A Methodology for Pattern Detection in Gait Behaviour using the i-Walker

Jonatan MORENO [a], Atia CORTÉS [b], Cristian BARRUÉ, [a], Ulises CORTÉS [a]

[a] *CS Department. Universitat Politècnica de Catalunya*
[b] *ESAII Department. Universitat Politècnica de Catalunya*

1. Introduction

The elderly population (aged 65+) is growing fast nowadays and is expected to be the social group with major increase in the near future. People over 70 years tend to suffer different diseases related to the aging process. For example, in the US 80% of elderly population suffers one chronic disease (*e.g.*, diabetes, Parkinson or Alzheimer among others), and 50% has at least two [4]. In most cases this leads to a loss of autonomy, community activity reduction and isolation. In addition, the frequency of falls increases with age and frailty level and are a main cause of unintentional injury [5]. A combination of biological factors and disease-related conditions are the main cause of most of the falls.

This will represent a challenge for the public health systems that will have to face a huge economic impact to deal with this demographic situation. This is already not sustainable in certain countries and will be a world-wide issue in a near future. One of the main objectives for the H2020 EU research program is to focus on the analysis of the causes and consequences of pathologies in order to find patterns that will support early detection of a disease or associated risks. Consequently the care community could take decisions on intervention and educational information to delay the physical or cognitive decline of the elderly and try to keep them independent as long as possible living in the community. The evolution of ICT tools in collaboration with medical knowledge has empowered the use of sensorized assistive devices that can provide ubiquitous, real-time information about a person's evolution in terms of physical and cognitive conditions.

In this work we aim to use an assistive device, the i-Walker [1], as a sensory platform that assists in elderly mobility that also is able to extract relevant information from the interaction that could be used as an early detection of symptoms related to aging or to some kind of pathology. We present a methodology to analyze the data measurements taken by the i-Walker and find patterns in *gait behaviour* with two different groups of people: healthy elderly people (not suffering from any disease or fall injury) and non-healthy elderly people. The second group were part of the pilot of the EU funded project I-DONT-FALL (IDF) [2].

Data has been collected, early 2015, from real in-patients at the *Fondazione Santa Lucia* (FSL), in Rome, and at the residence care center *Los Nogales* (LN) in Madrid following a strict inclusion protocol.

2. Methodology

For this work the data recorded as a result of the interaction of real users with the *i*-Walker while performing specific tasks is named *exercise*. For each exercise, the *i*-Walker produces a dataset which is composed by several Time Series (TS) representing each one specific sensor data (left normal force, right speed, *etc.*). A TS is a sequence of real values, in our case representing a sensor reading during the exercise. It could be defined as $T = (t_1, t_2 \cdots t_n), n \epsilon \mathbb{R}$

Therefore, the TS that represents the sensor data recorded in several exercise sessions are analyzed by finding relevant patterns that could be determinant to differentiate between various target populations by means of the gait behaviour that they present.

The methodology presented here targets to cluster the exercises into different groups regarding the patterns that they present. For each TS belonging to an exercise, patterns are extracted and a vectorized format that preserves the relevance of patterns is reached. By comparing all the resulting vectors a cluster relying on the relevant patterns is obtained.

2.1. Pattern Extraction

In order to extract the patterns given a TS we will apply *Symbolic Aggregate Approximation* (SAX) [3], a technique that has been widely used in the TS analysis domain. It allows dimensionality reduction and indexing with a lower-bounding distance measure. By applying SAX we are able to transform the given TS into a list of patterns.

First we define a window of length w that will be slided each iteration ss positions over the input TS. The values of the TS that falls inside the window will be normalized and we will run *Piece-wise Aggregate Approximation*(PAA). By specifying the spw parameter the PAA will reduce the values inside the window into spw values. Afterwards, an alphabet of cardinality a is defined and each of the resulting values of the PAA will be replaced by the corresponent alphabet symbol obtaining one pattern as a result. This procedure is repeated until the window reach the end of the TS obtaining a pattern at each iteration.

2.2. Infering Relevance

The relevance of the patterns is computed after they are extracted. Two factors have influence on the relevance: number of ocurrences and resolution. The patterns obtained as a result of the SAX methodology will all have the same resolution (specified by the cardinality of the alphabet). However, binary notation is used as an alphabet which allows us to generalize patterns. Notice that the pattern $0b00, 0b01$ can be generalized by taking the first bit of each symbol, that is $0b0, 0b0$. That is equivalent to execute SAX with $a = 2$ instead of $a = 4$.

In order to compute the relevance a tree is built from top to bottom going through the list of patterns. Starting with an empty tree (just the root node) each pattern is inserted iteratively. When a pattern has to be inserted the maximum generalization of it is taken and inserted into the tree. Therefore, intially the tree will have just one depth level. When a node at this level accumulates a certain number of patterns (specified by a threshold parameter) it has to be splitted into a deeper level of resolution. That is because the difference between the patterns that share the same generalization at resolution 1 but are different at resolution 2 could be significant. Once all the symbols are introduced a dynamic tree is obtained as a result that preserves the two factors that have influence into the relevance. Thus, the more occurrences a symbol has and the more deeply in the tree it falls (*e.g.*, more resolution is required) more relevant the pattern is.

Finally, a vector with a number of cells equal to the number of possible patterns that could be produced with the configuration used is built, *i.e.*, each cell represents each possible pattern. Thus, each position is filled with the computed relevance. That is, a cell representing a pattern that has reached the maximum possible resolution contains the number of occurrences of that pattern. The cells that represent patterns that have been not produced are filled with a 0. For the nodes that are a generalization of some patterns and have not been splitted (*i.e.*, the difference among that patterns is not relevant), the number of occurrences is spread out among the cells of the patterns that such node is representing.

2.3. Clustering

In order to perform the clustering a condensed distance matrix is computed. To fill the distance matrix the distance among the vectors that we want to analyze is computed. Afterwards, an Agglomerative Hierarchical Clustering (AHC) is executed to obtain the different clusters.

3. Results

This section outlines a first use case where the methodology has been applied. First, the exercises that are going to be clustered need to be picked. Inhere the dataset is composed by a subset of exercises produced at IDF project (those from *FSL* pilot) which belongs to users that present one or more pathologies. In addition, exercises produced by healthy users[1] from LN are used as well (see §1).

By applying the presented methodology it is expected that the exercises from healthy/non-healthy users falls into separate clusters. After executing the first phase of the methodology on the dataset for each variable belonging to an exercises the numerical vector is generated. The selection of the variables to be analyzed is required since in this first approach, the clustering works with a single variable. We selected *Right Normal Force* variable for the clustering, as it might be representative (*i.e.* the amount of leaning a user performs over the rollator).

---

[1]A user is considered healthy when it overcomes some medical tests and scales(*e.g.*, Barthel, Tinetti, *etc.*)

	True	False
True	3250	51
False	4	939

Table 1. Truth Table

		Precision	Recall
rnf	Healthy	0.998	0.984
	Non-Healthy	0.948	0.996

Table 2. Precision and Recall

Thus, all vectors that represent the relevance of patterns found in the *Right Normal Force* variable of each exercise are used to make the clustering.

Table 1 outlines the truth table of the results. The clustering process resulted into two clusters: non-healthy cluster (NHC) and healthy cluster (HC). Thus, 3250 from 3254 non-healthy users have fallen into NHC while 939 from 990 healthy users have fallen into HC. Table 2 outlines the corresponding precision and recall.

4. Conclusions

This work aimed to reveal whether the data obtained from *i*-Walker sensors contains relevant patterns that could be used to differentiate between some exercises performed by a target population. As shown in §3 the results are *promising*. The methodology has been able to differentiate the data used in this study. Since we already realized that certainly there are some patterns that could be linked to groups of exercises, we conclude that some data mining tasks as could be a classifier could be implemented on top of the presented methodology.

However, this is a first attempt to study the data obtainied from *i*-Walker sensors and therefore some aspects ought to be improved and there is some furhter analysis to be done (*e.g.*, the configuration of the different parameters, *etc.*)

References

[1] U. Cortés, A. Martínez-Velasco, C. Barrué, and R. Annicchiarico. AI based fall management services - the role of the i-Walker in I-DONTFALL. In *Advances in Artificial Intelligence - 11th Mexican International Conference on Artificial Intelligence, MICAI 2012, San Luis Potosí, Mexico, October 27 - November 4, 2012. Revised Selected Papers, Part I*, pages 395–406, 2012.

[2] Integrated prevention and Detection sOlutioNs Tailored to the population and Risk Factors associated with FALLs. http://www.idontfall.eu/, 2010.

[3] J. Lin, E. Keogh, L. Wei, and S. Lonardi. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.

[4] WHO. A strategy to prevent chronic disease in Europe. a focus on public health action. the CINDI vision. http://www.who.int, 2002.

[5] WHO. Global report on falls prevention in older age. http://www.who.int, 2007.