The technology change that has happened in Digital Media, how can we scale the authenticity of information on digital media?

Abstract

Checking the authentication of a news in online media has become one of the major issues today, it has tremendous real-world political and social impacts. However, Traditional machine learning and other AI based approaches for tackling fake news has been dramatically limited by the lack of labeled benchmark datasets.

There are many fake news detection approaches have been proposed, where most of them heavily rely on network-based techniques and AI techniques. However there are some flaws in the existing AI based solutions ,this is an attempt bridge this gap by proposing a Blended Crowdsource fake news detection approach. The aim is to combine man and machine resources to obtain a better result. Transparency of the model working as an opensource platform allows a more efficient analysis.

Introduction

To what extent the news on social media and so called "bonafide" news websites are real? Today More than 42% of Indians read news online. As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to seek out and consume news from social/digital media rather than traditional news organizations (newspapers/radio). It is extremely easy for anyone to post what they *incline* to and there is always a notion of creating panic or manipulate a person's decision or anything else that can have enduring repercussions, amidst this uncertainty and perplexity people tend to choose online sources for news information as:(i) It is very convenient, inexpensive and effortless to gobble news on digital media platforms when compared to traditional news media such as newspapers or television; (ii)It is easier to share with other readers online, Despite the advantages the quality of news is lower than the traditional news organisations, A large volumes of news articles with intentionally false information are produced for various purpose and greed.

Technology trends has made us possible access to content anywhere, anytime .The power of mobile and social is transforming how media is consumed and perceived. Continuous and instant access, particularly through sharing, empowers users to promote or destroy brands and institutions. Overnight stardom, reputational crises have become new phenomena that need to be managed.

Due to advancements in areas like image, video processing etc., news on digital media may not be authentic, the content posted may be fake or partially true. This may be due to reasons such as Miscommunication, rumours, political reasons, hatred, jealousy, promote products, for money revenge and so on

Definition
Technology can be most broadly defined as the entities, both material and immaterial, created by the application of mental and physical effort in order to achieve some value. In this usage, technology refers to tools and machines that may be used to solve real-world problems. It is a far-reaching term that may include simple tools, such as a crowbar or wooden spoon, or more complex machines, such as a space station or particle accelerator. Tools and machines need not be material; virtual technology, such as computer software and business methods, fall under this definition of technology.

Fake news, also known as **junk news**, **pseudo-news**, **alternative facts**, **false news** is a form of news consisting of deliberate <u>disinformation</u> or <u>hoaxes</u> spread via traditional <u>news media</u> (print and broadcast) or online <u>social media</u>. Digital news has brought back and increased the usage of fake news, or <u>yellow journalism</u>.

Fake news is the deliberate spread of misinformation through digital news media, Fake news can become indistinguishable from accurate reporting since it spreads so fast. One can download articles from sites, share information, re-share and add comments at the end leading to a false information widespread making it indistinguishable.

Authentic is that representing one's true nature or beliefs; true to oneself or to the person identified: entitled to acceptance or belief because of agreement with known facts or experience; reliable; trustworthy:

Significance

Using Social media/digital media for news update is a mixed blessing. On one hand it provides easy access with little or no cost and spread information in no time, however on the other hand, this provides an ideal place for the creation and spread of misinformation. Fake news can be extremely influential and has the ability to spread exceedingly fast, this can be difficult to rectify and may have ever lasting implications. People can rely their reasoning on what they are exposed to intentionally or unintentionally, establishing their own logic on lies and spreading more chaos, The biggest reason why false information is able to thrive continuously is that humans become a victim of Biased-Truth and as a society we have failed to teach everyone "<u>data and information literacy</u>". Some users of social media tend to be unaware that news, articles, posts are just published in order to influence people's decision. For example, a person believes in a particular idea whether it right or wrong, he/she tend to prove and uphold it with all possible news/Proofs (whether by using random articles from uncredible sites, posts from friends or anything that does agrees with their principles).We humans do not like to hear anything that contradicts what we like/believe this is how we function, One cannot help but favor what they like to hear .In addition, not only foes fake news negatively affect individuals, but it is also harmful to society, ruining the balance of news ecosystem

Echo chamber effect

Information can come from many different sources and perspectives. But when we're only hearing the same perspectives and opinions over and over again, we may be in something called an echo chamber. This may lead to our biased opinion unknowingly

Fundamental theories:

News-related theories: News related theories reveal the possible characteristics of fake news content compared to true news content. For instance, theories have proved that fake news potentially differs from the truth in terms like writing styles ,quality, quantity such as word counts and sentiments expressed, Attributes identified can be used to scale the authenticity of news using its writing style this can be achieved using supervised machine learning techniques.

User-related theories: Here theories investigate the characteristics of users involved in fake news activities, like posting liking, commenting and forwarding .

Existing Solutions:
- Image Analysis
- Natural Language Processing ,Sentiment analysis
- installing browser plug-ins (flags unreliable websites)
- reverse image search
- checking if the URL(or the source) is real
- Cross-check with a fact-checking organization

Inferences from analysis

- Checking for the creator :
  - The websites may have an "about us" link.
  - Doing a background check about the author.
  - Check for the website in other handles(like LinkedIn, twitter)
- Type of message/news:
  - Content.
  - if it can be found in multiple places.
  - if the website is updated regularly.
  - Is it about a fact or just opinion?
- Reason for the news:
  - Motivation.
  - if it is creating profit to anyone.
  - if the author/sources are being paid.
  - if it is sponsored

Perspectives:

Knowledge-based methods:

When detecting fake news from a knowledge-based perspective, one often uses a process known as fact-checking. Fact-checking, initially developed in journalism, aims to assess news authenticity by comparing the knowledge extracted from to-be-verified news content with known facts. In this section, we will discuss the traditional fact-checking (also known as manual fact-checking) and how it can be incorporated into automatic means to detect fake news i.e., automatic fact-checking.

Manual Fact-checking:

Broadly divided into expert-based and crowd-sourced fact-checking.

Expert-based Manual Fact-checking: it relies on domain experts as fact-checkers to verify the given news contents. Expert-based fact-checking is often conducted by a small group of highly credible fact-checkers, is easy to manage, and leads to highly accurate results, but is costly and poorly scales with the increase in the volume of the to-be-checked news contents.

The are many websites which have recently emerged to allow expert-based-checking better sever the public. Websites like PolitiFact (www.politifact.com) can be used as a fact/Standard for comparison ,it provide "the PolitiFact scorecard" which provides the statistics on authenticity distribution of all statements related to a specific topic.

Key features to look on are:
- Title
- Subtitle
- No. of images
- Body of the article
- Content and quality of the image
- Number of Readers
- Date and time when published

Website	Topics Covered	Content Analyzed	Assessment Labels
PolitiFact[3]	American politics	Statements	True; Mostly true; Half true; Mostly false; False; Pants on fire
The Washington Post Fact Checker[4]	American politics	Statements and claims	One pinocchio; Two pinocchio; Three pinocchio; Four pinocchio; The Geppetto checkmark; An upside-down Pinocchio; Verdict pending
FactCheck[5]	American politics	TV ads, debates, speeches, interviews, and news	True; No evidence; False
Snopes[6]	Politics and other social and topical issues	News articles and videos	True; Mostly true; Mixture; Mostly false; False; Unproven; Outdated; Miscaptioned; Correct attribution; Misattributed; Scam; Legend
TruthOrFiction[7]	Politics, religion, nature, aviation, food, medical, etc.	Email rumors	Truth; Fiction; etc.
FullFact[8]	Economy, health, education, crime, immigration, law	Articles	Ambiguity (no clear labels)
HoaxSlayer[9]	Ambiguity	Articles and messages	Hoaxes, scams, malware, bogus warning, fake news, misleading, true, humour, spams, etc.
GossipCop[10]	Hollywood and celebrities	Articles	0-10 scale, where 0 indicates completely fake news and 10 indicates completely true news

Fig. A Comparison among Expert-based Fact-checking Websites

Crowd-sourced Fact-checking Websites: Unlike expert based these websites are still in the early development stage, here users can upload articles, provide ratings for sentences within articles and choose tags that best describe the articles.

Automatic Fact-checking:

Mostly relying on information retrieval, Natural language processing (NLP) and machine learning techniques, as well as on network/graph theory .To review these techniques, a unified standard representation of knowledge is first presented that can be automatically processed by machines and has been widely adopted in related studies.
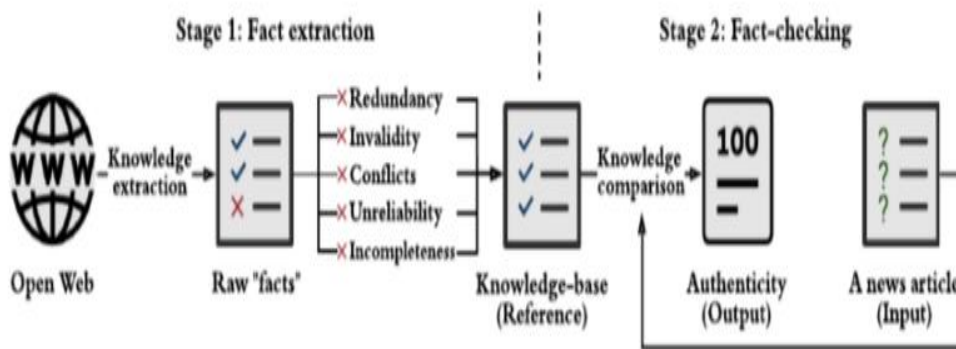
Fig. Automatic News Fact-checking Process

Fact Extraction:
Fact extraction can be said as preparing a dataset extracting information from various open sources and vectorizing, cleaning it in order to avoid redundancy, invalidity, conflicts, unreliability and incompleteness and make it suitable use it to train a model.
•Source: Author or publisher of the news article
•Headline: Short title text that aims to catch the attention of readers and describes the main topic of the article
•Body Text: Main text that elaborates the details of the news story; there is usually a major claim that is specifically highlighted and that shapes the angle of the publisher
•Image/Video: Part of the body content of a news article that provides visual cues to frame the story Based on these raw content attributes, different kinds of feature representations can be built to extract discriminative characteristics of fake news. Typically, the news content we are looking at will mostly be linguistic-based and visual based, described in more detail below.

Style-based methods:
Similar to knowledge-based fake news detection. style-based fake news detection also focuses on analysing the news content. However, knowledge-based methods mainly evaluate the authenticity of the given news, while style-based methods can assess news intention, i.e.,is there an intention to mislead the public or not. The intuition and assumption behind style-based methods is that malicious entities prefer

to write fake news in a "special" style to encourage others to read and convince them to trust.

Deception-oriented stylometric methods capture the deceptive statements or claims from news content. The motivation of deception detection originates from forensic psychology and various forensic tools including Criteria-based Content Analysis and Scientific-based Content. More recently, advanced natural language processing models are used in models like Deep syntax models implemented using probabilistic context free grammars (PCFG), with which sentences can be transformed into rules that describe the syntax structure. Based on the PCFG, different rules can be developed for deception detection, Rhetorical structure theory can be utilized to capture the differences between deceptive and truthful. Deep network models, such as convolutional neural networks (CNN), have also been applied to classify fake news.

Objectivity-oriented approaches capture style signals that can indicate a decreased objectivity of news content and thus the potential to mislead consumers, such as hyper partisan styles and yellow-journalism. Hyper partisan styles represent extreme behavior in favor of a particular political party, which often correlates with a strong motivation to create fake news. Linguistic based features can be applied to detect hyper partisan articles. Yellow-journalism represents those articles that do not contain well-researched news, but instead rely on eye-catching headlines (i.e., clickbait) with a propensity for exaggeration, scare-mongering, etc. Often, news titles will summarize the major viewpoints of the article that the author wants to convey, and thus misleading and deceptive clickbait titles can serve as a good indicator for recognizing fake news articles.

Style-based methods:
When detecting fake news from a propagation-based perspective, one can investigate and utilize the information related to the dissemination of fake news, e.g., how users spread it. Similar to style-based fake news detection, propagation based fake news detection is often formulated as a binary (or multi-label) classification problem as well, however, with a different input. Broadly speaking, the input to a propagation-based method can be either a (I) news cascade, a direct representation

of news propagation, or a (II) self-defined graph, an indirect representation capturing additional information on news propagation.

Proposed Solution

The approach mainly depends on collective intelligence and the "wisdom of crowds", The proposed method mainly focuses on crowdsourcing (rather than professional fact-checkers) to assess the reliability of news websites (rather than individual stories) and adjusting media platform's algorithm such that most authentic/voted contents are mostly visible. This approach is appealing because rating at the website level, rather than focusing on individual stories, does not require ratings to keep pace with the production of false headlines and because using laypeople rather than experts allows large numbers of ratings to be easily acquired. The main idea is to build a "opensource" where all the contents are authorised and especially **transparent** .

How is it different from normal crowdsourcing?
Crowdsourcing mainly focuses on only one source for a particular task that is the crowd/group of people, but here the concept is **blended crowdsourcing** which is a summation of Crowdsourcing and technology ,where man's and machine's computational power can be combined to get better results.
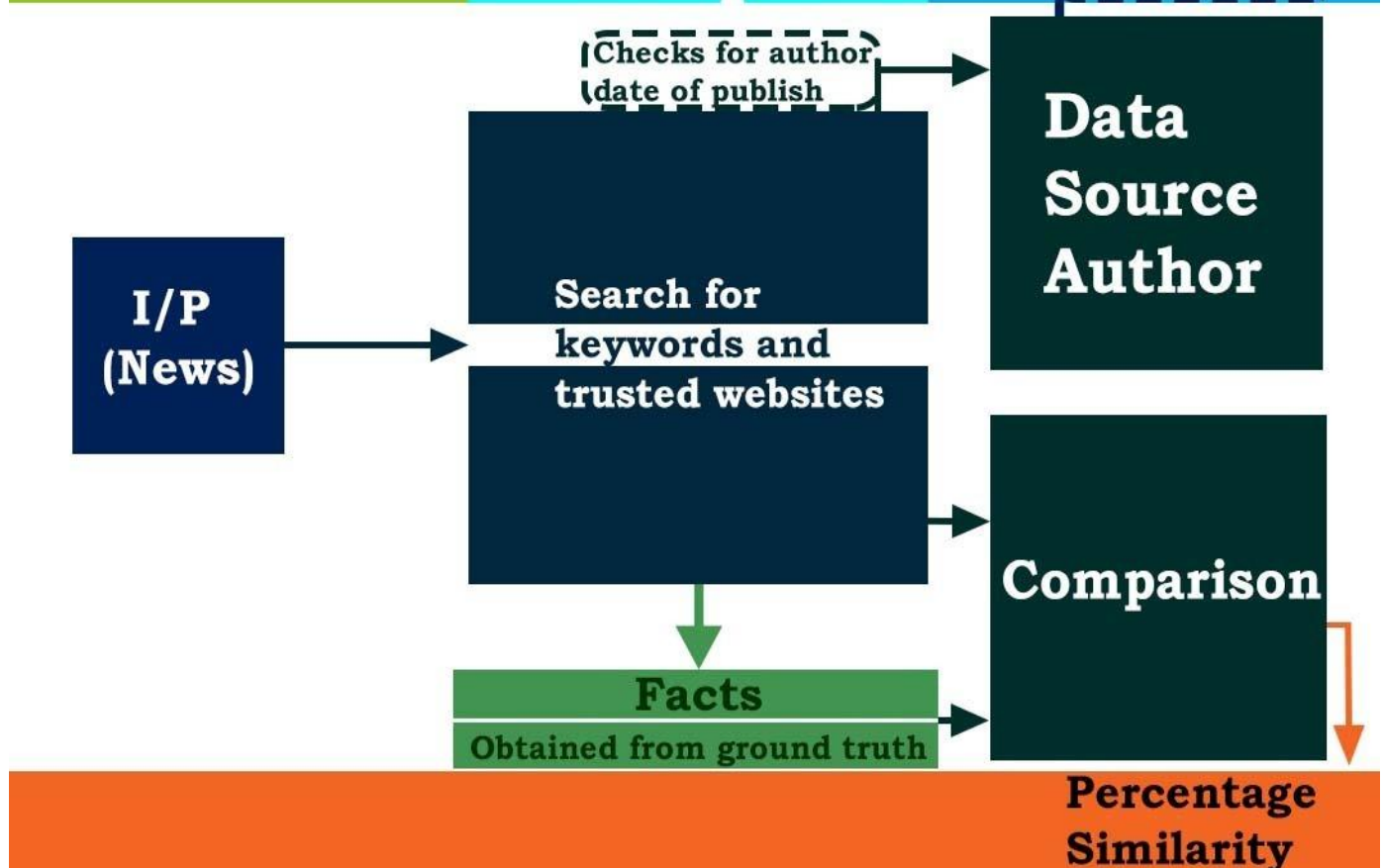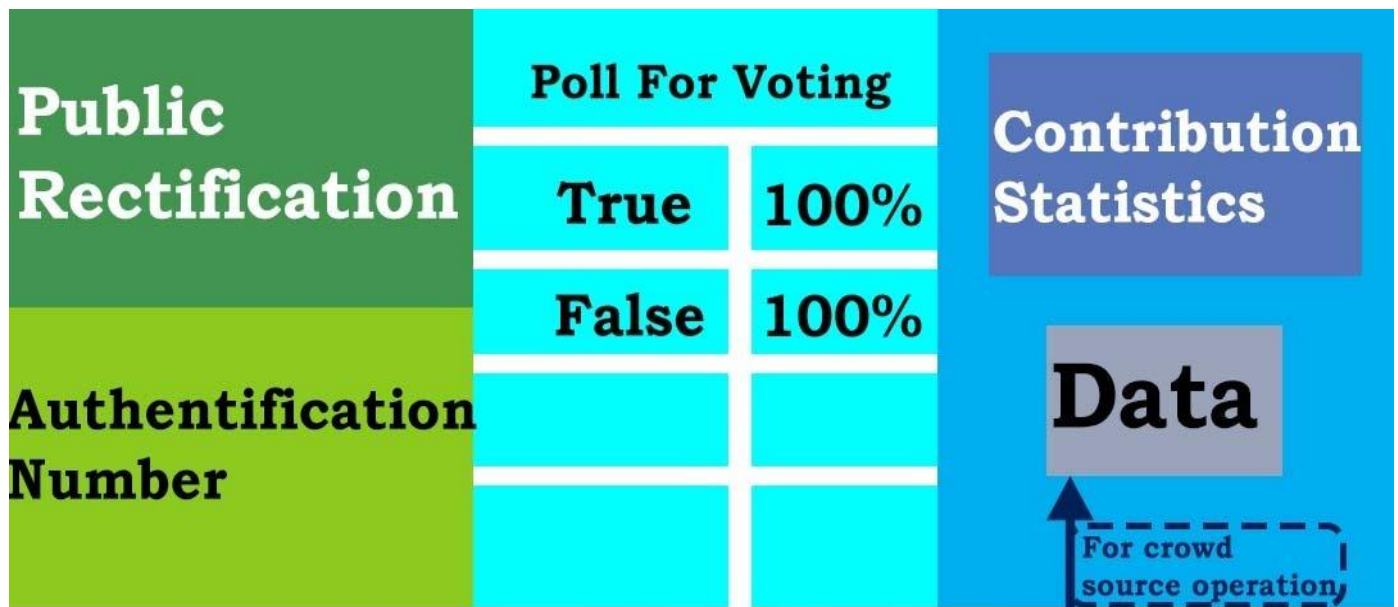
Model

There are multiple clusters used , one of them implements are news/information comparison without any machine learning/Natural language processing techniques

When given an input of news/information system searches web for similar/same news from selected web sources which are authentic and can be considered as standard and later compares both of them and gives output.
Also given a news/article it also gives author, sources, published date which can be used for future study and implementations.
Using a crowdsource for verification and voting along with necessary data and statistics. Users/contributors are verified and checked for their authentication (crowd can provide certain validations if needed).
Displaying contents and news which are most voted by the crowd and has better results in the comparison model.

Public Rectification	Poll For Voting		Contribution Statistics
	True	100%	
	False	100%	**Data**
Authentification Number			

For crowd source operation

Checks for author date of publish

I/P (News)

Search for keywords and trusted websites

Data Source Author

Comparison

Facts
Obtained from ground truth

Percentage Similarity

Conclusion and further study

In this paper, I have proposed a method for providing authentication of news in digital media using a blended crowdsource approach, news can be compared to obtain results, sources and other information about the given article can also be obtained, Crowdsourced information and other data obtained from other algorithms can be used to obtain a final result.

Crowd sourcing has shown to be a powerful technique for overcoming many challenges on data and information processing where current state of the art algorithms are still struggling. Combining algorithms and crowd sourcing can be more efficient in finding solutions for the current problem. However, there are quite few sub challenges to be tackled here, Like *Optimization* for the performance of algorithm along with human contributions, *Considerations* about achieving a higher quality with all the constraints and Making it *effective.*

Ultimately all the strategies would serve its purposes only if one is *information literate*. Every methodology would be of no value if one cannot/doesn't want to be open-minded about issues which we come across in the web pool every day.

References:

https://en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0.pdf
https://arxiv.org/pdf/1708.01967v3.pdf
https://www.merriam-webster.com/dictionary/authentic
https://en.wikipedia.org/wiki/Fake_news
https://www.dictionary.com/browse/authentic
https://edu.gcfglobal.org/en/digital-media-literacy/what-is-an-echo-chamber/1/
https://www.researchgate.net/publication/342762209_A_Survey_of_Fake_News_Fundamental_Theories_Detection_Methods_and_Opportunities
https://www.dqindia.com/whatsapp-message-coronavirus-vaccine-fake-truth/
https://www.stackoverflow.com
https://news.mit.edu/2019/reader-crowdsource-fake-news-0128#:~:text=Then%20consider%20using%20%E2%80%94%20yes%20%E2%80%94%20crowdsourcing,other%20kinds%20of%20online%20misinformation.