International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Online Public Dishonor Detection and Analysis via Social Media using Machine Learning Algorithms

Miss. Kolse Snehal J. and Prof. S. D. Jhondhale

Department of Computer Engineering Pravara Rural Engineering College, Loni Savitribai Phule Pune University Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Now social media data are increases very fast. In every area, social media data play an important role in every angle. Social media big data mining area welcomed by researchers. In current web word, range of users use social media and social network to read people connected information. Dishonoring behaviour of users on social media is major problem in today's life. It is seen that out of all remarks posted by users in a specific occasion on social media, greater part of them are probably going to embarrass the person in question. In this paper, dishonoring tweets identification and classification is carried out using machine learning algorithm. Dishonoring tweets are classify into five types: Offensive, correlation, condemning, strict/ethnic, joke on personal issue and each post/comment characterized into one of these types. At last, detection and classification of dishonoring tweets using machine learning algorithm is potential solution for online dishonor behaviour.

Keywords: *Dishonoring, online user interaction, public identification, text mining, classification,machine learning,Social Media.*

Introduction

Web based life, for example, Twitter, Facebook and instagram, is overwhelmingly the go-to stage for web clients to share their remarks or encounters towards certain point. It is a gold dig for the individuals who welcome the benefit of understanding open sentiment. Online Social Network (OSNs) are as often as possible overwhelmed with sharing comments against people or associations on their apparent bad behavior. At the point when a portion of these comments relate to target reality about the occasion, a sizable extent endeavors to censure the subject by passing speedy decisions dependent on bogus or mostly confirmed actualities. Restricted extension of fact checkability combined with the destructive idea of OSNs. That regularly converts these facts into dishonoring or public insult or trolling etc. Unnecessary talk as loathe discourse, tormenting, foulness, blazing, trolling etc., is all around examined in the survey. Then again, open disgracing, which is judgment of somebody who is in violation of accepted social norms to arouse feeling of blame in that person, has not pulled in a lot of consideration from a computational point of view. All things considered, these things are continually being on the rise for a few years. Open dishonoring have broad effect on essentially every part of unfortunate casualty's life. In broad daylight disgracing, a shammer is only here and there redundant rather than tormenting. Loathe discourse and obscenity are once in a while part of a disgracing occasion yet. There are various types of disgracing, for example, mockery and jokes, examination of the injured individual with some different people, and so forth.

The huge amount of comments which is often used to dishonor an almost unknown victim speaks. These is the viral nature of dishonoring events. For example, when Vivek oberoi, a public relations person tweeted one meme. Soon, a barrage of criticisms started pouring in, and the incident became one of the most talked about topics on social media and the Internet within day.

The organization of the paper is as follows section II gives the related work and limitations, section III gives proposed methodology, section IV gives results and discussion and last section concludes the paper with future work followed by references.

Review of Literature

Online social networks are often ooded with scathing remarks against individuals or businesses on their perceived wrongdoing. This paper studies three such events to get insight into various aspects of shaming done through twitter. An important contribution of our work is categorization of shaming tweets, which helps in understanding the dynamics of spread of online shaming events. It also facilitates automated

segregation of shaming tweets from non-shaming ones [1].

Two competing opinions regarding the role of social media platforms in partisan polarization. The "echo chambers" view focuses on the highly fragmented, customized, and niche-oriented aspects of social media and suggests these venues foster greater political polarization of public opinion. An alternative, which we term the "crosscutting interactions" view, focuses on the openness of the Internet and social media, with different opinions. This view thus argues that polarization would not be especially problematic on these outlets. Exploiting the variation among members of the U.S. House of Representatives in measured positions of political ideology, this study estimates the association between politicians' ideological positions and the size of their Twitter readership. The evidence shows a strong polarization on Twitter readership, which supports the echo chambers view. Lastly, we discuss the implications of this evidence for governments' use of social media in collecting new ideas and opinions from the public [2].

For detection of dishonour behavior two primary trigger mechanisms: the individual's mood, and the surrounding context of a discussion (e.g., exposure to prior trolling behavior). Through an experiment simulating an online discussion, we find that both negative mood and seeing troll posts by others significantly increases the probability of a user trolling, and together double this probability. To support and extend these results, we study how these same mechanisms play out in the wild via a data-driven, longitudinal analysis of a large online news discussion community. This analysis reveals temporal mood effects, and explores long range patterns of repeated exposure to trolling. A predictive model of trolling behavior shows that mood and discussion context together can explain trolling behavior better than an individual's history of trolling. These results combine to suggest that ordinary people can, under the right circumstances, behave like trolls [3].

A novel approach to the problem: our goal is to identify troll vulnerable posts, that is, posts that are potential targets of trolls, so as to prevent trolling before it happens. To this end, we define three natural axioms that a troll vulnerability metric must satisfy and introduce metrics that satisfy them. We then define the troll vulnerability prediction problem, where given a post we aim at predicting whether it is vulnerable to trolling. We construct models that use features from the content and the history of the post for the prediction. Our experiments with real data from Reddit demonstrate that our approach is successful in identifying a large fraction of the troll vulnerable posts [4].

To address the difficult task of sarcasm detection on Twitter by leveraging behavioral traits intrinsic to users expressing sarcasm. We identify such traits using the user's past tweets. We employ theories from behavioral and psychological studies to construct a behavioral modeling framework tuned for detecting sarcasm. We evaluate our framework and demonstrate its efficiency in identifying sarcastic tweets [5].

To check out whether or not public mood as measured from big-scale series of tweets published on twitter.com is correlated or even predictive of DJIA values. The consequences shows that modifications within the public temper nation can certainly be tracked from the content of large-scale Twitter feeds by way of instead simple textual content processing techniques and that such changes reply to a ramification of socio-cultural drivers in a exceptionally differentiated way [6].

Analysis of financial blogs and on-line news articles to expand a public mood dynamic prediction model for stock markets, referencing the perspectives of behavioral finance and the traits of online economic groups. A public mood time series prediction model is likewise provided, integrating features from social networks and behavioral finance, and uses huge information evaluation to assess emotional content material of commentary on modern inventory or economic issues to forecast changes for Taiwan stock index [7].

The software of deep recurrent neural networks to the challenge of sentence-stage opinion expression extraction. DSEs (Direct Subjective Expressions) consist of specific mentions of personal states or speech events expressing nonpublic states; and ESEs (Expressive Subjective Expressions) encompass expressions that imply sentiment, emotion, etc., with out explicitly conveying them [8].

Analysis of electoral tweets for extra subtly expressed facts such as sentiment (tremendous or bad), the emotion (pleasure, sadness, anger, and so forth.), the cause or reason behind the tweet (to point out a mistake, to aid, to ridicule, and so forth), and the style of the tweet (simple statement, sarcasm, hyperbole, and many others). There are sections: on annotating textual content for sentiment, emotion, fashion, and categories including cause, and on automatic classifiers for detecting those classes. Advantages are: Using a multitude of custom engineered features like those concerning emoticons, punctuation, elongated words and negation along with unigrams, bigrams and emotion lexicons features, the SVM classifier achieved a higher accuracy. Automatically classify tweets into eleven categories of emotions [9].

Representation of how large amounts of social media data can be used for large-scale open-vocabulary personality detection, evaluate which features are predictive of which personality dimension; and present a novel corpus of 1.2M English tweets (1,500 authors) annotated for gender and MBTI. Advantages are: The personality distinctions, namely INTROVERT–EXTROVERT (I–E) and THINKING–FEELING (T–F), can be predicted from social media data with high reliability. The large-scale, openvocabulary analysis of user attributes can help to improve classification accuracy [10].

Proposed Methodology

Our aim is to automatically classify posts/comments in the dishonoring five categories. In figure. 1, the main functional units involving automated classification of dishonoring comments are shown. Both labeled training set and test set of posts for each of the categories go through the preprocessing and feature extraction steps. The training set is used to train classifiers. The precision scores of the trained classifiers are next evaluated on the test set. Based on these scores, the classifiers are arranged hierarchically. A new post, after preprocessing and feature extraction, is fed to the trained classifiers and is labeled with the class of the first classifier that detects it to be positive. A tweet is deemed honoring if all the classifiers label it as negative.

A. Architecture

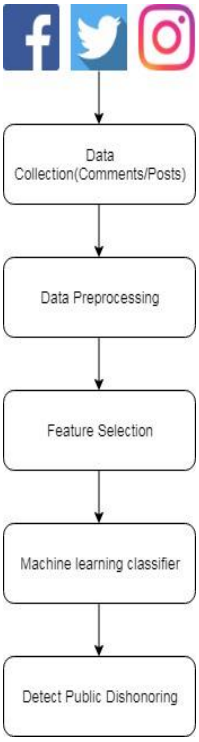

Fig. 1. Proposed System Architecture

1. Data Collection:
Wecollecttweet data from twitter using Twitter API(apps.twitter.com) 2. Preprocessing:
1. Stop word Removal- This technique removes stop words like is, are, they, but etc.
Initialize i,j
for i=1 to no of words in documents for j=1 no of words in stopword list
if
Words(i)==Stopwords(j)
then eliminate words(i) end if end for
2. Tokenization- This technique removes Special character and images.
Initialize feature vector bg feature =[0,0..0] for token in text.tokenize() do if token in dict then token idx=getindex(dict,token) bg feature[token idx]++ else continue end if end for
3. Stemming– Removes suffix and prefix and find original words for e.g.- 1. played – play 2.Clustering - cluster
The word w
Input = Normalize(input) if normalizeValidate(input) then return input; for each rule in rules do if input match with rule then
Stem = ExtractStem(input,rules) if not TestStemLength(Rule) then end for return input
3.Classification Algorithm:
Naive Bayes machine learning algorithm used for classification of text features.
Steps:
1. Given training dataset D which consists of documents belonging to different class say Class A and Class B 2. Calculate the prior probability of class A=number of objects of class A/total number of objects Calculate the prior probability of class B=number of objects of class B/total number of objects
3. Find NI, the total no of frequency of each classNa=the total no of frequency of class A Nb=the total no of frequency of class B
4. Find conditional probability of keyword occurrence givena class:
P (value 1/Class A) =count/ni (A)
P (value 1/Class B) =count/ni (B)
P (value 2/Class A) =count/ni (A)
P (value 2/Class B) =count/ni (B)
.....
....
P (value n/Class B) =count/ni (B)
5. Avoid zero frequency problems by applying uniform
distribution
6. Classify Document C based on the probability p(C/W)
a. Find P (A/W) =P (A)*P (value 1/Class A)* P (value
2/Class A)..... P(value n /Class A)
b. Find P (B/W) =P (B)*P (value 1/Class B)* P(value 2/Class
B)..... P(value n /Class B)
7. Assign document to class that has higher probability.

B. Block Diagram

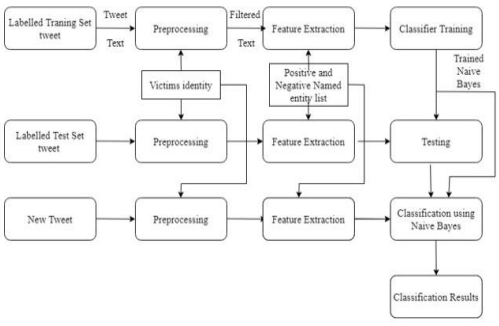

Fig. 2. Block Diagram

C. Algorithm

1. Naive Bayes Classifier

Ability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier. The Naive Bayes algorithm is called naive because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features. The Naive Bayesian classifier is based on Bayes theorem with the independence guess between predictors. A Naive Bayesian model is easy to form, with no critical iterative parameter computation which makes it particularly useful for very large datasets. Regardless of its simplicity, the Naive Bayesian classifier often does particularly well and is widely used because it often out performs more experienced classification methods.

D. Dataset

We obtain Twitter tweet and retweet data through our collaborations on research with Beijing Intelligent Starshine Information Technology Corporation, a leading big data collection and mining service provider in China. In this paper, each tweet or retweet is assigned a honoring and dishonoring tweet labels. We employ 15 raters to manually assign a dishonoring and honoring label for each tweet and retweet. Note that a lot of tweets are reposted without any added comments. We obtain a labeled dataset which contains over 100,000 tweets and retweets in total.

Results and Discussion

The simulation platform is used. It is built using Java framework on Windows platform. The system does not require any specific hardware to run; any standard machine is capable of running the application. We find dishonoring tweets and honoring tweets using naive bayes machine learning algorithm.

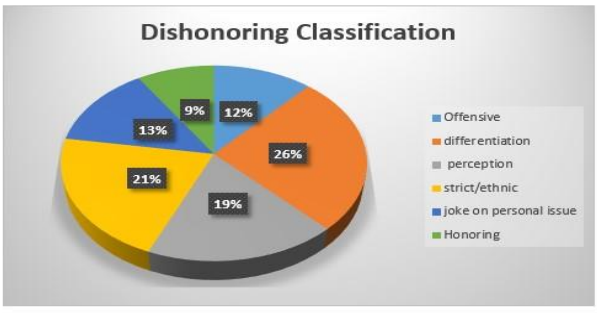

Fig. 3. Dishonoring Classification

Table 2:Dishonoring Classification

Sr No.	Dishonoring Type	Total Count
1	Offensive	18
2	differentiation	41
3	perception	29
4	strict/ethnic	33
5	joke on per-sonal issue	21
6	Honoring	14

Conclusion

In this paper, we proposed a potential solution for countering the menace of online public dishonoring in social media by categorizing shaming comments in five types, choosing appropriate features, and designing a set of classifiers to detect it. Instead of treating posts/comments as standalone utterances, we studied them to be part of certain dishonoring events. Dishonoring words can be mined from social media. In this proposed system allows users to find dishonoring comments with the data and their overall polarity in percentage is calculated using classification by machine learning. Potential solution for countering the menace of online public dishonoring in social media.

References

[1]. R. Basak, N. Ganguly, S. Sural, and S. K. Ghosh, "Look before you shame: A study on shaming activities on Twitter," in Proc. 25th Int. Conf. Companion World Wide Web, 2016, pp. 11–12.

[2]. S. Hong and S. H. Kim, "Political polarization on Twitter: Implications for the use of social media in digital governments," Government Inf. Quart., vol. 33, no. 4, pp. 777–782, 2016.

[3]. J. Cheng, C. Danescu-Niculescu-Mizil, J. Leskovec, and M. Bernstein, "Anyone can become a troll," Amer. Sci., vol. 105, no. 3, p. 152, 2017.

[4]. P. Tsantarliotis, E. Pitoura, and P. Tsaparas, "Defining and predicting troll vulnerability in online social media," Social Netw. Anal. Mining, vol. 7, no. 1, p. 26, 2017.

[5]. Rajadesingan, R. Zafarani, and H. Liu, "Sarcasm detection on Twitter: A behavioral modeling approach," in Proc. 8th ACM Int. Conf. Web Search Data Mining, 2015, pp. 97–106.

[6]. J. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," J. of Computational Science, vol. 2, no. 1, pp. 1-8, 2011.

[7]. J. Bollen, H. Mao, and A. Pepe, "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena," in Proc. of the 5th Int. AAAI Conf. on Weblogs and Social Media Modeling, 2011, pp. 450453.

[8]. S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets," Information Processing and Management, vol. 51, no. 4, pp. 480-499, 2015.

[9]. B. Plank and D. Hovy, "Personality Traits on Twitter — or— How to Get 1,500 Personality Tests in a Week," in Proc. of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 92–98.

[10]. X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang, "Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval," Proc. of the 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 912-921, 2015.