A COMPUTERIZED APPROACH FOR EMOTION RECOGNITION USING VOICE

R.A.D.G.D.KUMARA

189463P

Faculty of Information Technology

University of Moratuwa

Sri Lanka

August 2021

Msc in IT, Faculty of IT, University of Moratuwa

I the undersigned announce that this project report entitled **"A COMPUTERIZED APPROACH FOR EMOTION RECOGNITION USING VOICE**" is based on my work done while studying under the direction of **Mrs. Indika Karunarathne** .I affirm the statements made and the conclusions drawn are the result of my research work. I continue to confirm that

II. The work contained in this report is first and foremost done by me under the supervision of my supervisor.

III. The work has not yet been submitted to any other institution for any other degree / diploma / certificate at this university or from any University of India or abroad.

IV. We have followed the guidelines provided by the university in writing this report.

V. Whenever we use building materials (data, theoretical analysis, and documents) from other sources, we provide them with the appropriate credit in the report text and provide their details in the index.

Student Name R.A.D.G.D Kumara

Student ID-189463P

Supervisor Name- Mrs. Indika Karunarathne

Supervisor Signature

Date

Acknowledgement

This project would not have been possible without the support of many people. Many thanks to my adviser, Mrs. Indika Karunarathne, who read my numerous revisions and helped make some sense of the confusion. I thank the University of Moratuwa, Faculty on Information Technology for giving me a Dissertation Completion Fellowship to complete this work. And finally, I thank the parents, and the many friends who endured this long process with me, always giving of love and affection.

ABSTRACT—

The purpose of this paper is providing guess information about emotion status of human using voice cut, previous machine learning researches for emotion status of human using voice cut and accuracy and recall of each research and comparison of current research and previous related researches. deeply discuss how to use machine learning for emotion status of human using voice. Firstly, explain ML using many categories Supervised learning algorithms, Unsupervised learning algorithms like that logistic regression, random forest, svm, gaussianNB, decision tree and k-nn .Python scipy provide facilities for analyze voice and provide frequency spectrum and amplitude of voice . fast Fourier transform (FFT) is providing maximum amplitude from all frequencies and frequency of maximum amplitude. The mel scale is a scale of pitches that human earshot usually observes to be intermediate from each other. Maximum mel value, maximum amplitude from all frequencies, frequency of maximum amplitude, age of user and gender are used to find out emotion status human. Part of this research is analyzing various type of voices according to emotional situation. Then train and test model using Machine learning for predict what emotional situation of is given voice. Elicited emotional speech database are creating emotional situation artificially by accumulating fact from the talker. emotional states are identified by researcher after features are trained and evaluated by machine learning models. The best recognition rate is reported by SER between 75% and 82% on random forest and recognition rate between 75% and 77% on decision tree. From this result, we can see that Decision-making algorithms (random forest and decision tree) often perform better with our data base. Therefore, we concluded that the K-Nearest Neighbor (accuracy 50%) , svm (46%), GaussianNB (46%) and logistic regression (46%). Rule based algorithms are good for recognize emotion status using voice. Random forest is best machine learning algorithm for recognize emotion status using voice according to our experiment. Random forest has good test accuracy (between 80% to 90%).

Table of contents

List of figures

Chapter 01

1. Introduction

Emotions are frequently disorderly with state of mind and tempers, but the three reputations are not transposable. American Psychological Association (APA) are saying, emotion is well-define as “a complex reaction pattern, involving experiential, behavioral and physiological elements.” Emotions are identified by human as a how to beings contract with substances or environments they discovery himself important. Researcher can identify three part of Emotional practices: a subjective practice, a physiological reply and a interactive or sensitive reply.

emotional experience is increasing by feelings. Because a person is minding their experience and person mind is classified in the similar class as scarcity or agony. emotion is manufacturing feelings and may be predisposed by remembrances, opinions and further features.

A temper is demarcated by the APA as “any short-lived emotional state, usually of low intensity.” emotions vary from moods that’s why they lack impetuses and cannot identify strong initial point. For example, emotion of anger are generating by insults while an angry mood may rise deprived of deceptive origin.

task that defining emotions is not completed yet. theories are proposed by Many researchers about how to build up our emotions, and existing theories are challenged continuously. Still, researcher is wanted better knowledge to analyze when discovering the theme.

Emotions are based on biotic situations that combined with the nervous system transported on by neurophysiological variations variously allied with judgements, feelings, communication responses, and a degree of desire or irritation. Research on emotion is amplified expressively concluded the past two periods with various fields causative with psychology, neuroscience, sentimental neuroscience, endocrinology, medicine, antiquity, sociology of emotions, and computer science.

We are knew how to feel when have our heart beat increase with horror. This bodily reply is the result of the autonomic anxious system's reply to the emotion we were going through. The autonomic nervous system is handled our spontaneous bodily replies and orders our acute stress response. as reported by uncountable psychoanalyst, our physiological replies are likely how emotion are enabled us raise and endure as humans thru history.

Attractively, studies are discovered autonomic physiological replies have securest when a human facial expressions utmost faithfully look like the appearance of the sentiment they're facing. In other words, facial expressions show an important part in answering consequently to an emotion in a bodily sagacity.

Behavioral replies are important signal for identify that how we're feeling to others, but researches are discovered that they are stronger to behaviors happiness. study of Abnormal Psychology was invented when showing negative and positive emotional films, conquest of behavioral replies to emotion obligated physical stuffs on the participants. heart rates are raised by effects included. This guides that articulating behavioral replies to impetuses, both helpful and harmful, is best for your entire healthiness than holding those answers inside. Thus, there are benefits of smiling, laughing and uttering harmful emotions in a well technique.

The bodily and interactive replies are linked with emotions are demonstrated that emotion is much more than a psychological state. Emotion are moved our whole character and our health. Also, our aptitude to comprehend others' behavioral replies are played a main character inside emotional intelligence.

Emotion handle a important part in everyday relational human communications. Emotion is vital to your cogent to boot brainy choices. Its support for event and comprehends the emotional state of neighbors by transmission your feelings and providing reaction to others. Research is opened the strong part and emotion handle in maintaining human social interaction. Emotion was show carry enough information for the mental level of separate. This is opened up a novel research area named as automatic emotion recognition, taking main persistence to recognize and recover anticipated emotions.

In past researches, several strategies are identified emotional states using expression of face, voice record segments, bodily signals, etc. Many essential advantages are made by speech signals a better origin for sentimental calculating. as example, equated to several other biotic signals (e.g., electrocardiogram), dialog indicators are assimilated more willingly and thriftily.

This reason is general to novel thinker are absorbed in speech emotion recognition (SER). SER is purposed to discover the accurate emotional state of a talker from her speech clip segment. The field is acknowledged to growing finding new things are updated since last few years. There are several systems are detected the emotion of the person such as in the communicate with robot machines, audio investigation, web depended online class, profitable applications, medical lessons, acting, funding, call hubs, insubstantial systems, computer competitions, etc. For classroom transposition or web depended online class, evidence about the emotional level of participants are focused on the upgrading of instructing value. For example, instructor can learn SER system to understand how to handle lectures for prevent mistakes and understand how to take attention from students in class room or lecture room. It is very useful for lecturer because learning of student depend on how to feel learn things of student lecture room.

SER system divide to three main sections such as , select a virtuous emotional speech database, discovering operative features, and machine learning algorithms provide classifiers for use for prediction. When researcher analyze, key issue of the SER system is made by emotional feature abstract. Many researchers are tried to find important speech features which extract emotion information, like that power, pitch, formant frequency, Mel-frequency cepstrum coefficients (MFCC), and modulation spectral features (MSFs) . researchers wish to combine collection of features covering more emotional information. merging feature set is increased to top measurement and redundancy of speech features. it is made the learning development are difficulted for most machine learning algorithms and heaves the chance of overfitting. feature selection is important to reduce the dimensions redundancy of features.

In last few years, various classification algorithms are projected by researchers, as example, neural networks (NN),recurrent neural networks (RNN) Gaussian mixture model (GMM) and support vector machine (SVM).

Some researchers are also developed projects such as a modified brain emotional learning model (BEL) in (MLP) and (ANFIS) and has difficulties for speech emotion recognition. Another method is a multiple kernel Gaussian process (GP) classification. researcher was tried to mixing the linear kernel and (RBF) kernel. The Voiced Segment Selection (VSS) approach was planned in indentures with the voiced signal clip as the texture image processing feature which is different from the traditional method.

we was tried to develop a system for the identify of three emotional levels (angry, normal, and sadness). We are discovered the Time series, FFT and Mel spectrogram features and utilized these useful features to instruct six different machine learning paradigms (svm, random forest, decision tree , logistic regression, gaussianNB, k-nn,).

Chapter 02

2. Background and Motivation

Visual features are very famous in real world for many environments. Researchers are used visual features for analyze emotion status of human. User want one equipment for take a photo and time consuming is very less. Disadvantage is some time visual features are not provide accurate positions because some users are use same face position for all emotions Figure 2.1.

Figure 2. 1-face detection

Person has more emotional situations and now we analyze many emotional situations which relate with research. Anger is associated with bodily effects, including raised heart rate (Fig 02), raised blood pressure(Fig 03), and increased dose of adrenaline and noradrenaline. Drawbacks of this method is patient is went to clinic and face to experiments and this method want many equipment for procedure this method.

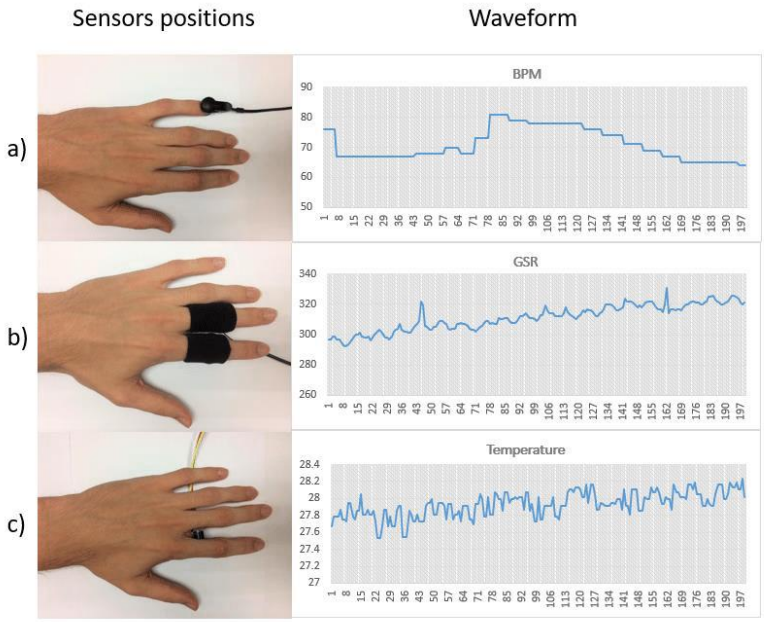

Figure 2. 2-waveform

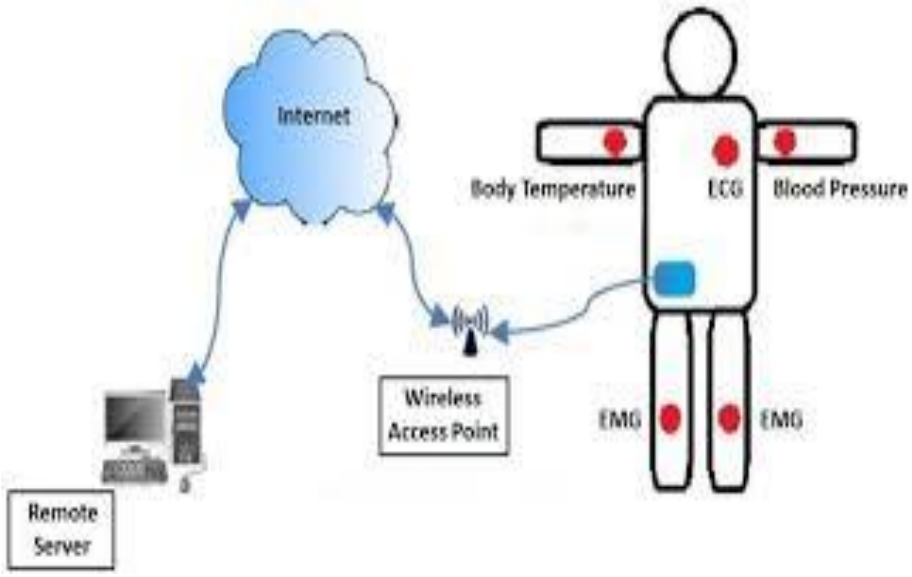

Figure 2. 3-sense of blood

Figure 2. 4

Sadness is feeling when following things are happening like disadvantage, loss, desolation, anguish, feebleness, displeasure and sorrowfulness.

An individual person has sadness then person is tried to noiseless or weary, and try to leave from neighbors. sadness could be converting to depression. Crying can be an indication of sadness can be identify easily when person were crying. Sadness is normal part of life, but this feeling happen frequent mean that could be indications of depression. Example is millions of American adults identified as depression patients every year. if patent is not managed depression in right way, depression disturb your day-to-day life and cause a life threat. Patient is infected by depression attack to his feel and changes body. Major depression can be a attack to health condition and damage quality of patient life.

depression Symptoms are contained awe-inspiring grief, grief, and a sense of guiltiness. This symptom can be explained as a feeling of blankness or fruitlessness. sometime Patient could not convert these feelings into words and themselves problematic for understand that symptoms now are inside of their mind and body.

Both Depression and stress are interrelated with them. Heart rate are speed up Heart rate are speed up Heart rate are speed up by Stress hormones and make blood vessels are tighten by Stress hormones, stress hormones are ready state of emergency. Sometime heart disease is leaded by Frequent Stress hormones. sometime immune system face negative impact from Depression and stress , patient can be more susceptible to poisons and sicknesses

Mind of person is very penetrating for emotional condition. Person face to more problem because they can't identify neighbor's emotional situation. Person want identify emotional satiation of their family members for good family life, emotional situation of employees of their company for success of company. Sometime human body does not provide any health issue but their mind is not in good state. This situation can't identify their neighbors. So people face to many problems without know the reason for this issue and their neighbors also face problems as a example family members office staff. Their neighbors have bad effects If person take wrong decision because of emotional issues. [2]

Purpose of this research paper is identify voice attributes (frequency spectrum and amplitude) according to above emotional situations and predict emotional situation according to voice attributes using machine learning algorithm.

Chapter 03

3. Aim and Objectives

System basically try to automatically understand and retrieve human emotions using voice

Objectives

- find out maximum amplitude with different motions
- compute the Fourier transform and find maximum individual frequency and maximum frequency's amplitude with different motions
- Compute Mel spectrogram and find maximum Mel value with different motions
- we identify how to classifiers are arranged features to recognize accuracy of emotions in speech

Chapter 04

4. Related works

Lucio Ciabattoni, Gabriele Foresi and Marianna Capecci are published research paper name as" A Machine-Learning Based Emotion Recognition System in Patients with Parkinson's Disease". Lucio Ciabattoni is planned to use 36 speakers who has sick (20 male patients and 16 patients) and age range varies among 58 and 80 years old and patients are defined by the clinician. [3]

Smartwatch signals are provided 27 features: in particular, cardiac frequency are related by 21 features (e.g. standard deviation's derivative and standard deviation), skin conductance is associated by 4 features (e.g. Standard deviation values and mean) and to skin temperature is related 2 features (e.g., maximum wrist temperature).

Table I: Accuracies of SVM trained with labels extracted from SAM scores

	0-class Producer Accuracy	1-class Producer Accuracy	Overall Accuracy
SVM-Linear	65.4%	89.5%	77.3%
SVM-RBF	83.7%	94.8%	91.3%
SVM-Polynomial	69.4%	89.7%	76.9%

Table II: Accuracies of SVM trained with labels extracted from the commercial software

	0-class Producer Accuracy	1-class Producer Accuracy	Overall Accuracy
SVM-Linear	50.8%	88.2%	76.4%
SVM-RBF	80.7%	90.2%	88.6%
SVM-Polynomial	52.5%	87.8%	74.8%

Mohit Wadhwa, Anurag Gupta and Prateek Kumar Pandey cut off the delivery of labels on feelings and sexuality and found that although the actualities are limited to six emotions namely. neutrality, happiness, sadness, anger, fear and disgust, the number of labels was too small for surprises and less important for frustration Shared data set from the first 5 factors are logically examined in relation to the following factors Gender distribution [3]

1-D format

These features initiate in audio clips are in matrix format. consider converting matrices to 1-D format by taking row and column methods. When making the first model the results found in the row of rows mean that they were better than most of the column means, so Mohit Wadhwa sustained with the 1-D list obtained by the line matrix methods of the feature. [3]

2D Data Format

The feature-matched y-axis feature is based on the n_mfcc or n_mels argument Mohit Wadhwa displaying when haul out information. The x-axis is based on the length of time the sound and the level of the sample that we are displaying while the element is being removed. obviously, we use different sample values in output according to their audio length. In either case, a larger or 5 second audio file was cut in 5 seconds and sampled at 16000 Hz and briefer clips were taken as samples in such a way that the audio length * of the sample size ruins 80000. In this way, we are able to maintain the matrix size of all audio clips without much detail.
Data sources are TESS (Toronto Emotional Speech Set):, Surrey Audio-Visual Emotion Expression):, RAVDESS: and Berlin. [3]
Youssef Serrestou , Catherine Cleder, Mohamed Mbarki, Kosai Raoof and Mohamed Ali Mahjoub tested the SER (Automatic Speech Emotion Recognition Using Machine Learning) program which covers four important steps. First step is a voice sample gathering. The second step includes a vector made by extracting features. As a next step, we strained to control which factors were most important to differentiate each emotion. These data points are presented in the machine learning classifier for viewing. [4]

In our study, we is discovered the first 12 orders of MFCC coefficients when speech signals were sampled at 16 KHz. For each directional coefficients, we include meanings, variations, general deviations, kurtosis, and skewness, and this is due to all other note frames. The vector of each MFCC element is 60-dimensional. [4]

Suraj Tripathi, Abhay Kumar, Abhiram Ramesh, Chirag Singh and Promod Yenigalla issued the title of the research paper as Deep Learning based Emotion Recognition System Using Speech and Text Features. Speech elements use as the Spectrogram and Mel-frequency

Cepstral Coefficients (MFCC) help to remember the emotion-related aspects of speech while the text helps to capture the semantic meaning, both of which contribute to different aspects of sensitivity. Suraj Tripathi is being examined for multiple Deep Neural Network (DNN) structures, which take different combinations of speech and textual features as input. Systematic network architecture gets high precision when it is associated with technologies in the standard database. The integrated MFCC-Text Convolutional Neural Network (CNN) model has proven to be very accurate in detecting emotions in the IEMOCAP data CNN Model with [5]

Spectrogram input (Model 2A)

Spectrogram size obtained and then scaled to Mel scale to find Mel-spectrograms. 128 Spectrogram coefficients per window are assisted in this model. Mel's frequency scale underscores the narrow end of the frequency spectrum over the complex one, thus imitating the cognitive abilities of the human ear Suraj Tripathi used python library such as a "librosa" , with the above mentioned lines. [5]

CNN Model with MFCC input (Model 3)

MFCC is a widespread speech feature widely exploited in many speech processing programs, we have decided to use the same in our tryouts to find emotions

Combined CNN Model based on both Text and Speech Features[5]

Spectrogram and MFCC model (Model 4A) [5]

Spectrogram and text model (Model 4B)[5]

MFCC and text model (Model 4C)[5]

The Spectrogram model declared in Model 2A provides a complete accuracy of 71.2%, while its varied variation mentioned in Model 2B gives a sensitivity accuracy of 71.3%. The MFCC-

based model also extends the results compared to the previously described models, coming in with a accuracy of 71.6%, approximately 2.5% earlier than the current results. [5]

Chapter 05

5. Methodology

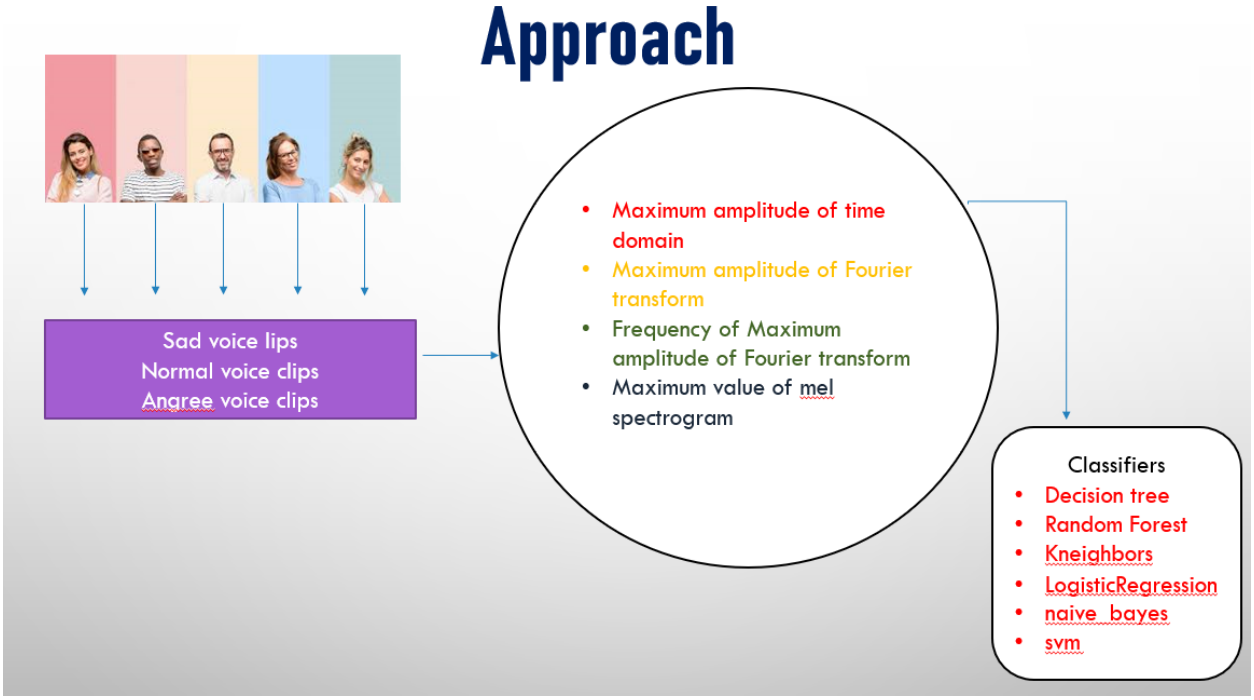

Figure 5. 1(approach)

Data preprocessing

speech database

example of speech database

Berlin database

- 535 voice clips
- 5 actors and 5 actresses

- . The INTER1SP Spanish emotional
- one actors and one actress.

- Basic emotion status →01 anger 02 sadness 03 joy 04 fear 05 disgust 06 surprise 07 neutral/normal

Toronto Emotional Speech Set

- young actress and old actress
- Size- 2800 voice cuts
- seven Basic emotion status.

Surrey Audio Speech Set

- 4 actors,
- size 480 voice cuts
- seven Basic emotion status.

Researcher want basic idea about different kind of databases which relate with SER when researcher select database, suitable speech database play key role for success research. Three type databases can utilize for SER.

1. Elicited emotional speech database- speaker is voiced artificially with different emotional situations

2.Actor based speech database: Trained and professional actors and actresses are voiced different emotional situations

3. Natural speech database: Real different emotional situations are recorded by researcher.

Frequency spectrum

Fourier transform

main downside of Fourier series is, focused to periodic signals only. It is mainly focus on signals such as periodic or nonperiodic, which we are unable to explain Fourier series. To cover this drawback, mathematical model is introduced by Fourier for translate time series to frequency domain which is it is name as Fourier transform'.

Fourier transform has in physics and engineering are provided various applications for Fourier transform as example RADAR, LTI systems, signal processing, stargazing. [6]

Fourier transform (FT) use a scientific translate that divides functions reliant on interplanetary or time into functions reliant on spatial or sequential frequency. Fourier transform is used to both the frequency domain illustration and the scientific calculation that connections the frequency domain illustration to a function of space or time.

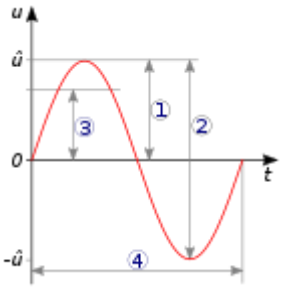

Figure 5. 2-time domain

A sinusoidal curve has following details
maximum amplitude explains using (1)
peak-to-peak explain using (2),
RMS explain using (3)
wave period explains using (4) (Figure 5.2 [7])

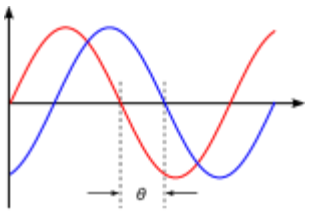

Figure 5. 3

Illustration of phase shift θ (Figure 5.3 [7]

)

.

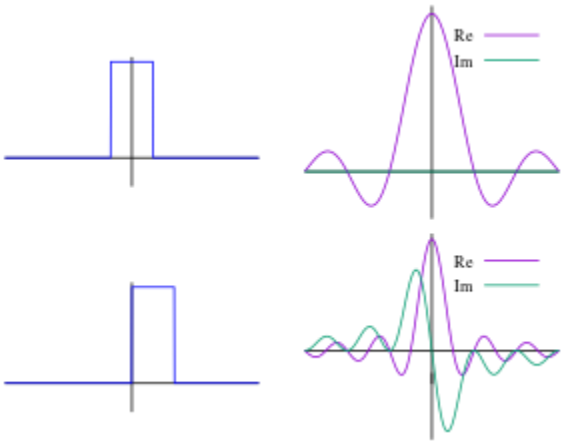

Figure 5. 4-Fourier transform

In the topmost diagram is explained Fourier transform $\hat{f}(\omega)$ and its unit pulse function $f(t)$, a function of frequency ω. Conversion of time domain is taken as difficult phase shifts inside frequency domain. In the second image display $g(t)$, a deferred unit pulse, together with the actual and invented portions of the Fourier transform.

One domain (time or frequency) is obtained by Linear operations by rather than consistent performance on another domain, sometimes quieter gain. The action of the difference in the time domain is similar to the increase in frequency so certain statistical alterations are easier to examine in the frequency domain.

Also, the difficulty in the domain of time is similar to the normal growth in the domain of frequency (see Convolution theorem). After performing the expected tasks, the effect conversion can be reverted to the time domain. Harmonic analysis is a regular study of the relationship between frequency and time periods, and types of tasks or tasks that are “simple” in one or the other, and have bottomless networks to several parts of modern mathematics. [7] Time-restricted activities have Fourier variations that are distributed across the conventional domain and vice versa, something known as the policy of hesitation. A key component of this system is the Gaussian work, which is critical to the concept of opportunity and statistics and to the study of physical conditions that reveal general distribution (e.g., spread). Fourier transformation of the Gaussian work into another Gaussian work [7]

fast Fourier transform (FFT)

Fourier transform transformer (FFT) is a procedure that figures Fourier transform (DFT) sequences, or its contrary (IDFT) variables. Fourier analysis converts the signal from its new domain (usually time or space) to an image that is in the frequency domain and vice versa. DFT is characterized by the fragmentation of prearrangement of values into different frequencies. This work is helpful in various pitches, but doing it directly from the definition is very much restrained and that is realistic. [8]

To do FFT (Fast Fourier Transform), just appear. The human ear benefits randomly and reluctantly from a strategy that requires years of mathematical intelligence to acquire. The ear creates a transformation by translating sound - pressure waves that move over time and into space - into a spectrum, the definition of sound as a sequence of volumes in different pitches. The mind then converts this information into visual sound. [9]

Similar modifications can be made by mathematical approach to the equivalent sound waves or almost any other mutable signal that varies with respect to time. The Fourier Transformation is a methodical tool used to make this transformation. Simply put, Fourier converts data that converts time domain waves into a frequency domain. Fourier modification attains this by dividing the format of the time-based waves in the order of sinusoidal terms, each with a different size, frequency, and phase. This development, in fact, alters the shape of the wave in a time zone that is problematic in accurately defining a functional series of sinusoidal functions that when combined together, mimic the emerging form of innovation. Setting the height of each sinusoidal word against its frequency produces a power spectrum, which is a response to a real wave form in the frequency range. [9]

Mel spectrogram

.

The mel scale is a scale that measure pitches that human can earshot usually observes to be intermediate from each other. As frequency rises, the interval, in hertz, among Mel scale values (or simply mel) increases. The name mel springs from melody and designates that the scale is grounded on the judgement among pitches. The mel spectrogram remaps the values in hertz to the mel scale.

The linear audio spectrogram is preferably suitable for tenders where all frequencies have identical status, while mel spectrograms are improved suitable for applications that need to model human earshot awareness. Mel spectrogram data is also matched for practice in audio classification applications.[10]

A mel spectrogram varies from a linearly mounted audio spectrogram in two ways:

A mel spectrogram logarithmically reduces frequencies upstairs a positive threshold (the corner frequency). For example, in the linearly scaled spectrogram, the vertical space between 1,000 and 2,000Hz is partial of the upright space between 2,000Hz and 4,000Hz. In the mel spectrogram, the space between those ranges is about the same. This scaling is analogous to human earshot, where we discovery it easier to differentiate between parallel low frequency sounds than similar high frequency sounds. A mel spectrogram calculates its output by increasing frequency-domain values by a filter bank.[10]

The sample builds the filter bank from a series of overlying triangular windows at a series of consistently spaced mels. The number of elements in a single frame in a mel spectrogram is equal to the number of filters in the filter bank.

The following image shows the linear audio spectrogram and the mel spectrogram of the same linearly growing and reducing tone. The tone starts at 20Hz, growths to 22,050Hz, and drops back to 20Hz. The image shows that the audio spectrogram represents the objective signal, but the mel spectrogram mirrors human observation, that is, the curve levels and signposts compact difference between high frequencies.[10]

Machine learning classifiers

In real world, “Machine Learning” is very famous name between researches and we use short form as ML, was invented for checking game of inspectors by machine. A computer program that can try to understand to pattern a training that is not clearly interacted by the author of the program. capability of illumination practice that the writer fully unmind sometime. three reasons are case to study: [11]

(1) program is utilized fact

(2) A metric which actions blameworthiness otherwise some structure of gap among the existing condition and expected situation

(3) A comment procedure which exploits computed fault to path the code segment to paradigm well performance in the succeeding procedures.

The machine learning algorithms are generally interested in three categories: [11]
1. Reinforcement learning algorithms
2. Unsupervised learning algorithms
3. Supervised learning algorithms

Supervised Learning algorithm is acted as black box that is build up outcomes after provided input. But programmer cannot identify process of black box. If programmer already identified pattern or ancient fact that mean programmer is contained the outcome collection for an input set, if programming depended from following rules name as supervised learning. Unsupervised Learning has not available labelled fact. Unsupervised learning is based on “clustering.” Reinforcement learning is a specific type of learning method is tried to manage contrarily from unsupervised and supervised methods.

static learning methods is used static labelled facts that used in systems. The systems are used facts that endlessly altering with time that name as dynamic methods. Individual category of

methods may be supervised learning algorithms, or unsupervised learning algorithms, but reinforcement learning algorithms are recurrently dynamic. [11]

Logistic Regression

Logistic regression is widely use in machine learning, Logistic regression categorized as supervised learning. It is used to predict categorical dependent variable under independent variable collection.

Logistic regression classifier is predicted result of a categorical dependent variable. Consequently, the harvest is definitely categorical or discrete answer. Answer is under either Yes / No, 0 / 1, true / False, etc.

.

Logistic Regression is almost identical to Linear Regression regardless of how it is used. Linear Regression is used to solve Regression difficulties, and Logistic regression is benefited to resolve classification problems.

classifier is practically match to Linear Regression no matter how it is utilized. Linear Regression classifier is answered to Regression questions, and Logistic regression classifier is helped to resolve division problems. [12]

Logistic regression classifier is not adjusting the regression line and it measure the "S" designed logistic function, predicting only two values (0 or 1).

Logistic Regression machine learning algorithm name as is an important because classifier contain the ability to create probabilities and categorize novel facts utilizing continuous and different data sets.

Logistic Regression can be used to separate annotations using a variety of facts and can simply control the most effective variables helped for classification. The image below introduces the function of logistic. [12](Fig 17)

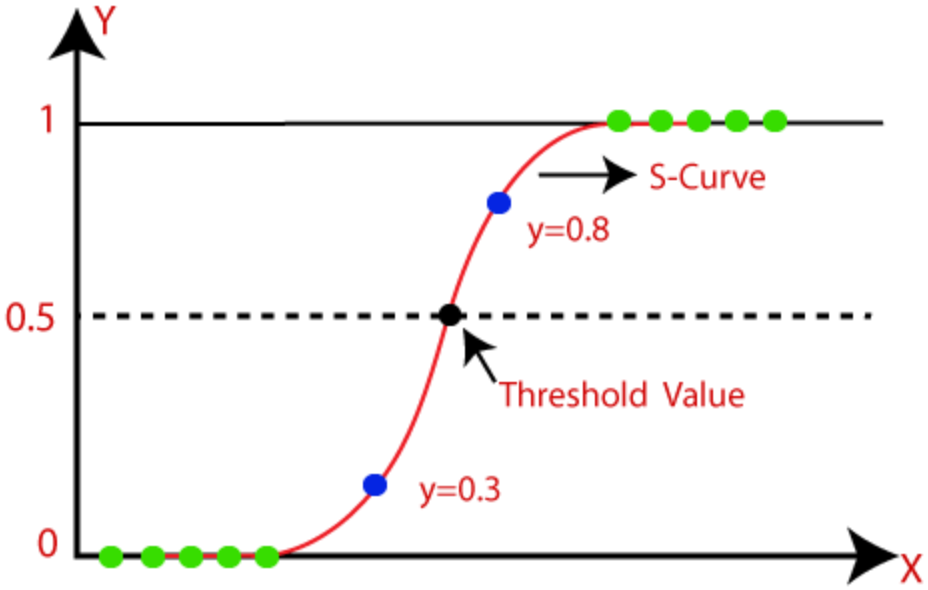

Figure 5. 5-sigmoid function

Logistic Function (Sigmoid Function):

The sigmoid function is the scientific function used to draw the predicted values in the predictions. Sets any real value to another value within the range 0 and 1. The setback value must be between 0 and 1, which can not go out of this limit, so it operates a curve comparable to the "S" form. The S-form curve is called the Sigmoid function or logistic function.

In the logistic regression, we use the concept of the limit/ threshold value, which defines the probability of 0 or 1. As the value above the limit value is straight to 1, and the value below the threshold value becomes 0. [12]

Assumptions for Logistic Regression:

Dependent variations should be categorized naturally. Independent variable should not be too cohesive. [12]

Advantages	Disadvantages
Logistic regression is provided facility work smoothly for execute, construe, and provide efficient manner for train.	If the number of definitions is fewer than the number of features, classifier should not be used, otherwise, they are among the overcrowded.
Logistic regression is not hoped about deliveries of classes around feature space.	model concept wise better for linear boundaries.
It can easily go to multiple stages (multi-country deferred) with acceptable class prediction tests.	A major restriction of model is the notion of coherence between dependent variations and independent variations.
Not only does it provide an measure of how to fit prediction (size coefficient), but also its hint direction (good or bad).	It can only be used to predict distinct activities. Therefore, the flexibility dependent on model is definite at a set number set.
It is too wild to separate unidentified records	Non-linear problems is not resolved by logistic regression for the reason that model is related a linear decision interface.
The Upright the accuracy when data sets is simple and well attained when the data collection is directly separated	Model is used average or no interrelationship among independent variables.
model coefficients are construed by it as feature status.	Neural Networks provide More power and dense for effortlessly beat this algorithm.

Table 02[13]

Decision Tree

Decision Tree is a Supervised learning method that can be used for both classification and Regression problems, but it is often preferred to solve Classification problems. A tree-structured classifier, where the internal bumps point to the features of the data collection, the branches indicate the rules of decision and the node of each leaf represents the result. [14]

For example, in the example beneath, decision-making trees captivate from data to evaluated a the sine curve for a set of decision rules if one. The subterranean the tree, the more complex the decisions are

In the decision tree, there are two areas, namely Decision Node and Leaf Node. Decision-making sites are used for decision making and have many branches, and Leaf nodes are the result of those decisions and do not contain other branches. [14]

Decisions or tests are reached on the basis of the features of the given data collection. It is a graphical representation of accepting all possible solutions to a problem / decision depending on the circumstances.

It is called the trunk tree because, compared to the tree, it begins with a root node, which grows on continuous branches and forms a tree-like structure. To order a tree discovery, we use the CART algorithm, which represents the Classification and Regression Tree algorithm. [14] The tree that decides essentially asks a question, and is based on the answer (Yes / No), and also divides the tree into subtrees.

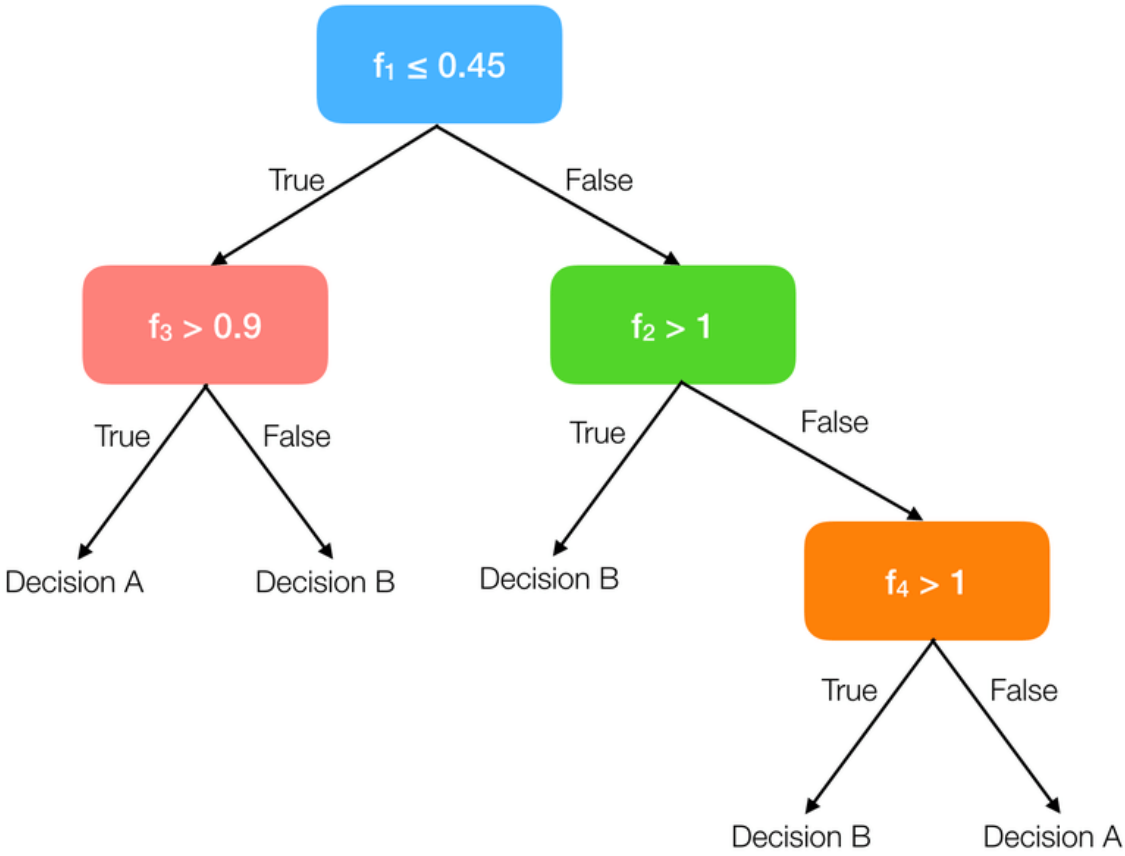

Figure 5. 6-Decision tree example

Why use Decision Trees?

There are many algorithms for machine learning, so choosing the best algorithm for the definite data and the problem is a major point to keep in mind when making a machine learning model. Below are two reasons for using the decision tree: [14]

Tree decisions often mimic human thinking when making a decision, so it is easy to identify. The over-the-top concept in the decision tree can be easily seen because it reflects a tree-like structure.

Advantages of the Decision Tree

• It is easy to see as it follows the same process where a person's track while creating any decision in real life.

• can be very good at resolving problems related to decisions.

• It is helpful to think about all the possible consequences of the problem.

• There are a few data purification requirements compared to other algorithms.

Disadvantages of the Decision Tree

• The cutting tree has many layers, which makes it compact.

• There can be an overdose problem, which can be solved using the Random Forest algorithm.

• On many classroom labels, the difficulty of computing the decision tree may arise

K-Nearest Neighbor

The K-Nearest Neighborhood is one of the humblest classifiers is depended on a supervised learning strategy. The K-NN algorithm uses contrasts between new case / data and available cases and adds a new case to the class that closely resembles the available groups.

The K-NN algorithm provides all existing data and classifies the new data point based on the comparison. This means that when new data looks at it it can be categorized into a neat collection class with the K-NN algorithm. [15]

The K-NN algorithm can be used for Regression and Classification but mainly for classification problems. K-NN is a non-parameter algorithm, which means it does not perform data transfer. It is also called the lazy student algorithm because it does not derive from a set of training rather it retains the database and during the classification, performs the action on the data collection. The KNN algorithm in the training phase simply retains the data collection and when it becomes new data, then classify those facts into a section that closely resembles the new data. [15]

Example: Suppose we have a picture of someone who looks like a cat and a dog, but we want to see if it is a cat or a dog. So in this identification, we can use the KNN algorithm, as it works in the same way. Our KNN model will get the same features of the new data set in images of cats and dogs and depending on the same features that will place it in the cat or dog category. [15]

How does K-Nearest Neighbor work?[15]

The K-NN residue can be defined in the root of the algorithm below: (Fig23 [15])

Step 1: Select the size K of the neighbors

Step 2: Calculate Euclidean distance K number neighbor

Step 3: Take your nearest K neighbors according to the Euclidean distance.

Step 4: Between these neighbors k, estimate the number of data points in each category.

Step 5: Assign points for new details in that section to the maximum number of neighbors.

Step 6: Our model is ready.

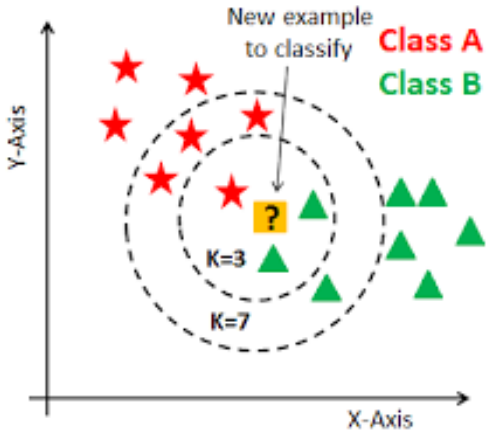

Figure 5. 7 [15]-KNN process

Advantages of KNN..

1. No Training Time: KNN is called Lazy Learner (Instance based). You get nothing during training. It does not initiate any discriminatory activity from the training details. In other words, there is no time to train. It maintains a training database and learns it only when making real-time forecasts. This makes the KNN algorithm much faster than other algorithms that need training

e.g. SVM, Linear Regression, etc.

2. Since the KNN algorithm does not require training prior to making predictions, new details may be added without seamstresses that will not reflect the accuracy of the algorithm.

3. KNN is very easy to use. There are only two limits required to use KNN namely K value and distance function (e.g. Euclidean or Manhattan etc.)

Disadvantages of KNN

1. It does not work well with large data collection: For larger networks, the cost of consolidating the distance between a new point and each current point is too high which minimum the recital of the algorithm.

2. It does not slog properly with large dimensions: The KNN algorithm does not slog properly with high dimensions data because with large size, it becomes a problem for the algorithm to measure the distance in each scale.

3.You want feature measurement: It is important to perform the feature measurement (calibration and standardization) before installing the KNN algorithm on any data set. If we do not do so, KNN may display incorrect predictions.

4. Kindliness to noise data, lost values and outliers: KNN is sensitive to noise in the database. It is important to pay for nonprofits and remove exporters.

Random Forest

Random Forest is a classification algorithm that covers different tree choices. It works by packing bags and inserting randomly when arranging each tree to make an effort to produce an unrelated forest of trees its group predictions are more accurate than any individual tree.
What do we need in order to guide our random jungle to make accurate class predictions?
We need features that are less predictable. After all, if we put garbage in there we will take out the trash.
Forest trees and especially their prognosis are important to be incompatible (or at least have a low interaction with each other). While the algorithm itself on the non-experimental aspect tries to establish this low-level relationship with us, the features we choose and the parameters we choose will also influence the final relationship. [16]

The Random Forest, as the name suggests, is a group of tree-based models suitable for random subsets of training details. Being a collaborative model, the main advantage of a random forest model is the abridged variance associated with single tree training. Since each tree in the group

is trained in a random set of standard training sets, the whole team is less likely to skip the training set. Forest-based random planning is one of the most straightforward types of many planning challenges.

Training of random forest models is based on the concept of integrated bootstrap, also known as wrap. Bagged-covered learning works as follows [17]

random sample (with additional inclusion) LL training examples from LL size training set.

train the tree-based model in the sample together in the first step.

The stairs above are repeated one by one

f TT trees that form a random forest. The number of trees in the forest, TT, is a hyperparameter, usually in the hundreds or thousands, depending on the size of the training set. It can be checked for verification error or wallet error. A bag error is a tree error in descriptions that were not part of the training sample used to train that tree. [17]

Steps 1 and 2 above apply to making any set of models based on bagging. In the case of random forests, detailed random production is presented in the form of elements.

Remind that tree-based models work by separating the top element in each section to form a node in the tree. With bag insertion, random forests make this step random by choosing one element among the top ones instead of the best feature in each category. [17]

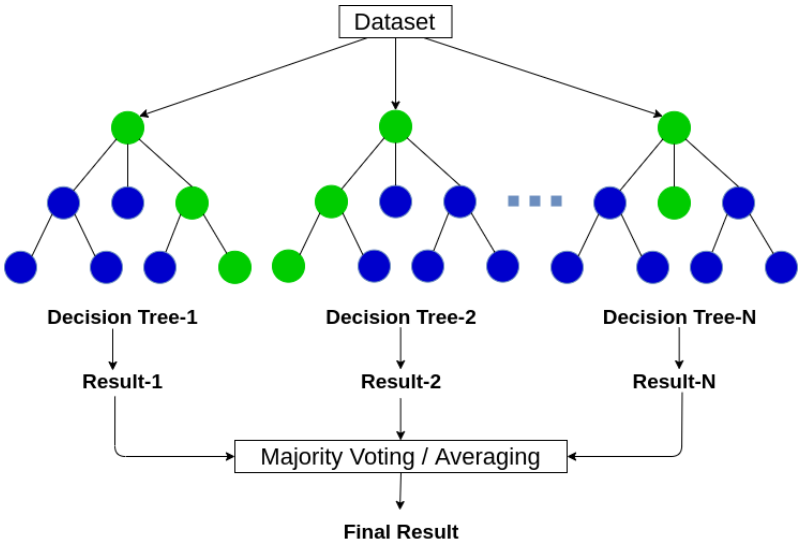

Figure 5. 8

Advantages and Disadvantages of Random Forest[18]

• Reduces overcrowding in decision trees and helps the advancement of accuracy

• It is plastic in both classification and regression problems

• Works properly with categorical and continuous values

• Sets missing data available in the data

• The prevalence of facts is not compulsory as it is done in a legal manner.

However, despite these advantages, a random forest algorithm also has some drawbacks.[18]

• It requires a lot of calculation power and heads as it counts many trees to cover its results.

• It also requires a lot of training time as it includes many decision trees that will control the phase.

• Due to the combination of decision trees, it is also difficult to interpret and fails to control the impact of each variation.

GaussianNB

Naive Bayes is an easy way to create classifiers: models that offer class labels in problematic situations, which appear to be carriers of feature values, where class labels are sorted from a limited set. There is no single algorithm for training such classifiers, but a family of algorithms based on a sense of belonging: all senseless Bayes categories accept that the value of a particular element is independent of the value of any other element, defined by class differences. For example, a fruit can be considered an apple if it is red, round, and about 10 cm wide. The inexperienced Bayes divider considers each of these factors to contribute independently to the probability that the fruit is an apple, regardless of whether there is a relationship between color, circumference and width. [19]

In some types of possible models, inexperienced Bayes classifier can be professionally trained in a supervised learning environment. In many real-world applications, the parameters test of the inexperienced Bayes models uses the high probability method; in other words, one can

work with an inexperienced Bayesi model without the possibility of tolerating the Bayesian or using any Basesi methods.

No matter how silly and cleverly simplified designs are designed, these useless Bayes clips have worked well in many real-world situations. In 2004, an analysis of the problem of Bayesian segregation revealed that there were plausible theoretical reasons for the seemingly impossible use of inexperienced Bayist separatists. [6] However, complete differences with other class algorithms in 2006 indicated that the Bayes arrangement was bypassed in other ways, such as developed trees or unplanned forests (Figure 29 [19]).

The advantage of the inexperienced Bayes is that they require only a small amount of training data to measure the required parameters in the separation.

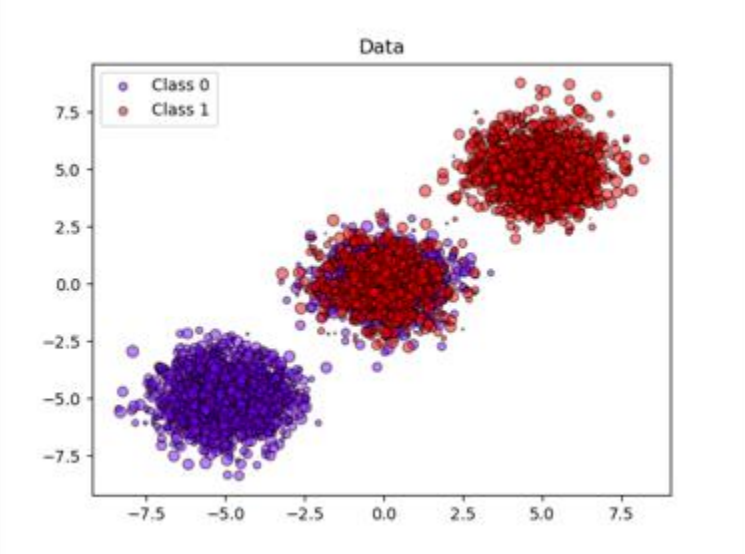

Figure 5. 9 [19]

Advantages of Naive Bayes[20]

This algorithm is predicted efficiently and class of a test dataset is predicted simply by naïve Bayes.

Researcher can divide to multi-class prediction problems without harm them.

Naive Bayes classifier want less training data than other classifier if model has independent features.

Naive Bayes algorithm extremely good for categorical input variables than numerical variables.

Disadvantages of Naive Bayes[20]

Naïve Bayes adopts that every features have not connection with each other its mean every feature is independent form others. in reality, researchers want find features that independent and this duty is very hard because most of time features are depended with each other's.

svm

A support vector machine is categorized as supervised machine learning algorithm that better for two-group classification problems. Inside of SVM model is able to classify new text after SVM model manipulate of labeled training fact for each group.

When we are compared with novel algorithms such as neural networks, SVM model is provided two main benefits: manipulation has capability to provide higher speed and provide better explanation with a portion of samples (in the thousands). SVM model is tried to provide better accuracy for text classification problems, where dataset has couple of thousands of tagged samples. [21]

When we compare set of examples of training , each one spot as setting to one group out of two groups, A model is figured by SVM training algorithm and model is received new examples to one group or the other, non-probabilistic binary linear classifier is made by it.

SVM plots training is provided examples to plots space so as to achievement the differentiation of the gap between the two categories. When New comers are in then marked into that same area and try fit to a category founded on which side of the gap they decrease and try to predict. [22]

When we compare with linear classification, SVMs has capacity for face to non-linear classification using the kernel trick, circuitously their inputs are plotted into high-dimensional feature spaces by SVM.

various real-world problems are solved by SVM:

text and hypertext categorization are used SVM in efficient manner.[7] shallow semantic parsing approaches are founded using support vector machines.

SVM is provided facilities for Classification of images. After just three to four round of relevance reaction, Investigational outcomes display that meaningfully higher search accuracy is attained by SVM than traditional query refinement schemes. This scenario same for image segmentation systems.

supervised SVM helpful for Classification of satellite data like SAR data .

using SVM is provided better response for Hand-written characters recognition.

The SVM algorithm is suited to biotic and other sciences. As example category of proteins with up to 90% of the blends grouped suitable way.

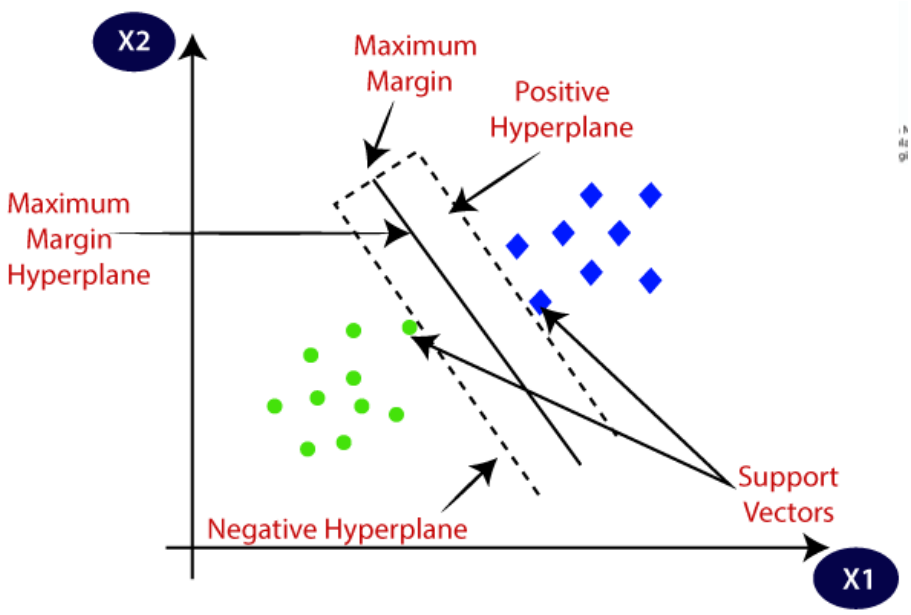

Figure 5. 10 [22]-svm example

SVM Advantages:

purely identify boundary between classes are provided better accuracy for svm.

high dimensional spaces are provided more actual by SVM.

SVM is efficient when number of dimensions is better when compare with number of samples.

SVM manage memory in effective manner

SVM Disadvantages:

huge data sets are not suited by SVM algorithm.

SVM are not provided best accuracy for data set which contain large scale noise as example target classes are overlapping.

data points are worked by support vector classifier.

Chapter 06

6. Implementation

Voice clip name format

Voice clip name format provide many details of voice owner. Voice clip name format file extension is .wav file format Voice clip name format provide voice owner detail such as gender and age.

First letter of format can be “g” or “b“. “g” mean female and “b” mean male. Third to filth letters provide emotion of voice owner. It can be “nor” or “ang” or “sad”. “nor” mean normal mood. “ang” mean angry mood. ”sad” mean sad mood. Seventh to eight letters provide age of voice owner.

Code segment

```
gender.append(j[0])
emotion.append(j[2:5])
age.append(j[6:8])
```

Figure 6. 1-file format decompose

Voice clip name format examples

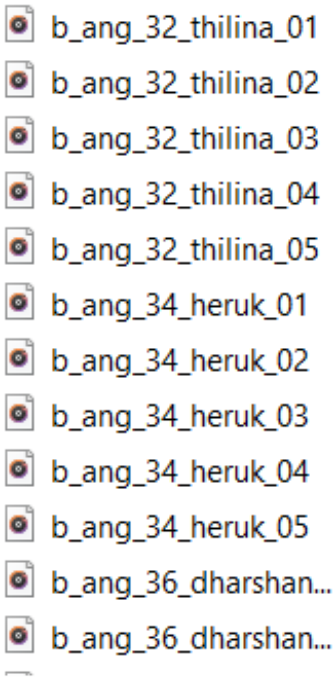

Figure 6. 2 file format

Voice clip features

Maximum amplitude

Python scipy wavfile provide amplitude of voice clip then maximum amplitude find out from amplitude array.

Code segment

```
y = signal[:, 0]  # use the first channel (or take their average, alternativ
t = np.arange(len(y)) / float(sr)
print ("amplitude is ",max(y))
```

Figure 6. 3-maximum amplitude

frequency_sepectrum is manually created function for frequencies and their amplitudes using FFT(fast Fourier transform).

Input parameters are

- t_series mean amplitude(time domain)
- Sr mean sample rate of time domain

Return parameters are

- Frq mean individual frequencies
- i_amplitude mean frequency's amplitude

process

x = fft(x) / n # fft computing and normalization

example

```
def frequency_sepectrum(x, sf):
    """
    Derive frequency spectrum of a signal from time domain
    :param x: signal in the time domain
    :param sf: sampling frequency
    :returns frequencies and their content distribution
    """
    x = x - np.average(x)  # zero-centering

    n = len(x)
    print(n)
    k = arange(n)
    tarr = n / float(sf)
    frqarr = k / float(tarr)  # two sides frequency range

    frqarr = frqarr[range(n // 2)]  # one side frequency range

    x = fft(x) / n  # fft computing and normalization
    x = x[range(n // 2)]

    return frqarr, abs(x)
```

Figure 6. 4-Frequency spectrum

frequency's amplitude

FFT provide amplitudes of each frequencies of voice clip. Now we try to find out maximum amplitude from all frequencies.

Code segment

```
frq, X = frequency_sepectrum(y, sr)#frequencies a
print ("maximum frequency is ",max(frq))
ts1 = max(frq).tostring()
frq_str=(np.fromstring(ts1, dtype=float))


frequency.append(frq_str)
print ("maximum amplitude is ",max(X))
ts2 = max(X).tostring()
occ=(np.fromstring(ts2, dtype=float))
```

Figure 6. 5- frequency's amplitude

Frequency of maximum amplitude

Now we try to find out frequency of maximum amplitude.

Code segment

```
max_fft_amp.append(occ)
import operator
max_index, min_value = max(enumerate(X), key=operator.itemgetter(1))
print ("frequency of maximum amplitude ",frq[max_index])
ts3 = frq[max_index].tostring()
occ_frq=(np.fromstring(ts3, dtype=float))
```

Figure 6. 6-Frequency of maximum amplitude

frequency_sepectrum is manually created function for frequencies and their amplitudes using FFT(fast Fourier transform).

Input parameters are

- t_series mean amplitude(time domain)
- Sr mean sample rate of time domain
- Return parameters are
- Frq mean individual frequencies
- i_amplitude mean frequency's amplitude

process

x = fft(x) / n # fft computing and normalization

example

```
def frequency_sepectrum(x, sf):
    """
    Derive frequency spectrum of a signal from time domain
    :param x: signal in the time domain
    :param sf: sampling frequency
    :returns frequencies and their content distribution
    """
    x = x - np.average(x)  # zero-centering

    n = len(x)
    print(n)
    k = arange(n)
    tarr = n / float(sf)
    frqarr = k / float(tarr)  # two sides frequency range

    frqarr = frqarr[range(n // 2)]  # one side frequency range

    x = fft(x) / n  # fft computing and normalization
    x = x[range(n // 2)]

    return frqarr, abs(x)
```

Figure 6. 7

frequency's amplitude

FFT provide amplitudes of each frequencies of voice clip. Now we try to find out maximum amplitude from all frequencies.

Code segment

```
frq, X = frequency_sepectrum(y, sr)#frequencies a
print ("maximum frequency is ",max(frq))
ts1 = max(frq).tostring()
frq_str=(np.fromstring(ts1, dtype=float))


frequency.append(frq_str)
print ("maximum amplitude is ",max(X))
ts2 = max(X).tostring()
occ=(np.fromstring(ts2, dtype=float))
```

Figure 6. 8- frequency's amplitude

Frequency of maximum amplitude

Now we are tried to calculate frequency of maximum amplitude.

Code segment

```
max_fft_amp.append(occ)
import operator
max_index, min_value = max(enumerate(X), key=operator.itemgetter(1))
print ("frequency of maximum amplitude ",frq[max_index])
ts3 = frq[max_index].tostring()
occ_frq=(np.fromstring(ts3, dtype=float))
```

Figure 6. 9- Frequency of maximum amplitude

Mel value

We select maximum Mel value from array.'

Input parameters

- time-series of audio
- audio time-series sampling rate

. now we are tried to Produce a Mel scale: Mel scale depend on frequency spectrum, and set n_mels in to 128 for evenly set apart frequencies(Fig 15)

Return parameters

- Mel spectrogram ---Y□MEL VALUE----X-□TIME

Code segment

```
sr, signal = wavfile.read(wave_file_path)
yy, srsr = librosa.load(wave_file_path)#time series and sampling rate
S = librosa.feature.melspectrogram(yy, sr=srsr, n_mels=128)#mel detail
```

Figure 6. 10-Mel value

Next calculate log scale (dB) using Mel detail.

Code segment

```
log_S = librosa.amplitude_to_db(S, ref=np.max)
print("mel value is ",np.max(S))
max_mel.append(np.max(S))
```

Figure 6. 11

Data features

Finally, we are created excel sheet with following details.

- Gender
- Age
- Amplitude (from time series)
- Frequency (time series)
- Maximum amplitude from FFT
- Frequency of Maximum amplitude from FFT
- Maximum Mel value
- Emotion

amplitude	gender	frequency	max_fft_amp	max_occurance_frequency	emotion	max_mel	age
24329	b	23999.08	271.801141	1053.361053	ang	802.5593	32
23576	b	23999.21	379.657605	400.0529171	ang	957.4911	32
25711	b	23999.26	307.317424	1303.656206	ang	1614.0065	32
25890	b	23998.34	554.3791	712.0119774	ang	1414.4512	32
24800	b	23998.62	457.543564	660.0836564	ang	1801.2346	32
21611	b	22049.52	208.511833	313.0782567	ang	428.1727	34
23463	b	22049.31	147.883748	340.3635333	ang	141.14482	34
21923	b	22049.36	172.249375	298.893715	ang	273.65735	34
22245	b	22049.34	212.445884	327.9672476	ang	378.4852	34
23340	b	22049.53	216.536156	331.892196	ang	670.1018	34
22667	b	23999.43	274.212018	628.9772727	ang	380.06406	36
27038	b	23999.61	126.171454	227.734375	ang	360.8742	36
24636	b	23999.43	167.925635	572.7272727	ang	692.70966	36
24775	b	23999.43	292.094733	450.5681818	ang	810.7126	36
19874	b	23999.52	128.038984	217.7884615	ang	436.81796	36
23324	b	22049.52	119.789488	290.4589844	nor	385.94357	32
16255	b	22049.35	210.275064	249.9156605	nor	375.26056	32
22571	b	22049.44	397.875795	276.2961648	nor	814.87933	32
11343	b	23998.84	173.554791	130.2149381	nor	72.508354	32
10665	b	23999.43	158.571898	131.02507	nor	122.10302	32
9927	b	23998.99	225.649665	134.0388007	nor	102.12695	32
13630	b	23998.71	169.534614	132.875555	nor	155.175	32
13752	b	22049.42	153.780788	250.250739	nor	126.020294	34
17288	b	22049.42	242.444801	250.250739	nor	290.60883	34
14138	b	23998.57	239.987108	538.7784345	nor	228.60931	34

Table 03

Machine learning classifiers

Logistic Regression

Code segment

Import Logistic Regression classifier

```
from sklearn.linear model import LogisticRegression
```

Figure 6. 12-Import Logistic Regression classifier

- Logistic Regression accuracy

```
logreg = LogisticRegression()
logreg.fit(X_train,y_train)
y_pred=logreg.predict(X_test)
#print(X_test)
#print(y_pred)
print("LogisticRegression Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

Figure 6. 13-Logistic Regression accuracy

Decision Tree

Code segment

Import Decision Tree Classifier

```
from sklearn.tree import DecisionTreeClassifier
```

Figure 6. 14-Import Decision Tree Classifier

Decision Tree accuracy

```
clf = DecisionTreeClassifier()
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
#print(X_test)
#print(y_pred)
print("DecisionTree Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

Figure 6. 15-Decision Tree accuracy

Random Forest

Code segment

- Import Random Forest Classifier

```
from sklearn.ensemble import RandomForestClassifier
```

Figure 6. 16-Import Random Forest Classifier

- Random Forest accuracy

Code segment

```
clf = RandomForestClassifier()
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
#print(X_test)
#print(y_pred)
print("RandomForest Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

Figure 6. 17-Random Forest accuracy

K-Nearest Neighbor

Code segment

- Import K-Neighbors Classifier

```
from sklearn.neighbors import KNeighborsClassifier
```

Figure 6. 18-Import K-Neighbors Classifier

- KNeighbors Accuracy

```
clf = KNeighborsClassifier(n_neighbors=3)
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
#print(X_test)
#print(y_pred)
print("KNeighbors Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

Figure 6. 19-KNeighbors Accuracy

GaussianNB

Code segment

- Import GaussianNB Classifier

```
from sklearn.naive_bayes import GaussianNB
```

Figure 6. 20-Import GaussianNB Classifier

- GaussianNB accuracy

```
clf = GaussianNB()
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
#print(X_test)
#print(y_pred)
print("GaussianNB Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

Figure 6. 21-GaussianNB accuracy

svm

Code segment

- Import svm Classifier

```
from sklearn import svm
```

Figure 6. 22-Import svm Classifier

- svm accuracy

```
clf = svm.SVC(kernel='linear') # Linear Kernel
clf = clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)
print(X_test)
print(y_pred)
print("svm Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

Figure 6. 23-svm accuracy

Chapter 07

7. Results and analysis

Now we are tried to analyze result which generate by machine learning classifiers, real results are presented and try to discuss what are the major things and differentiate of result among different classifiers. We are presented the recognition accuracy of using logistic regression, random forest , svm, gaussianNB, decision tree and k-nn.

test size

data type of train set is float, value can vary between 0.0 and 1.0 and split part of dataset to contain in the test split When size of test data set is depended on function call value and it is absolute number .it will be set to 0.3 in function call. it is mean 30% of data set use for train data set.

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=25)

train size

data type of train set is float, value can vary between 0.0 and 1.0 and split part of dataset to contain in the train split. When size of train data set mean number of tuples of main data set after separate test data set tuples and it is absolute number. If None, the value is remain data set of main data set after separate test data tuples from main data set. it is set to 0.7.it is mean 70% of data set use for train data set.

random state

data set is shuffled by random before applying the split for training data set and test data set. Both training data set and test data set is varied by multiple function calls because of Pass an int for reproducible output. we are set this number as 25.

Training accuracy

Different values are provided by rule-based algorithms. Figure 7.1 and figure 7.2 are examples for different values of rule-based algorithms. But k-nearest neighbor algorithm , GaussianNB algorithm, svm algorithm and Logistic regression algorithm provide constant values

```
RandomForest Accuracy: 0.8214285714285714
DecisionTree Accuracy: 0.75
KNeighbors Accuracy: 0.5
GaussianNB Accuracy: 0.4642857142857143
svm Accuracy: 0.4642857142857143

LogisticRegression Accuracy: 0.4642857142857143
```

figure 7. 1--Training accuracy example 01

```
RandomForest Accuracy: 0.7857142857142857
DecisionTree Accuracy: 0.75
KNeighbors Accuracy: 0.5
GaussianNB Accuracy: 0.4642857142857143
svm Accuracy: 0.4642857142857143

LogisticRegression Accuracy: 0.4642857142857143
```

Figure 7. 2-Training accuracy example 02

Test accuracy

We are used randomly selected ten rows for check test accuracy. System provided two different values for when we execute models several time. Fig 37 and Fig 38 display different two answers for this random rows.

Figure 7.3 has 90% test accuracy and Fig 38 provide 80% test accuracy

```
predicted value is  nor  original value nor
predicted value is  nor original value nor
predicted value is  ang original value ang
predicted value is  ang original value ang
predicted value is  ang original value ang
predicted value is  sad original value sad
predicted value is  sad original value sad
predicted value is  sad original value sad
predicted value is  sad original value nor
predicted value is  nor original value nor
```

Figure 7. 3-Test accuracy

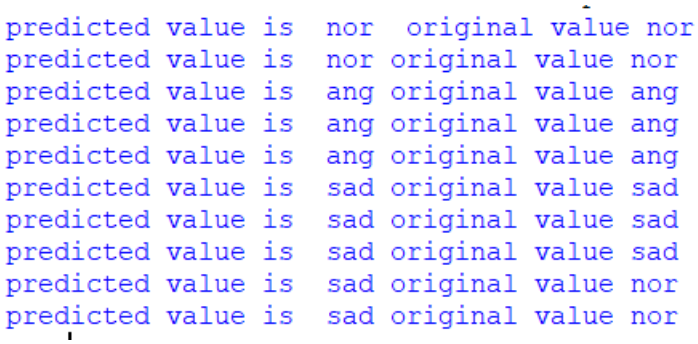

Figure 7. 4-Test accuracy

Chapter 08

8. Conclusion

automatic speech emotion recognition (SER) system is practiced six machine learning algorithms(logistic regression, random forest, decision tree, svm, GaussianNB, K-Nearest Neighbor) for predict different emotion status (angry, normal, sadness). Thus, six voice based features (fast Fourier transform (FFT) is providing maximum amplitude from all frequencies and frequency of maximum amplitude. The mel scale is a scale of pitches that human earshot usually observes to be intermediate from each other. Maximum mel value, maximum amplitude from all frequencies, frequency of maximum amplitude, age of user and gender are used to find out emotion status human) were extracted from our own database and we were tried to combine different voice features for acquire maximum accuracy. also, we was identified how to classifiers change according to voice features and how to features effect to outcome that accuracy of emotion recognition in speech.

voice features are trained and are evaluated by machine learning models to recognize emotional states. best recognition rate are reported by SER etween 75% and 82% on random forest and recognition rate between 75% and 77% on decision tree. From this result, we can see that Decision-making algorithms(random forest and decision tree) often perform better with our data base. Therefore, we concluded that the K-Nearest Neighbor (accuracy 50%) , svm (46%), GaussianNB (46%) and logistic regression (46%).

Rule based algorithms are good for recognize emotion status using voice. Random forest is best machine learning algorithm for recognize emotion status using voice according to our experiment. Random forest has good test accuracy (between 80% to 90%).

9. Reference

[1] https://online.uwa.edu/news/emotional-psychology/

[2]https://www.healthline.com/health/depression/effects-on-body#7

[3]https://www.analyticsinsight.net/speech-emotion-recognition-ser-through-machine-learning/

[4]https://www.intechopen.com/books/social-media-and-machine-learning/automatic-speech-emotion-recognition-using-machine-learning

[5]https://arxiv.org/ftp/arxiv/papers/1906/1906.05681.pdf

[6]https://www.tutorialspoint.com/signals_and_systems/fourier_transforms.htm

[7]https://en.wikipedia.org/wiki/Fourier_transform

[8]https://en.wikipedia.org/wiki/Fast_Fourier_transform

[9]https://www.dataq.com/data-acquisition/general-education-tutorials/fft-fast-fourier-transform-waveform-analysis.html

[10]https://developer.apple.com/documentation/accelerate/computing_the_mel_spectrum_using_linear_algebra

[11]Website title: http://dl.iranmohandes.com

[12]https://www.javatpoint.com/logistic-regression-in-machine-learning

[13]https://www.geeksforgeeks.org/advantages-and-disadvantages-of-logistic-regression/

[14]https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

[15] http://theprofessionalspoint.blogspot.com/2019/02/advantages-and-disadvantages-of-knn.html

Code segment

[16]https://towardsdatascience.com/understanding-random-forest-58381e0602d2

[17]https://the-learning-machine.com/article/ml/random-forest?gclid=Cj0KCQjwub-HBhCyARIsAPctr7zBlTUi6LppYL_yAvU1jUajLmnOcKn5-yRlIUJti2Fdsfv7YgaMxlgaAvD0EALw_wcB

[18]https://www.mygreatlearning.com/blog/random-forest-algorithm/

[19]https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[20]https://www.upgrad.com/blog/naive-bayes-classifier/

[21]https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/
[22]https://en.wikipedia.org/wiki/Support-vector_machine