Industrial Management Journal
Print ISSN: 2008-5885
Online ISSN: 2423-5369

Dynamic Pricing of Customer Classes in Rail Transportation Systems Using Deep Q Network Algorithm

Omid Niknami
MSc. Student, Department of Industrial Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran. E-mail: omid_niknami@modares.ac.ir

Elham Akhondzadeh Noughabi *
*Corresponding Author, Assistant Prof., Department of Information Technology, Faculty of Industrial Engineering and Systems, Tarbiat Modares University, Tehran, Iran. E-mail: elham.akhondzadeh@modares.ac.ir

Abstract

Objective

This research investigates the problem of dynamic pricing in rail transportation systems using advanced deep reinforcement learning techniques. The main goal is to optimize the revenue of railway transport companies by developing a ticket sales policy that dynamically adjusts ticket prices based on service classes. This approach allows rail transport companies to enhance revenue and profitability by accurately aligning prices with passenger demand.

Methods

To solve the problem of dynamic pricing, this research utilizes the Q deep network algorithm, which combines deep neural networks with Q-learning. Deep neural networks approximate Q values instead of using a costly Q table. The Q deep network algorithm is widely used due to its ability to learn optimal policies in complex environments. As reinforcement learning models are often too complex to analyze, numerical experiments and simulations are used to analyze different pricing strategies.

Results

The simulations demonstrate that the Q deep network algorithm successfully converges to a stable pricing policy. Various performance indicators were investigated, including such as total revenue, remaining capacity, average prices offered to customers, and the number of tickets sold in each service class. The algorithm showed improvement in the early stages and gradually achieved stability. The average total revenue converges to 225,000 after 5,000 iterations, indicating that the company earns an average of 225,000 monetary units from each train. The average residual capacity approaches zero after approximately 3,000 iterations,

indicating that the reinforcement learning agent learns to sell all available tickets to maximize total revenue. The average price index offered to customers stabilizes after approximately 7,500 iterations, indicating that the algorithm has converged to an optimal pricing policy. In this state, the average prices remain within the range of 680 to 700 monetary units, with no significant fluctuations observed. In other words, the reinforcement learning model has successfully converged based on the average proposed price index. Finally, after about 5,000 iterations, the average number of tickets sold for all service classes reaches a stable level. The average number of tickets sold for economy class is around 175 to 180 tickets, for business class is around 130 to 135 tickets, for special class is around 60 to 65 tickets, and for hotel class is around 23 to 25 tickets.

Conclusion

The findings of this study suggest that employing the Deep Q-Network algorithm in dynamic pricing can lead to substantial optimization in revenue management for railway transportation systems. The results of this research indicate that after approximately 7,500 iterations, the Q deep network algorithm reaches an optimal and stable policy with no significant changes in performance. It can be concluded that the use of the Q deep network algorithm in dynamic pricing can significantly improve the revenue management of rail transportation systems. This algorithm can learn and adapt to changing conditions, allowing for effective pricing policies to maximize revenue and determine the optimal number of tickets sold in each service class. The obtained findings can help rail transport companies improve pricing strategies and increase economic productivity.

Keywords: Dynamic pricing, Reinforcement learning, Rail transportation.

Citation: Niknami, Omid & Akhondzadeh Noughabi, Elham (2024). Dynamic Pricing of Customer Classes in Rail Transportation Systems Using Deep Q Network Algorithm. *Industrial Management Journal*, 16(4), 597-630. (*in Persian*)

Industrial Management Journal, 2024, Vol. 16, No 4, pp. 597- 630
Published by University of Tehran, Faculty of Management
https://doi.org/10.22059/IMJ.2024.377050.1008164
Article Type: Research Paper

Received: May 27, 2024
Received in revised form: October 10, 2024
Accepted: November 11, 2024
Published online: December 02, 2024

مدیریت صنعتی
شاپا چاپی: ۲۰۰۸-۵۸۸۵
شاپا الکترونیکی: ۲۴۲۳-۵۳۶۹

قیمت‌گذاری پویای کلاس‌های مشتریان در سیستم‌های حمل‌ونقل ریلی با استفاده از الگوریتم شبکهٔ عمیق Q

امید نیک‌نامی
دانشجوی کارشناسی ارشد، گروه مهندسی صنایع، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.
رایانامه: omid_niknami@modares.ac.ir

الهام آخوندزاده نوقابی*
* نویسنده مسئول، استادیار، گروه فناوری اطلاعات، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران.
رایانامه: elham.akhondzadeh@modares.ac.ir

چکیده

هدف: در این پژوهش به بررسی مسئلهٔ قیمت‌گذاری پویا در سیستم‌های حمل‌ونقل ریلی با استفاده از تکنیک‌های پیشرفته یادگیری تقویتی عمیق پرداخته شده است. هدف اصلی این پژوهش، بهینه‌سازی درآمد شرکت‌های حمل‌ونقل ریلی، از طریق ایجاد سیاست فروش بلیت است که بتواند قیمت بلیت را با در نظر گرفتن کلاس‌های خدماتی به‌صورت پویا تنظیم کند. این رویکرد به شرکت‌های حمل‌ونقل ریلی این امکان را می‌دهد تا با تنظیم دقیق‌تر قیمت‌ها، بر اساس تقاضای مسافران، درآمد بیشتری کسب کنند و سودآوری خود را بهبود بخشند.

روش: در این پژوهش، به‌منظور حل مسئله قیمت‌گذاری پویا، از الگوریتم شبکهٔ عمیق Q، یکی از الگوریتم‌های پُرکاربرد یادگیری تقویتی عمیق، بهره گرفته شده است. این الگوریتم یک الگوریتم پیش‌گامانه است که شبکه‌های عصبی عمیق را با یادگیری Q ترکیب می‌کند. در الگوریتم شبکهٔ عمیق Q، شبکه‌های عصبی عمیق وظیفهٔ تقریب مقادیر Q را به‌جای جدول پُرهزینهٔ Q برعهده دارند. یک شبکهٔ عصبی عمیق می‌تواند حالت فعلی و یک عمل ممکن را مشاهده کند و به‌صورت مستقیم مقدار Q را تقریب بزند. توانایی الگوریتم شبکهٔ عمیق Q در یادگیری سیاست‌های بهینه در محیط‌های پیچیده، این الگوریتم را به یک الگوریتم پُرکاربرد تبدیل کرده است. با توجه به این نکته که مدل‌های یادگیری تقویتی، اغلب برای تحلیل بیش از حد پیچیده به‌کار می‌روند، تحلیل استراتژی‌های قیمت‌گذاری متفاوت، فقط با آزمایش‌های عددی و شبیه‌سازی به‌دست می‌آیند.

یافته‌ها: نتایج شبیه‌سازی‌ها نشان می‌دهد که الگوریتم شبکهٔ عمیق Q، به‌طور موفقیت‌آمیزی به یک سیاست قیمت‌گذاری پایدار هم‌گرا تبدیل شده است. در این مطالعه، شاخص‌های عملکردی مختلفی مانند درآمد کل، ظرفیت باقی‌مانده، میانگین قیمت‌های ارائه‌شده به مشتریان و تعداد بلیت‌های فروخته‌شده در هر کلاس خدماتی بررسی شد. الگوریتم در مراحل اولیه با نرخ یادگیری بالا بهبود و به‌تدریج به پایداری و ثبات دست یافت. میانگین درآمد کل، پس از ۵۰۰۰ تکرار، به مقدار ۲۲۵ هزار هم‌گرا می‌شود. این مقدار نشان‌دهندهٔ آن است که این شرکت به‌طور متوسط، از هر قطار ۲۲۵ هزار واحد پولی درآمد کسب می‌کند. میانگین ظرفیت باقی‌مانده نیز پس از حدود ۳۰۰۰ تکرار، به مقدار صفر نزدیک می‌شود. هم‌گرایی نمودار میانگین ظرفیت باقی‌مانده به صفر، نشان‌دهندهٔ این است که عامل یادگیری تقویتی، به‌طور مؤثری یاد گرفته است که برای حداکثرسازی درآمد کل، می‌بایست تمامی بلیت‌های موجود را به فروش برساند. شاخص میانگین قیمت‌های پیشنهادی به مشتریان، پس از حدود ۷۵۰۰ تکرار به یک حالت پایدار می‌رسد، به این

معنا که الگوریتم به یک سیاست قیمت‌گذاری بهینه دست پیدا کرده است. در این وضعیت، میانگین قیمت‌ها در محدوده ۶۸۰ تا ۷۰۰ واحد پولی باقی می‌ماند و نوسان‌های چشمگیری مشاهده نمی‌شود. به عبارت دیگر، مدل یادگیری تقویتی بر اساس شاخص میانگین قیمت‌های پیشنهادی هم‌گرا شده است. در نهایت، پس از حدود ۵۰۰۰ تکرار، میانگین تعداد بلیت فروخته شده برای تمامی کلاس‌های خدماتی، به یک سطح تقریباً ثابت و پایدار می‌رسد. میانگین تعداد بلیت فروخته شده برای کلاس اکونومی، حدود ۱۷۵ تا ۱۸۰ بلیت، کلاس بیزینس حدود ۱۳۰ تا ۱۳۵ بلیت، کلاس ویژه حدود ۶۰ تا ۶۵ بلیت و کلاس هتل حدود ۲۳ تا ۲۵ بلیت به‌دست آمد.

نتیجه‌گیری: نتایج این پژوهش نشان می‌دهد که پس از حدود ۷۵۰۰ تکرار، الگوریتم شبکهٔ عمیق Q به یک سیاست بهینه و پایدار رسیده است و تغییرات چشمگیری در عملکرد مشاهده نمی‌شود. همچنین می‌توان نتیجه گرفت که استفاده از الگوریتم شبکهٔ عمیق Q در قیمت‌گذاری پویا، می‌تواند به بهبود چشمگیری در مدیریت درآمد سیستم‌های حمل‌ونقل ریلی منجر شود. این الگوریتم با قابلیت یادگیری و سازگاری با شرایط متغیر، قادر است که سیاست‌های قیمت‌گذاری مؤثری را با هدف حداکثرسازی درآمد به‌کار گیرد و همچنین، تعداد بهینهٔ بلیت‌های فروخته‌شده در هر کلاس خدماتی را تعیین کند. این دستاوردها می‌توانند به شرکت‌های حمل‌ونقل ریلی در بهبود راهبردهای قیمت‌گذاری و افزایش بهره‌وری اقتصادی کمک شایانی کنند.

کلیدواژه‌ها: قیمت‌گذاری پویا، یادگیری تقویتی عمیق، حمل‌ونقل ریلی.

استناد: نیک‌نامی، امید و آخوندزاده نوقابی، الهام (۱۴۰۳). قیمت‌گذاری پویای کلاس‌های مشتریان در سیستم‌های حمل‌ونقل ریلی با استفاده از الگوریتم شبکهٔ عمیق Q. *مدیریت صنعتی*، ۱۶(۴)، ۵۹۷- ۶۳۰.

مدیریت صنعتی، ۱۴۰۳، دوره ۱۶، شماره ۴، صص. ۵۹۷- ۶۳۰
ناشر: دانشکدهٔ مدیریت دانشگاه تهران
نوع مقاله: علمی پژوهشی

تاریخ دریافت: ۱۴۰۳/۰۳/۰۷
تاریخ ویرایش: ۱۴۰۳/۰۷/۱۹
تاریخ پذیرش: ۱۴۰۳/۰۸/۲۱
تاریخ انتشار: ۱۴۰۳/۰۹/۱۲
doi: https://doi.org/10.22059/IMJ.2024.377050.1008164

مقدمه

در فرایند خرید، قیمت یکی از عوامل اصلی است. انعطاف‌پذیری این عامل اهمیت نقش آن را در به حداکثر رساندن سود فروشنده دوچندان می‌کند که تأمین‌کنندگان و مدیران را به تمرکز بر قیمت‌گذاری سوق داده است. ظهور اینترنت فرصتی را برای مدیران شرکت‌های فناوری اطلاعات فراهم کرده است تا قیمت محصولات خود را در بازار رقابتی به‌صورت پویا تنظیم کنند. در این شرایط، اهمیت قیمت‌گذاری پویا افزایش یافته و نقش ابزارها و روش‌های هوشمند در این فرایند بیشتر شده است. علاوه‌براین، توسعهٔ یک استراتژی قیمت‌گذاری مناسب برای مدیران شرکت‌های فناوری اطلاعات یک مشکل دشوار و درعین‌حال یک مشکل اساسی است و عدم درک صحیح تصمیمات قیمت‌گذاری، به از دست رفتن فرصت‌های سودآور بازار منجر می‌شود. قیمت‌گذاری فرایندی چندبعدی است که تحت تأثیر محصول، حاشیهٔ سود و روابط با مشتری قرار می‌گیرد. درنتیجه بسیاری از شرکت‌های تولیدی و خدماتی برای قیمت‌گذاری مناسب محصولات خود با مشکلات جدی مواجهند (عادل‌نیا نجف‌آبادی، شکرچی‌زاده، نبی‌اللهی، خانی و رستگاری[1]، ۲۰۲۲).

قیمت‌گذاری پویا ابزار مدرن شرکت‌های حمل‌ونقل ریلی، هواپیمایی و اتوبوس‌رانی است که هدف آن افزایش درآمد ناشی از محاسبه به‌موقع تقاضای مسافر و تعدیل متوالی قیمت بلیت است. در این تنظیمات پویا، تصمیم‌گیری مسافران در مورد زمان خرید بلیت موردنظر برای صرفه‌جویی در هزینهٔ مفید است. تعیین این قیمت‌ها اصلی‌ترین چالش قیمت‌گذاری پویاست (استاوینوا، چونافِ و بوچِنینا[2]، ۲۰۲۱). قیمت‌گذاری پویا بر اساس عوامل مختلفی محاسبه می‌شود که بر آن تأثیر می‌گذارد. این عوامل ممکن است با زمان، تقاضا، شرایط آب‌وهوایی و فرهنگ تغییر کنند؛ بنابراین، تجزیه‌وتحلیل و اولویت‌بندی صحیح کلیه عوامل و پارامترها به‌صورت دوره‌ای، نقش مهمی در ساخت مدل‌های قیمت‌گذاری پویا ایفا می‌کند (ساهاران، باوا و کومار[3]، ۲۰۲۰).

در طول ۴۰ سال گذشته پیشرفت‌های زیادی در زمینهٔ مدیریت درآمد و قیمت‌گذاری پویا حاصل‌شده است. این رشته اکنون به‌خوبی تثبیت شده است، در حالی‌که صنایع هواپیمایی و هتلداری به‌اندازهٔ کافی در کانون توجه قرار گرفته‌اند، صنعت حمل‌ونقل ریلی کمتر در کانون توجه قرار گرفته است و تحقیقات اندکی در این حوزه موجود است (آرمسترانگ و مایسنر[4]، ۲۰۱۰). درآمد شرکت‌های حمل‌ونقل ریلی بیشتر از فروش بلیت‌های صندلی به دست می‌آید؛ بنابراین طبیعی است که شرکت سعی در ایجاد یک سیاست بهینه فروش بلیت برای دستیابی به حداکثر درآمد ممکن برای هر قطار داشته باشد (یانگ، ژو و یانگ[5]، ۲۰۱۲).

در دهه‌های اخیر، تحولات چشمگیری در زمینه یادگیری ماشین صورت گرفته که به‌طور چشمگیری بر حوزه‌های مختلف علمی و صنعتی تأثیر گذاشته است. یکی از کاربردهای برجستهٔ این تکنولوژی‌ها، قیمت‌گذاری پویاست که به‌دلیل پیچیدگی‌ها و تغییرات مداوم در بازارها، به رویکردهای هوشمند و تطبیقی نیازمند است. در این راستا، یادگیری ماشین، یادگیری تقویتی و یادگیری تقویتی عمیق به‌عنوان ابزارهای کارآمد و پیشرفته، جایگاه ویژه‌ای در بهینه‌سازی

1. Adelnia Najafabadi, Shekarchizadeh, Nabiollahi, Khani & Rastgari
2. Stavinova, Chunaev & Bochenina
3. Saharan, Bawa & Kumar
4. Armstrong & Meissner
5. Yang, Xu & Yang

استراتژی‌های قیمت‌گذاری پویا پیدا کرده‌اند (برتسیماس و کالوس[1]، ۲۰۲۰). در زمینهٔ قیمت‌گذاری پویا، یادگیری تقویتی می‌تواند با شبیه‌سازی سناریوهای مختلف بازار و تحلیل نتایج حاصل از هر استراتژی، به بهینه‌سازی قیمت‌ها کمک کند (ساتون و بارتو[2]، ۲۰۱۸). این روش به‌خصوص در شرایطی که داده‌های کافی برای آموزش مدل‌های سنتی وجود ندارد یا بازار به‌سرعت در حال تغییر است، می‌تواند بسیار مؤثر باشد.

در این راستا، تمرکز این تحقیق بر قیمت‌گذاری پویا برای کلاس‌های مختلف مشتریان با توجه به شرایط و عوامل تأثیرگذار روی قیمت و همچنین با هدف حداکثر شدن درآمد با استفاده از روش‌ها یادگیری تقویتی عمیق است. بر این اساس سؤال‌های پژوهش عبارت‌اند از:

۱. چگونه می‌توان قیمت‌گذاری پویا در سیستم‌های حمل‌ونقل ریلی را باهدف حداکثرسازی درآمد و با درنظر گرفتن کلاس‌های مختلف خدماتی (اکونومی، بیزینس، ویژه و هتل) با استفاده از روش‌های یادگیری تقویتی عمیق مدل‌سازی کرد؟

۲. چگونه می‌توان با استفاده از یادگیری تقویتی عمیق و الگوریتم شبکهٔ عمیق Q، یک مدل قیمت‌گذاری پویا برای سیستم‌های حمل‌ونقل ریلی در راستای حداکثرسازی درآمد ارائه کرد؟ چگونه می‌توان تعداد بهینه بلیت‌های فروخته شده در هر کلاس خدماتی را تعیین کرد تا درآمد کلی شرکت حمل‌ونقل ریلی حداکثر شود؟

اهداف این پژوهش را می‌توان در قالب سه مورد ذیل بیان کرد:

۱. مدل‌سازی قیمت‌گذاری پویا با در نظرگرفتن کلاس‌های مختلف مشتریان؛

۲. مدل‌سازی قیمت‌گذاری پویا با استفاده از یادگیری تقویتی عمیق و بررسی عملکرد الگوریتم شبکهٔ عمیق Q در سیستم‌های حمل‌ونقل ریلی؛

۳. به‌دست آوردن تعداد بهینه بلیت‌های فروخته شده در هر کلاس خدماتی.

جنبه‌های جدید و نوآوری که در این پژوهش وجود دارد، استفاده از روش‌های یادگیری تقویتی برای مسئله قیمت‌گذاری پویا با درنظر گرفتن کلاس‌های مشتریان در صنعت حمل‌ونقل ریلی است. روش‌های سنتی نمی‌توانند در بلندمدت قابل‌قبول باشند. با پیشرفت تکنولوژی، افزایش حجم داده‌ها و همچنین پیچیدگی بیشتر مسائل، برای پُرکردن شکاف‌های موجود، نیاز به روش یا روش‌هایی است که بتوانند جایگزین روش‌های سنتی شوند و قیمت‌های بلیت را به‌صورت پویا بادقت و عملکرد بهتری تخمین بزنند و به هر مشتری قیمت اختصاصی پیشنهاد دهند. بر اساس مرور ادبیات انجام شده، مقالات اندکی به بررسی مسئله قیمت‌گذاری پویا در سیستم‌های حمل‌ونقل ریلی با بهره‌گیری از روش‌های یادگیری ماشین پرداخته‌اند (جدول ۱). بر اساس دانش ما، این پژوهش نخستین پژوهشی است که از روش یادگیری تقویتی عمیق برای قیمت‌گذاری پویا در صنعت حمل‌ونقل ریلی استفاده می‌کند. همچنین در این پژوهش، برای نخستین بار کلاس‌های مختلف خدماتی شامل کلاس اکونومی، بیزینس، ویژه و هتل به‌عنوان بخش‌های مختلف مشتریان در صنعت حمل‌ونقل ریلی درنظر گرفته شده است.

1. Bertsimas and Kallus
2. Sutton and Barto

پژوهش حاضر به این صورت ساختاربندی شده است: در بخش دوم پیشینهٔ نظری بررسی می‌شود. بخش سوم به روش‌شناسی پژوهش اختصاص یافته است. در بخش چهارم یافته‌ها ارائه شده است و در بخش پنجم، نتیجه‌گیری بیان خواهد شد.

پیشینهٔ نظری پژوهش

مدیریت درآمد به نظریه و عمل مدیریت تقاضا با استفاده از ابزارهایی مانند قیمت‌ها یا دسترسی به محصولات بر اساس مدل‌های تقاضا اشاره دارد که باهدف حداکثرسازی سود یا درآمد انجام می‌شود. به‌عنوان یک حوزهٔ تخصصی، مدیریت درآمد از دههٔ ۱۹۷۰ در صنعت هواپیمایی آمریکا شکل گرفت. تکنیک‌های مدیریت درآمد به بهینه‌سازی درآمد محدود نمی‌شوند و برای حداکثرسازی اهداف دیگر مانند سود نیز استفاده می‌شوند. تمرکز مدیریت درآمد بر روی تنظیم قیمت‌ها با استفاده از سیستم‌های کامپیوتری هوشمندی است که به‌طور خودکار به فروش محصولات یا خدمات کمک می‌کنند و تقاضای مشتریان را پیش‌بینی می‌کنند. این پیش‌بینی‌ها به‌دنبال بهینه‌سازی تصمیمات مدیریت تقاضا استفاده می‌شوند. از زمان تدوین این مفهوم، آن را می‌توان در محدوده‌های متعددی از جمله حمل‌ونقل (قطارها، اجاره خودرو، کشتی‌ها، حمل‌ونقل بار)، هتل‌ها، رسانه و تبلیغات و سایر زمینه‌ها مشاهده کرد (اشتراوس، کلاین و اشتاینهارت[1]، ۲۰۱۸).

قیمت‌گذاری پویا یکی از اساسی‌ترین و رایج‌ترین ابزارهای مدیریت درآمد است. این ابزار شرکت‌ها را قادر می‌سازد تا با تطبیق بهتر عرضه با تقاضا، پاسخ به الگوهای تغییر تقاضا و دستیابی به تقسیم‌بندی مشتری، درآمد را افزایش دهند. قیمت‌گذاری پویا یک سازوکار قیمت‌گذاری است که در آن شرکت‌ها می‌توانند قیمت انتخابی خود را هرازچندگاهی به‌روز کنند. قیمت‌ها می‌توانند هفتگی، ماهانه، روزانه یا چندین بار در روز تغییر کنند و با انتقال از یک تراکنش به تراکنش دیگر به‌عنوان تابعی از اطلاعات خریدار، قیمت رقبا، یا موجودی باقی‌مانده باشند (ویتمن و بلوبابا[2]، ۲۰۱۹). روش معمول شرکت‌های حمل‌ونقل ریلی و هواپیمایی برای انجام مدیریت درآمد، کنترل موجودی صندلی است. شرکت هواپیمایی چندین قیمت یا کلاس برای همه صندلی‌های پرواز تعیین می‌کند، سپس تعداد فروش هر کلاس را با توجه به تقاضای بازار کنترل می‌کند. بدیهی است که هرچه تعداد کلاس‌های بیشتری تشکیل دهند، درآمد بیشتری می‌توانند کسب کنند (گائو، لی و فنگ[3]، ۲۰۲۲). قیمت‌گذاری پویا نیز یکی دیگر از روش‌های مدیریت درآمد است. برگر و فاچز[4] (۲۰۱۵) نشان دادند یک شرکت هواپیمایی که از قیمت‌گذاری پویا استفاده می‌کند و با شرکتی که مدیریت درآمد سنتی را به‌کار می‌گیرد، رقابت می‌کند، افزایش درآمد چشمگیری را به همراه خواهد داشت. در یک دوره کوتاه‌مدت، درآمد حاصل از قیمت‌گذاری پویا می‌تواند تا ۲۰ درصد افزایش یابد (برگر و فاچز، ۲۰۱۵). در مقایسه با حمل‌ونقل هوایی، حمل‌ونقل ریلی ظرفیت‌های بزرگ‌تر، توقف‌های میانی‌تر و برنامه‌های عملیاتی پیچیده‌تری دارد. در نتیجه، مسئله قیمت‌گذاری پویای حمل‌ونقل ریلی از نظر محاسباتی بسیار دشوارتر از صنعت هواپیمایی است. بنابراین، روش قیمت‌گذاری پویا خطوط هوایی را نمی‌توان به‌طور مستقیم برای مسئله قیمت‌گذاری پویا حمل‌ونقل ریلی اعمال کرد (وو، کین، کیو، زنگ و

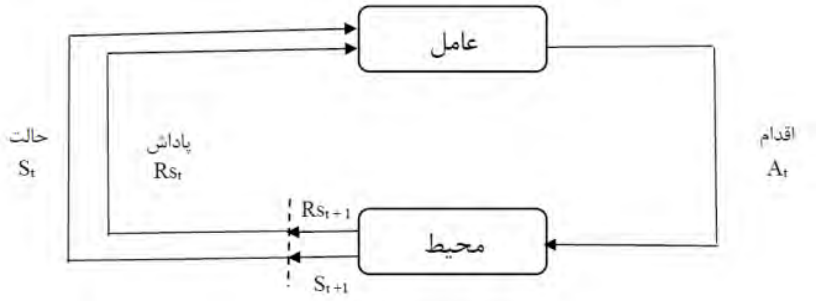
1. Strauss, Klein & Steinhardt
2. Wittman & Belobaba
3. Gao, Le & Fang
4. Burger & Fuches

یانگ[1]، ۲۰۱۹). در حمل‌ونقل ریلی مسافران، قیمت بلیت در مسیرهای مختلف اغلب ثابت است و با تغییر تعداد مسافران تغییری نمی‌کند. این ثابت بودن قیمت‌ها باعث می‌شود که برخی از قطارها بیش‌ازحد شلوغ شوند، در حالی که برخی دیگر خالی می‌مانند. از طرف دیگر، همهٔ مسافران، صرف‌نظر از میزان وفاداری یا استفاده‌شان از خدمات، بلیت‌ها را با قیمت یکسان خریداری می‌کنند. این موضوع نه‌تنها به شرکت راه‌آهن کمک نمی‌کند تا جریان مسافران را بهتر مدیریت کند، بلکه رقابت‌پذیری آن را در مقایسه با سایر روش‌های حمل‌ونقل مانند خطوط هوایی کاهش می‌دهد و در نتیجه بهبود درآمد شرکت راه‌آهن دشوار می‌شود (ژو، ونگ، لو و پان[2]، ۲۰۱۴).

یادگیری ماشین از جامعه علوم کامپیوتر سرچشمه گرفته و از مطالعه تشخیص الگو و نظریه یادگیری محاسباتی در هوش مصنوعی تکامل یافته است. به‌نوعی، یادگیری ماشین علم وادار کردن کامپیوترها به عمل‌کردن بدون برنامه‌ریزی صریح است. از دههٔ ۱۹۹۰ مقدار شایان توجهی از منابع و تحقیقات به توسعه الگوریتم‌های یادگیری پیچیده اختصاص داده‌شده است و آن‌ها در زمینه‌های مختلف به‌کار گرفته شده‌اند. در این مبحث مدل‌ها و رویکردهای مختلفی وجود دارد؛ اما معمول است که الگوریتم‌های یادگیری ماشین را به سه دستهٔ کلی بر اساس ماهیت سیگنال یادگیری یا بازخورد موجود در الگوریتم به سه دسته یادگیری نظارت‌شده، یادگیری بدون نظارت و یادگیری تقویتی دسته‌بندی می‌کنند. به یک عامل یادگیری تحت نظارت، باید حرکت صحیح برای هر حالتی که با آن روبه‌رو می‌شود، گفته شود؛ اما چنین بازخوردی به‌ندرت در دسترس است (راسل و نورویگ[3]، ۲۰۰۹). با این‌حال، یادگیری تقویتی، روشی است که توسط عوامل یادگیری ماشین برای یادگیری آنچه در غیاب نمونه‌های برچسب‌گذاری شده از کارهایی که باید انجام شود، انجام دهند، استفاده می‌شود.

یک سیستم یادگیری تقویتی به‌طور اساسی از دو بخش محیط و عامل تشکیل‌شده است. عوامل قادرند از طریق حسگرها ورودی‌ها را از محیط دریافت کرده و با استفاده از محرک‌ها روی محیط تأثیر بگذارند. به‌طورکلی، هر چیزی که در تعامل با عامل باشد و از آن دریافت یا به آن تأثیر داده شود، می‌تواند به‌عنوان محیط در نظر گرفته شود. در این چارچوب، عامل در تعامل با محیط خود، تجربه‌هایی کسب می‌کند و بر اساس آن‌ها بهبود می‌یابد. ازاین‌رو، یادگیری تقویتی یک فرایند تعاملی بین عامل و محیط است که به‌وسیله آن عامل می‌تواند اقدامات خود را برای بهبود عملکرد و تطابق با محیط بهبود بخشد. یادگیرنده در این رویکرد باید از پاداش‌های غیرمستقیم و با تأخیر یاد بگیرد که چگونه دنباله‌ای از اقدامات را انتخاب کند که پاداش تجمعی را به حداکثر برساند. معمولاً راه‌حلی که سریع‌ترین است (از نظر تعداد اقدامات) ارجحیت دارد؛ بنابراین، ترکیبی از پاداش‌های انباشته عمدتاً به‌گونه‌ای انتخاب می‌شود که پاداش‌های فوری به‌عنوان اولویت‌های اصلی موردتوجه قرار گیرند. این رویکرد به آزمون‌وخطا برای کشف راه‌حل‌های بهتر و بهبود استراتژی عمل عامل نیاز دارد. اصل اساسی یادگیری تقویتی در شکل ۱ نشان داده شده است (لیو، چن، لی، دوان و لی[4]، ۲۰۲۲؛ ساتون و بارتو، ۲۰۱۸).

1. Wu, Qin, Qu, Zeng & Yang
2. Zhu, Wang, Lv & Pan
3. Russell & Norvig
4. Liu, Chen, Li, Duan & Li



شکل ۱. مراحل یادگیری تقویتی

یکی از الگوریتم‌های یادگیری تقویتی عمیق، الگوریتم شبکهٔ عمیق Q[1] است. این الگوریتم یک الگوریتم پیش‌گامانه است که شبکه‌های عصبی عمیق را با یادگیری Q برای تقویت یادگیری ترکیب می‌کند. آسان‌ترین بهبودی‌ای که شبکه‌های عصبی عمیق می‌توانند به یادگیری Q ارائه دهند، عمل به‌عنوان تقریبگرهای مقادیر Q به‌جای جدول پُرهزینه Q است. یک شبکهٔ عصبی عمیق می‌تواند حالت فعلی و یک عمل ممکن را مشاهده کند و به‌صورت مستقیم مقدار Q را تقریب بزند. توانایی آن در یادگیری سیاست‌های بهینه در محیط‌های پیچیده، آن را به یک الگوریتم پُرکاربرد در این زمینه تبدیل کرده است. الگوریتم شبکهٔ عمیق Q برای آموزش از یک حافظهٔ بافر تکرار[2] استفاده می‌کند. این حافظه برای ذخیرهٔ تجربه‌های گذشته عامل استفاده می‌شود تا مجدداً از آن‌ها برای آموزش بهینه عامل استفاده شود. این ایده به‌عنوان تجربیات دوباره[3] شناخته می‌شود. الگوریتم برای اطمینان از اینکه حافظهٔ بازیابی تجربیات همیشه حاوی تجربیات جدیدتر و مرتبط با وضعیت فعلی باشد، از یک مدل صف اولین ورودی، اولین خروجی[4] استفاده می‌کند؛ به این معنا که زمانی که حافظهٔ بافر پُر می‌شود، تجربیات قدیمی‌تر از آن حذف می‌شوند و تجربیات جدیدتر به آن افزوده می‌شوند. این کار به‌عنوان حفظ یک حافظه محدود با آخرین و بهترین تجربیات برای آموزش استفاده می‌شود (مینه و همکاران، ۲۰۱۳).

پیشینهٔ تجربی پژوهش

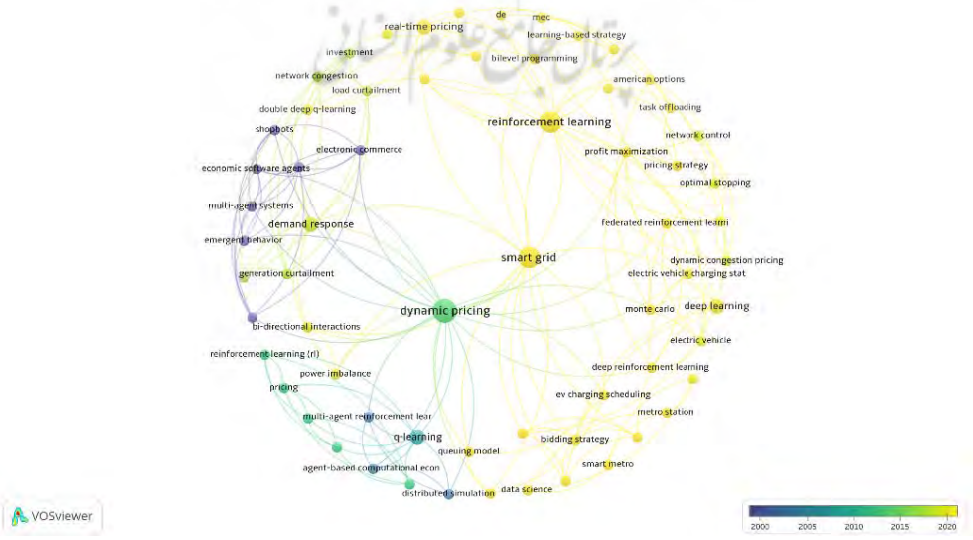
در دهه‌های اخیر، پیشرفت‌های قابل‌توجهی درزمینهٔ تکنولوژی و علم داده‌ها، به‌ویژه در حوزه‌های یادگیری ماشین و یادگیری تقویتی، ایجاد شده است. این پیشرفت‌ها امکان تحلیل و پردازش حجم وسیعی از داده‌ها را در زمان کوتاه‌تری فراهم کرده و الگوریتم‌های پیچیده‌ای را برای استخراج الگوهای پنهان و پیش‌بینی رفتارهای آینده معرفی کرده‌اند. به‌ویژه در زمینهٔ یادگیری ماشین و یادگیری تقویتی، توانایی سیستم‌ها برای یادگیری از داده‌ها و بهبود خودکار عملکرد در شرایط متغیر، موجب ایجاد راه‌کارهایی نوین برای حل مسائل پیچیده در مدیریت منابع، بهینه‌سازی فرایندها و تصمیم‌گیری شده است. یکی از روش‌های پیشرفته‌ای که در این راستا به‌کار گرفته می‌شود، قیمت‌گذاری پویاست. این روش با استفاده از الگوریتم‌های هوشمند و داده‌های لحظه‌ای، قادر است قیمت‌های بهینه را در هر زمان و با توجه به

1. Deep Q-Network
2. Replay buffer
3. Experience replay
4. FIFO

شرایط موجود تعیین کند. قیمت‌گذاری پویا نه‌تنها به شرکت‌ها این امکان را می‌دهد که از حداکثر ظرفیت منابع خود بهره‌برداری کنند، بلکه با توجه به تقاضای بازار، تغییرات اقتصادی و رفتاری مشتریان، می‌تواند درآمد را بهینه‌سازی کرده و تجربه مشتری را بهبود بخشد. این فرایند که به‌ویژه در صنایع حمل‌ونقل، هتلداری و خرده‌فروشی مورد توجه قرار گرفته، توانسته است نقش بسیار مؤثری در تغییر روش‌های سنتی قیمت‌گذاری ایفا کند و به شرکت‌ها کمک کند تا به سطح جدیدی از کارایی و رقابت‌پذیری دست یابند.

برای درک بهتر پیشرفت‌ها و تحولات مرتبط با قیمت‌گذاری پویا و یادگیری تقویتی، لازم است که ابتدا به مرور مطالعات قبلی در این زمینه پرداخته شود. برای مرور ادبیات این پژوهش، مقالات مرتبط از پایگاه‌های داده‌های معتبر شامل ScienceDirect، Springer، IEEE Xplore، wiley، Taylor & Francis Online و Google Scholar گردآوری شدند. جست‌وجو با استفاده از کلیدواژه‌های «قیمت‌گذاری پویا»، «حمل‌ونقل ریلی»، «سیستم‌های حمل‌ونقل»، «یادگیری ماشین»، «یادگیری تقویتی»، «یادگیری تقویتی عمیق» و همچنین نام الگوریتم‌های مختلف مرتبط با یادگیری تقویتی و یادگیری تقویتی عمیق انجام شد. بازهٔ زمانی مورد بررسی سال‌های ۲۰۰۰ تا ۲۰۲۴ بود. در مرحلهٔ اولیه، حدود ۲۰۰ مقاله شناسایی شد و پس از اعمال فیلترهایی نظیر ارتباط مستقیم با موضوع و استفاده از الگوریتم‌های یادگیری تقویتی در حوزه‌های مختلف، تعداد مقاله‌ها به ۴۱ مقاله کاهش یافت. این مقاله‌ها به‌عنوان مبنای بررسی ادبیات در این پژوهش استفاده شدند. ابتدا، یک تحلیل شبکه‌ای از مفاهیم کلیدی در حوزهٔ قیمت‌گذاری پویا و یادگیری تقویتی ارائه شده است که اطلاعات مفیدی مانند روابط بین مفاهیم، ارتباطات زمانی، میزان اهمیت و فراوانی در اختیار ما قرار می‌دهد. سپس، مقالات جمع‌آوری شده با دقت بیشتری بررسی و دسته‌بندی و در نهایت، پژوهش فعلی با مرتبط‌ترین مقالات در این حوزه مقایسه شدند.

شکل ۲ نشان‌دهندهٔ ارتباطات و هم‌بستگی‌های بین مفاهیم مختلف در این حوزه است و به ما کمک می‌کند تا روندهای تحقیقاتی موجود را شناسایی کرده و نقاط کمبود تحقیقات را مشخص کنیم.



شکل ۲. نمودار تحلیل هم‌بستگی مفاهیم مرتبط با قیمت‌گذاری پویا و یادگیری تقویتی

شکل ۲ نشان می‌دهد که مفاهیم قیمت‌گذاری پویا و یادگیری تقویتی به‌طور قابل‌توجهی با یکدیگر مرتبط هستند. این ارتباط قوی نشان‌دهندهٔ تمرکز عمدهٔ تحقیقات روی کاربردهای یادگیری تقویتی در تعیین قیمت‌های پویاست. رنگ زرد و سبز این خوشه‌ها نشان‌دهندهٔ این است که تحقیقات این حوزه، به‌طور پیوسته در حال گسترش و توسعه بوده‌اند و توجه زیادی را به خود جلب کرده‌اند. مفاهیم مرتبط با شبکهٔ برق نیز، به‌شدت به قیمت‌گذاری پویا و یادگیری تقویتی متصل هستند. در این زمینه، تحقیقات بسیاری روی استفاده از الگوریتم‌های یادگیری تقویتی برای مدیریت تقاضا، کاهش هزینه‌ها و بهینه‌سازی توزیع بار در شبکه‌های برق انجام شده است. این نمودار نشان می‌دهد که قیمت‌گذاری پویا با یادگیری تقویتی در حوزهٔ شبکهٔ برق، به‌شدت موردتوجه قرار گرفته و تحقیقات گسترده‌ای در این زمینه انجام‌شده است. این در حالی است که صنایع هواپیمایی و حمل‌ونقل ریلی که می‌توانند به‌طور چشمگیری از این فناوری بهره‌مند شوند، کمتر بررسی شده‌اند.

شایان ذکر است که مطالعاتی در زمینهٔ قیمت‌گذاری پویا با استفاده از یادگیری ماشین در حوزه حمل‌ونقل ریلی انجام شده‌اند که به بهبود تصمیم‌گیری و بهینه‌سازی درآمد شرکت‌های ریلی کمک می‌کنند (شان، لو، وو، ژائو و ژنگ[1]، ۲۰۲۴؛ کمندانی‌پور، یخچالی و توکلی‌مقدم[2]، ۲۰۲۳؛ یوسفی و پیشوایی[3]، ۲۰۲۲؛ جینگ، گائو، چن، ونگ و لی[4]، ۲۰۲۱). دسته‌بندی تحقیقات انجام شده در حوزه‌های کاربردی یادگیری تقویتی در شکل ۳ نمایش داده شده است. این دسته‌ها شامل کسب‌وکار الکترونیک (ناراهاری، راجو، راوی‌کومار و شاه[5]، ۲۰۰۵؛ راجو، ناراهاری و راوی‌کومار[6]، ۲۰۰۶)، خودروهای الکتریکی و ایستگاه‌های شارژ (الجعفری، جیاراج، کاثیرسان و تانیکانتی[7]، ۲۰۲۳؛ مقدم، یزدانی، ونگ، پارلولیت و شاه‌نیا[8]، ۲۰۲۰)، شبکهٔ هوشمند برق و انرژی (ژو و همکاران[9]، ۲۰۲۳؛ فرایجا و همکاران[10]، ۲۰۲۲؛ آویلا، هاردان، ژالیوا، القایلی و گویزانی[11]، ۲۰۲۲؛ جونگ[12]، ۲۰۲۲)، شبکه‌های بی‌سیم و مخابراتی (لیائو، کیائو، یو و لیو[13]، ۲۰۲۱)، حمل‌ونقل (پو، کانی، اونگ و گوه[14]، ۲۰۲۳؛ لی و اوکوسوری[15]، ۲۰۲۳؛ ساتو، سئو و فیوز[16]، ۲۰۲۲؛ گائو و همکاران، ۲۰۲۲؛ الکساندر و لینگ[17]، ۲۰۱۹؛ کالینز و توماس[18]، ۲۰۱۲؛ بوندوکس، نگوین، فیگ و آکونا آگوست[19]، ۲۰۲۰؛ گوساوی، بندلا و

1. Shan, Lv, Wu, Zhao & Zhang
2. Kamandanipour, Yakhchali & Tavakkoli Moghaddam
3. Yousefi & Pishvaee
4. Jing, Guo, Chen, Wang & Li
5. Narahari, Raju, Ravikumar & Shah
6. Raju, Narahari & Ravikumar
7. Aljafari, Jeyaraj, Kathiresan & Thanikanti
8. Moghaddam, Yazdani, Wang, Parlevliet & Shahnia
9. Xu et al.
10. Fraija et al.
11. Avila, Hardan, Zhalieva, Aloqaily & Guizani
12. Jung
13. Liao, Qiao, Yu & Liu
14. Poh, Connie, Ong, & Goh
15. Lei & Ukkusuri
16. Sato, Seo & Fuse
17. Alexander & ling
18. Collins & Thomas
19. Bondoux, Nguyen, Fiig & Acuna-Agost

داس[1]، ۲۰۰۲؛ پندی، ونگ و بویلز[2]، ۲۰۲۰) و سایر حوزه‌ها (کانگ، ژائو، چن و وی[3]، ۲۰۲۲؛ ژنگ، گان، لیانگ، جیانگ و چنگ[4]، ۲۰۲۱؛ کراشنینی‌کوا، گارسیا، ماستر و فرناندز[5]، ۲۰۱۹؛ رانا و الیویرا[6]، ۲۰۱۴) هستند.

شکل ۳. دسته بندی مقالات قیمت‌گذاری پویا

پس از بررسی دقیق و دسته‌بندی مقالات در حوزه قیمت‌گذاری پویا، مشاهده شد که الگوریتم‌های یادگیری تقویتی و یادگیری تقویتی عمیق در طیف گسترده‌ای از زمینه‌ها و صنایع برای طراحی و اجرای استراتژی‌های قیمت‌گذاری پویا به‌کار گرفته شده‌اند. این گستردگی کاربردها نشان‌دهندهٔ توانمندی و انعطاف‌پذیری این الگوریتم‌ها در مواجهه با چالش‌های قیمت‌گذاری پویا است. برای مثال، در حوزهٔ شبکهٔ برق، قیمت‌گذاری پویا به کمک یادگیری تقویتی توانسته است بهبودهای چشمگیری در مدیریت تقاضا، توزیع بار و بهینه‌سازی هزینه‌ها ایجاد کند. این روش‌ها با استفاده از داده‌های تاریخی و الگوریتم‌های پیشرفته، قادر به پیش‌بینی و تنظیم قیمت‌ها بر اساس شرایط جاری و پیش‌بینی‌های آتی هستند. با این‌حال، بررسی‌های اولیه نشان می‌دهد که استفاده از قیمت‌گذاری پویا با الگوریتم‌های یادگیری تقویتی در سایر صنایع همچون صنایع هواپیمایی و ریلی به‌مراتب کمتر مورد توجه قرار گرفته است. این امر می‌تواند به‌دلایل مختلفی از جمله پیچیدگی‌های خاص هر صنعت و نیاز به داده‌های وسیع و دقیق در این زمینه باشد.

جدول ۱ به مرور و دسته‌بندی مقالات مرتبط با استفاده از الگوریتم‌های یادگیری تقویتی و یادگیری تقویتی عمیق در طراحی و پیاده‌سازی استراتژی‌های قیمت‌گذاری پویا می‌پردازد. در این جدول، مقالات بر اساس متغیر حالت، عمل‌های اتخاذشده، تابع پاداش، نوع مدل‌سازی، الگوریتم‌های مورد استفاده، همچنین درنظر گرفتن یا درنظر نگرفتن بخش‌بندی مشتریان و روش ارزیابی نیز بررسی شده‌اند.

1. Gosavi, Bandla & Das
2. Pandey, Wang & Boyles
3. Cong
4. Zheng, Gan, Liang, Jiang & Chang
5. Krasheninnikova, García, Maestre & Fernández
6. Rana & Oliveira

جدول ۱. مرور ادبیات مقاله‌های قیمت‌گذاری پویا با روش‌های یادگیری تقویتی و یادگیری تقویتی عمیق

نویسندگان	حالت	عمل	پاداش (هدف)	مدل‌سازی	الگوریتم	بخش‌بندی مشتریان	روش ارزیابی
لی و اوکوسوری، ۲۰۲۳	اطلاعات درخواست‌ها، وسایل نقلیه فعال، میانگین دستمزد، سود سطح منطقه‌ای و شاخص زمانی	بردار ضرایب قیمت	سود	MDP	TD3		آزمایش‌های عددی
پو و همکاران، ۲۰۲۳	مرحله زمانی، اشغال پارکینگ بازیکن و اشغال پارک قابل مشاهده حریف	تخفیف	به حداکثر رساندن میزان اشغال پارکینگ، افزایش درآمد پارکینگ	POMDP	DNN		شبیه‌سازی
الجعفری و همکاران، ۲۰۲۳	ظرفیت باتری باقی‌مانده خودروی الکتریکی کاربران، زمان ورود و خروج خودروی الکتریکی.	قیمت	هزینه	MDP	MADNN		مطالعات موردی
ژو و همکاران، ۲۰۲۳	سهم بازار نیروگاه‌های مجازی در منابع انرژی تجدیدپذیر و ژنراتورهای حرارتی پس از دوره قرارداد قبلی	قیمت	رقابت با سایر نیروگاه‌های مجازی	MDP	DDPG		شبیه‌سازی
گائو و همکاران، ۲۰۲۲	تعداد صندلی‌های فروخته شده	قیمت	قیمت	MDP	Q-learning	✓	شبیه‌سازی
آویلا و همکاران، ۲۰۲۲	قیمت شبکه برق، قیمت خرید باتری، قیمت فروش باتری	قیمت	سود	MDP	DQN		آزمایش‌های عددی
زی و لوی[1]، ۲۰۲۲	سطح استفاده از سیستم ابری	قیمت	درآمد	MDP	Q-Learning, VpQ-Learning		آزمایش روی یک مجموعه داده در دنیای واقعی
ساتو و همکاران ۲۰۲۲	نرخ ورودی به گلوگاه، زمان انتظار و هزینه	افزایش یا کاهش عوارض	زمان انتظار بر اساس بازه زمانی و گلوگاه، میانگین زمان انتظار در تمام گلوگاه‌هایی که عوارض در آن‌ها دریافت می‌شود	MDP	DDPG		آزمایش‌های عددی
کانگ و همکاران، ۲۰۲۲	ارزش درک شده توسط کاربر	قیمت	سود و هزینه	MDP	Q-learning		شبیه‌سازی

1. Xie and Lui

نویسندگان	حالت	عمل	پاداش (هدف)	مدل‌سازی	الگوریتم	بخش‌بندی مشتریان	روش ارزیابی
فرایجا و همکاران، ۲۰۲۲	میانگین ساعتی کل مصرف انرژی	قیمت	سود	MDP	PPO		شبیه‌سازی
جونگ، ۲۰۲۲	تقاضای کل، نسبت شارژ انرژی به حداکثر ظرفیت سیستم ذخیره انرژی، زمان، هزینه هر واحد انرژی	مقادیر شارژ یا تخلیه	قیمت هر واحد برق ضرب در مقادیر شارژ یا تخلیه تقسیم بر چهار	MDP	DQN		-
ژو و همکاران، ۲۰۲۲	قیمت تسویه، قیمت خرده‌فروشی، مصرف انرژی	قیمت	سود	MDP	DQN		شبیه‌سازی
وان، کین، یانگ و کانگ[1]، ۲۰۲۱	تقاضای انرژی، مصرف انرژی، شاخص زمان	قیمت	رفاه اجتماعی	MDP	Q-learning		مطالعات موردی
دو و همکاران[2]، ۲۰۲۱	میانگین نسبت سیگنال به نویز بین هر ماینر و سرور محاسباتی لبه موبایل، نرخ ارسال هر ماینر در انتشار بلوک، مجموعه تخصیص شماره کانال، تخصیص بلوک منابع محاسباتی	قیمت	متوسط سودمندی منطقی برای همه ماینرها	MDP	A3C		شبیه‌سازی
ژنگ، وانگ، اوجلا و بات[3]، ۲۰۲۱	تقاضای برق و مصرف برق واقعی کاربر	قیمت	مبلغ قبض کاربر	MDP	Q-learning		شبیه‌سازی
باقرپور، مزینی و بدنوا[4]، ۲۰۲۱	اطلاعات وابسته به زمان، مقدار هدف نهادهای سرویس‌دهندهٔ بار، تقاضاهای مشتریان	قیمت	هزینه	MDP	Deep Contextual Bandits Algorithm		شبیه‌سازی
لیو، ژنگ و گویی[5]، ۲۰۲۱	پروفایل تقاضای بار از طرف مشتریان، قیمت عمده‌فروشی برق	قیمت	هزینه‌های مشتریان، سود خرده‌فروشان و پاداش جامع	MDP	DQN, A2C, Q-learning		آزمایش‌های عددی
لیائو و همکاران، ۲۰۲۱	شرایط شبکه، تقاضای منابع محاسباتی تجهیزات کاربر	قیمت	درآمد	MDP	Q-learning		شبیه‌سازی

1. Wan, Qin, Yu, Yang & Kang
2. Du et al.
3. Zhang, Wang, Aujla & Batth
4. Bagherpour, Mozayani & Badnava
5. Liu, Zhang & Gooi

نویسندگان	حالت	عمل	پاداش (هدف)	مدل‌سازی	الگوریتم	بخش‌بندی مشتریان	روش ارزیابی
ژنگ و همکاران، ۲۰۲۱	تعداد محصولات قدیمی باقی‌مانده در ابتدای هر دوره، تعداد محصولات جدید موجود در ابتدای هر دوره	قیمت	سود خالص	MDP	Q-Learning		شبیه‌سازی
عبدالرحمان و ژووانگ[1]، ۲۰۲۰	استفاده و کیفیت خدمات تمامی امکانات شارژ در زیرساخت شارژ خودروهای برقی پلاگین	قیمت	استفاده از تسهیلات شارژ	MDP	TD3		آزمایش‌های عددی
چن، لی، چن و لی[2]، ۲۰۲۰	زمان رسیدن کار	قیمت	سود	MDP	policy gradient		شبیه‌سازی
کیو، یه، پاپاداسکالوپولوس و استرباک[3]، ۲۰۲۰	قیمت‌های عمده‌فروشی بازار، تقاضای EV انعطاف‌ناپذیر و تقاضای خالص EV انعطاف‌پذیر	قیمت	سود کلی	MDP	PDDPG		مطالعات موردی
مقدم و همکاران، ۲۰۲۰	تقاضای خودروهای الکتریکی در هر بازه زمانی	قیمت	سود		Online RL Model Using AHC		شبیه‌سازی
پندی و همکاران، ۲۰۲۰	تعداد وسایل نقلیه در سلول، متعلق به یک کلاس در زمان	قیمت	درآمد، کل زمان سفر	POMDP	VPG, PPO		شبیه‌سازی
لو و همکاران[4]، ۲۰۲۰	ماه جاری سال، ترجیح مصرف کاربر، پیکربندی نرخ واحد انرژی	قیمت	رضایت مصرف کننده و هزینه قبض برق	MDP	Kernel Approximator-Based Batch Q-learning Algorithm Combined With Sampling And Data Processing Methods		آزمایش‌های عددی
ونگ، بی و ژنگ[5]، ۲۰۱۹	تقاضای شارژ باقی‌مانده و زمان پارک خودروی برقی	قیمت و نرخ شارژ	سود ایستگاه شارژ، بهره‌مندی مشتریان خودروهای برقی، رفاه اجتماعی	MDP	Hyperopia SARSA Algorithm		شبیه‌سازی

1. Abdalrahman & Zhuang
2. Chen
3. Qiu, Ye, Papadaskalopoulos & Strbac
4. Lu
5. Wang, Bi & Zhang

نویسندگان	حالت	عمل	پاداش (هدف)	مدل‌سازی	الگوریتم	بخش‌بندی مشتریان	روش ارزیابی
کراشنینی‌کوا و همکاران، ۲۰۱۹	درآمد مشتری، نگه داشت، نوع بخش‌بندی مشتری، قیمت بیمه برای مشتری	قیمت	درآمد	MDP, CMDP	VQQL		آزمایش‌های عددی
الکساندر و لینگ، ۲۰۱۹	تعداد بلیت‌های موجود، زمان	قیمت	درآمد	MMDP	Sarsa and Sarsa with eligibility traces	✓	آزمایش‌های عددی
لو، هانگ و ژنگ[1]، ۲۰۱۸	تقاضای انرژی و مصرف انرژی مشتریان	قیمت	مبلغ قبض مشتری	MDP	Q-learning		شبیه‌سازی
پندی و بویلز[2]، ۲۰۱۸	تعداد خودروهای هر کلاس در هر سلول در شروع زمان به‌روزرسانی عوارض	نرخ عوارض	درآمد	MDP	VFA		مقایسه الگوریتم VFA با سایر الگوریتم های ابتکاری
کیم، ژانگ، ون در شاار و لی[3]، ۲۰۱۵	تقاضای بار، دوره جاری، هزینه	قیمت	هزینه	MDP	Q-learning, PDS		آزمایش‌های عددی
رانا و الیویرا، ۲۰۱۴	ظرفیت باقی‌مانده	قیمت	درآمد	MDP	Q(λ)		شبیه‌سازی
کالینز و توماس، ۲۰۱۲	مرحله فعلی، قیمت فعلی حریف، تعداد صندلی‌هایی که آن‌ها فروخته‌اند	قیمت	درآمد	Markov game	Q-learning, SARSA and Monte Carlo		مطالعات موردی
راجو و همکاران، ۲۰۰۶	تعداد درخواست‌های عقب‌افتاده در صف یک، تعداد درخواست‌های عقب‌افتاده در صف دو، سطح موجودی انبار	قیمت	هزینه موجودی، هزینه سفارش مجدد	MDP	Q-learning	✓	شبیه‌سازی
ناراهاری و همکاران، ۲۰۰۵	سطح موجودی انبار، تعداد مشتریان در صف سفارش محصول	قیمت	سود	MDP	Q-learning	✓	شبیه‌سازی

1. Lu, Hong & Zhang
2. Pandey & Boyles
3. Kim, Zhang, Van Der Schaar & Lee

با مرور ادبیات قیمت‌گذاری پویا می‌توان نتیجه گرفت که اکثر پژوهش‌های انجام‌شده در مسئله قیمت‌گذاری پویا در حمل‌ونقل ریلی با مدل‌سازی ریاضی، برنامه‌ریزی پویا، تکنیک‌های مبتنی بر نظریهٔ بازی‌ها یا ترکیبی از روش‌های یادگیری ماشین و بهینه‌سازی برای اهداف حداکثرسازی سود و ایجاد تعادل میان عرضه و تقاضا انجام شده است و در حوزهٔ قیمت‌گذاری پویا در سیستم‌های حمل‌ونقل ریلی با استفاده از روش‌های یادگیری تقویتی و یادگیری تقویتی عمیق تحقیقات کمی وجود دارد (ساهاران و همکاران، ۲۰۲۰). جریان چشمگیری از ادبیات درباره قیمت‌گذاری پویا و یادگیری از جامعه علوم کامپیوتر پدید آمده است. به‌طورکلی، تمرکز این مقالات فقط بر ارائه یک تحلیل ریاضی از عملکرد سیاست‌های قیمت‌گذاری نیست، بلکه بر طراحی یک مدل واقع‌بینانه برای بازارهای الکترونیک و سپس استفاده از تکنیک‌های یادگیری ماشین و یادگیری تقویتی تمرکز دارند. یک مزیت این رویکرد این است که می‌توان تعداد زیادی از پدیده‌هایی که بر عرضه تأثیر می‌گذارند، مانند رقابت، نوسانات تقاضا و رفتار استراتژیک خریداران را مدل‌سازی کرد؛ همچنین، روش‌های سنتی نمی‌توانند به‌خوبی روش‌های یادگیری ماشین و یادگیری تقویتی این مسائل را مدل‌سازی کنند و با پیشرفت تکنولوژی، افزایش حجم داده‌ها و همچنین پیچیدگی بیشتر مسائل، نیاز به روش یا روش‌هایی است که بتوانند مدل‌های واقع‌بینانه‌تری نسبت به روش‌های سنتی بسازند و قیمت‌های بلیت را به‌صورت پویا با دقت و عملکرد بهتری تخمین بزنند و به هر مشتری قیمت اختصاصی پیشنهاد دهند (دن بوئر[1] ، ۲۰۱۵).

در سال‌های اخیر، الگوریتم‌های یادگیری تقویتی و یادگیری تقویتی عمیق به‌دلیل مزایای برجسته‌ای که در حل مسائل قیمت‌گذاری پویا دارند، به‌طور گسترده‌ای در پژوهش‌های این حوزه مورد استفاده قرار گرفته‌اند. همان‌طور که در شکل ۲ مقاله مشاهده می‌شود، بسیاری از مقالات اخیر از این روش‌ها بهره برده‌اند که نشان‌دهندهٔ برتری این روش‌ها نسبت به سایر روش‌های یادگیری ماشین و روش‌های سنتی است. با این حال، باید به این نکته نیز توجه داشت که بسیاری از این مقالات در حوزه صنعت برق و انرژی متمرکز بوده‌اند. این تمرکز عمدتاً به‌دلیل اهمیت بالای این صنعت نسبت به صنایع دیگر و نوع و ویژگی های مشتریان و به تناسب آن ذات مسئله قیمت گذاری پویا در مقایسه با حمل‌ونقل ریلی است. دلیل دیگر نیز فراهم بودن داده‌های مناسب و کافی در برخی صنایع نسبت به صنعت حمل‌ونقل ریلی است. در صنعت حمل‌ونقل هوایی نیز، مقالات متعددی از روش‌های یادگیری تقویتی و یادگیری تقویتی عمیق بهره برده‌اند. اما بررسی‌های ما نشان می‌دهد که در حوزه حمل‌ونقل ریلی، مقالات اندکی به موضوع قیمت‌گذاری پویا پرداخته‌اند و مقاله‌ای که از یادگیری تقویتی در این صنعت استفاده کرده باشد، یافت نشد. همچنین، در ادبیات مقالات فارسی نیز، پژوهشی که به قیمت‌گذاری پویا با استفاده از روش‌های یادگیری تقویتی در سیستم‌های حمل‌ونقل ریلی بپردازد، وجود ندارد.

نوآوری این پژوهش استفاده از روش‌های یادگیری تقویتی عمیق برای مسئله قیمت‌گذاری پویا در سیستم‌های حمل‌ونقل ریلی است. یادگیری ماشین و یادگیری تقویتی یکی از مهم‌ترین موضوعات پژوهشی در علوم کامپیوتر و مهندسی است که می‌تواند در بسیاری از رشته‌های مختلف به کار رود. این حوزه مجموعه‌ای از الگوریتم‌ها، روش‌ها و ابزارهایی را ارائه می‌دهد که امکان تجسم هوش در ماشین‌ها را فراهم می‌سازد. قدرت یادگیری ماشین و یادگیری

1. Den Boer

تقویتی در ابزارهای مدل‌سازی نهفته است که از طریق یک فرایند یادگیری که می‌تواند با مجموعه‌ای از داده‌ها که یک مسئله خاص را توصیف می‌کنند، آموزش داده شود و به داده‌های دیده نشده به شیوه‌ای درست پاسخ دهد (کاک و ارسلان[1]، ۲۰۲۱). یادگیری تقویتی به‌دلیل قابلیت اجرای جست‌وجوی آزمایشی و مبتنی بر خطا، امکان بهبود عملکرد عامل‌های یادگیری ماشین را به‌وسیله فراتر رفتن از محدودیت‌های یادگیری نظارت‌شده و بدون نظارت فراهم می‌آورد. این رویکرد قادر است در مواجهه با سناریوهای پیچیده‌تر، از جمله مواردی که اقدامات تنها بر پاداش فوری تأثیر نمی‌گذارند، بهبودهای معنی‌داری ارائه دهد که با استفاده از روش‌های دیگر یادگیری ماشین دست‌نیافتنی هستند (نیان، لیو و هووانگ[2]، ۲۰۲۰).

نوآوری دیگر این پژوهش، قیمت‌گذاری پویا برای کلاس‌های مختلف مشتریان است که در چند سال اخیر موردتوجه محققان قرار گرفته است. در حوزهٔ کسب‌وکار الکترونیک، مطالعاتی مانند ناراهاری و همکاران (۲۰۰۵) و راجو و همکاران (۲۰۰۶) از تکنیک‌های یادگیری تقویتی برای مدل‌های قیمت‌گذاری پویا و بخش‌بندی مشتریان در بازارهای خرده‌فروشی استفاده کرده‌اند. برخی از مطالعات دیگر مدل‌های قیمت‌گذاری را با در نظر گرفتن مشتریان کوته‌بین و استراتژیک ایجاد کردند و تأثیر نسبت‌های مختلف مشتری استراتژیک بر استراتژی‌های قیمت‌گذاری را موردبحث قرار دادند. دو و چن[3] (۲۰۱۷) در صنعت مد و پوشاک، تن[4] (۲۰۱۸) و گائو و همکاران (۲۰۲۲) در صنعت حمل‌ونقل هوایی، زنگ و ژنگ[5] (۲۰۱۵) و ژائو و همکاران (۲۰۱۷) نیز به صنعت خاصی اشاره نداشته‌اند. مقاله‌های ذکرشده بخش‌های مشتریان را به‌طور ویژه مشخص کرده‌اند و قابل‌تعمیم دادن به بخش‌های بیشتر نیست. در پژوهش دیگری الکساندر و لینگ (۲۰۱۹) مدلی را برای قیمت‌گذاری پویای بخش‌های مشتریان در صنعت هواپیمایی معرفی می‌کند که به بخش‌های بیشتر تعمیم‌پذیر است. مقالاتی مانند ژیائوکیانگ، لانگ و جین[6] (۲۰۱۷) و یان و همکاران[7] (۲۰۱۹)، به قیمت‌گذاری پویای بخش‌های مشتریان در صنعت حمل‌ونقل ریلی پرداخته‌اند. در این پژوهش، مشتریان به چهار بخش مختلف (اکونومی، بیزینس، ویژه و هتل) تقسیم‌بندی می‌شوند. مشتریان متقاضی هر کلاس خدماتی به‌عنوان یک بخش جداگانه در نظر گرفته و مدل‌سازی با در نظر گرفتن کلاس خدماتی به‌عنوان متغیر حالت انجام می‌شود. این نوع مدل‌سازی امکان به‌دست آوردن تعداد بهینه بلیت‌های فروخته شده برای هر کلاس خدماتی را فراهم می‌کند.

مدل مفهومی

یک فرایند یادگیری تقویتی که ویژگی مارکوفی داشته باشد را فرایند تصمیم‌گیری مارکوف می‌گویند و اگر فضاهای حالت و عمل محدود باشند، آن را فرایند تصمیم‌گیری مارکوف محدود می‌نامند. یک فرایند تصمیم‌گیری مارکوف محدود خاص با حالت و مجموعه اقدامات آن و با پویایی یک مرحله‌ای محیط تعیین می‌شود (ساتون و بارتو، ۲۰۱۸). در این تحقیق، مسئله قیمت‌گذاری پویا به‌عنوان یک فرایند تصمیم‌گیری مارکوف فرموله می‌شود؛ زیرا قیمت‌گذاری یک مسئله

1. Koc & Arslan
2. Nian, Liu & Huang
3. Du & Chen
4. Tan
5. Zeng & Zhang
6. Xiaoqiang, Lang & Jin
7. Yan et al.

تصمیم‌گیری در لحظه و در یک محیط تصادفی است. هدف این تحقیق تقریب سیاست قیمت‌گذاری است که درآمد را برای فروش موجودی معینی از محصولات در یک مهلت مشخص به حداکثر می‌رساند. در اینجا، یک مدل کلی برای تعداد ثابتی از محصولات یا خدمات یکسان ارائه شده است. اجزای اصلی فرایند تصمیم‌گیری مارکوف عبارت‌اند از:

- فضای حالت: S نشان‌دهندهٔ حالت سیستم شامل متغیرهای ظرفیت باقی‌مانده، کلاس بلیتی که مشتری قصد خریداری آن را دارد و اختلاف تاریخ خرید بلیت تا روز حرکت قطار است. متغیر ظرفیت باقی‌مانده بین صفر تا ۴۰۰ و متغیر اختلاف تاریخ خرید بلیت تا روز حرکت قطار بین صفر تا ۳۰ قرار دارد. همچنین متغیر کلاس بلیتی که مشتری قصد خریداری آن را دارد شامل ۴ کلاس خدماتی اکونومی، بیزینس، ویژه و هتل است.
- افق زمانی $t \in T\{0,1, \dots, m\}$ مجموعه‌ای از زمان‌های گسستهٔ متناهی است که در آن یک قیمت به یک مشتری پیشنهاد داده می‌شود.
- s_t وضعیت سیستم را در زمان t نشان می‌دهد که شامل ظرفیت باقی‌مانده، روز باقی‌مانده تا حرکت و کلاس بلیتی است که مشتری قصد خریداری آن را دارد.
- $A(s)$ نشان‌دهندهٔ مجموعه قیمت‌هایی است که فروشنده می‌تواند زمانی که سیستم در وضعیت s وجود دارد تعیین کند، $\alpha \in A(s)$ قیمتی برای وضعیت s است.
- احتمالات انتقال: $P_t(s_{t+1} | s_t, a_t)$ احتمال قرار داشتن در وضعیت s در زمان $t+1$ است، با توجه به اینکه در زمان t، در وضعیت s قرار داشته باشیم و قیمت a تنظیم شده باشد. با توجه به بدون مدل بودن الگوریتم شبکهٔ عمیق Q نیازی به تعریف احتمالات انتقال نیست و عامل از طریق تعامل مستقیم با محیط، سیاست بهینه را یاد می‌گیرد.
- R تابع درآمدی است که برای هر مرحله تصمیم‌گیری، در هر حالت و اقدام، یک عدد واقعی $r(s_t, a_t)$ را برای $s_t \in X$، $t \in T$ و $a_t \in A(s)$تعریف می‌کند که درآمد فوری مورد انتظار به‌دست‌آمده برای اجرای قیمت a را زمانی که سیستم در حالت s در زمان t باقی‌مانده است، مشخص می‌کند. در واقع درآمد تابعی از قیمت پیشنهادی توسط عامل است. درصورتی که مشتری تصمیم به خرید بلیت قطار با قیمت پیشنهادی توسط عامل گیرد، پاداشی به اندازه قیمت پیشنهادی به عامل داده می‌شود و اگر مشتری تصمیم به ترک سیستم بدون خرید بلیت گیرد، پاداش برابر صفر خواهد بود.

هدف، به حداکثر رساندن کل درآمد مورد انتظار ارائه‌شده در رابطه ۱ است که در آن $s \in X$ و E_π مقدار مورد انتظار با توجه به سیاست π هستند (ساتون و بارتو، ۲۰۱۸).

$$V^\pi(x_1) = E_\pi(r(s_1, a_1) + r(s_2, a_2) + \cdots + r(s_m, a_m)|s_1) \quad \text{(رابطه ۱)}$$

یک سیاست تابعی است که $\pi : s \to a_t$ قیمتی را که باید بر حالت سیستم مشخص می‌کند. در چارچوب فرایند تصمیم‌گیری مارکوف $V_t^\pi(s_t)$ نشان‌دهندهٔ پاداش کل مورد انتظار هنگام شروع از حالت s_t و پیروی از سیاست π است (یعنی $a_t = \pi(s_t)$ که در آن $\pi(s)$ نشان‌دهندهٔ اقدام انتخاب‌شده در حالت زمانی s است که از سیاست π پیروی می‌کنیم).

$Q_t^{\pi}(s_t, a_t)$ نشان‌دهندهٔ پاداش کل مورد انتظار تنزیل‌شده هنگام شروع از حالت s_t، انجام اقدام a_t در هنگام پیروی سیاست π است. یعنی Q_t^{π} تابع مقدار حالت ـ اقدام برای سیاست π در زمان t است. رابطهٔ ۲ نشان‌دهندهٔ رابطه بین $Q_t^{\pi}(s_t, a_t)$ و $V_t^{\pi}(s_t)$ است (ساتون و بارتو، ۲۰۱۸).

$$Q_t^{\pi}(s_t, a_t) = \sum_{s_{t+1} \in X} P_t(s_{t+1}|x_t, a_t)[r(s_t, a_t, s_{t+1}) + \gamma V_{t+1}^{\pi}(s_{t+1})] \quad \text{رابطه ۲)}$$

که در آن γ ضریب تخفیف است، $0 < \gamma < 1$ ازنظر تابع بهینه بلمن، رابطه ۳ برای s_t دلخواه، $Q_t^*(s_t, a_t)$ تابع مقدار بهینه برای هر جفت حالت ـ اقدام است. برای $t = \{1,2, \dots, m\}$ داریم:

$$Q_t^*(s_t, a_t) = \sum_{x_{t+1} \in X} P_t(s_{t+1}|s_t, a_t) \Big[r(s_t, a_t, s_{t+1}) + \gamma \max_{a_{t+1} \in A(s_{t+1})} Q_{t+1}^*(s_{t+1}, a_{t+1}) \Big] \quad \text{رابطه ۳)}$$

روش‌شناسی پژوهش

با توجه به این موضوع که مدل‌های یادگیری تقویتی برای تحلیل بیش‌ازحد پیچیده هستند، برای تحلیل مدل‌ها و به‌دست آوردن بینش‌هایی در مورد رفتار استراتژی‌های قیمت‌گذاری پویا، می‌بایست از ابزار شبیه‌سازی استفاده کنیم. آزمایش‌های عددی و شبیه‌سازی به ما اجازه می‌دهند تا عملکرد الگوریتم‌های مختلف یادگیری تقویتی را در شرایط و وضعیت‌های مختلف بررسی و بهترین رویکرد قیمت‌گذاری پویا را شناسایی کنیم. این آزمایش‌ها انجام می‌شوند تا مطمئن شویم که مدل‌های ما به‌درستی عمل می‌کنند و در صورت نیاز، تنظیمات و پارامترهای آن‌ها را بهبود دهیم.

ما یک شرکت حمل‌ونقل ریلی را در نظر می‌گیریم که در آن تقسیم‌بندی مشتریان بر اساس تمایل به پرداخت انجام می‌شود. فرض می‌شود مشتریان این شرکت حمل‌ونقل ریلی به ۴ گروه مختلف که هرکدام متقاضی یک کلاس خدماتی خاص هستند، تقسیم‌بندی شده‌اند. در ابتدای شبیه‌سازی فرض می‌شود که ظرفیت قطار برابر با ۴۰۰ و مشتریان از ۳۰ روز مانده به حرکت می‌توانند به خریداری بلیت قطار اقدام کنند. در هرروز تعدادی مشتری با تابع توزیع پواسون با پارامتر $\lambda = ۱۷$ وارد سیستم می‌شوند. سپس، یکی از چهار کلاس خدماتی (اکونومی، بیزینس، ویژه، و هتل) به‌صورت تصادفی و به مشتری تخصیص داده می‌شود. این تخصیص براساس درصدهای ارائه‌شده در جدول ۲ انجام می‌گیرد. برای مثال، زمانی که یک مشتری وارد سیستم می‌شود، با احتمال ۰/۴۷۸ متقاضی بلیت کلاس اکونومی، با احتمال ۰/۳۲۶ متقاضی بلیت کلاس بیزینس، با احتمال ۰/۱۴۵ متقاضی بلیت کلاس ویژه و با احتمال ۰/۰۵۱ متقاضی بلیت کلاس هتل خواهد بود. برای محاسبهٔ میزان تمایل به پرداخت، از توزیع‌های یکنواخت[1] استفاده می‌شود (الکساندر و لینگ، ۲۰۱۹). به‌طور مثال، مشتریان کلاس اکونومی تمایل به پرداختی بین ۳۰ تا ۵۰ واحد پولی دارند که این مقدار به‌صورت تصادفی از توزیع یکنواخت (۳۰, ۵۰) برای هر مشتری در این کلاس انتخاب می‌شود. به همین ترتیب، برای

1. Uniform

دیگر کلاس‌های خدماتی نیز مقادیر تمایل به پرداخت از توزیع‌های یکنواخت مشخص‌شده در جدول ۲ انتخاب می‌شوند. انعطاف‌پذیری مشتریان نسبت به قیمت نیز بر اساس توزیع نرمال[1] مدل‌سازی شده است (الکساندر و لینگ، ۲۰۱۹). به‌طور مثال، انعطاف‌پذیری مشتریان کلاس اکونومی با میانگین ۳۵۰ و انحراف معیار ۳۰ طبق توزیع نرمال (۳۵۰, ۳۰) تعیین می‌شود. این مقدار نشان‌دهندهٔ میزان انعطاف مشتری در برابر تغییرات قیمتی است. مشتریان متقاضی کلاس اکونومی کمترین میزان تمایل به پرداخت و مشتریان داخل کلاس هتل بیشترین میزان تمایل به پرداخت مطابق جدول۲ را دارند. درواقع به هر مشتری ۳ ویژگی به‌عنوان حالت تخصیص داده می‌شود که شامل کلاس خدماتی بلیتی که آن مشتری قصد خرید دارد، ظرفیت باقی‌مانده در هنگامی‌که مشتری اقدام به خرید می‌کند و تعداد روز باقی‌مانده به حرکت در هنگام خرید است.

جدول ۲. میزان تمایل به پرداخت و انعطاف پذیری مشتریان

کلاس مشتری	درصد مشتریان	تابع توزیع میزان تمایل به پرداخت	تابع توزیع انعطاف‌پذیری مشتری
مشتری کلاس اکونومی	۴۷/۸	Uniform (۳۰,۵۰)	Normal (۳۵۰,۳۰)
مشتری کلاس بیزینس	۳۲/۶	Uniform (۵۰,۸۰)	Normal (۷۰۰,۵۰)
مشتری کلاس ویژه	۱۴/۵	Uniform (۸۰,۱۵۰)	Normal (۱۰۰۰,۷۰)
مشتری کلاس هتل	۵/۱	Uniform (۱۵۰,۲۵۰)	Normal (۱۴۰۰,۹۰)

پس از مشخص شدن حالت سیستم، عامل با استفاده از رویکرد حریصانه ε یک قیمت را از فهرست {۳۲۰، ۴۲۰، ۵۲۰، ۶۲۰، ۷۲۰، ۸۲۰، ۹۲۰، ۱۰۲۰، ۱۱۲۰، ۱۲۲۰، ۱۳۲۰، ۱۴۲۰، ۱۵۲۰، ۱۶۲۰، ۱۷۲۰، ۱۸۲۰، ۱۹۲۰، ۲۰۲۰} انتخاب می‌کند و به هر حالت تخصیص می‌دهد. شایان ذکر است که این مجموعه قیمت‌ها بر اساس مشورت‌های گسترده با کارشناسان و خبرگان حوزه و همچنین با اجرای مکرر شبیه‌سازی‌های عملی، به‌دست آمده است. معیار تعیین مجموعه قیمت‌های ممکن بر اساس تحلیل‌های تجربی و نتایج به‌دست‌آمده از این شبیه‌سازی‌ها انتخاب شده است تا بهترین عملکرد و سازگاری با واقعیت‌های موجود در صنعت حمل‌ونقل ریلی ایران تضمین شود.

در ادامه با توجه به کلاس بلیت تخصیص داده‌شده به مشتری، میزان تمایل به پرداخت مشتری توسط تابع توزیع تخصیص داده‌شده در جدول ۲ به همان کلاس محاسبه می‌گردد. سپس، میزان تمایل به پرداخت مشتری با قیمت انتخاب‌شده توسط عامل مقایسه می‌شود. اگر قیمت پیشنهادی توسط عامل پایین‌تر از میزان تمایل به پرداخت مشتری باشد، پاداشی برابر با قیمت به عامل داده می‌شود و ظرفیت بلیت‌های باقی‌مانده به میزان یک واحد کاهش پیدا می‌کند. درصورتی‌که قیمت از میزان تمایل به پرداخت مشتری بیشتر باشد، به مشتری علاوه بر میزان تمایل به پرداخت یک مقدار دیگری تحت عنوان انعطاف‌پذیری مشتری در برابر قیمت تخصیص داده می‌شود. تحت این شرایط اختلاف قیمت و میزان تمایل به پرداخت مشتری محاسبه می‌شود و سپس با استفاده از رابطه ۴ احتمال خرید کردن مشتری به‌دست می‌آید.

1. Normal

رابطه ۴)
$$\text{احتمال خرید} = \left(-\frac{\text{قیمت پیشنهادی توسط عامل} - \text{میزان تمایل به پرداخت مشتری}}{\text{انعطاف پذیری مشتری}}\right) + ۱$$

رابطه ۴ بر اساس مبانی علمی و تجربی در حوزهٔ رفتار مصرف‌کننده و مدل‌های تصمیم‌گیری توسعه یافته است. این رابطه به‌طور خاص بیان می‌کند که هرچه قیمت پیشنهادی از میزان تمایل بـه پرداخـت مشـتری بیشـتر فاصـله بگیـرد، احتمال خرید کاهش می‌یابد. در این مدل، اختلاف بین قیمت و میزان تمایل به پرداخت مشتری به‌صورت خطی در نظـر گرفته شده و انعطاف‌پذیری مشتری به‌عنوان یک عامل تعدیل‌کننده وارد معادله می‌شود (الکساندر و لینـگ، ۲۰۱۹). ایـن روش به‌طور گسترده‌ای در مدل‌سازی رفتار مشتریان در شرایط مختلف استفاده می‌شود و به واقعیت‌های مشاهده‌شـده در رفتار مشتریان پاسخ می‌دهد. در این مدل، تصمیم‌گیری مشتری نه‌تنها بر اساس قیمت مطلق، بلکه بر مبنای تفاوت بـین قیمت پیشنهادی و ارزشی که مشتری برای کالا قائل است، انجام می‌گیرد. این رویکرد از جمله روش‌های شناخته‌شده در اقتصاد رفتاری و نظریه‌های تصمیم‌گیری است که برای تحلیل رفتار مشتریان در شرایط عدم قطعیت و ترجیحات متغیـر استفاده می‌شود.

مشتری با احتمال به‌دست‌آمده از رابطه ۴ تصمیم به خرید بلیت یا ترک سیستم می‌کند. اگر مشتری تصمیم به ترک سیستم بگیرد، عامل پاداش صفر دریافت می‌کند و ظرفیت بلیت‌های باقی‌مانده تغییر نمی‌کند. پس از اتمام بلیـت‌فروشـی به مشتریان واردشده به سیستم در روز جاری، شبیه‌سازی روز بعد آغاز می‌شود و از روز باقی‌مانده به حرکت به اندازهٔ یک واحد کم می‌شود. شبیه‌سازی تا زمانی ادامه پیدا می‌کند که به روز حرکت قطار برسیم. درصورتی‌که به روز حرکـت قطـار برسیم و هنوز ظرفیت قطار به صفر نرسیده باشد، بلیت‌فروشی در روز حرکت تا زمان حرکت قطار ادامه می‌یابـد و سـپس شبیه‌سازی با پایان می‌رسد. اگر تمامی بلیت‌ها قبل از رسیدن به روز حرکت به اتمام برسد یا به‌عبارت‌دیگر تمام ظرفیـت قطار زودتر از موعد فروش رود، شبیه‌سازی تا روز حرکت ادامه پیدا می‌کند؛ اما عامل بلیتی برای فروش در دسترس ندارد. بنابراین، عامل برای مشتریانی که پس از اتمام ظرفیت قطار به سیستم وارد می‌شوند، پاداش صفر دریافت می‌کند.

هایپرپارامترهای یادگیری تقویتی معمولاً از طریق فرایندهای تجربی و آزمایشی به دست می‌آیند. این فرایند شـامل تنظیم مکرر پارامترها و ارزیابی عملکرد مدل در شرایط مختلف است تا بهترین ترکیب برای دستیابی بـه نتـایج مطلـوب شناسایی شود. شبیه‌سازی را برای ۲۰۰۰۰ تکرار با پارامترهای جدول ۳ اجرا می‌کنیم.

جدول ۳. پارامترهای شبیه‌سازی

نام پارامتر	مقدار پارامتر	نام پارامتر	مقدار پارامتر
تعداد بلیت‌های موجود در ابتدای دوره	۴۰۰	مقدار اولیه ε	۱
زمان آغاز بلیت‌فروشی (روز مانده به حرکت)	۳۰	حداقل مقدار ε	۰/۱
کلاس‌های مشتریان	اکونومی، بیزینس، ویژه و هتل	نرخ کاهش ε	۰/۹۹۵
تعداد قسمت‌های شبیه‌سازی	۲۰۰۰۰	ضریب تخفیف γ	۱
نرخ یادگیری α	۰/۱	روش کاوش	ε حریصانه و کاهش ε

در این تحقیق، به‌منظور حل مسئله قیمت‌گذاری پویا، از الگوریتم شبکهٔ عمیق Q که یکی از الگوریتم‌های پرکاربرد یادگیری تقویتی عمیق است، بهره گرفته شده است. شایان ذکر است که در حوزه یادگیری تقویتی، انتخاب الگوریتم مناسب بستگی به ویژگی‌های خاص هر مسئله دارد. هیچ الگوریتمی به‌طور کلی برتر از سایرین نیست و هر کدام ممکن است در شرایط و مسائل مختلف عملکرد متفاوتی داشته باشند. الگوریتم شبکهٔ عمیق Q به‌دلیل ساختار شبکهٔ عصبی عمیق خود و قابلیت یادگیری از داده‌های پیچیده و حجیم، به‌ویژه در مواجهه با محیط‌های پیچیده و نیازمند تصمیم‌گیری‌های متوالی، به‌عنوان یکی از پرکاربردترین و موفق‌ترین روش‌ها شناخته می‌شود. علاوه‌بر مزایای فنی، شبکهٔ عمیق Q یکی از الگوریتم‌های پایه‌ای در حوزهٔ یادگیری تقویتی عمیق است که به‌طور گسترده در مقالات و پژوهش‌های اخیر مورد استفاده قرار گرفته است. بررسی‌ها نشان می‌دهد که پژوهش‌های متعددی در زمینهٔ قیمت‌گذاری پویا از این الگوریتم به‌دلیل مزایای قابل‌توجه آن بهره برده‌اند.

مدل شبکهٔ عصبی مورد استفاده در این پژوهش از نوع شبکهٔ عصبی پرسپترون چندلایه[1] و شبکهٔ عصبی کاملاً متصل[2] است که در چارچوب الگوریتم شبکهٔ عمیق Q پیاده‌سازی شده است. این مدل شامل چهار لایه چگال است. لایهٔ ورودی مدل اطلاعات مربوط به وضعیت فعلی عامل را دریافت می‌کند. این لایه به اولین لایهٔ چگال که دارای ۶۴ نورون است، متصل می‌شود. پس از آن، دو لایهٔ میانی، هر یک با ۱۲۸ نورون، نقش کلیدی در یادگیری الگوهای پیچیده و استخراج ویژگی‌های انتزاعی از داده‌های ورودی ایفا می‌کنند. در نهایت، لایهٔ آخر که دارای ۶۴ نورون است، به لایهٔ خروجی متصل می‌شود. در این مدل، برای فعال‌سازی نورون‌ها در لایه‌های میانی از تابع فعال‌ساز ReLU[3] استفاده شده است که به‌دلیل ویژگی‌های غیرخطی آن، به مدل امکان می‌دهد تا روابط پیچیده بین ورودی‌ها و خروجی‌ها را بهتر یاد بگیرد. در لایهٔ خروجی از تابع فعال‌ساز خطی استفاده شده است تا مقادیر Q به‌صورت مستقیم و بدون محدودیت در محدوده خاصی تولید شوند. این امر به شبکه امکان می‌دهد تا مقادیر پاداش‌های آینده را به‌درستی تخمین بزند. برای بهینه‌سازی این مدل، از بهینه‌ساز Adam با پارامترهای پیش‌فرض استفاده شده است که به‌دلیل کارایی و پایداری بالا، انتخاب مناسبی برای بهینه‌سازی شبکه‌های عصبی عمیق است. تابع هزینه مورد استفاده، میانگین مربعات خطا[4] است که اختلاف بین مقادیر پیش‌بینی‌شده توسط مدل و مقادیر واقعی را به حداقل می‌رساند. در فرایند آموزش شبکه، از یک حافظهٔ تجربه[5] با ظرفیت ۳۰۰۰ نمونه استفاده شده است که از هم‌بستگی بین نمونه‌ها جلوگیری کرده و به آموزش مؤثرتری منجر می‌شود. همچنین، به‌منظور حفظ پایداری در یادگیری، یک مدل هدف[6] طراحی شده است. برای توازن بین کاوش و بهره‌برداری، از استراتژی ε– حریصانه استفاده شده است که مقدار اولیه ε برابر با ۱ بوده و با نرخ ۰/۹۹۹۵ به‌تدریج کاهش یافته و تا حداقل مقدار ۰/۱ ادامه می‌یابد. این تنظیمات به شبکه اجازه می‌دهد تا به‌طور مؤثر استراتژی‌های بهینه را در فرایند بلیت‌فروشی بیاموزد و عملکرد خود را بهبود بخشد.

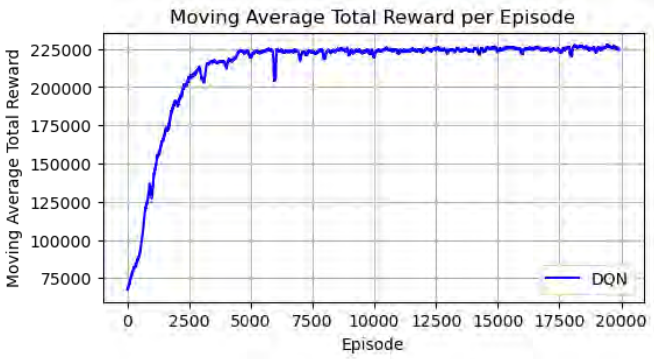
1. Multilayer Perceptron (MLP)
2. Fully Connected Neural Network
3. Rectified Linear Unit
4. Mean Squared Error
5. Experience Replay
6. Target Network

یافته‌های پژوهش

این بخش نتایج شبیه‌سازی‌های عددی را برای نشان دادن کارایی الگوریتم قیمت‌گذاری پویا پیشنهادی برای کلاس‌های مختلف مشتریان ارائه می‌کند. نقطه‌ضعف اصلی مدل‌های یادگیری تقویتی این است کـه مـدل‌هـا اغلـب بـرای تحلیـل، بیش‌ازحد پیچیده هستند و به‌این‌ترتیب، بینش‌هایی در مورد رفتار استراتژی‌های قیمت‌گذاری متفاوت تنها با آزمایش‌های عددی و شبیه‌سازی به‌دست می‌آیند (دن بوئر، ۲۰۱۵). در این مقاله نیز برای تحلیل مدل قیمت‌گذاری پویا از شبیه‌سـازی استفاده شده است.

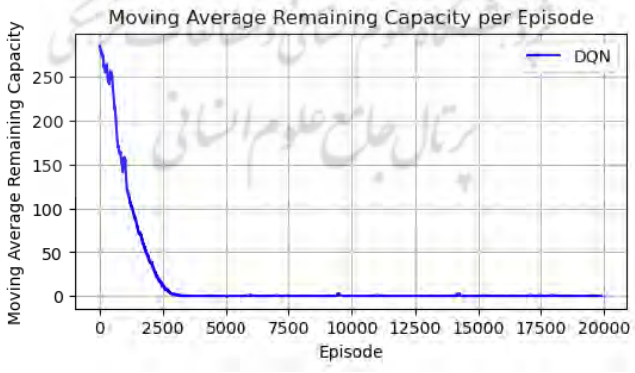
در این شبیه‌سازی، از الگوریتم شبکهٔ عمیق Q برای حل مسـئله قیمـت‌گـذاری پویـای کـلاس‌هـای مشـتریان در سیستم‌های حمل‌ونقل ریلی استفاده شده است. هدف، یافتن قیمت‌های بهینه برای هـر مشـتری ضـمن حـداکثر کـردن درآمد شرکت حمل‌ونقل ریلی است. برای تحلیل عملکرد الگوریتم شبکهٔ عمیق Q، نمودارهای «درآمد کـل»، «میـانگین قیمت ارائه‌شده به مشتریان»، «ظرفیت باقی‌مانده» و «تعداد بلیت فروخته‌شده برای هر کلاس خـدماتی» بـرای بررسـی هم‌گرایی این الگوریتم رسم شده‌اند. معیار هم‌گرایی به‌وسیلهٔ میانگین متحرک ارزیابی می‌شود که نشان‌دهندهٔ پیشرفت و پایداری الگوریتم در طول زمان است. نوسان‌های موجود در نمودارهای رسم شده را می‌توان با عدم قطعیـت در تقاضـا و میزان تمایل به پرداخت مشتریان توضیح داد. در نمودارهای رسم شده، هر تکرار میانگین بازه زمانی ۱۰۰ قسمت است.

شکل ۴ میانگین متحرک درآمد در هر دوره از فرایند یادگیری تقـویتی عمیـق را نشـان مـی‌دهـد. محـور y مقـدار میانگین متحرک درآمد کل را نشان می‌دهد، درحالی‌که محور x شمارهٔ دوره را نشان می‌دهد. در مراحل اولیـه یـادگیری، نمودار با مقدار کمی از درآمد آغاز می‌شود. درآمد کم در تکرارهای اولیه طبیعی است؛ زیرا مدل در ایـن مرحلـه هنـوز در حال کاوش محیط و شناسایی سیاست‌های بهینه برای کسب پاداش بیشتر است. با این حـال، پـس از تکرارهـای بیشـتر مشاهده می‌شود که شیب نمودار به‌سرعت افزایش می‌یابد. این افزایش سریع در دوره‌های اولیـه نشـان‌دهنـدهٔ یـادگیری مؤثر و پرسرعت مدل است؛ چراکه عامل در این مرحله از یادگیری، با استفاده از نرخ کاوش بالا، به‌سرعت استراتژی‌های مؤثرتری را کشف کرده و توانسته است پاداش‌های بیشتری کسب کند. ایـن مرحلـه از یـادگیری معمـولاً بـا آزمـایش و خطاهای متعدد همراه است که در نهایت به افزایش چشمگیر درآمد منجر می‌شـود. پـس از حـدود ۵۰۰۰ تکـرار، شـیب نمودار کاهش می‌یابد و منحنی به تدریج صاف و پایدار می‌شود. این روند نشان‌دهندهٔ هم‌گرایی فرایند یادگیری است، بـه این معنا که مدل به یک سیاست بهینه دست یافته و عملکرد آن به سطحی رسیده که دیگر با تکرارهـای اضـافی بهبـود قابل‌توجهی نخواهد یافت. به عبارت دیگر، در این مرحله، مدل توانسته است به یک سیاست باثبات و بهینه برسد که بـه آن امکان می‌دهد به‌طور مداوم پاداش‌های بالا و نسبتاً ثابت را کسب کند. نوسان‌های کـوچکی کـه پـس از هـم‌گرایـی مشاهده می‌شود، طبیعی هستند و ناشی از تغییرات تصادفی در محیط یا تنوع در پاسخ‌هـای عامـل باشـد. بـا ایـن حـال، میانگین پاداش‌ها در این مرحله حول مقدار تقریبی ۲۲۵ هزار باقی می‌ماند که نشان‌دهندهٔ پایداری عملکـرد مـدل اسـت. درواقع، این مقدار نشان‌دهندهٔ آن است که این شرکت به‌طور متوسط از هر قطـار ۲۲۵ هـزار واحـد پـولی درآمـد کسـب می‌کند.



شکل ۴. میانگین متحرک درآمد کل

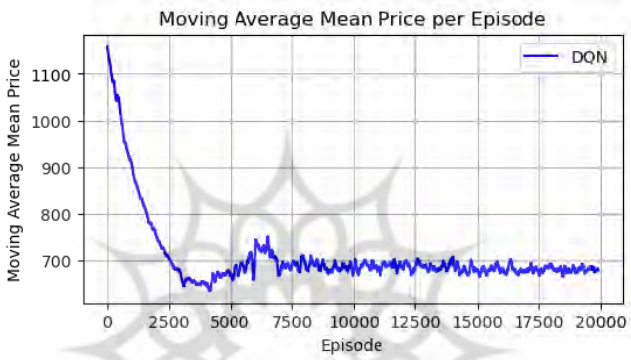
شکل ۵ نشان‌دهندهٔ میانگین متحرک ظرفیت باقی‌مانده در هر دوره را در طول یک فرایند یادگیری تقویتی عمیق است. محور y میانگین متحرک ظرفیت باقی‌مانده را نشان می‌دهد، درحالی‌که محور x عدد دوره را نشان می‌دهد. نمودار با ظرفیت باقی‌مانده بسیار بالا در حدود ۲۸۰ شروع می‌شود. این مقدار بالا در ابتدای فرایند آموزش قابل انتظار است؛ زیرا مدل در این مرحله هنوز سیاست بهینه‌ای برای قیمت‌گذاری یاد نگرفته است. با این حال، مقدار ظرفیت باقی‌مانده پس از تکرارهای اولیه به‌سرعت کاهش می‌یابد. این کاهش نشان می‌دهد که مدل در مراحل ابتدایی آموزش، به‌سرعت در حال یادگیری است و عامل با قیمت‌گذاری بهتر، توانسته تعداد بلیت بیشتری را به فروش برساند. پس از حدود ۳۰۰۰ تکرار، شیب نمودار به تدریج کاهش یافته و منحنی به یک حالت صاف و پایدار می‌رسد. در این نقطه، ظرفیت باقی‌مانده به صفر نزدیک می‌شود. عامل اکنون به‌طور مؤثری یاد گرفته است که برای حداکثرسازی درآمد کل، تمامی بلیت‌های موجود را به فروش برساند. این موضوع نشان می‌دهد که فرایند یادگیری هم‌گرا شده است و عملکرد مدل دیگر به‌طور چشمگیری با تکرارهای اضافی بهبود نمی‌یابد.



شکل ۵. میانگین متحرک ظرفیت باقی‌مانده

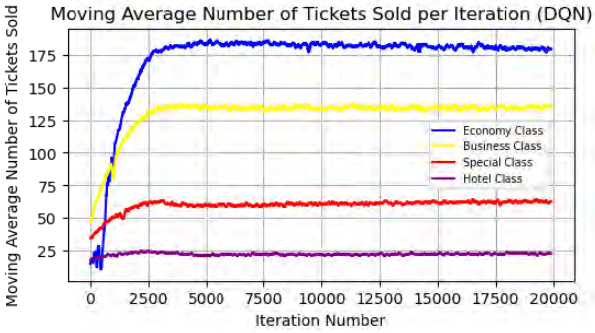
در شکل ۶، نمودار میانگین متحرک قیمت پیشنهادی در طول فرایند یادگیری تقویتی به تصویر کشیده شده است. در محور y، میانگین متحرک قیمت نمایش داده می‌شود و محور x نشان‌دهندهٔ تعداد دوره‌های یادگیری است. در ابتدا، نمودار با میانگین قیمتی نسبتاً بالا، در حدود ۱۱۵۰ شروع می‌شود. با پیشرفت فرایند یادگیری، مشاهده می‌شود که عامل

به‌سرعت قیمت‌ها را کاهش می‌دهد؛ این کاهش به‌وضوح در تکرارهای اولیه با یک منحنی رو به پایین تند نمایان است. این کاهش سریع نشان می‌دهد که عامل یادگیری تقویتی به‌طور پیوسته در حال تنظیم قیمت‌ها است تا به تعادلی بین قیمت و تقاضا برسد که به حداکثرسازی درآمد منجر می‌شود. با رسیدن به حدود ۷۵۰۰ تکرار، منحنی نمودار به یک حالت مسطح و پایدار می‌رسد و میانگین قیمت در حدود ۶۸۰ تا ۷۰۰ نوسان می‌کند. این تثبیت و صاف شدن منحنی نشان‌دهندهٔ هم‌گرایی فرایند یادگیری به یک سیاست قیمت‌گذاری پایدار است. نظارت بر میانگین متحرک قیمت با نشان دادن چگونگی تکامل استراتژی قیمت‌گذاری در طول فرایند یادگیری تقویتی، سایر معیارها را تکمیل می‌کند. یک نقطه قیمت رقابتی و پایدار نشان می‌دهد که سیستم به‌عنوان بخشی از راه‌حل کلی مدیریت درآمد، روی یک سیاست قیمت‌گذاری مؤثر هم‌گرا شده است.



شکل ٦. میانگین متحرک میانگین قیمت در هر دوره

شکل ۷ روند میانگین متحرک تعداد بلیت‌های فروخته‌شده در هر کلاس خدماتی را طی تکرارهای شبیه‌سازی نشان می‌دهد. محور y میانگین تعداد بلیت‌های فروخته‌شده در هر کلاس را نمایش می‌دهد، در حالی که محور x تعداد دوره‌های شبیه‌سازی را نشان می‌دهد. این نمودار اطلاعات ارزشمندی در مورد روند هم‌گرایی الگوریتم به تعداد بهینه بلیت‌های فروخته‌شده برای هر کلاس ارائه می‌دهد. کلاس اکونومی بیشترین تعداد بلیت فروخته‌شده را دارد که نشان‌دهندهٔ تقاضای بالای مسافران برای این کلاس است. از طرف دیگر، کلاس هتل کمترین تعداد بلیت فروخته‌شده را دارد که به‌دلیل قیمت بالاتر و تقاضای کمتر برای این کلاس است. با افزایش تعداد تکرارها، الگوریتم به‌تدریج به‌سمت یافتن تعداد بهینه بلیت‌های فروخته‌شده برای هر کلاس حرکت می‌کند. پس از حدود ۵۰۰۰ تکرار، نمودار برای تمامی کلاس‌ها به یک سطح تقریباً ثابت و پایدار می‌رسد که این هم‌گرایی موفقیت‌آمیز الگوریتم را نشان می‌دهد. در انتهای تکرارها، تعداد بلیت‌های فروخته‌شده برای هر کلاس به این صورت است: کلاس اکونومی حدود ۱۷۵ تا ۱۸۰ بلیت، کلاس بیزینس حدود ۱۳۰ تا ۱۳۵ بلیت، کلاس ویژه حدود ۶۰ تا ۶۵ بلیت و کلاس هتل حدود ۲۳ تا ۲۵ بلیت. این مسطح شدن منحنی‌ها در انتهای شبیه‌سازی نشان می‌دهد که الگوریتم به‌طور مؤثر توانسته است تعداد بهینه بلیت‌های فروخته‌شده برای هر کلاس را پیدا کند و به یک سطح پایدار برسد.



شکل ۷. میانگین متحرک تعداد بلیت‌های فروخته‌شده برای هر کلاس خدماتی

بحث و نتیجه‌گیری

در این مطالعه از «یادگیری تقویتی عمیق» در صنعت حمل‌ونقل ریلی استفاده شده است. این نوآوری، رویکردهای سنتی قیمت‌گذاری پویا را با رویکردی جدیدتر و هوشمندتر جایگزین کرده و نشان می‌دهد که الگوریتم‌های یادگیری تقویتی می‌توانند به‌طور مؤثری در حل مسائل پیچیده قیمت‌گذاری پویا به‌کار گرفته شوند. یک مزیت بارز یادگیری تقویتی در زمینهٔ قیمت‌گذاری پویا، توانایی این روش در انجام هم‌زمان تخمین پارامترها و بهینه‌سازی تصمیمات قیمتی است. در واقع، یادگیری تقویتی یک روش یادگیری است که در آن عامل از طریق تعامل با محیط، از تجربیات گذشته خود یاد می‌گیرد و با استفاده از یک تابع پاداش، تصمیمات بهینه را اتخاذ می‌کند (ساتون و بارتو، ۲۰۱۸)؛ این قابلیت نه‌تنها به افزایش دقت پیش‌بینی‌ها کمک می‌کند، بلکه به‌واسطهٔ بهینه‌سازی مداوم سیاست‌های قیمت‌گذاری، کارایی و سودآوری شرکت‌های حمل‌ونقل ریلی را به‌طور چشمگیری ارتقا می‌بخشد. علاوه‌براین، «کلاس‌های خدماتی مشتریان» به‌عنوان یکی از متغیرهای حالت در مدل قیمت‌گذاری پویا در صنعت حمل‌ونقل ریلی در نظر گرفته شده است. این رویکرد جدید که در سال‌های اخیر موردتوجه محققان قرارگرفته است، مدلی جامع‌تر برای قیمت‌گذاری فراهم می‌کند که علاوه‌بر متغیرهایی مانند ظرفیت باقی‌مانده و زمان حرکت، کلاس خدماتی را نیز در بهینه‌سازی قیمت‌ها لحاظ می‌کند. درنظر گرفتن این متغیر در مدل‌سازی، امکان شخصی‌سازی قیمت‌ها بر اساس ویژگی‌ها و رفتار مشتریان را فراهم کرده و در نهایت به افزایش رضایت مشتریان و سودآوری شرکت حمل‌نقل ریلی منجر می‌شود.

در این مطالعه، مسئله قیمت‌گذاری پویا به‌عنوان یک فرایند تصمیم‌گیری مارکوف فرمول‌بندی شده است و عوامل تأثیرگذار مانند ظرفیت باقی‌مانده، کلاس خدماتی موردنظر مشتری و زمان باقی‌مانده تا حرکت قطار به‌عنوان متغیرهای حالت در نظر گرفته شده‌اند. استفاده از الگوریتم شبکهٔ عمیق Q برای قیمت‌گذاری پویا امکان تنظیم خودکار و بهینه‌سازی قیمت‌ها را فراهم می‌کند. این امر موجب می‌شود تا شرکت‌های حمل‌ونقل ریلی بتوانند به‌صورت لحظه‌ای و بر اساس شرایط واقعی بازار، قیمت‌ها را تنظیم و حداکثر درآمد را کسب کنند. همچنین الگوریتم شبکهٔ عمیق Q، تعداد بهینه بلیت فروخته‌شده در هر کلاس خدماتی را به‌دست آورد که مقادیر به‌دست‌آمده شرکت حمل‌ونقل ریلی را در امر تقسیم ظرفیت کل قطار به کلاس‌های خدماتی که یکی از دغدغه‌های اصلی شرکت حمل‌ونقل ریلی است، یاری می‌کند. درمجموع، یافته‌های این مطالعه نشان دادند که روش‌های یادگیری تقویتی عمیق می‌توانند در حل مسائل پیچیده

قیمت‌گذاری پویا در صنعت حمل‌ونقل ریلی نیز مانند صنایع مختلف دیگر مورداستفاده قرار گیرند و جایگزین روش‌های سنتی شوند.

استفاده از مدل‌های قیمت‌گذاری پویا با استفاده از روش‌های یادگیری ماشین و یادگیری تقویتی، پیامدهای کاربردی و مدیریتی مهمی را برای مدیران کسب‌وکار فراهم می‌کند. این پیامدها می‌توانند به بهبود تصمیم‌گیری‌ها، افزایش کارایی و بهره‌وری و در نهایت، ارتقای عملکرد کسب‌وکار کمک کنند. این پیامدها عبارت‌اند از:

۱. قیمت‌گذاری پویا و هوشمند: همان‌طور که در شکل ۶ نشان داده شد، مدل یادگیری تقویتی توانسته است به‌طور مؤثری قیمت‌ها را تنظیم کرده و به یک سیاست پایدار در قیمت‌گذاری دست یابد. این امر نشان می‌دهد که استفاده از یادگیری تقویتی در سیستم‌های قیمت‌گذاری پویا می‌تواند به بهینه‌سازی قیمت‌ها بر اساس تقاضا و وضعیت بازار منجر شود. در نتیجه، شرکت‌های فعال در صنعت حمل‌ونقل ریلی مانند فدک، می‌توانند از این الگوریتم‌ها برای تنظیم قیمت‌ها استفاده کنند و درآمد خود را افزایش دهند، بدون اینکه تقاضای مشتریان کاهش یابد.

۲. تحلیل دقیق‌تر تقاضا: مدل‌های توسعه‌یافته در این تحقیق، به مدیران امکان می‌دهند تا الگوهای تقاضا را به‌طور دقیق‌تر تحلیل کرده و پیش‌بینی‌های دقیق‌تری از تغییرات تقاضا در بازار به دست آورند. این مدل‌ها امکان پاسخ‌دهی سریع به تغییرات بازار را فراهم می‌کنند و به مدیران این فرصت را می‌دهند تا به‌طور هوشمندانه‌تر و کارآمدتر تصمیم‌گیری کنند.

۳. حداکثرسازی فروش بلیت‌ها و استفاده بهینه از ظرفیت قطار: نتایج نشان می‌دهد که مدل یادگیری تقویتی پس از حدود ۳۰۰۰ تکرار توانسته است ظرفیت باقی‌مانده قطار را به صفر برساند؛ به این معنا که تمامی بلیت‌ها به فروش رسیده‌اند (شکل ۵). فروش تمامی بلیت‌ها نشان‌دهندهٔ توانایی مدل در تخصیص بهینه منابع و فروش کامل ظرفیت است. در عمل، شرکت‌های ریلی می‌توانند با بهره‌گیری از این نوع مدل‌ها، از منابع خود به‌طور کامل استفاده کرده و از هدررفت ظرفیت جلوگیری کنند.

۴. تقسیم‌بندی دقیق‌تر مشتریان: با در نظر گرفتن کلاس خدماتی به‌عنوان یک متغیر حالت، الگوریتم قادر است تمایل به پرداخت مشتریان را به شکلی دقیق‌تر یاد بگیرد و بر اساس آن، قیمت‌ها را متناسب با توان خرید مشتری تنظیم کند. در عین حال، یکی از اهداف اصلی مدل، فروش تمامی ظرفیت قطار است. بنابراین، مدل به‌گونه‌ای عمل می‌کند که علاوه‌بر فروش بلیت‌های کلاس‌های بالاتر و کسب درآمد بیشتر از این بخش، اطمینان حاصل شود که تمامی بلیت‌ها در سطوح خدماتی مختلف به فروش می‌رسند. این نتایج به مدیران شرکت‌های ریلی این امکان را می‌دهد که استراتژی‌های قیمت‌گذاری خود را متناسب با نیازها و توان خرید مشتریان هر کلاس خدماتی تنظیم کنند و به حداکثرسازی سود هر کلاس کمک کنند.

۵. کاهش ریسک در تصمیم‌گیری‌های استراتژیک: نمودار شکل ۴ نشان‌دهندهٔ روند میانگین متحرک درآمد است و شکل ۵ نشان‌دهندهٔ روند استفاده از ظرفیت قطارها است. این دو نمودار به مدیران نشان می‌دهد که با پیاده‌سازی این مدل، ریسک ناشی از قیمت‌گذاری نامناسب یا فروش ناکافی بلیت‌ها به حداقل می‌رسد. درنتیجه، مدیران می‌توانند تصمیمات آگاهانه‌تری در مورد قیمت‌گذاری و تخصیص ظرفیت بگیرند و از بهینه‌سازی عملکرد اطمینان حاصل کنند.

۶. توانمندسازی در برابر رقبا: با استفاده از تکنیک‌های پیشرفته و به‌روز در قیمت‌گذاری پویا، شرکت‌ها می‌توانند مزیت رقابتی خود را افزایش داده و در بازارهای پویا و رقابتی بهتر عمل کنند.

استفاده از این پیامدهای کاربردی و مدیریتی می‌تواند به مدیران کمک کند تا به‌طور مؤثرتر و کارآمدتر از مدل‌های قیمت‌گذاری پویا استفاده کنند و عملکرد کلی کسب‌وکار را بهبود بخشند.

محدودیت‌های پژوهش و پیشنهادهای پژوهش‌های آتی

تحقیقات انجام‌شده در زمینه قیمت‌گذاری پویا با استفاده از روش‌های یادگیری ماشین و یادگیری تقویتی، نشان می‌دهد که هنوز زمینه‌های بسیاری برای بهبود و گسترش این حوزه وجود دارد. برای محققان و مدیران کسب‌وکار، زمینه‌های تحقیقاتی آتی می‌تواند شامل موارد زیر باشد:

۱. افزایش میانگین میزان تمایل به پرداخت مشتریان با گذر زمان: در تحقیقات فعلی، یکی از چالش‌های اصلی این بود که مدل‌ها به تغییرات در تمایل به پرداخت مشتریان به‌مرور زمان حساس نبودند. مدل‌های موجود فرض می‌کردند که تمایل به پرداخت مشتریان ثابت است یا به‌طور ناگهانی تغییر می‌کند که به عدم توانایی در تطبیق با تغییرات تدریجی در رفتار مشتریان منجر شد. به‌منظور رفع این مشکل، پیشنهاد می‌شود که میانگین میزان تمایل به پرداخت مشتریان با گذر زمان افزایش یابد. این رویکرد می‌تواند مدل‌ها را تطبیق‌پذیرتر و پاسخ‌گوتر به تغییرات تدریجی در رفتار مشتریان کند (الکساندر و لینگ، ۲۰۱۹).

۲. بازنگری در فرضیهٔ فضای گسسته و محدود قیمت‌گذاری: یکی از محدودیت‌های عمده در تحقیقات ما این بود که مدل‌ها تنها قادر به تنظیم قیمت‌ها در یک مجموعهٔ گسسته و محدود از انتخاب‌ها بودند. این محدودیت، باعث می‌شد که مدل‌ها نتوانند قیمت‌گذاری دقیق و بهینه‌تری را انجام دهند، به‌خصوص در بازارهایی که نیاز به تغییرات ظریف و پیوسته قیمت دارند. پیشنهاد می‌شود که به‌جای استفاده از فضای گسسته، از فضای عمل پیوسته یا تقریب‌های با دقت بالا برای قیمت‌گذاری استفاده شود. این اقدام می‌تواند دقت و کارایی مدل‌ها را در تنظیم قیمت‌ها بهبود بخشد.

۳. مقایسهٔ الگوریتم‌های یادگیری تقویتی مختلف و استفاده از مدل‌های ترکیبی: در مقاله حاضر، تمرکز اصلی بر روی پیاده‌سازی و ارزیابی الگوریتم DQN برای مسئله قیمت‌گذاری پویا به‌عنوان یکی از الگوریتم‌های مرجع بوده است. پیشنهاد می‌شود که در تحقیقات آتی به بررسی و مقایسهٔ عملکرد الگوریتم DQN با روش‌های پیشرفته‌تری مانند Double DQN و Dueling DQN و سایر الگوریتم‌های یادگیری تقویتی و یادگیری تقویتی عمیق برای مسئله قیمت‌گذاری پویا در سیستم‌های حمل‌ونقل ریلی پرداخت. این مقایسه می‌تواند به درک بهتری از مزایا و معایب هر کدام از این الگوریتم‌ها در زمینه قیمت‌گذاری پویا کمک کند و به بهبود نتایج منجر شود. همچنین، بررسی مدل‌های ترکیبی که مزایای چندین الگوریتم را با هم ترکیب می‌کنند، می‌تواند به بهبود عملکرد و دقت مدل‌ها کمک کند.

منابع

مهرجو، سارا؛ عموزاد مهدیرجی، حنان؛ حیدری دهوئی، جلیل؛ رضوی حاجی آقا، سید حسین و حسین‌زاده، مهناز (۱۴۰۲). ارائه مدلی استوار برای قیمت‌گذاری پویا و مقایسه آن با قیمت‌گذاری ایستا در زنجیره‌های تأمین چندسطحی با رویکرد تئوری بازی‌ها. *مدیریت صنعتی*، ۱۵(۴)، ۵۳۴-۵۶۵.

References

Abdalrahman, A. & Zhuang, W. (2020). Dynamic pricing for differentiated PEV charging services using deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 23(2), 1415-1427.

Adelnia Najafabadi, H., Shekarchizadeh, A., Nabiollahi, A., Khani, N. & Rastgari, H. (2022). Dynamic pricing for information goods using revenue management and recommender systems. *Journal of Revenue and Pricing Management*, 21(2), 153–163.

Alexander, R. B. & Ling, J. S. (2019). *Multi-segment dynamic pricing for airline tickets using model-free reinforcement learning.*

Aljafari, B., Jeyaraj, P. R., Kathiresan, A. C. & Thanikanti, S. B. (2023). Electric vehicle optimum charging-discharging scheduling with dynamic pricing employing multi agent deep neural network. *Computers and Electrical Engineering*, 105, 108555.

Armstrong, A. & Meissner, J. (2010). Railway revenue management: Overview and models (operations research). *Department of Management Science, Lancaster University Working Papers*, (MRG/0019).

Avila, N., Hardan, S., Zhalieva, E., Aloqaily, M. & Guizani, M. (2022). Energy Pricing in P2P Energy Systems Using Reinforcement Learning. *arXiv preprint arXiv*:2210.13555.

Bagherpour, R., Mozayani, N. & Badnava, B. (2021). Improving demand-response scheme in smart grids using reinforcement learning. *International Journal of Energy Research*, 45(15), 21082-21095.

Bertsimas, D. & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3), 1025-1044.

Bondoux, N., Nguyen, A. Q., Fiig, T. & Acuna-Agost, R. (2020). Reinforcement learning applied to airline revenue management. *Journal of Revenue and Pricing Management*, 19(5), 332-348.

Burger, B. & Fuchs, M. (2005). Dynamic pricing—A future airline business model. *Journal of Revenue and Pricing Management*, 4(1), 39-53.

Chen, S., Li, L., Chen, Z. & Li, S. (2020). Dynamic pricing for smart mobile edge computing: A reinforcement learning approach. *IEEE Wireless Communications Letters*, *10*(4), 700-704.

Collins, A. & Thomas, L. (2012). Comparing reinforcement learning approaches for solving game theoretic models: a dynamic airline pricing game example. *Journal of the Operational Research Society*, 63(8), 1165–1173.

Cong, P., Zhou, J., Chen, M. & Wei, T. (2020). Personality-guided cloud pricing via reinforcement learning. *IEEE Transactions on Cloud Computing*, 10(2), 925-943.

Den Boer, A. V. (2015). Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science*, 20(1), 1-18.

Du, J., Cheng, W., Lu, G., Cao, H., Chu, X., Zhang, Z. & Wang, J. (2021). Resource pricing and allocation in MEC enabled blockchain systems: An A3C deep reinforcement learning approach. *IEEE Transactions on Network Science and Engineering*, 9(1), 33-44.

Du, P. & Chen, Q. (2017). Skimming or penetration: optimal pricing of new fashion products in the presence of strategic consumers. *Annals of Operations Research*, 257, 275-295.

Fraija, A., Agbossou, K., Henao, N., Kelouwani, S., Fournier, M. & Hosseini, S. S. (2022). A discount-based time-of-use electricity pricing strategy for demand response with minimum information using reinforcement learning. *IEEE Access*, 10, 54018-54028.

Gao, J., Le, M. & Fang, Y. (2022). Dynamic air ticket pricing using reinforcement learning method. *RAIRO-Operations Research*, 56(4), 2475-2493.

Gosavi, A., Bandla, N. & Das, T. (2002). A reinforcement learning approach to a single leg airline revenue management problem with multiple fare classes and overbooking. *IIE Transactions*, 34, 729–742.

Jing, Y., Guo, S., Chen, F., Wang, X. & Li, K. (2021). Dynamic differential pricing of high-speed railway based on improved GBDT train classification and bootstrap time node determination. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 16854-16866.

Jung, H. (2022). An optimal charging and discharging scheduling algorithm of energy storage system to save electricity pricing using reinforcement learning in urban railway system. *Journal of Electrical Engineering & Technology*, 17(1), 727-735.

Kamandanipour, K., Haji Yakhchali, S. & Tavakkoli-Moghaddam, R. (2023). Dynamic revenue management in a passenger rail network under price and fleet management decisions. *Annals of Operations Research*, 1-25.

Kamandanipour, K., Yakhchali, S. H. & Tavakkoli-Moghaddam, R. (2023). Learning-based dynamic ticket pricing for passenger railway service providers. *Engineering optimization*, *55*(4), 703-717.

Kim, B. G., Zhang, Y., Van Der Schaar, M. & Lee, J. W. (2015). Dynamic pricing and energy consumption scheduling with reinforcement learning. *IEEE Transactions on smart grid*, 7(5), 2187-2198.

Koc, I. & Arslan, E. (2021). Dynamic ticket pricing of airlines using variant batch size interpretable multi-variable long short-term memory. *Expert Systems with Applications*, 175, 114794.

Krasheninnikova, E., García, J., Maestre, R. & Fernández, F. (2019). Reinforcement learning for pricing strategy optimization in the insurance industry. *Engineering applications of artificial intelligence*, 80, 8-19.

Lei, Z. & Ukkusuri, S. V. (2023). Scalable reinforcement learning approaches for dynamic pricing in ride-hailing systems. *Transportation Research Part B: Methodological*, 178, 102848.

Liao, Y., Qiao, X., Yu, Q. & Liu, Q. (2021). Intelligent dynamic service pricing strategy for multi-user vehicle-aided MEC networks. *Future Generation Computer Systems*, 114, 15-22.

Liu, H., Chen, C., Li, Y., Duan, Z. & Li, Y. (2022). Chapter 1 - Introduction. In H. Liu, C. Chen, Y. Li, Z. Duan & Y. Li (Eds.), Smart Metro Station Systems (pp. 1–32). *Elsevier*.

Liu, Y., Zhang, D. & Gooi, H. B. (2020). Data-driven decision-making strategies for electricity retailers: A deep reinforcement learning approach. *CSEE Journal of Power and Energy Systems*, 7(2), 358-367.

Lu, R., Hong, S. H. & Zhang, X. (2018). A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. *Applied energy*, 220, 220-230.

Lu, T., Chen, X., McElroy, M. B., Nielsen, C. P., Wu, Q. & Ai, Q. (2020). A reinforcement learning-based decision system for electricity pricing plan selection by smart grid end users. *IEEE Transactions on Smart Grid*, 12(3), 2176-2187.

Mehrjoo, S., Amoozad Mahdirji, H., Heidary Dahoei, J., Razavi Haji Agha, S. H. & Hosseinzadeh, M. (2023). Providing a Robust Dynamic Pricing Model and Comparing It with Static Pricing in Multi-level Supply Chains Using a Game Theory Approach. *Industrial Management Journal*, 15(4), 534-565. (*in Persian*)

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv*:1312.5602.

Moghaddam, V., Yazdani, A., Wang, H., Parlevliet, D. & Shahnia, F. (2020). An online reinforcement learning approach for dynamic pricing of electric vehicle charging stations. *IEEE Access*, 8, 130305-130313.

Narahari, Y., Raju, C. V. L., Ravikumar, K. & Shah, S. (2005). Dynamic pricing models for electronic business. *Sadhana*, 30(2), 231–256.

Nian, R., Liu, J. & Huang, B. (2020). A review on reinforcement learning: Introduction and applications in industrial process control. *Computers & Chemical Engineering*, 139, 106886.

Pandey, V. & Boyles, S. D. (2018). Dynamic pricing for managed lanes with multiple entrances and exits. *Transportation Research Part C: Emerging Technologies*, 96, 304-320.

Pandey, V., Wang, E. & Boyles, S. D. (2020). Deep reinforcement learning algorithm for dynamic pricing of express lanes with multiple access locations. *Transportation Research Part C: Emerging Technologies*, *119*, 102715.

Poh, L. Z., Connie, T., Ong, T. S. & Goh, M. K. O. (2023). Deep reinforcement learning-based dynamic pricing for parking solutions. *Algorithms*, 16(1), 32.

Qiu, D., Ye, Y., Papadaskalopoulos, D. & Strbac, G. (2020). A deep reinforcement learning method for pricing electric vehicles with discrete charging levels. *IEEE Transactions on Industry Applications*, 56(5), 5901-5912.

Raju, C. V. L., Narahari, Y. & Ravikumar, K. (2006). Learning dynamic prices in electronic retail markets with customer segmentation. *Annals of Operations Research*, 143(1), 59–75.

Rana, R. & Oliveira, F. S. (2014). Real-time dynamic pricing in a non-stationary environment using model-free reinforcement learning. *Omega*, 47, 116–126.

Russell, S. & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3th ed.). Prentice-Hall, Upper Saddle River.

Saharan, S., Bawa, S. & Kumar, N. (2020). Dynamic pricing techniques for Intelligent Transportation System in smart cities: A systematic review. *Computer Communications*, 150, 603-625.

Sato, K., Seo, T. & Fuse, T. (2021). A reinforcement learning-based dynamic congestion pricing method for the morning commute problems. *Transportation Research Procedia*, 52, 347-355.

Shan, X., Lv, X., Wu, J., Zhao, S. & Zhang, J. (2024). Revenue management method and critical techniques of railway passenger transport. *Railway Sciences*, 3(5), 636-649.

Stavinova, E., Chunaev, P. & Bochenina, K. (2021). Forecasting railway ticket dynamic price with Google Trends open data. *Procedia Computer Science*, 193, 333–342.

Strauss, A. K., Klein, R. & Steinhardt, C. (2018). A review of choice-based revenue management: Theory and methods. *European Journal of Operational Research*, 271(2), 375–387.

Sutton, R. S. & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tan, M. (2018). Optimal Pricing for Tickets with Myopic and Strategic Passengers. *Ind. Eng. Manag*, 23, 107-115.

Wan, Y., Qin, J., Yu, X., Yang, T. & Kang, Y. (2021). Price-based residential demand response management in smart grids: A reinforcement learning-based approach. *IEEE/CAA Journal of Automatica Sinica*, 9(1), 123-134.

Wang, S., Bi, S. & Zhang, Y. A. (2019). Reinforcement learning for real-time pricing and scheduling control in EV charging stations. *IEEE Transactions on Industrial Informatics*, 17(2), 849-859.

Wittman, M. D. & Belobaba, P. P. (2019). Dynamic pricing mechanisms for the airline industry: a definitional framework. *Journal of Revenue and Pricing Management*, 18(2), 100–106.

Wu, X., Qin, J., Qu, W., Zeng, Y. & Yang, X. (2019). Collaborative optimization of dynamic pricing and seat allocation for high-speed railways: An empirical study from China. *IEEE Access*, 7, 139409-139419.

Xiaoqiang, Z., Lang, M. & Jin, Z. (2017). Dynamic pricing for passenger groups of high-speed rail transportation. *Journal of Rail Transport Planning & Management*, 6(4), 346-356.

Xu, H., Wen, J., Hu, Q., Shu, J., Lu, J. & Yang, Z. (2022). Energy Procurement and Retail Pricing for Electricity Retailers via Deep Reinforcement Learning with Long Short-term Memory. *CSEE Journal of Power and Energy Systems*, *8*(5), 1338-1351.

Xu, Z., Guo, Y., Sun, H., Tang, W. & Huang, W. (2023). Deep reinforcement learning for competitive DER pricing problem of virtual power plants. *CSEE Journal of Power and Energy Systems*.

Yan, Z., Zhang, P., Zhang, Y., Liu, H., Feng, C. & Li, X. (2019). Joint decision model of group ticket booking limits and individual passenger dynamic pricing for the high-speed railway. *Symmetry*, 11(9), 1128.

Yang, Q. Q., Xu, L. P. & Yang, Y. (2012). Dynamic Pricing for Multiple-Class High-Speed Railway on the Internet. *Applied Mechanics and Materials*, 253–255, 1263–1267.

Yousefi, A. & Pishvaee, M. S. (2022). A hybrid machine learning-optimization approach to pricing and train formation problem under demand uncertainty. *RAIRO-Operations Research*, *56*(3), 1429-1451.

Zeng, H. & Zhang, Y. (2015). Intertemporal Pricing of Substitutes under the Coexistence of Myopic and Strategic Consumers. *Syst. Eng*, 65, 33-39.

Zhang, P., Wang, C., Aujla, G. S. & Batth, R. S. (2021). ReLeDP: Reinforcement-learning-assisted dynamic pricing for wireless smart grid. *IEEE Wireless Communications*, *28*(6), 62-69.

Zheng, J., Gan, Y., Liang, Y., Jiang, Q. & Chang, J. (2021). Joint Strategy of Dynamic Ordering and Pricing for Competing Perishables with Q-Learning Algorithm. *Wireless Communications and Mobile Computing*, (1), 6643195.

Zhu, Y. T., Wang, F. Z., Lv, X. Y. & Pan, Y. (2014, August). Dynamic pricing for railway tickets with demand-shifted passenger groups. *In 2014 International Conference on Management Science & Engineering 21th Annual Conference Proceedings* (pp. 256-262). IEEE.