SNEH TEACHERS TRAINING COLLEGE, JAIPUR
"Foster Emotional Intelligence in Youth Through Education" (ICFEIYE-2024)
DATE: 15 April 2024
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 7.938

# Reviewing A Multimodal Sentiment Analysis for Achieving Accuracy in Image Content

Ms. Pooja Morey, M.E Computer Engineering, Student, Padm. Dr. V.B.Kolte Coe, Malkapur, Sgbau, India, moreypooja114@gmail.com
Prof. Y. B. Jadhao, M.E Computer Engineering, Faculty, Padm. Dr. V.B.Kolte Coe, Malkapur, Sgbau, India, ybjadhao@yahoo.com

## Abstract

In recent years, with the popularity of social media, users are increasingly keen to express their feelings and opinions in the form of pictures and text, which makes multimodal data with text and pictures the con tent type with the most growth. Most of the information posted by users on social media has obvious sentimental aspects, and multimodal sentiment analysis has become an important research field. Previous studies on multimodal sentiment analysis have primarily focused on extracting text and image features separately and then combining them for sentiment classification. These studies often ignore the interaction between text and images. Therefore, this project proposes a new multimodal sentiment analysis model. The model first eliminates noise interference in textual data and extracts more important image features. Then, in the feature-fusion part based on the attention mechanism, the text and images learn the internal features from each other through symmetry. Then the fusion features are applied to sentiment classification tasks. The experimental results on two common multimodal sentiment datasets demonstrate the effectiveness of the proposed model.

**Keywords: Data, Sentiment, Analysis, Classification, Multimodal**

## I. INTRODUCTION

The goal of image classification is to decide whether an image belongs to a certain category or not. Different types of categories have been considered in the literature, e.g. defined by presence of certain objects, such as cars or bicycles [7], or defined in terms of scene types, such as city, coast, mountain, etc. [12]. To solve this problem, a binary classifier can be learned from a collection of images manually labeled to belong to the category or not. Increasing the quantity and diversity of hand-labeled images improves Tags: desert,nature,landscape,sky Tags: rose, pink Labels: clouds, plant life, sky, tree Labels: flower, plant life Tags: india Tags: aviation, airplane, airport Labels: cow Labels: aeroplane. Example images from MIR Flickr (top row) and VOC'07 (bottom row) data sets with their associated tags and class labels. The performance of the learned classifier, however, labeling images is a time-consuming task. Although it is possible to label large amounts of images for many categories for research purposes [6], this is often unrealistic, e.g. in personal photo organizing applications. This motivates our interest in using other sources of information that can aid the learning process using a limited amount of labeled images.

### A. Multi-Modal Modelling

Multi-Modal Modeling of images and text combines semantic knowledge extracted from text with knowledge of spatial structures extracted from images. Models of this type learn joint representations of images and text. These joint representations have been used to relate images and text to improve search-and-retrieval, classification, and self-supervised learning.

### B. Self-Supervised Learning

As an alternative to fully human-supervised algorithms, recently, there has recently been a growing interest in self-supervised or naturally-supervised. These approaches make use of non-visual signals, intrinsically correlated to images, as a form of supervision for visual feature learning (Gomez et al., 2019). The prevalence of websites with images and loosely-related human annotations provides a natural opportunity for self-supervised learning.

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 7.938

Research has lately focused on joint image and text embeddings. Merging different kinds of data has motivated the possibilities of learning together from different kinds of data, which put more focus on the field of study where both general and applied research has been done. Sentiment analysis is a method, which classifies the given data due to having positive or negative opinion and being subjective or objective in general. In this case, the concept of emotions and opinions is coming in focus, the opinionated data is deeply analyzed to determine the strength of opinions, which is closely related to the intensity of emotions such as happy, fear, sad, anger, surprise etc. It is claimed that people's sentiments can be identified by examining their language expressions, and can be classified, according to the level of their strength. It is mostly used in advertisement placement, product benchmarking and market intelligence, and detection of company reputation or brand popularity and identification of fake or misinforming comments.

## II. LITERATURE SURVEY

Facial recognition is fundamental for a wide variety of security systems operating in real-time applications. Recently, several deep neural networks algorithms have been developed to achieve state-of-the-art performance on this task. The present work was conceived due to the need for an efficient and low-cost processing system, so a real-time facial recognition system was proposed using a combination of deep learning algorithms like FaceNet and some traditional classifiers like SVM, KNN, and RF using moderate hardware to operate in an unconstrained environment. Generally, a facial recognition system involves two main tasks: face detection and recognition. [1]

Traditionally the document classification was performed on the topic basis but later research started working on opinion basis. Following machine learning methods Naive Bayes, Maximum Entropy Classification (MEC), and Support Vector Machine (SVM) are used for sentiment analysis. The conventional method of document classification based on topic is tried out for sentiment analysis. The major two classes are considered i.e. positive and negative and classify the reviews according to that. [2]

Naïve Bayes is best suitable for textual classification, clustering for consumer services and Support Vector Machine for biological reading and interpretation. The four methods discussed in the paper are actually applicable in different areas like clustering is applied in movie reviews and Support Vector Machine (SVM) techniques is applied in biological reviews & analysis. Though field of opinion mining is latest technology, but still, it provides diverse methods available to provide a way to implement these methods. [3]

Based on prosody-based Automatic Personality Perception, i.e., on automatic prediction of personality traits attributed by human listeners to unknown speakers. The APP results show an accuracy ranging between 60 and 72 percent (depending on the trait) in predicting whether a speaker is perceived to be high or low with respect to a given trait. The most probable reason is that the corpus includes two categories of speakers (professional and nonprofessional ones) that differ in terms of characteristics typically related to the trait e.g. efficiency, reliability etc.[4]

## III. SYSTEM DESIGN

### A. Existing System

Traditionally the document classification was performed on the topic basis but later research started working on opinion basis. Following machine learning methods Naive Bayes, Maximum Entropy Classification (MEC), and Support Vector Machine (SVM) are used for sentiment analysis. The conventional method of document classification based on topic is tried out for sentiment analysis. The major two classes are considered i.e. positive and

negative and classify the reviews according to that. In [5], Naïve Bayes is best suitable for textual
classification, clustering for consumer services and Support Vector Machine for biological reading and interpretation. The four methods discussed in the paper are actually applicable in different areas like clustering is applied in reviews and Support Vector Machine (SVM) techniques is applied in biological reviews & analysis. Though field of opinion mining is latest technology, but still it provides diverse methods available to provide a way to implement these methods.

**B. Proposed System**

Speech signals convey not only words but also emotions. Various analysis and models had been submitted and explored for Textual Analysis but this analysis is incomplete due to ignorance of Sentiments involved and result may not be reliable and in addition Textual Analysis only focus on word content and thereby ignores the acoustic features of speech. Thus it needs analysis of Sentiments as well as text simultaneously. When applying a multi-modals tacked denoising autoencoder, the structure of the hidden nodes is modified with the number of classified input nodes, whereas the parameters remain the same. All single-modal data have a different input value depending on the number of parameters classified into each class, the middle hidden value of which is smaller than the input value.

**C. Dataset**

User's opinion is a major criterion for the improvement of the quality of services rendered and enhancement of the deliverables. Blogs, product review sites, micro blogs provide a good understanding of the reception level of the products and services.

A. Blogs contains reviews of many product and issue. These blogs contains large amount of opinionated text. Blogger opens a topic in discussion and records daily events for opinions, feelings, emotions etc, so it is very important to apply sentiment features over blog data [15].

B. Review sites site contains views of people about product or topic. These views are in form of comments which are in unstructured format. S, people read those comments and makes purchasing decision. E-commercewebsites1, 2, 3, 4 hosts millions of product reviews given by customer such data is used in sentiment analysis. Other than these the available are professional review sites5, 6 and consumer opinion sites on different topics.

C. Dataset- Most of the work in the field uses movie reviews data for classification. Movie review data is available as dataset. Other dataset which is available online is multidomain sentiment (MDS) dataset.

D.Micro-blogging- These are sites which allow user post small messages as a status e.g. tweets of tweeter. Sometimes these twits express opinions. So, Twitter messages are studied to classifysentiments.4.3 Technique used.

**IV. SENTIMENT ANALYSIS ALGORITHMS FALL INTO ONE OF THREE BUCKETS:**

**Rule-based:** these systems automatically perform sentiment analysis based on a set of manually crafted rules.

**Automatic:** systems rely on machine learning techniques to learn from data.

**Hybrid** systems combine both rule-based and automatic approaches.

**Rule-based Approaches** Usually, a rule-based system uses a set of human-crafted rules to help identify subjectivity, polarity, or the subject of an opinion. Counts the number of positive and negative words that appear in a given text. If the number of positive word appearances is greater than the number of negative word appearances, the system returns a positive sentiment, and vice versa. If the numbers are even, the system will return a neutral sentiment. Rule-based systems are very naive since they don't take into account how words

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Indexed, Double-Blind, Open Access, Peer-Reviewed, Refereed-International Journal, Impact factor (SJIF) = 7.938

are combined in a sequence. Of course, more advanced processing techniques can be used, and new rules added to support new expressions and vocabulary. However, adding new rules may affect previous results, and the whole system can get very complex. Since rule-based systems often require fine-tuning and maintenance, they'll also need regular investments.

**Automatic Approaches** Automatic methods, contrary to rule-based systems, don't rely on manually crafted rules, but on machine learning techniques. A sentiment analysis task is usually modeled as a classification problem, whereby a classifier is fed a text and returns a category, e.g. positive, negative, or neutral. Queries of products could be associated with the images, titles, and descriptions, returning a much richer set of data. Continued improvements in image classification model development have progressed the realm of computer vision centered on deep learning.

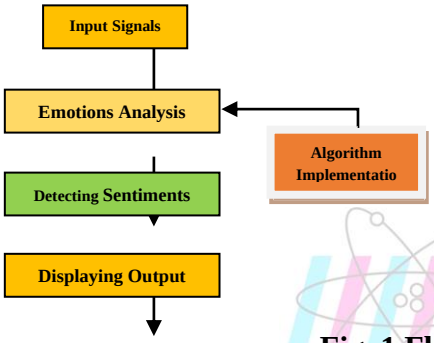


**Fig. 1 Flow of the proposed system**

In integration, we can integrate the results of all these modes. Facial expressions helps to understand the present emotion of human thus various methods have been proposed to identify typical part of face and movement of specific points associated with different emotional states. Often techniques used are sign judgement in which method describes the appearance in spite the meaning of shown behaviour, message judgement which addresses the interpretation of shown behaviour. One more technique is known as FACS (Facial Action Coding System) propped by Ekman which does manual labelling of facial behaviour, with respect to sign judgement. The movement of certain parts of facial muscles are encoded by FACS.

## V. CONCLUSION

Multi-modal Sentiment Analysis problem is a machine learning problem that has been a research interest for recent years. Though lot of work is done till date on sentiment analysis, there are many difficulties to sentiment analyser since Cultural influence, linguistic variation and differing contexts make it highly difficult to derive sentiment. The reason behind this is unstructured nature of natural language. The main challenging aspects exist in use of other modes; dealing with Multi-Modality entails the use of multiple media such as audio and video in addition to text to enhance the accuracy of sentiment analyzers. Textual emotional classification is done on basis of polarity, intensity of lexicons. Audio emotional Classification is done on basis of prosodic features. Video emotional Classification is done on basis postures, gestures etc. Infusion, we can integrate the results of all these modes; to get more accuracy. Future research could be dedicated to these challenges. So we are moving from uni-modal to multi-modal.

## REFERENCES

[1] Boiy, E. Hens, P., Deschacht, K. &Moens, M.F., "Automatic Sentiment Analysis in Online Text", In Proceedings of the Conference on Electronic Publishing(ELPUB-2007).

[2] J. Wiebe, T. Wilson, and C. Cardie. "Annotating expressions of opinions and emotions in language", Language Resources and Evaluation, 2005.

[3] Scott Brave and Clifford Nass, Emotion in Human- Computer Interaction. Retrieved from http://lrcm.com.umontreal.ca/dufresne /COM7162/EmotionHumanInteraction.pdf
[4] S. V. Bo Pang, Lillian Lee, "Thumbs up? Sentiment classifcation using machine learning techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), ACL, pp. 79– 86,July 2002.
[5] Pravesh Kumar Singh and Mohd Shahid Husain "Methodological study of opinion mining and sentiment analysis techniques" International Journal on Soft Computing (IJSC) Vol. 5, No. 1, February 2014
[6] Gelareh Mohammadi and Alessandro Vinciarelli "Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features", IEEE transactions on affective computing, vol. 3, no. 3, july-september 2012
[7] Litman, D.J. and Forbes-Riley, K., "Predicting Student Emotions in Computer-Human Tutoring Dialogues". In Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), July 2004
[8] Lee C M Narayanan, S.S., "Toward detecting emotions in spoken dialogs". IEEE Tran. Speech and Audio Processing, Vol. 13 NO. 2, March 2005
[9] Mozziconacci, S., "Prosody and Emotions". Int. Conf. on Speech Prosody. 2002 Danielle Lottridge, Mark Chignell, and Michiaki Yasumura," Identifying Emotion through Implicit and Explicit Measures: Cultural Differences, Cognitive Load, and Immersion", IEEE transactions on affective computing, vol. 3, no. 2, april-june 2012
[10] Mohammad Soleymani, Maja Pantic, Thierry Pun," Multimodal Emotional Recognition in Response to Videos", IEEE transactions Affective Computing , Vol.3,No.2, April-June 2012
[11] Gaurav Vasmani and Anuradha Bhatia "A Real Time Approach with Big data- A Survey", International Journal of Engg Sciences & Research Technology, Vol 3, Issue 9, September 2013
[12] LouisPhilippe, Morency Rada Mihalcea and Payal Doshi "Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web", ICMI'11, November 14–18, 2011, Alicante, Spain.
[13] Andrea Kleinsmith and Nadia Bianchi-Berthouze "Affective Body Expression Perception And Recognition:A Survey", IEEE transactions Affective Computing , Vol.4,No.1, January-March 2013s
[14] Gérard Dray, Michel Plantie , Ali Harb, Pascal Poncelet Mathieu Roche and François Trousset, "Opinion Mining From Blogs" International Journal of Computer Information Systems and Industrial Management Applications – IJCISIM Vol. 1 2009
[15] Charles B. Ward, Yejin Choi, Steven Skiena, "Empath: A Framework for Evaluating Entity-Level Sentiment Analysis", IEEE 2011
[16] Georgios Paltoglou and Michael Thelwall, "Seeing Stars of Valence and Arousal in Blog Posts" IEEE transactions on affective computing, vol. 4, no. 1, January-March 2013
[17] Shangfei Wang, Zhilei Liu, Zhaoyu Wang, Guobing Wu, Peijia Shen, Shan He, and Xufa Wang "Analyses of a Multimodal Spontaneous Facial Expression Database" IEEE transactions on affective computing, vol. 4, no. 1, January- March 2013
[18] Ashish Tawari and Mohan Manubhai Trivedi, "Speech Emotion Analysis: Exploring the Role of Context" IEEE transactions on multimedia, vol. 12, no. 6, October 2010
[19] Tal Sobol-Shikler and Peter Robinson, "Classification of Complex Information: Inference of Co-Occurring Affective States from Their Expressions in Speech", IEEE transactions on pattern analysis and machine intelligence, vol. 32,no.7,July 2010
**[20]** Felix Weninger, Jarek Krajewski, Anton Batliner, and Bjorn Schuller, "The Voice of Leadership:Models and Performances of Automatic Analysis in Online Speeches", IEEE transactions on affective computing, vol. 3, no. 4, October-December.2012