ruchansky, Seo, and Liu 2017). There can be many ill intentions behind creating and spreading fake news. These include defamation of personalities (Wang 2017), creating bias to change real-world event outcomes (Farajtabar et al. 2017), and decreasing trust in particular sections of social media.

Figure 1: Example of defamatory news (a) Elon Musk Gives Saudi Investors Presentation On New Autonomous Beheading Machine For Adulterers. Example of a bias inducing news (b) Trump says "America Has Not Been Stronger Or More United Since I First Opened My Eyes And Created The Universe".

Fake news is often written to defame certain famous personalities by spreading false information about them. These famous personalities could be politicians and movie stars. The LIAR (Wang 2017) dataset which contains labeled short real-world statements collected from Politifact, a fact checking website, contains examples of such defamatory news with reference to a diverse range of political personalities. It becomes important to stop the spread of such defamation so as to protect the reputation of these famous personalities. For example, the fake news shown in Figure 1(a) is an example of a fake news written to defame a certain personality.

Fake news can create a bias in the minds of people which in turn affects the outcome of important events like presidential elections, etc. This motivates one to stop the spread

of fake news to isolate event outcomes from bias. For example, the fake news shown in Figure 1(b) is an example of a fake news written to create a bias in the minds of people during the event of US Presidential Elections. Social media is the most easily accessible platform for news exchange. Fake news spread must hence be put to an end especially on social media.

Background and Previous Work

The fake news problem is quite old. Researchers have come up with various solutions belonging to different domains. The earliest solutions were purely using natural language processing for fake news detection (Castillo, Mendoza, and Poblete 2011) (Kwon et al. 2013). The lie detector (Mihalcea and Strapparava 2009) was one of the earlier major attempts in deception detection which used purely natural language processing techniques. Natural language processing based fake news detection depended on text data only and it's features (Gupta et al. 2014). For example, handcrafted rules could be written to point out explicit features such as large number of third person pronouns which were mostly common in fake news articles (Shu et al. 2018). However, explicit handcrafted features extracted from text data depends upon news content and the event context in which the news is generated (Ruchansky, Seo, and Liu 2017). Therefore, it is difficult to come up with discriminative textual features to get good detection results on fake news on new events. The next steps taken by the research community incorporated information from social networks in them. The social context of a news includes user interactions such as hastags, comments, reactions, retweets etc. (Shu et al. 2017).However, the shortcoming of social context based fake news detection lies in the noisy nature of these social features.

It is only recently that researchers have started using images along with text for the fake news detection task. Multimodal deep learning has previously been successfully applied to related tasks like visual question answering (Antol et al. 2015) and image captioning (Vinyals et al. 2015). With respect to fake news detection, TI-CNN (Yang et al.) (Text-Image Convolutional Neural Network) is a very recent work in which the authors scraped fake and real news generated during the US 2016 Presidential elections. The authors used parallel convolutional neural networks to find reduced representations for both image and text in a data point. Then, they merged the representations to find a combined feature representation for image and text which is used for classification. Rumour detection on microblogs (Jin et al. 2017) is another form of fake news detection. In this paper, the authors work with the Weibo (Jin et al. 2017) and Twitter (Boididou et al. 2015) datasets, obtained from Chinese authoritative news agencies and Twitter respectively. The authors proposed a multimodal fusion mechanism in which the image features are fused with the join features of text and social context produced by an LSTM(Long-Short Term Memory) network. They showed that images fused with neural attention from the outputs of the LSTM, the att-RNN mechanism performs well on multimodal rumour detection task.

Inspite of having so many existing techniques for fake news detection, the results produced are still not upto the mark. The problem of detecting fake news is hard primarily because of two reasons: (i) the scarcity of labeled data (Wang 2017) and (ii) deceptive writing style (Shu et al. 2017).

Contributions

In this research paper, we go beyond the existing work by presenting a large-scale dataset to help improve the performance of current fake news detection algorithms. We initially scrape The Wall Street Journal and The Onion to create our training dataset, termed as NewsBag, which has 215,000 news articles. The proposed dataset contain both news text and images. Since this training dataset is imbalanced, we use a data augmentation algorithm to create a bigger and approximately balanced training dataset, NewsBag++, containing around 589,000 news articles with both text and image data. To show a real world evaluation of our models, we scrape our testing set- NewsBag Test from completely different news websites. We use state-of-the-art text and image classification models in our experiments and also use the recently published Multimodal Variational AutoEncoder(MVAE)(Khattar et al. 2019) and FAKEDETECTOR(Zhang et al. 2018) for multimodal fake news detection. This is done by parallely training the networks with image and text input. However, we infer from our experiments that even very deep networks cannot generalize well to unseen and differently written news in the testing dataset. This shows the hardness of the fake news detection problem as fake news can vary with respect to writing style, news content, source etc. However, if seen from a relative point of view we show that it's a good idea to use multiple modalities of data from fake news detection. Our best multimodal model is a MVAE which beats our best single modality classification model, RCNN (Lai et al. 2015), by a significant margin. This provides inspiration for further work in the field of multimodal fake news detection.

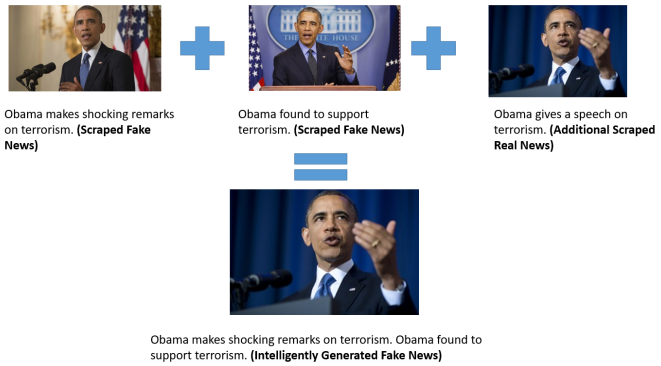


Figure 2: Example of fake news generation using Intelligent Data Augmentation Algorithm for generating fake news.

Dataset

The NewsBag dataset comprises of 200,000 real news and 15,000 fake news. The real news has been scraped from

Table 1: Comparison of existing datasets for Fake News Detection

Dataset	No. of real news articles	No. of fake news articles	Visual Content	Social Context	Public Availability
BuzzFeedNews	826	901	No	No	Yes
BuzzFace	1,656	607	No	Yes	Yes
LIAR	6,400	6,400	No	No	Yes
Twitter	6,026	7,898	Yes	Yes	Yes
Weibo	4,779	4,749	Yes	No	Yes
FacebookHoax	6,577	8,923	No	Yes	Yes
TI-CNN	10,000	10,000	Yes	No	Yes
FakeNewsNet	18,000	6,000	Yes	Yes	Yes
NewsBag Test	**11,000**	**18,000**	Yes	No	Yes
NewsBag	**200,000**	**15,000**	Yes	No	Yes
NewsBag++	**200,000**	**389,000**	Yes	No	Yes

the Wall Street Journal. The fake news have been scraped from the Onion which is a company that publishes articles on international, national, and local news. The Onion publishes satirical articles on both real and fictional events. We have manually asked several test subjects to go through the data and verify that the 15,000 articles picked by us are only those which cover fake events. However, since the NewsBag dataset is highly imbalanced we create NewsBag++, an augmented training dataset. The NewsBag++ dataset contains 200,000 real news and 389,000 fake news. The data augmentation algorithm used for generating new fake news given a ground truth set of fake and real news is described in the following section. Apart from NewsBag and NewsBag++, we create a NewsBag Test dataset for testing while training models on either of NewsBag or NewsBag++. The NewsBag Test dataset contains 11,000 real news articles scraped from The Real News and 18,000 fake news articles scraped from The Poke. We have used completely different sources of news for the NewsBag Test dataset so that we can understand how well a model trained on NewsBag or NewsBag++ generalises to unseen and differently written news.

Data Augmentation for Generating Fake News

The simplest idea to generate fake news would be to combine any two random news from the existing 15,000 fake news scraped from websites. However, this poses two problems. One, the two combined pieces of fake news may be totally irrelevant and hence make no sense together. This is not good for our research because we want fake news to be the way it is actually written by people. The second drawback is that the number of fake news images would be limited, since we would only be picking from the existing set of 15,000 images. This is not good with respect to training a robust model. So, we decide to come up with an intelligent data augmentation algorithm for generating fake news. Figure 2 shows an example of the same.

First of all, we scrape 170,000 additional real news from the Wall Street Journal besides the 200,000 real news we already have. Then, we get a bag-of-words representation for each news in this additional set of 170,000 real news. We get a bag-of-words representation for each news in our 15,000 fake news set as well . These bag-of-words representations are found after removing stop words from the respective news whose representation it is. Then, we do the following for mutliple iterations: Pick a random news from the 15,000 fake news set. Find all the fake news whose bag-of-words representation has an intersection above a threshold with the particular fake news picked from the 15,000 fake news set. Generate a new fake news by combining the text of each of these fake news with the fake news picked at first. Also, mark the pair so that it is never used for generation again. Find the real news from the additional 1,70,000 real news set whose bag-of-words representation has the largest intersection with the bag-of-words representation of this particular generated piece of fake news. Simply attach the image from this real news to the generated fake news.

Our augmentation algorithm generates fake news which is very similar to actual fake news written by people because of two main reasons. Firstly, the two fake news combined to generate a new one are very relevant to each other since their bag-of-words representation have the largest intersection with each other. This makes the generated news sound coherent and not completely senseless. And the second reason is that we attach an image from the real news whose bag-of-words representation has the largest in common with the bag-of-words representation of the generated fake news. This is actually the most intuitive way to write fake news since fake news writers must look for relevant real news images which can be attached to the fake news text they have written.

Nomenclature

We make our dataset publicly available in three different formats. The simplest is the Dataset Folder format which is commonly used by deep learning libraries like PyTorch. The image data is organised as two folders- fake and real. Each folder contains all images of that particular class. The same is the organisation for the text data.

FastText is a format used for data in text classification tasks. In the FastText Format, the three datasets- NewsBag Test, NewsBag and NewsBag++ exist as a text file each. Within the text file, each line represents a sample ie. two samples are separated by a newline character. Also, each line starts with __label__ followed by the target label for the sam-

Table 2: Analysis of the dataset

Textual Features/Dataset	NewsBag Test		NewsBag		NewsBag++	
	Fake	Real	Fake	Real	Fake	Real
Vocabulary Size (words)	29,571	25,286	40,897	124,243	109,006	124,243
Avg. number of characters per news	148	219	223	216	446	216
Avg. number of words per news	27	37	38	36	81	36
Avg. number of stopwords per news	9	11	13	11	27	11
Avg. number of punctuations per news	1	1	2	2	7	2

ple. This prefix allows models to retrieve the class for a given sample during training or testing. The actual sample follows the label prefix after separation by a space followed by a comma followed by a space. This format is very well suited for text classification as it requires very little extra memory to store every sample's label.

Google Colaboratory is an openly available tool for researchers which provides a Tesla K80 GPU backend. However, reading data folders from google drive with a lot of files or subfolders at the top level gives IO Error on Colab. Also, memory is limited on colab which calls for data compression. So, we provide our datasets- NewsBag.zip, NewsBag Test.zip and NewsBag++.zip in a format which we call the Google Colab format. We downsample our images to 28 by 28 so as to keep only the most useful visual information and limit memory requirement. We organise the text and images into numbered sub-directories with 500 text and image files each, respectively. The last subdirectory in the text and image folders may however have lesser than 500 files each. We prefix the label followed by a space to each filename to retrieve the target label during training or testing. Finally, we perform our experiments on Colab using this particular format and face no input/output errors.

Comparison with Other Existing Datasets

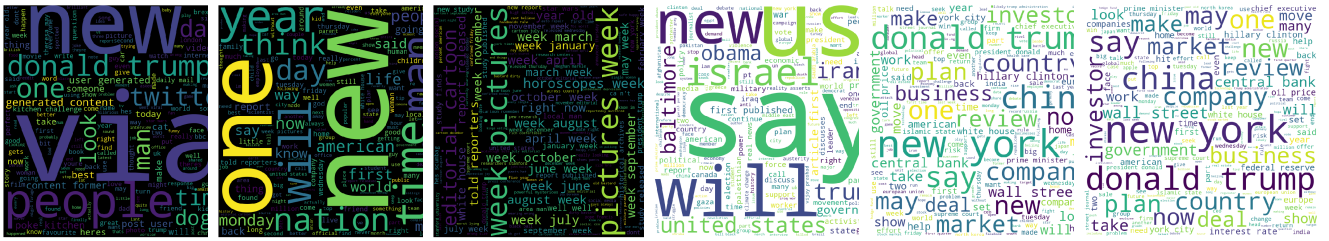
One of the main strengths of our database is it's size. Our NewsBag++ database stands at 589,000 data points with two classes- real and fake. This is an order of magnitude bigger than already existing fake news datasets. However, at the same time, the main weakness of our dataset is that it does not have any social context. By social context, we mean that there is no information on who is spreading the news on social media, what are the trends in the sharing of this news, what are the reactions and comments of users etc. This provides scope for further improvement where we can dig out the social context of news by searching similar posts on social media. Some of the already existing datasets for fake news detection are discussed below. Table 1 compares all the datasets.

- The FakeNews Net dataset (Shu et al. 2018) which is a recent work in fake news detection contains about 24,000 data points only. The main strength of this dataset is the presence of social context, for example, user reactions and comments etc.
- Similarly, the TI-CNN (Text-Image Convolutional Neural Network) (Yang et al.) also has only 20,000 data points. The fake news revolve around the 2016 US Presidential elections.
- BuzzFeedNews is a small dataset collected from Facebook. It has been annotated by BuzzFeed journalists. BuzzFace (Santia and Williams) is simply an extension of BuzzFeedNews. Both the datasets have content based on the US 2016 elections just like the TI-CNN dataset.
- The FacebookHoax (Tacchini et al. 2017) as the name suggests has hoaxes and non-hoaxes collected from few of Facebook's scientific and conspiracy pages respectively.
- The LIAR dataset(Wang 2017) is different from others because it is more fine-grained. Fake news are divided into fine classes- pants on fire, false and barely-true while real news are divided into fine classes- halftrue, mostly true, and true. This dataset contains real world short statements made by a diverse range of political speakers. It is collected from fact checking website Politifact, which uses manual annotation for the fine-grained classes.
- The Weibo dataset(Jin et al. 2017) is collected from Chinese authoritative news sources over a period of 4 years from 2012 to 2016. The annotation has been done by examining suspicious posts reported by credible users of the official rumour debunking system of Weibo.
- The Twitter dataset(Boididou et al. 2015) is collected from Twitter, originally for detecting fake content on Twitter. The data not only has both text and images but also additional social context information from twitter users.

We will observe that when we train a model on our augmented dataset vs training a model on these existing datasets, the accuracy achieved by the model trained on the augmented dataset is not hampered in comparison to the other datasets.

Analysis of the Dataset

In this section, we present key statistics about the NewsBag Test, NewsBag and NewsBag++ datasets. Each of these statistics can be used as handcrafted features that may be input to a machine learning model. However, one of the main reasons why fake news detection is hard is that these handcrafted features are not very discriminative. In other words, they are almost equal for both the classes- real and fake. This encourages the use of deep learning models which can learn hidden or latent features in the data. The significance, variation and lack of dicriminative property of the features for the



(a) (b) (c) (d) (e) (f)

Figure 3: Fake news word cloud representations for NewsBag Test, NewsBag and NewsBag++ are shown in black from (a)-(c) respectively. Real news word cloud representations for NewsBag Test, NewsBag and NewsBag++ shown in white from (d)-(f) respectively.

different datasets is described below. Table 2 summarises the analysis of the dataset.

Vocabulary is the set of unique tokens in a text dataset, also called the set of types. It is a very important indicator of the diversity of a dataset. But, in the case of both of our approximately balanced datasets- NewsBag Test and NewsBag++, the vocabulary size is almost equal for fake and real classes. This shows that fake and real news are equally diverse. For the NewsBag dataset, the vocabulary size is higher for the real news samples simply because of their larger number compared to fake news samples in the dataset.

We analyze the news content of the three datasets with respect to both the classes separately. Word Cloud representations reflect the frequency of words in a particular dataset. We make two interesting observations on the word cloud representations shown in Figure 3. Firstly, the word clouds of real news for all of the three datasets reflect important real word entities. For example, we can easily observe the highly frequent words Israel, New York and China in the word cloud representations of the real news of NewsBag Test, NewsBag and NewsBag++ respectively. On the other hand, fake news contain mostly words not related to important entities. For example, we see words such as new, one, week and pictures in the word clouds of the fake news in the NewsBag Test, NewsBag and NewsBag++ dataset respectively. This disparity between the word clouds of fake and real news emphasizes the fact that fake news do not have much real world content to speak about. They simply try to create news by using attractive words, for example, 'New' rule on tax payment etc. Another observation to make is that the NewsBag Test has noticeably different word cloud representations than our training datasets, NewsBag and NewsBag++. This is because we have scraped the NewsBag Test dataset from different websites (TheRealNews and ThePoke) while the training datasets contain news from Wall Street Journal and The Onion. We use different sources of news for the testing and training datasets so that we can observe how well our models generalize to unseen data points.

The length of the fake or real news in terms of the number of characters or words is once again dependent on the source of news. There is no fixed pattern. As we see, the NewsBag Test dataset has longer real news as compared to fake news, in contrast to the NewsBag dataset which has longer fake news. This is another reason why fake news detection is non-trivial. The length of the news (characters or words) is an example of a handcrafted feature which follows opposite pattern in our training (NewsBag or NewsBag++) datasets and testing(NewsBag Test) dataset. Features like these can actually fool the model. This is reflected in the baseline results we present in the experiments section, where we see the testing accuracy of some models to be less than random.

Stopwords and punctuations are least informative in a text. Just like the length of the news, we see that these features follow different patterns in real and fake classes, across different sources of news. Hence, these handcrafted features are also not suitable for classification.

Experiments

We train both single modality and multimodal models on our dataset. We show the training and testing accuracies for both NewsBag and NewsBag++. The test set is the same while training with either NewsBag or NewsBag++. All our experiments have been carried out on Google Colaboratory, an open source python notebook environment with a Tesla K80 GPU backend. The accuracies for each dataset and model are summarized in Table 3.

Single Modality - Text

We use the FastText data format for training our text classification models. The training setting for each model is described in detail below.

- FastText(Joulin et al.) is one of the simplest text classification methods known for it's efficiency. We use GloVe(Pennington, Socher, and Manning 2014) word embeddings which have 300 dimensional vectors, 2.2M types in vocabulary and 840B tokens. We train the model for 30 epochs with a learning rate of 0.5 and batch size 128.
- TextCNN(Kim) had improved the state-of-the-art in sentiment analysis and question classification. Here, we train the model for fake news classification. We use the same embeddings as in the case of FastText but we train the model with a slower learning rate of 0.3 and a smaller batch size of 64. We use convolutional kernels of sizes 3x3, 4x4 and 5x5. The model is trained for 15 epochs.

Table 3: Experiments carried out using NewsBag and NewsBag++ training sets

Model/Dataset	NewsBag		NewsBag++	
	Training Accuracy	Testing Accuracy	Training Accuracy	Testing Accuracy
fastText	0.95	0.46	0.98	0.52
TextCNN	0.96	0.51	0.98	0.46
TextRNN	**0.99**	0.51	0.99	0.43
RCNN	0.98	0.56	0.99	0.47
Seq2Seq (Attention)	0.98	0.48	**0.99**	0.45
Transformer	0.96	0.48	0.98	0.39
Deep Boltzmann Machine	0.81	0.32	0.60	0.31
Image ResNet	0.93	0.52	0.72	0.49
Image SqueezeNet	0.93	0.54	0.71	0.53
Image DenseNet	0.92	0.49	0.72	0.50
Multimodal Variational AutoEncoder	0.96	**0.71**	0.76	**0.62**
FAKEDETECTOR	0.96	**0.70**	0.74	**0.61**

- We use a bi-directional LSTM network for classification. The architecture is kept simple with only 2 hidden layers consisting of 32 units each. We use a maximum sentence length of 20 to enable faster training.
- Recurrent Convolutional Neural networks (Lai et al. 2015) capture context to learn better representations for words, thereby eliminating the requirement for handcrafted features. We train a simple RCNN with 1 hidden layer of size 64 using a dropout of 0.2. We keep the batch size as 128 and train the model for 15 epochs with a learning rate of 0.5.
- Neural Machine Translation (Bahdanau, Cho, and Bengio) is a recent approach for end-to-end machine translation. It uses an encoder-decoder architecture with a soft attention mechanism to align words better to each other. In order to use the sequence to sequence model(with attention), we use only the representation of a news article generated by the encoder for classification. The encoder architecture is a simple bi-directional LSTM with 1 hidden layer of size 32.
- Transformers (Vaswani et al.) eliminate the need for any RNN or CNN by using stacks of self-attention and position-wise feedforward neural networks for the machine translation task. The methodology to use transformer for fake news detection is the same as the sequence to sequence model. We use the self-attention and position-wise feedforward network in the encoder to get the data representation for classification.

Single Modality - Image

We use the Google Colaboratory data format for our image classification models. We show our results for very deep convolutional neural networks which have performed extremely well on image classification tasks.

- Restricted Boltzmann Machines (RBM's) have been successfully applied to the movie recommendation task earlier (Salakhutdinov, Mnih, and Hinton 2007). We present results from a Deep Boltzmann Machine based multimodal deep learning model (Srivastava and Salakhutdinov 2014). We first get a suitable representation for the image by minimizing the reconstruction loss and then classify on this reduced representation. The image pathway of the model consists of a stack of Gaussian RBMs with 3857 visible units, followed by 2 layers of 1024 hidden units. We train our model for 5 epochs with a batch size of 128.
- We use a ResNet(He et al. 2016) with 18 layers for classifying fake news on the basis of image only. ResNets have shown increase in accuracy and decrease in complexity in image classification tasks by learning residual functions with respect to the input layers. The final fully connected layer of the ResNet with 1000 dimensional output is replaced by another dense layer with 2 outputs to get the desired classification. We use a batch size of 128 and a learning rate of 0.01 decayed by a factor of 0.1 every 3 epochs. The model is trained for 7 epochs.
- We use SqueezeNet(Iandola et al.) as another model which takes less memory than AlexNet or ResNet, without sacrificing on accuracy. The training settings are kept same as ResNet. We see that when trained on our NewsBag dataset, SqueezeNet perform as good as ResNet. We use a bigger batch size of 256 for SqueezeNet.
- DenseNets (Huang et al. 2017) take the idea of feature propagation and feature reuse to the extreme which is the reason why they achieve good classification accuracy. For a given layer, the feature maps from all the previous layers are used as input, leading to a total K*(K + 1)/2 direct connections, where K is the number of convolutional layers. DenseNets are effective in reducing the vanishing gradients problem.

Mutliple Modality - Image and Text

The training of multimodal models is performed similar to the image only models. We have used the state of the art architectures used for fake news detection i.e MVAE (Khat-

tar et al. 2019) and FAKEDETECTOR (Zhang et al. 2018). In our experiments, we observe that multimodal algorithms significantly outperform the unimodal algorithms.

Inferences

The results summarised in the table indicate the hardness of the fake news detection problem. We observe that the training accuracies are very high for the NewsBag training set, irrespective of the modality of the model. In the case of NewsBag++, however, training accuracy for image modality only models and multimodal models is very low. On the other hand, text modality only models yield very high training accuracy even on NewsBag++. This leads us to infer that it is specifically the image modality of the data which is fooling the models in case of NewsBag++ training set. The reason behind this is that our custom intelligent data augmentation algorithm for NewsBag++ generation tries to generate realistic fake news by using images from the additional 170,000 real news, scraped from Wall Street Journal specifically for this purpose. This inference empirically verifies exactly how fake news writers can fool detection models by attaching real news images to their fake text content.

We also observe that irrespective of the training dataset and model used, the testing accuracies are very low. This is because when the source of news varies, as in the NewsBag Test and NewsBag/NewsBag++ datasets, even the very basic latent feature learnt by the model from the training set vary in the testing set, across classes. Even data augmentation using already available ground truth data, as in NewsBag++, does not seem to solve the problem of effective generalisation to unseen data. However, even on such unpredictable dataset, our best model- MVAE achieves about 20% improvement over random accuracy. We also observe that the augmented NewsBag++ dataset does not significantly hamper the performance when compared to NewsBag only, providing a scope to try further augmentation techniques resulting in improved results for fake news detection.

Conclusion

In this paper, we present NewsBag, a benchmark dataset for training and testing models for fake news detection. It is not only an order of magnitude larger than previously available datasets but also contains visual content for every data point. Our work brings forward the complexities involved in fake news detection due to unpredictable news content, the event context in which the news originated, author writing style, and news article sources. We show baseline results of state-of-the-art text classification and image classification models for single modality fake news detection. We also show results from multimodal fake news detection techniques. We indicate the hardness of the fake news detection problem by showing poor generalization capabilities of both single modality and multimodal approaches. We further support our claim about the non-trivial nature of the problem by presenting an augmentation algorithm which when used for fake news generation can fool very deep architectures, as empirically verified in our experiments. We infer that none of the single modality models achieve good improvement over a random coin toss. Multimodal approaches, however, achieve better performance by combining learning's from text and image modalities. Future work can be done in the direction of expanding the modality set for fake news detection datasets, for example, using social context, text, images, audio, and video for fake news detection. Also techniques like data augmentation which were applied by us can be tried to increase the size of training dataset and further improve the results of fake news detection.

References

Antol, S.; Agrawal, A.; Lu, J.; Mitchell, M.; Batra, D.; Zitnick, C. L.; and Parikh, D. 2015. Vqa: Visual question answering. In *Proceedings of the IEEE International Conference on Computer Vision*, 2425–2433.

Bahdanau, D.; Cho, K.; and Bengio, Y. Neural machine translation by jointly learning to align and translate. 2014.

Boididou, C.; Andreadou, K.; Papadopoulos, S.; DangNguyen, D.-T.; Boato, G.; Riegler, M.; and Kompatsiaris, Y. 2015. *et al. 2015*. Verifying Multimedia Use at MediaEval In MediaEval.

Castillo, C.; Mendoza, M.; and Poblete, B. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, 675–684. ACM.

Chen, Y.; Conroy, N. J.; and Rubin, V. L. 2015. News in an online world: The need for an automatic crap detector. In *Proceedings of the Association for Information Science and Technology 52(1*, 1–4.

Farajtabar, M.; Yang, J.; Ye, X.; Xu, H.; Trivedi, R.; Khalil, E.; Li, S.; Song, L.; and Zha, H. 2017. Fake news mitigation via point process based intervention. arxiv. preprint, (2017).

Gupta, A.; Kumaraguru, P.; Carlos, C.; and Meier, P. 2014. Tweetcred: Real-time credibility assessment of content on twitter. *Social Informatics:* 6:228–243.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. NV: Las Vegas.

Huang, G.; Liu, Z.; Weinberger, K. Q.; and van der Maaten, L. 2017. Densely connected convolutional networks. In *Proc. IEEE Conf. Comput*. Vis. Pattern Recog.

Iandola, F. N.; Han, S.; Moskewicz, M. W.; Ashraf, K.; Dally, W. J.; and Keutzer, K. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡0.5mb model size. 2016.

Jin, Z.; Cao, J.; Guo, H.; Zhang, Y.; and Luo, J. 2017. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *17). ACM, New York, NY, USA*, 795–816. Proceedings of the 25th ACM international conference on Multimedia (MM.

Joulin, A.; Grave, E.; Bojanowski, P.; and Mikolov, T. Bag of tricks for efficient text classification. 2016.

Khattar, D.; Goud, J. S.; Gupta, M.; and Varma, V. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *The World Wide Web Conference*, WWW '19, 2915–2921. New York, NY, USA: ACM.

Kim, Y. Convolutional neural networks for sentence classification. 2014.

Kwon, S.; Cha, M.; Jung, K.; Chen, W.; and Wang, Y. 2013. Prominent features of rumor propagation in online social media. *In Data Mining (ICDM)* 2013:1103–1108.

Lai, S.; Xu, L.; Liu, K.; and Zhao, J. 2015. Recurrent convolutional neural networks for text classification. In Press, A., ed., *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*, 2267–2273.

Mihalcea, R., and Strapparava, C. 2009. The lie detector: Explorations in the automatic recognition of deceptive language. *In Proceedings of the ACL-IJCNLP* 2009:309–312.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. *Empirical Methods in Natural Language Processing (EMNLP)* 1532–1543.

Ruchansky, N.; Seo, S.; and Liu, Y. 2017. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 797–806. ACM.

Salakhutdinov, R.; Mnih, A.; and Hinton, G. 2007. Restricted boltzmann machines for collaborative filtering. In *07), Zoubin Ghahramani (Ed.). ACM, New York, NY, USA,. DOI=*. Proceedings of the 24th international conference on Machine learning (ICML. 791–798.

Santia, G., and Williams, J. Buzzface: A news veracity dataset with facebook user commentary and egos in international aaai conference on web and social media. 2018.

Shu, K.; Sliva, A.; Wang, S.; Tang, J.; and Liu, H. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19(1):2017.

Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; and Liu, H. 2018. *FakeNewsNet: A Data Repository with News Content*. Social Context and Dynamic Information for Studying Fake News on Social Media.

Srivastava, N., and Salakhutdinov, R. 2014. Multimodal learning with deep boltzmann machines. *J. Mach* 15(1):2949–2980.

Tacchini, E.; Ballarin, G.; Vedova, M. L. D.; Moret, S.; and de Alfaro, L. 2017. Some like it hoax: Automated fake news detection in social networks. In *Proceedings of the Second Workshop on Data Science for Social Good (SoGood)*. Macedonia, 2017. CEUR Workshop Proceedings Volume 1960: Skopje.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. Attention is all you need. 2017.

Vinyals, O.; Toshev, A.; Bengio, S.; and Erhan, D. 2015. Show and tell: A neural image caption generator. *In Computer Vision and Pattern Recognition (CVPR)* 2015:3156–3164.

Wang, W. Y. 2017. Liar, liar pants on fire. *: A New Benchmark Dataset for Fake News Detection* 2067(10):P17–2067.

Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; and Ti-cnn, P. S. Y. Convolutional neural networks for fake news detection. 2018.

Zhang, J.; Cui, L.; Fu, Y.; and Gouza, F. B. 2018. Fake news detection with deep diffusive network model. *CoRR* abs/1805.08751.