Grenze International Journal of Engineering and Technology, Jan Issue

Fake News Identification using Pyramidal Network with Machine Learning Algorithms

Bhaskar Adepu
Associate Professor, Dept. of IT, Kakatiya Institute of Technology & Science, Warangal, INDIA.
Email: ab.it@kitsw.ac.in

***Abstract*—Technology has seen a rapid increase in the current years as compared to what existed a long time ago. Domains like artificial intelligence & machine learning, block chain development, cyber security has evolved into whole different era of expertise. Usage of social media like face book, twitter, instagram, youtube, flicker, whatsApp etc., has grown rapidly as well. With this, we have seen a rise in articulation of news articles which are usually untrue and misleading. The social media promoted in spreading of these fake news snippets across the world easily. Online social media users are easily infected by online fake news by deceptive words, which are having a profound effect on society offline. Various methods ranging from Naive Bayesian models to complex bi-directional recurrent neural networks have been used in this domain. With the growing research in the domain of Natural Language Processing (NLP) and computer vision, researchers have started to implement cross domain techniques. Several NLP techniques proved to be useful in computer vision applications and vice versa. Inspired from these, we propose a multilingual embedding approach followed by a pyramid based feature extractor that extracts key features of a news headline which are fed to several classifiers to determine its nature. Pyramid based networks have received huge praise and interest in the computer vision domain. Our model has shown better results than existing models.**

***Index Terms*— Fake News, Machine Learning, Pyramidal Network, Natural Language Processing, Social media.**

I. INTRODUCTION

Today the use of internet and social media has been increased tremendously. This in-turn increased the transfer of social information rapidly in the form of news articles, blogs etc. as the social media has many users and publishers. This paved a path to many social media users to spread the fake news rapidly. Now-a-days the spread of fake news has become the major issue as it may influence on people's sentiment or it may have political and financial benefits. It had become a major challenge to the researchers to classify between fake and genuine news in social media.

There are many cases which showed ill effects on society due to fake news and one such effects was during the COVID-19 pandemic outbreak. During this pandemic situation there were many fake news circulated in social media about the outbreak of COVID-19 in India especially after the commencement of Tabligi Jamaat meeting held in New Delhi. Most of the news circulated were reported as fake by many fact checking websites. This incident due to fake news propagation had shown many ill effects on one of the minority group in the country.

Grenze ID: 01.GIJET.9.1.64
© Grenze Scientific Society, 2023

To put a break to such incidents efficient fake news detection model has to be developed using deep neural network. In this research paper we are going to propose a novel supervised learning model which gives better results for detecting fake news[4][5][6].

A. Machine Learning Introduction

Machine learning is the one of the branch in Artificial Intelligence to work automatically or give the instructions to a system to perform an action. Here it works based on the data which we are giving to a system. Now a day's Machine Learning (ML) algorithms made a human to work simple. It helps in development of computer programs in advance model. In these days self-driving cars, hand written character recognition systems, stock market predictions are few machine learning applications. In self-driving cars, the car drives itself from the source to the destination without help of any member. While travelling from source to destination if any obstacle comes all of sudden, then the vehicle will slow down or stop. In the same manner handwriting recognition will helps us to reorganize a word or a number from the scripts of writer. In stock market it helps in sudden increment or decrement of share price it will help us to buy the share or not. ML is a domain in which experts intent to understand data and its behaviour, fit certain statistical models or if not complex models for future use. Deep learning is one of the advance concepts of machine learning. Machine learning algorithms are used to predict the future based on the past or historical data. Machine Learning algorithms are categorized into the following types:

1. Supervised learning Algorithms: Linear regression, Decision tree, Neural networks.
2. Unsupervised learning Algorithms: Various clustering algorithms like K-Means, Gaussian mixture models.
3. Reinforcement learning Algorithms: Utility learning, Q learning.
4. Semi supervised learning Algorithms: Graph based methods, Self-training.

II. Literature Survey

Many researchers have tried to detect the fake news in social media using various approaches and methodologies. We have implemented some of them and analyzed their results.

Ref.[1] has proposed a machine learning model with a new feature set and combined them with the existing classifiers for better prediction performances. The authors have collected 2282 Buzz feed news articles which are published during US 2016 elections. They have extracted various new features like lexical features, Psycholinguistic features etc and given them to the classifiers NB, KNN, RF, SVM, XGB and compared their performance in fact prediction. This type of approach has many advantages but it is not suitable for huge datasets.

Ref.[2] in their approach used deep learning models to detect the fake news in social media. In their work the author's have implemented a new deep learning model called Bi-directional LSTM to analyze a sequence from front to rear and rear to front. This approach gives the best results but it is a time consuming model.

Ref.[3] in their research work proposed a methodology for detecting objects using feature pyramids called as feature pyramid networks. The authors in their work they have implemented a architecture with connections in lateral which are used for constructing high-level semantic feature maps at all levels. The use of pyramid networks given them the best results which take very less time.

III. Proposed Classifiers

The algorithm that sorts information or categorizes information into labeled groups is known as classification model. To quote, we have filters to detect spam that search and identify news articles as "fake" or "real." Usually machine learning classification models are concrete implementation of pattern recognition. These blend in an essence of machine theory with a robust runtime application. They're more than just a sorting utility for grouping unlabeled instances into different groups or "mapping" them.

Many ML models often give out their interpretation as a probability score that allows developers to gain flexibility in choosing desired result. They are an essentially critical part of analysis in unsupervised learning, and classifiers are how systems classify and analyze unlabeled data in supervised or semi-supervised learning [7].

Neural Networks: To understand neural network, one can map its analogy to the incredibly complex part of human ecosystem, the brain. As shown in Fig.1, it has similar terminology of neurons which resemble the neurons of mind and kind of work in the similar light of functionality. They pass on the information they get, interpret to further similar structures for more understandability. In a way, it's very resembles statistical functions of curve and regression analysis, often cited in machine learning works [8][9]. A neural network

mimics the working of human complex brain by having a large structure of well organized neurons that pass information from each layer to another. They all function like small set of linear regression models. All the intermediate outputs are fed into units called activation units for several technical features like normalization etc. The last of the layers has elements that match to the input ways that show resemblance to it very efficiently.
To optimize the way they work or function , neural network keep tuning their middle layers until they come to a way that the weights perfectly or close to perfectly give the desire output. They are so powerful that they can even affect entire performance by trying to get the work done by getting desired output. They don't need any pre feature extraction technique because neural networks do this on their own in the hidden part of CNN layers architecture.

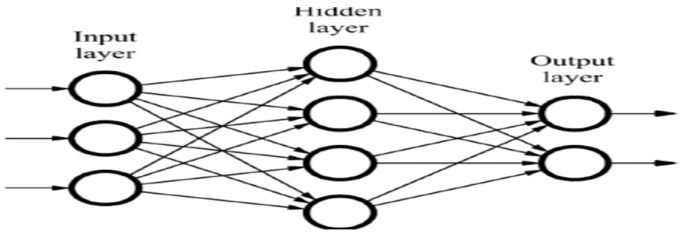

Figure 1: Neural Network

A. Text Transformations

ML techniques use features as rows data and attributes in columns. But text is in human understood language which isn't understood by machines. To tackle this, we convert our text into numbers that represent this text but since features are in 2D, we use vectors. This is called Text Transformation.
Converted text is in numbers which make it possible to analyze them and produce fruitful result that gives a meaning. Examples are whole documents or utterances in text analysis whose vectors are often identical in length. Each of the vector representation's properties is a function. When viewed together a document's features describe a multidimensional feature space that can be applied to machine learning methods [8][9].

B. Frequency Vectors

A rather straight forward approach to transformation is to use the frequency of words appearing in document as their representation in numbers as shown in the text. Each converted data is a vector with numbers which are the frequency of word tokens that appear in text documents. As shown in figure, we can either take the frequency count or the measured count with regard to the weights of the tokens.

C. One-Hot Encoding

When we consider models that rely on wide distributions, we observe that above method isn't that suitable since it ignores the semantics based on positioning of word in the text. This can cause huge redundancy and deviation from the desired goal while transformation and modeling. To tackle this, we come up with one hot-encoding which is a vector based representation that represents availability of tokens at respective positions in each text. Each vector is of dimension equaling the total number of tokens in the entire corpus. Each element of these vectors is an indication of the presence of that token at that position with either 1 or 0.

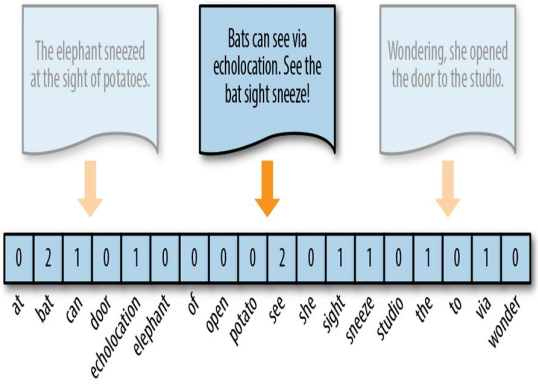

Figure 2 : Frequency Vector

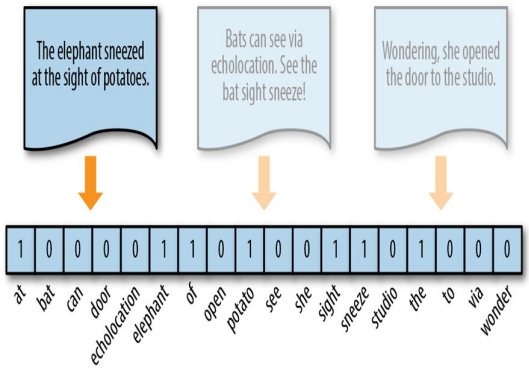

Figure 3 : One hot encoding

The above technique is incredibly well suited for documents of not that huge size. They tackle the problem posed by Frequency based methods. These are also suited because in DNNs, we observe values are batch normalized which means throughout the network, Boolean numerical like -1,0, 1 only appear. Calculations with these numbers when the input is also ones and zeroes is highly efficient since it isn't the problem of arithmetic operations but more of binary operations between zeroes and ones which are effectively fast and easier for any kind of machine to operate.

IV. Proposed System or Methodology

A. Word Embeddings

The above mentioned methods have been used for a very long time. However, the breakthrough in the world of Natural Language Processing (NLP) applications happened with the introduction of word embeddings. Previous methods did a great job at countering text and numerical problems but they failed at approaching a fixed spatial area for all the numbers to reside. Not approaching this space leads to a problem of ignoring texts which don't share the exact words but they semantically mean the same. Trigonometric measures of finding word similarity fail when the words don't match. This isn't the perfect way to process text documents with the increasingly development of human languages. Consider the words stethoscope and earphones. Both are similar in functioning although they are from completely different fields. The previous methods would definitely fail to identify their closeness based on the frequent appearance or inverse of it or just the count. In case of embeddings, embeddings show vectors with elements that represent attributes or qualities of the words and a measure of how much they exhibit that attribute. Earphones and stethoscope closeness can be found and leveraged in this case while it would fail in the other cases. This way we find word embeddings working lot better than the other methods.

Since making word embeddings requires huge corpus of data, we can't expect the available or the most popular ones from any just source. This can only be expected from pioneers like Microsoft or Apple or Google. Fortunately, Google happens to be the one to present us the world famous Word2Vec embedding. Word2Vec is the go-to vectoring method when it comes to word embeddings. It uses the CBOW method to make embeddings in a spatial area that exactly fits the representation of words about 1-10 million in count. Some other techniques developed by organizations like Face book, Amazon use DNNs to make embeddings. doc2vec uses a paragraph vector which is incredibly reliable as it does not depend on training targets but it learns in unsupervised way to leverage the constant size vectors. It tries to relate the words like pillow and bed closely while it distances the words like pillow and chair. Although its benefits, it really gives less regards to the positioning of words just like what we observe in m-gram models. Finally it gives an amazing outcome that has better generalization than the gram models and not really big size which makes it an incredible match to use in modeling techniques for automation processes and detection.

Feature Pyramid Networks: The method is quite simple, as shown in the following figure. All the layers are linked to each other be it any hierarchy i.e, the top down or bottom up and you'll find that this implementation actually gives an enrich semantic information that can be used for further textual modeling for detection.

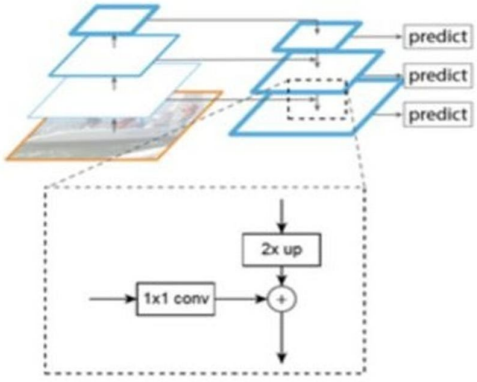

Figure 4: Feature Pyramid Network

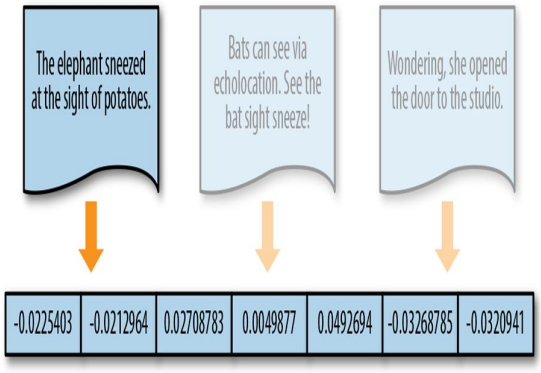

Figure 5 : Word Embedding

The algorithm layout is split into a set of third degree layers: the bottom-up convolution neural network (left in the upper figure), the top-down method (right in the upper figure), and the side-to-side relation between the

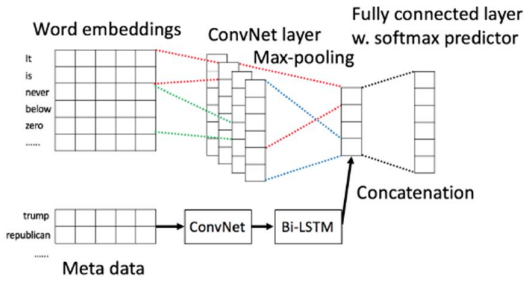

Figure 6: Proposed hybrid CNN for integrating text and meta data

features. In reality, the bottom-up component is the front phase of the CNN. In the front phase, the layout or say the representation of the function mapper changes once it parses certain levels, but not when parses other levels. Author will identify levels that do not change amount or quantity of feature mapping, so that every extraction comes out from the last level of each phase, so that a feature like resembling structure of pyramid can be created. Specifically, for ResNets, the author uses the activation output of the last residual structure of each point. The output of these residual modules is {CL2, CL3, CL4, CL5}, corresponding to the output of convol2, convol3, convol4 and convol5. Up sampling was used in the top-down process. Almost all up sampling uses the interpolation process that is, based on the pixels of the original image, a new interpolation variable is inserted between the pixels using an effective interpolation algorithm, thus increasing the size of the original image. By up-sampling a feature mapping, up-sampled feature mapping is the same size to next-layer feature map. The horizontal link between the sides is essentially to fuse the result of the up sampling with the feature map created from the bottom up. We perform a 1×1 convolution operation on a feature map of the corresponding layer created in the CNN, and fuse it with an up-sampled feature map to obtain a new feature map that combines the features of different layers with richer details. The objective of the 1×1 convolution operation here is to adjust the channels, and the requirements are the same as the channels of the latter layer. After the merge, the 3*3 convolution kernel will be used to convert each merge result, the purpose of which is to remove the aliasing effect of the up sampling, so you get a new feature map. You can use several new function maps to iterate layer by layer. Suppose the corresponding feature mapping results are PL2, PL3, PL4, PL5, and correspond to the convolution results CL2, CL3, CL4, CL5 from the bottom up. All levels in the pyramid system are divided into the classification layer (regression layer).

B. Traditional Classifiers

Support Vector Machines

This is the supervised learning model which is also collaborated with associated algorithms which is mostly used for classifier and regression based analysis. It constructs a n-dimensional plane in which each data item is pointed as a point i.e., each coordinate indicates the value which in-turn indicates a feature. This help in classifying by dividing the hyper plane into two different classes.

XGBOOST

XGBoost is the most prevailing machine learning algorithm. It stands for eXtreme Boosted trees. It is mostly used solving problems which include unstructured data like images etc. It is used to solve classification problems regression problems and also supports maximum number of programming languages to implement it. XGBoot is efficient because it is an ensemble learner. This classifier being an ensemble learner, it uses bagging and boosting.

Random Forest

Random Forest is also a supervised model. It is another ensemble technique which is used for classification and regression analysis. It contains multiple decision trees which are used as base learning models. It constructs the decision trees with data points and get the forecast from each and every point and picks the best solution from them.

V. EXPERIMENTATION AND RESULTS

Our idea is to use Feature Pyramid Network to extract features, we will not be training a classifier head to classify the text to binary labels. Our idea is also to make a hybrid pipeline where a Feature Pyramid Network would extract the features at lower levels and provide them to the simple yet efficient traditional machine

learning models hence not complicating the task too much but utilizing both statistical and deep learning methods for our classification. We have trained and evaluated models with and without features from Feature Pyramid Network. Our primary models include Support Vector Classifier, XGBoost classifier, Random Forest classifier, Logistic Regression.

Table I shows the description of Liar dataset which includes 12.8K human labeled short statements from politifact.com's API. Every statement is evaluated by a politifact.com editor for its truthfulness. After initial analysis, we identified duplicate labels, and merged the no-flip, full-flop, half-flip labels into true, false, half-true, labels respectively. Here we considered six fine-grained labels for rating of truthfulness such as pants-fire, false, barely true, half-true, mostly-true, and true. We sampled the statements from a variety of contexts/venues including news releases, TV/radio interviews, campaign speeches, TV ads, tweets, debates, face book posts, etc. Experimental results are depicted in Table II and Table III.

TABLE I: LIAR DATASET DESCRIPTION

Dataset Statistics	
Training data set size	10,269
Validation data set size	1,284
Testing data set size	1.283
Average statement length	17.9
Top 3 Speaker Affiliations	
Democrats	4,150
Republicans	5,687
None	2,185

TABLE II: DIFFERENT ML MODELS AND THEIR ACCURACY

Model	**Accuracy**
Logistic Regression	79.06%
Random Forest	99.58% (overfit)
XGBoost	82.23%
Support Vector Classifier	80.82%
XGBoost (with FPN)	**85.00%**
Support Vector Classifier (with FPN)	**87.50%**

Following table 3 shows the experimental results on the LIAR data set based on text only models and text with meta data like subject, speaker, job and context.

TABLE III: COMPARISON OF ML AND DL MODELS PERFORMANCE

Model	**Valid**	**Test**
SVM	0.258	0.255
Logistic Regression	0.257	0.247
Bi-LSTMs	0.223	0.233
CNNs	0.260	0.270
Hybrid CNNs		
Text with Subject	0.263	0.235
Text with Speaker	**0.270**	0.240
Text with Job	0.268	0.259
Text with Context	0.250	0.243
Text with All	0.247	**0.271**

VI. CONCLUSION AND FUTURE SCOPE

Conclusion: The approach is to use the following in a sequence. The news article is first pre-processed into word embeddings which are then fed to the pyramid network to get the word vectors. These word vectors are then fed to the traditional set of classifiers to get the classification result. The idea here is to extract and make use of features from the pyramid network which proved to be super useful in Computer Vision Applications. Results show that our technique improves accuracy by 7.5% on SVM Classifier and 5% on XGBoost Classifier. Hence, we can conclude that feature pyramid network is a useful technique in Fake News Classification.

Future Scope: This work would be enhanced to give still better accuracy by modifying the learning rate, threshold, amount or quantity of neurons in the hidden layer. It can also be enhanced to a multi class neural network.

REFERENCES

[1] Julio Reis, Fabricio Murai and Adriano Veloso, “Supervised Learning for Fake News Detection”, Intelligent Systems, IEEE 34(2):76-81, March 2019. DOI:10.1109/MIS.2019.2899143

[2] Pritika Bahad, Raj Kamal and Preeti Saxena, “Fake News Detection using Bi-directional LSTM-Recurrent Neural Network”, Procedia Computer Science 165:74-82, February 2020. DOI:10.1016/j.procs.2020.01.072

[3] Tsung-Yi Lin, Jonathon Shlens, Barret Zoph, “Revisiting ResNets: Improved Training and Scaling Strategies”, https://doi.org/10.48550/arXiv.2103.07579

[4] J. S. Sonawanee and D. R. Patil, "Prediction of heart disease using learning vector quantization algorithm," 2014 Conference on IT in Business, Industry and Government (CSIBIG), Indore, 2014, pp. 1-5.

[5] Z. Shen, M. Clarke, R. W. Jones and T. Alberti, "Detecting the risk factors of coronary heart disease by use of neural networks," Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1993, pp. 277-278.

[6] A. Dewan and M. Sharma, "Prediction of heart disease using a hybrid technique in data mining classification," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2015, pp. 704-706

[7] H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng and E. J. Lin, "HDPS: Heart disease prediction system," 2011 Computing in Cardiology, Hangzhou, 2011, pp. 557-560.

[8] M. G. Feshki and O. S. Shijani, "Improving the heart disease diagnosis by evolutionary algorithm of PSO and Feed Forward Neural Network," 2016 Artificial Intelligence and Robotics (IRANOPEN), Qazvin, 2016, pp. 48-53.

[9] S. A. Hannan, A. V. Mane, R. R. Manza and R. J. Ramteke, "Prediction of heart disease medical prescription using radial basis function," 2010 IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore, 2010, pp. 1-6.