RESEARCH ARTICLE

OPEN ACCESS

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 then 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.

International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 4, Jul – Aug 2017

- 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



Figure 1 shows Facebook extracter. 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 multinominal naive bayers 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].

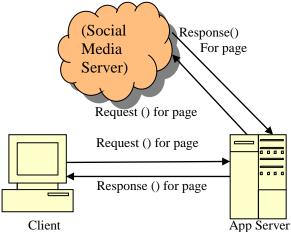


Fig. 2 Facebook analyzer

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/1456 34995501895/

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

International Journal of Computer Science Trends and Technology (IJCST) - Volume 5 Issue 4, Jul - Aug 2017

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)(1)

 $C_{map} = arg_{max} C \in C P(c/d)....(2)$

> All word positions in test document can be consider,

 $C_{NB} = \operatorname{argmax}_{cj \in c} P(cj) \pi_{i \in \text{positions}} P(xi/cj)$

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(4)

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

Where.

N-Total number of documents.

 $\{d \in D: t \in d\}$ -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.

International Journal of Computer Science Trends and Technology (IJCST) - Volume 5 Issue 4, Jul - Aug 2017

		* * *
l Support Narendra Modi Page id: 390976557590103	Narendra Modi fans from Karunadu	Narendra Modi - Prime Minister Of India
Show Posts	Page Id: 633444083378361 Show Posts	Page id: 551877531496760 Show Posts
I SUPPORT NARENDRA	Narendra Modi Fans : Karnataka	Die Hard Fans Of Narendra Modi
Page kl: 503940733076732	Page id: 427991917262845	Page id: 1474839259404026
Narendra Modi: The PM	Narendra Modi Live	Thailand Support
	Page 42: 399976557590183 Steam Posts I SUPPORT NARENDRA MODI Page 42: 309940733016732 Steam Posts.	Page d: 3987/855590183 Karunadu Shew Peds Page d: 8384488337661 Shew Peds Shew Peds I SUPPORT NARENDRA MODI Page d: 50394073376732 Narendra Modi Fans : Karnataka Page d: 427991917262845 Shew Peds Shew Peds

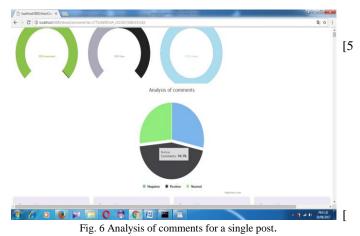
Fig. 4 Related pages for entered keyword.

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

C O localhost 5000/showPosts			\$
cebook Analyzer			
	All and a second se		
• •			
est id:	Post id:	Post id:	Post id:
7526890164_10159272683355165	177526890164,10159271949635165	177526890164_10159267891995165	177526890164_10159262493725165
7526890164_10159272683355165 eated time: 2017-08-			
7526890164_10159272683355165 eated time: 2017-08- 113-56:14+0000 essage:	177526890164,10159271949635165 Created time: 2017-08- 17710.04-08+0000 Message:	177526890164_10159267891995165 Created time: 2017-08- 16713-45-32+0000 Message:	177526890164_10159262493725165 Created time: 2017-08- 15T10:10:39+0000 Message
7526890164_10159272683355165 eated time: 2017-08- 113:56:14+0000 essage: को दोडमार जर क्षेत्र जल नम्ब म्या: इस सरकार	177526890164_10159271949635165 Created time: 2017-08- 17710.04.08+0000 Message: This months #ManeKiBaat will take	177526890164_10159267891995165 Created time: 2017-08 16713-45:32+0000 Message Met 6 bright officers of the Indian navy	177526890164, 10159262493725165 Created time: 2017-08- 15T10-10.3940000 Message Here are some glimpses from the Red
17526890164_10159272683355165 vated time: 2017-08- 1713.56.14-0000 essage: जोरी रोजनगर कर होत्र बल मन्य मा। इस सरकार	177526890164,10159271949635165 Created time: 2017-08- 17710.04-08+0000 Message:	177526890164_10159267891995165 Created time: 2017-08- 16713-45-32+0000 Message:	177526890164_10159262490725165 Created time: 2017-08- 15T10:10:39+0000 Message:
291 से 77526890164_10159272683355165 ब्रथ्मबर साम्य 2017 ट्रिक- 1713581449000 ब्रह्मब्रह्म सामे देवामर साम मार्ग्स द्वारात बेदोम्बर साम मार्ग्स 19100 2011/45115	177526890164,10159271949635165 Created time: 2017-86 17710 (4.4694000 Message: This months #Mannkillaat will take place on 27th August, Share your ideas for it, on the NM Mobile App. Call 1800- 11-7800 & record your message or share	177526890164,10159267891995165 Created stmc:2017-08 1611343240000 Message: Met 6 bright officers of the Indian navy who are going to circumnavigate the globe on the sailing vessel, NSV Tarini. This is the first-ever indian	177526890164, 10159262490725165 Created time: 2017-08- 15T10-10.39+0000 Message Here are some glimpses from the Red
17526890164_10159272683355165 vated time: 2017-08- 1713.56.14-0000 essage: जोरी रोजनगर कर होत्र बल मन्य मा। इस सरकार	1773/279490164, 101922/1949035165 Created time: 2017-080- 17710.04.08+0000 Message: This: months: atManetiBaat will take piloc eth 27th August. Share your ideas for it, on the NM Mobile Ago, 2011100- 11-7000 & record your message or share your ideas for ManerikBaat on the	177524090164,10159287691945165 Created time: 2017.06: 15113.45240500 Metasage: Met 6 bright efficers of the Indian navy who are going to circumanigate the globe on the saling varset, MNN Tarini. This is the first-ever Indian circumanigation of the globe by an all-	177526890164,10159262493725165 Greatest time: 2017-06 151710-103540000 Message: Here are some gimpses from the Red Fort this moming.
7528890164_10159272683355165 vated time: 2017-08- 113.56.14+0000 essage सारी डीजगार का दोप जन गण्ड गा। इस सरकार दुसाल बेरीजगार का गए हैं।	177526890164,10159271949635165 Created time: 2017-86 17710 (4.4694000 Message: This months #Mannkillaat will take place on 27th August, Share your ideas for it, on the NM Mobile App. Call 1800- 11-7800 & record your message or share	177526890164,10159267891995165 Created stmc:2017-08 1611343240000 Message: Met 6 bright officers of the Indian navy who are going to circumnavigate the globe on the sailing vessel, NSV Tarini. This is the first-ever indian	177526890164,10159262493725165 Greatest time: 2017-06 151710-103540000 Message: Here are some gimpses from the Red Fort this moming.

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.



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.

C Discalhout 5000/thowCo	mmentaltid=177526890164_80158910741215165 अलग नारारा मात्राज नाम आग गांव व साख्य व,आ जाता है। अर्थात कुल साखदी का तेक दल पति माह 38 बनीड 100 मांड है। और हर ते के इन सासदी की 465 करीड 60 माठा कपाय तेसन मराता में	ki kalakki tiki sharrini ne paytigi nami ayki Seadership ko natmastak pranam karte hal.Sir Magar anevale kal ke liye hame second front talyari kami hogi jo apke	This seems to be a Negative comment
	दिया जाता है। अर्थात भारत के विधायको और सांसदों के पीछे भारत का प्रति वर्ष 15 अरब 65 करोड़ 60 लाख रूपये खर्च होता है। ये तो सिर्फ	jaise determine, well confident honest aur intelligent hogi	Comment id:
	ूनके सुन्द केतन कहती की बात हुई। इनके अत्यार, सार्ट, सार्व, सार्व, सार्ट, कार्य, सार्ट, सार्ट, सीर सार्व्या, आर्ट, का का दाये की सारमान इनका ती है। अस्मित सार्वे आ का सार्वे का सार्ट का सार्ट हैं इस विधायकों, सेंदर साराटी पर, अब मार अधिक प्रकृत सुराधा में ती सार्ट, सार्ट्साकों में सी केतन पर, एक डिमायक की दी वीडियो के ती काल पर, एक डिमायक की दी वीडियो के ती काल पर, एक डिमायक की दी वीडियो की सार सार्टिया है। 3 रवीडा के सारन साराय सार सार्टिया है। 3 रवीडा के सारन साराय	Show Prediction	10158910741215164, 101589119041351 Creator ini 155700457375347 Author name: Jay Stah Creator time: 2017:36- 13171736:3140000 Message: Modi sab app ne jis hisab ak ketty mogan ko laterview me abrekommer ke bareme hamar responsibility ke bare om cikaar maximitary
	(25.000 स्मार्थ कोटन माल केटर की) त्या का 75 हाजर समये दोजा है। इस शिला के 4.562 शिपाकों की सुरक्षा कर बालाजन सार्थ के 9.476 कर बराई 22 साथ करित और है। इसी प्रकार साराटी कर सुरक्षा पर पति की 16.685 स्मार्थ कर साराटी है। 2. स्वेनी की सुरक्षा पाप्त नेतर, मांडिकी, मुदरानांडिय, प्रायानमंत्री के सुरक्ष के लिए लामाना 16.000 ज्यान ज्याजन के लिएन की त्यार हो।		ved ka reference diya irealiyek world guru ko shoha de dina tha Balke ek bar Mr. Puto hih ajke is answer se ek respect aur ajene ek world class liesdership dekin rake tha Sir ago cha gaye the. App to age badho Ame famari sathe chthye. Jai Jinendra
	जिनगर राजनामा कुस वायी लगभाग 776 करोड़ रुपया विदेशों हैं। इस प्रकार सन्तराधीन नेतराओं की सुरक्षा पर हर वर्ष लगभाग 20 अरव रुपये वाये मौते हैं। असीच वह नई स्वीरकोय पर क्या ये काम		Show Prediction This seems to be a Positive 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

https://en.wikipedia.org/wiki/Google_Analytics#Populari

- ty
 [3] https://en.wikipedia.org/wiki/Social_media_analytics
- [4] https://developers.facebook.com/*tools-and-support*/
- 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]

[2]

https://en.wikipedia.org/wiki/Natural_language_process ing.

- [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)

International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 4, Jul – Aug 2017

[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