Comparison of Classification Algorithms for Physical Activity Recognition*

Tomáš Peterek[1,2], Marek Penhaker[1,2], Petr Gajdoš[1,3], and Pavel Dohnálek[1,3]

[1] IT4 Innovations, VSB-Technical University of Ostrava, Czech Republic
[2] Department of Cybernetics and Biomedical Engineering, VSB-Technical University of Ostrava, Czech Republic
[3] Department of Computer Science, VSB-Technical University of Ostrava, Czech Republic
tomas.peterek@vsb.cz

Abstract. The main aim of this work is to compare different algorithms for human physical activity recognition from accelerometric and gyroscopic data which are recorded by a smartphone. Three classification algorithms were compared: the Linear Discriminant Analysis, the Random Forest, and the K-Nearest Neighbours. For better classification performance, two feature extraction methods were tested: the Correlation Subset Evaluation Method and the Principal Component Analysis. The results of experiment were expressed by confusion matrixes.

1 Introduction

Nowadays, cell phones do not serve us only as communication medium, they are equipped by powerful CPU and a GPU therefore it opens new opportunities for this field. The cell phones together with external sensors are possible to use as devices for control e.g. home appliances or they can be used for recording and pre-processing biomedical signals (ECG, EEG). Modern smartphones are equipped by embedded accelerometers which do not be necessarily used only for rotating the screen or playing games but they can be useful as a sensor of regular accelerometric data, which be used for many purposes. One of these purposes can be monitoring of physical activity. The smartphones allow preprocesing of data and due to powerful CPU perform sophisticated classification. The Smartphones with applications for physical activity recognition can be used in homes for elderly people, for recording circadian rhythms, or as fall detectors.

* This article has been elaborated in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 supported by Operational Programme Research and Development for Innovations funded by Structural Funds of the European Union and state budget of the 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.

A. Abraham et al. (eds.), *Innovations in Bio-inspired Computing and Applications,* Advances in Intelligent Systems and Computing 237,
DOI: 10.1007/978-3-319-01781-5_12,

2 Previous Work

The issue of physical activity recognition was solved many times by different classification approaches. e.g: the Support Vector Machine [1]. the Hidden Markov Model [2] [3], classification based on sparse representation [4], Bayes aproaches [5], the Neural Network [6][7], or the Linear Discriminant Analysis [8],[9].

3 Dataset

The dataset was downloaded from the UCI learning repository. The data were recorded by a smartphone Samsung Galaxy SII. 30 volunteers within ages of 19-48 years participated at this experiment. Each participant alternated six types of movement (walking, walking upstairs, walking downstairs, sitting standing, lying).The smartphones were placed on the volunteer's waists. Data were captured by 3-axis embedded accelerometer and 3 axial angular velocity sensor. The sampling frequency was set to 50 Hz. The data from accelerometers and gyroscope were filtered against noise and consequently they were divided segments into 2.56 seconds length with 50 % overlapping. Jerk signals were derived from these segments and magnitudes were calculated, and consequently The Fast Fourier Transform was applied on some these signals. Using this procedure ten different signals in time domain and four signals in frequency domain were obtained. The Jerks signals are derivative signals of regular accelerometric and gyroscopic data and they are used in many useful applications: dynamic motion aerial vehicle trace measurement, earthquake-resistant mechanisms of structures, mechanisms of high speed auto-control of machines, human responses in high speed moving vehicle and high-speed elevators. From each segment a few basic parameters were counted: (mean value, standard deviation, median of absolute deviation, max and min values, energy of segment, interquartile range, entropy, autorregresion coefficients, correlation coefficient between two signals, index of the frequency component with largest magnitude, weighted average of the frequency components to obtain a mean frequency, skewness of the frequency domain signal, kurtosis of the frequency domain signal, energy of a frequency interval within the 64 bins of the FFT of each window, angle between to vectors). By these statistical operations 561 features were calculated, which were used for classification. [1]

4 Classification Method

4.1 Random Forest

The Random Forest (RF) is very popular classification and regression algorithm. The RF algorithm belongs to ensemble learning methods. Likewise a regular forest consists of a number of trees, the RF algorithm consists of a number of classification or regression trees (CART). The algorithm does not use all features for the CART construction but only a few of them. The RF was designed by Leo Breiman in 2001.[12] In this paper, Breiman himself compare the RF with other

ensemble techniques and mentioned that this method has higher accuracy than e.g. Adaboost method. Since that time the RF is used in bioinformatics, medical informatics and so on. The algorithm is resistant to outliers, missing values, or noise. The RF stands out especially in simplicity of parameter tuning but the main problem is his interpretability. For optimal settings of the algorithm only two parameters have to be set: the number of trees in the forest and the number of variables in trees.

4.2 K-Nearest Neighbours

The K-nearest neighbours (KNN)is a non-parametric algorithm for classification. The KNN is one of the easiest method for data classification. The training set T and the testing sample x_i are given. The KNN classifier tries to find sample x_r from the training set T, with a minimal Euclidian distance to the testing sample. Better results are achieved if more than one sample from the training set are founded. This algorithm achieves satisfactory results but is not suitable for solving difficult tasks.

4.3 Linear Discriminant Analysis

Th Linear discriminant analysis (LDA) is a parametric classification technique. The LDA has a lot in common with the Principal Component Analysis (PCA) but with different that PCA does more for feature separation and LDA does more feature classification. The main aim of this method is to find linear combinations of features, which provide the best separation between classes. These combinations are called discriminant functions. The algorithm was developed by Fischer in 1931 but in his original form the algorithm is able to classify only to two classes. In 1988, the algorithm was improved for multiclass classification problem. There are n classes. The intra-class matrix can be calculated by

$$\hat{\Sigma}_w = S1 + ... + S_n = \sum_{i=1}^{n} \sum_{x \in c_i} (x - \hat{x_i}) (x - \hat{x_i})' \tag{1}$$

and inter-class matrix by an equation

$$\hat{\Sigma}_b = \sum_{i=1}^{n} m_i (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})' \tag{2}$$

where m_i is a number of samples in each class from the training set and x_i is the mean for each class and $\hat{x}$ is the total mean vector. Now, a linear transformation Φ should be suggested, in order to maximize Rayleigh coefficient, which is the ratio of determinants of inter-class and intra-class scatter matrixes.

$$J(\Phi) = \frac{\left| \Phi^T \hat{\Sigma}_b \Phi \right|}{\left| \Phi^T \hat{\Sigma}_w \Phi \right|} \tag{3}$$

The linear transformation Φ can be counted by solving equation

$$\hat{\Sigma}_b \Phi = \lambda \hat{\Sigma}_w \Phi \tag{4}$$

The last step is classification itself. The basic principle is measurement of metric or cosine distances between a new instances and centroid of classes. The new instances are classified acording to expression:

$$arg\, min\, d\left(z\Phi, \bar{x}_k \Phi\right) \tag{5}$$

5 Feature Extraction

All mentioned methods are sensitive to irrelevant features. Therefore the number of features has to be reduced so that only relevant features are preserved. There are three basic approaches:

- Feature transform: The basic idea of feature transform method is to transform the original feature space to a new feature space with smaller dimensions. There are algorithms such as: The Singular Value Decomposition, Non-negative Matrix Factorization (NNMF) or Principal Component Analysis (PCA), which are able to solve this task. The last named method showed promising results in similar classification problems, therefore it was also tested in this task.
- Feature filter carry out the feature selection process as a pre-processing step with no induction algorithm. The featerue filter are faster than wrapper and have better results in generalization.
- Feature wrapper achieves the best results but it is very time consuming especially if the number of features is high. The wrappers search for the best result in the whole space of possibly solution. It means that the wrapper methods have to design the classifier for every possible set of feature combination. In this case it is 2^{561} possible solutions. Therefore only first two approaches were tested.

5.1 Feature Filter – Correlation Feature Selection Method

The idea of feature filters is based on a hypothesis: *Good feature subsets contain features highly correlated with the classification, yet uncorrelated to each other.* The Feature filters try to decide which features should be included in the final subset and which should be ignored. The Correlation Feature Selection is typical feature filter method, which evaluates each subsets according to a hearuastic function:

$$M_s = \frac{k\overline{_{rcf}}}{\sqrt{k + k\left(k-1\right)k\overline{_{rff}}}} \tag{6}$$

where M_s is the merit of a subset S, which contains k features. Parameters $k_{\overline{rcc}}$ and $k_{\overline{rcf}}$ expres corelation between a feature and class respectively a feature and a feature. The acceptance of a feature depends on the extent to which it predicts classes in areas of the instance space not already predicted by other features. The numerator of equation provides information how a set of features is predictive of the class and the denominator provides informations how much redudancy there is among the features. [10]

5.2 Principle Components Analysis

The PCA tries to transform a set of observations of possibly correlated variables to a set of new uncorrelated variables called principal components. The PCA components can be counted by

$$X = YP \tag{7}$$

where X is a centering matrix and Y is an input matrix, P is a matrix of the eigenvector of the covariance vector matrix C, expressed by equation

$$C_x = P\Delta P^T \tag{8}$$

where P is orhonormal and Δ is a diagonal matrix of eigenvalues. The number of components which will be used for classification should cover more than 80 % of original variables dispersion. [11]

6 Testing and Results

The data were devided into a training part (70%) and a testing part (30%). Firstly, the full dataset was classified by the three algorithms. In the second part the dataset were transformed by the PCA and lastly the dataset was reduced by the CFS algorithm. The results were expressed by the confusion matrixes and precisions and recalls were counted for each class. For the CFS method the evolutionary search strategy was used. The parameters of the search strategy were set up on values which are shown in the Table 1. Using the CFS algorithm, 250 of features were screened out and the rest of them were used for classification. The PCA counted 10 principal components which represents over 90 % of original variables dispersion. Covererage of the original dispersion for each component are shown in the Fig. 1. These 10 principal components were used as a new feature vector. The results are expresed by the confusion matrixes and by values of the precision and the recall. The results for all classifiers and for both types of feature selection are shown in the Table 2, Table 3 and Table 4.

Table 1. The parameters of the evolutionary search strategy

Parameter	Value
Crossover probability	0.6
Generations	100
Mutation probability	0.01
Population Size	50

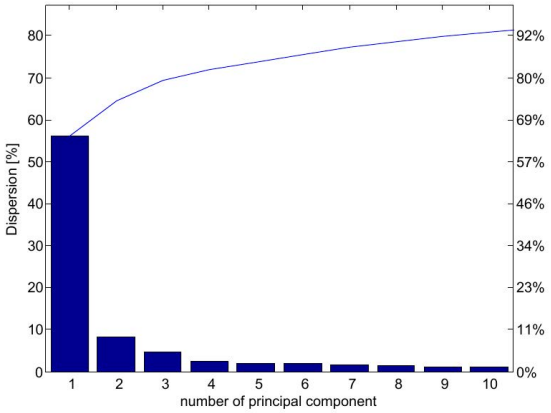

Fig. 1. The coverage of the original variable dispersion for the first ten principal components

Table 2. The confusion matrixes for the LDA

LDA	Walking	Upstairs	Downstairs	Sitting	Standing	Laying	Precision [%]
Walking	490	6	0	0	0	0	98.79
Upstairs	11	460	0	0	0	0	97.66
Downstairs	1	15	404	0	0	0	96.16
Sitting	0	1	0	435	55	0	88.59
Standing	0	0	0	22	510	0	95.86
Laying	0	0	0	0	0	537	100
Recall [%]	97.61	95.44	100	95.19	90.27	100	
LDA+CFSsubset	Walking	Upstairs	Downstairs	Sitting	Standing	Laying	Precision [%]
Walking	494	2	0	0	0	0	99.6
Upstairs	10	461	0	0	0	0	97.88
Downstairs	8	20	392	0	0	0	93.33
Sitting	0	1	0	396	94	0	80.65
Standing	0	0	0	65	467	0	87.78
Laying	0	0	0	0	0	537	100
Recall [%]	96.48	95.25	100	85.9	83.24	100	

Table 3. The confusion matrixes for the KNN

KNN	Walking	Upstairs	Downstairs	Sitting	Standing	Lying	Precision [%]
Walking	473	8	15	0	0	0	95.36
Upstairs	31	422	18	0	0	0	89.60
Downstairs	53	46	321	0	0	0	76.43
Sitting	0	2	0	389	99	1	79.23
Standing	0	0	0	88	451	0	84.77
Lying	0	0	0	3	1	533	99.26
Recall [%]	84.92	88.28	90.68	82.24	81.85	99.81	
KNN+CFSsubset	Walking	Upstairs	Downstairs	Sitting	Standing	Lying	Precision [%]
Walking	460	10	26	0	0	0	92.74
Upstairs	46	420	5	0	0	0	89.17
Downstairs	32	62	326	0	0	0	77.62
Sitting	0	1	0	378	112	0	76.99
Standing	0	0	0	41	491	0	92.29
Lying	0	0	0	0	0	537	100
Recall [%]	85.50	85.19	91.32	90.21	81.43	100	
KNN+PCA	Walking	Upstairs	Downstairs	Sitting	Standing	Lying	Precision [%]
Walking	110	108	43	96	66	73	22.18
Upstairs	128	56	66	100	57	64	11.89
Downstairs	75	52	67	77	50	99	15.95
Sitting	81	73	55	102	66	114	20.77
Standing	91	79	67	108	89	98	16.73
Lying	119	110	68	82	59	99	18.44
Recall [%]	18.21	11.72	18.31	18.05	23.0	18.1	

Table 4. The confusion matrixes for the RF

Random Forest	Walking	Upstairs	Downstairs	Sitting	Standing	Lying	Precision [%]
Walking	482	6	8	0	0	0	97.18
Upstairs	36	430	5	0	0	0	91.30
Downstairs	14	38	368	0	0	0	89.76
Sitting	0	0	0	422	69	0	85.95
Standing	0	0	0	54	478	0	89.85
Lying	0	0	0	0	0	537	100
Recall [%]	90.6	92.67	96.59	88.66	87.39	100	
Random Forest+CFS	Walking	Upstairs	Downstairs	Sitting	Standing	Lying	Precision [%]
Walking	478	6	12	0	0	0	96.37
Upstairs	56	402	13	0	0	0	85.35
Downstairs	7	41	372	0	0	0	88.57
Sitting	0	0	0	395	96	0	80.45
Standing	0	0	0	65	467	0	87.78
Lying	0	0	0	0	0	537	100
Recall [%]	88.35	89.53	93.7	85.87	82.95	100	
Random Forest+PCA	Walking	Upstairs	Downstairs	Sitting	Standing	Lying	Precision [%]
Walking	97	70	9	120	82	118	19.56
Upstairs	102	57	20	124	65	103	12.10
Downstairs	57	38	15	71	60	179	3.57
Sitting	80	59	33	69	82	168	14.05
Standing	95	108	174	4	3	148	15.79
Lying	73	77	28	109	84	161	27.19
Recall [%]	17.9	13.97	12.3	11.73	20.39	16.69	

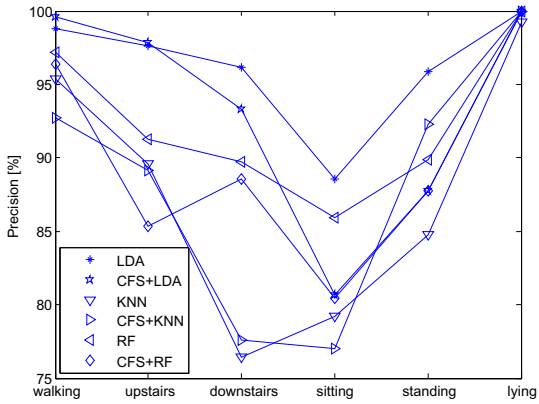

Fig. 2. The achieved values of precision for all classifiers

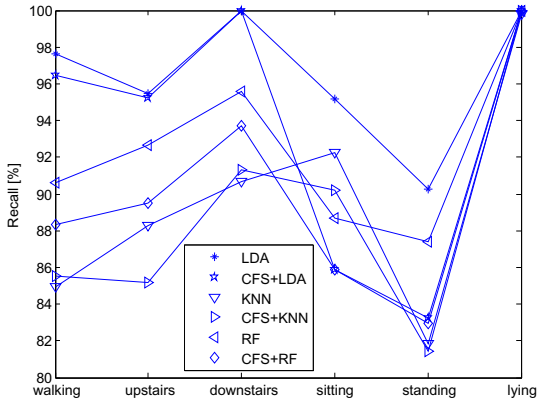

Fig. 3. The achieved values of recall for all classifiers

7 Conclusion

In this work three classification methods were tested for human physical activity recognition. The best classification performance was achived by the Linear Discriminant Analysis. The Random Forest and the K-NN methods were not able to achieve comparable results with the LDA. For simplification of classifier and possible performance classification improvement, two feature selection method were tested: the Corelation Feature Selection Method and the Principal Component Analysis. Reduction of the dataset by the CFS increased the precision of walking and walking upstairs but decreased the precision of the rest states. The most easist state for recognition was lying. The LDA and the RF achieved 100 % of accuracy. The PCA method completely failed because averagy

precission is about 20 %. It possible to say, that this method is not suitable for this task. Generaly, the LDA seems like very promising method for humans activity recognition , what is claimed in others works too.

References

1. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L.: Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. In: Bravo, J., Hervás, R., Rodríguez, M. (eds.) IWAAL 2012. LNCS, vol. 7657, pp. 216–223. Springer, Heidelberg (2012)
2. Li, A., Ji, L., Wang, S., Wu, J.: Physical activity classification using a single triaxial accelerometer based on HMM. In: IET International Conference on Wireless Sensor Network, IET-WSN, November 15-17, pp. 155–160 (2010), doi:10.1049/cp.2010.1045
3. Wu, J.-K., Dong, L., Xiao, W.: Real-time Physical Activity classification and tracking using wearble sensors. In: 2007 6th International Conference on Information, Communications & Signal Processing, December 10-13, pp. 1–6 (2007), doi:10.1109/ICICS.2007.4449890
4. Liu, S., Gao, R.X., John, D., Staudenmayer, J., Freedson, P.S.: Classification of physical activities based on sparse representation. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), August 28-September 1, pp. 6200–6203 (2012), doi:10.1109/EMBC.2012.6347410
5. Madabhushi, A., Aggarwal, J.K.: A Bayesian approach to human activity recognition. In: Second IEEE Workshop on Visual Surveillance (VS 1999), pp. 25–32 (July 1999), doi:10.1109/VS.1999.780265
6. Fang, H., He, L.: BP Neural Network for Human Activity Recognition in Smart Home. In: 2012 International Conference on Computer Science & Service System (CSSS), August 11-13, pp. 1034–1037 (2012), doi:10.1109/CSSS.2012.262
7. Khan, A.M., Lee, Y.-K., Kim, T.-S.: Accelerometer signal-based human activity recognition using augmented autoregressive model coefficients and artificial neural nets. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008, August 20-25, pp. 5172–5175 (2008), doi:10.1109/IEMBS.2008.4650379
8. Uddin, M.Z., Lee, J.J., Kim, T.-S.: Independent Component feature-based human activity recognition via Linear Discriminant Analysis and Hidden Markov Model. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2008, August 20-25, pp. 5168–5171 (2008), doi:10.1109/IEMBS.2008.4650378
9. Abidine, M.B., Fergani, B.: Evaluating C-SVM, CRF and LDA classification for daily activity recognition. In: 2012 International Conference on Multimedia Computing and Systems (ICMCS), May 10-12, pp. 272–277 (2012), doi:10.1109/ICMCS.2012.6320300
10. Hall, M.A.: Correlation-based feature selection for machine learning. The University of Waikato (1999)
11. Peterek, T., Krohova, J., Smondrk, M., Penhaker, M.: Principal component analysis and fuzzy clustering of SA HRV during the Orthostatic challenge. In: 2012 35th International Conference on Telecommunications and Signal Processing (TSP), July 3-4, pp. 596–599 (2012), doi:10.1109/TSP.2012.6256366
12. Breiman, L.: Random Forests. Mach. Learn. 45(1), 5–32 (2001), `http://dx.doi.org/10.1023/A:1010933404324`, doi:10.1023/A:1010933404324