

Type of Supervised Text Classification System for Unstructured Text Comments using Probability Theory Technique

Sreedhar Kumar S, Syed Thouheed Ahmed, Nisha Bai, Vinutha B A

Abstract: In this paper, an improved sentimental text analysis system called Probability Based Text Classifier (PBTC) is presented. It aims to train the existing unstructured text command set and to classify the sampled text command belongs into positive or negative polarity based on probability theory and supervised concepts. It consists of three stages pre-processed, training and classification. In the first stage, the proposed (PBTC) system identifies the relevant and irrelevant words in the unstructured text command set based on pre-determined text pattern model. In the second stage it identifies two dissimilar classes over the preprocessed text command set based on predetermined text pattern model and simple probability theory concepts. Next stage, the PBTC identifies the sample test text command without class label belong on which class based on Naive Bayer scheme and trained existing text command set. Experimental result shows that the proposed (PBTC) system is well suitable to train the unstructured text command set and classify the new text command belongs into positive or negative polarity with higher accuracy

Keywords: Classifier; Probability Based Text Classifier; Unstructured Text Command; Text Pattern; Supervised; Sentimental Analysis.

I. INTRODUCTION

Generally, the text classifier is a supervised concept that uses to identify distinct polarities such as positive or negative over the unstructured text commands or user review from the different online sources namely social media networks (facebook and twitter), online product sales and online services such as education, health, entertainment, travels etc., for deeper analysis and predication for future business development. Today's the social media networks plays a big role in the world society. For the simple example of the social media networks namely Twitter and facebook that are provided very big platform for the people to raise their views on various topics, current issues and things going on in the world^[1]. The system called sentiment analysis that uses to identify the polarity of a given unstructured text in the document, sentence, or feature and the expressed opinion in a document, a sentence or an

Revised Manuscript Received on September 22, 2019.

Sreedhar Kumar S*, Department of Computer Science and Engineering, Dr. T Thimmaiah Institute of Technology, VTU, Kolar Golad Field, Karnataka, India

Syed Thouheed Ahmed, Department of Computer Science and Engineering, Dr. T Thimmaiah Institute of Technology, VTU, Kolar Golad Field, Karnataka, India

Nisha Bai, Vinutha B A, Department of Computer Science and Engineering, Dr. T Thimmaiah Institute of Technology, VTU, Kolar Golad Field, Karnataka, India

*Corresponding author; *Email: sree.dr.2018@gmail.com

entity feature is positive, negative, or neutral. For a good example of the polarities that the sentiment classification looks, for instance, at emotional states such as "wrong", "enjoy", and "poor" etc.. The main application of the sentiment analysis is that to process and classify the customer related materials likely consumer reviews and survey responses, online social media and healthcare information for applications that range from marketing to customer service to clinical medicine based on supervised and unsupervised methods^[2,3,4,5]. This paper is presented a type of sentiment analysis system called probability based text classifier (PBTC) and it intentions to train the existing online unstructured user reviews or commands of consumer product and classify the recent user review belongs into which polarity. This paper is organized as follows: related work is discussed in Section 2. Section 3 contains details of the proposed PBTC system. The result and discussion is presented in Section 4. Conclusion drawn is in section 5.

II. RELATED WORK

In the last decades many of the authors reported distinct text classifiers techniques that intentions to analysis and categorize the unstructured text commands from different sources likely social media network, consumers commands of online product and reviewers commands of different online services such as movie, travels, bank, education, medication etc., which belongs to positive or negative polarity for future predication and business development. The comparative analysis is made by Heba et al.^[5] on sentiments analysis to exhibit how effective they are in missing sentimental data in twitter under various experimental environments. Studying purpose the STS (Standard Testing Sentiment Data base) is used. Experimental results indicate that naïve Bays for multinomial data perform other classification in tweet sentiment analysis. In^[6] the authors Mehra et al. have presented a procedure that used to analysis and categorize the user tweets into positive or negative class based Naïve Bayes and fuzzy classifiers. The author presented an experimental result by evaluating datasets and corresponding classification results shows that proposed system gives more efficiency in precision, Accuracy and Recurrence in classifying the tweets.

The algorithm Specific Action Rule Discovery on Grabbing Strategy (SARGS) is used by Jaishree Ranganathan et al. in^[7] to perform analysis of many meta actions generated for implementing an algorithm in spark driven system made by Apache and traditional Hadoop system. Both techniques use the twitter

social media data, find the results to perform sentimental analysis on corpus data of social network and time taken by both strategies is measured and their sub components of data processing. Spark system cautions fast compared to using meta action generating methods implemented by Hadoop MapReduce. The authors Trupthi et al.^[8] reported an automatic system based on interactive approach that used to predict the sentiments of the reviewers or tweets presented by people in social networks using Bigdata tools like hadoo. A comprehensive method is required to predict sentiment polarity, which supports to promote marketing and concentrate on sentimental analysis techniques, Sentiment classification based on feature and summarizing the opinion.

In this setup Naïve bays is less affecting by data thinning. An hybrid approach is used by Akshi Kumar et al.^[10] for corpus and dictionary methods to control orientation of the words in the tweets. In fourth year of the twitter sentimental analysis SemEval–2016 contain 4 tasks which comprise five sub tasks, significant content has been taken from the previous edition in three sub tasks. Beginning two sub tasks are returned from previous years and overall sentiments can be predicted, as well as the sentiment contained in the tweet topic. The three subtasks concentrate on basic sentiment using two variants in twitter task. Five point scale is used in first variant, here ordinary character is used for classification of tasks. Focus of the second variant is proper estimation of the effects of each class of interest. This task is called quantification in supervised learning technique. Review classification of customer's sentiments are analyzed by Geetika Gautam et.al^[12] it is helped for analyzing the information according to number of twitter tweets when customer opinions are mainly unstructured and there are there may be positively or negatively or in between of positive or negative. Pre-processed dataset is used to extract adjectives from data that contain some useful meaning called features, the extracted features list is used for applying machine learning oriented algorithms which used classification techniques called Naïve Bays, Entropy and support vector machine having word set made by semantic orientation and similarity of the content features. The performance of classifiers is measured based on precision, accuracy and recall parameters.

Machine learning based approach is designed by Abdeljalil et al.^[13] have implement and analyses sentiments in data. They have carried out twitter data sentiment classification, Two types of model have to be used: feature based, tree kernel based model and both models are used to perform analysis of features, which shows that very important features will combine the pre polarity in words and parts of speech in those words. Pre polarity POS specific features are introduced in Apporv Agarwal et.al^[14], it explores the tree kernel usage to eliminate the necessity of complex feature engineering. The modern feature and tree based kernel performs approximately same at that level, both perform at same baseline. It clearly shows procedure to connect to sentimental data in twitter and do the analysis using queries, experiments are executed on different kind of data like politics, humanities etc and applications will show the interesting outcomes. It proved that the normal sentiments exhibited by users in tweets are high.

Vishal A Kharde et al.^[15] concentrates on analyzing

sentimental data in twitter that is useful to do analyse on tweets information, Here user opinions are unstructured in nature and also heterogeneous, comparative analysis is performed after survey on existing techniques for mining opinion data including lexicon based and machine learning based strategies are used with evaluation parameters. Different machine learning algorithms are use like Entropy, Naïve Bayes and Support Vector Machine (SVM). Twitter sentiment classification technique using word embedding is developed by Duyu Tang et al.^[16] Sentiment Specific Word Embedding (SSWE) is used for encoding sentiment data in the sentence containing words. Methods for applying Sentiment Specific Word Embedding to a standard twitter data set classification shows that features in SSWE perform comparatively with handmade features found in top rated systems and by concentrating more on SSWE feature set performance can be improved further.

The twitter tweets on electronic consumables includes mobile, PC laptop etc., are analyzed by Neethu and Rajashree^[17] using an approach called machine learning. Performing analysis of sentiment data in the specific domain, it is required to find out the effectiveness of domain data in classifying the sentiment, a new feature list that is performed for classification of tweets according to positive, negative and exploit customer's opinion on produces.

The customer opinion on hotels is classified by Dietmar Grabner et al.^[18] using sentiment analysis explained the process used to extract the lexicons which are domain specific in nature. The lexicons are semantically related words use some corpus. The resultant lexicons used for sentiment analysis for producing the review classification. The standard tools for analyzing twitter sentiment data are used in^[19,20] different tools are incorporated for applying to 5 various test beds. The performance evaluation results are presented, along with an comprehensive analysis of errors are also used to spot the most important challenge faced by analysis tools of twitter sentiment data. The outcomes have good amount of suggestions for different stakeholder groups that include researchers of social media analysis, developers of NLP, managers of industry and professions using sentiments of social media site as input for making decisions. Another authors Jaishree Ranganathan et.al^[20] have reported to performed a comparative analysis of meta-actions generating algorithmic implementation in Apache Spark driven system and conventional Hadoop driven system using the Twitter social networking data and evaluate the results and performed corpus based sentimental analysis of social networking data, and test the total time taken by both the systems and their sub components for the data processing.

III. PROPOSED PBTC APPROACH

This section detail of the proposed system called Probability Based Text Classifier is described. It intentions to train the existing unstructured text commands and identifies the polarity or class of recent text command with finite number of words. The proposed PBTC system contains three stages, first stage, it pre-processes the



existing unstructured input text command set that collected from Social Media and Online Service through internet based on predetermined text pattern model. Next stage, it trains or categorizes the pre-processed unstructured input text command set into two dissimilar polarities likely positive and negative based on probability theory and predetermined polarity pattern model. In the last stage, the PBTC identifies the recent unstructured test text command without polarity that belongs into which class (Positive or Negative) based on Naïve Bayes classifier and trained text command set. The stages involved in the proposed PBTC system are illustrated in Fig 1.

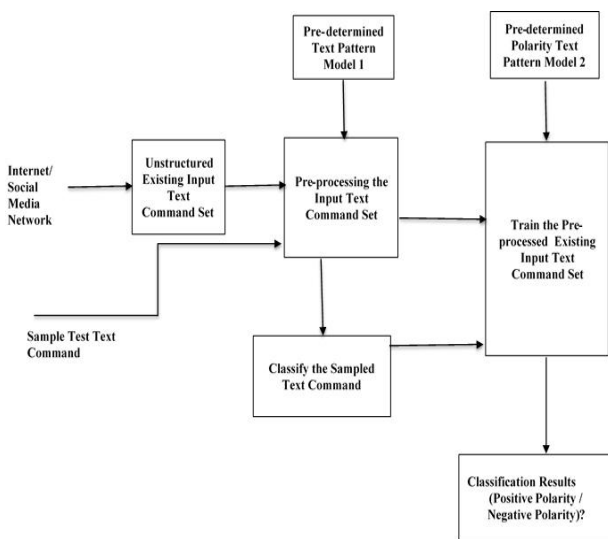


Fig. 1: Functional Diagram of Proposed PBTC System

3.1 Pre-processing Stage

In this stage, the proposed PBTC system, reduces number of words in each individual text command in the input text command set through the process of identifying the relevant (Subjective and Objective) and irrelevant (articles, proposition, connection) words and remove the irrelevant words based on predetermined pattern model. The pre-processing consists of two steps, in the first step, it collects the unstructured input text command set (T) from specific source likely social network or online user commands and is defined as $T = T_i$ for $i = 1, 2, \dots, n$, where T is the unstructured exiting input command set, n denotes the number of commands belongs to the input command set T and T_i represents the i^{th} command in the input command set. In the second step, it identifies the relevant and irrelevant words over each individual command in the input command set T based on predetermined pattern model 1 (PPM1) and simple existing word matching technique. Thus removes the irrelevant words in each individual command in the input command set and is defined as $\bar{T} = \bar{T}_i$ for $i = 1, 2, \dots, n$, where \bar{T}_i denotes the i^{th} preprocessed command belongs in to pre-processed input text command set \bar{T} with n commands. The predetermined text pattern model (PTPM1) contains finite number of irrelevant

keywords likely verbs, adjective, articles, proposition and conjunction.

3.2. Training Stage

In this stage, the proposed PBTC system, categorizes the pre-processed input text command set \bar{T} into two different polarities namely positive and negative classes based on probability theory and predetermined polarity text pattern model (PPTPM-2). The PPTPM-2 contains two different predetermined polarities namely positive and negative with finite number of respective Adjective key words It consists of three steps, in the first step, it measures the probability of positive polarity word $P(\text{"Positive Words"}|\bar{T}_i)$ P over the each individual text command in the input text command set $\bar{T} = \bar{T}_i$ for $i = 1, 2, \dots, n$, where \bar{T}_i denotes the i^{th} text command belongs in to pre-processed input text command set \bar{T} with n commands and it defined in the Equation (1) as:

$$P(\text{"Positive Words"}|\bar{T}_i) = \left\{ \left(\frac{\text{Number of Positive Words in } \bar{T}_i}{|\bar{T}_i|} \right), \forall \bar{T}_i \in \bar{T} \right\} \quad (1)$$

Where, $|\bar{T}_i|$ denotes the total number of words in i^{th} text command that belongs into input text command set \bar{T} . In the second step, it estimate the probability of negative word $P(\text{"Negative Words"}|\bar{T}_i)$ among the each individual text command in the input text command $\bar{T} = \bar{T}_i$ for $i = 1, 2, \dots, n$ and is defined in the Equation (2) as:

$$P(\text{"Negative Words"}|\bar{T}_i) = \left\{ \left(\frac{\text{Number of Negative Words in } \bar{T}_i}{|\bar{T}_i|} \right), \forall \bar{T}_i \in \bar{T} \right\} \quad (2)$$

In the last step, it identifies the i^{th} text command \bar{T}_i in the input text command set \bar{T} which belong to positive or negative polarity based on probability of positive word and negative word could calculate by Equation (1) and (2). If the probability of positive word is higher than probability of negative word, then that the i^{th} text command \bar{T}_i is placed into positive polarity otherwise negative polarity. Finally, repeat the steps from step 1 to step 3 until identifies all the text commands and its respective polarity in the input text command set \bar{T} .

3.3. Classification Stage

This stage, the proposed PBTC system classifies the new unstructured sample text command that belongs into positive or negative polarity based on existing Naïve Bayes scheme [9] and trained text command set \bar{T} . It consists of four steps, in the first step, it pre-processes the unstructured new text command (t) without polarity based on predetermined text pattern model (PTPM-1) which already discussed in the subsection (III.A). Next, it counts the number of words belong in the positive class and negative class respectively with finite number of existing commands over the trained text command set $\bar{T} = \bar{T}_i$ for

$i = 1, 2, \dots, n$. In the third step, it computes the conditional probability $P(p|\bar{t})$ positive polarity (p) over the pre-processed new text command \bar{t} based on the trained text command set \bar{T} , where p is the positive class that belongs to the trained text command set $p \subseteq \bar{T}$ and is defined in the Equation (3) as:

$$P(p|\bar{t}) = P(p) \prod_{j=1}^m P(\bar{t}_j|p) \tag{3}$$

Where, $P(p)$ is the prior probability of positive polarity in the trained text command set \bar{T} and is defined in the Equation (4) as:

$$P(p) = \left\{ \frac{R}{|\bar{T}|} \mid \text{where } 0 > R < n \right\} \tag{4}$$

Here, R denotes the number of trained text commands $\forall \bar{T}_k \in \bar{T}$ for $k = 1, 2, \dots, R$ that belongs in to positive class p in the trained text command set $p \subseteq \bar{T}$, \bar{T}_k is the k^{th} text command in the trained text command set \bar{T} and $P(\bar{t}_j | |p|)$ represents the probability of j^{th} text word \bar{t}_j in trained new text command \bar{t} that belongs in to positive polarity p in trained text command set \bar{T} and $|p|$ describes number of text words in the positive polarity p in trained text command. Similarly, in the next step, it calculates the conditional probability $P(N|\bar{t})$ negative polarity (p) over the pre-processed new text command \bar{t} based on the trained text command set \bar{T} , where N is the negative class that belongs to the trained text command set $N \subseteq \bar{T}$ and is defined in the Equation (5) as:

$$P(N|\bar{t}) = P(N) \prod_{j=1}^m P(\bar{t}_j | |N|) \tag{5}$$

where, $P(N)$ is the probability of negative polarity in the trained text command set \bar{T} with n existing text commands and is defined in the Equation (6) as:

$$P(N) = \left\{ \frac{E}{|\bar{T}|} \mid \text{where } 0 > E < n \right\} \tag{6}$$

Here, E denotes the number of existing text commands $\forall \bar{T}_k \in \bar{T}$ for $k = 1, 2, \dots, E$ that belongs in to negative class N in the trained text command set $N \subseteq \bar{T}$, \bar{T}_k is the k^{th} existing text command in the trained text command set \bar{T} and $P(\bar{t}_j | |N|)$ represents the probability of j^{th} text word \bar{t}_j in new text command \bar{t} that belongs in to positive polarity p in trained text command set \bar{T} and $|N|$ describes the

number of text words in the negative polarity N in trained text command set. Finally, it identifies the recent text command \bar{t} that belongs to the positive or negative class based on conditional probability of \bar{t} in positive class and negative class could calculate by Equation (3) and (5). If the conditional probability of \bar{t} in positive class $P(\bar{t}|p)$ is higher than conditional probability of \bar{t} in negative class $P(\bar{t}|N)$ then that the recent text command \bar{t} would be placed into positive polarity in the trained text command set \bar{T} otherwise it placed into negative polarity.

3.4. Algorithm for PBTC

Input: Unstructured Existing Text Command Set $T = \{T_1, T_2, \dots, T_n\}$, Recent Text Command t , PTPM-1, PPTPM-2

Output: Trained Existing Command set $\bar{T} = \bar{T}_i$, Classification Result of Text Command (\bar{t})

Begin

1. Pre-process the unstructured existing text command set $T = T_i$ for $i = 1, 2, \dots, n$ based on predetermined text pattern model1 (PTPM-1) as described in the subsection (3.1).
2. Split the pre-processed existing text command set $\bar{T} = \bar{T}_i$, with n commands into two distinct classes likely positive (p) and negative (N) based on PPTPM-2 and simple probability theory Equations (1) and (2) as expressed in the subsection (3.2).
3. Counts number of words in positive polarity ($|p|$) and negative polarity ($|N|$) over the text commands in the trained command set (\bar{T}).
4. Computes prior probability of positive class (p) over the trained existing text command set $\bar{T} = \bar{T}_i$ using Equation (4).
5. Finds prior probability of negative class (N) over the trained existing text command set $\bar{T} = \bar{T}_i$ using Equation (6).
6. Calculates conditional probability of positive class (p) in pre-processed recent text command \bar{t} based on trained text command set $\bar{T} = \bar{T}_i$ by Equation (3)
7. Estimates conditional probability of negative class (N) over the pre-processed recent text command \bar{t} based on trained text command set $\bar{T} = \bar{T}_i$ by Equation (5)
8. Identifies the text command \bar{t} belongs to the positive or negative class based on results of conditional probabilities $P(p|\bar{t})$ and $P(N|\bar{t})$ as described in the subsection (III. C).
9. Obtained the recent text command \bar{t} with class label in the trained text command set (\bar{T}) and update its size.

End

IV. RESULTS AND DISCUSSIONS

This section describes the detail of experimental result of proposed PBTC system that tested over the sample existing online user review command set. For the experimentation purpose, more than 100 unstructured consumer review text commands with different size taken from online related to mobile phone [21] and shown 12 commands only in the Table 1. The Table 1 shows that the contents of unstructured user



text commands $T = T_i$ for $i = 1, 2, \dots, n$, and number of words belong in the each individual text command, where T is the unstructured exiting user text command set, n denotes the number of user commands belongs to the input command set T and T_i represents the i^{th} user command in the input command set.

Table 1 Unstructured Existing Input Text (Online User Mobile Review) Commands

Comment Id	Unstructured Text Command Set (T_i)	Number of Words in (T_i)
1	Phone battery is draining	4
2	Phone has good battery	4
3	Phone has worst camera	4
4	Phone has good camera	4
5	Display of phone is worst	5
6	fast performance with face id and wireless charging	8
7	Phone is more expensive	4
8	Fantastic dual lens camera	4
9	Battery backup is poor	4
10	Phone hardware is faulty	4
11	Iphone has a faster processor	5
12	Phone has bigger storage option	5

In the first stage, the PBTC system pre-processes the unstructured user text command set T with n commands that presented in the Table 1 based on the predetermined text pattern model (PTPM-1) which shown in the Table 2. The Table 2 contains finite number of irrelevant text patterns likely Verbs and Conjunctions. The pre-processed user text command set \bar{T} with n commands and its respective are obtained in the Table 3.

Table 2 Predetermined Text Pattern Model (PTPM-1)

Irrelevant Text Patterns
is, the, with, are, there, has, had, their, it, and, but, they, three, this, that, then, will, should, could, can, did etc.

Table 3 Pre-processed Existing Input Text Command Set

Comment Id	Pre-processed Text Command Set (\bar{T}_i)	Number of Words in $ \bar{T}_i $
1	Phone Battery Draining	3
2	Phone Good Backup	3
3	Phone Worst Camera	3
4	Phone Good Camera	3
5	Display Phone Worst	3
6	fast performance face id wireless charging	6
7	Phone more expensive	3
8	Fantastic dual lens camera	4
9	Battery backup poor	3
10	Phone hardware faulty	3
11	Iphone faster processor	3
12	Phone bigger storage option	4

Next, the PBTC partitions the pre-processed user text command set \bar{T} into two polarities namely positive (p) and negative (N) based on probability of positive or negative adjectives belong in each individual command through predetermined polarity text pattern model (PPTPM-2) which discussed in the subsection (III. B). Table 4 shows the constructed (PPTM-2) including finite number of positive and negative polarity words respectively. Finally, the PBTC has identified six commands belongs into positive polarity (p) and another six commands comes under the negative polarity (N) in the command set \bar{T} based on probability of positive $P(p)$ and negative $P(N)$ words in the commands and results as obtained in the Table 5. Table 5 shown that the proposed PBTC system is properly trained the text command set \bar{T} through identified each individual text command and its respective polarity (p or N) .

Table 4 Predetermined Polarity Pattern Text Model (PPTM-2)

Positive (p) Polarity Patterns	Negative (N) Polarity Patterns
good, Intelligent, accepted, beautiful, easy, genius, fantastic, super, wow, amazing, happy, joy, delight, appreciative, admirable, fast, worthy, successful, bigger,	worst, draining, lazy, arrogant, impatient, annoying, fake, poor, dull, depression, jealous, expensive, faulty, unsatisfactory, wired

Table 6 Recent Unstructured User Text Command

Comment Id	Unstructured Recent Text Command (t)	Pre-processed Recent Text Command (\bar{t})
13	Battery Backup is worst and poor	Battery Backup worst poor
14	Phone has working good and well	Phone working good well

Table 5 Result of Trained (Classified) Existing Text Command Set

Comment Id	Pre-processed Text Command Set (\bar{T}_1)	Probability of Sentimental Polarity in (\bar{T}_1)		Classification Result (Positive (p) / Negative (N))
		Probability of positive (p) in (\bar{T}_1)	Probability of Negative (N) (\bar{T}_1)	
1	Phone Battery Draining	0	0.33	N
2	Phone Good Backup	0.33	0	P
3	Phone Worst Camera	0	0.33	N
4	Phone Good Camera	0.33	0	P
5	Display Phone Worst	0	0.33	N
6	fast performance face id wireless charging	0.166	0	P
7	Phone more expensive	0	0.33	N
8	Fantastic dual lens camera	0.25	0	P
9	Battery backup poor	0	0.33	N
10	Phone hardware faulty	0	0.33	N
11	Iphone working well	0.33	0	P
12	Phone bigger storage option	0.25	0	P

Table 7 Classification Result of Recent Text Command in Table

Comment Id	Pre-processed Recent Text Command (\bar{t})	Number of Words in (\bar{t})	Number of Words in Positive (p) Class in (\bar{T}_1)	Number of Words in Negative (N) Class in (\bar{T}_1)	Prior Probability of P(p)	Prior Probability of P(N)
13	Battery Backup worst poor	4	23	18	0.5	0.5
14	Phone working good well	4				

Table 8 Classification Result of Recent Text

Comment Id	Pre-processed Recent Text Command (\bar{t})	Number of Words in (\bar{t})	Prior Probability of P(p)	Prior Probability of P(N)	Conditional Probability of P(p \bar{t})	Conditional Probability of P(N \bar{t})	Classification Result (p/N)
13	Battery Backup worst poor	4	0.5	0.5	0.0	0.000037	N
14	Phone working good well	4			0.000013	0.0	p

In the last stage, the PBTC is tested sample recent user text commands (t) that belongs into which polarity (p or N) based on existing naive bayes scheme and trained text command set which obtained in the previous stage. Table 6 shows the sample unstructured recent text command (t) and its respective pre-processed



version (\bar{t}) . Initially, it counts the size of sample commands (\bar{E}) and number of words belong in each individual polarity in the trained command set (\bar{T}_1) . Next, it computes the prior probability of positive $P(p)$ and negative $P(N)$ polarities over the trained text command set (\bar{T}_1) and the results as obtained in the Table 7. After, it calculates the conditional probability of positive polarity $P(p|\bar{E})$ and negative polarity $P(N|\bar{E})$ in sample pre-processed text command (\bar{t}) and the results are obtained in the Table 8. Finally, the sample recent text command is classified into its respective polarity based on conditional probability of sentimental polarities in the command and the classification results are shown in the table 8. Experimental results shown in the Table 5, Table 7 and Table 8 that the proposed PBTC is well suitable to train the unstructured existing online user review text command set and classify the recent user review command into its respective sentimental polarities.

V. CONCLUSION

This paper, an improved sentimental text analysis system called Probability Based Text Classifier (PBTC) is designed. It intentions to train or categorize the unstructured existing text command set into two distinct classes likely positive and negative and also it classifies the sample recent text command into its respective sentimental polarity based on probability theory and supervised concept. First stage, it finds the relevant and irrelevant words in the unstructured user review text command set based on pre-determined text pattern model and then remove the irrelevant words. Second stage, it identifies two distinct sentimental polarities likely positive and negative over the preprocessed text command set based on predetermined polarity text pattern model and probability theory method. Last Stage identifies the sample recent text command without polarity label belong into which class based on Naive Bayes scheme and trained text command set. For the experimentation purpose, more than 100 unstructured consumer review text commands with different size taken from online related to mobile phone and tested by proposed PBTC system. Experimental result shown that the proposed (PBTC) system is well suitable to train the existing online user review text command set and classify the recent user review command into its respective sentimental polarities

REFERENCES

- Hobbs Jerry R, Walker Donald E, Amsler Robert A, "Natural Language Access to Structured Text," Proceeding of the 9th conference on computational linguistics (COLING'82), PP. 127-132, 1982.
- https://en.wikipedia.org/wiki/Sentiment_analysis
- Jayasanka, Sachira & Madhushani, Thilina & R. Marcus, E & A. A. U. Aberathne, I & Premaratne, Saminda. "Sentiment Analysis for Social Media", Information Technology Research Symposium, Volume: 4, 2013.
- Krishnamoorthy R, Sreedhar Kumar S, "An improved agglomerative clustering algorithm for outlier detection ", Applied Mathematics and Information Science, vol. 10, no. 3, pp. 1141-1154, 2016.
- Sreedhar Kumar S, Madheswaran M, Vinutha B A, Manjunatha Singh H, Charan K V, "A brief survey of unsupervised agglomerative hierarchical clustering schemes," International Journal of Engineering & Technology (UAE), vol. 8, no. 1, pp. 29-37, 2019.
- Ismail, Heba M., Saad Harous and Boumediene Belkhouche. "A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis", Research in Computer Science, vol. 110, pp. 71-83,

- 2016
- Mehra R, Bedi M K, Singh G, Arora R, Bala T and Saxena S, "Sentimental analysis using fuzzy and naive bayes", International Conference on Computing Methodologies and Communication (ICCMC), Erode, pp. 945-950, 2017.
- Ranganathan, Jaishree & S. Irudayaraj, Allen & Tzacheva, Angelina. "Action Rules for Sentiment Analysis on Twitter Data Using Spark", IEEE International Conference on Data Mining Workshops (ICDMW), 2017, DOI: 10.1109/ICDMW.2017.14.
- Trupthi M, Pabboju S, Narasimha G, "Sentiment analysis on twitter using streaming API", IEEE 7th International Advance Computing Conference (IACC), 2017.
- Ragupathy R, Lakshmana Phaneendra Maguluri, "Comparative analysis of machine learning algorithm on social media test," vol. 7, no. 2.8, pp. 284-290, 2018.
- Akshi Kumar, Teeja Mary Sebastian, "Sentiment Analysis on Twitter", International Journal of Computer Science Issues, vol. 9, no. 4, pp 372-378, 2012
- Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F. and Stoyanov, V. "SemEval-2016 task 4: Sentiment analysis in Twitter", In Proceedings of the 10th international workshop on semantic evaluation (semeval-2016) .pp. 1-18, 2016.
- Gautam, Geetika & Yadav, Divakar, "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis", 7th International Conference on Contemporary Computing, 10.1109/IC3.2014.6897213.
- Abdeljalil Elouardighi, Mohcine Maghfour, Hafdalla Hammia, Fatima-zahra Aazi. "A machine Learning approach for sentiment analysis in the standard or dialectal Arabic Facebook comments", 3rd international conference on Cloud Technologies and Applications (CloudTech), 2016.
- Apoorv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambow, Rebecca Passonneau, "Sentiment analysis of twitter data". Proceedings of the Workshop on Language in Social Media (LSM 2011), pp. 30-38, 23rd June 2011.
- Vishal.A.Kharde, Prof. Sheetal.Sonawane, "Sentiment analysis of twitter data: a survey of techniques", International Journal of Computer Applications , pp 5-15, DOI:10.5120/ijca2016908625, April 2016.
- Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B. "Learning sentiment-specific word embedding for twitter sentiment classification", Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 1555-1565, 25th June 2014.
- M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, pp. 1-5, 2013.
- Grabber D., Zanier M., Fliedl G., Fuchs M, "Classification of Customer Reviews based on Sentiment Analysis", Information and Communication Technologies in Tourism , pp 460-470, 2012.
- Abbasi, Ahmed, Ammar Hassan and Milan Dhar. "Benchmarking Twitter Sentiment Analysis Tools." Language Resources and Evaluation Conference, 2014.
- Ranganathan, Jaishree & S. Irudayaraj, Allen & Tzacheva, Angelina., "Action Rules for Sentiment Analysis on Twitter Data Using Spark", IEEE International Conference on Data Mining Workshops (ICDMW), pp 51-60, 2017. DOI: 10.1109/ICDMW.2017.14.
- Ahmed, Syed Thouheed, M. Sandhya, and Sharmila Sankar. "A Dynamic MooM Dataset Processing Under TelMED Protocol Design for QoS Improvisation of Telemedicine Environment." *Journal of medical systems* 43, no. 8 (2019): 257.
- Ahmed, Syed Thouheed, M. Sandhya, and Sharmila Sankar. "An Optimized RTSRV Machine Learning Algorithm for Biomedical Signal Transmission and Regeneration for Telemedicine Environment." *Procedia Computer Science* 152 (2019): 140-149.
- Patil, Kiran Kumari, and Syed Thouheed Ahmed. "Digital telemammography services for rural India, software components and design protocol." In *2014 International Conference on Advances in Electronics Computers and Communications*, pp. 1-5. IEEE, 2014.
- Thouheed, Syed, S. Ahmed, M. Sandhya, and S. Shankar. "ICT's Role in Building and Understanding Indian Telemedicine Environment: A Study." In *Information and Communication Technology for Competitive Strategies. Lecture Notes in Networks and Systems, vol 40*. Springer Singapore, 2019.