Project Report

Project Trainee: B Sachin Selvan

Paper Title: Fake News Detection on Social Media: A Data Mining Perspective

(<http://www.kdd.org/exploration_files/19-1-Article2.pdf>)

INTRODUCTION

As much of our lives are spent interacting online through social media platforms, more and more people are seeking out and consuming news from social media platforms rather than traditional media sources. The reason for this switch in  consumer behavior is inevitable given the digital age we are in. It is far quicker and less expensive to produce news on social media than traditional news media such as newspapers or television; and it is easier to share, comment on, and discuss the news with friends or other readers on social media.

Though fake news is not a new issue to carry out propaganda; it is purposefully written to mislead readers, making it difficult to detect purely based on news content alone. This problem can be dealt with by considering two aspects: characterization and detection.

FAKE NEWS CHARACTERIZATION

This section deals with the basic social and psychological theories regarding fake news. The article also discusses the various aspects of fake news in both traditional media as well as social media.

Definition of Fake News

Fake news is defined as articles that are intentionally and provably false, in order to mislead the users.

Fake News on Traditional News Media

Fake news media has evolved over time, from newsprint to radio or television and as recently transcended to online social media. Consumers frequently mistake fake news for real news due to cognitive biases that are part of human nature. Furthermore, it is difficult to correct a misperception once it has been formed.

According to ‘The Fake News Ecosystem’s Social Foundations’, the publisher’s power arises from two perspectives: (i) short-term utility: the motive to maximize profit, which is positively correlated with the number of customers reached; and (ii) long-term utility: ones reputation in terms of news authenticity.

Fake News on Social Media

Malicious Accounts on social media are used to distort online communities and provide misinformation. Trolling behaviors are heavily influenced by people’s moods and the context of online discussions, thus allowing fake news to spread easily. Cyborg users can circulate fake news in a combination that includes automated and human activities. Online communities become segmented and homogenous with limited information due to Echo Chamber Effect(this effect occurs when like-minded people form groups and begin to polarize their views).

FAKE NEWS DETECTION

Based on the characterization of news, the authors investigate the problem statement and proposed approaches for detecting fake news.

Problem Statement

The authors go over the mathematical formulation for detecting fake news on social media. It mainly consists of two parts: the Publisher and the Content. The Content ‘ca’ is made up of a number of attributes which represent the news article (such as the headline, text, image and so on).

The authors define Social News Engagements as a set of tuples E = e to represent the process of news spread over time among n users U = u1, u2,…,un and their corresponding social media posts P = p1, p2,…, pn. Each engagement eit = ui, pi, t represents a user ui spreading a news article an at time t using pi. We set t = N ull if the article a has no engagement yet, and thus ui represents the publisher.

Feature Extraction

Traditional news media primarily relies on News Content to detect fake news, whereas in social media, Social Context information can be used as additional  information to help detect fake news.

News Content Features

News content features can describe the meta data associated with a piece of news. Common linguistic features are frequently used to represent documents in Natural Language Processing (NLP) for a variety of task. To capture the deceptive patterns in writing styles to distinguish fake news, other features such as lying detection features are used. Clarity score, coherence score, diversity score and clustering score come under visual features that help in identifying fake news.

Social Context Features

Social context features are obtained from a client’s social engagements of news on social media platforms. To capture changes in posts over time, unsupervised embedding methods such as recurrent neural network (RNN) are used. Various can be built to extract features. The stance network is a network where the nodes represent all of the news tweets (relevant ones) and the edges represent the weights of similarity of views (stances).

Model Construction

The authors go through the details of model construction process for a variety of existing approaches. Based on inputs, existing methods are classified as Content Models and Social Context Models.

News Content Models

Existing approaches can be categorized as Knowledge-based and Style-based. Fake news publishers often have malicious intent to spread misleading information. Deception-oriented textual methods capture the deceptive statements or claims from news content. Partisanship styles represent extreme behavior in a particular political party, which often correlates with a strong motivation to create fake news.

Social Context Models

Social context models include relevant user social engagements, capturing auxiliary information from a variety of perspectives. To infer the truthfulness of original news articles, stance-based approaches use users’ views from relevant post contents. The views (stance) expressed by users in their posts can be represented explicitly or implicitly.

ASSESSING DETECTION EFFICACY

In this section, the authors discuss how to assess the performance of algorithms for fake news detection, focusing on the available datasets and evaluation metrics for the task.

Datasets

Online news can be collected from different sources, such as news agency homepages, search engines, and social media websites. However, manually determining the truthfulness of news is a challenging task. Some publicly available datasets are listed below:

* LIAR: it includes mostly short statements, rather than the entire news content. The dataset is collected from fact-checking website PolitiFact through its API. It includes 12,836 human-labeled short statements
* CRED-BANK: was originally collected for tweet credibility assessment, this is a large scale crowd sourced dataset of approximately 60 million tweets that cover 96 days starting from October 2015.
* BuzzFeedNews : This dataset comprises a complete sample of news published in Facebook from 9 news agencies over a week close to the 2016 U.S. election from September 19 to 23 and September 26 and 27. Every post and article is fact-checked by 5 BuzzFeed journalists. It contains 1,627 articles–826 mainstream, 356 left-wing, and 545 right-wing articles.

Evaluation Metrics

Various evaluation metrics have been used to assess the performance of machine learning algorithms for the detection of fake news. Precision measures the fraction of all detected fake news that is annotated as fake news. Higher the value for Precision, Recall, F1 and Accuracy, better is the performance. The Receiver Operating Characteristics (ROC) curve compares the classifier performance by examining the trade-off between the False Positive Rate (FPR) and the True Positive Rate (TPR). We draw the ROC cure by plotting the FPR on the x-axis and the TPR on the y-axis.

* True Positive (TP): when predicted fake news pieces are actually annotated as fake news.
* True Negative (TN): when predicted true news pieces are actually annotated as true news.
* False Negative (FN): when predicted true news pieces are actually annotated as fake news.
* False Positive (FP): when predicted fake news pieces are actually annotated as true news.

By formulating this as a classification problem, the following metrics are defined as,

* Precision = |T P| / |T P| + |F P|
* Recall = |T P| / |T P| + |F N|
* F1 = 2 · Precision · Recall / Precision + Recall
* Accuracy = |T P| + |T N| / |T P| + |T N| + |F P| + |F N|

OPEN ISSUES AND FUTURE RESEARCH

Fake news detection on social media is an emerging research area. The authors aim to point out research directions from a data mining perspective. The research directions are divided into four categories: Data-oriented, Feature-oriented, Model-oriented and Application-oriented.

CONCLUSION

With the growing popularity of social media, an increasing number of people consume news through social media rather than traditional news media. Social media has also been used to spread fake news, which has serious consequences for both individual users and the society. The authors investigated the fake news problem by reviewing existing literature in two stages: characterization and detection.