Analyzing Sepsis Health Outcomes Using Reinforcement Learning for MDP Dynamics

Andy Jin,[1] Jinpeng Song,[1] Sophia Sanchez[1]

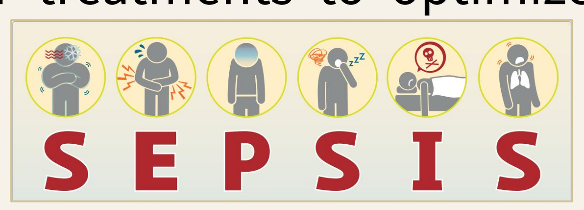
[1] *Computer Science Department, Stanford University*

Stanford ENGINEERING
Computer Science

Introduction

MOTIVATION

According to WHO, sepsis is estimated to affect more than 30 million people worldwide every year, potentially leading to 6 million deaths. We aim to provide physicians a data-driven approach on how to identify and administer treatments to optimize patient health outcomes.



PROBLEM DEFINITION: TWO PHASES

Inputs:

- 17,000 sepsis Boston General Hospitals patients
- 688 physiological and demographic features:
  - Treatments administered
  - Demographic/static
  - Lab values
  - Vital signs
  - Intake/output events
  - Time stamp

Problem Statement (Outputs):

1. **Construct an MDP** to specify sepsis transition dynamics using a generative model via the variational autoencoder (VAE)
2. **Deduce optimal treatment policies** given the health trajectories using deep Q-learning

Our MDP

State: Physiological and health indicators, per 4-hour timesteps—to capture contextual evidence

Action: 5x5 discrete space of potential medical interventions—dosage of intravenous fluid (IV) and the maximum vasopressor (VP)

Reward:

- **Non-terminal Timesteps:** Intermediate reductions in symptom severity—Sequential Organ Failure Assessment and Lactate levels.
- **Terminal Timestep:** Patient mortality in ICU

End State: Patient leaves ICU alive or dies in ICU

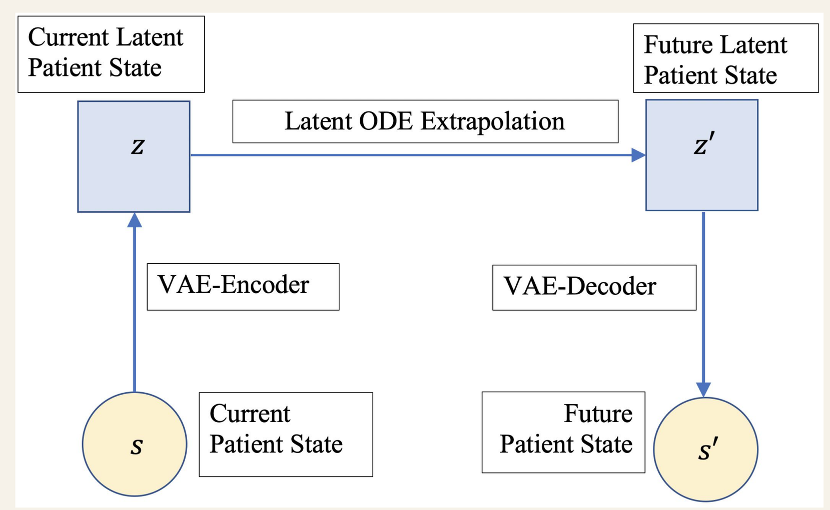
Model Implementation

POSTERIOR FOR VAE GENERATIVE MODEL

Goal: Estimate parameters θ (initialized to Gaussian with mean 0) that express predicted next patient state (z) given current state and data [1]

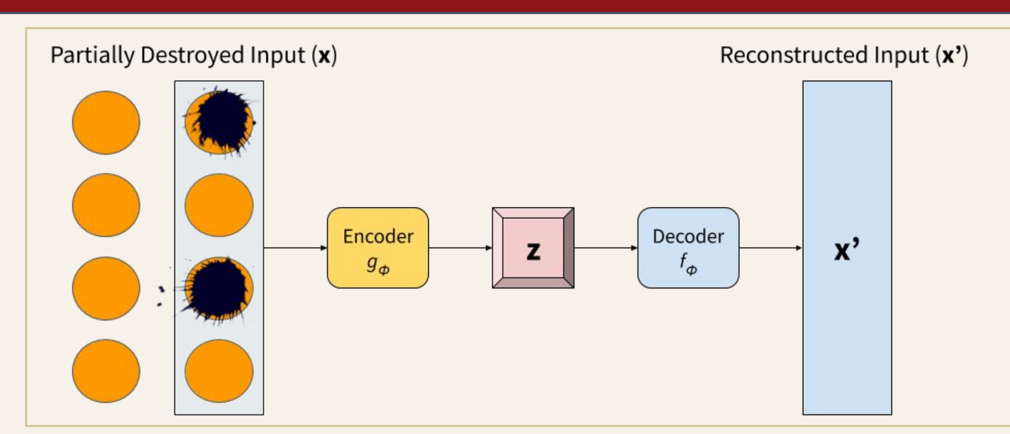
- $z_0 \sim p(z_0)$
- $z_0, z_1, ..., z_N = \text{ODE-Solve}(f_\theta, z_0, (t_0, t_1, ..., t_N))$
- each $x_i \overset{indep}{\sim} p(x_i|z_i)$ for $i = 0, 1, ..., N$

Data is **irregularly sampled** (treatments are not administered at consistent times), so we use Latent ODE-RNN [1] to approximate the latent space.



*Pre-activations in the RNN are based on initial-value solution to an ODE

THE VARIATIONAL AUTOENCODER



DEEP Q-LEARNING IMPLEMENTATION

Goal: Predict SOFA and mortality outcome for given patient state and treatment intervention:

$$\theta^* = \text{argmin}_\theta \mathbb{E}\left[(Q_{\text{target}} - Q(s, a; \theta))^2\right]$$

where Q_{target} is the discounted sum of rewards.

- Use **DQN** because state space is continuous [2]
- Use **Autoencoder** to expand the dimensions of state space
- Specifically use Dueling-DDQN to determine quality of state without knowledge of action [2].

Results and Evaluation

GENERATIVE MODELING EVALUATION

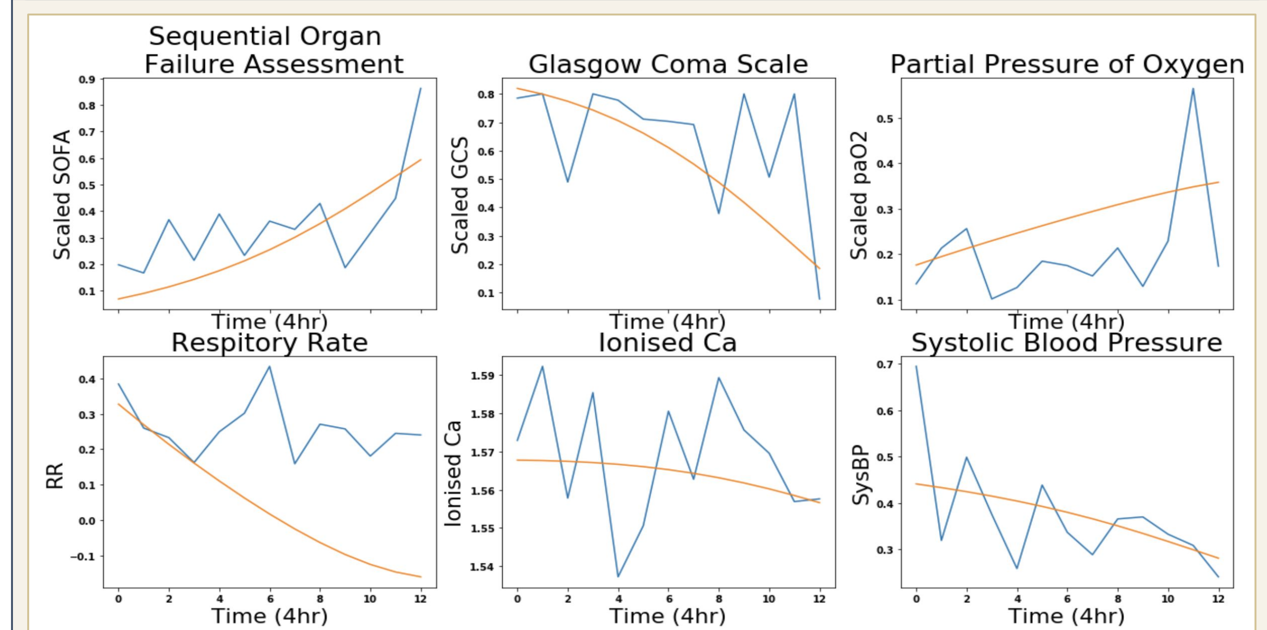


FIG. 1: Predicted state trajectories anchored at *t = 0*

DEEP Q-LEARNING OPTIMAL POLICIES

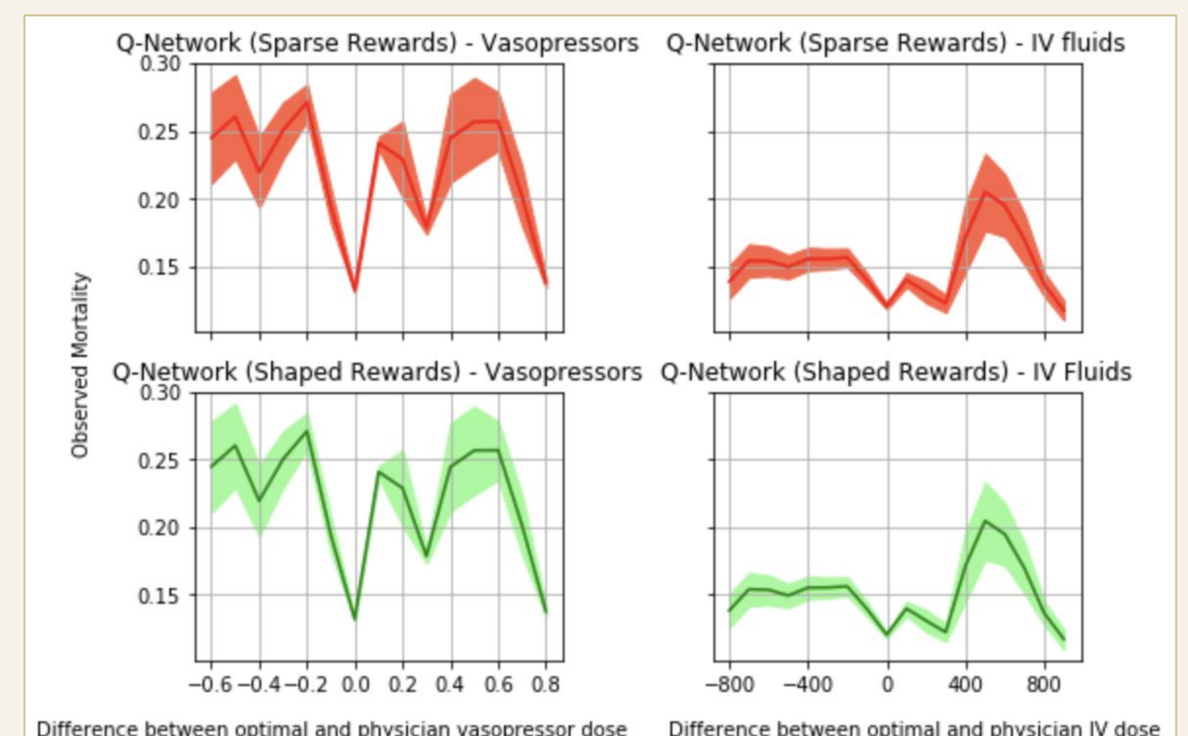


FIG. 2: Dosage given by clinician (solid line) vs. DQN (shaded area representing variance)

COMPARISON TO PHYSICIAN POLICIES

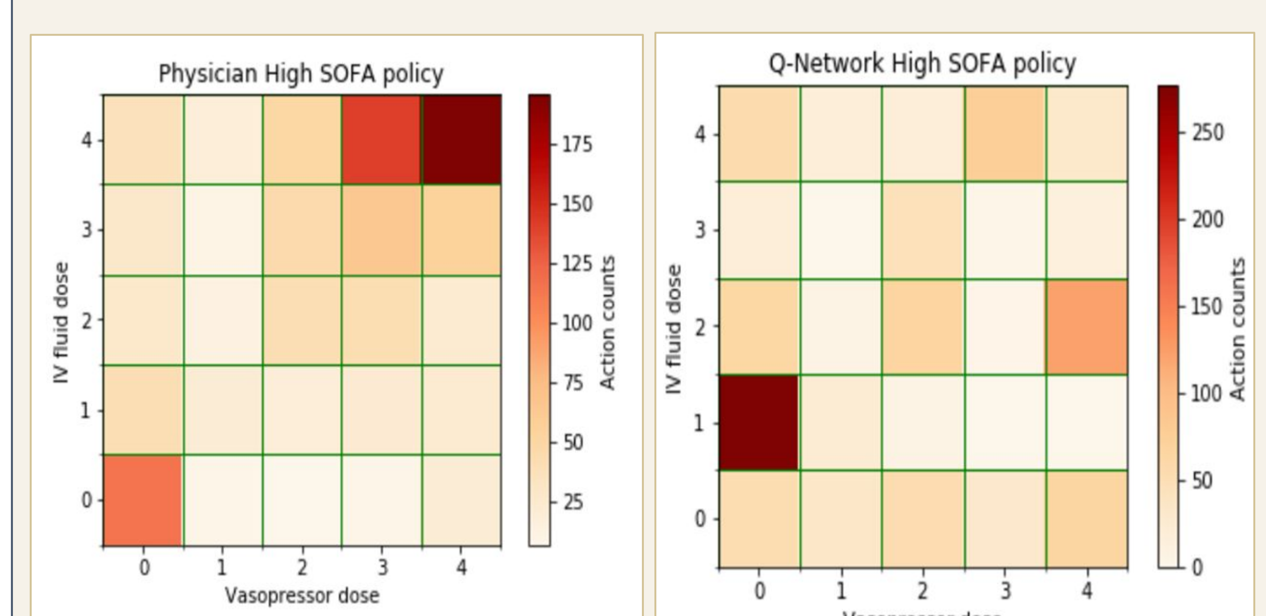


FIG. 3: IV and VP dosages given by physicians vs. recommended by DQN for high-SOFA patients

Discussion

ERROR ANALYSIS

Baseline: Our prediction of whether or not patient dies attains accuracy of ~0.65 with KNN (K = 4).

Oracle: What actually occurred in the patient. Specifically, the transition from a state given the action (which we recorded as a data point).

To evaluate our VAE and KNN, we calculated the MSE of the hold-out test set:

- **KNN**: MSE ~ 0.35
- **VAE**: MSE ~ 0.0041

DQN: Compare patient mortality given a deviation between physician policy and optimal policy. Generally, optimal mortality is at difference = 0.

DISCUSSION AND ANALYSIS

FIG. 1: Shows our VAE can approximate the patient state in extrapolation. However, limited in ability to generalize to later time-steps and patient-to-patient variability in treatment response

FIG. 2: We see that the closer the physician policy follows the the optimal policy, the greater the optimal survival.

FIG. 3: Shows the challenges of generalizing policies to High SOFA values, which occur less frequently

CHALLENGES AND FUTURE WORK

1. **Leveraging MDP from VAE:** Run *model-based* RL on generative model produced by VAE.
   - The VAE (Fig. 1) generates overly smooth predictions that do not precisely reflect noisy patient samples.
2. **Differential Privacy:** When training DQN autoencoder, add Gaussian noise to the SGD
   - Hard to generate robust privacy score and create an accurate graph of optimal policies

ACKNOWLEDGEMENTS

[1] Rubanova, et al. (2019). Latent ODE's for Irregularly-Sampled Time Series

[2] Raghu, et al. (2017). Continuous State-Space Models for Optimal Sepsis Treatment: A Deep Reinforcement Learning Approach.