Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 555–559 DOI: 10.15439/2024F3214 ISSN 2300-5963 ACSIS, Vol. 39

Disease Diagnosis On Ships Using Hierarchical Reinforcement Learning

Farwa Batool
Quaid-i-Azam University
Islamabad Pakistan
Email:farwabatool@ele.qau.edu.pk

Tehreem Hasan
Quaid-i-Azam University
Islamabad Pakistan
Email:tehreemhasan@ele.qau.edu.pk

Giancarlo Tretola
Department of Computer Engineering
Università Giustino Fortunato
Benevento Italy
Email: g.tretola@unifortunato.eu

Zaib Ullah
Department of Computer Engineering
Università Giustino Fortunato
Benevento Italy
Email: z.ullah@unifortunato.eu

Musarat Abbas
Quaid-i-Azam University
Islamabad Pakistan
Email: mabbas@qau.edu.pk

***Abstract*—Every year about 30 million people travel by ship worldwide often in extreme weather conditions and polluted environments and many other factors that impact the health of passengers and crew staff. Such issues require medical staff for passenger health care. We introduce a model based on Reinforcement learning(RL) which is used in the dialogue system. We incorporate the Hierarchical reinforcement learning (HRL) model with the layers of Deep Q-Network for dialogue oriented diagnosis system. Policy learning is integrated as policy gradients are already defined. We created a two-stage hierarchical strategy. We used the hierarchical structure with double-layer policies for automatic disease diagnosis. A double layer means it splits the task into sub-tasks named high-state strategy and low-level strategy. It has a user simulator component that communicates with the patient for symptom collection low-level agents inquire about symptoms. Once it's done collecting it sends results to the high-level agent which activates the D-classifier for the last diagnosis. When it's done its sent back by the user simulator to patients to verify the diagnosis made. Every single diagnosis made has its reward that trains the system**

I. Introduction

MARITIME TRANSPORTATION plays a vital role in global trade and passenger transport contributing to economic development and connectivity [14]. Maritime transportation is the backbone of global trade, as ships carry over 80 percent of trading goods worldwide [34]. Almost every industry is changing due to technology and new methods of operation, but the maritime sector is currently seeing this transition most quickly [26]. Further investigation provides insights into the function of innovative communications technology, including virtual telemedicine and secure radio expertise, and assesses their practicality in the context of emergency maritime medicine [8], [12]. There is always a need of medical facilities for passengers and crew members. One of the biggest challenge in it is timely and accurate diagnosis of disease.As ships have limited resources and lack of medical staff on board so we can not relay on traditional methods. So we move towards Machine learning and Artificial Intelligence (AI) to train system to do automatic diagnosis [2].

AI has emerged as a revolutionary force in many field like 5G vehicular networks [10], rehabilitation [24], MIMO communication [17] and also in healthcare, offering new methods to the way we do disease identification, its treatment, and tracking. The implementation of AI in healthcare is enhancing diagnostic accuracy [15]. Specially ,Hierarchical reinforcement learning (HRL) is a promising method to extend traditional reinforcement learning to solve more complex tasks [38]. Hierarchical reinforcement learning (HRL) provides more broad sprctrum to Rl, by offering a divide-and-conquer methodology. In this methodology, the intricate and challenging problems,are divided into multiple smaller problems. These divided problems are easier to solve and their solutions can be regenerative to solve other related problems. This methodology has preceding been successfully used to speed up many offline preparing and organising algorithms where the variables of the environment are known in advance [7]. Hierarchical reinforcement learning (HRL) is a layered algorithm based on RL. HRL has been evidenced to be efficient in challenges with deferred and infrequent rewards and minimizing the learning difficulty by splitting the long-term goal into stages [35]. The symptom collection process of multiple phases of consultation between the agent and the patient as a Markov decision process, and uses the reinforcement learning algorithm for training [30].

our contribution is implementing the HRL by assigning rewards to correct symptom query in result of agent collecting the symptom and relating it with certain disease.policy learning is integrated as policy gradients are already defined.As we are using hierarchical reinforcement learning it creates two stage hierarchical strategy, fist stage is high level strategy which triggers the low level strategy. Low level strategy have multiple agents working as symptoms checkers and disease classifiers. Each Agent is responsible for investigating certain types of diseases.At the end we have disease classifier which is responsible to check responses from all agents and conclude disease diagnosed. Every disease have relation with symptoms and symptoms are also related with more than one disease.So for achieving maximum accuracy its necessary to understand symptoms and narrow down options of diseases at every single question with dialogue simulator. Now on ships as we have limited medical staff so its doing diagnosis using HRL, in which we have Agents every single agent is specialized for certain field providing broad spectrum of diseases to be diagnosed. The paper organised as follows, firstly we have related work. As Reinforcement learning specifically hierarchical reinforcement learning is emerging and is popular for classification, So we mentioned worked done earlier.Secondly proposed framework model is which explains all components in the model that includes leader, agent, user simulator, d-classifier.Its shown in detail in figure 1.Thirdly we have benchmark models which describe all the best models we are comparing with.Lastly we have results and conclusions.

II. Related Work

This section outlines some related works on the use of reinforcement learning for healthcare problems.

Dynamic Treatment Regime (DTR) is has an importance in healthcare as well as for medical research. DTR are considered as sequence of alternative treatment paths and any of these treatments can be adapted depending on the patient's conditions [6]. Therefore, the authors in [22] apply a cooperative imitation learning approach to utilize information from both negative and positive trajectories to learn the optimal DTR. The given framework minimizes the chance of choosing any treatment that results in a negative outcome during the medical examination. However, the proposed work is not suitable to employ for the disease diagnosis on ships.

Online symptom checkers by [20] have been put into action to recognise the possible causes and treatments for diseases based on a patient's symptoms. The work in [11] uses deep RL for fast disease diagnosis. Similarly, authors in [25] utilize an approach of automatic development of a dialogue manager capable of doing goal-oriented dialogues for the health domain. While the work in [29] employs a hierarchical RL is used for automatic captioning the video.

A machine learning method upper confidence bound is utilized in [16] to assist patients during their medication process at home. Authors considered the cognitive and physical impairments of the patients in the training of the machine learning model. A similar work is also done in [5] but with the help of Thompson sampling method. However, these systems are useful to specific scenarios during medication at home.

An end to end multi-channel conversational interface for dynamic and co-operative target setting is developed in [29], which integrates collective reward (task/persona/sentiment) for task success, personalized augmentation and user-adaptive behavior. Furthermore, an automatic diagnostic system is designed in [27] by applying both evident and inherent symptoms utilized by the Deep-Q Network Reinforcement Policy.

Moreover, there are some AI based solutions for the continuous and remote monitoring of unpredictable health issues. Such a failure mode and effect analysis is given in [4] and [3] for a specific mobile health monitoring system. Both of these systems were designed to provide remote healthcare solutions but these are for certain cases and environments and cannot be generalised for other cases.

The works in [19] and [23] use AI techniques for risk management in nuclear medication department. The later will is the extension of former one and discuss the risk cases during examination at such departments. Although, the proposed systems are useful to avoid possible risk at nuclear medication departments but are not useful for healthcare solutions at ships. an End-to-End Knowledge-routed Relational Dialogue System (KR-DS) that enables dialogue management, natural language understanding, and natural language generation to cooperatively optimize via reinforcement learning is presented in [1]. [32]. Q-learning algorithm is used in [18] to create an optimal controller for cancer chemotherapy drug dosing. Major depressive disorder treatment is considered in [21]. The authors have utilized the strong transfer ability of HRL to build a cross-domain dialogue system, which learned shareable information in similar subdomains of different main domains to train a general underlying policy.

Hybrid and hierarchical RL methods gained significant attention in recent years [10]. The proposed work presents extended RL structure as hierarchical structure that has two-stage policies for automatic diagnosis. it has hierarchical structure with double layer policies for automatic disease diagnosis.Double layer means it splits the task into sub-tasks named as high-state strategy and low level strategy.User simulator communicates with patient for symptom collection low level agent inquire symptoms. Once its done collecting it sends results to high level agent which activates the D-classifier for last diagnosis.When its done its send back by user simulator to patients to verify diagnosis made.

III. Model Framework

The disease diagnosis model finds the policy π for the maximum reward. For disease diagnosis Markov decision process is used in which M = [S,A,P,γ] [9]. S is the state, S^h is state in high stage strategy, S^{li} is state in low-stage strategy, n is the number of low strategy agents. All states can be expressed as $S = S^h \bigcup \{S^{li}\}_{i=1}^n$. For actions, A^h is high stage agent's action ,A^{li} is low stage action. n is the number of low strategy agents. All actions are expressed as $A = A^h \bigcup \{A^{li}\}_{i=1}^n$ All dialogue rewards is shown by R. State transition model is shown by P. γ is the discount rate used to compute Q value function. The major aim is to optimize Markov decision process M = [S,A,P,Γ] and identify the policy π that elevate the cumulative discount reward for all (S,A).

In this paper, we extend simple RL structure into hierarchical structure that has two-stage policies for automatic diagnosis. Framework is shown in the Figure 1 it is hierarchical structure with double layer policies for automatic disease diagnosis. Double layer means it splits the task into sub-tasks named as high-state strategy and low level strategy. Idea is inspired by hospital consultation in real world. It works in a way that

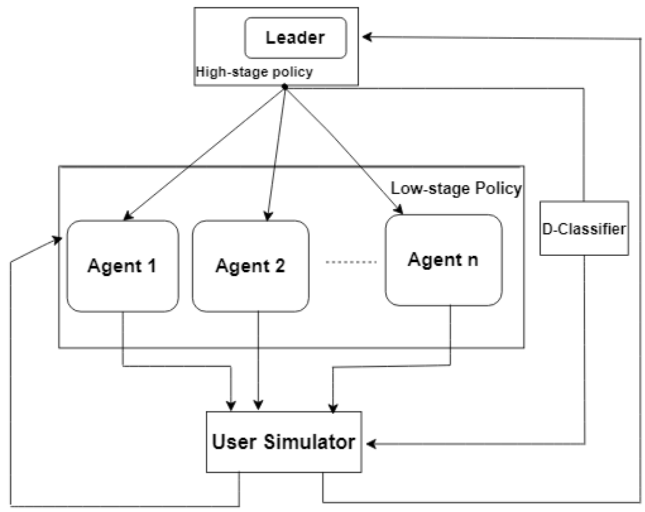

Fig. 1. Model Diagram

High-state Agent gets the current initial state as S_t, then it appoints a low level agent to communicate with user simulator for symptom collection. In Figure 1,it has four main parts Leader, Agent, User simulator and Disease classifier. Current initial state as S_t is encoded as a vector that depicts the level of each symptom and also about number of iterations necessary.

Consider a doctor that asks symptoms from patient. They will first consider that patient have certain disease and start asking related symptoms. Similar to that agent chooses a symptom to inquire the patient $A_t \in$ SṪhe possible user responses could be (true/false/unknown). If a_t is element in set of diseases, agent will inform user about diagnosis made, and diagnosis is made dialogue session would end and accuracy depends on correctness of diagnosis.

A. Strategy of Leader Model

In leader model its main task is to figure out if its activating the D-classifier or the agent to collect more symptoms. Once the leader activates the agent it will interact with user N number of cycles (dialogue rounds) until sub task is terminated. For action a_l^t the reward of the leader is r_t^l. Γ is the discount factor and $r_{t+t'}^e$ is the reward given by user simulator to low stage agent for current cycle. One reward is generated for disease classifier shown as r_t^e. In formula d is the action to activate the agent A^i.

$$r_t = \begin{cases} \sum_{t'=1}^{N} \Gamma^{t'} r_{t+t'}^e, & \text{if } a_t^l = A^i \\ r_t^e, & \text{if } a_t^l = d \end{cases} \quad (1)$$

The rewards obtained from user simulator will be aggregated as the reward of High-stage agent which is the high-stage reward calculated in equation(1).

B. Strategy of Agent Model

The objective of agent is to optimize the expected cumulative discounted reward. For that we use bellman equation in it Q-value function illustrates cumulative reward. In equation θ_l is the parameter of present policy network. Action of agent is shown by a_t^l and after taking action the next dialogue state is S_{t+1} to the policy π.

$$Q_l^\pi(s_t, a_t^l|\theta^l) = r_t^l + \mathbb{E}_{(S_{t+1},a_{t+1}^l)}[\Gamma_l^T Q_l^\pi(S_{t+1}, a_{t+1}^l|\theta^{l'})] (2)$$

The low-stage agent has task of compiling symptoms by taking to user simulator, which is activated by high-stage agent. The high level agent has layers of DQN and parameters of the network is shown byθ_l. The parameters keep updating in training by decreasing the mean-square error(MSE) between the Q-values of target network achieved and the Q-value of current one. That MSE is utilized as loss function of the advance policy network as shown in equation (3).

$$\begin{aligned} L(\theta^l) = \mathbb{E}[r_t^l + \Gamma_l^T \max l_{a_{t+1}} Q_l^*(S_{t+1}, a_{t+1}^l)|\theta^{l'}) \\ -Q_l^\pi(S_{t+1}, a_{t+1}^l|\theta^{l'})^2] \end{aligned} \quad (3)$$

In equation (3) first term is Q value of target network achieved and second one is Q value of present network.

C. User Simulator

The user simulator is the part of system that is responsible of communicating with agent and also contains the user aims in the data set. AT the start of every dialogue session it samples the aims randomly from training set. User aim hold two types of symptoms named as explicit and implicit symptoms. Explicit symptoms are provided to agent as initial input and with the help of that it will discover implicit symptoms while interacting with patient. During the interaction if it gets correct symptom then it will get reward as 1 , with incorrect symptom it will get reward of -1 and for an unknown symptom it will get reward of 0. Once its done collecting symptoms from patient low-stage agent activates high-stage agent and then the disease classifier for final classification of disease.

D. synthetic Dataset

On ships we have vast range of diseases that can occur, so having such big real world data set was almost impossible so we used synthetic data set available as Data/Fudan-Medical-Dialogue2.0 to show the effectiveness of HRL. In it every disease is linked with set of symptoms, not only that every single symptom has a probability for a certain disease. Now for identification process out of many symptoms in data set we choose any of explicit symptoms among those provided by the patient , that one symptom has more importance and rest of the symptoms are treated as implicit symptoms.

IV. BENCHMARK MODELS

First work is done on dialogue system which used task oriented disease diagnoses. It used one layer policy structure based DQN wich is called FLAT-DQN it has to do with choosing actions in each turn of dialogues[30]. After that there is a dialogue system for automatic medical diagnosis that communicates with patients to collect extra symptoms other than their self-reports and do automatic diagnose. It uses KR-DS that treats all diseases and all symptoms equally [33]. HDNO, a hierarchical reinforcement learning model, to improve performance and is validated on dialogue-based MultiWoz datasets [28]. HRL is a hierarchical reinforcement learning model which uses disease classifier for classification of symptoms separately [13]. GAMP, a model that integrates the generative adversarial network(GAN). Its policy was also DQN based used generator to generate action and a discriminator is there to check if its a good action taken on base of reward achieved [31]. HRL-pre-T , Its Hrl pre trained has two levels of policy just like us but one visible difference is that it trains the models separately and we train them together [11]. HRL is the model that used both real world and synthetic dataset and used in disease diagnosis [36]. KN-HRL is the enhanced model that creates the disease symptom relation matrix and do disease diagnosis based on patient's utterances [37].

V. Results and Conclusions

In order to check performance of our model we conduct experiment on same synthetic dataset.We did comparison of all models that includes Flat-DQN, KR-Ds, REFUEL, GAMP, HRL-Pre-T, KNHRL and Lastly our HRL model.Flat-DQN, KR-Ds, REFUEL performed almost similarly. Flat-DQN, KR-Ds are good models but performed best with the short dialogues. KNHRL and Lastly HRL model performed well but with less accuracy with larger dialogues. We present a comparison of all as given in Table 1.HRL(ours) used publicly available data set with the more medical knowledge in format of dialogues.It used disease symptom relation and symptom disease for training and testing both , also multiple rounds of dialogues with user simulator and patient and multiple layers of DQN which improves accuracy.

*Table*1

	Test Accuracy	Avg turns	Match rate
Flat-DQN	0.343	1.23	0.023
KR-Ds	0.357	6.24	0.388
REFUEL	0.416	4.56	0.161
GAMP	0.409	1.36	0.077
HRL-Pre-T	0.452	6.838	/
HRL	0.504	6.48	0.495
KNHRL	0.558	20.98	0.333
HRL (ours)	0.627	3.00	0.506

In future work we hope to gain more accuracy and collect some real world dataset.We think that with further more improvements this model can solve the problem of shortage of medical staff in the entire world.

Acknowledgment

This project has been partially funded by the "Programma Nazionale Ricerca, Innovazione e Competitività per la transizione verde e digitale 2021/2027 destinate all'intervento del FCS "Scoperta imprenditoriale" - Azione 1.1.4 "Ricerca collaborativa" - with the project SIAMO (Servizi Innovativi per l'Assistenza Medica a bOrdo) project number F/360124/01-02/X75.

References

[1] Qanita Bani Baker, Safa Swedat, and Kefah Aleesa. Automatic disease diagnosis system using deep q-network reinforcement learning. In *2023 14th International Conference on Information and Communication Systems (ICICS)*, pages 1–6, 2023.

[2] Mohamed Bakhouya, Roy Campbell, Antonio Coronato, Giuseppe de Pietro, and Anand Ranganathan. Introduction to special section on formal methods in pervasive computing, 2012.

[3] Marcello Cinque, Antonio Coronato, and Alessandro Testa. Dependable services for mobile health monitoring systems. *International Journal of Ambient Computing and Intelligence (IJACI)*, 4(1):1–15, 2012.

[4] Marcello Cinque, Antonio Coronato, and Alessandro Testa. A failure modes and effects analysis of mobile health monitoring systems. In *Innovations and advances in computer, information, systems sciences, and engineering*, pages 569–582. Springer, 2012.

[5] Antonio Coronato and Muddasar Naeem. A reinforcement learning based intelligent system for the healthcare treatment assistance of patients with disabilities. In *International Symposium on Pervasive Systems, Algorithms and Networks*, pages 15–28. Springer, 2019.

[6] Antonio Coronato, Muddasar Naeem, Giuseppe De Pietro, and Giovanni Paragliola. Reinforcement learning for intelligent healthcare applications: A survey. *Artificial Intelligence in Medicine*, 109:101964, 2020.

[7] Antonio Coronato and Giovanni Paragliola. A structured approach for the designing of safe aal applications. *Expert Systems with Applications*, 85:1–13, 2017.

[8] Jonathan S Dillard, William Maynard, and Rahul Kashyap. The epidemiology of maritime patients requiring medical evacuation: a literature review. *Cureus*, 15(11), 2023.

[9] Mario Fiorino, Muddasar Naeem, Mario Ciampi, and Antonio Coronato. Defining a metric-driven approach for learning hazardous situations. *Technologies*, 12(7):103, 2024.

[10] Mansoor Jamal, Zaib Ullah, Muddasar Naeem, Musarat Abbas, and Antonio Coronato. A hybrid multi-agent reinforcement learning approach for spectrum sharing in vehicular networks. *Future Internet*, 16(5):152, 2024.

[11] Hao-Cheng Kao, Kai-Fu Tang, and Edward Chang. Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[12] Umamah bint Khalid, Muddasar Naeem, Fabrizio Stasolla, Madiha Haider Syed, Musarat Abbas, and Antonio Coronato. Impact of ai-powered solutions in rehabilitation process: Recent improvements and future trends. *International Journal of General Medicine*, pages 943–969, 2024.

[13] Kangenbei Liao, CHENG ZHONG, Wei Chen, Qianlong Liu, Baolin Peng, Xuanjing Huang, et al. Task-oriented dialogue system for automatic disease diagnosis via hierarchical reinforcement learning. 2021.

[14] Luigia Mocerino, Fabio Murena, Franco Quaranta, and Domenico Toscano. Validation of the estimated ships' emissions through an experimental campaign in port. *Ocean Engineering*, 288:115957, 2023.

[15] Muddasar Naeem and Antonio Coronato. An ai-empowered home-infrastructure to minimize medication errors. *Journal of Sensor and Actuator Networks*, 11(1):13, 2022.

[16] Muddasar Naeem, Antonio Coronato, and Giovanni Paragliola. Adaptive treatment assisting system for patients using machine learning. In *2019 sixth international conference on social networks analysis, management and security (SNAMS)*, pages 460–465. IEEE, 2019.

[17] Muddasar Naeem, Antonio Coronato, Zaib Ullah, Sajid Bashir, and Giovanni Paragliola. Optimal user scheduling in multi antenna system using multi agent reinforcement learning. *Sensors*, 22(21):8278, 2022.

[18] Regina Padmanabhan, Nader Meskin, and Wassim M. Haddad. Learning-based control of cancer chemotherapy treatment**this publication was made possible by the gsra grant no. gsra1-1-1128-13016fromthe qatar national research fund (a member of qatar foundation). the findings achieved herein are solely the responsibility of the authors. *IFAC-PapersOnLine*, 50(1):15127–15132, 2017. 20th IFAC World Congress.

[19] Giovanni Paragliola, Antonio Coronato, Muddasar Naeem, and Giuseppe De Pietro. A reinforcement learning-based approach for the risk management of e-health environments: A case study. In *2018 14th*

international conference on signal-image technology & internet-based systems (SITIS), pages 711–716. IEEE, 2018.

[20] Yu-Shao Peng, Kai-Fu Tang, Hsuan-Tien Lin, and Edward Chang. Refuel: Exploring sparse features in deep reinforcement learning for fast disease diagnosis. *Advances in neural information processing systems*, 31, 2018.

[21] A.John Rush, Maurizio Fava, Stephen R Wisniewski, Philip W Lavori, Madhukar H Trivedi, Harold A Sackeim, Michael E Thase, Andrew A Nierenberg, Frederic M Quitkin, T.Michael Kashner, David J Kupfer, Jerrold F Rosenbaum, Jonathan Alpert, Jonathan W Stewart, Patrick J McGrath, Melanie M Biggs, Kathy Shores-Wilson, Barry D Lebowitz, Louise Ritz, George Niederehe, and for the STAR*D Investigators Group. Sequenced treatment alternatives to relieve depression (star*d): rationale and design. *Controlled Clinical Trials*, 25(1):119–142, 2004.

[22] Syed Ihtesham Hussain Shah, Antonio Coronato, Muddasar Naeem, and Giuseppe De Pietro. Learning and assessing optimal dynamic treatment regimes through cooperative imitation learning. *IEEE Access*, 10:78148–78158, 2022.

[23] Syed Ihtesham Hussain Shah, Muddasar Naeem, Giovanni Paragliola, Antonio Coronato, and Mykola Pechenizkiy. An ai-empowered infrastructure for risk prevention during medical examination. *Expert Systems with Applications*, 225:120048, 2023.

[24] Beata Sokołowska, Wiktor Świderski, Edyta Smolis-Bąk, Ewa Sokołowska, and Teresa Sadura-Sieklucka. A machine learning approach to evaluate the impact of virtual balance/cognitive training on fall risk in older women. *Frontiers in Computational Neuroscience*, 18:1390208, 2024.

[25] Milene Santos Teixeira, Vinícius Maran, and Mauro Dragoni. The interplay of a conversational ontology and ai planning for health dialogue management. In *Proceedings of the 36th annual ACM symposium on applied computing*, pages 611–619, 2021.

[26] Edvard Tijan, Marija Jović, Saša Aksentijević, and Andreja Pucihar. Digital transformation in the maritime transport sector. *Technological Forecasting and Social Change*, 170:120879, 2021.

[27] Abhisek Tiwari, Tulika Saha, Sriparna Saha, Shubhashis Sengupta, Anutosh Maitra, Roshni Ramnani, and Pushpak Bhattacharyya. Multimodal dialogue policy learning for dynamic and co-operative goal setting. In *2021 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2021.

[28] Jianhong Wang, Yuan Zhang, Tae-Kyun Kim, and Yunjie Gu. Modelling hierarchical structure between dialogue policy and natural language generator with option framework for task-oriented dialogue system. *arXiv preprint arXiv:2006.06814*, 2020.

[29] Xin Wang, Wenhu Chen, Jiawei Wu, Yuan-Fang Wang, and William Yang Wang. Video captioning via hierarchical reinforcement learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4213–4222, 2018.

[30] Zhongyu Wei, Qianlong Liu, Baolin Peng, Huaixiao Tou, Ting Chen, Xuanjing Huang, Kam-fai Wong, and Xiangying Dai. Task-oriented dialogue system for automatic diagnosis. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 201–207, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[31] Yuan Xia, Jingbo Zhou, Zhenhui Shi, Chao Lu, and Haifeng Huang. Generative adversarial regularized mutual information policy gradient framework for automatic diagnosis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01):1062–1069, Apr. 2020.

[32] Lin Xu, Lin Xu, Qixian Zhou, Qixian Zhou, , Ke Gong, Xiaodan Liang, Xiaodan Liang, Jianheng Tang, Jianheng Tang, Jianheng Tang, Lin Li, and Liang Lin. End-to-end knowledge-routed relational dialogue system for automatic diagnosis. *null*, 2019.

[33] Lin Xu, Qixian Zhou, Ke Gong, Xiaodan Liang, Jianheng Tang, and Liang Lin. End-to-end knowledge-routed relational dialogue system for automatic diagnosis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 7346–7353, 2019.

[34] Ran Yan, Dong Yang, Tianyu Wang, Haoyu Mo, and Shuaian Wang. Improving ship energy efficiency: Models, methods, and applications. *Applied Energy*, 368:123132, 2024.

[35] Qian Zhang, Tianhao Li, Dengfeng Li, and Wei Lu. A goal-oriented reinforcement learning for optimal drug dosage control. *Annals of Operations Research*, pages 1–21, 2024.

[36] Cheng Zhong, Kangenbei Liao, Wei Chen, Qianlong Liu, Baolin Peng, Xuanjing Huang, Jiajie Peng, and Zhongyu Wei. Hierarchical reinforcement learning for automatic disease diagnosis. *Bioinformatics*, 38(16):3995–4001, 07 2022.

[37] Ying Zhu, Yameng Li, Yuan Cui, Tianbao Zhang, Daling Wang, Yifei Zhang, and Shi Feng. A knowledge-enhanced hierarchical reinforcement learning-based dialogue system for automatic disease diagnosis. *Electronics*, 12(24), 2023.

[38] Qijie Zou, Xiling Zhao, Bing Gao, Shuang Chen, Zhiguo Liu, and Zhejie Zhang. Relabeling and policy distillation of hierarchical reinforcement learning. *International Journal of Machine Learning and Cybernetics*, pages 1–17, 2024.