A Survey: Sentiment Analysis Using Machine Learning Techniques for Social Media Analytics

N. Mahendran,
Assistant Professor, Dept of ECE
M. Kumarasamy College of Engineering,
Karur, India
mahe.sec@gmail.com

T. Mekala
Assistant Professor, Dept of CSE
M. Kumarasamy College of Engineering,
Karur, India
mekalathangavel@gmail.com

Abstract—The recent research play a vital role in sentiment analysis due to people shares their own positive and negative thoughts on social media platform. The field of sentiment analysis that analyzes feedback, customer opinion, sentiment evaluation, emotion from written language and attitudes. It is a decision making process to review the product how many are negative and positive. The main process in this analysis to recognize the convention of repudiation and the categorization of negative and positive sentiments received by the users or client in the social group. This research work presents a survey concerning the presently available research technique, problem and application that materialize in the area of sentiment analysis.

Keywords--- Sentiment Analysis, Feedback, Decision Making, Customer Opinion

I. INTRODUCTION

Recently people focus on social media platforms to share opinions and their own thoughts to day to day life, business, education, celebrity etc. People shared their both positive and negative opinions on social media platform. By using the sentiment analysis the client can know the feedback about the product before making a purchase. This analysis process a type of natural language processing for identifying the feel of the people about a particular topics or products.

Sentiment analysis named as sentiment classification, sentiment mining, review mining, subjectivity analysis, opinion mining or appraisal extraction and classification of polarity deals with the computational treatment of opinion and subjectivity in text. It intends to determine the decision of a speaker with regard to a confident topic. The position could reflect evaluation, judgment or decision, sentimental state or the intended disturbing communication. Furthermore, it should be an example that in this situation 'subjective' does not mean that something is not true. Following problems faced in sentiment analysis of social media [1]:

- Named Entity Recognition (NER) method - This process used to removing entities such as organization, locations and people from social media.
- Anaphora Resolution Process - This process used to deciding the problem of a noun phrase or pronoun.
- Parsing - This process used to identifying the subject and object of the sentence.
- Sarcasm - This process used to find the actually stand for verb.

Fig 1: Architecture of sentiment Analysis
II. THE PROCESS OF SENTIMENT ANALYSIS

A. Data Pre-processing

This method used to remove the noisy, incomplete and inconsistent data. Data pre-processed must be performing before any mining functionality. The following tasks are performed in pre-processing [3].

B. Eliminating URLs

In normal URLs not donate to examine the emotion in the informal text. For example consider the line “I logged in to www.Ecstasy.com website as I’m bored” the above line negative but because of the occurrence of the word delight it may become impartial and it’s a false guess. In order to eliminate this sort of failure message we operate a technique to eliminate URLs.

C. Questions

The question words are not contribute to polarity like where, what, who, which, how etc., such words are removed to reduce the complexity.

D. Removing Special Characters

This type of Special characters like, &(),[]{}\" detached in order to reduce inconsistencies through the task of polarity. For example “it’s good:” this type of special characters not detached sometimes the special typescript merge with the words and make individuals words occupied in the vocabulary, to conquer this difficulty we eliminate special characters.

E. Removal of Retweets

The process of retweeting doubling another user’s tweet and redistribution to other friends. This regularly occurs if a user likes another user’s tweet. Retweets usually shortened with “RT.” For example, we think about the Following tweet: Amazing! RT @rupertgrint Harry Place a major role in Film History http://bit.illy./Euxsi :).

III. DOCUMENT SENTIMENT CLASSIFICATION (DSC)

DSC approach consider the entire document as a single entity and classify the sentiment as positive (+) or negative (-) or neutral. This categorization identify that the prejudiced document communicates estimations on a individual objective and the estimations belong to a indivial produce or person or service. It is obvious that this supposition true for client reviews of yield credentials which frequently heart on one product and individual reviewer writes it. Supervised and unsupervised method of classification methods used in document-level sentiment classification method.

<table>
<thead>
<tr>
<th>Task No</th>
<th>Task</th>
<th>Document Sentiment Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task1</td>
<td>Entity extraction and categorization</td>
<td>Unique Entity</td>
</tr>
<tr>
<td>Task2</td>
<td>Aspect extraction and categorization</td>
<td>Cluster[8-10]</td>
</tr>
<tr>
<td>Task3</td>
<td>Opinion holder extraction and categorization</td>
<td>Extract opinion from structured data</td>
</tr>
<tr>
<td>Task4</td>
<td>Time extraction and standardization</td>
<td>Extract time from opinion</td>
</tr>
<tr>
<td>Task5</td>
<td>Aspect sentiment classification</td>
<td>Identify whether an opinion positive, negative or neutral</td>
</tr>
<tr>
<td>Task6</td>
<td>Opinion quintuple generation</td>
<td>Document format[11-12]</td>
</tr>
</tbody>
</table>

IV. Supervised Learning Techniques

This technique can be used to sentiment classification, such as
- Naive Bays Classification (NB)
- Maximum Entropy Rule (ME)
- Support Vector Machine (SVM)

A. Naive Bays algorithm

The statistical Bayesian algorithm characterizes a supervised learning technique, it supposes an originating probabilistic method and it permits us to confine improbability about the method in a honorable way by formative chances of the results. It solve predictive and diagnostic and problems. It was introduced by Bayes. So this process named after Thomas Bayes. This classification method provides prior knowledge, observed data and practical learning algorithms combined. It used to evaluate learning algorithm and provides a useful perspective for understanding. It computes explicit probabilities for hypothesis and it is robust to noise in raw data [2].
Algorithm 1: Naive Bayes algorithm

Input: M denotes message M={M_1,M_2...M_N}
Output: P denotes positive message P={P_11,P_12...}
P denotes negative message N={N_21,N_22...}

Begin

M={M_11,M_12...M_1N}  N=Naive Table

Do

Divide M into Words (W)

M_i={W_1,W_2.............W_n}
i=1,2,3,4....,n

While (W_i € N) then

Return positive and negative polarities;
Find word = log(positive polarity) - log(negative polarity);
End while;

Calculate total polarity message;

End

B. Maximum Entropy algorithm

ME models named as Gibbs, log linear, multinomial logic models, exponential and present a general purpose machine learning technique for prediction and classification which has been productively practical to fields as different as econometrics and computer vision. In general language processing, recently it utilized for sentence frontier discovery, parse selection, fraction of speech tagging, ambiguity declaration, and attribute-value grammars, to name presently a few applications. The main pros of this advance is their flexibility, it permit stochastic rule systems to be increased with additional semantic, pragmatic and syntactic features [5]. However, the arrangements function depends on cost. Most of the models need considerable processing possessions and very great amounts of understands training data set in order to exactly judgment these parameters. While parameter representation for this model theoretically straightforward, in execute ME models for feature natural language tasks usually pretty large, and regularly embrace more datasets of free parameters. Assessment of such more models not only expensive, but also, due to carefully scattered skin, vulnerable to round-off errors. Thus, precise, highly competent, scalable techniques required for opinion the parameters of sensible copies. This survey, we preserve a numeral of algorithms for estimating the characteristics of ME models, counting GIS-Generalized Iterative balance and MIS-Maximum Iterative Scaling, as well as universal function optimization methods such as conjugate gradient, ascent, and variable metric schemes. Astonishingly, the frequently used iterative sizing algorithms achieve quite faulty, and most of the test case troubles, a lowest memory limitations metric method outperformed the other selections [6].

Algorithm 2: Maximum Entropy classifier

Input: M denotes message M={M_1,M_2...M_N}
Output: P denotes positive message P={P_11,P_12...}
P denotes negative message N={N_21,N_22...}
Nu denotes Neutral messages Nu={Nu_31,Nu_32......}

Begin

M={M_11,M_12...M_1N}  Nt=Naive Table

Do

Divide M into Words (W)

M_i={W_1,W_2.............W_n}
i=1,2,3,4....,n

While (W_i € Nt) then

Return positive and negative polarities;
Find word = log(positive polarity) - log(negative polarity);
End while;

Calculate total polarity message;

End

C. Support Vector Machine

It defined as input and output format; vector space named as participation and positive or negative is an output. The word count increase, if increase the document size. Practically considerably more features created which can augment classification accuracy. SVM task to learn and simplify the input/output mapping process. In this text classification input is collection of documents and output process is their corresponding class. For example spam filter as an input and output is spam or no spam [7].

It is a collection of supervised learning process which has been used for regression classification and outlier’s detection. There are list of benefits for using SVM method such as: i) Effective in high dimensional space, ii) Used a support vectors function and memory efficient, iii) versatile

Table 2. Tag code and Name

<table>
<thead>
<tr>
<th>Tag Code</th>
<th>Tag Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Positive</td>
</tr>
<tr>
<td>-</td>
<td>Negative</td>
</tr>
<tr>
<td>&amp;</td>
<td>Both</td>
</tr>
<tr>
<td>/</td>
<td>Neutral</td>
</tr>
<tr>
<td>/</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Algorithm 3: Support Vector Machine

Input : M denotes message M={M1,M2...M1N} 
Output : P1 denotes positive message 
P={P11,P12,...} 
N2 denotes negative message 
N={N11,N12,...} 
Nu denotes Neutral messages 
Nu={Nu11,Nu22,...} 

Begin 
M={M11,M12...M1N} 
Nt=Naive Table 
do
Divide M into Words(W)
Mi={W1,W2,.........,Wn} 
i=1,2,3,4........n 
While(Wi ∈ Nt)then 
Return positive and negative polarities; 
End while; 
calculate word count; 
End

IV. PERFORMANCE EVALUATION

In this research paper we evaluate the supervised learning method to find solution of sentiment classification problem. We used SVM, ME, Entropy learning techniques to divide sentiment of product review documents. We test many features to find Bigrams, Unigrams, position of words and Parts of Speech (POS) used as features in these techniques. The fig 2 show that the best performance achieved in SVM classifier.

V. CONCLUSION

In this survey presents several machine learning techniques such as Support Vector Machine, Maximum Entropy classifier and Naive Bayes can achieve high accuracy for identifying sentiment when using this approach. Compare SVM with remaining algorithm it achieve high accuracy because it give effective in high dimensional space, versatile and also it used a support vectors function and memory efficient.

REFERENCES


Fig 2: Summary of Sentiment Analysis