International e-Conference on New Horizons And Multidisciplinary Applications In Science And Technology
In Association withInternational Journal of Scientific Research in Science and Technology
Volume 9 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Review on Challenges of Sentiment Analysis

Mr. Ram B. Ghayalkar[1], Prof. Dr. D. N. Besekar[2]
[1]Assistant Professor, Department of Computer Science, Shri R. L. T. College of Science, Akola, Maharashtra, India
[2]Professor, Department of Computer Science, Shri Shivali College of Arts, Commerce & Science, Akola, Maharashtra, India

ABSTRACT

Now in this internet era everything is available in the form of websites, blogs, social networks, e-commerce etc. so there is a importance of reviews, opinions, feedbacks by users. These feedbacks generated by users plays important role for business, individuals, governments. Here comes the role of Sentiment Analysis on the basis of feedbacks given by users. But there are several challenges facing the sentiment analysis and its evaluation process.
In this paper presents different perspectives of challenges occurs while finding the accuracy and extract subjective information from text for sentiment analysis and defining its polarity.

Keywords: Sentiment Analysis,emotion analysis, social media, Sarcasm, Multilingual data, text mining, Machine Learning

I. INTRODUCTION

Nowadays with the increasinguse of the internet a lot of information is available on the web which is about the different products, movies, books,technologies etc. People express their views, opinions etc on the different products,services,books etc. on the web or social media.
The sentiment found within comments, feedbackor critiques provide useful indicators form anydifferent purposes. Sentiment Analysis is a taskunder natural language processing which findsorientation of a person opinion or feelings over anentitySentiment analysis is an ongoing research field. InSentiment analysis based on the sentiment value it is decidedwhether the sentence is positive, negative or neutral. Thishelps a lot when need to rely on people's opinion.But there several challenges for analyzing correct sentiment behind the feedback, opinion so it must to understand and found solution on these challenges of sentiment analysis.[1][2][3]

II. CHALLENGES OF SENTIMENT ANALYSIS

❖ **Sarcasm**

Problem

People use irony and sarcasm in casual conversations and memes on social media. The act of expressing negative sentiment using backhanded compliments can make it difficult for sentiment analysis tools to detect the true context of what the response is actually implying. This can often result in a higher volume of "positive" feedback that is actually negative.

Solution

A top-tier sentiment analysis API will be able to detect the context of the language used and everything else involved in creating actual sentiment when a person posts something. For this, the language dataset on which the sentiment analysis model has been trained, needs to not only be precise but also massive.[4]

❖ **Polarity**

Problem

Words such as "love" and "hate" are high on positive (+1) and negative (-1) scores in polarity. These are easy to understand. But there are in-between conjugations of words such as "not so bad" that can mean "average" and hence lie in mid-polarity. Sometimes phrases like these get left out, which dilutes the sentiment score.

Solution

Sentiment analysis tools can easily figure out these mid-polar phrases and words in order to give a holistic view of a comment. In this context, a topic-based sentiment analysis can give a well-rounded analysis, but with aspect-based sentiment analysis, one can get an in-depth view of many aspects within a comment.[8]

❖ **Idioms**

Problem

Machine learning programs don't necessarily understand a figure of speech. For example, an idiom like "not my cup of tea" will boggle the algorithm because it understands things in the literal sense. Hence, when an idiom is used in a comment or a review, the sentence can be misconstrued by the algorithm or even ignored. To overcome this problem a sentiment analysis platform needs to be trained in understanding idioms. When it comes to multiple languages, this problem becomes manifold.

Solution

The only way this challenge can be met with sentiment analysis accuracy is if the neural networks in an emotion mining API are trained to understand and interpret idioms. Idioms are mapped according to nouns that denote emotions like anger, joy, determination, success, etc. and then the models are trained accordingly. Suffice to say, only then can a tool for analyzing sentiment give accurate insights from such text.

❖ **Negations**

Problem

Negations, given by words such as not, never, cannot, were not, etc. can confuse the Machine Learning model. For example, a machine algorithm needs to understand that a phrase that says, "I can't not go to my class reunion", means that the person intends to go to the class reunion.

Solution

A sentiment analysis platform has to be trained to understand that double negatives outweigh each other and turn a sentence into a positive. This can only be done when there is enough corpus to train the algorithm and it has the maximum number of negation words possible to make the optimum number of permutations and combinations.

- **Comparative sentences**

Problem

Comparative sentences can be tricky because they may not always give an opinion. Much of it has to be deduced. For example, when somebody writes, “the Galaxy S51is larger than the Apple iphone15”, the sentence does not mention any negative or positive emotion but rather states a relative ordering in terms of the size of the two phones.

Solution

Sentiment analysis accuracy can be achieved in this case when a sentiment model can compare the extent to which an entity has one property to a greater or lesser extent than another property. And then tie that to a negative or positive sentiment. This is not an issue of simply having a corpus of negative or positive sentiment-specific words, but in training the artificial intelligence machine to actually pull together information from its knowledge graph and analyze the relationship between entities, words, and emotions.

- **Multilingual data**

Problem

All the problems listed above get compounded when a mix of languages are thrown in. Each language needs a unique part-of-speech tagger, lemmatizer, and grammatical constructs to understand negations. Because each language is unique, it cannot be translated into a base language like say, English, to extract insights. A simple example being, if an idiom “like a fish takes to water” is translated into say, German, the idiom would have lost its meaning.

Solution

The only way these sentiment analysis challenges for multilingual data can be overcome is the hard way. This means that the sentiment analysis model needs to have a uniquely trained platform and named entity recognition model for each language. There is no shortcut to this because the model needs to be trained in each language manually by data scientists. This is a time-consuming process that needs precision and diligence. But the results are worth it because it will give you the highest sentiment analysis accuracy scores as possible. [5][6][7]

- **Audio-visual data**

Problem

Videos are not the same as text data. The challenge is not only that videos need to be transcribed but that they may have captions that need to be analyzed for brand logos. Social media videos also come with comments in addition to the video data.

Solution

A sentiment analyzer can give accurate insights from your data if it extracts information from video content as easily as from text data. For this, it needs to have a video content analysis model that can break down videos to extract entities and glean insights about customer opinion, product insights, and brand logos.

- **Emojis**

Problem

The problem with social media content that is text-based, like Twitter, is that they are inundated with emojis. NLP tasks are trained to be language specific. While they can extract text from even images, emojis are a language in itself. Most emotion analysis solutions treat emojis like special characters that are removed from the data during the process of text mining. But doing so means that companies will not receive holistic insights from the data.

Solution

To meet sentiment analysis challenges like this, a company needs to employ an emotion analyzer tool that can decode the language in emojis and not club them with special characters like commas, spaces or full stops. This in itself is a very advanced application where models like Repustate's are trained specifically for it. Data scientists first analyze whether people use emojis more frequently in positive or negative events, and then train the models to learn the correlation between words and different emojis. [8][9][10][17]

III. SENTIMENT CHALLENGES RELATED TO

- **Content-Related Challenges: Hashtags**

In the context of sentiment analysis,using emoticons such as :) and :(aspositive and negative labels, respectively,is one way of using distantsupervision. Hashtags are also widelyused for different machine learningtasks such as emotion identification. People use a plethora of hashtagsin their tweets about an election.Because of the dynamic nature of theelection domain, the quality, quantity,and freshness of labeled dataplay a vital role in creating a robustclassifier.

In election showed thathashtags were widely used for sarcasm, those hashtags as afeature for our classifier will decreaseaccuracy rather than increase it.

- **Content-Related Challenges: Links**

All existing techniques for tweet classifiersrely solely on tweet content andignore the content of the documentsthey point to through a URL. Those links arecrucial since without them, the tweetis often incomplete and inferring thesentiment is impossible or difficulteven for a human annotator.Therefore, we hypothesize thatincorporating the content, keywords,or title of the documents that a URLpoints to as a feature will increaseour performance. To the best of ourknowledge, there is no work on tweetclassification that expands tweetsbased on their URLs. However, linkexpansion has successfully beenapplied to other problems such astopical anomaly detection and distantsupervision.

- **Interpretation-Related Challenges: Sentiment Versus Emotion Analysis**

The study of sentiment has evolved tothe study of emotions, which has finergranularity. Positive, negative, andneutral sentiments can be expressedwith different emotions such as joyand love for positive polarity; anxietyand sadness for negative; and apathyfor neutral sentiment.

It is considered that emotion as a better criterionfor predicting people's actions, suchas voting, and usually there are significantemotional differences in thetweets that belong to the same polarity.[15]

- **Interpretation-Related Challenges: Vote versus Engagement Counting**

Most or all of the aforementionedchallenges affect the quality of oursentiment analysis approach. At the time of election, it is alsoimportant to correlate a user's onlinebehavior and opinion with that individual'sactual vote. First, the more a user tweets,the more reliably we can predict theuser's opinion. Second, highly activepeople are usually more influentialand more likely to actually vote inthe real world. That is why an electionmonitoring system should reportboth user-level normalized sentimentand tweet-level sentiment. The enduser analyzer must consider both factorsin prediction.

❖ **Trustworthiness-Related Challenges**

A social bot isa computer algorithm that automaticallygenerates content over socialmedia and tries to emulate and possiblychange public attitude. For thelast few years, social bots have inhabitedsocial media platforms. Similarto media reports,we also witnessbot wars between the two sides.

Research targeting pinpointingsources include use of supervisedstatistical models utilizing networkfeatures including retweets, mentions,and hashtag co-occurrence,20user features (such as language,geographical locations, account creationtime, and number of followersand followees), and timing features(such as content generation andconsumption, measuring tweet rate,and intertweet time distribution).Our effort to identify the sourcethat generates a tweet (checkingwhether or not it originates from anAPI) using a hybrid and empiricalapproach gave fairly good resultsas elsewhere. [11] [12]

IV. SOME OTHER TYPE OFSUBJECTIVITY DETECTION CHALLENGES CAN CONFUSE THE MODEL TRAINED FOR POSITIVE AND NEGATIVE CLASSIFICATION

1. The first challenge is that subjectivity detection itself is a subjective task, i.e., a piece of text may be neutral to some people but not to others.
2. The second challenge is improving the accuracy of subjectivity detection in short texts.
3. The third challenge is context dependency. Some words may be objective out of context but could assume subjectivity in a specific context or domain
4. The fourth challenge is reducing the computational cost of training features from a large vocabulary of words[13][14][15][16]

V. CONCLUSION

In this paper discussed the importance and effects of sentimentanalysis for decision making for individuals, company, organizations, government but there are several challenges in sentiment analysis and evaluation. These challenges become the obstacles for finding accurate, subjective analysis of given data. This paper presented moreover all type of challenges of sentiment analysis and also these area open for new researcher for further research in this field.

VI. REFERENCES

[1] . Pang B, Lillian Lee, “Opinion Mining and Sentiment Analysis”, Foundations and TrendsR_ inInformation Retrieval, Vol. 2, Nos. 1–2 (2008) 1–135, _c 2008.

[2] . Liu B., “ Sentiment analysis and opinion mining”, Synth Lect Human Lang Technol 2012, 5:1–167.

[3] . Feldman R., “ Techniques and applications for sentiment analysis”, Commun ACM 2013, 56:82–89.

[4] . F.A. Pozzia, E. Fersinib, E. Messinab, B. Liuc, “Challenges Of Sentiment Analysis In Social Networks:An Overview”, Sentiment Analysis in Social Networks, 2017 Elsevier Inc, http://dx.doi.org/10.1016/B978-0-12-804412-4.00001-2

[5] . Balamurali A R, Aditya Joshi Pushpak Bhattacharyya, “Cross-Lingual Sentiment Analysis for Indian Languages using Linked WordNets”, Proceedings of COLING 2012: Posters, pages 73–82, COLING 2012, Mumbai, 2012.

[6] . Mohammed Arshad Ansari, SharvariGovilkar, “Sentiment Analysis Of Mixed Code For TheTransliterated Hindi And Marathi Texts”, International Journal on Natural Language Computing (IJNLC) Vol. 7, No.2, April 2018.
[7] . Sonali Rajesh Shah, AbhishekKaushik, “Sentiment Analysis On Indian Indigenous Languages: A Review On Multilingual Opinion Mining”, arXiv:1911.12848 [cs.CL], 2019.
[8] . Ameen Abdullah, QaidAqlan, B. Manjula and R. LakshmanNaik, “A Study of Sentiment Analysis: Concepts, Techniques, and Challenges”, Proceedings of International Conference on Computational Intelligence and Data Engineering, Lecture Notes on Data Engineering and Communications Technologies 28, Springer Nature Singapore Pte Ltd. 2019.
[9] . Ganpat Singh Chauhan and Yogesh Kumar Meena, “DomSent: Domain-Specific Aspect Term Extraction in Aspect-Based Sentiment Analysis”, Smart Systems and IoT: Innovations in Computing, Smart Innovation, Systems and Technologies 141, Springer Nature Singapore Pte Ltd. 2020
[10] . Jian Yang, Juan Yang, “Aspect Based Sentiment Analysis with Self-Attention and Gated Convolutional Networks”, IEEE Xplore., 978-1-7281-6579-0/20, 2020 IEEE
[11] . MonirehEbrahimi, Amir HosseinYazdavar, and AmitSheth, “Challenges ofSentiment Analysis for Dynamic Events”, Published by the IEEE Computer Society, 1541-1672/17/$33.00 © 2017 IEEE IEEE INTELLIGENT SYSTEMS
[12] . ItiChaturvedia, Erik Cambria∗,a, Roy E. Welschb, Francisco Herrera, “Distinguishing between facts and opinions for sentiment analysis: Survey and challenges”, Information Fusion 44 (2018) 65–77, journal homepage: www.elsevier.com/locate/inffus, https://doi.org/10.1016/j.inffus.2017.12.006.
[13] . SoujanyaPoria, DevamanyuHazarika,NavonilMajumder, RadaMihalcea, “Beneath the Tip of the Iceberg: Current Challengesand New Directions in Sentiment Analysis Research”, IEEE Transactions on Affective Computing, 1949-3045 (c) 2020 IEEE. DOI 10.1109/TAFFC.2020.3038167.
[14] . Ahmed Alsayat a, NouhElmitwally, “A comprehensive study for Arabic Sentiment Analysis (Challenges and Applications)”, Egyptian Informatics Journal 21 (2020) 7–12, ScienceDirect, https://doi.org/10.1016/j.eij.2019.06.001.
[15] . Carlos A. Iglesias 1,*,† and Antonio Moreno, “Sentiment Analysis for Social Media, Appl. Sci. 2019, 9, 5037, www.mdpi.com/journal/applsci, doi:10.3390/app9235037.
[16] . DoaaMohey El-Din Mohamed Hussein, “A survey on sentiment analysis challenges”, Journal of King Saud University – Engineering Sciences (2018) by Elsevier B.V, 30, 330–338, http://dx.doi.org/10.1016/j.jksues.2016.04.002.
[17] . https://www.repustate.com/aspect-based-sentiment-analysis/?__hstc=753710.715644e62f51c700088e257ceff201f7.1632850888213.1632850888213.1632850888213.1&__hssc=753710.1.1632850888214&__hsfp=2887272311.