Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 155 (2019) 698–703

Procedia Computer Science

www.elsevier.com/locate/procedia

2nd International Workshop on Recent advances on Internet of Things: Technology and Application Approaches (IoT-T&A 2019)
August 19-21, 2019, Halifax, Canada

Human Activity Recognition: A Survey

Charmi Jobanputra[a] , Jatna Bavishi[b], Nishant Doshi[c*]

[a,b,c]*Pandit Deendayal Petroleum University, Raisan,Gandhinagar – 382007, India*

Abstract

Human Activity Recognition (HAR) has been a challenging problem yet it needs to be solved. It will mainly be used for eldercare and healthcare as an assistive technology when ensemble with other technologies like Internet of Things(IoT). HAR can be done with the help of sensors, smartphones or images. In this paper, we present various state-of-the-art methods and describe each of them by literature survey. Different datasets are used for each of the methods wherein the data are collected by different means such as sensors, images, accelerometer, gyroscopes, etc. and the placement of these devices at various locations. The results obtained by each technique and the type of dataset are then compared. Machine learning techniques like decision trees, K-nearest neighbours, support vector machines, hidden markov models are reviewed for HAR and later the survey for deep neural network techniques like artificial neural networks, convolutional neural networks and recurrent neural networks is also presented.

Keywords: Human Activity Recognition, Machine learning, Neural networks.

1. Introduction

The field of Human Activity Recognition (HAR) has become one of the trendiest research topics due to availability of sensors and accelerometers, low cost and less power consumption, live streaming of data and advancement in computer vision, machine learning, artificial intelligence and IoT.

In HAR, various human activities such as walking, running, sitting, sleeping, standing, showering, cooking, driving, opening the door, abnormal activities, etc. are recognized. The data can be collected from wearable sensors or accelerometer or through video frames or images. HAR can be extensively used in medical diagnosis. For keeping track of elderly people, HAR can be used. Crime rates can be controlled using HAR by monitoring. The smart home environment can be created by the daily activity recognition. Driving activities can be recognized and lead to safe travel. Military actions can be recognized using HAR.

The paper is divided into various state-of-the-art methods for human activity recognition and the challenges for activity recognition. Section 2 describes various state-of-the-art methods. Section 2.1 describes the review and comparison of machine learning methods for HAR such as decision trees, K-nearest neighbours(KNN), support vector machines(SVM) and hidden Markov model(HMM). Section 2.2 describes neural network models such as artificial neural networks(ANN), convolutional neural networks(CNN) and recurrent neural networks(RNN). Fig. 1.(a) summarizes the techniques for HAR. At last, in section 3, the open issues and challenges for activity recognition are described. Finally, section 4 presents the conclusion.

*Corresponding author. Tel.:+917923275458.
Email address: Nishant.Doshi@sot.pdpu.ac.in

1877-0509 © 2019 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2019.08.100

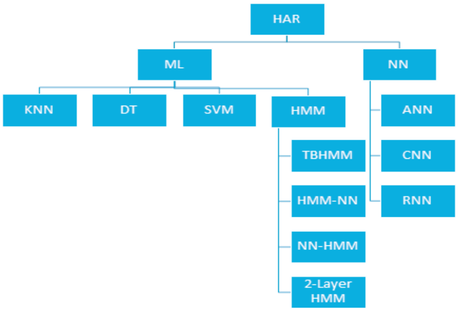

Fig. 1. (a) HAR state-of-the-art methods.

Fig 1. (a) Ref- ML: Machine Learning, NN: Neural Networks, KNN: K-Nearest Neighbour, DT: decision Tree, SVM: support Vector Machine, HMM: Hidden Markov Model, TBHMM: Threshold Based HMM, HMM-NN: HMM-Neural Networks, NN-HMM: Neural Networks-HMM, ANN: Artificial Neural Network, CNN: Convolutional NN, RNN: Recurrent NN.

2. State-of-the-art methods

2.1 Machine Learning Techniques:

Table 1. Summary of Decision Trees, K-Nearest Neighbours, and Support Vector Machine techniques

<table>
<tr><th rowspan="2">Reference + Method</th><th rowspan="2">Dataset</th><th rowspan="2">Model/Variation</th><th rowspan="2">Best Features</th><th colspan="6">Accuracy</th></tr>
<tr><th colspan="3">For test on sample set in %</th><th colspan="3">For 10 fold cross-validation in %</th></tr>
<tr><td rowspan="3">[4] + Iterative Dichotomiser 3 Decision Tree (ID3 DT)</td><td rowspan="3">Through accelerometer of smartphones</td><td>Vector(activity, position)</td><td>Mean</td><td colspan="3">73.72</td><td colspan="3">51.82</td></tr>
<tr><td>Position (decision tree for classification of the position of smartphone)</td><td>Fourier transforms coefficients.</td><td colspan="3">79.56</td><td colspan="3">61.31</td></tr>
<tr><td>Action(classify activity-position independent)</td><td>Maximum value</td><td colspan="3">88.32</td><td colspan="3">80.29</td></tr>
<tr><td rowspan="2">[3] + K-Nearest Neighbour (KNN)</td><td rowspan="2">online activity recognition system working on Android platforms</td><td>K-nearest neighbour(KNN) classification algorithm</td><td>-</td><td colspan="6">-</td></tr>
<tr><td>Clustered K-nearest neighbour</td><td>-</td><td colspan="6">92</td></tr>
<tr><td rowspan="4">[9] + Support Vector Machine (SVM)</td><td rowspan="4">Weizmann
UIUC1</td><td rowspan="4">-</td><td rowspan="4">-</td><td colspan="3">Discriminative Task</td><td colspan="3">Few Examples</td></tr>
<tr><td>L1AO</td><td>L1AAO</td><td>L1SO</td><td>FE-2</td><td>FE-4</td><td>FE-8</td></tr>
<tr><td>100</td><td>100</td><td>100</td><td>66.67</td><td>70.24</td><td>100</td></tr>
<tr><td>99.04</td><td>98.04</td><td>98.84</td><td>45.56</td><td>80.65</td><td>97.44</td></tr>
</table>

2.1.4 Hidden Markov Model

In [1], the author shows the HMM-based approach that uses threshold and voting to automatically segment and recognize complex activities.

Table 2. Summary Hidden Markov Model

<table>
<tr><td>Parameters</td><td>[1]</td><td>[2]</td></tr>
<tr><td>Dataset</td><td>Bulling et al.
Chen et al.
Anguita et al</td><td>Van Kasteren et al.
Tapia et al.</td></tr>
<tr><td>Activities</td><td>
<table>
<tr><td>Name</td><td>total of true labels (TT)</td><td>total of inferred labels(TI)</td></tr>
<tr><td>Walking</td><td>89.11</td><td>99.10</td></tr>
<tr><td>W.upstairs</td><td>93.21</td><td>97.34</td></tr>
<tr><td>W.dowstrais</td><td>97.62</td><td>81.19</td></tr>
<tr><td>Sitting</td><td>96.13</td><td>82.37</td></tr>
<tr><td>Standing</td><td>90.23</td><td>98.56</td></tr>
<tr><td>Lying down</td><td>86.59</td><td>95.88</td></tr>
</table>
<table>
<tr><td>Experiment</td><td>Accuracy</td></tr>
<tr><td>No filter</td><td>80.24</td></tr>
<tr><td>Filter</td><td>79.53</td></tr>
<tr><td>Continuous</td><td>81.21</td></tr>
<tr><td>Continuous and discrete</td><td>88.75</td></tr>
</table>
</td><td>
<table>
<tr><td rowspan="2">Name</td><td colspan="2">House-A</td><td colspan="2">House-B</td><td colspan="2">House-C</td></tr>
<tr><td>TT</td><td>TI</td><td>TT</td><td>TI</td><td>TT</td><td>TI</td></tr>
<tr><td>Breakfast</td><td>75.0</td><td>76.61</td><td>75.0</td><td>73.24</td><td>-</td><td>-</td></tr>
<tr><td>Brushing teeth</td><td>75.5</td><td>79.80</td><td>75.5</td><td>79.80</td><td>72</td><td>79.12</td></tr>
<tr><td>Dinner</td><td>78.0</td><td>79.75</td><td>78.0</td><td>75.00</td><td>-</td><td>-</td></tr>
<tr><td>Drinking</td><td>78.0</td><td>76.39</td><td>78.0</td><td>75.36</td><td>80</td><td>77.67</td></tr>
<tr><td>Leaving house</td><td>85.0</td><td>93.61</td><td>85.0</td><td>94.65</td><td>85</td><td>96.59</td></tr>
<tr><td>Others</td><td>50.0</td><td>33.99</td><td>60.0</td><td>36.23</td><td>50</td><td>26.67</td></tr>
<tr><td>Sleeping</td><td>80.5</td><td>93.38</td><td>80.5</td><td>84.55</td><td>80.5</td><td>75.09</td></tr>
<tr><td>Showering</td><td>77.0</td><td>74.90</td><td>77.0</td><td>79.96</td><td>76</td><td>78.75</td></tr>
<tr><td>Snack</td><td>72.0</td><td>86.64</td><td>-</td><td>-</td><td>65</td><td>74.71</td></tr>
<tr><td>Toileting</td><td>77.0</td><td>77.00</td><td>77.0</td><td>79.71</td><td>77</td><td>78.97</td></tr>
<tr><td>Dressing</td><td>-</td><td>-</td><td>80.0</td><td>89.38</td><td>75</td><td>90.36</td></tr>
<tr><td>Preparing BF</td><td>-</td><td>-</td><td>65.0</td><td>76.02</td><td>65</td><td>70.65</td></tr>
<tr><td>Preparing DN</td><td>-</td><td>-</td><td>66.0</td><td>76.74</td><td>65</td><td>70.65</td></tr>
<tr><td>Using dishwasher</td><td>-</td><td>-</td><td>76.0</td><td>83.51</td><td>-</td><td>-</td></tr>
<tr><td>Eating</td><td>-</td><td>-</td><td>-</td><td>-</td><td>75</td><td>72.81</td></tr>
</table>
</td></tr>
<tr><td>Recognition Models</td><td>Continuous Hidden Markov Model (cHMM)
Multiple Class Support Vector Machine (MC-SVM)</td><td>Conditional Random Field (CRF)
Hidden Markov Model (HMM)
2- Layer HMM
Naïve Bayes</td></tr>
<tr><td>Preprocessing</td><td>A Median filter and a 3rd order lowpass Butterworth filter with 20Hz frequency.</td><td>They Clearly not represented any preprocessing required on binary temporal data. But they mentioned about three different feature representations:
1. Raw. This feature uses the sensor data directly as it was collected from the sensor network. The value is 1 when the sensor fires and 0 otherwise.
2. Change Point (CP). This feature indicates when a sensor changes value. The value is 1 when a sensor state goes from zero to one or vice versa and 0 otherwise.
3. Last-Fired (LF). This feature indicates which sensor fired last. The sensor that changed state last continues to value 1 and changes to 0 when another sensor changes state.</td></tr>
<tr><td>Inputs:</td><td>
<table>
<tr><td>To Record data:</td><td>InvenSense MotionFitTM Kit</td></tr>
<tr><td>Sensor:</td><td>MPU-9150</td></tr>
<tr><td>Place of the sensors :</td><td>the left hand wrist</td></tr>
<tr><td>Daily activity mostly executed in:</td><td>59 m^2 flat</td></tr>
<tr><td>Recording sampling frequency:</td><td>50Hz</td></tr>
<tr><td>Activities are saved with:</td><td>3-axis accelerometer, 3-axis gyrometer data</td></tr>
</table>
</td><td>
<table>
<tr><td>Data saved in form of</td><td colspan="2">Binary temporal data</td></tr>
<tr><td rowspan="5">Sensors</td><td>Reed switch :</td><td>To measure doors and cupboards are open or closed.</td></tr>
<tr><td>Pressure mats:</td><td>To measure sitting on or lying in bed</td></tr>
<tr><td>Mercury contacts:</td><td>To detect the movement of objects</td></tr>
<tr><td>passive infrared (PIR):</td><td>to detect motion in a specific area,</td></tr>
<tr><td>float sensors</td><td>To measure the toilet being flushed</td></tr>
</table>
</td></tr>
</table>

2.2 Neural Network Techniques

2.2.1 Artificial Neural Network

In [8], the author describes a data acquisition module prototype developed by them, which gathers the data of the patient and recognizes abnormal status of the patient's health so that early treatment would be available.

For arm posture recognition,

- Input device: Accelerometer embedded in smart watch.
- Preprocessing: Filtering, normalization, feature extraction.

For body posture recognition.

- Input device: On chest.

A new dataset with different sets of accelerometer data and data from heart rate sensor was used to identify various activities in [8].

2.2.2 Convolutional Neural Network

In [6], the author describes human activity recognition through a very robust deep neural network technique that is convolutional neural network which can model the features effectively.

- Input device: Inertial Measurement Unit sensors and triaxial sensors.
- Placement of sensors: Two sensors each on left and right shank, two sensors centred on feet and one on lumbar region was placed.
- Sensor configuration: Single device, double device and triple device setups which used individual sensor data, in combination of two sensor data and adding third sensor data in combination of two respectively.
- Shape of window: (6(no. of sensors) $\times$ 204).

When observed, the combination of two or three sensors gave better results.

2.2.3 Recurrent Neural Network

Recurrent neural networks(RNN) recognizes the patterns which are separated by some intervals. Long Short Term Memory (LSTM) is a RNN architecture which models temporal sequences and has the capacity of memorizing the things. In [7], the author describes HAR using LSTM.

- Input device: Sensors embedded in houses at various locations.
- Data representation:
  1) Raw sensor data where the data from the sensor is directly used
  2) Last-fired sensor data which are the data received from the sensor that was fired last.

Configuration of LSTM was done as below:

Table 3.Summary of Above Methods

Reference + Dataset used + Implementation method	Activities classified	Network topology		Accuracy
		Parameters	Description	
[8] + Self - made dataset + ANN	Arm downwards Arm upwards Arm horizontal forward Arm horizontal backward Arm horizontal forward rotated upwards Arm horizontal forward rotated downwards	Network	2 layer Feedforward-Backpropagation	100%
		Activation function	Sigmoid	
		Input layer	3 neurons	
		Hidden layer	10 neurons	
		Output layer	6 neurons	
		Training algorithm	Levenberg-Marquardt	
		Performance evaluation	Mean squared error function	
[8] + Self - made dataset + ANN	Sitting Prone Left lateral recumbent Right lateral recumbent Supine	Network	2 layer Feedforward-Backpropagation	99.96%
		Activation function	Sigmoid	
		Input layer	3 neurons	
		Hidden layer	10 neurons	
		Output layer	5 neurons	
		Training algorithm	Levenberg-Marquardt	
		Performance evaluation	Mean squared error function	
[8] + Different dataset + ANN	Standing Supine Left lateral recumbent Right lateral recumbent Prone Walking (forward) Walking (backward) Running (forward) Running(backward)	Network	2 layer Feedforward-Backpropagation	99.08%
		Activation function	Sigmoid	
		Input layer	3 neurons	
		Hidden layer	10 neurons	
		Output layer	10 neurons	
		Training algorithm	Levenberg-Marquardt	
		Performance evaluation	Mean squared error function	
[6] + Otago Exercise dataset + CNN		Performance evaluation	5-fold cross validation	
		Three convolution layers with kernels of size	3 × 5, 2 × 4 and 2 × 2	
		Three max pooling layer with kernel size	3 × 3, 2 × 2 and 3 × 2	
		Three dense layers	500, 250 and 125 units	
		Activation function	ReLU function	
		Loss calculation	Cross entropy function.	
		Stochastic optimization method	Adam optimizer	
		Output layer Stochastic optimization method	m units, m: no. of activities in each group Adam optimizer	
		Batch size	1024	

3. Research gap and further challenges

- **During data collection:** If the data is to be collected through sensors, multiple sensors are to be worn by the person and the placement of sensors is an issue as it affects the results.
- **Feature extraction:** Extraction of principal features from sensor data is challenging.
- **Multiple persons:** If sensors are embedded in home environment, there can be multiple residents there and so to map the activities of multiple residents is a difficult task.
- **Time complexity and accuracy:** Different classification techniques give different time complexity and accuracy. It is frequently observed that if computational complexity of any classification model is less, it has somewhat poor accuracy as compared to the models where the accuracy is too good but computational complexity is less acceptable.
- **Real-time data:** Many results were calculated on the standard datasets which might vary when real-time dataset is used.
- **Multiple activities:** If the person performs more than one activity at the same time, recognition is difficult.
- **Vision based activity recognition:** With the live streaming of data and presence of crowd around, activity recognition may be difficult.
- **Location based activity recognition:** Outdoor locations can be traced through Global Positioning System (GPS) but indoor location is difficult to trace without embedding the sensors inside which creates multiple persons problem.
- **Sensor constraints:** We do not know if the sensor data is incorrect when faulty sensors are used.
- **Overfitting and underfitting:** Classification models like decision trees, neural networks can cause overfitting and SVM can cause underfitting when less training data is available. So, the method of implementation must be in accordance with the data.

4. Conclusion

In this survey, we carried out the comprehensive study of various tools and techniques which can be used in human activity recognition which included different machine learning algorithms and neural network techniques. The techniques were implemented on different datasets and they had varying observations depending upon the environmental conditions, type of data used such as accelerometer data, other sensor data, placement of sensors, methods of implementation. These techniques are compared on the basis of those contexts and also on the basis of computational complexities. Finally, challenges to human activity recognition are also presented. From this survey, we deduce that there is no single method which is best for recognition of any activity, hence in order to select a particular method for the desired application, one needs to take various factors into consideration and determine the approach accordingly. So, in spite of having numerous methods, some of the challenges still remain open and have to be resolved.

References

[1] Sarah Fallmann and Johannes Kropf. (2016) "Human activity recognition of continuous data using Hidden Markov Models and the aspect of including discrete data." *Intl IEEE Conferences* 121-126.

[2] M. Humayun Kabir, M. Robiul Hoque, Keshav Thapa, and Sung-Hyun Yang. (2016) "Two-Layer Hidden Markov Model for Human Activity Recognition in Home Environments." *International Journal of Distributed Sensor Networks*, Volume: 12 issue: 1

[3] Pinky Paul and Thomas George (2015) "An Effective Approach for Human Activity Recognition on Smartphone" I*EEE International Conference on Engineering and Technology*, pp 1-3.

[4] Lin Fan, and Zhongmin Wang. (2014)" Human activity recognition model based on decision tree." *International Conference on Advance Cloud and Big Data*, pp. 64-68.

[5] MS. Kanchan Gaikwad.(2012) " HMM Classifier for Human Activity Recognition ." *International Journal (CSEIJ)*, Vol.2,No.4

[6] Antonio Bevilacqua, Kyle MacDonald, Aamina Rangarej, Venessa Widjaya, Brian Caulfield, and Tahar Kechadi." Human Activity Recognition with Convolutional Neural Networks"

[7] Deepika Singh, Erinc Merdivan, Ismini Psychoula, Johannes Kropf, Sten Hanke, Matthieu Geist, and Andreas Holzinger.(2018) " Human Activity Recognition using Recurrent Neural Networks." ,267-274.

[8] Stefan Oniga, and József SütĘ. (2014)" Human activity recognition using neural networks." *15th International Carpathian Control Conference (ICCC)*

[9] K. G. Manosha Chathuramali and Ranga Rodrigo (2012) "Faster Human Activity Recognition with SVM" *The International Conference on Advances in ICT for Emerging Regions*, 197-203.