CONTEMPORARY TRENDS IN RESEARCH ON SENTIMENT ANALYSIS

Abstract

Sentiment analysis research has been started in the 2000s and now it is one of the most demanding fields in research. Sentiment analysis is a rapidly expanding domain in Natural Language Processing that seeks to derive subjective data from text. Sentiment analysis has gained attention in recent years for understanding human emotion or opinion expressed in the form of textual information. Over the previous few years researchers and practitioners, they talked about sentiment analysis with the help of deep learning models and machine learning algorithms. Sentiment analysis is divided into three stages aspect level, sentence level, and document levels. Sentiment analysis is spread in every domain such as political fields, marketing, and business purposes, and the most commonly used in social media. This book chapter is all about a thorough overview of sentiment analysis, the level of sentiment analysis, applications, methods, and techniques.

Keywords: Sentiment analysis, Algorithms, Natural Language Processing

Author

Silvia Das
Department of Computer Science & Engineering
RCC Institute of Information Technology
Kolkata, India
silviadas2199@gmail.com

I. INTRODUCTION

One of the most famous research areas in computer science is sentiment analysis. The definition of sentiment analysis can be defined as the ability to understand the emotions of users. Sentiment analysis is additionally known as opinion mining, emotion analysis, and opinion extraction. Opinion mining is the method of analyzing users' sentiments, opinions, and emotions from text data. Sentiment analysis involves analyzing the context, phrases, and words in the text to identify whether the sentiment or opinion expressed is positive, negative, or neutral. It also considers the feelings that are happy, angry, sad, etc. Sentiment analysis is a subfield of Natural Language Processing (NLP) and it is the most powerful application of NLP. Natural Language Processing has been extensively studied in data mining and text mining. NLP was introduced in the early 1940s, At this time the task was not as easy as people imagined, and some researchers have identified major issues in the development field of NLP. Later on 1960s basic Natural Language Processing was successfully developed. NLP is the field of Artificial Intelligence (AI) that is used to extract information from data. The main goal of Natural Language Processing is to understand various languages, which are used to process and extract information from them. Sentiment analysis is the most powerful tool for the analysis of data. Sentiment is described as an attitude and more about feeling.

Sentiment Analysis can be classified into various levels. Some common types are:

1. **Sentiment Analysis Based on Sentence:** The opinion of each sentence within a document is analyzed, and each sentence's polarity is determined. Sentence-level sentiment analysis is more about subjectivity classification.

2. **Sentiment analysis Based on Document:** It implements opinion on a single entity. Document-level sentiment analysis involves the overall sentiment of a whole document. It can determine any other piece of text and then show whether the result is positive, negative, or neutral. This type of sentiment analysis focuses on reviews and articles.

3. **Sentiment Analysis Based on Aspect:** It aims to detect the sentiment associated with different aspects of a text. Aspect-based sentiment analysis breaks down the text into different aspects and determines whether the text is positive, negative, or neutral. For example, "The saree looks beautiful but the fabric is cheap" In this sentence two aspects are there one is the positive as the appearance and another is the material is negative. ABSA provides a more detailed analysis.

4. **Sentiment Analysis Based on Emotion:** However traditional sentiment analysis focuses on classifying text as positive or negative or neutral, and emotion-based sentiment analysis focuses on psychological conditions like happiness, sadness, anger, and fear, rather than positive or negative sentiment.

II. SENTIMENT ANALYSIS APPLICATIONS

Reviews are important to organizations as well as business purposes because they always want public opinions about products and services. Nowadays, organizations are widely using social media for decision-making. If a customer wants to buy the product online, they don't need to ask their family or friends' opinions because there are many user

reviews in public forums on the web about the particular product. In recent years, sentiment analysis applications have been spread in every domain like customer support and feedback analysis, political analysis, market research, healthcare, hotel and restaurant reviews, and online advertisement. Thousands of companies including both start-ups and MNCs now operate in this field.

1. **Customer Support and Feedback Analysis:** Sentiment analysis identifying the needs of customers' requirement and recommending the proper thing. Pare D. J [17] stated that sentiment analysis is very useful in e-commerce service areas for emerging countries. Later Mackey T. k, Miner A., and Cuomo R.E. [16] explore the sentiment analysis framework that can be trained to analyze the essential information.

2. **Political Analysis:** Sentiment analysis detects the political content. The content can be text, videos, and images.

3. **Market Research:** In market research sentiment analysis analyzes the customer review feedback to understand market trends.

4. **Healthcare:** Patient reviews and feedback to understand the quality of healthcare services provided by the healthcare centre. Sentiment analysis also helps diagnose mental health problems.

5. **Online Advertisement:** Advertisers can use sentiment analysis to ensure their desired sentiments are well-received by the audience.

6. **Hotel and Restaurant Reviews:** To improve customer experience by analyzing reviews and feedback from platforms like Google My Business, Zomato, The Infatuation, and Trip Advisor.

III. RELATED WORK

Many researchers have contributed to sentiment analysis. A brief discussion of the work done previously on sentiment analysis is given in this section.

1. "A technique for adaptive aspect-based lexicons for sentiment classification" by Moulaei ME, Abadeh MS, Keshavarz H [11]. The authors defined techniques for the construction of dynamic lexicons to aid in the classification sentiment relying on their aspects a method primarily based totally on statistic and genetic algorithm.
2. Subhasini L.D.C.S., Li Y, Z Hang J., Atukorale A.S, and Wu Y [10] present "the results of a comprehensive review of contemporary opinion mining literature. It also covers how to extract text features from opinions with noise or uncertainly represent knowledge in opinions, and categorize them".
3. "Sentiment analysis has gained widespread acceptance in recent years, not just among researchers but also among business, governments and organizations" proposed by Sanchez-Rada JF, Iglesias CA [3].
4. Lighthart A, Catal C, and Tekinerdogan B [1] published "an overview on opinion mining in the earlier stage".

5. Piryani R, Madhavi D, and Singh VK [2] discuss the study topic from 2000 to 2015 and provide "a framework for computational processing in structured data with the primary goal of extracting views and identifying their moods".
6. "A brief overview of machine learning algorithms used in social media analysis" presented by Hangya V, Farkas R [9].
7. Balaji T.k, Annavarapu C.S.R., and Bablani A. [8] in their research paper "conducted a thorough examination of the several applications of social media analysis utilizing sophisticated machine learning algorithms".
8. "The effectiveness of Internet reviews" proposed by Yue L, Chen W, Li X, Zuo W, and Yin M [6].
9. Jain PK, Pamula R, and Srivastava G [7] discuss "machine learning applications that incorporate online reviews in sentiment categorization, predictive decision-making and the detection of false reviews".
10. "The problem of sentiment analysis and suggested potential directions" described by Yousif A, Niu Z, Tarus JK, Ahmad A and, Birjali, M., Kasri, M., & Beni-Hssane, A. [4, 5].

IV. METHODS

There are several methods and approaches to perform on sentiment analysis from traditional rule-based methods to advanced machine learning approaches.

1. **Rule-Based Methods:** Involves the predefined rules to describe the sentiment expressed in a piece of text. Rule-based methods rely on linguistic patterns and can provide reasonable results for context. It is a straightforward approach.

2. **Lexicon-Based Methods:** Processes depend on the idea that the textual semantic orientation is associated with the polarity of words, this is related to content words, phrases, and sentences. Lexicon-based methods are simple to implement.

3. **Supervised Methods:** The most powerful approach in sentiment analysis is supervised methods. Supervised methods deal with a large number of data and complex language patterns and deliver high performance.

V. SENTIMENT ANALYSIS IS A SUBFIELD OF NATURAL LANGUAGE PROCESSING

Sentiment analysis is a Natural Language Processing method used to determine the emotional tone of a given text. Mostly it interacts between computers and human language. Present-day NLP is based on a machine-learning algorithm. Sentiment analysis using NLP is a very promising area of research as well as applications. NLP relies on machine learning and deep learning.

1. **Feature Selection:** The goal is to take a relevant feature from a data set. Ritter A, Etzioni O, and Clark S [14] they present "open domain event extraction from Twitter". Razon A, and Barnden J [15] in their work said "automated text readability classification based on concept indexing with Parts-Of-Speech n-gram features". N-grams are contiguous sequences of n items from text. The combination of two features is called Bi-gram and the

combination of three features is called Tri-gram. In sentiment analysis Bi-gram and Tri-gram can help classify text, language model, etc.

2. **Feature Extraction:** Transforming raw text data into a numerical representation (from unstructured text data into structured data) is the method of feature extraction. Venugopalan M and Gupta D [13] discuss "The challenges to extract features from the text". The most popular techniques of feature extraction are Bag of Words (BOW), Term Frequency-Inverse Document Frequency (TF-IDF), and Parts Of Speech Tagging (POS).

    - **Bag of Words** is the simplest method used in NLP, bag of words creates all unique words in the corpus. The bag represents the text data and each document is represented as a vector.
    - **Term Frequency and Inverse Document Frequency (TF – IDF)** an extension of the bag of words in a document is TF-IDF. Term frequency follows a higher weight to words that appear more frequently in a document. The term's presence provides the value of either 0 or 1.
    - **Parts Of Speech Tagging** involves information about grammatical parts of speech in a text. The tagging assigns tags to words based on their syntactic roles in a sentence.

VI. ALGORITHM USED IN NLP

To perform an NLP task the most two popular algorithms are Support Vector Machine (SVM) and Naïve Bayes, They have different approaches to handling text data such as sentiment analysis.

1. **Naïve Bayes:** The definition of this algorithm is to classify text data into a predefined form and observe different features in each class. The algorithm calculates the probabilities of a document that belongs to a particular class and then assigns the highest probability. Lopamudra Dey, Sanjay Chakraborty, Beepa Bose, Anuraag Biswas, and Sweta Tiwari [12] they worked on "movie review and hotel review using Naïve Bayes Classifier".

2. **Support Vector Machine:** Machine learning algorithm used for binary classification, SVM is designed to find optimal hyperplanes that separate data points belonging to different classes.

VII. CHALLENGES

Sarcasm detection is quite challenging, as the actual meaning of the words may communicate one sentiment while the intended meaning is the opposite.

1. **Unstructured Data** suffer a predefined format, unstructured data doesn't follow the pattern. So this is also challenging in sentiment analysis.
2. **Subjectivity in Text** Comparison in positive, negative, and neutral expressions with varying tones makes it difficult to determine the sentiment.

VIII. CONCLUSION

In this book chapter, we discussed the overview of sentiment analysis, methodologies, applications, different levels of sentiment analysis, and the different studies provided in research papers. Sentiment Analysis has proven a useful resource in every industry, especially in the business field.

REFERENCES

[1] Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*, 1-57.
[2] Piryani, R., Madhavi, D., & Singh, V. K. (2017). Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. *Information Processing & Management*, *53*(1), 122-150.
[3] Sánchez-Rada, J. F., & Iglesias, C. A. (2019). Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. *Information Fusion*, *52*, 344-356.
[4] Yousif, A., Niu, Z., Tarus, J. K., & Ahmad, A. (2019). A survey on sentiment analysis of scientific citations. *Artificial Intelligence Review*, *52*, 1805-1838.
[5] Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134.
[6] Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, *60*, 617-663.
[7] Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer science review*, *41*, 100413.
[8] Balaji, T. K., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, *40*, 100395.
[9] Hangya, V., & Farkas, R. (2017). A comparative empirical study on social media sentiment analysis over various genres and languages. *Artificial Intelligence Review*, *47*, 485-505.
[10] Subhashini, L. D. C. S., Li, Y., Zhang, J., Atukorale, A. S., & Wu, Y. (2021). Mining and classifying customer reviews: a survey. *Artificial Intelligence Review*, 1-47.
[11] Mowlaei, M. E., Abadeh, M. S., & Keshavarz, H. (2020). Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications*, *148*, 113234.
[12] Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. *arXiv preprint arXiv:1610.09982*.
[13] Venugopalan, M., & Gupta, D. (2015, August). Exploring sentiment analysis on twitter data. In *2015 eighth international conference on contemporary computing (IC3)* (pp. 241-247). IEEE.
[14] Ritter, A., Clark, S., & Etzioni, O. (2011, July). Named entity recognition in tweets: an experimental study. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 1524-1534).
[15] Razon, A., & Barnden, J. (2015, September). A new approach to automated text readability classification based on concept indexing with integrated part-of-speech n-gram features. In *Proceedings of the international conference recent advances in natural language processing* (pp. 521-528).
[16] Mackey, T. K., Miner, A., & Cuomo, R. E. (2015). Exploring the e-cigarette e-commerce marketplace: Identifying Internet e-cigarette marketing characteristics and regulatory gaps. *Drug and alcohol dependence*, *156*, 97-103.
[17] Paré, D. J. (2003). Does this site deliver? B2B e-commerce services for developing countries. *The Information Society*, *19*(2), 123-134.

