Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
DOI: 10.22034/jscit.2026.576052.2229
Volume 15, Number 2, Summer 2026, pp. 78-91
Received: 02/16/2026, Revised: 07/05/2026, Accepted: 07/07/2026

A Markov Model-based and Deep Learning-based Electric Vehicle Charging Scheduling in Charge Stations

Samaneh Mohammadi-Jirgol[1], Mojtaba Valinataj[2*] and Fatemeh Zamani[3]
1- Faculty of Electrical and Computer Eng. Babol Noshirvani University of Tech. Babol, Iran.
2*- Faculty of Electrical and Computer Eng. Babol Noshirvani University of Tech. Babol, Iran.
3- Faculty of Electrical and Computer Eng. Babol Noshirvani University of Tech. Babol, Iran.
[1] s.mohammadi@stu.nit.ac.ir, [2*] m.valinataj@nit.ac.ir, and [3] zamani@nit.ac.ir

Corresponding author's address: Mojtaba Valinataj, Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran.

Abstract- The large-scale deployment and expansion of electric vehicles (EVs) significantly contribute to reducing fossil fuel consumption and environmental pollution. In these vehicles, optimal charging scheduling not only enables more efficient utilization of renewable energy resources, but also reduces charging costs and enhances power grid stability. However, fluctuations in electricity prices, variations in energy demand and charging power, and uncertainties regarding battery charging completion time pose major challenges for accurate prediction and optimal charging scheduling. In this paper, the EV charging scheduling problem is formulated as a Markov Decision Process (MDP), and a model-free Deep Reinforcement Learning (DRL)-based approach is proposed to derive an optimal charging strategy that supports real-time decision-making under uncertainty. The proposed model is trained and evaluated using real-world data from California. The results demonstrate that the proposed approach significantly reduces electricity costs while improving charging performance under uncertain conditions and volatile electricity prices. Compared to the baseline method, the proposed strategy achieves a 41% reduction in electricity costs, and compared to conventional methods, it yields higher cumulative rewards and lower cumulative charging costs, indicating its effectiveness and practical applicability for users at charging stations.

Keywords- Electric Vehicle, Charging Scheduling, Charging Station, Long Short-Term Memory (LSTM), Deep Reinforcement Learning.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
شناسه دیجیتال: 10.22034/jscit.2026.576052.2229
جلد ۱۵، شماره ۲، تابستان ۱۴۰۵، صفحه ۷۸-۹۱
دریافت: ۱۴۰۴/۱۱/۲۷، بازنگری: ۱۴۰۵/۰۴/۱۴، پذیرش: ۱۴۰۵/۰۴/۱۶

زمان‌بندی شارژ خودروهای برقی در ایستگاه‌های شارژ مبتنی بر مدل مارکف و یادگیری عمیق

سمانه محمدی جیرگل[۱]، مجتبی ولی‌نتاج*[۲]، فاطمه زمانی[۳]
۱- دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران.
*۲- دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران.
۳- دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران.
[1] s.mohammadi@stu.nit.ac.ir, [2*] m.valinataj@nit.ac.ir, and [3] zamani@nit.ac.ir

* نشانی نویسنده مسئول: مجتبی ولی‌نتاج، بابل، خیابان شریعتی، دانشگاه صنعتی نوشیروانی بابل، دانشکده مهندسی برق و کامپیوتر.

چکیده- تولید و گسترش انبوه وسایل نقلیه برقی به طور قابل توجهی به کاهش مصرف سوخت‌های فسیلی و حفظ محیط زیست کمک می‌کند. در این خودروها، زمان‌بندی مناسب شارژ نه‌تنها امکان بهره‌برداری مؤثرتر از منابع انرژی تجدیدپذیر را فراهم می‌سازد، بلکه با کاهش هزینه‌ها و افزایش پایداری شبکه برق، مزایای قابل‌توجهی نیز به همراه دارد. نوسانات قیمت برق، تغییرات در میزان تقاضای انرژی و توان شارژ و عدم اطمینان در مورد زمان تکمیل شارژ باتری، چالش‌های اساسی در زمینه پیش‌بینی و زمان‌بندی بهینه شارژ ایجاد می‌کنند. در این مقاله، مسأله زمان‌بندی شارژ وسایل نقلیه برقی به صورت یک فرآیند تصمیم‌گیری مارکف مدل‌سازی شده و یک روش مبتنی بر یادگیری تقویتی عمیق (DRL) و بدون نیاز به مدل برای یافتن راهبرد شارژ بهینه ارائه می‌شود که امکان تصمیم‌گیری بی‌درنگ را در شرایط عدم قطعیت فراهم می‌کند. مدل پیشنهادی بر روی داده‌های مربوط به ایالت کالیفرنیا آموزش داده و ارزیابی شده است. نتایج نشان می‌دهد که این رویکرد نه‌تنها منجر به کاهش قابل‌توجه هزینه برق می‌شود، بلکه عملکرد شارژ را نیز در شرایط عدم قطعیت و نوسانات قیمت برق بهبود می‌بخشد. روش پیشنهادی نسبت به روش پایه، منجر به ٪۴۱ کاهش در هزینه برق مصرفی شده و نسبت به روش‌های مرسوم، پاداش تجمعی آن بیشتر و هزینه تجمعی شارژ آن کمتر است که بیانگر کارایی مناسب و قابلیت کاربرد آن برای کاربران در ایستگاه‌های شارژ است.

واژه‌های کلیدی: خودروی برقی، زمان‌بندی شارژ، ایستگاه شارژ، حافظه طولانی کوتاه مدت، یادگیری تقویتی عمیق.

۱- مقدمه

با گسترش روزافزون استفاده از وسایل نقلیه برقی به‌عنوان جایگزینی برای خودروهای با سوخت فسیلی، زمان‌بندی بهینه فرآیند شارژ باتری‌ها به‌ویژه با توجه به افزایش نفوذ این وسایل در مقیاس‌های شهری و رشد ایستگاه‌های شارژ عمومی، به یکی از چالش‌های اساسی در مدیریت زیرساخت‌های شارژ تبدیل شده‌است [۱]. این مسأله در شرایطی که قیمت لحظه‌ای برق [۲]، تقاضای انرژی[۱] [۳] و وضعیت شبکه دارای نوسانات[۲] و عدم قطعیت باشند، پیچیده‌تر می‌شود و تصمیم‌گیری مبتنی بر فرضیات ساده را ناکارآمد می‌سازد. تصمیم‌گیری هوشمند برای زمان‌بندی شارژ می‌تواند منجر به کاهش هزینه‌ها، کاهش فشار روی شبکه برق و بهبود بهره‌وری کلی سیستم‌های شارژ شود که این موضوع از دیدگاه بهره‌برداران شبکه و کاربران نهایی حائز اهمیت است [۴, ۵]. در سال‌های اخیر، روش‌های مبتنی بر هوش مصنوعی به‌ویژه یادگیری تقویتی عمیق[۳]

به‌عنوان راهکاری موثر برای حل مسائل تصمیم‌گیری متوالی معرفی شده‌اند و در حوزه مدیریت انرژی و سیستم‌های قدرت نیز مورد توجه گسترده قرار گرفته‌اند [۶]. این الگوریتم‌ها با الگوبرداری از رفتار یادگیری انسان، قادرند بدون نیاز به مدل‌سازی محیط، به‌مرور زمان بهترین سیاست تصمیم‌گیری را از طریق تعامل با محیط بیاموزند که این ویژگی آن‌ها را برای کاربردهای بی‌درنگ مناسب می‌سازد [۶, ۷].

در این میان، استفاده از شبکه‌های عصبی بازگشتی مانند LSTM [۴] برای پیش‌بینی[۵] قیمت برق و مصرف انرژی، نقش مهمی در تقویت عملکرد سیستم زمان‌بندی دارد، زیرا این شبکه‌ها توانایی استخراج وابستگی‌های زمانی پنهان در داده‌های انرژی را دارند [۱, ۷]. ترکیب پیش‌بینی دقیق با تصمیم‌گیری بهینه می‌تواند اثربخشی سامانه‌های مدیریت انرژی را به‌طور چشم‌گیری افزایش دهد و عدم قطعیت در فرآیند تصمیم‌سازی را کاهش دهد [۷, ۸]. از طرف دیگر، مسأله زمان‌بندی شارژ در ابعاد واقعی نیازمند پردازش حجم بالایی از داده و انجام محاسبات پیچیده است که این موضوع طراحی ساختار مناسب مدل‌های یادگیری را ضروری می‌سازد. در چنین شرایطی، انتخاب ساختار مناسب شبکه عصبی، طراحی صحیح مدل یادگیری و ارزیابی دقیق عملکرد آن از اهمیت بالایی برخوردار است و تأثیر مستقیمی بر پایداری و همگرایی فرآیند یادگیری دارد [۴, ۵].

در این مقاله، با استفاده از داده‌های واقعی مربوط به ایالت کالیفرنیا و بهره‌گیری از ترکیب مدل‌های پیش‌بینی مبتنی بر LSTM با رویکرد یادگیری تقویتی عمیق، روشی کارآمد برای بهینه‌سازی فرآیند شارژ وسایل نقلیه برقی ارائه شده‌است که قابلیت پیاده‌سازی در سناریوهای عملی را دارد [۲, ۳, ۷]. اهمیت این پژوهش در آن است که با در نظر گرفتن عدم قطعیت‌های موجود در سیستم، امکان تصمیم‌گیری هوشمند و پایدار را برای فرآیند شارژ فراهم می‌سازد و به کاربران ایستگاه‌های شارژ این امکان را می‌دهد که به برنامه‌های شارژ باثبات‌تر، اقتصادی‌تر و هماهنگ با وضعیت انرژی دست یابند. سیستم پیشنهادی به‌عنوان یک ایستگاه شارژ هوشمند پیشرفته عمل می‌کند که علاوه بر پشتیبانی از فرآیندهای شارژ و دشارژ، قابلیت ارائه خدماتی نظیر پارکینگ هوشمند و تعویض باتری را نیز دارا بوده و بستری مفید و جامع برای توسعه سامانه‌های حمل‌ونقل هوشمند آینده فراهم می‌سازد.

نوآوری‌های این مقاله به شرح زیر خلاصه شده است:

- ادغام پیش‌بینی‌های مبتنی بر شبکه LSTM با رویکرد یادگیری تقویتی عمیق مبتنی بر شبکه Q به‌منظور بهینه‌سازی زمان‌بندی شارژ وسایل نقلیه برقی
- معرفی یک تابع پاداش چندجزئی به‌منظور دستیابی به بهینه‌سازی چندهدفه شامل کاهش هزینه، بهبود پایداری و افزایش کارایی شارژ
- طراحی یک بردار حالت توسعه‌یافته که علاوه بر اطلاعات انرژی، شامل داده‌های مربوط به پارکینگ هوشمند و وضعیت ایستگاه‌های شارژ است.
- ارائه راهکارهایی برای پایدارسازی فرآیند آموزش و تضمین همگرایی پاداش در رویکرد یادگیری تقویتی عمیق.

ادامه این مقاله بدین صورت است که در بخش دوم، مروری بر تحقیقات مرتبط پیشین ارائه می‌شود. در بخش سوم، مدل‌سازی مسأله و فرمول‌بندی آن تشریح شده و در بخش چهارم، روش پیشنهادی به طور کامل معرفی می‌گردد. نتایج شبیه‌سازی و ارزیابی عملکرد در بخش پنجم ارائه شده و در نهایت، جمع‌بندی و نتیجه‌گیری در بخش ششم بیان می‌شود.

۲- کارهای مرتبط

پیش‌بینی تقاضای شارژ و زمان‌بندی بهینه فرآیند شارژ یکی از موضوعات مهم در حوزه سیستم‌های انرژی و حمل‌ونقل هوشمند است. در برخی پژوهش‌های اولیه، از مدل‌های تصمیم‌گیری و روش‌های مبتنی بر شبیه‌سازی برای زمان‌بندی شارژ استفاده شده‌است. اگرچه این روش‌ها امکان بررسی رفتار تصادفی سیستم را فراهم می‌کنند، اما به دلیل پیچیدگی محاسباتی، توانایی محدودی در پیاده‌سازی در مقیاس‌های بزرگ و کاربردهای بی‌درنگ دارند [۹].

با پیشرفت روش‌های یادگیری ماشین، مدل‌های داده‌محور[۶] برای پیش‌بینی رفتار کاربران و تقاضای شارژ مورد توجه قرار گرفته‌اند. در این راستا، از روش‌های مبتنی بر شبکه‌های عصبی و مدل‌های سری زمانی برای پیش‌بینی زمان ورود خودروها، مدت زمان شارژ و میزان انرژی موردنیاز استفاده شده‌است [۱۰, ۱۱]. اگرچه این روش‌ها دقت پیش‌بینی را افزایش داده‌اند، اما در مواجهه با تغییرات سریع شرایط محیطی و رفتار غیرقابل پیش‌بینی کاربران، انعطاف‌پذیری محدودی دارند.

در ادامه، مدل‌های یادگیری عمیق به‌ویژه از نوع شبکه‌های عصبی بازگشتی[۷] مانند LSTM به دلیل توانایی در استخراج وابستگی‌های زمانی پنهان، به‌طور گسترده برای پیش‌بینی مصرف انرژی و تقاضای شارژ به کار گرفته شده‌اند مانند مراجع [۱۲, ۱۳]. همچنین، استفاده از مدل‌های مکانی-زمانی پیشرفته‌تر نظیر ConvLSTM برای بهبود دقت پیش‌بینی در مقیاس‌های زمانی کوتاه‌مدت و میان‌مدت گزارش

شده‌است مانند مراجع [۱۴, ۱۵]. با وجود عملکرد مناسب این مدل‌ها، تمرکز اصلی آن‌ها بیشتر بر پیش‌بینی بوده و به‌صورت مستقیم فرآیند زمان‌بندی بهینه شارژ را پوشش نمی‌دهند.

در سال‌های اخیر، الگوریتم‌های مبتنی بر یادگیری تقویتی و یادگیری تقویتی عمیق به‌عنوان ابزارهایی مؤثر برای حل مسائل تصمیم‌گیری متوالی در محیط‌های نامعین معرفی شده‌اند. در این مطالعات، الگوریتم‌هایی نظیر DQN[۸]، PPO[۹] و SAC[۱۰] برای بهینه‌سازی فرآیند شارژ و دشارژ وسایل نقلیه برقی به کار گرفته شده‌اند [۱۶-۱۸]. با این حال، بسیاری از این روش‌ها یا از پیش‌بینی دقیق شرایط آینده بهره نمی‌گیرند یا پیاده‌سازی آن‌ها در مقیاس‌های بزرگ با هزینه محاسباتی بالا همراه است، مانند روش‌های ارائه شده در [۱۹, ۲۰]. برخی پژوهش‌ها با ترکیب مدل‌های پیش‌بینی یادگیری عمیق و الگوریتم‌های یادگیری تقویتی، چارچوب‌های یکپارچه‌تری برای مدیریت شارژ ارائه داده‌اند. با این حال، این روش‌ها معمولاً اطلاعات ایستگاه‌های شارژ، محدودیت‌های عملی و شرایط واقعی شبکه را به طور کامل در نظر نمی‌گیرند [۲۱].

در مجموع، با وجود پیشرفت‌های قابل‌توجه در پژوهش‌های پیشین، همچنان نیاز به چارچوبی جامع وجود دارد که بتواند پیش‌بینی دقیق شرایط انرژی را با تصمیم‌گیری هوشمند بی‌درنگ ترکیب کند. در این راستا، پژوهش حاضر با هدف پاسخ‌گویی به این خلأ، رویکردی مبتنی بر تلفیق مدل‌های پیش‌بینی یادگیری عمیق و الگوریتم‌های یادگیری تقویتی برای زمان‌بندی بهینه شارژ وسایل نقلیه برقی ارائه می‌دهد.

۳- فرمول‌بندی مسأله

در این مقاله، مسأله زمان‌بندی شارژ و دشارژ وسایل نقلیه برقی برای افق زمانی روز آینده از دیدگاه کاربر مورد بررسی قرار می‌گیرد. هدف اصلی، کاهش هزینه کل انرژی مصرفی با در نظر گرفتن محدودیت‌های زمانی و میزان انرژی مورد نیاز وسیله نقلیه برقی است. با توجه به ماهیت تصادفی زمان ورود و خروج خودرو، میزان تقاضای روزانه انرژی و نوسانات قیمت برق، این مسأله در قالب یک فرآیند تصمیم‌گیری مارکف (MDP)[۱۱] مدل‌سازی می‌شود. در این چارچوب، تصمیم‌گیری در هر گام زمانی به‌گونه‌ای انجام می‌شود که پاداش تجمعی[۱۲] بیشینه گردد. برخلاف مدل‌های متداول MDP که عمدتاً از اطلاعات لحظه‌ای سیستم استفاده می‌کنند، در این پژوهش علاوه بر متغیرهای اصلی سیستم، اطلاعات ایستگاه شارژ و تاریخچه ۲۴ ساعته قیمت برق و مصرف انرژی نیز در بردار حالت لحاظ شده‌اند که موجب نمایش دقیق‌تر وضعیت محیط و بهبود کیفیت تصمیم‌گیری عامل می‌شود.

ساختار فرآیند تصمیم‌گیری مارکف به گونه‌ای است که تصمیم‌گیری در هر گام زمانی بر اساس موقعیت فعلی خودرو، قیمت لحظه‌ای برق و نوع شارژر (سریع یا کند) انجام می‌شود. علاوه بر این، در فرآیند تصمیم‌گیری، محدودیت‌های ایستگاه‌های شارژ از جمله ظرفیت محدود ایستگاه و وضعیت در دسترس بودن شارژرها به‌صورت شفاف در نظر گرفته شده‌است. هر خودرو تنها در صورتی مجاز به انتخاب یک ایستگاه و نوع شارژ مشخص است که ظرفیت کافی در آن ایستگاه وجود داشته باشد؛ در غیر این صورت، خودرو باید ایستگاه یا بازه زمانی جایگزین را انتخاب کند. به دلیل وجود این محدودیت‌های ظرفیتی، مسأله زمان‌بندی شارژ دارای ماهیت رقابتی است، به‌گونه‌ای که تصمیم شارژ هر خودرو می‌تواند گزینه‌های قابل انتخاب سایر خودروها را محدود کند. ساختار نمایش حالات و فرآیند تصمیم‌گیری در این مدل براساس چارچوب ارائه‌شده مرجع [۶] طراحی شده و در این پژوهش توسعه یافته‌است تا تمامی حالات و شرایط عملیاتی سیستم به‌طور جامع پوشش داده شوند. اجزای اصلی مدل MDP به شرح زیر تعریف می‌شوند:

۱) حالت: حالت سیستم در زمان t به صورت برداری از چندین متغیر به صورت زیر تعریف می‌شود:

$$s_t = (Start_{time}, End_{time}, SOC_t, SOC_{ideal}, ID, (P_{t-23}, \dots, P_t), (E_{t-23}, \dots, E_t)) \quad (1)$$

که در آن، $Start_{time}$ زمان شروع شارژ، End_{time} زمان پایان شارژ، SOC_t[۱۳] سطح شارژ فعلی باتری، SOC_{ideal} سطح شارژ مطلوب، $(P_{t-23}, \dots, P_t)$ توالی قیمت برق در ۲۴ ساعت گذشته و $(E_{t-23}, \dots, E_t)$ توالی مصرف انرژی در ۲۴ ساعت گذشته است. ID به عنوان شناسه ایستگاه شارژ به‌صورت یک مقدار عدد صحیح در بردار حالت، بیانگر وضعیت انتخاب ایستگاه و محدودیت‌های ظرفیت آن در هر گام زمانی است و نقش آن در مدل، اعمال محدودیت‌های محیط و تعریف فضای حالت گسسته است. با توجه به این که تمامی اطلاعات لازم شامل وضعیت باتری، زمان، قیمت و محدودیت ایستگاه در بردار حالت لحاظ شده‌است، شرط مارکف برقرار بوده و حالت فعلی اطلاعات کافی برای تصمیم‌گیری آینده را فراهم می‌کند. همچنین، متغیرهای پیوسته شامل سطح شارژ باتری (SOC)، قیمت برق و مصرف انرژی پیش از ورود به مدل با استفاده از روش Min-Max در بازه [۰،۱] نرمال‌سازی می‌شوند تا اثر اختلاف مقیاس داده‌ها کاهش یابد.

۲) عمل: عمل یا act_t نشان دهنده توان شارژ یا دشارژ خودرو در زمان t است که به صورت $-P \leq act_t \leq P$ تعریف می‌شود و در آن P

حداکثر توان مجاز است. مقدار مثبت act_t بیانگر شارژ، مقدار منفی act_t بیانگر دشارژ و act_t=۰ نشان‌دهنده عدم انجام عمل است. کاربر می‌تواند با شارژ باتری در ساعت کف قیمت و دشارژ در ساعات اوج قیمت، هزینه انرژی خود را کاهش دهد.

۳) انتقال حالت: انتقال حالت از S_t به حالت بعدی S_{t+1} به صورت زیر مدل‌سازی می‌شود:

$$s_{t+1} = f\,(s_t, act_t) \qquad (۲)$$

پس از انتخاب عمل (act_t) در زمان t، در هر گام زمانی، عامل یادگیرنده، بردار انتقال (s_t, act_t, rew_t, s_{t+1}) را ذخیره می‌کند که شامل حالت فعلی (S_t)، عمل انجام شده (act_t)، پاداش دریافتی (rew_t) و حالت بعدی (S_{t+1}) است. این تجربیات برای به‌روزرسانی مدل یادگیری و بهبود سیاست شارژ و دشارژ استفاده می‌شوند. مقدار SOC در هر گام زمانی بر اساس توان شارژ/دشارژ انتخاب‌شده توسط عامل و طول گام زمانی در محیط به‌روزرسانی می‌شود و به‌صورت افزایشی یا کاهشی در وضعیت بعدی سیستم اعمال می‌گردد. علاوه بر SOC، متغیرهای مرتبط با قیمت برق و مصرف انرژی نیز در هر گام زمانی مطابق با ساختار پنجره لغزان ۲۴ ساعته به‌روزرسانی شده و وارد حالت جدید سیستم می‌شوند، به‌گونه‌ای که اطلاعات زمانی جدید جایگزین مقادیر قدیمی‌تر شده و پویایی زمانی محیط حفظ می‌گردد.

۴) پاداش: تابع پاداش از دیدگاه کاربر در زمان t به صورت زیر تعریف می‌شود:

$$Rew_t = a.P_t.act_t - \begin{cases} b.C^{act_t}, & t = t_{Dep} \\ b.C^{act_t} - \delta(C_{max} - SOC_t)^2, & t \neq t_{Dep} \end{cases} \qquad (۳)$$

که در آن P_t قیمت برق در زمان t، act_t توان شارژ و دشارژ، SOC_t سطح شارژ باتری و C_{max} ظرفیت بیشینه باتری است. دو مولفه a و b ضرایب هزینه برق و جریمه سطح شارژ باتری را تعیین می‌کنند که در این مقاله برابر با یک در نظر گرفته شده‌اند. مقادیر a و b به‌صورت تجربی و بر اساس آزمون‌های مختلف تنظیم شده‌است، به‌گونه‌ای که مقدار انتخاب‌شده منجر به ایجاد تعادل مناسب بین هزینه انرژی و جریمه سطح شارژ باتری و همچنین پایداری فرآیند یادگیری گردد. اگر خودرو برقی در بازه انتخاب شده به موقع به ایستگاه شارژ برسد و به سطح شارژ مطلوب دست یابد، جریمه‌ای در نظر گرفته نمی‌شود؛ در غیر این صورت، به دلیل کمبود زمان، مؤلفه جریمه فعال شده و مقدار پاداش کاهش می‌یابد. مولفه جریمه، در صورتی فعال می‌شود که باتری خودرو تا زمان خروج به سطح شارژ مطلوب نرسیده باشد. پاداش در بازه زمانی کامل شارژ یعنی از زمان ورود خودرو (t_{Arr}) تا زمان خروج خودرو t_{Dep} محاسبه می‌شود.

ضریب جریمه (δ) بایستی دارای واحد دلار بر کیلووات‌ساعت مربع ($/kWh²) باشد. این انتخاب به این دلیل است که این ضریب در عبارت $(E_{max} - E_t)^2$ ضرب می‌شود که خود دارای واحد kWh^2 است؛ در نتیجه، حاصل‌ضرب آن دارای واحد دلار ($) خواهد بود. به این ترتیب، این مؤلفه از نظر ابعادی با سایر اجزای تابع پاداش مانند هزینه شارژ هم‌واحد می‌شود. قیود ظرفیت ایستگاه به‌صورت محدودیت محیطی اعمال شده و در صورت انتخاب ایستگاه غیرقابل دسترس، مقدار پاداش منفی یا اقدام جایگزین اعمال می‌شود.

۵) تابع عمل-مقدار[۱۴]: این تابع $Q(s,a)$ به صورت تکراری و با استفاده از قانون به‌روزرسانی الگوریتم یادگیری Q به‌روزرسانی می‌شود:

$$Q(s_t, act_t) \leftarrow (1-\alpha).Q(s_t, act_t) + \alpha((rew_t + \gamma. \max_{act'} Q(s_t', act_t')) \qquad (۴)$$

که در آن، α ضریب یادگیری جدول Q، rew_t پاداش دریافتی، γ ضریب تنزیل و $\max_{act'} Q(s_t', act_t')$ بیشترین مقدار تابع Q در زمان بعدی است.

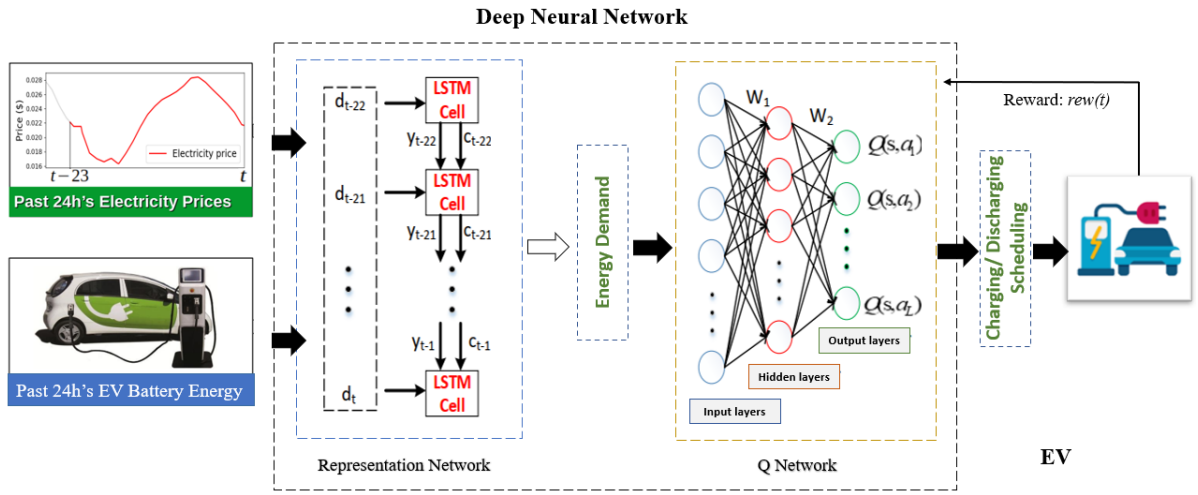
۴- روش پیشنهادی

تعیین سیاست بهینه برای بیشینه‌سازی پاداش تجمعی در شرایط عدم قطعیت رفتار کاربران و تغییرات قیمت برق دشوار است. برای حل این مسأله، در این مقاله از یک رویکرد یادگیری تقویتی مبتنی بر معادله بلمن ارائه‌شده در [۲۲] استفاده می‌شود، به‌طوری که تابع عمل-مقدار $Q(s,a)$ به‌صورت مکرر به سمت مقدار بهینه همگرا می‌شود، به‌گونه‌ای که با افزایش تعداد تکرارها ($i\to\infty$) مقدار آن به مقدار بهینه نزدیک می‌شود. در این روش، یک شبکه عصبی مقادیر $Q(s,a)$ را برای تمام اعمال شارژ و دشارژ با توجه به وضعیت فعلی سیستم (قیمت برق، میزان مصرف انرژی و تقاضای شارژ) تخمین می‌زند و در نهایت عملی که بیشترین مقدار تخمینی را دارد به‌عنوان برنامه بهینه انتخاب می‌شود.

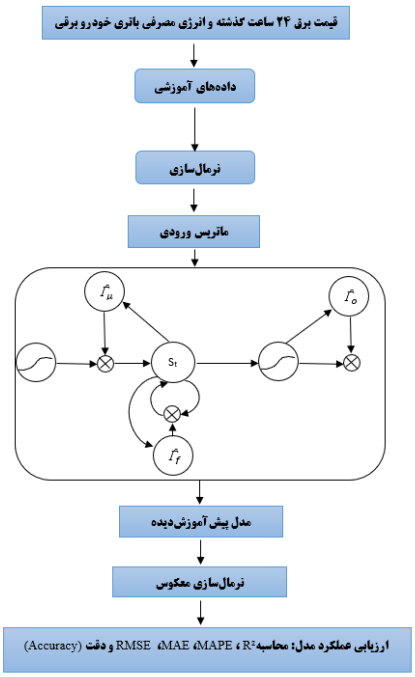
۴-۱- معماری شبکه عصبی عمیق

شمای کلی معماری پیشنهادی در شکل ۱ نشان داده شده‌است. ورودی‌های اصلی شامل قیمت برق و انرژی مصرفی در ۲۴ ساعت گذشته است. این معماری شامل دو زیر شبکه اصلی است:

۱) شبکه نمایش[۱۵]: این شبکه ویژگی‌ها را از داده‌های قیمت برق و مصرف انرژی با استفاده از شبکه LSTM استخراج می‌کند. شبکه LSTM از ۵ لایه تشکیل شده که شامل یک لایه ورودی، سه لایه



شکل ۱: معماری کلی روش پیشنهادی داده‌محور برای مدیریت شارژ و دشارژ وسایل نقلیه الکتریکی.



شکل ۲: چارچوب پیشنهادی برای شبکه پیش‌بینی مبتنی بر LSTM

مخفی و یک لایه خروجی است. در لایه ورودی، داده‌های قیمت برق و مصرف انرژی روز آینده به شبکه داده می‌شوند. لایه‌های مخفی که هر کدام دارای تعداد مشخصی نورون هستند، امکان استخراج وابستگی‌های زمانی موجود در داده‌ها را فراهم می‌کنند. در نهایت، لایه خروجی مقادیر پیش‌بینی‌شده (y_t) را تولید می‌کند. هر سلول LSTM شامل چند مؤلفه داخلی (دروازه‌ها[۱۶] و حالات) است که جریان اطلاعات را در طول زمان مدیریت کرده و یادگیری الگوهای متوالی را بهینه می‌سازند. دروازه فراموشی[۱۷] (Γf) تعیین می‌کند کدام اطلاعات از حالت قبلی کنار گذاشته شود. دروازه ورودی یا به‌روزرسانی (Γu) مشخص می‌کند چه مقدار اطلاعات جدید به حالت سلول (c_t) اضافه گردد و دروازه خروجی (Γo) کنترل می‌کند کدام بخش از حالت سلول به خروجی منتقل شود. در نهایت، حالت سلول حافظه بلندمدت شبکه را ذخیره و امکان نگهداری اطلاعات در گام‌های زمانی طولانی را فراهم می‌نماید. این فرآیند در بخش میانی شکل ۲ نشان داده شده‌است.

برای پیش‌پردازش داده‌ها، بهبود فرآیند آموزش و جلوگیری از حساسیت شبکه به مقادیر بزرگ، داده‌ها با استفاده از MinMaxScaler به بازه‌ی [۰،۱] نرمال‌سازی می‌شوند. سپس با استفاده از پارامتر پنجره لغزان[۱۸]، داده‌ها به توالی‌های زمانی مناسب برای ورود به شبکه تبدیل می‌گردند. در طول فرآیند آموزش، از تکنیک «توقف زودهنگام»[۱۹] به منظور جلوگیری از بیش‌برازش[۲۰] استفاده می‌شود و مقدار تابع خطا در هر دوره[۲۱] ذخیره می‌گردد. پس از اتمام دوره آموزش، مدل روی داده‌های آزمایش اعمال شده و نتایج پیش‌بینی‌شده از داده‌های آزمایش با مقادیر واقعی، با استفاده از معیارهایی مانند R^2، MAPE[۲۲]، MAE[۲۳] و RMSE[۲۴] ارزیابی می‌شوند. این فرآیند در بخش انتهایی شکل ۲ نشان داده شده‌است. ضریب تعیین (R^2) نشان می‌دهد که مدل تا چه حد توانسته تغییرات داده‌های واقعی را پیش‌بینی کند. مقدار آن بین ۰ و ۱ است و هرچه مقدار R^2 به یک نزدیک‌تر باشد، مدل دقت بالاتری دارد. این معیار مطابق معادله (۵) تعریف می‌شود:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y_i})^2} \qquad (۵)$$

معیار درصد خطای میانگین مطلق (MAPE) خطای پیش‌بینی را به‌صورت درصد بیان می‌کند و مناسب برای مقایسه خطا در داده‌هایی با مقیاس متفاوت است:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y_i}}{y_i}\right| \qquad (۶)$$

میانگین خطای مطلق (MAE) میانگین اختلاف مطلق بین مقادیر واقعی و پیش‌بینی‌شده را محاسبه می‌کند. مقدار پایین MAE به معنی پیش‌بینی دقیق‌تر مدل است:

$$MAE = \sum_{i=1}^{n} \frac{|y_i - \hat{y_i}|}{n} \quad (۷)$$

معیار ریشه میانگین مربعات خطا (RMSE) حساس به خطاهای بزرگ است و اختلاف‌های بزرگ‌تر را بیشتر جریمه می‌کند. مقدار پایین RMSE به معنی دقت بالاتر مدل است:

$$RMSE = \sqrt{\frac{\sum(y_i - \hat{y_i})^2}{n}} \quad (۸)$$

این معیارها ارزیابی جامعی از دقت پیش‌بینی و ویژگی‌های خطای مدل ارائه می‌دهند. جزئیات بیشتر مربوط به تعاریف ریاضی و فرمول‌بندی این معیارها در مرجع [۱۴] بیان شده‌است.

پارامترهای به‌کاررفته در شبکه پیش‌بینی LSTM در جدول ۱ نشان داده شده‌است. این تنظیمات بر اساس آزمایش‌های اولیه انتخاب شده‌اند تا توازن میان پایداری آموزش و دقت پیش‌بینی حفظ شود. پارامتر حذف تصادفی نورون‌ها[۲۵] به منظور جلوگیری از بیش‌برازش، ٪۲۰ از نورون‌ها در هر لایه را به صورت تصادفی غیرفعال می‌کند. پارامتر پنجره لغزان تعیین می‌کند که برای پیش‌بینی هر گام آینده، از چند داده زمانی قبلی استفاده شود؛ در این مقاله، از ۲۴ داده گذشته برای پیش‌بینی گام بعدی استفاده می‌شود. بهینه‌سازی وزن‌ها با استفاده از الگوریتم Adam انجام می‌شود. این الگوریتم با استفاده از گرادیان‌ها و نرخ یادگیری شبکه، گام‌های به‌روزرسانی پارامترها را به‌صورت تطبیقی تنظیم می‌کند. این کار به بهبود همگرایی پاداش تجمعی و افزایش سرعت آموزش کمک می‌کند و احتمال گیر افتادن در کمینه‌های محلی را کاهش می‌دهد. خروجی شبکه یک مقدار است (مانند قیمت برق در یک گام زمانی مشخص) و اندازه لایه‌های پنهان نشان‌دهنده تعداد نورون‌ها در هر لایه است که ظرفیت شبکه برای یادگیری و نگهداری الگوهای متوالی سری‌های زمانی را مشخص می‌کند.

۲) شبکه Q: برای پیاده‌سازی مدل پیشنهادی به منظور زمان‌بندی شارژ و دشارژ خودروهای برقی، از شبکه Q عمیق (DQN) استفاده شده‌است. در این مدل دو شبکه عصبی مجزا به کار گرفته شده‌است شامل شبکه ارزیاب که تصمیمات عامل را ارزیابی می‌کند و شبکه هدف که به‌طور دوره‌ای به‌روزرسانی می‌شود تا به پایدارتر شدن فرآیند یادگیری کمک کند. پس از استخراج ویژگی‌ها توسط شبکه LSTM، این ویژگی‌ها با داده میزان تقاضای قابل پاسخ‌گویی ایستگاه شارژ ترکیب شده و یک برنامه زمان‌بندی اولیه تولید می‌شود. سپس این ویژگی‌ها به شبکه Q داده شده و شبکه یک اقدام بهینه شارژ یا دشارژ را برای هر بازه زمانی انتخاب می‌کند.

شبکه Q شامل یک لایه ورودی، سه لایه پنهان با ۶۴ نورون و یک لایه خروجی است. لایه ورودی شامل قیمت برق پیش‌بینی‌شده، میزان انرژی مصرفی پیش‌بینی‌شده و داده‌های میزان تقاضا قابل پاسخ‌گویی است. لایه خروجی هفت مقدار گسسته تولید می‌کند که سطوح مختلف شارژ را نشان می‌دهد و به شبکه امکان می‌دهد الگوهای پیچیده را یاد گرفته و تابع Q را تقریب بزند.

در حالی که الگوریتم DQN به دلیل سادگی و سرعت همگرایی برای فضاهای عمل گسسته و محیط‌های کوچک مناسب است، روش‌های پیشرفته‌تری مانند PPO [۱۶] و SAC [۳] قادر به مدیریت فضاهای عمل پیوسته بوده و از پایداری بالاتر و عملکرد بهتر در محیط‌های پیچیده یا نویزی برخوردارند. با این حال، در این مقاله از DQN استفاده شده‌است، زیرا اقدامات شارژ خودروهای برقی گسسته‌اند و محیط نسبتاً کوچک است که آموزش سریع‌تر و پیاده‌سازی ساده‌تر مدل را امکان‌پذیر می‌سازد.

پارامترهای کلیدی مورد استفاده در فرآیند آموزش الگوریتم DQN را نشان می‌دهد. این مقادیر به‌گونه‌ای تنظیم شده‌اند که ضمن تضمین همگرایی مدل، از بیش‌برازش و ناپایداری در فرآیند یادگیری جلوگیری می‌کنند. پارامتر ε نقش کلیدی در تعادل میان اکتشاف[۲۶] و بهره‌برداری[۲۷] در طول آموزش دارد. در ابتدای فرآیند آموزش، عامل اغلب به‌صورت تصادفی عمل می‌کند تا محیط را بهتر

جدول ۱: پارامترهای شبکه LSTM

پارامتر	مقدار
اندازه دسته[۲۸]	۲۵۶
تابع هزینه	MSE-Loss
حذف تصادفی نورون‌ها	۰/۲
تابع فعال‌سازی[۲۹]	Linear
بهینه‌ساز	Adam
تعداد کل لایه‌ها	۵ (ورودی، ۳ لایه پنهان، خروجی)
اندازه لایه پنهان	۶۴
اندازه خروجی	۱
نرخ یادگیری شبکه	۰/۰۰۱
پنجره لغزان	۲۴

بشناسد. این پارامتر تعیین می‌کند که در شروع، چقدر از رفتار عامل تصادفی باشد. مقدار یک یعنی عامل در ابتدا فقط به دنبال اکتشاف محیط است. در طول آموزش، مقدار ε به تدریج و با نرخ ۰/۹۹۵ کاهش می‌یابد تا عامل بیشتر از دانش قبلی خود استفاده کند. با

این حال، مقدار ε نباید از حداقل مقدار تعیین‌شده یعنی ۰/۰۱ کمتر شود تا عامل همچنان بتواند فرآیند کاوش در محیط را ادامه داده و دانش جدیدی کسب کند. بنابراین، پس از هر دوره، مقدار ε طبق معادله زیر به‌روزرسانی می‌شود:

$$\varepsilon = \max(\varepsilon \times \varepsilon_DECAY_RATE, \varepsilon_MIN) \quad (۹)$$

شبکه هدف در فواصل زمانی مشخص وزن‌های شبکه خود را با وزن‌های شبکه ارزیابی[۳۰] به‌روزرسانی می‌کند. این کار باعث ثبات در فرآیند یادگیری می‌شود. حداکثر تعداد گام‌ها در هر دوره، بیشترین تعداد گامی است که عامل می‌تواند در هر دوره بردارد. در این مقاله، مقدار آن ۲۴ در نظر گرفته شده‌است که معادل یک برنامه ۲۴ ساعته برای شارژ و دشارژ خودروهای برقی است.

۴-۲- فرآیند آموزش

الگوریتم ۱ فرآیند آموزش شبکه عصبی عمیق برای تعیین سیاست بهینه شارژ و دشارژ خودروهای برقی را بیان می‌کند. قبل از شروع آموزش، محیط با مقادیر تصادفی برای هر عامل مقداردهی اولیه می‌شود. این مقادیر اولیه در جدول Q برای شبیه‌سازی تجربیات کاربر ذخیره شده و به الگوریتم اجازه می‌دهد تا یک سیاست بهینه را یاد بگیرد که شامل زمان ورود خودرو (t_{Arr})، زمان خروج خودرو (t_{Dep})، وضعیت فعلی شارژ باتری (SOC)، مقدار SOC مطلوب و وضعیت شارژ است.

پارامترهای شبکه عصبی عمیق با θ_t نشان داده می‌شوند. این پارامترها برای هر دو شبکه، یعنی شبکه نمایش و شبکه Q هستند و ابتدا به صورت تصادفی مقداردهی شده، سپس یک شبکه هدف با همان ساختار ایجاد می‌شود و پارامترهای آن (θ_target) به‌عنوان نمونه‌ای از θ قرار داده می‌شوند. شبکه هدف در طول آموزش به‌طور دوره‌ای با شبکه اصلی به‌روز می‌شود تا فرآیند آموزش پایدارتر شود. شبکه متناظر با بازه‌های زمانی حضور هر خودرو یعنی از زمان ورود خودرو t_{Arr} و تا زمان خروج در ایستگاه شارژ آموزش داده می‌شود، به‌طوری که طول هر دوره برابر $t_{Dep} - t_{Arr}$ است. در هر گام زمانی، یک عمل act_t، نمایانگر توان شارژ یا دشارژ است که طبق سیاست ε-greedy انتخاب می‌شود؛ به این صورت که با احتمال ε یک عمل تصادفی برای تشویق به اکتشاف، انتخاب می‌شود و با احتمال ε−۱ ، عملی که شبکه عصبی پیشنهاد می‌دهد، برای بهره‌برداری از دانش کسب شده انتخاب می‌گردد. پس از اجرای عمل انتخاب‌شده، پاداش متناظر rew_t با عمل دریافت و سیستم به وضعیت بعدی s_{t+1} منتقل می‌شود. هر انتقال (s_t,act_t,rew_t,s_{t+1}) در حافظه بازپخش[۳۱] ذخیره می‌گردد تا برای به‌روزرسانی پارامترهای شبکه مورد استفاده قرار

جدول ۲: پارامترهای شبکه Q

پارامتر	مقدار
اندازه دسته	۲۵۶
نرخ یادگیری شبکه (Lr)	۰/۰۰۱
اپسیلون	۱/۰
حداقل مقدار اپسیلون	۰/۰۱
نرخ کاهش اپسیلون	۰/۹۹۵
ضریب تنزیل (گاما)[۳۲]	۰/۹۹
تعداد تکرار برای به‌روزرسانی شبکه هدف[۳۳]	۵
اندازه حافظه	۱۰۰۰۰۰
حداکثر تعداد گام‌ها در هر دوره	۲۴
ضریب یادگیری (آلفا)	۰/۱
تعداد کل لایه‌ها	۵ (ورودی، ۳ لایه پنهان، خروجی)
اندازه لایه‌های پنهان	۶۴

الگوریتم ۱: آموزش شبکه عصبی عمیق

ورودی: قیمت برق، انرژی برنامه ریزی شده روزآینده، پاداش (rew)، حداکثر تعداد گام در هر دوره ($Max_steps = 24$) و دوره به روزرسانی شبکه هدف (T)

خروجی: پارامترهای شبکه عصبی θ

۱: مقداردهی اولیه تصادفی پارامترهای شبکه (θ)

۲: مقداردهی اولیه شبکه هدف: $\theta_{target} \leftarrow \theta$

۳: شروع حلقه دوره آموزش: برای هر دوره (Epoch = 1) تا (Max_steps) تکرار کن:

۴: دریافت حالت اولیه سیستم ($s_{t_{Arr}}$)

۵: شروع حلقه زمانی: برای هر گام زمانی از زمان شروع شارژ(t_{Arr}) تا زمان پایان شارژ (t_{Dep}) تکرار کن:

۶: انتخاب عمل (act_t) با استفاده از سیاست ε-greedy

۷: اعمال عمل انتخاب‌شده (act_t)، دریافت پاداش (rew_t) و انتقال به حالت سبعدی (s_{t+1})

۸: ذخیره انتقال ($s_t, act_t, rew_t, s_{t+1}$) در حافظه بازپخش (replay memory)

۹: نمونه‌گیری یک مینی‌بچ از حافظه بازپخش:
$F = \{(s_j, act_j, rew_j, s_{j+1})\}_{j=1}^{\#F}$

۱۰: به‌روزرسانی مقدار Q بر اساس رابطه:
$Q(s_t, act_t) \leftarrow (1-\alpha).\ Q(s_t, act_t) + \alpha\ ((rew_t + \gamma\ .\ \max_{act'} Q(s_t', act_t'))$

۱۱: محاسبه تابع خطا روی مینی‌بچ:
$\text{Loss} = \sum_{i=1}^{N} \frac{1}{N} \left(\left[\gamma\ .\ max_{act'} Q_{target}(s_i', act_i') + rew_i \right] - \left(Q_{eval}(s_i, act_i) \right) \right)^2$

۱۲: به‌روزرسانی پارامترهای شبکه:
$\theta_{t+1} \leftarrow \theta_t - \mu \cdot \nabla_{\theta_t} \text{Loss}(\theta_t)$

۱۳: هر T گام، همگام‌سازی شبکه هدف: $\theta_{target} \leftarrow \theta$

۱۴: انتقال به حالت بعدی: $s_t \leftarrow s_{t+1}$

۱۵: پایان حلقه زمانی

۱۶: پایان حلقه دوره آموزش

گیرد. در هر گام زمانی، یک دسته کوچک[۳۴] از این انتقال‌ها از حافظه بازپخش نمونه‌برداری شده و تابع خطا بر اساس اختلاف بین مقادیر Q پیش‌بینی شده و مقادیر Q هدف طبق معادله (۱۰) محاسبه می‌شود:

$$Loss = \sum_{i=1}^{N} \frac{1}{N} \Big([\gamma . max_{act'} Q_{target}(s_i', act') + rew_i] - (Q_{eval}(s_i, act_i)) \Big)^2 \quad (۱۰)$$

پارامترهای شبکه با استفاده از گرادیان نزولی بر اساس این خطا به‌روزرسانی می‌شوند (معادله ۱۱). در این معادله، ∇_{θ_t}. Loss (θ_t) گرادیان تابع خطا نسبت به تمام پارامترهای شبکه θ، شامل وزن‌ها و بایاس‌های هر لایه است. این گرادیان مشخص می‌کند که پارامترها چگونه باید تغییر کنند تا اختلاف بین مقادیر Q پیش‌بینی شده و مقادیر Q هدف کاهش یابد:

$$\theta_{t+1} \leftarrow \theta_t - \mu \cdot \nabla_{\theta_t} . Loss\ (\theta_t) \quad (۱۱)$$

علاوه بر این، شبکه هدف هر T گام با شبکه اصلی به‌روزرسانی می‌شود تا فرآیند آموزش پایدارتر شود. این روند به شبکه اجازه می‌دهد تا به تدریج یک سیاست بهینه برای برنامه‌ریزی شارژ و دشارژ خودروهای برقی یاد بگیرد.

۴-۳- زمان‌بندی روزانه شارژ/دشارژ خودروهای برقی

الگوریتم ۲ فرآیند پیشنهادی برای زمان‌بندی روزانه شارژ و دشارژ خودروهای برقی را نشان می‌دهد. ابتدا، پارامترهای شبکه عصبی از پیش‌آموزش‌دیده شده در الگوریتم ۱ بارگذاری می‌شوند. سپس، در هر گام زمانی از زمان ورود خودرو t_{Arr} تا زمان خروج t_{Dep}، اطلاعاتی مانند قیمت برق ۲۴ ساعت گذشته، انرژی برنامه‌ریزی‌شده روز آینده و میزان تقاضای انرژی استخراج می‌شود. شبکه نمایش ویژگی‌های قیمت برق و انرژی برنامه‌ریزی‌شده را استخراج کرده و آن‌ها را با داده‌های میزان تقاضای انرژی ترکیب می‌کند. در مرحله بعد، شبکه Q مقادیر عمل را محاسبه کرده و عمل بهینه را انتخاب می‌کند. عمل انتخاب‌شده به‌عنوان برنامه شارژ/دشارژ خودرو برای بازه زمانی مربوطه در نظر گرفته می‌شود.

۱-۱- تنظیمات آموزش و ارزیابی

چارچوب پیشنهادی که شامل مدل پیش‌بینی مبتنی بر LSTM و شبکه زمان‌بندی مبتنی بر DQN است، برای ۳۰۰۰ دوره آموزش و ارزیابی شد. در طول آموزش، مقدار اپسیلون به‌طور تدریجی از ۱ به ۰/۱ کاهش می‌یابد. در دوره‌های ابتدایی که مقدار اپسیلون برابر با ۱ است، مدل بیشتر به دنبال اکتشاف و امتحان گزینه‌های جدید است. اما در ادامه فرآیند آموزش، مقدار اپسیلون به‌تدریج با نرخ ۰/۹۹۵ کاهش یافته و در نهایت به مقدار ۰/۱ می‌رسد. این کاهش مرحله به مرحله اپسیلون باعث می‌شود مدل از مرحله اکتشاف به سمت بهره‌برداری برود و با استفاده از تجربیات قبلی، تصمیمات بهتری بگیرد.

تابع پاداش به گونه‌ای طراحی شده که با محدود کردن سطح شارژ اولیه باتری (SOC) در هنگام ورود به ایستگاه، هزینه عملیاتی را کاهش داده و جریمه را به حداقل برساند. این محدودیت، فرآیند یادگیری را به سمت برنامه‌های کم هزینه هدایت می‌کند. در نتیجه، پاداش تجمعی همگرا شده و مدل به انتخاب‌های مؤثرتری برای شارژ و دشارژ خودروهای برقی دست می‌یابد.

۵- نتایج تجربی

در این بخش، عملکرد چارچوب پیشنهادی زمان‌بندی شارژ و دشارژ خودروهای برقی مبتنی بر DQN با استفاده از داده‌های واقعی ایالت کالیفرنیا ارزیابی می‌شود. ارزیابی مدل در بیشینه‌سازی پاداش تجمعی و کاهش هزینه‌های برق تحت تنظیمات و پیکربندی‌های شبکه بررسی می‌شود. داده‌های ساعتی قیمت برق و انرژی مصرفی مربوط به سال ۲۰۲۲ هستند. برای هر ماه، ۲۰ روز اول برای آموزش و ۱۰ روز آخر برای آزمایش استفاده می‌شوند؛ این تقسیم‌بندی داده‌ها مطابق با استراتژی استفاده‌شده در مطالعات پیشین است.

الگوریتم ۲: زمان‌بندی شارژ و دشارژ روزانه وسایل نقلیه برقی

ورودی: حالت سیستم (s_t) شامل قیمت برق، انرژی برنامه‌ریزی‌شده روز آینده و میزان تقاضای انرژی

خروجی: برنامه زمان‌بندی شارژ/دشارژ خودرو برقی در بازه زمانی ($act_{t_{Arr}:t_{Dep}}$)

۱: بارگذاری پارامترهای شبکه آموزش‌دیده (θ) از الگوریتم ۱

۲: شروع حلقه زمانی: برای هر گام زمانی از (t_{Arr}) تا (t_{Dep}) تکرار کن:

۳: استخراج قیمت برق ۲۴ ساعت گذشته و انرژی برنامه‌ریزی‌شده روز آینده

۴: استفاده از شبکه استخراج نمایش (Representation Network) برای استخراج ویژگی‌ها از قیمت برق و انرژی برنامه‌ریزی‌شده

۵: ادغام این ویژگی‌ها با داده میزان تقاضای انرژی برای تشکیل بردار حالت (s_t)

۶: محاسبه مقدار Q برای تمام اعمال ممکن با استفاده از شبکهQ :

$$Q_{value} \leftarrow \mathrm{Q}\ (s_t,\ act_t; \theta)$$

۷: انتخاب عملی که بیشترین مقدار Q را دارد:

$$act_t \leftarrow argmax_{act}\ Q\ (s_t', act_t')$$

۸: اجرای عمل شارژ/ دشارژ خودرو برقی (act_t)

۹: پایان حلقه زمانی

۱۰: بازگرداندن کل برنامه زمان‌بندی ($act_{t_{Arr}:t_{Dep}}$)

۵-۱- تنظیمات اولیه

در این مقاله، یک خودروی برقی در چارچوب زمان‌بندی مبتنی بر DRL در نظر گرفته شده‌است که سطح شارژ باتری آن بین ۱ (حداقل شارژ) و ۲۴ (حداکثر شارژ) تغییر می‌کند. هنگام ورود به ایستگاه شارژ هوشمند، خودرو می‌تواند از میان هفت سطح گسسته شارژ/دشارژ که به صورت {۰/۰۳ ، ۰/۰۲ ، ۰/۰۱ ، ۰ ، ۰/۰۱- ، ۰/۰۲- ، ۰/۰۳-} مگاوات تعریف شده‌اند، یک سطح را انتخاب کند. مقادیر مثبت نشان‌دهنده شارژ، مقادیر منفی نشان‌دهند دشارژ و صفر نشان‌دهنده عدم انجام عمل است.

لایه خروجی شبکه DQN شامل هفت واحد است که متناظر با هفت سطح شارژ/دشارژ موجود است. پارامترهای شبکه با استفاده از توزیع نرمال مقداردهی اولیه شده و در طول آموزش با استفاده از الگوریتم بهینه‌سازی مناسب به‌روزرسانی می‌شوند. ضریب تنزیل (γ) برابر ۰/۹۹ در نظر گرفته شده‌است تا هم پاداش‌های فوری و هم پاداش‌های آینده در تصمیم‌گیری لحاظ شوند. کل این کار با استفاده از زبان برنامه‌نویسی پایتون و چارچوب یادگیری عمیق PyTorch توسعه یافته است.

۵-۲- داده‌های قیمت برق و مصرف انرژی

این داده‌ها مربوط به سال ۲۰۲۲ بوده و از وب‌سایت رسمی اپراتور مستقل سیستم برق ایالت کالیفرنیا (CAISO)[۳۵] [۲۳] دریافت شده‌اند. داده قیمت برق به‌صورت پنج دقیقه‌ای ثبت شده‌است، به‌گونه‌ای که در هر ساعت ۱۲ مقدار و در هر روز ۲۸۸ مقدار قیمت ثبت می‌شود. این داده‌ها شامل مقادیر مثبت، منفی و صفر هستند. مقادیر صفر و منفی معمولاً به‌عنوان داده‌های اتمی شناخته می‌شوند که کنترلشان دشوار است. با این حال، مقادیر منفی در برخی شرایط به پایداری شبکه کمک کرده و از نظر اقتصادی برای خریداران انرژی سودمند هستند. داده انرژی، نشان‌دهنده اطلاعات مربوط به برنامه‌ریزی روز آینده هستند. این داده‌ها میزان تقاضا نیروگاه‌ها، مصرف‌کنندگان و تولیدکنندگان برق را برای روز بعد تعیین می‌کنند. ساختار زمانی داده‌های انرژی مشابه داده‌های قیمت برق بوده و شامل ۲۸۸ مقدار برای هر روز است. این مجموعه داده در مطالعات پیشین مانند [۳, ۵, ۲۴, ۲۵] استفاده شده‌اند.

مقادیر مربوط به میزان تقاضای انرژی بر حسب مگاوات است. این مقادیر از داده‌های پیش‌بینی‌شده یک روزه CAISO[۳۶] دریافت شده و به‌صورت زمان‌بندی‌شده وضعیت میزان تقاضای پیش‌بینی‌شده را در طول شبانه‌روز مشخص می‌کنند. از این داده‌ها برای طراحی و پیاده‌سازی الگوریتم زمانبندی شارژ و دشارژ خودروهای برقی در ایستگاه شارژ پیشنهادی به جهت کاهش مصرف انرژی بهینه و هزینه‌های برق استفاده شده‌است.

برای جمع‌آوری داده‌های ایستگاه‌های شارژ خودروهای برقی در روش پیشنهادی از API سایت OpenChargeMap استفاده شده‌است تا اطلاعات مربوط به ایستگاه‌های شارژ در ایالت کالیفرنیا (شهر آدلانتو) جمع‌آوری گردد. در این بخش، پارامترهایی نظیر تعداد ایستگاه‌ها، نوع و ظرفیت شارژرهای ایستگاه‌ها، موقعیت جغرافیایی، تعداد نقاط شارژ و ... بررسی و دریافت می‌شود.

۵-۳- تقسیم‌بندی داده‌ها و آموزش مدل

مجموعه داده‌ها به زیرمجموعه‌های آموزش و آزمایش با نسبت‌های ٪۷۰ و ٪۳۰ تقسیم شده‌اند. این تقسیم‌بندی آموزش-آزمایش امکان ارزیابی قابلیت تعمیم مدل را فراهم می‌کند، همان‌طور که در مطالعات پیشین نیز نشان داده شده‌است (به‌عنوان مثال [۲۶-۲۸]، [۱۳]، [۵]، [۳]). مجموعه داده‌های آماده‌شده به‌عنوان ورودی به مدل LSTM برای پیش‌بینی قیمت برق و میزان انرژی مصرفی استفاده می‌شوند. سپس، پیش‌بینی‌های LSTM به الگوریتم زمان‌بندی مبتنی بر DQN داده می‌شود تا عملیات شارژ و دشارژ خودروهای برقی بهینه‌سازی شود.

۵-۴- عملکرد مدل LSTM برای پیش‌بینی قیمت برق و مصرف انرژی

در این بخش، عملکرد مدل شبکه عصبی LSTM برای پیش‌بینی قیمت برق و میزان مصرف انرژی ارزیابی می‌شود. پیش‌بینی قیمت برق در شکل ۳ نشان داده شده‌است. برای افزایش وضوح و نمایش بهتر مقایسه بین مقادیر واقعی و پیش‌بینی‌شده، محدوده افقی داده‌ها از ۰ تا ۱۵۰۰ در نظر گرفته شده‌است. طبق شکل ۳، شبکه LSTM توانسته تغییرات داده‌ها را به‌خوبی یاد بگیرد و پیش‌بینی کند. در بیشتر نقاط، مقدار پیش‌بینی‌شده به مقدار واقعی نزدیک است که نشان‌دهنده عملکرد مناسب مدل است. برای ارزیابی دقیق‌تر عملکرد مدل پیشنهادی، اختلاف بین مقادیر واقعی و پیش‌بینی‌شده برای قیمت برق محاسبه و به‌صورت جداگانه در شکل ۴ نشان داده شده است. نتایج مربوطه نشان می‌دهد که مقدار خطا در بازه ۲۰- تا ۲۶+ قرار دارد که بیانگر وجود انحراف محدود و کنترل‌شده در پیش‌بینی‌ها است. این موضوع نشان می‌دهد که مدل قادر است روند کلی تغییرات قیمت را با دقت مناسبی دنبال کند، هرچند در برخی نقاط دچار کم‌برآوردی یا بیش‌برآوردی جزئی می‌شود.

پیش‌بینی مصرف انرژی در شکل ۵ مشاهده می‌شود. مدل LSTM با ساختار پیشنهادی، برای پیش‌بینی مقدار داده در گام زمانی ۲۵ام از ۲۴ مقدار قبلی استفاده می‌کند. به همین دلیل پیش‌بینی‌ها بیشتر به میانگین مصرف انرژی در ۲۴ ساعت گذشته نزدیک هستند؛ به این صورت که وقتی داده‌های واقعی مصرف انرژی در بازه ۱۵۰۰ تا ۲۵۰۰ هستند، داده‌های پیش‌بینی‌شده در بازه ۱۸۰۰ تا ۲۳۰۰ قرار می‌گیرند که این به خاطر اثر پنجره زمانی ۲۴ ساعته است.

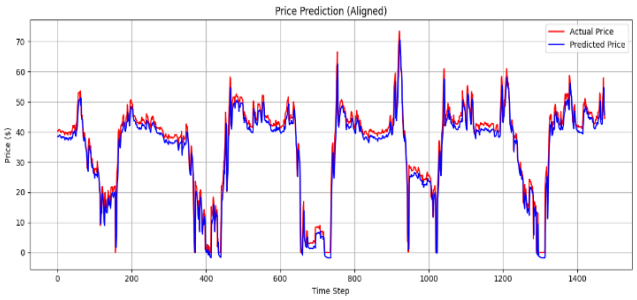
مشابه قیمت برق، اختلاف بین مقادیر واقعی و پیش‌بینی‌شده برای مصرف انرژی نیز محاسبه و به‌صورت جداگانه در شکل ۶ نشان داده شده است. مطابق این شکل، دامنه خطا در بازه ۴۸۳- تا ۶۴۷+ مشاهده می‌شود که با توجه به ماهیت نوسانی و گستره بزرگ‌تر مقادیر انرژی، این میزان خطا در مقایسه با مقیاس داده‌ها قابل قبول بوده و نشان‌دهنده توانایی مدل در یادگیری الگوهای پیچیده‌تر مصرف است. با وجود این نوسانات، روند کلی پیش‌بینی‌ها با مقادیر واقعی هم‌راستا بوده و مدل از پایداری مناسبی برخوردار است.

به‌طور کلی، نتایج نشان می‌دهد که مدل پیشنهادی ضمن حفظ دقت قابل قبول، توانایی مناسبی در تطبیق با تغییرات زمانی و شرایط غیرخطی موجود در داده‌ها دارد.

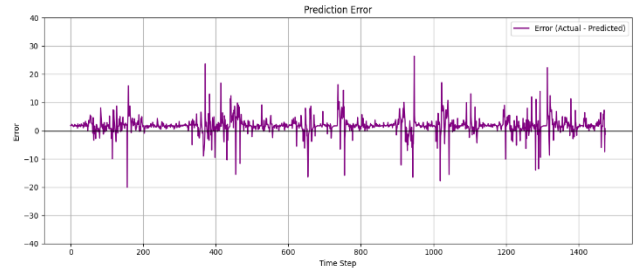
نتایج بدست آمده، با استفاده از معیارهای ارزیابی معرفی شده در بخش ۴-۱ شامل R^2، MAPE، MAE و RMSE بررسی شده‌اند. با توجه به وجود مقادیر صفر یا نزدیک به صفر در مجموعه داده، استفاده از MAPE قابل اعتماد نیست. بنابراین، یک شاخص خطای اصلاح‌شده به نام Safe-MAPE مطابق معادله (۱۲) تعریف و به‌کار گرفته شد و از مکمل آن نیز با عنوان Safe-MAPE-Accuracy استفاده شد. با افزودن یک عدد کوچک به مخرج معادله (۶)، این شاخص، تخمین خطای پیش‌بینی را مقاوم‌تر می‌کند و قابلیت اطمینان ارزیابی مدل را بهبود می‌بخشد. جزئیات بیشتر در مورد نسخه‌های دیگر Safe-MAPE در مرجع [۱۱] موجود است.

$$Safe_MAPE = \frac{100}{N}\sum_{i=1}^{n}\frac{|y_i - \hat{y_i}|}{max(|y_i| - |\hat{y_i}|, \in)} \qquad (۱۲)$$

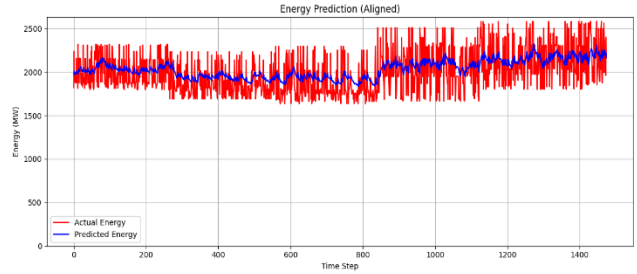
برای پیش‌بینی قیمت برق، بر اساس جدول ۳ و مطابق شاخص R^2، مدل توانسته است حدود ۹۲٪ از تغییرات داده‌های واقعی را به درستی پیش‌بینی کند. خطای نسبی اندازه‌گیری‌شده توسط شاخص Safe-MAPE حدود ۸٪ را نشان می‌دهد که مقدار قابل قبولی محسوب می‌شود. علاوه بر این، شاخص Safe-MAPE-Accuracy بیانگر آن است که مدل حدود ۹۲٪ از داده‌ها را با دقت مناسب پیش‌بینی کرده‌است. مقادیر MAE برابر با ۶ و RMSE برابر با ۲۴ واحد نیز دقت مناسب پیش‌بینی‌های مدل را تأیید می‌کنند.



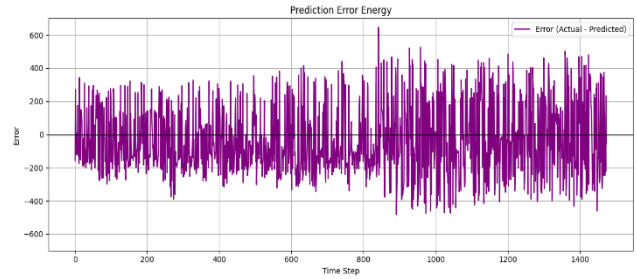
شکل ۳: پیش‌بینی قیمت برق با استفاده از مدل LSTM



شکل ۴: اختلاف مقدار واقعی و پیش‌بینی‌شده قیمت برق



شکل ۵: پیش‌بینی مصرف انرژی با استفاده از مدل LSTM



شکل ۶: اختلاف مقدار واقعی و پیش‌بینی‌شده مصرف انرژی

برای پیش‌بینی مصرف انرژی، مطابق جدول ۴، مدل توانسته است حدود ۵۰٪ از تغییرات داده‌های واقعی را پیش‌بینی کند. خطای نسبی اندازه‌گیری‌شده توسط شاخص Safe-MAPE حدود ۶٪ است که مقدار قابل قبولی محسوب می‌شود. مقدار MAE برابر با ۲۴۹ واحد و مقدار RMSE در حدود ۳۹۰ واحد است که نشان‌دهنده عملکرد مناسب مدل در پیش‌بینی مصرف انرژی، با وجود نوسانات موجود در داده‌های واقعی است.

با توجه به تابع خطا، شکل ۷ میانگین مربعات خطا را در طول آموزش مدل پیش‌بینی قیمت برق نشان می‌دهد. مقدار خطا در

جدول ۳: شاخص‌های ارزیابی برای پیش‌بینی قیمت برق

R^2	Safe_MAPE	Safe_MAPE _Accuracy	MAE	RMSE
۰/۹۱۸۷	۸/۲۲	۹۱/۸۸	۵/۷۲	۲۴/۴۳

جدول ۴: شاخص‌های ارزیابی برای پیش‌بینی مصرف انرژی

R^2	Safe_MAPE	Safe_MAPE _Accuracy	MAE	RMSE
۰/۵۰۰۳	۵/۹۴	۹۴/۰۶	۲۴۹/۱۹	۳۸۹/۸۸

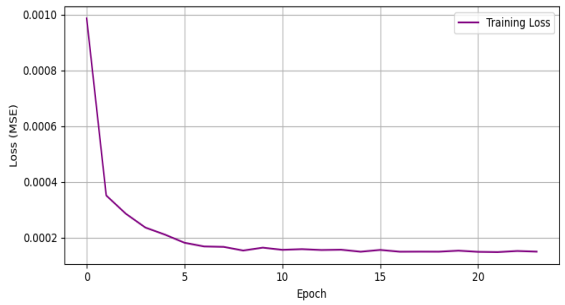
اولین دوره حدود ۰/۱٪ بوده که با ادامه فرآیند آموزش، خطا به‌تدریج کاهش یافت. به‌منظور جلوگیری از بیش‌برازش، از تکنیک توقف زودهنگام استفاده شد. در این‌جا، این تکنیک در دوره بیست و سوم فعال شد و در این مرحله، مقدار خطا به عدد بسیار کوچک ۰/۰۰۰۱۵ رسید. این موضوع نشان می‌دهد که مدل پیش از افت عملکرد، در زمان مناسبی متوقف شده‌است.

۵-۵- برنامه زمان‌بندی پیشنهادی ۲۴ ساعته برای شارژ خودروهای برقی

جدول ۵ نمونه‌ای از برنامه زمان‌بندی شارژ و دشارژ خودروهای برقی پیشنهادی را نشان می‌دهد. این برنامه به کاربران امکان می‌دهد تا قبل از ورود به ایستگاه یا شب قبل، زمان بهینه برای شارژ یا دشارژ خودروهای خود را انتخاب کنند، به گونه‌ای که هزینه شارژ کاهش یابد، مصرف انرژی در سطح مطلوب حفظ شود و ایستگاه بتواند تقاضای انرژی مورد نیاز را پاسخ دهد. ستون «قیمت برق» و «انرژی مصرفی» ارائه‌شده در جدول ۵ با استفاده از شبکه LSTM با سه لایه مخفی پیش‌بینی شده‌اند. ستون «میزان تقاضا» ظرفیت ایستگاه برای تأمین انرژی مورد نیاز شارژ خودروهای برقی در ایالت کالیفرنیا را نشان می‌دهد و مربوط به تمامی ایستگاه‌ها است.

۵-۶- عملکرد DQN برای زمان‌بندی شارژ خودروهای برقی

۱) ارزیابی پاداش تجمعی: پاداش تجمعی مدل DQN پیشنهادی برای تعداد متنوعی از دوره‌های آموزشی تا ۳۰۰۰ دوره ارزیابی و در نهایت به مقدار تقریبی ۲۱- دلار همگرا شد که نشان‌دهنده تثبیت رفتار مدل و یادگیری سیاست بهینه برای زمان‌بندی شارژ خودروهای برقی است. پاداش تجمعی پس از ۳۰۰۰ دوره آزمایش در شکل ۸ نشان داده شده است. در مقایسه با روش‌های ارائه شده در [۵]، [۲۰] و [۲۱] که به ترتیب به حدود ۲۶-، ۴۰- و ۱۵۱- دلار همگرا شدند، روش پیشنهادی عملکرد بهتری را نشان می‌دهد. برای مقایسه منصفانه، در همه این روش‌ها مقادیر عمل و داده‌های قیمت برق مشابه روش پیشنهادی در نظر گرفته شده‌اند.



شکل ۷: روند خطای اعتبارسنجی در مدل پیشنهادی LSTM برای پیش‌بینی قیمت برق

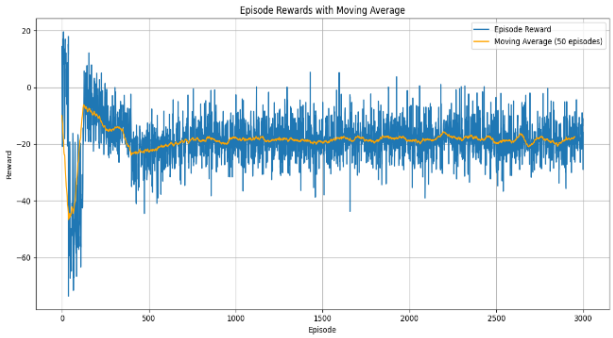
جدول ۵: نمونه‌ای از برنامه زمان‌بندی پیشنهادی شامل پیش‌بینی قیمت برق ۲۴ ساعته، مصرف انرژی و تقاضای شارژ خودروهای برقی

زمان (ساعت)	قیمت برق	انرژی	میزان تقاضا
۱	۷۷/۱۲۴	۲۰۲۰/۸۸۱	**۳۲۵۹/۹۵**
۲	۶۹/۶۰۵	۱۹۹۲/۳۷	**۳۱۲۳/۶۳**
۳	۶۷/۰۴۹	۱۹۸۰/۶۲۶	**۳۱۹۰/۵۱**
۴	۶۷/۴۲۹	۱۹۴۹/۷۹	**۳۳۱۶/۱۹**
۵	۷۰/۷۹۴	۱۹۵۲/۱۷۵	**۳۵۱۶/۹۵**
.	.	.	.
.	.	.	.
.	.	.	.

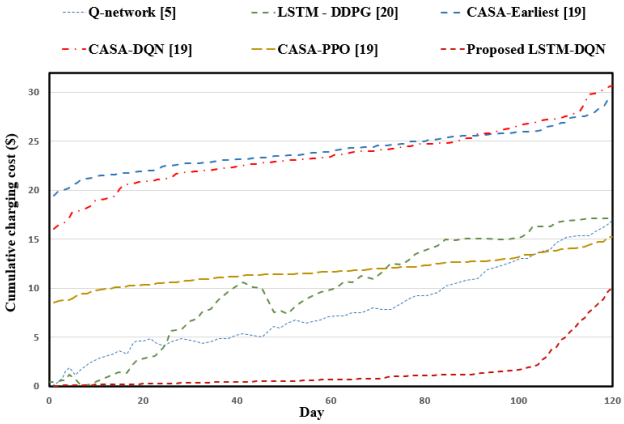
۲) ارزیابی هزینه تجمعی شارژ: عملکرد روش پیشنهادی به همراه روش‌های پیشین در طول ۱۲۰ روز آزمایشی ارزیابی شده است. برای هر روز، هزینه شارژ مطابق معادله (۱۳) محاسبه شده است:

$$Charging_cost = \sum_{t=t_{Arr}}^{t_{Dep}} P_t . act_t \qquad (۱۳)$$

شکل ۹ هزینه‌های تجمعی شارژ را برای روش‌های مختلف نشان می‌دهد. بایستی توجه داشت که در لحظه شروع شبیه‌سازی (روز صفر)، به دلیل تفاوت شرایط اولیه و سیاست‌های تصمیم‌گیری در الگوریتم‌های مختلف، مقادیر هزینه تجمعی بین روش‌ها متفاوت است. در برخی الگوریتم‌ها، اولین تخصیص منابع و اجرای وظیفه از همان ابتدای فرآیند شبیه‌سازی انجام می‌شود؛ بنابراین، هزینه تجمعی از همان نقطه ثبت شده و مقدار اولیه غیرصفر مشاهده می‌شود. در مقابل، در برخی روش‌ها ثبت هزینه با تأخیر بسیار کوتاهی آغاز شده و مقدار اولیه برابر صفر نمایش داده می‌شود. در این شکل، کاهش هزینه تجمعی شارژ با استفاده از روش پیشنهادی زمان‌بندی شارژ خودروهای برقی نسبت به روش‌های دیگر مشهود است. به‌عنوان مثال، روش پیشنهادی در انتهای دوره ۱۲۰ روزه به هزینه تجمعی شارژ برابر با ۱۰ دلار رسیده در صورتی که این هزینه



شکل ۸: پاداش تجمعی پیش‌بینی قیمت برق مبتنی بر روش پیشنهادی



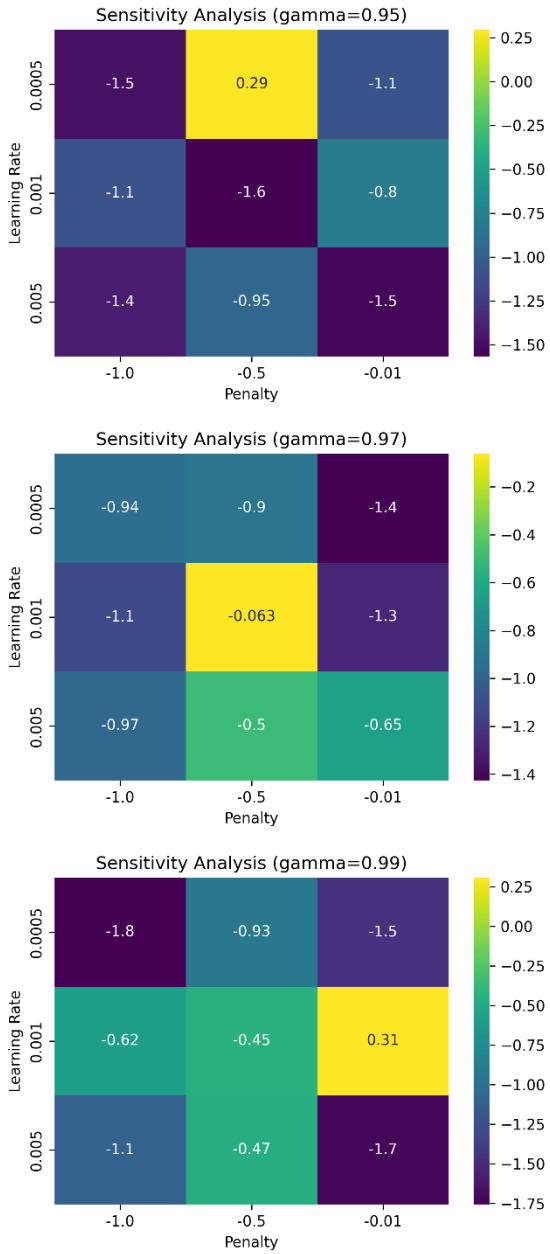
شکل ۹: هزینه‌های تجمعی شارژ روش پیشنهادی و روش‌های پیشین

برای روش پایه شبکه Q [۴, ۵] برابر با ۱۷ دلار بوده که نشان دهنده بهبود ۴۱ درصدی نسبت به آن است.

۵-۷- تحلیل حساسیت پارامترهای الگوریتم پیشنهادی

به‌منظور بررسی تأثیر پارامترهای یادگیری بر عملکرد مدل پیشنهادی، تحلیل حساسیت بر روی ضریب تنزیل (γ)، نرخ یادگیری (Lr) و ضریب جریمه تغییر عمل انجام شده است. برای این منظور، مجموعه‌ای از آزمایش‌ها به‌صورت جستجوی شبکه‌ای بر روی ترکیب‌های مختلف این پارامترها اجرا گردید و میانگین پاداش نهایی مدل در ۲۴ ساعت برای هر حالت محاسبه شد.

نتایج تحلیل حساسیت برای سه مقدار مختلف ضریب تنزیل (γ) در شکل ۱۰ نمایش داده شده است. این شکل نشان می‌دهد که عملکرد مدل به‌طور همزمان به نرخ یادگیری و ضریب جریمه در تابع پاداش وابسته است. با تغییر مقدار γ، ناحیه بهینه پارامترها جابه‌جا شده و مقدار بیشینه پاداش در نقاط متفاوتی از فضای پارامترها مشاهده می‌شود. بر اساس نتایج به‌دست‌آمده، بهترین عملکرد مدل در نرخ تنزیل ۰/۹۹ و نرخ یادگیری شبکه ۰/۰۰۱ حاصل شده‌است. در این حالت، با ضریب جریمه ۰/۰۱-، میانگین پاداش به حدود ۰/۳۱+ رسیده‌است. نتایج نشان می‌دهد که عملکرد مدل به تنظیم همزمان نرخ تنزیل، نرخ یادگیری شبکه و جریمه وابسته بوده و تغییر هر



شکل ۱۰: تحلیل حساسیت پارامترهای یادگیری تقویتی بر اساس نقشه حرارتی برای مقادیر مختلف ضریب تنزیل (γ)

یک از این پارامترها می‌تواند بر کیفیت یادگیری عامل تأثیرگذار باشد. به‌طور خاص، ترکیب γ=۰/۹۹، α=۰/۰۰۱ و جریمه ۰/۰۱- منجر به عملکرد مطلوب‌تر و همگرایی بهتر مدل نسبت به سایر تنظیمات بررسی‌شده، شده‌است. همچنین، این نتایج بیانگر حساسیت بالای مدل به هایپرپارامترها و اهمیت انتخاب مناسب آن‌ها برای دستیابی به رفتار پایدار و همگرایی مناسب است.

۶- نتیجه‌گیری

در این مقاله، مسأله زمان‌بندی شارژ خودروهای برقی با ادغام پیش‌بینی سری‌های زمانی و تکنیک‌های یادگیری تقویتی عمیق

مورد بررسی قرار گرفت. ابتدا، داده‌های تاریخی قیمت برق و میزان مصرف انرژی با استفاده از شبکه‌های LSTM آموزش‌دیده و داده روز آینده پیش‌بینی شدند. این نتایج به‌عنوان ورودی‌های مدل تصمیم‌گیری استفاده شدند. سپس، زمان‌بندی شارژ خودروهای برقی به‌صورت یک فرآیند تصمیم‌گیری مارکف مدل‌سازی شد. برای حل مسأله زمان‌بندی در شرایط عدم قطعیت قیمت برق و میزان تقاضا، یک شبکه DQN توسعه داده شد که از طریق تعامل با محیط، استراتژی شارژ بهینه را می‌آموزد. یک شبکه عصبی کاملاً متصل برای تخمین عمل-مقدار شارژ استفاده گردید که عمل با بالاترین مقدار مورد انتظار را انتخاب می‌کند. نتایج تجربی نشان داده است که روش پیشنهادی می‌تواند هزینه تجمعی شارژ را به‌طور مؤثر کاهش داده و در عین حال پاداش تجمعی را در طول زمان افزایش دهد. به این دلایل، روش پیشنهادی می‌تواند به عنوان راهکاری امیدوارکننده برای شارژ هوشمند خودروهای برقی مورد استفاده قرار گیرد و نتایج آن نشان‌دهنده کارایی قابل توجه در کاهش هزینه تجمعی شارژ و افزایش پاداش تجمعی است.

مراجع

[1] O. Sadeghian, A. Oshnoei, B. Mohammadi-Ivatloo, V. Vahidinasab, and A. Anvari-Moghaddam, "A comprehensive review on electric vehicles smart charging: Solutions, strategies, technologies, and challenges," *Journal of Energy Storage*, vol. 54, p. 105241, 2022. 10.1016/j.est.2022.105241.

[2] R. Zai, Y. Guo, Q. Liu, H. Sun, Q. Wu, and L. Xiao, "Knowledge Assisted Deep Reinforcement Learning for Electric Vehicle Charging Control," in *2022 IEEE 6th Conference on Energy Internet and Energy System Integration (EI2)*, 2022: IEEE, pp. 1882-1887. DOI: 10.1109/EI256261.2022.10116271.

[3] L. Yan, X. Chen, J. Zhou, Y. Chen, and J. Wen, "Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5124-5134, 2021. DOI: 10.1109/TSG.2021.3098298.

[4] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," IEEE transactions on smart grid, vol. 10, no. 1, pp. 841-851, 2017. DOI: 10.1109/TSG.2017.2753802.

[5] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time EV charging scheduling based on deep reinforcement learning," (in eng), *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5246-5257, 2018. DOI: 10.1109/TSG.2018.2879572.

[6] C. Zhang, Y. Liu, F. Wu, B. Tang, and W. Fan, "Effective charging planning based on deep reinforcement learning for electric vehicles," (in eng), *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 542-554, 2020. DOI: 10.1109/TITS.2020.3002271.

[7] A. Alsabbagh, H. Yin, and C. Ma, "Distributed electric vehicles charging management with social contribution concept," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3483-3492, 2019. DOI: 10.1109/TII.2019.2950460.

[8] Q. Xiao et al., "A deep reinforcement learning based charging and discharging scheduling strategy for electric vehicles," Energy Reports, vol. 12, pp. 4854-4863, 2024. 10.1016/j.egyr.2024.10.056.

[9] F. Neumann, "Optimal Scheduling of Electric Vehicle Charging in Distribution Networks," Technische Universität Berlin, Technische Universität Berlin, 2017. DOI: 10.13140/RG.2.2.13347.96805.

[10] R. Adam, K. Qian, and R. Brehm, "Electric vehicle user behavior prediction using Gaussian mixture models and soft information," in *2021 IEEE PES Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 2021: IEEE, pp. 1-5. DOI: 10.1109/ISGTAsia49270.2021.9715580.

[11] A. Orzechowski, L. Lugosch, H. Shu, R. Yang, W. Li, and B. H. Meyer, "A data-driven framework for medium-term electric vehicle charging demand forecasting," (in eng), *Energy and AI*, vol. 14, p. 100267, 2023. 10.1016/j.egyai.2023.100267.

[12] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," (in eng), *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2427-2439, 2019. DOI: 10.1109/TSG.2019.2955437.

[13] R. Luo *et al.*, "Deep learning approach for long-term prediction of electric vehicle (ev) charging station availability," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021: IEEE, pp. 3334-3339. DOI: 10.1109/ITSC48978.2021.9564633.

[14] F. Mohammad, D.-K. Kang, M. A. Ahmed, and Y.-C. Kim, "Energy demand load forecasting for electric vehicle charging stations network based on convlstm and biconvlstm architectures," *IEEE Access*, vol. 11, pp. 67350-67369, 2023. DOI: 10.1109/ACCESS.2023.3274657.

[15] S. Wang, A. Chen, P. Wang, and C. Zhuge, "Predicting electric vehicle charging demand using a heterogeneous spatio-temporal graph convolutional network," (in eng), *Transportation Research Part C: Emerging Technologies*, vol. 153, p. 104205, 2023. DOI: 10.1016/j.trc.2023.104205.

[16] J. Jin, S. Mao, and Y. Xu, "Optimal priority rule-enhanced deep reinforcement learning for charging scheduling in an electric vehicle battery swapping station," *IEEE Transactions on Smart Grid*, vol. 14, no. 6, pp. 4581-4593, 2023. DOI: 10.1109/TSG.2023.3250505.

[17] M. Boulakhbar, M. Farag, K. Benabdelaziz, M. Zazi, M. Maaroufi, and T. Kousksou, "Electric vehicles arrival and departure time prediction based on deep learning: the case of Morocco," in *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, 2022: IEEE, pp. 1-8. DOI: 10.1109/IRASET52964.2022.9738115.

[18] T. Qi, C. Ye, Y. Zhao, L. Li, and Y. Ding, "Deep reinforcement learning based charging scheduling for household electric vehicles in active distribution network," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 6, pp. 1890-1901, 2023. DOI: 10.35833/MPCE.2022.000456

[19] A. Zhang, Q. Liu, J. Liu, and L. Cheng, "CASA: cost-effective EV charging scheduling based on deep reinforcement learning," *Neural Computing and Applications*, vol. 36, no. 15, pp. 8355-8370, 2024. 10.1007/s00521-024-09530-3.

[20] S. Li *et al.*, "Electric vehicle charging management based on deep reinforcement learning," (in eng), *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 3, pp. 719-730, 2021. DOI: 10.35833/MPCE.2020.000460.

[21] S. Mishra, A. Choubey, S. V. Devarasetty, N. Sharma, and R. Misra, "An innovative multi-head attention model with BiMGRU for real-time electric vehicle charging management through deep reinforcement learning," *Cluster Computing*, vol. 27, no. 7, pp. 9993-10023, 2024. 10.1007/s10586-024-04494-4.

[22] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton University Press, 1958.

[23] OASIS. California independent system operator: oasis.caiso.com/mrioasis/logon.do.

[24] F. Zhang, Q. Yang, and D. An, "CDDPG: A deep-reinforcement-learning-based approach for electric vehicle charging control," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3075-3087, 2020. DOI: 10.1109/JIOT.2020.3015204.

[25] S. Wang, S. Bi, and Y. A. Zhang, "Reinforcement learning for real-time pricing and scheduling control in EV charging stations," (in eng), *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 849-859, 2019. DOI: 10.1109/TII.2019.2950809.

[26] L. Douaidi, S.-M. Senouci, I. El Korbi, and F. Harrou, "Predicting Electric Vehicle Charging Stations Occupancy: A Federated Deep Learning Framework," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, 2023: IEEE, pp. 1-5. DOI: 10.1109/VTC2023-Spring57618.2023.10199832.

[27] A. S. Khwaja, B. Venkatesh, and A. Anpalagan, "Short-term individual electric vehicle charging behavior prediction using long short-term memory networks," in *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, 2020: IEEE, pp. 1-7. DOI: 10.1109/CAMAD50429.2020.9209296.

[28] Z. Wan, H. Li, H. He, and D. Prokhorov, "A data-driven approach for real-time residential EV charging management," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 2018: IEEE, pp. 1-5. DOI: 10.1109/PESGM.2018.8585945.

پاورقی‌ها:

[1] Energy Demand
[2] Fluctuations
[3] Deep Reinforcement Learning (DRL)
[4] Long Short-Term Memory
[5] Forecasting
[6] Data-driven
[7] Recurrent Neural Networks
[8] Deep Q-Network
[9] Proximal Policy Optimization
[10] Soft Actor-Critic
[11] Markov Decision Process
[12] Cumulative Reward
[13] State Of Charge
[14] Action-value function
[15] Representation Network
[16] Gates
[17] Forget Gate
[18] Sliding Window
[19] Early Stopping
[20] Overfitting
[21] Episode
[22] Mean Average Percentage Error
[23] Mean Absolute Error
[24] Root Mean Squared Error
[25] Dropout
[26] Exploration
[27] Exploitation
[28] Batch Size
[29] Activation Function
[30] Evaluation Network
[31] Replay Buffer
[32] Discount factor
[33] Target Replace Iteration
[34] mini-batch
[35] California Independent System Operator
[36] Day-Ahead Demand Forecast