Discrete Switching-State Systems for Intervention and Outcome Predictions: Understanding Vasopressor Administration and Weaning in the ICU

Abstract

Vasopressor administration is common for controlling hypotension in intensive care units (ICUs), and changes in timing and dosage can have a large impact on patient outcomes. In this work, we use a switching-state autogressive model (SSAM) to model the multidimensional physiological timeseries of patients before, during, and after vasopressor administration. Using features derived from this SSAM, we achieve state-of-the-art results in predicting whether a patient will be administered a vasopressor, and the same features can be used to predict when a patient can be successfully weaned from the intervention.

1 Introduction

Decision-making in the intensive care unit (ICU) requires responding quickly to rapidly changing situations, but the efficacy of many interventions remains unquantified (Vincent and Singer 2010; Vincent 2013; Ospina-Tascón, Büchele, and Vincent 2008). The vast amounts of data that are collected in ICUs—vital signs, clinical notes, fluids, medications—suggest an opportunity for more data-driven decision-making. Many works have used these ICU measurements to predict in-hospital or 30-day mortality of patients in particular disease subgroups (Che et al. 2015; Caballero Barajas and Akella 2015; Ghassemi et al. 2014). However, these risk scores are of limited value to clinicians, who must make decisions of how and when to treat patients regardless of their underlying acuity.

This work takes an important step toward the *actionable* use of ICU data by modeling interventions in the ICU. We focus on vasopressors, a class of drug used to elevate mean arterial pressure. While vasopressors are commonly used in the ICU, few controlled clinical trials have documented improved outcomes from their use (Müllner et al. 2004), and it may even be harmful in some populations (DAragon et al. 2015). We first consider predicting the need for vasopressor administration. Knowing that a patient will need a vasopressor can help the clinical staff plan interventions. We also consider the more interesting question of whether a patient is ready to be weaned from the vasopressor. In particular, patients are often left on interventions longer than necessary because clinicians are attending to other patients; however, extended interventions and hospital stays can be both costly and detrimental to patient health (DAragon et al. 2015).

Unfortunately, analyzing data from the ICU is challenging: clinical signals are often irregularly sampled and contaminated by interference and human error. Strong modeling assumptions are needed to clean and impute the signals (Lasko, Denny, and Levy 2013; Marlin et al. 2012). However, imputation techniques can introduce noise into models (Liu et al. 2006), and generally do not account for the highly dependent temporal nature of the data (Marshall et al. 2010; Lee et al. 2012; Janssen et al. 2010). Switching dynamical systems models (Lehman et al. 2015; Quinn et al. 2009) have been used to impute signals, identify artifacts, discover physiological states in a variety of critical settings.

Our work follows in this vein of using switching state models to model the physiological state of the patient. However, unlike prior work (Lehman et al. 2015; Quinn et al. 2009), we build discrete autoregressive models rather than trying to model the continuous signal; (Joshi and Szolovits 2012; Ghassemi et al. 2015) demonstrated how such discretization can assist in predicting outcomes. We also focus on actionable predictions regarding interventions, rather than mortality. Finally, we consider a large number of signals, compressing them into a few latent physiological states.

This work makes the following contributions:

- We define three types of *clinically-useful* prediction tasks: predicting the need for vasopressive intervention (in an ungapped and gapped manner), and the more challenging problem of predicting when to wean a patient from vasopressor administration.
- We develop a switching-state autogressive model that achieves state-of-the-art predictions for both intervention onset tasks using only physiological signals in a large, public ICU data set. Prediction of weaning is a harder problem, resulting in a lower predictive power. In all tasks, our model improves prediction accuracy over baselines.
- We quantify unnecessary intervention time as well as discuss the clinical relevance of our discovered latent states.

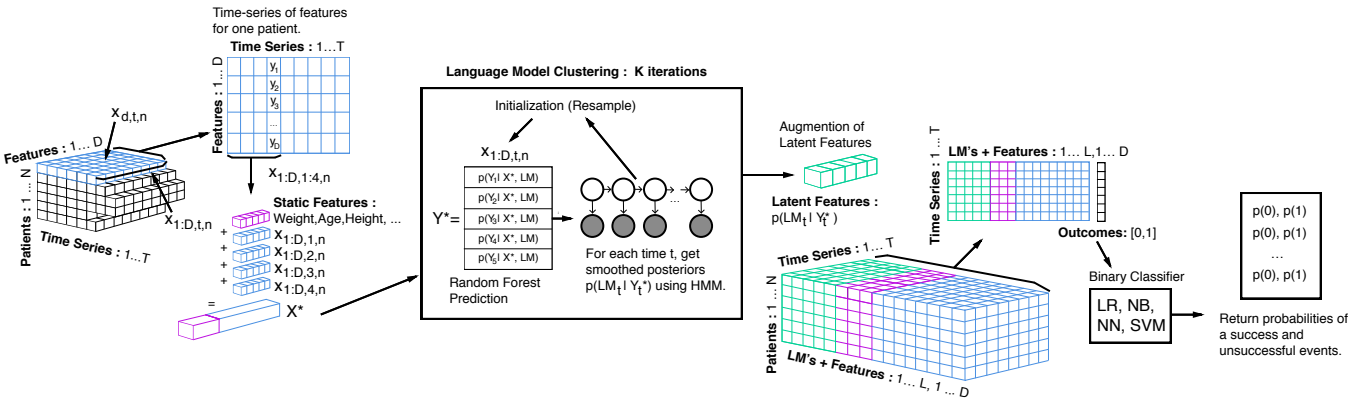

Figure 1: Overall graphical flow of experiment. (1) Baseline demographic features (e.g. age, sex, etc.), vital signs (heart rate, temperature, blood pressure, etc.), lab results (glucose levels, bicarbonate levels, etc.), and derived features (BMI) are extracted from the database for a filtered selection of patients. (2) For each patient, corresponding vital signs and lab results data are grouped into 4 hour blocks, flattened, and appended to the demographic features. (3) A switching-state autoregressive model is used the model the timeseries. (4) Latent features are then defined as the probability of each state at each time; these are appended to the features from step (2). (5) Given these features, a classifier is trained to predict the outcome of interest (e.g. vasopressor administration).

2 Physiological Model

Model Let the physiological signals $\boldsymbol{x}_t^n$ of a patient n at time t be a vector in $\mathbb{R}^P$ of P measurements, some of which may be missing. We learn a switching state autoregressive model (SSAM) in which we posit that there exists a hidden sequence $\{y_1^n, y_2^n, y_3^n, \dots\}$ generated from a hidden Markov Model (HMM) which controls the transition dynamics for the autoregressive model on the signals $\{\boldsymbol{x}_1^n, \boldsymbol{x}_2^n, \boldsymbol{x}_3^n, \dots\}$.

Specifically, we posit that the evolution of the sequence of observations $\{\boldsymbol{x}_1^n, \boldsymbol{x}_2^n, \boldsymbol{x}_3^n, \dots\}$ can be well-modeled by an autoregressive model $T_x(\boldsymbol{x}'(p)|x, \theta)$, where $\boldsymbol{x}'(p)$ is the p^{th} dimension of the signal at the current time, $\boldsymbol{x}$ is the signal at the previous time, and θ are some parameters associated with the model. In Section 3, we consider describe several options for the autoregressive transition model, including Naive Bayes and random forest classifiers.

We posit that these parameters θ are not constant for all time but are governed by a hidden physiological state sequence $\{y_1^n, y_2^n, y_3^n, \dots\}$, where each y_t^n is a discrete scalar in $\{1 \dots K\}$. Each state k is associated with a distinct set of parameters $\{\theta_{p,k}\}$, and we let $T_y(y'|y)$ be the probability of transitioning to state y' if the current state is y. Since there are K discrete states y, T_y is simply a KxK matrix. Let $T_x(\boldsymbol{x}'(p)|x, \theta)$ be the probability of observing the signal $\boldsymbol{x}'(p)$ given the signal at the previous time step $\boldsymbol{x}$. Since $\boldsymbol{x}$ may be high-dimensional, there are many choices for this transition function; we use θ to denote required parameters.

Then the generative process for the SSAM is given by

$$\begin{aligned} y_t^n &\sim T_y(\cdot|y_{t-1}^n) \\ \boldsymbol{x}_t^n(p) &\sim T_x(\boldsymbol{x}_t^n(p)|\boldsymbol{x}_{t-1}^n, \theta_{p,y_{t-1}^n}) \end{aligned}$$

The initial state y_0^n and measurements $\boldsymbol{x}_0^n$ are drawn from prior distributions π_y and $\pi_{\boldsymbol{x}}$. Thus, likelihood of the model is given by

$$\begin{aligned} L(\{y\}, \{\theta\}|\{\boldsymbol{x}\}) &= \prod_n^N \pi_y(y_0^n)\pi_{\boldsymbol{x}}(\boldsymbol{x}_0^n) \\ &\cdot \prod_{t=1}^{T_n} T_y(y_t^n \quad |y_{t-1}^n) \quad \prod_{p=1}^{P} T_x(\boldsymbol{x}_t^n(p)|\boldsymbol{x}_{t-1}^n, \theta_{p,y_{t-1}^n}) \end{aligned}$$

where L_n is the number of observations for patient n.

Inference The model above contains two sets of latent variables: the hidden physiological state sequences for each patient $\{y_1^n, y_2^n, y_3^n, \dots\}$ and the transition parameters $\theta_{p,k}$ for each measurement p and physiolotical state k. Our inference alternates between updating each of these sets of variables. We optimize the transition parameters $\theta_{p,k}$ and sample the physiological states $\{y_1^n, y_2^n, y_3^n, \dots\}$.

Given the state sequences for the patients $\{y_1^n, y_2^n, y_3^n, \dots\}$, we can split the patient data into K sets of tuples $\{(\boldsymbol{x}_{t-1}^n, \boldsymbol{x}_t^n)\}$ for which $y_{t-1}^n = k$. For each of these sets, we train the parameters $\theta_{k,p}$ of P classifiers to predict $\boldsymbol{x}_t^n(p)$ given $\boldsymbol{x}_{t-1}^n$. The training process will depend on the particular choice of classifier; classifiers such as Naive Bayes classifiers and random forests have standard implementations in many machine learning libraries. Importantly, these classifiers can be used to produce the probability of any particular output $\boldsymbol{x}_t^n(p)$.

Given the transition parameters $\{\theta_{p,k}\}$, we can update the state sequences using the standard forward-backward algorithm for HMMs (Rabiner and Juang 1986). We use a variant called forward-filtering backward sampling in which we first recursively compute the probabilities of each state y_t^n given

the data $\{\boldsymbol{x}_1^n, \boldsymbol{x}_2^n, \ldots, \boldsymbol{x}_t^n\}$ up to time t:

$$Pr(y_t^n | \{\boldsymbol{x}_1^n, \boldsymbol{x}_2^n, \ldots, \boldsymbol{x}_t^n\}, \{y_1^n, y_2^n, \ldots, y_{t-1}^n\}, \{\theta_{p,k}\}) \propto T_y(y_t^n | y_{t-1}^n) \prod_{p=1}^{P} T_x(\boldsymbol{x}_t^n(p) | \boldsymbol{x}_{t-1}^n, \theta_{p,y_{t-1}^n}) \quad Pr(y_{t-1}^n | \{\boldsymbol{x}_1^n, \boldsymbol{x}_2^n, \ldots, \boldsymbol{x}_{t-1}^n\}, \{y_1^n, y_2^n, \ldots, y_{t-2}^n\}, \{\theta_{p,k}\}) \quad (1)$$

and then sampling each state y_t^n from:

$$y_t^n \sim T_y(y_t^n | y_{t+1}^n)$$
$$Pr(\quad y_t^n \quad | \{\boldsymbol{x}_1^n, \boldsymbol{x}_2^n, \ldots, \boldsymbol{x}_t^n\}, \{y_1^n, y_2^n, \ldots, y_{t-1}^n\}, \{\theta_{p,k}\})$$

where the final state $y_{T_n}^n$ is simply sampled from equation 1.

3 Application to Predicting Vasopressor Administration and Weaning

3.1 Setting

Data The MIMIC II 2.6 database (Saeed et al. 2011) includes retrospective electronic medical records (EMRs) for 26,870 adult hospital admissions recorded between 2001 and 2008. Data was gathered from four ICUs at the Beth Israel Deaconess Medical Center: medical, surgical, coronary care unit, and cardiac surgery recovery unit.

Many ICU patients have a limited chance of survival, regardless of clinical intervention. Therefore, our cohort contains only those patients alive 30 days post-discharge and without orders for reduced care (e.g., "comfort measures only," "do not resuscitate," "do not intubate," or "CPR not indicated." Further, we focused on patients with between 12 and 96 hours of data. These criteria focus modeling efforts on situations in which clinical decisions about when to wean the patient should have a positive effect, rather than penalizing a classifier for the many situations in which a patient is taken off of life support. This results in 15,695 patients, 4,331 of which were administered vasopressors.

For these patients, we extracted the 10 most commonly reported static and dynamic variables (such as vital signs and lab results) as well as variables corresponding to vasopressor administration. Variables were standardized using the mean μ_p and standard deviation σ_p from the training set, and the resulting Z-scores were then zero-rounded to integer values in the range of -4:4. This procedure resulted in each of the vital and lab variables taking on 10 discrete values. Vasopressor administration variables were post-processed to recover continuous segments of administration and non-administration; the details of variable processing and extraction, as well as basic population statistics, are given in the supplementary materials.

Evaluation, Baselines, and Model Settings For each run, 30% of the patients were randomly held out for testing. To handle class imbalance, we downsampled negatives in the training set to achieve 30% positives. In both the training and test sets, for each time t, we considered windows of the previous 4 hours of data $\boldsymbol{X}_n[:, t-4 : t]$, where $\boldsymbol{X}_n[:, t] \in [-4, -3, \ldots 3, 4]$ contains the vector of 10 z-scored characters for hour t of patient n. Taken along with nine static admission variables (e.g., age, gender, etc.), this created a flattened vector of 49 "Raw" features.

Four baseline classifiers were trained: a linear-kernel support vector machine (SVMs), Naive Bayes (NB), and Logistic Regression (LR), and long short-term memories (LSTM). Standard packages and settings (Pedregosa et al. 2011) were used for the SVMs, NB, and LR classifiers; for the LSTM we used an existing implementation (Maclaurin and Duvenaud 2015) with 50 hidden units and 100 training iterations.

For the SSAM, we used 5 hidden states. To indicate a preference for staying in the same physiological state, we use a diagonal transition matrix $T(y, y')$ with $T(y, y) = 0.8$ and the remaining entries $T(y, y') = 0.05$ for $y \neq y'$. For the autoregressive models, we considered random forests (with 10 trees) and a Naive Bayes classifier. Inference was run for 45 iterations, starting with a random assignment of states to $\{y_t^n\}$. Tempering was used to used to avoid local optima. Once the model was learned, we used the filtered probabilities of each state $Pr(y_{t-3}^n) \ldots Pr(y_t^n)$ (computed from equation 1) from the last four hours as features for the downstream classifiers.

3.2 Predicting Vasopressor Administration Improved by SSAM Features

We examine the impact of the derived SSAM features as compared to standard physiological features in two prediction tasks, Vasopressor Administration and Vasopressor Weaning. For Vasopressor Administration, we examine an *ungapped* and *gapped* prediction setting, defined as:

1. Ungapped Vasopressor Administration: Predicting vasopressor administration within the next F hours: we make predictions for $t \in \{1, \ldots, t_n^{v_1} - F\}$ (or until the end of stay). However, now the outcome is labeled as positive only if there is vasopressor start in the next F hours.
2. Gapped Vasopressor Administration: Predicting vasopressor administration during F hours after a G hour gap. We make predictions for $t \in \{1, \ldots, t_n^{v_1} - F\}$ (or until the end of stay). Outcomes are labeled as positive only if the vaspressor starts in first F hours after a G hour gap.

We present *ungapped* results because it seems reasonable from a learning perspective to allow models to consider all data after the prediction window when assigning a label. However, if there was a need for an advanced "warning" prior to staffing or procedure decisions, *gapped* prediction would be a valuable option for actionable advice.

Table 1 compares the performance of baseline and autoregressive models for gapped and ungapped vasopressor administration, averaged over five repetitions. The first three classifiers show the baseline approach, which directly classifies using the concatenated feature vectors. The second group of classifiers uses only the probabilities of the states to make predictions; we see that these often approach the baseline performance. For example, the autoregressive model using Naive Bayes for word prediction is between 0.01 and 0.02 AUC of the baseline.

These results suggest that the latent features cluster the patients effectively into differentiable groups. The final group combines features from both the raw data and the switching-state autoregressive models. The combined approaches consistently improve 0.03 to 0.05 AUC above with

the baselines, with the Naive Bayes SSAM and logistic regression classifier having the best performance. While such performance gains may seem small, an increased of 0.05 could affect the treatment of thousands of patients annually in a large ICU.

Classifier Type	Ungapped (AUC)	Gapped (AUC)
Raw with LSTM	0.56 ($\pm$ 0.0043)	0.56 ($\pm$ 0.0058)
Raw with NB	0.87 ($\pm$ 0.0000)	0.82 ($\pm$ 0.0040)
Raw with SVM	0.88 ($\pm$ 0.0001)	0.83 ($\pm$ 0.0039)
Raw with LR	**0.89 ($\pm$ 1.1e-16)**	**0.83 ($\pm$ 0.0040)**
SSAM (RF) with SVM	0.61 ($\pm$ 0.1034)	0.60 ($\pm$ 0.0495)
SSAM (RF) with NB	0.76 ($\pm$ 0.0529)	0.65 ($\pm$ 0.0049)
SSAM (RF) with LR	0.81 ($\pm$ 0.0584)	0.66 ($\pm$ 0.0046)
SSAM (NB) with SVM	0.76 ($\pm$ 0.1104)	0.66 ($\pm$ 0.0747)
SSAM (NB) with LR	0.87 ($\pm$ 0.0073)	0.81 ($\pm$ 0.0124)
SSAM (NB) with NB	**0.87 ($\pm$ 0.0090)**	**0.83 ($\pm$ 0.0076)**
Raw+SSAM (RF) with SVM	0.69 ($\pm$ 0.0784)	0.76 ($\pm$ 0.0532)
Raw+SSAM (RF) with NB	0.88 ($\pm$ 0.0010)	0.82 ($\pm$ 0.0027)
Raw+SSAM (RF) with LR	0.92 ($\pm$ 0.0008)	0.86 ($\pm$ 0.0032)
Raw+SSAM (NB) with SVM	0.87 ($\pm$ 0.0565)	0.73 ($\pm$ 0.1083)
Raw+SSAM (NB) with NB	0.89 ($\pm$ 0.0011)	0.85 ($\pm$ 0.0052)
Raw+SSAM (NB) with LR	**0.92 ($\pm$ 0.0016)**	**0.88 ($\pm$ 0.0061)**

Table 1: AUC scores averaged over 5 repeated trials. SVM is linear kernel Support Vector Machine, NB is Gaussian Naive Bayes, and LR is Logistic Regression with L2 Norm. SSAM(NB) is the switching-state autoregressive model with a Naive Bayes predictor, and SSAM(RF) is the switching state autoregressive model with a random forest predictor.

3.3 Estimating Unnecessary Intervention Time and Predicting Vasopressor Weaning

In general, we note that predict weaning success is harder that predicting intervention onset for two key reasons. First, patient weaning is noted in multiple information sources, and thus may be unintentionally left out of the numerical flags we examine for vasopressor usage. Second, patients on a vasopressor may be physiologically ready to wean from intervention, but the vasopressor will remain for hours or days after due to staffing considerations, clinical judgment, or lack of familial support. If we were to use the previous *gapped* setting, this would create a large number of "fake" negative labels, whereby patients are predicted as successes because they could have been weaned earlier than they were.

For Vasopressor Weaning, we used the *ungapped* setting, and predict weaning success within the next F hours. As before, we consider $t \in \{t_n^{v_1}, \ldots, (t_n^{w_1} - F)\}$. However, now a successful wean is defined to be that no vasopressor administration was needed in the next F hours; this measure allows for the fact that a patient may be stable enough to "successfully" wean from a vasopressor, but requires another vasopressor administration some time later.

Quantitative Results: Predicting Successful Weaning Following the best results from Section 3.2, we trained three classifiers to predict successful weaning. The raw features obtained an AUC of 0.67 ($\pm$ 0.008), SSAM(NB) features were AUC 0.63 ($\pm$ 0.021), and Raw+SSAM(NB) features were AUC 0.71 ($\pm$ 0.005). While other work (Hug and Szolovits 2009) achieved higher AUCs (0.82), that work required the manual engineering of 438 features specific to the weaning task; we achieve very reasonable AUCs with only 10 phsyiological variables.

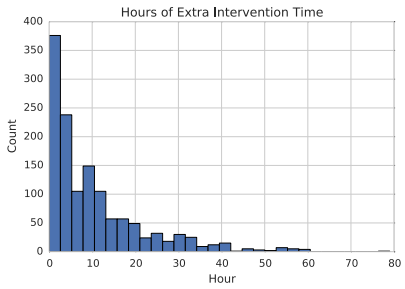

Figure 2: Histogram of excess time for which patients could have been successfully weaned according to the classifier.

Examining SSAM Probabilities for Early Weaning Our quantitative results above discriminate situations in which the clinician may have attempted to wean too early, causing the wean to be unsuccessful. However, clinicians report that patients are often left on interventions for much longer than necessary. Figure 2 shows the difference in time between predicted weaning times and actual weaning times. While a significant portion of patients were successfully weaned at the right time, the heavy-tail depicted suggests that many patients suffered from extended interventions.

A closer examination at a handful of patient examples from this histogram and match the predictions with information gathered from medical notes. We show three examples with a binarized predicted probability of a successful wean for that patient over time, and a figure showing the states that accompany each of those vectors at each time.

In Figure 3, a 72 year old man with coronary artery disease who was put on mechanical ventilation and pressors while in the ICU. The probability of a successful wean is low while the patient fails mechanical ventilation weaning early on in his stay, and immediately post-extubation. It's explicitly noted in his record at the lowest probability of wean that the patient is dependent on the vasopressors he is receiving. The patient stabilizes as the probability of wean success climbs, and the clinical staff actually begin to wean the patient near the highest predicted success in our estimates.

In Figure 4, a 62 year old male patient with a cardiac catheterization. The probability of successful wean remains low while patient is given a course of treatment and fluids, but struggles with a low central venous pressure (CVP) and increasing hematocrit (HCT). When the nursing staff notes an increasing need for vasopressors, the corresponding probability of a wean dips further. During recovery, our model's improved wean success matches the nurse's note that the pa-

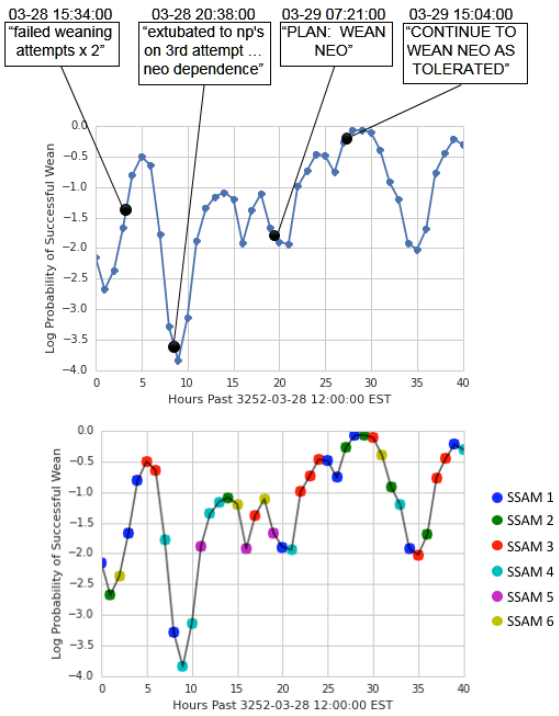

Figure 3: Probabilities of successful weaning and state over time for patient 10387.

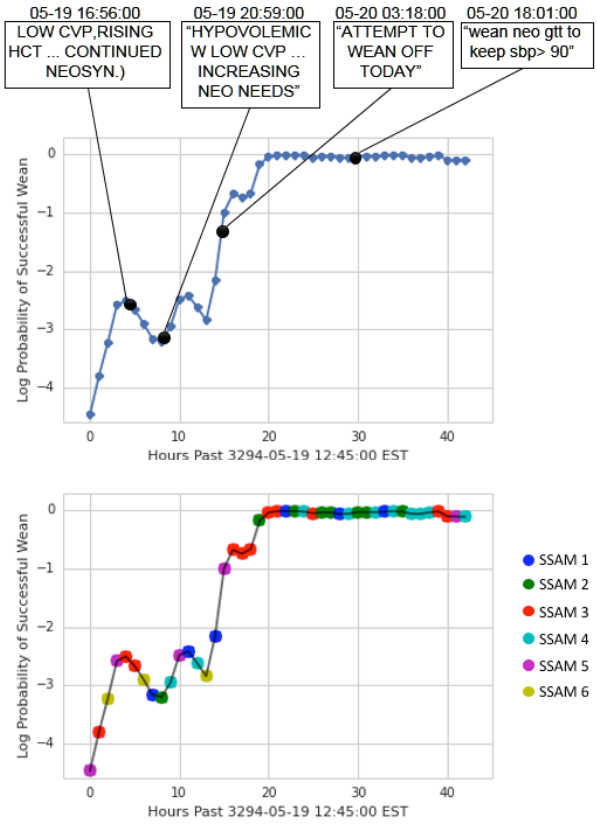

Figure 4: Probabilities of successful weaning and state over time for patient 11315.

tient should be weaned in the following day. In this case, the wean happens almost 10 hours after our model predicts that it could successfully have been done. However, this is likely due to clinical staff schedules, which vary widely in the ICU. For legal and ethical reasons, there is also a bias to maintain interventions in ICU patients rather than withdraw too early, even if a patient seems to be stable.

Finally we show a 65 year old man in Figure 5 who underwent a mitral valve replacement and coronary artery bypass graft. The patient tolerated the surgery well and was transferred to the cardiac surgery recovery unit for monitoring, where he maintained a stable condition. Based on the numerics data available, we were unable to find any indication that the patient was weaned, and thus we labeled this patient as an unsuccessful wean. However, the clinical notes indicate that the patient was successfully weaned from sedation within a the same day of his operation. In this case, we correctly predicted that the patient could be successfully weaned prior to his weaning, despite an incorrectly labeled example.

3.4 Clinical Relevance in Discovered States

We demonstrated that the SSAM features increased AUC in vasopressor administration and weaning prediction tasks. We theorize that this quantitative evidence is due to physiological models that are capturing physiological characteristics that are 1) relevant to interventions and intervention outcomes, but 2) not captured by raw physiological variables.

We first noted that certain states are associated with high and low probabilities in weaning prediction, e.g., as shown in patterns in Figures 3, 4, and 5).

We generalize these observation by counting the frequency with which any particular model was associated with correctly predicting successful or unsuccessful weans. Specifically, we looked at which SSAM generated the highest 1% of successful wean probabilities in the patients that were successful weans, and which SSAM generated the lowest 1% of successful wean probabilities in the patients that were unsuccessful weans.

As shown in Figure 6, we see an increased membership in SSAM 5 and 6 in those patients that had a high probability of a successful wean. On the other hand, data with a low probability of successful weaning in those patients who were not successful weans came more often from SSAM 1 and 3.

We look specifically at the physiological variables that correspond to these states by examining the transitions probabilities for variable values in SSAM 3 and 5. There are several interesting differences in these probabilities. In SSAM 5 transition probabilities for blood hematocrit values tended to stabilize from large abnormal values towards normalcy more often (8% vs. 5%). This could be indicative of patients who healthy enough to remove fluid resuscitation, so their hematocrit is responding with decreased blood viscosity. In SSAM 3 we observed that the respiration rate tends to stabi-

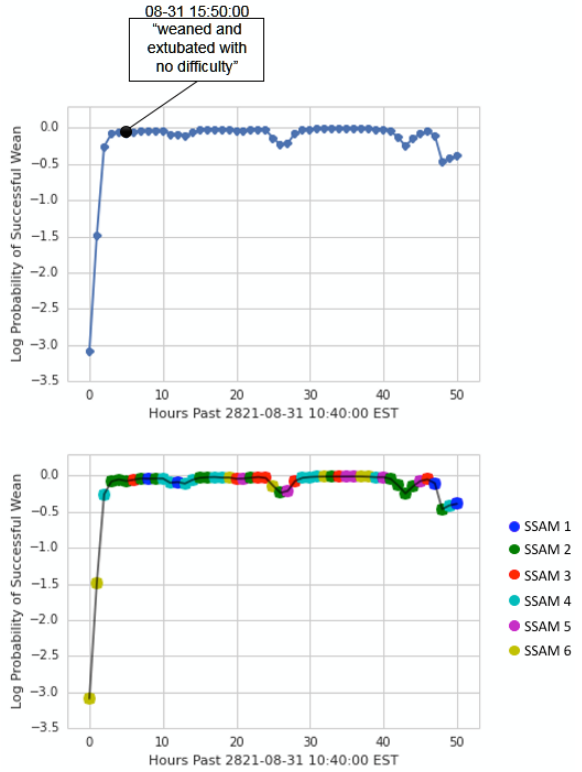

Figure 5: Probabilities of successful weaning and state over time for patient 3194.

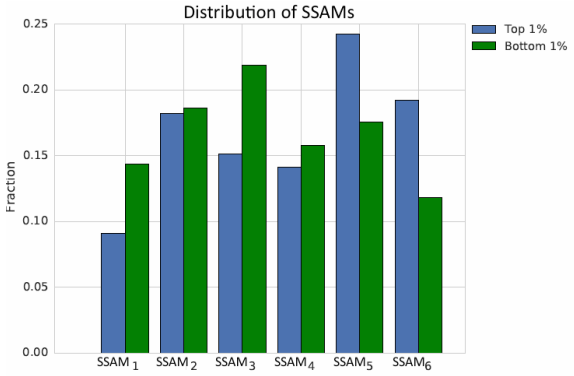

Figure 6: Histograms of the states across patients at time points of high (left) and low (right) probabilities of successful weans.

lize from low values towards normalcy more often (13% vs. 11%). This could indicate that SSAM 3 represents patients who eventually require some form of mechanical ventilation, which can cause more unsuccessful weaning patterns.

4 Discussion and Related Work

Much literature in clinical prediction has focused on aggregating manually-defined features to predict the risk of clinically significant events. For example, (Joshi and Szolovits 2012) aggregated physiological signals into organ-system focused clusters, and then aggregated these clusters to estimate patient physiological state in the ICU.

Switching dynamical systems models (Lehman et al. 2015; Quinn et al. 2009) have been used to impute signals, identify artifacts, discover physiological states in a variety of critical settings. Most of these works have focused on developing models for densely sampled, often one-dimensional data. Our work is distinct in that we consider higher dimensional data, and use discretization and binning to find relevant signals over longer time scales. (Other work has applied unsupervised method to discretized timeseries to discover anomalies (Syed and Guttag 2011) and patient similarities (Saeed and Mark 2006), but in a non-hierarchical fashion.) As noted in (Lin et al. 2007), timeseries symbolization creates many opportunities to analyze physiological data with the rich literature of techniques developed for discrete sequences; our data processing approach also makes it natural for us to consider rich, nonlinear transition models, such as random forests, rather than the linear dynamical systems approaches of the work above.

To predict vasopressor administration, (Fialho et al. 2013) reported an AUC of 0.79 ± 0.02, compared to 0.88 ± 0.0061 from our general hospital model using SSAMs. To our knowledge, we have the highest reported results for predicting vasopressor administration. (Hug and Szolovits 2009) selected 32 variables from a manually defined set over 438 clinically-guided features to predict successful weaning by 1-12 hours (AUC = 0.81, 0.82) and by 6-12 hours (AUC = 0.76, 0.83), where the second AUC represents looking only at patients who surved their hospital stay. While our AUCs are lower (0.71 ± 0.005), our approach did not use the large set of hand-engineered features; seeing whether our unsupervised physiological features improve prediction accuracy when combined with these engineered features is an interesting future direction.

5 Conclusions and Future Directions

In this work, we used a fully unsupervised switching-state autoregressive models to model the evolution of symbolized physiological timeseries. Using features from our SSAMs, we achieved state-of-the-art results in predicting whether a patient would receive a vasopressor administration; our model also discovered several features associated with successful weaning from vasopressors. There are many exciting directions for future work. By modeling the changing physiological state of the patient—rather than creating sets of hand-engineered features—we can take steps towards a unified approach to stratify risk for a variety of interventions and intermediate outcomes (such as mechanical ventillation or sepsis) , characterize the effects of dosages (such as (De Backer et al. 2010)) and multiple interventions. Such unified approach will also advance our understanding of patient physiology in critical care settings.

References

Caballero Barajas, K. L., and Akella, R. 2015. Dynamically modeling patient's health state from electronic medical

records: A time series approach. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, 69–78. New York, NY, USA: ACM.

Che, Z.; Kale, D.; Li, W.; Bahadori, M. T.; and Liu, Y. 2015. Deep computational phenotyping. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, 507–516. New York, NY, USA: ACM.

De Backer, D.; Biston, P.; Devriendt, J.; Madl, C.; Chochrad, D.; Aldecoa, C.; Brasseur, A.; Defrance, P.; Gottignies, P.; and Vincent, J.-L. 2010. Comparison of dopamine and norepinephrine in the treatment of shock. *New England Journal of Medicine* 362(9):779–789.

DAragon, F.; Belley-Cote, E. P.; Meade, M. O.; Lauzier, F.; Adhikari, N. K.; Briel, M.; Lalu, M.; Kanji, S.; Asfar, P.; Turgeon, A. F.; et al. 2015. Blood pressure targets for vasopressor therapy: A systematic review. *Shock* 43(6):530–539.

Fialho, A.; Celi, L.; Cismondi, F.; Vieira, S.; Reti, S.; Sousa, J.; Finkelstein, S.; et al. 2013. Disease-based modeling to predict fluid response in intensive care units. *Methods Inf Med* 52(6):494–502.

Ghassemi, M.; Naumann, T.; Doshi-Velez, F.; Brimmer, N.; Joshi, R.; Rumshisky, A.; and Szolovits, P. 2014. Unfolding physiological state: Mortality modelling in intensive care units. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 75–84. ACM.

Ghassemi, M.; Pimentel, M. A.; Naumann, T.; Brennan, T.; Clifton, D. A.; Szolovits, P.; and Feng, M. 2015. A multivariate timeseries modeling approach to severity of illness assessment and forecasting in icu with sparse, heterogeneous clinical data. In *Proc. Twenty-Ninth AAAI Conf. on Artificial Intelligence*.

Hug, C. W., and Szolovits, P. 2009. Icu acuity: real-time models versus daily models. In *AMIA Annual Symposium Proceedings*, volume 2009, 260. American Medical Informatics Association.

Janssen, K. J.; Donders, A. R. T.; Harrell, F. E.; Vergouwe, Y.; Chen, Q.; Grobbee, D. E.; and Moons, K. G. 2010. Missing covariate data in medical research: to impute is better than to ignore. *Journal of clinical epidemiology* 63(7):721–727.

Joshi, R., and Szolovits, P. 2012. Prognostic physiology: modeling patient severity in intensive care units using radial domain folding. In *AMIA Annual Symposium Proceedings*, volume 2012, 1276. American Medical Informatics Association.

Lasko, T. A.; Denny, J. C.; and Levy, M. A. 2013. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PloS one* 8(6):e66341.

Lee, C. H.; Arzeno, N. M.; Ho, J. C.; Vikalo, H.; and Ghosh, J. 2012. An imputation-enhanced algorithm for icu mortality prediction. In *Computing in Cardiology (CinC), 2012*, 253–256. IEEE.

Lehman, L.; Adams, R.; Mayaud, L.; Moody, G.; Malhotra, A.; Mark, R.; and Nemati, S. 2015. A physiological time series dynamics-based approach to patient monitoring and outcome prediction. *IEEE journal of biomedical and health informatics* 19(3):1068.

Lin, J.; Keogh, E.; Wei, L.; and Lonardi, S. 2007. Experiencing sax: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery* 15(2):107–144.

Liu, P.; Lei, L.; Yin, J.; Zhang, W.; Naijun, W.; and El-Darzi, E. 2006. Healthcare data mining: predicting inpatient length of stay.

Maclaurin, D., and Duvenaud, D. 2015. Autograd: Reverse-mode differentiation of native python.

Marlin, B. M.; Kale, D. C.; Khemani, R. G.; and Wetzel, R. C. 2012. Unsupervised pattern discovery in electronic health care data using probabilistic clustering models. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 389–398. ACM.

Marshall, A.; Altman, D. G.; Royston, P.; and Holder, R. L. 2010. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC medical research methodology* 10(1):7.

Müllner, M.; Urbanek, B.; Havel, C.; Losert, H.; Gamper, G.; and Herkner, H. 2004. Vasopressors for shock. *The Cochrane Library*.

Ospina-Tascón, G. A.; Büchele, G. L.; and Vincent, J.-L. 2008. Multicenter, randomized, controlled trials evaluating mortality in intensive care: Doomed to fail? *Critical care medicine* 36(4):1311–1322.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Quinn, J.; Williams, C. K.; McIntosh, N.; et al. 2009. Factorial switching linear dynamical systems applied to physiological condition monitoring. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31(9):1537–1551.

Rabiner, L. R., and Juang, B.-H. 1986. An introduction to hidden markov models. *ASSP Magazine, IEEE* 3(1):4–16.

Saeed, M., and Mark, R. 2006. A novel method for the efficient retrieval of similar multiparameter physiologic time series using wavelet-based symbolic representations. In *AMIA Annual Symposium Proceedings*, volume 2006, 679. American Medical Informatics Association.

Saeed, M.; Villarroel, M.; Reisner, A. T.; Clifford, G.; Lehman, L.-W.; Moody, G.; Heldt, T.; Kyaw, T. H.; Moody, B.; and Mark, R. G. 2011. Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database. *Critical care medicine* 39(5):952.

Syed, Z., and Guttag, J. V. 2011. Unsupervised similarity-based risk stratification for cardiovascular events using long-term time-series data. *Journal of Machine Learning Research* 12:999–1024.

Vincent, J.-L., and Singer, M. 2010. Critical care: advances and future perspectives. *The Lancet* 376(9749):1354–1361.

Vincent, J.-L. 2013. Critical care-where have we been and where are we going? *Critical Care* 17(Suppl 1):S2.