Sentiment Analysis - An Assessment of Online Public Opinion: A Conceptual Review

Dr. Garima Sainger[a]

[a]Assistant Professor, Institute of Business Management, GLA University, Mathura, India.
E-mail: garima.sainger@gla.ac.in

Abstract: This conceptual paper discusses sentiment analysis as a technique of research. It is a tool support decision for textual data collection and analysis available on the internet. It is also considered as a technique of data mining. It uses machine learning language to evaluate textual content. As a method of research, it is computational by nature and identify and categories opinions in the form of text. It targets a large data without any delay and hurdle and also facilitates the collection of data and its analysis. It helps domain leaders to collect real time data about emotions, opinion and attitude, without compromising, validity, reliability and generalizability. The paper also presents this as a way to divide quantitative and qualitative data through real time innovative ways of collection and analysis of data. The paper also discusses limitations one experience when applying this in their domain of research.

Keywords: Sentiment Analysis, Opinion, Social Networking Sites, Computational, Textual Data.

1. Introduction

Sentiment analysis, also known opinion mining, is the area of study that evaluates sentiment, opinion, attitude, analysis, appraisal and emotions of people for different services, products, organisations, individual, events and their respective features.

Sentiment analysis as a term appeared first in (Nasukawa and Yi, 2003), and opinion mining appeared first in (Dave, Lawrence and Pennock, 2003). While sentiments and opinions, researched earlier too (Wiebe, 2000; Tong, 2001; Das and Chen, 2001; Pang, Lee and Vaithyanathan, 2002; Turney, 2002; Morinaga et al., 2002;).

As a concept, it targets to determine positive, negative or neutral human feelings or opinions towards a product, service or information available. This is useful widely on Social networking sites like; Facebook, Twitter, Myspace & WhatsApp-Groups etc., and allows to gain a wider opinion of public for certain topics. This makes the decision making process faster and simpler than earlier due to the availability of real time situations such as people opinion and emotions, etc. It has become a popular area in decision making (Tawunrat and Jeremy, 2015) As per a survey by podium, online reviews affect their purchase decisions as per 93% of consumers.

In the 2012 presidential election to gather the opinion of the public about the announcement of policies and messages of the campaigns, the administration of Obama used sentiment analysis. This gave them the ability to overpower their opposition counterparts, thus emerging the winners for the second term. On the same, Donald Trump used the same method to understand how The Indigenous Americans thoughts about the immigrants, Muslims, African-Americans, Latinos & different cultural practices that came into implementation during The Obama Regime. This was through Social Media, hence using the above as the key to his succession in the powers.

Sentiment Analysis is defined as a process used to know about the mood behind a choice of words to know about the opinions, attitudes and emotions expressed within an outline mention. Communication through media affects leadership, decision making and strategies. It is also being highlighted by socialization theory that cognitive, affective and behavioural attitude of individuals also affected by Communication among them. The most popular social networking sites where people prefer to share their stories, experience, life styles and likes are Facebook, Twitter, Instagram and Myspace etc. Posting information could lead their friends to do the same or use their information to make a decision. This article contributes a varied opinion in social science research by highlighting the unique dimensions of Sentiment analysis and how the administration and leaders can benefit from the technique. The paper also highlights, the available literature in the direction of Sentiment Analysis and role of it in opinion capturing. But before going to discuss further one should understand first the actual meaning of Sentiment Analysis.

2. Sentiment Analysis: Concept

As an area of study it is originated from the disciplines of sociology, anthropology and psychology. Sentiment analysis is the "interpretation and classification of emotions (positive, negative and neutral) within text data using text analysis techniques (Fig:1). Sentiment analysis tools allow businesses to identify customer sentiment toward products, brands or services in online feedback".
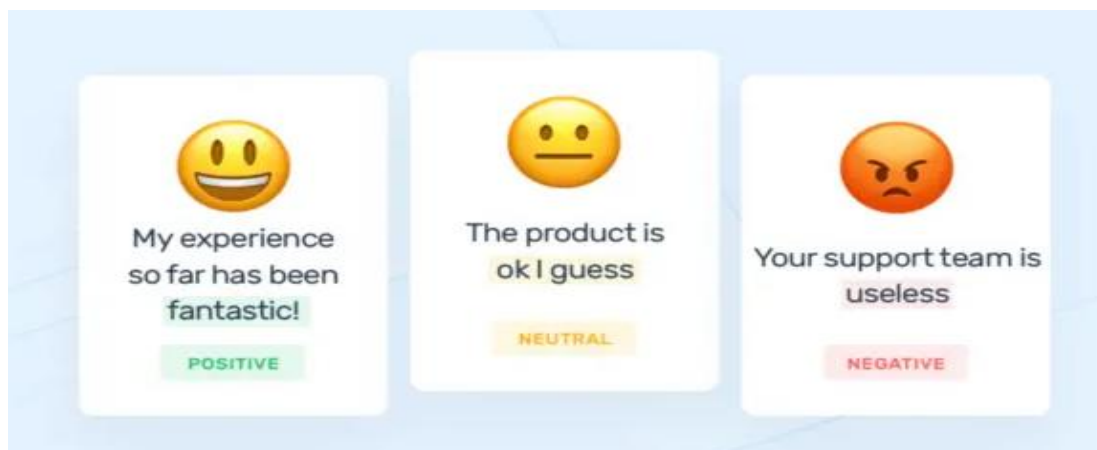

Figure 1. Sentiment Analysis
Source: monkeylearn.com

Sentiment Analysis is derived from the affective stance and appraisal theory, focuses on the role of emotions in cognitions shaping (Graziotin and Viking, 2016). Emotional assessment is a kind of assessment of a given context (good-bad) that focuses cognitive and bodily responses both (Kim, 2014).

However, the role of emotions in business functional areas are not new. For example, in marketing, behind purchase behaviour, customer emotions, acts as indirect motivators. For Politicians, entrepreneurs and administrators the concept of sentiment analysis is not new. By old surveys, customer comments and interview of focus group entrepreneurs analyzed sentiments. Analysis of human sentiments on the internet is much in practice now a day. Human emotions or convictions expressed on the internet, is a kind of attitude towards an event, object or situation generally displayed through different online media channels, with most widely used being social networking sites.

Political leaders use social networking sites such as twitter, Instagram and Facebook to convey their agendas and manifest to a large public and from their response and reactions they come to know about public thoughts and emotions towards their agendas and manifest. This enables them to come up with a plan or something that society finds beneficial and makes possible for them to outcast rivals. It is also being used to measure their popularity.

Sentiment analysis is known as analysis of online expression of an individual. It evaluates opinions and attitudes by using machine learning techniques on any area of interest or topic given. In the data mining filed, the two standpoints operational and functional exist to define the concept. "The functional aspects focus on practical uses of the method. Sentiment analysis as a process is described as that which categorizes a body of textual information to determine feelings, attitude and emotions towards a particular issue or object. There is another aspect of sentiment analysis in data mining, which basically focuses on the operations of the technique as a sub-field of computational linguistics". Sentiment analysis can be described as "An automated subjectivity analysis similar to opinion mining and appraisal extraction which focuses on extracting and classifying texts with machine language and computer programming" (Kumar and Sebastian, 2012). Though both perspectives discussed above are different, but their basic orientation is same. In another way, "Sentiment analysis is a data mining technique that uses natural language processing, computational linguistics and text analytics to identify and extract content of interest from a body of textual data".

Sentiment Analysis as a process, integrate different functions to develop wisdom from data based on text. This is a five stage process (Fig:2) starts with collection of data and lasts with output presentation. The process involves data sourcing through web content generated by user on social networking sites or any other online platform. The sourced data is examined and cleaned for subjective sentiments. Using a system of Polar classification these sentiments are further categorized into varied categories. Finally, interactive displays are used to present the results.

3. Process of Sentiment Analysis

Fig: 2 discusses about the process of sentiment Analysis. The different steps involved in the process are:
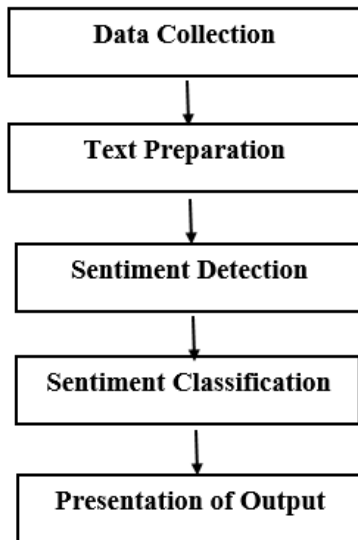
```
┌─────────────────────────┐
│     Data Collection     │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│     Text Preparation    │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│    Sentiment Detection  │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│  Sentiment Classification │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│   Presentation of Output │
└─────────────────────────┘
```

Figure 2. Process of Sentiment Analysis
Source: Liu, 2009

Data Collection

Content available on the internet for the years used by sentiment Analysis. The discussion, queries of users on social network sites like Facebook, Instagram and twitter as well as public forums like blogs, product review boards and discussion boards are points used for the data source. These data points are disorganized, huge and disintegrated on multiple portals very often. On the platforms of social media feelings and opinions are expressed in varied ways. The kind of vocabulary used, writing context and amount of details mentioned are different. Because of this, manual analysis is tedious and almost impossible to perform. But for data classification and extraction innovative text analytics and natural language processing (NLP) also employed along with sentiment analysis. After data extraction, it will be prepared for analysis.

Text Preparation

Before analysis cleaning the extracted data is known as text preparation. It involves recognizing and segregating textual content from non-textual one, along with data that can disclose the recognition of people, including: username, date, location, etc. Words that are not appropriate for the context of analysis are also excluded from the text database.

Sentiment Detection

This is the third stage of the process, use computational task to extract and appraise viewpoint from the textual dataset. For the purpose it includes only those sentences which have subjective expression and also examine sentences for subjectivity. It excludes sentences from further communication, convey objective communication and facts.

There are different levels at which sentiment detection can be done. It can be done either through complete document or by a sentence, phrase or a single term. The commonly used technique to do this consists of Unigrams, N-Grams, Negation and opinion words, etc. (Russel and Norvig; 2002). These techniques are:

Unigrams

This is one of the classical approaches often known as bag of words approach. Based on a single word frequency this approach represents each element as a feature vector.
N-Grams

This approach represents features of a document by multiple words in sequence (e.g.: words in pairs, triplets) which captures more context.

Lemmas

Rather than using literal words, it focuses on the use of synonyms. For example: better-good, best-good. This not only makes the classification task simpler, but also supports generalization. This involves the use of synonyms rather than the literal word. For example: better-good, best-good. This method reportedly makes the classification task easier as well as facilitates generalization. The research is this direction further suggests that meanings are not always necessarily synonyms, but when words are linked to their thesaurus meanings accuracy of sentiment classification reduce (Kushal et al., 2003).

Negation

This is an extension of N-gram methods where the phrases "I do not like this book" and "I like this book" would be treated similar under maximum classification techniques. But negation puts both terms into opposite groups. However, to model negation is not always easy. When ironies and sarcasm are used in a sentence, it is difficult to identify negation (Pang and Lee, 2008).

Opinion Words

Words used to describe people feelings and opinions are known as opinion words (nouns, adjectives, adverbs and verbs). They represent the absence and presence of a word and incorporated into a feature vector. These words represent subjectivity in a document. To find textual sentence making references to various features, attributes and objects is not uncommon. Sentiment analysis through mathematical algorithms can be used to extract and categories these objects, features and attributes. Which further assists in formally categorization, classification and data summarization which enhances precision and assists in the analysis stage.

4. Sentiment Classification

This stage is known as polarity classification. Each subjective sentence is classified into the classification groups textual dataset. On a continuum these groups usually represent two extreme points (positive- negative; good-bad; like- dislike etc.).

Binary and polar classification uses a wide variety of machine learning techniques. From observation and past experience machine learning with artificial intelligence targets at building computational models. This computer programming is used to understand and learn a particular set of data and predict or optimize some future criterion with the help of that acquired knowledge. The basic objective here is to develop a function suitable for forecasting a focused outcome- y (Dependent variable) using predefined input criteria or attributes-x. When outcomes are familiar, this is known as "Supervised learning". Statistical learning theory based support vector machine is another classification technique (Vapnik, 1995).

In this analysis, language of data is very important and can't be ignored. This uses English language for writing tools, techniques and literature. For the translation of multilingual nature, it presents a problem. While research also focuses on assimilating different languages to interest domain. Cross lingual adaptation is one of the challenges taking into consideration cultural idiosyncrasies (Hovy and Kim, 2006; Blitzer et al. 2007).

The bag of words method is one of the most basic method used to assign a weight or score on the basis of nature (positive or negative) and repetition of words in a text document. As score is calculated for each term, so as by taking the arithmetic sum or mean, score for whole document is calculated. Subjective assignment of a score is the simplest method for opinion document. This computes pseudo-expected" value of the document. Though this method is simple to comprehend and statistically grounded. This method is also questioned for not giving a choice to large volume data categorization. In addition to this, due to the diverse nature of human being, the reliability of this classification has also questioned.

Lexicons is another technique, perform as a connect among a language and the meaning presented by that language. It lists all words and their meaning in a particular language. For use in sentiment analysis a variety of lexicons have been designed. For the English Language WordNet is a lexical database, created by Princeton University in 1985. This database not only provides common meaning, but also merge words into

a set of synonyms categorized as synsets, but also through conceptual-semantic and lexical relations record relationship between synonym sets.

This synonym relationship is also used to compute the distance among words based on their relationship and diversity (Kamps et al., 2004). First, they categorize each word on the spectrum of positive-negative (+, -) and then on the basis of the length of the spectrum they compute the distance between words with closer words having shorter distance.

Another scoring method web search, with single word classification recognizes the contextual problems (Turney, 2002). For instance, "unpredictable "word has a favourable review of a movie, but the same word has unfavorable reviews for an automobile. To manage this problem, the method known as "tuples", use adverbs combined with verbs and adjectives combined with nouns. Tuples are segregated from reviews firstly and then semantic orientations of the segregated tuples are established. Lastly, for the whole document, average semantic orientations are computed. The search engine AltaVista is used to determine the semantic orientation of tuples, and ran two queries. First, it discusses about rate how many documents using the tuple "poor" and second rating the number of documents using tuple "excellent". If tuple appeared many times with "excellent" query, in comparison to "poor" then it is taken as a positive orientation while opposite in case of tuple appeared maximum times in "poor" query, then it would be treated as negative.

5. Presentation of Output

To convert the unstructured fragmented text into meaningful information is the general purpose of this analysis. Table, graphical displays like table, bar charts, line graphs and pie charts are the mainly used for the purpose. It segments polarity on frequencies, size, percentage and colour. For example, Table:1 and Fig:2 below represent ways of output presentation of sentiment analysis of twitter data.

Table 1. Sentiment Analysis Output

Query	Positive	negative	Neutral
Movie	53	11.1	35.8
politics	26.6	12.2	61.1
fashion	38.8	13.3	47.7
fake news	16.3	72.1	11.4
Justice	35.2	15.9	48.8
Humanity	36.9	33.3	29.7

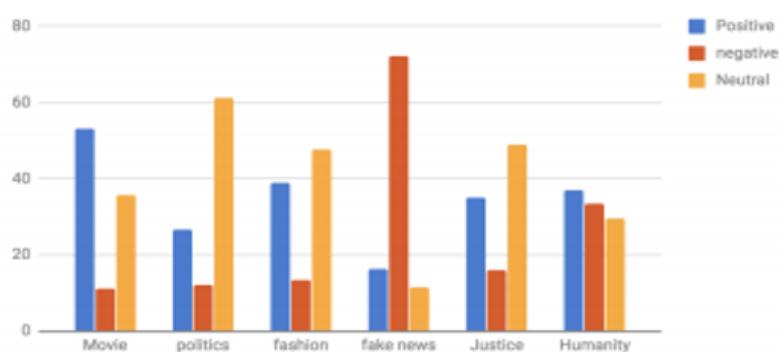**Source:** Sentiment Analysis of Twitter Data

Figure 3. Sentiment Analysis Output
Source: Sentiment Analysis of Twitter Data

6. Tools and Works in Sentiment Analysis

Existing online textual content available on sites like twitter, Facebook, Amazon, Epinion and Rotten Tomatoes have made researcher capable to avoid the use of manual annotation (Pang and lee, 2008). Many sentiment search engines exist to generate text results by running queries on any topic of interest. It code and categories results generally into two or three polar categories. Sentiment analysis becomes quite easy in presence of sentiment search engines. But the question of reliability and validity also raised in context of sentiment classification due to reviews available on sites like Epinion and Amazon have found to be positively skewed. Annotated databases and word lists are another tool available for sentiment analysis, categories words on the basis of emotions for example- aversive (negative) and attractive (positive). Sentiment analysis programs designed to categorize short textural documents include other tools.

7. Limitations

Though, sentiment analysis has become the simplest and easiest way to understand people's emotions, thoughts and attitudes towards a certain topic, but we have major issues that limit the above hence making it hard to understand the so called emotions, thoughts and attitudes. The following we have to elaborate and look at to make Sentiment Analysis more reliable and effective.

Linguistic, acts as a challenge to computational understanding, some words and phrases the computer can never understand or becomes difficult for the one using the method during data mining. As well, the users' response towards a certain topic might mean something different from what's being discussed. Use of emoji, especially on Facebook, Twitter and WhatsApp-Groups tends to give a challenge for understanding the emotions or the message being delivered.
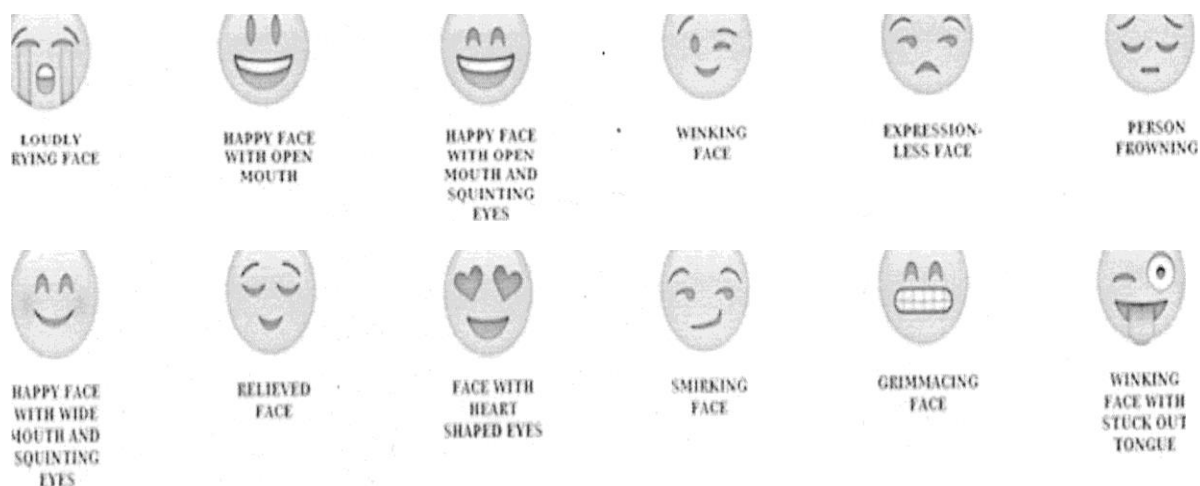

Figure 4. Emoji and Their Meanings
Source: Getty Images

Sarcasm, sometimes it's difficult to understand people on the message they post as sometime they may mean the opposite of what they portray i.e. (Irony). Nevertheless, for good reputation management and monitoring public relation sentiment Analysis is becoming popular as an important tool. If one uses it thoroughly, it may help to assess past performance and improve future results.

8. Conclusion

The present paper discusses sentiment Analysis as a technique of decision making, which is relatively new in the context of research. It facilitates the conversion of real time large volume of textual data into useful information. The paper also highlights Challenges faced in analysis and decision making by using textual data. Sentiment analysis helps to collect a large volume of qualitative information from participants without any external interference.

Sentiment analysis also helps to gather people opinion about a study, topic or leadership in real time. It eliminates individual biases and subjectivity and also offer a comprehensive and rigorous technique to interpret data in this new and challenging direction. Sentiment analysis has potential to revolutionize the decision making process, if it is integrated properly with existing styles of decision.

References

1. Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, 70-77.
2. Dave, K., Lawrence, S., & Pennock, D.M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *In Proceedings of the 12th international conference on World Wide Web*, 519-528.
3. Wiebe, J. (2000). Learning subjective adjectives from corpora. *in Proceedings of National Conference on Artificial Intelligence (AAAI)*, *20*(0).
4. Tong, R. (2001). Detecting and tracking opinions in on-line discussions. *In UCB/SIMS Web Mining Workshop*.
5. Sanjiv, D., & Chen, M. (2001). Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific finance association annual conference (APFA)*.
6. Bagheri, H., & Islam, M.J. (2017). Sentiment analysis of twitter data. *arXiv preprint arXiv:1711.10377*.
7. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.
8. https://www.gettyimages.in/illustrations/emoji?family=creative&mediatype=illustration&phrase=emoji &sort=mostpopular.
9. Turney, P.D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.
10. Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002). Mining product reputations on the web. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 341-349.
11. Tawunrat, C., & Jeremy, E. (2015). Chapter information science and applications, simple approaches of sentiment analysis via ensemble learning, volume 339 of the series lecture notes in electrical engineering, DISCIPLINES Computer Science, Engineering SUBDISCIPLINESAI. *Information Systems and Applications-Computational Intelligence and Complexity*.
12. Graziotin, D., Viking Mäntylä, M. & Kuutila, M. (2016). The Evolution of Sentiment Analysis-A Review of Research Topics, Venues, and Top Cited Papers. *arXiv*, arXiv-1612.
13. Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
14. Kumar, A., & Sebastian, T. M. (2012). Machine learning assisted sentiment analysis. In *Proceedings of International Conference on Computer Science & Engineering (ICCSE'2012)*, 123-130.
15. Liu, J., & Seneff, S. (2009). Review sentiment scoring via a parse-and-paraphrase paradigm. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 161-169.
16. Norvig, P.R., & Intelligence, S.A. (2002). *A modern approach*. Prentice Hall.
17. Kushal, A. M., Chanda, G., Shrivastava, K., Gupta, M., Sanyal, S., Sriram, T.V.N., & Banerjee, S. (2003). Multilevel modelling and rendering of architectural scenes. *In Eurographics (Short Papers)*.
18. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis Foundations and Trends in Information Retrieval, *2*.
19. *Vapnik, V.N. (1995). The Nature of Statistical Learning Theory.*
20. Kim, S.M., & Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, 1-8.
21. Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *In Proceedings of the 45th annual meeting of the association of computational linguistics*, 440-447.
22. Kamps, J., Marx, M., Mokken, R.J., & De Rijke, M. (2004). Using WordNet to measure semantic orientations of adjectives. *In LREC*, *4*, 1115-1118.
23. Turney, P.D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.
24. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis *Foundations and Trends in Information Retrieval*, *2*(1-2), 1-135.