ISSN: 2456-9348
Vol (02) _Issue (05) Impact Factor: 4.520

iJETRM
International Journal of Engineering Technology Research & Management

ACCELEROMETER BASED HEALTHCARE DEVICE USING NAÏVE BAYES CLASSIFIER

Ajay Krishnan MK*[1]
Nalini Nagendran[2]
*[1] VIT Vellore(School of Computer Engineering)
[2] VIT Vellore (School of Computer Engineering)

ABSTRACT

The domain of health-care is still relatively untouched by digitization. Doctors rely on ad-hoc methods of diagnosis since they have no way to share or record a patient's lifestyle history. People come from varied genetic backgrounds and follow different lifestyles and recording such data could be useful for the early diagnosis of critical conditions. One such metric that could be helpful is the amount of physical activity. A few startups have live activity tracking devices in the market but they are prohibitively expensive. The goal of the paper is to relay real time health statistics to the user and tabulate energy expenditure and effort based on the activity performed. The Naive Bayes model was used since it is computationally inexpensive and the features across three axes are independent

Keywords:

Activity Classification, Wearable, Data-Cleaning, Naive-Bayes Model, Cross-Validation, Health-Statistics

INTRODUCTION

The domain of health-care is still relatively untouched by digitization. Doctors rely on ad-hoc methods of diagnosis since they have no way to share or record a patient's lifestyle history. People come from varied genetic backgrounds and follow different lifestyles and recording such data could be useful for the early diagnosis of critical conditions. One such metric that could be helpful is the amount of physical activity. A few startups have live activity tracking devices in the market but they are prohibitively expensive. In this study, an alternate, low-cost method of activity tracking was explored using an Inertial Measurement Unit (MPU 9150) , which had three sensors on board: an accelerometer, a magnetometer and a three axis gyro. The sensors and a wireless network chip (NODE MCU) with a dual-core processor on board were mounted to a tennis wrist band for real time human activity recording and classification. The data collection was done with 3 test subjects over various activities for periods exceeding 30 minutes per activity. The goal of the paper is to relay real time health statistics to the user and tabulate energy expenditure and effort based on the activity performed. Raw sensor data is relayed to a UDP server running on a laptop/mobile device for processing. The study involved using a Naive Bayes model to classify data from the wearable device. Raw data was aggregated into a feature vector set with 8 time domain features for consideration. Data-cleaning and data classification was done on the raw-data. The Naive Bayes model was used since it is computationally inexpensive and the features across three axes are independent. Accuracy levels of 98.9% were achieved. Features were verified by cross validation over a test set and the performance of the model was evaluated using cross validation. Confusion matrix and graph plots of features were used to visually display results. An elementary graphical user interface was used to further display results in a relevant and useful manner. The wearable can provide real time statistics about physical activities

RELATED WORK

According to [1], Naive Bayes classifiers proves to be very effective with a few time and frequency domain features. [1] also suggests Incorporating additional features and/or activities through expanding the dimension of the feature vector and estimating feature probability density functions (PDFs) for new activities and/or features.However,it Employs TCP/IP which is reliable but not a viable option because of its three-way. acknowledgement principle since large number of packets are sent at a time
[2] proposes A digital low pass filter in order to separate the AC component from the DC component in each time series. For feature extraction, window overlapping is used where we subdivide the data set into smaller

IJETRM
International Journal of Engineering Technology Research & Management

subsets and window them individually. But the Accuracy is comparatively lower than those of other methodologies.
There is no provision to handle large sample sizes either.
[3] has adopted a method to Classify physical Activity in organic real-life conditions for physical activity research research utilizing data carried out by distributed cell phone accelerometer . Gradient boosting and Random Forrest methods were employed to predict sitting, jogging, standing and walking. The training models generated in view of the two classifiers were tested on accelerometer based data gathered from the cell phones of two subjects in organic living conditions. However,it faced difficulty in classifying extreme body movements. There were also readings that Misclassifies brisk paced walking to jogging. Instances where Struggles with the detection during transportation was also noted. [3] also requires the use of smartphones to be restricted during the data collection process.
[4] Employs an unsupervised method for recognizing physical activities using smartphone accelerometers. Features are extracted from the raw acceleration data collected by smartphones, then an unsupervised classification method called MCODE is used for activity recognition. The idea behind the implementation is to group similar actions together in clusters and decide on the misfits as and when they arrive. Due to this rigid nature of the approach, it is difficult to use it on large datasets.
[5] Proposes a OVO(one vs one) decomposition that employs the divide and conquer principle to give a group of diverse and simpler learners . The paper also goes on to propose a novel weighted combination for this decomposition to dynamically change itself thereby making the necessary adaptation. Active learning paradigm has been proposed to take into account the most important objects . The base learner that has been used is Naïve Bayes classifier which has the shortcoming of being efficient only for a small scale of classes. But the proposed ensemble has each classifier to perform only a binary task thereby making it very efficient Even though this addresses all the shortcomings of the previous methods, it is Very difficult to execute due to its nature of complexity.

METHODOLGY

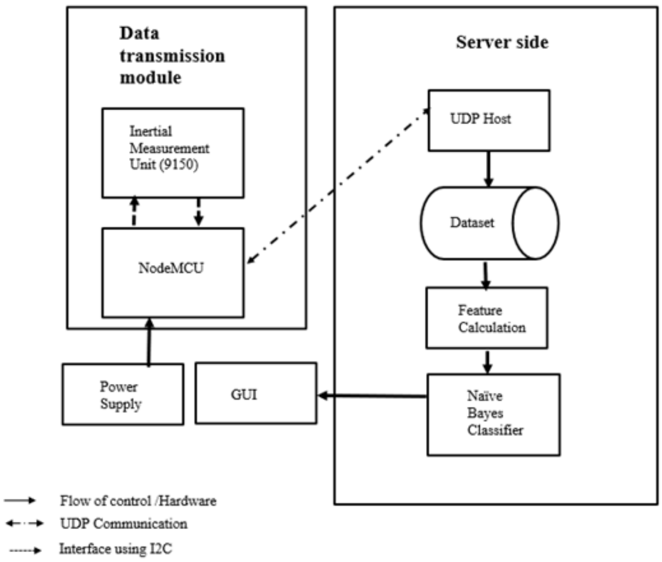

Fig. 1: Architecture Diagram

The communication between the NodeMCU processor and the Inertial Measurement Unit(IMU) was carried utilizing the I2C protocol at a speed of 400Mhz. A WiFi router is utilized to establish a common network to the NodeMCU and the PC running the UDP server. TCP has been overlooked since it requires every packet to be acknowledged which prompts a slower data exchange rate. The UDP server running on the PC gets readings from the processor and transfers them to a circular buffer which stores 400 of the latest values. Features are computed from these readings at a rate of 1Hz and fed to the Naive Bayes model, which is trained beforehand to predict the corresponding activity class. The prediction is then made to appear on a graphical interface dynamically. The normal testing rate that were accomplished in trial runs was around 100-150 examples for every second. It is conceivable to accomplish much higher testing rate utilizing all the more intense switches and furthermore MIMO recieving wires in both the transmitter and the recipient. However an example rate of 100 examples for each second is more or less expansive to over-compensate for the human body's reflex and movement speed.

Training on a Public Dataset

Before taking up the custom dataset, an open dataset from UC Irvine's Machine Learning Repository was utilized for assessment. This dataset contained readings from four accelerometers set on the chest, wrist, foot and waist. An aggregate of 13 activities were perceived. For this application, the required values were acquired from the accelerometer on the wrist and a subset of the activities, like being walking, cycling, running and being stationary.

Feature Consideration

After perusing various research papers concerning related paper, a rundown of features in the time and frequency domains was concluded. The features decided upon for the Naive Bayes classifier were maximum value, minimum value, standard deviation ,difference between maximum and minimum, mean absolute deviation, root mean square, median absolute deviation and interquartile range. These features were decided for their computational straightforwardness for free-life applications. The highlights were registered on a rolling window of 400 readings (2x test rate) for every axes. The subsequent feature matrix was put away in a file for later use. Roughly 70% of this dataset was utilized to train the classifier.

Calibration

Prior to its use, the sensor has to be tuned and calibrated.. Offsets should be found to discover a reference frame from which reliable precise readings could be acquired. The values for accelerometer were adjusted to the x,y,z plane correspondingly. The gyroscope was adjusted to demonstrate a 0 perusing on every one of the 3 axes and the magnetometer was lined up with the magnetic field of the Earth subsequent to eliminating hard iron biases to keep a steady reference frame for the IMU. The constant frame was essential for the filtering and sensor fusion algorithm which was utilized to acquire the orientation of the sensor frame as for Earth's reference frame. This calibration was done as such that the user is permitted to wear the device however he sees fit any orientation on either hand. The sensor will adjust in like manner to Earth's magnetic field to balance any adjustments in orientation and all readings will be procured from that reference frame .

Data Cleaning using Filtering Algorithms

Accelerometers don't experience the ill effects of drift yet their readings are exceptionally noisy There are numerous open source algorithms like Mathony an Madgwick filters, Kalman Filters and Complimentary filters. Kalman filters are the most precise of the above mentioned filters but it amounts to many matrix computations. The Madgwick channel was decided for this application since it's moderately more straightforward to execute and approaches Kalman filters in terms of efficiency .

Data Acquisition

Data from the NodeMCU remote network chip was sent to a UDP server running on a PC or mobile application. UDP was decided for its relative speed in transmitting information as it doesn't require ACK for the transmitted packets. Since the transfer rate came to upto 150 samples in each second, the loss in data due to untransmitted packets could be neglected and it was conceivable to work with that measure of erroneous data. TCP was at first considered however but due to its aforementioned ACK principle,it had to be neglected .In order to collect mass amounts of data ,volunteers had to perform three different activities . The volunteers in the test wore the gadget on their lower arm and did the following exercises; Running, Walking and, Standing Still. The readings from every movement were to be recorded and classified to populate the training set.

ISSN: 2456-9348
Vol (02) _Issue (05) Impact Factor: 4.520

iJETRM
International Journal of Engineering Technology Research & Management

Feature Engineering

Several time domain and frequency domain features were considered for training the model.

• Temporal Domain Features

i. Minimum and maximum and its difference
ii. Mean and Standard Deviation
iii. Median and Mean absolute deviation
iv. Interquartile Range
v. Root Mean Square
vi. Integrated value using trapezoidal approximation Mean crossings
vii. Pearson correlation coefficient between two signals

• Frequency Domain Features

i. Main Frequency Component
ii. Spectral Entropy and Relative Spectral

Entropy Energy of the signal in some frequency bands of interest

Most of the temporal domain features were used for the research as they are faster to compute in real time. A few of the features like mean, integral and correlation were not optimal since the values were similar for all the classes.

Naïve Bayes Classifier

Naïve Bayes is a straightforward approach for developing classifiers: models that appoint class labels to problem cases, denoted as vectors of feature values, where the class labels are drawn from some finite set. It isn't a solitary algorithm for training such classifiers, however a group of algorithms in view of a common principle: all Naïve Bayes classifiers work under the notion that the value of a specific component is independent of the value of some other component, given the class variable. Regardless of their Naïve outline and clearly distorted assumptions, Naive Bayes classifiers have worked great in numerous intricate real-life situations. Naïve Bayes classifier provides the edge of needing only a little measure of training data to evaluate the parameters essential for classification. Naive Bayes techniques are a bunch of supervised learning algorithms in light of applying Bayes' theorem with the "Naive" supposition of independence between each pair of features. Given a class variable y and a dependent element vector x1 through xn, Bayes' theorem expresses the following relationship:

P(y|x1, ..., xn) = P(y)P(x1, ..., xn|y) P(x1, ..., xn) (1)

Using the Naive assumption that all the features are independent:

P(xi |y, x1, ...xi−1, xi+1, ..., xn) = P(xi |y) (2)

The relation can be simplified to P(y|x1, ..., xn) = P(y) Qn i=1 P(xi |y) P(x1, ...xn) (3)

Since P(x1, ...xn) is constant given the input, the following classification rule can be used:

P(y|x1, ..., xn) ∝ P(y) Yn i=1 P(xi |y) (4)

yˆ = argmaxyP(y) Yn i=1 P(xi |y) (5)

RESULTS

Over the course of the paper, the prime aim of activity classification was realized across three stages.

The first stage involved indicating the x,y and z coordinates of the IMU as per its relative position and magnetic field. This was important because it indicated if the sensors were working soundly and responded to the sudden relative motion.

Vol (02) _Issue (05)
ISSN: 2456-9348
Impact Factor: 4.520

The second stage involved plotting the relative motion of the sensor along each of the x,y and z coordinates in a graph. This graph responded to each movement differently with the resulting waves having varying amplitudes and frequencies for Stationary, Walking and Running state of motion.
The third stage involved presenting the state of activity after classification to the user. This worked with the help of Naïve Bayes principle. There are a few shortcomings in this result because a short burst of action is mistaken for running while absolutely zero motion is required for the activity to be classified as walking.

Table 1: Coordinates and Activity Classification

Sample No.	X	Y	Z	Activity
1	1.04	0.76	0.87	0
2	1.78	0.87	1.35	1

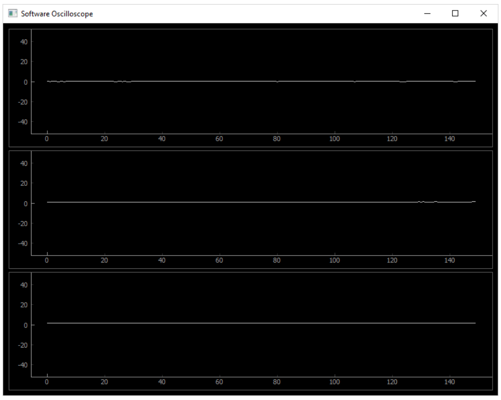

Fig 2. Stationary

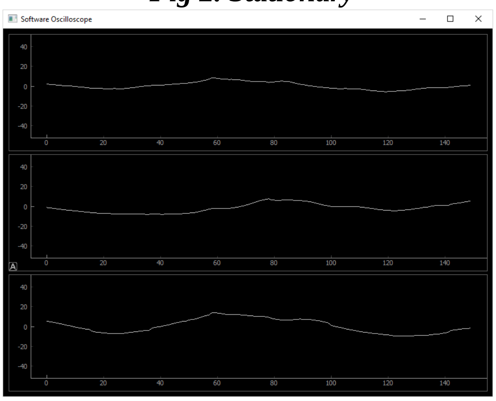

Fig 3. Walking

ISSN: 2456-9348
Impact Factor: 4.520

iJETRM
International Journal of Engineering Technology Research & Management

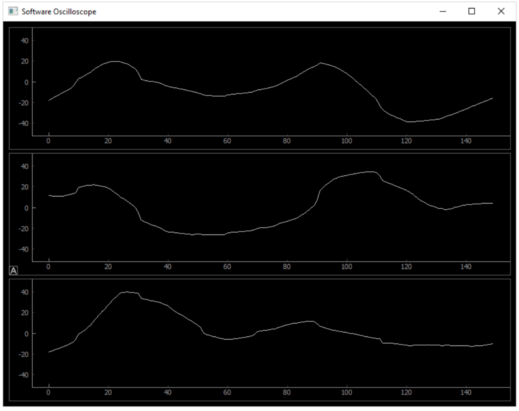

Fig 4. Running

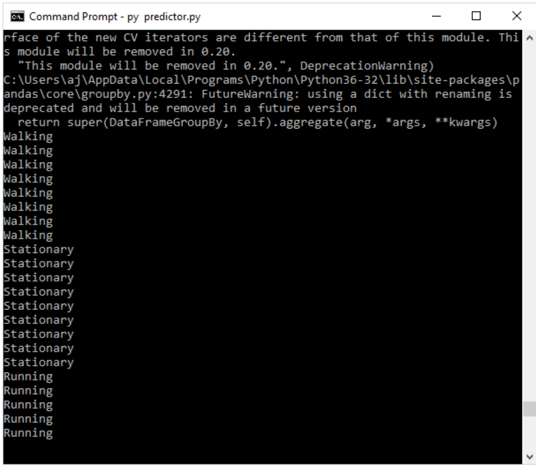

Fig 5. Activity Classification

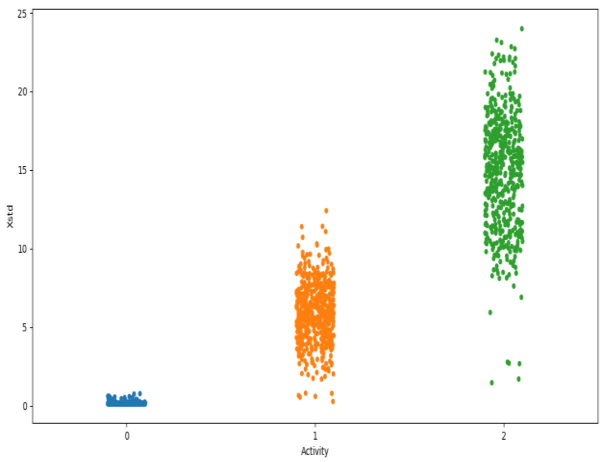

Fig 4.5 Standard deviation for each activity along x-axis

iJETRM
International Journal of Engineering Technology Research & Management

CONCLUSION

The frameworks accessible for this paper had the task of registering 150 examples per second and giving expectations at regular intervals of 2 seconds. It is conceivable to accomplish considerably higher prediction and sampling rates utilizing more powerful routers with MIMO and multi-threaded processing in the server side. There are likewise different sensors that could have incorporate into this gadget to make it more comprehensive and efficient, some of which could include incorporating an altimeter, blood oxygen sensor, heart beat sensor and so on. These would have brought about more particular and exact activity classification. For instance, a man being stationary demonstrates a sudden increase in heart rate action. This could imply that the individual is undergoing physical action or that he has a looming heart condition. The device is limited to only the power source on the user-end and the hardware on the receiving end. Visual investigation of the features indicated that parameters like mean, correlation and integral were nearly the same for all the three classes and were therefore causing misclassification errors.

FUTURE WORK

The applications of this gadget are wide. A case of a situation would be the scenario where a doctor could utilize this gadget to screen the wellbeing details of his patient and keep a recorded journal of his patient's month to month or yearly physical activities. This would help in endorsing custom-made medical advice to each patient.
Access to true class labels is limited, as every query would involve a human expert and cannot be done continuously during the course of sensor monitoring. This can be looked upon in subsequent research that can follow.

REFERENCES

[1] X. Long, B. Yin, and R. M. Aarts. Single-accelerometer-based daily physical activity classification. 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Sept 2009.

[2] Duc A. Tran Akram Bayat, Marc Pomplun. A study on human activity recognition using accelerometer data from smartphones. Procedia Computer Science, 34:450– 457, 2014.System Analysis Tools, An Overview. J. Clean. Product, 13 (12), 11651173, 2004

[3] Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results Kangjae Leea,⁎ , Mei-Po Kwanb , Illinois Informatics Institute, University of Illinois at Urbana-Champaign, National Center for Supercomputing Applications, MC-257, 1205 W Clark St., Urbana, IL 61801, USA Computers, Environment and Urban Systems 67 (2018) 124–131

[4] Yonggang Lu, Ye Wei ·Li Liu · Jun Zhong · Letian Sun · Ye Liu - Towards unsupervised physical activity recognition using smartphone accelerometers Published online: 9 January 2016 © Springer Science+Business Media New York 2016

[5] O].Bartosz Krawczyk - Active and adaptive ensemble learning for online activity. Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, USA recognition from data streams