TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 19, No. 3, June 2021, pp. 931~938
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v19i3.18319

Dialogue management using reinforcement learning

Binashir Rofi'ah, Hanif Fakhrurroja, Carmadi Machbub
School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Indonesia

Article Info

Article history:

Received Jul 14, 2020
Revised Oct 3, 2020
Accepted Oct 14, 2020

Keywords:

Dialogue management
Human-robot interaction
Knowledge growing
Reinforcement learning

ABSTRACT

Dialogue has been widely used for verbal communication between human and robot interaction, such as assistant robot in hospital. However, this robot was usually limited by predetermined dialogue, so it will be difficult to understand new words for new desired goal. In this paper, we discussed conversation in Indonesian on entertainment, motivation, emergency, and helping with knowledge growing method. We provided mp3 audio for music, fairy tale, comedy request, and motivation. The execution time for this request was 3.74 ms on average. In emergency situation, patient able to ask robot to call the nurse. Robot will record complaint of pain and inform nurse. From 7 emergency reports, all complaints were successfully saved on database. In helping conversation, robot will walk to pick up belongings of patient. Once the robot did not understand with patient's conversation, robot will ask until it understands. From asking conversation, knowledge expands from 2 to 10, with learning execution from 1405 ms to 3490 ms. SARSA was faster towards steady state because of higher cumulative rewards. Q-learning and SARSA were achieved desired object within 200 episodes. It concludes that reinforcement learning (RL) method to overcome robot knowledge limitation in achieving new dialogue goal for patient assistant were achieved.

Corresponding Author:

Hanif Fakhrurroja
School of Electrical Engineering and Informatics (SEEI)
Institut Teknologi Bandung
Ganesha No.10, Coblong, Bandung, Jawa Barat 40132, Indonesia
Email: hani002@lipi.go.id

1. INTRODUCTION

Betacoronavirus SARS-CoV-2 or Covid-19 outbreak is currently affecting almost all of the world with a total of 503,203 people infected or confirmed positive, and 22,340 of them died as of March 26, 2020 [1]. Since 11 March 2020, its status has changed to a pandemic. One of the conditions for a pandemic is the explosion of spread or the high number of cases that occur in a short period [2]. To reduce significant transmission rate, the mild-symptom patient must be isolated in the hospital until healed, which leads to stress [3]. The latest developments regarding companion/assistantship robots have been carried out [4], in order to support the mental health of patients, one of them is Silbot which has 6 activities such as waking the patient, checking the mood, reminding during meditation, checking safety, helping therapy, and emergencies. Silbot will care for mild-dementia patients [5].

Human-robot interaction (HRI) aims for extending robot functionality by making natural communication with human [6]. They [7] introduce algorithm from learning reward and human synchronously. The flexibility would gain from simultaneous learning; it gives a trainer ability to go in as desired and update reinforcement learning reward whereas it is still in progress. They [8] introduce learning scenario combination between practice and end-user critique, practice gives actual-world experience and

end-user/human critique whether good or bad label as input for loss function or unexpected value of candidate policies.

To build natural communication between human and robot for assistantships, our research team has prior works such as speech recognition [9], unclear pronunciation [10], robot walking and pattern generator [11, 12], robot path planning [13], speech and gesture recognition[14], multimodal interaction [15] and rule-based/scenario dialogue management [16]. However the problem of rule-based is unable to follow dialogue development, so the possibility of scenario mismatch is getting bigger, once the robot does not understand, no other choice for this method besides end the dialogue and gives generic answers like "I don't understand your commands" [17]. This paper takes part on overcome robot knowledge limitation to achieve new goal through flexible dialogue. We propose robot asking method to gather new knowledge from human feedback through conversation. The robot does not stop immediately because of not understand, but it will ask first and gather new knowledge with goal to take patient belongings. We also provide entertainment and emergency request to complement patient needs. This paper would discuss natural language processing both in understanding and generation with different sub-section, dialogue management method, also the hardware set-up for robot.

2. RESEARCH METHOD

2.1. Natural language understanding

This paper would focus on goal-driven dialogue. The function of natural language understanding (NLU) are extracting the raw voice until system got the information needed, and provide dialogue information for dialogue management. NLU includes identification of domain and intent, also semantic parsing [18]. Text will get several processes: Stopword process to remove unnecessary such as common words. Part of speech (POS) tagging process to get grammar tag with POStag_idn and use Indonesian tag set at [19]. We focused on tag VB (verb), NN (noun), and CD (cardinal number). We use Indonesia Tnt-Tagger for POS tag method and Indonesia IDPOSTAG corpus from [20]. Followed by stemming process to get root word by removing the affix [21]. Ended with storing process using JavaScript Object Notation (JSON) format.

2.2. Dialogue management

Dialogue management (DM) consists of state tracking and generates action. Approaches for dialogue management problems are graph-based dialogue, frame-based dialogue, statistical approach [22]. Human involvement in DM framework has been successfully carried out in previous studies [23]. The reward as feedback from expertise (can be formed in negative or positive rewards) was given to optimize policy on reinforcement learning (RL) in [24]. RL was still the main instrument for DM. RL is current mainstream technology in order to solve real-world problem with large-scale belief state space [18].

Before RL can be explained, it necessary to understand basic components used. A learner called an agent in RL studies its behavior by select actions in an environment [25]. At each time, the agent receives a representation of state s, while $s \in S$, where S is states. The agent pickups an action a, while $a \in A$, where A is a set of possible actions that the agent can take. As the return of its action, the agent receives reward r, while $r \in R$, and goes to new state s'. α is learning rate, γ is a discount factor, and π is policy that defines how an agent response from a specific state. The aim of an agent is selecting the optimal actions by maximizing its cumulative discounted reward.

In this paper, we use RL with temporal difference (TD) learning method. TD learning is a fusion of two benefits from Monte Carlo and dynamic programming as shown in (3), and (4). On one side, Monte Carlo methods have no model of environment's dynamics as shown in (1), so TD learns from raw experience. On the other side, dynamic programming (2) that no need waiting until the final outcome, so TD able to update estimates based on partially learned estimation [26]. Recall Monte Carlo:

$$MC: V^{\pi}(s) \leftarrow V^{\pi}(s) + \alpha[R(s) - V^{\pi}(s)], \alpha = \frac{1}{n_{R(s)}} \quad (1)$$

Recall dynamic programming:

$$DP: V^{\pi}(s) = \mathbb{E}_{\pi}[r_{t+1} + \gamma V^{\pi}(s_{t+1}) | s_t = s], \quad (2)$$

TD to make an update $V_{new}(s) \leftarrow V_{old}(s)$, Given (s, a, r, s'):

$$V_{new}(s) = (1-\alpha)V_{old}(s) + \alpha \underbrace{[r + \gamma V_{old}(s')]}_{TD\ Target} \quad (3)$$

$$V_{new}(s) = V_{old}(s) + \alpha \underbrace{[r + \gamma V_{old}(s') - V_{old}(s)]}_{TD\ Error} \quad (4)$$

The value function usually also called as state-value function $V(s)$ is the total amount of expected rewards that an agent can collect from that state to the end of the episode. The action-value function $Q(s,a)$ is total amount of expected rewards of taking an action from the state until the end of the episode. The way agent learns the best policy called update policy, and the way agent behaves called behavior policy. In this paper, we also implement two TD learning methods that are off and on policy (Q-learning and SARSA).

2.2.1. Q-learning

Absolute policy is used by agent in Q-learning to learn optimal policy, on the other hand, agent behaves with other policy. Because the behavior policy is different from update policy, so Q-learning is categorized as off-policy TD control. Q-value of Q-learning is shown in (5).

$$Q(S,A) \leftarrow Q(S,A) + \alpha[R + \gamma max_a Q(S',a) - Q(S,A)] \quad (5)$$

From (5) we have known that update policy $\gamma max_a Q(S',a)$ is different from behavior policy $Q(S,A)$. We use pseudocode from [26] to implement Q-learning in our python code as shown in Figure 1 (a).

2.2.2. State-action-reward-state-action (SARSA)

Agent in SARSA learns optimal policy and behaves with the same policy. Because the update policy and behavior policy are similar, so SARSA is categorized as on-policy. Q-Value of SARSA is shown in (6).

$$Q(S,A) \leftarrow Q(S,A) + \alpha[R + \gamma Q(S',A') - Q(S,A)] \quad (6)$$

From (6) we know that update policy $\gamma Q(S',A')$ and behavior policy $Q(S,A)$ also from pseudocode below, we know that $S \leftarrow S'$ and $A \leftarrow A'$ means update policy is the behavior policy. We use pseudocode [26] to implement SARSA in our python code as shown in Figure 1 (b).

Estimate $\pi \approx \pi_*$ with Q-learning (off-policy TD control)
Initialize $Q(s,a)$, for all $s \in S$, $a \in A(s)$ **Repeat** (each episode): Initialize S **Repeat** (each step): Choose A from S using π derived from Q Take action A, observe R, S' $Q(S,A) \leftarrow Q(S,A)$ $+ \alpha[R + \gamma max_a Q(S',a)$ $- Q(S,A)]$ $S \leftarrow S'$ **Until** S is terminal

(a)

Estimate $Q \approx q_*$ with SARSA (on-policy TD control)
Initialize $Q(s,a)$, for all $s \in S$, $a \in A(s)$ **Repeat** (each episode): Initialize S Choose A from S using π derived from Q **Repeat** (each step): Take action A, observe R, S' Choose A' from S' using π derived from Q $Q(S,A) \leftarrow Q(S,A)$ $+ \alpha[R + \gamma Q(S',A') - Q(S,A)]$ $S \leftarrow S'$ $A \leftarrow A'$ **Until** S is terminal

(b)

Figure 1. Pseudocode of RL with TD-policy: (a) Q-learning pseudocode and (b) SARSA pseudocode

2.3. Knowledge growing

Entertainment purpose consists of: *musik*/playing music audio, *dongeng*/playing fairy tale audio, *komedi*/playing comedy audio. Motivation purpose, consist of *motivasi*/Playing motivation audio. Emergency purpose: *memanggil perawat*/calling nurse, *keluhan*/reporting complaint of pain. Helping purpose, taking an object/patient's belongings. Researchers have emphasized on implementing robots that can imitate owning memory/knowledge to mitigate many social-robot challenges [27], some studies have exploit data, based on user profile [28, 29] to make memory-based adaptations. We implement robot asking during interaction to gather new information from human feedback, Figure 2 (a) is an example of additional knowledge.

Entertainment, motivation, and emergency need back-end intervention from admin to add appropriate content manually. In emergency, the robot behavior (calling nurse in fix place, where robot moves, and what robot talk) cannot be changed by user/patient. Meanwhile, helping is moving action from robot that depends on user/patient habit on locating his/her belongings, so it can be useful to use end-user experience. Only for this kind of action robot will grow up its knowledge.

The helping conversation can be seen in Figure 2 (a). Grey shades show the unknown of robot, then from the conversations that we proposed then appear words in cyan, yellow, and orange shades, that is new knowledge. New knowledge will be saved at Q-Table shown in Figure 2 (b) for further training, then RL both Q-learning and SARSA takes part on build new environment by calculating rewards at each action and state.

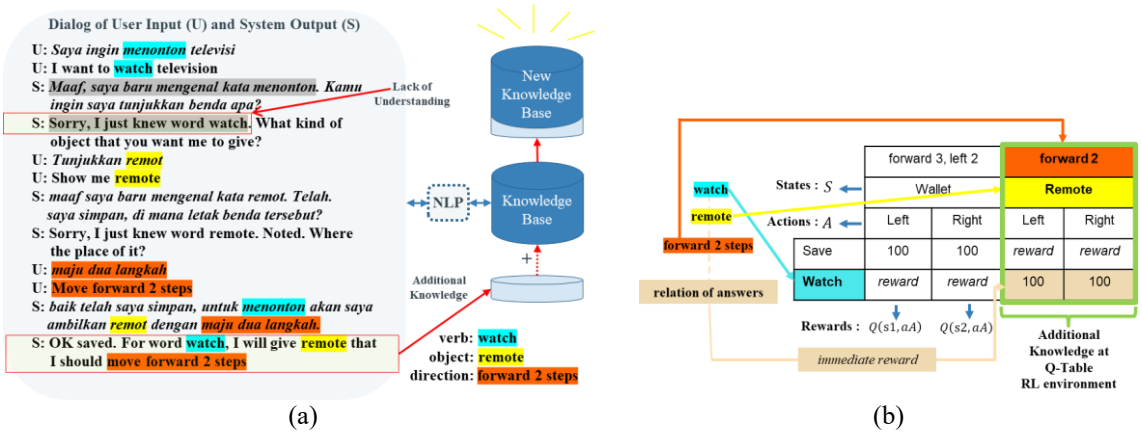


Figure 2. (a) Additional knowledge gathers from dialogue and (b) New knowledge stored in database

2.4. Natural language generation

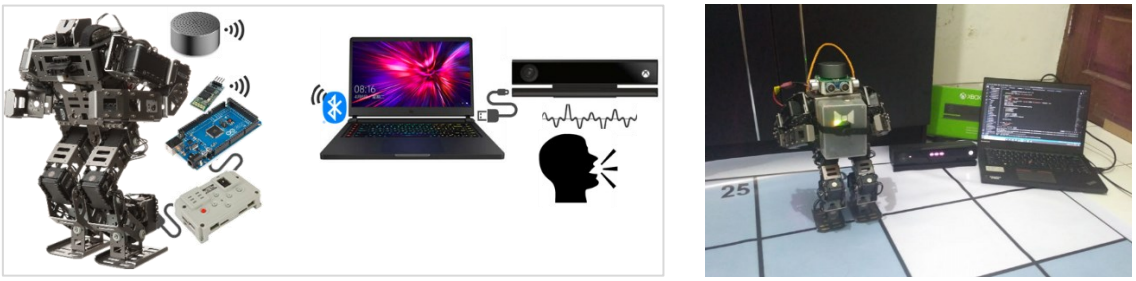
Natural language generation (NLG) is responsible to generate linguistic realization of the system's dialogue. The goal of NLG is to produce spoken that is easy for humans to understand. In this paper we had 3 response systems there are rejection, asking, and aborting. Once system found that all word in a sentence has no verb (listed on corpus) or unique words, system will reject and request to change with other new words until there is verb or unique word in that sentence. Asking response is started with searching verb in system database knowledge, if there is no similar verb then system will categorize it as new verb with no relation to object. The system will ask for object then searching the word in corpus, if there is object in the corpus then system will search in database. That is why some verbs can have one same object. After the system has new verb and new object, then system will ask for place, if system able to fulfill direction and iteration, then it will save as new knowledge. Aborting response is where the system will able to abort mission if user says *terima kasih*/thank you in the middle of asking conversation.

2.5. Humanoid robot

Bioloid grand prix (GP) is a humanoid robot equipped with CM-530 controller, and lithium battery for power supply [30]. We use modified Bioloid GP from [13] as previous project with an additional speaker mounted on top of the robot. Analog voltages from Arduino Mega 2560 [31] are converted to digital values by analog to digital converter (ADC) as a reference command for CM-530 that will be translated into robot movement. Robot movement consists of forward, backward, left, and right with its iteration.

2.6. System implementation

The hardware needed for this system is a microphone input (Kinect 2.0), processor (Laptop), controller (Arduino and CM-530), and output in the form of speakers and robots as shown in Figure 3 (a). In the hardware implementation, robot able to move everywhere without wire on cable as shown in Figure 3 (b). Speech output and robot movement control are sent from laptop to Arduino via bluetooth. We used Google speech recognition with id-ID (Indonesian language) to recognize and adjust ambient noise. Laptop powered by the Intel Core i5 processor, 8GB of memory. We use python language and RL algorithm builds on it. We equipped robot voice with speech registry from Windows called Microsoft Andika to give Indonesian voice and accent, also pronounce cardinal number in Indonesian. We set robot to talk 150 words per minute (WPM). The average speech rate for conversational is 120-150 WPM [31].



(a) (b)

Figure 3. System configuration: (a) design, and (b) implementation

3. RESULTS AND ANALYSIS

We conducted several experiments to see the performance of system. System configuration and environment as shown in Figure 3 (b). The performance will represent how fast robot execution, how accurate, and how knowledge growing. The experiment consists of entertainment execution, emergency execution, helping conversation with knowledge growing, policy behavior, and reward convergence.

3.1. Entertainment execution

This experiment gived us insight on execution time for single request. We implemented using 1 m fixed distance. Time counted right after translation from speech to text, we did it because the length of dialogues and the speed of people's speech rates varied. An average time was 3.74 ms. The slowest time occurs on purple shade with 8.23 ms. The fastest time was 0.86 ms with pink shade shown in Table 1.

Table 1. Time for entertainment execution

Dialogue	*Musik*/Music	*Dongeng*/Fairy tale	*Motivasi*/Motivation	*Komedi*/Comedy
[request]	2.99	2.00	3.92	3.91
nyalakanlah [request] (please turn on [request])	5.76	1.99	5.18	3.60
mainkanlah [request] (please play [request])	2.00	2.01	5.34	4.16
mainkan [request]nya (play the [request])	2.01	2.00	2.79	0.86
saya ingin mendengarkan [request] (I want to listen to [request])	2.00	1.03	7.47	4.90
[request] yang bagus (please the best [request])	8.23	2.00	6.35	6.81
saya bosan ingin [request] (I am bored, want to [request])	2.55	2.03	1.99	5.85
saya butuh [request] (I need [request])	3.91	2.00	7.10	3.91
[request]nya, tolong diputar (the [request], please play)	2.00	2.03	3.99	6.60
minta tolong [request] (please [request])	7.48	1.03	3.36	4.31
Average time level			3.74	

3.2. Emergency execution

In emergency situation, we asked robot to call nurse by talk unique word "*perawat (*nurse)". We use sentence *"panggilkan perawat* (call the nurse)", then robot will ask for complaint of pain. After conversation, robot will walk to the place where the nurse usually standby and describe complaint of pain to the nurse. On the other hand, the complaint will record on a report shown in Table 2.

3.3. Knowledge growing

In beginning there were only 2 verbs and 2 objects, then during this experiment, human gives unknown knowledge to robot. From the conversations, knowledge expanded to 10 verbs and 8 objects. We also tried different verbs related to same object. *Tonton*/watch and *lihat*/see have the same object that was remote. *Tulis*/write and *catat*/record have same object, that was pencil.

We separated Q-Table for Q-learning and SARSA because the rewards are different. Q-learning reward is shown in Table 3, the blue shades mean the highest reward that was connected between verb and

object on the Q-table. Whereas Table 4 was the final SARSA Q-Table which has more null/zero values. In SARSA policy, when state is terminal then reward will be grounded to zero. The highest rewards are shown in orange shades. We also did an experiment to execute training from 1 knowledge to 10 knowledge in 200 episodes. Every execution iterates 3 times for Q-learning and SARSA. From knowledge 1 until 5, time was varying, however from 6 knowledge, time consistently ramp up from 1405 until 3490 ms, and Q-learning needs more time than SARSA as shown in Figure 4.

Table 2. Report of emergency/complaint of Pain

Date Time	Complaint of Pain
8/11/2020 22:44	*saya pusing mual* (I feel dizzy and nausea)
8/12/2020 16:21	*saya nyeri* (I am in pain)
8/12/2020 16:21	*infus saya lepas* (My infusion peels off)
8/12/2020 16:22	*saya batuk darah* (I coughed up blood)
8/12/2020 16:32	*saya sesak nafas* (I am short of breath)
8/12/2020 16:33	*saya meriang* (I feel light-headed)
8/12/2020 16:33	*saya sakit perut tiba tiba* (I have a sudden stomachache)

Table 3. Q-Table for Q-Learning

	Wallet		Pencil		Blanket		Puzzle		Remote		Book		Barbell		Tissue	
	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R
Save	None	100	99	80	89	72	80	65	72	59	65	53	59	47	53	None
Write	None	92	100	100	99	80	89	72	80	65	72	59	65	53	59	None
Note	None	89	100	100	99	80	89	72	80	65	72	59	65	53	59	None
Sleep	None	88	73	99	100	100	99	80	89	72	80	65	72	59	65	None
Play	None	80	72	89	80	99	100	100	99	80	89	72	80	65	72	None
Watch	None	72	65	80	72	89	80	99	100	100	99	80	89	72	80	None
Read	None	65	59	72	65	80	72	89	80	99	100	100	99	64	85	None
See	None	72	65	80	72	89	80	99	100	100	99	79	89	71	80	None
Exercise	None	59	53	65	59	72	66	81	73	89	80	99	100	100	96	None
Clean	None	53	47	59	53	65	59	72	65	80	72	89	80	99	100	None

Table 4. Q-Table for SARSA

	Wallet		Pencil		Blanket		Puzzle		Remote		Book		Barbell		Tissue	
	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R
Save	None	100	99	0	89	0	80	0	72	0	65	0	59	0	53	None
Write	None	93	100	100	99	0	89	0	80	0	72	0	65	0	59	None
Note	None	91	100	100	99	0	89	0	80	0	72	0	65	0	59	None
Sleep	None	85	0	98	100	100	99	0	89	0	80	0	72	0	65	None
Play	None	80	0	89	0	99	100	100	99	0	89	0	80	0	72	None
Watch	None	72	0	80	0	89	0	99	100	100	99	0	89	0	80	None
Read	None	65	0	72	0	80	0	89	0	99	100	100	99	0	88	None
See	None	72	0	80	0	89	0	99	100	100	99	0	89	0	80	None
Exercise	None	59	0	65	0	72	0	80	0	89	0	99	100	100	81	None
Clean	None	53	0	59	0	65	0	72	0	80	0	89	0	99	100	None

Figure 4. Execution time as knowledge increase

3.4. Policy behavior

To know the movement of policy and actions taken in a certain state to reach appropriate object, we also taking plots for reward value at the end of episode (200th episode). As shown in Figure 5 reward shift towards *remot*/remote in the middle. Red shade means the lowest reward, where green shade is the highest reward (goal). The yellow shades describe the transition of reward value from the lowest to the highest.

	Wallet Dompet		Pencil Pensil		Blanket Selimut		Puzzle Puzzle		**Remote Remot**		Book Buku		Barbells Barbel		Tissue Tisu	
	left	right	left	right	left	right	left	right	**left**	**right**	left	right	left	right	left	right
Q-Learning	0.00	72.89	65.59	80.99	72.88	90.00	80.94	99.99	**100.00**	**100.00**	99.91	79.40	89.73	71.11	80.22	0.00
SARSA	0.00	72.86	0.00	80.98	0.00	89.99	80.94	99.99	**100.00**	**100.00**	99.97	0.00	89.94	0.00	80.87	0.00
	right		right		right		right		done		left		left		left	

Figure 5. Left and right policy direction to object "*remot*/remote"

3.5. Reward convergence

On this implementation, we want to know the performance of Q-learning and SARSA for every object in each final reward for 200 episodes. Start from 1 to 7 objects. It can be seen in Figure 6 that the SARSA cumulative reward was slightly higher than Q-Learning, which means that its algorithm was faster towards steady states because SARSA's policy does not explore all actions at each step so that it was focused to get the goal.

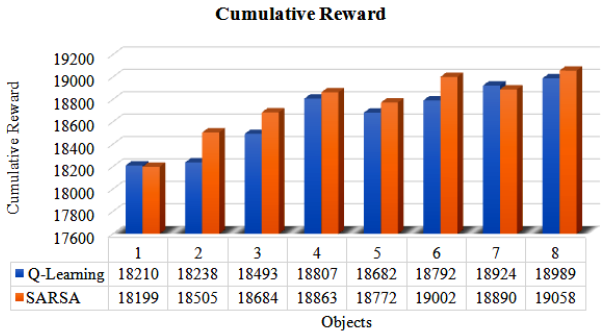


Figure 6. Cumulative reward between Q-learning and SARSA

4. CONCLUSION

From the experiment in previous section, it could be shown that the proposed system has the ability to expand from 2 to 10 knowledge. Additional knowledge affected to the time for learning execution that was getting longer from 1405 ms to 3490 ms. SARSA was faster towards steady state because of higher cumulative rewards. However, the difference between off and on learning can still be implemented, and the policy moves the action accordingly to achieve the desired object in 200 episodes. Equipped with entertainment feature to play music, fairy tale, motivation, and comedy request in fast average execution time of 3.74 ms. During emergency situation system able to call nurse and save 7 complaints of pain. It could be concluded that the method proposed in this paper successfully achieved the objective to overcome robot knowledge limitation in achieving new dialogue goal for patient assistant. For further research, dialogue classification and knowledge growing can be extended for chit-chat dialogue or non-goal driven.

REFERENCES

[1] C. F. Santos, "Reflections about the impact of the SARS-COV-2/COVID-19 pandemic on mental health," *Rev. Bras. Psiquiatr.*, vol. 42, no. 3, p. 329, 2020, doi: 10.1590/1516-4446-2020-0981.
[2] D. M. Morens, G. K. Folkers, and A. S. Fauci, "What Is a Pandemic?," *J. Infect. Dis.*, vol. 200, no. 7, pp. 1018-1021, 2009, doi: 10.1086/644537.
[3] Center for the Study of Traumatic Stress, "Psychological Effects of Quarantine During the Coronavirus Outbreak: What Healthcare Providers Need to Know," pp. 1-2, 2020.

[4] M. Gombolay *et al.*, "Robotic assistance in coordination of patient care," *Robot. Sci. Syst.*, vol. 12, 2016.
[5] M. Law *et al.*, "Developing assistive robots for people with mild cognitive impairment and mild dementia: A qualitative study with older adults and experts in aged care," *BMJ Open*, vol. 9, no. 9, 2019.
[6] J. Peltason and B. Wrede, "Modeling human-robot interaction based on generic interaction patterns," *AAAI Fall Symp. - Tech. Rep.*, vol. FS-10-05, pp. 80-85, 2010.
[7] W. B. Knox and P. Stone, "Reinforcement learning from simultaneous human and MDP reward," *11th Int. Conf. Auton. Agents Multiagent Syst. 2012, AAMAS 2012 Innov. Appl. Track*, vol. 1, pp. 528-535, 2012.
[8] K. Judah, S. Roy, A. Fern, and T. G. Dietterich, "Reinforcement learning via practice and critique advice," *Proc. Natl. Conf. Artif. Intell.*, vol. 1, pp. 481–486, 2010.
[9] K. Prepin and A. Revel, "Human-machine interaction as a model of machine-machine interaction: How to make machines interact as humans do," *Adv. Robot.*, vol. 21, no. 15, pp. 1709-1723, 2007.
[10] D. Handaya, H. Fakhruroja, E. M. I. Hidayat, and C. Machbub, "Comparison of Indonesian speaker recognition using vector quantization and Hidden Markov Model for unclear pronunciation problem," *Proc. 2016 6th Int. Conf. Syst. Eng. Technol. ICSET 2016*, pp. 39-45, 2017, doi: 10.1109/FIT.2016.7857535.
[11] Riyanto, W. Adiprawita, H. Hindersah, and C. Machbub, "Center of Mass based Walking Pattern Generator with Gravity Compensation for Walking Control on Bioloid Humanoid Robot," *2018 15th Int. Conf. Control. Autom. Robot. Vision, ICARCV 2018*, pp. 54-59, 2018, doi: 10.1109/ICARCV.2018.8580633.
[12] Riyanto, C. Machbub, H. Hindersah, and W. Adiprawita, "Slope balancing strategy for bipedal robot walking based on inclination estimation using sensors fusion," *Int. J. Electr. Eng. Informatics*, vol. 11, no. 3, pp. 527-547, 2019.
[13] M. Kusuma, Riyanto, and C. Machbub, "Humanoid Robot Path Planning and Rerouting Using A-Star Search Algorithm," *Proc. - 2019 IEEE Int. Conf. Signals Syst. ICSigSys 2019*, pp. 110-115, 2019.
[14] H. Fakhrurroja, A. Purwarianti, A. S. Prihatmanto, and C. Machbub, "Integration of Indonesian Speech and Hand Gesture Recognition for Controlling Humanoid Robot," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2018, pp. 1590-1595.
[15] H. Fakhrurroja, A. S. Prihatmanto, and C. Machbub, "Multimodal Interaction System for Home Appliances Control," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 15, 2020.
[16] D. A. Permatasari, H. Fakhrurroja, and C. Machbub, "Human-Robot Interaction Based On Dialog Management Using Sentence Similarity Comparison Method," *Int. J. Adv. Sci. Technol.*, vol. 10, no. 5, 2020.
[17] E. Merdivan, D. Singh, S. Hanke, and A. Holzinger, "Dialogue Systems for Intelligent Human Computer Interactions," *Electron. Notes Theor. Comput. Sci.*, vol. 343, pp. 57-71, 2019.
[18] X. Wang and C. Yuan, "Recent Advances on Human-Computer Dialogue," *CAAI Trans. Intell. Technol.*, vol. 1, no. 4, pp. 303-312, 2016.
[19] A. Dinakaramani, F. Rashel, A. Luthfi, and R. Manurung, "Designing an Indonesian part of speech tagset and manually tagged Indonesian corpus," *Proc. Int. Conf. Asian Lang. Process. 2014, IALP 2014*, pp. 66-69, 2014.
[20] F. Rashel, "Indonesia Tagged Corpus," *Designing an Indonesian Part of speech Tagset and Manually Tagged Indonesian Corpus*. [Online]. Available: https://github.com/famrashel/idn-tagged-corpus.
[21] A. S. Rizki, A. Tjahyanto, and R. Trialih, "Comparison of stemming algorithms on Indonesian text processing," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 17, no. 1, pp. 95–102, 2019.
[22] S. Yi and K. Jung, "A Chatbot by Combining Finite State Machine, Information Retrieval, and Bot-Initiative Strategy," *Alexa Price Proc.*, pp. 1-10, 2017.
[23] P. Shah, D. Hakkani-Tür, and L. Heck, "Interactive reinforcement learning for task-oriented dialogue management," *NIPS Work.*, no. i, p. 11, 2016.
[24] E. Ferreira and F. Lefèvre, "Reinforcement-learning based dialogue system for human-robot interactions with socially-inspired rewards," *Comput. Speech Lang.*, vol. 34, no. 1, pp. 256-274, 2015.
[25] S. Sendari, A. N. Afandi, I. A. E. Zaeni, Y. D. Mahandi, K. Hirasawa, and H. I. Lin, "Exploration of genetic network programming with two-stage reinforcement learning for mobile robot," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 17, no. 3, pp. 1447-1454, 2019.
[26] R. S. Sutton, *Reinforcement learning: an introduction*, Second. Cambridge, MA: The MIT Press, 2018.
[27] M. I. Ahmad, O. Mubin, and J. Orlando, "A systematic review of adaptivity in human-robot interaction," *Multimodal Technol. Interact.*, vol. 1, no. 3, 2017.
[28] M. I. Ahmad, O. Mubin, and J. Orlando, "Adaptive Social Robot for Sustaining Social Engagement during Long-Term Children–Robot Interaction," *Int. J. Hum. Comput. Interact.*, vol. 33, no. 12, pp. 943-962, 2017.
[29] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Empathic Robots for Long-term Interaction: Evaluating Social Presence, Engagement and Perceived Support in Children," *Int. J. Soc. Robot.*, vol. 6, no. 3, pp. 329-341, 2014.
[30] Robotis, "Robotis e-Manual Bioloid GP." [Online]. Available: https://emanual.robotis.com/docs/en/edu/bioloid/gp/#references. [Accessed: 06-Mar-2020].
[31] D. Barnard, "Average Speaking Rate and Words per Minute," 2018. [Online]. Available: https://virtualspeech.com/blog/average-speaking-rate-words-per-minute. [Accessed: 12-Jul-2020].