Measurement 104 (2017) 180–191

Contents lists available at ScienceDirect

Measurement

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

Probability distribution pattern analysis and its application in the Acute Hypotensive Episodes prediction

Dazhi Jiang [a], Chenfeng Peng [b], Yifei Chen [a], Zhun Fan [c,d], Akhil Garg [e,*]

[a] Department of Computer Science, Shantou University, 515063, China
[b] Interactive Entertainment Group, Tencent Technology (Shenzhen) Company Ltd., 518057, China
[c] Department of Electronic Engineering, Shantou University, 515063, China
[d] Guangdong Provincial Key Laboratory of Digital Signal and Image Processing, Shantou University, 515063, China
[e] Department of Mechatronic Engineering, Shantou University, 515063, China

ARTICLE INFO

Article history:
Received 19 January 2017
Received in revised form 28 February 2017
Accepted 16 March 2017
Available online 18 March 2017

Keywords:
Acute Hypotensive Episodes
Probability Distribution Pattern Analysis
Time series
Prediction

ABSTRACT

At present many hospitals have to deal with the patient's care and nursing for Acute Hypotensive Episodes (AHE) occurring in Intensive Care Units. AHE can cause fainting or shock suddenly, leading to irreversibility organ damage, and even death. Therefore, forecasting of occurrence of AHE is of practical value. However, the prediction of clinical AHE largely depends on the doctors' experience, which cannot guarantee the high rate of success. It is thus very meaningful for the clinical care to use appropriate methods to predict the AHE with an automatic and reliable method. In this study, a Probability Distribution Patterns Analysis (PDPA) method is presented to solve the time series prediction problem of AHE. In the first phase, the features are extracted from the PDPA in the global and integral time series, and the partial local time series in the fixed time window. In the second phase, the proposed algorithm combining Genetic Algorithm (GA) and Support Vector Machine (SVM), namely GA-SVM is adopted to select the vital features for the effective classification. In order to demonstrate the generality of our method, we also conduct experiments on a classical time series problem-Control Chart Patterns (CCPs) multi-class time series, which is a benchmark problem in the process control. For CCPs problem, the experimental results demonstrate that the proposed method outperforms several traditional methods. The obtained accuracy is 98.65%, which is superior to listed previous works using the same CCPs model. For AHE classification and forecasting, the methodology is applied in two data sets, a small data set (37 records) and a big one (2892 records). The test accuracy of 89.19%, sensitivity of 91.67%, specificity of 88% in the small data set, and a test accuracy of 80.76%, sensitivity of 78.19%, specificity of 81.51% in the big data set are achieved with the classification model.

1. Introduction

Depend on the urgency of the symptom, the hypotensive episode can be divided into two classes, Chronic Hypotensive Episode (CHE) and Acute Hypotensive Episode (AHE). CHE is defined as Systolic Arterial Blood Pressure (SABP) lower than 90 mmHg, and Diastolic Arterial Blood Pressure (DABP) lower than 60 mmHg. On the contrary, the AHE is defined as any period of 30 min or more during which at least 90% of the mean arterial pressure (MAP) measurements are at or below 60 mmHg. For the normal human blood pressure, the SABP is between 90 and 140 mmHg, and DABP is between 60 and 90 mmHg, and the MAP is between 70 and 105 mmHg [1].

The AHE is a series and common postoperative complication in the Intensive Care Units (ICU), which can damage the patient's organs and eventually, cause the patients to decease (If not promptly and proper treated in time). Thus, the early detection of AHE by computer becomes vitally important for clinical therapy and intervention in AHE. As provided in PhysioNet [2], the episodes of AHE can be predicted with reasonably high specificity (in 2009, the PhysioNet held a research competition of AHE prediction, which let the AHE automated prediction become a worldwide research topic). In this challenge, the data used to AHE prediction is blood pressure, which is a typical complex and non-linear time series data. Compared with other problems, the AHE prediction based on the time series is more complicated, because the human

* Corresponding author.
E-mail address: akhil@stu.edu.cn (A. Garg).

http://dx.doi.org/10.1016/j.measurement.2017.03.030
0263-2241/

body produces the blood pressure data quickly, heavily and dynamically.

Many studies have been done for building automatic prediction AHE system based on the machine learning and data mining methods. In these methods, neural network was frequently used to create the classifier for AHE prediction. For example, Henriques et al. developed a multi-models neural network for the forecast of the AHE [3]. In 2011, Teresa Rocha et al. [4] used the similar method to solve the AHE problem in a big data sets (the total number of samples is 2344, the proposed method obtained sensitivity 82.8%, and specificity 78.4%). Furthermore, Zhou et al. used a Chebyshev neural network to solve the challenge dataset (110 samples), and the result shows that the proposed method performs better than other solutions in AHE prediction [5]. In the thesis of M Ghassemi, the multi-variate neural network is the best method for AHE prediction which is tested in 1168 samples data sets. For each sample, the vital values are MAP, respiration rate, oxygenation level and heart rate. The yielded accuracy is 84% [6].

For other methods, Hoseinnia et al. presented a hybrid approach for predicting AHE, which is based on the wavelet transform and neural network. The wavelet transform is used to decompose the MAP, and the neural network is used to forecast the wavelet approximation coefficients [7]. Arasteh et al. presented to use Empirical Mode Decomposition (EMD) method to solve the AHE prediction. EMD is used to decompose the MAP time series into several Intrinsic Mode Functions (IMFs), and then some statistical features are extracted from the IMFs. Finally, the SVM is applied to the classification [8]. Lehman et al. [9] use a combination of Gaussian Mixture Model based clustering and K-Nearest Neighbours Classier on 227 patient records from the MIMIC II database using both Heart Rate and MAP measurements. They achieve an accuracy of 70% with 74% sensitivity and 60% specicity. Lee et al. [10], who use 1311 patient records using heart rate, MAP and clinical data from the MIMIC II database. Their algorithm uses 102 statistical, wavelet-based and clinical features and predicts the event 1 h in advance with 86% accuracy.

In our research group, Sun et al. predict the AHE with the Particle Swarm Optimizer (PSO) and K-means method. The method based on PSO and K-means is to extract the vital features of MAP data, then the SVM method is used to create the classifier model. The experiment is verified on the 2863 samples, and the best accuracy is 81.2% [11]. Jiang et al. used the EMD method to calculate patient's MAP data. Then, for features, the bandwidth of amplitude modulation, the frequency modulation and the power of IMFs were extracted. The total features are 5. A Multiple Genetic Programming (Multi-GP) is presented for the classification of the AHE [12,13]. Furthermore, in 2016, Jiang et al. transform the IMFs data into probabilistic distribution, then the statistic features, such as Peak, Mode, Skewness, Kurtosis and Shannon Entropy are extracted in different IMFs probabilistic distribution. Multi-GP was still used to create the classifier models, and the best achieved training and testing accuracy are 82.92% and 79.93% respectively [14].

According to the previous research, the features selection and classification method are two critical factors for improving the prediction performance of AHE. As a natural extension of previous work [14], this paper presents a generic methodology for time series prediction, which extracts features base on statistics and Probability Distribution Patterns Analysis (PDPA). Furthermore, in order to describe the dynamic distribution patterns in time series, the features are not only extracted from the global and integral time series, but also from the local and partial time series in the fixed time window. After features extraction, a method, which is combined with Genetic Algorithm and Support Vector Machine (GA-SVM), is presented to create classifier and select features simultaneously. All the features were extracted based on statistic so the process is very forthright and fast. This characteristic is especially important when the method is adopted in a real-time monitoring system of medical device. In addition, this method has a general applicability on the time series prediction and classification problem. In the experimental verification, for CCPs problem [15–17], the obtained accuracy is 98.65%, which is superior to listed previous works used the same CCPs model. For AHE classification and forecasting, the methodology is applied in the two data sets, the test accuracy of 89.19% in the small data set and 80.76% in the big data set are achieved from the classification model.

The rest of the paper is organized as follows: in Section 2, the experimental data sets and the prediction of AHE problem are briefly introduced. Section 3 describes the methodology for feature extraction and selection. Section 4 is the experiments verification, includes typical multi-class time series problem, small data sets of AHE and big data sets of AHE. Section 5 are conclusions and future research topic.

2. Data sets

As previous work, the data for the AHE experiment was collected from the Multi-Intelligent Monitoring in Intensive Care (MIMIC) II [18]. The data, where ware got from MIMICII database, regards a patient as a unit, and records the patients' vital signs, such as systolic arterial blood pressure (SABP), and diastolic arterial blood pressure (DABP). The SABP (Fig. 1 red[1] box) and DABP (Fig. 1 green circle) are the maximum pressure and the minimum pressure respectively. In this experiment, we focus on the mean arterial pressure (MAP), which is actually a combination of the SABP and DABP, and calculated as follows:

$$MAP = DABP + \frac{SABP - DABP}{3}$$

According to the above calculation method, the ABP data could be transformed into the MAP data as follows (see Fig. 2):

For AHE prediction, the validation set consists of two datasets, which are a small dataset and a big one. The small dataset was obtained from PhysioNet 2009 challenge [19], and the big dataset is downloaded from MIMICII [18]. In both datasets, instant T_0 is a marked stamp for the prediction. For the small dataset, in the training set, the data records contain all the data before and after instant T_0. In the testing set, the records are truncated at T_0 for the purpose of performance testing. Which means in the training set, every record contains the data obtained from 2 h before T_0 and 1 h after T_0. In the testing sets, the data is only collected from 2 h before T_0. In the small dataset, because some data are missing, only 48 records are selected as the training set, and 37 records as the testing set (two classification problems, AHE and NO_AHE problem, AHE means the patient will suffer in during the forecast widow and NO_AHE means no AHE symptom). The big dataset contains 2892 records. 600 records are selected randomly as the training set, which contains 300 AHE records and 300 NO_AHE records. The remaining is the testing set, which has 2292 records.

3. Methodology

Normally, the pattern in the systems and processes can be conveyed in the form of probabilistic distribution functions (PDFs) [20]. It inspired us to extract data PAP from time series to gain an insight into the underlying distribution pattern in the global and integral time series (such as the 2 h data before T_0 for AHE problem) and local distribution pattern in the local and partial time series in the fixed time window.

[1] For interpretation of color in Figs. 1 and 10, the reader is referred to the web version of this article.