Human-Aware Reinforcement Learning for Adaptive Human Robot Teaming

Saurav Singh
Electrical and Computer Engineering Department
Rochester Institute of Technology
Rochester, NY, USA
ss3337@rit.edu

Jamison Heard
Electrical Engineering Department
Rochester Institute of Technology
Rochester, NY, USA
jrheee@rit.edu

***Abstract*—Mistakes in high stress and critical multitasking environments, such as piloting an airplane and the NASA control room, can lead to catastrophic failures. The human's internal state (e.g., workload) may be used to facilitate a robot teammate's adaptations, such that the robot can interact with the human without negatively impacting overall team performance. Human performance has a direct correlation with workload states; thus, the human's internal workload state may be leveraged to adapt a robot's interactions with the human in order to improve team performance. A reinforcement learning-based paradigm that incorporates human workload states to determine appropriate robot adaptations is presented. Preliminary results using the proposed approach in a supervisory-based NASA MATB-II environment are presented.**

***Index Terms*—Soft Actor Critic; Workload; Human Robot Interaction; Reinforcement Learning**

I. Introduction

Humans working in high-stress and critical multitasking environments (e.g., piloting an airplane, search & rescue operations, and NASA control room) have to perform optimally due to a high failure cost. These stressful environments may increase the human workload considerably; causing an overload state and decreasing task performance [1]. The human may also become disengaged from the system if underloaded and fail to take timely action when needed. Task performance of such environments can be elevated by developing an adaptive human-robot teaming system where the robot adapts to the human's behavioral and workload states.

Overall workload can be divided into five components: cognitive, auditory, speech, visual and physical [2]. Physical workload can be further sub-divided into the fine motor, tactile, and gross motor [2]. Human workload can be subjectively estimated using questionnaires (e.g., the NASA-TLX [3]) or objectively estimated using physiological metrics [4]. Many researchers have combined such metrics using machine-learning algorithms to estimate human workload [5], [6].

Human workload has been used to adapt a robot's or system's autonomy level to increase team performance [7], [8]. However, these systems only consider a subset of the human's state (i.e., cognitive workload) and use rule-based control strategies (i.e., automate a task if the human is overloaded). Reinforcement learning may be used to determine more effective human-robot teaming control strategies [9]; however, these systems typically only consider the human's input's to the system and current task states. This paper focuses on augmenting a reinforcement learning agent's observation space with human workload state information in order to improve overall team performance. Additionally, a system aware of the human's multidimensional workload state information may permit individual adaptation strategies to emerge.

II. Algorithm

Reinforcement learning provides an approach where over time, a robot teammate can learn how its actions impact the human's workload and team performance. This work employs the Soft Actor-Critic (SAC) algorithm [10] to decide which task is automated in an adaptive autonomy paradigm. The Actor network, the two Critic networks, and the Value network each consisted of 2 fully connected layers with 64 neurons in each layer. The two critic networks and the Value network consisted of 1 neuron in the output layer and the Actor network used 2 neurons in the output layer (for the mean and standard deviation of the Gaussian distribution to sample the action). Relu and linear activation functions were used in the 2 fully connected layers and the output layers, respectively.

SAC requires three major components: *state*, *action*, and *reward*. Two state-space encapsulations are explored. **RL** encompasses task and interaction information and **RLH** augments **RL** with human workload information, as shown below:

$$S_{RL}(n) = \{I_1, I_2, \cdots, I_k, \cdots, I_K, a_{n-1}\} \quad (1)$$

$$S_{RLH}(n) = \{W_n, I_1, I_2, \cdots, I_k, \cdots, I_K, a_{n-1}\} \quad (2)$$

I_k represents the interaction information for the k^{th} task (time since last human interaction in seconds), K represents the total number of tasks, and a_{n-1} represents the last agent action taken. S_{RLH} augments S_{RL} with multidimensional workload information W_n, which consists of estimated overall workload w_o, cognitive w_c, auditory w_a, speech w_s, visual w_v, and physical w_p workload values. These values are estimated every 5-seconds using a workload assessment algorithm trained in a similar fashion as [11]. The incorporated algorithm uses a window size of 30 seconds consisting of the human's heart rate, heart rate variability, respiration rate, posture, syllables per second, speech pitch, speech intensity, and surrounding noise

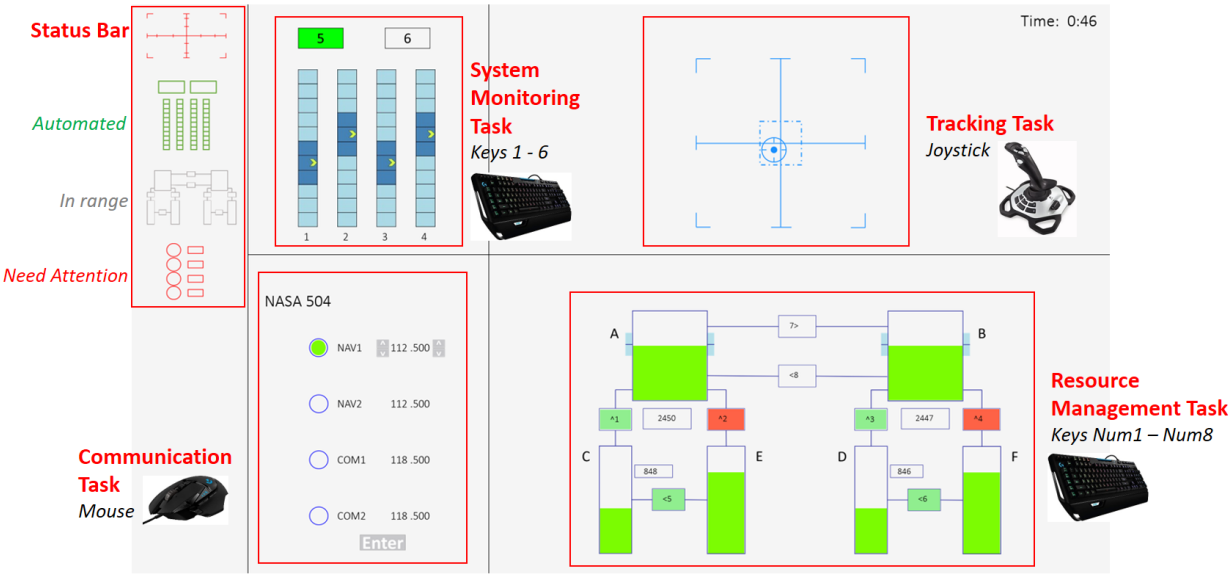

Figure 1. NASA Multi-Attribute Task Battery (MATB) Environment.

level. Features are then extracted and fed into a single deep network with five output neurons representing the human's cognitive, auditory, speech, visual, and physical workload states. These individual estimates are summed together to produce an overall workload estimate. Validating the incorporated workload assessment algorithm is out of scope for this work.

The SAC agent's continuous action space is discretized into $K+1$ possible actions (including no automation), where Task k is automated perfectly. A reward function r_k is developed for each individual task where poor task performance is penalized. Human idle time [12] (IT_h) is also penalized to prevent the agent from automating each task and disengaging the human. The total reward is computed by taking the weighted sum of the task rewards and the human idle time:

$$r = w_0 * r_0 + w_1 * r_1 + \cdots + w_K * r_K + IT_h \qquad (3)$$

The task weights were determined using average task rewards earned by an expert human (who has 20+ hrs. experience of working with the environment) ensuring that the task reward values are comparable. Data was collected with the expert going through the experiment without automation and the reciprocals of the magnitude of the overall average reward of each task (i.e., $w_k = 1/|mean(r_k)|$) were used as the task weights. This helped normalize the rewards of each task with respect to the expert's performance.

III. Experimental Design

Physiological, workload and task-related data were collected from two participants (Age: 22 and 58) in the NASA MATB-II task simulation environment [13] depicted in Figure 1. The participants were required to perform four tasks simultaneously: *Tracking*, *System Monitoring*, *Resource Management*, and *Communications*. The *Tracking* task consists of maintaining a crosshair on a target using a flight joystick, which impacts visual and physical workload. The performance metric is the root mean squared error between the position of the crosshair and the position of the target. The *System Monitoring* task consists of monitoring and resetting two alarm lights and four gauges using the keyboard, which increases the cognitive and visual workload. Performance is measured using response time to resetting an out-of-range light or gauge. The *Resource Management* task requires maintaining the fuel level of tanks A and B between the area marked by blue (between 2000 to 3000 units) by turning the appropriate pumps on/off using a keyboard, which increases cognitive and visual workload. Performance was measured as the percentage of time both tanks were in range. The *Communications* task required listening to audio commands: "NASA 504, NASA 504, turn you COM1 frequency to 128.450". The participant has to make the desired changes using the mouse and respond verbally: "This is NASA 504, COM1 set to frequency 128.450", which impacts auditory and speech workload. Response time to each request was used to measure performance.

The NASA MATB environment was modified to allow automation for each task, where automation was perfect. The status bar on the left of Figure 1 was also added for automation transparency. Each icon can turn red, grey, or green indicating that the task needs attention, the task parameters are in range, or the task is currently automated, respectively.

Workload was manipulated to be the independent variable to create three within-subject conditions: underload (UL), normal load (NL), and overload (OL). The participants were required to go through a 15-minute training session to get familiar with the task environment. This was followed by a 52.5 minute trial with a rule-based (RB) adaptive scheme that uses workload and task interaction data to automate the appropriate task, similar to the adaptive scheme used in [14]. The experiment was concluded with another 52.5-

Table I

CHANGE IN MEAN PERFORMANCE METRICS OF RLH AND RL ADAPTIVE AUTOMATION APPROACHES FOR THE HUMAN-ROBOT TEAMING ON NASA-MATB ENV. FROM BASELINES RB. BEST VALUES AMONG RLH AND RL IN EACH WORKLOAD CONDITION ARE REPRESENTED BY BOLD VALUES.

Workload Condition	Trial	Workload	Rewards	Tracking Error (pixels)	Gauges RT (seconds)	Lights RT (seconds)	Tanks in Range (%)	Comms RT (seconds)
Underload (UL)	RLH	0.08	**0.36**	**1.63**	-	2.82	**9.03**	-
	RL	**2.62**	-0.29	4.56	-	**0.17**	-4.16	-
Normal load (NL)	RLH	**-1.27**	-0.56	3.75	3.00	**-1.25**	-10.42	-0.55
	RL	9.58	**-0.02**	**1.11**	**-0.43**	-0.13	**0.7**	**-2.58**
Overload (OL)	RLH	**-6.42**	-3.16	18.25	**-1.00**	0.75	-70.30	**-1.79**
	RL	8.24	**0**	**1.99**	-0.27	**-0.09**	**0**	-0.3
Overall	RLH	**-2.5**	-1.15	7.82	-0.22	0.27	-23.56	**-1.56**
	RL	6.8	**-0.10**	**2.56**	**-0.32**	**-0.08**	**-1.17**	-0.71

Table II

AUTOMATION AND HUMAN INTERACTION TIMINGS WITH RLH AND RL ADAPTIVE AUTOMATION APPROACHES. MOST AUTOMATED TASKS BY RLH AND RL APPROACHES DURING EACH WORKLOAD CONDITION ARE REPRESENTED BY BOLD VALUES.

Workload Condition	Trial	Automation Time (seconds)				Interaction Time (seconds)			
		Tracking	Sys.Mon.	Res.Man.	Comms.	Tracking	Sys.Mon.	Res.Man.	Comms.
Underload (UL)	RLH	9	66	71	**275**	26	2	3	0
	RL	**86**	88	33	22	200	1	18	0
Normal load (NL)	RLH	22	66	127	**228**	29	11	2	7
	RL	**143**	74	55	22	148	7	13	11
Overload (OL)	RLH	0	33	110	**280**	48	13	1	27
	RL	**99**	91	66	11	121	47	0	49
Overall	RLH	31	165	308	**783**	103	26	6	34
	RL	**328**	253	154	55	469	55	31	60

minute session with either the RLH or RL approach as the between-subjects independent variable. Each trial consisted of seven consecutive 7.5-minute workload conditions (OL-UL-OL-NL-UL-NL-OL) to ensure that each workload transition was experienced. A 5-minute break occurred between the two trials in order to allow the participant's physiological signals to return to their resting state. The NASA Task Load Index (NASA-TLX) was completed after each trial to measure the subjective human workload [3] along with questions focused on trust and the participant's experience.

The SAC agents for the RLH and RL approaches were pre-trained with expert human interaction data. The interaction data collected from the participant during the Rule-Based trial was used to train the agent. The SAC agents are further trained during the first 29.5 minutes of the RLH/RL trial. The agent was not trained for the last 23 minutes of the RLH/RL trial so that the SAC agent's behavior is consistent. The data during this time was used for the evaluation of each agent paradigm.

IV. RESULTS & DISCUSSION

Preliminary results are presented for two reinforcement-learning adaptive automation approaches: SAC agent without human workload states (RL) and SAC agent with human workload states as part of the state space (RLH). Data from two participants (one with RLH and one with RL approach) was used to evaluate the two approaches. Table I shows the difference in the collected performance metrics for the RLH or RL approaches from their performances in the rule-based (RB) trial. Overall estimated workload for the RLH condition was lower than the workload during the respective RB trial; however, the mean reward was worse. The overall estimated workload for the RL condition was higher than the workload during the RB trial, but the rewards were similar. The RLH agent accumulated higher rewards during the underload condition than the RL agent, which achieved overall higher rewards. This result was not expected as the workload states provide more information to the SAC agent about the human teammate. Longer training times may be needed, as the multi-dimensional human workload states make the state space more complex.

Compared to RLH, the RL agent performed better in the tracking task with lower tracking error (lower is better), better in the system monitoring task with lower response time for lights and gauges (lower is better), better in resource management task with higher percent time when both tanks were in range (higher is better) but worse in the Communications task with higher response times to the audio commands (lower is better). An interesting trend regarding the automation behavior of the RLH agent was observed. Table II presents the automation and human interaction time by task and condition. The RLH agent automated the communications task the most throughout the trial. Accommodations made for the RLH participant may have impacted the results, as the volume of the audio commands was increased to the participant's comfort level. The high-volume audio commands were also picked up by the microphone used to compute the speech features for the workload estimates. This resulted in an abnormally high speech workload; influencing the agent to automate the com-

munications task. Automating the communication task more may be due to the agent using an individualized adaptation strategy, but more data is needed to be certain.

The RLH agent automating the communications task the most may have contributed to higher performance in the communications task as compared to the RL agent. On the other hand, the RL agent's automation decisions are more evenly spread out. The RL agent automated the tracking task the most, followed by the system monitoring task. Overall, the RLH agent automated at least one task for 1287 seconds out of the 1380 seconds (23 minutes) of data whereas the RL agent only automated tasks for a total of 790 seconds out of the 1380 seconds of data. This difference in automation time may be attributed to the RLH participant performing worse; hence, the agent automating more. Both participants interacted with the tracking task the most, due to it being the only task that requires constant physical interaction when not automated. However, it can be observed that the participant for the RLH trial had fewer interactions with the system in general which also might have contributed towards worse performance.

Since this is an ongoing study, data from more participants will be collected to better analyze the performance of the two proposed approaches. This is currently the main limitation of this work as drawing conclusions from the data of two participants (one with each approach) is not feasible. Another potential limitation of this work is the lack of algorithm training time. Additional training time may be needed for the agent to learn the relationship between human workload states.

V. Conclusion

This work presented an exploratory study focused on evaluating a human-aware reinforcement learning paradigm for adaptive human-robot teams. Two Soft Actor-Critic Reinforcement Learning-based approaches were presented that incorporate human interaction information and workload states to automate tasks in order to improve overall team performance. The human-aware SAC agent may have learned trends that are fine-tuned to the participant; however, more participant data is needed. Developing a human-aware reinforcement learning architecture may lead to robots or agents capable of tailoring their policies to a human teammate's internal state. Such an architecture may lead to more fluent team collaborations and improve team performance in dynamic task environments.

References

[1] C. D. Wickens, J. D. Lee, Y. Liu, and S. E. G. Becker, *An Introduction to Human Factors Engineering*, 2nd ed. Pearson Education, Inc., 2004.

[2] J. McCraken and T. Aldrich, "Implications of operator workload and system automation goals," U.S. Army Research Institution, Tech. Rep. ASI-479-024-84B, 1984.

[3] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Advances in Psychology*, pp. 139–183, 1988.

[4] S. Archer, M. Gosakan, P. Shorter, and J. Lockett, "New capabilities of the Armys maintenance manpower modeling tool," *Journal of the International Test and Evaluation Association*, vol. 26, no. 1, pp. 19 – 26, 2005.

[5] K. T. Durkee, S. M. Pappada, A. E. Ortiz, J. J. Feeney, and S. M. Galster, "System decision framework for augmenting human performance using real-time workload classifiers," in *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision*, 2015, pp. 8–13.

[6] M. Castor, *GARTEUR Handbook of Mental Workload Measurement*, ser. GARTEUR technical publications. Group for Aeronautical Research and Technology in Europe, 2003.

[7] S. Fuchs and J. Schwarz, "Towards a dynamic selection and configuration of adaptation strategies in augmented cognition," in *International Conference on Augmented Cognition*. Springer, 2017, pp. 101–115.

[8] D. B. Kaber and M. R. Endsley, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task," *Theoretical Issues in Ergonomics Science*, vol. 5, no. 2, pp. 113–153, 2004.

[9] S. Reddy, A. Dragan, and S. Levine, "Shared autonomy via deep reinforcement learning," in *Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018.

[10] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. PMLR, 2018, pp. 1861–1870.

[11] J. Heard, R. Heald, C. E. Harriott, and J. A. Adams, "A diagnostic human workload assessment algorithm for supervisory and collaborative human-robot teams," *ACM Transactions on Human-Robotic Interaction*, vol. 8, no. 2, pp. 1–30, 2019.

[12] G. Hoffman, "Evaluating fluency in human–robot collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, 2019.

[13] J. R. Comstock and R. J. Arnegard, "The multi-attribute task battery for operator workload and strategic behavior research," NASA Langley Research Center, Tech. Rep. NASA Tech. Memorandum 104174, 1992.

[14] J. Heard, J. Fortune, and J. A. Adams, "SAHRTA: A supervisory-based adaptive human-robot teaming architecture," in *IEEE Conference on Cognitive and Computational Aspects of Situation Management*, 2020.