Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 3, No. 1, 2023

Multimodal Emotion Recognition and Fluctuation: A Study on Sentiment Analysis of Online Public Opinion

Lin Sun

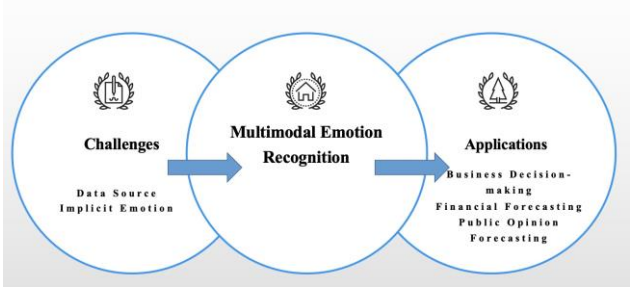
School of Government and Public Affairs, Communication University of China, 100024, Beijing, China

Abstract: Emotion is an important way for individuals to express their views on the Internet and an important variable that shapes public opinion. Considering the multimodality of data, such as text, picture and video, and the subtlety of emotional expression, a multimodal sentiment analysis model that addresses content involving difference senses, such as sight, hearing and touch at the same time is very necessary. This study outlines the basic steps, classification strategies and research methods of sentimental analysis and acknowledges the differences between sentimental analyses on text, picture and video. As multimodal sentiment recognition is still in its initial stage, there's still room for improvement in cross-disciplinary research on multimodal data of text, picture, audio, video in terms of weighted scoring, complex emotion and intensity recognition. It's concluded that future studies should focus on the intensity of different emotions, multimodal data fusion and how weighted scoring influences an emotion recognition model and explore application possibilities.

Keywords: Emotion; Online Public Opinion; Sentiment Analysis.

1. Introduction

In the post-truth era, public opinion has become more and more complex, including not only views but also emotions and attitudes. Emotion is an important way for people to express their opinions and measure whether the outside world meets their inner needs. Systematic emotion mining not only helps understand people's views and stances, but also has practical significance in business decision-making, forecasting, and emotion management However, as shown in Figure 1, sentiment analysis also faces challenges, such as the multimodality of data and the subtlety of emotional expression.



Figures 1. Challenges and Applications of Sentimental Analysis of Online Public Opinion

1.1. Challenges

1.1.1. Data Multimodality

At this stage, sentiment analysis is still text-based. Fine-grained sentiment analysis of text data mainly involves product reviews, service reviews, social media messages and other short texts characterized by short length and clear theme, which makes it easier for sentiment analysis. However, emotion data are not limited to text. Picture, video, audio and a combination of these types are very common. Moreover, as online interaction becomes an integral part of daily life, observing how people have conversations to express emotions and communicate has become an important means to conduct a dynamic sentiment analysis. Chen (2023) points out that sentiment analysis of conversations not only focuses on semantics and emotional expression, but also takes dialogues between contexts into account and promotes research innovations, such as multimodal information fusion, contextual modeling, and speaker modeling. [1]

1.1.2. Analysis of Implicit Emotions

Emotion is not always expressed in a straightforward manner; most emotions are scattered and hidden in factual descriptions and non-textual information. An objective statement without any emotional vocabulary may also reflect emotions. For example, a product review about how fragile a product is expresses negative feelings towards the product, but often ignored. The second scenario is that emotions are expressed in auxiliary information other than text, such as picture, video and audio. Emojis deserve extra attention here since they are implicit signs of emotions. Scholars（2005）have tried to propose ways to classify the emotions expressed by emojis, which adds another level to sentimental analysis.[2]

1.2. Applications

1.2.1. Business Decision-making

As online shopping takes up more of people's leisure time, sellers tend to study customer reviews, analyze their attitudes and emotions, delve deep into what influences their purchase decisions and predict their preferences. Moreover, a good understanding of the differences in how consumers feel before and after a purchase and the underlying mechanisms of evoking emotions in sales is the foundation for optimizing marketing decision-making afterwards. [3]

1.2.2. Financial Forecasting

Sentiment fluctuations can have an impact on investment behavior in financial markets. Using Opinion Finder and Google-Profile of Mood States (GPOMS), scholars classified sentiment expressions of Twitter users into positive and negative and found that they could predict the stock market changes 3-4 days in advance. The Calm Index, for example, was found to be in line with the trend of Dow Jones Industrial Average three days later, leading scholars to conclude that the sentiment of social media users towards stock markets could influence investment behavior and eventually decide market trends. [4]

1.2.3. Public Opinion Forecasting

Sentiment drives the evolution of online public opinion and may result in extreme behaviors. Scholars have often studied how emotion changes over time based on data mining to identify the trend of public opinion. For example, Zhang (2019) divided textual sentiment into 20 fine-grained sentiment categories and analyzed Weibo comments based on the MaxEnt model to predict online public opinion.[5] Li (2019) built a model of two foci, i.e., time and user, to predict the sentiments of multiple users and texts during a particular period of time.[6] Majumder N (2018) used CNN, open SMILE, 3D-CNN to extract feature vectors of text, audio and video, and GRU Networks to process these feature vectors to reduce information conflict and redundancy which are common shortcomings of traditional fusion techniques.[7]

However, emotions are embedded in multimodal data, such as text, picture and video and represented in more subtle and implicit ways. Cross-media presence and multiple sources are also potential challenges. In this regard, academia and industry tend to adopt multimodal fusion models for emotion recognition, extract feature indicators of different forms from text, picture, audio and video and perform fusion for standardized evaluation. This study will start with an overview of sentiment analysis of online public opinion, including the basic steps and sentiment classification and then compare recognition methods, aiming to improve multimodal sentiment analysis models and explore their application possibilities.

2. Literature Review

Sentiment analysis has long been popular in academia. Some scholars have used the methods of frequency–inverse document frequency (tf–idf) and data mining to examine the co-occurrence of words in comments and established the initial research themes as shown in Figure 2.[8]

Chinese scholars prefer research themes of online opinion monitoring, sentiment evolution and sentiment classification, while foreign scholars are more concerned with data mining. Research steps include sentiment information extraction, sentiment classification, sentiment retrieval and summarization, etc. The three main research approaches are machine learning, deep learning and sentiment analysis dictionaries. The early-stage research was dominated by the last approach and at a later stage, machine learning, deep learning and leveraging cognitive features for sentiment analysis gained popularity.

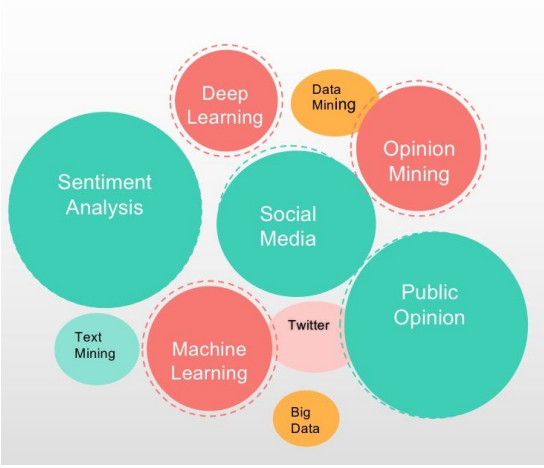


Figure 2. Top 10 Keywords in Literature of Sentimental Analysis of Online Public Opinion

2.1. Steps

Sentiment analysis mainly includes information extraction, sentiment classification, and sentiment retrieval and summarization. As shown in Figure 3, the first step is information extraction, which aims to identify information elements that reflect emotional tendencies, such as holder, target and opinion and provide data for the next step of sentiment classification. The second step is sentiment classification, which categorizes opinions based on emotional tendencies with the help of dictionaries and rules, and machine learning. The third step is sentiment summarization, which refines and condenses sentiment data.

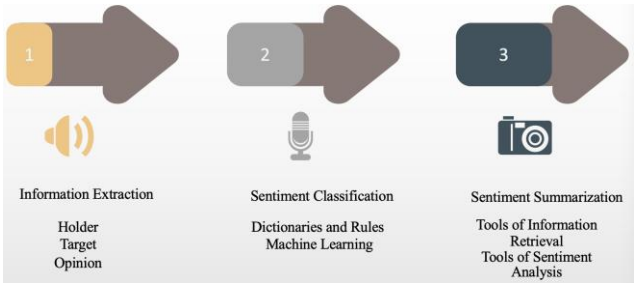


Figure 3. Steps of Sentiment Analysis

2.2. Classification

Emotions are not limited to two categories of positive and negative and different emotions coexist. Plutchik (2001) hypothesizes that the combination of two similar basic emotions leads to a complex emotion. Happiness and expectation, for example, equals optimism. [9] As a result, scholars focus on refining classification methods and diversifying levels of analysis to improve the speed and quality of recognition.

2.2.1. Sentiment Categories

Sentiment classification originated in the West, where the philosopher Descartes classified the “primitive” human emotions as surprise, happiness, hate, desire, joy, and sorrow, and the other emotions are refinements and combinations of these six emotions. Ekman (1992) proposed the use of facial expressions and physiological processes to identify emotional changes, and a theory of six basic emotions which are joy, sadness, anger, fear, disgust, and surprise. [10] On the basis of existing theories, Parrott (2001) constructed a three-layered tree structured model with the first layer as love, joy, surprise, anger, sadness, and fear. Eventually, after the second and third layer, 100 emotions are identified. [11]

As the Internet becomes more popular, there are more researches based on data from the Internet. In 2006, the Text Retrieval Evaluation Conference (TREC) first introduced “blog track”. Mishne (2005) used frequency in statistics and semantic features to classify emotions and identified 37 sentiment categories in blog posts. [12]

2.2.2. Level of Analysis

It is also necessary to determine the basic levels and units of analysis before the analysis begins. The aspects of emotions are understood within the chosen unit and the level of analysis determines the granularity of emotion recognition. For example, in terms of traditional text analysis, sentiment analysis can be divided into coarse-grained which means sentence-level or text-level and fine-grained which means target-based and aspect-based. Fine-grained analysis is more and more popular.

2.2.2.1. Coarse-grained Analysis

Text-level analysis is the most cursory type of sentiment analysis, assuming that there's only one kind of emotion present in the whole text which is either positive or negative. Sentence-level analysis divides a text into sentences. Even though the unit of analysis is smaller and the granularity is higher, the thinking behind it is no different. As for steps, sentence-level analysis includes judging whether a sentence is objective or subjective at the very beginning and then dividing subjective sentences into positive and negative. However, both sentence-level and text-level analyses are based on the assumption that there's only one kind of emotion within the unit of analysis, in conflict with the complexity and subtlety of emotions in the real world.

2.2.2.2. Fine-grained Analysis

Considering that coarse-grained analysis cannot handle situations where there are multiple targets and aspects, attribute-based fine-grained analysis which excels in mining opinions is an important shift. Therefore, it's necessary to conduct fine-grained analysis for the purpose of identifying the aspect of an entity and its corresponding sentiment lexicon. Fine-grained analysis tends to include three steps: first, identify the aspect of an entity and its corresponding sentiment lexicon; second, classify the orientation of the sentiment towards the aspect of the entity; third, summarize the results from classification. Fine-grained analysis can be divided into two kinds, i.e., entity-based sentiment analysis and aspect-based sentiment analysis.

In addition, apart from analysis of a single text, scholars also try to take a broader "event" as a unit of analysis and extract a target event and elements, such as event trigger word, event type, event argument and argument role from unstructured texts in natural languages with the help of event extraction methods to conduct research on complex real-life scenarios and predict public opinion. [13]

3. Methodology

3.1. Lexicon-based Approach

The lexicon-based approach requires glossaries or dictionaries as a basis to conduct sentiment classification and recognition. Here dictionaries refer to lists of words or phrases labeled with emotion tendencies that are selected by researchers. Researchers can calculate and identify the emotion tendencies in text conversations with mainstream dictionaries such as the WordNet-Affect and the NRC Emotion Lexicon. However, dictionaries are usually hand-curated, which causes shortcomings in flexibility and timeliness. Moreover, dictionaries are geographically and linguistically restricted, so their content is confined by a narrow pre-defined environment, posing considerable difficulties for promotion and application.

3.2. Machine Learning-based Approach

Machine learning-based approaches use manually labeled sentiment documents as a training set to create classifiers to speed up sentiment recognition. For example, Pang and Lee et al. (2002) pioneered the application of machine learning to text sentiment analysis, comparing the effectiveness of different machine learning methods of Support Vector Machine, Naive Bayes, and Maximum Entropy for sentiment classification, and concluded that Support Vector Machine classifiers are best out of three. [14]

3.3. Deep Learning and Natural Language Processing

Emotions exist through the whole process of expression, and some hidden emotions can only be understood properly with the help of contexts. In the face of complex and multiple analysis dimensions, deep learning methods are gaining traction in academia. These methods include Emotion Annotation, Word Embedding Learning, Long Short-Term Memory, Convolutional Neural Network, Recurrent Neural Network and so on, out of which RNN has an advantage in terms of contextual information and LSTM can solve the problem of vanishing gradients through its gating mechanism.

In the Internet era, the efficiency of hand-curated sentiment dictionaries has gradually decreased, and the accuracy of machine learning algorithms needs improvement due to the challenge of contextual coherence. In contrast, deep learning has a strong capacity for feature learning and addresses contextual information, but long training time and poor interpretability are its shortcomings.

4. Multimodality

In the Internet era, people express their emotions in a variety of ways. In addition to text recognition mentioned above, picture, audio, video and combinations of these forms are important materials for sentiment analysis, propelling a comprehensive evaluation system for multimodal emotion recognition to emerge. Taking the characteristics of picture, audio and video into consideration, the system dynamically adjusts the feature weight of each form of information by using feature fusion (early fusion) or decision fusion (late fusion) as appropriate. In order to comprehensively show the multimodal fusion process, below are common fusion methods and recognition methods of each modality.

4.1. Fusion Methods

Fusion methods are mainly divided into data-level fusion, feature-level fusion and decision-level fusion, depending on the processing stage at which fusion takes place. Data-level fusion refers to manually combines several sources of raw data to produce new raw data, which is likely to generate redundant information. Feature-level fusion refers to fusing feature vectors with the help of deep learning and manual selection, characterized by its flexibility and capacity for multimodal information complementarity. However, this method is a crude fusion of feature vectors of multimodal data, unable to adjust feature weights according to the characteristics of different modalities. Decision-level fusion goes a step further by inputting multimodal data into different classifiers for sentiment polarity assignment, which recognizes modal differences but does not allow for information complementarity.

Currently, multimodal research has become important and productive. Poria et al. (2016) used the generalized adaptive view-based appearance model (GAVAM) to identify features of facial expressions, used Open EAR to extract audio features and summarized text features for emotion recognition using concept extraction-based methods. [15] In addition, Zhao et al. (2022) pointed out that existing methods for multimodal aspect-level sentiment analysis focus one-sidedly on the interaction and ignore the inconsistency between image and text data. [16] To address this problem, their study proposed a Joint Aspect Attention Interaction

Network model to take the inconsistency and correlation between the two kinds of data into account.

4.2. Multimodal Recognition

Picture, video and audio as well as the traditional modality of text are the main categories for emotion recognition. As for pictures, there are three main techniques: the first is identifying underlying visual features, including elements such as color, texture and line; the second focuses on mid-tier semantic connotations, clarifying semantic features by analyzing adjectives and nouns; and the third is building deep neural networks with the help of deep learning algorithms. For example, Cao et al. (2016) designed a Visual Sentiment Topic Model (VSTM) based on the Visual Sentiment Ontology (VSO) to analyze the sentiment information implied by images. [17]

Meanwhile, scholars have also conducted systematic research on emotion recognition in the context of text-image and audio-visual fusion, introducing pre-processing of image, audio and other materials. For example, Huang (2022) suggested a method called Image Semantic Translation-based Sentiment Analysis for Image Fusion (ImaText-IST), which transforms images into image descriptions followed by sentiment polarity assignment.[18] Liu (2022) designed a combined attention graphic feature fusion module based on Transformer Encoder and attention mechanism, through which the information relevance of each word and picture in the text is calculated to improve the capacity for emotional representation of text features.[19]

5. Conclusion

Starting from the application value and challenges of sentiment analysis, this paper comprehensively analyzes the main steps and research methods of sentiment analysis and summarizes multimodal sentiment recognition methods suitable for analyzing public opinion in different forms, broadening research scope and the range of raw data that can be processed. Unfortunately, multimodal sentiment recognition is still in its initial stage. There's still room for improvement in cross-disciplinary research on multimodal data of text, picture, audio, video in terms of weighted scoring, complex emotion and intensity recognition. In the future, it is necessary to have a better understanding of the implications of variables such as emotion intensity and emotion conversion in the hope that sentiment analysis will play a bigger role in risk grading of a public opinion event.

References

[1] CHEN Xiaoting, Li Shi, "Survey on Emotion Recognition in Conversation," in Computer Engineering and Applications, vol. 3, J. 2023, pp.33-48.

[2] Read J, "Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification（Presented Conference Paper style)," presented at Proc of the 43rd ACL Student Research Workshop, Stroudsburg.

[3] Li Ran, Lin Zheng, Lin Hailun, Wang Weiping, Meng Dan, "Text Emotion Analysis: A Survey" in Journal of Computer Research and Development, vol. 55, J. 2018, pp. 30-52.

[4] Bolen J, Mao H, Zeng X, "Twiter mood predicts the stock market," in Journal of Computational Science, vol. 2, J.2011, pp.1 -8.

[5] Zhang M, Zheng R, Chen J, et al, " Emotional Component Analysis and Forecast Public Opinion on Micro-blog Posts Based on Maximum Entropy Model, " in Cluster Computing, vol. 3, J. 2019, pp.6295-6304.

[6] LiL, WuY, Zhang, etal, "Time+user Dual Attention Based Sentiment Prediction for Multiple Social Network Texts with Time Series, "IEEE-Access, vol.7, J. 2019, pp. 17644-17653.

[7] Majumder N, Hazarika D, Gelbukh A, et al, "Multimodal Sentiment Analysis Using Hierarchical Fusion with Context Modeling, "in Knowledge-based systems, vol.161, J.2018, pp.124-133.

[8] Shi Wei, Xue Guangcong, He Shaoyi, "Literature Review of Network Public Opinion Research from the Perspective of Sentiment "in Documentation, Information & Knowledge, vol.39, J. 2022, pp.105-118.

[9] Plutchik R, "The Nature of emotions, " in Philosophical Studies, vol.4, J.2001, pp. 393-409.

[10] Ekman P, "An Argument for Basic Emotions,"in Cognition and Emotion, vol.6, J.1992, pp.169-200.

[11] Parrott W G, Emotions in Social Psychology: Esential Readings. Oxford, UK: Psychology Press, 2001.

[12] Mishne G,"Experiments with Mood Clasification in Blog-Posts, "presented at Proc of the 28th ACM SIGIR Workshop on Stylistic Analysis of Text for Information Access, New York.

[13] Shen Lining, Yang Jiayi, Pei Jiaxuan, Cao Guang, Chen Gongzheng, "A Fine-grained Emotion Recognition Method Based on OCC-Model and Emotion Cause Event Extraction (Unpublished work style)" unpublished.

[14] Pang B, Lee L, "Thumbs up? Sentiment classification using machine learning techniques"in Empirical Methods in Natural Language Processing , vol.9, J.2022, pp. 79-86.

[15] Poria S, Cambria E, Howard N, et al, "Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content, "in Neurocomputing, vol.174, 2016, pp. 50-59.

[16] Zhao Yicheng, Wang Suge,"Image-text aspect emotion recognition based on joint aspect attention interaction (Unpublished work style) , " unpublished.

[17] CaoD,JiR,"VisualSentimentTopicModelBasedMicroblogImageSentimentAnalysis[J]. Multimedia Tools and Applications "in Multimedia Tools and Applications, vol. 75,2016, pp.8955-8968.

[18] HUANG Jian, WANG Ying, "Image-text fusion sentiment analysis method based on image semantic translation (Unpublished work style) , " unpublished.

[19] LIU Qiwei, LI Jun, GU Beibei, ZHAO Zefang, "TSAIE: Text Sentiment Analysis Model Based on Image Enhancement"in Frontiers of Data & Computing, vol. 4, J.2022, pp. 131-140.