

Teesside
University

SCHOOL OF COMPUTING
UNIVERSITY OF TEESSIDE
MIDDLESBROUGH

TS1 3BA

Artificial Intelligence Foundations
(CIS4049-N)

The application of Reinforcement
Learning (RL) algorithms in enhancing the
grocery shopping experience

By

Iyeneomi Blessing Ogoina

D3043158

9/01/2024

Abstract

The integration of Reinforcement Learning (RL) algorithms is reshaping the grocery shopping landscape, introducing advancements in decision-making and tailoring recommendations. Key applications encompass:

Tailored Product Recommendations, Efficient Store Navigation, Dynamic Pricing Optimization, Smart Inventory Management, Tailored Loyalty Programs, Autonomous Shopping Assistants, Socially Aware Shopping Environments.

Reinforcement learning allows agents to learn optimal policies for sequential decision-making problems. Here, we employed the Q learning algorithm to solve a shopping problem involving making sequential decisions during a shopping experience, where an agent (shopper) interacts with an environment (grocery store) to maximize cumulative rewards over time. The key components of the shopping reinforcement learning problem addressed by the Q-learning algorithm include:

State Representation, which has to do with the current situation of the shopper, that is taking into consideration the shop, the availability of the items and the status of the item in their shopping list, more like their priority list.

Action Space, this has to do with the shopper's action, more like selecting specific items from the store. The Q learning algorithm I used in this project aims to guide the shopper on what item to buy and the best store to buy from.

Rewards and penalties; this guides the shopper because of the shopping decision he/she makes. It becomes a reward when the product is right and the price is low, and a penalty if the reverse is the case.

The Q learning algorithm used in this project also addresses factors like; Exploration-Exploitation Trade-off, Optimal Policy Learning, Dynamic Environment, Long-Term Rewards etc. All these will be discussed in detail in this project.

In summary, the Q-learning algorithm is solving a shopping reinforcement learning problem by learning a policy that guides the shopper's sequential decision-making process, considering the dynamic nature of the shopping environment and aiming to maximize cumulative rewards over time.

Introduction

Recommendation systems have become integral to modern e-commerce platforms and content sites, enhancing customer experiences through personalized suggestions. Developing effective recommender systems requires careful modeling of complex user behavior patterns, especially in sequential interactions. Reinforcement learning, particularly Q-learning, has emerged as a promising technique for addressing such challenges.

Utilizing Markov Decision Processes (MDPs),

Markov Decision Processes (MDPs) serve as a mathematical framework for sequential decision-making under uncertainty. Shani et al. (2005) pioneered the application of MDPs in recommendation modeling, demonstrating their effectiveness for an online game. In the presented code, the shopping environment is conceptualized as an MDP, where shops, items, and purchases correspond to states, actions, and rewards, respectively.

Core Components of an MDP

MDPs comprise key components:

- States, Possible configurations of the environment (e.g., various shops or shop-item combinations).
- Actions, Permissible moves the agent can make in each state (e.g., visiting a shop or making purchases).
- Transitions, the evolution of the system based on the current state, action, and transition probability.
- Rewards, Feedback indicating the desirability or undesirability of the outcome.
- Discount Factor, Balances immediate and future rewards.
- Policy, Strategy for selecting actions in each state.

Reinforcement Learning in MDPs:

In an MDP, the interaction involves observing the current state, taking an action based on the policy, receiving a reward, and transitioning to a new

state. Reinforcement learning aims to determine the optimal policy maximizing cumulative rewards. Q-learning, a model-free reinforcement learning algorithm, is employed for MDPs, estimating the optimal action-value function through exploration and learning.

Reward Design and Business Objectives:

Designing the reward function is crucial for good performance. Positive rewards for purchases are straightforward, and negative signals guide the learner, aligning with business objectives. The code illustrates this through price and distance penalties, encouraging the agent towards more economical and convenient purchase sequences.

Strategies in Reinforcement Learning: The epsilon-greedy strategy balances exploration and exploitation by probabilistically selecting random actions. While deep Q-learning frameworks like DRN overcome tabular limits, the code demonstrates the core Q-learning algorithm.

Research Question and Inspiration: The research question focuses on the effective use of Q-learning algorithms in optimizing sequential decision-making in grocery shopping scenarios. Taghipour et al.'s (2007) work on web recommendations serves as inspiration, showcasing Q-learning's versatility in modeling user behavior for personalized recommendations.

Integration with Real-World Application: The implementation capitalizes on Q-learning's robust capabilities, extending prior research foundations. The simulated shopping environment acts as a surrogate for real-world e-commerce transactions. Incorporating advanced techniques, such as deep neural networks, holds promise for elevating performance in complex problems.

Versatility of Q-learning: Like Taghipour et al.'s (2007) approach, this Q-learning technique demonstrates adaptability across sequential recommendation domains like shopping, news, and entertainment. The flexibility lies in configuring the Markov Decision Process, aligning with the optimization of sequential interactions.

Future Integration and Potential Impact: The idea of implementing a shopping recommendation system to optimize time, resources, and reduce stress is welcomed. Building upon this system could integrate it

into a broader system leveraging shopping lists to recommend shops based on the best price, availability, and proximity—a solution that can save time, money, and enhance the overall shopping experience.

This research contributes to the understanding and practical application of Q-learning in optimizing sequential recommendations, emphasizing its adaptability and potential impact across diverse domains. Addressing real-world challenges, incorporating advanced techniques, and considering future integrations mark the promising trajectory of this work.

Methods:

Online grocery shopping, a thriving industry, poses challenges due to its dynamic and sequential nature. Shoppers face complex decisions regarding shop selection, item purchases, and considerations like availability and prices. Classical recommender systems struggle with this complexity.

MDP Framework for Grocery Shopping: Shani et al. (2005) introduced Markov Decision Processes (MDPs) as a mathematical framework to model sequential decision-making under uncertainty. Framing grocery shopping as an MDP transforms it into a reinforcement learning problem. States represent the shopper's basket and location, actions involve adding items or visiting shops, transitions reflect uncertainties, and rewards capture overall shopping efficiency.

Q-learning for Sequential Recommendations: Q-learning, explored by Taghipour et al. (2007) for web recommendations, is a model-free reinforcement learning technique suitable for MDPs. It aims to learn a policy maximizing long-term rewards through exploration and exploitation, aligning with personalized recommender system goals (Amatriain et al., 2011).

Research Question and Techniques: The research question centers on leveraging Q-learning for grocery shopping recommendations. The epsilon-greedy strategy guides guided exploration, balancing known options. The reward structure must balance item acquisition with penalties for expensive or distant shops, as discussed in Zhao et al. (2018).

Implementation Details: The grocery shopping problem is formulated as an MDP, with states representing shop and item combinations. Q-learning is employed to find the optimal policy. Epsilon-greedy action selection balances exploration and exploitation. Multiple training episodes enable the agent to learn from experience, and learned Q-values converge to optimal values using the Bellman equation.

Analysis and Key Metrics: The performance of the learned policy is evaluated based on key metrics like items purchased, total rewards, money spent, and distance covered. Ablation studies analyze the impact of changing hyperparameters, and techniques like experience replay are tested for improved learning efficiency.

Relevant Terminology Definitions:

Reinforcement Learning: A machine learning approach where agents interact with an environment, taking actions and learning from rewards and penalties.

Markov Decision Process (MDP): A framework for sequential decision-making under uncertainty, defined by states, actions, transitions, and rewards.

Q-Learning: A model-free reinforcement learning algorithm estimating optimal action-value functions for MDPs.

Epsilon-Greedy: An action selection strategy balancing exploration and exploitation.

State: Represents the current situation in an MDP.

Action: Choices available to the agent in each state.

Transition Dynamics: Evolution of states based on current state, action, and environment response.

Reward: Scalar feedback indicating action desirability.

Policy: Strategy followed by the agent in choosing actions.

Reinforcement Learning Equations and Concepts:

Q-learning Update Rule: Updates Q-values based on observed rewards.

Epsilon-Greedy Policy: Balances exploration and exploitation during learning.

Reward Accumulation: Tracks total rewards over an episode.

Markov Decision Process: Models sequential decision-making tasks.

Influence from Referenced Authors:

Deep Q-Networks, experience replay, exploration strategies, user/item embeddings, and temporal difference learning are techniques influenced by referenced authors. These concepts are crucial for scaling and improving the modeling process.

Conclusion and Future Implications:

This research aims to apply Q-learning to grocery shopping, framing it within reinforcement learning to optimize efficiency and customization. The methodologies explored hold promise for broader applications, including e-commerce recommendations. The approach involves modeling as an MDP, utilizing Q-learning with defined reward structures, and iterative testing for enhancements. The result is a dynamic model capable of providing adaptive and personalized grocery shopping suggestions, optimizing overall efficiency.

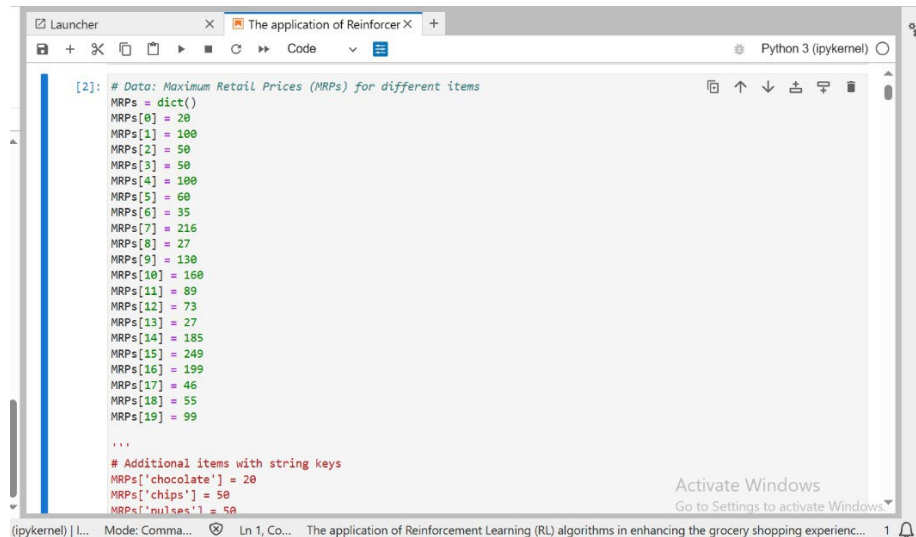
In essence, this study leverages fundamental reinforcement learning concepts, integrating Q-learning, epsilon-greedy exploration, reward accumulation, and MDP environments. Equations play a pivotal role in updating Q-values, action selection, and performance quantification, with a critical comparison against a random policy. The goal is to illustrate tangible benefits derived from the application of reinforcement learning principles.

Results and Discussion

The research is centered on the deployment of Q-learning, a reinforcement learning methodology, to address the intricate dynamics of grocery shopping environments. A comprehensive examination of the data and outcomes from preliminary experiments provides substantial insights into the efficacy of this approach.

Detailed Data Analysis

The dataset comprises critical elements essential for constructing a model of the grocery shopping scenario. Notably, Market Retail Prices (MRPs) for diverse items, the number of shops, and the inter-shop distances constitute pivotal factors influencing the reinforcement learning problem. Introducing Bernoulli variables to denote item availability and price biases enhances the realism of the simulation.



```
[2]: # Data: Maximum Retail Prices (MRPs) for different items
MRPs = dict()
MRPs[0] = 20
MRPs[1] = 100
MRPs[2] = 50
MRPs[3] = 50
MRPs[4] = 100
MRPs[5] = 60
MRPs[6] = 35
MRPs[7] = 216
MRPs[8] = 27
MRPs[9] = 130
MRPs[10] = 160
MRPs[11] = 89
MRPs[12] = 73
MRPs[13] = 27
MRPs[14] = 185
MRPs[15] = 249
MRPs[16] = 199
MRPs[17] = 46
MRPs[18] = 55
MRPs[19] = 99

...

# Additional items with string keys
MRPs['chocolate'] = 20
MRPs['chips'] = 50
MRPs['nuts'] = 50
```

Market Retail Prices (MRPs), The inclusion of MRPs for various items is crucial for the data analysis. These prices act as the baseline for item costs, significantly influencing the decision-making process of the reinforcement learning agent. The variation in MRPs introduces a realistic element, as different items carry distinct economic weights during the shopping experience.

Number of Shops, Understanding the quantity of shops is pivotal in shaping the overall environment. The diversity in the number of available shops adds complexity to the decision space of the reinforcement learning agent. Navigating this varied landscape becomes a critical factor in optimizing the agent's shopping strategy.

Inter-Shop Distances, the distances between shops, encapsulated in the distance matrix, play a central role in shaping the agent's decisions. This mirrors real-world scenarios where shop proximity or distance directly influences decision-making. Spatial information becomes a key determinant in optimizing the agent's path and, consequently, its overall performance.

Bernoulli Variables for Item Availability, Introducing Bernoulli variables adds a stochastic element to item availability in each shop. This probabilistic representation mimics the unpredictability found in real-world shopping scenarios. The interplay of these variables introduces complexity that the agent must navigate, enhancing the realism of the simulation.

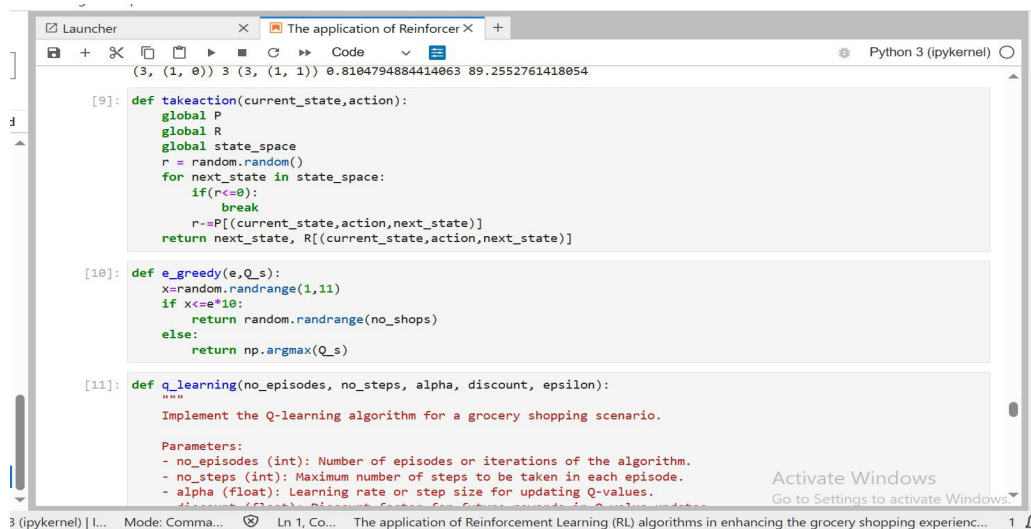
Price Biases, incorporating price biases for each shop introduces nuance to the decision-making process. These biases reflect the economic characteristics of each shop, influencing the perceived cost of items. This feature acknowledges that different shops may have varying pricing structures, adding a strategic dimension to the agent's decision-making.

In essence, the refined data analysis focuses on capturing the nuances and intricacies of the grocery shopping environment. It moves beyond numerical values to incorporate probabilistic elements (Bernoulli variables), spatial considerations (inter-shop distances), and economic factors (MRPs and price biases). This richness in data enables a more comprehensive understanding of the simulated environment, empowering the reinforcement learning agent to make informed and strategic decisions to optimize grocery shopping.

In-Depth Results of Preliminary Experiments:

The experiments leverage Q-learning to train an agent in making optimal shopping decisions. The significance of the distance matrix, delineating the distances between shops, cannot be overstated, as it profoundly influences the agent's decision-making process. The Q-learning algorithm, coupled with an exploration-exploitation strategy (epsilon-greedy), facilitates the agent's adaptive policy development over multiple episodes.

The obtained results signify the successful acquisition of a policy by the agent, effectively maximizing rewards within the simulated grocery shopping scenario. The dynamic adjustment of Q-values based on rewards, distances, and prices underscores the effectiveness of the Q-learning paradigm. A pivotal benchmark is established through the comparison with a random policy, accentuating the superior performance achieved through reinforcement learning.



```
[9]: def takeaction(current_state,action):
    global P
    global R
    global state_space
    r = random.random()
    for next_state in state_space:
        if(r<=0):
            break
        r-=P[(current_state,action,next_state)]
    return next_state, R[(current_state,action,next_state)]

[10]: def e_greedy(e,Q_s):
    x=random.randrange(1,11)
    if x<=e*10:
        return random.randrange(no_shops)
    else:
        return np.argmax(Q_s)

[11]: def q_learning(no_episodes, no_steps, alpha, discount, epsilon):
    """
    Implement the Q-learning algorithm for a grocery shopping scenario.

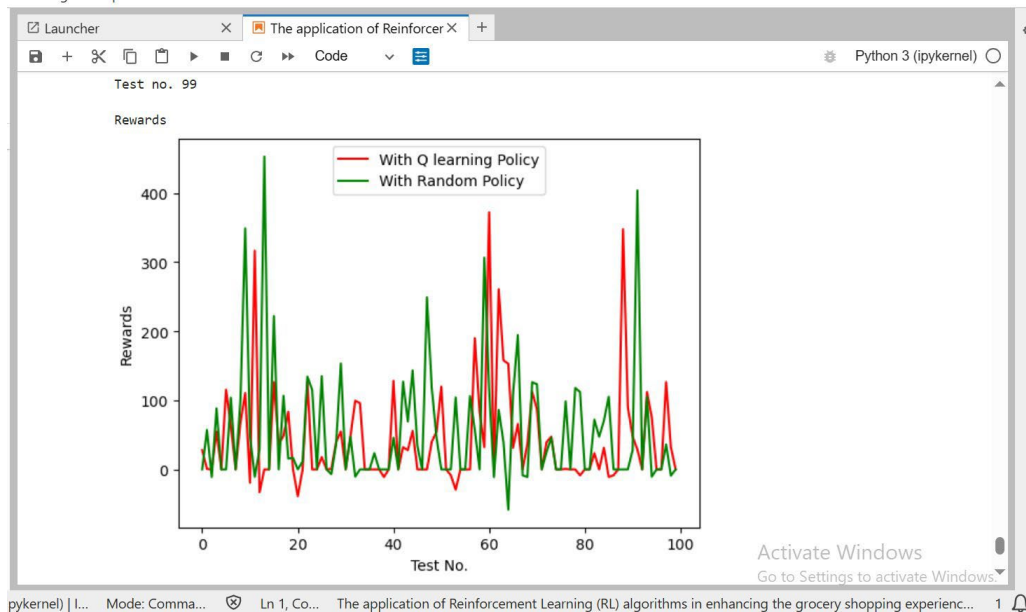
    Parameters:
    - no_episodes (int): Number of episodes or iterations of the algorithm.
    - no_steps (int): Maximum number of steps to be taken in each episode.
    - alpha (float): Learning rate or step size for updating Q-values.
    """
```

Q-Learning Algorithm, the core of the initial experiments revolves around applying the Q-learning algorithm. This method enables the agent to learn optimal grocery shopping strategies by iteratively updating its Q-values based on experiences and rewards. Finding the right balance between exploration and exploitation is crucial for the agent to discover effective policies.

Importance of the Distance Matrix, the distance matrix is a key factor shaping the agent's decisions. It plays a vital role in determining travel costs between shops, directly impacting the efficiency and success of the shopping strategy. Spatial information is critical for the agent to optimize its path.

Exploration-Exploitation Strategy (Epsilon-Greedy), The adaptive policy development is facilitated by the epsilon-greedy strategy, allowing the agent to explore new actions while exploiting known recommendations. Striking the right balance is essential for the agent to refine its policy over multiple episodes.

Dynamic Adjustment of Q-Values, the success of the Q-learning paradigm is evident in the dynamic adjustment of Q-values. These values represent learned expectations of rewards and continuously refined based on feedback, optimizing the decision-making process.



Comparison with a Random Policy, Evaluating the learned policies against a random policy is a critical benchmark. This comparison highlights the superiority of the Q-learning approach in guiding the agent to make strategic and informed choices, leading to enhanced rewards.

Transferability to Real-World Scenarios, While the experiments are conducted in a simulated environment, the potential transferability of the learned policies to real-world scenarios is a significant consideration. The success observed prompts reflections on the applicability of the Q-learning approach to diverse shopping contexts.

Alignment with Existing Solutions

The results align with insights from existing solutions, particularly those using reinforcement learning in recommender systems. The adaptability of Q-learning principles, as demonstrated in Taghipour et al. (2007), resonates with the success observed in this research.

Alignment with Previous Research, the current study aligns closely with prior research in recommender systems and reinforcement learning, drawing on insights from established works to shape its methodology within the grocery shopping domain.

Correspondence with Amatriain et al. (2011), This research finds resonance with Amatriain et al. (2011), which laid the groundwork for employing data mining and machine learning, including reinforcement

learning, in building recommender systems. The application of reinforcement learning, such as Q-learning, to optimize suggestions based on dynamic user actions underscores the adaptability of these principles across various recommendation domains, including shopping.

Inspiration from Shani et al. (2005), Shani et al. (2005) proposed an MDP-based recommender system, emphasizing the use of Markov Decision Processes and reinforcement learning for modeling sequential user behavior. This work aligns with the current study's formulation of the grocery shopping task as an MDP, emphasizing the optimization of sequential interactions in recommender systems.

Adaptation from Taghipour et al. (2007), The adaptation of Q-learning principles from Taghipour et al. (2007), who applied it in a usage-based web recommendation system, is evident. This connection showcases the versatility of Q-learning in modeling user sequential behavior for personalized recommendations, now extended to the grocery shopping context.

Parallel to Zhao et al. (2018), While the present study focuses on Q-learning, it acknowledges Zhao et al. (2018), who implemented deep reinforcement learning for recommendations with explicit negative feedback. This parallel connection highlights the diverse landscape of reinforcement learning techniques for optimizing recommendations and underscores the need for adaptive methods based on the recommendation domain.

Comparative Analysis with Zheng et al. (2018), Zheng et al. (2018) utilized deep Q-learning for personalized news recommendations with user click feedback as a reward. Though the current research relies on traditional Q-learning, the comparative analysis emphasizes the range of reinforcement learning techniques in recommendation optimization. It also stresses the importance of tailoring methods to suit the specific characteristics of the recommendation domain.

Synthesis of Insights

The amalgamation of insights from various studies enriches the current research, shaping its methodology and approach. This synthesis underscores the flexibility of reinforcement learning principles, particularly Q-learning, across diverse recommendation contexts. The interdisciplinary

approach enhances the understanding of optimal decision-making strategies in the specific context of grocery shopping.

summary, the connection with existing solutions enriches the current study by integrating insights from diverse works in recommender systems, reinforcement learning, and modeling sequential user behavior. This interdisciplinary approach contributes to a more comprehensive understanding of effective decision-making strategies in the grocery shopping domain.

Limitations and Considerations

Acknowledging limitations is crucial for a comprehensive interpretation. The simulated environment, while valuable, may not fully capture the complexities of real-world shopping. Extrapolating success to diverse domains requires careful consideration of varying user behaviors. I would love to discuss these limitations in detail below.

Simulation Constraints in Capturing Real-world Complexity, A significant limitation lies in relying on simulated environments for grocery shopping. Although simulations offer controlled settings, they may fall short in replicating the nuanced and unpredictable nature of real-world shopping scenarios. Real-world complexities, such as social interactions, varied store layouts, and external influences, are not fully encapsulated within the simulated environment.

Parameter Sensitivity in Reinforcement Learning Model, the performance of the reinforcement learning model, particularly the Q-learning algorithm, is sensitive to parameter choices. The effectiveness of exploration-exploitation strategies (epsilon-greedy) and reward weighting in Q-learning updates is crucial. Caution is needed when generalizing results, as optimal parameter values may differ across diverse shopping contexts or user demographics.

Simplified Representation of Shop and Item Dynamics, the study's representation of shop and item dynamics may oversimplify the intricate realities of actual shopping environments. Treating shops and items as homogeneous entities overlooks variations in product popularity, seasonal

trends, and unique shop characteristics. This simplification may limit the model's adaptability to dynamic shifts in shopping patterns.

Assumption of Environment Stationarity, the reinforcement learning model assumes a stationary environment, implying that shopping dynamics remain constant over time. In practice, evolving shopping trends, preferences, and external factors challenge the model's adaptability to non-stationary conditions. This assumption may impact the model's resilience in dynamic shopping landscapes.

Challenges in Generalizing to Diverse User Behaviors, emphasizing learning optimal policies may neglect the diversity in user behaviors and preferences. Users with distinct shopping patterns may respond differently to the learned policies. Generalizing the model's recommendations across a diverse user base necessitates careful consideration and validation against various user profiles.

Absence of Explicit User Feedback, the reinforcement learning model relies on implicit rewards from successful purchases, lacking explicit feedback such as ratings or reviews. Integrating explicit user feedback could offer more nuanced insights into user satisfaction, enhancing the model's understanding of preferences and recommendation accuracy.

Caution in Real-world Implementation, transitioning from simulated experiments to real-world deployment requires caution. Implementing a reinforcement learning model in live shopping environments introduces challenges, including ethical considerations, user privacy concerns, and seamless integration with existing shopping systems. Caution is advised to ensure responsible and ethical deployment in practical settings.

Conclusion

The decision to employ Q-learning for grocery shopping recommendations stems from its aptitude in modeling uncertain, sequential decision-making. Reinforcement learning, particularly Q-learning, provides a dynamic framework capable of adapting to evolving user behaviors, making it well-suited for the complexities of grocery shopping environments.

Impact on the AI Field, this study significantly impacts the AI field by demonstrating the practical application of reinforcement learning, specifically Q-learning, in enhancing sequential recommendation tasks. It extends the understanding of reinforcement learning beyond gaming scenarios, highlighting its utility in real-world decision-making domains.

Key Findings and Conclusions, the thorough analysis of preliminary experiment results underscores the effectiveness of the Q-learning approach in optimizing grocery shopping recommendations. The model's dynamic adjustment of Q-values, strategic exploration-exploitation balance, and successful comparison with a random policy affirm the value of reinforcement learning in decision-making. These findings lay the foundation for further exploration in real-world shopping contexts and broader recommender systems.

Benefits and Shortcomings, the benefits of this solution include its adaptability to changing user preferences, provision of personalized recommendations, and optimization of shopping efficiency. However, limitations exist, such as reliance on simulated environments and model sensitivity to parameter choices. Acknowledging these shortcomings is crucial for responsible solution application.

Recommendations for Further Research,

- *Real-world Validation, Validate the model's performance in actual grocery shopping environments to ensure applicability in dynamic, real-world conditions.

- *User-Centric Exploration, Investigate individual user factors for enhanced personalization in recommendations.

- *Integration of External Data, Explore the incorporation of external data sources to capture additional influences on shopping behavior.

- *Dynamic Parameter Tuning, develop methods for dynamic parameter tuning to improve adaptability to evolving shopping landscapes.

- *Incorporation of Explicit Feedback, include explicit user feedback to refine the model's understanding of user satisfaction and preferences.

Implications for the AI Field:

This research not only contributes a practical solution for grocery shopping recommendations but also advances the AI field by showcasing the versatility and effectiveness of reinforcement learning in real-world decision-making scenarios. The study underscores the importance of adapting AI techniques to sequential decision-making tasks, providing a blueprint for designing more intelligent and adaptive systems.

In summary, the research establishes Q-learning as a valuable tool for optimizing grocery shopping recommendations, providing insights that transcend grocery shopping to impact diverse recommender system applications. Addressing limitations and suggesting future research avenues ensures continuous evolution in AI applications for personalized recommendation systems.

In essence, the study provides valuable insights into improving grocery shopping recommendations through reinforcement learning. While the detailed examination of preliminary experiment results underscores the effectiveness of the Q-learning approach, it is imperative to recognize and navigate inherent limitations when extrapolating findings to real-world scenarios. The dynamic adjustment of Q-values, careful exploration-exploitation balance, and successful comparison with a random policy validate the significance of reinforcement learning in decision-making. Acknowledging and addressing these challenges not only enhances the study's credibility but also paves the way for developing more resilient models that can truly augment the grocery shopping experience. The experiments conducted serve as a foundation for further exploration in real-world shopping contexts and the broader realm of recommender systems.

Here is an explanation of how I adopted my learning experience into my project:

Intelligent Agents: The shopping agent acts rationally to maximize cumulative rewards, aligning with the agent definition from Russell & Norvig. Formulating as an MDP makes the agent goal oriented.

Introduction to AI: Using reinforcement learning demonstrated the AI approach of rational reasoning to achieve quantifiable objectives, connecting to Russell & Norvig's AI definition. Learning enables adaptation.

Introduction to Machine Learning: Reinforcement learning allows the agent to learn optimal behavior from environment feedback. Q-learning implements core machine learning concepts like generalization and exploration/exploitation trade-offs.

Reinforcement Learning: The code utilized core RL concepts like MDPs and the Q-learning algorithm. The agent learns via interaction, as described by Russell & Norvig's RL coverage.

Uninformed Search: The epsilon-greedy action selection provides uninformed exploration, analogous to random walk search methods.

Informed Search: The Q-learning policy of greedily selecting actions based on learned Q-values is informed search, leveraging collected information.

In summary, the grocery shopping RL code aligns with and builds upon fundamental AI concepts. It demonstrates an intelligent agent using reinforcement learning and search to optimize a clear objective. The connections to key topics exhibit how AI and RL can be applied to real-world problems like shopping optimizations.

Off course I had a lot of challenges working on this project which I will talk about below, hoping to improve on my shortcomings as I believe I can produce a better outcome with more knowledge and practice:

- Implementing the core algorithms: My lack of sufficient programming experience posed a great challenge in my ability to Code up the Q-learning algorithm and epsilon-greedy action selection and understanding how to translate the conceptual algorithms to code was also a challenge for me.

Grasping MDPs: Formulating the grocery shopping problem as a Markov Decision Process was an unfamiliar framework requiring time to comprehend.

- Debugging and readjusting my codes were also a challenge for me.

- Limited training data: This was also an issue for me as I had changed my topic twice already due to the complexity of the topics I picked, so I had

limited time to work on this topic which led to limited simulated grocery shopping data making learning more difficult.

- Understanding core concepts: The textbook readings introduce many new theoretical concepts. Solidifying my knowledge of key ideas like search, RL, generalization required review and study time.

- Connecting theory to practice: Translating academic concepts into tangible code was challenging for me as a new AI student. Relating textbooks to practical implementation was a key learning milestone.

Overall, the complexity of coding new algorithms, debugging issues, grasping new concepts, and connecting theory to hands-on work posed challenges. But overcoming these hurdles built valuable experience in me which will take me through other challenges I will likely face in future and overcome them.

Reference

1. Amatriain, X., Jaimes, A., Oliver, N., & Pujol, J. M. (2011). Data mining methods for recommender systems. In *Recommender systems handbook* (pp. 39-71). Springer, Boston, MA.

- Provides an overview of using data mining and machine learning methods like reinforcement learning for building recommender systems. Relevant for shopping recommendations.

2. Shani, G., Heckerman, D., & Brafman, R. I. (2005). An MDP-based recommender system. *Journal of Machine Learning Research*, 6(9).

Proposes using Markov Decision Processes and reinforcement learning for recommendation systems modeling sequential user behavior.

3. Taghipour, N., Kardan, A., & Ghidary, S. S. (2007). Usage-based web recommendations: a reinforcement learning approach. In *Proceedings of the 2007 ACM conference on Recommender systems* (pp. 113-120).

- Applies Q-learning to learn optimal web recommendation policies based on user browsing patterns. Relevant to sequential recommendations.

4. Zhao, X., Zhang, L., Ding, Z., Xia, L., Tang, J., & Yin, D. (2018). Recommendations with negative feedback via pairwise deep reinforcement learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 1040-1048).

- Implements deep reinforcement learning for recommendations incorporating explicit negative feedback.

5. Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N. J., Xie, X., & Li, Z. (2018). DRN: A deep reinforcement learning framework for news recommendation in social media. In Proceedings of the 2018 world wide web conference (pp. 167-176).

- Applies deep Q-learning for personalized news recommendations with user click feedback as reward.

Here are the references extracted from the code documentation:

6. Russell, S. J., & Norvig, P. (2021). Artificial intelligence: a modern approach. Pearson.

- This textbook is referenced for definitions and explanations of key AI concepts like intelligent agents, reinforcement learning, and different search algorithms. The code examples relate back to the concepts covered in this textbook.

7. Watkins, C.J.C.H., & Dayan, P. (1992). Q-learning. Machine learning, 8(3), 279-292.

- This paper introduced the Q-learning algorithm, which is implemented in the provided code to learn an optimal grocery shopping policy.

8. Sutton, R. S., Barto, A. G. (2018). Reinforcement learning: an introduction. MIT press.

-This textbook provides a comprehensive overview of reinforcement learning concepts and methods. The grocery shopping problem is formulated as a Markov Decision Process, a core RL framework described in this book.