GYMIC: An OpenAI Gym Environment for Simulating Sepsis Treatment for ICU Patients

Amirhossein Kiani
Stanford University
akiani@stanford.edu

Tianli Ding
Google Inc.
tding419@stanford.edu

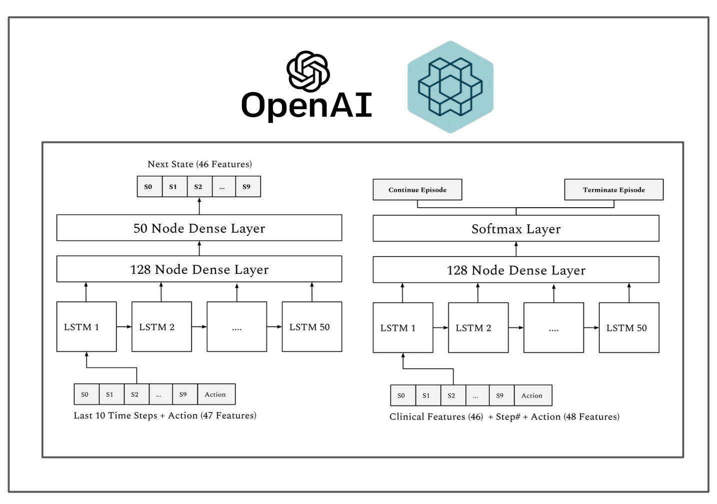
Peter Henderson
Stanford University
phend@stanford.edu

1 Introduction

Sepsis is a life-threatening illness caused by the body's response to infection and a leading cause of patient mortality. Aside from different dosage of antibiotics and controlling the sources of infection, sepsis treatment involves administering intravenous fluids and vasopressors. These procedures have however shown to have drastically different results on different patients and there is a lack of efficient real-time decision support tools to guide physicians. In this project we studied the application of reinforcement learning towards learning such policies from MIMIC which is an open patient EHR dataset from ICU patients. We built a custom OpenAI Gym environment to simulate the MIMIC Sepsis cohort and ran off-the-shelf OpenAI Baselines algorithms on our custom environment. We additionally replicated the work done by Raghu et al. on Off Policy Deep RL for learning Sepsis treatment policies.

2 Simulation Environment

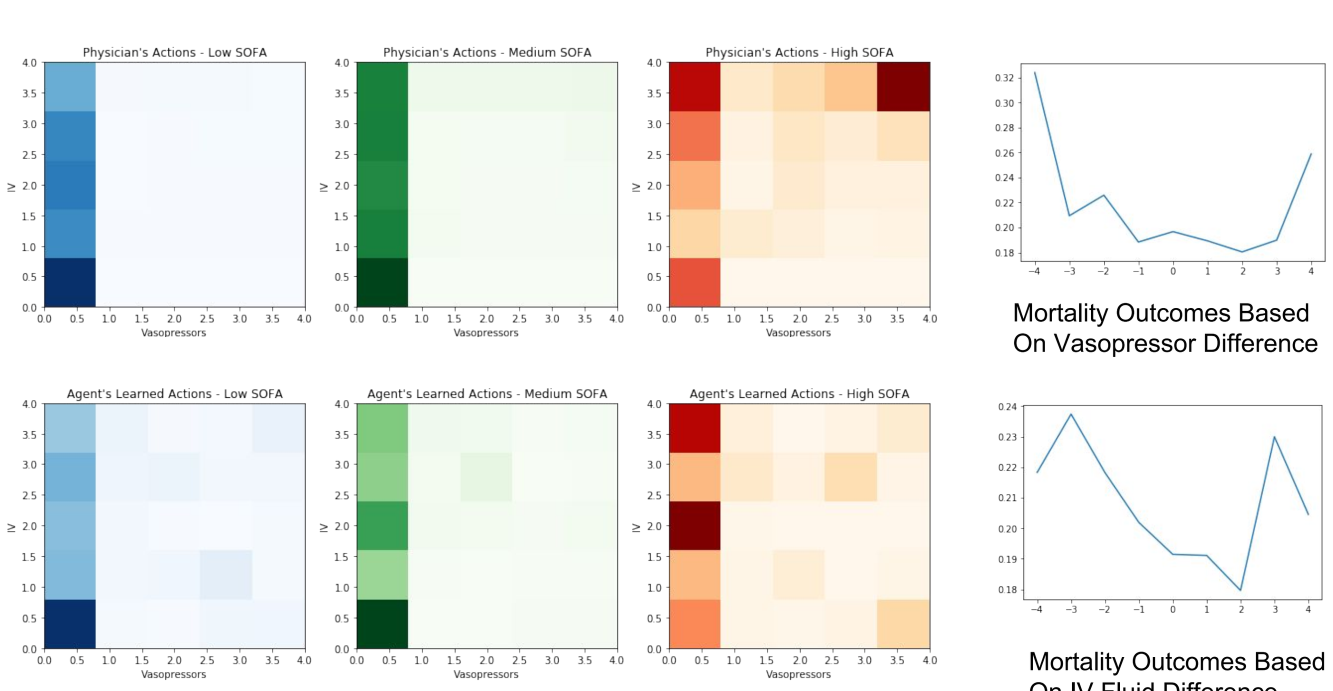
Our approach is centered on building a custom OpenAI Gym environment that simulates sepsis treatment trajectories in the ICU. This environment consists of four key modules:



(1) State Model: The input to this model consisted of the 46 normalized clinical features and the action value (0-24) for the 10 previous time steps (zero padded). The output of the model was the 46 normalized features for the next step.
(2) Episode Termination Model: A separate model was developed to detect episode transitions. The transitions were defined as two mutually exclusive cases of (1) terminating the episode or (2) continuing the episode.
(3) Episode Outcome Model: A third model with the same features and architecture as the Episode Termination Model was developed to predict two mutually exclusive outcomes of death or release from hospital. This model was used in the environment to decide the reward values at the end of each episode.
(4) OpenAI Gym Wrapper: We embedded the above models in an OpenAI Gym environment accessible on: https://github.com/akiani/gym-sepsis

3 Replication Study

We replicated the work done by Raghu et al. on Off Policy Deep RL for learning Sepsis treatment policies. We built a Dueling DQN Network that aside from minor differences matched the approach by Raghu et al. and were able to replicate their findings with minor mismatches due to potential differences in data processing and training procedure.



Mortality Outcomes Based On Vasopressor Difference

Mortality Outcomes Based On IV Fluid Difference

Raghu, A., Komorowski, M., Ahmed, I., Celi, L. A., Szolovits, P., and Ghassemi, M. Deep reinforcement learning for sepsis treatment. CoRR, abs/1711.09602, 2017. URL http://arxiv.org/abs/1711. 09602.

4 Dataset

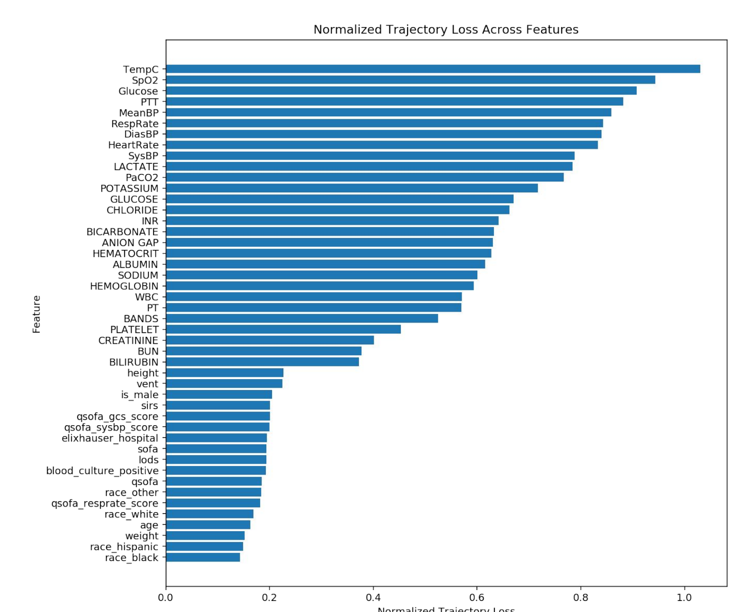
MIMIC is a large and comprehensive dataset consisting of de-identified health data associated with ~40,000 critical care patients. It includes features such as patient demographics, vital signs, laboratory tests, medications and medical interventions. We used **46 clinical features** to denote each state. For actions, we defined a discrete **5 x 5 action space** for the medical interventions spanning the space of intravenous (IV) fluid and maximum vasopressor (VP) dosage in a given 4 hour window. **Rewards were set to +15 for the last step in episodes where the patient was released from the hospital and to -15** for when the hospital stay had resulted in death of the patient. The data processing task was a major undertaking as it involved writing many SQL and R scripts and working with +60GB of data.

5 Simulator Results

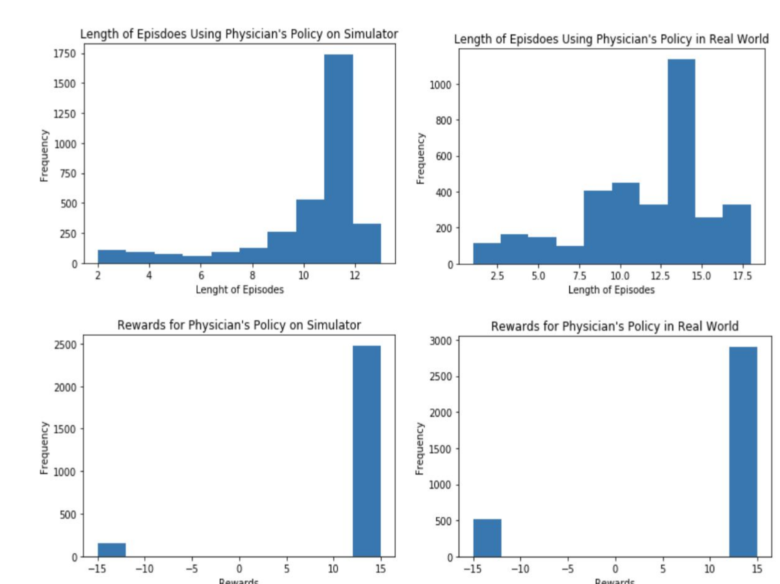
We measured our models' Accuracy and MSE based on the test dataset.

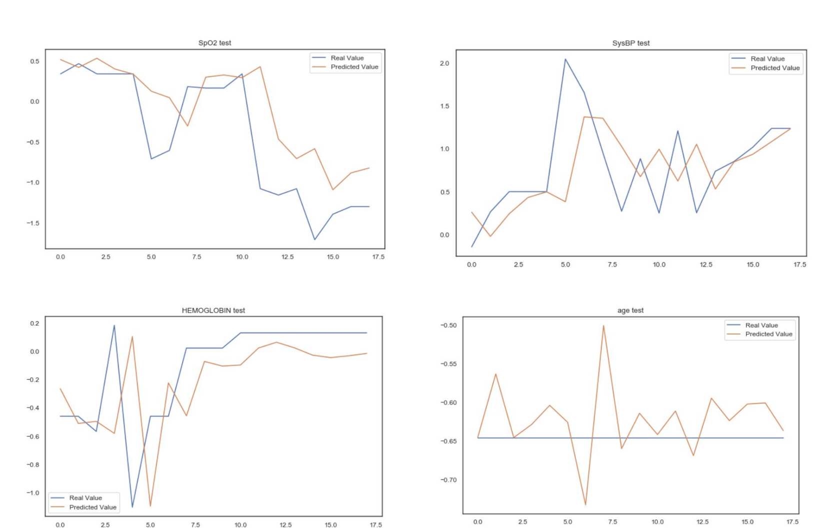
	Performance on Test Dataset	Metric
State Model	0.1161	Mean Squared Error
Episode Termination Model	97.20%	Accuracy
Outcome Prediction Model	86.31%	Accuracy

In order to measure the compounding error for the simulator, we defined the **Normalized Trajectory Loss** metric. This metric is defined as the mean squared error of the episodic values generated by the model when the physician's historical actions are played on a simulator initialized with identical initial state.

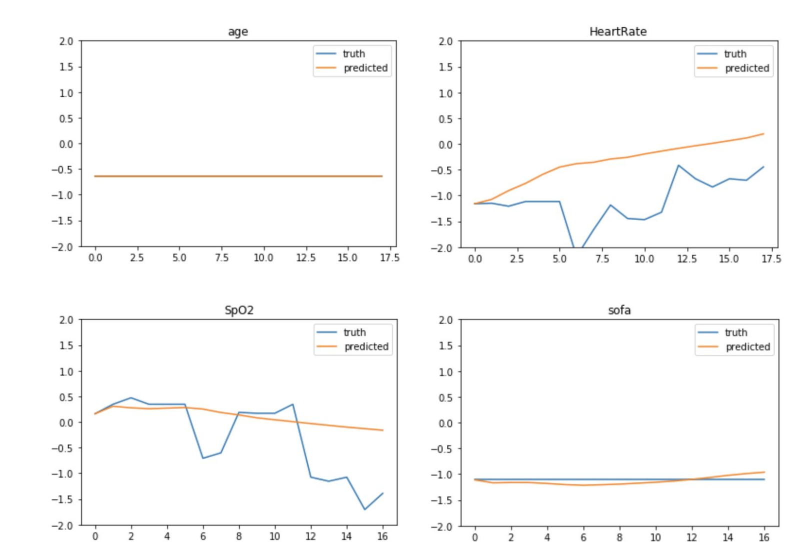


We also studied our simulator's performance by comparing its values to the real world data from our dataset. We did this both by predicting time series for each patient's features as well as rolling out physician policy.





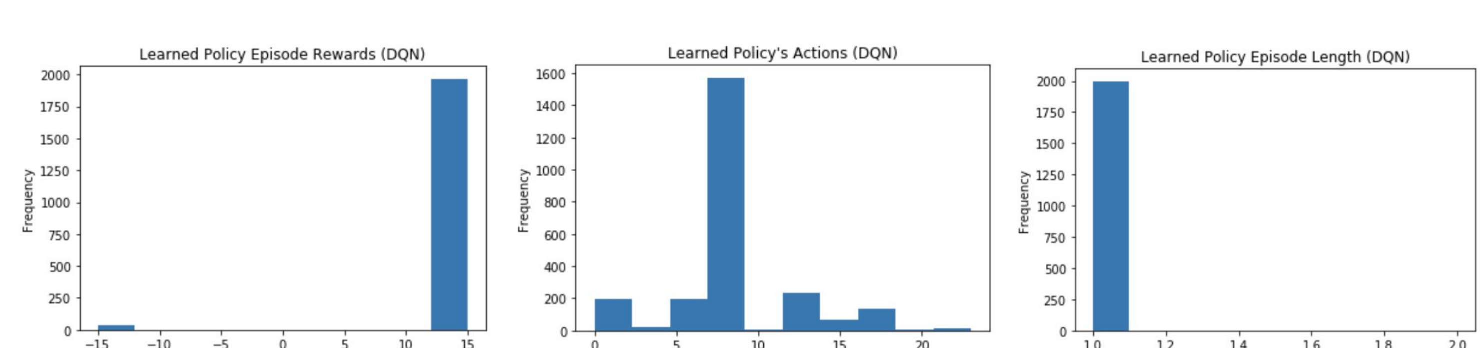
Time Series Prediction



Policy Rollout on Simulator

6 Training using OpenAI Baselines

To complete the evaluation of our simulator, we leveraged OpenAI Baseline's off the shelf algorithms to train two agents on top of our OpenAI gym environment. We chose the DQN method to learn a comparable policy to our off-policy learned method.



7 Future Direction

Potential additions to this work include: (1) leveraging Variational Autoencoders to denoise the data and improve the state model (2) building a stochastic model for the space that accounts of the uncertainties in the feature space (e.g. Deep Bayesian Neural Network) and (3) a visual render mode for the environment.