Analysis of the Clinical Utility of Algorithms in the 2009 PhysioNet/Computing in Cardiology Challenge for the Prediction of Acute Hypotensive Episodes

Grace M Mirsky[1], Maria S Guillem[2]

[1]Benedictine University, Lisle, IL, USA
[2]ITACA, Universitat Politècnica de València, Valencia, Spain

Abstract

An Acute Hypotensive Episode (AHE) typically indicates an impending life-threatening event for an ICU patient. Therefore, the accurate prediction of AHEs has clear clinical merit and was the focus of the 2009 Challenge. This paper analyzes if the competitors' methods actually provide clinical utility. This analysis was accomplished using the portion of the training set that included records not containing any AHEs. In this training set, the 10 hours prior to the forecast window were analyzed for 15 records using the algorithms described by Chen, Mneimneh, and Langley.

When Chen's method was evaluated using a sliding 5-minute window (15 records, 10 hours each), this method generated 884 false alarms. Mneimneh's algorithm produced 131 false alarms over the same 150 hours. Though both Chen and Mneimneh scored 10/10 in Event 1 and 36/40 in Event 2, these results indicate that these methods could never be used practically.

Although Langley's algorithm did not perform as favorably in the Challenge, 8/10 in Event 1 and 28/40 in Event 2, it had zero false alarms in the training set used in this study. This indicates that some potential may exist for a relatively simple method to be clinically useful, but performance must be considered both in terms of predictive accuracy as well as low false alarms rates.

1.   Introduction

An Acute Hypotensive Episode (AHE) typically indicates an impending life-threatening event for an intensive care unit (ICU) patient [1]. It is defined as a thirty minute time period in which the mean arterial pressure (MAP) is less than or equal to 60mmHg for at least 90% of that thirty minute window. An example of a blood pressure recording with an AHE is shown in Figure 1.

Since an AHE can indicate a wide variety of conditions, the accurate diagnosis of the underlying event must be made quickly in order to provide immediate intervention. Due to the severe nature of the possible outcomes (death or severe organ damage), accurate prediction of AHEs has clear clinical merit and has been studied by a number of authors with the goal of increasing the accuracy of prediction [2-6].
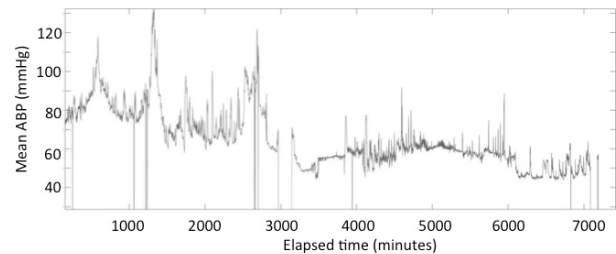

Figure 1. Example of drop in blood pressure which signifies an acute hypotensive episode.

1.1.   2009 Challenge

The accurate prediction of AHEs was the focus of the 2009 PhysioNet / Computing in Cardiology Challenge. The Challenge was comprised of two events in which participants competed: event one, which used test set A (ten patients), and event two, which used test set B (forty patients). The training set consisted of datasets from sixty patients, which were divided into four categories: H1, H2, C1, and C2. These designations are outlined in Table 1. These patient data were obtained from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) II database [7-9]. This database is distributed by PhysioNet and contains a range of physiological data recordings from ICU patients from 2001 to 2008.

The electrocardiogram (ECG) and arterial blood pressure (ABP) recordings were provided for each patient in the 2009 Challenge [1]. Each of these signals was sampled at 125 Hz. In addition, the heart rate, mean ABP, diastolic ABP and systolic ABP were sampled once per minute and provided in the training set for each patient. A one-hour forecast window was defined for each patient in which participants needed to identify whether or not an ABP would occur in that window, and at least ten hours of data prior to the forecast window were also provided.

Participants' algorithms were evaluated against a hidden holdout test set, and the number of submissions was limited to four attempts per participant for each event. Participants were instructed that in event one, precisely five out of the ten entries should be classified as belonging to H1. In event two, participants were instructed that ten to sixteen records should be classified as belonging to group H.

Table 1. Description of the four categories used in the training dataset [1].

Category	AHE in forecast window?	Additional details
H1	Yes	Received pressors
H2	Yes	Did not receive pressors
C1	No	No AHE during hospital stay
C2	No	AHE(s) occurred outside of forecast window

Out of nineteen submissions, the top three algorithms in the 2009 Challenge were submitted by Henriques, Chen, and Mneimneh [1]. Scores for the Challenge were reported as the number of correctly classified patients divided by the total number of patients in each event. Chen's algorithm scored 10/10 in event 1 and 36/40 in event 2. Mneimneh's algorithm scored 10/10 in event 1 and 36/40 in event 2. Henriques' algorithm scored 10/10 in event 1 and 37/40 in event 2. With these scores, Chen won event 1, and Henriques won event 2. There were actually several competitors who also scored 10/10 in event 1, but Chen was the first to submit an entry with a perfect score and was designated as the winner for that event.

Several of the top performers were very successful while using simplistic techniques based on a single observed parameter. Specifically, Chen's method used the mean of the Diastolic ABP over the 5-minute segment immediately before the forecast window [10]. The results were sorted and split according to the known distribution of expected AHEs per event in order to make the prediction. Secondly, Mneimneh's method classified an AHE by checking if the 20-minute segment prior to the forecast window had a $MAP < 71.1 mmHg$ [11].

Henriques' algorithm, however, was not based on a simple technique [12]. Rather, this algorithm utilized neural network multi-models in which the ABP signal prior to the forecast window is analyzed in terms of its correlation to a series of templates, which is used to predict the morphology of the signal in the forecast window. The predicted signal is then used to evaluate if an AHE will occur.

In addition to the three top scoring entries, the submission from Langley was also evaluated in this study, since it also followed the criteria of using a simple technique based upon a few parameters. This method involved examining a 30-minute window and determining if the AHE index exceeded 0.75, which indicated that more than 75% of the values in that window fell below 65mmHg [13].

1.2. False alarm analysis

While the Challenge evaluated the ability to predict an AHE over a given forecast window, only a single prediction per patient was required for the competition. In a realistic clinical scenario, however, patients would need to be continuously monitored, which creates the possibility of false alarms.

The presence of frequent false alarms in the ICU has been shown to reduce the quality of patient care [14,15]. High levels of noise in the ICU, resulting from false alarms, have been shown to sidetrack caregivers and irritate patients by frequently waking them up throughout the night [16,17]. In fact, the presence of false alarms is such an important problem that the focus of the 2015 PhysioNet / Computing in Cardiology Challenge was the reduction of false cardiac arrhythmia alarms in the ICU [18].

2. Methods

In order to evaluate the robustness of each of the methods submitted to the Challenge, 150 hours of data were selected for analysis. This set of data was comprised of the ten hours prior to the forecast window for the fifteen records in the training set which did not have any AHE's; these are records that were classified as type "C1" according to Table 1. Therefore, any AHE prediction algorithm executed over this data should optimally result in a prediction of zero AHE's, since the patients in this group did not experience any AHE's during their hospital stay.

For this study, each of the competitors' methods was implemented in MATLAB® [19] and executed over the selected 150 hours of data. The algorithms were evaluated every five minutes as soon as evaluation was possible in each record. A start-up time in which records could not be evaluated was necessary because, for instance, for Mneimneh's method, it involved the evaluation of a 20-minute segment. Therefore, the analysis obviously could not begin until at least twenty minutes had elapsed from the recording. After the initial twenty minutes, the algorithm was evaluated every five minutes. This technique was used for the robustness analysis since in real-life, a predefined forecast window is obviously not available. Patients must be monitored continuously; therefore, this technique provides a more realistic clinical

scenario because it results in a periodic analysis of the patient's current state.

3.   Results

When Chen's algorithm was evaluated every five minutes, this method generated 884 false alarms. Using Mneimneh's technique at the same interval for test, there were 131 false alarms. Though Langley's algorithm did not perform as favorably in the actual Challenge as the top scorers (8/10 in Event 1 and 28/40 in Event 2), it had zero false alarms in the 150 hour training set used in this study. The performance results are summarized in Table 2.

Henrique's method was very successful in the Challenge; however, it cannot be modified in a straightforward way to be analyzed for robustness using a sliding window, since the correlation analysis described in the study made use of all of the data prior to the forecast window. Therefore, the robustness could not have been fairly compared to the other methods, since using only a portion of the data prior to the forecast window (particularly for the sliding windows starting near the beginning of the recording) would have altered the presumed effectiveness of the method.

Table 2. Performance summary of each of the methods analysed for prediction of acute hypotensive episodes [1].

First author	Event #1	Event #2	False alarms
Chen	10/10	36/40	884
Mneimneh	10/10	36/40	131
Langley	8/10	28/40	0

4.   Discussion

Though both Chen and Mneimneh scored 10/10 in Event 1 and 36/40 in Event 2 [10,11], the results from the false alarm analysis indicate that these methods could never be used practically. An additional limitation of Chen's method is that it requires a priori knowledge of the number of AHEs in the test set. This is clearly an impossible assumption in any realistic situation, which further limits its clinical usefulness. Of course, these methods were not designed to be used clinically and were designed specifically for the competition. However, it is important to note that since a number of methods submitted to the Challenge exploit the knowledge of the forecast window, there is a significant limitation of how useful these algorithms would be in practice.

Langley's results were most interesting for this analysis, since the technique generated a perfect score of zero false alarms over the 150 hours analyzed. While Langley's performance in the Challenge was not as good

as the top scorers, this result indicates that some potential may exist for a relatively simple method to be clinically useful, since it did not generate an intolerable number of false alarms. Nonetheless, these results further assert that performance must be considered both in terms of predictive accuracy as well as maintaining low false alarms rates.

References

[1] Moody GB, Lehman LH. Predicting Acute Hypotensive Episodes: The 10th Annual PhysioNet/Computers in Cardiology Challenge. Computers in Cardiology 2009;36:541-544.

[2] Crespo C, McNames J, Aboy M, Bassale J, Ellenby M, Lai S, Goldstein B. Precursors in the Arterial Blood Pressure Signal to Episodes of Acute Hypotension in Sepsis. Proceedings of the 16th International EURASIP Conference BIOSIGNAL, 2002; 16; 206-8.

[3] Rocha T, Paredes S, Carvalho P, Henriques J, Harris M. Wavelet based time series forecast with application to acute hypotensive episodes prediction. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. IEEE, 2010; 2403-2406.

[4] Dernoncourt F, Veeramachaneni K, O'Reilly U. Gaussian process-based feature selection for wavelet parameters: Predicting acute hypotensive episodes from physiological signals. 2015 IEEE 28th International Symposium on Computer-Based Medical Systems. IEEE, 2015: 145-150.

[5] Janghorbani A, Arasteh A, Moradi, MH. Prediction of acute hypotension episodes using Logistic Regression model and Support Vector Machine: A comparative study. Iranian Conference on Electrical Engineering, IEEE, 2011;1-4.

[6] Jin K, Stockbridge N. Predicting acute hypotensive episodes from ambulatory blood pressure telemetry. *STATISTICS AND ITS INTERFACE* 2012;5(4):425-429.

[7] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000 (June 13);101(23):e215–e220. Circulation Electronic Pages: http://circ.ahajournals.org/cgi/content/full/101/23/e215.

[8] Saeed M, Lieu C, Raber G, Mark RG. MIMIC II: A massive temporal ICU patient database to support research in intelligent patient monitoring. Computers in Cardiology 2002; 29:641–644.

[9] MIMIC II Database. http://physionet.org/physiobank/database/mimic2db/.

[10] Chen X, Xu D, Zhang G, Mukkamala R. Forecasting acute hypotensive episodes in intensive care patients based on a peripheral arterial blood pressure waveform. Computers in Cardiology 2009;36:545-548.

[11] Mneimneh MA, Povinelli RJ. A rule-based approach toward the prediction of acute hypotensive episodes. Computers in Cardiology 2009;36:557-560.

[12] Henriques JH, Rocha TR. Prediction of acute hypotensive episodes using neural network multi-models. Computers in Cardiology 2009; 36: 549-552.

[13] Langley P, King ST, Zheng D, Bowers EJ, Wang K, Allen J, Murray A. Predicting acute hypotensive episodes from

mean arterial pressure. Computers in Cardiology Conference 2009; 36:553-556.

[14] Chambrin, MC. Alarms in the intensive care unit: how can the number of false alarms be reduced?. Critical Care-London 2001;5(4):184-188.

[15] Cvach, M. Monitor alarm fatigue: an integrative review. Biomedical Instrumentation & Technology 2012;46(4): 268-277.

[16] Konkani A, Oakley B, Bauld TJ. Reducing hospital noise: a review of medical device alarm management. Biomedical Instrumentation & Technology 2012;46(6): 478-487.

[17] Schmid F, Goepfert MS, Reuter DA. Patient monitoring alarms in the ICU and in the operating room. Critical Care 2013;17(2):216.

[18] Clifford GD, Silva I, Moody B, Li Q, Kella D, Shahin A, Kooistra T, Perry D, Mark RG. The PhysioNet/computing in cardiology challenge 2015: reducing false arrhythmia alarms in the ICU. Computing in Cardiology 2015;273-276.

[19] MATLAB Release 2016, The MathWorks, Inc., Natick, Massachusetts, United States.

Address for correspondence.

Prof. Grace M. Mirsky
Department of Computer Science
Benedictine University
5700 College Road
Lisle, IL, USA
gmirsky@ben.edu