

SIMULATING BOT TO DETECT HUMAN EMOTIONS USING NATURAL LANGUAGE PROCESSING

^[1] Rio Glen Lobo ^[2] Sambhav Jain, ^[3] Saraj Kishore,
^[4] Saurabh SR, ^[5] Dr. M Vinoth Kumar

^{[1][2][3][4]} BE Students, Department of Information Science and Engineering

^[5] Associate Professor, Department of Information Science and Engineering

^{[1][2][3][4][5]} Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India

Abstract

Texting, chatting, and SMS (short messaging service) have become an extremely popular medium of short and informal communication because they are very easy to send and receive. The urge to freely communicate with others is normal, humans are very social creatures who have a deep-seated instinct to form social groups, so that they may find security in the pack, hence this desire to have a healthy relationship is something we are born with.

However, this social and biological pressure to develop a friendship or an intimate relationship may lead to an individual in making a serious lapse of sound judgement when it comes to choosing the receiver of his emotions.

If we keep relying on our various forms of technologies and social media platforms to meet and socialize with new people across the globe, we may very end up in threatening and treacherous situations because there is really no feasible way to be sure of whom exactly we are talking to. Which is why if we have a well-defined chat bot with the ability to gauge human sentiment as well the intentions behind their messages and the ability to respond appropriately as well raise any necessary red flags about the communicating party would greatly ease the pressure of texting while safeguarding the user from malicious messages.

Keywords: SIMPBot, DialogFlow, React, SVM, OpenCV.

I. INTRODUCTION

Textual chat or SMS refers to any means of communication or social interaction over the internet which offers the instantaneous transmission of messages or media from the sender to the receiver. Chat rooms and online messaging platforms are currently the most popular ways for individuals to connect online to other individuals from across the globe. There may exist numerous reasons why a person would want to interact with a stranger on the internet in this modern world.

While numerous individuals enjoy the ease and simplicity of communicating via text, a few tend to enjoy its mystery as well as its anonymity. It is very easy to get overwhelmed and addicted in this new world of electronic communication if one tends to suffer from issues with self-esteem, self-worth depression, vanity or anxiety.

In the above instances, a certain amount of anonymity in the communicating medium allows for an individual to interact and socialize with somebody else without having to feel the pressures of social anxiety, peer pressure or social awkwardness that is often associated with in-person or face-to-face communication.

It does not matter how genuine or sincere the individual on the sender side of the chat may seem, there is absolutely no way to know if they are being honest. Sharing personal information with a stranger can make you vulnerable to various kinds of attack, such as emotional abuse, cyber bullying or other serious attacks such as phishing and other cyber attacks.

Victims never anticipate that they may be victimized, certain characteristics may increase the dangerous of someone taking advantage of you or even your friends and family through you. As genuine and kind as your intentions may seem, one should always exercise caution while using text chat to communicate with random individuals. In fact, a very recent statistic has shown that nearly 75% percent of teens, preteens and even adults are open to share intimate details about themselves or their families in return for few benefits, which

can vary from social validation, extortion and financial handouts. Furthermore, in 100% of the cases of online sexual predators luring teens, the teens have gone voluntarily to meet their assaulter.

We have come up with a rather innovative solution to this issue, we propose a chatbot which would act as an intermediary who would respond on behalf of the receiver.

SIMPbot is a unique chatbot that will be developed, the main objective is to create a Bot which observes and analyse the conversation of individuals recognizes patterns and try to capture any hateful, derogatory sentiments in the conversation, rates the individual on the basis of its speaking/writing ability and check if he/she passes a certain etiquette threshold. These features should help the user get a gauge of sender's intentions and make a decision of whether to carry on the conversation or not.

Additionally, ratings can be utilized to make amendments in the individuals or whether to give him/her access to a specific platform. This paper provides a brief overview of the approach to the above motive

II. EXISTING WORK

In this method, the author proposed a technique where we would need to prepare a dataset of sample tweets from the Natural Language Tool Kit package for NLP with various data scrubbing methods. After the dataset is ready to be processed, we would need to train the model on pre-classified tweets and use that model to classify the sample tweets into two sentiments, namely negative and positive sentiments.

Some prerequisites for this technique were to have Python 3 installed in the local computer and having NLTK module imported into the python interpreter. We then need to download the sample tweets from the NLTK package where the python interpreter stores it locally.

Now we need to make sense of the data that has been stored, therefore, the author used a technique of tokenization. Here, the data is split into smaller parts. This process takes a while as it isn't a simple splitting.

After tokenization, we now need to normalize the data where the word is converted into its canonical form. The author uses two techniques, lemmatization and stemming. After this, we need to remove noise from the data. Noise refers to the hyperlinks, special character or even punctuations.

The next step is to determine word density. It counts the words and the frequency of that word in a list. Following this, the data is prepared for the model. Naïve Bayes classifier is used to exercise the model. The final step follows, which is building and testing model. The model classifies the script into a particular sentiment, positive or negative tweets.

III. METHODOLOGY

Initially the data is pre-processed and is split on the basis of words and characters using a splitter in python, then the Word level input (embedding model) is applied to the words obtained from the splitter and one hot vector is used at the character level.

1. CHATBOT

Chatbot is a software designed to emulate the atmosphere of talking to a real individual online. It can be in a format of text or maybe in the format of text-to-speech. Its main motive is talk as a real individual would do in a real-life conversation. This requires a lot of tuning and testing. The term was first coined in 1994 by Michael Mauldin to describe these types of conversational programs.

Chatbots are used in various dialog systems for various services like customer support, routing and information gathering. Some chatbots uses word classification processes, natural language processing and artificial intelligence. Some chatbots simply scan for the general phrases and words from the databases or libraries.

1.1 ARCHITECTURE OF CHATBOT

The architecture of chatbot includes an Intent Classification module which identifies the intent of users' messages. Entity recognition module which extracts some structured bits of information from the message. The candidate response generator does all the specific calculations to process the user request. And finally, the response selector just scores all the response candidates and selects the most suitable response for the user.

1.2 CHATBOT WORKING

1) The messages will be provided by the user in the frontend (i.e., React).

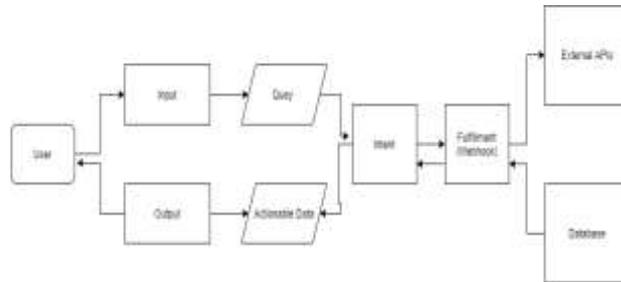


Figure 1: Intent and Entity Architecture

- 2) The messages will be transferred from the frontend to the Dialog Flow.
- 3) The message is categorized and matched to a corresponding intent (Intents are defined manually by developers in Dialog Flow)
- 4) We define the following actions for each intent in the fulfillment (Webhook).
- 5) When a particular intent is found by the Dialog Flow, webhook will use external APIs to find an appropriate response in the databases
- 6) The required response is sent back from the database to the webhook.
- 7) Webhook sends a formatted response to the intent.
- 8) Intent creates an actionable data which is then sent to the frontend.
- 9)The user gets the appropriate response from the simpbot.

Figure 2: Chatbot request/response Flowchart



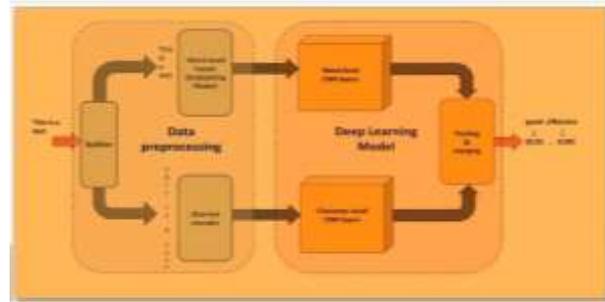


Figure 3: Flowchart of Text Detection

2. SPEECH DETECTION

Detection of abusive language is a difficult task to perform and working on a dataset to detect acceptable and unacceptable content is a much difficult task to perform.

2.1 WORD EMBEDDED MODELS

We have used the Word Embedded Model, in this all words of similar meaning have an analogous encoding. Embedding means 'dense' vector. Sentences are split into words and characters. Each word is represented as an array of numbers and Each character is represented as a binary array of numbers. One of the models that is developed by the word embedding approach is Word2Vec, it makes use of three layers which are the input layer, hidden layer and an output layer. Word2vec utilizes the two-model architecture to produce a distributed presentation of words: Continuous Bag-of-words (CBOW) or continuous skip-gram. In the above architecture, the model predicts the current word out of a window of surrounding context words. The given order of context words tends to not influence prognostication (bag-of-words assumption). In the continuous skip-gram architecture, the model uses the current word to prognosticate the neighbouring window of context words. The skip-gram architecture weighs close by context words more heavily than the distant context words. CBOW is faster while skip-gram is relatively slower but is prone to do a marvellous job for infrequent words.



Figure 4: Word2Vec

2.2 ONE-HOT

A one-hot is a group of bits among which the legal combinations of values are only those

with a single high (1) bit and all the others low (0). In natural language processing, a one-hot vector is a $1 \times N$ matrix (vector) used to distinguish between each of the given words word in their vocabulary from alternate words in the vocabulary. The vector mainly made up of 0 in all the cells with the exception of a single 1 in a cell used uniquely to recognize the word. One-hot encoding guarantees that machine learning does not assume that larger numbers are more valuable. For example, the value '8' is bigger than the value '1', but that does not mean that the value of '8' is more important than that of '1'. The same is true for words like: the value 'New York' is not more valuable than that of 'York'.

IV. DATASET

The Dataset used is obtained from Kaggle “Predicting the Type and Target of Offensive Post in Social Media by OLID”.

Apart from this, we have pre-trained dataset which was manually filled and made. Along with that we have all kind of Abusive words dataset to identify the abusive text from the text

We have made a dataset to identify the emotion behind the text which consists of more than 500 emotions.

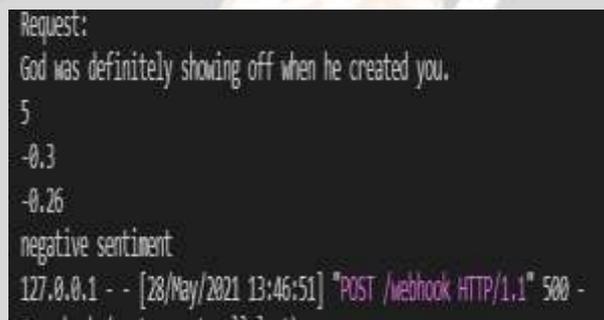


Figure 5: Negative Sentiment detection entered

In the DialogFlow, we have made intent which includes Bad language, flirt, Greeting, Hindi Abusive, Lewd Comments, Pickup lines, Trial.

V. RESULTS

After working on the pre-processed dataset with the model on word level and character level we got results on our various dataset, Bad language has performed 79.86% where we have performed on 900 negative sentiment out of 1000 mixed sentiments, Flirt has performed 81.04% where we have done it on 300 samples, Greetings has a given a result of 98%, Hindi Abusive has given a result of 82.65%, where on the other hand Lewd Comments and Pickup line shown result of approx. 79%.

Dataset	Results
Bad Language	79.86%
Flirt	81.04%
Greetings	98%
Hindi Abusive	82.65%
Lewd Comments	79%
Pickup Lines	79%

VI. CONCLUSION

The aim of SIMPBot is to create a bot which is able to observe the conversation and analyse whether the individual chatting is appropriate or not.

The bot tries to grasp any hateful, derogatory sentiment made by the individual and rate him/her on the aforementioned information. The rating achieved in the process can be utilised to make amendments in the individuals itself or whether to

allow that individual on a said specific app or platform. We strongly believe that future implementations of this bot will lead to a safer online environment in which conversations are monitored not for tracking, advertisements and other such money-making schemes but for the safety of our users.

VII. REFERENCES

- [1] Kaur, J., & Saini, J. R. (2014). Emotion detection and sentiment analysis in text corpus: a differential study with informal and formal writing styles. *International Journal of Computer Application*, ISSN, 0975-8887.
- [2] Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50, 723-762.
- [3] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2011). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 62(2), 419.
- [4] Escalante, H. J., Villatoro-Tello, E., Garza, S. E., López-Monroy, A. P., Montes-y-Gómez, M., & Villaseñor-Pineda, L. (2017). Early detection of deception and aggressiveness using profile-based representations. *Expert Systems with Applications*, 89, 99-111.
- [5] Bohra, A., Vijay, D., Singh, V., Akhtar, S. S., & Shrivastava, M. (2018, June). A dataset of Hindi-English code-mixed social media text for hate speech detection. In Proceedings of the second workshop on computational modeling of people's opinions, personality, and emotions in social media (pp. 36-41).]
- [6] Baruah, Arup, Ferdous Ahmed Barbhuiya, and Kuntal Dey. "IIITG-ADBU at HASOC 2019: Automated Hate Speech and Offensive Content Detection in English and Code-Mixed Hindi Text." FIRE (Working Notes). 2019.
- [7] Nobata, Chikashi, et al. "Abusive language detection in online user content." Proceedings of the 25th international conference on world wide web. 2016.
- [8] Ptaszynski, M., Lempa, P., Masui, F., Kimura, Y., Rzepka, R., Araki, K., ... & Leliwa, G. (2019). Brute-Force Sentence Pattern Extortion from Harmful Messages for Cyberbullying Detection. *Journal of the Association for Information Systems*, 20(8), 4.
- [9] Yang, J. F., & Shen, X. J. (2006, August). Research on key technologies of content-based erotic image filtering and its application. In *2006 International Conference on Machine Learning and Cybernetics* (pp. 3786-3792). IEEE.
- [10] Xin-Lu, W., Xiao-juan, L., & Xiao-bo, L. (2009, June). Nude image detection based on SVM. In *2009 International Conference on Computational Intelligence and Natural Computing* (Vol. 1, pp. 178-181). IEEE.
- [11] Zhu, H., Zhou, S., Wang, J., & Yin, Z. (2007, August). An algorithm of pornographic image detection. In *Fourth International Conference on Image and Graphics (ICIG 2007)* (pp. 801-804). IEEE.