ABSTRACT

In terms of e-commerce, recommender framework for the most part guides customers in a customized manner towards interesting / customized manner towards products in a large space of feasible options. To provide reliable recommendation, the recommender systems need to capture the exact customer's need and preference into a user profile. However, for complexes products/ services such as movies, music, news, user emotions play surprisingly critical roles in the decision making process. The traditional method of user profile system doesn't consider the impact of user emotion, the recommender systems cannot understand and capture the continually developing inclinations of a user. The Movie recommendation is based on notions with regards to a client ' s emotions and inclinations. This project additionally examines the System architecture and its implementation, as well as its evaluation procedure. We believe that our system provides improved recommendation to users because it enables the users to understand the relation between their emotional states and the recommended movies.

TABLE OF CONTENTS

Chapter No.	Title	Page N
	Abstract List of Figures List of Tables	vi viii ix
1.	INTRODUCTION	
	1.1 Outline of the Project	1
2.	LITERATURE SURVEY	2
3.	Aim And Scope of The Present Investigation	
	3.1 Purpose of the Project	4
	3.2 Project Architecture	5
4.	Experimental Or Materials and MethodsAlgorithms Used	
	4.1 Introduction	6
	4.2 Working Explanation	7
	4.2.1 Notations	7
	4.2.2 Dataset	7
	4.2.3 Convolutional Neural Networks	8
	4.2.4Support Vector Machine	10
	4.2.5Movie Recommender System	11
5.	Results And Discussion, Performance Analysis	13
6.	CONCLUSION and SUMMARY	16
7.	APPENDIX	
	A. Source Code	17
	B. Output	27
8	REFERENCES	28

LIST OF FIGURES

Figure No.	Title	Page No.
3.1	Architecture of the model	5
3.2	Architecture of the model	5
4.1	FER-2013	8
4.2	Architecture of traditional Neural Network	9
4.3	Convolutional Layer	9
4.4	Fully Connected	10
4.5	Linearly Separable Data Points	11
4.6	Best Hyper Plane	11
4.7	Movie Recommending Strategy	12
5.1	Accuracy and Loss	13
5.2	SVM vs CNN	13
5.3	Test Image 1	14
5.4	Test Image 2	14
5.5	Test Image 3	15
5.6	Test Image 4	15
6.1	Output	27

CHAPTER 1

INTRODUCTION

1.1 OUTLINE OF THE PROJECT

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

Image recognition is one of the tasks in deep neural networks.Neural networks are computing systems designed to recognize patterns. Their architecture is inspired by the human brain structure, hence the name. They consist of three types of layers: input, hidden layers, and output. The input layer receives a signal, the hidden layer processes it, and the output layer makes a decision or a forecast about the input data. Each network layer consists of interconnected *nodes (artificial neurons*) that do the computation.

Emotion recognition is the process of identifying human emotion. People vary widely in their accuracy at recognizing the emotions of others. Use of technology to help people with emotion recognition is a relatively nascent research area. Generally, the technology works best if it uses multiple modalities in context. To date, the most work has been conducted on automating the recognition of facial expressions from video, spoken expressions from audio, written expressions from text, and physiology as measured by wearables.

CHAPTER 2

LITERATURE SURVEY

“Recommender systems could be any framework which produce personalized suggestions as result or guides clients in a customized way useful products in a option space” [Burke 2000]. “The recommender systems are used by business in E-commerce for helping the client in their decision-making by recommending items and providing information” [Schafer et al. 2001]. The three main components of recommender systems are:

- The information of the product;
- The information that user should communicate with the system at the start of the recommending process; and
- An algorithm which combines the information of product and user together to generate a result.

There are five different recommendation techniques, but for advanced and subjective domains like film, music and news, the foremost appropriate techniques square measure cooperative, content-based, knowledge-based or the mixture of those techniques. These systems will establish cross-genre niche while not domain data.

The system makes suggestions by laying out the things that are liked by the client and the information of products to be suggested. The recommendations may be created even though the system has received a variety of ratings. However, content-based systems square measure restricted by the options that square measure expressly related to the objects that they suggest [Burke 2002]. For example, content based suggestion for film could be founded on composed materials of a few film: actors’ names, plot summaries, etc.

Moreover, it is impossible to consider all features associated to subjective andcomplex products such as movie, music or news. Anexample of content-based filtering system is Reel.com'sMovie Matches (www.reel.com).Knowledge-based recommendation systems suggestproducts based on inferences about a user's needs andpreferences. These systems have no start-up problemsand do not require user ratings. However, knowledgeacquisition is very difficult, for example the systemPickAFlick [Burke 2000].

Hybrid recommendation systems mix several recommendation methods to realize better execution with lesser downsides. The user profile not solely contains the general ratings of films, but also includes desired features of the movie. The system presumes that a user's perspective toward a selected feature would possibly vary with the amount of the feature in question. as an example, a user might need a high opinion of slapstick comedy in tiny amounts however not in huge amounts.

CHAPTER 3

AIM AND SCOPE OF THE PRESENT INVESTIGATION

3.1 PURPOSE OF THE PROJECT

The main goal of this project is to build a neural network model which gives an accuracy more than what we have received in previous works as mentioned in the literature review on the same dataset (i.e., FER2013) and be able to work dynamically for movie recommendation. The dataset chosen for the project is FER2013. The pre-trained models from transfer learning will be used to achieve the target. E.g.: Resnet50, MobileNetV2 etc. The model should be able to predict every emotion with similar accuracy and none of the emotions would be excluded. Here we have tried to build a model that should be able to do this on a real-time basis using a webcam or a camera and the time taken to generate the results be minimized as much as possible to facilitate real-life usage.

Disadvantages:

A few types of images the model tends to do poorly on include:

- The human in an unusual position
- Face appears against a background of a similar colour
- Brightness of the picture
- Scale variation (face is very large or small in image)

Advantages:

The model has higher prediction than the basic machine learning model.
This is simple algorithm built using simple and open-source packages like TensorFlow and SciPy library.

3.2 PROJECT ARCHITECTURE:

ARCHITECTURE OFTHE MODEL:

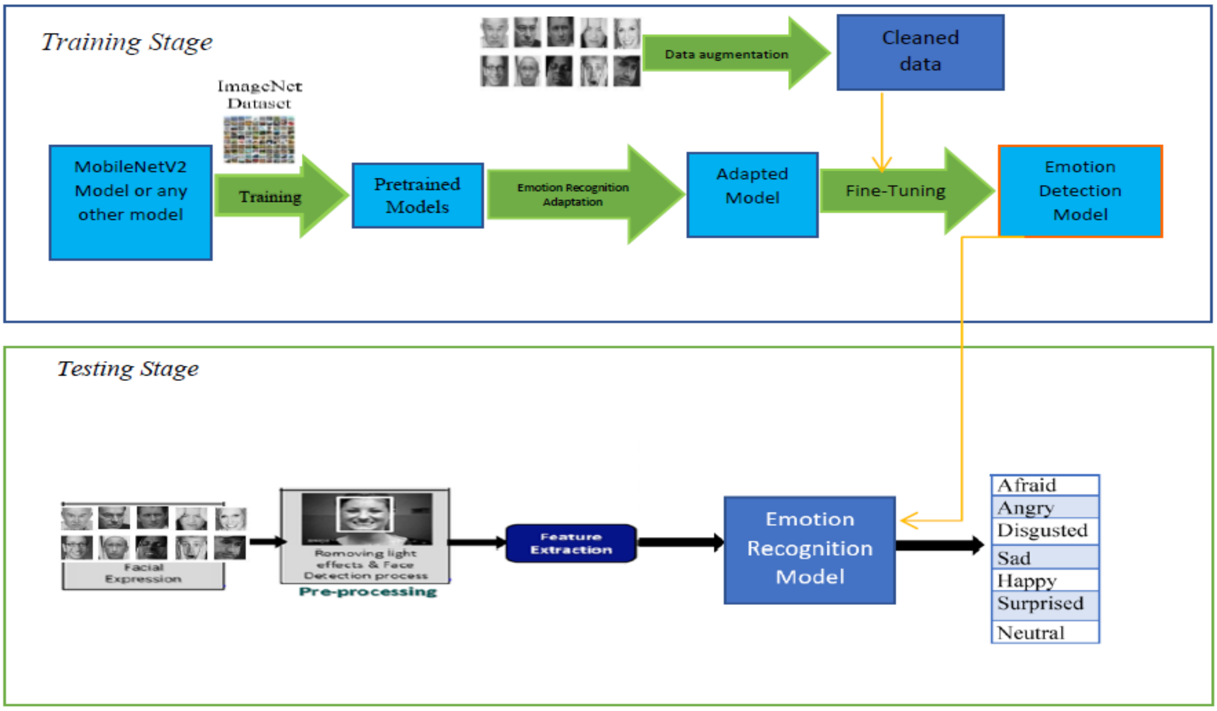

Figure 3. 1

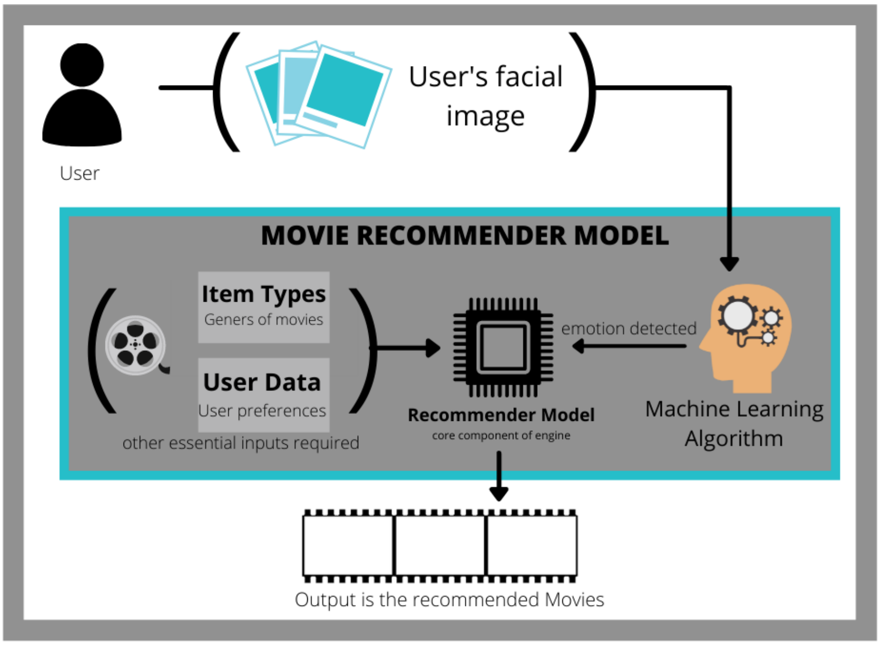

Figure 3.2

CHAPTER 4

EXPERIMENTAL OR MATERIALS AND METHODS

ALGORITHMS USED

4.1 INTRODUCTION

Image recognition is one of the tasks in deep neural networks.Neural networks are computing systems designed to recognize patterns. Their architecture is inspired by the human brain structure, hence the name. They consist of three types of layers: input, hidden layers, and output. The input layer receives a signal, the hidden layer processes it, and the output layer makes a decision or a forecast about the input data. Each network layer consists of interconnected *nodes (artificial neurons*) that do the computation.

Traditional neural networks have up to three hidden layers, deep networks may contain hundreds of them. So, to be able to recognize faces, a system must learn their features first. It must be trained to predict whether an object is X or Z. Deep learning models learn these characteristics in a different way from machine learning models. That’s why model training approaches are different as well. Each layer of nodes trains on the output (feature set) produced by the previous layer. So, nodes in each successive layer can recognize more complex, detailed features visual representations of what the image depicts. Such a “hierarchy of increasing complexity and abstraction” is known as *feature hierarchy.*

So, the more layers the network has, the greater its predictive capability.

4.2 WORKING EXPLANATION:

4.2.1 Notations:

$$(x, y) = \{\left(X^{(1)}, y^{(1)}\right), \left(X^{(2)}, y^{(2)}\right), \left(X^{(3)}, y^{(3)}\right), ..., \left(X^{(m)}, y^{(m)}\right)$$

Where $X \in R^{n_x}$, $y \in \{0, 1\}$

$$X = \left(\uparrow \ \uparrow \ \uparrow \ X^{(1)}\, X^{(2)} \ ... \ X^{(m)} \ \downarrow \ \downarrow \ \downarrow \ \right)$$

Here, (X(1), X(2), …,X(m) are columns)

$$y = [y^{(1)}, y^{(2)}, ..., y^{(m)}]$$

$$W = \left(\uparrow \ \uparrow \ \uparrow \ W^{(1)}\, W^{(2)} \ ... \ W^{(m)} \ \downarrow \ \downarrow \ \downarrow \ \right)^T$$

Where W is considered as the weight matrix usually initialized to 0.

$$A^{[1]} = \left(\uparrow \ \uparrow \ \uparrow \ a^{[1](1)}\, a^{[1](2)} \ ... \ X^{[1](m)} \ \downarrow \ \downarrow \ \downarrow \ \right)$$

Where A represents the Hidden Layer number, [i] represents the ith hidden layer number, and (i) represents the ith element number.

$$Z^{[1]} = \left(\uparrow \ \uparrow \ \uparrow \ z^{[1](1)}\, z^{[1](2)} \ ... \ z^{[1](m)} \ \downarrow \ \downarrow \ \downarrow \ \right)$$

Where Z matrix contains the finally calculated result of [i]th layer.

Dimensions:

Dimension of X: $m \times n_x$

4.2.2Dataset

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image.

The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

Created in 2013, FER-2013 dataset contains 35,887 grayscale 48x48-pixel images, with the images being stored in a spreadsheet where each image's pixel values are stored in cells per row. The images were obtained utilizing Google search and are later grouped per emotion classes, which are anger, disgust, fear, happiness, neutral, sadness, and surprise. As the dataset was built utilizing Google search, the images are in in-the-wild condition, with even very few images being animated characters.

The dataset originally had data distribution of 28,709 images for training and 3,589 images for public test, but after the competition ended, another 3,589 images which were used for private test were added to the dataset. The usages of FER-2013's data distribution vary among published researches, with each using public test images for different purposes, either as part of training set, validation set, or test set.

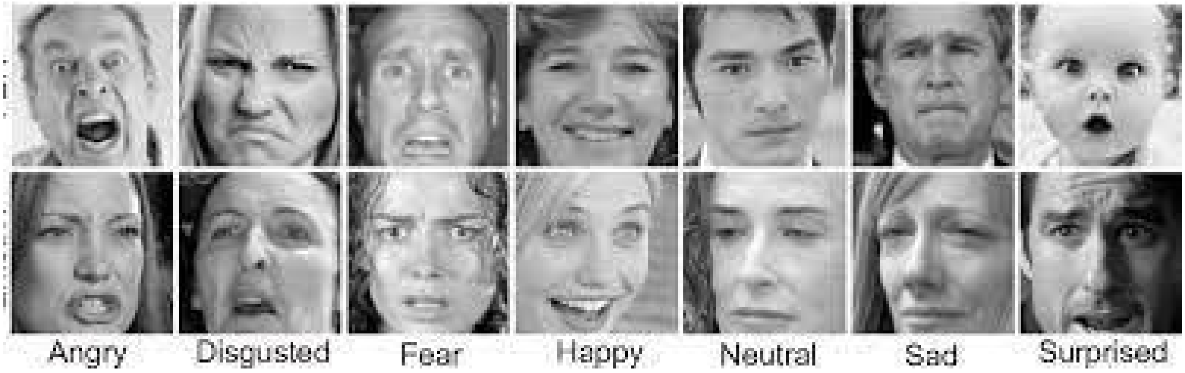

Figure 4. 1 FER-2013

4.2.3 Convolutional Neural Networks

*Architecture of a traditional CNN:*Convolutional neural networks, also known as CNNs, are a specific type of neural networks that are generally composed of the following layers:

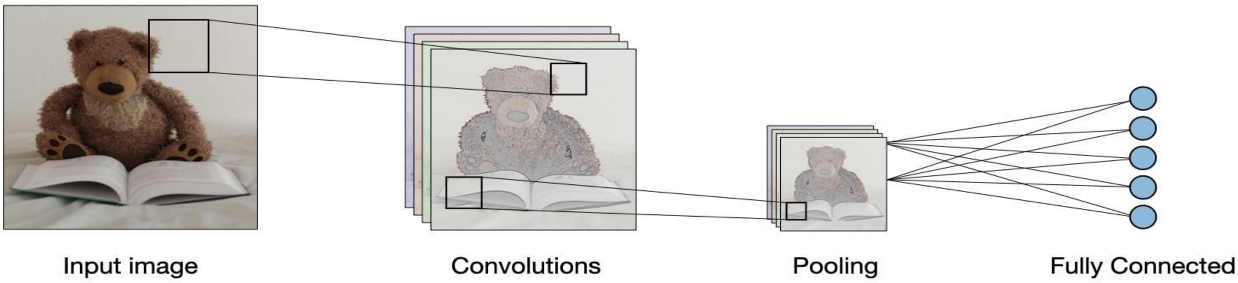

Figure 4. 2 Architecture of traditional Neural Networks

Types of layers: