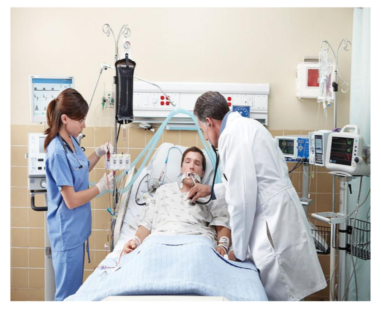
Learning Score Systems for ICU Mortality Prediction via Orthogonal Matching Pursuit

Aadirupa Saha, Chandrahas Dewangan, Harikrishna Narasimhan, Sriram Sampath, Shivani Agarwal

Department of Computer Science and Automation, Indian Institute of Science, Bangalore

ICU Patient Mortality Prediction



Estimating probability of patient survival/death in ICUs

- Monitoring quality of care
- Resource allocation
- Comparing ICUs across demographics
- …

ICU Score systems widely used in the US and Europe for this purpose

St. John's ICU data: A Case Study

St. John's Medical College Hospital, Bangalore, India

ICU patient data (2006-2014)

- 3499 patients
- 29 clinical observations

	AUC
APACHE-II	66%
LODS	63%

APACHE-II Score System

Features / Intervals / Scores

SCORE	4	3	2	1	0	1	2	3	4
Rectal Temp, °C	=41	39.0–40.9		38.5–38.9	36.0–38.4	34.0–35.9	32.0–33.9	30.0–31.9	=29.9
Mean blood pressure, mmHg	=160	130-159	110-129		70-109		50-69		=49
Heart rate	=180	140–179	[illegible]	[illegible]	[illegible]		55–69	40–54	=39
Respiratory rate	=50	35–49	[illegible]	25–34	12–24	10–11	6–9		=5
Arterial pH	=7.70	7.60–7.69		7.50–7.59	7.33–7.49		7.25–7.32	7.15–7.24	<7.15
Oxygenation									
If FIO2 > 0.5, use (A – a) DO2	=500	350–499	200–349		<200				
If FIO2 0.5, use PaO2					>70	61–70		55–60	<55
Serum sodium, meq/L	=180	160–179	155–159	150–154	130–149		120–129	111–119	=110
Serum potassium, meq/L	=7.0	6.0–6.9		5.5–5.9	3.5–5.4	3.0–3.4	2.5–2.9		<2.5
Serum creatinine, mg/dL	=3.5	2.0–3.4	1.5–1.9		0.6–1.4		<0.6		
Hematocrit	=60		50–59.9	46–49.9	30–45.9		20–29.9		<20
WBC count, 10^3/mL	=40		20–39.9	15–19.9	3–14.9		1–2.9		<1

Drawbacks of Score Systems

Not Adaptive!

- Often handcrafted by domain experts
- Tailored to a specific population
- Become suboptimal over time

Fixed set of clinical observations

- Not all observations available in a hospital

Standard ML methods

- Logistic Regression
- Support Vector Machine (+ Platt Scaling)
- Decision Trees
- …

Representation different from what clinicians prefer!

Our Contribution

A ML method for learning score system type prediction models for ICU mortality prediction:

- Adaptive!
- Easily interpreted by clinicians

Problem Setup

ICU Mortality Rate Prediction

Patient Training Sample: $((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N))$

Probability of death: $\hat{\eta}(\mathbf{x}) = \mathbf{P}(y = 1 \,|\, \mathbf{x})$

Computing Patient Mortality using Score Systems

Score Table

Feature Intervals

Feature i	$(a_1^i, a_2^i]$	$(a_2^i, a_3^i]$	$(a_3^i, a_4^i]$	…	$(a_m^i, a_{(m+1)}^i]$
	α_1^i	α_2^i	α_3^i	…	α_m^i

Scores

Severity score for a patient:

$$f_{\text{severity}}(\mathbf{x}) = \sum_{j=1}^{d} \sum_{k=1}^{m_j} \alpha_k^j \mathbf{1}(x_j \in (a_k^j, a_{k+1}^j])$$

Estimated patient mortality:

Parameters learnt using logistic regression

$$\hat{\eta}(\mathbf{x}) = \text{sigmoid}(\boxed{c} f_{\text{severity}}(\mathbf{x}) + \boxed{d})$$

Popular Score Systems: APACHE, SAPS, MPM, LODS, SOFA, etc…

Reformulation of Score Tables

Reformulated Score Table

Thresholds

Feature i	$(-\infty, a_1^i]$	$(-\infty, a_2^i]$	$(-\infty, a_3^i]$	…	$(-\infty, a_m^i]$
	α_1^i	α_2^i	α_3^i	…	α_m^i

Severity score for a patient:

Number of intervals / Scores / Thresholds

$$f_{\text{severity}}(\mathbf{x}) = \sum_{j=1}^{d} \sum_{k=1}^{\boxed{m_j}} \boxed{\alpha_k^j} \mathbf{1}(x_j \le \boxed{a_k^j})$$

Goal: Find a score table that minimizes logistic loss on training sample

Sparse Learning in a Blown-up Space

Cluster each feature into 'm' intervals

$$\underbrace{\mathbf{1}(x_1 \le a_1^1) \ldots \mathbf{1}(x_1 \le a_1^m)}_{\text{Original Feature 1}} \;\ldots\; \underbrace{\mathbf{1}(x_d \le a_d^1) \ldots \mathbf{1}(x_d \le a_d^m)}_{\text{Original Feature d}}$$

Sparse Logistic Regression!

Orthogonal Matching Pursuit Based Method

Iterate:

- Compute residual difference between estimated mortalities and true outcomes
- (Greedily) pick feature in blow-up space that best explains this difference
- Solve logistic regression problem over chosen features

Experimental Results

Data sets:

- St. John's data (3449 patients, 29 features)
- CinC data / MIMIC-II (4000 patients, 42 features)

Methods	AUC	Brier Score
LogitOMP-SS	70.15	0.1639
LOD	63.19	0.1724
Linear Logistic Regression	68.15	0.1664
Kernel Logistic Regression	69.00	0.1600
RankSVM + Platt Scaling	68.92	0.1668

Comparison with LOD on St. John's Data

Methods	AUC	Brier Score
LogitOMP-SS	70.47	0.1599
APACHE-II	66.07	0.1673
Linear Logistic Regression	70.47	0.1593
Kernel Logistic Regression	70.69	0.1582
RankSVM + Platt Scaling	70.67	0.1597

Comparison with APACHE-II on St. John's Data

Methods	AUC	Brier Score
LogitOMP-SS	94.32	0.0620
SAPS-II	88.02	0.0860
Linear Logistic Regression	91.20	0.0732
Kernel Logistic Regression	93.01	0.0688
RankSVM + Platt Scaling	93.13	0.0692

Comparison with SAPS-II on Cinc Data

Methods	AUC	Brier Score
LogitOMP-SS	86.67	0.0876
SOFA	81.19	0.0994
Linear Logistic Regression	84.53	0.0946
Kernel Logistic Regression	85.27	0.0921
RankSVM + Platt Scaling	85.49	0.0923

Comparison with SOFA on Cinc Data

Conclusion

	Interpretable by Clinicians?	Adaptive?
Static Score Systems	✓	✗
Standard ML Methods	✗	✓
LogitOMP-SS	✓	✓

References

1. Aurelie C. Lozano, Grzegorz Swirszcz, Naoki Abe. Group Orthogonal Matching Pursuit for Logistic Regression. AISTATS 2011.
2. Predicting Mortality of ICU Patients: The PhysioNet / Computing in Cardiology Challenge 2012.http://www.physionet.org/challenge/2012/.
3. W.A. Knaus, E.A. Draper, D.P. Wagner, and J.E. Zimmerman. APACHE II: A severity of disease classification system. Critical Care Medicine, 13(10):818–829, 1985.