Detailed Human Activity Recognition based on Multiple HMM

Mariana Abreu[1], Marília Barandas[1], Ricardo Leonardo[1] and Hugo Gamboa[2]

[1]*Associacao Fraunhofer Portugal Research, Rua Alfredo Allen 455/461, Porto, Portugal*

[2]*Laboratorio de Instrumentacao, Engenharia Biomedica e Fisica da Radiacao (LIBPhys-UNL), Departamento de Fisica, Faculdade de Ciencias e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal*

Keywords: Human Activity Recognition, Gesture Recognition, Smartphone Sensors, Feature Selection, Hidden Markov Models.

Abstract: A wide array of activities is performed by humans, everyday. In healthcare, precocious detection of movement changes in daily activities and their monitoring, are important contributors to assess the patient general well-being. Several previous studies are successful in activity recognition, but few of them provide a meticulous discrimination. Hereby, we created a novel framework specialized in detailed human activities, where signals from four sensors were used: accelerometer, gyroscope, magnetometer and microphone. A new dataset was created, with 10 complex activities, suchlike opening a door, brushing the teeth and typing on the keyboard. The classifier was based on multiple hidden Markov models, one per activity. The developed solution was evaluated in the offline context, where it achieved an accuracy of $84 \pm 4.8\%$. It also showed a solid performance in other performed tests, where it was tested with different detailed activities, and in simulations of real time recognition. This solution can be applied in elderly monitoring to access their well-being and also in the early detection of degenerative diseases.

1 INTRODUCTION

The inherent complexity of human behaviour tends to promote well-defined motions which are repeated on everyday basis. In this sense, several areas of the biomedical field could benefit from the recognition of detailed activities including health-care, elderly monitoring and lifestyle. Nowadays, smartphones possess multiple accurate sensors to better assist humans, which makes them prime candidates for monitoring human activities.

There are several applications with smartphone and wearable sensors, able to correctly discriminate between physical activities, such as *Walking* and *Sitting*. Furthermore, previous studies can successfully recognize complex activities like *Cooking* or *Cleaning*, recurring to numerous sensors (Kabir et al., 2016). To our knowledge, research is scarce when it comes to a more detailed discrimination with few sensors, such as between *opening a door* or *answer the phone*. These detailed activities are complex since they involve a physical state (standing, sitting) and the use of hands to perform a specific movement or interact with an object. We call them detailed, since they provide a more detailed information about the user, when compared to physical activities.

The motivation of this work is the development of a new solution, for detailed human activities recognition and monitoring, using only a sensing device and machine learning algorithms. This work includes the discrimination between several detailed activities and also their detection in a real time simulation. A higher definition of human activity monitoring, could enable a more detailed view of a subject's lifestyle and health.

In the past years, several studies are approaching the challenge of Human Activity Recognition (HAR) from different perspectives. The challenge associated with HAR is related to the amount of activities of interest and their characteristics. Lara *et al.* (Lara and Labrador, 2012) states that the complexity of the pattern recognition problem is determined by the set of activities selected. Even short activities such as *opening a door* or *picking up an object* have a broad variety of ways to be executed, which increases with the consideration of different users (Kreil et al., 2014).

For physical activity recognition, suchlike *walking* and *standing*, a high accuracy is achieved with smartphones, recurring mostly to the accelerometer (Machado et al., 2015). However, other strategies must be reckoned, for recognizing more complex activities, with similar body movements, such as open-

Abreu, M., Barandas, M., Leonardo, R. and Gamboa, H.
Detailed Human Activity Recognition based on Multiple HMM.
DOI: 10.5220/0007386901710178
In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019)*, pages 171-178
ISBN: 978-989-758-353-7

ing a door and opening a faucet. Previous works used sound to discriminate activities (Leonardo, 2018; B et al., 2016). With the integration of information from several sensors, a higher degree of discrimination could be achieved.

Given that Hidden Markov Model (HMM) enables the assimilation of the data temporal structure, it becomes an effective technique for classification (Cilla et al., 2009). The choice of multiple HMMs, one per activity, was inspired by some video recognition systems (Gaikwad and Narawade, 2012; Karaman et al., 2014). By having one model per activity, some time periods could be ignored in a continuous stream analysis. Also, at any moment, new activities could be added to the classification, allowing to personalize this tool. Furthermore, temporal sequences, such as daily routines, could be analyzed without the need of an extensive training set.

In the scope of HAR, the recognition performed could be of two types: offline and online. In the offline recognition, each activity sample is well-defined and isolated from other samples. Meanwhile, the online recognition happens in real time, where activities are directly interpreted in the time series. In this case, a sample up for testing could contain one activity, no activity or inbetween cases. In order to cope with these different scenarios, Tapia *et al.* (Tapia et al., 2004) defined some evaluation measures to distinguish totally wrong predictions from nearly right. Moreover, Cardoso *et al.* (Cardoso and Mendes-Moreira, 2016) adds that an activity could be a valid label if it is predicted in a significant amount, at least 30% of the true label.

The activities addressed in this work are complex and short in time, therefore, most of the acquisition signal is ignored and we are only interested in small activities that happen sporadically. Junker *et al.* (Junker et al., 2010) call this type of recognition as Activity Spotting. Kasteren *et al.* (van Kasteren et al., 2011) suggests what metrics should be used for evaluating cases such as ours, namely how to cope with imbalanced classes and what type of errors could occur in a continuous data stream.

In summary, from computer vision to pervasive sensing, many investigations have approached the challenge of human activities. Even so, there is a lack on previous studies when it comes to short detailed activities. Few evidences of such activities exist within the literature, therefore this work is an experiment in a poorly explored HAR field. To address these activities, the chosen classifier is based on multiple HMM and the recording device is a smartphone. Their recognition could expand the range of Activities of Daily Living (ADL) applications and improve current HAR systems.

2 PROPOSED METHOD

The framework created for activity recognition is based on the analysis of data from accelerometer, gyroscope, magnetometer and microphone sensors. The developed solution uses multiple HMM, one per each activity.

2.1 Signal Processing and Feature Extraction

The processing of tri-axial sensors (accelerometer, gyroscope and magnetometer) includes the extraction of all three axes (x, y and z) and also the overall magnitude. In order to address similar activities, such as open a door and open a faucet, sound was considered an important element. Therefore, we chose to combine the inertial sensors with the microphone. These sensors are recorded simultaneously and the starting and ending moments of each activity's repetition are annotated. The signal is segmented in windows of 250ms without overlap. For each 250ms window, over 30 different features were extracted, using a similar approach described in Figueira *et al.* (Figueira, 2016). The features come from temporal, spectral and statistical domains, which are calculated for all sensors and axes. The final output consists in a vector with 265 features for each 250ms window.

2.2 Feature Selection

In order to reduced the amount of features, two standard feature selection methods were tested. **Forward Selection (FS)** evaluates all features separately, choosing the one that leads to a higher performance, to join the set of best features (Cilla et al., 2009). Therefore, in each iteration, one new feature is added to the set, until the accuracy stops improving. On the other hand, in **Backward Elimination (BE)**, the process starts with the whole group of features (265) and one by one the features are removed if the accuracy increases without them. This process stops when accuracy starts decreasing (Li et al., 2015).

2.3 Classification

Hidden Markov Models are capable of interpreting time series. They are represented by a start distribution, a set of states and a set of observations (Machado et al., 2015). The states are associated through transition probabilities, while the observations are associated to the states through emission probabilities. Assuming a discrete clock, in each time the system will be in one specific state and will transit to another state (or itself), on the next iteration. Since the states are hidden, we infer the current state based on the current observations.

In this framework, we built one HMM per activity, where the set of observations is the set of best features. To classify a testing sample, we calculate the probability of each HMM to have generated that sample, using the Viterbi algorithm (Rabiner, 1989). Then, we can select the activity, corresponding to the most likely model, as the correct one.

2.4 Overall System Architecture

In Figure 1, an overall scheme of the framework developed is represented. In order to avoid over-fitting, the leave one user out cross validation method is used, to evaluate the classifier's performance. In summary, the data corresponding to the requested activities is extracted and its features are calculated. Then, it is split into a training group and a testing group by the leave one user out method. The training users will build the HMM for each activity whereas the testing user is used for decoding and prediction. By combining the scores of all users, we reach the final result.

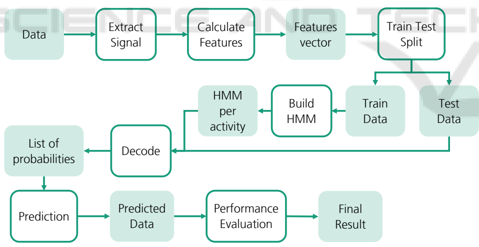

Figure 1: Overall Architecture.

3 EXPERIMENTS

To approach detailed activities, two different datasets were collected in order to test and evaluate the framework in both offline and online recognition.

3.1 Set Up

3.1.1 Sensor Placement

Since the majority of detailed activities are performed using the hands, we chose to place the sensing device on the wrist. The sensors of interest are built-in sensors on common a smartphone, therefore, our sensing device is a Samsung S5, which was attached to the dominant wrist with a wristband as it is shown on Figure 2. Nevertheless, in the final solution, the proposed assembly would be substituted by a smaller sensing device, such as a wearable or a smartwatch, to avoid discomfort to the user.

Figure 2: Smartphone attached to the wrist with a wristband.

3.1.2 Acquisition of Activities

Our dataset contains two types of detailed activities: continuous and isolated. Continuous activities, such as clapping hands or typing on the keyboard, can have different duration, from a few seconds up to minutes. On the other hand, isolated activities, such as opening a door or switching the light on, are usually not repeatable. The acquisition consisted in performing several repetitions of the same activity, where the start and ending of each repetition was annotated. Ten activities were selected, based on common activities we perform in our daily life: opening or closing a door (*Door*), opening or closing a window (*Window*), opening or closing a faucet (*Faucet*), turning the light on or off (*Light*), picking up the telephone (*Phone*), typing on the keyboard (*Keyboard*), mouse clicking and moving (*Mouse*), biting the nails (*NailBiting*), brushing teeth (*BrushTeeth*) and clapping hands (*Clap*).

3.2 Offline Activity Recognition

The first analysis was performed offline. In offline recognition, each sample only contains one activity, the difficulty relies on its correct prediction, among all possible choices.

3.2.1 Composition of the Offline Dataset

The offline dataset considered is composed by 8 users, who performed several repetitions of the activities described in Section 3.1.2. In Figure 3, the distribution of the dataset across all activities and users is shown.

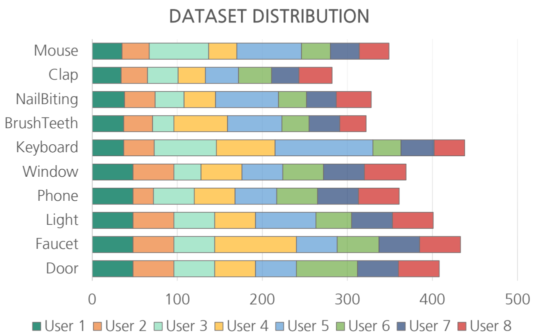

Figure 3: Distribution of the dataset across all 10 activities. In total, more than 3 hours of recordings compose the dataset, executed by 8 subjects, where each subject performed an average of 46 ± 6.0 repetitions per activity.

3.2.2 Feature Selection Method

The attempt to select the best set of features, was conducted by two different methods which were introduced in Section 2.2. In Figure 4, the behaviour of both methods for the first 7 iterations is shown. With 265 features we can reach an accuracy of $70 \pm 8.4\%$. BE eliminates 6 features, which leads to an end value of $72 \pm 6.9\%$ and a set of 259 best features. With only 7 best features, FS is able to reach $80 \pm 8.4\%$, which is considerably better than the result of BE. With FS, a final accuracy of $85 \pm 8.1\%$ was achieved, after 21 iterations, corresponding to a set of 21 best features.

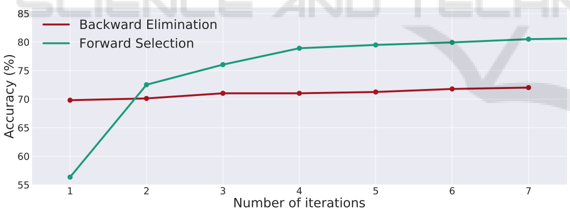

Figure 4: Evolution of accuracy for Forward Selection and Backward Elimination. The x axis is the number of iterations performed to choose or eliminate another feature. Both processes stop when accuracy stops improving.

For each iteration of FS, the feature chosen is the most relevant for the discrimination process. In Figure 5, an horizon plot shows the values of the first 5 features, for 3 repetitions of each activity.

From the horizon plot, the contribution of each feature to the recognition is clear, which demonstrates the value of selection methods. These first 5 features are *y_max*, *z_standard_deviation*, *sound_spectral_slope*, *y_spectral_variance* and *zgyr_spectral_kurtosis*. These first features belong already to three different sensors: accelerometer, microphone and gyroscope, which reinforces the idea of combining sound with inertial sensors for

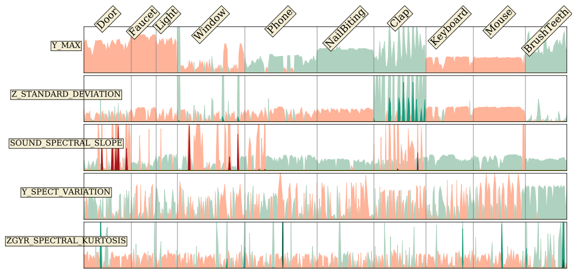

Figure 5: Horizon Plot for most relevant features ordered from top to bottom. The x axis contains 3 consecutive repetitions for each activity, while the y axis contains the values of the features. The different shades of green represent the positive values, while the negative values are in orange and red.

activity recognition. In fact, with Forward Selection, 21 features were selected, which came from all sensors considered (8 from accelerometer, 7 from the gyroscope, 3 from the magnetometer and 2 from the microphone) and all domains (6 statistical, 5 temporal and 10 spectral).

3.2.3 Feature Selection Criteria

The overall result achieved with FS presents a standard deviation of 8.1%. We decided to try another criteria to choose the best feature. Instead of using the arithmetic mean of all activities (overall accuracy), we combined the individual accuracy of each activity through the geometric mean. In Figure 6 we notice that both criteria behave similar. However in Figure 7 we see that the geometric mean leads to a lower standard deviation, based on choosing different features.

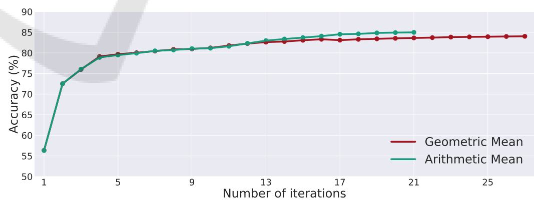

Figure 6: Evolution of accuracy through FS, using the criteria Arithmetic Mean and Geometric Mean. The x axis is the number of features added to the set. The first two features were the same, which explains the equal values.

The final result achieved with the geometric mean is $84 \pm 4.8\%$ which is more reliable than the previous result ($85 \pm 8.1\%$). With the geometric mean, the set of best features has a dimension of 27, still they belong to all sensors and domains, just as in the previous case.

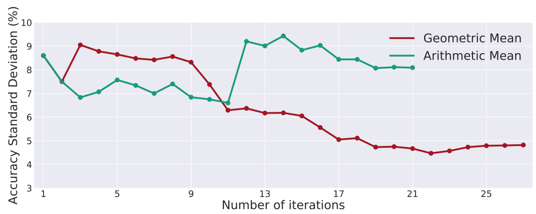

Figure 7: Standard deviation in each iteration of FS with Arithmetic Mean and Geometric Mean.

3.2.4 Number of States of Hidden Markov Models

When dealing with Markov models, one important parameter is the number of states. Usually a fixed number is given to all models (Gaikwad and Narawade, 2012). In this framework, we decided to choose the number for each HMM based on a clustering algorithm applied to the training data. The clustering algorithm used was hdbscan (McInnes et al., 2017). To validate N-based clustering, we compared it to using a fixed number of states (N=4). In Figure 8, both methods were applied with Forward Selection. The use of 4 states for all activities achieves a final accuracy of $83 \pm 5.9\%$ with 17 features, while varying the number of states is able to reach an accuracy of $84 \pm 4.8\%$ with 10 more features.

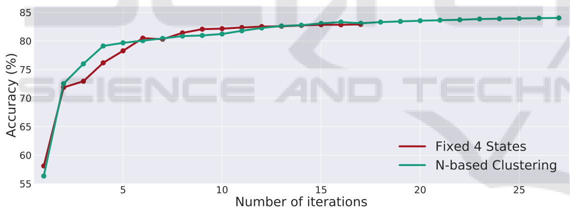

Figure 8: Evolution of accuracy with Forward Selection for a fixed number of states (Fixed 4 States) and for N-based Clustering. The x axis is the number of iterations performed, one feature is added to the set.

Despite the proximity of both results, the variation of the number of states allows to have a better perception on how many states should exist. Furthermore, N-based clustering showed an interpersonal invariability, within each activity. This process is only implemented in the training phase and therefore it does not jeopardize the time complexity of a real time application.

3.2.5 Analysis of Activities

Movements associated to the same object, like opening a door and closing a door, were acquired separately, but they were adressed as one activity, since the movement performed is similar. In Table 1, the accuracy for each activity, for the best result, is presented.

Activities *Door* and *Faucet* present a lower accuracy, while activities *NailBiting*, *Clap* and *BrushTeeth* have the best results.

Table 1: Final accuracy for each activity (%).

Door	*Faucet*	*Light*	*Window*	*Phone*
69 ± 17	70 ± 20	74 ± 22	85 ± 14	94 ± 11
NailBiting	*Clap*	*Keyboard*	*Mouse*	*BrushTeeth*
98 ± 3.9	98 ± 2.0	80 ± 25	89 ± 16	97 ± 4.0

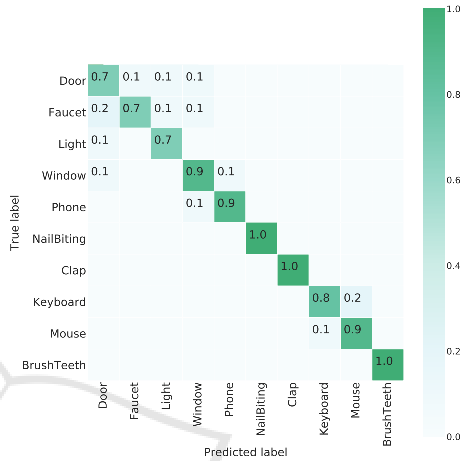

Figure 9: Confusion Matrix for all activities. This matrix is normalized and the values for the diagonal are presented on Table 1.

The confusion matrix in In Figure 9, show which are the activities more easily mistaken. The confusion observed can be explained by the similarity between movements, which is a good indicator of the classifier's performance: a *Keyboard* sample can be misplaced by *Mouse*, but not by any other activity.

3.2.6 New Activities

Up until now, we have shown the framework's ability to discriminate between ten short detailed activities. In fact, this framework was built specifically for them. This new experiment consisted on applying, from scratch, the same framework to a new set of activities. This new dataset was acquired exactly in the same way as the first, but the activities executed are different. Since this study was very brief, the dataset is smaller than the previous one. The activities considered are: *EatHand* - to take food into the mouth; *Writing* - to write on a notebook; *ReadBook* - to flip the pages of a book.

Once more, the set of best features was retrieved by Forward Selection. With 8 features, an accuracy of $94 \pm 1.9\%$ was achieved. Since only three activities were being classified, the challenge is easier when compared to the previous, with ten activities. Nevertheless, the high accuracy achieved sustains the adaptability of the framework developed to new activities, becoming a personalized tool for each user.

3.3 Online Activity Recognition

In online recognition, a continuous data stream is submitted for prediction. In this case, one testing sample could contain one activity, an empty sequence or both cases, which leads to two challenges: how to distinguish activities from empty periods; how to discriminate activities when we only see part of the activity.

3.3.1 Composition of the Online Dataset

To have an online activity monitoring simulation, three of the eight initial users performed all these continuous data streams: *Online 1* is composed by all activities except *BrushTeeth*; *Online 2* contains the activities *Door*, *Light*, *Window*, *NailBiting* and *Clap*; *Online 3* simulates the whole process of brushing teeth, which includes approximately 1 minute of *BrushTeeth* and several repetitions of *Faucet*. Besides containing the activities, the continuous data also presents *Empty* periods. In Figure 10, we can see the representation of each activity in each acquisition. The *Empty* class, represented by the walking man, is considerably predominant in all acquisitions, which highlights the irregularity of the dataset. The approximation to real life conditions, turns the activities into anomalies, sporadically occurring throughout the continuous signal, which is described as Activity Spotting (Junker et al., 2010).

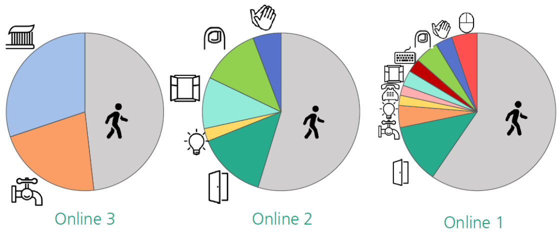

Figure 10: Dataset distribution for the continuous acquisitions *Online 1*, *Online 2* and *Online 3*.

3.3.2 Continuous Data Segmentation

If this solution is applied in an online activity recognition system, the signal is classified in real time through the use of the previously trained HMMs. The continuous stream is first segmented into 250ms windows (without overlap), from where the 27 features are extracted. However, it is still a multiple activity stream, which needs to be segmented into smaller samples for classification. Since some activities, such as *Light*, could have less than 3 seconds duration, we decided to segment into 2 second samples, with an overlap of 1.75 seconds. In each iteration, one sample is submitted for prediction. The next evaluation occurs 250 ms after the first and it evaluates the next 2 seconds.

3.3.3 Classifier Calibration

To distinguish between activity samples and empty samples, we calibrate the classifier based on the results of offline recognition. In the offline mode, the highest probability is associated to the most probable model, and can be saved as its result. Then, using a percentile, we define a threshold for each activity. The percentile 70 was chosen as our threshold, after testing several percentiles. A lower percentile would cause many false positives, while a higher percentile would miss some of the activities.

It is not necessary to show all labels predicted (every 250 ms), since it would be repetitive. Whenever a label is not *Empty*, while the previous label was, it means that an activity has started. Until the appearance of another *Empty* label, the whole period will be considered the same activity, which will correspond to the most repeated activity on the list of predicted labels.

3.3.4 Performance Evaluation

The three users that performed the online experiment, they achieved an offline accuracy of 81%, 79% and 85%. These values influence the results achieved in the online recognition. Given the unusual nature of this task, we divide the performance analysis in three: in Activity Spotting we are only interested in evaluating the ability of detecting activities, just like anomalies in the signal; in Activity True Prediction we consider only what happens inside activities, if they are well predicted or not; finally we analyze the results for each activity individually.

Regarding **Activity Spotting**, two main tests using F_1*score* were performed:

- *Empty Score* - The framework's ability to correctly predict *Empty* periods.
- *Anomaly Detector* - A binary classification (*Activity*/ *Empty*).

In Table 2, the results regarding Activity Spotting are presented. *Empty Score* is low, due to the interpretation of *Empty* periods as activities. Some adjustments should be implemented to improve its value. Nevertheless, our solution is able to spot activities within the continuous signal with an average accuracy for *Anomaly Detector* of $81 \pm 7.2\%$.

Table 2: Results of *Anomaly Detector* and *Empty Score*. The metric used is F_1*score*.

Anomaly Detector (%)	81 ± 7.2
Empty Score (%)	55 ± 7.5

In the analysis of **Activity True Prediction**, the correct prediction of the activity, when compared to the ground truth, is analyzed. We recur to three different performance metrics, to retrieve meaningful information from the results:

- *Substitution* (van Kasteren et al., 2011) - Percentage of activities which were classified as other activities.
- *Activity True Detector (Tapia et al., 2004)* - The average percentage of correct activity inside a true activity.
- *Top 2 Activity* - If the ground truth label was part of the top 2 most frequent of the predicted labels list.

The results in Table 3 are indicators of the classifier's proximity to a correct prediction. The value of *Substitution* ($20 \pm 8.2\%$) sustains a low misclassification of activities, while *Activity True Detector* indicates that, in average, $70 \pm 18\%$ of the activity is correctly predicted. Moreover, the high result for *Top 2 Activity* ($90 \pm 16\%$) indicates a high proximity of the classification to the ground truth.

Table 3: Results of *Activity True Detector*, *Top 2 Activity* and error *Substitution*.

Substitution (%)	20 ± 8.2
Activity True Detector (%)	71 ± 18
Top 2 Activity (%)	90 ± 16

3.3.5 Analysis of Activities

Further on, the activities are analyzed individually in terms of their precision, recall and F_1*score*. In this analysis, *Empty* moments are also considered.

In Table 4, the difference between some activities is notable. The activities *Nailbiting*, *Clap* and *BrushTeeth* achieved scores higher than 95%. This value is understandable based on the high results of offline accuracy. Besides *Light*, all activities presents an F_1*score* higher than 50%. The *Empty* periods were rightly classified if the user had is arm down, as in walking, which explains the high precision ($92 \pm 0.5\%$). The instant the user starts to raise its arm, the classifier identifies that movement as *Light*, resulting in a low precision and high recall for this activity. Furthermore, *Empty* periods also contained having the hand on the table, which is similar to *Mouse* and *Keyboard*. We also considered as *Empty*, predicted activities with less than 1 second. This process was helpful, but it also reduced the Recall of *Faucet* and *Door*, since this activities are often only partially classified. The overall results of precision, recall and F_1*score* are satisfying for further experiments. Still, some improvements can be performed in terms of Activity Spotting and Activity True Prediction.

Table 4: Results for each activity in Online Recognition.

	Precision (%)	*Recall* (%)	F_1*score* (%)
NailBiting	93 ± 9.6	99 ± 1.9	95 ± 5.7
Faucet	69 ± 40	48 ± 15	51 ± 24
Door	76 ± 11	46 ± 30	54 ± 28
Light	7.1 ± 5.5	97 ± 5.0	13 ± 9.4
Phone	42 ± 13	73 ± 38	53 ± 20
Window	62 ± 23	96 ± 3.9	71 ± 19
Clap	98 ± 3.7	96 ± 3.3	97 ± 2.7
Keyboard	76 ± 24	73 ± 24	73 ± 21
Mouse	66 ± 29	79 ± 37	72 ± 33
BrushTeeth	100 ± 0.0	100 ± 0.0	100 ± 0.0
Empty	92 ± 0.5	41 ± 7.2	55 ± 7.5
Total	71 ± 27	77 ± 23	74 ± 26

4 CONCLUSIONS

The major contribution of the present work is the ability to recognize short detailed activities, both offline and online. A dataset with 10 detailed activities was acquired for both contexts and an adaptive framework was created with multiple HMMs (one per each activity). The Forward Selection method was implemented to reduce the set of features. Even though this method is already used in previous studies, we applied a new criteria, which led to a lower standard deviation. Another contribution, is the N-based clustering approach to find the number of states in each HMM. The final contribution was the classifier's calibration for the online recognition, based on the offline results.

Despite the variability and similarity between the dataset's activities, we still achieved a final accuracy of $84 \pm 4.8\%$. This result was achieved with 27 features selected through Forward Selection, which came from different domains (statistical, temporal and spectral) and sensors (accelerometer, gyroscope, magnetometer and microphone), sustaining the important contribution of different sensors in activity recognition systems.

Furthermore, the solution was trained and tested with a totally new dataset, where it substantiate its ability to adjust from scratch to the data, to choose a new set of features and also reach great results in accuracy. Even though it is necessary to test with more activities, we are confident about the ability of this framework to adapt to any given activity, becoming a personalized tool for each user.

In online recognition, the solution underwent preliminary tests, using three of the eight initial users. The classifier was calibrated, by the percentile 70 of offline results, to allow the distinction between activities and *Empty* periods. The classifier's performance that the framework is able to detect activities within a continuous stream with an F_1score of $74 \pm 26\%$. To improve the classification inside true activities, the metric *Top 2 Activity* could be used as an additional criteria, for the prediction phase. To improve Activity Spotting, *Light* could serve as a trigger to identify the beginning of an activity, which was then combined to a binary classifier to perform *Empty/Activity* distinction.

The purpose of this work was to reach further than current recognition systems, and observe activities usually ignored or classified as *Walking (Door)* or *Sitting (Mouse and Keyboard)*. Moreover, the recognition of *BrushTeeth* could indicate if the time spent on this activity was adequate or if it was too short. Beyond that, the recognition of *NailBiting* could be helpful in the control of this impulse.

In the future, the dataset should be increased to more users. Also, other activities should be tested. Considering the application in a real live situation, our framework could be integrated into a wearable sensing device with an android interface.

ACKNOWLEDGEMENTS

This work was supported by North Portugal Regional Operational Programme (NORTE 2020), Portugal 2020 and the European Regional Development Fund (ERDF) from European Union through the project Symbiotic technology for societal efficiency gains: Deus ex Machina (DEM) [NORTE-01-0145-FEDER-000026]

REFERENCES

B, Y. F., Chang, C. K., and Chang, H. (2016). Inclusive Smart Cities and Digital Health. 9677:148–158.

Cardoso, H. and Mendes-Moreira, J. (2016). Improving Human Activity Classification through Online Semi-Supervised Learning. pages 1–12.

Cilla, R., Patricio, M. A., García, J., Berlanga, A., and Molina, J. M. (2009). Recognizing human activities from sensors using hidden markov models constructed by feature selection techniques. *Algorithms*, 2(1):282–300.

Figueira, C. (2016). Body Location Independent Activity Monitoring. Master's thesis.

Gaikwad, K. and Narawade, V. (2012). HMM Classifier for Human Activity Recognition. *Computer Science and Engineering*, 2(4):27–36.

Junker, H., Amft, O., Lukowicz, P., and Tröster, G. (2010). Gesture spotting with body-worn inertial sensors to detect user activities. 41(2008):2010–2024.

Kabir, M. H., Hoque, M. R., Thapa, K., and Yang, S.-H. (2016). Two-layer hidden markov model for human activity recognition in home environments. *International Journal of Distributed Sensor Networks*, 12(1):4560365.

Karaman, S., Benois-Pineau, J., Dovgalecs, V., Mégret, R., Pinquier, J., André-Obrecht, R., Gaëstel, Y., and Dartigues, J. F. (2014). Hierarchical Hidden Markov Model in detecting activities of daily living in wearable videos for studies of dementia. *Multimedia Tools and Applications*, 69(3):743–771.

Kreil, M., Sick, B., and Lukowicz, P. (2014). Dealing with human variability in motion based, wearable activity recognition. *2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PERCOM WORKSHOPS 2014*, pages 36–40.

Lara, O. D. and Labrador, M. a. (2012). A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys & Tutorials*, pages 1–18.

Leonardo, R. M. P. (2018). Contextual information based on pervasive sound analysis. Master's thesis.

Li, Z., Wei, Z., Yue, Y., Wang, H., Jia, W., Burke, L. E., Baranowski, T., and Sun, M. (2015). An Adaptive Hidden Markov Model for Activity Recognition Based on a Wearable Multi-Sensor Device. *Journal of Medical Systems*, 39(5).

Machado, I. P., Gomes, A. L., Gamboa, H., Paixão, V., and Costa, R. M. (2015). Human activity data discovery from triaxial accelerometer sensor: Non-supervised learning sensitivity to feature extraction parametrization. *Information Processing & Management*, 51(2):204–214.

McInnes, L., Healy, J., and Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11):205.

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.

Tapia, E. M., Intille, S. S., and Larson, K. (2004). Activity recognisation in Home Using Simple state changing sensors. *Pervasive Computing*, 3001:158–175.

van Kasteren, T. L. M., Alemdar, H., and Ersoy, C. (2011). Effective performance metrics for evaluating activity recognition methods. *Arcs 2011*, pages 1–34.