Automated Cyberbullying Detection using Clustering Appearance Patterns

This paper enhanced the Naïve Bayes classifier for extracting the words and examining loaded pattern clustering. The algorithm included two main methods: (1) creating partitions by iteratively relocating from entire datasets into clusters using k-mean clustering and (2) capturing any specific partition with the frequency of words with multinomial model feature vector and drawing the probability of words occurring in a document for predicting the eight classes. The proposed method resulted in increasing accuracy and reliability of an experiment. Texts are fed into cluster and discriminant analysis stage which is able to identify abusive texts. The abusive texts are then clustered by using K-Mean.

Naïve Bayes is used as classification algorithms to build a classifier from our training datasets and build a predictive model. Moreover, we also used Naïve Bayes to classify the abusive texts into one of the eight pre-defined categories. The categories include activities approach, communicative, desensitization, compliment, isolation, personal information, reframing, and relationship.

**Preprocessing**

Our approach consists of two main methods. The first method aims to clean and pre-process our datasets by removing non-printable and special characters, reducing the duplicate words and clustering the datasets. The second one concerns classification model to predict the text messages for preventing

Cyberbullying.

**Dataset**

The datasets of streaming texts are collected from two different types of sources: (1) posted messages by members in Perverted-justice used as training datasets, and (2) Twitter datasets from Stanford University as testing datasets. The main goal of our research is to improve accuracy of our automatic cyberbullying detection system and to process and analyze texts in nearby real time

The concept of a data stream is more suitable than a dataset. Generally, a stored dataset is a proper model when significant portions of the data are queried repeatedly, and updates are small and quite rare. In the opposite, a data stream is a fit model when a huge volume of data is arriving uninterruptedly and it is unreasonable to store the data in some forms of memory. In addition, data streams are applicable as a model of access to big datasets stored in secondary memory in which performance requirements need access through linear scans

**Classifier**

A Naïve Bayes classifier uses the concept of probability to classify large volume of data by finding models that differentiate classes of data .Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other features. The model can be applied to several kinds of problems, such as real time prediction, multi class prediction, and text classification. A Naïve Bayes technique uses machine learning and data mining techniques to filter unseen information.

**Clustering**

Clustering is a data mining technique to group data considered by data similarity. Each group of data is clustered based on the distance measured from data center and the deviation measurement. The popular clustering model is namely K-means. The clustering process is an unsupervised learning technique in which the clustering data is grouped without the target.

**Project overview**

First step, preparing dataset, data sources comes from the Cybercrime Data and Twitter across cluster networks from data streams.

Second step, generating clusters, data sources are clustered the features two categories of the messages as polite messages and abusive messages, which the contents of the messages are identified, based on a crime pattern and the normalized documents using K-means clustering technique. The K-means

Clustering is a method of vector quantization, originally from signal processing, that is widespread for cluster analysis in data mining. It aims to do a partition of n observations into k clusters in which each observation belongs to the cluster with the nearest mean for finding the word pattern frequency with

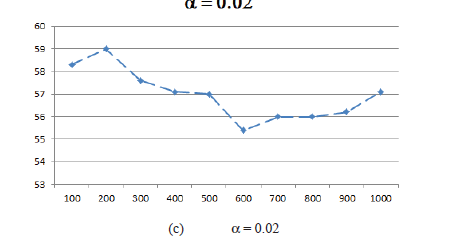
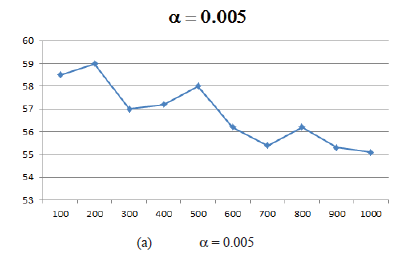
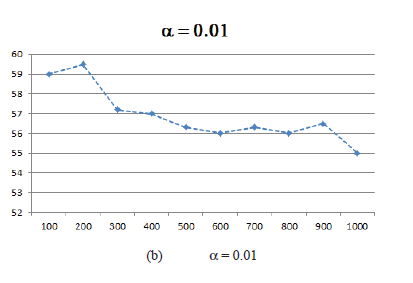
N-Gram in each sentence.

Third step, training data, the abusive partition is used to transform each into the feature extractor for classification technique using on Naïve Bayes. They assig the feature extractor into eight sub categories including; activities approach, communicative, desensitization, compliment, isolation, personal information, reframing, and relationship. The pairs of feature sets and labeled data are fed into the Naïve Bayes classifier to produce a model.

Final step, predicting data, the feature sets are fetched into the model to produce predicted labels in eight categories.

**Results**

They conducted a maximum number of iterations of 20 (to make a fair comparison) for this entire algorithm. Each experiment is running ten times. We set the threshold of relevant strength between two words to 0.4. For each sentence, we did training the data to optimize the sentiment word strengths that covered spelling correction, negating word list, and repeating letters or punctuation and emotions. The Kmeans clustering was used to split data sources into an abusive cluster.



From Figure 2 (a), (b) and (c), the maximum number of iterations is 100 and the feature set is 200 features. There are three values being tested in this works, which are 􀄮 = 0.005, = 0.01 and 􀄮 = 0.02. Naïve Bayes was used for outperforming without the extended feature set by a lower margin of 0.7%. It performs less well than N-gram *(“I love* *you”, “I love”, “love you” and “I”, “love”, you”)* with adding the pattern terms. Naïve Bayes performs the classification accuracy = 95.79% is already very high when using all features. All except accuracy are statistically significant difference. In terms of 􀄮 value, 0.02 tends to perform almost uniformly better than other values for this data set. When 􀄮 =

0.02, relevant measure will contribute more than redundant measure, which means that the extracting method will focus more on improving classification accuracy than reducing the number of features.

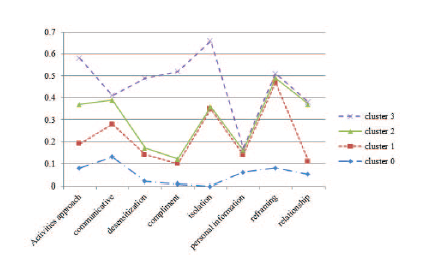


Figure 3 shows the results of a dataset that was collected during 2012-1013 and categorized into eight classes based on four clusters. We found that *k=4* produced the best results that follow from the Cyberbullying type (Child Exploitation, Money Laundering, Internet-based Drug Trafficking, and Mass Marketing Fraud). Based on the results in Figure 3, the isolation, communicative, and reframing were the three highest score on normalized activity count. For example, both of the contributions "are you alone" and "do you have many friends" are considered under the category of “isolation.”

The clustering step can reduce the amount of time needed to classify all the discussions,

and for the learning step they used a predefined corpora containing posted messages

identified by members of the Perverted Justice Foundation. Processes of the architecture

are distributed and orchestrated by the Apache Hadoop Yarn Ecosystem, and implemented

using the Apache Mahout libraries. Still, as stated the clustering process is not parallelized.

In addition, the use of K-means performs poorly on categorical data

Also, their study

did not consider other languages besides english which is also the language of the training

set. This point can limit the identification of other bullying discussions because bullies tend

to use more than often foreign words.