Author Profiling for Irony Detection based on Natural Language Processing Techniques

Ankan Sinha[1], Sumit Gupta[2]
[1]*Master of Engineering, Department of Computer Science & Engineering, University Institute of Technology, The University of Burdwan*
[2]*Project Supervisor & Assistant Professor Department of Computer Science & Engineering, University Institute of Technology, The University of Burdwan*

Abstract: The project is concentrated on profiling ironic authors in Twitter. Special emphases are given to those authors that employ irony to spread stereotypes, as an example, towards the community. The goal is going to be to classify authors as ironic or not reckoning on their number of tweets with ironic content. Therefore, given authors of Twitter along with their tweets, the goal is going to be to profile those authors who will be considered as ironic. Here during this project, various ML models are accustomed verify the info provided supported Homonyms to detect the Ironical users and distinguished them from the non-Ironical users, thus have obtained an optimal result which encourage us to proceed further within the near future supported this approach of using homonyms to represent an Ironical situation.

***Keywords*: Twitter, Irony, Homonyms, Machine Learning, LR, SVM, Accuracy**

1 INTRODUCTION

Online Communities or Social Network Sites (SNSs) are a perfect place for Internet users for sharing information about their daily activities and interests, publishing and accessing documents, photos, and videos. SNSs like Facebook, Twitter, YouTube, and Instagram give the power to make profiles, to possess an inventory of peers to interact with and to post and browse what others have posted. It comes as no surprise that, overall, SNSs - along with search engines - are among the foremost visited websites. Unfortunately, SNSs also are the best plaza for proliferation of harmful information. Cyberbullying, sexual predation, self-harm practices incitement is a few of the effective results of the dissemination of malicious information on SNSs. Many of those attacks are often carried by one individual, but they will be also managed by groups. The target of the trolls are often selected victims but, in some circumstances, the irony is often directed towards wide groups of people, discriminated for a few features, like race or gender. Such campaigns may involve an awfully sizable amount of stereotype people that are self-excited by hateful discussions, and such irony might find yourself with physical violence or violent actions.

However, irony detection is taken into account as a awfully difficult task, thanks to that it appears in various forms and all told styles of contexts. Irony often has ambiguous interpretations, and it often expresses the contrary to what's literally being said [Butler, 1953]. another excuse why irony may well be harder to detect in transcription than in spoken conversation is e.g., thanks to that information contained within the tone of the voice or facial expressions which will be important for understanding irony are lost. one more theory is that it takes longer to jot down on a computer than talking face-to-face, which people use that overtime to jot down more complex sarcasm than what they use normally. As we are entering the age of massive data there's an increasing amount of information being stored and processed on a daily basis, containing everything from customers shopping behavior to how different medicines act on patients. Handling this amount of information imply an automatic tool for data analysis, which is what Machine Learning provides.

Why is Irony Detection Interesting?

Except for being a remarkable field for human-computer interaction irony detection is additionally a crucial area in sentiment analysis and opinion mining. Understanding irony in text would enable a more

accurate picture of what's requested [Carvalho et al., 2009]. In treatment sarcasm detection could work as detection for brain injuries in an early stage. In research done at the University College of London they were ready to show that folks suffering brain injuries had impaired sarcasm comprehension compared to the control group [Channon et al., 2004]. Another important area is for security reasons to gauge whether a threat is real or not. the implications can be devastating if some ironic posts were to be taken seriously.

2 NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) refers to the branch of computer science and more specifically, the branch of computing or AI concerned with giving computers the flexibility to know text and spoken words in much the identical way kinsmen can. NLP combines computational linguistics rule-based modeling of human language with statistical, machine learning, and deep learning models [3]. Together, these technologies enable computers to process human language within the kind of text or voice data and to 'understand' its full meaning, complete with the speaker or writer's intent and sentiment.

3 ONLINE COMMUNITIES

An online community is simply a bunch of people coming together for a typical purpose, interest, or vision, and doing so via the web. Online communities typically use chat rooms, mailing lists, and forums as their primary mode of interaction [4].

4 AUTHOR PROFILING & IRONY DETECTION

4.1 Overview

Author profiling is that the analysis of a given set of texts to uncover various characteristics of the author supported stylistic- and content-based features, or to spot the author.

Characteristics analyzed commonly include age and gender, though newer studies have consider other characteristics like traits, personality, and occupation as well.

4.2 Irony

Irony may be a sophisticated type of language use that acknowledges a niche between the intended meaning and the literal meaning of the words. although it's a widely studied linguistic phenomenon no clear definition seems to exist [Filatova, 2012].

Irony is often divided into two broad categories: situational and verbal irony.

- The previous one is e.g. a cigarette company having non-smoking signs in the lobby.
- The latter one, which is considered during this master thesis, is most ordinarily defined as "saying the alternative of what you mean" [Butler, 1953] where the difference between the old chestnut and meaning is meant to be clear.

4.3 Homonyms

Homonyms are words that are spelled the identical and sound the identical but have different meanings. The word homonym derived from the word "homo" which implies "the same," and therefore the word "nym", which means "name." Thus, a homonym is a word with a minimum of two different meanings with alike sound and look. *Ex. Bark (Dog bark and Tree bark)*

5 PROPOSED WORK

5.1 Overview

We have built our workflow for the data collected and thus create our own way to implement our model.

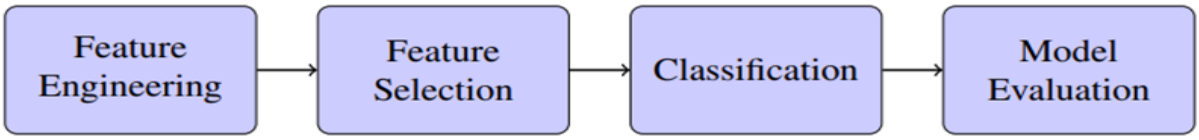

Fig 1: Text Categorization Pipeline

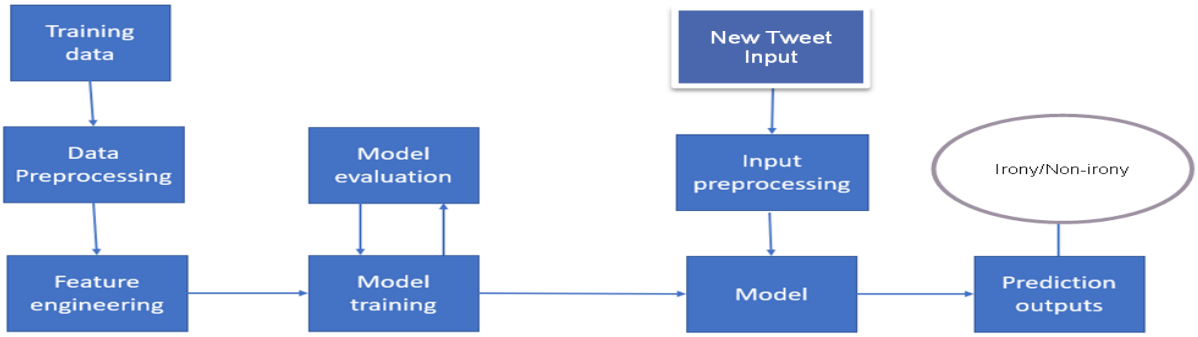

Fig 2: Workflow of our proposed System

6 METHODOLOGY

The different phases of the proposed methodology for Irony detection are described below.

6.1 Data Collection

Twitter may be a micro-blog that permits broadcasting of posts consisting of up to 140 characters and a post is often observed as a Tweet. it's possible to jot down on to other user by mentioning the opposite user's username prefixed with a @ character. That tweet will then come into view also to the mentioned users' followers. A tweet may contain a hashtag prefixing. Users may browse for hashtags; they're therefore normally used for putting a subject to the tweet or commenting on an incident etc. Other kind of data that are normal in tweets are URLs and pictures.

We downloaded the data set from *PAN Clef* [5]. The data was a set of 80,000 tweets of 400 users and a truth file which indicates 20 Users to be Ironical and 20 to be non-Ironical. Based on that we need to detect the Ironical user based on their tweet contents. Due to Machine Limitations, I have used 2000 tweets from each of the Ironical and Ironical users listed in the truth value as my training set and have tested 100 users i.e, 20000 tweets to find the result for Ironical and Non-ironical users.

The dataset contains:

- An XML file per author (Twitter user) with 200 tweets. The name of the XML file corresponds to the unique author id.
- A truth.txt file with the list of authors and the ground truth.

The format of the XML files is depicted in Fig. 3

```
<author lang="en">
<documents>
<document>Tweet 1 textual contents</document>
<document>Tweet 2 textual contents</document>
...
</documents>
</author>
```

Fig 3: Raw Dataset Format

6.2.1 Data Pre-processing:

We use a logistic regression with L1 regularization to reduce the dimensionality of the data. And then using Wordnet and Synsets, we find the various parts of speech for the key words and listed them.

6.2.2 Training of data:

- We then, count the no of Synsets for a word. The word whose synsets is greater than and equal to 2, we tag the word as homonym as per process.
- Now we count the no of homonyms per tweets, per user and no. of homonyms used upon total no. of words used by the users for the tweets, thus we get 3 features set for our training set.
- Again, we find the n grams for the tweets and count the same thus got another one feature for our set.
- Lastly for our fifth features set we calculate the total homonyms used in the 200 tweets per users.

Based on these features set we applied our ML model to train the data and finally obtain our training set for both the Irony and Non-Ironical users individually.

7 IMPLEMENTATION AND RESULT

We trained the final model using the entire dataset and used it to predict the user. We have used LR, Random Forest and SVM as our classifiers and thus compared the accuracy percentage among each other to obtain our final model. All modeling was performing using scikit-learn. The trained set so far, we have obtained from our given data set is as below,

TABLE-1: Homonyms frequency for each user

	User	Homonym/tweet for Ironic user	Homonym/tweet forNon-Ironic user	Homonyms/Ironic user	Homonym/Non-Ironic User
0	1	87.93	47.97	14.371905	12.035119
1	2	60.65	56.78	23.143519	14.033113
2	3	85.43	56.30	23.237668	21.097353
3	4	78.63	49.68	20.251030	15.721862
4	5	89.20	57.94	21.097353	11.090000
5	6	66.40	41.32	15.721862	14.230000
6	7	83.86	53.23	22.860000	10.860000
7	8	62.12	48.89	14.320000	9.870000
8	9	87.74	49.22	16.890000	14.670000
9	10	71.48	55.68	20.230000	17.230000

TABLE-2: Frequency and Mean of homonyms for all the users

	Homonym/tweet for Ironic user	Homonym/tweet forNon-Ironic user	Homonyms/Ironic user	Homonym/Non-Ironic User
mean	77.344000	51.70100	19.212334	14.083745
std	11.229019	5.20544	3.583483	3.390317
min	60.650000	41.32000	14.320000	9.870000
25%	67.670000	48.97250	16.013896	11.326280
50%	81.245000	51.45500	20.240515	14.131556
75%	87.162500	56.14500	22.419338	15.458896

7.1 Implementation

Here, we have used 100 +100 users as our testing set, i.e., We have considered a total of 20000 + 20000 each for ironic and non-Ironic detection of the users.

7.1.1 On implementing the LR model, below result is obtained,

TABLE-3: LR MODEL PREDICTION MATRIX

Prediction Matrix	Predicted	
	Ironic	Non-Ironic
Ironic	76	24
Non-Ironic	32	68

7.1.2 On implementing the RANDOM FOREST, below result is obtained,

TABLE-4: RANDOM FOREST PREDICTION MATRIX

Prediction Matrix	Predicted	
	Ironic	Non-Ironic
Ironic	78	22
Non-Ironic	25	75

7.1.3 On implementing the SVM, below result is obtained,

TABLE-9: SVM PREDICTION MATRIX

Prediction Matrix	Predicted	
	Ironic	Non-Ironic
Ironic	84	16
Non-Ironic	17	83

7.2 Accuracy comparison

	Precision	F1 Score	Accuracy
LR	0.7037	0.7308	0.7200
RF	0.7800	0.7685	0.7650
SVM	0.8400	0.8358	0.8350

TABLE -10: Accuracy comparison

Thus, we have obtained a maximum accuracy of 83.50% through our SVM model

8 CONCLUSION

Irony is a difficult phenomenon to define and is not monolithic. Our classifications of Irony tend to reflect our own subjective biases. People identify racist and homophobic slurs as hateful but tend to see sexist language as merely offensive. While our results show that people perform well at identifying some of the more egregious instances of Ironical and non-ironical, it is important that we are cognizant of the social biases that enter into our algorithms and future work aim to identify and correct these biases and also to increase our dataset quantity for a more precise result.

9 FUTURE WORKS

Primarily in this project we have successfully detected the Ironical tweets and thus the user from a social analysis and thus established our result. On further process we are looking forward to increasing our accuracy for the same by implementing a new algorithm and to cover up the images, videos and thus providing a more accurate analysis of the hate speeches which also helps us to detects the spreaders as well. Also, we are aiming on proving a live implementation of the work through an API thus to detect the same at the instance.

REFERENCES

[1] https://www.linkedin.com/pulse/natural-language-processing-ashik-kumar/

[2] https://blog.discourse.org/2021/08/online-community/

[3] Data set: https://pan.webis.de/clef22/pan22-web/author-profiling.html.

[4] Hajime Watanabe, Mondher Bouazizi, Tomoaki Ohtsuki, “Hate Speech on Twitter: A Pragmatic Approach to collect Hateful and Offensive Expressions and Perform Hate Speech Detection”.

[5] Rakshita Jain, Devanshi Goel, Prashant Sahu, Abhinav Kumar, Jyoti Prakash Singh, “Profiling Hate Speech Spreaderson Twitter”.

[6] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, Vasudeva Varma, “Deep Learning for Hate Speech Detection inTweets”.

[7] Thomas Davidson, Dana Warmsley, Michael Macy, Ingmar Weber, “Automated Htae Speech Detection and theProblem of Offensive Language”.

[8] Noman Ashraf, Abid Rafiq and Sabur Butt and Hafiz Muhammad Faisal Shehzad and Grigori Sidorov and Alexander Gelbukh, “YouTube Based Religious Hate Speech and Extremism Detection Dataset with Machine Learning Baselines”.

[9] Ashish Sureka, Ponnurangam Kumaraguru, Atul Goyal, Sidhart Chhabra, “Mining YouTube to Discover ExtremistVideos, Users and Hidden Communities”.

[10] Fabio Del Vigna, Andrea Cimino, Felice Dell’Orletta, Marinella Petrocchi, and Maurizio Tesconi, “Hate me, hate menot: Hate speech detection on Facebook”.

[11] ANAT BEN-DAVID, ARIADNA MATAMOROS-FERNÁNDEZ, “Hate Speech and Covert Discrimination on SocialMedia: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain”.

[12] Shakeel Ahmad1, Muhammad Zubair Asghar, Fahad M. Alotaibi and Irfanullah Awan,” Detection and classifcationof social media-based extremist afliations using sentiment analysis techniques”.

[13] Njagi Dennis Gitari, Zhang Zuping1, Hanyurwimfura Damien and Jun Long, “A Lexicon based Approach for HateSpeech Detection”.

[14] Erik Forslid Niklas Wikén, “Automatic irony- and sarcasm detection in Social media”.

[15] Jens Lemmens and Ben Burtenshaw and Ehsan Lotfi and Ilia Markov and Walter Daelemans, “Sarcasm Detection Using an Ensemble Approach