Classification Of Emotions From Social Media Data

1G. Bharanitharan, 2K. Vijayakumar, 3A. Kumarasan

1,2,3Department of computer science, SKP Engineering College Thiruvannamalai,

Received 28 January 2017; Accepted 22 March 2017; Available online 28 April 2017

Address For Correspondence:
G. Bharanitharan, Department of computer science, SKP Engineering College Thiruvannamalai,
E-mail: G. Bharanigansan208@gmail.com

Copyright © 2017 by authors and American-Eurasian Network for Scientific Information (AENSI Publication).
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

ABSTRACT
The increasing trend of online social networks in different domains, social network analysis has recently become the center of research. Online Social Networks have fetched the interest of researchers for its enormous amount of user generated content, including tweets, blog posts, and forum messages, is created. Researchers have stated that linguistics of the people either in the online or offline communication has strong correlations with their emotions. Twitter, the most popular micro blogging platform is used in this work for the task of emotion classification. The proposed approach includes a methodology for collecting a large number of tweets from Twitter and preprocessing them for removal of unwanted numbers, punctuations, and stop words. The filtered tweets are used to classify the emotions depending on the words which are extracted and processed using Natural Language Tool Kit (NLTK).

KEYWORDS: Social media, social media optimization (SMO), emotions, Block Lis Word Mixing.

INTRODUCTION
Social media is the collection of online communications channels dedicated to community based input, interaction, content-sharing and collaboration. Websites and applications dedicated to forums, microblogging, social networking, social bookmarking, social curation, and wikis are among the different types of social media. The various types of social media data include tweets, facebook posts, blog posts, discussion forums and online reviews.

Social media is becoming an integral part of life as social websites and applications proliferate. Most traditional online media include social components, such as comment fields for users. In business, social media is used to market products, promote brands connect to current customers and foster new business. Social media analytics is the practice of gathering data from blogs and social media websites and analyzing that data to make business decisions. The most common use of social media analytics is to mine customer sentiment to support marketing and customer service activities. Yet another application is to analyze the emotions of the people using social media data. The classification of emotion will be useful in understanding the human behavior particularly in crises situation such as floods, tsunami and earthquakes.

Social media marketing (SMM) takes advantage of social networking to help a company increase brand exposure and broaden customer reach. The goal is usually to create content compelling enough that user will share it with their social networks. One of the key components of SMM is social media optimization (SMO). Like search engine optimization (SEO), SMO is a strategy for drawing new and unique visitors to a website. SMO can be done in two ways: by adding social media links to content such as RSS feeds and sharing buttons, or by promoting activity through social media via status updates, tweets, or blog posts.

ToCite ThisArticle: G. Bharanitharan, K. Vijayakumar, A. Kumarasan., Classification Of Emotions From Social Media Data. Advances in Natural and Applied Sciences. 11(6):Pages:82-88
Social customer relationship marketing (CRM) can be a very powerful business tool. For example, establishing a Facebook page for a company allows people who like a particular brand to like the page, which creates a venue for communication, marketing and networking. Through social media sites, a user can follow conversations about a brand of his choice for real-time market data and feedback. From the customer’s perspective, social media makes it easy to tell a company and everyone else about their experiences with that company, irrespective of whether those experiences are good or bad. The business can also respond very quickly to both positive and negative feedback, attend to customer problems and maintain, regain or rebuild customer confidence.

Enterprise social networking allows a company to connect individuals who share similar business interests or activities. Internally, social tools can help employees’ access information and resources they need to work together effectively and solve business problems. Externally, public social media platforms help an organization stay close to their customers and make it easier to conduct research that they can use to improve business processes and operations.

Social media is also often used for crowdsourcing. Customers can use social networking sites to offer ideas for future products or tweaks to current ones. In Information Technology (IT) projects, crowdsourcing usually involves engaging and blending business and IT services from a mix of internal and external providers, sometimes with input from customers and/or the general public.

Opinion mining technique is an effective means of discovering public opinions. Various companies often use online or system based surveys to collect customer comments. Due to the emergence of social networking sites and applications, people tend to comment on their Facebook or tweet profile. Therefore the system based approach is not an efficient approach. Only a very small customer base can be reached and there is no guarantee that their answers in the survey are honest or not. Facebook, Twitter and all other social media sites are full of people’s opinions about products/services they use, comment about popular personalities and much more. Hence mining opinions about various subject matters from social media is a much more innovative approach for market analysis. A lot of research has been done on opinion mining from social media, most of which focuses on people’s sentiment towards various topics.

On the other hand, the integration of social media in the business world can also pose challenges. Social media policies are designed to set expectations for appropriate behavior and ensure that an employee’s posts will not expose the company to legal problems or public embarrassment. Such policies include directives for when an employee should identify himself as a representative of the company on a social networking website, as well as rules for what types of information can be shared.

Identifying human emotion plays a significant role in behavioral research and human emotions can best describe the character and intent of a person. Some of the emotions observed in human communication are anger, fear, happiness and sad. These emotions are conveyed in social media data either as textual words or as graphical symbols called emoticons and some of them are given below:

- **Anger:** :-(, :-
- **Fear:** :-(, :-(
- **Happy:** :D, : D, : D, :D
glad, happy, positive, relieved
- **Sad:** :-(, :-(, :-(, :-(

Emoticons are reliable indicators of sentiment. Social emotions are sometimes called moral emotions, because they play an important role in morality and moral decision making. Emoticons, such as :) :) :-( and :-(), are frequently used online in social media, blogs, forums, and other kinds of online social interactions. Because they are commonly used in online communications and they are often direct signals of sentiment, emoticons in text were widely used by researchers in tasks such as sentiment analysis as features to machine learning algorithms or as entries of sentiment lexicons for rule-based approaches.

Natural Language Processing (NLP) refers to the method of communicating with an intelligent system using a natural language such as English. NLP tools have the ability to map Natural Language sentences to their syntactic structure and their logic form. NLP techniques are used to classify emotions present in social media data.

The study of existing technique for emotion classification in social media data is presented below:

Thousands of human lives are lost every year around the globe, apart from significant damage on property, animal life, etc., due to natural disasters (e.g., earthquake, flood, tsunami, hurricane and other storms, landslides, cloudburst, heat wave, forest fire).

[1] focuses on reviewing the application of data mining and analytical techniques designed for (i) prediction, (ii) detection, and (iii) development of appropriate disaster management strategy based on the collected data from disasters. A detailed description of availability of data from geological observatories (seismological, hydrological), satellites, remote sensing and newer sources like social networking sites as Twitter was presented.
2 Related Work:

The existing OSN has a marvelous characteristic by which users can customize their wall by restricting some of their friends to comment/post on it. Moreover, it also provides another feature called “Block List”. Block List contains blocked users. If the user doesn’t want to share their ideas to his/her particular friend (or) if the user doesn’t wish to disclose their prevalence/details to their specific friends, then they can block permanently by this feature [Chau and Chen 2008], [Mooney and Roy 2000]. After they are blocked there will be no more relationship with each other in OSN.

Achieving customization in user wall, by restricting some users has lead to the growth of a new thorn called “Vulgarity of messages” [Sebastiani 2002]. To illustrate this issue, two years back, two college students from other state got arrested by the police since one student posted a scurrilous message about a reputed political party on another student’s wall. This news disseminated over OSN like a viral disease and many people commenced commenting on the matter with various unpleasing words [Vanetti et al. 2010]. Meanwhile, the government performed eavesdropping to this matter and took ridiculous action against the students who were the initiative and who made this as a sensational issue. At present, to avoid this kind of problem, users need to block such kind of friends. But those friends may be user’s relatives, close friends, well-wishers etc, who will also post harmless messages many times [Strater and Richter et al. 2007]. But no other solution existed in OSN to avoid this kind of blocking.

3 Existing System:

The study of existing technique for emotion classification in social media data is presented below:

Thousands of human lives are lost every year around the globe, apart from significant damage on property, animal life, etc., due to natural disasters (e.g., earthquake, flood, tsunami, hurricane and other storms, landslides, cloudburst, heatwave, forestfire).

[1] focuses on reviewing the application of data mining and analytical techniques designed for (i) prediction, (ii) detection, and (iii) development of appropriate disaster management strategy based on the collected data from disasters. A detailed description of availability of data from geological observatories (seismological, hydrological), satellites, remote sensing and newer sources like social networking sites as Twitter was presented.

4 Proposed System:

An extensive and in-depth literature study on current techniques for disaster prediction, detection and management was done and the results were summarized according to various types of disasters. Finally a framework for building a disaster management database for India hosted on open source Big Data platform like Hadoop in a phased manner has been proposed. The proposed system was designed to collect the tweets from Twitter and classify the emotions present in the tweets. The system could automatically distinguish between the five emotional classes, happy, fear, anger, sad and others on four topics discussed in media during 2015-2016. Python has been used along with NLTK for implementation of the proposed system for emotion classification.

5 System Architecture:

![System Architecture](image)

Fig. 1: System Architecture
6 Modules Description:

The first step was to collect a large set of tweets relating to four significant topics trending in news headlines during 2015-2016. For this purpose, the Python package tweetstream has been used to retrieve tweets related to the following four topics:

- cauvery
- rio
- flood
- election

The tweetstream package fetches tweets from Twitter’s streaming API in real-time. StreamingAPI gives several thousand tweets related to search HASHTAG. The snapshot presented in Fig 2 shows the Twitter data which are retrieved for the Hashtag ‘#cauvery’.

![Fig. 2: Twitter Data Collection for #cauvery](image1)

The collected twitter data is preprocessed to remove the unwanted numbers, punctuations, URLs and stop words, so as to enable emotions classification based on words. Preprocessing includes filtering, tokenization and constructing n-grams, which are explained, Fig 3 shows the preprocessing results for #cauverytwitter data.

![Fig. 3: Preprocessing](image2)
Fig 3 shows the result of tokenization, which is segmentation of text by splitting it by spaces and punctuation marks, to form a bag of words.

![Tokenization](image)

**Fig. 4: Tokenization**

The filtering process removes URL links, user names with symbol @ indicating a user name, Twitter special words (such as “RT”6), and emoticons and Fig 4 shows the filtered dataset.

![Filtered dataset](image)

**Fig. 5: Filtered dataset for #cauvery**

After automatic removal of non-English tweets, retweets, and duplicated tweets, approximately 50000 tweets remained. An average tweet in the dataset contains a wide variety of words including these communicating human emotions.

7 Algorithm:

Technique: Word Mixing

1. To combine (substances, elements, things, etc.) into one mass, collection, or assemblage, generally with a thorough blending of the constituents.

2. To put together indiscriminately or confusedly (often followed by up).
Input: $T_s$ - Training Set with binary class label values; $acl$ - No. of record for which $cl = cl_j$; 
$n$ - No. of training example for which $cl = cl_j$ 
Set 'X' to 3 [Arbitrary Constant] 
Set 'P' to 0.5 [priori estimate for $p(ai | cl_j)$] 
Get a[i], cl[j] where i,j are from 1 to n 
For i 1 to n do 
For j 1 to n do 
$P(ai | cl_j) = (acl + xp) / (x + n)$ 
End for 
End for 
Set cl[j] to “post” [class label value-1] 
For i 1 to n do 
For j 1 to n do 
If cl[j] equals post then 
Set $P(a(cl_j))$ to $P(cl_j) * P(ai | cl_j)$ 
Else 
Set $P(b(cl_j))$ to $P(cl_j) * P(ai | cl_j)$ 
End If 
End for 
End for 
Set MLC to argmax ($P(a(cl_j))$, $P(b(cl_j))$) 
Output: MLC [Most likelyhood classification] 

Conclusion: 
To identify tweets corresponding to the emotions, “happy”, “anger”, “fear”, “sad” and “other”, the word dictionaries have been used. Using Python API package, word matching and scoring has been done, and the Python Natural Language Toolkit (NLTK) has been used to identify the words. After the analysis, the preprocessed text is classified into 5 categories of emotions ie, “happy”, “anger”, “fear”, “sad” and “other”. Fig 5 shows the result of emotion classification on tweets from cauvery river water sharing issue. 

Fig. 5: Result of emotion classification on tweets from cauvery river water sharing issue.

The results of the experiments conducted on emotion classification using Twitter data collected using Twitter API on 4 news topics is presented in Table 2. The proposed method has been used to classify the news topics discussed in Twitter. The proposed system identified that in most of the cases, people had anger in all the issues. In #cauvery, 58.22% of tweets have been found to convey anger emotion. 21.08% of neutral emotions have been classified under "Others" category in #Rio. 49.11% of tweets from China Flood have been found to belong to anger category in #flood. Similarly 49.88% of the tweets from #election have been classified under “anger” category. 

Table 1: Emotion Classification Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Happy</th>
<th>Sad</th>
<th>Anger</th>
<th>Fear</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cauvery</td>
<td>13.70%</td>
<td>5.65%</td>
<td>58.22%</td>
<td>6.25%</td>
<td>16.17%</td>
</tr>
<tr>
<td>Rio</td>
<td>13.07%</td>
<td>10.53%</td>
<td>46.39%</td>
<td>8.92%</td>
<td>21.08%</td>
</tr>
<tr>
<td>Flood</td>
<td>7.74%</td>
<td>8.85%</td>
<td>49.11%</td>
<td>13.93%</td>
<td>20.35%</td>
</tr>
<tr>
<td>Election</td>
<td>13.16%</td>
<td>8.64%</td>
<td>49.88%</td>
<td>8.60%</td>
<td>19.71%</td>
</tr>
</tbody>
</table>
REFERENCES