



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2018IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 15th Nov 2018. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-07&issue=ISSUE-12](http://www.ijiemr.org/downloads.php?vol=Volume-07&issue=ISSUE-12)

Title: **A SURVEY ON SENTIMENT ANALYSIS**

Volume 07, Issue 12, Pages: 289–296.

Paper Authors

V.SIVA PARVATHI, YJN LAKSHMI

¹PVP Siddhartha Institute of Technology, Vijayawada

²PB Siddhartha College of arts and science, vijayawada



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

A SURVEY ON SENTIMENT ANALYSIS

¹V.SIVA PARVATHI, ²YJN LAKSHMI

¹Asst. Professor Dept of CSE, PVP Siddhartha Institute of Technology, Vijayawada

²Lecturer, Department of Computer Science, PB Siddhartha College of arts and science, vijayawada.

¹vallabhaneni.vsp@gmail.com, ²boppana.yl@gmail.com

Abstract Sentiment analysis or Opinion mining is one of the quickest developing fields with its call for and potential advantages growing every day. With the onset of the internet and modern technology, there has been a vigorous growth in the quantity of statistics. each character is capable of specific his/her personal ideas freely on social media. All of this facts may be analyzed and used that allows you to draw benefits and high-quality statistics. One such idea is sentiment analysis, here, the sentiment of the problem is taken into consideration and important facts is drawn out whether it be a product evaluation or his/her opinion on whatever materialistic. a few of such packages of sentiment evaluation and the method in which they're carried out are defined. moreover, the possibility of every of those works to impact any destiny work is considered and explained along with the analysis as to how the previous troubles in the equal area have been overcome.

Keywords: Sentiment Analysis, Opinion Mining, Product Review, Drug Related Classification, Forecasting Price.

INTRODUCTION

in this paper, an in depth analysis of numerous works associated with sentiment analysis is taken. The works in various fields inclusive of housing, stock trade, product assessment, drug response analysis and others are taken into consideration and explained. Sentiment evaluation has its packages in numerous fields. wherever there is a records attached to a subject there is always a sentiment. Sentiment evaluation is commonly taken into consideration to have either a wonderful, impartial or a negative end result. The emotional evaluation is occasionally merged with sentiment evaluation in which case multiple tuples come into use. Sentiment evaluation

basically involves classification, but the content material that actually consists of the sentiment or opinion need to be first recognized. The maximum essential problem in sentiment analysis is sentiment polarity categorization. this is tackled in every of the aforementioned works in a completely unique way so that it will improve the efficiency of evaluation in every of the cases. First, an green drug event extraction is enabled and its possibility via the use of sentiment analysis on statistics sets from Twitter is analysed. 2nd, the impact of the unfairness in traders' sentiment on the volatility of shares is explained. also, the manner wherein an

individual's opinion on social media influences the price of a inventory and the methods to forecast such shocks are analysed. moreover, based on the public sentiments, potential livable places are analysed and for this reason calibrated. Of the works mentioned, the analysis of the sentiment of a patron on a product through reviews has been the maximum approached place in sentiment analysis due to its capacity in both growing accuracy and enterprise advantages.

SENTIMENT ANALYSIS USING PRODUCT REVIEW DATA

Xing Fang and Justin Zhan

on this paper, Xing Fang and Justin Zhan try to talk about the most essential trouble in sentiment analysis, the sentiment polarity categorization, by means of thinking about a dataset containing over 5.1 million product opinions from Amazon.com with the goods belonging to four classes: beauty, books, electronics and home. The preceding papers in this location suggested to put off all the objective content material if you want to conduct sentiment evaluation however here, the subjective content is as a substitute extracted for destiny evaluation. The inputs taken are evaluations which contain the consumer info, overview, helpfulness and score. The score is considered because the floor fact for greater accurate analysis of the evaluation's sentiment. A max-entropy POS tagger is used which will classify the words of the sentence into forty six tags. an extra python program is specifically used to speed up this procedure. As a result, a total of 25 million adjectives, 22 million adverbs and 56 million verbs are

recognized, which normally generally tend to decide the sentiment. The negation words like no, now not, and greater are protected within the adverbs whereas Negation of Adjective and Negation of Verb are mainly used to discover the terms. 21,586 phrases are identified with a complete of zero.68 million. The set of rules also makes a listing of terms primarily based on prevalence. the subsequent are the numerous classification models which can be decided on for categorization: Naïve Bayesian, Random forest, and help Vector device. although this paper tackles the hassle of sentiment polarity categorization it still faces a couple of demanding situations and has its boundaries. One such being the curse of dimensionality in function vector formation which limits the range of dimensions and additionally forces to have the same quantity of dimensions. The overall performance of this technique is estimated by way of considering the average F1 rating. consequently future work might be benefited if these obstacles considered and thereby the accuracy and overall performance may be progressed.

EXPLORING PUBLIC PLACES FOR LIVEABLE PLACES BASED ON A CROWD-CALIBRATED SENTIMENT ANALYSIS MECHANISM

Linlin You and Bige Tuncer

With the measure of exchanges expanding step by step, the blockchain ends up cumbersome. Every hub needs to store all exchanges to approve them on the blockchain in light of the fact that they need to check if the wellspring of the present exchange is unspent or not. Additionally, because of the first confinement of square size and the time interim used to produce

another square, the Bitcoin blockchain can just process almost 7 exchanges for each second, which can't satisfy the prerequisite of handling a large number of exchanges continuously form. In the interim, as the limit of squares is little, numerous little exchanges may be postponed since mineworkers incline toward those exchanges with high exchange charge. There are various endeavors proposed to address the versatility issue of blockchain, which could be ordered into two sorts:

hub to work full duplicate of record, Bruce proposed a novel digital currency plot, in which the old exchange records are expelled (or overlooked) by the system [37].

A database named account tree is utilized to hold the parity of all non-void locations. Other than lightweight customer could likewise help settle this issue. A novel schem named VerSum [38] was proposed to give another way enabling lightweight customers to exist. VerSum enables lightweight customers to outsource costly calculations over vast data sources. It guarantees the calculation result is right through looking at results from numerous servers.

- Redesigning blockchain. In [39], Bitcoin NG (Next Generation) was proposed. The principle thought of Bitcoin-NG is to decouple regular square into two sections: key square

for pioneer race and microblock to store exchanges. The convention isolates time into epoches. In every age, mineworkers need to hash to produce a key square. Once the key square is created, the hub turns into the pioneer who is in charge of producing microblocks. Bitcoin-NG additionally

broadened the heaviest (longest) chain methodology in which microblocks convey no weight. Along these lines, blockchain is overhauled and the tradeoff between square size and system security has been tended to.

B. Privacy Leakage

Blockchain can protect a specific measure of security through the general population key and private key. Clients execute with their private key and open key with no genuine personality introduction. In any case, it is appeared in [40], [5] that blockchain can't ensure the value-based protection since the estimations of all exchanges and equalizations for every open key are freely noticeable. Moreover, the ongoing examination [41] has demonstrated that a client's Bitcoin exchanges can be connected to uncover client's data. In addition, Biryukov et al. [11] exhibited a strategy to connect client aliases IP addresses notwithstanding when clients are behind Network Address Translation (NAT) or firewalls. In [11], each customer can be particularly recognized by an arrangement of hubs it interfaces with. Notwithstanding, this set can be learned and used to discover the birthplace of an exchange. Various techniques have been proposed to enhance secrecy of blockchain, which could be generally ordered into two kinds:

- Mixing [42]. In blockchain, clients addresses are pseudonymous. Be that as it may, it is as yet conceivable to connect delivers to client genuine way of life the same number of clients make exchanges with a similar location much of the time. Blending administration is a sort of administration which gives secrecy by

exchanging assets from various info delivers to different yield addresses. For precedent, client Alice with deliver A needs to send a few assets to Bob with address B. In the event that Alice straightforwardly makes an exchange with input address An and yield address B, connection among Alice and Bob may be uncovered. So Alice could send assets to a confided in delegate Carol. At that point Carol exchange assets to Bob with numerous information sources c1, c2, c3, and so on., and different yield d1, d2, B, d3, and so forth. Weave's location B is likewise contained in the yield addresses. So it winds up harder to uncover connection among Alice and Bob. In any case, the go-between could be unscrupulous and uncover Alice and Bob's private data intentionally. It is likewise conceivable that Carol exchanges Alice's assets to her own location rather than Bob's location. Mixcoin [43] gives a basic strategy to keep away from unscrupulous practices. The middle person encodes clients' necessities including reserves sum and exchange date with its private key. At that point if the middle person did not exchange the cash, anyone could confirm that the delegate swindled. In any case, robbery is recognized yet at the same time not counteracted. Coinjoin [44] relies upon a focal blending server to rearrange yield delivers to avoid robbery. Furthermore, propelled by Coinjoin, CoinShuffle [45] utilizes decoding mixnets for address rearranging.

- Anonymous. In Zerocoin [46], zero-information confirmation is utilized. Excavators don't need to approve an exchange with advanced mark however to approve coins have a place with a rundown

of legitimate coins. Installment's source are unlinked from exchanges to anticipate exchange diagram examinations. Be that as it may, regardless it uncovers installments' goal and sums. Zerocash [47] was proposed to address this issue. In Zerocash, zero-information Succinct Non-intuitive Arguments of Knowledge (zk-SNARKs) is utilized. Exchange sums and the estimations of coins held by clients are covered up.

C. Selfish Mining

Blockchain is vulnerable to assaults of conniving narrow minded diggers. Specifically, Eyal and Sirer [10] demonstrated that the system is helpless regardless of whether just a little bit of the hashing power is utilized to swindle. In childish mining system, narrow minded diggers keep their mined squares without broadcasting and the private branch would be uncovered to people in general just if a few necessities are fulfilled. As the private branch is longer than the present open chain, it would be conceded by all diggers. Before the private blockchain publishment, legitimate mineworkers are squandering their assets on a futile branch while narrow minded excavators are mining their private chain without contenders. So narrow minded mineworkers have a tendency to get more income. In view of egotistical mining, numerous different assaults have been proposed to demonstrate that blockchain isn't so anchor. In willful mining [48], excavators could open up its gain by non-inconsequentially forming mining assaults with organize level shroud assaults. The trail-determination is one of the tenacious system that excavators still mine the squares regardless of whether the private chain is

deserted. However now and again, it can result in 13% gains in examination with a non-trail-obstinate partner. [49] demonstrates that there are childish mining methodologies that gain more cash and are productive for littler diggers contrasted with basic narrow minded mining. In any case, the increases are moderately little. Moreover, it demonstrates that assailants with under 25% of the computational assets can at present gain from narrow minded mining. To help settle the narrow minded mining issue, Heilman [50] displayed a novel methodology for legitimate diggers to pick which branch to take after. With arbitrary guides and timestamps, fair mineworkers would choose all the more new squares. In any case, [50] is helpless against forgeable timestamps. ZeroBlock [51] expands on the basic plan: Each square should be created and acknowledged by the system inside a most extreme time interim. Inside ZeroBlock, narrow minded mineworkers can't accomplish more than its normal reward.

V. POSSIBLE FUTURE DIRECTIONS

Blockchain has demonstrated its potential in industry and the scholarly community. We talk about conceivable future bearings as for four regions: blockchain testing, stop the inclination to centralization, huge information investigation and blockchain application.

A. Blockchain testing

As of late various types of blockchains show up and more than 700 digital forms of money are recorded in [52] up to now. Nonetheless, a few engineers may distort their blockchain execution to pull in financial specialists driven by the immense

benefit. Other than that, when clients need to consolidate blockchain into business, they need to know which blockchain accommodates their necessities. So blockchain testing component should be set up to test distinctive blockchains. Blockchain testing could be isolated into two stages: institutionalization stage and testing stage. In institutionalization stage, all criteria must be made and concurred. At the point when a blockchain is conceived, it could be tried with the concurred criteria to legitimate if the blockchain works fine as engineers guarantee. Concerning testing stage, blockchain testing should be performed with various criteria. For instance, a client who is accountable for online retail business thinks about the throughput of the blockchain, so the examination needs to test the normal time from a client send an exchange to the exchange is pressed into the blockchain, limit with regards to a blockchain square and so forth.

B. Stop the tendency to centralization

Blockchain is planned as a decentralized framework. Notwithstanding, there is a pattern that diggers are unified in the mining pool. Up to now, the main 5 mining pools together claims bigger than 51% of the aggregate hash control in the Bitcoin arrange [53]. Aside from that, egotistical mining system [10] demonstrated that pools with more than 25% of aggregate registering force could get more income than decent amount. Sane diggers would be pulled in into the egotistical pool lastly the pool could without much of a stretch surpass 51% of the aggregate power. As the blockchain isn't planned to serve a couple of associations, a

few strategies ought to be proposed to take care of this issue.

C. Big data analytics

Blockchain could be all around joined with enormous information. Here we generally classified the blend into two sorts: information administration and information examination. Concerning information administration, blockchain could be utilized to store essential information as it is appropriated and secure. Blockchain could likewise guarantee the information is unique. For instance, if blockchain is utilized to store patients wellbeing data, the data couldn't be altered and it is difficult to stole those private data. With regards to information investigation, exchanges on blockchain could be utilized for enormous information examination. For instance, client exchanging examples may be removed. Clients can anticipate their potential accomplices' exchanging practices with the examination..

EFFICIENT ADVERSE DRUG EVENT EXTRACTION USING TWITTER SENTIMENT ANALYSIS

Yang Peng, Melody Moh, and Teng-Sheng Moh

In this paper, Yang Peng, Melody Moh, and Teng-Sheng Moh discuss how the advancements of social media are being helpful to extract large datasets by using a drug-related classification and sentiment analysis to extract ADEs on Twitter. ADEs are adverse drug events. Even though the pharmaceutical companies perform many drug-related tests beforehand, when a drug is released into the market some ADEs will be unidentified. Through the above-mentioned method, a data of four months on

Twitter is collected so, as to capture the maximum number of ADEs. A simple and efficient pipeline is proposed to retrieve data from Twitter. The process of the pipeline is, the tweets from twitter are captured firstly and then the data is pre-processed (cleaned data is the output of data pre-processing). The drug classification is done for the cleaned data and the user opinion data is collected from which the ADEs are extracted. The captured tweets are stored in HIVE. Tweets are in JSON file and can, therefore, be stored in HIVE directly. They used python NLP tool for capturing tweets and Data pre-processing. For storing datasets of drug-related classification and tweets of sentiment analysis WEKA is used. Thus after thorough research on different tweets pipelines are built and they are compared to newly designed ones to extract numerous ADEs. As a result an average of 5 times of total number of ADEs, among them 20% are new ADEs. The proposed system may further include streaming more tweets from Twitter by using Topsy API, using more drugs as keywords in the experiments, applying Apache Spark for processing a lot of tweets. The proposed method may be applied to other areas such as daily consumer products for recognizing side effects and user opinions on them.

INVESTOR CLASSIFICATION AND SENTIMENT ANALYSIS

Arijit Chatterjee and Dr. William Perrizo

In this paper, Arijit Chatterjee and Dr. William Perrizo discuss the effect investors' bias has on the volatility of stocks in the market, sentiment analysis was done on tweets of the potential investors and also why they used Microsoft Azure over other



sentiment analyser tools. Twitter is one of the largest social media platforms with over 280 million active users with almost 500 million tweets created every

day. Some investors use Twitter to share their opinion on some ticker symbols every day, this paper discusses how these opinions of the investors affect the stock market. Investors are assumed to be sentiment driven. A top-down approach is used to make sure a stock is not overrated or underrated by the investors. The approach is based on two broad behavioral finance assumptions - sentiment and limits to arbitrage. Sentiment analysis is done on the tweets pulled from some selected investors' twitter feeds. They assign positive, negative and neutral sentiment scores to the ticker symbols from the pulled tweets by identifying "bad", "not good", "great" words in the tweets. For processing, the unstructured text in the tweets "Microsoft Azure" analyser tool is used. Because Microsoft Azure gave better results when compared to other analysis tools, this is shown in the four graphs in the paper where Microsoft Azure is compared with Stanford NLP Sentiment Analysis engine and another popular commercial tool. Through sentiment analysis, a particular user can understand the social sentiment score of a ticker symbol based on the discussions of key investors and make an informed decision about which stock to invest.

FORECASTING PRICE SHOCKS WITH SOCIAL ATTENTION AND SENTIMENT ANALYSIS

Li Zhang, Liang Zhang, Keli Zhao and Qi Liu

In this paper, the data from the Chinese Stock Market – SZSE and SSE are considered along with the social media activities in Weibo.com in order to extend recent studies on financial activities in social media and their impact on the stock market.

What makes this work stand out is the way in which the previous limitations, such as, inability to tackle practical problems in finance, lack of proper knowledge regarding the direction of the price shock and, were overcome by using this implementation. This work makes use of DSA – Degree of Social Attention, which has been introduced by the previous works to capture stock price

shocks. The method involves identifying the price shock as negative, near-zero or positive. These price shocks are essentially the difference between the expected value and actual value. Prior to estimating these price shocks, the social media activities are analysed and the features such as account information, future tracking, and response such as, like, repost, comment are considered. Furthermore, these details are cross-referenced to the account holder's actual activity and its effect on trading. The efficiency of this work is proved by evaluating the F-measure without DSA and with DSA to show the improvement. The following are the classifiers used: Naïve Bayes, Decision Tree, Random Forest, Logistic and LibSVM. Out of these classifiers, the Random Forest has the best performance and SVM the worst. Also, once the price shocks are estimated it is identified that the negative shocks lead to better accuracy than the positive shocks.

CONCLUSION

In this work, the latest developments in sentiment analysis are reviewed and the future possibilities for each of these developments are presented. Sentiment analysis is a field which is catching up in the recent years and its applications are subject to increase to a broader range in near future. This work is an attempt to create a basis with the help of which future works can be improved and also take a note of the challenges this field offers. The effectiveness of various approaches has been evaluated and shown.

REFERENCES

1. Xing Fang and Justin Zhan: Sentiment analysis using product review data
2. Linlin You and Bige Tuncer : Exploring public places for livable places based on a crowd-calibrated sentiment analysis mechanism
3. Yang Peng, Melody Moh, and Teng-Sheng Moh : Efficient Adverse Drug Event extraction using Twitter sentiment analysis
4. Arjit Chatterjee and Dr. William Perrizo : Investor classification and sentiment analysis
5. Li Zhang, Liang Zhang, Keli Xhao and Qi Liu : Forecasting price shocks with social attention and sentiment analysis