شیوه نوین برای حل مسئله تعمیرات و نگهداری با استفاده از الگوریتم یادگیری تقویتی

ابوالفضل قاسمی[۱]، سعید سید مهدوی[۲]، محمدحسین جاویدی[۳] و محمد باقر نقیبی سیستانی[۴]

[۱] آزمایشگاه تخصصی مطالعات سیستم و تجدید ساختار، دانشکده مهندسی، دانشگاه فردوسی مشهد، Abolfazl.ghasemi@Stu-mail.um.ac.ir
[۲] آزمایشگاه تخصصی مطالعات سیستم و تجدید ساختار، دانشکده مهندسی، دانشگاه فردوسی مشهد، Saeed.seyyedmahdavi@Gmail.com
[۳] آزمایشگاه تخصصی مطالعات سیستم و تجدید ساختار، دانشکده مهندسی، دانشگاه فردوسی مشهد، H-javidi@Ferdowsi.um.ac.ir
[۴] دانشکده مهندسی، دانشگاه فردوسی مشهد، Naghib@Yahoo.com

چکیده - تعمیرات و نگهداری واحدهای تولیدی از مهم‌ترین مسایل در بهره‌برداری سیستم‌های قدرت می‌باشد. استراتژی خارج کردن واحدهای تولیدی بمنظور انجام سرویس‌های برنامه‌ریزی شده و تعمیرات تاثیر مستقیم روی قابلیت اطمینان شبکه و درآمد شرکت‌های تولیدی دارد. با توجه به انگیزه افزایش سود شرکت‌های تولیدی در محیط تجدیدساختار یافته، در این مقاله روش جدیدی مبتنی بر تئوری بازیها و یادگیری تقویتی برای بررسی مسئله تعمیرات و نگهداری ارائه شده است. سپس با اعمال روش مذکور به یک سیستم، صحت و کارآمدی آن نشان داده شده است.

کلید واژه- تعمیرات و نگهداری، تجدید ساختار، تئوری بازیها، یادگیری تقویتی

۱- مقدمه

مسئله تعمیرات و نگهداری واحدهای تولیدی در سیستم قدرت از موارد مهمی است که بایستی مدنظر مالکان باشد. در محیط تجدید ساختار یافته سیستم قدرت، که هر مالک به دنبال حداکثرسازی سود خود می‌باشد، مسئله تعمیرات و نگهداری اینگونه مطرح می‌شود که هر مالک واحدهایش را بگونه‌ای از مدار خارج می‌کند که سود حاصل از این تصمیم‌گیری حداکثر گردد. البته، همراستا با این تصمیم‌گیری، بهره‌بردار مستقل سیستم نیز به دنبال تامین بار مورد نیاز و حفظ قابلیت اطمینان شبکه می‌باشد. در زمینه برنامه‌ریزی برای زمانبندی خروج واحدها از مدار، به منظور تعمیرات و نگهداری، مقالات زیادی منتشر شده است. در مرجع[۱] مسئله زمانبندی تعمیرات و نگهداری بصورت یک بازی استاتیکی مطرح شده که بازیگران، برای رسیدن به حداکثر سود، به یک نقطه تعادل نش می‌رسند. مرجع[۲] برای مسئله زمانبندی تعمیرات و نگهداری یک روش انعطاف‌پذیر را، با در نظر گرفتن قیود عدم قطعیت، با کمک روش فازی مطرح می‌کند. مرجع[۳] با استفاده از تئوری بازیها مسئله تعمیرات و نگهداری واحدهای تولیدی را بصورت یک بازی بسط داده است. این روش متشکل از یک بازی استاتیکی و یک زیر بازی دینامیکی می‌باشد که نهایتا، در پایان افق زمانی مدنظر، شرکت‌های تولیدی به نقطه تعادل نش می‌رسند. در مرجع [۴] نیز، نویسنده مسئله تعمیرات و نگهداری را با برنامه‌ریزی دینامیکی حل کرده است. در این مقاله، سریعتر رسیدن به دوره تعمیرات تا حد امکان نیز، بعنوان تابع هدف مسئله در نظر گرفته شده است. در مرجع [۵] ، برای برنامه زمانبندی تعمیرات و نگهداری روشی ارائه شده است که در آن محدودیت انتقال بین ناحیه‌ای و قیود احتمالی قابلیت اطمینان در نظر گرفته شده است. این روش حل بر پایه تجزیه بندر است. تابع هدف مورد نظر در این مرجع حداقل کردن هزینه‌های تعمیرات و نگهداری است. در طی سالیان اخیر، از روشهای زیادی برای حل مسئله تعمیرات و نگهداری بخصوص در محیط تجدید ساختاریافته استفاده شده است که برخی از آنها عبارتند از:

- روش برنامه‌ریزی خطی
- برنامه‌ریزی مختلط صحیح
- روش‌های تجزیه
- الگوریتم ژنتیک
- منطق فازی
- شبکه عصبی

در این مقاله، هدف شیوه جدیدی برای حل مسئله

تعمیرات و نگهداری ارائه گردیده است. در روش ارائه شده مسئله با طرح از دید تئوری بازیها و بکمک روش یادگیری تقویتی (Reinforcement Learning) حل می‌شود. ابتدا، در بخش ۲، مسئله از دید تئوری بازی بیان می‌شود. سپس در بخش ۳، روش یادگیری تقویتی به اختصار توضیح داده می‌شود. در بخش ۴، شیوه استفاده از یادگیری تقویتی در حل مسئله تعمیرات و نگهداری بیان شده است. نهایتا، در بخش ۵ نیز، صحت روش بر روی یک مساله نمونه، بررسی شده است.

۲- مسئله تعمیرات و نگهداری از دید تئوری بازیها

در محیط تجدید ساختار یافته، هر مالک به دنبال حداکثرسازی سود خود می‌باشد. اتخاذ تصمیم برای خارج کردن واحدها به منظور تعمیرات و نگهداری واحدها نیز از قاعده فوق مستثنی نیست. بازیگران در این بازی همان شرکت‌های تولیدی هستند. بازیگران بایستی بهترین زمان را برای خارج کردن واحدهای خود چنان در نظر بگیرند که سودشان حداکثر شود.

در اینجا، به بررسی بازی در دو حالت همکارانه و غیرهمکارانه شرکت‌ها برای تصمیم‌گیری آنها در برنامه‌ریزی تعمیرات و نگهداری واحدهایشان پرداخته شده است. منظور از بازی همکارانه این است که بازیگران با همکاری یکدیگر طوری برنامه‌ریزی می‌کنند که مجموع سود آنها بیشینه شود. در سمت مقابل بازی غیر همکارانه به حالتی اتلاق می‌شود که هر بازیگر صرفا به دنبال حداکثرسازی سود خود می‌باشد.

مسئله برنامه‌ریزی برای تعمیرات و نگهداری واحدها وابسته به بازار انرژی است یعنی، با تغییر قیمت بازار تصمیم گیری می‌تواند عوض شود.به همین دلیل این مسئله می‌تواند بعنوان یک بازی دینامیکی در نظر گرفته شود. با این حال دینامیکی بودن به معنای ترتیبی بودن این بازی نیست بلکه، همه بازیگران هم‌زمان تصمیم می‌گیرند و این بخش کار یک بازی استاتیکی است.

همانطور که اشاره شد؛ بازیگران این بازی شرکت‌های تولیدی در محیط تجدیدساختار یافته هستند. در این مقاله از حروف i وj برای نشان دادن شرکت تولیدی و واحد استفاده می کنیم. به عبارت دیگر منظور از Gij همان واحد j از شرکت تولیدی i میباشد که مقید به خارج شدن از مدار بمنظور تعمیرات و نگهداری در دوره dij می‌باشد. افق زمانی تعمیرات و نگهداری که ممکن است فرض شود یک سال (۳۶۵ روز) و یا یک ماه (۳۱ روز) است با حرف T نمایش داده می‌شود.

استراتژی این بازی این است که شرکت تولیدی واحدش را برای تعمیرات از مدار خارج می‌کند (y) و یا خارج نمی‌کند (n) که آنرا بصورت s={y,n} نمایش می‌دهیم.

در بازی غیر همکارانه (رقابت کامل)، هر شرکت بدنبال حداکثرسازی سود خود در پایان رقابت می‌باشد. در بازی همکارانه نیز، تابع مطلوبیت برای بازیگران حداکثر سازی سود کل در پایان بازی است. برای این کار، پس از مشخص شدن قیمت تسویه بازار، درآمد بازیگران و هزینه آنان مشخص می شود. سپس سود آنها بکمک آن محاسبه می شود[۶].

$$R_i = \sum_{j=1}^{N}\sum_{t=1}^{T} \pi^t q_{i,j}^t \qquad (۱)$$

$$C_{oi} = \sum_{j=1}^{N_i}\sum_{t=1}^{T}\left[a_{i,j} + b_{i,j} q_{i,j}^t + c_{i,j}(q_{i,j}^t)^2\right] \qquad (۲)$$

$$C_{mi} = \sum_{j=1}^{N_i} \omega_{i,j} \qquad (۳)$$

در روابط فوق R_i درآمد شرکت i ام است، π^t قیمت تسویه بازار در دوره t، C_{oi} تابع هزینه واحدها و C_{mi} هزینه مربوط به تعمیرات و نگهداری واحدها می‌باشد. تابع سود هر شرکت بصورت رابطه (۴) بیان می‌شود.

$$b_i = R_i - (C_{oi} + C_{mi}) \qquad (۴)$$

در بازی همکارانه هدف حداکثرسازی سود مجموع بازیگران است. پس حداکثر سود برابر است با:

$$B = \sum_{i=1}^{N_i} b_i \qquad (۵)$$

با این روش نهایتا بازیگران به یک تعادل در جبهه پرتو می‌رسند و حل این مسئله به سمتی می رود که، در پایان بازی، سود مجموع بازیگران حداکثر شود.

۳- روش یادگیری تقویتی

یادگیری تقویتی روشی موثر برای حل مسایل تصمیم‌گیری چند مرحله‌ای است. یک مسئله تصمیم گیری مارکوفی در برگیرنده تصمیم‌گیری یا انتخاب عمل a_k در مرحله یا حالت x_k سیستم است. با انتخاب عمل a_k، سیستم به حالت جدید x_{k+1} حرکت می‌کند. پاسخ یک مسئله تصمیم‌گیری چند مرحله‌ای مارکوفی، پیدا کردن ترتیبی از اعمال $a_0, a_1, a_2, \cdots, a_N$ برای رسیدن به هدف مورد نیاز است[۷]. در اینجا، مسئله برنامه ریزی تعمیرات و نگهداری واحدهای

تولیدی بصورت یک مسئله تصمیم‌گیری چند مرحله‌ای مارکوفی فرمول‌بندی می‌شود و به وسیله یک الگوریتم مبتنی بر $Q-Learning$ به حل آن پرداخته می‌شود.

الگوریتم مبتنی بر $Q-Learning$ به پیدا کردن ارزش Q است، که شامل ارزش جفت حالت-عمل (x,a) است، و معیاری از مناسب بودن انتخاب عمل a در حالت x می‌باشد. اگر $Q(a^*,x_k)\geq Q(a_k,x_k)$، آنگاه برای تمام اعمال ممکن در یک حالت، a^* بهترین عمل در آن حالت است.

الگوریتم با تخمین اولیه مقادیر ارزشهای Q آغاز می‌شود. هر بار که عامل تصمیم‌گیرنده با محیط تعامل می‌کند و به انتخاب عمل می‌پردازد، عملیات یادگیری صورت می‌گیرد. در روش $Q-Learning$ با پاسخی که بر اساس انتخاب جفت حالت-عمل از محیط دریافت می‌شود، مقادیر ارزشهای Q به روز می‌شوند. با انتخاب عمل، سیستم به حالت جدید می‌رود. بدین ترتیب، مجموعه‌ای از اعمال جدید برای سیستم مهیا می‌شود. با پیشرفت الگوریتم، مقادیر ارزشهای Q به مقادیر بهینه همگرا می‌شوند[۸].

حال، به بررسی چگونگی انتخاب عمل در هر حالت می‌پردازیم. در هر حالت، عملی وجود دارد که مقدار ارزش Q تا آن لحظه بیشترین است و به آن عمل، عمل حریصانه گفته می‌شود. این در حالیست که، امکان دارد مقدار تخمین تا آن لحظه اشتباه بوده باشد و عمل بهتری وجود داشته باشد. بنابراین، استراتژی انتخاب عمل باید به نحوی باشد که، علیرغم بهره‌برداری از عملی که بیشترین ارزش را دارد، به بررسی سایر اعمال نیز پرداخته شود. یک روش انتخاب اعمال روش $\varepsilon-greedy$ است . در این روش، عمل با بیشترین ارزش با احتمال $1-\varepsilon$ و یکی از اعمال دیگر با احتمال ε انتخاب می‌شوند. معمولا، در تکرارهای اول مقدار ε بزرگ انتخاب می‌شود تا همه اعمال مورد بررسی قرار گیرند. با پیشروی الگوریتم و در تکرارهای بالا، مقدار ε کاهش یافته تا شانس انتخاب اعمالی با ارزش کمتر کاهش یابد[۷-۹].

بنابراین، یادگیری تقویتی نیازمند مجموعه ای از حالات و اعمال جهت انتخاب است که، بر حسب سیگنال پاداش، مقدار ارزش Q را به روز می‌کند. اگر x_{k+1} حالت دستیابی شده از حالت x_k با انتخاب عمل a_k باشد، تابع ارزش جفت حالت-عمل $Q(x_k,a_k)$ بصورت زیر به روز می‌شود:

$$Q^{n+1}(x_k,a_k)=Q^n(x_k,a_k)+\alpha[g(x_k,a_k,x_{k+1}) +\gamma\times\max_{a'\in A}(Q(x_{k+1},a'))-Q^n(x_k,a_k)] \quad (۶)$$

که در رابطه فوق، A مجموعه اعمال و $g(x_k,a_k,x_{k+1})$ تابع پاداش مربوط به زوج حالت-عمل مربوطه است. ضریب α ضریب یادگیری و γ نرخ نزول برای تاثیر دهی پاداش‌های آینده است [۷].

۴- الگوریتم یادگیری تقویتی برای برنامه‌ریزی تعمیرات و نگهداری

در این بخش ابتدا مسئله برنامه‌ریزی تعمیرات واحدهای نیروگاهی بصورت یک مسئله تصمیم‌گیری چند مرحله‌ای، فرمول‌بندی می‌شود سپس الگوریتم یادگیری تقویتی برای حل این مسئله ارائه می‌شود.

در مسئله برنامه‌ریزی تعمیرات هدف، پیدا کردن برنامه بهینه برای خارج کردن واحدهای تولیدی با هدف بیشینه کردن مقدار سود در طی چندین بازه مشخص است. یادگیری تقویتی در این حالت یک مسئله با N مرحله تصمیم گیری است بطوریکه N در اینجا، تعداد بازه‌های زمانی مورد بحث می‌باشد. حالت‌های سیستم در هر یک از مراحل تصمیم‌گیری، به تعداد واحدهای تولیدی هر یک از شرکت‌ها بستگی دارد و برای هر یک از شرکت‌ها دارای 2^n حالت است که n تعداد واحدهای تولیدی هر شرکت است[۱۰]. هر حالت را می‌توان به صورت چندتایی (k,m,j) نشان داد که در آن، k شماره مرحله تصمیم‌گیری، m شماره واحد نیروگاهی و j یکی از 2^{n_m} حالت ممکن برای n_m واحد شرکت تولیدی شماره m است. شماره مجموعه کلیه حالت‌های موجود در مرحله k را می‌توان به صورت زیر نشان داد:

$$X_k=\{(k,1\cdots M,1\cdots 2^{n_m})\} \quad (۷)$$

بطوریکه در رابطه بالا، m تعداد واحدهای نیروگاهی و n_m تعداد واحد هر واحد نیروگاهی است.

در مرحله k، با انتخاب عمل a_k از مجموعه تمامی اعمال ممکن در A_k به مرحله k+1 می‌رسیم. این مجموعه اعمال ممکن در مرحله k، دربرگیرنده تغییر وضعیت شامل خارج یا وارد کردن هر یک از واحدهای تولیدی در مرحله k می‌شود. عمل a_k را می توان با زوج (m,i) نشان داد که در آن، m شماره شرکت تولیدی و i یکی از 2^{n_m} حالت ممکن برای n_m واحد نیروگاه شماره m است. بنابراین، مجموعه تمام اعمال ممکن برای مرحله k را می‌توان به صورت زیر نشان داد:

$$A_k=\{(1\cdots m,1\cdots 2^{n_m})\} \quad (۸)$$

سیستم یادگیری ،در هر انتخاب عمل، پاداشی دریافت می‌کند. پس از انتخاب عمل a_k، سیستم از حالت k به حالت

k+1 می‌رود. در اثر این تغییر حالت سیستم، وضعیت توزیع اقتصادی بار بین واحدهای نیروگاهی تغییر کرده و در نتیجه هزینه و درآمد هرکدام از شرکت‌ها نیز دچار تغییر می‌شود. بنابراین، در مسئله برنامه‌ریزی تعمیرات واحدهای نیروگاهی، سیگنال پاداش توابع ارزش هر یک از شرکت‌ها را می توان به صورت مقدار سود آنان در نظر گرفت که؛ این سود مجموع سود تمامی واحدهای آن شرکت تولیدی است.

بنابراین، سیگنال پاداش به صورت رابطه (۹) در نظر گرفته می‌شود:

$$r_k = g(x_k, a_k, x_{k+1}) = c_k(x_k) \qquad (۹)$$

برای انتخاب اعمال از مجموعه اعمال، از روش $\varepsilon - greedy$ استفاده می‌شود که در آن عمل با بیشترین ارزش با احتمال $1-\varepsilon$ ، و با احتمال ε از بین سایر اعمال یک عملگر انتخاب می شوند.

برای هریک از مراحل ۱ تا N، مقادیر توابع ارزش زوج حالت-عمل با رابطه زیر به روز می شوند:

$$Q^{n+1}(x_k, a_k) = Q^n(x_k, a_k) + \alpha \times [r_k + \gamma \times \max_{a' \in A}(Q(x_{k+1}, a')) - Q^n(x_k, a_k)] \qquad (۱۰)$$

هنگامی‌که سیستم به مرحله N برسد، دیگر مرحله بعدی وجود نخواهد داشت. برای به روزرسانی تابع ارزش زوج حالت-عمل در این حالت، از رابطه زیر استفاده می شود:

$$Q^{n+1}(x_k, a_k) = Q^n(x_k, a_k) + \alpha \times [r_k - Q^n(x_k, a_k)] \qquad (۱۱)$$

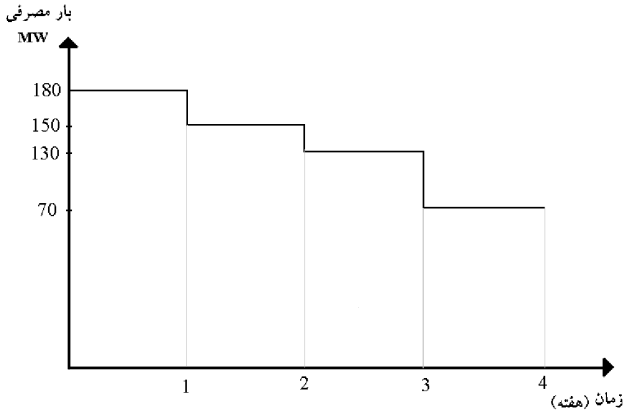
بمنظور شبیه‌سازی الگوریتم، باید مقادیری برای ضرایب ε، α و γ در نظر گرفته شوند. مقدار ε در تکرارهای اول برابر با ۰/۵ انتخاب شده است تا تمامی اعمال، به میزان لازم، مورد بررسی قرار بگیرند. با پیشرفت برنامه و در تکرارهای بالا، این مقدار کاهش می‌یابد. پارامتر یادگیری α تعیین می‌کند که تا چه میزان مقادیر ارزشهای Q در هر تکرار مورد به روزرسانی واقع می‌شوند. مقادیر کوچک ضریب یادگیری همگرایی را کندتر می کند در حالی که، مقادیر بزرگ ممکن است به حالت نوسانی منجر شود. در این مسئله با آزمون و خطا ضریب یادگیری ۰/۱ انتخاب شده است. در یادگیری تقویتی، مقدار پاداشی که در مراحل بعد بدست می‌آید ممکن است تاثیری متناسب با پاداشی که در این مرحله بدست می‌آید، نداشته باشد. با این حال در مسئله توزیع اقتصادی، پاداش های مرحله بعد با پاداش این مرحله تاثیر یکسانی دارد، به همین دلیل، نرخ نزول برابر یک در نظر گرفته می‌شود.

بعد از اینکه فرآیند یادگیری کامل شد، برای بدست آوردن توزیع متناظر با توان مورد نیاز، اعمالی با بیشترین ارزش حالت-عمل به صورت حریصانه در هر مرحله انتخاب میشوند که، این امر به برنامه تخصیص بهینه منجر می‌شود.

۵- شبیه سازی و استخراج نتایج

برای بررسی صحت و کارآیی الگوریتم ارائه شده، از شبکه نمونه مرجع [۳] استفاده شده است. در این نمونه، دو شرکت تولیدی که مشخصات آن‌ها در جدول (۱) آمده است، در بازار شرکت کرده‌اند. هر شرکت به دنبال حداکثر سازی سود خود است. در واقع، مسئله به صورت یک بازی غیر همکارانه مطرح شده است. در این تحقیق، مسئله برنامه ریزی تعمیرات و نگهداری واحدها به دو صورت بازیهمکارانه و غیر همکارانه مورد بررسی قرار گرفته است.

این برنامه ریزی برای چهار بازه زمانی هفتگی صورت میگیرد که منحنی مصرف در این بازه زمانی در شکل (۱) آورده شده است.



شکل (۱)- منحنی مصرف در دوره تعمیرات

برای بدست آوردن میزان درآمد واحدهای تولیدی نیاز به مشخص شدن قیمت بازار می‌باشد. در اینجا، قیمت بازار برابر با هزینه افزایشی واحدهای تولیدی در نظر گرفته می‌شود. برای پیاده سازی الگوریتم یادگیری تقویتی برای بازی همکارانه، تنها یک تابع ارزش برای زوجهای حالت-عمل، در هر بازه زمانی، در نظر گرفته شده است. برای حالتی که که تابع هدف حداکثر سازی سود هر شرکت تولیدی است یا به عبارت دیگر بازی غیرهمکارانه است، به هر شرکت تولیدی یک تابع ارزش تخصیص داده شده است که ارزش هر زوج حالت-عمل در هر بازه زمانی را نشان می‌دهد. ترتیب ورود و خروج واحدها به منظور تعمیرات و نگهداری در دو حالت بازی همکارانه و غیر همکارانه در جدول (۲) نشان داده شده است.

جدول(۱)- مشخصات واحدهای مربوط به شرکتهای تولیدی

شرکت تولیدی	واحد تولیدی	مدت زمان رفتن به تعمیرات (هفته)	حداقل تولید (MW)	حداکثر تولید (MW)	ضرایب تابع هزینه		
					aij	bij	cij
Genco1	۱	۱	۳۰	۱۲۰	۰/۰۰۵	۰/۱	۷/۵
	۲	۲	۲۰	۸۰	۰/۰۰۲۵	۰/۵	۴/۳
Genco2	۱	۱	۲۰	۱۰۰	۰/۰۰۲	۰/۶	۵/۲

جدول (۳)- مقایسه پاسخ برای بازی و روش حل متفاوت

نوع بازی و روش حل	وضعیت واحدها ۱=در مدار ۰=خارج از مدار				وضعیت سود شرکت‌ها	
	۱Genco		۲Genco		۱Genco	۲Genco
	واحد۱	واحد۲	واحد۱			
همکارانه RL	۱	۰	۱		۸۰/۷	۹/۷
	۱	۰	۱			
	۱	۱	۰			
	۰	۱	۱			
غیرهمکارانه RL	۱	۱	۰		۷۵/۷	۱۷/۴
	۱	۰	۱			
	۱	۰	۱			
	۰	۱	۱			
غیرهمکارانه MSU	۱	۱	۰		۷۵/۷	۱۷/۴
	۱	۰	۱			
	۱	۰	۱			
	۰	۱	۱			

در جدول (۳)، به مقایسه نتایج بدست آمده از الگوریتم پیشنهادی برای مسئله تعمیرات و نگهداری در دو بازی همکارانه و غیر همکارانه با نتایج حاصل از الگوریتم بازی استاتیک-بازی دینامیک ارائه شده در مرجع [۳] پرداخته شده است.

۶- نتیجه گیری و جمع بندی

در این مقاله با کمک یادگیری تقویتی و مفاهیم تئوری بازیها، به بررسی زمانبندی تعمیرات و نگهداری واحدهای نیروگاهی در محیط تجدید ساختار یافته پرداخته شده است. این مسئله برای دو حالت همکارانه و رقابت کامل بررسی و الگوریتمی برای تحلیل آن بوسیله یادگیری تقویتی ارائه شده است. با استفاده از یادگیری تقویتی، به بسیاری از مسیرهای غیر مطلوب در حین فرآیند یادگیری، ارزش کمی اختصاص داده می‌شود که در تکرارهای بعد، خود بخود موجب کم شدن احتمال انتخاب آنها می‌شود. صحت و کارآیی الگوریتم ارائه شده بر روی شبکه نمونه بررسی شده است. نشان داده شد که، الگوریتم پیشنهادی یک روش کاربردی برای حل مسئله برنامه ریزی ورود و خروج واحدهای تولیدی به منظور تعمیرات و نگهداری است.

جدول(۲)- زمان رفتن به تعمیرات واحدها در دوره مورد نظر در حالت الف)بازی همکارانه ب) بازی غیر همکارانه

شرکت تولیدی	واحد تولیدی	هفته اول	هفته دوم	هفته سوم	هفته چهارم
Genco1	واحد۱	-	-	-	تعمیرات
	واحد۲	-	تعمیرات	تعمیرات	-
Genco2	واحد۱	تعمیرات	-	-	-

(الف)

شرکت تولیدی	واحد تولیدی	هفته اول	هفته دوم	هفته سوم	هفته چهارم
Genco1	واحد۱	-	-	-	تعمیرات
	واحد۲	تعمیرات	تعمیرات	-	-
Genco2	واحد۱	-	-	تعمیرات	-

(ب)

مراجع

[1] D.Chattopadhyay, "A Game Theoretic Model for Strategic Maintenance and Dispatch Decisions", IEEE Transactions on Power Systems. Vol. 19, No. 4, PP. 2014-2021, Nov 2004.

[2] R. C. Leou, "A Flexible Unit Maintenance Scheduling Considering Uncertainties", IEEE Transactions on Power Systems, Vol. 16, No.3, PP. 552-559, Aug. 2001.

[3] J.Kim and J.Park, "A new game-theoretic approach to maintenance scheduling problems in competitive electricity markets"Power Engineering Society Summer Meeting,, vol. 3, PP. 1510 - 1515, 2002.

[4] T.N. Mohammadi and S.B.Hassanpour et all "A New Approach for Generation Maintenance Scheduling in a Deregulated Power System" TENCON 2005 IEEE, PP. 1 – 5, 2005.

[5] E.L.Silva and M.Morozowski, "Transmission constrained maintenance scheduling of generating units: a stochastic programming approach" Power Systems, IEEE Transactions on Vol. 10, PP. 695 – 701, 1995.

[6] M. Shahidehpour, H. Yamin, and Z. Li, "Market operations in electric power systems", [New York]: Institute of Electrical and Electronics Engineers, Wiley-Inter science, 2002.

[7] R.S.Sutton, A.G.Barto. "Reinforcement Learning: An Introduction", Cambridge, MA: MIT Press, 1998.

[8] T.P.Imthias Ahamed. "A Reinforcement Learning Approach to Unit Commitment Problem", Proceedings of National Power System Conference 2006.

[9] E.A.Jasmin, T.P.Imthias Ahamed, V.P.Jagathiraj. "A Reinforcement Learning Algorithm for Economic Dispatch considering transmission losses", TENCON 2008, IEEE region 10 conference.

[10] A.J.Wood, B.F.Woolenberg. "Power Generation and Control", John Wiley Sons, 2002.