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 a 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

Modern ICUs store terabytes of patient data in their databases but the current bedside technologies offer limited help in understanding the patient's pathophysiologic states. For example, what the current state-of-the-art monitoring technologies in ICUs show is the present value of the physiologic parameters such as the respiratory rate or the oxygenation level. But, what they do not show is the equally important information about the patient's previous hidden abnormal states that may be responsible for the patient's current condition and about the direction in which patient's physiology might be heading towards [**?**]. On the other hand, consider a modern navigation system like a GPS, it not only shows what path a driver took to reach a particular signpost but can also predict the traffic congestion probability, and also recalculate 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 make cost-effective quality decisions to improve patient outcomes. Understanding what trajectory a patient may follow and detecting possible abnormalities in physiologic conditions in real time can aid physicians in making timely and better informed decisions and, thereby, improving overall patient outcomes.

2 Proposal

In our work, we are interested in tracking the changing latent pathophysiologic states of a patient in real time. We define a latent pathophysiologic state of a patient 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 like *personality* in human psychology. There researchers analyze data to estimate abstract characteristics such as person's intelligence or a difficulty level of an educational test. These states, *intelligence* and *difficulty level*, 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 these latent states may arise naturally if the medical data can be effectively clustered and the underlying probability density can be estimated. We are currently working on probabilistic Markov models to find and track such latent patient states.

We are studying a previously preprocessed dataset [**?**] extracted from Multi-parameter Intelligent Monitoring for Intensive Care II (MIMIC II) database. The MIMIC II database has been 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. Our dataset contains 26,647 patients with over 1 million data records and 440 features. Each patient data record describes patient's physiological information, ventilator data, lab values, fluids and medications administered.

3 Related Work

Recently, Hug [**?**] 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 alarm physicians of a life-threatening event in advance [**?**]. Most of these works are learning supervised models that require high quality annotated datasets. On the other hand, in our work, we are interested in learning unsupervised probabilistic models using high-dimensional batch data, with millions of patient records.

The probabilistic latent models have shown promising success in other research fields. For example, in Natural Language Processing, topic models [**?**] learn hidden topics from a collection of documents and then use them to semantically index these documents. In collaborative filtering, matrix factorization[**?**] 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: Firstly, unlike the above latent models, the patient state cannot be assumed to be static as a patient may transition into multiple states over time. Secondly, the number of such hidden states is unknown. Thirdly, 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 [**?**] 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 points). 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 [**? ?**] to discriminate physiological patterns. Non-parametric Bayesian approaches [**?**] can identify the number of hidden cluster 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 the 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.