Optimizing Respiratory Management Using Markov Decision Processes: A Reinforcement Learning Approach on the MIMIC-IV Dataset

Vakula Venkatesh
Department of Mechanical Engineering
Stanford University
vakulav@stanford.edu

Amanda Rodriguez
Department of Biomedical Data Science
Stanford University
mandirod@stanford.edu

Abstract

In a critical care setting, maintaining vitals such as blood oxygen saturation (spO2) is essential to ensure positive patient outcomes. We present an adaptive decision-making framework for managing respiratory care, specifically targeting the optimal regulation of oxygen levels and respiratory rate. We use the MIMIC-IV data set[3], a freely accessible data set of electronic health records (EHR), to develop decision making policies that maximize patient safety and efficiency in oxygen management. Using Q-learning, we model the system as a Markov Decision Process (MDP) and evaluate policies for maximizing the expected reward. Our system uses information about SpO2 and frequency of respiratory movements as real-time data inputs to formulate optimal intervention measures.

1 Introduction

Effective decision-making is especially crucial in Intensive Care Units (ICUs) where patient conditions can change rapidly. Parameters such as blood oxygen saturation (SpO2) are central to respiratory management. Traditionally, medical decisions are made by clinicians based on their expertise and experience, but machine learning algorithms can be applied to automate and optimize these decisions.

Markov Decision Processes (MDPs) are a natural framework for decision-making in uncertain environments. MDPs provide a mathematical model for decision-making where outcomes are partially random and partially under the control of a decision-maker. In this paper, we explore the application of MDPs and reinforcement learning techniques for managing respiratory parameters such as SpO2 (oxygen saturation) and respiratory rate (RR) in critically ill patients. Specifically, we use the MIMIC-IV dataset[3] to simulate decision-making processe to determine the best actions based on observed physiological states to maintain SpO2 and RR within optimal ranges. The methodology uses Q-learning to identify a policy that maximizes long-term rewards by improving patient outcomes.

2 Related Work

Applying reinforcement learning to medical decision-making has is gaining more traction, but these techniques have found their place in medicine for years. Desautels et al. (2016)[1] developed machine learning models for predicting sepsis, a life-threatening condition in ICU patients, using the MIMIC-III dataset[2]. Their work demonstrated that reinforcement learning and other machine learning techniques could improve critical care predictions. Even earlier, Schaefer et al. (2005)[6] explored the use of MDPs for modeling medical treatment, particularly in optimizing drug dosages and ventilation settings. Additionally, Kochenderfer et al. (2022)[4] provided a comprehensive

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

framework for decision-making algorithms, with a focus on applications in medical environments, which aligns well with our work.

Our research extends these previous efforts by applying MDPs specifically to SpO2 management in the ICU, with an approach based on Q-learning for policy optimization. The MIMIC-IV dataset provides a valuable resource for simulating decision-making, as it includes detailed physiological measurements and clinical notes for ICU patients.

3 Methodology

3.1 Data Preprocessing

Analysis began with preprocessing SpO2 and RR data from the MIMIC-IV dataset. We filtered the data based on specific physiological parameters to isolate SpO2 and RR values. Then, we cleaned the to ensure that only valid SpO2 and RR values are retained. We merged the cleaned data to combine SpO2 and RR information by matching on patient identifiers. We lastly refined the dataset by removing unnecessary columns and renaming variables for clarity.

3.2 Markov Decision Process (MDP) Model

We model the decision-making process as a Markov Decision Process (MDP). The state space is defined as a combination of SpO2 and respiratory rate, represented as pairs of values. The possible actions are as follows:

- `increase_fio2` - Increase the fraction of inspired oxygen.
- `decrease_fio2` - Decrease the fraction of inspired oxygen.
- `increase_rr` - Increase respiratory rate.
- `decrease_rr` - Decrease respiratory rate.
- `change_mode` - Change the ventilator mode.
- `trigger_alarm` - Trigger an alarm for potential issues.
- `do_nothing` - No action, maintain current state.

The reward function is designed to reflect the goal of maintaining SpO2 and respiratory rate within optimal ranges. The optimal SpO2 range is between 92 and 100 percent, and the optimal RR range is between 12 and 20 breaths per minute.

3.3 Reward Structure

The reward function was designed to encourage actions that moved the patient towards optimal SpO2 and RR levels. States with values within the acceptable physiological ranges were assigned higher rewards, while deviations (e.g., hypoxemia or hypercapnia) incurred penalties.

The reward function evaluates SpO2 and respiratory rate relative to their optimal ranges. If SpO2 is too low (below 92%), actions like increasing FiO2 or respiratory rate are rewarded. If SpO2 is too high (above 100%), actions like decreasing FiO2 or respiratory rate are rewarded. If SpO2 falls in the optimal range (92-100%), the "do_nothing" action is rewarded.

For respiratory rate, if it is too low (below 12 breaths per minute), increasing respiratory rate is rewarded. If it is too high (above 20 breaths per minute), decreasing respiratory rate is rewarded. If the rate is within the optimal range, the "do_nothing" action is rewarded.

The function also penalizes inappropriate actions that move the state further from the target ranges. For instance, if SpO2 is too low and the action is to decrease FiO2, or if SpO2 is too high and the action is to increase FiO2, a large penalty is applied. This ensures that the agent learns to avoid actions that worsen the patient's condition.

3.4 Optimal Policy Characteristics

- Low SpO2 ($< 92\%$): The policy predominantly recommends increasing FiO2 to address oxygen deficiency. This aligns with clinical practices aimed at mitigating hypoxemia, which can have severe consequences if left unaddressed.
- High SpO2 ($> 100\%$): Actions to decrease FiO2 are prioritized to prevent oxygen toxicity, which can lead to cellular damage and other complications.
- Low RR (< 12 breaths/min): An increase in RR is recommended in such states, as slow breathing may result in inadequate gas exchange, leading to hypercapnia.
- High RR (> 20 breaths/min): A decrease in RR is suggested to prevent hyperventilation, which could lead to respiratory alkalosis and patient discomfort.
- Stable States: For states where SpO2 and RR are within their respective optimal ranges (92%-100% for SpO2 and 12-20 for RR), the policy favors the "Do Nothing" action, minimizing unnecessary adjustments and allowing the system to maintain homeostasis.

3.5 Q-Learning Algorithm

We employ Q-learning, a model-free reinforcement learning algorithm, to learn the optimal policy for oxygen management. The Q-values are updated iteratively using the Bellman equation:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right]$$

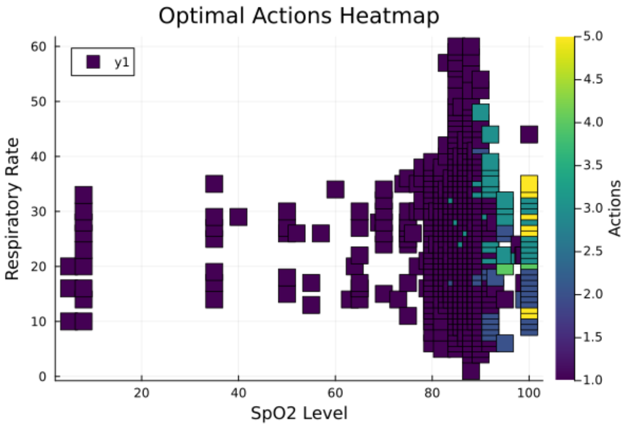
where: - $Q(s_t, a_t)$ is the Q-value for taking action a_t in state s_t, - r_{t+1} is the reward obtained after taking action a_t, - γ is the discount factor, and - α is the learning rate.

We run the Q-learning algorithm for 5000 episodes, each consisting of 1000 time steps. After training, the learned Q-values are used to derive the optimal policy.

4 Results

4.1 Performance Evaluation

After running 5000 episodes of Q-learning, we evaluate the policy by comparing the Q-values for each state-action pair. We observe that the learned policy maximizes rewards by choosing actions that maintain oxygen levels within the optimal range. The following plot shows the distribution of actions across the various states.



It can be seen that increasing FiO2 is prioritized more in scenarios where SpO2 is under the recommended limit, which reflects the clinical best practices for managing SpO2 in respiratory support.

5 Optimal Policy for Respiratory Management

The optimal policy derived for the respiratory management task is based on two critical parameters: **SpO2 (Oxygen Saturation)** and **Respiratory Rate**. The recommended action for each state combination of these parameters is the **increase in FiO2** (fraction of inspired oxygen), which is critical for ensuring proper respiratory support.

SpO2	Respiratory Rate	Recommended Action
92.0	13.0	increase_fio2
92.0	19.0	increase_rr
92.0	21.0	increase_fio2
92.0	20.0	increase_fio2
98.0	23.0	decrease_rr
98.0	16.0	increase_fio2
100.0	9.0	increase_rr
100.0	15.0	do_nothing

Table 1: Excerpt of the Optimal Policy

The table above shows a portion of the optimal policy where the action involves increasing FiO2 based on specific SpO2 and respiratory rate combinations. For a more comprehensive understanding, the complete policy continues across a wide range of values for both SpO2 and respiratory rate.

The policy consistently recommends increasing FiO2 in response to varying SpO2 levels and respiratory rates, suggesting that oxygen support needs to be augmented as either of these parameters moves outside of optimal ranges. Specific thresholds for each parameter (SpO2 and respiratory rate) trigger the corresponding FiO2 adjustments to ensure the patient receives adequate respiratory support.

This policy will be used to guide the management of patients' oxygenation needs based on real-time measurements of SpO2 and respiratory rate.

6 Discussion

Our results demonstrate that Q-learning has the potential to be applied successfully in the context of decision making related to respiratory management in an ICU. The model's reward function rewards actions that bring the patient's parameters closer to the target range while penalizing those that push the values further out of range. This ensures that the model learns to prioritize maintaining stability in SpO2 and respiratory rate levels, which is vital to ensure patient health. Furthermore, the Q-learning algorithm's iterative process of exploration and exploitation allows it to continuously improve its policy by balancing new, untested actions with the knowledge gained from past experiences.

The flexibility of Q-learning enables the model to adapt to varying patient conditions, as it can learn to adjust FiO2 and respiratory rates across different combinations of SpO2 and respiratory rate values. By leveraging historical data to inform decision-making, the model is capable of generating optimal policies that reflect clinical best practices for ventilator management. The learned policies show promising potential for real-time SpO2 management, reducing the risk of hypoxia or hyperoxia in patients. However, there are several limitations to our approach.

6.1 Prioritization of FiO2 Adjustments

One observation from the learned policy was the prioritization of increasing the fraction of inspired oxygen (FiO2) in response to respiratory distress or declining oxygen saturation levels. To contextualize this finding, we found a study whose objective consisted of determining whether respiratory rate measurements detect oxygen desaturation reliably. This study concluded that respiratory rate (RR) measurements correlate poorly with oxygen saturation (SaO2) measurements and do not reliably screen for desaturation. Specifically, patients with low SaO2 levels do not typically exhibit an increased RR, and conversely, increased RR is unlikely to indicate desaturation[5]. These insights suggest that RR alone may not be a dependable indicator for guiding respiratory interventions. In contrast, oxygen saturation levels provide a direct assessment of oxygenation status, justifying the policy's emphasis on FiO2 adjustments as a more targeted response.

This prioritization reflects the model's optimization of actions that directly address hypoxemia, thus enhancing its alignment with clinically effective strategies. The findings underscore the potential for reinforcement learning to capture nuanced relationships between physiological variables and therapeutic decisions.

6.2 Limitations

While the Q-learning model performs well in optimizing SpO2 management in controlled environments, there are several limitations:

- The model does not consider additional patient-specific factors such as age, underlying health conditions, or previous treatment history. These factors could significantly affect the appropriate interventions.
- Our model relies on the assumption that the environment is fully observable, but in practice, there are often uncertainties in clinical settings, such as incomplete or noisy data.
- The model was trained using a simulated environment, which may not fully capture the complexities of real-world ICU conditions. Further testing with real patient data is necessary to evaluate its generalizability.
- The policy optimization was based on a limited action space (FiO2 adjustments), but in practice, there are many other interventions and dynamic factors to consider, such as ventilator settings, medication adjustments, or the introduction of other clinical parameters.

7 Conclusion

In this paper, we applied Q-learning to optimize SpO2 management in ICU patients using the MIMIC-IV dataset. Our results highlight the potential of reinforcement learning in medical decision-making and provide a foundation for future research in critical care optimization. This model offers significant potential for reducing the risk of both hypoxia and hyperoxia, ultimately contributing to better patient outcomes. By continuously adapting to the patient's physiological states, the system can provide real-time recommendations that are more responsive than traditional methods. We hope that our approach can be expanded to more complex decision-making scenarios in healthcare, where timely interventions can save lives.

Future work can focus on refining the model by incorporating more complex patient data, such as the inclusion of additional physiological parameters like heart rate or blood gas measurements, which may provide a more comprehensive understanding of a patient's condition.

8 Individual Contributions

The development and completion of the project involved the equal contribution of all the team members (Amanda and Vakula). We carried out a modular approach to delegate project tasks and had consistent checkpoints to ensure we were on track. Amanda worked on researching available datasets for formulating the problem statement and modeling it. She also handled efficient documentation of the project updates, throughout. The team frequently met to come up with an achievable problem definition that could be modeled effectively. Vakula obtained the relevant parameters from the dataset on BigQuery, and carried out the implementation of the designed model on Julia and tested the model across varying parameters, episodes and steps. Obtaining the final optimal policy was through an equal division of tasks, with fair contribution from both ends.

9 Acknowledgements

We thank the developers and maintainers of the MIMIC-IV dataset for providing access to this valuable resource. We are also grateful to Dr Mykel Kochenderfer and the teaching team of CS238: Decision Making under Uncertainty for providing us with valuable guidance during the course of the project.

References

[1] T. Desautels, J. Calvert, J. Hoffman, and R. Mark. Prediction of sepsis in the icu using machine learning and clinical data. In *AIM*, 2016.

[2] A. E. W. Johnson, T. J. Pollard, L. Shen, L. W. H. Lehman, M. Feng, M. Ghassemi, and R. G. Mark. Mimic-iv: A freely accessible electronic health record dataset. *Scientific Data*, 7(1):1–8, 2020.

[3] Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Brian Gow, Benjamin Moody, Steven Horng, Leo Anthony Celi, and Roger Mark. Mimic-iv (version 3.1), 2024.

[4] Mykel J. Kochenderfer, Tim A. Wheeler, and Kyle H. Wray. *Algorithms for Decision Making*. MIT Press, 2022.

[5] W.R. Mower, C. Sachs, E.L. Nicklin, P. Safa, and L.J. Baraff. A comparison of pulse oximetry and respiratory rate in patient screening. *Respiratory Medicine*, 90(10):593–599, 1996.

[6] Andrew Schaefer, Matthew Bailey, Steven Shechter, and Mark Roberts. *Modeling Medical Treatment Using Markov Decision Processes*, pages 593–612. Science Direct, 01 2005.