International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Sentiment Analysis - Methods, Applications & Challenges

Ms. Swati B. Bhonde
Amrutvahini College of Engineering, Sangamner
Savitribai Phule Pune University, Pune, Maharashtra- India

Prof. Jayashree R. Prasad
Sinhgad College of Engineering, Pune
Savitribai Phule Pune University, Pune, Maharashtra- India

***Abstract* – With the excogitation of World Wide Web, Internet has become a coercive platform for people to express their views, opinions, emotions and sentiments about particular product, people and about life in general. So, when anyone wants to buy some product they prefer to go through several user reviews, blogs and discussion available in public forums. This opinionated textual data & word of mouth represents new & measurable source of information in many applications. Due to increased availability of online reviews in digital form, need to organize them in proper structured format is becoming necessary. Sentiment Analysis (SA) is one of the most widely studied applications in the era of Natural Language Processing (NLP) & Machine Learning (ML) which deals with identifying & extracting essential information from public opinion. This helps to make better decision in business intelligence, political campaigns & in product recommendation system. This paper presents survey on Sentiment Analysis, methods, applications and challenges. This review discusses methods for feature selection & sentiment classification in detail which will prove helpful for user in sentiment classification.**

***Keywords* – *Sentiment* Analysis (SA), Natural Language Processing (NLP), Opinion Mining (OM), Machine Learning (ML), Latent Semantic Indexing (LSI).**

I. INTRODUCTION

Now-a-days, people can freely impart their voice to the public by using social media like reviews, blogs, forum discussions, micro blogs & other social networking sites. All this valuable information caused SA techniques to move forward so that many complexities in unstructured data can be captured. As web is largest source of opinionated data, analysts are moving towards unsupervised approaches, which are really proving better than traditional algorithms. Goal of SA process is to identify whether given text is subjective or objective or it is positive or negative. Many times Opinion Mining (OM) & SA are used interchangeably. OM deals with extracting & analyzing opinion of people about some entity & SA deals with specific sentiment expressed in a text & then it tries to analyze it. This paper is organized as:

First section describes introduction to SA. In second section process of SA is presented.

Third section covers feature selection methods & sentiment classification techniques. Fourth section discusses about applications of SA. Fifth section highlights challenges in SA.

Sentiment or opinion is defined as a quintuple [1]- < o_j ; f_{jk}; so_{ijkl}; h_i; t_l >, where oj is a target object, f_{jk} is a feature of the object o_j , so_{ijkl} is the sentiment value of the opinion of the opinion holder hi on feature f_{jk} of object o_j at time t_l, so_{ijkl} is positive, negative, or neutral, or a more granular rating h_i is an opinion holder, t_l is the time when the opinion is expressed."

SA problem can be viewed as a classification problem where sentence is classified as positive, negative or neutral [2]. There are three levels of SA:

1) Document level: Classifies whole opinion document as positive or negative expression.

2) Sentence level: Sentiment expressed in sentence is classified either as a positive, negative or neutral.

3) Aspect/feature level: Classification of a sentiment with respect to specific aspect or feature of entities.

Users often express opinions about multiple aspects in same sentences. For example,

Story of this movie is not good but songs are very nice.

[Story – not good (negative) & songs – nice (positive)]

Objectives of SA are as follows:

- Detect opinion or sentiment words in a text.
- Identification of polarity expressed in sentence.
- Classification of polarity.

II. PROCESS OF SENTIMENT ANALYSIS

Figure 2.1 shows process of sentiment analysis [3].

Stepwise explanation of sentiment analysis process is as follows:

1. Unstructured Text as Input –

Input or data sets used in SA plays very significant role. Main source of this data is coming from reviews from social media sites for example- in political debate people's opinion about certain election candidate or political parties is expressed which can be helpful for prediction of results. These posts frequently contain short text which should be carefully handled and it needs special attention from programmer.

2. Data preprocessing:

Often input data is in raw format. Following techniques are used to clean this data:

a. Stopword removal
b. Apply Stemming algorithms
c. Use POS tagging
d. Use any dictionary tool to get meaning of word.

3. Sentiment term Identification –

It aims at extracting, analyzing emotions i.e. to decide whether given text contains any subjective contents or not. Output of this phase retrieves phrases which expresses some opinion.

This work is carried out as a part of BCUD project sponsored by Savitribai Phule Pune University, Pune.

International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

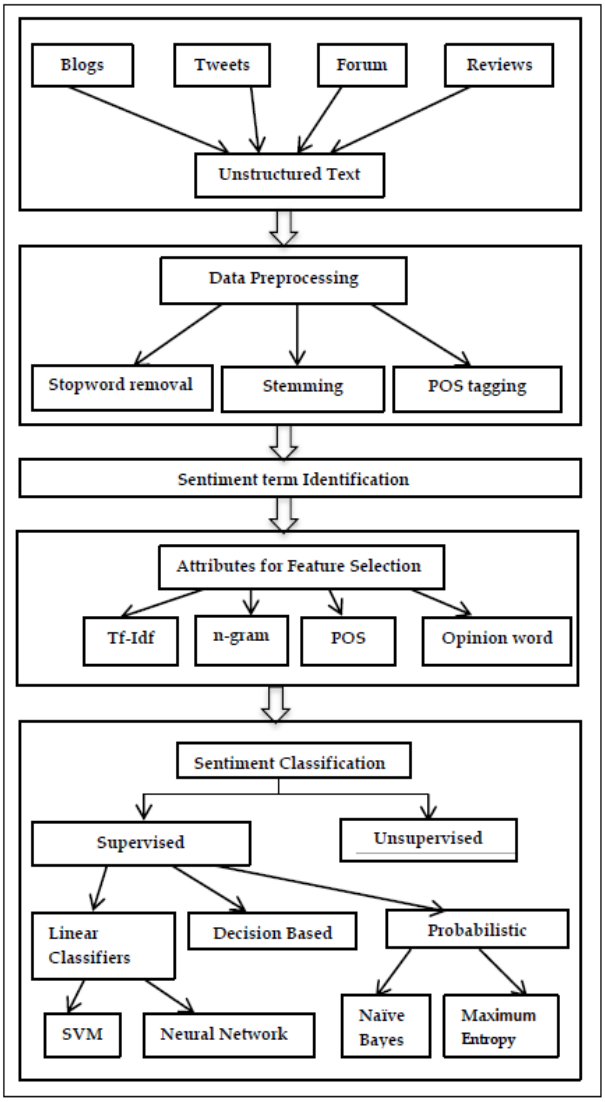

Fig.2.1. Sentiment analysis process of reviews

4. Feature selection methods:

Once a phrase that expresses opinion are detected then subsequent step is to predict sentiment orientation of what opinion holder is talking about. Feature can be implicit or explicit.

E.g. i) Sound quality of phone is better
Explicit Feature: sound quality
Sentiment: better (positive)
ii) This mobile easily fits into packet
Implicit feature: size
Sentiment: easily fits (positive)

Methods for feature selection are discussed in section-III.

5. Sentiment classification –

Problem of classification can be defined as, given a set of training records let's say, D= {X_1, X_2 ...., Xn} where each record is labeled to some class. This classification problem deals with finding only one class to which underlying record belongs. When there is only one class label, problem is called as hard classification problem & when an instance has a probabilistic value of label then it is called as soft classification problem is. Section-III discusses various method of sentiment classification.

III. Feature Selection Methods & Sentiment Classification

SA is considered as a sentiment Classification problem. First step here is to extract and select text features. Some of the features are as follows [4]:

1. Term presence:

In this process properties are extracted from data, because input data is too large to use in the classification. These features can be individual word or word n-gram & their frequency count [5]. It either gives binary weighting (0-if word appears, 1-if otherwise).

2. Term frequency:

The term frequency *(tf(w,d))* is the number of times that a word w occurs in a document d. Tf(w,d)=|{w€d}|.The tf-idf (term frequency-inverse document frequency) measure is a statistic that tells importance of a word in a set of document.

3. POS tagger:

The POS tagger is a method of marking up a word in a text corresponding to a particular part-of-speech. The idea here is that only a limited set of word in a sentence indicate the sentiment, referred to as sentiment-words. In English language POS examples are noun, verb, adverb and adjective etc.

4. Opinion words & phrases:

These are commonly used words used to express opinion including good or bad, like or dislike etc.

- *Feature selection methods:*

Feature selection ranks all features based on a measure of how much they are related to a class and removes all features below a specific predefined threshold [6]. If less number of features are selected then it improves efficiency of training and testing procedures also increases performance of classification as irrelevant features are removed. Figure 3.1 shows different feature selection methods, discussed as below:

i) PMI (Pointwise Mutual Exclusion) -

This measure is derived from information theory [7]. It is a formal way to model mutual information between features & classes. PMI i.e. MI(w) between word 'w' & class 'i' is defined on basis of level of co-occurrence between class i & word 'w'. Expected co-occurrence between class i and word w , on basis of mutual independence is given by, pi · F(w) and co-occurrence is given by F(w).pi(w)[8]. Here word 'w' is positively related to class 'i , when Mi(w) is greater than 0 & word is negatively correlated with class if Mi(w) is less than 0. PMI is a simple association, which can be used for unsupervised learning.

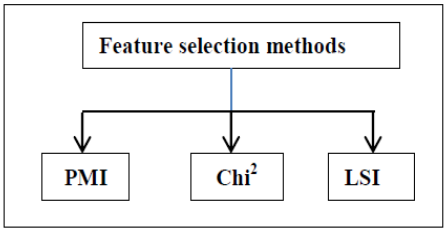

Fig.3.1. Feature selection methods

The classification is based on the average semantic orientation, which make use of reference words, for most positive and most negative association

ii) Chi-square () -

Let n be the total number of documents in the collection, pi(w) be the conditional probability of class i for documents which contain w, Pi be the global fraction of documents containing the class i, and F(w) be the global fraction of documents which contain the word w. Therefore, the chi-square -statistic of the word between word w and class i is defined as [9]:

Chi-square is better than PMI and is used in many applications. It considers only co-occurrence strength.

iii) Latent Semantic Indexing-

Feature selection methods focuses on reducing the dimensionality of the data by picking from the original set of attributes. Feature transformation methods create a smaller set of features as a function of the original set of features. LSI uses a technique called singular value decomposition to detect patterns in the relationships between the terms and concepts contained in an unstructured collection of text. LSI is popular feature transformation methods [10]. LSI method transforms the text space to a new axis system which is a linear combination of the original word features. Principal Component Analysis techniques are used to achieve this goal [11].

Main disadvantage of LSI is that it is unsupervised technique which has no knowledge of the class-distribution attributes. There are many other statistical approaches in feature selection like Hidden Markov Model & Latent Dirichlet Allocation. Both are used for separating entities in a review document from the subjective expression which are used for expressing polarities. LDA uses generative model that allows documents to be explained unobserved topics.

- *Sentiment classification:*

Basically it refers to find whether certain text represents opinion or fact of author & also most researchers have shown that there is a close relation between subjectivity classification and document sentiment classification. There are some approaches which are directly used for classifying whether document is subjective or objective [12]. To find overall similarity of a sentence of the document, following steps are used:

i) Use IR method to acquire document that describes only one topic.
ii) Calculate it's similarity scores with each sentence in those document & make average value.
iii) Assign sentence to the category for which average value is highest.

There are two basic approaches used in sentiment classification [3,25]:

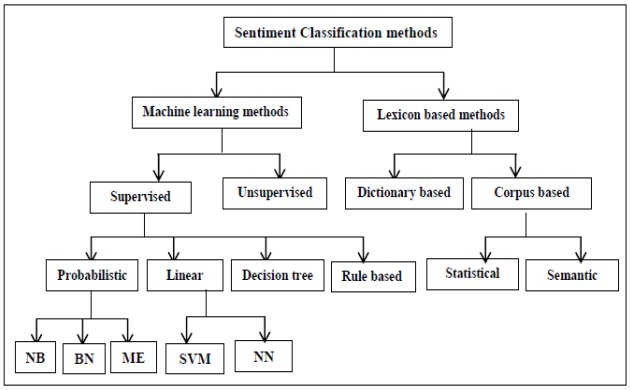

Fig.3.2. Sentiment classification methods

- *Machine Learning:*

Uses machine learning algorithm & linguistic features. They are further divided into supervised and unsupervised learning methods. Supervised methods uses large number of training documents & unsupervised methods are used when it is difficult to find these labeled training documents.

- *Lexicon based approaches:*

They depends on sentiment lexicon i.e. collection of well-known & pre-compiled sentiment terms. They are further divided into dictionary based approach & corpus based approach which uses some statistical approach to find sentiment polarity. Hybrid approach can be also used which is combination of both. Dictionary based approaches depends on finding out words that expresses opinion and then search dictionary for their synonym and antonyms. Corpus based approaches are different, they start with a seed list of words and then find other opinion words in a large corpus to find context specific meaning of the word. Following is the brief discussion of various algorithms used in both the approaches:

1. Machine learning:

The goal of machine learning is to design programs that learn and/or discover[3], i.e. automatically improve their performance on certain tasks. The result can be a 'learned' program which can carry out the task it was designed for. There are many algorithms used in classification for this which uses syntactic and linguistic features.

1.1 Supervised learning

This method assumes that labeled training dataset is already available. Few of the supervised classifiers are discussed in following section:

1.1.1 Probabilistic classifiers:

These classifiers uses mixture model for classification which assumes that all classes are components of mixture. All mixture components provide a generative model that provides probability of sampling the term for that component. Three famous probabilistic classifiers are as[25]:

- Naïve Bayes (NB)

The naïve Bayes algorithm uses Bayes theorem for predicting whether given feature belongs to particular class label[3]. It computes posterior probability of class based on distribution of words in the documents. It works with help of Bag-Of-Words approach in which position of

word is ignored. The formula P(label|Features) states the conditional probability of label given Feature. P(label) is probability of a label that a random feature sets the label. P(feature) is a probability that given set of feature is encountered. P(label|feature) is probability of feature set is classified as a label. It allows calculation of unknown conditional probability from known conditional probability in concerned with prior probability. Default assumption here is that presence of feature is not related to presence of any feature. Disadvantage of NB is that even small training set will suffice to train the model.

- Bayesian Network (BN)

This algorithm works with assumption that all features are independent of each other or fully dependent on each other. This tends to a Bayesian Network model – a graph in which random variables are nodes and conditional dependencies are edges[25]. As complexity of BN is very high it is not generally used.

- Maximum Entropy Classifiers (ME)

It is also called as Conditional Exponential Classifier. It uses encoding technique to convert labeled feature set to vectors and then weights are calculated for each feature to determine most relevant label for a feature.

1.1.2 Linear classifiers:

There are many linear classifiers attempting to detect good linear separator between different classes. Two of them are discussed below[3]:

- Support Vector Machine:

SVM exist in different form linear and non-linear. It is a supervised classifier. It is used when classes are linearly separable and line can be found which spilt the two classes perfectly. In practice the classes are usually not linearly separable; in such cases a higher order function can spilt the dataset. A function is applied to the data set which maps the point in the nonlinear data set to point in a linear data set. SVM is more suitable for text data as only few irrelevant features are there and they are also related with other features so can be easily separated by using hyperplane[13,14].

- Neural Network(NN):

In these algorithms neuron is a basic unit. Vector overline Xi is used to denote input to neuron, which is word frequencies in it^h document. Each neuron has some weight associated with it and it computes the function of it's input. We can also use multilayer neuron in which output of neuron in previous phase is passed as input to next stage.

1.1.3 Decision tree classifiers:

It uses hierarchical representation of training data space where condition on attribute value is used to divide the data[15]. Condition is whether word is present or absent. Process is repeated until leaf node contains minimum records used for the purpose of classification. Many methods like single attribute split[16], multi attribute split and discriminant-based multi attribute split can be used for performing split operation[17].

1.1.4 Rule Based Classifier:

Here set of rules are used for modeling data space[3]. LHS represents condition and RHS gives class label. Conditions are on presence of the term. Two most commonly used criteria are support and confidence. Support is exact number of instances in training data set relevant to rule. Confidence is conditional probability if RHS is satisfied then LHS is also satisfied.

1.2 Unsupervised learning:

This is another method of classification. Every time it's not possible to assign some class labels to training documents but it is very easy to collect these unlabeled documents. Unsupervised learning methods overcame this limitation. This approach collects documents, splits sentences into words and categorizes each sentence using list of words from each category depending on sentence similarity measures. Some of the methods refers to prior information as labeled features and use them directly for predicting unlabeled instances using generalized expectation criteria[25].

Many other unsupervised approaches also depends on semantic orientation using PMI, semantic spaces and distributional similarities between word and its polarity.

2. Lexicon Based approach:

Most of the input data available for sentiment analysis contains opinion words [3]. There can be positive words expressing like for the product/topic or can be negative words expressing dislike for product/topic. Words which are used for expressing opinion are called as "Opinion Lexicon". Following are three approaches used for collecting list of such opinionated words:

- Dictionary based approach:

Dictionary is important resource used in Sentiment Analysis. Set of opinion words is collected manually with their orientation which is linked with corpora like 'Wordnet' for their synonyms and antonyms. Newly found words are also attached to this and this list grows gradually. Main disadvantage of this approach is that it is difficult to get context specific meaning of the word directly.

Ex. "I deposited money in bank."

Here, bank refers to financial bank but it also shows other meaning - bank of river etc.

Logic is again required to find out current context of the document.

- Corpus-based approach:

This method eliminates problem of dictionary based approach to find out context sensitive meaning of word. Rather it depends on syntactic patterns along with list of keywords or seed words used to find out meaning of the word in a large corpus. Along with seed list many combinations of part-of-speech are also used to detect co-related sentences. Conditional Random Fields method was used for extracting opinion expressions[17]. Then taxonomy based approach was also used for extracting features based on domain specific resources used for capturing knowledge of opinion holder on a product. But this approach is not good as compare to dictionary based approach as it is not possible to create corpus of all domains. Therefore this approach is implemented using statistical or semantic technique as discusses follows:

- *Statistical approach:*

It deals with finding patterns or co-occurrence between seed opinion words. Documents available on web can be

used as corpus for developing a dictionary. Also they can be traverse to decide polarity of sentences. If word occurs in positive texts, its orientation is positive and vice versa[18]. If it has equal frequencies in all documents then it is neutral word. If word appears frequently within same context they can have same polarity. Thus polarity is initially unknown but it can be determined by using this approach.

- *Semantic approach:*

This approach uses sentiment values for detecting similarities between words. Many resources like WordNet, SentiWordnet can be used to calculate polarity of sentiments. This approach is widely used in building lexicon model for describing verbs, nouns, adverbs and adjectives. Main focus here is on finding orientation of opinion holder on particular aspect of entity. Statistical and semantic approach can be combined together to generate more accurate result.

- Lexicon based & NLP techniques:

In this method NLP techniques are merged with lexicon based approach to find out semantic relation and syntactical structure between sentences. In this approach each POS can have some value which is propagated through syntactic structure of parsed tree[19].

- Discourse information:

Discourse information is significantly important in field of SA. It can be among sentences or clauses described in sentences. Asher [20] has given five types of relations contrast, correlation, support, result and continuation with attached sentiment information and annotation. Work [21] has been done on finding intra-sentence discourse relation for eliminating ambiguities in polarity and to filter out polarity phrases to identify relationship.

IV. Applications of Sentiment Analysis

Some of the applications of sentiment analysis include online advertising, hotspot detection in forums, web blog author's attitude analysis, sentiment filtering etc[23].

a) For recommendation system:

For online shopping, people always prefer to go through many reviews and comments written by user on social media. Obviously the product having highest rating or all positive feedback from user is most likely to be selected by user. So, OM system can be used for recommendation of particular product.

b) For Ad Placement:

In online systems that display ads as sidebars, it is helpful to detect WebPages that contain sensitive content inappropriate for ads placement.. For more sophisticated systems, it could be useful to bring up product ads when relevant positive sentiments are detected, and perhaps more importantly, neglect the ads when relevant negative statements are discovered.

c) In Business Intelligence:

Web contains tremendous amount of information & for a user it is feasible to traverse through all contents provided on net. Using an Opinion Mining approach one can exploit web content in order to improve customer relationship management and provide better service to the customer by improving on the product quality and making the product personalized according to the customer view point. Business also wishes to understand online reviews to improve their product features.

d) For trend prediction

Organization can predict the trend by tracking public viewpoints. In stock market, analysis of sentiments related to whether the stock price will be higher or lower and help the investor to take decision related to buying or selling the stock.

e) For political domain

Opinion matters a great deal in politics. Sentiment analysis has specifically been proposed as a key enabling technology, allowing the automatic analysis of the opinions that people submit about pending policy or government-regulation, understanding what the voters is thinking, predicting the outcome of elections etc.

f) Application in smart home

Smart homes are going to be technology of future where based on current sentiment or emotion of the user, home will change it's ambience to create moody or peaceful information.

V. Challenges in Sentiment Analysis

SA is challenges task & required deep understanding of the problem some of the challenges are [22]:

1. Identifying subjective portion of text:

Same word can be treated as a subjective in one context, while it might be objective in some other. This makes difficulty of identify subjective portion of text.

E.g – Language of author was very crude - subjective
Crude oil is extracted from sea beds - objective.

2. Associating sentiment with specific keyword:

Sometimes sentences indicate extremely strong opinion, but it is difficult to find out source of these sentiments. Therefore association to keyword / phrase is extremely difficult[24].

Eg "Every time I saw '3-Idiots',I just enjoy it at fullest
and appreciate acting of all of them"
Here, it refer to movie, '3-Idiots' and
all of them – actor are referred .
It is also called as, co-reference resolution.

3. Sarcasm detection:

These types of sentences express negative opinion about target entity using positive words.

E.g. "Nice perfume" You must have spent so much on it. Here, though sentence contains all positive words overall sentiment of sentence is negative.

4. Thwarted Expression:

There are some sentences where minorities of text determine over all polarity of the documents [24].

E.g "XYZ is a brilliant film song are nice all actor are first grade & supporting cast is good as well & delivered good performance however it can't hold up". Sample Bag-of-word approach will fail here as most of the words are +ve but ultimate sentiment is -ve.

5. Indirect negation of sentiment:

Sentiment can be negated in simple way without using negative words like no, not.

E.g : "Avoid spicy food of outside "

Word "Avoid" - negates the sentence.

6. Order dependence:

In traditional text classification discourse structure doesn't play any role in classification as words are considered independent of each other, it is essential for SA[25].

E.g: "A is better than B" conveys exact opinion from "B is better than A"

7. Entity recognition:

Not everything in text talks about same entity. Separation of text about particular entity and then analysis of it's sentiment is needed..

E.g "I hate Nokia, but I like Samsung.".

Simple BOW approach will mark it as natural, but it carries specific sentiment for both entities present in the statement.

8. Identifying opinion holder:

All that written in text is not always opinion of author

For e.g. when someone quotes something for other it becomes difficult to identify source of that particular opinion. For example,

George accused his rival of overseeing a stagnant economy "the middle class has been crushed over last 4 years & jobs have been to scarce", former massachusetts governor said.

Even though comment given by george is -ve, this news item provides objective opinion .

9. Implicit sentiment:

Sometimes, word that indicates features is not always present in the sentence.

E.g 1. "This mobile is easily fits into packet" is actually speaking about size of mobile

2. "This book is expensive" is commenting on price or cost of book .

10. World knowledge:

Certain comments and statement might take reference of certain existing char/plays etc.to depict sentiment .our classifier need to be known to knowledge of the world to identify such comparison .

11. Use of short text:

Opinioner most of time uses abbreviations, lack of capitals, poor spelling, poor punctuation, poor grammar is pretty difficult to understand.

12. Detection of spam and fake reviews mainly through identification of duplicates, comparison of qualitative with summary review, detection of outliers are few things which we must consider.

13. Language, is still another challenges, most of the work done in SA is focused on English and Chinese Language.

VI. Conclusion

SA can be widely applied to any domain for classifying & summarizing reviews. Because of diversity of web source & tremendous data, it's necessary to analyze this data to predict opinion. Goal of this paper is to cover methods, application & challenges to directly enable opinion oriented information gaining system and to convey a reader about sense of information. Survey of different methods for feature selection & sentiment classification are discussed in detail which will prove helpful for the miner

Acknowledgement

I take this opportunity to extend my thanks to Savitribai Phule Pune University, for funding this project. My special thanks to Dr. Jayashree Prasad for her valuable guidance & kind support.

References

[1] Bing Liu, Bo Pang, "Sentiment Analysis and Opinion Mining"

[2] Wilson T, Wiebe J, Hoffman P., "Recognizing contextual polarity in phrase-level sentiment analysis", In: Proceedings of HLT/EMNLP; 2005.

[3] Walaa Medhat, Ahmed Hassan, Hoda Korashy , "Sentiment analysis algorithms and applications: A survey"

[4] Mrs. Vijyalaxmi M, Mrs Shalu Chopra, Mrs Sangeeta Oswal, Mrs Deepshikha Chaturvedi, "The How, When and Why of Sentiment Analysis "Departement of Information Technology, VESIT, Mumbai.

[5] Yelena Mejova, Padmini Srinivasan, "Exploring feature definition and selection for sentiment classifiers", In: Proceedings of the fifth international AAAI conference on weblogs and social media; 2011.

[6] Whitelaw Casey, Garg Navendu, Argamon Shlomo, "Using appraisal groups for sentiment analysis" , In: Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM); 2005. p. 625–31

[7] Cover TM, Thomas JA. , "Elements of information theory". NewYork: John Wiley and Sons; 1991

[8] Nádia F.F. da Silva a, Eduardo R. Hruschka a, Estevam R. Hruschka Jr.Tweet, "Sentiment analysis with classifier ensembles"

[9] Aggarwal Charu C, Zhai Cheng Xiang, "Mining Text Data", Springer New York Dordrecht Heidelberg London: _ Springer Science Business Media, LLC'12; 2012.

[10] Deerwester S, Dumais S, Landauer T, Furnas G, Arshman R. "Indexing by latent semantic analysis", JASIS 1990;41:391–407

[11] Xiaolin Zheng , Zhen Lin , Xiaowei Wang , Kwei-Jay Lin, Meina Song, "Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification"

[12] 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. p. 88–99

[13] Cortes C, Vapnik V. , "Support-vector networks", presented at the Machine Learning; 2009.

[14] Vapnik V. ,"The nature of statistical learning theory", new York; 2012

[15] Quinlan JR, "Induction of decision tree Machine Learn" 1986;1:81–106.

[16] Chakrabarti Soumen, Roy Shourya, Soundalgekar Mahesh V., "Fast and accurate text classification via multiple linear discriminant projections" , VLDB J 2003; 2:172–85.

[17] Lewis David D, Ringuette Marc., "A comparison of two learning algorithms for text categorization", SDAIR.

[18] Kim S, Hovy E. , "Determining the sentiment of opinions", In: Proceedings of interntional conference on Computational Linguistics (COLING'04); 2004

[19] Asher N, Benamara F, Mathieu Y.," Distilling opinion in Discourse: a preliminary study", presented at the COLING'08.

[20] Lafferty J, McCallum A, Pereira F. "Conditional random fields: probabilistic models for segmenting and labeling sequence data"

[21] Moreo A, Romero M, Castro JL, Zurita JM, "Lexicon-based comments-oriented news sentiment analyzer system", Expert Syst Appl
[22] Xianghua Fu, Guo Liu, Yanyan Guo, Zhiqiang Wang. "Multiaspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon", Knowledge Based System, 2013; 37:186–95."
[23] David Osimo1 and Francesco Mureddu, David Osimo1 and Francesco Mureddu , "Research Challenge on Opinion Mining and Sentiment Analysis"
[24] Jaganadh G, "Opinion Mining and Sentiment Analysis", 2012
[25] 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. p. 88–99.

AUTHOR'S PROFILE

Swati B. Bhonde

is graduated in Computer Engineering from College of Engineering, Pune in 2007. She had her masters in Computer Engineering from Bharati Vidyapeeth, Pune in 2011.
Currently she is working with Amrutvahini College of Engineering, Sangamner. Her interested areas of research are text mining, sentiment analysis & Natural language processing. This work is carried out as a part of BCUD project, sponsored by Savitribai Phule Pune University, Pune. Her professional membership includes ISTE, IAENG.

Jayashree Rajesh Prasad

graduated in CSE from North Maharashtra University in 1996 and completed M.E. in Computer Engineering from Pune University in 2004. She received her Ph.D. in CSE from Swami Ramananda Tirtha University, Nanded in 2014. She has a research project "Conversion of Gujrati Script to Speech", funded by BCUD (Savitribai Phule Pune University, Pune) to her credit. She works with Sinhgad College of Engineering, Pune. Her research interests are in the field of Soft Computing, pattern recognition and image processing. She is Life member of Computer Society of India, Life Member of Indian Society for Technical Education, Member of IAENG (International Association of Engineers) and Member of IACSIT (International Association of Computer Science and Information Technology).