MODEL BASED REINFORCEMENT LEARNING FOR ADAPTIVE HEALTHCARE DECISION SUPPORT SYSTEMS

K. Shyamala, K. Manjushree and K. Ramesh Babu
Department of Computer Science and Engineering, Mangalam College of Engineering, India

Abstract

In healthcare, the dynamic nature of clinical processes, coupled with the prevalence of complex diseases and evolving patient conditions, necessitates adaptable and personalized treatment approaches. While existing treatment recommendation systems rely heavily on rule-based protocols derived from clinical guidelines, these may overlook the nuances of individual patient cases, particularly in intensive care units (ICUs) where deviations from standard protocols could be beneficial. However, accessing reliable evidence, such as randomized controlled trials (RCTs), for ICU conditions can be challenging due to various factors including patient eligibility and limited positive findings from RCTs. In such contexts, leveraging large observational datasets and applying artificial intelligence (AI) and machine learning techniques, particularly reinforcement learning (RL), presents a promising avenue for aiding clinical decisions. RL algorithms aim to train agents to maximize cumulative rewards by learning optimal actions based on patient states and trajectories. Unlike traditional clinical protocols, RL policies offer more personalized approaches, capturing individual patient details. Multi-objective reinforcement learning further enhances decision-making by considering multiple objectives, such as cost and optimal path, simultaneously. By mapping state-action pairs to vector rewards, RL algorithms can effectively handle complex decision spaces and facilitate the selection of optimal actions.

Keywords:

Reinforcement Learning, Healthcare Prediction, Multi-objective Optimization, Machine Learning Algorithms, Intensive Care Units

1. INTRODUCTION

The increasing incidence of complicated diseases and the dynamic changes that occur in the clinical circumstances of patients are two factors that contribute to the dynamic nature of clinical processes in the health care industries. In most cases, the implementation of existing treatment recommendation systems is accomplished through the utilization of rule-based protocols that are created by medical professionals on the basis of evidence-based clinical standards or best practices [1-3]. These methods and guidelines might not take into account several comorbid conditions, which is another potential limitation [4]. Critically ill patients in an intensive care unit (ICU) may benefit from deviating from established treatment protocols and from customizing patient care through the use of methods that are not based on rules [5,6].

For the purpose of tailoring treatment to the specific needs of particular patients, medical professionals may consult randomized controlled trials (RCTs), systemic reviews, and meta-analyses as sources of potential information. On the other hand, randomised controlled trials might not be available or decisive for many intensive care unit conditions. The majority of patients who are admitted to intensive care units can also be too sick to participate in research studies [6]. In addition, only nine percent of therapy recommendations in the intensive care unit are based on randomized controlled trials [7], and the great majority of RCTs in critical care have outcomes that are poor [8]. We require more strategies, such as the utilization of big observational data sets, in order to assist clinical decision-making in intensive care units. Because they were collected in a data-rich environment, data from the intensive care unit (ICU) can be helpful for learning about patients. Afterwards, a substantial quantity of data can be included into artificial intelligence (AI) systems, which involve the utilization of computers to simulate human cognitive capabilities, as well as machine learning techniques, which involve the utilization of computer algorithms to carry out clinical activities without the requirement of particular instructions. In the intensive care unit (ICU), artificial intelligence and machine learning provide assistance with diagnosis [9]-[10], therapy, and resource management. Particularly well-suited for intensive care unit (ICU) environments is a machine learning technique known as reinforcement learning (RL), which takes into account the fluid character of severely ill patients.

Healthcare decision-making is a complex process that involves considering various factors such as patient history, current symptoms, available treatments, and medical guidelines. Traditional healthcare decision support systems (DSS) often rely on static rules or pre-defined algorithms, which may not adapt well to the dynamic and diverse nature of patient cases. This limitation can lead to suboptimal treatment recommendations and inefficiencies in healthcare delivery [12].

One of the major challenges in healthcare decision support systems is the lack of adaptability to changing patient conditions and evolving medical knowledge. Conventional approaches often struggle to incorporate the latest research findings and clinical insights into their decision-making processes. Moreover, they may not adequately account for individual patient preferences and unique contextual factors, leading to a one-size-fits-all approach that may not be optimal for every case [13]. As a result, there is a growing need for more flexible and adaptive systems that can continuously learn from new data and update their decision-making strategies accordingly. Model-based reinforcement learning presents a promising approach to address these challenges by leveraging mathematical models of patient dynamics and treatment outcomes to optimize decision-making over time.

When given the state-action trajectories of the patients, the primary goal of the RL algorithm is to train an agent that is capable of maximizing the cumulative future reward from the state-action pairs. When the agent notices a new state, it has the ability to carry out an action, and it has the ability to select the action that would result in the best possible long-term outcome (for example, survival). It is possible for the RL agent to select the most appropriate action in accordance with the condition of a patient when the agent has received adequate training. We refer to this process as acting in accordance with an optimal policy.

Comparable to a clinical protocol is the concept of a policy. In spite of this, a policy offers a number of advantages over a clinical protocol, one of which is that it is able to record more specific and individual medical information about each patient. One way to illustrate a policy is through the use of a table that represents the mapping of all potential states with actions. On the other hand, a policy could alternatively be represented by a deep neural network (DNN), in which case the DNN model would output the highest likelihood of an action if it were given the input of a patient's state. There are several different RL algorithms that can be used to train an optimal policy.

2. RL FOR HEALTHCARE PREDICTION

As a result of the fact that it takes into account both the optimal path and the cost as its goal function, multi-objective reinforcement learning is utilized in order to enhance the process of identifying agents or to locate the target query solution. Reinforcement learning, also known as agent learning, is a process in which an agent learns the environment in order to achieve a target by employing the most effective action. Immediately following the completion of the activity by the agent, the environment will send a feedback signal, which is referred to as the reward. For the purpose of achieving the highest possible reward signal, the agent and reward program links the environmental circumstance to the action.

In addition to enhancing the process of identifying agents or locating target query solutions, multi-objective reinforcement learning offers a robust framework for addressing the inherent trade-offs in healthcare decision-making. By simultaneously considering both the optimal path and the cost as objective functions, this approach enables decision support systems to navigate complex healthcare scenarios more effectively. Moreover, by utilizing vector rewards instead of scalar rewards, multi-objective reinforcement learning captures the nuanced relationships between various decision criteria, allowing for a more comprehensive evaluation of potential actions. Furthermore, the incorporation of Pareto-dominance relationships ensures that decision-making is guided by the principle of non-dominance, leading to the identification of solutions that are not inferior to others in any objective. This facilitates the exploration of diverse treatment options while avoiding suboptimal choices. Additionally, the action-selection function, guided by the scalarization function, enables the system to strike an appropriate balance between competing objectives, thereby providing clinicians with actionable insights that consider both cost-effectiveness and clinical efficacy. Moreover, the parameterization of the scalarization function allows for the customization of decision-making strategies based on the specific priorities and preferences of stakeholders. This flexibility ensures that the decision support system can adapt to different clinical contexts and patient populations, maximizing its utility across diverse healthcare settings. Overall, the integration of multi-objective reinforcement learning into healthcare decision support systems holds significant promise for improving patient outcomes, optimizing resource allocation, and advancing the delivery of personalized care.

The multi-objective reinforcement learning has rewards for each element using a vector rather than using a scalar reward. The vector value is stored as state (*s*) action (*a*) pair, which is represented as *Q*(*s*,*a*). The reinforcement or Q-learning to handle multiple objectives, i.e. cost and path calculation and vector value is applied over it. The vector value of actions is represents as *Q*(*s*) denotes the vector values of all actions in state *s*.

The core issue deals with selection of action using a vector value to form a policy. Here, the value of *Q*(*s*,*a*) is not a simple one, since it is a vector value and the action is considered maximal and optimal for first and second objective, respectively. The non-dominated multiple actions are represented as Pareto-dominance relationship [3]. The agents uses an action-selection function to define ordering on vector value, which allows greedy action given in Eq.(1). The action selection function (*U*) is used to find the preference path and reduced cost,

$$U\left(Q(s,a),Q(s,b)\right)=\begin{cases}+1 & \text{if } Q(s,a)\succ Q(s,b)\\ -1 & \text{if } Q(s,a)\prec Q(s,b)\\ 0 & \textit{otherwise}\, Q(s,a)\sim Q(s,b)\end{cases} \tag{1}$$

The scalarisation function (*f*) is implemented over action selection function (*U*) to map the values of vector to the values of vector, which is given by,

$$U\left(Q(s,a),Q(s,b)\right)=\begin{cases}+1 & \text{if } f\left(Q(s,a)\right)>f\left(Q(s,b)\right)\\ -1 & \text{if } f\left(Q(s,a)\right)<f\left(Q(s,b)\right)\\ 0 & \text{if } f\left(Q(s,a)\right)=f\left(Q(s,b)\right)\end{cases} \tag{2}$$

Hence, to maintain trade-off between the objective function i.e. cost and path, *f* is parameterized using the parameter in [3]. The scalarisation is estimated as linear weighted sum of cost and path as in Eq.(3) with weights w_o provides the importance of the following objective function:

$$f\left(Q(s,a)\right)=\sum_{o=1}^{n} w_o Q(s,a,o) \tag{3}$$

Since, the concept of action-reward is difficult for computation, linear scalarisation is applied, since it perform simplified operations. Since, it maintains the trade-off between the cost and optimal path.

3. MULTI-OBJECTIVE REINFORCEMENT LEARNING

The random access is used to produce the object set related to given query. This method is not ranked and it aids in providing better selection of attributes. Here, each service is associated with the cost for various services. Hence, the cost associated with random access resemble average service response time between the origin and target selection in database, which performs heavy computation. Hence, the calculation of cost function is used to determine the overall cost of target or tuple retrieval, which estimates the access between the origin and target.

The random access is used to find the top query combinations ($a \geq A$) in union of combinations [2]. The combinations are considered as top, since the scores experiences upper bounds, which is represented by,

$$u = f\left(\tau_1^{n_1}\left[S_1\right],....,\tau_m^{n_m}\left[S_m\right]\right) \tag{4}$$

Hence, the cost of tuple retrieval can be estimated, and this should dominate the computation cost of all combinations and its scores. Hence the services in domain specific service engine on non-additive model is given as,

$$C_m = \sum_{m=1}^{k} rc_m j_m\left(n_m\right) + h\left(t_m\right) \qquad (5)$$

where, rc_m is defined as unitary random access cost and $j_m(n_m)$ is defined as retrieved join attribute tuple.

4. RESULTS AND DISCUSSION

Numpy, Datetime, SciPy, Pandas, Matplotlib, and Scikit Learn are some of the libraries and packages that are utilized in this section for the purpose of carrying out the implementation. The complete implementation is carried out on a platform provided by Google Colab through the use of CPU runtime. The characteristics of the central processing unit (CPU) include an Intel Core i5 operating at 2.20 GHz. Datasets were obtained via the Kaggle repository, which included Johns Hopkins University and the World Health Organization among its contributors. These individuals are characterized by a number of symptoms, the most prominent of which are fever, cold, and cough.

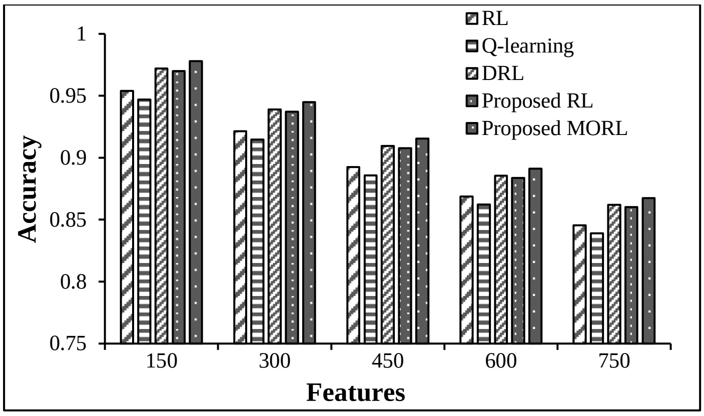

Fig.1. Accuracy

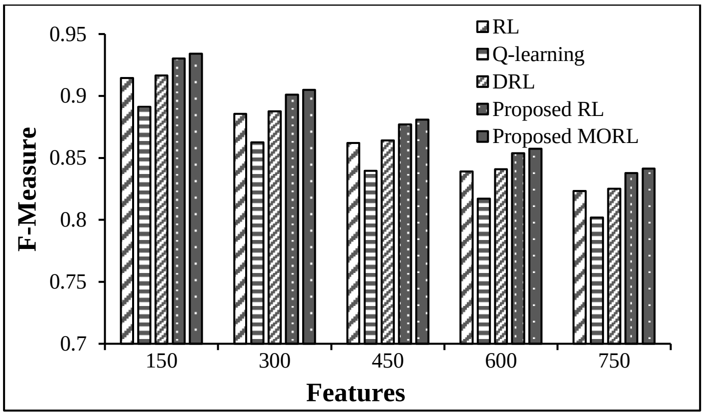

Fig.2. F-measure

The results of the accuracy comparison between the proposed RL Models and the various current RL models are displayed in Fig.1. The outcome of the simulation demonstrates that the proposed method obtains a higher rate of accuracy compared to other ways on the market.

The Fig.2 illustrates the results of the F-measure that was performed between the suggested RL Models and the existing RL models. With regards to the rate of F-measure, the results of the simulation demonstrate that the suggested technique achieves a higher rate than other methods.

5. CONCLUSION

Within this study, the RL model for healthcare prediction system, which incorporates five distinct machine learning algorithms, functions as a more effective classification model for predicting diseases in their earlier stages with greater accuracy. Through the utilization of RL, this meta-approach offers improved prediction performance while simultaneously merging the outcomes of classifiers. It has been demonstrated through simulation that the proposed strategy achieves a higher rate of prediction accuracy compared to models that are considered to be state-of-the-art. Based on the findings, it is evident that the RL model offers a greater rate of classification accuracy compared to other models that are currently implemented. A further conclusion that can be drawn is that the suggested method has a lower rate of classification error compared to the RL models that are already in use.

REFERENCES

[1] Ping Cao, Bailu Ye, Linghui Yang and Qing Pan, “Preprocessing Unevenly Sampled RR Interval Signals to Enhance Estimation of Heart Rate Deceleration and Acceleration Capacities in Discriminating Chronic Heart Failure Patients from Healthy Controls”, *Computational and Mathematical Methods in Medicine*, Vol. 2020, pp. 1-14, 2020.

[2] P. Dwivedi and M.K. Singha, “IoT based Wearable Healthcare System: Post COVID-19”, Proceedings of International Conference on the Impact of the COVID-19 Pandemic on Green Societies: Environmental Sustainability, pp. 305-321, 2021.

[3] H. A. Esfahani and M. Ghazanfari, “Cardiovascular Disease Detection using a New RL Classifier”, *Proceedings of International Conference on Futuristic Trends on Computational Analysis and Knowledge Management*, pp. 1011-1014, 2014.

[4] T. Vivekanandan and N.C.S.N. Iyengar, “Optimal Feature Selection using a Modified Differential Evolution Algorithm and its Effectiveness for Prediction of Heart Disease”, *Computers in Biology and Medicine*, Vol. 90, pp. 125-136, 2017.

[5] M. Sai Shekhar, Y. Mani Chand and L. Mary Gladence, “Heart Disease Prediction using Machine Learning”, *Lecture Notes in Electrical Engineering*, Vol. 708, No. 11, pp. 603-609, 2021.

[6] Z. Lou and G. Shen, “Reviews of Wearable Healthcare Systems: Materials, Devices and System Integration”, *Materials Science and Engineering: R: Reports*, Vol. 140, pp. 100523-100534, 2020.

[7] E. Dovgan, Y.C. Li and S. Syed Abdul, “Using Machine Learning Models to Predict the Initiation of Renal Replacement Therapy among Chronic Kidney Disease Patients”, *Plos One*, Vol. 15, No. 6, pp. 1-13, 2020.

[8] M. Abdar, M. Zomorodi-Moghadam, R. Das and I.H. Ting, "Performance Analysis of Classification Algorithms on Early Detection of Liver Disease", *Expert Systems with Applications*, Vol. 67, pp. 239-251, 2017.

[9] G.G. Wang, S. Deb and L.D.S. Coelho, "Elephant Herding Optimization", *Proceedings of International Symposium on Computational and Business Intelligence*, pp. 1-5, 2015.

[10] C. Saranya and G. Manikandan, "A Study on Normalization Techniques for Privacy Preserving Data Mining", *International Journal of Engineering and Technology*, Vol. 5, No. 3, pp. 2701-2704, 2013.

[11] Y. Chen, W. Xie and X. Zou, "A Binary Differential Evolution Algorithm Learning from Explored Solutions", *Neurocomputing*, Vol. 149, pp. 1038-1047, 2015.

[12] M.S. Amin, Y.K. Chiam and K.D. Varathan, "Identification of Significant Features and Data Mining Techniques in Predicting Heart Disease", *Telematics and Informatics*, Vol. 36, pp. 82-93, 2019.

[13] A.M. Sarhan, "Data Mining in Internet of Things Systems: A Literature Review", *Journal of Engineering Research*, Vol. 6, No. 5, pp. 252-263, 2023.