

Sentimental and Latent Analysis of Twitter based Product Reviews

R. Sathishkumar, T.M. Rino Simon, D. Abishake, M. Ajith Kumar, and Max Sherwin

Department of Computer Science Engineering,
Panimalar Engineering College, Chennai,
Tamilnadu, India

Copyright © 2017 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: Social Media Sites has evolved into an authoritative communication and information sharing tool used by billions of people around the Worldwide to post what is happening now in World. Social media platform permit post the different opinion many of the users. One of the Social media Twitter has turn out to be an important medium for peer interaction. The Existing system used the Sentiment analysis for the Product reviews and Different Classification algorithms for analysis the data in Social media. The Sentiment analysis has gained much notice in recent years for Product reviews in the Market. In this paper, proposed system is Latent analysis and Re tweet based Product reviews generally latent Attributes are Sex, age, regional basis, and political from Twitter user language .Latent attributes mainly helpful for advertising, personalization, and recommendation another Proposed Methodoly Re tweet based Product reviews. Re tweet reviews and Along with No of followers Count, tweet length, Hashtag Mention and Tweet Count are helpful to reviews the Product. The center of attention on expected retweet count of a tweet of an image linked. Data used in this study are online product reviews collected from Twitter.

KEYWORDS: Sentimental analysis, Latent attributes, Classification Algorithms, Retweet count, Hashtags.

1 INTRODUCTION

Sentiment analysis or opinion mining is an approach, thinking, or decision prompted by feeling. Sentiment analysis is additionally known as opinion mining,[1] understands the people's sentiments towards bound entities. Internet may be an ingenious place with relevance feeling information. From a user's aspect, individuals are able to post their own opinion and content through social media, like twitter, face book and blogs, or on-line social networking sites. From a researcher's perspective, several social media sites unharness their application programming interfaces prompting information assortment and analysis by researchers and developers.

Twitter presently has three totally different versions of arthropod genus offered specifically the rest API, the Search API, and also the Streaming API. With the remainder API, developers square measure able to gather standing information and user information; the Search API permits developers to question specific Twitter content, whereas the Streaming API is in a position to gather Twitter content in period of time. Moreover, developers will combine that API s to make their own applications [1]. Hence, sentiment analysis appears having a powerful fundament with the support of large on-line information. However, those sorts of on-line information have many flaws that probably hinder the method of sentiment analysis.

The primary flaw is that since individuals will freely post their own content, the standard of their opinions can't be warranted. For instance, rather than sharing topic-related opinions, on-line spammers post spam on forums. [1]Some spam are hollow in any respect, whereas others have unsuitable opinions conjointly referred to as pretend opinions The second flaw is that ground truth of such on-line information isn't continually offered. A ground truth is a lot of sort of a tag of an explicit opinion, indicating whether or not the opinion is positive, negative, or neutral.

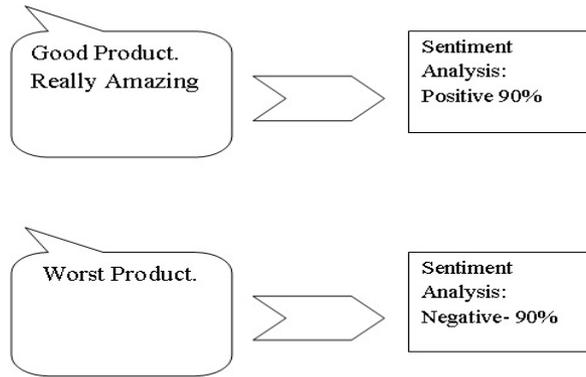


Figure 1 .Sentiment Analysis

Here, Haven on Demand has extracted a positive sentiment about the Good Products but a negative sentiment about Worst. In this case, it uses weights. Hate is obviously stronger than like and so it is weighted at 91%, while like has a weight of 72%.The Stanford Sentiment one hundred forty Tweet Corpus is one in all the datasets that has ground truth and is additionally public offered. The corpus contains one.6 million machine-tagged Twitter messages. Every message is labeled supported the emoticons discovered within the message. The Opinion Mining API is text to return the

Sentiment as negative, neutral and Positive. It contains a vocabulary of positive and negative words of different manner, and explains the patterns that describe how to combine these words to form positive and negative phrases. Sentiment analysis to valuable for the Product Analysis. The Sentimental Analysis helpful to review the Product.

```

{
  "positive": [
    {
      "sentiment": "Good",
      "topic": "Products",
      "score": 90,
      "original_text": "Good Product",
      "original_length": 16,
      "normalized_text": "Good Product",
      "normalized_length": 16
    }
  ],
  "Negative": [],
  "Aggregate": {
    "Sentiment": "positive",
    "Score": 90
  }
}
{
  "positive": [
    {
      "sentiment": "really like",
      "topic": "Products",
      "score": 0.8510865864332573,
      "original_text": "best product",
      "original_length": 11,
      "normalized_text": "Best Product",
      "normalized_length": 11
    }
  ]
}

```

Figure 2. Sentimental analysis Query

API Parameters:

Table 1. API Parameters

| Name | Type | Description |
|-----------|---------------|---|
| File | Array<binary> | File having the document to process. Multi POST |
| Reference | Array<String> | OnDemand reference obtained from either the Expand Container or Store Object API. passed to the API |
| Text | Array<String> | The text contains the Process |
| url | Array<String> | HTTP URL from which the document can be retrieved. |

2 EXISTING SYSTEM

a) Trending topic based classification:

Twitter is associate degree hugely widespread Social networking web site, Users will Post social info like breaking news, personal opinion regarding celebrities, and trending topics. Twitter account holder will post short messages referred to as tweets, short messages restricted by one hundred forty characters long and viewed by user's followers. Anyone World Health Organization chooses to possess other's tweets denote on one's timeline is named an acquaintance. Twitter has been used as a medium for period info dissemination and it's been utilized in numerous whole campaigns, elections, and as a print media. [6] The recognition of its use has been dramatically increasing. As of June 2011, regarding two hundred million tweets are being generated daily. once a brand new topic becomes widespread on Twitter, it's listed as a trending topic, which can take the shape of short phrases (e.g., #IPHONE7) or hashtags (e.g., #Best Product). What the Trend2 provides a frequently updated list of trending topics from Twitter. it's terribly fascinating to understand what topics are trending and what folks in different elements of the globe have an interest in. However, a really high share of trending topics are hashtags, a reputation of a personal, or words in different languages and it's typically tough to know what the trending topics are regarding. It's thus vital to classify these topics into general classes for easier understanding of topics and higher info retrieval. Classes: style, books, business, charity & deals, fashion, food & drink, health, holidays & dates, humor, music, politics, religion, science, sports, technology, TV & movies, other news, and other. Our goal is to assist users finding out info on Twitter to appear at solely smaller set of trending topics by classifying topics into general categories To classify trending topics into these predefined categories,[6] we have a tendency to propose 2 approaches: the well-known Bag-of-Words text classification, and victimization social network info.

b) Text based classification:

Text categorization is to make systems that are able to mechanically classify documents into classes. to make text classification systems, the bag of words illustration is that the most frequently used feature area. Its quality comes from its wide use within the field of knowledge retrieval and from the simplicity of its implementation. Yet, as within the bag of words illustration every dimension corresponds to the quantity of occurrences of the words in a very document, the task of classifying text documents into classes is tough as a result of the dimensions of the feature area is incredibly high. In typical issues, it unremarkably exceeds tens of thousands of words. Another side that hampers this task is that the undeniable fact that the quantity of coaching documents is many orders of magnitude smaller than the dimensions of the feature area .With the quick development of the net Technology, the analysis on text categorization has acquire a brand new stage, All kinds of strategies have consecutively got developed, including machine learning technique, has changing into the leading modality of text categorization, for instance Naïve theorem Classification, KNN , Support Vector Machine , Neural Network and Boosting then on. Some of these algorithms has already been realized and utilized in sensible systems, and created smart effects. Although Naive Thomas Bayes and therefore the k nearest neighbor's classifier are multi-class classifiers, the SVM are by default binary classifiers. Then, to handle multi-class issues, SVM sometimes depends on a 1 versus all strategy wherever as several binary classifiers as there are categories are trained. For example, within the case of a classification downside with n-classes, n one versus the remainder binary classifiers are trained. Therefore, once running experiments on advanced classification tasks involving over two-classes[15].

Literally scrutiny n SVM classifiers (for n classes) to single multi-class naive Thomas Bayes or k nearest neighbor's classifier. Additionally, to look at and compare the behaviors of the classifiers once experimental conditions are varied, these

conditions should be controlled exactly. Indeed, the properties of the coaching set will influence mostly the training talents of the classifiers, whereas in multi-class issues, it will be tough to know the actual influence of every category on the classifier's behaviors. Of the classifiers, whereas in multi-class issues, it will be tough to know the actual influence of every category on the classifier's behaviors. Therefore, in our state of affairs, we have a tendency to specialize in issues with solely two-classes. First, it allows United States of America to discard the influence of the multiclass aggregating rule within the case of SVM and therefore, to compare SVM a lot of fairly with naive Thomas Bayes and therefore the risk to regulate a lot of fastidiously the properties of the coaching set. In this regarded, so as to grant to each categories identical likelihood to be learned also, we have a tendency to solely studied things wherever the quantity of coaching instances is that the same in every category. Last, as binary issues are smaller.

c) Machine learning based classification:

Machine learning acts as a building block of various different fields of technology like computing, data processing, knowledge science, info retrieval and pattern recognition. Machine learning explores the analysis and development of algorithms that learn from the prevailing knowledge and draw inferences victimization that it can create predictions on the longer term knowledge properly. Instead of writing easy static codes, a model is developed victimization the datasets called coaching datasets. The event of the coaching datasets permits the computers to adopt information driven approach.

Datasets ends up in a lot of correct predictions. The first operate of machine learning rests on classification. Classification is that the method of Categorizing the info or a brand new observation into a category from a bunch of categories. The category is finished victimization the pattern determined within the coaching datasets whose class membership is already noted. Associate degree example would be distribution a given email into "spam" or "non-spam" categories or distribution a designation to a given patient as delineate by determined characteristics of the patient (gender, force per unit area, presence or absence of sure symptoms, etc.). Is one among the most important applications of Machine learning here, we have a tendency to be attending to explore 2 of the classification algorithms, Naïve Thomas Bayes and Support Vector Machines. So as to check algorithms we've got taken two issues. Within the initial example we have a tendency to be taking a theoretical state of affairs wherever our self driving automotive needs to decide whether or not to travel quickly or slow. So as to try and do that, we've got taken 2 factors that have an effect on the speed of the care- roughness and grade. The quick and slow are the categories and therefore the roughness and grade are the options. Victimization the coaching datasets the algorithms can predict whether or not the automotive can go quick or slow relying upon the worth of the grade and roughness. The Naïve Thomas Bayes can adopt a probabilistic approach whereas the SVM uses a non probabilistic approach.

d) Entity based classification:

Twitter could be a widespread micro-blogging service on the online, wherever folks will enter short messages that then come into view to other users of the service. The topics of those messages varies, there are lots of messages wherever the users specific their opinions regarding some firms or their product. These messages are an upscale supply of knowledge for firms for sentiment analysis or opinion mining[9] There's but an excellent obstacle for analyzing the messages directly: because the company names are typically ambiguous .

One desire initial to spot, that messages are associated with the corporate. During this paper we have a tendency to address this question. we have a tendency to gift numerous techniques for classifying tweet messages containing a given keyword, whether or not they are associated with a specific company thereupon name or not. That create use of company profiles that we have a tendency to create semi-automatically from external internet sources. Our advanced techniques take ambiguity estimations into consideration and conjointly mechanically extend the corporate profiles from the twitter stream itself [9] We have a tendency to demonstrate the effectiveness of our strategies through an in depth set of experiments. Moreover, we have a tendency to extensively analyze the sources of errors within the classification. The analysis not solely brings additional improvement, however conjointly allows using the human input a lot of expeditiously.

3 PROPOSED SYSTEM

Sentimental Analysis for Product Review:

1. Data Collection
2. Sentimental sentence Identification
3. Negotiation phase or Sentence Phase.

Data Collection:

Data used in these experiments is a collection of product reviews from Social Media sites like Twitter. From March, 2017, we collected, in total, millions of product reviews in which the products belong to the major categories: Mobiles, beauties, books, cars, appliances Etc. each review is rated in charts.

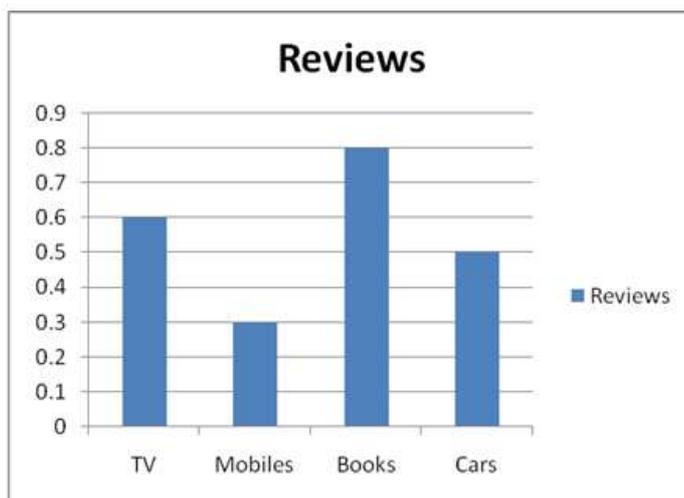


Fig 3. Data collection Reviews

Sentence phase identification:

The objective content is removed for sentiment analysis and all the subjective content was extracted for future analysis. The subjective content consists of all sentiment sentences. A sentiment verdict is the one that contains, at least, one positive or negative statement. The entire positive and negative were firstly separated English words. Every English word of a sentence has its role that explains how the word is used.[7] The syntactic roles are also known as the parts of speech. There are 8 parts of speech in English: verb, noun, pronoun, adjective, adverb, the preposition, Conjunction and the interjection.

Negotiation Phase:

- 1: **for** every hash tag Sentences **do**
- 2: **for** $i/i + 1$ as hash tag pair **do**
- 3: **if** $i + 1$ is a Negative Prefix **then**
- 4: **if** there is an adjective tag or a verb tag in next pair **then**
- 5: Negotiation of adjective Phrases $\leftarrow (i, i + 2)$
- 6: Negotiation of verb Phrases $\leftarrow (i, i + 2)$
- 7: **else**
- 8: **if** there is an adjective tag or a verb tag in the pair after next **then**
- 9: Negotiation of adjective Phrases $\leftarrow (i, i + 2, i + 4)$
- 10: Negotiation of verb Phrases $\leftarrow (i, i + 2, i + 4)$
- 11: **end if**
- 12: **end if**
- 13: **end if**
- 14: **end for**
- 15: **end for**
- 16: **return** Negotiation of adjective Phrases and Negotiation of verb Phrases.

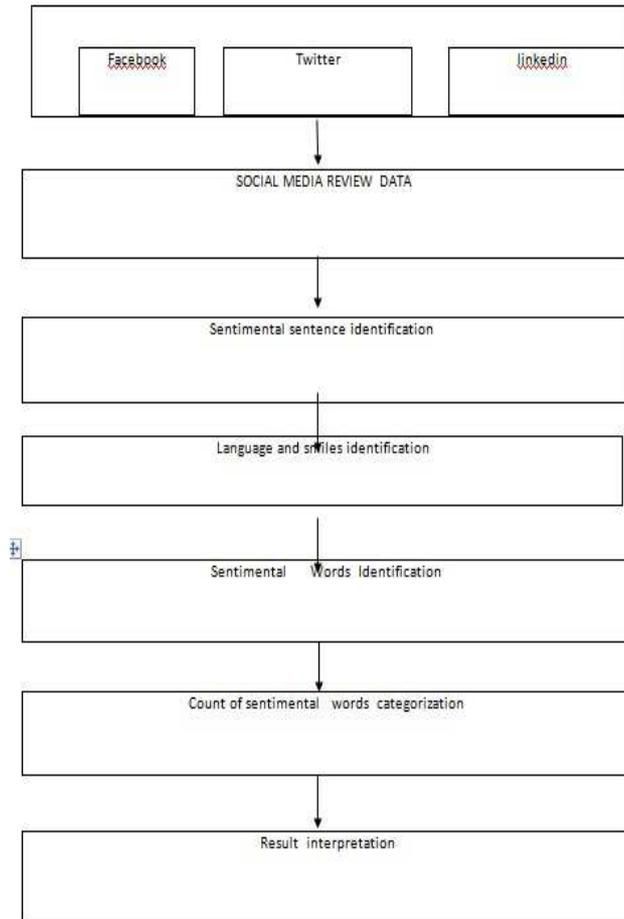


Fig 4. Flow diagram for Sentimental analysis

B) Latent Attributes:

Latent attributes are gender, age, and regional origin and political orientation. Latent attributes based Product review is Our Methodology. [3] Age is a real valued attribute, given the universal lack of ground truth for user age, we used the Twitter4J Java API.Users into two major demographic pools, and we take users below 30 and users above 30. [5]This binary categorization is nevertheless useful for user modeling and understanding. Gender helps to finds the User is male or female.

Table 2.Used Tweets

| Attributes | #No of Users per Class | No of Tweets |
|------------|------------------------|--------------|
| Gender | 500 | 6855000 |
| AGE | 1000 | 865545554 |
| Political | 200 | 5555485 |
| Religion | 100 | 4221155 |

Hash tag based Product review:

Hash tags are mostly keywords that are used to tag the tweets so users and Followers easily categorized and found by users and Current Hot news. For example Hash-tags are mentioned the hot news example #wesavejallikattu. [4] Given these use cases of hash tags, one would expect that interesting tweets are more likely to contain hash tags, the plot of below figure shows that tweets containing 1 to 2 hash tags are more expected to be retweeted than tweets without hash tags.

Table 3. Tweet Hash tags

| CHARATERISTIC | Value |
|--------------------------|-----------|
| Total Tweets | 6,240,996 |
| Cleared tweets | 6,232,587 |
| No hashtag tweets | 5,187,804 |
| At least one hashtag | 1,042,592 |
| One hashtag/tweet | 724,653 |
| Two hashtag/tweet | 132,965 |
| More than two hashtag | 69,900 |
| Total Hashtags in tweets | 24 |

Table 4. Popular Hash tags last Month

| No | Hashtag |
|----|------------------|
| 1 | #IPHONE7 |
| 2 | #SamsungA8 |
| 3 | #BESTPRODUCTASUS |
| 4 | #smartindia |
| 5 | #bbc |
| 6 | #teamfollowback |
| 7 | #news |
| 8 | #Northkorea |
| 9 | #love |
| 10 | #Neduvasal |

4 CONCLUSION

In this study, We Investigated Survey on Sentimental and Social Media based Product review. We use user databases of retweets for link prediction in the popular social media like Twitter. The sentimental Analysis or Opinion Mining is helps to study the People attitudes, Sentiments or human emotions certain entities, we take online based product review help of Social Medias like twitter. We hope that our survey is broad view and Discussions about sentimental analysis. While our research has established promising results on recommend modified hash tags, the scope of the research can be extended in several other directions in the future. We discuss the most prominent.

ACKNOWLEDGEMENT

This paper is based on research conducted by emerging technology of Information Retrieval .I am grateful for a number of friends and colleagues in encourages me to start a work, I thanks to Panimalar Engineering college for academic support. Finally I would like to thanks for all kept me going process.

REFERENCES

- [1] Sentiment analysis using product review data Xing Fang and Justin Zhan Otsuka *et al. Compu Social Networks (2016)* 3:3DOI 10.1186/s40649-016-0028-9
- [2] Retweets as a Predictor of Relationships among Users on Social Media Sho Tsugawa^{1*}, Kosuke Kit 2016
- [3] T. Bocklet, A. Maier, and E. N oth. Age determination of children in preschool and primary school age with gmm-based supervectors and support vector machines/regression. In TSD '08: Proceedings of the 11th international conference on Text, Speech and Dialogue, pages 253{260, Berlin, Heidelberg, .Springer-Verlag.
- [4] Tommasel, A.; and Godoy, D.; 2015. A novel metric for assessing user influence based on user behaviour. In: SocInf, pp. 15–21.
- [5] Ludu, P.S. 2015. Inferring Latent Attributes of an Indian Twitter User using Celebrities and Class Influencers, In Proceedings of the 1st ACM Workshop on Social Media World Sensors (SIdEWays'15). ACM, NewYork, NY,USA, 9-15.2016

- [6] Jean-Val'ere Cossu, Vincent Labatut, Nicolas Dugu'e. A Review of Features for the Discrimination of Twitter Users: Application to the Prediction of Offline Influence.2015
- [7] Schreiner T. New compete study: primary mobile users on twitter. 2011. <https://www.blog.twitter.com/2013/new-compete-study-primary-mobile-users-on-twitter>.
- [8] Heggstuen J. One in every 5 people in the world own a smartphone, one in every 17 own a tablet. 2013.
- [9] Entity-based Classification of Twitter Messages Surender Reddy Yerva_ International Journal of Computer Science and Applications c Technomathematics Research Foundation Vol. 9 No. 2, pp. 88 - 115, 2012
- [10] C.Nagarajan and M.Madheswaran – 'Analysis and Implementation of LLC-T Series Parallel Resonant Converter with Fuzzy controller'- International Journal of Engineering Science and Technology (IJEST), Applied Power Electronics and Intelligent Motion Control. Vol.2 (10), pp 35-43, December 2010
- [11] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.
- [12] C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.
- [13] MACHINE LEARNING COMPARISON OF CLASSIFICATION ALGORITHMS International Journal of Advanced Engineering and Global Technology / Vol-03, Issue-11, December 2015\
- [14] Y. S. Yegin Genc et.al, "Discovering context: Classifying tweets through a semantic transform based on wikipedia," in *Proceedings of HCI International*, 2011.
- [15] S. Kinsella, A. Passant,et.al "Topic classification in social media using metadata from hyperlinked objects," in *Proceedings of the 33rd European conference on Advances in information retrieval*, 2011, pp. 201–206.
- [16] J. Sankaranarayanan, H. Samet, et.al, "Twitterstand: news in tweets," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*,2009, pp. 42–51.
- [18] M. N. Hila Becker et.al "Beyond trending topics:Real-world event identification on twitter," in *Proceedings of AAAI*, 2011.
- [19] Go, R. Bhayani, et.al, "Twitter sentiment classification using distant supervision," 2009.
- [20] Pang, L. Lee, et.al "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. Association for Computational Linguistics, 2002, pp. 79–86. Weka 3: Data Mining Software in Java, <http://www.cs.waikato.ac.nz/ml/weka/>.