

شناسایی دینامیک غیرخطی بالگرد مبتنی بر داده‌های پروازی با استفاده از الگوریتم یادگیری Q

رضا غلامیان

دانشجوی دکتری، گروه مهندسی هوافضا، دانشکده فنی و مهندسی، دانشگاه اصفهان، اصفهان، ایران

حجت طائی

استادیار، گروه مهندسی هوافضا، دانشکده فنی و مهندسی، دانشگاه اصفهان، اصفهان، ایران

مریم ملک‌زاده

دانشیار، دانشکده مهندسی هوافضا، دانشگاه صنعتی شریف، تهران، ایران

چکیده

در این مقاله، شناسایی خارج از خط نرخ غلت بالگرد در حوزه زمان، با استفاده از داده‌های تست پروازی و الگوریتم یادگیری Q انجام شده است. هدف اصلی از این مقاله، توسعه یک روش هوشمند و خودآموز برای شناسایی حلقه باز رفتار دینامیکی بالگرد براساس ورودی تحریک فرکانس جاروب می‌باشد. در این راستا، الگوریتم یادگیری Q برای یادگیری و بهینه‌سازی مدل سیستم استفاده گردیده است. ابتدا داده‌های پروازی شامل ورودی‌های فرکانس جاروب به سیستم اعمال و خروجی‌های سیستم جمع‌آوری شده است، سپس با استفاده از الگوریتم مدل دینامیکی سیستم تخمین و مقایسه گردید. در نهایت اعتبارسنجی روش پیشنهادی با دو مجموعه داده تحت تحریک ورودی فرکانس جاروب و دابلت نیز صورت گرفت. هم‌چنین ارزیابی با روش‌های کلاسیک و مشابه یادگیری تقویتی از قبیل LS^* ، $MLP^\dagger$ ، $KF^\ddagger$ ، $SINDYc^\S$ و $SARSA^{**}$ نیز انجام شده است. نتایج حاصل از این رویکرد، دقت بالا و تطبیق پذیری خوب در شبیه‌سازی رفتار بالگرد را نشان می‌دهد. براساس اعتبارسنجی الگوریتم برای داده‌های تست پروازی در رویدادهای مختلف پروازی، این روش به‌طور مؤثری قادر به شناسایی تغییرات غیرخطی و پیچیدگی‌های مدل بالگرد می‌باشد که به‌عنوان یک ابزار مفید در طراحی و بهینه‌سازی کنترل‌کننده‌ها برای سیستم‌های پیچیده به‌کار می‌رود.

واژه‌های کلیدی: شناسایی سیستم، بالگرد، یادگیری Q، یادگیری تقویتی، داده‌های پروازی، ورودی فرکانس جاروب

Flight-Data-Driven Identification of Nonlinear Helicopter Dynamics Using Q-Learning

Reza Gholamian

Department of Aerospace Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran

Hojat Taei

Assistant Professor, Department of Aerospace Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran

Maryam Malekzadeh*

Associate professor, Department of Aerospace Engineering, University of Sharif, Tehran, Iran

Abstract

In this paper, offline time-domain identification of the helicopter roll-rate dynamics is performed using flight-test data and a Q-learning algorithm. The main objective of this study is to develop an intelligent and self-learning approach for the open-loop identification of helicopter dynamic behavior based on sweep-frequency excitation inputs. For this purpose, the Q-learning algorithm is employed to learn and optimize the system model. First, flight data obtained under sweep-frequency excitation, including the system inputs and corresponding outputs, are collected. The dynamic model of the system is then estimated using the proposed algorithm and its response is compared with the measured flight data. The proposed method is subsequently validated using two datasets obtained under sweep-frequency and doublet excitations. In addition, its performance is evaluated against classical and related reinforcement learning-based identification methods, including LS^* , $MLP^\dagger$, $KF^\ddagger$, $SINDYc^\S$, and $SARSA^{**}$. The results demonstrate high accuracy and good adaptability of the proposed approach in simulating helicopter behavior. Validation using flight-test data from different flight events shows that the proposed method can effectively identify nonlinear variations and complexities in the helicopter dynamics and can serve as a useful tool for the design and optimization of controllers for complex systems.

Keywords: System identification, helicopter, Q-learning, reinforcement learning, flight data, sweep-frequency input

* Least Square

† Multilayer perceptron

‡ Kalman Filter

§ sparse identification of nonlinear dynamics

** State Action Reward State Action

۱- مقدمه

شناسایی و مدل‌سازی سیستم‌های غیرخطی هواگردها، فرآیند پیچیده‌ای است که در آن رفتار دینامیکی وسیله پروازی تحت تأثیر ورودی‌های مختلف مورد تحلیل قرار می‌گیرد. در این راستا، روابط میان ورودی‌ها و خروجی‌ها به منظور حصول شبیه‌سازی دقیق از رفتار واقعی هواگرد در شرایط عملیاتی پیچیده به‌صورت غیرخطی مدل‌سازی می‌شود [۱].

در میان تمام حوزه‌های کاربردی اصلی شناسایی سیستم، وسایل پروازی به ویژه بالگردها و کوادروتورها اهمیت زیادی پیدا کرده‌اند [۲]. از نقطه نظر مدل‌سازی تحلیلی و همچنین شناسایی سیستم، بالگردها به دلیل اثرات جفت‌شدگی متقاطع^۱ آبرودینامیکی بالا و جفت‌شدگی روتور و همچنین کیفیت پایین اندازه‌گیری جریان هوا در سرعت پایین و شرایط پرواز ایستا^۲، سیستم‌های پیچیده‌تر و دشوارتری در نظر گرفته می‌شوند [۳].

یکی از مسائل مهم در فرایند شناسایی سیستم، انتخاب ورودی تحریک مناسب است. ورودی فرکانس جاروب^۳ بدلیل پوشش طیف وسیع فرکانسی، کاهش نویز نسبت به تحریک‌های تصادفی و ایمن بودن نسبت به ورودی‌هایی همچون ضربه یا پله، در این مقاله مورد استفاده قرار گرفته است. در ابتدا داده‌های حاصل از تست پروازی تحت تحریک فرکانس جاروب جمع‌آوری گردیده است [۴].

در فرایند شناسایی سیستم، انتخاب مدل مناسب بدلیل تأثیر قابل توجه در استفاده از دانش قبلی، توانایی بهبود نتیجه و تسهیل فرایندهای بعدی از جمله صحت‌سنجی و طراحی کنترل‌کننده، اهمیت دارد. یکی از روش‌های استخراج مدل‌های مناسب، رویکرد یادگیری تقویتی^۴ می‌باشد که تابع تبدیل حالت یک فرآیند تصمیم‌گیری مارکوف را با یادگیری یک تابع نگاشت از وضعیت فعلی و اقدام به حالت بعدی، به عنوان یک مسئله یادگیری نظارت‌شده به صورت مستقیم تقریب می‌زند [۵]. این رویکرد، بدلیل قابلیت حذف اثرات تاخیر، غیرخطی بودن بالا، غیر ایستا^۵ بودن و انباشتگی خطا در هنگام استفاده از پیش‌بینی‌های خود راه‌انداز^۶ در مقاله حاضر استفاده شده است [۶].

لی و شی در مرجع [۷] به مطالعه یک الگوریتم شناسایی سیستم مبتنی بر ترکیب حداقل مربعات وزن‌دار و الگوریتم بهینه‌سازی گرگ خاکستری بهبود یافته^۷ برای یافتن ساختار مدل دینامیکی بالگرد پرداختند. در مرجع [۸] ژانگ و هی فاز نشست یک بالگرد بدون سرنشین را در محیط‌های آشفته با ترکیب شناسایی پراکنده کنترل^۸ محور دینامیک غیرخطی با روش SINDYC^۹ و استراتژی کنترل تقلید تقویت عمیق به صورت خارج از خط ارزیابی کردند. لی و وانگ نیز در مرجع [۹] رویکرد مبتنی بر تابع پایه متعارف تعمیم‌یافته (GOBF)^{۱۰} یادگیری ماشین را برای طراحی کنترل‌کننده بالگرد بدون

سرنشین T-REX 800 بکار بردند. در مرجع [۱۰] مقایسه روش‌های یادگیری ماشین نظارت‌شده از جمله شبکه عصبی، بردار پشتیبان، رگرسیون طبقه‌بندی‌شده در کاربردهای هوایی مورد ارزیابی قرار گرفته است. آندرس و همکاران در مرجع [۱۱] با ترکیب یادگیری تقویتی و شناسایی سیستم به طراحی کنترل‌کننده برای بهینه‌سازی عملکرد و تطابق رفتار مدل بالگرد پرداختند. همچنین ایوبند و ارلند در مرجع [۱۲] کنترل‌کننده MPC را برای غلبه بر ترم‌های غیرخطی پکت پروازی بالگرد به جهت یادگیری آنلاین طراحی کردند. در مرجع [۱۳] شناسایی مدل USV با استفاده از یادگیری عمیق و روش LSTM مبتنی بر شبکه عصبی بازگشتی محاسبه و تخمین نرخ یاب و سرعت پرند انجام گردیده است. در مرجع [۱۴] منتر و تیشلر با هدف شناسایی مدل دینامیکی، آزمون‌های پروازی را بر روی بالگرد بدون سرنشین R-50 انجام دادند. مدل شناسایی‌شده در نهایت از دقت پیش‌بینی بالایی برخوردار بود و به‌عنوان پایه‌ای برای طراحی سامانه‌های کنترل پرواز و توسعه برنامه‌های شبیه‌سازی مورد استفاده قرار گرفت. خلیق و فهمی در مرجع [۱۵] رویکرد حوزه زمان شناسایی غیرخطی پارامترهای مدل را در طراحی کنترل‌کننده بالگرد بدون سرنشین در مقیاس کوچک تشریح نمودند. نتایج نشان می‌دهد که پاسخ کلی مدل غیرخطی شناسایی‌شده بالگرد با داده‌های پرواز واقعی مطابقت دارد. ولاداو و همکاران در مرجع [۱۶] با استفاده از روش شبکه عصبی بازگشتی المن^{۱۱} به شناسایی دینامیک بالگرد پرداختند. ملک زاده و همکاران در مرجع [۱۷] به طراحی کنترل پیش‌بین مدل خطی سازی فیدبک اصلاح شده مقید وسیله پروازی جهت تنظیم مانور در حضور گشتاورهای خارجی و اشباع عملکرد پرداختند.

براساس فعالیت‌های بررسی‌شده، این مقاله با تمرکز بر شناسایی خارج از خط سیستم، یک چارچوب جدید برای مدل‌سازی غیرخطی بالگرد مبتنی بر یادگیری تقویتی ارائه می‌کند. وجه تمایز روش پیشنهادی نسبت به رویکردهایی نظیر SINDYC^{۱۲}، GOBF^{۱۳} و شبکه عصبی بازگشتی المن، عدم وابستگی به مدل دینامیکی از پیش شناخته‌شده و عدم نیاز به استخراج مدل فضای حالت، کاهش مرتبه سیستم یا حذف اثرات محیطی است. از این رو، قابلیت روش‌های یادگیری تقویتی به‌عنوان یک رویکرد کاملاً داده‌محور در فرایند شناسایی سیستم و مدل‌سازی دینامیکی بالگرد مورد ارزیابی قرار می‌گیرد. نتایج حاصل می‌تواند در بهبود مدل‌سازی پروازی، طراحی کنترل‌کننده‌های هوشمند و ارتقای عملکرد سامانه‌های پروازی مؤثر باشد.

در روش پیشنهادی، داده‌های واقعی پروازی شامل ورودی‌های کنترلی و پاسخ‌های اندازه‌گیری‌شده سیستم به‌عنوان محیط یادگیری در نظر گرفته می‌شوند. الگوریتم یادگیری Q با مشاهده وضعیت لحظه‌ای سیستم، شامل نرخ غلت و تغییرات آن، مقدار تخمینی خروجی را تولید کرده و بر اساس میزان تطابق آن با داده واقعی، پاداش متناظر را دریافت می‌کند. تابع پاداش به‌گونه‌ای تعریف شده است که کاهش خطای تخمین منجر به افزایش پاداش شود؛ بنابراین، مدل از طریق یادگیری مبتنی بر تجربه و به‌روزرسانی تکراری مقادیر Q به‌صورت تدریجی بهبود می‌یابد. این فرایند بدون نیاز به مدل‌سازی دقیق ریاضی یا فرضیات خطی‌سازی، امکان استخراج رفتار دینامیکی سیستم را فراهم می‌کند.

¹ Cross coupling

² Hover

³ Frequency Sweep

⁴ Reinforcement learning

⁵ unstationary

⁶ Bootstrap prediction

⁷ Improved Grey Wolf Optimizer

⁸ control-oriented sparse

⁹ Sparse Identification of Nonlinear Dynamics with Control

¹⁰ generalized orthonormal basis function

¹¹ Elman Recurrent Neural Network

¹² sparse identification of nonlinear dynamics

¹³ Generalized Orthonormal Basis Function

شکل ۱- نمایی از بالگرد BO-105

جدول ۱- تعاریف و مقدار پارامترهای بالگرد BO-105 [۱۸]

متغیر	مقدار	واحد	توضیحات
Mass	۲۲۰۰	kg	جرم ب
I_x	۱۴۳۳	kg.m ²	ممان اینرسی حول محور طولی
I_y	۴۹۷۳	kg.m ²	ممان اینرسی حول محور عرضی
I_z	۴۰۹۹	kg.m ²	ممان اینرسی حول محور عمودی
I_{xz}	۶۶۰	kg.m ²	ممان اینرسی ضربی
$R_{main rotor}$	۴/۹۱	m	شعاع روتور اصلی
$N_{main rotor}$	۴	-	تعداد ملخ اصلی روتور اصلی
$chord_{main rotor}$	۰/۲۷	m	عرض پره روتور اصلی
$R_{tail rotor}$	۰/۹۵	m	شعاع روتور دم
$N_{tail rotor}$	۲	-	تعداد ملخ روتور دم
$chord_{tail rotor}$	۰/۱۷	m	عرض پره روتور دم

معادلات غیرخطی حرکت بالگرد برای متغیرهای سرعت خطی u, v, w و سرعت زاویه‌ای p, q, r در راستای محور X, Y, Z به صورت رابطه (۱) تعریف می‌شود [۲۰].

$$\begin{aligned} m\dot{u} &= -m(wq - vr) + X - mg\sin\theta \\ m\dot{v} &= -m(ur - wp) + Y + mg\cos\theta\sin\theta \\ m\dot{w} &= -m(vp - uq) + Z + mg\cos\theta\cos\theta \\ I_{xx}\dot{p} &= (I_{yy} - I_{zz})qr + I_{xz}(r + pq) + L \\ I_{yy}\dot{q} &= (I_{zz} - I_{xx})rp + I_{xz}(r^2 - p^2) + M \\ I_{zz}\dot{r} &= (I_{xx} - I_{yy})pq + I_{xz}(\dot{p} - qr) + N \end{aligned} \quad (1)$$

در رابطه (۱)، پارامترهای L, M, N به ترتیب ممان‌های آیرودینامیکی غلت، پیچ و یاو و $I_{xx}, I_{yy}, I_{zz}, I_{xz}$ ممان‌های اینرسی اصلی و I_{xx} ممان اینرسی ضربی بالگرد می‌باشد. معادلات سینماتیک چرخشی، که نرخ زوایای اوایل را به نرخ‌های زاویه‌ای دستگاه مختصات بدنی مرتبط می‌سازد در رابطه (۲) تعیین می‌گردد. رابطه (۲)، معادلات حالت غیرخطی برای زوایای اوایل غلت، پیچ، یاو بیان شده است [۲۰].

$$\begin{aligned} \dot{\phi} &= p + q \sin\theta \tan\theta + r \cos\theta \tan\theta \\ \dot{\theta} &= q \cos\theta - r \sin\theta \\ \dot{\psi} &= q \sin\phi \sec\theta + r \cos\phi \sec\theta \end{aligned} \quad (2)$$

داده‌های تاریخچه زمانی بالگرد BO-105 کانال عرضی پرواز ایستا، تحت تحریک فرکانس جاروب، برای انجام فرایند شناسایی و تخمین نرخ غلت انتخاب گردید. فرایند جمع‌آوری داده‌های شناسایی در بازه زمانی ۹۰ ثانیه و با گام زمانی ۰/۰۱ ثانیه انجام شده است [۴]. مجموع داده‌های جمع‌آوری شده شامل ۹۰۰۰ نمونه می‌باشد. در شکل (۲)، نمودار مانور ورودی تحریک سکان عرضی بالگرد و در شکل (۳) تغییرات نرخ غلت داده‌های تست پروازی آورده شده است.

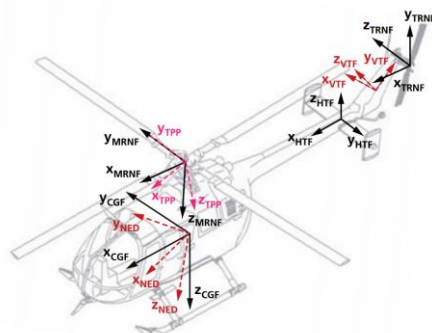
برخلاف کاربردهای متداول یادگیری تقویتی که مبتنی بر تعامل برخط با محیط واقعی یا شبیه‌سازی شده هستند، در این پژوهش الگوریتم یادگیری Q در قالب یک چارچوب شناسایی خارج از خط به کار گرفته شده است. در این ساختار، انتقال حالت‌ها و کنش‌ها از داده‌های ثبت‌شده ورودی-خروجی استخراج شده و فرآیند یادگیری به‌طور کامل بر روی داده‌های تاریخی پرواز انجام می‌شود. در نتیجه، ضمن حفظ ویژگی تطبیقی یادگیری تقویتی، نیاز به اکتشاف و تعامل برخط حذف شده و الگوریتم به ابزاری مناسب برای شناسایی داده‌محور دینامیک غیرخطی بالگرد تبدیل می‌شود.

در مقاله حاضر، ابتدا فرایند و روش تست پروازی و بررسی مشخصات بالگرد ارائه خواهد شد. سپس، به بررسی رویکرد مستقل از مدل الگوریتم یادگیری تقویتی، لزوم استفاده از آن، فرمول‌بندی و تنظیمات مربوطه پرداخته می‌شود. در ادامه نتایج حاصل از شناسایی سیستم ارزیابی و اعتبارسنجی الگوریتم انجام خواهد شد. در پایان نیز مدل تابع تبدیل حرکتی این بالگرد مبتنی بر داده‌های شناسایی شده جهت انجام فرایند صحنه‌گذاری و مقایسه روش‌های سنتی و تعامل محیطی تحت تحریک ورودی یکسان قرار خواهد گرفت.

۲- مدل‌سازی دینامیکی

۲-۱- معرفی و توصیف بالگرد BO-105

بالگرد BO-105 یک بالگرد دوموتوره سبک و چندمنظوره روتور اصلی چهارپره و روتور دم دومحوره است که توسط شرکت آلمانی (MBB)^۱ طراحی و ساخته شده است. این بالگرد به دلیل مانورپذیری بالا در مطالعات شناسایی سیستم و کنترل پرواز به‌طور گسترده مورد استفاده قرار می‌گیرد. بالگرد BO-105 مجهز به سیستم کنترل همپلتون و سیستم انتقال قدرت دو موتوره است که به آن امکان پرواز در شرایط مختلف و مقاوم در برابر خرابی را می‌دهد [۱۸]. سیستم کنترلی این بالگرد به گونه‌ای طراحی شده که خلبان از طریق سه سطح تحریک (اهرم^۲، پدال^۳ و سکان) قادر به تغییر مسیر، سرعت و ارتفاع باشد. اهرم برای کنترل ارتفاع بالگرد و سکان^۴ برای هدایت حرکت افقی افقی (جهت جلو، عقب و چرخش) استفاده می‌شود. پدال‌ها نیز به خلبان اجازه می‌دهند که جهت چرخش افقی^۵ بالگرد را کنترل کنند [۱۹]. در جدول (۱) مشخصات جرمی و ابعادی این بالگرد آورده شده است.



¹ Messerschmitt-Bölkow-Blohm

² Collective

³ Pedal

⁴ Cyclic

⁵ yaw

۳-۲- پیاده‌سازی الگوریتم یادگیری Q

الگوریتم یادگیری Q از یک جدول به نام "Q-table" استفاده می‌کند که در آن ارزش هر جفت حالت-کنش ذخیره می‌شود. این مقادیر Q نشان‌دهنده سودمندی مورد انتظار از انجام یک کنش معین در یک وضعیت معین و پیروی از سیاست بهینه پس از آن است. مقادیر جدول Q به طور دلخواه مقداردهی اولیه شده و با استفاده از تجربیات جمع‌آوری‌شده توسط عامل به طور مکرر بروزرسانی می‌گردند (رابطه (۳)).

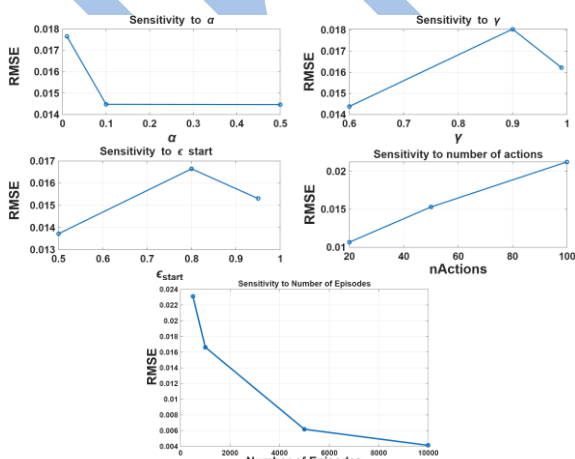
$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \times \max_{a'} Q(s', a') - Q(s, a)] \quad (3)$$

در رابطه (۳)، S حالت کنونی، a عمل انجام‌شده، r پاداش پس از عمل انجام‌شده، s' حالت جدید پس از کنش، a' کنش احتمالی حالت جدید است. نرخ یادگیری $0 \leq \alpha \leq 1$ و ضریب کاهش $0 \leq \gamma < 1$ می‌باشد. مقداردهی پارامترهای توصیف‌شده در رابطه (۳) به صورت جدول ۲ انجام شده است [۲۲].

جدول ۲- پارامترهای الگوریتم یادگیری Q

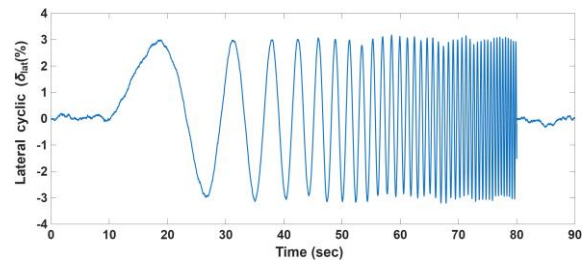
مقدار	پارامتر
۰/۹	γ
۰/۱	α
۰/۸	ϵ
۵۰۰۰	تکرار

پارامترها می‌توانند از نظر همگرایی، دقت یا پیچیدگی محاسباتی و موارد دیگر بر نتایج تأثیر بگذارند. به منظور بررسی بیشترین تأثیر پارامترهای الگوریتم بر روی عملکرد نهایی آن (مانند خطای RMSE، سرعت همگرایی یا پایداری)، تحلیل حساسیت انجام شده است. نتایج حاصل، در شکل (۵) نشان داده شده است.

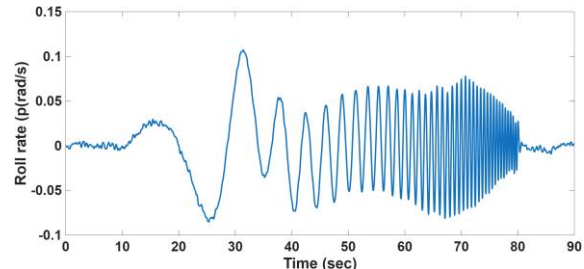


شکل ۵- نمودار حساسیت‌سنجی پارامترهای الگوریتم یادگیری Q

پارامتر α میزان وزن‌دهی الگوریتم به اطلاعات جدید را تعیین می‌کند؛ به‌گونه‌ای که مقادیرهای بزرگ‌تر باعث یادگیری سریع‌تر اما ناپایدارتر، و مقادیرهای کوچک‌تر موجب یادگیری آهسته اما پایدار



شکل ۲- ورودی تحریک



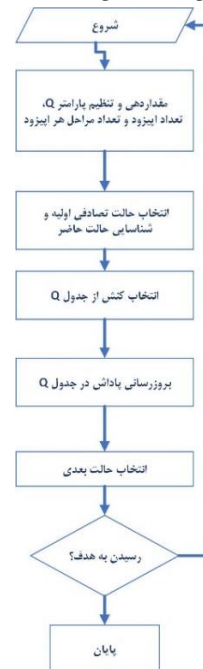
شکل ۳- تغییرات نرخ غلت داده‌های پروازی بالگرد BO-105

در ادامه به معرفی الگوریتم مورد استفاده جهت تخمین نرخ غلت بالگرد به صورت خارج از خط و براساس داده‌های تست پروازی پرداخته خواهد شد.

۳- فرمول‌بندی مسئله و پیاده‌سازی الگوریتم

۳-۱- مقدمه

یادگیری تقویتی یکی از روش‌های یادگیری ماشین مبتنی بر الگوی آزمون و خطا است که با استفاده از سیستم پاداش و مجازات، از بازخورد کنش‌ها برای بهینه‌سازی تصمیم‌گیری بهره می‌گیرد. در این میان، الگوریتم یادگیری Q بدون نیاز به مدل محیط، از طریق تعاملات متوالی، سیاست کنش بهینه را برای بیشینه‌سازی پاداش تجمعی در فرآیندهای تصمیم‌گیری مارکوف محدود تعیین می‌کند. روندنمای اجرای این الگوریتم در شکل (۴) نمایش داده شده است.



شکل ۴- روندنمای الگوریتم یادگیری تقویتی [۲۱]

$$predicted_p = \min(p_{norm}) + \frac{\max(p_{norm}) - \min(p_{norm})}{nActions} \times action \quad (4)$$

۴- پاداش بر اساس منفی خطای پیش‌بینی و مطابق با رابطه (۵) محاسبه می‌شود. این پاداش بیانگر میزان اختلاف بین مقدار پیش‌بینی‌شده و مقدار واقعی نرخ غلت است.

$$reward = -|p_{norm}(state) - predicted_p| \quad (5)$$

۵- انتقال به حالت بعدی: انتقال به حالت بعدی به سهولت انجام می‌شود و چنانچه داده‌ها به انتها رسیده باشند، به حالت اولیه برگشت داده می‌شود.

۶- بروزرسانی جدول Q: بروزرسانی جفت کنش-حالت با استفاده از فرمول بروزرسانی یادگیری تفاوت زمانی^۲ انجام خواهد شد (رابطه (۶)).

$$V(S_t) = V(S_t) + \alpha [V(S_{t+1}) - V(S_t)] \quad (6)$$

در این روش، برای بروزرسانی ارزش هر حالت در صورت انجام حرکت حریمانه، ارزش قبلی توسط یک ضریب مثبت به ارزش حالت بعدی نزدیک می‌شود. این بروزرسانی به الگوریتم این امکان را می‌دهد که سیاست‌های بهینه را یاد بگیرد.

۷- کاهش تدریجی ϵ : مقدار ϵ به آرامی کاهش می‌یابد تا به تدریج الگوریتم از اکتشاف به استفاده از سیاست بهینه منتقل شود.

۸- در هر تکرار، پاداش تجمعی آن تکرار ذخیره می‌شود. پس از پایان فرآیند آموزش، نرخ غلت پیش‌بینی‌شده بالگرد با استفاده از سیاست بهینه محاسبه می‌گردد، به‌طوری‌که این سیاست براساس انتخاب کنش با بیشترین مقدار Q در هر حالت تعیین می‌شود و سپس با داده‌های واقعی مقایسه می‌شود.

۹- ارزیابی: در نهایت، معیارهای ارزیابی مدل شامل ریشه میانگین مربع خطا^۳، میانگین خطای مطلق^۴، ضریب تعیین^۵ و درصد تطابق^۶ محاسبه می‌شود.

۴- نتایج

در طراحی ورودی تحریک برای سیستم‌های دینامیکی پیچیده مانند بالگردها، انتخاب نوع ورودی تحریک از اهمیت ویژه‌ای برخوردار است تا بتوان رفتار سیستم را به درستی تخمین زد. یکی از ورودی‌های تحریک متداول در شبیه‌سازی‌ها و تست‌های تجربی، ورودی فرکانس جاروب است که به طور خاص برای شبیه‌سازی پاسخ‌های فرکانسی بالگرد استفاده می‌شود. ورودی فرکانس جاروب به طور تدریجی فرکانس خود را در یک بازه مشخص تغییر می‌دهد و این امکان را

می‌شوند. ضریب کاهش γ اهمیت پاداش‌های آینده را مشخص می‌کند؛ و γ کوچک الگوریتم کوتاه نگر و بزرگ نشاندهنده بلندنظر اما حساس به نویز است. پارامتر ϵ میزان کاوش را مشخص می‌کند، موجب جستجوی گسترده‌تر و خطای بیشتر در مراحل اولیه می‌شود و مقدار بسیار کوچک آن احتمال گیر افتادن در سیاست‌های محلی را افزایش می‌دهد. بهترین عملکرد معمولاً با کاهش تدریجی ϵ در طول تکرارها حاصل می‌شود تا تعادل مناسبی بین کاوش و بهره‌برداری برقرار گردد. علاوه بر این، تعداد عامل‌ها یا تفکیک‌بندی حالت-کنش، تعیین‌کننده‌ی میزان دقت و وضوح تصمیم‌گیری در فضای یادگیری است. همچنین، افزایش تعداد تکرارهای آموزشی تأثیر مستقیمی بر بهبود عملکرد الگوریتم دارد، به‌طوری‌که مقدار RMSE به‌صورت یکنواخت کاهش یافته و پس از حدود ۵۰۰۰ تکرار، نرخ بهبود کندتر شده و الگوریتم به ناحیه اشباع همگرایی وارد می‌شود. این رفتار نشان می‌دهد که افزایش تعداد تکرارها در مراحل اولیه نقش کلیدی در کاهش خطا دارد، اما پس از رسیدن به یک سطح مشخص، بازده یادگیری کاهش یافته و بهبودهای جزئی‌تر حاصل می‌شود. علاوه بر این، تعداد حالت‌ها و تفکیک‌بندی فضای حالت-کنش نیز به‌عنوان یک عامل کلیدی در دقت تصمیم‌گیری نقش دارد؛ به‌طوری‌که افزایش بیش از حد آن می‌تواند منجر به افزایش پیچیدگی محاسباتی و کندی همگرایی شود، در حالی‌که انتخاب خیلی محدود آن دقت مدل را کاهش می‌دهد. بر همین اساس، انتخاب مقادیر پارامترهای جدول Q با توجه به این ملاحظات و بر مبنای تحلیل حساسیت انجام شده است. یکی از چالش‌های یادگیری تقویتی برقراری مصالحه بین اکتشاف و بهره‌برداری می‌باشد که این الگوریتم مبادله بین اکتشاف (انتخاب کنش‌های تصادفی برای کشف استراتژی‌های جدید) و بهره‌برداری (انتخاب کنش‌ها بر اساس دانش انباشته‌شده) را مدیریت می‌کند. تکنیک‌هایی مانند استراتژی تقریباً حریمانه^۱، که در آن عامل عمدتاً شناخته‌شده‌ترین کنش را انجام می‌دهد، اما گاهی اوقات یک اقدام تصادفی را امتحان می‌کند، اغلب تعادل بین این موارد را مدیریت می‌کنند [۲۲]. در این مسئله، تعداد حالات برابر با تعداد داده‌های نرمالایز شده تغییرات سکان عرضی بوده و ۵۰ کنش پیش‌بینی شده است.

حلقه آموزش فرآیند یادگیری در هر تکرار به شرح ذیل انتخاب می‌شود.

۱- انتخاب حالت اولیه: در هر تکرار، یک حالت اولیه به صورت تصادفی انتخاب می‌شود.

۲- انتخاب کنش با روش تقریباً حریمانه.

با روش تقریباً حریمانه اگر مقدار تصادفی کمتر از ϵ باشد، یک اقدام تصادفی (اکتشاف) انتخاب می‌شود. در غیر این صورت، اقدام با بیشترین Q-value برای حالت فعلی انتخاب می‌شود.

۳- پیش‌بینی نرخ غلت برای هر کنش بر اساس رابطه (۴) محاسبه می‌شود، به‌طوری‌که در این فرآیند، کنش‌ها به یک مقیاس نرمال از نرخ چرخش تبدیل می‌گردند.

² Temporal Difference

³ Root Mean Squared Error

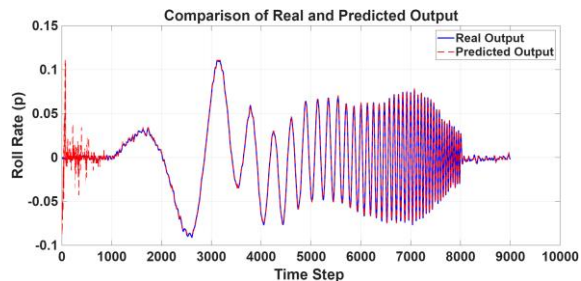
⁴ Mean absolute error

⁵ coefficient of determination

⁶ Fit percentage

¹ Epsilon-Greedy

علاوه بر این، به منظور ارزیابی بصری از کیفیت تطابق بین نرخ غلت واقعی و مقدار تخمین زده شده توسط الگوریتم، نمودار مقایسه‌ای در شکل (۷) ترسیم شده است. این نمودار نشان‌دهنده هم‌راستایی مناسب بین پاسخ‌های سیستم واقعی و خروجی الگوریتم بوده و بیانگر دقت قابل قبول آن در تخمین نرخ غلت در شرایط فاز ایستا می‌باشد.

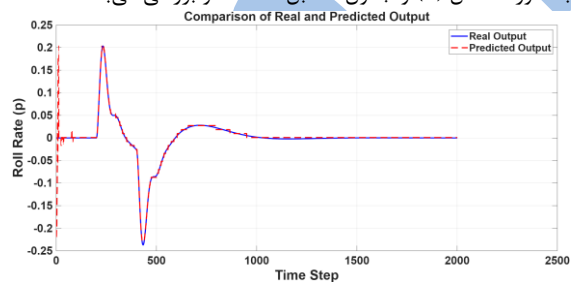


شکل ۷- اعتبارسنجی الگوریتم با استفاده از تاریخچه داده رویداد پروازی دوم (تحت فرکانس جاروب)

جدول ۴- نتایج داده‌های اعتبارسنجی رویداد پروازی دوم (تحت فرکانس جاروب)

پارامتر	مقدار
RMSE	۰/۰۰۴۹
MAE	۰/۰۰۱۵
R-squared	۰/۹۸۵۸
Fit Percentage	۹۸/۶۰
Reward	-۰/۰۰۲۵

حال جهت بررسی نقش ورودی تحریک در نتیجه شناسایی و اهمیت غنای محتوای طیفی ورودی، تاریخچه داده تست پروازی تحت ورودی دابلت مورد بررسی قرار گرفته است [۲۴]؛ نتایج شناسایی مدل به صورت شکل (۸) و جدول ۵ قابل مشاهده و بررسی می‌باشد.

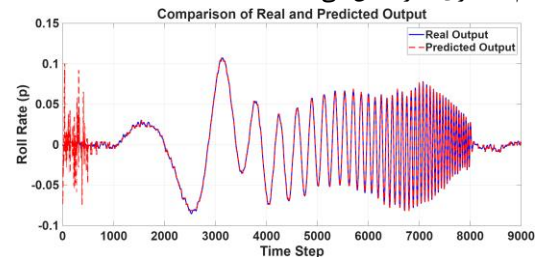


شکل ۸- اعتبارسنجی الگوریتم با استفاده از تاریخچه داده رویداد پروازی سوم (تحت ورودی دابلت)

جدول ۵- نتایج داده‌های اعتبارسنجی رویداد پروازی سوم (ورودی دابلت)

پارامتر	مقدار
RMSE	۰/۰۰۹۴
MAE	۰/۰۰۲۳
R-squared	۰/۹۵۴۸
Fit Percentage	۹۵/۶۸
Reward	-۰/۰۰۲۵

فراهم می‌آورد که سیستم دینامیکی به پاسخ‌های فرکانسی مختلف واکنش نشان دهد. در بالگردها، این ورودی می‌تواند به تحلیل و بررسی پاسخ سیستم به تغییرات فرکانسی مختلف در سطوح تحریک مختلف، از جمله حرکت‌های عمودی^۱، افقی^۲ و چرخش^۳ کمک کند [۲۳]. شکل (۶) با توجه به مطلوب بودن SNR تاریخچه زمانی، نمودار تناسب و تخمین نرخ غلت حلقه باز بالگرد تحت تحریک فرکانس جاروب توسط الگوریتم یادگیری Q را نشان می‌دهد.



شکل ۶- خروجی نرخ غلت داده پروازی و الگوریتم یادگیری Q رویداد پروازی اول (فرکانس جاروب)

همانطور که در شکل (۵) نشان داده شده است، الگوریتم یادگیری تقویتی توانسته است در ۹۰۰ گام زمانی، تطبیق ۹۸ درصدی را با خروجی نرخ غلت واقعی حاصل از تست پروازی داشته باشد که به معنای دریافت پاداش‌های مطلوب از سیستم است. مشخصات آماری توضیح داده شده، که مربوط به اجرای الگوریتم یادگیری Q می‌باشد، در جدول ۳ آورده شده است.

جدول ۳- نتایج تخمین نرخ غلت رویداد پروازی اول الگوریتم یادگیری Q

پارامتر	مقدار
RMSE	۰/۰۰۴۸
MAE	۰/۰۰۱۴
R-squared	۰/۹۸۵۶
Fit Percentage	۹۸/۵۸
Reward	-۰/۰۰۲۵

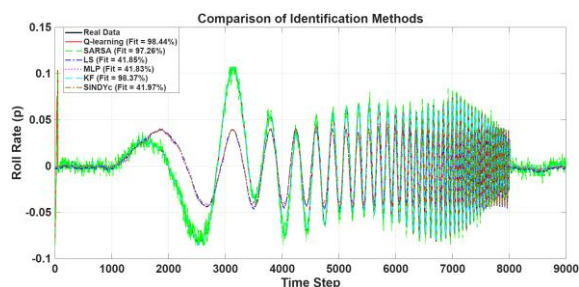
۴-۱- اعتبارسنجی مدل یادگیری Q

جهت اعتبارسنجی الگوریتم یادگیری Q، داده‌های فاز ایستا غلت رویداد دوم تست پروازی بالگرد BO-105 استفاده شده است [۲۴]. این داده‌ها به عنوان یک مرجع واقعی و معتبر جهت بررسی دقت تخمین‌های حاصل از الگوریتم به کار گرفته شده‌اند. در راستای تحلیل عملکرد الگوریتم، ابتدا داده‌های تجربی به صورت پیش‌پردازش شده و سپس برای آموزش و تست الگوریتم مورد استفاده قرار گرفتند. مقایسه خروجی‌های حاصل از الگوریتم با داده‌های واقعی، معیارهای آماری مناسبی نظیر میانگین خطای مطلق، ریشه میانگین مربعات خطا و ضریب تعیین به منظور سنجش میزان دقت و کارایی مدل به کار گرفته شده است. نتایج حاصل از این ارزیابی‌ها در جدول ۴ ارائه شده‌اند.

¹ Collective

² Cyclic

³ Pedals



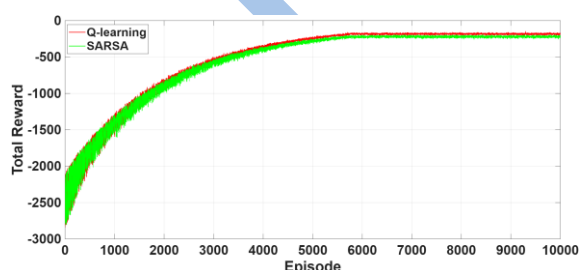
شکل ۹- اعتبارسنجی الگوریتم با روش‌های مشابه و کلاسیک براساس داده پروازی رویداد اول (فرکانس جاروب)

در جدول ۶ بررسی شاخص‌های آماری روش‌های بیان شده، ارائه می‌گردد.

جدول ۶- نتایج داده‌های اعتبارسنجی الگوریتم با روش‌های مشابه و کلاسیک

روش	RMSE	MAE	R-squared	درصد انطباق
یادگیری Q	۰/۰۰۴۸	۰/۰۰۱۳	۰/۹۸۴۴	۹۸/۴۴
SARSA	۰/۰۰۴۶	۰/۰۰۲۲	۰/۹۷۲	۹۷/۲۶
LS	۰/۰۳۰	۰/۰۲۱	۰/۴۱۸	۴۱/۸۵
MLP	۰/۰۳۲	۰/۰۲۲	۰/۴۱۸	۴۱/۸۳
KF	۰/۰۰۴۷	۰/۰۰۲۷	۰/۹۸۳	۹۸/۳
SINDYc	۰/۰۰۳۱	۰/۰۰۲۱	۰/۴۱۹	۴۱/۹۷

به‌منظور بررسی پایداری و همگرایی الگوریتم پیشنهادی، تغییرات پاداش تجمعی در طول تکرارهای یادگیری برای الگوریتم‌های یادگیری Q و SARSA در شکل (۱۰) ارائه شده است. همان‌طور که مشاهده می‌شود، پاداش تجمعی یادگیری Q پس از چند صد تکرار روند افزایشی یکنواختی را طی کرده و به یک مقدار ثابت نزدیک می‌شود که نشان‌دهنده همگرایی سیاست یادگیری است. همچنین نوسانات پاداش در مراحل پایانی بسیار کم بوده که بیانگر پایداری رفتار یادگیری در مواجهه با نویز داده‌های پروازی است.



شکل ۱۰- تغییرات پاداش تجمعی در طول تکرارهای یادگیری روش یادگیری Q و SARSA

در مقابل، الگوریتم SARSA با وجود همگرایی نسبی، نوسانات بیشتری در تکرارهای پایانی دارد که ناشی از ماهیت درون‌سیاستی و حساسیت بالاتر آن به خطاهای اندازه‌گیری است. بنابراین، یادگیری Q توانسته

بنابراین، با توجه به نتایج ارائه‌شده، الگوریتم یادگیری Q از دقت و اعتبار بالایی در شناسایی و تخمین مدل دینامیکی به‌صورت خارج از خط در حوزه زمان برخوردار است. الگوریتم پیاده‌سازی شده توانایی قابل‌توجهی در یادگیری سیاست مسئله و تخمین دقیق آن از خود نشان داده است.

۴-۲- ارزیابی الگوریتم با سایر روش‌ها

به‌منظور ارزیابی جامع عملکرد روش پیشنهادی، نتایج حاصل از الگوریتم یادگیری Q با شش روش مرجع شامل مدل حداقل مربعات، الگوریتم^۱ SARSA، مدل شبکه عصبی چندلایه، فیلتر کالمن و روش شناسایی دینامیکی مبتنی بر کتابخانه توابع پراکنده مقایسه شده است. شکل (۹) روند تغییرات نرخ غلت واقعی و مدل‌های تخمینی را تحت تحریک گسترده فرکانس جاروب نشان می‌دهد. مدل خطی LS صرفاً قادر به بازتولید بخش‌های خطی و کم‌دامنه پاسخ بوده و با افزایش غیرخطی بودن دینامیک بالگرد، انحراف قابل‌توجهی از داده‌های واقعی نشان می‌دهد. به‌طور مشابه، روش SINDYc علی‌رغم اتکا بر بازنمایی فشرده دینامیک‌های غیرخطی، در حضور نویز اندازه‌گیری و تحریک‌های گسترده، دقت محدودی داشته و در برخی بازه‌های زمانی دچار خطای دامنه و عدم تطابق فازی می‌شود. شبکه عصبی چندلایه (MLP) با وجود توانایی بالا در تقریب توابع غیرخطی، به دلیل وابستگی به داده‌های آموزشی و عدم استفاده از ساختار کنترلی مبتنی بر پاداش، در گذارهای فرکانسی و نواحی با تغییرات سریع دینامیکی، دچار خطای فاز و افت دقت می‌شود. در مقابل، فیلتر کالمن (KF) با بهره‌گیری از مدل دینامیکی سیستم و به‌روزرسانی بازگشتی تخمین حالت، عملکرد پایدارتری نسبت به روش‌های داده‌محور از خود نشان داده و در حضور نویز اندازه‌گیری، همگرایی مناسبی دارد. با این حال، وابستگی آن به مدل اولیه و خطی‌سازی محلی می‌تواند در شرایط غیرخطی شدید محدودیت ایجاد کند. در میان روش‌های مقایسه‌ای، الگوریتم SARSA با توجه به ماهیت درون‌سیاستی^۲ از همگرایی قابل قبولی برخوردار است، اما نسبت به نویز و اغتشاشات مسیر یادگیری حساس بوده و در برخی نواحی دامنه بالا، بخشی از نوسانات سیستم را به‌صورت تضعیف‌شده بازتولید می‌کند. در نهایت، الگوریتم یادگیری Q با استفاده از ساختار برون‌سیاستی^۳ و تعامل مستقیم با محیط، توانسته است رفتار سیستم را با دقت بسیار بالاتری در کل بازه فرکانسی بازتولید کند. نتایج کمی نیز نشان‌دهنده دقت بالای تطبیق مدل پیشنهادی با پاسخ واقعی و برتری آن نسبت به سایر روش‌های کلاسیک و داده‌محور است.

¹ State action reward state action

² On-policy

³ Off-policy

است با پایداری بیشتر و سرعت همگرایی بالاتر، سیاست کنترلی مناسب را برای تخمین نرخ غلت استخراج نماید.

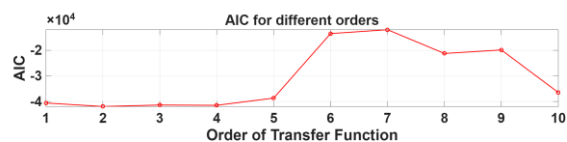
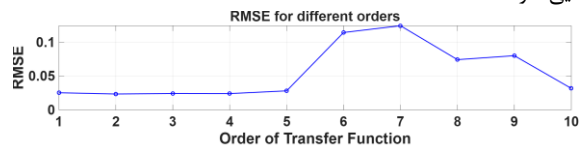
۳-۴- شناسایی پارامتری مدل کانال عرضی

دو رویکرد رایج برای شناسایی سیستم‌های پارامتری شامل استفاده از تابع تبدیل و مدل‌سازی در فضای حالت می‌باشد. در این مقاله، رویکرد مدل‌سازی مبتنی بر تابع تبدیل انتخاب شده است، چراکه این روش اطلاعات ارزشمندی در خصوص ویژگی‌های دینامیکی بنیادی سیستم فراهم می‌آورد و در بسیاری از کاربردهای کلیدی، از جمله تحلیل کیفیت کنترل و مدل‌سازی دینامیک پرواز به‌منظور طراحی سیستم‌های کنترلی، نقش مؤثری ایفا می‌کند. افزون بر این، مقایسه کیفیت تطبیق مدل‌های تابع تبدیل با سطوح مختلف پیچیدگی می‌تواند بینش‌هایی درباره مرتبه سیستم، میزان جفت‌شدگی میان ورودی‌ها و خروجی‌ها، و مقادیر اولیه معقول برای برخی از پارامترهای مدل ارائه دهد [۲۵].

تخمین تابع تبدیل سیستم با استفاده از داده‌های ورودی-خروجی نرمالایز شده و انتخاب بهترین مدل براساس معیارهای ارزیابی مختلف مانند ریشه میانگین مربعات خطا و AIC (معیار اطلاعات آکایکه) که در رابطه (۸) بیان شده، برای انتخاب مدل بهینه با توجه به تعداد پارامترها و کیفیت مدل انجام شده است. در این بخش، الگوریتم به بررسی مدل‌های مختلف با درجات^۱ مختلف پرداخته و بهترین مدل را بر اساس این معیارها انتخاب می‌کند. رابطه (۸) فرمول‌بندی AIC را نشان می‌دهد.

$$AIC = 2k - 2 \times \ln(\hat{L}) \quad (8)$$

در رابطه (۸)، K تعداد پارامترهای مدل و $\hat{L}$ تابع احتمال مدل است که از روی خطای RMSE به دست می‌آید. پس از تخمین و ارزیابی مدل‌ها با درجات مختلف، بهترین مدل بر اساس کمترین مقدار RMSE^۲ و AIC^۳ انتخاب می‌شود (شکل (۱۱)). برای این کار از توابع کمینه‌ساز استفاده می‌شود تا مدل‌هایی که کمترین خطا را دارند شناسایی شوند.



شکل ۱۱- معیارهای آماری انتخاب مرتبه مدل تابع تبدیل

فرم کلی تابع تبدیل $\frac{P}{\delta_{lat}}$ به صورت رابطه (۹) تعریف می‌شود [۴]. مرتبه مدل تابع تبدیل براساس موارد: حالت یکپارچه‌کننده زاویه غلت

(مرتبه اول)، غلت بدنه جفت شده / رگرسیون برآ افزا^۴ (مرتبه دوم)، پاسخ تاخیر/تقدم (مرتبه دوم)، دینامیک غلت هلندی (مرتبه دوم) و دینامیک عملگر (معادل تاخیر زمانی)، قابلیت افزایش تا مرتبه ۷ را دارد.

$$\frac{P}{\delta_{lat}}(s) = \frac{k\delta_{lat}L_{\beta_{1s}}e^{-\tau_{lat}s}}{\tau_f s^2 + s - \tau_f L_{\beta_{1s}}} \quad (9)$$

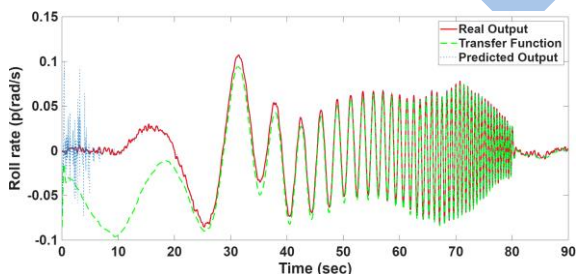
مدل تابع تبدیل تخمین زده شده که نتیجه آن در شکل (۸) آورده شده است، مرتبه دو بوده و به صورت رابطه (۱۰) می‌باشد.

$$\frac{P}{\delta_{lat}}(s) = \frac{0.7516s + 75.88}{s^2 + 10.84s + 96.53} \quad (10)$$

مدل مرتبه ۴ پاسخ دینامیک غلت نیز به فرم رابطه (۱۲) قابل تعریف می‌باشد.

$$\frac{P}{\delta_{lat}}(s) = \frac{0.8243s^3 + 70.14s^2 + 5.485s + 13.63}{s^4 + 9.938s^3 + 93.7s^2 + 5.551s + 17.24} \quad (12)$$

بر این اساس، با بهره‌گیری از مدل‌های تابع انتقال حاصل از تخمین‌های تحلیلی و مدل یادگیری تطبیقی مبتنی بر الگوریتم یادگیری Q، که به‌منظور بهینه‌سازی فرایند یادگیری به‌کار گرفته شده است، مقایسه‌ای جامع میان خروجی تابع تبدیل مربوط به پاسخ غلت در کانال عرضی بالگرد و داده‌های تجربی اندازه‌گیری شده صورت گرفته است. نتایج این مقایسه در شکل (۱۲) ارائه شده است و به‌وضوح دقت و صحت مدل پیش‌بینی شده را در مقایسه با داده‌های واقعی نشان می‌دهد.



شکل ۱۲- خروجی مدل تابع تبدیل و الگوریتم یادگیری Q با داده‌های پروازی رویداد اول

نتایج به خوبی مزیت روش‌های مبتنی بر هوش مصنوعی و تعامل محیطی را نشان می‌دهد. با توجه به پیچیدگی مانور فرکانس جاروب، تعامل و یادگیری خروجی حاصل از آن بسیار حائز اهمیت است.

۵- نتیجه‌گیری

در این مطالعه، شناسایی خارج از خط شش درجه آزادی کانال عرضی-سمتی نرخ غلت بالگرد BO-105 در فاز پرواز ایستا با استفاده از داده‌های تست پروازی و مانور تحریک فرکانس جاروب سکان عرضی، به‌طور مؤثر با بهره‌گیری از الگوریتم‌های یادگیری تقویتی و تعامل

^۴ Flap

^۱ Orders

^۲ Root mean square error

^۳ Akaike information criterion

- [1] Nelles O. Introduction to optimization. In Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes 2001 (pp. 23-34). Cham: Springer International Publishing.
- [2] Cai G, Chen BM, Lee TH. Unmanned rotorcraft systems. Springer Science & Business Media; 2011 Jun 1.
- [3] Zhou J, Liu S, Lu J, Liu X. An improved dynamic model identification method for small unmanned helicopter. Aircraft Engineering and Aerospace Technology. 2024 Feb 6;96(1):175-83.
- [4] Tischler MB, Rempke RK. Aircraft and rotorcraft system identification. Reston, VA: American Institute of Aeronautics and Astronautics; 2012 Aug.
- [5] Sutton RS, Barto AG. Reinforcement learning: An introduction. Cambridge: MIT press; 1998 Mar 1.
- [6] Martin H, Fernandez Vicente O, Perez S, Belfadil A, Ibanez-Llano C, Perozo Rondon FJ, Valle JJ, Arechalde Pelaz J. Reinforcement Learning in System Identification. arXiv e-prints. 2022 Dec:arXiv-2212.
- [7] Liu SY, Zhou J, Shi JY, Lu J. Small unmanned helicopter system identification based on the weighted least square method and improved grey wolf optimisation algorithm. The Aeronautical Journal. 2025 Jul;129(1337):1886-902.
- [8] Wei-Zhen Z, Zhen H, Zhang-Fan T. Reinforcement learning control design for perching maneuver of unmanned aerial vehicles with wind disturbances. Journal of Shanghai Jiao Tong University. 2024;58(11):1753-61.
- [9] Liu Z, Li J, Wang C, Yu R, Chen J, He Y, Sun C. System identification based on generalized orthonormal basis function for unmanned helicopters: A reinforcement learning approach. IEEE Transactions on Vehicular Technology. 2021 Jan 14;70(2):1135-45.
- [10] Maheshwari A, Davendralingam N, DeLaurentis DA. A comparative study of machine learning techniques for aviation applications. In 2018 Aviation Technology, Integration, and Operations Conference 2018 (p. 3980).
- [11] Martinsen AB, Lekkas AM, Gros S. Combining system identification with reinforcement learning-based MPC. IFAC-PapersOnLine. 2020 Jan 1;53(2):8130-5.
- [12] Bohn E, Coates EM, Moe S, Johansen TA. Deep reinforcement learning attitude control of fixed-wing uavs using proximal policy optimization. In 2019 international conference on unmanned aircraft systems (ICUAS) 2019 Jun 11 (pp. 523-533). IEEE.
- [13] Woo J, Park J, Yu C, Kim N. Dynamic model identification of unmanned surface vehicles using deep learning network. Applied Ocean Research. 2018 Sep 1; 78:123-33.
- [14] Mettler B, Tischler MB, Kanade T. System identification of small-size unmanned helicopter

محیطی مورد بررسی قرار گرفت. هدف این رویکرد، ارائه شبیه‌سازی دقیقی از رفتار دینامیکی بالگرد و شناسایی مدل‌هایی برای نرخ غلت در شرایط مختلف پروازی بود. علاوه بر این، اعتبارسنجی الگوریتم در فرآیند یادگیری به‌طور مطلوبی نشان داده شد.

با استفاده از داده‌های تست پروازی و الگوریتم‌های یادگیری تقویتی، نتایج شبیه‌سازی به‌خوبی با داده‌های واقعی مطابقت داشت و درصد تطابق به‌طور میانگین ۹۸٪ در شرایط مختلف پروازی برآورد شد. این تطابق بالا نشان‌دهنده دقت مدل‌های پیشنهادی و توانایی الگوریتم‌های یادگیری تقویتی در شبیه‌سازی رفتارهای پیچیده دینامیکی سیستم‌های غیرخطی، مانند نرخ غلت بالگرد، در طی فرآیندهای کنترل است.

ویژگی‌های آماری به‌دست‌آمده نشان می‌دهند که رویکرد یادگیری تقویتی استفاده‌شده در این مقاله، با توجه به ورودی پهن باند فرکانس جاروب با دامنه فرکانسی وسیع و سرعت تغییر فرکانس بالا، توانایی و دقت بسیار خوبی در فرآیند شناسایی سیستم و مدل‌سازی دینامیکی وسایل پروازی دارد.

در نهایت، این مطالعه نشان می‌دهد که الگوریتم‌های یادگیری تقویتی، به‌دلیل تعامل با محیط و یادگیری تدریجی، و همچنین عدم نیاز به مدل‌سازی دقیق از پیش، قادر به انطباق با تغییرات دینامیکی سیستم‌ها هستند. این الگوریتم‌ها موجب افزایش دقت پیش‌بینی رفتار سیستم، به‌ویژه در ترکیب با داده‌های تست پروازی و تعاملات محیطی، می‌شوند و می‌توانند به‌طور مؤثری در بهبود دقت شبیه‌سازی رفتار سیستم‌های دینامیکی پیچیده، از جمله نرخ غلت بالگرد، به کار گرفته شوند. این روش‌ها نه تنها به بهبود مدل‌سازی سیستم‌های پروازی کمک می‌کنند، بلکه در طراحی سیستم‌های کنترل تطبیقی و پیشرفته برای انواع مختلف هواپیماها و بالگردها نیز کاربرد دارند.

نتایج به‌دست‌آمده نشان‌دهنده کارایی بالای روش‌های یادگیری تقویتی در شناسایی و پیش‌بینی رفتارهای دینامیکی پیچیده در سیستم‌های پروازی است. در این مطالعه، از روش یادگیری Q برای شناسایی کانال غلت بالگرد استفاده شد که دقت مطلوبی را در پی داشته است. با توجه به تمرکز مقاله بر بررسی کارایی روش‌های مبتنی بر داده، ارزیابی با روش SARSA به‌عنوان معیار پایه و هم‌رده انجام گرفت. با این حال، برای توسعه‌های آینده، مقایسه روش پیشنهادی با الگوریتم‌های پیشرفته یادگیری تقویتی مبتنی بر شبکه‌های عصبی مانند DQN، PPO و Actor-Critic به‌منظور تحلیل جامع‌تر عملکرد در شرایط مختلف پروازی پیشنهاد می‌شود.

۶- نمادها

A	کنش
R	پاداش
Q	ارزش یک کنش
α	طول گام تخمین بازگشتی ارزش کنش

۷- مراجع

- dynamics. In Annual Forum Proceedings- American Helicopter Society 1999 May 25 (Vol. 2, pp. 1706-1717).
- [15] Khaligh SP, Fahimi F, Robert Koch C. A system identification strategy for nonlinear model of small-scale unmanned helicopters. *Journal of the American Helicopter Society*. 2016 Oct 1;61(4):1-3.
 - [16] Vladov S, Banasik A, Sachenko A, Kempa WM, Sokurenko V, Muzychuk O, Pikiewicz P, Molga A, Vysotska V. Intelligent Method of Identifying the Nonlinear Dynamic Model for Helicopter Turboshift Engines. *Sensors*. 2024 Oct 9;24(19):6488.
 - [17] Khodaverdian M, Malekzadeh M. Attitude stabilization of spacecraft simulator based on modified constrained feedback linearization model predictive control. *IET Control Theory & Applications*. 2023 May;17(8):953-67.
 - [18] Kerler M, Hönle J, Kau HP. Modeling of BO 105 flight dynamics for research on fuel savings due to single-engine operation.
 - [19] Ramanujam R, Rao S. Stability Analysis of Variable Geometry Helicopters. In 4th Asian/Australian Rotorcraft Forum 2015.
 - [20] Thomson DG, Bradley R. Recent developments in the calculation of inverse solutions of the helicopter equations of motion.
 - [21] Tischler MB. System identification requirements for high-bandwidth rotorcraft flight control system design. *Journal of Guidance, Control, and Dynamics*. 1990 Sep;13(5):835-41.
 - [22] Shethiya AS. Adaptive Learning Machines: A Framework for Dynamic and Real-Time ML Applications. *Annals of Applied Sciences*. 2024 May 17;5(1).
 - [23] Mark BT, Robert KR. Aircraft and rotorcraft system identification. AIAA, Virginia, USA. 2006.
 - [24] Tischler MB, Remple RK. Aircraft and rotorcraft system identification. Reston, VA: American Institute of Aeronautics and Astronautics; 2012 Aug.
 - [25] Geluardi S, Nieuwenhuizen FM, Pollini L, Bühlhoff HH. Frequency domain system identification of a light helicopter in hover. In 70th American Helicopter Society International Annual Forum (AHS 2014) 2014 May (pp. 1721-1731). Curran.