Off-Policy Deep Reinforcement Learning for Optimal Sepsis Treatment

Imran Ahmed and Aniruddh Raghu

Problem outline and goal

- Treatment policies for septic patients are suboptimal
  - Patients react variably to interventions
  - No universally agreed-upon treatment exists

- **Goal** - use observational data to discover treatment policies that improve chances of patient survival

- **Baseline** - mortality under physician treatment policy (13.7%)

Cohort

- MIMIC-III cohort - patients fulfilling Sepsis-3 criteria
  - Suspicion of infection
  - Evidence of organ dysfunction (SOFA score > 2)

- Include data from up to 24h preceding diagnosis
  - Time period around diagnosis is critical

- Outcome of interest - in-hospital mortality

Methods

- Formulation - continuous state-space Markov Decision Process (MDP)
  - **State** - continuous vector of patient's physiological measurements + vital signs
  - **Actions** - discretized over doses of vasopressors and IV fluids
  - **Rewards** - depend on model:
    - Sparse: at terminal timesteps, depending on outcome
    - Clinically-guided: also at intermediate timesteps

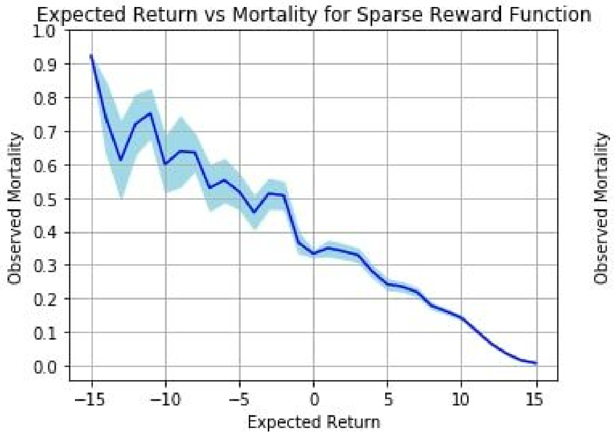
$$r(s_t, s_{t+1}) = C_0 1(s_{t+1}^{SOFA} = s_t^{SOFA} \ \& \ s_{t+1}^{SOFA} > 0) + C_1(s_{t+1}^{SOFA} - s_t^{SOFA}) + C_2 \tanh(s_{t+1}^{Lactate} - s_t^{Lactate})$$

- Finding optimal policy
  - Q-learning to find treatment policy that maximises expected **return** R (discounted sum of rewards)

Evaluation methodology

- Off-policy evaluation is hard!
- Use Doubly-Robust Value Estimator (Jiang and Li, 2015)
  - Accurately assess quality of learned policy using observational data from clinician actions
- Associate value estimate with mortality
  - Learn mapping between value estimates and mortality from observational data
  - Find value estimate of the learned policy
  - Use mapping to estimate expected mortality
- Qualitative policy evaluation

Results

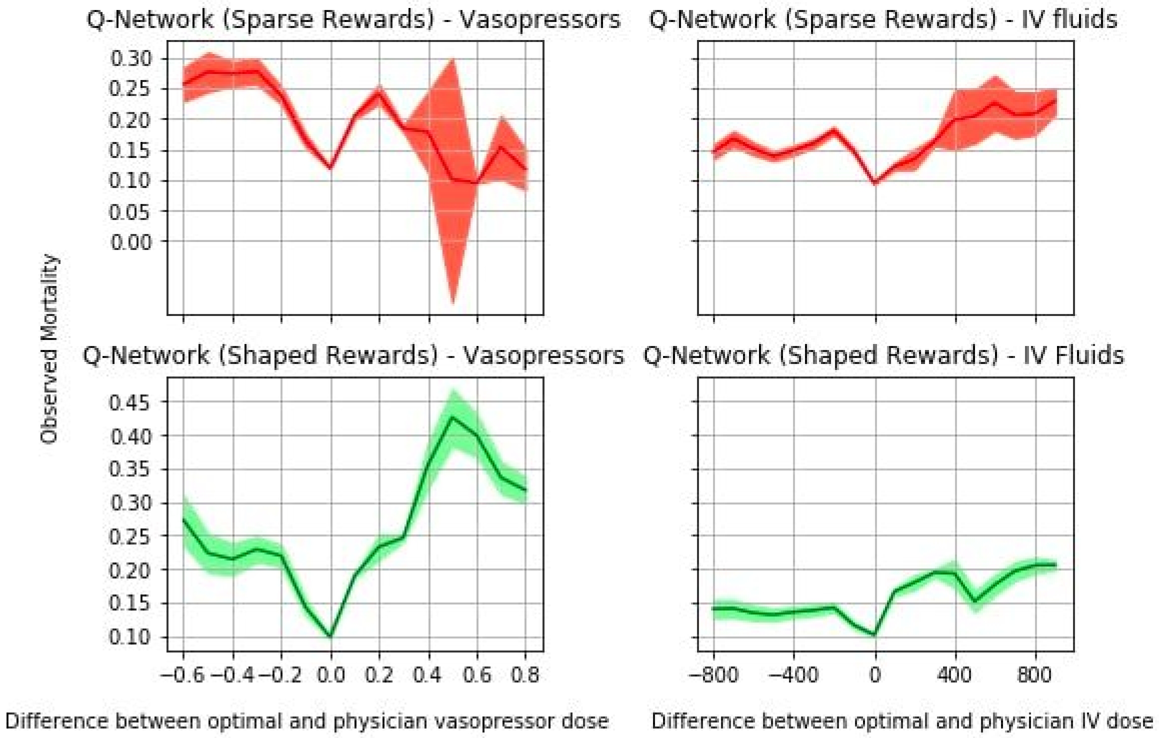


Policy	Expected Return	Estimated Mortality
Physician Sparse	11.17	$11.9 \pm 0.5\%$
Physician Shaped	11.04	$11.4 \pm 0.6\%$
Dueling DDQN Sparse	10.16	$12.8 \pm 0.5\%$
Dueling DDQN Shaped	13.3	$3.71 \pm 0.6\%$

Discussion and Future Work

- Learned improved treatment policies
  - Better than baseline
- Expected mortality is reduced
  - Our estimates are optimistic, but show promise
- Future work
  - Clinical insights into learned policies
  - Temporal aspects - recurrent Q networks, POMDPs

Other results



Other results