IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.6, Special Issue.11, Dec 2018 E-ISSN: 2347-2693

A Comprehensive Survey on Sentiment Analysis and Opinion Mining

R.Bhuvaneswari[1*], S. Ravichandran[2]

[1,2]Department of Computer Science, H.H.The Rajah's Collge (Autonomous), Pudukkottai, India

[*]*Corresponding Author: eshwari9878@gmail.com*

Available online at: www.ijcseonline.org

Abstract-- The colossal volume of data available on the internet media and the evolution of internet and social media, the internet user articulates their opinion, suggestion and views on the web. The comprehensive study of user opinion is quite imperative in the decision making processes. Unearthing the useful content from these opinion sources is a tedious and cumbersome task and this formed a new sector of research field named opinion mining and sentiment analysis. Opinion mining and sentiment analysis automatically extract and classify the user's opinion clearly. This paper focuses on the review of Opinion mining about a particular topic written in a natural language and classifies them as positive, negative or neutral based on the humans emotions involved in it.

Keywords:- Data source, classification techniques.

I. INTRODUCTION

In general the sentiment analysis is a broad analysis of user's opinions, sentiments, emotions, and mindsets about a particular product or sevice. This area is increasingly gaining momentum and importance in lot of business houses [1]. Sentiment analysis is defined as the Natural Language Processing (NLP) for discovering the attitude of the public regarding a product / topic or service. Sentiment analysis, also called opinion mining, is an important study that analyzes and discovers user's opinions, sentiments, mindsets, emotions, judgments related to products, services, organizations, individuals, personalities current issues,current events, and current topics. The sentiment analysis involves in creating an automatic method to collect and observe opinions about a product/service made in blogs, comments, reviews, forums or tweets. Sentiment analysis is tool which helps judging the success of new product or service and decides which versions of a product or service are popular among the users and this decision arrived will help the organization's to come up with a strong marketing strategy.

Sentiments and reviews are becoming more important mainly due to growth of e-commerce sector, which provides ample data source for analyzing opinions. Recently, customers on e-commerce website mostly depend on reviews and opinions posted by customers and users before taking a decision to buy a product. For example opinions given on e-commerce sites like Amazon, FlipKart can influence the customer's decision in buying products online.

DATA SOURCE

The input data for the opinion mining is found on various areas and this section identifies the data source required for the extraction of the opinion mining process.

A. Blogs

With the ever evolving technology and the increasing usage of the internet, blogging and blog pages grow substantially. The blog pages have become the common media to express individual personal opinions. Actually the bloggers update the daily activities occurred in their lives and feed their opinions, thoughts, and emotions in their blogs enabling others to reads.

Data Sets

The growing e-commerce sites provides the users with lot of input data which contains reviews about a particular product and service and also multi domain sentiment dataset available on Amazon.com, which includes like novels, gadgets, Electronics and Kitchen appliances with positive and negative reviews for each domain.

B. Review sites:

There are many review websites which provides the review of the users and this helps the customers a lot in buying a particular product. The opinions collected from these websites is highly rated and used to generate a picture about a product. But these reviews for product or services are commonly expressed in unstructured format and have to be preprocessed before mining the reviews.

C. Social media sites:

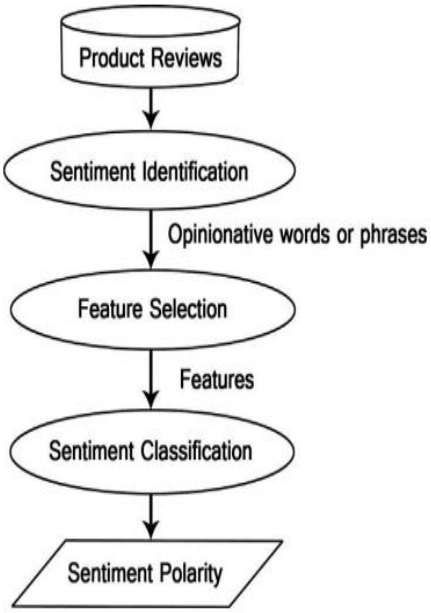
Social Media like Facebook and Twitter are one of the biggest forums where users can express opinions, suggestions and emotions. The user's expression on the social media is used by the companies, politicians and individuals to analyze and come up with better decision. The opinions in the social media will be in an unstructured format.

II. SENTIMENT ANALYSIS

Sentiment analysis is used to discover and analyze opinions and emotions from raw data. The sentiment classification is categorized into two forms namely, binary sentiment classification and multiclass sentiment classification. In binary sentiment it will be classified in two categories: positive and negative; and in multiclass sentiment it can be divided into more than two categories, such as positive, neutral and negative. To measure the strength of each sentiment, a range [-1, 1] is used to denote the sentiment value, where "-1" indicate the maximum negative degree and "1" indicate the maximum positive degree.

The data sets used in sentiment analysis are very crucial and the main sources of data are from the merchandise comments based on reviews. These comments and reviews are useful to the business as they provide complete detailed opinions regarding their product from the users or consumers. Sentiment analysis is not only used for product reviews but also used in stock markets news articles and political debates. A sample sentiment analysis process based on product review is shown in Fig.1.

FIGURE 1: SENTIMENT ANALYSIS PROCESS



TYPES OF SENTIMENT ALALYSIS

There are three main classification levels in Sentiment analysis: document-level, sentence-level, and aspect-level.

A. A.DOCUMENT LEVEL SENTIMENT ANALYSIS

The Document-level sentiment analysis targets to classify an opinion which is positive or negative. It ponders the complete text a basic data unit since it considers the full document as basic information unit and process at the Sentence level. The Sentiment Analysis aims to classify sentiment sent in every sentence. The foremost step is to check whether the sentence is subjective or objective. If the sentence found to be subjective, Sentence-level sentiment analysis confirm whether the sentence expresses positive or negative opinions that don't seem to be essentially subjective in nature.

B. SENTENCE LEVEL SENTIMENT ANALYSIS

As there is no difference between document level and sentence level classifications. Classifying text at the document level or at the sentence level doesn't provide the obligatory detail required opinions on all aspects of the entity that is required in several applications, to get the details.

C. ASPECT LEVEL SENTIMENT ANALYSIS

Aspect-level sentiment analysis targets to classify the sentiment with respect to the actual aspects of entities. The foremost step is to find the entities and their aspects. The opinion holders will provide completely different opinions based on numerous aspects of entities as shown in the example. "The camera quality of this mobile is not too good, but the battery life is long".

III. CLASSIFICATION TECHNIQUES

Sentiment Classification techniques can be categorized into three approaches namely [1], machine learning approaches, lexicon based approaches and hybrid approaches.

A. MACHINE LEARNING APPROACH

The main aim of machine learning analysis is to learn to recognize advanced patterns from the data and create logical support information to come up with the opinion. The basic function of machine learning is to develop algorithms which receive input raw data and predict an output based on the opinion of the users. The Machine Learning Approach (ML) applies the algorithms and uses linguistic options/choices. The machine learning approach is divided into supervised and unsupervised learning [9].

B. Supervised methodology

The supervised methodology uses many numbers of labeled trained documents for prediction. Supervised algorithms need human interaction at all the stages like providing input and analyzing the output, and finally need humans to find the accuracy levels during then prediction and training phase. But if the training is over, the algorithm can be applied to the new raw data for prediction.

C. Unsupervised methodology

The unsupervised method is not trained prior to the execution and this method is used for many complicated process tasks.

D. LEXICON BASED APPROACH

The Lexicon-based approach primarily depends on a sentiment lexicon, a set of well-known and precompiled sentiment terms. It is classified into dictionary-based approach and corpus-based approach that utilizes applied mathematics and linguistics methods to discover sentiment polarity.

E. HYBRID APPROACH

The hybrid approach combines plethora of approaches and commonly the lexicon based approach plays a major role in hybrid approach.

IV. OPINION/SENTIMENT COMPONENTS

There are three main components in opinion sentiments and they are, opinion holder (the person who comments), opinion object (Object on which the comment is expressed), Opinion orientation (Comment either positive or negative).

Opinion Holder → The battery life of this mobile is excellent

Opinion Object → battery life.

Opinion Orientation → Excellent

V. OPINION/ SENTIMENT TYPES

The opinion type is classified into two categories, namely

1.Regular Type : This type is referred simply as an opinion and sub-divided into two more categories,

a. **Direct Opinion:** A direct opinion defines an attitude expressed simply straight on an object or an object aspect. Example: The battery life of this mobile is excellent.
b. **Indirect Opinion:** It defines an opinion which is indirectly articulated on an object or an aspect. Example: After buying this product, I felt too bad.

2.Comparative type: A comparative opinion states a relation of similarities or diff erences between two or more entities. Example: "The camera in Iphone is better than the camera in Samsung".

VI. SURVEY BASED ON VARIOUS AUTHORS

The complete survey of work related to sentiment analysis is shown in the table 1, table 2, table 3 and various authors' research work along with their methods and approaches are studied and scrutinized [9].

TABLE 1: MACHINE LEARNING BASED WORK

MACHINE LEARNING APPROACH				
Algorithm	**Author**	**Publication**	**Accuracy level**	**Limitations**
SVM Naïve bayes Maximum entropy	Erik Boiy, Marie Fracine[2]	"A machine learning approach to sentiment analysis in multilingual web texts",2008	83 %	Limited range of annotated coaching examples area unit offered and also the texts area unit typically not grammatical.
Unsupervised (Association rule mining)	Kim schouten, et al. [6]	„Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data", IEEE xplore digital library,2017	67 %	A sensible limitation of this technique is that it needs multiple parameters.
Naïve bayes	Gayathri Deepthi, K. Sasi rekha [5]	"Opinion Mining and Classification of User Reviews in Social Media, International Journal of Advance Research in Computer Science and Management Studies, vol.2, pp. 37-41, (2014).	84 %	The accuracy level decreases with the increase in data size
SVM	Gauthami tripathi, Naganna. S, [7]	"Feature Selection and Classification approach for Sentiment Analysis", An International Journal, vol.2, pp.1-16, (2015).	85 %	-

TABLE 2: LEXICON BASED WORK

LEXICON BASED APPROACH				
Algorithm	**Author**	**Publication**	**Accuracy level**	**Limitations**
Sentiwordnet Senticnet Wordnet affect	Cataldo Musto, et al	"A comparison of lexicon based approaches for sentiment analysis of micro blog	71% 74% 62%	Word net affect was not good due to very small coverage of lexicon.

		posts"		
Dictionary based	Richa sharma, [10] sweta nigam, Rekha jain	"Mining of Product Reviews at Aspect Level"	65 %	Little bit slow during execution and the accuracy level not upto the mark

TABLE 3: HYBRID BASED WORK

HYBRID APPROACH				
Algorithm	**Author**	**Publication**	**Accuracy level**	**Limitations**
KNN SVM	Ankita Gupta, et al. [4]	"Sentiment analysis of tweets using machine learning approach", International journal of computer science and mobile computing,2017	76 %	-
SVM PSO	K. Uma Maheswari, S.P. Raja Mohana, G. Aishwarya Lakshmi [8]	"Opinion Mining using Hybrid Methods, International Journal of Computer Applications"	82%	-

VII. CONCLUSION

This survey paper discusses various approaches prevails to Opinion Mining and Sentiment Analysis and presents a detailed insight of different applications and potential challenges of Opinion Mining that makes it a cumbersome task. The state of the art of existing approaches has been described with respect to machine learning, lexicon based and hybrid approach.

REFERENCES

[1] Walaa Medhat, Ahmed Hassan, Hoda Korashy, Sentiment Analysis: A Survey, Ain shams Engineering Journal, vol.5, pp. 1093-1113, (2014).

[2] Lorraine Chambers, Erik Tromp, Mykola Pechenizki,Mohamed Medhat Gaber "Mobile sentiment analysis".

[3] Prof. Durgesh M. Sharma, Prof. Moiz M. Baig "Sentiment Analysis on Social Networking: A Literature Review"International Journal on Recent and Innovation Trends in Computing and Communication

[4] Ankita Gupta, et al. "Sentiment analysis of tweets using machine learning approach" International journal of computer science and mobile computing, 2017.

[5] Gayathri Deepthi, K. Sashi Rekha, Opinion Mining and Classification of User Reviews in Social Media, International Journal of Advance Research in Computer Science and Management Studies, vol.2, pp. 37-41, (2014).

[6] Kim Schouten, Onne van der Weijde, Flavius Frasincar, and Rommert Dekker "Supervised and unsupervised aspect category detection for sentiment analysis with co-occurancedata"IEEE transactions on cybernetics.

[7] Gautami Tripathi, Naganna.S, Feature Selection and Classification approach for Sentiment Analysis, An International Journal, vol.2, pp.1-16, (2015).

[8] K. Uma Maheswari, S.P. Raja Mohana, G. Aishwarya Lakshmi, Opinion Mining using Hybrid Methods, International Journal of Computer Applications, pp. 18-21, (2015).

[9] N. Sathyapriya, C. Akila, A Survey on Opinion Mining Techniques and Online Reviews, International Journal of Scientific Development and Research (IJSDR), vol.1, pp. 70-74, (2016).

[10] Richa Sharma, Shweta Nigam, Rekha Jain, Mining of Product Reviews at Aspect Level, International Journal in Foundations of Computer Science & Technology, vol.4, pp. 87-95, (2014)

