Exploring Sentiment Analysis: Applications, and Challenges —A Comprehensive Survey

Shwetha C H

Assistant Professor, Department of Computer Science & Engineering
Vivekananda College of Engineering and Technology, Puttur, Karnataka, India

shwethachethan22@gmail.com

---***---

Abstract - Sentiment analysis, also known as opinion mining, has emerged as a crucial field of research due to the exponential growth of user-generated content on various online platforms. This paper presents a comprehensive survey of sentiment analysis applications and challenges. It provides an overview of the different techniques employed for sentiment analysis, ranging from traditional machine learning approaches to more recent deep learning models. Furthermore, the survey examines the diverse applications of sentiment analysis across domains such as social media, e-commerce, customer reviews, and political analysis. Additionally, the paper highlights the major challenges and open research questions in sentiment analysis, including handling sarcasm, irony, and ambiguity, addressing data sparsity and imbalance issues, and ensuring cross-lingual and cross-domain generalization. By analyzing the existing literature, this survey aims to offer insights into the current state-of-the-art in sentiment analysis and provide directions for future research in this dynamic field.

Key Words: Sentimental Analysis, languages, text, multimode, opinion

1.INTRODUCTION

In recent years, sentiment analysis has garnered significant recognition and adoption, not only among researchers but also among businesses, governments, and organizations (Sánchez-Rada and Iglesias 2019). With the increasing prevalence of the Internet, it has become the primary source of global information. Countless users utilize various online platforms to express their viewpoints and opinions. To effectively monitor public sentiment and facilitate informed decision-making, it has become imperative to leverage user-generated data and analyze it automatically. Consequently, sentiment analysis has experienced a surge in popularity within research communities. This field is also referred to as opinion analysis or opinion mining. Recently, there has been significant growth in the field of sentiment analysis, with several research works focusing on various aspects of this task. Ligthart et al. (2021) published an overview of opinion mining in its early stages, while Piryani et al. (2017) discussed the topic from 2000 to 2015, proposing a framework for computationally processing unstructured data to extract opinions and identify their sentiments. Several surveys, such as those by Yousif et al. (2019) and Birjali et al. (2021), have provided insights into the challenges and potential directions in sentiment analysis. Research on sentiment classification, including the detection of opinion spam and fraudulent reviews, has been conducted by Soleymani et al. (2017), Yadav and Vishwakarma (2020), Yue et al. (2019), and Liu et al. (2012). Jain et al. (2021b) discuss the incorporation of online reviews in sentiment categorization, predictive decision-making, and the detection of false reviews using machine learning applications. Balaji et al. (2021) extensively examined various applications of social media analysis, employing sophisticated machine learning algorithms. Additionally, Hangya and Farkas (2017) provide a concise overview of the machine learning algorithms utilized in social media analysis.

The rise of social network sites has given rise to various fields dedicated to analyzing these networks and extracting valuable information from their content. One such field is sentiment analysis, which focuses on determining the sentiments expressed in text based on its content. As a subfield of natural language processing (NLP), sentiment analysis has garnered significant attention for its potential in informing decision-making based on public opinion. While there have been early works addressing sentiment analysis, it continues to evolve and develop even in the new millennium.

Sentiment analysis finds relevance in various real-world applications, such as product analysis, where it assists in identifying the components or qualities of a product that appeal to customers in terms of product quality. Subhashini et al. (2021) present a comprehensive review of contemporary literature on opinion mining, covering topics such as extracting text features from opinions with noise or uncertainty, representing knowledge in opinions, and categorizing them. Mowlaei et al. (2020) propose a technique for adaptive aspect-based lexicons for sentiment classification. They describe two strategies, one based on statistics and genetic algorithms, for constructing dynamic lexicons that assist in sentiment classification based on different aspects.

In the domain of reviews, sentiment analysis has been applied to various sectors, including hotels, airlines, healthcare, and the stock market (Zvarevashe and Olugbara 2018). Hotel reviews have been analyzed to gain a better understanding of customer preferences and dislikes. In the case of the stock market and cryptocurrencies, sentiment analysis has been utilized to determine market trends based on sentiment analysis (Valencia et al. 2019). Ahmad et al. (2019a) analyze sentiments of tweets across different domains. In the healthcare domain, sentiment analysis has seen increased application, including customer opinion analysis, customer satisfaction analysis, and other healthcare-related applications (Ruffer et al. 2020; Park et al. 2020; Cortis and Davis 2021; Arora et al. 2021; Baashar et al. 2020; Miotto et al. 2018).

The business sector has long leveraged sentiment analysis for various purposes, including reputation management, market research, competitor analysis, product analysis, and understanding customer voices (Ahmad et al. 2019a). Sentiment analysis continues to play a vital role in improving business operations and decision-making processes.

Sentiment analysis and natural language processing encounter several challenges, including informal writing styles, sarcasm, irony, and language-specific issues. Informal writing styles used by individuals can make sentiment analysis more challenging due to the presence of slang, abbreviations, and unconventional grammar. Sarcasm and irony pose significant difficulties as they involve the opposite or different meaning from what is explicitly stated.

Language-specific challenges arise due to variations in word meanings and orientations across different languages, contexts, and domains. Limited availability of tools and resources for sentiment analysis in all languages further exacerbates these challenges. Researchers have been actively addressing the detection of sarcasm and irony in text, which are considered critical challenges in sentiment analysis.

This work aims to analyze the various challenges associated with sentiment analysis, as well as the methodologies, applications, and algorithms employed in this field. Comparative data analysis will be presented using tables, flow charts, and graphs to provide a clear understanding of the task and its complexities. These visual representations will aid in simplifying the comprehension of the findings and insights derived from the analysis.

2. SENTIMENTAL ANALYSIS LEVEL

Sentiment analysis can be categorized into several levels based on the granularity of the analysis. These levels help in understanding the depth and scope of sentiment analysis. The common levels of sentiment analysis include:

A) Document-level Sentiment Analysis: At the document level, sentiment analysis aims to determine the overall sentiment expressed in a piece of text, such as an entire document, article, or review. It provides a high-level understanding of whether the sentiment is positive, negative, or neutral.

B) Sentence-level Sentiment Analysis: Sentence-level sentiment analysis focuses on analyzing the sentiment expressed within individual sentences. It aims to determine the sentiment polarity of each sentence, classifying it as positive, negative, or neutral. This level of analysis provides more granular insights into the sentiment within a document.

C) Aspect-based Sentiment Analysis: Aspect-based sentiment analysis delves deeper by identifying and analyzing sentiments related to specific aspects or features of a product, service, or topic. It aims to understand the sentiment associated with each aspect separately. For example, in a product review, aspect-based sentiment analysis can identify sentiments for aspects like price, usability, customer service, etc.

D) Entity-level Sentiment Analysis: Entity-level sentiment analysis focuses on determining the sentiment towards specific entities mentioned in the text, such as people, organizations, products, or brands. It identifies and analyzes the sentiment associated with each entity separately.

E) Fine-grained Sentiment Analysis: Fine-grained sentiment analysis goes beyond simple positive, negative, or neutral classifications. It aims to assign sentiment scores or intensities to express a more nuanced sentiment analysis. This level of analysis provides a detailed understanding of the varying degrees of sentiment, capturing subtle nuances and shades of opinion.

These levels of sentiment analysis provide different perspectives and insights, allowing for a more comprehensive understanding of the sentiments expressed in text data. The choice of the appropriate level depends on the specific analysis objectives and requirements of the application.

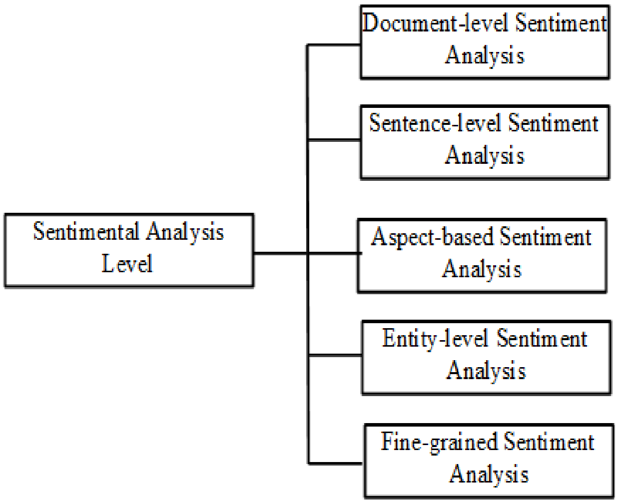

Fig 1: Levels of Sentimental Analysis

3.DATA COLLECTION AND FEATURE SELECTION

i. Data Collection

Data Collection in Sentiment Analysis: Data collection is a crucial step in sentiment analysis as it involves gathering the necessary text data for analysis. The data collected should be representative of the target domain or topic and should cover a wide range of sentiments to ensure the effectiveness of the sentiment analysis model. There are several approaches to data collection:

1. Manual Data Collection: In this approach, human annotators manually label text data with sentiment labels. This can be a time-consuming process, but it allows for fine-grained sentiment annotations.
2. Crowdsourcing: Crowdsourcing platforms like Amazon Mechanical Turk can be utilized to collect sentiment-labeled data from a larger pool of annotators. This approach helps in collecting a large volume of data more quickly, but it may have varying levels of annotation quality.
3. Pre-existing Datasets: Researchers can also use pre-existing datasets that have been previously labeled for sentiment analysis. These datasets are often publicly available and cover a wide range of domains, such as movie reviews, product reviews, social media posts, etc.

ii. Feature Selection

Feature Selection in Sentiment Analysis: Feature selection involves identifying the most relevant and informative features from the text data to represent sentiment. Effective feature selection is essential for improving the accuracy and efficiency of sentiment analysis models. Some common feature selection techniques include:

1. Bag-of-Words (BoW): The BoW approach represents text data as a collection of individual words or tokens. Features are selected based on the frequency of occurrence in the text corpus. Stop words (common words like "the," "is," etc.) may be excluded during feature selection.

2. n-grams: n-grams are sequences of n consecutive words in a text. They capture the context and ordering of words. Commonly used n-grams include unigrams (single words), bigrams (two consecutive words), and trigrams (three consecutive words).
3. Lexicon-based Features: Lexicons or sentiment dictionaries contain pre-defined lists of words with associated sentiment scores. Lexicon-based features involve matching words from the text to the sentiment lexicon and using the sentiment scores as features.
4. Word Embeddings: Word embeddings capture the semantic meaning of words by representing them as dense vectors in a high-dimensional space. Popular word embedding techniques like Word2Vec, GloVe, or FastText can be used to extract meaningful features.
5. Advanced Techniques: Advanced feature selection techniques, such as TF-IDF (Term Frequency-Inverse Document Frequency), word frequencies, and statistical measures like chi-square or mutual information, can be applied to select informative features.

The choice of feature selection technique depends on the nature of the data, the complexity of the sentiment analysis task, and the available computational resources. It is essential to experiment with different feature selection methods to find the most effective combination for accurate sentiment analysis.

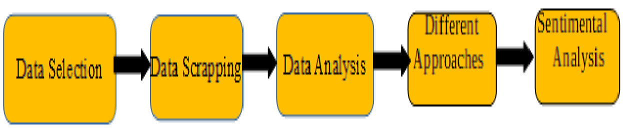

Fig 2: Sentimental Analysis Process

4. NEED OF SENTIMENTAL ANALYSIS

Sentiment analysis is highly valuable as it enables businesses to gain insights into how their customers perceive their brand. By automatically categorizing the emotions expressed in social media interactions, reviews, and other forms of data, organizations can make well-informed decisions. Sentiment analysis encompasses various methods and strategies that allow companies to analyze the sentiments of their customer base regarding a specific product or service. It helps in identifying the polarity of emotions (positive, negative, or neutral), determining the subject of discussion, and recognizing the entities expressing the sentiments. By leveraging sentiment analysis, businesses can effectively understand and respond to customer feedback, leading to improved customer satisfaction and informed decision-making. Sentiment analysis has become increasingly important due to the following reasons:

1. Customer feedback: Sentiment analysis helps businesses understand and analyze customer feedback, whether it's from social media, online reviews, surveys, or customer support interactions. By determining the sentiment behind customer opinions, companies can identify areas of improvement, address customer concerns, and enhance their products or services accordingly.
2. Brand reputation management: Sentiment analysis enables organizations to monitor and manage their brand reputation. By analyzing the sentiment expressed in social media posts, news articles, or online discussions, businesses can identify potential issues, respond to negative sentiment promptly, and take proactive measures to protect and enhance their brand image.
3. Market research and competitor analysis: Sentiment analysis allows companies to gain insights into customer preferences, opinions, and trends. By analyzing sentiment across different demographics or customer segments, businesses can identify emerging market trends, understand customer sentiment towards their own products as well as those of competitors, and make informed decisions regarding product development, marketing strategies, and market positioning.
4. Political analysis: Sentiment analysis is used in political campaigns and governance to gauge public opinion, monitor voter sentiment, and assess the effectiveness of political messages. By analyzing sentiment expressed on social media, news articles, or public forums, political organizations can adjust their messaging, focus on key issues, and understand the sentiments of voters.
5. Customer service and support: Sentiment analysis can be used in customer service and support systems to prioritize and categorize incoming customer queries or complaints based on their sentiment. This helps companies identify urgent issues, provide timely responses, and allocate resources effectively to address customer concerns.
6. Product feedback and improvement: Sentiment analysis helps businesses gather feedback on their products or services from various sources, including customer reviews, surveys, and online discussions. By analyzing sentiment, companies can identify patterns, uncover areas for improvement, and make data-driven decisions to enhance their offerings.

Overall, sentiment analysis is a valuable tool for understanding and harnessing the power of human emotions expressed in text. It enables businesses, organizations, and governments to gain insights, make informed decisions, and take proactive measures to improve their products, services, and overall customer experience.

5. APPLICATIONS OF SENTIMENT ANALYSIS

Sentiment analysis is a versatile technique with a wide range of applications across various domains. It can be used to analyze customer opinions, assess patient mental health based on social media posts, and much more. Moreover, advancements in technologies like Blockchain, IoT, Cloud Computing, and Big Data have further expanded the potential applications of sentiment analysis, making it applicable in almost any discipline. These technological advancements enable sentiment analysis to handle large-scale data, ensure data security, leverage real-time data streams, and utilize cloud-based infrastructure. As a result, sentiment analysis can be integrated into diverse fields, allowing for data-driven insights and decision-making in practically any industry or domain.

a. Product Review

With the rapid growth of e-commerce, there has been a significant increase in the number of products being sold and customer reviews being generated. Analyzing the sentiment expressed in these reviews can greatly assist customers in making informed decisions about their purchases (Paré, 2003). Sentiment analysis can be conducted at either the phrase level or the aspect level, as proposed by Schouten and Frasincar (2015). By performing sentiment analysis on product reviews, businesses can gain insights into customer opinions regarding their latest products, whether it be after launch or through the examination of comments and reviews. To focus on specific aspects of a product, such as food quality, service, or cleanliness, relevant keywords can be selected. By training a sentiment analysis framework, as suggested by Mackey et al. (2015), it becomes possible to identify and analyze the necessary information pertaining to these chosen features.

b. Business analysis

Sentiment analysis in the realm of business intelligence offers numerous advantages, enabling firms to leverage the data for product improvement, customer feedback analysis, and innovative marketing strategies. The primary application of sentiment analysis in business intelligence involves examining customers' perceptions of products or services. However, this practice is not exclusive to product producers; consumers can also employ sentiment analysis to review items and make well-informed decisions. The benefits of sentiment analysis in business intelligence are manifold. For instance, organizations can utilize the insights gained to enhance their products, evaluate customer feedback, and formulate effective marketing plans (Han et al., 2019). While sentiment analysis is commonly employed to analyze customers' perceptions of products or services, it extends beyond the purview of product manufacturers. Consumers themselves can leverage sentiment analysis to compare various products and make informed choices. In a study conducted over an eight-year period, Bose et al. (2020) analyzed food service reviews on Amazon.com, employing an emotion lexicon to classify sentiments into eight different emotions and two moods (positive and negative). Their findings revealed that sentiment analysis can identify customer behaviors and potential risks, ultimately leading to increased customer satisfaction.

c. Review analysis

The field of entertainment extensively relies on sentiment analysis. It is commonly used to analyze reviews of movies, shows, and short films in order to gauge the viewers' responses (Kumar et al., 2019). This not only assists viewers in making informed choices but also aids in the recognition and popularity of high-quality content. Sentence-level sentiment analysis (Lin and He, 2009) is widely employed in this domain to accurately determine the overall sentiment expressed in the reviews.

Similarly, the travel industry has been actively working on enhancing customer experiences by implementing machine learning and online consumer recommendation systems that leverage intelligent, data-driven decision-making techniques (Jain et al., 2021f). Furthermore, Jain et al. (2021e) discussed the classification of consumer decisions as positive or negative based on online reviews provided by valuable consumers. This approach helps the travel industry gain insights into customer feedback and improve their services accordingly.

d. Customer reviews

Sentiment analysis of hotel and restaurant reviews can be valuable for both customers and owners, aiding in informed decision-making and business improvement. Aspect-based sentiment analysis specifically focuses on identifying the aspects within reviews that receive the most positive or negative feedback. This analysis allows hotels and restaurants to target areas for improvement, benefiting both their reputation and customer satisfaction (Sann and Lai, 2020; Al-Smadi et al., 2018).

The application of sentiment analysis in the hospitality industry is particularly attractive due to its potential benefits. By analyzing reviews, customers can make better choices based on the sentiment expressed by previous guests. Additionally, owners and service providers can gain insights into specific aspects that garner positive or negative sentiments through ABSA (Akhtar et al., 2017). This enables them to focus on areas that require attention and make necessary improvements.

Ultimately, service providers stand to gain the most from sentiment analysis as they can identify the aspects receiving negative feedback and take steps to enhance those areas. By addressing customer concerns and improving upon shortcomings, hotels and restaurants can enhance customer experiences, boost their reputation, and ultimately increase profitability (Zhao et al., 2019).

e. Social media monitoring

Social data sentiment analysis enables continuous monitoring of customer sentiment around the clock, seven days a week, in real-time. This allows businesses to promptly respond and strengthen their brand image when negative discussions arise, while also capitalizing on positive mentions. Furthermore, this analysis provides a consistent and reliable source of information about customers, enabling businesses to track their progress over different seasons and make informed decisions based on the data.

Social media posts often present some of the most candid perspectives on products, services, and companies since individuals freely express their thoughts without being prompted. Customers feel compelled to share their feelings with the rest of the world, making social media a rich source of honest feedback.

6. CHALLENGES IN SENTIMENT ANALYSIS

Sentiment analysis, despite its valuable applications, faces several challenges that can impact the accuracy and effectiveness of the analysis. Here are some brief descriptions of the main challenges in sentiment analysis:

Subjectivity and Ambiguity: Sentiment analysis involves dealing with subjective language, which can vary from person to person. Understanding the nuances and different levels of sentiment expressed in text can be challenging. Ambiguity in language, such as sarcasm, irony, or figurative expressions, further complicates accurate sentiment analysis.

Contextual Understanding: The meaning of words or phrases can change depending on the context in which they are used. Sentiment analysis models need to accurately capture the contextual information to avoid misinterpretation and improve sentiment classification.

Data Noise and Preprocessing: Text data collected for sentiment analysis often contains noisy elements, such as spelling mistakes, abbreviations, slang, or grammatical errors. These noise elements can affect the performance of sentiment analysis models and require effective preprocessing techniques to handle them.

Domain and Genre Adaptation: Sentiment analysis models trained on one domain may not perform well when applied to a different domain or genre. Adapting sentiment analysis models to specific domains or genres, such as product reviews, social media posts, or news articles, is a challenge that researchers aim to address.

Handling Negation and Contrast: Negation words and phrases can change the sentiment polarity of a sentence. Recognizing negations and their impact on sentiment is essential for accurate sentiment analysis. Similarly, handling contrasting statements or conflicting sentiments within a text is another challenge that needs to be addressed.

Data Imbalance: Sentiment analysis datasets often suffer from class imbalance, where one sentiment class has a significantly higher number of samples compared to others. This can lead to biased models that perform poorly on minority sentiment classes.

Language and Multilingual Analysis: Sentiment analysis becomes more complex when dealing with multiple languages. Each language may have its own sentiment expressions, idiomatic phrases, and cultural nuances that require language-specific sentiment analysis techniques.

Lack of Annotated Data: Training accurate sentiment analysis models requires a significant amount of annotated data. However, creating large-scale annotated datasets is time-consuming and expensive, making it challenging to have diverse and representative training data for sentiment analysis.

Evaluation Metrics: Choosing appropriate evaluation metrics to assess the performance of sentiment analysis models is a challenge. Common metrics such as accuracy, precision, recall, and F1-score may not capture the nuances of sentiment analysis, leading to limited insights into model performance.

Ethical and Bias Concerns: Sentiment analysis models can inherit biases present in the training data, leading to unfair or discriminatory predictions. Addressing ethical concerns and ensuring fairness in sentiment analysis models is an important challenge that researchers are working on.

These challenges provide opportunities for further research and improvement in sentiment analysis techniques and methodologies. Researchers continue to explore innovative approaches to tackle these challenges and enhance the accuracy and applicability of sentiment analysis in various domains.

REFERENCES

[1]. Ahmad S, Asghar MZ, Alotaibi FM, Awan I (2019) Detection and classification of social media-based extremist affiliations using sentiment analysis techniques

[2]. Ahmad SR, Bakar AA, Yaakub MR (2019) A review of feature selection techniques in sentiment analysis

[3]. Akhtar N, Zubair N, Kumar A, Ahmad T (2017) Aspect based sentiment oriented summarization of hotel reviews

[4]. Balaji T, Annavarapu CSR, Bablani A (2021) Machine learning algorithms for social media analysis: a survey.

[5]. Behdenna S, Barigou F, Belalem G (2018) Document level sentiment analysis: a survey.

[6]. Annett M, Kondrak G (2008) A comparison of sentiment analysis techniques: Polarizing movie blogs. In: Conference of the Canadian Society for Computational Studies of Intelligence

[7]. Arora A, Chakraborty P, Bhatia M, Mittal P (2021) Role of emotion in excessive use of twitter during COVID-19 imposed lockdown in India.

[8]. Baashar Y, Alhussian H, Patel A, Alkawsi G, Alzahrani AI, Alfarraj O, Hayder G (2020) Customer relationship management systems (CRMS) in the healthcare environment: a systematic literature review

[9]. Duric A, Song F (2012) Feature selection for sentiment analysis based on content and syntax models.

[10]. Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining.

[11]. Eke CI, Norman AA, Shuib L, Nweke HF (2020) Sarcasm identification in textual data: systematic review, research challenges and open directions.

[12]. Bai X (2011) Predicting consumer sentiments from online text.

[13]. Birjali M, Kasri M, Beni-Hssane A (2021) A comprehensive survey on sentiment analysis: approaches, challenges and trends.

[14]. Bose R, Dey RK, Roy S, Sarddar D (2020) Sentiment analysis on online product reviews. In: Information and communication technology for sustainable development.

[15]. ***Cambria E, Das D, Bandyopadhyay S, Feraco A (2017) Affective computing and sentiment analysis. In: A practical guide to sentiment analysis.***