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 glycemic trajectories for severely ill septic patients via learning data-driven policies of deciding optimal targeted blood glucose level as clinicians’ 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 glycemic trajectories, which yielded a function indicating the relationship between real blood glucose values and 90-day mortality rates. This suggests that the learned optimal policy could reduce the patients’ estimated 90-day mortality rate by 6.3%, from 31% to 24.7%. The result demonstrates that the reinforcement learning with appropriate patient state encoding can potentially provide optimal glycemic trajectories 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 critical care.

● However, there are overwhelming variations in clinical conditions and physiological states among patients under critical care. This limits clinicians’ ability to perform appropriate glycemic control. In addition, clinicians sometimes may not be able to consider 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 the 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.

● RL has been used for treatment of schizophrenia [Shortreed 2011]; heparin dosing problem [Nemati 2016]; mechanical ventilation administration and weaning [Prasad 2016]; and 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].

● Fewer studies have utilized the RL approach to learn a better target for glycemic control.

Proposed Approach and Objectives

● Learn optimal policy to simulate personalized optimal glycemic trajectories, which are sequences of appropriate glycemic targets.

● The simulated trajectories are intended as a reference for clinicians to decide their glycemic control strategy, and to achieve better clinical outcomes.

● We hypothesized that the patient states, glycemic values, and patient outcomes can be modeled as a Markov decision process (MDP).

● Action = the glycemic value that lead to real clinical action.

● We explored RL approach to learn the policy for learning personalized optimal glycemic trajectories, and compared the prognosis of the trajectories simulated by the optimal policy to the real trajectories.

● The learned policy is intended as references for clinicians to adapt and optimize their care strategy, and to achieve better clinical outcomes.

Method

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 obtained by RL-learned policy.

Results

● Fewer studies have utilized the RL approach to learn a better target for glycemic control.

● Missed 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

● RL algorithm can potentially reduce around 6.3% of estimated mortality rate if we chose the appropriate patient state representations.

Conclusion

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