Dynamic Pricing For Restaurants Using Reinforcement Learning

Aashir Khan
Northeastern University
khan.aas@northeastern.edu

Sumer Shinde
Northeastern University
shinde.su@northeastern.edu

Yash Bhuptani
Northeastern University
bhuptani.y@northeastern.edu

Nathanael Chiang
Northeastern University
chiang.na@northeastern.edu

Aayush Jaiswal
Northeastern University
jaiswal.aa@northeastern.edu

Abstract

Dynamic pricing is a strategy for optimizing revenue in industries with fluctuating demand. Despite its proven efficacy in the travel and entertainment industries, its adoption in restaurants remains limited. Using reinforcement learning agents such as Deep Q-Learning (DQN) and Proximal Policy Optimization (PPO), we demonstrate the potential of implementing adaptive pricing in a simulated restaurant environment. By modeling customer behavior and other factors, the agents adjust menu prices in real-time to maximize revenue while maintaining price stability, offering a scalable solution for potentially improving profitability in the restaurant industry.

Introduction

The restaurant industry faces significant challenges in optimizing revenue due to varying demand patterns throughout different periods of the day. Traditional pricing models rely on static menus and predetermined price points, failing to capitalize on peak hours or incentivize dining during off-peak periods. While dynamic pricing has become normalized in industries like hotels and airlines, its implementation in restaurants remains complex due to diverse menu items, inventory constraints, and varying customer price sensitivities.

Reinforcement learning (RL) offers a promising solution to this challenge. As a branch of machine learning that trains agents to make sequential decisions by interacting with an environment and maximizing a reward signal, RL is well-suited for dynamic pricing tasks. In this project, we explore two RL approaches—Deep Q-Learning (DQN) and Proximal Policy Optimization (PPO)—within a simulation framework that balances revenue optimization with price stability. By comparing these models' performance, we aim to identify their strengths and limitations across different pricing scenarios, from high-margin to low-margin menu items.

Related Work

Dynamic pricing has leveraged reinforcement learning (RL) across various markets, from online marketplaces to energy sectors. Shartsis et al. (2018) demonstrated Q-learning's effectiveness in static demand environments, while our approach extends this using Proximal Policy Optimization (PPO) for continuous action spaces suited to restaurant pricing. Asadinejad and Rahmani's (2019) work in energy pricing highlighted the importance of managing multi-dimensional states and long-term objectives, principles we adapt to our domain.

The revenue management field has successfully applied RL to airlines and hospitality, with Karaesmen et al. (2019) using policy gradient methods for inventory and pricing optimization. We adapt these principles to the unique challenges of restaurants, where decisions span multiple items with shorter time horizons. Building on Zhang and Gupta's (2020) emphasis on customer simulation, we introduce a CustomerSimulator specifically designed for restaurant scenarios, implemented within a DynamicPricingEnv inspired by OpenAI Gym's framework (Brockman et al., 2016).

Problem Statement & Methods

Our framework for implementing dynamic pricing in restaurants involved training the two RL models (DQN and PPO) within a simulated restaurant environment. The simulation was created to replicate real-world complexities, including customer demand patterns, price elasticity, and inventory constraints. Studies have shown that customers would not order from restaurants as often if they were subject to dynamic pricing (Burke, 2023). However, we wanted to test whether that matters for the restaurant, as the goal of most restaurants is to maximize profits instead of serving the maximum number of customers.

<u>Data Preprocessing and Customer Simulation</u>

To create a simulation environment that accurately models real-world restaurant dynamics, we transformed raw data into a structured format suitable for simulation and training. Our approach integrates historical purchase patterns with sophisticated customer behavior modeling to create a comprehensive framework for dynamic pricing optimization. The foundation of our simulation relies on customer segmentation based on price sensitivity, time-dependent demand modeling, and practical operational constraints.

The simulation environment segments customers into three groups (price-sensitive, moderate, and premium) based on price elasticity, calculated as the ratio of percentage change in quantity to percentage change in price:

$$\text{Price Elasticity} = \frac{\%\Delta Quantity}{\%\Delta Price}$$

Time-dependent demand is modeled using Gaussian distributions for daily peaks (breakfast, lunch, dinner), weekday multipliers for traffic variations, and a Poisson distribution for stochastic purchase modeling, expressed as:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},\ k = 0, 1, 2, \ldots$$

where λ represents expected purchases based on demand, price, time, momentum, and availability.

The system implements key operational constraints including price bounds of 50-200% of base price, 70% random purchase probability, and inventory-limited availability. Sigmoid functions govern availability and price transitions, while the model accounts for segment-specific diminishing returns and purchase history momentum effects. Operating at hourly granularity, the simulator uses historical transaction data to calibrate peak/off-peak periods, customer density patterns, purchase frequencies, and short-term demand fluctuations, creating a comprehensive framework that enables the testing of dynamic pricing strategies while maintaining realistic market behavior.

Dynamic Pricing Environment and Reinforcement Learning Approaches

We designed a custom Gymnasium environment called DynamicPricingEnv to simulate restaurant pricing scenarios. This environment was created to be able to host a wide range of operational factors that influence pricing decisions. Key features were previous price, quantity sold, profit momentum, time of day, day of the week, base price, cost, and category encoding. These variables helped the agent to make informed decisions by considering multiple areas of restaurant operations, such as demand cycles, product-specific characteristics, and temporal dynamics.

The PPO model worked in a continuous action space, which allowed the agent to make fine-tuned price adjustments. This was most effective for small price changes that most likely impacted customer perception and satisfaction. In contrast, DQN used 41 discrete price levels between ±10%, which offered a simpler, more computationally efficient approach for managing pricing strategies, particularly for lower-margin items. Comparing the flexibility of PPO with the simplicity of DQN provided valuable insights into the trade-offs between model precision and computational overhead.

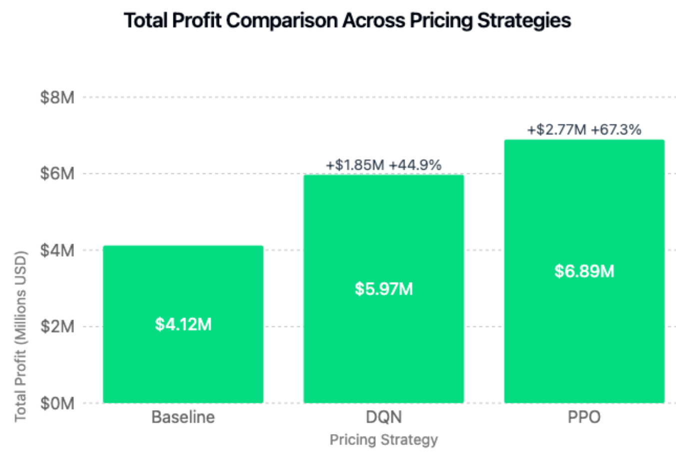
Both models used reward functions incorporating profit, price stability, profit momentum, customer elasticity, and inventory constraints. These components helped agents optimize short-term profitability while ensuring long-term stability and helping keep the current customers trustworthy of the prices. For instance, profit momentum encouraged the agent to capitalize on upward trends, while price stability discouraged excessive fluctuations that could steer away customers.

The PPO model utilized a multi-layer perceptron (MlpPolicy) with three hidden layers consisting of 512, 256, and 128 neurons. These layers, combined with ReLU activation functions, provided non-linearity to capture complex relationships within the data. Key hyperparameters included a learning rate of 3e-4, batch size of 256, and a GAE lambda of 0.95. To increase training efficiency, we used DummyVecEnv to run multiple instances of the environment in parallel, which increased the diversity of experiences available to the agent, leading to more learning. The DQN architecture was identical to PPO's, but we also added a replay buffer with a capacity of 1,000,000 state-action pairs. This stabilized the learning process by breaking correlations between consecutive experiences. This allowed the agent to repeatedly learn from impactful events, improving the overall quality of its decision-making process.

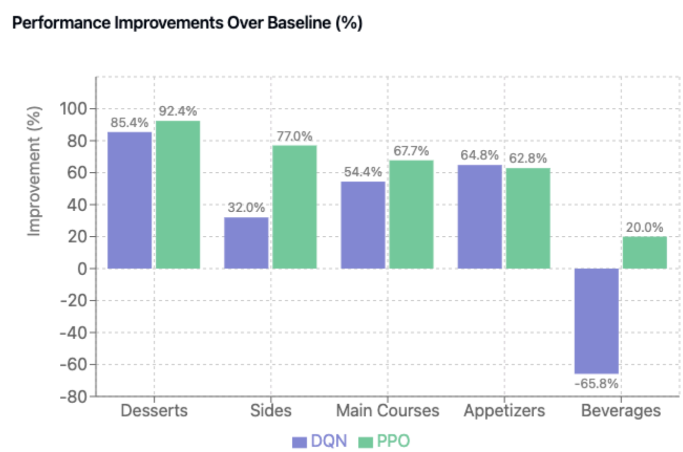
Data & Experiments

We collected the dataset from Kaggle. The dataset contains 72,998 entries (each representing a specific transaction in the restaurant) and 12 columns, including details like dates, times, order numbers, item IDs, categories, and pricing data. Key features include *Price, Cost,* and *Quantity* which are essential for reinforcement learning tasks like dynamic pricing.

Our analysis covered 68 distinct menu items across five categories, revealing significant improvements in profitability for both approaches. We ran the models on all items over a year. The baseline profit for our restaurant (no pricing changes) was $4.12M, the DQN agent profit was $5.97M, and the PPO agent profit was $6.89M. DQN achieved an average profit increase of 44.9% over baseline, while PPO demonstrated superior performance with a 67.3% increase, outperforming DQN by 22.4% on average.



The improvements varied notably across different product categories. Main Courses showed consistent improvements with both algorithms, with PPO achieving an average increase of 67.7% and DQN reaching 54.4%. Desserts demonstrated the highest percentage improvements overall, with PPO and DQN achieving 92.4% and 85.4% increases over baseline respectively. The Beverages category proved most challenging, with DQN performing significantly worse than baseline with a 65.9% decrease in profits, while PPO managed modest improvements with a 20.0% increase over baseline.



The baseline strategy maintained completely stable pricing with zero variance, as it kept prices fixed throughout. The DQN agent had a standard deviation of 1.51, indicating more aggressive and variable pricing adjustments. The PPO agent demonstrated more moderate price variations with a standard deviation of 1.01. PPO's more conservative price adjustments suggest it was able to achieve its superior profit performance (+67.3% over baseline) while maintaining 33% more stable pricing compared to DQN. This indicates PPO found a more efficient balance between maximizing revenue and maintaining price consistency.

Discussion

The PPO model demonstrated superior performance in generating higher profits and maintaining stable pricing across the board. It particularly excelled in optimizing prices for menu items with higher profit margins, such as seafood skewers, where moderate price adjustments did not significantly impact demand. Stability in pricing is critical in a restaurant setting, as frequent fluctuations can alienate customers and erode trust. PPO's ability to balance profitability with consistency makes it an excellent choice for these high-margin items.

In contrast, the DQN model showed promise in scenarios where lower-margin items, like beverages (e.g., fresh juice), were involved. Its aggressive exploratory approach allowed for more nuanced adjustments, leading to better results in these cases. Although DQN's pricing was less stable compared to PPO, its effectiveness in improving profits for low-margin items suggests that it can complement PPO when applied selectively.

Based on the analysis, a hybrid pricing strategy is recommended for the restaurant. PPO should be used for high-margin items to maintain stable profits, while DQN can be applied to low-margin items to capitalize on its strength in these scenarios. This combined approach has the potential to maximize overall profitability while addressing the unique characteristics of different menu items.

Challenges and Limitations

Our study faced several limitations worth noting. The training relied on historical data patterns without explicit modeling of external factors such as weather or events and had limited seasonal coverage. Model constraints included limiting price adjustments to ±10% per step and implementing a maximum price ratio of 2.0x base price, along with using a simplified customer behavior model. Dynamic pricing strategies, even when effective, must be implemented thoughtfully to avoid alienating customers. The simulations assumed customer behavior based on predefined parameters, which may not fully capture the complexity of real-world preferences and influences, such as competitor pricing or special events. Additionally, the models focused solely on profit maximization, without consideration of customer satisfaction or retention metrics.

Future Directions

Future work could expand the scope of the simulations to incorporate external factors like competitor pricing, loyalty programs, discount items or meal combos, and customer feedback loops. Integrating these elements would provide a more holistic view of pricing strategies and their impact. Further research could explore the development of hybrid pricing models that balance static and dynamic pricing strategies. For example, staples on the menu could remain at fixed prices, while premium or seasonal items could benefit from dynamic adjustments. Also, using feedback from customers could help refine pricing to align with customer expectations.

Conclusion

Dynamic pricing, when implemented strategically, offers significant potential for restaurants to maximize profits while adapting to demand fluctuations. The results of this project highlight the strengths of both PPO and DQN in achieving this goal. A hybrid pricing strategy, leveraging PPO for high-margin items and DQN for low-margin items, represents the optimal solution. This approach ensures profitability while maintaining the trust and loyalty of customers, paving the way for practical implementation in real-world scenarios.

Team Contributions

Aayush helped clean the data to prepare it for training. Nathanael helped with the customer simulation, using a Poisson distribution to simulate customer demand. Aashir helped in creating the simulator program to be fed into the models. Aayush, Aashir, and Yash helped create a plan on what agents to implement. Sumer helped with the PPO dynamic pricing agent while Aashir helped with the DQN agent and refining the PPO agent.

GitHub Repository

https://github.com/ashk0821/RestaurantDynamicPricing

References

- Asadinejad, A., & Rahmani, M. (2019). Dynamic pricing in energy markets using reinforcement learning. Energy Economics, 82, 170-180. https://doi.org/10.1016/j.eneco.2019.04.004
- Biswas, S., & Mahapatra, R. (2022). Fairness-aware reinforcement learning in dynamic pricing. Proceedings of the 15th International Conference on Artificial Intelligence and Data Science, 215-226.
- Brockman, G., Cheung, V., Pettersson, L., Schulman, J., Tang, J., & Ziegler, D. (2016). OpenAI Gym. arXiv preprint arXiv:1606.01540. https://arxiv.org/abs/1606.01540
- Burke, M. (2023). Dynamic Pricing Leads to Serious Losses: Busting the Top 3 Myths Surrounding This Emerging Tech. Capterra. https://www.capterra.com/resources/dynamic-pricing-for-restaurants/
- Chen, J., Li, X., & Xu, Y. (2018). Contextual bandits for pricing optimization. Proceedings of the 2018 International Conference on Machine Learning, 4630-4639.
- Karaesmen, Z., Koşar, S., & Özen, M. (2019). Application of reinforcement learning to pricing and inventory control in airlines. Transportation Science, 53(1), 44-59. https://doi.org/10.1287/trsc.2018.0872
- Raman, G., & Gao, F. (2020). Long-term effects of dynamic pricing strategies on customer satisfaction and loyalty. Journal of Marketing Research, 57(2), 320-338. https://doi.org/10.1177/0022243719892107