Contemporary techniques to manage of databases in gait analysis

Współczesne techniki zarządzania bazami danych z analizy chodu

dr inż. Małgorzata Syczewska*, dr inż. Piotr Wąsiewicz**

* Dept. Paediatric Rehabilitation, The Children's Memorial Health Institute, Warsaw, email: m.syczewska@czd.pl
** Institute of Electronic Systems, Warsaw University of Technology, Warsaw, email: pwasiewi@elka.pw.edu.pl

Key words: gait analysis, analytical techniques (analiza chodu, metody analityczne)

ABSTRACT

Data obtained from modern movement analysis systems are challenging to analyse. There are several reasons for that:

- large number of data obtained during the session;
- their multi-dimensionality,
- most of them are time-dependent;
- and often depend on each other.

During last years many different analytical techniques are used to deal with them in order to better understand the physiology and pathophysiology of the human movement (especially gait). This paper presents most commonly used and promising techniques.

STRESZCZENIE

Liczba danych otrzymywanych obecnie ze współczesnych systemów analizy ruchu jest trudna do analizy. Składa się na to kilka przyczyn:
- duża liczba danych otrzymywanych podczas jednej sesji analizy chodu;
- ich wielowymiarowość;
- większość z nich charakteryzuje się zależnością od czasu;
- często są od siebie wzajemnie zależne.

W ostatnich latach w literaturze przedmiotu pojawia się coraz więcej prac w których do opracowania danych pochodzących z analizy chodu stosowane są różne techniki analityczne. Ich zastosowane ma na celu lepsze zrozumienie fizjologii i patofizjologii ludzkiego ruchu (zwłaszcza chodu). Celem niniejszej pracy jest przegląd najpopularniejszych i najbardziej obiecujących z nich.

INTRODUCTION

The modern system for objective gait analysis provide large number of data from single session. Usually the gait analysis report comprise a series of graphs representing angle changes in 3-dimentional space during gait cycle of ankle, knee, and hip joints of both legs, orientation of pelvis in space, kinetic data (moments at the leg joints, ground reaction forces), spatio-temporal data (cadence, velocity, step length percentage of gait phases in gait cycle), and EMG data. These data is used for drawing clinically meaningful conclusions and planning further treatment of patient. Sometimes is also used for foreseeing the evolution of pathophysiological gait. The use of such amount of data for other purposes is challenging due to the following reasons:

- large amount of data as the output of single session;
- the multi-dimensionality of the data sets;

- most of the data from 3D gait analysis is time-dependent;
- some of the data (especially kinematic) are depended on each other.

There is a strong need in clinical setting to use gait data for other than single clinical decision making purposes:

- classification of gait patterns;
- assessment of the severity of gait deviations;
- monitoring of the patient status during long-time treatment;
- comparison of different treatment methods.

Therefore during last years different analytical techniques coming from other fields are transferred to gait analysis in order to solve the above described problems. The purpose of this paper is the presentation of these techniques and their application in management of gait data.

PRINCIPAL COMPONENT ANALYSIS (PCA)

The advantage of principal component analysis is its ability to reduce the dimensionality of the data sets, while keeping the original variability of the data [Chau, 2001, part1]. The algorithm used transforms the original variables X_i, i= 1, 2,.. n, into new orthogonal variables (or components) P_j, which preserve the variability of original variables. The new, principal components are linear combinations of the original variables:

$P_j = a_{j1}X_1 + a_{j2}X_2 + ... + a_{jn}X_n$, j= 1, 2,.. k=n,

The coefficients a_{ji} (called factor loadings) reflect the amount of variance in variable X_i captured by P_j. For a given data set the principal component analysis provides a unique solution.

The application of principal component analysis to gait analysis comprise of the following steps [Sadeghi et al., 2000]:

- choice of the gait parameters taken for the analysis;
- finding the covariance matrix of these parameters;
- calculation the eigenvalues and eigenvectors of this matrix;
- choice of the number of principal components used for further analysis; according to Kaiser criterion the components which account for more than 1 % of the variability should be taken, but usually the number of components is chosen on the criterion, that the sum of them account for most of the total variability (eg. 70 %).

Principal component analysis could be used for discrete sets of variables, but can be also used for analysis of the waveforms [Muniz & Nadal, 2009]. This feature is especially interesting for gait analysis, as many data (kinetic, kinematic, electromyography - EMG, or ground reaction force) are time-dependent and can be treated as variables. This approach has the advantage of using the information from the whole time-curve, and not only from certain minimal or maximal values during the gait cycle (or certain gait phases). Principal component coefficients could be used as factors characterizing the individual gait pattern, which differentiate the patient's gait from the normal reference set.

FUZZY LOGIC

Fuzzy logic enables creating from the continuous variable domain sets which interfere with each other. Classical sets are separated and a data point belongs only to one chosen set. Fuzzy set membership function can overlap with the another one at the particular data point. The data point membership value in such a set is somewhere between 0 and 1 this means it is not always equal to 0 or 1. Membership functions are simply fuzzy sets.

To generate such the sets the fuzzy clustering algorithm should be applied e.g. fuzzy c-means. It helps to categorize and classify the gait of subjects into healthy or pathological groups [Chau, 2001, part 1].

The human reasoning and derived from it rules are very often fuzzy and not very precise e.g. this man is a little awkward and that woman is quite habile. With fuzzy logic such the human experts rules can be applied to the gait data obtained e.g. from pressure sensors placed in shoes and as a result a gait stability index value is calculated after defining fuzzy sets and their final product.

When there is available only sensors data, then such rules can be generated by a custom algorithm: one rule for each possible combination of input data parameter fuzzy sets and chosen arbitrary output index parameter fuzzy sets [Biswas, 2008].

Fuzzy logic has also a few disadvantages:

- the output is not binary, it classifies data points from the continuous variable domain,
- it is quite arbitrary to determine a number of fuzzy sets and their shape and overlapping, it should be done a-priori
- relationship among different rules is unclear,
- design and computing algorithm can be hard to follow, though the fuzzy logic basis is quite simple.

NEURAL NETWORKS

The popularity of use of neural networks in automatic recognition of different gait patterns is constantly growing [Wu, et al., 2007, Chau, 2001]. The researcher use various sets of data (kinematic, ground reaction forces, foot pressure distribution data), various configurations of the networks and various learning stratergies. The advantage of the artificial neural networks is their flexibility, and non-linearity, which enables to model complicated non-linear dependencies, difficult to model with traditional analytical methods.

There are several shortcomings of this approach to gait data analysis:

- The large amount of the gait data require the selection of subset of data used for the analysis. This selection is arbitrarily done by the researcher and the final classification done by the network depend on the subset chosen.
- The network has to be trained on well defined pathological gait patterns, which should be distinct from one another. The new abnormal pattern could not properly classified, the networks have also problems to distinguish between the patterns of different origin, but close to each other.
- The networks cannot process directly the raw gait data, therefore some kind of pre-processing is required (normalization, fast Fourier transformation, rectification, averaging). The type and amount of pre-processing influence the training and performance of the networks, affecting the obtained results.
- The choice of the network architecture is crucial, as they are structure-dependent and hierarchy sensitive

Artificial neural networks usually consist of [Chau, 2001, part 2] inputs, outputs, and some processing between the two. This computational process is called hidden layers. The inputs should be selected from the large set of gait data in such a way, that they are independent variables. The training process of the networks adjust the internal parameters of the computational process according to the assumed criterion, i.e. prediction error falls within the preset interval.

Neural networks are not only use to classification task, but also for modelling. Their non-linearity encouraged the researchers to apply them to model the relationships between the EMG (representing muscle activity and force), and kinematic and kinetic parameters.

INDEXES

GGI

The Gilette Gait Index [Shuttle et al., 2000] (or normalcy index) was designed in order to express the gait pattern of a given subject by one number. This number could be considered as measure of distance between the set of discrete gait parameters of a patient from similar set of a healthy subject (or average of several such sets of healthy controls). The sets could be viewed as vectors in multidimensional space, in which the number of variables is equal to the space dimensions (and vector’s length). To avoid the error arising from two sources (correlation between some of the

variables, and different units in which they are expressed) the methods of multivariate statistics are used to calculate the distance. Through these methods the variables are uncorrelated, expressed in new uncorrelated coordinate system, and then the distance between the two vectors is calculated.
For calculation of GGI the 16 gait parameters were used: some spatio-temporal, other kinematc, describing range or peak angle at certain joints, in certain gait phases. The choice of the set and the number of the parameters used for calculation of GGI is usually based on clinical experience. Using similar method other types of indexes (based on different sets of variables) could be created. The choice of the variables could depend on the clinical picture of the pathology, which would be evaluated.

GDI

The development of GDI (gait deviation index) was based on the biometric method used for face identification [Schwartz, Rozumalski, 2008]. In this method the face is scanned, the data is first converted into greyscale and than to vectors. Further these vectors are processed using principal component analysis (described in earlier part of this paper). The resulting eigenvectors comprise most of the information and their linear combination is further used for face identification. In application of this method to gait analysis the digitised face was transferred to kinematic plots, and greyscale to joint angles.
In the calculation of GDI left and right side are processed separately. For example, the pelvic angles and hip angles in all three planes, knee flexion / extension and foot dorsi / plantarflexion angles (sagittal plane) together with foot progression angle (transversal plane) are incremented at 2 % interval in the gait cycle. This data form a single vector of length equal to 9 angles x 51 points = 459. This vector reflects the gait pattern of a given subject and can be used for further analysis and comparisons.
If there are two subjects who underwent gait analysis the results of these analyses could be summarized as two vectors, each representing individual subject gait pattern. Euclidean distance of these two vectors describe the similarity (or dissimilarity) of their gait.
The results of the gait analyses of a group of healthy control subjects could form a family of the vectors, and the average of these vectors is the mean vector describing the normal gait. In case of pathological gait the Euclidean distance between the patient's gait vector and the this normal vector reflects the amount of pathology. Based on this distance the Gait Deviation Index (GDI) is further calculated. The GDI higher or equal to 100 means that there is no gait pathology, each 10 point distance below 100 means the distance between the patient's gait and healthy group equal to one Standard Deviation distance.

Index of normality

This index reflects the normality (or pathology) of the gait pattern as one number [Chester et al., 2007]. The calculation of this index is based on seven other indexes. All these indexes have clinical interpretation. In the first step seven one-dimensional indices of gait are calculated, three kinematic and four kinetic. In the second step the final index is calculated.
The component indexes are as follows:

- sagittal index – quantifies the movement of the hip, knee and ankle in sagittal plane, the curves are replaced by the first sic Fourier components, and the principal frequency of the knee and ankle contribute to the calculation of the index;
- trunk index – this index is based on the relative movement of the trunk in three planes (lateral flexion, tilt, and rotation). Each curve was replaced by the mean and two first derivatives, forming a 9 element vector. The Mahalanobis distance describes its distance from the normalcy
- kinematic normal index – this index is created in a similar fashion to the Trunk Index and incorporates segments and planes that are not part of the original Sagittal Index of gait: hip adduction/abduction, knee varus/valgus, foot rotation, pelvic tilt, pelvic rotation, and pelvic obliquity. The angles are approximated as linear combinations of the normative mean angle

pattern, angular velocity pattern, and angular acceleration pattern. As a result a 18 element vector is formed, and further an index is calculated (using previous methodology).

- four kinetic indices – there are two moment and two power indices, as gait cycle is separated into stance and swing phases. Joint moment indices incorporate frontal and sagittal hip and knee moments with sagittal ankle moment, for a total of five joint moment curves. Joint power indices incorporate sagittal hip, knee, and ankle powers with frontal hip power, for a total of four joint power curves.

Following standard data-transformation protocols, square roots of the seven individual gait indices were used to calculate the overall gait index score.

CONCLUSIONS AND REMARKS

The mentioned data analysis methodologies are quite fundamental and prominent in the gait index prediction and classification research. While PCA can only unveil linear relationships, fuzzy clustering can reveal non-linear structures and grouping tendencies, but cannot handle time-varying data. On the other hand, neural networks can be robust to large variability in data and predict aproximate parameter values.

All these basic methods may be utilized together with more advanced ones e.g. factor analysis fractal dynamics, wavelet transform and many others in order to receive better quality gait indexes.

LITERATURE

Biswas A., Lemaire E.D., Kofman J. (2008) Dynamic gait stability index based on plantar pressures and fuzzy logic. Journal of Biomechanics 41: 1574 – 1581

Chau T. (2001) A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractals methods. Gait & Posture; 13: 49 - 66

Chau T. (2001) A review of analytical techniques for gait data. Part 2: neural network and wavelet methods. Gait & Posture; 13: 102 – 120

Chester V.L., Tingley M., Biden E.N. (2007) An extended index to quantify normality of gait in children. Gait & Posture; 25: 549 – 554

Muniz A.M.S., Nadal J. (2009) Application of principal component analysis in vertical ground reaction force to discriminate normal and abnormal gait. Gait &Posture; 29: 31 - 35

Sadeghi H., Prince F., Sadeghi S., Labelle H. (2000) Principal component analysis of the power developed in the flexion / extension muscles of the hip in able-bodied gait. Med Eng & Physics; 22: 703 – 710

Schwartz M., Rozumalski A. (2008) The gait deviation index: A new comprehensive index of gait pathology. Gait & Posture; 28: 351 – 357

Shuttle L.M., Narayanan U., Stout J.L., Selber P., Gage J.R., Schwartz M.H.. An index for quantifying deviations from normal gait. Gait & Posture; 11: 25 - 31

Wu J., Wang J., Liu L. (2007) Feature extraction via KPCA for classification of gait patterns. Hum Mov Sci; 26: 393 – 411