Institution of Electronics and Telecommunication Engineers

KOLKATA, CENTRE, Plot No J1-7, EP-Block, Sector V, Salt Lake, Electronics Complex, Kolkata 700091

A Quarterly News Letter

SIGNAL

Volume - 16 :: No - 2 :: April-June, 2019

CHAIRMAN'S MESSAGE

Mr. P. S. Biswas

The newsletter "SIGNAL" was out of circulation since 2005 and it has been revived by the EC since 2019 in its web version. This issue contains an advanced technology article on fake news detection written by one of the esteem fellow of IETE Dr. Pradosh K. Roy. Current activities of IETE has also been reflected in this issue.

Apart from being published in our website, this will be e-mailed to our IETE Kolkata members.

A special thanks to our Editor, Prof. J.K. Mandal, our Hon. Secretary, for taking initiative on this project.

We invite contributions/articles for this newsletter from our members. Your contribution may please be sent to secretaryietekolkata@gmail.com. If accepted, it shall be published in future edition.

With Greetings and Best Wishes

Partha S. Biswas

Members

- ⇒ Mr. P. S. Biswas, Hon. Chairman
- ⇒ Dr. Jyotsna Kumar Mandal, Secretary
- ⇒ Prof. P. R. Bandopadhyay, Treasurer
- ⇒ Mr. S. C. Rudra, Immediate Past Chairman, IETE Kolkata Centre
- ⇒ Mr. Anirban Guha, Vice Chairman
- ⇒ Sri Tapan Jyoti Sen, Committee Member
- ⇒ Mr. S. D Tiwary, Committee Member
- ⇒ Dr. A. K. Mukhopadhyay, Committee Member
- ⇒ Smt. Sangita Roy, Committee Member
- ⇒ Mr. Dibyandu Majumder, Committee Member
- ⇒ Prof. Dr. Sujit Biswas, Ex. Professor, Jadavpur University Co-opted Member
- ⇒ Mr. Soumya Roy, Ex. CGM Calcutta Telecom, Co-opted Member
- ⇒ Mr. Aniruddha Nag, Co-opted Member

EVENTS

INAUGURATION OF IETE STUDENTS' CHAPTER AND GUEST LECTURE ON STRATOSPHERE COMMUNICATION

14th August, 2019: The chief guest for the session and talk was Prof. (Dr.) J. K. Mandal (Hon. Secretary, IETE, Kolkata Section). The session was also enlightened by the presence of Dean, School of Engineering and Technology (Adamas University) and Hon. HOD, ECE, SOET, Adamas University. The pioneer 37 members of the ISF received their membership certificates and inspired their fellow students to be a part of the prestigious professional body (IETE).

During the latter half of the event Prof. (Dr.) J.K. Mandal delivered an informative and interactive guest lecture on Stratosphere Communication which was extremely interesting for the student and faculty members of the department. Assistant Prof. Jeet Banerjee received an appreciation certificate from IETE, New Delhi for his continuous efforts for establishment of ISF at Dept. of ECE, Adamas University. Hon. HOD (ECE) delivered his vote of thanks to the Hon. Guest, university administration, students, and the entire department. The session was completed with the handing over of an appreciation certificate to Hon. Guest speaker.

" You have to grow from the inside out. None can teach you, none can make you spiritual. There is no other teacher but your own soul. "

—Swami Vivekananda

EDITOR'S COLUMN

Dr. Jyotsna Kumar Mandal

The Managing committee took charge during third week of July 2018. Since then regular activities are being organized on behalf of IETE Kolkata Centre. The Executive Committee already conducted seven regular meetings and one Emergency EC meeting during this period. The decision has been taken to revive ISF branches and to initiate new branches. Initiatives has already taken in this regard and the process of motivating technical institutes are going on. Initiatives are taken to induct more new members. During this period two ISF programs have been organized. It has also been decided to provide grant in aid to organize ISF activity. During the period July 2018 to March 2019, the IETE Kolkata Centre has organized seven activities out of which two activities are done under the ISF Centres and five activities are organized at IETE Kolkata Centre.

As per decision of 7th EC meeting of IETE Kolkata Centre the EC has arranged to revive the Quarterly News Letter "SIGNAL" of IETE Kolkata Centre under the leadership of the Secretary, IETE as Editor. This is second in the series of revived version published in April 2019. Long tradition has been able to continue again from January 2019.

This issue contains one highly technical article written by highly experienced Fellow of IETE, Dr. Pradosh K. Roy. Besides this some activities have also been reflected. This volume number 16 and the issue number 2, April-June 2019. This redesigned newsletter will be uploaded in to the web portal of IETE and the same will be circulated to all members of IETE. The following sections are included into the News Letter "SIGNAL".

⇒ From Chairman's Desk

⇒ Editor's Column

⇒ IETE Activities

⇒ ISF Activities

⇒ Technical Article

⇒ Round the Globe

⇒ Technical News/Quiz

On behalf of the members of EC, I would like to thanks to all members and my scholars for taking positive role to reactivate the newsletter. Hope this will be a useful information and technical brochure for the esteem members. I would like to request our esteem members to participate in writing technical articles into this newsletter.

Best wishes

Jyotsna Kumar Mandal

Secretary, IETE Kolkata Centre

Editor, SIGNAL

"Forget not that the grossest crime is to compromise with injustice and wrong. Remember the eternal law: you must give, if you want to get."

– Subhash Chandra Bose

Machine Intelligence in the Detection of Fake News

Pradosh K. Roy , FIETE, B14,126 Ho Chi Minh Sarani, Calcutta 700008, roypkin@yahoo.com

Introduction : The bigram 'Fake News' , defined as 'fabricated information that mimics news media content in form but not in organizational process or intent'[1] ; is a growing menace and its detection defined as 'the task of categorizing news along a continuum of veracity, with an associated measure of certainty' [9] is believed to be a complex task. Cascading of fake news compared to the truth , in all categories of inoformation ; according to a recent analysis by Vosoughi et al. , is faster than the real news on social media websites and the effects are more pronounced for false political news. [2]. Consequently, it poses a grave threat not only to the reliability of certain media outlets but to the government and the society as well. It is reported that more than 62% of U.S. adults consume news on social media and there is a growing tendency in India to rely on social media . Moreover, the volume of disseminated information and the rapidity in which it is spread in social network sites make it practically impossible to assess the reliability of a news in a timely manner , highlighting the necessity for automatic hoax detection systems. Hence, for redesigning the information ecosystem in 21st century , automatic detection of fake contents is an emerging agenda for research and action . Readers interested in some recent researches in this context may refer to [3][4][5][6][9] and the references cited therein.

The objective of the communication is to introduce fake news detection i.e. information of doubtful veracity from Natural Language Processing [NLP] and Machine Learning [ML] perspectives and to exemplify the significance of statistical validation .

Data Sets: Labeled datasets available for experimentation are (i) Open Sources dataset having 9,408,908 articles of which 11,161 articles are in two categories, fake and reliable, [10] (ii) Kaggle dataset on fake news consisting 20,800 articles [11] and (iii) Github repository for fake and real news dataset by George McIntire , having two sections , headlines and text of the news [12].

Linguistic Cue plus Machine Learning Approach

For building a Machine Learning model, feature selection is of primary importance for optimum performance of the system.To reveal cues of deception in the contents and headlines , features normally used are :

(i) *n-grams Count Feature* , (ii) *tf-idf* :Term Frequency- Inverse Document Frequency , (iii) *Word Embedding* , (iv) *Sentiment Polarity Score* and (v) *Linguistic features* such as readability ease and lexical diversity. We define n-gram count feature, used for counting occurrences of n-grams in the title and body of the news as the ratio $ratio\ of\ unique\ n-gram = \frac{total\ unique\ n-gram}{total\ n-gram}$

Term frequency *tf(t,d)* is defined as $$tf(t,d) = 0.5 + 0.5.\frac{f_{t,d}}{\max\{f_{t',d}: t' \in d)\}}$$

Thus , *tf(t,d)* is raw count of the term in a document $f_{t,d}$ divided by $\max\{f_{t',d} : t' \in d\}$, the number of words in the document. Likewise, Inverse Document Frequency *(idf)* is the number of times a word occurs in a corpus of documents. This facilitates to understand which words are important . Usually the natural log - normalization of *idf* given by the following equation is used $idf(t,D) = log\frac{N}{|\{d \in D: t \in d\}|}$

Where N is the total number of documents in the corpus D , N = |D| , $|\{d \in D : t \in d\}|$ is the number of documents where the term t appears. Then *tf-idf* is calculated as *tf (t,d) . idf(t,D)* .

50 dimensional vector space representations of the words are given by *word embeddings* ; whereas *sentiment polarity* scores could be assessed by the Natural Language Toolkit [NLTK] for positive , negative and neutral sentiments. *Linguistic features* such as readability ease and lexical diversity, represent the texts statistically and also represent context of the sentences in terms of ease of reading it.

Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y). Classification belongs to the category of supervised learning where the targets also provided with the input data. Various ML classifiers used for fake news detection are : Random Forest (RF), Support Vector Classifier (SVC), Gaussian Naïve Bayes (GNB), AdaBoost (AB), K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP) and Gradient Boosting (XGB).

statistical validation: Cross validation is one of the model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. For binary classification of data , precision is the fraction of relevant instances among the retrieved instances, while recall (also known as sensitivity) is the fraction of

Cont..

relevant instances that have been retrieved over the total amount of relevant instances. Both precision and recall are therefore based on an understanding and measure of relevance. Experimental results from a recent study , confirm that the boosting algorithms viz. Gradient Boosting (XGB) and AdaBoost are performing better compared to other classifiers in terms of average accuracies and standard deviations [7]. Finally the test accuracy could be parameterized using F1 Score , harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0. F1 Score = [Precision-1 + Recall-1]/2 . High value of F1 score as shown below , statistically validate the accuracy of NLP and ML approach in fake news detection.

Dataset	Precision	Recall	F_1-Score
Open Source	0.92	0.92	0.92
Kaggle Data Set	0.93	0.94	0.94
Github Repository	0.89	0.87	0.89

Conclusion : As proliferation of internet and social websites have led to an exponential growth in opinion spams and fake news, in recent times, early detection of misinformation/disinformation is of primary importance for redesigning the 'information ecosystem in the 21st century. Though linguistic cue appears to be promising , there is an alternative Network Approach in which 'network information, such as message metadata or structured knowledge network queries can be harnessed to provide aggregate deception measures' [8]. Both forms typically include ML techniques for training classifiers to suit the analysis. Though research on fake news detection is in the primary stage , it is envisaged that eventually it would lead to the creation of a 'system and culture having values that promote truth'[1].

Acknowledgment : The author is thankful to the Dr. V.K. Shrivastava, Asia Pacific Institute of Information Technology, Panipat, Haryana for extending institutional hospitality in November, 2018 ; where the final results were discussed .

References

[1] David Lazer, Matthew Baum, Yochai Benkler, Adam Berinsky, Kelly Greenhill, Filippo Menczer, Miriam Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven Sloman, Cass Sunstein, Emily Thorson, Duncan Watts, Jonathan Zittrain , The Science of Fake News : Science, 359(6380), 1094-1096 (2018)

[2] Soroush Vosoughi, Deb Roy, and Sinan Aral., The spread of true and false news online, Science, 359 (6380), 1146–1151 (2018)

[3] Horne BD ; Adali S. , This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news : Paper presented at: The 2nd International Workshop on News and Public Opinion at ICWSM; Montreal, Canada (2017) : https://arxiv.org/abs/1703.09398

[4] Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass , Preslav Nakov : Predicting Factuality of Reporting and Bias of News Media Sources (2018) , https://arxiv.org/abs/1810.01765

[5] Ver´onica P´erez-Rosas, Bennett Kleinberg, Alexandra Lefevre, Rada Mihalcea , Automatic Detection of Fake News , Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, August 20-26, 3391–3401 (2018).

[6] Miguel ´A. Carreira-Perpi˜n´an , CSE176 Introduction to Machine Learning , Lecture notes ,University of California, Merced; November 28, 2016

[7] Arvinder Pal Singh, M. Fernandez, S. Choubey , Mahima Goel , Performance Comparison of Machine Learning Algorithms in Fake News Identificaion , Proceedings of ICACDS 2019. Ghaziabad, CCIS Springer 2019 [in press]

[8] Niall J. Conroy, Victoria L. Rubin, and Yimin Chen , Automatic Deception Detection: Methods for Finding Fake News , *ASIST 2015, November 6-10*, 2015, St. Louis, MO, USA.

[9] Combating Fake News: An Agenda for Research and Action , Final Report , Northeastern Uni versiy & Harvard Kennedy School, May 2017.

[10]Opensource Dataset : http://www.opensources.co/

[11] Kaggle Dataset, https://www.kaggle.com/jruvika/fake-news-detection

[12] Github Repository, https://github.com/GeorgeMcIntire/fake_real_news_dataset

Inauguration of IETE students' Forum and Lecture on Stratosphere Communication