International Journal of Innovative Research in Engineering
Volume 4, Issue 3(May-June 2023), PP: 115-118.
www.theijire.com **ISSN No: 2582-8746**

Optimization of Ads Using Reinforcement Learning and Comparison of Algorithms

Sharan Nishanth.M[1], Giridharadhayalan.M[2], Karthi Raja.S[3], Yuvaraj.E[4]
[2]Asst. Professor, Department of Mechatronics, Dr. Mahalingam College of Engineering and Technology, Pollachi, Tamilnadu, India.
[1,3,4] Department of Mechatronics, Dr. Mahalingam College of Engineering and Technology, Pollachi, Tamilnadu, India.

How to cite this paper:
Sharan Nishanth.M[1], M.Giridharadhayalan[2], Karthi Raja.S[3], Yuvaraj.E[4], "Optimization of Ads Using Reinforcement Learning and Comparison of Algorithms", IJIRE-V4I03-115-118.

https://www.doi.org/10.59256/ijire.2023040362

Abstract: *Ads optimization is the process of maximizing the effectiveness and profitability of advertising campaigns by improving targeting, messaging, and delivery strategies. This involves using data-driven techniques to analyze user behavior, identify key performance metrics, and optimize ad campaigns to achieve specific business goals, such as increasing conversions or reducing acquisition costs. Ads optimization can be applied to various types of advertising, including search engine marketing, social media advertising, display ads, and video ads. Common techniques used in ads optimization include A/B testing, machine learning algorithms, and predictive modeling. Ads optimization has become an essential component of modern digital marketing, as it allows advertisers to achieve higher ROI and better engage with their target audience.*

Key words: *Upper Confidence Bound; Ads Optimization; Thompson Sampling; Reinforcement Learning*

I.INTRODUCTION

Reinforcement learning-based ads optimization is a new discipline in digital marketing that uses machine learning methods to boost the effectiveness of advertising campaigns. In order to improve the effectiveness and efficiency of advertising, this research aims to investigate the concepts and uses of reinforcement learning in ad optimization. The project will start by outlining the basics of reinforcement learning and some of its uses in various fields. It will then go into detail about the various methods and algorithms that may be used to optimize adverts using reinforcement learning. Additionally, the project will investigate how targeting, messaging, and bidding can all be improved using reinforcement learning. The project will contain case studies and actual examples from various industries to show the practical applications of reinforcement learning for advertisements optimization. Using machine learning techniques for ad optimization raises several ethical issues, including fairness and transparency. These issues will also be covered. Readers will have a thorough comprehension of reinforcement learning in ads optimization and the strategies employed to provide the best outcomes in digital advertising by the project's conclusion. Additionally, by using reinforcement learning, they will be given the skills and resources they need to enhance their own marketing initiatives and increase their return on investment.

II.LITERATURE SURVEY

1) Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. Machine Learning, 47(2-3), 235-256.

Reinforcement learning policies face the exploration versus exploitation dilemma, i.e., the search for a balance between exploring the environment to find profitable actions while taking the empirically best action as often as possible. A popular measure of a policy's success in addressing this dilemma is the regret, that is the loss since the globally optimal policy is not followed all the time. One of the simplest examples of the exploration/exploitation dilemma is the multi-armed bandit problem. Lai and Robbins were the first ones to show that the regret for this problem has to grow at least logarithmically in the number of plays. Since then, policies which asymptotically achieve this regret have been devised by Lai and Robbins and many others. In this work we show that the optimal logarithmic regret is also achievable uniformly over time, with simple and efficient policies, and for all reward distributions with bounded support.

2) Sutton, R. S., &Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.

When we consider the nature of learning, the notion that we acquire knowledge by interacting with our surroundings is likely the first that comes to mind. Infants don't have explicit teachers when they play, wave their arms, or gaze about, but they do have a direct sensory relationship to their surroundings. Using this relationship yields a plethora of knowledge about cause and effect, about the results of activities, and about what to do to accomplish goals.These interactions surely play a significant role in how we learn about our surroundings and ourselves throughout our lives. We are keenly aware of how our actions affect our environment, whether we are learning to drive a car or to hold a conversation, and we try to control what happens by changing our behavior. A fundamental tenet of nearly all theories of learning and intelligence is learning from interaction.

3) Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research, 4, 237-285.

The theory of reinforcement learning offers a normative explanation1, based on psychological2 and neuroscientific3 perspectives on animal behaviour, of how agents may maximise their influence over an environment. Agents, however, are faced with a challenging task: they must generate effective representations of the environment from high-dimensional sensory inputs, and they must be able to employ reinforcement learning effectively in conditions that are close to real-world complexity. to apply knowledge from previous situations to new ones, employ them. Unexpectedly, reinforcement learning, and hierarchical sensory processing systems seem to work in harmony to help humans and other animals resolve this issue.4,5, the former supported by a wealth of brain data that demonstrates striking similarities between the phasic signals released by dopaminergic neurons and temporal difference reinforcement learning algorithms3. Although reinforcement learning agents have seen some success across several domains6,7,8, their applicability has previously been restricted to domains where useful features can be manually created or to domains with fully observed, low-dimensional state.

III.SYSTEM REQUIREMENTS

The hardware and operating system requirements for this software are listed here because it is software and must run on hardware and operating systems. Any version of Windows, Linux, or Mac OS can be used, hence it is platform-neutral. Any version of Windows, Linux, or Mac OS can be used, hence it is platform-neutral. Vs code community, for this to work, you must have Python installed on your PC.

We used a variety of cutting-edge technologies for this project, each of which will be examined in this chapter along with a full explanation of why it was chosen. The project's modules and features will be project's modules and features will be explanation. Let's first examine the language utilized in this project, though. We selected Python because it is a very recent language and has a lot of capabilities like machine learning and computer vision.

IV.RESEARCH METHODOLOGY

Set the issue forth: This entails deciding on the key performance indicators (KPIs) that will be used to gauge the effectiveness of the advertising campaign as well as the business objectives and target market.

Data gathering and pre processing: This entails gathering information on user activity, ad impressions, and other pertinent factors that will be utilised to reinforce

Model selection entails picking the best reinforcement learning algorithm for the task at hand, such as Thompson sampling or Upper Confidence Bound (UCB), based on the nature

To choose the optimal advertising to present to consumers in real-time, the model must be trained using the data that has been gathered. Through trial and error, the algorithm will hone its skills while also getting feedback from the system based on how well it performs.

Implementation and testing: Following training, the model can be put into practise in a real-world setting and tested to see how it does. To evaluate the efficacy of the advertisements, this can entail performing A/B testing or other kinds of experiments.

Optimization and refinement: Based on the results of testing, the reinforcement learning algorithm may need to be refined and optimized to further improve its performance. This may involve adjusting parameters, changing the training data, or implementing other strategies to fine-tune the model.

V.IMPLEMENTATION

The thorough design is actually converted into functional code during the project's implementation phase. The phase's goal is to convert the design into the best possible solution in an appropriate language used in programming. This chapter goes over the project's implementation elements and gives details on the programming language and development environment that were used. In addition to giving a summary of the primary components of the project and their orderly progression, it also does so. Following are the duties needed for the implementation phase: Investigation of the system and its limitations. The creation of transitional methods. An assessment of the changeover procedure. Making the best decisions possible when choosing the platform. Selecting the right language for application development

VI.WORKING PRINCIPLE

The Upper Confidence Bound (UCB) and Thompson Sampling are two popular algorithms used in multi-armed bandit problems, including ads optimization. Here's a brief explanation of how each algorithm works:

Upper Confidence Bound (UCB):

The UCB algorithm is built on the notion of striking a balance between exploitation and exploration or finding the best alternative. The algorithm chooses the advertisement with the highest upper confidence bound (UCB) value by maintaining a probability distribution over.

The expected reward is added to a confidence interval that takes into consideration how many times the option has been chosen to get at the UCB value. The confidence interval makes certain that less-explored possibilities are prioritized higher, and the algorithm gradually moves towards choosing the option with the largest expected return. Classifier that makes use of eye aspect ratio. Simple Euclidean math is used to do this.

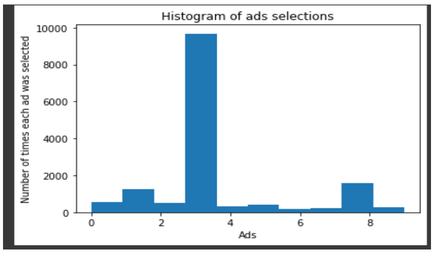
Thompson Sampling: The Thompson Sampling method similarly strikes a balance between exploration and exploitation by maintaining a probability distribution over the anticipated benefits of each option. The Thompson Sampling algorithm, in contrast to UCB, selects a sample from the probability distribution rather than relying on a confidence interval when choosing an option.

To determine the expected reward for each option, the algorithm first assigns a prior distribution to each one. The distribution is then updated considering the actual reward. The algorithm chooses the option with the highest sample value by taking samples from the posterior distributions of all the options. With this strategy, the algorithm can investigate less-tried options while giving those with a higher expected reward more weight.

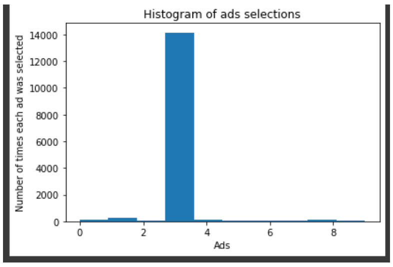
VII.COMPARISON RESULT

Since the model for facial recognition in this project has already been trained, there is no need for dataset training. The 4shape_predictor_68_face_landmarks.dat is used to find faces in frames and images. The result is dependent on the condition of the algorithm's artificial object, the driver's eyes. Depending on how the eyes are placed, the following outcome is produced.

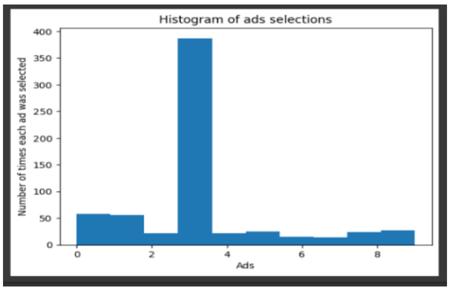
RESULTS OF COMPARISION



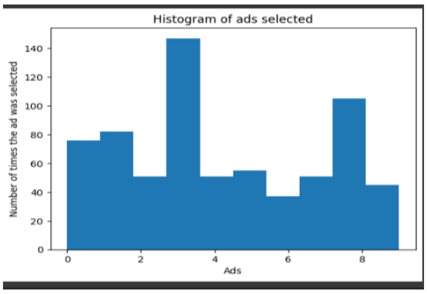
Iteration 1 of Upper Confidence Bound



Iteration of Thompson Sampling



Final Iteration of Upper Confidence Bound



Final Iteration of Thompson Sampling

VIII.TESTING METHODOLOGIES

Multi-armed Bandit Testing: Multi-armed bandit testing is a variation of A/B testing that allows for dynamic allocation of traffic to different variations. Instead of splitting traffic equally between two versions, multi-armed bandit testing allocates traffic based on the expected performance of each variation, with more traffic being directed to the variation that is expected to perform better.

Contextual Bandit Testing: Contextual bandit testing is a more advanced form of multi-armed bandit testing that takes into account the context of each user interaction. In ads optimization, contextual bandit testing can be used to dynamically select the best ad to display based on the user's behavior, interests, demographics, and other variables.

IX.CONCLUSION

In conclusion, ads optimization using reinforcement learning algorithms such as Upper Confidence Bound and Thompson Sampling can be an effective way to improve the performance of online advertising campaigns. By balancing exploration and exploitation, these algorithms enable advertisers to make informed decisions about which ads to display to users, based on the expected reward.

However, in order to successfully implement ads optimization, it is important to have a thorough understanding of the available data, the specific goals and KPIs of the campaign, and the testing methodologies that can be used to evaluate the effectiveness of different ad variations.

Moreover, while the use of reinforcement learning algorithms can improve the efficiency of ad campaigns, it is important to note that they are not a substitute for effective targeting, engaging ad creatives, and a well-designed landing page. A successful ads optimization strategy requires a holistic approach that takes into account all aspects of the campaign, from ad placement and targeting to user engagement and conversion.

Overall, with the right approach, ads optimization using reinforcement learning algorithms can help advertisers improve the performance of their ad campaigns, drive more traffic to their websites, and ultimately increase their return on investment (ROI).

X.FUTURE SCOPE

Integration with other AI technologies: Ads optimization could be further enhanced by integrating it with other AI technologies, such as natural language processing, image recognition, and predictive analytics. This could enable more targeted and personalized ads that are tailored to individual users' interests.

Multichannel ads optimization: With the proliferation of multiple online channels, such as social media, mobile apps, and websites, ads optimization could be extended to optimize across multiple channels. This would require new algorithms that can balance the performance of ads across different channels and provide a seamless user experience.

Continuous learning: Ads optimization algorithms could be further improved by incorporating continuous learning mechanisms that allow them to adapt to changing user behaviors and preferences over time. This could include the use of deep reinforcement learning techniques that can learn from past experiences and improve their decision-making over time.

References

[1]. *Kuleshov, V., &Precup, D. (2014). Algorithms for multi-armed bandit problems. Journal of Machine Learning Research, 15(1), 95-127.*

[2]. *Lattimore, T., &Szepesvári, C. (2020). Bandit algorithms (Vol. 1). Cambridge University Press*

[3]. *Chapelle, O., & Li, L. (2011). An empirical evaluation of Thompson sampling. In Advances in neural information processing systems (pp. 2249-2257).*

[4]. *May, A., &Korda, N. (2012). Thompson sampling for 1-dimensional exponential families with unknown variance. In Advances in Neural Information Processing Systems (pp. 3034-3042).*

[5]. *Russo, D. (2016). A Tutorial on Thompson Sampling. Foundations and Trends® in Machine Learning, 9(4-5), 338-403*