Computers in Biology and Medicine 67 (2015) 74–82

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Manifold ranking based scoring system with its application to cardiac arrest prediction: A retrospective study in emergency department patients

Tianchi Liu [a], Zhiping Lin [a], Marcus Eng Hock Ong [b,c], Zhi Xiong Koh [b], Pin Pin Pek [b], Yong Kiang Yeo [a], Beom-Seok Oh [a], Andrew Fu Wah Ho [d], Nan Liu [b,e,*]

[a] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[b] Department of Emergency Medicine, Singapore General Hospital, Singapore
[c] Health Services and Systems Research, Duke-NUS Graduate Medical School, Singapore
[d] SingHealth Emergency Medicine Residency Program, Singapore Health Services, Singapore
[e] Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore

ARTICLE INFO

Article history:
Received 18 May 2015
Accepted 1 October 2015

Keywords:
Cardiac arrest
Machine learning
Scoring system
Manifold ranking
Emergency medicine

ABSTRACT

Background: The recently developed geometric distance scoring system has shown the effectiveness of scoring systems in predicting cardiac arrest within 72 h and the potential to predict other clinical outcomes. However, the geometric distance scoring system predicts scores based on only local structure embedded by the data, thus leaving much room for improvement in terms of prediction accuracy.
Methods: We developed a novel scoring system for predicting cardiac arrest within 72 h. The scoring system was developed based on a semi-supervised learning algorithm, manifold ranking, which explores both the local and global consistency of the data. System evaluation was conducted on emergency department patients' data, including both vital signs and heart rate variability (HRV) parameters. Comparison of the proposed scoring system with previous work was given in terms of sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV).
Results: Out of 1025 patients, 52 (5.1%) met the primary outcome. Experimental results show that the proposed scoring system was able to achieve higher area under the curve (AUC) on both the balanced dataset (0.907 vs. 0.824) and the imbalanced dataset (0.774 vs. 0.734) compared to the geometric distance scoring system.
Conclusions: The proposed scoring system improved the prediction accuracy by utilizing the global consistency of the training data. We foresee the potential of extending this scoring system, as well as manifold ranking algorithm, to other medical decision making problems. Furthermore, we will investigate the parameter selection process and other techniques to improve performance on the imbalanced dataset.

1. Background

Prediction of clinical outcomes of patients in intensive care units (ICUs) is of high importance. For example, identifying patients who are at low risk of developing adverse outcomes without admission to ICU can help ration precious and expensive critical care resources. Recently, prediction of cardiac arrest within 72 h has drawn researchers' attention because it allows clinicians to make timely and safe clinical decisions based on evidence-based prognostication to ensure the correct patients receive the appropriate monitoring and treatment.

Traditionally, outcome prediction is made by intensive care clinicians based on explicit judgment. However in the past decades, several scoring systems have been developed and widely used in ICUs [1], for instance, the Acute Physiology and Chronic Health Evaluation

Abbreviations: APACHE, Acute physiology and chronic health evaluation; AUC, Area under the curve; ECG, Electrocardiogram; FP, False positive; FN, False negative; HRV, Heart rate variability; ICU, Intensive care unit; *k*-NN, *k*-nearest neighbors; LOOCV, Leave-one-out cross-validation; MPM, Mortality probability model; MRSPA, Manifold ranking-based scoring prediction algorithm; NPV, Negative predictive value; PPV, Positive predictive value; ROC, Receiver operating characteristics; SAPS, Simplified acute physiology score; SVM, Support vector machine; TN, True negative; TP, True positive

* Corresponding author at: Department of Emergency Medicine, Singapore General Hospital, Outram Road, Singapore 169608, Singapore. Tel.: +65 6321 3590; fax: +65 6321 4873.
E-mail addresses: liut0012@e.ntu.edu.sg (T. Liu), ezplin@ntu.edu.sg (Z. Lin), marcus.ong.e.h@sgh.com.sg (M.E.H. Ong), koh.zhi.xiong@sgh.com.sg (Z.X. Koh), pek.pin.pin@sgh.com.sg (P.P. Pek), ykyeo@ntu.edu.sg (Y.K. Yeo), bsoh@ntu.edu.sg (B.-S. Oh), sophronesis@gmail.com (A.F.W. Ho), liu.nan@sgh.com.sg (N. Liu).

http://dx.doi.org/10.1016/j.compbiomed.2015.10.001
0010-4825/

(APACHE) [2], Simplified Acute Physiology Score (SAPS) [3], and Mortality Probability Model (MPM) [4]. Each of these scoring systems typically predicts one specific kind of clinical outcome. For example, both APACHE and SAPS predict mortality. These traditional scoring systems lack specificity in the prediction of cardiac arrest. Therefore there is a need to develop a new and efficient scoring system specific for cardiac arrest prediction.

Another challenge for cardiac arrest prediction is the selection of patients' feature variables, which highly impacts the accuracy of prediction. Much research has been done to identify efficient indicators for prediction. For instance, Krizmaric et al. [5] have identified and correlated several variables (arrival time, initial $EtCO_2$, final $EtCO_2$, etc.) with out-of-hospital cardiac arrest, while patients' vital signs such as body temperature and blood pressure are correlated with in-hospital cardiac arrest. Churpek et al. [6] have successfully derived a model using vital signs to predict in-hospital cardiac arrest. However, researches have recently shown that not all vital signs are useful in the prediction of clinical outcomes [7]. Liu et al. [8] identified heart rate variability (HRV) as a new effective variable for prediction of mortality when used in combination with some other vital signs. It is not known what the best combination of feature variables for accurate cardiac arrest prediction is.

One study aiming to identify efficient predictors for cardiac arrest and to develop a scoring system for prediction is the geometric distance scoring system by Liu et al. [9]. The results suggested that the combination of HRV parameters and vital signs is a strong and efficient model for cardiac arrest prediction. This scoring system computes a risk score for a point (in feature space representing a patient) based on both its geometric distances from the point neighbors and the predicted binary outcome generated by support vector machine (SVM), which is a generalized classifier with learning algorithm widely used for data mining. With the advancement of computation technology, it is now possible to use machine learning techniques to solve various bio-medical problems, for example, improving the accuracy and functionality of prediction models [10], identifying relevant factors [5] and discovering new ways to utilize medical signals [11]. Nonetheless, each machine learning technique has its own drawbacks and limitations; the performance of the geometric distance scoring system still has room for improvement.

In this study, we aim to develop a novel manifold ranking based scoring system, and test it on an application of cardiac arrest prediction. The geometric distance scoring system makes use of only local structure embedded by the data, whereas the proposed scoring system explores both the local and global consistency of the data and thus has potential to further improve the prediction accuracy. Manifold ranking, a semi-supervised graph-based learning algorithm used for information retrieval, plays the key role in the development of the proposed scoring system. This study shows that manifold ranking has potential to be used in other bio-medical applications.

2. Methods

2.1. Dataset: balanced data vs. imbalanced data

The dataset used in this paper was original collected by the Department of Emergency Medicine, Singapore General Hospital, Singapore for an observational clinical study between November 2006 and December 2007. Ethics approval from the Institutional Review Board (IRB) with a waiver of patient consent for the study was obtained.

All patients were initially triaged by a nurse, and those with Airway, Breathing, Circulation problems, or thought to be possibly unstable and needing close monitoring are routinely put on ECG monitoring using the LIFEPAK 12 defibrillator/monitor (Physio-Control, Redmond, WA). Lead II ECGs sampled at 125 Hz were extracted as text files for HRV analysis using CODE-STAT Suite data review software (version 5.0, Physio-Control) and proprietary ECG extraction software. Charts were included for review if they had an ECG recording showing sinus rhythm and were excluded if they were in nonsinus rhythm (ventricular or supraventricular arrhythmias). A list of HRV parameters used in this study is shown in Table 1.

Demographic data and vital signs (systolic blood pressure, diastolic blood pressure, respiratory rate, heart rate, Glasgow coma scale, temperature, pain score, and oxygen saturation) were obtained from hospital records; clinical outcomes, e.g., cardiac arrest within 72 h, were recorded for patients of interests. The 72 h outcome was heuristically defined as it is commonly used in the emergency department as a cutoff time point. The core idea of the proposed algorithm is to compute a point's risk score based on its position both with respect to its neighbors and in the global manifold. For an unbiased score prediction process, a balanced training dataset, which consists of equal number of samples with positive and negative outcome, is required.

However, most of the medical data is highly imbalanced. For example, the collected dataset used in this study consists of a majority group of 973 negative samples (94.9% patients without cardiac arrest within 72) and a minority group of 52 positive samples (5.1% patients with cardiac arrest within 72 h), giving an imbalance ratio of about 19. In this study, experiment on balanced data was first conducted followed by experiment on imbalanced data. The balanced dataset is made up of all 52 positive cases and another 52 randomly selected negative cases.

To derive a balanced training dataset out of an imbalanced one, the ensemble-based system [12,13] is adopted, where two commonly used methods are under-sampling and over-sampling. In this paper we implemented the under-sampling strategy. In brief, ensemble-based system is used to partition the majority training class (negative samples in our study) into M non-overlapping groups randomly. Each group has almost the same number of samples as the minority class. By combining the M majority class with the minority class, M balanced training ensembles are created. Risk score prediction of every testing data is done for M times based on the M balanced training ensembles. M risk scores are generated and the mean is assigned to the testing sample as

Table 1
List of HRV parameters and their definitions.

HRV parameter (unit)	Definition
aRR (s)	Average width of the RR interval
STD (s)	Standard deviation of all RR intervals
avHR (beats/min)	Average of the instantaneous heart rate (HR)
sdHR (beats/min)	Standard deviation of the instantaneous HR
RMSSD (s)	Root mean square of differences between adjacent RR intervals
NN50 (count)	Number of consecutive RR intervals differing by more than 50ms
pNN50 (%)	Number and percentage of consecutive RR intervals differing by more than 50ms
Triangular index	Total number of all RR intervals divided by the height of the histogram of intervals
TINN	Baseline width of a triangle fit into the RR interval histogram using a least squares
LF power (ms^2)	Power in low frequency range 0.04–0.15 Hz
HF power (ms^2)	Power in high frequency range 0.15–0.40 Hz
Total power (ms^2)	Total power estimated from RR intervals
LF norm (n.u.)	LF power in normalized units: LF/(Total power − VLF) × 100
HF norm (n.u.)	HF power in normalized units: HF/(Total power− − VLF) × 100
LF/HF	Ratio of LF power to HF power

VLF: Very low frequency power in range $\leq$ 0.04 Hz.