Advanced Data Analysis with Python
Final Report

Sadra Safadoust

January 2023

1 Introduction

Recently, the scientific community has become increasingly interested in human activity recognition. Human activity recognition deals with the problem of identifying the activities and emotions of humans using the measurements captured by different kinds of sensors. It has many applications in health care, fitness and preventive care. In this work, we try to recognise human activity using different machine learning methods and compare the results. We use two datasets that have been captured with different sensors and are used for identifying different kinds of human activities.

2 Related Work

There are many works that focus on human activity recognition. These works use different datasets that have been captured by different kinds of sensors. The most commonly used sensors are those that are small and low-cost such as accelerometers and gyroscopes. [9], [8], and [14] use these sensors and [6] and [12] use Magnetometer, another small and low-cost device, in their studies. Environmental sensors are another category of sensors that have been widely used. Among this category, [17] and [10] use temperature sensors, while [4] use humidity sensors. Medical equipment, such as ECG are used in [11] and [18]. Other devices such as GPS and WiFi have also been utilised in works such as [7] and [3] respectively.

Using the above mentioned sensors, several publicly available datasets have been released for human activity recognition. In this section we briefly mention some of them. Opportunity dataset [16] uses body sensors, objects sensors and ambient sensors to detect 18 human activities. PAPAM2 [15] exploits Colibri wireless IMUs for human activities. The number of activities in this dataset is also 18, however they are different from the activities in the opportunity dataset. ActiveMiles [1] uses smartphone accelerometer and gyroscope in uncontrolled environments to detect 7 activities of daily life.

Note that there are many more works or datasets in the literature. For a comprehensive list of such previous works, sensors and dataset please refer to surveys such as [5] and [19] and [13].

3 Methodology

In this work, we use different machine learning methods on two different datasets and evaluate the performance of different methods on these datasets. In this section, we first describe the datasets used:

3.1 Datasets

Wearable Physiological Measurements: This dataset was proposed in [11]. It uses measurements from Electrocardiogram (ECG), Thoracic Electrical Bioimpedance (TEB) and Electrodermal Activity (EDA). Measurements of ECG are influenced by the level of stress, emotions and the activity, while TEB is an indicator of the breathing function. Using EDA, we can measure sweat gland activity on the skin, which is controlled by the sympathetic nervous system, making it a suitable source of data for emotion recognition.

This Wearable Physiological Measurements Dataset, which we will call the "**Wearable Dataset**" for short, measures 4 different activities: 1) neutral activity 2), emotional activity, 3) mental activity, and 4) physical activity. In order to elicit the first two activities, segments from several films were used. Concretely, the "Earth" documentary was used for the neutral activity, while "Life is beautiful" (1997) by Miramax, "American History X" (1998) by Savoy Pictures [62], and "I am legend" (2007) by Warner Bros were used for the emotional activity. The mental activity was elicited by playing Tetris and several mental arithmetic games. In order to elicit physical activity, participants went up and down the stairs for five minutes.

There are 40 participants in this dataset, each of which have 112 instances. Each instance consists of 533 features. 174 features from ECG signal, 151 features from TEB, 104 features from EDA in the arm, and 104 from EDA in the chest. This labels of this dataset is completely balanced and out of 4480 instances, each of the four classes composes exactly 1120 instances. Out of 40 subjects, we randomly select 32 participants for our training split, 4 for validation and 4 for testing split.

UCI HAR Dataset: This dataset has was released in [2] and uses the embedded accelerometer and gyroscope sensors of a smartphone (Samsung Galaxy S II) and captures 3-axial linear acceleration and 3-axial angular velocity.

We will also call this dataset, the "**Smartphone Dataset**", to emphasise on its difference from the previous dataset. It captures 6 different activities: 1) walking, 2) walking upstairs, 3) walking downstairs, 4) sitting, 5) standing, and 6) laying. This dataset consists of 30 participants within an age bracket of 19-48 years. Out of the 30 participants, the data of 9 participants (2947 instances) is used for testing and the data of 21 participants (7352 instances) can be used for training, where each instance consists of 561 attributes. We randomly select 16 of 21 participants (5609 instances) and use as our training set, and use the remaining 5 participants (1743 instances) as the validation set.

3.2 Classification

In this section we talk about the methods that we used on these two datasets. We used several machine learning and deep learning models on in order to classify the data. In all of the methods, we first normalise the data using the mean and standard deviation of the training set. Then for each method, we trained the model only on the train set and performed validation and hyperparameter tuning on the validation set and selected the best model. Then we evaluated the best models of each method on the never-before-seen test set.

We first mention the methods and the different hyperparameter and choices involved. The results are summarised in Table 1.

Ridge Classifier: We used the Ridge classifier with different values of α which controls the regularization strength. We tried values of $\{0, 0.1, 1, 3\}$ for α and found that the values of $\alpha = 0$ and $\alpha = 0.1$ perform the best for the wearable and the smartphone datasets, respectively.

Support Vector Machine Classifier: Here we first used a Support Vector Machine (SVM) with a linear kernel. We tried different values of the regularization parameter C which is equivalent to $\frac{1}{2\alpha}$ in the ridge classifier. We tried values of $\{0.1, 1, 3, 10\}$ and found that $C = 1$ and $C = 0.1$ works best for the wearable and the smartphone datasets, respectively.

We also tried SVMs with different kernels. Concretely, we trained SVMs with poly, rbf, and sigmoid kernels and found that the sigmoid kernel performs the best on the validation set of wearable dataset, while the poly kernel performed the best on the smartphone dataset.

AdaBoost Classifier: AdaBoost is an ensemble classifier. After fitting a classifier on the data, it fits additional copies of the classifier on the same data, focusing on the hard cases.

K-Nearest Neighbour Classifier: For this classifier, we try different values of K. We tried values of $\{1, 3, 5, 7, 9\}$ on both of the datastes. $K = 7$ achieved the best performance for the wearable dataset, while we found that for the smartphone dataset, $K = 9$ achieved the best performance.

Decision Tree Classifier: We tried different criterions for the decision tree classifier. Namely, we tried “gini”, “entropy”, “log loss” and found that ‘log loss” achieves the best accuracy on the wearable dataset, and “gini” achieves the best accuracy on the smartphone dataset.

Method	Wearable	Smartphone
AdaBoost	58.3	53.1
KNN	62.1	87.0
Nonlinear SVM	69.6	90.9
Decision Tree	70.5	79.3
Random Forest	75.0	89.8
MLP	75.7	92.9
Ridge	<u>77.2</u>	<u>93.8</u>
Linear SVM	**79.7**	**94.1**

Table 1: Classification accuracies of differnet methods on the test sets of Wearable and Smartphone datasets. To see the hyperparameters of each method please see Sec 3.2. For each dataset, the best performance is shown in **bold** and the second best method is <u>underlined</u>.

Random Forest Classifier: Random forest, similar to AdaBoost is an ensemble classifier. Several decision tree classifiers are fitted to various subsamples of the training dataset, and averaging is used to improve predictive accuracy and control overfitting.

Neural Network Classifier: As a deep learning model, we trained an MLP classifier. We tried different number of layers with different number of neurons in each layer. Concretely, we tried four different architectures. a) An architecture with a hidden layer of size 1024, b) two hidden layers of size 1024 and 128, c) three hidden layers of size 1024, 512, 128, and d) three hidden layers of size 1024, 1024, 128. For the nonlinear activation functions we tried relu, leaky relu, sigmoid and tanh activation functions. We also tried Stochastic Gradient Descent and Adam optimisers. For each optimiser we tried learning rates of 0.001, 0.0001, and 0.00001. Therefore, we trained 96 models on each dataset and selected the best one according to the performance on the validation set. On the wearable dataset, we found that the network with 3 hidden layers of size 1024, 1024, 128 with tanh activation layer which was trained an adam optimiser with learning rate of 0.001 performed the best. On the smartphone dataset, the network with two hidden layers with leaky relu activation function, trained with adam optimizer using a learning rate of 0.001 performed the best.

4 Results

4.1 Comparision

The results of our experiments are summarised in Table 1. It can be seen that in both datasets, linear SVM performs the best, followed by the ridge classifier and then MLP. In both datasets we see that the ensemble method AdaBoost performs the worst. Nonlinear SVM performs well on the smartphone dataset while it does not perform that good on the wearable dataset. We can also

Sensor	Linear SVM	Ridge
ECG	<u>60.9</u>	<u>59.8</u>
TEB	**71.9**	**71.9**
EDA (arm)	45.3	<u>59.8</u>
EDA (hand)	50.7	50.2
EDA (both)	48.4	57.6

Table 2: Classification accuracies using only the data from a single sensor. For each method, the best performance is shown in **bold** and the second best method is <u>underlined</u>.

see that except for AdaBoost, all of the other methods perform better on the smartphone dataset than on the wearable dataset.

4.2 Single Sensor

Remember that the wearable datasets consists of ECG measurements, TEB measurements, EDA measurements in arm, and EDA measurements in hand. We want to see how much each of these sensors contribute to the activity classification. Therefore, we take the two top performing methods, Linear SVM and Ridge, and train them separately on each of the sensors. The hyperparameters here are the same as the previous one found in Sec 3.2. We show the results in Table 2. We can see that the TEB achieves the best performance among the sensors. We also see that for Linear SVM, using EDA in hand is more accurate than EDA in arm, while for the ridge classifier it is the other way around. It is also interesting that combining the EDAs, makes the performance of the best working EDA worse. Concretely, when using linear SVN, using EDA in hand alone we can achieve 50.7% accuracy, while combining it with EDA in arm reduces it to 48.4%. Also in ridge classifier, we can achieve 59.8% accuracy using EDA in arm alone, whereas if we combine it with EDA in hand, the performance degrades to 57.6%.

5 Conclusions

In this work, we trained several machine learning models on two different datasets for human activity recognition and compared their performances. We saw that the linear SVM and the ridge classifier outperformed the other methods including the deep learning models. Which shows that deep learning models are not the best models for every datasets, especially these kinds of tabular data. We showed that using machine learning models and different kinds of sensors, e.g. wearable smartphone sensors, we are able to accurately predict the activity that the human is involved in. We also evaluated the contribution of each sensor alone in the final activity classification.

References

[1] Active miles. http://hamlyn.doc.ic.ac.uk/activemiles/activemiles.html.

[2] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra Perez, and Jorge Luis Reyes Ortiz. A public domain dataset for human activity recognition using smartphones. In *Proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning*, pages 437–442, 2013.

[3] Sourav Bhattacharya and Nicholas D Lane. From smart to deep: Robust activity recognition on smartwatches using deep learning. In *2016 IEEE International conference on pervasive computing and communication workshops (PerCom Workshops)*, pages 1–6. IEEE, 2016.

[4] Debraj De, Pratool Bharti, Sajal K Das, and Sriram Chellappan. Multimodal wearable sensing for fine-grained activity recognition in healthcare. *IEEE Internet Computing*, 19(5):26–35, 2015.

[5] Florenc Demrozi, Graziano Pravadelli, Azra Bihorac, and Parisa Rashidi. Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. *IEEE Access*, 8:210816–210836, 2020.

[6] Eiman Kanjo, Eman MG Younis, and Chee Siang Ang. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. *Information Fusion*, 49:46–56, 2019.

[7] Isah A Lawal and Sophia Bano. Deep human activity recognition using wearable sensors. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 45–48, 2019.

[8] Li Liu, Yuxin Peng, Ming Liu, and Zigang Huang. Sensor-based human activity recognition system with a multilayered model using time series shapelets. *Knowledge-Based Systems*, 90:138–152, 2015.

[9] Ye Liu, Liqiang Nie, Li Liu, and David S Rosenblum. From action to activity: sensor-based activity recognition. *Neurocomputing*, 181:108–115, 2016.

[10] Abdul Kadar Muhammad Masum, Arnab Barua, Erfanul Hoque Bahadur, Mohammad Robiul Alam, Md Akib Uz Zaman Chowdhury, and Mohammed Shamsul Alam. Human activity recognition using multiple smartphone sensors. In *2018 international conference on innovations in science, engineering and technology (ICISET)*, pages 468–473. IEEE, 2018.

[11] Inma Mohino-Herranz, Roberto Gil-Pita, Manuel Rosa-Zurera, and Fernando Seoane. Activity recognition using wearable physiological measurements: Selection of features from a comprehensive literature study. *Sensors*, 19(24), 2019.

[12] Natalia Neverova, Christian Wolf, Griffin Lacey, Lex Fridman, Deepak Chandra, Brandon Barbello, and Graham Taylor. Learning human identity from motion patterns. *IEEE Access*, 4:1810–1820, 2016.

[13] Henry Friday Nweke, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-Garadi. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion*, 46:147–170, 2019.

[14] Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1):4–21, 2016.

[15] Attila Reiss and Didier Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers*, pages 108–109. IEEE, 2012.

[16] Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Kilian Förster, Gerhard Tröster, Paul Lukowicz, David Bannach, Gerald Pirkl, Alois Ferscha, et al. Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh international conference on networked sensing systems (INSS)*, pages 233–240. IEEE, 2010.

[17] Abdulhamit Subasi, Dalia H Dammas, Rahaf D Alghamdi, Raghad A Makawi, Eman A Albiety, Tayeb Brahimi, and Akila Sarirete. Sensor based human activity recognition using adaboost ensemble classifier. *Procedia computer science*, 140:104–111, 2018.

[18] Md Zia Uddin, Mohammed Mehedi Hassan, Ahmed Alsanad, and Claudio Savaglio. A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. *Information Fusion*, 55:105–115, 2020.

[19] Yan Wang, Shuang Cang, and Hongnian Yu. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137:167–190, 2019.