



A Hybrid Reinforcement Learning and Adaptive Fuzzy Sliding-Mode Control Strategy for Switching Systems

ARTICLE INFO

Article Type

Original Research

Authors

Safari H.¹,
Khanmirza E.^{1*},
Davaei Markazi A.¹

How to cite this article

Safari H., Khanmirza E., Davaei Markazi A., A Hybrid Reinforcement Learning and Adaptive Fuzzy Sliding-Mode Control Strategy for Switching Systems. Modares Mechanical Engineering; 2025;25(02):87-95.

¹ School of Mechanics, Iran University of Science and Technology, Tehran, Iran.

*Correspondence

Address: School of Mechanics, Iran University of Science and Technology, Tehran, Iran.

khanmirza@iust.ac.ir

Article History

Received: November 1, 2024

Accepted: April 15, 2025

ePublished: April 26, 2025

ABSTRACT

This article investigates the challenging problem of simultaneous control of discrete and continuous inputs in switching systems. Switching systems, which are a special type of hybrid systems, require advanced control methods that can simultaneously manage continuous and discrete inputs due to their hybrid nature. An innovative approach is introduced in this paper, combining Adaptive Fuzzy Sliding Mode Control (AFSMC) with Reinforcement Learning (RL). This method not only manages both types of inputs simultaneously but also operates adaptively and robustly with minimal model information, performing learn and optimize online. To evaluate the performance and verify the proposed algorithm, the two-tank system is selected as a benchmark example in this field. The simulation results showed that the tank level tracking error is reduced to less than 1 cm despite the noise in the measurement with a standard deviation of 0.005 and also the sudden change of the system parameter. Additionally, the number of valve position changes decreased to 6 after 1000 episodes, indicating a significant reduction in switching frequency and an improvement in system stability. This algorithm achieves desired objectives with lower control costs compared to non-hybrid methods (management of discrete and continuous inputs). Furthermore, this approach can serve as a scalable framework for controlling other complex systems with combined inputs across various engineering domains.

Keywords Adaptive-Fuzzy-Sliding Mode Control, Reinforcement Learning, Switching System, Two Tank System

CITATION LINKS

1- modelling and control approach for a type of mixed logical dynamical system using in chilled water system of refrigeration system. 2- Stability Analysis and Voltage Control in the Power System Based on the Hybrid Automata Model. 3- Practical hybrid model predictive control for electric pneumatic braking system with on-off solenoid valves. Proceedings of the Institution of Mechanical Engineers, Part D: 4- A unified switching control framework for continuous robot-assisted training. 5- Challenges to control: A collective view. 6- Stabilization of the networked hybrid systems with communication constraints and its applications. 7- Stable MPC Design for ... 8- Adaptive fuzzy output-feedback tracking control for switched nonstrict-feedback nonlinear systems with prescribed performance. 9- Modeling of three-tank hybrid system using Mixed Logical Dynamical formalism. 10- Robust Optimal Control of Uncertain Nonlinear Switched System using Approximate Dynamic Programmin. 11- ... 12- Continuous-discrete reinforcement learning for hybrid control in robotics. 13- Control of switched systems with delayed states: A direct AFSMC approach. 14- Design of state-dependent switching with dwell time constraint for interval positive switched systems. 15- Control of Switched Linear Systems: Adaptation and Robustness. 16- Adaptive fuzzy sliding-mode control for multi-input multi-output chaotic systems. 17- Q-learning. Mach Learn. 18- Reinforcement learning: An introduction. 19- Explicit model-predictiv e control of hybrid dynamical systems: Application to a two-tank system. 20- Liquid level tracking control of three-tank systems.

طراحی رویکرد کنترلی مبتنی بر یادگیری تقویتی و روش تطبیقی-فازی-لغزشی جهت کنترل سیستم‌های سوئیچینگ

حمیدرضا صفری^۱، اسماعیل خان‌میرزا^۲، امیرحسین دوائی مرکزی^۱

^۱ دانشکده مکانیک، دانشگاه علم و صنعت ایران، تهران، ایران

چکیده

این مقاله به بررسی مسئله چالش برانگیز کنترل همزمان ورودی‌های گسسته و پیوسته در سیستم‌های سوئیچینگ، می‌پردازد. سیستم‌های سوئیچینگ که نوع خاصی از سیستم‌های ترکیبی هستند، به دلیل ماهیت ترکیبی خود، نیازمند روش‌های کنترلی پیشرفته‌ای هستند که بتوانند به‌طور همزمان ورودی‌های پیوسته و گسسته را مدیریت کنند. در این مقاله، یک رویکرد نوآورانه با ترکیب کنترل تطبیقی-فازی-مد لغزشی (AFSMC) و یادگیری تقویتی (RL) معرفی شده است. این روش نه تنها قابلیت مدیریت همزمان این دو نوع ورودی را دارد، بلکه قادر است با داشتن حداقل اطلاعات از مدل، به‌صورت تطبیقی و مقاوم عمل کرده و در حالت آنلاین به یادگیری و بهینه‌سازی بپردازد. برای ارزیابی عملکرد و راستی آزمایی این الگوریتم، سیستم دو مخزن به عنوان یک نمونه معیار در این حوزه انتخاب شده است. نتایج شبیه‌سازی نشان داد که خطای ردیابی سطح مخازن با وجود نویز در اندازه‌گیری با انحراف معیار ۰/۰۰۵ و همچنین تغییر ناگهانی پارامتر سیستم، به کمتر از ۱ سانتی‌متر کاهش می‌یابد. همچنین، تعداد تغییرات وضعیت شیرها به ۶ بار بعد از ۱۰۰۰ اپیزود (episode) می‌رسد که نشان‌دهنده کاهش چشمگیر تعداد سوئیچینگ و افزایش پایداری سیستم است. این الگوریتم با کاهش هزینه کنترلی در مقایسه با روش‌های غیر ترکیبی (مدیریت همزمان ورودی گسسته و پیوسته)، به اهداف مطلوب دست می‌یابد. همچنین، این رویکرد می‌تواند به عنوان یک چارچوب قابل توسعه برای کنترل سایر سیستم‌های پیچیده با ورودی‌های ترکیبی در حوزه‌های مختلف مهندسی مورد استفاده قرار گیرد.

کلیدواژه‌ها: کنترل تطبیقی-فازی-مد لغزشی، یادگیری تقویتی، سیستم سوئیچینگ، سیستم دو مخزن

تاریخ دریافت: ۱۴۰۳/۰۸/۱۱

تاریخ پذیرش: ۱۴۰۴/۰۱/۲۶

*نویسنده مسئول: khammirza@iust.ac.ir

۱- مقدمه

در سال‌های اخیر، سیستم‌های ترکیبی به یکی از حوزه‌های تحقیقاتی برجسته و جذاب برای مهندسان تبدیل شده است [1]. زیرا در عمل، بیشتر سیستم‌های دینامیکی ماهیت ترکیبی دارند و در صنایع و زمینه‌های مختلف، مانند سیستم‌های الکتریکی [2]، خودرو [3]، سیستم‌های رباتیک [4]... مشاهده می‌شوند. مفهوم سیستم‌های ترکیبی اولین بار در [5] مطرح شد. این عبارت به سیستم‌هایی اطلاق می‌شود که در آن دینامیک پیوسته و گسسته باهم در تعامل باشند. این مورد در بسیاری از سیستم‌های تکنولوژیکی که در آن‌ها تصمیم‌گیری منطقی و اقدامات کنترلی با فرآیندهای فیزیکی پیوسته ترکیب می‌شوند، صادق است. از آنجایی که این سیستم‌ها روز به روز در حال پیچیده‌تر شدن

هستند، نیاز شدیدی به نظریه دقیق سیستم‌های ترکیبی وجود دارد که امکان مدل‌سازی، تحلیل و کنترل سیستم‌هایی را فراهم می‌کند که در آن عناصر منطقی و فرآیندهای پیوسته با هم در تعامل باشند.

موضوع مقاله حاضر، کنترل نوع خاصی از سیستم‌های ترکیبی می‌باشد که به سوئیچینگ سیستم‌ها معروف هستند. سیستم‌های سوئیچینگ کلاس خاصی از سیستم‌های ترکیبی هستند که از چندین زیر سیستم و یک قانون کلیدزنی تشکیل شده است که زیرسیستم‌های فعال را در هر لحظه مشخص می‌کند. به عبارت دیگر، سیستم‌های سوئیچینگ شامل تعدادی دینامیک‌های پیوسته هستند که به آن‌ها زیر سیستم گفته می‌شود و همچنین، شامل سیگنال کلیدزنی می‌باشند که نحوه ارتباط این زیرسیستم‌ها را با همدیگر تعیین می‌کند. در واقع، این سیگنال کلیدزنی می‌باشد که در هر لحظه مشخص می‌کند کدام یک از زیرسیستم‌ها فعال باشند.

مقالات متعددی در زمینه کنترل و تجزیه و تحلیل سیستم‌های ترکیبی و سوئیچینگ موجود است که در ادامه بررسی خواهند شد. سیستم‌های هیبریدی شبکه‌ای که شامل ترکیب دینامیک‌های پیوسته و گسسته هستند، چالش‌های منحصربه‌فردی در طراحی و پایداری دارند. یان هی و همکاران [6]، راهکارهایی برای تثبیت این سیستم‌ها با در نظر گرفتن محدودیت‌های ارتباطی و تأخیرات شبکه‌ای ارائه کرده‌اند. این پژوهش از تکنیک‌های پیشرفته نظیر کوانتیزاسیون لوگاریتمی و الگوریتم‌های بهینه‌سازی برای بهبود عملکرد و کاهش هزینه‌های کنترلی بهره می‌گیرد. چنین روش‌هایی می‌توانند زمینه‌ساز توسعه سیستم‌های هوشمند و پایدار در حوزه‌های متنوع مهندسی باشند. علما و همکاران [7]، به بررسی شرایط پایداری نمایی حلقه-بسته برای سیستم‌های ترکیبی توصیف شده با مدل مرکب-منطقی-دینامیکی و کنترل پیش‌بین مدل پرداخته‌اند. این تحقیق استفاده از توابع لیاپانوف برای تضمین پایداری سیستم‌ها را معرفی می‌کند و نشان می‌دهد که این روش بهبودهایی در کاهش پیچیدگی‌های محاسباتی و پیاده‌سازی در مقایسه با روش‌های سنتی ارائه می‌دهد.

در مقاله [8]، یک طرح کنترل ردیابی خروجی-بازخورد فازی تطبیقی برای کلاسی از سیستم‌های غیرخطی سوئیچ‌شونده تک ورودی و تک خروجی در فرم بازخورد غیر دقیق با عملکرد تجویز شده و سوئیچینگ دلخواه پیشنهاد شده است. ژانگ و همکاران [1]، سیستم آب سرد تهویه مطبوع مرکزی را به عنوان یک سیستم ترکیبی در نظر می‌گیرند. تنظیم دینامیکی ظرفیت خنک‌کننده بر اساس سیستم ترکیبی می‌تواند به کنترل دقیق دما و صرفه جویی در زمان واقعی انرژی منجر شود. این مقاله یک نوع اصلاح‌شده جدید از سیستم دینامیک منطقی مختلط را پیشنهاد می‌کند که عملکرد مدل را افزایش می‌دهد و انعطاف‌پذیری سوئیچینگ و اثر کنترل را بهبود می‌بخشد. یاکوبی و همکاران [9] از روش مدل‌سازی

ورودی‌های گسسته و پیوسته را ندارند. یکی از خلأهای جدی در این حوزه، نبود روش‌های مستقل از مدل برای کنترل سیستم‌های سوئیچینگ است، که نیازمند مدیریت همزمان ورودی‌های گسسته و پیوسته در محیط‌های متغیر و پیچیده هستند.

در این مقاله، روشی نوین مبتنی بر یادگیری تقویتی ارائه شده است که از ترکیب الگوریتم Q-learning با کنترل تطبیقی-فازی-مد لغزشی استفاده می‌کند. این ترکیب منحصربه‌فرد چندین ویژگی کلیدی و نوآورانه را ارائه می‌دهد:

۱. کنترل مستقل از مدل: الگوریتم پیشنهادی با حداقل اطلاعات از مدل سیستم کار می‌کند و به صورت آنلاین تطبیق می‌یابد، که آن را از سایر روش‌های وابسته به مدل متمایز می‌کند.

۲. مدیریت همزمان ورودی‌های ترکیبی: روش ارائه‌شده به طور همزمان ورودی‌های پیوسته و گسسته را مدیریت می‌کند و پایداری سیستم را در شرایط سوئیچینگ تضمین می‌نماید.

۳. افزایش پایداری با کاهش تعداد سوئیچینگ: شبیه‌سازی‌ها نشان داده‌اند که الگوریتم پیشنهادی تعداد تغییرات سوئیچینگ را به ۶ بار در ۱۰۰۰ اپیزود کاهش داده و خطای ردیابی سطح آب را به کمتر از ۱ سانتی‌متر (با توجه به bound_1) می‌رساند.

۴. کاهش هزینه کنترلی: الگوریتم از یادگیری تقویتی برای اتخاذ تصمیمات بهینه در سوئیچینگ بهره می‌گیرد و از افزایش هزینه کنترلی جلوگیری می‌کند، که در روش‌های مشابه کمتر به آن توجه شده است.

در ادامه این مقاله به شرح زیر سازماندهی شده است: بخش ۲ به بیان مسئله می‌پردازد. بخش ۳ الگوریتم پیشنهادی RL-AFSMC شرح داده می‌شود. در بخش ۴ دینامیک یک سیستم سوئیچینگ معیار، سیستم دو مخزن بیان شده و نتایج شبیه‌سازی ارائه خواهد شد و در نهایت در بخش ۵ نتیجه‌گیری بیان می‌گردد.

۲- بیان مسئله

سیستم سوئیچینگ غیرخطی m-ورودی و m-خروجی آفین با معادلات زیر، در نظر گرفته می‌شود [14].

$$\dot{x}^{(n)} = F_{\sigma(t)}(x) + G_{\sigma(t)}(x)u + d \quad (1)$$

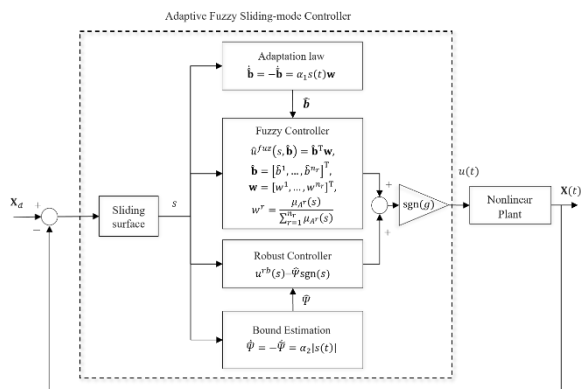
$$x = [x, \dot{x}, \ddot{x}, \dots, x^{(n-1)}]^T \in \mathbb{R}^n \quad (2)$$

که در آن u بردار ورودی کنترلی پیوسته، n مرتبه سیستم، d بردار اغتشاشات خارجی غیر قابل اندازه‌گیری، $F_{\sigma(t)}(x)$ توابع هموار و احتمالاً نامعلومی از x ، $G_{\sigma(t)}(x)$ یک ماتریس قطری و ثابت است که می‌تواند نامعلوم باشد با فرض اینکه $G_{\sigma(t)}^{-1}$ وجود داشته باشد. همچنین، $\sigma(t) \in \{1, \dots, k\}$ نماد سیگنال سوئیچینگ (ورودی

منطقی-دینامیکی ترکیبی برای سیستم‌های ترکیبی سه مخزن استفاده کردند. آن‌ها یک چارچوب پیش‌بین مبتنی بر تکنیک‌های برنامه‌ریزی عدد صحیح ارائه دادند که با ترکیب دینامیک‌های پیوسته و گسسته، به بهینه‌سازی عملکرد و کاهش خطا در سیستم کمک می‌کند. نتایج نشان داد که این روش قابلیت کنترل دقیق سیستم‌های پیچیده صنعتی را دارا بوده و پایداری سیستم را تضمین می‌کند. در مقاله‌ای که توسط بلوچ‌زاده و همکاران [10] ارائه شده است، یک رویکرد کنترل بهینه و مقاوم برای سیستم‌های سوئیچینگ غیرخطی با زیرسیستم‌های تحمیل شده و عدم قطعیت‌ها معرفی شده است. این مقاله به طراحی سیگنال سوئیچینگ و ورودی کنترل به‌منظور حداقل کردن تابع هزینه می‌پردازد. برای حل مشکلات ناشی از استفاده از برنامه‌ریزی دینامیکی تقریبی (ADP) در سیستم‌های سوئیچینگ با عدم قطعیت، یک مدل مناسب معرفی شده است. در این رویکرد، ابتدا عدم قطعیت‌ها با استفاده از یک کنترل کننده مقاوم زمان تاخیر جبران می‌شود و سپس سیگنال سوئیچینگ و ورودی کنترل با استفاده از ADP طراحی می‌گردد. چهاردولی و همکاران [11]، مسئله پایداری محدود زمانی سیستم‌های سوئیچینگ غیرخطی با استفاده از پسخور اکید و انتگرال‌گیرهای توان کسری را بررسی می‌کنند. توان‌های کسری در هر زیرسیستم متفاوت بوده و این امر ساخت تابع لیاپانوف مشترک را دشوار می‌کند. برای حل این مشکل، با تعریف یک انتقال مختصات مناسب و افزودن ترم‌های توانی انتگرال، یک قانون کنترلی فیدبک طراحی می‌شود که پایداری محدود زمانی هر زیرسیستم را تضمین می‌کند. سپس، با استفاده از قضیه تابع لیاپانوف مشترک، نشان داده می‌شود که این قانون کنترلی می‌تواند پایداری حلقه بسته سیستم را تحت هر سوئیچ دلخواه تضمین کند. نتایج شبیه‌سازی برای سیستم سوئیچینگ مرتبه ۳ نیز اثربخشی روش پیشنهادی را تایید می‌کند. نوئورت و همکاران [12]، با ارائه رویکرد یادگیری تقویتی مسائل کنترل سیستم‌های ترکیبی که توانایی کنترل همزمان ورودی‌های پیوسته و گسسته را دارا باشد، مورد بحث قرار می‌دهند. دوائی مرکزی و همکاران [13]، بر روی کنترل کلاسی از سیستم‌های غیرخطی چند ورودی-چند خروجی با تاخیر در حالت، همراه با شرایط سوئیچینگ که به حالت‌های تاخیری بستگی دارد، تمرکز دارند. ایده اصلی در این کار در نظر گرفتن تغییرات دینامیک ناشی از جابجایی بین زیرسیستم‌ها به عنوان عدم قطعیت می‌باشد. کنترل توسط روش تطبیقی-فازی-مد لغزشی توسعه یافته برای سیستم زیر سطحی انجام می‌شود.

با توجه به مطالعات پیشین، سیستم‌های ترکیبی و به‌ویژه سیستم‌های سوئیچینگ از جمله چالش‌های مطرح در زمینه مدل‌سازی و کنترل هستند. تاکنون روش‌های مختلفی برای مدیریت این سیستم‌ها ارائه شده است، اما اکثر آن‌ها یا به مدل‌های پیچیده و کامل نیاز دارند یا توانایی مدیریت همزمان

کننده‌ی فازی می‌باشد. بلوک دیاگرام نحوه عملکرد کنترل کننده تطبیقی-فازی-مد لغزشی در شکل ۱ مشخص شده است.



شکل (۱) بلوک دیاگرام روش کنترل تطبیقی-فازی-مد لغزشی

۳-۲- الگوریتم RL-AFSMC

در بخش قبل کنترل کننده تطبیقی-فازی-مد لغزشی به عنوان کنترل کننده مناسب جهت کنترل زیرسیستم‌های یک سیستم سوئیچینگ معرفی شد؛ اما نکته‌ای که باید در نظر گرفت، این کنترل کننده تنها قادر به کنترل زیرسیستم‌ها، بدون در نظر گرفتن عامل سوئیچ می‌باشد. به عبارتی دیگر، این کنترل کننده نقشی در رخداد تغییر زیر سیستم فعال را ندارد و صرفاً با توجه به زیر سیستمی که فعال می‌باشد تلاش به تولید ورودی کنترلی پیوسته جهت کنترل همان زیر سیستم فعال می‌نماید و توانایی تغییر زیر سیستم به زیر سیستم دیگر را ندارد. در این مقاله همان‌طور که در بخش ۲ بیان شد، رخداد سوئیچ در سیستم مورد بررسی باید وابسته به الگوریتم کنترلی باشد یعنی کنترل کننده توانایی تغییر زیر سیستم فعال را داشته باشد زیرا با توجه به هدف کنترلی، در هر لحظه از زمان زیر سیستمی که منجر به ارضاء بهتر اهداف کنترلی می‌شود می‌تواند متفاوت باشد. بنابراین، باید از الگوریتمی استفاده نمود که توانایی تغییر زیر سیستم فعال با توجه به هدف کنترلی را داشته باشد (روش پیشنهاد شده، استفاده از روش یادگیری تقویتی برای تولید ورودی کنترلی گسسته) و همزمان بتواند در زیر سیستم فعال فعلی عملیات کنترل را نیز اداره کند (که این وظیفه بر عهده کنترل کننده تطبیقی-فازی-مد لغزشی است). الگوریتمی که برای حل این مسئله ارائه می‌شود ترکیب روش یادگیری تقویتی با کنترل کننده تطبیقی-فازی-مد لغزشی می‌باشد. الگوریتم پیشنهادی علاوه بر کنترل هر یک از زیر سیستم‌های فعال، توانایی تصمیم‌گیری درباره زمان رخداد سوئیچ بین زیر سیستم‌های مختلف را نیز خواهد داشت.

در این رویکرد از الگوریتم یادگیری Q که مهم‌ترین الگوریتم یادگیری تقویتی و مستقل از مدل می‌باشد و به نوعی یک روش off-policy TD است استفاده می‌شود. این الگوریتم برای اولین بار

کنترلی گسسته) می‌باشد که در هر لحظه از زمان t تصمیم می‌گیرد کدام زیر سیستم آفین، فعال باشد و k تعداد زیرسیستم‌ها را نشان می‌دهد [15].

فرضیات حاکم بر مسئله از قرار ذیل می‌باشد.

(۱) سیستم سوئیچینگ از نوع وابسته به کنترل کننده می‌باشد.

(۲) توابع $F_{\sigma(t)}$ و $G_{\sigma(t)}$ پیوسته و محدود هستند.

(۳) $G_{\sigma(t)}^{-1}$ وجود داشته باشد.

(۴) مسیر $x(t)$ در همه جا پیوسته است.

هدف، طراحی کنترل کننده‌ای هوشمند برای سیستم سوئیچینگ (۱) می‌باشد تا بتواند با کنترل همزمان سیگنال سوئیچینگ و ورودی پیوسته، خطای ردیابی را کاهش دهد و اهداف از پیش تعیین شده را عملی سازد.

۳-۳- ارائه رویکرد کنترلی RL-AFSMC

۳-۳-۱- کنترل کننده تطبیقی-فازی-مد لغزشی

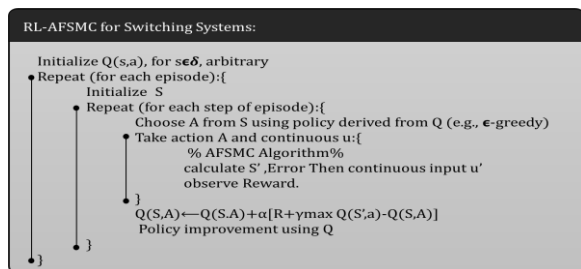
در این بخش به معرفی کنترل کننده تطبیقی-فازی-مد لغزشی می‌پردازیم که در الگوریتم پیشنهادی وظیفه تولید ورودی کنترلی پیوسته را خواهد داشت. در طراحی این کنترل کننده، مفهوم کنترل مد لغزشی با استراتژی کنترل فازی ترکیب شده است تا یک کنترل کننده مد لغزشی-فازی مستقل از مدل به دست آید. دو دیدگاه مختلف در رابطه با طراحی سیستم‌های کنترلی تطبیقی-فازی-مد لغزشی وجود دارد: ۱- دیدگاه مستقیم؛ ۲- دیدگاه غیر مستقیم. الگوریتم‌های کنترلی مستقیم از جهاتی نسبت به الگوریتم‌های غیرمستقیم آن موثرترند زیرا احتمال مواجه شدن با پدیده تکینگی در روش‌های مستقیم کمتر است. همچنین به این دلیل که در روش غیر مستقیم نسبت به روش‌های مستقیم از الگوریتم‌های فازی بیشتری برای تقریب استفاده می‌شود، همگرایی در روش مستقیم سریع‌تر اتفاق می‌افتد. بنابراین در این بخش، یک الگوریتم کنترلی مستقیم تطبیقی-فازی-مد لغزشی برای کنترل سیستم‌های چند ورودی-چند خروجی غیر خطی با اطلاعات حداقلی از مدل، ارائه می‌شود. ساختار کنترلی این الگوریتم شامل سیستم فازی برای تقریب کنترل کننده ایده‌آل مد لغزشی و کنترل کننده مقاوم برای جبران تفاوت میان کنترل کننده فازی و کنترل کننده ایده‌آل می‌باشد. پارامترهای سیستم فازی و کران عدم قطعیت‌های سیستم توسط قوانین تطبیقی تنظیم می‌شوند که بر پایه‌ی پایداری لیاپانوف استخراج می‌شوند که همگرا شدن خطای ردیابی به صفر را تضمین می‌کنند. اثبات پایداری این روش در مقاله [16] بررسی شده است.

در این روش قانون کنترلی به صورت رابطه زیر می‌شود [16].

$$u_i = \text{sgn}(g)[\hat{u}_i^{fuz}(s_i, \hat{b}_i) + u_i^{rb}(s_i)] \quad (3)$$

که کنترل کننده فازی $\hat{u}_i^{fuz}$ تقریب‌گر کنترل کننده ایده‌آل و u_i^{rb} جبران کننده اختلاف بین مقدار کنترل کننده ایده‌آل و کنترل

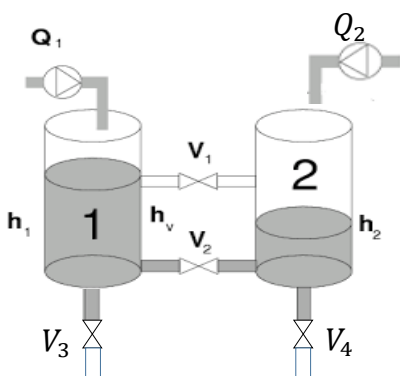
می باشد تا کنترل کننده AFSMC بتواند به همگرایی رسیده و از پدیده چترینگ (chattering) دوری کند.
الگوریتم پیشنهادی RL-AFSMC در شکل ۳ نشان داده شده است.



شکل ۳) الگوریتم روش RL-AFSMC برای سیستم های سوئیچینگ

۴- بررسی مورد مطالعاتی

همانطور که در شکل ۴ نشان داده شده است، سیستم معیار از دو مخزن (شماره ۱ و شماره ۲) و دو پمپ (شماره ۱ و شماره ۲) تشکیل شده است. پمپ ۱ و پمپ ۲ مایع را با دبی Q_1 و Q_2 بصورت ورودی پیوسته به ترتیب وارد مخزن ۱ و ۲ می کنند. همچنین، در اینجا چهار شیر با نام های V_1, V_2, V_3, V_4 وجود دارد که بصورت کلید زن هستند یعنی فقط می تواند بسته یا باز باشند. مخزن ۱ و مخزن ۲ از طریق شیر ۱ و ۲ نیز با یکدیگر در ارتباط هستند.



شکل ۴) شماتیک سیستم معیار دو مخزن [19]

دینامیک سیستم را می توان با مجموعه ای از معادلات دیفرانسیل ذیل بیان کرد [20]:

$$\begin{aligned} S\dot{h}_1 &= Q_1 - Q_{12} - Q_{10} - Q_{12h} \\ S\dot{h}_2 &= Q_2 + Q_{12} - Q_{20} + Q_{12h} \end{aligned} \quad (5)$$

که در آن

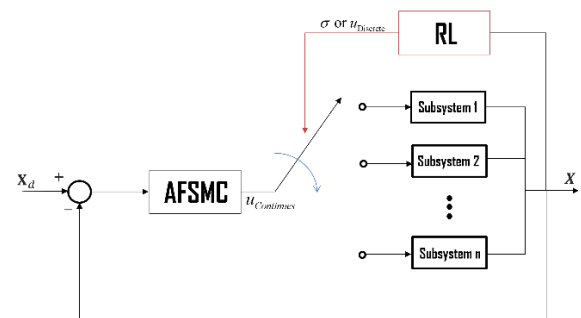
$$\begin{aligned} Q_{12h} &= V_1 \cdot az_1 \cdot S_n \cdot \text{sgn}(\max(h_1, h_v) - \max(h_2, h_v)) \cdot \sqrt{2g|\max(h_1, h_v) - \max(h_2, h_v)|} \\ Q_{12} &= V_2 \cdot az_2 \cdot S_n \cdot \text{sgn}(h_1 - h_2) \cdot \sqrt{2g|h_1 - h_2|} \\ Q_{10} &= V_3 \cdot az_3 \cdot S_n \cdot \sqrt{2gh_1} \\ Q_{20} &= V_4 \cdot az_4 \cdot S_n \cdot \sqrt{2gh_2} \end{aligned} \quad (6)$$

توسط واتکینز [17] ارائه شد. در این الگوریتم ارزش های بهینه مستقل از سیاست رفتاری به صورت زیر تخمین زده می شوند [18].

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha [R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)] \quad (4)$$

در اینجا، تابع ارزش حالت-عمل آموخته شده Q ، مستقل از سیاستی که دنبال می شود، مستقیماً q^* (تابع ارزش حالت-عمل بهینه) را تقریب می زند.

با توجه به خصوصیات سیستم های سوئیچینگ که در بخش ۲ شرح داده شد و همچنین نیاز این مسئله جهت کنترل و مدیریت آنلاین با داشتن اطلاعات حداقلی از مدل سیستم، ارائه چنین روشی با دارا بودن تمام خصوصیات روش های AFSMC و همچنین یادگیری تقویتی (یادگیری Q) مورد نیاز می باشد. بنابراین، الگوریتم پیشنهادی RL-AFSMC با توجه به توانمندی های طراحی شده برای آن، قادر به کنترل همزمان ورودی های پیوسته و گسسته بصورت آنلاین و با داشتن حداقل اطلاعات از مدل، خواهد بود. در شکل ۲ نمایش بلوک دیاگرامی این روش نشان داده شده است.



شکل ۲) بلوک دیاگرام روش RL-AFSMC

همانطور که در شکل ۲ مشاهده می شود، دو نوع بلوک AFSMC و یادگیری تقویتی (RL) داریم که هریک وظیفه تولید یک نوع خاص از ورودی کنترلی را دارند. بلوک AFSMC وظیفه تولید ورودی کنترلی پیوسته را دارد و بلوک RL ورودی کنترلی گسسته (سیگنال کلیدزنی) را تولید می کند. حال برای ارائه الگوریتم RL-AFSMC، در الگوریتم یادگیری Q و در بخش محیط آن تغییراتی ایجاد می کنیم. بدین منظور، کنترل کننده AFSMC را که وظیفه اعمال ورودی کنترلی پیوسته را دارد، به عنوان بخشی از محیط در نظر می گیریم و پاداش و جریمه را با توجه به حضور و فعالیت این کنترل کننده تعیین می کنیم. بدین گونه که کنترل کننده AFSMC بدون در نظر گرفتن عامل سوئیچ کننده، تلاش به کنترل زیر سیستم فعلی می کند. از طرفی الگوریتم یادگیری Q همواره در حال ارزیابی خروجی محیط با توجه به اهداف از پیش تعریف شده توسط کاربر می باشد. بنابراین، در زمان لازم الگوریتم یادگیری Q اقدام به سوئیچ برای نیل به اهداف را انجام خواهد داد. نکته حائز اهمیت در این الگوریتم پیشنهادی، مربوط به مدت زمان اقامت در هر زیر سیستم

$$e_1, e_2 = \begin{cases} 1 & \text{if } x_{1,2} - x_{1,2}^d \geq bound_1 \\ 0 & \text{if } |x_{1,2} - x_{1,2}^d| < bound_1 \\ -1 & \text{if } x_{1,2} - x_{1,2}^d < -bound_1 \end{cases} ; bound_1 = 1(cm)$$

$$e_3, e_4 = \begin{cases} 1 & \text{if } x_{1,2} - h_v \geq bound_2 \\ 0 & \text{if } |x_{1,2} - h_v| < bound_2 \\ -1 & \text{if } x_{1,2} - h_v < -bound_2 \end{cases} ; bound_2 = 1(cm) \quad (۸)$$

جدول ۱) پارامترهای الگوریتم کنترلی و سیستم دینامیکی دو مخزن

پارامتر	مقدار
T_{end}	۱۲۰ (s)
$time_step$	۰/۱ (s)
$X(0)$	[۸۰, ۴۰] (cm)
Q_{max}	۰/۰۳ (Lit/s)
h_{max}	۱۲۰ (cm)
h_v	۷۰ (cm)
S	۰/۷۸۵۴ (m^2)
S_n	۰/۰۰۷۸ (m^2)
az_1	۰/۶
az_2	۰/۶
az_3	۰/۹
az_4	۰/۹
g	۹/۸۱
$\lambda_{1,2}$	۰/۱
$b_{1,2}(0)$	[۰, ۰, ۰, ۰]
$\psi_{1,2}$	۰
α_1^{AFSMC}	۰/۸
α_2^{AFSMC}	۰/۰۱
$Episode$	۱۰۰۰
γ	۰/۸
α^{RL}	۰/۵
$\epsilon(\epsilon - greedy)$	۰/۹۸
$\rho(Reduction\ factor)$	۰/۹۸
$mean$	۰
Std	۰/۰۰۵

همچنین فضای عمل شیرها را نیز به دو حالت بسته و یا باز با مقادیر ۰ و ۱ گسسته می‌کنیم. با توجه به اینکه در سیستم چهار شیر وجود دارد، بنابراین کل حالات ممکن از آرایش شیرها، تعداد $۲^۴ = ۱۶$ خواهد بود. همچنین کل حالات ممکن از وضعیت سیستم با توجه به متغیرهای $e_1, \dots, e_4$ تعداد $۳^۴ = ۸۱$ حالت خواهد بود. بعد از گسسته‌سازی محیط، نوبت به تعریف پاداش و جریمه در الگوریتم یادگیری تقویتی می‌رسد. پاداش و جریمه مطابق با رابطه ذیل تعریف می‌شود.

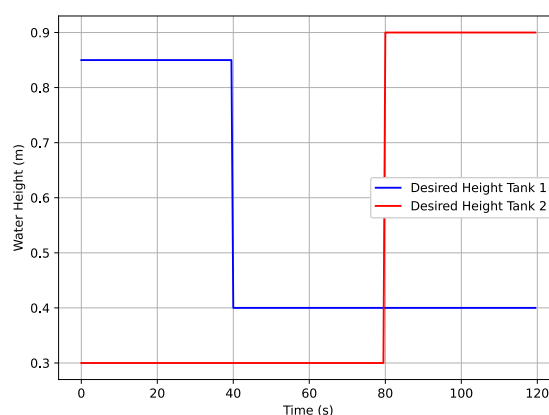
می‌باشد و $x = [h_1, h_2]$ بردار متغیر حالت و $u = [Q_1, Q_2, V_1, V_2, V_3, V_4]$ بردار ورودی کنترلی است بگونه‌ای که:

$$u = [u_{continuous}, u_{discrete}]$$

$$u_{continuous} = [Q_1, Q_2]$$

$$u_{discrete} = [V_1, V_2, V_3, V_4] \quad (۷)$$

مقادیر پارامترهای سیستم در جدول ۱ آورده شده است. هدف از کنترل این سیستم، ردیابی سطح آب در مخزن ۱ و ۲ با توجه به مسیر شکل ۵ می‌باشد. همچنین جهت بررسی مقاوم بودن الگوریتم به نویز و تغییر پارامتر سیستم، در اینجا نویز سفید با مقادیر میانگین و انحراف معیار مشخص شده در جدول ۱ به اندازه‌گیری ارتفاع سطح آب اعمال می‌شود. همچنین، مقدار پارامتر az_3 بعد از ثانیه ۴۲ به مقدار ۰/۴۵ کاهش پیدا می‌کند.



شکل ۵) مقادیر مطلوب سطح آب در مخازن جهت ردیابی

پارامتر p در جدول ۱ به عنوان ضریب کاهش معرفی می‌گردد. این پارامتر بعد از هر اپیزود (episode) در پارامتر ϵ ضرب می‌شود و هدف از این کار افزایش احتمال انتخاب عمل گریزی (greedy) با گذشت هر اپیزود می‌باشد که نهایتاً منجر به تولید سیاست قطعی خواهد شد.

با توجه به خصوصیت روش یادگیری Q ابتدا به گسسته‌سازی محیط می‌پردازیم. از اینرو، برای محیط یادگیری، چهار متغیر حالت جهت تعیین وضعیت سیستم در نظر می‌گیریم. این چهار متغیر عبارتند از:

e_1 : خطای ردیابی مربوط به مخزن ۱ که دارای سه مقدار $[-۱, ۰, ۱]$ می‌باشد.

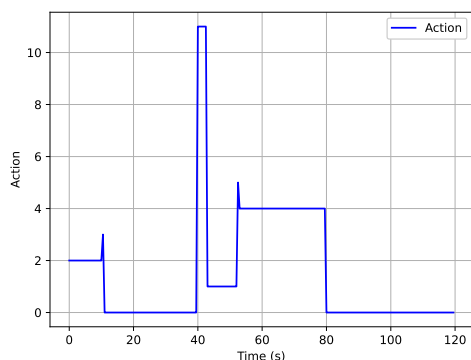
e_2 : خطای ردیابی مربوط به مخزن ۲ که دارای سه مقدار $[-۱, ۰, ۱]$ می‌باشد.

e_3 : وضعیت ارتفاع سطح آب در مخزن ۱ نسبت به ارتفاع h_v که دارای سه مقدار $[-۱, ۰, ۱]$ می‌باشد.

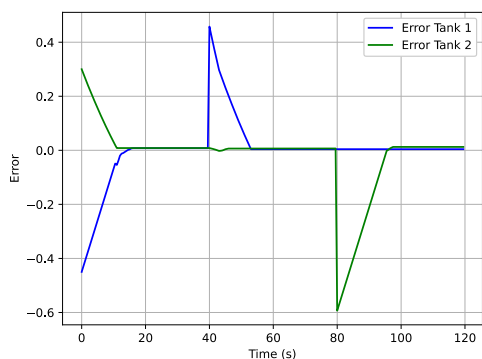
e_4 : وضعیت ارتفاع سطح آب در مخزن ۲ نسبت به ارتفاع h_v که دارای سه مقدار $[-۱, ۰, ۱]$ می‌باشد.

که داریم:

جریان‌های ورودی Q_1 و Q_2 اعمال شده و صرفاً توانایی افزایش ارتفاع سطح آب در صورت منفی بودن مقدار خطای مخزن را دارد. در مقابل، ورودی کنترلی گسسته که شامل باز و بسته شدن شیرها است، علاوه بر ایفای نقش در کاهش ارتفاع سطح آب مخازن در صورت مثبت بودن مقدار خطا، با بهره‌گیری از یادگیری سیاست مناسب، آرایش بهینه‌ای از وضعیت شیرها را انتخاب می‌کند تا به ورودی کنترلی AFSMC در افزایش ارتفاع سطح آب مخازن در شرایط خطای منفی کمک کند. این تعامل هماهنگ بین ورودی‌های گسسته و پیوسته، دقت و کارایی سیستم کنترلی را به میزان قابل‌توجهی بهبود می‌بخشد.



شکل ۸) ورودی کنترلی گسسته (سوئیچ بین آرایش مختلف شیرها) در اپیزود ۱۰۰۰



شکل ۹) خطای ردیابی در اپیزود ۱۰۰۰

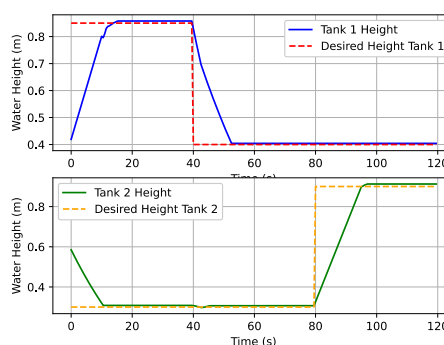
با توجه به توضیحات ارائه شده، در نمودار شکل ۶ در هر زمان که ارتفاع سطح آب نسبت به ارتفاع مطلوب کمتر می‌باشد کنترل‌کننده AFSMC تلاش کرده و با اعمال حداکثر ورودی دبی سطح آب را بالا آورده و به سطح مطلوب می‌رساند. از طرفی، در هر زمان که ارتفاع سطح آب نسبت به ارتفاع مطلوب بیشتر باشد، کنترل‌کننده RL با اتخاذ آرایشی مناسب از باز یا بسته بودن شیرها، میبایست منجر به کاهش سریع ارتفاع شده تا به سطح مطلوب برسد، ولی چنین اتفاقی در اپیزود ۱ رخ نمی‌دهد و سیستم دارای خطای زیادی خواهد بود. این موضوع بدین علت می‌باشد که عامل در ابتدای یادگیری قرار داشته و هنوز تعاملی با محیط نداشته است، بنابراین سیاست بهینه یافت نشده است. بدیهی است در

$$Reward = \begin{cases} -20|Error| - 10U_{AFSMC} + NotChangingValves & \text{if } |Error| > bound_1 \\ 150 - 10U_{AFSMC} + NotChangingValves & \text{if } |Error| \leq bound_1 \end{cases} \quad (9)$$

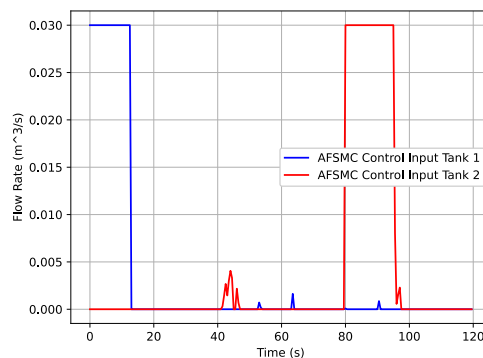
در رابطه فوق، عبارت اول منجر به کاهش خطای ردیابی خواهد شد زیرا با کاهش خطا در زمان فعلی نسبت به زمان قبلی، جریمه کمتری تولید می‌شود. عبارت دوم نیز منجر به انتخاب آرایشی از شیرها خواهد شد که باعث کاهش دبی ورودی به سیستم می‌شود و افزایش آن منجر به تولید جریمه خواهد شد. عبارت NotChangingValves یک پارامتر پاداش می‌باشد که منجر به کاهش سوئیچ‌های پی در پی می‌شود و عامل سعی می‌کند آرایشی را در نظر بگیرد که منجر به کمترین سوئیچ در طول شبیه‌سازی شود (این مورد جهت دوری از کاهش زمان اقامت اعمال شده است).

۵- جمع‌بندی و نتایج راستی آزمایی

حال به جمع‌بندی و بحث درباره نتایج حاصل از شبیه‌سازی می‌پردازیم.



شکل ۶) ردیابی سطح آب در اپیزود ۱۰۰۰



شکل ۷) ورودی کنترلی پیوسته (ورودی دبی) در اپیزود ۱۰۰۰

نمودارهای ارائه‌شده نتایج شبیه‌سازی در اپیزود ۱۰۰۰ را با وجود نویز اندازه‌گیری که منجر به خطای اندازه‌گیری $\pm 1cm$ می‌شود، و همچنین تغییر پارامتر az_3 از $0/9$ به مقدار $0/45$ از ثانیه ۴۲ به بعد را نمایش می‌دهند. شکل ۶، نمودار ردیابی سطح آب در مخازن ۱ و ۲ را نشان می‌دهد. ورودی کنترلی پیوسته که توسط کنترل‌کننده تطبیقی-فازی-مد لغزشی (AFSMC) مدیریت می‌شود، از طریق

با توجه به توضیحات مطرح شده در بخش قبل، مشاهده شد که الگوریتم پیشنهادی به شکل مناسبی قادر به کنترل و مدیریت این نوع از سیستم‌های خاص می‌باشد و همچنین، این الگوریتم توانست این نوع از سیستم را با داشتن اطلاعات حداقلی از مدل و بصورت آنلاین و تطبیقی یاد گرفته و کنترل کند.

همچنین، یکی از چالش‌های اصلی این تحقیق، تحلیل و اثبات پایداری سیستم حلقه بسته با در نظر گرفتن شرایط سوئیچینگ با استفاده از الگوریتم پیشنهادی است. این موضوع همچنان در دست بررسی است و در تحقیقات آینده به تفصیل مورد ارزیابی و حل قرار خواهد گرفت.

تاییدیه اخلاقی: محتویات علمی حاصل فعالیت‌های پژوهشی نویسندگان بوده و صحت نتایج آن نیز بر عهده نویسندگان مقاله است. همچنین تایید می‌شود که این مقاله در نشریه دیگری چاپ نشده و یا در دست بررسی برای چاپ نمی‌باشد.

تعارض منافع: در فرآیند انجام و گزارش، بیطرفی رعایت شده و تعارض منافعی وجود نداشته است.

منابع مالی: منابع مالی توسط نویسندگان تامین شده است.

فهرست علائم

بردارهای خروجی فازی	$b_{1,2}(0)$
شتاب گرانش (m/s^2)	g
ارتفاع سطح آب در مخزن i (m)	h_i
ارتفاع شیر ۱ از سطح مخزن‌ها (m)	h_v
دبی ورودی به مخزن i (m^3/s)	Q_i
شیر i (۰ یا ۱)	V_i
دبی ورودی از مخزن ۱ به ۲ از طریق V_2 (m^3/s)	Q_{12}
دبی ورودی از مخزن ۱ به ۲ از طریق V_1 (m^3/s)	Q_{12h}
دبی خروجی از مخزن i (m^3/s)	Q_{i0}
سطح مقطع مخزن ۱،۲ (m^2)	S
سطح مقطع لوله‌های اتصال (m^2)	S_n
میانگین نویز سفید	$mean$
انحراف معیار نویز سفید	Std
علائم یونانی	
نرخ یادگیری مربوط به b_i	α_1^{AFSMC}
نرخ یادگیری مربوط به ψ_i	α_2^{AFSMC}
ضریب جریان در شیر j ($j=1,2,3,4$)	α_{z_j}
اندازه گام الگوریتم RL	α_{RL}
پارامتر سیاست $\epsilon - greedy$	ϵ
سیگنال سوئیچینگ	σ
پارامتر متغیر لغزشی	$\lambda_{1,2}$
حداکثر عدم قطعیت فازی	$\psi_{1,2}$
ضریب کاهش	ρ
نرخ تخفیف	γ

این شرایط عامل بصورت تصادفی مجموعه زیادی از عمل‌ها را انجام می‌دهد که این امر موجب افزایش خطا و همچنین چترینگ در ورودی کنترلی پیوسته خواهد شد. علت بوجود آمدن پدیده چترینگ در اپیزود ۱ شبیه‌سازی، تعداد سوئیچینگ بالا در کسری از زمان می‌باشد و به این علت که زمان اقامت در این حالت بسیار پایین می‌باشد، کنترل‌کننده AFSMC فرصت لازم برای رسیدن به پاسخ حالت ماندگار مناسب را نیافته و قبل از ناپدید شدن اثرات گذرا، سیستم مجدد سوئیچ می‌کند.

همچنین، تصادفی بودن سیاست در اپیزود ۱ منجر به این می‌شود که کنترل‌کننده پیوسته، تلاش کنترلی بیشتری (افزایش زمان رسیدن به حالت مطلوب) برای نیل به هدف داشته باشد در صورتی که با اتخاذ سیاست بهینه با توجه به هدف کنترلی تعریف شده، سیاست بهینه منجر به کاهش ورودی کنترلی پیوسته نیز خواهد شد.

اما بعد از گذشت ۱۰۰۰ اپیزود، عامل بهینه‌ترین سیاست را اتخاذ کرده و بعد از آن سیاست قطعی می‌شود. البته باید در نظر داشت که با توجه به توانایی یادگیری آنلاین این الگوریتم، چنانچه تغییری در محیط مشاهده شود، عامل سعی در یادگیری مجدد خواهد نمود و این امر قابلیت مهم این الگوریتم می‌باشد.

از نمودارهای بدست آمده در بالا می‌توان مشاهده کرد که ردیابی مسیر به درستی انجام شده و خطا به کمتر از ۱ سانتی‌متر (با توجه به پارامتر $bound_1$) رسیده است (شکل ۶ و ۹). از طرفی عامل با اتخاذ سیاستی مناسب منجر به کاهش تعداد سوئیچ در سیستم (شکل ۸) و همچنین کاهش تلاش (هزینه) کنترلی ورودی پیوسته (دبی) شده است (شکل ۷).

۶- بحث و نتیجه‌گیری

در این مقاله، الگوریتمی برای کنترل سیستم‌های سوئیچینگ طراحی شد که قادر به کنترل همزمان ورودی‌های گسسته و پیوسته است و با داشتن حداقل اطلاعات از مدل، به‌صورت تطبیقی و مقاوم در برابر نویز و تغییر پارامتر عمل می‌کند. ایده اصلی، بر پایه ترکیب کنترل‌کننده تطبیقی-فازی-مد لغزشی با نوع خاصی از یادگیری تقویتی، یعنی یادگیری Q ، شکل گرفته است. برای این ترکیب، تغییراتی در الگوریتم یادگیری Q اعمال شد و کنترل‌کننده AFSMC، که وظیفه تولید ورودی کنترلی پیوسته را بر عهده دارد، در بخش محیط این الگوریتم جایگذاری شده است. این تغییر منجر به طراحی الگوریتمی شد که قابلیت کنترل سیستم‌های سوئیچینگ با ورودی‌های همزمان گسسته و پیوسته را دارد. به منظور ارزیابی کارایی الگوریتم پیشنهادی، سیستم دو مخزن، که به عنوان یک مسئله معیار در بسیاری از تحقیقات برای بررسی الگوریتم‌های کنترل به‌کار می‌رود، مورد استفاده قرار گرفت. شبیه‌سازی‌های انجام‌شده بر روی این سیستم و نتایج آن در بخش ۴ ارائه شده‌اند.

منابع

- 13- Markazi AH, Khodadadi M. Control of switched systems with delayed states: A direct AFSMC approach. *IET Control Theory & Applications*. 2023;17(1):1-8.
- 14- Luan F, Wang J. Design of state-dependent switching with dwell time constraint for interval positive switched systems. *IEEE Access*. 2018;6:47836-42.
- 15- Yuan S. Control of Switched Linear Systems: Adaptation and Robustness. 2018.
- 16- Poursamad A, Markazi AH. Adaptive fuzzy sliding-mode control for multi-input multi-output chaotic systems. *Chaos, Solitons & Fractals*. 2009;42(5):3100-9.
- 17- Watkins, C.J.C.H., Dayan, P. *Q-learning*. *Mach Learn* 8, 279-292 (1992).
- 18- Sutton, R.S., *Reinforcement learning: An introduction*. A Bradford Book, 2018.
- 19- Saidi E, Hammi Y, Douik A. Explicit model-predictive control of hybrid dynamical systems: Application to a two-tank system. In 2015 IEEE 12th International Multi-Conference on Systems, Signals & Devices (SSD15) 2015 (pp. 1-6). IEEE.
- 20- Yu S, Lu X, Zhou Y, Feng Y, Qu T, Chen H. Liquid level tracking control of three-tank systems. *International Journal of Control, Automation and Systems*. 2020;18(10):2630-40.
- 1- Zhang Y, Chu X, Liu Y, Liu Y. A modelling and control approach for a type of mixed logical dynamical system using in chilled water system of refrigeration system. *Mathematical Problems in Engineering*. 2019;2019(1):7131380.
- 2- Forouzesh F, Eslami M, Jafari Shahbazzadeh M. Stability Analysis and Voltage Control in the Power System Based on the Hybrid Automata Model. *International Transactions on Electrical Energy Systems*. 2023;2023(1):5037957.
- 3- Zhao J, Wu X, Song Z, Sun L, Wang X. Practical hybrid model predictive control for electric pneumatic braking system with on-off solenoid valves. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*. 2024;238(10-11):3253-66.
- 4- Zhang M, Wang C, Zhang C, Li P, Liu L. A unified switching control framework for continuous robot-assisted training. *IEEE/ASME Transactions on Mechatronics*. 2023;29(4):2743-55.
- 5- Levis A. Challenges to control: A collective view-- Report of the workshop held at the University of Santa Clara on September 18-19, 1986. *IEEE Transactions on Automatic Control*. 1987;32(4):275-85.
- 6- He Y, Lin S, Sun J. Stabilization of the networked hybrid systems with communication constraints and its applications. *International Journal of Control, Automation and Systems*. 2023;21(2):464-74.
- 7- علما علیرضا، شاصادقی مختار، رضائی امین. طراحی کنترل کننده پیش بین مدل پایدار ساز برای سیستم های هایبرید مرکب منطقی دینامیکی: رویکرد تابع لیاپانف مبتنی بر نرم بی نهایت.
- 8- Wang Y, Chang Y, Alkhateeb AF, Alotaibi ND. Adaptive fuzzy output-feedback tracking control for switched nonstrict-feedback nonlinear systems with prescribed performance. *Circuits, Systems, and Signal Processing*. 2021;40:88-113.
- 9- Yaakoubi H, Haggège J. Modeling of three-tank hybrid system using Mixed Logical Dynamical formalism. In 2022 5th International Conference on Advanced Systems and Emergent Technologies (IC_ASET) 2022 (pp. 55-60). IEEE.
- 10- Baluchzadeh M, Karimpour A, Pariz N. Robust Optimal Control of Uncertain Nonlinear Switched System using Approximate Dynamic Programmin. *Modares Journal of Electrical Engineering*. 2016;16(4).
- 11- چهاردولی، ح.، پایداری مطلق و محدود زمانی سیستم های سوئیچینگ به فرم پسخور اکید با انتگرال گیرهای توان کسری، سی و یکمین همایش سالانه بین المللی مهندسی مکانیک ایران و نهمین همایش صنعت نیروگاهی ایران. ۱۴۰۲.
- 12- Neunert M, Abdolmaleki A, Wulfmeier M, Lampe T, Springenberg T, Hafner R, Romano F, Buchli J, Heess N, Riedmiller M. Continuous-discrete reinforcement learning for hybrid control in robotics. In Conference on Robot Learning 2020 (pp. 735-751). PMLR.