Sensor and Feature Selection for an Emergency First Responders Activity Recognition System

Sebastian Scheurer*,‡ Salvatore Tedesco† Kenneth N. Brown* Brendan O'Flynn†

*Insight Centre for Data Analytics
Dep. of Computer Science
University College Cork
Cork, Ireland

†Tyndall National Institute
University College Cork
Cork, Ireland

***Abstract*—Human activity recognition (HAR) has a wide range of applications, such as monitoring ambulatory patients' recovery, workers for harmful movement patterns, or elderly populations for falls. These systems often operate in an environment where battery lifespan, power consumption, and hence computational complexity, are of prime concern. This work explores three methods for reducing the dimensionality of a HAR problem in the context of an emergency first responders monitoring system. We empirically estimate the accuracy of k-Nearest Neighbours, Support Vector Machines, and Gradient Boosted Trees when using different combinations of (A)ccelerometer, (G)yroscope and (P)ressure sensors. We then apply Principal Component Analysis for dimensionality reduction, and the Kruskal-Wallis test for feature selection. Our results show that the best combination is that which includes all three sensors (MAE: 3.6%), followed by the A/G (MAE: 3.7%), and the A/P combination (MAE 4.3%): the same as that when using the accelerometer alone. Moreover, our results show that the Kruskal-Wallis test can be used to discard up to 50% of the features, and yet improve the performance of classification algorithms.**

I. Introduction

Human activity recognition (HAR) systems have a wide range of applications, such as monitoring ambulatory patients, workers for movement patterns associated with repetitive strain injury (RSI), or elderly populations for dangerous falls [1]. A substantial body of this work uses data obtained from wearable inertial measurement units (IMUs), e.g., accelerometers, gyroscopes, and magnetometers, to train machine learning classification algorithms, such as Support Vector Machines (SVM) or k-Nearest Neighbours (kNN), to discriminate among the activities of interest [2]. In this framework, the classification algorithm typically operates on a set of features that have been extracted along a sliding window from a data-set of many labelled trials of the activities of interest.

HAR systems often operate in constrained environments where battery lifespan and power consumption are a prime concern. In this context, it is important to reduce computational complexity—and hence power consumption—where possible. In this paper, we explore three different approaches for reducing the dimensionality, and hence computational complexity, of the HAR inference problem: sensor selection, dimensionality reduction, and feature selection.

We estimate the accuracy of kNN, SVM, and Gradient Boosted Trees (GBT) [3, ch. 10] when using 7 possible sensor combinations. The best combination is then used in two more experiments, in which unsupervised dimensionality reduction via Principal Component Analysis (PCA), and supervised feature selection via the Kruskal-Wallis (K-W) test are applied to simplify the HAR inference problem. To our knowledge, this is the first application of the K-W test for feature selection in a HAR context. Using these data we show that a subject-independent Mean Absolute Error (MAE) as low as 4.3% can be achieved with just one sensor (or 40 features), and that a simple feature selection, such as the K-W test, can be used to discard 50% of the features and yet improve performance.

II. Methods

In [4], we compared kNN, SVM and GBT using data describing 17 human activities in the context of an emergency first responders monitoring system [5] developed as part of the European SAFESENS (Sensor Technologies for Enhanced Safety and Security of Buildings and its Occupants) project. The data-set consisted of 65 features which had been extracted from acceleration and angular velocity signals collected by a wearable IMU that was attached to the chest and equipped with barometer, 3D accelerometer, and 3D gyroscope.

For this work, in addition to the accelerometer and gyroscope features described in [4], the data-set was enriched with features extracted from the pressure signal. Because pressure is directly related to altitude, the collected signal is bound to separate some activities with high accuracy at the place where the data were collected—but the generalisation fails if the sensor is moved to a different altitude. In order to avoid potential biases, we first calculated the mean pressure over all samples of the "standing" position in the data-set and subtracted it from the pressure signal. Then, the pressure signal was subjected to the same pre-processing and feature extraction—with exception of the pairwise correlation features—procedure as the raw gyroscope signal in [4].

This publication has emanated from research supported by a research grant from Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289, the European Regional Development Fund under grant number 13/RC/2077-CONNECT, and the European funded project SAFESENS under the ENIAC program in association with Enterprise Ireland (IR20140024).

‡Corresponding author email: sebastian.scheurer@insight-centre.org.

Using these data, we conducted three experiments. In each, we estimated the generalisation error for the three classifiers using the algorithm parameters and estimation procedure described in [4]. In the first experiment we compared the predictive value of different sensor combinations by evaluating all possible combinations via leave-one-subject-out cross-validation (LOSO CV). The best combination was then used to run the second and third experiment, each of which evaluated (via LOSO CV) a different method for reducing the dimensionality of the HAR inference problem. In one of them we applied PCA for dimensionality reduction, retaining only the number of Principal Components (PC) required to explain 10%, 30%, 50%, 70%, and 90% of the total variance. In the other, the K-W test was applied for feature selection.

The K-W test is a non-parametric statistical test against the null hypothesis that the tested samples were generated by the same distribution. We leveraged the K-W test for supervised feature selection by applying it to each of the features—which had been partitioned into 17 disjoint samples according to the target class for this purpose—in turn. Then, the features were ranked according to the K-W test statistic, and the top 10th, 30th, 50th, 70th, and 90th percentile was retained as inputs for the inference algorithm.

III. Results & Discussion

Tables I–III each list the MAE and its standard error (SE)—calculated across the 11 folds of the LOSO CV and subsequently averaged over the 17 target classes—as well as the standard deviation (SD) among the target classes from one of the three experiments. The MAE estimates (with precision SE) the generalisation error we can expect on data from unseen individuals, while the SD serves as a measure of how much the MAE varies among the 17 target classes. Each of the entries in these tables summarises a set of class-wise MAEs. Three examples of these are shown in Table IV which lists class-wise MAEs that resulted when using the K-W test to select feature subsets of varying sizes as input for the GBT algorithm.

The results for each combination of the (A)ccelerometer, (G)yroscope, and (P)ressure sensor are given in Table I. The results from the PCA experiments are shown in Table II, where the cumulative percentage of variance explained is given by the first column and the corresponding number of components (n) by the second. The results from our experiments with K-W feature selection are given in Table III, where the percentile that is being retained is given by the first column and the corresponding number of features (n) by the second.

According to the results shown in Table I, the best combination is indeed that which includes all three sensors (72 features), where the best performance (MAE: $3.6\% \pm 0.9\%$) was achieved with the GBT algorithm. However, comparable (MAE: $4.3\% \pm 1\%$) performance can be obtained using only one sensor, namely the accelerometer; thus retaining 40 of the 72 features and reducing the dimensionality by 44%. In contrast, neither the gyroscope (25 features) nor the pressure sensor (7 features) appears to be very useful on its own. A particularly bad choice for a single-sensor HAR system is the pressure sensor, especially considering that a dummy model, which makes predictions solely based on the class proportions, results in an MAE of 10.3%. The best two-sensor combination is clearly that of accelerometer and gyroscope ($A\,G$), whose performance is very close to that of the $A\,G\,P$ combination, while using only 65 (86%) of the 72 features. Furthermore, while differences among classifiers that are based on the same sensor combination are well below any of their underlying estimates' precision, there is a visible gap separating combinations that include the accelerometer from those that do not.

The results from our PCA experiments in Table II show that it can maintain an MAE below 5%, while reducing the dimensionality of the three-sensor ($A\,G\,P$) inference problem beyond what is feasible by simply discarding sensors. An average MAE of 4.7% is obtained with only 9 PCs (explaining 10% of the total variance), but even retaining as many as 40 PCs (explaining 90% of the total variance) the performance does not approach that of the $A\,G\,P$, or even the $A\,G$ (65 features), combination in Table I.

As the K-W feature selection experiments in Table III show, we can improve, albeit only marginally, on the best combination from Table I if we retain as few as half of the features, thereby halving the inference problem's dimensionality from 72 to 36 features and—assuming the algorithm's

TABLE I
MAE (± SE) AND SD (ALL IN %) FOR ALL SENSOR COMBINATIONS

	GBT MAE	GBT SD	SVM MAE	SVM SD	kNN MAE	kNN SD
A G P	3.6 ± 0.9	2.3	3.8 ± 0.8	2.0	4.2 ± 0.8	2.6
A G	3.7 ± 0.9	2.2	3.9 ± 0.8	2.1	4.2 ± 0.8	2.5
A	4.3 ± 0.9	2.3	4.6 ± 0.9	2.2	5.0 ± 0.9	2.5
A P	4.3 ± 1.0	2.3	4.6 ± 1.0	2.2	5.0 ± 1.0	2.6
G P	5.8 ± 1.0	3.3	6.1 ± 1.0	2.7	6.4 ± 1.1	3.1
G	6.1 ± 1.0	2.9	6.5 ± 1.0	2.5	6.5 ± 1.0	2.9
P	10.4 ± 1.8	6.0	10.5 ± 1.2	4.9	10.5 ± 1.6	5.2

TABLE II
MAE (± SE) AND SD (ALL IN %) WHEN USING PCA

%	n	GBT MAE	GBT SD	SVM MAE	SVM SD	kNN MAE	kNN SD
10	1	9.4 ± 1.1	3.9	9.5 ± 1.1	3.9	9.4 ± 1.1	3.8
30	3	8.0 ± 1.2	3.2	8.4 ± 1.1	3.2	8.1 ± 1.1	3.1
50	9	4.7 ± 0.9	2.7	5.0 ± 0.9	2.5	4.9 ± 1.0	2.6
70	21	4.5 ± 0.8	2.7	4.5 ± 0.8	2.5	4.7 ± 0.9	2.7
90	40	4.4 ± 0.8	2.6	4.1 ± 0.8	2.3	4.4 ± 0.8	2.6

TABLE III
MAE (± SE) AND SD (ALL IN %) WITH K-W FEATURE SELECTION

%	n	GBT MAE	GBT SD	SVM MAE	SVM SD	kNN MAE	kNN SD
10	7	6.6 ± 1.0	2.8	7.1 ± 1.0	2.9	6.6 ± 1.0	2.8
30	21	4.8 ± 1.1	2.5	4.8 ± 0.9	2.6	5.2 ± 1.2	2.7
50	36	3.5 ± 0.9	2.1	3.9 ± 0.9	2.2	3.8 ± 0.8	2.2
70	50	3.5 ± 0.9	2.2	3.8 ± 0.9	2.2	3.7 ± 0.8	2.2
90	64	3.5 ± 0.8	2.3	3.7 ± 0.8	2.0	3.9 ± 0.8	2.4

TABLE IV
MAE (± SE) FOR GBT WITH K-W, RETAINING DIFFERENT PERCENTILES

Percentile	30	50	70
All 4s	6.0 ± 2.4	6.0 ± 2.4	6.0 ± 2.4
Crawl H & K	2.3 ± 0.9	1.7 ± 0.5	1.6 ± 0.5
Crawl M	2.7 ± 0.8	2.0 ± 0.5	1.8 ± 0.5
Crouch	5.2 ± 0.8	5.1 ± 0.8	5.2 ± 0.8
Duck walk	1.6 ± 0.5	1.5 ± 0.5	1.2 ± 0.5
Fall	0.8 ± 0.2	0.8 ± 0.2	0.7 ± 0.2
Jump off	1.9 ± 0.5	1.7 ± 0.5	1.7 ± 0.5
Jump on	1.9 ± 0.4	1.7 ± 0.5	1.6 ± 0.5
Lie	4.8 ± 2.4	4.3 ± 2.4	4.3 ± 2.4
Run	7.1 ± 1.6	3.6 ± 1.1	3.3 ± 1.0
Run down	6.8 ± 0.9	3.7 ± 0.8	3.8 ± 0.7
Run up	7.7 ± 1.4	2.1 ± 0.6	2.0 ± 0.5
Sit	7.1 ± 1.3	6.4 ± 1.2	6.5 ± 1.2
Stand	8.6 ± 1.6	8.4 ± 1.4	8.5 ± 1.4
Walk	3.9 ± 0.8	3.0 ± 0.7	3.1 ± 0.7
Walk down	7.8 ± 1.0	4.7 ± 0.7	4.7 ± 0.7
Walk up	5.5 ± 0.7	2.9 ± 0.4	2.8 ± 0.4
Average	4.8 ± 1.1	3.5 ± 0.9	3.5 ± 0.9

time complexity is linear or worse in the number of features—at least halving the run-time. If we decrease the number of features further, we observe deteriorating performance, as expected, for all three algorithms—most notable in the case of kNN—and we might expect that moving in the other direction and increasing the number of features would have the opposite effect, namely to improve performance. However, our data show that this is not necessarily the case. While SVM does indeed improve its performance marginally—starting with an MAE of 3.9% when using 50%, to 3.8% when using 70%, to 3.7% when using 90% of the features—GBT, instead, maintains a stable MAE of 3.5%, regardless if 50%, 70%, or 80% of the features are being retained; and kNN achieves its best performance when using 70% of the features—if passed a larger percentage, its performance begins to deteriorate.

The class-wise MAEs shown in Table IV further illustrate what happens when we increase the percentile of features that is being retained, using the GBT results as an example. What stands out is that the reduction of the average MAE that can be seen when moving from the 30th to the 50th percentile can be attributed mainly to the significantly reduced MAE from the three running (“Run”, “Run down”, and “Run up”), as well as the “Walk down” and “Walk up” activities, and—to a much lesser extent—the “Walk” (horizontally) activity. This means that at least some of the features that are in the 50th, but not in the 30th percentile, are useful for discriminating among these activities. It also means that there is little benefit from using more than 30% of the features for applications, such as fall detection, where fine-grained distinctions like these can be of little practical concern or impact.

We conclude our discussion with a summary of the K-W ranked percentiles illustrated in Fig. 1. The 10th percentile, amounting to 10% of the accelerometer (Acc), and 12% of the gyroscope (Gyro) features, consists of the SD of the x and y, and the inter-quartile range (IQR) of the Acc y axes; as well as the SD of the x and y, and IQR of the Gyro x axes. The 30th

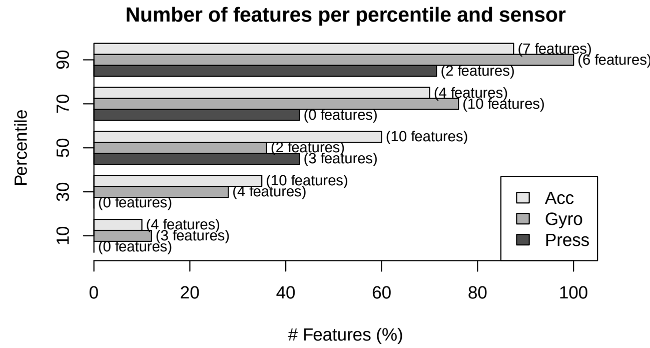

Fig. 1. Number of features per sensor and K-W percentile

percentile adds IQR, SD, and signal magnitude area features amounting to 27% of the Acc, and 18% of the Gyro features not present in the 10th percentile. The 50th percentile contains all types of features that had been extracted, except pairwise correlations, adding 38% of the Acc, 11% of the Gyro, and 68% of the pressure features not present in the 30th percentile. The 70th percentile adds 25% of the Acc, and 63% of the Gyro features not present in the 50th percentile. The 90th percentile adds what are mostly peak power frequency, spectral entropy, and pairwise correlation features, amounting to 58% of the Acc, 100% of the Gyro, and 50% of the pressure features that were not present in the 70th percentile.

IV. CONCLUSION

We showed that the single best sensor (among the three evaluated) for HAR is the accelerometer, resulting in an MAE of $4.3\% \pm 0.9\%$ when used with the GBT algorithm. At the other extreme we found the pressure sensor, which resulted in an MAE of 10.4%, no better than what we would get when merely guessing the proportion of activities (classes) in the data-set. The sensor combination that achieved the best results was that with accelerometer, gyroscope, and pressure, with an MAE of $3.6\% \pm 0.9\%$, closely followed by the accelerometer/gyroscope combination with an MAE of $3.7\% \pm 0.9\%$. Moreover, our results showed that a simple univariate feature selection method such as the Kruskal-Wallis test can be used to reduce the complexity of a HAR inference problem by as much as 50% while not only maintaining, but even improving the performance of HAR inference algorithms.

REFERENCES

[1] O. D. Lara and M. A. Labrador. “A Survey on Human Activity Recognition using Wearable Sensors”. In: *Commun. Surveys Tuts.* 15.3 (2013).

[2] A. Bulling, U. Blanke, and B. Schiele. “A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors”. In: *Comput. Surv.* 46.3 (2014).

[3] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. 2nd ed. Springer, 2009.

[4] S. Scheurer, S. Tedesco, K. N. Brown, and B. O’Flynn. “Human Activity Recognition for Emergency First Responders via Body-Worn Inertial Sensors”. In: *Int. Conf. Wearable & Implant. Body Sensor Netw.* IEEE, 2017.

[5] S. Tedesco, J. Khodjaev, and B. O’Flynn. “A novel first responders location tracking system: Architecture and functional requirements”. In: *Mediterranean Microw. Symp.* IEEE. 2015.