

FAKE NEWS DETECTION: STATE-OF-ART APPROACHES, CHALLENGES AND RESEARCH DIRECTIONS

Makhan Kumbhkar¹, Prachee Vengurlekar²

Asst. Professor, Christian Eminent College, Indore, M.P., India¹

Student, Institute of Engineering and Technology, DAVV, Indore, M.P., India²

Abstract:

“We live in the age of information chaos”. One can undoubtedly say in this day and age, data also known as news too few is more valuable than money itself. This news should be in genuine structure which is typically found in debased form. Driving us to have a critical requirement for an identity of genuine news from any conceivable phony news is the most challenging research thread. News, being a type of data can be abstract to the evidences and hotspot for its credibility. As a human, one can undoubtedly recognize genuine news from counterfeit news with the assistance of one's natural ability to conclude rationale and stunning wellspring of the data piece. Simply that one necessity few confided in sources to check for current realities and fantasies. Be that as it may, on an ongoing premise, there is a desperate requirement for some product which can nip such 'bogus news' in its bud. Fundamentally being a piece of Information Retrieval, this research domain is taking up a great deal of consideration from specialists worldwide to concoct a continuous answer for such an issue. In this article we have checked and dissected many case study, survey papers and summarized this paper to give the peer users a short thought of what counterfeit news is, its various flavors in the news range, its attributes and distinguishing proof fundamental. Additionally, we have provided various insights on the tools used for fake news detection in real time, challenges and future research directions to the researchers.

Keywords: *Fake News, survey, identification, real news, fake news detection tools, types of fake news.*

1. INTRODUCTION: *“We live in the age of information disorder”*

Understanding information disorder Terminologies

The term “fake news” was rejected by Wardle in the year 2015, with a criticism that the term is “wholly inadequate” to describe the consequences associated with it. The British government in 2017 October decided to no longer use the term “fake news” in its official documents and a recommendation is sent to the “House of Commons' Digital, Culture, Media and Sport Committee” regarding the same. However, the recent review study states “fake news” with a broad consideration as fabricated information.

Wardle distinguishes the three new terminologies to describe information disorder as follows:

1. **Dis-information**
2. **Mis-information and**
3. **Mal-information**

The Venn diagram shown in the in the figure 1: a content of fake news represents the three different information types in fake news and their relationship. Misinformation is totally associated with falseness without a harmful intent whereas mal-information is associated with true content but with an intent to harm. Disinformation is the intersection of the both.

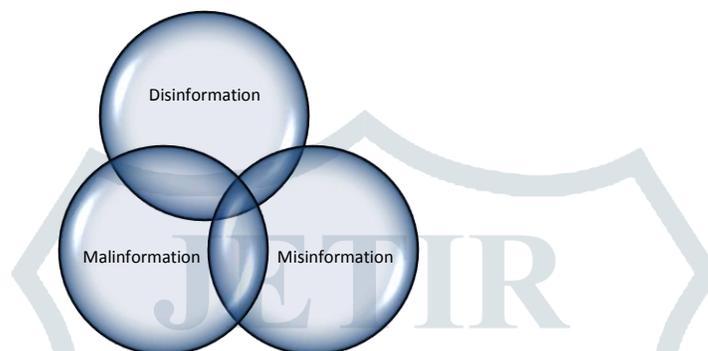


Figure 1: Contents of fake news

- **Disinformation:** The information is created by the people and is shared in public with bad and harmful intent. People often create such dis-information with a view to earn money, to have the political impact regionally or globally: or to cause unnecessary trouble in the society or individual.
- **Mis-information:** The false information shared in public but not with a harmful intent. Here the person sharing the content is unaware of the false content and thus not realize the mis-information communicated thereby. The socio-psychological aspect of the person is the main motivation in sharing the false content, believing that they are actually helping the community network.
- **Mal-information:** The information shared to the public is genuine but the intent to share is to cause harm to the individual or society at the large.

The key features presented in this paper are the introduction to fake news, terminologies used in the ecosystem of fake news detection, types of fake news. State-of-art methods classified based on the type of data associated with fake news. The rest of the paper is organized with the sections as follows: In section II, we present a brief overview of the various researchers work on the fake news detection, section III presents about types of fake news .Section IV discuss on fake news detection using visual, linguistic and non linguistic approaches .Section V presents the tools to detect the fake news and finally section VI highlights the research directions and challenges for young researchers in the fake news detection area.

II.LITERATURE SURVEY

These days, fake news can be made and spread effectively through the numerous online media stages, coming about in a broad true effect. A work by Alvaro Figueira et al., (Alvaro Figueir) Demonstrates and describes how bogus data multiplies on friendly stages and why it prevails with regards to misdirecting per users are basic to foster effective calculations and apparatuses for their initial recognition. A new flood of exploring in this space has intended to resolve the central questions utilizing strategies dependent on AI, profound learning, include designing, chart mining, and picture and video examination, along with recently made informational collections and web administrations to recognize deluding content. Samadhi Sharma et al.,(Samadhi Sharma et al)and team members in their article have checked and analyzed many research articles along with many survey articles and summed up a paper so as to provide the readers with a short idea of what fake news is, it's different flavors in the news spectrum, its characteristics and identification basic. Also included the different methods used by prior researchers in the same field. Using few researches as examples they learned about the basics of those methods used in fake news identification. The future aspects are also included in this article along with the challenges one faces while doing research in this very field. Bashar Al Asaad et al.,(Bashar Al Asaad et al) proposed a supervised machine learning approach for fake news detection using Scikit-learn library functions and python programming language. An endeavor to check the news articles validity relying upon their qualities. They executed a calculation joining a few characterization techniques with text models. It performed well, and the exactness results were moderately fulfilling. Z Khanam et al.,(Z Khanam et al) work makes an investigation of the exploration identified with counterfeit news location and investigates the customary machine learning models to pick the best, to make a model of an item with directed AI calculation, that can characterize counterfeit news as evident or bogus, by utilizing devices like python Scikit-learn, NLP for text based investigation. A work by Monther Aldwairi et al., (Monther Aldwairi, 2016) provides a solution to the users to detect and further filter the social media sites with false and the misleading information. state-of-the-art methodologies for fake news detection are presented by the researchers Wissam Antoun et al.,(Wissam Antoun et al) in their article. Gaurav Bhatt et al .,(Gaurav Bhatt et al .,2017) proposed a noval method that combines the neural,statistical and external factors to provide a solution to the challenging fake news detection task in real time. Ray Oshikawa et al ., (Ray Oshikawa et al) in their work presented a research directions at the fine and coarse grained level for the researchers and also expressed the insights on the use of natural language for fake news detection task. Their work focused on various challenging aspects on the theme with various different techniques in hand to solve the problem of fake news ecosystem.

III.TYPES OF FAKE NEWS

With the three types of information disorders as discussed in the section I of the paper,we now highlight seven frequently used cateogries of fake news staed by many researchers in the fake news ecosystem.This understanding

helps the young researchers to know the complexity associated with problem domain and also understand the challenges thereby. The seven different types of fake news are represented in the figure 2.

Types of Fake News



Figure

2: Seven types of Mis - & Disinformation

1. **Satire:** A type of fake information that has a potential impact to fool a person and may be wrongly interpreted as the factual information. The information is mainly focused on the content and makes use of sarcasm associated with it. Satire does not cause any harmful effects to the society and is created only for the entertainment purpose.
2. **False connection:** A type of fake information where the content of the information is not supported by the headline info, visual info and the captions associated with the news. It is mainly a result of poor journalism trying to relate the unordered attributes only with the view of extracting financial benefit.
3. **Misleading Content:** A type of fake information that is used to target individual. The information here is framed against a particular issue or an individual person with a harmful intention.
4. **False Context:** A type of fake information that includes the false information based on contextual genuine content and is also shared in public with harmful intention.
5. **Imposter Content:** A type of fake information that is created by impersonating the real sources.
6. **Manipulated content:** A type of fake information that narrates a complete different imagery story presenting genuine information.
7. **Fabricated Content:** A type of fake information that is created with 100% false content with a harmful intention to deceive an individual or society at the large.

IV. FAKE NEWS DETECTION – VISUAL, LINGUISTIC AND NON LINGUISTIC

Visual-based sort of phony news utilizes content that coordinates different types of media which includes graphical portrayal, for example, Photoshopped pictures and recordings. Visual news that gets the attention of the watchers are chiefly posted on stages like web-based media and media locales. Facebook, Instagram also, Twitter are well known instances of regularly utilized web-based media to post and share online content along these lines course to numerous different clients. Over 70% of its clients use them as day by day news sources to get the most recent and fastest updates. Media locales are worked by content media companies and their substance center around a wide scope of visuals and plan their destinations dependent on style and user's advantage.

Linguistic-based type of fake news is in the form of text or string content and generally analyzed by text linguistics. Its content largely focuses on text as a communication system and includes characteristics like tone, grammar, and pragmatics that allow discourse analysis. Examples of linguistic-based platforms are blog sites, emails and news sites. Blog sites are managed by users and the content produced is unsupervised which considers it easy to receive wrong information. Email is another medium where its users can receive news and this poses as a challenge to detect and validate their authenticity. It is known that hoaxes, spam and junk mails are infamously spread through emails. Popular news websites too can generate their own content and attract users with their authentic presence

V. TOOLS TO DETECT FAKE NEWS

Table 1: List of fighting disinformation tools

FIGHTING DISINFORMATION TOOLS	TOOL TYPE	STATUS	INTENDED USERS	COST	TOOL FOCUS & METHODOLOGY
BAD NEWS	Educational	Fully operational	Public	free	Process focused and online game
BOT SENTINEL	Spam detection	Fully operational	Public	free	Content and Process focused, Machine Learning & AI Based
CLAIM BUSTER	Verification	Fully operational	Public	free	Content focused, Machine Learning & AI Based
DIRT PROTOCOL	Verification	In process	Public	free	Process focused, Block Chain Management
EXIFDATA	Verification	Fully operational	Public	free	Content focused,

VI. RESEARCH DIRECTIONS & CHALLENGES

- An ideal challenge that most of the researchers today come across is the possibility of the fake claim to turn as the real news in the future. This is most important issue associated with problem solving in the domain with the risk factor involved during addressing.
 - All the methods available in real time are purely based on the creditability of the journalism and the primary source of the information.
 - Availability of data sets that facilitate to be working in all the approaches with a wide variety of news information.
 - Fake detection methods must be reliable & feasible to provide the aspect of tagging with in real time.
 - Needs to present a comparative analysis study with some pure authentic information sources, so as check the fact and provide authenticity of the information.
 - To identify and analyze every aspect of the news article and investigate the information disorder.
- There are a lot more such difficulties that outcome in issues in examination or advancement of phony news identification strategy or framework. This leads the specialists and engineers in consistent circumstances that cause them to understand that the exploration is still a long way from completed around here.

CONCLUSION

With the current situation progressing one can undoubtedly tell that in this day and age no one is anything short of a simply post writer what they mean to. There are no rules on which their realness can be based. Additionally, there are some devilish components in this general public that have effectively evolved bots that send constant spam and phony news to online media for their own advantage or simply some good times. Online media has turned into a center point for counterfeit news that should be revised before this phony news render these web-based media futile. The shortcomings in the fake news detection ecosystem are:

- Human brain disseminates the inbuilt logic function to detect the authenticity of the information, but it is possible to devise an automated machine to detect fake news yet a difficult and challenging task in real time.
- Issues associated with the methods or software's available in market currently like realorsatire.com provide a function to check one new article to be real or fake. There is a gap in the research domain for a web crawler or bots or an answering machine that can randomly check the news articles on social media.

To sum up the points, the domain of fake news detection is a ocean, such a ocean where the researchers can apply the learning of artificial intelligence, machine learning, deep learning, natural language processing and propose

the feasible automated or semi-automated solution in real time which can replace the manual state-of-the-art methods.

REFERENCES

1. Fetterly, D., Manasse, M., & Najork, M. (2004). Spam, damn spam, and statistics: using statistical analysis to locate spam web pages. In Proceedings of the seventh workshop on the Web and databases (WebDB) (pp. 1–6), Paris, France, June 2004.
2. Gan, Q., & Suel, T. (2007). Improving web spam classifiers using link structure. In AIR Web'07: proceedings of the 3rd international workshop on adversarial information retrieval on the web (pp. 17–20), New York, NY, USA. New York: ACM.
3. Ntoulas, A., Najork, M., Manasse, M., & Fetterly, D. (2006). Detecting spam web pages through content analysis. In Proceedings of the World Wide Web conference (pp. 83–92), Edinburgh, Scotland, May 2006.
4. Ciampaglia, G., Shiralkar, P., Rocha, L., Bollen, J., Menczer, F., & Flammini, A. (2015). Computational fact checking from knowledge networks.
5. Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). News in an Online World: The Need for an “Automatic Crap Detector”. In The Proceedings of the Association for Information Science and Technology Annual Meeting (ASIST2015), Nov. 6–10, St. Louis.
6. Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., & Mittal, A. (2018, April). Combining neural, statistical and external features for fake news stance identification. In *Companion Proceedings of the The Web Conference 2018* (pp. 1353-1357).
7. Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. (2010). Who is tweeting on Twitter: Human, Bot, or Cyborg? in the Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC '10, pp. 21– 30.
8. Cook, D., Waugh, B., Abdipanab, M., Hashemi, O., Rahman, S. (2013). Twitter Deception and Influence: Issues of Identity, Slacktivism and Puppetry