Survey on Fake news and Fake profiles detection techniques

Dr. Minal Moharir 1, BinduShree N 2
1 Associate Professor, R V College of Engineering, Bangalore
2 Student, R V College of Engineering, Bangalore

Abstract

The rapidity and low cost of news sharing has made social media one of the key platforms for people to view and absorb news. The propagation of regular access media channels such as social media streams, news forums and web publications has made it impossible for credible information sources to be established, thereby raising the need for computational resources capable of offering data on the authenticity of online content. However, these social media platforms are also a means of misleading news, which has a negative effect on people and culture. Social media includes online news to be conveyed to the public and online accounts on social networks such as Facebook, Twitter, etc. The situation is critical when fake news is communicated across. Fake accounts are created to access personal information and harm people. To prevent such disaster activities, the techniques are developed to identify such fake news and fake profiles on online social networks. In this paper, four techniques are reviewed and compared in terms of accuracy for detection of fake news and profiles. The techniques are Decision tree, Logistic regression, Neural network and K-Nearest Neighbor (KNN). Among these four algorithms are Logistic regression performs better according to the review.

Keywords: Social media, Online social networks, Fake news and profiles, Decision tree, Logistic regression, Neural network, KNN.

1. INTRODUCTION

With different reasons, such as political concerns, the volume of false news rises to generate popular sentiments by spreading falsified facts. In these days, multiple attempts are being made to improve the false news identification system. The influence and scope of modern social media has been exploited by the false news globally to successfully disseminate information that not only degrades people's trust in the mainstream press and journalism, but manipulates public attitudes and viewpoints. Fake news is an incredible challenge because of the slight distinction between fake news and actual news. Despite Web evolving exponentially, news is being transmitted increasingly quickly and more efficiently. It produces many false news which confuses the perception of the reader. In order to create a stable social media climate, misleading reporting needs to be eliminated and the origins of inaccurate social discussed, so that the origins of misinformation has to be tracked. There can be different reasons why false news is spreading. The first is because of the lack of people's awareness. The readers do not know the accuracy of the media and the reliability of the reporting. It will affect the public adversely. The vast dissemination of fake news, which was reported to manipulate elections and undermine democracy, has become a big global concern. Media, social, cognitive and computer scientists are working to research the diverse factors of the global dissemination of online disinformation and the creation of solutions, whereas social media and search engines continue to use countermeasures.

People connect and create relationships with others through social networks. Google, Facebook, LinkedIn and Twitter are the most widely visited blogs on social networks. Millions of websites draw millions of people and the majority of them have incorporated these websites as major part of their lives. Because a wide range of social networks and an intelligence center are available, this has become a probable channel for attackers to use or attack. Different websites have different ways to prevent these threats, but they cannot be avoided as each day they discover new strategies for targeting. False profile formation on social networks is considered more harmful than any other form of cybercrime. Just before the individual is informed about incorrect development about accounts, this fraud has to be identified. A variety of algorithms and methods for identifying fake profiles have been proposed. This paper reviews the four proposed algorithms for detection of fake news and fake profiles in online social media and network. The four effective techniques discussed here are Decision tree, Logistic
regression, Neural network and KNN. These methods are undergone into comparative analysis and Logistic regression turns out to be better performed with respect to the study.

2. LITERATURE SURVEY

Shuo Yang et. al., [1] investigated how we can spot false news. They treat news facts and the integrity of people as hidden random variables and use social media contributions to define their perspectives on the authentic of reporting. They use a Bayesian network model to capture the reliance on news facts, perceptions of users and reputation of users. In order to simultaneously assess validity and the legitimacy of participants, an optimal Gibbs sampling approach is introduced. The proposed approach is tested using two datasets, and the results show that the suggested algorithm meets the unsupervised criterion.

Veronica Perez-Rosas et. al., [2] concentrated on automatic false information detection in online journalism. They had a double commitment. Initially, they introduced two new data sets covering seven separate news fields for false news identification. Furthermore, they carried out a set of learning’s to build reliable news sensors. However, they carried out statistical research to classify fake news automatically and manually. For the role of false news identification they built computing tools and models. In one dataset, manual and crowd sourcing annotations are collected, and the second is extracted from the internet. Utilizing these datasets, they examined many language properties that mostly occur in false content and create fake news detectors based on linguistic features.

Sanjeev Dhawan et. al., [3] concentrated on Facebook for false profile identification. Facebook is the one of the most popular social media platform, on which one can exchange comments, pictures and videos. However it is hard to see if the new person is real or not. It might be a malicious person. Web includes so many user-accessed third-party apps. When the user wants to use some data from third parties, then the user will authorize the client to access those profiles. Whenever users approve the submission, otherwise they can view personal details, such as the name, email address and list of contacts, etc. Hackers often create these apps and force the user to install this malicious software. The customer has to share his personal information with App and run malicious apps. Hacker uses the sensitive details of the user and places ransomware on the account screen.

This work provides a variety of approaches for identifying unauthorized users and for eliminating false profiles for applications. Each strategy was explored in terms of pros and cons. This was accompanied by the effects of these approaches with respect to implementations and output parameters.

Yasyn Elyusufi et. al., [4] emphasizes on identification of fake social media accounts. Analytical methods can be categorised to recognise fake social media accounts as approaches for the study of profile data. Until evaluating data sets, the model introduced applied data preprocessing techniques. A methodology has been used to define and reduce the non-significant properties of datasets. For data sets of false and legitimate consumers the proposed model was trained independently using supervised computer algorithms. This analysis will inspire us to work with minimal knowledge on social networks and to find ways to better choose from accurate data. In fact, we can try to identify effective strategies when less knowledge is accessible to specific methods in other fields.

Maniraj et. al., [5] has been developed innovative method for detecting counterfeit accounts. They used gradient boost algorithm with a 3 attribute decision tree. Those characteristics include spam commentary, automated behaviour and the rate of participation. In order to identify false accounts accurately, they pooled machine and data science learning. The advancement in the development of fake identities has made existing structures redundant. The variables that are transmitted to the new system are unpredictable.

Muhammad Syahmi Mokhtar et. al., [6] proposed a false news detection model using a Logistic Regression to determine the Logistic Regression's efficiency as a fake news detection model and create a web-based framework that supports news or news URLs. Logistic regression was used in this analysis for the identification of false news. This initiative relates to and applies the framework for model creation. This work is not the similar as designing a particular traditional software system because it focuses on machine learning model creation. Machine learning takes a long time for model preparation and model validation and a decent and good dataset consistency as well.

There are some challenges that are hard to face in creating this false news identification model. As a
result of some practical reasons, this model is not an outstanding model. Even though the model is outstanding, but it can only be validated with excellent precision, which does not provide new fresh results.

Sherry Girgis et. al., [7] designed a cluster was to determine whether or not the news article is incorrect, thus approaching the trouble with the help of RNN theoretical models from the solely deep learning viewpoint. By adding them to the data collection, they shall show the disparity and evaluate the results. In order for researchers to increase learning through deep learning with a broad data set, word modularizing is used for extracting features or signals to differentiate between syntactical word and semantic word relationships.

Smruthi et. al., [8] used a hybrid model to detect the presence of fraudulent accounts on the basis of machine learning algorithms. The evaluation method demonstrated the power of the new system in terms of the precision of the identification of fake profiles. There are several tools available on Twitter to identify fraudulent accounts. As the processing of datasets is a repetitive job on Facebook, it has been done very least to find fake profiles. The research discussed in this work addresses the findings of the identification of false accounts using Facebook capabilities. The output was measured utilizing supervised algorithms of machine learning, and the average precision of 80 per cent was achieved.

3. COMPARATIVE ANALYSIS

The algorithms used for detecting fake news and fake profiles are represented in below figure 1.

**Figure 1: Fake news and fake news detection techniques**

A decision tree is a method for decision-making that utilizes a branch-like decision model and its potential impacts, involving random events, asset costs and value. One approach is to demonstrate an algorithm containing only conditional declaration power. In this case, decision trees take the value that contains more forged accounts, to keep track of the rate of success. The initial tree is performed utilizing a root node and artificial action because the child node together with fake messages remarks as different node. While the root node, spam comments and interaction rate are corresponding nodes the second tree retains artificial operation. Spam messages, artificial events and interaction levels are used as root nodes to form the third tree.

The logistic model is used in mathematics to predict the possibility of a particular entity or an actual occurrence like fail/pass, loss/win, dead/live. This can be expanded to model various types of events like when a picture includes a human, a horse, a lion, etc. Increasing image reference will have a probability of 0 to 1. Logistic Regression has the ability to create calculations in text classifications. Logistic regression demonstrated effective classification efficiency on the basis of current studies. Furthermore, an approach called stancedetection is introduced to progress model output precision. This model, on the basis of research performed, provides excellent precision when building this false news model in a stance identification strategy.

Recurrent Neural Network (RNN) is a type of Neural Network in which the output from previous step is taken as input to the current step. All the inputs and outputs are independent of each other in conventional neural networks, but cases exist where it is required to calculate the next word of a sentence using the previous words. Because of this RNN came into picture that solves this issue with the aid of a Hidden Layer. The chief and most significant feature of RNN is Hidden state, which retain information.

KNN is a non-parametric approach used for regression and classification when identifying patterns. The feedback is the nearest examples of instruction in the practical field of both situations. Feedback, whether KNN is used for regression or classification depends on the outcome. The performance is a class member of k-NN classification. An entity is graded by a vote for its neighbors, which assigns the entity to a most ordinary category of its nearest neighbors. When \( k = 1 \), the entity is literally allocated to the closest neighbor’s group. The performance is the value of
property of the object in k-NN regression. This sum is the combination of the nearest k values.

The detailed summary of four algorithms is given in table 1 below. The four algorithms are compared in terms of accuracy as shown in table 2. The accuracy is obtained from the review perspective of techniques for different datasets.

Table 2: Comparison of techniques

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision tree</td>
<td>0.64</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>0.69</td>
</tr>
<tr>
<td>Neural network</td>
<td>0.66</td>
</tr>
<tr>
<td>KNN</td>
<td>0.63</td>
</tr>
</tbody>
</table>

4. CONCLUSION

The proficient machine learning techniques are analyzed for detection of fake news and fake profiles on social networks such as Facebook, Tweeter, etc. The survey outcomes Logistic regression is better in classification and detection. The table 2 shows the research values for accuracy. Decision tree and Neural network performs better in some case of datasets, where as KNN gives better results when combined with some other classification algorithms. This survey work highlights that Logistic regression achieves better performance compared to other techniques.

Table 1: Description of techniques

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Applications</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision tree</td>
<td>A decision tree is a block diagram framework with an &quot;evaluation&quot; on an attribute for every inner node, and every node reflects the consequence of the test, and every node presents a class name. The root-to-leaf routes are grouping rules.</td>
<td>Decision trees find applications in situations where the actions are out of control for the decision maker.</td>
<td>Classification and Clustering</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>Logistic regression is a mathematical model that uses a logistic equation in its base form, but there are also more complex extensions. For regression analysis, logistic regression measures the logistic function parameters.</td>
<td>Logistic regression is being utilized in a variety of areas, involving medical fields, machine learning, social sciences etc.</td>
<td>Classification and Regression</td>
</tr>
<tr>
<td>Neural network</td>
<td>A neural network is a system or chain of neurons or an artificial neural network comprising of simulated neurons or nodes in the current context.</td>
<td>These are used in data compression and processing, decision making, novelty detection etc.</td>
<td>Classification</td>
</tr>
<tr>
<td>KNN</td>
<td>The KNN model uses the functional similarity to forecast the new data points values, meaning that the new data point is allocated a value dependent on the closeness of the training set points.</td>
<td>This has applications in the field of genetics, data compression, trade and industry forecasting.</td>
<td>Classification and Regression</td>
</tr>
</tbody>
</table>
5. REFERENCES


