Human Activity Recognition for Emergency First Responders Via Body-Worn Inertial Sensors

Sebastian Scheurer*, Salvatore Tedesco†, Kenneth N. Brown* and Brendan O'Flynn†
*Insight Centre for Data Analytics, Dep. of Computer Science; †Tyndall National Institute
University College Cork, Cork, Ireland
sebastian.scheurer@insight-centre.org

***Abstract*—Every year over 75 000 firefighters are injured and 159 die in the line of duty. Some of these accidents could be averted if first responders had better information about the information on the ground. The SAFESENS project is developing a novel monitoring system for first responders designed to provide response team leaders with timely and reliable information about their firefighters' status during operations, based on data from wireless inertial measurement units. In this paper we investigate if Gradient Boosted Trees (GBT) could be used for recognising 17 activities, selected in consultation with first responders, from inertial data. By arranging these into more general groups we generate three additional classification problems which are used for comparing GBT with k-Nearest Neighbors (kNN) and Support Vector Machines (SVM). The results show that GBT outperforms the other two approaches for three of these four problems, and that it leads to a generalisation error that is more balanced across the target activities than SVM.**

I. Introduction

Every year over 75 000 firefighters are injured and 159 die in the line of duty [1]. Some of these accidents could be averted, and many ameliorated, if firefighters (and other first responders) had better information about the situation on the ground as it unfolds. The SAFESENS (Sensor Technologies for Enhanced Safety and Security of Buildings and its Occupants) project [2] is developing a location-tracking and monitoring system for firefighters and other first responders which will make that information available to them.

The system monitors firefighters via wireless inertial measurement units (WIMU), integrated into the self-contained breathing apparatus' (SCBA) straps. Data are streamed from the WIMU to a smartphone, carried by each firefighter, where an application buffers the data for 10 seconds before transmitting them as one batch to the command & control center (CCC). In the CCC the data are used to show the officers where their firefighters are and timely clues about what they are doing. The system is designed to work reliably in the harsh and unpredictable conditions of emergency situations, and be resilient if pre-deployed infrastructure fails.

In this paper, we investigate the use of Support Vector Machines (SVM), k-Nearest Neighbors (kNN) and Gradient Boosted Trees (GBT), for recognising up to 17 different activities relevant for monitoring first responders during emergency response operations. To our knowledge, this is the first application of GBT to a human activity recognition (HAR) problem. We show that GBT are able to distinguish among these activities, using gyroscope and accelerometer data from a single IMU, with subject-dependent and -independent mean absolute errors of less than 1% and 4%, respectively.

II. Related works

Human activity recognition based on inertial data has been a fruitful line of enquiry for more than a decade [3], [4]. Although much of the work during this time has been dedicated to recognising activities associated with daily living, especially for monitoring the elderly or other populations at risk, there have been attempts to extend HAR to more dynamic activities and environments. Zhang et al. [5], for example, present an indoor location tracking system for emergency first responders, each of whom is required to wear six IMUs, that maintains a model (based on Kalman filters) of the tracked bodies. Williamson et al. [6], instead, use Gaussian Staircase and Partial Least Squares regression to estimate the loads carried by first responders and soldiers from IMU data.

Statistical and machine learning techniques have been used widely in HAR applications. Among the most popular and successful are SVM and kNN, both of which are frequently used as a baseline when evaluating other learning algorithms. GBT [7] is a learning algorithm that combines many simple decision or regression trees—known as *base learners* in this context—into a powerful ensemble through a flexible and widely used technique known as boosting [8]. Although HAR algorithms that use the boosting technique exist [9], [10], the GBT algorithm itself has not been evaluated in this context.

III. Methods

To investigate if kNN, SVM, and GBT could be effective for monitoring firefighters we selected 17 activities in consultation with collaborating firefighters: two types of crawling (on hands & knees, and military style on one's stomach), duck walking, falling, two types of jumping (on and off a chair), three types of running and walking (horizontally, and up and down the stairs) and five static postures: being on one's hands and knees (all fours), standing, sitting, crouching, and lying down (e.g., after falling). These are used to form three additional HAR problems of decreasing complexity. The first of these, the "move-type/lie" problem, consists of seven target classes (activities): Crawling, duck walking, falling, lying down, running, walking, and the static postures. The next problem, the "move-type" problem, differs from the move-type/lie problem in that lying down is used as an additional static posture. The fourth problem is discriminating between falls and non-falls.

A. Experimental design and data acquisition

We recruited 11 volunteers (all male, age: 20–34) via email and word of mouth, each of whom met the firefighter eligibility criteria: aged between 18 and 37, height of at least 166 cm, a body-mass index (BMI) of 20–30, no problems with eyes, ears or teeth, and of healthy and robust physical constitution. Volunteers were invited, one at a time, to our lab in the buildings of the Tyndall National Institute, where we instructed them to perform several supervised trials of each activity.

To simulate some of the constraints imposed by the firefighting gear we asked participants to wear heavy boots, and carry 13 litres of water in a backpack to simulate the weight of the SCBA that firefighters carry during operations. One of the backpack's shoulder straps served to hold the WIMU in place. The WIMU, developed as part of the SAFESENS project, is equipped with sensors for barometric pressure, humidity, temperature (both internal and external), and a tri-axial accelerometer, gyroscope and magnetometer. The sensor data can be transmitted wirelessly (via BLE), or written to an SD card. We recorded our data to the SD card at a sampling rate of 30 Hz. Other materials used were a chair for jumping on and off, a treadmill, and an inflatable mattress for falling and lying down.

To aid with labelling the data, trials were timed by the experimenter, and participants instructed to tap the WIMU before and after each trial. To avoid potential bias in the data, the sequence in which the tasks were performed by each participant was randomised. For each repetition (trial), participants were further instructed to enact a (randomly chosen) variant of the task.

1) Falling: Participants were instructed to stand beside the mattress, then fall onto it, lay still for a moment, get up, and assume the starting position. For each trial, participants were instructed to fall either *forward*, or to the *side*.

2) Jumping: Participants were instructed to stand in front of (or on) the chair, jump onto (or off) it, pause for a moment, and finish by getting back in the starting position.

3) Horizontal walking, crawling, and duck-walking: Participants were instructed to move around the hallway and room in the specified manner for one minute per trial, or until they felt exhausted. For each walking trial they were instructed to walk at either *slow*, *regular*, or *fast* speed.

4) Horizontal running: Represented through two trials: running on the treadmill (to capture running at a steady velocity), and in the hallway (to capture turns and more realistic accelerations). For treadmill running, participants were asked to run at *slow* (7 km/h), *regular* (10 km/h) or *fast* (12 km/h) speed for 90 seconds. For hallway running, participants were instructed to run from one end of the hallway to the other at either *slow*, *regular* or *fast* speed. As many hallway running trials as needed were performed to obtain a total of 90 seconds of data from each participant.

5) Walking, and running, up and down the stairs: Participants were instructed to position themselves at the top or bottom of the staircase, then walk (or run) down (or up) the stairs, stop, and return to the starting position. We performed as many trials as necessary to obtain 90 seconds of data from each participant who were instructed to ascend or descend at either *slow*, *regular*, or *fast* speed.

6) Static: The static tasks, or postures, are standing, sitting, crouching, being on ones hands and knees (all 4s), and lying down. For these tasks participants were instructed to assume the specified position for one minute per trial. All the static tasks, with the exception of crouching and being on all fours, had designated variants. For standing they were to either stand *upright*, *bent* forward with hands on knees, or *leaning* against the wall (with shoulder or back). For sitting the variants were to either sit in normal position on a *chair*, upright on the *floor*, or on the floor with the back or shoulder(s) *leaning* against the wall. Finally, for lying the variants were to lie either face-down on one's *front*, or on the *side*.

B. Data pre-processing

The collected data was prepared as follows. First, the coordinate systems are aligned to conform to the same notion of up and down. Then a median-filter with a window size of 3 is applied to smooth the signal. Next, the (median-filtered) signal is resampled to a constant (its mean) sampling frequency. If the original signal is not sampled with a constant frequency due to potential hardware limitations, then the resampled signal will contain gaps. These gaps were filled by linear interpolation, which is a reasonable approximation in the absence of further information.

In the next step, we replace each of the accelerometer signal's channels (x, y, z) with two derived features, namely its gravity and body component. The accelerometer captures acceleration from two sources: the earth's gravitation, and the movement of the IMU and its wearer. Because it is those movements we are interested in, it is a good idea to separate the two components. This is achieved following the approach described in [11]. The original signal contains no additional information and is not used further. Finally, the signal is segmented into 3-second sliding windows with one second overlap. The resulting data-set consists of 16 621 windows (instances), distributed as follows: on all fours: 5.8%, crouch: 4.4%, sit: 9.8%, stand: 8.6%, lie: 6%, crawl (hands & knees): 6.1%, crawl (mil.): 5%, duck walk: 4%, fall: 0.6%, jump off/on: 0.9% each, run hallway: 2.9%, run treadmill: 9.4%, run up: 5.6%, run down: 5.7%, walk horizontally: 5.9%, walk down: 8.2%, and walk up: 10.2%.

C. Feature extraction

We selected seven time-domain (mean, sample standard deviation, skew, kurtosis, inter-quartile range, signal magnitude area, and pairwise correlations between each sensor's x, y, and z channels) and two frequency-domain (spectral power entropy and peak power frequency) features that have proven useful in previous HAR applications, and extracted these from the gyroscope, and the gravity and body acceleration. Most of these features are statistical (e.g., mean, skew) and follow their usual definitions.

The signal magnitude area (SMA) combines multiple channels into a single measure of the signal's cumulative magnitude relative to the signal's duration. It has proven useful in previous HAR work, particularly for distinguishing between periods of activity and rest [11]. The SMA was extracted from the gyroscope, and the body and gravity acceleration signals.

Both the Spectral Power Entropy (SPE) and the Peak Power Frequency (PPF) are popular frequency-domain features. Both rely on a uniform sampling rate, and an estimate of the power spectral density (PSD). For the SPE this was estimated via the periodogram, for the PPF via Welch's method. The SPE was then calculated following [12].

D. Algorithm tuning

We used the same procedure to separately tune each algorithm for each of the four problems. The procedure has been designed, following current best practices from the HAR and machine learning literature, to minimise the likelihood of setting an algorithm's parameters to a set of values that leads to a large generalisation error. It requires us to define a resampling method for estimating the generalisation error, and an objective function (metric) that measures it. We used the mean absolute error (MAE) as the metric to be minimised and leave-one-subject-out cross-validation (LOSO-CV) as the resampling method for estimating it. We chose the MAE because, unlike performance metrics such as Accuracy, it does not depend on the misclassification costs. We further have to specify the parameters and corresponding set of values that should be searched by the procedure.

The procedure's first step is to randomly split the data-set into a 70% development set, and a 30% validation set. Then, using the development set only, it estimates the train and test error using the chosen resampling method (LOSO-CV) and metric (MAE). These estimates are then used to choose the best parameter settings following a minimax approach: the parameter settings with the lowest upper 95% confidence interval (C.I.) are selected. The selected settings are then used to train the algorithm on the full development set, and calculate its MAE on the validation set. This quantity is then compared against the C.I. to validate the parameter settings for the algorithm.

IV. Results and Discussion

For kNN, the tuning procedure was applied over the values 2, 5, 10, 20, 40, 80, and 160 for k, the size of the neighborhood, and both weighted (by the inverse distance) and unweighted voting was considered. The procedure leads to $k = 2$ and weighted voting regardless of the problem. For SVM, the tuning procedure was applied over the values 10^{-9}, 10^{-6}, 0.001, 1, and 1000 for γ, the coefficient for the Radial Basis Function (RBF) kernel, and 0.01, 1.778, 316, 56 234, and 10^7 for C, the penalty term. The procedure leads to $\gamma = 0.001$ regardless of the problem, and to $C = 10^7$ for the move-type and move-type/lie problem, C = 316.228 for the 17-activity problem, and C = 1.778 for fall detection.

The GBT is an ensemble of trees and as such depends on a tree induction algorithm. We use the Classification and Regression Tree (CART) algorithm [13, chap. 9] for this purpose. Other tree-induction algorithms, namely C4.5 and C5.0, exist, but if trees are shallow (which generally leads to better results in boosted ensembles), their ability to prune trees is unlikely to make a significant impact. Trees are kept simple by imposing a maximum of 16 leafs per tree, 9 features per split, and a minimum of 11 samples per leaf. To further safeguard against overfitting, each tree is restricted to a 30% sample.

For GBT, the tuning procedure was applied over the values 0.02, 0.04, 0.06, 0.08, and 0.1 for α, the learning rate, and over values 50–1600 for M, the number of boosting iterations. The procedure leads to α = 0.02 for the fall detection and the 17-activity, and α = 0.1 for the move-type and move-type/lie problem. The procedure suggests 1600 iterations regardless of the problem but we found that the loss gradient flattens considerably after about 200 iterations. Beyond 600 iterations the improvements are marginal at best and no longer justify the additional computing time. We used 750 iterations in our final GBT.

Using these settings, we estimated each algorithm's subject-dependent/-independent performance via 11-fold cross-validation (CV), and LOSO-CV, respectively. The results for the three multi-class problems are listed in tables I–III. The results for the fourth problem (fall detection) are as follows: the subject-dependent (CV) MAE (and Accuracy) scores for GBT, SVM and kNN are 0.06% (99.96%), 0.05% (99.98%), and 0.02% (99.99%), respectively. The corresponding subject-independent (LOSO) scores are 0.17% (99.86%), 0.12% (99.92%), and 0.1% (99.92%).

Looking at these results in combination, we note that all three classifiers are able to discriminate among the targeted activities accurately, with the class-wise subject-dependent MAE ranging from 0.02% to 5.57%, and the subject-independent MAE from 0.1% to 11.76%. GBT performs the best on the three multi-class problems, followed closely by SVM. However, despite the small difference (0.2–1%) between their average scores, GBT's class-wise MAEs are more balanced across the target classes. For fall detection, the order is reversed and kNN outperforms SVM by about 0.02%, GBT by about 0.05%. However, falls are difficult to simulate, and this may be an artefact of the experimental design.

V. Conclusion

We showed that Gradient Boosted Trees (GBT) can be used to recognise up to 17 human activities for monitoring first responders during operations, with subject-independent and -dependent accuracy of over 73% and 97%, respectively. By merging the 17 activities into more general groups we obtained four classification problems which we used for tuning and benchmarking three popular machine learning algorithms, namely GBT, kNN and SVM. The results show that GBT tends

TABLE I
MAE AND OVERALL ACCURACY (%) FOR THE 17-ACTIVITY PROBLEM

	LOSO GBT	LOSO SVM	LOSO kNN	CV GBT	CV SVM	CV kNN
All 4s	4.04	4.74	5.99	0.26	1.60	2.51
Crawl H & K	1.69	1.43	1.88	0.10	0.27	0.25
Crawl Mil.	2.12	1.74	2.17	0.19	0.39	0.41
Crouch	6.16	5.77	7.22	0.33	2.57	3.42
Duck walk	1.59	1.27	1.04	0.08	0.23	0.11
Fall	0.21	0.25	0.10	0.05	0.10	0.02
Jump off	0.81	1.06	0.84	0.35	0.48	0.26
Jump on	0.60	0.76	0.58	0.23	0.29	0.13
Lie	2.65	3.24	3.82	0.10	1.05	1.24
Run	5.18	5.38	5.58	1.41	1.84	1.26
Run down	4.05	4.21	4.44	1.08	1.52	1.31
Run up	3.24	3.98	4.39	1.00	1.22	0.94
Sit	6.84	8.47	9.74	0.30	4.25	4.35
Stand	8.53	10.45	11.76	0.43	5.55	5.57
Walk	4.78	4.80	6.03	0.71	1.25	1.53
Walk down	6.07	5.17	7.54	1.24	1.72	2.33
Walk up	3.78	3.89	6.01	0.61	0.91	1.17
Mean MAE	3.67	3.92	4.65	0.50	1.48	1.58
Accuracy	73.29	72.46	61.49	97.68	93.23	88.77

TABLE II
MAE AND OVERALL ACCURACY (%) FOR THE MOVE-TYPE/LIE PROBLEM

	LOSO GBT	LOSO SVM	LOSO kNN	CV GBT	CV SVM	CV kNN
Crawl	1.48	1.10	2.00	0.10	0.23	0.33
Duck walk	1.51	0.92	1.04	0.06	0.16	0.11
Fall	0.17	0.17	0.1	0.05	0.07	0.02
Jump	0.75	0.80	0.83	0.29	0.41	0.27
Lie	3.10	3.45	3.82	0.11	1.19	1.24
Run	3.84	5.76	6.77	1.02	2.05	1.98
Static	3.81	4.49	5.34	0.15	1.32	1.71
Walk	4.18	5.89	7.45	0.87	1.85	2.31
Mean MAE	2.35	2.82	3.42	0.33	0.91	1.00
Accuracy	90.99	90.8	86.74	98.9	97.92	96.79

TABLE III
MAE AND OVERALL ACCURACY (%) FOR THE MOVE-TYPE PROBLEM

	LOSO GBT	LOSO SVM	LOSO kNN	CV GBT	CV SVM	CV kNN
Crawl	1.38	1.08	2.00	0.10	0.22	0.33
Duck walk	1.48	0.91	1.04	0.06	0.15	0.11
Fall	0.16	0.17	0.10	0.04	0.07	0.02
Jump	0.75	0.79	0.83	0.29	0.40	0.27
Run	4.00	5.74	6.77	0.99	2.04	1.98
Static	0.71	1.10	1.65	0.04	0.18	0.51
Walk	4.35	5.79	7.45	0.85	1.83	2.31
Mean MAE	1.83	2.23	2.83	0.34	0.70	0.79
Accuracy	94.03	93.86	90.41	99.02	98.62	97.74

to fewer misclassifications, distributed more evenly among the target classes, than kNN or SVM.

VI. Acknowledgements

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

References

[1] N. N. Brushlinsky, M. Ahrens, S. V. Sokolov, and P. Wagner. *World Fire Statistics*. Technical report 21. Center of Fire Statistics, International Association of Fire and Rescue Services, 2016.

[2] S. Tedesco, J. Khodjaev, and B. O'Flynn. "A novel first responders location tracking system: Architecture and functional requirements". In: *Mediterranean Microwave Symp.* IEEE. 2015.

[3] O. D. Lara and M. A. Labrador. "A Survey on Human Activity Recognition using Wearable Sensors". In: *IEEE Communications Survey & Tutorials* 15.3 (2013).

[4] A. Bulling, U. Blanke, and B. Schiele. "A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors". In: *ACM Comp. Surv.* 46.3 (2014).

[5] R. Zhang, F. Hoflinger, and L. Reindl. "Inertial sensor based indoor localization and monitoring system for emergency responders". In: *Sensors* 16.2 (2013).

[6] J. R. Williamson, A. Dumas, G. Ciccarelli, A. R. Hess, B. A. Telfer, and M. J. Buller. "Estimating Load Carriage from a Body-worn Accelerometer". In: *Int. Conf. on wearable and Implantable Body Sensor Networks*. IEEE, 2015.

[7] J. H. Friedman. "Greedy function approximation: A gradient boosting machine". In: *The Annals of Statistics* 29.5 (2001).

[8] R. E. Schapire. "The Boosting Approach to Machine Learning: An Overview". In: *Workshop on Nonlinear Estimation and Classification*. MSRI. Springer, 2003.

[9] U. Blanke and B. Schiele. "Daily Routine Recognition through Activity Spotting". In: *Int. Symp. on Location- and Context-Awareness*. 2009.

[10] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. "A Hybrid Discriminative/Generative Approach for Modeling Human Activities". In: *Int. Joint Conf. on Artificial Intelligence*. 2005.

[11] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler. "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring". In: *Trans. on Information Technology in Biomedicine* 10.1 (2006).

[12] M. Ermes, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen. "Detection of Daily Activities and Sports with Wearable Sensors in Controlled and Uncontrolled Conditions". In: *Trans. on Information Technology in Biomedicine* 12.1 (2008).

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