International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Human Activity Recognition based on Phone Sensors

Kowshik S[1], Anurag A[2], Nikhil Kumar Reddy Obili[3]

[1-3]*School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India*

Abstract: *Modal Sensor data like Light, Pressure and Accelerometer are an important factor in day-today context aware services in smartphone apps. Our model aims to recognize common modes of locomotion and transportation from data acquired by the sensors in a smartphone. We propose a convolutional neural network-based pipelining. The activity recognition is done on the Sussex-Huawei Locomotion- Transportation (SHL) Recognition Challenge Dataset. The model shows promising results of recognizing activities Still, Walk, Run, Bike, Car, Bus, Train and Subway.*

Index terms: Convolutional Neural Network, Event classification, Transportation Mode Recognition.

Introduction:

The mode of transportation of most smartphone users plays a key role in most contextual information that uses the users mobility during travel. Using multimodal data gives the applications a smarter, lighter and more intelligent for service adaptation focusing on human activity monitoring. In recent years, many studies have recognized modes of transport from motion and GPS from embedded sensors in smartphones. Most of such studies employ the use of classical Machine Learning and Deep Learning models for the inference of transportation mode.

The state of the art in motion-based transportation recognition performance was established in the SHL recognition challenge 2018. The outcomes reveal that approaches based on motion sensors struggle distinguishing between distinct transportation modes of similar classes.

Therefore, it is fairly hard to compare such studies based on the choice of modalities being used. Light and Pressure are a few key modalities being used in generations of smartphones. The advantages include less attenuation in signals in different kinds of environments and fairly low power consumption resulting low battery usage and cheaper implementation. The main challenge in this field are the factors to distinguish similar transport modes such as rail and subway which may have same settings for light and pressure.

In this paper, we aim to answer the question: Can light and pressure be used to recognize transportation? We present different pipelines to recognize eight modes of transportation activities (Still, Walk, Run, Bike, Car, Bus, Train, Subway) from sensors of smartphones and evaluate the performance.

Related work:

A multi- view or multiple knowledge approach where the base layer consists of 9 machine learning models (Fully connected DNN, Random Forest, Gradient Boosting, Extreme Gradient Boosting, SVM, Ad Boosting, KNN, Gaussian Naïve Byes and Decision Tree). The Sensor data frequency is downsized to 50Hz. From the 30 features of 50Hz frequency, 1696 features were calculated belonging to time domain, frequency domain and features for sequential analysis. A DNN of 2 fully connected dense layers with 256 and 128 neurons was found to be optimum. After tuning the hyper parameters of the ML models using randomized 2-fold parameter search, the extracted features were fed to the base layer. The output probabilities of these ML models were fed to the Meta layer consisted of seven ML algorithms (Random Forest, Gradient Boosting, SVM, ad Boosting, KNN, Gaussian Naïve Byes and Decision Tree). The first quarter of the data was used to train and the second quarter was used for evaluation. Here again the randomized 2-fold parameter search was used for hyper parameter tuning. The stochastic HMM finds and corrects errors in the classified sequence of labels by improving the accuracy of the final output by mapping similar patterns in data to similar labels. This can be done once the entire sequence of output label has been predicted or when we need to iteratively predict the final element to the predict future ones.

Based on previous research the data is divided into one second window of non-overlapping 100 samples and Features in both time and frequency domain features are extracted. Direction independent magnitude values of accelerometer, gyroscope, magnetometer, linear accelerometer, and gravity are extracted from time series data extending the time series features to 25 and further common measures of statistical dispersion like Max, Min, Mean, Standard deviation, Interquartile, range, Kurtosis, Variance are found for each for a window size of 100 samples giving a total of 175 features for time domain analysis. The direction independent composite data is used for frequency analysis. Initially noise removal is done with the help of a bandpass filter and then Fast Fourier

transform of 1 second time interval is taken to give 50 points out of which the real valued frequency domain data is kept. Further noise removal is done using median and Gaussian filters respectively. Highest magnitude of the FFT window and corresponding frequency, spectral entropy of the FFT window, total energy of the FFT window, sub-band energy, band energy ratio and spectral entropy, cepstral coefficients and their 1st-order differences were the frequency domain features found to give a high accuracy. The adopted multistage model uses the random forest algorithm to differentiate between static and dynamic activity. At the second stage the static activities are classified as still, car, bus, or rapid-transit activity while the dynamic activities are classified as walk, run, or bike activity. The third stage under the rapid-transit activity is for differentiating between train and subway activity. The final output is based on voting. The window time of majority vote does not remain the same and is greedily determined by the ratio of dynamic activity and static activity in the sensor data for one minute. Higher dynamic to static activity ratio indicates a larger time window to be used (1 minute) else a shorter (20 sec) time window is used. A time independent 5- fold cross validation was used for the train-test split for evaluation. It was found that out of Random Forest, SVM and Neural networks, random forest gave the highest accuracy when measured using the F-1 metric (F-1 score 92%). When this algorithm is used for a multistage pipeline model with voting, the final outcome had an accuracy of 97% for recognizing 8-types of activities.

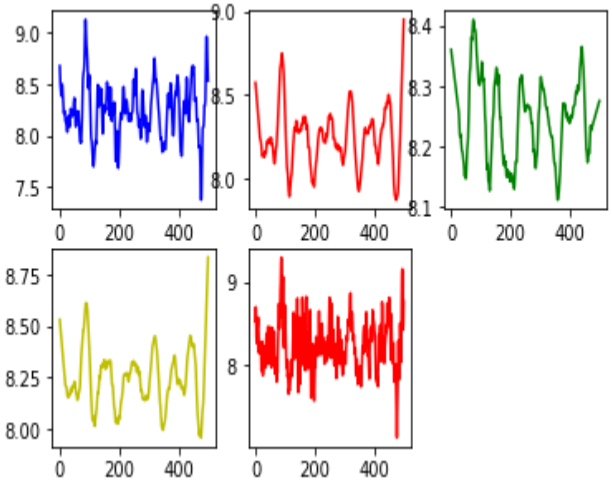

Fig: Frequency graphs of modals.

Similar to the [1], the sample size is downsized 50Hz for faster computational purposes. Three types of Virtual sensor streams are calculated, magnitude of sensors with three axes, de rotated sensors to avoid over fitting specific orientations and Euler angles as the individual directional components might influence the result. Feature extraction is done for each one-minute interval as it was found to be optimal and the label for each window is the most frequent label in that window. Measures of statistical distribution are found from the time domain namely minimum, maximum, autocorrelation, number of samples above/below the mean and the average difference between two sequential data samples.

The features of the frequency domain have been found using the Power spectral density (PSD). Three largest magnitudes, normalized energy, normalized entropy for differentiating between signals with similar energy, binned distribution and Skewness and kurtosis were calculated. A relatively large number of features have been extracted with the help of fresh library and hence the highly correlated ones are removed keeping the Pearson coefficient as a measure. The greedy "wrapper" algorithm was implemented using the random forest classifier to remove the redundant data and make the computation more feasible. It is noticed that the de-rotated magnetometer had the most significant features, followed by those from accelerometer and gyroscope. It was found that the Extreme Gradient Boosting model (XGB) was the best performing. The tree model with the tree booster parameter with tree booster was chosen and the hyperparameters were tuned iteratively. First tree specific boosting parameters and regularization parameters are adjusted with a high learning rate and then the learning rate is brought down to an optimal level. The final outcome when measured with the F-score metric, the XGB optimized model had a 90.2% accuracy and 93.7% accuracy when the train-test split was 90% and 10% respectively.

Literature Survey:

Authors and Year (Reference)	Title (Study)	Concept/ Theoretical model/ Framework	Methodology used/ Implementation	Dataset details/ Analysis	Relevant Finding	Limitations/ Future Research/ Gaps identified
Zhenghua Chen, Chaoyang Jiang_, and Lihua Xie Fellow, IEEE	A Novel Ensemble ELM for Human Activity Recognition Using Smartphone Sensors	a novel ensemble ELM approach using smartphone sensors with GRP for the initialization of input weights of base ELMs.	novel ensemble ELM with the assist of Gaussian random projection (GRP) for the initialization of input weights of base ELMs. Creating data diversity to boost perfomance	dataset includes three-dimensional acceleration and gyroscope of frequency 50Hz collected using a Samsung Galaxy SII smartphone. Dataset using a Huawei 20 Pro smartphone.	Proposed Ensemble ELM has 97:35% accuracy on the first dataset and 98.88% accuracy with the new dataset.	Combining various sensors to identify specific activities. Developing a semi supervised algorithm for conveniently working with unlabeled dataset similar to Active learning
Francisco Javier Ordóñez * and Daniel Roggen	Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition	DeepConvLSTM: a deep learning framework composed of convolutional and LSTM recurrent layers		open dataset – Skoda-activities of assembly-line workers in a car production environment of sample rate of 98hz OPPORTUNITY dataset for activity recognition of sample rate 30Hz. It is divided into activities belonging to Task A (modes of locomotion) and Task B (gesture recognition).		Transfer learning approach based on these models to perform Activity recognition on large-scale data. with pattern recognition.
Murad, A., & Pyun, J. Y. (2017). Deep recurrent neural networks for human activity recognition. *Sensors*, *17*(11), 2556.	Deep Recurrent Neural Networks for Human Activity Recognition	Cascaded Bidirectional and Unidirectional LSTM-based DRNN Model	The first layer is bidirectional, whereas the upper layers are unidirectional. The number of hidden unidirectional layers is a hyperparameter that is tuned during training	UCI-HAD USC-HAD Opportunity Daphnet FOG - Dataset containing movement data from patients with Parkinson's disease (PD) who suffer from freezing of gait (FOG) symptoms. Skoda	Model is able to extract more discriminative features and capture the temporal dependencies between input samples in activity sequences by exploiting DRNN functionality.	Experimentation on large scale complex human activities, as well as exploring transfer learning between diverse datasets. And also Investigating resource efficient implementation of a DRNN for low-power devices
Charissa Ann Ronao, Sung-Bae Cho	Human activity recognition with smartphone sensors using deep learning neural networks	Deep convolutional neural networks	Using a multi-layer convnet with alter-nating convolution and pooling layers, features are automatically extracted from raw time-series sensor data with lower layers extracting more basic features and higher layers deriving more complex ones and classifying them accordingly		Convents make use of local dependency of time-series 1D signals, and the translation invariance and hierarchical characteristics of activities. They also	experimenting with a combination of convnet and SVM, incorporating frequency convolution together with time convolution, using a different error function, or including cross-channel pooling in place of normal max-pooling.

Methods:

Random forest classification is a gathering learning strategy for grouping, relapse and different assignments that works by developing a huge number of choice trees at preparing time and yielding the class that is the method of the classes or mean forecast of the individual trees.

K-nearest-neighbors classification is a non-parametric strategy used for gathering and relapse. In the two cases, the data contains the k closest planning models in the segment space. The yield depends upon whether k-NN is used for gathering or backslide: In k-NN gathering, the yield is a class investment. A thing is portrayed by a larger part vote of its neighbors, with the article being allotted to the class most essential among its k nearest neighbors (k is a positive number, routinely little). If k = 1, by then the article is only allotted to the class of that single nearest neighbor. In k-NN backslide, the yield is the property estimation for the thing. This value is the ordinary of the estimations of k nearest neighbors.

In neural frameworks, *Convolutional neural networks* (ConvNets or CNNs) is one of the principal orders to do pictures affirmation, pictures game plans. Articles area, affirmation faces, etc., are a part of the districts where CNNs are by and large used. Technically, significant learning CNN models to get ready and test, every data picture will go it through a movement of convolution layers with channels (Kernels), Pooling, totally related layers (FC) and apply SoftMax ability to arrange a thing with probabilistic characteristics some place in the scope of 0 and 1. Convolution of a picture with various channels can perform activities, for example, edge recognition, obscure and hone by applying channels.

Long-Short-Term-Memory systems – generally just called "LSTMs" – are an uncommon sort of RNN, equipped for adapting long haul dependencies. LSTMs are unequivocally intended to dodge the long haul reliance issue. Recollecting data for extensive stretches of time is for all intents and purposes their default conduct, not something they battle to learn. All intermittent neural systems have the type of a chain of rehashing modules of neural system. In standard RNNs, this rehashing module will have a straightforward structure, for example, a solitary tanh layer. LSTMs likewise have this chain like structure, however the rehashing module has an alternate structure. Rather than having a solitary neural system layer, there are four, interfacing in an exceptionally unique way.

Methods used and Results:

Algorithms	Training	Validation Accuracy
Simple feature extraction and using random forest for classification	86.3	-
NN Classification	93	68
CNN 1D (Overfitting)	97	72
LSTM	85	78
CNN+LSTM+1D	89	73

Conclusion:

Accurate prediction of the activity is a tedious task when data is superfluous. A global accuracy of 94% is attained. The results sometimes overfit for larger parameters in naïve algorithms.

References:

[1] Chen, Z., Jiang, C., & Xie, L. (2018). A Novel Ensemble ELM for Human Activity Recognition Using Smartphone Sensors. IEEE Transactions on Industrial Informatics, 15(5), 2691-2699.

[2] Morales, F. J. O., & Roggen, D. (2016, September). Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations. In Proceedings of the 2016 ACM International Symposium on Wearable Computers (pp. 92-99). ACM.

[3] Murad, A., & Pyun, J. Y. (2017). Deep recurrent neural networks for human activity recognition. Sensors, 17(11), 2556.

[4] Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert systems with applications, 59, 235-244.