WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

PERSONALITY PREDICTION

Prof. Bhavana.B[1], Ophelia[2], Nisha[3], Pallavi[4], Priya[5]

[1]Dept. of Computer Science and Engineering, Guru Nanak Dev Engineering College, Bidar
[2,3,4,5]Dept. of Computer Science and Engineering, Guru Nanak Dev Engineering College, Bidar

ABSTRACT

Computer vision is an important area in the field of computer science which aids the machines in becoming smart and intelligent by perceiving and analyzing digital images and videos. A significant application of this is Activity recognition – which automatically classifies the actions being performed by an agent. Personality prediction aims to apprehend the actions of an individual using a series of observations considering various challenging environmental factors. This work is a study of various techniques involved, challenges identified and applications of this research area.
The smartphone has become quite ubiquitous and an indispensable part of our lives in the modern day. It has many sensors which capture several minute details pertaining to our activities. So, it is but inevitable that human desire creeps in to augment and improve one's own actions by studying such behaviors captured through the instrumentalities of the smartphone. In this context, study of data on human activities captured through accelerometer and gyroscope get primal significance. In this paper, we have attempted to apply machine learning and deep learning techniques on a publicly available dataset. Initially, classification algorithms like K-Nearest Neighbor's and Random Forest are applied. The classification accuracies observed are 90.46% and 92.97% respectively. Using benchmark feature selection and dimensionality reduction techniques does not improve the model accuracies to a large extent – with reported accuracies of 91.48% and 92.56% respectively. However, on employing deep neural network techniques, an accuracy of 97.32% is achieved, which indicates suitability of deep learning techniques over traditional machine learning techniques for the task of personality prediction using mobile sensor data.

Keywords *- Python, Django, model.*

1. INTRODUCTION

Personality Prediction, a sub domain of vision related applications, is the ability to identify and recognize the actions or goals of the agent, the agent can be any object or entity that performs action, which has end goals. The agent can be a single agent performing the action or group of agents performing the actions or having some interaction. One such example of the agent is human itself and recognizing the activity of the person called as Personality Prediction.

The goal of persons activity recognition is to automatically analyze ongoing activities from an unknown video (i.e. a sequence of image frames). In a simple case, where a video is segmented to contain only one execution of a person's activity, the objective of the system is to correctly classify the video into its activity category [1].

Personality prediction is one of the active research areas in computer vision as well as human computer interaction. The wide variety and unexpected pattern of activities performed by person, poses the challenge to the methodology adopted for recognition.

There are various types of person's activities. They are conceptually categorized as – gestures, actions, interactions, and group activities. Gestures are elementary movements of a person's body parts, and are the atomic components describing the meaningful motion of a person. Actions are single-person activities that may be composed of multiple gestures organized temporally.

Interactions are Person activities that involve two or more persons and/or objects. Finally, group activities are the activities performed by conceptual groups composed of multiple persons and/or objects. Approaches implemented should cater to all the requirements. The above listed category of activities is known as Usual Activities.

Any activity which is different from the defined set of activities is called as Unusual Activity. These unusual activities occur because of mental and physical discomfort. Unusual activity or anomaly detection is the process of identifying and detecting the activities which are different from actual or well-defined set of activities and attract human attention.

2. PROBLEM STATEMENT

Understanding the activities of human from videos is demanding task in Computer Vision. Identifying the actions being accomplished by the human in the video sequence automatically and tagging their actions is the prime functionality of intelligent video systems. The goal of activity recognition is to identify the actions and objectives of one or more objects from a series of examination on the action of object and their environmental condition. The major applications of Personality prediction vary from Content-based Video Analytics, Robotics, Human-Computer Interaction, Human fall detection, Ambient Intelligence, Visual Surveillance, Video Indexing etc. This paper collectively summarizes and deciphers the various methodologies, challenges and issues of personality prediction systems Variants of Personality prediction systems such as Human Object Interactions

and Human-Human Interactions are also explored. Various benchmarking datasets and their properties are being explored. The Experimental Evaluation of various papers are analyzed efficiently with the various performance metrics like Precision, Recall, and Accuracy.

3. PROPOSED IDEA

I. EXISTING SYSTEM

Personality prediction (HAR) is generally evaluated based on accuracy and computational cost. In order to effectively recognize different activities, previous works have tried to extract hand-crafted features from the signals of the accelerometer and gyroscope [7], and applied different classical machine learning approaches like Support Vector Machine [8], Random Forest [9] etc. for classification. Sharma et al. have used neural networks (ANN) [4], while Khan has used decision trees to classify basic activities [3]. In order to reduce the computational cost, some researchers have used feature selection [5] before applying classification model. However, most of these works have employed methodologies that use certain selected and hand-crafted features which not only increase the computational cost but also make them difficult to compare. Some research has also been carried out for HAR using deep learning. Duffner, Berlemont, Lefebvre and Garcia (2014) have used deep convolutional neural network (DCNN) which uses accelerometer and gyroscope data together to automatically extract features for activity recognition

II. PROPOSED SYSTEM

- The purpose of this thesis is to reinforce the Activity video performance.
- First, a raw Activity video is taken as AN input. The Activity video supply may be a live camera, recorded file or the other Activity video capturing device.
- During this raw Activity video, there'll be some blur or noise elements. To get rid of this blur or noise elements, preprocessing is completed on the Activity video.
- An Activity video may be pictured a collection of frames. To perform the preprocessing steps, frames square measure extracted one by one from the Activity video.
- Frame Extraction: These methodologies also are known as Preprocessing. During this approach extract image from applied input Activity video. Thence offer numerous sample image then to changing image especially size format. Image to convert grey scale: during this step input sample image like as colored image to changing grey scale image. This process provides a color distinction input image sample.
- Classifier - Logistic Regression: Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable. In a binary logistic regression model, the dependent variable has two levels – 0/1 or pass/fail. The output of LR model is a probability. We can select a threshold value, usually the threshold value is set to 0.5 in a binary classification model.
- Preprocessing: Given a activity video sequence, the image silhouettes square measure extracted by background subtraction and thresholding very manner virtually just like the approach to create gait illustration insensitive to the gap between the camera and additionally the topic, we tend to size each activity video silhouette into 64×44 per the middle of mass of each silhouette.
- Watershed Segmentation: The Watershed work on might be particular technique for segmenting digital footage that uses a form of region growing technique supported an image gradient.

4. MODULES

A) DATA COLLECTION:

Machine learning is the new big thing in the world of computer science. The motivation behind this project is to implement machine learning algorithms in real-world data sets so that their accuracy can be studied and effective conclusions can bedrawn. In this task, we develop AI models for "Personality prediction Using Smartphones Data Set" from UCI (University of California Irvine) online storehouse. This informational index has been gathered from chronicles of 30 human subjects caught by means of cell phones empowered with installed inertial sensors. Many AI courses utilize this information for educating purposes. This is a multi-arrangement issue. The informational collection has 10,299 lines and 561 segments. There are thirty volunteers of age group 18-50 years and examinations done in that. Every individual performed physical activities like WALKING,WALKING_UPSTAIRS,WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a cell phone on the midsection. Utilizing its implanted accelerometer and gyrator, we caught 3-pivotal straight increasing speed and 3- hub rakish speed at a steady rate of 50Hz. The analyses have been video- recorded to name the information physically

B) DEEP LEARNING NETWORKS:

The conventional machine learning methods rely on the hand engineering features and whose performance will be based on the selection of appropriate features. Advancement in technologies and due to the evolution of high-performance architectures, deep learning methods are widely being used nowadays. Manual feature engineering has been passed on to the deep neural networks which have the capability of choosing the features automatically with respect to the underlying samples.

C) VIDEO ANALYSIS:

A state of art Recurrent Neural Network (RNN), LSTM is popularly used to model time series data and vastly applied in language modelling, speech recognition, language translation, Image captioning and also in autonomous driving systems. This deep architecture was designed to handle vanishing gradient problem and avoid long-term dependency problem with the help of three gates namely input, output and

forget gates. Each cell is controlled by these three gates that accepts input for the current time step and the output from previous step. The model will learn to predict the human activity one time-step at a time.

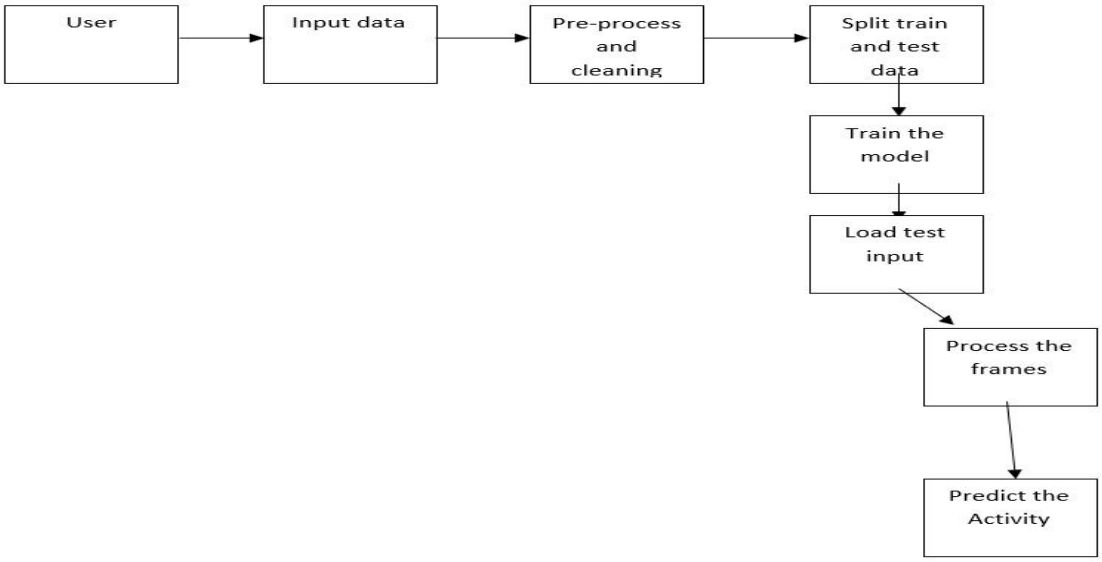
5. DATASET DESCRIPTION

We have used the standard dataset named, Personality prediction database [6], which was created from inertial sensors embedded in a waist-mounted smartphone. This was collected from 30 volunteers and are to sense the activities such as Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing and Laying. The two sensors used were embedded accelerometer and gyroscope to capture -axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. Pre-processing and sampling were applied on the captured data by applying noise and then by using fixed-width sliding windows of 2.56 seconds with 50% overlap. Butterworth low-pass filter was used to separate the gravitational and body motion components of the sensor into body acceleration and gravity data. Filter with 0.3 Hz cut-off frequency was used to have the gravitational force which has only low frequency components. Variables from time and frequency domain were used to construct the vector of features from each window.

Exploratory Data Analysis (EDA) 1) Studying Patterns: In many cases, a candidate's pattern of behavior in regard to his/her activities can be predicted. For instance, a person leading a rather sedentary lifestyle will show pronounced effects while sitting or laying whereas an objectively active candidate will show fluctuations while performing activities like walking or standing. It is important to clarify that both of the above candidates are extremes of a spectrum. So, to study a more or less general behaviors, the patterns of behaviors in the principal components for the activities are looked at.
2) Studying Correlations: There could be factors that influence the similarity or dissimilarity seen in the behaviors patterns for various candidates while performing different activities. However, it would be incorrect and absurd to state that if two features are correlated, then one causes the other. This is because correlation does not imply causation. From the correlation matrix (Fig. 6), it can be inferred that accelerometer readings in the x and y directions, and gyroscope readings in the y and z directions, are significantly positively correlated. Moreover, the gyroscope reading in the y direction and the accelerometer reading in the z direction are significantly negatively correlated. This could be because of many factors: positioning of the smartphones, a general fashion of performing activities among the subjects, movement in one direction affecting that in another direction etc. Deciphering these causes is beyond the scope of the present work since the objective here is to solely study the patterns of behaviors in data.

Principal Component Analysis (PCA): PCA is employed to convert the set of observations of possibly correlated features, into a set of values of linearly uncorrelated variables [5].. Each of the principle components is chosen in such a way that it describes most of the available variance. All these components are orthogonal to each other. The dataset used for this study has 561 variables which is quite a high dimensionality. It can be reduced to 2 or 3 dimensions using PCA so that we can plot and hopefully understand the data better. The activities are characterized in numbers as 1. Walking 2. Walking Upstairs 3. Walking Downstairs 4. Sitting 5. Standing 6. Laying The explained variance tells how much information can be attributed to each of the principal components. This is important since while converting high dimensional space to 2- dimensional space, we are bound to lose some variance.

6. SYSTEM ARCHITECTURE

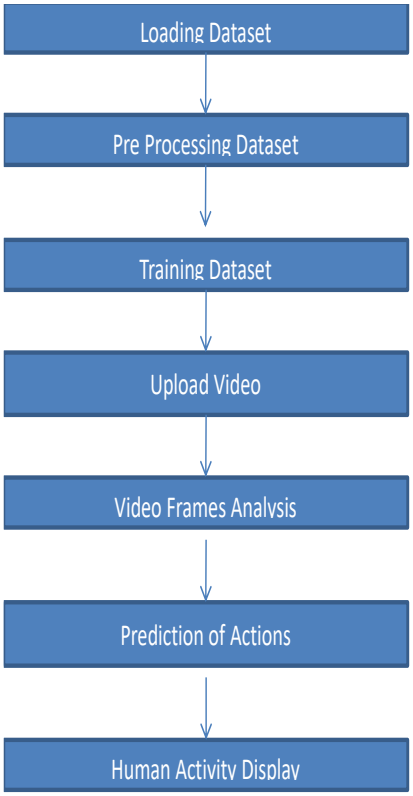


7. SYSTEM DESIGN

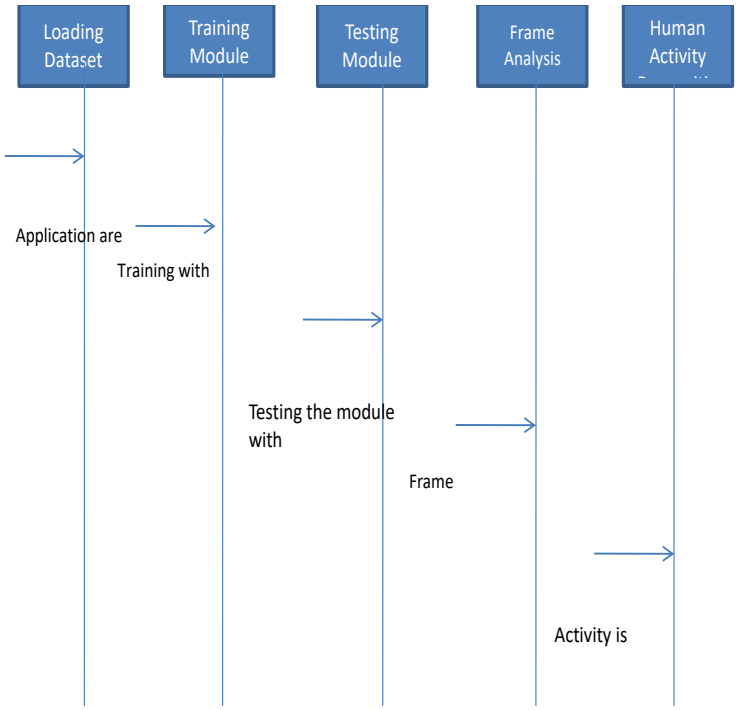
System Design is the next development stage where the overall architecture of the desired system is decided. The system is organized as a set of sub systems interacting with each other. While designing the system as a set of interacting subsystems, the analyst takes care of specifications as observed in system analysis as well as what is required out of the new system by the end user.

As the basic philosophy of Object-Oriented method of system analysis is to perceive the system as a set of interacting objects, a bigger system may also be seen as a set of interacting smaller subsystems that in turn are composed of a set of interacting objects. While designing the system, the stress lies on the objects comprising the system and not on the processes being carried out in the system as in the case of traditional Waterfall Model where the processes form the important part of the system.

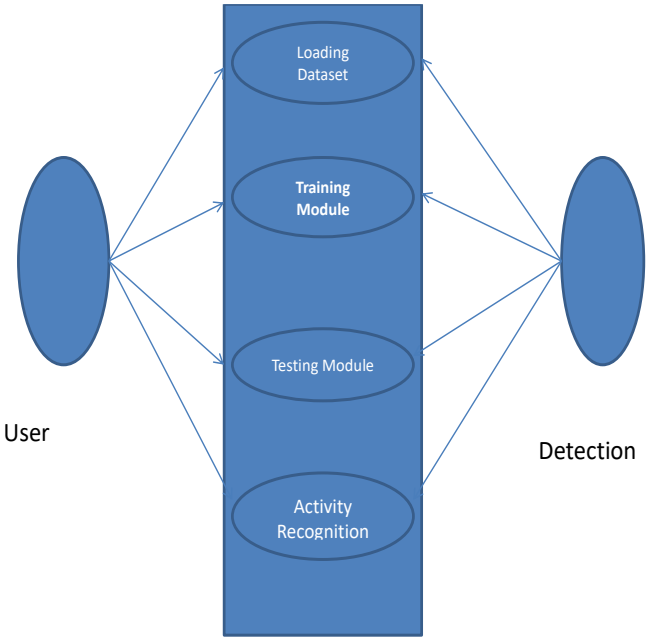
DATA FLOW DIAGRAM:



SEQUENCE DIAGRAM:



USE CASE DIAGRAM:



8. TECHNOLOGY

PYTHON:

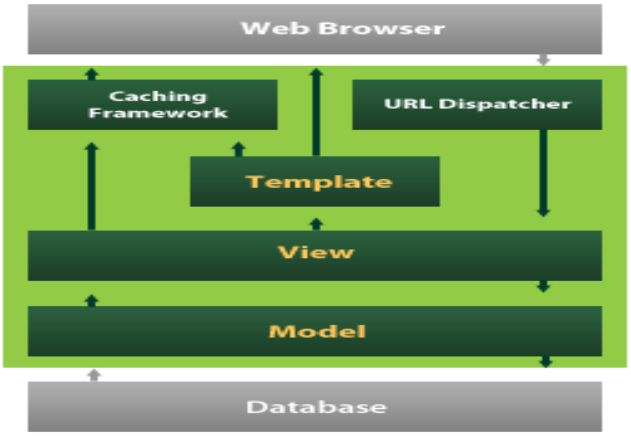
Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. An interpreted language, Python has a design philosophy that emphasizes code readability (notably using whitespace indentation to delimit code blocks rather than curly brackets or keywords), and a syntax that allows programmers to express concepts in fewer lines of code than might be used in languages such as C++or Java. It provides constructs that enable clear programming on both small and large scales. Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations. CPython is managed by the non-profit Python Software Foundation. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library

DJANGO:

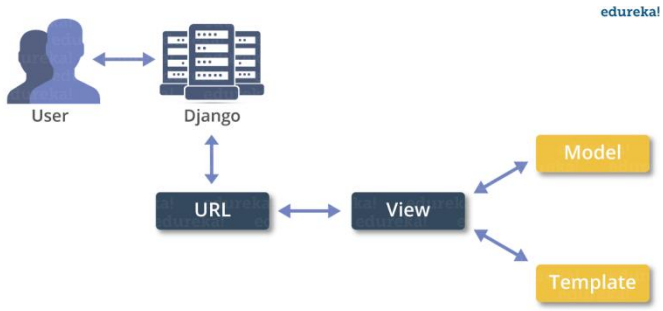
Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It’s free and open source.

Django's primary goal is to ease the creation of complex, database-driven websites. Django emphasizes reusability and "pluggability" of components, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models.

Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models



Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models



9. CONCLUSION

Human activity analysis is a popular activity in the growing industry and we have applied different machine learning algorithm. Comparative study performed among the applied various techniques kNN, SVM, Random forest, Neural Networks, Logistic regression and Naïve Bayes. In them, Logistic Regression and neural network gave good results whereas Naive Bayes result was not good. The implementation of Neural Network on Python gave better results than the one provided in the Orange tool. The limitations of this work is though the efficiency of neural network is good, the model is not dynamic. The inability of getting trained with real time data will force us to train the model every time new data comes. In future, these results can be used for making smart watches and similar devices which can track a user's activity and notify him/her of the daily activity log. They can also be used for monitoring elderly people, prison inmates, or anyone who needs constant supervision.

REFERENCES

[1] K. Sumathi, N. S. Lakshmi, and S. K. Kundhavai, "Reviewing the impact of smartphone usage on academic performance among students of higher learning." in International Journal of Pure and Applied Mathematics, vol. 118, 2018.

[2] M. A. Osman, A. Z. Talib, Z. A. Sanusi, T. Shiang-Yen, and A. S. Alvi, "A study of the trend of smartphone and its usage behavior in malaysia," in International Journal on New Computer Architectures and Their Applications. IEEE, 2012.

[3] N. Gupta, S. Garg, and K. Arora, "Pattern of mobile phone usage and its effects on psychological health, sleep, and academic performance in students of a medical university." in National Journal of Physiology, Pharmacy and Pharmacology, vol. 6, 2016.

[4] S. Ranasinghe, F. A. Machot, and H. C. Mayr, "A review on applications of activity recognition systems with regard to performance and evaluation." in International Journal of Distributed Sensor Networks, 2016.

[5] J. T. Sunny and S. M. George, "Applications and challenges of personality prediction using sensors in a smart environment." in International Journal for Innovative Research in Science and Technology, vol. 4, 2015.

[6] P. R. Woznowski, R. King, W. Harwin, and I. Craddock, "A personality prediction framework for healthcare applications: ontology, labelling strategies and best practices." in International Conference on Internet of Things and Big Data, 2016.

[7] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for personality prediction using smartphones." in Esann, 2013.

[8] "Personality prediction on smartphones using a multiclass hardware-friendly support vector machine," in International workshop on ambient assisted living. Springer, 2012, pp. 216–223.

[9] T. Peterek, M. Penhaker, P. Gajdos, and P. Dohn ˇ alek, ´ "Comparison of classification algorithms for physical activity recognition," in Innovations in bio-inspired computing and applications. Springer, 2014, pp. 123–131.