A Reinforcement Learning and IoT based System to Assist Patients with Disabilities

Muddasar Naeem[1][a], Antonio Coronato[2], Giovanni Paragliola[2] and Giuseppe De Pietro[2]
[1] *Universita' Degli Studi di Napoli Parthenope, Napoli, Italy*
[2] *ICAR-CNR, Napoli, Italy*

Keywords: Artificial Intelligence, Disability, Healthcare, IoT, Pill Reminder, Reinforcement Learning.

Abstract: One of the important aspect of clinical process is to complete treatment according to given plan. The successful completion of this task is more challenging when a person have some physical or mental disability and requires resources and man power for personalized treatment and care. We can mitigate this problem by an intelligent guidance and monitoring system who can assist elderly persons and patients in their treatment schedule. Reinforcement learning and IoT systems have received considerable credit of significant contribution in healthcare over last few years, could be suitable choice for said objective. We propose a pill reminder system using Bayesian reinforcement learning assisted with IoT devices to help people (having mental and/or physical disability) in their treatment plan. The proposed intelligent system is able to successfully communicate with the person through a suitable audio, visual and textual message. The proposed pill-reminder system has been demonstrated for a specific treatment plan of a hypertension patient.

1 INTRODUCTION

An important aspect of treatment for all stake holders connected to healthcare sector in continuous improvement to patient treatment which ensure provision of satisfaction to patients (Ross et al., 1993). Success to this objective is not dependent on single factor instead it depends on many factors like trained personal, quality care centers, medical and nuclear examination equipment, use of suitable medication (Gullapalli N Rao, 2002)and use of modern technology i.e Artificial Intelligence (AI) and Machine Learning (ML). We have witnessed the revolution of AI and transformation it has brought to our living style (Patel et al., 2009), (A Testa,). Some of appealing applications are: diagnostic systems (Ling et al., 2017), virtual assistants , personalized treatment (Petersen et al., 2018), *DTR*, a multi-step clinical decision processes ((Lavori, 2004), (Chakraborty,)), medical imaging (Sahba et al., 2006), dialogue systems and chat-bots (Kearns et al., 2011), risk management ((Giovani Paragliola, 2019), (Antonio Coronato, 2019)), control systems (Prasad et al., 2017) and rehabilitation (Reinkensmeyer et al., 2012), (G Paragliola,). That is why billion dollars have been investing on use of AI and ML especially Reinforcement Learning (RL) in healthcare.

RL is an ML approach much more focused on goal-directed learning from interaction, than other approaches of machine learning i.e. supervised and unsupervised learning (Sutton and Barto, 1998). In the last decade, we have seen much application on the use of RL in healthcare departments. This is due to the similar objective of RL algorithms and clinicians. This is to say that the goal of doctors is to find an optimal sequence of treatments for a particular patient. This scenario is in accordance with the major objective to RL which is to find an optimal policy for a given problem in a given environment. Hence, RL has achieved considerable success to help clinicians in optimizing and personalizing treatment sequences (Thall and Wathen, 2005) (Murphy et al., 2006).

Normally, caregivers are being used to assist patients and elderly persons in their activities of daily living (ADL) through the use of cues, signals and verbal reminders. Then we saw the development of a computer-based solution ' Cognitive Ortosis for Assisting activities in the Home (COACH)' (M.A.Sc. et al., 2001) to assist dependent persons, an agent based platform for task distribution in virtual environments (A Coronato,). Few research work on use of computer vision technology to assist patients are:

[a] https://orcid.org/0000-0003-0815-4883

Naeem, M., Coronato, A., Paragliola, G. and De Pietro, G.
A Reinforcement Learning and IoT based System to Assist Patients with Disabilities.
DOI: 10.5220/0009783804910498
In *Proceedings of the 5th International Conference on Internet of Things, Big Data and Security (IoTBDS 2020)*, pages 491-498
ISBN: 978-989-758-426-8

(Wren et al., 1997), (Oliver et al., 2004), (D. F. Llorca, 2007), (Llorca et al., 2011).

In this century, we have moved to the use of AI and ML in the healthcare sector e.g predictive analysis, development of intelligent robotic caregivers, making the ICU smarter (Suresh et al., 2017). The systems based on ML decision making approaches like 'Partially Observable Markov Decision Process (POMDP)' are: the assisted cognition project (H. Kautz, 2002), a situation-aware system for the detection of motion disorders of patients with autism spectrum disorders (A Coronat0,) aware home project (Mynatt et al., 2000), the adaptive house (Mozer, 2005), nursebot project (Pineau et al., 2003) and automated hand-washing assistance (30, 2010). Most of these works assist elder persons and patients with dementia in one of ADL like hand washing.

The contribution of the present work is to assist patients of any disease and elderly persons having any mental or physical disabilities like audio and visual instead of assisting with physical activities. We modeled the problem as an MDP and provide a solution by using Bayesian Thompson sampling.

The proposed intelligent system first sends a reminder to patients according to one's treatment plan and then guide the patient to specific medicine through an appropriate type of message. We have also demonstrated our work on a practical treatment plan which was advised to a patient of hypertension. The message type is very critical and the choice of message is performed by considering a person's skills i.e. physical and mental. After learning, the RL agent can choose a suitable type of audio message or visual message or text message depending upon one's physical and mental abilities.

The next part of the paper is organized as follows. Section 2 presents a brief introduction of RL followed by the system model section, results and discussion and finally the conclusion of work.

2 BACKGROUND

Markov Decision Process (MDP) is the central concept of all RL problems. The goal of the RL algorithms is to find the solution to an MDP. Ans MDP model has the following components:

-Set of states: $S = \{s_1, s_2, s_3, ...s_n\}$

-Set of actions: $A = \{a_1, a_2, a_3, ...a_n\}$

-Transition model: $T(s_t, a, s_{t+1})$

-Reward R

Reward and transition model depend upon on current state s_t, selected action a_t and resulting state s_{t+1}.

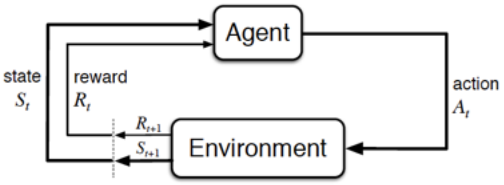

Figure 1: The Reinforcement Learning problem.

The target of RL algorithms is to interact with a given world either with some prior knowledge i.e reward and transition model (model-based RL e.g Dynamic Programing which includes value iteration and policy iteration) or without any prior knowledge (model-free RL). Examples of model-free algorithms are Monte Carlo and Temporal difference (TD) learning. Q-learning and SARSA are widely used as TD algorithms (Russell and Norvig, 2009). After many repetitive interactions with the environment, the **agent** learns the characteristics of the environment. The target of RL agent is to find an optimal **action** out of available actions in each **state**. Optimal action returns best desired numerical **reward** to the agent. An agent choose an action in each state which results in a **policy** π. An optimal policy maximizes the aggregated future reward for a specific problem. The working framework of RL is shown in figure 2.

An RL agent should maintain a balance between exploitation and exploration while learning i.e. an equilibrium between maximizing reward from already known useful actions or to explore new moves which even give better rewards. In exploitation, the priority of the agent is to select the best action based on his learning and knowledge and in exploration, the agent attempts actions in a stochastic way to improve his experience and learning to get more reward. Bayesian methods can be a solution to the exploitation-exploration dilemma due to their ability to capture uncertainty in learned parameters and avoid over-fitting (Welling and Teh, 2011). Few famous methods used for Bayesian approximations are Myopic (Dearden et al., 2013) and Thompson Sampling (Strens, 2000) which will be used in present work.

3 SYSTEM MODEL

This section presents the proposed approach based on Bayesian RL agent, planner and checker (IoT system) to assist patients and elderly persons at home having one or a combination of more than one disability. We consider audio, visual disabilities and condition of patient working memory and attention. The goal of the RL agent is to provide assistance to the patient according to his/her treatment plan through a

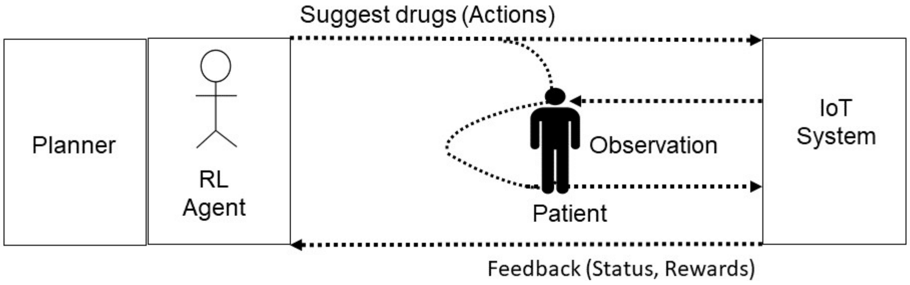

Figure 2: The proposed model.

suitable message. Figure 2 shows the proposed model which has three main components such as planner, tutor (agent) and IoT system as checker or observer.

The planner sends a reminder to both tutor and checker according to the advised treatment plan. The tutor's role is most important which is to choose a suitable choice of message for communication with the patient. The tutor considers the physical and mental abilities of the patient before selecting the way of communication. The role of the checker is to monitor the patient takes the right drug. It observes the status and sends it along with the reward to the Tutor

The audio, visual, working memory and attention abilities which tutor must consider are defined below:

Auditory Perception

- Audio Skill (AU_s) in $[0->1]$: The audio skill plays an important role in the treatment process of any patient and reduces dependency on caregivers. Our intelligent reminder system takes account of audio disability before choosing a message. If a patient has a lower skill of audio perception then the probability that he/she can listen to an audio message will be low and consequently the chance of discontinuity of in planned treatment. In that case, we have to check for other skills of the patient like visual perception.

Visual Perception

- Visual Skill (VS_s) in $[0->1]$: It is another important skill that must be considered when assisting patients and the elderly population. The lower visual skill means that the probability to view a visual message will be minimal. Our agent has alternative options to select in case someone has a week of visual perception.

Working Memory

- Working Memory High Skill (WM_{hs}) in $[0->1]$: The lower the skill, the lower the probability to understand a "complex" message. For example, if a patient's working memory high skill is in good condition then he can understand audio or visual scientific messages depending on his/her audio and visual disability level.

- Working Memory Low Skill (WM_{ls}) in $[0->1]$: The lower the skill, the lower the probability to understand "simple" audio or visual message.

Table 1: Basic Messages.

Label	Message Type
C1	Audio Scientific Message
C2	Audio Simple Message
C3	Visual Pill-Box Image
C4	Visual Pill-Box and pill Image
C5	Scientific Textual Message
C6	Simple Textual Message

Attention

- Attention Skill (AT_s) in $[0->1]$: The lower the skill, the higher the probability to ignore the message. The attention of the patient is critical in the successful completion of the task. It's feedback which when positive, motivates our agent to select and send messages to the patient.

We define six types of messages as shown in table 1 by using the first four skills. For example, a suitable message for the patient who has a good audio ability and working memory is a scientific audio message e.g 'take 'Adalat Crono 30mg' pill'. When having reasonably good audio skills but with lower working memory then a simple audio message is a better choice instead of a scientific message i.e. 'take the pill from the red, white, green, etc box'. Similarly, when a patient's audio skills are not good but visual ability is good then the agent has options to select an image of pill-box or image of a pill or a scientific or simple text message depending on the condition of working memory.

The **planner** sends a reminder to the RL agent on a scheduled hour with the name of the medicine. It wakes up both tutor and checker when it is time for the patient to take a drug In the current simulation setup, we use a timer to keep track of hours and day of the treatment plan. In the next version, we will use the system input to keep track of hours and days.

Then **tutor** i.e. the agent chooses a single or combination of the audio, visual and textual message. After receiving a message from the RL agent, the possible directions for the patient is going to the right pill-box or wrong pill-box as graphically shown in figure 3.

The checker monitors the patient's actions through suitable IoT devices and sends feedback to the RL agent (Muddasar Naeem, 2020). It checks the drug that the patient is going to take, observes the status and sends it along with the reward to the Tutor. The reward for the RL agent to guide a patient to the right box is one and zero otherwise .

Before sending a message, the agent checks whether a patient is attentive or no and in case, the patient is not attentive, the agent waits and sends repetitive alerts to catch the attention of the patient. Once the patient gets attentive then the agent will start sending a suitable message by considering the audio, visual and working memory skills.

In figure 3, large open circles are states and small solid circles are action nodes for each state-action pair. This is a finite MDP model (Sutton and Barto, 1998) and the resulting destination state depends on the probability of the current state-action pair. The patient will be directed to the right box if he/she understands the message sent by the agent. The probability that a patient will understand a message or ignore a message is calculated as:

$$\begin{gathered}\text{Probability that the message understood } = P_{mu} \\ P_{mu} = f\left(AU_s, VS_s, WM_{hs}, WM_{ls}\right) \\ \text{Probability that the message ignored} = P_{mi} \\ P_{mi} = f\left(AT_s\right)\end{gathered} \tag{1}$$

Equations 3 and 1 indicate the probability that a patient understands a message is a function dependent on the audio, visual and working memory abilities of that patient. Similarly, the probability that a patient will ignore a message depends on the attention status of the patient. This can be further elaborated in equations 2 and 3.

$$\begin{gathered}P_{mu} = \min(1, C1 * AU_s * WM_{hs} + C2 * AU_s * WM_{ls} + \\ C3 * VS_s * WM_{hs} + C4 * VS_s * WM_{ls} + \\ C5 * VS_s * WM_{hs} + C6 * VS_s * WM_{ls})\end{gathered} \tag{2}$$

$$P_{mi} = 1 - AT_s \tag{3}$$

The state value and action-value function for an RL problem using the Bellman equation can be written as shown in equations 4 and 5 respectively.

$$V\pi(s') = \sum_a \pi(s,a) \sum_{s'} p(s'|s,a)[R(s,s',a) + \gamma V^{\pi}(s')] \tag{4}$$

$$Q^{\pi}(s,a) = \sum_{s'} p(s'|s,a)[R(s,s',a) + \gamma Q^{\pi}(s',a')] \tag{5}$$

The target of value function approaches is to calculate state and action-value function and then drive the optimality policy through maximum value function in every state. Optimal state value function using Bellman optimally equation 6 is given as:

$$V^{\pi*}(s') = \max_{a \in A(s)} \sum_{s'} p(s'|s,a)[R(s,s',a) + \gamma V^{\pi*}(s')] \tag{6}$$

By using equation 6, we can write optimal equation for our model. For example, the Bellman optimality equation at wait and message selection states abbreviated as w and MS respectively, may be written as shown in equations 7, 8 and 9.

$$\begin{gathered}V^*(w) = \underbrace{\max}_{a} p(w|w,a)[r(w,a,w) + \gamma V^*(w)] + \\ p(MS|w,a)[r(MS,a,w) + \gamma V^*(MS)]\end{gathered} \tag{7}$$

$$\begin{gathered}V^*(w) = \underbrace{\max}_{a} P_{mi}[r(w,a,w) + \gamma V_*(w)] + \\ (1 - P_{mi})[r(MS,a,w) + \gamma V_*(MS)]\end{gathered} \tag{8}$$

$$\begin{gathered}V^*(MS) = \underbrace{\max}_{a} P_{mu}[r(RB,a,MS) + \gamma V_*(RB)] + \\ (1 - P_{mu})[r(WB,a,MS) + \gamma V^*(WB)]\end{gathered} \tag{9}$$

$$Q^{\pi^*}(s,a) = \sum_{s'} P(s'|s,a)[R(s,s',a) + \gamma * \max_{a'} Q^{\pi*}(s',a')] \tag{10}$$

The Bellman equation we need to solve for Bayesian RL is given in 11.

$$V^{\pi*}(x,b) = \max_a \sum_{x'} Pr(x'|x,b,a)[X'_r + \gamma V^{\pi*}(x', b_{xax'})] \tag{11}$$

Where X, b and $Pr(x'|x,b,a)$ represents set of states, distribution (belief) over the unknown θ used for exploration and transition probabilities respectively. Each selected message will guide a patient to the right box with a certain probability. The higher the probability the more likely the person reached the right box. This unknown probability θ is modeled based on our initial probabilities given in 2 and 3.

Based on initial probabilities, we model the posterior distribution of θ using Bayes rule as follows:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} \tag{12}$$

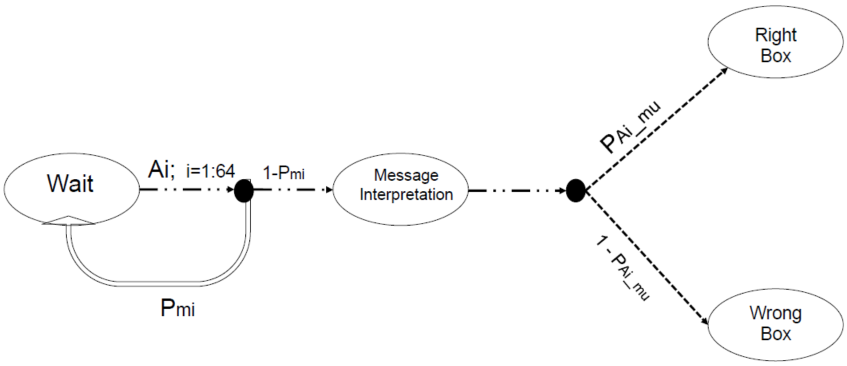

Figure 3: The Reinforcement Learning block.

The terms $P(x|\theta)P(\theta)$ and prior $P(\theta)$ are likelihood function which follows Bernoulli distribution and Beta distribution. Sutton considers a similar problem of multi-armed bandit in his book (Sutton and Barto, 1998) and he used Gaussian distribution. As $P(\theta)$ Beta distributed and $P(x|\theta)P(\theta)$ is Bernoulli distributed, the term $P(\theta|x)$ is also Beta distributed which is to say that when a patient reached to right box then posterior will become $Beta(\alpha+1,\beta)$ and if patient reached to wrong box then posterior will be $Beta(\alpha,\beta+1)$. We use Thompson sampling (Prasad et al., 2017) to solve the problem of exploration-exploitation in which for each message, the probability θ is sampled from the prior and then the message with the highest sampled probability is selected. Figure 5 shows the convergence of 12 to different numbers of iterations.

4 DISCUSSION

We have implemented the proposed study in general and have demonstrated our system for a specific treatment plan which was advised by a clinician to a patient of hypertension as shown in table 2. The working of the proposed model is shown in figure 4. The information from the planner to the RL agent consists of drug name and time hour. Then the tutor (RL agent) by using Thompson sampler decides how to communicate with the patient about the scheduled drug. As can be seen in figure 4, that agent can send one or combination of the image of a scheduled medicine, image of pill, scientific or simple name of message through text or audio.

To explain better working of proposed system and figure 4, let consider the table 2. According to this treatment plan, the patient of hypertension need to take drug **Adalat Crono 30mg** at $08:00$ and $20:00$ every day for a period of forty days. So at $08:00$ and $20:00$ each day, the system first learns the status of the patient's mental and physical skills separately i.e. system does not use the information he learned at $08:00$ for a drug which patient will take at $20:00$. In the second step, the system decides how to communicate drug information to the patient. More precisely, at $08:00$ and $20:00$, the system has to choose one or combination of more than one out of the following six.

- **Scientific Audio Message** as: take the 'Adalat Crono 30mg'.

- **Simple Audio Message** as take 'the medicine from the white, red and yellow box' (this is a sample message, one can set it to a more preferred and better way).

- **Scientific Visual Message** as: simply send the box image of the 'Adalat Crono 30mg' as can be seen in figure 4.

- **Simple Visual Message** as: simply send the pill image of the 'Adalat Crono 30mg' as can be seen in figure 4.

- **Scientific Text Message** as: a text message like this: "take 'Adalat Crono 30mg'." will appear on screen e.g mobile phone or tablet of pepper robot

- **Simple Text Message** as: a text message like this: "take 'the medicine from the white, red and yellow box'." will appear on screen e.g mobile phone or tablet of pepper robot.

The next question is the choice of RL algorithms for the current work. We have briefly introduced most of the RL algorithms in the background section. We have not compared our work with previous work as according to our best knowledge to date, no work of such nature exists before. This will be the first work of its kind. So we tested different algorithms for our work and set two performance metrics. One is the Average Utility Distribution (AUD) of each 63 messages at message interpretation state and second is to convergence time i.e number of iterations need to get experimental probabilities given the initial probabilities.

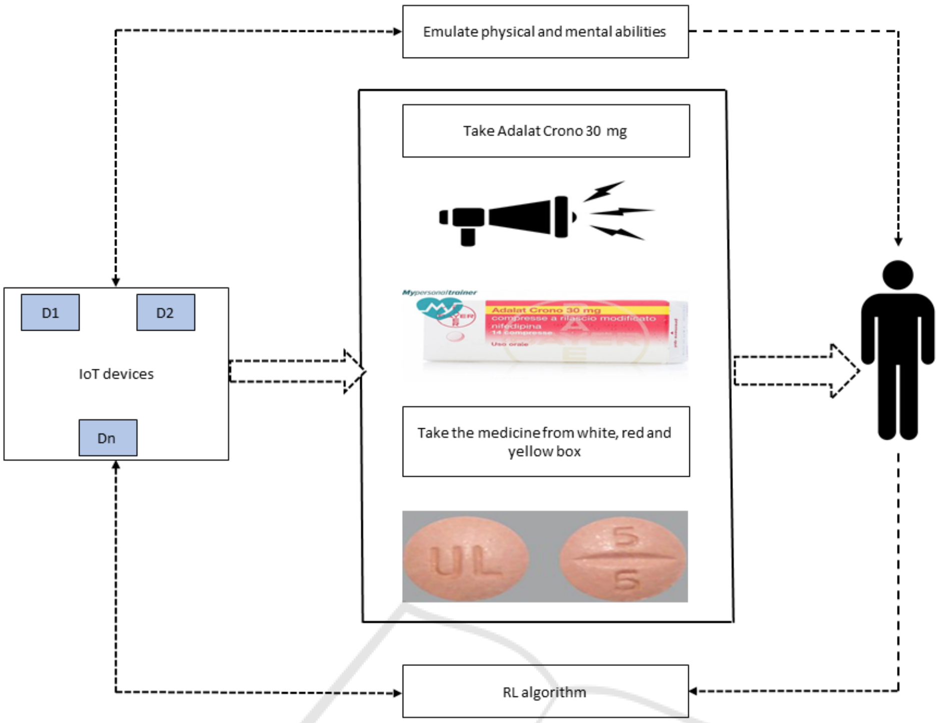
Emulate physical and mental abilities
Take Adalat Crono 30 mg
Adalat Crono 30 mg
Take the medicine from white, red and yellow box
D1
D2
IoT devices
Dn
RL algorithm

Figure 4: Working of the proposed model.

Table 2: Treatment plan for a patient of hypertension.

Medicines	Time
Sp/ Duoplavin 75+100 mg	1 cp at 14:00 after lunch
Sp/ Torvast 80 mg	1 cp at 22:00
Sp/ Bisoprololo 1.25 mg	1 cp at 08:00
Sp/ Adalat crono 30 mg mg	1 cp at 08:00 and 22:00
Sp/ Prefolic 15 mg	1 cp at 14:00
Sp/ Sideral forte	1 cp at 13:00
Sp/ KCl retard 600 mg	1 cp $*$ 3
Sp/ Humulin R	4 UI after breakfast, 6 UI after lunch and dinner
Sp/ Spiriva respimat	1 puff at 18:00
Sp/ Pulmaxan 200 mcg	1 cp before breakfast

Moreover, as we have sixty-three available messages to choose one and in the future, this list can get even bigger, it is important to have a delicate balance between exploration and exploitation as explained in the background section. Here we use Average Aggregated Reward (AAR) and 'Root Mean Square Error (RMSE) to measure exploitation and exploration respectively. When both RMSE and AAR are low then the agent is doing exploration and exploit already known actions when both RMSE and AAR have high values.

As can be seen in table 3, the results of Thompson sampling with Boltzman, Epsilon-decreasing, Epsilon-greedy, Greedy, Random, and Softmax are being compared and Thompson sampling has a comparatively better average reward. Similarly, in figure 5, we can see most of the actions need only a few trials to reach an estimate of experimental probabilities from initial probabilities. Furthermore, Thompson sampling maintains a decent balance between exploitation and exploration as evident from row two and three of table 3.

5 CONCLUSIONS

We have proposed a RL based system to provide clinical support to patients having audio or visual or both disabilities. The pill-reminder system is able to assist

Table 3: Comparison of AUD, AAR and RMSE of different algorithms.

Algo $\Longrightarrow$	Boltzman	Thompson	Softmax	Eps-Dec	Eps-Greedy	Random	Greedy
AUD $\Longrightarrow$	0.0156	0.9190	0.0642	0.04025	0.0668	0.4019	0.5199
AAR $\Longrightarrow$	44.17	38.55	45.42	44.35	45.5	36.39	32.76
RMSE $\Longrightarrow$	0.7522	0.2593	0.7023	0.7265	0.6986	0.3531	0.4557

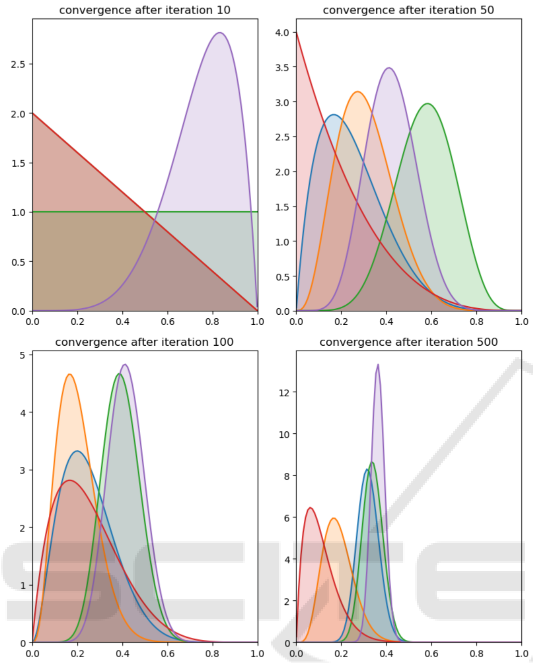

Figure 5: Convergence at Different Numbers of Iterations.

patients with single audio, visual, textual or combination of more than one message choice out of six types of message in their treatment plan. Depending upon a patient's disability, RL agent send an appropriate type of message according to one's schedule time of treatment. In next version of our tutoring system project, we will include implementation of proposed work on interrelated computing devices i.e IoT system and will try to generalize the proposal to more complex scenarios.

ACKNOWLEDGEMENTS

This work is partially supported by AMICO project which has received funding from the National Programs (PON) of the Italian Ministry of Education, Universities and Research (MIUR):
code ARS0100900(DECREEN. 1989,26 July 2018).

REFERENCES

(2010). Automated handwashing assistance for persons with dementia using video and a partially observable markov decision process. *Computer Vision and Image Understanding*, 114(5):503 – 519. Special issue on Intelligent Vision Systems.

A Coronat0, G. De Pietro, G. P. A situation-aware system for the detection of motion disorders of patients with autism spectrum disorders. In *Expert systems with applications 41 (17), 7868-7877*.

A Coronato, G De Pietro, L. G. An agent based platform for task distribution in virtual environments. In *Journal of Systems Architecture 54 (9), 877-882*.

A Testa, A Coronato, M. C. J. A. Static verification of wireless sensor networks with formal methods. In *2012 Eighth International Conference on Signal Image Technology and Internet*.

Antonio Coronato, Giovani Paragliola, M. N. G. D. P. (2019). Reinforcement learning-based approach for the risk management of e-health environments: A case study.

Chakraborty, B., . M. E. E. Statistical methods for dynamic treatment regimes.

D. F. Llorca, F. Vilarino, J. Z. G. L. (2007). A multi-class svm classifier for automatic hand washing quality assessment. *Proc. of the British Machine Vision Conference*.

Dearden, R., Friedman, N., and Andre, D. (2013). Model-based bayesian exploration. *CoRR*, abs/1301.6690.

G Paragliola, A. C. Gait anomaly detection of subjects with parkinson's disease using a deep time series-based approach. In *IEEE Access 6, 73280-73292*.

Giovani Paragliola, M. N. (2019). Risk management for nuclear medical department using reinforcement learning.

Gullapalli N Rao, L. V. P. (2002). How can we improve patient care?.

H. Kautz, L. Arnstein, G. B. O. E. D. F. (2002). An overview of the assisted cognition project. In *Proceedings on the 2000 Conference on Universal Usability*. AAAI-2002 Workshop on Automation as Caregiver: The Role of Intelligent Technology in Elder Care, Edmonton.

Kearns, M. J., Litman, D. J., Singh, S. P., and Walker, M. A. (2011). Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. *CoRR*, abs/1106.0676.

Lavori, P. W., . D. R. (2004). Dynamic treatment regimes: practical design considerations. clinical trials. 1(1).

Ling, Y., Hasan, S. A., Datla, V. V., Qadir, A., Lee, K., Liu, J., and Farri, O. (2017). Learning to diagnose: Assimilating clinical narratives using deep reinforcement learning. In *IJCNLP*.

Llorca, D. F., Parra, I., Sotelo, M. Á., and Lacey, G. (2011). A vision-based system for automatic hand washing quality assessment. *Machine Vision and Applications*, 22(2):219–234.

M.A.Sc., A. M., P.Eng., Ph.D., G. R. F., P.Eng., Ph.D., J. C. B., and C.Eng. (2001). The use of artificial intelligence in the design of an intelligent cognitive orthosis for people with dementia. *Assistive Technology*, 13(1):23–39.

Mozer, M. C. (2005). *Lessons from an Adaptive Home*, chapter 12, pages 271–294. John Wiley Sons Ltd.

Muddasar Naeem, Antonio Coronato, G. P. G. D. P. (2020). A cnn based monitoring system to minimize medication errors during treatment process at home.

Murphy, S. A., Oslin, D. W., Rush, A. J., and Zhu, J. (2006). Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders. *Neuropsychopharmacology*, 32(2):257–262.

Mynatt, E. D., Essa, I., and Rogers, W. (2000). Increasing the opportunities for aging in place. In *Proceedings on the 2000 Conference on Universal Usability*, CUU '00, pages 65–71, New York, NY, USA. ACM.

Oliver, N., Garg, A., and Horvitz, E. (2004). Layered representations for learning and inferring office activity from multiple sensory channels. volume 96, pages 163–180. Elsevier.

Patel, V. L., Shortliffe, E. H., Stefanelli, M., Szolovits, P., Berthold, M. R., Bellazzi, R., and Abu-Hanna, A. (2009). The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine*, 46(1):5 – 17. Artificial Intelligence in Medicine AIME' 07.

Petersen, B. K., Yang, J., Grathwohl, W. S., Cockrell, C., Santiago, C., An, G., and Faissol, D. M. (2018). Precision medicine as a control problem: Using simulation and deep reinforcement learning to discover adaptive, personalized multi-cytokine therapy for sepsis. *CoRR*, abs/1802.10440.

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., and Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results.

Prasad, N., Cheng, L., Chivers, C., Draugelis, M., and Engelhardt, B. E. (2017). A reinforcement learning approach to weaning of mechanical ventilation in intensive care units. *CoRR*, abs/1704.06300.

Reinkensmeyer, D. J., Guigon, E., and Maier, M. A. (2012). A computational model of use-dependent motor recovery following a stroke: Optimizing corticospinal activations via reinforcement learning can explain residual capacity and other strength recovery dynamics. *Neural networks : the official journal of the International Neural Network Society*, 29-30:60–9.

Ross, C. K., Steward, C. A., and Sinacore, J. M. (1993). The importance of patient preferences in the measurement of health care satisfaction. *Medical care*, pages 1138–1149.

Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition.

Sahba, F., Tizhoosh, H. R., and Salama, M. M. A. (2006). A reinforcement learning framework for medical image segmentation. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pages 511–517.

Strens, M. J. A. (2000). A bayesian framework for reinforcement learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pages 943–950, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Suresh, H., Hunt, N., Johnson, A. E. W., Celi, L. A., Szolovits, P., and Ghassemi, M. (2017). Clinical intervention prediction and understanding using deep networks. *CoRR*, abs/1705.08498.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Thall, P. F. and Wathen, J. K. (2005). Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. *Statistics in Medicine*, 24(13):1947–1964.

Welling, M. and Teh, Y. W. (2011). Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pages 681–688, USA. Omnipress.

Wren, C. R., Azarbayejani, A., Darrell, T., and Pentland, A. (1997). Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19:780–785.