pHealth 2020
B. Blobel et al. (Eds.)

doi:10.3233/SHTI200631

A Machine Learning Approach for Human Activity Recognition

Angelos PAPOUTSIS [1], Giannis BOTILIAS, Petros KARVELIS, Chrysostomos STYLIOS
Department of Informatics and Telecommunications
University of Ioannina
Kostakioi, 47150 Arta, Greece

Abstract. Human Activity Recognition (HAR) is becoming a significant issue in modern times and directly impact the field of mobile health. Therefore, it is essential the designing of systems which are capable of recognizing properly the activities conducted by the individuals. In this work, we developed a system using the Internet of Things (IoT) and machine learning technologies in order to monitor and assist individuals in their daily life. We compared the data collected using a mobile application and a wearable device with built-in sensors (accelerometer and gyroscope) with the data of a publicly available dataset. By this way, we were able to validate our results and also investigate the functionality and applicability of the wearable device that we choose for the Human Activity Recognition problem. The classification results for the different types of activities presented using our dataset (99%) outperforms the results from the publicly database (97%).

Keywords. Activity recognition, sensors, machine learning, predictive methods, health

Introduction

Over the last decade, a significant amount of interest exists in analyzing people's interests and daily activities. The analysis of human behavior is the key to a better understanding of human needs, with the most relevant areas are those of well-being and providing assistance in cases of need. Human behavior modelling can be achieved through a process called Human Activity Recognition (HAR). Using sensors that produce various types of signals and machine learning models, it is possible to recognize daily habits such as walking, running, sitting, lifting, climbing/descending stairs, cycling, and more [1].

Throughout the years, human activity recognition has attracted the research community attention. In general, there exists efforts that make use of the machine learning and deep learning algorithms [2]. Besides the different types of algorithms that are used the main difference between them lies in the manner that features get extracted from the raw data. The Machine learning methods heavily rely on a field expert who uses techniques from the time and frequency domain in order to extract heuristic handcrafted features, also known as "shallow" features [3]. On the other hand, deep learning methods can extract features directly from the raw data [2]. An interesting approach was presented

[1] Angelos Papoutsis, Department of Informatics and Telecommunications, University of Ioannina, Kostakioi, 47150 Arta, Greece, E-mail: papoutsis@kic.uoi.gr.

in [4], where the authors developed a semi-supervised approach to facilitate the annotation of labelled data.

Different classification methods, such as Support Vector Machines (SVM) [5], k-Nearest Neighbor (k-NN) [6], and Decision Trees (DT) [7] are used from the machine learning area, in order to identify a broad range of activities [2]. For example, Zebin, Scully, & Ozanyan [8], used SVM in order to recognize human activities using accelerometer and gyroscope time series data collected from several volunteers. They explored different feature extraction methods to speed up the recognition process and reached a high classification accuracy. They presented an accuracy of 96.7% in the classification task. In another effort [9], the authors utilized the KNN using a smartphone-based accelerometer sensor. After the extraction of several features, an accuracy of 97.97% was achieved. The authors stated out that KNN is an exceptional algorithm with high accuracy and low statistical error. Last Dewi, & Chen [10] used a special decision tree algorithm called Random Forest (RF) and compared it with three classifiers such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Linear Discriminant Analysis (LDA) in order to recognize human activities using different features. Random forest outperforms the other classifiers with an accuracy of 98.57%.

Classification methods from the deep learning area such as Convolutional Neural Networks (CNNs) [11] and Long Short-Term Memory (LSTM) [12,13] are commonly using for the identification of human activities [2]. Xu, Yang, & Liu [14] used a CNN model in order to recognize six different activities. The data derived from a smartphone three-axis accelerometer sensor, and the results presented a high accuracy of 91.97%, which outperforms the accuracy of the support vector machine algorithm (82.27%). In another effort [15], the authors used a Long Short-Term Memory model using data from a controlled setting, as well as under field conditions. The results showed that the model is robust enough as it produced a high accuracy (88.6%) and outperform other efforts from the literature.

The goal of our work is to design a system that directly applies to the emerging sector of mobile health and uses the key elements of IoT and machine learning technologies. The primary scope of this system is to monitor and provide assistance for people that may experience problems in their health.

1. Methods

1.1. Learning / Training phase

In order to classify the activities conducted by the users, we used deep Recurrent Neural Networks (RNNs) and especially a subcategory of them called Long-Term Memory Networks (LSTMs). They initially proposed by Hochreiter [12] and improved in 2000 by Gers [13]. In [16], the authors stated that LSTM is suitable for human activity recognition, and they do not require some expert domain knowledge in extracting the features from the raw data. This type of model can process multiple different sequences of input data, like the three axes of the accelerometer and gyroscope sensor.The model consists of two LSTM layers with 64 neurons each, and between them, we make use of the dropout technique in order to reduce the overfitting problem on the training data. We also utilized the L2 regularization with a loss of 0.0015 in order to force the weights of our network to be small and so make our model simpler. Finally, two dense, fully connected layers used in order to interpret the features from the LSTM model and an output layer in order to perform the final predictions.

For performance analysis, we used Scikit-learn (Version 0.22.2), which is a Python-based machine learning toolkit. We utilize the leave one subject out methodology in order to obtain different datasets in the train and test part each time.

Additionally, we used Adam optimizer [17] to optimize the network with a learning rate of 0.0025 and the sparse categorical cross-entropy loss function as we deal with a multi-class classification problem. The neural network trained for 100 epochs. We also used a batch size of 32 samples meaning that 32 windows of data will be exposed to the model before the weights updated.

1.2. Real-World Dataset

After extensive research, we decided to work with an available online dataset called Real-world dataset [18]. The dataset consists of 15 subjects (8 males and 7 females) that perform eight activities (downstairs, upstairs, jumping, lying, standing, sitting, running/jogging, walking). Six different sensors were used during the data collection (accelerometer, GPS, gyroscope, light, magnetic field, and sound). As regards the placement of the sensors, the researchers took into consideration seven different positions on the subject's body (chest, forearm, head, shin, thigh, upper arm, and waist). The activities recorded, using smartphones and a smart-watch. Each subject performed each activity roughly 10 min except for jumping due to the physical exertion (~1.7 min). The data collected with a sampling rate of 50HZ. In our experiment, we did not consider the whole dataset. More specifically, we used five activities (downstairs, upstairs, standing, sitting, walking) performed by six subjects. From the available data, we used the one that were collected from the wrist position with the use of accelerometer and gyroscope sensors. Table 1 depicts the number of activities conducted by one subject (1393 windows of data).

Table 1. Number of Activities for the Real-World Dataset

Activity	Number of activities
Downstairs	226
Sitting	301
Standing	322
Upstairs	226
Walking	318

1.3. Our Dataset

In order to build our dataset, we designed a mobile application for android smartphones in order to collect data from a wearable device via Bluetooth. The data from the wearable device were stored in a data storage environment built with the MongoDB, a NoSQL database management system. A smartphone device acts as a bridge between the wearable device and the database. As for the wearable device after an extensive search, we decided to use the Ambient Lab™ model, called MetaMotionR (MMR), which is capable of capturing motions and environmental sensor data in real-time [19]. The device was adjusted on the wrist of six users. The users performed five different activities like walking, sitting, standing, upstairs, and downstairs while the data collected with a frequency of 50 HZ for 10 min each using the accelerometer and gyroscope built-in sensors. Table 2 depicts the number of activities conducted by one subject (1400 windows of data).

Table 2. Number of Activities for our Dataset

Activity	Number of activities
Downstairs	250
Sitting	268
Standing	296
Upstairs	290
Walking	296

1.4. Classification Metrics

The overall performance of a classification model is measured by a set of metrics that present, in mathematical terms, how reliable the model is in the HAR process [2]. The most common metrics are accuracy, precision, recall, and F-measure [20]. Using the following values, we defined the metrics:

- True Positives, (TP): The number of positive instances that were classified as positive.
- True Negatives, (TN): The number of negative instances that were classified as negative.
- False Positives, (FP): The number of negative instances that were classified as positive.
- False Negatives, (FN): The number of positive instances that were classified as negative.

The accuracy is commonly used to summarize the overall classification performance of all classes:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

The precision is the ratio of correctly classified positive instances to the total number of instances classified as positive:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

The recall is the ratio of correctly classified positive instances to the total number of positive instances:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

The F-measure combines precision and recall in a single value:

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

2. Results

Below we present the results that we obtained by using our dataset. The LSTM model that we used presented a high recognition accuracy on the activities and also a high total accuracy. After the training, we report an accuracy of 99% (Table 3).

Table 3. Classification Results for our Dataset

	Precision (%)	**Recall (%)**	**f1-score (%)**
Downstairs	0.99	0.99	0.99
Sitting	1.00	0.99	1.00
Standing	0.99	1.00	1.00
Upstairs	0.99	0.99	0.99
Walking	1.00	1.00	1.00
Accuracy			**0.99**

We also trained with the same model (LSTM) and the same parameters as before the data from the Real-World dataset. The total accuracy of this model was 97% (Table 4). As a result, the model accuracy with our dataset outperforms the model accuracy with the Real-World dataset. The results also revealed that the initial selection of the MMR devices was an excellent choice as the sensors from the device provides us with high quality data.

Table 4. Classification Results for Real-World Dataset

	Precision (%)	**Recall (%)**	**f1-score (%)**
Downstairs	0.96	0.96	0.96
Sitting	0.99	0.99	0.99
Standing	0.96	0.98	0.97
Upstairs	0.97	0.96	0.97
Walking	0.97	0.97	0.97
Accuracy			**0.97**

3. Conclusions

In this work, long-term memory networks were implemented to classify human activities. The results of the classification methods were analyzed using four performance evaluation: precision, recall, f-measure and accuracy rate. The dataset collected with the MMR sensor compared with a dataset obtained using the Real-World Dataset. The best result obtained from the MMR dataset with a total classification accuracy of 99%. The quality of the IoT system developed heavily influence the high accuracy that we achieved during our experiments. Also, the results showed that LSTM models are highly applicable to the Human Activity Recognition problem.

References

[1] Banos O, Galvez J-M, Damas M, Pomares H, Rojas I. Window Size Impact in Human Activity Recognition. Sensors. 2014Sep;14,4:6474–99.
[2] Lima WS, Souto E, El-Khatib K, Jalali R, Gama J. Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview. Sensors. 2019;19,14:3213.
[3] Wang J, Chen Y, Hao S, Peng X, Hu L. Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters. 2019; 119:3–11.
[4] Bota P, Silva J, Folgado D, Gamboa H. A Semi-Automatic Annotation Approach for Human Activity Recognition. Sensors. 2019;19,3:501.
[5] Support Vector Machines. Dordrecht: Springer; 2008. Svm.
[6] Cover T, Hart P. Nearest neighbor pattern classification. IEEE transactions on information theory 13(1), 21-27. 1967; knn.
[7] Quinlan JR. Induction of decision trees. Machine learning. 19861; 1: 81-106.
[8] Zebin T, Scully PJ, Ozanyan KB. Inertial Sensor Based Modelling of Human Activity Classes: Feature Extraction and Multi-sensor Data Fusion Using Machine Learning Algorithms. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering eHealth 360°. 2016 Jan;306–14.
[9] Mandong AM, Munir UM. Smartphone Based Activity Recognition using K-Nearest Neighbor Algorithm. In Proceedings of the International Conference on Engineering Technologies, Konya, Turkey (pp. 26-28).
[10] Dewi C, Chen R-C. Human Activity Recognition Based on Evolution of Features Selection and Random Forest. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). 2019.
[11] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86,11:2278–324.
[12] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997;9,8:1735–80.
[13] Gers F, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. 9th International Conference on Artificial Neural Networks: ICANN 99. 1999.
[14] Xu W, Pang Y, Yang Y, Liu Y. Human Activity Recognition Based on Convolutional Neural Network. 2018 24th International Conference on Pattern Recognition (ICPR). 2018.
[15] Milenkoski M, Trivodaliev K, Kalajdziski S, Jovanov M, Stojkoska BR. Real time human activity recognition on smartphones using LSTM networks. 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2018.
[16] Ordóñez F, Roggen D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors. 2016;16,1:115.
[17] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.
[18] Sztyler T, Stuckenschmidt H, Petrich W. Position-aware activity recognition with wearable devices. Pervasive and Mobile Computing. 2017; 38:281–95.
[19] Giannis Botilias, Track My Health: An IoT Approach for Data Acquisition and Activity Recognition pHealth 2020, Prague, Czech Republic, 14 – 16 September 2020, (To be published or In press).
[20] Lara OD, Labrador MA. A Survey on Human Activity Recognition using Wearable Sensors. IEEE Communications Surveys & Tutorials. 2013;15,3:1192–209.