International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Human Movement Recognition System using R

Ajay Agarwal, Ashutosh Sharma, Amit Kumar Gupta, Vikas Goel

Abstract: With the proliferation of ubiquitous computing, the desire to make everyday life smarter and easier is growing more and more. Human Activity Recognition (HAR) is the result of a similar motivation. By recognizing user activity, HAR enables a wide range of comprehensive IT applications. To contribute to the multifarious applications proposed by HAR, it is essential to plan the appropriate activities. The simplest of the problems is using the wrong data manipulation and the execution of the prediction using erroneous algorithms interfering with the performance of the HAR system. R has proven to be a powerful and flexible tool for data mining and analysis. Here, we analyze the set of data extracted from UCI (University of California, Irvine) dataset using R. As a result of analysis any activity performed by participants will be recognized. As a sample, we are extracting data of 30 volunteers aged 19 to 48, each carrying a Smartphone at the waist. They perform various activities and record the data. Using the confusion matrix to apply on Support Vector Machine algorithm, we extract energy needed to perform activities, the frequency of each domain, etc., from dataset and display the results; standing, lying, or sitting. In other words, we classify the activities to be done by participants. Its applications include surveillance systems, patient monitoring systems and various systems, including interaction between people and electronic devices. This document will drive future research in more productive areas.

Index Terms: Human Activity Recognition, R Tool, Smartphone, Support Vector Machine Algorithm.

I. INTRODUCTION

The recognition of human activity plays an important role in dialogue and interpersonal relationships between humans. This is to provide information on the identity, personality and psychological state of the person. Human ability to recognize human activity is one of the subjects of scientific research in computer vision and machine learning [1]. The recognition of human activity (HAR) is an important task for Ambient Intelligence (AmI) systems because they are sensitive, responsive and adaptable [2]. It can recognize the human condition of human behavior and can be used as input to other systems. For example, in healthcare, you can detect abnormalities in individual movements and treat accordingly. It can be used to detect abnormal behavior on security and to prevent theft and other criminal acts. Activity-based activity [3] aims to understand the state of users and the environment using heterogeneous sensors. When these sensors were attached to the subject's body, the ideal device for many tasks related to physiological signals [4] was a cell phone. In addition to basic phones, smart phones, the next generation of mobile phones, offer features such as multitasking and the deployment of various sensors. They follow the activities in a transparent manner, learn from them and then help us make better decisions for future actions [4]. This is one of the important concepts on which Ambient Intelligence (AmI) relies. In this article, we plan to use smart phones with potential applications. Using an accelerometer and a gyroscope installed on the Smartphone, the data is recorded and used for analysis. An easy-to-use graphical user interface helps users accomplish their tasks and improves the efficiency of expert users. R is a powerful statistical programming language that makes it easy to use the latest statistical methods using thousands of additional packages available on the R Archive Network (CRAN) full download server [5]. Use the R programming with. The user must find the name of the function that performs their task and remember the name and option of the name of the variable and the argument that gives it. The HAR dataset [6] comes from the UCI archives and contains factors such as body acceleration and gravitational acceleration, which help to understand the movement in depth. The details of the dataset will be explained later.

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
Ajay Agarwal*, Information Technology, KIET Group of Institutions, Ghaziabad, India.
Ashutosh Sharma, Information Technology, IMS-Ghaziabad, Ghaziabad, India.
Amit kr. Gupta, Computer Application, KIET Group of Institutions, Ghaziabad, India.
Vikas Goel, CS&E, Ajay Kumar Gaarg Engineering College, Ghaziabad, India

II. MOTIVATION

Building a model to learn and analyze human movements in real time for evaluation is a difficult task. So we need research like this. Several approaches already exist to tackle the problem of recognition of human activity. Some use video, others portable sensors, such as an accelerometer, a digital compass, an angular speed sensor. The problem with these approaches is that it is expensive or requires fixed infrastructure. Therefore, smart phones use smart phones because they can easily use smart phones, which are relatively inexpensive and easy to use. R is open source; it is the first choice to analyze with more than 8000 packages. In addition, the social impact of this topic is great. If sports scientists and doctors can study personal movements that better understand their users and should help find new learning and new results. Therefore, it is essential in my aspect to study this subject and discover new discoveries for others.

III. PURPOSE

This Research is made to analyze the situation in real life. This provides a better approach for processing large data associated with different scenarios. The Human Activity Certified dataset specifically has the following objectives.

• Check the results to determine if the recognition is suitable for real life.
• Introduce the main concept of human activity recognition and its application to real world problems such as surveillance systems and patient monitoring systems. Focuses specifically on the field of motor disorders.
• To better understand user behavior in applications, imagine ways to contextualize physical activity.
• The main objective is to compare accuracy and identify the most appropriate classification model for an accurate analysis.

IV. RELATED WORK

Previous research on the recognition of human activity using accelerometers shows that it is possible to classify several attitudes and activities in real time. In [9], we developed a two-layer model combining a Gaussian mixing model and a first-order Markov model and classified activities such as sitting, walking, cycling and subway. Only one 3-axis accelerometer installed on the cylinder was used. In [10], we classified the activities carried out in the carpentry workshop with an accuracy of 84.4% by combining the data of the three accelerometers and the two microphones placed at different places of the body. They modeled most activities using a single hidden Gaussian Markov model. The number of hidden states modeling each activity was selected by visual inspection. In [11], several classification algorithms of 20 types of physical activity were evaluated from data acquired using 5 biaxial accelerometers, and the overall recognition rate was 84%.
In [12], the authors classify six different activities with an accelerometer and are commonly used to mount electronic devices with an accuracy of 16.7% to 92.8% depending on the position of the accelerometer 6. It is ranked in two different positions of the body and functions to use.

V. METHOD

The necessary packages are as follows.
It is to remember that all these packages are available in R *version 3.5.1*and higher.
Useful for drawing graphics in the Graphic Output
library (ggplot2)
Helps read data from a dataset; CSV file
library (readr)
Helps for svm library; to apply the SVM algorithm
library ("e1071")
After downloading the package, run the following command to start importing data with RStudio.
The train dataset includes 563 variables and 2947 observations, and the test dataset includes 563 variables and 7352 observations.

Steps

Step 1. #Extracting training data separated by blank. The train dataset includes 563 variables and 2947 observations.
traindata.df <- read.csv (paste ("train.csv", sep = ""))
Step 2. #Extracting test data separated by blank. The test dataset includes 563 variables and 7352 observations.
testdata.df <- read.csv (paste ("test.csv", sep = ""))
Step 3. #The database is attached to the R search path. So objects in the database can be accessed by simply giving their names.
attach (traindata.df)
attach (testdata.df)
Step 4. #Load psych package, a general purpose toolbox.
library (psych)
Step 5. #describe provides concise statistical description.
describe. (traindata.df)
describe (testdata.df)
Note: To verify, command dim(data) provides 10299 observations.
Step 6. #After a thorough review of the dataset, we combine the train and test data sets and create a single data set with a simple name for easy access. The combined dataset includes 563 variables and 10299 observations.
data <-rbind (traindata.df, testdata.df)
Step 7. #Applying naming transformation to entire dataset.
nameVec <- make.names (names (data), unique = TRUE)
names (data) <- nameVec
Step 8. #Subdivide the data in a ratio of 70: 30, with 70 =(10299*.7=7209) being the training data set and 30 being the test data set.
traindata <- data [1: 7209,]
testdata <- data [- c (1: 7209),]
pc.pvar <- pc.var / sum (pc.var)
Step 9. #Applying Principal Components Analysis (PCA) to reduce number of dimensions. Center= TRUE==> the data will be centered and cale=TRUE==>scaled before the analysis.
pc <- prcomp (traindata [, - 563], center = TRUE, scale = TRUE)
pc.var <- pc $ sdev ^ 2
pc.pvar <- pc.var / sum (pc.var)
Draw the graph:

Step 10. #Plotting Cummulative proportions of Principal Components to decide number of components to be taken into consideration
xlab and ylab=label of x-axis and y-axis, type=both points and lines, main ➔heading and color=”red”.
plot (cumsum (pc.pvar), xlab = "Principal component", ylab = "cumulative Proportion of variance explained", type = 'b', main = "Principal Components proportions", col = "red ")

#Add horizontal line.
abline (h = 0.95)

#Add vertical line.
abline (v = 100)

See snapshot at Figure 1 as output of PCA.

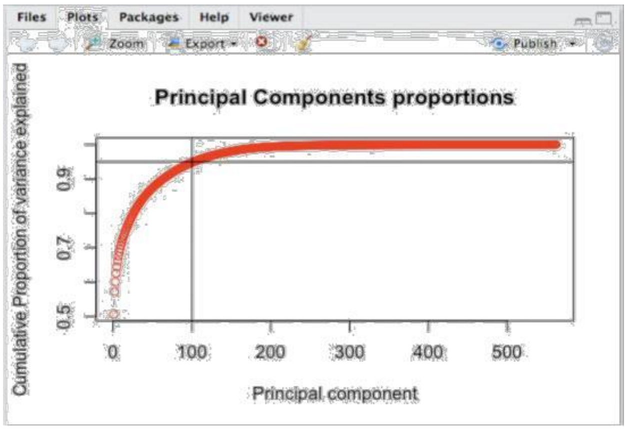

Figure 1: Explanation of Main Component Graph and Distributed Cumulative Ratio

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

OPEN ACCESS

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Select the first 100 main variables:

Step 11. # Graph predicts that first 100 principal components provides 95% of variance in dataset

train.data <-data.frame (activity = traindata $ Activity, pc $ x)

# Selecting first 100 principal components
train.data <-train.data [, 1: 100]

Step 12. #Training our model with SVM (Support Vector Machine) algorithm. We will use this model because it is the most efficient classification algorithm for large datasets.

svm_model <- svm (activity ~, data = train.data)

Step 13. #Preparing testing data for modelling with PCA(Principal Component Analysis)

test.data<-predict(pc,newdata=testdata)
test.data<-as.data.frame(test.data)
test.data<-test.data[,1:100]

Step 14. #Predicting testing data with train SVM model

result<-predict(svm_model,test.data,type="class")

Step 15. #Generating Confusion Matrix

test.data$Activity=testdata$Activity
references<-test.data$Activity

VI. CONCLUSION

We have analyzed the set of data extracted from UCI using R which become a subject of recognition of human activity. We found that the SVM algorithm is best suited to get the best results from the HAR dataset. We showed how to analyze the activities of the subject, from the recording of the activity to obtain the result. This shows how doctors or supervisors can use this technique to investigate candidates and use them easily to solve problems. In addition, an accuracy of 93.8% indicates the suitability of the model and guarantees the best results of use. We classify the activities to be done. This document will drive future research in more productive areas.

AppendixExperimental Setup and Data Collection

t<-table(references,result)

See snapshots at Figure 2.

```
> test.data$Activity=testdata$Activity
> references<-test.data$Activity
> t<-table(references,result)
> t
                     result
references            LAYING SITTING STANDING WALKING WALKING_DOWNSTAIRS
  LAYING                 534       0        0       0                  3
  SITTING                  4     419       63       0                  4
  STANDING                 0      29      500       1                  2
  WALKING                  0       0        0     477                 19
  WALKING_DOWNSTAIRS       0       0        0       3                409
  WALKING_UPSTAIRS         0       0        0      21                 23
                     result
references            WALKING_UPSTAIRS
  LAYING                             0
  SITTING                            1
  STANDING                           0
  WALKING                            0
  WALKING_DOWNSTAIRS                 8
  WALKING_UPSTAIRS                 427
```

Figure 2. Confusion Matrix

VII. RESULT

Step 16. Calculating error from confusion matrix
accuracy <- (t[1,1]+t[2,2]+t[3,3]+t[4,4]+t[5,5]+t[6,6])/sum(t)
AccuracyRate <- accuracy*100
c("Accuracy", AccuracyRate)

The output of step 23 is: "Accuracy" "93.8581608415338". This ensures that we used the correct model and it can be implemented in our day-to-day lives for any type of activity recognition.
The experiments were conducted on groups of 30 volunteers aged 19 to 48 years. Each person has done various activities such as walking, walking upstairs, walking downstairs, sitting, standing, sleeping at the waist with a Smartphone. The experiments were videotaped to facilitate data labeling [7]. The Samsung Galaxy S2 Smartphone is used for experimental purposes because it includes accelerometers and gyroscopes to measure linear acceleration and angular velocity, respectively, at a constant speed sufficient to capture the movements of the human body. We also use an app to easily retrieve Google Play store data. The recognition process begins with the acquisition of the sensor signal and is then preprocessed by applying a noise filter and then sampled in a fixed window according to the calculation. Next, look for the frequency component of the signal using the fast Fourier transform. Finally, they are used as entries of SVM classifiers trained for activity recognition.

R Studio

R Studio is a free open source open source IDE for R, which is a free software environment for statistical computing and graphical programming languages. RStudio was founded by JJ Allaire, the creator of the ColdFusion programming language. It is written in C ++ programming language and uses the Qt framework for the graphical user interface. R Studio has two editions. The program is an R Studio desktop that runs locally as a normal desktop application. R Studio Server can access R Studio using a web browser while running on a remote Linux server (see Figure 3).

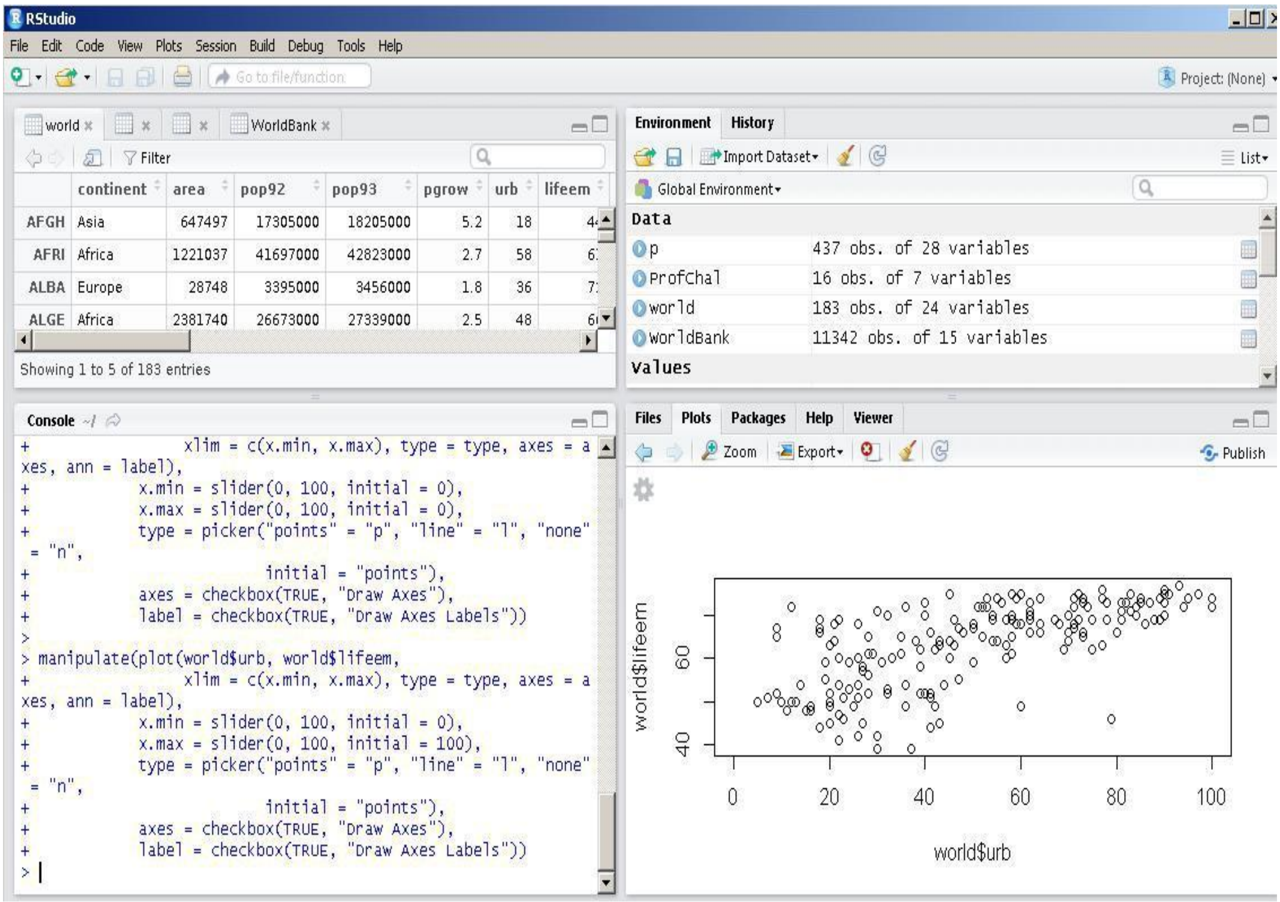

Figure 3. R Studio window

We used R Studio Desktop Open Source and developed RStudio Desktop to run on desktop computers (Windows, macOS, Linux). RStudio Desktop includes a variety of robust tools for the console, a syntax highlight editor for direct support for code execution, tracing, history viewing, debugging, and space management working. A typical R Studio window looks like the figure 6 below. It consists of 4 sections:

1. In the "R script" section at the top left, to view the script.
2. Section R of the console under the section Script R. Here we execute the function to get the result.
3. Environment section R at the top right. It includes all datasets used in the Research.
4. Use the graphical output section under the R environment to display the graph and deepen understanding.

Confusion Metrics

A confusion matrix, also called a contingency table, provides a complete overview by summarizing the results of the classification. By tabulating the predicted classes and the actual classes, we show the individual results of each class. Table 1 shows the confusion matrix and its components.

Table 1. Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

TP-It shows the true positive which is the number of positive records correctly predicted by the TP model.TN-It shows the true negative which is the number of negative records correctly predicted as negative by the TN model.

FP-It indicates a false detection, which is the number of negative records correctly predicted to be positive by the FP model.

FN-It shows false negatives, which is the number of positive records falsely predicted as negative by FN - model

Each of these measures has its own meaning depending on the research questions. We can define several standardized measures from the confusion matrix

- Accuracy – It is the total number of correct predictions proposed by the model which includes the positive and negative predictions.

Accuracy = number of correct predictions/ total number of prediction

i.e.

$$= \frac{TP + TN}{TP + TN + FP + FN}$$

- Recall or True Positive Rate – It is the proportion of positive classes identified correctly by the model.

Recall = number of correct positive predictions / total number of positive cases

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

$$= \frac{TP}{TP + FN}$$

- Precision – It is the fraction of positive cases correctly identified over all the positive cases predicted
  Precision = number of correct positive predictions/ total number of positive predictions
  $$= \frac{TP}{TP + FP}$$

Support Vector Machine (SVM)

In data analysis or decision-making science, in most cases we encounter situations in which we have to classify the data according to specific dependent variables. Support solutions for this applicable method. Logistic regression, random forest algorithm, Bayesian algorithm, etc. SVM is a machine learning technique that separates data from categories (ie margins).

- **Maximum Margin Classifier:**

Maximum margin classifiers are an easy way to separate separable data (or via a separable plane in a multidimensional space). We can see why it cannot be applied to the data.

- **Support for the vector classifier:**

SVC is an extension of the maximum margin that allows for incorrect classification.

- **Support for the vector machine:**

The SVM vector support vector is an additional extension of SVC to support nonlinear boundaries. Different definitions are clearly distinguished, but people prefer to avoid complications by calling SVM.

The following elements are separable linearly, for example linearly after the data set, and can be classified using any linear separation technique shown in Figure 4.

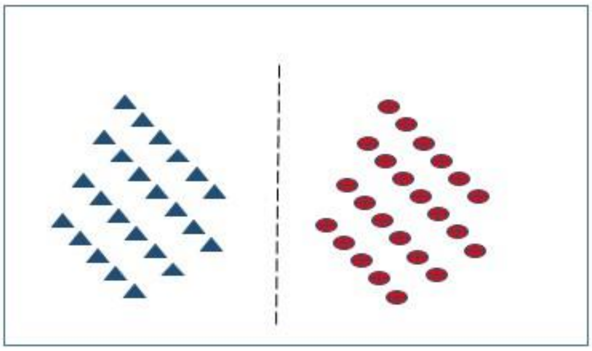

Figure 4. Linear Classification of dataset

But, What if we have data something like given in the following example?

In the below example Figure 5, there is no separating boundary which can divide this data in the current dimensions (1D) but can there be any boundary which can promote the data in higher dimensions?

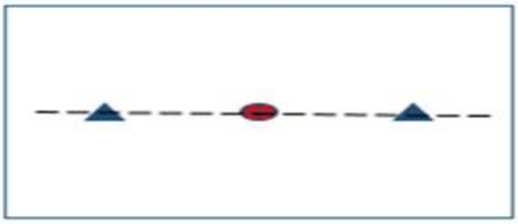

Figure 5. Another data set with no linear seprate boundary

Maximum Margin Classifier

Before we start let's first understand "What is Hyper plane?" In a P-Dimensional space, a hyper plane is a flat subspace having dimensions "P-1" i.e., in a 2-D space a hyper plane will be a line-

$B_0+B_1X_1+B_2X_2=0$

While in an m-dimensional space-

$B_0+B_1X_1+B_2X_2+\ldots+B_mX_m=0$

A hyper plane divides the space into two parts –

$B_0+B_1X_1+B_2X_2+\ldots+B_mX_m< 0$, and

$B_0+B_1X_1+B_2X_2+\ldots+B_mX_m>0$

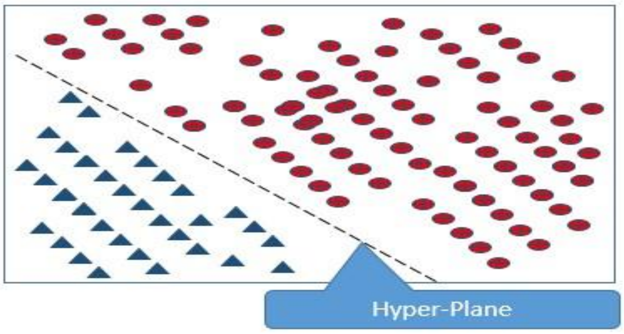

Figure 6. Hyper-Plane

An example of hyper-plane shown is in Figure 6 above. As we can see that there can be multiple possible such hyper planes which divide the data. Consider the following illustration in Figure 7 for this point. This represents a chaotic situation in which an analyst chooses one. However, this is where the maximum margin classifier appears. The maximum margin classifier helps one choose the best solution from a set of possible solutions. Our goal is to identify the separation hyper plane furthest from the observation. When calculating the vertical distance between each point of the training data set and the separation hyper plane, the optimal solution has a plane with a maximum margin.

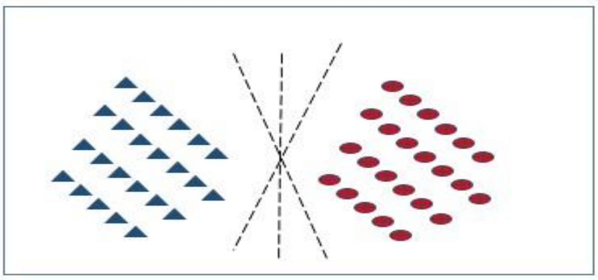

Figure 7. Multiple possibilities of Hyper-Plane

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

When inserting a slab to separate the data, the optimum solution has the maximum width and the center line of the slab is called the hyper plane shown in Figure 8.

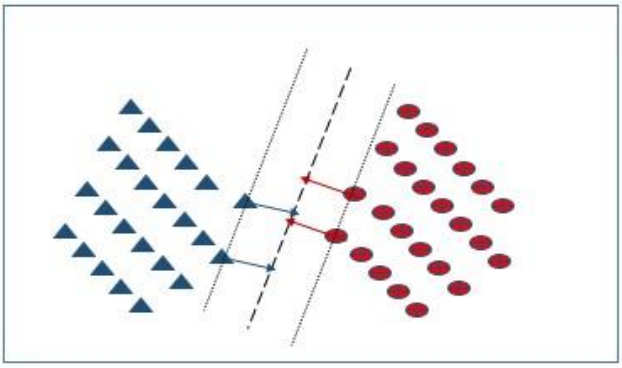

Figure 8. Center line of the slab is called the hyper plane

In the example shown in Figure 8, we can see that there are 4 points which are nearest to the boundary or are defining boundary, these points are called "Support Vectors".

REFERENCES

1. What is R? https://www.r-project.org /about.html, accessed on 4/9/2017 at 10:30 AM.
2. https://archive.ics.uci.edu/ml/ datasets/ accessed on 5/10/2017 at 11 AM.
3. Video showing how readings were recorded http://www.youtube.com/watch?v=XOEN9W05_4A, accessed on 5/10/2017 at 12 AM.
4. What is RStudio? https://en.wikipedia.org/ wiki/RStudio, accessed on 10/10/2017 at 12 AM.
5. Mumtaz, A., Zhang, W., and Chan, A. B., "Joint motion segmentation and background estimation in dynamic scenes," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Columbus, OH), pp. 368–375, 2014.
6. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz., "Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine.", in Vitoria-Gasteiz, Spain,, pp 1-8, 2013.
7. Davies, N., Siewiorek, D.P., Sukthankar, R., "Activity-based computing.", IEEE Pervasive Computing, Volume 7 issue 2,pp. 20–21, 2008.
8. Diane J. Cook, Juan C. Augusto, and Vikramaditya R. Jakkula., "Ambient intelligence: Technologies, applications, and opportunities." Pervasive and Mobile Computing, Volume 5 issue 4, pp. 1-8, 2007.
9. U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In International Workshop on Wearable and Implantable Body Sensor Networks, pp 1-50 April 2006.
10. P. Lukowicz, J. A. Ward, H. Junker, M. Stager, G. Troster, A. Atrash, and T. Starner. Recognizing workshop activity using body worn microphones and accelerometers. In Proceedings of the 2nd International Conference on Pervasive Computing, pp18–22, April 2004.
11. L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In Proceedings of the 2nd International Conference on Pervasive Computing, pp 1–17, April 2004.
12. R. W. DeVaul and S. Dunn. Real-time motion classification for wearable computing applications. Technical report, MIT Media Lab, Dec 2001.

AUTHORS PROFILE

Ajay Agarwal received B Tech degree from IET Lucknow, 1979. He did Master of Engineering with honors (Computer Science and Engineering) from M.N.N.I.T. Allahabad in 1995 and Ph.D. from I.I.T. Delhi in 2006. He is a versatile professional with a strong academic background His area of research includes mobile adhoc networking, soft computing, data base management system. He has received many awards. He has published many papers in reputed journals and conferences. He is an author of a book. Presently he is working as Professor in Information Technology Department at KIET Group of Institutions, Ghaziabad. [ajay.agarwal@kiet.edu]

Ashutosh Sharma is MCA, M.Tech. (IT) and currently working as an Assistant Professor in IT Department of IMS-Ghaziabad. He is having experience of more than 14 years in both teaching and Industry. His areas of Interest are Computer Networks, Advanced Computer Networks, WAN Routing Protocols, LAN Technologies & Client Server Technologies.
[**ashutosh.sharma@imsuc.ac.in**]

Amit Kumar Gupta received MCA degree from Kurukshetra University, Haryana 2003. He received his Ph.D. (CS) from Bundelkhand University, Jhansi, 2014. He is currently an Associate Professor in the Department of Computer Applications at Kiet Group of Institutions affiliated from AKTU, Lucknow. His research interests include Neural Networks, Fuzzy Logic, Reliability, Artificial Intelligence, mobile computing, wireless computing, Computer Network. [amit.gupta@kiet.edu]

Vikas Goel B.Tech. degree in Information Technology from VBS Purvanchal University, Jaunpur, India, in 2001. He received his master degree (M.Tech.) in Computer Science from Shobhit University, Meerut, India, in 2009. He received his Ph.D. (CSE) from Uttarakhand Technical University, Uttarakhand, 2017. He is currently an Associate Professor in the Computer Science & Engineering department at Ajay kumar Engineering College affiliated from AKTU, Lucknow. His research interests include mobile computing, wireless computing, broadcasting data in mobile devices, distributed computing.
[goelvikas@akgec.ac.in]
goelvikas

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication