

# Human Activity Recognition with Wearable Devices: A Symbolic Approach

Angelo Cenedese, Luca Minetto, Gian Antonio Susto\* and Matteo Terzi

Dept. of Information Engineering  
University of Padova  
(Italy)

---

## ABSTRACT

In the context of activity recognition, wearable devices are nowadays the preferable hardware thanks to their usability, user experience and performances; at the same time, these devices present limitations in terms of computational capability and memory, which force the algorithm design to be at the same time efficient and simple. In this work, we adopt Symbolic Aggregate Approximation (SAX), a symbolic approach for information retrieval in time series data that allows dimensionality and numerosity reduction; SAX is employed here, in combination with 1-Nearest Neighbor classifier, to identify activity phases in continuous repetitive activities from inertial time-series data. The proposed approach is validated on a cross-country skiing dataset and on a daily living activities dataset.

---

Keywords: *Activity Recognition, Machine Learning, Symbolic Aggregate Approximation, Time-Series Learning, Wearable Devices*

Paper received 1/09/2016; received in revised form 28/10/2016; accepted 7/11/2016.

## 1. Introduction

Activity Recognition (AR) is a prominent research area with a wide range of applications to home automation (Belgioioso, Cenedese, Grillo, Fraccaroli & Susto, 2014) gaming (Gowing et al., 2014), sport (Cenedese, Susto & Terzi, 2016) and health care (Clifton, Clifton, Pimentel, Watkinson & Tarassenko, 2013) to cite a few. In particular, the rapid diffusion of IMUs (Inertial-Measurement Units) has allowed, in recent years, the development of compact sensor-equipped devices (e.g. smartwatches and smartphones), which lead efficient monitoring of human activities to be feasible and to have a strong impact on the quality of life (Clifton et al., 2013); on

---

Cite as:

Cenedese A., Minetto L., Susto G.A. & Terzi M. (2016) Human Activity Recognition with Wearable Devices: A Symbolic Approach. <i>PsychNology Journal</i> , 14(2-3), 99 – 115. Retrieved [month] [day], [year], from <a href="http://www.psychnology.org">www.psychnology.org</a> .
---

\*Corresponding Author:  
Gian Antonio Susto  
Department of Information Engineering, University of Padova,  
via Giovanni Gradonigo 6, 35131 Padova (Italy)  
E-mail: [gianantonio.susto@dei.unipd.it](mailto:gianantonio.susto@dei.unipd.it)

the other hand, wearable devices present some limitations in terms of computational capability and memory, which force the algorithm design to be, at the same time, efficient and simple (Cenedese, Susto, Belgioioso, Cirillo & Fraccaroli, 2015). In such systems, IMU-generated data may be obtained for example from accelerometers, gyroscopes and magnetometers.

In this context, the AR problem is strictly connected to Gesture Recognition (GR); in fact, activities may be seen as compositions of gestures that are executed in a continuous time window (Tran & Trivedi, 2012). Due to the vastness of application scenarios, it may be helpful to categorize AR problems into three main types:

- *continuous-repetitive* - activities that are continuous and composed by repeated gestures with a periodic behavior within the same activity type;
- *continuous-spot* - continuous activities with non-repetitive gestures;
- *isolated* - activities composed by isolated gestures.

This work is focused on the continuous-repetitive type (Morris, Saponas, Guillory & Klener, 2014), which is typical of sports (rowing and swimming for example) and health monitoring applications.

AR problems are usually tackled by Machine Learning (ML) approaches (Morris et al., 2014; Tran & Trivedi, 2012) and they are considered as classification problems: the activity or gesture in exam  $\tau_i$  has to be associated with one of the a-priori defined  $K$  possible classes of activities/gestures  $\tau = \{\tau_i\}_{i=1}^K$ . The main challenge in applying ML tools in AR problems is to translate the informative content contained in the IMU-generated time series into a static format that can be handled by ML classifiers (Ravi, Dandekar, Mysore & Littman, 2005); the aforementioned procedure is called Feature Extraction, a phase that may be time consuming and may lead to information loss (Susto, Schirru, Pampuri & Mclloone, 2016).

In this work, we employ Symbolic Aggregate approXimation (SAX) (Lin, Keogh, Lonardi & Chiu, 2003), a technique to efficiently map time-series into strings; SAX, used in combination with a simple classifier, like 1-Nearest Neighbor (Friedman et al. 2009), allows to directly deal with time-series inputs (without a Feature Extraction phase) and to have a parsimonious solution in terms of complexity.

The rest of the paper is organized as follows: in Section 2, the SAX-based AR solution will be presented; results are validated in Section 3 on a normal day activity dataset and on a Cross-Country Skiing dataset. Final remarks will be provided in Section 4.

## 2. AR Algorithm and SAX

The main phases of the proposed algorithm are listed below:

1. *Gesture Identification* - identification and isolation of a single gesture. There exist various approaches to deal with this task (Belgioioso et al., 2014); in this work, we focus on the following phases, assuming that a Gesture Identification routine is providing the activities separated in single gestures.
2. *Symbolic representation* - gesture is mapped to a string representation through SAX technique.
3. *Gesture classification* - the algorithm classifies the gesture exploiting the features extracted through the SAX procedure.
4. *Activity classification* - following a sliding window of length  $l_{SW}$ , the classification of the activity in each segment is done by majority vote starting from the classified gestures within that window.

Phases 2, 3 and 4 are described in more details in the following.

### 2.1 Symbolic Representation

In this phase, IMU-generated data are mapped into a symbolic domain using SAX to yield dimensionality and numerosity reduction (Lin et al., 2003), fundamental features for wearable AR solutions.

SAX has been used in several fields of application, such as classification and clustering problems applied on telemedicine time series (Androulakis, 2005), entomological problems (Kasetty, Stafford, Walker, Wang & Keogh, 2008), mechanical systems (Harris, 2013) and anomaly detection (Carbone, 2014).

SAX representation also allows distance measures to be defined on the symbolic domain. Specifically, SAX allows a time series of arbitrary length  $n$  to be reduced to a string of arbitrary length  $w$ , with  $w < n$ , and typically  $w \ll n$ ; the ratio of  $n$  to  $w$  is known as *compression rate*. The string of length  $w$  is composed by  $w$  characters from the alphabet set, whose size is also an arbitrary integer  $\alpha$ , where  $\alpha > 2$ .

The discretization procedure is unique: in order to transform the raw time series into the symbolic strings we need an intermediate representation; first, the data is transformed into the *Piecewise Aggregate Approximation (PAA)* representation and then the PAA representation is symbolized into a discrete string; the procedure will be detailed in the following Sections. There are two important advantages to doing this (Lin et al., 2003):

- **Dimensionality Reduction:** the dimensionality reduction power of PAA is the well-defined and well-documented (Tran & Trivedi, 2012; Yi & Faloutsos, 2000);
- **Lower Bounding:** proving that a distance measure between two symbolic strings is a lower bound for the true distance between the original time series. The key observation that allows to prove lower bounds is to concentrate on proving that the symbolic distance measure bounded from below the PAA distance measure.

In order to simplify the reading, Table 1 summarizes the major notation used in this section.

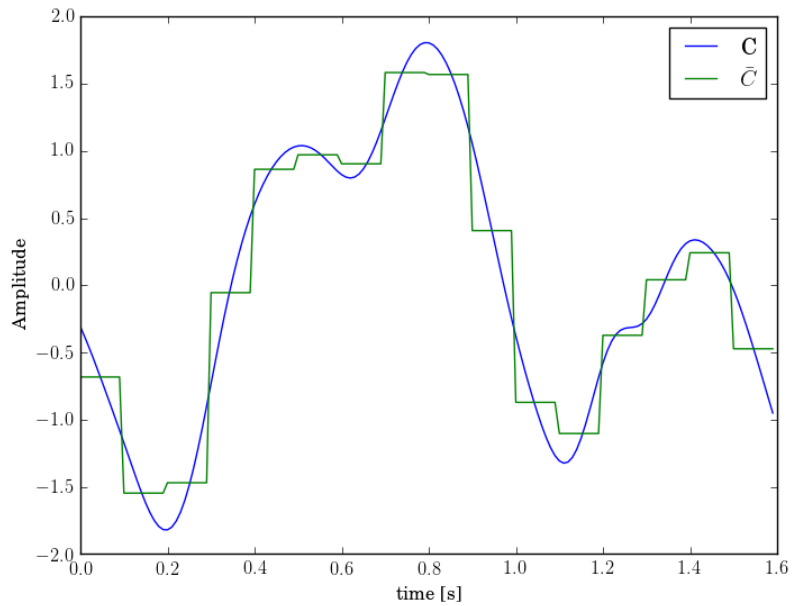
Notation	Meaning
$n$	Time-series length
$C = c_1, \dots, c_n$	Time-series
$w$	Number of PAA segments representing $C$
$\bar{C} = \bar{c}_1, \dots, \bar{c}_w$	PAA of the time-series $C$
$\alpha$	Alphabet cardinality (size)
$\hat{C} = \hat{c}_1, \dots, \hat{c}_w$	Symbolic representation of $C$

**Table 1.** Meaning of the main notation used in this section.

### 2.1.1 PAA Dimensionality Reduction

The concept behind the PAA is that a time series  $C$  of length  $n$  can be represented in a  $w$ -dimensional space by a vector  $\bar{C} = \bar{c}_1, \dots, \bar{c}_w$ .

In order to do this, the original data (time-series and subsequences) is first *normalized* to have zero mean and unit standard deviation, then is divided into  $w$  equal by sized *frames* and the mean value of the data  $\bar{c}_i$  falling within the  $i$ -th frame is calculated. A visual example of PAA signal's approximation is illustrated in Figure 1.



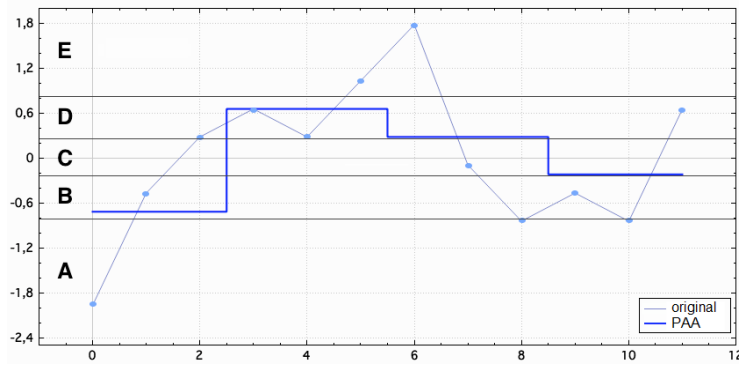
**Figure 1.** An example of PAA approximation of a signal.  $C$  is the original signal while  $\bar{C}$  is its PAA approximation. In this example  $w = 10$  and  $f_s = 100$  Hz.

### 2.1.2 Discretization

To obtain a discrete representation, a further transformation must be applied to the PAA signal. It is desirable to have a discretization technique that will produce symbols with equiprobability. This is easily achieved if we supposed that the time series in exam have a Gaussian distribution; in (Lin et al., 2003) it is demonstrated that the aforementioned assumption is reasonable. Hence, given an alphabet cardinality  $\alpha$ , the *breakpoints* for the discretization can be simply determined by finding the  $\alpha + 1$  points that will produce  $\alpha$  equal-sized areas under the Gaussian curve.

Formally, the breakpoints are a sorted list of numbers  $B = \beta_0, \dots, \beta_\alpha$  such that the area under the Gaussian curve from  $\beta_i$  to  $\beta_{i+1}$  is equal to  $1/\alpha$ . Obviously,  $\beta_0$  and  $\beta_\alpha$  are defined as  $-\infty$  and  $+\infty$  respectively.

In summary, a time-series can be discretized in the following manner. First, time-series is normalized, then a PAA is obtained from the original time series and then all PAA coefficients that are below the smallest breakpoint are mapped to the symbol  $a$ , all coefficients greater than or equal to the smallest breakpoint and less than the second smallest breakpoint are mapped to the symbol  $b$  etc. The concatenation of these subsequent symbols that represent the signal is called *word*. In Figure 2 an explanatory example of the discretization step is provided.



**Figure 2.** In the example, with  $n = 12$ ,  $w = 3$  and  $\alpha = 5$ , the time series is discretized by obtaining a PAA approximation and then using predetermined breakpoints  $\beta$  is mapped map into SAX symbols; the time series is mapped into the word BDDC.

### 2.1.3 Distance Measure

After introducing the *SAX representation*, a new distance measure can be defined on it. The most common distance measure for time series is the *Euclidean distance*, but is not the only one. For the subsequent considerations, let  $Q = q_1, \dots, q_n$  and  $C$  be two time-series of the same length  $n$  and let  $\hat{C}$  and  $\hat{Q}$  be their SAX symbolic representations. The SAX distance is defined as:

- **Euclidean Distance:**

$$D(Q, C) = \sqrt{\sum_{i=1}^n (q_i - c_i)^2} \#(1)$$

- **PAA Distance:**

$$D_{PAA}(\bar{Q}, \bar{C}) = \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w (\bar{q}_i - \bar{c}_i)^2} \#(2)$$

- **SAX Distance:**

$$D_{SAX}(\hat{Q}, \hat{C}) = \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w \bar{d}(\hat{q}_i, \hat{c}_i)^2} \#(3)$$

Equation (3) represents a proved lower bounding approximation of the Euclidean distance between the original subsequences  $Q$  and  $C$ .

In practice, the  $\bar{d}(\cdot, \cdot)$  function can be implemented by using a *lookup table* and can be defined by the following expression:

$$f(\hat{q}_i, \hat{c}_i) = \begin{cases} 0, & |\hat{q}_i - \hat{c}_i| \leq 1 \\ \beta_{MAX(\hat{q}_i, \hat{c}_i)-1} - \beta_{MIN(\hat{q}_i, \hat{c}_i)} & otherwise \end{cases} \quad \#(4)$$

where  $\beta_x$  are the breakpoints of the symbol  $x$ .

The distance between two SAX representations of a time series requires looking up the distances between each pair of symbols, squaring and summing them, taking the square root and finally multiplying by the square root of the compression rate.

In conclusion, it can be noticed that there is a clear tradeoff between the parameter  $w$  controlling the number of approximating elements, and the value  $\alpha$  controlling the granularity of each approximating element. The SAX technique is highly data dependent thus it's difficult to determine a tradeoff analytically, but it must be found empirically.

## 2.2 Gesture Classification

The classification phase is an on-line procedure aiming to assign one of the  $K$  possible classes (gestures/activity types available for the problem at hand) to the observation in exam. One of the most employed approaches to deal with classification is k-Nearest Neighbours (k-NN) (Friedman et al., 2009): with k-NN a new observation is classified as the class most represented within the group of  $k$  'closest' tagged observations available in an historical dataset. It is clear how this approach strongly relies on the availability of a metric distance, which is of no trivial to be defined when dealing with time series data of, possibly, different lengths.

In this work, the distance defined in (3) is employed by a 1-NN classifier, being 1 a typical choice for the size of the neighborhood  $k$  when dealing with complex, time series-related input data (Pękalaska, Duin & Paclik, 2006). The k-NN classifiers are lazy learning approaches (Friedman et al., 2009), where all the computational cost is done on-line with the search for the  $k$  closest tagged observations in the historical dataset: in order to make the algorithm affordable in the wearable framework, templates are defined. *Templates* are strings that act as unique representative of the problem classes  $\tau = \{\tau_i\}_{i=1}^K$ : the 1-NN classifier will make comparison only with the templates by computing at most  $K$  distances. In our approaches  $\tau_i$  is chosen as the most

represented word within the  $i$ -th class but other solutions could also be employed. In the following, we discuss the choice of parameters  $w$  and  $\alpha$ .

### 2.2.1 Setting the word length

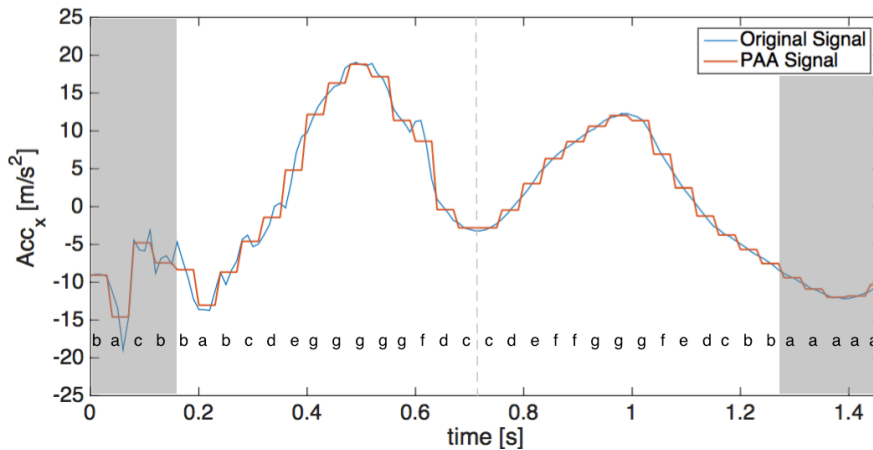
The word length  $w$  is closely connected to the signal PAA. In fact, the choice of  $w$  depends on the dynamics we want to capture: if a signal presents high dynamics and gestures in different classes have similar shapes,  $w$  should have a small value. On the other hand, in order to maximize the compression rate  $w$  should be as high as possible. Thus, we select  $w$  in such a way to trade-off between the two exposed objectives.

When working on gestures, it is necessary to consider the period information. For this reason, even if the word length is fixed, the number of samples considered in each PAA frame can vary with the signal period. This is not a problem in the proposed framework because we are only interested on *shape information*. So, the length of each PAA frame is proportional to the gesture period, but usually differs from one gesture to another. For this reason, we calculate the number of samples in each PAA frame as  $n/w$  rounded down to the nearest integer. In most cases  $n$  is not divisible by  $w$  and then the rest of this fraction indicate the number of samples not considered in the PAA approximation. To avoid information loss, we use the period information as follow:

1. We calculate the PAA frame length  $l$  as  $\frac{n}{w}$  rounded down to the nearest integer.
2. We calculate the rest of  $\frac{n}{w}$ ,  $r$ , that is an integer between 0 and  $w - 1$ .
3. We calculate how many other PAA frames could be created with the  $r$  excess samples as  $\frac{r}{l}$  rounded down to the nearest integer.
4. We calculate the rest of  $\frac{r}{l}$ , that is an integer between 0 and  $l - 1$ . The last rest represents the real number of samples neglected that, at worst, is still significantly less than  $w - 1$ . In fact, with this method it's obvious that by increasing  $w$  then the final number of excess samples tends to 0. In this way, all gesture signals could be approximated with a PAA signal formed by  $w + \left\lfloor \frac{r}{l} \right\rfloor$  frames of equal size.

In this manner, however, the number of frames varies from a gesture to another and depends on the period, leading some gestures to have word lengths longer than  $w$ . We overcome this problem by selecting the *central substring* of fixed length  $w_1$ : even if for some gesture of length  $n$  is non-divisible by  $w$ , and so the word is longer than  $w$ , we

can simply extract the  $w_1$  central characters. This procedure has the by-product of removing noise which is present at the borders of a gesture. In Figure 3. In the example  $w = 30$ ,  $w_1 = 28$ ,  $\alpha = 7$ . A DP (see next section) gesture with  $n = 148$  that is approximated with 37 PAA frames. The word "bacbbabcdegggggfdccdeffggfedcbbaaaa" is the SAX string that represents the signal in Figure; from this the  $w_1$  central characters are extracted. These are the characters that correspond to the lighten portion of the signal. an example of the selection of the central substring is given.



**Figure 3.** In the example  $w = 30$ ,  $w_1 = 28$ ,  $\alpha = 7$ . A DP (see next section) gesture with  $n = 148$  that is approximated with 37 PAA frames. The word "bacbbabcdegggggfdccdeffggfedcbbaaaa" is the SAX string that represents the signal in Figure; from this the  $w_1$  central characters are extracted. These are the characters that correspond to the lighten portion of the signal.

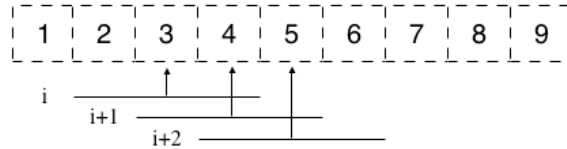
### 2.2.1 Setting the Alphabet Size

The alphabet size  $\alpha$  entails the discretization on the amplitude domain. If a high resolution is required, alphabet size  $\alpha$  should be chosen high enough. In literature, it has been heuristically proven that the most effective values are 3-4 (Keogh, Lin & Fu, 2005).

### 2.3 Activity Classification

Activities can be seen as composition of gestures. Thus, in order to consider a group - of subsequent gestures, we use a sliding window approach that allows to evaluate the sequence of the gestures related to the training session simulation and already classified. To do so, we consider a sliding window of fixed size  $l_{sw}$  that includes  $l_{sw}$  subsequent gestures at a time. Calculating the mode between the classes of these  $l_{sw}$

gestures, we find the activity that represents the window in exam; we memorize that the subject has performed this activity for the period of the central gesture of the window. After that, we shift the sliding window of one position, considering other  $l_{sw}$  gestures and so on. In Figure 4 we illustrate an example of three sliding window shift, considering  $l_{sw} = 3$ .



**Figure 4.** Examples of sliding window in three subsequent shift on the gestures sequence;  $l_{sw} = 3$ . Notice that the arrowed lines point to the central gesture of each window.

### 3. Experimental Results

The presented work has been tested on two datasets:

- *HAR dataset* - a reduced version of the UCI Human Activity Recognition (HAR) Using smart-phones Dataset (Anguita, Ghio, Oneto, Perez & Ortiz, 2013; Reyes-Ortiz, Oneto, Samà, Parra & Anguita, 2016) where three continuous-repetitive normal day activities have been examined. The dataset represents a three classes AR problem, where the activities are walking (WLK), walking upstairs (WUS) and walking downstairs (WDS); the dataset includes experiments that were carried out by 30 people where all the participants were wearing a smartphone, Samsung Galaxy S II, on the waist during the experiment execution. 3-axial linear acceleration and 3-axial angular velocity have been captured at a constant rate of 50Hz using the embedded accelerometer and gyroscope of the device.
- *Cross-Country Skiing dataset* - a dataset of Cross-Country Skiing where 3 different styles were performed by 8 skiers; the three styles are
  1. double poling (DP), where both poles are used in parallel by the skier;
  2. diagonal stride (DS), where the poles are used in succession;
  3. kick-double-pole (KDP), a variant of DP, where an asymmetrical kick is performed by the skier.

Analyzed athletes had different *skill levels*, from recreational (R) to achiever (A). Athletes were wearing a smart-watch placed on the wrist; 3-axial linear

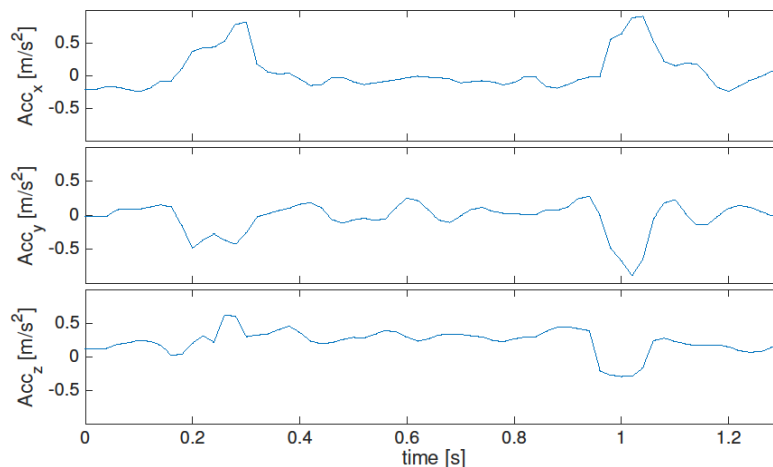
acceleration, 3-axis angular velocity, 3-axis magnetic field, have been captured at a constant rate of 100Hz using the embedded accelerometer, gyroscope and magnetometer of the device.

In order to use the aforementioned described approach only one IMU-generated data stream has to be considered, which represents an advantage in terms of time complexity and power consumption (fundamental in resource-constrained devices); hence, we focus on *acceleration* signals, which best capture signal variability.

In the following, we discuss the results obtained with the HAR dataset and the Cross-Country Skiing dataset, respectively.

### 3.1. Experimental Results: HAR dataset

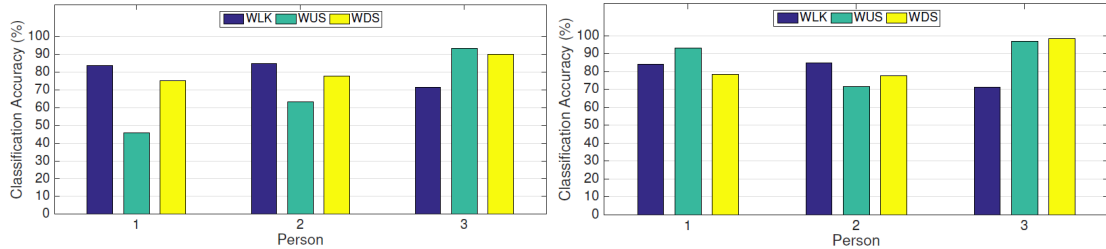
The time series in Figure 5 is an example of a WDS gesture; after testing the classification performances on a training dataset, the *z-axis* has been chosen. The sampling rate is 50 Hz and the minimum period for a gesture is almost 1 s, therefore, it doesn't seem reasonable to choose a value of  $w > 20$ . Furthermore, the signal doesn't appear highly affected by noise and so we can set  $w_1 = 18$ . The results showed in the following are obtained with  $\alpha = 7$ .



**Figure 5.** HAR Dataset - Example of a typical WDS Gesture. The informative content of the Gesture seems to be subdivided between the three axes.

The classification accuracy on the GR problem is reported on the left panel of Figure 6, where the classification accuracy is reported as a mean value in a 3-fold cross-validation procedure (Friedman et al., 2009) separately for each person. It has emerged how WUS and WDS classes are quite similar when monitoring the

acceleration on the z-axis; for this reason, a second experiment has been performed where WUS and WDS are considered as a single class: the classification accuracy is reported in the right panel of Figure 6 and shows the benefit introduced by this solution.



**Figure 6.** Classification accuracy on a test dataset with  $w = 20$ ,  $\alpha = 7$ : on the left panel WUS and WDS are considered as separated classes (**Case 1**), while on the right one they are considered as a single class (**Case 2**).

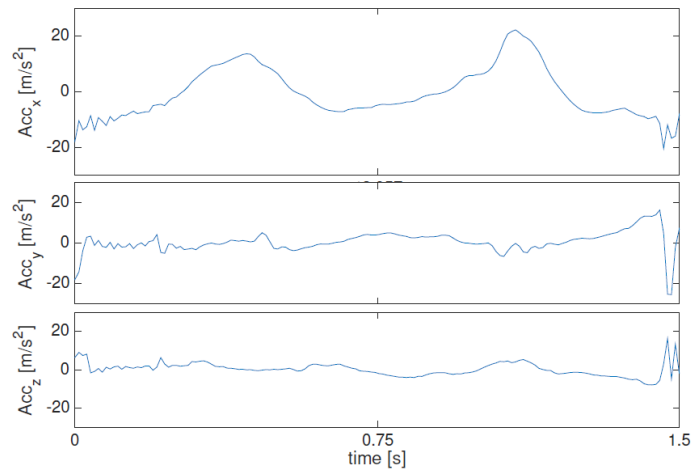
Finally, in Table 2, results for 3 persons in terms of the AR problem are reported in terms of activity time (AT): namely, how much time of the total AT has been correctly classified; also in this case, results are reported for both experiments (the one where WUS and WDS are considered as separated classes and the one where they are as a single class). Results are reported for different values of the window length  $l_{sw}$ , proving how longer windows lead to a better algorithm performance.

Person	$l_{sw}$	Case 1	Case 2
		Correctly Recognized AT (%)	Correctly Recognized AT (%)
1	3	3' 30" / 4' 48" (72%)	4' 07" / 4' 48" (86%)
	5	3' 42" / 4' 48" (77%)	4' 15" / 4' 48" (89%)
	7	3' 49" / 4' 48" (80%)	4' 26" / 4' 48" (92%)
2	3	3' 05" / 3' 42" (81%)	3' 14" / 3' 42" (87%)
	5	3' 06" / 3' 42" (84%)	3' 14" / 3' 42" (87%)
	7	3' 09" / 3' 42" (85%)	3' 15" / 3' 42" (88%)
3	3	3' 18" / 3' 40" (90%)	3' 22" / 3' 40" (92%)
	5	3' 28" / 3' 40" (95%)	3' 29" / 3' 40" (95%)
	7	3' 29" / 3' 40" (95%)	3' 29" / 3' 40" (95%)

**Table 2.** AR performances for HAR dataset reported in terms of AT for different values of  $l_{sw}$ .

### 3.2. Experimental Results: Cross-Country Skiing

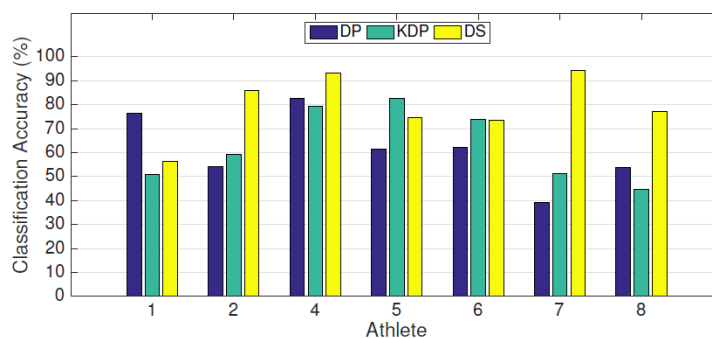
In this dataset, we consider only the x-axis acceleration data because it is the most informative signal in the dataset in exam. In fact, axes y and z data streams represent a signal with a high noise level. For this reason, we chose not to considerate acceleration along y and z. In Figure 7 an example of DP gesture is shown where it is apparent the higher informative content in the x-axis w.r.t. other acceleration signals.



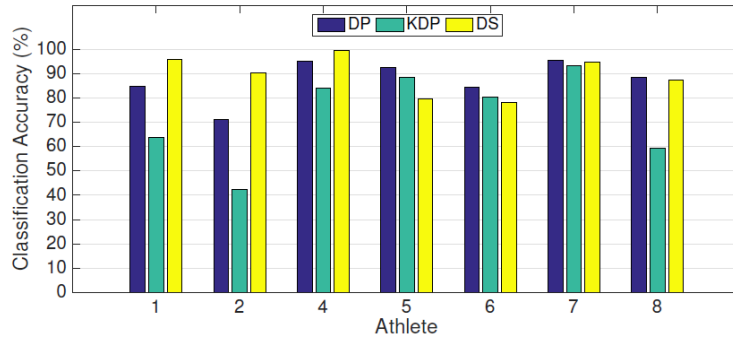
**Figure 7.** Cross-Country Dataset - Example of typical DP gesture. The most informative content of the Gesture seems to be contained in the x-axis.

In this application, we consider as fundamental that the PAA signal evolution follows at the best the original signal evolution, therefore we chose a length word  $w = 30$  and  $\alpha = 7$ : the chosen value of  $w$  is high enough to guarantee good adherence with the original signal, but with a reasonable dimensionality reduction effect.

As explained, the styles DP and KDP differs for the presence of a kick performed by the skier, a movement that is supposedly difficult to be observed from the device position (wrist); this hypothesis has been verified experimentally by considering two classification configurations: DP and KDP are considered, respectively, as different gestures (**Case 1**) and as the same gestures (**Case 2**). Results, in term of accuracy, of these gesture classification experiments are reported in Figure 8 and Figure 9.



**Figure 8.** Cross-Country Dataset – (**Case 1**). In this experiment:  $w = 30$ ,  $w_1 = 25$  and  $\alpha = 7$ . Percentage of Classification accuracy for the seven Athletes.



**Figure 9.** Cross-Country Dataset – (**Case 2**). In this experiment:  $w = 30$ ,  $w_1 = 25$  and  $\alpha = 7$ . Percentage of Classification accuracy for the seven Athletes are here reported. In the **Case 2** formulation of the AR problem, the cardinality of the classification problem is 2; however, the classification accuracy reported here for class (DP/KDP) is divided into the 2 original styles to provide more insights on the classification capabilities.

After gesture classification, we test performance for the activity classification problem in terms of AT.

We tested this approach with different sliding window size:  $l_{sw} = 3$ ,  $l_{sw} = 7$  and  $l_{sw} = 11$  on two athletes (4 and 7). Values of  $l_{sw}$  greater than 11 seem too elevated because we consider reasonable that after 11 gestures of one technique, an athlete change activity. In Table 3, for each Athlete, we summarized three experiments executed considering the gesture classification with the two configurations: three classes DP, KDP and DS (**Case 1**) and two classes DP/KDP (we consider DP and KDP as the same gesture) and DS (**Case 2**). It can be noticed that the Activity recognition works better increasing  $l_{sw}$ . In case 1, for Athlete 4 the activity recognition works better than Athlete 7; it means that in order to obtain great results of activity recognition, we need a classification accuracy  $> 50\%$  in each class. For case 2, it can be noticed that the total activity time is recognized for  $l_{sw} > 7$  for both the Athlete; in this case in fact, the classification accuracy is  $> 50\%$  in each class.

Athlete	$l_{sw}$	Case 1	Case 2
		Correctly Recognized AT (%)	Correctly Recognized AT (%)
4	3	10' 41" / 11' 41" (91.4)	11' 33" / 11' 41" (98.8)
	7	11' 15" / 11' 41" (96.2)	11' 41" / 11' 41" (100)
	11	11' 19" / 11' 41" (97.5)	11' 41" / 11' 41" (100)
7	3	6' 42" / 11' 36" (57.7)	11' 29" / 11' 36" (99)
	7	6' 59" / 11' 36" (60.2)	11' 36" / 11' 36" (100)
	11	7' 06" / 11' 36" (61.1)	11' 36" / 11' 36" (100)

**Table 3.** AR performances for Skiing dataset reported in terms of AT for different values of  $l_{sw}$ .

#### 4. Conclusions and Future Works

In this work a symbolic-based solution for AR problem applied to resourced-constrained devices is proposed. The approach is based on SAX, a symbolic representation used in several fields of application that allows dimensionality reduction. The work has been tested on datasets related to continuous-repetitive activities; on the tested datasets, the proposed approach achieved good classification accuracy, however it has been shown how similar movements (the pair WUS-WDS and the pair DP-KDP) are difficult to be recognized. In order to improve the algorithm performance a follow-up of this work could be the creation of a metrics that allows to measure the SAX distance considering more than one IMU-generated time series or a classification approach that takes into account multiple SAX distances computed on different IMU-generated time series.

#### 5. References

- Androulakis, I. P. (2005). New approaches for representing, analyzing and visualizing complex kinetic mechanisms. *Computer Aided Chemical Engineering* (20), 235-240.
- Anguita, D., Ghio, A., Oneto, L., Parra Perez, X., & Reyes Ortiz, J. L. (2013, April). A public domain dataset for human activity recognition using smartphones. Presented at the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium *Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 437-442).
- Belgioioso, G., Cenedese, A., Cirillo, G. I., Fraccaroli, F., & Susto, G. A. (2014, December). A machine learning based approach for gesture recognition from inertial measurements. *Presented at the 53rd IEEE Conference on Decision and Control*, Los Angeles, California.
- Carbone, F. (2014). Ricerca di anomalie su serie temporali di temperature basata su rappresentazione simbolica [Temperature Time-Series Anomaly Detection based on Symbolic Representation]. M.Sc. Thesis, University of Padova (In Italian).
- Cenedese, A., Susto, G.A., Belgioioso, G., Cirillo, G. I. & Fraccaroli, F. (2015) Home Automation Oriented Gesture Classification from Inertial Measurements. *IEEE Transactions on Automation Science and Engineering*, 12(4), 1200-1210.

- Cenedese, A., Susto, G.A., & Terzi, M. (2016, June). A parsimonious approach for activity recognition with wearable devices: an application to cross-country skiing. Presented at *the 15<sup>th</sup> European Control Conference*, Aalborg, Denmark.
- Clifton, L., Clifton, D. A., Pimentel, M. A., Watkinson, P. J., & Tarassenko, L. (2013). Gaussian processes for personalized e-health monitoring with wearable sensors. *IEEE Transactions on Biomedical Engineering*, 60(1), 193-197.
- Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The elements of statistical learning* (Ed. 2). Berlin: Springer series in statistics.
- Gowing, M., Ahmadi, A., Destelle, F., Monaghan, D. S., O'Connor, N. E., & Moran, K. (2014). Kinect vs. low-cost inertial sensing for gesture recognition. Presented at the 20th *International Conference on Multimedia Modeling*, Dublin, Ireland.
- Harris, B. W. (2013). Anomaly detection in rolling element bearings via two-dimensional Symbolic Aggregate Approximation (Doctoral dissertation, Virginia Tech).
- Kasetty, S., Stafford, C., Walker, G. P., Wang, X., & Keogh, E. (2008). Real-time classification of streaming sensor data. *Presented at the 20th IEEE International Conference on Tools with Artificial Intelligence*, Dayton, Ohio.
- Keogh, E., Lin, J., & Fu, A. (2005, November). Hot sax: Efficiently finding the most unusual time series subsequence. *Presented at the 5th IEEE International Conference on Data mining, Houston, Texas*.
- Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. In Mohammed J. Zaki, Charu C. Aggarwal (Eds) *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery* (p. 2-11). New York: ACM Press.
- Morris, D., Saponas, T. S., Guillory, A., & Kelner, I. (2014, April). RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. *Presented at the 32<sup>nd</sup> annual ACM conference on Human factors in computing systems*, Toronto, Canada.
- Pękalska, E., Duin, R. P., & Paclík, P. (2006). Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, 39(2), 189-208.
- Ravi, N., Dandekar, N., Mysore, P., & Littman, M. L. (2005, July). Activity recognition from accelerometer data. *Presented at the 20th National Conference on Artificial Intelligence*, Pittsburgh, Pennsylvania.

- Reyes-Ortiz, J. L., Oneto, L., Samà, A., Parra, X., & Anguita, D. (2016). Transition-aware human activity recognition using smartphones. *Neurocomputing*, 171, 754-767.
- Susto, G. A., Schirru, A., Pampuri, S., & McLoone, S. (2016). Supervised aggregative feature extraction for big data time series regression. *IEEE Transactions on Industrial Informatics*, 12(3), 1243-1252.
- Tran, C., & Trivedi, M. M. (2012). 3-D posture and gesture recognition for interactivity in smart spaces. *IEEE Transactions on Industrial Informatics*, 8(1), 178-187.
- Yi, B. K., & Faloutsos, C. (2000, September). Fast Time Sequence Indexing for Arbitrary Lp Norms. *Presented at the 26th International Conference on Very Large Data Bases*, Cairo, Egypt.

