GAIT FEATURE SELECTION IN WALKER-ASSISTED GAIT USING NSGA-II AND SVM HYBRID ALGORITHM

Maria Martins[1], Cristina Santos[1],Lino Costa[1], Anselmo Frizera[2]

[1]Minho University, [2]Federal University of Espirito Santo

ABSTRACT

Nowadays, walkers are prescribed based on subjective standards that lead to incorrect indication of such devices to patients. This leads to the increase of dissatisfaction and occurrence of discomfort and fall events. Therefore, it is necessary to objectively evaluate the effects that walker can have on the gait patterns of its users, comparatively to non-assisted gait. A gait analysis, focusing on spatiotemporal and kinematics parameters, will be issued for this purpose. However, gait analysis yields redundant information and this study addresses this problem by selecting the most relevant gait features required to differentiate between assisted and non-assisted gait. In order to do this, it is proposed an approach that combines multi-objective genetic and support vector machine algorithms to discriminate differences. Results with healthy subjects have shown that the main differences are characterized by balance and joints excursion. Thus, one can conclude that this technique is an efficient feature selection approach.

Index Terms**—**Evolutionary algorithms, Walker-assisted gait, SVM, NSGA-II, Rehabilitation.

1. INTRODUCTION

Walker-assisted rehabilitation is becoming popular, since walkers present characteristics that make use of the residual capabilities of its users enhancing motor strength capabilities [1]. Therapists evaluate assisted motor function based on visual information and personal expertise. Gait evaluation is manually assessed and the final clinical decisions are empirical depending on the physician experience. This leads to different interpretations of the evolution of the patients' treatment.

Therefore, it is necessary to better characterize the interaction between walker-patient, as well as the benefits that walkers can bring to their gait. Patients who need to use walkers suffer from balance and posture problems that change their gait pattern. These changes are defined by features related to gait cycle, trunk bending, loss of balance, joints abnormal orientation, abnormal walking base, small/large steps, problems with propulsion, and others.

In order to be able to verify differences in such specific features, and thus quantify changes due to walker treatment, one needs to access differences and similarities between the corresponding relevant features. These include spatiotemporal and kinematics parameters of the gait.

Automatic identification and classification of the patient's gait performance with the use of walkers has not been attempted before, as far as the authors know.

Thus, the first step is to identify through classification which subset of gait features is the most suited one to discriminate between assisted and non-assisted gait. Before implementing the classifier, feature selection is an important pre-processing step for pattern recognition. In this process, less discriminatory features are eliminated, leaving a subset of the original features which retains sufficient information to discriminate among classes. Thus, it can discard irrelevant and redundant information that affect the classifier's performance [2].

However, to develop an efficient approach for extracting useful information from gait features is a challenging task. Gait data are characterized by its high dimensionality and the existence of high correlation between features [2].

Feature selection techniques based on machine learning have been a focus of methodological development in recent years [3]. These techniques are often used to reduce the dimension of features.

Evolutionary techniques were intensively used for feature selection to solve the combinatory problem and to provide one optimal solution with the maximum classification performance [4-5]. Genetic algorithms are a valid and efficient method to deal with this problem and proved to be more efficient than classical methods developed for feature selection [6]. Recently introduced multi-objective approach was applied in many fields [6]. A SVM-based on multi-objective optimization with the aim of minimize the risks of the classifier and the model accuracy was proposed by Bi [7]. Igel [8] implemented the same framework but with the aim of minimize the number of features of the model. Hamdani et al. [9] optimized simultaneously the number of features and the global error obtained by a neural network classifier using NSGA-II (Nondominated Sorting Genetic Algorithm-II), thus creating classifiers' ensembles. This evolutionary technique was also applied in unsupervised learning by Morita et al [5] in handwritten word recognition. However, to the authors' knowledge these approaches have not been applied to the field of gait and walkers' rehabilitation.

This paper proposes to assess quantitative gait information from assisted-gait through an automatic acquisition and selection of gait features. The main aim is to propose a gait identification approach for assisting in the decision making of the diagnosis of a patient. This will lessen the dependency and load of the clinicians.

A preliminary study with healthy subjects walking with and without a walker with forearms supports is presented. Acquisition of gait parameters was achieved using a motion camera system. It is applied the proposed feature selection technique based on the combination of NSGA-II, which is among the latest developed algorithms for multi-objective optimization [10], and SVM (based classifier) algorithms. This technique is able to identify the redundancies present on the overall gait data. Therefore, the important gait features to discriminate between the assisted and non-assisted gait are selected. So, the objective is to find the best non-dominated solution which contains more discriminant features that separate the class volumes so that the conditions (assisted and non-assisted gait) can be effectively distinguished by: minimizing the number of features and maximizing the accuracy (ACC), Mathews Correlation Coefficient (MCC), and F1 Score (F1) of the SVM classifier. A combination of 3 objectives between these 4 metrics will be tested, to find the best combination that finds the best optimal solution.

The motivation of this work is to evaluate the performance and rehabilitation of the patient in a clinic using a walker with forearms supports, by selecting the most relevant parameters that are affected by an assisted gait ambulation.

2. METHODS

5.1. Participants

In total 35 healthy young (age range 23-27 years) voluntary subjects participated in this study. Written informed consent for publication was obtained from all individuals. Work approved by Hospital of Braga Centre of Ethics.

5.2. Experimental Setting

Tests were performed using the VICON 612 motion analysis system (http://www.vicon.com/) connected to six video cameras at a frequency of 200 Hz.

Two conditions were considered: unassisted and assisted gait with a walker with forearm supports. Subjects were fitted with fifteen reflecting spherical markers according to the marker set-up described by Vaughan et al [11]. All subjects were barefoot and asked to walk in a 10 m straight-forward path with a self-preferred walking speed, 3 times each condition. The walker upper base height was adjusted for each subject. The subjects' elbows had to be flexed approximately 90º. Thus, subjects should assume a standing upright position with their forearms placed in the supports of the walker.

5.3. Gait Parameters

The gait features selected for this study can be grouped into 2 categories: (1) spatiotemporal and (2) joint kinematics. Since subjects are considered to present no motor dysfunctions, symmetrical gait was assumed [11]. The calculated gait features are listed in fig. 1. These features were those commonly reported in gait analysis [11] and they were calculated through the Vicon 612 analysis system. Median over the performed three trials was calculated for each subject feature and formed the basis for all subse1quent analysis. The median was chosen over the mean since it is more resistant to measurement errors. Custom Matlab Software was used for all Vicon data processing and gait features' calculation.

5.4. Feature Selection based on NSGA-II-SVM

NSGA-II is designed both for discovering good features subsets and for final feature selection and classification. This will be done by determining the best compromise between the two/three conflicting objectives. For that purpose, Support Vector Machine-based classifier (SVM) is used to ensure the fitness evaluation of each candidate feature subset by classifying them during the successive generations.

NSGA-II starts from a random population of binary individuals (chromosome) representing the subset of features for classification. In order to compare the individuals, the population is sorted based on the domination relation according to several (two or three) conflicting SVM classification performance criteria.

First, it is important to select subsets of the data to be used as training and test in the classification stage. In this study, a SIXfold cross-validation (CV) resampling approach is used to construct the learning and test sets for the SVM-based classifier. Initially, the two-group samples (assisted and non-assisted gait) are randomly divided into six non-overlapping subsets of roughly equal size, respectively. A random combination of the subsets for the two groups constitutes a test set (6 sets) and the total remaining subsets are used as the learning set (6 sets). Thus, the SIXfold CV resampling produces a total of 36 pairs (6x6 combinations) of learning and test sets. Each individual of the population is evaluated over the 36 pairs, i.e. SVM is executed 36 times, and then is calculated an average of these 36 results. Second, the fitness of each individual is computed according to SVM classification (kernel based on Gaussian radial basis functions). In this binary classification problem, it is not guaranteed the existence of a simple hyperplane as a separating criterion for the gait parameters considered. Therefore, a kernel based on Gaussian radial basis functions is adopted for SVM. Three metrics are adopted as evaluation criteria of the performance of each feature subset: Accuracy (ACC), Mathews Correlation Coefficient (MCC), and F1 Score (F1).

Since it is considered here a two-class prediction problem (binary classifier), the outcomes are labeled either as positive (p) –assisted ambulation- or negative (n) – unassisted ambulation. There are four possible outcomes from a binary classifier. If the outcome from a prediction is *p* and the actual value is also *p*, then it is called a true positive (TP); however if the actual value is *n* then it is said to be a false positive (FP). Conversely, a true negative (TN) has occurred when both the prediction outcome and the actual value are *n*, and false negative (FN) is when the prediction outcome is *n* while the actual value is *p*.

Name	Variable
Step width (m)	Width
Step length (m)	Step_l
Cadence (step/min)	CAD
Stance phase(%)	stance
Swing phase (%)	swing
Double support (%)	DS
Average Speed (m/s)	Speed
Step time (s)	Step_t
Ankle plantarflexion maximum (degrees)	APF
Ankle dorsiflexion maximum (degrees)	ADF
Ankle range of motion during gait cycle (degrees) in the sagittal plane	ATy
Maximum flexion of the knee (degrees)	KF
Maximum extension of the knee (degrees)	KE
Knee range of motion during gait cycle (degrees) in the sagittal plane	KT
Maximum flexion of hip (degrees)	HF
Maximum extension of hip (degrees)	HE
Hip range of motion during gait cycle (degrees) in the sagittal plane	HT
Maximum abduction of the hip (degrees)	Hab
Maximum adduction of the hip (degrees)	Had
Hip range of motion during gait cycle (degrees) in the frontal plane	HTx
Foot maximum progression deviation (interior rotation) (degrees)	AI
Foot maximum progression deviation (exterior rotation) (degrees)	AE
Foot range of progression deviation during gait cycle (degrees) in the transverse plane	ATz
Range of motion of sacrum (height) (m)	T10z
Lateral flexion (right) of the trunk (degrees)	SR
Lateral flexion (left) of the trunk (degrees)	SL
Pelvic lateral range of motion (degrees) in the frontal plane	ROMlat
Pelvic maximum flexion (degrees)	SF
Pelvic maximum extension (degrees)	SE
Range of motion of Ext/Flex of the trunk (degrees) in the sagittal plane	ROMFlexExt

Fig. 1. Gait Features.

Accuracy, *ACC*, is the most common and simplest measure to evaluate a classifier. It is just defined as the degree of right predictions of a model:

$$\mathrm{ACC} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FN+FP}} \quad (1)$$

An accuracy of 100% means that the measured values are exactly the same as the given values.

Matthews Correlation Coefficient (*MCC*) is a metric used in machine learning as a measure of the quality of binary classifications. It takes into account true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of very different sizes.

$$\mathrm{MCC} = \frac{\mathrm{TP\times TN - FP\times FN}}{\sqrt{\mathrm{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}} \quad (2)$$

This coefficient can be seen as a correlation coefficient between the observed and predicted binary classifications. It outputs a value between −1 and +1. A coefficient of +1 represents a perfect prediction, 0 no better than random prediction and −1 indicates total disagreement between prediction and observation.

F1 Score (*F1*) is a measure of a test's accuracy. It considers both the PC and the SST of the classification to compute the score:

$$\mathrm{F1} = 2 \times \frac{\mathrm{PC\times SST}}{\mathrm{PC+SST}} \quad (3)$$

F1 can be interpreted as a weighted average of the PC and SST, where an *F1* reaches its best value at 1 and worst score at 0.

The maximization of these performance metrics plus the minimization of the number of features allow comparing feature subsets. Consequently, better feature subsets have a greater chance of being selected to form a new subset through crossover and mutation. Crossover combines different features from a pair of subsets into a new subset and mutation changes some of the values (thus adding or deleting features) in a subset randomly.

The NSGAII-SVM algorithm is an iterative process in which each successive generation is produced by applying genetic operators to the members of the current generation. In this manner, good subsets are "evolved" over time until the stopping criteria are met. The flowchart of the method, implemented in Matlab, is presented in Fig. 2.

(1) Calculate gait features: read the matrix P x G from database, where P is the number of participants and G is the number of gait features.

(2) Generate parent population P_0: Generate *N* individuals (parent population) randomly. Each individual is a fixed-length string with a G-length of bits of either 1 or 0 (binary-coded). Features are binary-coded within each string as either presence (1) or absence (0).

(3)Parent population evaluation: (i) Fitness Calculation- each solution in population representing a combination of features is evaluated in terms of the evaluation criteria (one/two of the metrics presented in eq. (1-3) and the number of selected features). The population is sorted according to the domination relation; (ii) Crowding Distance- crowding distance is calculated for each individual. The crowding distance is a measure of how close an individual is to its neighbors. Large average crowding distance will result in better diversity in the population. To compute crowding distance for an individual, we average the distances to its immediate neighbors along the same front in every dimension (dimensions correspond to objective functions). Then, put a rank value based on its nondomination level.

(4) Generate child population Q_0: form a child population Q_0 on the basis of P_0 by performing the followed genetic operators (i) Selection- Selection operator in NSGA-II is composed of picking child population from the parent population with the same size. The binary tournament selection [12] runs a tournament between two individuals and selects the winner. It can obtain better result than the methods of proportional and genitor selection. Hence, binary tournament selection is adopted to select the next generation individual; (ii) Crossover Operator- Crossover combines two parents, to form children, for the next generation. Then a

scattered crossover is used [12]. This type of crossover creates a random binary vector. So, the genes are selected from the first parent where the vector is a 1, and from the second one where the vector is a 0, and combines the genes to form the first child, and vice versa to form the second one; (iii) Mutation Operator- Adaptive Feasible Mutation adds a randomly generated number to each element in the child population. The direction (positive or negative) of the random number is adaptive with respect to the last successful or unsuccessful generation. The feasible region is bounded by the relative constraints and inequality constraints (0 and 1) [12].

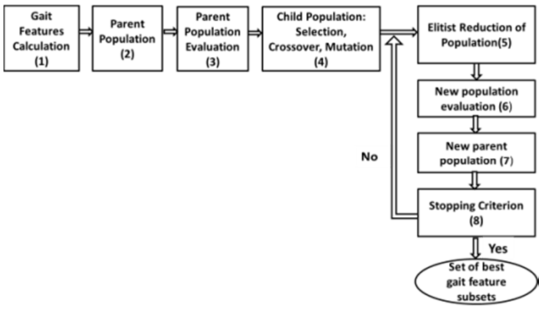

Fig.2 Flow chart of the proposed NSGAII-SVM combination.

(5) Elitist Reduction of Population: At the *t*th generation, produce population R_t of size N by integrating parent population P_t with child population Q_t.
(6) R_t population evaluation: The new population R_t is sorted on the basis of domination and evaluated as described in (3). Assign a corresponding rank.
(7) Create new parent population P_{t+1}: by filling the highest ranked front set until the size of the population size exceeds N.
(8)Stopping criterion verification: Goes to step 5 until the stopping criterion is satisfied.

3. RESULTS AND DISCUSSION

This section shows the experimental details and evaluation of the results of NSGAII-SVM for different combinations of the presented fitness objectives: number of features (NF), ACC, MCC and F1. The used parameters on the NSGA-II are presented in Table 1. The population size was 100 individuals. The evolution process ends if 100 generations are performed and/or fitness values reach zero and/or stall generations limit reaches 10. The methodology proposed will be used in the classification and discrimination of two conditions: (1) normal unassisted gait and (2) assisted gait with the use of a walker device with forearm supports.

The optimal set will contain the gait feature solutions that are trade-offs between objectives. The aim is to find the minimum number of gait features (NF) that, simultaneously, can maximize one/two of the three selected metrics (ACC, MCC and F1). Therefore, six different combinations of metrics give rise to the following multi-objective problems: (a) Bi-objective problems: Max ACC and Min NF; Max MCC and Min NF; Max F1 and Min NF; (b)Three-objective problems: Max ACC, Min NF and Max MCC; Max ACC, Min NF and Max F1; Max MCC, Min NF and Max F1.

Number of participants (P)	35x2
Size of population (N)	100
Length of the chromosome(G)	30
Stopping Criterion	100 generations
	All objectives values = 0
	Stall Gen. Limit = 10
Crossover probability	0.8

Table 1. NSGA-II parameters for the features subset selection.

Table 2 shows the achieved results. The Pareto-optimal frontier presents only three compromise solutions. For all combinations of objectives, the NSGAII-SVM found an optimal solution with only one feature (underlined), but presents better performance measures with two features (italic) and the maximum performance measures values with three features. However, with three parameters the performance reaches the maximum value for all the performance measures. It can be seen that three different combinations were found: DS, HE, SL; T10z, HE, SL; DS, HE, KT. All these combination have in common the feature HE. These parameters are in accordance with other studies that evaluated the effects of assisted gait [13].

It can be seen that hip extension (HE) is the parameter that can classify the two conditions with high performance of the classifier.

Since all the objective combinations achieved good results, the authors have to select one combination. In order to assist this decision task, one will analyze if any of the objectives is redundant and as such should be discarded.

A Principal Component Analysis (PCA) [2] enables to evaluate the relationship between objectives for the solution found. In combination ACC, MCC and NF, Component 1 (PC1) and Component 2 (PC2) explain approximately 99% of the variance. In combination MCC, NF and F1, PC1 and PC2 explain approximately 98% of the variance and PC1 and PC2 of ACC, F1 and NF explain approximately 98%.

A biplot representation of the PCA results is provided in Fig. 3 in order to further inspect the relations between objectives. Fig. 3a) enables to see the positive correlation between ACC and MCC, as well as an independency between MCC and NF; and ACC and NF (an angle of ≈ 90º). A similar relationship is observed in fig. 3b), where it exists a slighter positive correlation between MCC and F1 and independency between these and NF. An even slighter positive correlation is verified between F1 and ACC, and independency between these and NF, in fig. 3c).

The objectives ACC and MCC are closely related indicating the existence of a larger degree of redundancy. Therefore, one of these two objectives can be discarded with a smaller loss of information. Despite the other two combinations also showing a positive correlation between MCC/F1 and F1/ACC, they have less redundancy.

Thus, if we want to choose three objectives, the best combination seems to be ACC/NF/F1 and MCC/NF/F1.

However, we can have only a combination of two metrics, by eliminating one of the metrics based on its redundancy. MCC/NF is the chosen combination since these two objectives are less correlated (angle of ≈ 90º).

Since we are interested in finding a trade-off between the number of variables and a good performance of the classifier, we choose the solution with two features: HE and DS, because it still has good metrics values for less features.

This preliminary study with this technique indicates that sagittal plane movement in hip and balance are sufficient to discriminate differences between assisted and non-assisted gait.

Fitness Combination	Best evaluation features	NF	ACC	MCC	F1
ACC/NF	HE	1	93,19%		
	DS,HE	2	98,89%		
	T10z,HE,SL	3	100%		
MCC/NF	HE	1		0,8857	
	DS,HE	2		0,9814	
	DS,HE,SL	3		1	
F1/NF	HE	1			0,9389
	DS,HE	2			0,9889
	T10z,HE,SL	3			1
ACC/NF/MCC	HE	1	93,19%	0,8857	
	DS,HE	2	98,88%	0,9814	
	DS,HE,SL	3	100%	1	
ACC/NF/F1	HE	1	93,19%		0,9389
	DS, HE	2	98,89%		0,9889
	T10z,HE,SL	3	100%		1
MCC/NF/F1	HE	1		0,8974	0,944
	T10z,HE	2		0,9814	0,9888
	DS,HE,KT	3		1	1

Table 2. Results of the NSGA-II-SVM combination using the selected metrics as evaluation criterion.

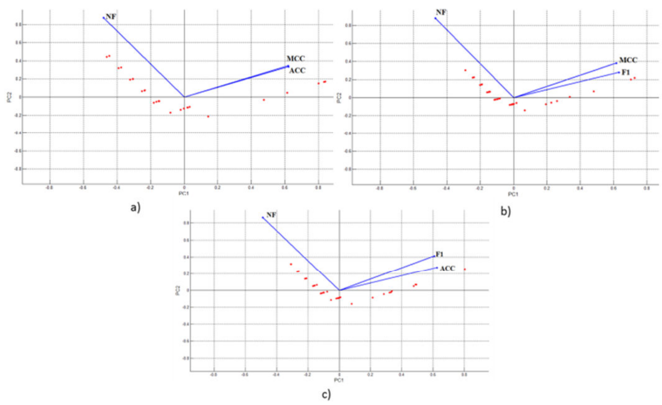

Fig. 3 PCA biplot for verifying Objective Correlation: a) Combination *ACC/MCC/NF*, b) Combination *MCC/NF/F1*; c) Combination *ACC/NF/F1*.

4. CONCLUSIONS

This work is intended as a proof of concept study made with healthy young volunteers to verify which gait parameters are the most important to detect differences between the use and non-use of a walker with forearm supports. A multi-objective NSGAII-SVM approach was applied. In order to evaluate the classification performance it was considered four types of metrics: ACC, MCC, F1 and NF.

It was chosen the objective combination MCC/NF as the best one to distinguish between the two study conditions. So, it was also concluded that assisted and non-assisted gait can be differentiated with the use of: HE and DS.

Moreover, future study will be made with elders or other type of patients to compare with the results of this paper.

ACKNOWLEDGMENT

This work has been supported by FCT – Fundação para a Ciência e Tecnologia in the scope of the project: PEst-OE/EEI/UI0319/2014. Work supported by Portuguese Science Foundation (grant SFRH/BD/76097/2011).

REFERENCES

[1] Martins, M; Frizera, A; Santos, C; Ceres, R. Assistive Mobility Devices focusing on Smart Walkers: Classification and Review. Robotics and Autonomous Systems, 2011. 10.1016/j.robot.2011.11.015.

[2] Olney, S., Griffin, M., McBride, I. 1998 "Multivariate Examination of Data from Gait Analysis of Persons with Stroke". Journal of the American Physical Therapy Association. 78:814-828.

[3] Saeys, Y., Inza, I., Larrañaga. A review of feature selection techniques in bioinformatics. Bioinformatics Advance Access, August 2007.

[4] Lee C, Landgrebe DA. Decision boundary feature extraction for neural networks. IEEE Trans Neural Networks 1997;8:75–83.

[5] Morita M, Sabourin R, Bortolozzi F, Suen CY. Unsupervised feature selection using multi-objective genetic algorithms for handwritten word recognition. ICDAR 2003;2:666–70.

[6] Sun Y, Yin L. A genetic algorithm based approach for 3D face recognition. Comput Imaging Vision 2007;35:95–118.

[7] Bi J (2003) Multi-Objective Programming in SVMs, Proceedings of the Twentieth International Conference on Machine Learning, ICML-2003, Washington DC.

[8] Igel C (2005) Multi-Objective Model Selection for Support Vector Machines, in C.A. Coello et al. (Eds.), EMO 2005, pp. 534-546, Springer-Verlag Berlin Heidelberg.

[9] Hamdani TM, Won JM, Alimi AM, Karray F. Multi-objective feature selection with NSGA II. Adapt Nat Comput Algorithms 2007;4431:240–7.

[10] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 2002;6:182–97.

[11] Vaughan C, Davis B, Connor J. Dynamics of human gait. Kiboho publishers, 2nd ed.

[12] J. C. Spall, Introduction to stochastic search and optimization, A John Wiley & Sons, Inc., Hoboken, New Jersy, 2003.

[13] Ishikura, T.2001. "Biomechanical Analysis of weight bearing force and muscle activation levels in the lower extremities during gait with a walker". Acta Med. Okayama,vol.55, nº2, pp. 73-82.