

Semantic Analysis and Report Generation for Social Content

Ms. Sheela A Badagi^[1], Mrs. Shanta Kallur^[2]

Department of Computer Science & Engineering
K.L.E.Institute of Technology, Hubballi-30
Karnataka - India.

ABSTRACT

With rapid development of e-commerce the merchants selling the products, later ask their opinions. Most customers express their opinions on various kinds of entities like products/ services. Extracting the useful insights from internet based application like facebook helps to create and exchange of user generated content to communicate with others. The reviews have rich source of information for decision making and sentimental analysis. Sentimental analysis refers to a classification problem where the main focus is to predict the polarity of words and classify them into positive, negative and neutral feelings, with the aim of identifying attitude and opinions. In this paper, we develop a web application which will derive social content for a company, service or a product to take business decisions on semantic analysis.

Keywords: — *Semantic analysis, Opinion mining, Sentimental classification, Polarity, Social Media.*

I. INTRODUCTION

Semantic analysis is the study of language, it is the process of relating syntactic structure. The idea of transformational generative grammar is introduced by the syntactic structures. This method uses some of the phrase rules which can break down the sentences into more than one parts, and its level of clauses, sentences and paragraphs to the level of writing as whole to their language independent language. Social media or social content is computer mediated technologies that helps to creation and sharing of information, ideas and other forms of expression via networks.

Now a days, More and more products are sold on the web due to rapid expansion of e-commerce. So it has become common approach for merchants to enable their customers to review and give suggestions (opinions) on the products to which they are purchased [1]. If each product have a more number of comments then it may difficult for manufacturer to maintain the information ,so for that mining of the interested product features, and identifying the opinion of the sentence and display the result as positive, negative or neutral comments.

A. Related Work

Social media uses some of the concepts like computer science, data mining, machine learning, natural language processing and algorithms. Social media is the internet based application that helps to create and exchange the user generated content. It is easy for the users to communicate with each other. Some of the challenges to data mining are vast, noisy, unstructured and distributed type of data. Because in data mining the data is in the form of unstructured, noisy, distributed, it is not in a exact format which, the users are interested. Now a days data is everywhere, mining of

interested data is of a big task. The social media data is everywhere, overloaded in two forms,

- Information Overload
- Interaction Overload

Information overload includes photos, blogs. Interaction overload includes the comments, followers, posts. Data mining from the social media can help to get millions of users opinions and behaviours for product, i.e

- Product sentiments
- Human behaviour

With respect to particular product, the product contains the positive feedback from the users or negative feedback from the users else neutral feedback. Based on the reviews developers are going to improve their business. To enhance their business our project provides Semantic analysis and report generation for social content i.e

- Extract the data from social media (facebook), select the interested fields (posts, comments, likes), interpret the comments for individual page that is positive or negative or neutral, predict aggregation of comments and represent them by using pai chart.

Social media like facebook analyzer extract the facebook data by using the facebook graph API into web facebook entities ,i.e posts, likes, sentiments of comments.

B. Objectives

- Extraction of page data from social media like facebook by using Graph API.
- Tell the presence of the Facebook pages.
- Number of posts count, fans count, comments for a selected page.

- Number of comments count, shares count, likes count for a selected post.
- Classification of extracted data based on keywords.
- Extracting useful insights as sentimental analysis.
- Sentimental analysis of individual comment on page, that is positive, negative or neutral.
- Sentimental analysis of aggregation of comments on a page.

C. Aim Of The Proposed System



Fig. 1 Facebook Extracter

Figure 1 shows Facebook extractor. The facebook analyzer tool is used to extract the facebook page information by using the graph API.

- Extract data from the social media site Facebook.
- Analyzing the extracted data with TF-IDF vectorization and multinomial naive bayes algorithm.
- Presentation : get the users interacted information such as; number of posts ,fans count, comments for selected post, comments, shares, total number of likes per post , prediction of comment to tell positive, negative or neutral.

II. DATA EXTRACTION PROCESS

This process extracts the Data from the facebook by using graph API. Facebook graph API explorer is used to Test, create, and authenticate API calls and debug responses [4].

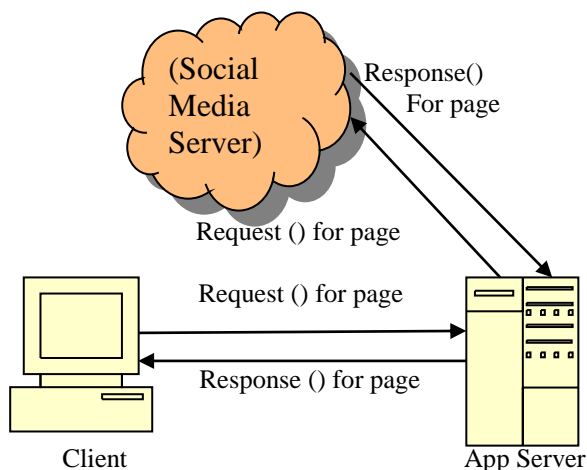


Fig. 2 Facebook analyzer

Figure 2 shows Facebook analyzer. The proposed system tells about over all objectives of this paper, system design contains 3 components as follows,

- Client
- Application Server
- Social media server to access facebook.

It involves some of the steps as follows,

- The client requests to application server to get page information.
- The application server requests for facebook server.
- Facebook server responses for requested info to Application server.
- Client gets the requested information from Application server.

The proposed system can help various business applications such as e-commerce application to get the money and to save the money.

D. Graph API

Facebook graph API can help to generate access token, by using this url,

- <https://developers.facebook.com/tools/explorer/145634995501895/>

E. Generation Of Access Token

To select access token within graph API explorer, select get user access token, user permissions like,

- User name
- User location
- User posts
- User likes etc..

F. Access Token Information

Access token gives information about, Application name.

- User of that application.
- Validation of that application interms of true and false.
- Expiration time of that application in Indian standard format.
- Scopes tell about the fields which are accessible by the user.

III. SENTIMENT ANALYSIS

Sentiment analysis (Opinion analysis) can be achieved with the help of Sentiment classification [5]. Emotions are the positive or negative reactions. Human communication is under the control of emotions and opinions [10]. While communicating, humans can express their opinions by face to face. But now a days, e-commerce growing faster, and all are using internet to share the information [6]. Sentiments can be classified based on,

- Emotions
- Keywords
- Sentiwordnet

In this paper, sentiments can be classified with the help of positive and negative keywords. Positive keywords are like, gud, super, excellent, efficient and so on. Negative keywords like bad, worst, cheat, not good, etc... by using the multinomial naive bayer's algorithm we can predict the polarity of the keywords is positive or negative. In other cases, the same keywords give the different meaning in different situation.

Example: Battery life of the Samsung cell is too long (positive comment).

Samsung mobile takes too long to start (negative comment).

G. Natural Language Processing(NLP)

It is the field of interaction between computer and human (natural) language [7].

NLP contains

- NLG(Natural language generation)
- NLU(Natural language understanding)

The natural language generation (NLG) converts computer database language to human readable form. The natural language understanding (NGU) is the identification of single semantic from multiple semantics by using NLP.

H. Uses Of NLP Algorithms

NLP algorithm based on machine learning algorithms, how the data can be analyzed more is the model accuracy. NLP algorithms are useful to create the advanced applications and prototypes.

- **Summarize blocks of text:** Ignoring irrelevant information, only extracting the useful information.
- **Identify the type of entity extracted:** Identify the named entity recognition.
- **Sentimental analysis:** Identify the sentiment of a string from text, that is positive, negative or neutral.

I. NLP Examples

- **Sentimental analysis:** Based on stand ford (tool) NLP can help to identify the opinion (positive, negative, neutral [9]).
- **Social media analysis:** Extract subjective information about particular object based on polarity of object, identify the trends of public opinion.

J. Multinomial Naive Bayers Algorithm

Naive bayers algorithm is used for text classification.

- For a document d and a class c, equation(1) gives the conditional probability is,

$$P(c/d)=P(d/c)p(c)/p(d) \dots\dots\dots(1)$$

$$C_{map}=\arg\max_{c \in C} P(c/d) \dots\dots\dots(2)$$

- All word positions in test document can be consider,

$$C_{NB}=\arg\max_{c_j \in C} P(c_j) \prod_{i \in C_{positions}} P(x_i/c_j) \dots\dots\dots(3)$$

With the help of Tf-Idf vectorizer and naive bayers algorithm we are going to classify the text within document.

K. Tf-Idf Vectorizer

Tf-Idf can be successfully used for stop-words filtering in various subject fields including text summarization and classification.

- Tf(Term Frequency) tells about the number of times a term occurs in a document is called its term frequency.
- Idf(inverse Document Frequency)tells about weight of terms that occur very frequently in the document[8].

Equation (4) gives the Term frequency weight is,

$$tf(t,d)=f_{t,d} \dots\dots\dots(4)$$

Equation (5) gives the Inverse document frequency weight is,

$$idf(t,D) = \log N / |d \in D : t \in d| \dots\dots\dots(5)$$

Where.

N-Total number of documents.

{d ∈ D : t ∈ d}-Number of documents where term appears.

IV. DISCUSSION AND RESULTS

Facebook analyzer is a web application. It searches and analyzes the facebook pages for entered keyword. If pages available then it displays all related pages. It also shows posts, fans count for that page. And for each post it displays number of comments, number of shares and count of likes.

Eg: Go to browser and enter the url as: localhost 5000.

Keyword entered is: Narendra Modi

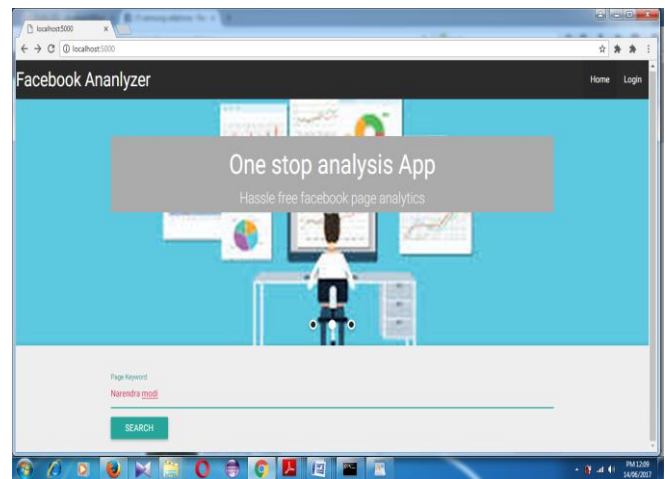


Fig. 3 Show Pages for entered keyword.

The Fig 3 shows the facebook analyzer, that searches the pages for entered keyword by using, any of the browsers.

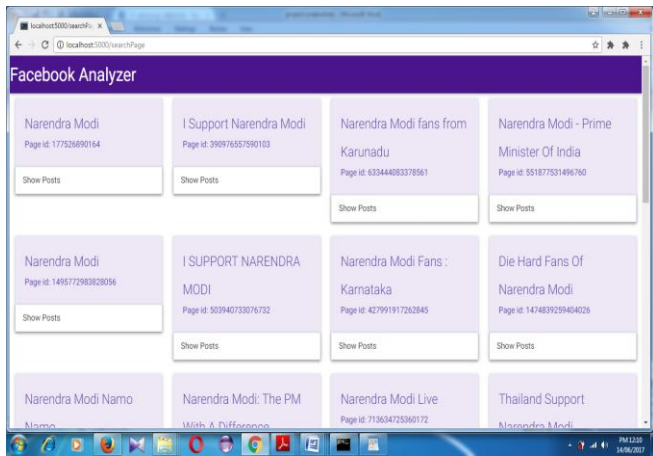


Fig. 4 Related pages for entered keyword.

The Fig 4 shows all related pages for given keyword Narendra modi. i.e, Narendra modi, I support Narendra modi, Narendra modi fans from karunadu so on.

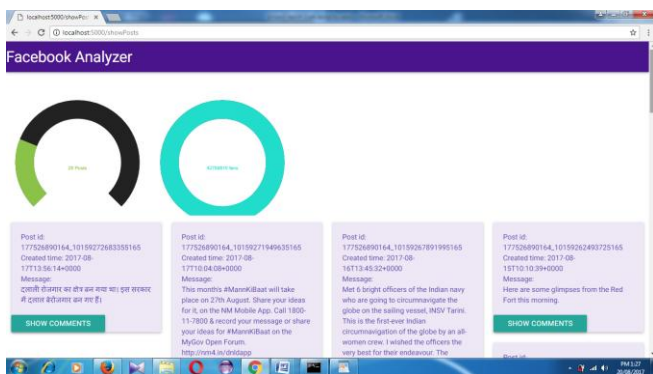


Fig. 5 Show posts, post counts, fans counts.

The above figure 5 gives the post count, fans count for that selected page. each post contains post id, created time and messages. The messages may be in all languages. But it can predict only English comments.

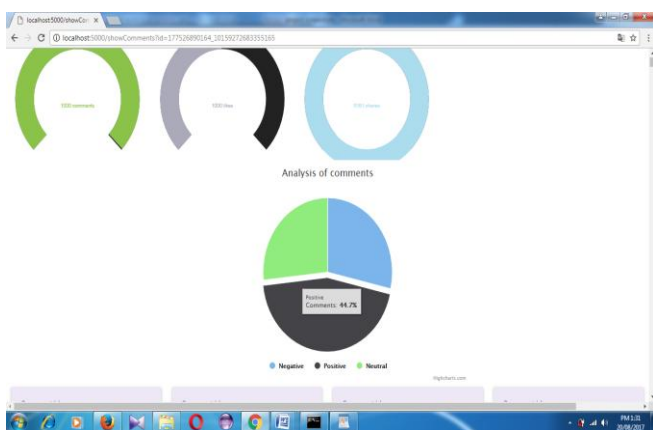


Fig. 6 Analysis of comments for a single post.

The above figure 6 gives the number of comments, shares, likes, along with, all comments can be displayed. Prediction of these individual comments helps to provide the sentimental analysis. Aggregation analysis of these comments can be reported by using chart.

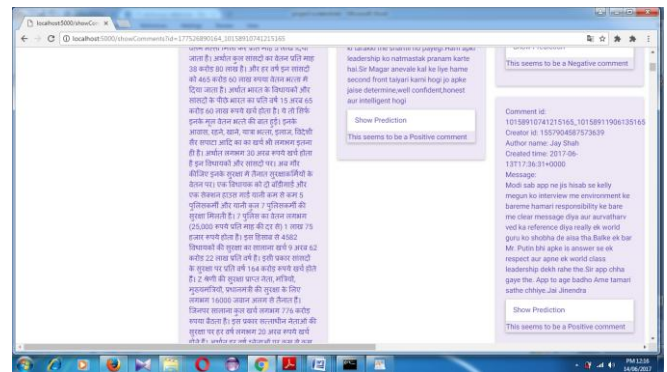


Fig.7 Prediction of single comment.

The above figure 7 shows the, Each comment can be predicted as positive, negative and neutral as shown above.

V. CONCLUSION

This project helps to analyze various user requests as positive, negative, neutral. It even aids, getting aggregation of comments, which may help to take a decision for a given query. This System provides system owner or user to predict upon future plan of trends.

REFERENCES

- [1] [1] Minqing Hu and Bing Liu Department of Computer Science University of Illinois at Chicago 851 South Morgan Street Chicago, IL 60607-7053 {mhu1, liub}@cs.uic.edu
- [2] https://en.wikipedia.org/wiki/Google_Analytics#Popularity
- [3] https://en.wikipedia.org/wiki/Social_media_analytics
- [4] <https://developers.facebook.com/tools-and-support/>
- [5] International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Volume 1 Issue 10 (November 2014)
- [6] International Journal of Computer Applications Technology and Research Volume 4– Issue 6, 419 - 424, 2015, ISSN:- 2319–8656
- [7] https://en.wikipedia.org/wiki/Natural_language_processing
- [8] <https://en.wikipedia.org/wiki/Tf%E2%80%93idf>.
- [9] International Journal of Advanced Research in Computer Science and Software Engineering (Volume 5, Issue 5, May 2015)

- [10] Information Sciences Institute University of Southern California 4676Admiralty Way Marina del Rey, CA 90292-6695 skim@isi.edu Information Sciences Institute University of Southern California 4676 Admiralty Way Marina del Rey, CA 90292-6695 hovy@isi.edu Proceedings of the COLING conference, Geneva, 2004