2020

A Comparative Study of Machine Learning Approaches for Human Activity Recognition

Loknath Sai Ambati
Dakota State University

Omar El-Gayar
Dakota State University

A Comparative Study of Machine Learning Approaches for Human Activity Recognition

Loknath Sai Ambati
Dakota State University

Omar F. El-Gayar
Dakota State University

5-28-2020

A Comparative Study of Machine Learning Approaches for Human Activity Recognition

Loknath Sai Ambati
loknathsai.ambati@trojans.dsu.edu

Omar El-Gayar

A Comparative Study of Machine Learning Approaches for Human Activity Recognition

Loknath Sai Ambati
Dakota State University
LoknathSai.Ambati@trojans.dsu.edu

Omar El-Gayar
Dakota State University
Omar.El-Gayar@dsu.edu

ABSTRACT

The goal of this project is to study the performance of Machine Learning (ML) techniques used in Human Activity Recognition (HAR). Specifically, we aim to 1) evaluate and benchmark the performance of various ML techniques used for HAR against established ML performance metrics using multiple datasets, and 2) map the characteristics of various HAR datasets to appropriate ML techniques. From a theoretical perspective, the research will shed light into the strengths and weaknesses of various ML techniques that can provide insights into future research aimed at improving these techniques for HAR. From a practical perspective, the research provides guidance into the applicability of various ML techniques to HAR datasets. Overall, studies into improving HAR performance could lead to a significant improvement in the self-care and self-management interventions. These improvements would open doors for creative innovations in healthcare and other commercial applications that require the detection of human activity.

Keywords

Human Activity Recognition, Machine Learning, Performance, Healthcare, Benchmark.

INTRODUCTION

Wearable technology has gained much attention in recent years (Iqbal et al. 2018). Prominent applications include self-care aimed at improving health and well-being, and self-management which focuses on facilitating the management of chronic diseases. Most HAR applications employ Machine Learning (ML) techniques (Meyer et al. 2016) with varying degrees of success. The HAR data is unique and different compared to other types of data. Characteristics such as sensor type, placement of sensor position on body and preprocessing steps contribute to the uniqueness of the HAR data. Further, it is not clear, which ML techniques works best for HAR, creating the need for further comparative studies.

The main focus of research studies in HAR is on improving and developing novel models for rare activities in unique environments. Although, recent studies attempted to address the performance of machine learning models on HAR (Baldominos et al. 2019; Nabian 2017), there remains a gap for a comprehensive study to benchmark the performance of various ML techniques when applied to different HAR datasets. Accordingly, the objective of this research is to compare and contrast the performance of various ML techniques using multiple HAR datasets.

The significance of the project is two-fold: From a theoretical perspective, the research will shed light into the strengths and weaknesses of various ML techniques that can provide insights into future research aimed at improving these techniques for HAR. From a practical perspective, the research provides guidance into the applicability of various ML techniques to HAR datasets. Overall, studies into improving HAR performance could significantly support self-care and self-management interventions. These improvements would open doors for creative innovations in healthcare and other commercial applications that require the detection of human activity.

The rest of the paper is organized as follows: a brief literature review is presented in section 2, following the introduction. Section 3 describes the methodology Section 4 and Section 5 summarizes and discusses the results by comparing and contrasting with extant literature. Finally, section 6 concludes the paper with a summary of key contributions, limitations, directions for future research.

LITERATURE REVIEW

The processing level of the raw data obtained from the sensors follows a typical Activity Recognition Chain (ARC) model (Bulling et al. 2014) for classifying human activities as shown in Figure 1. The ARC model involves sampling of raw data, preprocessing, segmentation, feature extraction, and finally, classification. Usually, one requires deep domain expertise to make it to the feature extraction step. Due to this reason, most researchers rely on domain experts for feature extraction and feature engineering. Machine learning and deep learning techniques are utilized to classify human activities using the resultant features (Saha et al. 2018).

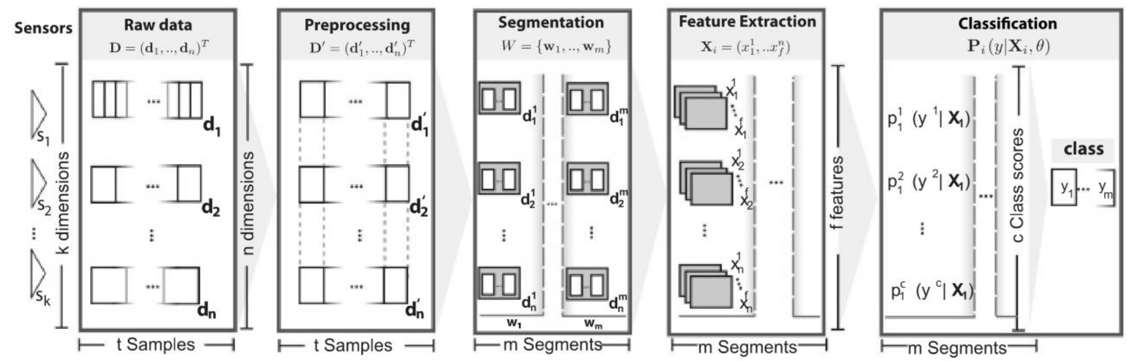

Figure 1: Activity Recognition Chain (ARC) Model (Bulling et al. 2014)

Most of the HAR literature deals with preprocessing and classifying the data using common ML techniques (Baldominos et al. 2019; Jain and Kanhangad 2018; Nabian 2017; Ronao and Cho 2017; Sousa et al. 2017). These types of analyses compare various ML techniques to identify the most suitable technique for a HAR dataset. Despite some attempts to compare various ML techniques on multiple HAR datasets (Dohnálek et al. 2014; Li et al. 2018), there is a need for a comprehensive study to evaluate and benchmark the performance of various ML techniques with different HAR datasets and map the characteristics of various HAR datasets to appropriate ML techniques. These characteristics can assist in building a better framework for different HAR applications.

METHODOLOGY

Datasets:

We relied on the University of California, Irvine (UCI) data repository to obtain HAR datasets. These datasets were used in prior research (Gaikwad et al. 2019; Garcia-Ceja and Brena 2015) (Anguita et al. 2013; Nakano and Chakraborty 2017), and are thus suitable for use for benchmarking various ML algorithms. Table 1 presents the data sets and their characteristics.

Dataset	Sensors	Sensor Position	Activities performed	Dataset description	Sampling Frequency
Pamap2	3 Colibri wireless IMUs (inertial measurement units) and BM-CS5SR (HR monitor) – Accelerometer, Gyroscope, magnetic sensor and temperature sensor.	wrist, chest and side ankle.	lying, sitting, standing, walking, running, cycling, Nordic walking, watching TV, computer work, car driving, ascending stairs, descending stairs, vacuum cleaning, ironing, folding laundry, house cleaning, playing soccer and rope jumping.	9 subjects (1 female and 8 male) aged 27.22 (+-) 3.31 years performed the 12 mandatory activities and 6 optional activities for 2-3 minutes.	100 samples/sec
Mhealth	accelerometer, a gyroscope, a magnetometer and ecg (Shimmer2 [BUR10] wearable sensors).	chest, right wrist and left ankle	L1: Standing still (1 min), L2: Sitting and relaxing (1 min), L3: Lying down (1 min), L4: Walking (1 min), L5: Climbing stairs (1 min), L6: Waist bends forward (20x), L7: Frontal elevation of arms (20x), L8: Knees bending (crouching) (20x), L9: Cycling (1 min), L10: Jogging (1 min), L11: Running (1 min), L12: Jump front & back (20x)	10 volunteers of diverse profile performed 12 physical activities for about 1 min	50 samples/sec
SWELL	accelerometer, a gyroscope, a magnetometer, and a linear acceleration sensor (Samsung Galaxy SII (i9100) smartphone).	upper arm, wrist, two pockets, and belt position	walking, sitting, standing, jogging, biking, walking upstairs and walking downstairs	10 participants performed 7 activities for 3-4 minutes. All are male with ages 25-30.	50 samples/sec

Table 1. Dataset Characteristics

Analysis:

The data is inspected for duplicates, null values, and data imbalance. We compare different ML techniques commonly used for HAR using the datasets from the UCI data repository. Techniques considered are naïve Bayes, logistic regression, decision tree, decision tree with entropy, random forest, and gradient boosting decision tree algorithm. Although the application of neural network based techniques are popular they often run into overfitting problems in the case of HAR (Jobanputra et al. 2019). Further, run time is over ten minutes per dataset on an intel i7 8^{th} generation system (considering large HAR datasets), which would increase exponentially with Artificial Neural Networks (ANN). Hence, ANN were not included in this research. We use accuracy as the metric to evaluate the ML techniques. Accuracy is highly recommended for evaluating the performance of machine learning models when dealing with HAR (Li et al. 2018).

RESULTS:

The datasets are preprocessed and standardized into a format where each independent column represents the sensor data (except timestamp and subject ID), dependent column represents the activity performed and each row representing the sample values from each sensor for a particular sampling time. Once, all the three datasets are standardized into a similar format after the preprocessing, we split each dataset randomly into training (70%) and test data (30%). Then we train each dataset with the selected ML techniques. We evaluate each ML technique using accuracy for every dataset, respectively.

In all the datasets, 3D accelerometer, 3D Gyroscope and 3D Magnetometer are common. Pamap2 has a heart rate monitor and temperature sensor as unique sensors while Mhealth has ECG as unique sensor and SWELL has linear acceleration sensor as unique. SWELL uses smartphone to collect the data, Mhealth uses wearable sensors and Pamap2 uses wireless IMU's to collect the data.

Dataset	**Naïve Bayes**	**Logistic Regression**	**Decision Tree**	**Decision Tree with Entropy**	**Random Forest**	**XGBoost**
Pamap2	90.19%	92.03%	99.93%	99.96%	**99.99%**	**99.99%**
Mhealth	52.12%	73.82%	91.11%	91.84%	**94.1%**	93.41%
SWELL	87.98%	85.57%	97.61%	97.79%	99.57%	**99.92%**

Table 2. Accuracy scores for each dataset with respect to ML techniques

As shown in Table 2 the Pamap2 dataset has the highest accuracy values when compared to other datasets for all ML technique. Naïve Bayes technique performed better than logistic regression in the case of Mhealth compared to the other two datasets. Similarly, decision tree with entropy performed better than decision tree technique using all three datasets and Random forest technique performed better than decision tree with entropy in all three cases. The accuracy values remained same for random forest and XGBoost for Pampap2, Random forest performed better than XGBoost for SWELL and XGBoost performed better than random forest for Mhealth.

DISCUSSION

Overall, all ML algorithms performed better when using the Pamap2 dataset. A possible explanation is the inclusions of two additional sensors (heart rate monitor and temperature sensor) in this dataset. Another possible explanation is the higher sampling rate (100Hz sampling frequency rather than 50Hz). In essence, the improvement in results across all algorithms is associated with the higher sampling frequency and number of sensors utilized to collect the HAR data. Commonly used sensor positions in all datasets are chest, wrist, and ankle. The belt and pocket position in SWELL have made a positive impact on the ML accuracy scores.

When naïve Bayes and logistic regression techniques are compared, naïve Bayes performed better only with the SWELL dataset. The unique features of SWELL data are the usage of smartphone to collect the data and the additional sensor positions on the body to collect the data. Given this context, we can consider that naïve Bayes technique performs better with HAR data collected from smart phone and the number of sensor positions on the body. Decision tree techniques have outperformed naïve Bayes and logistic regression. Ensemble models such as random forest and XGBoost techniques outperformed traditional machine learning techniques in the case of HAR.

CONCLUSION

In this research, we evaluated the performance of various ML techniques used for HAR using classification accuracy on multiple datasets and tried to map the characteristics of various HAR datasets to appropriate ML techniques. Keeping all the HAR data characteristics constant, the accuracy of ML techniques is dependent on sampling frequency, amount of data collected from the sensors and number of sensors utilized to collect the HAR data. The factors that might affect the ML performance are sensors used to collect the data, sensor position on body, activities performed, number of participants performing the activities and the frequency of sampling at which the data is collected. The number of sensors and sampling frequency had a positive impact on the HAR data quality in general. The belt, pocket positions and smart phone data had a positive impact on naïve Bayes technique. Overall, there is no single ML technique that works best for all HAR. Accordingly, we need to consider different techniques based on the factors and characteristics of the HAR data, respectively. Future research could examine the performance on additional HAR datasets to improve the understanding of the relation between the HAR data characteristics and the performance of various ML algorithms. Further attention is warranted on examining HAR datasets with wide variety of characteristics and including various ML techniques especially neural network-based techniques. The validity of the research can be increased by performing different preprocessing steps to the datasets and optimizing the hyperparameters of the various ML techniques.

REFERENCES

1. Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J. L. 2013. "A Public Domain Dataset for Human Activity Recognition Using Smartphones," in *ESANN*.

2. Baldominos, A., Cervantes, A., Saez, Y., and Isasi, P. 2019. "A Comparison of Machine Learning and Deep Learning Techniques for Activity Recognition Using Mobile Devices," *Sensors* (19:3), p. 521. (https://doi.org/10.3390/s19030521).

3. Bulling, A., Blanke, U., and Schiele, B. 2014. "A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors," *ACM Comput. Surv.* (46:3), 33:1–33:33. (https://doi.org/10.1145/2499621).

4. Dohnálek, P., Gajdoš, P., and Peterek, T. 2014. "Human Activity Recognition: Classifier Performance Evaluation on Multiple Datasets," *Journal of Vibroengineering* (16:3), pp. 1523–1534.

5. Gaikwad, N. B., Tiwari, V., Keskar, A., and Shivaprakash, N. C. 2019. "Efficient FPGA Implementation of Multilayer Perceptron for Real-Time Human Activity Classification," *IEEE Access* (7), pp. 26696–26706. (https://doi.org/10.1109/ACCESS.2019.2900084).

6. Garcia-Ceja, E., and Brena, R. 2015. "Building Personalized Activity Recognition Models with Scarce Labeled Data Based on Class Similarities," in *Ubiquitous Computing and Ambient Intelligence. Sensing, Processing, and Using Environmental Information*, Lecture Notes in Computer Science, J. M. García-Chamizo, G. Fortino, and S. F. Ochoa (eds.), Springer International Publishing, pp. 265–276.

7. Iqbal, Z., Ilyas, R., Shahzad, W., and Inayat, I. 2018. "A Comparative Study of Machine Learning Techniques Used in Non-Clinical Systems for Continuous Healthcare of Independent Livings," in *2018 IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE)*, , April, pp. 406–411. (https://doi.org/10.1109/ISCAIE.2018.8405507).

8. Jain, A., and Kanhangad, V. 2018. "Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors," *IEEE Sensors Journal* (18:3), pp. 1169–1177. (https://doi.org/10.1109/JSEN.2017.2782492).

9. Jobanputra, C., Bavishi, J., and Doshi, N. 2019. "Human Activity Recognition: A Survey," *Procedia Computer Science* (155), The 16th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2019),The 14th International Conference on Future Networks and Communications (FNC-2019),The 9th International Conference on Sustainable Energy Information Technology, pp. 698–703. (https://doi.org/10.1016/j.procs.2019.08.100).

10. Li, F., Shirahama, K., Nisar, M. A., Köping, L., and Grzegorzek, M. 2018. "Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors," *Sensors* (18:2), p. 679. (https://doi.org/10.3390/s18020679).

11. Meyer, J., Schnauber, J., Heuten, W., Wienbergen, H., Hambrecht, R., Appelrath, H., and Boll, S. 2016. "Exploring Longitudinal Use of Activity Trackers," in *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, , October, pp. 198–206. (https://doi.org/10.1109/ICHI.2016.29).

12. Nabian, M. 2017. "A Comparative Study on Machine Learning Classification Models for Activity Recognition," *Journal of Information Technology & Software Engineering* (7:4), pp. 4–8. (https://doi.org/10.4172/2165-7866.1000209).

13. Nakano, K., and Chakraborty, B. 2017. “Effect of Dynamic Feature for Human Activity Recognition Using Smartphone Sensors,” in *2017 IEEE 8th International Conference on Awareness Science and Technology (ICAST)*, , November, pp. 539–543. (https://doi.org/10.1109/ICAwST.2017.8256516).

14. Ronao, C. A., and Cho, S.-B. 2017. “Recognizing Human Activities from Smartphone Sensors Using Hierarchical Continuous Hidden Markov Models,” *International Journal of Distributed Sensor Networks* (13:1), p. 1550147716683687. (https://doi.org/10.1177/1550147716683687).

15. Saha, S. S., Rahman, S., Rasna, M. J., Zahid, T. B., Islam, A. K. M. M., and Ahad, M. A. R. 2018. “Feature Extraction, Performance Analysis and System Design Using the DU Mobility Dataset,” *IEEE Access* (6), pp. 44776–44786. (https://doi.org/10.1109/ACCESS.2018.2865093).

16. Sousa, W., Souto, E., Rodrigres, J., Sadarc, P., Jalali, R., and El-Khatib, K. 2017. “A Comparative Analysis of the Impact of Features on Human Activity Recognition with Smartphone Sensors,” in *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*, WebMedia ’17, New York, NY, USA: ACM, pp. 397–404. (https://doi.org/10.1145/3126858.3126859).