CS447 Literature Review: FAKE NEWS DETECTION

RAJAN PATEL,
rajansp2@illinois.edu

December 12, 2022

Abstract

Fake news has become one of the biggest problems today, and the rise of social media has only fueled the fire. This document is a literature review of a few research papers reflecting our progress in tackling the Fake News problem. Detecting fake news is time-consuming and costly as it requires knowledgeable personnel to verify every fact manually. This attribute is one of the main reasons to automate the task of fake news detection based on available information. To better detect fake news, it is crucial to understand different aspects of Fake News - linguistic features, the intent of the author, and its social media impact.

1 Introduction

Fake News refers to the intentional or unintentional passing of false information under the hood of real news. It is observed that politicians bend the truth to gain the trust of people, however, it is not just them. News channels use satire in their headlines to cater to the keenness of the readers, and scammers use hoaxes like "*Tesla inviting people to give bitcoins for double the amount in exchange*" to scam millions of dollars of the readers and so on. Hence, it is crucial to curtailing the spread of fake news and spread awareness about the publisher's intention. This paper aims to review the work done by some researchers in the field by providing an overview of their research studies. This review aims to address a few basic questions about fake news detection; 1. How well do different NLP models perform in detecting fake news? 2. How well do these models perform for languages other than English, 3. How far have we gotten?

2 Background

Fake News is broadly used interchangeably with other related terms such as fact-checking and rumor-detection. However, the fine line separating fake news from others is the term **News**. To better understand these different terms and the amount of effort required in building an automatic fake news detection model, the paper discussed in **Section 3** provides an overview of some of the famous benchmark models including their datasets. Classifying news as fake or real mainly depends on the intent of the author and the quality of the information. Subsequently, we shall move beyond classifying the news as either fake or real. **Section 4** addresses this aspect of fake news detection by exploiting the linguistic features utilized in painting different shades of truth-truth, hoax, satire, and propaganda. Increasing use of social media and the ability to share information without proper checks has given mainstream news media a run to retain customers. To win the competition, people as well as institutions often fall prey to utilizing clickbait headlines. Clickbaits partially contribute to

the notoriety of Fake News. Even though fake news is a worldwide problem, a lot of research work to automate the detection task is carried out for the English language only. The paper discussed in **section 5** acknowledges this problem by providing evidence of clickbait's existence in Hindi Media through a preliminary study. This section raises an alarming concern that even established news media make good use of clickbait against the common belief of they not doing so. With its recent arrival, Transformer Model (**?**) has changed the way NLP studies are undertaken. To show the recent progress, **Section 6** discusses a novel approach that not only detects fake news effectively but also tackles the problem of timely detection before fake news becomes widespread. Finally, **Section 7** provides concluding remarks to the questions raised in the 1 section.

3 A Survey on Natural Language Processing for Fake News Detection (Oshikawa et al., 2018)

3.1 Paper Summary

A lot of studies have been performed to automate the task of fake news detection using NLP. This paper summarizes the progress by surveying significant benchmark methods, introducing the challenges faced, and the pros and cons of each method. Fake news is often confused with other related terms such as fact-checking, rumor detection, stance detection, and sentiment analysis. These terms are discussed briefly in the paper, and then the author tries to classify different research studies according to these terms.

Most researchers formulate the task of fake news detection as either binary or multi-class classification, but some also devise it as a regression task. The difficulty of converting the discrete ground truth labels to numeric scores attributes to the reason for not using a regression approach.

Building a solid fake news detection model requires a quality dataset. Most of the present datasets can be categorized into three categories: claims (a few sentences), entire articles, and Social Networking Services (posts on social media) datasets. Various available datasets are shown in Table 1, and the pros and cons of each are discussed in detail in the paper. As of 2020, it is worth noting that the LIAR and the FEVER are one of the best datasets available within the claims category. Similarly, FakeNewsNet is one of the best datasets for classifying entire articles, and Buzzface, an enriched version of BuzzFeednews, is one of the best datasets in the SNS category.

The collection of the dataset step is always followed by preprocessing the data to ready it for use with different models (discussed below). Preparation usually consists of processing words into features using Word Count (LIWC), Term Frequency (TF), and Term Frequency-Inverse Document Frequency (TF-IDF). Other methods include tokenization, stemming, and generalization. Similarly, word sequences are processed into embedding vectors using pre-trained models such as word2vec and GloVe.

The Methods employed by researchers to classify news broadly fall into three categories: Machine Learning Models, Rhetorical Approach, and Evidence Collecting Models. Machine Learning models consist of Non-neural networks models such as SVM, Naive Bayes Classifiers, LR, and RFCs and neural network models such as RNNs and LSTMs, CNNs, and Multi-Source Multi-Class Fake news Detection framework (MMFD, a combination of CNN and LSTM). The rhetorical approach methods use rhetorical structure theory (RST) to analyze the framework of the text, such as circumstances, evidence, and purpose identifying the coherence and structure of the text in terms of Fake news. Often, RST is combined with the Vector Space Model (VSM) to convert news text into vectors explaining the results obtained by RST. Lastly, the Evidence Collecting models are based on

Recognizing Textual Entailment method that recognizes sentences that reinforce or undermine the input data such as news articles to predict its veracity. However, these methods can only be applied to datasets including evidence, such as FEVER and Emergent datasets containing links supporting the claims.

Claims	**Entire Articles**	**SNS**
PolitiFact[1]	FakeNewsNet[3]	BuzzFeedNews[5]
Channel4.com[2]	BS-Detector[4]	BuzzFace[7]
LIAR		Some-Like-It-Hoax[6]
FEVER		Pheme
Emergent		CredBank

Table 1: Different Available Datasets For Fake News Detection

Furthermore, this paper concentrates on studies carried out on the three datasets: LIAR, FEVER, and FAKENEWSNET, and compares the performance based on experiments carried out in the original paper. As expected, CNN models perform better than non-neural network models, and the LSTM model outperforms CNN models because of its ability to retain dependencies longer. Further, adding additional features such as meta-data, annotator's verdict, and social engagements improves the accuracy of models. Bhattacharjee's 2-class label NLP Shallow Deep CNN model, UNC-NLP Nie's Semantic Matching Network, and Della's HC-CB-3 and Deligiannis' GCN are the best performing models on LIAR, FEVER, and the FAKENEWSNET datasets respectively with some models having accuracy over 90%.

The paper further suggests some improvements for future tasks by providing a critique of these common methods. Some of the suggestions are as follows.

- It is not pragmatic to categorize statements into plain true or false categories. This becomes evident by the increase in confidence of annotators in using intermediate labels such as Mostly True, Mostly False, or Mix.
- Current Models on multi-class classification do not focus much on the order of labels when classifying. For example, Mostly True and True are treated as the same in learning methods and this can be improved in future tasks.
- It is inaccurate to establish a publisher's authenticity purely based on the fact that they have posted any fake news. As a matter of fact, on websites such as PolitiFact and GossipCop[8], the majority of publishers flagged as unauthentic have published just one fake article. Conversely, publishers with the absence of fake news articles are considered authentic. Therefore, there should be a score to establish the level of authenticity.

[1] https://www.politifact.com/
[2] https://www.channel4.com/news/factcheck/
[3] https://github.com/KaiDMML/FakeNewsNet
[4] https://github.com/bs-detector/bs-detector
[5] https://github.com/BuzzFeedNews/2016-10-facebook-factcheck
[6] https://github.com/gabll/some-like-it-hoax
[7] https://github.com/gsantia/BuzzFace
[8] https://www.gossipcop.com

- Some datasets just validate the links associated with articles rather than the content itself. This is wrong. We should verify the content of the entire article for the truthfulness and preferably add a truthfulness score to individual statements.

3.2 Critique

This paper performs a sound review of the progress made by different researchers in the field of Fake News Detection using NLP techniques. The author discusses various aspects of fake news from the point of related terms to the pros and cons of different benchmark methods. This paper helps set up future tasks by providing the pros and cons of various methods and datasets. Suggestions like adding an attention mechanism on LSTM to obtain higher accuracy, adding Meta-data and other information to improve the quality of models, using multi-class classification, gathering improved contents of the dataset, etc., to generate better models are provided. Although we may have moved past LSTM Attention models to Transformer-based models, this paper has done a solid survey of NLP techniques before 2020 and given some good suggestions to overcome shortcomings independent of the model's choice.

4 Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking Rashkin et al.

4.1 Paper Summary

Words have considerable power in shaping people's beliefs and opinions((Rashkin et al.)) and thus the problem of fake news in media and political discourse. This paper reviews the language style of news media in the context of political fact-checking and fake news detection. The paper aims to exploit linguistic characteristics of news to distinguish fake news from real ones. Figure 1 shows news content as a function of two parameters -the author's intent and the veracity of the news article. Many times, it is difficult to distinguish news between fake and real territories, and hence, the suggestion of using a 6-point scale (taken from the PolitiFact website) to rate the truthfulness of political fact-checking using available data from the PolitiFact website.

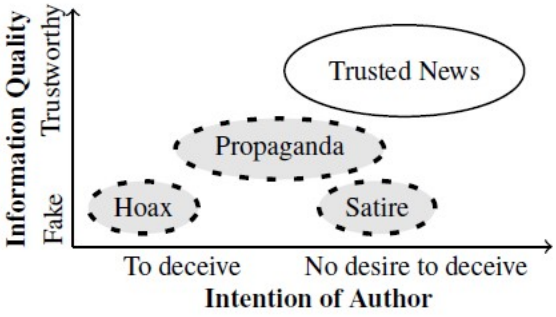

Figure 1: Types of news articles categorized based on their intent and information quality.

Fake News Analysis: Based on the level of truthfulness, fake news can be categorized into three different categories as follows:

- Satire: mimics real news but still cues the reader that it is not meant to be taken seriously.
- Hoax: convinces readers of the validity of a paranoia-fueled story.
- Propaganda: misleads readers so that they believe a particular political/social agenda.

News media often use satire to cater to the humor of the readers, while politicians use propaganda that mixes truth and lies to convince the readers.

The data obtained from previous works in communication theory and stylistic analysis in computational linguistics is tokenized using NLTK((Steven Bird and Loper.)) to compute the per-document count for each lexicon and report averages per article of each type. Lexicons here are the Linguistic Inquiry and Word Count (LIWC), strongly and weakly subjective words with a sentiment lexicon, and a lexicon for hedging and intensifying lexicons based on Wiktionary. These words dramatize, sensationalize, and obscure the information, and their presence in the news content is measured.

It is observed that fake news uses more first-person and second-person pronouns than real news. This can be attributed to the fact that trustworthy editors are more diligent to remove personal language. It is also observed that more modifier words (words to exaggerate) are used in fake news while comparative words, assertive words, words from *hear* category, and numbers are more observed in real news.

The task of news reliability prediction splits the 23k articles into a training set of 20k articles and a test set of 3k articles. As the articles of the train and test set are from different sources, the model is trained to classify without relying on news sources. N-gram (up to tri-gram) tf-idf feature vectors are used for training on the Max-Entropy classifier with L2 regularization. 20% of the training samples are used to develop an in-domain set, and the model achieved a 65% F1 score against the out-of-domain test set.

Data	Sources	Random	MaxEnt
Dev	in-domain	0.26	0.91
Test	out-of-domain	0.26	0.65

Table 2: F1 scores of 4-way classification of news reliability.

50 highest weighted n-gram features were examined for each class- hoax, trusted, satire, and propaganda, with the MaxEnt classifier.

Predicting Truthfulness: Truthfulness is rated on a 6-point basis scale- True, Mostly True, Half True, Mostly False, False, and Pants-on-fire(absurdly false). This scale allows for identifying the facts between completely true and completely false. Out of 4.366 labeled statements taken from the PolitiFact website, most of them are labeled as neither absolutely true nor false by the website on a 6-points scale. For the task of truth prediction, the samples are split into three categories- 2,575 samples of the training set, 712 of samples the development set, and 1,074 samples of the test set trained on an LSTM model taking a sequence of words as input.

1. Evaluation Metrics:
    - F1-score
2. Baseline models with two variants of inputs - one with word tf-idf and another with LIWC concatenated to tf-idf vectors

- MaxEntropy
- Naive Bayes

3. Model
    - Model: Theano and Keras-based LSTM
    - Embedding Layer: GLOVE with 100-dim embedding size
    - Input: a sequence of words as input to LSTM is concatenated with LIWC features before undergoing the activation
    - Output: Binary and Multi-class classification labels.
4. Hyperparameters:
    - hidden state: 300-dim
    - Batch size: 64
    - Epochs: 10
    - Optimizer: Adam Optimizer

Results and Discussion: Refer to Tables 4 and 5 for result. As observed in the results, the chosen LSTM model outperforms the other models when using text as inputs, but the performance gain is not much. Conversely, when combining text with LIWC vector, other models perform better. Also, the LSTM model doesn't outperform other models in the test set. Also, it is evident that adding LIWC information to the text as the input increases the performance for the Naive Bayes model but not for the MaxEnt model or the LSTM model.

The authors have discussed good relevant work in the field of *deception detection* and *fact-checking fake news detection* which I chose to skip in the summary here but suggest having a look at if interested.

4.2 Critique

The good thing about the paper is that the authors took the challenging task of understanding the linguistic style of fake/deceptive news. They acknowledged the fact that most of the time, the news is not just black and white, but rather comes in shades of grey. The overall performance of neither model is in the satisfactory range which cues that the task is a challenging one and needs more effort. A lower F1 score in a multi-class classification could be a result of 6 label classifications- a suggestion would be to reduce the number of labels and see the performance of the models. Questions regarding the methodology such as any studies performed to idealize hyperparameters, and any comparison with other majority baseline models are raised. Also, LSTM being a neural network model, requires a larger input size for training, and results improve with larger training sets. I wonder if the poor performance of the LSTM model is because of the smaller train set and if increasing the train size will better the performance. Overall, this piece of work is important as it lays the road to utilizing the lexical characteristics of information along with its statistical data to determine the truthfulness of political news.

5 Clickbait in Hindi News Media: A Preliminary Study Kaushal and Vemuri

5.1 Paper Summary

The advent of social media caused traditional trusted news media to compete with a lot of new unorthodox news aggregators. As a result of this competition, a lot of news headlines are used as clickbaits - sensationalized and made catchy, to attract news readers. This paper discusses the impact of clickbait on news media, analyzes the news in terms of clickbait and certain linguistic features, and extends its application to the Hindi-language news media.

The data for this preliminary study was prepared by tracking the tweets from established 5 Hindi news sources for a period of 4 months(May 2020 to Aug 2020). The data were cleaned to remove instances of polls, graphics (images and videos), cartoons, mentions, and hashtags at the end of the tweets to promote them on the platform. Instances of hashtags and mentions were not removed when they are an integral part of the headlines. Tweets' interactions like retweets and favorites were also tracked to understand the impact of clickbaits on readers' behavior. With the help of 7 independent annotators, news were annotated on a 5-point scale measuring the intensities of clickbaits (0 = not click-bait to 1 = strongly click-bait). These obtained click-bait scores are compared against tweets' interaction on Twitter, word count, and POS (part of speech) tags to find the correlation between them. A Hidden Markov Model-based POS tagger was used following the Viterbi Algorithm (Ekbal et al., 2007). To isolate POS tags' correlation with clickbait score, POS tags were normalized per word count. The D'Agostino-Pearson Test was conducted on all distributions in the dataset.

Results and Discussion: The average clickbait score for the 5 Hindi news sources was 0.433 with an std. dev. of 0.30. It is found that out of the 100 most retweeted news headlines, 21 tweets had a high clickbait score ≥ 0.75 whereas 36 tweets had a score ≤ 0.25. It was observed that clickbait scores had a positive correlation with the tweet's interaction parameters; replies(0.25), retweets(0.19), and favorites (0.18), negative correlation with word count(-0.39). A mixed correlation was observed between clickbait score and POS tags; WQ(0.44), VAUX(0.18), NNP(0.17), DEM(-0.21), INJ(-0.18), SYM(-0.25) and no significant correlations with other POS tags.

As seen above, 21% of mainstream Hindi media used clickbait. A similar aggravated trend is seen for mainstream English media; up to 47.56% of social media posts by mainstream English media were clickbaits (Rony et al., 2017).

5.2 Critique

Not many researchers work on languages other than English, and this paper tries to extend that work to the Hindi Language by reviewing the extent of clickbaits in Hindi News Media. The preliminary research is very limited to only seven annotators ranking clickbait scores and utilizing only the top 20 tweets from each of 5 news sources (total of 100) to check the existence of clickbaits in Hindi News. Putting aside the limitations, this paper has set a good starting point to understand the extents of clickbait that can be extended to languages beyond Hindi. An interesting outcome of this study was that shorter news is more likely to be used as clickbaits for the Hindi Language, which is contrary to the observation for the English language (Chakraborty et al., 2016). This outcome needs to be studied further.

6 Automatic Fake News Detection in Political Platforms – A Transformer-based Approach Raza

6.1 Paper Summary

As the title hints, this paper focuses on fake news detection in political news using a transformer-based approach. The author proposes a novel approach utilizing social context with a modified transformer architecture to detect fake news in the early stage before it is widespread. Transformer (Vaswani et al.) is the latest state-of-the-art model gaining popularity in different tasks of NLP such as text classification and detection methods. It is a self-sufficient two-step architecture that uses the whole lexical data as one input, understands sequential relations between words, and outputs future behavior based on past inputs.

News or information consists of a hierarchy of three stakeholders - The publisher, the news content, and the users; the information travel in the same order as noted. Fake news is made sensational and hence spreads faster than real news, making it more critical to curbing them at the earliest stage.

The content of fake news is often unclear and deceptive, with lots of capital letters, punctuations, or emotional words. Whenever these patterns alone are not sufficient in determining fake news accurately, other features like social context and temporal information can be used. Social context (Shu et al.) refers to users' interactions like, share, subscribe, comment, follow, etc. Temporal information relates to time-related information. The contents of a news item are news ID or news title and other auxiliary items like news body, source, author, time of publication (temporal information), etc. This paper considers both the news context (main information referred to as headline and side information) and the social context (tweets), and temporal information to detect fake news.

The contributions of the papers are summarized as follows:

- A novel Transformer model that considers news content and associated side information for the task of fake news detection.
- Incorporated side information in the main model. In addition to the lexical data, non-lexical data such as words and categorical data is also modeled. Used a multi-head attention mechanism to attend to different parts of the information noted above.
- Utilized position encoding (Devlin et al.) representing words' order (value of word and its temporal position in a sentence) in the Transformer model to achieve the goal of early fake news detection.

The working system is tested against real-world data consisting of news articles from a variety of sources and using social context from Twitter. Ablation studies are also performed showing the relevance of both news content and social content in determining fake news patterns. It is deemed that more the inclusion of information the better the fake news detection task will be.

The task of fake news is considered both a binary classification (fake or real) and a multi-class classification (fake, mixed, or real). The novel transformer-based model, **Faker**, is a modified pre-trained Bidirectional Encoder Representations from Transformer (BERT) model to also include side information in addition to the main information. Each news item N is represented by its headline and side information, SI (aux news content, social context, and temporal information). Thus, any information is represented by a sequence of words on $N = \{n_1, n_2..., n_l\}$ accompanied by side information $SI = \{s_1, s_2..., s_l\}$, where l is the length. Side information can be lexical or non-lexical whereas main information is just lexical.

The first layer is the embedding layer, which takes input as a sequence of words consisting of N and SI. The [CLS] token is added at the start of the sequence and later used for the class label prediction. The token and segment embedding of the BERT model is used to generate syntax and semantics representation of each word. Position encoding of words is used to capture the temporal information in sequences. The timestamp of a news publication determines the position value of each word.

The output from the first layer is then fed to the first layer of the twelve layers of the encoder block. The output of the encoder block is news, $\tilde{N} = \{\tilde{n}_1, \tilde{n}_2 ..., \tilde{n}_l\}$ and side information, $S\tilde{I} = \{s\tilde{i}_1, s\tilde{i}_2 ..., s\tilde{i}_l\}$. Each of these Vectors $\tilde{\text{N}}$ and $\text{S}\tilde{\text{I}}$ is then passed to a **Fusion Block** where each piece of information is represented with a token - x for the textual word, nu for the numeric word, and c for the categorical word. The gating mechanism is performed here on the non-lexical data (nu and c) to produce a new non-lexical vector h as

$$h = g_c \circ (W_c c) + g_n u \circ (W_n unu) + b_h$$

where g_c and g_{nu} are gating vector, W denotes weight matrices, b denotes bias vector, and $\circ$ is element-wise multiplication. The non-lexical feature i is fused with x using an activation function R and then it goes into h. g_i, gating vectors are defined as

$$g_i = R(W_{gi}[i||x] + b_i)$$

The obtained h is combined with lexical vector X by weighted summation to get the fused sequence m as:

$$m = x + \alpha h$$

where x is a text feature and α is a normalizing vector to dampen the magnitude of h within a range. The α is obtained as

$$\alpha = min(\frac{||x||_2}{||h||_2} * \beta, 1)$$

where the $||x||_2$ and $||h||_2$ denote the l_2 norms of x and h, and hyperparameter β is selected during the validation process. Attention is applied to lexical and non-lexical vectors to form the final fused representation $\bar{n}$. Similarly, output from each fusion block is calculated for each word $\bar{n}_i$. The new sequence $\bar{N} = \{\bar{n}_{CLS}, \bar{n}_1, \bar{n}_2, ..., \bar{n}_l\}$ is then fed as input to the next Encoder block. A similar process is followed to get the output at the end of the Second Encoder block $\bar{\bar{N}} = \{\bar{\bar{n}}_{CLS}, \bar{\bar{n}}_1, \bar{\bar{n}}_2, ..., \bar{\bar{n}}_l\}$. Again, the first token, [CLS], is important to determine the class label-real or fake.

Dataset and Experiment: With the aim to focus on political news, the author considered only the events related to the 2020 US Presidential elections from a relatively broad set NELA-GT-2020 dataset. The resulting parsed dataset included 294,504 events related to the 2020 US Election across 403 sources between Jan 1st, 2020, and Dec 31st, 2020. The ground truth labels are taken from Media Bias/Fact Check website. Over 400,000 embedded Tweets found in news articles make up the social context of the dataset.

News articles without any labels are sorted based on Article IDs and timestamps. Authors used distant supervision(Mintz et al.) to assign labels to each news story. Per suggestions in the NELA-GT-18 paper (Norregaard et al.), weak distant labels may be imprecise and partial but can be used to create a robust predictive model as reflected in the results of recent work(Horne et al.).

The under-sampling technique is used to handle data imbalance caused post-labeling process (around 37k fake, 12.5k real, and 32k mixed labels) by removing records from the majority class.

Feature	Description	Format
Article ID*	Article identifier	Integer
News title	Headline of news	Text
News source *	News Source (e.g., CNN, TheOnion)	Categorical
News content *	News Body	Text
Author *	Author of article	Categorical
URL *	URL of the article	Text
Publication timestamp*	Publication time as UNIX timestamp	Integer
Tweet ID *	ID of tweet	Integer
Embedded tweet*	Raw data from tweets (on news)	Text

Table 3: Dataset features, * is side information (Raza)

1. Evaluation Metrics:
    - Accuracy **ACC**,
    - precision **Prec**,
    - recall **Rec** and F1-score **F1**, and
    - area under curve **AUC**
2. Baseline Models: A broad range of selected models are as follows:
    - Fake-news detection methods - TriFN, Declare, Grover
    - Transformer-based methods - BERT, GPT-2, VGCN-BERT
    - Other Methods - SVM, DeepWalk
3. Hyperparameters:
    - max length of news and Tweets: 500 words; padded and truncated accordingly
    - Dimensionality: 768
    - Batch size: 8
    - Dropout rate: 0.25
    - Epochs: 10
    - Learning rate: 1e-3
    - Optimizer: Adam Optimizer

Results and Discussion: Refer to Table 6 for results. Overall, it is observed that the proposed Faker model outperforms the baseline models in both the tasks of Binary Classification and Multi-Class Classification. Also, it is noted that the best-performing baseline model is the **TriFN** model which is performing at 85% of the proposed model.

Authors used different sampling ratios, $\theta\varepsilon\{0.2, 0.4, 0.6, 0.8, 1.0\}$, of the training samples and compared their results with other baseline models. It is found that the **Faker** model consistently outperforms other baseline models by at least 5-30% and higher in inferring binary labels and multi-class labels respectively. The authors also discussed the confusion matrix obtained by running

tests on 4000 samples for Binary classification. It is observed that Faker model has high accuracy (96.3%), high precision (94.8%) and high recall (98.81%).

For checking the effectiveness of early detection of fake news, authors deemed that the test time shall be within a few minutes from the time of news creation. Further, to check the model's efficiency in detecting news in a delayed setting, the authors decided on a range of time up to 120 minutes from the time of generation. As expected, it is observed that all the models do a better job of fake news detection with more time (except for the SVM model) which shows that more data helps in better classification. Furthermore, it is also observed that the **Faker** model consistently outperforms other models in all time ranges. Authors attribute this good performance to the position-aware mechanism which learns hidden patterns from the sequence of news data and tweets to classify the news articles. Subsequently, the model pays more attention to the data that reflect the truthfulness of news articles with respect to temporal patterns.

Authors also performed ablation studies to determine the importance of each portion of the data-headline, social context, news content body, and news source. It is seen that the social context on its own is not a good parameter to classify fake news but definitely adds value to the data when used with news content as observed in results with and without social context. Further, the news content body plays a more important role in detecting fake news than the headline itself as seen from the comparison of results from the Faker model without body content and the Faker model without a headline. This reflects that the body content has more information than the headline. The authors also explain that the selected hyperparameters shown in section **Methodology** gave them the best model.

The authors have discussed good relevant work in the field which I chose to skip in the summary here but suggest having a look at if interested. The paper finished with possible future scope based on recommendations received. They are

1. Include more social context such as friends' network, propagation paths, and implicit users' feedback.

2. Another recommendation is to combat data and concept drift.

3. Data labeling scheme can be investigated because of the possibility of incorrectly labeled data leading to data biases.

6.2 Critique

The biggest strength of this paper is that the author takes on the challenging task of timely detection of fake news and succeeds in obtaining a decent score compared to other available models. Other pros include a well-documentation with precisely all the information presented to replicate it. The author went through a series of iterations to get the best hyperparameters. Like Rashkin et al., the author of this paper also treated the truthfulness of news as a multi-class label. The strength of the FAKER model is that it is based on the Transformers Model (a very recent development in AI), and many researchers would be curious to further the studies carried out in this paper. This paper bolsters the claim made by other researchers that additional information improves the performance of the model in detecting fake news. Both the direction laid out by the author and the performance of the Faker model are promising.

7 Discussion

Fake news certainly has become one of the biggest challenges with the rise of social media. There are multiple assets under the umbrella of fake news- hoaxes, satire, or propaganda, and the information can be divided into categories of truth, lies, and mixed information. There are multiple ways to distinguish fake news from real ones - from basic ones like trusted sources and authors to more advanced ones like exploiting the linguistic style of news content, and social context. The term Fake News Detection became widespread somewhere around the mid-2010s, particularly around the 2015 US elections. Fake news detection can be seen as a two-part problem- one is detecting information as fake and the second is timely detection for critical matters. A good progress has been achieved in the first part of binary classification of fake news, but not in multi-class classification of fake news. This is because current models do not account for distance between adjacent labels (false vs mostly false) when learning. The second part is the most challenging part of detecting fake news in the early stage with minimal features and curtailing its propagation before it becomes widespread.

Researchers have moved past RNN and LSTM to more advanced language models and methods like BERT, GPT2, attention, etc. Section 6 discusses one of the latest models in the field by providing a novel approach to fake news detection task and also attempts to detect fake news timely. Fake news is a widespread problem, but very little progress is achieved for global languages other than English. This is evident from the shortage of research papers in the field for the Hindi language. Section 5 briefly discusses similar traits of little progress are observed for fake news detection in other languages. Hence, it is safe to say that significant efforts are required for languages other than English.

8 Conclusion

This paper aims to check the progress of automatic fake news detection and automatic detection in the Hindi language. Upon reviewing some of the recent research works in the field, it is evident that fake news detection is a challenging task, and timely detection further adds to it. This challenge is due to the lack of a satisfactory training dataset possessing ground truth labels because verifying truth on a mass scale is both time-consuming and costly. As expected, neural network-based models outperform machine learning-based statistical models. We have reached a point where neural-network models can classify news as either true or false with more than 90% accuracy. The challenge now lies in classifying the news into finer sub-labels - mostly true, mixed, mostly false, etc. Timely detection of fake news yet remains a prominent challenge. Also, it can be said that fake news detection is still in the early phase for languages other than English. Lastly, there are suggestions discussed in paper summaries which will help in future tasks.

References

Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding.

Asif Ekbal, Samiran Mandal, and Sivaji Bandyopadhyay. 2007. Pos tagging using hmm and rule-based chunking.

Benjamin D. Horne, Jeppe Norregaard, and Sibel Adali. Robust fake news detection over time and attack.

Vivek Kaushal and Kavita Vemuri. Clickbait in Hindi news media : A preliminary study.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. Distant supervision for relation extraction without labeled data.

Jeppe Norregaard, Benjamin D. Horne, and Sibel Adali. Nela-gt-2018: A large multi-labelled news dataset for the study of misinformation in news articles.

Ray Oshikawa, Jing Qian, and William Yang Wang. 2018. A survey on natural language processing for fake news detection.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. Truth of varying shades: Analyzing language in fake news and political fact-checking.

Shaina Raza. Automatic fake news detection in political platforms - a transformer-based approach.

Md Main Uddin Rony, Naeemul Hassan, and Mohammad Yousuf. 2017. Diving deep into clickbaits: Who use them to what extents in which topics with what effects? *CoRR*, abs/1703.09400.

Kai Shu, Suhang Wang, and Huan Liu. Beyond news contents: The role of social context for fake news detection.

Ewan Klein Steven Bird and Edward Loper. Natural language processing with python.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need.

Model	2-CLASS		6-CLASS	
	text	+ LIWC	text	+ LIWC
Majority Baseline	.39	-	.06	-
Naive Bayes	.44	.58	.16	.21
MaxEnt	.55	.58	.20	.21
LSTM	.58	.57	.21	.22

Table 4: Model performance on the PolitiFact validation set. (Rashkin et al.)

MODEL	FEATURE	2-CLASS	6-CLASS
Majority Baseline		.39	.06
Naive Bayes	text + LIWC	.56	.17
MaxEnt	text + LIWC	.55	.22
LSTM	text + LIWC	.52	.19

Table 5: Model performance on the Politifact test set.(Rashkin et al.)

Model /Metric	**TriFN**	**Grover**	**Declare**	**BERT**	**VGCN-BERT**	**GPT2**	**SVM**	**DeepWalk**	**Faker**
Binary Classification									
ACC	0.695	0.602	0.579	0.690	0.652	0.602	0.459	0.620	**0.824**
F1	0.660	0.598	0.552	0.612	0.635	0.609	0.468	0.610	**0.768**
AUC	0.698	0.678	0.577	0.619	0.632	0.648	0.430	0.542	**0.804**
Multi-class Classification									
ACC	0.675	0.582	0.559	0.660	0.650	0.582	0.400	0.519	**0.810**
F1	0.640	0.580	0.540	0.591	0.605	0.589	0.456	0.598	**0.750**
AUC	0.680	0.660	0.563	0.601	0.632	0.636	0.420	0.529	**0.780**

Table 6: Results of all models using Binary and Multi-class classification (Raza)