ISSN No:-2456-2165

Sentiment Analysis Approach for Opinion Estimation from Text: A Review

Nikita Dipak Bachhav
Department of Computer Engineering
Pillai College of Engineering
New Panvel, Navi Mumbai, India

Prashant Nitnaware
Department of Computer Engineering
Pillai College of Engineering
New Panvel, Navi Mumbai, India

Abstract:- Sentiment Analysis is a language processing technique that categorizes textual expressions into positive, negative or neutral. This paper provides overview of sentiment analysis and its powerful algorithms to extract the actual emotions from the given text, along with the applications and its important tools. This paper also represents comparative study of all algorithms and their application on social media platforms.

Keywords:- *Sentiment Analysis; Polarity; Machine Learning.*

I. INTRODUCTION

With persistently increasing online media usage and feedback system, every sector or organization gets comments and reviews in the form of text. As text is very powerful way to express opinion, to extract its real emotion though automation, sentiment analysis provides various algorithms. These opinions provide insights about user's experience and attitude towards products of the company and possible future interests. Sentiment analysis has some important notations as follows: (1) Subjective and Objective - Subjective: "Detective Pikachu is wonderful movie", the sentiment in this sentence is 'wonderful', and therefore this sentence is subjective. Now, check another one i.e. objective: "Ryan Reynolds is a voice over artist of Detective Pikachu" this sentence does not has any sentiment, this is representing a fact, and thus this sentence will be categorized as objective. (2) Polarities - Positive: "I love new dress", it's showing positive emotion. Negative: "The party, we attended last night was really horrible", showing negative emotion. Neutral: "I usually get tired in evening" it's a fact, without any specific emotion. (3) Sentiment level - Document: In this, single polarity is given i.e. either positive, negative or neutral to the entire document. Sentence: In this level, classification of the document is done based on every sentence. Evaluation of each sentence is done distinctly and then it is categorized as negative, positive or neutral. Thus, a single document has many sentences and each sentence has its individual polarity. Phrase: This level deals with the fragmentation of document into sentence and then defragmentation of each sentence into phrases. After that assigning the polarities to every phrase as positive, negative or neutral. This technique has also been named as aspect based analysis [1].

II. SENTIMENT ANALYSIS PROCEDURE

Process of sentiment analysis comprises of mainly three important phases, as shown in figure 1.

A. Data collection

Sentiment analysis can be done on any text data. Data can be gathered and extracted from available website or any other available source [2]. Only subjective data will be collected for processing and objective data will be eliminated.

B. Pre-processing

Data may contain some useless features and therefore cleaning and filtering is required before processing. This step involves, removing URL, filtering, questions, removal of special character and stop words, reduction or lemmatization, tokenization (in this step a complete string is divided into multiple sections such as symbols, keywords, phrases, words, and other features called tokens) and feature selection.

C. Classification

Classification is used in sentiment analysis for classification of data that falls under 3 categories i.e. either positive, negative or neutral. Two highly used approaches are, (1) Subjective lexicon, and (2) Machine learning approach. Many sub-approaches / algorithms falls under these two main approaches, shown in figure 2. Sentiment analysis helps to properly judge the feedbacks and tackle the choice-overload problem [3]. The results can be used in recommendation systems to offer better services to the customers [4].

III. APPLICATIONS OF SENTIMENT ANALYSIS

A. Educational sectors

Educational sectors are always need to maintain or achieve higher educational standard as well as student gratification as per government or regulating bodies. Online Feedback System based on robust sentiment analysis technique provides a means for measurement of level of satisfaction in students and lecturers.

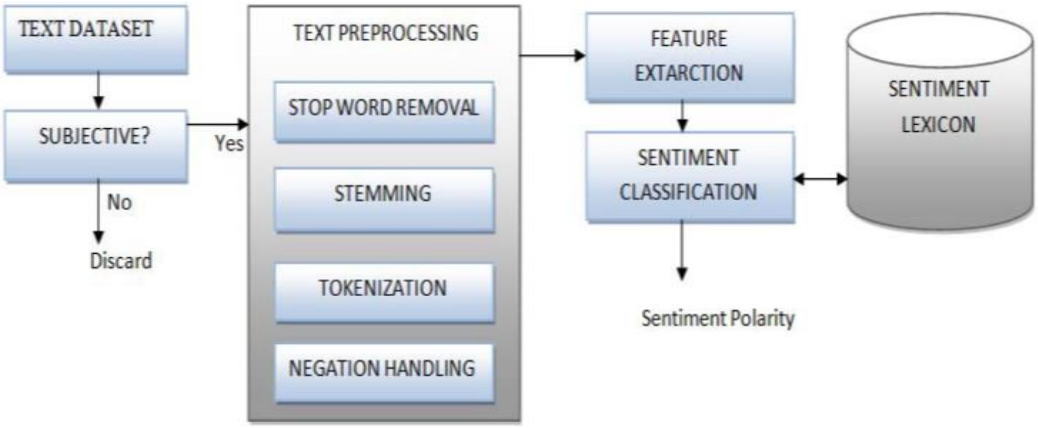

Fig 1:- Process of Sentiment Analysis. *(source [5])*

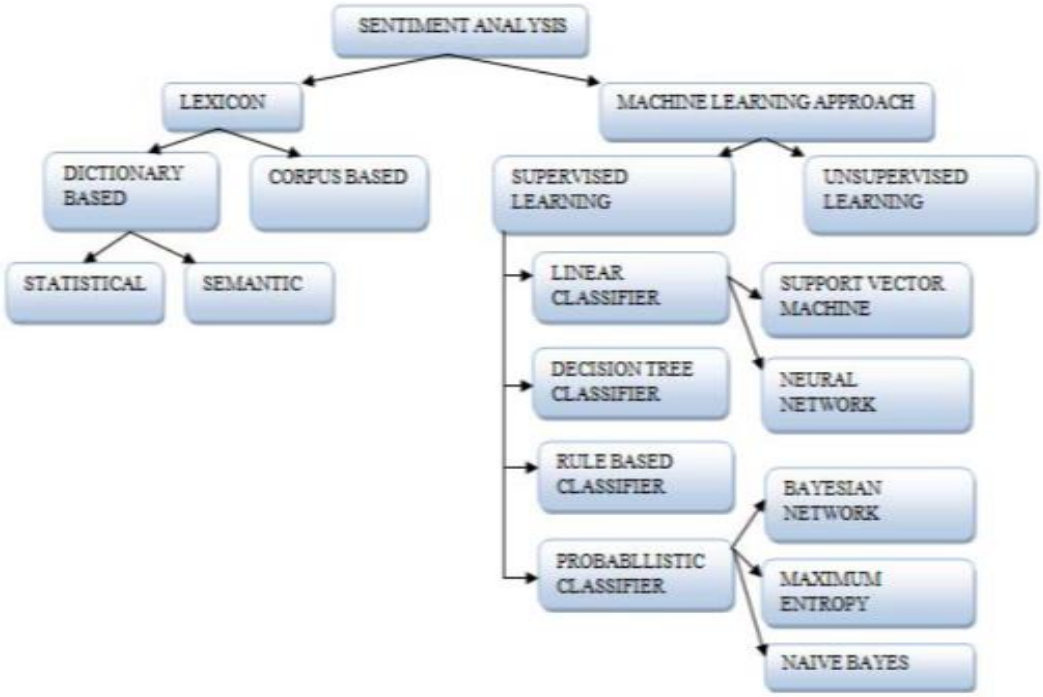

Fig 2:- Algorithms of Sentiment Analysis, *(source [6])*

B. Online stores

Online business or web based services seek ways to stand firm in competition and provide customer satisfaction to get in turn customer loyalty. The feedback system is the comprehensive way to understand consumer opinions. As it is general approach to give reviews after buying any item from e-commerce site, visiting newly opened restaurants and movie reviews, such reviews must be analyzed to understand customer's behavior for particular product.

C. Economic sectors

Online economic sectors continuously update the information about financial investments, stock market and fund management. People who follows such forums always post comments and feedbacks. Analysis of such feedback will help newly joined user to take accurate and beneficial decision before investing, so that it could minimize the risk.

D. Social media

Social media platforms are widely used ones. People are using this platform to conduct surveys and create awareness. Analysis of such system must be done to forward and spread essential information and discard fake ones.

E. Companies:

Companies give appraisals based on employee's performance. In this case feedback analysis system will help to evaluate the performance of employees from feedback collection and analysis process to make correct appraisal decision.

IV. TOOLS

Natural language processing (NLP) is popularly used mechanism for extraction of information and their classification from text. There are many open source (freely available and modifiable) tools are available for implementation of NLP. They are being used for sentiment analysis of available data on web [7], some of them but not limited to, are as follows:

A. Natural Language ToolKit

This toolkit (NLTK) is a text processing toolkit that works on multiple phases like, categorization, tokenization, ending, labelling, analyzing, etc. It gives ease in connectivity with 50 plus corpora and verbal resources [1].

B. Opinion Finder

It helps to identify discrete sentences along with the identification of the subjectivity of each sentence. The result is comprises of the conclusion regarding subjectivity along with words that actually representing the expressions on which positive, negative or neutral guesses have been made [1].

C. Open NLP

This, machine learning technique based toolkit is useful for natural language processing application on text. The NLP tasks, such as tokenizing, named extractor, parser, and conference resolution, chunkier and part of speech tagger. For formation of advanced processing services on text, these tasks are essential [1].

D. Web Fountain

This tool performs the unavoidable processes required for sentiment analysis, like text collection, indexing and sending queries etc. High performance of this tool leads to its claim for application on distributed platforms.

E. Weka

This is again a machine learning based tool and also comes with constraint of having JAVA programming language which is required for the implement of this tool. For data monitoring, it provides easy access with GUI interface. Almost every important machine learning algorithms are included in this tool, such as linear regression, classification, pre-processing, clustering etc. [1].

F. Ling Pipe

For linguistic processing of the text, this technique is used. It includes algorithms like clustering, classification, and feature extraction.

G. Opinion Observer

For analysis and comparison of opinions this tool is applied. It produces result in graphical format for easy understanding and analysis [1].

V. ANALYSIS AND CONCLUSION

Every sentiment analysis algorithm has unique features and applications. Performance of these approaches on different platforms have been studies by many research work. A summary is given in table 1. It shows the application of algorithms on different platforms.

Dimensions										
			Review				Social Media			Prediction analysis
			Movie	Product	Restaurant	Other	Facebook	Twitter	Other	
Approaches	Machine Learning	Supervised	*							*
		Unsupervised		*		*				
	Lexicon based									
			*							*
Algorithm used	Support vector machine		*		*			*		*
	Maximum Entropy			*						
	Naïve Baiyes's						*			
	Neural network			*		*				
	Decision tree					*				
	Regression					*				
	Association Rule					*				
Granularity	Document									
	Sentence		*				*			
	Aspect			*						*
	Word				*					

Table 1:- Performance of Sentiment Analysis Algorith(Source 12)

REFERENCES

[1]. Bing Liu, Sentiment Analysis and Subjectivity, Handbook of NaturalLanguage Processing, 2010
[2]. D V Nagarjuna Devi, Chinta Kishore Kumar and Siriki Prasad, "Feature Based Approach for Sentiment Analysis by Using Support Vector Machine",IEEE 6th International Conference on Advanced Computing,2016.
[3]. Christopher C. Yang ,Y.C. Wong, Chih-Ping Wei,"Classifying Web Review Opinions for Consumer Product Analysis", ICEC '09, August12-15, 2009, Taipei, Taiwan
[4]. Subhabrata Mukherjee, Pushpak Bhattacharyya, "Feature Specific Sentiment Analysis forProduct Reviews", 13th International Conference, CICLing 2012, New Delhi, India, March 11-17, 2012,Proceedings, Part I
[5]. Heeyoung Lee, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu and Dan Jurafsky, "Deterministic coreference resolutionbased on entity-centric, precisionrankedrules", Computational Linguistics 39(4), 2013.
[6]. Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky, "Stanfords Multi Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task", InProceedings of the CoNLL-2011 Shared Task, 2011.
[7]. Khin Phyu Phyu Shein et al.,"Sentiment Classification basedon Ontology and SVM Classifier", 2010 Second International Conference on Communication Software and Networks IEEE, 2010.
[8]. Sun Yueheng et al., "Automatic Sentiment Analysis for Web User Reviews", 1st International Conference on Information Science and Engineering (ICISE2009).
[9]. Cheng Mingzhi et al.,"A Random Walk Method for Sentiment Classification", Second International Conference on Future Information Technology and Management Engineering, IEEE, 2009.
[10]. Mostafa Al Masum Shaikh et al.,"An Analytical Approachto Assess Sentiment of Text", IEEE, 2007.
[11]. Dr. Arvind K Sharma et al., "Web Opinion Mining Techniques and Tools for Finding User's Opinion", In the 5th International Conference on Innovative Research in Engineering Scienceand Management (ICIRESM-16) at India International Centre, Max Mueller MargS, New Delhi, India
[12]. Bansal, N., & Singh, A. (2016). *A review on opinionated sentiment analysis based upon machine learning approach. 2016 International Conference on Inventive Computation Technologies (ICICT)*