افزایش سودآوری بازار شبکه های هوشمند برق با تکنیک یادگیری تقویتی عملگر-نقاد

اکرم بیگی[1] امین اکبریان[2]

۱- استادیار- دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی، تهران

akrambeigi@sru.ac.ir

۲ – دانش آموخته کارشناسی ارشد- دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی، تهران

a.akbarian@sru.ac.ir

چکیده: بازار شبکه‌های هوشمند برق پیچیده و پویاست. کارگزاران که واسطه‌گران فروش برق بین خرده‌فروشی‌ها و عمده‌فروشی‌ها هستند به‌صورت گسترده‌ای در بازارهای جدید شبکه‌های هوشمند به کار گرفته می‌شوند. به‌علت پیچیدگی و توزیع‌شدگی ذاتی بازار در شبکه‌های هوشمند رویکردهای استفاده از سیستم‌های چندعامله برای حل مسائل آن مناسب است. در این رویکردها می‌توانیم عامل‌های خودمختاری داشته باشیم که به‌صورت ۲۴ ساعته درحال تبادل اطلاعات با دیگر عامل‌ها هستند. این عامل‌ها با چالش های اساسی شامل الگوی مصرف متنوع مشتریان، تغییر قیمت با توجه به الگوی مصرف مشتریان و میزان مصرف برق در طول شبانه روز مواجه‌اند. هدف ما در این مقاله این است که ضمن مدل کردن اجزای بازار برق با سیستم‌های چندعامله، با ارائه روشی مبتنی بر یادگیری عامل‌ها سودآوری در بازار شبکه های برق را افزایش دهیم. در روش پیشنهادی ابتدا مساله تنوع مصرف مشتریان را با انجام یک روش خوشه‌بندی متوالی مناسب داده‌های سری زمانی پردازش می‌کنیم. سپس برای هر گروه خوشه‌بندی شده به صورت مجزا یک روش یادگیری تقویتی سیاست فعال با عنوان یادگیری تقویتی عملگر- نقاد به کار می‌بریم. درنهایت تاثیر تغییر پاداش را در سود حاصله ارزیابی می‌کنیم و برای هر خوشه تعرفه‌ای مطابق با زمان مصرف مربوطه به صورت ساعتی ارائه می‌دهیم.

واژه‌های کلیدی: شبکه های هوشمند، انرژی های تجدیدپذیر، بازار تعرفه، یادگیری تقویتی، خوشه بندی.

نوع مقاله: پژوهشی

DOI: 10.52547/jiaeee.19.1.245

تاریخ ارسال مقاله: ۱۳۹۸/۱۱/۱۴

تاریخ پذیرش مشروط مقاله: ۱۳۹۹/۰۷/۰۲

تاریخ پذیرش مقاله: ۱۳۹۹/۱۱/۵

نام نویسنده‌ی مسئول: دکتر اکرم بیگی

نشانی نویسنده‌ی مسئول: ایران – تهران- لویزان- دانشگاه تربیت دبیر شهید رجایی- دانشکده مهندسی کامپیوتر

۱- مقدمه

شبکه هوشمند[1] یک سیستم الکتریکی تکامل یافته است که تقاضای برق را به شیوه پایدار، قابل اعتماد و اقتصادی مدیریت می کند و روی زیرساخت‌های پیشرفته ساخته و تنظیم شده است. به واسطه این زیرساخت‌ها، انسجام تمام مجموعه‌های درگیر در این شبکه برای انتقال به سمت منابع پایدار انرژی تسهیل می شود. از نظر تئوری، از یک سو شبکه های هوشمند مکانیزمی پایدار و کارآمد را برای مدیریت سیستم‌های تولید و توزیع برق ارائه داده و از سوی دیگر بستری منعطف و پویا برای تعرفه مصرف انرژی توسط مشتریان ایجاد می‌کنند [۱]. نمایندگی‌های توزیع برق که در شبکه هوشمند کار می‌کنند، که ما از آنها به عنوان کارگزاران[2] یاد می‌کنیم، می‌توانند عدم تعادل عرضه و تقاضا در بازار را از طریق استراتژی های قیمت گذاری پویا نشان دهند؛ در حالی که همزمان هزینه های سربار مشتریانشان با خرید عمده انرژی از شرکت‌های تولیدکننده بزرگ کاهش می‌یابد [۲]. با این وجود، چالش های متعددی در عملیاتی شدن این استراتژی وجود دارد، مانند مدیریت نوسان بالای سناریوهای عرضه و تقاضا، ترغیب ذینفعان با انگیزه های پنهانی و مدیریت خطاهای هوشمندسازی موجودیت های شرکت کننده. یکی از اهداف مهم در این حوزه حین بهبود تعادل در وضعیت عرضه و تقاضا، افزایش سودآوری برای کارگزاران است که در این پژوهش به آن می پردازیم.

سالهای زیادی در بازار عمده‌فروشی برق، شرکت‌های تولیدکننده فقط به دنبال رقابت با یکدیگر برای فروش انرژی الکتریکی به مشتریان بودند تا از این طریق سود بیشتری کسب کنند و هیچ مکانیزمی برای امکان مشارکت طرف تقاضا به ویژه برای واحدهای ساختمانی مسکونی با مصرف برق نسبتاً زیاد وجود نداشت [۳]. برخلاف سیستم‌های توزیع برق سنتی که جریان توزیع انرژی الکتریکی در آن یک طرفه است، در بازار شبکه های هوشمند برق جریان سیستم دوطرفه است. درچنین سیستم‌های دو طرفه‌ای کنش و تراکنش‌های مصرف‌کننده‌ها، توصرف‌کننده‌ها[3] (موجودیتی که هم تولیدکننده و هم مصرف‌کننده انرژی است) و عرضه کننده های انرژی با یکدیگر در تعامل هستند [۴]. تقاضاهای متنوع انرژی، تغییرات قیمت و مهاجرت مشتری ها بین تعرفه‌های مختلف کارگزارها سبب ایجاد پیچیدگی در بازار شبکه های هوشمند می‌شود. برای غلبه بر این پیچیدگی و پویایی انواع مختلف کارگزاران برق، از الگوریتم‌های هوش مصنوعی متنوعی برای مدل کردن و حل مسائل داد و ستد بین اجزای شبکه استفاده شده است.

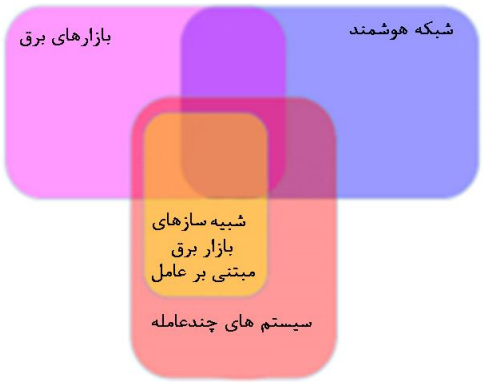
شبکه هوشمند از بازار تعرفه، بازار عمده فروشی و ابزار توزیع تشکیل شده است. در بازار تعرفه محلی، مصرف کنندگان (برای مثال خانوارها) برق را خریداری می‌کنند. تولیدکنندگان (بعنوان مثال ژنراتور خورشیدی) برق را از طریق خرده فروشی به کارگزاران خاص می‌فروشند. کارگزاران به طور خاص، قراردادهای تعرفه را برای جذب مشتری به منظور توسعه سبد برق خود منتشر می‌کنند. در بازار عمده فروشی نیروگاه‌ها انرژی تولید شده را با روش‌های معمولی (به عنوان مثال با زغال سنگ) به کارگزارها می فروشند. کارگزارها انرژی برق را برای تحویل در آینده خریداری می‌کنند یا می‌فروشند. همچنین واحدها و تاسیسات توزیع مانند ایستگاه‌های فرعی و ایستگاه‌های ذخیره انرژی پاسخگوی تعادل عرضه و تقاضا در زمان واقعی هستند. به عنوان مثال، هنگامی که شکاف برق (ایجاد قطعی در یکی شبکه‌های توزیع انرژی محلی) در نمونه کارهای کارگزار پدیدار می‌شود، واحدتوزیع، منبع اضطراری را تأمین می‌کند و هزینه‌های آن را پرداخت می‌کند. کارگزاران سنتی نیروی برق را از بازار عمده فروشی تهیه می‌کنند. با کاهش میزان منابع زغال سنگ و منابع نفتی، انرژی‌های تجدیدپذیر جایگزین روش‌های متداول تولید برق می‌شوند. بنابراین، عمده ترین کار کارگزاران آینده خرید انرژی تجدیدپذیر محلی توزیع شده برای جلب رضایت مصرف کنندگان خود است. در بازار عمده فروشی به علت تنوع تولیدکننده‌های انرژی، قیمت‌ها، کمیت‌ها و همچنین به علت حضور انرژی‌های تجدیدپذیر، پایداری‌های متنوعی از تولیدکنندگان و مصرف‌کنندگان به نمایش در می‌آید. در بازار خرده‌فروشی به دلیل وجود مشتری‌های مختلف، نیازمندی‌های مختلفی روی قیمت، کمیت، کیفیت و زمان مصرف انرژی وجود دارد.

۱-۱- انگیزه پژوهش

از آنجایی که جمعیت جهان در حال افزایش است تقاضای انرژی مازاد (سالانه ۳۵ درصد تا سال ۲۰۴۰ تخمین زده شده است) باید به طور مداوم عرضه شود تا توسعه اقتصادی در بازار شبکه های هوشمند برق پایدار بماند [۵]. با این حال، برای تطبیق تقاضای پیش بینی شده، جنبه بهره‌وری انرژی باید به دقت مورد بررسی قرار گیرد. پیشرفت‌های اخیر شبکه‌های هوشمند انرژی نوید بهبودهای چشمگیری را در بهره‌وری انرژی می‌دهد. عمده این پیشرفت‌ها در توسعه تولیدکننده‌های غیرمتمرکز و تکنولوژی‌های ذخیره‌سازی است [۶]؛ به گونه‌ای که مشتریان بتوانند در داد و ستد دوطرفه غیرمتمرکز بازار برق شرکت کنند و تا حد زیادی نیازهای خود را بدون ایجاد فشار بیشتر به شبکه برآورده سازند.

از مزایای تغییر بازار برق از حالت انحصاری به غیرانحصاری و شرکت مشتریان (مصرف‌کنندگان بازار سنتی برق) در خرید و فروش انرژی، اصلاح الگوی مصرف است. یک راه حل بهره وری انرژی با هدف کاهش هزینه های برق و حفاظت از محیط زیست ارائه سرویس های خاص مانند تجهیزات کم مصرف و سایر ابزارهای بهبود کارایی است. راه حل دیگر برای اصلاح الگوی مصرف تغییر زمان استفاده از برق به طور داوطلبانه توسط کاربران با سیاست های تعیین قیمت انرژی است [۷]. اما این امر مدیریت قیمت و تعرفه‌گذاری را با چالش اساسی روبه‌رو می‌کند. یکی از رویکردهای حل مساله در این حوزه به کارگیری عامل‌های هوشمند تحت عنوان کارگزار است. در این حالت می‌توان یک هم‌ارزی و تعادل بین عرضه و تقاضای انرژی ایجاد کرد. شکل (۱)

جایگاه رویکردهای مبتنی بر سیستم‌های چندعامله و ابزارهای موجود در این حوزه را در حل مسائل شبکه‌های هوشمند و بازار برق نشان می دهد.



شکل(۱): ارتباط سیستم چند عامله و بازار شبکه‌های هوشمند

در شکل (۱)، بخش شبکه هوشمند و بازارهای شبکه های برق بیشتر مرتبط با تعرفه‌های متناسب با زمان استفاده[۴] (TOU) و مدیریت سمت تقاضا[۵] (DSM) است. تاکید بر این دو رویکرد طراحی شبکه‌های هوشمند را با چالش تعداد زیاد شرکت‌کنندگان (مشتری) غیرهمگن که دارای نفع شخصی هستند روبرو می‌کند. از این‌رو نیاز به استفاده سیستم‌های چندعامله ضروری به نظر می‌رسد. در حوزه مدل‌سازی با سیستم‌های چندعامله نیز نیازمند طراحی محیطی غنی از تکنیک‌هایی هستیم که می‌توانند به درک و تحلیل پیچیدگی‌های بازار برق کمک کنند.

محققان مشکلاتی را که ممکن است در بخش‌های کلیدی شبکه‌های هوشمند در آینده بوجود آید را در این حوزه‌ها دسته‌بندی می‌کنند: مدیریت عرضه و تقاضا، ماشین‌های الکتریکی، نیروگاه‌های مجازی، شبکه‌های خودترمیم[۶] و همچنین بوجود آمدن مشتریانی که هم در کار تولید و هم در کار مصرف برق هستند [۸,۹]. سودمندسازی عامل‌های تجاری و متعادل‌سازی عرضه و تقاضا به گونه‌ای که باعث کاهش مصرف انرژی در زمان اوج مصرف شود مسئله‌هایی هستند که در این مقاله مورد تحلیل و بررسی قرار گرفته اند. همانطور که اشاره شد با توجه به پیچیدگی و توزیع شدگی ذاتی بازار برق که شامل موجودیت‌های خودمختاری است که در تمام طول شبانه‌روز در حال فعالیت هستند، استفاده از روش‌های متداول یادگیری تقویتی برای مدل سازی مسئله که عمدتا مبتنی بر چارچوب فرایند تصمیم‌گیری مارکوف است، نتیجه‌ای جز چالش نفرین ابعاد و همچنین واریانس بالای بین حالت‌های مختلف ندارد. بنابراین در این پژوهش ما روشی مبتنی بر مبتنی بر سیستم‌های چندعامله و یادگیری عملگر-نقاد ارائه می دهیم تا ضمن مقابله با موارد گفته شده، هم سودآوری بیشتری برای کارگزاران کسب کنیم و هم بازه طولانی‌تری در طول شبانه‌روز میزان عرضه و تقاضا در وضعیت تعادل باشد. حضور مشتریان در بازار پیچیده شبکه‌های هوشمند برق نیاز به یک الگوریتم تعرفه‌گذاری هوشمند ساعتی را که بهره‌وری را تضمین کند ضروری می‌سازد. به طور کلی می توان انگیزه‌های پیاده سازی الگوریتم پیشنهادی را در صرفه‌جویی و کاهش در هزینه، افزایش سود مشتریان و بهبود بهره‌وری کلی سیستم دانست.

۱-۲- پژوهش‌های مرتبط

لازم به ذکر است که تحقیقات روی بازار شبکه‌های هوشمند در بخش وسیعی از حوزه پایداری محاسباتی نیز تاثیرگذار است. پایداری محاسباتی یک شاخه‌ی میان رشته‌ای است که هدف از آن به کارگیری تکنیک‌هایی از رشته‌های کامپیوتر، دانش اطلاعات، تحقیق در عملیات (پژوهش‌های عملیاتی)، ریاضیات کاربردی و آمار برای متعادل‌سازی (هم‌ارزی) نیازهای محیطی، اقتصادی و اجتماعی برای توسعه پایدار است [۱۰].

یکی از اولین کارگزارانی که تعرفه‌های متناسب با زمان استفاده را به کار گرفت Mercator بودکه با استفاده از تعرفه‌های TOU با ۲ یا ۳ نرخ روزانه استفاده می‌شد. تا آن زمان شبیه‌سازها تنها مشتریان ثابت را مدل می‌کردند و اثر تعرفه های TOU نمی توانست در حضور مشتریان با تقاضای متغیر مورد آزمایش قرار گیرد. Mercator اولین عامل تجاری رقابتی بود که از تعرفه‌های متغیر استفاده کرد [۴] و [۱۱]. Mercator از دو نوع استراتژی برای تعریف تعرفه استفاده کرد: ۱) راهبرد مدل‌سازی تعرفه ۲) راهبرد به‌روز تعرفه. هر دوی آنها به هدف بهینه سازی در نظر گرفته شده بودند که در آن اهداف کارگزار نیز افزایش سود و تثبیت سهام بازار برای مشتری مناسب بود. Mercator از الگوریتم روش بهینه‌سازی ازدحام ذرات[۷] برای جستجوی تعرفه‌ای با بالاترین سود پیش‌بینی شده بهره‌گیری کرده است. به این منظور فضای مساله را با ذرات چهار تا شش بعدی که هر کدام نشان‌دهنده تعرفه‌ای بودند مدل کرده‌اند. هدف نیز جستجوی تعرفه‌ای با بالاترین سود پیش‌بینی شده برای کارگزار تعریف شده است. لازم به ذکر است که محدویت‌های فضای جستجوی ذرات بر سهم بازار Mercator تاثیر داشته است.

پژوهش [۱۲] را می توان جزء اولین تحقیق‌ها در خصوص استفاده از TOU توسط عاملین مستقل در یک مقیاس بزرگ و دقیق به‌شمار آورد. در این کار، شبیه‌سازی واقع بینانه ای از بازارهای رقابتی خرده فروشی مستقل و مشتریان کارگزار با تقاضای متغیر با استفاده از تکنیک LATTE انجام شده است. LATTE یک چارچوب کلی است که می تواند به روش های مختلفی مقدار دهی اولیه شود و از اینرو برای تنظیمات خاص بازار شبکه های هوشمند مناسب است. همچنین در این تحقیق به موضوع مشتریان فرصت طلب[۸] نیز پرداخته شده است. آنها مشتریانی هستند که از کاهش نرخ انرژی که به خاطر تغییر تقاضای سایر مشتریان انجام گرفته است سود می برند بدون آنکه تقاضا و مصرف انرژی خود را کاهش دهند و همچنین می‌توانند تا

۹۰درصد از سود حاصل از قیمت گذاری در لحظه[۹] (RTP) بهره‌مند گردند[۶].

عامل Cwibroker [۱۳] از دو راهبرد تعرفه مختلف بهره می‌برد. در انحصار دوگانه فروش (بازاری که دو شرکت تقریبا دارای همه سهم بازار هستند)، این راهبرد تعرفه با الهام از الگوریتم تلافی[۱۰] مورد استفاده قرار گرفت. در انحصار چندجانبه از تعرفه‌ای استفاده کرد که در آن کاندیدایی با تعرفه‌های با نرخ ثابت ایجاد کرد و به تخمین میزان بهره آینده پرداخت. ایده سود تعرفه تخمین زده، به بهینه سازی سودمندی LATTE شباهت دارد، هر چند به نظر می‌رسد به روش متفاوتی به اجرا رسیده است. گزارش شده است که این راهبرد در رابطه با انحصار چندجانبه کارآمد نبوده و راهبرد مبتنی بر اکتشاف[۱۱] برای بهبود اجرا در این ساختار مناسب بوده است.

کارگزار Aston [۱۴] از فرایند تصمیم‌گیری مارکوف[۱۲] (MDP) استفاده کرد تا روند مناقصه‌ای را به صورت عمده‌فروشی انجام دهد. همچنین از یک SMDP[۱۳] (یک سیستم پویا که حالت‌هایش به صورت تکرارهای تصادفی تجدیدنظر شده‌اند) جداگانه برای روند انتخاب تعرفه استفاده کرد. در بخش MDP، Aston به صورت فرضی یک مدل گسسته پایه[۱۴] برای ارزش پاکسازی عمده فروشی (برابری بازار در عرضه و تقاضا پاکسازی نامیده می‌شود) ایجاد کردکه در آن ۲۰ حالت ممکن از داده‌های اولیه مسئله به صورت برون‌خط[۱۵] ساخته شده بودند.

یک مدل بازار برق با قیمت گذاری پویا و مصرف انرژی در یک میکروشبکه در [۱۵] مطالعه شده است که از الگوریتم Q-learning برای یادگیری کاهش هزینه سیستم برای مشتریان استفاده می کند. در این مدل که مبتنی بر سیستم های چندعامله است، هر مشتری می تواند برای زمان بندی مصرف انرژی خود براساس قیمت خرده فروشی مشاهده شده با هدف به حداقل رساندن هزینه پیش بینی شده، مدل سیستم خود ایجاد کند.

در پژوهش‌های ذکر شده از MDP به صورت سنتی و دستی برای مدل سازی بازار تعرفه شبکه های هوشمند استفاده می‌کنند و تنظیمات آن را به صورت دستی و مکاشفه‌ای انجام می دهند. این مسئله باعث کندی و نفرین ابعاد شده و علاوه برآن به علت گسسته سازی فضای مسئله باعث از دست رفتن بعضی از داده‌های مساله و در نتیجه انتشار تعرفه غیر دقیق و نامنطبق بر نیاز مشتری می‌شود.

در [۱۶] یک ساختار یادگیری تقویتی جهت تعیین استراتژی پیشنهاد قیمت مالک‌های تک‌ژنراتوری و چندژنراتوری ارائه گردیده است. در ساختار پیشنهادی، عاملان بازار با یادگیری تقویتی، به آموزش مناسبی برای سطح تغییرات پیشنهاد قیمت خود دست یافتند به نحوی‌که هر تغییر رفتار دیگر آنها با توجه به نوع رفتار سایر عاملان، باعث کاهش سودشان می‌شد و این به معنای رسیدن به نقطه تعادل نش است. ایراد بزرگ روش‌های مبتنی بر تعادل نش این است که برای بدست آوردن کارایی بالا نیاز به اطلاعات کاملی از عامل‌ها (کارگزار) است، درصورتی‌که محیط بازار تعرفه محیطی با عامل‌های دارای نفع شخصی است که تمایلی به اشتراک دانش ندارند.

در [۱۷] عاملی معرفی شده است که سه هدف عمده را دنبال می‌کند: (۱) تخمین درست و کارای تقاضای انرژی مشتری‌ها، (۲) بدست آوردن انرژی مورد نیاز در بازار عمده‌فروشی مطابق با نیاز مشتری‌ها با کمترین قیمت و (۳) فروش انرژی خریداری شده به مشتری‌ها با یک قیمت مناسب به گونه‌ای که دارای سودمندی بالایی باشد و در عین حال مشتری‌ها را به خود جذب کند. این عامل برای برآورده‌سازی هدف اول یک از روش داده‌کاوی استفاده می‌کند که در آن مشتری‌های مختلف بر اساس الگوی مصرفشان خوشه‌بندی شده و سپس استراتژی یک روز جلوتر[۱۶] ، به صورت ساعتی میزان تقاضای مشتری‌هایی که به تعرفه‌شان پیوسته‌اند را تخمین می‌زند. همچنین برای هدف دوم از یک MDP برای تخمین انرژی مورد نیاز در بازار عمده‌فروشی استفاده می‌کند. برای هدف سوم نیز یک روش یادگیری تقویتی مستقل برای مشتری‌ها با الگوهای مصرف مختلف را به کار می‌گیرد که میزان قیمت انرژی را برای آن ها بهینه می‌سازد. در روش مذکور یادگیری با ناظر، بدون ناظر و تقویتی با یکدیگر ترکیب شده و یک مدل سیستماتیک ارائه شده است که با کارایی بالایی می تواند با پویایی شبکه‌های هوشمند خود را وفق دهد. این روش مانند روش‌های پیشین از گسسته سازی و ساخت دستی MDP استفاده می کند. همانطور که گفته شد این امر کارایی و دقت جواب‌ها را کاهش می‌دهد و اطلاعات مفیدی از داده‌ها هنگام گسسته‌سازی از بین می‌برد.

در پژوهش [۱۸] یک مدل‌سازی عامل محور رفتارتولیدکننده ناهگون را مشخص و زمینه تعامل و یادگیری را در یک محیط پویا فراهم می‌آورد. مکانیزم حراج و تسویه یک بازی تکرارشونده است و از نوع همکارانه فرض شده است زیرا عامل‌ها دارای تابع مشترک هستند، اما در زمینه سهم بازار رقابت می‌کنند. از آنجایی که مدل پیشنهاد شده تهی است به عبارت دیگر دارای حافظه نیست، با تکرار بازی‌های مستقل تصمیم گیری انجام می شود. در این مقاله با فرض ثابت بودن بازی موجودیت بهره‌بردار قیمت‌گذاری ساعتی در بازار را انجام می‌دهد. هدف این پژوهش بدست آوردن بهینه سراسری است به گونه‌ای که تقاضای مصرف با حداقل هزینه مصرف برآورده شود و انرژی ترکیبی از انواع انرژی باشد. نهاد بهره‌بردار دو وظیفه اجرای حراج و تسویه را انجام می دهد و برای اجرای این هدف از استراتژی یادگیرنده استفاده می‌کند که ترکیب استراتژی حریصانه و تصادفی است و اساس آن تضمین میزان درجه حرارت در معادله بولتزمان[۱۷] است.

در [۱۹] از الگوریتم استکلبرگ برای مدیریت انرژی در روز آینده بر اساس تئوری بازی استفاده شده است. در این مدل یک رهبر که شرکت توزیع کننده است وجود دارد و مشترکین و عامل‌ها پیشروهای این مسئله هستند. در این روش مشترکین با توجه به سیگنال قیمت، مصرف بار پاسخگوی خویش را زمان‌بندی می کنند. زمانی که الگوریتم (بازی) به تعادل می رسد رفاه اجتماعی (سود شرکت) بیشینه می

شود. در این روش شرکت توزیع برنامه کاهش ماکزیمم بار را از طریق انعقاد قرارداد تشویقی اجرا می‌کند. در این روش برای رسیدن به نقطه قویا یکنوا از الگوریتم‌های توزیع آسنکرون مبتنی بر بهترین پاسخ استفاده می شود. روش مدیریت انرژی در این پژوهش سه بخش دارد: ۱) تعیین قیمت روز پیش ۲) تعیین جریمه ۳) زمان‌بندی برای شارژ و تخلیه شارژ باطری و پاسخگویی به بار شبکه برق. این روش پروفایل عرضه شبکه را مسطح‌تر کرده و ماکزیمم عرضه را کاهش می‌دهد که در نهایت منجر به کاهش صورتحساب مشترکین می شود.

در مقاله [۲۰]، یک طرح تشویقی برای خرده فروشان در راستای ایجاد یک برنامه پاسخگویی موثر به تقاضا در طول اوج تقاضا با هدف به حداقل رساندن ریسک مالی، پیشنهاد شده است. در این الگوریتم یک قیمت انگیزشی مطلوب با توجه به شرایط بازار ارائه می شود تا مشتریان را به کاهش تقاضای برق خود در یک ساعت مشخص در ساعت های اوج تشویق کند. در این روند از یک رویکرد خطی استفاده شده است و قیمت مطلوب برای تشویق در نقطه ای که افت سود در حداقل سطح باشد تعیین می شود. در این الگوریتم میزان قیمت مطلوب به شدت به میزان مشارکت مشتریان حساس است.

پژوهش [۲۱] تکیه اصلی را به سمت تقاضا معطوف کرده است. رویکرد اصلی در این پژوهش اجرای برنامه پاسخگویی به بار با استفاده از روش‌های ریاضی است. در این روش نهاد تعیین بار مدل پیشنهاد شده از یک مدل دو سطحی استفاده می‌کند تا سود این نهاد و مصرف کننده به صورت همزمان بیشینه شوند. از طرف دیگر با روش KTT[۱۸] استفاده می‌کند تا مسئله دو سطحی را به تک‌سطحی تبدیل کرده و بتواند از روش ریاضی برای حل آن به جای روش ابتکاری استفاده کند. سطح بالا مربوط به نهاد تعیین بار و سطح پایین مرتبط با جمع کننده‌هاست. برای ارزیابی مدل از دو طرح استفاده می کند طرح اول با حضور مصرف‌کنندگان در برنامه پاسخ‌گویی بار است وطرح دوم بدون حضور آنان است که در آن تمامی بارها قیمت بازار خرده فروشی را می‌پردازند. نتایج نشان می دهد که می‌تواند پرداختی مصرف کنندگان را کاهش دهد. یکی از مشکلات مدل پیشنهاد شده این است که نهاد تعیین بار داری ارتباط یک‌طرفه با منابع تجدیدپذیر است و فقط توانایی خرید را داراست و در صورتی که بخشی از بار قطع شود، قیمت پرداختی بارهای غیر منعطف همان قیمت بازار خرده فروشی می شود.

همانطور که گفته شد، با توجه به مشکلاتی نظیر نفرین ابعاد و گسسته‌سازی داده مسئله که منجر به واریانس بالا در سیاست‌ها می‌شود ارائه الگوریتمی که بتواند به صورت پیوسته و خودکار محیط مسئله را مدل‌سازی کرده و از نفرین ابعاد نیز تا حدی زیادی دوری کند امری ضروری است. تکنیک یادگیری عمیق عملگر-نقاد یکی از روش‌های یادگیری تقویتی است که در آن عملگر تصمیم می گیرد که کدام عمل انتخاب شود و نقاد می گوید عمل انتخاب شده چقدر خوب است و چگونه باید تنظیم شود [۲۲].

در [۲۳] مروری شده است بر پژوهش‌هایی که از یادگیری عمیق تقویتی در حوزه‌های مختلف سیستم‌های قدرت استفاده کرده اند که از آن جمله می توان به حوزه‌های مدیریت انرژی، پاسخگویی به تقاضا، بازار برق و کنترل عملیاتی را نام برد. یکی از تحقیقاتی که در این مقاله مرور شده است در حوزه کنترل و بهینه کردن هزینه برای مشتریان بازار برق است که در [۲۴] آمده است. این مقاله یک استراتژی مطلوب مبتنی بر یادگیری عمیق تقویتی برای به حداقل رساندن هزینه برق و مشکلات مدیریت انرژی مسکونی بدون اطلاع از بار واقعی خانوار و قیمت برق ارائه داده است.

۱-۳- دستاورد و نوآوری

مدل مبتنی بر یادگیری عمیق تقویتی عملگر-نقاد با توجه به ترکیب سیاست و تابع ارزش می‌تواند کارایی قابل قبولی را در مساله مورد بررسی به‌دست دهد. دستاورد و نوآوری‌های روش پیشنهادی ما ارائه استراتژی قیمت‌گذاری کارآمدی برای توسعه کارگزاران در بازار خرده‌فروشی برق بر اساس گام‌های طراحی شده زیر است:

الف) انجام الگوریتم خوشه بندی متوالی مشتریان با استفاده از معیار فاصله زمانی پویا[۱۹] (DTW)

ب) استفاده از یادگیری تقویتی عملگر-نقاد که ترکیبی از روش‌های مبتنی بر سیاست و ارزش است.

ج) به کارگیری روش اصلاح پاداش برای افزایش سودآوری مشتریان

در ادامه چگونگی ارائه و انجام مراحل مذکور با جزییات شرح داده خواهند شد. نتایج به‌دست آمده از انجام آزمایش‌ها بهبود عملکرد را با به کارگرفتن این روش نشان می دهند.

۲- روش پیشنهادی

در این تحقیق یک استراتژی قیمت گذاری کارآمد برای توسعه کارگزاران در بازار خرده فروشی برق بر اساس یادگیری تقویتی عملگر-نقاد و خوشه‌بندی متوالی ارائه شده است. از یک سو، روش‌های مبتنی بر ارزش مانند یادگیری Q از همگرایی ضعیف رنج می برند، زیرا فقط در فضای ارزش کار می کنند و تغییر جزئی در تخمین ارزش می تواند فضای سیاست را کاملا تحت تاثیر قرار دهد. از سوی دیگر، روش‌های مبتنی برسیاست، مانند جستجوی سیاست و گرادیان، مستقیماً در فضای سیاست کار می‌کنند و منحنی‌های یادگیری نرم‌تری دارند. در روش‌های مبتنی بر سیاست حداکثرسازی تابع پاداش مورد انتظار با جستجو در فضای سیاست است. مشکل عمده این روش‌ها ضمانت بهبود عملکرد با هر بروزرسانی است (به دلیل تمایل به همگرا شدن در بهینه محلی). همچنین از واریانس بالای بین حالت‌های مختلف و ناکارآمدی در نمونه برداری رنج می‌برند. روش عملگر-نقاد از دو روش فوق الذکر نقاط قوت را استخراج کرده و در پیاده سازی الگوریتمش استفاده می کند: به کارگیری عملگر (به روزرسانی گرادیان سیاست) و

یک منتقد خوب (به عنوان مثال تابع ارزش). به کارگیری این تکنیک‌ها به عملگر-نقاد اجازه می دهد تا از طریق به روزرسانی تفاضل موقتی[۲۰] (TD) در هر مرحله از کارایی بیشتری برخوردار باشد. با توجه به سیاست فعال[۲۱] بودن این الگوریتم حتی در موارد غیرخطی نیز همگرایی تضمین شده است. علاوه‌براین، با توجه به اینکه در سیاست‌های پالایش شده جستجو می‌کند واریانس پایین‌تری در مقایسه با روش‌هایی که صرفا از روش‌های مبتنی بر سیاست یا روش‌های مبتنی بر تابع ارزش هستند دارد.

برای معرفی روش پیشنهادی ابتدا به تعریف مدل مسئله در بازار تعرفه برق و تعریف فرمول‌های مورد نیاز مسئله پرداخته می‌شود. روش پیشنهادی بر روی بازار تعرفه، عمده فروشی و بازار توزیع متمرکز شده است. همچنین، استراتژی قیمت گذاری کارگزار برای به حداکثر رساندن درآمد و تعادل عرضه و تقاضا بررسی شده است.

۲-۱- مدل سازی مسئله

مکانیسم استفاده کارگزاران به طور گسترده برای مشتریانی که علاقه‌مند به دستیابی به سیاست‌های بلندمدت به منظور کاهش هزینه‌ها یا کسب سود هستند در شبکه هوشمند استفاده می‌شود. یک کارگزار از یک سو با مشکلات مشتری ها مانند تعادل عرضه و تقاضا مواجه است و از سوی دیگر باید مسائل مربوط به رقابت با سایر کارگزاران برای به حداکثر رساندن سود خود را مد نظر داشته باشد. اجزای کلیدی بازار شبکه هوشمند به صورت زیر است:

۱) مجموعه مصرف کنندگان نیروی برق با C نشان داده می‌شود که در آن $C = [C_i, i = 1,2,..N]$ و هر C_i گروهی از مصرف‌کنندگان با الگوی برق مصرفی مشابه است. مصرف‌کنندگان هنگامی که قراردادهای تعرفه مربوطه را انتخاب می‌کنند، توسط کارگزاران پذیرفته می‌شوند.

۲) مجموعه تولیدکنندگان نیرو با $P = [P_i, i = 1,2,..M]$ نشان داده می‌شوند. هر P_i نماینده یک نوع از تولیدکنندگان از همان روش تولید است. تولیدکنندگان انرژی را از طریق تعرفه برق به کارگزاران می‌فروشند.

$$A = \begin{bmatrix} Maintain; Lower; Raise; \\ Revert; Inline; MinMax \end{bmatrix} \quad (۳)$$

۳) کارگزاران نیز با $B = [B_i, i = 1,2,..k]$ نشان داده می‌شوند. آنها واسطه بین مصرف‌کنندگان و تولیدکنندگان بوده و به دنبال دریافت سود در بازارهای برق هستند. کارگزاران شکاف بین مصرف و تولید را با بدست آوردن یا متوقف سازی تعهدات تولید جبران می‌کنند. مشتریان فعلی کارگزاران، منزلتشان[۲۲] (فرایند ترکیب کردن موجودی‌های مالی) را متشکل از مصرف‌کنندگان $\psi_{t,C}$ و تولیدکنندگان $\psi_{t,P}$ در بازه زمانی t تشکیل می‌دهند، که به صورت بلادرنگ توسط واحد توزیع اجرا شده است.

۴) عملگر سرویس O امکانات فیزیکی را برای شبکه منطقه‌ای و بهره برداری از شبکه الکتریکی در زمان واقعی اداره می‌کند. در آغاز هر ساعت، کارگزاران براساس وضعیت بازار تعرفه‌ها را منتشر می‌کنند. سپس مشتریان تعرفه‌ها را انتخاب می‌کنند و اپراتور خدمات تعهدات برق را طبق نمونه کارهای کارگزاران یعنی $\psi_t = \psi_{t,C} \cup \psi_{t,P}$ تحویل می‌دهد. در پایان ساعت فعلی، بازار تعرفه، سود و عدم توازن کارگزاران را محاسبه می‌کند.

چنین فرایندی را می‌توان به عنوان یک فرایند تصمیم گیری

$$M^{B_L} = \langle S, A, P, R \rangle \quad (۱)$$

مارکوف (MDP) مدل کرد [۲۵]. به طور عمومی، یک MDP برای کارگزار یادگیرنده تقویتی را که با BL نشان می دهیم می‌تواند به صورت رابطه (۱) تعریف شود:

در این رابطه، S مجموعه‌ای از وضعیت‌هاست و هر S_i، کارگزاران و نمایه سابقه[۲۳] اعمال مشتری در مراحل قبل را رمزگذاری می‌کند. A مجموعه‌ای از اقدامات است و هر a_i عملی است که قیمت کارگزار را در لیست زمانی بعدی تعیین می‌کند. همچنین، $P(s,a) \to s'$ یک تابع احتمال انتقال حالت از s به s' است هنگامی که یک عامل عمل a را اجرا می‌کند. $r \in R$ یک پاداش فوری است که سود کارگزاران دریافت شده در بازه زمانی فعلی را نشان می‌دهد. همچنین، $\Pi = S \to A$ استراتژی قیمت گذاری است که $\pi(S)$ مشخص می‌کند که کدام B_L باید حالت S را انتخاب کند. بازار تعرفه B_l منطقی است اگر:

$$P_{t,C}^{min} \geq P_{t,P}^{max} + \mu_L \quad (۲)$$

در رابطه (۲)، $P_{t,C}^{min}$ و $P_{t,P}^{max}$ به ترتیب حداقل قیمت‌های تعرفه مصرف‌کننده و حداکثر قیمت تعرفه تولیدکننده همه کارگزاران به جز B_L را نشان می‌دهند. μ_L سود حاشیه ذهنی[۲۴] است که B_L انتظار دارد. وضعیت منزلت، وضعیت تعادل عرضه و تقاضا را در نمونه کارهای B_L توصیف می‌کند و مقادیر آن بصورت مادون عرضه[۲۵]، مافوق عرضه[۲۶] و متعادل[۲۷] تعریف می‌شود. همچنین می‌توان وضعیت فعلی B_L را با دو ویژگی حداقل قیمت‌های تعرفه مصرف‌کننده و حداکثر قیمت تعرفه تولیدکننده تعریف کرد. مجموعه‌ای از عمل‌های A به شرح زیر است:

هر عامل نحوه تنظیم تعرفه فعلی خود را با قیمت زمانبندی بعدی تعیین می‌کند. محدوده قیمت در بازه [0.01,0.20] تعریف می شود که دامنه واقع بینانه‌ای از قیمت برق در بازار جهانی است. کمترین واحد قیمت 0.01 است [۲۶]. تعریف هرعمل به شرح زیر است:

- Maintain: انتشار همان قیمت‌های گذشته
- Lower: کاهش قیمت مصرف‌کننده و تولیدکننده به اندازه 0.01
- Raise: افزایش قیمت مصرف‌کننده و تولیدکننده به اندازه 0.01
- Revert: تنظیم قیمت به اندازه 0.01 به سمت میانه

$$m_t = [\frac{1}{2}(P_{t,C}^{max} + P_{t,P}^{min})] \quad (۴)$$

- Inline: تعیین قیمت جدید مصرف‌کننده و تولیدکننده بصورت:

$$P_{t+1}^{B_L}, C = \left[m_t + \frac{\mu_L}{2}\right] and\ P_{t+1}^{B_L}, P\ [m_t - \frac{\mu_L}{2}] \quad (۵)$$

- MinMax : تعیین قیمت جدید مصرف‌کننده و تولیدکننده بصورت:

$$P_{t+1}^{B_L}, C = P_{t,C}^{max}\ and\ P_{t+1}^{B_L} = P_{t,P}^{min} \quad (۶)$$

انتقالات $S \times A \rightarrow S$ توسط بازار تعرفه و پاداش کارگزاران B_k با رابطه (۷) محاسبه می‌شود:

$$r_t^{B_k} = P_{t,C}^{k}\psi_{t,C} - P_{t,P}^{k}\psi_{t,P} - \phi_t$$
$$\phi_t = \begin{cases} \phi - (\psi_{t,C} - \psi_{t,P}), if\ \psi_{t,C} \geq \psi_{t,P} \\ \phi + (\psi_{t,C} - \psi_{t,P}), if\ \psi_{t,C} \geq \psi_{t,P} \end{cases} \quad (۷)$$

در رابطه (۷)، $\psi_{t,C}$ و $\psi_{t,P}$ نمایانگر میزان مصرف و تولید فعلی مشتریان در سبد محصولات B_k است و Φ_t هزینه عدم تعادل B_k در زمان t است. علاوه براین اگر وضعیت منزلت فعلی مافوق عرضه باشد، برق اضافی با قیمت φ_+ به O فروخته می‌شود و اگر مادون عرضه باشد، این نیرو را از O با قیمت φ_- خریداری می‌کند.

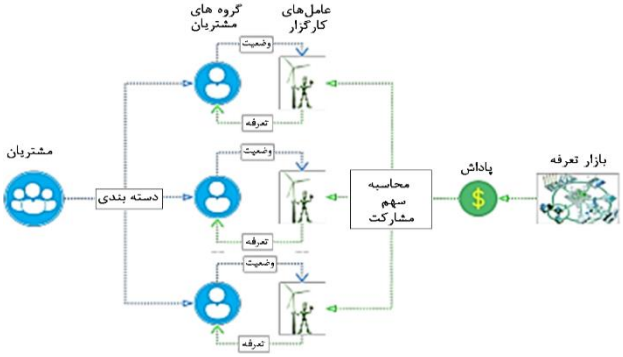
۲-۲- دسته بندی مشتریان

شکل (۲) رویکرد مبتنی بر استراتژی چندعامله در بازار برق را نشان می‌دهد. مشتریان با الگوهای مختلف مصرف برق در گروه‌های مختلفی قرار می‌گیرند و برای هر کدام استراتژی متناسب با الگوی مصرفشان اجرا می‌شود.

سیگنال‌های پیوسته خام از بازار خرده فروشی مانند قیمت تعرفه کارگزار می‌توانند به طور مستقیم یک حالت را بسازند به جای اینکه به صورت دستی از ویژگی‌های گسسته وضعیت محدوده قیمت و وضعیت منزلت تشکیل شوند. علاوه براین برای تعریف دقیق اطلاعات وضعیت، می‌توان در چندین دور از اطلاعات گذشته استفاده کرد. وضعیت یک نوع مشتری را می‌توان بصورت رابطه (۸) تعریف کرد:

$$S = < P_t, U_t, R_t \mid t = 1,2,...,T > \quad (۸)$$

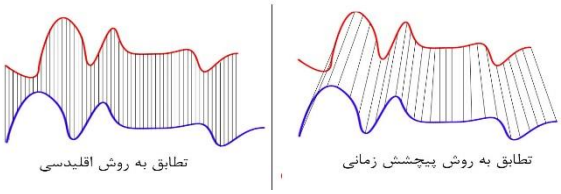
در این رابطه P_t مجموعه‌ای از قیمت‌های تعرفه کارگزاران برای این نوع از مشتری در شکاف زمانی t است. U_t مصرف برق متوسط در یک گروه از مشتریان در زمان t و R_t نرخ پذیرش به تعرفه این نوع از مشتریان است.



شکل (۲): رویکرد چندعامله بازار شبکه‌های هوشمند

انتشار تنها یک تعرفه برای همه مصرف‌کنندگان کافی نیست. به عنوان مثال، حتی اگر فقط مصرف‌کنندگان خانگی در بازار تعرفه در نظر گرفته شوند، به دلیل شیوه‌های مختلف زندگی و مصرف‌های مختلف، الگوهای مصرف برق آنها متفاوت است. بنابراین استفاده از چندین عامل برای انتشار تعرفه‌های مربوطه برای گروه‌های مختلف مصرف‌کننده می‌تواند تعادل عرضه و تقاضا را بهتر تسهیل کند. در روش پیشنهادی، مصرف‌کنندگان طبق الگوی مصرف برق آنها دسته‌بندی می‌شوند. در این روش با توجه به اینکه مصرف برق از نوع داده‌های سری زمانی است، کارگزاران الگوریتم K-Means را با معیار فاصله زمانی پویا (DTW) به کار می گیرند.

کارگزاران می‌توانند از روش‌های مختلف خوشه‌بندی استفاده کنند، اما با توجه به اینکه داده‌های مصرف برق از جنس سری زمانی هستند و داده‌ها طول متفاوتی دارند، به کار گرفتن روش DTW [۲۷] برای خوشه بندی داده‌ها مفید و کارا است. بنابراین کارگزاران با استفاده از معیار فاصله زمانی پویا در K-Means الگوریتم مناسبی برای اندازه‌گیری شباهت بین دو دنباله زمانی که طول متفاوتی دارند، در اختیار دارند. ایده اصلی DTW مقایسه دو سیگنال با طول متفاوت با ساختن نقاط تطابق چند به چند و یک به چند است به گونه‌ای که فاصله مجموع می‌تواند بین دو سیگنال کمینه شود. یک مثال از این موضوع در شکل (۳) آمده است.



شکل(۳) : مقایسه روش اقلیدسی با پیچشش زمانی [۲۷]

به عبارت دیگر K-Means منحنی‌های توالی را با توجه به شباهت آنها پیچ‌وتاب می‌دهد. سپس فاصله بین نقاط مربوطه را به ترتیب مطابقت بهینه و نه به ترتیب زمان محاسبه می‌کند. بعد از خوشه‌بندی،گروه‌هایی از کاربران بدست می آیند که الگوهای مصرف برق یکسانی دارند، حتی اگر گاهی اوقات مصرف آنها از لحاظ زمانی در یک بازه نباشد.

۲-۳- اصلاح پاداش

با توجه به گروه‌های خوشه بندی شده مشتریان، هر یک از آنها می‌توانند با توجه به روند انتشار تعرفه‌ها به یک کنترل یادگیری تقویتی مستقل اختصاص داده شوند. بنابراین کارگزار پیشنهادی به عنوان یک سیستم چندعامله، به جای ترکیبی از عامل‌های مستقل در نظر گرفته می‌شود. نکته اصلی نحوه محاسبه مقدار سهم هر کارگزار فرعی توسط r_t است. به این منظور در روش پیشنهادی، محاسبه می کنیم که اگر مشتریان خاصی که توسط کارگزار فرعی i اداره می شوند را لحاظ

نکنیم چه میزان ضرر ایجاد می‌شود. در رابطه (۹) سهم هر زیرکارگزار در نظر گرفته می‌شود:

$$r_t^i = r_t - (\sum_{j \neq i} P_t^j \Psi_{t,C}^j - \sum_{k \neq i} P_t^k \Psi_{t,P}^k - \phi_t^i), j \in C, K \in P \quad (۹)$$

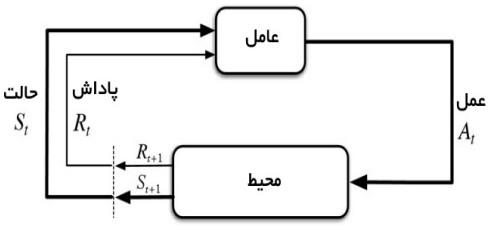
در رابطه فوق، i نوعی مشتری است که توسط کارگزار فرعی i مربوطه شارژ شده است (به آن مشتری مقدرای برق داده شده است). r_t نیز توسط رابطه (۹) محاسبه می‌شود. $\Psi_{t,C}^j$ مصرف کلی از مصرف‌کنندگان j در زمان t را نشان می‌دهد. $\Psi_{t,P}^k$ بیانگر کل خروجی تولیدکنندگان k در زمان t است. همچنین، P_t^j قیمت تعرفه جاری کارگزاران برای C_j و P_t^k قیمت تعرفه جاری برای P_k است. ϕ_t^i هزینه عدم تعادل فعلی است که با رابطه (۱۰) محاسبه می شود:

$$\phi_t = \begin{cases} \phi - \left(\sum_{i \neq j} \psi_{t,C} - \sum_{k \neq i} \psi_{t,P} \right), if \sum_{i \neq j} \psi_{t,C} \geq \sum_{k \neq i} \psi_{t,P} \\ \phi + \left(\sum_{i \neq j} \psi_{t,C} - \sum_{k \neq i} \psi_{t,P} \right), \text{otherwise} \end{cases} \quad (۱۰)$$

با اصلاح شکل پاداش r_t^i برای هر کارگزار فرعی i، آنها سیاستهای خود را با متناسب با سهم مشارکت خود در بازار تعرفه به روز می‌کنند. بنابراین، اگر یک کارگزار فرعی عمل بدی را انتخاب کند، اما در مقابل پاداش کلی افزایش یابد،کارگزار فرعی از انتخاب این عمل با سهم مشارکت منفی، جلوگیری می کند.

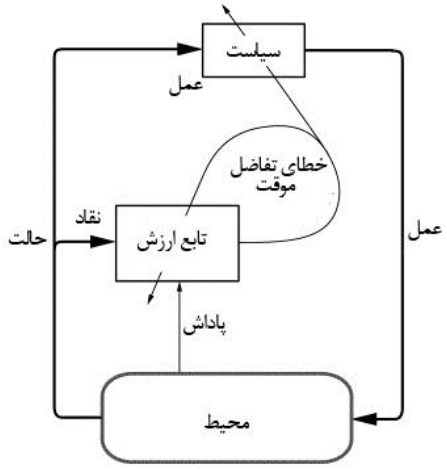
۲-۴- الگوریتم یادگیری

یادگیری تقویتی گونه‌ای از روش‌های یادگیری ماشین است که یک عامل را قادر به یادگیری در محیطی تعاملی با استفاده از آزمون و خطاها و استفاده از بازخوردهای اعمال و تجربیات خود می‌سازد [۲۵]. یادگیری تقویتی در مقایسه با یادگیری غیرنظارتی دارای اهداف متفاوتی است. در حالیکه هدف در یادگیری غیرنظارتی پیدا کردن مشابهت‌ها و تفاوت‌های بین نقاط داده محسوب می‌شود. در یادگیری تقویتی هدف پیدا کردن مدل داده مناسبی است که پاداش تجمعی کل را برای عامل بیشینه می‌کند. شکل (۴) ایده اساسی و عناصر درگیر در یک مدل یادگیری تقویتی را نشان می‌دهد.



شکل (۴): مدل یادگیری تقویتی

در این روش یادگیری ساختار حافظه جداگانه‌ای هم برای سیاست و هم برای تابع ارزش در نظر گرفته می‌شود. از آنجاییکه معماری عملگر-نقاد از اصول یادگیری تقویتی تفاضل موقتی استفاده می‌نماید، قابلیت پیاده سازی به صورت زمان حقیقی در طی مسیر سیستم را دارا است [۲۲]. در این معماری ساختار سیاست به عنوان عملگر شناخته می‌شود زیرا از آن برای تولید عمل استفاده می‌شود و ساختار تابع ارزش به عنوان نقاد شناخته می‌شود زیرا برای نقد اعمال انجام گرفته توسط عملگر به کار گرفته می‌شود. یادگیری در معماری عملگر-نقاد به صورت سیاست فعال است به این معنی که نقاد باید در سیاستی که توسط عملگر دنبال می‌شود یادگیری را به طور همزمان انجام دهد. در طول یادگیری در هر گام زمانی نقاد یک خطای تفاضل موقت را تولید کرده و براساس آن یادگیری در عملگر و نقاد انجام می-شود. این امر در شکل (۵) آمده است.



شکل (۵): مدل یادگیری تقویتی عملگر-نقاد

بعد از اجرای هر عمل، حالت جدید محیط توسط نقاد با رابطه (۱۱) ارزیابی شده و تعیین می‌شود که آیا حالت محیط بهتر شده یا خیر؟

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \quad (۱۱)$$

در صورت مثبت بودن δ_t تمایل برای انتخاب عمل δ_t انجام شده و باید تقویت شود. در صورت منفی بودن تمایل برای انتخاب عمل انجام شده باید کاهش یابد. در این تحقیق برای افزایش سرعت یادگیری روش اثر شایستگی[۲۸] برای به روزرسانی ارزش حالت‌های مختلف محیط بکار گرفته شد. در روش اثر شایستگی ارزش حالت‌های مختلف محیط توسط رابطه های (۱۲) و (۱۳) به روز می‌شوند:

$$V(s_t) \leftarrow V(s_t) + \alpha\delta_t e_e(s_t), \quad 0<\alpha<1 \tag{۱۲}$$

$$et(s_t) = \begin{cases} \gamma\lambda e_{t-1}(s), if \ \ s \neq s_t \\ \gamma\lambda e_{t-1}(s)+1, if \ \ s = s_t \end{cases} \quad 0 \le \gamma, \lambda \le 1 \tag{۱۳}$$

در رابطه (۱۲) و (۱۳)، α نرخ یادگیری، γ نرخ تخفیف، λ میزان تاثیرپذیری ارزش حالت‌های ابتدایی اپیزود از ارزش حالت‌ها و سیگنال‌های انتهایی محیط است. برای $\lambda = 0$ فقط یک حالت از محیط در گام زمانی t مقدار شایستگی غیر صفر دارد و بنابراین فقط ارزش آن حالت به روز می شود. برای λ مثبت، عامل باید در هر گام زمانی پیش بینی‌ها و آثار شایستگی را برای تمام حالات به روز نماید و به همین دلیل پیاده سازی از نظر محاسباتی برای $\lambda > 0$ سنگین‌تر از $\lambda = 0$ است؛ مخصوصاً در مواقعی که فضای حالت بزرگ باشد. به هر حال، استفاده از λ مثبت سرعت یادگیری را به طور قابل ملاحظه‌ای افزایش می‌دهد. در این پژوهش، پس از انجام آزمایش‌های مختلف مقدار α برابر ۰,۲، مقدار γ برابر ۰,۹۰ **و** مقدار λ برابر ۰,۸۵ انتخاب شده اند. با توجه به اینکه سرعت همگرایی در مسائل بلادرنگ از اهمیت بالایی برخوردار است، تنظیم نرخ یادگیری به نحوه صحیح و درست باعث افزایش همگرایی می‌شود. از اینرو با توجه به اینکه نوع داده‌ها سری زمانی است، انتخاب نرخ یادگیری پایین دقت تخمین و سرعت همگرایی را بهبود می‌بخشد.

احتمال انجام اعمال مختلف توسط سیاست $\varepsilon - greedy$ محاسبه می‌شود که ε نشان دهنده میزان تمایل عامل برای کاوش ارزش اعمال مختلف در حالت‌های مختلف محیط است. هرچه میزان ε به یک نزدیک‌تر باشد، سیاست عامل به انتخاب تصادفی نزدیک‌تر و تمایل عامل به کاوش اعمال مختلف افزایش می‌یابد. هرچه میزان ε به صفر نزدیک‌تر باشد، سیاست عامل به انتخاب حریصانه نزدیک‌تر و تمایل عامل به کاوش کاهش می‌یابد:

$$\pi_t(s,a) = pr\{a_t = a | s_t = s\} = \begin{cases} 1-\varepsilon+\frac{\varepsilon}{|A_s|}, \text{ if a = argmax}_{a' \in A_s} P(s,a') \\ \frac{\varepsilon}{|A_s|}, \quad \text{else} \end{cases} \tag{۱۴}$$

$$0 \le \varepsilon \le 1$$

در حالت S و عمل a با احتمال ε -1 محیط جدید دقیقا مانند محیط قبلی رفتار می‌کند. با احتمال ε از میان اعمالی که احتمال برابری دارند به صورت تصادفی یکی را انتخاب می کند.

$$p(s_t, a_t) \leftarrow P(s_t, a_t) + \beta\delta_t \tag{۱۵}$$

در رابطه (۱۵) مقادیر پارامترهای سیاست در عملگر هستند که در طول یادگیری تغییر می‌کنند و نشان دهنده تمایل برای انتخاب هر عمل که a در حالت محیط s است. تقویت و یا تضعیف تمایل برای انتخاب هر عمل می‌تواند توسط افزایش یا کاهش $p(s_{t,a}t)$ در زمان‌های مختلف انجام شود. در رابطه (۱۵) ، β پارامتر طول گام بوده و دارای یک مقدار مثبت است. مقدار β در این پژوهش برابر ۹۰۰ انتخاب‌شده است که نسبت به سایر مقادیر نتیجه بهتری را در طول آزمایش‌های ما به همراه داشت.

در الگوریتم عملگر-نقاد :

۱- پارامترهای سیاست θ و عملگر را مقدار دهی اولیه می کنیم
۲- برای هر تکرار آموزش تحت سیاست فعلی، M مسیر (دنباله‌ای از سیاست‌ها) را نمونه برداری می‌کنیم.
۳- گرادیان سیاست را حساب می‌کنیم.
۴- برای هر مسیر تابع سودمندی را محاسبه نموده و برای تخمین زننده گرادیان از آن استفاده می‌کنیم.
۵- در انتها همه تخمین‌های انجام شده را جمع می‌کنیم.
۶- سپس پارامترهای نقاد ϕ را در روشی مشابه آنچه در بالا گفته شد آموزش می‌دهیم.

در این الگوریتم ما به صورت پایه‌ای در تلاش هستیم تا تابع سودمندی را تقویت کنیم. تلاش عمده الگوریتم قرار دادن تابع سودمندی در نقطه مینیمم است. الگوریتم پیشنهادی درشکل (۶) قابل مشاهده است.

Actor-Critic Algorithm

Initialize policy parameters θ, critic parameters ϕ
For iteration= 1,2,... do
  Sample m trajectiries under the current policy
  $\Delta\theta \leftarrow 0$
  for i= 1,...,m do
    for t= 1,...T do
      $A_t = \sum_{t' \ge t} \gamma^{t'-t_{r_t^i}} - V_\phi(s_t^i)$
      $\Delta\theta \leftarrow \Delta\theta + A_t \nabla_\theta \log\left(a_t^i \middle| s_t^i\right)$
  $\Delta\phi \leftarrow \sum_i \sum_t \nabla_\phi \left\| A_t^i \right\|^2$
  $\theta \leftarrow \alpha\Delta\theta$
  $\phi \leftarrow \beta\Delta\phi$
End For

شکل (۶): الگوریتم پیشنهادی

همانگونه که مشخص است ما در تلاشیم تا تابع سودمندی را به گونه‌ای تنظیم کنیم تا به معادله بلمن[۲۹]، که همان دست یافتن به حالت بهینه دارای بیشترین پاداش دریافتی است، نزدیک‌تر شویم. یادگیری و بهینه‌سازی بین تابع سیاست و همچنین تابع نقاد را به طور پیوسته تکرار می‌کنیم و در هر تکرار گرادیان ها را به روز می کنیم. این الگوریتم به دلیل اینکه از ترکیب سیاست و تابع ارزش استفاده می‌کند نسبت به سایر الگوریتم‌های یادگیری تقویتی که یا فقط از تابع

ارزش یا تابع سیاست استفاده می‌کنند نتایج بهتری را به دست می‌دهد.

۳- آزمایش ها, نتایج و تحلیل

یکی از مجموعه داده‌های معروف برای بازار تعرفه پایگاه داده شبکه هوشمند انرژی می‌توان به داده‌های شرکت UCI [۲۸] اشاره کرد که به صورت دقیقه‌ای ثبت می‌گردند ولی به علت اینکه روی یک مصرف‌کننده توجه دارند مناسب بستر سیستم‌های چندعامله نیستند. یکی دیگر از مجموعه داده‌هایی که به علت جامعیت آن در پژوهش‌های مختلفی مورد استفاده قرار گرفته‌اند، داده‌های منتشر شده توسط کمپانی انگلیسی Power Networks است[۲۹]. این مجموعه از داده گستره متنوعی از مصرف مشتریان برق خانگی را شامل می‌شود که در بازه‌های نیم ساعته ثبت شده‌اند. این مجموعه شامل یک میلیون داده است که از سال ۲۰۱۱ تا ۲۰۱۴ ذخیره شده‌اند و هر رکورد دارای ۴ ویژگی است. در این مقاله نیز از داده‌های این مجموعه استفاده شده است. ویژگی‌های پایگاه داده مذکور در جدول (۱) آمده است. برای اینکه تقریب خوبی از مجموع کل داده‌های این مجموعه برای انجام آزمایش‌ها در این مقاله داشته باشیم، به صورت تصادفی دو هزار نمونه را استخراج کرده‌ایم. با توجه به تصادفی بودن و پویایی انتخاب نمونه‌ها، می‌توان این تعداد نمونه را تقریب مناسبی از توزیع داده‌های اصلی در همه مصرف‌کنندگان دانست. از این رو می‌توان گفت داده‌های شبیه‌سازی شده به اندازه کافی برای بازنمایی داده‌های دنیای واقعی دقیق هستند.

جدول(۱): ویژگی‌های پایگاه داده

ردیف	ویژگی هر نمونه	نرخ محاسبه آن
۱	مصرف انرژی	کیلووات بر ساعت
۲	تاریخ	تاریخ و زمان ثبت
۳	زمان	ساعت
۴	ID مشتری	-----

به منظور ارزیابی روش پیشنهادی، ابتدا به بررسی و مقایسه عملکرد یادگیری تقویتی براساس Q-learning و عملگر-نقاد پرداخته ایم و پس از آن، تاثیر روش پیشنهادی بر سود حاصله از بازار شبکه‌های هوشمند در طول زمان را بررسی کرده ایم. در انجام آزمایش‌ها، نتیجه الگوریتم یادگیرنده عملگر-نقاد با مدل‌های یادگیر DQN و همچنین با چهار رقیب پایه‌ای که در جدول (۲) آمده‌اند مقایسه کرده‌ایم. الگوریتم‌های این جدول به این دلیل برای مقایسه انتخاب می‌شوند که ضرورت استفاده از یک مدل یادگیر در بازار تعرفه را نشان دهند.

جدول(۲): رقیب های پایه ای مسئله

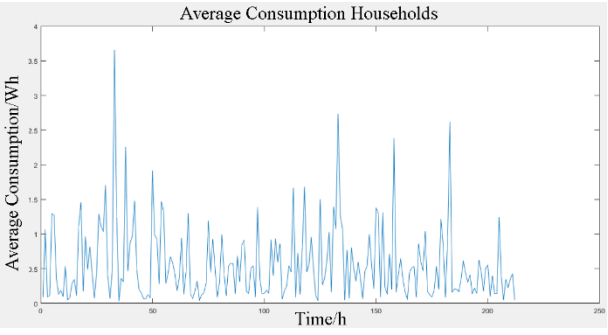
ردیف	عنوان رقیب	استراتژی
۱	کارگزار متعادل شده	با اجرای عملیات‌های بالا بردن و کاهش دادن سعی در ایجاد تعادل بین عرضه و تقاضا را دارد
۲	کارگزار حریص	اجرا کردن عملیات کمینه بیشنه زمانی که وضعیت نرخ قیمت منطقی است و اجرا کردن در خط روی قیمت وقتی که وضعیت نرخ قیمت معکوس شده است
۳	کارگزار تصادفی	انتخاب عملیات به صورت تصادفی انجام می‌شود
۴	کارگزار ثابت	همیشه عملیات نگهداری را انجام می‌دهد

آزمایش‌های انجام شده در این پژوهش روی یک سیستم با مشخصات RAM 4 GB و CPU پنج هسته‌ای و در محیط برنامه نویسی MATLAB 2018 انجام شده است. جدول (۳) مشخصات پارامترهای مساله را نشان می دهد.

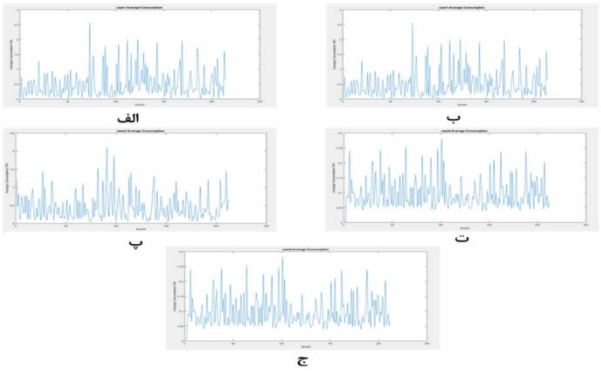
جدول (۳) : پارامترهای مسئله

پارامترهای مسئله	مقدار
β	900
α	0.2
γ	0.9
Λ	0.85

نمودار شکل (۷) میانگین برق مصرفی کل مشتریان و نمودار شکل (۸) میانگین مصرف برق را در ۵ خوشه حاصله نشان می‌دهد. با خوشه‌بندی مصرف مشتریان ما می‌توانیم توزیع جمعیت مشتریان را به دست آوریم که در زمینه پیش بینی مصرف در ساعت آتی به ما کمک می‌کند.



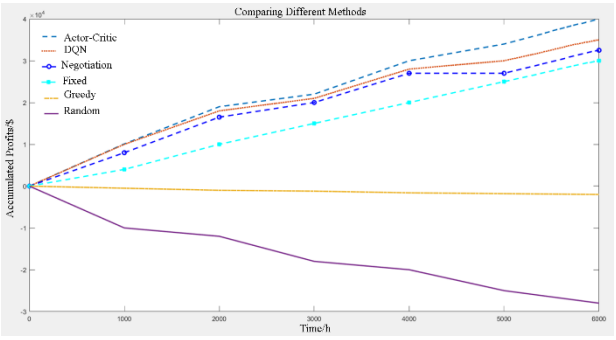
شکل (۷) : میانگین مصرف انرژی در کل داده‌های نمونه



شکل (۸): میانگین مصرف انرژی در ۵ خوشه بدست آمده

خوشه‌بندی از دو جهت حائز اهمیت است: اول اینکه همانطور که قبلا اشاره شد به علت تنوع الگوی مصرف مشتریان نمی‌توان برای همه آنها تعرفه ثابتی انتشار داد و ثانیا با خوشه‌بندی می‌توان پیک مصرف انرژی هر خوشه را شناسایی کرده و در آن زمان تعرفه گران‌تری صادر کرد که هم باعث اصلاح الگوی مصرف و هم باعث سودآوری کارگزار خواهد شد. در این مقاله، به مقایسه سود خالص روش پیشنهادی با سایر روش‌های DQN، Negotiation، Fixed (ثابت)، حریصانه و تصادفی پرداخته شده است.

روش DQN به علت اینکه از یک یادگیری عمیق تقویتی در ساختار خود استفاده می کند [۲۶]. روش یادگیری مبتنی بر مذاکره [۳۰] به دلیل اینکه در مقالات به عنوان یک معیار[۳۰] مطرح شده است در این مقاله با روش پیشنهادی ما مورد مقایسه قرار گرفته است. همانطور که درشکل (۹) نشان داده شده، روش پیشنهادی نسبت به سایر روش‌ها میزان سوددهی بیشتری دارد. روش‌های DQN، Negotiation و Fixed به ترتیب در رده‌های بعدی قرار دارند. روش حریصانه متمایل به مقادیر منفی است. روش تصادفی نیز کاملاً به سمت پایین و منفی متمایل است. روش تصادفی به دلیل اینکه عمل‌ها را بدون در نظر گرفتن تاثیر آنها در محیط به صورت شانسی انتخاب می‌کند در بین رقبا تقریبا شانسی نزدیک به صفر برای ارائه تعرفه جذاب دارد. دلیل اینکه ما از روش تصادفی برای مقایسه استفاده می‌کنیم این است که نشان دهیم بررسی حالت محیط بعد از هر عمل و بررسی بازخورد آن عمل روی محیط‌هایی که داری نفع شخصی بدون اشتراک دانش هستند اهمیت به سزایی دارد.



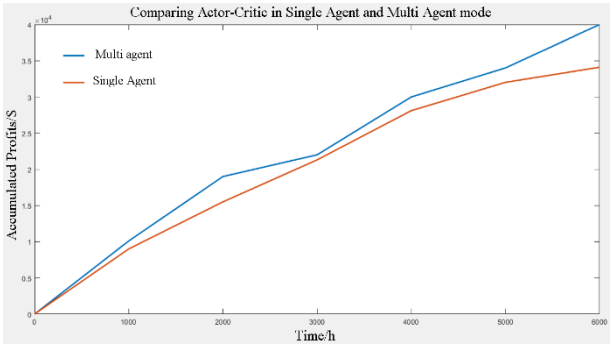
شکل (۹) : میزان سود روش‌های مختلف

روش پیشنهادی ما، با توجه به ترکیب سیاست و تابع ارزش می‌تواند با محیط انطباق پیدا کرده و استراتژی کارا و یادگیری بهتری از مدل یادگیرنده DQN که برای یادگیری استراتژی از یک شبکه عصبی عمیق و تابع ارزش Q و نیز مدل یادگیری Negotiation که برای یادگیری استراتژی از مذاکره بین عامل‌ها استفاده می کند داشته باشد.

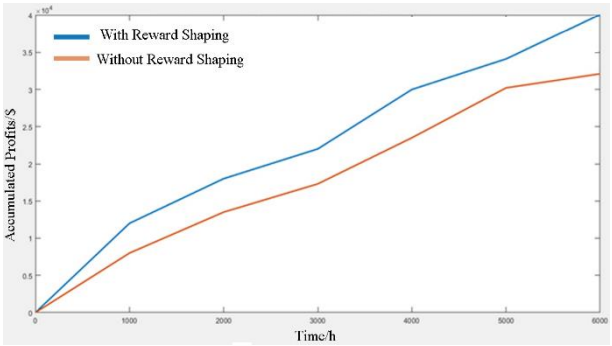
جدول (۴) : مقایسه اعمال انجام شده برای متعادل سازی بازار و میانگین خالص سود حاصله

روش‌ها	مافوق‌عرضه (kWh)	مادون‌عرضه (kWh)	میزان عدم‌تعادل	سود
حریصانه	۱۴۸۴۸۸	-۴۳۰۲۴۲	۵۷۸۷۳۰	-۱۳۰۹۵۰
تصادفی	۶۱۷۴۳۶	-۴۰۰۱۷۲	۱۰۱۷۶۰۸	-۲۴۶۵۶۲
Negotiation	۲۴۵۵۶۴	-۲۵۶۸۲۶	۵۰۲۳۹۰	۲۱۲۱۸۲
DQN	۲۷۵۵۰۳	-۲۲۶۷۰۱	۵۰۲۲۰۴	۲۷۲۱۸۲
روش پیشنهادی	۲۹۴۶۸۴	-۲۰۵۵۹۳	۵۰۰۲۷۷	۳۰۵۴۷۸

جدول (۴) سود خالص را در روش‌های مختلف مقایسه کرده است. همانطور که مشاهده می شود روش پیشنهادی ما سود بیشتری نسبت به سایر روش‌ها به دست آورده است و همچنین مقدار عدم‌تعادل کمتری را نیز داراست. شکل (۱۰) نتایج انجام آزمایش‌ها برای محاسبه سود خالص را در دو وضعیت تک‌عامله و چندعامله نشان می‌دهد. همانطور که در شکل (۱۰) نشان داده شده است روش تک‌عامله با وجود اینکه عملکرد آن به سمت سود آوری مثبت گرایش دارد اما نسبت به روش چندعامله عملکرد ضعیف‌تری دارد. زیرا با توجه به پیچیدگی و پویایی محیط مسئله استفاده از بستری مبتنی بر سیستم چندعامله با عامل‌های خودمختاری که ۲۴ ساعته و هفت روز هفته در فعالیت هستند ضروری است. در بستر شبیه‌ساز چندعامله ما از ۷ عامل استفاده کرده ایم که بر اساس نتایج ما بهترین حالت از لحاظ سودآوری نسبت به زمان اجرای الگوریتم است.

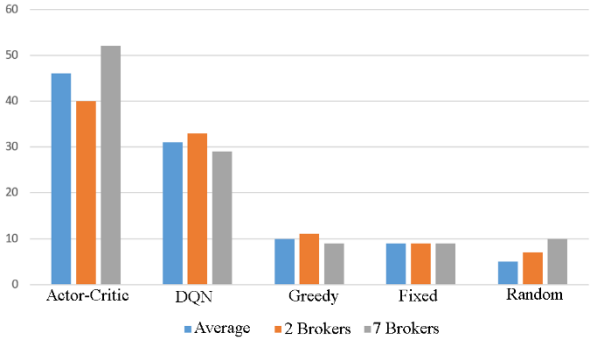


شکل (۱۰) : میزان سود روش‌های تک عامله و چندعامله



شکل (۱۱) : میزان سود با اعمال تابع اصلاح پاداش

همچنین اصلاح پاداش نسبت به روشی که شکل پاداش تغییر نمی کند توانایی بهتری در یادگیری استراتژی بهینه دارد که این موضوع در شکل (۱۱) نشان داده است. برای اصلاح پاداش از رابطه (۹) استفاده می‌کنیم. اصلاح پاداش برای ساختار چندعامله بررسی شده است. نمودار آبی نشان دهنده روند سودآوری الگوریتم بعد از اصلاح پاداش است. شکل (۱۲) نشان دهنده میزان تعرفه‌های پذیرش شده توسط مشتریان است.



شکل (۱۲) : درصد عضویت کاربران به تعرفه های ارائه شده

همان گونه که در شکل (۱۲) نشان داده شده است روش پیشنهادی ما، تعرفه هایی را که ارائه داده است که برای مشتری‌ها جذابیت بیشتری داشته و آن را پذیرفته اند. همچنین شکل (۱۲) نشان می‌دهد که روش پیشنهادی ما در حضور تعداد عامل‌های بیشتر یعنی زمانی که پیچیدگی محیط بالا می رود و عامل‌های رقابت کننده بیشتری در محیط هستند به عبارت دیگر یعنی زمانی که به محیط واقعی نزدیک‌تر هستیم عملکرد بهتری دارد و تعداد بیشتری از مشتری‌ها را به تعرفه‌های صادر شده خود جذب می کند.

با توجه به مفروضات مدل ارائه شده، فرایند تعیین قیمت را می‌توان یک بازی ایستا فرض نمود که هر ساعت تکرار می شود. در این بازی هر عامل قیمت پیشنهادی خود را به‌طور مستقل اعلام نموده و دلال با توجه به تقاضای برق مصرفی، بهترین پیشنهادها را انتخاب می نماید. در این میان، با اینکه فرایند یادگیری سبب تغییر در استراتژی تصمیم سازی عامل می گردد، اما روی مکانیزم بازار (محیط) که مستقل از عامل است، تأثیری ازلحاظ پویایی ندارد. لذا ایستا و تکراری بودن صرفاً مربوط به محیط و مکانیزم حراج و خرید و فروش است. لازم به ذکر است که یادگیری سبب پویائی رفتار عامل می‌شود، اما مکانیزم های اجرائی بازار را تغییر نمی دهد. بنابراین با لحاظ این مفروضات و قیود، این مدل می تواند به شکل بازی همکارانه توصیف شود، که عامل ها در تابع پاداش مشترک هستند، اما در زمینه‌ی سهم بازار رقابت دارند. دربازی همکارانه، عامل ها به دنبال سود بهینه همه‌جانبه هستند و برای نیل به آن برخی قواعد باید لحاظ شود. قوانین به‌نوعی است که به‌جای صرفاً بیشینه نمودن سود هر عامل، یک تابع هدف تعریف می‌شود که سود مجموعه را نیز تأمین نماید.

۴- جمع بندی و پیشنهاد کارهای آتی

هدف این پژوهش بررسی تاثیر یادگیری تقویتی توزیع شده در عملکرد عامل‌های بازار برق جهانی است. مدل نرم افزاری بازار برق جهانی که یک مدل چندعامله است در پژوهشگاه‌ها سال‌هاست به صورت نرم‌افزار اصلی شبیه‌سازی و هدایت بازار مورد استفاده قرارگرفته است. مشخصه‌های این نرم‌افزار براساس داده‌های واقعی تنظیم شده و نتایج شبیه سازی نزدیک به عملکرد واقعی بازار برق جهانی است. در مساله بررسی شده، هرعامل قیمت پیشنهادی خود را به طور مستقل اعلام نموده و دلال با توجه به تقاضای برق مصرفی، بهترین پیشنهادها را انتخاب می‌نماید. در این روش یک مدل یادگیری تقویتی عملگر-نقاد برای تعرفه‌گذاری قیمت در بازار شبکه‌های هوشمند برای متعادل‌سازی عرضه و تقاضا در بستر سیستم‌های چندعامله ارائه شد. در روش ما سود حاصل برای عامل کارگزار نسبت به سایر روش‌های مقایسه شده افزایش داشته و تعداد تعرفه‌های بیشتری برای مشتریان جذاب بوده است که مشتریان آنها را پذیرفته اند.

در این بین موضوع مشتریان فرصت طلب مسئله مهمی است که در پژوهش های مرتبط مورد غفلت قرار گرفته است. این مشتریان می‌توانند از سود حاصل از قیمت گذاری در لحظه بسیار سود ببرند. در ادامه این پژوهش قصد ما بر این است که با تعریف یک مکانیزم پاداش و هدیه برای تغییر تعرفه و کاهش مصرف در پیک انرژی باعث شویم تا فقط مشتری‌هایی که مصرف برق خود را کاهش داده‌اند از آن بهره‌مند شوند. ارائه این مدل و همچنین استفاده از بستر کنتورهای دیجیتال باعث می شود مشتریان فرصت طلب با احتمال بالاتری شناسایی و جریمه شده و انگیزه آنها برای استفاده مجانی کمتر شود.

مراجع

[1] Sharifi, R., Fathi, S.H., Anvari-Moghaddam, A., Guerrero, J.M. and Vahidinasab, V., 2018, February. An economic customer-oriented demand response model in electricity markets. In 2018 IEEE International Conference on Industrial Technology (ICIT) (pp. 1149-1153). IEEE.

[2] Babic, J. and V. Podobnik. Adaptive bidding for electricity wholesale markets in a smart grid. in AAMAS Workshop on Agent-Mediated Electronic Commerce and Trading Agents Design and Analysis (AMEC/TADA 2014). 2014.

[3] Sharifi, R., Anvari-Moghaddam, A., Fathi, S.H., Guerrero, J.M. and Vahidinasab, V., 2019. An optimal market-oriented demand response model for price-responsive residential consumers. Energy Efficiency, 12(3), pp.803-815.

[4] Ketter, W., J. Collins, and M.d. Weerdt, The 2018 power trading agent competition. ERIM Report Series Reference, 2017.

[5] Bayram, I.S., Shakir, M.Z., Abdallah, M. and Qaraqe, K., 2014, December. A survey on energy trading in smart grid. In 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp. 258-262). IEEE.

[6] Shabanzadeh M, Sheikh-El-Eslami M, Haghifam M. Profitability Analysis of Coalition of Virtual Power Plants and Load Aggregators in Active Distribution Networks . Journal of Iranian Association of Electrical and Electronics Engineers. 2018; 15 (3) :45-58

[7] Sharifi, R., Anvari-Moghaddam, A., Fathi, S.H. and Vahidinasab, V., 2019. A flexible responsive load economic model for industrial demands. Processes, 7(3), p.147.

[8] Nosratpoor H, Zanganeh A. Optimal Self-healing of Smart Distribution Grids Based on Spanning Trees to Improve System Reliability. Journal of Iranian Association of Electrical and Electronics Engineers. 2019; 16 (1) :91-101

[9] Ramchurn, S.D., Vytelingum, P., Rogers, A. and Jennings, N.R. Putting the'smarts' into the smart grid: a grand challenge for artificial intelligence. Communications of the ACM, 55(4), pp.86-97, 2012.

[10] Gomes, C.P., Computational sustainability: Computational methods for a sustainable environment, economy, and society. The Bridge, 2009. 39(4): p. 5-13.

[11] Stavrogiannis, L.C. and Mitkas, P.A. Evaluation of Market Design Agents: The Mertacor Perspective. In Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets (pp. 211-225). Springer, Berlin, Heidelberg, 2009.

[12] Urieli, D. and P. Stone. Autonomous electricity trading using time-of-use tariffs in a competitive market. in Thirtieth AAAI Conference on Artificial Intelligence. 2016.

[13] Liefers, B., J. Hoogland, and H. La Poutré, A successful broker agent for power tac, in Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets. 2014, Springer. p. 99-113.

[14] Kuate, R.T., et al. An intelligent broker agent for energy trading: an MDP approach. in Twenty-Third International Joint Conference on Artificial Intelligence. 2013.

[15] Kim, B.-G., Zhang, Y., van der Schaar, M., & Lee, J.-W. Dynamic Pricing and Energy Consumption Scheduling With Reinforcement Learning. IEEE Transactions on Smart Grid, 7(5), 2187–2198, 2016.

[۱۶] رمضانیان لنگرودی، م.، میرحسینی مقدم، س. م.، علیزاده، ب.، استفاده از روش یادگیری رقابتی برای قیمت‌دهی استراتژیک شرکت‌های تولید بر اساس LMP در بازار برق. مجله مهندسی برق دانشگاه تبریز، دوره ۴۷، شماره ۲، صفحه ۵۳۷-۵۴۹، تابستان ۱۳۹۶.

[17] Wang, X., M. Zhang, and F. Ren, A hybrid-learning based broker model for strategic power trading in smart grid markets. Knowledge-Based Systems, 2017. 119: p. 142-151.

[۱۸] اصغری اسکوئی، م. ر.، فلاحی، ف.، دوستی‌زاده، م.، مشیری، س.، کاربرد یادگیری تقویتی در یک مدل‌سازی عامل‌محور برای بازار عمده‌فروشی برق ایران، پژوهشنامه اقتصاد انرژی ایران، سال ۷، شماره ۲۵، صفحه ۱-۴۰، زمستان ۱۳۹۶.

[19] Shojaei A, Moallem M, Manshaei M H. Social Optimal Energy Management with the Presence of the battery Storages in Smart Grid. Journal of Iranian Association of Electrical and Electronics Engineers. 2020; 17 (2) :123-133

[20] Teimourzadeh Baboli P. Designing Incentive-based Demand Response Program for Minimizing Financial Risk of Retailer during Peak Period. Journal of Iranian Association of Electrical and Electronics Engineers. 2019; 15 (4) :93-102.

[21] Karimi H, Jadid S. Real -Time Pricing Design Considering Uncertainty of Renewable Energy Resources and Thermal Loads in Smart Grids . Journal of Iranian Association of Electrical and Electronics Engineers. 2019; 16 (1) :1-10

[22] Lowe, R., et al. Multi-agent actor-critic for mixed cooperative-competitive environments. in Advances in Neural Information Processing Systems. 2017.

[23] Zhang, Z., Zhang, D., & Qiu, R. C. Deep reinforcement learning for power system applications: An overview. CSEE Journal of Power and Energy Systems, 6(1), 213-225, 2019.

[24] Wan, Z., Li, H., & He, H. (2018, July). Residential energy management with deep reinforcement learning. In 2018 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.

[25] Sutton, R. S., and Barto A. G. Reinforcement learning: An introduction. MIT press, 2018.

[26] Yang, Y., Hao, J., Sun, M., Wang, Z., Fan, C. and Strbac, G., July. Recurrent Deep Multiagent Q-Learning for Autonomous Brokers in Smart Grid. In IJCAI (Vol. 18, pp. 569-575), 2018.

[27] Keogh, E. and C.A. Ratanamahatana, Exact indexing of dynamic time warping. Knowledge and information systems, 2005. 7(3): p. 358-386.

[28] https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption. visited Dec 2018.

[29] https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households visiited dec 2018.

Reddy, P. Semi-Cooperative Learning in Smart Grid Agents. PhD Thesis. CMU-ML-13-114 Carnegie Mellon University, December 2013.

زیرنویس ها

[1] Smart Grid
[2] Broker
[3] Prosumer
[4] Time of Use
[5] Demand Side Management
[6] Self-healing Networks

[7] Particle Swarm Optimization
[8] Free-riding Customer
[9] Real Time Pricing
[10] Tit-For-Tat
[11] Heuristic-based Strategy
[12] Markov Decision Process
[13] Semi Markov Decision Process
[14] Underlying Discrete Model
[15] Offline
[16] Day-ahead Strategy
[17] Boltzmann
[18] Karush–Kuhn–Tucker
[19] Dynamic Time Warping
[20] Temporal Difference
[21] On-Policy
[22] Portfolio
[23] Profile History
[24] Subjective Margin
[25] Short-supply
[26] Over-supply
[27] Balanced
[28] Eligibility Traces
[29] Bellman
[30] Benchmark