Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/bbe

Original Research Article

Gait patterns classification based on cluster and bicluster analysis

J. Pauk [a,b,*], K. Minta-Bielecka [a]

[a] Bialystok University of Technology, Bialystok, Poland
[b] Glenrose Rehabilitation Hospital, Edmonton, Canada

ARTICLE INFO

Article history:
Received 15 August 2015
Received in revised form
1 March 2016
Accepted 7 March 2016
Available online 19 March 2016

Keywords:
Hemiplegia
Gait patterns
Clustering
Biclustering
Joints moments
Spatial–temporal gait parameters

ABSTRACT

Gait patterns of hemiplegia patients have many potential applications such as assistance in diagnosis or clinical decision-making. Many techniques were developed to classify gait patterns in past years; however, these methods have some limitations. The main goal of the study was to present the performance evaluation results of the new biclustering algorithm called KMB. The second objective was to compare clustering and biclustering methods. The study was performed based on the gait patterns of 41 hemiplegia patients over 12 months post-stroke, at the age of 48.6 ± 19.6 years. Spatial–temporal gait parameters and joint moments were measured using motion capture system and force plates. Clustering and biclustering algorithms were applied for data consisting of joint moments of lower limbs. The obtained results of this study based on joint moments, clustering, and biclustering can be applied to evaluate patient condition and treatment effectiveness. We suggest that the biclustering algorithm compared to clustering algorithms better characterizes the specific traits and abnormalities of the joint moments, especially in case of hemiplegia patients.

1. Introduction

Hemiplegia is a condition caused by damage to the part of the brain that controls movements of the limbs, trunk, etc. that affects a right or left side of the body. The injury to the left side of the brain causes a right hemiplegia and injury to the right side of the brain results in left hemiplegia [1,2]. While typical gait sequences are repetitive and exhibit periodic behavior, hemiplegia patients usually have problem with gait symmetry [3]. Hemiplegic gait describes altered gait patterns that have been affected by deformity, muscle weakness, impaired motor control, and pain. Deviations of normal gait patterns are usually observed during both swing and stance phases, and they require systematic evaluation for assessment of functional compensations [3]. Accurate identification of the gait patterns from hemiplegia patients has many potential applications such as assistance in diagnosis or clinical decision-making. The application of intelligent classification systems may enable clinicians to differentiate gait patterns into clinically significant categories. In recent years, many techniques were developed to classify gait patterns, thus making it possible to automatic analysis based on the gait parameters [4–11]. Among the various methods, neural networks models and support vector machines

---

* Corresponding author at: Bialystok Technical University, Wiejska 45C, 15-351 Bialystok, Poland.
E-mail addresses: j.pauk@pb.edu.pl (J. Pauk), kasia.minta@gmail.com (K. Minta-Bielecka).

http://dx.doi.org/10.1016/j.bbe.2016.03.002

(SVM) have previously been applied to identify hemiplegia patients. Clustering, a basic tool used in machine learning and patterns classification, is also a very popular technique used to analyze gait patterns. This method involves separating different objects into groups of similar objects called clusters such that the cluster variations are minimized and the variations between clusters are maximized [12,13]. In recent years, various clustering methods have been applied to resolve numerous classification problems [14,15]. In patterns classification, clustering is usually applied to find natural groupings in the data. The main limitation of standard clustering approaches is that they are very sensitive to variations in the gait trials, especially in the hemiplegia patients. The technique that overcomes the limitations of standard clustering approaches may be biclustering, which, in case of the gait patterns analysis, groups both patients and samples simultaneously. Cheng and Church [16] are among the pioneers in introducing the concept of biclustering in data analysis. Their biclusters were based on uniformity criteria, and a greedy algorithm was developed to discover biclusters. Biclustering has been also applied for dimensionality reduction of databases via automatic subspace clustering of high dimensional data, and has been used in a number of other areas [17,18]. The concept of associating biclustering with biological data analysis allows identification of sets of patients sharing consistent expression patterns over the subsets of samples [16–18]. Unfortunately, this method has some limitations. First, it is not easy to present all biclusters simultaneously. Second, a bicluster can show complicated coherent patterns. Finally, the criterion to evaluate a biclustering algorithm depends on the structures of biclusters to be detected [19]. This gave motivation to search for a new efficient algorithm such as proposed in this paper KMB, which can be used to analyze gait patterns. The main goal of this paper was to present the performance evaluation results of the new biclustering algorithm called KMB. The second objective was to compare clustering and biclustering methods.

2. Material and methods

2.1. *Subjects*

The study included the gait patterns of 41 hemiplegia patients over 12 months post-stroke. Inclusion criteria for hemiplegia patients characteristics were applied, such as first ever stroke, ischemia or intra-cerebral hemorrhage assessed by computerized tomography or magnetic resonance imaging, occurrence of stroke at least 12 months before enrolment in the study, ability to walk minimum 10 m independently, absence of prior treatments of orthopedic surgery, and cognitive ability to follow the study instructions. Patients were excluded if they had received botulinum injections or a phenol nerve block in the hemiplegic lower limb within the last 6 months before the evaluation, and if they had any other medical disorders, which might have adversely affected their gait patterns. The patient's level of functional independence was assessed by neurologists through the modified Barthel index (BI) [8]. The maximum score of 100 means that the patient is fully independent in physical functioning. The lowest score of 0 corresponds to the totally dependent, bedridden person. The study was approved by local ethics committee. All subjects received full information about the study before giving signed informed consent.

2.2. *Measuring spatial–temporal gait parameters and joint moments*

Gait parameters were measured using a motion tracking system (Motion Analysis Corp., USA) and two AMTI force platforms (Advanced Mechanical Technology, Inc., USA). Temporal parameters were estimated using sample frequency of 1000 Hz, whereas spatial–temporal gait parameters were estimated using sample frequency of 120 Hz. Each subject was asked to walk barefoot at habitual speed. The measurements were repeated to obtain at least three valid walking trials. Mean values of stride duration (s), velocity (m/s), stance duration (%), stride length (m), and step length (m) of both affected and unaffected side were calculated. The spatial–temporal gait parameters have been summarized in Table 1. The hip, knee, and ankle joint moments were computed using an inverse dynamic approach [3] and were normalized to the body mass. The signals were initially low-pass filtered using a Butterworth filter.

2.3. *Agglomerative hierarchical clustering*

The data set included 177 gait cycles of joint moments for 41 hemiplegia patients. A one-way clustering analysis was considered to be reliable and stable for a data set with continuous variables. The data set is divided into many small sub-clusters, which are then combined into a specified number of clusters using an agglomerative hierarchical clustering algorithm (AHC). In the clustering algorithm, the Ward's method has been applied as agglomeration method, and the Euclidean distance as a dissimilarity measure.

2.4. *A new biclustering algorithm KMB*

In this article we proposed the new algorithm called KMB, which was based on two existing biclustering methods such as node-deletion algorithm [16] and node-addition algorithm [18]. Examination results were represented as a matrix $A = (R,C)$ with rows R corresponding to walking trials and columns

Table 1 – Spatial–temporal gait parameters for affected and unaffected side in hemiplegia patients.

Gait parameters	Hemiplegia		*p*-Value
	Affected side	Unaffected side	
Stride duration (s) Mean (SD)	1.48 (.44)	1.51 (.56)	$p = .306$
Velocity (m/s) Mean (SD)	.73 (.41)		
Stance duration (%) Mean (SD)	63.82 (7.71)	68.13 (8.66)	$p < .05$
Stride length (m) Mean (SD)	.93 (.35)	.93 (.37)	$p = .479$
Step length (m) Mean (SD)	.48 (.16)	.47 (.19)	$p = .316$