Autonomous coach for physical exercise games - Robot agent

Maria Carolina Varandas Roque
carolina.roque@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

January 2021

Abstract

Nowadays, the world's population is ageing. The downfall of physical and cognitive capabilities is common due to advanced age or specific diseases. Physical activity emerges as an important tool in the maintenance of the elderly's quality of life, since their independence relies on the capacity to maintain themselves autonomous in their daily routines. However, the senior population are not very keen on physical exercise, for a variety of reasons, such as, fear of injuries, lack of motivation and company, or the shortage of activities for their age group. Over time, some strategies have been implemented to minimize the lack of commitment demonstrated by the eldery. Exergames, virtual coaches and robot-assisted training are the ones that stand out the most. In these approaches, the physical exercises are projected to the senior population, taking into consideration their limitations and restraints. One of the most important factors is the feedback provided during the activities, which stimulates and motivates their participation. The aim of this thesis is to add motivational feedback in a set of specific exergames. The data was obtained from several sessions with the elderly playing the exergames, supported by physiotherapists. Through the Learning from Demonstration approach and the kNN, CRF and HCRF algorithms, it is intended to generate a classifier that predicts which feedbacks should be given in certain game situations. Even though the results were not optimal, the accuracy of some feedbacks was positive. Finally, suggestions are delineated for future work.

Keywords: elderly, motivation, exergames, physical exercise, feedback.

1. Introduction

The world is facing the generalized ageing of the population [1]. For the first time in human history, in 2018, the number of people over 65 years old surpassed the number of children under five years of age [2]. Beyond that, in 1990, only 54 million people in the world were aged 80 or more, while in 2019, this number reached 143 million people. It is projected that in 2050 this number will increase to 426 million people [2]. Thus, the need to help and support this age group has become a focus of concern and interest for governments worldwide [1].

Due to the advanced age, the human body loses mental and motor capabilities, which leads to mobility limitations, added fragilities, and other health problems. In this regard, it is essential to adopt a healthy lifestyle, to which physical exercise is intrinsic. According to Larsen et al. [3], the sport activity can effectively reduce the weakness and the loss of physical condition of the elderly.

However, there are a great number of barriers that keep the elderly away from sports practice. Aside from the motor difficulty in performing the exercises, there is also the lack of interest and company, health debilitations, fear of injuries and the shortage of activities directed to their age group [3].

In order to motivate the population to exercise, several approaches have emerged over time, among those, the exergames. These games are played by executing actual physical movements. Larsen et al. [3] reveal a positive impact of these games on the physical condition of the elderly. Furthermore, and according to the same source, the feedback given during the exergames is one of the factors that make them more motivating. As stated by Burgers et al. [4], the feedback can have a substantial influence in the motivation of the players, depending on how it is delivered, the user's interpretation and how much it is related to their behaviour. The increase of the players' motivation can enhance their adherence to the physical exercises [5].

Therefore, this work aims to add a verbal feedback feature in a set of exergames, specifically designed for the elderly, with the goal of making them physically active. To achieve this, a variety of classification models were studied, to predict the motivational feedback to be transmited to the elderly, at certain moments of the game.

2. Background

Over the past few years, a variety of approaches have been developed to motivate people, particularly the elderly, to engage in physical exercise. These include serious games, robot-assisted training and virtual coaches.

2.1. Serious Games

A serious game is described as a game whose main objective is not leisure. However, it can also be fun at the same time [6]. Serious games have been developed in various fields, including education, training, well-being, health care, among others.

Exergames can be considered serious games since they have a greater purpose than just entertainment. The main characteristic of these games is that they require several physical movements to be played. In that manner, they are an alternative to traditional physical exercise.

As the practice of sport is so important in society, exergames can be a reliable vehicle of physical exercise, contributing to improving the quality of life and satisfaction of the population [7]. Besides, they can also help counteract the low participation of older people in physical activity, due to their motivating and stimulating characteristics.

2.2. Robot-Assisted Training

Advances in robotics have made it possible to apply robots as a tool of assistance to man and can also be specific to a particular target audience, from children, elderly, students, to patients who have suffered a stroke. Robot-assisted training is a growing area of research in the field of Human-Robot Interaction. This area studies how robots can interact with humans, helping them to perform determined tasks of physical and cognitive training [8]. These systems can be applied in the most diverse areas, from post-stroke rehabilitation to therapy with children with Autism Spectrum Disorders, to cognitive training for patients with dementia or Alzheimer's disease, among other applications.

2.3. Virtual Coaches

Virtual coaches have also emerged with the evolution of technology. Their goal is to continuously monitor the activities of their users and their environment, in order to detect situations where intervention is needed, providing immediate assistance [9]. One of the applications of these devices is in helping patients who leave the hospital and do not have the necessary training to continue the health care they have received until then. For example, in the handling of newly prescribed medical devices or following more complex medical regimes. Virtual coaches can act in this regard, ensuring compliance with rehabilitation instructions and the following of new medical regimes.

2.4. Example - Interative Exercise Coaching System

Although several fitness applications have already been launched over time, many of these have not been designed for the elderly, but rather for the younger population. For this reason, they are not suited to the limitations and difficulties characteristic of older people.

To this end, Ofli et al. [10] developed a training system that leads their users through a set of video exercises, tracking and measuring their movements, while providing real-time feedback on user performance, and recording it over time. This system aims to improve the flexibility, balance, strength and endurance of the elderly, in order to reduce the risk of falls and promote their independence in daily activities. In this system, a Kinect camera is used, which allows the skeletal reconstruction of the user, from 20 joints. Thus, it is possible to have an estimate of the player's body posture in real time, used to evaluate his performance, to count successfully performed repetitions and to send feedback alerts at the necessary moments.

Each exercise begins with a video of a trainer indicating its health benefits and making a demonstration of the movements required for the task. After this initial step, the players start the exercise, following the feedback that is given throughout the game. In the end, the user's performance is recorded and summarised. There is also a history which allows the visualization of a monthly summary or the player's last 10 sessions.

Ofli et al. have studied various exercises, one of which is the Buddha's Prayer exercise that consists of lifting and lowering the arms, always with the palms of the hands together [10]. This exercise is to be done while seated and its objective is to raise the hands as much as possible without separating the palms. The relevant joints in this exercise are marked in orange in Figure 1.

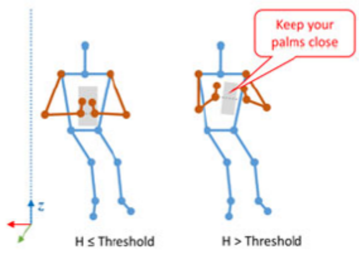

Figure 1: Verification of Buddha's Prayer exercise movements. Source: Ofli et al. [10].

Several measurements are made throughout the game to ensure there is sufficient lifting of the arms and the palms are not separated. If they separate and that distance exceeds a certain threshold, the system sends a message to alert the user, as illustrated in Figure 1.

Figure 2 provides an example of how feedback is delivered to the user. At the top of the screen, it can be seen the number of the exercise, its name and the number of successful repetitions. The performance bars illustrate the user's performance and their length evaluate the user's current pose, i.e. the bars are dynamic and constantly swinging according to the player's posture.

Figure 2: Feedback screen displayed during the games. Source: Ofli et al. [10].

As shown in Figure 2, feedback messages and alerts such as "sit tall" and "keep your feet on the ground" also appear on the screen. These messages are also transmitted via audio. Additionally, short sounds (e.g. "ding") are emitted whenever the repetition of movements is successfully performed.

In this study, 6 elderly participated and all of them found the feedback screen useful during the exercise and enjoyed receiving the audio motivation during the tasks. They also found the initial demonstrations helpful for the correct performance of the exercises. Although this study had some limitations in the number of participants and the duration of the tests, it was possible to hypothesize the potential of these interactive exercise systems in promoting physical activity in the elderly.

3. Study Definition

For this work, it was used the Portable Exergame Platform for Elderly People (PEPE) presented in Figure 3. This system consists of an augmented reality game platform and aims to promote the practice of sports in the elderly, contributing to their autonomy in the daily routine tasks, while simultaneously providing an entertainment activity [11]. The exergames are projected on the floor, and the player's movements are captured through a Kinect sensor, allowing the control of the game elements [12].

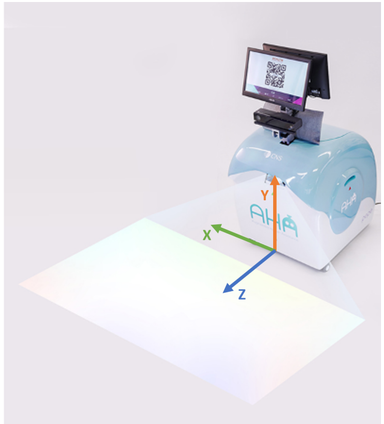

Figure 3: PEPE robotic system, with the representation of the referential. Adapted from: AHA [11].

These exergames aim to train motor skills (balance, agility and flexibility), aerobic endurance and muscular strength. In this system there are 5 exergames in total: Grape Stomping, ExerFado, ExerPong, Toboggan Ride and Rabelos. The first three concentrate on exercising the lower limbs, while Rabelos and Toboggan Ride focus on the trunk and arms.

3.1. Games Data

The data used, for this work, was collected in 5 visits to the "Residência Sénior de Belverde", Seixal, a nursing home, where 22 elders, aged over 70, participated along with 5 physiotherapists. The elderly were selected by the physiotherapists, according to their mental capacities, since it was required from the chosen players to understand the game and its objective.

In total, 158 games were played: 34 of the Toboggan Ride game, 39 of the Grape Stomping, 34 of the ExerPong, 40 of the Rabelos and 11 of the ExerFado.

During each game, the PEPE system executes 60 reads per second. Each read collects data of the current state of the game and the player's position. As it might be expected, the parameters that characterize the current state of the game vary according to its objective, meaning that each game has its own set of parameters. Concerning the position of the player, at each read, records are made for the 25 joints detected by Kinect V2. The identification of joints is shown in Figure 4, and their measurements are the same for the five exergames. The referential, from which the measurements are performed, is depicted in Figure 3.

Therefore, for each joint, the following data is recorded:

- JointType - the name of the joint, which allows its identification;

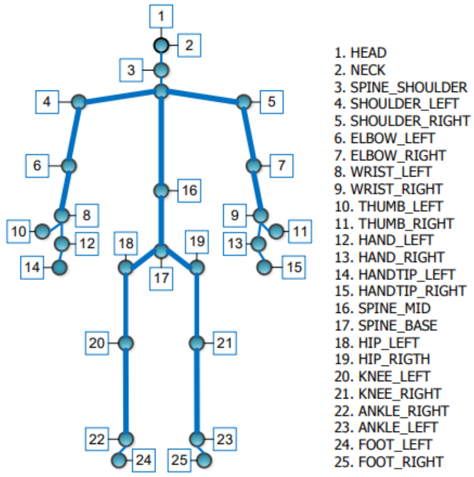

Figure 4: Joints identified by Kinect V2. Source: Ahmed et al. [13].

- PositionX, PositionY, PositionZ - XYZ joint coordinates in meters, where Y corresponds to the height of the joint in relation to the ground plane, X refers to the lateral deviation in relation to the vertical plane of the camera's centre and Z indicates the distance between the joint and the camera plane;

- OrientationX, OrientationY, OrientationZ, OrientationW - Rotation of the joint referential relatively to the referential present in Figure 3, in quaternions.

3.2. Feedbacks

All exergames were filmed in order to record the player's performance, together with the verbal feedback provided by the physiotherapists, always ensuring the protection of the identity of all participants. Therefore, 158 videos were recorded and transcribed, to obtain in writing every instruction transmitted to the elderly.

Based on the studies of Johnson et al. [14] and Burgers et al. [4], the feedback was divided into three types: explanatory, corrective and motivational. To each one of them it was established a definition. The explanatory feedback gives the player information on how he must proceed to achieve the correct answer; provides clues to improve its performance and explains why a determined answer was either right or wrong. Regarding the corrective feedback, it only indicates whether the player's response, at a specific moment in the game, is correct or incorrect, and does not provide any additional information. Lastly, the motivational feedback refers to encouragement messages so that the player does not give up and manages to complete the game.

All the transcriptions were divided and labelled with these three distinct types of feedback.

3.3. Selection of exergame and feedback type

The aim of this work is to encourage the elderly to take physical exercise by playing exergames, so it was decided to replicate only the motivational feedback. Note that, on its own, the feedback is already considered as motivation in these games. However, it was important to emphasize the messages which have a motivational nature and, for this reason, these were the focus of the work.

All messages classified with the motivational feedback label were collected and counted. As it was expected, there are a great number of words and expressions that have the same meaning, even though they slightly differ. Thus, some of these messages were grouped, accordingly with their semantics.

Subsequently, and taking into account that each exergame collects different parameters, it was decided to study only the Rabelos, since it had the highest number of executions, being played a total of 40 times. The purpose of this game is to row a Rabelo boat and collect the wine barrels, while deviating from the rocks. This game exercises the upper limbs of the participants and requires a lateral rotation of the trunk.

The parameters specific to this game, collected at each read, are described in the following points:

- TimeInSeconds - hour of the day, in seconds;

- SessionTimeInSeconds - Time passed since the beginning of the game, in seconds;

- PlayerPositionX, PlayerPositionZ, PlayerRotationY - XZY coordinates of the boat inside the game. The Z-direction is positive along the river, starting at 0. The X-axis has its origin in the centre of the river and assumes positive values to the right and negative values to the left. The parameter PlayerRotationY is the Y coordinate, in quaternions, referring to the boat's rotation;

- PickingLeft - Its values can be 0 or 1, being 1 if the gesture of catching the barrels to the left is executed successfully, and 0 otherwise;

- PickingRight - Its values can be 0 or 1, being 1 if the gesture of catching the barrels to the right is executed successfully, and 0 otherwise;

- MovingForward - Its values can be 0 or 1, being 1 if the boat is advancing in the current frame, and 0 otherwise;

- Rows - Number of paddings performed so far;

- Barrels - Number of barrels gathered so far;

- RocksHits - Number of rock collisions so far.

However, not all the parameters corresponding to the player position and of the game itself were used to perform the intended replication. Since the game Rabelos implies the execution of movements of the upper limbs, identical to the actions of rowing and collecting barrels, the chosen parameters were the ones with incidence in the arms, while the ones regarding the lower limbs were ignored.

Thus, the joints chosen were the hands, wrists and elbows of both sides. For each joint, it was considered its XYZ coordinates and its rotation regarding the referential illustrated in Figure 3. Concerning the parameters of the game itself, it were only considered the number of rows performed until a given moment (Rows) and the number of barrels collected until the time instant at cause (Barrels).

The 7 feedback sets with more samples were the ones selected to replicate in this study because it is convenient to have enough data to apply the algorithms. Hence, these feedbacks are synthesized and labelled, with numbers from 1 to 7, in Table 1.

Table 1: Counting of the feedbacks used.

Label	**Feedbacks (in portuguese)**	**Counting**
1	Bora, Bora lá, Vamos embora.	128
2	Continue, Continua a remar, Continue em frente.	161
3	Reme/Rema, Remar, Rodar, Reme bem, Temos que remar, Tem que remar.	50
4	Vamos lá,Vá lá, Vamos remar, Vamos a isso.	210
5	Isso mesmo, É isso, É isso mesmo, Assim é que é.	45
6	Boa.	148
7	Isso.	109

To align the transcribed feedback messages with the games, it was necessary to write down the time instant in which these messages were delivered, preferably in the same order of magnitude in which the data are read, that is, in milliseconds.

3.4. Methodology

In this work, it is intended to implement the ability to replicate the motivational feedback given by physiotherapists, when the elderly are playing the above mentioned exergames. To achieve this goal, it is necessary to create a mapping that allows the system to reproduce a certain feedback from the current state of the game and the player. This mapping is entitled policy [15]. For this purpose, it is required to know in advance the responses of the physiotherapists to the various states of the game.

Several approaches can be used to learn a policy, including Learning from Demonstration and Reinforcement Learning [15]. In the first one, a policy is developed from examples, which correspond to sequences of pairs action-state, recorded during teacher demonstrations. The algorithms employed in this technique use these sequences to generate a policy that reproduces the desired behaviour.

In this thesis, it was chosen the Learning from Demonstration approach because it revealed advantages over the challenges faced by the other methods, meaning that there is no need to determine an extra function that provides rewards to the system, such as in Reinforcement Learning, neither exist inaccuracies characteristic of traditional approaches.

In order to clarify concepts, the teacher demonstrations refer to the physiotherapist responses to different game situations, being the teacher a physiotherapist.

3.4.1. Learning from Demonstration

Formally, Learning from Demonstration consists of a set of states S and actions A, mapped through a probabilistic transition function. A teacher reproduces certain demonstrations D that the agent should be able to imitate. Generally, the state of the world S is not fully observable and, therefore, the agent only has access to an observed state Z, through the mapping $M : S \to Z$. However, in this work, the system is able to collect all the information of the game and the player, so the state of the world is fully observable, i.e., the state Z corresponds exactly to the state S. The policy π selects the action to execute, from the observation of the world state $\pi : Z \to A$.

In Figure 5 the derivation and execution of the policy is schematized. In a first phase, the policy is derived from the demonstrations of a teacher and then selects the action to be performed according to the current state of the world.

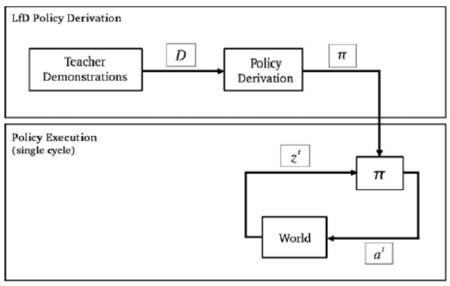

Figure 5: Policy derivation and execution in Learning from Demonstration. Source: Argall et al. [15].

There are several procedures for developing a policy based on demonstrations. Figure 6 shows a categorization of these approaches. Note that the path 1-2-3 identified in the figure represents the approaches adopted in this work.

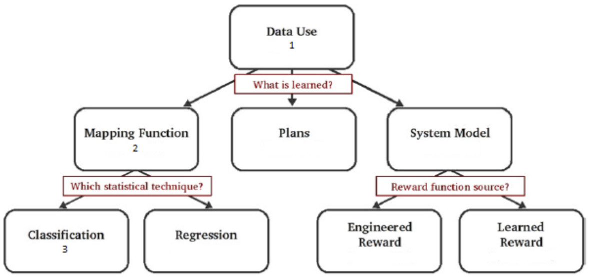

Figure 6: Categorization of approaches for learning a policy. Adapted from: Argall et al. [15].

The derivation of policies in the Learning from Demonstration approach can be done through three distinct methods: mapping function, plans, and system model, as identified in Figure 6. The approach selected in this work for the generation of a policy is the mapping function, since the system aims to imitate the teacher demonstrations, without the need of previous planning, with pre and post conditions (Plans), nor the creation of a transition model of the states of the world (System Model), which can lead to several inaccuracies in the system. This technique is considered deterministic because it is given all the necessary information for the elaboration of the policy. Its objective is based on the calculation of a policy that allows to make a generalization from the set of demonstrations, in order to acquire valid solutions also for states that are similar or close to those studied.

Overall, the techniques used in the mapping function are divided in 2 categories: classification, if the output of the algorithm is discrete, and regression, if the output is continuous. Both are represented in Figure 6. The classification algorithms categorize the input data into discrete classes, in the attempt to group similar values. In the context of this work, this approach will be used since it is intended to obtain a classification of the input data, and these data and the respective output of the algorithm are discrete.

There are several classification algorithms that can be used in this approach [16]. The k-Nearest Neighbors (kNN) and the Conditional Random Fields (CRF) algorithms were chosen for their advantages and simplicity. Besides these, the Hidden Conditional Random Fields (HCRF) algorithm is also used, which is considered an extension of CRF. Regarding the kNN algorithm, it is considered one of the simplest and oldest approaches, used in pattern classification [17]. Even so, it obtains very competitive results compared to other classification algorithms. Its performance depends fundamentally on the distance metric used to identify its nearest neighbours. Therefore, this metric must be adapted to each problem individually [17], which will be analysed later in this work.

4. Results and Discussion

In order to perform the tests, it was necessary to firstly generate the data files to be studied. Afterwards, the kNN, CRF and HCRF classification algorithms were applied to these files.

4.1. Pre-processing of data

About 20 files with different sizes were developed to apply the kNN and CRF algorithms and they can be grouped in 3 parts.

In the first part, there are files with time intervals of the same value before and after the instant of time that defines the beginning of the feedback. The range of values tested extends from 200 to 2000 milliseconds. It was not considered necessary to test values below 200 milliseconds due to the reaction time of physiotherapists [18]. On the other hand, it was assumed that the physiotherapist would not take more than 2 seconds to react to the behavior of the elderly. Consequently, the values analysed in this part were 200, 250, 300, 350, 500, 1000, 1500 and 2000 milliseconds, before and after the instant of time that defines the beginning of each feedback.

In the second part, it was maintained the same set of values of 200, 250, 300, 350, 500, 1000, 1500 and 2000 milliseconds before the instant of time that marks the beginning of the feedback and 0 milliseconds in the interval after that instant, because when the feedbacks are spoken, they refer to something that has already happened, i.e., that precedes the feedback.

In the third part, it was taken into account the error that could exist when the alignment between the video files of the games and the messages resulting from the transcriptions was made. For this reason, an error of 300 milliseconds was assumed, which might have occured at the moment corresponding to the establishment of each game's beginning. Thus, the time intervals analysed in this part observe only the 300 milliseconds after that instant. Regarding the values that precede it, the intervals 300, 500, 750, 1000 and 1500 milliseconds were considered, which are never lower than the error established.

4.2. KNN Results

The *fitcknn* function from MATLAB allows the determination of the most appropriate distance metric and the optimal number of nearest neighbours to consider for each input file. As a result of this function, it was found that for the majority of tests this distance was the city block and only 1 nearest neighbour should be considered.

Hence, the results of these tests revealed that the accuracies obtained are relatively close, varying in a range between 0.1726 and 0.1890. The standard deviations were considered low compared to the calculated means, reaching a maximum value of 0.0133. Either with the part 1 or part 3 files, accuracy values were always higher than 0.18, which demonstrates the possible relevance of the time interval after the start of the feedback.

4.3. CRF Results

To test this algorithm, the 20 data files were also used. Additionally, 3 different values were tested for the *windowSize* parameter, being them 0, 1 and 3. The *windowSize* value defines the amount of data that is used, before and after the current observation to predict its classification. A total of 60 tests were performed with this algorithm.

With *windowSize* = 0, the accuracy results are between 0.1814 and 0.2133. For *windowSize* = 1, the accuracies obtained are in the 0.1802 to 0.2259 range, and finally, with *windowSize* = 3, results between 0.1802 and 0.2242 were achieved.

In general, the range of values was similar for the 3 *windowSize* values, and the same happened for the standard deviation results. Comparing the outcomes of the 60 tests, it can be verified that in 50% of the cases the highest accuracies were achieved with *windowSize* = 1, followed by *windowSize* = 3, with the best results in 35% of the cases. It is also possible to notice that accuracies tend to increase when the intervals before and/or after the feedback instant increase in value.

Analysing in more detail the accuracies of each of the labels, it is noted that there is a preference for those classes that have a higher number of samples, and those that have fewer samples were barely used in predicting the data.

4.4. HCRF Results

In the application of this algorithm only 2 files were utilised due to the extension of the tests performed in each one of them. So, the ones chosen were those that cover the time intervals of [300,300]ms and [500,300]ms. The range of values applied to test the *nbHiddenStates* and *windowSize* parameters was 1, 2, 3 and 5 for the first and 1, 3 and 5 for the second.

This algorithm was applied in two different situations. In the first, there was no manipulation of the data, while in the second, the data used in the training of the classifier was replicated. In other words, the number of samples in the training sets for each of the labels was balanced in order to eliminate the tendency/preference for certain classes. This replication aims to obtain better results than those achieved in the first case (without replication).

Therefore, the results achieved for each of the files are displayed in Figures 7 and 8. These values refer to the results obtained with the test sets, with and without replication of the data used in the training of the classifier (note figure legends).

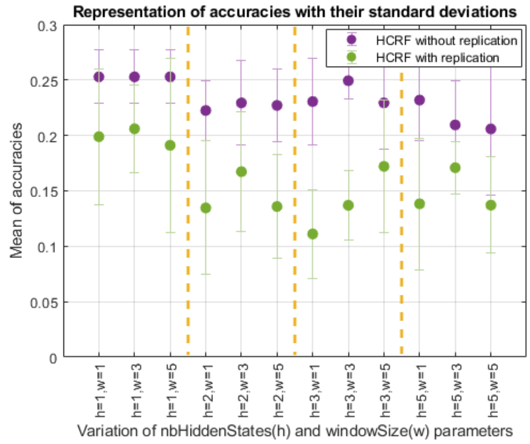

Figure 7: Accuracy results of [300,300]ms file.

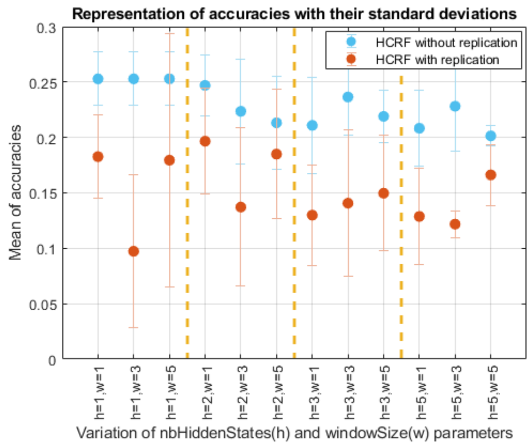

Figure 8: Accuracy results of [500,300]ms file.

In the case of the [300,300]ms file without replication, it can be seen that the accuracy results are quite identical, ranging from 0.205 to 0.253. Regarding their standard deviations, they range from 0.016 to 0.06, reaching even relatively high values, taking into account the obtained mean results. In the case with replication, the accuracy values oscillated more than the ones indicated in the previous situation. This variation comprises mean values between 0.111 and 0.206. The standard deviations also reached relatively high values, compared to the respective accuracy means, and encompass results between 0.0238 and 0.0783.

With the data file [500,300]ms without replication, it was observed that all accuracy values are greater than 0.20, oscillating more specifically in a range between 0.201 and 0.253. Their standard deviations show a variation between 0.009 and 0.048. In the case of data replication, there was an oscillation in the results of the accuracy means between 0.0973 and 0.1966, never reaching the value of 0.20, and a variation between 0.0119 and 0.1147 in the values of the standard deviation.

Examining in more detail the specific accuracies of each label, it can be verified that in the case without manipulation of data, there was a noticeable preference for the prediction with labels 2, 4 and 6, with emphasis on label 4. This is due to the fact that these are the classes with the highest number of samples in the training sets. In the situation where the data was replicated, there was an improvement in this "preference", since the choice of labels in the prediction is more distributed.

4.5. Comparison of kNN, CRF and HCRF results

For the same set of data files, the kNN algorithm revealed accuracies between 0.17 and 0.19, never reaching values above this maximum limit, while the CRF algorithm registered accuracy results between 0.18 and 0.22, for any value of *windowSize* tested. So, it can be concluded that the CRF algorithm demonstrated a better performance in its predictions than the kNN algorithm.

Focusing only on the CRF algorithm, it was noted that when the *windowSize* takes the value 1, it is obtained accuracy values above 0.21 for 9 times. On the other hand, when this parameter takes the value 0, only 1 result higher than 0.21 is counted and 2 results when the *windowSize* takes the value 3. Thus, it can be stated that the CRF algorithm with *windowSize* = 1 obtained better results in general than the other two values tested for this parameter.

In order to facilitate the comparison with the HCRF algorithm, the results were summarized in the Tables 2 and 3. The accuracy mean indicated in the CRF algorithm corresponds to the highest achieved among the 3 *windowSize* values tested. For the HCRF, the highest and the lowest accuracy mean are displayed with the respective combination of parameters. The parameter *nbHiddenStates* is represented by h and the *windowSize* by w.

From Table 2, it can be observed that the worst value obtained with the HCRF is quite close to the best achieved by the CRF. Regarding the highest result reached by HCRF, it is possible to verify that it is far from the others and an accuracy of 0.24 was never achieved with any of the files studied with the CRF algorithm.

Concerning what was obtained with the [500,300]ms file, displayed in Table 3, it can be seen that the results are quite identical to those revealed in Table 2. The worst result of the HCRF algorithm is close to the one achieved by CRF and the best is also close to 0.24, distancing itself from the remaining ones.

Additionally, and as expected, in both cases the kNN algorithm presented the worst performance of the 3 classification algorithms studied.

Table 2: Accuracies of the [300,300]ms file with the 3 algorithms.

Classification algorithms	**Parameters**	**Accuracy mean**
kNN	-	0.1866
CRF	$w = 3$	0.2083
HCRF	$h = 5$, $w = 5$	0.2059
	$h = 3$, $w = 3$	0.2494

Table 3: Accuracies of the [500,300]ms file with the 3 algorithms.

Classification algorithms	**Parameters**	**Accuracy mean**
kNN	-	0.1835
CRF	$w = 3$	0.2059
HCRF	$h = 5$, $w = 5$	0.2014
	$h = 2$, $w = 1$	0.2470

4.6. Comparison of HCRF results with and without replication

Analysing the Figures 7 and 8, it can be observed that, in both cases, the highest accuracies were always achieved without data replication. Furthermore, these results demonstrate a smaller oscillation of their values and reveal lower standard deviations. Comparing the size of the data files, the advantage between any of the files is not perceptible. Both have accuracies between 0.20 and 0.25, and standard deviations with similar intervals.

The main objective of replicating the data in the training sets was not achieved. However, as a result of the analysis of the accuracies of each label individually, it was possible to consider more classes during the prediction process. This means that the balance in the number of samples per label, allowed the classifier not to focus too much on specific labels, ignoring the others. Due to this fact, replication revealed an advantage over the application of the HCRF algorithm without any manipulation.

4.7. Other approaches

Other strategies could have been tested in the attempt to obtain better accuracy values. Besides the replication of the training set, the elimination of labels with the lowest number of samples could have been implemented, as well as the reduction of the number of samples from the training sets to their minimum value, or even the replication but with a small amount of noise, to avoid a full repetition of specific samples of the training data.

Consequently, the best combination of parameters was selected, for each one of the files indicated in the Tables 2 and 3, and these 3 additional cases were tested: HCRF with replication and noise, HCRF with training set reduction and HCRF with label elimination. The results obtained for the new tests are presented in Tables 4 and 5.

Table 4: Results of the new strategies for the [300,300]ms file.

Strategies	**Modifications**	**Accuracy mean**
HCRF (without manipulation)	-	0.2494
HCRF with replication (with and without noise)	Training set	0.1370
	Training set with noise	0.1954
HCRF with reduction	Training set	0.1555
HCRF with label elimination	All data	0.2754

Table 5: Results of the new strategies for the [500,300]ms file.

Strategies	**Modifications**	**Accuracy mean**
HCRF (without manipulation)	-	0.2470
HCRF with replication (with and without noise)	Training set	0.1966
	Training set with noise	0.2588
HCRF with reduction	Training set	0.2094
HCRF with label elimination	All data	0.2938

As can be seen, in both cases, the best results were obtained with the elimination of the labels with the smallest number of samples. The reduction of the data from all training sets was not advantageous, because the difference between the number of samples was significant and, most likely, the training sets became too small, leading to the training of the classifier to be ineffective. On other hand, the noise revealed an improvement in accuracy compared to the case with the full replication of the data. This reveals that data replication with values of features close to the real ones was advantageous for the prediction of the classifications. Thus, if the 7 labels were to be maintained, this would be the most suitable approach.

By comparing the size of the files, it is now possible to verify that, in general, the file covering a longer interval of time achieves better accuracy results than the other file studied, with a smaller interval.

5. Conclusions

The aim of this work was to generate a model of a classifier to predict the motivational feedbacks to be transmitted to the elderly at certain moments of a game. The data was obtained from sessions where the elderly played the exergames while the physiotherapists provided the necessary indications. The motivational feedbacks selected for replication correspond to those with a higher number of samples, in order to contribute to the training of the classifiers.

The study was based on 3 different classification algorithms: kNN, CRF and HCRF. After an analysis of the results achieved with each of the algorithms, it is concluded that the one with the best performance is the HCRF.

Afterwards, changes were made in the application of the HCRF algorithm, which consisted in replicating the data used in the training of the classifier. Although there were no improvements in the accuracy of the results, this replication revealed that, with the balance of the number of samples per label, there is a more uniform distribution of the labels used in the prediction, not giving preference to those with a higher number of samples.

It was also verified that each data file revealed different results, depending on the parameter set *nbHiddenStates* and *windowSize* selected. The same combination of these parameters may not be ideal for the same file, with different training and test sets. Hence, it is necessary to test several values on these parameters in order to achieve a combination that produces good results.

Other approaches could have been developed to improve the accuracies achieved. Among them, it is highlighted the elimination of labels with a lower number of samples, which reached the highest accuracy values in this work. The results obtained, when eliminating the labels 3 and 5, confirmed that the discrepancy between the number of samples per label was the main reason which led to a preference for those with a higher number of samples, disregarding the others. However, and keeping the 7 labels, it is concluded that the best option would be to perform the replication of the training data with a small amount of noise. Therefore, the data replicated are close but not exactly the same as the originals, which proved to be an advantage, as the accuracy results were higher for this case, compared to the others which also used the 7 labels.

Although an overall accuracy of 0.2588 may seem low, it is higher than predicting labels randomly. More specifically, if the prediction of the 7 labels were random, the probability of success would be, approximately, 0.14 ($1/7 \approx 0.1428571$), so this classifier is already demonstrating an advantage.

5.1. Future work

Further decisions could have been taken when selecting study data. For example, the choice of parameters from the games and joints read by Kinect could have been different from the one selected. Additionally, it could also be tested more or less parameters and analyse the influence that this would have on the results. Beyond that, more tests can be performed with the HCRF algorithm for data files with other dimensions, which was not part of the experiments performed in this study. The set of parameters used in the HCRF algorithm (*windowSize* and *nbHiddenStates*) can also be extended, especially if the data file size is increased. It is also suggested to further analyse the latest approaches indicated in subsection 4.7, since some of them seemed promising, in terms of improving the accuracy of the classifier.

References

[1] J. A. G. Marin, K. F. Navarro, and E. Lawrence. Serious games to improve the physical health of the elderly: A categorization scheme. In *CENTRIC 2011, The Fourth International Conference on Advances in Human-Oriented and Personalized Mechanisms, Technologies, and Services*, pages 64–71. IARIA XPS Press, 2011.

[2] U. N. DESA. World population prospects 2019: Highlights. *New York (US): United Nations - Department for Economic and Social Affairs, Population Division*, 2019.

[3] L. H. Larsen, L. Schou, H. H. Lund, and H. Langberg. The physical effect of exergames in healthy elderly - a systematic review. *Games for Health Journal*, 2(4):205–212, 2013.

[4] C. Burgers, A. Eden, M. D. van Engelenburg, and S. Buningh. How feedback boosts motivation and play in a brain-training game. *Computers in Human Behavior*, 48:94–103, 2015.

[5] T. Baranowski, K. Bower, P. Krebs, C. Lamoth, and E. J. Lyons. Effective feedback procedures in games for health. *Games For Health Journal*, 2(6):320–326, 2013.

[6] D. Djaouti, J. Alvarez, J. P. Jessel, and O. Rampnoux. *Origins of Serious Games*, pages 25–43. Springer, Londres, 2011.

[7] N. Zeng, Z. Pope, J. E. Lee, and Z. Gao. A systematic review of active video games on rehabilitative outcomes among older patients. *Journal of Sport and Health Science*, 6(1):33–43, 2017.

[8] K. Tsiakas, M. Kyrarini, V. Karkaletsis, F. Makedon, and O. Korn. A taxonomy in robot-assisted training: Current trends, needs and challenges. *Technologies*, 6(4):1–19, 2018.

[9] D. Siewiorek, A. Smailagic, and A. Dey. Architecture and applications of virtual coaches. *Proceedings of the IEEE*, 100(8):2472–2488, 2012.

[10] F. Ofli, G. Kurillo, Štěpán Obdržálek, R. Bajcsy, H. B. Jimison, and M. Pavel. Design and evaluation of an interactive exercise coaching system for older adults: Lessons learned. *IEEE Journal of Biomedical and Health Informatics*, 20(1):201–212, 2016.

[11] Augmented Human Assistant (AHA). `http://aha.isr.tecnico.ulisboa.pt/`. Acedido a 2020-08-20.

[12] M. Čaić, J. Avelino, D. Mahr, G. Odekerken-Schröder, and A. Bernardino. Robotic versus human coaches for active aging: An automated social presence perspective. *International Journal of Social Robotics*, 12(4):867–882, 2020.

[13] F. Ahmed, P. P. Paul, and M. L. Gavrilova. Dtw-based kernel and rank-level fusion for 3d gait recognition using kinect. *The Visual Computer*, 31(6-8):915–924, 2015.

[14] C. I. Johnson, S. K. T. Bailey, and W. L. V. Buskirk. *Designing Effective Feedback Messages in Serious Games and Simulations: A Research Review*, pages 119–140. Springer, Cham, 2017.

[15] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 2009.

[16] J. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, page 282–289. Morgan Kaufmann Publishers Inc., 2001.

[17] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10(9):207–244, 2009.

[18] J. Shelton and G. P. Kumar. Comparison between auditory and visual simple reaction times. *Neuroscience and Medicine*, 1(1):30–32, 2010.