A Review on Sentiment Analysis Approaches

Vaibhav Wanode
Samrat Ashok
Technological Institute,
Vidisha

Shrikant Kurmi
Samrat Ashok
Technological Institute,
Vidisha

Udit Rajput
Samrat Ashok
Technological Institute,
Vidisha

Pankaj Pandit
Samrat Ashok
Technological Institute,
Vidisha

ABSTRACT

Sentiment analysis has become one of the most profound research areas with the increasing growth of social media on web. Nowadays, millions of users exchange their views, ideas, expressions, feelings, opinions on social media like twitter and Facebook. Sentiment analysis or opinion mining mainly focuses on classification and prediction of people's opinion about a specific target. It involves classifying text documents or sentences based on the opinion expressed being positive or negative about a given topic. The task of sentiment analysis seems similar to text classification, but it faces many challenges that have motivated focused research in this domain. To automate the sentiment analysis task, various machine learning and lexicon based techniques have been proposed in literature. Though these techniques have been widely used for sentiment classification, these techniques failed to achieve the best results in terms of accuracy and to resolve all the challenges. So, there is need to develop new automated techniques that resolve all the challenges and give the best performance

General Terms

Sentiment Analysis, Opinion Mining

Keywords

Supervised machine learning, Lexicon based approach

1. INTRODUCTION

Sentiment analysis or opinion mining is the automated and computational study of people's sentiments, expressions, attitudes and emotions towards a particular target. The target object is any individual, event or topic. The reviews written by people cover these topics. The two terms sentiment analysis or opinion mining can be used interchangeably. However, some researchers stated that both these expressions have slightly different notions [12]. Opinion mining extracts and analyses people's opinion about an entity while sentiment analysis identifies the sentiment expressed in a text then analyses it. Therefore, Sentiment analysis aims to automate the task of finding opinions, identifying the sentiments they express, and then classifying the sentiment polarity. People's opinion plays a crucial role in decision making in various domains. If one wants to buy a particular product, he or she may want to know other's opinion before purchasing the product. In real world, organizations and businesses want to find their consumers feedback about their product or services. In recent years, sentiment analysis applications have spread through many domains from recommendation systems, Ad placements, and trend prediction to healthcare and politics. Recent years witnessed the explosive growth of social media (like blogs, reviews, forums, comments and postings on social networking sites) on the web. Most of the organizations are using these contents to make the decisions. Nowadays theorganizations are not dependent on opinion polls, surveys and focus groups due to the huge amount of data available publicly. The task of mining opinion is formidable due to the need to check the individual web sites. It is very difficult, for a human reader to identify the relevant sites and extract the opinions in them. Therefore, there is need of automated sentiment analysis. Most of the organizations are using their own analysis tools to find the opinions of the consumers. Two approaches are widely used for opinion mining and sentiment analysis: 1) Machine learning based 2) Lexicon based. The machine learning based approach uses various supervised and unsupervised learning algorithms for sentiment classification. While the Lexicon based techniques use a lexicon dictionary with sentiment words related to specific domain for sentiment classification. The dictionary contains the information about the polarity of each word, whether the words are positive or negative. The word in the sentence can be matched with the words in dictionary to determine their polarity. Some researchers combined the machine learning and lexicon based techniques [7, 11]. The various sentiment analysis techniques are categorized as shown in Fig 1.

2. SENTIMENT ANALYSIS APPROACHES

The Sentiment classification is a task of classifying a target unit in a document to positive (favorable) or negative (unfavorable) class. There are three main classification levels [27]:

1. document level: classifies an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit (talking about one topic);
2. sentence-level: classifies sentiment expressed in each sentence. If the sentence is subjective it classifies it in positive or negative opinions;
3. aspect-level: classifies the sentiment with respect to the specific aspects of entities. Users can give different opinions for different aspects of the same entity.

At document level it is possible to classify whether a whole users opinion expresses a positive or negative sentiment. For example, given a product/service review, it is possible to determine whether the review expresses an overall positive or negative opinion about the product/service. The sentence determines whether each sentence expresses a positive or negative. While the entity/aspect level, instead of looking at language construction (sentences, phrases, paragraphs, clauses etc), directly focus on the opinion itself. It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion). The document sentiment classification approach is used by [6] that classify movie reviews by using supervised machine learning method. In [28] the authors used the semantic orientation of words defined by [8] and several information from the Web and thesaurus. They achieved 85% accuracy with and the semantic orientation of words and the lemmatized word unigram. While the study provided by [29] used the sentence level classification approach. It considered word dependency trees as features for sentence-wise sentiment polarity classification. On the contrary the study conducted by [8] determined the relationship between a polarity-unknown word and a set of selected manually seeds for classifying the polarity-unknown word into positive or negative class. Also the study provided by [11] to extract sentiment polarities by using expressions such as or "fast but inaccurate" or "beautiful and smart". In [30] a survey on different methods of sentiment analysis available in literature related to product reviews (such as machine learning, semantic orientation, opinion polling, holistic lexicon-based approach etc.) is carried out. The survey underlines that sentiment analysis/opinion mining play vital role to make decision about product /services. Another survey on approaches used for sentiment analysis is provided in [31] in which three approaches for performing sentiment extraction are described:

- subjective lexicon approach: is a list of words to witch is assigned a score that indicates its nature in terms of positive, negative or objective;
- n-gram modeling approach: that can use uni-gram, bi-gram, tri-gram or combination of these for the sentiments classification;
- machine learning approach: performs the semi and/or supervised learning through the extraction of the features from the text and learn the model.

While [32] analyse three types of techniques for Sentiment Classification:

i. machine learning approach,
ii. lexicon based approach and
iii. hybrid approach.

The machine learning approach is used for predicting the polarity of sentiments based on trained as well as test data sets. It applies the ML algorithms and uses linguistic features. The main advantage of this method is the ability to adapt and create trained models for specific purposes and contexts, its main disadvantage is the low applicability of the method on new data because is necessary the availability of labeled data that could be costly or even prohibitive. It can use supervised and unsupervised methods. The machine learning uses a supervised approach when there is a finite set of classes (positive and negative). This method needs labeled data to train classifiers [6]. In a machine learning based classification a training set is used by an automatic classifier to learn the different characteristics of documents, and a test set is used to validate the performance of the automatic classifier. The unsupervised methods are used when it is difficult to find labeled training documents. Unsupervised learning does not require prior training in order to mine the data. Unsupervised approaches to document-level sentiment analysis are based on determining the semantic orientation (SO) of specific phrases within the document. If the average SO of these phrases is above some predefined threshold the document is classified as positive, otherwise it is deemed negative. Among the machine learning approaches the most used are: (i) Bayesian Networks: it is a probabilistic approach that models relationships between features in a very general way. It is based on directed acyclic graph in which nodes are variables and arcs represent the dependence between variables (ii) Naive Bayes Classification: it is an approach particularly suited when the dimensionality of the inputs is high. Despite its simplicity, it can often outperform more sophisticated classification methods (iii) Maximum Entropy: this method is mostly used as alternatives to Naive Bayes classifiers because it does not assume statistical independence of the random variables (features) that serve as predictors. The principle behind Maximum Entropy is to find the best probability distribution among prior test data. (iv) Neural Networks: this model is based on a collection of natural/artificial neurons uses for mathematical and computational model analysis (v) Support Vector Machine: it is a supervised learning model which analyzes data and patterns that can be used for classification and regression analysis. The basic idea behind this is to find a maximum margin hyper plane represented by vector. It finds an optimal solution. While the lexicon-based approach does not need any prior training in order to mine the data. It uses a predefined list of words, where each word is associated with a specific sentiment. They are based on the counting of positive and negative words. These methods vary according to the context in which they were created. Lexical don't need labeled data, but is hard to create a unique lexical-based dictionary to be used for different contexts. For example slang used in Social Networks is rarely supported in lexical methods. Among the lexicon-

based approaches the most used are: (i) Dictionary based approach: it is a method that translates a word by word as a dictionary without correlating the meaning of words between them (ii) Novel Machine Learning Approach: it integrates important linguistic features into automatic learning (iii) Corpus based approach: it has been widely used to explore both written and spoken texts in order to assign a sentiment factor of words that depend on frequency of their occurrences (iv) Ensemble Approaches in sentiment classification: it increases classification accuracy by combining arrays of specialized learners. The study provided by [34] gives an example of the lexicon based approach applied on a morphologically rich language: Urdu It focuses on the sentence grammatical structures, besides to the morphological structure of the words. For the analysis, two types of grammatical structures (adjective phrases as Senti-Units and nominal phrases as their targets) are extracted and then linked. For the extraction and linking two parsing methods have been implemented: shallow and dependency parsing. In [35] the authors focused on the lexicon-based approach for Arabic sentiment analysis by building the main two components of the lexicon-based sentiment analysis approach: the lexicon and the sentiment analysis tool. The study provided a guide for the researchers in their on-going efforts to improve lexicon-based sentiment analysis.

2.1 Tools for sentiment Analysis

There are many studies that provide methods and tools used for sentiments analysis. The most used tools for detecting the feelings polarity (negative and positive affect) of a message is based on the emoticons. Emoticons are face-based and symbolize sad or happy feelings, although there are a wide range of non-facial variations. To extract the feelings polarity from emoticons, different set of common emoticons can be used (http://messenger.msn.com/Resource/Emoticons.aspx; http://www.cool-smileys.com/text-emoticons; http://messenger.yahoo.com/features/emoticons). Therefore, emoticons have been often used in combination with other techniques for building a training dataset in supervised machine learning techniques [36]. Another method is the Linguistic Inquiry and Word Count [37] that allows analysing not only positive and negative but also emotional, cognitive, and structural components of a text based on the use of a dictionary containing words and their classified categories. For example, the word "agree" belongs to the word categories: assent, affective, positive emotion, positive feeling, and cognitive process. This software is available at http://www.liwc.net/. Happiness Index [38] is a sentiment scale that uses the popular Affective Norms for English Words (ANEW) [39]. It gives scores for a given text between 1 and 9, indicating the amount of happiness. The authors calculated the frequency that each word from the ANEW appears in the text and then computed a weighted average of the valence of the ANEW study words. Another tool is the SentiStrength (http://sentistrength.wlv.ac.uk/Download) that is considered by [40] "the most popular stand-alone sentiment analysis tool". It uses a sentiment lexicon for assigning scores to negative and positive phrases in text. For identifying the feeling polarity several key classifiers are proposed [41, 42]. In [43] the SentiWordNet (at http://sentiwordnet.isti.cnr.it/.) tool is described. SentiWordNet is a lexical resource publicly available for supporting sentiment classification and opinion mining applications. It is based on an English lexical dictionary called WordNet [44] that ghaters adjectives, nouns, verbs etc. into synonym sets called synsets.

Each synset is associated to three numerical scores Pos(s), Neg(s), and Obj(s) which indicate how positive, negative, and "objective" (neutral) the terms contained in the synset are. The scores, which are in the values of [0, 1] and add up to 1, are obtained using a semi-supervised machine learning method. The tool, used in opinion mining, is based on WordNet an English lexical dictionary that collect nouns, verbs, adjectives and other grammatical classes into synonym sets (synsets) [44]. Another tool is the PANAS-t [45]. The tool consists of an adapted version of the Positive Affect Negative Affect Scale (PANAS) [46], method used in psychology. The PANAS-t tracks increases or decreases in sentiments over time; it is based on a large set of words associated with eleven moods: joviality, assurance, serenity, surprise, fear, sadness, guilt, hostility, shyness, fatigue, and attentiveness. This method computes the score for each sentiment for a given time period as values between [−1.0, 1.0] to indicate the change. The open source tool SailAil Sentiment Analyzer (SASA) [47] was evaluated with 17,000 labeled tweets on the 2012 U.S. Elections. It was evaluated also by the Amazon Mechanical Turk (AMT), where "turkers" were invited to label tweets as positive, negative, neutral, or undefined. The SASA python package version 0.1.3 is available at https://pypi.python.org/pypi/sasa/0.1.3. In [45] the authors developed a new sentiment analysis method that combines the different described approaches in order to provide the best coverage and competitive agreement. They implemented a public Web API, called iFeel (http://www.ifeel.dcc.ufmg.br), which provides comparative results among the different sentiment methods for a given text. In [48], the authors described SenticNet a tool that explores artificial intelligence and semantic Web techniques. The tool explores artificial intelligence and semantic Web techniques. It uses Natural Language Processing (NLP) techniques to infer the polarity of common sense concepts from natural language text at a semantic level, rather than at the syntactic level. SenticNet was tested to measure the level of polarity in opinions of patients about the National Health Service in England [48]. SenticNet version 2.0 is available at http://sentic.net/. In [49, 50] EWGA and FRN tools are used. The EWGA tool uses an entropy-weighted

genetic algorithm for an efficiently selection of features for sentiment classification using a wrapper-model. While the FRN uses a feature relation network considering two syntactic n-gram relations: parallel relations and subsumption [51]. Sentiment140 formerly known as Twitter Sentiment discovers the positive and negative opinions and sentiment of a brand, product, or topic on Twitter. This tool uses classifiers built from machine learning algorithms. Unlike other tools that show aggregated numbers which makes it difficult to assess how accurate their classifiers are, this tool is able to classify individual tweets. The NRC Hashtag Sentiment Lexicon (version 0.1) is a list of words with associations to positive and negative sentiments. The lexicon is distributed in three files: unigrams-pmilexicon.txt, bigrams-pmilexicon.txt, and pairs-pmilexicon.txt. The NRC Emotion Lexicon is comprised of frequent English nouns, verbs, adjectives, and adverbs annotated for eight emotions (joy, sadness, anger, fear, disgust, surprise, trust, and anticipation) as well as for positive and negative sentiment.

3. Classification of sentiment analysis approaches

Starting from the analysis provided in the previous sections a classification of sentiment analysis approaches with respect to features/techniques and advantages /limitations is provided in Table 1.

Machine learning based approach uses classification technique to classify text; it consists of two sets of documents: training and a test set. The training set is used for learning the differentiating characteristics of a document, while the test set is used for checking how well the classifier performs. The features of machine learning based approach for sentiment classification are:

- term presence and their frequency: that includes uni-grams or n-grams and their presence or frequency.
- part of speech information: used for disambiguating sense which is used to guide feature selection.
- negations: has the potential of reversing sentiments opinion words/phrases: that expresses positive or negative sentiments.

The lexicon based approach uses sentiment dictionary with opinion words and match them with the data for determining polarity. There are three techniques to construct a sentiment lexicon: manual construction, corpus-based methods and dictionary-based methods. The manual construction is a difficult and time-consuming task. Corpus-based methods can produce opinion words with relatively high accuracy. Finally, in the dictionary based techniques, the idea is to first collect a small set of opinion words manually with known orientations, and then to grow this set by searching in the WordNet dictionary for their synonyms and antonyms.

SENTIMENT CLASSIFICATION APPROACHES		FEATURES/ TECNIQUES	ADVANTAGES AND LIMITATIONS
Machine learning	Bayesian Networks Naive Bayes Classification Maximum Entropy Neural Networks Support Vector Machine	Term presence and frequency Part of speech information Negations Opinion words and phrases	ADVANTAGES the ability to adapt and create trained models for specific purposes and contexts LIMITATIONS the low applicability to new data because it is necessary the availability of labeled data that could be costly or even prohibitive
Lexicon based	Dictionary based approach Novel Machine Learning Approach Corpus based approach Ensemble Approaches	Manual construction, Corpus-based Dictionary-based	ADVANTAGES wider term coverage LIMITATIONS finite number of words in the lexicons and the assignation of a fixed sentiment orientation and score to words
Hybrid	Machine learning Lexicon based	Sentiment lexicon constructed using public resources for initial sentiment detection Sentiment words as features in machine learning method	ADVANTAGES lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain LIMITATIONS noisy reviews

Finally, in the hybrid approach, the combination of both the machine learning and the lexicon based approaches has the potential to improve the sentiment classification performance. There are some advantages and limitations in using these different approaches depending on the purpose of the analysis. We provide an overview of the main. The main advantage of machine learning approaches is the ability to adapt and create trained models for specific purposes and contexts, while the limitation is that it is difficult integrating into a classifier, general knowledge which may not be acquired from training data. Furthermore, learnt models often have poor adaptability between domains or different text genres because they often rely on domain specific

features from their training data. Lexicon-based approaches have the advantage that general knowledge sentiment lexicons have wider term coverage, however these approaches have two main limitations. Firstly, the number of words in the lexicons is finite, which may constitute a problem when extracting sentiment from very dynamic environments. Secondly, sentiment lexicons tend to assign a fixed sentiment orientation and score to words, irrespective of how these words are used in a text. The main advantages of hybrid approaches are the lexicon/learning symbiosis, the detection and measurement of sentiment at the concept level and the lesser sensitivity to changes in topic domain. While the main limitation is that reviews are with a lot of noise (irrelevant words for the subject of the review) are often assigned a neutral score because the method fails to detect any sentiment. Moreover a classification of tools for sentiment analysis according to the different techniques used for sentiment analysis is given in Table 2.

Table 2. Tools for sentiment analysis

TOOLS FOR SENTIMENT ANALYSIS	TECNIQUES USED BY TOOLS
EMOTICONS	Emoticons contained in the text
LIWC	Dictionary and sentiment classified categories
SentiStrengh	LIEC dictionary with new features to strength and weak sentiments
Senti WordNet	Lexical dictionary and scores obtained by semi-machine learning approaches
SenticNet	Natural language processing approach for inferring the polarity at semantic level
Happiness Index	Affective Norms for English Words (ANEW) and scores for evaluating happiness in the text
AFINN	Affective Norms for English Words (ANEW) but more focused on the language used in microblogging platforms.
PANAS-t	Eleven-sentiment psychometric scale
Sentiment140	API that allows classifying tweets to polarity classes positive, negative and neutral.
NRC	Large set of human-provided words with their emotional tags.
EWGA	Entropy-weighted genetic algorithm
FRN	Feature relation network considering syntactic n-gram relations

The different approaches and tools analysed in this paper can be applied in different fields such as: business, politic, public actions and finance. On considering the business field many studies have been developed in the area of reviews of consumer products and services. There are many websites that provide automated summaries of reviews about products, such as Google Product Search. In [47], the authors developed SumView, a Web-based system that summarizes the product reviews and customer opinions. It integrates review crawling from Amazon.com, automatic product feature extraction along with a text field where users can input their desired features, and sentence selection using the proposed feature-based weighted non-negative matrix factorization algorithm. The most representative sentences are selected to form the summary for each product feature [47]. In the business domain context, sentiment analysis is also used for brand reputation, online advertising and on-line commerce. Sentiment analysis is used to monitor the reputation of a specific brand on Twitter and/or Facebook. Tweetfeel is an application that performs real-time analysis of tweets that contain a given term (http://www.tweetfeel.com). With respect to on-line advertising, it has become one of the major revenue sources of the Web ecosystem. In this context, recent applications of sentiment analysis are in Blogger Centric Contextual Advertising [52] and dissatisfaction oriented online advertising [53] which refer to the development of personal ads in any blog page, chosen in according to company business interests. Another application of sentiment analysis in the business domain is represented by the on-line commerce. The assumption is that consumers value others' opinions about restaurants, travel and stores guide each consumer in searching, hence Bing- and Google-computed star ratings. An important study in this context is developed by [54] that provided a Senti-lexicon for restaurant reviews. With respect to the politic domain, the voting advise applications represent an important application of sentiment analysis. It enables campaign managers to track how voters feel about different issues and how they relate to the speeches and actions of the candidates. An analysis of tweets related to the 2010 campaign can be found at ttp://www.nytimes.com/interactive/us/politics/2010-twittercandidates.html. In this context, sentiment analysis is also used to clarify politicians' positions, such as what public figures support or oppose, to enhance the quality of information that voters have access to [24].

Sentiment analysis is also used within the public actions implementation process. In this context sentiment analysis gives an important contribution also in monitoring real-world events. Linguists at UT Austin use sentiment analysis to get a read on the rapidly evolving situation in the middle east. In particular sentiment analysis for monitoring critical information about earthquake locations and magnitude, riot locations; the monitoring help policy makers to minimise damage in areas which are expected to be affected next by such events. Another important application of sentiment analysis is the monitoring of the opinions that people submit about pending policy or government-regulation proposals [55, 56] and legal matters sometimes also known as "blawgs" [57]. A new emerging sentiment analysis domain is represented by the modern intelligent transportation systems (ITSs). For the completeness of ITS space, it is necessary to collect and analyze the public

opinions exchange. The first attempt to apply sentiment analysis on the area of traffic is represented by the study developed by [58] where a traffic sentiment analysis (TSA) has been developed. The TSA allows analysing the traffic problem in a humanizer way. Finally, with respect to the finance domain, millions of financial news are circulating daily on the Web and financial markets are continuously changing and growing. There are numerous news items, articles, blogs, and tweets about each public company. A sentiment analysis system can use them and aggregate the sentiment about them as a single score that can be used by an automated trading system, such as the system The Stock Sonar (http://www.thestocksonar.com) that shows graphically the daily positive and negative sentiment about each stock alongside the graph of the price of the stock. In [59], the authors have investigated the Arizona Financial Text (AZFinText) system a financial news article prediction system, and has paired it with a sentiment analysis tool. They have found that the financial news articles have a direct impact on influencing the prices of commodities and shares. The analysis of finance news and opinions also allow monitoring the financial risk. The study conducted by [47] uses the finance- specific sentiment lexicon to model the relations between sentiment information and financial risk. In specific, they formulate the problem as two different prediction tasks: regression and ranking. In Table 3 a schematization of sentiment analysis applications is provided.

Table 3: Sentiment analysis applications

SENTIMENT ANALYSIS APPLICATIONS
BUSINESS
Consumers voice
Brand reputation
Online advertising: Blogger Centric Contextual Advertising Dissatisfaction oriented online advertising
On-line commerce
POLITIC
Voting advise applications
Clarification of politicians' positions
PUBLIC ACTIONS
Real-world events monitoring
Legal matters "blawgs"
Policy or government-regulation proposals
Intelligent transportation systems
FINANCE
Prices of commodities and shares evolution
Financial risk individuation

4. CONCLUSION

The paper starting from the analysis of different studies provided in the literature, provides a classification of (i) sentiment classification approaches with respect to features/techniques and advantages /limitations (ii) tools for sentiment analysis with respect to the different techniques used for sentiment analysis. The sentiment classification approaches can be classified in (i) machine learning (ii) lexicon based and (iii) hybrid approach. The machine learning approach is used for predicting the

polarity of sentiments based on trained as well as test data sets. While the lexicon based approach does not need any prior training in order to mine the data. It uses a predefined list of words, where each word is associated with a specific sentiment. Finally in the hybrid approach, the combination of both the machine learning and the lexicon based approaches has the potential to improve the sentiment classification performance. On considering the tools used for sentiments analysis, the most used tools for detecting the feelings polarity (negative and positive affect) are discussed in the paper: Emoticons, LIWC, SentiStrengh, Senti WordNet, SenticNet, Happiness Index, AFINN, PANAS-t, Sentiment140, NRC, EWGA and FRN. The different approaches and tools analysed in the paper can be applied in different fields such as: business, politic, public actions and finance. In the business domain sentiment analysis is mainly used for detecting consumer's voice, brand reputation and the online advertising and on-line commerce trend. With respect to the politic domain, the voting advise represents an important application of sentiment analysis that is also used to clarify politicians' positions enhancing the quality of information that voters have access to. Sentiment analysis is also used within the public actions implementation process. In this context it gives an important contribution in monitoring real-world events. Another important application of sentiment analysis is the monitoring of the opinions that people submit about pending policy or government-regulation proposals and legal matters. A new emerging sentiment analysis domain is also represented by the modern intelligent transportation systems. Finally with respect to the finance domain, the sentiment analysis is used to detect the trend of prices of commodities and shares evolution and the financial risks. A future challenge in applying sentiment classification approaches and tools for sentiment analysis of posts in social media is to overcome the ambiguity that actually represents particular problem since it is not easily make use of coreference information. Typically the analysed posts contain irony and sarcasm, which are particularly difficult to detect. So an evolution of approaches and tools is required to overcome this limitation.

5. REFERENCES

1. Liu, B. 2006. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer.
2. Liu, B. 2006. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer.
3. Wiebe, J. & Riloff, E. 2005. Creating subjective and objective sentence classifiers from unannotated texts. Computational Linguistics and Intelligent Text Processing, 2005, pp. 486-497.

4. Nasukawa ,T. & Yi, J. 2003. Sentiment analysis: capturing favorability using natural language processing. In Proceedings of the 2nd international conference on Knowledge capture, October 23–25, 2003. (pp. 70–77). Florida, USA

5. Morinaga, S., Yamanishi, K., Tateishi, K., Fukushima, T. 2002. Mining product reputations on the web. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 341-349.

6. Pang, B., Lee, L., Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. Proc. of 7th EMNLP, pp.79-86.

7. Tong, R.M. 2001. An operational system for detecting and tracking opinions in on-line discussion. In Proceedings of SIGIR Workshop on Operational Text Classification.

8. Turney, P. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classifcation of reviews. In Proceedings of the 40th ACL, pp. 417-424.

9. Wiebe, J. (2000) Learning subjective adjectives from corpora. In Proceedings of National Conference on Artificial Intelligence.

10. Wilson, T., Wiebe, J., Hoffmann, P. 2009. Recognizing contextual polarity: An exploration of features forphraselevel sentiment analysis. Computational Linguistics, 35(3), pp. 399-433.

11. Hatzivassiloglou, V. & McKeown, K.R. 1997. Predicting the semantic orientation of adjectives. In Proceedings of the 8th conference on European chapter of the association for computationnal linguistics Madrid, Spain, pp.174-181.

12. Pang, B., & Lee, L. 2004. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting of the Association for Computational Linguistics (ACL), pp. 271–278. Barcelona, Spain

13. Yi, J., Nasukawa, T., Niblack, W., Bunescu, R. 2003. Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. In Proceedings of the 3rd IEEE international conference on data mining (ICDM 2003), November 19–22, 2003, pp. 427-434 Florida, USA.

14. Hiroshi, K., Tetsuya, N., Hideo, W. 2004. Deeper sentiment analysis using machine translation technology. In Proceedings of the 20th international conference on computational linguistics (COLING 2004), August 23- 27, pp. 494-500, Geneva, Switzerland.

15. Bollen, J., Mao, H., Zeng, X. 2011. Twitter mood predicts the stock market. Journal of Computational Science, 2(1), pp. 1-8.

16. Abbasi, A., Fu, T., Zeng, D., Adjeroh, D. 2013. Crawling Credible Online Medical Sentiments for Social Intelligence. Proceedings of the ASE/IEEE International Conference on Social Computing

17. Fu, T., Abbasi, A., Zeng, D., Chen, H. 2012. Sentimental Spidering: Leveraging Opinion Information in Focused Crawlers. ACM Transactions on Information Systems, 30(4), 24

18. Abbasi, A., Chen, H., Salem, A. 2008. Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums. ACM Transactions on Information Systems, 26(3), 12.

19. Pang, B., & Lee, L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, Vol 2 (1-2).

20. Niu, Y., Zhu, X., Li, J., Hirst, G. 2005. Analysis of polarity information in medical text. In Proceedings of the American Medical Informatics Association 2005 Annual Symposium.

21. Balahur, A., Kozareva, Z., Montoyo, A. 2009. Determining the polarity and source of opinions expressed in political debates

22. Ku, L.W., Li, L.Y., Wu, T.H., Chen, H.H. 2005. Major topic detection and its application to opinion summarization. In Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR), Salvador, Brasil.

23. Bansal, M., Cardie, C., Lee, L. 2008. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In Proceedings of the International Conference on Computational Linguistics (COLING), 2008. Poster paper, pp.15-18.

24. Terveen, L., Hill, W., Amento, B., McDonald, D., Creter, J. 1997. PHOAKS: A system for sharing recommendations. Communications of the Association for Computing Machinery (CACM), 40(3), PP. 59-62

25. Kim, S.M. & Hovy, E. 2006. Automatic identification of pro and con reasons in online reviews. In Proceedings of the COLING/ACL Main Conference Poster Sessions, pp. 483-490

26. Medhat, W., Hassan, A., Korashy, H. 2014. Sentiment analysis algorithms and applications: A survey, Ain Shams Eng.

27. Mullen, T., & Collier, N. 2004. Sentiment Analysis using Support Vector Machines with Diverse Information Sources. Proc. of 9th EMNLP, pp. 412-418.

28. Kudo, T., & Matsumoto, Y. 2004. A Boosting Algorithm for Classification of Semi-Structured Text. In EMNLP, Vol. 4, pp. 301-308.

29. Jebaseeli, A. N., & Kirubakaran, E. 2012. A survey on sentiment analysis of (product)

reviews. International Journal of Computer Applications, 47(11).

30. Kaur, A., & Gupta, V. 2013. A survey on sentiment analysis and opinion mining techniques. Journal of Emerging Technologies in Web Intelligence, 5(4), 367- 371.