INFORMATICS JOURNALS

107

SHRI-LEKH (Multi-disciplinary Research Journal) Vol 1(2), December 2025, p. 107-114

Online ISSN: 3139-5694
DOI:10.18311/SHRI-LEKH/2025/v1i2/53990

Opinion Mining in the Era of Social Media: Techniques, Challenges, and Future Directions

Priyanka Saurabh Sharma[1]* and R. K. Dhuware[2]

[1]Department of Computer Science, J. M. Patel Arts, Commerce and Science College, Bhandara – 441904, Maharashtra, India

[2]Department of Computer Science, Dhote Bandhu Science College, Gondia – 441614, Maharashtra, India

Abstract

Opinion mining, also known as sentiment analysis, is a subfield of Natural Language Processing (NLP). This field focuses on identifying and extracting subjective information from text, with the explosion of social media, online reviews, and user-generated content. Opinion mining has become a key technology for researchers, businesses, governments, healthcare, education, finance, etc. This paper focuses on the overall history and evolution of opinion mining and provides a comprehensive review of current techniques, discusses major challenges, and outlines future directions in the field. This paper also gives some ideas about datasets and benchmarks. This paper aims to present opinion mining in the era of social media. This paper walks through and highlights the basic concept of opinion mining, the history of opinion mining, tools used to mine the opinions, opinion mining applications and issues faced in mining the opinions. The significance of this paper is also to the latest approach and outlines future directions.

Keywords: Datasets and Benchmark, Machine Learning, Opinion, Opinion Mining, Sentiment Analysis, Sentiment Classification, Machine Learning

INTRODUCTION

In a world where it's incredibly easy to connect with or hear from anyone, understanding what people really think is a powerful skill. Here, one hidden tool used by researchers, businesses, and decision-makers is opinion mining. It is a branch of Natural Language Processing (NLP) that extracts subjective information from text, including sentiments, views, and emotions. In essence, it enables computers to recognise the emotional tone of written language and identify whether a passage conveys neutral, positive, or negative emotions. It uses NLP techniques, machine learning, and data mining to process text from various sources like customer feedback, online reviews, and social media. It helps organisations understand customer sentiment, brand perception, and overall public opinion on various topics. The aim of opinion mining is not just overall opinion but also the specific aspects or entities that people are expressing opinions about. One of the most difficult aspects of opinion mining is feature weighting, which is critical for accurate classification. Opinion mining techniques are used to analyse the feelings, emotions, opinions, feedback and concerns of several users in the form of reviews.

Brief History and Evolution of the Field

A new kind of society where people may freely connect and exchange ideas and opinions has emerged as a result of the internet's increasing accessibility.

*Email: priyankasaurabhkumarsharma@gmail.com

Opinion Mining began in 1982, when people sought an automated method of text interpretation rather than relying on their own brains to process natural language (Anbananthen K. & Elyasir A., 2013). Using subjective and objective sentences as indicators, scanning the text with example questions, and treating the psychological component as a crucial component of natural language processing are all employed in this step of text interpretation.

The evolution of the text annotation area, which is characterised as labelling and tagging words while reading any type of material, is depicted in the second phase timeline. Because corpora are used at the analysis step, text annotation takes the place of conventional text interpretation approaches.

The third phase, which began in the early 2000s, focuses on the distinction between subjective and objective methods that yield more accurate polarity categorisation at the document or phrase level and provide detailed or summary results, respectively. This process is known as the extraction technique. In the third phase, opinion mining is advanced to a new application period and placed at the centre of the internet to evaluate unstructured text across several domains and contexts, a process known as web data mining.

In the fourth phase of the timeline, the progressive stage of opinion mining was fostered by web data mining, which created web applications to assist in analysing user comments and feedback on various websites to assess goods and services in a variety of fields, including social networking, politics, healthcare and entertainment.

The final stage of opinion mining development shows a popular method called the lexical approach, which enhances opinion mining algorithms and applications and creates new features by utilising lexical resources such as WordNet, thesaurus and SentiWordNet. Multilingual and cross-lingual sentiment analysis were employed in this phase.

CURRENT TRENDS AND TECHNIQUES IN OPINION MINING

Opinion mining techniques or approaches are used to mine the reviews or opinions in a standard way. Some opinion mining techniques are:

Lexicon-Based Methods

Lexicon-based methods are a category of techniques in NLP and sentiment analysis that rely on predefined lists of words called lexicons to determine the sentiment or meaning of a piece of text. The method conducts opinion mining by identifying an opinion lexicon that examines the textual data. Lexicon-based methods can categorise opinions in two ways. The first one is a dictionary-based approach, and the second one is a corpus-based approach. These methods assume that each word in a lexicon has an associated sentiment score (e.g., positive, negative, neutral). The overall sentiment of a text is derived by aggregating the scores of the words it contains. This method does not require labelled datasets, so no training data is needed. This method is operated on rules and dictionaries instead of learning from data. Works of Lexicon-Based Methods are:

Lexicon Creation: A lexicon is created or selected. For example: SentiWordNet, Automatic Finn Intensity Nielsen (AFINN), Valence Aware Dictionary and Entiment Reasoner (VADER), Linguistic Inquiry and Word Count (LIWC).

Text Preprocessing: The input text is cleaned and tokenised into words.

Sentiment Scoring: In Sentiment scoring, each word in the text is in the lexicon, its score is added to a cumulative total and if words not in the lexicon are ignored.

Aggregation: It is the final statement where "Final sentiment = sum (or average) of all matched word scores" here, applying weighting, negation handling, intensifiers, etc., is optional (Zhoa Z. *et al.*, 2023).

In Lexicon-Based Methods there is no need for large training datasets and quick to implement and easy to use even performs well on short, informal texts (e.g., tweets) but in this method context sensitivity is limited, struggles with sarcasm, irony, or domain-specific languages and new or uncommon words are ignored means fixed vocabulary exists (Rajgor D. & Barot M., 2017)

MACHINE LEARNING-BASED METHODS

Machine Learning is a popular and useful technique in opinion mining because of its capacity to handle enormous volumes of data and its automatic implementation. (Rani M. & Sumathy S., 2017). Machine Learning-Based Methods refer **to algorithms** that solve specific problems by learning patterns from data, rather than being explicitly programmed with rules. These methods are widely used across various domains like image recognition, natural language processing, finance, healthcare, and more. The main types of Machine Learning (ML)-based methods are:

S**upervised Learning:** Supervised learning relies on labelled data for both training and testing documents, indicating that a target value (i.e., label) is essential for executing supervised learning (Singh, P. & Husain, M, 2014). Classification is a widely used supervised data mining method or opinion mining approach. A key issue with the supervised learning method lies in the amount and quality of the data. These algorithms may fail if the training data is either inadequate or biased. Supervised Learning's common algorithms are Logistic Regression, Support Vector Machines (SVM), Linear Regression, Decision Trees, Neural Networks, and k-Nearest Neighbours. Some applications of supervised learning are Spam detection, fraud detection, and image classification (Emam A. *et al.*, 2018).

Unsupervised Learning: Unsupervised methods involve a learning system for inputs, without any expected output values provided. It utilises unlabeled text data. Unlike supervised learning, there is no need for a target value (i.e., label) to perform unsupervised learning. This learning technique gains knowledge through observation. Clustering is a commonly employed technique in unsupervised learning (Rajgor D. & Barot M., 2017). Clustering groups together objects that share similar characteristics into a cluster or group. Common Algorithms used by this learning are k-Means Clustering, Hierarchical Clustering, Principal Component Analysis (PCA), and Autoencoders. Some applications of this learning are market segmentation, anomaly detection, data compression, etc.

Semi-Supervised Learning: A semi-supervised strategy is an innovative machine learning technique for extracting insights from opinions, especially useful in scenarios with limited labelled data, and it delivers promising outcomes (Dhokrat *et.al.*, 2015). The purpose of the semi-supervised learning method is to develop a classifier that integrates both labelled and unlabelled data. Self-training, co-training and label propagation techniques are used for that. This is mostly used in learning in medical diagnosis (limited labelled data) and speech analysis.

Reinforcement Learning: When we interact with an environment, receiving rewards or penalties at that time, we deal with reinforcement learning. This learning is used to learn optimal actions to maximise cumulative reward. This technique uses some core Elements like Agent, Environment, State, Action and Reward. Some popular algorithms are Q-learning, Deep Q Networks (DQN) and Policy Gradient Methods. Reinforcement learning is used in robotics, game playing (e.g., AlphaGo) and recommendation systems.

Deep Learning Methods

Deep learning has revolutionised the field of opinion mining by enabling systems to automatically learn features from data without manual engineering and to better capture complex linguistic patterns.

A more recent development in the broader subject of ML is deep learning, which has already been effectively applied to data mining. Artificial Neural Networks (ANNs) with more layers than conventional ANNs are typically used in deep learning models. Previously, traditional methods like Naive Bayes, SVM relied on manually engineered features and had limitations in understanding complex language patterns. Deep learning models overcome this by automatically learning features, capturing context and syntax, handling negation, sarcasm, and multi-word expressions better than traditional methods. Some methods used by deep learning methodologies are:

Recurrent Neural Networks (RNNs): These are good for sequential data, and they can capture dependencies between words over time. It is especially useful with Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) variants.

Convolutional Neural Networks (CNNs): This is originally used in vision, but is effective for text classification and helps learn local patterns (e.g., n-grams) via filters.

Attention Mechanisms: These allow the model to focus on the most relevant parts of the text, often paired with RNNs or Transformers.

Transformer-Based Models (BERT, RoBERTa, DistilBERT, GPT): These are retrained language models that understand language deeply, which are used for fine-tuning on sentiment datasets to achieve state-of-the-art performance. It can handle long-range dependencies and context better than RNNs or CNNs.

MULTIMODAL OPINION MINING

This is the process of analysing and extracting sentiments, emotions, or opinions from multiple types of data modalities — such as text, audio, images, and video — rather than relying solely on one (usually text). This allows for a richer and more accurate understanding of opinions, especially in contexts like social media, customer feedback, movie reviews, and video blogs (vlogs) (Zhoa Z. *et al.*, 2023). Applications of Multimodal Opinion Mining are social Media Monitoring (text + images + emoji), Customer Service (speech + text), Human-Robot Interaction, Movie/TV Reviews (video + subtitles + facial expressions), Healthcare (facial analysis + voice for emotion detection), etc. Some Multimodal Opinion Mining Techniques are:

Feature Extraction per Modality: We use NLP techniques like BERT, LSTM for text; Pitch, tone, MFCCs, energy for audio and Facial expressions, gestures using CNNs or OpenPose for visual.

Fusion Strategies: This involves three stages. The first one is early fusion, which combines raw features from all modalities early, late fusion, which performs separate analysis for each modality and then merges the results, and the last one is hybrid fusion, which is a mix of both early and late fusion (Zhoa Z. *et al.*, 2023).

Datasets and Benchmarks

The largest challenge in conducting research is gathering data, and tasks like sentiment analysis and opinion mining are too challenging due to the abundance of information available online and the difficulty of gathering that data and extracting an opinion from such vast amounts of data (Singh, P., & Husain, M., 2014). Therefore, we talked about some of the data sources that are available, as well as the techniques that are used to extract the sentiments and opinions from the given text. Numerous data source exists online, such as blogs, microblogs, online posts, News feeds, Forums, review sites, etc.

Data Collection: In this part, the information to be examined, text cleaning using text cleaning tools, and the data can be processed and made ready for analysis, examining the information and understanding the result.

Blogs: Blogs are simply a user's online journal or personal space where they may express their thoughts and opinions on any subject they want.

Online Reviews: On the internet, various review sites are available where you can examine online reviews of any product before buying it.

Microblogging: Microblogs allow users to share small pieces of content, including single images, brief phrases, or links to videos. These are the main reasons that microblogs are so popular.

Online Posts: In online posts, people share their own thoughts, opinions, images, videos, views, likes, dislikes, comments and other information about particular subjects.

Forums: A message board, often known as an Internet forum, is an online discussion platform where users can engage in discussions through posted messages.

APPLICATION OF OPINION MINING

Opinion mining is a powerful technique for extracting and analysing subjective information from text, and has a wide range of applications across various sectors.

Business and Marketing: Monitoring public opinion about a brand on social media, news stories, and review websites; and analysing consumer feedback and reviews to identify preferences, problems, and new trends (Rani M. & Sumathy S., 2017). In order to understand preferences, pain points, and developing trends, opinions are used to analyse customer evaluations and comments. Based on these reviews, areas for product and service improvement are identified, and the impact of advertising campaigns is assessed by examining public sentiment.

Customer Service: Understanding customer wants and expectations by examining feedback from several channels, as well as recognising customer emotions in support tickets, emails, and chat logs to prioritise and address issues efficiently (Laxmi L. *et al.*, 2025).

Social Media Analysis: Tracking new trends and interesting subjects on social media, keeping an eye out for possible crises and taking preventative measures, and examining competitor talks on social media to learn about their advantages and disadvantages.

Political Analysis: Anticipating election results by evaluating public sentiment reflected on social media and news reports, as well as examining public attitudes towards government policies and initiatives.

Financial Analysis: Examining news articles, social media content, and financial statements to anticipate fluctuations in stock prices and recognise possible threats to investments based on public opinion and news occurrences.

Healthcare: Examining social media and online discussion platforms for negative drug reactions and safety issues while gaining insight into patient experiences and satisfaction with healthcare services (Rajgor D. & Barot M., 2017).

Education: Analysing student feedback on courses, instructors, and the overall learning experience and evaluating the effectiveness of educational materials and resources.

By leveraging opinion mining techniques, businesses, organisations, and individuals can gain valuable insights, make informed decisions, and improve their overall performance.

Challenges in Opinion Mining

Some issues and challenges of opinion mining affect the results of opinion mining.

Domain Adaptation: The most significant obstacle that opinion mining and sentiment analysis encounter is the reliance on specific domains for sentiment terminology.

A particular feature set might achieve excellent results in one domain while performing poorly in another.

Fake Reviews and Bots: Both justifiable and spam content can be found on the internet (Seerat B. & Azam F., 2012). Before processing, this spam content should be removed for efficient sentiment classification. This can be accomplished by looking for duplicates, spotting outliers, and taking the reviewer's reputation into account (Rani M. & Sumathy S., 2017).

Disproportion of Opinion Mining Software: Opinion mining software is quite expensive, thus only large corporations and the government can currently afford it. It goes above and beyond what the average citizen would anticipate to ensure that everyone benefits. This should be accessible to everyone.

Integration of Opinion With Implicit and Behaviour Data: The opinion words must be integrated with implicit data for sentiment analysis to be successful. The true behaviour of sentiment words is determined by the implicit data.

Multilingual Opinion Mining: These factors make it difficult to accurately analyse opinions expressed in different languages and across different cultural contexts.

Blended Sentences: Blended Sentences means reviews that are positive in one context might be negative from another point of view. For example, if a client comments that a particular wireless earbuds' sound is amazing, it would be a favourable view. However, if the buyer complains that the sound is too loud, it would be a bad impression.

Way of Expressing the Opinion: People don't express their ideas in the same manner. Each person has their own unique opinion because everyone thinks and expresses themselves differently.

Use of Abbreviations and Short Forms: People using social media more frequently for conversation and to express their opinions through shortcuts or abbreviations Seerat B. & Azam F., 2012). The use of abbreviations, special symbols is increasing day by day, so finding opinions from that is too challenging. For example, Thx for Thanks, OMG for Oh my God, b4 for before, bcoz for because, NC for no comment, TY for thank you, etc.

Typographical Errors: Sometimes, typographical errors cause problems while extracting opinions. These errors can involve misspellings. They are a common form of human error, often occurring when someone types too quickly or isn't paying close attention.

Orthographic Words: To express their happiness and excitement, people use orthographic terms to convey their joy and enthusiasm words e.g. Word Toooo….. Cute…, I am Soooo Haappy, or if they are in a hurry, they stress the words, e.g. comeeee fasssssst I am waitttttnggg.

Natural Language Processing Overheads: The sentiment analysis tool was hampered by natural language overheads such as ambiguity, co-reference, Implicitness, inference, etc.

FUTURE DIRECTIONS

Real-Time Opinion Tracking: Real-time opinion tracking, also known as trading or prediction markets, is a real-time platform where users predict the outcome of future events and trade their opinions on those predictions (Seerat B. & Azam F., 2012). These platforms are gamified and allow users to express their views on various topics, like current events, sports, or even stock prices, and then buy or sell positions based on their analysis and the information available (Qiang Li *et al.*, 2024).

Sentiment-Aware Language Models: Traditional sentiment analysis methods often overlook the sensitivity to emotional changes in text, which limits their applicability in real-world scenarios. These days, a deep learning based Sentiment Aware Analysis

Model (SAAM) uses the potent representational capabilities of large-scale language models to tackle this problem.

Few-Shot and Zero-Shot Sentiment Analysis: Large Language Model (LLMs) can be adapted to new tasks in NLP using zero-shot and few-shot learning strategies, which need little to no training data. With Zero-shot learning, a model uses its previously learned information to complete a job without any prior examples. By giving the model a limited number of task-specific examples to instruct it, few-shot learning improves performance.

Ethical Concerns and Bias in Models: Ethical issues and biases present in AI models pose considerable challenges, arising from prejudiced training data, defective algorithms, and even human biases integrated into the system (Razali N. *et al*., 2021). Such biases may result in discriminatory results, reinforce societal disparities, and diminish public confidence in AI systems. Tackling these problems necessitates thorough attention to fairness, transparency, and accountability during the entire AI lifecycle.

Explainable Sentiment Analysis: Conventional sentiment analysis models, particularly those based on deep learning, often function as "black boxes," which complicates the process of discerning the reasons behind their classification of a text as positive, negative, or neutral. Explainable sentiment analysis, commonly referred to as Explainable Sentiment Analysis (XSA), is a domain dedicated to increasing the transparency and comprehensibility of sentiment analysis models. This allows users to understand the basis on which the model makes its sentiment predictions. XSA seeks to remedy this issue by offering explanations for sentiment predictions, thereby fostering trust and enabling users to detect biases or inaccuracies in the model's reasoning.

CONCLUSION

Opinion mining is an emerging area of data mining used to extract knowledge from a vast amount of customer reviews, comments and feedback and reviews on any product or product. The key insights in this paper history of opinion mining and highlight the current technologies and the vast potential of opinion mining as a research and application field. This report presents the various Opinion Mining techniques and datasets, and benchmarks. This paper summarises the issues and challenges of opinion mining that affect the results of opinion mining.

REFERENCES

Anbananthen, K. & Elyasir, A. (2013). Evolution of opinion mining, *Australian Journal of Basic and Applied Sciences, 7*(6): 359-370.

Dhokrat, A., Khillare, S., & Mahender, C. N. (2015). Review on techniques and tools used for opinion mining, *International Journal of Computer Applications Technology and Research, 4*(6), 419 - 424. https://doi.org/10.7753/IJCATR0406.1001

Emam, A. & Alzahrani, M. (2018). Opinion mining techniques and tools: A case study on an Arab newspaper, *International Conference on Computational Science and Computational Intelligence(CSCI)*, https://doi.org/10.1109/CSCI.2017.49

Laxmi, L., Ali, B. M., Ali, K. Dhana S., Rafiq, M., Shernazarov, I., Othman, N. & Khan, M. (2025). Opinion mining in e-commerce: Evaluating machine learning approaches for sentiment analysis, *Result in Control And Optimization 19*, https://doi.org/10.1016/j.rico.2025.100575

Qiang, Li, Zhao, F., Zheng, J., Xiang, H., Ouyang, H., & Zhao, L. (2024). Enhancing emotion perception in sentiment analysis through large-scale language models, *International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. https://doi.org/10.1109/ICNC-FSKD64080.2024.10702242

Rajgor, D., & Barot, M. (2017). Opinion mining: Classification, techniques, challenges, *International Journal of Current Engineering and Scientific Research (IJCESR)4*(7).

Rani, M., & Sumathy, S.(2017), Analysis on various machine learning based approaches with a perspective on the performance, *IEEE Innovations in Power and Advanced Computing Technologies* (IEEE Press, Vellore, 2017), pp. 1-7. https://doi.org/10.1109/IPACT.2017.8244998

Razali, N., Nur Atiqah., Malizan,, Nor Asiakin., Hasbullah., Muslihah, Wook., Norulzahrah Mohd Zainuddin., Khairul Khalil Ishak., Suzaimah Ramli., & Sazali Sukardi. (2021). Opinion mining for national security: Techniques, domain application, challenges and research opportunities, *Journal of Big Data 8*(150). https://doi.org/10.1186/s40537-021-00536-5 PMid: 34900516 PMCid: PMC8642766.

Seerat, B., & Azam, F. (2012). Opinion mining: Issues and challenges (A survey), *International Journal of Computer Applications, Foundation of Computer Science (FCS), NY, USA, 49*(9), https://doi.org/10.5120/7658-0762

Singh, P., & Husain, M. (2014). Methodological study of opinion mining and sentiment analysis techniques, *International Journal on Soft Computing (IJSC). 5*(1). https://doi.org/10.5121/ijsc.2014.5102

Zhoa, Z., Liu, W., & Wang, K. (2023). Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning, *Journal of Big Data 10,*155. https://doi.org/10.1186/s40537-023-00837-x