Table of Contents

ABSTRACT

Communication through voice is one of the main components of affective computing in Computing in human-computer interaction. In this type of interaction, properly comprehending the meanings of the words or the linguistic category and recognizing the emotion included in the speech is essential for enhancing the performance. In order to model the emotional state, the speech waves are utilized, which bear signals standing for emotion such as happy, sad, fear, neutral. This project is aiming to design and develop speech based emotion reaction (SER) prediction system, where different emotions are recognized by means of Convolutional Neural Network (CNN) classifiers. Spectral features extracted is mel-frequency cepstral (MFCC). Librosa package in python language is used to develop proposed algorithm and its performance is tested on taking Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) samples to differentiate emotions such as happiness, surprise, anger, neutral state, sadness, fear etc. Feature selection (FS) was applied in order to seek for the most relevant feature subset. Results show that the maximum gain in performance is achieved by using CNN.

Problem Statement

Emotions play a vital role in communication, the detection and analysis of the same is of vital importance in today’s digital world of remote communication. Emotion detection is a challenging task, because emotions are subjective. We define a SER system as a collection of methodologies that process and classify speech signals to detect emotions embedded in them. Such a system can find usein a wide variety of application areas like interactive voice based-assistant or caller-agent conversation analysis. In this study we attempt to detect underlying emotions in recorded speech by analyzing the acoustic features of the audio dataof recordings. In this project, we will predict the emotion in the speech of a person’s audio on the given dataset using CNN and deep learning algorithms. The dataset consists of 12,800 audio files of 12 male and 12 female voices with different emotions like happy, anger, sad, surprise, neutral, fear, disgust. The major goal of the proposed system is understanding Convolutional Neural Network, and predictingEmotion based on model.

LIST OF FIGURES

1.INTRODUCTION

The human brain is an intricate organ that has been a lasting inspiration for research in Artificial Intelligence (AI). The neural networks in brain had the capability of learning all concepts from experiencing low level information and is remembers them which are processed by sensory periphery. Learning language,understanding speech, and recognizing faces are some examples that manifest the remarkable power of the human brain in learning high-level concepts. The main goal of AI is to develop intelligent systems that are able to generate rational thoughts and behaviour similar to human thoughts and performance. Emotion plays a significant role in daily interpersonal human interactions. There are several modalities for expressing human emotions like body-posture, facial expression & voice. Out of which speech is very significant in expressing emotions. In order to communicate effectively with people, the systems need to understand the emotions in speech. A lot of machine learning algorithms have been developed and tested in order to classify these emotions carried by speech. The aim to develop machines to interpret paralinguistic data like emotion, helps in human-machine interaction. The approach for speech emotion recognition (SER) primarily comprises two phases known as feature extraction and features classification phase. The first phase Feature extraction is the key part in the Speech Emotion Recognition. The quality of the features directly influences the accuracy of classification results. Typically, the Feature Extraction method designs handcraftfeatures based on acoustic features of speech. The second phase includes feature classification using linear and non-linear classifiers. The most commonly used linear classifiers for emotion recognition include the Maximum Likelihood Principle (MLP) and Support Vector Machine (SVM) and Convolution Neural Network (CNN). Usually, the speech signal is considered to be non-stationary. Hence, it is considered that non-linear classifierswork effectively for SER.

1.1 Machine Learning

Machine Learning is the most popular technique of predicting or classifyinginformation to help people in making necessary decisions. Machine Learning algorithms are trained over instances or examples through which they learn from past experiences andanalyse the historical data.

Simply building models is not enough. You must also optimize and tune the modelappropriately so that it provides you with accurate results. Optimization techniques involvetuning the hyperparameters to reach an optimum result.

As it trains over the examples, again and again, it is able to identify patterns in orderto make decisions more accurately .Whenever any new input is introduced to the ML model, it applies its learned patterns over the new data to make future predictions. Based on the final accuracy, one can optimize their models using various standardized approaches. In this way, Machine Learning model learns to adapt to new examples and produce better results.

Simply building models is not enough. You must also optimize and tune the modelappropriately so that it provides you with accurate results. Optimization techniques involve tuning the hyperparameters to reach an optimum result.As it trains over the examples, again and again, it is able to identify patterns in orderto make decisions more accurately .Whenever any new input is introduced to the ML model, it applies its learned patterns over the new data to make future predictions. Based on the final accuracy, one can optimize their models using various standardized approaches. In this way, Machine Learning model learns to adapt to new examples and produce better results.

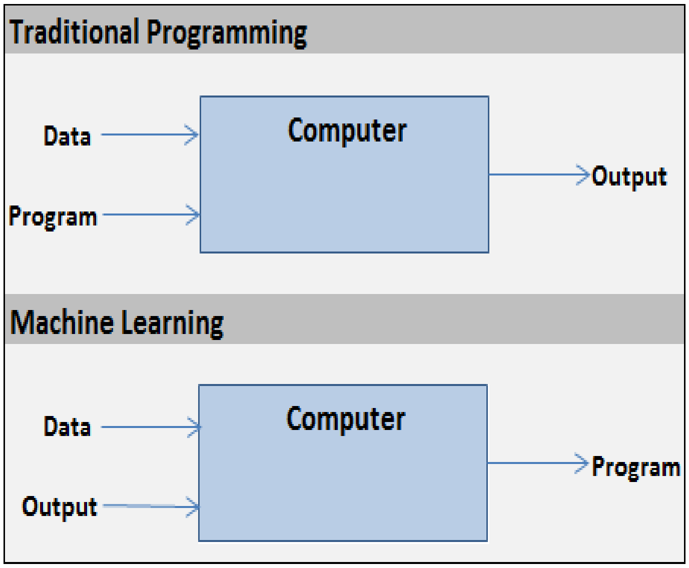

Fig 1.1 Machine Learning vs Traditional Programming

1.2 Types of Learning

Machine Learning Algorithms can be classified into 3 types:

1. Supervised learning
2. Unsupervised Learning
3. Reinforcement Learning

1.2.1 Supervised Learning

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data and on basis of that data machines predict the output the labelled data means some input data is already tagged with the correct output.In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as student learns in the supervision of the teacher.

Supervised learning is the most popular paradigm for machine learning. It is the easiest to understand and the simplest to implement. It is the task of learning a function thatmaps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value(also called the supervisory signal). A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. Supervised Learning is very similar to teaching a child with the given data and that data is in the form of examples with labels, we can feed a learning algorithm with these example- label pairs one by one, allowing the algorithm to predict the right answer or not. Over time, the algorithm will learn to approximate the exact nature of the relationship between examples and their labels. When fully trained, the supervised learning algorithm will be ableto observe a new, never-before-seen example and predict a good label for it.

Most of the practical machine learning uses supervised learning. Supervised learning is where you have input variable (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.Y = f(x)

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for the data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. Supervised learning is often described as task oriented. It is highly focused on a singular task, feeding more and more examples to the algorithm until it can accurately perform on that task. This is the learning type that you will most likely encounter, as it is exhibited in many of the common applications like Advertisement Popularity, Spam Classification, face recognition**.**

data, unsupervised learning, also known as self-organization, allows for modelling of probabilitydensities over inputs.

Unsupervised machine learning algorithms infer patterns from a dataset without reference to known, or labelled outcomes. It is the training of machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data. Unlike supervisedlearning, no teacher is provided that means no training will be given to the machine. Therefore, machine is restricted to find the hidden structure in unlabelled data by our-self.For example, if we provide some pictures of dogs and cats to the machine to categorized, then initially the machine has no idea about the features of dogs and cats so it categorize them according to their similarities, patterns and differences. The Unsupervised Learning algorithms allows you to perform more complex processing tasks compared to supervised learning. Although, unsupervised learning can be more unpredictable compared with other natural learning methods.Unsupervised learning problems are classified into two categories of algorithms:

- **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.
- **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

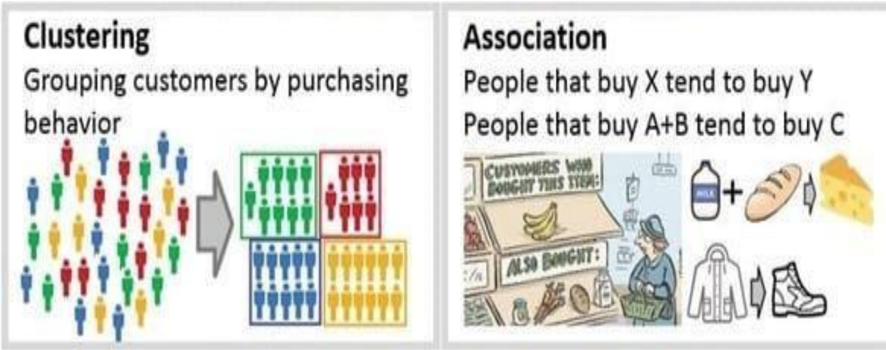
Clustering
Grouping customers by purchasing behavior
Association
People that buy X tend to buy Y
People that buy A+B tend to buy C
CUSTOMERS WHO BOUGHT THIS ITEM:
ALSO BOUGHT:
MILK

Fig1.2.2: Clustering & Association