

Study on Video Interview Analysis

Rani Bane¹ Yashoda Eknarayan² Prasad Ambekar³

^{1,2,3}Department of Information Technology

^{1,2,3}Vasantadada Patil Pratishthan's College of Engineering and Visual Arts, Mumbai, Maharashtra, India

Abstract— Due to the rapid growth of the population, the number of individuals opting for job opportunities has also increased. Now, it is the role of a company to distribute these opportunities to the deserving candidates. It has become difficult for a company to select suitable candidate that better fits their requirements. Similarly, it is also difficult for candidates to grab these opportunities as they may lose them due to the biased selection process by human intelligence. So there is a need of the system which can handle entire recruiting process and which will be beneficial for both companies and candidates. With the immense advancement in technology, variety of job recruitment systems are available which gears up the recruiting process. Even though these systems have automated the recruiting process, they still have few drawbacks. These systems have eliminated the manual work upto some extent but still there's a need of the system that will eliminate it upto greater extent. Such system can be developed using web application integrated with different machine learning algorithms, where candidates can give interview through video recordings. Then these videos will be analysed by the system where various facial expressions and sentiments can be detected and voice can also be analysed. Accordingly, candidates will be selected and rejected based on the above statistics and that will enhance the recruitment process and manual work can be eliminated upto greater extent.

Keywords: Convolution Neural Network, Facial Emotion Recognition, Voice Analysis, Video Analysis

I. INTRODUCTION

Job Recruitment is considered as a big challenging process for a company. Traditionally, hiring procedure used to involve a lots of people, paper work and time. With the assistance of Information Technology and automated softwares this processes can be automated. So, to cut this cost and to save the time required for the recruitment, lots of companies have shifted from traditional recruitment process to the automated recruitment process. Currently, there are multiple approaches are available that helps in hiring procedure, such as various job portals and job recruitment softwares that provides the virtual environment for conducting the interviews. Using these job portals companies can include available job vacancies and candidates can apply for the post they desire. But even after this, the company has to hold time consuming meetings to go through these applicants. So, clearly, this is not enough, as recruitment process involves lots of tests, multiple rounds of interviews for which the company needs to do the manual work. Nowadays, even though the interviews are being online, such as audio conferences over telephone or through video calls, the manual work still needs to be done by Human Intelligence, as they need to monitor all these interviews. So, by considering the drawbacks of available resources, a new system can be developed using deep learning algorithms, Natural Language Processing [NLP] algorithm and other

various machine learning algorithms. Using these algorithms the videos that are recorded by virtual interviews can be analysed considering various parameters such as Facial Expression, Voice analysis and Sentiment Analysis etc., thereby deciding selection and rejection of candidate.

II. LITERATURE REVIEW

Video Processing for Facial Expression detection, Voice analysis and sentiment analysis can be done using deep learning algorithms, machine learning algorithm and some other algorithms. For this, training dataset and testing dataset must be prepared accurately and then applying suitable algorithm on the dataset results must be generated.

- [1] The authors [1] introduced a TAN classifier that learns the dependencies between different facial features. This TAN classifier can be used for facial expression recognition from the recorded videos. They also presented algorithm that can be used to find best TAN structure, using which they obtained improved results over simple NB models. They concluded that, by reducing facial expression recognition problem to mood recognition problem the classification results are more precised.
- [2] Authors Laurent Son Nguyen, Daniel Gatica-Perez used dataset of 939 conversational English-speaking video resumes hosted on YouTube making their work to constitute the first computational study on video resumes. They have analyzed conversational video resumes, in which they analyzed formation of personality and hirability impressions. work
- [3] Using MTCNN Algorithm authors [3] proposed emotion analysis on video recordings. They have used Cohn Kanade dataset from which 31000 images are used for training and 1006 images are used for testing. Training was done by considering 4952 images of angry emotion, 5120 of fear, 8989 images of happy, 6077 images of sad, 4002 images of surprise, 6197 images of neutral emotion. The model is tested with 106 images of different emotions. Their study can be used in customer behavior analysis, video surveillance, counseling, business advertising etc.
- [4] The paper [4] is more of Literature Review paper. In which authors presented comparative study on different techniques of video based facial and emotions recognition using different methods. As Facial expression and emotion recognition plays very important role in daily communication the authors have provided advantages, disadvantages and accuracy performance of the methods studied from 2015 to 2016. The current methods research problems are defined in this paper.
- [5] Authors have used the Natural Language Processing [NLP] approaches to identify the candidate concepts from job posts and resume. They proposed an automatic online recruitment system based on coupling multiple semantic resources and statistical concept-relatedness

measures. Then, statistical based concept-relatedness measures are utilized to refine the lists of extracted concepts.

III. ALGORITHM

A. Convolution Neural Network

Convolution Neural Network (CNN) is an algorithm of Deep Learning which takes images as an input. It then assigns learnable weights and biases to various characteristics or objects in the image and is able to differentiate one from the other. CNN can learn these characteristics or objects and it can process them accordingly. The pre-processing required in a CNN is much lower as compared to other classification algorithms. CNN algorithm converts the images into a different configuration which can be easy to process, without criticizing the prior characteristics of an original image so that the result will be more precise.

CNN algorithm consists of multiple layers, in which Convolution and Pooling layers are the layers of convolution process in which the features of an image are extracted and then this output is given to further layers.

- 1) Convolution Layer: This is the first layer that is used to extract the various features from the input images. This layer performs convolution mathematical operation between the input image and a filter of a particular size. The output of this layer is referred as "Feature map". This feature map gives us information about the image (corners and edges). Later, this feature map is given as an input to other layers to learn several other features of the input image.
- 2) Pooling Layer: Convolutional Layer is followed by a Pooling Layer. It acts as an intermediate between Convolution Layer and FC layer. It decreases the size of an convoluted image to reduce the computational cost. This is performed by decreasing the connections between layers and independently operates on each feature map.
- 3) Fully Connected Layer: The Fully Connected (FC) layer consists of the biases and weights. In this layer, the input images from the previous layers are flattened, that is their size is reduced. Then these flattened vector images undergoes few more FC layers, performing the mathematical functions operations on flattened images.
- 4) Dropout Layer: This layer is used to avoid Overfitting in training dataset.

CNN algorithm is the efficient way which helps to extract different features of images and then classify them accordingly, thus providing the more precise results.

IV. DATASET

A. Dataset FER2013

For Facial Expression detection FER2013, Kaggle Dataset is used. This kaggle dataset was prepared by Aaron Courville and Pierre-Luc Carrier. The training set consists of 28,709 images. The public test set used consists of 3,589 images. Using this dataset, the expressions which can be detected are 'Fear', 'Disgust', 'Angry', 'Happy', 'Sad', 'Neutral' and 'Surprised'. The dataset consists of images which are of 48x48 pixel and it has grayscale images of faces. The faces

in the images are automatically registered. Faces are more or less centered. Images occupies about the same amount of space in every image in the dataset. Using CNN algorithm these images can be categorized into seven types of emotions.

The dataset needs to be divided into training dataset(train.csv) and testing dataset(test.csv). Training dataset contains two columns, "emotion" and "pixels". The "emotion" column contains a number ranging from 0 to 6, representing seven emotions. The "pixels" column contains a string surrounded in quotes for each image. test.csv contains only the "pixels" column. Using these the emotions of an image can be predicted.

It is an open Source dataset, which is very efficient for facial expression analysis generating more precise results.

V. CONCLUSION

This study concludes that Deep Learning algorithm like CNN can be used for analysis of multiple features such as facial expression detection and voice analysis.

In CNN algorithm, the greyscale images undergoes the convolution and pooling layer multiple times, thereby providing more accurate results of expression detection.

Whereas, Sentiment analysis can be done using machine learning algorithms. Using the proper dataset like FER2013 and dividing the dataset into proper proportions as training dataset and testing dataset and then applying the suitable algorithm to it, more accurate and precise results can be generated.

REFERENCES

- [1] Ira Cohen, Nicu Sebe, Ashutosh Garg, Michael S. Lew, Thomas S. Huang. "Facial Expression Recognition from Video Sequences." 0-7803-7304-9/02/\$17.00 2002 IEEE
- [2] Laurent Son Nguyen, Daniel Gatica-Perez "Hirability in Wild: Analysis of Online Conversational Video Resumes" Ieee Transactions on Multimedia, Vol. 18, No. 7, July 2016.
- [3] Thenmozhi S, Smeeta Mahabaleshwara Hegde "Emotion Extrication and Analysis on Videos" 2019 IEEE International Conference on Cloud Computing in Emerging Markets (CEEM).
- [4] Husam Salih, Lalit kulkarni. "Study of Video Based Facial Expression and Image Recognition Methods." International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC 2017).
- [5] Aseel B. Kmail, Mohammed Maree, Mohammed Belkhatir, Saadat M. Alhashmi. "An Automatic Online Recruitment System based on Exploiting Multiple Semantic Resources and Concept-relatedness Measures" 2015 IEEE 27th International Conference on Tools with Artificial Intelligence.