INFS 4020 – Big Data Concepts

Assignment 1 – Using reinforcement learning to assist retailers with dynamic pricing

Wangjun Shen

2022/09/08

Contents

Introduction

Due to the development of the Internet and the development of e-commerce, customers can more easily obtain information related to goods and services, Changes in commodity prices can also have a significant impact on consumers' shopping decisions in a short period of time, thereby affecting retailers' earnings. The goal of an enterprise is to maximize its own interests. Therefore, the enterprise will consider some factors (such as holidays, competitors' prices, etc.) to adjust the price of its own products, which can be considered a classic example of reinforcement learning in the real world. In reinforcement learning, the agent will adjust its next behaviours based on the feedback (award) it takes in the environment, and finally obtain a decision that maximizes its long-term benefits. This coincides with the retailer's goal of adjusting the selling price of its products to gain more revenue (Yin & Han, 2021).

Overview of Dynamic Pricing

The reduction of time and space costs for buyers and sellers to participate in transactions in the retail market is one reason for the shift in pricing from static to dynamic, and another reason is the elimination of information barriers and the increase of competitors brought about by the development of the Internet. For retailers, they need to be able to respond to changes in the "environment" to adjust product prices in a timely manner, so that they will not fall behind in competition with competitors in the era of developed Internet.

Dynamic pricing is a way for retailers to adjust commodity prices according to their own supply capacity, changes in commodity costs and changes in user demand in an attempt to obtain more benefits. (Elmaghraby & Keskinocak, 2003).

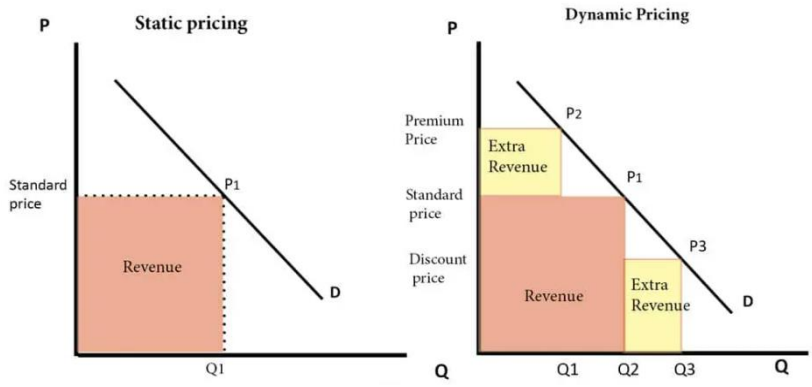


Diagram 1 Static Pricing vs. Dynamic Pricing (Pettinger, 2019)

Overview of Reinforcement Learning

Reinforcement learning interacts with the environment in a "trial and error" way, and learns the optimal strategy by maximizing the accumulated reward (Rao, 1998), it learns by interacting with the environment and based on the immediate reward signal obtained during the interaction, in order to maximize the expectation of accumulating positive reward, which is an important branch of machine learning.

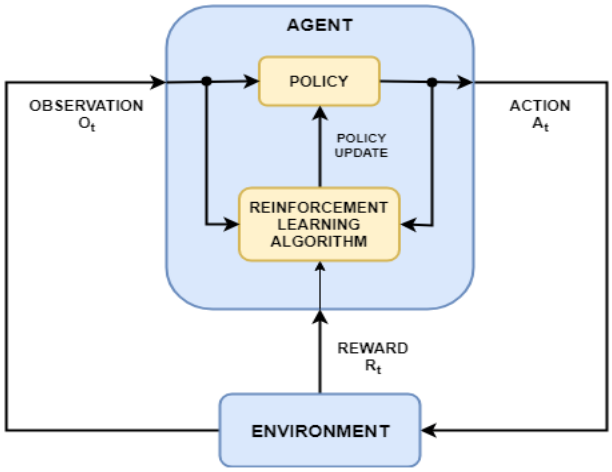


Diagram 2 Reinforcement Learning Diagram

Reinforcement learning architectures can be broken down into four main parts:

a) Policy: Ways to act and adjust decisions
b) Environment: The external situation of the agent in the activity
c) Result feedback: What are the consequences of the decisions and actions the agent is making?
d) Evaluation function: Used to evaluate the outcome and thus decide whether to "reward" or "punish"

For the research of reinforcement learning in dynamic pricing, the research is usually based on the following two environmental models:

a) Markov Decision Process (MDP): a mathematical model of sequential decision, which is used to simulate the achievable random policies and rewards of an agent in an environment where the system state has Markov properties (Sutton & Barto, 2018).
b) Semi-MDP (SMDP): an extension of Markov Decision Processes for modelling stochastic control problems. Different from Markov decision process, each state of semi-Markov decision process has a certain sojourn time, and the sojourn time is a general continuous random variable (Jewell, 1963).

There are two types of algorithms for reinforcement learning:

a) Value function-based: choosing the best action in the sate
b) Policy-based: Learning through state-action random mapping functions

For Value function-based, academic research typically uses two learning algorithms as shown below:

a) Q-learning algorithm: a model-free algorithm (Mnih et al., 2015).
b) SARSA algorithm: An algorithm that attempts to find optimal decisions by iterating over state-action value functions, used in the context of reward functions and state transition probability positions. The convergence speed of the SARSA algorithm is slow, so a conservative strategy should be adopted for the algorithm (Lin & Kim, 1991).

Using Reinforcement Learning in Dynamic Pricing

For retail market transactions, it can be mapped from display problems to reinforcement learning:

REINFORCEMENT LEARNING	RETAIL MARKET
ENVIRONMENT	refers to the trading market
STATUS	The lowest price of the current market, inventory, whether it is a holiday or a special day and other factors
AGENT	Dynamic Pricing Algorithm
RESPONSE	Raise or lower prices, lower or cancel shipping
REWARD	Total profit from dynamic pricing by agent

When the reinforcement learning method is used to solve the dynamic pricing problem, factors such as the number of suppliers (single-supplier, multi-supplier), the environment model and the selection algorithm need to be considered. Whether traditional or reinforcement learning-based multi-vendor dynamic pricing research usually assumes the number of suppliers in the market, generally divided into single-supplier and multi-supplier. For multi-vendor research, most of them assume that there are two suppliers in the market, and there is some kind of competitive relationship between the two suppliers. In the research based on reinforcement learning, it is represented as two agents, and there is mutual influence and competition between them.

Kutschinski et al (2003) studied the competition among multiple suppliers in the e-commerce market, but still continued the assumption of two suppliers in terms of the number of suppliers, this study showed that sellers who adopted Q-learning were able to adopt almost the best response price strategy against those who used fixed price strategies:

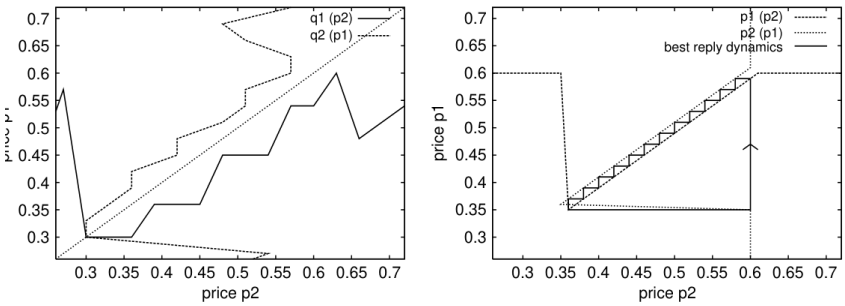


Diagram 3 Dynamic Pricing vs. Fixed Pricing (Kutschinski et al., 2003)

Rana and Oliveira proposed a method for establishing an expert system to solve the dynamic pricing problem in the case of only one supplier using the SARSA algorithm. The premise of this research is that the decision-making process conforms to the MDP, and the profit growth is realized through the dynamic pricing algorithm in the simulation experiment (Rana & Oliveira, 2015).

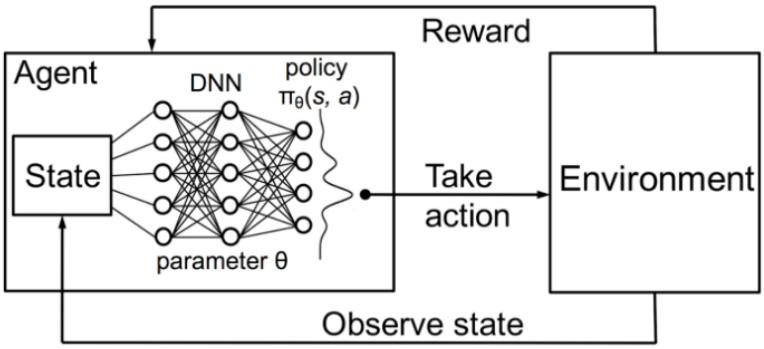
Limitations and Issues to Using Reinforcement Learning

There are disadvantages to using dynamic pricing. The following table compares the advantages and disadvantages of dynamic pricing:

Advantages	Disadvantages
higher profit	User dissatisfaction and churn
Respond effectively to competition from others	The two sides are caught in a vicious competition
be able to acquire some new customers	Dissatisfaction and loss of old customers

Although the strategy of using dynamic pricing allows retailers to adjust the price of goods to obtain higher profits in the short term, the loss of some customers and the vicious PR incidents caused by abnormal prices may lead to lower profits in the long term.

Although there are several other algorithms, the Q-learning algorithm and SARSA are mainly used. The Q-Learning algorithm has been widely used because of its good results. However, the Q-learning algorithm belongs to the tabular algorithm and has a relatively good learning effect for small-scale and discrete systems, but for continuous large-scale systems, there will be cases where the convergence speed is slow or cannot be converged. An effective solution to this problem is to combine deep learning and reinforcement learning, so that the trained model can have the perception ability of deep learning and the decision-making ability of reinforcement learning:



Deep Reinforcement Learning Loop Diagram (Mao et al., 2016)

A successful example of deep reinforcement learning algorithm is alpha go, which also has good development in natural language processing, robot control, etc., but more research is needed in the field of dynamic pricing to determine whether it is really efficient.

One of the advantages of reinforcement learning is that it does not require a lot of data or even data to train the model, but this is also a disadvantage when using reinforcement learning to dynamically price retailers, because retailers actually hold a lot of labelled data, the value of this data is not effectively utilized.

In addition, reinforcement learning is based on the premise that the entire process is a Markov decision process or a process that approximates a Markov decision process, but the problem is that the user's behaviours is affected by many factors, which makes this premise untenable in the real world. Whether a consumer decides to consume not only depends on the previous state of the consumer, but also the previous state and other consumption states will also affect the current state.

[word count: 1361]

References

Andrew, A. (1999). REINFORCEMENT LEARNING: AN INTRODUCTION by Richard S. Sutton and Andrew G. Barto, Adaptive Computation and Machine Learning series, MIT Press (Bradford Book), Cambridge, Mass., 1998, xviii 322 pp, ISBN 0-262-19398-1, (hardback, £31.95). *Robotica, 17*(2), 229-235. doi:10.1017/S0263574799211174

Elmaghraby, W., & Keskinocak, P. (2003b, October). Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions. *Management Science*, *49*(10), 1287–1309. https://doi.org/10.1287/mnsc.49.10.1287.17315

Jewell, W. S. (1963). Markov-Renewal Programming. I: Formulation, Finite Return Models. *Operations Research*, *11*(6), 938–948. http://www.jstor.org/stable/167834

Lin, C. S., & Kim, H. (1991). CMAC-based adaptive critic self-learning control. *IEEE Transactions on Neural Networks*, *2*(5), 530–533. https://doi.org/10.1109/72.134290

Mao, H., Alizadeh, M., Menache, I., & Kandula, S. (2016, November 9). Resource Management with Deep Reinforcement Learning. *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*. https://doi.org/10.1145/3005745.3005750

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015, February 25). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Pettinger, T. (2019, June 12). Dynamic Pricing. *Economics Help*. Retrieved September 9, 2022, from https://www.economicshelp.org/blog/148008/economics/dynamic-pricing/

Rao, R. P. N. (2000). Reinforcement Learning: An Introduction; RS Sutton, AG Barto (Eds.); MIT Press, Cambridge, MA, 1998, 380 pages, ISBN 0-262-19398-1, $42.00.

Rana, R., & Oliveira, F. S. (2015, January). Dynamic pricing policies for interdependent perishable products or services using reinforcement learning. *Expert Systems With Applications*, *42*(1), 426–436. https://doi.org/10.1016/j.eswa.2014.07.007

Sutton, R. S., & Barto, A. G. (2018, November 13). Reinforcement Learning, second edition: An Introduction (Adaptive Computation and Machine Learning series) (second edition). Bradford Books.

Yin, C., & Han, J. (2021). Dynamic Pricing Model of E-Commerce Platforms Based on Deep Reinforcement Learning. *Computer Modeling in Engineering &Amp; Sciences*, *127*(1), 291–307. https://doi.org/10.32604/cmes.2021.014347