A New Method of Dimensionality Reduction for Large Time Series Applied to Accelerometer Wristbands' Signals

Alihuén García Pavioni[1], Beatriz López Ibáñez[1]
[1] *Exit Grup, University of Girona, Catalonia, Spain*
alihuen.garcia@udg.edu, beatriz.lopez@udg.edu

Keywords: accelerometers wristbands, high-dimensional time series, time series classification, dimensionality reduction, feature extraction, behavior recognition, signal state changes

Abstract: Feature extraction for high-dimensional time series has become a topic of great importance in recent years. In the medical field, the information needed to predict emotions, stress, epileptic seizures, heart attacks, and other diseases, can be provided by body sensors in the form of time series signals. This work intends to provide a way for these devices to save the relevant information, using little storage memory, by defining a new feature extraction method. The method proposed in this work relies on the relevant data associated with the "changes" in the time series. These changes are identified according to the conditional probabilities of passing from one state to another during the time series, as well as the "relevance" of each state. We show the results of this method with an experiment based on accelerometers data recorded by the ©ActiGraph wGT3X-BT wristband to recognize sedentary behavior. After applying this method, it was achieved to reduce time series frames of dimension 360, to vectors of dimension 12; while the accuracy of their classification was 84%.

1 INTRODUCTION

In the medical field, the study of selecting relevant data from the time series has become of fundamental importance in the last years: the electrocardiogram, the respiration, the skin temperature, the blood volume pulse, and the electrodermal activity are all represented by time series, and provide the information needed to predict emotions, stress, epileptic seizures, heart attacks, and other diseases (Montesinos et al., 2019; Shoeb and Guttag, 2010; Wang et al., 2014; Ravish et al., 2014). The usage of new technologies for measuring these body signals, such as smart wristbands, makes it necessary to reduce the cost of storage without losing the relevant data.

This work intends to provide a way for these devices to save the relevant information, using little storage memory, by defining a new feature extraction method, called *State Changes Representation for Time Series* (SCRTS). Different to other methods in the literature, the SCRTS, relies on the relevant data associated with the "changes". These changes are identified according to the conditional probabilities of passing from one state to another during the time series, as well as the "relevance" of each state, which provide the information needed to represent or characterize a time series in several contexts.

To test our method, we conduct an experiment with people in our lab, trying to determine when they were in the office or when they were not, based on their behaviors described through the time series provided by the usage of ©ActiGraph wGT3X-BT wristbands accelerometers.

2 LITERATURE SURVEY

Time series feature extraction (TSFE) is essential for machine learning effectiveness in time series problems, on the one hand because it reduces the dimension of the feature space, and on the other, because it has a significant impact on the final results, since it transforms the input data in vectors easier to interpret.

There exist many different TSFE methods. The Fourier Transform (Bracewell and Bracewell, 1986) and the Wavelet Transform (Shensa, 1992) could be very useful applied to time series composed mostly of periodic waves, like it happens with the ECG signals, or other signals related to the light, the electricity, the image, or the sound, among others. But when it comes to analyse time series that does not present a periodic behavior, such as the data extracted from an accelerometer, this methods may not work well.

There exist other classical statistical methods for feature extraction like the Singular Value Decomposition (SVD)(Cadzow et al., 1983), the Principal Component Analysis (PCA) (Wold et al., 1987), or the Linear Discriminant Analysis (LDA) (Izenman, 2013). These methods use Linear Algebra tools for reducing the information of the input matrix. They work well for static data, but when it comes to time series, it may not work properly. We are interested in selecting the features that contain information about the changes, and we consider that this information is in its conditional probabilities related to the states.

In (Zhou and Chan, 2015) a method called Multivariate Time Series Classifier (MTSC) is proposed, using conditional probabilities to create a measure that allows to discover some intra and inter-temporal patterns. In this work we make use of the conditional probabilities as well, but instead of seeking for these intra and inter-temporal patterns, we seek for other features that could give a description of how much the signal has changed inside of each period, relating these changes with the activity made by the wristband's user.

In the particular case of feature extraction for classifying accelerometer signals, in (Preece et al., 2008) is presented a comparison of 14 methods based on the wavelet transform and other well-known time- and frequency-domain signal characteristics. In (Long et al., 2009) is compared a Bayesian classification with that of a Decision Tree based approach. In (Mathie et al., 2004) is presented a framework structured around a binary decision tree in which movements were divided into classes and sub-classes at different hierarchical levels.

The method proposed in this work, the SCRTS, divide the range of possible time series' values into different "states", and extract the information of how these states "change" (that is to say, which values take the conditional probabilities), and how much importance, or "weight", these states had in the frames that we want to classify.

3 METHODOLOGY

Fig. 1 shows the different steps of the methodology, detailed in this section.

3.1 Data Collection

A time series is a succession of data measured in time and arranged chronologically. As we made use of accelerometers to obtain the data, we refer to the values of the time series as *vector magnitudes*, which is the original name used, and we refer to a vector magnitude value as VM. We refer as τ to the time frequency with which these vectors magnitudes are displayed by the device.

3.2 Division Into Frames

Every time signal was divided in frames of T minutes. So far, every frame is represented by a vector of dimension d equal to the number of samples in it. Then, we denote each frame as a vector F such that:

$$F = (VM_i)_{i \leq d}. \tag{1}$$

Therefore, in the experiment presented in this work, as every value is given every ten seconds (i.e., $\tau = 10$ seconds), if for example we choose our $T = 60$ minutes, we have that each frame has a dimension of $d = 360$.

3.3 Discretization

There are many different methods for *time series temporal discretization* (Moskovitch and Shahar, 2015; Azulay et al., 2007), which involves to obtain a sequence of *states* from a numerical time series. These states are domain dependent. This means that, if we look at the time series, with axis y being the vector magnitude values and axis x the time, then the y axis is divided by several *cut points* into n intervals, each of them representing a state.

Formally, given a set of cut points

$$CP = \{cp_0, cp_1, \ldots, cp_n\}, \tag{2}$$

we can generate a set Σ of n states, each state $S_i \in \Sigma$ representing an interval of the vector magnitudes values made by the cut points, as follows:

$$\begin{aligned} S_1 &= [cp_0, cp_1), \\ S_2 &= [cp_1, cp_2), \\ &\vdots \\ S_n &= [cp_{n-1}, cp_n], \end{aligned} \tag{3}$$

where cp_0 and cp_n are the time series' minimum and maximum values respectively.

We say that a vector magnitude VM *is of state* S_i, when $VM \in S_i$, that is to say, when $cp_{i-1} \leq VM \leq cp_i$. Consistently, given a set of states Σ, we can represent each frame F by a sequence of states

$$S(F) = \{S^1, S^2, \ldots, S^d\}, \tag{4}$$

where $S^t \in \Sigma$, and the supra-index t indicates the chronological position of the state in the frame.

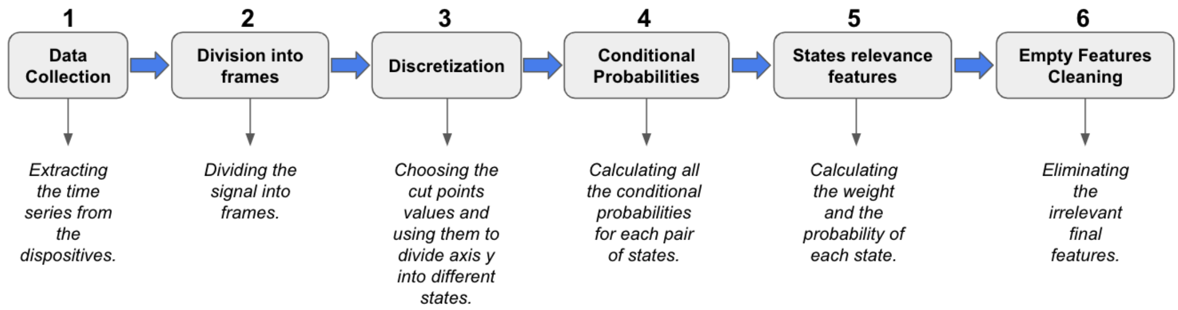

Figure 1: SCRTS steps.

3.4 Conditional Probabilities

Given a frame F of our time-signal represented by $S(F) = \{S^1, S^2, \ldots, S^d\}$, the *conditional probability* of getting state $S^{t+1} = b$ after being in state $S^t = a$, with $a, b \in \Sigma$, is defined as

$$\operatorname{Prob}(S^{t+1} = b \mid S^t = a) = \frac{\operatorname{Frec}(a,b)}{\operatorname{Frec}(a)}, \tag{5}$$

such that $1 \leq t \leq d-1$, being $\operatorname{Frec}(a,b)$ the number of times that a is followed by b in $S(F) = \{S^1, S^2, \ldots, S^d\}$, and $\operatorname{Frec}(a)$ the number of times that a appears in $\{S^1, S^2, \ldots, S^{d-1}\}$.

Therefore, we calculate every conditional probability for each frame, which gives us a total of n^2 features per frame in each case. Thus, we use these features for making a new vector for representing the information contained in frame F. We call $C(F)$ to refer to the vector of all the conditional probabilities of F.

$C(F)$ reflect the "jumps" from one state to another, giving a description of the changes or the "stays", showing which jumps were more common in F, and which states stay longer without changing.

3.5 States Relevance Features

But $C(F)$ does not give any information about the states "relevance" in the time series, or which of the states have appear a greater number of times. Though, if we want to create a vector that contains most of the state's changes relevant features, then we should probably include relevant data regarding to the states appearance in the frame. To that end, we make usage of two features: the *state probability*, and the *state weight*.

3.5.1 State Probability

For each $S_i \in \Sigma$, the *probability of state* S_i to come out in frame F is:

$$P(S_i) = \frac{\operatorname{Frec}(S_i)}{d}. \tag{6}$$

Therefore, we refer to the set of all the state probabilities of a frame F as $P(F)$, that is to say

$$P(F) = \{P(S_1), P(S_2), \ldots, P(S_n)\}. \tag{7}$$

3.5.2 State Weight

As we already said in (3), for every state $S_i \in \Sigma$, we have a cp_{i-1} and a cp_i indicating the rang for a vector magnitude VM to be labelled as state S_i. Then, we define the *midpoint* of each state $S_i \in \Sigma$ as

$$\operatorname{mid}_i = \frac{(cp_i + cp_{i-1})}{2}. \tag{8}$$

The *distance from the midpoint to the top* of state S_i is

$$\operatorname{dis}_i = |\operatorname{mid}_i - cp_i| \ \left(= |\operatorname{mid}_i - cp_{i-1}|\right). \tag{9}$$

Now, for every VM belonging to a state S_i we can define the *normalized inverted distance* of VM to it's respective midpoint mid_i of S_i, as

$$\operatorname{NID}_i(VM) = \frac{-|\operatorname{mid}_i - VM|}{\operatorname{dis}_i} + 1. \tag{10}$$

For the reader familiar with statistics, the normalized inverted distance is similar to the z-score function (Kreyszig, 2009). The difference is that the normalized inverted distance is like making $1 - $z-score, but instead of dividing by the standard deviation, in the normalized inverted distance we divide by the distance from the midpoint to the top of its respective state interval.

The importance of the NID_i is that it gives and idea of how "weighted" VM is for state S_i. If VM lies in the midpoint of the state S_i, the NID_i is 1, which is the maximum value possible; and the further it lies from the midpoint, the lower the NID_i is, being 0 at the lower and upper values of S_i that is to say,

$$\operatorname{NID}_i(\operatorname{mid}_i) = 1; \tag{11}$$

$$\mathrm{NID}_i(\mathrm{cp}_i) = \mathrm{NID}_i(\mathrm{cp}_{i-1}) = 0. \qquad (12)$$

Thus, if we sum all the NID_i's of all the vector magnitudes laying in a state $S_i \in \Sigma$ for a frame F, and we normalize the result, then we have a notion of how much "weight" or relevance has S_i in F. Let's say that $Q(S_i) = \{VM_1, VM_2, \ldots, VM_q\}$ is the set of all the vector magnitudes of the frame F laying in state S_i, then, we define the *weight* of state S_i in F as

$$W(S_i) = \begin{cases} \dfrac{\sum_{j=1}^{q} \mathrm{NID}_i(VM_j)}{d}, & \text{if } Q(S_i) \neq \emptyset; \\ 0, & \text{if } Q(S_i) = \emptyset; \end{cases} \qquad (13)$$

where d is the amount of vector magnitudes in $F = (VM_j)_{j \leq d}$. We refer to all the state weights of a frame F as $W(F)$, that is to say

$$W(F) = \{W(S_1), W(S_2), \ldots, W(S_n)\}, \qquad (14)$$

with n being the number of states in Σ (as we already said in (3)).

Though, the dimension of these vectors depend on the number of states. Let's call $dim(V)$ to the function that returns the dimension of a vector V, then

$$dim(C(F)) = n^2; \qquad (15)$$
$$dim(P(F)) = dim(W(F)) = n. \qquad (16)$$

Finally, we call as the *representation vector* $R(F)$ to the vector containing the features selected to represent F according to our method. These features are: $C(F)$, $P(F)$ and $W(F)$. Though, the dimension of $R(F)$ is

$$dim(R(F)) = n^2 + 2n. \qquad (17)$$

3.6 Empty Features Cleaning

The SCRTS is, as the name says it, a method for representing time series. This is to say that our goal is to extract all the relevant data of all the frames involved so they can be used as a matrix for a machine learning algorithm. This matrix has every $R(F)$ of each frame F as rows. So each column of the matrix represents a different feature of the frames. Therefore, there is a column for each conditional probability, one for every weight, etc. If one of these columns has more than a 75% of zeros means that the features of the column are not relevant for representing the time series and it could bring some noise for the machine learning performance, then we delete it. This process will reduce the dimension of the training-test matrix even more. As a result, the representation vector of each frame will reduce its dimensionality.

4 EXPERIMENTAL SETUP

To test our method, we conduct an experiment with people in our lab, trying to determine when they were in the office or when they were not, based on their behaviors described through the time series provided by the usage of ©ActiGraph wGT3X-BT wristbands accelerometers.

Eight PhD's and master students working at the same office in the University of Girona, wore these wristbands on their skillfully wrist for a week, which were programmed to record a measure every ten seconds (i.e., $\tau = 10$ seconds).

The subjects were asked to take note of their office check-in and check-out times for each day using the wristband. Next, when dividing the signal into frames of T minutes duration, each frame was labeled with 1 if the subject was more than half of that time at the office, or with 0 otherwise.

We chose Freedson Adult 1998 cut points provided by ActiLife (Freedson et al., 1998) for the discretization, which give us a total of 5 states (i.e., $n = 5$) [1].

4.1 Classification

The classification of frames between classes $\{0,1\}$ was made using sequential Artificial Neural Networks (ANN) [2]. It had 8 hidden layers of 12 nodes and the Relu activation function in each layer. The output layer was a dense layer with 2 nodes and the Sigmoid activation function. The optimizer was Adam with a learning rate of 0.001 units; the loss function was the binary crossentropy and the number of epochs was 40.

All the data frames were randomly shuffle together and split into the training set (75%) and the test set (25%). We applied random oversampling (Ling and Li, 1998) to level the quantity of frames in the training set labeled with 1 with the ones labeled with 0. The accuracy, the true positive rate (TPR) and the true negative rate (TNR), were calculated using the test set. This procedure was executed 20 times, and the final results (presented in next section), were calculated as the average of the results obtained in each of the 20 performances.

[1] We also tried with the cut points provided by Actilife called *Freedson Adult VM3 2011, Trost Toddler 2011,* and *Troiano 2008*, but Freedson Adult 1998 were the ones which gave us better results in this experiment.

[2] We also tested other architectures, as LSTM, with similar results.

5 RESULTS

We applied the SCRTS to the data of all the wristbands together. The results achieved have been compared with the ones obtained from the same physiological signals without any feature extraction method, that is, using the raw data directly. Different frame lengths were explored: $T = 15$, $T = 30$ and $T = 60$ minutes. The results with the final dimension of the vectors representing the frames for the classification (Dim.), the accuracy (Acc.), the true positive rate (TPR) and the true negative rate (TNR) are provided in Table 1.

The first aspect to notice is that working with the SCRTS gave much better results than working with the raw data, specially for long periods. One other thing is that the dimensionality reduction with the SCRTS is considerable compared to the raw data. In the best case ($T = 60$), the dimension of each frame was reduced from 360 to 12 with the SCRTS, while the accuracy was 84%, the TPR 81%, and the TNR 84%.

In Table 2 it is shown the results obtained after applying the SCRTS to the data of each wristband (W) individually, with $T = 60$ minutes.

It could be seen that the SCRTS also works well in the classification of the time series individually.

6 CONCLUSIONS AND FUTURE WORK

In this work we proposed a new method for dimensionality reduction, the SCRTS, based on representing how the signal information changes according to different states. In particular, state changes are modeled with conditional probabilities, state probabilities and *state weights*. This method has been shown to work very well in a long time series classification problem. The classification for 60 minutes frames gave an accuracy of 84%, a TPR of 81%, and a TNR of 85%, while showing a lot of effectiveness for storage, since it reduced the original data of dimension 360, to a vector of dimension 12. The difference between doing the classification with the row data and doing the classification with the data after applying the SCRTS, shows that this technique not only reduces the cost of storage, but also works considerably better for classifying long time periods.

This experiment was done with accelerometers wristbands, which return one-channel time series per user. In futures works we will try this technique with the data collected by other wearable devices measuring other body features, such as the electrocardiogram, the respiration, the skin temperature, the blood volume pulse or the electrodermal activity. These wearable devices return multi-channel time series for each user, which add some complexity to our technique that require some new treatments.

ACKNOWLEDGEMENTS

This work was carry out with the support of the Generalitat de Catalunya 2017 SGR 1551, and funded by the Grants for the recruitment of new research staff (FI), provided by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR).

We also thanks to the collaboration and advice of the members of the ExiT Grup from the University of Girona, to which the authors of this work belong.

REFERENCES

Azulay, R., Moskovitch, R., Stopel, D., Verduijn, M., De Jonge, E., and Shahar, Y. (2007). Temporal discretization of medical time series-a comparative study. In *IDAMAP 2007 workshop*.

Bracewell, R. N. and Bracewell, R. N. (1986). *The Fourier transform and its applications*, volume 31999. McGraw-Hill New York.

Cadzow, J. A., Baseghi, B., and Hsu, T. (1983). Singular-value decomposition approach to time series modelling. In *IEE Proceedings F (Communications, Radar and Signal Processing)*, volume 130, pages 202–210. IET.

Freedson, P., Melanson, E., and Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Medicine & science in sports & exercise*, 30(5):777–781.

Izenman, A. J. (2013). Linear discriminant analysis. In *Modern multivariate statistical techniques*, pages 237–280. Springer.

Kreyszig, E. (2009). Advanced engineering mathematics, 10th eddition. Accessed: 26-04-2021.

Ling, C. X. and Li, C. (1998). Data mining for direct marketing: Problems and solutions. In *Kdd*, volume 98, pages 73–79.

Long, X., Yin, B., and Aarts, R. M. (2009). Single-accelerometer-based daily physical activity classification. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6107–6110. IEEE.

Mathie, M., Celler, B. G., Lovell, N. H., and Coster, A. (2004). Classification of basic daily movements using a triaxial accelerometer. *Medical and Biological Engineering and Computing*, 42(5):679–687.

Montesinos, V., Dell'Agnola, F., Arza, A., Aminifar, A., and Atienza, D. (2019). Multi-modal acute stress recognition using off-the-shelf wearable devices. In

Table 1: Results comparison using the SCRTS and the raw data

	T (min.)	Dim.	Acc. (%)	TPR (%)	TNR (%)
SCRTS	15	8	79	83	78
	30	8	81	82	81
	60	**12**	**84**	**81**	**85**
Raw Data	15	90	81	42	86
	30	180	80	42	86
	60	360	16	9	97

Table 2: Results of SCRTS applied to each wristband individually

	W1	W2	W3	W4	W5	W6	W7	W8
Acc. (%)	84	77	84	84	89	83	89	78
TPR (%)	56	91	81	96	98	94	67	89
TNR (%)	86	74	84	82	87	80	96	77

2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 2196–2201. IEEE.

Moskovitch, R. and Shahar, Y. (2015). Classification-driven temporal discretization of multivariate time series. *Data Mining and Knowledge Discovery*, 29(4):871–913.

Preece, S. J., Goulermas, J. Y., Kenney, L. P., and Howard, D. (2008). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, 56(3):871–879.

Ravish, D., Shanthi, K., Shenoy, N. R., and Nisargh, S. (2014). Heart function monitoring, prediction and prevention of heart attacks: Using artificial neural networks. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pages 1–6. IEEE.

Shensa, M. J. (1992). The discrete wavelet transform: wedding the a trous and mallat algorithms. *IEEE Transactions on signal processing*, 40(10):2464–2482.

Shoeb, A. H. and Guttag, J. V. (2010). Application of machine learning to epileptic seizure detection. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 975–982.

Wang, X.-W., Nie, D., and Lu, B.-L. (2014). Emotional state classification from eeg data using machine learning approach. *Neurocomputing*, 129:94–106.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.

Zhou, P.-Y. and Chan, K. C. (2015). A feature extraction method for multivariate time series classification using temporal patterns. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 409–421. Springer.