International conference on "Recent Advances in Interdisciplinary Trends in Engineering & Applications

Instagram Sentiment Analysis :Opinion Mining

Shweta Gangrade[a], Nirvishesh Shrivastava[b], Jayesh Gangrade[a,b,∗]

[a]CSE Department, IES IPS Academy, Indore, India
[b]CSE Department, IES IPS Academy, Indore, India

Abstract

In recent times, studies on sentiment analysis are being actively conducted by implementing natural language processing technologies for analyzing subjective data such as opinions and attitudes of users expressed on the Web, blogs, and social networking services (SNSs). Conventionally, to classify the sentiments in texts, most studies determine positive/negative/neutral sentiments by assigning polarity values for sentiment vocabulary using sentiment lexicons. However, in this study, sentiments are classified based on Thayer's model, which is psychologically defined, unlike the polarity classification used in opinion mining. In this paper, as a method for classifying the sentiments, sentiment categories are proposed by extracting sentiment keywords for major sentiments by using hashtags, which are essential elements of Instagram. By applying sentiment categories to user posts, sentiments can be determined through the similarity measurement between the sentiment adjective candidates and the sentiment keywords. The test results of the proposed method show that the average accuracy rate for all the sentiment categories was 90.7%, which indicates good performance. If a sentiment classification system with a large capacity is prepared using the proposed method, then it is expected that sentiment analysis in various fields will be possible, such as for determining social phenomena through SNS.

Keywords: Instagram, sentiment, opinion, hashtags, Social networking services (SNSs).

1. Introduction:

Recently, social network service (Social Network Service, SNS), along with the use of smartphones, as a part of life, various social networks has emerged. There are three types 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, 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 specific topics such as car interests or hobbies. Tumblr, which is a customized service, Pinterest, Instagram (7-years-old).SNS has been developed in various forms Global et al.(2014), Park B et al.(2014).The spread of this SNS is related to the interests of the individual by enabling information sharing between entities, analyze your interests or feelings using timed posts research Nam M et al.(2014), Kim J et al.(2014), Lee C et al.(2013) has been carried out, but most existing. It is in the study using the universalized SNS. Therefore in this study, one of the 3rd generation SNS, Instagram we want to classify users' emotions as objects signs. An Instagram 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. In this paper, unlike the polarity classification used in un-mining classify emotions 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 classify 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 analyze your main emotions.

All the composition of this paper is as follows. In Chapter 2, SNS Existing studies using emotion analysis using data and a hash tag will be described. In Chapter 3, the method of classification and analysis of user's emotions. In this paper, as a method of analysis, and the suggested hashtag-based emotion category. How to apply emotions to your posts? In chapter 4, to evaluate the performance, and finally conclusion and future studies are presented in Chapter 5.

2. **Related Studies:**

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 analysis has been underway Kim D et al.(2012), Chang M et al.(2012), Rosa K et al.(2011). 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 A ll. Kim D et al.(2012), 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 Chang M et al.(2012), 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. Information about the situation can also be shared. On these hashtags there are representative studies utilizing his features Rosa K et al.(2011), Wang X et al.(2011), Gunawardena N et al.(2013).The study of Rosa K et al.(2011) supervised the message of Twitter classification method, it can be used as an important indicator to express experimental results are shown. The study of Wang X et al.(2011) 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 and the hash tag is an important element of emotional analysis. It is experimentally proved that it is. The study of Gunawardena N et al.(2013) 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 it will suggest an emotion category.

3. Experiment and Results:

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 effectiveness 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 displays the number of pictures over which the analysis has been done, number of words in the captions along with the 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 profile.

```
#!C:\Python27\python.exe
print ("Content-Type: text/html\n")

from selenium import webdriver
from collections import Counter
from selenium.webdriver.common.keys import Keys
import emoji
import urllib3.request as ul
import time
import sys

username = open("post.txt","r").read()
driver=webdriver.Chrome("C:/chromedriver.exe")
driver.get("https://www.instagram.com/"+username+"/")
p='//*[@id="react-root"]/section/main'

while True:
    last_height = driver.execute_script("return document.body.scrollHeight"
    driver.execute_script("window.scrollTo(0, document.body.scrollHeight);"
    time.sleep(2)
    new_height = driver.execute_script("return document.body.scrollHeight")
    if new_height==last_height:
        break


driver.implicitly_wait(100)

ee=driver.find_elements_by_xpath(p)
zz=[ee.get_attribute('outerHTML') for ee in ee]
```

Fig. 1. A part of python code of Instagram Analysis

When we go through it and run it. We have to give input to the script as the Instagram username of the user to fetch all the hashtags, emoji, etc uploaded by him on the photos over Instagram platform.
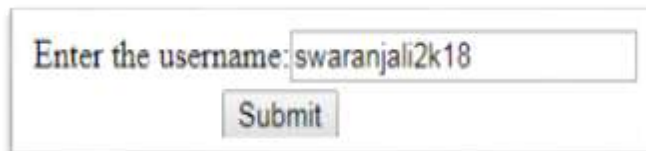
Fig. 2. A glimpse of the form to be filled with username

Once the user 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).

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.

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 filer 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.

For Example: User: swaranjali2k18 has uploaded a photo and used some captions along with hashtags and emoji in it. Let us look how it looks over Instagram to the users, how it looks in raw format and how it looks after the being filtered.

Fig. 5. Picture uploaded by user over Instagram

As, shown in figure 5 'swaranjali2k18' has uploaded a picture and captioned it as: "1st Day of SWARANJALI 2018 being displayed in city bhaskar Indore emoji emoji." The raw data format as shown in figure 6.

Fig. 6. Caption in Raw Data Format

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.

Somewhat like this:

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"

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

Words are the words in caption used by that specific user (swaranjali2k18 in this case). Now in the last image let us see the overall output that how many times all the distinct words are being enlisted by the user.
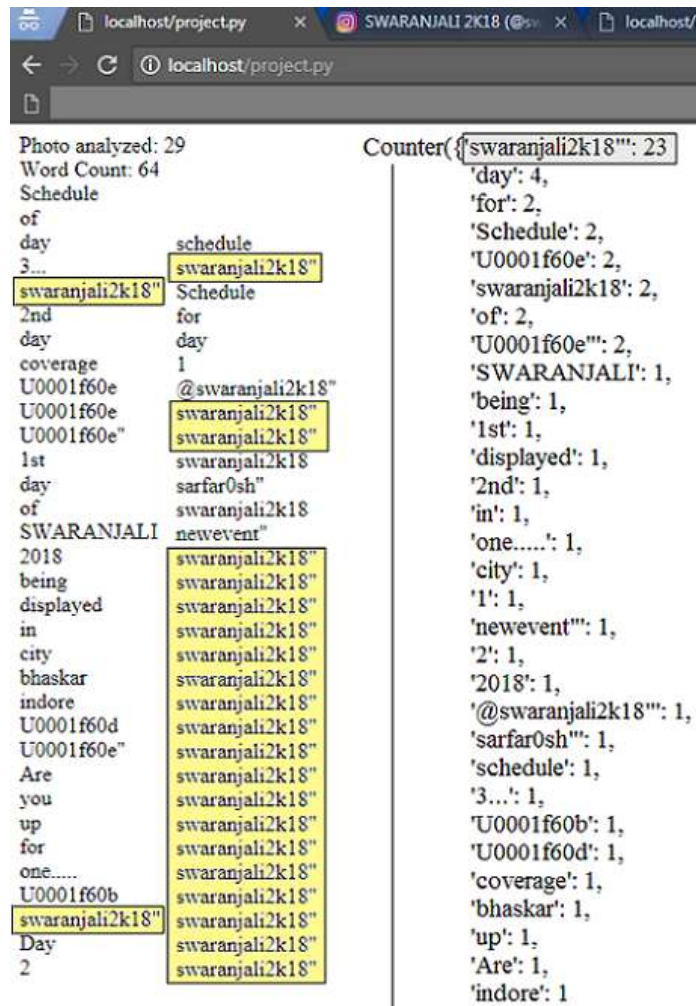
Fig. 8. Overall output along with the frequency count of the user.

As we can see 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 kind of approaches are being used by organizations to meet the 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 demand of the person even before the persons knows by himself.

Applications and Tools:

Some of the applications of sentiment analysis includes online advertising, polling, etc. Online advertising has become one of the major revenue sources of today's Internet ecosystem. Sentiment analysis find its recent application in dissatisfaction oriented online advertising Guang Qiu(2010) and Blogger-Centric Contextual Advertising (Teng-Kai Fan, Chia-Hui Chang ,2011), which refers to the assignment of personal ads to any blog page, chosen in according to bloggers' interests.

When faced with tremendous amounts of online information from various online forums, information seekers usually find it very difficult to yield accurate information that is useful to them, where useful information is quickly exposed to those seekers. Nan Li (2010) used sentiment analysis approach to provide a comprehensive and timely description of the interacting structural natural groupings of various forums, which will dynamically enable efficient detection of hotspot forums. In order to identify potential risks, it is important for companies to collect and analyze information about their competitors' products and plans. Sentiment analysis find a major role in competitive intelligence to extract and visualize comparative relations between products from customer reviews, with the interdependencies among relations taken into consideration, to help enterprises discover potential risks and further design new products and marketing strategies.

Opinion summarization summarizes opinions of articles by telling sentiment polarities, degree and the correlated events. With opinion summarization, a customer can easily see how the existing customers feel about a product, and the product manufacturer can get the reason why different stands people like it or what they complain about. Ku, Liang, and Chen (2006) investigated both news and web blog articles. Algorithms for opinion extraction at word, sentence and document level are proposed. The issue of relevant sentence selection is discussed, and then topical and opinionated information are summarized. Opinion summarizations are visualized by representative sentences. Other applications includes online message sentiment filtering-mail sentiment classification, web blog author's attitude analysis etc.

Conclusion:

In the past few years, there has been a huge growth in the use of social-photo sharing platforms such as Instagram. Spurred by that growth, companies and media organizations are increasingly seeking ways to mine Instagram for information about what people think and feel about their products and services i.e. detection of their sentiments.

Sentiment detection has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications. There are likely to be many other applications that has not been discussed yet. It is found that sentiment classifiers are severely dependent on domains or topics. From the above work it is evident that neither classification model consistently outperforms the other, different types of features have distinct distributions. It is also found that different types of features and classification algorithms are combined in an efficient way in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification performance.

In future, more work is needed on further improving the performance measures. Sentiment analysis can be applied for new applications. Although the techniques and algorithms used for sentiment analysis are advancing fast, however, a lot of problems in this field of study remain unsolved. The main challenging aspects exist in use of other languages, dealing with negation expressions; produce a summary of opinions based on product features/attributes, complexity of sentence/ document, handling of implicit product features, etc. More future research could be dedicated to these challenges.

References:

GlobalWebIndex, GWI Social Summary, 2014.

B.S. Park, "Global SNS Present Condition and Implications," Korea Information Society, Development Institute, Information & Communications policy, Vol. 26, No. 2, pp. 22-34, 2014.

M.J. Nam, J.I. Kim, and J.H. Shin, "A User Emotion Information Measurement using Image and Text on Instagram-Based," Journal of Korea Multimedia Society, Vol. 17,No. 9, pp. 1125-1133, 2014.

J.I. Kim, D.J. Choi, B.K. Ko, E.J. Lee, and P.K. Kim, "Extracting User Interests on Facebook," International Journal of Distributed Sensor Networks, Vol. 2014, Article ID 146967, pp. 1-5, 2014.

C.H. Lee, D.H. Choi, S.S. Kim, and S.W. Kang, "Classification and Analysis of Emotion in Korean Microblog Texts," Journal of Korean Institute of information Scientists and Engineers: Databases, Vol. 40, No. 3, pp. 159-167,2013.

D.Y. Kim, Y. Wu, and H.R. Park, "Constructingan Evaluation Set for Korean Sentiment Analysis Systems Incorporating the Category and the Strength of Sentiment," The Journal of the Korea Contents Association, Vol. 12,No. 11, pp. 30-38, 2012.

M.S. Chang, "Empirical Sentiment Classification using Psychological Emotions and Social Web Data," Journal of Korean Institute of Intelligent Systems, Vol. 22, No. 5, pp. 563-569, 2012.

K.D. Rosa, R. Shah, B. Lin, A. Gershman, and R. Frederking, "Topical Clustering of Tweets,"Proceeding of the ACM SIGIR : SWSM, pp. 1-8, 2011.

X. Wang, F. Wei, X. Liu, M. Zhou, and M.Zhang, "Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment ClassificationApproach," Proceedings of the 20thACM International Conference on Informationand Knowledge Management, pp. 1031-1040,2011.

N. Gunawardena, J. Plumb, N. Xiao, and H.Zhang, Instagram Hashtag Sentiment Analysis,University of Utah, Utah, 2013.