Cyberbullying and Fake Account Detection in Social Media

Authors:

Prof. Richard Joseph, Jayesh Samtani, Sagar Sidhwa, Somesh Tiwari, Riya Wadhwani

Affiliation: Mumbai University, BE Computer Engineering, India.

INTERNATIONAL CONFERENCE ON ADVANCES IN SCIENCE AND TECHNOLOGY - ICAST 2021

Introduction

- Cyber Crime and Bullying have increased on Social Networking sites with having more than 50 Crores active users until now, so the misuse of the Online Social Platform had taken place in several times for e.g Bullying Someone by sending the harmful messages ,spreading of the harassment messages by using the fake accounts, using the abusive words on the social media etc

- In a recent report it was found that nearly 25% of People, especially teens and young adults are finding new ways to bully one another over the Internet and parents don't know that their child has been involved in a cyberbullying incident.

- A Preventive measure to STOP the above crimes caused a need for different Machine Learning algorithms for detection of the Cyber Crime and Bullying and the fake accounts so as to report these issues to the system immediately and to stop the crimes to increase in future and develop a secure online environment.

Research Question

- What is the contribution to ongoing theory building and intervention/prevention efforts against cyberbullying?
- What effects do child victims of cyber bullying last into the victim's adulthood?
- How does cyberbullying affect self esteem in adolescents?

Methodology

- The proposed approach contains three main steps namely Preprocessing, features extraction and classification step.
- In the preprocessing step from the Toxic dataset we had used the parameters - Toxic , Sever Toxic,Obscene,Insult,Threat,Identity hate and from the fake account dataset the parameters used are - Name, Status Count, Followers Count, Friends Count ,Url, Time Zone, Listed Count ,Screen Name ,Profile Bio,Location.
- we clean the data by removing the noise and unnecessary text.
- The preprocessing step is done in the following: -
- **Tokenization**
- **Lowering text**
- **Stop words and encoding cleaning**
- The second step of the proposed Model is the features extraction step. In this step the textual data is transformed into a suitable format applicable to feed into machine learning algorithms

Methodology

- The last step in the proposed approach is the classification step where the extracted features are fed into a classification algorithm to train, and test the classifier and hence use it in the prediction phase. We will use classifiers, namely, SVM (Support Vector Machine), Naive Bayes,Random Forest,Decision Tree , Logistic Regression.

- Accuracy of different algorithms will be Compared to get the best possible result.

- If offensive text is Found in the Post the details of users such as IP address,latitude ,longitude,ISP will be stored.

- For the fake profile detection the detection process starts with the selection of the profile that needs to be tested. After selection of the profile the suitable attributes ie., features are selected on which the classification algorithm is being implemented ,the attributes extracted are passed to the trained classifier.

Models and Techniques

Stochastic Gradient Classifier

- SGD Classifier is a linear classifier (SVM, logistic regression, a.o.) optimized by the SGD. These are two different concepts. While SGD is a optimization method, Logistic Regression or linear Support Vector Machine is a machine learning algorithm/model. You can think of that a machine learning model defines a loss function, and the optimization method minimizes/maximizes it.

- The word ‘stochastic‘ means a system or a process that is linked with a random probability. Hence, in Stochastic Gradient Descent, a few samples are selected randomly instead of the whole data set for each iteration. In Gradient Descent, there is a term called “batch” which denotes the total number of samples from a dataset that is used for calculating the gradient for each iteration. In typical Gradient Descent optimization, like Batch Gradient Descent, the batch is taken to be the whole dataset. Although, using the whole dataset is really useful for getting to the minima in a less noisy and less random manner, but the problem arises when our datasets gets big.

- So, in SGD, we find out the gradient of the cost function of a single example at each iteration instead of the sum of the gradient of the cost function of all the examples.

Models and Techniques

Random Forest

- Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction
- There are two stages in Random Forest algorithm, one is random forest creation, the other is to make a prediction from the random forest classifier created in the first stage. The whole process is shown below, and it's easy to understand using the figure.
- Here the author firstly shows the Random Forest creation pseudocode:
  1. Randomly select "K" features from total "m" features where k << m
  2. Among the "K" features, calculate the node "d" using the best split point
  3. Split the node into daughter nodes using the best split
  4. Repeat the a to c steps until "l" number of nodes has been reached
  5. Build forest by repeating steps a to d for "n" number times to create "n" number of trees

Experimental Setting

1. we have taken dataset for cyberbully from kaggle and also real time data from twitter API and did some preprocessing(like tokenization, lowering text, then stop word removal) .

2. After this we have extracted the features like we have seperated all the categories of text and then created balanced training and testing dataset so that model will not be overfit.

3. After this we trained different models like (Logistic Regression, SVM, Random Forest, KNNB, BernoulliNB, Multinomial DB).Then we created a TFIDF Vectorizer and calculated F1 score of each model and from that we selected Model which gives best accuracy for each type of comment type.

Results and Discussion

User id no:	11
User name:	Sagar
User email:	sagar@gmail.com
Title:	Result
Content:	son of a bitch
Toxic:	100.0%
Severe toxic:	99.0%
Obscene:	99.0%
Insult:	100.0%
Threat:	47.0%
Identity hate:	32.0%

Results and Discussion

Timezone:	Asia/Kolkata
Continent code:	AS
Country code:	IN
Country:	India
Region:	Maharashtra
City:	Ulhasnagar
Organization:	AS141300 Vrd Webservices Pvt Ltd
Organization name:	Vrd Webservices Pvt Ltd

Results and Discussion

```
[41] df.head(3)
```

	user	listed_count	followers_count	favorite_count	statuses_count	friends_count
0	TDataScience	1281	75385	60	20993	1722
1	sidhuwrites	113	78729	44	59410	2214
2	NVIDIAHPCDev	1107	52163	72	7408	734

```
[19] # RANDOM FOREST

     rf_classifier = RandomForestClassifier(n_estimators=100, max_depth=2, random_state=0)
     rf_classifier.fit(X_train, y_train)
     train_predictions = rf_classifier.predict(X_train)
     prediction = rf_classifier.predict(X_test)
```

```
[42] j=rf_classifier.predict([[20993,75385,1722,60,1281,2,0]])
     #statuses_count', 'followers_count', 'friends_count','favourites_count', 'listed_count', 'sex_code', 'lang
```

```
print(j)
```

```
[0]
```

Conclusions and Future Research

- In this project, we proposed an approach to detect Cyberbullying and comment classification as toxic,obscene,threat,insult,identity hate and Fake Account Detection using machine learning techniques.
- We have evaluated our model on Different ML Algorithms and we have also used Countvectorizer for features extraction By using machine learning algorithms to its full extent.
- We will eliminate the need for manual prediction of a fake account, which needs a lot of human resources and is also a time-consuming process.

Future Research

If a post contains a normal text and a web page link ,our system will identify the web link as the simple text and will calculate the percentage of all the categories.we can use web crawling method to scrap the text from the web page and calculate the percentage of all the categories.

References

[1] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE:synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, vol.16, pp. 321–357, 2002.

[2] I. Jolliffe, Principal Component Analysis, 2002.View at: MathSciNet.

[3] S. Sperandei, “Understanding logistic regression analysis,” Biochemia Medica, vol. 24, no. 1, pp. 12–18, 2014.

[4] R. Kohavi, “A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” in Proceedings of the in 14th international joint conference on Artificial intelligence,pp. 20–25, 1995.

[5] N. E. Willard, “Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress”, Research Press, 2007.

Appendices

- Increase of Cyber Crime and Bullying at Social Networking Sites.
- Need of Detection and Removal of Fake Accounts.
- To Provide a Secure Environment for Users of any Age.
- To Apply Machine Learning to remove the Cyber Crime and Bullying.
- Secure System Entry.
- Operations Layout should be well Maintained.