طراحی پایدارساز PSS3B بر اساس الگوریتم KH و Q-learning برای میراسازی نوسانات فرکانس‌پایین سیستم قدرت تک‌ماشینه

عبداله یونسی[۱] حسین شایقی[۲] عادل اکبری مجد[۳] یاشار هاشمی[۴]

۱- دانشجوی دکتری- گروه مهندسی برق، دانشکده فنی و مهندسی- دانشگاه محقق اردبیلی- اردبیل- ایران
a.younesi@ieee.org

۲- استاد- گروه مهندسی برق، دانشکده فنی و مهندسی- دانشگاه محقق اردبیلی- اردبیل- ایران
hshayeghi@gmail.com

۳- دانشیار- گروه مهندسی برق، دانشکده فنی و مهندسی- دانشگاه محقق اردبیلی- اردبیل- ایران
akbarimajd@gmail.com

۴- دانشجوی دکتری- گروه مهندسی برق، دانشکده فنی و مهندسی- دانشگاه محقق اردبیلی- اردبیل- ایران
yashar_hshm@yahoo.com

چکیده: هدف از این مقاله استفاده از روش یادگیری تقویتی به منظور تولید سیگنال مکمل برای بهبود عملکرد پایدارساز سیستم قدرت است. یادگیری تقویتی یکی از شاخه‌های مهم یادگیری ماشین در مبحث هوش مصنوعی بوده و روش کلی حل مسائل فرایند تصمیم‌گیری مارکو (MDP) است. در این مقاله یک روش کنترلی مبتنی بر یادگیری تقویتی به نام Q-learning ارایه و به منظور بهبود عملکرد پایدار ساز سیستم قدرت سه باند (PSS3B) در یک سیستم قدرت تک‌ماشینه مورد استفاده قرار گرفته است. بدین منظور ابتدا پارامترهای پایدار ساز سیستم قدرت سه باند تحت نقاط مختلف بهره برداری با بهینه‌سازی تابع هدف مبتنی بر مقادیر ویژه توسط الگوریتم بهینه سازی جدید KH تنظیم شده و سپس توسط الگوریتم یادگیری تقویتی ارایه شده بر اساس روش Q-learning به صورت بلادرنگ کارایی آن بهبود می‌یابد. از ویژگی اساسی پایدارساز پیشنهادی مبتنی بر یادگیری تقویتی سادگی و عدم وابستگی آن به مدل سیستم و تغییرات نقاط کار بهره برداری است. برای ارزیابی کارایی پایدار ساز سیستم قدرت سه باند مبتنی بر یادگیری تقویتی پیشنهادی نتایج آن با پایدار ساز سیستم قدرت معمولی و پایدار ساز سیستم قدرت سه باند طراحی شده با الگوریتم KH تحت نقاط کار مختلف با هم مقایسه می شود. نتایج شبیه‌سازی بر اساس شاخص‌های عملکردی نشان می‌دهد که پایدار ساز سیستم قدرت پیشنهاد شده در این مقاله عملکرد بهتری نسبت به دو روش دیگر از لحاظ کاهش زمان نشست و میرا نمودن نوسانات فرکانس پایین دارد.

کلمات کلیدی: پایدار ساز سیستم قدرت سه باند، یادگیری تقویتی، Q-learning

تاریخ ارسال مقاله : ۱۳۹۳/۰۴/۳۱

تاریخ پذیرش مشروط مقاله: ۱۳۹۴/۰۵/۱۷

تاریخ پذیرش مقاله: ۱۳۹۴/۱۰/۲۳

نام نویسنده‌ی مسئول: دکتر حسین شایقی

نشانی نویسنده‌ی مسئول: ایران – اردبیل – انتهای خیابان دانشگاه – دانشگاه محقق اردبیلی – دانشکده‌ی فنی – گروه مهندسی برق

[Downloaded from jiaeee.com on 2026-11-20]

[DOR: 20.1001.1.26765810.1396.14.3.6.1]

Journal of Iranian Association of Electrical and Electronics Engineers Vol14 No.3 Fall2017

۱- مقدمه

پایداری سیستم‌های قدرت یکی از مهمترین جنبه‌های عملکرد شبکه برق است، چرا که باید فرکانس و ولتاژ سیستم قدرت همواره در مقادیر نامی خود باشند حتی تحت اغتشاشات بزرگ مانند افزایش ناگهانی بار، خاموشی ناگهانی یک ژنراتور و یا خارج شدن یک خط انتقال در حین وقوع خطا. سیستم‌های قدرت را می‌توان سیستم‌های بزرگ و بهم پیوسته که دینامیک بسیار پیچیده‌ای دارند تصور کرد. ارتباط بین اجزای مختلف سیستم‌های قدرت نوسانات مختلفی را به کل سیستم تحمیل می‌کند. در این بین نوساناتی که دارای فرکانس پایین (بین 0.2 – 3.0 Hz) هستند از اهمیت زیادی برخوردارند. زیرا زمانی که این نوسانات شروع می‌شوند تا مدت زمان طولانی ادامه پیدا می‌کنند. گاهی اوقات در صورت عدم میرا ساز مناسب دامنه آن‌ها بزرگ شده و باعث نا پایداری سیستم قدرت می‌شوند. همچنین این نوسانات محدودیت‌های زیادی به قابلیت انتقال توان سیستم قدرت تحمیل می‌کنند[۱]. به منظور بهبود میرایی نوسانات سیستم قدرت، ژنراتور را به PSS تجهیز می‌کنند که توانایی میرا کردن این نوسانات را داراست. در [۲] نویسندگان PSS را برای سیستم قدرت تک‌ماشینه به صورت مقاوم طراحی کرده‌اند. در [۳] و [۴] نویسندگان برتری پایدار ساز سیستم قدرت چند باند را بر پایدار ساز معمولی در سیستم تک‌ماشینه و چند ماشینه نشان داده‌اند. در [۵] یک روش هوشمند مبتنی بر شبکه‌های عصبی و هوش مصنوعی مورد استفاده قرار گرفته است که به طور کامل جایگزین PSS و AVR می‌شود. این روش کنترلی گرچه مقاوم و تطبیقی است ولی برای پیاده سازی در عمل سخت‌افزار بسیار پیچیده ای لازم دارد که استفاده از آن را در عمل غیر ممکن می‌کند. در [۶] نویسندگان با ادغام ویژگی‌های شبکه‌های عصبی و منطق فازی، PSS مبتنی بر شبکه عصبی و منطق فازی را برای میرا سازی نوسانات سیستم قدرت طراحی کرده اند که روش بسیار پیچیده‌ای است و طراحی دشواری دارد.

محققان بسیاری بر نیاز به یک آموزش هوشمند و نظام مند برای کنترل سیستم قدرت تاکید کرده‌اند. عامل‌های هوشمند که می‌توانند در هر لحظه قدرت تصمیم‌گیری خود را به روز رسانی کنند[۷]، [۸]، [۹]. این نیاز را می‌توان به کمک یک روش محاسباتی برای یادگیری به نام یادگیری تقویتی برطرف کرد. منظور از بکارگیری یادگیری تقویتی جهت طراحی کنترل کننده این است که عامل‌های خودکار و هوشمند در هر حالت از سیستم بتوانند تصمیم بگیرند و در راستای افزایش پاداش طولانی مدت خود در هر حالت عمل بکنند. در دهه اخیر، یادگیری تقویتی جایگاه ویژه‌ای در کنترل سیستم قدرت پیدا کرده‌است و با موفقیت در مباحث پایداری سیگنال کوچک، پایداری ولتاژ ، پایداری گذرا و مباحث بازار برق بکار گرفته شده است. در [۱۰] اصول بکارگیری یادگیری تقویتی در کنترل پایداری سیستم قدرت بررسی شده و کارایی این روش در کنترل TCSC ، برای بهبود نوسانات توان عبوری بین دو ناحیه سیستم چهار ماشینه نشان داده شده است. از نتایج این مقاله می‌توان دریافت که یادگیری تقویتی قابل اعمال به هر سیستمی با هر اندازه بزرگی و پیچیدگی دینامیکی است و می‌توان با کوچکترین شناخت از سیستم آن را کنترل کرد. همچنین این روش کنترلی مقاوم است و با تغییر شرایط سیستم خود را تطبیق می‌دهد. در [۱۱] نویسندگان دو کاربرد از یادگیری تقویتی را نشان داده‌اند. در کاربرد اول از یادگیری تقویتی برای تنظیم کردن بهره پایدار ساز سیستم قدرت معمولی استفاده شده است و در کاربرد دوم یادگیری تقویتی به طور کامل جایگزین پایدار ساز سیستم قدرت شده است، هردو کاربرد نشان می‌دهند که یادگیری تقویتی می‌تواند مکمل و یا جایگزین مناسبی برای پایدار ساز سیستم قدرت باشد. در [۱۲] از یادگیری تقویتی برای کنترل توان راکتیو استفاده و با روش‌های احتمالاتی CLF مقایسه شده است که نتایج برتری یادگیری تقویتی را ثابت می‌کنند. [۱۳] کاربرد یادگیری تقویتی را در مبحث بازار برق نشان می‌دهد.

در این مقاله سه نوع کنترل برای پایداری سیستم قدرت تک‌ماشینه بررسی می‌شود. پایدار ساز سیستم قدرت معمولی (CPSS)، پایدار ساز سیستم قدرت سه باند (PSS3B) و پایدار ساز سیستم قدرت سه باند همراه یادگیری تقویتی (PSS3B+RL) پایدار ساز PSS3B ابتدا به روش مقاوم تحت نقاط مختلف بهره برداری با بهینه‌سازی تابع هدف مبتنی بر ضریب میرایی و نسبت میرایی مدهای الکترومکانیکی ناپایدار و با میرایی ضعیف توسط الگوریتم بهینه سازی جدید KH طوری طراحی می شود که مدهای الکترومکانیکی ناپایدار و با میرایی ضعیف را به ناحیه مشخصی از صفحه مختلط انتقال دهند و سپس کارایی کنترلی آن در سیستم غیر خطی بر اساس الگوریتم یادگیری تقویتی پیشنهادی Q-learning به صورت بلادرنگ برای میرا نمودن هرچه بهتر نوسانات فرکانس پایین بهبود می‌یابد. از یادگیری تقویتی جهت تولید سیگنال مکمل برای بهبود عملکرد پایدار سیستم قدرت سه باند استفاده شده‌است. استراتژی کنترلی پیشنهادی ویژگی‌های پایدار ساز سیستم قدرت سه باند و یادگیری تقویتی مبتنی بر Q-learning را درهم آمیخته و منجر به ساختار کنترلی ساده و انعطاف‌پذیر شده و روش قدرتمندی برای میرایی نوسانات فرکانس پایین و بهبود پایداری دینامیکی سیستم قدرت محسوب می‌شود. پایدار ساز سیستم قدرت معمولی به روش جبران کننده فاز طراحی شده است. سپس نتایج با یکدیگر مقایسه و برتری روش کنترلی پیشنهادی از منظر فراجهش، فروجهش، زمان نشست و معیارهای عملکردی ITAE، ISTSE و ISE نشان داده شده است.

۲- انواع پایدار سازهای سیستم قدرت

پایدار ساز سیستم قدرت یک کنترل فیدبک الکترونیکی از سیستم تحریک واحد تولید است که وظیفه آن میرا کردن نوسانات و افزایش

Journal of Iranian Association of Electrical and Electronics Engineers - Vol.14- No.3 Fall 2017

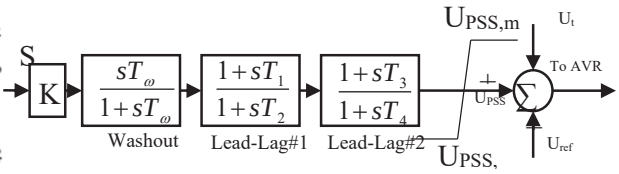
[Downloaded from jiaeee.com on 2026-11-20]

[DOR: 20.1001.1.26765810.1396.14.3.6.1]

حد پایداری زاویه روتور سیستم قدرت با مدوله کردن ولتاژ تحریک ژنراتور است [۱۴]. IEEE مدل‌های مختلفی برای PSS تعریف کرده است که در این مقاله مدل PSS معمولی و PSS3B مورد بررسی قرار می‌گیرد.

۲-۱- پایدار ساز سیستم قدرت معمولی (CPSS)

شکل(۱) مدل IEEE یک پایدار ساز PSS را نشان می‌دهد. ورودی این پایدار ساز سیگنال تغییرات سرعت زاویه‌ای می‌باشد که اصطلاحاً به آن CPSS گفته می‌شود.



شکل(۱): ساختار پایدار ساز CPSS

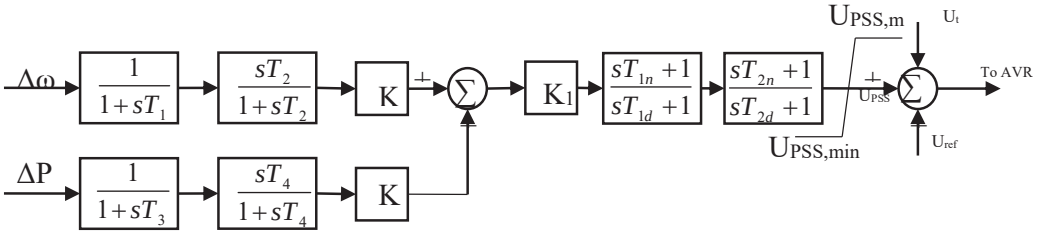
این بلوک دیاگرام شامل بلوک washout می‌باشد که باعث کاهش پاسخ بیش از حد تحمل سیستم در هنگام بروز اغتشاش بزرگ می‌شود. چون PSS باید با تغییرات سرعت، گشتاور الکتریکی در فاز ایجاد کند، از بلوک پیشفاز-پسفاز در PSS استفاده می‌شود. تعداد بلوک‌های پیشفاز-پسفاز بستگی به ماهیت سیستم دارد. در این مقاله دو بلوک برای آن فرض می‌شود. میزان میرایی PSS توسط بهره K_ω ایجاد می‌شود. این پایدار ساز بسیار حساس به نویز بوده و همواره حاوی نوسانات پیچشی است. در این مقاله، برای سادگی T_1 برابر T_3 و T_2 برابر T_4 فرض شده‌اند و پایدار ساز به روش جبران فاز طراحی شده‌است[۱۵]. داده‌های آن در جدول(۱) نشان داده شده است.

جدول(۱): داده های مربوط به CPSS مورد استفاده در این مقاله

نام	K_ω	T_ω[s]	T_1[s]	T_2[s]	U_{PSS}^{min}[pu]	U_{PSS}^{max}[pu]
مقدار	12.5	5	0.0738	0.028	-0.15	0.15

۲-۲- پایدار ساز سیستم قدرت PSS3B

مدل IEEE پایدار ساز PSS3B در شکل۲ نشان داده شده است. پایدار ساز PSS3B از دو ورودی تغییرات توان الکتریکی (ΔP) و سرعت تغییرات زاویه روتور ($\Delta\omega$) استفاده می‌کند. در این پایدار ساز، T_1 و T_3 ثابت‌های زمانی مبدل و T_2 و T_4 بیانگر ثوابت زمانی فیلتر پیچشی در دو کانال می‌باشند. بهره مطلوب پایدار ساز با تنظیم مقادیر K_1، K_2 و K_3 بدست می‌آید. همچنین T_{1n}، T_{1d}، T_{2n} و T_{2d} ضرایب جبرانگرهای فاز پایدار ساز می‌باشند. در خروجی پایدار ساز نیز از محدود کننده ولتاژ تحریک استفاده شده است[۱۶]. در ساختار شکل۲، T_2 مساوی T_4 و برابر ۱۰ و همچنین برای ساده‌تر شدن مسئله بهینه‌سازی مقادیر T_{1n}=0.02، T_{1d}=0.01، T_{2n}=0.03 و T_{2d}=0.01 فرض شده‌اند. بنابراین برای پایدار ساز PSS3B پارامترهای K_1، K_2، K_3، T_1 و T_3 قابل تنظیم اند.



شکل(۲): ساختار پایدار ساز PSS3B

مسئله بهینه‌سازی تنظیم پارامترهای این پایدار ساز تعریف و بوسیله الگوریتم KH حل می‌شود.

۳- مروری اجمالی بر الگوریتم Krill herd

الگوریتم KH ساختاری مشابه به PSO دارد ولی طبق [۱۷] عملکرد آن نسبت به PSO بسیار بهتر بوده و به راحتی نیز قابل اعمال است. بنابر این در این مقاله از این الگوریتم برای حل مسئله بهینه‌سازی استفاده شده است. کریل نام نوعی جانور سخت پوست است که در آب‌های تمام جهان یافت می‌شود. این جانوران به صورت دسته‌های بسیار بزرگ همواره در حال حرکت هستند و الگوریتم KH با الهام از قوانین حاکم بر حرکت دسته‌جمعی این جانوران بدست آمده است. در الگوریتم KH فرض بر این است که حرکت هر ذره کریل از سه فاکتور تاثیر می‌پذیرد: حرکت ناشی از ذره‌های دیگر، فعالیت کاوش‌گرانه (جستوجو برای غذا) و انتشار تصادفی. معادله لاگرانژ نشان داده شده در فرمول(۱)، این فرض را به صورت ریاضی مدل می‌کند.

$$\frac{dX_i}{dt} = N_i + F_i + D_i \qquad (۱)$$

[Downloaded from jiaeee.com on 2026-11-20]

[DOR: 20.1001.1.26765810.1396.14.3.6.1]
Journal of Iranian Association of Electrical and Electronics Engineers Vol14 No.3 Fall2017

که در آن N_i حرکت ناشی از ذره‌های دیگر، F_i حرکت کاوش‌گرانه و D_i انتشار فیزیکی ذره iام است. حرکت ناشی از ذره‌های دیگر طبق فرمول‌های(۲) و(۳) بیان می‌شود.

$$N_i^{new} = N^{\max}\alpha_i + \omega_n N_i^{old} \quad (۲)$$

که در آن

$$\alpha_i = \alpha_i^{local} + \alpha_i^{target} \quad (۳)$$

در روابط فوق، N^{max} بیشترین سرعت القا شده، ω_n عددی بین [0,1]، ضریب اینرسی حرکت القا شده، N_i^{old} حرکت القا شده قبلی، α_i^{local} تاثیر محلی است که توسط همسایه‌ها اعمال می‌شود و α_i^{target} تاثیر جهت هدف است که توسط بهترین ذره اعمال می‌شود. حرکت کاوش-گرانه بر اساس دو متغییر اصلی فرمول‌بندی می‌شود، اولی موقعیت غذا و دومی تجربه قبلی از موقعیت غذا است. این جابجایی توسط فرمول-های (۴) و (۵) تعریف می‌شود.

$$F_i = V_f \beta_i + \omega_f F_i^{old} \quad (۴)$$

که در آن

$$\beta_i = \beta_i^{food} + \beta_i^{best} \quad (۵)$$

در روابط فوق، V_f سرعت کاوش، ω_n عددی بین [0,1]، ضریب اینرسی کاوش، β_i^{food} ضریب جذب غذا و β_i^{best} بهترین موقعیت تجربه شده ذره iام تا به حال است. انتشار فیزیکی ذره‌های کریل، یک فرایند تصادفی فرض می‌شود و طبق فرمول (۶) بیان می‌شود.

$$D_i = D^{\max}\delta \quad (۶)$$

که در آن، D^{max} بیشترین سرعت انتشار است و δ یک بردار تصادفی جهت‌دار بین [-1,1] است. در نهایت موقعیت جدید ذره کریل iام در لحظه $t+\Delta t$ با فرمول (۷) محاسبه می‌شود.

$$X_i(t+\Delta t) = X_i(t) + \Delta t \frac{dX_i}{dt} \quad (۷)$$

لازم به ذکر است که پارامتر ثابت Δt بسیار مهم است و باید با توجه به مسئله بهینه سازی به دقت تعیین شود. چون مقدار این کمیت به فضای جستوجو بستگی دارد می‌توان آن را به کمک فرمول (۸) تعیین کرد.

$$\Delta t = C_t \sum_{j=1}^{NV} (UB_j - LB_j) \quad (۸)$$

که در آن، NV تعداد کل متغیرها، UB_j و LB_j حد بالا و حد پایین متغییر j ام هستند، C_t یک ثابت بین [0,2] می‌باشد که به ذره‌های کریل اجازه می‌دهد فضای جستوجو را به دقت جستوجو کنند. فلوچارت الگوریتم KH با کادر آبی رنگ در شکل (۴) نشان داده شده-است. برای جزئیات بیشتر در باره الگوریتم KH به [۱۷] مراجعه شود.

۴- تابع هزینه

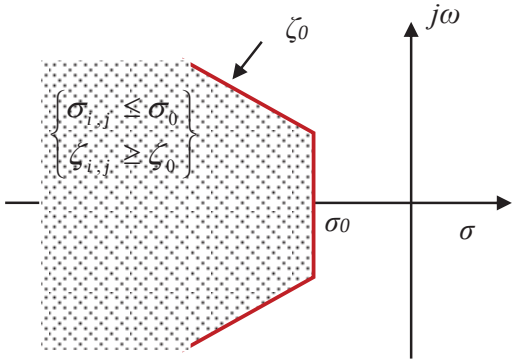
برای مقاوم بودن پایدار ساز در مقابل تغییرات نقاط کار سیستم، بهینه‌سازی با توجه به تغییرات P_t، Q_t و X_e در محدوده‌های تعریف شده انجام شده‌است. نقاط کاری استفاده شده برای بهینه‌سازی به صورت زیر هستند:

- توان اکتیو (P_t): از 0.4 تا 1 با گام‌های 0.1 .
- توان راکتیو (Q_t): از 0.2- تا 0.5 با گام‌های 0.1 .
- راکتانس خط (X_e): از 0.2 تا 0.7 با گام‌های 0.1 .

نحوه محاسبه تابع هزینه اینگونه است که برای هر نقطه کاری، سیستم خطی‌سازی می‌شود، مقادیر ویژه سیستم حلقه بسته بدست می‌آیند و تابع هدف با استفاده از مقادیر ویژه ناپایدار و یا کمتر میرا شده سیستم، که نیاز به جابجایی به سمت چپ صفحه مختلط دارند، محاسبه می‌شود. فرمول (۹) تابع هدف مورد استفاده در این مقاله را نشان می‌دهد.

$$J = \sum_{j=1}^{np} \sum_{\sigma_{i,j} \geq \sigma_0} \left[\sigma_0 - \sigma_{i,j}\right]^2 + a \sum_{j=1}^{np} \sum_{\zeta_{i,j} \leq \zeta_0} \left[\zeta_0 - \zeta_{i,j}\right]^2 \quad (۹)$$

که در آن np تعداد نقاط کار، σ قسمت حقیقی مقادیر ویژه، ζ ضریب میرایی و a ضریب وزنی می‌باشند. در رابطه ۱، a=۱۰، σ_0=-۱ و ζ_0=۱۰٪ درنظر گرفته شده است. شکل(۳) مفهوم تابع هدف فرمول (۹) را به خوبی توصیف می‌کند. برای جزئیات بیشتر به [۱۸] مراجعه شود.



شکل(۳): ناحیه مشخص شده برای تابع هدف

فرآیند طراحی پایدار ساز PSS3B را می‌توان به صورت مسئله بهینه سازی با محدودیت‌های زیر در نظر گرفت:

[Downloaded from jiaeee.com on 2026-11-20]

[DOR: 20.1001.1.26765810.1396.14.3.6.1]

$$\begin{aligned} &\text{minimize } J \\ &\text{subject to:} \\ &K_1^{\min} < K_1 < K_1^{\max}, K_2^{\min} < K_2 < K_2^{\max} \\ &K_3^{\min} < K_3 < K_3^{\max}, T_1^{\min} < T_1 < T_1^{\max} \\ &T_3^{\min} < T_3 < T_3^{\max} \end{aligned} \qquad (۱۰)$$

نتایج حاصل از حل مسئله بهینه‌سازی فرمول (۱۰) در جدول(۲) نشان داده شده است.

جدول(۲): نتایج حاصل از بهینه سازی PSS3B

نام	K_1	K_2	K_3	T_1	T_3
مقدار	2.0304	13.7193	0.5	0.1002	0.01

۵- یادگیری تقویتی (Reinforcement Learning)

یادگیری تقویتی روشی است که در آن یک یا بیشتر از یک عامل در تعامل با محیط یک سیاست کنترلی بهینه را در جهت محقق شدن یک هدف از پیش تعیین شده یاد می‌گیرند. سیاست بهینه به معنی انتخاب بهترین عمل، در بین عمل‌های موجود برای هر موقعیت عامل در محیط است. در حالت کلی عامل دانش اولیه‌ای از محیط پیرامون خود ندارد و سیاست کنترلی را با یک روش سعی و خطا یاد می‌گیرد. روش‌های یادگیری تقویتی می‌توانند هر سیستم غیر خطی را بدون ساده سازی کنترل کنند. Q-learning یکی از شناخته شده‌ترین روش‌های یادگیری تقویتی است که در این مقاله مورد استفاده قرار می‌گیرد. دلیل این استفاده سادگی و عدم وابستگی این روش به مدل سیستم می‌باشد. از ویژگی‌های برجسته کنترل کننده‌های مبتنی بر Q-learning می‌توان عدم وابستگی به مدل سیستم، مقاوم در برابر تغییر شرایط بهره برداری و عدم قطعیت پارامترهای سیستم، کنترل تطبیقی و سادگی در پیاده سازی را بر شمرد. این روش کنترلی می‌تواند به صورت مکملی مناسب برای روش‌های کنترلی سنتی بکار گرفته شود،که در این مقاله از این ویژگی استفاده شده است و با بهره‌گیری از Q-learning و ایجاد یک سیگنال مکمل کارایی پایدار ساز سیستم قدرت افزایش پیدا کرده است. یادگیری تقویتی فرض می‌کند که محیط (سیستم کنترلی) به حالت‌های محدودی تقسیم شده است و با مجموعه‌ی $\{S\}$ نشان داده می‌شود. در هر گام t، عامل خود را در حالت s_t سیستم می‌بیند و کنش a را از بین مجموعه‌ای از کنش‌های موجود $\{A\}$ انتخاب می‌کند. عامل به محض انجام کنش از محیط پاداش می‌گیرد. پاداش داده شده به نحوی تعریف می‌شود که میزان رضایت‌مندی از کنش انجام شده را نشان دهد. سپس عامل خود را در موقعیت جدیدی از سیستم s_{t+1} می‌بیند و دوباره کنش مناسب را انتخاب کرده و همین روند تا زمان برآورده شدن هدف تعیین شده ادامه می‌یابد. هدف یادگیری تقویتی، پیدا کردن یک سیاست، یک نگاشت بین حالت‌ها و کنش‌های سیستم است، که در نتیجه آن پاداش طولانی مدت کاهش یافته، بیشینه شود. پاداش طولانی مدت کاهش یافته سیستم به صورت زیر داده می‌شود:

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

که در آن، γ عددی بین $[0,1]$ است و ضریب کاهش نام دارد، این ضریب اهمیت پاداش‌های آینده را در تصمیم‌گیری نشان می‌دهد. اگر مقدار آن 0 در نظر گرفته شود، پاداش‌های بعدی در روند تصمیم‌گیری بی‌تاثیر خواند بود و اگر برابر 1 فرض شود، پاداش‌های بعدی نیز در تصمیم‌گیری تاثیر می‌گذارند. یادگیری تقویتی دارای یک تابع ارزش است که در Q-learning تابع Q نامیده می‌شود و عبارت است از :

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

که در آن، π سیاست کنترلی، s حالت فعلی، a کنش انتخابی و r پاداش دریافتی از محیط است. روش یادگیری تقویتی سیاست بهینه π^* را طوری پیدا می‌کند که مقدار تابع Q فرمول (۱۲) بیشینه شود. در روش Q-learning از یادگیری تقویتی در هر گام زمانی، فرمول (۱۲) در تعامل عامل با محیط، به روز رسانی می‌شود. رابطه به روز رسانی، معادله بهینه بلمن نامیده شده و با فرمول (۱۳) داده می‌شود.

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

که در آن α، بین $[0,1]$ بوده و ضریب تضعیف نام دارد که نشان دهنده مقدار واقعی خطاست. در جدول(۳) گام‌های پیاده سازی الگوریتم Q-learning به طور خلاصه نشان داده شده است.

برای انتخاب کنش در هر حالت از الگوریتم شبه حریصانه استفاده می‌شود، به این صورت که به احتمال ε-۱، کنش با Q بیشتر انتخاب می‌شود و با احتمال ε از بین همه کنش‌ها یکی به تصادف انتخاب می‌شود.

۶- پارامترهای یادگیری تقویتی برای کاربرد در این مقاله

الف) حالت‌ها

هدف اصلی از بکاربردن پایدار ساز، میرا کردن نوسانات فرکانس پایین سیستم قدرت است. به عبارت دیگر پایدار ساز باید نوسانات توان و یا نوسانات فرکانس را میرا کند. بنابر این می‌توان از ΔP و یا $\Delta\omega$ و یا ترکیبی از این دو به عنوان حالت‌های سیستم استفاده کرد. در این مقاله، $\Delta\omega$ به عنوان حالت در نظر گرفته شده است. بازه 0.01- تا 0.01+ (در سیستم پریونیت) به ۱۲ قسمت تقسیم شده و هر قسمت یک حالت سیستم را نشان می‌دهد. بنابر این مجموعه حالت‌ها را می‌توان به صورت زیر تعریف کرد:

[Downloaded from jiaeee.com on 2026-11-20]
[DOR: 20.1001.1.26765810.1396.14.3.6.1]
Journal of Iranian Association of Electrical and Electronics Engineers Vol14 No.3 Fall2017

$$S = \{(-\infty, -0.01], (-0.01, -0.0082], (-0.0082, -0.0064], (-0.0064, -0.0045], (-0.0045, -0.0027], (-0.0027, -0.0009], (-0.0009, 0.0009], (0.0009, 0.0027], (0.0027, 0.0045], (0.0045, 0.0064], (0.0064, 0.0082], (0.0082, 0.01], (0.01, +\infty)\} \quad (۱۵)$$

در مجموعه حالت‌ها، حالت [0.0009,0.0009-) به عنوان حالت عادی در نظر گرفته می‌شود.

ب) کنش‌ها

تعریف مجموعه کنش‌ها بسیار پیچیده است و از اهمیت فراوانی برخوردار است. با توجه به محدود کننده‌هایی که در خروجی پایدار ساز قرار می‌دهند، می‌توان محدوده این کنش‌ها را تخمین زد. با توجه به مرجع [۱۹] می‌توان تخمین زد که کنش‌ها در بازه [۰/۲ ۰/۲-] قرار دارند. برای سادگی مجموعه کنش‌ها به صورت زیر تعریف می‌شود:

$$A = \{-0.2, 0, 0.2\} \quad (۱۶)$$

ج) پاداش

به طور کلی چون هدف از طراحی پایدار ساز میرا کردن نوسانات فرکانس و توان می‌باشد، انتخاب پاداش مورد نظر برابر قرار داده شده است با فاصله $\Delta\omega$ از نقطه 0 در دو گام زمانی t و t-1 در این مقاله فرض شده است که هر کنش به مدت 50 میلی ثانیه به سیستم اعمال می‌شود و سپس حالت بعدی و پاداش محاسبه می‌شوند. بنابر این پاداش به صورت زیر تعریف شده است:

$$\mathrm{Re}ward_t = \sum_{k=t-1}^{t} \Delta\omega(k) \quad (۱۷)$$

همچنین مقادیر α، ε و γ به ترتیب 0.02، 0.05 و 0.98 در نظر گرفته شده‌اند. در شکل(۴)، مدل سیستم قدرت مورد مطالعه و نحوه اعمال یادگیری تقویتی به پایدار ساز سیستم قدرت، جهت بهینه کردن عملکرد آن نشان داده شده است.

۷- نتایج شبیه‌سازی

برای نشان دادن عملکرد روش کنترلی پیشنهادی، شبیه سازی‌های کامپیوتری به کمک نرم‌افزار متلب صورت گرفته است. سیستم مورد مطالعه در شکل(۴) نشان داده شده‌است و اطلاعات اجزای آن به طور کامل در [۱۰] موجود می‌باشد. شبیه‌سازی‌ها برای یک خطای سه فاز ۱۰۰ میلی‌ثانیه در باس ژنراتور، در شرایط بهره‌برداری متفاوت انجام شده است. شکل‌های (۵) و (۶) نتایج حاصل از وقوع یک خطای سه فاز ۱۰۰ میلی ثانیه در باس ژنراتور را در دو وضعیت کاری متفاوت نشان می‌دهند.

جدول(۳): پیاده‌سازی گام به گام الگوریتم Q-learning

آ. پیدا کردن سیاست بهینه

آ-۱. تعریف مجموعه حالت‌ها، کنش‌ها و پاداش.

آ-۲. تعیین مقدار γ، α و ε.

آ-۳. مقدار دهی اولیه $Q(s,a) = 0$ برای تمامی حالت‌ها و عمل‌ها.

آ-۴. برای هر بار اجرا (Episode)

آ-۴-۱. حالت فعلی سیستم (s) را محاسبه کن.

آ-۴-۲. تا زمان رسیدن به هدف تکرار کن.

آ-۴-۲-۱. با استفاده از الگوریتم شبه حریصانه (ε-greedy)، کنش a را از بین کنش‌های موجود برای حالت s انتخاب کن.

آ-۴-۲-۲. کنش a را انجام بده و پاداش r و حالت بعدی را بگیر.

آ-۴-۲-۳. تابع Q را طبق رابطه زیر به روز رسانی کن.

$$Q(s,a) = Q(s,a) + \Delta Q \quad (۱۴)$$

آ-۴-۲-۴. حالت بعدی سیستم را در حالت فعلی قرار بده.

پایان تکرار. (برو به آ-۴-۲)

پایان یک اجرا. (برو به آ-۴)

ب. اجرا کردن سیاست بهینه

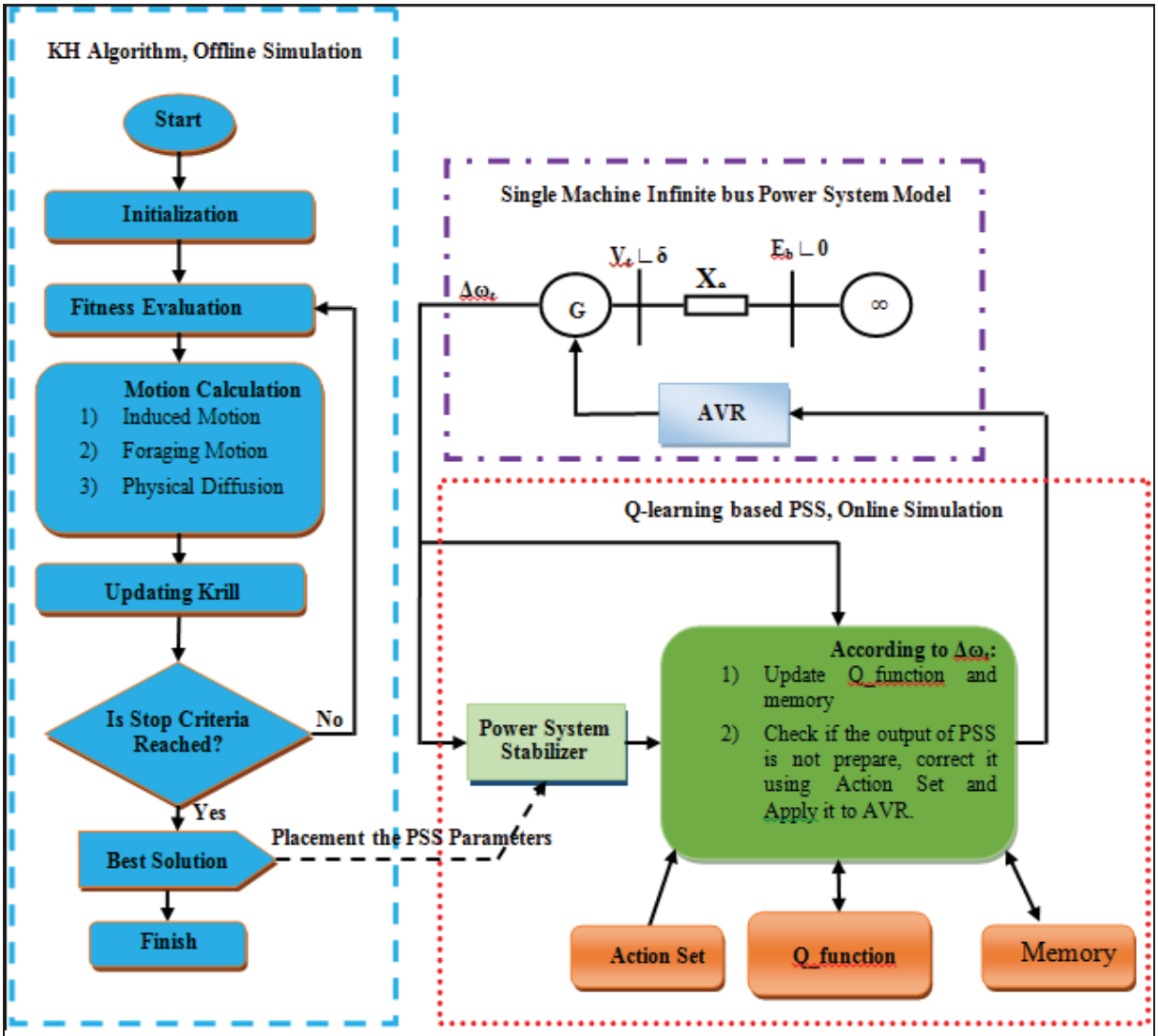
ب-۱. برای حالت فعلی سیستم، کنشی را که مقدار تابع Q برای آن بیشترین است، انتخاب کن.

ب-۲. حالت بعدی سیستم را پیدا کرده و به جای حالت فعلی قرار بده.

ب-۳. به ب-۱ رفته و همین روند را ادامه بده.

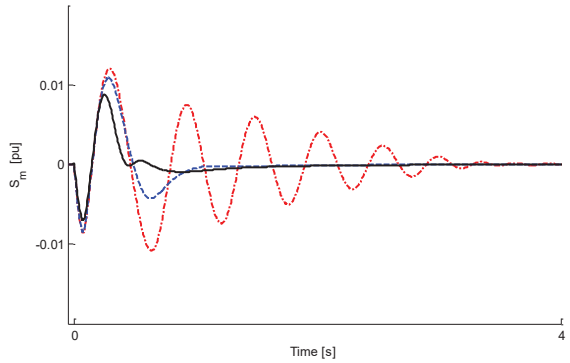
[Downloaded from jiaeee.com on 2026-11-20]

[DOR: 20.1001.1.26765810.1396.14.3.6.1]



شکل(۴): مدل سیستم قدرت مورد مطالعه و روش پیشنهادی برای بهینه کردن عملکرد پایدار ساز سیستم قدرت به وسیله Q-learning و الگوریتم Krill heard . خط‌چین آبی: فلوچارت الگوریتم KH، نقطه‌چین قرمز: پایدار ساز مبتنی بر Q-learning، خط‌نقطه بنفش: مدل سیستم قدرت مورد مطالعه

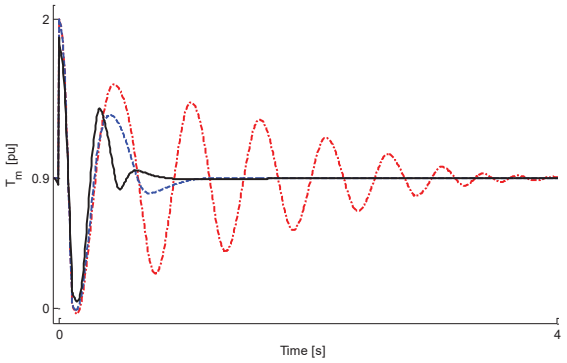
شکل‌ها به واضحی نشان می‌دهند که روش کنترلی پیشنهادی، مستقل از شرایط بهره‌برداری بسیار کاراتر از پایدار ساز معمولی و پایدار ساز PSS3B است. در ادامه برای مقایسه عملکرد پایدار سازها در وضعیت‌های کاری متفاوت نتایج شبیه‌سازی برای چهار وضعیت کاری متفاوت مورد بررسی آماری قرار گرفته است و معیارهای فراجهش، فروجهش، زمان نشت، ITAE، ISTSE و ISE برای تغییرات Δω، در جدول(۴) محاسبه شده‌اند. با توجه به داده‌های جدول۴ می‌توان نتیجه گرفت که روش کنترلی مکمل، مبتنی بر Q-learning مستقل از وضعیت‌های کاری متفاوت عملکرد PSS3B را بهبود بخشیده است. مقدار فراجهش در هر چهار وضعیت کاری مورد بررسی تقریبا ٪۲۰ نسبت به حالت PSS3B بدون کنترل مکمل، بهبود پیدا کرده‌است. مقدار فروجهش در وضعیت‌های کاری ۱ و ۲ تقریبا ٪۱۶ و در وضعیت‌های کاری ۳ و ۴ تقریبا ٪۱۳ بهبود پیدا کرده است. زمان نشست در وضعیت‌های کاری ۱ و ۳ حدود ٪۰/۱ ، در وضعیت کاری ۲ حدود ٪۷ و در وضعیت کاری ۴ به میزان ٪۳۳ بهبود پیدا کرده‌است.



الف

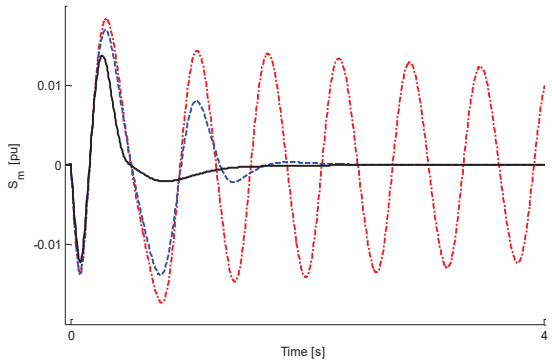
[Downloaded from jiaeee.com on 2026-11-20]

[DOR: 20.1001.1.26765810.1396.14.3.6.1]
Journal of Iranian Association of Electrical and Electronics Engineers Vol14 No.3 Fall2017

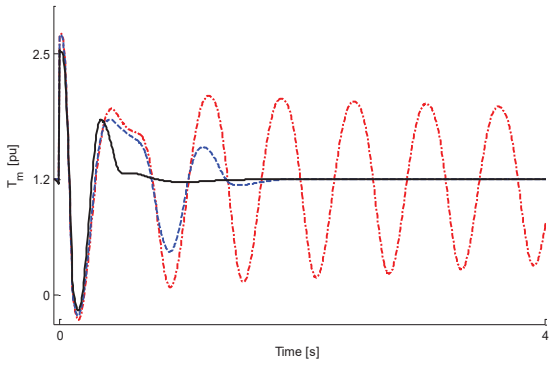


ب

شکل (۵): نتایج حاصل از خطای سه فاز ۱۰۰ میلی‌ثانیه در وضعیت کاری $Pt = 1.2$، $Qt = 0.2$، $Xe = 0.7$ و $Tm = 0.9$. (الف) S_m، (ب) T_m.
یکپارچه : PSS3B+RL، خط‌چین : PSS3B، خط‌نقطه : CPSS



الف



ب

شکل(۶): نتایج حاصل از خطای سه فاز ۱۰۰ میلی‌ثانیه در وضعیت کاری $Pt = 0.7$، $Qt = 0.5$، $Xe = 0.3$ و $Tm = 1.2$. (الف) S_m، (ب) T_m.
یکپارچه : PSS3B+RL، خط‌چین : PSS3B، خط‌نقطه : CPSS

معیار ITAE در وضعیت‌های کاری ۱ و ۲ تقریبا ۴۰٪ و ۳ و ۴ تقریبا ۶۰٪ بهبود پیدا کرده‌است. معیار ISTSE نیز در همه وضعیت‌های کاری به ترتیب ۵۸، ۶۳، ۷۰ و۷۲ درصد بهبود پیدا کرده‌است. به همین ترتیب معیار ISE نیز برای چهار وضعیت کاری به ترتیب ۹۵، ۵۶، ۶۹ و ۷۱ درصد بهبود پیدا کرده‌است. بنابر این شکل‌های ۵ و ۶ و همچنین تحلیل داده‌های جدول۴ ثابت می‌کنند که روش کنترلی پیشنهادی مبتنی بر یادگیری تقویتی بسیار کاراتر از روش‌های دیگر است.

جدول (۴): مقایسه عملکرد انواع PSS در وضعیت‌های کاری مختلف

	وضعیت کاری	معیار	CPSS	PSS3B	PSS3B+RL
۱	$T_m = 0.9, P_0 = 1.2, Q_0 = 0.2, X_e = 0.7$	%OS	1.2	1.09	**0.88**
		%US	1.08	0.842	**0.7**
		T_s [s]	3.2784	1.0101	**1.009**
		ITAE	0.1262	0.0401	**0.0248**
		ISTSE	0.0082	0.0024	**0.001**
		ISE	7.08e-5	2.2e-5	**1e-6**
۲	$T_m = 0.9, P_0 = 1.2, Q_0 = 0.2, X_e = 0.7$	%OS	1.19	1.08	**0.87**
		%US	1.06	0.8376	**0.6956**
		T_s [s]	3.2483	1.4599	**1.3556**
		ITAE	0.1224	0.0399	**0.0247**
		ISTSE	0.0078	0.0022	**0.0008**
		ISE	6.77e-5	2.26e-5	**9.97e-6**
۳	$T_m = 1.2, P_0 = 0.7, Q_0 = 0.5, X_e = 0.3$	%OS	1.83	1.69	**1.37**
		%US	1.73	1.39	**1.22**
		T_s [s]	10.9621	1.5069	**1.3502**
		ITAE	0.4227	0.1026	**0.0410**
		ISTSE	0.0557	0.0106	**0.003**
		ISE	3.98e-4	9.72e-5	**2.85e-5**
۴	$T_m = 1.2, P_0 = 1, Q_0 = 0, X_e = 0.5$	%OS	1.84	1.7	**1.37**
		%US	1.76	1.42	**1.23**
		T_s [s]	10.5223	1.7864	**1.3388**
		ITAE	0.4214	0.1027	**0.0405**
		ISTSE	0.00554	0.0107	**0.0029**
		ISE	9.96e-4	9.74e-5	**2.81e-5**

۸- نتیجه‌گیری

در این مقاله، پایدار ساز سیستم قدرت PSS3B توسط الگوریتم هوشمند KH طراحی گردید. به منظور طراحی مقاوم کنترل کننده تابع هدف مورد استفاده مقادیر ویژه ناپایدار و یا کم میرا شده سیستم را به سمت پایداری و میرا شدن شیفت داد. بعد از تنظیم بهینه پارامترهای PSS3B عملکرد آن توسط الگوریتم یادگیری تقویتی ارایه شده بر اساس روش Q-learning به صورت بلادرنگ کارایی آن بهینه گردید. از ویژگی اساسی پایدارساز ترکیبی پیشنهادی مبتنی بر یادگیری تقویتی سادگی و عدم وابستگی آن به مدل سیستم و مقاوم بودن آن به تغییرات نقاط کار بهره برداری و اختلالات وارده بر سیستم است.. بعد از اعمال شیوه کنترلی پیشنهادی، عملکرد سه نوع روش کنترلی یعنی پایدار ساز سیستم قدرت معمولی، پایدار ساز PSS3B و PSS3B+RL با تحت اغتشاش قرار دادن سیستم در شرایط بهره-برداری مختلف و اعمال خطای سه فاز مورد ارزیابی قرار گرفت که نتایج شبیه سازی نشان می‌دهد که روش کنترلی پیشنهادی با در هم آمیختن ویژگی های پایدار ساز سیستم قدرت سه باند و یادگیری تقویتی مبتنی بر Q-learning منجر به ساختار کنترلی ساده و انعطاف‌پذیر شده و توانایی بالایی در میرا کردن نوسانات فرکانس پایین و بهبود پایداری دینامیکی سیستم قدرت دارد. برای نشان دادن برتری روش پایدارساز پیشنهادی، نتایج شبیه‌سازی در وضعیت‌های کاری مختلف از حیث مقادیر فراجهش، فروجهش، زمان نشست، ITAE، ISTSE و ISE با هم مقایسه شدند و نتایج به روشنی برتری روش پیشنهادی را تایید می‌کند. بنابر این می‌توان نتیجه گرفت که یادگیری

[Downloaded from jiaeee.com on 2026-11-20]
[DOR: 20.1001.1.26765810.1396.14.3.6.1]

تقویتی می‌تواند مکمل و حتی جایگزین مناسبی برای پایدار سازهای سیستم قدرت باشد.

ضمایم

فرمول (۱۸) معادلات دینامیکی سیستم قدرت مورد مطالعه را نشان می‌دهد.

$$\begin{aligned}
&\frac{d\delta}{dt} = \omega_B S_m \\
&\frac{dS_m}{dt} = \frac{1}{2H}\left[-DS_m + T_m - T_e\right] \\
&\frac{dE'_d}{dt} = \frac{1}{T'_{q0}}\left[-E'_d - (x_q - x'_q) i_q\right] \\
&\frac{dE'_q}{dt} = \frac{1}{T'_{d0}}\left[-E'_q + (x_d - x'_d) i_d + E_{fd}\right] \\
&\frac{dE_{fd}}{dt} = \frac{1}{T_a}\left[K_a(V_{ref} + V_s - V_t) - E_{fd}\right] \\
&\begin{bmatrix} \upsilon_d \\ \upsilon_q \end{bmatrix} = \begin{bmatrix} E'_d \\ E'_q \end{bmatrix} - \begin{bmatrix} 0 & x'_q \\ -x'_d & 0 \end{bmatrix} \begin{bmatrix} i_d \\ i_q \end{bmatrix} \\
&\begin{bmatrix} i_d \\ i_q \end{bmatrix} = \begin{bmatrix} 0 & X_e \\ -X_e & 0 \end{bmatrix}^{-1} \left(\begin{bmatrix} \upsilon_d \\ \upsilon_q \end{bmatrix} + E'_b \begin{bmatrix} \sin\delta \\ -\cos\delta \end{bmatrix} \right)
\end{aligned} \qquad (۱۸)$$

معادلات معیارهای مورد استفاده در این مقاله در رابطه۱۹ نشان داده شده‌اند.

$$\begin{aligned}
ITAE &= \int_0^{t_{sim}} t\,|\Delta\omega| dt \\
ISTSE &= \int_0^{t_{sim}} t^2 \Delta\omega^2 dt \\
ISE &= \int_0^{t_{sim}} \Delta\omega^2 dt
\end{aligned} \qquad (۱۹)$$

مراجع

[1] Anderson, M. and Fouad, A. A., Power system control and stability, Ames: IA: Iowa State Univ. Press, 1977.

[2] Dehghani M, Nikravesh S, Karrari M. “Decentralized Robust Power System Stabilizer Design”, Journal of Iranian Association of Electrical and Electronics Engineers, Vol. 4, No. 1, pp. 36-43, 2007.

[3] Khodabakhshian, A., Hemmati R. and Moazzami M., "Multi-band power system stabilizer design by using CPCE algorithm for multi-machine power system," Electric Power Systems Research, Vol. 101, pp. 36-48, 2013.

[4] He, P., Wen, F., Ledwich, G., Xue, Y. and Wang, K., "Effects of various power system stabilizers on improving power system dynamic performance," Electrical Power and Energy Systems, Vol. 46, pp. 175-183, 2013.

[5] Farahani, M., "A multi-objective power system stabilizer," IEEE Transactions on Power Systems, Vol. 28, No. 3, pp. 2700-2707, 2013.

[6] Malik, O. P. and Hariri, A., "Power system stabilizer based on a self-learning adaptive network fuzzy inference system," Transactions of the Institute of Measurement and Control, vol. 24, no. 2, pp. 153-173, 2002.

[7] Taylor, C. W., "Response-based, feedforward wide-area control," in NSF/DOE/EPRI Sponsored Workshope on Future Research Directions for Complex Interactive Networks, Washington DC, USA, 2000.

[8] Liu, C. C., Jung, J., Heydt, G. T. and Vittal, V., "The strategic power infrastructure defense (SPID) system," IEEE Control System Magazine, pp. 40-52, 2000.

[9] Diu, A. and Wehenkel, L., "EXaMINE-Experimentat on of a monitoring and control system for managing vulnerabilitis of the european infrastructure for electrical power exchange," in IEEE PES Summer Meeting, Chicago, USA, 2002.

[10] Ernst, D., Glavic, M. and Wehenkel, L., "Power system stability control: Reinforcement learning framwork," IEEE Transaction on Power Systems, Vol. 19, No. 1, pp. 427-435, 2004.

[11] Yu, T. and Zhen, W. G., "A reinforcement learning approach to power system stabilizer," in IEEE Power & Energy Society General Meeting, Calgary, AB, 2009.

[12] Vlachogiannis, J. G. and Hatziargyriou, N. D., "Reinforcement learning for reactive power control," IEEE Transaction on Power Systems, Vol. 19, No. 3, pp. 1317-1325, 2004.

[13] Naduri, V. and Das, T. K., "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.

[14] Safari, A., Shayeghi, H., Jalilzadeh, S. , “Robust Coordinated Design of UPFC Damping Controller and PSS Using Chaotic Optimization Algorithm”, Journal of Iranian Association of Electrical and Electronics Engineers, Vol. 12, No. 3, pp. 55-62, 2015.

[15] Singh, R., "A novel approach for tuning of power system stabilizer using genetic algorithm," Master of Sience dissertaition, Department of Electrical Engineering, Indian Institute of Science, Bangalor, India, 2004.

[16] IEEE recommended practice for excitation system models for power system stability studies. [Performance]. IEEE Standard 421.5-2005, 2006.

[17] Gandomi, A. H. and Alavi, A. H., "Krill herd: A new bio-inspired optimization algorithm," Commun Nonlinear Sci Number Simulat, Vol. 17, pp. 4831-4845, 2012.

[18] Abdel-Magid, Y. L. and Abido, M. A., "Optimal multiobjective design of robust power system stabilizers ssing genetic algorithms," IEEE Transaction on Power Systems, Vol. 18, No. 3, pp. 1125-1132, 2003.

[19] Padiyar, K. R., Power System Dynamics, Giniraj Lane, Sultan Bazar, Hyderabad: BS Publications, 2008.

[Downloaded from jiaeee.com on 2026-11-20]

[DOR: 20.1001.1.26765810.1396.14.3.6.1]

Journal of Iranian Association of Electrical and Electronics Engineers Vol14 No.3 Fall2017