Political Sentiment Analysis using Hybrid Approach

Vikash Nandi[1], Suyash Agrawal[2]

[1] M.tech Scholar, Department of Computer Science and Engineering, Rungta College of Engineering and Technology, Chhattisgarh, India, vikashnandi@gmail.com

[2] Reader, Department of Computer Science and Engineering, Rungta College of Engineering and Technology, Chhattisgarh, India, suyash.agrawal1983@gmail.com

---***---

Abstract – *We take various decisions in our daily life every day while shopping, investing money, dealing with others, choosing our favorite superstar, celebrities, electing our ministers and these decisions are being made on the basis of our past life experiences. It is said by someone "who learns from his own mistakes is smart, but who learns from others mistake is genius". Opinion Mining enables us to use others past experiences for taking right decisions. In sentiment analysis we take reviews from social networks and process those reviews in such a manner that we can understand writer's opinion, which will help us in making strategies in future. Mining opinions from natural language is a critical task because it requires a deep knowledge of implicit and explicit, regular and irregular express-ions, and grammatical rules of the language. Our research work is based on politics; Sentiment Analysis can help the political parties in making future strategies, their party promotion and can help in predicting the election results. In elections large amount of efforts, time and money spent. At the time of elections social networks gets flooded with the online discussions about political parties and political celebrities, lots of controversial discussions and debates are held over social networks. All these discussions give us opportunity to use it as a resource for study and analysis.*

In our research we are using twitter as our data source and are applying hybrid approach of sentiment analysis; it combines both the Lexical Dictionary based approach with the features of Support vector machine learning classifier. Lexical approach works on bag of words where test dataset words are matched with preset dictionary words for mining while SVM is a supervised learning classifier which extracts features from test data and on the basis of those features classification is done.

Key Words: **Sentiment Analysis, Opinion Mining, Artificial Intelligence, Machine learning algorithm, Support Vector Machine, Natural language processing, Lexical Orientation, Part of Speech.**

1. INTRODUCTION

Due to the easy availability of web resources in digital form; much of the current research is focusing on the area of Sentiment Analysis and Opinion Mining. Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of opinion, attitude and emotions towards an entity. The two expressions SA and OM are co-related, they express a mutual meaning. However Mikalai and Themis, 2012 [3] et al. stated that OM and SA have slightly different notions. OM analyzes people's opinion for an entity, while SA identifies the sentiment expressed through a piece of text. The aim of SA is to find opinions, identify sentiments, and then classify those sentiments into different categories. An accurate system for predicting sentiments could enable us, to extract opinions from the internet and predict social behavior, political trends, and emerging parties from a particular geographical location. Bakliwal, Foster, van der Puil, O'Brien, Hughes[6] et al. sentiment analysis of political tweets using subjective-lexical-based during the general election 2011 in Ireland. There are various challenges that we face while sentiment analysis; people don't always express their feelings in a same way, a single sentence can be positive for one instances and may be negative for other instance, there are millions of sentence combinations possible, wrong spelling, and handling negations, intensifiers and spams are also challenging. Medhat, Hassan and Korashy, 2014[2] et al. there are three important classification levels in SA: Document-level, Sentence-level, and Aspect-level Sentiment Analysis. The first step in SA is to identify whether the sentence is subjective or objective. Subjective sentence carries sentiments while objective sentence doesn't carry any sentiment it signifies absence of sentiment; usually considered as neutral and also being ignored in the analysis. Koppel and Schler[34] stated that neutral sentences are also important for an effective sentiment analysis. Wilson et al. [26] have pointed out that sentiment expressions are not necessarily subjective in nature. Barhan and Shakhomirov, [24] stated in the definition of opinion that opinions are of two types first is the simple opinion (e.g. 'X' is good) and second one is the comparative opinion (e.g. 'X' is better than 'Y'). Usually texts are classified into three groups of sentiments positive, negative, and neutral. Bakliwal, Foster, van der Puil, O'Brien, Hughes[6] et al. have categorised tweets into 6 types pos(positive), neg(negative), mix(both positive and negative), neu(neutral), nen(non written in English), non(non-relevant or irrelevant). The assigned sentiment will reflect the whole text.

1.1. Text Collection

Data collection is the process of gathering information from a targeted source in a systematic manner which is then processed for outcome. SA is mainly done with the product reviews; these analyses are for manufacturing, marketing companies and business holders to take right decisions on the bases of the user's experiences and requirements. Our SA project is for the Political background and for that we require political based comments; comments related to politics, famous politician, political parties, current political issues, elections etc. These reviews are collected from the microblogging sites. In search of a microblogging website the first name that comes in mind is the Twitter. The success behind the twitter is its' easy accessibility and free-format of messaging service. Twitter provides short messaging service having maximum length of 140 characters and therefore users use lots of abbreviation, #Hashtags, Emoticons to talk more in fewer sentences. Twitter allows getting data for study and research purposes; we can get relevant data by searching text in the search field. To collect data from Twitter we require developers API. Twitter requires user register a developer account and provides authentication details when he queries the API. Once the user account is verified the user will be issued with the authentication detail which allows access to the API.

1.2. Text Pre-processing

Pre-processing enables cleaning of raw data and makes it ready for the classifier. Remove URL links (e.g. http://www.exapmles.com), user names (e.g. @AnupamSharma), remove articles (e.g. 'a', 'an', 'the') form dataset that don't carry any sentiments. Alexander Pak, Patrick Paroubek [22] et al. to increase the efficiency of the classification, we should neglect common n-grams, i.e. n-grams don't strongly indicate any sentiment nor indicate objectivity of a sentence. Kouloumpis, Wilson and Moore, 2011[32] et al. aimed at using Hashtags and Emoticons in the sentiment analysis to improve efficiency.

1.3. Natural Language Processing (NLP)

Natural language processing is the branch of computer science and linguistic computation which is concerned with the analysis of natural language. Using NLP a computer can understand the human language and respond accordingly by using artificial intelligence. The processing of natural language involves statistical learning, learning from corpus. Corpus is a document or a set of documents that help to learn inference rules and find features for language processing. Sentiment Classification can be roughly divided into Machine Learning approach and in Lexical Based approach. The Machine learning approach applies machine learning algorithms, training database and feature selection while Lexical Based approach relies on the statistical lexicons, a collection of known precompiled term database, which is also known as Dictionary-based approach. To improve efficiency Hybrid approach is used; Hybrid approach combines both the above approaches. Rudy Prabowo and Mike Thelwall [33] et al. is a combined approach for sentiment analysis, in this paper authors have combined the rule-based classification, supervised learning and machine learning techniques.

1.3.1. Supervised Machine Learning Techniques

Supervised machine learning techniques, two sets of documents are needed; training dataset and test dataset. Training dataset is available for all kinds of classes in which we want to classify our dataset. With the help of those training datasets a relationship is found which will be used for classification of documents. Researchers have proved that supervised learning technique is better in performance than unsupervised lexicon based method. The most important point in Machine Learning is the feature selection. The collected dataset is used to extract features that will be used by the program to train its classifier. The frequency of word occurrence is more suitable feature, but the overall document's sentiment may not necessarily be indicated through the repeated use of keywords. Pang 2002 et al. have achieved better results by using a term presence rather than term frequency. Erik Cambria and Bebo White, 2014 [30] et al. have drawn a graph of NLP system performance that describes the evolution of NLP research through three different eras of curves from 1950 to 2000, from 2000 to 2050 and beyond 2050. The graph sows that the algorithmic approach is matching the best path condition for language processing.

1.3.2. Lexical Orientation Dictionary Based Approach

Lexical Based approach works on an assumption that the collective polarities of a document are the summation of polarities of the individual words or phrases. The Statistical orientation approach of Sentiment analysis is "unsupervised learning" because it does not require any prior training to classify the data. Instead, it measures how far a given word is inclined towards positive and negative sentiment. Much of the researches in unsupervised sentiment classification make use of LEXICAL resources available online. Kamps et al (2004) focused on the use of lexical relations in sentiment classification. In the dictionary based approach a dictionary of sentiment bearing words are collected along with their polarity score, is used to classify the text into positive, negative and neutral opinion. This word list is grown by searching words in the well known corpora WordNet. Strapparava and Valitutti [28] used WordNet-Affect. Baccianella, Esuli, and Sebastiani [36] et al. used SentiWordNet 3.0 as Lexical source database for analysis.

After creation of dictionary now the processing of NLP starts. At first collected dataset is tokenized; the whole document is chopped into individual words. Words are sent to POS tagger for labialization, here each words is labialized with its grammatical composition. After POS tagging we match words with dictionary entities and if matching conditions is true then that words is assigned with sentimental score (which is assigned to the dictionary words), here the sentiment score of each words of a sentence are added together to get the sentiment polarity of a sentence. This process if repeated for each sentences of whole document. The limitation of such algorithms is the fact that they can process only the information that they can see.

1.4. POS (Part Of Speech)

Part of speech is also known as word classes or lexical categories. The process of classifying words into their parts of speech and label them accordingly is known as part-of-speech tagging, POS-tagging, or simply tagging. In part of speech tagging the tokens are associated with grammatical composition of our natural language. Each word of a language has certain meaning and role in grammatical composition. In POS tagging the tagger defines the meaning and role of each word in grammatical context. Tagger helps in sentiment extraction as it tells the classifiers which part of text will have sentiment and which not. Most POS taggers are trained from NLTK (Natural Language Tool-Kit).

The NLTK is suit of modules, datasets, tutorials and exercises for processing natural language. NLTK is a leading platform for building programs that work with natural language data. It provides an easy to use interface with a suite of text processing libraries for classification, tokenization, stemming and tagging, parsing and semantic reasoning. NLTK is easily available for all major operating systems like Windows, Mac OS, and Linux. Best of all, NLTK is a free, open source project. Steven Bird and Edward Loper, 2004 [29] et al. was about the information of NLTK (Natural Language Tool-Kit).

2. PROBLEM IDENTIFICATION

In this sentiment analysis project data pre-processing is the most important task. A good qualitative data can increase the efficiency while inconsistent and incomplete data can reduce the efficiency and effectiveness of the work.

Spam- Any message, comment, friend request, wall post, profile generated by malicious persons with intent to harm others in financial, mental, social aspects is known as spam. It is usually seen that a lot of spam tweets come along with our actual dataset that we don't want, which spread wrong information. Spams are texts that are not genuine, the objective of spammers is to spread rumour, they write either too much positive or too negative about an issue and send those spams again and again in the social media to achieve their aim. These spams must be detected and removed during the cleaning process. We can also discard tweets with abusive words; this is easy in lexical based approach.

Intensifiers- Some words intensifies the meaning of succeeding words are known as Adverb; an Adverb is a word used to add something to the meaning of a verb, an adjective, another adverb or any other part of speech. In general we use adverbs to intensify the polarity of text; in natural language processing adverbs are also known as 'Intensifiers'. Intensifiers are very challenging for an accurate opinion mining; so those must be taken into consideration while sentiment digging. Suppose a sentence "Modi is a very popular celebrity around the world", if these words "very" and "popular" are taken individually and then we process it then this would be a wrong calculation. Here the word "very" is intensifying the meaning of "popular" hence both must be treated collectively.

Negation- Negation is a condition were a word oppose something; when negate word is attached with any regular sentiment it reverses the meaning of that regular word. Generally we use negation to oppose the succeeding term. Negation is conveyed by common negation words (e.g. not, neither, nor). For example; the word 'good' expresses a positive attitude whereas word 'not good' is clearly a negative word whereas if you will treat both words separately then both words will neutralize themselves. Negation is a very common language construction that affects the polarity of succeeding words (as it reverses the polarity) and therefore, needs to be taken into consideration. Research in the field text mining has shown that there are various other words that invert the polarity of an opinion expressed, such as valence shifters, connectives or models. "I find the performance of the new laptop is less practical", is an example for valence shifter, "market says it is a great mobile, but I fail to see why", shows the effect of connectives. An example of using model is, "In theory, the mobile should have worked even under water also". From these examples we can understand that, negation is difficult yet important aspect of sentiment analysis.

3. METHODOLOGY

Sentiment Analysis is a field of study which analyses people's opinion, attitude and perception towards any entity; entities may be products, personalities, issues etc. SA is the process of determining whether a piece of text is positive or negative. In recent years, it's been a hot topic in both academia and industry. Thanks to the massive popularity of social media which provide a constant availability of textual data. It helps decision makers in many ways like they can get exact public response without conducting any market campaigning, surveys. It will help a political party to know his party image in public. The figure below shows the flow of control in the Lexical approach; the flow starts from microblogging website; which is the source of our text dataset; we collect text reviews form website and store those review for future processing. In the cleaning part we take data from database and remove unwanted contents from it and make it ready for Natural Language Processing; where our text document get classified into sentiments of different categories like positive, negative and neutral.

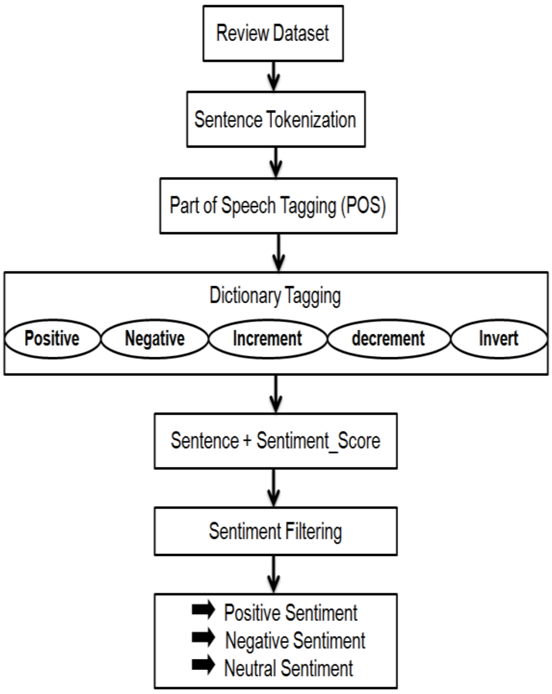

Fig-1. Lexical orientation- Dictionary based approach

Dictionary Tagging:

Dictionary is the heart of Lexical based approach; it holds sentiment words that will be matched with test dataset. Dictionary has two fields one is the word and other is sentiment score that defines the level of polarity. Usually 2 dictionaries are created by researchers for positive and negative words. We have created 3 more; the need for creating 3 extra dictionaries is to solve the conditions of intensifiers and negation condition. Each dictionary is stored separately in file system. We have collected more than 3500 words for our dictionary. Polarity score is number from 1 to 5 for positive words and from -1 to -5 for negative words describes the level of polarity. If a word is highly intense and polar, showing strong sentiment then accordingly a number closer to 5 would be assigned to that word and if a word is less intense having weak sentimental meaning then accordingly a number closer to 1 is assigned to that. Sentiment score 0 signifies that word is neutral (having no sentiment). We have created 5 separate dictionaries; each file is created for one type of words. Positive dictionary having positive words, Negative dictionary has negative words, Increment dictionary has adverbs which increment the positivity (Increment words will doubles the score of succeeding word by multiplying score with 2.0, Decrement dictionary has adverbs which increment the negativity or decrement Positivity (Decrement words will make half the score of succeeding word by dividing score by 2.0.) and Invert dictionary has negate words which invert the meaning of succeeding word (In the case of negation the program will multiply the succeeding term with (-1) to invert the score.)

Filtering:

Filtering is done after getting the sentences with their sentimental score we will filter-out those sentences into four categories as Positive, Negative, Neutral and Spam.Then those sentences are stored into four different files, files names are according to the type of sentences stored into it. Those files will be required for the final sentiment calculation and will be served as our training dataset for new test data. Sentences with sentiment score more than zero are positive sentences, less than zero are negative sentences, score exactly equal to zero are neutral sentences, Sentences with sentiment score less than -20 or greater than 20 are taken as spams. The spams are the sentences which are either too positive or too negativity and those are spread for the purpose of strongly support or strongly oppose any issue, the objective of spammers is to spread rumour in the social media.

Sentiment Calculation:

Sentiment calculation is done for each tweet, first the tweets are tokenized and each token (tokenized word) is matched with dictionary words and on true conditions, appropriate sentiment score is assigned to the tokens (from dictionary) and all such scores are summed together for each tweet. Hence we will have tweets with their scores, at last all tweets scores are added to get the final score of the document which will give the result whether the document is positive or negative. Counters are used to count each type of tweet, counters are incremented by one when preset condition is satisfied; this will give us numbers of each type of tweet. To represent sentiments in an effective way we can calculate percentage polarity. From the calculations we can determine the sentiment of each sentence as well as overall sentiment of the document.

$$\text{Positive Sentiment }\% = \frac{\text{Number of Positive Tweets}}{\text{Total Number of Tweets} - \text{number of Spam Tweets}} * 100\%$$

$$\text{Negative Sentiment }\% = \frac{\text{Number of Negative Tweets}}{\text{Total Number of Tweets} - \text{number of Spam Tweets}} * 100\%$$

$$\text{Neutral Sentiment }\% = \frac{\text{Number of Neutral Tweets}}{\text{Total Number of Tweets} - \text{number of Spam Tweets}} * 100\%$$

Machine Learning Feature Extraction with SVM:

In machine learning classification, items are represented by their features. In our project, documents are represented by their words, so we are using words as features. We have used *'vectorizer'* to translate the input text documents into vectors of features. We want to give appropriate feature weights to different words, and TF-IDF is one of the most used weighting schemes used in text analysis applications. We convert the test data into a matrix of TF-IDF (Term Frequency Inverse Document Frequency) features; it counts the occurrences of features. In vectorization we will discard those words which occur very rarely and also words with very high occurrence, like 'is, the, are' are stop words which are frequently used in any document. They don't have any sentimental value; so these words could be discarded. Words with rarely occurrence in test data will also be neglected. At first we will create the vocabulary list (i.e. the list of words/features) and the feature weights from the training dataset. Second, we simply apply those features on the test dataset using the same vocabulary as the training data. Training and test datasets are stored into two different set of folders one consists of positive words and other consisting negative words, each folders is having 90% training data and 10% test data, this method is called cross-validation.

Support Vector Machine:

SVM is a discriminative classifier considered the best text classification method. The support vector machine is a statistical classification method proposed by Vapnik. An SVM classifies data by finding the best hyperplane that separates all the data points of one class from other class. The best hyperplane for an SVM means the one plane with the maximum possible margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. Sometimes we find situations where a linear hyperplane is not applicable to classify different classes because data points of different classes are closely overlapped with each other; in this situation a non-linear classification hyperplane is the solution. SVM have achieved a great success in text categorization. In sentiment analysis SVM is widely used by many researchers the reason is that text data are ideally suited for SVM because of the sparse nature of text. Multiple variants of SVM have been developed in which Multi class SVM is used for Sentiment classification (Kaiquan Xu, 2011). SVM performs better when feature space in increased Vohra and Teraiya, 2013 [8].

As explained in Pang et al. (2002) given a category set, C = {+1, -1} and two pre-classified training sets, a positive set, $T_r^+ = \sum_{i=1}^{n}(di, +1)$ and a negative sample set $T_r^- = \sum_{i=1}^{n}(di, -1)$. SVM finds a hyperplane in between these two sets that separates both with a maximum margin (or the largest possible distance from both sets). At processing step each training sample is converted into a real vector, x_i that represents features of associated document, d_i. Hence $T_r^+ = \sum_{i=1}^{n}(xi, +1)$ represents a positive sample set and $T_r^- = \sum_{i=1}^{n}(di, -1)$ represents a negative sample set. In an ideal condition we expect to find a clear separation between positive and negative sample sets, in this regard that features found in one class do not appear in another class, or the features position do not cross over each others' hyperplane. One document may contain features that appear in both classes; this would result a wrong vector side. To overcome this problem SVM allows multi-class classification.

Linear Support Vector Classification:

This linear SVC is implemented by liblinear class; it has more flexibility and scale better in large numbers of samples. This class supports both dense and sparse input and the multiclass support is handled according to one-versus-the-rest scheme. The objective of Linear SVC is to fit to the data you provide, returns a best fit hyperplane that categorize your data. After getting the hyperplane you can then feed some features to your classifier to see what the "predicted" class is.

4. RESULT AND DISCUSSION

In our SA project we have used hybrid approach; a combination of Lexical Dictionary based approach with the Machine Learning feature extraction approach. By combining these two approaches we have achieved a high efficiency. At first we proceed with review dataset and applied Lexical Dictionary based classification, from which we get classified sentiment output in three sentiments as positive, negative and neutral. Output of lexical approach is shown in the table below.

Table1. Output of Lexical approach

Approach	Total Tweets	Pos	Neg	Neu	Spam
Lexical Approach	295	122	81	92	0
	Percent (%)	41.35	27.45	31.18	0

This table shows that by applying sentiment extraction on a file having 295 tweets using lexical approach we get 122 positive tweets, 81 negative tweets, 92 neutral tweets and 0 spams. From these we have calculated percentage involvement of each type, which will provide the sentiment of the document. We have stored those classified results in different files which will be treated as input for the next machine learning approach. Neutral sentiments have been ignored as this type of files doesn't have any sentiments in it. Positive and negative files are used as training samples for new test data which have formed the base for feature extraction by SVM. When we apply SVM on our test data we get result as shown in the table 1.2 below.

Table2. Accuracy of SVM in precision, recall and f1-score measure

Classifier Type	Training Time	Prediction Time	Sentiment	Precision	Recall	F1-score
SVC (Kernel=linear)	9.9734	0.9816	Negative	0.90	0.92	0.91
			Positive	0.92	0.90	0.91
			avg/total	0.91	0.91	0.91
LinearSVC	0.1157	0.0004	Negative	0.91	0.94	0.93
			Positive	0.94	0.91	0.93
			avg/total	0.93	0.93	0.93

The results indicate that SVM with LinearSVC improves the classification effectiveness. From the table above we can see the result of SVM; the average result shows an efficiency of 91 and 93% which is tremendous for Sentiment Analysis project. The Linear Support Vector Classifier perform better than the SVC with Kernel=linear and also the time taken by Linear SVC is fractional as compare to the other one.

5. CONCLUSION

This paper has illustrated an effective sentiment analysis that is performed on political reviews of population collected from Twitter. In this project we have presented Lexical approach and Feature based approach. Used support vector machine as sentiment classifier that classifies test data according to related sentiments, as positive, negative and neutral. Such an analysis could provide valuable views that can be effectively used by political parties and leaders for their party promotion and this can give a significant lead in promotion. With this study we can conclude which party is having the maximum number of follower and what are public opinions towards any particular party. Throughout the duration of this project many different data analysis tools were employed to collect, clean and mine sentiment from the dataset.

It is also found that lexical and features based classification algorithms are combined in an efficient way in order to overcome their individual drawbacks and get benefits from each other's merits, and finally enhance the sentiment classification performance. With the use of hybrid approach we have achieved an accuracy of 93% which is good enough for a sentiment analysis system. The main challenging aspects exist in dealing with negation and intensifier expressions; processing negation plays an important role in sentiment analysis. Many previous studies adopted a simple technique that to take negation words as simple negative words, which will produce a wrong output. We have stored negation words in separate database i.e. in Invert database and on getting such conditions we reverse the sentimental polarity of the sentence. In the same manner we have also satisfied intensifiers by adding or subtracting the sentiment polarity score of text.

As the future work, we plan to adapt our sentiment analysis system to languages other than English, collect corpus data and compare the characteristics of the corpus across different languages. We plan to use the obtained data to build a multilingual sentiment classifier and will also try to make the spam filtering method more affective by applying an early spam detection technique so that we will get only valid information.

6. REFERENCE

[1] Bing Liu, "*Sentiment Analysis and Opinion Mining*", Morgan and Claypool Publishers, May 2012.

[2] Walaa Medhat, Ahmed Hassan and Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Shams Engineering Journal (2014) 5, 1093–1113

[3] Tsytsarau Mikalai, Palpanas Themis, "Survey on mining subjective data on the web". Data Min Knowl Discov 2012; 24:478–514

[4] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, vol.2, No1-2(2008)1-135

[5] Bo Pang and Lillian Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", Proceedings of ACL

[6] Akshat Bakliwal, Jennifer Foster, Jennifer van der Puil, Ron O'Brien, Lamia Tounsi and Mark Hughes, "Sentiment Analysis of Political Tweets: Towards an Accurate Classifier", LASM 2013, Atlanta, Georgia, June 13 2013

[7] De Wang, Danesh Irani, and Calton Pu, "A Social-Spam Detection Framework", Anti-Abuse and Spam Conference Sept 1-2, 2011, Perth, Australia Copyright 2011 ACM 978-1-4503-0788-8/11/0

[8] S.M. Vohra and J.B. Teraiya, "A Comparative Study of Sentiment Analysis Techniques", ISSN: 0975–6760 nov 12 To oct 13 | volume – 02, issue – 02

[9] Kaiquan Xu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, "Mining comparative opinions from customer reviews For Competitive Intelligence", Decision Support Systems 50 (2011)

[10] Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, "Using wordnet to measure semantic orientation of Adjectives", 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal, 2004

[11] Ting-Chun Peng and Chia-Chun Shih , "An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs", IEEE/WIC/ACM august 2010, ISSN 2151-9617

[12] Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," Computational Intelligence, vol. 22, pp. 110–125, 2006

[13] Polanyi and A. Zaenen, "Contextual lexical valence shifters", AAAI technical report SS-04-07, 2004

[14] Liu, H. (2004). MontyLingua: "An end-to-end natural language processorwith common sense"

[15] G. Adda, J. Mariani, J. Lecomte, P. Paroubek, and M. Rajman, 1998, "The GRACE French part-of-speech tagging evaluation task"

[16] Abhishek Bhola, "Twitter and Polls: Analyzing and estimating political orientation of Twitter users in India General#Elections2014", arXiv:1406.5059v1[cs.SI] Jun14.

[17] Ethem Alpaydin. 2004. "Introduction to Machine Learning (Adaptive Computation andMachine Learning)". The MIT Press

[18] Hayter Anthony J. 2007, "*Probability and Statistics for Engineers and Scientists". Duxbury, Belmont, CA, USA*

[19] Alec Go, Lei Huang, and Richa Bhayani, 2009, "Twitter sentiment analysis", Final Projects from CS224N for Spring 2008/2009 at the Stanford Natural Language Processing Group

[20] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury, 2009, "Micro-blogging as online word of mouth branding", In *CHI EA '09 New* York, NY, USA, ACM

[21] John Blitzer, Mark Dredze, Fernando Pereira, "Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification", Department of Computer and Information Science University of Pennsylvania

[22] Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", Universit´e de Paris-Sud, Laboratoire LIMSI-CNRS, Bˆatiment 508,F-91405 Orsay Cedex, France

[23] Namrata Godbole, Manjunath Srinivasaiah,Steven Skiena, "LargeScale Sentiment Analysis for News and Blogs". Stony Brook University, Stony Brook, ICWSM'2007 Boulder, NY 11794-4400, USA

[24] Anton Barhan, Andrey Shakhomirov. "Methods for Sentiment Analysis of twitter messages". Saint-Peterburg, State University of Aerospace Instrumentation, Saint-peterburg, Russia

[25] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". In *ICML '01: pages 282–289, San*

[26] Wilson T, Wiebe J, Hoffman P, "Recognizing contextual polarity in phrase-level sentiment analysis", In: Proceedings of HLT/ EMNLP, 2005

[27] Liu B. "Sentiment analysis and opinion mining". Synth Lect Human Lang Technol, 2012

[28] Carlo Strapparava and Alessandro Valitutti, "WordNet-Affect: an Affective Extension of WordNet"

[29] Edward Loper and Steven Bird, "NLTK: The Natural Language Toolkit". arXiv:cs/0205028v1 [cs.CL] 17 May 2002

[30] Erik Cambria and Bebo White, "Jumping NLP Curves: A Review of Natural language Processing Research", IEEE computational intelligence magazine |1556-603X, MAY 2014

[31] Hatzivassiloglou, V., McKeown, K.R, "Predicting the semantic orientation of adjectives", NJ, USA, Association for Computational Linguistics (1997)

[32] Jalaj S. Modha, Gayatri S. Pandit and Sandip J. Modha, "Automatic Sentiment Analysis for Unstructured Data", 2013, IJARCSSE ISSN: 2277 128X

[33] Rudy Prabowo and Mike Thelwall, "Sentiment Analysis: A Combined Approach"

[34] Moshe Koppel and Jonathan Schler, "The importance of Neutral Examples for Learning Sentiments"

[35] Jonghyuk Song, Sangho Lee and Jong Kim, "Spam Filtering in Twitter using Sender-Receiver Relationship"

[36] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani, "SENTIWORDNET 3.0: An Enhanced Lexical Resource.