International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 03, 2020
ISSN: 1475-7192

Ensemble Feature Selection to improve the classifier Performance in Sentimental Analysis

[1]Mr. M. Gunasekar ,[2]S. Naveen Kumar ,[3]K. Sakthi Gnanesh ,[4]T.A.Salman Syed Mukthar ,[5]K.Shribalaji

ABSTRACT-- *Pre-trained word embedding's are used in several downstream applications as well as for constructing representations for sentences, paragraphs and documents. One improvement area is reducing the dimensionality of word embedding. Reducing the size of word embedding can improve their utility in memory-constrained devices, benefiting several real world applications. Therefore, in this paper, we focus on how to classify textual information and it consist of online comments from Wikipedia talk page edits where an unsupervised learning approaches are used to obtain better performance of sentimental analysis. To this end, we first analyse the dataset for pre-training by using a phenomenon called glove word embedding, and giving some unique dimensions to each comments. Then we reduce the dimensions of the comments using dimensionality reduction approach and propose an iterative algorithm called t-SNE to visualise the high dimensional data. Finally, a Bidirectional LSTM model is built using keras to classify the sentences into appropriate types of toxicity. To the best of our knowledge, this work is first to the study of negative online behaviours, like various types of toxic comments.*

Keywords-- *Sentimental Analysis,Word Embedding Dimensionality Reduction, Visualization, Toxicity*

I. INTRODUCTION

Holding constructive and inclusive conversations online is a vital task for providers of platforms. Automatic detection of toxic remarks, such as hate speech, assaults and threats, may lead to positive discussions.

Additionally , new legislation to enforce the deletion of illegal content in less than 72 hours have been implemented in some European countries. Recent research on the topic tackles common problems in natural language processing, such as long-range dependencies or misspelled and idiosyncratic words. The solutions suggested include carefully bidirectional, recurrent neural networks and the use of pre-trained embedding terms. It is vital for future work to know which problems are already being solved by state-of-the-art classifiers and for which problems are still vulnerable to error in current solutions. This examination was done by python.

II. RELATED WORK

[1] *Assistant Professor Department Of Information Technology M Kumarasamy College Of Engineering, Tamil Nadu, India*

[2] *B.Tech Student Department Of Information Technology M Kumarasamy College Of Engineering, Tamil Nadu, India*

[3] *B.Tech Student Department Of Information Technology M Kumarasamy College Of Engineering, Tamil Nadu, India*

[4] *B.Tech Student Department Of Information Technology M Kumarasamy College Of Engineering, Tamil Nadu, India*

[5] *B.Tech Student Department Of Information Technology M Kumarasamy College Of Engineering, Tamil Nadu, India*

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

Text arising from immersive online contact conceals many threats, such as fake news, online abuse and toxicity[6]. Toxic comment is not only verbal abuse but a rude, insulting or something likely to make someone leave a conversation. The personal assault, online threats and bullying attitudes may also be called negative commentary.Unfortunately, it is a common trend in the world of web and causes many problems and the risk is increased with the growth of social media sites and the proliferation of online communication. The Wikimedia foundation suggested that 54 per cent of those who had encountered online were Abuse has resulted in a reduction in participation in a given project.A 2014 Pew Report also highlights that 73 percent of adult internet users saw someone harassed online and 40 percent experienced that personally. While efforts are being made to improve the safety of online environments based on crowdsourcing voting schemes or the ability to denounce a comment, these techniques are inefficient in most cases and fail to predict a potential toxicity.

The first layer of NN architectures integrates the one-hot token representations into a lower-dimensionality vector space, which it then fine-tunes by back-propagation. This layer yields word embedding as the first layer's weights, and is generally referred to as the embedding layer. In recent years, word embedding learnt without supervision has seen considerable progress in various NLP tasks.Embedding layer pre-training using a dedicated word embedding technique such as Word2Vec (Mikolov et al . , 2013) or Glove (Pennington et al . , 2014) has proved to be an efficient method of spreading broad corporate information. We compile a collection of pre-trained word embedding from various online sources, and use them to initialize the classifiers' word embedding layer.Such pretrained embeddings differ in text source, size, amount of tokens next to the training method.

GloVe 7 Pennington et al . (2014) show that the ratio of the probabilities of co-occurrence of two terms (rather than their probabilities of co-occurrence itself) is what comprises information and therefore try to encode this information as vector difference.To accomplish this, they propose a weighted objective of lesser squares which aims directly to reduce the difference between the dot product of two word embedding and the logarithm of their number of co-occurrences. We use GloVe embedding trained on three separate text sources: Twitter text (200 dimensions), WikiNews (300 dimensions), and Popular Crawl Corpus(300 dimensions) general web text.

One area of improvement is reduction of word embedding dimensionality. Reducing the size of word embedding in memory-constrained devices can improve their utility, benefiting several applications in the real world. In this study, we present a novel technique that effectively incorporates PCA-based reduction in dimensionality to create efficient lower-dimensional word embedding.Empirical tests on several benchmarks show that our algorithm effectively reduces the size of the embedding while achieving comparable or (more often) better performance than original embedded content.

Neural networks for the classification of toxic comment use recurrent neural network (RNN) layers, such as long-term memory (LSTM) or Gated Recurrent Unit (GRU) layers, similar to other text classification tasks. Standard neuronal networks suffer from the problem of fading gradients.With the increasing number of time measures, back-propagation over time may cause the gradients used for weight updates to become vanishingly small. With gradients close to zero, no updates are made to neural network weights and therefore there is no training process. With the aid of gates, layers LSTM and GRU overcome the vanishing gradient problem.The status of each cell is transmitted to the next cell and gates that control changes to those states. For an unspecified number of time measures, long-range dependencies can be transmitted if the gates block changes in the states for the respective

cells. An extension to standard layers of LSTM and GRU are bidirectional layers of LSTM or GRU, which process the sequence of words in the correct and reverse order.

III. PROPOSED SYSTEM

The emphasis is on researching online negative attitudes, such as derogatory comments (i.e. remarks that are disrespectful, insulting or otherwise likely to cause others to exit a discussion). There are so many templates providing toxicity by using the Perspective API.Nonetheless, the modern frameworks also make mistakes because they do not allow users to pick the categories of toxicity they are interested in seeing (for example, some sites may be fine with profanity but not for certain toxic material types).Compared to existing models of Insight, the model should be better able to discern various sources of hostility, such as bullying, obscenity, threats and identity-based hatred. Changes to the current model will ideally allow the online dialogue to become more positive and respectful. The Sentimental Analysis Model used in this project is for analysing toxic comments online.The main challenge is to classify the remarks into various forms of toxic remarks, such as poisonous, serious poisonous, harmful, threatening, assault and identity hate. The inclusion of the word Golve pertains to our dataset. Terms are then translated into vector formulae. We were using an unsupervised approach for research.We suggested an important method called reduction of dimensionality to lift the proportions out of huge volumes. For data exploration and visualization of closely related dimensions we used Distributed Stochastic Neighbor Embedding (t-SNE) algorithms.In the end, a Bidirectional LSTM model is built using keras to classify the sentences into correct categories. Finally we used the Bidirectional LSTM model to classify the phrases into different toxicity groups.

IV. IMPLEMENTATION

The implementation contains pre-training and analysis of the dataset.

The implementation of this project involves the process of loading the dataset, classifying the data and presenting visual representation.

A. Algorithm:

We have used the following Algorithm

1) t-SNE Algorithm:

We used t-SNE(Distributed Stochastic Neighbour Embedding) in this proposed sense. We also had very good visualization tests for data, because we used pre-trained weights. When we train ourselves to embed this technique is important.

B. Module Description:

The following system flow is the procedure used in our analysis

Received: 03 Dec 2019 | Revised: 21 Jan 2020 | Accepted: 18 Feb 2020

Work Flow

1) Word embedding – Glove Technique:

GloVe is a pattern dependent upon count. Count-based models learn the vectors with a co-occurrence on matrix counts, reducing the dimensionality. Formulate this matrix into a lower-dimensional matrix of terms and attributes, where each row gives a representation of the vector for each word. Pre-processing the count matrix is achieved by standardizing the counts and smoothing them for long.

The corpus, as we are all aware, is a raw document, and thus must be pre-processed before our model is fed. I used Glove pre-trained Word embedding to construct raw data vectors which were provided. Designed here the Term Vectors have 100 dimensions. Which is every word has been defined as a vector of 100-dimensions.

2) Descriminator – Dimensionality Reduction:

Improving the word vectors that have been pre-trained will make the research process more accurate. The removal of the word embedding dimensionality is one area of transition. Reducing the size of word embedding in memory-constrained systems would improve its performance, benefiting many applications in the real world.Reducing the dimensionality of the widely used pre - trained word embedding (word2vec, glove, fast text) and demonstrating that we can easily control half of the usual dimensionality of the embedding space geometry.

- The prediction problem is given by,

$$w_i^T \cdot \tilde{w}_j + b_i + \tilde{b}_j = \log X_{i,j}$$

bw and bc are bias term

- The objective of the function

$$J = \sum_{i,j=1}^{V} f(X_{i,j}) \, (w_i^T \cdot \tilde{w}_j + b_i + \tilde{b}_j - \log X_{i,j})^2$$

f(Xi,j) is a weighting function to penalize rare co-occurrence

- The model generate two sets of words vectors W and $\tilde{W}$
- W and $\tilde{W}$ are equivalent and differ only as a result of their random initialization

Two set of vector should be perform equivalently

- Authors proposed to use (W + $\tilde{W}$)/2 to get word vector

Components:

There are two components of dimensionality reduction:

•Feature selectionIn this we are attempting to find a subset of the original set of variables or features that can be used to model the issue..

•Feature extraction: This reduces the data to a lower dimension space in a high dimensional feature space, i.e. a space with lesser no. of dimensions.

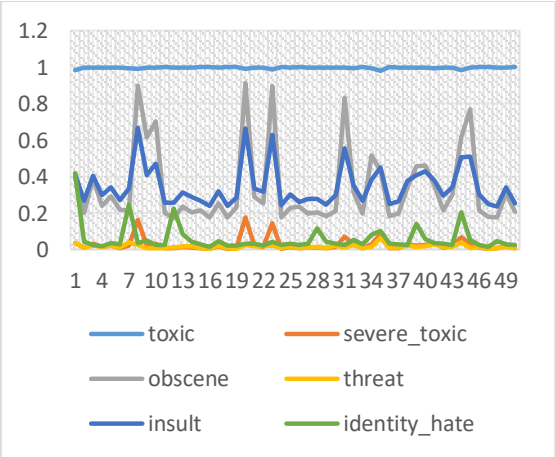
In this project, before using them, we need to imagine our word embedding vectors to see if similar terms are placed next to each other. Since our word vectors have 100 dimensions, visualizing them is not easy. So, we used an unsupervised (Dimensionality Reduction) approach to reduce the dimensions to 2 and then map these terms.

3) Classifier:

Bidirectional LSTMs are given at Keras via the wrapper of the Bidirectional layer. This wrapper takes a standard layer (e.g., first layer of LSTM) as a reference. It also helps you to decide the merge mode, which is how to combine forward and backward outputs before moving on to the next step. Long Short Term Memory Networks (LSTM) are an RNN subclass, Specialized in the long-term processing of information. More over the Bidirectional LSTMS hold the contextual details in both directions. Finally, a Bidirectional LSTM model is built using keras to classify the sentence into correct toxicity categories.

V. RESULTS

Hence the project result includes visualization of the sample datasets, which can be interpreted by all individuals. Datasets which all people can understand. The statistical outcome of analyzing the data shows the likelihood of each statement in the dataset having various types having toxic levels.The statistical result also provides further visualization of closest meaningful words. Depending on the classification of harmful, extreme harmful, offensive, personality hatred, danger and provocation, a statistical review is carried out.



VI. CONCLUSION AND FUTURE WORK

In this paper , the problem of posting toxicity comments on the Wall of Wikipedia has been regulated. Using the T-distributed Stochastic Neighbor Embedding(t-SNE) algorithm, the high dimensional data was

visualised. In one specific statement, we can estimate the probability of toxic categories. In doing so, we can see the vulgarity level inside a group of comments.Yet existing models only classify negative comments. This work also aims to capture the kinds of toxic comments. Reduction of Dimensionality is used to reduce statement dimensions. Bidirectional Short term memory model is used to classify the sentences into appropriate toxicity categories.In our method we achieve the 98.9012 level of accuracy. We plan to analyze the technique of emotional classification such as Sad, joy, fear , anger, Surprise with the toxic comments in future study. We intend to incorporate not only the framework of the Wikipedia but also of the social media site.

REFERENCES

1. Revati Sharma, Meetkumar Patel.September 2018. Toxic Comment Classification Using Neural Networks and Machine Learning:International Advanced Research Journal in Science, Engineering and Technology 5,9
2. Betty van Aken, Julian Risch, Ralf Krestel, and Alexander Löser.2017. Challenges for Toxic Comment Classification: An In-Depth Error Analysis.
3. Mujahed A. Saif, Alexander N. Medvedev, Maxim A. Medvedev, and Todorka Atanasova. 11 December 2018. Classification of Online Toxic Comments Using the Logistic Regression and Neural Networks Models: AIP Conference Proceedings 2048.
4. Spiros V. Georgakopoulos, Sotiris K. Tasoulis, Aristidis G. Vrahatis, and Vassilis P. Plagianakos. 2018. Convolutional Neural Networks for Toxic Comment Classification. In SETN ’18: 10th Hellenic Conference on Artificial Intelligence, July 9–12, 2018, Patras, Greece. ACM, New York, NY, USA.
5. Björn Gambäck and Utpal Kumar Sikdar. 2017. Using convolutional neural networks to classify hate speech. In ALW1@ACL.
6. Anna Schmidt and Michael Wiegand. 2017. A survey on hate speech detection using natural language processing. In SocialNLP@EACL.
7. Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts Xingyou Wang1 , Weijie Jiang2 , Zhiyong Luo3. http://www.aclweb.org/anthology/C16-1229