Computational Intelligence in Electrical Engineering
Vol. 14, No. 3, 2023
Research Paper

Optimal Decentralized Energy Management of Electrical and Thermal Distributed Energy Resources and Loads in Microgrids Using Reinforcement Learning

Razieh Darshi[1], Saeed Shamaghdari [2], Aliakbar Jalali [3], Hamidreza Arasteh[4]
[1] Ph.Dd student, Faculty member of the University of Science and Technology, Tehran, Iran
[2] Associate professor, Faculty member of the University of Science and Technology, Tehran Iran
[3] professor, Faculty member of the University of Science and Technology, Tehran Iran
[3] The member of Dept. of Power Systems Planning and Operation, Niroo Research Institute, Tehran, Iran

Abstract:
In this paper, a decentralized energy management system is presented for intelligent microgrids with the presence of distributed resources using reinforcement learning. Due to the unpredictable nature of renewable energy resources, the variability of load consumption, and the nonlinear model of batteries, the design of a microgrid energy management system is associated with many challenges. In addition, centralized control structures in large-scale systems increase computational volume and complexity in control algorithms. In this paper, a fully decentralized multi-agent structure for a microgrid energy management system is proposed and the Markov decision process is used to model the stochastic behavior of agents in the microgrid. Electrical and thermal distributed resources, batteries, and consumers are considered intelligent and independent agents. They have the learning ability to explore and exploit the environment in a fully decentralized manner and achieve their optimal policies. The proposed method for hourly microgrid management is model-independent and based on learning. The method maximizes the profits of all manufacturers, minimizes consumer costs, and reduces the dependence of the microgrid on the maingrid. Finally, using real data from renewable energy sources and consumers, the accuracy of the proposed method in the Iranian electricity market is simulated and verified.

Keywords: Multi-agent energy management system, Reinforcement learning, Markov decision process, Microgrid, Electrical and thermal distributed energy resources.

https://doi.org/10.22108/isee.2022.133209.1556

مقاله پژوهشی

مدیریت انرژی غیرمتمرکز بهینه منابع و بارهای پراکنده الکتریکی و گرمایی در ریزشبکه‌ها با استفاده از یادگیری تقویتی

راضیه دارشی[۱]، سعید شمقدری*[۲]، علی اکبر جلالی[۳]، حمیدرضا آراسته[۴]

۱- دانشجوی دکتری دانشکده مهندسی برق، دانشگاه علم و صنعت، تهران، ایران
raziehdarshi@yahoo.com

۲- دانشیار دانشکده مهندسی برق، دانشگاه علم و صنعت، تهران، ایران
shamaghdari@iust.ac.ir

۳- استاد دانشکده مهندسی برق، دانشگاه علم و صنعت، تهران، ایران
drjalali@iust.ac.ir

۴- استادیار، عضو گروه پژوهشی برنامه‌ریزی و بهره‌برداری سیستم‌های قدرت پژوهشگاه نیرو، تهران، ایران
harasteh@nri.ac.ir

چکیده: در این مقاله، یک سیستم مدیریت انرژی غیرمتمرکز برای ریزشبکه‌های هوشمند باحضور منابع پراکنده با استفاده از یادگیری تقویتی ارائه می‌شود. مسئله طراحی سیستم مدیریت انرژی ریزشبکه‌ها به دلیل ویژگی‌های پیش‌بینی‌ناپذیر خروجی منابع تجدیدپذیر، متغیربودن بارمصرفی و مدل غیرخطی باتری‌ها به‌منظور ذخیره و تأمین انرژی با چالش‌های زیادی روبه‌رو است. علاوه بر این، استفاده از ساختارهای کنترل متمرکز در سیستم‌های ابعاد وسیع، به بروز مشکلاتی ازقبیل افزایش حجم محاسباتی و پیچیدگی در الگوریتم‌های کنترلی منجر می‌شود. در این مقاله، ضمن ارائه یک ساختار کاملاً غیرمتمرکز چندعامله برای سیستم مدیریت انرژی، از پروسه‌های تصمیم‌گیری مارکوف برای مدل‌سازی رفتار تصادفی عامل‌ها در ریزشبکه استفاده می‌شود. منابع پراکنده الکتریکی و گرمایی، باتری و مصرف‌کنندگان، عامل‌های هوشمند و مستقل در نظر گرفته می‌شوند که دارای توانایی یادگیری هستند و پس از اکتشاف محیط و بهره‌برداری به‌صورت کاملاً غیرمتمرکز، سیاست بهینه خود را به دست می‌آورند. روش ارائه‌شده برای برنامه‌ریزی ساعتی ریزشبکه، یک روش مستقل از مدل و مبتنی بر یادگیری است که ضمن بیشینه‌کردن سود کلیه تولیدکنندگان، هزینه مصرف‌کنندگان را کمینه و از وابستگی ریزشبکه به شبکه اصلی نیز می‌کاهد. درنهایت با استفاده از داده‌های واقعی از نیروگاههای انرژی تجدیدپذیر در ایران و داده‌های خرید و فروش انرژی در بازار برق ایران، دقت روش پیشنهادی شبیه‌سازی و ارزیابی می‌شود.

واژه‌های کلیدی: سیستم مدیریت انرژی چندعاملی، یادگیری تقویتی، تصمیم‌گیری مارکوف، ریزشبکه، منابع تولید پراکنده الکتریکی و گرمایی.

۱- مقدمه

یکی از مهم‌ترین تغییرات درحال انجام در شبکه‌های قدرت، انتقال از منابع انرژی متمرکز سنتی به منابع انرژی پراکنده است. منابع انرژی پراکنده با توجه به فواید زیست‌محیطی بسیار آنها نقش کلیدی در تولید انرژی پاک و پایدار دارند [۱]. منابع تولید پراکنده می‌توانند انتشار کربن، تلفات ارسال توان و هزینه ساخت زیربناها را کاهش دهند [۲]. ریزشبکه‌ها شبکه‌های قدرت مقیاس کوچک و خودحمایت‌گر هستند که از منابع انرژی پراکنده استفاده

[۱] تاریخ ارسال مقاله: ۱۴۰۱/۰۱/۲۰
تاریخ پذیرش مقاله: ۱۴۰۱/۰۹/۲۷
نام نویسندهٔ مسئول: سعید شمقدری
نشانی نویسندهٔ مسئول: ایران، دانشگاه علم و صنعت، تهران، دانشکده مهندسی برق

می‌کنند. منابع انرژی پراکنده می‌تواند شامل منابع انرژی تجدیدپذیر مانند توربین‌های بادی و پنل‌های خورشیدی، منابع انرژی تجدیدناپذیر ازجمله دیزل ژنراتورها و سیستم‌های ذخیره انرژی باتری باشد. ریزشبکه‌ها می‌توانند در دو مد متصل به شبکه اصلی یا غیرمتصل (جزیره‌ای) کار کنند [۳].

ریزشبکه‌ها قادرند قابلیت اطمینان شبکه را بهبود ببخشند و انرژی با کیفیت و پایدار را تأمین کنند؛ اما برنامه‌ریزی و عملکرد آنها به دلیل عدم‌قطعیت ناشی از پیش‌بینی بار مصرفی و توان خروجی انرژی‌های تجدیدپذیر با چالش‌های زیادی مواجه است. اگرچه ریزشبکه‌ها با مهیاکردن یک مسیر انعطاف‌پذیر برای ترکیب منابع انرژی پراکنده تجدیدپذیر به شبکه‌های قدرت، نقش مهمی در مسیر مدرنیته‌کردن آنها دارند، منابع تجدیدپذیر با توجه به شرایط مختلف جوی، متغیر و پیش‌بینی‌ناپذیر است؛ برای مثال، ماژول‌های فتوولتائیک، تنها در حضور تابش خورشید می‌توانند برق تولید کنند. همچنین، مزارع بادی نیز در حضور باد کافی قادر به تولید برق هستند. علاوه بر مشکلات ناشی از وجود عدم‌قطعیت در ریزشبکه‌ها، لازم است از وابستگی ریزشبکه‌ها به شبکه اصلی نیز کاسته شود؛ به نحوی که سود کلیه واحدهای تولید انرژی افزایش و هزینه مصرف‌کنندگان داخلی در ریزشبکه‌ها نیز کاهش یابد. سیستم‌های ذخیره انرژی باتری به‌منظور ذخیره و تأمین برق در زمان‌های مختلف در ریزشبکه‌ها استفاده می‌شوند. حداکثر میزان انرژی قابل شارژ یا تخلیه در یک زمان مشخص در یک باتری وابسته به قابلیت ذخیره انرژی، حالت شارژ فعلی (SOC) و مشخصات ذاتی آن است. همچنین، SOC فعلی باتری با رفتار شارژ/ تخلیه قبلی مشخص می‌شود؛ بنابراین، مدیریت انرژی باتری به یک مسئله تصمیم‌گیری ترتیبی برای یک سیستم دینامیکی تبدیل می‌شود که تصمیمات قبلی بر گزینه‌های موجود آتی تأثیرگذار است. باتری‌ها دارای خواص متغیربازمان و غیرخطی هستند؛ از این رو برنامه‌ریزی شارژ و تخلیه آنها چالش‌های جدیدی به سیستم مدیریت انرژی اضافه می‌کند و نیازمند الگوریتم‌های کنترلی پیشرفته‌تری است [۴].

روش‌های مدیریت انرژی هوشمند به‌تازگی مورد توجه بسیاری قرار گرفته است. این روش‌ها برای حل مشکلات ناشی از عدم‌قطعیت و تغییرپذیری در ریزشبکه‌ها و ایجاد توان قابل اعتماد برای شبکه‌های قدرت استفاده می‌شود.

روش‌های مختلفی در سال‌های اخیر برای حل مشکل عدم‌قطعیت در مدیریت انرژی ریزشبکه‌ها توسعه داده شده است [۵]. در [۶] یک استراتژی کنترل برای عملکرد هماهنگ ریزشبکه‌ها در یک سیستم توزیع ارائه می‌دهد. اپراتور شبکه توزیع و هر ریزشبکه به‌عنوان یک واحد مجزا با توابع هدف منحصربه‌فرد در نظر گرفته می‌شود؛ به طوری که هزینه‌های عملیاتی کمینه شود. مسئله به‌عنوان یک مسئله دوسطحی تصادفی فرموله شده است. در سطح بالا، اپراتور شبکه و در سطح پایین، ریزشبکه‌ها در نظر گرفته شده است. در [۷] یک مسئله تصادفی برای برنامه‌ریزی انرژی ریزشبکه‌ها و حل چالش‌های عملیاتی بارهای کنترل‌پذیر و منابع انرژی تنظیم می‌کند؛ به طوری که هزینه‌های عملیاتی مورد انتظار ریزشبکه و تلفات توان را به حداقل رساند و با طبیعت متناوب انرژی‌های تجدیدپذیر سازگار است. در [۸] یک مدل بهینه‌سازی محدب مقاوم برای مدیریت انرژی ریزشبکه‌ها ارائه شده است. در حالت متصل به شبکه اصلی، هزینه انرژی‌های وارده به شبکه، بارهای تأمین‌شده با منابع انرژی پراکنده و باتری کمینه می‌شود. در حالت جزیره‌ای بارهای تأمین‌نشده، با در نظر گرفتن اولویت مصرف به حداقل می‌رسد. در روش‌های بهینه‌سازی لیاپانف، کنترل بهینه یک سیستم پویا ازطریق تابع لیاپانف به دست می‌آید. در [۹] به توسعه یک روش مدیریت انرژی برخط برای کارکرد زمان حقیقی ریزشبکه‌ها با در نظر گرفتن پخش بار و قیود عملکردی سیستم پرداخته شده است. مدیریت انرژی برخط به‌عنوان مسئله تصادفی پخش بار بهینه، مدل‌سازی و از بهینه‌سازی لیاپانف استفاده شده است. در مقاله [۱۰]، روش بهینه‌سازی دو مرحله‌ای ارائه شده است. در مرحله اول، برنامه ساعتی با استفاده از برنامه‌ریزی روز آینده انجام می‌گیرد. در مرحله دوم، پخش بار اقتصادی و مبادله انرژی با استفاده از بهینه‌سازی لیاپانوف به‌صورت زمان حقیقی

برنامه‌ریزی می‌شود. در [۱۱] در ساختار ریزشبکه از دو کنترل‌کننده مرکزی برای ریزشبکه و شبکه گاز استفاده شده است. برای کاهش هزینه‌های بهره‌برداری، مسئله خرید و فروش انرژی در ریزشبکه با استفاده از یک مدل خطی عدد صحیح‌آمیخته، مدلسازی و به کمک نرم‌افزار GAMS مسئله فوق حل شده است.

در روش‌های فوق فرض می‌شود مشخصات پروسه‌ها و متغیرهای تصادفی با مدل‌های پیش‌بینی‌کننده یا مقادیر مورد انتظار مربوط به آنها موجود است. وابستگی این روش‌ها به یک مدل تخمین‌گر منجر شده است تا دقت این روش‌ها به دقت مدل تخمین زده شده وابسته باشد. یادگیری تقویتی روش‌های مستقل از مدل برای حل مسائل کنترل بهینه سیستم‌های دینامیکی ارائه می‌دهد که نیازی به وجود خواص تصادفی پروسه‌های تصادفی نیست.

در [۱۲]، سیستم مدیریت انرژی یک ساختمان برای کاهش پیک مصرف انرژی با استفاده الگوریتم فراابتکاری دومرحله‌ای پیاده‌سازی شده است. هزینه و مصرف انرژی با حضور انرژی‌های تجدیدپذیر با استفاده از الگوریتم ژنتیک در [۱۳] کمینه شده است. در [۱٤]، از الگوریتم بهینه‌سازی بویایی کوسه و گرگ خاکستری برای ارزیابی تأثیر پاسخگویی بار در ریزشبکه‌ها استفاده شده و هزینه تولید و تلفات شبکه کمینه شده است. گرچه روش‌هایی که از الگوریتم‌های ابتکاری استفاده می‌کنند به مدل ریاضی نیازی ندارند و برای بهینه‌سازی جهانی برنامه‌ریزی غیرخطی و غیرهموار، به‌طور منعطف‌تر و موثرتری عمل می‌کنند، بیشتر این روش‌ها قادر به یادگیری تقلیدی و ذخیره دانش قبلی نیستند و در هر مرحله یک جمعیت جدید به‌صورت تصادفی انتخاب می‌شود؛ به همین دلیل، زمان اجرا برای محاسبه نقطه بهینه زیاد می‌شود [۱۵].

از نظریه بازی‌ها نیز در طراحی سیستم مدیریت انرژی ریزشبکه‌ها استفاده می‌شود [۱٦]. در مقاله [۱۷]، برنامه‌ریزی روز آینده ریزشبکه‌ها و شرکت توزیع با استفاده از نظریه بازی‌ها پیاده‌سازی شده است. برنامه‌ریزی تقاضا و سیستم ذخیره انرژی به‌عنوان یک مسئله بهینه‌سازی چندتابع هدفه مدلسازی شده است. خروجی منابع تجدیدپذیر با استفاده از [۱۸] تخمین زده و به کمک تکنیک نظریه بازی در مقاله [۱۹] مسئله مدیریت انرژی حل شده است. در مقاله [۲۰]، با استفاده از تئوری بازی‌های غیر مشارکتی، مدیریت انرژی چندعامله یک ریز شبکه با حضور منابع انرژی تجدیدپذیر و بارهای فصلی پیاده‌سازی شده است. در مقاله [۲۱]، برنامه‌ریزی یک شبکه قدرت در حالت متصل به شبکه با استفاده از مدل نظریه بازی‌های مشارکتی و غیرمشارکتی به دست آمده است. توربین‌های بادی، پنل خورشیدی و باتری‌ها بازیکن در مسئله در نظر گرفته شده است. وجود نقطه تعادل نش با تحلیل تقعر توابع بازده و مدل عدم قطعیت اثبات شده است. برای یافتن نقطه تعادل در مقالات [۲۰, ۲۱]، از یک روش جستجو تکراری [۲۲] استفاده شده است. همچنین، از پیش‌بینی سرعت باد، شدت تابش خورشید و تقاضای بار و الگوریتم ازدحام ذرات نیز استفاده شده است. در مسائل فوق برای حل مسئله به روش نظریه بازی‌ها، سیستم تنها شامل یک حالت در نظر گرفته می‌شود. در نظریه بازی‌ها یک فرض اساسی وجود دارد که بازیکن‌ها دارای توانایی یادگیری و سازگاری یکسانی هستند. بیشتر الگوریتم‌های نظریه بازی به مدلی از محیط نیاز دارند؛ زیرا از انتقال، T و پاداش، r استفاده می‌کنند. هدف این الگوریتم‌ها نیز محاسبه ارزش تعادلی بازی (یعنی پاداش‌های وزن‌دار مورد انتظار برای هر یک از عوامل)، به جای یافتن سیاست‌های تعادلی است. این بدان معنی است که آنها اغلب الزامات قوی برای رفتار همه عوامل ایجاد می‌کنند. درمقابل، الگوریتم‌های یادگیری تقویتی فرض می‌کنند جهان ناشناخته است و فقط مشاهدات و پاداش اعمال آنها از محیط موجود است. هدف در روش‌های یادگیری تقویتی این است که عامل خط‌مشی خود را در راه‌حل تعادل بازی بیابد و معمولاً الزامات کمتری برای رفتار عامل‌های دیگر ایجاد می‌کند [۲۳]؛ به همین دلیل به‌تازگی به استفاده از روش‌های یادگیری تقویتی توجه زیادی شده است. روش‌های یادگیری تقویتی در مقایسه با روش‌های یادگیری نظارتی و غیرنظارتی در حیطه انرژی، قابلیت‌های جالبی را در زمینه

کاربردهای کنترلی ارائه می‌دهد [۲٤]. در سیستم‌های کنترلی، با توجه به اینکه به دست آوردن اطلاعات اولیه از محیط و مدل‌سازی سیستم در عمل بسیار سخت است، یادگیری تقویتی قادر است روش‌هایی مستقل از مدل برای حل مسائل دارای عدم‌قطعیت ارائه دهد. این روش‌ها با استفاده از ذخیره اطلاعات تجربه‌شده قبلی، عملکرد عامل یادگیرنده را بهبود می‌بخشند. در [۲۵]، برای مدیریت انرژی ساختمان از یک روش داده‌محور مبتنی بر شبکه عصبی و یادگیری Q استفاده شده است. یک تکنیک برنامه‌ریزی پویا تطبیقی براساس یادگیری تقویتی برای کنترل یک ریزشبکه هوشمند در [۲٦] توسعه داده شده است. در [۲۷]، مدیریت انرژی چندعامله غیرمتمرکز برای بارهای الکتریکی در یک ریزشبکه با استفاده از یادگیری Q پیاده‌سازی شده است. در [۲۸] یادگیری تقویتی سلسله‌مراتبی برای محاسبه سیاست بهینه توسعه داده شده است. اگرچه روش ارائه‌شده برای حل معضل ابعاد در روش‌های یادگیری تقویتی سودمند است، به یک نوع سیاست بهینه محلی به نام سیاست بهینه بازگشتی همگرا می‌شود [۲۹]. پخش بار اقتصادی توزیع‌شده در [۳۰] با استفاده از الگوریتم یادگیری تقویتی مشارکتی براساس اطلاعات دریافتی از عامل‌های همسایه و با استفاده از استراتژی Diffusion به دست آمده است.

با استفاده از الگوریتم یادگیری Nash Q بار درخواستی واحدهای ریزشبکه تخصیص داده شده و سود هرکدام بیشینه شده است [۳۱]. الگوریتم یادگیری Nash Q یک توسعه از الگوریتم یادگیری Q معمولی برای سیستم‌های چندعامله غیرمشارکتی است [۳۲]. در الگوریتم یادگیری Nash Q، یک عامل نه‌تنها پاداش خود، پاداش و اعمال سایر عامل‌ها را نیز دریافت می‌کند. در واقعیت، موجودبودن اطلاعات مربوط به اعمال و پاداش‌های سایر عامل‌ها برای همه عامل‌های مصرف‌کننده و تولیدکننده یا حتی برای یک سیستم مرکزی نیز به‌راحتی امکان‌پذیر نیست. همچنین، با افزایش تعداد عامل‌های یادگیرنده، سایز تابع Nash Q زیاد می‌شود؛ درنتیجه، زمان اجرا بسیار زیاد و انجام محاسبات پیچیده می‌شود.

روش یادگیری تقویتی عمیق از تلفیق روش‌های یادگیری عمیق با روش یادگیری تقویتی به دست می‌آید. در این روش‌ها، با تخمین توابع ارزش و توابع سیاست با استفاده از شبکه‌های عصبی عمیق مشکل ابعاد توابع Q حل می‌شود. در [۳۳]، یک شبکه عمیق Q برای حل مسائل با تعداد زیاد سنسورهای ورودی توسعه داده شده است. برنامه‌ریزی زمان حقیقی یک ریزشبکه در [۳٤] با استفاده از شبکه‌های عصبی عمیق برای تخمین توابع ارائه شده است. در [35]، برای افزایش انعطاف‌پذیری و قابلیت اطمینان یک ریزشبکه باحضور منابع تجدیدپذیر، از الگوریتم بهینه‌سازی Proximal Policy براساس یادگیری تقویتی عمیق و شبکه‌های عصبی مرکزی critic بهره گرفته شده است. برای کمینه‌کردن هزینه‌های عملیاتی ژنراتورها و کاهش هزینه‌های خرید انرژی از شبکه اصلی در [۳٦]، توان اکتیو و راکتیو ژنراتورهای معمولی و مقدار توان شارژ/تخلیه باتری با استفاده از روش یادگیری تقویتی عمیق محاسبه شده است. در مقاله [۳۷]، با استفاده از یک روش یادگیری تقویتی عمیق مبتنی بر سیاست با فضای تصمیم‌گیری و حالت پیوسته، هزینه‌های عملیاتی ریزشبکه شامل هزینه تبادل انرژی با شبکه توزیع‌شده، هزینه عملیاتی ریزتوربین و هزینه عملیاتی باتری کمینه شده است. به‌تازگی توجهی زیادی به ترکیب روش‌های یادگیری تقویتی و یادگیری عمیق شده است. روش‌های عمیق به کمک روش‌های یادگیری تقویتی می‌آید تا مشکل ابعاد برای محاسبه تابع Q با تعداد عامل‌های زیاد را برطرف کند؛ اما همچنان سایر مشکلات موجود در روش‌های متمرکز در این روش‌ها برطرف نشده است. در روش‌های یادگیری عمیق لازم است اطلاعات مربوط به همه عامل‌ها شامل اعمال و پاداش‌ها برای یک واحد کنترل‌کننده مرکزی موجود باشد. بیشتر روش‌های ارائه‌شده برای مدیریت انرژی ریزشبکه‌ها از یک ساختار کنترل مرکزی استفاده می‌کنند. در کنترل‌کننده‌های متمرکز، یک واحد به‌عنوان کنترل‌کننده انتخاب می‌شود و مسئولیت مدیریت سایر واحدها را برعهده دارد. در ساختار متمرکز، واحد کنترل‌کننده با همه عامل‌ها در ارتباط است؛ در حالی

که در ساختار کنترلی توزیع‌شده، کنترل‌کننده‌ها با همسایگی‌های خود در ارتباط هستند؛ اما در ساختار کنترل‌کننده غیرمتمرکز ارتباطی بین کنترل‌کننده‌ها به‌صورت جداگانه نیست [۳۸, ۳۹]. هر دو ساختار توزیع‌شده و غیرمتمرکز به‌عنوان سیستم کنترلی چندعامله در نظر گرفته می‌شود. در سیستم‌های قدرت ابعاد وسیع، هنگامی که واحدهای تولید برق در شبکه پراکنده شده‌اند و ارتباطات، محدود و غیرقابل اعتماد هستند، استفاده از ساختارهای کنترلی متمرکز با چالش‌های جدی روبه‌رو می‌شود [۴۰]. در ساختارهای کنترلی متمرکز، هر خطایی در کنترل‌کننده باعث خرابی کل سیستم می‌شود. برای کاهش سایز و پیچیدگی کنترل متمرکز از روش‌های کنترل غیرمتمرکز استفاده می‌شود. این ساختار کنترلی، یک معماری کنترلی ساده با توانایی بالا برای تضعیف اغتشاشات را فراهم می‌کند.

در جدول ۱، به‌طور خلاصه روش‌های مدیریت انرژی با استفاده از یادگیری تقویتی مقایسه شده است. در بیشتر مقالات به مدیریت انرژی در زمینه بارهای الکتریکی پرداخته شده و به بارهای گرمایی توجه نشده است. برنامه‌ریزی انرژی در بسیاری از این روش‌ها به‌صورت متمرکز یا توزیع‌شده انجام می‌گیرد. همچنین، همان‌طور که مشاهده می‌شود در بیشتر مقالات روشی برای مدیریت مصرف و بهینه‌سازی سود مصرف‌کنندگان و تولیدکنندگان با ارائه پیشنهاد قیمت به‌صورت هم‌زمان انجام نشده است. در برخی مقالات، طول عمر باتری در نظر گرفته شده است؛ اما در هیچ‌کدام تأثیر مدل باتری بر محاسبه تعداد تعویض باتری لحاظ نشده است.

بنابراین، در این مقاله، با ارائه یک ساختار کامل غیرمتمرکز به طراحی سیستم مدیریت انرژی ریزشبکه‌ها پرداخته می‌شود. عامل‌ها بدون دردسترس بودن اطلاعات عامل‌های همسایه و تنها با دریافت حالات محیط به یادگیری و آموزش می‌پردازند. طراحی سیستم به‌صورت مستقل از مدل و با استفاده از یادگیری تقویتی انجام می‌شود. در این پژوهش، مدیریت انرژی هر دو نوع بار الکتریکی و گرمایی در نظر گرفته می‌شود. فرض می‌شود ریزشبکه شامل منابع انرژی پراکنده گرمایی و الکتریکی، سیستم ذخیره انرژی باتری و بارهای مصرفی الکتریکی و گرمایی است. سیستم به گونه‌ای طراحی می‌شود که سود کلیه منابع افزایش یابد، هزینه مشتریان کمینه شود و از وابستگی ریزشبکه به شبکه اصلی کاسته شود. همچنین، با در نظر گرفتن طول عمر باتری، هزینه ناشی از تخریب باتری کمینه می‌شود.

به‌طور خلاصه نوآوری‌های اصلی مقاله به‌صورت زیر خلاصه شده‌اند:

۱. ارائه یک ساختار غیرمتمرکز برای سیستم مدیریت انرژی چندعامله ریزشبکه‌ها با حضور منابع انرژی پراکنده الکتریکی و گرمایی، سیستم ذخیره انرژی باتری و بارهای مصرفی الکتریکی و گرمایی؛

۲. طراحی یک روش مستقل از مدل مبتنی بر یادگیری تقویتی برای برنامه‌ریزی ساعتی سیستم فوق بدون دردسترس بودن مدل عدم‌قطعیت موجود در عرضه و تقاضا؛

۳. توانایی ارائه قیمت و تصمیم‌گیری بر مقدار توان خروجی (به‌جز منابع انرژی‌تجدیدپذیر) با عامل‌های تولیدکننده برای بهینه‌سازی میزان سود فروش و هزینه‌های عملیاتی؛

٤. مدیریت و کاهش هزینه مصرف‌کنندگان الکتریکی و گرمایی و کاهش وابستگی ریزشبکه به شبکه اصلی؛

۵. در نظر گرفتن مدل طول عمر باتری و کمینه‌کردن هزینه‌های ناشی از تخریب باتری؛

٦. استفاده از داده‌های واقعی منابع انرژی تجدیدپذیر و مصرف‌کنندگان، برای مقایسه و ارزیابی دقت روش پیشنهادی در شبکه برق ایران.

در همین راستا در بخش دوم، ساختار ریزشبکه معرفی می‌شود. در بخش سوم، طراحی سیستم مدیریت انرژی با استفاده از یادگیری تقویتی ارائه می‌شود. در بخش چهارم، شبیه‌سازی و نتایج بررسی می‌شوند. در قسمت آخر نتیجه‌گیری ارائه می‌شود.

جدول (۱): مقایسه تحقیقات انجام‌شده در زمینه مدیریت انرژی ریزشبکه‌ها

مقاله	بار الکتریکی	بار گرمایی	ساختار غیر متمرکز	طول عمر باتری	مدیریت بار	ارائه پیشنهاد قیمت	مدیریت تولیدکنندگان (هزینه تولید و سود فروش)	مقایسه
[۲۵]	✓	×	✓	×	✓	×	×	✓
[۲۶]	✓	×	×	✓	✓	×	×	✓
[۲۷]	✓	×	✓	×	✓	✓	✓	×
[۲۸]	✓	×	✓	×	✓	×	× فقط هزینه تولید	✓
[۳۰]	✓	×	×	✓	×	×	× فقط هزینه تولید	✓
[۳۴]	✓	×	×	×	×	×	× فقط هزینه تولید	✓
[۳۱]	✓	✓	×	×	×	×	×	×
[۱۶]	✓	×	✓	×	✓	×	×	×
[۳۵]	✓	×	×	×	✓	×	×	✓
[۳۶]	✓	×	×	×	×	×	× فقط هزینه تولید	✓
روش پیشنهادی	✓	✓	✓	✓	✓	✓	✓	✓

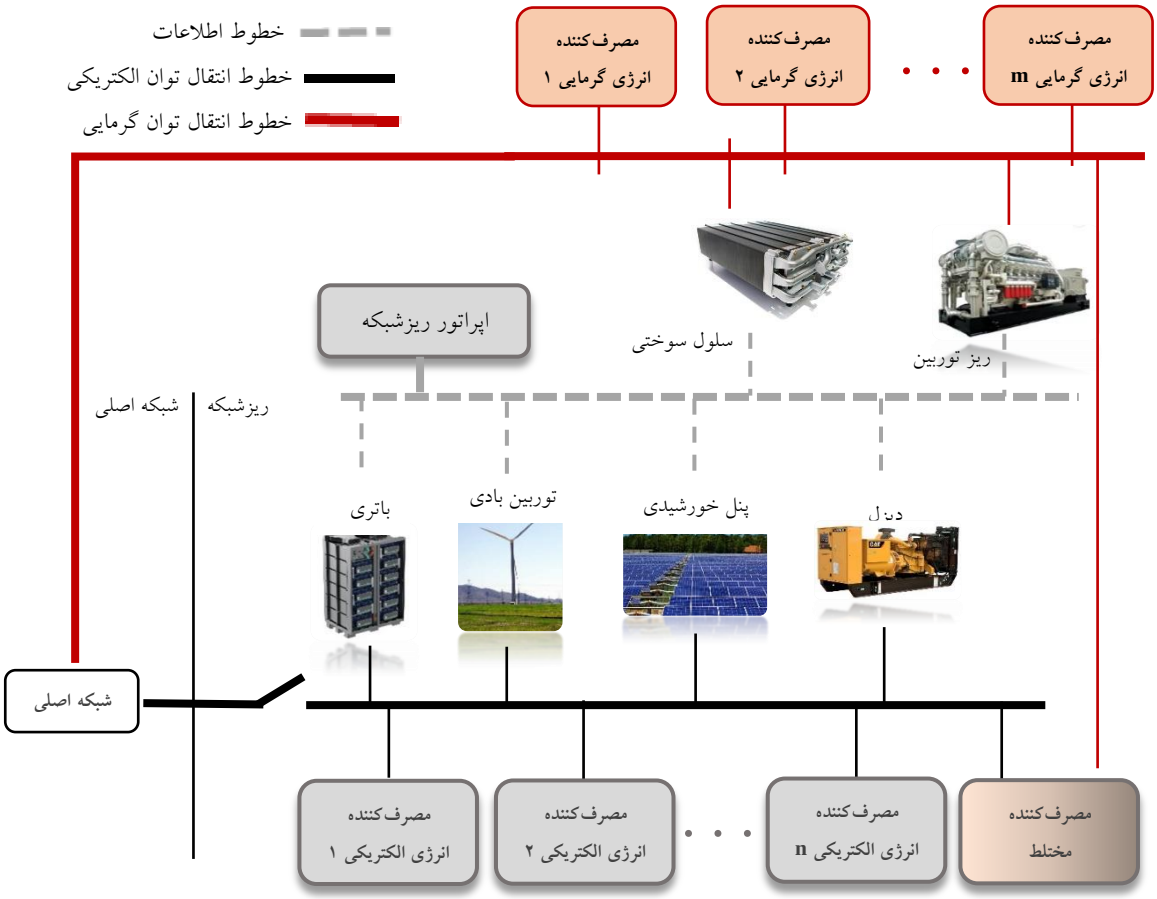
۲- ساختار ریزشبکه

۲-۱- ریزشبکه

به یک شبکه قدرت مقیاس‌کوچک، ولتاژ پایین و خودکار که منابع انرژی پراکنده و بارها را به هم متصل می‌کند، ریزشبکه گفته می‌شود. منابع انرژی پراکنده شامل انرژی‌های تجدیدپذیر، انرژی‌های تجدیدناپذیر و باتری است. ریزشبکه‌ها در دو حالت متصل به شبکه و جزیره‌ای کار می‌کنند [۳]. در حالت کلی فرض می‌شود ریزشبکه‌ها به شبکه اصلی متصل‌اند [۴۱, ۴۲]. ریزشبکه‌ها ازطریق نقطه اتصال مشترک (Point of Common Coupling) به شبکه اصلی متصل می‌شوند. در مد متصل به شبکه، ریزشبکه‌ها تعادل تقاضا و تأمین را با فروش انرژی اضافی به شبکه اصلی و خریدن کسری انرژی از آن تأمین می‌کنند

یکی از اهداف مهم در مسئله مدیریت انرژی ریزشبکه‌ها، کاهش وابستگی آنها به شبکه اصلی است؛ بنابراین، سیستم مدیریت انرژی ریزشبکه‌ها باید به گونه‌ای طراحی شود که علاوه بر افزایش سود عامل‌ها، وابستگی ریزشبکه به شبکه اصلی نیز کاهش یابد.

بارها در ریزشبکه‌ها به دو دسته کنترل‌پذیر و غیرکنترل‌پذیر تقسیم می‌شوند. بارهای غیرکنترل‌پذیر مانند سیستم‌های مراکز درمانی و وظایف ضروری در صنعت هستند که می‌باید در زمان تقاضا تأمین شوند. درواقع، بارهای غیرکنترل‌پذیر انعطاف‌ناپذیر نسبت به زمان بوده‌اند و نمی‌توان آنها را در طول زمان جابه‌جا کرد؛ اما بارهای کنترل‌پذیر دارای قابلیت حذف یا انتقال به زمان‌های کم‌باری هستند؛ حتی می‌توان برخی از بارهای کنترل‌پذیر را کاهش داد. شکل ۱ ساختار یک ریزشبکه متشکل از پنل‌های خورشیدی، توربین بادی، دیزل ژنراتور، سلول سوختی الکتریکی و گرمایی، میکروتوربین گرمایی و الکتریکی، باتری و تعدادی بار محلی الکتریکی و گرمایی را نشان می‌دهد. اپراتور ریزشبکه (Microgrid Operator) یک عامل کنترل‌کننده سطح بالا در ریزشبکه‌های قدرت در نظر گرفته می‌شود.



شکل (۱): ساختار ریزشبکه

۲-۲- توابع هدف:

هدف سیستم مدیریت انرژی برای یک ریزشبکه، بیشینه‌کردن سود کلیه عامل‌های داخل شبکه در یک مدت زمان طولانی است [۸, ۲۷]؛ به همین دلیل، سود کلی عامل تولیدکننده iام برای مدت زمان طولانی به‌صورت زیر تعریف می‌شود:

$$max\ F_i = \sum_{t=1}^{\infty} \gamma^t * [Pr_i(t) \times P_i^{mic}(t) + S_p(t) \times P_i^{main}(t) - C_i^{op}(P_i^{mic}(t) + P_i^{main}(t))] \quad (۱)$$

به طوری که t بازه زمانی است که در آن رابطه فوق محاسبه شده است. پارامتر γ نرخ تخفیف گفته می‌شود و مقداری بین صفر و یک دارد. این پارامتر بیان‌کنندهٔ ارزش فعلی پاداش‌های آتی است. زمانی که γ به یک نزدیک می‌شود، به تولیدکنندهٔ سودهای آینده به‌شدت توجه می‌کند.

$P_i^{mic}(t)$ و $P_i^{main}(t)$ به‌ترتیب توان فروخته‌شده از ژنراتور iام به ریزشبکه و شبکه اصلی در بازه زمانی t است. $Pr_i(t)$ قیمت پیشنهادی فروش انرژی از ژنراتور iام به ریزشبکه است. $S_p(t)$ حالت محیط است که قیمت خرید انرژی توسط شبکه اصلی از ریزشبکه است. $C_i^{op}(\cdot)$ تابع هزینه‌های عملیاتی ژنراتور iام است. تابع هزینه به‌صورت عملی محاسبه می‌شود [۴۳].

تابع هدف سیستم‌های ذخیره انرژی باتری در ادامه آورده شده است:

$$max\ F_b = \sum_{t=1}^{\infty} \gamma^t * [(Pr_b(t) \times P_b^{mic}(t) + S_p(t) \times P_b^{main}(t)) - Pr_m(t) \times P_b^{input}(t) - Ex(t)] \quad (۲)$$

به طوری که‌ترم اول سود حاصل از فروش انرژی، ترم

دوم هزینه ناشی از خرید انرژی و ترم سوم هزینه ناشی از طول عمرباتری است. در هر بازه زمانی، باتری می‌تواند خریدار یا فروشنده انرژی باشد. $Pr_b(t)$ قیمت پیشنهادی فروش انرژی از باتری است. $P_b^{mic}(t)$ و $P_b^{main}(t)$ به‌ترتیب توان فروخته‌شده از باتری به ریزشبکه و شبکه اصلی در بازه زمانی t است P_b^{input} و Pr_m مقدار توان خریداری‌شده از باتری و قیمت بازار برق است. سیاست سیستم‌های ذخیره انرژی باید به گونه‌ای باشد که انرژی را در زمان‌های با قیمت پایین خریداری و در پیک مصرف به شبکه بازگرداند. $Ex(t)$ هزینه ناشی از کاهش طول عمر و تخریب باتری است که در اثر شارژ و تخلیه باتری رخ می‌دهد [۳۰]. اهداف عامل‌های مصرفی کمینه‌کردن هزینه‌ها است که به‌صورت زیر محاسبه می‌شود:

$$min\ F_c = \sum_{t=1}^{\infty} \gamma^t [Pr_m(t) \times \left(L_i^{NC}(t) + \beta(t) \times L_i^C(t)\right) + \mu \times (1 - \beta(t)) \times L_i^C(t)] \quad (۳)$$

$L_i^{NC}(t)$ و $L_i^C(t)$ به‌ترتیب بارهای غیرکنترل‌پذیر و بارهای کنترل‌پذیر در بازه زمانی t است. $\beta(t)$ نسبت بار خاموش‌شده به بارهای کنترل‌پذیر است. μ ضریب میزان نارضایتی در مصرف‌کنندگان به‌ازای حذف بار است. مقدار آن بستگی به نوع مصرف‌کننده و میزان اشتیاق آنها در مدیریت و بهینه‌سازی مصرف و هزینه‌های خود دارد.

۲-۳- قیود مسئله

برای تضمین قابلیت اطمینان و امنیت یک شبکه می‌باید توان مورد تقاضا توسط تولیدکننده‌ها در تمامی زمان‌ها تأمین شود. به‌منظور دستیابی به تعادل توان الکتریکی و تنظیم فرکانس شبکه به‌طور پیوسته از روش‌های مدیریت ذخیره استفاده می‌شود. در حالتی که ریزشبکه متصل به شبکه اصلی باشد، تأمین فرکانس و ذخیره عملیاتی توسط شبکه اصلی به‌عنوان یک ژنراتور اصلی تأمین می‌شود. برای حل مسئله مدیریت ذخیره در حالت جزیره‌ای، روش‌های کنترل مشارکتی برای تولیدکننده‌های پراکنده توسعه داده شده است [۴۴-۴۶]. در حالت متصل به شبکه اصلی، قید تعادل توان به معنای برابری میزان توان تولیدی با میزان بارهای مصرفی است [۳۰]؛ بنابراین، برای بارهای الکتریکی قید تعادل توان به‌صورت زیر تعریف می‌شود:

$$\sum_{i=1}^{n} Load_i^E = P_w + P_{PV} + P_d + P_b + P_{MT}^E + P_{FC}^E + P_{main}^E \quad (۴)$$

$Load_i^E$ میزان تقاضای بار الکتریکی عامل مصرفی iام و n تعداد عامل‌های مصرفی الکتریکی است. P_w، P_{PV}، P_d، P_b، P_{MT}^H، P_{FC}^H و P_{main}^E به‌ترتیب توان تولیدی الکتریکی توربین بادی، پنل‌های خورشیدی، دیزل ژنراتور، باتری، ریزتوربین، سلول سوختی و شبکه اصلی است. قید تعادل توان برای بارهای گرمایی نیز به‌صورت زیر تعریف می‌شود:

$$\sum_{i=1}^{m} Load_i^H = P_{MT}^H + P_{FC}^H + P_{main}^H \quad (۵)$$

$Load_i^H$ میزان تقاضای بار گرمایی عامل مصرفی iام و m تعداد عامل‌های مصرفی گرمایی است. P_{MT}^H، P_{FC}^H و P_{main}^H به‌ترتیب توان تولیدی گرمایی ریز توربین، سلول سوختی و شبکه اصلی است.

قیود ظرفیت بیان‌کنندهٔ بازه عملیاتی ژنراتورهای پراکنده است و دارای محدوده زیر است [۳۱]:

$$P_i^{min} \leq P_i(t) \leq P_i^{max} \quad (۶)$$

به طوری که توان خروجی ژنراتور پراکنده i در بازه زمانی t با $P_i(t)$ بیان می‌شود. P_i^{min} و P_i^{max} به‌ترتیب کمینه و بیشینه توان خروجی ژنراتور i است.

SOC بیان‌کنندهٔ سطح شارژ باتری نسبت به ظرفیت آن است. قید فنی زیر به‌منظور جلوگیری از شارژ و تخلیه بیش از حد سیستم ذخیره انرژی باتری اعمال می‌شود [۳۰]:

$$SOC_{min} \leq SOC(t) \leq SOC_{max} \quad (۷)$$

به طوری که SOC_{min} و SOC_{max} کمینه و بیشینه حالت شارژ باتری است. در این پژوهش SOC در بازه $[0.2,\ 0.8]$ محدود شده است تا از آسیب به باتری جلوگیری شود.

۳- طراحی سیستم مدیریت انرژی ریزشبکه‌ها با استفاده از یادگیری تقویتی

۳-۱- یادگیری تقویتی

اصلاح فعالیت‌ها براساس عمل و عکس‌العمل در تعامل با محیط، یادگیری تقویتی نامیده می‌شود. به یادگیری تقویتی، یادگیری براساس اعمال نیز گفته می‌شود. در این روش، یک عامل با اثرگذاشتن بر محیط و گرفتن پاسخ از آن، تلاش می‌کند اعمال و سیاست‌های کنترلی خود را بهبود ببخشد؛ درنتیجه، می‌تواند بازخورد بهتری از محیط دریافت کند. در یادگیری تقویتی، عامل با دریافت پاداش و عدم پاداش یا تنبیه در تعامل با محیط، سعی در اصلاح اعمال خود دارد.

الگوریتم‌های یادگیری تقویتی براساس این ایده ساخته شده‌اند که تصمیمات کنترلی درست می‌بایست در حافظه سیستم توسط سیگنال تقویتی باقی بماند بطوری که در دفعات بعدی احتمال استفاده از آنها بیشتر باشد.

یادگیری تقویتی از ساختار رسمی پروسه‌های تصمیم‌گیری مارکوف استفاده می‌کند و ارتباط بین یک عامل یادگیرنده و محیط را با استفاده از حالات، اعمال و پاداش توصیف می‌کند [۴۷]. در هر بازه زمانی t، عامل یادگیری تقویتی قادر است تا حالات محیط، S_t، را مشاهده نماید و براساس حالات مشاهده شده اعمال، A_t، را انجام دهد. در یک بازه زمانی بعد، به‌عنوان نتیجه عمل خود، عامل یک پاداش عددی، R_{t+1}، را دریافت می‌کند و به حالت جدید، S_{t+1}، می‌رود. بنابراین با استفاده عمل و عکس العمل با محیط، یک عامل یاد می‌گیرد اعمالی را انتخاب کند که پاداش خود را بیشینه کند. پاداش یک عدد است که با استفاده از تابع پاداش محاسبه می‌شود و مطابق با هدف مسئله یادگیری تقویتی تعریف می‌شود. هدف عامل هوشمند، بیشینه‌کردن کلیه پاداش‌های دریافتی در یک زمان طولانی است؛ حتی اگر پاداش لحظه‌ای منفی باشد؛ مانند رفتار بشر، عامل یادگیری تقویتی ممکن است پاداش لحظه‌ای را فدا کند تا در درازمدت پاداش تجمعی بیشتری به دست آورد.

در پروسه‌های تصمیم‌گیری مارکوف، احتمال هر ارزش ممکن برای حالت S_t و پاداش R_t، فقط به حالت و عمل لحظه قبل، S_{t-1} و R_{t-1}، بستگی دارد. به عبارت دیگر، به حالات و اعمال اولیه وابسته نیست. حالت سیستم باید شامل کلیه اطلاعات مربوط به همه جنبه‌های عمل و عکس‌العمل محیط و عامل در گذشته باشد؛ بنابراین، یک حالت با این مشخصات دارای خاصیت مارکوف است.

یک پروسه تصمیم‌گیری مارکوف محدود شامل مجموعه اعمال، حالات و پاداش $(S,\ A,\ R)$ با تعداد المان‌های محدود است. به‌منظور مدل‌سازی مسئله یادگیری تقویتی، دینامیک پروسه‌های تصمیم‌گیری با استفاده از تابع توزیع احتمال شرطی p به‌صورت زیر مشخص می‌شود [۴۷]:

$$p(s', r|s, a) \doteq Pr[S_t = s', R_t = r|S_{t-1} = s, \quad p : S \times R \times S \times A \to [0,1] \quad (۸)$$

برای همه $s', s \in S, r \in R,\ and\ a \in A(s)$. تابع شرطی $p(s', r|s, a)$ احتمال ارزش متغیرهای تصادفی R_t و S_t در لحظه t است که فقط به حالت و عمل لحظه قبل وابسته است. به عبارت دیگر، زمانی که حالت و عمل قبلی داده می‌شود، مدل، حالت و پاداش بعدی را محاسبه می‌کند.

ارزش مورد انتظار جمع وزن دار پاداش‌ها در حالت s با سیاست π، تابع ارزش گفته می‌شود و به‌صورت زیر تعریف می‌شود [۴۷]:

$$v_\pi(s) \doteq E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} |S_t = s\right], \text{for all } s \in S \quad (۹)$$

پارامتر γ، نرخ تخفیف است. پروسه‌های تصمیم‌گیری مارکوف باید بین پاداش‌های لحظه‌ای و آینده مبادله کند. نرخ تخفیف ارزش فعلی پاداش‌های آتی را مشخص می‌کند. این پارامتر مشخص می‌کند عامل یادگیری تقویتی چقدر به پاداش‌های آینده دور نسبت به پاداش‌های لحظه‌ای اهمیت دهد و زمانی که پارامتر γ به یک نزدیک می‌شود، عامل یادگیرنده به پاداش‌های آینده به‌شدت توجه می‌کند.

سیاست هر عامل مشخص می‌کند چگونه یک عامل در یک حالت مشخص عمل کند. درواقع یک نگاشت از حالات مشاهده‌شده به اعمال انجام‌شده در آن حالات است.

تابع ارزش – عمل، $Q_\pi(s, a)$، بیان‌کنندهٔ ارزش عمل

انجام‌شده در حالت s تحت سیاست π است و به‌صورت زیر بیان می‌شود [٤٧]:

$$Q_\pi(s,a) \doteq E_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1}\,|S_t=s,A_t=a\right] \quad (١٠)$$

به طوری که $Q_\pi(s,a)$ ارزش مورد انتظار جمع پاداش‌های وزن‌دار (با ضریب تخفیف) در حالت s، عمل a و تحت سیاست π است.

برای حل مسئله یادگیری تقویتی، عامل یادگیرنده باید سیاستی را محاسبه کند که مقدار زیادی پاداش در طول زمان زیادی به دست آورد. یک سیاست، سیاست بهینه گفته می‌شود، اگر جمع مورد انتظار پاداش‌ها بزرگ‌تر یا مساوی سایر سیاست‌ها باشد. سیاست‌های بهینه، تابع ارزش - عمل بهینه یکسانی را به اشتراک می‌گذارند و به‌صورت Q_* نمایش داده می‌شود. اگر یک عامل، تابع Q_* را برای هر حالت s داشته باشد، می‌تواند به‌راحتی هر عملی را که $Q_*(s,a)$ را بیشینه کند، پیدا کند.

۳-۲- روش مستقل از مدل یادگیری Q:

حتی اگر یک مدل دقیق و کامل از دینامیک محیط موجود باشد، محاسبه سیاست بهینه با حل معادله بهینگی بلمن به‌راحتی امکان‌پذیر نیست. برای محاسبه سیاست بهینه از قانون بروزرسانی یادگیری Q استفاده می‌شود [٤٨، ٤٩]. یادگیری Q اساساً یک پروسه تصمیم‌گیری مارکوف است. فرض کنید از حالت S با اعمال A به حالت بعدی S انتقال داده شود و پاداش R دریافت شود؛ بنابراین، قانون به‌روزرسانی یادگیری Q به‌صورت زیر تعریف می‌شود [٤٨]:

$$\begin{aligned} &Q^{k+1}(s_t,a_t) \\ &= (1-\alpha)Q^k(s_t,a_t)+\alpha[R_{t+1} \\ &+ \quad \gamma\max_a Q^k(s_{t+1},a)] \end{aligned} \quad (١١)$$

به طوری که $\gamma\in[0,1)$ نرخ تخفیف و $\alpha\in[0,1]$ نرخ یادگیری است. تابع بهینه Q با استفاده از تابع Q یاد گرفته شده در رابطه ١١ به‌صورت مستقیم تخمین زده می‌شود. یادگیری Q یک روش یادگیری تقویتی مستقل از مدل است. در صورتی که همه جفت حالت - عمل به‌صورت پیوسته به‌روزرسانی و گسسته‌سازی شود، با احتمال یک به مقدار ارزش - عمل بهینه همگرا می‌شود. هر عامل در تمامی حالات می‌باید به‌صورت تکرارپذیری آزمایش شود تا یک تخمین معتبر از پاداش مورد انتظار به دست آید.

۳-۳- مدیریت انرژی ریزشبکهٔ پیشنهادی

منابع انرژی پراکنده و مشترکین برق، عامل‌های مستقل و هوشمند در نظر گرفته می‌شوند. عامل‌ها دارای توانایی یادگیری هستند و می‌توانند با انتخاب تصمیمات درست، سود خود را بیشینه کنند. عامل‌های یادگیری تقویتی با استفاده از بازخوردهای اعمال و تجربیات خود، سیاست بهینه را کشف می‌کنند. با توجه به خاصیت متغیر با زمان خروجی منابع انرژی تجدیدپذیر و تصادفی‌بودن مقدار بار مصرفی، از پروسه‌های تصمیم‌گیری مارکوف برای مدل‌سازی رفتار تصادفی عامل‌ها در ریزشبکه استفاده شده و سیاست بهینه عامل‌ها با الگوریتم مستقل از مدل یادگیری Q به دست آمده است. فرض شده است شبکه در مد متصل به شبکه اصلی کار می‌کند.

هدف مسئله یادگیری تقویتی، بهینه‌کردن توابع هدف (١)- (٣) است؛ به طوری که قیود (٤)-(٧) نیز برآورده شود. حالات، اعمال و پاداش در ادامه به‌تفصیل توضیح داده می‌شوند.

حالات: برای کلیه عامل‌ها به‌جز باتری، حالات شامل (t,S_s,S_p) است. t بازه زمانی، S_s قیمت برق به هنگام فروش برق به شبکه اصلی و S_p قیمت خرید برق از شبکه اصلی است. عامل باتری علاوه بر حالات فوق دارای یک حالت اضافی شامل سطح شارژ باتری است که مقدار آن بین ٠ تا ١٠٠ درصد تغییر می‌کند.

اعمال: مجموعه اعمال برای ریزتوربین و سلول سوختی شامل مقدار توان الکتریکی تولیدشده، مقدار توان گرمایی تولیدشده، قیمت پیشنهادی فروش انرژی الکتریکی و گرمایی به ریزشبکه است. دیزل ژنراتور نیز دارای قدرت تصمیم‌گیری بر میزان توان تولیدی الکتریکی و قیمت

پیشنهادی فروش انرژی است. مجموعه اعمال منابع انرژی‌های انرژی تجدیدپذیر شامل توربین بادی و پنل خورشیدی فقط شامل قیمت پیشنهادی فروش انرژی است؛ زیرا توان خروجی منابع انرژی‌های تجدیدپذیر کاملاً وابسته به شرایط آب‌وهوایی است و متغیرهای تصادفی هستند. برای بیشینه‌کردن سود منابع تجدیدپذیر فرض می‌شود آنها در حالت ردیابی نقطه حداکثر توان هستند. در صورتی که قیمت پیشنهادی منابع انرژی تجدیدپذیر با قیمت پیشنهادی منابع غیرتجدیدپذیر برابر باشد، اولویت فروش با منابع انرژی تجدیدپذیر است. ریزشبکه ابتدا توان تولیدشده با توربین بادی و پنل خورشیدی را خریداری می‌کند و در صورت نیاز از انرژی سایر منابع استفاده می‌کند. اگر مقدار توان تولیدی منابع انرژی بیش از نیاز ریزشبکه باشد، عامل‌ها می‌توانند انرژی مازاد را مستقیماً به شبکه اصلی بفروشند؛ اما با توجه به اینکه میزان قیمت خریداری‌شده توسط شبکه اصلی بسیار کمتر از قیمت فروش انرژی است، کلیه عامل‌های تولیدی می‌باید به نحوی آموزش ببینند که با ارائه قیمت صحیح در بازار رقابتی برق، بتوانند توان تولیدشده خود را در داخل ریزشبکه بفروشند. درواقع، این موضوع کمک می‌کند تا ریزشبکه‌ها توان مورد نیاز خود را به جای خرید از شبکه اصلی از تولیدکننده‌های داخلی خریداری کنند؛ درنتیجه، ضمن افزایش سود تولیدکنندگان داخلی، از وابستگی ریزشبکه به شبکه اصلی نیز کاسته می‌شود. مجموعه اعمال باتری شامل حالت شارژ یا تخلیه، میزان توان مبادله‌شده و قیمت پیشنهادی است. در حالت شارژ توان باتری، منفی و در حالت تخلیه، مثبت است. میزان تقاضا فرض می‌شود یک متغیر تصادفی با تابع توزیع نمایی است. می‌توان تقاضا را به دسته بارهای غیرکنترل‌پذیر و کنترل‌پذیر تقسیم کرد. بروی دسته اول کنترلی نیست و در زمان تقاضا می‌باید برآورده شوند؛ اما میزان حذف بارهای کنترل‌پذیر با توجه به علاقه‌مندی آنها برای شرکت در مدیریت هزینه‌ها و الگوی مصرف قابل کنترل بوده و جزء مجموعه اعمال عامل‌های مصرف‌کننده است.

پاداش‌ها: با توجه به اینکه هدف مسئله یادگیری تقویتی بیشینه‌کردن توابع هدف (۱)–(۳) است، پاداش لحظه‌ای به گونه‌ای تعریف می‌شود که توابع فوق را بیشینه کند؛ به همین دلیل، برای منابع انرژی پراکنده پاداش میزان سود لحظه‌ای حاصل از فروش انرژی است. پاداش مصرف‌کنندگان منفی صورتحساب برق مصرفی است. پاداش لحظه‌ای باتری در حالت شارژ منفی می‌شود و در این حالت باتری ممکن است برای جلوگیری از دریافت پاداش منفی هیچ‌گونه فعالیتی را انجام ندهد و پاداش دریافتی‌اش صفر شود. به‌منظور جلوگیری از تنبل‌شدن باتری، توابع پاداش لحظه‌ای باتری در حالت شارژ و تخلیه به‌صورت زیر تعریف و از یک ضریب اصلاح $\propto$ نیز استفاده می‌شود:

$$\begin{cases} k^{soh}\dfrac{Pr_b(t)P_b^{mic}(t)+S_p(t)P_b^{main}(t)}{profit_b^{max}}, Discharge \\ k^{soh}\propto\left(1+\dfrac{Pr_m(t)P_b(t)}{profit_b^{max}}\right) \quad , Charge \\ profit_b^{max}=Pr_b^{max}*P_b^{max} \end{cases} \quad (۱۲)$$

ضریب $\propto$ مقداری بین صفر و یک دارد. به نحوی تنظیم می‌شود که سود باتری بیشینه شود. اگر یک باتری به‌خوبی آموزش ببیند، سود آن مقداری مثبت است. اگر سود باتری منفی شود، درواقع باتری انرژی را با قیمت بالا خریداری و با قیمت پایین فروخته است؛ بنابراین، باید آموزش باتری به نحوی صورت پذیرد که درنهایت سود باتری مثبت شود و میزان پاداش حاصل از فروش انرژی از هزینه خرید انرژی کمتر نشود. در هر لحظه باتری می‌تواند مصرف‌کننده یا تولیدکننده باشد و در آن واحد نمی‌تواند دارای هر دو حالت باشد. k^{soh} بیان‌کنندهٔ هزینه‌های ناشی از کاهش طول عمر باتری است. برای محاسبه k^{soh}، ابتدا هزینه تخریب و کاهش طول عمر باتری به‌صورت زیر محاسبه می‌شود [۳۰]:

$$Ex(t)=\rho_b*\big(SOH(t+1)-SOH(t)\big) \quad (۱۳)$$

$\rho_b(t)$ ضریب تخریب و متناسب با هزینه باتری است. SOH (State Of Health) حالت سلامت باتری است و مقداری بین صفر و یک دارد [۳۰]:

$$SOH(t+1)=SOH(t)+(1-k^{soh})SOH(t) \quad (۱۴)$$

$1-k^{soh}$ برای دو حالت شارژ و تخلیه باتری به‌صورت زیر محاسبه می‌شود [۵۰, ۵۱]:

$$\begin{cases} a^h * [(-c_b * P_b(t))^{\beta^h} + \eta^h]^{-1}. \text{ charg} \\ a^h * [(-d_b \cdot (P_b(t)/SOH(t))^{\alpha^h})^{\beta^h} + \eta^h]^{-1}. \text{ dis} \end{cases} \quad (۱۵)$$

به طوری که پارامترهای ثابت a^h، β^h و η^h وابسته به مشخصه تخریب باتری هستند. ضریب c_b براساس مشخصه شارژ باتری محاسبه می‌شود. همه پارامترهای فوق با استفاده از تست‌های عملی به دست می‌آیند. P_b مقدار توان شارژ و تخلیه باتری است. سایر پارامترهای α^h و d_b نیز با استفاده از تست‌های عملی محاسبه می‌شود [۳۰].

الگوریتم مدیریت انرژی ریزشبکه شامل دو مرحله است. در مرحله اول، عامل‌ها تابع Q را یاد می‌گیرند. توان خروجی توربین بادی و پنل خورشیدی با استفاده از توابع توزیعWeibull و Beta به‌ترتیب مدل می‌شوند [۵۲, ۵۳]. پارامترهای توزیع Weibull و Beta از داده‌های خروجی منابع انرژی تجدیدپذیر شبکه قدرت تخمین زده می‌شود. از مدل خروجی توربین بادی و پنل خورشیدی صرفاً برای تولید داده عملی بیشتر برای استفاده در مراحل یادگیری و ارزیابی الگوریتم استفاده می‌شود. در مرحله دوم توابع Q تخمین زده شده ارزیابی می‌شوند. اپراتور شبکه، یک عامل حراج‌کننده در نظر گرفته می‌شود که قیمت بازار و میزان توان خریداری‌شده از ریزشبکه را مشخص می‌کند. به‌طور خلاصه، روش مدیریت انرژی ریزشبکه در الگوریتم ۱ آورده شده است.

٤- شبیه‌سازی

در این بخش، سیستم مدیریت انرژی پیشنهادی برای ریزشبکه هوشمند با استفاده از الگوریتم یادگیری تقویتی و داده‌های خروجی انرژی‌های تجدیدپذیر و داده‌های دریافتی از بازار برق ایران شبیه‌سازی شده است. داده‌های واقعی شامل توان خروجی توربین بادی و پنل خورشیدی با همکاری پژوهشگاه هوا خورشید دانشگاه فردوسی مشهد و شرکت برق منطقه‌ای در بهار و تابستان ۱۳۹۹ به‌صورت ساعتی جمع‌آوری شده‌اند. در جدول ۲، پارامترهای مجموعه داده‌ها برای ٢٤ ساعت شبانه روز ارائه شده‌اند. مقدار خروجی در جدول ۲ نرمالیزه شده است. ریزشبکه پیشنهادی مطابق شکل ۱ متشکل از منابع انرژی گرمایی و الکتریکی، باتری و بارهای مصرفی الکتریکی و گرمایی است. مشخصات منابع پراکنده مطابق جدول ۳ است. ٤ عامل مصرف‌کننده بار الکتریکی، ۳ عامل مصرف‌کننده بار گرمایی و یک عامل مصرف‌کننده بار الکتریکی و گرمایی به‌ترتیب به ظرفیت ۸، ٤ و ۸ کیلووات در ریزشبکه در نظر گرفته شده‌اند. ظرفیت دیزل ژنراتور به نسبت ظرفیت منابع تجدیدپذیر کمتر در نظر گرفته شده است تا استفاده از منابع غیرتجدیدپذیر در شبکه‌های قدرت محدودتر شود. مجموع میزان توان تولیدی از مجموع میزان توان مصرفی نیز به دلیل مدیریت مصرف کمتر در نظر گرفته شده است.

الگوریتم ۱:

// مقداردهی اولیه

مقداردهی اولیه پارامترهای یادگیری و K_1

مقداردهی $Q_0=0$

مقداردهی k_1

// یادگیری

برای $k=1:k_1$

هر عامل حالت‌های محیط را مشاهده می‌کند.

میزان تقاضا و خروجی منابع تجدیدپذیر با استفاده توابع توزیع احتمال انتخاب می‌شود.

هر عامل یک عمل را به‌صورت تصادفی انتخاب می‌کند.

هر عامل پاداش خود را مشاهده می‌کند.

جدول Q هر عامل به‌روزرسانی می‌شود.

پایان

// ارزیابی

مقداردهی اولیه k_2

برای $k=1:k_2$

هر عامل حالات محیط را مشاهده می‌کند.

مقدار تقاضا و خروجی منابع انرژی تجدیدپذیر مشخص می‌شود.

هر عامل بهترین عمل را از جدول Q خود انتخاب می‌کند.

هر عامل میزان پاداش خود را مشاهده می‌کند.

میزان سود هر عامل، محاسبه و عملکرد سیستم ارزیابی می‌شود.

پایان

مصرف‌کنندگان قادرند به میزان حداکثر ۷۰ درصد مصرف خود را مدیریت کنند. ۳۰ درصد باقیمانده به‌عنوان بار ضروری در نظر گرفته می‌شود که در زمان تقاضا حتماً تأمین می‌شوند.

یک روز به ۲٤ بازه زمانی یک‌ساعته تقسیم شده است. در هر بازه، نرخ خرید و فروش از شبکه اصلی در بازه ۱۵۰-۱۲۰۰ ریال بر کیلو وات ساعت قرار دارد. با توجه به میزان خرید و فروش انرژی در بازار برق ایران در سایت IREMA [٤٥] میزان قیمت پیشنهادی توسط تولیدکنندگان محدوده بین ۲۰۰-۱۳۰۰ ریال بر کیلووات ساعت تعیین شده است.

روش ارائه‌شده تحت چهار سناریو: بدون یادگیری - یادگیری تولیدکنندگان - یادگیری مصرف‌کنندگان - یادگیری همه عامل‌ها ارزیابی شده است. عملکرد هر عامل طی چهار سناریو، هر سناریو به‌مدت ۸۰ روز و جمعاً به‌مدت ۳۲۰ روز شبیه‌سازی شده است. در ۸۰ روز اول هیچ‌گونه یادگیری وجود ندارد و کلیه بارهای درخواستی در همان زمان برآورده می‌شود و منابع انرژی پراکنده به‌صورت تصادفی یک عمل را انتخاب می‌کنند. در ۸۰ روز دوم، تنها منابع پراکنده آموزش دیده‌اند و دارای توانایی تصمیم‌گیری هوشمند هستند. در ۸۰ روز سوم، تنها عامل‌های مصرفی توانایی یادگیری دارند و در ۸۰ روز آخر کلیه عامل‌ها آموزش دیده‌اند. فاز آموزش به‌مدت ۱۰،۰۰۰ روز اجرا شده و فاز ارزیابی برای هر سناریو به مدت ۱۰ روز و جمعاً به‌مدت ۸۰۰ روز شبیه‌سازی شده است. میانگین نتایج ارزیابی الگوریتم یادگیری تقویتی بروی سیستم مدیریت انرژی فوق در شکل‌های ۲ الی ۱۰ نمایش داده شده است. در این قسمت، مدل تخریب باتری در نظر گرفته نشده است. میانگین مقدار سود و توان کلیه عامل‌ها برای چهار سناریوها در جدول ٤ نمایش داده شده است. در جدول ٤، هزینه برای یک عامل مصرف‌کننده و توان شامل کل بار درخواستی در ریزشبکه است.

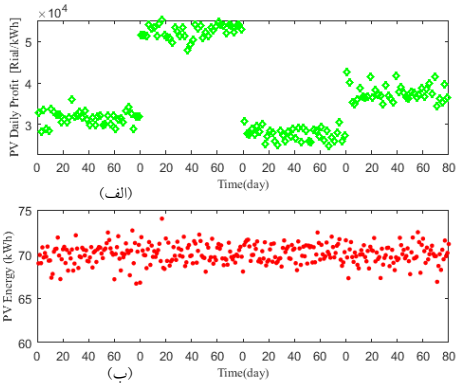
با توجه به شکل ۲ و۳، اگرچه میانگین خروجی توربین بادی و پنل خورشیدی در سناریو دوم (۸۰ روز دوم) و در سناریو چهارم (۸۰ روز چهارم) تقریباً تغییری نداشته، سود آنها به‌طور درخور توجهی افزایش یافته است؛ زیرا در این سناریوها منابع تولیدکننده دارای توانایی تصمیم‌گیری هستند و می‌توانند تصمیمات هوشمندانه‌تری اتخاذ کنند. در شکل ٤ الی ٦، میانگین روزانه سود و توان دیزل ژنراتور، سلول سوختی و ریزتوربین نشان داده شده است. با توجه به آموزش عامل‌های تولیدکننده در سناریو دوم و چهارم، سود عامل‌های دیزل ژنراتور، سلول سوختی و ریزتوربین در این سناریوها نیز افزایش یافته است. نسبت سود به تولید در دیزل ژنراتور در سناریو اول و چهارم به‌ترتیب ۲۳۹ و ۲٥٤ است؛ بنابراین، اگرچه تولید در سناریو چهارم افزایش پیدا کرده، نسبت سود به تولید (طبق جدول ٤) برای دیزل‌ژنراتور نیز افزایش یافته است. درواقع دیزل ژنراتور توانسته است به‌طور هوشمندانه تولید خود را به ساعتی که درخواست و هزینه خرید بالا هست، منتقل کند. همچنین، این عامل با ارائه قیمت معقول برای پیشنهاد فروش انرژی توانسته است انرژی بیشتری در داخل ریزشبکه به فروش برساند و سود خود را افزایش دهد. نسبت سود به تولید در سناریو اول و چهارم به‌ترتیب برای سلول سوختی ۲۲٥ و ۲۹۱ و برای ریزتوربین ۱۷۳ و ۲٤۸ است؛ بنابراین، همانند دیزل ژنراتور نیز این عامل‌ها توانسته‌اند با اکتشاف و بهره‌برداری از محیط در حین آموزش تصمیمات بهینه‌تری اتخاذ کنند. شکل ۷ نتایج شبیه‌سازی باتری را نشان می‌دهد. همان‌طور که مشاهده می‌شود در سناریوهایی که باتری آموزش داده شده است، سود آن مثبت و در سایر زمان‌ها منفی است. سود منفی به این معنا است که باتری در بیشتر زمان‌ها، انرژی را با هزینه بالا خریداری و هنگامی که قیمت برق پایین بوده، اقدام به فروش کرده است.

جدول (۲): پارامترهای توابع توزیع احتمال Beta و Weibull برای مدلسازی توان خروجی پنل خورشیدی و توربین بادی.

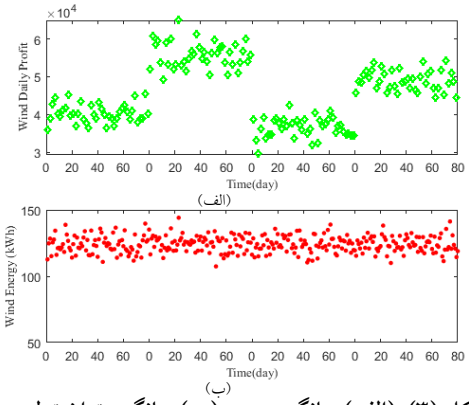
Hour	Weibull parameter (a,b)		Beta parameter (a,b)	
1	0.565282	2.032585	0.019076	10000000
2	0.568418	2.290494	0.019076	10000000
3	0.5185	1.998599	0.019076	10000000
4	0.549884	1.692578	0.019076	10000000
5	0.577571	1.628994	0.019076	10000000
6	0.58976	1.899791	0.019076	10000000
7	0.605378	2.135819	0.009781	0.293334
8	0.597392	1.44144	7.895576	18.8233
9	0.561618	1.165105	5.990502	6.199672
10	0.474976	0.840914	9.22076	4.262388
11	0.492669	1.00863	26.1694	5.562986
12	0.476329	1.134804	2.451834	0.310795
13	0.445124	1.005857	1.932275	0.066813
14	0.419577	0.88521	7.61E+13	1.27E+13
15	0.419628	0.997423	25.38013	6.917985
16	0.460638	1.097642	95.42691	68.56888
17	0.459141	1.060481	3.88E+14	5.17E+14
18	0.53346	1.170101	6.438968	32.54034
19	0.64936	1.376157	0.000455	0.107348
20	0.716691	2.589421	0.000455	0.107348
21	0.70835	2.524027	0.019076	10000000
22	0.697165	2.909614	0.019076	10000000
23	0.671788	2.851862	0.019076	10000000
24	0.604807	1.991805	0.019076	10000000

جدول (۳): ظرفیت واحدهای تولید انرژی

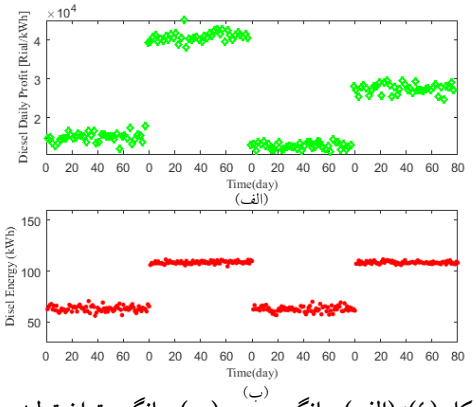
DER	Wind	PV	BESS	MT	FC	Diesel
P_{rated}(kW)	۱۰	۱۰	۵	٦	٦	۵



شکل (۲): (الف) میانگین سود، (ب)میانگین توان تولیدی روزانه پنل خورشیدی

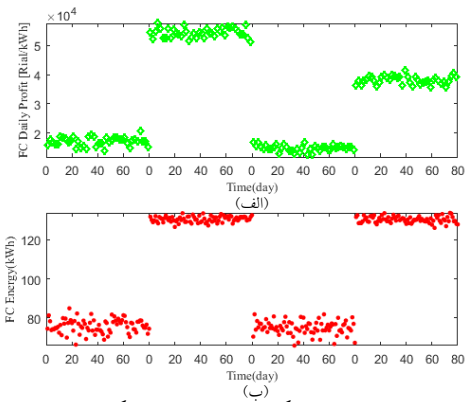


شکل (۳): (الف) میانگین سود، (ب) میانگین توان تولیدی روزانه توربین بادی

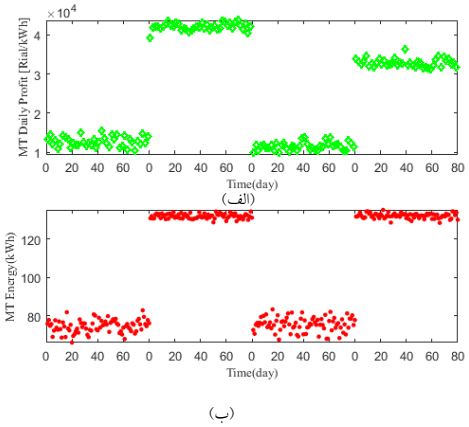


شکل (٤): (الف) میانگین سود، (ب) میانگین توان تولیدی روزانه دیزل ژنراتور

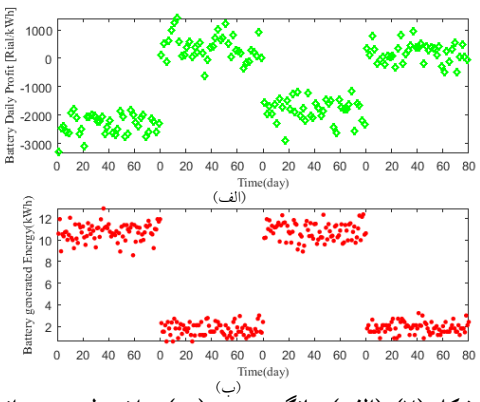
در سناریو چهارم نسبت به سناریوی دوم سود عامل‌های تولیدی کاهش یافته است؛ زیرا در حالت چهارم مصرف‌کنندگان نیز دارای قابلیت تصمیم‌گیری هوشمند هستند. در شکل ۸ و ۹ نتایج مصرف‌کنندگان الکتریکی و گرمایی به ترتیب نمایش داده شده‌اند. در سناریو سوم و چهارم مصرف‌کنندگان آموزش دیده‌اند.



شکل (٥): (الف) میانگین سود، (ب) میانگین توان تولیدی روزانه سلول سوختی

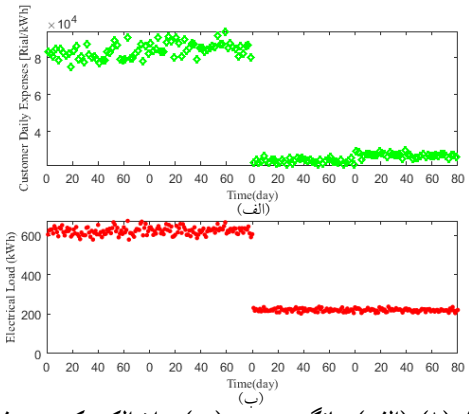


شکل (٦): (الف) میانگین سود، (ب) میانگین توان تولیدی روزانه ریزتوربین

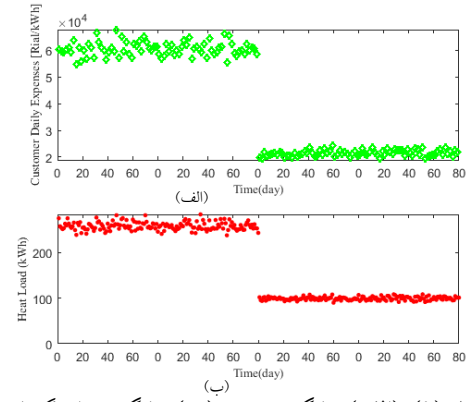


شکل (۷): (الف) میانگین سود، (ب) توان تولیدی روزانه باتری

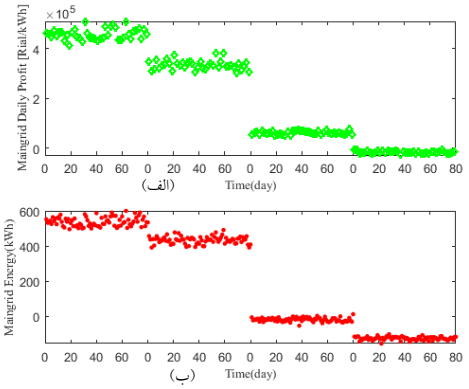
برای مقایسهٔ صحیح بین سناریوها از مقایسه نسبت هزینه به مصرف در سناریو اول و دوم با سناریو سوم و چهارم استفاده می‌شود. برای مصرف‌کنندهٔ الکتریکی این نسبت در سناریو اول تا چهارم به‌ترتیب ۱۳۲، ۱۳۷، ۱۱۰ و ۱۲٤ است. کاهش این نسبت‌ها به این معنا هست که عامل مصرف‌کننده توانسته است به مدیریت مصرف بپردازد و مصرف خود را در زمان‌هایی که قیمت برق بالا هست، کاهش دهد و در زمان‌هایی که قیمت پایین است، بیشتر کند و نیازهای خود را برآورده کند.



شکل (۸): (الف) میانگین هزینه، (ب) توان الکتریکی مصرفی روزانه مصرف‌کننده الکتریکی



شکل (۹): (الف) میانگین هزینه، (ب) میانگین توان گرمایی مصرفی روزانه مصرف‌کننده گرمایی



شکل (۱۰): (الف) میانگین سود روزانه شبکه اصلی، (ب) میانگین توان روزانه تحویل داده شده به ریزشبکه

با انتخاب عدد ۱۰ به‌عنوان ضریب نارضایتی (μ) عامل‌ها توانسته‌اند بین کاهش هزینه‌ها و به تبعِ کاهش مصرف و نیز ایجاد نارضایتی و عدم راحتی خود مصالحه کنند. با مقایسه این نسبت برای عامل‌های گرمایی در سناریوهای مختلف مطابق جدول ٤، مشاهده می‌شود توضیحات فوق برای مصرف‌کننده گرمایی نیز صادق است (شکل ۹). اگرچه هزینه مصرف‌کنندگان در سناریو چهارم نسبت به سناریو سوم اندکی افزایش یافته است، سود تولیدکنندگان در سناریو چهارم به طور درخور توجهی نیز رشد کرده است. با توجه به اینکه در یک ریزشبکه هدف این است که هم سود تولیدکنندگان افزایش یابد و هم هزینه مصرف‌کنندگان کاهش یابد، این اختلاف هزینه قابل اغماض است. همچنین، در سناریو چهارم از وابستگی ریزشبکه به شبکه اصلی کاسته شده است (رجوع شود به شکل ۱۰). مطابق شکل ۱۰، به‌مجرد اینکه تعداد عامل‌های بیشتری در ریزشبکه آموزش داده می‌شود، سود شبکه اصلی کمتر می‌شود. در سناریو چهارم حتی سود منفی شده است. سود منفی به معنای این است که سود حاصل از فروش انرژی به ریزشبکه از هزینه انرژی خریداری‌شده از ریزشبکه کمتر است. همچنین، مشاهده می‌شود در سناریو آخر توان خریداری‌شده از شبکه اصلی نیز منفی است؛ یعنی مجموع توان دریافتی از شبکه اصلی از مجموع توان داده شده به شبکه اصلی کمتر شده است؛ درنتیجه، وابستگی ریزشبکه به شبکه اصلی نیز به‌طور چشمگیری کاهش یافته است.

در شکل ۱۱ مقدار ساعتی سود/ هزینه و توان مصرفی/تولیدی عامل‌های ریزشبکه نشان داده شده است. پنل خورشیدی فقط در طول روز از ساعت ۷ الی ۱۸ در تابستان قادر است انرژی تولید کند. در سایر زمان‌ها، توان خروجی و سود پنل خورشیدی صفر است. خروجی توان توربین بادی تقریباً در ساعات شبانه‌روز یکسان است؛ زیرا این نمودار میانگین خروجی یک توربین بادی طی ۸۰۰ روز را نشان می‌دهد. در ساعات اوج مصرف بین ساعت ۱۲ الی ۲۰ به دلیل درخواست بیشتر، قیمت انرژی نیز زیادتر شده و بنابراین سود توربین بادی و سایر تولیدکننده‌ها شامل دیزل ژنراتور، ریزتوربین و سلول سوختی نیز افزایش یافته است. طبق انتظار در ساعات اوج مصرف، هزینه و توان مصرفی عامل‌های مصرفی نیز افزایش یافته است. روش پیشنهادی یک روش مدیریت انرژی ساعتی است. در مقاله [٥٤] برای محاسبه مصرف انرژی در آینده، با کمک الگوریتم لونبرگ –

مارکوارت شبکه‌های عصبی، مصرف انرژی به‌صورت کوتاه‌مدت پیش‌بینی شده است.

در این قسمت، هزینه ناشی از کاهش طول عمر باتری و میزان تخریب آن پس از هر استفاده، محاسبه و نتایج با حالت بدون مدل تخریب مقایسه می‌شود. قبل از در نظر گرفتن مدل تخریب باتری، تعداد دفعات تعویض باتری در مدت ۸۰۰ روز به‌طور میانگین ۲۳,۱ بار است. به دلیل تخریب حاصل از شارژ و تخلیه بیش از حد و استفاده نادرست از باتری، تعداد تعویض باتری زیاد شده است؛ بنابراین، با توجه به قیمت اولیه باتری و تعداد دفعات زیاد تعویض باتری در قسمت قبل، لازم است مدل تخریب باتری در نظر گرفته شود. پس از اضافه‌کردن هزینه ناشی از تخریب باتری در تابع پاداش (رابطه ۱۲)، تعداد تعویض باتری برای ۸۰۰ روز به‌طور میانگین به مقدار ۰,۸ کاهش یافته است. اگرچه سود باتری کمتر شده و نزدیک به صفر است، با توجه به کاهش تعداد دفعات باتری قابل اغماض است؛ زیرا هزینه خرید باتری بسیار زیاد است و نیز وجود باتری‌ها برای تأمین بارهای ضروری در هنگام قطع برق ضروری است.

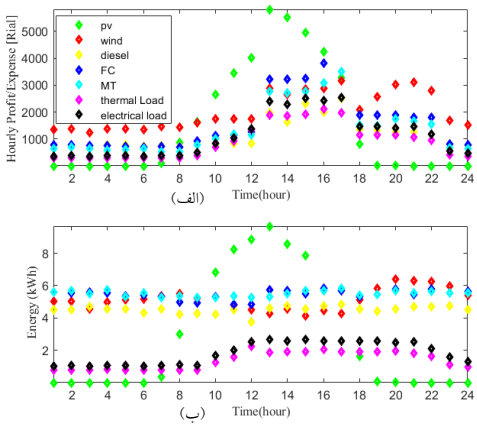
روش پیشنهادی با الگوریتم مونت‌کارلو [۵۵] مقایسه شده است. روش مونت‌کارلو بر مبنای کسب تجارب بسیار و شبیه‌سازی زیاد انجام می‌شود و درنتیجه، تخمینی که از تابع Q به دست می‌آورد، ادعا می‌شود به مقدار بهینه خیلی نزدیک می‌شود [۴۷]؛ به همین دلیل، روش مناسبی برای مقایسه است و این قابلیت را دارند که در سیستم‌های با ساختار غیرمتمرکز پیاده‌سازی شوند. جدول ۵ نتایج شبیه‌سازی را نشان می‌دهد. با توجه به جدول ۵، سود باتری در این روش منفی شده و باتری نتوانسته است به‌خوبی آموزش ببیند. همچنین، سود دیزل ژنراتور و سلول سوختی نیز نسبت به حالت قبل کاهش یافته است.

جدول (٤): میانگین نتایج الگوریتم مدیریت انرژی ریزشبکه (بعد از ۸۰۰ روز اجرا)

	سود/ هزینه روزانه (ریال)				انرژی تولیدی/ مصرفی روزانه (کیلووات ساعت)			
سناریو	اول	دوم	سوم	چهارم	اول	دوم	سوم	چهارم
PV	31284	52206	27318	37318	70	70	70	70
Wind	40117	55431	36919	48966	124	125	124	124
Diesel	15042	40462	13086	27448	63	108	63	108
FC	17092	54000	15378	37795	76	130	76	130
MT	12766	41595	11885	32737	74	131	76	132
Elec. load	81896	85706	24893	27321	621	625	226	221
Heat load	60608	60308	21582	21751	258	257	100	99
Battery	-2333	324	-1746	142	10.7	1.7	10.7	2.0
Maingrid	457950	335280	64976	-14606	539	435	-15	-125

جدول (٥): میانگین نتایج شبیه‌سازی مدیریت انرژی ریزشبکه با استفاده از روش [۵۵] (بعد از ۸۰۰ روز اجرا)

	سود/ هزینه روزانه (ریال)				انرژی تولیدی/ مصرفی روزانه (کیلووات ساعت)			
سناریو	اول	دوم	سوم	چهارم	اول	دوم	سوم	چهارم
PV	31366	55515	26766	39581	70	70	70	70
Wind	39913	64538	36618	57912	124	124	124	124
Diesel	14959	47637	12969	27127	63	118	64	118
FC	17109	48588	16249	29429	76	142	77	142.4
MT	13097	54168	11799	40907	75	143	76	143
Elec. load	81897	86893	21640	26492	623	619	201	197
Heat load	60276	61025	17850	19474	257	256	85	84.8
Battery	-2399	-9.5	-1607	-14.2	10.7	0.18	10.6	0.16
Maingrid	441970	280610	23115	-60870	485	351	-108	-241



شکل (۱۱): (الف) میانگین سود/هزینه ساعتی، (ب) میانگین توان تولیدی/مصرفی ساعتی پنل خورشیدی، توربین بادی، دیزل، سلول سوختی، ریز توربین و بارهای گرمایی و الکتریکی

سود سایر عامل‌ها افزایش یافته است. برای مقایسه عادلانه دو روش از شاخص مقایسه Fairness Factor (FF) در مقاله [۲۷] استفاده شده است. در این شاخص، سود ریزشبکه با توجه به سود کلیه عامل‌های تولیدی و مصرفی محاسبه می‌شود. مقدار شاخص FF در سناریوی چهارم برای روش مونت‌کارلو ۱٫۶۳ است و برای روش ارائه‌شده در این مقاله ۱٫۸۷ است. شاخص FF برای روش مونت‌کارلو به‌طور درخور توجهی از روش پیشنهادی کوچک‌تر است؛ بنابراین، از مقایسه مقدار فاکتور FF در این دو روش، می‌توان نتیجه گرفت سود ریزشبکه در روش پیشنهادی با در نظر گرفتن سود کلیه عامل‌ها بهبود یافته است.

۵- نتیجه‌گیری

در این مقاله، یک روش جدید غیرمتمرکز برای مدیریت انرژی الکتریکی و گرمایی ساعتی یک ریزشبکه پیشنهاد شد. در این روش، با در نظر گرفتن عدم‌قطعیت در تقاضای بارهای الکتریکی و گرمایی، انرژی تجدیدپذیر و قیمت برق، یک سیستم مدیریت انرژی مستقل از مدل با استفاده از یادگیری تقویتی ارائه شد. برخلاف روش‌های سنتی مبتنی بر مدل که نیازمند یک تخمین‌گر عدم قطعیت است، این روش براساس یادگیری است و نیاز به یک مدل صریح از عدم قطعیت ندارد. دردسترس بودن اطلاعات، برای یک واحد کنترل‌کننده مرکزی یا حتی برای عامل‌های همسایه، در عمل سخت است. با افزایش ابعاد شبکه‌های قدرت این مشکل بیشتر می‌شود؛ بنابراین، با استفاده از روش غیرمتمرکز پیشنهادی، مشکلات ناشی از پیچیدگی ارتباطات و محاسبات برطرف شد. عملکرد روش ارائه‌شده تحت چهار سناریو: بدون یادگیری، یادگیری تولیدکنندگان، یادگیری مصرف‌کنندگان و یادگیری همه عامل‌ها شبیه‌سازی شد. برای ارزیابی مدل پیشنهادی از داده‌های واقعی توربین بادی و پنل خورشیدی و داده‌های بازار برق ایران استفاده شد. در قسمت شبیه‌سازی مقاله نشان داده شد سود کلیه واحدهای تولیدی افزایش، هزینه مصرف‌کنندگان کاهش و رضایت‌مندی آنها افزایش پیدا کرده است. همچنین، روش ارائه‌شده، از وابستگی ریزشبکه به شبکه اصلی نیز کاسته است.

علاوه بر آن، قابلیت پیاده‌سازی روش پیشنهادی به‌صورت ساعتی برای مدیریت انرژی ریزشبکه‌ها نشان داده شده است. در پایان، اثبات همگرایی روش پیشنهادی به جواب بهینه یا نزدیک به بهینه به‌عنوان کارهای آتی پیشنهاد می‌شود.

سپاسگزاری

بدین‌وسیله از حمایت‌های پژوهشی پژوهشگاه نیرو و همچنین، از حمایت‌های پژوهشگاه هوا خورشید دانشگاه فردوسی مشهد برای جمع‌آوری داده‌های عملی، صمیمانه تشکر و قدردانی می‌کنم.

مراجع

[1] M. F. Akorede, H. Hizam, and E. Pouresmaeil, "Distributed energy resources and benefits to the environment," *Renewable and sustainable energy reviews*, Vol. 14, No. 2, pp. 724-734, 2010.

[2] L. Mariam, M. Basu, and M. F. Conlon, "Microgrid: Architecture, policy and future trends," *Renewable and Sustainable Energy Reviews*, Vol. 64, pp. 477-489, 2016.

[3] L. G. Meegahapola, D. Robinson, A. Agalgaonkar, S. Perera, and P. Ciufo, "Microgrids of commercial buildings: Strategies to manage mode transfer from grid connected to islanded mode," *IEEE Transactions*

algorithms," *Automatica*, Vol. 104, pp. 90-101, 2019.
[17]B. Javanmard, M. Tabrizian, M. Ansarian, and A. Ahmarinejad, "Energy management of multi-microgrids based on game theory approach in the presence of demand response programs, energy storage systems and renewable energy resources," *Journal of Energy Storage*, Vol. 42, p. 102971, 2021.
[18]S. A. Mansouri, A. Ahmarinejad, M. S. Javadi, and J. P. Catalão, "Two-stage stochastic framework for energy hubs planning considering demand response programs," *Energy*, Vol. 206, p. 118124, 2020.
[19]S. Rao, "Game theory approach for multiobjective structural optimization," *Computers & Structures*, Vol. 25, No. 1, pp. 119-127, 1987.
[20]M. R. B. Khan, R. Jidin, and J. Pasupuleti, "Multi-agent based distributed control architecture for microgrid energy management and optimization," *Energy Conversion and Management*, Vol. 112, pp. 288-307, 2016.
[21]S. Mei, Y. Wang, F. Liu, X. Zhang, and Z. Sun, "Game approaches for hybrid power system planning," *IEEE Transactions on Sustainable Energy*, Vol. 3, No. 3, pp. 506-517, 2012.
[22]A. S. Chuang, F. Wu, and P. Varaiya, "A game-theoretic model for generation expansion planning: problem formulation and numerical comparisons," *IEEE transactions on power systems*, Vol. 16, No. 4, pp. 885-891, 2001.
[23]M. Bowling and M. Veloso, "An analysis of stochastic game theory for multiagent reinforcement learning," Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science, 2000.
[24]A. Hernandez-Matheus *et al.*, "A systematic review of machine learning techniques related to local energy communities," *Renewable and Sustainable Energy Reviews*, Vol. 170, p. 112651, 2022.
[25]X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A Multi-agent Reinforcement Learning based Data-driven Method for Home Energy Management," *IEEE Transactions on Smart Grid*, 2020.
[26]G. K. Venayagamoorthy, R. K. Sharma, P. K. Gautam, and A. Ahmadi, "Dynamic energy management system for a smart microgrid," *IEEE transactions on neural networks and learning systems*, Vol. 27, No. 8, pp. 1643-1656, 2016.
[27]E. Foruzan, L.-K. Soh, and S. Asgarpoor, "Reinforcement learning approach for optimal distributed energy management in a microgrid," *IEEE Transactions on Power Systems*, Vol. 33, No. 5, pp. 5749-5758, 2018.
[28]F.-D. Li, M. Wu, Y. He, and X. Chen, "Optimal control in microgrid using multi-agent reinforcement learning," *ISA transactions*, vol. 51, no. 6, pp. 743-751, 2012.
[29]T. G. Dietterich, "Hierarchical reinforcement learning with the MAXQ value function decomposition," *Journal of artificial intelligence research*, Vol. 13, pp. 227-303, 2000.
[30]W. Liu, P. Zhuang, H. Liang, J. Peng, and Z. Huang, "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," *IEEE transactions on neural networks and learning systems*, Vol. 29, No. 6, pp. 2192-2203, 2018.
[31]L. Yang, Q. Sun, D. Ma, and Q. Wei, "Nash Q-
on Sustainable Energy, Vol. 5, No. 4, pp. 1337-1347, 2014.
[4] L. Lei, Y. Tan, G. Dahlenburg, W. Xiang, and K. Zheng, "Dynamic Energy Dispatch Based on Deep Reinforcement Learning in IoT-Driven Smart Isolated Microgrids," *IEEE Internet of Things Journal*, 2020.
[5] H. Shayeghi, E. Shahryari, M. Moradzadeh, and P. Siano, "A survey on microgrid energy management considering flexible energy sources," *Energies*, Vol. 12, No. 11, p. 2156, 2019.
[6] Z. Wang, B. Chen, J. Wang, M. M. Begovic, and C. Chen, "Coordinated energy management of networked microgrids in distribution systems," *IEEE Transactions on Smart Grid*, Vol. 6, No. 1, pp. 45-53, 2014.
[7] W. Su, J. Wang, and J. Roh, "Stochastic energy scheduling in microgrids with intermittent renewable energy resources," *IEEE Transactions on Smart grid*, Vol. 5, No. 4, pp. 1876-1883, 2013.
[8] J. S. Giraldo, J. A. Castrillon, J. C. López, M. J. Rider, and C. A. Castro, "Microgrids energy management using robust convex programming," *IEEE Transactions on Smart Grid*, Vol. 10, No. 4, pp. 4520-4530, 2018.
[9] W. Shi, N. Li, C.-C. Chu, and R. Gadh, "Real-time energy management in microgrids," *IEEE Transactions on Smart Grid*, Vol. 8, No. 1, pp. 228-238, 2015.
[10]W. Hu, P. Wang, and H. B. Gooi, "Toward optimal energy management of microgrids via robust two-stage optimization," *IEEE Transactions on smart grid*, Vol. 9, No. 2, pp. 1161-1174, 2016.
[11]S. Haddadipour, V. Amir, and S. Javadi Arani, "A strategy proposing the simultaneous purchase and sale of electricity to exploit of a multi-agent micro-grid energy market," *Computational Intelligence in Electrical Engineering*, Vol. 11, No. 4, pp. 93-110, 2020.
[12]S. Umetani, Y. Fukushima, and H. Morita, "A linear programming based heuristic algorithm for charge and discharge scheduling of electric vehicles in a building energy management system," *Omega*, Vol. 67, pp. 115-122, 2017.
[13]V. J. Gutierrez-Martinez, C. A. Moreno-Bautista, J. M. Lozano-Garcia, A. Pizano-Martinez, E. A. Zamora-Cardenas, and M. A. Gomez-Martinez, "A heuristic home electric energy management system considering renewable energy availability," *Energies*, Vol. 12, No. 4, p. 671, 2019.
[14]A. Seifi, M. H. Moradi, M. Abedini, and A. Jahangiri, "Assessing the impact of load response on microgrids with the aim of increasing the reliability and stability of network voltage by examining the uncertainty in the production of renewable resources," *Computational Intelligence in Electrical Engineering*, Vol. 12, No. 1, pp. 87-98, 2021.
[15]X. S. Zhang, T. Yu, Z. N. Pan, B. Yang, and T. Bao, "Lifelong learning for complementary generation control of interconnected power grids with high-penetration renewables and EVs," *IEEE Transactions on Power Systems*, Vol. 33, No. 4, pp. 4097-4110, 2017.
[16]Z. Hu, M. Zhu, P. Chen, and P. Liu, "On convergence rates of game theoretic reinforcement learning

[47]S. Richard, B. SUTTON, and G. Andrew, *Reinforcement learning: an introduction*. MIT press, 2018.
[48]C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3, pp. 279-292, 1992.
[49]C. J. C. H. Watkins, "Learning from delayed rewards," 1989.
[50]B. Aksanli and T. Rosing, "Optimal battery configuration in a residential home with time-of-use pricing," in *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2013: IEEE, pp. 157-162.
[51]D. Doerffel and S. A. Sharkh, "A critical review of using the Peukert equation for determining the remaining capacity of lead-acid and lithium-ion batteries," *Journal of power sources*, Vol. 155, No. 2, pp. 395-400, 2006.
[52]G. Bowden, P. Barker, V. Shestopal, and J. Twidell, "The Weibull distribution function and wind power statistics," *Wind Engineering*, pp. 85-98, 1983.
[53]S. Trashchenkov and V. Astapov, "The applicability of zero inflated beta distributions for stochastic modeling of PV plants' power output," in *2018 19th International Scientific Conference on Electric Power Engineering (EPE)*, 2018: IEEE, pp. 1-6.
[54]R. Darshi, M. A. Bahreini, and S. A. Ebrahim, "Prediction of Short-Term Electricity Consumption by Artificial Neural Networks Levenberg-Marquardt Algorithm in Hormozgan Province, Iran," in *2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 2019: IEEE, pp. 1-4.
[55]Y. Du and F. Li, "Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning," *IEEE Transactions on Smart Grid*, Vol. 11, No. 2, pp. 1066-1076, 2019.
learning based equilibrium transfer for integrated energy management game with We-Energy," *Neurocomputing*, Vol. 396, pp. 216-223, 2020.
[32]J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *Journal of machine learning research*, Vol. 4, No. Nov, pp. 1039-1069, 2003.
[33]V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *nature*, Vol. 518, No. 7540, pp. 529-533, 2015.
[34]Y. Ji, J. Wang, J. Xu, X. Fang, and H. Zhang, "Real-time energy management of a microgrid using deep reinforcement learning," *Energies*, Vol. 12, No. 12, p. 2291, 2019.
[35]K. Deshpande, P. Möhl, A. Hämmerle, G. Weichhart, H. Zörrer, and A. Pichler, "Energy Management Simulation with Multi-Agent Reinforcement Learning: An Approach to Achieve Reliability and Resilience," *Energies*, Vol. 15, No. 19, p. 7381, 2022.
[36]M. H. Alabdullah and M. A. Abido, "Microgrid energy management using deep Q-network reinforcement learning," *Alexandria Engineering Journal*, Vol. 61, No. 11, pp. 9069-9078, 2022.
[37]C. Guo, X. Wang, Y. Zheng, and F. Zhang, "Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning," *Energy*, Vol. 238, p. 121873, 2022.
[38]M. Andreasson, D. V. Dimarogonas, H. Sandberg, and K. H. Johansson, "Distributed PI-control with applications to power systems frequency control," in *2014 American Control Conference*, 2014: IEEE, pp. 3183-3188.
[39]A. Akbarimajd, M. Olyaee, B. Sobhani, and H. Shayeghi, "Nonlinear multi-agent optimal load frequency control based on feedback linearization of wind turbines," *IEEE Transactions on Sustainable Energy*, Vol. 10, No. 1, pp. 66-74, 2018.
[40]V. C. Gungor *et al.*, "Smart grid technologies: Communication technologies and standards," *IEEE transactions on Industrial informatics*, Vol. 7, No. 4, pp. 529-539, 2011.
[41]Y. Li and Y. W. Li, "Power management of inverter interfaced autonomous microgrid based on virtual frequency-voltage frame," *IEEE Transactions on Smart Grid*, Vol. 2, No. 1, pp. 30-40, 2011.
[42]Q. Jiang, M. Xue, and G. Geng, "Energy management of microgrid in grid-connected and stand-alone modes," *IEEE transactions on power systems*, Vol. 28, No. 3, pp. 3380-3389, 2013.
[43]V. Vahidinasab, "Optimal distributed energy resources planning in a competitive electricity market: Multiobjective optimization and probabilistic design," *Renewable energy*, Vol. 66, pp. 354-363, 2014.
[44]M. Q. Wang and H. Gooi, "Spinning reserve estimation in microgrids," *IEEE Transactions on Power Systems*, Vol. 26, No. 3, pp. 1164-1174, 2011.
[45]A. Cagnano, A. C. Bugliari, and E. De Tuglie, "A cooperative control for the reserve management of isolated microgrids," *Applied energy*, vol. 218, pp. 256-265, 2018.
[46]H. Zhang, H. Sun, Q. Zhang, and G. Kong, "Microgrid Spinning Reserve Optimization with Improved Information Gap Decision Theory," *Energies*, Vol. 11, No. 9, p. 2347, 2018.