The Indian Journal Of Basic and Applied Research
Volume 1 |Issue 1| June 2015
Online & print@www.ijbar.co.in

Opinion Mining on Twitter Feeds

Varsha Verma[1], Kunal Epan[2]

ABSTRACT:

Opinion Mining is the field of study that analyses people's opinions, views and emotions from text. It is one of the most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. And with the help of technological advances, we now have a huge volume of opinionated data recorded in digital form for analysis. Opinion Mining systems are being applied in almost every business and social domain because opinions are central to almost all human activities and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the choices we make, are largely conditioned on how others see and evaluate the world. For this reason, when we need to make a decision we often seek out the opinions of others. This is true not only for individuals but also for organizations. Twitter is an active medium to express opinions and views of the public in general. We present a system that will classify the most recent tweets extracted according to a search query into positive and negative classes, thus generating an overall opinion of the public. We will be using classification algorithms in order to classify the tweets into the positive and negative categories.

INTRODUCTION:

Opinion Mining, commonly referred to as sentiment analysis is used to extract the information from a text with respect to its polarity to the given subject. The goal of such a system is to effectively process subjective information. It identifies opinion bearing words, which are usually adjectives which define the nature and polarity of the text. In general, opinions can be expressed about anything, a product, a service, a topic, an individual, an organization or an event. It allows classification of the text or feature, i.e. whether the polarity of the opinion is positive, negative or neutral. Given a set of tweets, the system classifies them into positive and negative classes. This is clearly a classification learning problem. But it is also different from topic based text classification. In topic based classification, topic related words are important. However, in this case, opinion related words are important for us, e.g. good, bad , worst, excellent, horrible, etc. On the basis of these opinion words and proper weightage to the features, one can obtain the overall sentiment of the tweet related to the topic.

METHODOLOGY:

In this project, we will be doing Opinion Mining on tweets pertaining to a particular search topic. The most recent tweets related to the topic will be searched on twitter, after which their polarity will be classified as positive or negative. As a result, the current overall sentiment among the public related to that particular topic can be obtained by applying this on a large set of tweets extracted from the public at a particular time. For the classification task, a Twitter Sentiment Corpus containing 1.6 million classified tweets is used as a training set to build our classifier. It is to be used to build a classifier which will classify the tweets searched as positive and negative. On the basis of this classified data on the search dataset, we need to determine the overall public opinion of the search query topic. Given a tweet **T**, searched from the most recent tweets from the search query, classify it into either of the two classes – positive or negative. Repeat the process for the entire set of current tweets and thus get the overall opinion of the search query among the public.

Literature Survey:

OPINION MINING TECHNIQUES:

1. PROBABILISTIC CLASSIFIERS:

Probabilistic classifiers use mixture models for classification. The mixture model assumes that each class is a component of the mixture. Each mixture component is a generative model that provides the probability of sampling a particular term for that component. Three of the most famous probabilistic classifiers are discussed here.

1.1 NAÏVE BAYES CLASSIFIER (NB):

The Naïve Bayes Classifier is the simplest and most commonly used classifier. It computes the posterior probability of aclass based on the distribution of words in a document. P(label) is the prior probability of a label or the likelihood that a random feature set the label. P(features|label) is the prior probability that a given feature set is being classified as a label. P(features) is the prior probability that a given feature set is occurred.

1.2 BAYESIAN NETWORK (BN) :

The main assumption of the NB classifier is the independence of the features. The other extreme assumption is to assume that all the features are fully dependent. This leads to the Bayesian Network model which is a directed acyclic graph whose nodes represent random variables, and edges represent conditional dependencies. BN is considered a complete model for the variables and their relationships.

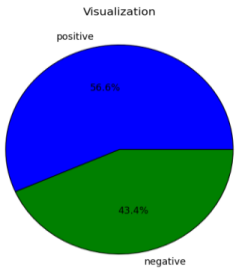
RESULTS AND DISCUSSIONS:

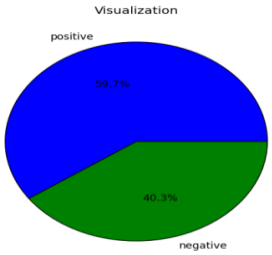
METHOD	ACCURACY(%)	TRAINING DATA SIZE (number of tweets)	TEST DATA SIZE (number of tweets)
Naïve Bayes'	69.97%	20,000	500
Maximum Entropy	73.36%	5,000	500

The training for Maximum Entropy is much more resource consuming than that of Naïve Bayes' because of the exponential calculation of Entropies of its features. Hence, training takes a greater time for Maximum Entropy Model. Since it doesn't consider independence of features, unlike Naïve Bayes', it is likely to give better results than that of Naïve Bayes, as evident from the results of the Stanford Test dataset.

TOPIC	POSITIVE SENTIMENT (%) NB	NEGATIVE SENTIMENT (%) NB	POSITIVE SENTIMENT (%) ME	NEGATIVE SENTIMENT (%) ME	NET SENTIMENT
Phil Hughes	42.645	57.354	32.713	67.286	NEGATIVE
Ebola	44.390	55.610	36.569	63.430	NEGATIVE
Interstellar	56.61	43.39	59.72	40.28	POSITIVE

```
RT @amabirdman: Pretty sure this was the inspiration behind the spaceship in Interstellar :P http://t.co/kj2OjHrnrS
Just saw Interstellar but now how will I ever be able to think about anything besides time dilation and black holes ever again
"Tars, you da real mvp  #Interstellar"
"Interstellar.. Dang, seriously mind blown"
We just watched interstellar with john's parents and hoLY SHIT THAT WAS GREAT
The organ crescendos in the Interstellar music &gt;&gt;&gt;&gt;
RT @onemilerunboy: Interstellar turn up with @whatssophanny @brandenteguh @Zach_RG15 @jilliaaannnn and jackkkkkk
" @Reading ""A Vector Alphabet of Interstellar Travel by Yoon Ha Lee | http://t.co/ONz7J9FSwJ"" http://t.co/iaMq86xA8K"
RT @CaribbeanAccent: Well. #Interstellar is going to either blow your mind in an epic way or leave you confused and possibly pissed. #Inter…
Pretty sure this was the inspiration behind the spaceship in Interstellar :P http://t.co/kj2OjHrnrS
RT @MaddieTayRich05: INTERSTELLAR WAS BRILLIANT. PURE BRILLIANCE.
#Interstellar GREAT MOVIE !!
The most incredible SciFi in a long time. http://t.co/hVWN9Son1G
Still wanting to see Interstellar.....
Mocking jay and interstellar is so disappointing
RT @shelbygwalt: Interstellar is by far the best movie I've seen in a long time.
Go watch interstellar
INTERSTELLAR WAS THE BEST MOVIE EVER
RT @janimall: Let the deep space thoughts begin #Interstellar
"RT @Interstellar: Love and Gravity, how science and emotion mingle to create a powerful atmosphere in #Interstellar. http://t.co/VVRBS4OBOf…"
RT @gentlemanabroad: #Interstellar unlikely to change his mind... Prestige Author Slams Nolan's Dark Knight Trilogy http://t.co/Qvp4QSmXHd …
After seeing interstellar it is the second best movie I have ever seen only behind the hobbit
Interstellar got me fucked up...
"Shoutout to all the Physics I learned this semester, it made the movie Interstellar all the more interesting "
Interstellar was fucking incredible
"RT @RachelRizzo: If u havent yet seen #Interstellar, stop what ur doing now &amp;amp; go see it. Also, as someone obsessed w movie soundtracks, thi…"
"RT @LeeAnnMiller10: Interstellar was so ridiculously good, everyone should see it"
"Somewhere ""Beyond the Wormhole"" Albert Einstein &amp;amp; Stanley Kubrick are giving Interstellar 2 thumbs up."
Hypnotizing spinning moment..whew!! #interstellar http://t.co/GyEyNosKId #nolan #zimmer #bgm #spectacular
"Holy shit, interstellar was a fucking great ass movie! Highly recommend it but NO bathroom breaks"
Pacific Science Center hangouts before Interstellar. It was... http://t.co/MS1ztHpMyG http://t.co/ct9dThE69H
Interstellar turn up with @whatssophanny @brandenteguh @Zach_RG15 @jilliaaannnn and jackkkkkk
HELLO YES IF IGO SEE INTERSTELLAR WILL I CRY
RT @FaZeJev: I just got home from seeing Interstellar... I can't look at life the same.
RT @mackfrere: Christopher Nolan does it again #Interstellar
"#Interstellar should be shown in all high school science classes as a ""fun end of the year movie."""
"People hyped up #Interstellar way too much, I'd give it 3 out of 5 stars. #NotSoStellar"
My brain is fried after watching Interstellar #trippyshit
"Interstellar was even better the second time I watched it. If you haven't seen it, you need to. It's incredible."
Interstellar was just made for people to watch it twice
Interstellar is fave movie ever 
I thought Interstellar was 99% Bullshit.  But it was very inspiring to think about.
Post interstellar http://t.co/ubCcEIxAB7
RT @Zach_Seay: MATTHEW MCCONAUGHEY COULDVE SPENT AN HOUR AND A HALF ON THAT WATER PLANET IN INTERSTELLAR + IN DAT TIME UVA STILL WOULDNT HA…
```





References:

1. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval Vol. 2, No 1-2* (2008) 1–135c © 2008 Bo Pang and LillianLee.
2. Predicting the semantic orientation of adjectives. *Published in EACL '97 proceedings of the 8th conference on European chapter of the Association for Computational Linguistics Pages 174-181.* Stroudsburg PA, USA © 1997 Vasileios Hatzivassiloglou, Kathleen R. Mckeown.
3. Twitter sentiment classification using distant supervision. *Published in CS224N Project Report, Stanford Page 1-12 2009.* Alex Go, Richa Bhayani, Lei Huang.
4. Thumbs up? Sentiment analysis using machine learning techniques. *Published in proceedings of EMNLP Pages 79-86 2002.* Bo Pang, Lillian Lee, Shivakumar Vaithyanathan.
5. Twitter as a corpus for sentiment analysis and opinion mining. *Published in proceedings of LREC 2010.* Alexander Pak, Patrick Paroubek.
6. Stanford Twitter Sentiment Corpus - http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip

.