A Competitive Approach for Human Activity Recognition on Smartphones

Attila Reiss, Gustaf Hendeby and Didier Stricker

Original Publication:

Attila Reiss, Gustaf Hendeby and Didier Stricker, A Competitive Approach for Human Activity Recognition on Smartphones, 2013, ESANN2013, 455-460. ISBN: 978-87419-081-0

From the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013), 24-26 April, Bruges, Belgium

A Competitive Approach for Human Activity Recognition on Smartphones

Attila Reiss[1], Gustaf Hendeby[2,3] and Didier Stricker[1]

1 — German Research Center for Artificial Intelligence (DFKI)
Trippstadter Str. 122, 67663 Kaiserslautern, Germany

2 — Competence Group Sensor Informatics, Division of Sensor and E/W Systems
Swedish Defence Research Agency (FOI), 581 11 Linköping, Sweden

3 — Dept. of Electrical Engineering, Linköping University
581 83 Linköping, Sweden

Abstract. This paper describes a competitive approach developed for an activity recognition challenge. The competition was defined on a new and publicly available dataset of human activities, recorded with smartphone sensors. This work investigates different feature sets for the activity recognition task of the competition. Moreover, the focus is also on the introduction of a new, confidence-based boosting algorithm called ConfAdaBoost.M1. Results show that the new classification method outperforms commonly used classifiers, such as decision trees or AdaBoost.M1.

1 Introduction

The recognition of basic physical activities — such as sitting, walking or running — is a well researched topic [2, 6]. Current research in the area of activity recognition focuses among other things on personalization, on increasing the number of activities to be recognized, and on smartphone-based mobile realizations. Activity monitoring systems using sensors fitted in modern smartphones (accelerometers, gyroscopes, etc.) have the benefit that the user does not have to wear additional sensor components, thus such systems are nonintrusive and have high acceptance in everyday life. On the other hand, drawbacks of such systems are the lower recognition accuracy and reliability. The work in [10] showed that the recognition performance is significantly lower in systems relying only on one sensor position on the user's body (*e.g.* the torso) than in systems with multiple sensor locations (*e.g.* additional to the torso, sensors also placed on the user's upper and lower limbs). Another important issue with smartphone-based systems is the device's position and orientation: either these are fixed — thus losing some of the system's flexibility — or their constant change must be compensated, *e.g.* by developing features robust to sensor displacement [3].

Only few publicly available datasets exist in the research field of human activity monitoring. The recent release of the PAMAP2 dataset [11, 12] addresses this issue, providing data recorded from 9 subjects wearing 3 inertial measurement units and a heart rate monitor, and performing 18 different physical activities. A new public dataset is presented in [1], introducing data recorded from smartphone sensors. Moreover, an activity recognition competition has been proposed on this new dataset, animating the research community to compare existing methods, and to develop new algorithms in order to improve on the recognition performance of human activities. This paper describes the development of a competitive approach participating in the activity recognition competition, focusing on different feature subsets and a new classification algorithm.

2 Methods

This section describes the basic conditions of the competition: the parameters of the dataset, and the provided feature set. Moreover, in order to increase the accuracy of activity recognition, a new boosting algorithm is presented based on concepts and ideas of existing boosting techniques.

2.1 Dataset

The activity recognition competition is defined on a new, publicly available dataset of daily human activities. The dataset was recorded with 30 subjects, performing 6 different activities: walking, ascending stairs, descending stairs, sitting, standing and lying. The embedded 3D-accelerometer and 3D-gyroscope of a waist-mounted smartphone (Samsung Galaxy S II) were used to collect data at 50 Hz. More information about the dataset can be found in [1].

The created dataset was randomly partitioned into a training and test set for the competition, containing 70% and 30% of all samples, respectively. The training part includes activity labels (ground-truth) for each sample, whereas the test data only provides the patterns. The goal of the competition is to predict the unknown activity labels of the test samples. To achieve the highest possible overall accuracy on the test part, standard k-fold cross-validation (CV) is used on the training set to select the most promising methods. This validation technique simulates best the goal of measuring the performance on a randomly selected test set which is excluded from the training procedure.

2.2 Feature sets

The competition dataset does not provide raw sensory data, but directly a large set of extracted features. For this, the sensor signals of both the accelerometer and the gyroscope were pre-processed, then sampled in fixed-width sliding windows of 2.56 seconds and 50% overlap [1]. A feature vector is extracted from each window by computing variables in both time and frequency domain, on different components of the acceleration and angular velocity signals.

In total, the dataset consists of 561 features per activity instance. However, these features have different importance in the task of recognizing physical activities. Previous work has investigated what type of sensors is more valuable, and using the data of a given sensor what kind of extracted features are more valuable for distinguishing different activities. It is commonly accepted that accelerometers are the most informative sensors for activity recognition [6]. Concerning the features extracted from raw signal data, features such as mean, standard deviation, energy, or entropy have proven to be useful [2, 9, 10]. Based on these results, two subsets of the entire feature set were defined:

'Small' feature set: this set only uses features extracted from acceleration

Algorithm 1 ConfAdaBoost.M1

Require: Training set of N instances: $(\underline{x}_i, y_i)$ $i = 1, \ldots, N$ ($\underline{x}_i$: feature vector, $y_i \in [1, \ldots, C]$)
New instance to classify: $\underline{x}_n$

procedure TRAINING($(\underline{x}_i, y_i)$ $i = 1, \ldots, N$)
  Assign equal weight to each training instance: $w_i = \frac{1}{N}$, $i = 1, \ldots, N$
  for $t \leftarrow 1, T$ **do**
    Fit weak learner on the weighted dataset: $f_t(\underline{x}) \in [1, \ldots, C]$
    Compute the confidence of the prediction that instance $\underline{x}_i$ belongs to the predicted class: p_{ti}, $i = 1, \ldots, N$
    Compute error e_t of model on weighted dataset: $e_t = \sum_{i:y_i \neq f_t(\underline{x}_i)} p_{ti} w_i$
    if $e_t = 0$ or $e_t \geq 0.5$ **then**
      Delete last $f_t(\underline{x})$ and terminate model generation.
    end if
    Compute $\alpha_t = \frac{1}{2} \log \frac{1-e_t}{e_t}$
    for $i \leftarrow 1, N$ **do**
      $w_i \leftarrow w_i e^{\left(\frac{1}{2} - \mathbb{I}(y_i = f_t(\underline{x}_i))\right) p_{ti} \alpha_t}$
    end for
    Renormalize the weight of all instances so that $\sum_i w_i = 1$
  end for
end procedure

procedure PREDICTION($\underline{x}_n$)
  Set zero weight to all classes: $\mu_j = 0$, $j = 1, \ldots, C$
  for $t \leftarrow 1, T$ **do**
    Predict class with current model: $[c, p_t(\underline{x}_n)] = f_t(\underline{x}_n)$, where $p_t(\underline{x}_n)$ is the confidence of the prediction that instance $\underline{x}_n$ belongs to the predicted class c
    $\mu_c \leftarrow \mu_c + p_t(\underline{x}_n)\alpha_t$
  end for
  The output class is $\arg\max_j \mu_j \quad j = 1, \ldots, C$
end procedure

data. It includes the time domain features mean, standard deviation, and correlation; the frequency domain features energy, entropy, mean, and maximum frequency; and inclination. In total, this feature set contains 26 out of the original 561 features.
'Large' feature set: in addition to the 'small' feature set, this set also includes features extracted from gyroscope data. Moreover, some additional features are included such as skewness, kurtosis, and the energy in different frequency bands. In total, this feature set contains 128 out of the original 561 features.
These two feature subsets are compared to the entire set of features, referred to as **'full' feature set** in this paper. Results from the comparison of the 3 feature sets are presented in Section 3.

2.3 Classification algorithms

The selection of an adequate classifier for solving activity recognition classification problems has been widely investigated the past decade. Previous work showed that decision tree based classifiers, especially boosted decision trees, usually achieve high performance on named classification tasks [10, 12]. Therefore, the C4.5 decision tree classifier [7] and the AdaBoost.M1 (using C4.5 as weak learner) algorithm [4] are applied on the competition dataset. Moreover, a new boosting method called ConfAdaBoost.M1 is compared to the well known classifiers. ConfAdaBoost.M1 (*cf.* Algorithm 1) is a confidence-based extension of the AdaBoost.M1 algorithm. It is a direct multiclass classification technique, using

Table 1: Comparison of classification results using different feature sets.

	C4.5 decision tree	AdaBoost.M1	ConfAdaBoost.M1
'Small' feature set	92.79%	97.63%	98.30%
'Large' feature set	94.14%	98.91%	99.17%
'Full' feature set	93.55%	98.67%	99.29%

the information about how confident the weak learners are in the prediction of the instance's classification. This idea proved to be beneficial in previous work for the binary classification case (*cf.* the Real AdaBoost algorithm in [5]), and when extending the prediction step of the original AdaBoost.M1 algorithm (*cf.* [8]). However, ConfAdaBoost.M1 is the first multiclass boosting algorithm using the confidence information in both the training and prediction steps.

The main idea of ConfAdaBoost.M1 can be described as follows. In the training part of the algorithm the weak learner's confidence of the classification is returned for each instance (line 5), and is then used to compute the new weight of that instance: the more confident the weak learner is in a correct classification or misclassification the more the weight will be reduced or increased, respectively (line 12). Moreover, the confidence values are used in the prediction part of the algorithm: the more confident the weak learner is in a new instance's prediction the more it counts in the output of the combined classifier (line 21).

3 Results and discussion

The accuracy of all different combinations of the 3 defined feature sets and the 3 classifiers is shown in Table 1. The results are obtained with 10-fold CV on the training part of the competition dataset. For both boosting algorithms the number of boosting iterations is set to 100. It is clear that the ConfAdaBoost.M1 classifier outperforms the other two methods, higher accuracy was achieved with this new boosting technique on all 3 feature sets. Overall, best performance (99.29%) was achieved with the ConfAdaBoost.M1 classifier using all features.

The accuracy of 98.30% with ConfAdaBoost.M1 using the 'small' feature set is noteworthy. This means that although only less than 5% of the original features are used (26 out of 561), the observed accuracy loss is only 1%. Moreover, the 'small' feature set only relies on the acceleration signal, the gyroscope is not used. Therefore, from a practical point of view, it should be considered only using this very limited feature set: it highly reduces the training time of classifiers, as well as requires significantly less computations during online classification. The latter property is especially beneficial for mobile human activity monitoring, where computational resources (*e.g.* of smartphones) might be limited.

Fig. 1 compares the new ConfAdaBoost.M1 algorithm to AdaBoost.M1, using the 'full' feature set. The accuracy of AdaBoost.M1 levels off at significantly fewer iterations, whereas the ConfAdaBoost.M1 algorithm benefits from the fact that it reaches the stopping criterion (*cf.* line 7 of Algorithm 1) much later, thus can further increase the accuracy with more iterations. Moreover, it should be noted that although the classification problem consists of a large amount of features and up to 500 boosting rounds are used — thus the size of the classifier

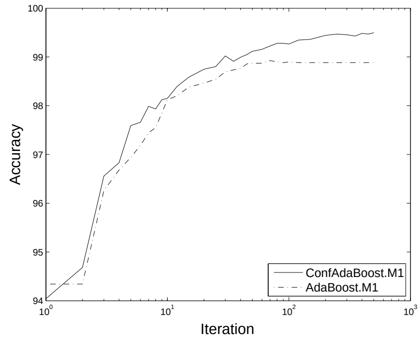

Fig. 1: Comparison of ConfAdaBoost.M1 to AdaBoost.M1: accuracy [%] of the classifiers with respect to the number of boosting iterations.

is very large — no overfitting is observed. Thus ConfAdaBoost.M1 adopts one of the beneficial characteristics of boosting: it rarely overfits a classification task.

The results in this section show that the highest overall accuracy is obtained with the ConfAdaBoost.M1 classifier, when the number of boosting rounds is set to 500 and the entire feature set is used. With 10-fold CV, an accuracy of 99.50% is achieved on the training part of the database, thus similar performance is to be expected on the test set as well, as pointed out in Section 2.1. Table 2 shows the confusion matrix of the best performing classifier. The only noticeable confusion occurs between the postures sitting and standing. In activity recognition systems (*e.g.* in [10]) it is a common restriction that these postures form one activity class, since an extra sensor on the thigh would be needed to reliably differentiate them [2, 6]. The analysis of the trained classifier reveals that a large amount of decision nodes are created to distinguish between sitting and standing. On the other hand, the separation of the posture lying is basically trivial due to the different orientation of the accelerometer. Moreover, the differentiation of the 3 postures from the 3 locomotive activities is a simple task as well: a basic step detection is sufficient, which can be derived from *e.g.* the standard deviation of the vertical acceleration signal.

4 Conclusion and future work

This paper presents a competitive approach for smartphone-based human activity recognition. A new confidence-based boosting method (ConfAdaBoost.M1) is introduced, and experiments show that it outperforms commonly used classifiers. Moreover, this work also reflects on the necessity of using large feature sets: especially in mobile systems — where limitations in computational resources exist — using only a few selected features should be considered due to the minimal performance loss observed with the 'small' feature set. The competition dataset did not provide subject information along with the feature vectors, thus only subject dependent training was possible. However, as shown in previous work (*e.g.* in [12]), the results obtained this way are highly optimistic in the estimation of real life scenarios. Therefore, if subject information would be added

Table 2: Confusion matrix of the ConfAdaBoost.M1 classifier with 500 boosting rounds, using the 'full' feature set. The results are obtained on the training part of the database, with 10-fold CV.

Annotated activity	**Recognized activity** 1	2	3	4	5	6
1 - walk	1223	1	2	0	0	0
2 - asc. stairs	0	1073	0	0	0	0
3 - desc. stairs	0	0	986	0	0	0
4 - sit	0	0	0	1264	21	0
5 - stand	0	0	0	12	1362	0
6 - lie	0	0	0	0	0	1407

to the dataset in a future release, this work should be extended with subject independent training as well. Moreover, this work should also be extended to deal with sensor displacement, in order to develop more robust solutions for smartphone-based systems. However, it is not possible to consider this issue using only the provided dataset [1], since the smartphone was worn by all subjects at the same, fixed position during data collection.

Acknowledgements This work has been performed within the project ActivityPlus, funded by the "Stiftung Rheinland-Pfalz für Innovation" under contract number 961 - 386261/1028.

References

[1] D. Anguita, A. Ghio, L. Oneto, X. Parra and J. L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. In *International Workshop of Ambient Assisted Living (IWAAL)*, 2012.

[2] M. Ermes, J. Pärkkä, and L. Cluitmans. Advancing from offline to online activity recognition with wearable sensors. In *30th Annual International IEEE EMBS Conference*, Vancouver, Canada, pp. 4451–4454, 2008.

[3] K. Förster, P. Brem, D. Roggen, and G. Tröster. Evolving discriminative features robust to sensor displacement for activity recognition in body area sensor networks. *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Melbourne, Australia, pp. 43–48, 2009.

[4] Y. Freund, and R. E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139, 1997.

[5] J. Friedman, T. Hastie, and R. Tibshirani. Additive Logistic Regression: A Statistical View of Boosting. Annals of Statistics, 28:337–407, 2000.

[6] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and I. Korhonen. Activity classification using realistic data from wearable sensors. *IEEE Trans. Inf. Technol. Biomed.*, 10(1):119–128, 2006.

[7] J. R. Quinlan. C4.5: Programs for Machine Learning. *Morgan Kaufmann*, 1993.

[8] J. R. Quinlan. Bagging, boosting and C4.5. In *13th National Conference on Artificial Intelligence (AAAI)*, Portland, OR, pp. 725–730, 1996.

[9] N. Ravi, N. Dandekar, P. Mysore, and M. Littman. Activity recognition from accelerometer data. In *17th Conference on Innovative Applications of Artificial Intelligence (IAAI)*, Pittsburgh, PA, pp. 1541–1546, 2005.

[10] A. Reiss and D. Stricker. Introducing a Modular Activity Monitoring System. In *33rd Annual International IEEE EMBS Conference*, Boston, MA, pp. 5621–5624, 2011.

[11] A. Reiss and D. Stricker. Creating and Benchmarking a New Dataset for Physical Activity Monitoring. In *5th Workshop on Affect and Behaviour Related Assistance (ABRA)*, Crete, Greece, 2012.

[12] A. Reiss and D. Stricker. Introducing a New Benchmarked Dataset for Activity Monitoring. In *IEEE 16th International Symposium on Wearable Computers (ISWC)*, Newcastle, UK, 2012.