A Multinomial Logistic Regression Approach for Arrhythmia Detection

Omar Behadada, University of Tlemcen, Tlemcen, Algeria
Marcello Trovati, Edge Hill Universitiy, Ormskirk, United Kingdom
Georgios Kontonatsios, Edge Hill Universitiy, Ormskirk, United Kingdom
Yannis Korkontzelos, Edge Hill Universitiy, Ormskirk, United Kingdom

ABSTRACT

Cardiovascular diseases are the leading causes on mortality in the world. Consequently, tools and methods providing useful and applicable insights into their assessment play a crucial role in the prediction and managements of specific heart conditions. In this article, we introduce a method based on multi-class Logistic Regression as a classifier to provide a powerful and accurate insight into cardiac arrhythmia, which is one of the predictors of serious vascular diseases. As suggested by our evaluation, this provides a robust, scalable, and accurate system, which can successfully tackle the challenges posed by the utilisation of big data in the medical sector.

KEYWORDS

Big Data, Knowledge Extraction, Multinomial Logistic Regression, Text Mining

INTRODUCTION

In the last report of World Health Report 2013(Yu & Chou, 2008), cardiovascular diseases are considered as one of most worrying health issues and the largest cause of mortality in the world. In particular, the detection of cardiac arrhythmia is a very promising area, because premature ventricular contraction (PVC) has been shown to be an effective predictor of sudden death (Behadada, Trovati, Chikh, Bessis, & Korkontzelos, 2016a). Furthermore, the continuous creation and availability of data provide new challenges as well as real opportunities in this field.

This article is based on a previous study (Behadada et al., 2016a) introducing an initial implementation of a Multinomial Logistic Regression (MLR) as a classifier of cardiac arrhythmia. The objective of MLR is to learn the posterior probability distributions of each class, in order to create a robust and accurate knowledge-based system, which provides a crucial insight into arrhythmia detections from a variety of data information sources (Behadada, Trovati, Chikh, & Bessis, 2016). MLR has been shown to have a variety of advantages compared to other techniques, which include no need for a linear relationship between the dependent and independent variables, and despite being similar to linear regression, it can be more easily interpreted (Tabachnick, Fidell, & Osterlind, 2001).

However, one of the limitations of (Behadada et al., 2016a) is that the feature selection method, i.e., the process which identifies the most informative and discriminative features of arrhythmia classification, was carried out manually based on discussions with medical experts. To demonstrate...
that this choice was indeed appropriate, we have utilised and expanded the method in (Behadada et al., 2016a) to determine and assess how influential such features are with respect to cardiac arrhythmia. In other words, this will allow determining whether any of the concepts related to the feature selection process is indeed relevant to it.

The dataset analysed in the article is part of ongoing research, aiming to integrate a variety of information from potentially huge unstructured sources, including Doppler and Mir images (Behadada et al., 2016).

RELATED WORK

Existing approaches for the automatic detection of cardiac arrhythmia employ different machine learning methods including Support Vector Machines (Asl, Setareh dan, & Mohbibi, 2008), Decision Trees (Exarchos et al., 2007), Artificial Neural Networks (Acharya, Bhat, Iyengar, Rao, & Dua, 2003; Kara & Okandan, 2007), fuzzy classifiers (M. Tsipouras, Goletsis, & Fotiadis, 2004) and ensemble machine learning models (M. G. Tsipouras & Fotiadis, 2004) to name a few.

Acharya et al. (2003), investigated the use of Artificial Neural Networks (ANN) trained via back propagation and a fuzzy classifier to automatically classify heart rate signals into 8 different classes (i.e., normal, pre-ventricular contraction, complete heart block, sick sinus syndrome, left bundle branch block, ischaemic/dilated cardiomyopathy, atrial fibrillation, and ventricular fibrillation). For evaluation purposes, they employed the MIT-BIH arrhythmia database and sampled 1,000 cases for each classification category. The results that they obtained showed that the proposed machine learning models achieve a classification accuracy of approximately 80%-85% with the fuzzy classifier slightly outperforming the ANN.

Exarchos et al. (2007) developed a hybrid machine learning model which integrates decision trees with a fuzzy classifier. The hybrid model firstly uses decision trees to automatically extract association rules between the input features. The association rules are subsequently used as input to the fuzzy classifier. The hybrid classification model was evaluated on two closely related tasks, namely ischaemic and arrhythmic beat detection. Experimental results indicated that the hybrid classification model achieves a very high classification accuracy of 92% and 96% for the ischaemic and arrhythmic beat detection, respectively.

Asl et al. (2008) proposed a machine learning classification system to automatically identify six frequently occurring types of cardiac arrhythmia. The authors argued that the input heart rate signals, used as features in the classification system, are non-linearly separable (i.e., there is no linear combination of the features that discriminates samples of different classes). To alleviate non-linearity of the input features, they used generalised discriminant analysis (GDA), a dimensionality reduction technique that embeds the initial features into a lower dimensional but linearly separable feature space. As a classification model, they explored the use of a Support Vector Machines (SVM) with a Radial Basis Function (RBF). The RBF kernel function is known to better address non-linearly separable classification problems. Experimental evidence determined that the proposed classification model, which combines a non-linear dimensionality technique with SVM, substantially outperforms previously employed ANN models.

Tsipouras et al. (2004) presented an ensemble of neural networks for automatic arrhythmia detection. In their approach, they used multiple neural networks trained on different combinations of heart rate features. A set of manually constructed rules was then used to combine the output of the neural networks. Experiments conducted showed that the ensemble neural network obtained a robust performance of 87% sensitivity and 89% specificity. Moreover, the authors demonstrated that their proposed model is able to identify a wide range of arrhythmia types, albeit the performance decreased for some specific types (e.g., left bundle branch block and right bundle branch block beats). One limitation of the ensemble neural network model is that the classification performance depends upon fine tuning a large set of hyperparameters (e.g., number of hidden layers, dimensionality of hidden
On Construction of Cluster and Grid Computing Platforms for Parallel Bioinformatics Applications
www.igi-global.com/article/construction-cluster-grid-computing-platforms/52587?camid=4v1a

Service-Oriented Development of Workflow-Based Semantic Reasoning Applications
www.igi-global.com/article/service-oriented-development-of-workflow-based-semantic-reasoning-applications/104763?camid=4v1a