Fake News Detection on Social Media: A Data Mining Perspective

# Kai Shu; Amy Sliva; Suhang Wang et al.

## 2017

We present the details of mathematical formulation of fake news detection on social media

# Scholarcy Highlights

* 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 media rather than traditional news organizations
* We present the details of mathematical formulation of fake news detection on social media
* Note that few papers exist in the literature that detect fake news using social context features. Because we believe this is a critical aspect of successful fake news detection, we introduce a set of common features utilized in similar research areas, such as rumor veracity classification on social media
* We focus on the available datasets and evaluation metrics for this task
* Accuracy = |T P | + |T N | (5) |T P | + |T N | + |F P | + |F N |. These metrics are commonly used in the machine learning community and enable us to evaluate the performance of a classifier from different perspectives
* We introduced the basic concepts and principles of fake news in both traditional media and social media

# Scholarcy Summary

## Introduction

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 media rather than traditional news organizations.

The reasons for this change in consumption behaviors are inherent in the nature of these social media platforms: (i) it is often more timely and less expensive to consume news on social media compared with traditional news media, such as newspapers or television; and (ii) it is easier to further share, comment on, and discuss the news with friends or other readers on social media.

Given the prevalence of this new phenomenon, “Fake news” was even named the word of the year by the Macquarie dictionary in 2016

## Objectives

We aim to point out the differences between these areas and fake news detection by briefly explaining the task goals and highlighting some popular methods.

Fake news detection on social media is a newly emerging research area, so we aim to point out promising research directions from a data mining perspective

## Results

Evaluation MetricsTo evaluate the performance of algorithms for fake news detection problem, various evaluation metrics have been used.

By formulating this as a classification problem, we can define following metrics, Precision = |T P | (2) |T P | + |F P |.

F 1 = 2 · P recision · Recall (4) P recision + Recall Accuracy = |T P | + |T N | (5) |T P | + |T N | + |F P | + |F N |.

These metrics are commonly used in the machine learning community and enable us to evaluate the performance of a classifier from different perspectives.

Precision measures the fraction of all detected fake news that are annotated as fake news, addressing the important problem of identifying which news is fake.

Because fake news datasets are often skewed, a high precision can be achieved by making fewer positive

## Conclusion

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media.

Social media has been used to spread fake news, which has strong negative impacts on individual users and broader society.

We explored the fake news problem by reviewing existing literature in two phases: characterization and detection.

We introduced the basic concepts and principles of fake news in both traditional media and social media.

We reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction.

We further discussed the datasets, evaluation metrics, and promising future directions in fake news detection research and expand the field to other applications