Available in:
Journal.isrc.ac.ir

Journal of Space Science, Technology & Applications (Persian)

Vol. 4, No. 1, pp.: 64-77
2024

DOI:
10.22034/jssta.2024.446359.1153

Article Info

Received: 2024-03-03

Accepted: 2024-05-18

Keywords

Low-Thrust , Equinoctial orbital elements , Reinforcement Learning , Actor, Critic networks , Agent

How to Cite this article

Hamed. Soleymani, M. Bakhtiari, K. Daneshjou, ” Low-Trust Trajectory Design from LEO to GEO Using Reinforcement Learning”, *Journal of Space Science, Technology and Applications*, vol 4 (1), p.: 64-77, 2024.

Original Article

Low-Trust Trajectory Design from LEO to GEO Using Reinforcement Learning

Hamed Soleymani, Majid Bakhtiari, Kamran Daneshjou

1.* PhD student in aerospace engineering at Iran University of Science and Technology, Tehran, Iran, h_soleymani99@nt.iust.ac.ir (Corresponding Author)
2. School of Advanced Technologies - Iran University of Science and Technology, Tehran, Iran, bakhtiari_m@iust.ac.ir
3. Faculty member of Iran University of Science and Technology, Tehran, Iran, kjoo@iust.ac.ir

Abstract

In this paper, In the preliminary phase of space mission design, the selection of the spacecraft's trajectory is critical. This study simulates the dynamics of low-thrust orbital transfers within a two-dimensional orbital plane, employing a set of ordinary differential equations to represent the continuum of the spacecraft's orbital elements. These elements are encapsulated by six orbital parameters, manipulated under a defined thrust vector strategy within the action space, adhering to a specified policy framework. An agent, trained via a reinforcement learning algorithm within an actor-critic network architecture, is tasked with executing a low-thrust transfer between Low Earth Orbit (LEO) and Geostationary Orbit (GEO). The algorithm dynamically adjusts the spacecraft's trajectory, informed by initial orbital conditions and mission-specific constraints, to derive an optimal thrust angle trajectory and corresponding adjustments in orbital elements for the maneuver. To validate the algorithm's efficacy and robustness, a comparative analysis is conducted by implementing an alternative transfer mode at a varied orbital altitude. Additionally, the study explores the impact of adjusting the algorithm's degradation coefficient hyperparameter on the learning efficacy. Conclusively, the findings suggest that the agent, once adequately trained within the specified dynamical model, is capable of autonomously executing analogous orbital transfers. This is achieved without necessitating reiteration of the dynamical simulations, contingent solely upon the stipulation of initial and terminal orbital parameters

مقاله پژوهشی

طراحی مسیر فضاپیمای تراست پایین از مدار لئو به ژئو با استفاده از یادگیری تقویتی

حامد سلیمانی*[۱]، مجید بختیاری[۲]، کامران دانشجو[۳]

۱- دانشجوی دکترا هوافضا دانشگاه علم و صنعت ایران، تهران، ایران h_soleymani99@nt.iust.ac.ir
۲- عضو هیئت علمی دانشگاه علم و صنعت ایران، تهران، ایران bakhtiari_m@iust.ac.ir
۳- عضو هیئت علمی دانشگاه علم و صنعت ایران، تهران، ایران kjoo@iust.ac.ir
* نویسنده مسئول

چکیده

در فاز اولیه طراحی مأموریت‌های فضایی، انتخاب دقیق مسیر فضاپیما از اهمیت بالاتری برخوردار است. در این پژوهش، دینامیک انتقال مداری تراست پایین دایروی صفحه‌ای بر اساس معادلات دیفرانسیل اعتدالی به‌عنوان محیط پیوسته برای متغیرهای مسئله که شش عنصر مداری اعتدالی یک فضاپیما هستند، شبیه‌سازی می‌شود. بردار رانش به عنوان فضای عمل تعریف شده و تحت یک سیاست انتخاب و به محیط اعمال می‌شود. عامل توسط الگوریتم یادگیری تقویتی شبکه بازیگر-منتقد برای انجام انتقال مداری تراست پایین از مدار لئو به مدار ژئو آموزش داده می‌شود. مسیر فضاپیما توسط الگوریتم مطابق با شرایط اولیه و قیود ماموریت جستجو می‌شود و در نهایت، پروفیل زاویه تراست مطلوب و تغییرات عناصر مداری مرتبط برای یک حالت مانور انتقال مداری به دست خواهند آمد. برای بررسی دقت و اعتبار الگوریتم در نتایج حالت اول مانور مداری، حالت دوم با اندازه تراست متفاوت پیاده سازی می شود. همچنین تأثیر تغییر فراپارمتر ضریب تنزل الگوریتم بر روند یادگیری نیز بررسی می‌شود. در نهایت با در نظر گرفتن نتایج، عامل آموزش دیده در محیط دینامیک مسئله می‌تواند مأموریت‌های مشابه را بدون نیاز به شبیه‌سازی مجدد دینامیک مسئله و پارامترهای آن و فقط با تعیین شرایط اولیه و نهایی، با موفقیت به انجام برساند

دسترس‌پذیر در نشانی:
Journal.isrc.ac.ir

دو فصلنامه
علـوم، فـناوری و
کاربردهـای فضـایی

سال چهارم، شماره ۱، صفحه ۷۷-۶۴
بهار و تابستان ۱۴۰۳

DOI:
10.22034/jssta.2024.446359.1153

تاریخچه داوری

دریافت: ۱۴۰۲/۱۲/۱۳

پذیرش: ۱۴۰۳/۰۲/۲۹

واژه‌های کلیدی

تراست پایین، عناصر مداری اعتدالی، یادگیری تقویتی، عامل، شبکه بازیگر، منتقد

نحوه استناد به این مقاله

حامد سلیمانی، مجید بختیاری، کامران دانشجو، "طراحی مسیر فضاپیمای تراست پایین از مدار لئو به ژئو با استفاده از یادگیری تقویتی"، *دوفصلنامه علوم، فناوری و کاربردهای فضایی*، جلد چهارم، شماره اول، صفحات ۷۷-۶۴، ۱۴۰۳.

۱- مقدمه

مسئله طراحی مسیرهای فضایی بهینه برای کاهش مصرف سوخت و زمان انتقال، موضوع اساسی در حوزه هوافضا است. به‌منظور دستیابی به مسیرهای بهینه سازگار با پارامترهای کلی مأموریت، مانند سیستم پیشرانش و جرم سوخت در دسترس، باید یک مسئله بهینه‌سازی جامع و پیچیده حل شود. این مسئله از دهه ۱۹۵۰ مورد توجه قرار گرفته و توسط پیشگامانی همچون لاودن، میل، کان وی، لیتمن و ادلبام بررسی شده است. مسئله انتقال مداری تراست پایین به دلیل پیچیدگی خاصی که دارد، نیاز به روش‌های عددی دارد و به‌طور تحلیلی قابل حل نمی‌باشد. استفاده از پیشران سوزش پیوسته به‌طور گسترده‌ای برای ملاقات مداری[۵]، نگهداری مدار[۶]، انتقال مدار و مأموریت‌های فضایی بین سیاره‌ای موردمطالعه قرارگرفته است. سیستم‌های پیشرانش تراست پایین می‌تواند امکان‌پذیری‌های مأموریت را با استفاده کارآمدتر از سوخت، بسیار بهبود ببخشد که به نوبه خود اجازه افزایش نسبت بار محموله به وزن کل فضاپیما و یا کاهش قابل‌توجه حجم فضاپیما و در نتیجه ابعاد و سایز پرتابگر را می‌دهد. هنگام استفاده از پیشرانه تراست پایین برای انجام انتقالات مداری، ضربه ویژه[۷] بالا ترکیب‌شده با یک نیروی ویژه کوچک منجر به زمان انتقال بسیار طولانی می‌شود.[۱]

برای یافتن مسیر بهینه، نیاز به حل مسئله پیچیده ای است که «مسئله دو مقدار مرزی[۸]» نامیده می شود. روش های کلاسیک برای حل مسئله کنترل بهینه به طور کلی به سه دسته روش های مستقیم، غیر مستقیم و ترکیبی تقسیم می‌شوند. در روش‌های مستقیم، با گسسته کردن مسیر به قوس‌های کوتاه با قدر و جهت رانش ثابت و اعمال شرایط مرزی مناسب، مسئله پیوسته به یک مسئله برنامه‌ریزی غیرخطی تبدیل می‌شود.[۷] روش‌های غیرمستقیم از تکنیکی استفاده می‌کنند که به عنوان اصل حداقل پونتراگین[۹] شناخته می‌شود، که مسئله را به یک مسئله ارزش مرزی دو نقطه‌ای تبدیل می‌کند. از سوی دیگر روش های ترکیبی، بهره گیری از مزایای هر دو روش ذکر شده است. در سال های گذشته مطالعات زیادی در زمینه بهینه سازی زمان انتقال و مصرف سوخت (مینیمم کردن مقدار تغییرات سرعت) در انتقال های ضربه ای انجام شده است[۱۸و۱۹]. در حوزه تراست پایین، انتقال حداقل زمان ژئوسنکرون برای فضاپیماهای تراست پایین با استفاده از روش‌های پیوسته توسط کایلو و همکاران [۲] بررسی شد. آن‌ها مسیرهای انتقال چندگانه چرخه سوخت را با استفاده از روش تک تیراندازی و روش هیبریدی پاول بهینه کردند. میچا کیم [۳]یک رویکرد جدید برای بهینه‌سازی مسیرهای تراست پایین پیوسته، با تمرکز بر یک الگوریتم بهینه‌سازی مبتنی بر روش‌های غیرمستقیم پیشنهاد کرد و یک الگوریتم تبرید تطبیقی[۱۰] را برای تولید حدس اولیه معرفی کرد. رایان راسل [۴] از حساب تغییرات و نظریه بردار آغازگر برای انتقال زمین-ماه با رانش کم با چرخش های متعدد استفاده کرد. در این کار، با استفاده از جستجوی جهانی برای راه‌حل‌های روش غیرمستقیم محلی، به‌یک جبهه بهینه پرتو برای زمان پرواز و جرم نهایی دست یافت. در سال ۲۰۱۳ یک مطالعه مقایسه ای بر اساس روش های غیر مستقیم برای حل مسئله کنترل بهینه غیرخطی برای مسیر تراست پایین فضایی توسط نوابی و همکاران[۱۶] ارائه شد. که لی هوانگ [۵] بر بهینه‌سازی انتقال‌های تراست پایین به نقاط لاگرانژی تمرکز کرد و از برنامه‌ریزی درجه دوم متوالی برای بهینه‌سازی مسیرهای انتقال از مدارهای پارکینگ به مدارهای لاگرانژی استفاده کرد. بروس کانوی و مارو پونتانی [۶]یک روش بهینه سازی ازدحام ذرات[۱۱] غیرمستقیم را پیشنهاد کردند، که پیش نیازهای ضروری بهینه سازی را با یک الگوریتم اکتشافی یکپارچه می کند. جلناتان عزیز و همکاران [۸]یک تبدیل سلندمن[۱۲] با برنامه‌ریزی پویای دیفرانسیلی[۱۳] را برای بهینه‌سازی معادلات حرکت فضاپیما اعمال کرد و متغیر مستقل را به زاویه مدار برای بهینه‌سازی در اطراف اجسام سه‌یاره‌ای تغییر داد. ویلیامز و کاورستون-کارول [۹] با استفاده از الگوریتم ژنتیک[۱۴]، مسئله حداقل سوخت را برای فضاپیمای پیشران الکتریکی خورشیدی مطالعه کردند. فکور و همکاران[۲۰] در سال ۲۰۱۹ مسئله کنترل بهینه مسیر تراست پایین دایروی از مدار لئو به ژئو را با ترکیب روش تحلیلی و الگوریتم جمعیتی کلونی زنبور مصنوعی[۱۵] برای حالت های کمترین زمان، کمترین تلاش کنترلی مطالعه کرده اند. در [۱۰]، ژنبو وانگ یک استراتژی بهینه سازی محدب برای حل مسائل انتقال بهینه تراست پایین، با تاکید بر بهره‌وری سوخت، توسعه داد. در واقع هر دو روش مستقیم و غیر مستقیم

ماهیت محلی دارند. این بدان معنی است که راه حل بهینه ای که آن‌ها می‌توانند شناسایی کنند به نخستین راه حل موجود بستگی دارد. نتیجه عددی به دست آمده از تکنیک های قطعی به طور معمول در مجاورت حدس است [۶]. در نتیجه بررسی روش‌های حل مسئله کنترل بهینه که در بالا بیان شد، می‌توان گفت که انتخاب بین این روش‌ها به نیازهای خاص مسئله از جمله پیچیدگی قیود، الزامات دقت عددی و امکان‌سنجی حدس اولیه بستگی دارد. همچنین روش‌های مبتنی بر جمعیت مانند الگوریتم ازدحام ذرات و ژنتیک مسائل مقادیر اولیه هستند و برای همگرایی نیاز به راه‌حل اولیه خوبی دارند. از آنجایی که مسئله انتقال مدار تراست پایین-یک مسئله مقدار مرزی دونقطه‌ای است و رسیدن به حالت نهایی مطلوب مهم است، استفاده از الگوریتمی لازم است که بتواند این فرآیند را برآورده کند. علاوه بر این، الگوریتم ازدحام ذرات و ژنتیک می توانند در حداقل های محلی به خصوص در مسائل دارای فضاهای حالت بسیار پیچیده‌یا چند حالته به دام بی‌افتند، عملکرد آن‌ها می تواند با افزایش ابعادیا پیچیدگی مسئله تضعیف شود، به تنظیمات پارامترهایشان حساس هستند.

این اشکالات تقریباً توسط روش‌های یادگیری ماشین پوشش داده می‌شوند و باعث می‌شوند که این روش‌ها در مسائل ابعاد بالا و بسیار پیچیده استفاده شوند، در حداقل‌های محلی به دام نیافتند و در کاربردهای خودکار و خودران استفاده شوند. یکی از شاخه های مهم در یادگیری ملشینی، یادگیری تقویتی است که در سیستم های مختلف مانند ماشین‌ها و ربات‌های صنعتی استفاده می‌شود. این نوع یادگیری به دلیل ماهیت تعاملی‌اش، بیشتر شبیه به یادگیری از محیط اطراف است که در انسان و حیولنات دیده می شود. در زمینه‌یادگیری تقویتی، عامل برای دستیابی به اهداف خود نیازی به دانش قبلی از دینامیک محیط ندارد. این ویژگی، الگوریتم‌های یادگیری تقویتی را قادر می‌سازد تا به طور مؤثری با مسائل مدل‌سازی تحلیلی پیچیده مقابله کنند. از این رو، این ویژگی امکان آموزش کارآمد عامل را فراهم می کند و آن را برای استفاده در موقعیت های مختلف با دینامیک مسئله مشابه قابل تطبیق می کند.

در این پژوهش برای تحلیل عملکرد این نوع الگوریتم های یادگیری ماشین به عنوان یک کار پیشگام در زمینه انتقالات مداری به ارتفاع بالا، از الگوریتم شبکه بازیگر- منتقد استفاده می شود. دینامیک انتقال مداری تراست پایین دایروی هم‌صفحه بر اساس معادلات دیفرانسیل اعتدالی به عنوان محیطی برای تعامل عامل که فضایی پیوسته برای متغیرهای مسئله که شش عنصر مداری اعتدالی یک فضاپیما هستند، شبیه سازی می شود. بردار رانش به عنوان فضای عمل تعریف شده و تحت یک سیاست تصادفی انتخاب و به محیط اعمال می‌شود. عامل توسط الگوریتم شبکه بازیگر-منتقد آموزش داده می‌شود تا بتواند انتقال مداری تراست پایین موردنظر را انجام دهد. مسیر بهینه توسط این الگوریتم مطابق با شرایط اولیه و قیود ماموریت(مانند اندازه نیروی تراست) جستجو می‌شود و در نهایت پروفیل های زاویه تراست بهینه برای دو حالت مانور انتقال مداری و یک حالت مانور تغییر زاویه میل مدار به دست خواهند آمد.

۲- یادگیری ماشین؛ یادگیری تقویتی

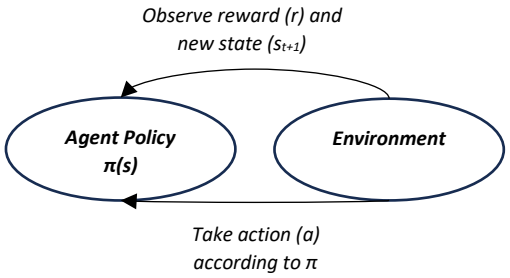
هر مقاله یادگیری ماشین به طور کلی به سه دسته تقسیم می‌شود: یادگیری نظارت شده، یادگیری بدون نظارت و یادگیری تقویتی. آخرین دسته، یادگیری تقویتی، که هدف این پژوهش است، به یادگیری ماشین اجازه می دهد تا برای مسائلی که هیچ داده ای در دسترس نیست استفاده شود. یادگیری تقویتی عیوب کلی روش های قبلی را برطرف کرده و بر روی طیف وسیعی از مسائل کنترلی پیاده سازی شده است. سیستم یادگیری در یادگیری ماشین عامل نامیده می‌شود و رفتار بهینه آن عامل با یک نگاشت از ناحیه حالت X به ناحیه عمل A به صورت $S: X \rightarrow A$ تعریف می شود. سیاست یک اصطلاح کلی است که یک عامل برای انجام یک عمل در یک محیط از آن پیروی می‌کند. مقدار پاداش عددی است که به یک عامل می گوید که‌یک عمل دریک انتقال حالت چقدر خوب است. [۱۱]

فرآیند تصمیم گیری مارکوف چارچوبی است که برای تعریف مسائل یادگیری تقویتی استفاده می شود. در یادگیری تقویتی، یک عامل در محیطی با هدف به حداکثر رساندن ارزش پاداش قرار می گیرد. عامل با انجام یک عمل تعیین شده توسط یک سیاست با محیط تعامل دارد. وقتی حالت خاصی حاصل شد، پاداشی توسط محیط به عامل داده می شود. ارزش به عنوان بازگشت مورد انتظار در حالت s در هنگام پیروی از یک خط

مشی π تعریف می شود. تابع ارزش استخراج شده و مشتق شده از معادله بلمن، کل پاداش دریافتی با انجام‌یک عمل در‌یک حالت خاص را بیان می کند. از نظر‌یک فرآیند مارکوف محدود، تابع ارزش را می توان به صورت زیر تعریف کرد:

$$v_\pi = E_\pi[G_t|S_t = s] = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}|S_t = s\right] \qquad (۱)$$

فرآیند تصمیم‌گیری مارکوف در شکل (۱) نمایش داده شده است.[۱۱]



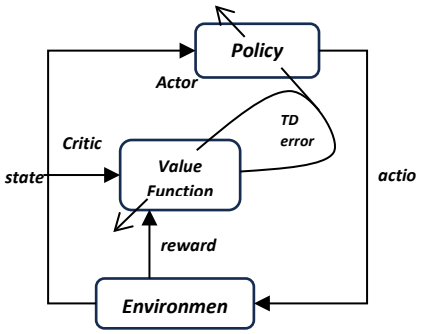
شکل ۱. فرایند تصمیم‌گیری مارکوف[۱۱]

E_π مقدار مورد انتظار یک متغیر تصادفی است زمانی که یک عامل از یک سیاست خاص π پیروی می‌کند؛ R مقدار پاداش در یک حالت در زمان t است و γ یک ضریب تنزل[۱۱] است که مقدار آن بین صفر و ۱ است. ضریب تنزل نشانه از میزان تأثیر پاداش‌های آینده بر تابع ارزش است.[۱۱]

گرادیان سیاست[۱۱] یک عمل را بر اساس یک سیاست پارامتری شده بدون یادگیری تابع ارزش انتخاب می‌کند. با روش های گرادیان سیاست، سیاست باید پیوسته و قابل انتگرال‌گیری باشد. این اجازه می دهد تا روش های گرادیان سیاست در فضاهای عمل پیوسته اعمال شوند. دو الگوریتم اصلی که گرادیان های سیاست را اعمال می کنند عبارتند از REINFORCE، یک گرادیان سیاست مبتنی بر مونت کارلو و روش‌های Actor-Critic. این الگوریتم‌ها را می‌توان در فضاهای عمل پیوسته اعمال کرد. [۱۱]

الگوریتم های منتقد بازیگر نشان داده شده در شکل ۲، روش‌های تفاضل زمانی و گرادیان سیاست را برای کاهش اشکالات دو نوع الگوریتم ترکیب می‌کنند. الگوریتم‌های بازیگر-منتقد شامل بازیگری است که عملی را بر اساس یک سیاست انتخاب می‌کند درحالی که سیاست توسط منتقد با استفاده از یک تابع ارزش ارزیابی می‌شود. الگوریتم گرادیان سیاست قطعی عمیق[۱۱] یکی از چندین الگوریتم‌های بازیگر-منتقد است. به طور معمول، منتقد‌یک تابع ارزش حالت[۱۱] است. با انتخاب شدن‌یک عمل، منتقد وضعیت جدید را ارزیابی می‌کند تا مشخص کند که آیا اوضاع بهتر یا بدتر از حد انتظار پیش رفته است. این ارزیابی خطای اختلاف زمانی است که به آن خطای تفاضل زمانی[۱۱] می گویند:[۱۱]

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \qquad (۲)$$

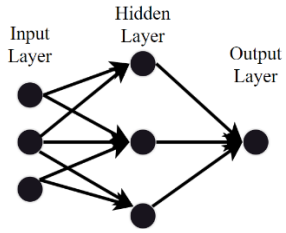


شکل ۲. معماری شبکه بازیگر- منتقد[۱۲]

که در آن $V(s_t)$ تابع ارزش فعلی است که توسط منتقد پیاده‌سازی شده است. این خطای تفاضل زمانی می تواند برای ارزیابی عمل انتخاب شده a_t، عمل انجام شده در حالت s_t استفاده شود. اگر خطای تفاضل زمانی مثبت باشد، نشان می‌دهد که تمایل به انتخاب a_t باید برای آینده تقویت شود، در حالی که اگر خطای تفاضل زمانی منفی باشد، نشان می دهد که این تمایل باید ضعیف شود. [۱۱]

مدل منتقد تحت عنوان یک شبکه عصبی نمایش داده می‌شود که ورودی‌های آن حالت و عمل یک عامل و خروجی آن‌یک مقدار q منفرد است.[۱۱]

یک شبکه عصبی پیش‌خور از لایه‌هایی تشکیل‌شده است که از طریق وزن‌ها به یکدیگر متصل شده‌اند. به‌طورکلی از یک‌لایه ورودی، یک یا چندلایه پنهان و یک‌لایه خروجی تشکیل‌شده است که در شکل ۳ نشان داده شده است.



شکل ۳. شبکه عصبی پیش‌خور[۱۱]

۳- دینامیک مسئله انتقال مداری تراست پایین

کنترل یک فضاپیما را می‌توان با فرآیند تصمیم‌گیری مارکوف توصیف کرد، به این معنا که در هر مرحله زمانی فضاپیما حالت خاصی دارد و باید تصمیم گرفت که چه اقدامی انجام دهد تا درنهایت به یک هدف برسد. [۱۳]

در مانورهای تراست پایین، یک نیروی خارجی **F** به فضاپیما معادله حرکت اضافه می‌شود:[۱۴]

$$\ddot{\mathbf{r}} = \frac{\mu}{|\mathrm{r}|^3}\mathbf{r} + \frac{\boldsymbol{F}}{m} \qquad (۳)$$

جایی که m جرم فضاپیما و μ ثابت گرانشی استاندارد جسم مرکزی است.[۱۴]

دینامیک غیرخطی مانورهای تراست پایین را می‌توان با استفاده از معادلات تغییرات گاوس توسط معادلات برحسب عناصر مداری کپلری بیان کرد که در مرجع ۶ بیان شده است. اما یکی از معایب استفاده از عناصر مداری کپلری، تکینگی درشیب مداری یا خروج از مرکز صفر است. لذا به سراغ عناصر مداری اعتدالی می‌رویم. مجموعه‌ای از عناصر مداری هستند که برای تحلیل و طراحی مسیر بهینه مفید هستند. آن‌ها برای مدارهای دایروی، بیضی و هذلولی معتبر هستند. این معادلات اعتدالی اصلاح‌شده، هیچ تکینگی برای خروج از مرکز صفر و زاویه شیب مداری برابر با ۰ و ۹۰ درجه را نشان نمی‌دهند و می‌توان از این تکینگی‌ها جلوگیری کرد.

$$p = a(1 - e^2) \qquad (۴)$$

$$f = e\,cos(\omega + \theta) \qquad (۵)$$

$$g = e\,sin(\omega + \theta) \qquad (۶)$$

$$h = tan\left(\frac{i}{2}\right) cos(\Omega) \qquad (۷)$$

$$k = tan\left(\frac{i}{2}\right) \sin(\Omega) \qquad (۸)$$

$$L = \Omega + \omega + \theta \qquad (۹)$$

حال معادلات دیفرانسیل معمولی مرتبه اول گاوس برحسب عناصر اعتدالی بازنویسی می‌شود:[۱۱]

$$\frac{dp}{dt} = \frac{2p}{w}\sqrt{\frac{p}{\mu}}F_s \qquad (۱۰)$$

$$\frac{df}{dt} = \sqrt{\frac{p}{\mu}}\left[\sin L\,F_r + [(w+1)\cos L + f]\frac{F_s}{w} - (h\sin L - k\cos L)\frac{gF_w}{w}\right] \qquad (۱۱)$$

$$\frac{dg}{dt} = \sqrt{\frac{p}{\mu}}\left[-\cos L\,F_r + [(w+1)\sin L + g]\frac{F_s}{w} + (h\sin L - k\cos L)\frac{gF_w}{w}\right] \qquad (۱۲)$$

$$\frac{dh}{dt} = \sqrt{\frac{p}{\mu}}\frac{s^2F_w}{2w}\cos L \qquad (۱۳)$$

$$\frac{dk}{dt} = \sqrt{\frac{p}{\mu}}\frac{s^2F_w}{2w}\cos L \qquad (۱۴)$$

$$\frac{dL}{dt} = \sqrt{\mu p}\left(\frac{w}{p}\right)^2 + \frac{1}{w}\sqrt{\frac{p}{\mu}}(h\sin L - k\cos L)\,F_w \qquad (۱۵)$$

که در آن F_s و F_r و F_w به ترتیب نیرو تراست در جهت‌های شعاعی، مماسی و نرمال بر صفحه مداری هستند. جایی که r در امتداد جهت شعاعی مدار، s جهت محیطی و w در امتداد بردار ممنتوم زاویه‌ای است.[۱۱]

$$s^2 = 1 + h^2 + k^2 \qquad (۱۶)$$

$$w = 1 + f\cos L + g\sin L \qquad (۱۷)$$

در این بررسی برای تحلیل عملکرد این نوع از الگوریتم‌های یادگیری ماشین و ایجاد یک پژوهش اولیه و هم‌چنین به‌عنوان یک کار اولیه در زمینه انتقال مداری ارتفاع‌های بالا با این دسته از روش‌ها برای محققان دیگر، این پیاده‌سازی از حالت بدون مدل استفاده می‌شود. در اینجا موارد مهم و مدنظر برای انجام فرآیند طراحی مسیر تغییرات نیم قطر اصلی، خروج از مرکز و تغییرات زاویه نیروی رانش رانشگر است.

تابع پاداش (مثبت برای یک پاداش، منفی برای یک جریمه) پاداش‌هایی را برای قرار گرفتن در یک حالت خاص یا برای انجام عملی در یک حالت خاص مشخص می‌کند. تابع پاداش به‌طور ضمنی هدف یادگیری را بیان می‌کند. به‌گونه‌ای که تابع پاداش به اینکه چطور باید سیستم (یعنی فرآیند تصمیم‌گیری مارکوف) کنترل شود سمت‌وسو می‌دهد.[۱۵]

تابع پاداش می‌تواند یکی از چالش‌برانگیزترین انتخاب‌ها در هنگام طراحی یک محیط باشد. کیفیت تابع پاداش می‌تواند تعیین کند که عامل چقدر سریع به یک‌راه حل همگرا می‌شود. دادن یک پاداش پراکنده به عامل ممکن است زمان اجرا را به میزان قابل‌توجهی افزایش دهد و همگرایی عامل را به یک‌راه حل دشوار کند.

تابع پاداش در این تحقیق بر اساس اختلاف مجذور بین حالت هدف و حالت فعلی است که در زیر نشان داده‌شده است:

$$r_p = \frac{\sqrt{\left(p_{targ} - p(t)\right)^2}}{p_{targ}} \quad (۱۸)$$

$$r_f = \sqrt{\left(f_{targ} - f(t)\right)^2} \quad (۱۹)$$

$$r_g = \sqrt{\left(g_{targ} - g(t)\right)^2} \quad (۲۰)$$

$$r_h = \sqrt{\left(h_{targ} - h(t)\right)^2} \quad (۲۱)$$

$$r_g = \sqrt{\left(k_{targ} - k(t)\right)^2} \quad (۲۲)$$

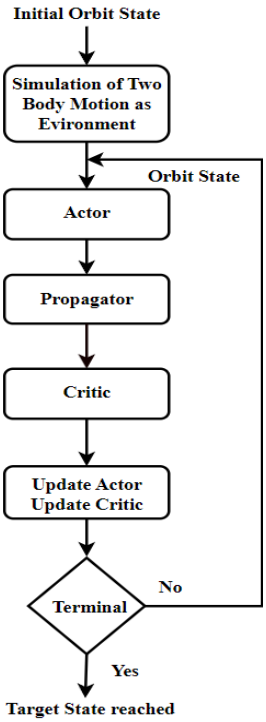
$$r = -\left(\alpha_p r_p + \alpha_f r_f + \alpha_g r_g + \alpha_h r_h + \alpha_k r_k\right) \quad (۲۳)$$

ضرایب α_p، α_f، α_g، α_h و α_k برای مقیاس بندی وزن نسبی عناصر مداری هستند. برای اینکه عامل یک سیاست خوب را یاد بگیرد، مقیاس بندی مناسب تابع پاداش لازم است. با توجه به تعریف تابع پاداش و اینکه مانور موردنظر افزایش ارتفاع یا اصلاح زاویه میل مدار باشد، ضرایب مربوط به عناصر مداری اعتدالی مؤثر در مأموریت باید بالاتر از سایر ضرایب تنظیم شوند، زیرا تابع پاداش به طور ضمنی هدف یادگیری را بیان می‌کند. از آنجایی که عموما در مانور های انتقال مدار هدف رسیدن به ارتفاع و جهت گیری مشخص شده مدار نهایی است و فقط حالت مداری در نظر گرفته می شود، عنصر مداری آنومالی حقیقی را می توان حذف کرد. بنابراین، عنصر مداری زاویه آنومالی حقیقی θ در هیچ‌یک از مأموریت ها هدف قرار نمی گیرد‌یا در تابع پاداش گنجانده نمی شود. این ضرایب در جدول ۱ ارائه شده است.

۴- نتایج

برای فضاپیما، تابع ارزش تابعی از تفاضل بین حالت فعلی و نهایی و هدف فضاپیما است. در محیط یادگیری تقویتی یک الگوریتم بر مبنای مدل بازیگر- منتقد، عناصر مداری اعتدالی ۶ گانه به‌عنوان پارامترهای دخیل در الگوریتم یا به عبارت بهتر حالت‌های مسئله خواهند بود. هدف این است که این عامل با تعامل با محیط بتواند بهترین دنباله از همین حالت‌ها را پیدا کند که یک مسیر برای انتقال مداری از مدارهای پایین به مدارهای بالاتر به‌حساب آید. در بخش شبکه بازیگر کار به این صورت است که بر اساس بازه‌ای که برای مقدار بزرگی نیروی رانش که توسط کاربر بر مبنای قیود و الزامات مأموریت انتخاب و یک دسته عمل به عامل داده می‌شود. عامل به‌صورت تصادفی از بین آن‌ها عملی را انتخاب کرده و به محیط مسئله اعمال می‌کند. از سوی دیگر شبکه منتقد که وظیفه ارزیابی تابع ارزش حلت مسئله را دارد، در هر اپیزود فرآیند آموزش، با ارزیابی‌هایی که از مقدار ارزش هر حالت انجام‌شده بر اساس عمل انتخاب‌شده توسط سیاست عامل،سعی می‌کند مقادیر به‌دست‌آمده به مقدار پاداش میانگین نزدیک‌تر کند و درنهایت نیز با برقراری شرایط توقف و رسیدن به نقطه نهایی با توجه به شرایط مرزی تعیین‌شده، الگوریتم را متوقف می‌کند و درنهایت پروفیل زاویه تراست بهینه برای انجام این انتقال حاصل خواهد شد. دراین‌بین شرط توقف مسئله با توجه به اینکه مأموریت مانور افزایش مدار باشد‌یا‌مانوری برای تغییر خروج از مرکز‌یا زاویه میل تعیین می‌شود. شکل ۶ پیاده سازی الگوریتم را در نرم افزار نمایش می دهد.

لازم به ذکر است که با توجه به ساده و ملموس بودن عناصر مداری کپلری در مباحث مکانیک مداری، و همچنین بیان بهتر نتایج خروجی حاصل از الگوریتم در انتقال مداری موردنظر، خروجی ها هم در قالب عناصر مداری اعتدالی و هم کپلری نمایش داده خواهد شد.



شکل ۴. معماری نرم افزاری

هدف اصلی که مانور افزایش مدار به مدار هم ارتفاع مدار ژئو است، در دو حالت با ارتفاع یکسان اما مقدار تراست متفاوت برای مدار اولیه به‌منظور بررسی عملکرد و دقت الگوریتم انجام می‌شود:

۱) مدار اولیه در ارتفاع ۱۰۰۰ کیلومتری از سطح زمین و تراست ۳ نیوتن

۲) مدار اولیه در ارتفاع ۱۰۰۰ کیلومتری از سطح زمین و تراست ۴ نیوتن

۱) حالت نخست مانور افزایش مدار

مشخصات مانور افزایش مدار در جدول ۳ ارائه شده است. سپس با ایجاد تغییر در فراپارامترهای الگوریتم، میزان تأثیر تغییرات آن‌ها بر فرآیند جستجو، بهره‌برداری و اهمیت‌دهی به پاداش‌های آینده بر مبنای ضریب‌های تنزل، نرخ یادگیری و بررسی خواهد شد. لازم است مقادیر نوسان برای عناصر مداری موردنظر ایجاد شود. هنگامی‌که تفاوت بین حالت مدار هدف و وضعیت مدار فعلی مطابق مقادیر نوسان مجاز مشخص شده در جدول ۲ باشد، اپیزود با موفقیت خاتمه می‌یابد.

جدول ۱. ضرایب تابع پاداش و مقادیر نوسان عناصر مداری

ضرایب تابع پاداش		مقدار نوسان مجاز عناصر مداری	
α_p	۱۰	p(km)	۱۰
α_f	۵	f(rad)	۰.۰۱
α_g	۵	g(rad)	۰.۰۱
α_h	۱	h(rad)	۰.۰۰۱
α_k	۱	k(rad)	۰.۰۰۱

جدول ۲. مشخصات حالت اول مانور افزایش مدار

	حالت اول مانور مداری	حالت دوم مانور مداری
جرم فضاپیما(kg)	۳۰۰۰	۳۰۰۰
مقدار تراست(N)	۳	۴
ارتفاع مدار اولیه(km)	۱۰۰۰	۱۰۰۰
ارتفاع مدار نهایی(km)	۳۵۷۸۶	۳۵۷۸۶

با درنظرگرفتن به مقدار نوسان مجاز عناصر مداری، جدول ۴ به وضوح نتایج به دست آمده از الگوریتم در انتقال تراست پایین را نشان می دهد که به طور منطقی قابل قبول هستند. برای هردو حالت مانور انتقال مداری، پارامترهای شبکه های عصبی مطابق جدول ۵ تنظیم شده است.

جدول ۳. مقادیر عناصر مداری حاصل از الگوریتم یادگیری تقویتی در حالت اول مانور افزایش مدار

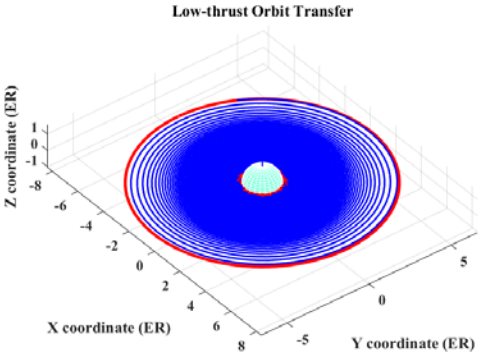
عناصر مداری کپلری			عناصر مداری اعتدالی			زمان انتقال(روز)
	اولیه	نهایی		اولیه	نهایی	
a(km)	7.3781e+03	4.2168e+04	p = a	7.3781e+03	4.2168e+04	۴۹.۴۱
e	0.0	0.0128	f(rad)	0	-0.0041	
i(deg)	0.0	0.0	g(rad)	0	0.0112	
ω(deg)	0.0	2.0577e+02	h(rad)	0	-4.9928e-09	
Ω(deg)	0.0	0.0	k(rad)	0	2.1745e-10	

جدول ۴. پارامترهای شبکه عصبی بازیگر منتقد در مانور افزایش مدار

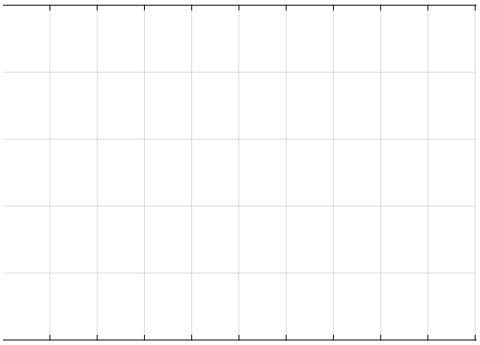
شبکه منتقد			شبکه بازیگر			
حالت سوم	حالت دوم	حالت اول	حالت سوم	حالت دوم	حالت اول	
۳۶۰	۳۶۰	۳۶۰	۲۸۰	۲۸۰	۲۸۰	تعداد گره لایه پنهان اول
۲۸۰	۲۸۰	۲۸۰	۲۰۰	۲۰۰	۲۰۰	تعداد گره لایه پنهان دوم
ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	تابع فعال ساز
۰.۰۱	۰.۰۱	۰.۰۱	۰.۰۰۱	۰.۰۰۱	۰.۰۰۱	نرخ یادگیری
۰.۹۵	۰.۹۵	۰.۹۰	۰.۹۵	۰.۹۵	۰.۹	ضریب تنزل

در شکل ۶، نمودار زرد، تعاملات عامل در محیط و پاداش‌های جمع آوری شده از عامل را در اپیزودهای متوالی نشان می‌دهد. نوسان اولیه به این دلیل است که عامل سعی می کند بین فضای محیط را بیشتر مورد جستجو قرار دهد تا دنباله‌ای مطلوب برای حالت‌ها در این مأموریت پیدا کند. در نهایت، فرآیند با برآوردن شرایط توقف و رسیدن به حالت‌ها به حالت‌های مداری نهایی مطلوب به پایان می‌رسد. شکل ۷، مدارات انتقال در انجام این مأموریت را نشان می دهد که مدار قرمز رنگ همان مدار ژئو است که با موفقیت فضاپیما به آن رسیده است. با توجه به فرضیات انتقال مداری درنظرگرفته شده، بردار تراست وارده در این انتقال مماس بر مسیر حرکت فضاپیما به دست آمده است که در شکل ۸ نشان داده شده است. تغییرات نیم قطر اصلی بیضی و خروج از مرکز مدارات انتقال نیز به ترتیب در شکل های ۹ و ۱۰ نمایش داده شده است که می‌توان دید به ارتفاع مداری و خروج از مرکز مدار ژئو دست پیدا کرده است.

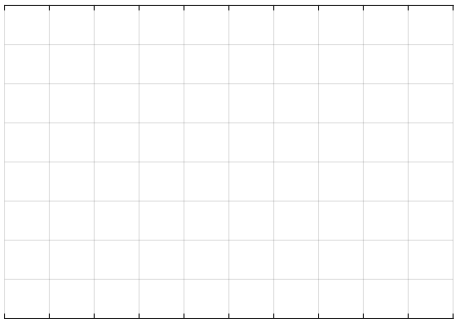
شکل ۶. تابع پاداش در حالت اول مانور افزایش مدار



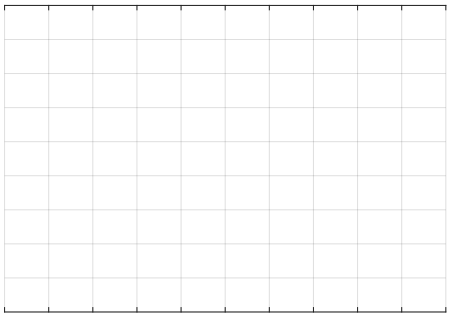
شکل ۷. مدارات انتقال در حالت اول مانور افزایش مدار



شکل ۸. پروفیل زاویه انحراف تراستر در حالت اول مانور افزایش مدار



شکل ۹. تغییرات نیم قطر اصلی در حالت اول مانور مداری



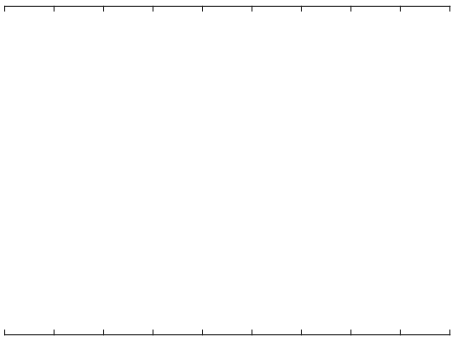
شکل ۱۰. تغییرات خروج از مرکز مدارات در حالت اول مانور مداری

۲) حالت دوم مانور افزایش مدار

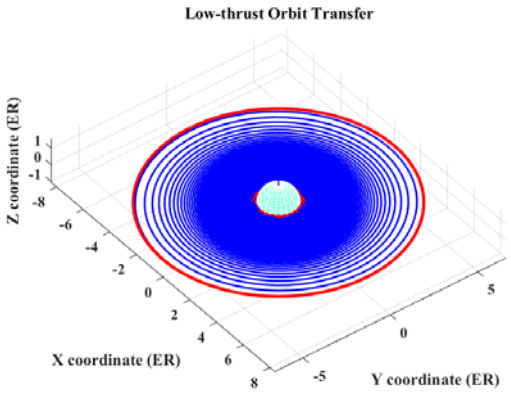
در حالت دوم، همانطور که در جدول ۵ مشخص است، مقدار ضـریب تنزل به ۰.۹۵ افزایش داده می شـود تا در ضـمن انجام مأموریت، تأثیر پاداش های آینده بر عملکرد الگوریتم نیز بررسـی شود. جدول ۶ به وضوح نتایج به دست آمده از الگوریتم در انتقال تراسـت پایین مربوطه را نشـان می دهد که به طور منطقی قابل قبول هستند. نمودار پاداش حالت دوم در شکل ۱۱ قابل مشاهده است. با افزایش این ضریب، تمرکز و سرعت عامل بر پاداش های آینده بیشـتر شـده اسـت و با توجه به اینکه در حالت اول یکبار عامل آموزش دیده اسـت، فرآیند همگرایی و جمع آوری پاداش مثبت که به معنی مطلوب بودن حلت های به وجود آمده برای این انتقال مداری است، با سرعت و دقت بهتری پیش رفته است. در نهایت عامل با جمع آوری حداکثر پاداش مثبت با موفقیت به شـرایط مدار ژئو دسـت پیدا می کند. در شـکل ۱۲ مشـاهده می شـود که فضـاپیما با موفقیت در مدار مطلوب قرار گرفته اسـت. منحنی زاویه تراستر نیز نشان می دهد که در تمام مسیر مماس بر مسیر حرکت ماهواره اعمال شده است تا در نهایت شرایط نیم قطر اصلی و خروج از مرکز مدار ژئو مطابق شـکل های ۱۴ و ۱۵ برآورده شود. مقدار خروج از مرکز در مدار نهایی همان‌طور که در شـکل ۱۴ مشخص است به مقدار ۰/۰۱۱۷۶ می‌رسد و همچنین نوسـان آن درنهایت کنترل می‌شـود که با توجه به تابع پاداش و هدف تعریف‌شده برای مسئله قابل‌قبول می‌باشد.

جدول ۵. مقادیر عناصر مداری حاصل از الگوریتم یادگیری تقویتی در حالت دوم مانور افزایش مدار

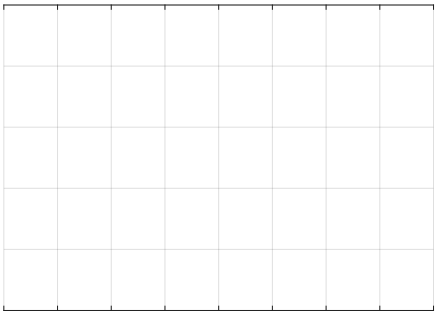
عناصر مداری کپلری			عناصر مداری اعتدالی			زمـــان انتقال(روز)
	اولیه	نهایی		اولیه	نهایی	
a(km)	7.3781e+03	4.2167e+04	p = a	7.3781e+03	4.2168e+04	۳۷.۱۰
e	0.0	0.0117	f(rad)	0.0	-0.0041	
i(deg)	0.0	0.0	g(rad)	0.0	0.0112	
ω(deg)	0.0	1.7074e+02	h(rad)	0.0	-4.9928e-09	
Ω(deg)	0.0	1.7844e+02	k(rad)	0.0	2.1745e-10	



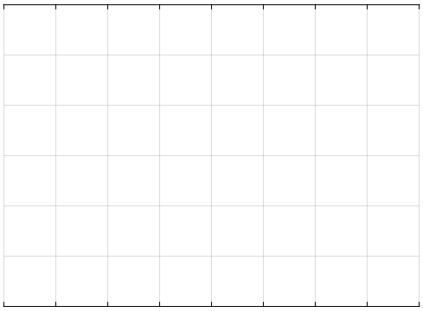
شکل ۱۱. نمودار پاداش در حالت دوم مانور افزایش ارتفاع مدار



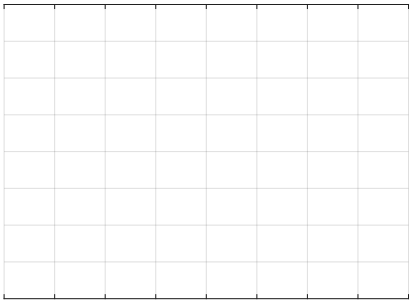
شکل ۱۲. مدارات انتقال در حالت دوم مانور افزایش مدار



شکل ۱۳. پروفیل زاویه انحراف تراستر در حالت دوم مانور افزایش مدار



شکل ۱۴. تغییرات خروج از مرکز در حالت دوم مانور افزایش مدار



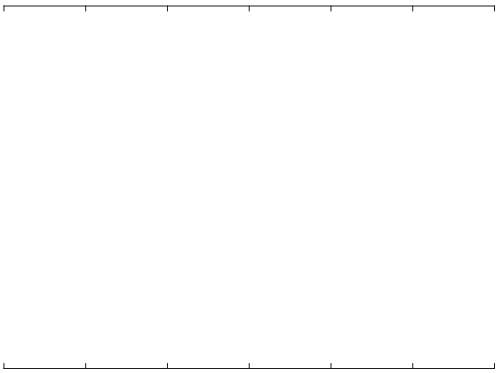
شکل ۱۵. تغییرات نیم قطر اصلی در حالت دوم مانور افزایش مدار

۳) حالت سوم مانور افزایش مدار

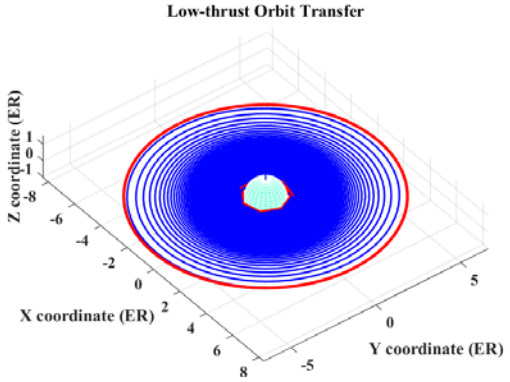
در حالت سوم با هدف بررسی دقت و صحت الگوریتم در نتایج، ارتفاع مداری به ۱۵۰۰ کیلومتری از سطح زمین افزایش یافته است. مشخصات شبکه ها و فراپارمترهای الگوریتم طبق جدول ۵ مانند حالت دوم تنظیم شده است. مقادیر عناصر مداری مدار نهایی در جدول ۶ نمایش داده شده است. با توجه به اینکه عامل در دو حالت قبل با دینامیک مسئله یادگیری را انجام داده است، در این جا همانطور که از شکل ۱۶ مشخص است عامل با استفاده از تعاملات قبلی، با دریافت پاداش های مثبت و بدون داشتن تعاملات و پاداش منفی(شیب مثبت) توانسته است فرآیند یادگیری را با موفقیت و دقت بهتری نسبت به تو دو حالت قبل به انجام رساند و دنباله ای از حالت ها را برای مسیر تراست پایین مورد نظر مطابق شکل ۱۷ ترسیم نماید. مقادیر تغییرات زاویه تراست، نیم قطر اصلی و خروج از مرکز مدار های انتقال تراست پایین در شکل های ۱۸ تا ۲۰ نمایش داده شده است.

جدول ۶. مقادیر عناصر مداری حاصل از الگوریتم یادگیری تقویتی در حالت سوم مانور افزایش مدار

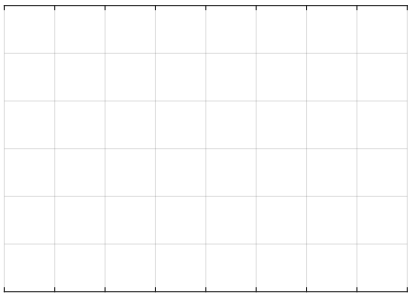
زمان انتقال(روز)	عناصر مداری اعتدالی			عناصر مداری کپلری		
	نهایی	اولیه		نهایی	اولیه	
۳۵.۰۵	4.2168e+04	7.8781e+03	p = a	4.2167e+04	7.8781e+03	a(km)
	-0.0041	0.0	f(rad)	0.0117	0.0	e
	0.0112	0.0	g(rad)	0.0	0.0	i(deg)
	-4.9928e-09	0.0	h(rad)	1.7074e+02	0.0	ω(deg)
	2.1745e-10	0.0	k(rad)	1.7844e+02	0.0	Ω(deg)



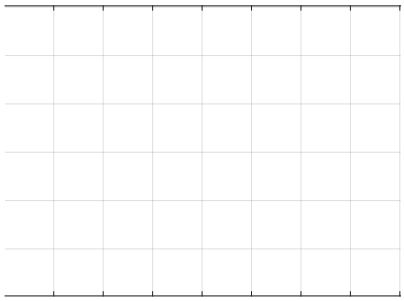
شکل ۱۶. نمودار پاداش در حالت سوم مانور افزایش ارتفاع مدار



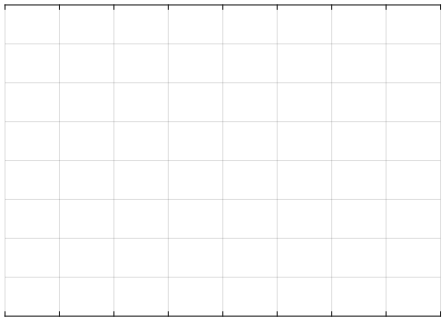
شکل ۱۷. تغییرات نیم قطر اصلی در حالت دوم مانور افزایش مدار



شکل ۱۸. پروفیل زاویه انحراف تراستر در حالت سوم مانور افزایش مدار



شکل ۱۹. تغییرات خروج از مرکز در حالت سوم مانور افزایش مدار



شکل ۲۰. تغییرات نیم قطر اصلی در حالت سوم مانور افزایش مدار

۵- نتیجه‌گیری

در مطالعات گذشته، استفاده از الگوریتم های یادگیری ماشین به خصوص یادگیری تقویتی در طراحی مسیر فضاپیمای تراست پایین در مأموریت های مانور مداری مانند ملاقات مداری، انتقال مداری و غیره مورد توجه قرار نگرفته است. در این مقاله، دینامیک مسئله طراحی مسیر فضاپیمای تراست پایین تحت قالب معادلات دیفرانسیل اعتدالی بر حسب عناصر مداری اعتدالی برای مدارهای دایروی هم صفحه بیان شد. با توجه به اینکه در مانور انتقال مداری هدف اصلی رسیدن به ارتفاع و خروج از مرکز مدار موردنظر است، تابع پاداش به صورت مجموع وزن دهی شده عناصر مداری اعتدالی بیان شد. سپس الگوریتم شبکه بازیگر-منتقد برای سه حالت مانور افزایش مدار با دو نیروی تراست متفاوت به منظور نمایش عملکرد صحیح الگوریتم پیاده سازی شد. اثر تغییر فراپارامتر ضریب تنزل نیز بر روند یادگیری عامل به منظور تحلیل حساسیت الگوریتم نشان داده شد. حالت سوم افزایش ارتفاع از ۱۵۰۰ کیلومتری به مدار ژئو برای نمایش دقت و مقایسه نتایج به دست آمده در حالت های قبل پیاده‌سازی شد. در نهایت با بررسی نتایج به دست آمده از سه شبیه سازی انجام شده، عملکرد این نوع از الگوریتم های یادگیری ماشین در انجام مأموریت های انتقال مداری به خوبی نمایش داده شد. مقادیر مهم عناصر مداری مانند نیم قطر اصلی و خروج از مرکز مدار ها که در مانور افزایش مدار لئو به ژئو اهمیت بالایی دارد در هر سه حالت در محدوده مقادیر صحیح تعیین شده و منطقی خود قرار دارند. در حالت اول عامل به دلیل عدم آگاهی از دینامیک مسئله نیاز به طی کردن اپیزود ها و تعامل بیشتر با محیط دارد که این موضوع در نمودار پاداش مربوط به آن مشاهده و توصیف شد. در

ادامه حالت های دوم و سوم عامل با استفاده از تجربیات قبلی خود بهتر و سریعتر توانسته است به هدف تعیین شده با استفاده از تابع پاداش تعریف شده برسد که به خوبی در شکل های نمودار پاداش و عناصر مداری هریک از این حالت ها قابل مشاهده است. از طرفی نیز می توان دید که با تغییر شرایط اولیه مانند نیروی تراست و ضرایب تنزل و یادگیری الگوریتم دقت این روش در مقایسه با روش های کنترل بهینه [۱۶] بیشتر است و حساسیت آن به این تغییرات نیز کمتر می باشد. همچنین روش‌های یادگیری ماشین جستجو های جهانی را برای یافتن راه حل انجام می دهند که در مقایسه با سایر روش های بیان شده در مرور بر ادبیات یک مزیت محسوب می شود.

جدول ۷. مقایسه دو رویکرد کنترل بهینه و یادگیری تقویتی[۱۷]

یادگیری تقویتی	کنترل بهینه
یک بهینه سازی تصادفی	یک بهینه سازی قطعی
زمان اجرای بسیار سریع برای سیاست آموزش دیده	زمان اجرا نامحدود به جز موارد خاص مانند محدودیت های تحدب
هیچ محدودیتی در نمایش پویایی مشکل وجود ندارد	دینامیک باید تحت یک معادله دیفرانسیل معمولی نشان داده شود.
حلقه بسته (هدایت و کنترل یکپارچه)	حلقه باز (برای ردیابی مسیر بهینه به یک کنترلر نیاز دارد)

مزیت اصلی و قابل توجه الگوریتم های یادگیری ماشین که باعث شده است تمرکز محققان به استفاده بیشتر از این روش ها معطوف شود این است که این نوع از الگوریتم ها بعد از آموزش دیدن بر روی دینامیک و محیط مسئله می توانند برای شرایط اولیه و قیود مختلفی که بر مأموریت حاکم می شود به کارگرفته شوند و دیگر نیاز به شبیه سازی و حل مجدد و کامل مسئله و تنظیم پارامترهای آن نیست که این خاصیت با بیان خودکار و خودران بودن این دسته از الگوریتم ها شناخته می شود.

از سوی دیگر، یکی از چالش‌های مهمی که روش یادگیری تقویتی با آن مواجه و بخش مهم الگوریتم است، تعیین تابع پاداش متناسب با دینامیک مسئله مربوطه است. چراکه زمانی که تابع پاداش خیلی پراکنده و دور از هدف اصلی مسئله باشد نه‌تنها به نتایج دقت کافی را ندارد بلکه زمان زیادی نیز صرف یک مسیر اشتباه خواهد شد. همچنین در زمینه بهینه‌سازی، زمانی که اهداف مسئله از یک به دو یا بیشتر تبدیل می‌شود، تعریف و تنظیم کردن تابع پاداش و ضرایب آن و همچنین پارامترهای الگوریتم دشوار و پیچیده می‌شود که نیازمند تسلط کامل بر دینامیک مسئله و عملکرد الگوریتم در این دسته از مسائل است.

همچنین به عنوان یک کار مستقل و جذاب بعدی، می توان در کنار مسئله طراحی مسیر، پارامتر زمان و یا مصرف سوخت را نیز در معادلات و فرآیند پیاده سازی الگوریتم در نظر گرفت و با تنظیم تابع پاداش و شبکه های عصبی الگوریتم بازیگر منتقد این موضوع را نیز مورد بررسی و تحلیل قرار داد که به نوبه خود یک نیاز مهم در عملیات های واقعی مأموریت های مداری به حساب می آید.

تعارض منافع

هیچ‌گونه تعارض منافع توسط نویسندگان بیان نشده است.

مراجع

[1] Robert E. Pritchett, “Numerical methods for low-thrust trajectory optimization”, Purdue University West Lafayette, Indiana, August 2016.

[2] J. B. Caillau, J. Gergaud, and J. Noailles. 3D geosynchronous transfer of a satellite: Continuation on the thrust. Journal of Optimization Theory and Applications, 118 (3):541–565, September 2003.

[3] Mischa Kim, “Continuous Low-Thrust Trajectory Optimization: Techniques and Applications”,Doctor of Philosophy Dissertation in Aerospace Engineering ,April 18, 2005

[4] Ryan P. Russell, “Primer Vector Theory Applied to Global Low-Thrust Trade Studies”, California Institute of Technology, Pasadena, California 91109 , JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS, 2007 , DOI: 10.2514/1.22984

[5] Weipeng Li Hai Huang,"Optimization of low-thrust transfers to libration point orbits", Aircraft Engineering and Aerospace Technology: An International Journal (2016), Vol. 88 Iss 1 pp. 16 – 23

[6] Mauro Pontani · Bruce Conway, “Optimal Low-Thrust Orbital Maneuvers via Indirect Swarming Method” Accepted: 29 October 2013, Springer Science+Business Media New York 2013, DOI 10.1007/s10957-013-0471-9

[7] P. Enright and B. A. Conway, “Discrete approximations to optimal trajectories using direct transcription and Nonlinear Programming,” Journal of Guidance Control and Dynamics, Vol. 15, 09 1992, 10.2514/3.20934.

[8] Jonathan D. Aziz, Jeffrey S. Parker, Daniel J. Scheeres, Jacob A. Englander, “Low-Thrust Many-Revolution Trajectory Optimization via Differential Dynamic Programming and a

[19] Navaee, M., & Sanati, M. (2013). Optimal Impulsive Orbital 3D Maneuver with or without Time Constraint. *Amirkabir Journal of Mechanical Engineering*, *44*(2), 53-69.
[20] Fakoor, M., Sadeghi, S., & Bakhtiari, M. (2020). Investigation of low thrust optimal orbital transfer from LEO to GEO considering circular orbits. *The Journal of the Astronautical Sciences*, *67*, 77-97.

Sundman Transformation", American Astronautical Society, part of Springer Nature 2018
[9] S. N. Williams and V. Coverstone-Carroll. Mars missions using solar electric propulsion. Journal of Spacecraft and Rockets, 37(1):71–77, January–February 2000.
[10] Zhenbo Wang and Michael J. Grant, "Minimum-Fuel Low-Thrust Transfers for Spacecraft: A Convex Approach" School of Aeronautics and Astronautics, Purdue University, West Lafayette, Indiana, 47907-2045, USA, DOI 10.1109/TAES.2018.2812558, IEEE Transactions on Aerospace and Electronic Systems 2018
[11] Daniel S. Kolosa, "A Reinforcement Learning Approach to Spacecraft Trajectory Optimization", Western Michigan University, Dissertations. 3542,2019
[12] M.P. van Hoorn**,** "OPTIMIZING AIR-TO-AIR MISSILE GUIDANCE USING REINFORCEMENT LEARNING", Master of Science thesis in Aerospace Engineering, Delft University of TechnologyMarch 26, **2019**
[13] T.A.H. Kranen, "Low-Thrust Gravity Assist Trajectory Optimisation using Evolutionary Neurocontrol", Dissertation TU Delft Aerospace Engineering, 2019**.**
[14] Howard D. Curtis, **"***Orbital Mechanics for Engineering Students***"**, Third Edition, season 6, p.g 344, 2014
[15] Richard S. Sutton and Andrew G. Barto,Reinforcement Learning:An Introduction, MIT Press, Cambridge, MA, 1998A Bradford Book
[16] M. Navabi, E. Meshkinfam, "Space low-thrust trajectory optimization utilizing numerical techniques, a comparative study," 2013 6th International Conference on Recent Advances in Space Technologies (RAST), IEEE, 2013, pp. 303-307, doi: 10.1109/RAST.2013.6581222.
[17] Brian Gaudet, Roberto Furfaro, Richard Linares, "Reinforcement Learning for Angle-Only Intercept Guidance of Maneuvering Targets", Aerospace Science and Technology,Volume 99, April 2020
[18] Navabi, M., & Sabatifar, M. (2010). Optimal Impulsive Maneuver Between Elliptical Coplanar-Noncoaxial Orbits. *Space Science and Technology*, *3*(1), 67-74.

[i] Orbital Rendezvous
[ii] orbit maintenance
[iii] specific impulse(Isp)
[iv] Two-boundary value problem
[v] Pontryagin
[vi] Adaptive Simulated Annealing
[vii] Particle Swarm Optimization
[viii] Sundman
[ix] differential dynamic programming
[x] Genetic Algorithm
[xi] Artificial Bee Colony (ABC) algorithm
[xii] initial value problems
[xiii] Reinforcement Learning(RL)
[xiv] Actor-Critic networks
[xv] Supervised Learning (SL)
[xvi] Unsupervised Learning (UL)
[xvii] Agent
[xviii] Policy
[xix] reward value
[xx] Value
[xxi] expected return
[xxii] discount factor
[xxiii] policy gradients
[xxiv] Deep Deterministic Policy Gradients (DDPG)
[xxv] state-value function
[xxvi] temporal difference error
[xxvii] Gauss's Variational equations
[xxviii] equinoctial orbital elements
[xxix] model-free
[xxx] Exploration
[xxxi] Exploitation