A Framework for fake news detection on Social Media: A Probabilistic Reasoning Approach

Presenter:

Md. Samiullah,

Research Officer, Murdoch Children Research Institute

Problem Definition

Challenges

- Increased consumption of fake news through social media
- Difficult to identify solely by humans
- Human-Technology collaboration is needed

Research Gap

- Very few studies exist using Bayesian Network for fake news identification
- Most of the existing research are based on a single topic as user profile or social context or new content
- A framework of Bayesian Network is not yet available

Solution

- A BN with independent modules of the Bayesian network will give the combined effect of
  - user profile,
  - social content and
  - news content features
- The BN framework is modifiable and extendable with further findings or expert elicitation.

Literature

- Fake news detection using naive Bayes classifier with only support count [1]
- The correlation between user profiles and fake/real news [2]
- Various features for automatic detection of fake news presented [3]
- AI and ML techniques are used to implement binary classification of different news content as well as to verify credibility of some news website [4]
- A comprehensive overview of negative impact of online fake news with features of users, content, and context for classifying fake news. [5]
- A new set of features and the prediction performance measure for automatic detection of fake news. [6]

Literature …

- A comprehensive survey of fake news detection methods along with the identification of datasets and various artificial intelligent methods [7]
- A probabilistic graphical model to automatically infer true records and source quality without any supervision. [8]
- The social and computer science research regarding belief in fake news and the mechanism by which it spreads. [9]
- Assessing the credibility of a given set of tweets. Specifically, analysis of microblog postings related to “trending” topics, and classify them as credible or not credible, based on features extracted from them. [10]

Limitations of existing studies

- Deterministic approaches
  - Most of the studies use ML models to classify a news as true or false, rather than providing a probabilistic reasoning.
  - The studies on probabilistic approaches mainly use Naive Bayes algorithm.
- Scope for adding new information
  - In case of the previous studies, it is difficult to add new information or existing knowledge to the proposed approaches.

Objectives

- To detect fake news on social media platforms in order to mitigate the spread of false information.
- To develop a framework for fake news detection that can be
  - Modified or expanded.
  - Enriched and validated by experts
- To create scope for incorporating expert elicitation or existing knowledge in future.

A probabilistic framework

- Query based framework
    - The framework follows the approach of asking questions like humans.
    - Several questions can be asked e.g., Who is the source of the news? What is the news about etc.
    - Each question acts as a module consisting of several features.
        - Several states are added for each module determined by these features.
- Bayesian network framework
    - A BN framework is developed based on the “questions” where the news authenticity will be conditionally dependent on the questions.
    - News authenticity is the query node, and the probability of the states is determined based on the provided evidence and the obtained posterior.

3W approach : Introduction

- 3W approach is a solution for detecting fake news based on the proposed probabilistic framework.
    - 3W stands for 3 questions beginning with ‘W’ -> Who, What and When.
    - The three questions will be asked about a news in order to determine its credibility
- Why we choose 3W approach ?
    - These three questions are the important parts of a news on which its credibility is dependent.
    - Features to deal with these questions are handy in existing datasets.
    - No other existing approaches has used the combination of the features based on the three questions.

3W approach : Modules

Module	Description
Who? *Who are responsible in spreading the news?*	• The credibility of the user posting information about the news • The credibility of the original or referred URL mentioned in the post
What? *What is in the news? What are people talking about the news?*	• News post content • The credibility of the original or referred URL mentioned in the post • User sentiments
When? *Temporal information about the news*	• Contemporary information from other posts in the same social media • Contemporary external knowledge

Table 1 : Modules for 3W approach.

3W approach : Feature list

Who module	What module	When module
domain_name *has_author* *website_age* *is_credible_site* *registration_age* *tweet_count* *same_topic_tweet_count* *follower_count* *followee_count* *is_verified* *has_description*	*Similarity with headline* *# of characters* *# of words* *Flesch index* *# of '?'* *# of '!'* *# of uppercase letters* *# of emoticons* *# of external urls* *# of favorites* *# of comments* *# of retweets* *# of hashtags* *tweet_sentiment* *comment_sentiment*	*contains_trending_topics* *contains_trending_hashtags* *contains_google_trending_topics* *contemporary_other_tweets_opinion* *contemporary_other_social_media_opinion*

Table 1 : Features used in 3W approach.

3W approach : Bayesian Network

- Each modules has a node (last level or child or output node)
  - Each node has two states - High (H) and Low (L).

- News authenticity as query node
  - News authenticity have two states - True (T) and Fake (F).
  - The probability of the states is determined by the evidence (states of the three modules.)

- Conditional probability table
  - The conditional probabilities of all combinations of the three states and news authenticity state are estimated from dataset (explained in next slides).

3W approach : Bayesian Network Model

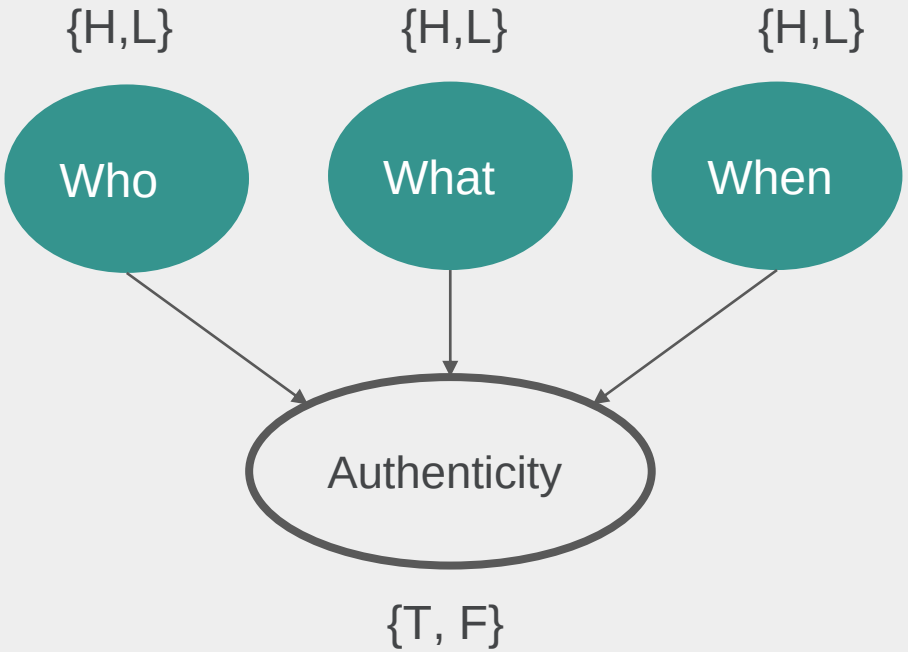

Figure 1 : 3W Bayesian Network (high-level view).

3W approach : Process

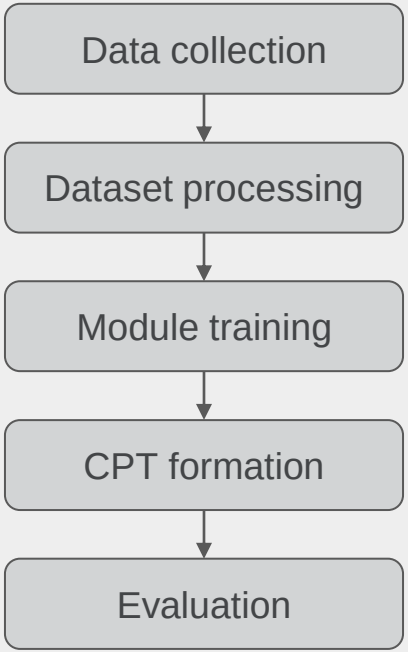

Figure 2 : Procedure steps.

References

1. Granik, Mykhailo, and Volodymyr Mesyura. "Fake news detection using naive Bayes classifier." *2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON)*. IEEE, 2017.
2. Shu, Kai, Suhang Wang, and Huan Liu. "Understanding user profiles on social media for fake news detection." *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*. IEEE, 2018.
3. Reis, Julio CS, et al. "Supervised learning for fake news detection." *IEEE Intelligent Systems* 34.2 (2019): 76-81.
4. Kumar, Vinit, et al. "Fake News Detection using Machine Learning and Natural Language Processing." *2021 International Conference on Technological Advancements and Innovations (ICTAI)*. IEEE, 2021.
5. Zhang, Xichen, and Ali A. Ghorbani. "An overview of online fake news: Characterization, detection, and discussion." *Information Processing & Management* 57.2 (2020): 102025.
6. Reis, Julio CS, et al. "Supervised learning for fake news detection." *IEEE Intelligent Systems* 34.2 (2019): 76-81.
7. Narang, Poonam, and Upasana Sharma. "A Study on Artificial Intelligence Techniques for Fake News Detection." *2021 International Conference on Technological Advancements and Innovations (ICTAI)*. IEEE, 2021
8. Zhao, Bo, et al. "A bayesian approach to discovering truth from conflicting sources for data integration." *arXiv preprint arXiv:1203.0058* (2012).
9. Lazer, David MJ, et al. "The science of fake news." *Science* 359.6380 (2018): 1094-1096.
10. Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. "Information credibility on twitter." *Proceedings of the 20th international conference on World wide web*. 2011.

Any Questions