International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 06th Feb 2023. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=ISSUE-02

DOI: 10.48047/IJIEMR/V12/ISSUE 02/14

Title **Human Activity Classification based on Smartphone Data using ML Techniques**

Volume 12, Issue 02, Pages: 100-104

Paper Authors
Ch.Praneeth, Sujan Chowdary P, Sairam M, Venkata Naga Janardhan K, Sai Sankar S

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Human Activity Classification based on Smartphone Data using ML Techniques

[1]**Ch.Praneeth**, Asst.Professor, IT Department, PVP Siddhartha Institute of Technology
[2]**Sujan Chowdary P**, IT Department, PVP Siddhartha Institute of Technology
[3]**Sairam M**, IT Department, PVP Siddhartha Institute of Technology
[4]**Venkata Naga Janardhan K**, IT Department, PVP Siddhartha Institute of Technology
[5]**Sai Sankar S**, IT Department, PVP Siddhartha Institute of Technology

Abstract
Today's technology has rendered machine learning a buzzword, and it is evolving very quickly. Even though we are unaware of it, machine learning is a part of our daily lives. Human activity recognition has wide applications worldwide in Human Service System. In this paper we design a sturdy activity recognition system based on smart phone data collection. The system uses embedded accelerometer and gyroscope to collect 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50hz. Activities are classified using various classifiers namely k-nearest neighbor classifier, SVM classifier, Decision tree classifier, Random Forest classifier and Naïve Baye's classifier. We implemented passive learning algorithms to compare the accuracy between them.

Keywords: Machine learning, Classification, Accuracy of a classifier.

Introduction

The ultimatum for understanding human activities have grown in medical field, especially in elder care support, rehabilitation assistance and mental disorders. A huge number of resources can be conserved if custodians are alerted by sensors in case of abnormal behavior detected in patients. Many studies have efficiently identified activities using wearable sensors with low error index, but the majority of the previous works are done in the laboratories with constraints. Observations from multiple sensors achieve low error-rate, but the complex constraint is not feasible in practice. This project uses feasible and commercially available smartphones as sensors to identify human activities[5]. The growing popularity and computational power of smartphone make it an ideal tool for comfortable body-attached sensors. Smartphones have built-in sensors such as accelerometer or gyroscope. Observations have shown that gyroscope can help activity recognition even though its contribution alone is not as efficient as accelerometer. Because gyroscope cannot be accessed in cellphones as accelerometer, our paper only uses readings from a 3-dimensional accelerometer. Contrastingly from many other works before, we relaxed the constraints of attaching sensors to fixed body position with fixed device orientation. In our design, the phone can be placed at any place around waist such as jacket pocket and pants pocket, with arbitrary orientation. These are the most common positions where people place their mobile phones. The aim of this paper is to design a light weight and accurate system on smartphone that can recognize human activities. Through testing and comparing various learning algorithms, we find the best fit for our system in terms of accuracy on a smartphone.

Within Machine learning there are two approaches[1]. They are Supervised learning and Unsupervised learning. The machine learns while being watched over in supervised learning. It includes a model that can make predictions using data that has been labeled. A dataset that has been labeled means that you already know the intended response. Machine learning algorithms are used in unsupervised learning to examine and group unlabeled data sets. Without the aid of humans, these algorithms find hidden patterns in data. Supervised learning can be further divided into two types called Classification and Regression. When an output variable has two or more classes and is categorical, classification is

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

employed and when the output variable has a real or continuous value, regression is employed.

The Human Activity Recognition dataset was built from the recordings of 30 study participants performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors[4]. The objective is to classify activities into one of the six activities such as walking, jogging, moving upstairs, downstairs, running, sleeping performed.

Method

In this we used different machine learning models and compared their respective result accuracy to the given data set[6]. The machine learning models used for the experiment are k-nearest neighbor classifier, SVM classifier, Decision tree classifier, Random Forest classifier.

Before implementing the algorithms we normalized the data in the range of 0 to 1 i.e, feature scaling based normalization so that it helps the gradient descent to work more efficiently as there is more uniform data point across the plot.

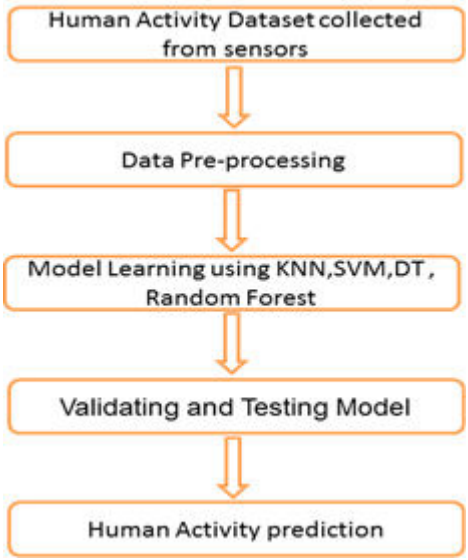

Figure 1: Human Activity Recognition System

KNN Classifier

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category[2].

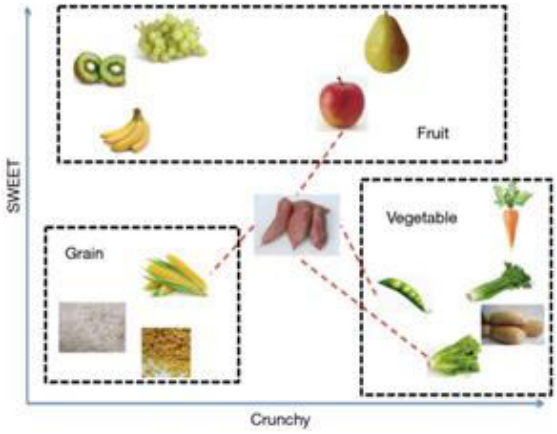

Figure 2: KNN classifying a new sample

SVM Classifier

Support Vector Machine(SVM) is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog[3]. On the basis of the support vectors, it will classify it as a cat.

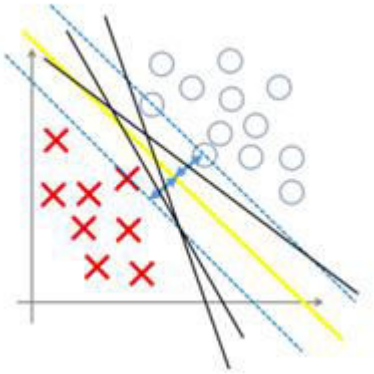

Figure3 : Hyper plans that separate training samples

Decision Tree Classifier:

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset.

It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

Suppose there is a candidate who has a job offer and wants to decide whether he should accept the offer or Not. So, to solve this problem, the decision tree starts with the root node (Salary attribute by Attribute selection measure). The root node splits further into the next decision node (distance from the office) and one leaf node based on the corresponding labels. The next decision node further gets split into one decision node (Cab facility) and one leaf node. Finally, the decision node splits into two leaf nodes (Accepted offers and Declined offer).

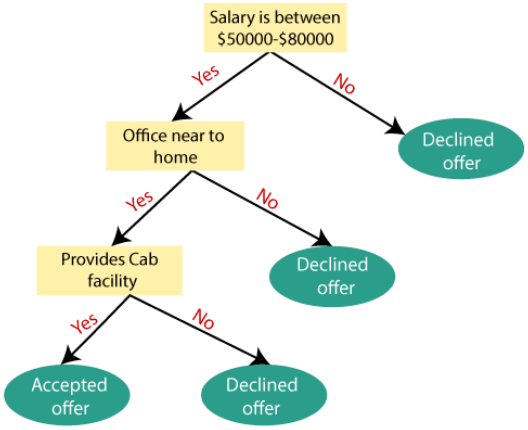

Figure 4: KNN classifying a new sample

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Random Forest Classifier:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

Suppose there is a dataset that contains multiple fruit images. So, this dataset is given to the Random Forest Classifier. The dataset is divided into subsets and given to each decision tree. During the training phase, each decision tree produces a prediction result, and when a new data point occurs, then based on the majority of results, the Random Forest classifier predicts the final decision.

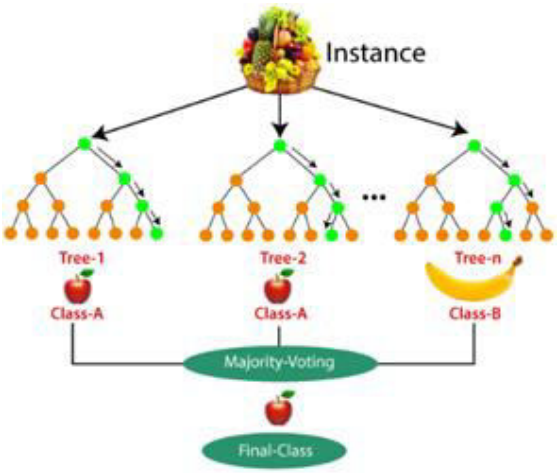

Figure 5: Random forest classifying a new sample

Results:

The training set of data is used to teach the model and help it discover any hidden characteristics or trends. The validation set is a set of data that isn't part of the training set and is used to verify the performance of our model as it's being trained. We can adjust the model's hyper-parameters and configurations using the data from this validation procedure. It functions similarly to a critic telling us whether or not our training is on the right track. After training is complete, the model is tested using a different collection of data called the test set. It offers a neutral measure of the final model's performance in terms of precision, accuracy, etc.

In this paper we test the accuracy of the dataset against validation and test datasets using different algorithms and the result of the models are as below:

Model /Accuracy	Validation data	Test data
KNN	0.9678150498640072	0.9632819582 95557
SVM	0.9419764279238441	0.1425178147 2684086
Decision Tree	0.9378966455122394	0.1951136749 2365117
Random Forest	0.9768812330009066	0.1425178147 2684086

Table 1: Comparing accuracy of models with validation dataset and test dataset.

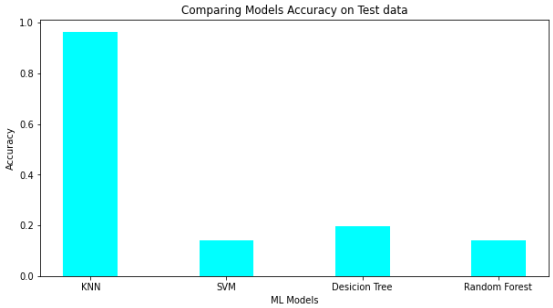

Figure6: Comparing accuracy of models with test dataset.

From the Observation, the KNN classifier has more accuracy while the SVM classifier has the low accuracy on the human activity classification data set.

Conclusion:

In light of the user's input, this study offered a method for identifying and forecasting human activity utilising machine learning algorithms such

Decision tree classifier, Random Forest classifier, K-Nearest Neighbors, and Support Vector Machines. As a result, the suggested techniques also result in a comparison of various machine learning methods for multiclass classification. We saw that the KNN classifier had excellent accuracy for the data on human activity recognition.

References:

1. M. Habib, Z. Wang, S. Qiu, H. Zhao and A. S. Murthy, "Machine Learning Based Healthcare System for Investigating the Association Between Depression and Quality of Life," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 5, pp. 2008-2019, May 2022, doi: 10.1109/JBHI.2022.3140433.
2. Zhang Z. Introduction to machine learning: k-nearest neighbors. Ann Transl Med. 2016 Jun;4(11):218. doi: 10.21037/atm.2016.03.37. PMID: 27386492; PMCID: PMC4916348.
3. Zhao., The Development and Application of Support Vector Machine Jun 2021 J. Phys.: Conf. Ser. 1748 052006
4. Zahraa S. Abdallah, Mohamed Medhat Gaber, Bala Srinivasan, and Shonali Krishnaswamy., Activity recognition with evolving data streams: A review. Comput. Surv. 51, 4 (2018),
5. Bayat, A., Pomplun, M., Tran, D.A.: A study on human activity recognition using accelerometer data from smartphones. Procedia Comput. Sci. 34, 450–457 (2014).
6. CORSTON-OLIVER, Simon, Michael GAMON, and Chris BROCKETT (2001): A Machine Learning Approach to the Automatic Evaluation of Machine Translation. Proceedings of ACL, pp. 140-147.