A Survey on Sentimental Analysis Techniques and its Usage in Recommendation Systems

Tinashe Rufu
School of Information Sciences and Technology.
Harare Institute of Technology
Harare, Zimbabwe

Tawanda Mudawarima,
School of Information Sciences and Technology.
Harare Institute of Technology
Harare, Zimbabwe

Stanford Mavuchi
School of Information Sciences and Technology
Harare Institute of Technology
Harare, Zimbabwe

Abstract:- The widely anticipated use of natural language processing (NLP) and machine learning is sentiment analysis (SA). The introduction of Web 2.0 resulted in this industry expanding significantly. People have long used the Internet as a forum to share their thoughts, feelings, and judgments about things, other people, and life in general. As a result, the Internet is currently a vast supply of humorous written data that expresses opinions. The automatic classification of opinionated writing as being negative, positive, or neutral is a crucial function of sentiment analysis. This study compared methods, techniques used in sentimental analysis, and its application in recommendation systems. Since the advent of recommender systems, this field of study has become increasingly important.

***Keywords; -** Sentimental Analysis, Recommendation System, Emoji Analysis, Sarcasm Detection, Machine Learning.*

I. INTRODUCTION

Analytics has become a fuel engine of the twenty first century time and data is a key part of the building [8]. Every industry uses data which can be found in image format, text, numeric and videos. Information is infinite because it can be observed in the universe. As the amount of data increases, so do the needs of the organization. On average, 8.23 tweets are sent every second, 510 comments are sent every minute, 294,000 status updates are made on social media site Facebook on an hourly basis, and the international store chain Walmart process larger than one million transactions of customers on an hourly basis [8]. Nowadays, social sites and media platforms like Whatsapp, Twitter have become a primary source of communication and information exchange. With millions of users sharing their thoughts and opinions on a daily basis, social Medias provides a huge amount of big data that can be examined to gain insights into various aspects of society, including public sentiment towards different topics. Sentiment analysis refers to the method of automatically extracting and identifying information that is subjective from text, has emerged as a valuable tool for analyzing social media data.

There are basically three types of data in big data which include unstructured, structured and semi-structured. Structured data are raw information with specific repeating patterns. This pattern makes managing, reading and processing data easy for any application, examples include MySQL and Oracle [15].Unstructured data is illogical or unrelenting pattern type of data for example video, feedback, text, audio, Facebook [15].Semi-structured data can be described as data without a schema or view and data that can describe itself, example Sensor data, NoSQL, CSV, XML and JSON are considered semi-structured [15].The paper examined types of recommendation systems which are divided into two , collaborative filtering and content based filtering.A better approach is the socalled hybrid analytical ap proach, where the two analytical methods are useful at differe nt levels of agreement. The concept of analysis is done at three levels, Document level, Sentence level and Physiological level.

II. LITERATURE REVIEW

A. Advisory System

The advisory system also known as recommender system is systems that provide recommendation to user based on ratings and opinions [14]. Recommender systems are tools and technologies that give recommendations of products/services that may be useful to users. Recommendations are made based on product profiles and user profiles created using various methods and algorithms. Know users' interests and find unique and relevant content created for them based on their preferences and value models [14].

a) Document level Sentiment Analysis

The distribution of all information is based on the opinion of each person [11]. The problem is on classification of text. A person writes an article to express his opinion about an organisation. The real challenge is that at text level not every sentence needs to comment on a topic. It is important to distinguish between subjectivity and objectivity in the classification process. Editing should not contain false statements. Data classification at the data level can be done using both supervised and unsupervised learning. Techniques such as Support Vector Machine, Bayesian Classifiers, and Maximum Entropy are used.

Semantic information can be used to determine the meaning of inferred words. Therefore, classification of data level theory has both advantages and disadvantages. In fact, data can broadly describe how people feel about a particular organization [11].

b) Sentence level sentiment analysis

The polarity of each auxiliary time is determined .In some cases; data classification is not suitable for everyone's needs. In early days of sentence level analysis many efforts focused on finding content and most of the methods used supervised learning. This can be divided into two tasks, determining which sentence represents the original sentence and then determine the star type or write each sentence true or bad. While many objective sentences can convey emotion, they cannot express positive or negative emotions or decisions. Many ideas, facts, and concepts can be found in one sentence.

c) Feature-based or aspect-level sentiment analysis

It provides additional and more detailed information for example it shows whether people like a product or not. Online reviews in most known organizations rely on relevant blog information. Another interesting idea to solve problem in sentiment analysis is use of ideas to create graphic model of the text [16]. The model performs graph based queries to determine appearances of all text and uses segmentation records to extract related images [11].

Theory classification methods can be divided into machine lea rning methods, dictionary methods and hybrid methods as sho wn in Table.

Table 1: Classification Methods

Sentiment approaches	**Classification**	**Features/techniques**	**Advantages and limitations**
Machine learning (ML)[13]	supervised and unsupervised learning	**Supervised-** • SVM • K Nearest Neighbour • Naves Bayes • Logistic Regression • Deep Neural Network • Convolutional Network • Recurrent Neural Network **Unsupervised-** • K Means • Generative adversely Network • Restricted Boltzmann Machine • Auto Encoder	**Advantages** • Eliminates manual entries • Accurate • Unsupervised learning methods' detection performance is often lower to that of supervised learning methods. • **limitations** • The fundamental impediment to supervised learning is a lack of adequate labelled data. • Manually labelling data is costly and time intensive for those who are not tagged
Lexicon Based[10]	Dictionary Based Corpus based (Statistic or semantic)	• Dictionary Based • Corpus based Manual	**Advantage:** • Coverage of wide spectrum of words - **Limitations:** • Orientation of sentiments is fixed and score of words
Hybrid[10]	Machine learning lexicon based	• lexicon built with public data resource • Sentiment words used as features in the ML method	**Advantage:** • Sentiment detection and measurement that is less susceptible to change in the issue domain **Limitations:** • Noisy reviews

B. Individuals and communities.

In [14], the writer proposed a recommendation system which is personalized based on automata learning and sentimental analysis making usage of the LASA framework, with the framework recommending areas near the user's current geographical location. The system calculated an average penalty using a probability vector. The system's benefits include personalized recommendations due on the system user previous knowledge and improved effectiveness, but there are some security concerns. In [1], the Keyword-Aware Service Recommendation (KASR) approach was presented for customized suggestions using user-based

ISSN No:-2456-2165

collaborative filtering techniques implemented on Hadoop for scalability. The evaluation was done using the Jaccard coefficient and cosine similarity metric, and the approach was found to be superior to conventional methods in terms of scalability and effectiveness. However, there are drawbacks, such as not dividing user reviews into positive and negative feedback and the Jaccard coefficient approach's unreliability in not accounting for textual sentiments. In [17], the author proposed a social framework that uses machine learning and sentiment analysis methods to make personalization and recommendations ranking based on a user's preferences and first choices. The framework removes unwanted data, enabling faster services and better results, but the framework accuracy could increase by application of either or both context-based and preference-based searching.

In [2], the writer proposed a recommendation system based on Bayesian inference that uses ratings shared with friends to calculate similarity scores for ranking. The proposed system was found to be better than the existing trust-based recommendation, with benefits such as greater accuracy through friends' recommendations and solving the difficulty of a huge number of elements in recommendation using collaborative filtering. However, the model has drawbacks like cold initiation and sparseness of ratings issues. In [18], the investigator made a proposal of a technique to get user feedback in situations where most users may fail to give feedback with the method of automatically extracting data about the specific product from Twitter. The writer made comparison of text classifier methods and techniques and came to a result of concluding Support Vector Machine to have accuracy of 82% with it being used as classifier for text. The merits of the technique are greater accuracy and rating, even with a sparse of user’s direct information, but its usage in other domains needs further examination.

In [19], the writers proposed a recommender system for game videos based on client-side usage, independent of user count or audiovisual consumption, avoiding computation, concurrency, and security issues that arise when using central server methodologies. However, the system cannot be used to recommend specific video clips. In [20], an individualized probabilistic travel suggestion model was proposed that uses individuals' traits and images to mine demographics for travel landmarks and paths and thus benefits personalized trip suggestion services. The merits include producing more pleasing results and enhancing group recommendations through travel preference analysis, but investigating more competitive recommendation models is necessary. In [21], the authors presented separate approaches to dealing with text features which includes the model of bag-of-words, negation handling, adverbs and adjectives controlling, threshold hurdling word occurrences, and substitute knowledge using WordNet. The following were examined: K-means clustering, decision trees, maximum entropy, and naive Bayes, based on the accuracy of classifying instant memos by semantic significance. The merits of this study include evaluation of the fit of separate feature selection and algorithms for learning for the classifying comments based on polarity and subjectivity.

C. Emojies

Emojis have become increasingly popular on social media as a means for users to express their emotions and opinions. As a result, researchers have become interested in using sentiment analysis on emojis to better understand user sentiments. Studies have explored the use of emojis in sentiment analysis, with a particular focus on Twitter. In [22], proposed a method for improving the sentiment analysis of tweets by incorporating emojis and emoticons. Their findings showed that adding emojis as an additional feature can improve the accuracy of sentiment analysis, especially for tweets with ambiguous sentiment. Similarly, [23] suggested that using emojis in a sentiment analysis model can enhance the model's performance, particularly for tweets with complex emotions. Sentiment analysis of emojis is an emerging field that has potential applications in social media monitoring, customer feedback analysis, and opinion mining. These studies provide valuable insights and methods for incorporating emojis into sentiment analysis models, improving the accuracy and effectiveness of sentiment analysis on social media data.

Fig 1: Some of the images and their meanings [28]

Emoji	**Name**	**Meaning**
	Face with Tears of Joy	Extreme happiness or laughter
	Red Heart	Love (red by default, but meaning is same for any colour)
	Person with Folded Hands	Prayer, thank you, and sometimes a high five
	Thumbs Up Sign	Well done, good job, or approval
	Loudly Crying Face	Uncontrollable tears, perhaps due to sadness or joy

Sarcasm is a linguistic device that involves expressing the opposite of what is actually meant. It is commonly used on social media platforms like Twitter and Facebook, allowing users to express their genuine feelings about a subject without being overtly negative. Detecting sarcasm in sentiment analysis is a crucial area of study, as it can enhance the accuracy of sentiment analysis algorithms and provide more detailed insights into social media discussions. There are two primary approaches to detecting sarcasm in sentiment analysis: machine learning algorithms and rule-based methods. Machine learning algorithms like support vector machines (SVMs), decision trees, and neural networks can be trained on labeled datasets of sarcastic and non-sarcastic tweets to identify patterns and characteristics that indicate sarcasm. In one study, [4] utilized deep learning to achieve high accuracy rates in sarcasm detection in tweets.

On the other hand, rule-based methods, such as lexical and syntactic analysis, employ a set of rules or heuristics to identify words or phrases that are frequently utilized in sarcastic language, such as negation and exaggeration. In [5] proposed a hybrid model that combines rule-based and machine-learning methods for sentiment analysis and sarcasm detection in Arabic tweets. Recently, researchers have explored the use of multimodal data, such as text and emojis, for sarcasm detection. Emojis can convey a range of emotions and attitudes that text alone cannot capture. In [22] incorporated emoji's and emoticons to enhance the accuracy of sentiment analysis in tweets, while [3] investigated the use of emojis to improve sentiment analysis on Twitter. Detecting sarcasm in sentiment analysis is a complex and challenging research area that necessitates the application of advanced machine learning and rule-based methods. Recent studies have exhibited promising results in employing multimodal data such as text and emojis to improve sarcasm detection accuracy. Further research is required to develop more robust and scalable algorithms that can handle the dynamic nature of social media conversations.

In [3], the author proposed a recommendation which is Bayesian-based for social networks to be carried online. With this, grades are shared with colleagues. Probabilities (conditional) are used to calculate rating similarity. A ranking is made based on the similarity score. It shows that Bayesian inferred advice is better than advice based on existing beliefs. The main advantages were on high accuracy through suggestions from your friends and solved the problem of large elements in the joint filter proposal requirements. The major drawbacks included price problem and cold initiation. In [5] proposed a recommendation system for video games broadcast on the network and included everything from the Olympic Games. Tips were used only by customers, regardless of the number of users or audiovisual consumption. This avoids the computational, compatibility, and security problems that arise when using a centralized server methodology in situations of huge user base for example, Olympic sports events. A special probabilistic travel proposal model was proposed [6]. Using personality traits and images is an effective way to mine demographics for travel destinations and routes, and is useful for personal travel recommendation services in this area. The benefits include group recommendations by analyzing travel preferences.

In [7], proposed a precise analysis of the accuracy and usage of machine learning methods in classifying immediate memories based on their semantic value. They proposed different approaches to extract text features, such as bag-of-words models, error resolution, phrase and adjective control, word occurrence thresholding, and WordNet replacement. Performance is measured using accuracy of the following techniques, K-clustering, maximum entropy, decision trees, and simplex bins. It evaluated the fitness of learning algorithms in selecting different features and classifying them based on subjectivity and polarity of interpretation.

III. ANALYSIS OF TECHNIQUES

Table 2 summarised sentiment analysis techniques with their associated limitations.

Table 2: Summary of Techniques

Title Issues	**Addressed**	**Proposed Technique**	**Limitations**
[24]	Sentiment Analysis of Twitter Data	Conventional machine learning algorithms	Limited to twitter
[12]	. Sentiment Analysis of Image with Text Caption	Deep Learning Techniques	Limited to only images from food sources
[25]	Improves accuracy text SA	CNN combined with SVM	Only for Chinese
[26]	polarity shift	Opinion lexicon based and naïve Bayes Algorithm	Polarity of phrase can differ from polarity of word into it

[27]	Novel adaptability approach	Dictionary based approach	Limited to twitter data it does not distinguish the impact degree of the different metrics in order to accentuate a feeling
[9]	improving the accuracy by extending the labelled data	semi supervised model based on dynamic threshold and metaclassifier	Limited base learner

IV. CONCLUSION AND FUTURE RESEARCH

Recent researches have focused on the problem of machine learning sentiment analysis. Although some significant efforts have been made in this area, a fully automated and customized system has not yet been demonstrated. This is because natural language is not structured. New research shows how sentiment analysis can be used to create more accurate recommendation systems. Classification was done on some recommender system methods based on sentiment analysis. The goal of future research will be to improve traditional methods, algorithms to improve accuracy of recommender system's recommendations and predictions. Further research is certainly needed to further improve outcome assessment. Sentiment analysis is adept for a variety of purposes. On the other hand, although the methods and algorithms used for sentiment analysis are developing rapidly, many challenges in the field remain unanswered. Among the most difficult aspects are using extra language, interacting with negative words, production features, sentence or document complexity, internal production features, etc. Depending on design considerations, this challenge may be the subject of further research in the future. Business decision making is often based on sentiment analysis, also known as brainstorming. Many businesses and organizations base all business decisions solely on customer reviews. Many methods can be used to analyze sentiment. After reviewing all these documents, there is no interest in languages other than English and Chinese due to lack of resources and research in other languages. Future challenges include using sentiment analysis to improve capabilities and deep learning to analyze more social media data. Creating dynamic bookmarks will also be a priority.

ACKNOWLEDGMENTS

I am grateful to the department of School of Information Science and Technology of Harare Institute of Technology for their cooperation and suggestions.

REFERENCES

[1]. Sandra Garcia Esparza, Michael P. O‟Mahony, Barry Smyth, Mining the real-time web: A novel approach to product recommendation, Elsevier Journal - Knowledge-Based Systems 29 (2012) 3–11.
[2]. Cloud recommendations Assessment-Based Learning Automation for nMachine Learning-Based Sentiment Analysis of Social Data (researchgate.net)
[3]. X. Yang, Y. Guo, Y. Liu, "Suggestions Based on Bayesian Inference in Online Social Networks", IEEE Operations in Parallel and Distributed Systems, Vol. 24, Issue 4, p. 642-651, 2013
[4]. Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., & Wright, B. (2020). Sarcasm detection using machine learning algorithms in Twitter: A systematic review. International Journal of Market Research, 62(5), 578–598. https://doi.org/10.1177/1470785320921779
[5]. M. Alduan, F. Alvarez, J. Menendez and O. Baez, "Game Video-Based Recommendation Systems User Perception," in Multimedia IEEE Transactions, vol.14, No.6, p. 1546-1557, 2013.
[6]. Yan-Ying Chen, An Jung Cheng, "Travel Advice Based on Mining Characteristics and Travel Group Types from Community Contributed Photos" IEEE Transactions in Multimedia, ib. 15 6, October 2013
[7]. Gasparetto, A.; Marcuzzo, M.; Zangari, A.; Albarelli, A. A Survey on Text Classification Algorithms: From Text to Predictions. *Information* 2022, *13*, 83. https://doi.org/10.3390/info13020083
[8]. Applications and problems of evaluation theory: A survey – IJERT
[9]. Han Yan, Liu Yan., Jin, Z "Sentiment Analysis of Semi-Supervised Learning: Dynamic Thresholding and Multiple Classifier Based Models", Phau ntawv Journal of Neural Computing and Applications, pp1-13 (2019). https://doi.org/10.1007/s00521-018-3958-3 Computing Intell, Vol. 0 (2012)
[10]. Rodrigo Moraes, João Francisco Valiati, Wilson P. Gavião Neto "Document-level sentiment classification: an empirical comparison between SVM and ANN" Expert Syst Appl, 40 (2013), pp. 621-633
[11]. Rahul Tejwani, "Sentiment Analysis: A Survey" Hoda Mohamed

[12]. Chaubey PK, Arora TK, Raj KB, Asha GR, Mishra G, Guptav SC, Altuwairiqi M, Alhassan M. Sentiment analysis of images with text subtitles using deep learning techniques. The Neuroscience of Computational Intelligence. 25 Lub Rau Hli 2022; 2022: 3612433. Ministry of Interior: 10.1155/2022/3612433. PIB: 35795734, PMC: 9252700.
[13]. Liu, H.; Lang, B. Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. Appl. Sci. 2019, 9, 4396. https://doi.org/10.3390/app9204396
[14]. Suggestion Systems – Wikipedia
[15]. https://www .geeksforgeeks.org/
[16]. Wei jin, Srihari," Mining Hidden Associations in Text Corpora through Concept Chain and Graph Queries "
[17]. Shrote, Khushboo R., and A. V. Deorankar. "Review based service recommendation for big data." Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2016 2nd International Conference on. IEEE, 2016.
[18]. Shunmei Meng, Wanchun Dou, Xuyun Zhang, Jinjun Chen," KASR: A Keyword-Aware Service Recommendation Method on Map Reduce for Big Data Applications" IEEE Transactions On Parallel and Distributed Systems, TPDS-2013-12-1141.
[19]. G. Kang, J. Liu, M. Tang, X. Liu and B. cao, "AWSR: Active Web Service Recommendation Based on Usage History," 2012 IEEE 19th International Conference on Web Services (ICWS), pp. 186-193, 2012.
[20]. Zibin Zheng, Xinmiao Wu, Yilei Zhang, Michael R. Lyu, Fellow, and Jianmin Wang," QoS Ranking Prediction for Cloud Services" IEEE Transactions On Parallel and Distributed Systems, Vol. 24, No. 6, June 2013.
[21]. Meghana Ashok, Swathi Rajanna and Pradnyesh Vineet Joshi "A Personalized Recommender System using Machine Learning based Sentiment Analysis over Social Data." Conference on Electrical, Electronics and Computer Science IEEE 2016
[22]. Verma, Jai Prakash, Bankim Patel, and Atul Patel. "Big data analysis: recommendation system with Hadoop framework." Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on. IEEE, 2015.
[23]. Burke, Robin. "Hybrid recommender systems: Survey and experiments." User modeling and user-adapted interaction 12.4 (2002): 331-370.
[24]. Yili Wang, Jiaxuan Guo, Chengsheng Yuan, Baozhu Li: Sentiment Analysis of Twitter Data
[25]. Chen Y, Zhang Z (2018) Research on text sentiment analysis based on CNNs and SVM. In: 13th IEEEconference on industrial electronics and applications (ICIEA), pp. 2731–2734
[26]. S M Shuhidan, S R Hamidi, S Kazemian et al., "Sentiment Analysis for Financial News Headlines using Machine Learning Algorithm[M]", Proceedings of the 7th International Conference on Kansei Engineering and Emotion Research, 2018
[27]. El Alaoui, I., Gahi, Y., Messoussi, R. et al. A novel adaptable approach for sentiment analysis on big social data. J Big Data 5, 12 (2018). https://doi.org/10.1186/s40537-018-0120-0
[28]. The 100 Most Popular Emojis Explained (makeuseof.com)