International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 13
October 2017

Implementation of Classification Technique Framework for Sentiment Analysis

K.SRAVANA KUMARI

Lecturer, Government Degree College,affiliated to Satavahana University,Telangana,India

ABSTRACT: Sentiment Analysis (SA) is an ongoing field of research in text mining field. SA is thecomputational treatment of opinions, sentiments and subjectivity of text. Sentiment analysis is for this reason used to categorise these statements asa high-quality one or a bad one. There are diverse advantages of Sentiment Analysis. It makes the consumer aware ofthe diverse superb and terrible functions of any product. It allows the users in powerful decision making.Furthermore, SA facilitates agencies to searching for comments from those evaluations and alleviate their products/serviceswherever necessary. For instance, while someone plans to buy a mobile phone, he tends to examine a couple ofoverview web sites to study the critiques that the alternative clients have written. In this way, the consumer can get anconcept approximately the capabilities that he may additionally recall as crucial.

KEYWORDS-Sentiment analysis, Opinion mining, Feature extraction, pre-processing, pattern generation.

I. INTRODUCTION

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational take a look at of human beings's reviews, attitudes and emotionsin the direction of an entity. The entity can constitute individuals, activitiesor topics. These topics are most probably to be protected by usingevaluations. The expressions SA or OM are interchangeable.They express a mutual which means. However, a few researchersstated that OM and SA have barely distinctive notions [1].

Opinion Mining extracts and analyzes people's opinion aboutan entity at the same time as Sentiment Analysis identifies the sentimentexpressed in a text then analyzes it. Therefore, the target ofSA is to discover reviews, pick out the emotions they express,after which classify their polarity as proven in Fig. 1.Sentiment Analysis can be taken into consideration a classification method as illustrated in Fig. 1. There are 3 essential categorytiers in SA: record-degree, sentence-stage, and thing-stageSA. Document-degree SA objectives to classify an opinion reportas expressing a high-quality or terrible opinion or sentiment.Itconsiders the entire document a simple data unit(talking approximately one subject matter). Sentence-degree SA goals to categorisesentiment expressed in each sentence. The first step is todiscover whether the sentence is subjective or goal. If thesentence is subjective, Sentence-stage SA will determinewhether or not the sentence expresses advantageous or poor evaluations.

Wilson et al. [2] have talked about that sentiment expressions are not necessarily subjective in nature. However, there is noessential distinction among document and sentence degreeclassifications because sentences are simply brief files [3].Classifying text at the file degree or on the sentence degreedoes not provide the necessary detail wanted evaluations on allfactors of the entity which is wanted in many applications,to gain these information; we need to visit the aspect level.Aspect-stage SA ambitions to categorise the sentiment with respect tothe specific aspects of entities. The first step is to identify theentities and their components. The opinion holders can give different evaluations for exceptional aspects of the same entity like thissentence ''The voice great of this cellphone isn't always excellent, however thebattery existence is lengthy''. This survey tackles the first kinds ofSA.

The data sets utilized in SA are an critical issue in this area.The major resources of statistics are from the product evaluations. Thesereviews are essential to the business holders as they are able to takeenterprise choices in keeping with the evaluation outcomes of customers'critiques approximately their products.

The opinions assets are mainlyassessment websites. SA isn't always most effective applied on product critiques however canadditionally be carried out on inventory markets [4,5], news articles, [6] orpolitical debates [7]. In political debates as an example, we may want tofigure out people's reviews on a certain election candidates orpolitical parties. The election outcomes also can be anticipated frompolitical posts. The social community web sites and micro-bloggingweb sites are considered a excellent supply of facts becausepeople share and speak their critiques about a certain subject matterfreely. They are also used as facts sources inside the SA method.

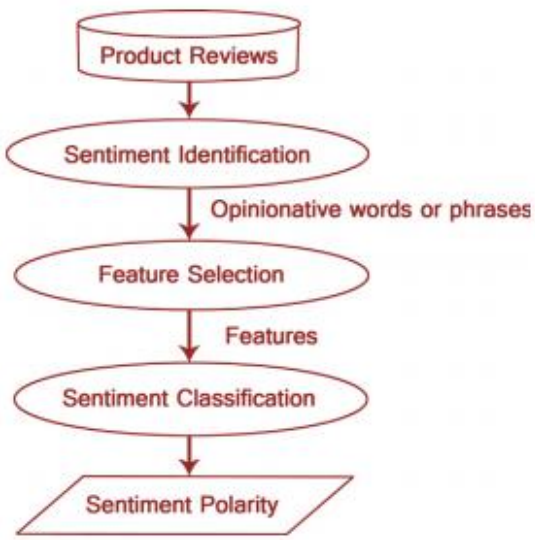

Figure 1 Sentiment analysis process on product reviews

There are many applications and enhancements on SAalgorithms that have been proposed inside the last few years. This survey objectives to present a closer look on these enhancements and tosummarize and categorize a few articles offered on this fieldin keeping with the numerous SA strategies. The authors have accumulated fifty-4 articles which provided important upgrades to the SA discipline recently. These articles cowl a hugestyle of SA fields. They have been all published inside the previous fewyears. They are categorized consistent with the target of the object illustrating the algorithms and facts used in their paintings.According to Fig. 1, the authors have discussed the FeatureSelection (FS) techniques in information together with their associatedarticles regarding some originating references. The SentimentClassification (SC) techniques, as proven in Fig. 2, arediscussed with greater info illustrating associated articles andoriginating references as well.This survey may be useful for brand spanking new comer researchers on thisdiscipline as it covers the maximum famous SA strategies and applications in a single studies paper. This survey uniquely gives asubtle categorization to the diverse SA strategies that'snot discovered in different surveys. It discusses also new associated fieldsin SA which have attracted the researchers currently and their corresponding articles. These fields encompass Emotion Detection(ED), Building Resources (BR) and Transfer Learning (TL).

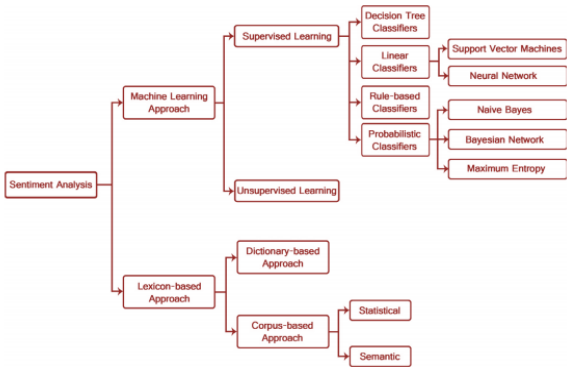

Figure 2 Sentiment classification techniques.

Emotion detection objectives to extract and analyze emotions, whilethe feelings will be explicit or implicit within the sentences.Transfer mastering or Cross-Domain classification is involvedwith analyzing information from one area after which the usage of theconsequences in a target area. Building Resources ambitions at creatinglexica, corpora in which opinion expressions are annotatedaccording to their polarity, and once in a while dictionaries. In thispaper, the authors supply a more in-depth appearance on these fields.There are severa quantity of articles presented every yearwithin the SA fields. The range of articles is increasing throughyears. This creates a want to have survey papers that summarize the recent studies traits and directions of SA.

The contribution of this survey is vast for many reasons. First, this survey provides sophisticated categorization ofa big quantity of latest articles in step with the techniquesused. This angle may want to

assist the researchers who are acquaintedwith positive strategies to use them in the SA field and pick outthe ideal method for a sure software. Second,the diverse techniques of SA are classified with brief detailsof the algorithms and their originating references. This canhelp new comers to the SA discipline to have a wide ranging viewat the entire subject. Third, the to be had benchmarks records unitsare discussed and categorized according to their use in sureprograms. Finally, the survey is superior with the aid of discussing theassociated fields to SA along with emotion detection, constructingassets and switch gaining knowledge.

II. BACKGROUND WORKS

This framework consists of three main steps [1]. The first step being data collection, followed by preprocessingof the data collected. The last step is the classification which categorizes the data processed into either positiveor negative. Fig. 3 gives the basic overview of sentiment analysis framework.

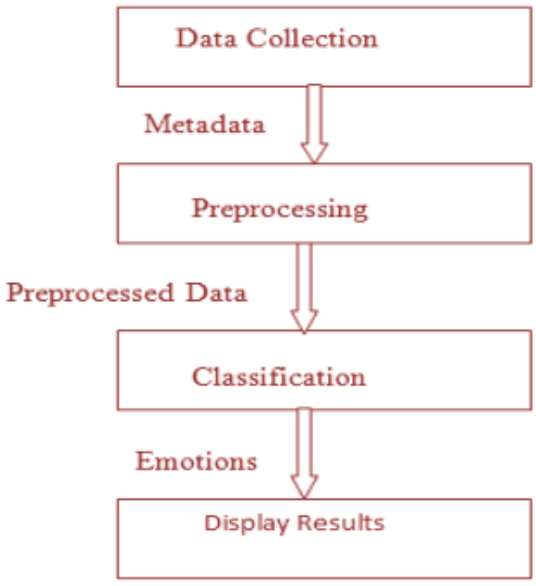

Figure3 Sentiment Analysis Framework [1]

A. Data Collection Sentiment Analysis: It can be done on any data. The data can either be collected fromany data set or can be extracted from any website. Data set is available online with thousands of reviews alongwith the label of positive and negative. On the other hand, extracting data from web is a lengthy task but one canperform sentiment analysis on the data of their own choice.

B. Pre-Processing Data: It can be extracted from the web contains several syntactic features that may not be usefuland therefore data cleaning and filtering needs to be done. In order to remove the unprocessed data, this stepneeds to be performed. It is imperative to preprocess all the data to carry out further functionalities. The variouspre-processing steps involved are given as below:

1) Removing URLs URLs are of no use while performing sentiment analysis and can sometimeslead to false analysis. For example "I have logged in to www.happy.com as I am bored... This sentence isnegative but because of there is one positive word in the url, it becomes neutral thus leading to a wrongprediction. To avoid the chances of false prediction, URLs must be removed.

2) Filtering Repeated letters in words like "thankuuuuu" are often used to show the depth ofexpression. However, these words are absent in the dictionary hence the extra letters in the word needs tobe eliminated. This is done on the basis of a rule that a letter cannot repeat itself more than three timesand if there is such letter that will be eliminated.

3) Questions Words like "what", "which", "how" etc., does not contribute to polarity and thussuch words must be removed in order to reduce the complexity.

4) Removing special characters In order to remove discrepancies during the Sentiment Analysisprocess, special characters like '[] {} 0/' should be removed. For example "it's good:" If these charactersare not eliminated before performing sentiment analysis, they will get combined with the words and thosewords will not be recognized. To avoid the situation, removal of such characters is important.

5) Removing Stop words and emoticons Stop words are words that should be excluded in orderto proceed with the SA process. Stop words don't carry as much meaning, such as determiners andprepositions (in, to, from, etc.) and thus needs to be filtered. Most of the times, while writing a review,people tend to use emoticons in order to express their feelings better. Although, these emoticons help inbetter

understanding of the emotions but while performing Sentiment analysis, this can mislead andpredict wrong.

6) Lemmatization or stemming Lemmatization and stemming aims to reduce inflectional andrelated forms of a word to a common base forms. Stemming achieves its goal correctly most of the timeby removing the ends of the words. Whereas, lemmatization does the same process properly with the useof a vocabulary and morphological analysis of words.

7) Tokenization refers to splitting the sentence into its desired constituent parts. It is an importantstep in all NLP tasks.

8) Feature selection it finds a reduced set of attributes that provides a suitable representation ofthe database given a certain analysis to be performed. This is necessary because the excessive use ofslangs, ironies and language mixtures makes the classification task easy.

C. Classification: It is a technique which classifies data into various categories. Classification is also used inthe field of Sentiment Analysis in order to classify data into three classes namely positive, negative and neutraland based on that the sentiment analysis process is completed. The classification task requires a pre-classifieddatabase sample, called training set, which is used to train and generate a classifier. It also helps in comparingnew unlabeled data to be classified. The classifier accuracy is highly dependent upon such training data. Thereare different classifiers available to perform the same and are discussed below, but Naive Bayes classifier is theone which is most commonly used for classification of data in Sentiment Analysis.

III. APPROACH

Many strategies have been used to investigate sentiments from dataset of various class andsize. The techniques used in the beyond are: Naïve Bayes algorithm, Support Vector Machine, Neural network andmany others. The experiments performed the use of those strategies have shown and proved their efficiency. Butthere are many open regions of research. Many researchers have used a hybrid method for the sentiment analysis.They have blended numerous algorithms to attain higher results. Akshi et al.Have used Neural Network toperform sentiment analysis on tweets. Neural Networks provide numerous advantages over other techniques. Theyhave distinguished functions like adaptive mastering, fault tolerance, parallelism and generalization.ANNs are able to studying and that they need to gain knowledge of. There are several mastering techniques :

Supervised Learning: It involves a trainer this is pupil than the ANN itself. For instance, the teacher feeds a few instance information about which the instructor already knows the solutions.For instance, sample spotting. The ANN comes up with guesses at the same time as recognizing. Then the teacherpresents the ANN with the solutions. The network then compares it guesses with the teacher's "correct" solutionsand makes adjustments in step with mistakes.

Unsupervised Learning: It is required whilst there may be no instance information set with recognized solutions. Forinstance, Searching for a hidden pattern. In this case, clustering i.e. Dividing a hard and fast of elements into companiesin line with some unknown pattern is performed based on the prevailing data units present.

Reinforcement Learning: this strategy built on observation. The ANN makes a decision by observingits environment. If the observation is negative, the network adjusts its weights to be able to make a differentrequired decision the next time.

In my proposed work we will work with supervised Learning model of ANN. For my dataset we will take aset of feedback given by users. The set of feedback can be labelled as: Excellent, Good, Average and Poor.The ANN needs a learning algorithm to effectively calculate weights for the neurons to fire. In this paper we planning to use an optimization algorithm for adjusting weights in the neural network. The algorithm that using in Firefly Algorithm. Firefly algorithm is a metaheuristic proposed by Xin-She Yang and inspiredby the flashing behavior of fireflies [11]. The pseudo code of the algorithm is given below.

In firefly algorithm, there are two important variables, which is the light intensity and attractiveness. Firefly is attracted toward the other firefly that has brighter flash than itself. The attractiveness is depended with the light intensity.

Firefly Algorithm

Objective function $f(\boldsymbol{x})$, $\boldsymbol{x} = (x_1, ..., x_d)^T$
Generate initial population of fireflies $\boldsymbol{x}_i$ $(i = 1, 2, ..., n)$
Light intensity I_i *at* $\boldsymbol{x}_i$ *is determined by* $f(\boldsymbol{x}_i)$
Define light absorption coefficient γ
while *(t <MaxGeneration)*
for $i = 1 : n$ *all n fireflies*
  for $j = 1 : n$ *all n fireflies (inner loop)*
  if $(I_i < I_j)$, *Move firefly i towards j;* **end if**
    Vary attractiveness with distance r via $\exp[-\gamma r]$
    Evaluate new solutions and update light intensity
  end for j
end for i
Rank the fireflies and find the current global best $\boldsymbol{g}_*$
end while
Postprocess results and visualization

Figure 4 Pseudo code for Firefly Algorithm (Yang, 2010).

The light intensity thus attractiveness is inversely proportional with the particular distance r from the light source. Thus the light and attractiveness is decrease as the distance increase.

$$I(r) = I_0 e^{-\gamma r^2} \quad (1)$$

I = light intensity,
I_0 = light intensity at initial or original light intensity,
γ = the light absorption coefficient
r = distance between firefly i and j

Attractiveness is proportionally to the light intensity seen by the another fireflies, thus attractiveness is β

$$\beta = \beta_0 e^{-\gamma r^2} \quad (2)$$

β_0 = Attractiveness at r is 0
The distance between two fireflies can define using Cartesian distance

$$r_{ij} = |x_i - x_j| = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (3)$$

Firefly i is attracted toward the more attractive firefly j, the movemeunt is defined as

$$\Delta x_i = \beta_0 e^{-\gamma r_{ij}^2} \left(x_j^t - x_i^t\right) + \alpha \varepsilon_i, \quad x_i^{t+1} + \Delta x_i \quad (4)$$

In equation (4), the first term is for attraction, γ is the limitation when the value is tend to zero or too large. If γ approaching zero ($\gamma \to 0$), the attractiveness and brightness become constant, $\beta = \beta_0$. In another word, a firefly can be seen in any position, easy to complete global search. If the γ is nearing infinity or too large ($\gamma \to \infty$), the attractiveness and brightness become decrease. The firefly movements become random. The implementation of firefly algorithm can be done in these two asymptotic behaviors. While the second the term is for randomization, as α is the randomize parameter. The ε_i can be replace by ran - 1/2 which is ran is random number generated from 0 to 1.

Our proposed algorithm will have the following steps:-
1) Collection of dataset
2) Pre-processing and cleaning of data
3) Calculate the relative occurrence of words in the dataset
4) Creation of neural network structure with each word being assigned a node
5) Re-adjusting the weights of the neural network by using the firefly algorithm
6) Calculation of the accuracy of the generated result

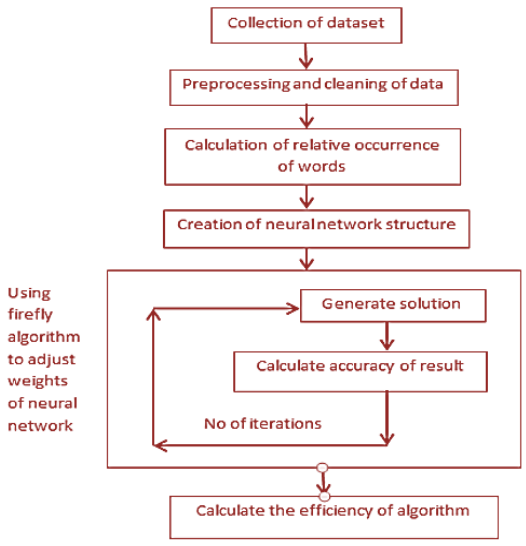

Figure 5 Proposed Sentiment Analysis Framework

IV. CONCLUSION

In this paper, it is been determined that sentiment analysis isthe efficient technique to examine the user behavior. Thesentiment analysis contains the four steps and in this workimprovement in the feature extraction phase is done usingthe pattern based technique.In this paper algorithm that we will be using in Firefly Algorithm. Firefly algorithm is a metaheuristic proposed by Xin-She Yang and inspiredby the flashing behavior of fireflies.

REFERENCES

[1] H. Sinha and A. Kaur, A Detailed Surveyand Comparative Study of Sentiment Analysis Algorithms .IEEE, 2016.

[2] Federico Neri, Carlo Aliprandi, Federico Capeci, MonstserratCuadros, Tomas, "Sentiment Analysis on social media",IEEE/ACM International conference on Advances in Social Networks analysIs and mining,2012,pp. 919-926.

[3] Ana C. E.S.Lima, Leandro N.de Castro., "Automatic Sentiment Analysis of twitter Messages", Publisher IEEE, 2012, pp.52-57.

[4] Min Wang, Hanxio Shi., "Research on Sentiment Analysis Technology and Polarity Computation of Sentiment words",Publisher IEEE, 2010, pp.331-334.

[5] ZHU Nanli, ZOU Ping, LlWeign, CHENG Meng., "Sentiment Analysis: A Literature Review", in proceedings of the 2012IEEE ISMOT, pp. 572-576.

[6] Seyed-Ali Bahrainian, Andreas Dengel., "Sentiment Analysis and Summarization of Twitter Data", IEEE 161hInternational Conference on Computational Science and Engineering, 2013,pp. 227-234.

[7] Seyed-AliBahrainian, Andreas Dengel., "Sentiment Analysis using Sentiment Features", IEEE/WIC/ACM InternationalConferences on Web Intelligence(WI) and Intelligent Agent Technology(lAT),20I3,pp. 26-29.

[8] Sunil Kumar Khatri, HimanshuSinghal, PrashantJohri., "Sentiment analysis to predict Bombay Stock Exchange UsingArtificial Neutral Network, Publisher IEEE,2014.

[9] Rui Xia, et al., "Dual Sentiment Analysis: Considering Two Sides of One Review", Transactions on Knowledge and DataEngineering, Vol. 27,No.8, Publisher IEEE, 2015, pp. 2121-2133.

[10] Vee W.LO, Vidyasagar POTDAR., "A review of opinion mining and Sentiment Classification Framework in SocialNetworks", 3rolEEE International Conference on Digital Ecosystems and technologies, 2009, pp. 396-40I.

[11] A. Kumar and R. Khorwal., "Firefly Algorithm for Feature Selection in Sentiment Analysis," SpringerLink, 2017.

[12] QiuGuang, He Xiaofei, Zhang Feng, Shi Yuan, Bu Jiajun, ChenChun. DASA: dissatisfaction-oriented advertising based onsentiment analysis. Expert SystAppl 2010;37:6182–91.

[13] Lu Cheng-Yu, Lin Shian-Hua, Liu Jen-Chang, Cruz-LaraSamuel, Hong Jen-Shin. Automatic event-level textual emotionsensing using mutual action histogram between entities. ExpertSystAppl 2010;37:1643–53.

[14] NeviarouskayaAlena, Prendinger Helmut, Ishizuka Mitsuru.Recognition of Affect, Judgment, and Appreciation in Text. In:Proceedings of the 23rd international conference on computational linguistics (Coling 2010), Beijing; 2010. p. 806–14.

[15] Bai X. Predicting consumer sentiments from online text.DecisSupport Syst 2011;50:732–42.