Final Data Science Project (67814) - Analyzing Decision Making in ICUs

Guided by Prof. Michael Beil, Prof. Sigal Sviri and Mr. Gal Hyams

December 14, 2022

1 Abstract

Sepsis is a deadly critical that accounts for 30% of intensive care admissions[2]. The gold standard for stabilization of patients in septic shock is with the drug Norepinephrine (NOR). Guidelines specify stabilizing patients around target value of 65 mmHg Mean Arterial Pressure (MAP) using NOR[3], but do not account for specific patient trajectory and how to do it. Previous research showed high variability in sepsis treatment patterns [4].

In this work we show additional evidence for the variability in sepsis treatment across different Intensive Care Units (ICUs) in the same hospital using the MIMIC-IV[8] database. In particular, we show that for $^6/_8$ of MAP categories (bins) there exists a signifciant ($p < 0.05$) difference in mean NOR dosage between the Medical and Surgical ICUs (MICU, SICU). Given those differences, we looked to harness Offline Reinforcement Learning to establish a NOR dosage optimal policy that is seeking to stabilize patients around the recommended target value for MAP (65 mmHg).

Offline RL models for sepsis treatment have been developed in the past, most knowingly in 2018 work published by Komorowski et al [5]. However, those works focused on limiting mortatlity (long-term reward) and not short-term patient stabilization. Our model uses a simplified state space, which consists only of binned MAP measurements.

Training the model on MIMIC did not produce explainable results, due to the potentially and variable large temporal difference between observations (MAP measurements) and the actions (NOR doses). However, when trained with data from the eICU database[9], which have a higher temporal resolution (MAP every 5 minutes instead of every hour in MIMC), the model produced much more reasonble results, which had a higher Mutual Information between MAP value and dosage. Creating a reliable MAP stabilizing RL-model holds potential for improving the treatment given to sepsis patients in ICUs.

2 Introduction

2.1 Medical Background and Past Works

Sepsis is a life-threating condition that may arise as a response to an infection. It often requires ICU admission. Hence, it accounts for 30% of ICU admissions[1]. sepsis will often be charactherized with a drop in blood pressure, which in turn will reduce the blood supplied to vital organs. It is a critical condition, with an estimated mortality rate of 38% [2].

Norepinephrine (NOR)[1] is a drug that constricts the blood vessels in the body, thus increasing the blood pressure (drugs with this property are known as Vasopressors). NOR is considered to be the gold standard when treating sepsis, with a goal stabilization a parient's Mean Arterial Pressure (MAP[2]) near the target value of 65 mmHg. [3]

However, apart from this rule of thumb, there are no patient-level guidelines to instruct treatment for a specific patient. For example, even though there is an explicit target goal MAP, there is no regards for the specific patient's MAP trajectory to reach that goal. Thus, it shoud come as no surprise that there is a high variablity in sepsis treatment patterns [4]

It is only natural that attempts at harnessing Artifcial Intelligence for this task will come about. Such was the vast work published in 2018 Nature Medicine (*The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care,* **Komorowski et al** [5]). In this work, a Reinforcment Learning (see 2.4.3) based model was developed to suggest optimal treatment patterns: mainly doses of Vasopressors and IV fluids. The target of the model was to reduce the patients' mortality as much as possible, be it in the ICU or 90-day mortatlity. The model included a set of 48 patient variables: from demographics, to vitals signs and labratory results.

Although promisingly claiming: "*mortality was lowest in patients for whom clinicians' actual doses matched the AI decisions",* the work raised criticism and concern regarding applying the model in actual clincal use. For example in the following figure taken from the manuscript :

[1] NOR for short, may sometimes be refered to as noradrenaline (NA, NE)

[2] Mean Arterial Pressure is a measure of the average pressure in the arteries during one cardiac cycle. Estimated from the Systolic and Diastolic BP: MAP $= {}^{(2\cdot \text{ diastolic BP})+\text{systolic BP}]}/_{3}$

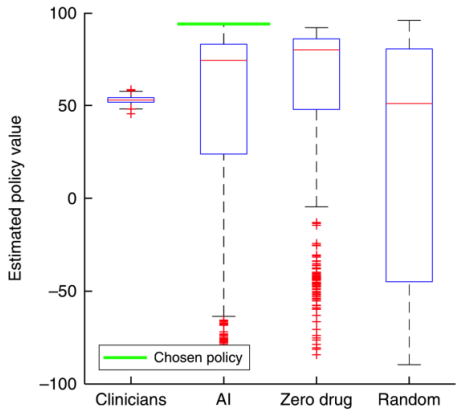

Figure 1: Estimated Policy Values boxplots comparing the AI model with clinicans, zero-drug policy and random policy [5].

it is apparent that the AI preformed better than human clinicians. However, it is alarming to see that on average, **no interference policy was actually better,** in terms of estimated policy value, than both the AI and the human clincian's policy.

This issue, as well as a recreation of the results and comparing the AI's suggestions to clincian's descisions for indvidual patient's trajectories are discussed in great lengths in the 2019 review by Jeter, Josef, Shashikumar, & Nemati [6] . One crucial keypoint of criticism discussed by Jeter et al is the sole focus of the 2018 work on long term rewards: hospital and 90-days morbidity. It disregarded the importance of intermediate rewards in the short term, in particular stablizing a patient's vitals and mainting their MAP in the recommended values, which bares more resemblance to a clinican's methodology in decsion making.

2.2 Data

The abovementioned works were trained on MIMIC-III (Medical Information Mart for Intensive Care) - a publicly available database of patients admitted to the Beth Israel Deaconess Medical Center (BIDMC) in Boston, USA, published in 2016[7]. In March 2021 a new version of MIMC was released - MIMIC-IV. It contains deidentified data of 50,048 patients admitted to an intensive care unit [8].

In addition, we used the eICU Collaborative Research Database. It is a large, comprehensive dataset of critical care data collected from over 200 ICUs at multiple hospitals across the United States, made available by Philips Healthcare. The database covers patients who were admitted to critical care units in 2014 and 2015, and was made publicly available at 2018 [9].

2.3 Objectives in our work

Our overall goal in this work was to understand desicion making in the ICU. The main objective is to model sequential decision-making, i.e. time-dependent trajectories of decisions. As a first step, we wanted to give additional evidence for the lack of uniform sepsis treatment patterns in the ICU. Secondly we sought to create a simple, interpretablemodel for decision making in the task of stabilizing a sepsis patient to a given Mean Arterial Blood Pressure (MAP) value.

2.3.1 Treatmeant Variability Assessment between ICU units

As mentioned above, evidence regarding the variability in treatment patterns has been published, like the work preformed by Bray et al (2020)[4]. This work analyzed the variability by comparing sepsis treatment protocols in different medical units in the UK's healthcare system (NHS Trusts). Considerable variation was found: from the factors that define a patient as "high risk", to treatment steps recommended in a patient's pathway.

Even though MIMIC does not contain any identifying data on which clinician gave a specific treatment, we seeked to assess the sepsis treatment variablity evident in the data. A possible mediator was to look at the treatment variability across between different ICU units. Two particular units interesed proposed by our project guides were the Medical ICU (MICU) and the Surgical ICU (SICU). Those two units differ in both their patient populations (SICU hosts patients before / after surgeries) and the training of medical staff (medicine vs surgery / anaesthesia).

We sought to find evidence for differences in treatment decisions - which we limited to the NOR dosage that patients with similar MAP recieved in the different units. Results and methodoloy are detailed in section 3.1 on page 6.

2.3.2 Basic MAP Stabilization Recommendation Model

We sought to create a simple interpreteble Reinforcement Learning model to recommend short term MAP stabilization of sepsis patients. Unlike the 2018 Nature's Clinician, this model will not seek to minimize long-term morbidity, but will look to keep a sepsis patient's MAP value at a stable range.

2.4 Mathematical Background

2.4.1 Permutations Test

Permutations Test is a statistical test useds to determine whether two observations were drawn from the same distribution. It is a non-parametric approach and does not require knowing the distribution of the data. The test is performed by randomly taking two groups of the distribution and computing the test statistic (e.g difference of means) with shuffeled labels. The p-value is calculated as the fraction of randomly labeled iterations that had a more extreme test statistic value than that of the original dataset.

2.4.2 Kolmogorov Smoirnov Test

Additional non-parametric test to determine whether two samples come from the same underlying distribution. The KST works by calculating CDFs (Cumulative Distribution Function) for the two samples and comparing the maximal difference between them. The maximal difference is known as the KS statistic: $D_{KS} = \sup_x |F(x) - G(x)|$, which can be used to generate p-values for the hypothesis $H_0 : F = G$ against $H_1 : F \neq G$.

2.4.3 Reinforcement Learning

Reinforcement Learning (RNL) is a branch of Machine Learning, which corresponds to a set of problems and solutions. The key factors which distinguish it from Supervised and Unsupervised Learning is that in RNL, an agent interacts with the enviroment in a dynamic way, looking to optimzie a reward function. A usual RNL process follows the procedure:

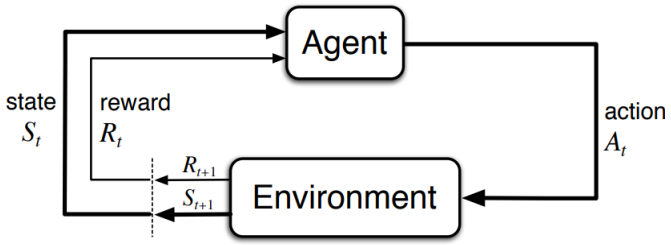

Figure 2: The agent–environment interaction in reinforcement learning. Taken from Sutton & Barto, 2018 [12]

During the training phase of RNL, the agent learns which action to take at a given state, to maximize the reward.
We will notate the set of all possible actions with a capital $\mathcal{A}$, and set of all possible states -$\mathcal{S}$. A few important definitions for our context (taken from [12]) are:

Definition 1 (Policy function)**.** a function that dictates which action should be taken at a given state, denoted as $\pi : \mathcal{S} \to \mathcal{A}$.

Definition 2 (Value function)**.** an estimation of the expected cumaltive reward if one followed the policy π starting from state $s \in \mathcal{S}$.

$$v_\pi(s) = E_\pi\left(\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s\right)$$

where E_π is to emphasize that the expected value is taken with regards to following policy π from state s. R_k is the return recieved at the k'th step of the process, and $\gamma \in (0, 1)$ is a discount factor - a hyperparameter used to discount the value of longer termed reward which ensures the convergence of the series).

Definition 3 (Action-Value function (Q-Function))**.** A function which denotes the value of taking action $a \in \mathcal{A}$ at state $s \in \mathcal{S}$ and following policy π thereafter:

$$q_\pi(s, a) = E_\pi\left(\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s, A_t = a\right)$$

3 Results

3.1 SICU - MICU Comparision

We divided the inputs data to groups (binning) according to their last blood pressure measurement before receiving the Norepinephrine:

$$MAP\ RANGES = \left[\underbrace{(0, 49)}_{b_0}, (50, 59), \ldots, (80, 89), \underbrace{(90, 200)}_{b_7}\right]$$

After doing so, we want to check for each group the hypothesis that the treatment will be the same across different units, against the conjecture that treatment varies. Let μ_{M,b_j} denote the expected value of NOR dose in the Medical ICU unit for the group with bp in the range denoted as b_j, and likewise for μ_{S,b_j}(for Surgical ICU).

We tested for each of the 8 hypothesis separately:

$$H_0 : \mu_{M,b_j} = \mu_{S,b_j}$$
$$H_1 : \mu_{M,b_j} \neq \mu_{S,b_j}$$

3.1.1 Permutations Test

We started with a baseline permutation test (15,000 permutations), which provided the following results:

	0	1	2	3	4	5	6	7
bp_range	(0, 49)	(50, 59)	(60, 64)	(65, 69)	(70, 74)	(75, 79)	(80, 89)	(90, 200)
p_val	0.825667	0.0012	0.138	0.012733	0.0116	0.010533	0.003933	0.0002
is pval < 0.05	False	***True***	False	***True***	***True***	***True***	***True***	***True***

Figure 3: P-Values for each bin testing $H_0 : \mu_{M,b_j} = \mu_{S,b_j}$against $H_1 : \mu_{M,b_j} \neq \mu_{S,b_j}$ using 15,000 permutations. See Notebook "permutation_test.ipynb" for reproduction .

It is evident that in almost all groups there has been a significant difference between the expected NOR dose between the units for patients with the same MAP values. This is true for all bp ranges apart from two: $(0, 49)$ which doesn't have enough samples for statistically significant p-value, and $(60, 64)$,, which is very close to the target value.

3.1.2 KST comparision

Following is a comparision of the PDF for each of MAP bins, alongside their KST p-values.

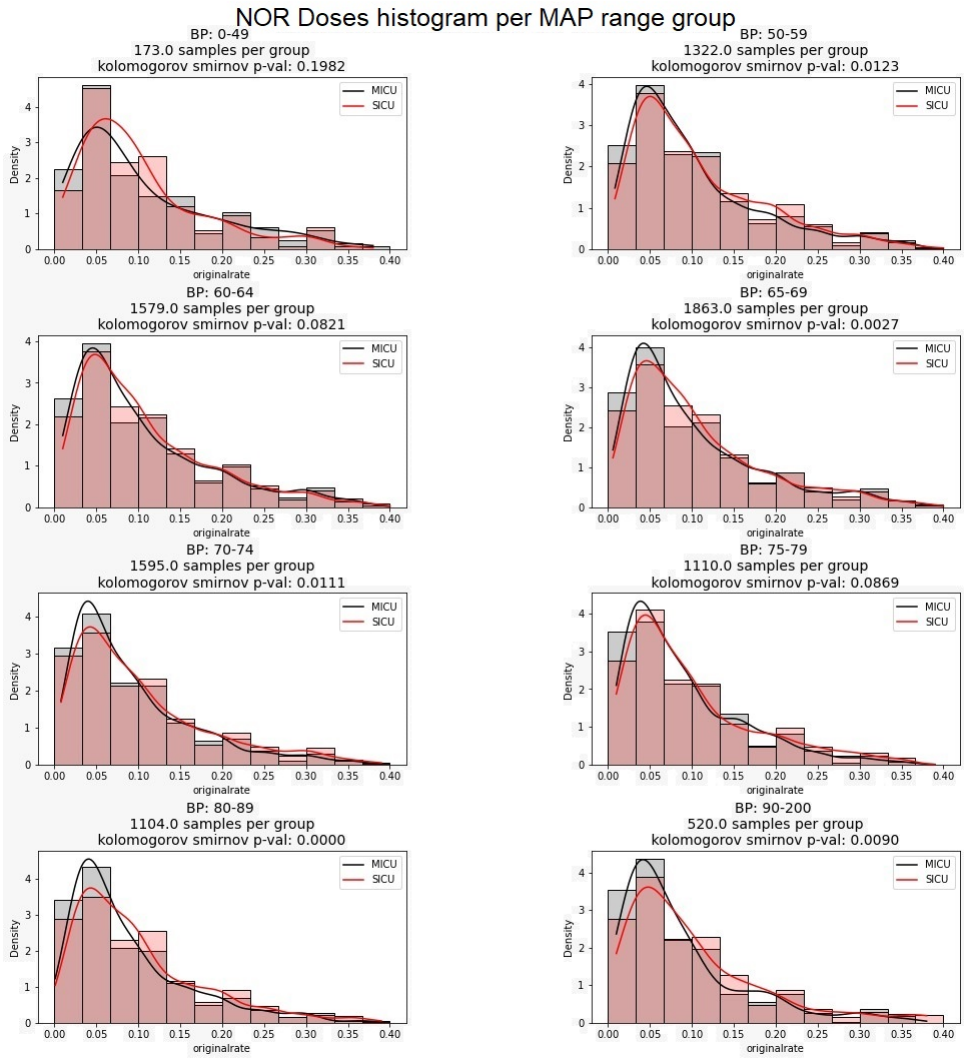

Figure 4: Probability Density Function of the original NOR rate in each BP group, with division to MICU and SICU, alongside KST p-val and the amount of samples in each group. For reproduction - run ks_test.ipynb notebook

Similarly to the p-values generated using the permutation test, in most groups $p - val < 0.05$, evident that there is a significant difference in the means of the NOR dosage given to patients with similar BP between the units.

Additional resolution achieved in this method is the shape of the distributions we see that the peak of the distribution of MICU tends to be lower than SICU, which indicates a higher range of dose values given by the doctors in MICU (In SICU doctors tend to give dosage which is close to 0.05 across all groups).

Two groups that did not have a significant difference are the first group (0-49) which doesn't contain enough samples for statistical significance, and two groups near the target area of 65 - $(60 - 64)$ and $(75 - 79)$.

In both units, the dosage variabilty changes according to the prior BP measurement. Around the edges the variability is higher, while in the 50-90 portion the variability is lower. This could explain the high p-value for the first bin (0-49).

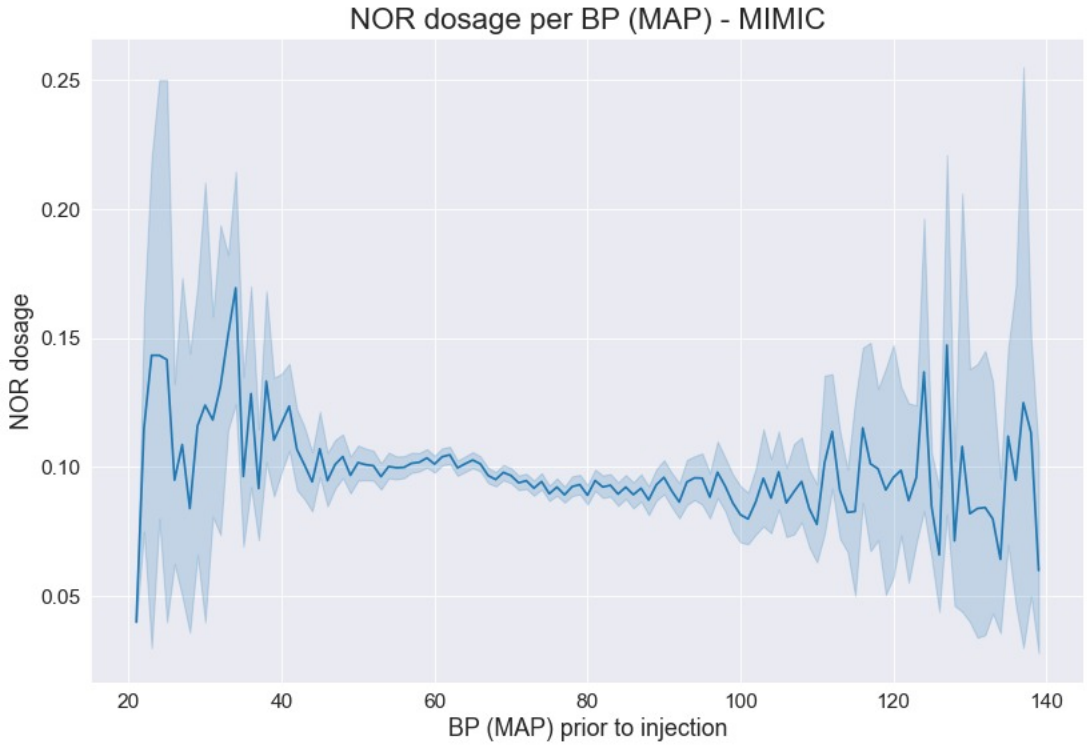

Figure 5: Illustration of the dosage variability per prior BP in the MIMIC database. For reproduction: 'databases_variability.ipynb'

3.2 Reinforcement Learning

3.2.1 Setting and Terminology

We've implemented the RNL framework for our task at hand: keeping the patient's MAP value at a prespecified target MAP. The formulation is natural for this problem:

- Agent - the caretaker, which decides which action to take.
- States - represented as the MAP bin of the patient. We defined the bins like before: $\mathcal{S} = \left[\underbrace{(0,49)}_{b_0}, (50,59), \ldots, (80,89), \underbrace{(90,200)}_{b_7}\right]$.
- Actions - NOR dosage. In order to define a finite set of possible doses, we decided to limit the set of possible actions to $\mathcal{A} = \{0, 0.01, \ldots 0.4\}$

- Reward function - we decided to define ahead of training the goal of maximizing the amount of time a patient spents in the recommended area of 65 MAP. Therfore, the reward function was defined to be the negative square distance from 65 $R_t = -\|s_t - 65\|^2$ (negative since we are looking to maximze the reward function in RNL algorithms).

- Policy - determines which NOR dosage to give to a patient at a given state. The agent's goal is to learn the optimal policy in terms of the cumlative reward generated from the reward function.

- Episode - The trajectory of states, actions and rewards for a single patient. Starting from his arrival to the ICU (or the initial data point available) to the last.

3.2.2 Offline RNL

In classic RNL setting, the agent has the freedom to knowingly make explorative non-optimal desicions with the goal of widening its knowledge on the enviroment. For our task and data this is neither possible (since we have a static dataset) nor ethical (doing so would mean knowingly taking wrong choices in patient treatment). The solution is Offline RNL - the agent learns *off-policy,* from dataset $\mathcal{D}$ that was collected prior training according to *any* policy, and does not interact with the enviroment in an online setting [13].

3.2.3 Markovian Assumption

As is the case in many RNL based models, our model assumes the decision process represented in the data adheres to the Markov property. This assumptios fits the reality of decisions about NOR in shock patients well. The caretaker is mainly, if not exclusively interested in adjusting the blood pressure with a drug (NOR) that does not accumulate.

Formally, given a trajectory of states (i.e MAP values) $s_1, \ldots, s_t$,and an action (NOR doseage) $a_t \in \{0, 0.01, \ldots, 0.4\}$,the next state s_{t+1} will depend solely on the action and the previous state. That is, the transition function $\tau : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ satifies:

$$\tau\left(s_{t+1}, a_t \mid s_0, \ldots, s_t\right) = \tau\left(s_{t+1}, a_t \mid s_t\right)$$

3.2.4 Patient Simulator and Transition Probabilites Estimation

Our enviroment was repesented via a ***Patient Simulator*** class. Each instance of this class has a MAP state, and can be interacted with via a *give_ dose* API . *Give_ dose* will change the patient's MAP according to an estimation of the transition probabilities - $\hat{\tau}$. A natural way to estimate $\hat{\tau}$ is by sampling the training data after conditioning. We'll estimate $\hat{\tau}$ as

the marginal probability of a state s_{t+1} given the previous s_t and action a_t. Formally, given dataset $\mathcal{D}$, define the subsets:

Definition 4. $\mathcal{D}_{*|s_t,a_t}$- All entries in the dataset where action a_t was taken following the state s_t.

Definition 5. $\mathcal{D}_{s_{t+1}|s_t,a_t}$- Entries in the dataset that are equal to s_{t+1}, and followed the tuple (s_t, a_t).

The natural estimator for the transition probability will be:

$$\hat{\tau} = (s_{t+1}, a_t \mid s_t) = \frac{\left|\mathcal{D}_{s_{t+1}|s_t,a_t}\right|}{\left|\mathcal{D}_{*|s_t,a_t}\right|}$$

Remark 1. This estimator is a generalization of the MLE for transition matrix in a classic Markov Chain (without actions). It can be proven that in a Markov Chain model $X_1, \ldots, X_n$ where one looks to find the transition probability $p_{ij} = P\left(X_{t+1} = j \mid X_t = i\right)$,the MLE estimator is $\hat{p}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$, where n_{ij} is the number of instances of the tuple of states (X_i, X_j) across trajectories, and the summation in the denominator is across all possible states [15]. If we were to define a new state set from the cartesian product of the states and actions sets $\mathcal{S}' := \{(s, a) \in \mathcal{S} \times \mathcal{A}\}$, our $\hat{\tau}$ will converge with the MLE estimator for the transition probability in that model.

In practice, ***Patient Simulator*** does not calculate $\hat{\tau}$ explicitly. It samples the next state from the group $\mathcal{D}_{*|s_t,a_t}$, which implies the transition probability $\hat{\tau}$ over the states. Therefore, states for which $\mathcal{D}_{s_{t+1}|s_t,a_t} = \emptyset$ will have zero probability.

3.2.5 Initialization and Termination of Epsiodes

Two important cases left to be considered were how to start and end an episode in the patient simulator. In order to be as close to the data as possible, we decided to sample the starting point randomly from all states which appear as the starting point for a patient in $\mathcal{D}$. Likewise, when we sample using the transition function $\hat{\tau}$ a state s_{t+1} which happens to be a terminal state for a particular patient, we will terminate the episode.

3.2.6 Model Training Algorithm

We choose to train our RNL model using Q-Learning Monte Carlo algorithm. The main idea of the algorithm is to iterativly estimate the state-action function $q_\pi(s, a)$, calculate the optimal policy π using it, play an episode using the policy π, and update the q function using the newly aquired data from the episode. We also used the *first-visit variation* of the

algorithm, which is an optimization variant that skips re-evaluating the same state-action pairs in the same episode (for effiency). This variation is widely studied, and is referenced in page 128 in [12].
The algorithm assumes (and takes as an input) an episode simulator which is able to generate a series of states, action and following rewards : $\{(s_t, a_t, r_t)\}_{t=0}^{k}$ (s_k being the terminal state). In our setting, the Episode simulator is implemented via the class *Patient Simulator*. Follows is a psedo-code of the training algorithm:

Algorithm 1 First Visit Q-Learning Monte Carlo

Input: Episode Simulator, N-Episodes (integer)
Output: Optimal Policy π^* and Estimated Value function for that policy $\hat{v}_{\pi^*}$

1. **Initialization:**
   (a) Initialize π_0 randomly.
   (b) Initialize Q function that evaluates each state-action q value to be 0:
   $$q = \{(s, a) \to 0 \mid (s, a) \in \mathcal{S} \times \mathcal{A}\}$$
   (c) Initialize empty returns hashtable, that maps each pair to a list of returns produced from this pair.
   $$\boldsymbol{returns} = \{(s, a) \to [\ldots] \mid (s, a) \in \mathcal{S} \times \mathcal{A}\}$$
2. **Iteration:** For each $i = 1, ..., N - Episodes$:
   (a) Initialize an empty hashtable of seen_before_states $\{\ldots\}$.
   (b) Generate a series of states, actions and rewards $\{(s_t, a_t, r_t)\}_{t=0}^{k}$ via the Episode Simulator with policy π_{i-1}.
   (c) Iterate over the epsiode's trajectory - For $t = 0, \ldots, k$:
      i. Check if the pair $(s_t, a_t) \in$seen_before_states. If it has been seen move to next t. Else add (s_t, a_t) to seen_before_states .
      ii. Add r_t to the returns hashtable list that matches the pair (s_t, a_t)
      iii. Update $q(s, a)$ to be the average of returns list that matched the pair (s_t, a_t) including return r_t.
   (d) Update the policy $\pi_i(s) = arg\max_{a \in \mathcal{A}} \{q(s, a)\}$.
3. **Termination**: Final policy is $\pi_{N-episodes}$. State-value function can be inferred using $q_{\pi_{N-episodes}}$: $v_{\pi_{N-episodes}}(s) = \max\{q(s, a) \mid a \in \mathcal{A}\}$. Algorith returns the pair: $\pi_{N-episodes}, v_{\pi_{N-episodes}}$

We used an existing implementation by Colin Skow[14] and fitted it to our settings' needs.

3.2.7 Results From Training the Model

After filtering the MIMIC data to clear decisions (see 5.1.6) we ran the Q-Learning Monte Carlo Algorithm for 1000 episodes, and ploted the resulting policy & value functions:

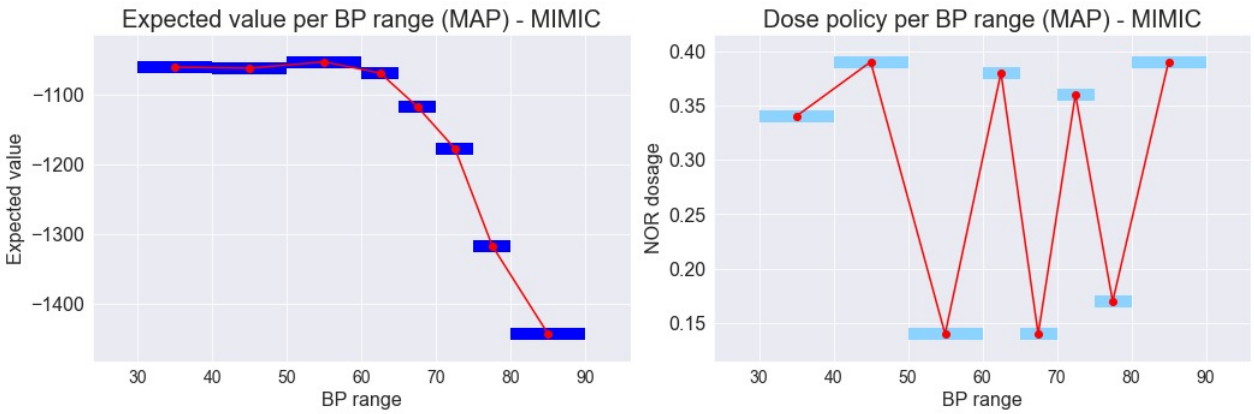

Figure 6: MIMIC trained: Estimated BP Value function after $1,000$ iterations, with discount factor $\gamma = 0.9$. For reproduciton code see rnl_resuls.ipynb notebook.

The policy appears to be random. The Estimated category value has more order to it, reflecting the reward function. The lower bins have more value since there is a higher probability to get to the target of 65, while lowset estimated bin value is around 80-90, reflecting areas were the probability to go back to the target value is the lowest.

3.3 eICU RNL

The eICU offers a higher time resolution of MAP measurements. While patients in MIMIC have an average of one measurement per hour, in the eICU time resolution is every five minutes. As the number of samples available in the interval between each NOR injection is larger, the variance before each NOR injection is smaller:

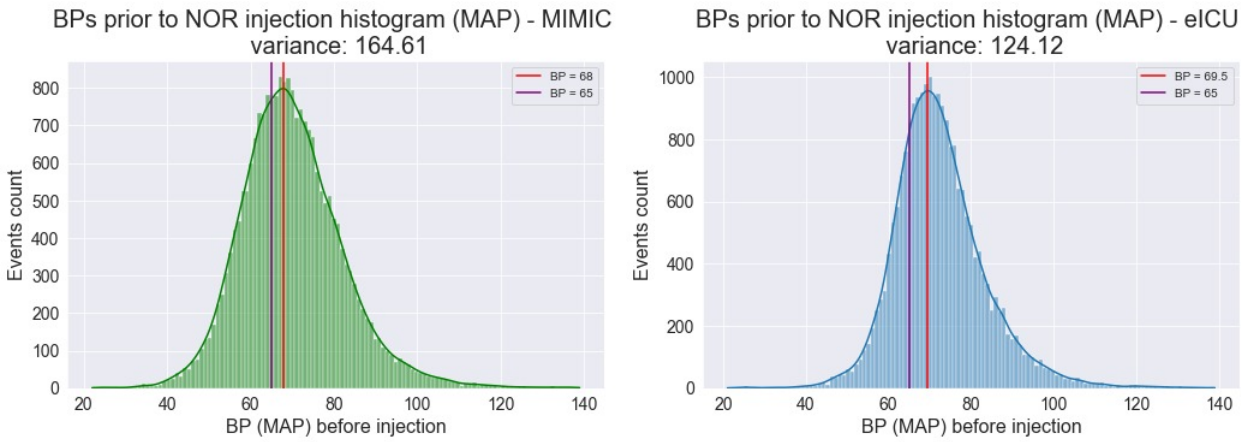

Figure 7: Histogram of BPs Prior to NOR injection, comparision of MIMIC and eICU. Both have a gaussian bell curve with similar means (marked in red). However the variance in eICU is smaller, which hints it may be more suitable for learning using simulation in the RNL setting. For reproducion run: 'databases_variability.ipynb'

This encouraged us to process the eICU data to fit our RNL framework as well, and train our model using this data.

The results were more plausible than those obtained from MIMIC:

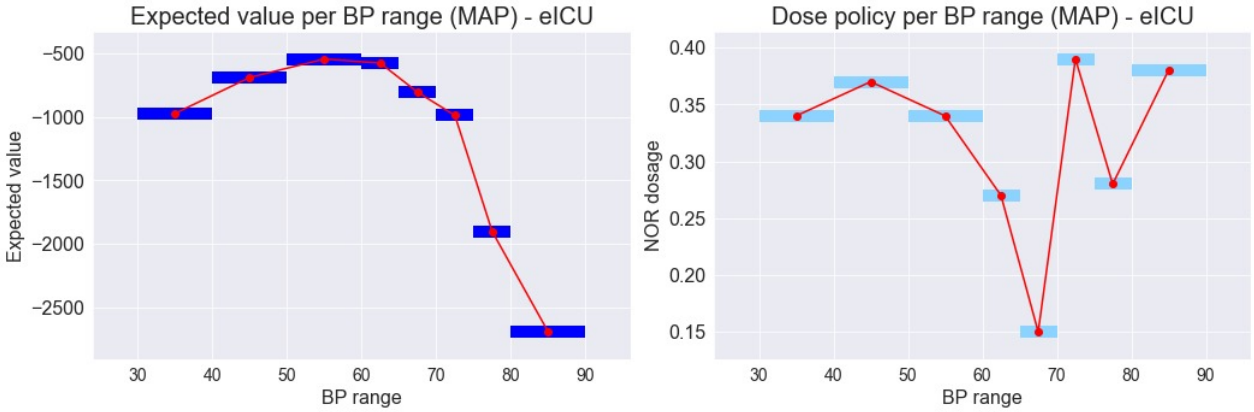

Figure 8: eICU trained: Estimated BP Value function after $1,000$ iterations, with discount factor $\gamma = 0.9$. For reproduciton code see rnl_results.ipynb notebook

It looks that the policy that was obatined with eICU is more reasonable than MIMIC: Recommends higher dosage in the low BP range, and as we get close to 65 lower doses. The exception is the right handside of the BP ranges, where the model suggests higher doses. This is unreasonable and may be due to symmetric reward function chosen (see 4).

3.3.1 Evaluation Metric

While it is clear to the observer , it is hard to quantify the *"reasonablity"* of the eICU policy vs the MIMIC. One attempt at quantifiying this is to look at the relation between the BP range and the recommended dosage. We would expect that a "good" policy will be closely related to the BP value observed. However, the relation should not be linear and may not be even monotonic. A measure of the information one variable provides on the other is Mutual Information. Defined as :

$$MI(X,Y) = \sum_{x} \in X \sum_{y} \in Y p(x,y) log \frac{p(x,y)}{p(x)p(y)}$$

MI quantifies the amount of information that one random variable contains about another. More specifically, it measures the reduction in uncertainty about one random variable given knowledge of the other. In our case, the MI of BP and the MIMIC policy was 0.801, while the MI of BP and the eICU policy was 1.147, suggesting that for eICU, BP values provided a more significant uncertainty reduction regarding the policy (see rnl_results.ipynb notebook for code).

4 Discussion

4.1 MICU - SICU Comparision

As expected, we have seen major differences between the doses given to patients with similar MAP. In both the KST and Permutations test, most groups had significant differences, apart from the bin which didn't have enough samples - $(0, 49)$, and bins near the target value. This finding supports previous research on sepsis treatment variability, showing that there is no uniform policy for treatment, even within the same hospital. Examing the CDF of NOR doses for each MAP group provided additional insights to the differences between SICU and MICU, and showed that there is higher treatment variability in the MICU on patients with the similar MAP.

It is important to note that the tests split the patients solely on their MAP values, and did not split to internal cohorts within each bin. It is only natural that patients in SICU and MICU will have different symptoms and physiology, and therefore may require different treatment. A finer test of treatment variability could have been comparing the decisions different doctors made within the same unit, and perfably in treating the same patient. This could not have been done using MIMIC, as it doesn't hold any data on the doctor who made the desicion.

The high variability in sepsis treatment, even within the same hospital, indicates potential for the effectiveness of an AI-based model. Such model will not be limited to the treating expirence a single human clinican can obtain in a lifetime, and could make desicion based on huge train datasets.

4.2 RNL

We've estabilshed a simple, short-term RNL model, looking to stabilize a patient near the target MAP value of 65. Training the model on MIMIC data provided poor results, which did not pass the reasonability test. The model makes the Markovian assumption, which likely does not hold in low time resolution data like MIMIC (BP mesaurement every hour). However, training the model on the eICU data, which had a higher time resolution (mesaurement every 5 min on average) showed much more plausible results, and generated a better Mutual Information between the BP and Policy recommended dose.

The initial target of our study was to preform behavioral analysis - find out what is the target MAP that doctors in different units aim to. In this study we found that **in order to learn behaviors, we need a close temporal resolution** - between the MAP value observed and the desicion (NOR dose) that followed. As a proof of this, we showed that learning from a higher resolution data gives more plausible policies.

Focusing on a short-term model holds potential for clincal applications. It uses the key factor crucial for maintaining the patient alive in the state of Septic shock - the MAP, and could be relevant to situations where only the short goal of stabilizing the patient is important. Simple models are also easier to train, as the state space $\mathcal{S}$ is smaller ($\mathcal{S}$ grows exponentially with each variable added).

4.3 Conclusions Based on the Results

In conclusion, there is evidence for differences in sepsis treatment patterns between MICU and SICU units. A necessary condition for creating a baseline AI model that is able to recommend sepsis treatment regimes is to have a high temporal resolution between observations (MAP measurements) and desicions (NOR injections).

In this regard, eICU is more suited for the task than MIMIC, as the temporal resolution it provides is higher. An indicator for this is the much more plausible results obtained when training the RNL model on the eICU data. Creating a MAP stabilizing AI-model holds potential for improving the treatment given to sepsis patients at ICUs.

4.4 Limitations

1. **Markovian Assumtpion**: the model assumed the Markovity of the decision process. Although simplifying, this isn't an unreasonable assumption: the half life of NOR is about a minute, so the decisions' results are instant. However, as discussed before, **this only holds if the data reflects this time resolution**.

2. **Patient filtering:** We filtered patients with comorbidities as well as patients with vasopressor medcine overlaps (see 5.1.1), to prevent biases in decision-making (change of dose decision that was triggered by a different medicine in effect). In a practical setting this assumption will not hold.

3. **Patient Simulation:** As discussed, we trained the model *in-silico*. The patient simulation did not limit the trajectory to an actual patient, but used bins of all the patients in the data. Thus, the simulator was invariant to the fact that observations came from different patients, but only took their MAP bin into account. Thus the simulator could have also been biased by patients with a lot of measurements.

4. **Reward function:** Our reward function was a symmetric ℓ_2 distance from the target value of 65 MAP. In practice, no such equilibrium exists: MAP values of 45 and 85 are very different from a clinical standpoint and should not be rewarded equally.

4.5 Future Work

The model is far from being clinically relevant and applicable. A crucial first step will be to compare the model's suggestions on actual patients to the decisions made by a clinician, and check if they are within a reasonable range. Next, one should dive into the areas of disagreement between the clinician and the AI, and use the hindsight test to verify who was right (e.g - AI suggested higher dose, the patient's BP did not rise, clinician had to adjust and vice-versa).

In addition, several important adjustments are required:

1. **Adding variables to the model :**

   (a) Add the fluids the patient receives as an additional variable.

   (b) Take comorbidities into account (as well as other medicines that are taken simultaneously).

2. **Markovian Relaxation**: Consider taking several MAP entries back, not just most recent. Doing so could relax the strong Markovian assumption.

3. **Fine-tune the RNL hyperparameters**: Number of Episodes, γ (discount factor), ε (percentage of explorative decisions in the training phases).

4. **Smooth Sampling in Simulation**. During simulation phase, instead of sampling from patients with the exact same bins, consider sampling from all bins, only with lower probability.

5. **Different reward function**. We choose for simplicity a symmetric ℓ_2 distance from 65, which limits the model (see 4). A more accurate reward function should be defined to better reflect the clincian's point of view in terms of which MAP value is the most stable.

6. **Termination conditions**: Terminate patients simulation that reach unreasonable MAP values, not only based on the data.

7. **Modern Neural Network Architecture:** The model could be reshaped to match the modern NN approaches for handling sequential data - the Transformer model. It could include attention (for faster computations) and long term memory, which could also help to relax the Markovian assumption.

After a stable model passes the human comparision test, it could be used for behavioral analysis. In particular, Inverse RNL: Inferring which target values achieve the most reward, which is relevant in understanding the current treatment patterns used in practice.

5 Resources and Methods

5.1 Preprocessing

In order to work nimbly, we created a subset of the relevant data from all the possible data that MIMIC-IV offers. The subset was creaded by applying the following folters:

5.1.1 Patients Filters

1. **Sepsis ICD codes** - Keep only the following ICD (International Classification of Disease) codes:

    - 99592 - Severe sepsis, ICD version 9
    - 99591 - Sepsis, ICD version 9
    - R652 - Severe sepsis, ICD vesion 10

- R6520 - Severe sepsis without septic shock, ICD version 10
- R6521 - Severe sepsis with septic shock, ICD version 10

2. **Age** - Keep only patients between ages 20 to 90.

5.1.2 Stays Filters

1. **Unit** - Keep only stays that start and end in the same unit.
2. **Length** - Keep only stays of at least one day.

5.1.3 Chart Events Filters

1. **HR and BP events** - Keep only the events of blood pressure and heart rate, with the following MIMIC codes:
    - 225312 - ART BP Mean
    - 220052 - Arterial Blood Pressure mean
    - 220181 - Non Invasive Blood Pressure mean
    - 220045 - Heart Rate

5.1.4 Input Events Filters

1. Norepinephrine and its alternatives:
    - 221906 - Norepinephrine
    - 221662 - Dopamine
    - 221289 - Epinephrine
    - 221749 - Phenylephrine
    - 229617 - Epinephrine
    - 229630 - Phenylephrine (50/250)
    - 229631 - Phenylephrine (200/250)_OLD_1
    - 229632 - Phenylephrine (200/250)
    - 229789 - Phenylephrine (Intubation)
    - 222315 - Vasopressin

5.1.5 Interim summary of data filters

By applying all the filters we got 754 MB that we could load into the memory freely instead of 70 GB of MIMIC data.

5.1.6 Decisions Filter

The main part of the project is analyzing the decision making of the medical teams. Since the most important part is the human decision, The idea of decision filter is to keep only events of doses that were taken by a medical authority, and were not recorded automatically. In contrast to the previous filters, in which data was kept because of accurate condition that was fulfilled, the case of decisions filter is different. There is no field in the data that indicates if the dose recorded due to a decision of a human or not.
Since there is no such field, we need to define what decision is by other propereties we do have:

1. **FinishedRunning status description** - there are 5 status descriptions in the input event files, and one of them is "FinishedRunning". by the documentation of MIMIC-IV that status means: "The delivery of the item has finished (most frequently, the bag containing the compound is empty)"[10]. It infers that those records are not an active decision of a human. Deeper investigation of those records shows that indeed the rate of those records is not one that human will decide to type manually (like 0.080032155, 0.10005264, 0.3417207). Therefore, we defined those doses as non-desicion doses.

2. **Short gap between two doses that the first was stopped or paused** - there are specific status descriptions in the input event files that implies that the dose was paused or stopped by the caretaker. Sometimes, there is a record of a successive dose that comes immediately after and with same or very close dose rate (for example 0.2 and 1.9999). In this case we defined the successive dose as non-decision dose.

3. **Successive doses with almost same dose rate -** there are cases in which the first dose have an endtime that is the same as the starttime of the successive dose, and in addition the dose rate is almost the same. In those cases we defined the successive dose as non-decision dose.

5.1.7 Medicine overlap filter

In many cases a patient get more than one vasopressors simultaneously. This situation leads to a difference in the dose rate of Norepinphrine between two patients with the same symptoms. In order to solve this issue, we filtered out cases in which there are overlap between

vasopressors. The vasopressors that we take into account are the those were mentioned previously. 5.1.4

References

[1] Bennett SR. Sepsis in the intensive care unit. Surgery (Oxf). 2015 Nov;33(11):565-571. doi: 10.1016/j.mpsur.2015.08.002. Epub 2015 Oct 9. PMID: 32287818; PMCID: PMC7143675.

[2] Martin CM, Priestap F, Fisher H, Fowler RA, Heyland DK, Keenan SP, Longo CJ, Morrison T, Bentley D, Antman N; STAR Registry Investigators. A prospective, observational registry of patients with severe sepsis: the Canadian sepsis Treatment and Response Registry. Crit Care Med. 2009 Jan;37(1):81-8. doi: 10.1097/CCM.0b013e31819285f0. PMID: 19050636.

[3] Dugar S, Choudhary C, Duggal A. Sepsis and septic shock: Guideline-based management. Cleve Clin J Med. 2020 Jan;87(1):53-64. doi: 10.3949/ccjm.87a.18143. Epub 2020 Jan 2. PMID: 31990655.

[4] Bray A, Kampouraki E, Winter A, Jesuthasan A, Messer B, Graziadio S. High Variability in Sepsis Guidelines in UK: Why Does It Matter? Int J Environ Res Public Health. 2020 Mar 19;17(6):2026. doi: 10.3390/ijerph17062026. PMID: 32204395; PMCID: PMC7142432.

[5] Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med. 2018 Nov;24(11):1716-1720. doi: 10.1038/s41591-018-0213-5. Epub 2018 Oct 22. PMID: 30349085.

[6] Jeter, R., Josef, C., Shashikumar, S., & Nemati, S. (2019). Does the" Artificial Intelligence Clinician" learn optimal treatment strategies for Sepsis in intensive care?. arXiv preprint arXiv:1902.03271.

[7] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.

[8] Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2022). MIMIC-IV (version 2.1). PhysioNet. https://doi.org/10.13026/rrgf-xw32.

[9] The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG and Badawi O. Scientific Data (2018). DOI: http://dx.doi.org/10.1038/sdata.2018.178. Available from: https://www.nature.com/articles/sdata201817

[10] MIMIC-IV documentation. Note: https://mimic.mit.edu/docs/iv/modules/icu/inputevents/

[11] Toufen C Jr, Franca SA, Okamoto VN, Salge JM, Carvalho CR. Infection as an independent risk factor for mortality in the surgical intensive care unit. Clinics (Sao Paulo). 2013;68(8):1103-8. doi: 10.6061/clinics/2013(08)07. PMID: 24037005; PMCID: PMC3752640.

[12] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

[13] Levine, S., Kumar, A., Tucker, G., & Fu, J. (2020). Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems. ArXiv, abs/2005.01643.

[14] Skow, Colin, Coding Demos from the School of AI's Move37 Course https://github.com/colinskow/move37.

[15] Carnegie Malon University, Cosma Shalizi, Statistics 36-462, Spring 2009, Lecture 6 https://www.stat.cmu.edu/~cshalizi/462/lectures/06/markov-mle.pdf