International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Analysis of User's Behavioral Pattern Using Sentiment Analysis: A Survey

Farah Jamal[1], Kavita Agrawal[2]

Department of CSE, Integral University, Lucknow, India

Associate Professor, Department of CSE, Integral University, Lucknow, India

Abstract: ***People express themselves by giving suggestions, opinions, feedback or ideas about the product. People use "social media" platforms such as twitter, facebook, blogs etcto express their views. Sentiment analysis becomes more pivotal due to availability of large amount of information on social media .Sentiments from social media such as facebook or twitter provide most up-to-minute and comprehensive information. The study of user's behavior helps firms and organization to improve their marketing strategy. Users opinion that are present on social networking sites will help business analyst/specialist to provide better option to the users and thus motivating the users to buy the products. The aim is to determine the expressed reviews of customer using supervised learning and deep learning algorithm. This will help in understanding the various issues of customer like how he feels, think and select between different alternatives.***

Keywords: Sentiment Analysis, Hybrid, Deep Learning, Naïve Bayes, Decision Tree

1. Introduction

In order to compete with today's market; the business analyst (organization) needs to analyze user's behavior regarding their product. It is very necessary to analyze the user opinion or interest for different brands.So sentiment analysis is a process to evaluate user behavior, in order to determine the user interest regarding any particular brand or topic. It is used to analyze the polarity of the users response i.e. favorable(positive), non-favorable(negative) or neutral. It is a field of Natural Language Processing and is very critical. It also helps in determining the emotions behind the series of words or sentences.

"Sentiment analysis is a mechanized process of wrenching out of knowledge from the views expressed by the people on social media platforms". In my work, we will be using sentiment analysis for mining of the text and then extracting important data that will help the organization to understand the opinion of the customers.This will not only help the organizations to see what people think of their product but also the opinions of users for their competitors. The ability to quickly understand consumer's attitude and react accordingly helps in increasing the productivity of the firm.

An application of sentiment analysis is very vast. Since it helps us to gain an overview of wider public opinion behind certain topics or products, therefore it is very helpful in social media monitoring.

The fundamental steps for performing sentiment analysis comprises of data collection, pre-processing, feature extraction, sentiment detection and classification using simple computation or machine learning approaches. We can have three class sentiment classification i.e. positive, negative or neutral.

2. Sentiment Analysis Scope

Document Level: Full document is analyzed. At this level, opinion of the 'overall document' is determined to get the sentiment.

Sentence Level:- Individual sentence is analyzed. At this level, opinion of 'single sentence' is determined to get the sentiment.

Sub-sentence level:-Sub-phrase within a sentence is analyzed. At this level, opinion of 'sub-phrase' in a sentence is determined.

3. Sentiment Analysis Classification

1) **Machine learning:-** It is an "application of AI". It gives capability to system to imbibe things automatically and allows software applications to become more accurate without being explicitly programmed. They can identify the texts which are expressing sentiments.
   a) ***Supervised Learning Techniques***:- Support Vector Machines(SVM), Bayesian Network, Naïve Bayes Classification, Decision Tree are extensively used techniques.
   b) ***Unsupervised Learning Techniques***:-Clustering Algorithm, Matrix factorizations are unsupervised learning techniques.
2) **Deep Learning Techniques:** It is a machine learning approach that fit computer to do things instinctively as human can do i.e. grasp by examples. In deep learning, computer models are prepared to implement classification function directly from the content. The asset of deep learning is that their progress improves as the size of data expand. We will be applying Long-Short term memory (LSTM).It is a supervised deep neural network. LSTM is an extension of RNN. RNN has 'limited or short-range' memory and in consolidation with LSTM they have 'persisting or long-range' memory.

LSTM enables RNN to remember input for long period of time.

3) **Lexicon Based:** It is a simple and practical approach. It is a collection of known and precompiled sentiments. There are two methods in this approach:
   a) ***Dictionary Based:*** -A set of words is collected manually.
   b) ***Corpus Based:*** -These are further divided into semantic and statistical data. The tweets that we will collect will come under semantic data and the keywords that we will create according to product will come under statistical data.

4) **Hybrid:-** It is the combination of both machine learning and lexicon based approach. Accuracy and precision of data can be improved by combining both.

We will be using Naïve Bayes classifier for classification, decision tree for precision as it gives graphical description of possible solutions to a decision based on certain condition and Long-Short Term Memory (LSTM) for predicting the outcome.

4. Literature Survey

The following papers are surveyed in this section:-

(Shamal, 2018)[1] in their paper titled "Sentiment Analysis using Token2Vec and LSTMs"proposed a Long Short-Term Memory based deep learning approach to perform sentiment analysis of user reviews on social media on electronics domain. The paper has also introduced the concept of Token2Vec and has reached an average accuracy of 88.2%. Thus it proves that LSTM-RNN with Token2Vec is much better that existing techniques.

(Wint, 2018)[2] in their paper titled "Deep Learning based Sentiment Classification in Social Network Services Datasets" used hybrid architecture of Convolutional Neural Network(CNN) and Bidirectional LSTM. It combines the strength of both CNNs and BLSTM. CNN is used for extracting different positive or negative word from SNS data whereas BLSTM is used to produce sentence level representation. Two kinds of SNS data is used in this paper: product review data and social network sites data. They concluded that they got best performance by using their H2CBi models with Word2Vec for product review dataset and H2CBi model with Word2Vec and fastText for social network site datasets.

(B.Thejaswini, 2018)[3] Suresh in their paper titled "Study of user's behavior in Structured E-Commerce Websites" proposed the use of temporal logic and model techniques as an alternative to data mining techniques. They proposed a methodology for using it in a structured e-commerce websites. The main emphasis is given on the user's behavior using web server log file prediction. Web server logs are pre-processed to extract the detailed traces. The business analyst can use a set of (predefined) temporal logic patterns to formulate queries that could help them to discover and understand the way clients use the website. The queries can check the existence of complex causality relationships between events, based on the website structure and contents of user's actions.

(Yue, 2018) [4] in their paper titled "A Survey of sentiment analysis in social media" focuses on presenting typical method from different perspective for categorizing sentiment analysis i.e. task-oriented, granularity oriented and methodology-oriented. They explored the different types of data tools that can be used in research analysis and suggested their strength and limitations. They presented an overview of multimodal sentiment analysis(MSA).They established a common terminology that helps people from different background to easily understand and laid a foundation for research in sentiment analysis.

(L, 2018)[5]in their paper titled "Sentiment Analysis of Customer Reviews on Laptop Products for Flipkart"used ROCK and CART algorithm for the classification and analysis of product reviews for flipkart website. This paper has analyzed the fundamentals of opinion mining. The approach consists of Extraction, Clustering and Classification. The Laptop details are extracted by using flip kart product API. Using product API, we can fetch the brand name, reviews, rating and other related things for product. Positive and negative words are classified from the reviews and then it calculates percentage of positive and negative words.

(Zainuddin, 2017)[6] in their paper titled "Hybrid sentiment classification on twitter aspect-based sentiment analysis" proposed a hybrid approach for sentiment classification. They present a principal component analysis (PCA), latent semantic analysis (LSA) and random projection (RP) in their paper. The implementation of this hybrid approach improved the accuracy performance from the existing sentiment classification method.

(Lang, 2017)[7] in their paper titled "Understanding Consumer Behavior with Recurrent Neural Network" proposed an approach to apply RNNs to predict the consumer behavior in E-Commerce website. They proposed that applying RNN is better than existing vector-based approach. RNN helps in linking individual actions directly to predict in an intuitive way. In this paper the RNN is applied on the real data of large European online fashion platform and advantage of RNN over other approach. They concluded that applying RNN will provide improved empirical performance, better prediction explanations and reduced feature engineering.

(Sharma, 2017)[8] in their paper titled "Opinion Mining and Sentiment Analysis on Customer Review Documents- A Survey" [8] proposed that opinion mining and sentiment analysis is an emerging field of data mining technique used to extract the knowledge from a huge volume of customer comments, feedback and reviews on any product or topic etc. They used three techniques:-Supervised Learning techniques, unsupervised learning techniques and case-based reasoning as opinion mining techniques. These techniques can be used in opinion spam detection, purchasing product or services, quality improvement in product or service, marketing research, policy making and decision making.

(Ribeiro, 2016)[9] in their article "SentiBench- a benchmark comparison of state-of-the-practice sentiment analysis methods" aims at presenting a comparison between twenty-four sentiment analysis methods. The evaluation is based on eighteen labeled data sets that includes movie and product reviews, messages on social networks, and opinions in news articles. This paper suggests that sentiment analysis cannot be used as 'mass-produced' methods. It suggests that companies and researchers must perform experiments with different methods before applying a method because same social media text can be interpreted differently based on sentiment method choice.

(Mehta, 2016)[10] Jain in their paper titled "Sentiment Mining and Related Classifiers" covers all the fundamental details of opinion analysis. This paper discuss the existing work on sentiment analysis and opinion mining of customer assessments and reviews online. This paper also explains the workflow of opinion mining process, recent trends and information regarding sentiment analysis. From paper we can conclude that based on data accumulation different algorithms execute differently. Some of the algorithms perform very well and some of the algorithms do not perform up to the mark depending on user's requirement. None of the algorithm can be considered superior over other.

(M.Jose, 2016)[11] in their paper titled "A semantic graph based approach on interest extraction from user generated texts in social media" puts forward an idea to utilize the social networking or micro-blog profiles for generating an interest graph that could be used as a knowledge base for these systems where user interests are required. A design for recommender system that could work along with this interest extraction method was also proposed and implemented. This system could perform faster by using data dumps of DBpedia instead of using its services online. The proposed method proved to be effective in developing interest profiles of Twitter users by using the texts they post as tweets. The resultant graph of the system could be used by a number of online services which require knowing the tastes and interests of users. This paper proposes a semantic graph based method to identify the likes and interests of users by analyzing their twitter feeds. Unlike the other conventional methods there is no need to track the user activity on the Internet or conduct exclusive surveys and ratings to collect explicit ideas from the user

(Batrinca, 2014)[12] in their paper titled "Social media analytics: a survey of techniques, tools and platforms" have written this paper for researchers seeking to analyze the wealth of social media. It presents an extensive review of software tools for social networking media, blogs, newsgroups, chat and news feeds. Due to the availability of web-based application programming interfaces (APIs) provided by Twitter and Facebook analyzing social media feeds for sentiment analysis, has become a major research and business activity. This paper surveys some of the social media software tools, and for completeness introduced social media scraping, data cleaning and sentiment analysis.

(Song, 2013)[13] in their paper titled "Multiple Categorization of products: Cognitive modeling of customers through social media data mining" analyzed the user opinion in online space to psychologically know the user preference towards the product. It analyzes the complex preferences regarding diverse product and brands and provides a new model for inter-market connection. This study focused on connections between individual information by constructing network of user opinions .They provide more scientific approach to the structures of online market from management information systems, human-computer interactions and digital marketing.

(Lima, 2012)[14] in their paper titled "Automatic Sentiment Analysis of Twitter Messages" proposed a sentiment analysis system with automatic training based on tweets containing either emoticons or sentiment based words. These sets were used to categorize the tweets that could not be classified automatically. The technique chosen to classify the unlabeled tweets was the naïve Bayes algorithm. Three different approaches (emoticon-based, word-based and hybrid) with different criteria for automatic classification were assessed. The results suggested that the combination of techniques provide improved accuracy.

5. Comparative Study

Title of Paper	*Study of Paper*	*Pros*	*Cons*	*Result*
Sentiment Analysis using Token2Vec and LSTMs[1] (Shamal, 2018)	They proposed the use of LSTM and Token2Vec for analysis of user reviews on social media	The model is able to analyze the reviews which consists of emojis and social acronyms	The model cannot detect and analyze sarcasm in user comments.	The paper has reached an average accuracy of 88.2%
Deep Learning based Sentiment Classification in Social Network Services Datasets[2] (Wint, 2018)	Use of hybrid architecture of Convolutional Neural Network (CNN) and Bidirectional LSTM. Two kinds of SNS data is used in this paper: product review data and social network sites data.	Combines the strength of both CNNs and BLSTM.	Only small and balanced datasets were used. Filter region size was same in both CNN layers	They concluded that they got best performance by using their H2CBi models with Word2Vec for product review dataset and H2CBi model with Word2Vec and fastText for social network site datasets.
Study of user's behavior in Structured E-Commerce Websites[3] (B.Thejaswini, 2018)	Proposed the use of temporal logic and model techniques as an alternative to data mining techniques.	Pros:- It provides casual relation among events of user trace. The proposed system can deploy different parallel servers with different parts of the log and can	Cons:- The analyst has to use pre-defined logic patterns to formulate query.	The business analyst can use a set of (predefined) temporal logic patterns to formulate queries that could help him to discover and understand the way clients usethe website.

		execute the queries in parallel.		
A Survey of Sentiment Analysis in Social Media[4] (Yue, 2018)	This paper focuses on presenting typical method for categorizing sentiment analysis i.e. task-oriented, granularity oriented and methodology-oriented.	Pros:- Established a common terminology that help people from various background knowledge to easily understand and laid a foundation for research in sentiment analysis.	Cons:-Multi-modal system is recommended which is less used in present scenario.	This survey categorized and classified sentiment analysis researches from different perspective as well as explored different types of data and tools that can be used in research analysis and suggested their strength and limitations.
Sentiment Analysis of Customer Reviews on Laptop Products for Flipkart[5] (L, 2018)	Used ROCK and CART algorithm for the classification and analysis of laptop product reviews for flipkart website	Pros:-This work in future can be used n taking reviews from multiple website.	Cons:-It provides less information on the relationship between the predictors and response.	This work classifies positive and negative words from reviews and calculates the percentage of positive and negative words to know which product have more percentage of positive review
Hybrid sentiment classification on twitter aspect-based sentiment analysis[6] (Zainuddin, 2017)	A new hybrid approach was proposed to perform finer grained sentiment analysis	Pros:-This research focus on aspect based sentiment analysis on twitter in order to perform finer grained analysis.	Cons:-Analysis was done only on tweets.	It improves the performance accuracy from existing sentiment classification method by 76.55, 71.62 and 74.24%.
Understanding consumer behavior with Recurrent Neural Network[7] (Lang, 2017)	To predict consumer behavior by applying RNNs approach in e-commerce	Reduced feature engineering, improved empirical performance and better prediction explanations.	Consumers future behavior is only predicted. Individual tastes and level of product is not predicted.	By using RNN empirical performance is improved, reduced feature engineering and better prediction explanations.
Opinion Mining and Sentiment Analysis on Customer Review Documents- A Survey[8] (Sharma, 2017)	To survey opinion mining with different levels of architecture, tools, techniques, and comparative study of different techniques and challenges.	Improvement of existing sentiment word identification algorithm and developing fully automatic tools for analysis.	There were various challenges while doing analysis which needs to be sorted out.	They founded that more innovative and effective techniques should be invented in order to overcome the present challenges faced by Sentiment Analysis and Opinion mining.
Sentiment-a benchmark comparison of state-of-the-practice sentiment analysis methods [9] (Ribeiro, 2016)	Thorough comparison of twenty-four sentence-level sentiment analysis method using standard data sets that span different types of data sources.	The methods code and datasets used in this article are open.	After research they were not able to conclude that which sentiment analysis method is better.	It suggests that researchers and companies must perform experiment with different methods before applying any method.
Sentiment Mining and Related Classifiers: A Review[10] (Mehta, 2016)	It covers all the details of opinion analysis and comprises of current and forthcoming scope of sentiment analysis	Explained opinion mining process, applications and recent trends of sentiment analysis.	They were unable find that which algorithm is superior.	This paper provides a comprehensive understanding of various opinion mining classifiers.
A semantic graph based approach on interest extraction from user generated texts in social media[11] (M.Jose, 2016)	To utilize the social networking or micro-blog profiles for generating an interest graph that could be used as a knowledge base for these systems where user interests are required	The use of semantic relations among the topics helped in highlighting the core areas of interest. This system can perform better and faster by using data dumps of DBpedia.	Sometimes suggestions become annoying and are useless because of poor knowledge base.	This method is purely based on the texts that a user leaves in a particular social network website or a micro blog. It also put forward the design for a recommender system that can work along with the proposed interest extraction method.
Social media analytics: a survey of techniques, tools and platforms.[12] (Batrinca, 2014)	This paper surveys some of the social media software tools, and for completeness introduced social media scraping, data cleaning and sentiment analysis.	Provides an overview for scientists seeking to utilize social media scraping and analytics in their business or research.	Techniques presented may not be valid till now.	It presents a comprehensive review of software tools for social networking media, wikis, blogs, newsgroups, and news feeds.
Multiple Categorizations of products: Cognitive modeling of customers through social media modeling[13] (Song, 2013)	To analyze user opinion in online space to learn preferences toward various product psychologically categorized by user.	They provided a head start for uncovering the construction of boundaries of the market in the cognition model of actors.	Analysis was concentrated on the model of whole network rather than econometric analysis.	This study focused on connectedness between individual information units by constructing network of user opinions.
Automatic Sentiment Analysis of Twitter Messages[14] (Lima, 2012)	They proposed a sentiment analysis system with automatic training based on tweets containing either emoticons or sentiment based	The combination of techniques provided improved accuracy.	Only for tweets that contains emoticons or sentiment based word.	Three different approaches (emoticon-based, word-based and hybrid) with different criteria for automatic classification were assessed. The

	words.			results suggested that the combination of techniques provide improved accuracy.

6. Conclusion

Sentiment analysis is emerging as a challenging field with diverse application. Different approaches have been introduced till now. This paper surveys the different approaches of categorizing and analyzing sentiments of users on social media. From our review, we concluded that different algorithms execute differently on the basis of data they collected. In this paper we have examined different researches with their pros and cons. The use of LSTM with other techniques has helped in getting better performance yet with different cons. Like use of LSTM with Token2Vec cannot detect and analyze sarcasm. So till now, none of the algorithm is exceptionally superior over other in all context.

References

[1] Achira Jeewaka Shamal, Rankothge Gishan Hiranya Pemathilake, Sachith Paramie Karunathilake and Gamage Upeksha Ganegoda. (2018), "Sentiment Analysis using Token2Vec and LSTMs" in International Conference on Advances in ICT for Emerging Regions.

[2] Zar Zar Wint, Yuki Manabe and Masayoshi Aritsugi. (2018), "Deep Learning based Sentiment Classification in Social Network Services Datasets" in IEEE/ACIS 3rd International Conference on Big Data, Cloud Computing, Data Science & Engineering, pp 91-96

[3] B.Thejaswini, Reddi Durga Sree and Karamala Suresh. (2018), "Study of user's behavior in structured E-commerce Websites" in International Journal of Scientific Research & Engineering Trends, vol-4.

[4] Lin Yue, Weitong Chen, Xue and Wanli Zuo. (2018), "A survey of sentiment analysis in social media" in Springer-Verlag London Ltd. ,pp 1-47

[5] Janhavi N L, Santhosh Kumar K L and Jharna Majumdar. (2018), "Sentiment Analysis of Customer Reviews on Laptop Products for Flipkart" in International Research Journal of Engineering and Technology (IRJET),vol-5.

[6] Nurulhuda Zainuddin, Ali Selamat, Roliana Ibrahim. (2017), "Hybrid sentiment classification on twitter aspect-based sentiment analysis" in Springer Science+Business Media,LLC,part of Springer Nature.

[7] Tobias Lang and Matthias Rettenmeier. (2017), "Understanding consumer behavior with RNN" in Int. Workshop on Machine Learning for Recommender Systems.

[8] Surya Prakash Sharma, Dr Rajdev Tiwari and Dr Rajesh Prasad. (2017), "Opinion Mining and Sentiment Analysis on Customer Review Documents- A Survey" in International Journal of Advanced Research in Computer and Communication Engineering,vol-6.

[9] Filipe N Ribeiro, Matheus Araujo, Pollyanna Goncalves, Marcos Andre Goncalves and Fabricio Benevenuto. (2016), "Sentiment-A benchmark comparison of state-of-the practice sentiment analysis methods" in EPJ Data Science a Springer Open Journal.

[10] Rehee Mehta and Dr. Shaily Jain. (2016), "Sentiment mining and related classifiers: A Review" in IOSR Journal of Computer Engineering, vol-18, pp 50-54.

[11] Lijo M. Jose and Rahamathulla K.(2016), "A semantic graph based approach on interest extraction from user generated texts in social media" in IEEE International Conference on Data Mining and Advanced Computing.

[12] Bogdan Batrinca, Philip C and Treleaven. (2014), "Social media analytics: a survey of techniques, tools and platforms" in Springer, pp 89-116.

[13] Gil-Young Song, Youngjoon Cheon, Kihwang Lee, Heuiseok Lim, Kyung-Yong Chung and Hae-Chang Rim.(2013), "Multiple Categorization of products: Cognitive Modeling of customers through social media data mining" in Springer Verlag London, pp 1387-1403.

[14] Ana C.E.S. Lima and Leandro N.de Castro. (2012), "Automatic Sentiment Analysis of Twitter Messages" in IEEE Fourth International Conference on Computational Aspects of Social Networks (CASoN).