Optimizing Medical Treatment for Sepsis in Intensive Care: From Reinforcement Learning to Pretrial Evaluation

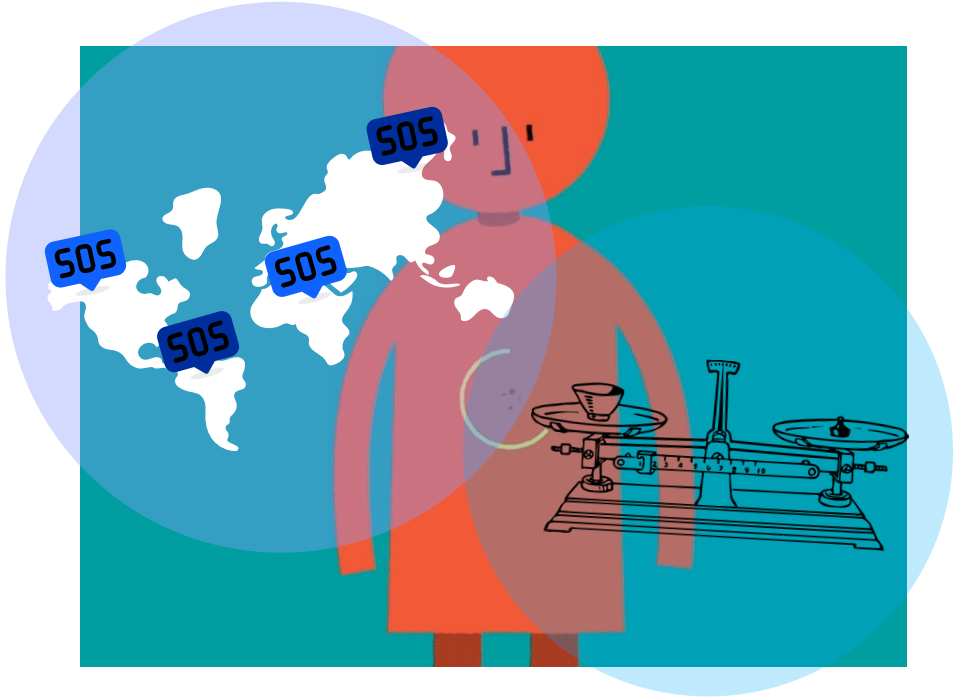
by Li, Albert-Smet, Faisal

Weiqing Wang[1, 2]
Cong Wei[1]

1. Department of Computer Science, University of Toronto.
2. Worldwide Ecosystem, International Business Machines (IBM)

The aim is to establish a policy for prospective clinical testing in general, then

Why Sepsis?

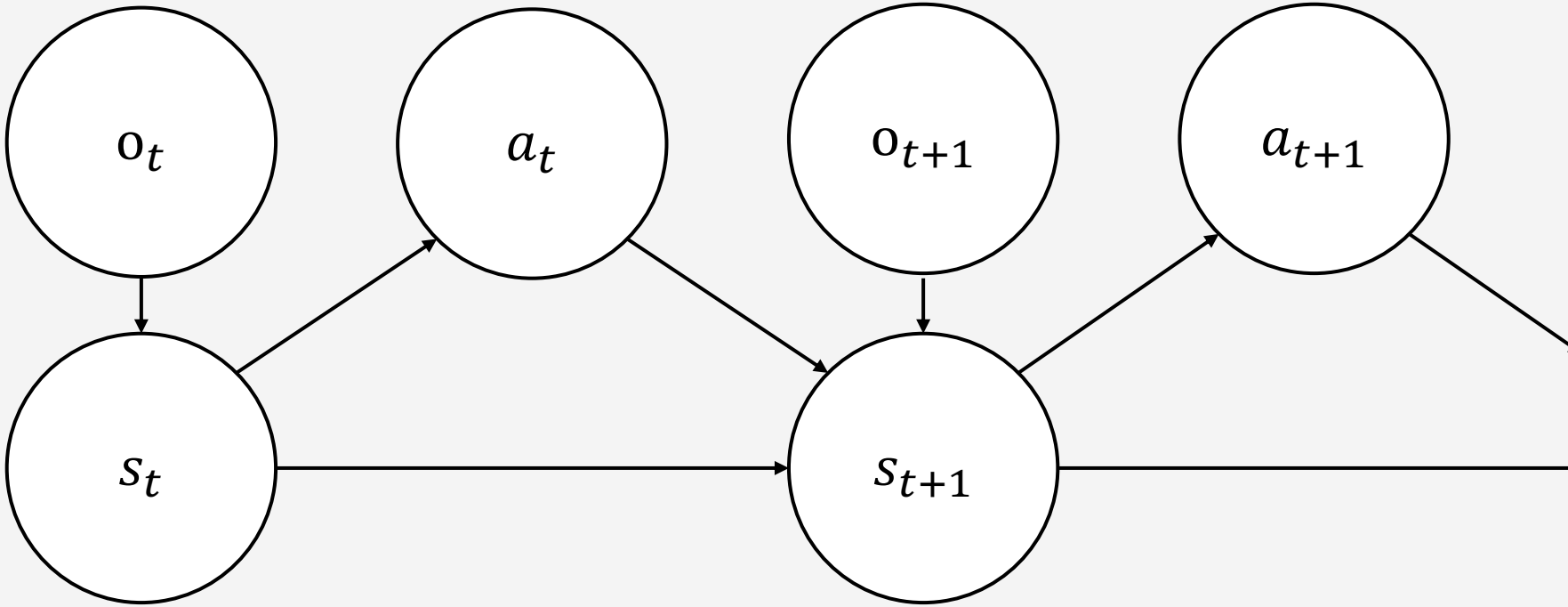


IBM

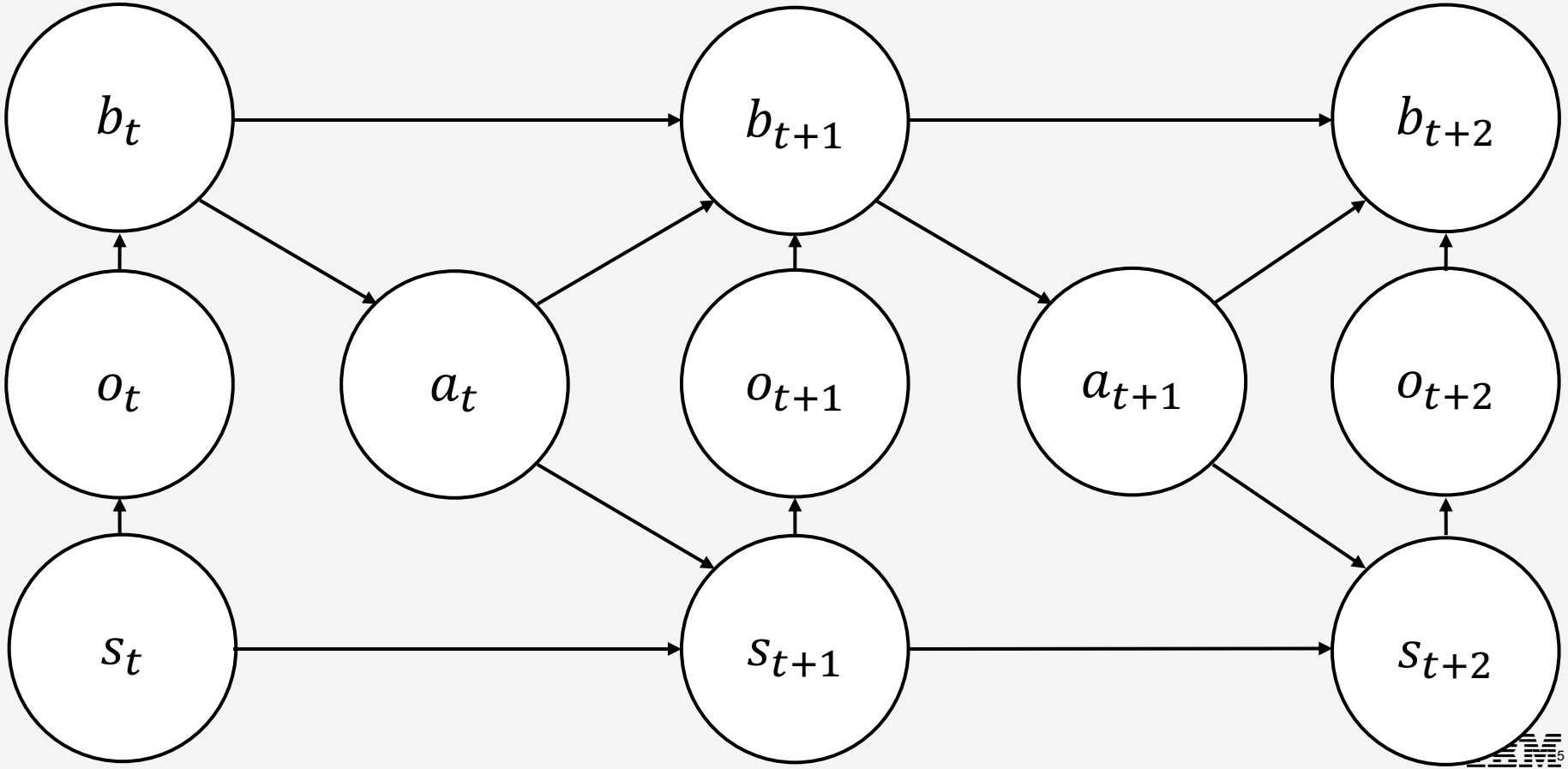
Why Reinforcement Learning?

IBM

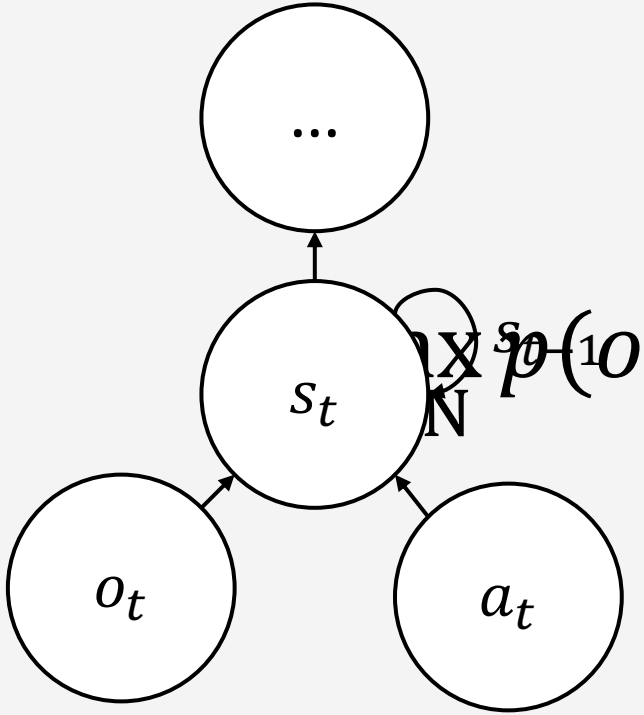
Hidden Markov Decision Process



Partially Observable Markov Decision Process



$$o_{t+1} = g(o_0, a_0, \dots, o_{t-1}, a_{t-1}, o_t, a_t)$$
$$= h(s_{t-1}, a_t, o_t)$$



$$p(o_{t+1}|s_t, a_t)$$

Policy Optimization

Value-based

- Train $Q(S, a)$.
- Policy:
  $\underset{a \in \mathcal{A}}{\text{argmax}}\, Q(S, a)$

Policy Based

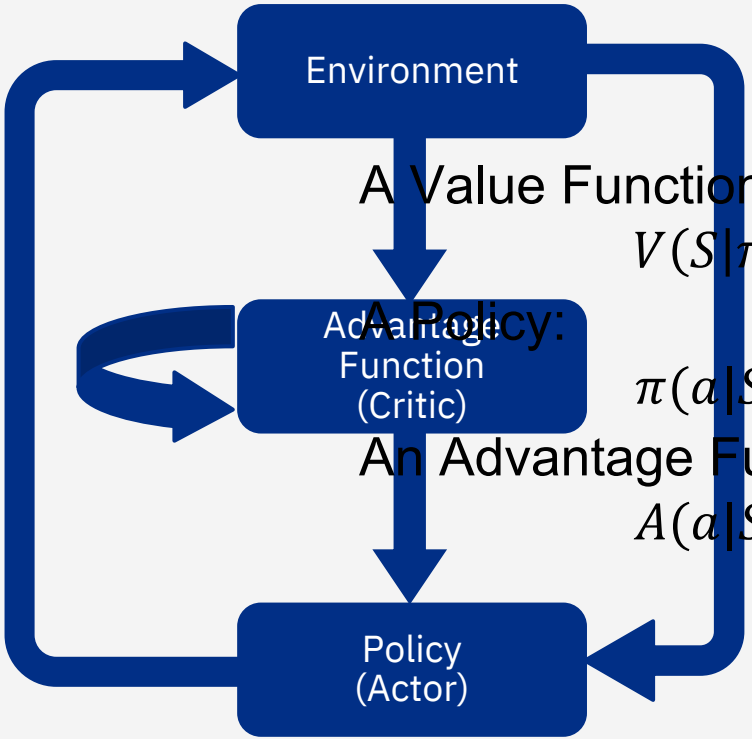
- π is stochastic.
- Choose π that maximises $V^{\pi}(S_0)$

Model Based

- Agent Evaluation
- Agent Training
- World Model

IBM

Advantage Actor Critic (A2C)



A Value Function:

$V(S|\pi)$

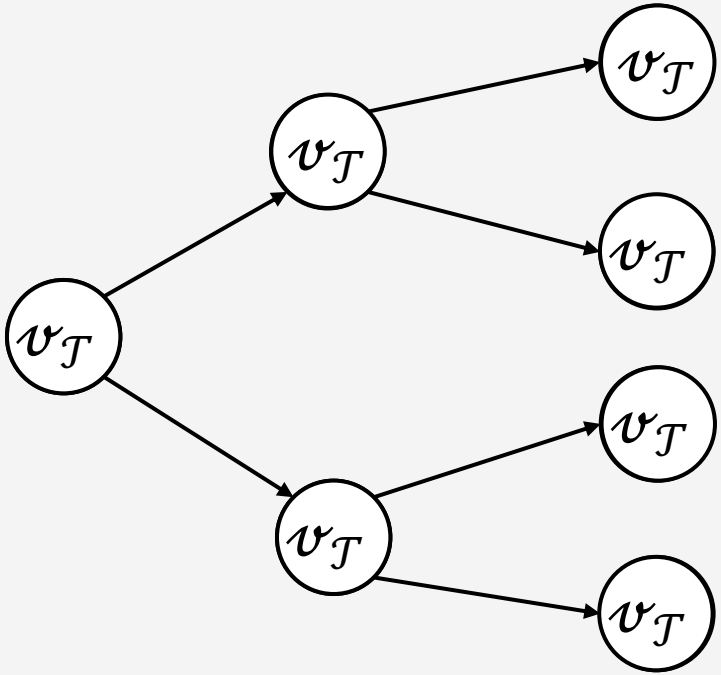
A Policy:

$\pi(a|S)$

An Advantage Function:

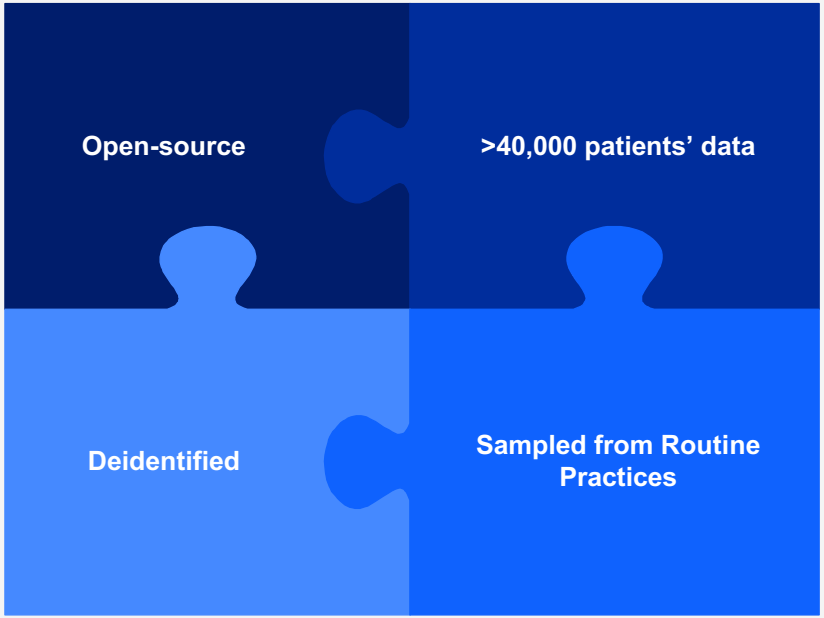
$A(a|S)$

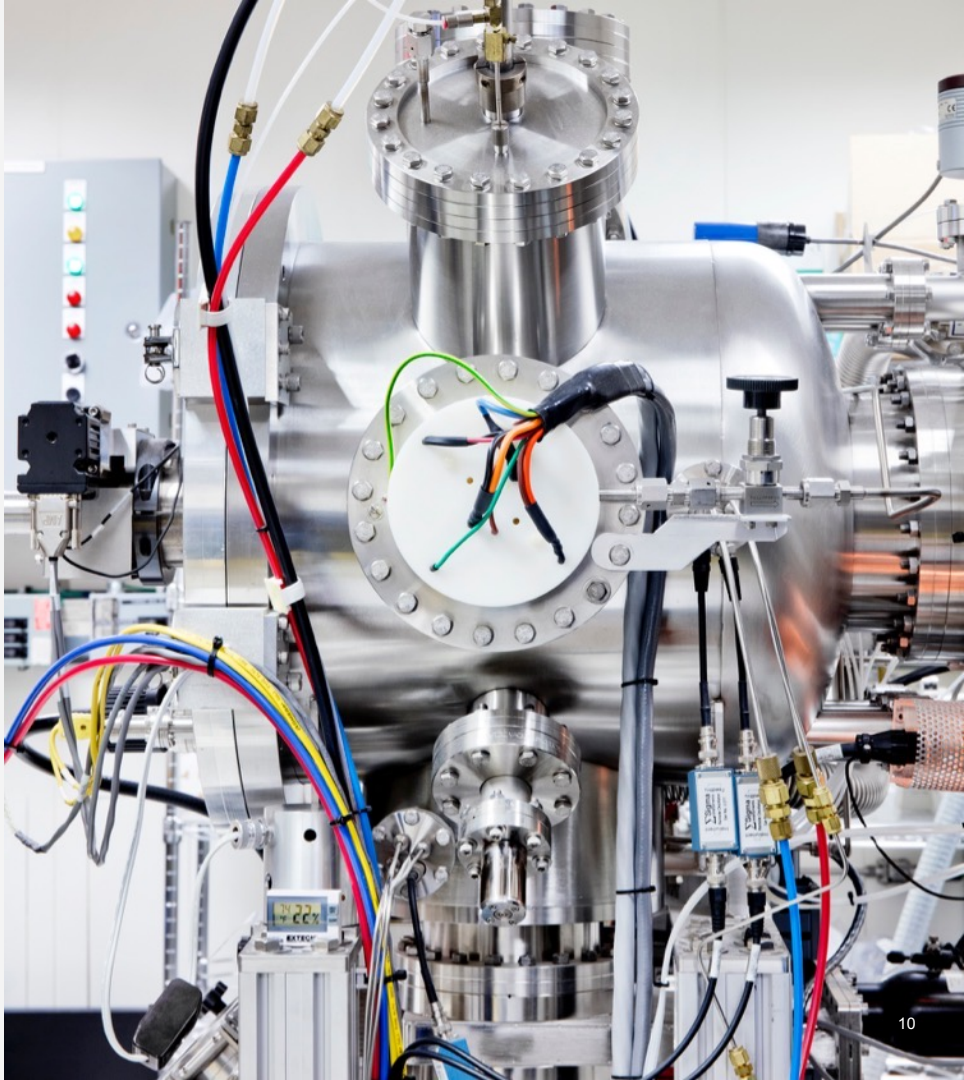
Heuristic Tree Search



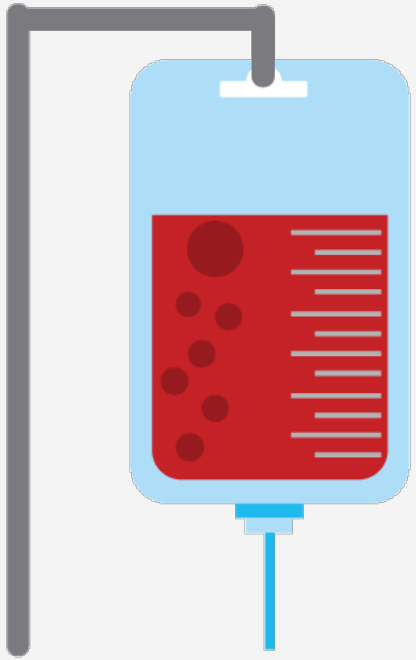
Preclinical Training Dataset

Medical Information Mart for Intensive Care (MIMIC)





The Action Space

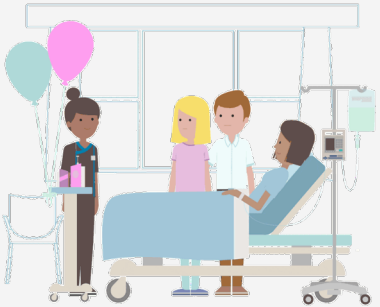


Maximum dose of vasopressors administered.

The total volume of intravenous fluids injected over each hourly period.

Reward Function

Survival



-10
Death

+10
Discharge

Pre-clinical Testing

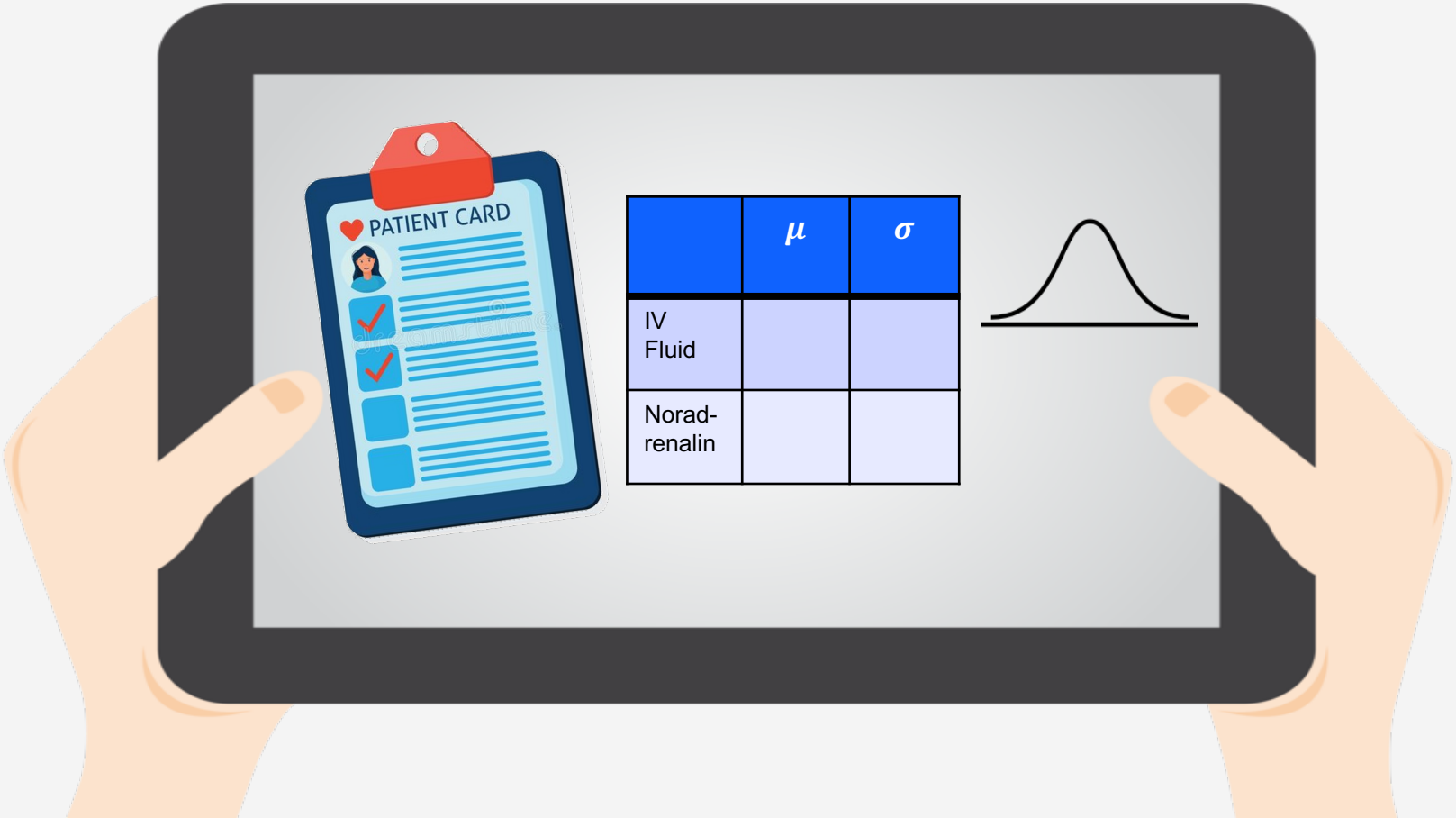
Consistency?

Whether AI is consistent with human clinicians?

Model-agnostic

A novel framework is proposed.

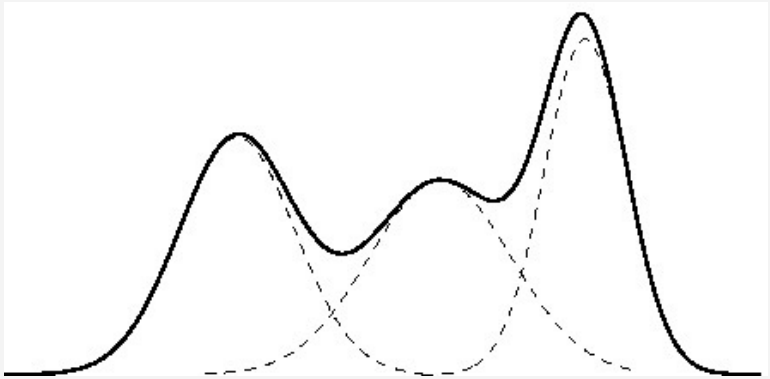
Pre-clinical Testing

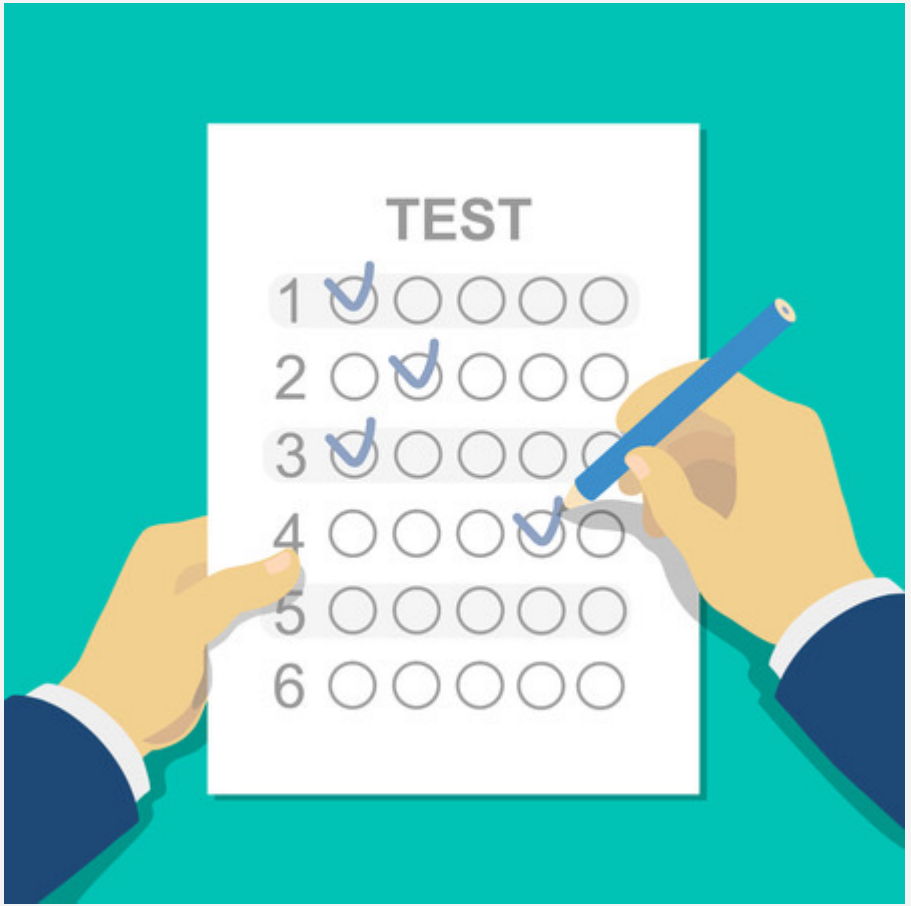


	μ	σ
IV Fluid		
Norad-renalin		

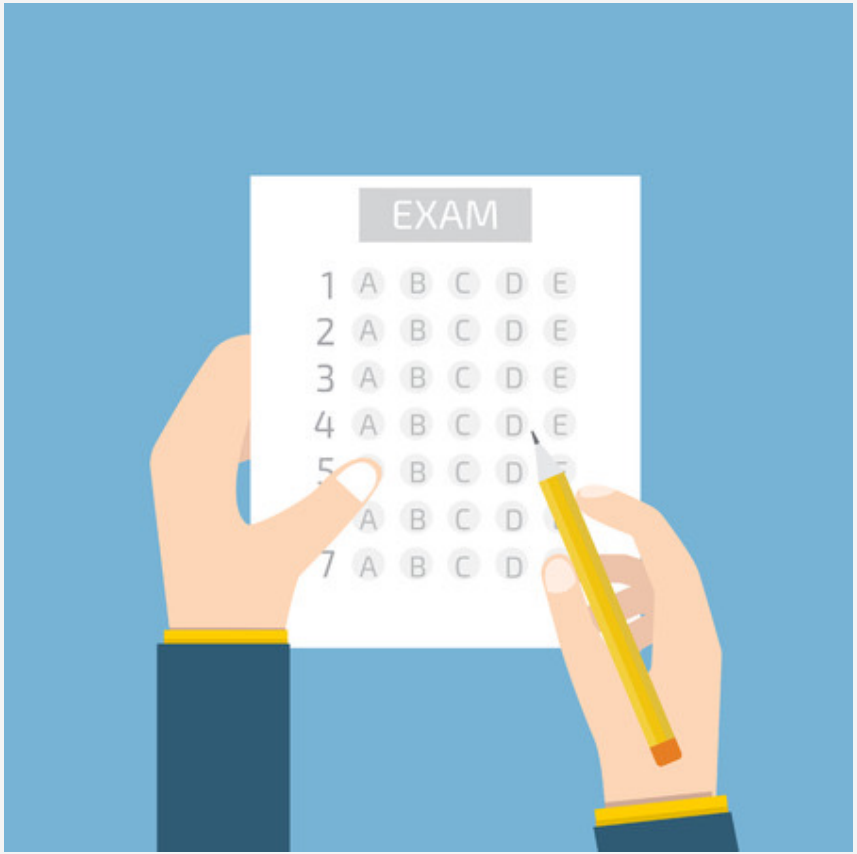
Evaluation – P-score

$$P_{\text{score}}(a) = \frac{1}{N}\sum_{i=1}^{N} f_{\mu_i,\sigma_i}(a)$$





IBM



Evaluation – C-score

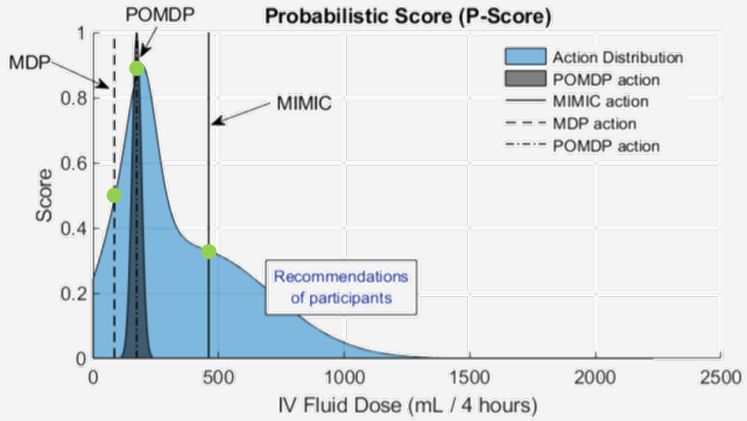
$$C_{\text{score}}(a) = \hat{p}(\mu_* - 3\sigma_* \le a \le \mu_* + 3\sigma_*)$$

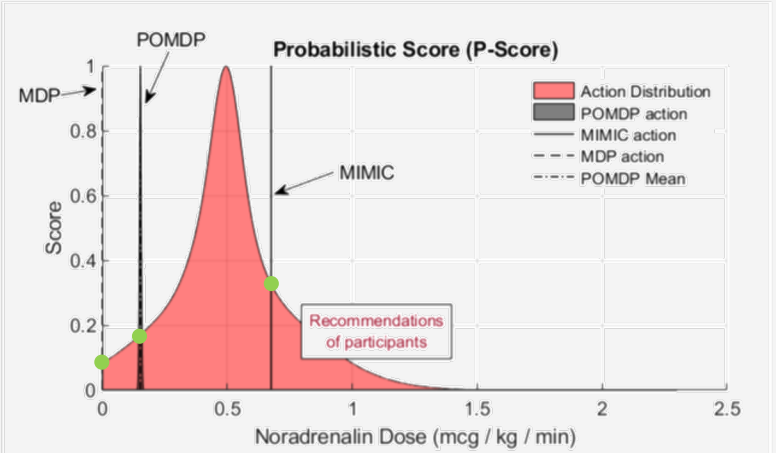
$$= \frac{1}{N}\sum_{i=1}^{N} I(\mu_i - 3\sigma_i \le a \le \mu_i + 3\sigma_i)$$

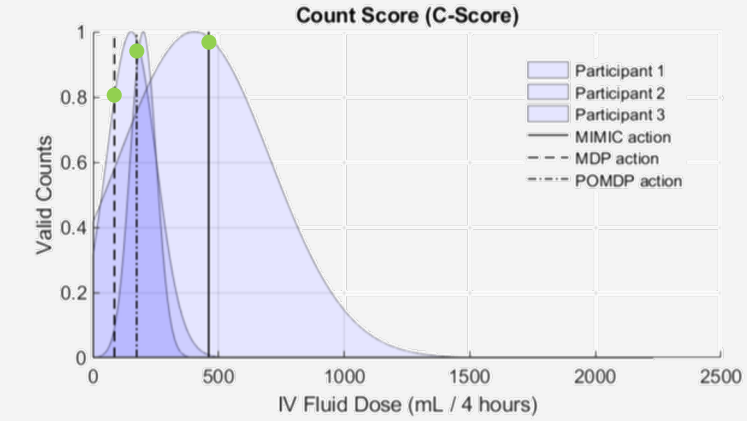
$$\approx p(\mu_* - 3\sigma_* \le a \le \mu_* + 3\sigma_*)$$

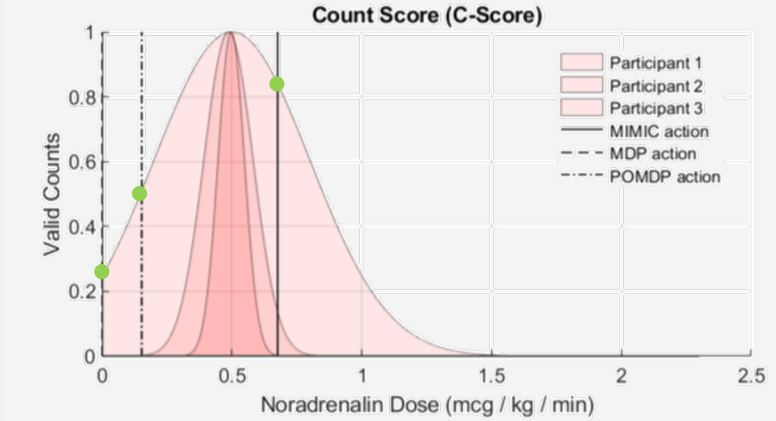
IBM

Result


POMDP
MDP
MIMIC
Probabilistic Score (P-Score)
Action Distribution
POMDP action
MIMIC action
MDP action
POMDP action
Recommendations of participants
Score
IV Fluid Dose (mL / 4 hours)


POMDP
MDP
MIMIC
Probabilistic Score (P-Score)
Action Distribution
POMDP action
MIMIC action
MDP action
POMDP Mean
Recommendations of participants
Score
Noradrenalin Dose (mcg / kg / min)


Count Score (C-Score)
Participant 1
Participant 2
Participant 3
MIMIC action
MDP action
POMDP action
Valid Counts
IV Fluid Dose (mL / 4 hours)


Count Score (C-Score)
Participant 1
Participant 2
Participant 3
MIMIC action
MDP action
POMDP action
Valid Counts
Noradrenalin Dose (mcg / kg / min)

Result

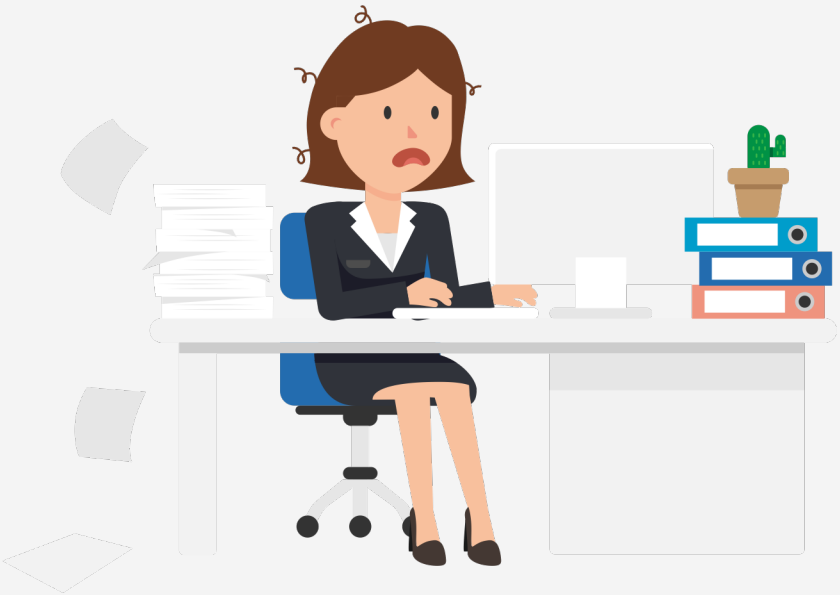
Table 1: This table contains the average validation scores observed when three separate intensive care consultants evaluated 10 patient trajectories from the MIMIC dataset. The best score for each row is shown in bold. The column labels stand for: the original action recorded (MIMIC), a discrete and non-time dependant model (MDP) by Komorowski et al. (2018), and the model we propose in this work (POMDP).

ACTION	SCORE	MIMIC	MDP	POMDP
IV Fluids	P-Score	**0.454**	0.434	0.448
	C-Score	0.578	0.622	**0.644**
	Zero Count	0.133	0.100	**0.033**
Vasopressors	P-Score	**0.584**	0.286	0.488
	C-Score	**0.700**	0.467	0.644
	Zero Count	**0.033**	0.233	**0.033**

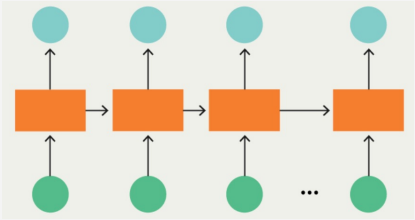
Strength

- ✓ POMDP fits Medical Setting
- ✓ An accurate state representation
- ✓ Preclinical Validation Framework

Limitations



- ✓ Large Dataset Required
- ✓ Small Number of Participants
- ✓ Pitfall of RNN



IBM

Thank you

Weiqing Wang

—

weiqing.wang@utoronto.ca | Weiqing.Wang@ibm.com

Cong Wei

—

cong.wei@mail.utoronto.ca

Appendix: Actor Critic

$$A(S,a) = \begin{cases} \delta_t & \text{last nonterminal step} \\ \gamma c_t \max(A(s_{t+1}, a_{t+1}), 0) + \delta_t & \text{otherwise} \end{cases}$$

- δ_t is the off-policy temporal difference error for V at time t.
- c_t is the contracted Importance Sampling Ratio.
- γ is the discount factor.

Appendix: Actor Critic

$$s^* = \underset{s \in \mathcal{F}}{\operatorname{argmax}}\, \gamma^{D(s)-1} \prod_{d=0}^{D(s)-1} \hat{p}\left(s^{d+1} \middle| s^d, a_{\mathcal{T}}(s^d)\right) \delta_{\mathcal{T}(s^0, s)}(s^d)$$

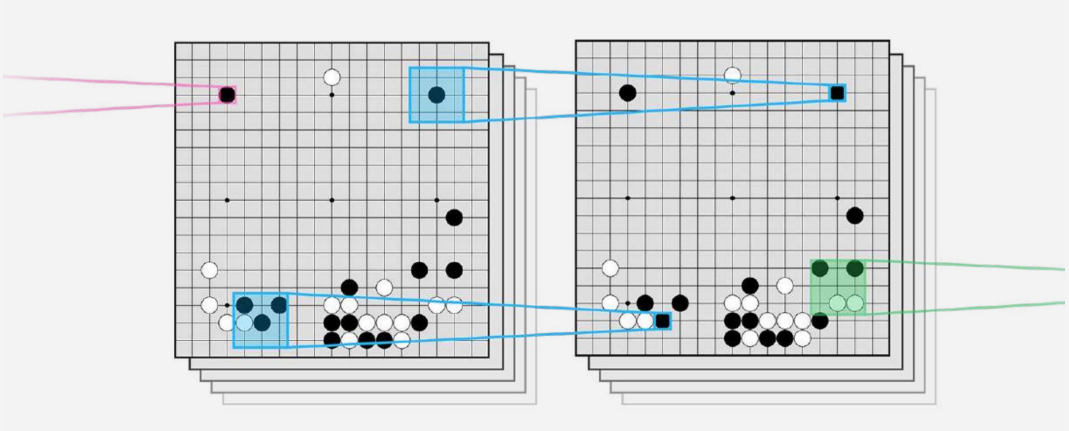
- $D(s)$ is the depth of the node.
- δ_x is the Dirac delta function concentrated at x.
- $a_{\mathcal{T}}(s^d)$ deterministic tree policy in s^d.

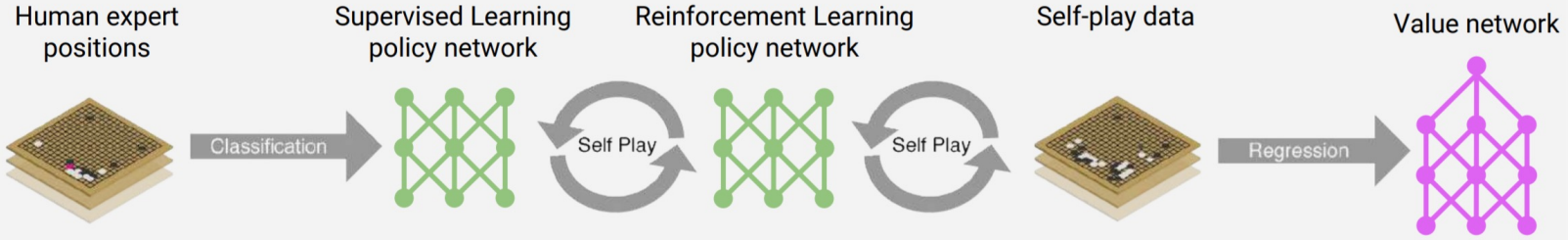
Appendix: Actor Critic

$$a_{\mathcal{T}}(s^d) = \underset{a \in \mathcal{A}}{\operatorname{argmax}}\, r(s^d, a) + \gamma \sum_{s^{d+1} \in C(s^d, a)} \hat{p}(s^{d+1} | s^d, a) V(s^{d+1})$$

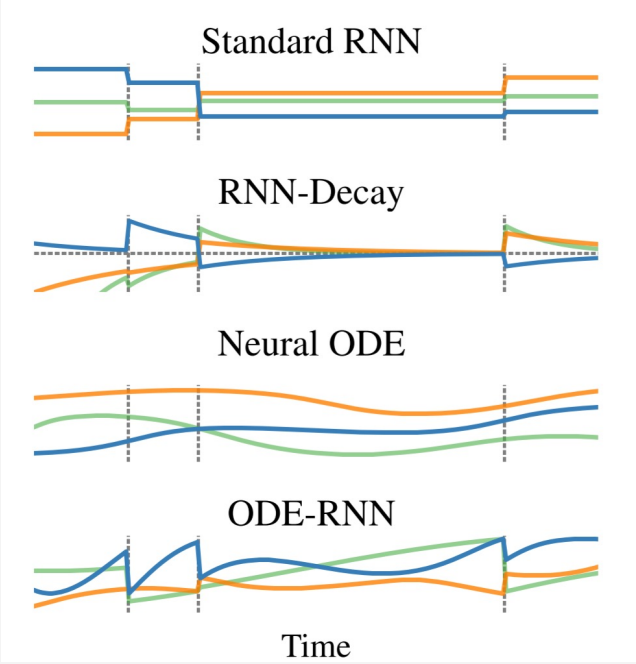
$$v_{\mathcal{T}}(s^d) = \begin{cases} V(s^d) & s^d \text{ is a leaf} \\ r\left(s^d, a_{\mathcal{T}}(s^d)\right) + \gamma \sum_{s^{d+1} \in C\left(s^d, a_{\mathcal{T}}(s^d)\right)} \hat{p}(s^{d+1} | s^d, a_{\mathcal{T}}(s^d)) v_{\mathcal{T}}(s^{d+1}) & \text{otherwise} \end{cases}$$

Appendix: Limitation

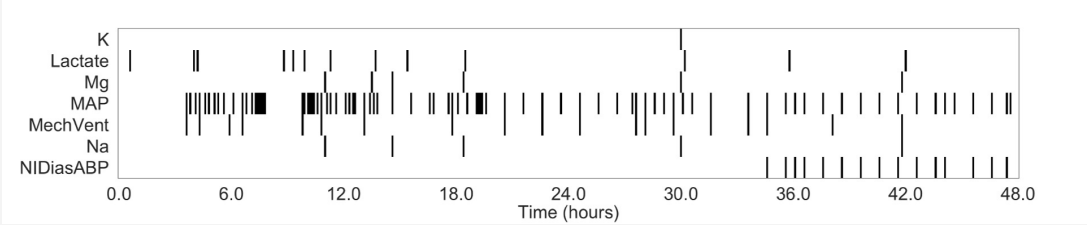




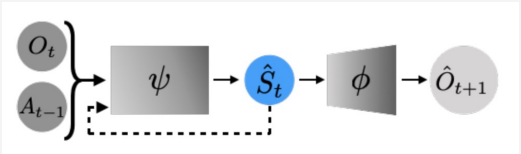
Appendix: Pitfall of RNN



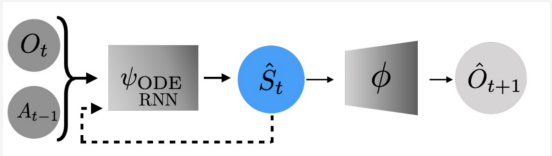
Hidden State Trajectories.



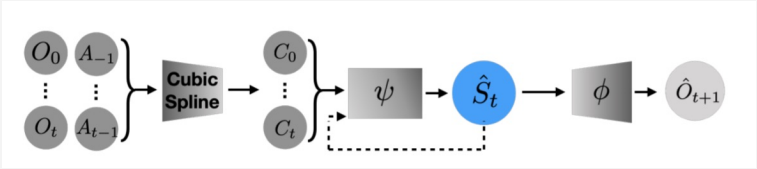
PhysioNet Challenge 2012 dataset.



Basic RNN Autoencoder (RNN).



ODE-RNN.



Neural Controlled Differential Equations.