JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Twitter Sentiment Analysis using Supervised Machine Learning Algorithms

Shruti Kulkarni[1], Anisaaara Nadaph[2] and Dr. Geetika Narang[3]
ME Student[1], Assistant Professor[2] and Professor[3]
Department of Computer Engineering
KJ's Trinity College of Engineering and Research, Pune INDIA
kulshruti.25@gmail.com[1], anisaaranadaph.tcoer@kjei.edu.in[2] and geetikanarang.tcoer@kjei.edu.in[3]

Abstract: Social networking has become a distinguished platform within the digital age. In sentiment analysis, categorizing tweets into polarity categories may be a common task. the foremost advanced approaches to the present drawback use supervised machine learning models that are learned victimization manually annotated examples. Twitter sentiment analysis permits businesses to trace public angle regarding their product and events in real time. The text pre-processing of Twitter information is that the initial stage in sentiment analysis. the bulk of surviving Twitter sentiment analysis focuses on extracting new sentiment parts. The goal of this analysis is to indicate the way to mix matter info from Twitter conversations with sentiment diffusion patterns to enhance sentiment analysis on Twitter information. thereto purpose, we tend to examine sentiment spreading by staring at a development referred to as sentiment reversal, and that we discover many intriguing properties of the sentiment reversals. Then taking into consideration the interactions between matter info in Twitter messages and the sentiment diffusion patterns, we tend to gift SentiDiff, associate degree unvarying system for predicting sentiment polarity explicit in Twitter tweets. To our data, this can be the primary study to use sentiment diffusion patterns to help within the improvement of Twitter sentiment/emotion analysis. intensive tests on real-world information show that, in comparison to progressive matter information based on sentiment analysis algorithms.

Index Terms **- Twitter, sentiment analysis, text pre-processing, Diffusion of sentiment, social networks, feature fusion, and graph analysis.**

I. INTRODUCTION

The impact of on-line social networks, as well as their ascent, has revolutionized the patterns of human interaction and data diffusion among the population in fashionable society in recent years. Sentiment Analysis in micro blogging has fully grown in quality over the previous decade. individuals communicate regarding their daily life's exploitation social media platforms like Twitter, wherever users' posts cover a large vary of topics. several analysis publications on Sentiment Analysis offer attention-grabbing approaches for classification algorithms, and the importance of pre-processing before and through the feature choice method is generally acknowledged. during this context, pre-processing refers to the method of improvement and making ready texts for classification. it's true that unstructured writings on the Internet, like those found on Twitter, include considerable levels of noise. we tend to describe knowledge that contains no significant data for the analysis at hand, i.e., Sentiment Analysis, as noise.
Twitter has adult to become one amongst the world's most well-liked social media sites. attributable to the immense quantity of knowledge on the market from Twitter, extracting people' feeling polarities sent in tweets has become outstanding analysis issue. many techniques are developed to supply political election plans by analyzing Twitter users' sentiment assessments on political parties and politicians, as an example. Twitter sentiment analysis is additionally utilized by businesses as a fast and effective approach to trace however folks feel concerning their merchandise and types. The goal of sentiment analysis on Twitter knowledge is to reason a Twitter message's sentiment polarity as positive, neutral, or negative. ancient text sentiment analysis ways are used on to perform Twitter sentiment analysis. Twitter messages, like news broadcasts and book items, square measure often transient and imprecise. because of the informal nature of Twitter postings, there square measure a lot of slangs, acronyms, misspelt words, and modal particles. As a result, once typical text sentiment analysis algorithms square measure accustomed forecast the sentiment polarities of Twitter posts, their accuracy suffers considerably. several distinctive sentiment analysis approaches for Twitter messages are developed to beat this challenge. totally supervised ways and distantly supervised ways is loosely classified into 2 sorts.

Pre-processing approaches will facilitate solve the matter of rackety labels in sentiment analysis. Recent analysis, however, has confirmed that no viable pre-processing approaches exist for all the datasets and algorithms. once a pre-processing strategy that works for one rule and one dataset is applied to a different dataset or rule, the performance of sentiment analysis might suffer. In general, each absolutely supervised and distantly supervised systems for the Twitter sentiment analysis largely specialize in matter info from Twitter messages, and thanks to the distinctive characteristics of Twitter messages, they're unable to get adequate results. The supervised ways attempt to develop sentiment classifiers exploitation knowledge that has been expressly classified and sentiment lexicons. one in all the key drawbacks of supervised approaches is that manually building the sentiment lexicons and labelling knowledge is long and labor intensive, and as a result, most methods' sentiment lexicons and tagged knowledge square measure of deficient to make sure glorious performance. what is more, supervised algorithms oft have confidence hand-loomed options and determining the way to produce effective options remains a troublesome task. From knowledge with rackety labels like emoticons and hash tags, the distantly supervised algorithms learn sentiment classifiers. several study communities have become a lot of inquisitive about sentiment diffusion, that is primarily involved with understanding however sentiments in social networks impact info diffusion. By clicking the retweet button inside a tweet, users on Twitter will republish another user's tweet and share it with their own followers. Users will add a comment to a tweet once reposting it, and the original tweet are going to be announce aboard it. Tweets and retweets will thereby offer info concerning their writers' sentiment extremes a couple of topics.

As a result, we will check at however sentiment polarity differs from a tweet to its retweets to check however sentiment dispersion works on Twitter. the most effective planning is to fuse matter info from the Twitter messages and sentiment diffusional info during a supervised learning architecture, thanks to the shortcomings of existing Twitter sentiment analysis solutions that solely contemplate matter info and the shut comparison between sentiment diffusion patterns and sentiment polarities of Twitter messages.

II. LITERATURE SURVEY

H. Li et al: during this paper, Quantitative analysis, language analysis, and open secret writing were used to analyse the utilization of social media by immigrant-focused non-profits. This study, which incorporates interviews with participants UN agency run such social media accounts, demonstrates however immigrant-focused NPOs use social media in an exceedingly hostile political atmosphere by using 3 vital strategies: 1) spreading info on immigration-related problems and policies; 2) encouraging individuals to affix collective efforts to impact the political atmosphere; and 3) partaking in dialogue with external stakeholders like politicians, the media, and different teams.

D. Paul et al: during this work, the investigator examines that younger generations area unit still heavily dependent on Twitter. As a result, it principally represents the voice of the younger generation. it'll be not possible to urge an entire image till Twitter users become well-known across generations. per exit polls, the previous generation and non-graduates favour Republican, and their share is statistically important. a lot of applied math tools and also the ability to increasingly update varied models among Compass area unit among the continuing and future comes we're performing on. to get the multilingual resolution on sentiment analysis, Compass' default sentiment model is substituted with another language's sentiment classification model, as an example, French sentiment classification model.

F. Bravo-Marquez et al: ASA may be a new approach that uses previous info from associate degree opinion lexicon15 to provide artificial coaching information for Twitter sentiment analysis from unlabelled corpora. employing a predefined polarity lexicon, the approach annotates tweets per the polarity of their words and generates balanced coaching information by sampling and averaging tweets containing terms of a similar polarity.

J. Zhao et al: during this paper, Twitter sentiment polarity classification is influenced by six totally different pre-processing approaches. On 5 Twitter datasets, we have a tendency to do a series of experiments with four classifiers to ascertain however effective varied pre-processing ways area unit. The removal of URLs, stop words, and numerals features a minor impact on classifier performance; in addition, exchange negation and increasing acronyms will increase classification accuracy. As a result, deleting stop words, numerals, and URLs reduces noise whereas having no impact on performance. In sentiment analysis, exchange negation is useful. For the Twitter sentiment classification downside, we elect the foremost relevant pre-processing approaches and have models for many classifiers.

S. Symeonidis et al: during this paper, an outsized range of pre-processing approaches that had ne'er been evaluated as compared analysis before, and place them to the take a look at in 2 datasets. On the premise of accuracy, every technique was tested in four representative machine learning algorithms. what is more, supported the results, we have a tendency to classified carrying out classes and tallied the amount of characteristics for every technique. Finally, associate degree ablation study was dispensed on all of the approaches, in addition because the superior ones, to envision however they interacted. we'll take a look at these techniques on datasets from several areas, like news articles and products or moving picture reviews, in future studies and add new options.

X. Zhang et al: during this work, the present topological measurements supported propagation path failed to justify the efforts of people to induce their info current suitably and fairly, so the researchers examined the special processes for info propagation on Twitter. they found that the bulk of users acted as middleware within the info dissemination method. to precise the link between their engagement and also the impact they received, the study instructed a mensuration mistreatment 3 measures: user engine, user excitement, and user length.

F. Shanghai dialect et al: during this paper, the researchers area unit operating to unravel the matter of extraordinarily unbalanced learning from the point of view of feature learning. they provide UCML, a singular technique. this is often the primary plan to introduce the conception of MFL into the context of unbalanced learning. They take a look at 5 severely unbalanced datasets from a range of application sectors. UCML beats progressive extremely unbalanced learning algorithms, in step with the findings. The results of the experiments show that 3 key parts of our strategy area unit effective. They conjointly discovered that our methodology is a lot of proof against giant imbalance ratios.

A. Joulin et al: during this paper, the investigator presents a basic baseline text classification algorithmic program. Our word options, in contrast to the unsupervised learned word vectors from word2vec, will be averaged along to get acceptable sentence representations. Fast Text achieves performance reminiscent of recently according deep learning-inspired approaches in numerous workloads whereas being considerably quicker. though deep neural networks have much better representational power than shallow models in theory, it's unclear whether or not basic text categorization problems like sentiment analysis area unit the most effective thanks to assess them. We'll create our code public so others will develop on prime of it.

I. Chaturvedi et al: during this paper, for multi-label classification of nodes in social networks, researchers instructed a replacement fuzzy convolutional deep walk. In terms of F-measure, our projected technique outperforms the baseline by up to fifteen. Deep convolutional neural networks area unit being thought of for learning communities of node embeddings effectively. in addition, we have a tendency to use a fuzzy feedback controller to capture the sequence of connected nodes within the network. Deep learning heuristically approximates the network's weights through contrastive divergence, leading to higher accuracy than baselines. the strategy is computationally economical and needs way less quality than customary ways. Finally, at every convolutional kernel within the deep model, we will show the communities as extremely activated nodes.

P. Badjatiya et al: during this work, the investigator looked into the utilization of deep neural network architectures for hate speech detection and discovered that they outperformed existing ways considerably. once deep neural network models' embeddings were incorporate with gradient boosted call trees, the most effective accuracy values were obtained. we have a tendency to will investigate the role of user network options within the task within the future.

III. Existing System

The majority of extant Twitter sentiment analysis focuses on extracting new sentiment components. The question of a way to opt for a pre-processing approach, on the opposite hand, is left unreciprocated. One disadvantage of those ways is that the high price of knowledge annotation. because of the distinctive properties of Twitter messages, the bulk of existing ways for Twitter sentiment analysis entirely judge matter data from tweets and square measure unable to produce satisfactory results. though recent analysis has incontestable that sentiment dissemination patterns square measure closely associated with the sentiment polarities of Twitter tweets, current techniques largely specialise in the matter content of Twitter messages, ignoring sentiment diffusion data.

IV. Proposed System

The goal of this analysis is to point out a way to mix matter data from Twitter conversations with sentiment diffusion patterns to enhance sentiment analysis on Twitter knowledge. to it purpose, we tend to examine sentiment spreading by gazing a development referred to as sentiment reversal, and that we discover many intriguing properties of sentiment reversals. Then, taking into consideration the interactions between matter data in Twitter messages and sentiment diffusion patterns, we tend to gift SentiDiff, associate unvaried system for predicting sentiment polarity declared in Twitter tweets. To our information, this is often the primary study to use sentiment diffusion patterns to assist within the improvement of Twitter sentiment analysis. the subsequent square measure the main contributions of this paper: We investigate sentiment reversal, the development during which a tweet and its retweet have distinct sentiment polarities, to explore sentiment diffusion on Twitter. The options of sentiment reversals square measure investigated, and a sentiment reversal prediction model is projected. We propose SentiDiff, associate unvaried methodology that considers the inter-relationships between matter data in Twitter messages and sentiment dispersion patterns to forecast the sentiment polarity of every message. To assess the performance of our instructed methodology, we tend to run variety of tests. The experimental results counsel that our projected SentiDiff methodology aids progressive matter information-based sentiment analysis algorithms.

Advantages: -

• It improves the performance of sentiment analysis on Twitter.
• It improves sentiment classification accuracy.

System Architecture

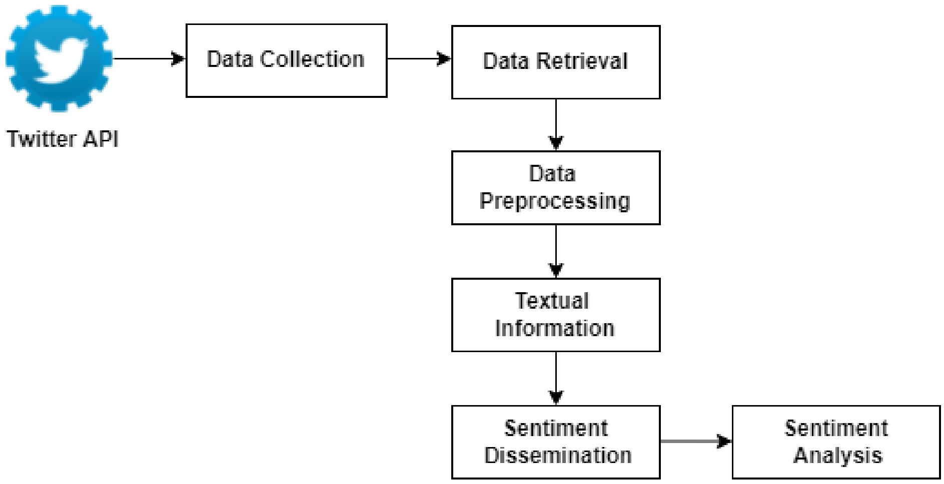

Fig. 1. Architecture System

Algorithm (Support Vector Machine):

To classify the twitter emotions, support vector machines (SVMs) are used. The SVM Support Vector Machine is basically a classifier with two types of class boundaries: linear and non-linear. SVMs work by creating a hyperplane between data that represents the class to which they belong. The main goal is to train the machine with the known data and then use that data to train the SVM to discover the ideal hyperplane that delivers the largest distance to the nearest training data point in any class. procedure:

Step 1: Read the testing tweet's text and the trained features.
Step 2 Examine all the tweet's test features, as well as all the pull features.
Step 3: Take a close look at the kernel.
Step 4: Use both features to train the SVM and examine the results.
Step 5: Using a trained SVM as classifier, classify the observations.

V. Result and Discussion

The total accuracy of current algorithms and support vector machine methods is shown in this section. Therefore, when compared to previous approaches, this works and produces better sentiment analysis findings. The following is the interpretation of the experimental results: TPe: True Positive (correctly predicted number of instances) FPe: False Positive (wrongly predicted number of instances), TNe: True Negative (correctly predicted number of instances when not needed) FNe False Negative (instances) accuracy = TPe + TNe ÷ TPe + FPe + TNe + FNe precision = TPe ÷ TPe + FPe reproduction based on this parameter. Rate = TPe ÷ TPe + FNe. F1 Measure = 2 × precision rate × recall rate ÷ precision rate + recall rate.

Table 1. Comparison

Sr. No.	**Existing System**	**Proposed System**
Algorithm	Naïve Bayes	Support Vector Machine
Precision	60.2%	65.4%
Recall	85.5%	89.7%
Accuracy	82.7%	88.9%

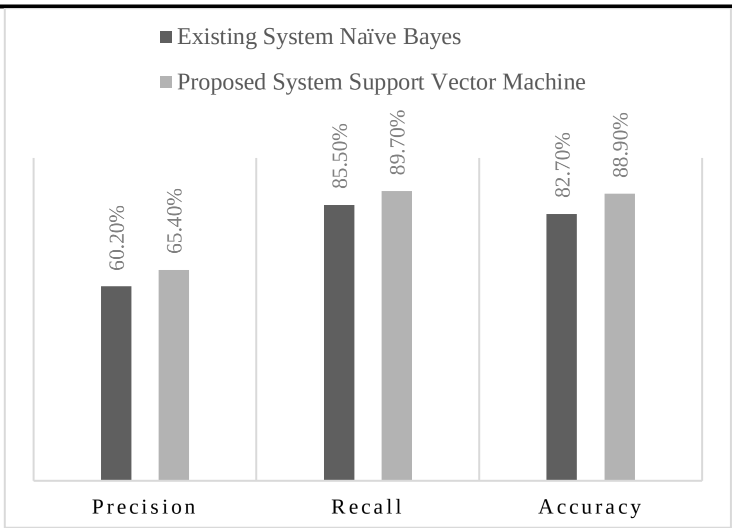

Fig. 2. Comparison Graph

VI. CONCLUSION

In this article, we'll look at ways to combine text information from Twitter tweets with sentiment propagation models to improve sentiment analysis performance for Twitter data. To this end, we study the propagation of mood by studying a phenomenon known as mood inversion and uncover some interesting properties of mood inversion. Next, we will develop SentiDiff, an iterative algorithm that predicts the polarity of emotions represented by Twitter tweets, considering the interaction between textual information in Twitter messages and the diffusion patterns of emotions. To our knowledge, this is the first time that an emotion diffusion model has been used to improve sentiment analysis on Twitter. Extensive testing on real-world datasets shows it to be superior to state-of-the-art text sentiment analysis algorithms. In the future, we will analyze the difference in emotional spread patterns between topics and con-sider topic information from Twitter tweets when combining text and emotion spread data.

References

[1] H. Li, L. Dombrowski, and E. Brady, "Working toward empowering a community: How immigrant-focused non-profit organizations use twitter during political conflicts," in Proceedings of the 2018 ACM Conference on Supporting Groupwork. ACM, 2018, pp. 335–346.

[2] D. Paul, F. Li, M. K. Teja, X. Yu, and R. Frost, "Compass: Spatio temporal sentiment analysis of us election what twitter says!" in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2017, pp. 1585–1594.

[3] F. Bravo-Marquez, E. Frank, and B. Pfahringer, "Annotate-sampleaverage (asa): A new distant supervision approach for twitter sentiment analysis," in 22nd European Conference on Artificial Intelligence (ECAI), vol. 285. IOS Press, 2016, pp. 498–506

[4] J. Zhao and X. Gui, "Comparison research on text pre-processing methods on twitter sentiment analysis," IEEE Access, vol. 5, pp. 2870–2879, 2017.

[5] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," Expert Systems with Applications, 2018.

[6] X. Zhang, D.-D. Han, R. Yang, and Z. Zhang, "Users participation and social influence during information spreading on twitter," PloS one, vol. 12, no. 9, p. e0183290, 2017.

[7] F. Wu, X.-Y. Jing, S. Shan, W. Zuo, and J.-Y. Yang, "Multiset feature learning for highly imbalanced data classification." in AAAI, 2017, pp. 1583–1589.

[8] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, vol. 2, 2017, pp. 427–431.

[9] I. Chaturvedi, S. Cavallari, E. Cambria, and V. Zheng, "Learning word vectors in deep walk using convolution," in Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference, 2017, pp. 323–328.

[10] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep learning for hate speech detection in tweets," in Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 2017, pp. 759– 760.