Prognostic Physiology: Modeling Patient Severity in Intensive Care Units Using Radial Domain Folding

Rohit Joshi, PhD, Peter Szolovits PhD
Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA

Abstract

Real-time scalable predictive algorithms that can mine big health data as the care is happening can become the new "medical tests" in critical care. This work describes a new unsupervised learning approach, radial domain folding, to scale and summarize the enormous amount of data collected and to visualize the degradations or improvements in multiple organ systems in real time. Our proposed system is based on learning multi-layer lower dimensional abstractions from routinely generated patient data in modern Intensive Care Units (ICUs), and is dramatically different from most of the current work being done in ICU data mining that rely on building supervised predictive models using commonly measured clinical observations. We demonstrate that our system discovers abstract patient states that summarize a patient's physiology. Further, we show that a logistic regression model trained exclusively on our learned layer outperforms a customized SAPS II score on the mortality prediction task.

Introduction

Early recognition of clinical deterioration is an important problem because seventy percent of adverse events, which occur in about 16% [2] of hospital admissions, are preventable. The signs of clinical instability often precede an actual cardiac arrest or an unexpected critical event by a mean of 6.5 hours [4]. Buist et al. [4] estimates that early recognition of decline in a patient's baseline condition leads to a 50% reduction in the occurrence of cardiac arrest in general hospital wards, resulting in a decrease in overall hospital mortality. Though predictive scoring systems are gaining popularity in critical care, currently they are used in a very limited way, typically only to support staffing and census predictions even though data mining has been applied to ICU medical data for over two decades [6,15,16, 22]. Some examples of ICU predictive scoring systems include: the SAPS II [14] and APACHE [12] scores as mortality predictors, and the sequential organ failure assessment (SOFA) [1] and multi organ dysfunction score (MODS) [1] for organ failure prediction.

Most research in ICU data mining can be best classified as *smart computation approaches with sparse sensing assumptions*. For example, most research initiatives rely on building supervised machine learning models under an assumption of a resource-limited ICU environment and aim to select the fewest and best commonly measured clinical predictors for a particular outcome [1,12,14,16]. Work in multivariate unsupervised learning has predominantly used commonly measured signals such as the heart rate or the oxygenation level to generate clusters of similar patient states [19]. Scaling to consider many more clinical variables is not only computationally challenging, but it also becomes hard to define an appropriate number of informative physiological clusters without the advice of practicing physicians [7]. Not surprisingly, there has been a dearth of published literature on high-dimensional multivariate unsupervised learning in Critical Care [3]. In this work, we assume an opposite scenario: a modern ICU with massive health data collection facilities, with a need for a scalable data analytic framework for evidence-based medicine.

We propose to use unsupervised learning in a very different way, which we believe can dramatically improve the way physicians visualize patients' evolving clinical states. We introduce a novel clustering algorithm, radial domain folding, which learns lower dimensional abstractions in an organ-specific manner from routinely generated patient data. Then, we train a predictive model on our unsupervised feature layer to recognize and track critical conditions in ICUs in real-time. Our method takes advantage of the clinical knowledge that detailed measurements of sets of parameters are most useful to provide insight into the functioning of specific organ systems, but that overall patient mortality is best predicted by learning to aggregate the patterns of abnormalities in individual organ systems. Our proposed system differs from existing approaches to predictive modeling in two main ways:

1. **We remove the feature selection step at the level of clinical observations:** Feature selection is at the heart of every predictive model. However, when feature selection is performed at the level of clinical observations, the resulting predictive algorithm is constrained to a particular task. As a result, we see different predictors used

in scoring systems for organ severity and for mortality. Consequently, an organ severity prediction task is not technically a sub-task for mortality prediction, even though they share similar foundations and common characteristics of inferring patient severity. For example, it is confusing to see that the SOFA score uses Serum Creatinine for renal severity prediction whereas the SAPS II score does not consider Creatinine levels to be a significant predictor of mortality, but rather considers a patient's Urea levels (BUN) and Urine Output. Further, SAPS II does not consider coagulation parameters such as platelet count for mortality prediction, whereas SOFA includes coagulation factors for calculating overall patient severity. In our system, feature selection is more appropriate on the learned unsupervised feature layer.

2. **We do not perform clustering in a traditional uniform way**: Theoretically, it can be easily shown that clustering using traditional methods over high-dimensional big data is a computationally hard problem. Kshetri [11] demonstrated empirically that standard k-means (probably the most efficient clustering algorithm), without parallelism, fails on approximately 50,000 data instances in R [20] on a 192GB machine in a high dimensional MIMIC II clinical dataset [18]. To scale to a million data instances, Kshetri's greedy algorithm [11] uses chunks of 30-40,000 rows of matrices iteratively. Moreover, the standard dimensionality reduction techniques, such as principal component analysis, do not work with incomplete datasets in which some values are missing. Our proposed clustering approach groups similar ICU patients based on abnormalities in specific organ systems and scales up to millions of patients' data instances.

Big Data: A Challenge to Clinical Decision Making

Figure 1 depicts a patient's ICU time course in multiple dimensions from the MIMIC II database [18]. Abbreviations around the periphery show some of the commonly measured clinical variables for understanding a patient's health. The axis (-8, 8) shows the number of standard deviations from the normal range of each parameter (0 being normal). The line width captures time variation. This patient was admitted with heart failure, having a past history of chronic kidney failure. His health parameters at ICU admission are depicted by the red line, the thinnest line. As expected, some of the parameters are missing. The thickest line, pink, shows the health

status of the patient near the discharge time. As evident from Figure 1, it is difficult to visualize, by the values of the individual parameters, if the patient's overall condition actually improved with time. Imagine a million such lines for thousands of patients. *How can an algorithm then create patient profiles by considering varied aspects and varied lengths of thousands of patients' hospital stays to provide individualized predictions in real time?* In contrast to the current patient profiling systems, our system learns complex physiological concepts such as heart states or kidney states in real time from the data. We also hypothesize that organ-severity prediction is a sub-task of mortality prediction.
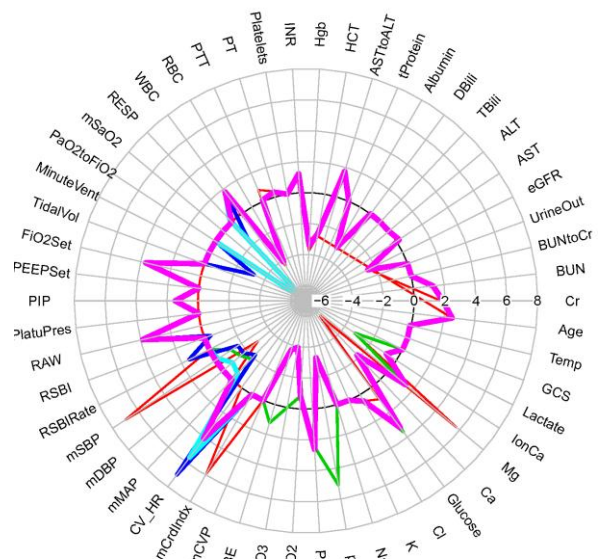

Figure 1: Big data: a scalability challenge to clinical decision making

Radial Domain Folding (RDF)

We present a new multivariate clustering approach, **Radial Domain Folding (RDF),** that generates a layered grouping of patient states. We assume that each patient state (normally, the collection of data about a patient at a particular time) is associated with a set of data elements x_{ij} where i identifies the measurement and j identifies the patient state. E.g., we might have that $x_{24,300}$ is the value of the 24th measured parameter, say the serum sodium, for the 300-th patient state**.** From the medical literature, we know that abnormalities in certain parameters are most closely related to the state of specific organ systems, types of therapy and patient histories (which we call domain foci). Therefore, we also assume that each patient state may be analyzed at two levels:

1. **Focus-specific clustering:** How each focus (organ system, therapy type, or patient history) can be assessed in terms of the directions and magnitudes of the deviations from normal of the data elements that bear on that focus, and

2. **Disease-state clustering:** How the overall patient can be characterized as a function of the abnormalities noted in each focus.

For both the focus-specific and the overall disease-state levels of analysis, we perform two sub-analyses:

a) We abstract the data from the previous layer to represent the direction and magnitude of abnormalities, and

b) We cluster the resulting abstract patient states into a modest number of similar patient states that we believe correspond to different types and degrees of illness.

We implement this method using three distinct layers, described in the sections following.

Layer 0: Abstraction of primary data

We first abstract each data point to a pair of $\langle m_i, d_i \rangle$ where m_i is the scaled magnitude of that point's deviation from normal and d_i is a direction of deviation. For each numerical data item, x_i, we normalize the value to something like a z-score, z', in which all values within the normal range of the variable are normalized to zero. Let $x_{i,L}$ and $x_{i,H}$ be the low and high ends of the normal ranges of variables x_{ij}.

$$z(x_{ij}) = x_{ij} - \mu_i / sd(x_{ij}) \qquad (1)$$

$$z'(x_{ij}) \qquad (2)$$

$$= \begin{cases} 0, & if\ z(x_{i,L}) < z(x_{ij}) <\ z(x_{i,H}) \\ z(x_{ij}) - z(x_{i,H}), & if\ z(x_{ij}) > z(x_{i,H}) \\ z(x_{ij}) - z(x_{i,L}), & if\ z(x_{ij}) < z(x_{i,L}) \end{cases}$$

where μ_i is the mean of the x_{ij} and $sd(x_{ij})$ is the their standard deviation. Magnitude m_{ij} is just $|z'(x_{ij})|$. Direction d_{ij} is defined as *1, 0, -1*, depending on whether $z'(x_{ij})$ is positive, zero or negative, respectively.

Our data also contain qualitative values, such as the reason for ICU admission, aspects of medical history, etc. For now, we have not included these in the severity or direction calculations, though we plan to develop methods for doing so.

Layer 1: Focus-specific clustering

Our second step is to cluster the abstractions of variables that are relevant to a each specific focus, f_k. First, for each focus f_k the subset of measurements that bear on it is given by

$$D(k) = \left\{ i \middle| \begin{matrix} focus\ f_k\ is\ influenced\ by\ x_{ij} \\ for\ each\ patient\ state\ j \end{matrix} \right\} \qquad (3)$$

and is assumed to be given by background medical knowledge.

Step1a: $\langle magnitude, direction \rangle$ abstraction:

We form clusters separately for the magnitudes of abnormalities and for their directions. Unlike in the general case of clustering arbitrary data, in our case we know that the cluster near zero magnitude of abnormality along every component data direction is special — it corresponds to the well patient. Therefore, we compute the sum of squares of the normalized deviations defined by Equation 2 for each patient state as a distance measure from the normal zero magnitude cluster, and then cluster the patient states using hierarchical clustering over this one-dimensional measure

$$M_{kj} = \sum_{i\ \in D(k)} z'(x_{ij})^2 \qquad (4)$$

We perform this clustering very efficiently by sampling only a small fraction of our data to create the clusters and assigning all other patient states to the nearest cluster center in this one-dimensional representation[1]. We order the resulting clusters by mean degree of abnormality, thus associating the clusters with an increasing measure of severity.

The resulting clusters indicate how abnormal the patient is in relation to a particular focus, but this method collapses the specific nature of that abnormality, so that two patient states may share a common M but arrive there by very different abnormalities in the underlying data. For this reason, we compute a second clustering for f_k based on the directions of abnormality of the individual data elements associated with that focus. For these directions, we code the values *1, 0, -1* of each variable using a Jaccard representation, *10, 00, 01*, concatenate the direction representations for all the parameters relevant to focus f_k, and then compute a

---

[1] Although we could use a sophisticated adaptive method to determine the optimal number of clusters, we chose 6, given as an exogenous parameter

hierarchical clustering using a Jaccard score [21] over this representation. This uses a distance function that is the ratio of the number of digits at which two vectors mismatch divided by the sum of those mismatches plus the number of matches at *1* positions. The distance is *1* if no digits match and zero if they all do.

In our domain, the number of unique direction vectors for patient instances tends to be small, so we are able to compute the full distance matrix and apply hierarchical clustering efficiently to create *8* direction clusters[2] using the frequency counts of the unique direction vectors as starting points. This is like starting the weighted hierarchical tree construction in the middle rather than with a singleton set. We order the resulting clusters by their weighted mean distance from the normal direction vector, which is equivalent to counting the number of 1s in the Jaccard representation of directions. This creates a second, different measure of severity, based on the number of data element that are abnormal rather than the total degree of their abnormality.

Step 1b: focus-specific severity:

The result of focus-specific abstraction assigns each patient state to a magnitude cluster that indicates how severely abnormal that focus is, and a direction cluster that indicates the combination of data abnormalities that led to that severity. We have thus created an abstract representation for each focus of each patient state. For each focus of each patient state, we thus obtain an assignment to one of six clusters for severity and one of eight for direction, or 48 total possibilities. We perform hierarchical clustering using a squared Euclidean distance on these focus-specific abstractions, using the 48 unique possibilities and their frequency counts as starting points. We order the resulting clusters by the average severities given by their magnitude and direction input clusters, and thus create an aggregate measure of focus severity.

Layer 2: Disease-state clustering

Step2a: $\langle magnitude, direction \rangle$ abstraction:

We currently use 8 foci, therefore we characterize each patient state by the identity of eight clusters for severity of each focus. To characterize the overall nature of a patient state, we now apply a data abstraction and clustering algorithm similar to that described for Steps 1a and 1b, above.

---

[2] The number of clusters is, as before, a tunable parameter of the method, where we have empirically determined that 8 seems to do well

First, we apply formulas 1 and 2 to the focus-specific severities to again generate abstracted versions of these inputs.

To calculate the aggregate magnitude of disease severity, we take the average magnitudes of the clusters assigned to the magnitudes of each focus as our input data. Thus, for each overall patient disease state, we have eight inputs, being the severities of the focus-specific clusters. We apply formula 4, and again apply hierarchical clustering to find patient disease states that are of similar severities.

To cluster directions of abnormality among the different foci, we use a scheme similar to that used earlier to cluster directions of abnormality within individual foci, but with differences in detail. Because our input variables at the disease state level have no negative values (i.e., one cannot have negative abnormality in any focus), instead of using a *1, 0, -1* scale to represent direction, we determine a "normal or nearly-normal" class, a "somewhat abnormal" class, and a ``highly abnormal'' class, giving us a *0, 1, 2* representation and a Jaccard encoding of *00, 01, 11*. Direction distances are computed using this representation, and we again form direction clusters for the aggregate disease state in a manner similar to what we did for each focus.

Step 2b: disease-state severity:

As was the case for individual foci, this abstraction methods yields an assignment at the overall disease level of each patient state to one of a set of severity clusters aggregated from the severities of the various foci and another aggregated from the directions of the various foci. We then compute an overall patient disease state severity using the method of Step 1b.

Results and Discussion

For the data available on patients in the MIMIC II critical care database [18], we identified a set of foci based on different organ systems, therapy types and patient history. Each clinical variable is assigned to a focus. Table 1 shows a few foci and a few of the clinical variables used in each focus. We studied a previously preprocessed dataset [10] of approximately ten thousand patients with a million chart events from the MIMIC II database. Each patient's ICU stay is represented as multiple chart events or data records at an hourly time interval. Each record is nurse-verified and contains over two hundred clinical parameters. The missing parameter values were filled by repeating the last known value until a reasonable limit assumption, as described in Hug et al [9, 10]. We now show that our algorithm leads to the discovery of abstract patient states that summarize a patient's physiology.

Domain Focus	Clinical Variables
Kidney	Creatinine (Cr), BUN, BUN/Cr, Urine Out/Hr/Kg, eGFR
Liver	Bilirubin, AST, ALT, Albumin, ...
Cardiovascular	MAP, HR, CVP, Cardiac Index, ...
Respiration	RR, SpO2, FiO2, PEEP, PIP, ...
Hematology	Hgb, RBC, WBC, INR, Platelets,...
Electrolytes	Na, K, Mg, Ca, Glucose, ...
Acid-base	PaCO2, pH, CO2, Base Excess, ...
General	GCS, Age, Temp, ...
Medication Type	Diuretic, Antiarrhythmic, Antiplatelet, Sympathomimetic, ...
Chronic	AIDS, Metastatic Carcinoma, Hematologic Malignancy
Location Unit	Surgical, Medicine, Trauma, ...
EKG	Rhythm types, PVC, ...

Table 1: Domain foci

Patient Severity Visualization Using RDF

Complex clinical data can lie in over a 100-dimensional space and, as shown in Figure 1, it is difficult to get an intuitive feel for what the data looks like. Figure 2 shows the learned severity graph in different organ systems (Step 1b) using the *Radial Domain Folding* algorithm during the ICU time course of the patient in Figure 1. The radial axes capture organ severities from 1 (being normal) to 8 (being the worst). The number of severities, 8, is an exogenous parameter indicating the number of clusters of severity to be found for each organ. Different colored lines depict different time points during the patient's ICU stay, similarly to Figure 1. The line width captures time progression.

At the time of admission, shown by the red line, this patient's cardiac state was grouped into a high severity cluster 7 (he had heart failure). His electrolytes and lung status were severe too (elytes cluster 7; lung cluster 6). The rest of the information was missing. The patient's lung status worsened quickly (transition from lung cluster 6 to lung cluster 7). Near the ICU discharge time, the pink line, this patient's cardiovascular, lung and electrolytes status had improved (as shown by the respective transitions from a higher severity state to a lower severity state). The patient's kidney status remained the same through his ICU stay (he had a history of a chronic kidney failure). This progress and improvement in
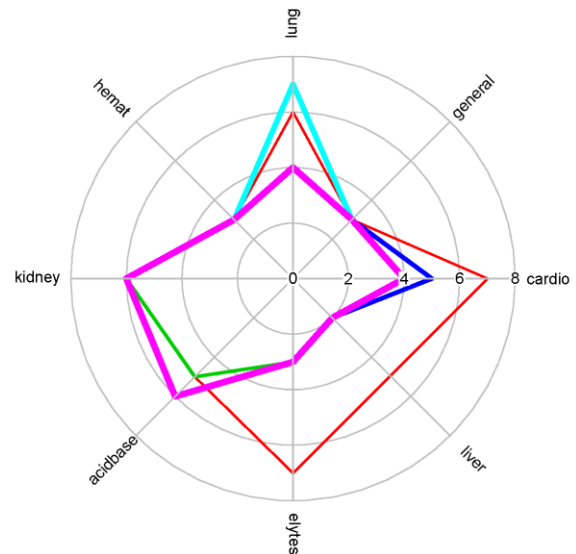

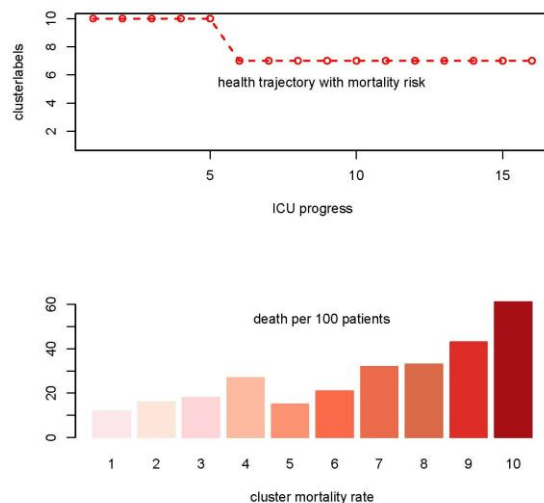
Figure 2: Focus Severity Graph using RDF

Figure 3: Overall Disease Severity using RDF

the patient's health status over time was difficult to visualize in Figure 1.

Figure 3 shows the patient trajectory in terms of the RDF Layer 2 (Step 2b) summarization of overall physiological health and the mortality rates associated with ten learned health clusters, numbered in order of increasing severity. This patient gradually moves from a high mortality rate cluster 10 (at admission) to a lower mortality rate cluster 7 (near the time of discharge). By automatically discovering physiologically meaningful clusters, our algorithm enables a physician to visualize any patient's evolving clinical condition.

Figure 2 and 3 also show that it is relatively simple to define an appropriate number of clusters in our framework. If desired, one can also adopt a complex statistical technique, such as in Kshetri [11], to estimate the appropriate number of clusters mathematically. Further, in comparison to Kshetri's greedy k-mediods approach [11] that took several hours to cluster only 40,000 data instances from our dataset on a 192GB machine, our new system is fast. The runtime depends on the focus. For example, RDF clustering on the kidney focus with over a million data instances took about 90 seconds in a non-parallel R-based implementation on the same machine.

Theoretically, our algorithm clusters a patient's health status in *sub-linear time* [17]. In the context of machine learning, our algorithm takes an approach similar to manifold learning [5] and our algorithm exploits computational advantages resulting from transforming each individual focus locally into a Euclidean plane, a 2-D manifold. Our algorithm is fast because our low-dimensional manifold representation can be learned extremely efficiently and consequently, the later steps are also learned fast as only a core set of representative data points, a tiny fraction of the complete clinical set, is good enough to compute approximate patient clusters. In contrast, standard clustering algorithms, such as k-means, hierarchical clustering, and non-parametric Bayesian clustering [8] are extremely slow on large high-dimensional clinical dataset because their complexity increases tremendously with an increase in the size of input data. Simply sampling few points to speed up standard clustering algorithm is not a good approach because there are no guarantees that the samples represent the entire space. Finding a "core" set of representative points using a low-rank matrix projection is a challenging problem that can potentially give sub-linear time speed ups [17] to a clustering problem. Our algorithm presents one such approach.

Further, the empirical results in Kshetri [11] suggest that using clustering algorithms over all the features is also a naïve approach in terms of visualizing big high-dimensional clinical data. For example, clustering all features is equivalent to learning directly our Step 2b of RDF Layer 2, the overall health status graph, as shown in Figure 3. In contrast, our approach offers finer granularity visualizations of underlying organ-severities and their temporal transitions. Our algorithm, being fast, can both pre-compute clusters and conduct clustering when patient data arrive in real time.

Real-time Mortality Prediction using RDF Layers

In the context of machine learning, our learning method is an unsupervised feature learning approach and can also be viewed as a preliminary feature extraction step, after which pattern recognition algorithms are applied. Therefore, we investigate the potential clinical value of our new abstractions by studying the performance of logistic regression (LR) classifiers built from these values compared to more traditional classifiers built from the original clinical data. We selected Logistic regression (LR) because we wanted to make a fair comparison between the LR-based gold standard SAPS-II model often used in ICUs and our learned lower-dimensional RDF layer; and to evaluate that the performance gain is due to the better informative features learned and not due to the difference in classifier. Further, Hug, in this PhD work [10], trained several models and demonstrated that LR, though considered not a highly sophisticated classifier, gives state-of-the-art results on the same dataset. We compare LR trained on: a) the degree and direction abnormalities of different foci (RDF Layer 1:Step1a) together with qualitative information, such as ICU service location unit and past chronic diseases; b) the severities of the various foci (RDF Layer 1:Step 1b) and the same qualitative information; c) the 50 best clinical features selected using an information gain method [13] from over 200 features using a feature selection method; d) the SAPS II features; and e) a customized SAPS-II gold standard model (SAPS II score by approximating the score of the missing "type of admission" field using the "service location unit"[9, 10]). We hypothesize that incorporating *heart is failing* should have a better discriminatory power than simply knowing that *blood pressure is low*.

We followed a performance evaluation strategy similar to that described in Hug et al [9]. The data were divided into training and test data using a 70/30 split strategy with ~12% of expired patients in both training and test data. We performed five-fold cross-validation and repeated the evaluation five times. Figure 5 shows that both our new classifiers achieve surprisingly high Area under the ROC Curve (AUC) of 0.89. Hug et al [9, 10] has earlier shown that their best classifier achieves an AUC of 0.87, while an approximation to the SAPS II model achieves an AUC of 0.81 on the same dataset. Our algorithm achieves a similar AUC to that of Hug's best classifier, but without incorporating specialized predictive variables representing summaries of an observation over time. We believe that the main advantage in our type of approach is that these results show that our layered representation can directly be
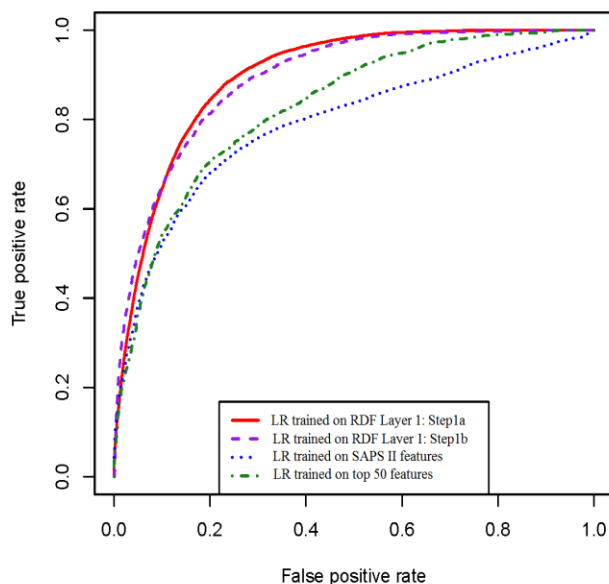
Figure 4: Mortality Prediction Comparison

Figure 5: Mortality Prediction Comparison on High Severity Patients

used as an intermediate representation to create dynamic models of patient state transitions to predict impending adverse events or to forecast the course of disease progression given an intervention. We pursue this as future work.

Mortality Prediction: High Severity Patients

The patient in Figure 3 was admitted with high severity but recovered within a few days of ICU treatment. The current predictive models, such as SAPS II, are trained on admission data and are agnostic to ICU treatment strategies. Understanding real-time mortality risk in high severity patients is important to infer patients' responses to therapies and treatments.

To assess performance on the high severity group, we sorted all ten thousand patients into decreasing order based on their day one pseudo-SAPS II scores. Then, we selected the two thousand highest scoring patients. These patients had a minimum pseudo-SAPS II score of 50. The mortality rate was about 22% in this high severity group (nearly double that of the whole sample). We evaluated the performance of two of the above classifiers using a similar strategy as above. We compared LR trained on the RDF Layer 1 (Step 1b), the focus severities, with that of an approximate SAPS II model, which follows the Hug et al. [9] strategy to replace the missing "type of admission" field by the location unit indicators.

Figure 5 shows that the approximate SAPS II model achieved an AUC of 0.77. In comparison, LR trained on focus severities (RDF Layer 1 Step 1b) achieved an AUC of 0.91.
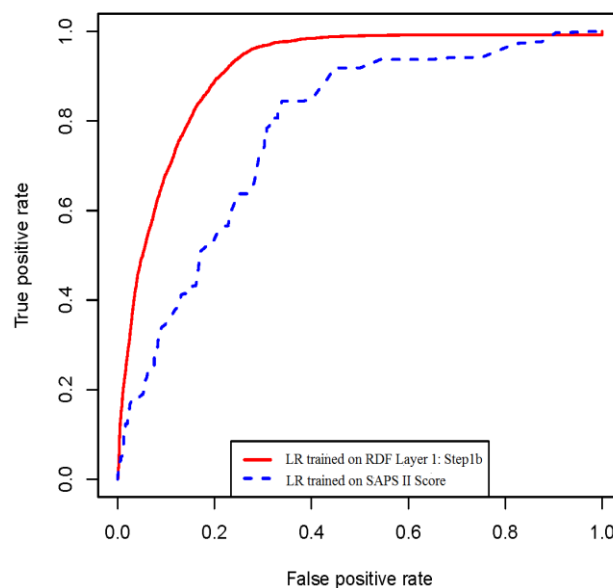
Limitations

To model the holistic view of a patient's ICU stay, we had to make certain assumptions in dealing with missing values and mixed data types (categorical and real-valued). For example, we assumed an organ to be normal and the missing observations to be in the normal range if all the relevant clinical variables of an organ were missing. Such strategies are often used in severity of illness scores in order to improve the model's coverage at potential cost to model performance [15]. Further, we separated categorical and real-valued variables into different foci. One extension could be to use a generalized distance metric to overcome this limitation. Another interesting extension could be to automatically allocate features to domain foci through an optimization process on training data. We also observe that different runs of our clustering algorithm produce different clusters because of the sampling introduced in the RDF algorithm. In our experience, this can lead to assignment of slightly different severity scores to cases near cluster boundaries, and to small variations in the mortality statistics shown in Figure 3. This behavior seems innate to sampling methods and generates only small differences in our results.

Conclusion

This work describes a scalable data analytic framework that provides prognostic previews of patients' clinical conditions in real-time. By computing similarities among the patients on the

basis of organ systems abnormalities, we show that it is possible to use a scalable unsupervised learning approach to summarize a patient's physiology in a holistic way.

Our framework exploits the availability of massive data sets using an outcomes-free approach, and consequently it enables a variety of clinical care applications, ranging from health profiling, triage, informed staffing and operational decisions, to real-time therapy selection. Unfortunately, there has not been much work in creating such "richer" representations for better situational awareness of patients' critical conditions in ICUs. We hope our paper will spur more research in this area.

Acknowledgments

References

1. Bota DP, Melot C, Ferreira FL, Ba VN and Vincent JL. The multiple organ dysfunction score (MODS) versus the sequential organ failure assessment (SOFA) score in outcome prediction. *Intensive Care Med*, 28(11):1619–1624, 2002.

2. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med*, 324:370–376, 1991

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

4. Buist MD, Moore GE, Bernard SA, Waxman BP, Anderson JN and Nguyen TV. Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: preliminary study. *Br Med J*, 324:387–390, 2002

5. Cayton L, Algorithms for manifold learning, *University of California, San Diego, Tech. Rep.* CS2008-0923, 2005

6. Chang RW. Individual outcome prediction models for intensive care units. *Lancet*, 2(8655):143–146, 1989

7. Cohen MJ, Grossman AD, Morabito D, Knudson MM, Butte AJ and Manley GT. Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis. *Critical Care*, 14, 2010

8. Heller, KA, Bayesian Hierarchical Clustering, *Intl. Conf. on Machine Learning*, 2005

9. Hug CW and Szolovits P. ICU Acuity: Real-time Models versus Daily Models. *AMIA,* 260-264, 2009

10. Hug C, Detecting Hazardous Intensive Care Patient Episodes Using Real-time Mortality Models, *PhD thesis*, Massachusetts Institute of Technology, 2009

11. Kshetri K. Modeling Patient States in Intensive Care Patients. *Master's thesis*, Massachusetts Institute of Technology, 2011.

12. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*, 100(6):1619–1636, 1991

13. Guyon, I and Elisseeff A, An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research*, 3:1157-1182, 2003

14. Le Gall JR, Lemeshow S, and Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA*, 270(24):2957–2963, 1993

15. Ohno-Machado L, Resnic FS and Matheny ME. Prognosis in critical care. *Annu Rev Biomed Eng*, 8:567–599, 2006

16. Pang BC, Kuralmani V, Joshi R, et al. A hybrid outcome prediction model for severe traumatic brain injury, *Journal of Neurotrauma*, 24(1): 136-146, 2007

17. Rubenfeld R, Sublinear time algorithms, *Proceedings of the International Congress of Mathematicians*, 2006

18. Saeed M, Lieu C, Raber G and Mark RG. MIMIC II: a massive temporal ICU patient database to support research in intelligent patient monitoring. *Comput Cardiol*, 29:641–644, 2002

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

20. R Development Core Team. R: A language and environment for statistical computing, 2011 http://www.R-project.org

21. Xu R and Wunsch D. Survey of Clustering Algorithms, *IEEE Trans. on Neural Networks*, 16(3):645-678, 2005

22. Zhang Y and Szolovits P. Patient-specific learning in real time for adaptive monitoring in critical care. *Journal of Biomedical Informatics*, 41(3):452-460, 2008