Computers in Biology and Medicine 54 (2014) 32–36

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Time series for blind biosignal classification model ☆

Derek F. Wong, Lidia S. Chao *, Xiaodong Zeng, Mang-I Vai, Heng-Leong Lam

Department of Computer and Information Science, University of Macau, Av. Padre Tomás Pereira Taipa, Macau S.A.R., China

ARTICLE INFO

Article history:
Received 21 November 2013
Accepted 10 August 2014

Keywords:
Blind biosignal classification
Dynamic time warping (DTW)
Time series clustering
Machine learning
Bioinformatics

ABSTRACT

Biosignals such as electrocardiograms (ECG), electroencephalograms (EEG), and electromyograms (EMG), are important noninvasive measurements useful for making diagnostic decisions. Recently, considerable research has been conducted in order to potentially automate signal classification for assisting in disease diagnosis. However, the biosignal type (ECG, EEG, EMG or other) needs to be known prior to the classification process. If the given biosignal is of an unknown type, none of the existing methodologies can be utilized. In this paper, a blind biosignal classification model (*B²SC Model*) is proposed in order to identify the source biosignal type automatically, and thus ultimately benefit the diagnostic decision. The approach employs time series algorithms for constructing the model. It uses a dynamic time warping (DTW) algorithm with clustering to discover the similarity between two biosignals, and consequently classifies disease without prior knowledge of the source signal type. The empirical experiments presented in this paper demonstrate the effectiveness of the method as well as the scalability of the approach.

1. Introduction

The term 'biosignal' refers to all kinds of signals that can be measured and monitored from biological beings, including electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG), and electrooculography (EOG). The signals are the recordings of the electrical activity of the heart, brain, muscles, and eyes, respectively, which can be used in making diagnostic decisions. Research into ECG classification has proliferated since the discovery of the automatic recognition of electrocardiogram waves by Stallman and Pipberger [1]. After this discovery, a rule-based rough-set decision system was developed to generate an inference engine for ECG classification from different standard time-plane features [2]. Then, ECG classification was optimized by applying a feature selection method with a new criterion function index to improve performance [3]. ECG signal classification using parallel genetic algorithms and neural networks [4], as well as using block-based neural networks, was proposed [5]. Although EEG classification is more difficult due to the high dimensional feature space, a certain amount of research is underway in order to benefit the biomedical community. For example, Subasi [6] developed a Mixture of Expert (ME) network structure to improve the accuracy of epileptic seizure detection in EEG, and the overall predictive performance was superior to any of the individual experts. An automatic recognition method for Alzheimer's disease (AD) with single-channel EEG recording using combined genetic algorithms (GA) and artificial neural networks (ANN) was proposed by Cho et al. [7]. Similar signals are obtained in EMG and EOG classification, however, the complex nature of the signals often means that their analysis and classification is difficult [8]. Lucas et al. employed a support vector machine (SVM) approach to classify multi-channel surface EMG signals. The experimental results showed that their method is suitable for real-time applications [9]. Then, the authors established an EMG signal classification system to discriminate finger motions by using linear neural networks [10]. It showed promising performance for classifying motions based on biosignal patterns. Recently, Alkan and Günay proposed a surface EMG signal classification system, which used five discriminant functions and an SVM classifier [11]. Their system is used to classify EMG signals in prosthetic arm control. Furthermore, EOG is an efficient measurement technique to detect eye movement for human activity recognition as shown by the use of EOG signals for the realization of a Human Computer Interface (HCI) device [12], which is able to recognize a patient's eye movements and thus restore some communication abilities. Güven and Kara [13] used Artificial Neural Networks to analyze EOG signals, which can be used to diagnose subnormal vision. Then Bulling et al. proposed a method for analyzing repetitive eye movement patterns [14]. They used EOG signals to detect three basic eye movement types (saccades, fixations, and blinks) and proved that the recognition methodology could successfully identify five office activities.

☆This is a special focus paper published in connection with the 2nd International Conference on Biomedical Engineering and Biotechnology (iCBEB 2013), held in Wuhan, China, October 11-13, 2013.

* Corresponding author.
E-mail address: lidiasc@umac.mo (L.S. Chao).

http://dx.doi.org/10.1016/j.compbiomed.2014.08.007
0010-4825/

However, all of the aforementioned studies concentrated on a single type of biosignal, where the source of the biosignal was known (ECG, EEG, EMG, or EOG) to the classifier, and the classification was conducted based on the nature of the specific type of biosignal. However, to our knowledge, there is not any literature about automatically identifying biosignals, which would be useful when the type of biosignal is unknown. To bridge this gap, this study proposes an automatic approach for identifying an unknown biosignal and typing it as ECG, EEG etc., in addition to conventional biosignal classification. The approach incorporates a DTW (dynamic time warping) algorithm of time series classification, which is capable of identifying a specific disease or symptom regardless of the type of source biosignal. The organization of this paper is as follows: in Section 2, the proposed approach and its three major phases are presented in detail. After that, the empirical results are discussed in Section 3, followed by the conclusion and future work sections.

2. Blind biosignal classification model

An emerging trend is that not only do professionals use medical technologies and tools, but also patients and others [15]. The rapid development of innovative information and communication technologies has led to home-based healthcare, which is feasible and preferable, particularly for the elderly and patients with chronic diseases who self-manage their health at home rather than in a hospital. However, home-based users of medical technology are usually non-professional and not well trained, so they may incorrectly attach ECG sensors to their bodies for example. Furthermore, multiple biosignals may be acquired simultaneously and mixed or combined into one, which may result in unknown or ambiguous source biosignals with only their temporal waveforms or time-frequency patterns known, but not their types [16]. Inspired by these facts, a blind biosignal classification model (*B²SC Model*) is proposed, which serves as a partial implementation of an integrated multi-function biosignal measurement device for such home-based healthcare paradigms. *B²SC Model* can automatically identify an unknown (blind) mixed biosignal for further detailed analysis and diagnosis, and its purpose is multifold: (1) enabling non-skilled home-based users to operate a multi-function biosignal acquisition device easily for monitoring and managing their health conditions; (2) helping to improve the diagnostic capability of a disease or symptom without knowing the exact type of source biosignal; and (3) allowing the use of only one device and one sensor which could adapt to multiple biosignals. *B²SC Model* is composed of three major phases: biosignal template construction, template optimization and management, and pattern matching. The details of each phase will be described below.

2.1. Biosignal template construction

As previously discussed, the quality and quantity of a training dataset in a classification problem severely affects the efficiency and performance of a classification algorithm. The biosignal template construction phase aims at constructing a library of biosignal templates using all available raw data from biosignals, and such templates will be ultimately used for guidance in accurately classifying an unknown biosignal. The raw biosignals usually have long signal lengths and the datasets often have large file sizes, meaning that it is impossible for the raw data to be used directly in training the classification model. Moreover, a single type of biosignal may have diverse patterns. Fig. 1 shows an example of this: all four patterns are ECG signals. The first pattern indicates a normal ECG signal, and the other three reveal the different stages of ischemic heart disease. Therefore, the construction of a variety of biosignal templates as training datasets is critical for the subsequent task of classification.

The benchmark datasets, downloaded from the MIT-BIH Arrhythmia Database [17], the CHB-MIT Scalp EEG Database [18] and the EMG Datasets Repository [19] were used to produce the ECG, EEG, and EMG templates, respectively. The MIT-BIH Arrhythmia Database provides standard test materials for the evaluation of arrhythmia detectors as well as for basic research in cardiac dynamics. The CHB-MIT Scalp EEG Database collects the EEG recordings from pediatric patients who suffered from intractable epilepsy. Furthermore, the CAP (Cyclic Alternating Pattern) Sleep Database records EMG and EOG signals as well as EEG signals. It is intended to provide a useful number of carefully annotated examples of CAP in a representative variety of pathophysiologic contexts, which is vital for the study of CAP dynamics, and in particular for the development and evaluation of CAP analysis. In this paper, we use ECG as an example, to describe the entire process of the *B²SC Model*. EEG or EMG datasets are used for construction and training in exactly the same way as ECG signals.

The essential tasks of template construction include biosignal segmentation and template formation. Segmentation partitions each ECG signal into small pieces so that each piece contains only one cycle of an ECG recording, representing one candidate template. As a result, there are in total 240,000 one-cycle ECG signals generated from the entire MIT-BIH Arrhythmia Database. All of these one-cycle ECG signals are used as the source for generating the library of ECG templates. Nonetheless, the number of candidate biosignal templates determines the efficiency and performance of the subsequent classification tasks: too few templates for each category may lead to underfitting, which misclassifies some important biosignal patterns; but too many templates may cause redundancy and wastage of resources. To this end, the existing ECG categories defined by Jenkins and Gerred [20] are used as guidelines to determine the groups of templates, which describe nine major common categories for ECG signals. We have ignored the minor subcategories in this study. Similarly, the general categories specified by Niedermeyer and da Silva [21] and Khushaba et al. [19] are used to determine the classes for EEG and EMG in this study. Therefore, in the template formation stage, all candidate ECG templates are roughly divided into nine groups. The widely used clustering methodology, *k*-nearest neighbors (*k*NN) algorithm [22], was employed for classifying a template by the major label amongst its *k*-nearest neighbors in the ECG database.

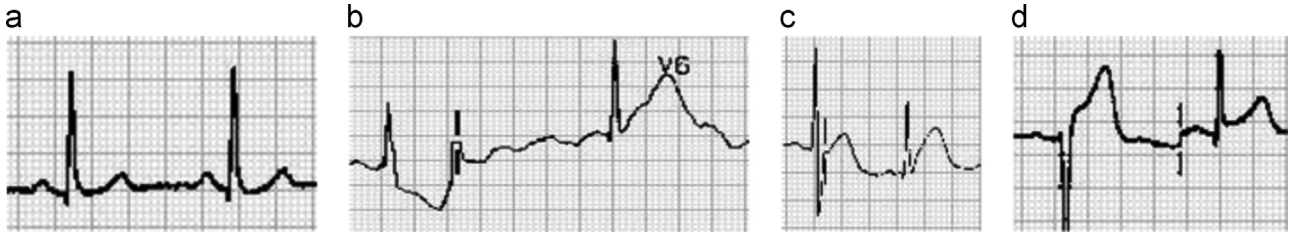

Fig. 1. Samples of different patterns from ECG signals: (a) normal ECG; (b) acute inferior myocardial infarction; (c) acute posterior myocardial infarction, and (d) acute anterior myocardial infarction.