Sentiment Analysis: A Review

M. Edison¹, A. Flora², D. Vinnarasi³, A. Aloysius⁴

¹Research Scholar in Computer Science St. Joseph’s College (Autonomous), Trichy
edi.muthu01@gmail.com

²Research Scholar in Computer Science St. Joseph’s College (Autonomous), Trichy
florafrancis24@gmail.com

³Research Scholar in Computer Science St. Joseph’s College (Autonomous), Trichy
vinnarasiti@gmail.com

⁴Assistant Professor in Computer Science St. Joseph’s College (Autonomous), Trichy
Aloysius1972@gmail.com

Received 16 June 2017; Accepted 10 July. 2017

Abstract
Sentiment Analysis be situated a big domain to study people’s opinion. Nowadays people are expressing their emotions differently by social media network. Most of the users are posting their emotions with proper and improper text. In this paper, review of the sentiment analysis concepts, methods and mechanisms for solving the issues discussed with different aspects. Especially, pre-processing, mathematical and statistical mechanism are used in lexicon based approach for refereed papers, which was handled in sentiment analysis. Most of the research articles in sentiment analysis used pre-processing algorithm and methods to predict the results with a help of statistical analysis. Especially, lexicon and machine learning approaches are used by the researchers in this domain, which was assisted for classification and measured polarity of the sentiments.

Key Words: Sentiment Analysis, Sentiment Classification, Lexicon Based Approach, Social Media Network, Opinion Mining.

1. Introduction
Nowadays many people are expressing their personal interest, feeling and opinion through social media network (SMN) such as twitter and facebook etc. Twitter is one of the most popular SMN in this world. People can express their emotions in a textual formation in the SMN. Especially, Twitter is allowed 140 characters limitation to post comments by user. Each and every utterance has a distinct meaning. Mainly, people express their feeling in short form of text such as smiles called as emoticons and acronyms [1], which are complex issues for understanding. Whereas, the data are an ambiguous so which are very difficult to classify in different classes with different levels. Before that the ambiguous data must be cleaned while the cleaned data gives original data, which are helped to determine the polarity like positive, negative and neutral [2]. Variety of analysis gives univariate and multivariate method for selecting the features. The sentiment classifier method to classify the data in different aspects. Sentiment classification could be done by researchers with variety of approaches and levels. The approaches are lexicon based approach and machine learning approaches, which are best approach in sentiment analysis (SA).

2. LITERATURE REVIEW:
Sentiment analysis is to recognize and express emotions digitally which presents in the lexicon based approach for sentiment classification to detect the scores and the slags used in tweet [1]. Shuigui Haang et al [2] have proposed Sentiment model which represents based on emoticons classification. It’s built orientation model of sentiment words with the orientation of emoticons, and train the model with SVM classifier. It was high efficient way to automatically classify the orientation of emoticon. Seongik park et al [3] proposed a method to build a...
thesaurus lexicon using dictionary-based approach for the sentiment classification. It was used to increase availability of posts and used to increase accuracy of the sentiment classification. Stefano Baccianella et al [4] proposed SENTIWORDNET 3.0 lexical resource explicitly devised for supporting sentiment classification and opinion mining applications. It was evaluating SENTIWORDNET 3.0 against framework of WORDNET 3.0 manually annotated for positivity, negativity and neutrality and the result indicate accuracy improvement. Fei Jiang et al [5] suggested about emoticons have been widely employed to express different types of moods, emotions and feelings in microblog environments. In existing work used many emoticons that convey clear emotional meanings as noisy sentiment labels. The proposed work of emoticon space model (ESM) that leverage more emoticons to construct word representations from huge amount of unlabeled data. It has identified subjectivity, polarity and emotions in microblog environment. LU Xing et al [6] proposed machine learning method to determine the sentiment polarity of short Chinese texts, which is represented result of experiments show the progress in the accuracy. Monisha kanakaraj et al [7] have submitted NLP based approach to enhance the sentiment classification by adding semantics in feature vectors and thereby using ensemble methods of classification.

3. SENTIMENT ANALYSIS

Definition

“It is the computational study of people’s opinions, appraisals and emotions toward entities, events and their attributes. Opinions are important because whenever we need to make decision – we listen to other’s opinions” [8].

Sentiment Analysis refers to the use of natural language processing, text analysis and computational linguistics to process the data and determine the data what it has been represented. SA is widely applied for reviews of the customer materials such as reviews, survey responses, social media and health care materials.

SA is classifying the polarity of a given text at the sentence, document or feature/aspect level these levels are expressed opinion in a sentence level, a document or an entity feature/aspect level with different classes like: positive, negative or neutral. Beyond that sentiment classification for scaling system is commonly associated with sentiment classes.

Sentence Level

Sentence level approach goes to the sentences to determine whether the sentence communicated as positive, negative, or neutral opinion. Usually neutral means no opinion. This level of analysis is closely related to subjectivity classification which decides sentences called objective sentences. The factual information from sentences called subjective sentences that prompt subjective views and opinions. However, the subjectivity is not equivalent to sentiment as many objective sentences can imply opinions [8].

Document Level

The document level approach is to classify whether the whole document expresses as positive or negative sentiment. The system concludes whether the review states an overall positive or negative opinion about the product. This task is commonly Sentiment Analysis and Opinion Mining known as document-level sentiment classification. This level of analysis assumes that each document expresses opinions on a single entity. Thus, it is not appropriate for documents which appraise or compare numerous entities [8].

Aspect Based Level

The aspect based level approach carries analysis of a particular product feature. It does not disclose what exactly people like or dislike but it makes people realize the opinion about the product entity. Aspect level performs fine-grained analysis is also called feature level. This approach directly expresses the sentiment which consists of positive, negative and neutral. For example, the sentence “The iPhone’s call quality is good, but its battery life is short” clearly has a positive tone for the first part, the second part is having a negative opinion. It cannot be assumed that the sentence is entirely positive or entirely negative. This might be evaluated in two aspects, how the aspect level is working, qualitative and quantitative analysis [8].

4. LEXICON BASED APPROACH

Semantic orientation is a measurement of subjective and objectivity opinion. It usually captures the evaluative factor and potency or strength towards a subject topic, person, or idea. There are two main approaches are exist the problem for extracting
sentiments automatically. The lexical-based approach involves the calculating orientation of words or phrase in the document. The text classification approach involves the building classifier from labeled instance of texts or sentences [9].

Dictionaries for lexical-based approaches could be annotated manually, as we describe automatically using seed words to expand the list of words. Most of the lexicon-based research has focused on using adjectives as indicators of the statistical text classification which builds support vector machine classifier trained on a particular data set using features such as unigrams or bigrams, and with or without part-of-speech labels. The polarity of these words actually makes some sense in context sequels.

Microblogging service data is a sample source for opinion mining scopes. Twitter and other microblogging services act as a platform for marketing and make social relations with people. There are two major approaches used by the researchers to extract sentiment automatically i.e. Lexicon-based and text classification approach. In the first case documents so is calculated from the words or phrase.

The second case classifiers are annotated instances of text sentences which is described as a machine learning approach. Lexicon-based scoring of tweets using the R language is working based on simple average sentiment score [10]. Major of the work in sentiment analysis is based on binary classification i.e. dividing reviews or blogs into “positive” and “negative” classes. This approach is integrate different lexicons and dictionary of tweets.

In this paper, presents lexicon based review for sentiment classification. The sentiment classification to detect score for the particular slangs which are extracted from tweets. Text based message (tweets) up to 140 characters are allowed to post in twitter, which are constraint due to short text, slang, abbreviation and poor structure of sentences. So, these kind of data/text/tweets are very difficult to analysis. Two major approaches are used to analyse the data in Semantic orientation. Subjectivity text identification extracting opinionated word from the source text and detect the slang transaction in each classes. The sentiment score has been assigned according to their classes as +1, -1 then the sentiment scores are measured by using SentiWordNet (SWN) [11, 12]. Lexicon based methods provide efficiency by manually developed word affective word and valence value. The examples of the acronyms and emoticons tables are shown below in Table. 1 and 2.

<table>
<thead>
<tr>
<th>ACRONYMS</th>
<th>DICTIONARY LEXICON</th>
</tr>
</thead>
<tbody>
<tr>
<td>gr8</td>
<td>great</td>
</tr>
<tr>
<td>gud</td>
<td>Good</td>
</tr>
<tr>
<td>Wel</td>
<td>Well</td>
</tr>
<tr>
<td>5n</td>
<td>Fine</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMOTICONS</th>
<th>DICTIONARY LEXICON</th>
</tr>
</thead>
<tbody>
<tr>
<td>:)</td>
<td>Happy</td>
</tr>
<tr>
<td>(:</td>
<td>Sad</td>
</tr>
<tr>
<td>:-)</td>
<td>Joy</td>
</tr>
<tr>
<td>(:-)</td>
<td>sorrow</td>
</tr>
</tbody>
</table>

5. SENTIMENT BUILDING
Authors are using many tricks and methods for improving the text classification to determine the polarity. Mainly, people using emotional icon in text to express their feelings differently. The emoticons and acronyms are biggest issue for the sentiment classification. The lexicon based approach is helped to build a sentiment building and helped to evaluate the sentiment with a classes and score [13].

Nowadays the users are used new modification model to overcome the drawback and allows the optimal cost ratio via language evolution. In statement analysis plays many important role for the orientation of statement analysis. Emoticons are widely used in the social network for good interaction and visualization [14, 15]. This is the effective way for the oriented the emoticons. People use social media like facebook, twitter, youtube, and so on. NLP
named as Natural Language Processing based approach to adding additional semantic. It helps in various fields as prediction and analysis and so on. Natural Language Processing goal is to finding the semantic meaning of the text context and analyzing the text information. The comparative result is illustrated below in Table. 3.

### Table 3: Comparative Performance Results of the Refereed Papers

<table>
<thead>
<tr>
<th>S.NO</th>
<th>AUTHORS</th>
<th>PAPER &amp; YEAR</th>
<th>ACCURACY (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alexander et.al</td>
<td>Exploiting Emoticons in Sentiment Analysis – 2013 On a sentence level</td>
<td>59.5%</td>
</tr>
<tr>
<td>2</td>
<td>Fazal et.al</td>
<td>Lexicon-Based Sentiment Analysis in the Social Web – 2014 On multi-class classification</td>
<td>87%</td>
</tr>
<tr>
<td>3</td>
<td>Zhenhua et.al</td>
<td>An Unsupervised Method for Short-Text Sentiment Analysis Based on Analysis of Massive Data – 2015 On normal text</td>
<td>58.4%</td>
</tr>
<tr>
<td>4</td>
<td>Hussam et.al</td>
<td>Lsislf: Feature Extraction and Label Weighting for Sentiment Analysis in Twitter – 2015 On system rank third participants</td>
<td>64.27%</td>
</tr>
<tr>
<td>5</td>
<td>Saprativa et.al</td>
<td>Sentiment Analysis using Cosine Similarity Measure – 2015 On a 5 class sentiment classification</td>
<td>68.46%</td>
</tr>
<tr>
<td>6</td>
<td>Ayushi et.al</td>
<td>IIIT-H at SemEval 2015: Twitter Sentiment Analysis The good, the bad and the neutral – 2015 On improved acronyms and emoticons</td>
<td>59.83 and 67.04%</td>
</tr>
<tr>
<td>7</td>
<td>Xiao SUN et.al</td>
<td>Hybrid Model Based Sentiment Classification of Chinese Micro-Blog</td>
<td>63.7%</td>
</tr>
<tr>
<td>8</td>
<td>FazalMasudKundi et.al</td>
<td>Lexicon-Based Sentiment Analysis in the Social Web</td>
<td>92%</td>
</tr>
<tr>
<td>9</td>
<td>Emma Haddi et.al</td>
<td>Role of Text Pre-Processing in Sentiment Analysis</td>
<td>93.5%</td>
</tr>
</tbody>
</table>

### 6. CONCLUSION

In this work discussed about the sentiment analysis or opinion mining to cover techniques and approaches that promise to directly enable opining mining information seeking system. Most human would be able to quickly interpret that the person was being sarcastic. By applying this extract contextual understanding to the sentence, we can easily identity the sentiment polarity. Especially with a help of the techniques to be improved the reliability thesaurus lexicon. These applications and techniques are aided to apply in the sentiment analysis and get the powerful reflection for it. The techniques are extracted from mathematical and statistical model which helps to practice quickly understand and decision making for the analysis.

### REFERENCES:


AUTHOR PROFILE

M. Edison is a research scholar in the department of Computer Science, St. Joseph’s College (Autonomous), Tiruchirappalli, Tamil Nadu, India. He is doing his Doctor of Philosophy in the area of Big Data. He has published research articles in his research area and he has attended many workshops, conferences and he has acted as a resource person for national workshops.

A. Flora is a research scholar in the department of Computer Science, St. Joseph’s College (Autonomous), Tiruchirappalli, Tamil Nadu, India. She is doing her Master of Philosophy in the area of Big Data. She has published research articles in her research area and she has attended many workshops and conferences.

D. Vinnarasi is a research scholar in the department of Computer Science, St. Joseph’s College (Autonomous), Tiruchirappalli, Tamil Nadu, India. She is doing her Master of Philosophy in the area of Big Data. She has published research articles in her research area and she has attended many workshops and conferences.

Aloysius is working as an Assistant Professor in the Department of Computer Science, St. Joseph’s College (Autonomous), Tiruchirappalli, Tamil Nadu, India. He has 16 years of experience in teaching and research. He has published many research articles in the National / International conferences and journals. He has acted as a chairperson for many national and international conferences. Currently, eight candidates are pursuing Doctor of Philosophy Programmed under his guidance.