Sentiment Analysis Approaches on Different Data set Domain: Survey

Shailendra Kumar Singh[1], Sanchita Paul[2] and Dhananjay Kumar[1]

[1] *PG- Student,* [2]*Assistant Professor*
Department Of CSE
Birla Institute Technology, Mesra, Ranchi, India
sks.it2012@gmail.com; *sanchita07@gmail.com; dhananjay.sharma48@gmail.com*

Abstract

The growth of social website and electronic media contributes vast amount of user generated content such as customer reviews, comments and opinions. Sentiment Analysis term is referred to the extraction of others (speaker or writer) opinion in given source material (text) by using NLP, Linguistic Computation and Text mining. Sentiment classification of product and service reviews and comments has emerged as the most useful application in the area of sentiment analysis.

This paper focuses on the comparative study (1997 – 2012) of different sentiment classification techniques performed on different data set domain such as web discourse, reviews and news articles etc. The most popular approaches are Bag of words and feature extraction used by researchers to deal with sentiment analysis of opinion related to movies, electronics, cars, music etc. The sentiment analysis is used by manufacturers, politicians, news groups, and some organization to know the opinions of customer, people, and social website users.

Keywords: Opinion, Sentiment Analysis, Sentiment Classification, Sentiment Classification Techniques, Product Reviews, Social Issues

1. Introduction

In this scientific era, internet provides huge volume of information. Most of the people share their opinions over internet by using social networking sites in form of textual data. These textual data are publically available over internet & has a great impact in building opinions about a particular entity, object or political activities among the users of social media. Shared information is generally in form of reviews, articles, posts, news etc. Today most of the people preferred electronic media and internet over printed media. Especially in electronics and film sectors, customers used to write their reviews about products or their features. Social media user may find others opinion by collecting & analyzing their reviews about different features of the product. In this way one can find the best products, one that meets their needs by comparing the one products' features, price, size and quality etc. with other products. Moreover companies can find out the strengths, weakness and limitation of their products or those of competitors [1]. In this way, manufacturers solved the reported problems, valuable for product development, consumer relationship management and marketing and use the business intelligence behind the analysis for future investments. Traditionally, manufacturers conduct consumer feedback and surveys in manually for this purpose. While that was well-designed surveys which where provides quality estimations, but

it was costly especially if the size of gathered survey data was in large volume [2].The analyzing and summarizing of others opinions, expressed in huge opinionated text data is a very interesting new field for researchers. This area of research is called Sentiment Analysis or Opinion Mining of others [3].

From sentiment and opinion mining perspective, usually there are two kinds of textual information facts and opinions. Facts refer to the objective statements about the nature of a product, while opinion describes attitude, appraisals and emotions extraction of a product, service, topic or an issue [4].The Sentiment Analysis is also called Interdisciplinary field that crosses artificial intelligence, natural language processing and text mining. Sentiment analysis has emerged as a Subfield of Text Mining because most of opinions are available in the text format and its processing is easier than other formats [5]. Sentiment analysis first time appeared in the literature in 1990 and then it became a major research topic in 2000. The extensive research on automatic text analysis for sentiment, such as sentiment classifiers [6 , 7, 8, 9,10], affect analysis [11, 12] , automatic survey analysis [13, 8],opinion extraction [14], recommender systems [15], subjectivity detection, sentiment prediction, text analysis for opinions, extracting product features, extraction of customers opinion.

As per the technical perspective, there are two main approaches for SA, such as Bag of words (BOW) and Feature based Sentiment (FBS) [16]. In the BOW approach, the syntactic & semantic information between words are last. This approach is not useful in opinions mining of products & their features. While FBS approach is used for analyzing the sentiment of products & their features. The main aim of this paper is to show, which sentiment classification technique used on what type of data set. Basically machine learning techniques were used for FBS and others for BOW. But all these methods classify the sentiment polarity of given documents either as positive or negative sentiment.

2. Sentiment Classification

The Sentiment classification is presented in various formats in different domains. Positive/negative, good/bad, like/dislike, buy/don't buy, recommended/not recommended, excellent/boring (film), support/against [5], optimistic/pessimistic [9], favorable/unfavorable [17]. Sentiment classification may be done at different levels. In Document level – whole document classify either into positive or negative class. Sentence level – classifies sentence into positive, negative or neutral class. Aspect or feature level – identifying & extracting product features from the source data.

2.1. Sentiment Classification Techniques

The sentiment classification techniques are categorized into three categories. They are machine learning algorithms, link analysis methods, and score- based approaches. Machine learning algorithms are applicable to sentiment analysis mostly belongs to supervised classification. In machine learning techniques, there are two sets of data: training set and a test set. Training set data/documents are used by automatic classifier to learn the differentiating characteristics of documents, while a test set is used to check how well the classifier performs [3].A number of machine learning approaches are used to classify the reviews. These techniques are Support Vector Machines (SVM), Naïve Bayes (NB), and Maximum Entropy (ME). Machine learning approaches starts from collecting training dataset, then to train a classifier on the training data. Once a supervised classification technique is selected, then an important step: decision to make is feature selection. Then only,

supervised classifier tells us how documents are represented [3]. There are some common features that are used in sentiment classification. They are given below.

- **Terms and their frequency:**

These include unigram or n-gram and their frequency or presence. Pang et al. (2000) [6] claim that unigrams gives better results than the bi-grams in movie review sentiment analysis. While Dave et al. (2003) [21] claim that tri-grams and bi-grams give better polarity classification results for product- review.

- **POS information :**

In POS tagging, each term in sentence will be assigned a label, which used to represent its position/role in the grammatical context. So, with POS tags, we can identify adjective and adverbs which are used as sentiment indicators [19].

- **Negations:**

Negation has potential of reversing a sentiment [46] so, these feature must take into account.

- **Opinion words and phrases:**

The opinion word and opinion phrase are used to extract positive / negative sentiments. There are   approaches such as lexicon-based or statistical-based, are used to determine the semantic orientation of an opinion words and phrases. While Hu and Lui et al. (2004) [29] used Word Net to determine the sentiment polarity of extracted adjective as positive or negative polarity.

The most commonly machine learning algorithm used are Support Vector Machine (SVM) and Naïve Bayes (NB). Support Vector Machine has been commonly used for sentiment analysis of movie reviews [6, 32, 39], while Naïve Bayes used to reviews and web discourse [6, 32, 25]. SVM have performed better than NB [6]. Other algorithms are also used in prior sentiment classification studies such as Winnow [34] and AdaBoost [40].

Link analysis technique for sentiment classification of link-based metrics and features. For web-site opinions Efron (2004) [25] used co-citation analysis, But Agarwal, R. *et al.*, (2003) [20] used message-send/reply link feature to classify sentiments in USENET newsgroups.

Score-based methods are generally used to classify message sentiment based on the total sum of all positive or negative sentiment polarity. Nasukawa and Yi (2003) [17], Yie *et al.*, (2003) [2], Fei *at al.*, (2004) [26] used phrase pattern matching, which requires checking text for manually created, polarized phrase tags (positive or negative). +1 value assigned for Positive phrase and -1 for Negative phrase. All messages are classified on the basis of Positive sum and negative sum value, positive sum value means positive sentiment and negative means negative sentiment.

2.2. Sentiment Analysis Features

In the previous sentiment analysis studies, the sentiment Analysis features are classified into four kinds such as syntactic, link-based, semantic, and stylistic features. The syntactic attributes are used with semantic features as a set of features for sentiment extraction. Syntactic features include part of speech (POS) tags [6, 2, 27], POS tags n-grams [6, 27] and punctuation. Also syntactic features include phrase patterns, which were created by using POS tag n-gram patterns [21, 2, 26]. The research found that phrase patterns e.g. "n+aj"

(noun followed by (+) adjective), so this phrase patterns represent (+) sentiment, but "n+dj" (noun followed by (-) adjective), represent (-) sentiment [26].

In Semantic features, polarity addition or strength -related scores to words and phrases are assigned by using fully automatic annotation or manual/semi-automatic techniques. In 1997 Hatzivassiloglou et al. [18] proposed a semantic orientation (SO) method. Later extended by Turney, (2002) [19], that automatically computed the SO score for each word/phrase by using a mutual information calculation technique. In score computation, we take the mutual information between a phrase and the word like word "excellent" and then subtracting the mutual information (polarity) between the same phrase and the word like "poor". Later Semantic Orientation approach was analyzed by using latent semantic analysis [51].

The sentiment lexicons are generated by manually or semi-automatically [10, 26, and 40] Riloff *et al.*, [22] constructs sets of objective nouns, weak subjectivity and strong subjectivity. Appraisal groups [39] are another effective method for annotating semantics to words/phrases. The lexicons are developed, while based on appraisal theory [52]. Appraisal group include polarity of phrase, graduation, orientation and attitude. In 2005 Whitelaw, C. *et al.*, [39] was able to find more accuracy on a movie reviews corpus by using appraisal groups, several outperforming previous studies [33], the use of syntactic features [6] and the automated mutual-information based approach [19].

Link-based features include link/citation analysis to extract sentiments of online articles and web documents [46]. The Opinion web pages mostly share similar sentiment to each other [25]. Since the link-based features are very less used, so it is difficult to know that how effective they may be used for sentiment classification.

Stylistic features used structural and lexica attributes in different previous authorship/stylometric studies [53, 54, and 46]. Gamon (2004) [27] used sentence length as a lexical features for sentiment analysis of feedback surveys. It was unclear that how stylistic features were used as sentiment classifiers for movies/products features review, while style markers have been found that commonly used in particular web discourse [54, 46].

2.3. Sentiment Analysis Domains

Sentiment classification techniques applied on different data set types such as Reviews, Web Discourse and News Articles. The reviews include movie reviews, product features reviews and music reviews [14, 6, 19].Product features reviews are complex, because a review of a single person can have both positive and negative sentiment about a specific feature of the product. Sentiment extraction of movie reviews is very interesting, because movie reviewers present their opinion in large summaries and use complicated literary devices such as rhetoric and sarcasm.

Web discourse sentiment analysis includes social website, News groups, and comments (face book, twitter). This domain usually extracts sentiment of particular issues/topics like global warming, gun control and politics [20, 25]. Robinson (2005) [55] evaluated sentiments of social issues like World Trade Center Attack on 9/11 in three different organization in the United States, Brazil and France. Some authors have performed sentiment analysis on news articles [2, 40].

2.4. Reduction of Features for sentiment classification

There are different manual and automated approaches have been used to select essential attribute feature. Gamon (2004) [27] and Jeonghee Yi et al. (2003) [2] used log likelihood to

select important attributes from a large features space. He used this technique to improve accuracy and focused on selected subset of sentiment discriminators.

3. Sentiment Analysis Tasks

Most of the researchers focus on specific tasks: sentiment analysis of words [18], subjective expressions [40, 31], subjective sentences [32] and topic [2, 17, 36]. These approaches find sentiment at opinion/facts level and used to enhance the effectiveness of a sentiment classification Pang & Lee (2004) [32]. Choi et al. (2005) [42] focus on finding the sources of opinions (e.g. Finding who play a crucial role the person or organizations in influencing other individuals' opinion) instead of carrying out a sentiment classification. There was various types of Data sources where used that include The Multi-Perspective Question Answering (MPQA) corpus, user feedback, The Wall Street Journal (WSJ) corpus and the Document Understanding Conference (DUC) corpus.

The following parameters are used for the comparative study of different existing works in Sentiment Analysis. They are –

A. Sentiment Analysis Features

- **Syntactic (F1)** - Words/POS tag n-grams, Phrase patterns.
- **Semantic (F2)** – Positive and Negative tags, appraisal groups, semantic orientation.
- **Link based (F3)** – Web links, send/reply patterns, document citations.
- **Stylistic (F4)** –Lexical and structure measures of style.

B. Sentiment Classification Techniques

- **Machine Learning (T1)** – Support Vector Machine, Naïve Bayes, Maximum Entropy.
- **Link Analysis (T2)** - Citation analysis, message send/reply patterns.
- **Similarity score (T3)** – Lexicon, Sentiment pattern, sentiment score counts, etc.

C. Sentiment Analysis of Data Set Domains

- **Reviews (D1**) - product, movie and music reviews.
- **Web discourse (D2)** – Social Website (face book, twitter, YouTube).
- **News articles (D3)** – Online articles and web pages.

D. Accuracy, Precision and Recall

We can see the formula to compute the accuracy, precision and recall values [57].

Table 1. Confusion Table [57]

	Machine (Yes)	Machine (No)
Human (Yes)	tp	Fn
Human (No)	fp	tn

Precision (P)= $t_p/(t_p + f_p)$; **Recall (R)**= $t_p/(t_p + f_n)$; **Accuracy (A)**= $t_p + t_n / (t_p + t_n + f_p + f_n)$

E. Data Size and Data Source

Data size means the number of sentence/ expression/feedback/review on which techniques are applied for sentiment analysis.

Data source mean that from which place (e.g. website, movie review, web pages, journals, customer feedback) data sets have been taken.

Table:-2 Summery of Existing Works in Sentiment Analysis

Authors	Features				Reduce Features	Techniques			Domains			Accuracy	Precision	Recall	Data size	Data source
	F_1	F_2	F_3	F_4		T_1	T_2	T_3	D_1	D_2	D_3					
Hatzivassiloglou & Mckeown,1997 [18]	√				No			√		√		78.1-92.4	N/A	N/A	657 adj (+) 679 adj (-)	WSJ Corpus
P. Subasic et al., 2001[11]	√	√			No			√			√					
R.M.Tong, 2001[10]	√	√			No			√	√							
S. Morinaga et al, 2002[14]	√				Yes			√	√							
Pang ,Lee et al, 2002[6]	√				No	√			√			77-82.9	N/A	N/A	700 (+) 700 (-)	IMDB
Turney, 2002 [19]	√	√			No			√	√			65.8-84	N/A	N/A	240 (+) 170 (-)	Movie review
Agrawal ,R. et al, 2003[20]	√		√		No	√	√			√						
Dave , K., et al, 2003 [21]	√				No	√		√	√			88.9	N/A	N/A	N/A	Product review
Nasukawa & yi; 2003 [17]	√	√			No			√			√	94.3	N/A	28.6	118 (+) 58(-)	Web page
							√					94.5	N/A	24	255	Camera review
Riloff, K., et al, 2003[22]		√		√	No	√					√					
Jeonghee Yi et al, 2003 [2]	√	√			Yes			√	√			85.6	87	56	735(+) 4227(-)	Digital camera review
											√	90-93	86-91	N/A		Web pages
Yu & Hatzivassilogol	√	√			No	√		√			√					

Reference														
0u, 2004[23]														
Bieineke, P., et al, 2004[24]		√		No	√		√	√						
Efron, M., et al., 2004[25]	√		√	No	√	√			√						
Fei, Z., et al, 2004[26]		√		No		√	√								
Gamon, 2004[27]	√		√	Yes	√			√		77.5	N/A	N/A	N/A	Customer feed back	
Grefenstette et al, 2004[28]	√	√		No		√		√							
Hu & Liu, 2004 [29]	√	√		No		√	√								
Kanayama et al, 2004[30]	√	√		No		√	√								
Kim & Hovy, 2004 [31]		√		No	√			√		75.6-77.9	N/A	79.8	N/A	DUC Corpus	
Pang & lee, 2004 [32]	√	√		No	√	√	√			86.4-87.2	N/A	N/A	1000(+) 1000(-)	Movie review	
Mullen & collier,2004[33]	√	√		No	√			√							
Nigam & Hurst, 2004 [34]	√	√		No	√			√							
Wiebe et al, 2004[35]	√		√	Yes	√		√	√	√						
Hiroshi et al, 2004[36]	√	√		No		√	√			89-100	N/A	43	200	Camera review	
Liu et a.l , 2005[37]	√	√		No		√	√								
Mishne , 2005[38]	√	√	√	No	√			√							
Whitelaw et al, 2005[39]	√	√		No			√								
Wilson et al. 2005 [40]	√	√		No	√				√	73.6-75.9	68.6-72.2/ 74.0-77.7	45.3-56.8/85.7-89.9	13183 Expression	MPQA Corpus	
Pang & lee 2005 [41]	√		√	Yes	√			√		66.3	N/A	N/A	5006	Movie review	
Choi et al 2005 [42]		√		No		√			√	N/A	70.2-82.4	41.9-60.6	N/A	MPQA corpus	

Ng, V., et al, 2006 [43]	√	√			Yes	√			√								
Riloff, E., et al, 2006 [44]	√				Yes	√		√		√							
Konig & Brill, 2006 [45]	√				No	√		√	√				>91	N/A	N/A	1000(+) 1000(-)	Movie review
									√				<72	N/A	N/A	N/A	Customer Feedback
Abbasi et al , 2008 [46]	√			√		√			√				90-60-90.80	95.55		1000 message	US & middle Eastern Web forum postings
Zhang et al , 2011[47]	√	√			No	√		√	√				68.7	85.4	82.7		Twitter
Fang et al , 2011 [48]	√				No	√		√	√				56.2	66.8	N/A	2718 sentence	Multi domain sentiment Dataset
Mudinas et al 2012 [49]	√				No	√		√	√				85.09	82.3	N/A	1056 Review	Movie Review
Akshik et al , 2012 [50]	√				No			√		√			N/A	N/A	N/A		

4. Discussion & Future Scope

Since in Table 2, NLP and Pattern based techniques are used for the product review [2, 23, 36] and performed with 85.6 – 100 accuracy. While in the case of machine learning accuracy range varies between 56.2 and 91.0. So, it is true that NLP and Pattern based techniques performed better than the Machine Learning for sentiment analysis of product review. The Probabilistic technique [32] has accuracy range between 75.6 and 81.2. In previous works related to sentiment analysis are performed over the product and their features data set [19, 21, 17, 2, 32, 36, 41, 45, 49], while few papers in the field of Social issues [1] has been published.

Hence the recent emerging area of interest is sentiment analysis of social issues. Now a day most of the research scholars have been working on Twitter and YouTube comments data set. To perform sentiment analysis the most and common source of data set are web pages, social web site like face book , twitter, YouTube *etc*. There is a vast scope for research scholars to increase the accuracy level up to some extent by using well designed sentence structure.

Table 3. Future Directions Over Existing Works (Enhancing the Accuracy)

Author	Domain	Accuracy (%)	Technique	Objective	Future Scope
Jeonghee Yi et al., 2003 [2]	Web Document	91-93	NLP	Feature extraction	1) Need full parsing to provide better sentence structure analysis. 2) Increase the level of automation to handle the semantics accurately for validation of more advanced sentiment.
	Product Review	85.6			
M.Karamibekr et al.,2012 [1]	Social Issues	65	POS, Pattern based	Extraction of Verb & opinion term	1) Extract Main verb. 2) Increase the dictionaries of opinion terms & opinion verbs. 3) Use Hybrid Techniques. 4) Identify the ambiguous sentence and reduce them.

References

[1] M. Karamibekr and Ali A. Ghorbani, "Sentiment Analysis of Social Issues", International Conference on Social Informatics (IEEE), **(2012)**.

[2] J. Yi, T. Nasukawa, R. B and W. Niblack. "Sentiment Analyzer: Extracting sentiments about a given topic using Natural language Processing Techniques", Proceedings of the 3rd IEEE International Conference on Data Mining, **(2003)**.

[3] S.M.Vohra and J.B.Teraiya. "A comparative Study of Sentiment Analysis Techniques", Journal JIKRCE, vol. 2, no. 2, pp. 313-317, **(2013)**.

[4] B. Pang and L. Lee, "Opinion mining and sentiment analysis", Foundations and trends in Information Retrieval, vol. 2, **(2008)**, pp. 1-135.

[5] V.P.H Binali and W. Chen. "A state of the art opinion mining and its application domains". In IEEE International Conference on Industrial technology, **(2009)**, pp. 1-6.

[6] B. Pang, L. Lee and S.Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques", Proc. Of the 2002 ACL EMNLP Conf., **(2002)**, pp. 79-86.

[7] M.Hearst. "Direction-based text interpretation as an information access refinement". Text –Based Intelligent Systems, **(1992)**.

[8] W.Sack. "On the computation of point of view". In Proc. of the 12th AAAI Conf., **(1994)**.

[9] S.Das and M.chen. "Yahoo! For amazon : Extracting market sentiment from stock message boards". In Proc. Of the 8th APFA, **(2001)**.

[10] R.M. Tong, "An Operational System for detecting and tracking opinions in on-line discussion", SIGIR Workshop on Operational Text Classification, **(2001)**.

[11] P. Subasic and A. Huettner, "Affect analysis of text using fuzzy semantic typing", IEEE Trans, On Fuzzy Systems, Special Issue, **(2001)**.

[12] C. Whissell, "The dictionary of affect in language", Emotion: Theory, Research, and Experience, pp. 113–131.

[13] H. Li and K. Yamanishi. "Mining from open answers in questionnaire data". In Proc. of the 7th ACMSIGKDD Conf., **(2001)**.

[14] S. Morinaga, K. Yamanishi, K. Teteishi, and T. Fukushima. "Mining product reputations on the web". In Proc. of the 8thACM SIGKDD Conf., **(2002)**.

[15] L. Terveen,W. Hill, B. Amento, D. McDonald, and J. Creter, "PHOAKS: A system for sharing recommendations", CACM, vol. 40, no. 3, pp. 59–62, **(1997)**.

[16] B. Liu, Sentiment analysis and subjectivity, Handbook of Natural Language Processing, **(2010)**.

[17] T. Nasukawa and J. Yi. "Sentiment analysis: capturing favorability using natural language processing", Proceedings of the 2nd international conference on Knowledge capture, **(2003)**, pp. 70–77.

[18] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives", Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics, pp. 174– 181, **(1997)**.

[19] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of the 40th Annual Meetings of the Association for Computational Linguistics, Philadelphia, PA, pp. 417–424, (2002).

[20] R. Agrawal, S. Rajagoplalan, R. Srikant and Y. Xu, "Mining news groups using networks arising from social behavior", Proceedings of the 12th International World Wide Web Conference (WWW), (2003), pp. 529–535.

[21] K. Dave, S. Lawrence and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", Proceedings of the 12th International Conference on the World Wide Web (WWW), (2003), pp. 519–528.

[22] E. Riloff, J. Wiebe, and T. Wilson, "Learning subjective nouns using extraction pattern bootstrapping", Proceedings of the 7th Conference on Natural Language Learning, (2003); Edmonton, Canada.

[23] H. Yu and V. Hatzivassiloglou. "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences", Proceedings of the Conference on Empirical Methods in Natural Language Processing, (2003).

[24] P. Beineke, T. Hastie, and S. Vaithyanathan, "The sentimental factor: Improving review classification via human-provided information", Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, (2004).

[25] M. Efron, G. Marchionini and J. Zhiang, "Implications of the recursive representation problem for automatic concept identification in on-line government information", Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIST) SIG-CR Workshop, (2004).

[26] Z. Fei, J. Liu, and G. Wu, "Sentiment classification using phrase patterns", Proceedings of the 4th IEEE International Conference on Computer Information Technology, (2004).

[27] M. Gamon, "Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis", Proceedings of the 20th International Conferenceon Computational Linguistics, (2004).

[28] G. Grefenstette., Y.Qu, J. G. Shanahan and D. A. Evans, "Coupling niche browsers and affect analysis for an opinion mining application", Proceedings of the 12th International Conference Recherche d'Information Assistee par Ordinateur, pp. 186–194, (2004).

[29] M. Hu and B. Liu, "Mining and summarizing customer reviews", Proceedings of the ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, (2004).

[30] KANAYAMA, H. Kanayama, T. Nasukawa and H. Watanabe "Deeper sentiment analysis using machine translation technology", Proceedings of the 20th International Conference on Computational Linguistics, pp. 494–500, (2004).

[31] S. Kim and E. Hovy, "Determining the sentiment of opinions", Proceedings of the 20th International Conference on Computational Linguistics, (2004).

[32] B. Pang and L Lee, "A sentimental education: Sentimental analysis using subjectivity summarization based on minimum cuts", Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, (2004).

[33] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources", Proceedings of the Empirical Methods in Natural Language Processing (EMNLP) Conference, (2004); Barcelona, Spain.

[34] K. Nigam and M. Hurst "Towards a robust metric of opinion", Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text, (2004).

[35] J. Wiebe, T. Wilson, R. Bruse, M. Bell and M. Martin, "Learning subjective language.Comput. Linguist", vol. 30, no. 3, (2004), pp. 277–308.

[36] K. Hiroshi, N. Tetsuya and W. Hideo, "Deeper sentiment analysis using machine translation technology", Proceedings of the 20th international conference on computational linguistics (COLING 2004) , (2004); Geneva, Switzerland.

[37] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the Web", Proceedings of the 14th International World Wide Web Conference (WWW), (2005), pp. 342–351.

[38] G. Mishne, "Experiments with mood classification", Proceedings of the 1st Workshop on Stylistic Analysis of Text for Information Access, (2005); Salvador, Brazil.

[39] C. Whitelaw, N. Garg and S. Argamon, "Using appraisal groups for sentiment analysis", Proceedings of the 14th ACM Conference on Information and Knowledge Management, (2005), pp. 625–631.

[40] T. Wilson, J. Wiebe and P. Hoffman, "Recognizing contextual polarity in phrase-level sentiment analysis", In Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, (2005); British Columbia, Canada.

[41] B. Pang and L. Lee, "Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales", Proceedings of the 43rd annual meeting of the Association for Computational Linguistics (ACL), University of Michigan, USA, (2005).

[42] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, "Identifying sources of opinions with conditional random fields and extraction patterns", Proceeding of the conference on empirical methods in natural language processing (EMNLP 2005), **(2005)**;Vancouver, B.C., Canada.

[43] V. Ng, S. Dasgupta and S.M.N. Afrin, "Examining the role of linguistic knowledge sourcesin the automatic identification and classification of review", Proceedings of the COLING/ACL Conference, **(2006)**; Sydney, Australia.

[44] E. Riloff, S. Patwardhan, and J. Wiebe, "Feature subsumption for opinion analysis", Proceedings of the Conference on Empirical Methods in Natural Language Processing, **(2006)**; Sydney, Australia.

[45] A. C. K ̈onig and E. Brill, "Reducing the human overhead in text categorization, Proceedings of the 12th ACM SIGKDD conference on knowledge discovery and data mining, **(2006)**; Philadelphia, Pennsylvania, USA.

[46] A. Abbasi , H. Chen and A. Salem, "Sentiment Analysis in Multiple Languages", Feature Selection for Opinion Classification in Web Forums ACM Transactions on Information Systems, vol. 26, no. 3, **(2008)**.

[47] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu and B. Liu, "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis", Technical report, HP Laboratories, **(2011)**.

[48] Ji Fang and Bi Chen, "Incorporating Lexicon Knowledge into SVM Learning to Improve Sentiment Classification", Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), **(2011)**.

[49] A. Mudinas, D. Zhang, M. Levene, "Combining lexicon and learning based approaches for concept level sentiment analysis", Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, **(2012)**; New York, NY, USA.

[50] A. Kumar and T. M. Sebastian, "Sentiment analysis on twitter", IJCSI, vol. 9, no. 3, **(2012)**.

[51] P, D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association", Trans. Inf. Syst., vol. 21, no. 4, 315–346, **(2003)**.

[52] J. R. Martin and P. R. R. White, "The Language of Evaluation: Appraisal in English", **(2005)**; Palgrave, London,.

[53] O. de Vel, A. Anderson, M. Corney, and G. Mohay, "Mining e-mail content for author identification forensics, Rec. 30, 4, 55–64, 2001.

[54] R. Zheng, J. Li, Z. Huang and H. CHEN, "A framework for authorship analysis of online messages: Writing-Style features and techniques", J. Amer. Soc. Inf. Sci. Technol, vol. 57, no. 3, **(2006)**, pp. 378–393.

[55] L. Robinson, "Debating the events of September 11th: Discursive and interactional dynamics in three online", J. Comput. Mediat. Commun, vol. 10, no. 4, **(2005)**.

[56] J. Khairnar and M. Kinikar, "Machine Learning Algorithms for Opinion Mining and sentiment classification", IJSRP, vol. 3, no. 6, **(2013)**.

[57] R. Prabowo and M. Thelwall, "Sentiment Analysis: A Combind Approach", Journal of Informetric - Elsevier, **(2009)**.

Authors

Shailendra Kumar Singh obtained his B.Tech degree in Information Technology from UPTU (Lucknow) in 2012. Currently he is pursuing Master of Engineering in Software Engineering from Birla Institute Of Technology, Mesra, Ranchi (India).

Dr.Sanchita Paul is an Assistant Professor in Department of Computer Science & Engineering, Birla Institute of Technology, Mesra, Ranchi (India). She obtained B.E, M.E and Ph.D Degree in Computer Science and Engineering. Her area of Interests is Artificial Intelligence, Cloud Computing, Bioinformatics, NLP, Automata Theory, Design and Analysis of Algorithms.

Dhananjay Kumar obtained his B.Tech degree in Computer Science & Engineering from RGPV, Bhopal. Currently he is pursuing Master of Technology in Information Technology from Birla Institute Of Technology, Mesra, Ranchi (India).