Literature Review of Text and Multimodal Sentiment Analysis

Yutong Li

School of Computing, Beijing Institution of Technology, Beijing, China
1120231924@bit.edu.cn

Abstract: Sentiment analysis, also known as opinion mining, is a crucial branch of Natural language processing, which focuses on recognizing, extracting, and quantifying sentiment tendencies, emotional intensity and specific emotion types in textual data. With the rapid development of the internet and communication, analyzing sentiment contained in textual data becomes important and crucial for understanding public opinion, consumer behavior, and emotional trends. This paper provides a comprehensive review of sentiment analysis in the range of its application, evolution, task types, methodology and future development by analyzing the literature of this field. Sentiment analysis has developed from traditional lexicon-based methods to modern deep learning methods like CNN, RNN and transformer model, which have significantly improved accuracy and robustness. This paper also discussed challenges in sentiment analysis like sarcasm detection and cross-lingual analysis, and proposed potential solutions. The findings aim to provide comprehensive insight for researchers and contribute to innovations in sentiment analysis.

Keywords: sentiment analysis, artificial intelligence, multimodal, literature review

1. Introduction

Sentiment analysis, also known as Opinion Mining, is an important branch of Natural language Processing (NLP). It aims to automatically recognize, extract and quantify sentiment tendency, emotional intensity and concrete emotion type. With the rapid growth of internet, user-generated content on social media platforms has become diverse and rich, leading to a circumstance that sentiment analysis is indispensable to understanding public opinion, consumer behavior and emotional trends. Its application has spanned a wide range of fields such as recommendation systems, healthcare and medical systems and marketing research.

In the early stages, scholars analyze sentiment by approaches based on statistics and probability and emotion dictionaries, like SentiWordNet and AFINN [1-2]. Nowadays, as artificial intelligence technology and large language models evolve rapidly, modern tools like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Transformer models like BERT have been used in sentiment analysis and have significantly improved the accuracy and robustness [3]. From another perspective, scholars originally take text only into consideration to analyze what sentiment is conveyed by the sentences. This article aims to provide a comprehensive overview of sentiment analysis by exploring its common methods, different task types and applications. By using literature analysis approach, this article highlights the evolution of the field, discusses current techniques, and identifies emerging trends and challenges with the hope of offering valuable insights and guidance for researchers and ultimately contributing to the continued development and innovation in sentiment analysis.

2. Introduction of sentiment analysis

Sentiment analysis is a process, technique or method to automatically determine the attitude or emotional predisposition of an opinion holder in a text to a certain topic.

2.1. Application of sentiment analysis

According to the definition, it can be easily detected that sentiment analysis has a wide range of applications in various fields.

Take recommendation systems for example, companies collect users' daily usage behavior and comments and then use sentiment analysis methods to detect whether the user likes or dislikes this video or advertisement so they can adjust different strategies to advertise to different users. Combined with big data technology, companies can easily draw users' paintings, and put proper advertisements accordingly. Another example is healthcare and medical system, on one hand, sentiment analysis helps medical workers to analyze patients' attitudes towards the curative effect and find latent requirements. On the other hand, sentiment analysis is very useful in psychological diseases. Doctors record dialogues between patients and doctors and then use sentiment analysis to estimate patients' mentation and suit the remedy to the case. In addition, by using sentiment analysis, the government can monitor public opinions and take prompt actions to keep the situation under control.

2.2. Importance of sentiment analysis

Sentiment analysis has significant importance in the world as the data has driven most parts of the digital world. It provides scholars the ability to decode unstructured textural data generated by daily usage. Sentiment analysis enables policymakers and researchers to gain actionable insights into public opinion by automatically identifying and quantifying emotions, opinions, and attitudes expressed in text. Apart from applications in society, sentiment analysis can also contribute to the development of artificial intelligence in natural language understanding.

As digital communication continues to grow, the ability to accurately and efficiently analyze sentiment will remain a critical tool for transforming raw data into meaningful insights, fostering better communication, and driving innovation across industries.

3. Task type of sentiment analysis

Sentiment analysis encompasses a wide range of task types that cater to different levels of granularity and application. These tasks range from basic polarity detection to analyzing how intensive the emotion is and from what emotion will the text express to what exactly the emotion is. These tasks represent different purposes and solve specific problems in order to understand human emotions and opinions. This article will explore these four task types mentioned above.

3.1. Polarity

Polarity classification is the most fundamental task in sentiment analysis and is widely studied. In a word, this task aims to categorize text into three aspects: positive, negative and neutral. It is essential for applications like product review research because it's critical to detect whether customers' feedback is favorable or unfavorable. In early stages, polarity classification relied on sentiment lexicons and a rule-based system [4]. Nowadays with the neural network development, machine learning and deep learning models like BERT (Bidirectional Encoder Representations from Transformers) and LSTM (Long Short-Term Memory) achieved higher accuracy [5].

3.2. Intensity

Intensity analysis seeks deeper into emotions by quantifying the strength of emotions expressed in text instead of providing only binary or ternary views of sentiment like polarity detection. This task uses a continuous range (like -1 to 1) or ordinal scale (like very negative, negative, neutral, positive, very positive) to represent emotional strength. It is usually used in situations where the degree of sentiment matters, such as measuring customer enthusiasm for a new product or the severity of complaints. Regression models and deep learning methods are common tasks used to estimate sentiment intensity.

3.3. Timing information

Timing information focuses on understanding how the emotion evolves over time. This task is usually combined with polarity detection methods to model temporal sentiment patterns. It’s crucial for applications like stock market prediction and brand reputation management. Because tracking sentiment trends can provide early warnings or opportunities for intervention. Common predictive models like recurrent neural networks (RNNs) and transformers are used to forecast future sentiment trends based on historical data.

3.4. Refined emotion

Refined emotion analysis goes beyond basic polarity and intensity to identify specific emotional states, such as joy, anger, sadness, fear, and surprise. It is usually used in mental health monitoring for timely intervention. It can detect signs of depression and anxiety on social media platforms to provide nuanced emotional expressions that can provide insights into an individual’s psychological state. Since this task requires more accuracy, refined emotion detection relies on emotion lexicons, such as the NRC Emotion Lexicon, or machine learning models trained on annotated datasets. In addition, deep learning approaches, such as multi-label classification models, have shown promise in capturing complex emotional nuances.

4. Common technology and methods of sentiment analysis

Thanks to the development of the theory of computation and the diverse data generated on the internet, sentiment analysis has evolved significantly over the years. There are various interesting methods employed in this field and are approximately divided into two categories: language-based methods and multimodal-based methods. This article explores these two primary methods and introduces their techniques, merits and demerits.

4.1. Language-based methods

Language-based methods focus on plain textual data to understand and interpret emotions and sentiments contained in the text. This kind of method has been widely studied and applied due to the abundance of text data from chatting platforms and media.

4.1.1. Rule-based method

The rule-based method relies on sentiment lexicons and grammatical rules, which are predefined. These lexicons distribute scores to each word like -1,1,0 representing negative, positive and neutral.

A document or sentence polarity is determined by calculating the scores of each word in it. Formula and algorithm can be applied in the process of computing the final score of the text. The rule-based method is practical for sentence and feature-level analysis. It is interpretable and requires no training data, which means an unsupervised approach. While it is more comprehensible to machine learning, the rule-based method depends heavily on manual work. In addition, the meaning of a word changes with the context. That is to say, a word like “high” can be positive when describing the salary and can be negative in the phrase “high and mighty”. So, the lexicon is lack of agility.

4.1.2. Machine learning method

Machine-learning method is often used for tasks of classification. This method requires a large scale of data. If these data are labeled, a supervised learning method can be used. Common supervised learning methods like support vector machine (SVM) and multiplayer perceptron (MLP) can be trained on the labeled data and do not need a lot of top-level design. If these data are unlabeled, an unsupervised learning method can be used. While the supervised method is mainly used in sentiment analysis, when the data is difficult to label or the cost is too much, scholars have no choice but to choose unsupervised learning. A common method in unsupervised learning is clustering. It divides data into different clusters where each cluster represents a similar point of view and therefore has the ability to classify the polarity of the text.

4.1.3. Deep learning method

Deep learning is developed on the basis of machine learning, including deep neural networks (DNN), convolutional neural networks (CNN) and recurrent neural networks (RNN). The term “deep” means that there is a large number of layers of perceptron in the network. By using a residual connection, deep learning has achieved great results. These models are unnecessary with the predefined lexicon or predesigned feature extraction methods to work, they can learn features and rules by themselves instead. But unlike rule-based methods, these models lack interpretability.

4.2. Multimodal-based methods

In the previous stage, sentiment analysis focuses on plain textural data to identify what emotions and opinions are expressed in the text. But in order to make artificial intelligence act like a human and judge sentiment more accurately, the AI should get access to other modality information like visuals or sounds. By integrating multiple pieces of information, this method can provide a more comprehensive understanding of emotion and sentiment.

4.2.1. Facial expression analysis

Countenance conveys a lot of information about people’s emotions (sad, happiness, anger and so on) and attitudes toward things. In previous research, scholars found that babies show various expressions without being taught [6]. This gives us reason to believe that facial expressions do express human sentiment. From probability models like the hidden Markov model (HMM) to neural networks like CNN, there are many methods to recognize what expression is expressed in a photo or video. By tracking the features of facial expressions and calculating them with an algorithm, models can distinguish different emotions.

4.2.2. Voice and speech analysis

The vocal aspect carries a lot of information during communication in daily life. Usual models can not get access to vocal data, but if we ignore how the message was delivered and focus on the content only, we may fully misunderstand the sentiment contained in the utterance. Irony speech can be an obstacle on our way to identifying emotion exactly. Mel-frequency cepstral coefficients (MFCCs) are often used to extract acoustic features and prosodic features in order to obtain the pitch and rhythm of the speech. RNNs and LSTM are also used in vocal analysis to analyze the timing information of the speech.

4.2.3. Gesture and body language analysis

Gestures and body language also carry out our emotions when communicating. Models will extract features of gestures and body language by bone tracking, joint position and body posture representations. By using deep learning models and multimodal fusion methods, gesture and body language can offer a more holistic view of emotions.

4.2.4. Physiological signal analysis

When someone is angry, his heart must beat faster, his breath must be heavier and maybe sweats. What is said above aims to demonstrate that emotion is not only contained in external expressions but also reflected in internal physiological signals. This kind of data can be collected by wearable devices or sensors and then analyzed by models to draw conclusions about the present state of psychology. One of the problems is whether physiological signals can clearly distinguish between different emotions. Fear and anger can both cause the heart to race. By collecting multi-dimensional data like brain siganls and take timing information into consideration can solve this problem [7].

4.2.5. Multimodal fusion methods

Multimodal fusion methods mainly integrate at least two types of data for analysis. The techniques of this method include early (feature) fusion and later (decision) fusion. Feature fusion is mostly studied on signal processing problems, for example, how to align visual signals with vocal signals into the model, while decision fusion is how to synthesize results to get the final answer after processing various types of signals. On the other hand, the acceptance of multi-mode signals requires high cost and large computing power requirements, and how to reduce the cost is also a research direction. By using this method, the accuracy of the model can be significantly improved: the accuracy of the human-processed data (such as the laboratory environment) can be improved by 12.7%, while the natural scene data is improved by about 4.59%. In the MOUD data set, the multi-modal feature fusion can reduce the error rate of emotion recognition by 10.5% [8].

5. Challenges and potential solutions

As sentiment analysis continues to evolve, several challenges occur from the complexity of human emotions and the diversity of information. It is crucial to address these problems to advance this field and build more robust, accurate, and inclusive sentiment analysis models. Some key challenges and potential solutions are discussed below.

5.1. Sarcasm and irony

Sarcasm and irony mean saying or writing the opposite of what someone means, or of speaking in a way intended to make someone else feel stupid or show them that he is angry. This makes sentiment analysis more complex because a positive word can mean the exact opposite. Spitale et al. found that certain facial action units and acoustic characteristics of speech are key indicators of irony expression, which means that multimodal detection can effectively recognize sarcasm and irony [9].

5.2. Cross-lingual and cross-culture analysis

The development of the internet and communications technology making people closer than ever before, leading to language and culture blending and melting together. multi-language mixing and using idioms from other cultures to express emotions appears in the network corpus. This presents several difficulties for sentiment analysis. Barnes J et al. proposed a bilingual embedding model that make projection from a source language to a target language instead of just translating the output of the source language model and performed great on the test [10].

6. Conclusion

This article provides an overview of sentiment analysis, definition, applications and introduces common types of tasks and their correlative method in sentiment analysis. Some rule-based methods and machine learning methods have shown their potential ability to analyze sentiment and they did perform great on tests. Some challenges faced are discussed as well as potential solutions. This paper lacks some comprehensiveness in analyzing the methodology of sentiment analysis like robustness. Multimodal aspects deserve more discussion and study. Future studies can further explore the application of multimodal sentiment analysis, especially in cross-language and cross-cultural settings. In addition, sentiment analysis combined with physiological signals is also a worthy research direction.

References

[1] S. Baccianella, A. Esuli, F. Sebastiani, SentiWordNet, Sentiwordnet 3.0: An enhanced lexical resource for se ntiment analysis and opinion mining, in: Proc. Int. Conf. Lang. Resour. Eval. {LREC} 2010, 17-23 2010, Eu ropean Language Resources Association, Valletta, Malta, 2010, pp. 1-5, http: //www.lrec-conf.org/proceeding s/lrec2010/pdf/769_Paper.pdf.

[2] Zezawar, T.K., Aung, N.M., Sentiment analysis of students' comment using lexicon based approach. IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 2017.

[3] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding, 2018, pp. 1-16, http://arxiv.org/abs/1810.04805.

[4] HU M, LIU B. Mining and summarizing customer reviews[C/OL]//Proceedings of the tenth ACM SIGKDD i nternational conference on Knowledge discovery and data mining. 2004. http://dx.doi.org/10.1145/1014052.1 014073. DOI:10.1145/1014052.1014073.

[5] LI W, QI F, TANG M, et al. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification[J/OL]. Neurocomputing, 2020: 63-77. http://dx.doi.org/10.1016/j.neucom.2020.01.006. DOI:10.1016/j.neucom.2020.01.006.

[6] C.E. Izard. Innate and universal facial expressions: Evidence from developmental and cross-cultural research. Psychological Bulletin, 115(2): 288-299, 1994.

[7] R. W. Picard, E. Vyzas, J. Healey. "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE Transactionas on Pattern Analysis and Machine Intelligence, vol. 23, pp. 1175-1191, 2001.

[8] Poria S, Cambria E, Bajpai R, Hussain A. A review of affective computing: From unimodal analysis to multimodal fusion.Information Fusion, 2017, 37: 98-125.

[9] Micol Spitale, Fabio Catania, and Francesca Panzeri. 2024. Understanding Non-Verbal Irony Markers: Mac hine Learning Insights Versus Human Judgment. In International Conference on Multimodal Interaction (IC MI '24), November 04--08, 2024, San Jose, Costa Rica. ACM, New York, NY, USA 9 Pages. https://doi.org/1 0.1145/3678957.3685723

[10] Barnes J, Klinger R. Embedding projection for targeted cross-lingual sentiment: Model comparisons and a real-world study[J]. Journal of Artificial Intelligence Research, 2019, 66: 691-742-691-742.