INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

Sentiment Analysis in Videos: A Brief Survey

Sarthak Zende, Atharva Wadkar, Aayush Giri, Mujahid Hussain, Prof. S.G.Shinde

Student, Dept. of Computer Engineering, Shrimati Kashibai Navale College of Engineering, SPPU,

Pune, Maharashtra, India

Student, Dept. of Computer Engineering, Shrimati Kashibai Navale College of Engineering, SPPU,

Pune, Maharashtra, India

Student, Dept. of Computer Engineering, Shrimati Kashibai Navale College of Engineering, SPPU,

Pune, Maharashtra, India

Student, Dept. of Computer Engineering, Shrimati Kashibai Navale College of Engineering, SPPU,

Pune, Maharashtra, India

Professor, Dept. of Computer Engineering, Shrimati Kashibai Navale College of Engineering, SPPU,

Pune, Maharashtra, India

ABSTRACT: Sentiment Analysis has been very effective in processing large quantum of data available on the internet to find meaningful insights and polarity from the data. Many industries make use of sentiment analysis such as banking, healthcare, Government, the Stock market, Hospitality, and Retail Markets. Lately, the use of sentiment analysis has been explored for Speech-based polarity detection. This paper surveys sentiment analysis and discusses its three approaches namely Machine learning, Lexicon based, and Hybrid approach. The high level of dependency on the internet and technology nowadays has led to new revenue streams and commercial enterprise models for businesses, but with this arises new gaps and opportunities for companies to exploit. Companies like Uber, Netflix, Meta, Twitter, and Google have been known to possess tremendous amounts of data and have been able to harness meaningful market insights using user data gathered across their services. Sentiment analysis is also helpful in hate speech detection across social media but with the advent of newer social media formats, it has become increasingly difficult to work on different formats of data. There has been less work done on sentiment analysis historically especially sentiment analysis based on Audio and video format if there's any, which has not been researched a lot. Sentiment Analysis validates data information, and data science focuses on deriving meaningful insights from data for problem-solving. Both of these technologies employ algorithms to control interactions with unique data segments.

KEYWORDS: Sentiment Analysis; Machine learning; Lexicon; Hybrid approach; data science

I. INTRODUCTION

Sentiment Analysis (SA) is an ongoing field of exploration in the textbook mining field. SA is the computational treatment of opinions, sentiments, and subjectivity of textbooks.Sentiment Analysis operations are delved into and presented compactly in this check. These papers are distributed according to their benefactions in colorful Sentiment Analysis ways. That makes it pivotal for businesses, artists, content generators, and enough anyone differently on the internet to dissect what people are saying about them and what they offer. It provides vital information on what they can change and ameliorate. Hence, Sentiment Analysis is going to be a significant aspect to help associations and individualists grow. The main thing is to prize the meaningful data and dissect the observation of druggies to apply sentiment because substantially peoples use this way to find others' views for tone- satisfaction in opinion mining. still, some issues raised in opinion mining(OM) like as many words are representing allegorical meanings from a different perspective.

Machine Learning Approach:
Another important concept used in sentiment analysis is data mining. Data mining is a fashion of feting distinct, salutary and interesting patterns in huge quantities of data. The process of analysis involves the identification of if/ also

combinations, followed by the operation of support and confidence criteria. Support refers to the frequency of a particular item that occurs in a dataset, while the number of circumstances of if/ also combinations, which are accurate determines the confidence. Some other data mining parameters include soothsaying, Clustering, Path or Sequence Analysis, and Bracket. The machine literacy system in the sentiment bracket depends on the popular data processing machine literacy styles. Machine literacy styles are popular currently, and perfection is fairly high. Because of the diversity of semantic expression, wordbook matching has a large error, but the machine literacy styles won't be such a problem. The scenes of machine literacy styles are different, they can complete both private and objective brackets or negative emotion brackets and do not need to go into the words, rulings, or alphabet position as well as the wordbook matching. This textbook will carry on the machine literacy system of emotion analysis, and introduce the concrete contents and the evaluation styles of colorful machine literacy styles. SVM and the naïve bayesian model are some of the models we can use in the machine literacy approach

Lexicon based Approach: The lexicon-based approach involves calculating exposure for a document from the semantic exposure of words or expressions in the document. In wordbook-grounded sentiment analysis, words in textbooks are labeled as positive or negative(and occasionally as neutral) with the help of a so-called valence wordbook. Take the expression " Good people occasionally have bad days. ". A valence wordbook would label the word " Good " as positive; the word " bad " as negative; and conceivably the other words as neutral. Once each word in the textbook is labeled, we can decide on an overall sentiment score by counting the figures of positive and negative words and combining these values mathematically. A popular formula to calculate the sentiment score

$$StSc = \frac{\text{number of positive words} - \text{number of negative words}}{\text{total number of words}}$$

If the sentiment score is negative, the text is classified as negative. It follows that a positive score means a positive text, and a score of zero means the text is classified as neutral.

Hybrid approach: Hybrid models are an amalgamation of Machine learning models with other techniques or soft computing/ optimization methods to improve the sentiment analysis approach**.** It involves combining AI techniques to achieve better results than the machine learning approach as it is a combination of machine learning and different techniques it gives better results. A sentiment analysis system at the judgment position, using a combination of sentiment dictionaries, NLP essential tools, and fuzzy sets ways, should perform the same or better than moment's accepted textbook bracket supervised machine literacy algorithms when the ultimate are utilized in insulation.

II. RELATED WORK

In paper [1] it states a comprehensive overview of the last update in machine learning. Many recently proposed algorithms enhancements and their applications are investigated and briefly explored. The results of the survey are presented briefly in this survey paper. In this survey paper, Sentiment Analysis identifies the sentiment expressed in a text and then analyses it as positive or negative. And there are different methodologies used which are broadly classified into two main categories Machine learning approach and lexicon-based approach. The furthermore subdivision is shown in this figure.
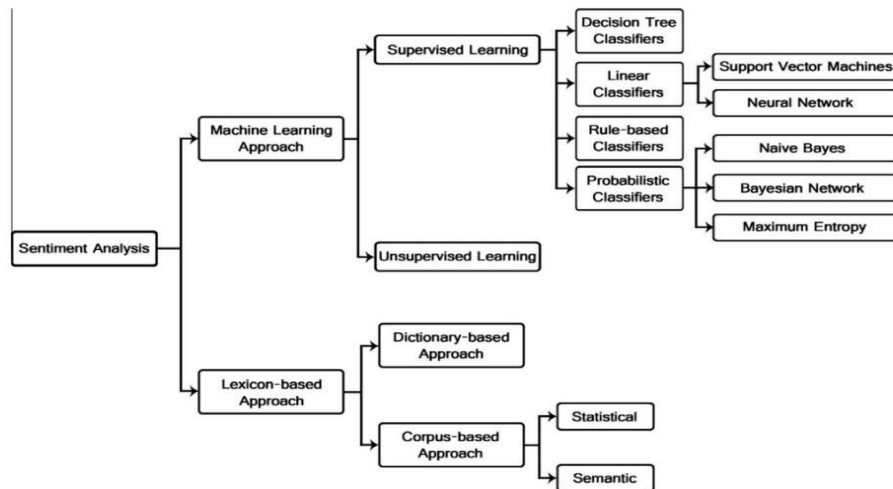
Figure 1: Sentiment Classification Techniques

Fifty-four of the recently published and cited articles were categorized and summarized in this survey paper. These articles gave contributions to many SA-related fields that use SA techniques for various real-world applications. Naive Bayes and Support Vector Machines are the most frequently used ML algorithms for solving SC problems. The most common lexicon source used is WordNet which exists in languages other than English. Building resources, used in SA tasks, is still needed for many natural languages. The interest in languages outside English is growing as there is still a lack of resources and research concerning these languages. The SA process has attracted researchers recently and still needs some enhancements.

In [2] The main aim of this survey paper was to develop a Multimodal Sentiment Analysis domain that can extract the information in terms of the emotions of the people regarding a specific product from any of the three mediums (text, audio, or video) can lead to the better technological aspects in the market. The objective of the research was to build a system that can identify the sentiment categorized into six types: anger, joy, disgust, sadness, fear, and surprise of a video. The system developed depicts how much of each of these sentiments is present in a particular input. In this process first, the entire video clip was cut into multiple shorter clips based on time. Next, the shorter video clips are turned into screenshots which are taken at the frequency of one frame per second. And these clips are split into audio clips and then audio clips are extracted to form text files. After extraction output is presented to the user on the bases of early mentioned category.

In [3] we explored sub-symbolic representations gained from semantic concepts to gain insights into the emotional and contextual information provided by video transcriptions. Furthermore, we have successfully leveraged the derived features to automatically classify video segments regarding arousal and valence as well as 10 domain-specific speaker topics. This used the semantics in a more unsupervised way to explore the content of videos by clustering and in combination with high-level feature sets of other modalities (e.g., face and voice features) for multi-modal modeling. They performed a sentiment analysis study focusing on topics and emotions in video car reviews from YouTube. Thereby, in this they aim at a more in-depth exploration of the spoken word, hence, they make use of transcriptions. Therefore, they used two common ways to utilize text computationally: a) understand the meaning of words from their symbolic representation through knowledge-based and statistical approaches, and b) learn a continuous vector space (embeddings) from the symbolic space of words. The figure below is the pipeline process of sentiment analysis from the video, which is briefly explained.

In [4] we performed a vital task of sentiment analysis for classifying the opinions/reviews and sentiments automatically into three classes positive, negative and neutral by using different methods described by researchers. And the different methodologies used are broadly classified into two main categories: Machine learning approach and lexicon-based approach. Different algorithms are used such as naïve Bayes, vector support machines, neural networks and etc. SVM (support vector machines) is presenting high performance as compared to other algorithms but it also has some limitations.

In [5] we focused on sentiment analysis using machine learning for airline datasets. In this, the conclusion is derived by finding the polarity of a particular review whether it is poor, average, or excellent for the airline dataset and 0 or 1 for the other three datasets used in this process. The entire task was performed using Python. The rich corpus of data sets was used to find out various trends and patterns effectively. The variety of datasets, that were taken into our consideration assisted the study and application of different algorithms like Logistic Regression, KNN (k - Nearest

Neighbours), Support Vector Machine Classifier, Decision Tree Classifier, Random Forest Classifier, AdaBoost Classifier, AdaBoost Classifier and bagging Classifier. And also assisted in creating a successful model for predicting outputs.

In [6] they speak about the problems in this era of microblogging, wherein people proportion their personal perspectives or tweets on social media and analyze the sentiment of people based totally on their perspectives to take on-the-spot step strategies. Which can be taken for committing sentiment analysis like machine learning classifiers, Polarity-based sentiment analysis, and deep learning models which can be used to categorize users' tweets as having "positive" or "negative" sentiment. To detect sentiment and text simultaneously a probabilistic model is used based on LDA, this model is unsupervised and the preliminary results are promising. The Paper uses the concept of polarity in Phrase level sentiment analysis. It first determines whether the expression is neutral or polar then removes the ambiguity of the polar expressions. The main motive is to be able to pinpoint positive or negative sentiments. To gain information from unstructured forms of data, we first need to process them using various methods.

The paper [7] proposes a human sentiment analysis model that can perform sentiment analysis on any data set. To perform sentiment analysis on a given set of data, the model employs a dictionary-based approach known as the "lexicon method." The author uses the SentiWordnet lexicon dictionary to derive the sentiment polarity associated with each word. The model consists of five steps: data collection (dataset), pre-processing, feature selection and extraction, sentiment classification, and sentiment score calculation. For future related work, the author thinks about how the accuracy and efficiency of the system can be enhanced by using a sentiment dictionary that would be able to classify emotions not only as positive or negative but also as moods such as angry, boring, happy, and sad. In [8] the author proposed a lexicon-based approach to calculate the sentiments of the fans by using the NRC Emotion Lexicon before, during, and after the match. At the end, they deduced a one-to-one correspondence between players' performances and fans' sentiments and showed how those sentiments varied over a period of time.

This research [9] discusses a hybrid approach to sentiment analysis (SA) that combines a hybrid standard classification with semantic rules and fuzzy sets. The Movie Review Dataset, a publicly accessible dataset, is used for the testing. The hybrid system model that has been developed operates with a high degree of accuracy and precision. It enhanced the outcomes obtained before using the same dataset using Naive Bayes (NB) and Maximum Entropy (ME). As a result, the paper's initial problem statement—that a hybrid technique using semantics and fuzzy sets should be able to perform better—is satisfied. The paper also managed to distinguish between various polarity strengths. SentiWordNet development over time should enhance the proposed system with time.

A hybrid technique is utilised in publication [10] to identify sentiment analysis from news comments. This study uses a hybrid strategy to identify polarity using a sentiment lexicon, and then uses the lexicon's findings to train machine learning (M.L.) algorithms. Support Vector Machine (SVM) and K-Nearest Neighbor are the machine learning methods utilised (kNN). The paper's findings demonstrate that SVM outperforms kNN when evaluating news comments. In order to provide more accurate sentiment labels, the article suggested developing a domain-specific lexical dictionary and adding a number of opinion criteria.

III. CONCLUSION AND FUTURE WORK

In this paper, we surveyed and discussed Speech Sentiment Analysis and its three main approaches. Basically, we concluded three approaches in this work for analysis like machine learning approach, lexicon-based approach, and hybrid approach, and respectively came to various results of sentiment analysis and understood through working of the various models. Using machine learning we detected sentiment and text simultaneously a probabilistic model is used based on LDA, the model is unsupervised and the preliminary results are acceptable. With the lexicon-based method, the lexicon dictionary is used to derive the sentiment polarity associated with each word, and emotions are categorized. In Hybrid models, an amalgamation of Machine learning models with other techniques or soft computing/ optimization methods is used to improve the results.

The work done by us mainly focuses on text analysis where a small dataset is prepared however the dataset can be improved and a bigger dataset can be achieved for analysis. And for future work, the sentiment analysis can be extended to be performed on audio, video, and even on emojis which would be a better option for this research.

REFERENCES

1. Walaa Medhat , Ahmed Hassan , Hoda Korashy . In 2014 Ain Shams Engineering Journal 5,pages 1093-1113.
2. Dr. Ashwini Rao, Akriti Ahuja, Shyam Kansara, Vrunda Patel ,"Sentiment Analysis on User-generated Video, Audio and Text". In 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS).

3. Lukas Stappen, Erik Cambria, Bjorn W. Schuller ,” Sentiment Analysis and Topic Recognition in Video Transcriptions”. In IEEE Intelligent Systems, March/April 2021.
4. Shamsa Umar, Maryam, Fizza Azhar, Ghulam Samdani, Sayyam Malik,” Sentiment Analysis Approaches and Applications: A Survey”. In International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 1, July 2018.
5. Sentiment Analysis using Machine Learning Techniques on Python,” Sentiment Analysis using Machine Learning Techniques on Python”. In Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3.
6. Abbi Nizar Muhammad, Saiful Bukhori, Priza Pandunata,” Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes – Support Vector Machine (NBSVM) Classifier”. In Proc. ICOMITEE 2019, October 16th-17th 2019, Jember, Indonesia.
7. Singh, V., Singh, G., Rastogi, P., & Deswal, D. (2018). Sentiment Analysis Using Lexicon Based Approach. 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC). doi:10.1109/pdgc.2018.8745971
8. Agarwal, A., & Toshniwal, D. (2018). Application of Lexicon Based Approach in Sentiment Analysis for short Tweets. 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE). doi:10.1109/icacce.2018.8441696.
9. Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to sentiment analysis. 2016 IEEE Congress on Evolutionary Computation (CEC). doi:10.1109/cec.2016.7744425.
10. Mukwazvure, A., & Supreethi, K. . (2015). A hybrid approach to sentiment analysis of news comments. 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions). doi:10.1109/icrito.2015.735928210.1109/icrito.2015.7359282.

BIOGRAPHY

Sarthak Zende is a B.E student in the Computer Engineering Department, at Shrimati Kashibai Navale College of Engineering, SPPU (formerly Pune University). He is in the final year of his Bachelor's in Computer Engineering and will pass out in 2023. His research interests are Machine Learning, Natural Language Processing, Artificial Intelligence, Algorithms, Data science, etc.

Atharva Wadkar is a B.E student in the Computer Engineering Department, at Shrimati Kashibai Navale College of Engineering, SPPU (formerly Pune University). He is in the final year of his Bachelor's in Computer Engineering and will pass out in 2023. His research interests are Machine Learning, Data Science, Data Structures and Algorithms, and Web development.

Aayush Giri is a B.E student in the Computer Engineering Department, at Shrimati Kashibai Navale College of Engineering, SPPU (formerly Pune University). He is in the final year of his Bachelor's in Computer Engineering and will pass out in 2023. His research interests are Design and analysis of Algorithms, Data Analytics, DBMS, Front End Web development.

Mujahid Hussain is a B.E student in the Computer Engineering Department, at Shrimati Kashibai Navale College of Engineering, SPPU (formerly Pune University). He is in the final year of his Bachelor's in Computer Engineering and will pass out in 2023. His research interests are Web 3.0, Blockchain Technology, DBMS, Full stack Web development, etc.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462   📞 6381 907 438   ✉ ijircce@gmail.com