An Overview of Classification Techniques for Human Activity Recognition

P Dohnálek[1, 2], P Gajdoš[1, 2], T Peterek[2, 3] and V Snášel[1, 2]
[1] Department of Computer Science, FEECS
[2] IT4Innovations, Centre of Excellence
VŠB – Technical University of Ostrava, 70833 Ostrava, Czech Republic

E-mail: tomas.peterek@vsb.cz

Abstract. In this paper, both classic and less commonly used classification techniques are evaluated in terms of recognizing human activities recorded in the PAMAP2 dataset that was created using three inertial measurement units. Seven algorithms are compared in terms of their accuracy performance with the best classifier being based on the Orthogonal Matching Pursuit algorithm that has been modified to remove the limitation of the number of training vectors per class present in its original version. The overview shows that human activities as defined by the PAMAP2 dataset can be recognized reliably even without any prior data preprocessing.

1. Introduction

Human activity recognition is one of the more recent research topics that recently gained on popularity and focus of both academic and commercial researchers. Since human activity monitoring has a broad range of applications, like homecare systems, prisoner monitoring, physical therapy and rehabilitation, public security, military uses and others, motivation to create a reliable human activity recognition system is great.

Generally, approaches to recognizing activities can be divided into two groups - sensor-based and vision-based. Sensor-based systems use various sensors that are attached to the subject being monitored. Vision-based activity recognition systems, on the other hand, try to eliminate the need for sensors and attempt to recognize subject's behavior from images and video sequences. Both approaches have their challenges arising from their nature. While sensor-based systems require classification algorithms to be as speedy as possible in order to be implemented in low-power wearable devices, accurate and reliable vision-based systems are still a challenge no matter the computation power. This paper focuses on sensor-based systems, one of which was used to create the Physical Activity Monitoring for Aging People (PAMAP2) dataset that this paper elaborates on.

1.1. Current Approaches in Activity Recognition

In general, in activity recognition authors attempt to recognize static states (lying, sitting, standing, etc.), dynamic states (walking, running, etc.) and/or transition states (i.e. standing to walking). Data preprocessing to improve the classification accuracy is common [17]. Classification methods currently widely used in the area are based both on classic algorithms like the Classification And Regression Tree (CART) [9] or k-Nearest Neighbor (k-NN) [4] and more advanced techniques like the Adaptive Neuro Fuzzy Interference System (ANFIS) [3] or Iterative Dichotomiser 3 (ID3) [5] and others.

1.2. The PAMAP2 dataset

The PAMAP2 dataset contains data of nine healthy human subjects, each subject wearing three inertial measurement units (IMUs) by Trivisio, Germany and a heart rate monitor. Each of the three IMUs measures temperature and 3D data from an accelerometer, gyroscope and magnetometer. The data is sampled at 100Hz and transmitted to PC via a 2.4GHz wireless network. Subjects wore one IMU on the dominant wrist, one on the dominant ankle and one on the chest. Detailed information on the

[3] Corresponding author

dataset can be found in [15] and [16].

The methods have been tested on all 9 test subjects in the dataset, labeled in the dataset as *subject101* through *subject109*. Data of all these subjects consists of 2872532 measurements, each containing 54 values. The description of the values is available in the dataset documentation. Some values can be missing, indicated with a NaN (Not a Number) value. Every NaN value was replaced with a zero.

The activities performed are *lying*, *sitting*, *standing*, *walking*, *running*, *cycling*, *Nordic walking*, *ascending stairs*, *descending stairs*, *vacuum cleaning*, *ironing* and *rope jumping*. Transition activities were discarded. Since some measurements contain only NaN values, these measurements were discarded as well. In total, 1942746 activity measurements were used. From each measurement, irrelevant values, that is the orientation of each IMU and timestamp, were removed. Since most heart-rate values were NaNs due to different operating frequency, they were also removed. As a result, each measurement contains 39 values. These values were not preprocessed any further, the classifiers were tested on raw sensor data as provided by the dataset.

2. Classifiers being evaluated

To provide an overview of learning algorithms with application to human activity recognition, eight distinct classifiers were tested, including the above mentioned k-NN and CART. Also, the OMP classifier as defined in [12] was evaluated against a custom modification that significantly improves the reliability of the recognition. Other classifiers are Linear Discriminant Analysis (LDA) [14], Quadratic Discriminant Analysis (QDA) [11] and Nearest Centroid Classifier (NCC) [1]. The following subsections provide a brief informal description for each of the classifiers.

2.1. k-Nearest Neighbors

k-NN is a non-parametric algorithm, meaning that it makes no assumptions about the structure or distribution of the underlying data, thus being suitable for real-world problems that usually do not follow the theoretical models exactly. The method is also considered to be a lazy learning algorithm as it performs little to no training during computation. As a result, the method uses the whole training dataset during classification. k-NN is well known for its simplicity, speed and generally good classification results in applications like bioinformatics [2].

2.2. Nearest Centroid Classifier

An extremely fast classifier. The approach is similar to that of k-NN, but instead of k closest training samples the method picks the label of the class whose training samples' mean (centroid) is closest to the signal query. The speed and simplicity of the algorithm is compensated by low classification performance. Therefore the classifier is usually coupled with one or more data preprocessing techniques. In many implementations the method has been successfully used to create pattern recognition systems in bioinformatics [6].

2.3. Classification and Regression Tree

This algorithm classifies a sample according to groups of other samples with similar properties. During training, the training data is continuously divided into smaller subsets (tree nodes). When the divisions are finished, the samples are clustered together according to their properties. Testing samples are then evaluated against certain conditions in each node and propagated throughout the tree. When the sample reaches a leaf node, it is then assigned the class to which the samples in that node belong. In this paper, a binary tree with logical conditions was used. CARTs are still under extensive research and can be used even as part of larger algorithmic structures [8].

2.4. Linear Discriminant Analysis

LDA is a well-known technique used to identify sample clusters in a given set of data. It attempts to divide clusters (data classes) with a linear function so that the classes are as distant from each other as possible, but at the same time keeping the distance between individual data samples in a single class minimal. The method assumes that the data in each class is normally distributed, but it has still been successfully applied in many problems of automatic recognition, for example in image feature extraction [7].

2.5. Quadratic Discriminant Analysis

As the name suggests, QDA is very closely related to LDA with the exception that QDA does not assume the normal distribution of the data in each class. Instead of the linear function that separates the classes from each other, the function used by QDA is quadratic and can be considered to be a generalization of LDA. Due to greater time complexity, the method is not as widely used as LDA, but it is still feasible in many applications, including the already-mentioned bioinformatics [13].

2.6. Orthogonal Matching Pursuit

Well described in [10], OMP is an iterative sparse approximation algorithm that reduces data into a given number of sparse coefficients and thus can be considered a dimensionality reduction algorithm. Given an overcomplete dictionary of observations, for each observation to be classified OMP picks a number of the best fitting observations from the dictionary and uses them to compute the sparse coefficients. Those are then checked against the dictionary itself for similarity and classified. Originally, the classifier required the number of training observations for each class to be the same. The algorithm was modified so that this limitation is no longer present and evaluated (OMP2 in tables).

Table 1. Recognition accuracy for each of the classifier.

	Training set size				
Classifier	10%	20%	30%	40%	50%
OMP2	98.27	99.14	99.56	99.43	99.60
3-NN	97.51	98.72	99.31	99.29	99.54
CART	98.87	98.49	99.15	99.11	99.37
OMP	95.85	97.56	98.29	98.39	99.43
QDA	75.33	75.37	73.62	73.66	74.68
LDA	61.84	61.16	58.98	58.65	61.34
NCC	51.09	50.63	48.88	46.60	49.59

3. Experiments

The following section describes the dataset used to evaluate the performance of the classifiers as well as the process of the evaluation and its results.

3.1. Experimental settings

The execution of some of the algorithms can be customized through execution parameters which, for these experiments, were set according to the best empirical speed/accuracy ratio. The k-NN algorithm's k parameter was set to 3. The number of sparse coefficients s computed by OMP was 10. For CART, default MATLAB settings was used. All of the algorithms were implemented in the latest version of MATLAB. The entire dataset was divided into a training and a testing set. Experiments were performed on 5 different settings where the training set was a 10%, 20%, 30%, 40% or 50% portion of the dataset.

Given the varied time complexity of the individual classifiers, they were divided into two groups according to the speed with which they finished. Algorithms considered in this paper to be high-speed are LDA, QDA and NCC whose training and testing phase took only several seconds total. The rest is considered to be low-speed as finishing an entire experiment was a matter of tens of minutes to hours, depending on the training set size. For this reason, the number of observations used in the testing phase of a run was limited to 10000. As all compared classifiers are purely deterministic, it was sufficient to run the experiment with each setting only once.

Table 2. Classification accuracy with regards to individual actions (30% training set).

	Classifier						
Actions	OMP2	3-NN	CART	OMP	QDA	LDA	NCC
lying	100	100	99.80	100	88.95	90.28	90.48
sitting	100	99.83	98.90	99.66	83.52	80.03	83.43
standing	99.75	99.75	99.63	99.63	82.27	67.49	58.99
walking	99.27	99.02	99.27	95.36	60.67	44.54	25.57
running	99.86	98.56	99.00	99.71	75.72	31.75	26.29
cycling	99.78	100	99.57	98.92	91.58	86.18	66.52
N. walking	98.54	97.97	99.92	95.54	58.93	34.42	8.20
asc. stairs	98.81	99.15	97.46	97.63	52.20	48.98	24.58
des. stairs	99.56	99.12	96.70	96.70	52.75	35.39	14.94
vacuum cl.	100	99.90	98.89	99.20	76.26	58.85	53.52
ironing	99.57	99.46	99.46	99.14	88.20	71.14	72.10
rope j.	100	99.32	99.55	100	65.77	53.38	49.55

3.2. Results

The classification accuracies given as percentual success rate are shown in Table 1 where the classifiers are sorted according to their success in the descending order. It can be seen that in almost every case the modified version of the OMP classifier (OMP2) is superior to the other classifiers, the exception being the 10% training set where CART performs better. For training set sizes of 30% and higher, OMP2 becomes very closely followed by k-NN which is, in turn, only slightly better than CART. The original OMP, while providing satisfactory results, was the least successful of the low-speed algorithms. At 10% training set, the difference between OMP and OMP2 was the most significant at 2.42%. When the training set size was set to 50%, all low-speed methods provided very accurate recognitions.

For high-speed classifiers the recognition accuracies were not very impressive, once again proving that lower time complexity usually impacts precision. NCC was the fastest of the classifiers, but also the least successful. Following was, as was expected, LDA with slower computation, but better results. The same holds true for QDA.

How accuracy is dependent on the training set can be seen in Figure 1 for the low-speed classifiers and Figure 2 for the high-speed ones. For the more sophisticated, low-speed algorithms the increase of training set size does indeed benefit the recognition accuracy. That is not the case of the high-speed classifiers where the training set size practically did not matter or made the accuracy worse. That is the expected behavior because the observation distribution in the dataset does not change with the amount of observations, meaning it is very difficult if not impossible to reliably cluster observations together according to their classes. This also explains why Figure 2 suggests that larger training sets actually diminish the accuracy performance.

Since the experiments show that 30% can provide very satisfactory results, it is reasonable to consider this training set size a good compromise between speed and accuracy. For this reason, Table 2 elaborates on the results for this training set size. It shows the percentual success rates for each classifier with regards to each of the actions to be recognized. *Lying* and *sitting* came out as activities fairly easy to recognize with any of the evaluated classifiers while distinguishing *Nordic walking* was a task too difficult for high-speed classifiers. This is to be expected due to the similarity of *Nordic walking* to simple *walking*. Low-speed classifiers performed very well in recognizing every activity when OMP2's worst result was misclassifying only 1.46% of total *Nordic walking* observations. k-NN managed to drop below 99% only in a single activity. CART provided fairly consistent and satisfactory results, although 2.86% deficiency in *descending stairs* against OMP2 becomes noteworhty.

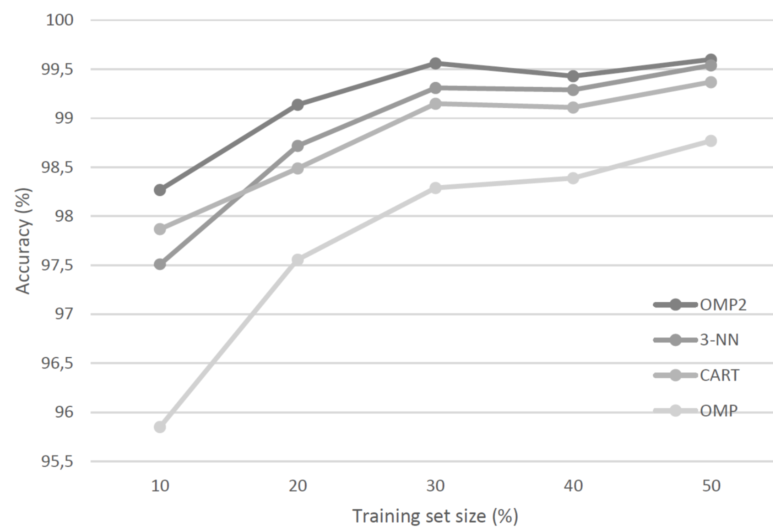

Figure 1. Accuracy dependency on the size of the training set for low-speed classifiers.

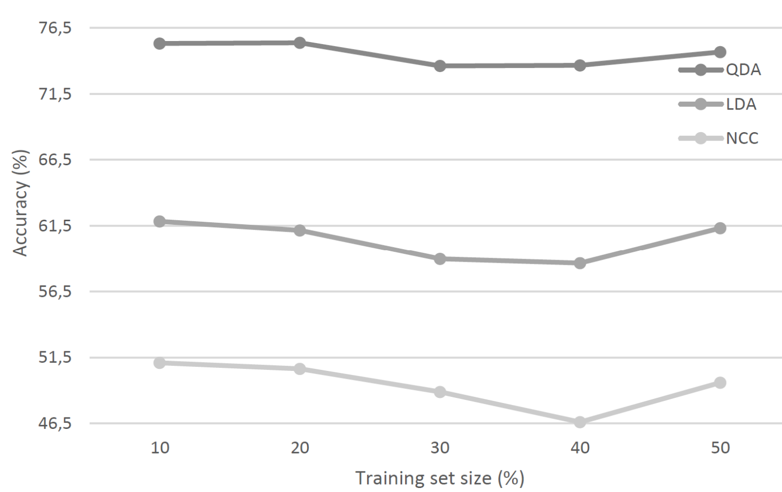

Figure 2. Accuracy dependency on the size of the training set for high-speed classifiers.

4. Conclusion

This paper evaluated several classification techniques and presented their success rates in human activity recognition without any prior preprocessing. Given the sensor technology that was used to create the PAMAP2 dataset, it was shown that activities performed in the database can be recognized reliably and with very high precision. In terms of recognition accuracy, the presented modification of the OMP classifier was shown to perform the best, however the precision comes at the price of significant time complexity. The fastest of the algorithms was NCC, but its recognition accuracy is not sufficient for practical use. From the speed/accuracy ratio perspective, k-NN seems to be the most reasonable choice as its accuracy performance is superseded by OMP2 only closely, but k-NN has a significant edge in computation times. For this reason, the main focus of future work in this area should be making the classifiers more efficient or finding a suitable preprocessing technique that would enable high-speed classifiers to provide better results.

Acknowledgements
The article has been elaborated in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 funded by Structural Funds of the European Union and state budget of the Czech Republic. The work is partially supported by Grant of SGS No. SP2013/70, VŠB – Technical University of Ostrava, Czech Republic. This work was also supported by the Bio-Inspired Methods: research, development and knowledge transfer project, reg. no. CZ.1.07/2.3.00/20.0073 funded by Operational Programme Education for Competitiveness, co-financed by ESF and state budget of the Czech Republic.

References
[1] Alan R Dabney and John D Storey 2007 Optimality driven nearest centroid classification from genomic data. *PLoS ONE* **2(10)**
[2] Paul Horton and Kenta Nakai 1997 Better prediction of protein cellular localization sites with the it k nearest neighbors classifier *Ismb* **5** 147–52
[3] Jyh-Shing Roger Jang and Chuen-Tsai Sun 1997 *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*
[4] Luciana C Jatoba, Ulrich Grossmann, Chistophe Kunze, Jorg Ottenbacher and Wilhelm Stork 2008 Context-aware mobile health monitoring: Evaluation of different pattern recognition methods for classification of physical activity *30th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society* 5250 –53
[5] D Lavanya and K Usha Rani 2011 Performance evaluation of decision tree classifiers on medical datasets *International Journal of Computer Applications* **26(4)** 1–4
[6] Ilya Levner 2005 Feature selection and nearest centroid classification for protein mass spectrometry *BMC bioinformatics* **6**
[7] Ming Li and Baozong Yuan 2005 2d-lda: A statistical linear discriminant analysis for image matrix *Pattern Recognition Letters* **26(5)** 527–32
[8] Andy Liaw and Matthew Wiener 2002 Classification and regression by randomforest *R news* **2(3)** 18–22
[9] Wei-Yin Loh 2011 Classification and regression trees *WIREs Data Mining Knowl Discov* **1** 14–23
[10] Thomas Blumensath and Mike E Davies 2007 On the difference between orthogonal matching pursuit and orthogonal least squares
[11] Santosh Srivastava, Maya R Gupta and Béla A Frigyik 2007 Bayesian quadratic discriminant analysis *J. Mach. Learn. Res.* **8** 1277–305
[12] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry and Yi Ma 2009 Robust face recognition via sparse representation *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31** 210–27
[13] MQ Zhang 1997 Identification of protein coding regions in the human genome by quadratic discriminant analysis *Proceedings of the National Academy of Sciences* **94(2)** 565–8
[14] Fujin Zhong and Jiashu Zhang 2013 Linear discriminant analysis based on l1-norm maximization *IEEE Transactions on Image Processing* **22(8)** 3018–27
[15] A Reiss and D Stricker 2012 Creating and benchmarking a new dataset for physical activity monitoring *The 5th Workshop on Affect and Behaviour Related Assistance (ABRA)*
[16] A Reiss and D Stricker 2012 Introducing a new benchmarked dataset for activity monitoring *The 16th IEEE International Symposium on Wearable Computers (ISWC)*
[17] A M Khan, Young-Koo Lee, S Y Lee and Tae-Seong Kim 2010 A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer *IEEE Transactions on Information Technology in Biomedicine* **14(5)** 1166–72