e-ISSN : 0976-5166
p-ISSN : 2231-3850

A SURVEY ON SENTIMENT ANALYSIS

S.Madhusudhanan[1]
Research Scholar,AnnaUniversity,Chennai, India
E-mail:ssmadhu80@gmail.com

Dr.M.Moorthi[2]
Professor, Department of Electronics and Communication Engineering,
Prathyusha Engineering College, Chennai, India.
E-mail: moorthi.ece@prathyusha.edu.in

Abstract - The growth of web 2.0 provides a great medium for people to share opinions, comments and emotions. Web Opinion Mining or Sentiment analysis is one of the tasks in text mining that aims to develop system to automatically extract, identify and classify user's opinion from text written in natural language, user generated content or user generated media. Organizations are interested to get feedback on their products and customer service for business intelligence. Individuals are also interested in other's opinion for decision making. This survey presents the details of recent works available in the literature for the field of Sentiment Analysis. The existing techniques are grouped into categories based on the methods. The aim of this survey is to provide a summary of current research activities on this area and implementation of various useful techniques applied on sentiment analysis.

Keywords: Machine Learning,Natural Language processing,NavieBayes,Opinion Mining, Sentiment Analysis,SVM

1. Introduction

Today reviews or comments plays an impact on customer procuring through e-commerce websites. This sharing provides attitude, emotion, or reaction about customer. The comments may be about goods, or services or any related things. To make decision on the availability of opinion rich and huge volume of information (Example comments in Amazon, Flipkart, Twitter, Facebook etc.,). we need an intelligent system for learning opinions. This analysis is known as Sentiment Analysis or Opinion Mining. It will help the individuals, Organizations, and Government to know what the attitude of public about their particular product or service is [12] [48]. Opinion mining is a task which combines Natural Language Processing (NLP) and machine learning techniques to analyze text as positive, negative or neutral. For example," I had an Intel XOLO Q1100 for about 2 years. It works brilliantly, durable and reliable. Its display is beautiful and the phone is fast and perfect size to fit into my pocket", is a positive opinion. Opinions may be Direct and Indirect. The expression of sentiment on some objects is referred as Direct Opinions. For instance," Sony Xperia S is excellent phone with excellent Camera Quality and Gaming", is a positive opinion for Sony mobile phone. Indirect opinions are comparing two or more objects with similarities and differences. For example, "Intel XOLO Q1100 is far better than iPhone. I look at the customization, ease of use, menus, and speed everything". In the above example, the author compares the features of mobile phones.

Subjectivity Detection is a technique to determine opinion as subjective or objective expression from a piece of text. For instance, (1) Digital Camera is a good device for taking photographs. (2) The quality of picture on this camera is good. Both the sentences contain sentiment bearing words good, despite first sentence is an objective or factual sentence (i.e., does not convey any sentiment) whereas second one depicts opinion about that camera, is a subjective sentence. Sentiment Classification is to organize the subjective sentence as positive, negative or neutral from the document, also known as polarity classification. Sentiment Summarization provides sentiment summary at aspect level.

The applications of Opinion Mining are: Brand Sentiment analysis helps to understand the tastes, preferences and customer patterns by mining unstructured data from blogs and social media. Competitor analysis is also important for organizations to compare with their peers and able to know their strength and weakness of their products. In marketing intelligence, business organizations collect feedback from customers through email or social media and analyze which aspects of the product or service they are having difficulty. This type of analysis is known as complaint analysis which detects new problems faced by the customers. In Audio and Video processing, opinion mining procedures are used as an input feature for text to speech synthesis, and online video analysis. In Financial industry, opinion mining is used to predict stock market and to analyze it. Government will take decision based on opinion polls collected from social web sites to know their strength and weakness.

Major challenges are addressed in various research works [12] [48] [61] [65]: Entity Identification is an important task in opinion mining. A sentence may contain multiple entity, the opinion mining system needs to identify on which entity the opinion is expressed. Opinion Holder Detection is a task of detecting opinion topics and opinion holder. Opinion Classification determines whether the opinion of the sentence is positive, negative or neutral. Opinion spam Detection is one of the major task used to identify the bogus opinions in reviews and forums. Sarcasm identification is a common technique that a sentence may contain implicit opinion without the presence of any opinion bearing words, identifying such sentence is a major issue in opinion mining.

The objective of this work provides methodologies and recent developments of Sentiment Analysis that can be applied in day to day activities. This paper is organized as Section I Introduction to Opinion mining. Section II System Architecture is discussed in detail manner. Section III discussion about Opinion Mining tasks. Section IV Future Challenges and finally with Conclusion.

2. System Architecture

Steps involved in opinion mining are shown in Fig. 1. (i) preprocess the review text. (ii) Transformation of Opinion Text (iii) Dimensionality Reduction Techniques (iv) Modelling Techniques and (v) Sentiment Analysis

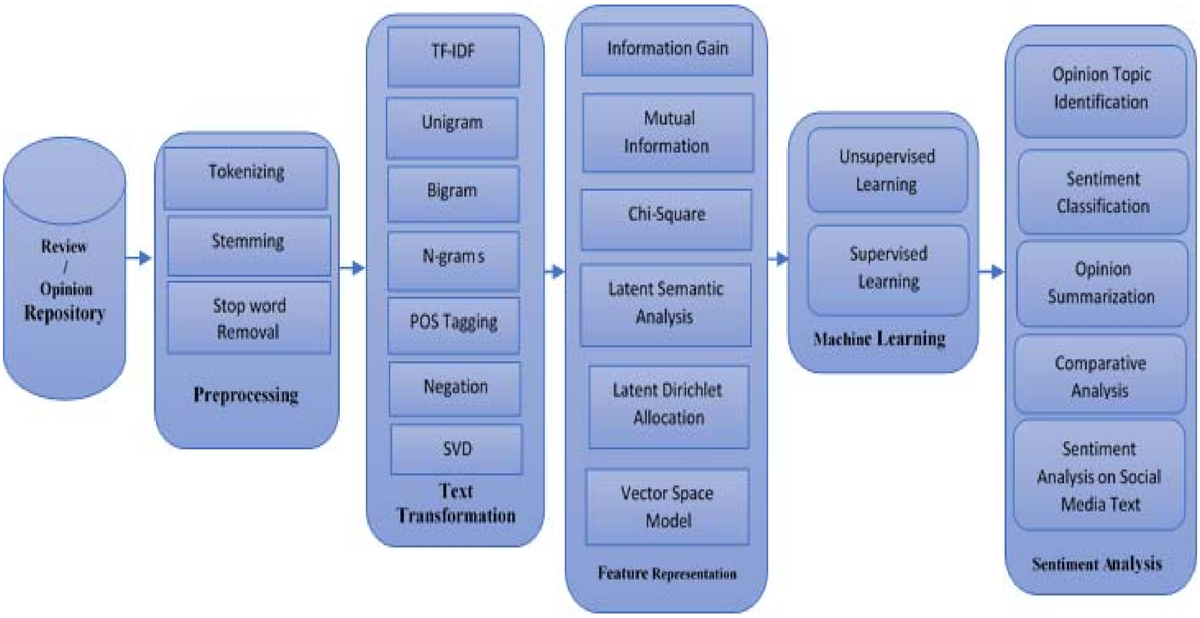

Fig 1 - Various Steps involved in Opinion Mining

2.1 Preprocessing

Tokenization refers to the procedure of splitting a set of text into meaningful words (stems), phrases or symbols.Stop words are function words like prepositions, articles, conjunctions and pronouns, providing language structure instead of content. These terms do not affect category discriminations. Additionally, common words like ‘a’ and ‘of’, may be removed as they recur frequently so that it is not discriminating for a specific class. Generic terms are detected by a threshold on the quantity of documents the term appears in, for instance, if it is present in more than 50% of the texts, or through the provision of a stop word list. Stop words are language as well as field-specific. On the basis of the classification tasks, removal of terms that are crucial predictors may be risked, for instance, the term ‘can’ discriminates between aluminium as well as glass recycling.

Word stemming is a rough pseudo-linguistic procedure which discards suffixes for reducing words to their stem. For instance, the words searching, searched, searches may be conflated to one stem – search. The common practice of stemming or lemmatizing, combining several word forms like plural or verb conjugation into one singular word decreases features number to be regarded.

Within Parts of Speech (PoS), the total contents of the text are denoted by unigram as well as N-gram, and are split into two categories: the first comprising single terms known as unigram, the second comprising multi-words known as N-gram. Those features with greatest relevance are regarded for sentiment classification.

2.2 Transformation of Opinion Text

In opinion mining, features are selected by the words, terms or phrases that contain opinion as positive or negative.To improve the performance, an efficient feature selection techinque is required to apply after preprocessing. Some of the text representations are (a) TF-IDF (Term Frequency - Inverse Document Frequency) [21],[63] which consider input text as a feature vector by calculating the term frequencies of the whole documents (b) Position of information can be encoded into a feature. Unigram feature [9],[23],[71] represents a single token of word; bigram feature [5], [23],[76],[78] represents two tokens of words and more than two tokens represents *n*-grams [28],[33] feature. (c)Part-of-Speech Tagging(POS) [64] is a crude form of word sense disambiguation and is used to identify the part of speech on the preprocessed text. In most of research work, adjectives have been used as a feature. (d) Word based features are generated from parsing a syntax tree which is used to identify relationships between sentiment expression and the subject term. (e) Negation words are indirectly considered as a feature for text segmentation.(f) Some time topic information can be employed as a feature. (g) Singular Value Decomposition(SVD)[5] is used as an input for Latent semantic analysis(LSA)[49].SVD is employed to reduce the dimension space of document vector by transforming text into matrix form.

2.3 Dimensionality Reduction Techniques

The high dimensionality of data are not useful for learning process. To capture relevant features different techniques are applied to reduce it to low dimension features(also known as Dimensional Reduction technique).

2.3.1 Latent Dirichlet Allocation (LDA)

LDA is a probabilistic document model where the documents are considered as a mixture of latent topics[18],[57],[74]. For each topic T, the model finds a conditional Probability distribution $P(W|T)$ where W as word.Finally the word-topic is formed as matrix in which word meanings are showed in the row.

2.3.2 Vector Space Models (VSM)

VSM seeks to model words directly [53],[73]. It explicitly learns semantic word vectors by applying SVD to factor on to the term document co-occurrence matrix. The matrix values are normalized before applying SVD. A k-dimensional representation for a given word is obtained through the entries corresponding to the k largest singular values of the words basis in the factored matrix.To increase the performance of retrieval and categorization, we can use TF-IDF weighting scheme to transform the values.

2.3.3 Latent Semanticc Analysis (LSA)

LSA is a statistical technique for extracting and representing the similarity of meaning of words*(Manning et al., 2008)*.It uses SVD method to construct a low-rank LSA approximation C_k .Let r be the rank of the M X N matrix,then SVD of C can be

$$C = U\sum V^T \tag{1}$$

where U be the M X M matrix whose columns are the orthogonal eigenvectors of CC^T ,and V be the N X N matrix whose columns are the orthogonal eigenvectors of C^T . $C.C^T$ is a transpose of a matrix C.

2.3.4 Chi-Square (Chi)

In statistics, χ^2 feature [1],[49],[53],[73] selection is applied to two independent events A and B if $P(AB) = P(A)P(B)$ or, equivalently, $P(A|B) = P(A) and P(B|A) = P(B)$ .The occurances of the term and occurence of the class is selected as a feature. Then ranking can done with the following Eq.2

$$\chi^2(\mathsf{D},t,c) = \sum_{e_t\in\{0,1\}}\sum_{e_c\in\{0,1\}}\frac{(N_{e_te_c}-E_{e_te_c})^2}{E_{e_te_c}} \tag{2}$$

where e_t and e_c are defined as in mutual information equation.N is the observed frequency in D and E is the expected frequency.

2.3.5 Mutual Information (MI)

MI of two random variables is a quantity that measures the mutual dependence of two random variables [1],[48],[49],[72].MI measures presence/absence of a term contributes to make the correct classification decision on C,

$$I(U;C) = \sum_{e_t\in\{0,1\}}\sum_{e_c\in\{0,1\}} P(U=e_t, C=e_c)\log_2\frac{P(U=e_t, C=e_c)}{P(U=e_t)P(U=e_c)} \tag{3}$$

where U is a random variable that takes values $e_t = 1$(the document contains term t) and $e_t = 0$ (the document does not contain t), and C is a random variable that takes values $e_c = 1$ (the document is in class c) and $e_c = 0$(the document is not in class c).

2.3.6 Information Gain(IG)

IG calculates the relevance of a feature based on the probability of the word exist in a particular class[1],[48],[53].

$$IG(D,t,c) = H(p_D) - \sum_{x \in \{D_{t^+}, D_{t^-}\}} \frac{|x|}{|D|} H(p_x) \tag{4}$$

where H is entropy, D is the training set, and D_{t^+}, and D_{t^-} are the subset of D with term t, and the subset of D without term t, respectively.

Table 1: Summary of recent work related to Sentiment Analysis

Reference & Year	Contribution	Text Representation	Classifier Used	Application
[21],2013	Reevaluates objective sentiment words in SentiWordNet for efficient Sentiment classification.	Term frequency	SVM	Movie
[63],2013	Language independent feature weighting method is proposed for Sentence level Subjectivity detection.	Pruned Inverse Class frequency	Fuzzy Control System	
[19],2012	Develop a review-summarization system in mobile environment	Bag-of-word	SVM and LSA based Feature	
[9],2004	The minimum cut algorithm based on graph was employed to help sentiment classification	Unigram	SVM	
[64],2004	Automatic Polarity Classification of Movie Reviews	Filtered POS Tagging	Naive Bayes, Markov model	
[78],2013	A domain-dependent sentiment lexicon is constructed using supervised learning method for sentiment classification	Unigram, Bigrams	SVM	Product
[23],2013	Sentiment Thesaurus is developed and used for Cross Domain Sentiment Classification	Unigram Bigrams	Sentiment Sensitive Thesaurus	
[18],2012	A Weakly supervised learning method is proposed to classify document level sentiment classification and conjunction with topic detection	LDA with collapsed Gibbs Sampling	Joint Sentiment Topic Model	
[10],2011	A new embedding mechanism of n-grams called "latent n-grams" is used to model n-gram features and supervised learning technique is employed for sentiment classification.	Bag-of-n-grams	SVM	
[76],2010	Lexicon Enhanced method for sentiment classification	Unigram, Bigrams	SVM	
[74],2013	A novel approach is proposed to jointly extract aspect and aspect dependent sentiment lexicon from online customer reviews.	LDA with collapsed Gibbs Sampling	Joint Aspect model	Restaurant Hotel
[33],2013	Starlet approach is developed for multi document summarizes from evaluative text	n-grams	A*Search	
[7],2014	A hybrid approach of extracting opinion from social media text to predict the outcome of election result	Principal Component Analysis	SVM, Naive Bayes, Maximum Entropy, ANN	Twitter

[71],2013	A framework has developed for analyzing and visualizing the public sentiment by dictionary based and machine learning based approach	Unigram	Naive Bayes	
[5],2013	A framework is developed to improve Text-To-Speech synthesis which automatically synthesize expressive speech at the sentence	Bigrams	Associative Relational Network, LSA	
[56],2013	Dynamic Sentiment Lexicon is constructed for improving Sentiment Analysis.	Bag-of-Words	Ad boost with Logistic Model	Cancer
[28],2014	Develop a sentiment ontology to conduct sentiment analysis of online opinion post.	n-grams	SVM	Stock Market
[62],2015	A Dual prediction algorithm is proposed for improving Sentiment Classification.	DSA model	Prediction Algorithm	Product Review
[25],2015	A Joint Sentiment Framework for Sentence Level Sentiment Classification	Candidate Generation Model	Segmentation Ranking Model & Classification Model	Twitter & Movie Review
[24],2016	A neural network based is proposed for Sentiment Analysis	Word Features	Neural Network Algorithm	Twitter
[29],2016	A framework called SWIMS to determine the feature weight for Sentiment Analysis	Part of Speech, PMI and Chi Square	SVM	Large Movie Review, Multidomain, & Cornell Movie Review
[30],2016	A framework called MOMS to determine the feature weight for Sentiment Analysis	MOMS Model	SVM	
[45],2017	Machine Learning Techniques are used to classify reviews	Unigram, Bigram& Trigram	SVM, NB, &ME	Rotten Tomatoes, Movie Review
[55],2017	A rule-based classification scheme is proposed to enhance Sentiment Analysis	Part of Speech	RuleBased Classifier	Drug, Car & Hotel
[75],2017	Multi-text summarization technique is used to find the most informative sentences	Part of Speech	K-medoids Clustering Algorithm	Hotel

2.4 Modelling Techniques

Modelling techniques are applied to built and predict data accurately based on selected features.There are two main approaches-Lexicon based and machine learning based methods[4],[12],[48],[49],[53],[65].Lexicon based approach involves calculating semantic orientation (SO) of words or phrases in a document. The creation of Dictionaries can be done manually or automatically using seed words. The adjective words are extracted from the given text then its corresponding SO values are calculated from the lexicon dictionary. Examples are WordNet,SentiwordNet, and Emotion library. Machine Learning is a process of learning meaningful information from data by a machine with less effort of human intervention. Types of Machine learning are (1) Supervised Learning and (2) Unsupervised Learning.Summarizes of recent work related to opinion mining are shown in Table. 1.

2.4.1Supervised learning : This requires labelled training data and classify it as a binary classification problem i.e., positive and negative [5],[45],[55],[75],[76]. Given the training data, the system learns a classification model by one of the common classification algorithms such as Naive Bayes (NB), Support Vector Machine (SVM), and Maximum Entropy (ME).

2.4.1.1 Navie Bayes Classifier(NB)

Naive Bayes is a probabilistic learning method that consider terms independently [5],[7],[64],[71]. Given a collection of N documents d_j where j= 1 to N, then each document is represented as a sequence of T terms $d_j = t_1, t_2, ..., t_T$, the probability of a document d_j occurring in class c_k is given by Eq.5

$$P(c_k \mid d_j) = P(c_k)\prod_{i=1}^{T} P(t_i \mid c_k) \qquad (5)$$

where $P(t_i \mid c_k)$ is the conditional probability of term t_i.

In opinion mining, NB Classifier helps to predict the polarity of data based on baye's theorem. Thus the conditional probability of a word with positive and negative is calculated by Eq.6

$$P(Sentiment \mid Sentence) = \frac{P(Sentiment)P(Sentence \mid Sentiment)}{P(Sentence)} \quad (6)$$

and frequency of each classes are calculated by Eq.7

$$P(Word \mid Sentiment) = \frac{P(No.of\ \ words\ occur\ in\ class + 1)}{P(No.of\ \ words\ belonging\ to\ a\ class + Total\ \ No.of\ \ words)} \quad (7)$$

The strength of the NB classifier is easy to model,requires small amount of training data with short computational time to estimate the parameters,and achieves good results on various applications.Due to independent assumptions of variables,NB has loss of accuracy.

2.4.1.2 Maximum Entropy (ME)

Maximum Entropy (ME) classification [7],[12],[49],[53] is effective for natural language processing applications and generally performs better than Naive Bayes for text classification. In ME, the $P(c \mid d)$ is estimated in exponential form as in Eq.8,

$$P_{ME}(c \mid d) = \frac{1}{Z(d)} exp(\sum_i \lambda_{i,c} F_{i,c}(d,c)) \quad (8)$$

Where, Z (d) is a normalization function. Fi,c is a feature/class function for feature fi and class c, as in Eq.8,

$$F_{i,c}(d,c') = \begin{cases} 1 & n_i > 0 \,\&\, c' = c \\ 0 & otherwise \end{cases} \quad (9)$$

As per Eq. 9, a feature or class function is given as output if the bigram appears else it is considered as negative. In ME, relationship between the features are not considered as compared to Naive Bayes, and thus performs better when conditional independence assumption are not met.

2.4.1.3 Support Vector Machine (SVM)

Support vector machines (SVM) attempts to separate positive and negative training samples by finding the optimal possible surface (hyperplane/decision boundary) to divide the samples. The linear hyperplane is so placed that the data is separated with maximum margin [37],[39],[49],[53]. For the linearly separable data, the data points that lie on hyperplane margin are known as support vector points and the rest of the points are ignored. SVMs are efficient in dealing with learning tasks where the number of features are large with respect to the number of training instances.

In the context of Opinion mining,Support Vector Machines (SVMs) are supervised learning methods used for classification.Many researchers have reported that SVM is widely used in sentiment classification [4],[7],[9],[10],[21],[22],[28],[48],[78].For a given set $C = \{+1,-1\}$ and two pre-defined training sets, i.e., a positive set, $T_r^+ = \sum_{i=1}^n (d_i,+1)$ and negative sample set $T_r^- = \sum_{i=1}^n (d_i,-1)$,the SVM finds a hyperplane that separates the largest possible from both sets.Each training sample is converted into a vector x_i for its corresponding document d_i in pre-processing step.Then the positive and negative sample set is given by $T_r^+ = \sum_{i=1}^n (x_i,+1)$ and $T_r^- = \sum_{i=1}^n (x_i,-1)$ .Then classification task is used to discover which side of the hyperplane a test sample falls into.SVM provides very good performance on experimental results and low dependency on data set.For extremely nosiy data, SVM fails to correctly classify the samples due to ambiguity and sparseness in the test samples.

2.4.1.4 Artifical Neural Network (ANN)

ANN is based on the idea of perceptron and it models the output as nonlinear function from the input of linear features. ANN is a multilayer perceptron with input layer, hidden layer and an output layer. All the input nodes are connected with each other which contain perceptron variables and all nodes are connected with hidden layers also. The neurons in hidden layers may be attached with other hidden layers or the output layers. In output layer, one neuron is present for binary predication or more in case of non-binary prediction. The efficiency of ANN depends on input and activation functions of neurons, network architecture and the weight of each input connection [7],[24],[25],[37],[48],[49],[53],[63]. The behavior of ANN is based on the current values of the weights when the first two aspects are fixed. During training, the weights are initally set to random values, and the training instances are repeatedly run through network. The values for the input are varied and output of the network is compared with the desired output. In case of error, the weigths are adjusted slightly in the direction such that the output values are closer to desired output. An advantage of this technique is that it does not impose any sort of restriction with repect to the starting data and also it is a robust technique for the presence of nosiy data. The disadvantage is, it requires more time for execution due to lot of parameters involved on a number of layers. To overcome the sentiment polarity issues, Duyu Tang,et al [24] have developed a neural network algortihm by encoding sentiment information together with context of word. The propose model wuld learn the sentiment embeddings effectively.

2.4.2 Unsupervised Learning approach is used to determine the semantic score or semantic orientation (SO) of term within the document [55],[[69],[70],[79]. If the average SO value is above threshold value, then document is considered as positive otherwise it is negative. In this appraoch, labelled data is not required. The SO of a given word is estimated based on the difference of its strength of its association with a set of positive words and negative words. Point wise mutual information (PMI) value represents the strength of the semantic association between words. For a term t, and seed term sets (Sp for positive set and Sn for negative set), then the orientation value O(t) is given by:

$$O(t) = \sum_{t_i \in S_p} PMI(t,t_i) - \sum_{t_i \in S_n} PMI(t,t_i) \qquad (10)$$

3. Tasks in Opinion Mining

3.1 SentimentClassification

It classifies each document d_i in document D is labelled as positive or negative category. Most of the prior works focused on binary classification problem.(i.e.,identifying opinion as Positive or Negative).But it is often useful to identify document as Strongly Positive, Positive, Neutral, Negative and Strongly Negative by comparing several opinion reviews and rank them for accurate classification. This type of classification is called multiclass classification. In general, there are various levels in sentiment classification:-Document Level Sentiment Classification, Sentence level Sentiment Classification and Aspect Level Sentiment Classification [12], [26], [29], [30], [45],[48], [55], [61], [65].

3.1.1 Document Level Sentiment Classification

Document level sentiment classification considers the whole document as opinionated document and classifies it as positive or negative. Most recent works on sentiment analysis are shown in Table.I. A mincut-based algorithm was proposed by authors [9], to classify each sentence as subjective or objective.A sentence graph was constructed to classify the document as positive or negative.The objective sentences are not considered so as to improve sentence based subjectivity classification. The authors [69],[70] used unsupervised approach to classify a document as positive or negative by determining SO of the term within the document. To determine SO of a given word is the difference between the PMI (Pointwise Mutual Information) of the phrase with two sentiment words ’excellent’ and ’poor’. PMI (P, W) measure the statistical dependence between the phrase P and the word W based on their co-occurrence in a given corpus or over the web. If the average SO value is above threshold value, then document is consider as positive otherwise it is negative. It does not require any labelled data. To enhance the performance of sentiment classification using SentiWordNet, the authors [76] proposed an approach by combining machine learning technique and semantic orientation method into one framework. Three features Content-free, Content-specific and sentiment features are extracted from review after three features are built, SVM Classifier is used to classify the dataset. The proposed method achieves better classification. Abbasi et. al [1] proposed a feature selection method IG and GA (Genetic Algorithm). These two methods have combined into a new algorithm called Entropy Weighted Genetic Algorithm (EWGA) to improve the accuracy of sentiment classification.A neural network model is employed for document vector representation by the quthors Dang et al.[26]. The sentence representation is learned by the convolutin neural network then the semantics of sentences and their relation are encoded in doucment representation with gated recurrent neural network.

3.1.2 Sentence Level Sentiment Classification

Sentence Level Sentiment Classification is a task to classify a sentence or a clause of the sentence as subjective or objective.For a subjective sentence, it classifies as positive, negative or neutral opinion [5],[25],[59],[60],[63].In document level classification,a whole document is considered and classify whether the overall document is expressed as positive, negative or neutral, where as in sentence level, a single document may contain multiple opinions,so the documents are broken down into sentences, then sentences are evaluated by utilizing lexical method in order to determine their semantic orientation.For example:*"It’s been a week since I got my Galaxy S4 and this thing is super awesome.My previous phone was the S3 which I loved to death and this time I’m super excited to have the S4 in my hands. I had absolutely zero issues with the S3 but the S4 looks* *so sexy with the metal edges"*.It contains multiple opinions about Galaxy S4 mobile phone on above example and can classify it as positive.

The author Yang et al. [77] proposed a simple strategy based on feature selection for sentence level classification where highly ranked features are selected. A classifier was built using these features and then applied to target domain. McDonald et al.[50] presented a hierarchical sequence learning model similar to conditional random fields (CRF) to jointly learn and infer sentiment at both the sentence-level and the document-level. The results showed that learning of both levels jointly improved classification accuracy. A rule based approach is proposed by authors[3]. The method classifies subjective and objective sentences from reviews, blogs and comments.Bag of Sentences (similar to bag-of-words) is constructed from reviews, POS tagging is applied, and only subjective sentences are used for further processing. SentiWorNet is used for word semantic score calculation of each words, then the rules are constructed for sentence level polarity calculation.Finally the opinion strength for both sentence and feedback level is calculated.A joint segmentation and classification framework for sentence level sentiment analysis by the authors Duyu Tang et al.[25].The proposed work creates segmentation at sentence level and find the polarity simultaneously. Sentiment polarity between phrases and words are handled efficiently by the proposed method.

3.1.3 Aspect Level Sentiment Classification

It has following steps; first step is to identify and extract object features that have been commented on by an opinion holder (e.g.Picture Quality,battery life). In second step polarity of opinions are determined then feature synonyms are grouped on the third step. Fourth step is to train the classifier using small amount of data set. Fifth step is to test the dataset using training data. Finally we can classify it as positive,negative or neutral [16],[20],[29],[30],[36],[44],[62],[72],[74]. For example,consider the **Nikon COOLPIX S3300** review from Amazon. *"I just got this camera yesterday and have been taking pictures in all different kinds of lighting. This camera has done great in even a dark room. I am very pleased with how quiet it is which will be great when shooting video's.Haven't been able to get a great super close up yet but I think that is due to my lack of understanding on just how to use all the functions of the camera just yet.I'm not a professional photographer but as a capture every moment mom of two I have taken thousands upon thousands of pictures.I've been through many camera's (Samsung,Kodak,Fuigi)to name a few none have come close to the quality of picture of this Nikon. I am very happy with this camera so far.I love how compact it is and the fun colors it comes in."* The reviewer gave positive feedback about the picture quality of the digital camera and compare the features with other products.

A lexicon-based algorithm for aspect level sentiment classification is proposed in their work [52], the method determines the sentiment orientation of a sentence. The sentiment orientation of a sentence has been calculated by summing up the orientation scores of all sentiment words in the sentence. The sentiment score +1 is assigned to positive sentiments and -1 is assigned as negative sentiments. The authors [43] proposed supervised learning method to identify several specific types of opinions. A semi-supervised learning algorithm [2] was used to learn from a small set of labeled sentences and a large set of unlabeled sentences. The aspect model based on unsupervised classification is proposed by the authors [41]. The goal is to discover customer satisfaction on a particular product, service, or business. Their work uses unlabelled data for aspect identification. It takes set of reviews and some predefined aspect as input to identify polarity of each aspects from each review. Multi-aspect segmentation algorithm is proposed to split multiple aspect sentences into multiple single-aspect units. Opinion polling generation is a final task which gives quantitative indications of user as positive and negative opinions about product or business. Hownet lexicon is used to find the polarity of each aspect expressed in each review.The authors Farhan Hassan Khan et al. [29] have developed a framework called SWIMS which performs subjectivity detection by determining the feature weights and a lexical approach is used for polarity selection on different part of speech simultaneously To achieve high accuracy, a model called 'Intelligent model selection' is also proposed for cross valkidation. A semi-supervised framework for feature selection called Multi-Objective Model Selection(MOMS) is proposed by the authors Farhan Hassan Khan el at.[29]. By utilizing SVM, the feature weights are determined and the proposed model is used to enhance the classification performance.

3.2 Cross-Domain Sentiment Classification

In Sentiment classification ,Classifier trained for one domain couldn't achieve same result one on other domain, because of diffferent words used to construct opinion . So domain adaptation method is needed to solve the above issue. The original domain is termed as source domain, and the new domain as target domain. In [71], the authors employed structural correspondence learning (SCL) for domain adaptation. During training, labelled reviews from a source domain and unlabelled reviews from both the source and target domains were used. SCL identifies a set of *m*features or pivot features occuring in both domains. During classification, the combined features are used to classify the original and the target domains.

A spectral feature alignment (SFA) algorithm aligns domain-specific words from different domains into unified clusters was presented by the authors [57]. It is similar to SCL and uses labelled examples from source domain and unlabeled examples from target domain. A bipartite graph is constructed showing the domain-independent words and the domain-specific words and a domain specific word is linked to a domain-independent word if they co-occur. The authors [23] proposed a sentiment sensitive distributional thesaurus using labelled data from source domain and unlabelled data from source and target domains. The proposed method performs significantly better on a benchmark dataset containing Amazon user reviews.

3.3 Sentiment Analysis on Social media text

Mining online social content has lot of challenges compared with normal reviews, because of very short message, no verbose on interaction, using colloquial words, no specific topic, may vary from political to daily context, numerous and misspelling [25],[32],[34],[35],[66],[67],[71]. For example, 95% of exchange content in social network at least one abbreviation (such as ”gr8”for”great”, ”5” for ”fine”, ”ni8”for ”night” and ”tnx”for ”Thanks”) of Standard English. Machine learning technique and dictionary-based approach are implemented by the authors.It doesn’t require any reference corpus and any training data also. In their work three dataset from real-world social websites like Digg, Twitter and MySpace are employed. Three machine learning approaches such as Naive Bayes, Maximum Entropy, and SVM using unigram as features are compared with lexicon-based classifier. In [34], sentiment classification of Twitter postings (or tweets) was investigated. The authors took a supervised learning approach was used for classification, and with the traditional features, hashtags, smileys, punctuations were used as features. Results show that the performance was effective in classification.

3.4 Opinion Topic Identification

Topic modelling is an unsupervised learning method that specifies a probabilistic procedure by which documents can be generated. The clusters with a topic and its probability is formed as an output, for the entire document. Probabilistic Latent Semantic Analysis(pLSA) [38] and LDA [11] are utilized for this. For sentiment analysis, the basic model (e.g., LDA) can be used to find the aspects and sentiments combinely.Topic modelling can be applied to extract aspects words and sentiment words.The authors [51] have proposed a first joint model known as Topic Sentiment Mixture model (TSM) by integrating sentiment into pLSA to detect sentiment from all the topics.

The authors Bordy and Elhadad [16] have proposed a method to classify aspects using topic model and then find aspect-specific sentiment words by considering adjectives words.A hybrid model by combining ME and LDA have proposed by the authors[80] where syntactic features are used to separate aspect words and opinion specific words. The authors [40] proposed a Aspect and Sentiment Unification Model (ASUM). In their work, the algorithm detects sentiment coupled aspect words with respect to different sentiments. A model based on weakly supervised sentiment classification is proposed by the authors [18] known as Joint Sentiment Topic model (JST) based on extension of LDA and also proposed a parameterized JST known as Reverse-JST. In the porposed method classify as if the probability of a positive-sentiment document is greater than the negative-sentiment document then document textid is classified as positive and vice versa. It also detects sentiment-topic pairs under different sentiment labels.

A Joint Aspect Sentiment (JAS) model is proposed by the authors [74] to extract meaningful description about the aspect, and aspect-level classification .The proposed method gives detail information in terms of aspect.

3.5 Opinion Summarization

In online comments, many comments are lengthy and only few sentences have contains opinions for some popular products. So it is impossible to make decision for a customer to know about the product by a glance. Opinion summarization is a task to summarize opinion by providing sentiment polarity, degree and some related information about that product. So customers can easily get conclusion and also product manufacturer can know the pros and cons on that product from the opinion summary.Hu and Liu [52] proposed a work on customer reviews. Their work involves three tasks: (1) extracts the features; (2) classify as positive or negative from the extracted features; and (3) finally Summary is generated from the classified information. Chien-Liang et al.,[19] have proposed review summarization system in mobile environment. First the system retrieves movie review from the database. Machine Learnining technique is applied to classify it as positive or negative. Then LSA based filtering approach is used to provide summary on the classified content.Hua et al. [75] have proposed a novel multi-text summarization techniques for identyfying the top-k most informative sentences of hotel reviews by implementing k-mediods clustering algorithm. The proposed algorithm partitioned sentences into k groups. The medoids from that groups are used for final summarization result.

3.6 Comparative Analysis

In business intelligence, now a day’s organizations are collected feedback about their product from the customer through online. When more and more people express their recommendation, then product receives good response rapidly. Comparative analysis is a task to visualise and compare customer’s opinions of different products. So user can analyse strength and weakness about that product. For a potential customer, he / she may see a visual side-by-side and feature-by-feature comparison of consumer opinions on the product that helps him/her to decide which product has to buy. For example,consider **Dell Inspiron i15RV** (personal computer) review taken from amazon.*"We switched to Dell computers for our medical staff after using Toshibas for years until the Toshibas became problematic. The Dell is sturdy and reasonably fast for basic computing and daily needs"*.In the above review, the performance of Toshiba computer is compared with Dell Inspiron computer. Liu, Hu and chen [14] proposed a framework to compare and analyze opinions. A prototype system called Opinion Observer is implemented. Two tasks were performed on visualization technique; (1) Extract the product features; (2) from the extracted feature, classify it as positive or negative.One can envisage and relate opinions of different product. The proposed system provides a comparative statement in detail manner.

4. Future Directions

A review of opinion mining tasks with number of feature selection and classification techniques are discussed in the previous sections.Some of the further work that need to be explored are listed below for further research

- When opinions can be expressed in indirect form, common sense knowledge is required to recognize the sentiment.To analyze the complex sentences which indirectly convey opinions is a one of the challenging task in Sentiment Analysis.
- Sentiment analysis on social media content is still a big challenge due to noisy text(i.e.,presence of colloquial words).
- Most of the sentiment classification work considered only subjective sentences and neglect objective sentence. So the objective sentences are need to included for training the classifer.
- Opinion Question Answering(Opinion QA) is a one of the challenges in opinion mining.
- Few works have done on identifying sarcasm sentences. So a system is required to develop and integrate sarcasm sentences with opinion mining task.
- A more domain independent sentiment lexicon has to be developed for multi domain and multi-class sentiment classification.
- The multimodal sentiment analysis can be used to mine the useful information,when users post comments in audio or audiovideo formats instead of text.
- Developing Recommentation systems based on opinion analysis, is also one of the challenges in opinion mining.

5. Conclusion

In an intelligent information system, machine learning techniques and NLP plays a vital role in major areas like text mining, opinion mining, information retrieval, computational analysis and statistical analysis.In this survey, so far various methods of opinion mining has been discussed and attempted to provide an overview on this. Most of the current research work has been done on an unstructured text only. This can be extended to deep analysis on speech synthesis and online video sentiment analysis.

References

[1] A.Abbasi, H. Chen, and A. Salem, Sentiment Analysis in Multiple Languages: Feature Selection for Opinion classification in Web Forums, ACM Trans. Information Systems, vol. 26, no. 3, pp. 1-34, 2008.
[2] A.Aue and M. Gamon, Customizing Sentiment Classifiers to New Domains: A Case Study, Proc. Recent Advances in Natural Language Processing (RANLP), 2005.
[3] Aurangzeb khan, Baharum Baharudin, and Khariullah Khan, Sentiment Classification Using Sentence-level Lexical Based Semantic Orientation of Online Reviews, Trends in Applied Sciences Research, 2011, Pp 1141-1157, ISSN 1819-3579.
[4] Alpaydin, E.,Introduction to machine learning. Cambridge, Mass: Mit Press,2004.
[5] Alexandre Trilla and Francesc Alias, Sentence-Based Sentiment analy- sis for Expressive Text-to-Speech, IEEE Transaction on Audio, Speech and Language Processing,Vol21,No2,Feburary 2013.
[6] Aldo Gangemi, et al, Frame-Based Detection of Opinion Holders and Topics: A Model and a Tool, IEEE Computational Intelligenence Magazines,page(s):20 -30, February 2014
[7] Anjaria, Malhar, Guddeti, & Ram Mohana Reddy, Influence factor based opinion mining of Twitter data using supervised learning, IEEE international Conference,Page(s):1-8, January 2014
[8] Baccianella,S., Esuli,A., And Fabrizio, S. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the Annual Conference on Language Resourcesand Evaluation (LREC10).
[9] B. Pang and L. Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, Proc. 42th Ann. Meeting on Assoc. for Computational Linguistics (ACL), pp. 271- 278, 2004.
[10] D. Bespalov, B. Bai, Y. Qi, A. Shokoufandeh, "Sentiment Classification Based on Supervised Latent n-gram Analysis",20th ACM Conference on Information and Knowledge Management, October 2011
[11] D. Blei, A. Ng, and M. Jordan, Latent Dirichlet Allocation, Journal of Machine Learning Research, 3:9931022, January 2003.
[12] Bo Pang and Lillian Lee. 2008. Opinion Mining and Sentiment Anal- ysis, Now Publishers Inc.
[13] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, Thumbs up? Sen- timent Classification using Machine Learning Techniques, Proceedings of EMNLP 2002.
[14] B. Liu, M. Hu, and J. Cheng, Opinion Observer: Analyzing and Comparing Opinions on the Web, Proc. Intl Conf. World Wide Web, pp. 342-351, 2005.
[15] Bradley,M. and Lang, P. 1999. Affective Norms For English Words (Anew): Stimuli, Instruction Manual And Affective Ratings. Tech. Rep., Gainesville, Fl. The Center For Research In Psycho physiology, University Of Florida.

[16] Brody S, Elhadad N. An Unsupervised Aspect Sentiment Model for Online Reviews [C], Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics: June 1-6, 2010. Los Angeles, California, USA, 2010: 804-812.
[17] Blitzer,John Mark Dredze,and Fernando Pereira, Domain Adaptation with structural Corresponding learning,Conference on Emprical Meth- ods in Natural Language Processing,2006
[18] Chenghua Lin, Yulan He, Richard Everson, Weakly Supervised Joint Sentiment-Topicdetection From Text, IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 6, June 2012
[19] Chien-Liang Liu Wen-Hoar Hsaio ; Chia-Hoang Lee ; Gen-Chi Lu ; Jou, E, Movie Rating and Review Summarization in Mobile Environment, IEEE Journal & Magazines, May 2012, Page(s):397 - 407
[20] Chung,C. and Pennebaker, J. he Psychological Function Of Function Words. In Social Communication:Frontiers Of Social Psychology, K. Fiedler, Ed., 343359,2007.
[21] Chihli Hung and Hao-Kai Lin,Using Objective Words in SentiWordNet to improve Word-of-Mouth Sentiment Classification, IEEE Intelligent Systems, Page(s): 47 54, March 2013.
[22] Cristianini, N., & Shawe-Taylor, J., An introduction to support vector machines and other Kernel-based learning methods (1st ed.). Cambridge University Press.2008.
[23] Danushka Bollegala, David Weir and John Carroll, Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus, IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 8,Page(s):1719 1731, August 2013.
[24] Duyu Tang, Bing Qin, Furu Wei, Li Dong, Ting Liu, Ming Zhou ”Sentiment Embeddings with Applications to Sentiment Analysis ”,IEEE Transactions on Knowledge and Data Engineering,Page(s): 496 - 509 ,Volume: 28, Issue: 2, Feb. 1 2016.
[25] Duyu Tang, Bing Qin, Furu Wei, Li Dong, Ting Liu, Ming Zhou ,”A Joint Segmentation and Classification Framework for Sentence Level Sentiment Classification”,IEEE/ACM Transactions on Audio, Speech,and Language Processing , Page(s): 1750 - 1761,Volume: 23, Issue: 11,Nov. 2015.
[26] Duyu Tang, Bing Qin, Furu Wei, Li Dong, Ting Liu, Ming Zhou,”Document Modeling with Gated Recurrent Neural Network for Sentiment Classification”,Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 14221432,Lisbon, Portugal, 17-21 September 2015
[27] K. Dave, S. Lawrence, and D. M. Pennock, ”Mining The Peanut Gallery: Opinion Extraction And Semantic Classification Of Product Reviews ” In Proceedings Of The 12th Www, 2003.
[28] Desheng Dash Wu, et al, A Decision Support Approach for Online Stock Forum Sentiment Analysis, IEEE Transactions on Systems, Man,Cybernetics Systems, Page(s): 1 -11, 2014
[29] Farhan Hassan Khan, Usman Qamar, and Saba Bashir, SWIMS: Semi Supervised Subjective Feature Weighting and Intelligent Model Selection for Sentiment Analysis, Elsevier Knowledge-Based Systems, Pages 97- 11,Feburary 2016.
[30] Farhan Hassan Khan, Usman Qamar, and Saba Bashir,”Multi-Objective Model Selection (MOMS)-based Semi-Supervised Framework for Sentiment Analysis”,Springer Link ,Cognitive Computation,Volume 8, Issue 4, pp 614628,August 2016.
[31] K. Frantzi, S. Ananiadou, and H. Mima, Automatic Recognition of Multi-Word Terms: The C-Value/Nc-Value Method, Intl J.Digital Libraries, Vol. 3, Pp. 115-130, 2000.
[32] Georgios Paltoglou and Mike Thelwall,” Twitter,Myspace,Digg:Unsupervised Sentiment Analysis in Social Media”,Acm Transactions On Intelligent Systems and Technology,Vol. 3,No. 4,Article 66,Publication Date:September 2012.
[33] Giuseppe Di fabbrizio,Ahmet Aker and Robert Gaizauskas,”Summarizing online Reviews using Aspect Rating Distributions and Language Modelling”,IEEE Intelligent Systems,Page(s):28-37,May 2013.
[34] Go, A.,Huang,L.,and Bhayani, R. Twitter Sentiment Analysis. Tech. Rep., Stanford Natural Language Processing Group,2009.
[35] Harvey, M. 2010. Facebook Ousts Google In Us Popularity. The Sunday Times (3/17/10).
[36] V. Hatzivassiloglou and K. R. Mckeown, ”Predicting The Semantic Orientation Of Adjectives” In Proc. Of 8th Eacl, 1997
[37] Haykin, S. ,Neural networks: A comprehensive foundation (2nd ed.). Upper Saddle River, NJ: Prentice Hall. 1998.
[38] T. Hofmann, Probabilistic Latent Semantic Indexing, Proc. 22nd Ann.Intl Acm Sigir Conf. Research And Development In InformationRetrieval, Pp. 50-57, 1999.
[39] Huang, T. M., Kecman, V., & Kopriva, I. , Kernel based algorithms for mining huge data sets: Supervised, semi-supervised, and unsupervised learning. Studies in computational intelligence (Vol. 17). Secaucus, NJ,USA: Springer,2006.
[40] Jo Y, Oh A H. Aspect and Sentiment Unification Model for Online Review Analysis, Proceedings of the 4th ACM International Conference on Web Search and Data Mining: February 9-12, 2011. Hong Kong, China. ACM Press, 2011: 815-824.
[41] Jingbo Zhu, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou,Matthew Ma,Aspect-Based Opinion Polling From Customer Reviews, IEEE Transactions On Affective Computing, Vol. 2, No. 1, January-March 2011
[42] A.Kennedy and D. Inkpen,Sentiment Classification Of Movie Reviews Using Contextual Valence Shifters,Computational Intelligence, Vol. 22, No. 2, Pp. 110-125, 2006.
[43] M. Kim and E. Hovy,Extracting Opinions, Opinion Holders, And Topics Expressed In Online News Media Text,Proc. Workshop Sentiment and Subjectivity In Text, 2006.
[44] N. Kobayashi, K. Inui, And Y. Matsumoto,Extracting Aspect evaluation And Aspect-Of Relations In Opinion Mining,Proc. Joint Conf. Empirical Methods In Natural Language Processing And Computational Natural Language Learning, Pp. 1065-1074, 2007.
[45] Prayag Tiwari ,Brojo Kishore Mishra,Sachin Kumar,Vivek Kumar,” Implementation of n-gram Methodology for Rotten Tomatoes Review Dataset Sentiment Analysis”, International Journal of Knowledge Discovery in Bioinformatics (IJKDB),Volume7,Issue-1,Jan-June2107.
[46] T. Li, Y. Zhang, and V. Sindhwani, A Non-Negative Matrix Trifactorization Approach To Sentiment Classification With Lexical Prior Knowledge, Proc. Joint Conf. 47th Ann. Meeting Of The Acl and The Fourth Intl Joint Conf. Natural Language Processing Of The Afnlp, Pp. 244-252, 2009.
[47] C. Lin And Y. He, Joint Sentiment/Topic Model for Sentiment Analysis, Proc. 18th Acm Conf. Information And Knowledge Management (Cikm), Pp. 375-384, 2009.
[48] Liu, B., Web data mining: Exploring hyperlinks,2nd edition, New York: Springer,2011.
[49] Manning, C. D., Raghavan, P., & Schtze, H, Introduction to information retrieval. New York, NY, USA: Cambridge University Press. 2008.
[50] McDonald Ryan,Kerry Hannan, et al.,Strucutured models fpr fine-tocoarse sentiment analysis,in proceeding of Annual Meeting of the Association for Computational linguistics(ACL-2007),2007.
[51] Mei Qiaozhu, Ling Xu, Wondra M, et al. Topic Sentiment Mixture:Modeling Facets and Opinions in Weblogs[C], Proceedings of the 16th International Conference on World Wide Web: May 8-12, 2007. Banff,Alberta, Canada. ACM Press, 2007: 171-180.
[52] Minqing Hu and Bing Liu, Mining and Summarizing Customer Reviews. Proceedings Of The Acm Sigkdd International Conference On Knowledge Discovery & Data Mining (Kdd-2004, Full Paper), Seattle, Washington, Usa, Aug 22-25, 2004
[53] T.M. Mitchell, Machine Learning.New York: Mcgraw-Hill, 1997

[54] T. Minka, Estimating A Dirichlet Distribution, Technical Report, Mit, 2003.
[55] Muhammad Zubair Asghar , Aurangzeb Khan, Shakeel Ahmad, Maria Qasim,Imran Ali Khan,"Lexicon-enhanced sentiment analysis framework using rule-based classification scheme",PLoS ONE 12(2),February 23, 2017
[56] Nir Ofek, et al, Improving Sentiment Analysis in an Online Cancer Survivor Community Using Dynamic Sentiment Lexicon , IEEE International Conference on Social Intelligence and Technology (SOCIETY,Page(s):109-113), May 2013
[57] Pan, Sinno Jialin,et al.,, Cross-domain sentiment classification via spectral feature alignment , international conference of World Wide Web(WWW),2010.
[58] Pak, A. and Paroubek, P. 2010. Twitter As A Corpus For Sentiment Analysis And Opinion Mining. In Proceedings Of The Annual Conference On Language Resources and Evaluation(Lrec10).
[59] Read, J. 2005. Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classication. In Proceedings of The Acl Student Research Workshop. Association For Computational Linguistics, Morristown.
[60] E. Riloff, J. Wiebe, and T. Wilson, Learning Subjective Nouns Using Extraction Pattern Bootstrapping, Proc. Seventh Conf. Natural Language Learning At Hlt-Naacl, 2003.
[61] Ronen Feldman,Techniques and applications for sentiment analysis, ACM Communication, Vol: 56, No.4, Pages: 82-89, April 2013.
[62] Rui Xia; Feng Xu; Chengqing Zong; Qianmu Li; Yong Qi; Tao Li," Dual Sentiment Analysis: Considering Two Sides of One Review" ,IEEE Transactions on Knowledge and Data Engineering,Pages: 2120 -2133, Volume: 27, Issue: 8,Year- 2015.
[63] Samir Rustamov Mark A. Clements, Sentence-Level Subjectivity Detection Using Neuro-Fuzzy Models In Proceedings of the 4th Workshop on Association for Computational Linguistics, pages 108114, June 2013.
[64] Salvetti.F,Lewis.S.,& Reichenbach.C (2004),Automatic Opinion Polarity Classification of Movie Reviews,Colorado Research in Linguistic, June 2004.
[65] Tang, H., Tan, S., & Cheng, X., A survey on sentiment detection of reviews, Expert Systems with application, 10760- 10773,2009
[66] Thelwall,M.,Buckley,K.,Paltoglou,G., and Cai, D. 2010. Sentiment Strength Detection In Short Informal Text. J.Amer.Soc.Inf.Sci.Technol.61, 12.
[67] M. Thelwall, K. Buckley, & G. Paltoglou, Sentiment in twitter events, Journal of the American Society for Information Science and Technology 62 (2) (2011) 406 418.
[68] I.Titov and R. Mcdonald, A Joint Model Of Text And Aspect Ratings For Sentiment Summarization, Proc. Assoc. Computational LinguisticsHuman Language Technology (Acl-Hlt), Pp. 308-316, 2008.
[69] Turney, Peter D. 2002. Thumbs Up Or Thumbs Down? Semantic Orientation Applied To Unsupervised, Classification Of Reviews. Proceedings Of The 40th Annual Meeting Of The Association For Computational Linguistics (Acl'02), Philadelphia, Pennsylvania, Usa, July 8-10, 2002. Pp 417-424. Nrc 44946.
[70] Turney,P.D. and Littman, M. L. 2002. Unsupervised Learning of Semantic Orientation From A Hundred-Billion-Word Corpus. Cor Cs.Lg/0212012.
[71] Vu Dung Nguyen ; Varghese, B. ; Barker, A., The royal birth of 2013: Analysing and visualising public sentiment in the UK using Twitter,In IEEE International Conference on Big Data, Page(s): 46 54,October 2013.
[72] Wilson,T.,Wiebe,J., and Hoffmann, P. 2009. Recognizing Contextual Polarity: An Exploration Of Features For Phrase-Level SentimentAnalysis. Comput. Linguist. 35, 399433.
[73] Weiss, S. M., Indurkhya, N., & Zhang, T. (2010). Fundamentals of predictive text mining. London; New York: Springer-Verlag [74] XU Xueke, CHENG Xueqi, TAN Songbo, LIU Yue, SHEN Huawei, Aspect-Level Opinion Mining of Online Customer Reviews Communications, China , IEEE Journals & Magazines, 2013.
[74] Ya-Han Hua, Yen-Liang Chenb, Hui-Ling Choub,"Opinion mining from online hotel reviews A text summarization approach",Elsevier Information Processing & Management,Pages 436449,Volume 53, Issue 2, March 2017.
[75] Yan Dang, Yulei Zhang, and Hsinchun Chen, A Lexicon-Enhanced Method For Sentiment Classification: An Experiment On Online Product Reviews, Ieee Intelligent System, Pp. 46-53, 2010.
[76] Yang,Hui,LuoSi, and Jamie Callan, Knowledge transfer and opinion detection in the TREC blog track,2006.
[77] You Li, et al., Constructing Domain-Dependent Sentiment Lexicons Automatically for Sentiment Analysis,Information Technology Journal, Asian Network for Scientific information,2013,990-996
[78] T. Zagibalov and J. Carroll, Unsupervised Classification Of Sentiment And Objectivity In Chinese Text, Proc. Third Intl Joint Conf. Natural Language Processing, Pp. 304-311, 2008.
[79] Zhao,Wayene Xin,Jing Jiang, Hongfei Yan, and Xiaoming Li, Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid,Conference on emprical Methods in Natural Language Processing,2010.

Author Profile

Madhusudhanan S is currently pursuing Doctoral research at Anna University, Chennai, India. He received M.Tech degreee in Computer Science and Engineering from SASTRA University,Tanjore,India. His research interests include Machine Learning,Text Mining, Opinion Mining.

Dr.M.Moorthi has completed his BE degree at Arulmigu Meenakshi Amman College of Engineering, Kanchipuram, in Electronics and Communication Engineering in the year 2001 and M.E in Medical Electronics in the year 2007 at Anna University,Guindy Campus,Chennai,India.He received his PhD degree at Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University, Kanchipuram.He has 13years of teaching experience and he is currently working as a Professor in the Department of Electronics and Communication Engineering at Prathyusha Engineering College, Chennai. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), Indian Society for Technical Education (ISTE),IETE.He has been the reviewer for 2012 and 2013 IET image processing.His research interests are image segmentation, image compression, neural network, fuzzy logic, Artificial Intelligence.