

# International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: [www.ijarcsms.com](http://www.ijarcsms.com)

## *An Application for Sentiment Analysis Based on Expressive Feature in the Sentence*

**Shital S. Dabhade<sup>1</sup>**

CSE Department  
Abha Gaikwad – Patil College of Engg  
Mohgaon, Nagpur - India

**Prof. Sonal S. Honale<sup>2</sup>**

CSE Department  
Abha Gaikwad – Patil College of Engg  
Mohgaon, Nagpur - India

*Abstract: Our daily life has forever been influenced by what folks suppose. concepts and opinions of others have forever affected our own opinions, because the net plays associate degree more and more important role in people's social lives, it contains additional and additional data regarding their opinions and sentiments. The distillation {of knowledge of data of data} from this immense quantity of unstructured information, also called opinion mining and sentiment analysis. Nowadays, with the speedy evolution of good phones, mobile applications (Mobile Apps) became essential components of our lives. However, it's tough for customers to stay track and perceive the app sphere as a result of new apps square measure coming into market a day. Such an outsized quantity of apps appears to be an excellent chance for patrons to shop for from a large choice varies. But, initial they need to know what the apps do, however square measure they viewed by different customers then they need to get the apps to use on their good phones. Typically, on-line client reviews contain 2 components, ratings and matter comments. Rating indicates the analysis of client experiences employing a numeric scale; however matter comments square measure capable of telling additional perceptive stories that the ratings cannot. it's terribly difficult for a possible user to scan all of them to create a choice. Also, app developers have difficulties find out a way to improve the app performance supported overall ratings alone and would profit by understanding the thousands of matter comments. In planned approach, user reviews square measure summarized and options square measure extracted from the apps mentioned within the reviews. Then, IP approach is employed for writing rules then sentiment analyzer is employed for the analysis of sentiments that square measure arising through the matter comments of users.*

*Keywords: sentiment analysis, nlp approach, Machine Learning, android Apps.*

### I. INTRODUCTION

Sentiment Analysis may be a linguistic communication process and knowledge Extraction task that aims to get writer's feelings expressed in positive or negative comments, queries and requests, by analyzing an outsized numbers of documents. typically speaking, sentiment analysis aims to work out the perspective of a speaker or a author with regard to some topic or the musical notation of a document. In recent years, the exponential increase within the net usage and exchange of belief is that the propulsion behind Sentiment Analysis these days. the net may be a immense repository of structured and unstructured information. The analysis of this information to extract latent belief and sentiment may be a difficult task.

The analysis of sentiments is also document primarily based wherever the sentiment within the entire document is summarized as positive, negative or objective. It will be sentence primarily based wherever individual sentences, bearing sentiments, within the text square measure classified. reserves will be phrase primarily based wherever the phrases during a sentence square measure classified in step with polarity. Sentiment Analysis identifies the phrases during a text that bears some sentiment. The author could discuss some objective facts or subjective opinions. it's necessary to tell apart between the 2. reserves finds the topic towards whom the sentiment is directed. A text could contain several entities however it's necessary to seek out the entity towards that the sentiment is directed. It identifies the polarity and degree of the sentiment. Sentiments

square measure classified as objective (facts), positive (denotes a state of happiness, cloud nine or satisfaction on a part of the writer) or negative (denotes a state of sorrow, dejection or disappointment on a part of the writer). the feelings will additional tend a score supported their degree of quality, nega or sound judgment.

Generally speaking, sentiment analysis aims to work out the perspective of a speaker or a author with regard to some topic or the discourse polarity of a document. The perspective is also his or her judgment or analysis, affectional state (that is to mention, the emotion of the author once writing), or the supposed emotional communication (that is to mention, the emotional impact the author needs to possess on the reader). Basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level — whether or not the expressed opinion during a document, a sentence or associate degree entity feature/aspect is positive, negative, or neutral. Advanced, "beyond polarity" sentiment classification appearance, as an example, at states like "angry," "sad," and "happy".

Natural Language process may be a in theory motivated vary of process techniques for analyzing and representing present texts at one or additional levels of linguistic analysis for the aim of achieving human-like language process for a spread of tasks or applications. many components of this definition square measure as follows. first of all the inaccurate notion of 'range of process techniques' is critical as a result of there square measure multiple strategies or techniques from that to settle on to accomplish a specific style of language analysis. 'Naturally occurring texts' will be of any language, mode, genre, etc. The texts will be oral or written. the sole demand is that they be during a language employed by humans to speak to at least one another. Also, the text being analyzed shouldn't be specifically made for the aim of the analysis, however rather that the text is gathered from actual usage.

The notion of 'a level of linguistic analysis' refers to the very fact that there square measure multiple forms of language process well-known to be at work once humans manufacture or comprehend language. it's thought that humans unremarkably utilize all of those levels since every level conveys differing types of which means. however varied IP systems utilize completely different levels, or mixtures of levels of linguistic analysis, and this is often seen within the variations amongst varied IP applications. This additionally ends up in abundant confusion on the a part of non-specialists on what IP very is, as a result of a system that uses associate degreey set of those levels of research will be same to be an NLP-based system. The distinction between them, therefore, may very well be whether or not the system uses 'weak' IP or 'strong' IP. 'Human-like language processing' reveals that IP is taken into account a discipline among AI (AI). And whereas the complete lineage of IP will rely upon variety of different disciplines, since IP strives for human-like performance, it's acceptable to think about it associate degree AI discipline. 'For a spread of tasks or applications' points out that IP isn't sometimes thought-about a goal in and of itself, except maybe for AI researchers. For others, IP is that the means that for accomplishing a specific task. Therefore, Information Retrieval (IR) systems square measure only if utilizes IP, likewise as artificial intelligence (MT), Question-Answering, etc.

## II. RELATED WORK

Jiawen Liu, Mantosh Kumar Sarkar and Goutam Chakraborty gave the IP rule-based model for humanoid apps ;it is rigorously designed in SAS® Sentiment Analysis Studio twelve.1 for predicting sentiments in check information. Tianxi Dong, Jonghyun Kim, Zhangxi Maya Lin decribes the text analysis provided in SAS® Text manual laborer to predict the and feature-based ratings for on-line application reviews. Alexandre Trilla and Francesc Alías, they need analysis to enhance the state of the art Text-To- Speech (TTS) synthesis studies each the process of input text and also the ability to render natural communicative speech. M. Sakthivel, G. Hema describes that sentiment analysis applies linguistic communication process techniques and linguistics to extract data regarding sentiments expressed by authors and readers a couple of specific subject, therefore serving to users in creating sense of big volume of unstructured net information. Erik Wales, Björn Schuller, Catherine Havasi gave the 2 fields that square measure sometimes combined beneath an equivalent umbrella or maybe used as synonyms. each fields use {data mining data process} and linguistic communication processing (NLP) techniques to find, retrieve, and

distill data and opinions from the globe Wide Web's large matter data. Changbo Wang, Zhao Xiao, Yuhua Liu, YanruXu, Aoying Chou, and Kang Zhang they introduced a replacement image system for analysing, visualizing and confirmatory the feelings of net users on public topics. They used a text-based sentiment mining technique and a model-driven prediction approach to investigate the general public sentiments on hot topics. Akshi Kumar, Teeja Madonna Sebastian this paper illustrates the analysis space of Sentiment Analysis and its latest advances. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and mineral Jou during this paper the author describes Sentiment classification is applied to the movie reviews, and rating information is based on sentiment-classification results. MaralDadvar, Claudia Hauff, Franciska de Jong they investigate the problem of determining the polarity of sentiments in movie reviews when negation words, such as not and hardly occur in the sentences. They examine how different negation scopes (window sizes) affect the classification accuracy.

### III. METHODOLOGY

The basic idea of the proposed system is to develop natural language based Android App Polarity analysis system for reviews which is capable of evaluating the reviews and provides result in appropriate form. Proposed architecture for this research work is as shown in Figure 1.

#### A. REVIEWS COLLECTION

The reviews are collected from the google app store and real time reviews are also taken by providing the free space form to the user. The reviews from the real time datasets are grouped under training datasets and test datasets.

#### B. POLARITY, FEATURE IDENTIFICATION

In this step the review is taken as an input. The inputted review is considered one by one for the polarity identification step. The purpose of this step is to filter all the unrequired words that are neither related to sentiment analysis nor feature analysis. Here, the positive and negative words from the considered reviews are identified. The identification step is considering the reviews and with the help of sentence splitter, splitting the review one by one as an input.

Parsing: Parsing means identifying the words and grouping them according to the need of the application or system. The reviews are parsed one by one for identifying the polarity and the feature. The parsing process provides us with the array form that is helpful for computational purpose (coding). The reviews are broken down in the form of a single array.

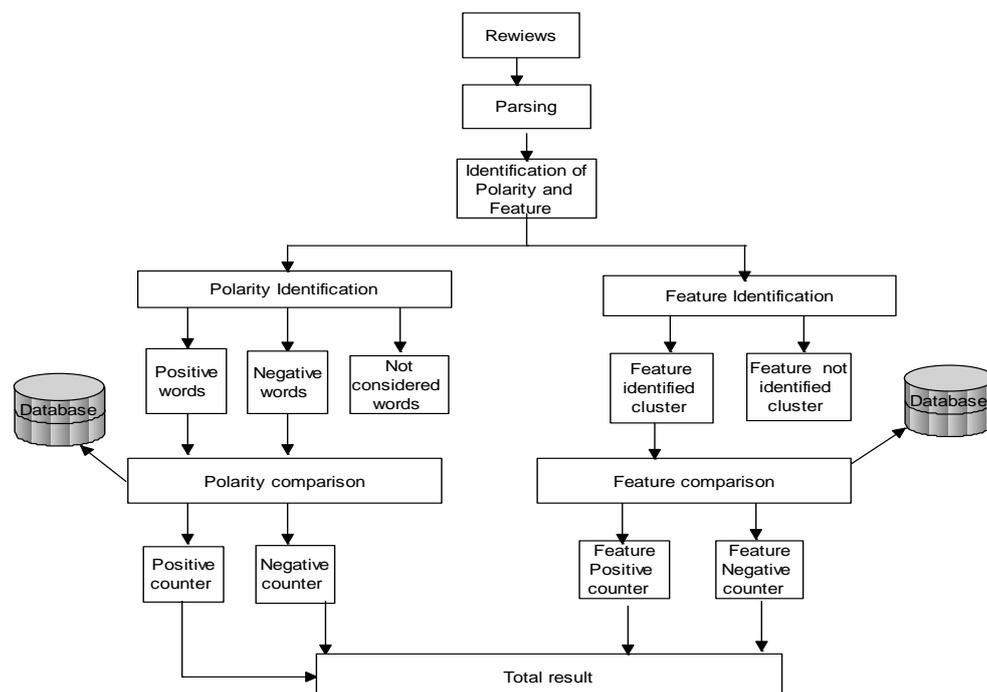


Fig1. Proposed System

### C. POLARITY COMPARISON

The above identification output provides us with the important polarity and the feature words only. Now in this step the above output polarity and feature words are compared with the help of rule based approach. The rule based approach deals with the designing of the rules by observing the review data. Generally the five rules have been defined by observing the common pattern in the reviews. The rule based approach works with comparing the identified word with the defined datasets with the help of word matching process. In this process the identified words are matched with the dataset and only complete word matching are allowed for the purpose of comparison. The words with short cuts and wrong spell are not considered for the performance of the system. Example: The input words “very good” from the above step is compared with all the polarity dataset. Now at the time of comparison there are various words that increase the polarity of the sentences but, a drawback is that these words should come just before the polarity word. These words are known as the polarity enhancement word (i.e. very, too, so .....etc). For comparing the identified words (feature, polarity) the comparing algorithms that is used is as follows:

The first algorithm used explains the polarity word comparison from the dataset that are available. Here the rules are checked whether the review contains very not, not, very along with positive or negative word or simply positive or negative words and then accordingly positive and negative counters are increased or decreased. The second algorithm checks the feature present in the reviews and then accordingly increases or decreases the counters of the cluster head’s positive or negative respectively and if it is not present then it returns back to other review present.

### D. DATASET

For the above comparison purpose we required some datasets that will help in classifying the polarity in terms of positive and negative and classifying only related features of the app. The proposed system consists of 4 dataset they are as follows:

1. Positive Dataset: This dataset is created for comparing all the positive words. In this dataset all the positive words as per our knowledge are kept in this dataset. The dataset can be incremented and decremented as per the words are known.
2. Negative Dataset: This dataset is created for comparing all the negative words. In this dataset all the negative words as per our knowledge are kept in this dataset. The dataset can be incremented and decremented as per the words are known.
3. Dictionary: The Dictionary consist of all the positive and negative words as per our knowledge mark with the tag words then this tag words are searched in the above datasets(positive / negative)for comparing the polarity of the words.
4. Feature Clustering Dataset: In this dataset, features are clustered in the form of clustering approach. The clustering approach works in the following manner.

- Deciding the cluster head (Gui, Performance, Security)
- Distributing the features in the above cluster head group
- Comparing the feature clustering dataset with the feature words.

### E. POLARITY RESULT

The rules efficiency are tested by finding out the polarity words manually and comparing both the results a near to some value is obtained. The polarity results provides us with the positive , very positive , negative , very negative counts of the words present in the inputted reviews.

### F. FEATURE CLUSTER IDENTIFICATION

The identified feature is compared with the feature cluster dataset. The feature is correctly recognized and the polarity of the feature is calculated by using the polarity comparison algorithm. The counters of the feature are incremented and decremented based on the polarity words.

### G. FEATURE COMPARISON

The features are compared with the feature dataset by using the above feature comparison algorithm. Based on that polarity is identified. The feature that are not matched with the feature dataset having the problem of feature absentee or spell mistake.

### H. FEATURE RESULT

All the identified features with the polarity count are calculated and the individual cluster head counts are taken out for judging the feature capability in the android app.

### I. TOTAL RESULT

The two steps first polarity result and second feature result are combined together to get the result for the user as well as developer. The result facilities user to correctly judge which app to used and which to not. The developer by seeing the result are in a state what the problem is exactly in the system and which features should be improved in order to make better performance of the app.

## IV. EXPERIMENTAL RESULT ANALYSIS

The results of the opinion analysis are computed by using the parameters precision and recall. Precision is one measure of the effectiveness of some computer applications for finding search words, candidate terms, and other items. Precision is a measure of the proportion of the results of a computer application that are considered to be pertinent or correct. Recall is one measure of the effectiveness of some computer applications for finding search words, candidate terms, and other items. Recall is a measure of the proportion of all possible correct results of a computer application that the application actually produces.

For example, suppose you are using a computer application to search for terms in a document that has 80 terms in it. (You know because you counted them.) If the application finds 55 of these terms, then the recall of the application is 55 out of 80, or 0.62. In this system these two parameters are evaluated for checking the accuracy of the system. The rule based approach of the system is tested with 150 reviews of the three apps and the following frequencies of the rules are calculated for simplification the results are manually checked with the minor changes. Table 1 represents the rule based frequency approach in the three apps.

Table 1: Rule based frequency calculation

Sr no	Rules (Frequency)	Angry birds	Whas App	Toi
1	Very Positive	11	4	4
2	Positive	22	27	20
3	Very Negative	12	7	3
4	Negative	14	12	26

The precision ratio of the correct and the incorrect words are calculated and shown in the table no. 2. The graph represents the precision value of the three testing app. It can be observed that the polarity word sequence is near to same in all the three apps review.

Table 2: Polarity precision calculation

Sr no	App Name	Correct Word	Incorrect Word	Precision
1	Angry Birds	58	9	86%
2	Whats App	50	10	83.33%
3	Toi	53	11	82.81%

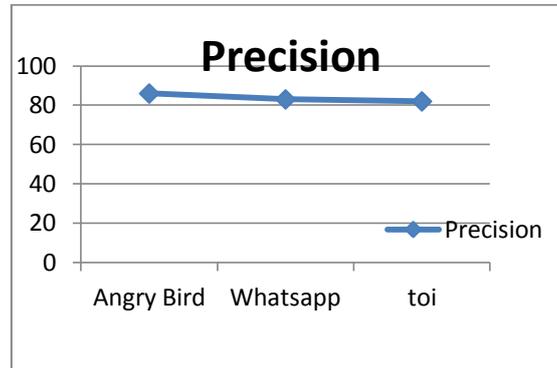


Figure 2: polarity precision graph

The app features are identified and to check which sentence contains the feature part are only considered in the feature comparison phase. So for this the recall value will provide us with the actual sentence provide with the feature. The Table 3 shows the sentence that contains feature and the recall value is calculated.

Table 3: Feature Recall calculation

Sr. No	App name	Feature considered sentence (correct+incorrect)	Feature not considered sentence	Recall
1	Angry birds	28	50	56%
2	Whats App	28	50	56%
3	Toi	29	50	58%

The feature precision ratio of correctly identifying the feature cluster is considered for the three apps and the precision value is calculated in the Table 4. The figure3 provides us with the precision graph that shows that the system is able to correctly identify the feature in the cluster head.

Table 4: Feature precision calculation

Sr. No	App name	Feature considered correctly	Feature considered incorrectly	Precision
1	Angry birds	23	5	82.14%
2	Whats App	25	3	89.28%
3	Toi	26	3	89.65%

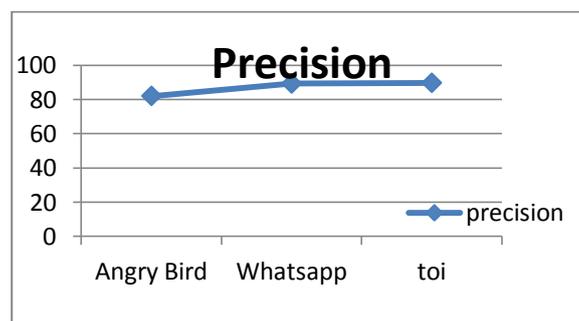


Figure 3: Feature precision graph

## V. CONCLUSION

The proposed system is capable of 1) Providing user with space to write their own reviews, 2) Determining the polarity and features of the app based on the considered reviews, 3) Providing the user with the accurate result generated by the proposed system, 4) Providing the app developer with the list of the features (Positive/Negative aspects) in order to improve the app performance all these objectives that are mentioned in the problem definition. The system is evaluated by the precision and recall parameter of the polarity as well as feature in the result part. The result is sufficient up to the mark for judging the Android App and the developers are also able to predict the problem and the improvement needed in the app for its popularity within less time.

## References

1. Jiawen Liu, Mnatosh Mantosh Kumar Sarkar and GoutamChakraborty, "Feature-based Sentiment Analysis on Android App Reviews Using SAS® Text Miner and SAS® Sentiment Analysis Studio," SAS Global Forum 2013.
2. Tianxi Dong, Jonghyun Kim, Zhangxi Lin, "Predicting Application Review Rating with SAS® Text Miner," Paper 269-2013, SAS Global Forum 2013.
3. Alexandre Trilla and Francesc Alias, "Sentence-Based Sentiment Analysis for Expressive Text-to-Speech," IEEE Transactions on Audio, Speech, And Language Processing, Vol. 21, No. 2, February 2013.
4. M. Sakthivel , G. Hema, " Sentiment Analysis Based Approaches for Understanding User Context in Web Content," IJCSMC, Vol. 2, Issue. 7, July 2013, pg.231 – 239.
5. Erik Cambria, BjornSchuller, Catherine Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," IEEE Transactions on Opinion mining, 2013.
6. LisetteGarcía-Moya, HenryAnaya-Sánchez, and Rafael Berlanga-Llavori, "Retrieving Product Features and Opinions from Customer Reviews," IEEE Transactions on Opinion mining 2013.
7. Changbo Wang, Zhao Xiao, Yuhua Liu, YanruXu, Aoying Zhou, and Kang Zhang, "SentiView: Sentiment Analysis and Visualization for Internet Popular Topics," IEEE Transactions on Human-Machine Systems, Vol. 43, No. 6, November 2013.
8. Akshi Kumar, Teeja Mary Sebastian, "Sentiment Analysis: A Perspective on its Past, Present and Future," IJ. Intelligent Systems and Applications, 2012, 10, 1-14 Published Online in MECS September 2012.
9. G. Vinodhini, RM. Chandrasekaran, "Sentiment Analysis and Opinion Mining," IJARCSSE Volume 2, Issue 6, June 2012.
10. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery 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.
11. Maral Dadvar, Claudia Hauff, Franciska de Jong, "Scope of Negation Detection in Sentiment Analysis," <http://wordnet.princeton.edu/>, October 2010.
12. JeongheeY, Tetsuya Nasukawa, Razvan ,Bunescu, Wayne Niblack, "Sentiment Analyzer:
13. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval,vol. 2, nos. 1–2, 2008, pp. 1–135.
14. B. Liu, Sentiment Analysis and Opinion Mining, Morgan and Claypool, 2012.
15. E. Cambria and A. Hussain, Sentic Computing: Techniques, Tools, and Applications, Springer, 2012.