Medical Engineering & Physics 32 (2010) 1145–1160

Contents lists available at ScienceDirect

Medical Engineering & Physics

journal homepage: www.elsevier.com/locate/medengphy

A fuzzy decision tree-based SVM classifier for assessing osteoarthritis severity using ground reaction force measurements

S.P. Moustakidis [a], J.B. Theocharis [a,*], G. Giakas [b,c,1]

[a] Aristotle University of Thessaloniki, Department of Electrical and Computer Engineering, Division of Electronics and Computer. Eng., 54124 Thessaloniki, Greece
[b] University of Thessaly, Department of Physical Education and Sports Science, Greece
[c] Centre of Research and Technology – Thessaly (CERETETH), Institute of Human Performance and Rehabilitation (InHuPeR), 42100 Trikala, Greece

ARTICLE INFO

Article history:
Received 11 January 2010
Received in revised form 4 August 2010
Accepted 12 August 2010

Keywords:
Osteoarthritis detection
Decision trees
GRF signals
Feature selection
Class grouping
Support vector machines
Wavelet packet

ABSTRACT

A novel fuzzy decision tree-based SVM (FDT-SVM) classifier is proposed in this paper, to distinguish between asymptotic (AS) and osteoarthritis (OA) knee gait patterns and to investigate OA severity using 3-D ground reaction force (GRF) measurements. FDT-SVM incorporates effective techniques for feature selection (FS) and class grouping (CG) at each non-leaf nodes of the tree structure, which reduce the overall complexity of DT building and alleviate the overfitting effect. The embedded FS and CG are based on the notion of fuzzy partition vector (FPV) that comprises the fuzzy membership degrees of every pattern in their target classes, serving as a local evaluation metric with respect to patterns. FS is driven by a fuzzy complementary criterion (FuzCoC) which assures that features are iteratively introduced, providing the maximum additional contribution in regard to the information content given by the previously selected features. A novel Wavelet Packet (WP) decomposition based on the FuzCoC principles is also introduced, to distinguish informative and complementary features from GRF data. The quality of our method is validated in terms of statistical metrics drawn by confusion matrices, such as sensitivity, specificity and total classification accuracy. In addition, we investigate the impact of each GRF component. Finally, comparative results with existing techniques are given, demonstrating the efficacy of the suggested approach.

1. Introduction

Walking is the body's natural means of moving that is strongly dependent on a series of interactions between two multi-segmented limbs and the total body mass. Humans strive to retain the walking capability even in the presence of severe impairment. The resulting gait pattern is a mixture of normal motion and compensatory reactions of adjacent muscles. Identification of such walking malfunctions requires subjective evaluation of an experienced clinician whereas the assessment of more complex cases necessitates laboratory measurements. The exponential growth of computer technology in recent years and the updated instrumentation has facilitated patient's examination, adding greater precision and providing information from multiple correlated factors. Meeting these demands, this study focuses on developing a fully automatic and objective computer method for the discrimination between normal and osteoarthritic gaits and the assessment of the osteoarthritis (OA) severity grade using ground reaction force measurements.

Biomedical problems, such as OA, that comprise high-dimensional, noisy and very complex data, may be an ideal application area for computational intelligent (CI) algorithms. Unlike statistical analysis, these techniques are capable of capturing non-linear interactions between multiple variables, forming more powerful analytical tools. Given the good performance of CI methods, it is believed that this type of analysis is worth pursuing with potentials to be applied in clinical applications. Their clinical use could take a form of an evaluation, ideally before and after some intervention, to assess the degree of improvement in patient's gait function. Along this direction, different approaches have been developed in the past, solving complex classification problems such as neural networks and statistical discriminant analysis.

Decision trees (DT) [1] have recently attracted considerable interest, since they are powerful learning techniques capable of organizing the information extracted for a training dataset in a hierarchical structure. A DT uses a recursive partitioning technique where the initial data set is finally categorized into pattern groups with similar characteristics. The evolution of DTs starts with the well-known *ID3* (Interactive Dichotomiser 3) suggested by Quinlan [2]. In the original *ID3*, the entropy measure is used to establish

* Corresponding author. Tel.: +30 2310 996343; fax: +30 2310 996447.
E-mail addresses: theochar@eng.auth.gr (J.B. Theocharis), ggiakas@pe.uth.gr (G. Giakas).
1 Tel.: +30 24310 47010; fax: +30 24310 47042.

1350-4533/$ – see front matter © 2010 IPEM. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.medengphy.2010.08.006

the information gain of each attribute. The feature that minimizes the impurity of the generated subsets is retained for data partitioning and the process is repeated for each one of the obtained subsets. C4.5 is an extension and revision of the ID3 algorithm which utilizes an improved information gain-ratio measure to segment attributes, incorporates a pruning process to avoid overfitting, while at the same time is capable of handling missing value attributes.

The main shortcoming of crisp DT methods (such as ID3, C4.5) is that quantitative data is partitioned into crisp sets by applying recursively sharp decision boundaries. This implies that small changes in the attribute values of unseen patterns may result in misclassifications. To address the problem associated with crisp decisions, various researchers introduced fuzzy decision trees (FCT) induction algorithms [3]. FCTs handle cognitive uncertainties by allowing gradual transitions between attribute values. Several methodologies [4–7] have followed the approach where the attributes are first prefuzzified and then used as input to an ID3-based decision tree. However, these methods are time consuming and require a human expert to determine the appropriate set of membership functions. FCT algorithms have also been addressed [8–10], incorporating cognitive uncertainties within the tree construction process. In these studies, the fuzzy entropy [11] and the non-specificity measure are used to assess the vagueness and ambiguity uncertainties. An alternative approach has also been proposed [12–15], whereby a fuzzy inference framework is applied on a pre-generated crisp decision tree. A tuning stage is next employed to adjust the membership functions and the inference parameter values, leading to a highly optimized rule base with good generalization ability.

Extending the DT methodology, a new family of hierarchical classifiers was recently proposed, incorporating Support Vector Machines (SVMs), mainly due to their high generalization capabilities. Typically, a multi-class SVM classifier reduces the complex problem to multiple binary classification problems that can be solved separately. One-against-all (1AA) [16], all-against-all (AAA) [17], direct acyclic graph SVM (DAGSVM) [18] and error correcting output codes (ECOC) [19] are among the most common decomposition strategies proposed in the literature. Their main criticism is that an a priori decomposition of the problem is performed disregarding the properties of the particular data under consideration. To this end, alternative decomposition strategies have also been examined, by disposing SVM classifiers in a hierarchical architecture. These DT-based structures are composed of nodes and ramifications where the internal nodes correspond to SVM classifiers and the leaf nodes represent the problem classes. Since structure determination strongly depends on the specific partitioning technique applied to each node, various decomposition criteria are investigated, such as class distance metrics [20,21], balanced subsets [22] and confusion matrixes [23].

A novel DT-based multi-class SVM is proposed in this paper that combines the high generalization ability of SVMs with the DTs' advantage of organizing the extracted knowledge in human-interpretable structures. Our method integrates a class grouping (CG) technique in each internal node with the goal to accomplish a coarse discrimination between the more separable classes at the upper nodes of the tree, whereas the overlapping classes are examined at the lower nodes. Furthermore, at each internal SVM node, the suggested partition technique incorporates an efficient Feature Selection (FS) process that is capable of identifying compact subsets composed of strongly relevant and non-redundant (complementary) features. The distinguishing characteristics of the proposed method against other CI approaches of the literature are described as follows: (a) The resulting DT-tree structure is highly interpretable, comprising a set of comprehensible sub-problems implemented on features valuable for clinical validation of the OA problem. (b) Traditional classifiers solve the OA problem globally, disregarding the data distribution and the degree of confusion between the classes. Contrarily, in our classifier suitable class groups are hierarchically formed according to their mutual confusion. (c) The incorporated FS and CG processes simplify the overall complexity of the classifier, increase its generalization capabilities and provide clinical evidence related to the cause of the pathology considered. (d) Unlike the existing DT methods that utilize global partitioning criteria, the aforementioned CG and FS techniques rely on a local evaluation measure with respect to patterns realized by the so-called fuzzy partition vector (FPV). Associated with each feature, FPV is constructed by invoking a fuzzy class allocation scheme that assigns membership grades to every class. Hence, FPV shows the membership degrees distribution along the patterns, allowing an efficient manipulation of features and feature spaces. (e) Owing to the embedded FS, our approach is able to cope with high-dimensional feature spaces. To this end, instead of confining to heuristic features from GRF records, we perform wavelet packet decomposition of GRF signals with the goal to identify more discriminating characteristics leading to better OA classification results. The superiority of our approach is verified against other CI approaches of the literature, in terms of total classification accuracy and statistical parameters drawn by confusion matrices.

The rest of the paper is organized as follows. The OA problem is introduced in Section 2, along with a literature review and the GRF data acquisition procedure. In Section 3 the DT's objectives are presented whereas Section 4 introduces the notion of FPV, including useful definitions in the fuzzy domain. Sections 5 and 6 deal with the proposed WP determination and FS, respectively, while in Section 7 the class grouping methodology is analytically described. In Section 8 we elaborate on the suggested DT algorithm and provide an illustrating example on the OA dataset. Performance and comparative results are presented in Section 9, whereas conclusions are drawn in Section 10.

2. Problem statement: assessing osteoarthritis

Osteoarthritis (OA) is a degenerative joint disease caused by the breakdown and eventual loss of the cartilage of one or more joints. Clinical manifestations of OA may include joint pain, stiffness, creaking and locking of joints causing differentiations in human's gait. Within the context of automatic and objective OA diagnosis and assessment, several studies have addressed gait pattern classification. An artificial neural network, proposed in [24], utilizes a data set comprising ground reaction forces, cadence and walking velocity to discriminate healthy subjects from subjects with ankle, knee or hip arthrosis. GRF parameters along with their chronological incidence of occurrence were also used in [25], supplying a multilayer perceptron trained on a group of 20 subjects (10 solid arthrodesis and 10 healthy). Beynon et al. [26] introduced a method based on the Dempster–Shafer theory of evidence that is able to produce an objective analysis of biomedical data collected in a clinical motion-analysis setting. Five characteristics, including the vertical ground reaction force, were chosen to classify subjects into either a normal or osteoarthritic (OA) knee-function group.

In a recent study [27], the authors investigated different features such as the coefficients of a polynomial expansion and coefficients of a wavelet decomposition extracted from GRF measurements. A nearest neighbor classifier was employed to distinguish between asymptotic (AS) and osteoarthritis (OA) knee gait patterns. The best discrimination rate achieved was 91% from a group of 42 participants (16 AS and 26 OA). The use of SVMs has been recently explored for automated detection of various gait abnormalities, assessment of the overall gait function after a certain treatment or intervention and classification of different clinical populations. In [28], the authors investigated the application of SVM to classify