

Instagram Analysis: Sentiment Analysis | Opinion Mining

Nirvishesh Shrivastava

RGPV, Dept. of Computer Science & Engineering
Institute of Engineering & Science IPS Academy
Indore, India

niirvishesh@gmail.com

Abstract –

Now-a-days people say that “Data is the new Oil.” Indeed it is. Today’s world is driven by this new oil. But wonder what it would be if we cannot make sense out of the data. In today’s era, enormous amount of data is being engendered by the Internet users among the whole world and likely to be flooded with more. The fix is to extract the essence from the data and make it useful for the people. Many tech giant companies are onto it by various mechanism like Hadoop, opinion mining, sentiment analysis, etc. Studies over sentiment analysis are being deliberately and positively conducted by enacting natural language processing technologies for subjective analysis such as attitude and opinion of persons exhibited on the web, social networking services (SNS), blogs etc. Instagram Analysis is the approach to extract the opinion of an Instagram user by the photos and videos uploaded by him/her. It uses sentiment analysis with opinion mining approach and presents the interest of the user.

Keywords -

Emoji, Hashtag, Instagram, Instagram username, Opinion, Sentiment.

Therefore in this paper, one of the 3rd generation SNS, Instagram we want to classify users' emotions as objects signs. An Instagram photo is an image that you want to share. It is possible to shoot through a when uploading a simple hashtag (hashtag) you can express your feelings by posting a typical third-generation SNS that can interact with other users to be. If a sentiment classification system with a large capacity is prepared using the presented method, then it is expected that sentiment analysis in various fields will be possible, such as for determining social phenomena through SNS.

In this paper, despite the anti-thesis categorization used in un-mining classify sentiment based on psychologically defined emotions and analyzed emotions used in actual Instagram. The hashtag of the Instagram category. Research on existing emotion analysis in order to organize the emotion of the text. And a positive value for the emotional vocabulary, the emotions were judged by judging the emotions, categories allow you to examine your main emotions that can solve ambiguity about subjective feelings.

2. Related Studies:

1. Introduction:

Recently, social network service (Social Network Service, SNS), along with the usage of smartphones, as a part of life, various social networks has emerged. There are three categories of SNS. The first-generation SNS Online communicating around offline connections like cyworld's mini-homepage or blog was in the form of. In the second generation SNS, together with Twitter or Facebook short message in real time format service is becoming popular. Various with the emergence of social networks, in the existing SNS that continues communication sharing is focused on distinct concepts such as interests or hobbies.

Tumblr; which is a customized service; Pinterest, Instagram (7-years-old). SNS has been developed in various forms [1,2]. The spread of this SNS is related to the interests of the single individual by enabling information sharing between entities, analyze your interests or feelings using timed posts research [3,4,5] has been conveyed, but most existing. It is in the study using the universalized SNS.

2.1 Sentiment Analysis

Emotional analysis refers to textual information such as sentences or documents. Analyzing the opinions, attitudes, feelings, etc. of the people expressed. In the past, polarity classifies the word as positive and negative extremes, but recently, using SNS data there are many researches on emotion scrutiny has been underway [5,6,7]. The seven emotions 'anger', 'confusion', 'depression', 'fatigue' analyzing the emotional characteristics of each genre, through empirical analysis, the nature of the data and that it can be applied to practical applications all. [6], Twitter and Cyworld C log, domestic portalNaver (Naver) movie review comment data, the seven categories of positive subcategories, the details of negative subcategories and classified into 15 categories.

In [7], psychological emotion based on the analysis of the text of Twitter after extracting the emotion information frame, define it basically suggesting a way to modify the sensibility of the emotion, In addition to restructuring the emotions that serve as a basis for positive and negative 9 positive, 10 negative, 4 neutral. The existing emotion classification

method is SNS of the text data, based on Thayer's emotional model, to classify users' emotions using hashtags do.

2.2 Hashtag (Hashtag)

A hashtag uses a specific word after the '#'Facebook, Instagram, and other keywords on the SNS. It is a function introduced so that it can be searched conveniently. User by posting a hashtag on a post, you can express your interests, your emotions. Data about the situation can also be shared. On these hashtags there are representative studies utilizing his features [8, 9, 10]. The study of [8] supervised the message of Twitter classification method, it can be used as an important indicator to express experimental results are shown.

The study of [9] based on the graph model for emotional analysis. We propose a hash tag emotion classification approach. At this time, tag type to express subjective opinion on the subject. Is an emotional hash tag, a subject hash tag, a table the emotional subject hash? The graph model appears as three categories of tags, we decided to determine the polarity of emotions through the co- and the hash tag is a vital element of emotional analysis. It is experimentally proved that it is.

The study of [10] the hashtag function is important for the rise of social networks as a role, and in Instagram buy a Naive Bayes classifier we attempted to classify the hashtags by emotion classification. Therefore in this paper, we utilize the features of these hash tags I would like to suggest an emotion category.

3. Experiment:

Our goal of this experiment is two-fold. First, extract all the words used in the captions of the images uploaded by the user including 'Hashtags' and 'Emoji' from his Instagram account. Second, to evaluate the productiveness of the features from section for opinion mining in Instagram captions and to get the frequency of each distinct word used by the user in all the images.

The algorithm used in the Instagram Analysis presents the number of images or videos over which the analysis has been done, number of words in the captions including hashtags and emoji, the words in the sequence, in the way, they were uploaded and the last the frequency of the words that how many times a word has been used by the user in his/her images and videos.

The first part of the experiment extracts all the information which the person has entered in his profile in the form of captions, hashtags, emoji, etc. That information is stored in a text file which is highly unorganized which can be seen in the later part of this experiment. The second part extracts the essence out of that unorganized data by implementing various steps defined in the Instagram Analysis algorithm. And gives us the useful information about the data.

The sequence of the analysis of the user is depicted in figure1.

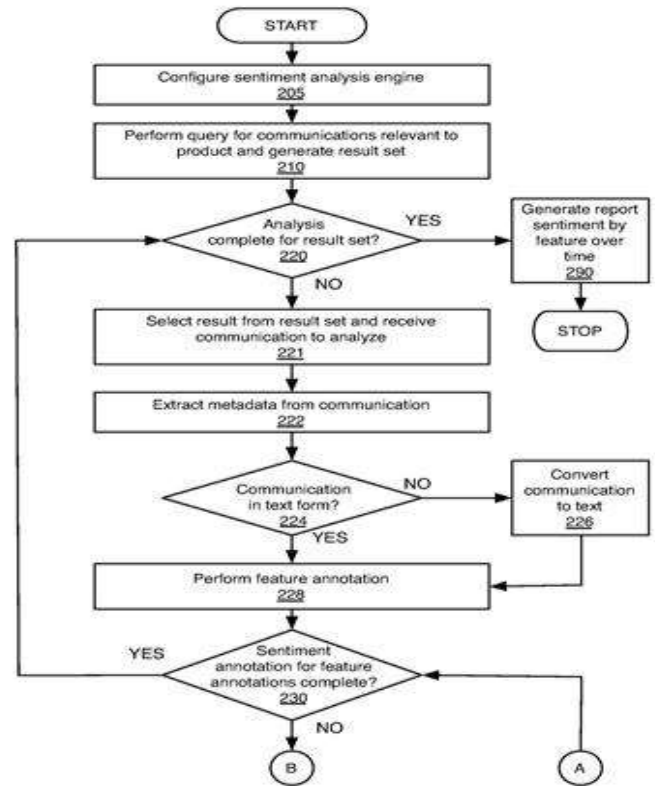


Fig. 1. The sequence of the analysis of the user.

```

#!/usr/bin/perl
use strict;
use warnings;
use LWP::Simple;
use Data::Dumper;

my $url = "https://www.instagram.com/";
my $username = "your_username";
my $password = "your_password";

my $url .= $username;

my $response = get($url);

if ($response =~ /password/) {
    print "Password is required\n";
} else {
    print "No password required\n";
}

my $username = "your_username";
my $password = "your_password";

my $url .= $username;

my $response = get($url);

if ($response =~ /password/) {
    print "Password is required\n";
} else {
    print "No password required\n";
}

my $username = "your_username";
my $password = "your_password";

my $url .= $username;

my $response = get($url);

if ($response =~ /password/) {
    print "Password is required\n";
} else {
    print "No password required\n";
}
  
```

Fig. 2. A part of python code of Instagram Analysis

Before the python script gets executed, it needs to get an input i.e. the Instagram username of the person to fetch all the hashtags, emoji, captions uploaded by him on the photos over Instagram platform.

Fig. 2. A glimpse of the form to be filled with username

Once a person has entered the username in this form. The html takes this username as input and gives it to the script. The script then runs the code with that specific username and displays all the photos uploaded by the user (condition being the profile of the user should be public as we cannot breach the security of Instagram as per its policies for private Instagram accounts).

The flow of the program will be like this:

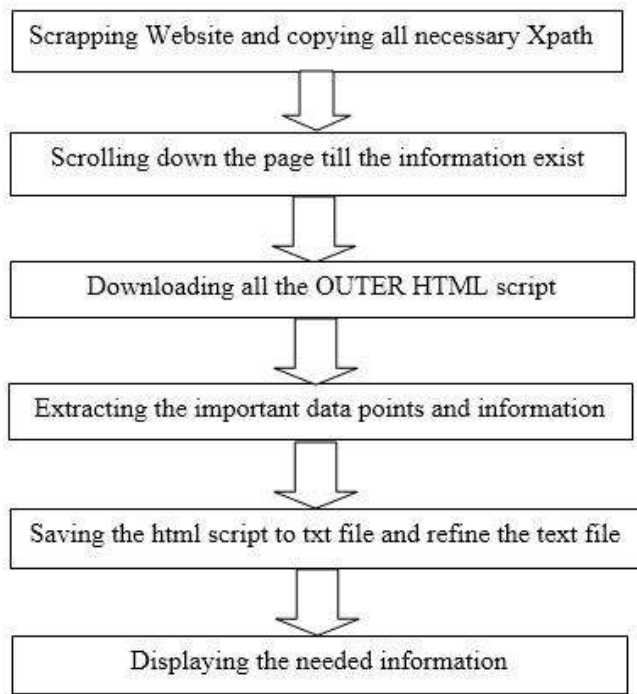


Fig. The flow of the program

The view of the Instagram user after the script open it through the WebDriver:



Fig. 3. Profile of the Instagram user (Ex: swaranjali2k18)

After traversing through all the photos of the user, i.e. when the script reaches to the end of the photos by maximizing the height of the scroller, it fetches all the captions used by the user and stores them in a file, that data that has been generated by the user is in raw format, it is the 'OUTER HTML'. Somewhat like this.

```

<main class=" _8u1zq _2v/9o " role=" main"><article class=" _mesns">
class=" _b0acm"><div class=" _82odm"><canvas class=" _15vpm" width=
style="transform: translate(-9px, -9px); position: absolute; top: 0px; left:
170px;"></canvas><span class=" _pg23k " style="width: 152px; height:
src="https://scontent-bom1-
1.cdninstagram.com/vp/94444a4074c42f0f0b7bf1a5ca9cd642/5B61DD25
19/s150x150/27878555_271782373361571_6308220593091117056_n.jpg
class=" _o6mpc"><div class=" _ienqf"><h1 class=" _rf3jb notranslate"
title="swaranjali2k18">swaranjali2k18</h1><a class=" _ncrqg " href="/
%2Fswaranjali2k18%2F&source=follow" rel="nofollow"><span cl
class=" _qv64e _gexxb_r9b8f _njrw0 ">Follow</button></span><
class=" _bnq48 "><span class=" _t98z6 "><span class=" _fd86t ">31</spa
class=" _bnq48 "><span class=" _t98z6 "><span class=" _fd86t " title="2:
followers</span></li><li class=" _bnq48 "><span class=" _t98z6 "><span
following</span></li></ul><div class=" _tb97a"><h1 class=" _kc4z2">SW
  
```

Fig. 4. Raw Data generated by the Python Code

After the data is being generated by the python code of the Instagram user, the crux is to extract the meaningful data of our use from this raw format whose glimpse is depicted in the figure 4. After refining all the unnecessary stuffs from the raw data file through various means of filter as in this case filter means splitting the code from keywords as 'class', 'id', 'src', 'img alt', etc after storing them in various arrays in python. And at last accessing the filtered data for our use.



swaranjali2k18 • Follow

swaranjali2k18 1st day of SWARANJALI 2018
 being displayed in city bhaskar indore □□

As, it can be seen that ‘swaranjali2k18’ has uploaded a picture and captioned it as: “1st Day of SWARANJALI 2018” being displayed in city bhaskar Indore emoji”. Now let us look how it looks in the raw data format.

```
src= https://content-bom1-  
Lcdninstagram.com/vp/1cc9d14d130f018d2d332aee90aa72d45B4F7736f451.2885-  
15/c35/e198.0.435.435/27892460_198256974092710_4821735847919353856_n.jpg" style=""><div><div  
:class="sj7dy"><div><div><a><div><div class="mck9w_gvnoze_tnlp"><a  
href="jp/BkHfjxjK1.2/taken-by=swaranjali218"><div class="e3l2"><div class="4rbun"><img  
alt="1st day of SWARANJALI 2018 being displayed in city bhaskar indore U'0001f60dU'0001f60e"  
:class="2diSp" decoding="auto" sizes="293px" srcset="https://content-bom1-  
Lcdninstagram.com/vp/ee81471706156044efc2de55f23408e5B5F57CA/451.2885-  
15/c150/c150/c35/e87.0.367.367/27894040_195035857756123_1712227289946652672_n.jpg
```

This is how caption look in the Raw Data format ‘highlighted with yellow color’. This is quite not readable all the time for all persons. The main essence is to extract useful data from it to the readable format for all.

In the figure 7, the highlighted part of the image with yellow color is the caption that was displayed in the images 5 and 6 and the rest of the words are the words in caption used by that specific user (swaranjali2k18 in this case).

Text	Frequency
Photo analyzed: 29	
Word Count: 64	
Schedule	
of	
day	schedule
3...	swaranjali2k18"
swaranjali2k18"	Schedule
2nd	for
day	day
coverage	1
U0001f60e	@swaranjali2k18"
U0001f60e	swaranjali2k18"
U0001f60e"	swaranjali2k18"
1st	swaranjali2k18
day	sarfar0sh"
of	swaranjali2k18
SWARANJALI	newevent"
2018	swaranjali2k18"
being	swaranjali2k18"
displayed	swaranjali2k18"
in	swaranjali2k18"
city	swaranjali2k18"
bhaskar	swaranjali2k18"
indore	swaranjali2k18"
U0001f60d	swaranjali2k18"
U0001f60e"	swaranjali2k18"
Are	swaranjali2k18"
you	swaranjali2k18"
up	swaranjali2k18"
for	swaranjali2k18"
one.....	swaranjali2k18"
U0001f60b	swaranjali2k18"
swaranjali2k18"	swaranjali2k18"
Day	swaranjali2k18"
2	swaranjali2k18"

Figure 7. Refined caption along with the total words used by the user in the pictures uploaded.

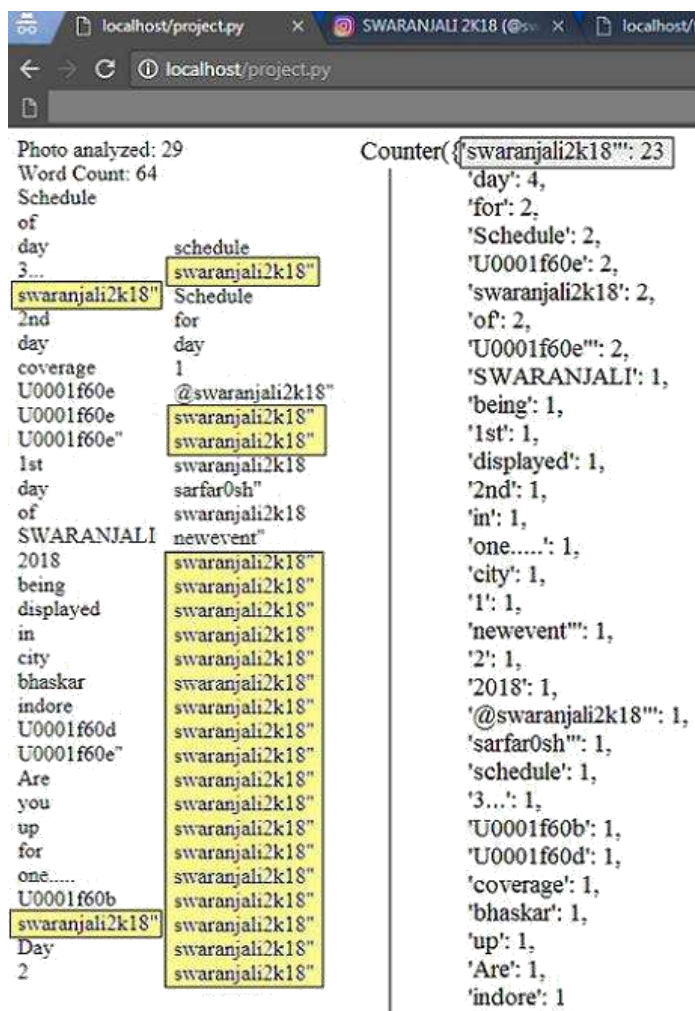


Figure 8. Overall output along with the frequency count of the user.

It can be seen from the figure 8, that the word in the right upper corner highlighted with grey color 'swaranjali2k18' has been used most of the times by the user i.e. "23". The highlighted portions with yellow color are the frequency of that word 'swaranjali2k18'. We can count it as well that it appears 23 times. Similarly, 'day' has been used 4 times by the user and so on.

By these means we can explore the user in terms of his thought process that what are his various interests and opinions about product, politics, favors, etc. Therefore, these kind of approaches are being used by organizations to meet the demand of the person even before the persons knows by himself.

4. Test and Results:

By implementation of this Instagram Analysis algorithm, it has been found by various experiments on various users that the efficiency of the algorithm is 94%.

Some of the experiments are:

USERNAME	INTEREST	ACCURACY
journey2taste	Food	100%
swaranjali2k18	Swaranjali	100%
shivya	Travel	95%
gramiksha_	School	85%
Indorehd	Travel	90%

Table1. Interest of the user

The table1 above depicts the interest of the users predicted by the algorithm by the words used in the captions in the form of hashtags, words, emoji, etc.

5. Applications and Tools:

Some of the applications of sentiment analysis involves online advertising, polling, etc. Online advertising has grown up as one of the major gaining sources of today's Internet ecosystem.

When faced with tremendous amounts of online information from various online forums, information seekers usually find it very difficult to yield accurate information which is useful to them, where useful information is quickly exposed to those seekers. In order to identify potential risks, it is important for companies and industries to collect and analyze information about their competitors' products and plans.

Sentiment analysis finds a major role in competitive intelligence to extract and visualize comparative relations between from customer reviews to help the enterprise in discovering the probable danger in the products and further draft alternate products and marketing approach. Summarization of the opinion encapsulates opinion of the objects by telling corresponding events and sentiment polarities. With the help of it, a production house in no time can see the how their consumers feel about their product and its ups and downs in the market.

Instagram Analysis is based on Python script which uses various libraries such as; selenium, emoji, collections, time, urllib3. It also requires a web driver which helps in taking that username of the person to the web browser along with the whole link of the Instagram user. The python version used in the whole algorithm is Python2.7 which was found much compatible while dealing with emoji.

6. Conclusion:

As we are seeing that, there has been tremendous growth in the social photo sharing platforms such as Instagram, Facebook, etc. Spurred by this growth media groups and industries are finding various paths to mine the data produced by these platforms to get a favorable result for their organization about how the consumers are feeling about their product and the reviews of their product by detecting the sentiments of the users.

A wide variety of applications are based on Sentiment detection in information systems including real time applications in classifying reviews.

Today's world which produces an enormous amount of data regularly with a lots of variety and veracity, it is obligatory to make a sense out of that data. Instagram is an emerging platform these days which produces a lot of data which when utilized in a proper manner can lead to rightful and meaningful outcome for the present generation and in the longer run as well.

In future, more work is desired on increasing the performance measures. Sentiment analysis can be applied for new applications. Although the techniques and algorithms used for opinion mining or sentiment analysis are advancing fast, but, a lot of problems in this field of study remains complicated and unsolved such as dealing with negation expressions, complexity of sentence/document, etc. More future research could be dedicated to these challenges.

7. References

- [1] GlobalWebIndex, GWI Social Summary, 2014.
- [2] Park, B. S. "Global SNS Present Condition and Implications." *Korea Information Society, Development Institute, Information & Communications policy* 26.2 (2014): 22-34.
- [3] Nam, Minji, Jeongin Kim, and Juhyun Shin. "A user emotion information measurement using image and text on Instagram-based." *Journal of Korea Multimedia Society* 17.9 (2014): 1125-1133.
- [4] Kim, Jeongin, et al. "Retracted: Extracting User Interests on Facebook." *International Journal of Distributed Sensor Networks* 10.6 (2014): 146967.
- [5] Lee, Cheolseong, et al. "Classification and analysis of emotion in Korean microblog texts." *Journal of Korean Institute of Information Scientists and Engineers: Databases* 40.3 (2013): 159-167.
- [6] Kim, Do-Yeon, Yong Wu, and Hyuk-Ro Park. "Constructing an evaluation set for Korean sentiment analysis systems incorporating the category and the strength of sentiment." *The Journal of the Korea Contents Association* 12.11 (2012): 30-38.
- [7] Chang, Moon-Soo. "Empirical Sentiment Classification Using Psychological Emotions and Social Web Data." *Journal of Korean Institute of Intelligent Systems* 22.5 (2012): 563-569.
- [8] Rosa, Kevin Dela, et al. "Topical clustering of tweets." *Proceedings of the ACM SIGIR: SWSM* (2011).
- [9] Wang, Xiaolong, et al. "Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach." *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011.
- [10] Gunawardena, Nipun, et al. "Instagram hashtag sentiment analysis." *University of Utah CS530/CS630 Conference of Machine Learning*. 2013.