International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1952101

Influencing Opinions through False Online Information : A Study

Baldev Singh
Department of Computer Science and Information Technology, Lyallpur Khalsa College, Jalandhar, India

ABSTRACT

Online Social media generates lot of information now-a-days. It is not legitimate information so there are the chances of fake and false information produced using social media. It is very alarming that majority of the people getting news from social media which is very much prone to false information in comparison to traditional news media which is very dangerous to the society. One of the primary reasons to influence opinion through false information is to earn money, name or fame. In this study, the focus is on to highlight false information generated through fake reviews, fake news and hoaxes based on web & social media. It summarized various False information spreading Mechanisms, False Information Detection Algorithms, Mining Techniques for Online False Information to detect and prevent false online information.

Keywords : Fake News, Tweets, Data Mining, False Information Detection Algorithms

I. INTRODUCTION

Internet is a wide source of information. An information is legitimate or not, it is not always easy to predict. Although some of the sources are legitimate sources that may belong to public, public limited, private limited or even private concerns. But the online private sources of information cannot be easily predicted whether they provide true information. It means to say that they may or may not be trustworthy. Now-a-days mechanics of Internet is totally changed. In later period, Internet was used by majority of users as a consumer and vary few were the producers which means to say that majority of the users get information or required available tools through the usage of the Internet but today's percentage of the category of producers (providing online information, services and other tools etc.) is increased many fold. Now-a-days, it is to be decided by the users whether Internet is Boon or Bane. It depends upon the usage whether it is used properly or excessively.

Now-a-days, Internet is a catalytic tool of opinion building as millions of tons of information spread to millions of user in minutes. Opinions of the Internet users is very important from different aspects whether these are political, social, religious or business etc. Right information leads to good results or say good opinions of the users but opinions made through false information [1,2]may lead to disastrous results. It is very alarming that majority of the people getting news from social media which is very much prone to false information in comparison to traditional news media [3,5] which is very dangerous to the society. One of the primary reasons to influence opinion through false information is to earn money, name or fame. In this study, the focus is on fake reviews, fake news and hoaxes based on web and social media.

False information spreading Mechanisms

False information is generated either intentionally or unintentionally but leads to negative influences. False information leads to fictitious positive sentiments.

Large scale fake and false information is generated through bots and sock puppets[4,6,7] by way of fake accounts. The basic purpose of such false information is to create illusion of public consensus. Bots can use Tweeter to retweet in bulks and following to each other to influence the opinion. Similarly Sockpuppets acts through online discussions and influence the opinion and disagree with those community members who opposes. These bots [8,19] and sockpuppets performances as bad actors. The humans are the poor judges when we talk about false information as the metric to find the authenticity is not versatile, it is mainly judgement basis. The false information is tried to present as true information by the bad actors on the basis of some aspects such as

- By presenting as well-referenced
- Same false contents through multiple channels
- Same false contents by multiple times
- By way of well-written presentation
- Lack of education of information consumers
- Confirmation biases of information consumers

False information can be categorized on the basis of Intent of the author, Knowledge (opinion based or fact based), the main actors' role is played in spreading false information by using social media is through the following mechanism:

(i) Tweets and retweets
(ii) Likes
(iii) Shares (information shares online)
(iv) Comments
(v) Subscribe to get more and more false information
(vi) Rating
(vii) Referencing
(viii) Spreading behavior
(ix) Short bursts of false information in groups
(x) Blogs
(xi) Articles
(xii) Self-motive advertisements
(xiii) Information propagation by using artificial intelligence in posting patterns

II. LITERATURE REVIEW

Some studies are contributed here that are part of the literature review. Researchers have considered various aspects to analyze the false information. The analysis made by different researchers is categorized as stylistic analysis [9], sentiment analysis [10], venue based analysis [11], Text size based analysis, psycholinguistic analysis [12], deceptive opinion spam index [13], Temporal characteristics based research aspects of false information are made by some of the researchers which are based on the timestamping like Inter-arrival times of successive reviews [Shah et al. [14], Hooi et al. [15], and Ye et al. [16], many fraudulent check-ins/reviews [17], repeatedly post[18], group structure based reviews [Mukherjee et al. [19].

Group structure-based reviews: The review analysis is based on the ratio of number of reviews made for number of at least **m** common products to the total number of reviews for an associated product. Fraudster groups of large group size and having higher support count fundamentally leads to false information. Text size-based analysis: Short text based opinions tends to false information basically, having chances of more extreme values like (opinion on point scale 1-5) e.g. 1 or 5 or near to extreme values. Fake news based false information detection parameters are also suggested by the researchers. Like it is general opinion that fake new have longer title , more proper nouns, lesser number of stopwords, more capitalized and catchy words, less explained information, satire news[20].

Propagation characteristics leads to opinion influencing. The study by [Gupta:32-22] revealed that only 30 users contributed about 90% of the

retweets made in context of fake images. These were made during hurricane Sandy through Twitter. Another study made by [21] revealed that 1.3 million rumor tweets as well. That study is also related to the study made by [22]. False information propagated through cascade. Through cascades [23], false information is shared and shared on Facebook and other social platforms. False information is basically spreads at greater depth comparative to the actual or real news. Bot accounts propagated false information using follower-followee network of Twitter and this propagation steadily increased by way of rebroadcasting network.

Spam reviews: Although it is not the subject of all to identify reviews whether these are spam reviews leading to false information or non-spam reviews. Following are some indicators of spam reviews identified by researchers:

- Strange patterns
- Strange language
- Strange timing
- Strange numbers
- Strange facts
- Obvious threats
- Obvious guidelines violations

Following table shows the statistics of local consumer review survey made by BrightLocal. It highlights the consumers' behavior and opinion towards consumers reviews.

Table 1: Key Statistics of Local Consumer Review Survey [32]

86% of consumers read reviews for local businesses. Survey includes age group of people aged 18-34).
Consumers read an average of 10 online reviews before feeling able to trust a local business.
40% of consumers only take into account reviews written within the past 2 weeks – up from 18% last year.
57% of consumers will only use a business if it has 4 or more stars.
80% of 18-34 year olds have written online reviews – compared to just 41% of consumers over 55.
91% of 18-34 year old consumers trust online reviews as much as personal recommendations.
89% of consumers read businesses' responses to reviews.

Variety of topics shared on social media and there is lot of stuff that have false information as tweets, reviews, shares, comments etc. Following figure-1 shows trends of fake news and tweets demonstrated by Vosoughi et al.' in their research.

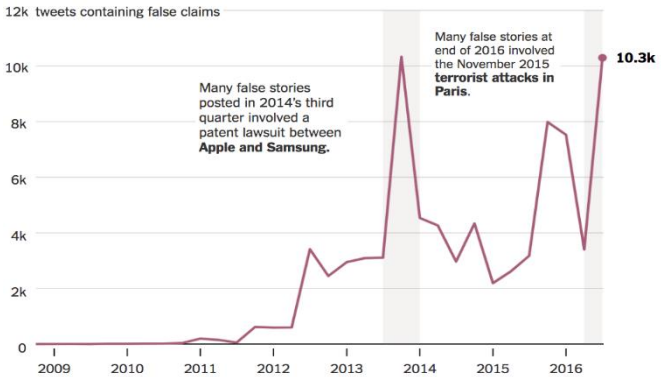

Figure 1 : Fake stories Tweet trends [28]

1. False Information Detection Algorithms

Although false information detection is a tedious and probabilistically true task but various methods and algorithms are devisee by researchers from time to time. These methods and algorithms are implemented or to be implemented based upon various contexts of false information retrievals. The false information retrievals are based on various characteristics Textual characteristics, User characteristics, Network characteristics, Propagation characteristics [24] and Debunking characteristics. These categories of characteristics encompasses false blogs, fake news, false cascades, hoaxes, and rumors etc.

The broad categories of algorithms for false information detection are mainly Feature engineering [25,29], modeling and graph based. Efficient features are developed and used for false information detection in In Feature engineering algorithms. These features may be individual or joint features. Graph-based algorithms [27,30] are based on the study of information spreaded on the basis of certain story and then use of botnet retweets in near identical time. The type of algorithms are known as modeling-based algorithms [26,28]that is bases on empirical observation of edges to create information propagation models. Its outcome also pinpoint that the detected information is really false information or there are anomalies in false information detection so that these anomalies can be eliminated.

Following tables highlights research made by researchers through the use of different categories of algorithms for false information detection.

Table 2 : Category of Algorithms and Associated detection algorithms Models

Sr.No.	Category of Algorithms	Some Associated detection algorithms Models
1	Feature-based	Sparse Additive Generative Model (SAGE) , SVM classifiers Three-class classification
2	Graph-based	Edge distributions, Dense block detection, Co-clustering, A priori algorithm
3	Modeling-based	Using Feature based and Modeling based models:- Time series analysis, Multidimensional Time Series, Correlation, Co-clustering

2. Mining Techniques for Online False Information

Data mining is the process to generate new information which is useful for some inferences that are used in decision making by various types of organizations. As our study is related to false information detection because it adversely influence the opinions of online information consumers which are made on the basis of fake news, fake tweets, fake information sharing through social sites etc. Data mining techniques are able to extrapolate patterns and generate new knowledge that is useful for decision making. Although statistical techniques and machine learning are also useful for most of such cases.

Various data mining techniques that can be used for detection of false online information made through web and social media are summarized below:

1. Tracking patterns. By the use of learning online patterns of information, inferences can be taken outcome as false or true information.
2. Classification: Classification techniques of data mining can be used for classifying the information (false or not) on the basis of defined features or patterns although this technique is more complex. Classification can be made on the basis of history of tweets, reviews, retweets, background of tweeters, information producers etc.
3. Association. Association technique can be applied by using the specific events of correlation about different patterns, reviews and other attributes of information available.
4. Outlier detection. Outliers may be the anomalies. Assume that some reviews or tweets are flooded over social sites to make opinion at a particular time period that can be false information which can be used for influencing opinion.
5. Clustering. Clustering is techniques that can be used generate clusters on the basis of various factors/features of opinion building.
6. Regression. Regression technique can be used for planning and modeling purpose. As regression leads to set up relationship on the basis of dependency so various features of considered online information can used to determine whether the information is false or not.
7. Prediction. Prediction technique of data mining is helpful to permit about the false information on the basis of historical trends of reviews, tweets that can be associated with time period like election days in a country.

III.CONCLUSION

Information generated using social media is increased many fold. The information is legitimate or not leads to false information. Various motives are there for generating false information like social, political, business that are attempted to fulfil by way of malicious resources like manual or electronic tweets, reviews, flooding information on social sites in the form of text, images, audio or video on Facebook, LinkedIn etc. sites. In this study, various resources of false information along with social media mechanisms of false information are presented. Various algorithms are highlighted in Literature Review section that are proposed by researchers for finding the false information flooded online but there is no specific algorithm, technique or model cannot be implemented because flooding of information resources are of different types and of different patterns. There are open avenues to research based on various areas like feature engineering, machine learning, data mining, big data analysis, natural processing etc. to detect and prevent false online information.

IV. REFERENCES

[1]. Alexander Smith and Vladimit Banic. Fake news: How a partying macedonian teen earns thousands publishing lies. NBC News, 2016.
[2]. Neil Shah. FLOCK: Combating astroturfing on livestreaming platforms. In Proceedings of the 26th International Conference on World Wide Web, 2017.
[3]. Craig Silverman. This analysis shows how viral fake election news stories outperformed real news on facebook. Buzzfeed News, 16, 2016.
[4]. Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Technical report, National Bureau of Economic Research, 2017.

[5]. Solomon E Asch and H Guetzkow. Effects of group pressure upon the modification and distortion of judgments. Groups, leadership, and men, pages 222–236, 1951.

[6]. Tim Berners-Lee, Robert Cailliau, Jean-François Groff, and Bernd Pollermann. World-wide web: The information universe. Internet Research, 20(4):461–471, 2010.

[7]. Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. Everyone's an influencer: quantifying influence on twitter. In Proceedings of the 4th ACM International Conference on Web Search and Data Mining, 2011.

[8]. Alex Beutel, Kenton Murray, Christos Faloutsos, and Alexander J Smola. CoBaFi: collaborative bayesian filtering. In Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014.

[9]. Tanushree Mitra, Graham P Wright, and Eric Gilbert. A parsimonious language model of social media credibility across disparate events. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. ACM, 2017.

[10]. Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy. Towards a general rule for identifying deceptive opinion spam. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014.

[11]. Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.

[12]. James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. Mahway: Lawrence Erlbaum Associates, 71(2001):2001, 2001.

[13]. Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 us presidential election online

[14]. Neil Shah, Alex Beutel, Bryan Hooi, Leman Akoglu, Stephan Gunnemann, Disha Makhija, Mohit Kumar, and Christos Faloutsos. Edgecentric: Anomaly detection in edge-attributed networks. In Proceedings of the 2016 IEEE International Conference on Data Mining Workshops, 2016. , Vol. 1, No. 1, Article . Publication date: April 2018.

[15]. Bryan Hooi, Neil Shah, Alex Beutel, Stephan Günnemann, Leman Akoglu, Mohit Kumar, Disha Makhija, and Christos Faloutsos. Birdnest: Bayesian inference for ratings-fraud detection. In Proceedings of the 2016 SIAM International Conference on Data Mining.SIAM, 2016.

[16]. Junting Ye, Santhosh Kumar, and Leman Akoglu. Temporal opinion spam detection by multivariate indicative signals. In Proceedings of the 10th International AAAI Conference on Web and Social Media, 2016.

[17]. Amanda J Minnich, Nikan Chavoshi, Abdullah Mueen, Shuang Luan, and Michalis Faloutsos. Trueview: Harnessing the power of multiple review sites. In Proceedings of the 24th International Conference on World Wide Web, 2015.

[18]. Yuming Lin, Tao Zhu, Xiaoling Wang, Jingwei Zhang, and Aoying Zhou. Towards online review spam detection. In Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014.

[19]. Arjun Mukherjee, Bing Liu, and Natalie Glance. Spotting fake reviewer groups in consumer reviews. In Proceedings of the 21st International Conference on World Wide Web. ACM, 2012.

[20]. Victoria L Rubin, Niall J Conroy, Yimin Chen, and Sarah Cornwell. Fake news or truth? using satirical cues to detect potentiallymisleading news. In Proceedings of the 14th Annual Conference of the North American Chapter of

the Association for Computational Linguistics: Human Language Technologies, 2016.

[21]. Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. Hoaxy: A platform for tracking online misinformation. In Proceedings of the 25th International Conference Companion on World Wide Web, 2016.

[22]. Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In Proceedings of the 22nd International Conference on World Wide Web. ACM, 2013.

[23]. Adrien Friggeri, Lada A Adamic, Dean Eckles, and Justin Cheng. Rumor cascades. In Proceedings of the 8th International AAAIConference on Web and Social Media, 2014.

[24]. Fangtao Li, Minlie Huang, Yi Yang, and Xiaoyan Zhu. Learning to identify review spam. In Proceedings of the International Joint Conference on Artificial Intelligence, 2011.

[25]. Vlad Sandulescu and Martin Ester. Detecting singleton review spammers using semantic similarity. In Proceedings of the 24th international conference on World Wide Web. ACM, 2015.

[26]. Brian Skyrms. Signals: Evolution, learning, and information. Oxford University Press, 2010.

[27]. Kumar, Srijan & Shah, Neil. (2018). False Information on Web and Social Media: A Survey.

[28]. Online]. Available at: https://www.nytimes.com/2018/03/08/technology/twitter-fake-news-research....

[29]. Supasorn Suwajanakorn, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Synthesizing obama: learning lip sync from audio. Journal of ACM Transactions on Graphics, 36(4):95, 2017.

[30]. Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Alfaro. Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506, 2017.

[31]. Alex Beutel. User behavior modeling with large-scale graph analysis. 2016.

[32]. Online available at: https://www.brightlocal.com/learn/local-consumer-review-survey/

Cite this article as :

Baldev Singh, "Influencing Opinions through False Online Information : A Study", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 443-449, March-April 2019. Available at doi : https://doi.org/10.32628/CSEIT1952101
Journal URL : http://ijsrcseit.com/CSEIT1952101