

نوع مقاله: پژوهشی

تاریخ پذیرش: ۱۴۰۰/۹/۲۴

تاریخ دریافت: ۱۳۹۹/۱۱/۱۳

مدیریت مصرف انرژی خانگی با استفاده از یادگیری تقویتی چندعاملی

علی فروتنی^۱، محمد رستگار^{۲*}

^۱ کارشناسی ارشد، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز، شیراز، ایران

Ali.forootani@shirazu.ac.ir

^۲ دانشیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه شیراز، شیراز، ایران

Mohammadrastegar@shirazu.ac.ir

چکیده: افزایش مصرف انرژی الکتریکی، مسئله‌ای است که همواره به‌عنوان یکی از چالش‌های تأمین‌کنندگان برق مطرح بوده است. به‌دنبال افزایش مصرف، برنامه‌های پاسخ‌گویی بار که سعی در مدیریت مصرف انرژی با اهدافی نظیر کاهش هزینه‌ها و افزایش قابلیت اطمینان دارند، بیش از پیش مورد توجه قرار گرفته‌اند. از طرفی هوشمندسازی مصرف‌کنندگان، امکان بهره‌گیری هرچه بیشتر از هوش مصنوعی برای مدیریت انرژی را میسر ساخته است. این مقاله روشی برای مدیریت مصرف انرژی خانگی با هدف کمینه کردن قبض برق و نارضایتی مشترک ارائه می‌دهد. با تفکیک بارهای خانه به سه دسته بارهای غیرقابل کنترل، قابل جابه‌جایی و قابل کنترل، یادگیری تقویتی چندعاملی با الگوریتم *Q-Learning* راهکاری است که در این مقاله برای اتخاذ تصمیمات بهینه دربارهٔ هریک از وسایل خانه در نظر گرفته شده است. به‌دلیل ماهیت الگوریتم *Q-Learning*، روش پیشنهادی در این مقاله برخلاف روش‌های برنامه‌ریزی عدد صحیح امکان افزودن وسایل بیشتری از خانه و حل مسئله‌های پیچیده‌تری را داراست. پیاده‌سازی روش پیشنهادی این مقاله در بخش مطالعهٔ عددی منجر به کاهش قبض برق مشترک تا ۲۴/۸٪ گردید. همچنین، نتایج حاصل از اعمال روش ارائه‌شده حاکی از صحت عملکرد آن است.

واژه‌های کلیدی: مدیریت مصرف انرژی، یادگیری تقویتی، هوش مصنوعی، پاسخ‌گویی بار.

* نویسنده مسئول

• این مقاله برگزیده کنفرانس SJC 2020 بوده که پس از تکمیل و داوری مجدد پذیرفته شده است.

۱. مقدمه

بر اساس مطالعات صورت گرفته در مرجع [۱]، پیش‌بینی می‌شود که تا سال ۲۰۴۰، مصرف انرژی الکتریکی ۲۸٪ افزایش یابد. از این رو نیاز به برنامه‌های مدیریت مصرف انرژی به‌عنوان راه‌حلی برای اجتناب از افزایش هزینه‌های تولید انرژی و همچنین کاهش مصرف انرژی الکتریکی در ساعات اوج مصرف بیش از پیش احساس می‌شود.

در این راستا، در سال‌های اخیر مدیریت مصرف انرژی بر پایه برنامه‌های پاسخ‌گویی بار منجر به تنظیم و مدیریت پویای مصرف انرژی شده است. منظور از تنظیم و مدیریت انرژی، جابه‌جا نمودن بار یا تنظیم توان مصرفی آن با در نظر گرفتن قیمت لحظه‌ای برق و همچنین قابلیت اطمینان شبکه است. به‌طور کلی پاسخ‌گویی بار به دو دسته قیمت محور و تشویق محور تقسیم می‌شود. پیاده‌سازی پاسخ‌گویی بار چه از سمت مصرف‌کننده و چه از سمت بهره‌بردار شبکه، مزایای متعددی را برای هر دو به‌همراه دارد که از جمله آن‌ها می‌توان به کاهش قبض برق مصرف‌کننده، کاهش هزینه بهره‌برداری و ریسک شبکه، کاهش انتشار گازهای گلخانه‌ای و افزایش بازدهی شبکه اشاره کرد [۲].

اما صرف پیاده‌سازی یک الگوریتم بهینه‌سازی برای کمینه کردن مصرف انرژی و قبض برق با جابه‌جایی یا کاهش مصرف، روشی عملی جهت پیاده کردن آن در دنیای واقعی نیست زیرا به‌رغم اهمیت هزینه مصرف انرژی برای مشترک، آسایش و راحتی مشترک نیز نقش بسزایی در تمایل وی برای پیاده کردن برنامه پاسخ‌گویی بار دارد. یکی از اساسی‌ترین پارامترهای تعیین‌کننده رضایت خاطر مشترک، رضایت دمایی خانه است [۳]. علاوه بر رضایت دمایی، فاصله گرفتن توان مصرفی برخی وسایل خانه نظیر سیستم روشنایی از حداکثر توان مصرفی، از دیگر مواردی است که ناراضی‌ی مشترک را به‌همراه دارد [۴]؛ در نتیجه، برقرار کردن یک مصالحه بین مصرف انرژی بارهای خانه و آسایش مشترک از جمله چالش‌های حل مسئله پاسخ‌گویی بار برای یک خانه هوشمند است. بارهای خانه را می‌توان به سه دسته بارهای غیرقابل کنترل مثل یخچال، بارهای قابل جابه‌جایی مثل ماشین لباس‌شویی و بارهای قابل کنترل مثل سیستم سرمایشی و روشنایی تفکیک کرد. علاوه بر موارد فوق، خودرو برقی را نیز می‌توان بار قابل کنترلی تصور نمود که علاوه بر کنترل توان شارژ و دشارژ، می‌توان به باتری خودرو به چشم یک سیستم ذخیره انرژی نگاه کرد. مشخصاً هرچه وسایل بیشتری از یک خانه هوشمند درگیر حل گردد، شباهت فضای حل مسئله به واقعیت و

البته پیچیدگی حل مسئله افزایش می‌یابد. یادگیری تقویتی که یکی از دسته‌های اصلی یادگیری ماشین می‌باشد، روشی است که بدون نیاز به اطلاعات کامل از محیط و به‌دلیل سادگی در پیاده‌سازی، به‌عنوان راه‌حلی برای پیاده‌سازی و حل مسائلی که پیچیدگی بالایی دارند، مطرح می‌شود [۳].

در سال‌های اخیر مطالعات گسترده‌ای در زمینه مدیریت انرژی و پاسخ‌گویی بار با استفاده از هوش مصنوعی صورت گرفته است. ترکیب یادگیری تقویتی و شبکه‌های عصبی عمیق و همچنین استفاده از الگوریتم Q-Learning روشی است که به‌ترتیب در مقاله [۵] و [۶] مورد استفاده قرار گرفته است. مرجع [۷] با تأکید بر ضعف روش‌های متداول بهینه‌سازی در مدیریت انرژی یک خانه هوشمند نظیر ثابت در نظر گرفتن شرایط محیط، با بهره‌گیری از یادگیری تقویتی، به یک کنترل هدفمند برای وسایل خانه دست یافته است. استفاده از شبکه‌های عصبی Convolutinal و MLP برای تخمین زدن Q-table، نوآوری است که در مرجع [۸] برای کنترل بهینه آب‌گرمکن و پمپ حرارتی یک خانه هوشمند ضمن کاهش قبض برق و ناراضی‌ی مشترک، مورد استفاده قرار گرفته است. در مراجع [۱۱] و [۱۲] به مسئله قیمت‌گذاری بهینه در پاسخ‌گویی بار یک شبکه هوشمند از طریق Q-Learning پرداخته شده است. بهره‌گیری از معماری شبکه عصبی بازگشتی LSTM برای پیش‌بینی قیمت برق و استفاده از این قیمت پیش‌بینی شده برای مدیریت انرژی خانگی با کمک یادگیری تقویتی، نوآوری نویسندگان در مرجع [۱۵] برای کمینه کردن مصرف انرژی الکتریکی است. با فرض مجهز بودن خانه به اینترنت اشیا، ذخیره‌سازی و پردازش اطلاعات، مراجع [۱۶] و [۱۷] به حل مسئله مدیریت مصرف انرژی خانگی پرداخته‌اند. صرف‌نظر کردن از بسیاری از بارهای خانه، نویسندگان مراجع [۱۸] و [۱۹] را قادر ساخته است تا با ترکیب کردن الگوریتم Q-Learning و منطق فازی، به‌ترتیب به مدیریت مصرف انرژی بارهای قابل جابه‌جایی و یک آب‌گرمکن هوشمند دست یابند. بهینه‌سازی مصرف سیستم سرمایشی/گرمایشی و همچنین آب‌گرمکن از موضوعاتی است که مورد توجه نویسندگان متعددی در این زمینه بوده و پژوهش‌های گسترده‌ای در زمینه بهره‌گیری از روش‌های مختلف یادگیری تقویتی به انجام رسیده است [۲۰-۲۴].

محبوبیت استفاده از یادگیری تقویتی به‌دلیل دو قابلیت اساسی آن است: اول اینکه یادگیری تقویتی با تعامل با محیط مسئله، اقدام به پیدا کردن عمل بهینه می‌کند. این موضوع که در شکل (۱) نشان

توان مصرفی ساعتی این وسایل با توجه به قیمت لحظه‌ای برق می‌نماید. به صورت خلاصه، مهم‌ترین نوآوری‌های این مقاله بدین شرح است:

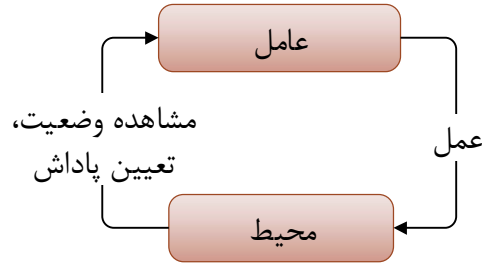
- به‌کارگیری یادگیری تقویتی چندعاملی به‌منظور مدیریت مصرف انرژی ساعتی بر مبنای پاسخ‌گویی بار قیمت محور برای یک خانه هوشمند.
- تفکیک وسایل به سه دسته بارهای غیرقابل کنترل، قابل جابه‌جایی و قابل کنترل و مدیریت بهینه هر یک از این وسایل به صورت غیرمتمرکز.
- توجه هم‌زمان به مصرف انرژی مشترک و نارضایتی مشترک در تابع هدف مسئله.

در ادامه این مقاله، بخش ۲ مقاله به توضیح الگوریتم Q-Learning استفاده‌شده، مدل‌سازی وسایل خانه و فرمول‌بندی پاسخ‌گویی بار یک خانه پرداخته است. در بخش ۳، سناریوهایی برای آزمایش روش پیشنهادی به‌همراه نتایج و بحث پیرامون آن‌ها ارائه شده است و در نهایت در بخش ۴، جمع‌بندی ارائه شده است.

۲. مدل‌سازی و فرمول‌بندی مسئله

در این مقاله، به مدیریت مصرف انرژی با بهره‌گیری از هوش مصنوعی پرداخته می‌شود. یکی از پرکاربردترین شاخه‌های هوش مصنوعی، حوزه یادگیری ماشین است که خود به سه دسته یادگیری نظارت‌شده، یادگیری بدون نظارت و یادگیری تقویتی تقسیم می‌شود. همچنین از سال ۲۰۱۳، افزایش تمایل پژوهشگران به استفاده از یادگیری تقویتی در مدیریت انرژی خانگی قابل مشاهده است [۲۶]. به بیان ساده، در هر مسئله یادگیری تقویتی، هدف در آموزش دادن یک عامل به‌منظور اخذ تصمیمات بهینه خلاصه می‌شود. آموزش دیدن عامل از طریق تعامل با یک محیط^۱ است که این محیط، فضا و چهارچوب مسئله را مشخص می‌کند. انجام هر عمل^۲ توسط عامل در هر وضعیت^۳ از محیط، موجب اختصاص یافتن پاداش و جریمه به عامل می‌شود. این پاداش و جریمه مطابق با طراحی سیگنال پاداش^۴ تعیین می‌شود. در مسئله مدیریت مصرف انرژی خانگی، برای هر یک از وسایل خانه هوشمند، یک عامل به‌منظور اخذ تصمیمات بهینه برای روشن/خاموش کردن و همچنین تعیین سطح توان مصرفی وسایل می‌بایست آموزش ببیند. محیط حل مسئله، در واقع خانه مورد مطالعه به‌همراه فرضیات مسئله، یعنی

داده شده، بدین معنی است که برای اعمال آن نیازی به اطلاعات محیط نیست [۲۵]؛ دوم اینکه یادگیری تقویتی در قیاس با سایر روش‌های متداول بهینه‌سازی، از سادگی بیشتری در پیاده‌سازی برخوردار است [۲۵].



شکل (۱): طرح کلی تعامل در یادگیری تقویتی

با وجود استفاده از انواع روش‌های یادگیری تقویتی در پژوهش‌های صورت گرفته، همچنان دو محدودیت در مطالعات فوق دیده می‌شود: اول اینکه اکثر مقالات، مطالعه خود را روی تعداد معدودی از وسایل خانگی انجام داده‌اند؛ بدین معنی که صرفاً به مدل کردن بارهای قابل کنترل پرداخته شده و در موارد ترکیب این بارها با بارهای قابل جابه‌جایی نیز به دلیل افزایش پیچیدگی مسئله، یک نوع بار قابل کنترل در کنار بارهای قابل جابه‌جایی مدل شده است و برخی بارهای قابل کنترل مهم نظیر خودرو برقی کنار گذاشته شده‌اند. محدودیت دومی که در مطالعات مشاهده می‌شود، این است که رضایت خاطر مشترک از حیث دما، توان مصرفی، ساعت کارکرد و فاصله گرفتن زمان روشن شدن بارهای قابل جابه‌جایی از زمان مطلوب کاربر، موضوعی است که همگی در کنار هم دیده نشده‌اند.

این مقاله یک الگوریتم برای مدیریت مصرف انرژی ساعتی بر اساس پاسخ‌گویی بار قیمت محور برای وسایل مختلف خانگی ارائه می‌دهد. این کار از طریق پیاده کردن الگوریتم Q-Learning صورت می‌گیرد که یکی از روش‌های تصمیم‌گیری بهینه در یادگیری تقویتی است. با توجه به اعمال این روش برای هر وسیله خانگی به صورت مجزا، رویکرد پیشنهادی در این مقاله در گروه روش‌های غیرمتمرکز مدیریت انرژی خانه هوشمند قرار می‌گیرد. نقطه قوت رویکرد غیرمتمرکز در شکستن هزینه پردازشی بسیار زیاد بهینه‌سازی یک خانه است که با کمک یادگیری تقویتی چندعاملی، مسئله به صورت مجزا برای هر وسیله حل می‌شود. این مقاله با در نظر گرفتن سه نوع بار مختلف برای یک خانه یعنی بارهای غیرقابل کنترل مثل یخچال، بارهای قابل جابه‌جایی مثل ماشین لباس‌شویی و ماشین ظرف‌شویی و بارهای قابل کنترل مثل سیستم سرمایشی، روشنایی و خودرو برقی، اقدام به بهینه‌سازی وضعیت روشن/خاموش بودن و

1. Environment
2. Action
3. State
4. Reward signal

۲.۱. Q-Learning

متداول‌ترین راه برای مدل کردن یک مسئله یادگیری تقویتی، ایجاد قالبی برای مسئله به فرم فرایند تصمیم‌گیری مارکوف است [۲۸] که پژوهشگر را قادر می‌سازد تا از طریق گسسته‌سازی زمان، امکان حل مسئله در چهارچوب یک مسئله تصمیم‌گیری بهینه را فراهم کند. یک فرایند تصمیم‌گیری مارکوف متشکل از یک مجموعه پنج‌مانه به قرار مقابل $\langle S, A, P, r, \gamma \rangle$ می‌باشد. این پنج مانه به ترتیب بیانگر وضعیت، عمل انجام‌شده توسط عامل، ماتریس احتمال انتقال، سیگنال پاداش و نرخ آینده‌نگری عامل هستند. در این مقاله، یکی از روش‌های بی‌نیاز از مدل مورد استفاده قرار گرفته است. همان‌طور که از اسم آن مشخص است، در روش‌های بی‌نیاز از مدل، هیچ مدل ریاضی از محیط در اختیار نیست و همچنین پاداش قرار گرفتن در یک وضعیت نیز از قبل مشخص نیست و تنها با قرار گرفتن در هر وضعیت، پاداش آن تعیین می‌شود. از جمله محبوب‌ترین روش‌های بی‌نیاز از مدل می‌توان به SARSA و Q-Learning اشاره کرد [۲۶]. بی‌نیاز از مدل بودن الگوریتم Q-Learning، این روش را قادر می‌سازد تا بدون داشتن اطلاعات اولیه و خصوصی از خانه، مدیریت انرژی را حل کند. در الگوریتم Q-Learning، یک ماتریس که هر درایه از آن، بیانگر ارزش انجام یک عمل بخصوص در یک وضعیت بخصوص می‌باشد، قابل تصور است. اساس فرمول‌بندی الگوریتم Q-Learning بر پایه به‌روزرسانی ارزش هر عمل در هر وضعیت بر اساس پاداش آنی قرار گرفتن در آن وضعیت و همچنین بیشترین ارزش وضعیت جدید ناشی از عمل انتخاب‌شده می‌باشد که در رابطه (۱) نشان داده شده است.

$$Q(S_t, a_t) = Q(S_t, a_t) + \alpha[r + \gamma Q(S_{t+1}, a_t) - Q(S_t, a_t)] \quad (1)$$

که در آن α, γ, S_t, a_t به ترتیب بیانگر نرخ یادگیری، ضریب آینده‌نگری الگوریتم، وضعیت در لحظه t و عمل در لحظه t هستند. ضریب آینده‌نگری، در واقع میزان توجه عامل به ارزش بلندمدت و کوتاه‌مدت را مشخص می‌کند. با توجه به رابطه (۱)، با صفر قرار دادن این پارامتر، ضریب Q-Value متعلق به گام زمانی بعد صفر خواهد شد. با این رویکرد، عامل نهایتاً پس از اتمام آموزش، عملی را انتخاب خواهد کرد که دارای بیشترین پاداش آنی باشد. درحالی‌که ممکن است اتخاذ عملی دیگر، پاداش آنی کمتری نسبت به انتخاب سایر اعمال داشته باشد اما در بلندمدت به جواب بهتری برای مسئله منجر گردد. از طرف دیگر، اختصاص دادن مقادیر نزدیک به یک، منجر به ایجاد دیدی صرفاً آینده‌نگر برای عامل خواهد شد. به صورت خاص در مسئله مدیریت مصرف انرژی

قیمت برق و مشخصات عملکرد وسایل است. نحوه طراحی سیگنال پاداش در بخش بعد، به تفصیل شرح داده شده است. اما الگوریتم‌های مبتنی بر یادگیری تقویتی به طرق مختلف قابل دسته‌بندی و تشریح هستند. می‌توان این الگوریتم‌ها را بر اساس نحوه آموزش عامل به دو دسته روش‌های بازیگر صرف^۱ و منتقد صرف^۲ تفکیک کرد. در دسته بازیگر صرف، هدف نهایی از آموزش عامل، یاد گرفتن احتمال انتخاب هر عمل مجزا در تمامی حالات ممکن است. در صورت آموزش صحیح عامل، در نهایت بهترین عمل که عملی با بیشترین پاداش می‌باشد، می‌بایست بیشترین احتمال انتخاب شدن را به خود اختصاص دهد. در دسته روش‌های منتقد صرف، هدف از آموزش دادن عامل، یاد گرفتن ارزش بلندمدت اتخاذ هر عمل در هر وضعیت و در نتیجه انتخاب اعمال با بیشترین ارزش بلندمدت از بین سایر اعمال ممکن است. دسته سومی نیز تحت عنوان الگوریتم‌های بازیگر-منتقد^۳ برای یادگیری تقویتی قابل تصور است که به ادغام این دو رویکرد پرداخته و کاربردهای بخصوص خود را دارد [۲۷].

دیگر نگاه متداول در دسته‌بندی روش‌های یادگیری تقویتی، تفکیک این الگوریتم‌ها به دو دسته روش‌های مدل محور و بی‌نیاز از مدل است. در روش‌های مدل محور، یک رابطه ریاضی شبیه به تابع تبدیلی دقیق برای محیط مسئله تعریف می‌شود که نتیجه آن، تشکیل ماتریسی تحت عنوان ماتریس احتمال انتقال است. در این دسته از روش‌ها، پیش از شروع مسئله، محیط حل مسئله و احتمال تغییر وضعیت برای عامل می‌بایست از پیش تعیین شده باشد. دسته دوم، روش‌های بی‌نیاز از مدل هستند. نقطه قوت اصلی این دسته از روش‌ها، عدم نیاز به تعریف تابع تبدیلی دقیق از محیط مسئله است. در واقع با این رویکرد، عامل به مرور و با تجربه عمل‌های متفاوت در شرایط مختلف، به یادگیری محیط می‌پردازد. امروزه روش‌های بی‌نیاز از مدل، از محبوبیت بالایی در حل مسائل مدیریت انرژی نظیر مدیریت انرژی مصرف انرژی خانگی برخوردارند [۲۶]. الگوریتم استفاده‌شده در این مقاله، الگوریتم Q-Learning است که یک روش بازیگر صرف و بی‌نیاز از مدل است. بدین معنی که در این الگوریتم، عامل بدون دانستن اطلاعات کامل از محیط، اقدام به یاد گرفتن ارزش بلندمدت هر عمل از طریق تجربه عمل‌های متفاوت می‌کند. در ادامه به بیان جزئیات این الگوریتم پیشنهادی برای حل مسئله مدیریت انرژی پرداخته شده است.

1. Actor only
2. Critic only
3. Actor-Critic

تعیین یک مقدار پیش فرض کوچک برای ϵ ، عامل اقدام به تولید یک عدد تصادفی با تابع چگالی احتمال یکنواخت در محدوده صفر تا یک می‌کند. در صورتی که عدد تولید شده از ϵ مقدار بزرگ‌تری داشته باشد، عامل اقدام به انتخاب عملی با بیشترین Q -Value از بین عمل‌های ممکن می‌کند؛ اما در صورتی که عدد تولید شده از ϵ مقدار کوچک‌تری داشته باشد، عامل به صورتی تصادفی عملی را انتخاب می‌کند. این روش باعث تجربه کردن اعمالی با شانس انتخاب کمتر در حالت عادی و در نتیجه جلوگیری از توقف روند آموزش عامل در نقطه کمینه محلی می‌گردد. این روند تا اتمام یک ۲۴ ساعت تکرار می‌شود (حلقه درونی در شکل ۲). سپس تمامی مراحل مجدداً از ابتدا تکرار خواهد شد، با این تفاوت که عامل مسئله نسبت به تکرار قبل با تجربه‌تر شده چراکه نتیجه کوتاه‌مدت و همچنین بلندمدت (در طی ۲۴ ساعت) اعمالی راکه در تکرار قبل انتخاب نموده است می‌داند. این روند به صورت مکرر تکرار خواهد شد تا جایی که عامل یادگیری تقویتی، همه اعمال متفاوت را در شرایط مختلف و تأثیر بلندمدت اعمال مختلف بر یکدیگر در بلندمدت را تجربه کرده باشد.

۲.۲. مدل‌سازی وسایل خانگی

برای مدل کردن وسایل یک خانه هوشمند، در قدم اول وسایل خانه به دو دسته بارهای پاسخ‌گو و غیرپاسخ‌گو تقسیم می‌شود. بارهای غیرپاسخ‌گو، وسایلی را شامل می‌شوند که کنترلی روی نحوه و زمان مصرف آن‌ها وجود ندارد، مانند یخچال. همچنین برخی وسایل دیگر مانند تلویزیون نیز به دلیل ماهیت غیرهوشمند و همچنین وابستگی زیاد به رفتار مشترک، جزء بارهای غیرپاسخ‌گو در این مقاله فرض شده‌اند که مطابق فرضیات مرجع [۲۹] است. با توجه به غیرقابل کنترل بودن عملکرد این دسته از بارها، انرژی مصرفی توسط این دسته از وسایل، دقیقاً برابر با نرخ نامی مصرف انرژی آن‌ها در نظر گرفته می‌شود.

در نتیجه، هزینه مصرف انرژی این دسته از وسایل برابر خواهد

بود با:

$$C_{Base Load} = C_t \cdot E_{Base Load} \quad (2)$$

که C_t قیمت برق در ساعت t می‌باشد.

خانگی، ممکن است تصمیم به خاموش کردن سیستم سرمایشی در ساعتی بخصوص، به دلیل قیمت بالای برق در آن ساعت، دارای جریمه کوتاه‌مدت کمتری نسبت به روشن بودن با توان مصرفی پایین داشته باشد؛ اما این کار ممکن است به گرم شدن بیش از حد خانه و نیاز به مصرف انرژی بیشتر در ساعات بعد باشد و منجر به هزینه برق بیشتر شود. از این رو در مسائل مدیریت مصرف انرژی، مقدار ضریب آینده‌نگری غالباً در محدوده $0/9$ قرار داده می‌شود [۳۱]. دیگر پارامتر قابل تنظیم در رابطه (۱)، α می‌باشد که در واقع بیانگر گام یادگیری و بهروزرسانی Q -Value در الگوریتم است. به این دسته از پارامترهای قابل تنظیم، hyperparameter اطلاق می‌شود. در واقع hyperparameter به پارامتری گفته می‌شود که قسمت جدایی‌ناپذیر یک الگوریتم یادگیری ماشین است و بسته به نوع مسئله، توسط محقق تعیین می‌شود [۲۹]. متداول‌ترین راه برای تنظیم hyperparameter در یک الگوریتم مبتنی بر یادگیری ماشین، استفاده از آزمون و خطا و یا تجربیات محقق است [۳۱]. در مسئله مدیریت مصرف انرژی خانگی، اختصاص دادن مقادیر بزرگ نظیر $0/5$ به این پارامتر، موجب عبور عامل از نقطه بهینه و نتایج نامناسب خواهد شد. از طرف دیگر اختصاص دادن مقادیر بسیار کوچک به این پارامتر باعث کند شدن بیش از حد و بعضاً متوقف شدن روند آموزش می‌شود. از این رو در این مقاله، مقدار $0/1$ به این پارامتر اختصاص داده شده است.

شکل (۲) روند طی شده برای آموزش عامل با الگوریتم Q-Learning را نمایش می‌دهد. همان طور که در این شکل مشخص است، برای آموزش عامل مربوط به هریک از وسایل مدل‌شده در یادگیری تقویتی، ابتدا محیط مسئله و شرایط حاکم نظیر قیمت برق، مشخصات مصرفی و همچنین پارامترهای یادگیری تقویتی نظیر نرخ یادگیری و ضریب آینده‌نگری تعیین می‌شوند. در قدم بعد، ماتریس Q -Table با توجه به ابعاد فضای حل مسئله برای هر عامل شکل می‌گیرد. برای مثال، برای Q -Table متعلق به سیستم سرمایشی، با فرض اینکه در طی ۲۴ ساعت شبانه‌روز، دمای داخل خانه می‌تواند مقداری بین ۱۶ تا ۲۵ درجه سانتی‌گراد به خود اختصاص دهد (فارغ از دمای مطلوب ساکنان) و همچنین متصور شدن پنج سطح مصرف انرژی متفاوت برای این وسیله، یک ماتریس 24×5 در ۵ متصور است. در قدم بعد، عامل اقدام به دریافت اطلاعات موجود در ساعت جاری می‌کند. حال نوبت به اتخاذ یک عمل در وضعیت موجود می‌رسد. روش استفاده شده در این مقاله برای اتخاذ عمل بهینه، روش ϵ -greedy می‌باشد. در این روش با

دارد، در رضایت مشترکین در نحوه روشن کردن این وسایل است. با فرض اینکه حالت مطلوب مشترک این است که بار قابل جابه‌جایی در بازه $T1$ و $T2$ روشن گردد، عبارت نارضایتی مشترک برای بار قابل جابه‌جایی به فرم زیر تعریف شده است:

$$discomfort_{sh} = |T_{on} - T1| \text{ if } T_{on} < T1 \quad (6)$$

$$discomfort_{sh} = |T_{on} - T2| \text{ if } T_{on} > T2 - T_{op} \quad (7)$$

که T_{on} ساعت روشن کردن وسیله و T_{op} مدت مورد نیاز وسیله برای روشن ماندن پس از شروع به کار است.

دسته‌سومی از بارها که به مدل‌سازی آن‌ها در این مقاله پرداخته می‌شود، بارهای قابل کنترل است که میزان مصرف انرژی آن‌ها در هر ساعت قابل تنظیم است. یکی از بارهای قابل کنترل که سهم قابل توجهی در مصرف انرژی الکتریکی یک خانه دارد، سیستم سرمایشی/گرمایشی است. در این مقاله فرض شده است که یک سیستم تهویه هوا (سرمایشی/گرمایشی)، می‌تواند پنج سطح مصرف متفاوت داشته باشد؛ به عبارت دیگر:

$$E_{air\ conditioner} = [E_0, E1, E2, E3, E_{max}] \quad (8)$$

با توجه به رابطه (۸)، محاسبه هزینه انرژی مصرفی ساعتی توسط سیستم تهویه به قرار زیر است:

$$C_{air\ conditioner} = C_t \cdot E_{air\ conditioner} \quad (9)$$

محدودیتی که در رابطه با مدل کردن سیستم تهویه مطرح است، در بحث رضایت دمایی مشترک است. با فرض اینکه قرار گرفتن در دمایی بیش از $temp_{max}$ و همچنین دمایی کمتر $temp_{min}$ موجب نارضایتی مشترک می‌شود، می‌بایست دما در ساعت t حتماً در این محدوده قرار گیرد.

$$Temp_{min} < Temp_t < Temp_{max} \quad (10)$$

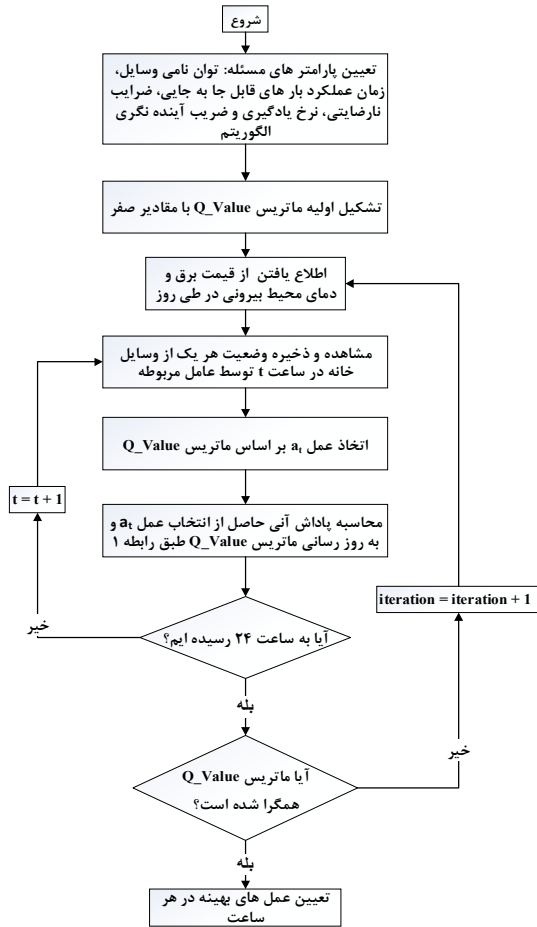
دمای لحظه‌ای داخل خانه، تابع پارامترهای متعددی نظیر توان مصرفی سیستم تهویه، دمای بیرون خانه و روابط ترمودینامیک حاکم بر این قضیه می‌باشد. از این رو رابطه تعیین کننده دمای خانه به قرار زیر می‌باشد [۳۰]:

$$Temp_{t+1} = \epsilon_{air} \cdot Temp_t + (1 - \epsilon_{air}) \cdot (Temp_{outdoor,t} + \eta_{ac} \cdot E_{air\ conditioner} / K_{air}) \quad (11)$$

در رابطه فوق، ϵ_{air} بیانگر اینرسی دمایی محیط، K_{air} نماینده ضریب انتقال حرارت سیستم تهویه، η_{ac} بیانگر ضریب عملکرد سیستم تهویه و $temp_{outdoor,t}$ نیز دمای خارج از خانه در ساعت t را نشان می‌دهد.

با داشتن دمای حاصل در هر ساعت، ترم نارضایتی دمایی مشترک عبارت است از:

$$air\ conditioner\ discomfort = |Temp_t - Temp_{min}| \text{ if } Temp_t < Temp_{min} \quad (12)$$



شکل (۲): روند پیشنهادی آموزش عامل

$$E_{Base\ Load} = E_{Rated} \quad (3)$$

برای رسیدن به یک مطالعه جامع، مناسب است تا بارهای پاسخ‌گو نیز به دو دسته بارهای قابل جابه‌جایی و بارهای قابل کنترل تقسیم شوند. بارهای قابل جابه‌جایی، بارهایی هستند که فقط زمان روشن/خاموش کردن آن‌ها قابل تنظیم است. شایان ذکر است که از لحظه‌ای که تصمیم به روشن کردن بار قابل جابه‌جایی گرفته شود، این وسیله می‌بایست به مدت زمان عملکرد خود روشن بماند. برای این دسته از وسایل، متغیر تصمیم‌گیری دودویی Z به‌گونه‌ای در نظر گرفته شده که در صورت روشن کردن وسیله، این متغیر برابر با «۱» و در صورت خاموش کردن آن، برابر با «۰» باشد. Z بیانگر عمل انجام‌شده توسط عامل یادگیری تقویتی است. بنابراین انرژی ساعتی و هزینه معادل آن به ترتیب در روابط (۴) و (۵) نشان داده می‌شود.

$$E_{Shiftable\ Load} = Z \cdot E_{Rated} \quad (4)$$

$$C_{Shiftable\ Load} = C_t \cdot E_{Shiftable\ Load} \quad (5)$$

محدودیتی که برای روشن کردن این دسته از وسایل وجود

$$air\ conditioner\ discomfort = |Temp_t - Temp_{max}| \quad (13)$$

if $Temp_t < Temp_{max}$

دیگر بار قابل کنترل در نظر گرفته شده در این مقاله، سیستم روشنایی خانه است. مشابه سیستم سرمایشی، برای توان مصرفی سیستم روشنایی خانه هوشمند، بازه‌ای از توان‌های مصرفی به قرار زیر در نظر گرفته شده است:

$$E_{Lighting} = [E_1, E_2, E_3, E_{max}] \quad (14)$$

با توجه به رابطه (۱۴)، هزینه انرژی مصرفی توسط سیستم روشنایی به قرار زیر است:

$$C_{Lighting} = C_t \cdot E_{Lighting} \quad (15)$$

برای مدل کردن نارضایتی مشترکین در کنترل بهینه سیستم روشنایی، فرض شده که مطلوب این است که روشنایی خانه در حداکثر حالت ممکن یا به عبارتی حداکثر توان مصرفی قرار داشته باشد [۳۱]. بدین ترتیب فاصله از این حد مطلوب به عنوان عدم رضایت به صورت زیر تعریف می‌شود:

$$discomfort_{Lighting} = |E_{Lighting} - E_{max}| \quad (16)$$

آخرین وسیله‌ای که در این مقاله به مدل‌سازی آن اقدام شده، خودرو برقی است. همان طور که در مرجع [۳۲] پیشنهاد شده است، فرض شده که شارژر هوشمند خودرو برقی، توان‌های گسسته‌ای را برای باتری خودرو فراهم می‌کند. بیشتر پژوهش‌های مرتبط با مدیریت مصرف انرژی خانگی که به مدل‌سازی خودرو برقی اقدام کرده‌اند، این وسیله را فقط به عنوان یک مصرف‌کننده انرژی الکتریکی در نظر گرفته اند. اما در این مقاله علاوه بر در نظر گرفتن دو مد شارژر با حداکثر توان نامی و نصف توان نامی، دو مد شارژر نیز برای شارژر هوشمند خودرو در نظر گرفته شده به گونه‌ای که عامل خودرو برقی بیاورد تا در صورتی که فروش انرژی به شبکه سودمند باشد، اقدام به شارژ باتری خود کند. از این رو برای کنترل توان خودرو برقی، مجموعه‌ای از توان‌ها به قرار زیر تعریف می‌شود:

$$E_{EV} = [-E_{max}, -E_1, 0, +E_1, +E_{max}] \quad (17)$$

محاسبه هزینه توان مصرفی توسط شارژر خودرو برقی، مانند سایر بارهای قابل کنترل است:

$$C_{EV} = C_{EV} \cdot E_{EV} \quad (18)$$

با توجه به توان انتخاب شده توسط عامل در هر ساعت، وضعیت شارژ باتری به صورت زیر خواهد بود [۷]:

$$SOC_{new} = SOC_{previous} + E_{EV} / E_{max} \quad (19)$$

دو محدودیت اساسی در خصوص مدل‌سازی خودرو برقی مطرح است: محدودیت اول، بحث رضایت مشترک است که متناسب با

فاصله توان شارژ خودرو با حداکثر توان شارژ در رابطه (۲۰) مدل می‌شود. محدودیت دوم نیز در بحث استهلاک باتری خودرو است که مطابق مدل ارائه شده در مرجع [۳۳]، در رابطه (۲۱) مدل می‌شود.

$$discomfort_{EV} = |E_{EV} - E_{max}|^2 \quad (20)$$

$$degradation_{EV} = cost_{batt} \cdot M_k \cdot E_{EV} / E_{max} \quad (21)$$

که $cost_{batt}$ هزینه کل باتری و M_k نیز نشان‌دهنده شیب خطی در نظر گرفته شده برای تقریب زدن تبعیت استهلاک باتری از سیکل شارژ و دشارژ است.

۳.۲. تابع هدف مسئله

در این مقاله از الگوریتم Q-Learning برای آموزش عامل‌های مسئله مدیریت مصرف انرژی استفاده شده است. از این رو برای آموزش عامل‌ها، می‌بایست توابع هدف در سیگنال پاداش و خطای عامل مربوط به هر وسیله از خانه هوشمند گنجانده شود. به دلیل اجتناب از تکرار و وجود شباهت در روند تعیین سیگنال پاداش و خطا، فرم کلی طراحی این سیگنال که برای تمامی وسایل مدل شده در قسمت قبل صادق است، در رابطه (۲۲) نشان داده شده است.

$$R = \beta_1(P_{appliance}) + \beta_2(discomfort_{appliance}) \quad (22)$$

شایان ذکر است که عبارت دوم رابطه (۲۲) برای خودرو برقی، جمع نارضایتی مشترک (رابطه ۲۰) و استهلاک باتری (رابطه ۲۱) است [۳۳]. در رابطه فوق، ضریب β_1 وزن عبارت هزینه مصرف انرژی هر وسیله (عامل) و β_2 وزن عبارت رضایت مشترک است. طبیعی است که این ضرایب برای هر یک از وسایل، مقادیر مختلفی را با توجه به اهمیت وسیله و همچنین همگرایی مسئله به خود اختصاص می‌دهند که در مطالعات عددی بدان پرداخته خواهد شد.

۳.۳. مطالعات عددی

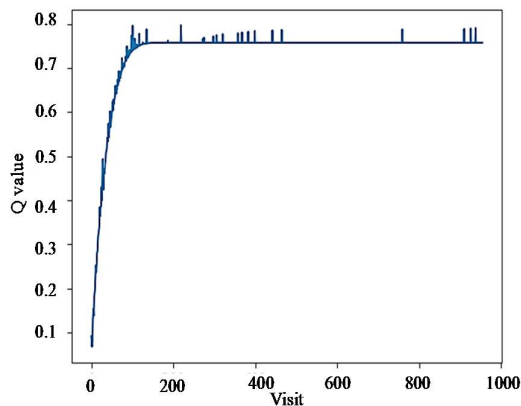
در این بخش با مطالعه سه سناریوی مختلف و بحث در نتایج آن‌ها به بررسی الگوریتم پیشنهادی در این مقاله پرداخته می‌شود. محیط شبیه‌سازی و ادیتور استفاده شده در این مقاله، به ترتیب python و spyder هستند.

۳.۱. فرضیات مسئله

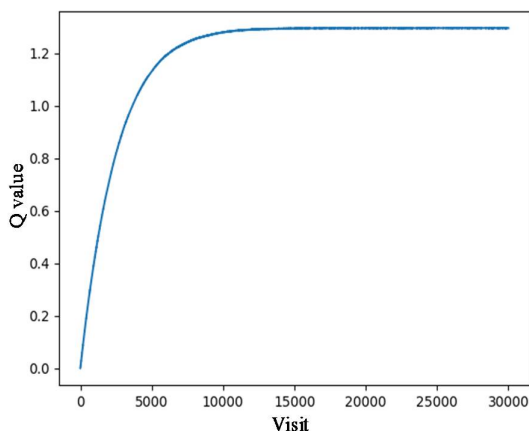
قیمت برق برای ۲۴ ساعت از منبع [۳۴] اخذ شده است که در شکل (۳) مشاهده می‌شود. ضریب آینده‌نگری γ در الگوریتم Q-Learning برابر ۰/۹ در نظر گرفته شده تا قابلیت توجه به پاداش‌های بلندمدت به عامل مسئله اضافه گردد. نرخ یادگیری α نیز برابر ۰/۱ فرض شده است. به طور کلی هفت وسیله از خانه در این مطالعه در نظر گرفته شده که مشخصات این هفت وسیله در جدول (۱) نمایش داده

شده است.

چهار ساعت در روز تلویزیون روشن بوده که این چهار ساعت به صورت تصادفی و با تابع چگالی احتمال نرمال تولید شده‌اند. برای ماشین لباسشویی و ظرفشویی نیز اعدادی تصادفی در بازه زمانی مورد رضایت کاربر به عنوان زمان روشن شدن انتخاب شده‌اند. سیستم سرمایشی به گونه‌ای تنظیم شده که دما همیشه در حدود ۱۸ درجه سانتی‌گراد نگه داشته شود. برای سیستم روشنایی فرض شده که این سیستم از ساعت ۶ تا ۱۲ با توان ۰/۴ کیلووات، ۱۲ تا ۱۸ با توان ۰/۶ و از ۱۸ تا ۲۳ با توان حداکثر توان یعنی ۰/۸ کیلووات فعال است. در نهایت برای خودرو برقی نیز ساعت ورود و خروج راننده و همچنین درصد شارژ باتری به هنگام ورود به خانه با تابع توزیع احتمال نرمال تولید و در قدم بعد فرض شده است که با حداکثر توان قابل شارژ، باتری به حالت شارژ کامل رسانده شده و سپس شارژ قطع می‌شود. نتایج حاصل از این رویکرد در جدول (۲) نمایش داده شده است.



شکل (۴): نمودار همگرایی عامل سیستم روشنایی

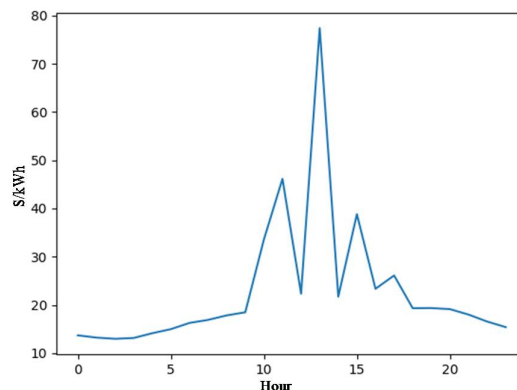


شکل (۵): نمودار همگرایی عامل ماشین لباسشویی

جدول (۲): نتایج سناریو اول (بدون مدیریت مصرف انرژی توسط

۲.۳. نتایج و بحث در نتایج

ابتدا برای بررسی همگرایی الگوریتم Q-Learning، در شکل (۴) و (۵) نمودار Q-Value مربوط به عامل سیستم روشنایی به نمایندگی از بارهای قابل کنترل و همین طور عامل ماشین لباسشویی به نمایندگی از بارهای قابل جابه‌جایی نمایش داده شده‌اند. نکته اینکه با توجه به اشکال (۴) و (۵) دیده می‌شود که عامل سیستم روشنایی پس از حدود ۲۰۰ بار مشاهده هر حالت به همگرایی رسیده ولی عامل بار قابل جابه‌جایی پس از حدود ده هزار مشاهده حالت‌ها به همگرایی رسیده است. در ادامه به بررسی سناریوهای در نظر گرفته شده در این بخش پرداخته شده است.



شکل (۳): قیمت برق در ۲۴ ساعت از منطقه هوستون آمریکا [۳۳]

جدول (۱): مشخصات وسایل خانه هوشمند مطالعه شده

بازه استفاده	توان (kW)	وسیله	نوع بار
۰-۲۳	۰/۲	یخچال	غیر قابل جابه‌جایی
۰-۲۳	۰/۳	تلویزیون	غیر قابل جابه‌جایی
۹-۱۶	۱/۵	ماشین لباسشویی	قابل جابه‌جایی
۹-۱۲	۱/۶	ماشین ظرفشویی	قابل جابه‌جایی
۰-۲۳	[۰/۶، ۰/۹، ۱/۲]	سیستم سرمایشی	قابل کنترل
۰/۳			
۶-۲۳	[۰/۴، ۰/۶، ۰/۸]	سیستم روشنایی	قابل کنترل
[۰/۲			
۱۸-۶	-۳، ۰، +۳، +۶	خودرو برقی	قابل کنترل
[۰/۶،			

• سناریو اول: مصرف بدون مدیریت انرژی

در این سناریو، مصرف انرژی و قبض برق مشترک بدون استفاده از یادگیری تقویتی و بر اساس رفتار تصادفی مشترک برای هر یک از وسایل محاسبه شده است. فرضاً برای یخچال فرض شده که در طول ۲۴ ساعت می‌باشد. برای تلویزیون نیز فرض شده است که مجموعاً

ساعت، هزینه روشن بودن وسیله در ساعات آتی و میزان ناراضیاتی حاصل از تأخیر در روشن کردن آن میسر شده است. در خصوص سیستم روشنایی و سرمایشی نیز، عامل مربوط توانسته است تا با وجود ترتیب اثر دادن رضایت مشترک، به گونه‌ای رفتار کند که کاهش مصرف انرژی در کنار جلب رضایت مشترک فراهم شود. جدول (۴) نمایانگر دمای خانه با توجه به توان در نظر گرفته شده توسط عامل برای سیستم سرمایشی است که بازه ۶ صبح تا ۱۸ بعد از ظهر به عنوان نمونه نمایش داده شده‌اند.

هزینه مصرف (\$)	مجموع مصرف (kWh)	وسيله
۱/۱۲	۴/۸	یخچال
۰/۳۵	۱/۲	تلویزیون
۰/۷۷۸۵	۳	ماشین لباس شویی
۰/۸۳۰۴	۳/۲	ماشین ظرف شویی
۳/۱۷	۱۳/۸	سیستم سرمایشی
۲/۱	۹/۴	سیستم روشنایی
۲/۱۲	۳۰+ شارژ و ۱۵- دشارژ	خودرو برقی (سناریو دوم)
۳/۱۲۳	۱۸+ شارژ	خودرو برقی (سناریو سوم)
۱۰/۴۶۸	۵۰/۴	کل وسایل (سناریو دوم)
۱۱/۴۷۲	۵۳/۴	کل وسایل (سناریو سوم)

بدیهی است که میزان درصد کاهش مصرف انرژی و قبض برق، وابستگی بسیار زیادی به نحوه مدل کردن ناراضیاتی مشترک، وزن این عبارت در تابع هدف هر وسیله و مشخصات مصرف وسایل در نظر گرفته شده دارد. برای مثال، ضریب β_2 در عبارت مربوط به تأخیر در زمان شروع به کار بار قابل جابه‌جایی و همین‌طور عبارت رضایت دما سیستم تهویه، هر دو برابر با $0/1$ و ضریب β_1 این وسایل برابر با 1 در نظر گرفته شده‌اند. این ضرایب جزء هاپرپارامترهای مسئله محسوب می‌شوند و با آزمون و خطا به دست آمده‌اند. شایان ذکر است که نسبت یک به ده برای β_1 و β_2 به معنی ارزش دهبرابری قبض برق نسبت به رضایت مشترک نیست چراکه مقیاس عددی توابع تأثیر مهمی در تصمیم‌عامل یادگیری تقویتی دارد. در این مقاله دمای مطلوب ساکنان منزل بین ۱۸ تا ۲۲ درجه سانتی‌گراد فرض شده است. با توجه به جدول (۴) دیده می‌شود که با در نظر گرفتن مقدار ضریب β_2 سیستم سرمایشی برابر $0/1$ ، نهایتاً تا $0/4$ درجه سانتی‌گراد از دمای ۲۲ درجه انحراف مشاهده می‌شود و هیچ‌گاه نیز دمای خانه به زیر ۱۸ درجه نمی‌رسد.

جدول (۴): عملکرد ساعتی سیستم سرمایشی در سناریو سوم برای

هزینه مصرف (\$)	مجموع مصرف (kWh)	وسيله
۱/۱۲	۴/۸	یخچال
۰/۳۵	۱/۲	تلویزیون
۱/۴۸۶	۳	ماشین لباس شویی
۱/۵۸	۳/۲	ماشین ظرف شویی
۴/۲۷	۱۸/۳	سیستم سرمایشی
۲/۵۸	۱۰	سیستم روشنایی
۳/۸۷	۱۸	خودرو برقی
۱۵/۲۵۶	۵۸/۵	کل

• سناریو دوم: مدیریت انرژی با روش پیشنهادی بدون در نظر گرفتن استهلاك باتری خودرو

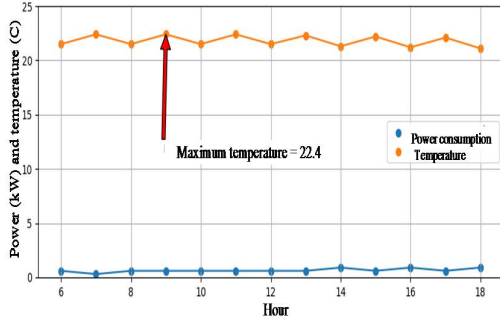
در این سناریو، روش پیشنهادی مدیریت مصرف انرژی ساعتی با کمک یادگیری تقویتی چندعاملی روی خانه مورد مطالعه اجرا شده است. در قدم بعد، نتایج فرمان کنترلی عامل هریک از وسایل روی وسیله مربوطه پیاده‌سازی و سپس به محاسبه انرژی مصرفی و قبض برق مشترک پرداخته شده است. همچنین به منظور مشارکت ذخیره‌ساز خودرو برقی در خرید و فروش برق، از اثر استهلاك باتری به دلیل شارژ و دشارژ متعدد صرف‌نظر شده است.

• سناریو سوم: مدیریت انرژی با روش پیشنهادی با در نظر گرفتن استهلاك باتری خودرو

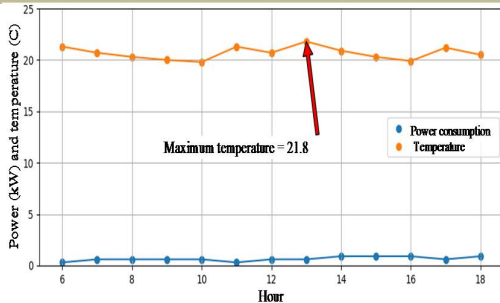
در این سناریو، مشابه سناریوی دوم عمل شده، با این تفاوت که استهلاك باتری خودرو نیز در نظر گرفته شده است. انتظار می‌رود که در این حالت، عامل خودرو به دلیل در نظر گرفتن اثر شارژ و دشارژ متوالی در استهلاك باتری، تمایل کمتری برای مشارکت در فروش برق از خود نشان دهد. همان‌طور که ذکر شد، صرفاً نتایج عامل خودرو برقی در سناریو ۲ و ۳ با یکدیگر متفاوت‌اند و سایر وسایل نتایج مشابهی در این دو سناریو دارند. از این‌رو نتایج ادغام‌شده سناریوی دوم و سوم به استثنای خودرو برقی به صورت تفکیک‌شده برای هر سناریو، در جدول (۳) نتایج نمایش داده است.

با توجه به جداول (۲) و (۳) دیده می‌شود که یادگیری تقویتی چندعاملی، به گونه‌ای مدیریت مصرف انرژی ساعتی یک خانه هوشمند را انجام داده است که نسبت به حالت بدون استفاده از این الگوریتم، به ترتیب $31/3\%$ و $24/8\%$ کاهش در هزینه مصرف انرژی مشاهده می‌شود. همچنین نکته جالب توجه اینکه عامل ماشین لباس شویی و ظرف شویی، توانسته به رغم ثابت بودن انرژی مصرفی آن‌ها در هر سناریو، به هزینه کمتری در سناریوی دوم و سوم منجر شود. این امر از طریق توجه عامل یادگیری تقویتی به قیمت در هر

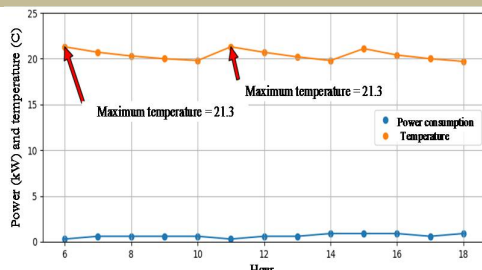
در سه حالت با ضرایب مختلف نارضایتی دما آموزش داده شده است. در این سه حالت، مقدار ضریب نارضایتی دما به ترتیب برابر با ۰/۱، ۰/۲ و ۰/۵ قرار داده شده است. نتیجه حاصل برای دمای لحظه‌ای و توان مصرفی در هر ساعت برای این سه حالت در قالب شکل‌های (۶)، (۷) و (۸) نمایش داده شده است.



شکل (۶): نمودار دمای لحظه‌ای خانه و توان مصرفی سیستم سرمایه‌ی به‌ازای $\beta_2=0.1$



شکل (۷): نمودار دمای لحظه‌ای خانه و توان مصرفی سیستم سرمایه‌ی به‌ازای $\beta_2=0.2$



شکل (۸): نمودار دمای لحظه‌ای خانه و توان مصرفی سیستم سرمایه‌ی به‌ازای $\beta_2=0.5$

همان‌طور که در شکل (۶) نشان داده شده و در سناریوی سوم هم بدان اشاره شد، در صورت اختصاص دادن مقدار ۰/۱ به ضریب نارضایتی، تا ۰/۴ درجه سانتی‌گراد از حداکثر دمای مطلوب در نظر گرفته شده برای ساکنان (۲۲ درجه سانتی‌گراد) انحراف مشاهده می‌شود. با افزایش این ضریب به ۰/۲، انتظار می‌رود که افزایش مصرف انرژی الکتریکی به‌منظور جلب بیشتر رضایت مشترک روی دهد. از این رو در شکل (۷) مشاهده می‌شود که حداکثر دمای خانه هوشمند در این شرایط از ۲۲/۴ به ۲۱/۸ درجه کاهش پیدا کرده

بازه ۶ صبح تا ۱۸ بعد از ظهر

ساعت	توان	دمای سابق	دمای جدید
۶	۰/۶	۲۲/۴	۲۱/۵
۷	۰/۳	۲۱/۵	۲۲/۴
۸	۰/۶	۲۲/۴	۲۱/۵
۹	۰/۳	۲۱/۵	۲۲/۴
۱۰	۰/۶	۲۲/۴	۲۱/۵
۱۱	۰/۶	۲۱/۵	۲۲/۴
۱۲	۰/۶	۲۲/۴	۲۱/۵
۱۳	۰/۶	۲۱/۵	۲۲/۳
۱۴	۰/۹	۲۲/۳	۲۱/۳
۱۵	۰/۶	۲۱/۳	۲۲/۲
۱۶	۰/۹	۲۲/۲	۲۱/۲
۱۷	۰/۶	۲۱/۲	۲۲/۱
۱۸	۰/۹	۲۲/۱	۲۱/۱

در خصوص خودرو برقی با توجه به جدول (۳) دیده می‌شود که در سناریوی دوم که از استهلاک باتری صرف‌نظر شده، عامل خودرو علاقه‌مند به فروش انرژی بوده (۱۵ کیلووات ساعت) که این موضوع باعث کاهش قابل توجه هزینه خودرو برقی شده است که دور از انتظار نیز نبود. اما نکته جالب توجه درباره خودرو برقی، در مقایسه سناریوی اول و سوم است. با مشاهده انرژی مصرف‌شده برای شارژ باتری در سناریوی اول و سوم دیده می‌شود که در هر دو، ۱۸ کیلووات ساعت انرژی برای شارژ باتری استفاده شده اما در سناریوی سوم این مصرف انرژی یکسان منجر به هزینه کمتری گردیده است. این موضوع ناشی از آن است که در سناریوی سوم، به‌جای شارژ یکبار باتری، با توجه به ساعات ارزان‌تر بودن قیمت اقدام به شارژ باتری شده است.

گفتنی است که با توجه به تفاوت در وسایل، توابع هدف مدل‌شده، مشخصات نامی وسایل خانه و قیمت برق در مطالعات پیشین، امکان مقایسه دقیق بین نتایج الگوریتم‌های استفاده‌شده در این دسته از پژوهش‌ها میسر نیست. اما با مقایسه درصد کاهش قبض برق حاصل‌شده در این کار با سایر پژوهش‌هایی که اقدام به حل مسئله مدیریت مصرف انرژی خانگی از طریق Q_Learning نموده‌اند، مشاهده می‌شود که روش پیشنهادی در این مقاله، نتایجی نزدیک به کارهای مشابه [۲۵ و ۳۱] را حاصل کرده است.

۳.۳ آنالیز اثر ضریب نارضایتی مشترک

در سناریوی سوم به‌صورت خاص عملکرد سیستم سرمایه‌ی در صورت اختصاص دادن مقدار ۰/۱ به ضریب نارضایتی مشترک در قالب جدول (۴) نمایش داده شد. در این قسمت، هدف بررسی تأثیر این ضریب در شرایط مختلف است؛ از این رو عامل سیستم سرمایه‌ی

هوشمند، انواع عوامل تأثیرگذار در رضایت مشترک نظیر رضایت دما، تأخیر در زمان روشن کردن وسایل، رضایت از حیث روشنایی منزل و شارژ باتری خودرو، همگی در کنار هم دیده شده و یک برنامه مدیریت مصرف انرژی به صورت ساعتی حاصل گردد. نتایج حاصل از مطالعات عددی حاکی از آن است که روش پیشنهادی این توانایی را دارد تا با کمترین اختلال در رضایت مشترک، $24/8\%$ قبض برق مشترک را کاهش دهد. همچنین آنالیز حساسیت ضریب نارضایتی مشترک نشان داد که می توان با تنظیم مقدار این ضریب، به سطوح مختلفی از رضایت مشترک و طبیعتاً مصرف انرژی الکتریکی دست یافت. همان طور که انتظار می رفت با افزایش این ضریب، جلب بیشتر رضایت دمایی مشترک و همچنین افزایش بیشتر مصرف انرژی الکتریکی مشاهده شد.

است. شایان ذکر است که در این حالت، سیستم سرمایه‌ی به جای $13/8$ کیلووات ساعت (حالت اول)، $15/89$ کیلووات ساعت انرژی مصرف کرده است. در حالت سوم، مقدار ضریب نارضایتی برابر با $0/5$ در نظر گرفته شده است. در این شرایط انرژی مصرفی سیستم سرمایه‌ی و دمای خانه تغییر محسوسی نسبت به حالت دوم نداشته‌اند؛ به استثنای اینکه حداکثر دمای خانه هوشمند از $21/8$ درجه سانتی‌گراد به $21/3$ درجه کاهش پیدا کرده که طبق شکل (۸) در ساعات ۶ و ۱۲ این دما مشاهده می‌شود.

۴. نتیجه گیری

در این مقاله روشی برای مدیریت مصرف انرژی خانگی به صورت ساعتی ارائه شد. بهره‌گیری از یادگیری تقویتی چندعاملی، نویسندگان را قادر ساخت تا ضمن مدل کردن همهٔ مسائل حائز اهمیت در خانهٔ

مراجع

- [1] U. E. I. Administration, "International energy outlook 2017", [Online]. Available: <https://www.eia.gov/outlooks/ieo/>.
- [2] Wen, Z., O'Neill, D. and Maei, H., "Optimal demand response using device-based reinforcement learning", in IEEE Transactions on Smart Grid, Vol. 6, No. 5, pp. 2312-2324, 2015.
- [3] Yu, L., et al., "Deep Reinforcement learning for smart home energy management", in IEEE Internet of Things Journal, Vol. 7, No. 4, pp. 2751-2762, 2020.
- [4] Wan, Z., Li, H. and He, H., "Residential energy management with deep reinforcement learning", 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, pp. 1-7, 2018.
- [5] Wang, B., Li, Y., Ming, W. and Wang, S., "Deep reinforcement learning method for demand response management of interruptible load", in IEEE Transactions on Smart Grid, Vol. 11, No. 4, pp. 3146-3155, 2020.
- [6] Lu, R., Ho Hong, S., "Incentive based demand response for smart grid with reinforcement learning and deep neural network", in Applied Energy, Vol. 236, pp. 937-949, 2019.
- [7] Liu, Y., Zhang, D. and Gooi, H. B., "Optimization strategy based on deep reinforcement learning for home energy management", in CSEE Journal of Power and Energy Systems, Vol. 6, No. 3, pp. 572-582, 2020.
- [8] Ruelens, F., et al., "Direct load control of thermostatically controlled loads based on sparse observations using deep reinforcement learning", CSEE Journal of Power and Energy Systems, Vol. 5, pp. 423-432, 2019.
- [9] Zhang, D., Han, X. and Deng, C., "Review on the research and practice of deep learning and reinforcement learning in smart grids", in CSEE Journal of Power and Energy Systems, Vol. 4, No. 3, pp. 362-370, 2018.
- [10] Huang, X., Hong, S. H., Yu, M., Ding, Y. and Jiang, J., "Demand response management for industrial facilities: a deep reinforcement learning approach", in IEEE Access, Vol. 7, pp. 82194-82205, 2019.
- [11] Lu, R., Hog, S. H. and Zhang, X., "A dynamic pricing demand response algorithm for smart grid: reinforcement learning approach", in Applied Energy, Vol. 220, pp. 220-230, 2018.
- [12] Du, Y. and Li, F., "Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning", in IEEE Transactions on Smart Grid, Vol. 11, No. 2, pp. 1066-1076, 2020.
- [13] Mocanu, E., et al., "On-Line building energy optimization using deep reinforcement learning", in IEEE Transactions on Smart Grid, Vol. 10, No. 4, pp. 3698-3708, 2019.
- [14] Ye, Y., Qiu, D., Wu, X., G. Strbac and Ward, J., "Model-Free real-time autonomous control for a residential multi-energy system using deep reinforcement learning", in IEEE Transactions on Smart Grid, Vol. 11, No. 4, pp. 3068-3082, 2020.
- [15] Li, H., Wan, Z. and He, H., "A deep reinforcement learning based approach for home energy management system", 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, pp. 1-5, 2020.
- [16] Mohammadi, M., Al-Fuqaha, A., Guizani, M., and Oh, J., "Semisupervised deep reinforcement learning in support of iot and smart city services", in IEEE Internet of Things Journal, Vol. 5, No. 2, pp. 624-635, 2018.
- [17] Tai, C., Hong, J. and Fu, L., "A real-time demand-side management system considering user behavior using deep q-learning in home area network", 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, pp. 4050-4055, 2019.
- [18] Alfaverh, F., Denaï, M. and Sun, Y., "Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management", in IEEE Access, Vol. 8, pp. 39310-39321, 2020.
- [19] Al-jabery, K., Wunsch, D. C., Xiong, J. and Shi, Y., "A novel grid load management technique using electric water heaters and Q-learning", 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, pp. 776-

- 781, 2014.
- [20] Al-jabery, K., et al., "*Demand-Side management of domestic electric water heaters using approximate dynamic programming*", in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 36, No. 5, pp. 775-788, 2017.
- [21] Al-jabery, K., Xu, Z., Yu, W., Wunsch, D. C., Xiong, J. and Shi, Y., "*Demand-Side management of domestic electric water heaters using approximate dynamic programming*", in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 36, No. 5, pp. 775-788, 2017.
- [22] Ruelens, F., et al., "*Reinforcement learning applied to an electric water heater: from theory to practice*", in IEEE Transactions on Smart Grid, Vol. 9, No. 4, pp. 3792-3800, 2018.
- [23] Patyn, C., Ruelens, F. and Deconinck, G., "*Comparing neural architectures for demand response through model-free reinforcement learning for heat pump control*", 2018 IEEE International Energy Conference (ENERGYCON), Limassol, pp. 1-6, 2018.
- [24] Al-jabery, K., Xu, Z., Yu, W., Wunsch, D. C., Xiong, J. and Shi, Y., "*A novel grid load management technique using electric water heaters and Q-learning*", 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, pp. 776-781, 2014.
- [25] Xu, X., et al., "*A Multi-Agent reinforcement learning-based data-driven method for home energy management*", in IEEE Transactions on Smart Grid, Vol. 11, No. 4, pp. 3201-3211, 2020.
- [26] Vázquez-Canteli, R. J. and Nagy, Z., "*Reinforcement learning for demand response: a review of algorithms and modeling techniques*", in Applied Energy, Vol. 235, pp. 1072-1089, 2019.
- [27] Sutton, S. R. and Barto, A., *Introduction to reinforcement learning* (No. 4). MIT press Cambridge, 1998.
- [28] Mason, K. and Grijalva, S., "*A review of reinforcement learning for autonomous building energy management*", in Comput Electr Eng, Vol. 78, pp. 300–12, 2019.
- [29] Mathew, A., Roy, A. and Mathew, J., "*Intelligent residential energy management system using deep reinforcement learning*", in IEEE Systems Journal, 2020.
- [30] Wu, Z., Zhou, S., Li, J. and Zhang, X., "*Real-Time scheduling of residential appliances via conditional risk-at-value*", in IEEE Transactions on Smart Grid, Vol. 5, No. 3, pp. 1282-1291, 2014.
- [31] Lu, R., Hong, S. H. and Yu, M., "*Demand Response for home energy management using reinforcement learning and artificial neural network*", in IEEE Transactions on Smart Grid, Vol. 10, No. 6, pp. 6629-6639, 2019.
- [32] Sun, B., Huang, Z., Tan, X. and Tsang, H. K. D., "*Optimal scheduling for electric vehicle charging with discrete charging levels in distribution grid*", IEEE Transactions on Smart Grid, Vol. 9, No. 2, pp. 624–634, 2018.
- [33] Wan, Z., Li, H., He, H. and Prokhorov, D., "*Model-Free real-time ev charging scheduling based on deep reinforcement learning*", in IEEE Transactions on Smart Grid, Vol. 10, No. 5, pp. 5246-5257, 2019.
- [34] <http://www.energyonline.com/Data/GenericData.aspx?DataId=4>