Guiding the management of sepsis with deep reinforcement learning

Stephen Pfohl
Stanford University
spfohl@stanford.edu

Ben Marafino
Stanford University
marafino@stanford.edu

1. Introduction

Sepsis is an acute medical condition which principally manifests as a disproportionate immune response to infection. It is characterized by widespread inflammation and coagulopathy, both of which lead to global tissue hypoperfusion and eventually, shock and end-stage organ failure [9]. Mortality rates among patients with sepsis are high: approximately 20 to 30% of patients with sepsis die, and this rate approaches 70% among patients experiencing septic shock [19]. As of 2014 in the United States, sepsis accounts for over 1.5 million cases and 250,000 deaths annually [2], and may be linked to as many as half of all in-hospital deaths nationwide [12].

Despite the grave toll of sepsis, no specific protocols for its treatment have yet been validated in large, randomized multi-center studies, nor are any drugs marketed specifically for treating sepsis. Current approaches are primarily supportive and focus on early cardiorespiratory resuscitation followed by antibiotic therapy, with the aim of maintaining hemodynamic stability and thus organ perfusion [1]. Such resuscitation can be achieved with the administration of intravenous fluid, with or without vaspressor medications which act to increase blood pressure and oxygen delivery to tissues, and is often carried out in a goal-directed manner.

For example, critical care physicians will commonly order fluid resuscitation for patients in the early stages of sepsis, with the goal of meeting predefined hemodynamic targets. However, while the increasing prevalence of such *goal-directed therapy* appears to be associated with declining sepsis mortality over the past decade [16], it remains an "one-size-fits-all" strategy. With the increased proliferation of electronic health records (EHRs), the current bounty of clinical data could could be leveraged to further personalize treatment protocols for patients with sepsis, potentially informing existing protocols and decision support systems, thus improving outcomes.

2. Context and related work

Data-driven approaches to tackling the problem of sepsis have attracted considerable interest in recent years. Notably, EHR data have been used to develop early-warning systems to predict sepsis onset among inpatients with high sensitivity and specificity [8]. However, the problem of sequential decision-making and decision support in sepsis, and in critical care medicine more generally, do not appear to be as well studied. Raghu *et al.* [18] applied deep reinforcement learning to learn and evaluate optimal treatment policies for sepsis using MIMIC-III data. Nemati *et al.* [15] also applied deep RL to learn heparin dosing policies, while Prasad *et al.* [17] used fitted Q-iteration (FQI) – a form of Q-learning – to learn policies for weaning ICU patients off of mechanical ventilation; both these approaches also used MIMIC data.

More broadly, reinforcement learning has also been applied to sequential decision problems in healthcare outside of the inpatient setting. Ernst *et al.* [5] used fitted Q-iteration to learn optimal treatment strategies for patients with HIV; such strategies are complex to learn, due to antiviral resistance, leading to patients being cycled on and off of therapy over yearly timescales. Escandell-Montero *et al.* [6] again used FQI to optimize erythropoietin-stimulating agent dosing among hemodialysis patients with anemia, with the goal of stabilizing patients' hemoglobin levels and minimizing side effects. Zhao *et al.* [22] used Q-learning to select from among strategies for treating non-small cell lung cancer from clinical trial data.

3. Dataset

3.1. Overview

For this project, the data will be drawn from the Medical Information Mart for Intensive Care-III (MIMIC-III) database, [11] which contains data collected from the ICU stays of 38,597 unique adult patients from between 2001 and 2012 at the Beth Israel Deaconess Medical Center (Boston, MA). MIMIC-III exhibits several characteristics that make it particularly attractive for this task. In particular, it allows for reasoning over temporally-dense streams of clinical time series of laboratory measurements, vital signs, bedside monitor waveforms, and fluid and medication infusions alongside contextually relevant demographic and di-

agnostic features in coded fields and free-text clinical notes. As MIMIC-III is publicly available and frequently used in clinical applications of machine learning [7], the results of this project may be reproduced by others. In line with recent clinical guidelines, we will use the Sepsis-3 consensus definition of sepsis [20] to identify patients who have developed sepsis and to refine our cohort.

In cases where the data in MIMIC-III may prove insufficient, the Stanford Translational Research Integrated Database Environment (STRIDE), comprising data for 2 million patients, may be used for unsupervised and transfer learning of meaningful embeddings of diagnoses (ICD-9), procedures (CPT), and medications (RxNorm, DRG), as in [3, 14].

4. Methods

4.1. Overview

We propose to primarily improve upon previous work in this domain. In particular, we aim to leverage model-free methods such as the Deep Q-Network (DQN) approach [21] to evaluate policies with off-line data stored in the electronic health record. We propose to implement all algorithms in Python using the PyTorch library.

As in [17, 18], we model the state of a patient as a partially observable Markov decision process (POMDP) over a continuous state space, where the primary state variables of interest are those that are most relevant over the hourly timescales on which decisions are made in critical care – namely, physiological measurements, laboratory tests, and fluid input/output events. In order to potentially improve the quality and generalizability of the learned policies, we propose several major extensions to the prior work:

1. **Continuous action spaces.** The action spaces defined in prior works have been limited in that they only consider a small, discretized grid of possible actions with low time resolution. In particular, in [18], the authors formulated the action space as a 5×5 grid over volumes of IV fluid and the maximum vasopressor dose administered over a 4-hour window. Expanding the action space to include a broader range of possible fluid or medication interventions will enable a richer space of potentially viable policies to be explored. Given that such an expansion of actions considered increases the size of the action space exponentially, it may be worthwhile to cast this problem as one of continuous control, [13] and to explore the development of an embedded action space [4].

2. **Reward shaping.** In prior work [18], the reward signal was defined only at the end of an episode, which was defined as patient discharge: a negative reward applied if the patient expired; otherwise, the reward was taken to be positive. This formulation induces a sparse reward signal – the *credit assignment problem* – and the learning process may be improved by modifying the reward function, for example, as in [17], to consider both short- and long-term rewards. Short-term rewards could be derived from surrogate measures of organ dysfunction and hypotension and take into account the the risk of transitions to severe sepsis and septic shock among patients in the early stages of sepsis. Possible such measurements include the shock index (HR/SBP), $ScvO_2$, and serum lactate levels, among others. Sensitivity analyses could be conducted to determine the extent of policy invariance (if any) as the result of this form of reward shaping.

3. **Unsupervised learning of patient histories as a source of adjunctive state data.** A patient's prior history of diagnoses, procedures, and medications may prove informative in guiding short-term treatment policies. Given that these features are heterogeneous, high-dimensional, and unlikely to change over hourly timescales, we propose that a low-dimensional, distributed representation of historical coded sequences be learned in an unsupervised manner from, for example, MIMIC-III or an auxiliary dataset such as STRIDE, and appended to the state vector used during reinforcement learning.

4.2. Evaluation

The evaluation of learned policies is intrinsically hard for off-policy models and hard in general for retrospective electronic health record data due to biases associated with its collection. As in [18], we aim to evaluate the learned policies with Doubly Robust Off-policy Value Evaluation [10] and compare the learned policies to the observed physician policies in MIMIC-III in terms of the ability to reduce mortality on held out data.

5. Conclusion

Even as treatment protocols continue to improve, the human toll of sepsis remains unacceptably high. The tools of reinforcement learning can be used to develop methods that aid clinical decision support for critical care medicine physicians treating sepsis, and to assist researchers in optimizing and evaluating candidate treatment strategies. We propose to build on, and improve upon, prior attempts to tackle this problem. In particular, our approach will be able to scale to employ continuous action spaces of treatment strategies, and more clinically relevant reward functions; and will incorporate latent representations of patient state. We hope that this work, taken with that of others, will ultimately lead to tools to make better decisions in order to reduce the burden of sepsis.

References

[1] Angus, D. and Poll, T.V.D. Severe sepsis and septic shock. *New England Journal of Medicine*, 369:840–51, 2008.

[2] Centers for Disease Control and Prevention. Data Reports: Sepsis. URL https://www.cdc.gov/sepsis/datareports/index.html.

[3] Choi, E. et al. Multi-layer Representation Learning for Medical Concepts. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pages 1495–1504, New York, New York, USA, 2016. ACM Press.

[4] Dulac-Arnold, G. et al. Deep Reinforcement Learning in Large Discrete Action Spaces. 12 2015.

[5] Ernst, D. et al. Clinical data based optimal STI strategies for HIV: a reinforcement learning approach. *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 667–672, 2006.

[6] Escandell-Montero, P. et al. Optimization of anemia treatment in hemodialysis patients via reinforcement learning. *Artificial Intelligence in Medicine*, 62(1):47–60, 9 2014.

[7] Harutyunyan, H. et al. Multitask Learning and Benchmarking with Clinical Time Series Data. *SIGKDD 2017*, (17): 1–16, 2017.

[8] Henry, K.E. et al. A targeted real-time early warning score (TREWScore) for septic shock. *Science Translational Medicine*, 7(299):122–299, 8 2015.

[9] Jacobi, J. Pathophysiology of sepsis. *American Journal of Health-System Pharmacy*, 59(SUPPL. 1):1435–1444, 2002.

[10] Jiang, N. and Li, L. Doubly Robust Off-policy Value Evaluation for Reinforcement Learning. 11 2015.

[11] Johnson, A.E. et al. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3:160035, 2016.

[12] Liu, V. et al. Hospital Deaths in Patients With Sepsis From 2 Independent Cohorts. *Jama*, 312(1):90, 2014.

[13] Metz, L. et al. Discrete Sequential Prediction of Continuous Actions for Deep RL. 5 2017.

[14] Miotto, R. et al. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific Reports*, 6:26094, 5 2016.

[15] Nemati, S., Ghassemi, M.M. and Clifford, G.D. Optimal medication dosing from suboptimal clinical examples: A deep reinforcement learning approach. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, volume 2016-Octob, pages 2978–2981. IEEE, 8 2016.

[16] Nguyen, H.B. et al. Early goal-directed therapy in severe sepsis and septic shock: insights and comparisons to ProCESS, ProMISe, and ARISE. *Critical Care*, 20(1):160, 12 2016.

[17] Prasad, N. et al. A Reinforcement Learning Approach to Weaning of Mechanical Ventilation in Intensive Care Units. 2017.

[18] Raghu, A. et al. Continuous State-Space Models for Optimal Sepsis Treatment: a Deep Reinforcement Learning Approach. *Mlhc*, 68, 2017.

[19] Russel, J.A. The current management of septic shock. *Minerva medica*, 99(5):431–58, 2008.

[20] Singer, M. et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *Jama*, 315(8): 801, 2016.

[21] Zhan, Y., Ammar, H.B. and Taylor, M.E. Theoretically-grounded policy advice from multiple teachers in reinforcement learning settings with applications to negative transfer. In *IJCAI International Joint Conference on Artificial Intelligence*, volume 2016-Janua, pages 2315–2321. Nature Research, 2 2016.

[22] Zhao, Y. et al. Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics*, 67(4):1422–1433, 12 2011.