Learning Score Systems for ICU Mortality Prediction via Orthogonal Matching Pursuit

Aadirupa Saha[*], Chandrahas Dewangan[†], **Harikrishna Narasimhan**[*], Sriram Sampath[‡], Shivani Agarwal [*]

[*]Indian Institute of Science, Bangalore, India
[†]Veveo India Pvt. Ltd., Bangalore, India
[‡]St. John's Medical College Hospital, Bangalore, India

Predicting Patient Mortality in Intensive Care Units

Estimating probability of patient survival/death in ICUs

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

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

Intensive Care Unit patient data

- 3499 patients with 29 clinical observations (2006-2014)

Applied score systems popular in US and Europe

	AUC
APACHE-II	66%
LOD	63%

Apache-II Score System

Scores

Intervals

Clinical Observations / Features

	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					55–69	40–54	=39
Respiratory rate	=50	35–49				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

(Knaus et al., Critical Care Medicine, 1985)

Drawbacks of Score Systems

- Not adaptive
  - Often handcrafted by domain experts
  - Tailored to a specific population (Score systems built using western patient data known to perform poorly on Indian patients; e.g., Sampath et al., 1999)
- Fixed set of clinical observations
  - Not all observation 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 models for ICU mortality prediction

- Adaptive!
- Easily interpreted by clinicians

Outline

- Score systems
- Learning score systems using OMP
- Experiments

ICU Mortality Rate Prediction

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

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

Score Table

Feature Intervals

Feature 1	$(a_1^1, a_2^1]$	$(a_2^1, a_3^1]$	$(a_3^1, a_4^1]$	...	$(a_{m1}^1, a_{(m1+1)}^1]$
	α_1^1	α_2^1	α_3^1	...	$\alpha_{m_1}^1$

Scores

Feature 2	$(a_1^2, a_2^2]$	$(a_2^2, a_3^2]$	$(a_3^2, a_4^2]$		$a_{(m2+1)}^2]$
	α_1^2	α_2^2	α_3^2	...	$\alpha_{m_2}^2$

Feature 3	$(a_1^3, a_2^3]$	$(a_2^3, a_3^3]$	$(a_3^3, a_4^3]$	...	$(a_{m3}^3, a_{(m3+1)}^3]$
	α_1^3	α_2^3	α_3^3	...	$\alpha_{m_3}^3$

.

.

.

Feature d	$(a_1^d, a_2^d]$	$(a_2^d, a_3^d]$	$(a_3^d, a_4^d]$	...	$(a_{md}^d, a_{(md+1)}^d]$
	α_1^d	α_2^d	α_3^d	...	α_{md}^d

Computing Patient Mortality Rate

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:

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

Parameters learnt using logistic regression

Popular Score Systems

APACHE -II (Knaus et al., Critical Care Medicine, 1985)

SAPS-II (Le Gall et al., JAMA, 1993)

MPM-III (Higgins et al., Critical Care Medicine, 2007)

LOD (Le Gall et al., JAMA, 1996)

SOFA (Vincent et al., Intensive Care Medicine, 1996)

…

Not Adaptive!

Outline

- Score systems
- Learning score systems using OMP
- Experiments

Score Table: Reformulation

Thresholds

Feature 1

$(-\infty, a_1^1]$	$(-\infty, a_2^1]$	$(-\infty, a_3^1]$	…	$(-\infty, a_{m1}^1]$
α_1^1	α_2^1	α_3^1	…	$\alpha_{m_1}^1$

Feature 2

$(-\infty, a_1^2]$	$(-\infty, a_2^2]$	$(-\infty, a_3^2]$	…	$(-\infty, a_{m2}^2]$
α_1^2	α_2^2	α_3^2	…	$\alpha_{m_2}^2$

Feature 3

$(-\infty, a_1^3]$	$(-\infty, a_2^3]$	$(-\infty, a_3^3]$	…	$(-\infty, a_{m3}^3]$
α_1^3	α_2^3	α_3^3	…	$\alpha_{m_3}^3$

$\vdots$

Feature d

$(-\infty, a_1^d]$	$(-\infty, a_2^d]$	$(-\infty, a_3^d]$	…	$(-\infty, a_{m_d}^d]$
α_1^d	α_2^d	α_3^d	…	$\alpha_{m_d}^d$

Score Table: Reformulation

Scores/Coefficients Thresholds

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

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

Score Table: Reformulation

Scores/Coefficients

Thresholds

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

?

Obtained by clustering each feature into intervals

Sparse Learning in Blown-up Space

Sparse Logistic Regression

Original Feature 1

Original Feature d

$\mathbf{1}(x_1 \leq a_1^1)$	$\dots$	$\mathbf{1}(x_1 \leq a_1^{m_1})$	$\dots$	$\mathbf{1}(x_d \leq a_d^1)$	$\dots$	$\mathbf{1}(x_d \leq a_d^{m_d})$

Sparse Learning in Blow-up Space

Original Feature 1 | | Original Feature d

$$\mathbf{1}(x_1 \le a_1^1) \;\Big|\; \ldots \;\Big|\; \mathbf{1}(x_1 \le a_1^{m_1}) \;\Big|\; \ldots \;\Big|\; \mathbf{1}(x_d \le a_d^1) \;\Big|\; \ldots \;\Big|\; \mathbf{1}(x_d \le a_d^{m_d})$$

Orthogonal Matching Pursuit (OMP)

Iterate:

- Compute residual difference between estimated mortality rates and true outcomes
- (Greedily) pick coordinate in blow-up space that best explains this difference
- Solve logistic regression problem over chosen coordinates

Lozano et al. , AISTATS 2011

LogitOMP-SS

Initialize:

$\mathcal{I}_0 = \phi$

$\eta_0(i) = \frac{1}{2}, \forall i \in [N]$

$t = 1$

Loop

$r_t(i) = \eta_{t-1}(i) - \mathbf{1}(y_i = 1), \forall i \in [N]$

$$(j_t, a_t) = \underset{(j,a)\in\mathcal{P}\setminus\mathcal{I}_{t-1}}{\operatorname{argmax}} \frac{|\mathbf{b}(j,a)^\top \mathbf{r}_t|}{||\mathbf{b}(j,a)||_2}$$

If $\left(\frac{|\mathbf{b}(j_t,a_t)^\top \mathbf{r}_t|}{||\mathbf{b}(j_t,a_t)||_2} \leq \epsilon\right)$

Break

End If

$\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{(j_t, a_t)\}$

$$(\widehat{\boldsymbol{\alpha}}_{1:t}, \widehat{\beta}) = \underset{\boldsymbol{\alpha}_{1:t}\in\mathbb{R}^t, \beta\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^N \log\left(1+e^{-y_i\left(f_{\boldsymbol{\alpha}_{1:t}}(\mathbf{x}_i)+\beta\right)}\right) + \frac{\lambda}{2}\|\boldsymbol{\alpha}_{1:t}\|_2^2,$$

$$\text{where } f_{\boldsymbol{\alpha}_{1:t}}(\mathbf{x}) = \sum_{\tau=1}^t \alpha_\tau \mathbf{1}\left(x_{j_\tau} \leq a_\tau\right)$$

$\eta_t(i) = \sigma_{\text{sigmoid}}\left(\sum_{\tau=1}^t \widehat{\alpha}_\tau \mathbf{1}(x_{i,j_\tau} \leq a_\tau) + \widehat{\beta}\right), \forall i \in [N]$

$t = t + 1$

End Loop

Output: $\widehat{\eta}_S(\mathbf{x}) = \sigma_{\text{sigmoid}}\left(\sum_{\tau=1}^{t-1} \widehat{\alpha}_\tau \mathbf{1}(x_{j_\tau} \leq a_\tau) + \widehat{\beta}\right)$

Sparse Learning in Blow-up Space

Original Feature 1				Original Feature d		
$\mathbf{1}(x_1 \leq a_1^1)$	$\ldots$	$\mathbf{1}(x_1 \leq a_1^{m})$	$\ldots$	$\mathbf{1}(x_d \leq a_d^1)$	$\ldots$	$\mathbf{1}(x_d \leq a_d^{m})$
α_1^1	$\ldots$	α_1^m	$\ldots$	α_d^1	$\ldots$	α_d^m

$$\underbrace{\alpha_1^1 \quad \ldots \quad \alpha_1^m \quad \ldots \quad \alpha_d^1 \quad \ldots \quad \alpha_d^m}_{\text{Learned scores/coefficients}}$$

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

Outline

- Score systems
- Learning score systems using OMP
- Experiments

Experiments

- Data sets:
  - St. John's data (3449 patients, 29 features)
  - CinC data / MIMIC-II (4000 patients, 42 features)
- Baseline score systems:
  - APACHE-II
  - SAPS-II
  - SOFA
  - LOD
- Baseline ML methods:
  - Linear/Kernel logistic regression, RankSVM

Comparison with LOD Score System

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

St. John's data

Comparison with APACHE-II Score System

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

St. John's data

Comparison with SAPS-II Score System

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

Cinc Data

Comparison with SOFA Score System

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

Cinc Data

Group Sparse Variant

- Often desirable to use models that yield good prediction accuracy with a small number of clinical observations
- Pick groups of feature-threshold pairs at each iteration

Group Sparse Variant

No. of Features	AUC	Brier Score
10	63.95	0.1699
15	65.15	0.1684
20	**65.93**	**0.1673**
APACHE-II (27 features)	**66.07**	**0.1673**

St. John's Data

Conclusion

	Interpretable by Clinicians?	Adaptive?
Static Score Systems	✓	✗
Standard ML Methods	✗	✓
Proposed Method	✓	✓