Università di Genova
DIPARTIMENTO INFORMATICA, BIOINGEGNERIA, ROBOTICA E INGEGNERIA DEI SISTEMI

Computer Science Workshop
PhD program in Computer Science and Systems Engineering

Improving Activity Recognition While Reducing Misclassification of Unknown Activities

Andrea Fasciglione, Maurizio Leotta, Alessandro Verri

Catching Traditional Activity Recognition Systems Weaknesses in Real-World Scenarios

Introduction

Wearable devices are nowadays widely studied for purposes like fall detection or for tracking the overall level of daily physical activity on both healthy people and patients suffering from different diseases.

Moreover, in the last few years, many studies have been proposed regarding activity recognition using wearable devices. This would be useful for specialists for monitoring the activities performed by subjects and their real physical abilities in a noncontrolled scenario. Different works are present in the literature dealing with the activity recognition, but most are limited to recognizing a predefined set of activities [1].

Most of the literature regarding activity recognition does not face the problem of classifying activities of interest performed among activities that are not, i.e., *unknown* activities for the classifier. We have proposed a novel method, aimed at improving the activity recognition accuracy in presence of *unknown* activities (i.e., in real-world settings). The approach is based on an ensemble of different classifiers, a filtering procedure, and a final voting mechanism.

Approach

Here we will describe in detail the proposed approach that allowed to reduce the number of misclassifications of unknown activities and to increase the overall accuracy of the classification of the activities of interest. The approach we propose in our paper [2] is an improvement of the baseline approach we presented in our previous paper [3]. In the following, we will briefly describe the *baseline approach* and the employed *dataset.* Moreover, we will describe the multi-classifier extension and the two additional used techniques: (1) detecting and removing outliers and (2) using an ensemble method with a voting system.

Dataset and Previous Work
In our previous paper [3], we described a novel dataset and an approach for recognizing a set of ADL. Concerning the dataset employed, it contains the recordings of 17 ADL, performed by 8 different healthy subjects, while wearing wearable devices on their body. In our analyses, we have only considered accelerometer data obtained from wrist device. In addition to the 17 ADL, the original raw data published online in conjunction with our previous work also contain unlabeled data recorded between the execution of the various activities: we relabeled all this data with the generic label “other”.

To resume the baseline approach from our previous work, the first step concerns extracting the feature vectors and the associated labels, by using a sliding window approach, considering only the accelerometer data. During the features extraction, the sliding window passes over the data and, for each axis (X, Y, Z), different measures are extracted from the data contained in the window obtaining a total of 24 features.
After extracting the features, and split our data into training and test set (75% and 25% of data, respectively), we trained an SVM model with the data and then test its accuracy to discern the 17 ADL.

Limitations. The focus of the baseline approach was limited to recognize a predefined set of activities from a test set containing only well-defined activities. When we tried to adopt such an approach to real-world data, we faced a more challenging issue: the activities of interest are mixed in-between a plethora of not well-defined other activities performed during the day. For this reason, in this study, we focus our attention on discerning our activities of interest also from "other" data.

Using Multiple Classifiers
Analyzing the results obtained by the Baseline Approach (SVM-based), we have reasoned about the fact that every Machine Learning algorithm can behave differently on various portions of the same dataset. Distinct classifiers can have strengths and weaknesses, and regarding weaknesses, the problem leads to obtaining erroneous classifications. We had also in mind the principle of ensemble learning, i.e., combining multiple classifiers to potentially increase the accuracy, efficiency, and robustness w.r.t. the single classifier.

Our approach is based on the idea of ensemble learning and combines the results of different classifiers:

1) a Support Vector Machine model (SVM)
2) a Decision Tree model (DT)
3) a Random Forest model (RF)
4) a k-Nearest Neighbour model (k-NN)
5) a Gaussian Naive Bayes model (GNB)

We have tried to combine algorithms based on different foundations and theories to avoid the pitfalls that a solution based on a single classifier can face. Therefore, starting from an input dataset containing 24 features, we executed five classifiers instead of only one and obtained as output five lists of candidate activities.

Detecting and Removing Outliers
By analyzing the predicted labels obtained from every single classifier, we discovered that often many transients outliers appear in the list of predicted labels. In our case, we would call outliers those few samples that are predicted as label j in the middle of a long list of labels k.

For example, according to predictions, while performing the *walking* activity for a minute, a subject would have started to *brush teeth* for 0.5 only seconds: this kind of behaviour is due to short movements that are wrongly associate by the classifier to another activity (e.g., brush teeth). By considering a minimal reasonable length of an activity we tried to detect these outliers and remove them, by "smoothing" the list of labels that we can obtain from each of the considered algorithms. To filter out the outliers, we have used a method based on a sliding window, consisting of computing the mode over a set of labels.

Voting System
The final step of our approach takes as input the five lists of labels already filtered from outliers and produces a final unique list of labels by relying on a voting mechanism. Since we have decided that each classifier has equal weight, we simply take the majority. For each sample, we set as the final label the label that is present more than three times in the classifiers’ predictions for that sample; if the most present prediction has no majority, we set "other" as the final label for that sample.

Comparison Method
When comparing the accuracy of the proposed method, we will not only compare the obtained results by using the F1 score: we will also consider as measures (1) "other“ data correctly labelled as "other" (TP Other for short) and (2) average amount of data regarding activities of interest that are wrongly labelled as "other" (Avg FP Other for short).

Considered metrics help us to measure the problem of misclassification and to make comparisons among different approaches. Every model that we have created during our tests was subject-dependent: we trained a model on training data of a subject *s* and we tested that model on the test data of the same subject *s*. We then averaged over all eight subject’s data.

Results & Forthcoming Research

In the context of activity recognition and about reducing the amount of misclassified samples when dealing with real-life scenarios and daily recordings, we have proposed a method able to increase compared to a simple SVM-based classifier. The approach, in fact, based on an ensemble method with a voting system and on a technique able to filter outliers has been able not only to increase the amount of "other" data correctly classified, but also to increase the overall accuracy measured as F1 Score and at the same time to reduce the amount of data that were wrongly classified as "other".

With respect to the previous SVM-based baseline approach, our current proposal is able to increase the F1 Score of 5.5%, the amount of TP Other has increased for a total of 11.1% (equivalent to a reduction of 28.5% of misclassified data), and average FP Other has decreased of 15.5%. The observed differences for both F1 Score and TP Other are statistically significant.

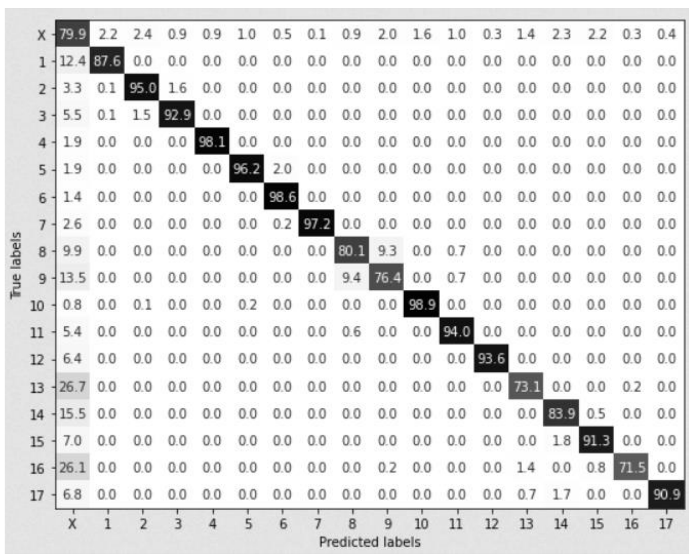

Figure 1:
Confusion matrix obtained after improvements

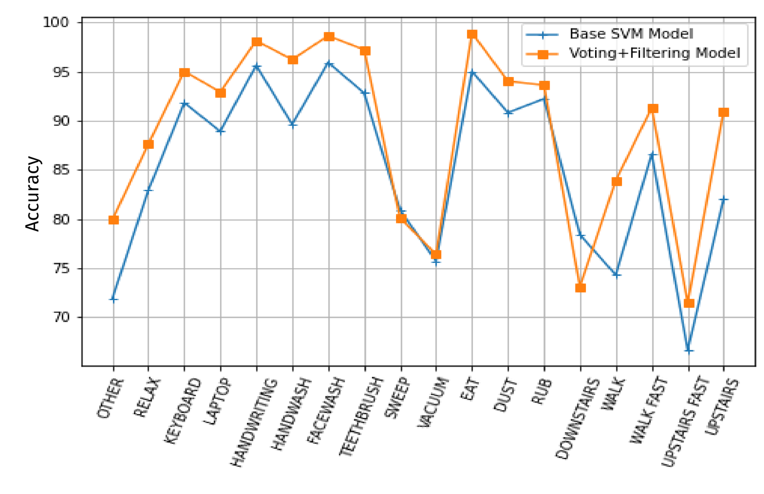

Figure 2:
Improvements of results for each considered ADL from baseline to novel approach

In our future work, we plan to test the capabilities of our proposed approach by using other datasets publicly available online and to compare our results with those achievable with methods present in the literature. Additional tests could be made by varying the combination of machine learning algorithms used in the voting step.

References

[1] S. Zhang, A. V. Rowlands, P. Murray, and T. L. Hurst, “Physical activity classification using the genea wrist-worn accelerometer.” Medicine and science in sports and exercise, vol. 44, no. 4, pp. 742–748, 2012
[2] A. Fasciglione, M. Leotta, and A. Verri, "Improving Activity Recognition while Reducing Misclassification of Unknown Activities”, 2021, Sept., accepted for 30th IEEE Int. Conf. on Enabling Technologies: Infrastructure for Collaborative Enterprises.
[3] M. Leotta, A. Fasciglione, and A. Verri, “Daily living activity recognition using wearable devices: A features-rich dataset and a novel approach,” in Proceedings of 25th International Conference on Pattern Recognition Workshops (ICPR 2021 Workshops), ser. LNCS, A. e. a. Del Bimbo, Ed. Springer, 2021, vol. 12662. [Online]. Available: https://doi.org/10.1007/978-3-030-68790-8_15

Acknowledgements

This work is an activity of the Software Engineering for Healthcare Lab, a joint laboratory between the University of Genova and Janssen Pharmaceuticals (Johnson & Johnson group). We want to show our gratitude to Massimo Raineri, Giacomo Ricca, and Salvatore Saporito of Janssen Pharmaceuticals for the support provided.

Wanna use our dataset? ;)

You can find it on Harvard Dataverse by looking for «*Daily Living Activity Recognition Using Wearable Devices: A Features-rich Dataset and a Novel Approach*» or at the following links:

- https://sepl.dibris.unige.it/2020-DailyActivityDataset.php
- https://doi.org/10.7910/DVN/G23QTS

.... And please, cite us! ☺

CONTACTS

Andrea Fasciglione
andrea.fasciglione@edu.unige.it

Maurizio Leotta
maurizio.leotta@unige.it

Alessandro Verri
alessandro.verri@unige.it

CSW