Multi-Label Classification in Human Activity Recognition: A Comparison Between Logistic Regression and k-Nearest Neighbors

Caíque Santos Lima
{[redacted]@dac.unicamp.br}

Departamento de Engenharia de Computação e Automação (DCA)
Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (Unicamp)
Campinas, SP, Brasil

Abstract — Human activity recognition (HAR) has lately become a field of high interest within medical, sporty, aging care, rehabilitation scenarios. The ability to automatically recognize a person's context can be done through gadgets used in everyday life, especially wearables. HAR systems process the signals from the sensors present in these devices and extract features to train machine learning (ML) models capable of recognizing the activities performed by the user. The applications range from fitness tracking apps to solutions capable of calling the emergency if the user has suffered a serious accident. In this paper, two supervised ML approaches were compared using a public dataset for recognition of four selected everyday activities.

Keywords — Feature engineering. Human activity recognition. Machine Learning. Multiclass classification. Wearables.

1. Introduction

The behavioral context of a person depends on different circumstances, i.e., *is the person alone or accompanied? is he/she indoors or outdoors? what kind of activity is the person doing?* and so on. Human activity recognition (HAR) has become a area of study of high interest within medical, sporty, aging care, rehabilitation scenarios. The ability to automatically recognize a person's context is very beneficial in many domains and this can be done through gadgets (e.g. smartphones, smartwatches, etc). The emergence of wearable devices has enabled applications that help monitoring sedentary behavior, protecting people from Office Workers Syndrome (OWS), tracking physical activity, sleep monitoring, elderly and specific care [3]. These applications range from a pop-up that notifies the user that he/she has already spent too much time sitting down and it is time to move, to calling an emergency if the person has suffered a serious accident.

The most commonly activities related to HAR research are *walking*, *running*, *biking*, *jogging*, *remaining still*, *walking upstairs*, and *walking downstairs* [6]. But the set of activities are not restricted to just these, other activities with a high level of complexity, such as *working* and *studying*, are also included. For a simple activity set, that includes only the activities of moving and not moving — *walking* and *sitting*, for example — can be used inertial sensors to measure proper acceleration, orientation and angular velocity of the body. For more complex scenarios — *is the person at home? is the person with friends?* — context signals may be needed. In these cases, additional signals such as GPS and audio may be used, although such sensors may pose user privacy issues.

In addition to technical complexity, HAR is intrinsically connected to ethical aspects. There is a fine line between the benefits provided by these applications and the privacy and integrity of user data. To deal with this issue, a learning paradigm emerges seeking to address the problem of data governance and privacy by training algorithms collaboratively without exchanging the data itself. This new approach is known as Federated Learning (FL) [5]. FL allows training a global model without moving data beyond the firewalls of the edge devices. Instead, the machine learning (ML) process occurs locally at each participating federated entity and only model characteristics (e.g., parameters, gradients) are shared. In this regard, FL can be applied to mitigate the risk of data exposure in HAR systems and preserve users' privacy.

According to [8], behavioral context is wide and complex, in other words, people interact with their gadgets in different manners and do not focus on a single activity. An activity like running can have different contexts: outside, indoors on a treadmill, at the gym, at home, alone, with friends, and so on. People perform activities in a different manner, due fitness, gender, age, etc. In such a broad setting, [10] points out six challenges to be considered: (1) the selection of the attributes to be measured; (2) the construction of a portable, unobtrusive, and inexpensive data acquisition system; (3) the design of feature extraction and inference methods; (4) the collection of data under realistic conditions; (5) the flexibility to support new users without the need of re-training the system; and (6) the implementation in mobile devices meeting energy and processing requirements.

As discussed in [3], the general pipeline encompass-

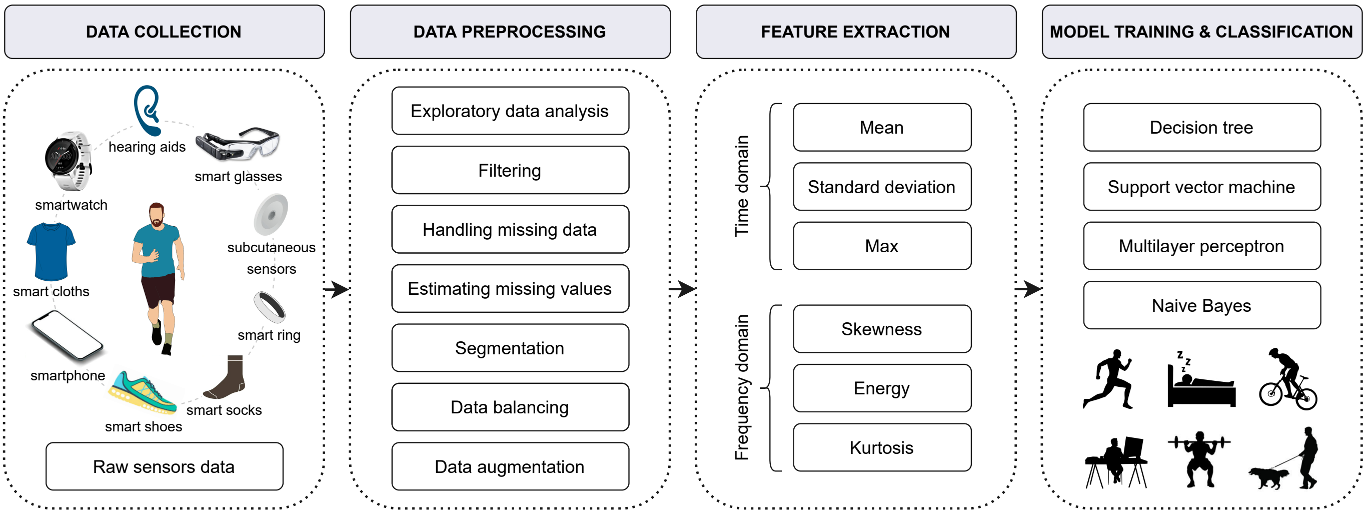

Figure 1. The general pipeline of HAR systems based in ML approaches.

ing HAR systems can be basically arranged into four main steps: *data collection*, *data preprocessing*, *feature extraction* and *model training and classification*, as depicted in Figure 1. In the first step, the raw data are gathered from a smart shoes — which registers movement, rotation, speed, impact and weight distribution in the gait [1] — or even from a smartphone kept in your pocket, for example. Then, in step two, the raw data is cleaned and prepared. In the third and fourth stages, characteristic features are extracted and the model training process is carried out to infer the activities, respectively.

Automatic HAR can be performed using a fusion of different types of sensors embedded in gadgets. If the person is moving, the system then uses a ML algorithm to estimate whether the person is jogging, running or cycling. On the other hand, if the system is not moving, another algorithm can be used to determine whether the person is sitting up watching TV or lying down asleep. More details on how signals are processed and how algorithms were developed and assessed in this work are described in the section 2. The performance of these algorithms are properly presented and discussed in section 3. Section 4. concludes the paper. And, finally, future perspectives are presented in the section 5.

2. Materials and Methods

The design of any HAR system depends on the activities to be recognized [10]. In other words, the choice of which variables are relevant to the problem is associated with the context of each activity to be recognized. For a simple activity set, that includes only the activities of moving and not moving, thresholding the standard deviation of 3-axis acceleration magnitude can be enough to reach an accuracy of 99.44%, as demonstraded in [9]. On the other hand, for more complex activity sets, ML models for context recognition can be applied, as shown in [8].

The training stage of these models requires a time series of measured attributes from individuals performing each activity. In turn, the time series are split into time windows to apply feature extraction thereby filtering relevant information in the raw signals. Then, ML methods are used to generate an activity recognition model from the data of extracted features. Besides, for testing, data are collected during a time window, which is used to extract features. Such feature set is evaluated in the priorly trained ML model, generating a predicted activity. Figure 2 shows the common steps involved in the processes described above.

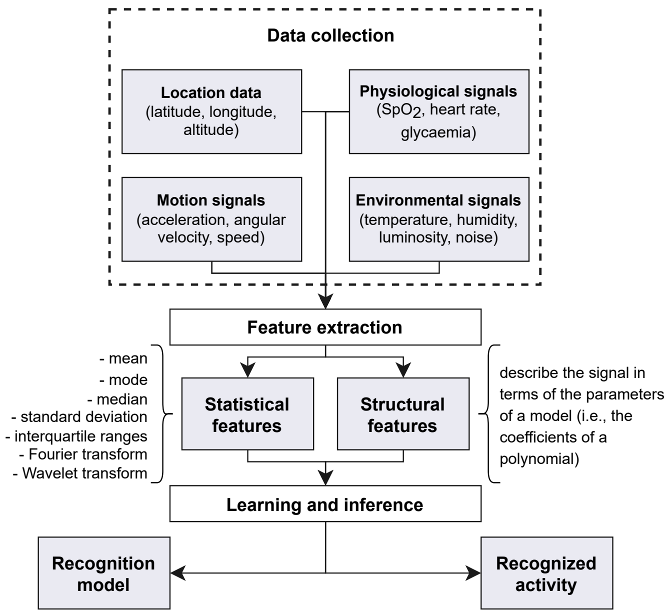

Figure 2. Framework for training and testing HAR systems based on wearable devices.

In this work, the *ExtraSensory Dataset* was used to assess the proposed models. This public dataset is made up of a total of 308,320 labeled examples (in minutes) from sixty volunteers carrying out 51 daily activities. Table 1 details statistics (over 60 participants) on the amount of data collected. Every example contains measurements from the user's personal smartphone and from a smartwatch that the researchers provided for some of them. Not all the sensors were available at all times, as Androids and iOS devices were used [7]. Table 2 specifies details about the sensors.

	Range	Mean (SD)
Age (years)	18-42	24.7 (5.6)
Height (cm)	145-188	171 (9)
Weight (kg)	50-93	66 (11)
Body mass index (kg/m^2)	18-32	23 (3)
Labeled examples	685-9706	5139 (2332)
Additional unlabeled examples	2-6218	1150 (1246)
Average applied labels/example	1.1-9.7	3.8 (1.4)
Participation duration (days)	2.9-28.1	7.6 (3.2)

Table 1. Statistics over the 60 participants in the *ExtraSensory Dataset* [7] (SD: standard deviation).

In order to evaluate classification performance of the two approaches used in this work — *logistic regression* and *k-nearest neighbors* (k-NN) algorithms —, four activities were selected: *sitting*, *walking*, *sleeping* and *lying down*. Volunteer data were subdivided into *training* and *test* sets. In the direction of evaluating the ability of models to generalize to unseen examples — which is one of the main objectives for HAR systems to be used by different users, as discussed in [4] —, the algorithms were trained with data from 59 participants and later tested with data from the remaining subject (ID 098A72A5). The data used has over 263,000 labeled examples from sixty participants. Every example represents one minute and has measurements from various sensors previously presented in Table 2. Figure 3 presents the distribution of examples in each of the selected classes in both used sets.

The sensors used were diverse and include *low frequency sensors* (light, air pressure, humidity, temperature, proximity), *audio properties* (max absolute value of recorded audio), *location* (latitude, longitude, altitude, speed, accuracies), *magnetometer* (tri-axial direction and magnitude of magnetic field), *accelerometer* (tri-axial direction and magnitude of acceleration), *watch accelerometer and compass* (tri-axial acceleration and watch heading in degrees), *phone state* (app status, battery state, Wi-Fi availability on the phone, time-of-day) and *gyroscope* (rate of rotation around phone's 3 axes). From these sensors, a total of 225 statistical features were extracted, such as: *mean*, *standard deviation*, *percentiles*, *entropy*, *energy* among others.

Sensor	Raw measurements	Examples	Users
Accelerometer	3-axis (40 Hz)	308,306	60
Gyroscope	3-axis (40 Hz)	291,883	57
Magnetometer	3-axis (40 Hz)	282,527	58
Watch accelerometer	3-axis (25 Hz)	210,716	56
Watch compass	heading angle (var)	126,781	53
Location	long-lat-alt (var)	273,737	58
Location precomputed	location variability (1 pe)	263,899	58
Audio	13MFCC (46 ms frames)	302,177	60
Audio power	1 pe	303,877	60
Phone state	1 pe	308,320	60
Low frequency sensors	1 pe	308,312	60

Table 2. The sensors in the *ExtraSensory Dataset* [7] ("1 pe" means sampled once per example, "var" means variable sampling rate — gathering updates whenever the value changes).

Figure 3. Time for every selected context label in each set.

The statistical features were standardized by removing the mean and scaling to unit variance in order to enhance algorithms performance. After preparing the data, the proposed models were properly trained. The first was a multiclass classifier based on logistic regression. The other model was implemented using the k-NN method, where a search was made for the k closest training examples that minimized the average error over the training data. The two supervised approaches were compared against test data in terms of *accuracy* (percentage of how many selected activities each model was able to classify correctly), *sensetivity* (it represents the proportion of correctly classified examples among the total number of examples of each selected activity, also known as the true positive rate — TPR), *specificity* (it represents the proportion of correctly classified examples which do not belong to the class that one wants to predict among the non-target activities, also known as the true negative rate — TNR), *balanced accuracy* (i.e., the average sensetivity obtained for each class) and *precision* (it corresponds to the proportion of patterns of the positive class, or target activity, correctly classified in relation to all examples assigned to the positive class).

3. Results and Discussion

As per the heatmap shown in Figure 4, for examples present in the dataset, there is strong correlation between certain selected activities. For example, *lying down* comes always with *sleeping*, so it is possible to see a strong correlation (0.83) between these activities. On the other hand, *sleeping* and *walking* show low correlation (–0.13), since they are activities that do not happen concurrently, in normal situations. This type of analysis is particularly interesting, especially in less obvious relationships, as it can help to check the correlation between sensor features to be used in the models. Since there is a high correlation between sleeping and lying down, different sensors can be explored to characterize these activities: a light sensor can be useful to help distinguish them, since people tend to sleep in dark environments and usually they sit in bright environments (in the kitchen, in the office, in the classroom), for example.

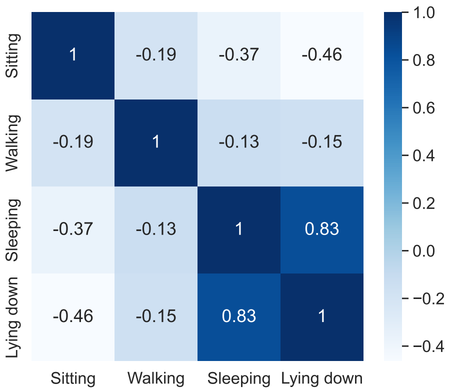

Figure 4. Correlation between selected context labels.

To visualize the training data, besides a general nonlinear dimension reduction, the dimension reduction technique UMAP [2] was used. The `n_neighbors` parameter for this projection, which controls how UMAP balances local versus global structure in the data, is 15. The `min_dist` applied was 0.1, the function of this parameter is to provide the minimum distance (euclidean, in this case) apart that points are allowed to be in the low dimensional representation. The dimensionality of the reduced dimension space is 2. The UMAP projection of selected activities from the *ExtraSensory Dataset* onto the training set is shown in Figure 5 below. This figure shows two larger clusters, although it is not possible to see a high separability as these clusters are composed of multiple activities. In this case, one can explore the combination of sensors and features extraction that are more appropriate to separate such activities. For example, audio and light signals may be more relevant to distinguish between walking and sleeping activities, since the former usually occurs in brighter and noisy environments, while the latter occurs in places with opposite characteristics.

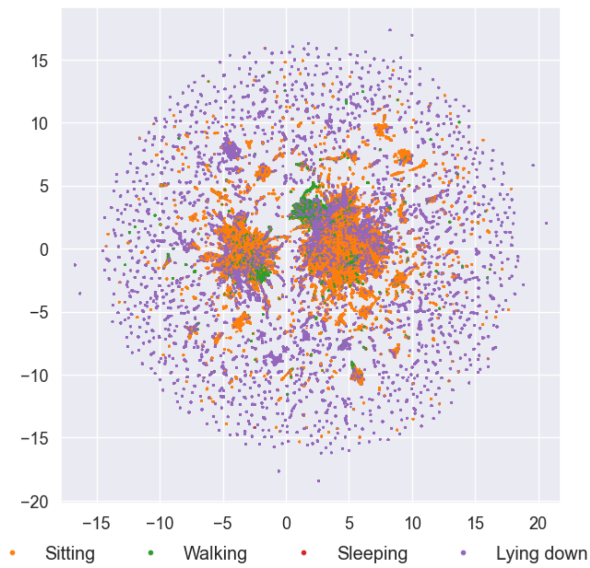

Figure 5. UMAP projection of selected context labels.

Regarding training time, as expected, the logistic regression method was faster than k-NN. The former took just 55.6s and the latter took 2h 16min 56s. k-NN model had an average performance of 0.81 ± 0.13 (sleeping) and 0.8 ± 0.15 (lying down) in the evaluated metrics. Using the logistic regression method, the average performance was 0.8 ± 0.08 (sleeping) and 0.88 ± 0.05 (lying down). On the other hand, two activities were particularly challenging for both models around TPR and precision: sitting and walking. k-NN model had an average performance of 0.73 ± 0.07 (sitting) and 0.76 ± 0.22 (walking) in the evaluated metrics. Using the logistic regression method, the average performance was 0.61 ± 0.09 (sitting) and 0.71 ± 0.28 (walking). The detailed result of every model for each selected activity can be seen in Figure 6.

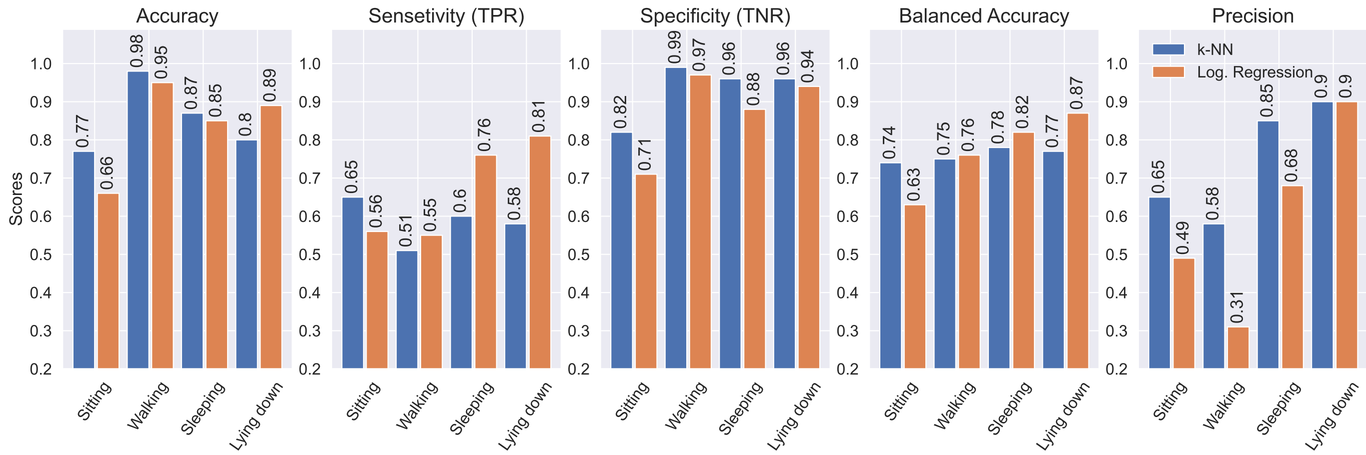

Figure 6. Performance of the models evaluated.

4. Conclusion

A person's behavioral context depends on several factors and ranges from a simple daily task performed by a person to a more complex context. In this scenario, HAR systems have been developed to be applied in different well-being fields: *health*, *sports*, *aging care*, *rehabilitation* and many others. These systems are powered by ML models that use data from sensors present in gadgets used in everyday life, especially wearables. Data is processed and important features are extracted and then used for training models that are embedded in these devices. These applications may be able to recognize whether a person is sleeping, walking, cooking or even sedentary. In this paper, two supervised ML approaches were compared using a public dataset for recognition of four selected everyday activities. In addition to demonstrating the performance of the assessed techniques, insights from data analysis were also discussed and some considerations were raised.

5. Future Works

During exploratory data analysis in this work, it was noted that some of the activities were more challenging to discriminate. As discussed in the section 3., proper sensing data can simplify the activity classification process. For example, the watch accelerometer can be a good indicator for specific hand-motion activities, like *typing on the computer* or *cooking*, while audio might better predict environmental contexts like *in class* or *at a party*. In this sense, one can explore the combination of sensors and feature extraction along with a cross-validation process in model training in order to seek even more accurate performance in the HAR context. Likewise, models based on deep learning could also be evaluated, also considering the hardware constraints that wearable devices tend to have, such as processing capacity, storage and battery.

References

[1] Continental AG. Connected footwear can smart shoe technology change the way you run?, 2022.

[2] McInnes, Leland et al. Umap: Uniform manifold approximation and projection for dimension reduction, 2018.

[3] Mekruksavanich, Sakorn et al. Enhanced Hand-Oriented activity recognition based on smartwatch sensor data using LSTMs. *Symmetry*, 12(9):1570, September 2020.

[4] Morales, Jafet et al. Physical activity recognition by smartphones, a survey. *Biocybernetics and Biomed. Engineering*, 37(3):388–400, January 2017.

[5] Rieke, Nicola et al. The future of digital health with federated learning. *NPJ Digital Medicine*, 3:119, September 2020.

[6] Shoaib, Muhammad et al. A survey of online activity recognition using mobile phones. *Sensors*, 15(1):2059–2085, 2015.

[7] Vaizman, Yonatan et al. Recognizing detailed human context in-the-wild from smartphones and smartwatches. *IEEE Perv. Comp.*, 16(4):62–74, October 2017.

[8] Vaizman, Yonatan et al. Context recognition In-the-Wild: Unified model for Multi-Modal sensors and Multi-Label classification. *Proc. ACM Interact. Mob. Wearable Ub. Tec.*, 1(4):1–22, January 2018.

[9] Wang, Yi et al. A framework of energy efficient mobile sensing for automatic user state recognition. In *Proce. of the 7th Intern. Conf. on Mobile Sys., Applic. and Serv.*, MobiSys '09, page 179–192, New York, NY, USA, 2009. ACM.

[10] Oscar D. et al Lara. A survey on human activity recognition using wearable sensors. *IEEE Commun. Surveys & Tutorials*, 15(3):1192–1209, 2013.