IJCSN - International Journal of Computer Science and Network, Volume 3, Issue 6, December 2014
ISSN (Online): 2277-5420 **www.IJCSN.org**
Impact Factor: 0.274

Opinion Mining from Social Networks

[1] Khyati Dave, [2] Surbhi Chandurkar , [3] Ashika Sinha

[1, 2, 3] Department of Computer Engineering, Savitribai Phule Pune University,
G. H. Raisoni Institue of Engineering & Technology
Pune, Maharashtra, India

Abstract - Online social networks have become a popular communication tool for the masses. People are constantly sharing their opinions on social media. Therefore, social networking websites are a rich source for opinion mining. By analyzing the various opinions expressed on such sites, we can determine how well a product is doing in the current market. In this paper, we propose a system which uses the popular micro-blogging website Twitter for mining user opinions about products or services. Our proposed system presents an approach to extract data from Twitter and perform linguistic analysis on it. Combining the techniques of artificial intelligence and natural language processing, we attempt to classify the opinion as positive, negative or neutral. Using word dependencies and Part Of Speech tagging, we analyze data to track the opinions expressed on social media on the given subject. This paper proposes a system that can give the user information in the form of graphs and charts, about the opinions of the other users on the product or service.

Keywords - *Opinion, parser, quintuple, sentiment, Twitter*

1. Introduction

In the last 10 years or so, online social networks have seen a tremendous rise in popularity. There are millions of people participating in online social networking sites such as Twitter and Facebook. There are an estimated 900 million active users in Twitter at the moment. These social networking sites also play a role in providing space for social discussions, business promotions and even government programs. These users, be it individuals or workgroups, share opinions, knowledge as well as interest with the other individuals and workgroups on the site. With the increasing number of participants in social networking sites, the analysis and evaluation of the information is also increasing in complexity. In the real world, businesses and organizations always want to find consumer or public opinions about their products and services. Individual consumers also want to know the opinions of existing users of a product before purchasing it. In the past, when an individual needed opinions, he/she asked friends and family. When an organization or a business needed public or consumer opinions, it conducted surveys, opinion polls, and focus groups. In this new information age, where thoughts and opinions are shared so prolifically through online social networks, tools that can make sense of the content of these networks are paramount. In order to make best use of this information, we need to be able to distinguish what is important and interesting. There are obvious benefits to companies, governments and so on in understanding what the public think about their products and services, but it is also in the interests of large public knowledge institutions to be able to collect, retrieve and preserve all the information related to certain events and their development over time.

With the explosive growth of social media (e.g., reviews, forum discussions, blogs, micro-blogs, Twitter, comments, and postings in social network sites) on the Web, individuals and organizations are increasingly using the content in these media for decision making. Nowadays, if one wants to buy a consumer product, one is no longer limited to asking one's friends and family for opinions because there are many user reviews and discussions in public forums on the Web about the product. For an organization, it may no longer be necessary to conduct surveys, opinion polls, and focus groups in order to gather public opinions because there is an abundance of such information publicly available. However, finding and monitoring opinion sites on the Web and distilling the information contained in them remains a difficult task. Automated sentiment analysis systems are thus needed.

Sentiment analysis, also called *opinion mining*, is the field of study that analyzes people's opinions, sentiments,

evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [1]. Social web analysis is all about the users who are actively engaged and generate content. This content is dynamic, rapidly changing to reflect the societal and sentimental fluctuations of the authors as well as the ever-changing use of language. Sentiment classification is a special case of text categorization problem, where the classification is done on the basis of attitude expressed by the authors [5]. Sentiment analysis requires a deep understanding of the document under analysis because the concern here is how the sentiment is being communicated.

2. Literature Survey

Recently, techniques for opinion mining have begun to focus on social media, combined with a trend towards its application as a proactive rather than a reactive mechanism [2]. Understanding public opinion can have important consequences for the prediction of future events. Almost all the work on opinion mining from Twitter has used machine learning techniques. Previous attempts aimed to classify arbitrary tweets on the basis of positive, negative and neutral sentiment, constructing a simple binary classifier which used n-gram and POS features and trained on instances which had been annotated according to the existence of positive and negative emoticons [4]. The existing systems, HootSuite and Bottlenose, are social media management systems that analyze social media and business data to detect trends for brands. These websites allow their users to identify social buying signals, send global or geo targeted messages across a number of social networks, visualize brand performance, and track trend intelligence, brand health and competitor performance. But these systems have the following disadvantages:

I. Expensive
II. Tightly brand specific, not for personal use.

Since these systems are expensive, have more brand-centric approach and limited classification, we propose a new system that is free, more trend-centric to data division and has a better brand analysis providing a method for an automatic collection of opinions about products from social networking sites which can be used to perform sentiment classification among positive, negative and neutral sets. This system can be used for overall review of a product or service, for political and social opinions, for better marketing strategies and better audience targeting.

Our work differs from earlier work in four main aspects:

I. Our focus is not on classifying each review as a whole but on classifying each sentence in a review.
II. We give more consideration/importance to the language properties of the sentence and in understanding the sentence constructs, for each sentence we recognize the subjects of the feeling and the feature being described.
III. We concentrate on the effects of conjunctions and sentence constructions which have not been researched for sentiment analysis.
IV. Our method does not need a training set since it depends on linguistic analysis.

3. Architecture

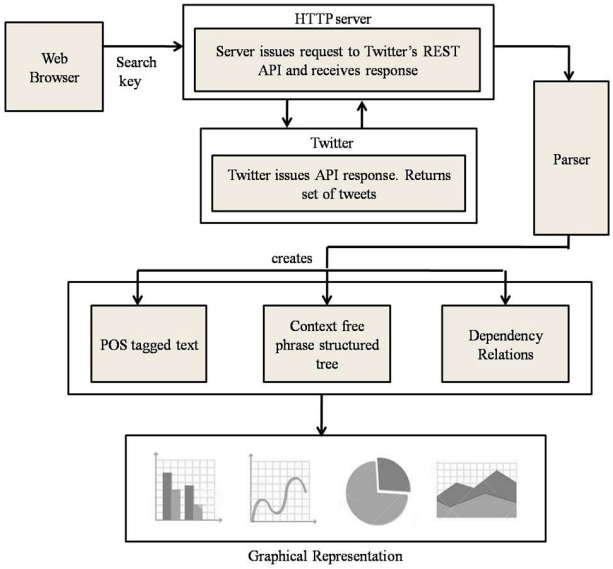

Figure 1: Proposed Architecture

3.1 Browser

The user accesses the project as a website from any standard web browser. Since the proposed project will be implemented in Java, the only prerequisite is that the browser must have Java Runtime Environment enabled for it to work. The user inputs the search query from via an input field. That search key will be forwarded to the server.

3.2 HTTP Server

The search key is sent to an HTTP server. The server sends an HTTP request to Twitter's Application

Programming Interface. The HTTP request carries the search key as a parameter. The parameter is then used to search Twitter's timeline for the given search key. The server also receives the HTTP response to the request from the Twitter's API. The response received will be in the form of a tweet set extracted by Twitter's API from the data stream.

3.3 Twitter API

In our system, we propose to use Twitter's REST API. It provides programs the access to read and modify data from the Twitter stream. Our project requires only reading the data and then doing further evaluations on it. We do not attempt to post of modify the Twitter stream. The REST API consists of a set of APIs that provide basic functions to access Twitter data. The Twitter Search API is a part of Twitter's v1.1 REST API. It allows us to query the indices of tweets (recent or popular). The API provides a multitude of query operators which allow the program to search the stream keeping certain constraints on the query. It also allows the program to send queries using additional parameters such as result type, geo location, language etc.

3.4 Parser

In our project, we propose to use a natural language parser. A parser decomposes a digital sentence into its basic grammatical structure. It analyses the phrase structure of the sentence and identifies the subject and the object of the sentence. In this project we propose to use the Stanford NLP Parser. The parser takes a sentence as input and outputs dependency relations known as Stanford dependencies and phrase structure trees [3]. These provide a source to modify the tweet into parts which define relationships between the words and in turn define whether a tweet is subjective enough to derive an opinion from.

3.5 Graphical Representation

Once the dependencies are created and text is tagged with each word's respective part of speech, the sentiment calculation can begin. We attempt to classify each tweet into a quintuple to represent an opinion in its most basic form. The quintuple is of the following form:

$$OQ = \{E_i, A_{ij}, Si_{jkl}, H_k, T_l\} \quad (1)$$

where,
E_i : Name of entity
A_{ij} : An aspect of Ei
S_{ijkl} : Sentiment value on aspect Aij
H_k : Opinion holder
T_l : Time when opinion was expressed
OQ : Opinion Quintuple

The sentiment value is calculated from the collected corpus which it is represented using graphs and charts.

4. Algorithms

4.1 POS Tagging Algorithm

In this algorithm, the user inputs the query which is then passed to the Twitter's Application Programming Interface. The Application Programming Interface will respond with a set of tweets matching the query. These set of tweets are then passed to the parser which parses each tweet and assigns POS tags to each sentence. The algorithm is as follows:

i. Scan the sentence word by word.
ii. For each word, analyze relative position in sentence and determine POS
iii. Assign part of speech from standard list iii. Create phrase tree based on POS tags
iv. Create dependency relationships based on word index and tag

4.2 Evaluation Algorithm

This algorithm generates the opinion quintuples from the tweets. Based on the generated sentiment values, graphs regenerated to depict the overall sentiment.

The algorithm is as follows:

i. For each tweet T_i:
   - Extract $entity_i$, $aspect_i$, $sentiment\_value_i$, $opinion\_holder_i$, $time_i$
   - Add_to_quintuple($entity_i$, $aspect_i$, $sentiment\_value_i$, $opinion\_holder_i$, $time_i$)
   - Calculate average sentiment value for quintuple.
ii. Compare sentiment values and plot results

Example:

a) Query:
"I like Microsoft Lumia's user interface."

b) POS tagged sentence:

I/PRP like/VBP Microsoft/NNP Lumia/NNP 's/POS user/NN interface/NN ./.

Where,

PRP: Personal Pronoun
VBP: Singular Verb Phrase Present Tense
NNP: Proper Noun
POS: Possessive
NN: Singular Noun

c) Context-free Phrase Tree:

```
(ROOT
  (S
    (NP (PRP I ))
    (VP (VBP like )
      (NP
        (NP (NNP Microsoft ) ( NNP Lumia ) (POS 's ))
        (NN user) (NN interface)))
(. .)))
```

d) Dependency Relation:

nsubj (like-2, I-1)
Root (ROOT-0, like-2)
nn (Lumia-4, Microsoft-3)
poss (interface-7, Lumia-4)
possessive (Lumia-4, 's-5)
nn (interface-7, user-6)
dobj (like-2, interface-7)

Where,

nsubj: Nominal Subject
nn: Noun Compound Modifier
poss: Possession Modifier
possessive: Possessive Modifier
dobj: Direct Object [3,6]

5. Mathematical Model

Let S be the system such that

S = {U, AP, TW, P, OQ, GR}
(2)

U : User which enters the query
AP : Twitter API which extracts data from the stream based on the query
TW : The tweets in the Twitter stream
P : An English language parse that assigns parts of speech to each tweet and produces POS tagged text, context free phrase structured tree and dependency relations
OQ : The opinion quintuple generated from the tweets.

Activities:

Activity 1:

The user's query is passed to the API. Let Un be the nth user entering the query:

f(Un) -> AP
(3)

Activity 2:

The API searches the stream and returns all the tweets matching the query. Let $\{TW_1,TW_2,TW_3 \ldots TW_{10}\}$ be the set if tweets extracted:

AP -> {{TW1,TW2,TW3...TW10} Є TW
(4)

Activity 3:

The parser will parse each tweet and assign POS tags to each sentence. Let f(POS) be the POS tagging function.

P -> f(POS) Є TWi
(5)

Activity 4:

The parser will generate dependency relations. Let f(DR) be the dependency relation generating function.

P -> f(DR) Є TWi
(6)

Activity 5:

The system will generate opinion quintuples from the tweets.

S -> OQ {Ei, Aij, Sijkl, Hk, Tl} Є TW
(7)

Activity 6:

Based on the sentiment values, graphs are generated to depict overall sentiment. Let f(genGraph) be the graph generating function.

S -> f(genGraph) } Є OQ (8)

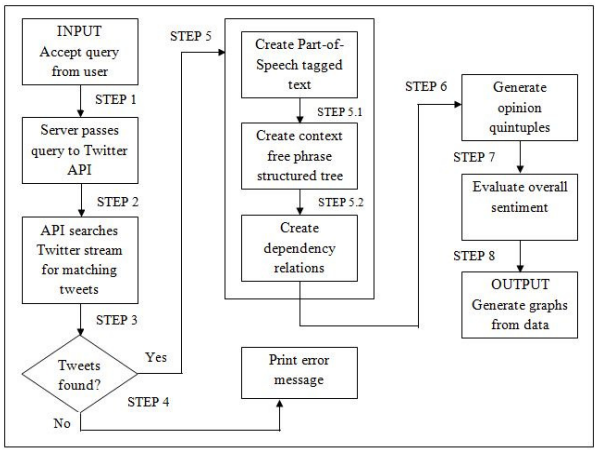

Figure 2: Steps in the process of sentiment analysis in our project.

6. Conclusions

Social networks have become a major role of communication. The large amount of data and opinions on micro-blogging websites makes them a rich source for opinion mining and sentiment analysis. This proposed project, though still in progress, provides promising initial results. We aim to mine useful opinions from social networks and analyze them to determine how a product is performing in the market. By using a parser we attempt to perform various Natural Language Processing steps to derive meaning from the statements extracted. Using the tools described in this paper, we can successfully create a system which mines opinions from the Twitter stream and classifies them as positive, negative and neutral sentiments. As future work, we plan to go beyond Twitter, expand to other social networking websites, blog posts and comments on E-commerce websites.

Acknowledgments

We are thankful to our college G. H. Raisoni Institute of Engineering and Technology, Pune for giving us an opportunity to prepare a paper on "OPINION MINING FROM SOCIAL NETWORKS". We would like to express our deep sense of gratitude and special thanks towards our esteemed guide and Head of Computer Department, Mrs. Deeksha Bhardwaj, for giving us this splendid opportunity to select and present this project. We deeply thank our principal, Dr. R. D. Kharadkar and all the staff members of the department for their immense support towards the realization of the project. We are also thankful to Dr. Parag Kulkarni, Director of Anomaly Solutions Pvt. Ltd. for sponsoring our project.

References

[1] Bing Liu, "Sentiment Analysis and Opinion Mining" Morgan & Claypool Publishers, May 2012.

[2] Luciano Barbosa and Junlan Feng, "Robust Sentiment Detection on Twitter from Biased and Noisy Data" AT&T Labs – Research, 2010

[3] Marie-Catherine de Marnee and Christopher D. Manning, "Stanford typed dependencies manual", September 2008, Revised for the Stanford Parser v. 3.3 in December 2013.

[4] Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining" Universit´e de Paris-Sud, Laboratoire LIMSI-CNRS, France, 2010

[5] Amitava Das and Sivaji Bandyopadhyay, "Subjectivity Detection using Genetic Algorithm 2011

[6] Marie-Cathrine de Marneffe, Bill MacCaitney and Christopher D.Manning." Generating typed dependency parses from phrase structure parses". In LREC 2006.

Khyati Dave Currently pursuing Bachelor of Engineering from G. H. Raisoni Institute of Engineering and Technology, Pune affiliated to Savitribai Phule Pune University. Opinion MIning from Social Networks is our final year project guided by Mrs. Deeksha Bhardwaj, Head of Computer Department and sponsored by Dr. Parag Kulkarni, Director of Anomaly Solutions Pvt. Ltd. in the academic year 2014-2015.

Surbhi Chandurkar Currently pursuing Bachelor of Engineering from G. H. Raisoni Institute of Engineering and Technology, Pune affiliated to Savitribai Phule Pune University. Opinion MIning from Social Networks is our final year project guided by Mrs. Deeksha Bhardwaj, Head of Computer Department and sponsored by Dr. Parag Kulkarni, Director of Anomaly Solutions Pvt. Ltd. in the academic year 2014-2015.

Ashika Sinha Currently pursuing Bachelor of Engineering from G. H. Raisoni Institute of Engineering and Technology, Pune affiliated to Savitribai Phule Pune University. Opinion MIning from Social Networks is our final year project guided by Mrs. Deeksha Bhardwaj, Head of Computer Department and sponsored by Dr. Parag Kulkarni, Director of Anomaly Solutions Pvt. Ltd. in the academic year 2014-2015.