ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 3 | Issue 6 | December 2024

www.ijirid.in

A Systematic Research on Emotion Recognition from Facial Expressions Using Machine Learning Techniques

Mohini Narkhede[1], Prof. Pallavi P. Rane[2], Prof. Nilesh N. Shingne[3]

[1]*PostGraduate Student, Rajarshi Shahu College of Engineering, Buldhana, (M.S), India*

[2]*Assistant Professor, Rajarshi Shahu College of Engineering, Buldhana, (M.S), India*

[3]*Assistant Professor, CS&IT, Sanmati Engineering College Washim, (M.S), India*

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 emotion recognition sentiment datasets demonstrate the effectiveness of the proposed model.*

Keywords: Opinion, Data, Sentiment, Classification, Features.

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, or defined in terms of scene types, such as city, coast, mountain, etc. 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, this is often unrealistic, e.g. in personal photo organizing applications. [2]

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. The prevalence of websites with images and loosely-related human annotations provide a natural opportunity for self-

DOI 10.5281/zenodo.14790039

supervised learning. This differs from previous image-text embedding methods in that the goal is to learn generic and discriminative features in a self-supervised fashion without making use of any annotated dataset.

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. A Deep Visual-Semantic Embedding Model proposes a pipeline that, instead of learning to predict ImageNet classes, learns to infer the Word2Vec representations of their labels. By exploiting distributional semantics of a text corpus of every word associated with an image provides inferences of previously unseen concepts in the training set. Semantically relevant predictions make this model valuable even when it makes errors. These errors are generalized to a class outside the labeled training set. [3]

II. PROJECT OBJECTIVES

- To analysed the techniques used for sentiment data and also demonstrate that how individual model works. Extracting the sentiments from different input modes is achieved by different classifying techniques.
- To find out different input modes to generate the model for analysis.
- To analysed challenges of our proposed system and integration of different modes and its effect on emotion reorganization system.

III. LITERATURE REVIEW

Alejandra Sarahi Sanchez-Moreno et al state that 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]

HANG DU, et all state that Face recognition (FR) is an extensively studied topic in computer vision. Among the existing technologies of human biometrics, face recognition is the most widely used one in real-world applications. With the great advance of deep convolutional neural networks (DCNNs), the deep learning based methods have achieved significant improvements on various computer vision tasks, including face recognition. In this survey, we focus on 2D image based end-to-end deep face recognition which takes the general images or video frames as input, and extracts the deep feature of each face as output. We provide a comprehensive review of the recent advances of the elements of end-to-end deep face recognition. Specifically, an end-to-end deep face recognition system is composed of three key elements: face detection, face alignment, and face representation. [2]

Madan Lal et al state that with the rapid growth in multimedia contents, among such content face recognition has got much attention especially in past few years. Face as an object consists of distinct

DOI 10.5281/zenodo.14790039

features for detection; therefore, it remains most challenging research area for scholars in the field of computer vision and image processing. [3]

IV. MOTIVATION

Presently large amount of data is available on social networking sites, product review sites, blogs, forums etc. This data holds expressed opinions and sentiments. The volume, variety, velocity are properties of data, whether it comes from the Internet or an enterprise resource planning system, sentiment analysis system should get the data and analyze it. Due to the large volume of opinion rich web resources such as discussion forum, review sites, blogs and news corpora available in digital form, much of the current research is focusing on the area of sentiment analysis and opinion mining. Expression of any sentiment is a mixture of text, prosody, facial expression, body posture etc. Thus only text input cannot fully represent a sentiment. In image categorization the goal is to decide if an image belongs to a certain category or not. A binary classifier can be learned from manually labeled images; while using more labeled examples improves performance, obtaining the image labels is a time consuming process. [4]

V. PROPOSED METHODOLOGY

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. [5]

The present schemes are dealing with only facial image which holds good quality of resolution. But to be practical this is not the case indeed existing system may fail due to several reasons such as low resolution, fast movement of subject etc. Another issue with respect to present system in how to reliable extract there linguistic and paralinguistic features from the image data with many features that has been from the emotion detection within a certain image quality, but such research activity were not satisfied to overcome the traditional challenges. There is need of an automatic extraction of linguistic information from the image data. Spoken contents can be extracted from speech signals with the help of speech to text converter.

In order to implement an effective multimodal, the fusion of different modes must be done with suitable joint feature vectors which is composed of various features of different modalities with differ in time scales. Furthermore to make multimodal effective and realistic other parameters can be considered with respect to human entity such as age, gender. One mode doesn’t give sufficient solution. There is need to consider other modes also such as audio/video. Social media are a huge untapped source of user opinion for various products and services. 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. Speech signals convey not only words but also emotions.

DOI 10.5281/zenodo.14790039

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 on focus is on word content and thereby ignores the acoustic features of speech. Thus it needs analysis of Sentiments as well as text simultaneously. [6]

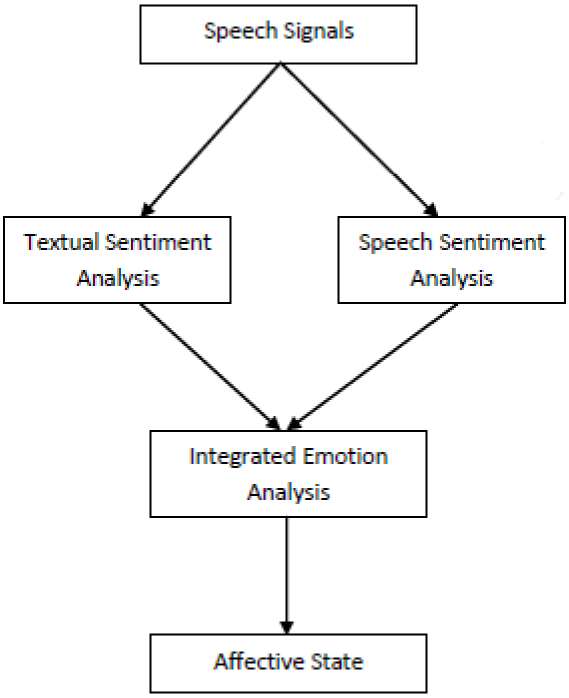
Speech Signals
Textual Sentiment Analysis
Speech Sentiment Analysis
Integrated Emotion Analysis
Affective State

Figure 1: Flow of the Proposed System

VI. ALGORITHM USED

6.1) SVM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. [7]

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

SVM can be of two types:

- ***Linear SVM:*** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- ***Non-linear SVM:*** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

“Support Vector Machine” (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate. [8]

6.2) Classifier Used (Haar Cascade)

Haar Cascade is a machine learning-based approach where a lot of positive and negative images are used to train the classifier. Positive images – These images contain the images which we want our classifier to identify. Negative Images – Images of everything else, which do not contain the object we want to detect.

The algorithm can be explained in four stages:

- Calculating Haar Features
- Creating Integral Images
- Using Adaboost
- Implementing Cascading Classifiers

It’s important to remember that this algorithm requires a lot of positive images of faces and negative images of non-faces to train the classifier, similar to other machine learning models.

6.2.1 Calculating Haar Features:

The first step is to collect the Haar features. A Haar feature is essentially calculations that are performed on adjacent rectangular regions at a specific location in a detection window. These features can be difficult to determine for a large image. This is where integral images come into play because the number of operations is reduced using the integral image

6.2.2 Creating Integral Images:

Without going into too much of the mathematics behind it (check out the paper if you’re interested in that), integral images essentially speed up the calculation of these Haar features. Instead of computing at every pixel, it instead creates sub-rectangles and creates array references for each of those sub-rectangles. These are then used to compute the Haar features. It’s important to note that nearly all of the Haar features will be irrelevant when doing object detection, because the only features that are important are those of the object.

6.2.3 Adaboost Training

Adaboost essentially chooses the best features and trains the classifiers to use them. It uses a combination of “weak classifiers” to create a “strong classifier” that the algorithm can use to detect objects. Weak learners are created by moving a window over the input image, and computing Haar features for each subsection of the image. This difference is compared to a learned threshold that separates non-objects from objects. Because these are “weak classifiers,” a large number of Haar features is needed for accuracy to form a strong classifier.

DOI 10.5281/zenodo.14790039

6.2.4 Implementing Cascading Classifiers:

The cascade classifier is made up of a series of stages, where each stage is a collection of weak learners. Weak learners are trained using boosting, which allows for a highly accurate classifier from the mean prediction of all weak learners. Based on this prediction, the classifier either decides to indicate an object was found (positive) or move on to the next region (negative). Stages are designed to reject negative samples as fast as possible, because a majority of the windows do not contain anything of interest.

It's important to maximize a low false negative rate, because classifying an object as a non-object will severely impair your object detection algorithm. A video below shows Haar cascades in action. The red boxes denote "positives" from the weak learners.

While face recognition has been around in one form or another since the 1960s, recent technological developments have led to a wide proliferation of this technology. This technology is no longer seen as something out of science fiction movies like Minority Report. With the release of the iPhone X, millions of people now literally have face recognition technology in the palms of their hands, protecting their data and personal information. While mobile phone access control might be the most recognizable way face recognition is being used, it is being employed for a wide range of use cases including preventing crime, protecting events and making air travel more convenient. [9]

VII. APPLICATIONS

7.1) Prevent Retail Crime

Face recognition is currently being used to instantly identify when known shoplifters, organized retail criminals or people with a history of fraud enter retail establishments. Photographs of individuals can be matched against large databases of criminals so that loss prevention and retail security professionals can be instantly notified when a shopper enters a store that prevents a threat. Face recognition systems are already radically reducing retail crime. According to our data, face recognition reduces external shrink by 34% and, more importantly, reduces violent incidents in retail stores by up to 91%.

7.2) Unlock Phones

A variety of phones including the latest iPhone are now using face recognition to unlock phones. This technology is a powerful way to protect personal data and ensure that, if a phone is stolen, sensitive data remains inaccessible by the perpetrator.

7.3) Smarter Advertising

Face recognition has the ability to make advertising more targeted by making educated guesses at people's age and gender. Companies like Tesco are already planning on installing screens at gas stations with face recognition built in. It's only a matter of time before face-recognition becomes an omni-present advertising technology.

DOI 10.5281/zenodo.14790039

7.4) Find Missing Persons

Face recognition can be used to find missing children and victims of human trafficking. As long as missing individuals are added to a database, law enforcement can become alerted as soon as they are recognized by face recognition—be it an airport, retail store or other public space. In fact, 3000 missing children were Discovered In Just Four Days Using Face Recognition In India!

7.5) Help the Blind

Listerine has developed a ground breaking facial recognition app that helps the blind using face recognition. The app recognizes when people are smiling and alerts the blind person with a vibration. This can help them better understand social situations.

7.6) Protect Law Enforcement

Mobile face recognition apps, like the one offered by FaceFirst, are already helping police officers by helping them instantly identify individuals in the field from a safe distance. This can help by giving them contextual data that tells them who they are dealing with and whether they need to proceed with caution. As an example, if a police officer pulls over a wanted murderer at a routine traffic stop, the officer would instantly know that the suspect may be armed and dangerous, and could call for reinforcement.

VIII. CONCLUSION

Face emotion recognition and 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. [10]

REFERENCES

[1] Efficient Face Recognition System for Operating in Unconstrained Environments Alejandra Sarahi Sanchez-Moreno, Jesus Olivares-Mercado, Aldo Hernandez-Suarez, Karina Toscano-Medina, Gabriel Sanchez-Perez and Gibran Benitez-Garcia, J. Imaging 2021, 7, 161. https://doi.org/10.3390/jimaging7090161

[2] The Elements of End-to-end Deep Face Recognition: A Survey of Recent Advances Hang Du, Hailin Shi, Dan Zeng, Xiao-Ping Zhang, Tao Mei Accepted for publication in ACM Computing Surveys Computer Vision and Pattern Recognition (cs.CV)

[3] The Elements of End-to-end Deep Face Recognition: A Survey of Recent Advances HANG DU, HAILIN SHI, DAN ZENG, XIAO-PING ZHANG, TAO MEI, JD AI arXiv:2009.13290v4 [cs.CV] 27 Dec 2021, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 6, 2018.

DOI 10.5281/zenodo.14790039

[4] Study of Face Recognition Techniques: A Survey Madan Lal, Kamlesh Kumar, Rafaqat Hussain Arain, Abdullah Maitlo, Sadaquat Ali Ruk, Hidayatullah Shaikh

[5] Research on Face Recognition Classification Based on Improved GoogleNet Zhigang Yu, Yunyun Dong, Jihong Cheng, Miaomiao Sun , and Feng Su Hindawi Security and Communication Networks Volume 2022, Article ID 7192306, 6 pages https://doi.org/10.1155/2022/7192306

[6] 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).

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

[8] Scott Brave and Clifford Nass, Emotion in Human- Computer Interaction. Retrieved from http://lrcm.com.umontreal.ca/dufresne /COM7162/EmotionHumanInteraction.pdf

[9] 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.

[10] 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

DOI 10.5281/zenodo.14790039