ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

IndiSent Analysis in Twitter using Machine Learning Methods

Neelima, Dr. Ela Kumar

M.Tech (CSE) Student, IGDTUW Kashmere Gate, Delhi, India

Professor, Department of CSE, IGDTUW Kashmere Gate, Delhi, India

ABSTRACT: This paper presents a new idea for sentiment analysis in twitter, especially for the Indian users. Sentiment analysis is one of the best tool to measure sentiments of the users hidden behind their text. But it is possible that sentiments are not analysed successfully due to some barriers.Indeed, the task of automatic sentiment recognition in online text becomes more difficult for all the aforementioned reasons like limited size of character i.e. 140, unlimited spelling mistakes, slang words and different languages. In our research, the primary and underlying idea is that the fact of knowing how people feel about certain topics can be considered as a classification task and removing the language barrier using Google Translator. Twitter is used for the collection of data corpus in multilingual. Collected raw dataset is transformed into standard language i.e. English, and used for polarity classification of data corpus. Hence, we present how sentiment analysis can assist different languages, by Google translator and experimental results on the Naive Bayes Classifier and Maximum Entropy classifier and comparison the results of these two. Experimental results show that our proposed techniques are efficient and performs better than previously proposed methods. We worked with Hinglish, however, the proposed technique can be used with any other language.

KEYWORDS: Google translator, data corpus, lexicon, SentiWordNet, sentiment analysis, twitter, Naïve Bayes Classifier and Maximum Entropy Classifier.

I. INTRODUCTION

With the explosive growth of the social media content on the Internet in the past few years, people now express their views on almost anything in discussion. There are many microblogging websites like Twitter, Facebook, and Tumbler etc.Twitter has become a very popular communication tool among Internet users and it is one of the most open and simplest platform to share their opinions on different topic. Nowadays people used to review the comments and posts of the customer before purchasing any product. So, we can say that social media consists hidden opinions of the users. Finding the opinion sites and monitoring them on the web is difficult task. Thus there is a need for automatic opinion mining and summarization systems. Sentiment is the opinion, emotion, feeling, attitude, thoughts or behaviour of the user. And Sentiment Analysis is a method for identifying the ways in which sentiment is expressed in texts. Specifically, it is an analysis of the opinions and emotions hidden behind the text in form of comment, post, tweet, reply etc. Now, what is IndiSent Analysis? Indisent Analysis is made up of two terms: India + Sentiment Analysis. Therefore IndiSent Analysis means Sentiment Analysis of Indian tweets. It is difficult to analysed Hindi tweets because system treats Hindi as junk and discarded.

There are three levels of sentiment analysis in [1] , Coarse level in which sentiment of whole document is considered, second is Mid-level sentiment analysis in which sentiment of the sentence is determine and last Fine level deals with attribute level sentiment analysis.

In Neethu [2] Symbolic techniques or Knowledge base approach and Machine learning techniques are the two main techniques used in sentiment analysis. Knowledge base approach requires a large database of redefined emotions and an efficient knowledge representation for identifying sentiments. Machine learning approach makes use of a training set to develop a sentiment classifier that classifies sentiments. Since a predefined database of entire emotions is not required for machine learning approach, it is rather simpler than Knowledge base approach.

In [3, 4], we study how twitter can be used for sentiment analysis purposes. We show how to use Twitter as a corpus for sentiment analysis and opinion mining. We use twitter for the following reasons:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

a) Huge data corpus
Twitter contains an enormous number of text posts and it grows every day. The collected data corpus can be arbitrarily large. Because increasing number of people begin to use twitter as a main way to express their attitudes and they can put forth their reviews at any time using computers as well as cell phones, data should be processed with high speed.

b) Diversity of contents
The contents of twitter is not limited to any specific topics like politics or economy, because people can write a review about any aspect of their lives. Twitter's audience varies from regular users to celebrities, company representatives, politicians, and even country presidents. Therefore, it is possible to collect text posts of users from different social and interests groups. Twitter's audience is represented by users from many countries. Although it is possible to collect data in different languages.

c) Real time Analyzation
When the tweet is published, others can read it almost at the same time. Update and deletion are also real time. And now sentiments are also classified in real time.

d) Short length of Tweets
Tweet is usually very short, even just a shortsentence. It avoids describing every details andfocus on personal sentiments.

Besides the challenges traditional sentiment analysis systems face following additional difficulties:

a) Short Length of text.
b) Spelling Mistakes.
c) Special tokens like URLs, emoticons.
d) Diversity of content.
e) Different style of Language.
f) Multilingual content.
g) Slang words

II. RELATED WORK

A. Related Research of Microblog all over the world

Munmum [5] reported an interesting finding that Moods play a critical role in our everyday lives; they fundamentally direct our attention and responses to environment, frame our attitudes, and are vital to maintaining healthy social relationships.
Social networking sites play vital role in our day to day life. Author [6] and [7] have shown the sentiments analysis using Facebook. Sentiment analysis of twitter messages has recently received great attention from both research and industry. [8] analysed over 150,000 twitter posts, and found out that 19% of the posts mention a brand name, and 20% of these contained some expression of brand sentiments, and 50% of these had positive, and 33% were critical of the company or product.

B. There are two approaches to perform the task of sentiment analysis.

According to [9] they are categorized as: Machine Learning (Supervised) or Natural language Processing (Unsupervised).

1) Natural Language Processing/Symbolic Technique (Unsupervised)

Much of the research in unsupervised sentiment classification using symbolic techniques makes use of available lexical resources. In [6] Symbolic techniques that focuses on the force and direction of individual words (the so-called "bag of words" approach). In that approach, relationships between the individual words are not considered and a document is represented as a mere collection of words.

We can divide the symbolic technique into two subparts [10]:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

a) Dictionary Based Technique: The idea of dictionary-based approach is first to collect some small set of seed words with their known opinion polarities and then to iteratively extend them with the help of online dictionary e.g. WordNet.WordNet database consists of words connected by synonym relations. WordNet is the popular lexical resource used to determine the overall sentiment, sentiments of every word is determined and those values are combined with some aggregation functions.
There exist a number of affective lexicons [11] for English, Spanish, German and other languages. However, only a few of such resources are available for French.

b) Corpus Based Technique: This approach uses as its bases the syntactic or co-occurrence patterns and the predefined set of seed words with its polarities. SentiWordNet is a widely-used English sentiment lexical resource that was generated by automatically annotating the synsets of the WordNet, where each synset received the three scores indicating to which extend the respective synset is to be regarded positive, negative or objective. The sum of all three scores always equals one.

2) Machine Learning (Supervised)

A number of machine learning techniques like Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM) are used to classify reviews.
In [12] it is found that Naive Bayes works well for certain problems with highly dependent features. This is surprising as the basic assumption of Naive Bayes is that the features are independent.
In [13] the machine learning approach on average claims an accuracy rate of 83%. The work uses a machine learning approach for classification. The f-measure is used as metric for evaluation, and claims efficiency up to 70%.

C. **Other models based on Sentiment Analysis**

Diffusion Estimation model is given by [14] predicts the sentiment in each user next tweet on topicusing these variables (and several other variables not mentioned here) and a support vector machine classifier.
In [8] a probabilistic generative model specifies a stochastic procedure by which data can be generated, usually making reference to unobserved random variables that express latent structure. Statistical inference probability distributions over latent variables are computed (higher order uncertainty approximation) conditioned on a given dataset.
In [15] Sentiment Polarity Identification approach represents an efficient combination of techniques that are focused on constructing an efficient structured representation of the natural language text, using words that contain polarity and eliminating words that do not have discriminative power in the classification process.
VIKOR [1] which is a compromise ranking method of the multi criteria decision making (MCDM) approach, customer satisfaction for mobile services can be accurately measured by a sentiment-analysis scheme that simultaneously considers maximum group utility and individual regret.
It is shown in [16] SiLA (Similarity Learning Algorithm) is a variant of Voted Perception algorithm. It learns diagonal, symmetric or square matrices depending on the problem to solve. Although the main algorithm is online, there is also an online to batch conversion. The online update iteratively improves the similarity matrix A whenever the current similarity matrix fails to correctly classify an example.
A Self Learned framework[17] where prior knowledge from a generic sentiment lexicon is used to build a classifier where preferences on expectations of sentiment labels of those lexicon words are expressed using generalized expectation criteria. Pseudo-labeled documents by this classifier are used to automatically acquire domain-specific feature words whose word-class distributions are estimated and are subsequently used to train another classifier by constraining the model's predictions on unlabeled instances.
HMM-LDA(Hidden markov model- Latent Dirichlet Allocation) [18] a set of new feature selection schemes that use a Content and Syntax model to automatically learn a set of features in a review document by separating the entities that are being reviewed from the subjective expressions that describe those entities in terms of polarities. By focusing only on the subjective expressions and ignoring the entities, we can choose more salient features for document level sentiment analysis.

III. PROPOSED METHODOLOGY

The main methodology for IndiSent Analysis is divided into two levels, in which level 0 showing the basic steps of Sentiment Analysis and level 1 showing the advanced steps of IndiSent Analysis.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

1) Level 0 is followed by four basic steps of Sentiment Analysis: Data Extraction, Preprocessing of data, Creation of feature vector, and Classification algorithm shown in figure 1.

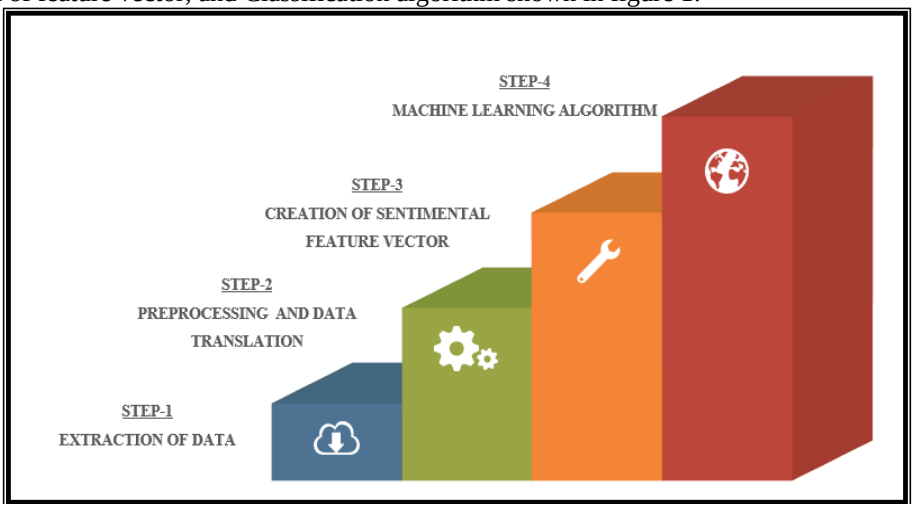


Figure 1. Level 0 diagram of IndiSent Analysis

2) Advanced steps of IndiSent Analysis is given in level 1 shown in figure 2.

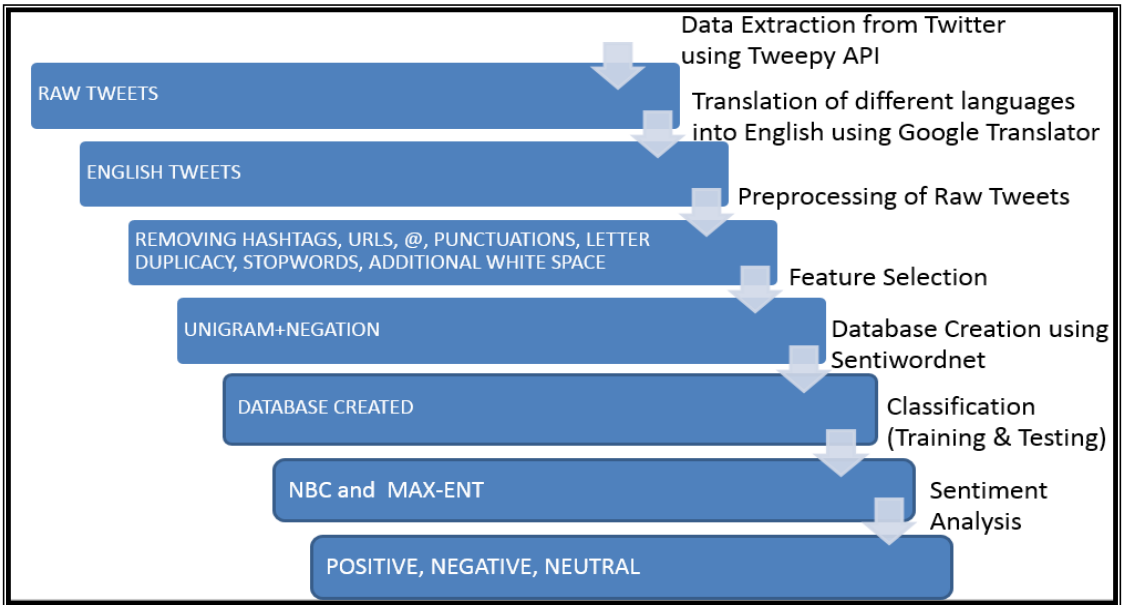


Figure 2. Level 1 diagram of IndiSent Analysis

A. Data Extraction

Data is extracted from the twitter in form of tweets using tweepy api. In this step, data corpus is collected in raw form, full of impurities and unwanted words.

For the Data collection and preparation we are using the Twitter API 2.3.0 for our data collection. The API class provides access to the entire twitter RESTful API methods. The Tweepy API offered by the Twitter service especially for python. Tweepy is open-sourced and enables Python to communicate with Twitter platform and use its API.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Tweepy supports accessing Twitter via Basic Authentication and the newer method, OAuth. Twitter has stopped accepting Basic Authentication so OAuth is now the only way to use the Twitter API.
Two databases are developed for the experiment. One database is implemented with the Google Translator and second one is without google Translator. Purposes of using two different databases is to determine that the proposed system is work good than the previous one or not?

B. Translation

Raw data is present in different languages e.g., In India mostly people like to tweet in English and Hindi and Hinglish (Combination of two). So, First step of cleaning of data is to convert all the languages into single one. And English is international language, so we translate all the raw data into English using google translator.
In SA it is concluded that, one of the major problem is that more than 50% tweets are treated as junks. From the previous study it has concluded that, in the country like Germany, France and India etc. where they use to prefer their native language instead of English. Researchers have to first build a dictionary for their native language and give score correspond to the existing dictionary. And then do some more practice. Instead of doing other process we first convert raw data corpus present in any language (in our case: Hindi) into English language using Google translator.

C. Preprocessing

Now, all the junks are cleaned and preprocess in the step of preprocessing. In this step we are doing: Removing # tags, removing @, removing urls, removing special symbols like punctuation marks, most important removal of stop words and additional white space. A full overview of the pre-processing methods is given in Table 1.

Table 1: Features that are usually removed from the tweets.

WWW. / https /	URL	Typically a link
@	Mention	Lag to mention another user
#	Hash tag	Used to tag a tweet
yaaaaaaahooooo	Letter duplicity	Sign of excessive joy or sad
My Name Is Neelima	Words in Upper and lower case	Used to write something
! ? . , “ ‘ -	Punctuation	Used for special purpose
N ee lim a	Additional white space	Misspelling or slang word
RT	Retweet	Reposting another’s tweet

a) Replacement of URL

There might be a chance that URLs are relevant for the sentiment. Not necessarily the value of the URL itself, but the fact that there are references to URLs. To make these features more informative for the machine learning algorithms, a pre-processing method was implemented for eliminating them. I don't intend to follow the short urls and determine the content of the site, so we can eliminate all of these URLs via regular expression matching or replace with generic word URL.

b) Replacing of @

This means that a user name like @username is replaced by the generic word AT_USER was chosen as it is very unlikely that it would be a part of the original.

c) Replacing of hashtags

#hashtag hash tags can give us some useful information, so it is useful to replace them with the exact same word without the hash. E.g. #nike replaced with 'nike'.

d) Reducing Letter Duplication

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

By reducing and normalizing excessive letter duplications like "I'm soooooo happyyyyyyyy!!!!!!", the vocabulary is reduced and the classification can perform better. However, there might be that the word "sooooo" is more sentimentally charged than the proper form of the word "so". The sentence "I'm soooooooo happyyyyyyyy!!!!!!" is more positive polar than the statement "I'm so happy!". To reduce the vocabulary but not lose information, all duplicates of more than two consecutive characters are reduced to exactly two duplications. So "I'm soooooooo happyyyyyyyy!!!!!!" would be changed to "I'm soo happyy!!!", and the same goes for "I'm sooooooooooooooo happyyyyyyyy!!!!!!" This way the additional sentiment is preserved.

e) Convert into Lower Case

It is necessity that all words are treated at the same level. So, all the tweets are converted into lower case.

f) Removal of Punctuation sign

All the special symbols, we can say punctuation marks have to be removed from here. Because they represent no sentiments and treated as a junk. So it is better to remove all there symbol and clean the data.

g) Remove additional white space

User are doing many mistakes while writing tweets. Removal of additional white space means that removal of multiple white space with the single white space. So that data is presented in a single stream of line without any extra white space and every single word is differentiated from others.

h) Removal of RT

RT is retweet. Retweet is defined as reposting of other user's tweet. If tweet on the same topic has already been added them retweet is iterated as a junk and discarded because of no further use.

D. Feature Words Generation

In this paper, selection of the feature word is based on the three techniques: Unigram, Negation and Part of speech (using NLTK). In the feature word selection step, preprocessed data are segmented into unigram (n-gram) and every single word is treated as a feature. Negation are attached with the unigram method to improve accuracy and determine the real sentiments.

a) Unigrams

Unigram is the classic approach to feature word selection, in which each document is represented as a feature vector, where the elements indicate the presence (or frequency) of a word in the document. In other words, the document is represented by its keywords [19].

b) Attaching Negation

In [15] the negation word "not" can change the polarity of an entire sentence. The sentence "I am happy" is obviously positive, and "I am not happy" negative. If using unigrams, the features of the second sentence would have been (I; am; not; happy) and for the first one (I; am; happy). Both having the word "happy" in them, and if that were to be a significant feature for positive classification, the negative sentence could be classified as positive. By attaching the negation word to the preceding and the following word, the features will also reflect the change in polarity. So the features for the negative sentence would be (I; am; not; not happy).

E. Database Creation

After the selection of feature words, database is created correspond to different sentiments. This database is created with the help of Sentiwordnet dictionary. Database of feature word is created using Lexicon resources. Data is segregated into three: positive, negative and neutral and automatic labeling is done by the sentiwordnet.

The Sentiword line also includes a positive score and a negative score that is basically what we have to use in our system to calculate if a tweet is positive or negative. Checking if a tweet has words on this dictionary and summing all those scores.

F. Classifier

After successfully implementation of all the steps, Classification procedure is begin. Naïve Bayes and Maximum entropy are used to classify the text. First data is train through the two classifier and then testing has to be done.

Actual sentiments are determine through the best classifier in the form of negative, positive and neutral.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

In this experiment we tested the impact of training set size on classifier performance. We took a random sample of 10% from the dataset and used it as a test set. Then we trained different classifiers on different sized portions of the remaining 90%. In the experiment, we compare the Naive Bayes (NB) and Maximum Entropy classifier, each trained with the full feature set.

a. Naïve Bayes Classifier

The Naïve Bayes (NB) classifier is based on Bayes rule, a practical Bayesian learning model that is easy to understand and implement. The Bayes rule allows us to determine this probability of any event. It is the probabilistic approach to the text classification. Here the class labels are known and the goal is to create probabilistic models, which can be used to classify new texts. It is specifically formulated for text and makes use of text specific characteristics. The NB classifier is based on the assumption that all the attribute values are conditionally independent given the target value of the instance. There are inbuilt library for Naïve Bayes in NLTK. The detailed description of this algorithm can be found here [10]:

$$v_{NB} = \arg\max_{v_j \in V} P(v_j) \prod_i^n p(a_i|v_j) \qquad \text{Equation (1)}$$

Here P is the probability, 'a' is the feature word [a1, a2, a3…. an] and v is class label [positive, negative, and neutral].

b. Maximum Entropy

Maximum Entropy (MaxEnt) is a multinomial logistic regression model that allows for classification with more than two discrete classes. The principle in MaxEnt is to model all that is known and assume nothing about that which is unknown. In other words, if you have some knowledge about a domain, choose a model that is consistent with the knowledge, but otherwise as uniform as possible. In [12] the MaxEnt models are feature-based, and in binary classification scenarios it is the same as general logistic reasoning. Unlike NB it has no assumptions of conditionally independence, and can therefore be used with feature selection methods like n-grams and extended unigrams (unigrams with negation support). There is also inbuilt library for Maximum Entropy in NLTK. But NLTK do not support SVM anymore. In [2], no assumptions are taken regarding the relationship between features. This classifier always tries to maximize the entropy of the system by estimating the conditional distribution of the class label. The conditional distribution is defined as

$$P_{lamda}(y|X) = \frac{1}{Z(X) \exp\{\sum_i lamda_i f_i(X,y)\}} \qquad \text{Equation (2)}$$

'X' is the feature vector and 'y' is the class label. Z(X) is the normalization factor and $Lamda_i$ is the weight coefficient. Fi(X, y) is the feature function which is defined as

$$f_i(X,y) = \begin{cases} 1, & X = x_i \text{ and } y = y_i \\ 0, & otherwise \end{cases} \qquad \text{Equation (3)}$$

IV. EVALUATION

We considered two databases (Database1 and Database2) for the evaluation of the proposed system. Database 1(D1) consist tweets in the form of English + Hindi language, extracted from the twitter using Tweepy API and then preprocessed with the aforementioned method Goggle translator.
And the second Database (D2) is also same as the first, extracted through Tweepy API and then preprocessed. But here, no translator is used for the translation of languages.

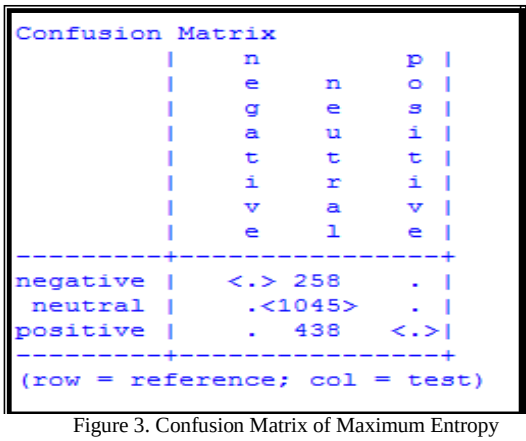
Confusion Matrix of Max-Ent and Naïve Bayes showing true positive, true negative, false positive and false negative box [D1].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

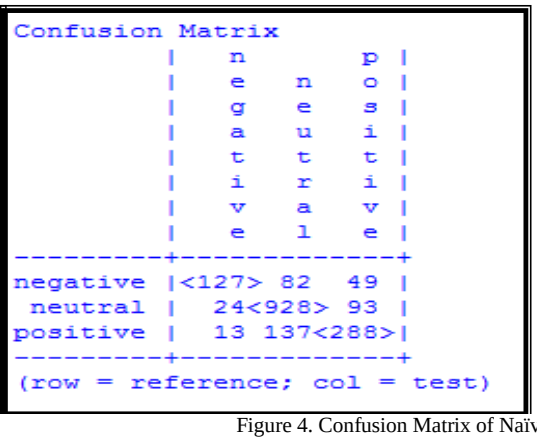
(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015



```
Confusion Matrix
          |    n          p |
          |    e    n     o |
          |    g    e     s |
          |    a    u     i |
          |    t    t     t |
          |    i    r     i |
          |    v    a     v |
          |    e    l     e |
----------+-----------------+
negative  |   <.> 258     .  |
 neutral  |    .<1045>    .  |
positive  |    .   438   <.> |
----------+-----------------+
(row = reference; col = test)
```

Figure 3. Confusion Matrix of Maximum Entropy



```
Confusion Matrix
          |    n          p |
          |    e    n     o |
          |    g    e     s |
          |    a    u     i |
          |    t    t     t |
          |    i    r     i |
          |    v    a     v |
          |    e    l     e |
----------+-----------------+
negative  |<127> 82   49    |
 neutral  |  24<928>  93    |
positive  |  13  137<288>   |
----------+-----------------+
(row = reference; col = test)
```

Figure 4. Confusion Matrix of Naïve Bayes Classifier

The comparative study of various classifier for Sentiment Analysis in twitter is presented by various parameters like precision, recall and f-measure shown below:

Table 2: Comparison of Naïve Bayes and Maximum Entropy using Google Translator (Database1)

		NAÏVE BAYES	MAXIMUM ENTROPY
PRECISION	POSITIVE	0.66	0.0
	NEUTRAL	0.80	1.0
	NEGATIVE	0.77	0.0
RECALL	POSITIVE	0.65	None
	NEUTRAL	0.88	0.60
	NEGATIVE	0.49	None
F-MEASURE	POSITIVE	0.66	None
	NEUTRAL	0.82	0.79
	NEGATIVE	0.69	None
ACCURACY	MEAN OF THREE	0.77	0.60

Table 3: Comparison of Naïve Bayes and Maximum Entropy without Google Translator (Database2)

		NAÏVE BAYES	MAXIMUM ENTROPY
PRECISION	POSITIVE	0.43	1.0
	NEUTRAL	0.60	0.0
	NEGATIVE	0.62	0.0
RECALL	POSITIVE	0.58	0.37
	NEUTRAL	0.32	None
	NEGATIVE	0.63	None
F-MEASURE	POSITIVE	0.46	0.54
	NEUTRAL	0.51	None
	NEGATIVE	0.62	None
ACCURACY	MEAN OF THREE	0.52	0.37

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

V. CONCLUSION

We tested two different base machine learning algorithms: Naïve Bayes, and Maximum Entropy for sentiment. From experimenting with these setups, we found that machine learning algorithms perform well when we used Google translator with the data corpus. Our experiments show that Naïve Bayes classifier perform best over MaxEnt in classification. Given the confusion matrices and results on same data corpus with different techniques.

As negation were proven informative features, support for "NOT" as feature would be an improvement. By reducing the training set, we obtained better classification for negative and positive tweets separately, but when the number of instances was reduced to a perfect balance between positive, neutral and negative classes, the total number of tweets was too low to train a well-performing classifier. A larger dataset would reveal if training with more balanced data would give a better performing classifier.

VI. FUTURE WORK

An obvious way to extend this work would be to add other classification algorithms to the, e.g., Conditional Random Fields or more elaborate ensembles. There are also several features and feature selection methods that could be investigated and a less naïve way of handling negation. Rather than the simple treatment of negation used here, an approach to automatic induction of scope through a negation detector [20] could be used. Relational features could also be added, as shown by Karlgren et al. [21] and Johansson and Moschitti [22].

REFERENCES

[1] Daekook Kang, Yongtae Park. Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications 41 (2014) 1041–1050.
[2] Neethu M S Rajasree R. Sentiment Analysis in Twitter using Machine Learning Techniques. 4th ICCCNT 2013 July 4 - 6, 2013, Tiruchengode, India IEEE – 31661.
[3] Zhen Niu, Zelong Yin, Xiangyu Kong. Sentiment Classification for Microblog by Machine Learning. 2012 Fourth International Conference on Computational and Information Sciences.
[4] Alexander Pak, Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. M.tech Thesis, LREC, 2010.
[5] Munmun De Choudhury. You're Happy, I'm Happy: Diffusion of Mood Expression on Twitter. Proceedings of HCI KOREA 2015.
[6] Christos Troussas, Maria Virvou. Sentiment analysis of Facebook statuses using Naive Bayes classifier for language Learning. Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on 10-12 July 2013 Page(s):1 – 6, IEEE.
[7] Alvaro Ortigosa, Jose M. Martin, Rosa M. Carro. Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior 31 (2014) 527–54.
[8] Asli Celikyilmaz1. Dilek Hakkani-Junlan. Probabilistic Model-Based Sentiment Analysis of Twitter Messages. 2010 IEEE.
[9] Dr. Sachin A. Kadam, Mrs. Shweta T. Joglekar. Sentiment Analysis: An Overview. IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 4, Aug-Sept, 2013.
[10] Kateryna Rybina. Sentiment analysis of contexts around query terms in documents. M.Tech thesis, 2012.
[11] Alexander Pak, Patrick Paroubek. Twitter for Sentiment Analysis: When Language Resources Are Not Available. 2011 22nd International Workshop on Database and Expert Systems Applications.
[12] Mikael Brevik Øyvind Selmer.Classification and Visualization of Twitter Sentiment Data. M.Tech thesis, NTNU Open.
[13] Raisa Varghese Jayasree. Aspect Based Sentiment Analysis using Support Vector Machine Classifier. 2013 IEEE.
[14] Vadim Kagan, V.S. Subrahmanian, and Andrew Stevens, Sentimetrix. Using Twitter Sentiment to Forecast the 2013 Pakistani Election and the 2014 Indian Election. 2015 IEEE Ieee Intelligent Systems Published by the IEEE Computer Society.
[15] Raluca-Sonia Chiorean, Mihaela Dinşoreanu, Daciana-Ioana Faloba. Sentiment Polarity Identification using Machine Learning Techniques. 2013 IEEE.
[16] Hasnat Ahmed, Muhammad Asif Razzaq, Ali Mustafa Qamar. Prediction of Popular Tweets Using Similarity Learning. 2013 IEEE .
[17] Yulan He , Deyu Zhou. Self-training from labeled features for sentiment analysis. Information Processing and Management 47 (2011) 606–616.
[18] Adnan Duric, Fei Song. Feature selection for sentiment analysis based on content and syntax models. Decision Support Systems 53 (2012) 704-711.
[19] Erik Boiy; Pieter Hens; Koen Deschacht; Marie-Francine Moens. Automatic Sentiment Analysis in On-line Text. Proceedings ELPUB2007 Conference on Electronic Publishing – Vienna, Austria – June 2007.
[20] Isaac G. Councill, Ryan McDonald, and Leonid Velikovich. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In Proceedings of the workshop on negation and speculation in natural language processing, pages 51–59, 2010.
[21] Jussi Karlgren, Gunnar Eriksson, Magnus Sahlgren, and Oscar Täckström. Between bags and trees–constructional patterns in text used for attitude identification. In Advances in Information Retrieval, pages 38–49. Springer, 2010.
[22] Richard Johansson and Alessandro Moschitti. Relational features in fine grained opinion analysis. Computational Linguistics, (Early Access): 1–37, 2012.