ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

A One-Vs-One Classifier Ensemble with Majority Voting for Activity Recognition

B. Romera-Paredes[1], M. S. H. Aung[1], N. Bianchi-Berthouze[1]

[1]University College London, UCL Interaction Center (UCLIC)
WC1E 6BT, London, UK

Abstract. A solution for the automated recognition of six full body motion activities is proposed. This problem is posed by the release of the Activity Recognition database [1] and forms the basis for a classification competition at the European Symposium on Artificial Neural Networks 2013. The data-set consists of motion characteristics of thirty subjects captured using a single device delivering accelerometric and gyroscopic data. Included in the released data-set are 561 processed features in both the time and frequency domains. The proposed recognition framework consists of an ensemble of linear support vector machines each trained to discriminate a single motion activity against another single activity. A majority voting rule is used to determine the final outcome. For comparison, a six "winner take all" multiclass support vector machine ensemble and $k-$Nearest Neighbour models were also implemented. Results show that the system accuracy for the one versus one ensemble is 96.4% for the competition test set. Similarly, the multiclass SVM ensemble and $k-$Nearest Neighbour returned accuracies of 93.7% and 90.6% respectively. The outcomes of the one versus one method were submitted to the competition resulting in the winning solution.

1 Introduction

The use of Machine learning (ML) as an approach to interpret human biomechanical information has increasingly been investigated in recent years. For example, the survey by Lai et al. [4] describes an overview of computational intelligence methods applied to a wide variety of gait related recognition problems. ML can be particularly powerful in settings where there is limited biomechanical data and where full body apparatus, such as inertial measurement unit (IMU) body suits or multiple camera based systems, are not practical. Therefore robust and accurate recognition systems which are dependent only on a small number of devices are highly desirable. Additionally, if the activities of interest are well defined and can be reliably labeled, then a small number of optimally placed IMUs with supervised ML is a fast, minimally invasive and inexpensive solution.

In Aung et al. [2] a comparison between a supervised ML model based on Gaussian Mixture Models with other bespoke rule based methods was made. In that study gait events were classified using a single tri-axial accelerometer. Only the ML method was found to be robust to varying the attachment location of the sensor on the leg. In the study by Anguita et al. [1] a multiclass Support Vector Machine (SVM) framework classified six full body activities with an overall accuracy of 89%. In that study feature vectors containing 17 variables derived

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

from accelerometer and gyroscope signals acquired by a single smart phone. In the next section the data collected and presented in Anguita et al. [1] is outlined and forms the basis of this study. In Section 3, we discuss our rationale to solve this classification problem followed by a description of our proposed framework. Section 4 shows comparative results between the proposed ensemble, a comparable multiclass SVM and $k-$Nearest Neighbour ($k-$NN). Finally we conclude and discuss the findings in Section 5.

2 Data Description

Originating from the study by Anguita et al. [1] the Activity Recognition (AR) data-set was publicly released. This data-set forms the classification problem for the AR competition within the European Symposium on Artificial Neural Networks 2013 (ESANN). The released data-set contains a training set of 7352 instances each categorized into one of six motion classes: *sitting, standing, laying, walking, walking upstairs* and *walking downstairs*. A test set containing 2947 instances with initially unreleased labels was also provided for the final competition submission. The training set had been randomized thus mixing the samples from each of the 30 participants. The participant identifiers were not released.

The provided set of features were derived from tri-axial accelerometric and gyroscopic signals, captured using a single Samsung Galaxy S2 phone worn on the waist sampled at 50Hz. The acquired signals had been pre-processed using a median filter and a 3^{rd} order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Additionally, the acceleration signals were further separated into body and gravity signals using another low pass Butterworth filter with a corner frequency of 0.3 Hz. The third order derivatives (jerk) of each of the signals along each axis were also provided along with their corresponding magnitudes using an Euclidean norm. The Fast Fourier Transform (FFT) of the signals and their corresponding magnitudes were also provided. Each of these resultant signals were applied to 17 functionals calculated using a sliding window of 2.56s with a 50% overlap. Overall this led to feature vectors containing 561 variables for each instance.

3 Rationale and Methodology

The rationale that underpins the proposed framework begins with the assumption that the three non walking activities (*standing, laying* and *sitting*) are likely to have very different spatio-temporal characteristics to the three walking based activities (*walking, walking upstairs* and *walking downstairs*). A non-walking activity should be readily separable against any of the walking activities. However it would be much more difficult to discriminate between classes that have similar motion characteristics, for example *sitting* and *standing* still. Therefore, we consider the use of One Versus One models (OVO) that are specifically trained to separate a single class against another. We have investigated linear classifiers

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

for two reasons. First, a preliminary visual analysis was conducted (e.g. see figure 1) which showed that even considering only the first 3 principal components, the classes were almost linearly separable. Second, the dimensionality of the data is much smaller than the number of instances, thus working directly with the original features is more efficient. If the OVO modules work well then a simple unweighted Majority Voting (MV) rule should be sufficient to determine the final outcome. If our assumptions hold the expected distribution of votes should occur as follows: 5 votes for the true class, 4 and 3 votes for the two similar false classes respectively and the three dissimilar classes with $0-2$ votes each.

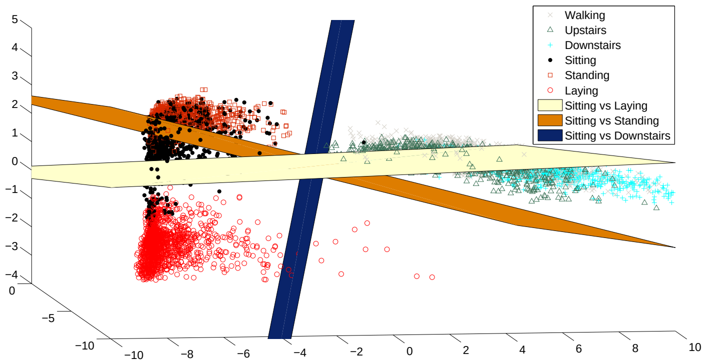

Fig. 1: Visualization of the training data with three OVO decision boundaries projected onto the three principal components.

3.1 One Versus One Support Vector Machines

The proposed OVO ensemble consists of 15 linear single class SVMs [3] each trained only on the instances relevant to the two OVO classes of interest a and b. Thus giving subsets of the data:

$$D_{a,b} = \left\{ (\mathbf{x_i},\, y_i) \,|\, \mathbf{x_i} \in \mathbb{R}^p,\, y_i = \begin{cases} 1 & \text{if class}\,(\mathbf{x_i}) = a \\ -1 & \text{if class}\,(\mathbf{x_i}) = b \end{cases} \right\}_{i=1}^{n},$$

where $\mathbf{x_i} \in \mathbb{R}^p$ denotes the vector of features for instance i. The number of training instances n is determined by the number of instances that belong to class a or b from the N training instances available for all classes. Soft margin linear SVM models were generated according to the following:

$$\begin{array}{ll} \min\limits_{\mathbf{w},b,\xi} & \frac{1}{2}\left\|\mathbf{w}\right\|^2 + C\sum\limits_{i=1}^{n}\xi_i \\ s.t: & y_i\left(\mathbf{w}^\top \mathbf{x_i} + b\right) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \;\forall i \in \{1 \dots n\} \end{array}$$

where $\mathbf{w}$ and b represents the weight and bias parameters of the linear model. C is a parameter which controls how the model fits the input data and needs to

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

be tuned beforehand. The full set of features provided in the database was used leading to $p = 561$. No further feature selection or processing was implemented. The model was optimized using Sequential Minimal Optimization [3].

3.2 One Vs All Support Vector Machines and k-Nearest Neighbour

Six linear SVMs were trained to classify each of the six labels on a one versus all (OVA) basis. We apply the full unaltered feature set and the same learning scheme as described in the previous section. The main difference here being that the six classifiers take $n = N$ training instances in the following form:

$$D_a = \left\{ (\mathbf{x_i},\, y_i) \,|\, \mathbf{x_i} \in \mathbb{R}^p,\, y_i = \begin{cases} 1 & \text{if class}\,(\mathbf{x_i}) = a \\ -1 & \text{if class}\,(\mathbf{x_i}) \neq a \end{cases} \right\}_{i=1}^{n}$$

For final outcome of the OVA framework we employ a "winner take all" rule; taking the highest value from the six models as final. Additionally, a k-Nearest Neighbour model is also implemented for further comparison.

4 Results

We have compared the performance of the three methods described. The parameters that needs to be preset are C for SVM-based approaches and k for $k-$NN. The training scheme consisted of randomly splitting the original training set of 7532 instances into a new training set of $N = 7000$ instances and test set of 352 instances. This process was repeated 800 times for a predefined value of the parameters ($C \in \{10^{-2}, 10^{-1}, 10^{0}, 10\}$ for SVMs and $k \in \{1, 2 \ldots, 15\}$ for $k-$NN). The optimal value was obtained by choosing the maximum average performance over all iterations. This optimal value was found to be $C = 10^{-1}$ for both SVM approaches and $k = 10$ for $k-$NN.

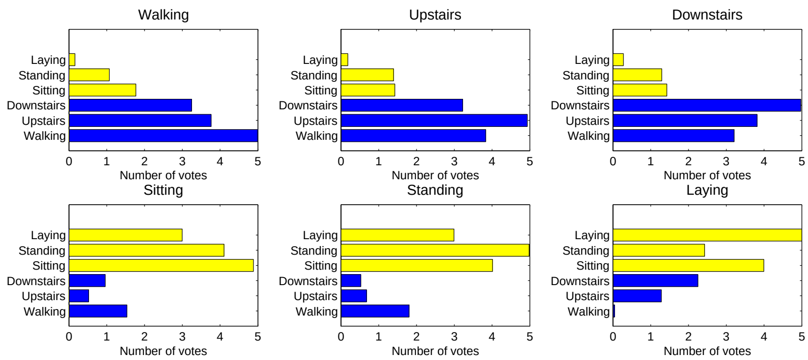

Fig. 2: The histograms show the mean voting distribution output by the OVO framework for the test set. Each subplot contains the vote breakdown for the instances in each class. The dark shades indicate walking related classes, the lighter shade indicates non-walking related classes.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

The confusion matrices in Table 1 show the outcomes for the OVO, OVA and $k-$NN methods respectively taken from the competition test set of 2947 instances[1]. The abbreviated column labels indicate the classes as follows: Wa - *walking*, Up - *walking upstairs*, Do - *walking downstairs*, Si - *sitting*, St - *standing* and La - *laying*.

Activity	Wa	Up	Do	Si	St	La	*Accuracy %*
Walking	**493**	0	0	3	0	0	99.40
Upstairs	28	**443**	0	0	0	0	94.06
Downstairs	2	6	**412**	0	0	0	98.10
Sitting	0	2	0	**433**	56	0	88.19
Standing	0	0	0	9	**523**	0	98.31
Laying	0	0	0	0	0	**537**	100
Precision %	94.26	98.23	99.28	97.96	90.33	100.00	**96.40**
Activity	Wa	Up	Do	Si	St	La	*Accuracy %*
Walking	**496**	0	0	0	0	0	100.00
Upstairs	41	**430**	0	0	0	0	91.30
Downstairs	37	2	**381**	0	0	0	90.71
Sitting	8	2	0	**448**	33	0	91.24
Standing	3	0	0	37	**492**	0	92.48
Laying	0	0	1	1	20	**515**	95.90
Precision %	84.79	99.08	99.74	92.18	90.28	100.00	**93.72**
Activity	Wa	Up	Do	Si	St	La	*Accuracy %*
Walking	**486**	0	10	3	0	0	97.98
Upstairs	36	**427**	8	0	0	0	90.66
Downstairs	51	38	**331**	0	0	0	78.81
Sitting	0	4	0	**409**	78	0	83.30
Standing	0	0	0	47	**485**	0	91.17
Laying	0	0	0	2	2	**533**	99.26
Precision %	84.82	91.04	94.84	89.30	85.84	100.00	**90.63**

Table 1: Confusion matrix for the One-Vs-One SVM framework (top), One-Vs-All SVM framework (middle) and k-NN model (bottom).

5 Discussion

The distributions of votes as shown in figure 2 generally follow the expectation described in the rationale. For each case the true class repeatably receives 5 votes, the false similar classes $3-4$ votes and the dissimilar classes less than 2 votes. The only exception to this trend being the votes for *walking downstairs* when the true class is *laying*, here a dissimilar class received more than 2 votes.

[1]The ground truth labels for the competition test set had been released by the time of this paper's composition but were not released during the development phase. Therefore the results of this unseen test set are given instead of the mean of the iterated 352 evaluation instances during the training.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

However even in this case the votes for the true class were unaffected. The confusion matrices support the hypothesis of separability between the dissimilar classes, with relatively very few false positives between them for all three frameworks. As expected significantly more false positives exist between classes that are similar, mostly notably between *sitting* and *standing;* and also between the walking classes. Though in these cases overall the OVO ensemble performs better compared to the OVA system and $k-$NN, thus leading to highest overall accuracy score of 96.4%.

The release of the AR data-set will facilitate many further investigations in both ML system development and also in the understanding of human movement. An interesting factor not considered in the competition and this study is the difference between the subjects. The participant identifiers were not initially released for the purpose of the challenge but will be useful in determining features that are idiosyncratic and those that are common to the cohort. Previous works by [7, 5, 6] specifically investigate these factors and utilize them in the development of motion recognition systems. To test this the AR data-set could be split by subjects and evaluation could be done on unseen subject specific instances.

Acknowledgements

The authors of this study are funded by the Engineering and Physical Sciences Research Council, UK (EP/H017178/1); part of the Pain rehabilitation: E/Motion-based automated coaching project (www.emo-pain.ac.uk).

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)*, Vitoria-Gasteiz, Spain, December 2012.

[2] M. S. H. Aung, S. B. Theis, P. J. Kenney, L., D. Howard, R. W. Selles, A. H. Findlow, and J. Y. Goulermas. Automated detection of instantaneous gait events using time frequency analysis and manifold embedding. *IEEE in Transactions on Neural Systems & Rehabilitation Engineering*, In press, 2013.

[3] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge University Press, 2000.

[4] D. T. H. Lai, R. K. Begg, and M. Palaniswami. Computational Intelligence in Gait Research: A Perspective on Current Applications and Future Challenges. *IEEE Transactions on Information Technology in Biomedicine*, 13(5):687–702, SEP 2009.

[5] B. Romera-Paredes, A. Argyriou, Bianchi-Berthouze N., and M. Pontil. Exploiting Unrelated Tasks in Multi-Task Learning. *Artificial Intelligence*, XX, 2012.

[6] B. Romera-Paredes, M. S. H. Aung, M. Pontil, P. Watson, and Bianchi-Berthouze N. FG 2013 Willimans, A. C de C. Transfer learning to account for idiosyncrasy in face and body expressions. *Automatic Face and Gesture Recognition (FG)*, 10th, 2013.

[7] M.a.O. Vasilescu. Human motion signatures: analysis, synthesis, recognition. *Object recognition supported by user interaction for service robots*, 3:456–460, 2002.