International Journal of Electronics, Communication & Soft Computing Science and Engineering
ISSN: 2277-9477, Volume 4, Issue 6

A Comprehensive Review of Sentiment Analysis: Techniques, Approaches, Applications and Challenges

Mr. Vaibhav Gajananrao Ingole
Lecturer in Information Technology, Government Polytechnic, Amravati, India-444603
Email: - vingole42@gmail.com

Received on: 12 Jan 2019 **Revised on: 15 Feb 2019** **Published on: 18 Feb 2019**

***Abstract* --** Sentiment analysis, or opinion mining, focuses on detecting, extracting, and categorizing subjective information within textual data. This includes opinions, attitudes, and emotions. One of the main drivers of the surge in sentiment analysis over the last two decades has been the emergence of social media, review websites, blogs, and online discussion forums. Each of these contains an abundance of data composed of interpersonal opinions and have created a significant need for automated systems to analyze sentiments. Initially, sentiment analysis focused on lexicon-based and classical machine learning methods. Recently, there has been a shift toward representations and architectures based on deep learning. It examines core sentiment analysis tasks at the document, sentence, and aspect level, as well as analysis of sentiments within social media. It reviews lexicon-based methods, along with sentiment analysis using Naïve Bayes, Support Vector Machines, and Maximum Entropy, as well as recursive and convolutional neural networks and distributed document representations. It analyses challenges including negation, sarcasm, ambiguity, domain adaptation, multilingual text, code-mixed languages, noisy social media data, and a lack of annotated data. It also identifies areas for future research focusing on code-mixed and multilingual sentiment analysis, where current methods are largely inadequate.

Keywords: Sentiment Analysis, Opinion Mining, Natural Language Processing, Machine Learning, Deep Learning, Aspect-Based Sentiment Analysis, Social Media, Text Classification

1. INTRODUCTION

The Internet and connectedness over the Web have transformed how and what people say. Customers have shifted how they communicate their experiences. They now post reviews, messages, blogs, and other content on social media and forums. Many of these posts contain opinions, attitudes, and feelings. These 'natural' forms of communication are difficult to process and represent a challenge for opinion extraction.

Sentiment analysis is the computation of opinions, feelings, attitudes, and emotions in written form. Pang and Lee position opinion mining and sentiment analysis as the computational treatment of opinion-oriented information, and emphasize its value for information access and decision making [1]. Liu and Zhang also position sentiment analysis as the computation of opinions, emotions, and attitudes towards entities and individuals, towards topics and issues, and towards events and phenomena [2].

In comparison to traditional information retrieval, the focus is on what the document is about, and in sentiment analysis, what the author thinks about that particular topic is of interest. A sentiment analysis system would have to understand that the phrase 'battery backup' is negative, and that 'camera quality' is positive, in the review "The camera quality is excellent, but the battery backup is disappointing."

Early work in sentiment classification was

essentially an extension of traditional text classification methods. Pang, Lee, and Vaithyanathan showed that Naive Bayes, Maximum Entropy, and Support Vector Machines could be used for sentiment classification and that classification of sentiment differed from conventional classification of topics.

After that, the focus of sentiment analysis was on classification of increasing granularity. This included detection of subjectivity, classification of sentiment at the level of sentences, extraction of aspects, identification of aspect-level polarities, and detection of opinion holders, as well as classification of emotions, comparative opinions, and opinion spam [1], [2].

With the focus of machine learning shifting to representation learning, sentiment analysis also shifted from reliance on engineered features. Word embeddings learned via Word2Vec captured semantic relationships between words and initiated a shift toward learning sentenc eembeddings [5]. Further improvements in the area of representation learning for sentiment classification came from Recursive Neural Tensor Networks and Convolutional Neural Networks [7], [8].

We attempt to capture the major advancements in sentiment analysis up to 2018, and identify core approaches, assess their performance, and elaborate on the challenges that remain unsolved.

2. EVOLUTION OF SENTIMENT ANALYSIS

2.1 Early Sentiment Classification

Pang, Lee, and Vaithyanathan's 2002 study is pioneering in this field. The authors examined sentiment classification with movie reviews and compared Naïve Bayes, MaxEnt, and SVMs. Unlike topic classification, which can be done comparatively easily with machine learning methods, sentiment classification relies on much more context and nuance. Pang, Lee, and Vaithyanathan's study showed that positive and negative reviews could be classified with existing machine learning methods, but that sentiment classification is more challenging than topic classification. This study acknowledged and advanced the field of sentiment analysis within NLP

2.2 Development of Opinion Mining

Pang and Lee gave a detailed survey of opinion mining and sentiment analysis in 2008 [1]. Their study encompassed sentiment classification, subjectivity, opinion summarization, resources, and ancillary applications. Their survey also delineated several important concepts: subjective vs. objective text, positive vs. negative sentiment, document vs. sentence-level sentiment, opinion expression vs. implied opinion, and opinion giver vs. opinion recipient.

Sentiment Analysis Tasks

Sentiment analysis encompasses multiple related tasks, not a single classification problem, as previously assumed.

2.3 Document-Level Sentiment Analysis

Document-level sentiment analysis involves classification of entire documents. For example, the document “This laptop is excellent. The performance is outatting and the display is beautiful.” would be classified as positive.

$S(d) \in \{\text{Positive, Negative, Neutral}\}$

where S(d) denotes the sentiment assigned to document d.

Document-level analysis is well-suited for product reviews because a review typically expresses an overall opinion about a product. This method can be misled, however, by reviews containing multiple opinions, for example, “The camera takes excellent photos, but its battery life is poor.”

2.4 Sentence-Level Sentiment Analysis

Sentence-level sentiment analysis assigns sentiment to individual sentences. Analyzing the following sentence, “The movie is not good,” should not assign a positive sentiment because the word “good” is used. An effective system should account for negation, contrast, and other sentence-level phenomena.

2.5 Aspect-Based Sentiment Analysis

Aspect-Based Sentiment Analysis (ABSA) focuses on sentiment toward specific aspects of an entity. Consider the sentence, "The phone has an excellent camera but a poor battery life." The expected output is Camera: Positive and Battery: Negative. SemEval-2014 Task 4 set a foundational benchmark for aspect-based sentiment analysis in the domains of laptop and restaurant reviews [9].

2.6 Social-Media Sentiment Analysis

Social media posts are often short, informal, and noisy. A satisfactory analysis of sentiment in social media posts should account for hashtags, emojis, misspellings, slang, abbreviations, URLs, and user mentions. Go, Bhayani, and Huang used the concept of distant supervision to classify sentiment on Twitter using emoticons as noisy labels [10].

3. Major Approaches to Sentiment Analysis

Sentiment analysis approaches can be categorized as lexicon-based, traditional machine-learning, and deep-learning methods.

3.1 Lexicon-Based Approach

A lexicon-based approach to sentiment analysis relies on a dictionary of words or sentences predefined to contain sentiment polarities. A simple sentiment score can be calculated as: $\text{Score}(T) = \Sigma\ \text{Polarity}(w_i)$ where w_i is a sentiment-bearing word.

Positive, negative, and neutral classes are assigned based on whether Score(T) is greater than, less than, or equal to zero, respectively. This model lacks sophistication to capture the nuances of language. SO-CAL, introduced by Taboada et al., accounts for polarity, sentiment strength, negation, and intensification.

This model has several advantages including the need for small, unlabeled datasets, simplicity, low computational cost, real-time processing, and interpretability. It has "sins of inclusion" including dependence on the domain, inability to decipher sarcasm and other forms of context, limited vocabulary, and difficulty with code-mixing and multilingual text.

4. Machine-Learning Approaches

In machine-learning approaches to sentiment analysis, sentiment patterns are learned through examples. The general framework is Text → Preprocessing → Feature Extraction → Classification → Sentiment.

4.1 Naive Bayes

Naive Bayes is a classification framework with probabilistic foundations that is commonly used for text classification. For a document d and sentiment class c:

$$P(c|d) = P(d|c)P(c) / P(d)$$

The class with the highest posterior probability is selected. Naive Bayes has a strong reputation for text classification; however, the assumption of feature independence is rarely satisfied in natural language.

4.2 Support Vector Machine

Another traditional algorithm for sentiment classification that is very successful is SVM. With sentiment classification, the goal is to train the algorithm to identify the optimal boundary that separates the sentiment classes with the given training examples.

$$f(x) = w^T x + b$$

Where w is the weight vector and b is the bias. SVM can work with sparse text representations and is compatible with TF-IDF features. Pang et al. cited SVM as a successfully implemented traditional algorithm in their early experiments for sentiment classification.

4.3 Maximum Entropy

Also known as logistic regression in a binary case, Maximum Entropy calculates the class probabilities as:

$$P(c|x) = e^{(w^T x)} / \Sigma c' e^{(w^T' x)}$$

The model is useful when multiple linguistic features contribute to sentiment classification.

5. Feature Engineering

Traditional sentiment analysis systems rely heavily on feature engineering.

5.1 Bag-of-Words

The Bag-of-Words model represents a document based on word occurrence ignoring the order of words in a sentence.

5.2 TF-IDF

Term Frequency-Inverse Document Frequency assigns importance to terms based on their frequency in a document and their distribution across other documents.

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t)$$

$$IDF(t) = \log(N/df(t))$$

TF-IDF is commonly used with traditional classifiers such as SVM and logistic regression.

5.3 N-Grams

N-grams are sequences of words. Negation can be modeled using N-grams; Unigram: not; Bigram: not good; Trigram: not very good. Using higher order N-grams can lead to dimensionality explosion.

6. DEEP-LEARNING APPROACHES

Sentiment analysis was vastly improved with the introduction of deep learning due to models being able to learn representations from text with little to no preprocessing.

6.1 Word Embeddings

Word embeddings use vector notation to represent words. Mikolov and colleagues created Word2Vec to represent words in a way that captures their semantic and syntactic relations. Word embeddings include distributed representations, where words with similar meanings have similar vector representations. Compared to sparse Bag-of-Words vectors, distributed representations provide better semantic information.

6.2 Paragraph Vector

Le and Mikolov created a way to add distributed representations to sentences and documents, also known as Paragraph Vector, where the goal is to represent variable-length texts with fixed-length vector representations. Their work improved text classification and sentiment analysis.

6.3 Recursive Neural Networks

Socher and colleagues created a fine-grained sentiment annotated corpus, the Sentiment Treebank, and proposed a Recursive Neural Tensor Network for sentiment analysis. The corpus and the model focus on understanding the sentiment of constituent parts that make up larger phrases and sentences. The Sentiment Treebank was a significant offering in the field of sentiment research.

6.4 Convolutional Neural Networks

Kim applied Convolutional Neural Networks (CNNs) to sentence classification and demonstrated that simple CNN architectures can be effective for sentence-level classification. A standard architecture for a CNN-based sentiment model consists of Word Embeddings, Convolution, Pooling, and a Fully Connected Layer, with a Softmax activation function at the end. Local patterns such as “very good” and “poor customer service” can be learned by CNNs.

6.5 Recurrent Neural Networks

RNNs and LSTMs are well-suited for sequential text. Compared to a Bag-of-Words representation, an LSTM can capture context and relate words across a longer sequence.

7. Comparison of Sentiment Analysis Approaches

The following table summarizes the major approaches seen in the literature.

Approach	Features Data Context	Training InterpretabilityMajor

SVM TF-IDF/N-grams Required Medium Medium Feature engineering

Maximum Entropy Linguistic features Required Medium Medium Feature dependency

Word2Vec + ML Word embeddings Large corpus + labels Medium Low Static representations

CNN Learned embeddings Required Medium–High Low Limited long-range context

RNN/LSTM Sequential embeddings Required High Low Computational complexity

Recursive Neural Network Tree-based representations Required High Medium Requires parsing/tree structure

Aspect-based methods Aspect + sentiment features Usually required High Medium Aspect identification

8. Sentiment Analysis Resources and Datasets

Progress in sentiment analysis has been propelled by the development of benchmark datasets.

8.1 Movie Reviews

Movie reviews were among the first datasets used in sentiment classification. Pang et al. showed the efficacy of traditional machine learning methods with this dataset [3].

8.2 Stanford Sentiment Treebank

The Stanford Sentiment Treebank introduced the first fine-grained annotations of sentiment at the phrase level and facilitated research on sentiment at the tree level and beyond [7].

8.3 Twitter Data

Twitter data offers a large-scale domain for sentiment analysis. However, Twitter data is difficult to work with due to the short message length, informal language, spelling and grammar variations, and the use of slang, hashtags, URLs, and emojis. Go et al. performed distant supervision using emoticons as noisy labels [10].

8.4 SemEval Datasets

SemEval offers standardized tasks for sentiment analysis. SemEval-2014 Task 4 was the first major step in assessing aspect-based sentiment analysis for both extraction and classification of sentiment polarity [9].

9. Rule-Based Social-Media Sentiment Analysis

In their work, Hutto and Gilbert described VADER, a concise, rule-based system for sentiment analysis of social media text [12]. VADER is based on a lexicon for sentiment analysis augmented with grammar and syntax for sentiment analysis. It proved that sentiment analysis of social media can be approached with a set of linguistic rules.

10. Major Challenges in Sentiment Analysis

There are many remaining challenges in sentiment analysis, in spite of the advances made in this area of natural language processing (NLP).

10.1 Negation

Examples like good, not good, not bad, and never excellent, illustrate how sentiment can be reversed or modified by negation. Word-based approaches often fail to analyze such cases properly.

10.2 Sarcasm

Among all the challenges, identifying sarcasm is perhaps the most difficult. An example like “Great! Another two-hour delay,” contains a word that is positive, but the sentiment is obviously negative.

10.3 Context Dependence

The word "light" can be positive or negative depending on the context. For a laptop, "light" is a positive adjective, but for a sound system, "light" is a negative adjective. Therefore, a sentiment analysis system must have a context awareness.

10.4 Domain Dependence

Sentiment words can have different meanings in different domains. A model trained for one domain will likely not generalize to another domain.

10.5 Multilingual Sentiment Analysis

Sentiment analysis resources for languages other than English are still developing. There are various obstacles such as a lack of annotated datasets and sentiment lexicons, morphological complexity, different word order, and a lack of pretrained resources.

10.6 Code-Mixed Text

An example of code-switching is seen in the sentence, "Movie khup awesome aahe." Traditional monolingual sentiment analysis models will not be able to classify sentiment in code-mixed text due to the mixing of different languages and grammars.

10.7 Emoji and Emoticons

Emoji and other emoticons have been increasingly used to express sentiment in posts. Traditional NLP pipelines remove nonsynicious characters, which can lead to loss of sentiment in a post.

10.8 Class Imbalance

Datasets can contain a disproportionate amount of positive reviews. This can lead to misleading conclusions based on accuracy scores alone. In this case, precision, recall, and F1 score should be considered.

11. Applications of Sentiment Analysis

There are many uses of sentiment analysis in different fields such as e-commerce, monitoring and analyzing brands and customers, political and financial analysis, healthcare, and education.

12. Evaluation Metrics

The evaluation of sentiment analysis models mainly depends on accuracy, precision, recall, and F1 score.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

In cases of class imbalance in sentiment classification, macro-averaged F1 score is more informative than accuracy.

13. Research Gaps Identified from the Literature

The literature reviewed has identified important research gaps including: limited resources for multilingualism, inadequate code-mixed sentiment resources, the difficulty of encapsulating context for sentiment, the need for aspect-level analysis, opaque deep models, domain adaptation, and unstructured social media data.

14. Future Research Directions

Subtopics for future research include multilingual sentiment models, Indian language sentiment analysis, code-mixed sentiment analysis, explainable sentiment analysis, hybrid models, and real-time sentiment analysis. A possible architecture is Text → Language Identification → Preprocessing → Lexicon Features + Embeddings → Neural Model → Aspect Detection → Sentiment Classification → Explanation.

CONCLUSION

Sentiment analysis has become a major research area including representation learning, deep neural networks, aspect-based analysis, and social-media analytics, having advanced beyond simple lexicon-based and traditional machine-learning techniques.

The initial studies in the area demonstrated the potential of machine learning for sentiment classification at the document level with Naïve Bayes, Maximum Entropy, and SVM algorithms. Lexicon-based methods showed the potential of sentiment classification with dictionaries and rules, especially in the absence of annotated data.

Deep learning and the introduction of distributed representations of words changed sentiment analysis systems dramatically. Representations of semantics and compositional structures became amenable to neural network architectures, including RNNs and CNNs. At the same time, finer-grained analysis of sentiment with respect to aspects of products, services, and entities became possible with aspect-based sentiment analysis.

There are a number of challenges, including negation, sarcasm, domain dependence, multilingual text, noisy social media, and code-mixed language. These include, but are not limited to, Indian languages with a relative lack of resources, as well as multilingual code-mixed communication.

REFERENCES

[1] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[2] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis," in Mining Text Data, C. C. Aggarwal and C. Zhai, Eds. Boston, MA, USA: Springer, 2012, pp. 415–463.

[3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification Using Machine Learning Techniques," in Proceedings of EMNLP, 2002, pp. 79–86.

[4] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," Computational Linguistics, vol. 37, no. 2, pp. 267–307, 2011.

[5] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and Their Compositionality," in Advances in Neural Information Processing Systems, 2013, pp. 3111–3119.

[6] Q. V. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," in Proceedings of the 31st International Conference on Machine Learning, PMLR, vol. 32, 2014, pp. 1188–1196.

[7] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," in Proceedings of EMNLP, 2013, pp. 1631–1642.

[8] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proceedings of EMNLP, 2014, pp. 1746–1751.

[9] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in Proceedings of the 8th International Workshop on Semantic Evaluation, 2014, pp. 27–35.

[10] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification Using Distant Supervision," CS224N Project Report, Stanford University, 2009.

[11] B. Schuller, A. E.-D. Mousa, and V. Vryniotis, "Sentiment Analysis and Opinion Mining: On Optimal Parameters and Performances," WIREs Data Mining and Knowledge Discovery, vol. 5, pp. 255–263, 2015.

[12] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, 2014, pp. 216–225.