Recommender systems enhancement using deep reinforcement learning

Berdiyev Shuhrat
School of Computing
Asia Pacific University of echnology and Innovation (APU)
Kuala Lumpur, Malaysia
tp058898@mail.apu.edu.my

Chandra Reka Ramachandran
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
Chandra.reka@staffmail.apu.edu.my

Zailan Arabee Abdul Salam
School of Computing
Asia Pacific University of Technology and Innovation (APU)
Kuala Lumpur, Malaysia
zailan@apu.edu.my

***Abstract*— Recommendation systems provide users with personalized recommendations, stand out as systems aiming to provide appropriate and efficient services. Traditional recommendation systems provide suggestions to their users with static approaches and do not include user preferences that change over time in suggestion strategies. In this study, a comprehensive review and comparison of recommendation systems that can adaptively develop suggestion approaches according to changing user preferences and learn user preferences is presented. In this research other approaches and solutions with frameworks of Deep Reinforcement Learning will be compared. In all articles that is reviewed Markov Decision Process (MDP) were used as a solution for dynamic recommendations and long-term rewards for the users. Proposed frameworks like DEERS, DRR etc. will be reviewed.**

Keywords*— Deep Reinforcement Learning (DRL), recommendation systems, Markov Decision Process (MDP), DEERS, DRR.

I. Introduction

Recommendation Systems are in demand in online world, due to its possibility to keep users' interest and/or improve sales by identifying users interest providing them opportunity to buy it. Various techniques like content-based filtering, collaborative filtering, hybrid-based, and model-based filtering [1-3] are exists to build recommendation system.

In general, steps of the recommendation system are considered static, mainly because models are trained with the dataset collected by user logs. This trained model considers the wishes of the users so far but does not considers the possibility that the preferences of the user base may change over time. Once trained model is implemented it will not interact with user, and so it will recommend things according to the old dataset. Therefore, two main problems can be identified. 1) Consideration of recommendation as static process. 2) Paying attention on recommended item clicked or not clicked and not paying attention to long-term contribution. This research will look for solution to enhance recommender system by considering two limitation that mentioned in the above paragraph. Reinforcement Learning (RL) has great potential to tackle these limitations [4] in the recommender systems. There are several frameworks proposed using DRL where environment and agent interaction will be used where environment is users with their wishes and agent is recommendation system itself.

II. DIFFERENT APPROACHES

A. Non RL based recommendaions

Many approaches like content-based recommendation, collaborative filtering, hybrid-based approach, and models-based recommendation systems that are mentioned by [1-3] are static approach or uses time as a feature [4].

Movie recommendation system have been done by [1]. In order to provide users (who have films previously rated) with acceptable recommendations, author integrated collaborative philter techniques with the neural networks with a User-User matrix. Recommendation systems have become a major part of the lives of all. With the immense number of films released across the globe each year the lack of a correct recommendation also makes people miss out on some wonderful arts work. To make the right recommendations it is important to bring machine learning based recommendation systems into action. It illustrated that content-based recommendation systems which may not seem very successful on their own but can solve the cold start issues which collaborative filtering methods face if operating independently in combination with collaborative techniques. Author therefore conclude that it is necessary to consider different approaches to the recommendation engine in order to develop a hybrid engine which overcomes and multiplies the shortfalls in these independent approaches. Where there are limitations when independent approaches to a film recommendation framework are merged, they allow users to make the correct film recommendations.

Similar works can be seen but all of them are using static approaches, so possibility of user preference change in the future are not considered.

B. RL based recommendations

[5] explores probabilistic approaches to decision-making, e.g., Strategies Bayesian and Boltzmann, along with different deterministic exploration policies, such as greedy and credit approaches and random approaches. It discusses also how the importance of exploration in profound enhancement learning will enable scientists and technology to develop their exploration strategies. Better analysis can allow the agent to make better choices and deal with the environment optimally. Reinforcement Learning (RL) agent needs to take decisions based on experience, based on its past experience which has proven ideal for the achievement of rewards, to improve the education, balance exploration and exploitation. Bayesian Dropout is contrasted with different

approaches to exploration. In the OpenAI gym setting - CartPole - all of these exploration methods are contrasted and deployed with Deep Q-Network. Bayesian dropout was better than all other methods of discovery.

[6] suggested that higher-frequency operating strengthening algorithms are useful, by analyzing general rules in the field of optimal control. Then established the relationship between small action differences, estimated Q values errors and declining output in increasing frequencies in typical enhancement learning algorithms. Author named it the reinforcement dilemma that disappears. He has explored two state-of-the-art approaches to solve the problem. The first was the use of agents upgraded to an optimal-preserving and action-increasing operator, by the vantage learning operator. In the second method a dual network architecture was used to provide deeper enhancement learning agents, generating robust action value estimates. Also identified new parameters that are influenced by control theory to assess agents' output. The results of this study illustrate the fact that the mixture of agents and concepts, older and simpler, can lead to successful performers.

[7] proposed a recommending system using deep reinforcement learning and achieved higher accuracy in the recommendation than current methods. Moreover, by checking each choice adequately and receiving incentives (i.e. feedback), they declared a bandit algorithm advances learning. When the number of things to be learned is increasing regularly it is virtually impossible to learn everything. For a new service user, the problem of collecting adequate data is the same as the problem with collaborative filtration. Authors propose a system of guidelines focused on profound strengthening training to address this issue. A neural multi-layer network updates the value function in deep reinforcement learning. They were able to make suggestions with a view to the state transition by improving learning. The following contributions have also been made. Firstly, they could generate stores suggested as vectors for the distributed representation of the N dimension by depth strengthening learning. Therefore, relative to explicitly suggested bandit algorithms, high Mean Reciprocal Rank (MRR)s and recall values could be learned efficiently. Secondly, on the basis of the similarity of the recommendation to the clicked store vector in this system, authors were able to reward themselves even if they did click on a store other than the recommended store. It could therefore also be studied offline. Third, shop vectors were generated from experiments, so it could also be suggested for newly listed shops that Latent Dirichlet Allocation LDA vectors could be made.

[8] are proposing a new structure for news recommendation on Deep Reinforcement Learning. The diverse nature of news features and user expectations means that online personalized news recommendation is a major challenge. Therefore, they propose a Deep Q-Learning-based recommendation system that can specifically model potential rewards in order to overcome the above described obstacles. In order to find enticing news for consumers, a successful discovery approach is also implemented. Extensive tests on the offline data set and online production environment for a commercial news application are performed and the superior performance of methods has been demonstrated.

Usually tailored film recommendation algorithms often support a static view of the recommendation process and only take account of current rewards. Thus, the dynamic transition between users and objects is hard to adjust. [9] suggest a film recommendation framework based on Deep Reinforcement Learning to better adapt the complex property to changing the distribution or interest for users. First to establish baseline, they take the natural DQN algorithm. Secondly, Double DQN is used to overcome overestimation and in effect minimize error within the paradigm of nature DQN. Furthermore, authors use stratified sampling to accelerate convergence, not random sampling. In the end, experimental results show by testing on the MovieLens dataset that this algorithm is superior to conventional algorithms. Experiments have shown that the recommendation accuracy can be enhanced by this process and that this algorithm is close to recent algorithms.

III. PROPOSED FRAMEWORKS

Several frameworks are introduced using Deep Reinforcement Learning. Dynamic recommendation and long-term reward are considered in those frameworks.

Recommend systems allow consumers to orient themselves in the wide variety of products, resources, and events. In a set of suggestions and user reviews, a user communicates with the recommender. The principle that previous interactions impact later ones, and the value of the interaction series can be formed and resolved by improved learning using Markov's Decision Processes (MDP).

A. *MDP Based*

[4] proposed an environment with a single interface that will allow to use the same sequence data to compare different methods of recommendation and different algorithms. Also carried out a comprehensive parameter analysis on the well-known MovieLens dataset for Deep Deterministic Policy Gradient (DDPG) methods. This is a very exciting and vast path for science. Next are the key benefits of adding RL to the recommendation process:

- Takes the process of recommendations for long term gain as complex and optimal.
- Shapesthe impact of the user state recommendations.
- The MDP formalism is versatile and, in this sense, various situations can be easily modelled.

Basically, MDP formalism and DDPG adoption will be considered in this approach.

$$\max_{\theta} \mathbb{E}[Q(s, a_{\theta}(s))]$$

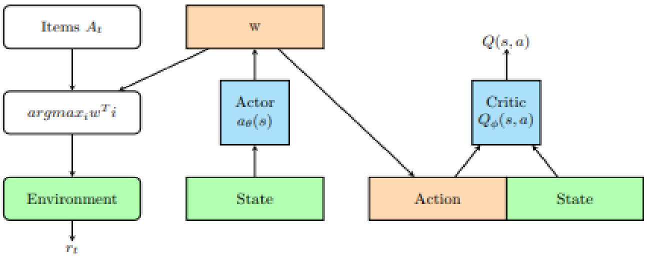

Fig. 1. DDPG [4]

B. *DRR*

According to [10] a profound DRR system for reinforcement learning to perform the recommendation work. The DRR paradigm considers the recommendation a series of decisions and adopts a "Actor-Critical" learning program to model the user-recommending systems interactions that can be complex adaptation as well as long-term awards. A status monitor module is further integrated into DRR, recording interactions of objects with users directly. Three systems are designed for instantiation. Extensive studies are performed in four real-world datasets both offline and online. The findings show that the DRR approach proposed exceeds cutting-edge competitors.

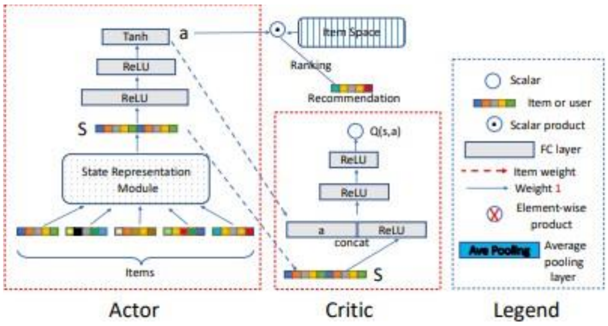

Fig. 2. DRR Framework [10]

DRR Framework consist of three part:

- The Actor Network
- The Critic Network
- The State Representation Module

C. Self-Supervised Q-Learning (SQN) and Self-Supervised Actor-Critic (SAC)

A variety of variables including long-term user participation, multiple user experiences like clicks, transactions etc. must be considered during session or sequential recommendations. The latest state-of-the-art controlled methods cannot accurately model them. A promising way is to cast sequence recommendation as a reinforcement learning (RL) problem. The preparation of the agent by encounters with the environment is an essential part of RL approaches. [11] suggests sequential recommendation self-supervised reinforcement learning. He improves standard guidance models with two performance layers: one for self-monitored learning and one for RL. The RL component serves as a regulator to direct the controlled layer to unique incentives (for example, suggesting products that can lead to purchasing rather than clicking) while the auto- controlled layer with a cross-entropy loss provides high gradient signals for parameter updates. Author proposed two frameworks based on this approach, namely Self-controlled Q-learning (SQN) and Self-Supervised Actor-Critical (SAC). Also combined four state of the art recommendation models with the proposed frameworks. The efficiency of our method is demonstrated by experiments on two real-world data sets.

D. Introspection Framework

A significant number of explanatory literatures on artificial intelligence (XAI) arises from feature relevance techniques to describe an in-depth neural network performance (DNN). However, it was not thoroughly studied to determine how XAI methods would help explain models beyond classification tasks, for example reinforcement training (RL). [12] reviewed recent work towards Explanatory Reinforcement Learning (XRL), an area which with different audiences in general, needs legal, responsible, and credible algorithms to use as a subfield of Explanatory Artificial Intelligence. When it is important to justify and clarify the actions of the agent in critical circumstances, enhanced clarification, and understandability of RL models can help to gain empirical insight into the internal functioning of the still black box. Author assessed the studies which relate explanation directly to RL. These studies are divided into two categories, as explained: transparent algorithms and post-hoc explanation. Also explored the most popular XAI works on how the latest trends in RL could theoretically explain the challenging current and future daily problems.

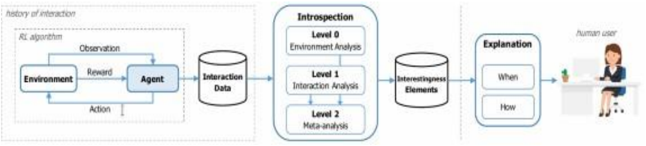

Fig. 3. Introspection Framework [12]

E. Common Framework of DRL

A review on current literature and an analysis of different issues is under way to discuss various studies: applications, methods used, researchers' difficulties and obstacle reduction recommendations. [13] in this thesis applies to all related studies about the manipulation and solutions of objects based on learning reinforcement. The problem of object grasp is a big problem of manipulation. Object grasping includes detection systems, methods, and tools to allow effective and easy training of agents. Several studies have suggested the key elements of the world and agent are the object grasping and its subtypes. In comparison to other review papers, there are different observations on deep learning-based manipulation in this review article. For researchers and practices, the outcomes of this thorough analysis of deep reinforcement training in the field of manipulation may be useful since they are able to speed up critical guidance. The implementation of a robotic Deep-RL approach has significantly enhanced the environmental and agent's learning functionality and the core elements of the RL concept. Due to the growing demands on the fulfilment of complex tasks, deep learning was combined with RL to transform robotic learning using the most useful features in different robotic applications for robotic tasks.

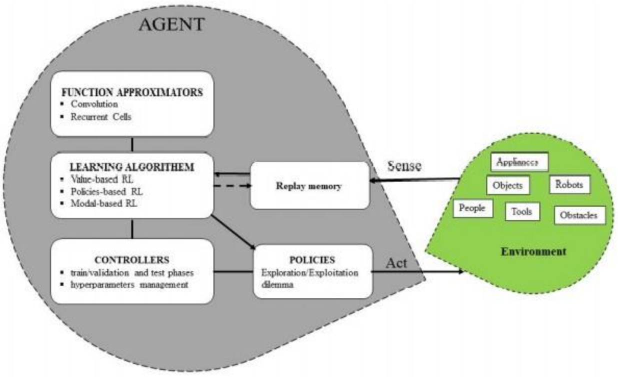

Fig. 4. Common Framework of DRL [13]

Fig 4. demonstrates common work principle of Deep Reinforcement Learning and its overall process.

IV. NEGATIV OR TRUST BASED RECOMMENDATION

Recommended systems play an important role in alleviating the overload issue by proposing custom products or services to consumers. The overwhelming majority of conventional recommendation systems regard the process as a static procedure and make recommendations based on a set strategy.

According to [14] new recommendation framework that can constantly improve its strategies through user experiences. To automatically learn the optimal strategies through the recommendation of test and error items and reinforcements from feedback obtained from users, authors model sequential interactions between users and a recommendation system as a Markov Decision Process (MDP) and exploit reinforcement learning (RL). Feedback from users can be positive and negative, and it can increase suggestions from all feedback forms. However, there is much more negative feedback than a positive feedback, and it is difficult to implement it at the same time because positive feedback could be buried by negative feedback. In this research, authors establish a new method for incorporating them into the proposed DEERS system. The experimental findings based on real-world data on e-commerce indicate the efficiency of the system proposed. More details such as dwell time can be saved in the user experience log and used within our system to obtain stronger negative feedback. Its' work principal is shown in the Fig 5.

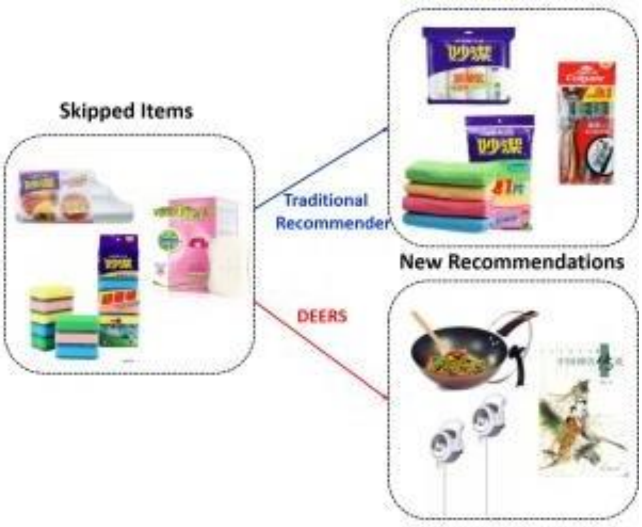

Fig. 5. Impact of negative feedback on recommendations [14]

A trust-based recommendation is an important application for a Social Network for human-computer interaction. But previous studies typically assume that the confidence value among users is permanent, unable to respond quickly to the complex changes in user confidence and preference. In fact, there is a difference between real assessment and expected assessment, which align with trust values, after receiving the recommendation. [15] a confidence boost approach through enhancement learning focused upon the dynamics of trust and the evolving mechanism of trusts between users. The Recursive Lowest Squares (RLS) algorithm is used to consider the complex effect on user confidence of assessment differences. In addition, a Deep Q-Learning (DQN) reinforcement approach is being studied to simulate the learning process of user expectations and to increase trust. Experiments suggest that authors may easily respond to the changes in user expectations using their approach to recommending systems. This process is more reliable on advice compared to other methods.

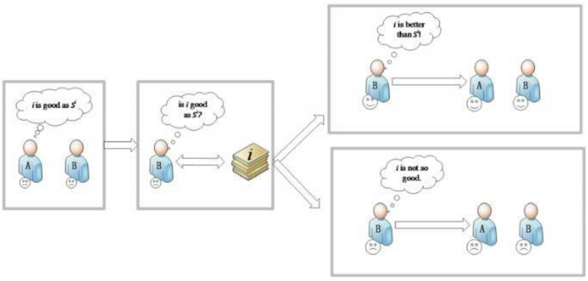

Fig. 6. Trust Based Recommendation [15]

Fig 6. shown that user A recommends to user B within interest area, and once B accepts recommendation it checks value with own expected value, so depending on the result user B trust to user A or not.

V. AREAS OF RECOMMENDATION SYSTEMS

A good place to stay is one of the first things to do while planning the tour. Reserving a hotel online can be a daunting job for any destination, with thousands of hotels to choose from. We agreed to work on the task of recommending hotels to users, since these circumstances were significant. We used the Expedia data collection, which includes several features that help us understand the process by which a user can select certain hotels over others. The aim of this hotel recommendation work consists of predicting and recommending that a consumer is more likely to reserve five hotel clusters with hundreds of different clusters.

[16] categorized the problem as a classification problem for several groups and maximized the likelihood of each cluster and finally picked up the top five clusters. Of the 38 million data points author has provided, the most critical and challenging aspect of implementing the solutions was to construct and extract significant features. Because of the size of the data, data exploration took a long time to extract features which appear to have a great impact on hotel cluster predictions. After using many models and technologies, it concluded that the best results for Ensemble Learning with Data Leak models are 0:496. This again demonstrates the synergistic effect of combining many weak learners. Most of methods involved the classification of clusters according to their expected class probabilities, which appear rational.

[17] discuss five evaluation elements of the related papers. There has been a coding scheme which includes: the e- learning system metrics for the evaluating algorithms, the recommendation filtering technology, the recommendation process phases, and the system's learning outcomes. Research shows that most e-learning systems are mainly adaptive, and precision is a key assessment indicator for recommendation algorithms. The most popular recommendation filtering technology is hybrid filtering in existing e-learning recommendation systems. Most studies recognize that the data collection phase is a significant

process. Lastly, two main indicators will achieve the learning effects of the recommender system: affections and correlations. An important part of the process of recommendations including explicit input and implicit feedback is the information gathering phase. The explicit feedback indicates some information, including user expectations, that can be explicitly accessed while the implicit feedback implicitly predicts the user's interest in the item based on user behavior. The system will reliably advise the user on desired details or content with these reviews.

Today, almost every product or piece of knowledge can be found on the Internet. The high volume of information returned by an internet query includes philters that can validate and rank the options. Recommendation systems (RSs) are a software tool to qualify and recommend solutions available to meet the needs and desires of the user.

[18] explores some relevant RSS applications in different fields such as video, music, e-commerce websites, news, etc. It also explores different methods of filtering, for example collaborative, content-based and hybrid. Recommender Framework extracts valuable data from the vast amount of knowledge provided by consumers according to their desires and wishes. RS follows three main stages for customer recommendations such as knowledge collection process, learning phase, prediction and recommendation phase in order to recommend goods or objects.

[3] identified that in the ecosystem of information and e-commerce, recommendation systems constitute a significant component. They are a powerful way to allow users to filter wide areas of information. Almost years of collaborative filtering research has led to a variety of algorithms and a traditional set of tools and software to test their performance. Field research works towards a deeper understanding of how suggested technology can be embedded in some fields. The variety of personalities expressed by numerous recommending algorithms shows that the suggestion is not a single problem. Recommenders face unique problems in domain specific tasks, knowledge needs and item domains, and a design and assessment of recommendation must be carried out based on the tasks of the user to be assisted. Successful implementations should start with an appropriate overview and targets of potential users. This analysis is the basis for the device creator to select an algorithm and to incorporate it into the user interface. This article explains what author did on the course project, including the working information and strategies of our video recommender. A video recommender system will be built in this project. The basic functions provide video suggestions for individual users. Additional enhancements such as testing and assessment on various datasets are also made. In addition, the framework is coupled with advanced algorithms, such as collaborative filtering, coefficient of correlation and checking.

VI. CONCLUTION

This research reviewed various techniques for recommender systems and where this system is used. Focus is to solve the limitations of existing techniques. Different type of frameworks and approaches was introduced which are deal with static process of previous systems.

Frameworks like DRR and DEERS taking account immediate and long-term rewards which satisfy the limitation of user preferences change in over time. So, dynamic process of recommendation system can be achieved by applying DRL. Beside that DEERS framework was considering negative feedback of users to get accurate result in recommendation. Also trust-based recommendation approach was reviewed, and this system was built upon users trust which calculated with value of expectation on the recommended item.

REFERENCES

[1] P. Praveen, P Goud and S. Parmar, “Movie Recommendation System Approaches”, International Journal of Engineering Research & Technology (IJERT) vol. 9, pp. 870-872, April, 2020.

[2] Y. Kumar and N. Kumari, “Movie Recommendation System”, Journal of Interdisciplinary Cycle Research vol. XII, pp. 847- 85, September, 2020.

[3] P. Shah and S. Sanghvi, “Video Recommender System”, 01 June, 2020, Available: https://www.researchgate.net/publication/342425128_Video_Recom mender_System, [Accessed Nov. 01, 2020].

[4] Anton Dorozhko, “Reinforcement Learning for Long-term Reward Optimization in Recommender Systems,” Available at: https://dorozhko-anton.github.io/assets/trackrecords/msc/rl_recsys_draft_dorozhko_ant on.pdf , 2019, [Accessed: Nov. 01, 2020].

[5] A. J. Rehman and P. Tomar, “Decision-Making in Reinforcement Learning”, 01 June, 2019, Available: https://www.researchgate.net/publication/333530541_Decision-Making_in_Reinforcement_Learning, [Accessed Nov. 01, 2020].

[6] S. Ravi, “Reinforcement Lerning Across Timescales”, Master of Science Thesis, Maritime and Materials Eng., Delft University of Technology, 11 August, 2017, [Accessed: Nov. 01, 2020].

[7] I. Munemasa and Y. Tomomatsu,”Deep Reinforcement Learning for Recommender Systems”, 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 6-7 March, 2018, pp. 226-233. Available: https://ieeexplore.ieee.org/abstract/document/8350761, [Accessed: Nov. 01, 2020].

[8] G. Zheng, F. Zhang, Z. Zheng, Y. Xiang, N. J. Yuan, X. Xie and Z. Li, “DRN: A Deep Reinforcement Learning Framework for News Recommendation”, WWW '18: Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23-27 April, 2018, pp. 167-176, Available: https://dl.acm.org/doi/abs/10.1145/3178876.3185994, [Accessed: Nov. 03, 2020].

[9] Z. Zhao and X. Chen, “Deep Reinforcement Learning based Recommended System using Stratified Sampling”, IOP Conf. Ser.: Mater. Sci. Eng. Vol. 466, Nanjing, China, 17-19 August, 2018, Available: https://iopscience.iop.org/article/10.1088/1757-899X/466/1/012110/meta, [Accessed: Nov. 03, 2020].

[10] F. Liu, R. Tang, X. Li, W. Zhang, Y. Ye, H. Chen, and Y. Zhang, “Deep Reinforcement Learning based Recommendation with Explicit User- Item Interaction Modeling”, arXiv:1810.12027v3 [cs.IR], 29 October, 2019, Available: https://arxiv.org/abs/1810.12027, [Accessed Nov. 03, 2020].

[11] X. Xin, A. Karatzoglou, I. Arapakis and J. M. Jose, “Self-Supervised Reinforcement Learning for Recommender Systems”, 01 June, 2020, Available: https://www.researchgate.net/publication/342093511_Self-Supervised_Reinforcement_Learning_for_Recommender_Systems, [Accessed Nov. 06, 2020].

[12] A. Heuillet, F. Couthouis and A. N. Rodriguez, “Explainability in Deep Reinforcement Learning”, 01 August, 2020, Available: https://www.researchgate.net/publication/343711800_Explainability_in_Deep_Reinforcement_Learning, [Accessed Nov. 06, 2020].

[13] M. Q. Mohammed, K. L. Chung and C. S. Chyi, “Review of Deep Reinforcement Learning-Based Object Grasping: Techniques, Open Challenges, and Recommendations”, IEEE access vol. 8, pp. 178450- 178481, 9 October, 2020,

Available: https://ieeexplore.ieee.org/abstract/document/9210095, [Accessed Nov. 07, 2020].

[14] X. Zhao, L. Xia, L. Zhang, J. Tang, Z. Ding and D. Yin, "Recommendation with Negative Feedback via Pairwise Deep Reinforcement Learning," KDD'18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Minin, London, United Kingdom, August 19-23, 2018, ACM, New York, NY, USA, pp. 1040-1048. https://doi.org/ 10.1145/3219819.3219886.

[15] F. Qi, X. Tong, L. Yu and Y. Wang, "Personalized Project Recommendations: using Reinforcement Learning", EURASIP Journal on Wireless Communications and Networking, vol. 280, December com.ezproxy.apiit.edu.my/article/10.1186/s13638-019-1619-6,[Accessed: Nov. 07, 2020].