Bare Demo of IEEEtran.cls for Computer Society Journals

Miguel Perez-Xochicale

Abstract—The following template is considered for future submission of my research. I believe that any advance in science should be freely released in order that anyone who is interested in the idea can reproduce the current advances. To share, to modify and to improve are the bacis principles for pushing the the boundaries of any discipline in science.

Index Terms—Activity Recognition; On-Body Inertial Sensors; Motor Skill Assessment; Human-Robot Interaction

1 INTRODUCTION

ACCORDING to Bulling *et al.* [1] the common challenges in HAR using body-worn sensors are: (i) *intraclass variability* which occurs when an activity is performed differently either by a single person or several people. For example, gait patterns may be more dynamic in the morning after sleep than in the evening after a day full of activities; (ii) *interclass similarity* occurs when the sensor data is very similar. For example, in recognising dietary activity, drinking water or coffee entails the same arm movements [2]; and (iii) *the NULL class problem* occurs when ambiguous activities are irrelevant for the recognition methods which leads to wrong classification of the activities [3]. Recently, Bulling *et al.* [1] reviewed advances in the Activity Recognition Chain (ARC) using body-worn sensors. The general ARC comprehend five stages (data acquisition, signal preprocessing, segmentation, feature extraction and selection training and classification) of which the applied technique in each stage depends on the activity to recognise.

Given the case of the variability in dance activities, it is hypothesised that there are three possible reasons for variation: (i) inherent noise in body-worn sensors, (ii) inherent properties of the activity itself and (iii) differences in the people performing the activity, e.g., gender, anthropometry or level of skills.

1.1 Research Questions

For this PhD, the time-delay embedding and PCA methods have proven to be a reliable method for feature extraction in HAR [4], [5], it is therefore hypothesised that these methods might be suitable to learn the variability of human activities. Therefore, the following research questions will be addressed:

1) In the light of limitations of time-delay embedding and PCA, which other non-linear analysis tools would be suitable to explore
2) The variability in different human activities and use them as a features for machine learning algorithms?

2 TIME-DELAY EMBEDDING

The aim of the time-delay embedding, also known as Takens Theorem [6], is to reconstruct a $k-$dimensional manifold M of an unknown dynamical system $s(t)$ from a time series $x(t)$ with discrete observations at given time points t. Time-delay embedding assumes that the time series is a sequence $x(t) = h[s(t)]$, where $h : M \to \mathbb{R}$ is a measurement function in the unknown dynamical system, being $x(t)$ measurable.

Thus, the time delay reconstruction is defined as: $\overline{x}(t) = (x(t), x(t-\tau), ..., x(t-(m-1)\tau))$ where m is the embedding dimension and τ is the embedding time-delay. $\overline{x}(t)$ defines a map $\Phi : M \to \mathbb{R}^m$ such that $\overline{x}(t) = \Phi(s)$. Similarly, $y(t) = \Psi[\overline{x}(t)]$ is a n-dimensional vector where $\Psi : \mathbb{R}^m \to \mathbb{R}^n$ is a further transformation. Figure 1 illustrate the time delay reconstruction process. For details, see the work of Uzal *et al.* [7].

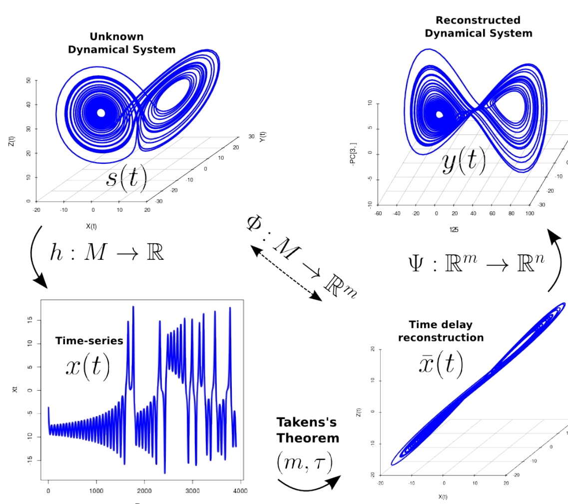

Fig. 1. The reconstruction problem. The figure is based on the work of Uzal *et al.* [7].

2.1 Embedding Parameters m and τ

Given any time series $x(t)$, the time delay reconstruction system, $\overline{x}(t)$, is easy to implement. For this work, Cao's method [8], a modification of the False Nearest Neighbours (FNN) algorithm, and mutual information algorithm by

2.1.1 Minimum Embedding Dimension m_{min}

Cao's method [8] for computing the minimal embedding dimension is based on the mean values $E1(d)$ and $E2(d)$ in which d is a given embedding dimension value.

$E1(d)$ is used to obtain the minimal dimension m_{min} and stops changing when the time series comes from an attractor. We computed $E1(d)$ values for $1 \leq \tau \leq 10$ to exemplify the minor dependency of τ given periodic, chaotic and random time series.

The second of these values, $E2(d)$, is used to distinguish deterministic signals from random signals in which case the $E2(d)$ values will be approximately equal to 1 for any d. Similarly, we computed $E2(d)$ values for periodic, chaotic and random time series, to exemplify the no significant dependency on τ, where $1 \leq \tau \leq 10$.

Cao's method is a modified version of the FNN method, and $E1(d)$ and $E2(d)$ values are only dependant on m and τ [8].

2.1.2 Minimum Time-delay Embedding τ_{min}

The method of choosing the minimum Time-delay embedding, τ_{min}, was proposed by Fraser *et al.* in which the first minimum of the mutual information graph is chosen to estimate the minimal time-delay embedding parameter. The local minimum for the Chaotic series is $\tau_{min} = 18$. On the other hand, for random time series the mutual information plot have no local minimum and values are monotonically decreasing which means that $\tau_{min} = 1$. However, further research has to be done when data comes from a periodic time series since its minimum in the mutual information plot appears to be at $\tau_{min} = 3$.

3 The Activity Recognition Chain

Bulling *et al.* [1] reviewed the state of the art of HAR using body-worn inertial sensors.

The first stage of the ARC is the raw data collection from several sensors attached to different parts of the body. Sensors data over a given time, s_i, provide multiple values d^i, (e.g. d^1, d^2, d^3 for 3-D acceleration referred to as x, y and z direction)

$$s_i = (\mathbf{d}^1, \mathbf{d}^2, \dots \mathbf{d}^t),\text{for } i = 1, \dots, k \tag{1}$$

where k denotes the number of sensors.

In the preprocessing stage of the ARC, raw multivariable time series are transformed into a pre-processed time series $D' = (d'_1, \dots, d'_n)^T$, where d'_i is one dimension of the data for the preprocessed time series and n is the number of total data dimensions. Different methods for the preprocessing tasks may be applied to the raw data (e.g. synchronisation, calibration, unit conversion, normalisation, resampling, denoising or baseline drift removal [1]).

The stage of data segmentation identifies segments within the continuous data stream that are likely to have information about activities. The segmentation stage creates a set of segments w_m such that

$$W = \{w_1, \dots, w_m\}, \tag{2}$$

where m correspond to the number of segments. Since the segmentation of the data is a difficult problem, there are various methods in the literature to tackle this problem: sliding window, energy-based segmentation, rest-position segmentation, additional sensors and external context sources.

In the feature extraction stage, a feature extraction function F reduces the signals D' into segmented signals W. The total number of features X_i is the feature space.

$$X_i = F(D', w_i) \tag{3}$$

In the literature on activity recognition, different methods for feature extraction can be found including signal-based features, body model features, event-based features, multilevel features or automatic feature ranking and selection.

Machine learning tools have been used in HAR over the last 15 years so as to describe, analyse and predict human activities [1]. However, the chosen approach is subject to computational complexity, recognition performance or latency. Generally for the learning stage, a training data set $T = \{X_i, y_i\}_{i=1}^N$ is computed prior to the classification with N pairs of feature vectors X_i and ground truth labels y^i (possible activities to recognise). For this stage, model parameters θ can be learned to decrease the classification error on T. Then, with the trained model T, each feature vector X_i is mapped to a set of class labels $Y = \{y^1, \dots, y^c\}$ with scores $P_i = \{p_i^1, \dots, p_i^c\}$:

$$p_i(y \mid X_i, \theta) = I(X_i, \theta),\text{for } y \in Y \tag{4}$$

and inference method I. Finally, the classification output y_i is computed with the maximum score P_i

$$y_i = \operatorname*{argmax}_{y \in Y, p \in P_i} p(y|X_i, \theta) \tag{5}$$

The most common classification algorithms are: decision trees, Bayesian models, domain transform, fuzzy logic, Markov models, support vector machines (SVM), artificial neural networks (ANN) and ensembles of classifiers [9].

Similarly, when the recognition of activities can miss, confuse or falsely recognise activities that did occur, several metrics can be used to optimise the classification. Some of the metrics are confusion matrices, accuracy, precision, recall, and F-scores, decision-independent Precision-Recall or receiver operating characteristic curves (ROC curves) [1].

4 Artificial Signals

To understand the possible sources of variability in dance activities, I followed the method of Hammerla *et al.* [10] which might be useful to model: i) noise in sensors, ii) properties of activities and iii) properties of people.

The proposal of Hammerla *et al.* [10] is aimed to examine the effects of variability in the precision of motion (additive noise) and in the strategy of motion (structural noise) of activities using artificial signals.

Additive noise is normalised noise with variance σ_a^2 added to the sinusoid signal S:

$$S^a = S + \mathbf{N}(0, \sigma_a^2) \tag{6}$$

Structural noise is a sinusoid signal distorted with different variance in frequency and amplitude σ_s^2 and window length w_s. Algorithm 1 describes the creation of structural noise. To make the data less redundant for possible variations of environmental conditions or body-worn sensor mobility in users, the data is whitened (i.e. data is normalised to have zero mean and unit variance)

Algorithm 1 Structural Noise

Input: time-series S^a, variance σ_s^2, window length w_s
Output: Structurally distorted signal S^s
for $j = 1$ to L, $j = j + w_s$ **do**
  $\mathbf{u'} \leftarrow \mathbf{N}(0, \sigma_s^2)$
  S^a = sinusoid with frequency $|\mathbf{u'}|$ and variance σ_a^2 of length w_s
  $S^s_{j \rightarrow j+w_s} = S^s_{j \rightarrow j+w_s} + S^a \times \sigma_s^2$
end for
$S^s = whiten(S^s)$
return S^s

By varying both σ_a^2 and σ_s^2, it is possible to simulate and control the additive noise and the structural noise in the structure of the human activity. For example, low values of σ_a^2 are associated with precise movements while low values of σ_s^2 correspond to a well chosen strategy for a motion.

5 Future Work

To raise the bar in the field of human activity recognition, the plan for the next six months is:

- **Sep. 2015 (11th)** Using the sawing data, present the limitations of time-delay embedding and PCA and propose improvements for the method. I am also planning to investigate other non-linear analysis tools that would be suitable to explore the variability of dance activities.
- **Oct. 2015 (12th)** Work towards a submission in Measuring Behavior 2016 and Augmented Human 2016 conferences.
- **Nov. 2015 (13th)** Submit works in the Measuring Behavior 2016 and Augmented Human 2016 conferences.
- **Dec. 2015 (14th)** Search for an appropriate journal and work towards a submission.
- **Jan. 2016 (15th)** Submit a journal publication

Acknowledgments

Miguel Perez-Xochicale gratefully acknowledges the studentship from the National Council for Science and Technology (CONACyT) Mexico from November 2014 to November 2017 to pursue his postgraduate studies at University of Birmingham.

References

[1] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn intertial sensors," *ACM Computing Surveys*, pp. 1–33, 2014.
[2] O. Amft, "Automatic dietary monitoring using on-body sensors: Detection of eating and drinking behavior in healthy individuals," Ph.D. dissertation, ETH Zurich, 2008.
[3] ——, "Self-taught learning for activity spotting in on-body motion sensor data," in *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*, June 2011, pp. 83–86.
[4] J. Frank, S. Mannor, and D. Precup, "Activity and Gait Recognition with Time-Delay Embeddings Time-Delay Embeddings," *AAAI Conference on Artificial Intelligence*, pp. 407–408, 2010.
[5] A. Samà, F. J. Ruiz, N. Agell, C. Pérez-López, A. Català, and J. Cabestany, "Gait identification by means of box approximation geometry of reconstructed attractors in latent space," *Neurocomputing*, vol. 121, pp. 79–88, 2013.
[6] F. Takens, "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence, Warwick 1980*, vol. 898, pp. 366–381, 1981. [Online]. Available: http://dx.doi.org/10.1007/bfb0091924
[7] L. C. Uzal, G. L. Grinblat, and P. F. Verdes, "Optimal reconstruction of dynamical systems: A noise amplification approach," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 84, no. 1, 2011.
[8] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, pp. 43–50, 1997.
[9] O. D. Lara and M. a. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
[10] N. Hammerla, T. Ploetz, P. Andras, and P. Olivier, "Assessing motor performance with pca," *International Workshop on Frontiers in Activity Recognition using Pervasive Sensing*, pp. 18–23, 2011.