Representation and Reinforcement Learning for Personalized Glycemic Control in Septic Patients

Wei-Hung Weng[1], Mingwu Gao[2], Ze He[3], Susu Yan[4], Peter Szolovits[1]

[1] CSAIL, MIT | [2] Philips Connected Sensing Venture | [3] Philips Research North America | [4] Massachusetts General Hospital

Abstract

Glycemic control is essential for critical care. However, it is a challenging task since there has been no study on personalized optimal strategy for glycemic control. This work aims to learn personalized optimal blood glucose trajectories for severely ill septic patients, and provide a data-driven policy as clinicians' decision making reference. We encoded patient states using a sparse autoencoder and adopted a reinforcement learning paradigm using policy iteration, to learn the optimal policy from data. We also estimated the expected return following the policy learned from the recorded blood glucose trajectories, which yielded a function indicates the relationship between real glucose values and the 90-day mortality rate. By following the learned optimal policy, the patients' estimated 90-day mortality rate could be reduced by 6.3%, from 31% to 24.7%. The result demonstrates that the reinforcement learning paradigm with appropriate patient state encoding can potentially optimize blood glucose management for critical care, and allow clinicians to design a personalized strategy for glycemic control in septic patients.

Background

Motivation

- Critically ill patients have the issue of poor glucose control, which includes the presence of dysglycemia and high glycemic variability.
- Current clinical practice follows the guidelines suggested by the NICE-Sugar trial to control the blood sugar level for cricital care.
- However, there are overwhelming variations in clinical conditions and physiological states among patients under critical care, and limit clinicians' ability to perform appropriate glycemic control. Clinicians sometimes even may not be able to take into account of the issue of glycemic control.
- To help clinicians better address the challenge of managing patients' glucose level, we need a personalized glycemic control strategy that can take into account of variations in patients' physiological and pathological states.

Reinforcement Learning (RL) in Clinical Domain

- RL is a potential approach for the scenario of sequential decision making with delayed reward or outcome.
- RL also has the ability to generate optimal strategies based on non-optimized training data.
- For treatment of schizophrenia [Shortreed 2011]. Heparin dosing problem [Nemati 2016]. Mechanical ventilation administration and weaning [Prasad 2016]. Sepsis treatment [Raghu 2017].
- Related to glycemic control, some studies utilize RL and inverse RL to design clinical trials and adjust clinical treatments [Bothe 2014].
- Less studies have utilized the RL approach to learn a better target laboratory values as references for clinical decision making.

Proposed Approach

- We hypothesized that the blood glucose values trajectory can be modeled as a Markov decision process (MDP).
- We explored the RL approach to learn the policy for optimal personalized blood glucose value range using retrospective data, and compared the expected return (mortality rate) of RL-simulated blood glucose value trajectories and the glucose trajectories in the real data.

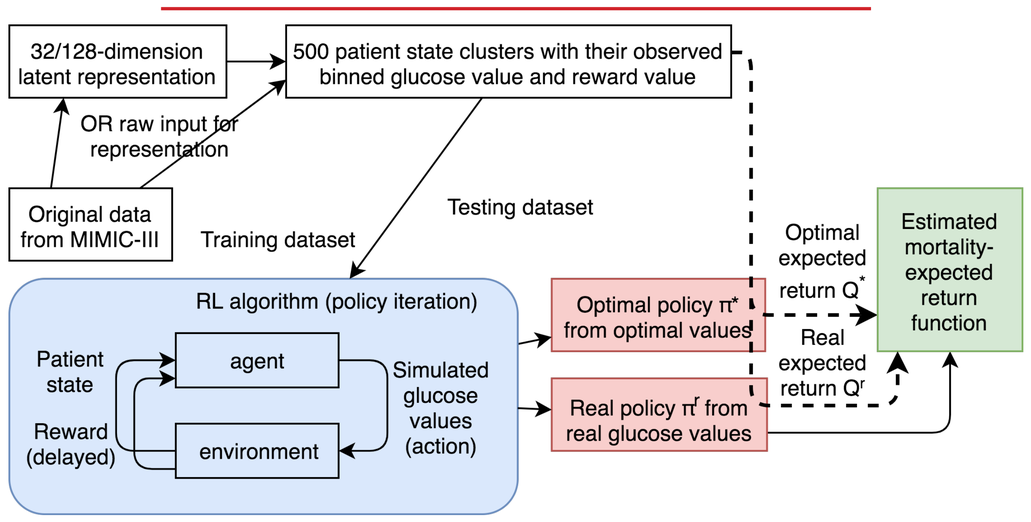
Objective

- Learn optimized blood glucose values trajectories, which are sequences of appropriate targeted glycemic ranges.
- The learned policy is intended as references for clinicians to adapt and optimize their care strategy, and to achieve better clinical outcomes.

Reference

- Bellman. Dynamic Programming. 1957.
- Bothe et al. The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. Expert Review of Medical Devices, 2014.
- Howard. Dynamic Programming and Markov Processes. 1960.
- Nemati et al. Optimal medication dosing from suboptimal clinical examples: A deep reinforcement learning approach. In IEEE Engineering in Medicine and Biology Society, 2016.
- Ng. Sparse autoencoder. 2011. https://web.stanford.edu/class/archive/cs/cs294a/cs294a.1104/sparseAutoencoder.pdf.
- Prasad et al. A reinforcement learning approach to weaning of mechanical ventilation in intensive care units. arXiv 2017.
- Raghu et al. Continuous state-space models for optimal sepsis treatment - a deep reinforcement learning approach. arXiv 2017.
- Shortreed et al. Murphy. Informing sequential clinical decision-making through reinforcement learning: an empirical study. Machine Learning 2010.
- Silver. Reinforcement Learning Lecture 3: Planning by Dynamic Programming. 2015.

Methods



Patient State Encoding

- Raw features vs. sparse autoencoder-encoded features [Ng 2011].
- 500 state clusters by k-means clustering.

Policy Evaluation / Iteration

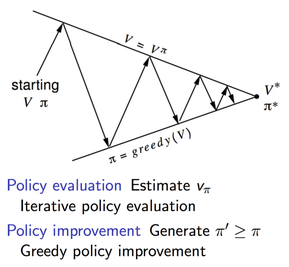
- Learn optimal policy & evaluate on real trajectories.
- 90-day mortality rate = f (expected return)
- Compute and compare the estimated mortality rate of real and optimal glucose trajectories obtain by RL-learned policy.

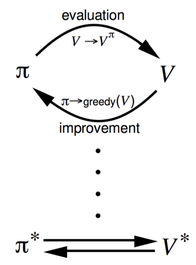
[Courtesy: David Silver]

$$v_{k+1}(s) = \sum_{a \in \mathcal{A}} \pi(a|s)(\mathcal{R}_s^a + \gamma \mathcal{P}_{ss'}^a v_k(s'))$$

$$Q^\pi(s,a) = \mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v^\pi(s')$$

$$\pi'(s) = \arg\max_{a \in \mathcal{A}} Q^\pi(s,a)$$





Experiment

Data Source and Study Cohort

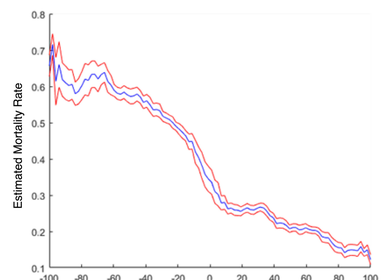
- 5,565 septic patients in MIMIC-III version 1.4.
- Sepsis-3 criteria to identify patients with sepsis.
- Exclude: age < 18, SOFA < 2, not first ICU admission.
- Diabetes: (1) ICD-9, (2) pre-admission HbA1c $> 7.0\%$, (3) admission medication, and (4) history of diabetes in the free text
- Data were collected at one hour interval.
- Missing values: linear and piecewise constant interpolation

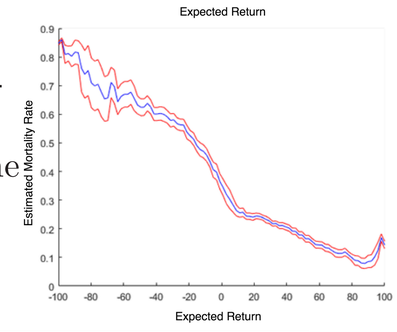
RL Settings

- Reward: 90-day mortality (+100 / -100)
- Action: discretized glucose levels (11 bins) as the proxy of real actions
- State: Total 46 normalized variables (patient level variables, blood glucose related variables, periodic vital signs)

Result

- The data distribution of learned expected return, which is the rescaled Q-value, is negatively correlated with mortality rate with high correlation value.
- The learned expected return reflects the real patient status well.
- The optimal policy learned by the policy iteration algorithm can potentially reduce around 6.3% of estimated mortality rate if we chose the appropriate patient state representations.





	Real trajectory		Optimal trajectory	
Representation	Expected return	Estimated mortality	Expected return	Estimated mortality
Raw features	10.04	31.00%	36.42	27.29%
Autoencoder	8.75	31.08%	32.49	24.75%

Conclusion

We utilized the RL algorithm with representation learning to learn the personalized optimal policy for better predicting target glucose levels from retrospective data. The proposed method may reduce the mortality rate of septic patients, and can potentially assist clinicians to optimize the real-time treatment strategy at dynamic patient state levels with a more accurate treatment goal, and ultimately leading to optimal clinical decisions. Future works include applying continuous state approach, different evaluations, and applying to different clinical decision problems.