Feature Extraction for Sentiment Classification on Twitter Data

Amit G. Shirbhate[1], Sachin N. Deshmukh[2]

[1, 2]Department of Computer Science and Information Technology, Dr. B. A. M. University Aurangabad, Maharashtra, India

Abstract: *In the last couple of years, there has been a rapid growth in the use of social networking websites. It is a medium having a huge amount of information where users can view the other users opinion, that are classified into different sentiment classes, which are increasingly growing as a key component in decision making. In this paper, we take one such popular microblog called Twitter. These tweets sometimes express opinions about different topics. People post real time messages about their opinions on different topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. In this paper, we introduce a novel approach for automatically classifying the sentiment of "tweets" into positive, negative and neutral sentiment. Experimental evaluations show that our proposed techniques are efficient and perform better than previously proposed methods. In our research, we worked with English language however; the proposed technique can be used with any other language.*

Keywords: Classifier, Feature Selection, Microblogging, Sentiment Analysis.

1. Introduction

With the quick development of Web 2.0, users generate tremendous quantities of online information in the forms of reviews, blogs, tweets, etc. This abundance of data includes users" opinions about events, products, or people. It provides new opportunities for establishments and companies to understand their consumers and improve product quality as well as enhance their competitiveness [1].

Microblogging today has turned into an exceptionally well known specialized service among Internet clients. Every day, Millions of users share opinions on various aspects of life. Hence microblogging sites contain huge data for opinion mining and sentiment analysis. Authors of those messages write about their life, share opinions on variety of topics and discuss current issues. Due to a free format of messages and an easy availability of microblogging platforms, internet users tend to move from traditional communication tools (such as traditional blogs or mailing lists) to microblogging services. As large amount users post about products and services they use, or express their political and religious views, microblogging sites become valuable sources of people"s opinions and sentiments. Such information can be productively utilized for marketing or social studies.

Twitter currently receives over 190 million tweets and manufacturing companies receive over 10 thousand web products in a day, in which people share their opinion regarding a wide range of products, their features and services. Sentiment analysis also referred to as Opinion mining, is the field of study that analyses people"s opinion, sentiments, attitudes, evaluations, and emotions through social media data. These topics are most likely to be covered by reviews. Sentiment analysis in reviews is the process of investigating product reviews on the internet to determine the feeling or overall opinion about a product [2].
The objective of Sentiment Analysis is to find opinions, recognize the sentiments they express, and then classify with respect to their polarity. Sentiment Analysis can be considered as a classification process. There are three main classification levels in Sentiment Analysis: sentence-level, document-level, and aspect-level. Sentence-level aims to classify sentiment expressed in each sentence. The initial step is to identify whether the sentence is subjective or objective. When the sentence is subjective, Sentence-level Sentiment Analysis will figure out whether the sentence expresses positive or negative opinions. Document-level Sentiment Analysis aims to classify an opinion document as expressing a positive or negative opinion. Aspect-level aims to classify the sentiment with respect to the specific aspects of entities. The opinion holders can define different opinions for different aspects of the same entity like this sentence, "the voice quality of Samsung phones is not good, but the battery life is long"[3].

Different phases of opinion mining system are: opinion extraction, analysis, classification, scaling and summarization. Opinion extraction is the first phase of opinion mining process, which is performed after all opinions are loaded by user. Each new phase relies on previous one and output of previous phase will act as an input to next phase. Opinion extraction is first phase and all the other phases depend on it. If extraction is not relevant then opinion mining procedure will be diverted and not all the extracted features will be topic-related [4].

There are a few difficulties in Sentiment analysis. The first is a sentiment word which is considered as positive in one circumstance and negative in another circumstance. A second challenge is that people generally express opinions in a different way. Most of the conventional text processing depends on the way that little contrasts between two pieces of text don't change the meaning so much. In Sentiment analysis, however, "I'm very happy today" is far different from "Today, I'm not happy". People can be contradictory about their statements. Most of the reviews contain positive and negative comments, which is sometimes manageable by analyzing sentences only one at a time. However, in the more informal medium like twitter, people most likely to merge different opinions in the same sentence which is easy for a human to understand, but very difficult for a computer to parse [5], [6].

2. Related Work

There has been a wide range of research done on sentiment analysis, from rule-based, bag-of-words approaches to machine learning techniques. Two main research directions of opinion mining operate on either the document level or the sentence level. Both document level and sentence level classification methods are usually based on the identification of opinion words or phrases [7].

Sentiment analysis has been turned into a field of interest for many researches as number of blogs and social networks increased rapidly. In [8], a very broad outline of the existing work was introduced in which, the authors describe existing methods and approaches for an opinion-oriented information retrieval. In [9], authors used blogs to constructcorpora for sentiment analysis and also used emotion symbols for allocating to blog posts as indicators of users" mood. They used SVM and CRF classifiers to classify sentiments at the sentence level and afterwards determined the overall sentiment of the document, by assuming the sentiment of the last sentence of the document as the sentiment at the document level. For the sentiment classification, author in [10] collected texts containing emoticons from Usenet newsgroup,in which emoticons are used to create a training set. The dataset was divided into "positive" and "negative" having the texts with happy emoticons and sad or angry emoticons respectively. They used SVM and Naive Bayes to train the dataset, and obtain up to 70% of accuracy on the test set. In [11], authors utilized Twitter to collect training data and then to perform a sentiment analysis, in which they developed corpora by utilizing emoticons to acquire "positive" and "negative" samples, and then also utilize different classifiers. The authors got best result with the Naive Bayes classifier using a mutual information measure for feature selection. The authors were able to obtain up to 81% of accuracy on their test set. However, this strategy displayed a terrible performance with three classes "positive", "negative" and "neutral" [12].

In [13], author did a magnificent job for sentiment classification on Twitter data. To train a model, they utilized polarity predictions as noisy labels and use 1000 manually labeled tweets for tuning and another 1000 manually labeled tweets for testing. However, they don't specify how they gather their test data. For the classification, they utilized different syntax features of tweets like retweet, hashtags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS of words [14].

3. Sentiment Analysis in Twitter

Sentiment analysis is a type of natural language processing for analyzing the mood of the public about a specific product or subject. Opinion Mining and Sentiment Analysis are the branches of Text Mining which refer to the process of extracting nontrivial patterns and interesting information from unstructured script documents. We can say that they are the expansion to data mining and knowledge discovery. Opinion Mining and Sentiment Analysis concentrate on polarity identification and emotion recognition correspondingly. Opinion Mining has higher attractive potential than data mining, as it is the most natural type of

storing the information in text format. It is much complex task than data mining because it needs to manage unstructured and fuzzy information [8], [15].

3.1 Data Description

Twitter is a social networking service that lets its users to post real time messages, called tweets. Tweets have many unique characteristics. Twitter, with nearly 600 million users and over 250 million messages per day, has rapidly turned into a gold mine for organizations to monitor their reputation and brands by extracting and analyzing the sentiment of the Tweets posted by the public about their remarks, markets, and other contenders. Performing Sentiment Analysis on Twitter is complicated than doing it for large reviews. This is because the tweets are very short and mostly contain slangs, emoticons, hash tags and other twitter language [16].

There is no large public available data set of Twitter tweets with sentiment, so we use Twitter API to collect data. The Twitter API has a parameter that specifies in which language you want to retrieve tweets andwe set this parameter to English. We acquire15000 tweets of three distinct categories i.e. camera, mobile phone, and movie (5000 tweets for each categories). Then we performed preprocessing on tweets.

3.2 Preprocessing

In the tweets, people use acronyms, make the spelling mistake, use emoticons and the other characteristics that express the social meaning. Particularly the other characteristics includes the used of "@" and "#" symbols in tweets, where "@" symbol is act as the target and used to refer the other users on microblog while "#" symbol represents the content of the tweet. Emoticons are the pictorials representation that express people"s mood.

In this paper, we utilized three dictionaries for preprocessing the twitter data. 1) Stop Word Dictionary: Stop word dictionary recognizes a stop words in the tweets. 2) Emoticon Dictionary: It identifies emoticons with their emotional state to extract the emotion contain in the tweets. We prepared the emoticon dictionary by labeling 170 emoticons listed on Wikipedia with their emotional state. For example, ":=(" is labeled as negative whereas ":)" is labeled as positive [14]. 3) Acronym Dictionary: It is used to restore the acronym in the tweets. We used an acronym dictionary from an online resource. For example, omg is translated to oh my god.

We pre-process the remaining tweets as follows [14], [17]: A) All the words are transformed into lower case. B) Replace all the emoticons with their sentiment polarity with reference to the emoticon dictionary. C) Targets (e.g. "@AmitS") are replaced with "USER". D) Remove the tweets having few words (set limit is equal to 5). E) Remove the URL from the tweets. F) Replace a sequence of repeated characters by one character, for example, convert cooooooool to col. G) Split each tweets into the words/features to find its Part-of-Speech (POS) tag (i.e. Verb, Noun, Adjective, etc.). H) Each word is checked, whether a word is a Stop word or not by using a Stop word Dictionary.

According to all previous preprocessing steps, all words are transformed into (word, POS tag, English-word, Stop-word), where English-word distinguishes whether the word is an English word or not (EN represents English Word, and NEN represents Non- English Word) and Stop-word recognizes whether the word is a stop word or not (ST represents Stop Word, and NST represents Non-Stop Word). For example, (iPhone, NN, NEN, NST) [17].

3.3 Feature

Some of the features are as follows [15], [18].

3.3.1 Terms Presence and Frequency: These features are individual words or n-grams words with their frequency counts. It either uses the term frequency weights or gives binary weighting to the words.

3.3.2 Parts of speech (POS): Finding descriptive-words/adjectives from the content, as they are important indicators of opinions.

3.3.3 Opinion words and phrases: These are words generally used to express opinions including good or bad, like or hate. On the other hand, some phrases express opinions without using opinion words.

3.3.4 Negations: The presence of negative words might change the opinion orientation like not good is equivalent to bad.

4. Sentiment Classification

Opinion Mining and Sentiment Analysis are the branches of Text Mining which refer to the process of extracting interesting information and nontrivial patterns from unstructured script documents. Though Sentiment Analysis and Opinion Mining may appear to be identical as a traditional text mining, but it varies because of following facts. Sentiment Classification (SC) is the binary polarity classification which deals with a relatively small number of classes [15]. Sentiment classification is simple task compared to text auto categorization. While Opinion mining displays numerous extra tasks other than sentiment polarity detection.

4.1 Feature Selection Methods

In the feature selection, the issue is to acquire a considerable reduction of the features, without affecting the performance of the classifier. Feature Selection methods can be divided into lexicon-based methods that need human annotation, and statistical methods are automatic and more frequently used methods. The feature selection techniques treat the documents either as Bag of Words (BOW's), or as a string which retains the sequence of words in the document. Because of its simplicity for the classification process, BOW is used regularly. The most common feature selection step is the removal of stop-words and stemming [18]. Some of the regularly used statistical methods in the feature selection are as follows:

4.1.1 Pointwise Mutual Information (PMI)
PMI is a measure of association used in the information theory and statics. In the PMI, association between an unknown features and Positive/Negative seed terms such as "good" or "bad" are used to recognize sentiments of that features. Authors in [19] proposed a verb oriented sentiment classification approach for social domains. The proposed approach concentrates on the verb as they are center component of an opinion. Authors in [20] use the PMI between the keywords to identify similar words. They process the PMI between the noun phrases in each domain. Given that there are set of words that appear only once, and only the noun phrases present more than twice are considered. After computing the PMI score, only the sets of keywords with PMI scores more than a threshold are linked by skip edges [16], [24].

4.1.2 Chi-Square (X^2)
The chi-square test is used to figure out if there is a significant difference between the expected frequencies and the observed frequencies in one or more domains. Chi-square is a statistical test which is commonly used to compare observed data with the data that we hope to obtain according to a specific hypothesis. The chi-square test is always testing null hypothesis, which expresses that there is no significant difference between the expected and observed result. Authors in [21] conducted cross tabulation and chi square analyses for all the datasets, to compare sentiment analysis score with its comparing star rating. Authors in [22] compared the approach (denoted as STD) with the Chi-square test based approach (Chi-Square). Authors proved that the STD approach is significantly better than Chi-Square approach in both precision and recall [16].

4.2 Sentiment Classification Techniques

Sentiment Classification techniques can be differentiated into machine learning approach, lexicon based approach and hybrid approach [25]. The Machine Learning Approach (ML) applies the well-known ML algorithms and uses semantic features. The Lexicon-based Approach depends with respect to sentiment lexicon, which is nothing but a collection of known and precompiled sentiment terms. It is partitioned into dictionary-based approach and corpus-based approach which utilize statistical or semantic methods to find sentiment polarity. The hybrid Approach combines both approaches/methodologies and is exceptionally basic with sentiment lexicons playing a key part in the majority of methods. The different approaches and the most famous algorithms of SC are represented in Figure.1 [18].

4.2.1 Machine Learning Approach
The text classification methods using ML approach can be differentiated into supervised and unsupervised learning methods [23]. In general, the machine learning approach relevant to sentiment analysis mostly belongs to supervised classification and text classification techniques in particular. Thus, it is called "supervised learning" [24]. In a machine learning based classification, couple of documents sets is required i.e. training set and a testing set. A training set is utilized by aclassifier to learn the differentiating characteristics of documents, and a test set is utilized to validate the performance of the classifier [25]. Various machine learning techniques have been used to classify the reviews. Machine learning techniques like Naive Bayes (NB), maximum entropy (ME), and support vector machines (SVM) have made extraordinary success in text

categorization. In the natural language processing area, the other most well-known machine learning methods are K-Nearest neighborhood, ID3, C5, centroid classifier, winnow classifier, and the N-gram model [26].

(a) Naive Bayes (NB)

The Naive Bayes classifier is the simplest and most generally utilized classifier. Naive Bayes classification model computes the posterior probability of a class, depends on the distribution of the words in the document. The model works with Bag-of-Word feature extraction which avoids the position of the word in the document. It uses Bayes Theorem to predict the probability that a given feature set belongs to a particular label [29].

$$P\left(\frac{label}{feature}\right) = \frac{P(label)*P\left(\frac{feature}{label}\right)}{P(feature)} \quad (1)$$

P(label) is the prior probability of a label or the likelihood that a random feature set the label. P(features|label) is the prior probability that a given feature set is being classified as a label. P(features) is the prior probability that a given feature set is occurred. Given the Naive assumption which states that all features are independent, the equation could be rewritten as follows [18]:

$$P\left(\frac{label}{feature}\right) = \frac{P(label)*P\left(\frac{f1}{label}\right)*......*p\left(\frac{fn}{label}\right)}{P(feature)} \quad (2)$$

(b) Maximum Entropy (ME)

The Maximum-Entropy(Maxent) Classifier also known as a conditional exponential classifier, which converts labelledfeature sets into the vectors using encoding. This encoded vector is then used to calculate weights for each feature that can be combined to determine the most likely label for a feature-set [30]. This classifier is parameterized by a set of X{weights}, which is used to merged the joint features that are generated from a feature-set by an X{encoding}. Specifically, the encoding maps each C{(feature-set, label)} pair to a vector. The probability of each label is then computed using the following equation [18]:

$$P\left(\frac{fs}{label}\right) = \frac{dotprod\ (weights, encode\ (fs, label))}{sum\left(\begin{array}{c}dotprod\ (weights, encode\ (fs, l)), \\ for\ linlabels\end{array}\right)} \quad (3)$$

(c) Support Vector Machine (SVM)

Text data are perfectly suited for SVM classification because of the sparse nature of text, in which few features are irrelevant, however they have a tendency to be correlated with each other and usually organized into linearly separable categories [31]. SVM can build a nonlinear decision surface in the original feature space by mapping the data instances non-linearly to an inner product space where the classes can be linearly separated with a hyper-plane [32], [33].

4.2.2 Lexicon-based Approach

The lexicon-based approach based on finding the opinion lexicon which is used to analyze the text. Opinion words are utilized in many sentiment classification tasks [18], [23]. Positive opinion words are utilized to express some desired states, while negative opinion words are utilized to express some undesired states [34]. There are also opinion phrases and expression which together are called opinion lexicon. The lexicon-based approach is partitioned into two methods. The dictionary-based approach based on finding opinion seed words, and then searches the dictionary of their equivalent synonyms and antonyms. The corpus-based methodology starts with a seed list of opinion words, and then finds other opinion words in a large corpus to get opinion words with context specific orientations. This should be possible by utilizing statistical or semantic methods [18], [23], [35].
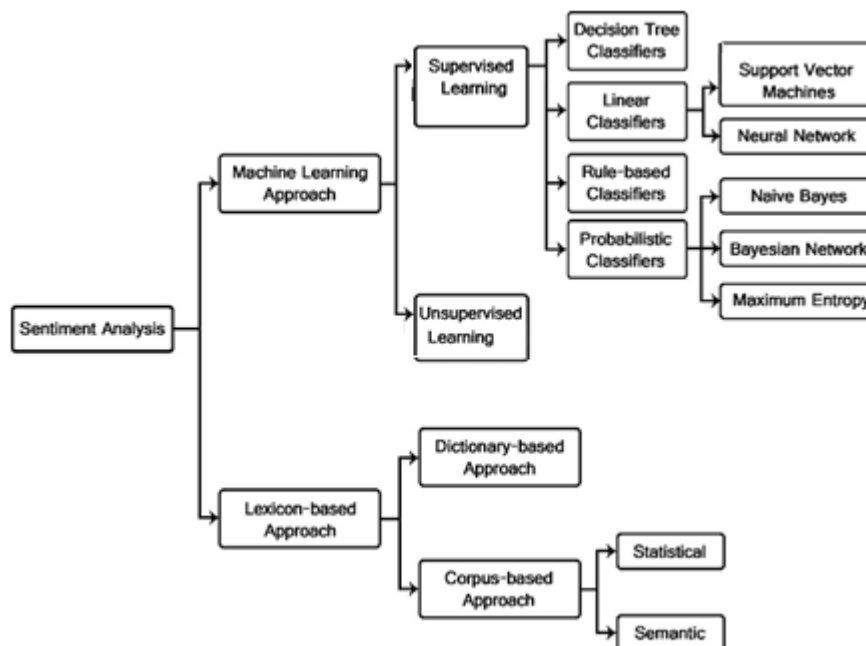

Figure 1: Sentiment Classification Technique

Paper ID: NOV161677

2186

5. Results

Table 1 shows the output of the preprocessing steps. In Preprocessing, we performed all the steps that are given in the section 3.2 for three different datasets i.e. Camera, Mobiles and movies.

Table 1: Datasets details

Sr. No	Name of Dataset	No. of Tweets before Preprocessing	No. of Tweets after Preprocessing
1	Camera	5000	2156
2	Mobiles	5000	2355
3	Movies	5000	1950

Table 2 and Table 3 shows Training data and Testing data respectively, which we labeled manually. Because labeling the tweets is very time consuming job, the amount of data might not be very much currently.

Table 2: Training Data

Name of Dataset	Training Data			
	No. of Tweets	Positive Tweets	Negative Tweets	Neutral Tweets
Camera	1900	732	443	725
Mobile	2100	776	453	871
Movies	1700	675	379	646

Table 3: Testing Data

Name of Dataset	Testing Data			
	No. of Tweets	Positive Tweets	Negative Tweets	Neutral Tweets
Camera	256	91	75	90
Mobile	255	102	50	103
Movies	250	113	52	85
Total	761	306	177	278

From the different machine learning algorithms (section 4.2.1); we used Naive-bayes Classifier to determine the sentiment of the tweets on Testing Data. Table 4 shows the result of Testing Data using Naive-Bayes Classifier. In this Table 4, we treat positive and negative tweets as Opinion and neutral tweets as a Non-Opinion.

Table 4: Result of Naive-Bayes classifiers

	Opinion (483)	Non Opinion (278)
Opinion	431	223
Non Opinion	52	55

From the different Feature Selection methods (Section 4.1), we used Mutual Information Feature Selection method for extracting the different features from the tweets. We used Naïve-Bayes Classifiers and MI Feature selection to train multi classifiers. Table 5 represents the results of MI feature selection on test data.

Table 5: Feature Selection using MI

	Camera (256)	Mobile (255)	Movies (250)
Positive	224	227	222
Negative	32	27	28
Neutral	0	1	0

6. Conclusion and Future Work

In our research, we have presented a method that automatically collects tweets from twitter using Twitter API. Then on these collected tweets, weperform preprocessing for our desired results and after that we performed POS tagging by using Tree-Tagger. We can use some of tweets to train a sentiment classifier. Our classifier is able to determine positive, negative and neutral sentiments from tweets. The classifier is based on the multinomial Naive Bayes classifier that uses Unigram and POS tags as features. It is found that different types of features and classification algorithms are combined in order to increase the performance, so we combined Naive bayes classifier with Mutual Information Future selection algorithm. By using Naïve bayes classifier, our system got 66%, 62.3% and63.2% results for camera, mobile and movies dataset respectively and the overall accuracy of Naive-Bayes Classifiers on complete Testing Data is 63.8%. By using MI feature selection along with Naïve-Bayes classifiers, our system got 87.5%,89% and 88.8% results for camera, mobile and movies dataset respectively and the overall accuracy of MI along with Naïve-Bayes classifiers is 88.4%.

Although the techniques and algorithms used for sentiment analysis are advancing fast, but there is a lot of problems in this field. The main challenging aspects exist in use of other languages, complexity of sentence/document, dealing with negation expressions, handling of implicit product features, etc.In future, more work is required like to deal with unlabeled training data for further improving the performance.

References

[1] Zhi-Hong Deng, Kun-Hu Luo, Hong-Liang Yu, "A Study of Supervised Term Weighting Scheme For Sentiment Analysis", Expert Systems with Applications: An International Journal, Volume 41, Issue 7, June 2014, pg.3506-3513.

[2] Ming Hao, ChristionRohrdantz, HalldorJanetzko, "Integrating Sentiment Analysis and Term Associations with Geo-Temporal Visualizations On Customer Feedback Streams", Published in SPIE Proceedings Vol. 8294, 25 January 2012.

[3] Bing Liu, "Sentiment Analysis and Opinion Mining", Synthesis Lectures on Human Language Technologies 5(2012), no.1, pg.1-167.

[4] RamandeepSandhu, Rahul Mehta, "Applying Opinion Mining to Organize Web Opinions", International Journal of Computer Science, Engineering and Applications (IJCSEA) Volume 1, No. 4, August 2011, pg.82-89.

[5] D. Sai Krishna, G AkshayKulkarni, A. Mohan, "Sentiment Analysis-Time Variant Analytics", International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 3, March 2015, pg.466-472.

[6] PranaliTumsare, Ashish .S. Sambare, Sachin .R. Jain. "Opinion Mining In Natural Language Processing Using Sentiwordnet And Fuzzy", International Journal of Emerging Trends & Technology in Computer

Science (IJETTCS) Volume 3, Issue 3, May – June 2014.

[7] Poornima Singh, Gayatri S Pandi (Jain), "Opinion Mining on Social Media: Based on Unstructured Data", International Journal of Computer Science and Mobile Computing IJCSMC, Vol. 4, Issue. 6, June 2015, pg.768–777.

[8] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval Vol. 2, No 1-2 (2008), pg.1–135.

[9] Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen, "Emotion Classification Using Web Blog Corpora", In WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence 2007,Washington, DC, USA. IEEE Computer Society,pg.275-278.

[10] Jonathon Read, "Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification", In ACL,The Association for Computer Linguistics 2005,pg.43-48.

[11] Alec Go, Lei Huang, RichaBhayani, "Twitter Sentiment Analysis", Final Projects from CS224N for Spring 2008/2009 at the Stanford Natural Language Processing Group.

[12] Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", pg.1320-1326.

[13] Luciano Barbosa, JunlanFeng, "Robust Sentiment Detection on Twitter from Biased and Noisy Data", Proceedings of the 23rd International Conference on Computational Linguistics: Posters 2010, pg.36–44.

[14] ApoorvAgarwal, BoyiXie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, "Sentiment Analysis of Twitter Data", Proceedings of the workshop on Languages in Social Media, LSM'11, pg.30-38.

[15] Kishori K. Pawar, R. R. Deshmukh, "Twitter Sentiment Analysis: A Review", International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015 9 ISSN 2229-5518, pg.957-964.

[16] Smt. Shubhangi D Patil, Dr Ratnadeep R Deshmukh, "Review of Twitter sentiment analysis", International Journal of Scientific & Engineering Research, Volume 5, Issue 10, October-2014 ISSN 2229-5518, pg.1616-1623.

[17] Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social Media Data", 2013 IEEE 14th International Conference on Mobile Data Management, pg.91-96.

[18] WalaaMedhat, Ahmed Hassan, HodaKorashy, "Sentiment Analysis Algorithms and Applications: A Survey", Ain Shams Engineering Journal (2014) 5, pg.1093–1113.

[19] MostafaKaramibekr, Ali A. Ghorbani, "Verb Oriented Sentiment Classification", 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.

[20] MinjieZheng, Zhicheng Lei, LIU Yue, Xiangwen Liao, Guolong Chen, "Identify Sentiment-Objects from Chinese Sentences Based on Skip Chain Conditional Random Fields model", 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing.

[21] ParisaLak, OzgurTuretken," Star Ratings Versus Sentiment Analysis - A Comparison of Explicit and Implicit Measures of Opinions", 2014 47th Hawaii International Conference on System Science, pg.796-205.

[22] KekeCai, Scott Spangler, Ying Chen, Li Zhang, "Leveraging Sentiment Analysis for Topic Detection", 2008 IEEE/WIC/ACM Internation-al Conference on Web Intelligence and Intelligent Agent Technology, pg.265-271.

[23] Shailesh Kumar Yadav, "Sentiment Analysis and Classification: A Surve", International Journal of Advance Research in Computer Science and Management Studies Volume 3, Issue 3, March 2015, pg.113-121.

[24] AishwaryaBhole, Prof. V. D. Thombre, "Review of Sentiment Classification Methods and Opinion Mining: The Future Roadmap", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 3 Issue 3, March – 2014, pg.243-248.

[25] Miss. Siddhi Patni, Prof. AvinashWadhe, "Comparative Study of Available Technique for Detection in Sentiment Analysis", International Journal of Computational Engineering Research Vol, 03 Issue, pg.73-77.

[26] G.Vinodhini, RM.Chandrasekaran , "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 6, June 2012, pg.282-292.

[27] Peter D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 27th annual meeting on Association for Computational Linguistics, pg.76-83.

[28] Diana Maynard, Adam Funk, "Automatic Detection of Political Opinions in tweets", In: Proceedings of the 8th international conference on the semantic web, ESWC'11; 2011, pg.88–99.

[29] Vidisha M. Pradhan, Jay Vala, PremBalani, "A Survey on Sentiment Analysis Algorithms for Opinion Mining", International Journal of Computer Applications (0975 – 8887) Volume 133 – No.9, January 2016, pg.7-11.

[30] RichaBhayani, Alec Go, Lei Huang, "Twitter Sentiment Classification using Distant Supervision".

[31] Joachims T, "Probabilistic analysis of the rocchio algorithm with TFIDF for text categorization", In: Presented at the ICML conference; 1997, pg.143-151.

[32] Aizerman M, Braverman E, Rozonoer L, "Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning", 1964, pg.821–837.

[33] Bo Pang, Lillian Lee,ShivakumarVaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques".

[34] SnehaVishwakarma, Suresh Kumar, "A Lexical Approach for Tweets Sentiment Classification", International Journal of Scientific Research volume: 4, Issue: 10, ISSN No. 2277-8179, pg.242-245.

[35] DudhatAnkitkumar, Prof. R. R. Badre, Prof. MayuraKinikar, "A Survey on Sentiment Analysis and Opinion Mining", International Journal of Innovative Research in Computer and Communication

Engineering, Vol. 2, Issue 11, November 2014, pg.6633-6639.

Author Profile

Amit Gangadhar Shirbhate Received B.E in Computer Science and Engineering from JNEC, Dr. B.A.M.University, Aurangabad in 2014. Currentlypursuing M.Tech in Computer Science and Engineering from Department of Computer Scienceand IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India.

Dr. Sachin N. Deshmukh completed his Ph.D. and M.E inComputer Science and Engineering. He is currentlya Professor in Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India. His research Area is Text mining, Social Web mining and Intension Mining. He is a member of Adhoc Board of Studies in BioInformatics and Liberal arts of Dr. B. A. M. University Aurangabad and Adhoc Board of Computer Science at Shivaji University Kolhapur.