**Project – Initial Report – Human Activity Recognition (HAR)**

1. **Abstract**

Human Activity Recognition (HAR) is classifying the activity of a person using responsive sensors that are affected by human movement. Both users and capabilities(sensors) of smartphones increase and users usually carry their smartphones with them. These facts make HAR more important and popular. This work focuses on the recognition of human activity using smartphone sensors using different machine learning classification approaches. Data retrieved from smartphones’ accelerometers and gyroscope sensors are classified to recognize the human activity. Results of the approaches used are compared in terms of efficiency and precision.

1. **Initial report**
2. Literature review

* Paper 1 - Bulbul, E., Cetin, A., & Dogru, I. A. (2018). Human Activity Recognition Using Smartphones. 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT).

Introduction - Smartphones are the most useful tools in our daily life and with the advancing technology they get more capable day by day to meet customer needs and expectations. An accelerometer has been standard hardware for almost all smartphone manufacturers. Since there is a meaningful difference in characteristics between data retrieved from these sensors, many features could be generated from these sensors' data to determine the activity of the person that is carrying the device. In this study, a dataset consisting of signals from the accelerometer and gyroscope of a smartphone carried by different men and women volunteers while doing different activities are classified using different machine learning approaches.

Dataset - Dataset consists of signals from a smartphone carried by 9 individuals performing 6 different activities. Activities performed are listed below with their corresponding codes.

• WALKING

• CLIMBING UP THE STAIRS

• CLIMBING DOWN THE STAIRS

• SITTING

• STANDING

• LAYING

Signals are recorded with a sampling rate of 50Hz and stored as time series for each dimension so 6 different signals were obtained (3 are from the accelerometer and the other 3 are from the gyroscope). The noise was filtered using median and 20Hz Butterworth filters to get more precise results. A second 3hz Butterworth filtering was applied to eliminate the effect of gravity in accelerometer signals. Values then normalized to (-1,1) interval. Euclid magnitudes of the values of 3 dimensions were calculated to merge 3-dimensional signals into one dataset. Finally, class codes (activity codes) given above for each row are added at the end of them among with the number that is given to each individual. In the end, the dataset consists of 2947 records with 561 features.

Proposed method - Supervised machine learning is used to recognize activity from dataset records. Different supervised machine learning models are designed using different classification approaches. Designed models first trained with training data that consists of 80% of the total dataset and then tested with the rest. Classification precision of models is tested and observed using 5-fold cross-validation. Methods used for classification are as follows:

• Decision Trees

• Support Vector Machines

• K-nearest neighbours (KNN)

• Ensemble classification methods

o Boosting

o Bagging

o Stacking

* Paper 2 - Wang, H., Zhao, J., Li, J., Tian, L., Tu, P., Cao, T., … Li, S. (2020). Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep Learning Techniques. Security and Communication Networks, 2020, 1–12.

Introduction - Human behavior recognition (HAR) is the detection, interpretation, and recognition of human behaviors, which can use smart health care to actively assist users according to their needs. Human behavior recognition has wide application prospects, such as monitoring in smart homes, sports, game controls, health care, elderly patients care, bad habits detection, and identification. It plays a significant role in in-depth study and can make our daily life smarter, safer, and more convenient. This work proposes a deep learning-based scheme that can recognize both specific activities and the transitions between two different activities of short duration and low frequency for health care applications.

Dataset - This paper adopts the international standard Data Set, Smartphone-Based Recognition of Human Activities, and Postural Transitions Data Set to conduct an experiment, which is abbreviated as HAPT Data Set. The data set is an updated version of the UCI Human Activity Recognition Using popularity Data set. It provides raw data from smartphone sensors rather than preprocessed data and collects data from accelerometer and gyroscope sensors. In addition, the action category has been expanded to include transition actions. The HAPT data set contains twelve types of actions. A total of 815,614 valid pieces of data are used here.

Proposed method - The overall architecture diagram of the method proposed in this paper contains three parts. The first part is the preprocessing and transformation of the original data, which combines the original data such as acceleration and gyroscope into an image-like two-dimensional array. The second part is to input the composite image into a three-layer CNN network that can automatically extract the motion features from the activity image and abstract the features, then map them into the feature map. The third part is to input the feature vector into the LSTM model, establish a relationship between time and action sequence, and finally introduce the full connection layer to achieve the fusion of multiple features. In addition, Batch Normalization (BN) is introduced, in which BN can normalize the data in each layer and finally send it to the Softmax layer for action classification.

* Paper 3 - Agarwal, P., & Alam, M. (2020). A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices. Procedia Computer Science, 167, 2364–2373.

Introduction - Here the architecture for the proposed Lightweight model is developed using Shallow Recurrent Neural Network (RNN) combined with Long Short Term Memory (LSTM) deep learning algorithm. then the model is trained and tested for six HAR activities on resource-constrained edge devices like RaspberryPi3, using optimized parameters. The experiment is conducted to evaluate the efficiency of the proposed model on the WISDM dataset containing sensor data of 29 participants performing six daily activities: Jogging, Walking, Standing, Sitting, Upstairs, and Downstairs. And lastly, the performance of the model is measured in terms of accuracy, precision, recall, f-measure, and confusion matrix and is compared with certain previously developed models.

Dataset - Here Android smartphone having an inbuilt accelerometer is used to capture tri-axial data. The dataset consists of six activities performed by 29 subjects. These activities include walking, upstairs, downstairs, jogging, standing, and sitting. Each subject performed different activities by carrying a cell phone in the front leg pocket. A constant Sampling rate of 20 Hz was set for the accelerometer sensor. A detailed description of the dataset is given in table 1 below.

Total no of samples: 1,098,207   
Total no of subjects: 29   
Activity   Samples: Percentage   
Walking   4,24,400   38.6%  
Jogging   3,42,177   31.2%  
Upstairs   1,22,869   11.2%  
Downstairs   1,00,427   9.1%  
Sitting   59,939   5.5%  
Standing   48,397   4.4%

Proposed method - The working of the Lightweight RNN-LSTM-based HAR system for edge devices. The accelerometer reading is partitioned into fixed window size T. The input to the model is a set of readings (x1, x2, x3,…….,xT-1, xT) captured in time T, where xt is the reading captured at any time instance t. These segmented window readings are then fed to the Lightweight RNN-LSTM model. The model uses the sum of rule and combine output from different states using a softmax classifier to one final output of that particular window.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Title | Author | Algorithm | Dataset | Accuracy |
| 2018 | Human Activity Recognition Using Smartphones | Erhan BÜLBÜL, Aydın ÇETİN, İbrahim Alper DOĞRU | Decision trees, SVM, KNN, Ensemble techniques | Nil | 91.7, 99.4, 97.5 |
| 2020 | Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep Learning Techniques | Huaijun Wang, Jing Zhao, Junhuai Li, Ling Tian, Pengjia Tu, Ting Cao, Yang An, Kan Wang, Shancang Li | CNN-LSTM | HAPT | 95.8 |
| 2020 | A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices | Preeti Agarwal, Mansaf Alam | RNN-LSTM | HAPT | 95.75 |

1. Problem definition

Human activity recognition is the problem of classifying sequences of accelerometer data recorded by specialized harnesses or smartphones into known well-defined movements.

It is a challenging problem given the large number of observations produced each second, the temporal nature of the observations, and the lack of a clear way to relate accelerometer data to known movements.

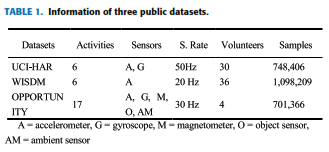
Classical approaches to the problem involve hand crafting features from the time series data based on fixed-size windows and training machine learning models, such as for ensembles of decision trees. The difficulty is that this feature engineering requires deep expertise in the field.

Recently, deep learning methods such as recurrent neural networks and one-dimensional convolutional neural networks or CNNs have been shown to provide state-of-the-art results on challenging activity recognition tasks with little or no data feature engineering.

1. Dataset exploration

Source of dataset

|  |  |
| --- | --- |
| Name | Source |
| UCI | <https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones> |
| WISDM | <https://www.cis.fordham.edu/wisdm/dataset.php> |
| OPPORTUNITY | <https://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition> |



* Dataset description

1. UCI-HAR

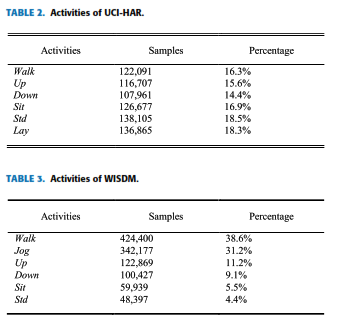
The UCI-HAR dataset was built from the recordings of 30 subjects aged 19-48 years. During the recording, all subjects were instructed to follow an activity protocol. And they wore a smartphone (Samsung Galaxy S II) with embedded inertial sensors around their waist. The six activities of daily living are standing (Std), laying (Lay), walking (Walk), walking downstairs (Down), and walking upstairs (Up). In addition, this dataset also includes postural transitions that occur between the static postures: standing to sitting, sitting to standing, sitting to laying, laying to sitting, standing to laying, and laying to standing. Specifically, in this paper, only six basic activities were selected as input samples due to the percentage of postural transitions being small. The experiments had been video-recorded to manually label the data. Finally, the researchers captured 3-axial acceleration and 3-axial angular velocity data at a constant rate of 50Hz. According to statistics, the number of samples in this dataset is 748406, and the detailed information was shown in Table 2.

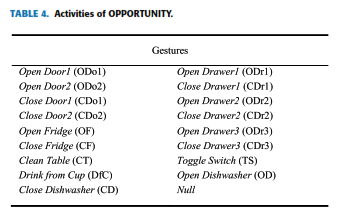
1. WISDM

The WISDM dataset has a total of 1098209 samples, and the percentage of the total samples associated with each activity was shown in Table 3. It can be seen that WISDM is an unbalanced dataset. Activity walking takes up the most, reaching 38.6% while standing only accounts for 4.4%. Its experimental object consists of 36 subjects. These subjects performed certain daily activities with an Android phone in their front leg pockets. The sensor used is an accelerometer with a sampling frequency of 20 Hz. It is also a built-in motion sensor of the smartphone. Six activities were recorded: standing (Std), sitting (Sit), walking (Walk), upstairs (Up), downstairs (Down), and jogging (Jog). The data collection was supervised by a dedicated person to ensure the quality of data.

1. OPPORTUNITY

The OPPORTUNITY dataset was collected in a sensor-rich environment, which includes 17 complex gestures and modes of locomotion. Overall, it contains recordings of four subjects who perform morning activities in daily life scenes. Different modalities of sensors had been integrated into the environment, objects, and the body. In terms of the sensor setting, the OPPORTUNITY challenge guidelines were adopted. We only considered the sensors on the body, including 5 inertial measurement units on the sports jacket, 2 InertiaCube3 sensors on the feet, and 12 Bluetooth 3-axis acceleration sensors. During the recording, five activities of daily living (ADL) sessions and one drill session was conducted for each subject. Each sensor axis is considered as a separate channel, resulting in an input space of 113 channels in size. Specifically, these sensors have a sampling rate of 30 Hz. In this paper, we focused only on the recognition of sporadic gestures. Thus, this is an 18-class (including the Null class) segmentation and classification problem. The gestures included in this dataset were summarized in Table 4.



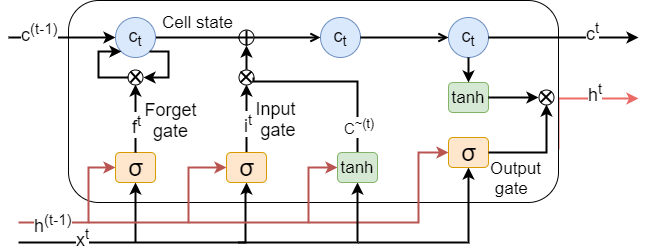


* Feature list

'-XYZ' is used to denote 3-axial signals in the X, Y, and Z directions.

|  |  |
| --- | --- |
| Feature name | Type |
| tBodyAcc-XYZ | Numerical |
| tGravityAcc-XYZ | Numerical |
| tBodyAccJerk-XYZ | Numerical |
| tBodyGyro-XYZ | Numerical |
| tBodyGyroJerk-XYZ | Numerical |
| tBodyAccMag | Numerical |
| tGravityAccMag | Numerical |
| tBodyAccJerkMag | Numerical |
| tBodyGyroMag | Numerical |
| tBodyGyroJerkMag | Numerical |
| fBodyAcc-XYZ | Numerical |
| fBodyAccJerk-XYZ | Numerical |
| fBodyGyro-XYZ | Numerical |
| fBodyAccMag | Numerical |
| fBodyAccJerkMag | Numerical |
| fBodyGyroMag | Numerical |
| fBodyGyroJerkMag | Numerical |

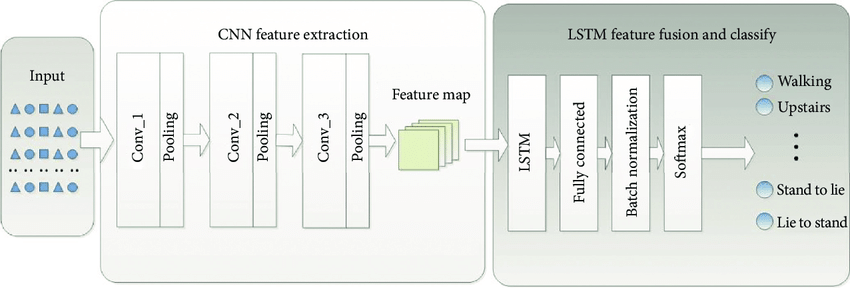
1. Architecture



Long Short Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does not give an efficient performance. LSTM can by default retain the information for a long time. It is used for processing, predicting, and classifying based on time-series data.

1. **Forget Gate:** The information that is no longer useful in the cell state is removed with the forget gate. Two inputs *x\_t* (input at the particular time) and *h\_t-1* (previous cell output) are fed to the gate and multiplied with weight matrices followed by the addition of bias. The resultant is passed through an activation function which gives a binary output. If for a particular cell state the output is 0, the piece of information is forgotten and for output 1, the information is retained for future use.
2. **Input gate:** The addition of useful information to the cell state is done by the input gate. First, the information is regulated using the sigmoid function and filters the values to be remembered similar to the forget gate using inputs *h\_t-1* and *x\_t*. Then, a vector is created using *tanh* function that gives an output from -1 to +1, which contains all the possible values from h\_t-1 and *x\_t*. At last, the values of the vector and the regulated values are multiplied to obtain the useful information
3. **Output gate:** The task of extracting useful information from the current cell state to be presented as output is done by the output gate. First, a vector is generated by applying tanh function to the cell. Then, the information is regulated using the sigmoid function and filter by the values to be remembered using inputs *h\_t-1* and *x\_t*. At last, the values of the vector and the regulated values are multiplied to be sent as an output and input to the next cell.

1. Proposed pipeline



1. **References**

<https://www.niser.ac.in/~smishra/teach/cs460/2021/project/21cs660_group22/>

<https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/?ref=lbp>