Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 3 Issue 5, May - 2016

Survey: Computational Study Of Public Moods From Social Media

Arslan Ali Raza
Qurtuba University of Science and Information Technology, D.I.Khan, Pakistan
arslanali_raza@yahoo.com

Bashir Ahmad
Qurtuba University of Science and Information Technology, D.I.Khan, Pakistan
bashahmad@gmail.com

***Abstract*—Sentiment Analysis is the most prominent field of Machine Learning and the problem of Natural Language Processing. The aim of sentiment analysis system is to predict the moods and sentiment of public expressed in the form of text from social media. This study exposes the variety of issues for Sentiment analysis related to Machine Learning and Natural Language Processing. The major goal of this Survey is to highlight the substantial research in the field of Sentiment Analysis. The challenges, data sources, tasks, levels, applications and recent trends of sentiment analysis are covered in this review paper.**

Keywords—Machine Learning, Natural Language Processing, Sentiment Analysis

1. Introduction

Sentiment Analysis and Opinion Mining is the field of Data Mining and problem of Natural Language Processing. The aim of Sentiment analysis is to track the public moods towards different entities belonging from various fields of life; basically it is the sub field of Machine Learning. Sentiment analysis is the rapid growing research area in current scenario because thousands of views, suggestions, opinions, and sentiments about different real world entities are available on rising channel namely “Social Media”. This huge amount of available opinionative data is the key source for various organizations as the quality and selling rate of a product directly depends on public suggestions and views. This new rising field has provided noticeable facilities not only to organization but also to political intelligence and social psychologists.

In this new age of web, users express their feelings on this novel growing media, more specifically the microblogs are designed especially for the exchange of short text information. Microbloging sites allow its users to post a short text messages for communication. Through these sites the online users exchange their suggestions, views and thoughts with each other in very frank and informal way in the form of unstructured text. These online networking sites have changed the whole world by connecting the peoples belonging from various regions with each other. Infact Microblogs are proved as electronic word of mouth. The feelings, appraisals, attitudes and emotions shared on these social sites are treated as sentiments or opinions. Sentiment Analysis is the computational study of sentiments and opinions. Inspite of the proliferation of microbloging sites the informal communication style has created several challenges. Following are some common challenges;

Unstructured Text: Sentiments and opinions are expressed mostly in unstructured way on microbloging sites.

Dual Sense: Opinion with dual polarity: In some context the polarity of a term is positive while in another situation the same term may have negative polarity.

Negation: Negation handling is another key challenge in this field.

Slang Detection: People express their feelings in short, noisy form due to the restriction of length of text in microbloging sites. So slang detection is another important challenge in SA.

Context dependency: Mostly sentiment terms are context dependent, so analyzing sentiment of whole sentence / document according to specific context is also the challenge of this field.

Previous research shows that some challenges have been solved with appropriate solutions but some other challenges still need to be unfolded like Spam detection, Cross Domain Classification, Sarcasm Identification, Multilingual Text Classification, Domain Independency and Aspect based efficient SA. In this Survey our focus is to highlight the new growing field Sentiment Analysis, its challenges, techniques, levels, tasks and areas. The major aim of this paper is to unfold the notion of Sentiment analysis in NLP. Comparative analyses of various approaches in the field of SA are presented in literature review. The rest of the paper is organized as follows: Section 1 presents the introduction and challenges of SA, Section 2 presents Literature Survey (introduces various techniques used for SA), Section 3 presents Data source, Section 4 presents Tasks of SA, Section 5 introduces levels of SA, and Section 6 presents Conclusion.

2. Literature Review

The term sentiment analysis was first used in 2003 by Nasukawa and Yi [14] and Opinion mining by Dave et al in 2003 [15]. The field of sentiment analysis is introduced recently by researchers' community and growing rapidly due to the availability of source data on web. This research field has changed the way of analysis and provides ease to various organizations. The aim of sentiment analysis is to detect the public moods digitally about specific entity for better evaluation. SA systems digitally recognize the polarity of each word whether a single sentiment term is expressing positive opinion or negative. These polarities have great importance for various surveyors. The words opinion [16], Sentiment [21], appraisal [17], Attitude [19], Valence [20], Semantic orientation [10,22] and Polarity [18] are used to express related but not the same concepts. Basically Sentiment analysis is the computational study of opinions, sentiments conveyed in the form of text. In each sentiment analysis process the first step is the subjectivity classification. The subjectivity classification is the process of classifying the whole text into subjective (opinion bearing terms/ polar terms) and objective (Facts). J. Wiebe et al. [31] defined that the subjectivity is the linguistic expression of public opinions, emotions, speculation and attitudes. Much research in the field of sentiment analysis has been done on product and movie reviews. Let's take a view of previous research performed by different researchers and academicians. The main approaches used for sentiment analysis are Machine Learning and Lexicon Based Approach. In machine learning approach the whole text is classified using various classification techniques and machine learning algorithms like Naïve Bayes, Bayesian Network, Maximum Entropy and Support Vector Machine. Analyzing public Sentiments using machine learning generally based on supervised machine learning approaches. The above algorithms are actually the supervised machine learning algorithms. The machine learning classification needs two sets for classification (1) Training data sets (2) Test data sets. Sentiment classification approaches are depicted in Fig. 1 for the sake of better understanding. Some other machine learning techniques in NLP are N-gram model, Centroid classifier, C5, ID3, winnow classifier and KNN. These machine learning techniques are used for different level of sentiment analysis as Naïve Bayes algorithm is widely used in document level sentiment analysis [12]. These classifiers are the family of simple probabilistic classifiers based on implementing Bayes theorem with naïve assumption between the various features. The basic aim is to estimate the probabilities provided in a test document by applying the probabilities of words and categories. Bo Pang et al.[17] used Naïve Bayes classifier for the classification of product reviews. Kang and Yoo [27] proposed an improve NB classifier for solving the tendency problem which achieves higher accuracy for positive classification. Support Vector Machine is a statistical classification method proposed by Vapnik [24] in 1979 that examine the data and recognize pattern used for regression and classification, it is a machine learning algorithm and is used widely for the classification of product and movie reviews. The SVM algorithm with Bag of Words is presented by Whitelaw et al [25] for the classification of movie reviews. Previous research shows that this classifier outperforms all other classifiers [13]. Kaiquam and Xu [11] used multi class SVM for the classification of sentiments. Pang et al. [17] Experimented SVM, NB and ME for the classification of movie reviews and their finding shows that SVM performed better than all other ML algorithms with bag of words features. Gautami Tripathi et al. [32] investigated the behavior of SVM and Naïve Bayes for feature selection. KNN classifier requires test example for learning. Basically it relies on the category label attached to the training document that is same to the test document. KNN idea is quite simple and effective in many data mining application. The limitation of ML approaches in sentiment analysis is that these all approaches are domain specific and they not work well on domain independent data. So in case of domain independency another approach for sentiment analysis is used namely 'Lexicon based approach'. In lexicon based approach the semantic orientation of document/sentence is evaluated by calculating the sum of score of each opinion bearing words and phrases. As in lexicon each opinionative word has some score and this score suggests whether a given word is positive or negative because the score are assigned according to the nature of word. (E.g. bad = -1 and good = 1). These approaches generally depend on precompiled list of opinionative terms. Lexicon performs well in domain independent sentiment analysis. Turney et al. [22] experiment SA on product reviews using lexicon based approach and produced satisfactory results. The context handling problem is solved efficiently through lexicon based approach. Taboada et al. [26] handled the context using lexicon based sentiment analysis. This approach works well in slang detection problem. Hossam S et al. [23] investigated the public moods from colloquial Arabic text. Harb et al. [30] used features like adjectives and adverbs for the analysis of movie reviews. G.Wei and F. Sebastiani [1] performed prevalence estimation task for tweet sentiments. They performed experiments on eleven tweet sentiment classification datasets through two learners using seven quantification specific algorithms. Aurangzeb et al.[10] proposed a rule based sentiment classifier for the classification of reviews and comments. Their findings show that the proposed classifier achieved 87% accuracy at feedback and 83% at sentence level for comments and 97% at feedback and 86% at sentence level for customer reviews.

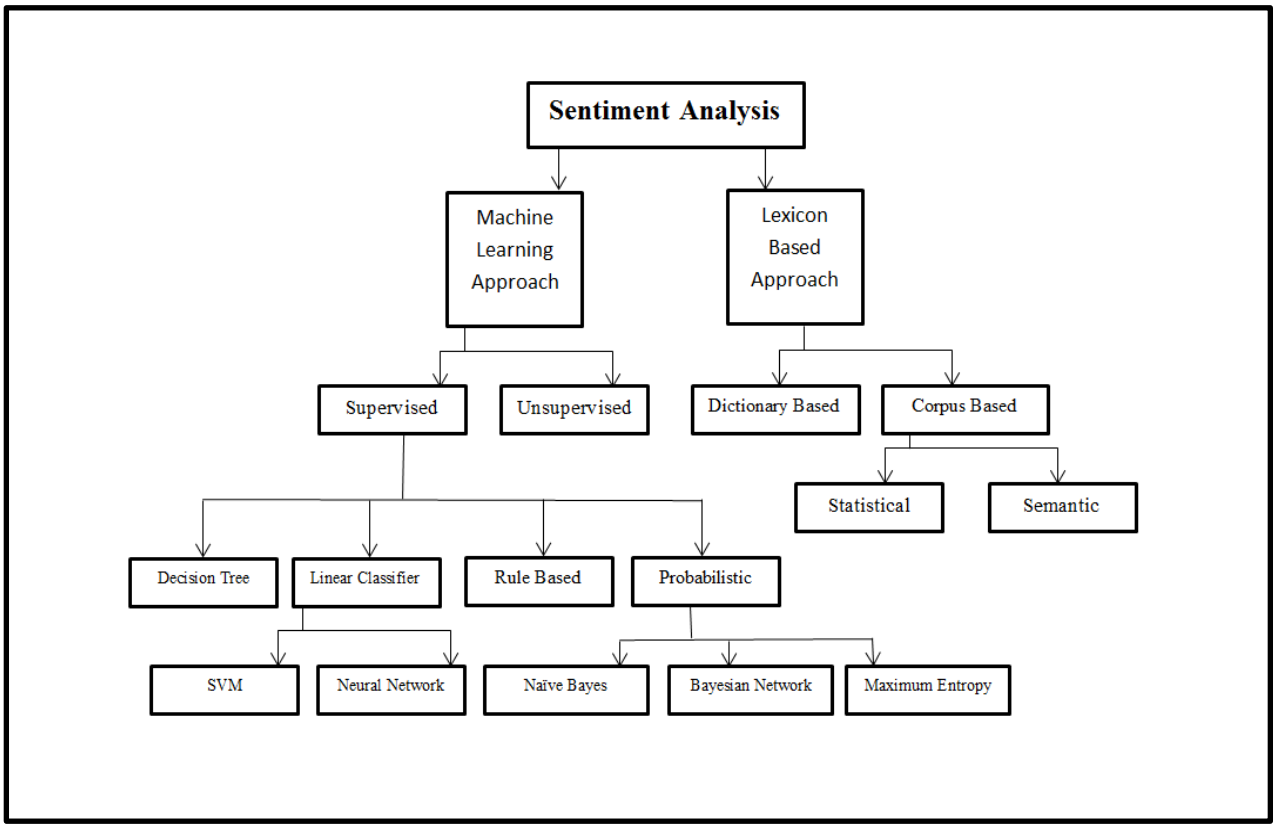

Fig.1. Sentiment Analysis Approaches

3. Data Sources

The proliferation of Web 2.0 gave birth to new field Sentiment analysis in research and academia. In fact this field has gained much attention after the great success of social media as this media today is considered to be an outstanding communication channels for various communities. The blogs, review sites, microblogs and forums are the valuable data sources in sentiment analysis process. These all are the applications of Web 2.0.

3.1. Blogs

Blogs are sites that are used to express one's personal suggestions, opinions and sentiments. The owner of blog namely blogger record the daily activities, gossips and events discussion on their blog. Mostly these blogs contain the discussion about various products, services and events.

3.2. Forums

The online discussion sites in which people from various regions can exchange their views in the form of posted messages are known as forums and sometimes called message board. Forums are used mainly for mining public sentiments about particular entities as these are generally dedicated to a single topic.

3.3. Review Sites

Websites in which the people post reviews about organizations, products, events, services and even other people like stars and celebrities are review sites. Most of the research in sentiment analysis has been done on data sets of review sites as these sites are valuable source for product and movie reviews. Kerstin Denecke [08] performed multi-domain SA on four different product types: Books, DVDs, Kitchen appliances and electronics using amazon product reviews.

3.4. Microblogs

Microblog is a type of blog that allows its users to post short messages. Microblogs are today the well-known communication channels among different communities like education, sports, politics and business. This channel provides full duplex communication, as we can send and receive messages instantly on these sites. Some well-known microbloging sites are: Tumblr, Meetme, Myspace, Plurk and Twitter. Twitter is the most popular microbloging site; it enables its users to publish text up to 140 characters "tweet". Tweets contain opinionative information with some emotion bearing symbols "emoticons" e.g. ☺ happy and ☹ sad etc. Due to the short text restriction of microblogs most of tweets are slangy and noisy in nature so a proper preprocessing is required before text analysis.

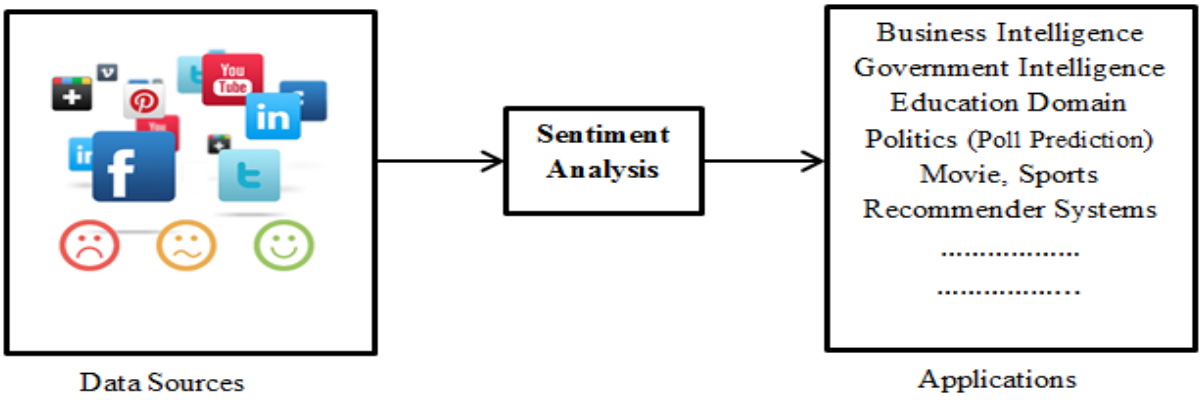

Fig.2. General Model for SA Process

4. Tasks of Sentiment Analysis

Sentiment analysis itself is the task of text mining as it is the sub field of data mining in which the public sentiments and their semantic orientation is calculated. The basic aim of Sentiment analysis or opinion mining is the detection of public sentiments or views about desired entities. Sentiment analysis process comprises of two tasks namely Subjectivity classification and Sentiment classification.

4.1. Subjectivity Classification

Social Media contents consist of objective and subjective information. Extracting useful opinionative information from whole text is called subjectivity classification. ($T_s UT_o = T$). The opinionative sentences are the relevant data in sentiment analysis process while the factual information is irrelevant data. The factual information is known as objective data. Much research has been done on subjectivity classification. Bing Liu [29], Wiebe et al [6, 28], Bo pang et al.[9] experimented subjectivity methods for sentiment analysis.

4.2. Sentiment Classification

Once the subjectivity classification task is completed the subjective text is passed to the next phase for the detection of polarity (whether a given text is positive or negative). Sentiment classification is the process of classifying a text into positive and negative according to their score. The text can be classified into Binary or Multi clauses. Aspect and Feature based sentiment analysis is the sub task of sentiment classification. In feature based sentiment analysis the polarities about the attributes of entities are determined. Here the task is to find whether a specific feature of desired entity is liked or disliked. The attributes or features of an object are considered as aspect e.g. the *picture quality* of camera, *voice* of cell phone and *speed* of car.

5. Levels of Sentiment Analysis

The existing research on Sentiment analysis shows that researchers have worked on three levels of granularities Document level, Sentence level and Entity /Aspect/ Phrase level.

5.1. Document Level

The process of detecting sentiments from whole review about single entity is considered as document level sentiment analysis. A document is the whole review. The advantage of document level sentiment analysis is that in this level of analysis the semantic score of whole document is evaluated whether a given review is expressing the positive opinion or negative about concerned entity. But the difficulty lies when the sentences in a document are irrelevant or stating mixed opinion.

5.2. Sentence Level

The sentence level sentiment analysis is very much similar to document level sentiment analysis here in this level of analysis sentence is considered as basic unit for which the whole sentiment is to be determined. In case of simple sentences this level of classification is applicable but in case of complex sentences this classification level fails. More over some objective sentences also contain opinions that are ignored in sentence level sentiment analysis. Wilson et al. [10] Performed sentiment analysis for strong and weak opinion clauses. Previous research shows that this level of sentiment analysis is better than document level at blogs and forums. Aurangzeb et al.[7] used sentence level lexical based semantic orientation for the classification of online reviews.

5.3. Entity/ Aspect level

The document and word level sentiment analysis only detect whether an entity is liked or disliked but they never provide such fine grained analysis, which attributes were liked or disliked in the given review. In such cases aspect level sentiment analysis is applied as this level of analysis provide the pinpoint information about each aspect of particular entity that were discussed in a single review or text. In past the aspect level was considered as feature level or feature based opinion mining [37]. Mostly this level of sentiment analysis is applicable on product reviews because the opinions about product are based on the specific features or attributes. Generally speaking attributes are considered to be the aspect of an entity. The phases includes in aspect level sentiment analysis are; (i) Aspect extraction (ii) Opinion Prediction (iii) Sentiment classification (iv) Summarization. In word level sentiment analysis mostly adjectives are treated as feature but verb, adverb and nouns are also used as features.

6. Applications of Sentiment Analysis

The sentiment analysis approach is applicable almost in every field of life. The social media is growing rapidly among various communities. Peoples are turning toward this media progressively. Today it is possible to analyze the opinion and behavior of a customer toward a specific product “What the other customers are discussing about that product which they want to buy”. SA applications are endless it is used in the monitoring of social media, tracking customer reviews, competitors , online advertising, email filtering, search engines, business ,education, marketing, politics, government intelligence and survey responses. Individuals and organizations from various domains are taking advantage of Sentiment Analysis systems. Gamon et al [4] experimented a prototype system for mining topic and sentiments from customer reviews. German elections 2009 [3] and US Congressional elections [5] also predicted with amazing accuracy. Elyasir et al. [2] performed sentiment analysis in education domain. They compared various Malaysian Universities through their proposed framework.

7. Conclusion

The growing nature of social media sites has changed the analyses trends of public opinions and sentiments. Every second user utilizes social media to express his feelings and emotions. The views, feelings, speculations and appraisals expressed on social sites are treated as sentiments or opinions. These sentiments are analyzed for making better decisions. Sentiment Analysis or Opinion Mining is the computational study of public moods or sentiments with respect to some topic. In this survey paper we have reviewed Sentiment Analysis tasks, levels, approaches, data sources, applications and challenges. Although this field has become popular among Machine learning, NLP and Data Mining researchers but still there are some problems need to be unfolded. Furthermore, there is need of fully automated and highly efficient system for fine grained analysis. We strongly believe that this survey paper will be beneficial to new researchers in the field of sentiment analysis and opinion mining.

References

[1] Wei Gao, Fabrizio Sebastiani, “From classification to quantification in tweet sentiment analysis”, Social Network Analysis and Mining, Vol 6, no 15 Springer-Verlag Wien, 2016

[02] A. M. H. Elyasir, K. S. M. Anbananthen “Opinion Mining Framework in the Education Domain” International Journal of Social, Human Science and Engineering Vol:7 No:4, 2013

[3] A. Tumasjan, T. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment, " in Proc. of 4th ICWSM. AAAI Press, pp. 178-185, 2010

[4]. Gamon, Michael, Anthony Aue, Simon Corston-Oliver, and Eric Ringger. Pulse: Mining customer opinions from free text. Advances in Intelligent Data Analysis VI: p. 121-132, 2005

[5]. A. Livne, M. Simmons, E. Adar, and L. Adamic, "The party is over here: Structure and content in the 2010 election, " in Proc. of 5th ICWSM, 2011

[6] J.M. Wiebe, “Learning subjective adjectives from corpora,” Proceedings of the National Conference on Artificial Intelligence, pp. 735-741, 2000

[7] Khan, A., “Sentiment Classification by Sentence Level Semantic Orientation using SentiWordNet from Online Reviews and Blogs,” International Journal of Computer Science & Emerging Technologies, vol. 2, no. 4, 2011.

[8]. K. Denecke, Are SentiWordNet scores suited for multi-domain sentiment classification? In Proceedings of the 4th International Conference on Digital Information Management (ICDIM ’09), pages 32–37, Ann Arbor, MI, USA, 2009.

[9] Bo Pang and Lilian Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In ACL-2004, 2004.

[10] Wilson, Theresa, Janyce Wiebe, and Rebecca Hwa. Just how mad are you? Finding strong and weak opinion clauses. In Proceedings of National Conference on Artificial Intelligence (AAAI-2004). 2004.

[11] Kaiquan Xu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, “Mining comparative opinions from customer reviews for Competitive Intelligence”, Decision Support Systems 50 743–754, 2011

[12] Melville, Wojciech Gryc, “Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification”, KDD‟09, June 28–July 1, 2009, Paris, France.Copyright ACM 978-1-60558-495-9/09/06, 2009

[13] Rui Xia , Chengqing Zong, Shoushan Li, “Ensemble of feature sets and classification algorithms for sentiment classification”, Information Sciences 181 1138–1152, 2011

[14] T. Nasukawa, “Sentiment Analysis□: Capturing Favorability Using Natural Language Processing Definition of Sentiment Expressions,” pp. 70–77, 2003.

[15] K. Dave, I. Way, S. Lawrence, and D. M. Pennock, “Mining the Peanut Gallery□: Opinion Extraction and Semantic Classification of Product Reviews,” 2003.

[16]S.M. Kim and E. Hovy, “Automatic detection of opinion bearing words and sentences,” Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), , pp. 61-66, 2005

[17] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, pp. 79-86 , 2002

[18] J.M. Wiebe, “Identifying subjective characters in narrative,” Proceedings of the 13th conference on Computational linguistics-Volume 2, pp. 401-406. , 1990

[19] S. Argamon, K. Bloom, A. Esuli, and F. Sebastiani, “Automatically determining attitude type and force for sentiment analysis,” Human Language

Technology. Challenges of the Information Society, pp. 218-231. , 2009
[20] L. Polanyi and A. Zaenen, “Contextual valence shifters,” Computing attitude and affect in text: Theory and applications, pp. 1-10, 2006
[21] S.M. Kim and E. Hovy, “Determining the sentiment of opinions,” Proceedings of the 20th international conference on Computational Linguistics, pp. 1367-1374, , 2004
[22] P. Turney, “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews,” Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL’02), pp. 417-424, 2002
[23]. Hossam S. Ibrahim , Sherif M. Abdou and Mervat Gheith”Sentiment Analysis for modern standard Arabic and colloquial” International Journal on Natural Language Computing (IJNLC) Vol. 4, No.2,April 2015
[24] Cortes, C. and Vapnik, V. Support vector networks. Machine Learning, 20:273–297. APPLICABLE ALGEBRA IN ENGINEERING COMMUNICATION AND COMPUTING, , 1995
[25] Whitelaw, C., and Patrick, J. "Selecting Systemic Features for Text Classification," in Proceedings of AAAI Fall Symposium on Style and Meaning in Language, Art, and Music, 2004.
[26] Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. Computational Linguistics, 37(2): p. 267-307, , 2011.
[27] Kang Hanhoon, Yoo Seong Joon, Han Dongil. Senti-lexicon and improved Naı¨ve Bayes algorithms for sentiment analysis of restaurant reviews. Expert Syst Appl 2012
[28] Wiebe, Janyce, Rebecca F. Bruce, and Thomas P. O’Hara. Development and use of a gold-standard data set for subjectivity classifications. In Proceedings of the Association for Computational Linguistics, AC L-1999
[29] Hu, Minqing and Bing Liu. Mining and summarizing customer reviews. In Proceedings of ACM
SIGKDD International Conference on Knowledge Discovery and Data Mining), 2004.
[30] A. Harb, M. Planti, M. Roche, “Web Opinion Mining: How to extract opinions from blogs? To cite this version: Web Opinion Mining: How to extract opinions from blogs? Categories and Subject Descriptors.” , 2008
[31] Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. Learning subjective language. Computational Linguistics, 30(3):277–308, 2004
[32]. Gautami Tripathi and Naganna S. “Feature Selection and classification approach for Sentiment Analysis” Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.2, June 2015