Modeling Latent Pathophysiologic States

Rohit Joshi **Kanak Kshetri** **Choong-Hyun Lee** **Peter Szolovits**

Computer Science and Artificial Intelligence Laboratory,
Massachusetts Institute of Technology, Cambridge, MA 02139
{rjoshi,kanak,chl,psz}@mit.edu

Abstract

The large quantity of data collected in Intensive Care Units (ICUs) provides the possibility of building intelligent models to recognize and track patterns of critical conditions. We present the idea and discuss the challenges in building probabilistic models from batch data to predict the course of latent physiologic conditions.

1 Introduction

Even though modern ICUs store terabytes of patient data, the current bedside technologies offer limited help in understanding a patient's pathophysiologic states. For example, while monitors are capable of displaying the present value of a physiologic parameter like respiratory rate or oxygenation level, they do not show information about the patient's hidden abnormal states that might be responsible for the current condition or the direction in which the patient's physiology is heading [1]. In contrast, a modern navigation system like a GPS not only shows the path a driver took to reach a particular signpost but is also capable of predicting the traffic congestion and recalculating the predicted path if the driver chooses to change his course. Even without such technological assistance, critical care physicians are expected to digest and interpret these stacks of data instantly, and to make cost-effective quality decisions to improve patient outcomes. Understanding the trajectory a patient may follow and detecting possible abnormalities in physiologic conditions in real time can aid physicians in making timely and informed decisions and thus improve overall patient outcomes.

2 Proposal

We wish to track the changing latent pathophysiologic states of a patient in real time. We define a latent pathophysiologic state as an abstract hidden state which is learnt from the data and is assumed sufficient to understand a patient's behavior for a particular condition. Our definition is similar to the concept of an abstract trait in human psychology, such as a person's *intelligence* or the *difficulty level* of a test. These states may not exist in any physical or physiological sense but they are still sufficient to understand people's behavior: for example, in predicting the performance of a particular student on another set of tests. We believe that similar latent states may arise naturally if medical data can be effectively clustered and the underlying time-varying probability density can be estimated. We are currently working on probabilistic Markov models to find and track such latent patient states.

We study a previously preprocessed dataset [3] extracted from the Multi-parameter Intelligent Monitoring for Intensive Care II (MIMIC II) database [6], collected from multiple surgical and medical ICUs of a Boston-area hospital and like many batch databases, it is high dimensional and has partially missing and erroneous data values. In our work, we are interested in learning states and transitions at a longer time scale. So, instead of using high-resolution waveform signals, we decided to use nurse-verified hourly summarized data for all parameters. Our extracted dataset contains $26,647$ patients with over 1 million data records and 440 parameters. The patient's ICU stay is represented as multiple data records collected at an hourly time interval. Each record describes the patient's physiological information, ventilator data, lab values, fluids and medications administered, and other related features. We restrict patient data to the first 7 ICU days. At present, we have not exploited clinical data from the narrative text records or signal data from continuous monitoring.

3 Related Work

Recently, Hug [3] provided a real-time acuity score and showed that his methodology provides an elegant way for continuous risk assessment of ICU patients. In the medical data processing field, there have also been decades of research to detect trends from the time series data, to warn physicians of a life-threatening event in advance [9]. Most of these works learn supervised models that require high quality annotated datasets. In our work, we are interested in learning unsupervised probabilistic models using high-dimensional batch data, from thousands of patient records.

The probabilistic latent models have shown promising success in other research fields. For example, in Natural Language Processing, topic models [8] learn hidden topics from a collection of documents and then use them to semantically index these documents. In Collaborative Filtering, matrix factorization [7] approaches can identify the latent profiles of users and movies from the large Netflix dataset and then make recommendations on an unseen combination of a user and a movie.

In our case, it is difficult to identify latent patient states for three main reasons: First, unlike the latent models mentioned above, the patient state cannot be assumed to be static because a patient may transition into multiple states over time. Second, the number of such hidden states is unknown. Third, many existing clustering and probabilistic generative algorithms do not scale well to the size and high-dimensional nature of our dataset. For example, a popular hierarchical agglomerative clustering approach [2] requires prior specification of the number of clusters, which is difficult even for an experienced physician. Moreover, the algorithm is extremely computationally expensive, as it needs to recompute an $N \times N$ matrix several times for every new cluster assignment (N being the number of data records). Dynamic Bayesian Networks and Markov Models also need the number of states to be defined in advance. Some recent works have succeeded in applying complex Markov Models with explicit pre-defined set of states [4, 5] to discriminate physiological patterns. Non-parametric Bayesian approaches [10] can identify the number of hidden clusters without prior specification, but the current approaches work only on a very small dataset.

4 Conclusion

The understanding and tracking of the latent pathophysiologic states is a challenging task that can improve future health care delivery systems. Our work is relevant not only to the ICU domain but also to other related fields such as home monitoring, fitness devices and disability care.

References

[1] T Buchman. Novel representation of physiological states during critical illness and recovery. *Critical Care*, 14:127, 2010.

[2] M Cohen, A Grossman, D Morabito, et al. Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis. *Critical Care*, 14:R10, 2010.

[3] C Hug. *Detecting Hazardous Intensive Care Patient Episodes Using Real-time Mortality Models*. PhD thesis, MIT, 2009.

[4] L Peelen, NF De Keizer, E Jonge, et al. Using hierarchical dynamic bayesian networks to investigate dynamics of organ failure in patients in the intensive care unit. *J biomed Inform*, 43(2):273–86, 2010.

[5] J Quinn and C Williams. Physiological monitoring with factorial switching linear dynamical systems. *Probabilistic Methods for Time-Series Analysis*, 2010.

[6] M. Saeed, C. Lieu, G. Raber, and R. G. Mark. MIMIC II: a massive temporal ICU patient database to support research in intelligent patient monitoring. *Computers in Cardiology*, 29:641–644, 2002.

[7] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *NIPS*, 2008.

[8] M Steyvers and T Griffiths. Probabilistic topic models. *Handbook of Latent Semantic Analysis*, 2007.

[9] C Tsien. *Automated Detection of Alarmable Trends*. PhD thesis, MIT, 2000.

[10] F Wood, TL Griffiths, et al. A non-parametric bayesian method for inferring hidden causes. In *UAI*, 2006.