کنترل‌کننده مقاوم تطبیقی بار فرکانس مبتنی بر یادگیری تقویتی برای یک سیستم قدرت به‌هم‌پیوسته شامل SMES

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

۱، ۲، ۳ و ۴- گروه مهندسی برق - دانشکده فنی و مهندسی - دانشگاه محقق اردبیلی - اردبیل - ایران
[۱]akbarimajd@gmail.com, [۲]hshayeghi@gmail.com, [۳]hamidmohammadnejad70@gmail.com, [۴]a.younesi@ieee.org

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

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

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

A. Akbarimajd[1], Associate Professor; H. Shayeghi[2], Professor, H. Mohammad Nejad[3], MSc Student, A. Younesi[4], PhD Student

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

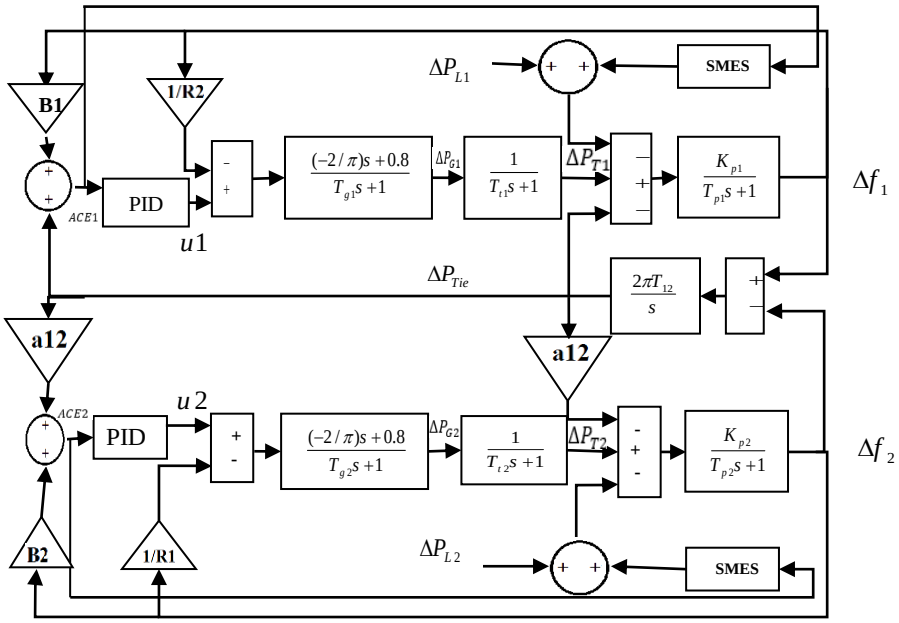
سیستم قدرت از اجزای مختلفی تشکیل شده است که انرژی الکتریکی را در ابعاد وسیعی انتقال می‌دهد. زمانی بین تولید و مصرف تعادل برقرار می‌شود که فرکانس سیستم به‌خوبی کنترل شود. اگر مقدار تولید توان از مقدار مصرف کم‌تر باشد فرکانس سیستم کاهش می‌یابد. هرچه سیستم بزرگ‌تر باشد؛ تغییرات بار، اثر کم‌تری روی فرکانس سیستم می‌گذارد [۱]. هدف از طراحی کنترل‌کننده 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 باطری‌ها پیشنهاد کردند. واحد ذخیره انرژی مغناطیسی ابررسانا[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 به‌هم‌پیوسته و در [۳] از الگوریتم جستجوی فاخته[۳] (CSA) برای تنظیم پارامترهای کنترلی SMES و کنترل‌کننده 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 اصلاح‌شده[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 دریچه 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.

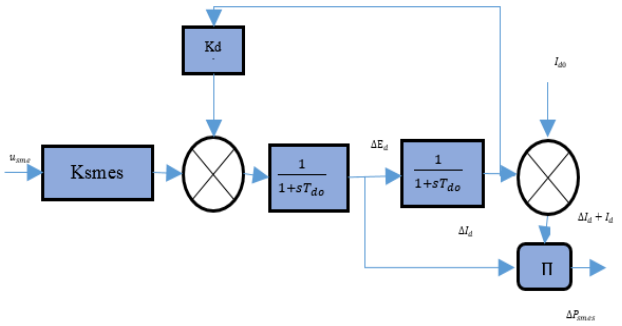
یادگیری تقویتی[۵] (RL) به‌عنوان یک روش کنترلی بهینه مورد توجه قرار گرفته است. در یادگیری تقویتی هدف، پیدا کردن سیاست بهینه یعنی یافتن کنش بهینه در هر حالت از سیستم است. در این روش، عامل سیاست کنترلی بهینه، خود را در تعامل با محیط (سیستم تحت کنترل) یاد می‌گیرد [۶]. یادگیری تقویتی در دهه اخیر در کنترل 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 ناحیه سیستم چهارماشینه کارایی این روش در کنترل TCSC[۶]، 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 [۱۱،۱۲]، کنترل بهینه منابع تجدیدپذیر برای سیستم‌های ریزشبکه هوشمند توسط روش یادگیری Q انجام شده است. هدف این مقاله، طراحی هماهنگ کنترل‌کننده PID با SMES با استفاده از الگوریتم MTLBO و بهبود عملکرد آن با استفاده از روش یادگیری تقویتی به‌منظور کم کردن نوسانات فرکانس ناشی از تغییرات بار و غلبه بر تأخیر زمانی ناشی از گاورنر می‌باشد. چالش اصلی در طراحی کنترل‌کننده‌ها با استفاده از یادگیری Q آن است که فضای جستجوی ماتریس **Q** بزرگ است؛ درنتیجه در مقالاتی که با استفاده از روش یادگیری Q به تنظیم پارامترهای کنترل‌کننده 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 اصلاح‌شده 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 کنترل‌کننده و $R1$ و $R2$ نشان‌دهنده ضریب تنظیم سرعت گاورنر به p.u.Hz؛ T_{G1} و T_{G2} ثابت سرعت گاورنر؛ ΔP_{G1} و ΔP_{G2} تغییر در مقادیر توان گاورنر؛ T_{T1} و T_{T2} ثابت زمانی توربین؛ ΔP_{T1} و ΔP_{T2} تغییر در مقدار توان توربین؛ ΔP_{Tie} تغییر در توان خط ارتباطی؛ K_{PS1} و K_{PS2} ضریب مقاومت سیستم قدرت؛ T_{PS1} و T_{PS2} ثابت زمانی سیستم قدرت؛ T_{12} ضریب سنکرون و Δf_1 و Δf_2 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.



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



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.

$$G_g = \frac{K_1 + \frac{K_2}{\pi}s}{1+sT_g} \qquad (۱)$$

در این رابطه K_1=۰/۸ و K_2=-۰/۲ 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- الگوریتم MTLBO برای تنظیم اولیه ضرایب کنترل‌کننده‌ها

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

۳-۱- مروری بر الگوریتم MTLBO

الگوریتم بهینه‌سازی مبتنی بر تدریس - یادگیری یکی از روش‌های قدرتمند و پرطرفدار در بهینه‌سازی در سال‌های اخیر، در بسیاری از رشته‌های مهندسی است که توسط رائو و همکاران در [۱۵] معرفی شده است. این روش به مانند اکثر الگوریتم‌های بهینه‌سازی الهام‌گرفته‌شده 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 ارائه‌شده 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 $N*D$ 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.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1D} \\ x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{ND} \end{bmatrix} \qquad (۲)$$

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

$$\mathbf{M}_{\mathrm{d}}=[m_1,m_2,m_3,...,m_d] \quad (۳)$$

که m_i میانگین دانش‌آموزان برای موضوع iام است.

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

$$\mathbf{M}_{\mathrm{diff}}=\mathrm{rand}(0,1)\times[\mathbf{X}_{\mathrm{best}}-TF\times\mathbf{M}_{\mathrm{d}}] \quad (۴)$$

در رابطه (۴) TF ضریب تدریس است که می‌تواند به‌صورت تصادفی یک یا دو باشد، $\mathrm{rand}(0,1)$ 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.

$$\mathbf{X}_{\mathrm{new}}=\begin{cases}\mathbf{X}_{\mathrm{old}}+\mathbf{M}_{\mathrm{diff}} \; if \; f(\mathbf{X}_{\mathrm{old}}+\mathbf{M}_{\mathrm{diff}})\le f(\mathbf{X}_{\mathrm{old}})\\ \mathbf{X}_{\mathrm{old}} \quad if \; f(\mathbf{X}_{\mathrm{old}}+\mathbf{M}_{\mathrm{diff}})>f(\mathbf{X}_{\mathrm{old}})\end{cases} \quad (۵)$$

فاز یادگیرنده: در این مرحله، یادگیرنده دو دانش‌آموز به‌صورت تصادفی انتخاب می‌کند و تلاش می‌کند دانش خود را به‌وسیله 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.

$$\mathbf{X}_{\mathrm{learner\ phase}}=\begin{cases}\mathbf{X}_{\mathrm{i}}+r(0,1)(\mathbf{X}_{\mathrm{i}}-\mathbf{X}_{\mathrm{j}})\\ if \; f(\mathbf{X}_{\mathrm{i}})\le f(\mathbf{X}_{\mathrm{j}})\\ \mathbf{X}_{\mathrm{i}}+r(0,1)(\mathbf{X}_{\mathrm{j}}-\mathbf{X}_{\mathrm{i}})\\ if \; f(\mathbf{X}_{\mathrm{i}})>f(\mathbf{X}_{\mathrm{j}})\end{cases} \quad (۶)$$

$$\mathbf{X}_{\mathrm{new}}=\begin{cases}\mathbf{X}_{\mathrm{learner\ phase}} \; if \; f(\mathbf{X}_{\mathrm{learner\ phase}})\le f(\mathbf{X}_{\mathrm{old}})\\ \mathbf{X}_{\mathrm{old}} \; if \; f(\mathbf{X}_{\mathrm{learner\ phase}})>f(\mathbf{X}_{\mathrm{old}})\end{cases} \quad (۷)$$

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 بردار جهش تولید می‌شود و به‌صورت رابطه (۸) فرمول‌بندی می‌شود.

$$\mathbf{X}_{\mathrm{mut}}=\mathbf{X}_{\mathrm{rand1}}+r(0,1)(\mathbf{X}_{\mathrm{rand2}}-\mathbf{X}_{\mathrm{rand3}}) \quad (۸)$$

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

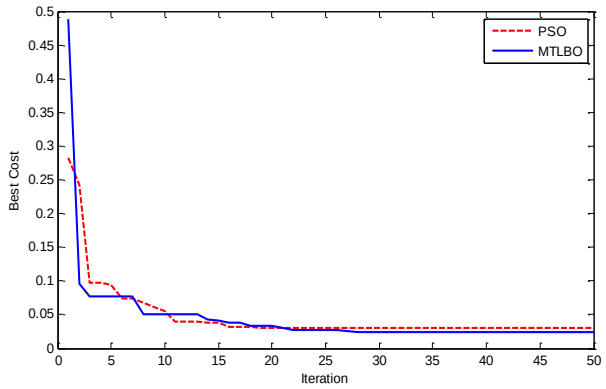
$$\mathbf{X}_{\mathrm{newmut}}=\begin{cases}\mathbf{X}_{\mathrm{mut}} \; if \; r_1(0,1)\ge r_2(0,1)\\ \mathbf{X}_{\mathrm{i}} \quad \mathrm{otherwise}\end{cases} \quad (۹)$$

$$\mathbf{X}_{new}=\begin{cases}\mathbf{X}_{newmut} \; if \; f(\mathbf{X}_{newmut})\le f(\mathbf{X}_{old})\\ \mathbf{X}_{old} \quad if \; f(\mathbf{X}_{newmut})>f(\mathbf{X}_{old})\end{cases} \quad (۱۰)$$

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.

۳-۲- تنظیم ضرایب اولیه PID و SMES

در این بخش الگوریتم MTLBO برای تنظیم پارامترهایPID و SMES به کار می‌رود. در این مقاله یک عضو (دانش‌آموز) جواب مسئله بهینه‌سازی و دروس (متغیرهای مسئله) ضرایب کنترل‌کننده‌های 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 استفاده‌شده 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 فراجهش 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 کنترل‌کننده PID و SMES طبق این معیار بهینه شده است. در جدول ۱ مقادیر بهینه‌شده K_p، K_i، K_d، K_{SMES} و K_{id} آورده شده و با مقادیر بهینه‌شده 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 دو 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 یادگرفته‌شده برای انجام این روش باید بهترین کنش را از بین کنش‌های قابل اجرا در هر موقعیت انتخاب کند.

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

	پارامترهای کنترل‌کننده‌ها		ITAE	ITSE×۱۰^{۴}	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 $\{\mathbf{s}\}$ نشان می‌دهند. یادگیری تقویتی با انتخاب هر کنش از مجموعه $\{\mathbf{a}\}$ 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 کاهش‌یافته را بیشینه کند. پاداش طولانی‌مدت کاهش‌یافته به‌صورت رابطه (۱۱) داده می‌شود.

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \qquad (۱۱)$$

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 به‌صورت رابطه (۱۲) داده می‌شود.

$$\mathbf{Q}^{\pi}(s,a) = E_{\pi}\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \middle| s_t = s, a_t = a\right\} \qquad (۱۲)$$

در این رابطه π سیاست بهینه اتخاذشده r پاداش و a کنش انتخابی است. سیاست باید طوری انتخاب شود که مقدار $\mathbf{Q}$ 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.

$$\Delta\mathbf{Q} = \alpha\left[r_{t+1} + \gamma \max_{a} \mathbf{Q}(s_{t+1}, a_{t+1}) - \mathbf{Q}(s_t, a_t)\right] \qquad (۱۳)$$

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

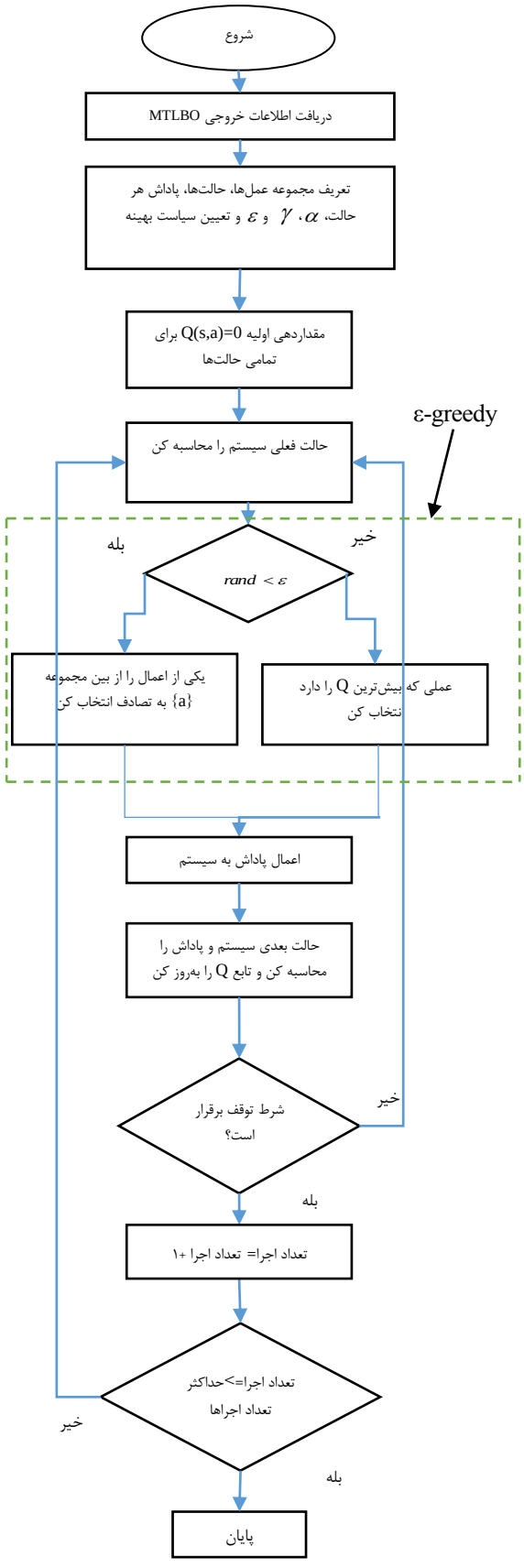
برای سیستم مورد مطالعه دو عامل برای یادگیری در نظر گرفته شده است. یک عامل برای تغییر ضرایب کنترل‌کننده ناحیه یک و یک عامل هم به همین ترتیب برای ناحیه دو لحاظ شده است؛ که این عامل‌ها برای رسیدن به بیش‌ترین پاداش مربوط به ناحیه خود و به‌طور مستقل از هم تلاش می‌کنند. در ادامه به تعریف مجموعه عمل‌ها، حالت‌ها، 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 و K_{SMES} سعی در یادگیری محیط را دارد. در این مقاله برای هر کدام از ضرایب کنترل‌کننده K_p، K_i، K_d و K_{SMES} پنج کنش a={۰/۹, ۱, ۱/۱, ۱/۲} در نظر گرفته شده است. که این کنش‌ها به‌صورت ضرب در مقدار خروجی کنترل‌کننده‌های مورد نظر اعمال می‌شوند. به دلیل اینکه یادگیری تقویتی تک‌عاملی بر روی هر ناحیه از سیستم مورد مطالعه اعمال شده است؛ کنش‌ها به‌صورت رابطه (۱۳) در نظر گرفته می‌شود.

به‌طور نمونه اگر عامل موجود در ناحیه یک دارای ضرایب کنترل‌کننده $(K_p, K_i, K_d, K_{smes})$=(۲, ۱/۵, ۱, ۵) باشد با اعمال کنش a=(۱/۱, ۱, ۱/۲, ۰/۹) ضرایب کنترل‌کننده در ناحیه یک به‌صورت $(K_p, K_i, K_d, K_{smes})$=(۲/۲, ۱/۵, ۱/۲, ۴/۵) تغییر پیدا می‌کند.

$$\{a\} = \begin{bmatrix} [\ ۰/۹ & ۰/۹ & ۰/۹ & ۰/۹\ ] \\ [\ ۰/۹ & ۰/۹ & ۰/۹ & ۱\ ] \\ [\ \vdots & \vdots & \vdots & \vdots\ ] \\ [\ ۱/۲ & ۱/۲ & ۱/۲ & ۱/۲\ ] \end{bmatrix} \qquad (۱۳)$$

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 انتخاب‌شده 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 $\Delta\dot{\omega} \le 0$ در بازه [۰/۰۰۵ -۰/۰۲۵]، ۱۰۰۰ حالت به‌صورت محدوده بازه‌ای برابر و ۲ حالت برای زمانی‌که منحنی از این محدوده خارج می‌شود، به‌صورت زیر در نظر گرفته شده است. در کل با توجه به رابطه (۱۳) ۲۰۰۴ حالت برای هر



شکل ۴: فلوچارت الگوریتم یادگیری Q خارج از خط

عامل در نظر گرفته شده است.

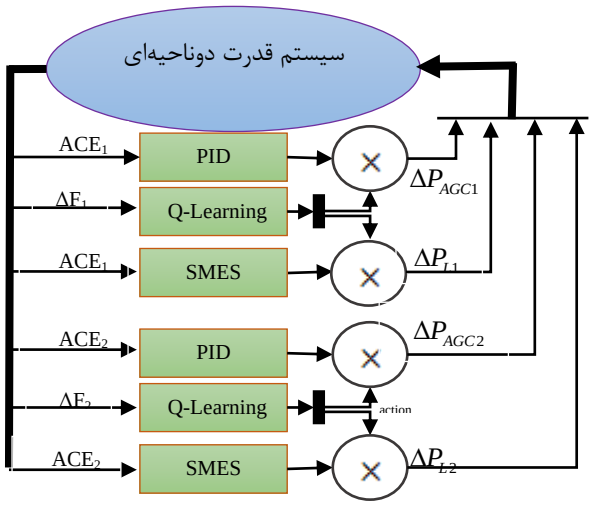
پاداش: در این مقاله چون هدف بهبود عملکرد کنترل‌کننده $\Delta\omega$ است که توسط الگوریتم MTLBO طراحی شده است؛ بنابراین پاداش‌ها هم متناسب با این هدف در نظر گرفته شده‌اند و هر موقع عامل کنشی را انجام داد که نتیجه بهتری نسبت به کنترل‌کننده طراحی‌شده با الگوریتم پیشنهادی داد پاداش مثبتی برای عامل در نظر 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 به‌صورت زیر در نظر گرفته می‌شود:

$$\{s\}=\begin{cases} s=1 & if \quad \Delta\omega<-0.025 \& \Delta\dot{\omega}>0 \\ s=2 & if \quad -0.025\le\Delta\omega<-0.024975 \& \Delta\dot{\omega}>0 \\ & \vdots \\ s=1001 & if \quad 0.00497\le\Delta\omega<0.005 \& \Delta\dot{\omega}>0 \\ s=1002 & if \quad \Delta\omega\ge0.005 \& \Delta\dot{\omega}>0 \\ s=1003 & if \quad \Delta\omega<-0.025 \& \Delta\dot{\omega}\le0 \\ s=1004 & if \quad -0.025\le\Delta\omega<-0.02497 \& \Delta\dot{\omega}\le0 \\ & \vdots \\ s=2003 & if \quad 0.00497\le\Delta\omega<0.005 \& \Delta\dot{\omega}\le0 \\ s=2004 & if \quad \Delta\omega\ge0.005 \& \Delta\dot{\omega}\le0 \end{cases} \qquad (۱۴)$$

$$P=10^{5}\left\|\Delta\omega_{\mathrm{TLBO}}\right|-\left|\Delta\omega_{\mathrm{RL}}\right\| \qquad (۱۵)$$

$$\begin{aligned} r=\{&if \quad \Delta\omega_{\mathrm{RL}}>0.005 \,\mathrm{or}\, \Delta\omega_{\mathrm{RL}}<0.025\Rightarrow r=-100 \\ &\mathrm{elseif}\,\Delta\omega_{\mathrm{TLBO}}>0 \;\Delta\omega_{\mathrm{RL}}>0\Rightarrow r=P(\Delta\omega_{\mathrm{RL}}-\Delta\omega_{\mathrm{TLBO}}) \\ &\mathrm{elseif}\,\Delta\omega_{\mathrm{TLBO}}<0 \;\Delta\omega_{\mathrm{RL}}<0\Rightarrow r=P(\Delta\omega_{\mathrm{TLBO}}-\Delta\omega_{\mathrm{RL}}) \\ &\mathrm{elseif}\ \left|\Delta\omega_{\mathrm{TLBO}}\right|<\left|\Delta\omega_{\mathrm{RL}}\right|\Rightarrow r=-P\frac{1}{\left|\Delta\omega_{\mathrm{RL}}\right|+0.001} \\ &\mathrm{elseif}\ \left|\Delta\omega_{\mathrm{TLBO}}\right|>\left|\Delta\omega_{\mathrm{RL}}\right|\Rightarrow r=P(\frac{1}{\left|\Delta\omega_{\mathrm{RL}}\right|+0.001}) \\ &\mathrm{end}\} \end{aligned} \qquad (۱۶)$$

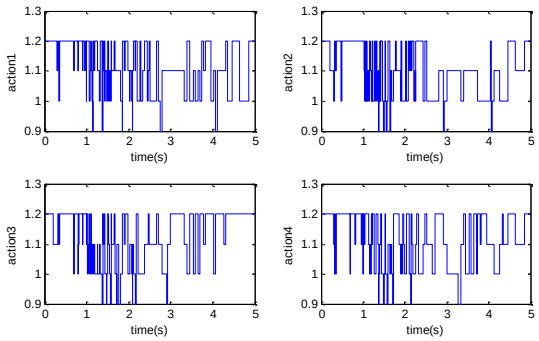
پارامترها: ε ، α و γ به‌ترتیب ۰/۰۳، ۰/۰۶ و ۰/۹۷ در نظر گرفته شده است. در شکل ۵ نحوه اعمال یادگیری به سیستم دوناحیه گرمایی، نشان داده شده است.



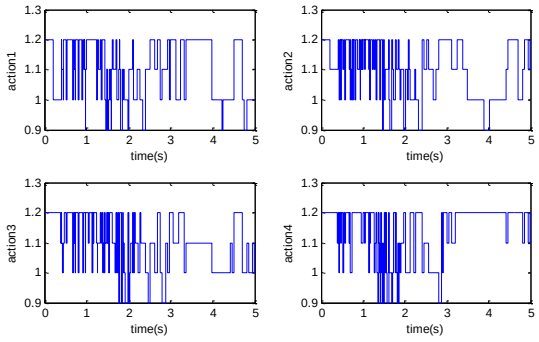
شکل ۵: نحوه اعمال RL به سیستم قدرت مورد مطالعه

۵- نتایج شبیه‌سازی

شبیه‌سازی‌ها به‌کمک نرم‌افزار MATLAB صورت گرفته است. برای یادگیری کامل محیط توسط عامل برای هر یک از ناحیه‌های سیستم مورد مطالعه افزایش بار به‌صورت 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 کنترل‌کننده ناحیه یک توانسته انحرافات فرکانسی مربوط به ناحیه یک را به‌خوبی میرا کند و عامل ۲ در کنترل‌کننده ناحیه ۲ نیز تأثیر بهتری بر سیگنال کنترلی نسبت به کنترل‌کننده‌های مقایسه‌شده را دارد. همچنین تغییر بار ۲٪ و ۱٪ مقدار نامی به‌ترتیب 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، ISTE، ISE، 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.



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



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

۶- نتیجه‌گیری

هدف از این مقاله طراحی کنترل‌کننده بار فرکانس مقاوم و تطبیقی برای یک سیستم قدرت چندناحیه‌ای مجهز به SMES است. در این راستا ابتدا الگوریتم MTLBO برای تنظیم اولیه پارامترهای کنترل‌کننده 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 به‌منظور 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 به‌سادگی 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 کنترل‌کننده مبتنی بر یادگیری تقویتی، شاخص‌های عملکرد حوزه زمان محاسبه شده و مقایسه شده‌اند که برتری روش کنترل پیشنهادی را ثابت می‌کنند.

ضمائم

ضمیمه الف

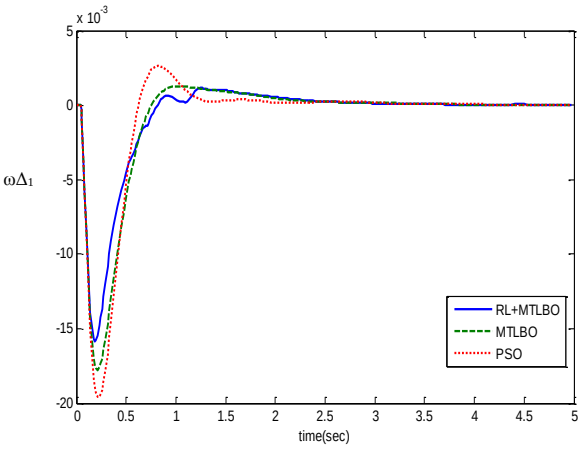
پارامترهای سیستم مورد مطالعه

f= ۶۰ Hz.
TH1, TH2 = ۰/۰۸ s.
TT1, TT2 = ۰/۳ s.
Tr1, Tr2 = ۱۰ s.
T12 = ۰/۰۸۶ pu MW/rad.
Tp1, Tp2 = ۲۰ s.
Kr1, Kr2 = ۰/۵
KP1, KP2 = ۱۲۰ Hz/pu MW.
R1, R2 = ۲/۴ Hz/MW.
B1, B2 = ۰/۴۲۵ MW/Hz.

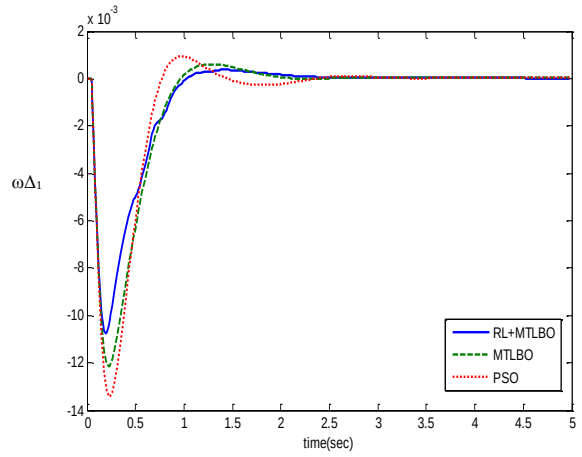
ضمیمه ب

پارامترهای واحد SMES

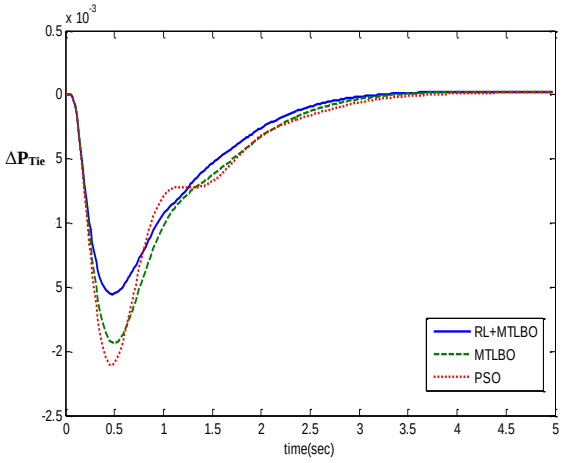
L = ۲/۶۵ H
TDC = ۰/۰۳ s
KSMES = ۱۰۰ kV/unit MW
Kid = ۰/۲ kV/kA
I_{d0} = ۴/۵ kA



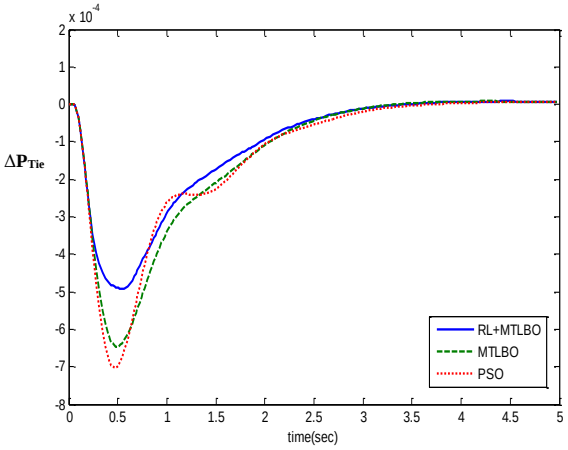
شکل ۸: تغییرات فرکانس در ناحیه ۱ برای تغییر بار ۳ درصد در ناحیه ۱



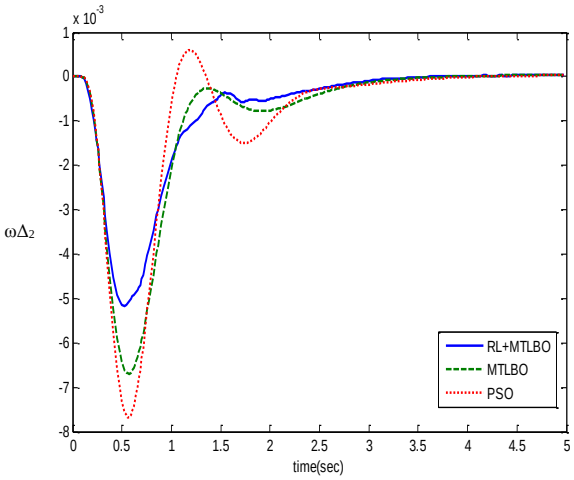
شکل ۱۱: تغییرات فرکانس در ناحیه ۱ برای تغییر بار ۲ درصد در ناحیه ۱ و تغییر بار ۱ درصد در ناحیه ۲



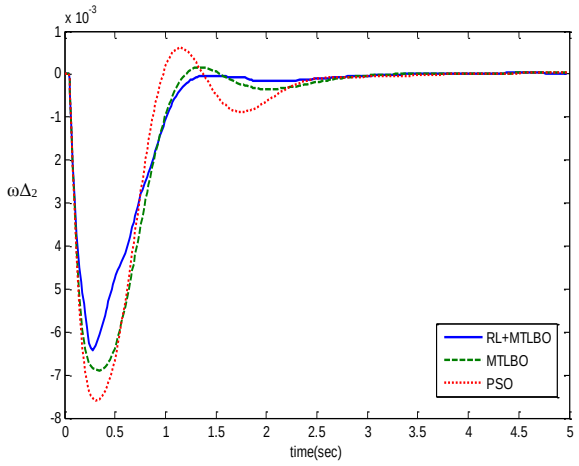
شکل ۹: تغییرات توان در خط ارتباطی برای تغییر بار ۳ درصد در ناحیه ۱



شکل ۱۲: تغییرات توان در خط ارتباطی 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



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

جدول ۲: مقایسه معیارهای ارزیابی حوزه زمان برای روش‌های کنترلی

	معیار		ΔP_{L1} = ۰/۰۱ p.u ΔP_{L2} = ۰/۰۰ p.u	ΔP_{L1} = ۰/۰۰ p.u ΔP_{L2} = ۰/۰۲ p.u	ΔP_{L1} = ۰/۰۳ p.u ΔP_{L2} = ۰/۰۱ p.u
RL+MTLBO	ITAE		۰/۰۰۳۰	۰/۰۰۵۹	۰/۰۰۸۸
	ITSE×۱۰^{۶}		۲/۵۷۶۹	۹/۷۱۱۶	۲۷/۱۴۵
	IAE		۰/۰۰۴۳	۰/۰۰۸۳	۰/۰۱۳۷
	Peak overshoot ×۱۰^{۳}	ΔF_1	۰/۴۴۷۴	۰/۰۲۰۳	۰/۰۰۹۳
		ΔF_2	۰/۰۰۸	۰/۶۹۰۶	۰/۳۳۹۰
		ΔP_{Tie}	۰/۰۰۷۲	۱/۰۳۳۷	۰/۵۱۵۴
	Peak undershoot ×۱۰^{۳}	ΔF_1	-۵/۴۱۷۵	-۳/۴۵۳۶	-۱/۷۵۸۳
		ΔF_2	−۱/۸۲۱۳	−۱۰/۷۲۷۹	−۵/۳۶۰۲
		ΔP_{Tie}	−۰/۵۳۳۱	−۰/۰۱۴۱	−۰/۰۰۷۶
MTLBO	ITAE		۰/۰۰۳۴	۰/۰۰۶۸	۰/۰۱۰۰
	ITSE×۱۰^{۶}		۳/۵۶۱۵	۱۴/۲۵۱	۴۰/۶۰۶
	IAE		۰/۰۰۴۸	۰/۰۰۹۷	۰/۰۱۶۱
	Peak overshoot ×۱۰^{۳}	ΔF_1	۰/۴۶۸۶	۰/۰۲۰۹	۰/۰۱۰۴
		ΔF_2	۰/۰۰۹۹	۰/۸۳۸۹	۰/۴۱۸۵
		ΔP_{Tie}	۰/۰۰۸۱	۱/۲۹۱۲	۰/۶۴۵۶
	Peak undershoot ×۱۰^{۳}	ΔF_1	−۵/۹۱۹۵	−۴/۴۶۴۳	−۲/۲۳۲۱
		ΔF_2	−۲/۲۳۲۷	−۱۱/۸۴۲۳	−۵/۹۲۴۵
		ΔP_{Tie}	−۰/۶۴۵۷	−۰/۰۱۶۲	−۰/۰۰۸۱
PSO	ITAE		۰/۰۰۴۳	۰/۰۰۸۷	۰/۰۱۲۳
	ITSE×۱۰^{۶}		۷/۹۷۴۰	۳۱/۸۷۲	۸۷/۳۰۷
	IAE		۰/۰۰۴۹	۰/۰۰۹۹	۰/۰۱۶۱
	Peak overshoot ×۱۰^{۳}	ΔF_1	۰/۸۷۳۱	۰/۳۹۸۲	۰/۱۹۸۹
		ΔF_2	۰/۱۹۹۱	۱/۷۴۷۱	۰/۸۷۲۰
		ΔP_{Tie}	۰/۰۰۸۲	۱/۴۰۳۳	۰/۷۰۱۸
	Peak undershoot ×۱۰^{۳}	ΔF_1	−۶/۵۴۷۰	−۵/۱۱۳۷	−۲/۵۵۶۵
		ΔF_2	−۲/۵۵۵۰	−۱۳/۰۹۶۱	−۶/۵۳۹۶
		ΔP_{Tie}	−۰/۷۰۲۰	−۰/۰۱۶۷	−۰/۰۰۸۳

ضمیمه ج

روابط معیار خطاهای استفاده‌شده در این مقاله

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

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

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

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

در این روابط t_{sim} زمان شبیه‌سازی است.

ضمیمه د

پارامترهای الگوریتم بهینه‌سازی PSO

C1=۲

C2=۲

W=۰/۹۹

مراجع

[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. Chaine 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، رئوف 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 بین المللی 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 Programing 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 framwork," *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.

[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 AVRs considering DGs," *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.

زیرنویس‌ها

[1] Superconducting Magnetic Energy Storage

[2] Pattern Search

[3] Cuckoo search algorithm

[4] Modified Teaching–Learning Based Optimization

[5] Reinforcement Learning

[6] Thyristor Controlled Series Capacitor

[7] Constrained Load Flow

[8] Particle Swarm Optimization