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

Classification Levels, Approaches, Tools, Application and Challenges in Sentimental Analysis- A Survey

S. Thulasi Bharathi[1*], S. Charles[2]

[1]Department of Information Technology, St.Joseph's College, Trichy, Tamilnadu, India
[2]Department of Information Technology, AI Musanna College of Technology, Muladdah Musanna, Oman

[*]*Corresponding Author: srimas88@yahoo.co.in*

Available online at: www.ijcseonline.org

***Abstract*—** Sentiment analysis is an application of natural language processing. It is also known as emotion extraction or opinion mining. Sentiment analysis or opinion mining is the computational study of opinions, sentiments and emotions. Opinions are usually particular expressions that designate people's sentiments, judgments' or approach toward entities, events and their properties. In general, opinions can be expressed on anything, e.g., a product, a service, an individual, an organization, an event, or a topic. In this paper, the SA classification levels, approaches are discussed. It also reports about various categories of tools used to process the sentimental analysis data. And various application and challenges in sentimental analysis are explained.

***Keywords*—** Senitmental Analysis, NLP, Opinion Mining.

I. INTRODUCTION

An Opinion is a belief or view which was formed about a product or a thing by a majority of people, not necessarily based on fact/knowledge. It generally refers to what a person thinks about something or opinion is a subjective belief, in turn the result of emotion or facts interpretation [1]

Opinion Mining (OM), also called as Sentiment analysis, is a natural language processing type to find common mood or views about a product or topic.

SA involves collecting and examining opinions about services or products from various blog posts, tweets, reviews and comments. Sentiment analysis is very useful in the areas of economic and marketing. This helps to determine the product's success/failure rate. It also helps in analyzing the success strategy.

With the growth of social media (forum discussions, reviews, blogs, comments and postings in social network sites, micro-blogs, Twitter) on the Web, organizations and individuals are using content in such media to make decisions. In general, the overall polarity of the context and the writer's sentiment about the aspect is identified using opinion mining.

II. SENTIMENTAL ANALYSIS PROCESS

The sentiment analysis is a complex process that involves 5 different steps to analyze sentiment data. The sentiment analysis process steps were discussed in [2] as follows

2.1 SENTIMENT ANALYSIS PROCESS STEPS:

- **Data Collection**

The first step of sentiment analysis is to collect data from user generated content contained in microblogs, online discussion-forums, social network working media, online portals. These data are confused, expressed in different ways by using different vocabularies, slangs, context of writing and emoticons. Natural language processing is a difficult task. Therefore, text analytics and NLP(natural language processing) are used to extract and classify the raw data.

- **Text Preparation**

This process consists in cleaning the extracted data before analysis. Non-textual contents and contents that are irrelevant for the analysis are identified and eliminated.

- **Sentiment Detection**

The extracted sentences of the reviews and opinions are examined. Sentences with subjective expressions whichexpresses some personal feelings or beliefs (opinions, beliefs and views) are retained.Sentences with objective communication (facts, factual information) are discarded.

- **Sentiment Classification**

In this step, subjective sentences are classified in positive, negative or neutral.

Visualizing the data

The main objective of sentiment analysis is to convert unstructured data into meaningful information. After completing the analysis, the results can be visualized using charts and graphs. The frequency(time) can be analyzed and can be visualized graphically by constructing a sentiment time line with the chosen value (frequency, percentages, and averages) over time.

2.2 SENTIMENT CLASSIFICATION LEVELS

The Sentiment classification is a task of classifying a target unit in a document to positive (favorable) or negative (unfavorable) class. There are three main classification levels [3]

2.2.1 Document Level Classification

In this type of classification, sentiment is extracted from the entire review, and a whole opinion is classified. The goal is to classify a review as positive, negative, or neutral.

Example "I bought a smartphone a few days ago. It is such a nice phone, although a little large. The touch screen is cool. The voice quality is clear too. I simply love it!"

"Document level classification works best when the document is written by a single person and expresses an opinion/sentiment on a single entity".

2.2.2 Sentence Level Classification

The sentence level sentiment analysis determines whether each sentence expresses a positive, negative or neutral opinion, for a product or service review. This type is used for reviews and comments that contain one sentence and written by the user. Sentence –level classification involves two steps

- Subjectivity classification of a sentence into one of two classes:

Objective and subjective

- Sentiment classification of subjective sentences intotwo classes:
- Eg 1: I called her through phone yesterday, but the line is not clear.

In the above example, "line is not clear" is the opinion.

Positive and negative

Eg 2: I called her through phone yesterday.

From the above examples, example 1 is subjective sentence which gives an opinion "the line is not clear" and so the sentence is subjective. Example 2 is objective sentence, which gives the factual information alone.

2.2.3 Entity and Aspect Level Classification

Aspect level is the opinion mining and summarization based on feature. The classification concerns by identifying and extracting product features from the source data. This type is used when we need sentiments about desired aspect/feature in a review.

The following is the steps for finding the sentiments in the aspects [4]

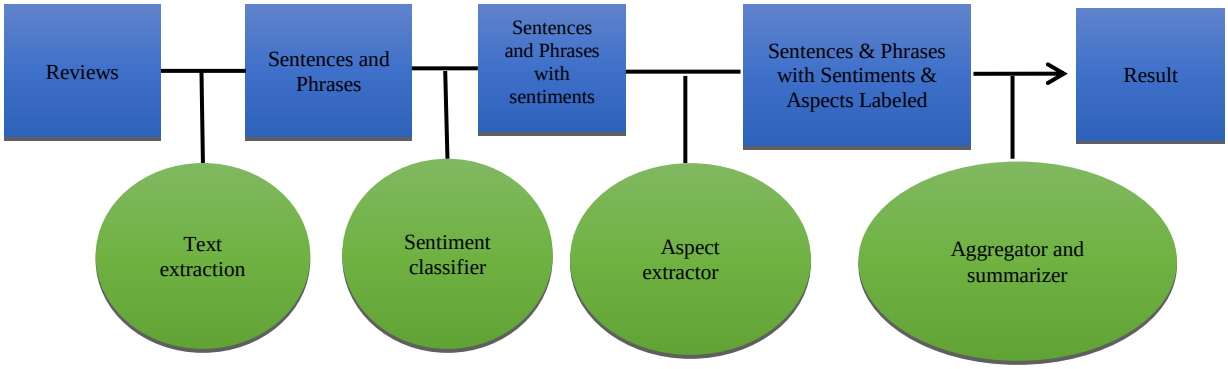

Fig 1: Steps in Finding the Sentiments in Aspect Level

2.3 SENTIMENT CLASSIFICATION APPROACHES:

The sentiment classification approaches can be classified in:[5]

(i) Machine Learning
(ii) Lexicon Based And

The categorization is as shown in figure 2 Sentiment Analysis Approaches.

2.3.1 MACHINE LEARNING APPROACH

The machine learning approach is used for envisaging the polarity of sentiments based on trained as well as test data sets. It applies the Machine Learning algorithms and uses semantic features. [6] The machine learning approaches the most used are:

(i) Bayesian Networks: It is a probabilistic approach that models relationships between features in a very general way. These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables.
(ii) Naive Bayes Classification: it is an approach particularly suited when the dimensionality of the inputs is high. Despite its simplicity, it can often outperform more sophisticated classification methods

(iii) Maximum Entropy: this method is mostly used as alternatives to Naive Bayes classifiers because it does not assume statistical independence of the random variables (features) that serve as predictors. The principle behind Maximum Entropy is to find the best probability distribution among prior test data.

(iv) Neural Networks: this model is based on a collection of natural/artificial neurons uses for mathematical and computational model analysis

(v) Support Vector Machine: it is a supervised learning model which analyzes data and patterns that can be used for classification and regression analysis.

2.3.2 LEXICON BASED APPROACH

Lexicon based approach extracts the sentiments from the text. The process of assigning a positive or negative label to a text that detects the text's opinion towards its main subject and it involves in calculating the sentiment polarity for a review using the SEMANITC ORIENTATION of the text, words or phrases given in the review. **Lexicon Based Approach is categorized as [3]**

(i) Dictionary-based approach: The dictionary is the domain specific i.e. Polarity of the words are set according to a specific domain (shopping blogs, movie reviews etc) to know products review.

In the dictionary based techniques, a small set of opinion words are collected manually with known orientations, and then to grow this set by searching in the WordNet dictionary for their synonyms and antonyms.

(ii) Corpus based approach: The corpus based approach begins with a basic list of opinion words and tofind the other opinion words in a large quantity to help in finding the opinion words with context specific orientations. E.g. extract sentiments from microblog data.

Corpus based approach further more classified as statistical and semantic approach. In statistical approach, co-occurrences of words are calculated to identify sentiment. In semantic approach, terms are represented in semantic space to discover relation between terms[7]

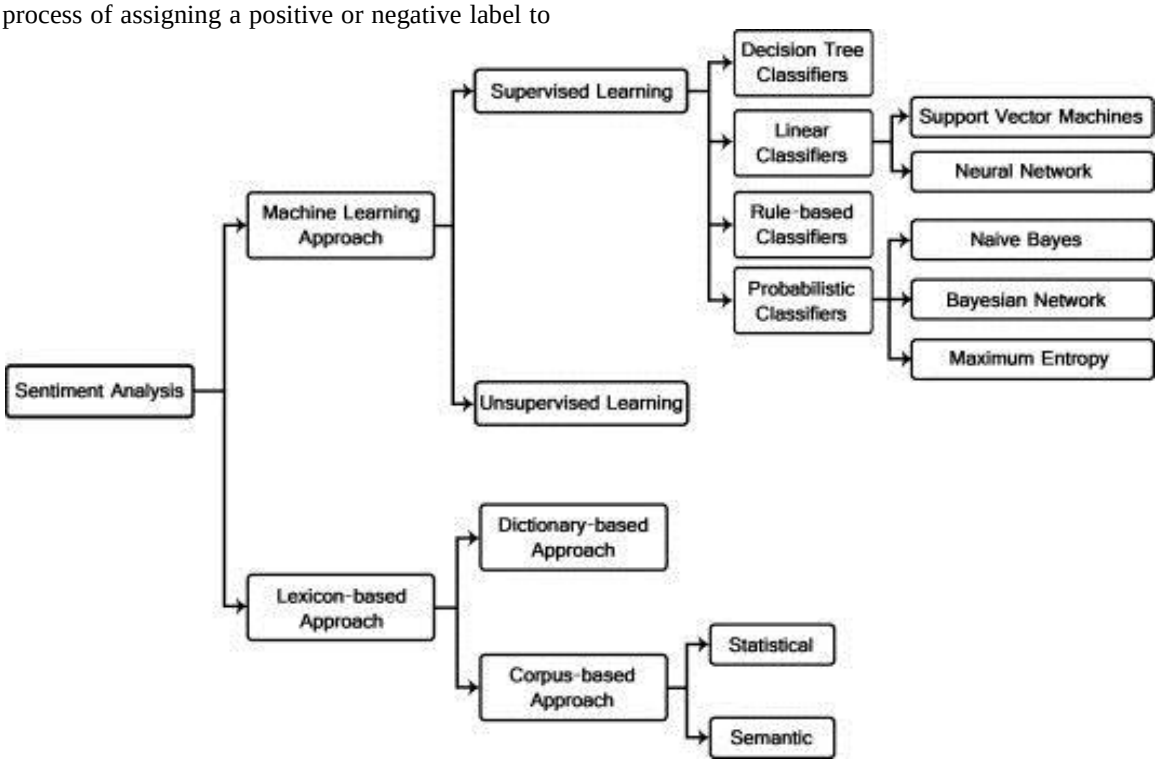

Figure 2. Sentiment classification techniques, [3]

2.4 SENTIMENTAL ANALYSIS TOOL:

Sentiment Analysis tool — process a set of search results for a given item, generating product attributes (quality, features etc.) and aggregating opinion. It is an automatic extraction of knowledge from the opinions posted by the public in the social networking media on a particular topic/problem. Sentiment analysis tools are used in different fields such as politics, finance, business etc. a large number of sentiment analysis tools are designed to do text analytics. Tools along with details about their application and extensions are provided in table 1. [8]

Table 1 : TEXT ANALYSIS TOOLS- CATEGORY

Tool name	Application	Web source	Extension for Sentimental Analysis
Lexalytics	natural language processing(NLP)	https://www.lexalytics.com/ technology/sentiment	Saliance
IBM Watson Alchemy API	NLP, Text Analytics, Content Recommendation	https://www.ibm.com/watson/ developercloud/alchemylanguage.html	AlchemyAPI
Provalis Research Analytics Software	Text analysis, Content Analysis	https://provalisresearch.com/	WordSat
SAS Text Miner	Text analysis, Ontology, Sentiment Analysis, NLP	https://www.sas.com Nil	Nil
Sysomos	Social media monitoring, Text Analysis	https://sysomos.com	Media Analysis Platform (MAP)
Expert system	Semantic Search, NLP, Conent Analysis	www.expertsystem.com	Cogito
Rapid miner	Social media analysis, Market Search	https://rapidminer.com	Setiment Analysis
Clarabridge	Social media Analysis, Sentiment Analysis	www.clarabridge.com/ text-analytics	Nil
Luminoso &bitext	Text analysis Bitext Sentiment Analysis, Concept Extraction, Text Analysis	https://luminoso.com https://www.bitext.com	Luminoso Compass
AYLIEN	Text analysis, NLP, Concept Extraction	aylien.com	Nil
Averbis Nil	Text analytics, Information Discovery	https://averbis.com/en	Nil
Buzzlogix	Sentiment Analysis, Social Media Monitoring	https://buzzlogix.com	Nil
VisualText	NLP,Text Analytics	www.textanalysis.com	Nil
Semantria	Text analysis by API and Excel plugin, Sentiment Analysis	https://semantria.readme.io/	Nil
General Sentiment	Sentiment Analysis, Text mining, Social media analytics	www.generalsentiment.com	Nil
Abzooba	Social media monitoring, Text Analysis	www.abzooba.com/	XPRESSOInsight
Medallia	Social media monitoring, Text mining	www.medallia.com	Nil
Synapsify	Social media monitoring, Text mining	www.gosynapsify.com	Snapify core API
Etuma	Social media monitoring, Sentiment Analysis	https://www.etuma.com/	Nil

2.5 APPLICATIONS

Sentimental analysis has been applied in various areas like business intelligence, market prediction, social media etc. The following list gives the applications of sentimental analysis. [9]

- **Market and FOREX rate prediction.**

Sentiment analysis provides insight to financial market for investors to recognize, retort and reply to market opinions. To scrutinize and observe the overall sentiment of financial market sentiment analysis or opinion mining can be utilized.

- **BUSINESS ANALYTICS**

Product vision or demand is determined by sum of the discussions about, advertising, public relations and corporate messaging. Sentiment analysis is useful in predicting how company's product name, product or service is being alleged through web.

- **SOCIAL ADVERTISING OVER MICROBLOGS**

Online advertising Sentiment analysis or Opinion Mining plays a vital role in marketing and advertising domains. Promoter has the duty to examine popularity and performance of ads that was posted on site by the public. After reviewing the comments of web users, promoters can decide what type of advertisements should put up on which type of web pages.

- **POLITICS**

In the time of political events, sentimental analysis is utilized to track how voters think about various political issues and to study the relation to the speeches and the actions of the candidates. [10]

- **EDUCATION FIELD**

The students' performance can be predicted or analyzed by monitoring the students' learning style where sentimental analysis can be used. Course opinion, infrastructure of the learning environment, teaching and learning issues can be addressed which benefits the students. The response can be collected via social media such as Twitter, Face book, or e-learning sytem and hence the teaching –learning process can be improved by analyzing sentimental analysis over it. [11]

2.6 CHALLENGES

Some of the major challenges faced in the area of sentiment analysis.

- Domain-oriented Issues

Existing researches are trying to overcome domain dependence challenge using domain transfer [12] where small amount of training data is labeled from the new domain which is called the target domain where it used for testing the original/source domain training dataset.

- **Challenges related to opiniondata[**13]
  - Comparative opinion
  - Subjective words not expressed any opinion
  - Objective words implicitly expressed opinion
  - Negation handling
  - Sarcasm and ironic detection

- **Challenges with respect to Computations**[14]
  - Incremental Approach
  - Parallel Computing for massive data
  - Ironic and sarcastic statements Improving the precision of algorithm
  - Grammatically Incorrect Words
  - Real time opinion mining
  - Demand of automation

III. CONCLUSION

Opinions play an important role for individuals as well as companies to read the curiosity of a customer about their needs and interest towards products. Thus, any individual can take the decision to buy a product or any companies can provide what exactly customer want and establish better customer relationship by solving their problems through their opinions. In this paper, different classification levels and its approaches of sentiment analysis were reconnoitred. In order to accomplish the aim of this paper, the challenges with respect to data and computations are discussed. In addition, various areas of applications are also discussed. This study provides a groundbreaking aspect for researchers to study and apply sentiment analysis in various domains.

REFERENCE

[1] Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis–An Assessment of Peoples' Belief: A Survey. International Journal.

[2] AlessiaD'Andrea, Fernando Ferri, PatriziaGrifoni, TizianaGuzzo(2015) Approaches, Tools and Applications for Sentiment Analysis Implementation, International Journal of Computer Applications(09758887),vol.125)

[3] Medhat, W., Hassan, A., Korashy, H. 2014. "Sentiment analysis algorithms and applications: A survey", Ain Shams Eng.

[4] Ashish Katrekar, AVP, Big Data Analytics " an introduction to sentimental analysis", GlobalLogic Inc, www.globallogic,com. – FIGURE 1.

[5] Maynard, D., & Funk, A. 2011. Automatic detection of political opinions in tweets. In: Proceedings of the 8th international conference on the semantic web, ESWC'11, p. 88-99.

[6] Pang, B., Lee, L., Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. Proc. of 7th EMNLP, pp.79-86.

[7] Michael W. Berry, Soft Computing in Data Science, First International Conference, Scds 2015, Putrajaya, Malaysia,

September 2-3, 2015, Proceedings (Communications in Computer and Information Science)

[8] Vishal vyas , Uma, "An Extensive study of Sentiment Analysis tools and Binary Classification of tweets using Rapid Miner", 6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India, Procedia Computer Science 125 (2018) 329–335.

[9] K. Ravi, V. Ravi , "A survey on opinion mining and sentiment analysis: tasks, approaches and applications" Knowledge-Based Systems (2015), doi: http://dx.doi.org/10.1016/j.knosys.2015.06.015

[10] Lucas Montesano's; S. Juan Pablo Rodrguez ; Marcos Orchard ; Susana Eyheramendy," Sentiment analysis and prediction of events in TWITTER", CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) , Page(s):903 - 910 , Oct 2015.

[11] Akshi Kumar, Prakhar Dogra and Vikrant Dabas,"Emotion Analysis of Twitter using Opinion Mining",IEEE,978-1-4673-7948-9/15,2015

[12] B. Lue, Sentiment Analysis and Opinion Mining (Morgan & Claypool Publishers, 2012).

[13] Osamah A.M Ghaleb ,Anna Saro Vijendran, "THE CHALLENGES OF SENTIMENT ANALYSIS ON SOCIAL WEB COMMUNITIES" international journal on advanced researchin science and engineering vol 6, issue 12, dec 2017

[14] Harshali P. Patil1 and Mohammad Atique "Applications, Issues and Challenges in Sentiment analysis and Opinion Mining– A User's perspective" international journal of control theory and aplications. International science press Vol 10. Number 19 2017.

