Predicting Acute Hypotensive Episodes: The 10th Annual PhysioNet/Computers in Cardiology Challenge

GB Moody, LH Lehman

Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

This year's PhysioNet/Computers in Cardiology Challenge aimed to stimulate development of methods for identifying intensive care unit (ICU) patients at imminent risk of acute hypotensive episodes (AHEs), motivated by the possibility of improving care and survival of these patients. Participants were asked to forecast the occurrence of an AHE up to an hour in advance, in two groups of ICU patient records from the MIMIC II Database, drawing on data that included at least 10 hours of physiologic waveforms, time series, and accompanying clinical data prior to the one-hour forecast window. In event 1, most participants were able to identify without errors, in a group of 10 high-risk patients receiving pressor medication, which five of the patients experienced AHEs during the forecast window. In event 2, participants were able to classify correctly as many as 37 (93%) of a diverse group of 40 patients, including nearly all of those who experienced AHEs.

1. Introduction

Among the most critical events that occur in intensive care units, acute hypotensive episodes require effective, prompt intervention. Left untreated, such episodes may result in irreversible organ damage and death. Timely and appropriate interventions can reduce these risks. Determining what intervention is appropriate in any given case depends on rapidly and accurately diagnosing the cause of the episode, which might be sepsis, myocardial infarction, cardiac arrhythmia, pulmonary embolism, hemorrhage, dehydration, anaphylaxis, effects of medication, or any of a wide variety of other causes of hypovolemia, insufficient cardiac output, or vasodilatory shock. Often the best choice may be a suboptimal but relatively safe intervention, simply to buy enough time to select a more effective treatment without exposing the patient to the additional risks of delaying treatment.

In cooperation with Computers in Cardiology, PhysioNet[1] hosts an annual series of Challenges, inviting participants to address significant open problems of clinical or scientific interest. The goal of this year's Challenge was to encourage the discovery of effective methods for identifying at least a subset of patients at imminent risk of AHEs.

For the purposes of this Challenge, we defined an AHE as an interval in which at least 90% of the non-overlapping one-minute averages of the arterial blood pressure waveform (MAP) were in the *acute hypotensive range* during any 30-minute window within the interval. We defined the acute hypotensive range to include MAP measurements no greater than 60 mmHg, and (because signal loss or artifact may result in anomalously low MAP) no less than 10 mmHg. Thus AHEs by this definition may contain short intervals of signal loss or MAP in the normotensive range, but these can be no longer than three minutes in any half hour, and any AHE must contain at least 27 minutes of MAP measurements in the acute hypotensive range.

Data for this year's Challenge come from the MIMIC II project[2], which has been collecting a representative sample of physiologic signals and time series, and accompanying clinical data, from patients in all of the ICUs of a major teaching hospital for several years. Its deidentified records, comprising the MIMIC II Database[3], have been and are continuing to be contributed to PhysioNet, which makes them available to the research community worldwide.

In December 2008, the MIMIC II Database contained 2320 complete adult patient records (including both recorded physiologic signals and time series, and accompanying clinical data). Each such record covers at least one entire ICU stay, typically including signals and time series with durations of several days with few or no interruptions. Arterial blood pressure was recorded in 1237 patients (53% of the 2320 available cases). During their ICU stays, 511 patients (41% of 1237) experienced one or more recorded AHEs. In-hospital mortality of patients with AHE (193 of 511, or 37.8%) was more than twice that of patients whose arterial blood pressure (ABP) was monitored and who did not experience AHEs (129 of 726, or 17.8%). To the extent that one might forecast AHEs in the ICU, increased vigilance and better opportunities for planning appropriate interventions may result, leading to improved care and survival of patients at risk of these events.

Challenge participants were asked to develop automated

techniques for predicting AHEs up to an hour in advance in selected ICU patient records, using any data available before the *forecast window* for each record. For each case, we chose a specified time, T_0, at least 10 hours after the beginning of the MAP time series. Participants used the portion of each record occurring before T_0 to predict if an AHE would begin during the hour following T_0, the forecast window.

2. Methods

We selected previously posted MIMIC II patient records, augmented for the first time with deidentified clinical data, for participants to use while training their algorithms. For testing their algorithms, we selected previously unposted patient records of subjects who had AHEs, and others who did not, and provided truncated versions of these records on PhysioNet.

2.1. Initial selection criteria

Not all MIMIC II records include all of the data elements needed for this Challenge. Records chosen for the Challenge's training and test data sets included:

- At least 10 hours of data before T_0, and at least one hour of data after T_0.
- ECG and ABP signals sampled at 125 Hz. (Two records in the training set did not include these signals, however.) Additional signals are available in virtually all cases.
- Time series of vital signs sampled once per minute (in the training set) and once per second (in the test sets). These include heart rate and mean, systolic, and diastolic ABP. Most records include a variety of additional vital-signs time series, most often including respiration rate and SpO_2.
- Clinical data entered into the ICU medical information systems (records of observations, measurements, and interventions performed in the ICU). These include intravenous medications and fluids as well as other medications ordered.
- Results of laboratory tests, records of medications ordered, and other data gathered in the hospital but outside of the ICU.

Patients whose MIMIC II records meet the criteria above are assigned to a group (H or C) and a subgroup (H1, H2, C1, or C2):

- *H*: those with an AHE beginning within the forecast window (the one-hour period following T_0)
  - *H1*: those who received pressor medication (dobutamine, dopamine, epinephrine, isoproterenol, norepinephrine, phenylephrine, or vasopressin)
  - *H2*: those who did not receive pressor medication
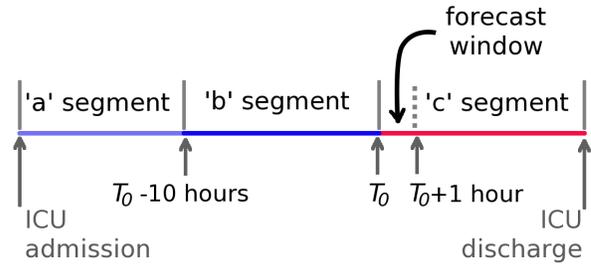- *C*: those with no AHE within the forecast window

Figure 1. Segmentation of test set records (see text).

- *C1*: those with no documented AHE at any time during their hospital stay
- *C2*: those who had one or more AHEs before or after the forecast window

2.2. The training set

We selected a training set of 60 cases from the set of MIMIC II records meeting the initial selection criteria and available in December 2008, and chose a T_0 for each case. The training set consisted of 15 records from each of groups H1, H2, C1, and C2. The classification of each case, and the data before and after T_0, were available for study.

2.3. The test sets

In April 2009, a set of records for an additional 1345 adult patients was contributed to PhysioNet from the MIMIC II project. These records incorporated several long-anticipated technical improvements over those included in the training set, including higher temporal resolution for most of the time series including MAP, higher amplitude resolution for the physiologic signals, and a larger number of simultaneously recorded signals.

From the newly available records meeting the initial selection criteria, we selected 50 for the Challenge test sets, and chose a T_0 for each case. Each selected record was divided into an 'a' segment including all data available more than 10 hours before T_0, typically beginning at or shortly following the patient's admission to the ICU; a 10-hour 'b' segment beginning at the end of the 'a' segment and ending at T_0; and a final 'c' segment beginning at T_0 and ending at the patient's discharge from the ICU (see Figure 1). The 'a' and 'b' segments were available to Challenge participants, although many chose to use the 'b' segments only. The 'c' segments (which include the forecast window), and the complete original MIMIC II patient records from which these test set records were derived, were withheld for the duration of the Challenge; they were first posted in September 2009.

Test set A, which was selected for event 1 of the Challenge, consisted of records from 5 H1 patients and 5 C1 patients, all of whom received pressors.

Although not explicitly stated as a selection criterion, none of the H1 or C1 patients whose records were included in the training set received pressors before T_0, and it was originally intended that this should be true in test set A as well. In selecting the test sets, however, it became apparent that group H1 cases without pressors before T_0 were less common than anticipated. For test set A, therefore, we chose *only* records of H1 and C1 patients who received pressors, so that, as in the training set, the presence of pressors per se does not indicate to which group a record belongs.

Test set B, used in event 2 of the Challenge, consisted of records from 14 H patients and 26 C patients, a proportion that very roughly matches the observed incidence of AHE among MIMIC II patients with MAP time series (as noted, 41%). Some, but not all of the patients represented in this set were among those receiving pressors, as in test set A.

2.4.   Challenge events

Event 1 focused on distinguishing between two groups of ICU patients who are receiving pressor medication: patients who experience an acute hypotension episode, and patients who do not. These two groups represent extremes of AHE-associated risk. Successful methods for separating these populations may lead to finding indices that are prognostic of AHE in these individuals. Participants were told that exactly five of the patients in test set A experienced AHEs in the forecast window.

Event 2 aimed to address the broad question of predicting AHE in a population in which about a third of the patients experience AHE. It is likely that a variety of methods can be used to identify different subsets of the patients at risk; for example, those who have had previous documented AHE (especially if more than once) may be relatively easy to identify, on the basis of a priori knowledge of their pathophysiology or of their response to medication. The potential benefits of finding AHE predictors for even a modest subset of the at-risk patients may be significant, if improvement in outcome can be shown to follow from increased vigilance and preparation for effective intervention in these patients. Participants were told that between 10 and 16 patients in test set B experienced AHEs in the forecast window.

The number of correct classifications was reported after each entry, and participants were permitted to make up to four entries in each event, allowing them a limited opportunity to improve their results.

3.   Results

Nineteen individuals and teams from 15 nations submitted entries to one or both of the Challenge events.

A goal of these Challenges is to encourage development of open-source solutions to the Challenge problems, which can jump-start follow-on studies, including creation of hybrid algorithms combining the strengths of complementary approaches. For this reason, we present awards not only for the best solution overall in each event, but also for the best open-source solution in each event. In this Challenge, the best solutions overall were open-source solutions.

The second entry received (only 27 hours after the test sets were posted) was the winning open-source and overall entry in event 1, from Xiaoxiao Chen of Michigan State University. His perfect score in event 1 was later matched by Niels Wessel, Hagen Malberg, Jingyu Yan, Florian Jousset, Franco Chiarugi, Jorge Henriques, Kun Jin, Fayyaz-ul-Amir Afsar Minhas, Thomas Ho Chee Tat, Pierre-Alexandre Fournier, Mohamed Mneimneh, and Dieter Hayn.

In event 2, the winning open-source and overall entry was from Jorge Henriques of the University of Coimbra, Portugal, with 37 correct classifications of a possible 40, including correct classifications of 13 of the 14 patients who experienced an AHE in the forecast window. Two participants (Xiaoxiao Chen and Mohamed Mneimneh) submitted entries with 36 correct classifications; these were also open-source entries.

Among the top 20 entries in event 2, records 224 and 222 were most often missed (in 15 and 12 entries respectively). Record 224 contains an AHE that follows a sudden drop in MAP from 65 mmHg at $T_0 + 35$ minutes to 60 mmHg six minutes later, and eventually to 45 mmHg at $T_0 + 56$ minutes. In record 222, MAP is declining just before T_0, but rapidly rises from 66 mmHg to 98 mmHg in the first minute after T_0, in response to administration of IV fluids. This is followed by a rapid decline to 63-70 mmHg in the next 5-10 minutes, following a pattern that can also be seen during the previous 10 hours of this record. This time, however, an AHE begins at about $T_0 + 25$ minutes, reaching a minimum of 54 mmHg.

Records 227 and 235 were most often incorrectly identified as likely to be followed by AHEs (by 18 and 13 of the top 20 entries respectively). MAP in record 227 is steady at 62-63 mmHg for most of the hour following T_0, and dips briefly to 53 at $T_0 + 56$ minutes, returning to 70 mmHg a few minutes later; it drops below 60 mmHg for about 15 minutes starting at $T_0 + 72$ minutes (beyond the forecast window). In record 235, MAP is stable for the hour following T_0, then declines sharply from 70 mmHg at $T_0 + 72$ minutes to 63 mmHg in less than a minute, and then drops into the acute hypotensive range a few minutes later (beyond the forecast window).

4.   Discussion and conclusions

The results obtained by most of the Challenge participants are excellent by any measure, yet it should be understood that perfect forecasting of AHE is neither possible (there will always be unanticipated events) nor necessary in order for AHE forecasting to have clinical utility. In this respect, the results exceed our expectations.

Methods used by many of the Challenge participants are discussed in their papers in this volume. It is clear from the results that excellent forecasting of AHE is possible with reference to the ABP signal or even the MAP time series alone, as Chen[4] and Henriques[5] showed in achieving the best results in the two challenge events, although other information from the patient records was used by other participants. In this respect, we had hoped to see more use of other information, both to inform the interpretation of the ABP and to provide robust forecasting in the context of ABP signal loss or degradation. These aims remain topics for future work.

The most common false positive cases contain episodes that, by less restrictive definitions of AHE and of the forecast window, might have been called true positives. In practice, it would have been appropriate for an AHE forecast to identify these patients as likely to experience AHEs.

It is also interesting to observe that a combination of the strategies employed by the most successful participants can yield even better performance in event 2 than was obtained using any single method. For example, a meta-forecast that predicts AHE if any of the top 3 algorithms (those of Henriques, Chen, and Mneimneh[6]) do so would classify all 14 H patients correctly (but with 5 false positives). A meta-forecast that predicts AHE only if all three of these algorithms agree would classify all of the C patients correctly, missing only two H patients.

Following the conclusion of the Challenge in September, the missing portions of the Challenge dataset were posted on PhysioNet, to support followup studies. See http://physionet.org/challenge/2009/ for access to the Challenge datasets, the contributed code for the open-source division entries, and additional information about the Challenge.

Acknowledgements

References

[1] 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.

[2] 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.

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

[4] 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.

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

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

Address for correspondence:

George B. Moody
MIT E25-505A, Cambridge, MA 02139 USA
george@mit.edu