

Online Fake News Prediction on Social Media Using Machine Learning

Mayuri Shinde¹ Milindkumar Vaidya²

¹PG. Student ²Assistant Professor

^{1,2}Department of Computer Engineering

^{1,2}AVCOE College, Maharashtra, India

Abstract— Basically recommendation system generates based on profiles of users news benefits based on their past historical browsing behavior for such users who connected with the system recently as well as explicitly allowed web history. To produce personalized news recommendations, combine the information filtering mechanism with the user profiles experienced with the current collaborative filtering mechanism. To build a customized news recommendation system, use the popular micro blogging service using Facebook. The proposed research provides online news recommendation using hybrid machine learning algorithm. System initially deals with Natural language Processing (NLP) to extract the features and train the module respectively. The system can recommend the news based on user personalized history, various dataset have been evaluate to measure the performance analysis of system which provides better prediction accuracy than other recommendation systems.

Keywords: Facebook and Twitter, Recommendation for Personalized Data, Recommendation Programs, User Profile

I. INTRODUCTION

There are a number of people having profiles on social media platforms (SMPs) are growing, thus hiding their identity for malicious purposes. Over the last few years, ONLINE social networks have seen both the number of users and the amount of information shared explosively rise. Users may use these sources of messages to connect, share, discover and disseminate information. Some of those services provide social connections (Facebook and Twitter, for example). Others (YouTube and Flickr, for starters) are used for sharing content. Knowing the actions of consumers at such pages is one of the major research challenges. System Uses Twitter's Social Network as our case study. In July 2006, Twitter, one of the most successful social networking micro blogging sites, was introduced with around 328 million active monthly users and around 500 million active user's posts per day.

II. RELATED WORK

We use different techniques to classify the text such as Rule-based, Neural Network, Decision Trees and Machine Learning. There are also some tricks and classifications which are based on machine learning. The basic concept of these techniques is to classify news type using the trained classifier which can automatically predict some of the predefined classes with the news type. The concept of probability is used by Naïve Baye's Algorithm. We train the module with the Bayesian rule of probability and the parameter in Naïve Baye's. After that the text document is represented in the form of a bag of words where each word is

considered to be independent of the other this primarily degrades the output of that method.

III. EXISTING METHODOLOGY

This section provides literature review as well as some existing methodologies which already done by previous authors.

According to [1] the event-based approach based on consumer curiosity used by LeMeNo for News Recommendation. The network of suggestions is based on both news and consumer preferences. Based on machine learning techniques news articles are recommended such as grouping similar articles, predicting their content, topic similarity & keyword extraction. Based on the time spent reading an article, the system learns user interests whether the user likes the article as well as the user-specified rates of interest in various subjects. Our framework incorporates multiple approaches to news recommendations to further improve the likeliness of users to recommend a relevant article.

According to [2] evaluates some of the most Machine learning techniques are commonly used to automatically identify Nepali data, particularly Naive Bayes, SVM and Neural Networks. The method is being experimented with a self-created Nepali News Corpus with 20 different categories and a total of 4964 posts, gathered online by crawling various national news portals. Functionality dependent on TF-IDF is derived to train and examine the models from the preprocessed documents.

According to [3] Social Poisson factorization (SPF), a Probabilistic model incorporating social network information into a standard factorization method SPF applies to the algorithmic suggestion a social aspect. It provides a robust method to test SPF data and shows that it outperforms rival methods on six datasets in the real world data sources include a social reader.

According to [4] Privacy risks Similar to numerous emerging and influential automation patterns, including internet customization, behavioral profiling and location-based customization. Program analyzes user behaviors about privacy and personalization as well as technologies that can help reduces the risks to privacy. Program ends with a review that describes risks and technical solutions as well as places at the nexus of personalization and privacy for further study. Such structures will help programmers and analysts place the data protection issues in perspective of solutions when designing customization systems.

According to [5] Active approach to creating an organized user profile that highlights the transient essence of active user behavior. The user profile is collected from diverse, heterogeneous data sources, documenting dynamic consumer activity over time, to reliably represent changing

desires. To collect specific user data and incorporate the suggested "3D User Profile," natural language processing methods, machine learning and semantic interface technologies were used. Our approach often supports user profiles generated as structured data, so that other customized recommendation systems and Semantic Linked Open Data applications can use them to provide smart, personalized services.

According to [6] the recommendation system is part of the information retrieval area, the data mining class and the machine learning class. Recommendation tools play a central role in the ecommerce market today. Recommenders systems generally alert customers of items like books, DVD's, images, electronic products, and much more. Recommendation services help users receive tailor-made reviews help users make the right decisions regarding their online transactions, increase sales and redefine web browsing experience for users, keep customers, and enhance their shopping experience.

According to [7] a user profile model to define user preferences that are multi-perspective. Then system discuss the degree of user preferences for historical news and propose a method for calculating historical news ' preferential weight based on the user's reading behavior and news popularity. This approach may create user profiles more effectively. System also provides a dynamic news recommendation method that takes into account the preferences of both short-term and long-term users. Recommendation based on content: the recommendation system attempts to find news with content similar to the news the user has read. According to [8] is a platform to improve user interaction and familiarity with Networks Communications. It initially applies a mechanism that better subscribes the customer through a dynamic, customized recommendation system that gives users the most suitable tweets. Trend Fusion, a ground-breaking tool used by social media to improve user feedback. This analyzes forecasts the regional distribution of patterns in the social network and suggests the most interesting trends for the consumer.

According to [9] In Google News, personalized news notification program. The Recommendation system creates accounts for consumers who are signed in with news interests and expressly enabled Web history based on their past click behavior. System first conducted a large-scale analysis of anonymizing Google News users by clicking logs to understand how the interest in news for users changes over time. In this paper we built a Bayesian system based on the log study to predict users ' current news priorities from the actions of that user and the news patterns shown in all users ' activity.

According to [10] Customized news system recommendation technology. In particular, the Research work has suggested a shared hybrid filtering algorithm based on news reviews to meet the demand for the personality of the users and ease the data sparse problem. Through strengthening the correlation coefficient function through incorporating news hot parameters when measuring user similarity. But in this paper the hybrid recommendation algorithm is used to predict user ratings to make non-zero user rating matrix.

IV. PROPOSED SYSTEM DETAILS

Because of the multi-dimensional nature of fake news, recognizing the category of news is not so easy. This is the reason the proposed strategy is a mix of Naïve Baye's classifier, Support Vector Machines, and semantic investigation.

A. System Architecture

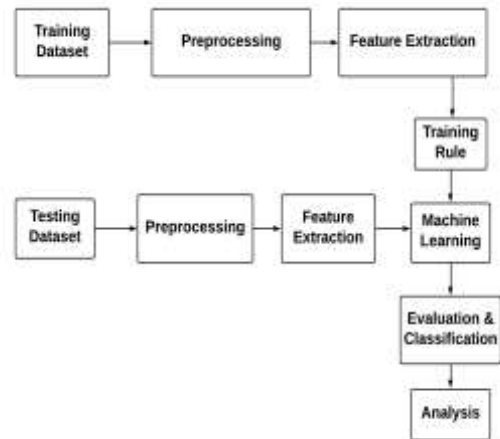


Fig. 1: Proposed System architecture

1) Data Acquisition:

First extract all the information for different Social Media accounts based on certain parameter is extracted from API and then divide the data into test and train data. We go through the following steps to detect whether the news is fake or real.

2) Data Pre-processing:

Then we will apply various preprocessing steps such as lexical analysis, Stop word removal, steaming (porter's algorithm), Index term selection and data cleaning in order to make our dataset proper.

3) Lexical analysis:

Lexical analysis separates the input alphabet into 1) Word character (e.g. the letters a-z) and 2) word separators (e.g. space, newline, and tab).

4) Stop word Removal:

Stop word removal refers to the removal of words that occur most frequently in documents.

5) Steaming:

Steaming replaces all the variants of a word with a single stem word. Variants include plurals, gerund forms (ing forms), third person suffixes, past tense suffixes, etc.).

6) Data Training:

We compile Artificial as well as real time using online news data and provide training with any machine learning classifier.

7) Testing with machine learning:

We predict online news using any machine learning classifier, weight calculator for real time or synthetic input data accordingly.

8) Analysis:

We demonstrate the accuracy of proposed system and evaluate with other existing system.

B. Algorithm Design

Input: Feature of training rules TrainFeatures[], features for test record TestFeatures[]

Output: highest Similarity weight for class label.

Step 1: Read all training rules from DB for each (Record R into Train[])!=Null.

Step 2: items [] split(R)

Step 3: items1 [] split(TestF)

Step 4: Calculate Weight(DB [i], items1)

Step 5: Return w.

C. Mathematical Model

Let SModel, be the proposed system which is characterized as below

$Smodel = \{\{Iset\}, \{Input_t, Input_s, Input_{st}, Input_{tw}, Input_{tr}, Input_{ts}, Input_{fs}, Input_{cs}, Input_{tp}, Input_r\}, \{R\}\}$

Which is define =

$Iset \rightarrow$ Input data collected for training and testing

$Input_t \rightarrow$ web document metadata extracted from triplets.

$Input_s \rightarrow$ Apply stop word removal on extracted data.

$Input_{st} \rightarrow$ Then applies = Porter's Stemming on stop word removal data called as lemmas features.

$Input_{tw} \rightarrow$ calculate the Feature using TF-IDF method.

$Input_{tr} \rightarrow$ Training data Set

$Input_{ts} \rightarrow$ Test data Set

$Input_{fs} \rightarrow$ Extract the features set from training dataset called as normalized features set.

$Input_{cs} \rightarrow$ Apply proposed machine learning algorithm and calculate the similarity weight.

$Input_{tp} \rightarrow$ apply optimization algorithm for ordered the best result according to achieved weight.

$R \rightarrow$ News prediction based on available category.

V. DATASET USED

For this research we collected data set from online social media using twitter API. Using this API we extract various existing news as well as currently posted information by different users. We downloaded around 2000 samples to evaluate the proposed system using supervised learning algorithms. The data splitting mechanism has use as 10 fold cross-validation.

Total Size	2000
Training Samples	1450
Testing Samples	650

Table 1: Dataset description downloaded using twitter API

VI. CONCLUSION

In this work, we have shown a practical approach for the identification of fake news on Social media. Proposed system describes a personalized based news recommendation from social media. It helps the users to believe in particular news which is spread over the Social Media. In future the efficiency and accuracy of the system can be enhanced to a certain level.

REFERENCES

[1] Khandelwal, Dhruv, et al. "LeMeNo: Personalised News Using Machine Learning." 2018 *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. IEEE, 2018.

[2] Shahi, Tej Bahadur, and Ashok Kumar Pant. "Nepali news classification using Naïve Bayes, Support Vector Machines and Neural Networks." 2018 *International Conference on Communication information and Computing Technology (ICCICT)*. IEEE, 2018.

[3] Chaney, Allison JB, David M. Blei, and Tina Eliassi-Rad. "A probabilistic model for using social networks in personalized item recommendation." *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 2015.

[4] Toch, Eran, Yang Wang, and Lorrie Faith Cranor. "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems." *User Modeling and User-Adapted Interaction* 22.1-2 (2012): 203-220.

[5] Krishnan, Gokul S., and S. SowmyaKamath. "Dynamic and temporal user profiling for personalized recommenders using heterogeneous data sources." 2017 *8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2017.

[6] Vaidya, Nayana, and A. R. Khachane. "Recommender systems-the need of the ecommerce ERA." 2017 *International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2017.

[7] Zhu, Zhiliang, et al. "A Dynamic Personalized News Recommendation System Based on BAP User Profiling Method." *IEEE Access* 6 (2018): 41068-41078.