

کنترل کننده مقاوم تطبیقی بار فرکانس مبتنی بر یادگیری تقویتی برای یک سیستم قدرت به هم پیوسته شامل SMES

عادل اکبری مجد^۱، دانشیار؛ حسین شایقی^۲، استاد؛ حمید محمدنژاد^۳، دانشجوی کارشناسی ارشد؛ عبدالله یونسی^۴، دانشجوی دکتری

۱، ۲، ۳ و ۴- گروه مهندسی برق - دانشکده فنی و مهندسی - دانشگاه محقق اردبیلی - اردبیل - ایران

^۱akbarimajd@gmail.com, ^۲hshayeghi@gmail.com, ^۳hamidmohammadnejad70@gmail.com, ^۴a.younesi@ieee.org

چکیده: هدف از این مقاله استفاده از یادگیری تقویتی برای طراحی کنترل کننده‌های PID و SMES مقاوم و تطبیقی برای کنترل بار فرکانسی در یک سیستم قدرت دو ناحیه‌ای حرارتی است. ابتدا تنظیم پارامترهای کنترل کننده‌های PID و SMES به صورت یک مسئله بهینه‌سازی مدل شده توسط الگوریتم تدریس - یادگیری اصلاح شده حل می‌شود. سپس عملکرد هم‌زمان آن‌ها با استفاده از الگوریتم پیشنهادی مبتنی بر یادگیری تقویتی بهینه می‌گردد. کنترل کننده‌های مبتنی بر یادگیری، ساختاری ساده و قابل فهم داشته و به راحتی به هر سیستمی قابل اعمال هستند. در نهایت برای ارزیابی عملکرد کنترل کننده پیشنهادی شبیه‌سازی‌های کامپیوتری توسط نرم‌افزار MATLAB صورت گرفته است و مقایسه روش‌های کنترلی طبق معیارهای حوزه زمان، ITAE، IAE، ITSE و مقدار فراجهدش و فروجهش، نشان از برتری روش ارائه شده را دارد.

واژه‌های کلیدی: کنترل بار فرکانس، الگوریتم بهینه‌سازی مبتنی بر تدریس - یادگیری اصلاح شده، یادگیری Q، SMES.

Robust Adaptive Load Frequency Controller Based on Reinforcement Learning in an Inter-Connected Power System

A. Akbarimajd¹, Associate Professor; H. Shayeghi², Professor, H. Mohammad Nejad³, MSc Student, A. Younesi⁴, PhD Student

1, 2, 3, 4- Department of Electrical Engineering, Faculty of Engineering, University of Mohaghegh Ardabili, Ardabil, Iran
E-mails: ¹akbarimajd@gmail.com, ²hshayeghi@gmail.com, ³hamidmohammadnejad70@gmail.com, ⁴a.younesi@ieee.org

Abstract: The aim of this paper is using reinforcement learning for designing of robust and adaptive PID and SMES controllers to load frequency control in a two area thermal power system. Thus, in first setting of PID and SMES controller parameters formulated as an optimization problem and solved using teaching-learning optimization algorithm. Then the simultaneous performance of designed controllers improved using proposed reinforcement learning based controller. Simple and understandable structure and easy to use are distinguished advantages of q-learning based controllers. In order to evaluate the performance of the proposed controller, computer simulations have been done by using MATLAB software. Simulation results verified that the proposed q-learning based controller exhibits much better performance from the conventional optimization based controllers from viewpoint of time domain performance indices like over shoot, under shoot, ITAE, ITSE, and IAE.

Keywords: LFC, MTLBO, Q-learning, SMES.

تاریخ ارسال مقاله: ۱۳۹۴/۰۹/۰۱

تاریخ اصلاح مقاله: ۱۳۹۴/۱۲/۱۵ و ۱۳۹۵/۰۳/۰۲

تاریخ پذیرش مقاله: ۱۳۹۵/۰۵/۲۵

نام نویسنده مسئول: عادل اکبری مجد

نشانی نویسنده مسئول: ایران - اردبیل - خیابان دانشگاه - دانشگاه محقق اردبیلی - دانشکده فنی و مهندسی - گروه مهندسی برق.

۱- مقدمه

ناحیه سیستم چهارماشینه کارایی این روش در کنترل^۶ TCSC، نشان داده شده است. از این مقاله نتیجه گرفته می‌شود که یادگیری تقویتی بر روی هر سیستمی با هر اندازه و پیچیدگی دینامیکی قابل اجرا است و می‌تواند بدون شناخت زیادی از سیستم، آن را کنترل کند. همچنین این روش کنترلی مقاوم بوده و با تغییر شرایط سیستم خود را مطابق با سیستم تطبیق می‌دهد. در [۹] یادگیری تقویتی برای کنترل توان راکتیو و با روش‌های احتمالاتی محدود پخش بار^۷ (CLF) مقایسه شده است. در [۱۰] الگوریتم یادگیری Q (نسخه‌ای از یادگیری تقویتی) برای حل مسئله کنترل بهینه باتری برای سیستم‌های انرژی خانگی پیشنهاد شده است. در [۱۲، ۱۱]، کنترل بهینه منابع تجدیدپذیر برای سیستم‌های ریزشبه هوشمند توسط روش یادگیری Q انجام شده است. هدف این مقاله، طراحی هماهنگ کنترل کننده PID با SMES با استفاده از الگوریتم MTLBO و بهبود عملکرد آن با استفاده از روش یادگیری تقویتی به منظور کم کردن نوسانات فرکانس ناشی از تغییرات بار و غلبه بر تأخیر زمانی ناشی از گاورنر می‌باشد. چالش اصلی در طراحی کنترل کننده‌ها با استفاده از یادگیری Q آن است که فضای جستجوی ماتریس Q بزرگ است؛ در نتیجه در مقالاتی که با استفاده از روش یادگیری Q به تنظیم پارامترهای کنترل کننده می‌پردازد پیچیدگی فضایی و زمانی الگوریتم بالا است و در این مقاله برای غلبه بر این مشکل ابتدا فضای جستجوی عامل توسط بهینه‌سازی با الگوریتم بهینه‌سازی مبتنی بر تدریس - یادگیری اصلاح شده کاهش داده شده است. سپس عامل در این محیط شروع به جستجوی عمل بهینه برای هر موقعیت محیط می‌کند. تابع هدف برای الگوریتم MTLBO معیار خطای ITAE بوده و نتایج حاصل از این الگوریتم با الگوریتم بهینه‌سازی اجتماع ذرات^۸ (PSO) مقایسه شده است.

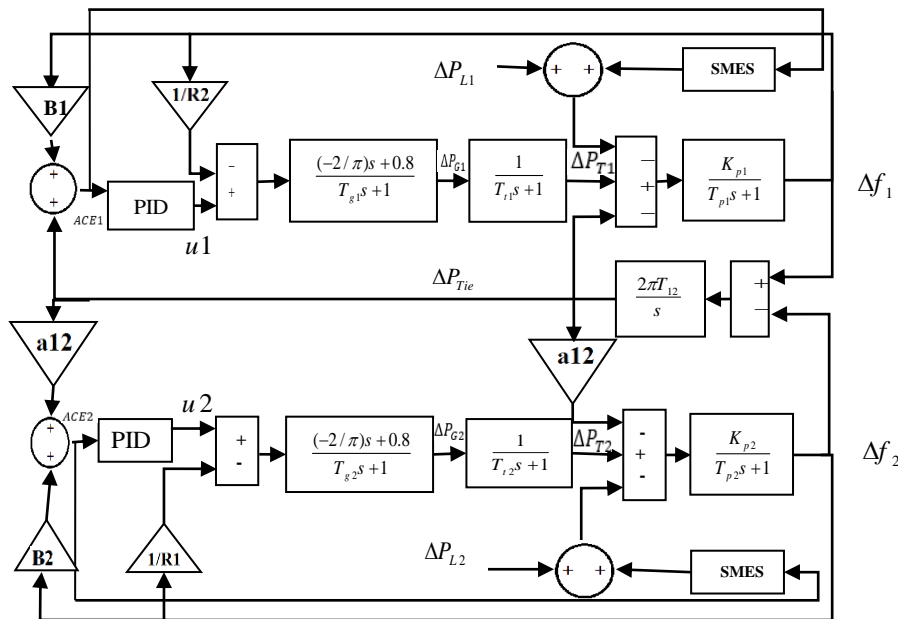
۲- سیستم مورد مطالعه

سیستم مورد مطالعه پیشنهادی از ارتباط دو ناحیه تشکیل شده است. هر دو ناحیه شامل یک سیستم گرمایی است که در شکل ۱ نشان داده است. $B1$ و $B2$ نشان‌دهنده بایاس فرکانسی با پارامترهای $ACE1$ و $ACE2$ که خطای کنترلی هر ناحیه هستند و $u1$ و $u2$ خروجی کنترلی هر کنترل کننده و $R1$ و $R2$ نشان‌دهنده ضریب تنظیم سرعت گاورنر به p.u.Hz ؛ $TG1$ و $TG2$ ثابت سرعت گاورنر؛ $\Delta PG1$ و $\Delta PG2$ تغییر در مقادیر توان گاورنر؛ $T1$ و $T2$ ثابت زمانی توربین؛ $\Delta PT1$ و $\Delta PT2$ تغییر در مقدار توان توربین؛ $\Delta PTie$ تغییر در توان خط ارتباطی؛ $KPS1$ و $KPS2$ ضریب مقاومت سیستم قدرت؛ $TPS1$ و $TPS2$ ثابت زمانی سیستم قدرت؛ $T12$ ضریب سنکرون و $\Delta f1$ و $\Delta f2$ تغییرات فرکانس سیستم به Hz است. مقادیر پارامترهای مربوط به سیستم مورد مطالعه در ضمیمه (الف) آورده شده است. باند راکد سرعت گاورنر فقط برای واحدهای حرارتی تعریف می‌شود. ناحیه راکد گاورنر تأثیر زیادی در پاسخ دینامیکی مکانیسم کنترل بار فرکانس دارد.

سیستم قدرت از اجزای مختلفی تشکیل شده است که انرژی الکتریکی را در ابعاد وسیعی انتقال می‌دهد. زمانی بین تولید و مصرف تعادل برقرار می‌شود که فرکانس سیستم به خوبی کنترل شود. اگر مقدار تولید توان از مقدار مصرف کم‌تر باشد فرکانس سیستم کاهش می‌یابد. هرچه سیستم بزرگ‌تر باشد؛ تغییرات بار، اثر کم‌تری روی فرکانس سیستم می‌گذارد [۱]. هدف از طراحی کنترل کننده بار-فرکانس کم کردن نوسانات فرکانس و تغییرات بار در شرایط عادی و در صورت بروز اغتشاش در شبکه است. با ظهور ادوات الکترونیک قدرت با عملکرد سریع و ذخیره‌سازهای انرژی، محققان روش‌های زیادی را برای بهبود نوسانات خطوط ارتباطی همانند به‌کارگیری SMES و باتری‌ها پیشنهاد کردند. واحد ذخیره انرژی مغناطیسی ابرسانا^۱ (SMES) باعث می‌شود که نوسانات توان در خط ارتباطی کاهش پیدا کند. طراحی SMES برای کنترل بار فرکانس دوناحیه‌ای با روش‌های مختلفی انجام شده است. در [۲] از الگوریتم جستجوی الگو^۲ (PS) برای تعیین پارامترهای بهینه SMES و PID در یک سیستم قدرت به هم پیوسته و در [۳] از الگوریتم جستجوی فاخته^۳ (CSA) برای تنظیم پارامترهای کنترلی SMES و کنترل کننده PI در یک سیستم دوناحیه‌ای گرمایی استفاده شده است. در این روش‌ها سعی بر این بوده که انحراف فرکانس در حالت دینامیکی و ماندگار کاهش یابد.

پارامترهای کنترلی SMES و PID در محدوده وسیعی تغییر می‌کنند. به همین دلیل یافتن مقادیر بهینه این پارامترها با استفاده از یادگیری تقویتی، یک فرآیند زمان‌بر است لذا در این مقاله برای کاهش زمان شبیه‌سازی ابتدا مقادیر پارامترهای کنترلی با استفاده از یک الگوریتم هوشمند تا حد امکان به مقادیر بهینه نزدیک می‌شوند. برای این منظور از الگوریتم بهینه‌سازی یادگیری مبتنی بر تدریس - یادگیری اصلاح شده^۴ (MTLBO) استفاده می‌شود. نویسندگان زیادی در مراجع مختلف این الگوریتم را به کار برده‌اند. در [۴] برای جایابی بهینه خازن در شبکه توزیع و در [۵] برای پخش بار اقتصادی با در نظر نقطه اثر درجه واحدهای تولیدی استفاده شده است. ویژگی مهم الگوریتم پیشنهادی این است که هیچ پارامتر کنترلی برای روند بهینه‌سازی ندارد و بنابراین اجرای آن ساده بوده و تنها نیاز به پارامترهای رایج کنترل مانند اندازه جمعیت و تعداد نسل دارد.

یادگیری تقویتی^۵ (RL) به عنوان یک روش کنترلی بهینه مورد توجه قرار گرفته است. در یادگیری تقویتی هدف، پیدا کردن سیاست بهینه یعنی یافتن کنش بهینه در هر حالت از سیستم است. در این روش، عامل سیاست کنترلی بهینه، خود را در تعامل با محیط (سیستم تحت کنترل) یاد می‌گیرد [۶]. یادگیری تقویتی در دهه اخیر در کنترل سیستم‌های قدرت جایگاه خاصی یافته است. در [۷] کاربرد یادگیری تقویتی در مبحث بازار نشان داده شده است. در [۸] اصول به‌کارگیری یادگیری تقویتی در پایداری سیستم قدرت مورد بررسی قرار گرفته است و برای بهبود نوسانات توان عبوری بین دو



شکل ۱: بلوک دیاگرام سیستم قدرت دوناحیه‌ای گرمایی همراه با واحد SMES

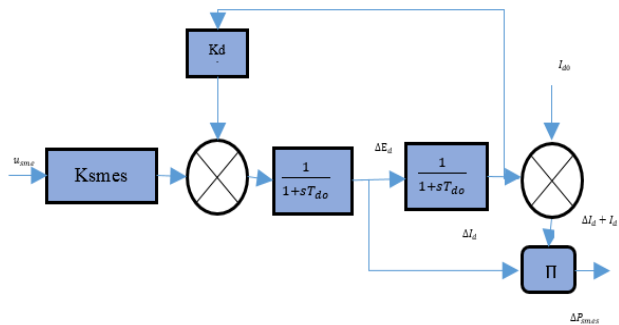
۳- الگوریتم MTLBO برای تنظیم اولیه ضرایب کنترل کننده‌ها

در این بخش ابتدا مروری بر الگوریتم MTLBO انجام می‌شود و سپس نحوه طراحی ضرایب کنترل کننده‌ها با استفاده از آن تشریح می‌گردد.

۳-۱- مروری بر الگوریتم MTLBO

الگوریتم بهینه‌سازی مبتنی بر تدریس - یادگیری یکی از روش‌های قدرتمند و پرتعداد در بهینه‌سازی در سال‌های اخیر، در بسیاری از رشته‌های مهندسی است که توسط راتو و همکاران در [۱۵] معرفی شده است. این روش به مانند اکثر الگوریتم‌های بهینه‌سازی الهام گرفته شده از طبیعت می‌باشد. ابتدا جمعیتی به‌عنوان دانش‌آموزان یک کلاس انتخاب می‌شود و بهترین دانش‌آموز به‌عنوان مدرس این جمعیت انتخاب می‌شود و مدرس سعی دارد با تأثیری که بر یادگیری دانش‌آموزان می‌گذارد، دانش دانش‌آموزان را بهبود بدهد. دانش‌آموزان مطابق با کیفیت آموزش ارائه شده کسب دانش می‌کنند. علاوه بر این، دانش‌آموزان با بحث و تبادل نظر بین خودشان سعی در ارتقای سطح علمی خود دارند. بردار اولیه برای جمعیت دانش‌آموزان طبق رابطه (۲) است که ابعاد این بردار $N \times D$ است که N تعداد دانش‌آموزان و D تعداد موضوعات درسی (متغیرهای مسئله) می‌باشد. این الگوریتم از دو مرحله اصلی و یک مرحله جهش که برای همگرایی بیشتر مطابق با [۱۶] به کار می‌رود، تشکیل شده است که در ادامه، هر فاز توضیح داده شده است.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1D} \\ x_{21} & x_{22} & \dots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{ND} \end{bmatrix} \quad (2)$$



شکل ۲: بلوک دیاگرام واحد SMES

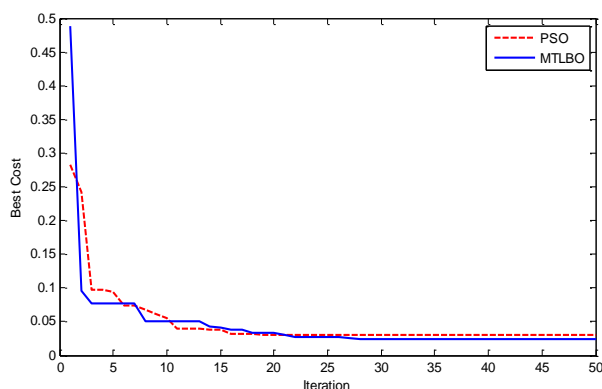
تأثیر باند راکد گاورنر به دامنه انحراف فرکانسی بستگی دارد. اگر انحراف کوچک باشد؛ ممکن است در محدوده باند راکد قرار گیرد و در نتیجه کنترل سرعت غیرفعال خواهد شد. همچنین اثر غیرخطی پس‌زنی مربوط به باند راکد تمایل به ایجاد نوسانات سینوسی ماندگار دارد. تابع تبدیل گاورنر با لحاظ باند راکد به‌صورت رابطه (۱) آورده شده است.

$$G_g = \frac{K_1 + \frac{K_2 s}{\pi}}{1 + sT_g} \quad (1)$$

در این رابطه $K_1 = 0.8$ و $K_2 = -0.2$ مطابق با [۱۳] در نظر گرفته می‌شود. همچنین بلوک دیاگرام کنترلی واحد SMES که در هر ناحیه کنترلی برای کاهش عدم تطابق آنی بین تولید و بار استفاده شده، در شکل ۲ نشان داده شده است. توضیحات کامل در مورد واحد SMES در مرجع [۱۴] موجود است. همچنین پارامترهای مربوط به آن در ضمیمه (ب) داده شده است.

۳-۲- تنظیم ضرایب اولیه PID و SMES

در این بخش الگوریتم MTLBO برای تنظیم پارامترهای PID و SMES به کار می‌رود. در این مقاله یک عضو (دانش‌آموز) جواب مسئله بهینه‌سازی و دروس (متغیرهای مسئله) ضرایب کنترل کننده‌های PID و SMES در نظر گرفته شده‌اند. بهینه‌سازی با هدف مقاوم بودن سیستم برای انواع مقادیر تغییرات بار انجام شده است. معیار خطای ITAE (انتگرال زمان ضرب در مقدار مطلق خطا)، ITSE (انتگرال زمان در مقدار مربع خطا) و IAE (انتگرال مطلق خطا) از جمله توابع هزینه‌ای است که برای بهینه‌سازی استفاده می‌شود [۱۸]. روابط معیار خطاهای استفاده شده در ضمیمه (ج) داده شده است. معیار خطای ITAE زمان نشست را بهتر از دو معیار IAE و ISE کاهش می‌دهد [۱۷]. همچنین مقدار پیک فراجش را کاهش می‌دهد [۱۸]. بنابراین طبق [۱۸] معیار خطای ITAE بهترین تابع هزینه در مطالعات کنترل بار فرکانس را دارا است. در این مقاله از معیار EATI برای بهینه‌سازی استفاده شده است. بنابراین مقادیر کنترل کننده PID و SMES طبق این معیار بهینه شده است. در جدول ۱ مقادیر بهینه شده K_d ، K_i ، K_p ، K_{SMES} و K_{id} آورده شده و با مقادیر بهینه شده آن‌ها توسط الگوریتم PSO مقایسه شده است. در [۱۹] روش کار الگوریتم PSO و پارامترهای آن در ضمیمه (د) آورده شده است. در شکل ۳ نحوه همگرایی تابع هدف در روش مقایسه شده که نشان‌دهنده سرعت بیش‌تر الگوریتم MTLBO در بهبود تابع هدف می‌باشد. این مقادیر برای مقدار تغییر بار ثابت ۱٪ بار نامی در هر دو ناحیه بهینه شده است.



شکل ۳: مقایسه نحوه همگرایی الگوریتم PSO و MTLBO

۴- کنترل تطبیقی مبتنی بر یادگیری تقویتی

۴-۱- الگوریتم یادگیری تقویتی

یادگیری تقویتی روشی است که در آن یک یا چند عامل جهت رسیدن به یک هدف معین یک سیاست بهینه کنترلی را یاد می‌گیرند. سیاست بهینه یادگرفته شده برای انجام این روش باید بهترین کنش را از بین کنش‌های قابل اجرا در هر موقعیت انتخاب کند.

فاز مدرس: مدرس تلاش می‌کند تا میانگین نمره کلاس را به سمت خود بکشد. بردار میانگین برای هر دانش‌آموز به صورت رابطه (۳) است.

$$\mathbf{M}_d = [m_1, m_2, m_3, \dots, m_d] \quad (3)$$

که m_i میانگین دانش‌آموزان برای موضوع i ام است.

اختلاف بین میانگین یک موضوع دانش‌آموزان و مدرس به صورت رابطه (۴) است.

$$\mathbf{M}_{diff} = \text{rand}(0,1) \times [\mathbf{X}_{best} - TF \times \mathbf{M}_d] \quad (4)$$

در رابطه (۴) TF ضریب تدریس است که می‌تواند به صورت تصادفی یک یا دو باشد، $\text{rand}(0,1)$ یک عدد تصادفی بین صفر و یک است.

عضو جدید طبق رابطه (۵) تعیین می‌شود یعنی در صورتی پذیرفته می‌شود که بهتر از عضو قبلی باشد.

$$\mathbf{X}_{new} = \begin{cases} \mathbf{X}_{old} + \mathbf{M}_{diff} & \text{if } f(\mathbf{X}_{old} + \mathbf{M}_{diff}) \leq f(\mathbf{X}_{old}) \\ \mathbf{X}_{old} & \text{if } f(\mathbf{X}_{old} + \mathbf{M}_{diff}) > f(\mathbf{X}_{old}) \end{cases} \quad (5)$$

فاز یادگیرنده: در این مرحله، یادگیرنده دو دانش‌آموز به صورت تصادفی انتخاب می‌کند و تلاش می‌کند دانش خود را به وسیله اثر متقابل دو دانش‌آموز بر همدیگر افزایش دهد. یادگیرنده تقویت می‌شود اگر دانش‌آموزان دانش بیش‌تری تحویل دهند. یادگیرنده طبق روابط (۶) و (۷) اگر وضعیت بهتری نسبت به قبل داشته باشد پذیرفته می‌شود.

$$\mathbf{X}_{learner\ phase} = \begin{cases} \mathbf{X}_i + r(0,1)(\mathbf{X}_i - \mathbf{X}_j) & \text{if } f(\mathbf{X}_i) \leq f(\mathbf{X}_j) \\ \mathbf{X}_i + r(0,1)(\mathbf{X}_j - \mathbf{X}_i) & \text{if } f(\mathbf{X}_i) > f(\mathbf{X}_j) \end{cases} \quad (6)$$

$$\mathbf{X}_{new} = \begin{cases} \mathbf{X}_{learner\ phase} & \text{if } f(\mathbf{X}_{learner\ phase}) \leq f(\mathbf{X}_{old}) \\ \mathbf{X}_{old} & \text{if } f(\mathbf{X}_{learner\ phase}) > f(\mathbf{X}_{old}) \end{cases} \quad (7)$$

در رابطه (۶) i و j دو دانش‌آموز تصادفی است که $i \neq j$ و $r(0,1)$ یک عدد تصادفی بین صفر و یک است.

فاز جهش: در این فاز در هر تکرار یک جهش تولید می‌شود و به صورت رابطه (۸) فرمول‌بندی می‌شود.

$$\mathbf{X}_{mut} = \mathbf{X}_{rand1} + r(0,1)(\mathbf{X}_{rand2} - \mathbf{X}_{rand3}) \quad (8)$$

که \mathbf{X}_{rand1} ، \mathbf{X}_{rand2} و \mathbf{X}_{rand3} سه دانش‌آموز غیرتکراری و تصادفی است که برای یادگیرنده i ام انتخاب می‌شود و $r(0,1)$ یک عدد تصادفی بین صفر و یک است. سپس برای این فاز روابط (۹) و (۱۰) داده می‌شود:

$$\mathbf{X}_{newmut} = \begin{cases} \mathbf{X}_{mut} & \text{if } r_1(0,1) \geq r_2(0,1) \\ \mathbf{X}_i & \text{otherwise} \end{cases} \quad (9)$$

$$\mathbf{X}_{new} = \begin{cases} \mathbf{X}_{newmut} & \text{if } f(\mathbf{X}_{newmut}) \leq f(\mathbf{X}_{old}) \\ \mathbf{X}_{old} & \text{if } f(\mathbf{X}_{newmut}) > f(\mathbf{X}_{old}) \end{cases} \quad (10)$$

که $r_1(0,1)$ و $r_2(0,1)$ دو عدد تصادفی بین صفر و یک است. دانش‌آموز اصلاح‌یافته جدید در صورتی پذیرفته می‌شود که بهتر از دانش‌آموز قبلی باشد.

جدول ۱: مقادیر پارامترهای کنترل کننده‌های PID و SMES و مقایسه عملکرد دو روش بهینه‌سازی

	پارامترهای کنترل کننده‌ها	ITAE	ITSE×۱۰ ^۶	IAE	
MTLBO	K _p	-۳/۳۹۳	۰/۰۰۳۲	۸/۸۰۵۷	۰/۰۰۶۸
	K _i	-۱۰			
	K _d	-۰/۴۱۷			
	K _{smes}	۲/۰۵			
	K _{id}	۰/۲۰۸			
PSO	K _p	-۴/۳۴۸۱	۰/۰۰۳۳	۹/۰۰۵۴	۰/۰۰۷۰
	K _i	-۱۰			
	K _d	-۰/۴۵۹۱			
	K _{smes}	۲/۰۴۸			
	K _{id}	۰/۲۲			

به دلیل اینکه محیط برای عامل یا عامل‌ها شناخته شده نیست برای شناختن محیط باید ابتدا از روش آزمون و خطا استفاده شود. در واقع هدف از یادگیری تقویتی دستیابی به بیشترین پاداش ناشی از اثر متقابل با محیط با روش سعی و خطا است. در یک دسته‌بندی می‌توان یادگیری تقویتی را به دو نوع وابسته به مدل و مستقل از مدل تقسیم‌بندی کرد [۲۰]. در این مقاله از نوع مستقل از مدل استفاده شده است. دلیل انتخاب این روش برای بهینه کردن عدم وابستگی این روش به مدل سیستم و سادگی این روش در پیاده‌سازی است. در یادگیری تقویتی، محیط به حالت‌های محدودی تقسیم می‌شود که آن را با مجموعه $\{s\}$ نشان می‌دهند. یادگیری تقویتی با انتخاب هر کنش از مجموعه $\{a\}$ در هر گام t و با دریافت پاداش R حالت جدیدی (s') را تجربه می‌کند و این روش تا زمانی که عامل یا عامل‌ها به هدف مورد نظر برسند ادامه پیدا می‌کند. در حقیقت هدف پیدا کردن سیاست بهینه‌ای است که پاداش طولانی‌مدت کاهش یافته را بیشینه کند. پاداش طولانی‌مدت کاهش یافته به صورت رابطه (۱۱) داده می‌شود.

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (11)$$

که در آن γ ضریب کاهش نامیده می‌شود و بین صفر و یک است. یادگیری تقویتی دارای یک تابع ارزش است که با Q در یادگیری Q نشان می‌دهند و به صورت رابطه (۱۲) داده می‌شود.

$$Q^{\pi}(s, a) = E_{\pi} \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a \right\} \quad (12)$$

در این رابطه π سیاست بهینه اتخاذ شده r پاداش و a کنش انتخابی است. سیاست باید طوری انتخاب شود که مقدار Q بیشینه شود. رابطه (۱۲) باید در هر گام در تعامل با محیط به‌هنگام شود. رابطه (۱۳) فرمول به‌هنگام‌رسانی است که به آن رابطه بهینه بلمن گفته می‌شود.

$$\Delta Q = \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right] \quad (13)$$

در رابطه (۱۳) α عددی بین صفر و یک است و ضریب تضعیف نامیده می‌شود. همچنین برای انتخاب کنش بهینه ابتدا به صورت خارج از خط از سیاست شبه‌حریصانه استفاده می‌شود [۲۱]. در سیاست شبه‌حریصانه عامل یادگیر به احتمال $\epsilon - 1$ کنشی که دارای بیشترین

مقدار می‌باشد و به احتمال ϵ یک کنش از بین کنش‌های تعریف شده در هر موقعیت را انتخاب می‌کند. در این سیاست ϵ عددی بین صفر و یک است. در شکل ۴ فلوچارت مربوط به روش یادگیری Q خارج از خط داده شده است. بعد از اتمام یادگیری خارج از خط، یادگیری به صورت برخط با سیاست حریصانه بر روی سیستم اعمال می‌شود. در سیاست حریصانه عامل فقط کنشی که بیشترین مقدار را در ماتریس Q دارد، انتخاب می‌کند. برای جزئیات بیشتر به [۲۲] مراجعه فرمایید.

۴-۲- عناصر یادگیری تقویتی به‌کاررفته

برای سیستم مورد مطالعه دو عامل برای یادگیری در نظر گرفته شده است. یک عامل برای تغییر ضرایب کنترل کننده ناحیه یک و یک عامل هم به همین ترتیب برای ناحیه دو لحاظ شده است؛ که این عامل‌ها برای رسیدن به بیشترین پاداش مربوط به ناحیه خود و به‌طور مستقل از هم تلاش می‌کنند. در ادامه به تعریف مجموعه عمل‌ها، حالت‌ها، پاداش هر حالت، α ، γ و ϵ برای عامل پرداخته می‌شود.

مجموعه کنش‌ها: تعریف مجموعه کنش‌ها از اهمیت به‌سزایی برخوردار است و طوری باید تعریف شود که باعث ناپایداری سیستم در همه حالت‌ها نشود. هر عامل با تغییر ضرایب K_p ، K_i ، K_d ، K_{smes} و K_{id} سعی در یادگیری محیط را دارد. در این مقاله برای هر کدام از ضرایب کنترل کننده K_p ، K_i ، K_d و K_{smes} پنج کنش $\{0/9, 1, 1/1, 1/2, a\}$ در نظر گرفته شده است. که این کنش‌ها به صورت ضرب در مقدار خروجی کنترل کننده‌های مورد نظر اعمال می‌شوند. به دلیل اینکه یادگیری تقویتی تک‌عاملی بر روی هر ناحیه از سیستم مورد مطالعه اعمال شده است؛ کنش‌ها به صورت رابطه (۱۳) در نظر گرفته می‌شود.

به‌طور نمونه اگر عامل موجود در ناحیه یک دارای ضرایب کنترل کننده $(K_p, K_i, K_d, K_{smes}) = (2, 1/5, 1, 5)$ باشد با اعمال کنش $a = (1/1, 1, 1/2, 0/9)$ ضرایب کنترل کننده در ناحیه یک به صورت $(K_p, K_i, K_d, K_{smes}) = (2/2, 1/5, 1/2, 4/5)$ تغییر پیدا می‌کند.

$$\{a\} = \begin{bmatrix} 0/9 & 0/9 & 0/9 & 0/9 \\ 0/9 & 0/9 & 0/9 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix} \quad (13)$$

مجموعه حالت‌ها: با توجه به هدف سیستم کنترل که کاهش نوسانات فرکانسی و توان خط ارتباطی و در نتیجه افزایش پایداری سیستم است از $\Delta \omega$ یا ΔP_{tie} و یا مشتقات این مقادیر می‌توان به‌عنوان محیط یادگیری عامل استفاده کرد. در این مقاله از $\Delta \omega$ و $\Delta \dot{\omega}$ مربوط به ناحیه‌ای که عامل در آن وجود دارد؛ به‌عنوان محیط عامل انتخاب شده و در تمامی روابط منظور از $\Delta \omega$ ، تغییرات ω در ناحیه‌ای که عامل موجود است، می‌باشد. برای انتخاب بهتر کنش‌ها و یادگیری بیشتر محیط برای هر دو حالت $\Delta \dot{\omega} > 0$ و $\Delta \dot{\omega} \leq 0$ در بازه $[0.05, -0.25]$ ، حالت به صورت محدود بازه‌ای برابر ۲ حالت برای زمانی که منحنی از این محدوده خارج می‌شود، به صورت زیر در نظر گرفته شده است. در کل با توجه به رابطه (۱۳) 2004 حالت برای هر

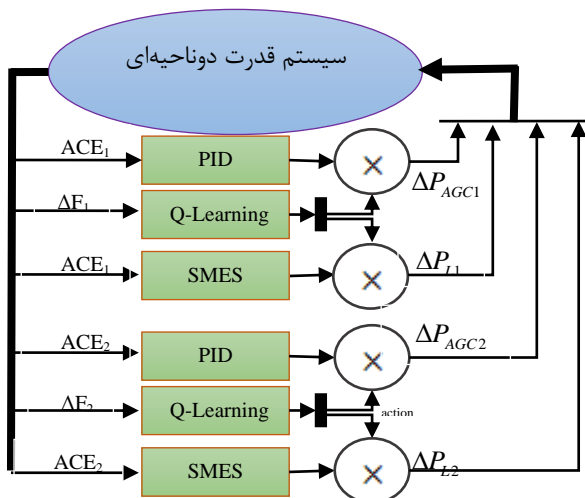
پاداش: در این مقاله چون هدف بهبود عملکرد کنترل کننده $\Delta\omega$ است که توسط الگوریتم MTLBO طراحی شده است؛ بنابراین پاداش‌ها هم متناسب با این هدف در نظر گرفته شده‌اند و هر موقع عامل کنشی را انجام داد که نتیجه بهتری نسبت به کنترل کننده طراحی شده با الگوریتم پیشنهادی داد پاداش مثبتی برای عامل در نظر گرفته می‌شود. بزرگ بودن این پاداش متناسب است با بهتر بودن نتیجه $\Delta\omega$ که گرفته می‌شود. همچنین با ضرب کردن در P که طبق روابط (۱۵) و (۱۶) برای اینکه در نتیجه نسبی بهتر در فاصله‌های بیش‌تر از صفر پاداش بیش‌تری داشته باشد، انجام شده است؛ تابع پاداش برای این سیستم به صورت زیر در نظر گرفته می‌شود:

$$s = \begin{cases} 1 & \text{if } \Delta\omega < -0.025 \ \& \ \Delta\dot{\omega} > 0 \\ 2 & \text{if } -0.025 \leq \Delta\omega < -0.024975 \ \& \ \Delta\dot{\omega} > 0 \\ \vdots & \\ 1001 & \text{if } 0.00497 \leq \Delta\omega < 0.005 \ \& \ \Delta\dot{\omega} > 0 \\ 1002 & \text{if } \Delta\omega \geq 0.005 \ \& \ \Delta\dot{\omega} > 0 \\ 1003 & \text{if } \Delta\omega < -0.025 \ \& \ \Delta\dot{\omega} \leq 0 \\ 1004 & \text{if } -0.025 \leq \Delta\omega < -0.02497 \ \& \ \Delta\dot{\omega} \leq 0 \\ \vdots & \\ 2003 & \text{if } 0.00497 \leq \Delta\omega < 0.005 \ \& \ \Delta\dot{\omega} \leq 0 \\ 2004 & \text{if } \Delta\omega \geq 0.005 \ \& \ \Delta\dot{\omega} \leq 0 \end{cases} \quad (14)$$

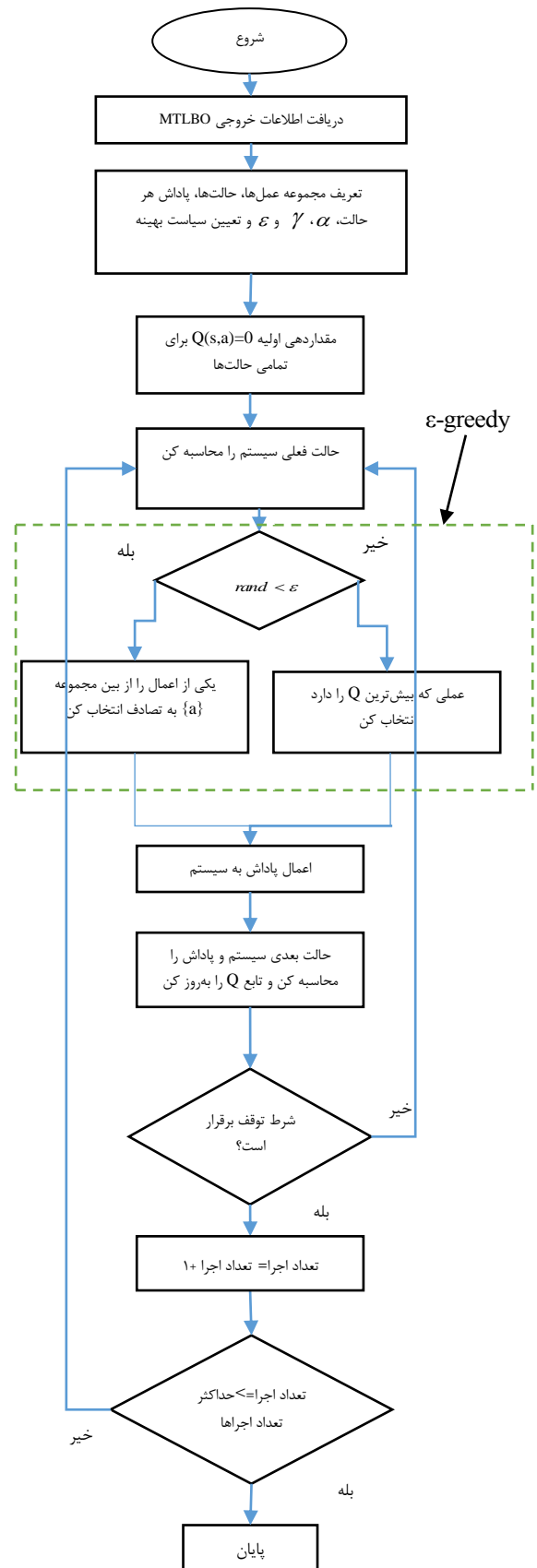
$$P = 10^5 \left| |\Delta\omega_{TLBO}| - |\Delta\omega_{RL}| \right| \quad (15)$$

$$r = \begin{cases} -100 & \text{if } \Delta\omega_{RL} > 0.005 \ \text{or } \Delta\omega_{RL} < -0.025 \\ P(\Delta\omega_{RL} - \Delta\omega_{TLBO}) & \text{elseif } \Delta\omega_{TLBO} > 0 \ \& \ \Delta\omega_{RL} > 0 \\ P(\Delta\omega_{TLBO} - \Delta\omega_{RL}) & \text{elseif } \Delta\omega_{TLBO} < 0 \ \& \ \Delta\omega_{RL} < 0 \\ -P \frac{1}{|\Delta\omega_{RL}| + 0.001} & \text{elseif } |\Delta\omega_{TLBO}| < |\Delta\omega_{RL}| \\ P \frac{1}{|\Delta\omega_{RL}| + 0.001} & \text{elseif } |\Delta\omega_{TLBO}| > |\Delta\omega_{RL}| \end{cases} \quad (16)$$

پارامترها: α ، ϵ و γ به ترتیب ۰/۰۳، ۰/۰۶ و ۰/۹۷ در نظر گرفته شده است. در شکل ۵ نحوه اعمال یادگیری به سیستم دوحاحیه گرمایی، نشان داده شده است.



شکل ۵: نحوه اعمال RL به سیستم قدرت مورد مطالعه



شکل ۴: فلوچارت الگوریتم یادگیری Q خارج از خط

عامل در نظر گرفته شده است.

۵- نتایج شبیه‌سازی

شبیه‌سازی‌ها به کمک نرم‌افزار MATLAB صورت گرفته است. برای یادگیری کامل محیط توسط عامل برای هر یک از ناحیه‌های سیستم مورد مطالعه افزایش بار به صورت پله‌ای دامنه محدود و تصادفی به مدت ۱۰۰ ثانیه داده شده است. تغییر در عمل‌های عامل‌های ۱ و ۲ در شکل‌های ۶ و ۷ برای تغییر بار ۳٪ مقدار نامی در ناحیه یک نشان داده شده است. نتایج شبیه‌سازی برای این تغییر بار در شکل‌های ۸، ۹ و ۱۰ نشان از این دارند که عامل مربوط به کنترل کننده ناحیه یک توانسته انحرافات فرکانسی مربوط به ناحیه یک را به خوبی میرا کند و عامل ۲ در کنترل کننده ناحیه ۲ نیز تأثیر بهتری بر سیگنال کنترلی نسبت به کنترل کننده‌های مقایسه شده را دارد. همچنین تغییر بار ۲٪ و ۱٪ مقدار نامی به ترتیب برای ناحیه یک و دو اعمال شده است و نتایج در شکل‌های ۱۱، ۱۲ و ۱۳ نشان می‌دهند که عامل‌های یادگیر باعث افزایش مقاوم بودن سیستم برای انواع تغییرات بار شده است. سه تغییر بار نیز مورد بررسی آماری قرار گرفته و معیارهای ITAE، JSTE، JSE، OS و US برای تغییرات $\Delta\omega$ در هر ناحیه و توان عبوری از خط ارتباطی در جدول ۲ محاسبه شده‌اند. داده‌های جدول ۲ نشان می‌دهند که روش پیشنهادی یادگیری تقویتی، یک روش کارآمد برای میرا کردن انواع انحرافات بار و کارایی بیشتری از روش‌های دیگر مبتنی بر الگوریتم‌های بهینه‌سازی داشته است.

۶- نتیجه‌گیری

هدف از این مقاله طراحی کنترل کننده بار فرکانس مقاوم و تطبیقی برای یک سیستم قدرت چندناحیه‌ای مجهز به SMES است. در این راستا ابتدا الگوریتم MTLBO برای تنظیم اولیه پارامترهای کنترل کننده هماهنگ با SMES به کار گرفته شده است. MTLBO، یک الگوریتم فراابتکاری بر مبنای تدریس - یادگیری می‌باشد که مهم‌ترین ویژگی آن، سادگی و سرعت بالای آن است. در ادامه به منظور بهبود عملکرد بلادرنگ آن، یک روش کنترلی مبتنی بر یادگیری تقویتی پیشنهاد شده است. روش پیشنهادی مبتنی بر یادگیری تقویتی پیوسته تصمیمات خود را ارزیابی کرده و دانش خود را از محیط تحت کنترل به‌هنگام می‌کند، بنابراین رفتاری انطباقی دارد. روش کنترلی پیشنهادی برای کنترل، نیاز به فرض اولیه قوی نسبت به سیستم ندارد و بدون دانش اولیه می‌تواند به سادگی هر سیستمی را کنترل کند. نتایج حاصل از شبیه‌سازی، عملکرد بسیار خوب روش کنترلی پیشنهادی را در مقایسه با طرح‌های کنترلی سنتی نشان می‌دهد. همچنین در ادامه برای برجسته کردن توانایی کنترل کننده مبتنی بر یادگیری تقویتی، شاخص‌های عملکرد حوزه زمان محاسبه شده و مقایسه شده‌اند که برتری روش کنترل پیشنهادی را ثابت می‌کنند.

ضمائم

ضمیمه الف

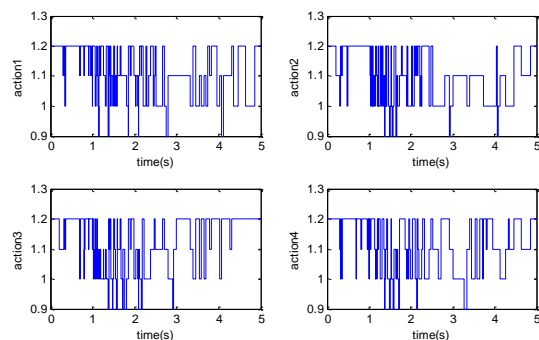
پارامترهای سیستم مورد مطالعه

- $f = 60 \text{ Hz}$.
- $TH1, TH2 = 0.08 \text{ s}$.
- $TT1, TT2 = 0.3 \text{ s}$.
- $Tr1, Tr2 = 10 \text{ s}$.
- $T12 = 0.086 \text{ pu MW/rad}$.
- $Tp1, Tp2 = 20 \text{ s}$.
- $Kr1, Kr2 = 0.5$
- $KP1, KP2 = 120 \text{ Hz/pu MW}$.
- $R1, R2 = 2/4 \text{ Hz/MW}$.
- $B1, B2 = 0.425 \text{ MW/Hz}$.

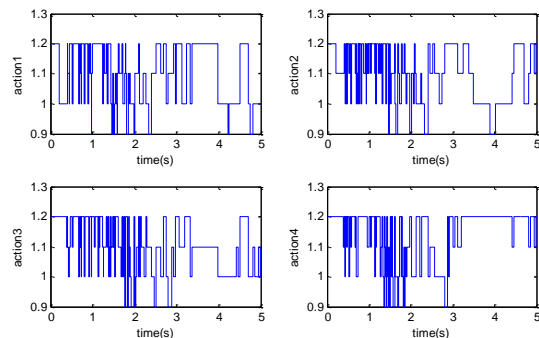
ضمیمه ب

پارامترهای واحد SMES

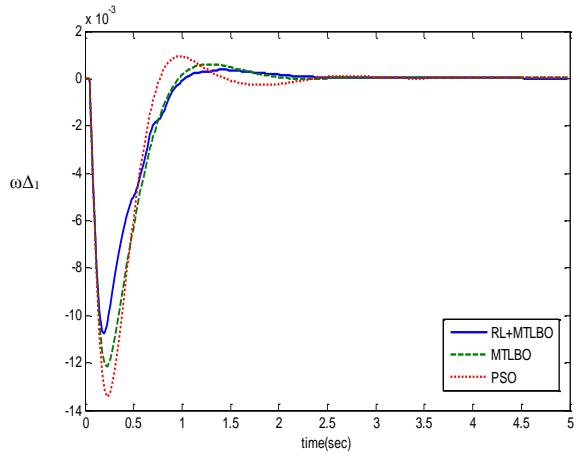
- $L = 2/65 \text{ H}$
- $TDC = 0.3 \text{ s}$
- $KSMES = 100 \text{ kV/unit MW}$
- $Kid = 0.2 \text{ kV/kA}$
- $I_{d0} = 4/5 \text{ kA}$



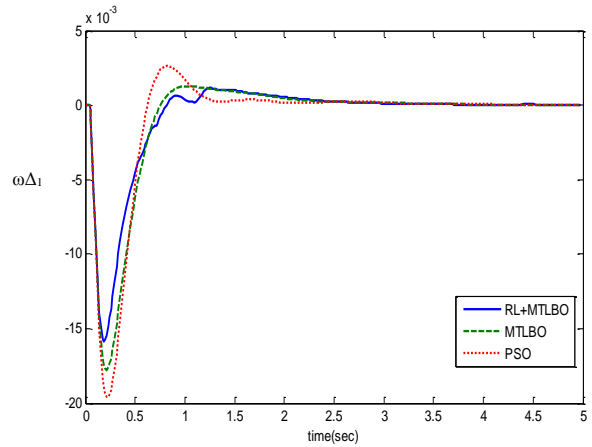
شکل ۶: نحوه تغییر ضریب مربوط به هر یک از کنترل کننده‌های ناحیه یک برای تغییر بار ۳ درصد در ناحیه یک



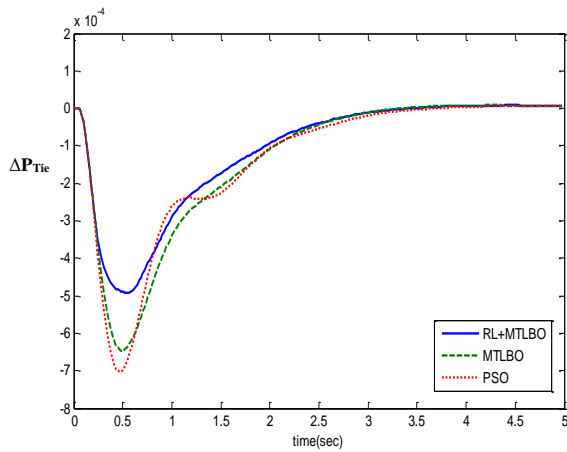
شکل ۷: نحوه تغییر ضریب مربوط به هر یک از کنترل کننده‌های ناحیه دو برای تغییر بار ۳ درصد در ناحیه یک



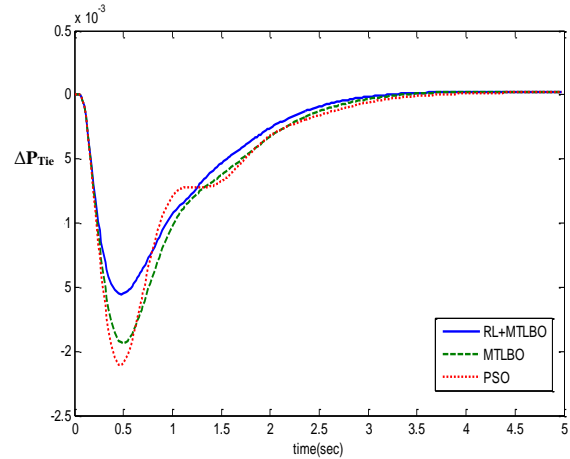
شکل ۱۱: تغییرات فرکانس در ناحیه ۱ برای تغییر بار ۲ درصد در ناحیه ۱ و تغییر بار ۱ درصد در ناحیه ۲



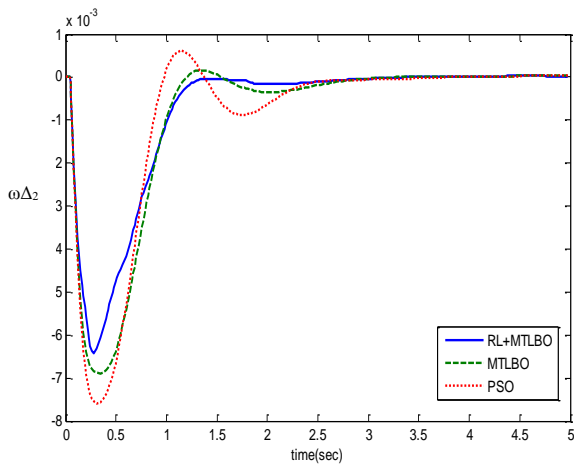
شکل ۸: تغییرات فرکانس در ناحیه ۱ برای تغییر بار ۳ درصد در ناحیه ۱



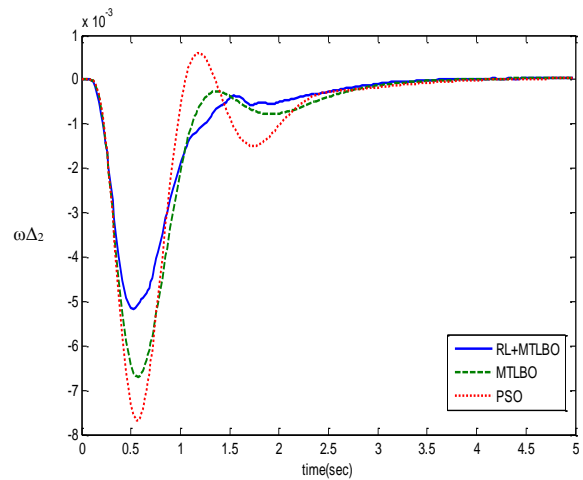
شکل ۱۲: تغییرات توان در خط ارتباطی برای تغییر بار ۲ درصد در ناحیه ۱ و تغییر بار ۱ درصد در ناحیه ۲



شکل ۹: تغییرات توان در خط ارتباطی برای تغییر بار ۳ درصد در ناحیه ۱



شکل ۱۳: تغییرات فرکانس در ناحیه ۲ برای تغییر بار ۲ درصد در ناحیه ۱ و تغییر بار ۱ درصد در ناحیه ۲



شکل ۱۰: تغییرات فرکانس در ناحیه ۲ برای تغییر بار ۳ درصد در ناحیه ۱

جدول ۲: مقایسه معیارهای ارزیابی حوزه زمان برای روش‌های کنترلی

معیار		$\Delta P_{L1} = 0.1 \text{ p.u}$ $\Delta P_{L2} = 0.0 \text{ p.u}$	$\Delta P_{L1} = 0.0 \text{ p.u}$ $\Delta P_{L2} = 0.2 \text{ p.u}$	$\Delta P_{L1} = 0.2 \text{ p.u}$ $\Delta P_{L2} = 0.1 \text{ p.u}$	
RL+MTLBO	ITAE	0.030	0.059	0.088	
	ITSE $\times 10^6$	2/5769	9/7116	27/145	
	IAE	0.043	0.083	0.137	
	Peak overshoot $\times 10^{-2}$	ΔF_1	0/4474	0/203	0/093
		ΔF_2	0/008	0/6906	0/3390
		ΔP_{Tie}	0/0072	1/0237	0/5154
	Peak undershoot $\times 10^{-2}$	ΔF_1	-5/4175	-3/4526	-1/7583
		ΔF_2	-1/8213	-10/7279	-5/3602
		ΔP_{Tie}	-0/5331	-0/0141	-0/0076
MTLBO	ITAE	0.034	0.068	0.100	
	ITSE $\times 10^6$	3/5615	14/251	40/606	
	IAE	0.048	0.097	0.161	
	Peak overshoot $\times 10^{-2}$	ΔF_1	0/4686	0/209	0/104
		ΔF_2	0/0099	0/8289	0/4185
		ΔP_{Tie}	0/0081	1/2912	0/6456
	Peak undershoot $\times 10^{-2}$	ΔF_1	-5/9195	-4/6642	-2/2321
		ΔF_2	-2/2227	-11/8423	-5/9245
		ΔP_{Tie}	-0/6457	-0/0162	-0/0081
PSO	ITAE	0.043	0.087	0.123	
	ITSE $\times 10^6$	7/9740	31/872	87/307	
	IAE	0.049	0.099	0.161	
	Peak overshoot $\times 10^{-2}$	ΔF_1	0/8731	0/3982	0/1989
		ΔF_2	0/1991	1/7471	0/8720
		ΔP_{Tie}	0/0082	1/4033	0/7018
	Peak undershoot $\times 10^{-2}$	ΔF_1	-6/5470	-5/1127	-2/5565
		ΔF_2	-2/5550	-13/0961	-6/5396
		ΔP_{Tie}	-0/7020	-0/0167	-0/0083

$$ITSE = \int_0^{t_{sim}} t \cdot [|\Delta f_1|^2 + |\Delta f_2|^2 + |\Delta P_{Tie}|^2] dt$$

ضمیمه ج

روابط معیار خطاهای استفاده شده در این مقاله

در این روابط t_{sim} زمان شبیه‌سازی است.

ضمیمه د

پارامترهای الگوریتم بهینه‌سازی PSO

C1=2

C2=2

W=0/99

$$ITAE = \int_0^{t_{sim}} t [|\Delta f_1| + |\Delta f_2| + |\Delta P_{Tie}|] dt$$

$$ISE = \int_0^{t_{sim}} [|\Delta f_1|^2 + |\Delta f_2|^2 + |\Delta P_{Tie}|^2] dt$$

$$IAE = \int_0^{t_{sim}} [|\Delta f_1| + |\Delta f_2| + |\Delta P_{Tie}|] dt$$

مراجع

- [12] D. Fuselli, F. D. Angelis, M. Boaro, D. Liu, Q. Wei, S. Squartini and F. Piazza, "Action dependent heuristic dynamic programming for home energy resource scheduling," *International Journal of Electrical Power and Energy Systems*, vol. 48, pp. 148-160, 2013.
- [13] R. K. Sahu, S. P. Panda and U. K. Rout, "DE optimized parallel 2-DOF PID controller for load frequency control of power system with governor dead-band nonlinearity," *Electrical Power and Energy Systems*, vol. 49, pp. 19-33, 2013.
- [14] R. J. Abraham, D. Das and A. Patra, "Automatic generation control of an interconnected hydrothermal power system considering superconducting magnetic energy storage," *Electrical Power and Energy Systems*, vol. 29, pp. 571-579, 2007.
- [15] R. V. Rao, V. J. Savsani and D. P. Vakharia, "Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems," *Comput-Aided Design*, vol. 43, pp. 303-15, 2011.
- [16] H. Hosseinpour, T. Niknam and S. I. Taheri, "A modified TLBO algorithm for placement of AVR's considering DG's," *In the Proceeding of Power System Conference*, pp. 1-8, 2011.
- [17] H. Shabani, B. Vahidi and M. Ebrahimpour, "A robust PID controller based on imperialist competitive algorithm," *ISA Transactions*, vol. 52, pp. 88-95, 2013.
- [18] R. K. Sahu, S. Panda and P. C. Pradhan, "Design and analysis of hybrid firefly algorithm-pattern search based fuzzy PID controller for LFC of multi area power systems," *Electrical Power and Energy Systems*, vol. 69, pp. 200-212, 2015.
- [۱۹] محمد مؤمنی، مهدی آقاصرام، وحید شاکر، شهرام جمالی و مهدی نوشیار، «ارائه یک فیلتر جدید برای حذف نویزهای ضربه‌ای و ترکیب فیلتر پیشنهادی با الگوریتم PSO به منظور کشف و دفاع در برابر حملات سیل‌آسای SYN»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۶، شماره ۱، ۱۳۹۵.
- [20] D. Ernst, M. Glavic and L. Wehenkel, "Power systems stability control: reinforcement learning framework," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 427-435, 2004.
- [21] F. Lewis and D. Liu, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, Wiley-IEEE Press, 2013.
- [22] A. Gosavi, *Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning*, Dordrecht, Kluwer Academic Publishers, 2003.
- [1] P. Kundur, *Power System Stability and Control*, McGraw-Hill, New York, 1994.
- [2] M. Farahani and S. Ganjefar, "Solving LFC problem in an interconnected power system using superconducting magnetic energy storage," *Physica C*, vol. 487, pp. 60-66, 2013.
- [3] S. Chaîne and M. Tripathy, "Design of an optimal SMES for automatic generation control of two-area thermal power system using Cuckoo search algorithm," *Journal of Electrical Systems and Information Technology*, pp. 1-13, 2015.
- [۴] بلال محمدی کله سر، رؤف حسن‌پور، احمد فلاح خوشبخت و فرزاد مرتضوی، «مکان‌یابی خازن در شبکه توزیع با در نظر گرفتن مدل بار ثابت و مؤثر با الگوریتم مبتنی بر تدریس-یادگیری اصلاح‌یافته»، بیست و هشتمین کنفرانس بین‌المللی برق/ایران، ۷ صفحه، تهران، ۱۳۹۲.
- [5] V. Sharmiladeve, K. Geetha, S. Geethanjali and K. Keerthivasan, "Modified teaching learning based optimisation for economic dispatch considering valve point effects," *Australian Journal of Basic and Applied Sciences*, vol. 9, pp. 176-182, 2015.
- [6] L. Bushoniu, D. Ernst, B. D. Schutter and R. Babuska, "Approximate reinforcement learning: an overview," *In the Proceedings of the IEEE Symposium on Programming and Reinforcement Learning*, pp. 1-8, 2011.
- [7] V. Naduri and T. K. Das, "A reinforcement learning model to assess market power under auction-based energy pricing," *IEEE Transaction on Power Systems*, vol. 22, no. 1, pp. 85-95, 2007.
- [8] D. Ernst, M. Glavic and L. Wehenkel, "Power system stability control: Reinforcement learning framework," *IEEE Transaction on Power Systems*, vol. 19, no. 1, pp. 427-435, 2004.
- [9] J. G. Vlachogiannis and N. D. Hatziargyriou, "Reinforcement learning for reactive power control," *IEEE Transaction on Power Systems*, vol. 19, no. 3, pp. 1317-1325, 2004.
- [10] T. Huang and D. Liu, "A self-learning scheme for residential energy system control and management," *Neural Computing and Applications*, vol. 22, pp. 259-269, 2013.
- [11] M. Boaro, D. Fuselli, F. D. Angelis, D. Liu, Q. Wei and F. Piazza, "Adaptive dynamic programming algorithm for renewable energy scheduling and battery management," *Cognitive Computation*, vol. 5, pp. 264-277, 2013.

زیرنویس‌ها

⁵ Reinforcement Learning⁶ Thyristor Controlled Series Capacitor⁷ Constrained Load Flow⁸ Particle Swarm Optimization¹ Superconducting Magnetic Energy Storage² Pattern Search³ Cuckoo search algorithm⁴ Modified Teaching-Learning Based Optimization