Predicting Patient's Trajectory of Physiological Data using Temporal Trends in Similar Patients: A System for Near-Term Prognostics

Shahram Ebadollahi, Jimeng Sun, David Gotz, Jianying Hu, Daby Sow, Chalapathy Neti
IBM T.J. Watson Research Center, Hawthorne, NY.

Abstract

Providing near-term prognostic insight to clinicians helps them to better assess the near-term impact of their decisions and potential impending events affecting the patient. In this work, we present a novel system, which leverages inter-patient similarity for retrieving patients who display similar trends in their physiological time-series data. Data from the retrieved patient cohort is then used to project patient data into the future to provide insights for the query patient. The proposed approach and system were tested using the MIMIC II database, which consists of physiological waveforms, and accompanying clinical data obtained for ICU patients. In the experiments we report the effectiveness of the inter-patient similarity measure and the accuracy of the projection of patients' data. We also discuss the visual interface that conveys the near-term prognostic decision support to the user.

1 Introduction

The task of prognosis is an important component of the process of clinical care. It is about predicting the future health status of the patient and the probable course of her health indicators [6, 8]. Oftentimes clinicians are concerned about the near term potential trajectory of a number of *Key Patient Indicators* (KPIs). In this work we are primarily interested in providing insights to the clinicians into the near-term prognosis and projections of those KPIs by leveraging inter-patient similarities. We present a prototype of our system MITHRA, which stands for MIning Temporal Health Records for Advanced prognostic decision support, in the context of ICU patient care.

Figure 1 illustrates the concept behind our approach for making such predictions. Clinically similar patients are retrieved for a query patient whose trajectory of KPIs are available up to the present time. Using the temporal characteristics of a cohort of retrieved patients, the KPIs of the query patient are projected into the future. A user of the system can also select the subset of patient KPIs that are most relevant to the context of the task in the care process to have the system perform the above steps based on the selected subset. The two main challenges for accomplishing this task are: (1) Devising similarity measures that can reflect the clinical proximity among patients by taking into account the domain knowledge; and (2) Projecting patients' data into future based on the characteristics of the data in the similar patient cohort.

In this paper, we report the implementation of this concept in the context of monitoring patients in the Intensive Care Unit (ICU). ICUs are data rich environments where patients are continuously being monitored for several aspects of their health. Alerts that can indicate the incidence of an imminent adverse condition based on the behavior of patients temporal data are important support mechanism for physicians in this environment. Accompanying those alerts with insight regarding the likely behavior of patient KPIs can further qualify and clarify them. In this setting, our goal is to retrieve patients who display similar evolution patterns in their ICU data to the patient being monitored and use the *future* trend of the cohort of similar patients to predict if the patient being monitored is going to experience a medical event within a specific time horizon. The insight provided to the clinician through the pro-
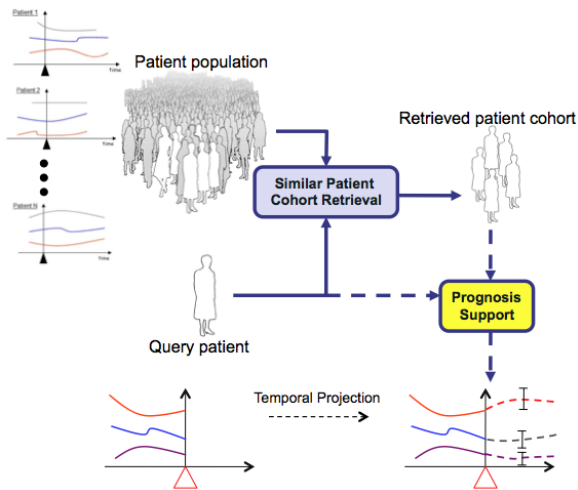
1

Figure 1: Retrieving patients based on their clinical similarity to a query patient and using the retrieved patients to project the evolution of patient's clinical characteristics.

jections of the patient's physiological data into the future could further clarify and qualify the generated alerts.

The proposed approach and system were tested using the *MIMIC II* database, which consists of physiological waveforms, and accompanying clinical data obtained for ICU patients [1].

The most similar and relevant work to ours is the one reported by Saeed and Mark [7]. They employed a multi-resolution description scheme for physiological temporal ICU data and used unsupervised similarity metrics and a K-Nearest Neighbor algorithm to predict the occurrence of the event. Our approach is not only capable of producing such kinds of classifications, but also can go one step further and provide the trajectory of the patient physiological data in the future time intervals.

2 Methodology

The key components of MITHRA for ICU setting are: physiological data stream processing for efficient data preprocessing and feature detection, patient similarity metric design incorporating medical domain knowledge, prognosis prediction based on similar patients, and fi-

nally, a visualization environment for point-of-care decision support.

2.1 Physiological Stream Processing and Feature Extraction

To cope with the large amount of patient physiological data generated by ICU monitoring devices, we leverage a state of the art stream computing platform [2] to perform missing value imputation, clinical event detection, and feature extraction.

Taking advantage of the efficient window operators provided by the stream processing system, sliding windows are used to detect the occurrence of clinically significant events. For example, we follow the rule defined by PhysioNet [5] to detect Acute Hypotensive Episode (AHE) events. This rule checks if 90% of the mean Arterial Blood Pressure (ABP) data stream has values below 60 in a 30-minute window. All events are then persisted for the training of similarity metrics performed by the supervised metric learning module.

Features are also computed using the sliding window operator, in which every sliding window generates a feature vector. In the current system, we use the top-F wavelet coefficients as the features.

2.2 Similarity Metric Design

When a physician looks for similar patients in a database, the similarity is often based not only on quantitative measurements such as the ones obtained from patient monitoring devices and through Electronic Health Records (EHR), but also the physician's judgment of clinically relevant factors and what she deems to be clinically similar. To capture this clinically relevant notion of similarity, we need to learn a distance metric that can automatically adjust the importance of each numeric feature by leveraging the domain knowledge. We have incorporated a method called *Locally Supervised Metric Learning* (LSML) to address this requirement.

Formally, quantitative measurements of a patient are represented by a N-dimensional feature vector $\mathbf{x}$. The domain knowledge could be captured as labels on some of the patients. With this formulation, our goal is to learn a generalized Mahalanobis distance between patient $\mathbf{x}_i$ and patient $\mathbf{x}_j$ defined as:

2

$$d_m(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{P}(\mathbf{x}_i - \mathbf{x}_j)} \qquad (1)$$

where $\mathbf{P} \in \mathbb{R}^{N \times N}$ is called the *precision matrix*. The key is to learn the optimal $\mathbf{P}$ such that the resulting distance metric has the following properties: 1) *Within-class compactness:* patients of the same label are close together; 2) *Between-class scatterness:* patients of different labels are far away from each other. Note that the same label reflects characteristics that are believed to lead to clinical similarity. In the evaluation, we use the event AHE as the label.

To formally measure these properties, we use two kinds of neighborhoods defined for each patient x_i: The *Homogeneous neighborhood* of $\mathbf{x}_i$, denoted as $\mathcal{N}_i^o$, is the k-nearest patients of $\mathbf{x}_i$ with the same label. The *Heterogeneous neighborhood* of $\mathbf{x}_i$, denoted as $\mathcal{N}_i^e$, is the k-nearest patients of $\mathbf{x}_i$ with different labels. We define the local compactness and scatterness around point x_i as the sum of squared distances within its homogeneous neighborhood and heterogeneous neighborhood, respectively. We then formulate an objective function to find a $\mathbf{P}$ that minimizes the local compactness and maximizes the local scatterness over all patients [9].

2.3 Prognosis Based on Similar Patients

When a query patient with available observations up to a decision point is presented to MITHRA, the stream processing components are first applied to extract features from an *assessment window*. The features are then used to retrieve the K most similar patients using the similarity metric learned during off-line data analysis.

During the retrieval process, temporal alignment is performed between the query patient and each candidate patient to identify the window in the candidate patient's history that best matches the query patient's assessment window. The alignment is currently carried out by "sliding" the query patient's assessment window against the candidate patient. The position along the candidate patient's temporal sequence that yields the smallest distance between the two windows (computed using the metric learned during offline analysis phase) is then identified as the anchor point.

Once the n reference patients are retrieved and properly aligned, the observations on these patients after the anchor point can be used to predict the prognosis (future measurements) of the query patient. Let $\mathbf{x}_i = \{x_{i1}, x_{i2}, ..., x_{iw}\}$ represent measurements from one sensor for the reference patient i in the window preceding the anchor point, and $\mathbf{y} = \{y_1, y_2, ..., y_w\}$ represent the measurements from the same sensor for the query patient in the window preceding the decision point. The linear regression model takes the form:

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + ... + \beta_n \mathbf{x}_n, \qquad (2)$$

where parameters β_i, $i = 0, 1, ...n$ are solved by the least squares estimator. Under the assumption that relationship between the reference patients and the query patient remains the same before and after the decision/anchor point, the predicted measurements for the query patient in the window succeeding decision point, $\bar{\mathbf{y}}$, can be computed by plugging in on the right hand side the observed measurements from the reference patients in the window succeeding anchor point, $\bar{\mathbf{x}}_i, i = 1, ..., n$. In other words:

$$\bar{\mathbf{y}} = \beta_0 + \beta_1 \bar{\mathbf{x}}_i + \beta_2 \bar{\mathbf{x}}_2 + ... + \beta_n \bar{\mathbf{x}}_n. \qquad (3)$$

2.4 Visualization

Prognostic data for a query patient is conveyed to clinicians through a modified ICU monitor. The monitor display includes both (1) predicted signal measurements and (2) alerts for imminent adverse conditions that are expected based on our analysis of the similar patient cohort. For each physiological signal, a time-series of predicted measurements is visualized to the right of the corresponding historical plot. For predicted values, accuracy is conveyed via a confidence interval rendered both above and below the predicted measurements as shown in Figure 2. Predicted events are integrated into the visual display as well. Textual alerts are displayed in the top left corner of the monitor and highlighted in red. Signals that led to the alert condition are also highlighted.

The monitor also provides clinicians with the ability to modify the set of KPIs used to select the similar patient cohort. By default, the system uses all physiological signals. However, the interface allows clinicians to customize the set of signals used to calculate the near-term prognosis based on the context. This setting can be changed at any time by the clinician.
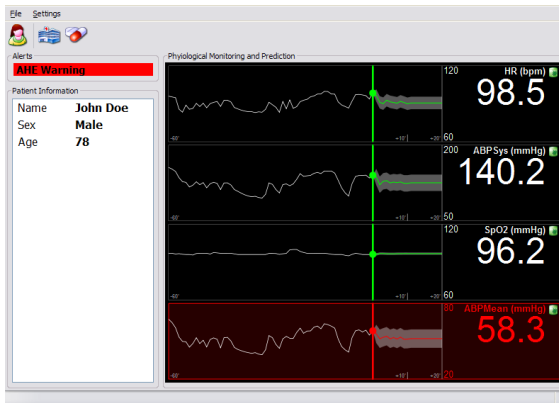
3

Figure 2: An ICU monitor visualizing near-term prognostic information along side historical physiological data.

3 Experiments

Experiments were carried out using physiological data for 1500 patients downloaded from the MIMIC II database [1]. The physiological streams for each patient include mean ABP measure, systolic ABP, diastolic ABP, Sp02 and heart rate measurements. Every sensor is sampled at 1 minute intervals.

3.1 Similar patient retrieval

To evaluate the supervised metric learning scheme, we partitioned 1500 patients into two groups *H* or *C*. Those in group H (590 patients) experienced AHE events, whereas those in group C (910 patients) did not experience any AHE.

For each patient in group C, we extracted a 2-hour window centered around a random timestamp T_0. For each patient in group H, we extracted a 2-hour window centered around T_0, such that in the hour after T_0, the patient experienced AHE. We used 80% of those patients for training and 20% for testing.

Methods of comparison   With the existing rule-based streaming event detection [5], AHE can not be detected until at least 30 minutes after T_0, because the sliding window operator needs to have enough samples in order to detect the event. We treat T_0 as the decision point and evaluate the effectiveness of our similarity measure by its ability to predict the onset of AHE in the one hour window after T_0 based on the data before T_0.

As the baseline, we compare our method against the winning method used in the 10th Annual PhysioNet/Computers in Cardiology Challenge, as presented in [3], referred to as *Challenge09* below. Since in that method only the mean ABP stream was used, we also used only that stream in the similarity measure evaluation to ensure fair comparison.

More specifically, in our method LSML, the wavelet coefficients of the 1-hour window from mean ABP were computed. We used Daubechies-4 Wavelet [4] and kept the top-10 coefficients as a feature vector. We thus obtained 1500 m-dimensional feature vectors where $m = 10$.

Classification and Retrieval Performance   The performance metrics we used include k-NN *classification error rate* and *precision@10* retrieval results. The *precision@10* of a query point is computed by retrieving 10-nearest points with a specific distance metric and then computing the percentage of those retrieved points having the same label as the query patient.

Table 1 shows both classification results in accuracy and retrieval results in precision@10 using 5-NN classifier. We observe that our LSML method consistently outperforms the baseline method.

Table 1: Classification and Retrieval Accuracy

	Challenge09	**LSML**
Classification (accuracy)	0.4982	**0.8551**
Retrieval (precision@10)	0.5230	**0.7998**

3.2 Prognosis Accuracy

To evaluate the accuracy of our prognosis analysis module we used the same data as in the previous evaluation for similarity measure. Again, we used 80% patients for training, and 20% for testing. The objective was to predict the mean ABP value on a per minute basis for the hour after T_0 for a query patient. For each AHE patient, we set T_0 as the timestamp when AHE occurs, and we use the 1-

4

hour window before T_0 as the query window, and evaluate prognisis accuracy over the 1-hour window after T_0.

Traditional approaches for prognosis only use current patient condition without systematic ways of leveraging other similar patients. To simulate traditional decision making, we used the following baseline methods:

- *Mean value* uses simply the mean of the measurements in the query window.
- *AR* builds an autoregression model to forecast the values after T_0. We used AR(5) in our experiment.

For evaluation we use the relative error rate defined as:

$$ e = \frac{\sum (x_i - \hat{x}_i)^2}{\sum_i x_i^2}. $$

Figure 3 shows the prediction relative error rate increases as the prediction horizon increases. Our similarity-based method significantly outperforms both baselines, which confirms the benefit of retrieving similar patients to help assess the query patient's prognosis.
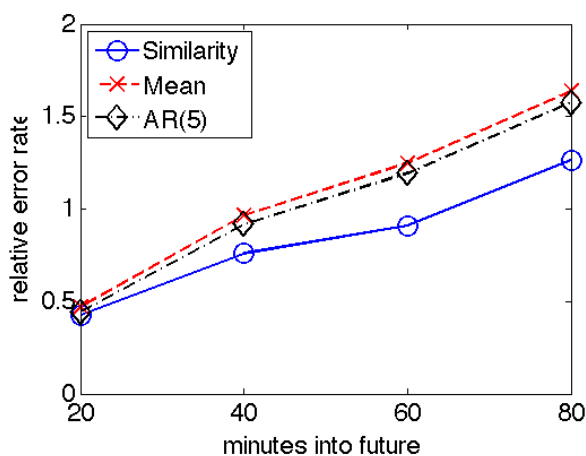

Figure 3: Patient Prognosis Accuracy vs. prediction horizon.

4 Conclusion

We have presented a prototype of MITHRA, which is our system for near-term prognostics, for the ICU patient care.

This system provides important insights to clinicians in an ICU setting. The system leverages stream processing techniques for efficient data pre-processing and feature extraction, a supervised patient similarity metric capable of incorporating domain knowledge, and visualization techniques for providing decision support at the point of care. Given a query patient, the similarity metric is used to retrieve a cohort of similar patients. The data of the patients in this cohort is then used to project the query patient's data into the future using correlation-based techniques. The projections are then visualized along with alerts and other patient indicators for providing insights to the clinicians. The methodology has been tested using physiological data from 1500 patients with promising results. While the experiments were carried out using AHE events, because of the existence of publicly available data, our methodology is completely event agnostic by design and is expected to be applicable to other events. The limitations of this work are: 1) we assume that accurate and sufficient label information is provided by the physician, where such label might not be available in general; 2) we only allow a single label per patient, while in practice there are cases of multiple labels per patient due to comorbidity. Future work includes investigating ways to address the above limitations and also providing comprehensive results on evaluation of the user interface.

References

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

[2] L. Amini, H. Andrade, R. Bhagwan, F. Eskesen, R. King, P. Selo, Y. Park, and C. Venkatramani. SPC: A distributed, scalable platform for data mining. In *Workshop on Data Mining Standards, Services and Platforms, DM-SSP*, Philadelphia, PA, 2006.

[3] X. Chen, D. Xu, G. Zhang, and R. Mukkamala. Forecasting acute hypotensive episodes in intensive care patients based on peripheral arterial blook pressure waveform. In *Computers in Cardiology (CinC)*, 2009.

[4] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

[5] http://www.physionet.org/challenge/2009/. Physionet/computers in cardiology challenge 2009: Predicting acute hypotensive episodes.

[6] M. G. M. Hunink and P. P. Glasziou. *Decision Making in Helath and Medicine - Integrating Evidence and Values*. Cambridge University Press, 2006.

[7] M. Saeed and R. Mark. A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations. In *American Medical Informatics Association*, 2006.

[8] H. C. Sox, M. A. Blatt, M. C. Higgins, and K. I. Marton. *Medical Decision Making*. American College of Physicians (ACP), 2007.

[9] F. Wang, J. Sun, T. Li, and N. Anerousis. Two heads better than one: Metric+active learning and its applications for it service classification. In *ICDM*, 2009.

5