

# A STUDY ON TECHNIQUES AND ARCHITECTURE OF SENTIMENT ANALYSIS

**Surbhi Goyal**

*M.Tech. Student, Department of Computer Science and Applications  
Jagannath University, Jaipur*

## Abstract

*Today most of the products and the services are available online so that direct interaction between the service provider and the customer is established. Now the feedback of the products and the services is taken in the form of online reviews. Sentiment Analysis is the technique which comes under opinion mining to analyse these reviews and concludes the review sentiment. In this paper, a study on different types of sentiment analysis techniques is defined along with the exploration of the architecture of sentiment analysis.*

*Keywords: Sentiment Analysis, Emotion, Feature Analysis*

## I. Introduction

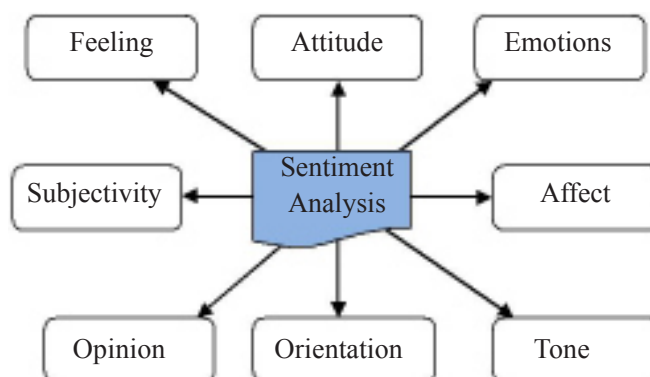
Sentiments or the emotions are the expressive information that are generated itself when a human presents information in different forms. The facial expression, human gesture, speech utterances are some of the common aspects of human information expression. Today, the user is very much involved with the worldwide web especially with regard to social media. While posting blogs or tweets the user also includes his emotions and expressions in the form of text and icons. These emotions or expressions are based on the word or the phrase selection done by the web user. Some of the web service providers or the companies also accept the user feedback or the review to analyze the popularity of the site, product or the service. Once this kind of textual review is accepted by the service provider, an analysis can be performed under different dimensions based on the type of information incorporated in it. These features and the feature dimensions can vary according to the application areas [1,2,7,9]. Some of the common forms of emotions represented by different researchers along with impact classes are listed in table 1.

**Table 1: Emotion Classes**

<b>Emotion</b>	<b>Impact Class</b>
Excited, Awesome, Elated, Enthusiastic, Strong, Wonderful Wow, Warmhearted	Strongly Positive Affect
Aroused, Astonished Surprised	Strongly Engaged

Content, Happy, Kindly Satisfied, Pleased	Pleasantness
Dull, Drowsy, Sleepy Fine, Better, Ok	Low Positive Affect
At rest, calm, relaxed Placid	Low Negative Affect
Quiescent, Quiet Still	Disengagement
Blue, grouchy, lonely Sad, sorry, unhappy not good, bad	Unpleasantness
Worst, scornful, bitter, hostile, fearful, distressed	High Negative Affect

These emotion or expression classes are used to identify user interest in the product or the service so that the service/product provider can take requisite action to increase his sales. The identification of the emotion class from the review, post or the feedback is called sentiment analysis. Sentiment analysis is one of the key emerging research areas that include the integrated concepts of text mining and natural language processing. The sentiments or emotions are observed under different aspects shown in figure 1. The high level classification of sentiment analysis is considered in two main aspects called the positive and the negative sentiments. There are number of existing approaches to perform the topic oriented classification [12,13].



**Figure 1: Different Aspects of Sentiments**

Two major categories of sentiment classification are binary sentiment classification and multi-class classification. The binary classifications identify sentiments in terms of positive and negative classification. In case of multi-class classification, the strength of classification category is also identified.

## Opinion Mining

Opinion Mining is the process of text mining to arrive at some conclusions from the reviews or feedbacks retrieved from users or customers. The opinions are generally taken to collect the thoughts on various aspects of the product including technical details. These opinions or feedback can be taken from the expert reviews as paid review because free users generally comment with less commitment. Once the opinions are collected, the filtrations over these reviews are performed to identify the most relevant information in the form of opinion class or the sentiment class[14][15]. In this section, the study and the analysis of the sentiment analysis is defined. Different aspects, classes and application areas are explored. In section II, the work done by the earlier researchers in the area of sentiment classification is discussed and explored. In section III, a standard model for sentiment classification is presented. In section IV, the different classification approaches for sentiment classification are discussed.

## II. Existing Work

Sentiment classification is one of the broad segments used by many applications in different areas. One of the major application areas is the analysis of movie reviews done by different web users. Some of the work done by earlier researchers in the area of sentiment identification based on movie review analysis is being shown in this section. In year 2006, Qiang Ye et al. has used the semantic analysis based approach to perform sentiment classification for Chinese movie reviews. The author defined an automated system for opinion mining under two main classes using the machine learning algorithm. Semantic rules are adapted by the researcher to perform sentiment analysis and classification [1]. Another semantic analysis based intelligent work on sentiment extraction was defined by Richard Colbaugh et al. in year 2010. Authors performed lexicon based analysis on social media contents on user posts and handled many challenges while performing the web information extraction. The commutation approach presented in this work includes the text information extraction. The computation is performed on the word information and the knowledge extraction process. Author also presented a case study based on the related work to show the effectiveness of the work [2]. An adaptive model based on probabilistic analysis for sentiment extraction was presented by Xiaohui Yu in year 2010. This model was utilised to analyze sales performance and increase sale performance. This adaptive model extracts the sentiment features from the review analysis and enhances the data capabilities so that the performance analysis can be done effectively. The work includes the case study performed on movie domain so that system performance can be improved [3].

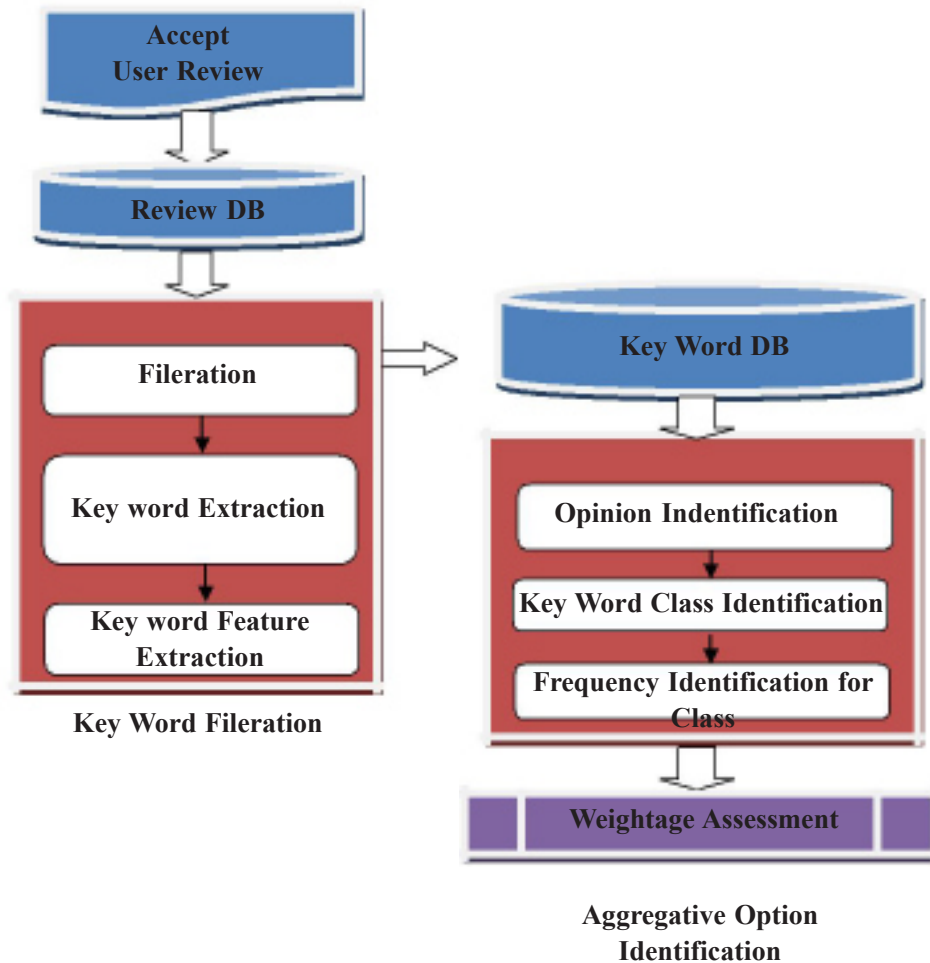
A case study based work on review mining was performed by Xiaohui Yu et al. in year 2012. The author presented the work on movie domain to increase sales performance. The researcher performed the statistical rule mining along with word ranking and the recommender system. The characterization is performed under time series analysis so that the analysis set will be reduced and optimization will be obtained. The author also presented an experimental case study which shows the accurate sentiment classification [4]. In year 2012, Chien Liang Liu et al. defined a review summarization based analysis and the rating identification approach based on feature extraction. Author used the SVM classification approach to perform level 1 classification and later on polarity classification was utilised to generate the movie ranking. The review classification was implemented on extracted movie features so that preferred aspect based analysis was performed [5].

In year 2011, Kristin Glass et al. performed social media content extraction for sentiment analysis to identify security aspects. Author presented a text classification model using a bipartite graph to perform work on words and documents. Later, the sentiment labels were applied over it to perform effective classification. The researcher defined a conjugate gradient method to present a semi-supervised classification approach [6]. Another method on sentiment analysis was presented by Hasan et al. In this method, proximity patterns were identified along with proximity types and the relation. This extracted information was then classified into different sentiments using supervised and statistical methods. The author showed that statistical learning is more accurate in terms of results [7]. Hogenboom et al. presented feature strength analysis approach under different aspects to automate sentiment analysis. The author used negation based approach to identify the influence of the scope [8].

Ming Hao has presented a topic based twitter post analysis for sentiment detection. Author defined this analytical approach under different vectors such as keyword density, stream analysis and influence. Author presented a cell based structured analysis to identify the influential opinion. The author defined the results in the visual analysis by area wise segmentation so that the sentiment relation to the specific population area was identified [9]. Chales B. Ward et al. has defined a framework for the text monitoring and sentiment signal analysis for real world classes to perform text classification. Author performed the phrase level and document level analysis so that large scale forecasting will be performed. Author standardized the work under different application areas such as business, stock market, movies etc. [10]. Another automated work on sentiment analysis was performed by Martin Wollmer et al. for audio visual context evaluation. Author defined the annotation analysis approach and performed summarization to reduce feature size. This hybrid feature analysis was classified using SVM approach and prediction approach to predict the sentiment [11]. A textual review analysis and classification approach was presented by Mothami [12] using SVM approach. The author used the multi theme document for the analysis and performed layered analysis.

### **III. Reserch Methodology**

In this section, the standard architecture for the sentiment analysis has been defined. The reliability of any data mining operation depends on the dataset itself. In this kind of mining, the data is collected from end users that include general information as well as technical. To perform the effective sentiment analysis, multiple reviews are considered. The number of review considerations also defines the accuracy obtained from the review process. The basic architecture of the sentiment analysis is shown in figure 2.



**Figure 2: Sentiment Analysis Framework**

As shown in the figure, once the reviews are collected from users or the customer, a review database is generated. Now filtration on this review DB is performed. At the initial stage, the keyword extraction from the reviews is derived. This extraction process includes the elimination of the stop words and identification of the keywords. From these keywords, the adjectives and the other opinion oriented keywords are identified. After this stage, filtered opinion keywords dataset is generated.

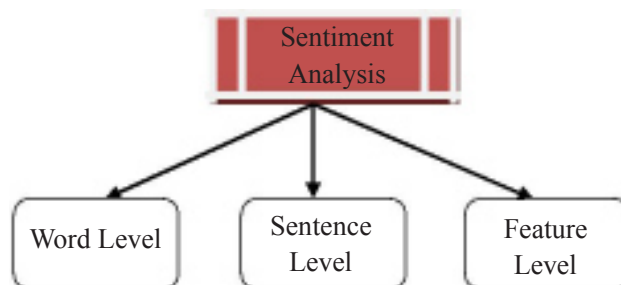
At the second level, the work on this keyword database is performed. At this stage, the opinion identification is defined as the combination of three main stages. In the first stage, the opinion class is identified for each keyword. Later, number of opinion keywords in each class is identified and based on this, the weight age is assigned to each belonging opinion class. An opinion class that will have more number of keywords will be assigned by higher weightage. Finally, the aggregative weightage is applied on each opinion class and the collective decision is taken. This decision is considered in the form of opinion class identification as the final result.

#### IV. Types of Sentiment Analysis

Sentiment analysis is basically used to identify the conclusion of any review or the feedback by performing the sentiment encoding. This analysis is generally based on a single word, adjective or phrase. The example of the sentiment analysis is listed here under

*“Review: The Movie screenplay was awesome”*

Here “awesome” defines that the movie review is positive and that the reviewer liked the film. But this single word phenomenon is always not accurate, because some of the technical reviews are multiline and having different aspects about different movie qualities. These qualities include the direction, music, screenplay, star cast, story etc. Each category is then analyzed under the respective review class. Once the initial pre-processing is done, there are number of approaches to perform the sentiment analysis. The most common categories of review analysis are shown in figure 3.



**Figure 3: Type of Sentiment Analysis**

##### A) Word Level Sentiment Analysis

It is mostly used as an effective sentiment analysis technique. An effective encoding is done between the sentiment words and the class. Such as

*(Brilliant, Awesome, Very Good) => Positive Sentiment*

*Sentiment Words*

*Sentiment Class*

There are number of databases that represent the adjective and the respective class. This adjective extraction comes under the lexical analysis of the review. The class generation is a kind of clustering. In the general form, two classes are formed to identify the positive and the negative reviews. The reliability of this approach depends on the adjective or the sentiment word set. The word set must include all the synonyms and antonyms relative to the word. Here the synonym represents the positive sentiments and the antonym represents the negative sentiments.

##### B) Sentence Level Sentiment Analysis

In this sentiment analysis approach, different levels of granularity are analyzed over the review. A rule based analysis is required to perform the sentence based sentiment identification. These rules include the negation rules extraction approach. It means the sentence or the review including negative words such as no, not and

never are used to represent the negative perspective of the sentiment. Some of the verbs that represent a negative sense also represent the negative reviews such as “stop”, “problem” etc. These verbs are also analyzed in different verb forms and combinations. Some of the examples of the sentence level sentiments are given in table 2.

**Table 2: Sentence Level Sentiment**

Type	Example	Sentiment
Negation Negative	No Problem	Positive
Negation Positive	Not Good	Negative
Negation Neutral	Will not work	Negative

### C) Feature Level Sentiment Analysis

It is one of the most intelligent analyses of movie reviews. This analysis process defines feature identification from the review. This feature is compared from the review set and based on which the orientation score is identified. Each positive feature is assigned with positive weightage and the negative feature is assigned with negative weightage. Once all the features are collected, the aggregation on the weightage is performed to identify the overall featured score. If the score is positive, the review is considered positive otherwise it is considered negative. The feature analysis approach is based on the statistical or the mathematical formula based on which overall prediction of the sentiment feature is carried out.

## IV. Conclusion

In this paper, a study oriented work is outlined with regard to sentiment analysis. This analysis process is defined in terms of different approaches for sentiment analysis. The work also includes the exploration of sentiment analysis.

### Works Cited

- [1] Q. Ye, W. Shi and Y. Li, “Sentiment Classification for Movie Reviews in Chinese by Improved Semantic Oriented Approach”. Proceedings of the 39th Hawaii International Conference on System Sciences, pp 1-5, 2006.
- [2] R. Colbaugh and K. Glass. “Estimating Sentiment Orientation in Social Media for Intelligence Monitoring and Analysis”, International Conference on Intelligence and Security Informatics, pp 135 - 137, 2010.
- [3] X. Yuy, Y. Liu and A. Anz, “An Adaptive Model for Probabilistic Sentiment Analysis,” IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp 661-665, 2010.
- [4] X. Yu, Y. Liu, J. X. Huang and A. An, “Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain”, IEEE Transactions on knowledge and Data Engineering Vol. 24, No. 4, April 2012.

- [5] C Liu, W. Hsaio, C. Lee, G. Lu, and E. Jou, Movie Rating and Review Summarization in Mobile Environment, IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, Vol. 42, No. 3, May 2012.
- [6] K. Glass and R. Colbaugh, Estimating the Sentiment of Social Media Content for Security Informatics Applications, IEEE International Conference on Intelligence and Security Informatics, pp 65 - 70 , 2011
- [7] S.M. S. Hasan and D. A. Adjero, Proximity-Based Sentiment Analysis, 4<sup>th</sup> International Conference on Applications of Digital Information and Web Technologies, pp 106 - 111, 2011
- [8] A. Hogenboom, P. V. Iterson, B. Heerschop, F. Frasinca, and Uzay-Kaymak, Determining Negation Scope and Strength in Sentiment Analysis, pp 2589 - 2594 , 2011.
- [9] M. Hao, C. Rohrdantz, H. Janetzko and U. Dayal, Visual Sentiment Analysis on Twitter Data Streams, IEEE Conference on Visual Analytics Science and Technology, pp 277 – 278, 2011
- [10] C. B. Ward, Y. Choi, S. Skiena and E. C. Xavier, Empath: A Framework for Evaluating Entity-Level Sentiment Analysis, International Conference & Expo on Emerging Technologies for a Smarter World, pp 1 - 6, 2011.
- [11] M. Wöllmer, F. Weninger, T. Knaup, and B. Schuller, YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context, IEEE Intelligent Systems, pp 46 – 53, 2013.
- [12] K. Mouthami, .K. Nirmala Devi and V. M. Bhaskaran, Sentiment Analysis and Classification Based On Textual Reviews, International Conference on , pp 271 – 276, 2013.
- [13] V.K. Singh, R. Piryani, A. Uddin and P. Waila, Sentiment Analysis of Movie Reviews and Blog Posts, IEEE International Advance Computing Conference, pp 893 – 898, 2013.
- [14] S. Kamath, S A. Bagalkotkar, A. Khandelwal, S. Pandey and K. Poornima, Sentiment Analysis Based Approaches for Understanding User Context in Web Content, International Conference on Communication Systems and Network Technologies, pp 607-611, 2013.
- [15] V.K. Singh, R. Piryani, A. Uddin and P. Waila, Sentiment Analysis of Movie Reviews, International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing, pp 712 - 717, 2013.