IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.6, Special Issue.11, Dec 2018 E-ISSN: 2347-2693

A Survey on Sentimental Analysis Techniques

K.Lino FathimaChinnaRani[1*], M. Kasthuri [2], N.Solomon PraveenKumar [3]

[1,2,3]Dept. of Computer Science, Bishop Heber College, Tiruchirappalli, Tamil Nadu, India.

[*]*Corresponding Author: lino.judith@gmail.com*

Available online at: www.ijcseonline.org

***Abstract*—** Sentiment analysis is one of the fastest developing research areas in computer science, which is helpful to analyze people's opinions, sentiments, evaluations, attitudes and emotions from written language. It is widely studied in data mining, web mining, and text mining. This survey paper presents a comprehensive study on various recently used sentiment analysis techniques. The main target of this survey paper is to give full image of sentimental analysis techniques and the related field with brief details. The cluster of datasets given as an input and the accuracy level is checked by using discourse relations. The limitations and features are also discussed.

***Keywords*—** Sentiment analysis, Discourse relations, Baseline algorithms, Text mining.

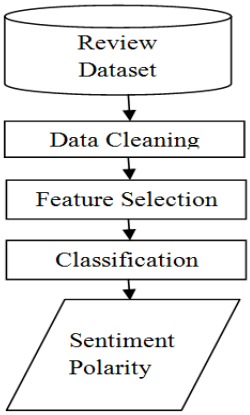
I. INTRODUCTION

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's beliefs, attitudes and sensation toward an entity. The entity can signify persons, events or topics. These subjects are most likely to be enclosed by reviews. The two expressions Sentiment Analysis or Opinion Mining are exchangeable. They prompt a mutual meaning. However, certain researchers specified that Opinion Mining and Sentiment Analysis have slightly diverse concepts. Opinion Mining extracts and examines people's view about an entity while Sentiment Analysis identifies the sentiment conveyed in a text then evaluates it. Therefore, the goal of Sentiment Analysis is to find opinions, recognize the sentiments they express, and then categorize their polarity.

Linguistic treating is associated to area of human computer collaboration. The duty of identifying opinion of review is called sentiment analysis. The growing importance of sentiment analysis coincides with the growth of social media such as reviews forum discussions, blogs, micro-blogs, Twitter, and social networks [1].

Opinion may be positive, negative or neutral polarity. There are three different level of Sentiment analysis they are document level, sentence level and entity-aspect level. Complete opinion is to be considered in document level whether opinion of specific sentence is to be measured in sentence level. Focus is directly on opinion itself is called entity and aspect level.

The following diagram shows the phrases of sentiment analysis process flow [17].



Fi.g 1.1

1.1 Review Dataset

The Review Dataset can be constructed by the customers. Customers post their review in comment, forum or blog.

1.2Data cleaning

Pre-processing the data is the process of cleaning and preparing the text for classification.The whole process involves several steps: online text cleaning, white space removal, expanding abbreviation, stemming, stop words removal, negation handling and finally feature selection. In this step, we identify the sentiment phrase.

Sentiment identification is also known as sentiment Detection, Recognition and Aspect identification [14]. Sentiment detection is a subfield of natural language preprocessing and a significant process of text analysis which the primary target identifying the presence of emotions

within the specified text.

The ability to distinguish what type of thing included in that sentence called the sentiment recognition. Finally the sentimental aspect identification considered in details of the object expresses opinions in the text.

1.3 Feature selection

Feature selection is the process of selecting a subset of relevant features for used in sentiment classification. It involves in two processes.

i) Identify the parts of the document that will contribute to positive or negative sentiments.

ii) Combine these parts of the document in ways that increase the odds the document.

Using feature selection helps to discard redundant or irrelevant data. There are various current features existing in feature selection such as Terms presence and frequency [24], Parts of speech [22], Negations[10], Feature weighting [6], etc.

1.4 Classification

The definite task of Sentiment classification is a text classification where objective is to organize a text according to the sentimental polarities of opinions if contains. Example: favorable or unfavorable, positive or negative. This mission has received substantial interests in the computational linguistic community owed to its potential application.

1.5 Sentiment Polarity

Sentiment polarity (positive vs. negative) of a text relies on sentiment lexicons, that is, large collections of words, each annotated with its own positive or negative orientation (i.e., prior polarity). The overall sentiment of a text is therefore computed upon the prior polarity of the contained words.

II.SENTIMENT ANALYSIS TECHNIQUES

Large datasets are available on-line today, they can be numerical or text file and they can be structured, semi-structured or non-structured. The following Fig 2.1 represents sentiment analysis techniques.

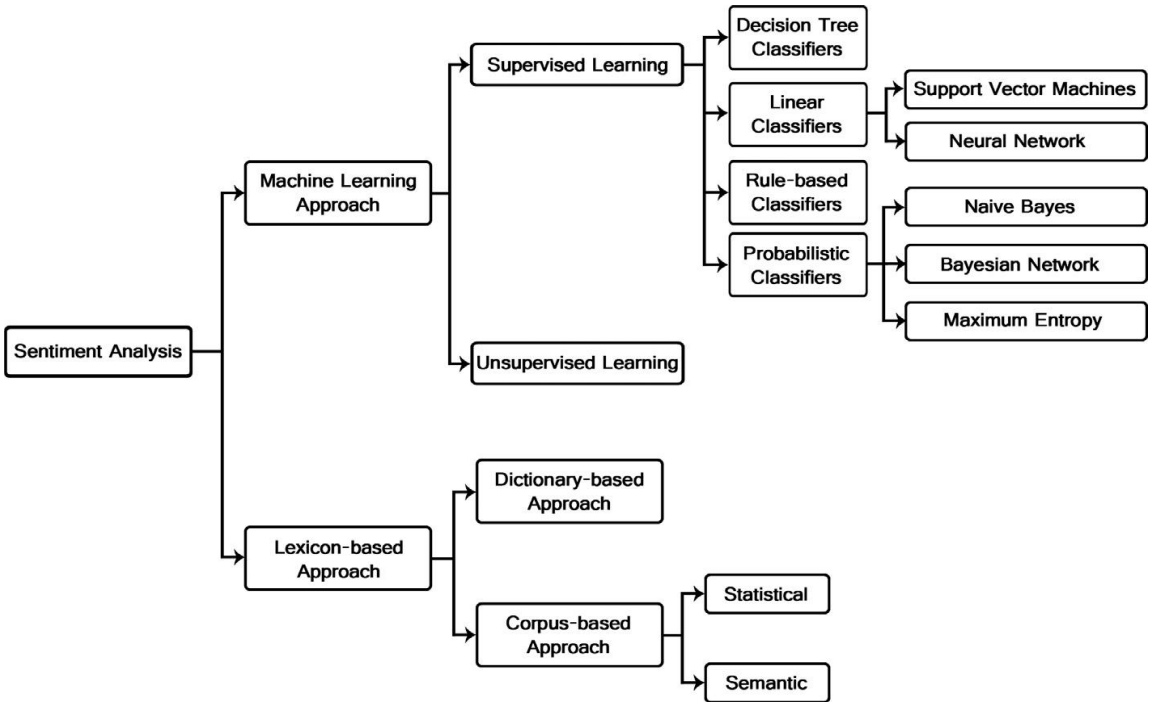


Fig. 2.1 Sentiment Analysis Techniques

Sentiment classification is tracking the mood of public about particular product or event or topic. Many different information retrieval techniques and tools have been proposed according to different data types. Sentiment classification, also known as opinion mining, is to identify and extract subjective information in source materials, which can be positive or negative. Using appropriate mechanisms and techniques, this large amount of data can be processed into information to support decision making. Sentimental Analysis Techniques mainly classified in two types:

2.1 Machine Learning approach

2.2 Lexicon based Approach

2.1. MACHINE LEARNING APPROACH

Machine Learning (ML) approach based on the well-known ML algorithms to resolve the SA as a regular text classification problem that makes use of syntactic and/or linguistic properties [22]. There are two steps in Sentiment classification of machine learning. First one is to abstract feature and store in feature vector and second one is to train feature vector by using classification algorithms.

Machine learning is further classified into two categories.

2.1.1 Supervised Learning

2.1.2 Unsupervised learning [28].

2.1.1 Supervised learning

The supervised learning methods depend on the existence of categorized training documents [29].Supervised machine learning techniques are used for classified document or sentences into finite set of class i.e., into positive, negative and objective. Training data set is available for all kind of classes. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize the training data.

Supervised classification algorithms are,

2.1.1.i) Decision Tree

2.1.1.ii) Linear classifiers

2.1.1.iii) Rule based classifiers

2.1.1.iv) Probabilistic classifiers

Supervised learning technique is relies on labeled dataset which is delivered as input to train the model and this model is applied to test data to produce output.

2.1.1.i) Decision tree classifiers

Decision tree classifier provides a hierarchical decomposition of the preparation data space in which a condition on the attribute value is used to divide the data. The circumstance or predicate is the presence or absence of one or more words. The separation of the data space is done recursively until the leaf nodes contain specific minimum numbers of records which are used for the purpose of categorization [26].

2.1.1.ii) Linear classifiers

Linear classifiers: Given ={X1 . . . . Xn} the normalized document word frequency, vector ={a1 . . . . an} is a vector of linear coefficients with the same dimensionality as the feature space, and b is a scalar; the output of the linear predictor is defined as p= + b, which is the output of the linear classifier. The predictor p is a separating hyper plane between different classes. There are many kinds of linear classifiers; among them is Support Vector Machines (SVM) which is a form of classifiers that attempt to determine good linear separators between different classes. Two of the most famous linear classifiers are discussed in the following subsections.

2.1.1.ii.a) Support vector Machine

2.1.1.ii.b) Neural Network

2.1.1.ii.a) SupportVector Machines Classifiers(SVM)

The main principle of SVMs is to decide linear separators in the search space which can best divide the different classes. There are 2 classes x, o and there are 3 hyper planes A,B and C. Hyper plane A provides the best division between the classes, because the normal expanse of any of the data points is the leading, so it represents the maximum margin of separation. This method is used to analyze the complete vectorized data and the key idea behind the training of model is to find a hyper plane [5].

2.1.1ii.b) Neural Network (NN)

Neural Network consists of several neurons where the neuron is its basic unit. The inputs to the neurons are denoted by the vector over line Xi which is the word frequencies in the document. There are set of weights, which are associated with each neuron used in order to compute a function of its inputs.

Multilayer neural networks are used for non-linear boundaries. These multiple layers are used to induce multiple piecewise linear boundaries, which are used to approximate enclosed regions belonging to a particular class. The outputs of the neurons in the earlier layers feed into the neurons in the later layers. The training method is more complex because the errors require being back-propagated over different layers.

2.1.1.iii) Rule-based classifiers

In rule based classifiers, the data space is modeled with a set of rules [30]. The left hand side represents a condition on the aspect set expressed in disjunctive normal form while the right hand side is the class label. The conditions are on the term occurrence. Term absence is rarely used because it is not useful in sparse data .There are numbers of criteria in order to generate rules, the training phase construct all the rules depending on these criteria.

The most two common criteria are support and confidence. The support is the absolute number of instances in the training data set which are relevant to the rule. The Confidence refers to the conditional possibility that the right hand side of the rule is satisfied if the left-hand side is satisfied.

2.1.1.iv) Probabilistic classifiers

Probabilistic classifiers use combination models for the classification. The mixture model assumes that each class is a component of the mixture. Each mixture component is a generative model that provides the probability of sampling a particular term for that component. These kinds of classifiers are also called generative classifiers. Three of the most

famous probabilistic classifiers are discussed in the next subsections.

2.1.1.iv.a) Naïve Bayes Classifier
2.1.1.iv.b) Bayesian Network
2.1.1.iv.c) Maximum Entropy

2.1.1.iv.a) Naive Bayes Classifier (NB)

The Naive Bayes classifier is the simplest and most commonly used classifier. It is used to predict the probability for a given tuple to belong to a particular class. It is used because of its easiness in both during training and classifying steps[8].

Naıve Bayes classification model computes the posterior probability of a class, based on the distribution of the words in the document. The model works with the BOWs feature extraction which ignores 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.Phrase-level opinion mining is also well-known as aspect based opinion mining. It is used to extract most important aspects of an item and to predict the orientation of each aspect from the item reviews[7].

2.1.1.iv b)Bayesian Network (BN)

The main assumption of the NB classifier is the independence of the features. The other extreme assumption is to assume that all the features are fully dependent. This leads to the Bayesian Network model which is a directed acyclic graph whose nodes represent random variables, and edges represent conditional dependencies. BN is considered a complete model for the variables and their relationships. In Text mining, the computation complexity of BN is very expensive.

2.1.1.iv.c) Maximum Entropy Classifier (ME)

It is one of the classifiers that is commonly used to power up our Machine Learning Application Programming Interface.The Max Entropy classifier is a probabilistic classifier which belongs to the class of exponential models.The Max Entropy does not assume that the features are conditionally independent of each other. The MaxEnt is based on the Principle of Maximum Entropy and from all the models that fit our training data, selects the one which has the largest entropy. The Max Entropy classifier can be used to solve a large variety of text classification problems such as language detection, topic classification, sentiment analysis and more.

2.2 LEXICON BASED APPROACH

Opinion words are engaged in many sentiment classification tasks. Positive opinion words are used to convey some desired states, while negative opinion words are used to convey some undesired states. There are some opinion phrases and idioms which together are called opinion lexicon. There are three main approaches in order to compile or collect the opinion word list [9].

Manual approach is very time consuming and it is not used alone. It is usually combined with the other two automated approaches as a final check to keep away from the mistakes that resulted from computerized methods. The two automated approaches are presented in the following subsections.

2.2.1 Dictionary based Approach
2.2.2 Corpus based Approach

2.2.1 Dictionary based Approach

The main scheme of the dictionary-based approach is a small set of opinion words is collected manually with known orientations. Then, this set is grown by searching in the well-known corpora WordNet or thesaurus for their synonyms and antonyms[4]. The newly found words are added to the seed list then the next iteration starts. The iterative process stops when no new words are found. After the process is completed, manual check can be carried out to remove or correct errors.

2.2.2 Corpus Based Approach

The Corpus-based approach helps to solve the problem of finding opinion words with situation precise orientations. Its methods depend on syntactic patterns or patterns that occur together along with a seed list of opinion words to find other opinion words in a huge corpus [25]. They started with a list of seed opinion adjectives, and used them along with a set of linguistic constraints to identify additional adjective opinion words and their orientations. The constraints are for connectives like AND, OR, BUT, EITHER-OR.

There are also adversative expressions such as but, however which are indicated as opinion changes. In order to decide if two conjoined adjectives are of the same or different orientations, knowledge is applied to a large corpus [25]. Then, the links between adjectives form a graph and clustering is performed on the graph to create two sets of words: positive and negative.

The corpus-based approach is performed using statistical approach or semantic approach as illustrated in the following subsections:

2.2.2.a) Statistical Approach
2.2.2.b) Semantic Approach
2.2.2.c) Discourse Information

2.2.2.a). Statistical Approach

Finding co-occurrence patterns or seed opinion words can be done using statistical techniques. This could be done by deriving posterior polarities using the co-occurrence of adjectives in a corpus. It is possible to use the whole set of indexed documents on the web as the corpus for

the dictionary construction.

This situation overcomes the problem of the unavailability of some words if the used corpus is not huge enough .The polarity of a word can be recognized by studying the occurrence frequency of the word in a huge annotated corpus of texts . If the word occurs more often among positive texts, then its polarity is positive. If it occurs more regularly among negative texts, then its polarity is negative. If it has equivalent frequencies, then it is a neutral word.

The similar opinion words often appear together in a corpus. This is the main surveillance that the state of the art methods are based on. Therefore, if two words appear together frequently within the same context, they are likely to have the similar polarity. Therefore, the polarity of an unidentified word can be resolute by calculating the relative regularity of co-occurrence with another word.

2.2.2.b) Semantic approach

The semantic approach gives sentiment values directly and relies on different values for computing the similarity between words. This principle gives parallel sentiment values to semantically close words. Word Net for example provides different kinds of semantic relationships among words used to calculate sentiment polarities. Word Net could be used too for obtaining a list of sentiment words by iteratively expanding the preliminary set with synonyms and antonyms and then determining the sentiment polarity for an unknown word by the comparative count of positive and negative synonyms of this word.

The semantic advance is used in many applications to build a lexicon model for the description of verbs, nouns and adjectives to be used in SA The model described the detailed subjectivity relations between the actors in a sentence expressing separate attitudes for each actor. These subjectivity relations are labeled with information relating to both the identity of the attitude holder and the orientation (positive vs. negative) of the attitude.

2.2.2.c) Discourse information

The importance of conversation in SA has been rising recently. Discourse information can be found either among sentences or among clauses in the same sentence [18]. Five types of rhetorical relations are used. They are Contrast, Correction, Support, Result, and Continuation with attached sentiment information for annotation [19]. The proposed concepts called opinion frame, the components of opinion frames are opinions and are the relationships between their targets. They have improved their work and investigated design choices in modeling a discourse scheme for improving sentiment classification.

Rhetorical Structure Theory (RST) describes how to crack a text into spans, each indicating a meaningful part of the text [21]. The proposed a framework that performed document SA (partly) based on a document's discourse structure which was obtained by applying RST on sentence level. They hypothesized that they can improve the presentation of a sentiment classifier by splitting a text into important and less important text spans. The lexicon based used for classification of movie reviews. Their results showed improvement in SC accuracy compared to a baseline that does not take discourse structure into account [20].

III. REVIEW OF LITERATURE

This section is helpful to show literature review about recent sentiment analysis techniques. Table 3.1 shows recent sentiment analysis techniques applied in various domains, adapted techniques, data set used, reached accuracy, advantages and Limitations.

TABLE 3.1 SENTIMENT ANALYSIS TECHNIQUES, ADVANTAGES AND LIMITATIONS

Year	Author	Title	Technique	Dataset	Accuracy	Advantages	Limitations
2002	Pang, B., Lee, L. and Vaithyanathan, S.	Sentiment classification using machine learning techniques.	SVM MaxEnt NB	Movie Review	82.9 % 81 % 81.5%	i) SVM gives the best result.	i) Sentiment categorization is difficult ii) Naive Bayes classifiers results must be improved
2009	Go, A., Bhayani, R. and Huang, L.	Twitter sentiment classification using distant supervision.	NB MaxEnt SVM	Twitter	82.7% 83% 82.2%	i) Machine learning approach gives the best accuracy level.	i) Unigram technique does not improve the performance.
2013	Shahana, P.H. and Omman	Evaluation of Features on Sentimental Analysis.	NB	Customer Review	92.37%	i) The unigram performance is better as compare to the bigram.	i) In bigram and other techniques, the accuracy is not reached up to the level.
2013	Mouthami, K., Devi, K.N. and Bhaskaran, V.M	Sentiment analysis and classification based on textual reviews. In Information	Fuzzy classifier	Movie Review	good	i) Fuzzy classifiers give high score on interpretability. ii) Faster deployment in	i) The accuracy can further be improved by applying specific strategies to generate

		Communication and Embedded Systems				a large number of attributes.	the Fuzzy space for each monitored parameter.
2014	Gautam, G. and Yadav, D.,	Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In Contemporary Computing	WordNet SVM MaxEnt NB	Twitter	89.9% 85.5% 83.8% 88.2%	i)Machine learning approaches improves the accuracy level by applying SVM and NB techniques	i) Lexicon based approaches are not implemented.
2014	Khan, F.H., Qamar, U. and Javed, M.Y.,	A visual sentiment analysis framework. In Information Society	Unsupervised	Twitter	80% up	i)Three different classification techniques, imoticon analysis, BOW and senti WordNet are used to increases the precision, f-measure and accuracy.	i)The base line techniques are not implemented
2014	da Silva, N.F., Hruschka, E.R. and Hruschka, E.R.,	Tweet sentiment analysis with classifier ensembles. Decision Support Systems	Ensemble Classifier (RF, NB, LR, SVM)	Sanders Twitter Stanford Twitter OMD Twitter HC Twitter	84.89% 87.20% 76.81% 78.35%	i) Feature hashing are discussed for the better representation of tweets. ii) The classifier ensembles can improve classification accuracy.	i)Unigram, bigram techniques are not implemented.
2015	Jeyapriya, A. and Selvi, K.	Extracting aspects and mining opinions in product reviews using supervised learning algorithm. In Electronics and Communication Systems.	SVM	Movie Review	83%	i)SVM increases the execution speed of the model and also the result accuracy.	i) It fails to select relevant features in cases where there are strong feature interaction.
2015	Tripathy, A., Agrawal, A. and Rath, S.K	Classification of Sentimental Reviews Using Machine Learning Techniques	SVM NB	Movie Review	94% 89.5%	i) Confusion matrix is constructed based on the accuracy. ii) Algorithms are applied on polarity data set and gives best result.	i) For unigram, bigram, trigram, fourgram, fivegram classification techniques decreases the accuracy.
2015	Bhadane, C., Dalal, H. and Doshi, H	Sentiment analysis: Measuring opinions. Procedia Computer Science	SVM	Customer Review	78%	i) Removing stop words step in pre-processing leads to higher accuracy.	i) Accuracy must be tested using linear kernel.
2016	Hemant M.Banker, Deepak Gupta	Rule based Approach for sentiment Analysis of Traffic Data	Rule based approach	Traffic data set	100%	i) Result will be increased by applying rules on the data. So it gives more accurate results.	i) Rules applied as per location and updating user result and by using map provide new way for travelling.
2017	Karthik Sripathi	Oscars 207 – Text mining & Sentimental Analysis	SAS technique	Twitter	83%	i) It gives an idea to identify the people reaction towards unwanted events during the stage show.	i)This analysis can be extended to build a text predictive model where is a scope of predicting the sentiments towards unwanted events
2018	Penubaka Balaji D.Haritha O.Nagaraju	An overview on opinion mining techniques & Sentiment Analysis	LDA, topic model ling, extrinsic &intrinsic feature selection Algorithms	Customer review	93%	i)The base line algorithms and proposed Algorithm are discussed in order to improve the performance	i)The new classification techniques yet to be implemented for higher accuracy.

IV.CONCLUSION

This survey paper presents an overview on the recent techniques in sentiment analysis, adopted techniques; data set used advantages and limitations. The recently published articles were categorized and summarized. These articles provideuseful contributions to various fields which adopts Sentiment Analysis techniques for different real-world applications. From the literature review it is observed that existing sentiment analysis techniques have its own

advantages and limitation. In order to overcome the limitations available in the existing sentiment analysis techniques, if anyone will propose new sentiment analysis technique with hybridize some other techniques like Stemming, Data structures, and then it is very useful to different domains.

From the literature review, it is clear that the enhancements Sentiment Classification and Full Search algorithms are still an open field for research. The interest in languages other than English in this field is growing as there is still a lack of resources and researches concerning these languages. Information from micro-blogs, blogs and forums as well as news source, is widely used in sentiment analysis recently. This media information plays a great role in expressing people's feelings, or opinions about a certain topic or product. Using social network sites and micro-blogging sites as a source of data still needs deeper analysis.

REFERENCES

[1] Liu. B,"Sentiment analysis and opinion mining", *Synthesis Lectures On Human Language Technologies*,vol.5,no.1, pp.1-167,2012

[2] Pang B. Lee, L. and Vaithyanathan, S., ,"Sentiment classification using machine learning techniques", *Proceedings Of The Acl-02 Conference On Empirical Methods In Natural Language Processing*,vol.10,pp.79-86,2002.

[3] Chaovalit P. and Zhou L.,"Movie review mining:A Comparison Between Supervised And Unsupervised Classification Approaches",*Proceedings of the 38th Hawaii International Conference on System Sciences*,pp.112c,2005.

[4] Miller G.A,"WordNet a lexical database for english", *Communications Of The Acm*,vol.38,no.11, pp.39-41,1995.

[5] Tripathy A, Agrawal A and Rath S.K,"Classification of Sentimental Reviews Using Machine Learning Techniques", *Procedia Computer Science*,vol.57,no., pp.821-829,2015.

[6] Shahana P.H. and Omman B,"Evaluation of Features on Sentimental Analysis", *Procedia Computer Science*,vol.46, pp.1585-1592,2015.

[7] Jeyapriya A. and Selvi K,"Extracting aspects and mining opinions in product reviews using supervised learning algorithm", *Electronics and Communication Systems*,pp.548-552,2015.

[8] Kanakaraj M. and Guddeti R.M.R,"Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques", *Semantic Computing*, pp.169-170,2015.

[9] Mouthami K, Devi K.N. and Bhaskaran V.M,"Sentiment analysis and classification based on textual reviews", *Information Communication and Embedded Systems*,pp.271-276,2013.

[10] Bhadane C., Dalal H. and Doshi H,"Sentiment analysis: Measuring opinions", *Procedia Computer Science*,vol.45,pp.808-814,2015.

[11] Gautam G. and Yadav D,"Sentiment analysis of twitter data using machine learning approaches and semantic analysis", *Contemporary Computing*, pp.437-442, 2014.

[12] Zhou X, Tao X., Yong J and Yang Z,"Sentiment analysis on tweets for social events", *Computer Supported Cooperative Work In Design*, pp.557-562,2013.

[13] Go A, Bhayani R. and Huang L,"Twitter sentiment classification using distant supervision", *Cs224N Project Report*,pp.12,2009.

[14] Khan F.H., Qamar U. and Javed M.Y,"Sentiview: A visual sentiment analysis framework", *Information Society*,pp.291-296,2014.

[15] da Silva N.F., Hruschka E.R. and Hruschka E.R,"Tweet sentiment analysis with classifier ensembles", *Decision Support Systems*,vol.66,pp.170-179,2014.

[16] Michael W. Berry,"Soft Computing in Data Science", *Communications in Computer and Information Science*, pp.2-3, 2015.

[17] Patel S.N. and Choksi M.J.B,"A Survey of Sentiment Classification Techniques", *Journal for Research*,vol.1,pp.1,2015.

[18] Heerschop B., Goossen F., Hogenboom A., Frasincar F., Kaymak U. and de Jong F,"Polarity analysis of texts using discourse structure", *Proceedings of Cimk*,pp.1061-1070,2011.

[19] Hernault H., Bollegala D. and Ishizuka M,"A semi-supervised approach to improve classification of infrequent discourse relations using feature vector extension", *Proceedings of The Emnlp*,pp.399-409,2010.

[20] Joty S Carenini G. and Ng R.T,"A novel discriminative framework for sentence-level discourse analysis", *Proceedings of Emnlp*,vol.,no., pp.904-915,2012.

[21] Bill Mann"An Introduction to Rhetorical Structure Theory (RST),vol.8,no.3, pp.243-281,1999.

[22] Narayanan R., Liu, B. and Choudhary A,"Sentiment analysis of conditional sentences", *Proceedings Of Emnlp*,pp.180-189,2009.

[23] Pang B. and Lee L,"Opinion mining and sentiment analysis", *Information Retrieval*,vol.2,no.1, pp.1-135,2008.

[24] Polanyi L. and Zaenen A,"Contextual valence shifters", *Computing Attitude and Affect*,, pp.,2006

[25] Qiu, L. Zhang, W. Hu C. and Zhao K,"a self-supervised model for sentiment classification", *Proceedings Of Cikm*,, pp.929-936, 2009.

[26] Somasundaran S., Namata J. and Getoor L,"Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification", *Proceedings Of The Enmlp*,pp.170-179,2009.

[27] M. Kasthuri, Dr. S. Britto Ramesh Kumar, "A Framework for Language Independent Stemmer Using Dynamic Programming", In: *International Journal of Applied Engineering Research* (IJAER), Print ISSN 0973-4562, Volume.10, Number.18, pp.39000-39004 Online ISSN1087-1090, 2015.

[28] M. Kasthuri, Dr. S. Britto Ramesh Kumar, "PLIS: Proposed Language Independent Stemmer for Information Retrieval Systems Using Dynamic Programming", *2016 World Congress on Computing and Communication Technologies*, ISBN: 978-1-5090-5573-9, pp.132-135, IEEE, 2016.

[29] M. Kasthuri and Dr. S. Britto Ramesh Kumar, "Multilingual Phonetic Based Stem Generation", *Proceedings of the Second International Conference on Emerging Research in Computing, Information Communication and Applications* (ERCICA-2014), ELSEVIER Science :: Technology India, ISBN:9789351072 607, Volume.1, pp.437-442, 01-02 August, 2014.

[30] Dr.M.Kasthuri and Dr.S.Britto Ramesh Kumar, "PLIS: Proposed Language Independent Stemmer Performance Evaluation", In: *International Journal of Advanced Research in Computer Science & Technology* (IJARCST 2017), Volume.5, Issue.4, ISSN: 2347-8446 (Online), ISSN:2347-9817 (Print), Oct-Dec, 2017.

Authors Profile

Prof. Dr. M.Kasthuri is working as an Assistant Professor in the Department of ComputerApplications, Bishop Heber College,Tiruchirappalli, Tamil Nadu, India. Shehad completed her Doctorate of Philosophyin Computer Science in June 2017 atBharathidasanUniversity, Tiruchirappalli.She has published a number of National and International level research papersrelated to Web Mining and Stemming concepts. She has completed UGC sponsored Minor ResearchProject entitled as Language Independent Stemmer.

Prof. K.Lino Fathima Chinna Rani is working as an Assistant Professor in the Department of Computer Applications,Bishop Heber College,Trichirappalli, Tamil Nadu,India. She had completed her Master of Information Technology in June 2008 at Bharathidasan University,Tiruchirappalli. She had Completed her Master of Philosophy in Computer Science in September 2009 at Bhararthidasan University,Tiruchirappalli.

Prof. N.Solomon Praveen Kumar is working as an assistant professor in the department of Computer Applications, Bishop Heber College,Tiruchirappalli,Tamil Nadu, India. He had completed his Master of Computer Applications in June 2009 at Anna Univrsity,Chennai.He had Completed his Master of Philosophy in Computer Science in septmber 2012 at Bhararthidasan University, Tiruchirappalli. He has eight years of teaching experience in computer Applications field since January 2010 to till date.

