ISSN: 2476-6909; Modares. Mech. Eng. *2026;26(05):383-399.*

Modares Mechanical Engineering

DOI: 10.48311/mme.2025.117147.82868
journal homepage: mme.modares.ac.ir

Design and Simulation of an Autonomous System Using Reinforcement Learning for Articulated Vehicle Pose Adjustment

Moein Qanbari Senjegani, Majid Sadedel*

Mechanical Engineering Department, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article Type
Original Research

Article History
Received: October 19, 2025
Revised: November 16, 2025
Accepted: November 24, 2025
ePublished: March 04, 2026

ABSTRACT

Automatic posture adjustment is a key application in the domain of intelligent vehicles, playing a fundamental role in enhancing safety and optimizing vehicle maneuvering operations. Articulated Vehicles (AVs), due to their high degrees of freedom and the complex, nonlinear dynamics resulting from the joint between the tractor and trailer, present a more challenging control problem compared to rigid vehicles. The objective of this research is to design and simulate an automatic control system for articulated vehicle posture adjustment utilizing a Deep Reinforcement Learning (DRL) .framework. This system can also serve as a foundation for more advanced applications, such as autonomous parking In this study, the precise modeling of articulated vehicle dynamics and the jackknifing phenomenon was initially carried out. The developed model was validated using the specialized software, TruckSim. Subsequently, to reduce computational complexity, the learning process was segmented into two distinct phases: maneuver preparation and final posture adjustment. For training the intelligent agent, Deep Deterministic Policy Gradient (DDPG) and Twin Delayed DDPG (TD3) algorithms were employed, which were optimized with neural networks comprising three to five hidden layers. Evaluation results indicated that the TD3 algorithm, owing to its superior ability to maintain the stability of the learning process, outperformed DDPG. Ultimately, the proposed control system, with the optimal structure for each phase, achieved success rates of 96.6% in the preparation phase and 94.6% in the final adjustment phase, thereby confirming the high efficiency and reliability of the DRL-based system in addressing the control challenges of articulated vehicles.

Keywords: Motion Planning, Articulated Vehicle, Reinforcement Learning, Deep Deterministic Policy Gradient (DDPG), Twin Delayed Deep Deterministic Policy Gradient (TD3)

How to cite this article
Qanbari Senjegani M, Sadedel M., Design and Simulation of an Autonomous System Using Reinforcement Learning for Articulated Vehicle Pose Adjustment. Modares Mechanical Engineering; *2026;26(05):383-399.*

*Corresponding author's email: majid.sadedel@modares.ac.ir
*Corresponding ORCID ID: 0000-0002-0285-8460

شاپا: ۷۹۰۹-۲۴۷۶؛ مهندسی مکانیک مدرس. ۳۹۹-۳۸۳:(۰۵)۲۶؛۱۴۰۵

مهندسی مکانیک مدرس

DOI: 10.48311/mme.2025.117147.82868

صفحه اصلی مجله: mme.modares.ac.ir

طراحی و شبیه‌سازی سیستم خودکار مبتنی بر یادگیری تقویتی برای تنظیم موقعیت خودروهای مفصلی

معین قنبری سنجگانی، مجید ساده‌دل*

گروه مهندسی مکانیک، دانشگاه تربیت مدرس، تهران، ایران

اطلاعات مقاله

نوع مقاله

مقاله پژوهشی

تاریخچه مقاله

دریافت: ۱۴۰۴/۰۷/۲۷

بازنگری: ۱۴۰۴/۰۸/۲۵

پذیرش: ۱۴۰۴/۰۹/۰۳

ارائه آنلاین: ۱۴۰۴/۱۲/۱۳

چکیده

تنظیم موقعیت خودکار یکی از کاربردهای کلیدی در حوزه خودروهای هوشمند است که در افزایش ایمنی و بهینه‌سازی عملیات مانور وسایل نقلیه نقش اساسی دارد. خودروهای مفصلی به دلیل برخورداری از درجه آزادی بالا و دینامیک پیچیده و غیرخطی ناشی از مفصل بین کشنده و تریلر، کنترل دشوارتری نسبت به خودروهای صلب دارند. هدف از این پژوهش، طراحی و شبیه‌سازی یک سیستم کنترل خودکار برای تنظیم موقعیت خودروهای مفصلی با بهره‌گیری از چارچوب یادگیری تقویتی عمیق است. این سیستم می‌تواند به عنوان زیربنایی برای کاربردهای پیشرفته‌تر نظیر پارک خودکار مورد استفاده قرار گیرد. در این مطالعه، ابتدا مدل‌سازی دقیق دینامیک حرکت خودروی مفصلی و پدیده قیچی‌شدن انجام گرفت و مدل توسعه‌یافته با استفاده از نرم‌افزار تراک‌سیم صحت‌سنجی شد. سپس، فرآیند یادگیری به منظور کاهش پیچیدگی محاسباتی، به دو فاز مجزا (آماده‌سازی مانور و تنظیم نهایی موقعیت) تقسیم شد. برای آموزش عامل هوشمند در این دو فاز، از الگوریتم‌های گرادیان سیاست قطعی عمیق (DDPG) و گرادیان سیاست قطعی عمیق دوگانه تأخیری (TD3)، که با شبکه‌های عصبی شامل ۳ تا ۵ لایه نهان بهینه‌سازی شدند، استفاده گردید. نتایج ارزیابی‌ها نشان داد که الگوریتم TD3، به دلیل توانایی بالاتر در حفظ پایداری فرآیند یادگیری، عملکرد بهتری نسبت به DDPG ارائه می‌دهد. در نهایت، سیستم کنترل پیشنهادی با ساختار بهینه برای هر فاز، موفقیت‌هایی به ترتیب ۹۶.۶٪ در فاز آماده‌سازی و ۹۴.۶٪ در فاز تنظیم نهایی را کسب کرد که کارایی بالا و قابلیت اطمینان سیستم مبتنی بر یادگیری تقویتی عمیق در مواجهه با چالش‌های کنترلی خودروهای مفصلی را تأیید می‌کند.

کلیدواژه‌ها: برنامه‌ریزی حرکت، خودروی مفصلی، یادگیری تقویتی، گرادیان سیاست قطعی عمیق

نحوه ارجاع به این مقاله

قنبری سنجگانی معین، ساده‌دل مجید، طراحی و شبیه‌سازی سیستم خودکار مبتنی بر یادگیری تقویتی برای تنظیم موقعیت خودروهای مفصلی، مهندسی مکانیک مدرس. ۳۹۹-۳۸۳:(۰۵)۲۶؛۱۴۰۵

*پست الکترونیکی نویسنده عهده‌دار مکاتبات: majid.sadedel@modares.ac.ir

*شناسه ارکید نویسنده عهده‌دار مکاتبات: 0000-0002-0285-8460

۱- مقدمه

پیشرفت‌های شگرف در حوزه سیستم‌های حمل و نقل هوشمند (Intelligent Transportation Systems - ITS)، مهندسی وسایل نقلیه را به طور بنیادی دگرگون ساخته است. این تحولات، که با هدف بهبود مستمر ایمنی جاده‌ای، کاهش مصرف انرژی و افزایش کارایی عملیاتی حمل و نقل صورت پذیرفته‌اند، نتایج قابل ملاحظه‌ای در پی داشته‌اند. در کانون این پیشرفت‌ها، سیستم‌های کمک‌راننده پیشرفته (Advanced Driver Assistance Systems - ADAS) نقش محوری ایفا می‌کنند؛ این سیستم‌ها با ادغام حسگرهای پیشرفته، عملگرهای دقیق و الگوریتم‌های کنترلی پیچیده وظایف حیاتی رانندگی از جمله جلوگیری از برخورد، حفظ خط و تنظیم موقعیت خودکار (Automatic Posture Adjustment) را تسهیل نموده‌اند.

با این وجود، تنظیم موقعیت خودکار، به دلیل محدودیت‌های فضایی، حضور موانع پویا و الزامات مانورهای با دقت بالا، همچنان به عنوان یکی از چالش‌برانگیزترین قابلیت‌های وسایل نقلیه هوشمند قلمداد می‌شود. اگرچه مطالعات گسترده‌ای برای خودروهای سواری انجام شده است، اما وسایل نقلیه مفصلی (Articulated Vehicles - AVs) نظیر تریلرهای کشنده و اتوبوس‌های مفصلی کمتر مورد توجه قرار گرفته‌اند. این وسایل نقلیه از چندین بدنه صلب تشکیل شده‌اند که از طریق مفاصل مکانیکی به یکدیگر متصل شده‌اند، در نتیجه، درجات آزادی اضافی و رفتار سینماتیکی غیرخطی (Nonlinear Kinematic Behavior) پیچیده‌ای را معرفی می‌کنند. بنابراین تنظیم موقعیت این دسته از وسایل نقلیه از نظر فنی بسیار دشوار بوده و ریسک‌های ایمنی ذاتی مانند پدیده قیچی‌شدن (Jackknifing) و عدم هم‌ترازی جانبی را به همراه دارد. با توسعه یک سیستم کارآمد برای حل این چالش، می‌توان بستر لازم برای کاربردهای پیشرفته‌تر نظیر پارک خودکار این وسایل را فراهم نمود.

تنظیم موقعیت متداول وسایل نقلیه مفصلی به شدت به تخصص و تجربه اپراتور و کنترل دستی وابسته است. این وابستگی اغلب به ناکارآمدی عملیاتی، افزایش ریسک برخورد و طولانی شدن زمان مانورها در شرایط عملیاتی منجر می‌شود. این چالش‌ها به ویژه در محیط‌هایی نظیر مراکز لجستیکی، سایت‌های صنعتی و انبارها، که فضای محدود بوده و موقعیت‌یابی دقیق وسایل نقلیه برای فرایندهای بارگیری و تخلیه امری ضروری است، حیاتی می‌باشند. این ضرورت، لزوم فوری توسعه سیستم‌های کنترلی هوشمند را برجسته می‌سازد که قادر به برنامه‌ریزی و اجرای ایمن، کارآمد و قابل اعتماد مانورهای تنظیم موقعیت برای وسایل نقلیه مفصلی باشند.

برای پرداختن مؤثر به چالش‌های فوق‌الذکر، یادگیری تقویتی (Reinforcement Learning - RL) به عنوان یک روش کنترلی امیدوارکننده برای وسایل نقلیه خودکار مطرح شده است. برخلاف استراتژی‌های کنترل کلاسیک که متکی بر مدل‌سازی صریح و قواعد از پیش تعریف‌شده هستند، یادگیری تقویتی یک عامل را قادر می‌سازد تا اقدامات بهینه را از طریق تعامل مستقیم با محیط خود بیاموزد. طبیعت داده‌محور و اکتشافی یادگیری تقویتی، آن را برای مدیریت دینامیک‌های پیچیده و تصمیم‌گیری تحت عدم قطعیت بسیار مناسب می‌سازد. به طور خاص، یادگیری تقویتی عمیق، که قدرت یادگیری تقویتی را با توانایی نمایش ویژگی‌های شبکه‌های عصبی عمیق ادغام می‌کند، توانایی بالایی در مدیریت فضاهای حالت با ابعاد بالا نشان داده و به رویکردی کاربردی برای توسعه استراتژی‌های کنترل وسایل نقلیه مفصلی تبدیل شده است.

این مقاله به طراحی، توسعه و شبیه‌سازی یک سیستم تنظیم موقعیت خودکار مبتنی بر یادگیری تقویتی عمیق می‌پردازد که به طور خاص برای مدیریت سینماتیک پیچیده وسایل نقلیه مفصلی تدوین شده است. به منظور مدیریت پیچیدگی ذاتی فرایند تنظیم موقعیت، چارچوب پیشنهادی از یک استراتژی مدولار مبتنی بر فاز استفاده می‌کند که وظیفه را به دو مرحله متوالی تقسیم می‌کند: (۱) موقعیت‌یابی و آماده‌سازی به سمت ناحیه هدف و (۲) هم‌ترازی نهایی با محور هدف آموزش مدل با استفاده از الگوریتم‌های پیشرفته یادگیری تقویتی عمیق، یعنی گرادیان سیاست قطعی عمیق (DDPG) و گرادیان سیاست قطعی عمیق دوقلو تأخیری (TD3)، صورت پذیرفته است. همچنین، توابع پاداش سفارشی و معماری‌های شبکه عصبی به منظور تضمین یادگیری ایمن، پایدار و کارآمد معرفی شده‌اند.

سهم کلیدی این پژوهش شامل موارد زیر است:

- ادغام موفق سینماتیک پیچیده وسایل نقلیه مفصلی، از جمله پدیده قیچی‌شدن، در یک چارچوب کنترل مبتنی بر یادگیری تقویتی عمیق؛ این نوآوری با مدل‌سازی دقیق دینامیک غیرخطی و ادغام آن در محیط یادگیری، چالش‌های ایمنی ذاتی خودروهای مفصلی را حل می‌کند.
- ارائه یک روش‌شناسی آموزش مرحله به مرحله ساختار یافته برای تسهیل همگرایی و بهبود کارایی در مسئله پیچیده تنظیم موقعیت؛ این رویکرد با تقسیم فرآیند به فازهای آماده‌سازی و تنظیم نهایی، پیچیدگی محاسباتی را کاهش می‌دهد و نسبت به روش‌های تک‌مرحله‌ای قبلی کارآمدتر است.
- انجام تحلیل مقایسه‌ای عملکرد دقیق بین الگوریتم‌های DDPG و TD3 به منظور شناسایی کارآمدترین استراتژی کنترل برای این وظیفه تخصصی؛ نتایج نشان می‌دهد TD3 با حفظ پایداری یادگیری، واریانس Q-value را کمتر می‌کند و در سناریوهای واقعی پایدارتر عمل می‌کند، که این تحلیل در کارهای مشابه وجود نداشته است.

در نهایت، این تحقیق با هدف پیشبرد استقرار عملی سیستم‌های تنظیم موقعیت خودکار در سناریوهای پیچیده و واقعی شامل وسایل نقلیه حمل و نقل مفصلی است.

۲- پیشینه تحقیق

تنظیم موقعیت خودکار خودروهای مفصلی یکی از کاربردهای کلیدی در وسایل نقلیه هوشمند محسوب می‌شود که مستقیماً به افزایش راحتی راننده، کارایی ترافیکی و بهبود ایمنی کمک می‌کند. در طول سال‌ها، روش‌ها و سیستم‌های گوناگونی برای پشتیبانی از مراحل مختلف فرآیند تنظیم موقعیت توسعه‌یافته‌اند، شامل ادراک، تصمیم‌گیری و کنترل.

مسئله تنظیم موقعیت، به ویژه در محیط‌های پویا و به طور گسترده با روش‌های مبتنی بر قواعد و بهینه‌سازی مورد بررسی قرار گرفته است. تکنیک‌های کلاسیک مانند استنتاج فازی [۱]، پارامترسازی مسیر چندجمله‌ای با بهینه‌سازی الگوریتم ژنتیک [۲] و برنامه‌ریزی مسیر با استفاده از الگوریتم کلونی مورچه ترکیبی شده با منحنی‌های بزیه [۳] برای تولید مسیرهای صاف و قابل اجرا به کار گرفته شده‌اند. همچنین برخی پژوهش‌ها بر بهینه‌سازی مبتنی بر چارچوب تکرارشونده سبک‌وزن برای پارک خودکار با موانع نامنظم [۳] و مدل‌سازی مبتنی بر نرم‌افزار برای پارک خودروهای سنگین مفصلی [۴، ۵] بر جنبه‌های هندسی و ایمنی تمرکز کرده‌اند. با این حال، این روش‌ها اغلب خاص وظیفه هستند، نیاز به تنظیم دستی دارند و فاقد سازگاری با سناریوهای غیرمنتظره می‌باشند.

یادگیری تقویتی، به ویژه یادگیری تقویتی عمیق، در این حوزه کاربرد فراوانی دارد. یادگیری تقویتی به عاملان اجازه می‌دهد تا استراتژی‌های تنظیم موقعیت را به طور خودکار از تعامل با محیط بیاموزند و وابستگی به قواعد تعریف‌شده را کاهش دهند. ژانگ و همکاران [۶] یک چارچوب ترکیبی یادگیری تقویتی با استفاده از DQN برای پارکینگ موازی پیشنهاد کردند. سانگ و همکاران [۷] از یادگیری تقویتی عمیق مبتنی بر مدل با جستجوی درخت مونت کارلو برای مانورهای پیچیده در فضاهای تنگ بهره بردند. این مدل‌ها کارایی یادگیری بهبودیافته و نرخ موفقیت بالاتری نسبت به برنامه‌ریزی حرکت سنتی نشان دادند. وانگ و همکاران [۸] الگوریتم ASAP-RL را معرفی کردند که با پارامترسازی مهارت‌های حرکتی و بهره‌گیری از دانش متخصص، کارایی یادگیری تقویتی را در برنامه‌ریزی حرکت خودروهای خودران در ترافیک متراکم افزایش می‌دهد. سعیدی و ساده‌دل [۹] پیاده‌سازی الگوریتم ناوبری مبتنی بر رفتار را بر روی ربات موبایل چهارچرخ فرمان‌پذیر بررسی کردند که مستقیماً به ناوبری در وسایل نقلیه چرخ‌دار مربوط می‌شود. در مطالعه‌ای توسط ملائی و همکاران [۱۰]، یک چارچوب چندلایه برای رانندگی خودکار در بزرگراه با استفاده از یادگیری تقویتی توزیعی ارائه شد که از الگوریتم‌هایی مانند FQF و IQN برای بهبود تصمیم‌گیری در شرایط ترافیکی پیچیده بهره می‌برد. ریزه‌وندی و همکاران [۱۱] نیز به تصمیم‌گیری خودروهای خودران در بزرگراه‌ها پرداختند و نشان دادند که الگوریتم‌های یادگیری تقویتی عمیق می‌توانند رفتار خودرو را در تغییرات سریع ترافیک مدیریت کنند و امنیت و کارایی را افزایش دهند. ادریسی و همکاران [۱۲] مسیریابی هوشمند وسایل نقلیه در شبکه ترافیکی پویا را با یادگیری تقویتی بررسی کردند و نشان دادند که این روش می‌تواند ترافیک را متعادل و زمان سفر را کاهش دهد.

خودروهای مفصلی، مانند تریلرهای کشنده و اتوبوس‌ها، چالش‌های اضافی به دلیل سینماتیکی غیرخطی، دینامیک مفصلی و خطر قیچی‌شدن ایجاد می‌کنند. تحقیقات اولیه این مسائل را از طریق مدل‌سازی اضافی و طراحی سیستم کنترل مورد بررسی قرار دادند. برای نمونه، تیان‌لانگ‌یل و همکاران [۱۳] یک مدل دینامیکی غیرخطی برای خودروهای مفصلی ارائه کردند و پایداری تحت شرایط بار متفاوت را تحلیل نمودند. جونگ و همکاران [۱۴] یک سیستم تخمینی زاویه لغزش جانبی با استفاده از فیلترهای کالمن دوگانه متغیر با زمان برای بهبود دقت کنترل فرمان طراحی کردند. چندین کار نیز طراحی کنترل‌کننده برای سیستم‌های مفصلی را بررسی کرده‌اند، مانند کنترل خطی با تنظیم بهره و کنترل غیرخطی مبتنی بر نظریه لیاپانوف [۱۵]. این مطالعات امکان‌پذیری نظری کنترل خودروهای مفصلی، به ویژه در مانورهای عقب‌گرد را تأیید می‌کنند. با این حال، عملکرد آنها در محیط‌های پیچیده و پویا همچنان محدود است.

ون‌ده‌وندل و همکاران [۱۶] که یادگیری تقویتی را برای برنامه‌ریزی مسیر در مانور پارک عقب‌گرد بررسی کردند و آئرتسن و همکاران [۱۷] که برنامه‌ریزی مسیر زمان واقعی برای وسایل نقلیه تجاری مفصلی خودران ارائه دادند، بر جنبه‌های عملی تمرکز دارند. چنگ [۱۸] رویکرد یادگیری تقویتی عمیق برای برنامه‌ریزی مسیر طولی آنلاین در ترافیک مختلط را پیشنهاد داد. ندامانی و همکاران در [۴] و [۵] مدل‌سازی راننده مبتنی بر محاسبات نرم برای پارک خودکار وسایل نقلیه سنگین و پارک خودکار خودروهای مفصلی با ANFIS را تحلیل کردند. پی و همکاران [۲۰] مسیریابی خودسازمان‌یافته برای وسایل نقلیه خودران با یادگیری تقویتی عمیق را معرفی کردند. یه و همکاران [۲۰] مروری بر الگوریتم‌های یادگیری تقویتی عمیق برای برنامه‌ریزی حرکت و کنترل وسایل نقلیه خودران ارائه دادند. چنگ و همکاران [۲۱] یادگیری تقویتی ایمن انتها به انتها با توابع مانع برای وظایف کنترل پیوسته ایمنی-بحرانی را پیشنهاد کردند. لیو و وانگ [۲۲] از TD3 برای برنامه‌ریزی حرکت بدون برخورد در ناوبری ربات استفاده کردند. زو و همکاران [۲۳] روش تصمیم‌گیری رانندگی خودران انتها به انتها مبتنی بر الگوریتم TD3 بهبودیافته در سناریوهای پیچیده را ارائه دادند. ژانگ و همکاران [۲۴] تولید مسیر پارک خودکار با یادگیری تقویتی عمیق را پیشنهاد دادند. آناگنوستارا و همکاران [۲۵] یادگیری تقویتی عمیق و یادگیری تقلیدی برای شبیه‌سازی پارک خودران را ترکیب کردند.

واهدان [۲۶] برنامه‌ریزی مسیر پویا برای وسایل نقلیه خودران با مدل TD3 یادگیری تقویتی تطبیقی را ارائه داد. آتارد [۲۷] یادگیری تقویتی برای ناوبری خودران وسایل نقلیه مفصلی را در پایان‌نامه کارشناسی ارشد بررسی کرد. ون‌دام و همکاران [۲۸] تخصیص کنترل ایمن برای وسایل نقلیه سنگین مفصلی را در سمپوزیوم کنترل وسیله نقلیه پیشرفته بررسی کردند. وانگ و همکاران [۲۹] روش تصمیم‌گیری و برنامه‌ریزی مسیر وسیله نقلیه هوشمند مبتنی بر یادگیری تقویتی عمیق در فضای فرنت را ارائه دادند.

در حالی که یادگیری تقویتی عمیق اثربخشی خود را در سیستم‌های تنظیم موقعیت متداول نشان داده است، کاربرد آن در تنظیم موقعیت خودروهای مفصلی همچنان کمتر بررسی شده است. روش‌های موجود معمولاً فاقد مدولاریتی برای مدیریت تنظیم موقعیت چندمرحله‌ای هستند و به ندرت دینامیک‌های خاص وسایل نقلیه مانند محدودیت‌های زاویه مفصل را در نظر می‌گیرند. این پژوهش یک چارچوب یادگیری تقویتی مدولار و مبتنی بر مراحل برای تنظیم موقعیت خودروهای مفصلی با استفاده از DDPG و TD3 پیشنهاد می‌دهد. با مدل‌سازی دینامیک‌های خاص وسایل نقلیه و تقسیم وظیفه به زیروظایف، این روش پایداری یادگیری و عملکرد را بهبود می‌بخشد. در مطالعات پیشین مرتبط با مسیریابی تنظیم موقعیت خودکار خودروهای مفصلی، اغلب از روش‌هایی مانند بهره‌گیری از داده‌های جمع‌آوری‌شده از رانندگان متخصص (که هزینه‌بر و زمان‌بر است و عملکرد را به سطح متخصصان محدود می‌کند) یا مسیرهای از پیش تعریف‌شده بر اساس دانش قبلی انسانی (مانند روش‌های هندسی، جستجوی کیستیوری و کنترل فازی) استفاده شده است. این رویکردها معمولاً بهترین مسیر ممکن را ارائه نمی‌دهند، وابسته به سناریوهای خاص هستند، فاقد سازگاری با محیط‌های پویا می‌باشند و ممکن است بهینه‌سازی همزمان پارامترهایی مانند ایمنی و کارایی را نادیده بگیرند. نوآوری‌های کلیدی در این پژوهش عبارتند از:

- به کارگیری دو روش از کاربرد یادگیری تقویتی عمیق در تنظیم موقعیت خودروهای مفصلی؛ برخلاف کارهای قبلی که یادگیری تقویتی را فقط برای خودروهای معمولی استفاده کرده‌اند در این پژوهش مدل دینامیکی خودروی مفصلی که علاوه بر برخورد دچار حادثه قیچی‌شدن هم می‌شود به کار گرفته شده است.
- بهینه‌سازی همزمان بازدهی، ایمنی (شامل جلوگیری از پدیده قیچی‌شدن و برخورد با موانع) و راحتی مانور (مانند صافی مسیر و کاهش تنش‌های مکانیکی) به عنوان پارامترهای کلیدی در عملیات تنظیم موقعیت خودروهای مفصلی در نظر گرفته می‌شوند. این عوامل در طراحی سیستم‌های کنترل برای تضمین عملکرد ایمن و کارآمد در فضاهای محدود ضروری هستند و در کاربردهای مشابه مانند پارک خودکار (که نیازمند مسیریابی دقیق است و شباهت‌های ساختاری زیادی با تنظیم موقعیت دارد) نیز برجسته می‌باشند.

۳- مدل سازی سینماتیکی و اعتبارسنجی یک وسیله نقلیه مفصلی برای تنظیم موقعیت خودکار

۳-۱- مقدماتی بر مدل‌سازی سینماتیکی

وسایل نقلیه مفصلی، از یک کشنده (Tractor) و یک تریلر (Trailer) تشکیل شده‌اند که توسط یک مفصل لولایی به هم متصل شده‌اند. به دلیل رفتار سینماتیکی غیرخطی و با ابعاد بالا، چالش‌های قابل توجهی را برای سامانه‌های تنظیم موقعیت خودکار ایجاد می‌کنند. در مقایسه با وسایل نقلیه متداول، سیستم‌های مفصلی یک درجه آزادی اضافی یعنی زاویه مفصل (Articulation Angle) را معرفی می‌کنند که مانورها را، به ویژه در عملیات مقدماتی، پیچیده‌تر می‌سازد. از این رو، تدوین یک مدل سینماتیکی دقیق برای شبیه‌سازی رفتار وسیله نقلیه به شیوه‌ای آگاه از فیزیک و همچنین آموزش یک عامل یادگیری تقویتی تحت شرایطی که هم واقع‌بینانه باشد و هم از نظر محاسباتی کارآمد، امری حیاتی است.

۳-۲- فرضیات مدل‌سازی

برای انجام فرایند یادگیری عملیات تنظیم موقعیت خودروی مفصلی توسط عامل، یک مدل ریاضی نیاز است تا با دریافت ورودی‌های کنترلی وضعیت بعدی خودرو را تعیین کند. برای این کار از یک مدل دوبعدی با برخی فرضیات برای ساده‌سازی استفاده شده است. در ادامه هر یک از این فرضیات تشریح شده‌اند.

- حرکت بدون لغزش لاستیک‌ها: باتوجه به اینکه سرعت خودرو در فرایند تنظیم موقعیت بسیار کم است، فرض می‌شود که لاستیک‌های خودرو بدون لغزش حرکت می‌کنند. این بدان معناست که هیچ انحرافی از مسیر پیش‌بینی‌شده ناشی از لغزش لاستیک‌ها وجود ندارد. این فرضیه امکان استفاده از مدل‌های ساده‌تر را فراهم می‌آورد.
- مدل‌سازی دوبعدی: مدل‌سازی در فضای دوبعدی انجام می‌شود. بنابراین تغییرات در محور عمودی نادیده گرفته شده و حرکت تنها در صفحه x-y تحلیل می‌شود. همچنین سطح تنظیم موقعیت کاملاً صاف و بدون پستی‌وبلندی فرض می‌شود. این فرض باعث حذف اثرات ناشی از تغییرات ارتفاع و شیب در مسیر حرکت خودرو می‌شود.
- ساختار صلب خودرو: کامیون و تریلر صلب در نظر گرفته شدند و هیچ تغییر شکلی در این بخش‌ها رخ نمی‌دهد.
- محدودیت‌های زاویه فرمان: زاویه چرخ‌های جلو خودرو بر اساس یکی از خودروهای مرسدس بنز در محدوده ۵۲- تا ۵۲ درجه در نظر گرفته می‌شود که از طراحی فیزیکی و محدودیت‌های مکانیکی خودرو تبعیت می‌کند [۳۰].
- محدودیت سرعت: حرکت به دو صورت روبه جلو و عقب با سرعت‌های محدود انجام می‌شود. به طور مشخص، سرعت

حرکت روبه جلو تا ۱.۵ متر بر ثانیه و سرعت حرکت روبه عقب تا ۰.۸ متر بر ثانیه فرض شده است.

- نادیده گرفتن نیروهای محیطی: عوامل خارجی نظیر باد، اصطکاک‌های ناشی از هوا یا سایر تأثیرات محیطی در مدل‌سازی لحاظ نمی‌شوند.
- مدل تک‌محور: به جای استفاده از جفت چرخ‌ها در هر محور، تنها یک چرخ مرکزی به عنوان جایگزین هر جفت چرخ در نظر گرفته شده است. این تغییر باعث ساده‌تر شدن معادلات حرکت و کاهش تعداد متغیرهای مورد نیاز برای مدل‌سازی می‌شود. این تغییر به مدل تک مسیر مشهور است که در بسیاری از کاربردهای مدل‌سازی وسایل نقلیه به کار می‌رود.
- استفاده از محورهای معادل: برای ساده‌سازی فرایند، به جای درنظرگرفتن مجموع محورهای واقعی، از تقریبی مدل تک‌محور معادل برای کشنده و تریلر استفاده شده است. این رویکرد سینماتیکی خودرو مفصلی را تسهیل کرده و امکان استخراج معادلات وضعیت (State Equations) نهایی، که در شکل ۱ نمایش داده شده، را فراهم می‌آورد.

۳-۳- هندسه وسیله نقلیه و چارچوب‌های مرجع

شکل ۱ هندسه وسیله نقلیه مفصلی را با نقاط مرجع کلیدی نشان می‌دهد:

- نقطه وسط محور جلوی کامیون: ft
- مفصل کامیون و تریلر : rt
- مرکز محور عقب تریلی: rs

حالت وسیله نقلیه توسط موارد زیر تعریف می‌شود:

- موقعیت: (xt,yt)، (xs,ys)
- جهت‌گیری (زوایای هدایت): φ_s ,φ_t
- زاویه مفصلی: $\gamma \triangleq \varphi_t - \varphi_s$

ورودی‌های کنترلی عبارتند از:

- سرعت خطی v
- زاویه فرمان θ

۳-۴- استخراج معادلات سینماتیکی

برای ثبت حرکت وسیله نقلیه، از فرض عدم لغزش بهره گرفته شده و محدودیت‌های سرعت در مختصات محلی استخراج می‌گردند. شرط عدم لغزش به این معناست که مؤلفه سرعت جانبی در نقاط تماس تایرها باید صفر باشد.

با استفاده از مفهوم مرکز چرخش لحظه‌ای (ICR)، سرعت زاویه‌ای کشنده به صورت زیر تعریف می‌شود:

$$\dot{\varphi}_t = \frac{v}{L_t} sin(\theta) \qquad (۱)$$

بنابراین:

$$\dot{\gamma} \triangleq \dot{\varphi}_t - \dot{\varphi}_s \qquad (۲)$$

فهرست متغیرهای مدل در جدول ۱ ارائه شده است.

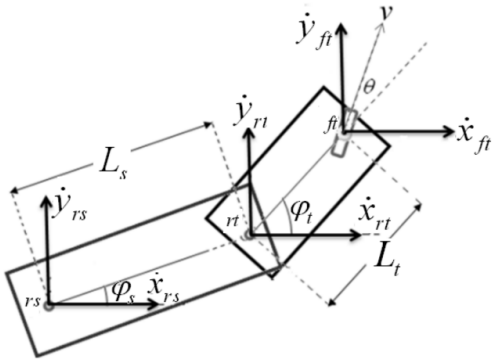
جدول ۱ متغیرها و مولفه‌های مدل سینماتیکی

Table 1 Variables and Parameters in the Kinematic Model

Symbol	Description
v	Linear velocity (control input)
θ	Steering angle (control input)
Lt	Distance from front axle to articulation joint
Ls	Distance from joint to trailer rear axle
φ_t	Heading angle of tractor
φ_s	Heading angle of trailer
γ	Articulation angle
ft ,rt, rs	Reference points for position tracking

۳-۵- تشخیص قیچی‌شدن و فرمول‌بندی ریاضی

یکی از حالت‌های شکست حیاتی در وسایل نقلیه مفصلی، قیچی‌شدن است که زمانی رخ می‌دهد که زاویه مفصلی از آستانه‌ای فراتر رود که بازیابی غیرممکن است. در این نقطه، حرکت چرخشی تریلی بر کشنده غالب می‌شود و سیستم ناپایدار دینامیکی می‌گردد (شکل ۲).



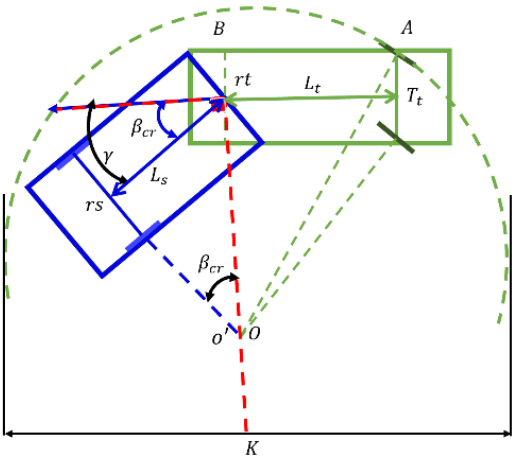
شکل ۱ محور مختصات محلی برای تعیین موقعیت کامیون و تریلر

Fig. 1 Local coordinate axis for determining the position of the truck and trailer

این شرط به طور هندسی با هم‌ترازی مراکز چرخش محورهای کشنده و تریلی مشخص می‌شود. زاویه بحرانی که قیچی‌شدن آغاز می‌شود توسط معادله زیر تعیین می‌شود:

$$\gamma_{cr} = arcsin\left(\frac{L_s}{\sqrt{\left(\frac{K-f}{2}\right)^2 - L_t^2} - \frac{T_t}{2}}\right) \qquad (۳)$$

در جدول ۲، متغیرها و پارامترها در مدل سینماتیکی ارائه شده‌اند.



شکل ۲ وضعیت خودرو در هنگام وقوع پدیده قیچی‌شدن

Fig. 2 Articulated Vehicle Configuration During Jackknifing

هنگامی که $\gamma > \beta_{cr}$ ، اپیزود در شبیه‌سازی خاتمه می‌یابد و پاداش منفی به عامل یادگیری تقویتی اختصاص می‌یابد تا از وقوع پدیده قیچی‌شدن جلوگیری کند.

جدول ۲ متغیرها و مولفه‌های مدل سینماتیکی

Table 2 Variables and Parameters in the Kinematic Model

Symbol	Description
γ_{cr}	Critical jackknifing angle
T_t	Truck wheelbase length
K	Twice the turning radius of the outer tire
f	Outer tire thickness
o'	Rotation center of the trailer's rear wheel
o	Rotation center of the trailer's front-outer wheel

۳-۶- صحت‌سنجی مدل خودروی مفصلی

برای اعتبارسنجی دقت مدل سینماتیکی استخراج‌شده، از آزمون گوزن (بر أساس استاندارد ایزو۲-۳۸۸۸ - یک مانور استاندارد برای ارزیابی پایداری وسیله نقلیه در حین اجتناب از موانع) استفاده شده‌است. آزمون شامل سه بخش فرمان اصلی است:

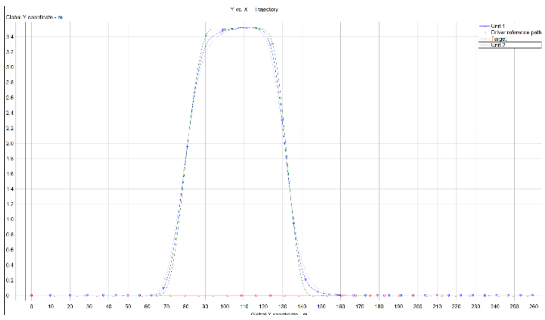
- ورود خط مستقیم
- تغییر خط ناگهانی
- بازگشت به خط اصلی

آزمون گوزن هم در نرم‌افزار تراک‌سیم (Trucksim) (شکل ۳) و هم در Python مبتنی بر معادلات سینماتیکی پیشنهادی، شبیه‌سازی شد.

در انجام آزمون، سرعت خودرو ثابت و برابر ۱/۵ متر بر ثانیه است. برای تعیین زاویه چرخ‌ها به‌عنوان ورودی شبیه‌سازی دوبعدی، ابتدا آزمون گوزن در نرم‌افزار تراک‌سیم اجرا شد. سپس زاویه چرخ‌ها استخراج شده و به‌عنوان ورودی به شبیه‌سازی صورت‌گرفته در نرم‌افزار پایتون ارائه شد.

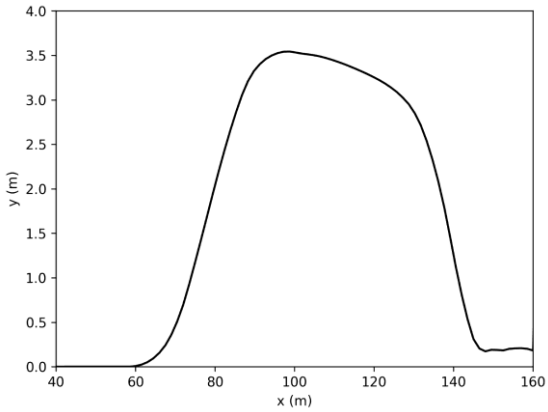
شکل ۳ شبیه‌سازی انجام شده در نرم‌افزار تراک‌سیم

Fig. 3 Simulation Performed in TruckSim Software



شکل ۴ مسیر پیموده‌شده در نرم‌افزار تراک‌سیم

Fig. 4 Motion Path in the TruckSim Environment



شکل ۵ مسیر شبیه‌سازی شده در نرم‌افزار پایتون

Fig. 5 Motion Path in the Python Environment

شکل‌های ۴ و ۵ مسیرهای مسیر هر دو شبیه‌سازی را نشان می‌دهند. با مقایسه مسیر طی شده و محاسبه فاصله نقطه‌به‌نقطه هر دو مسیر با یکدیگر، میانگین مربعات خطا کمتر از ۱/۸۵ است. برای انجام این محاسبه ابتدا به ازای هر نقطه در مسیر شبیه‌سازی پایتون، نزدیک‌ترین نقطه در مسیر شبیه‌سازی تراک‌سیم تشخیص داده می‌شود. سپس فاصله هر نقطه با نقطه نظیر محاسبه شده و با رابطه زیر میانگین مربعات خطا محاسبه خواهد شد. این میزان خطا باتوجه‌به اینکه گام زمانی تراک‌سیم یک میلی‌ثانیه و گام زمانی شبیه‌سازی صورت‌گرفته برای یادگیری یک‌دهم ثانیه است (تا عملیات یادگیری با سرعت مناسبی انجام شود)، مقدار مناسبی است.

۴- پیاده‌سازی یادگیری تقویتی برای تنظیم موقعیت خودروی مفصلی

۴-۱- یادگیری تقویتی در تنظیم موقعیت خودکار

یادگیری تقویتی یک حوزه از یادگیری ماشینی است که به طراحی و آموزش عامل‌های هوشمند برای انجام وظایفی مختلف می‌پردازد. در مدل‌های تنظیم موقعیت خودکار خودرو، یادگیری تقویتی وظیفه دارد که ورودی‌های کنترلی مانند سرعت و زاویه فرمان را به مدل ارائه داده و خودرو را در مسیری دلخواه به هدف هدایت کند. به عبارتی، عامل یادگیری تقویتی نقش کنترل‌کننده را ایفا می‌کند و با استفاده از اطلاعات حسگرها، تصمیمات مناسب را اتخاذ می‌کند تا خودرو به طور ایمن و کارآمد به مقصد برسد.

در عملیات تنظیم موقعیت خودکار، عوامل مختلفی از جمله بازدهی مانور، راحتی سرنشینان، ایمنی خودرو و وضعیت نهایی موقعیت اهمیت دارند. همچنین، زمان مورد نیاز برای تنظیم موقعیت نیز به عنوان یک معیار مهم در نظر گرفته می‌شود. عامل یادگیری تقویتی با در نظر گرفتن تمام این اهداف، یاد می‌گیرد که چگونه بهترین تصمیم‌ها را برای رسیدن به آن‌ها اتخاذ کند.

۴-۲- ساختار فضای حالت

بردار حالت ارائه شده به عامل یادگیری تقویتی باید کلیه اطلاعات لازم و مرتبط برای تصمیم‌گیری را به‌نحوی کدگذاری کند که از افزونگی پرهیز شود. در وسایل نقلیه مفصلی، این بردار شامل مجموعه‌ای از ویژگی‌های موقعیتی، هندسی و وضعیت عملیاتی است که به شرح زیر دسته‌بندی می‌شوند:

- حالت کشنده:
  - موقعیت دکارتی: (xt,yt)
  - زاویه هدایت: φ_t
  - سرعت رو به جلو: v
- حالت تریلی:
  - موقعیت محور عقب: (xs,ys)
  - زاویه هدایت: φ_s
- هندسه مفصلی:
  - زاویه مفصلی: γ
  - فاصله و جهت‌گیری به هدف: $d, \varphi_{sg}, \varphi_{tg},$

بردار حالت کلی به صورت زیر نشان داده می‌شود:

$$s = [x_t, y_t, \varphi_t, x_t, y_t, \varphi_t, d, \varphi_{sg}, \varphi_{tg},] \quad (۴)$$

برای اطمینان از تعمیم و همگرایی سریع‌تر، تمام ورودی‌ها به محدوده‌های ثابت نرمال‌سازی می‌شوند.

۴-۳- فضای عمل و ساختار سیاست

بردار اقدام a شامل دستورات کنترل با مقادیر پیوسته است:

- زاویه فرمان: $\theta \in [-52°, 52°]$
- سرعت خطی: $v \in [-0.8°, 1.5°]$

انتخاب فضای عمل پیوسته به دلیل نیاز به دقت بالا صورت گرفته است، زیرا اعمال گسسته ممکن است منجر به نوسان، حرکت نامنظم یا همگرایی زیربهینه شود.

سیاست (Policy) مورد استفاده، قطعی (Deterministic) است و یک عمل منفرد را برای هر حالت تولید می‌کند، که این ویژگی مورد نیاز الگوریتم‌های DDPG و TD3 می‌باشد. شبکه بازیگر (Actor Network) نگاشت سیاست را به‌صورت زیر تقریب می‌زند:

$$a_t = \mu(s \mid \theta^{\mu}) \quad (۵)$$

که در آن θ^{μ} پارامترهای قابل یادگیری شبکه را نشان می‌دهد.

۴-۴- فرمول‌بندی تابع پاداش

تابع پاداش یکی از ارکان اساسی در یادگیری تقویتی است، زیرا رفتار عامل را هدایت می‌کند و به آن می‌آموزد که چه اقداماتی مطلوب و چه اقداماتی نامطلوب هستند.تابع پاداش ترکیبی چند هدفی به این صورت تعریف می‌شود:

$$\begin{aligned} R_{total} = {} & k_{collision}R_{collision} \\ & + k_{jacknife}R_{jacknife} \\ & + k_{deviation}R_{deviation} \\ & + k_{goal}R_{goal} \\ & + k_{distance}R_{distance} \\ & + k_{comfort}R_{comfort} \\ & + k_{efficiency}R_{efficiency} \end{aligned} \quad (۶)$$

در ادامه هر یک از پاداش‌ها به صورت مجزا مورد بررسی قرار می‌گیرند.

۴-۴-۱- پاداش هدف

پس از رسیدن موفق‌آمیز به ناحیه موقعیت تعیین‌شده، پاداش مثبت به عامل اعطا می‌گردد. علاوه بر این، یک پاداش ثانویه بر اساس کیفیت هم‌ترازی بین کشنده و تریلر اعمال می‌شود تا اطمینان حاصل شود که وسیله نقلیه با حداقل خطای مفصلی به موقعیت صحیح می‌رسد:

$$R_{goal1} = \begin{cases} +50 & if\ target \\ 0 & else \end{cases} \quad (۷)$$

$$R_{goal2} = \begin{cases} 5 \times |\varphi_s - 90| + 5\,|\varphi_t - 90| & if\ target \\ 0 & else \end{cases} \quad (۸)$$

$$R_{goal} = R_{goal1} + R_{goal2} \quad (۹)$$

این امر اطمینان می‌دهد که وسیله نقلیه به موقعیت صحیح با خطای مفصلی حداقل برسد.

۴-۴-۲- جریمه برخورد

ایمنی بالاترین اولویت را در تنظیم موقعیت خودکار دارد. اپیزود در صورت بروز شکست‌های ایمنی بلافاصله خاتمه می‌یابد و جریمه‌های منفی بزرگ اختصاص داده می‌شوند:

$$R_{collision} = \begin{cases} -50 & if\ collision \\ 0 & else \end{cases} \quad (۱۰)$$

۴-۴-۳- جریمه قیچی‌شدن

برای جلوگیری از قیچی‌شدن، شرطی که زاویه مفصلی از حد ایمن فراتر رود، عامل در صورت تجاوز زاویه مفصلی از آستانه بحرانی جریمه می‌شود. اپیزود خاتمه یافته و جریمه زیر اعمال می‌گردد:

$$R_{deviation} = \begin{cases} -50 & if\ jacknife \\ 0 & else \end{cases} \quad (۱۱)$$

۴-۴-۴- جریمه انحراف بیش از حد

در صورتی که فاصله بین وسیله نقلیه و هدف غیرقابل قبول بزرگ شود، یا اگر این فاصله به‌صورت مداوم در ۱۰ گام گذشته افزایش یافته باشد، اپیزود ناکارآمد تلقی شده و خاتمه می‌یابد. این شرط از محاسبات هدررفته بر روی مسیرهایی با احتمال موفقیت کم جلوگیری می‌کند:

$$R_{deviation} = \begin{cases} -50 & if\ deviation \\ 0 & else \end{cases} \quad (۱۲)$$

۴-۴-۵- پاداش شکل‌دهی مبتنی بر فاصله

این پاداش عامل را تشویق می‌کند تا به طور مداوم به هدف نزدیک‌تر شود. در هر گام زمانی، اگر وسیله نقلیه نزدیک‌تر از گام قبلی به هدف باشد، پاداش مثبت اختصاص می‌یابد. برعکس، اگر دورتر شود، پاداش منفی برای منع اقدامات ناکارآمد یا نادرست داده می‌شود. علاوه بر این، یک پاداش استاتیک مبتنی بر فاصله فعلی به هدف شامل می‌شود. هر دو پاداش توسط فاصله اولیه در ابتدای اپیزود نرمال‌سازی می‌شوند تا عدالت در موقعیت‌های شروع متفاوت تضمین شود. این طراحی دوجزئی هم پیشرفت مداوم و هم همگرایی دقیق به هدف را انگیزه می‌دهد.

$$R_{distance1} = \begin{cases} +2 & if\ closer \\ -2 & else \end{cases} \quad (۱۳)$$

$$R_{distance2} = 100 \times \frac{-d}{d_0} \quad (۱۴)$$

$$R_{distance} = R_{distance1} + R_{distance} \quad (۱۵)$$

۴-۴-۶- جریمه راحتی

به منظور در نظر گرفتن راحتی سرنشینان در حین تنظیم موقعیت خودکار، یک تابع جریمه مبتنی بر راحتی معرفی می‌شود. این جزء برای بهبود کیفیت سواری در پیاده‌سازی‌های عملی مهم است. بر اساس استانداردISO 2631-1:1997، شدت کلی لرزش با استفاده از میانگین مربع ریشه‌ای وزن (RMS) شتاب در سه جهت متعامد ارزیابی می‌شود:

$$a_v = \sqrt{k_x^2 a_{wx}^2 + k_y^2 a_{wy}^2 + k_z^2 a_{wz}^2} \quad (۱۶)$$

که a_{wx}، a_{wy} و a_{wz} شتاب‌های RMS وزنی در محورهای مربوطه هستند و kx، ky و kz ضرایب وزنی تعریف‌شده توسط ISO-۲۶۳۱ ۱- هستند. از آنجایی که شبیه‌سازی در این مطالعه در یک صفحه دوبعدی انجام می‌شود، شتاب عمودی a_{wz} ناچیز در نظر گرفته شده و از محاسبات حذف می‌شود.

بر اساس مقدار محاسبه‌شده a_v، یک جریمه راحتی گسسته اختصاص می‌یابد:

$$R_{comfort} = \begin{cases} 0, a_v < 0.315 \\ -1, 0.315 \le a_v < 0.63 \\ -2, 0.63 \le a_v < 1 \\ -3,1 \le a_v < 1.6 \\ -5,1.6 \le a_v < 2 \\ -7, 2 \le a_v \end{cases} \quad (۱۷)$$

۴-۴-۷- پاداش کارایی زمانی

برای تشویق سریع‌تر رسیدن به مقصد، یک پاداش مبتنی بر زمان معرفی می‌شود. این مؤلفه تنها پس از رسیدن به هدف فعال می‌شود و نسبت معکوس با تعداد گام‌های گرفته‌شده دارد:

$$R_{efficiency} = \begin{cases} min(60 - t, 0) & if\ target \\ 0 & else \end{cases} \quad (۱۸)$$

اهمیت پاداش و ضرایب ki بر اساس فاز آموزش تنظیم می‌شوند.

۴-۵- الگوریتم‌های یادگیری

در میان الگوریتم‌های یادگیری تقویتی عمیق برای وظایف کنترل پیوسته، روش‌هایی مانند PPO (Proximal Policy Optimization)، SAC (Soft Actor-Critic)، DDPG و TD3 کاربردی هستند. PPO به عنوان یک الگوریتم مبتنی بر سیاست جاری، پایداری بالایی در فرآیند آموزش دارد و در وظایف ساده‌تر (مانند فضاهای حالت کوچک) عملکرد خوبی نشان می‌دهد. SAC نیز مبتنی بر سیاست خارجی است اما با افزودن حداکثر آنتروپی به سیاست، کاوش بهتری در محیط‌های تصادفی ارائه می‌دهد و می‌تواند در جلوگیری از حالات ناپایدار (مانند قیچی‌شدن) مقاوم‌تر عمل کند. DDPG یک الگوریتم مبتنی بر سیاست خارجی و قطعی است که برای فضاهای عمل پیوسته طراحی شده و کارایی مناسبی در کنترل دینامیک‌های غیرخطی دارد. TD3 به عنوان نسخه بهبودیافته DDPG، با ویژگی‌هایی مانند منتقدان دوقلو و به‌روزرسانی‌های تأخیری، مشکلات بیش‌تخمین‌زنی را کاهش می‌دهد و پایداری بیشتری ارائه می‌کند. هر دو الگوریتم DDPG و TD3 به دلیل پایداری در وظایف کنترل پیوسته، برای مدیریت دینامیک‌های غیرخطی و فضای عمل با ابعاد بالا در تنظیم موقعیت وسیله نقلیه مفصلی مورد استفاده قرار می‌گیرند. در روابط زیر قانون‌های به‌روزرسانی TD3 و DDPG ارائه شده است.

$$y = r + \gamma \cdot \min\left(Q_1(s', a'),\ Q(s', a')\right) \quad (۱۹)$$

$$y = r + \gamma Q'\left(s', \mu'(s' \mid \theta^{\mu'}) \mid \theta^{Q'}\right) \quad (۲۰)$$

۴-۶- معماری‌های شبکه عصبی

در شبکه عصبی ارائه شده، هر عامل یادگیری تقویتی از ساختار زیر استفاده می‌کند:

- شبکه بازیگر:
  - ورودی: بردار حالت (ابعاد = ۸–۱۲)
  - لایه‌های پنهان: ۳–۵ لایه، فعال‌سازی‌های ReLU

- خروجی: ۲ نورون (سرعت، فرمان)، مقیاس‌شده از طریق tanh/sigmoid
- شبکه منتقد:
  - ورودی: الحاق [حالت، اقدام]
  - لایه‌های پنهان: ۳–۵ لایه، پیکربندی مشابه
  - خروجی: مقدار Q اسکالر
- ابتداسازی:
  - ابتداسازی Xavier/He
  - نرمال‌سازی دسته در لایه‌های اولیه
  - نرخ‌های یادگیری: 10^{-3} (بازیگر)، 10^{-4} (منتقد)

۴-۷- تنظیمات ابرپارامترها

برای هر یک از الگوریتم‌ها ابرمولفه‌های مختلفی وجود دارد که بر روی عملکرد آن‌ها تأثیر زیادی می‌گذارند. در فرایند یادگیری تقویتی، تنظیم دقیق ابرمولفه‌ها تأثیر مستقیمی بر سرعت و کیفیت یادگیری عامل دارد. این مؤلفه‌ها رفتار الگوریتم را شکل می‌دهند و تنظیم درست آن‌ها می‌تواند منجر به بهبود عملکرد در مسائل پیچیده شود. جدول ۳ عناصر اساسی یادگیری تقویتی را ارائه می‌دهد.اکتشاف به تدریج با پیشرفت آموزش کاهش می‌یابد تا بهره‌برداری از سیاست‌های آموخته‌شده را ترجیح دهد.

جدول ۳ مولفه‌های کلیدی یادگیری تقویتی

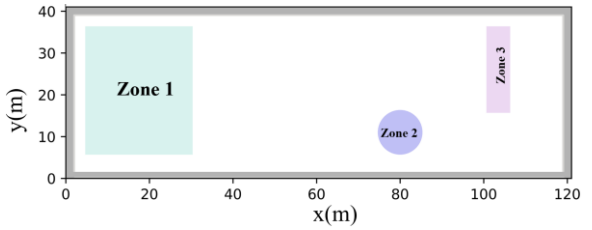
Table 3 Key Components of Reinforcement Learning

Parameter	Value
Replay buffer size	10^6
Batch size	125
Discount factor	0.99
Soft update	0.005
Action noise	$\mathcal{N}(0, 0.1^2)$ Gaussian
TD3 delay steps	2
Target noise (TD3)	$\mathcal{N}(0, 0.2^2)$ *Clipped*
Epsilon decay	$\varepsilon_0^{ne} and\ \varepsilon_0 = 0.9994\ or\ 0.9999$

۴-۸- استراتژی یادگیری چندفازه

باتوجه‌به پیچیدگی و دشواری حرکت خودروی مفصلی از منطقه شروع به محل نهایی تنظیم موقعیت و نیاز به پردازش قوی و گسترده برای یادگیری، این فرایند به دو بخش مجزا تقسیم شده است تا آموزش با سرعت بیشتری انجام شود. مطابق شکل ٦ سه منطقه برای این تنظیم موقعیت در نظر گرفته شده است. همچنین دو مسیر حرکت عبارت‌اند از:

- حرکت آماده‌شدن برای مانور: خودرو از منطقه ۱ با مکان و جهت‌گیری مشخص شروع به حرکت کرده و به منطقه ۲ می‌رسد.
- حرکت تنظیم جهت خودرو در راستای محور هدف: خودرو مسیر خود را از منطقه ۲ آغاز می‌کند تا در منطقه ۳ به‌صورت موازی با محور y قرار گیرد.



شکل ۶ مناطق در نظر گرفته شده برای فازهای حرکت خودرو

Fig. 6 Defined Zones for Vehicle Movements

محدوده آغاز و پایان هر فاز در جدول ٤ ارائه شده است.

جدول ۴ محدوده اغاز و پایان هر فاز (مکان و جهت‌گیری)

Table 4 Spatial Boundaries (Position And Heading Angles) For The Start And End Regions Of Each Phase

Phase	Start Goal Region	φ_s (deg)	φ_t (deg)	y_{ft} (m)	x_{ft} (m)
1	Start	[-30,30]	[-60,60]	[30,60]	[15,35]
	Finish	[-30,30]	[-30,30]	A circle of radius 5 located at (85, 35)	
2	Start	[-30,30]	[-30,30]	[30,40]	[70,80]
	Finish	[80,100]	[80,100]	[40,60]	[101,106]

ضرایب پاداش برای وظیفه تنظیم موقعیت مفصلی به دقت تنظیم شده‌اند تا اولویت‌های متغیر هر فاز حرکت را بازتاب دهند: در فاز اول که بر نزدیک شدن ایمن به ناحیه هدف از شرایط اولیه مختلف تمرکز دارد، جریمه‌های برخورد، قیچی‌شدن مفصل و انحراف، همراه با رسیدن به هدف و حفظ فاصله صحیح، همگی ضرایب متوسط (۳) دریافت می‌کنند. با ورود به فاز دوم که هم‌ترازی جهت‌گیری وسیله نقلیه حیاتی می‌شود، ضریب قیچی‌شدن مفصل به ۵ افزایش یافته و جریمه‌های برخورد و انحراف برای تضمین رفتار ایمن دنبال کردن منحنی با وزن ۴ اهمیت بیشتری پیدا می‌کنند. در هر دو فاز راحتی و کارایی زمانی با وزن حداقل و ثابت ۰.۵، به عنوان اولویت‌های ثانویه نسبت به نتایج ایمنی و عملکرد حیاتی در نظر گرفته می‌شوند.

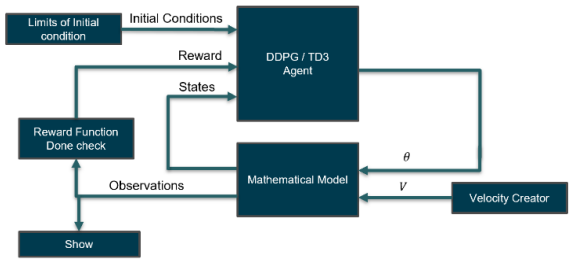
در این بخش یک چارچوب کنترل یادگیری تقویتی جامع برای تنظیم موقعیت خودکار وسیله نقلیه مفصلی را توصیف شد. با بهره‌برداری از استراتژی آموزش مرحله‌به‌مرحله، توابع پاداش شکل‌یافته و الگوریتم‌های قوی مانند TD3 و DDPG، سیستم سیاست‌های کنترل تطبیقی و تعمیم‌پذیر می‌آموزد. بخش بعدی اعتبارسنجی تجربی را از طریق شبیه‌سازی‌های با وفاداری بالا و مقایسه عملکرد در چندین سناریوی آزمون ارائه می‌دهد و مقایسه عملکرد در چندین سناریوی آزمون ارائه می‌دهد.

۵- چارچوب شبیه‌سازی و نتایج تجربی

۵-۱- مروری بر پلتفرم شبیه‌سازی

ایجاد یک سیستم شبیه‌سازی برای عملیات تنظیم موقعیت خودرو مفصلی با استفاده از مدل‌سازی تقریبی و الگوریتم‌های یادگیری تقویتی شامل طراحی و اجرای چندین کلاس پایتون بوده است (شکل ۷). این کلاس‌ها به شرح زیر عمل می‌کنند:

- کلاس مدل ریاضی: هسته شبیه‌سازی، مسئول مدل‌سازی دینامیک حرکت خودروی مفصلی (بر اساس ورودی‌های سرعت و زاویه فرمان)، محاسبه موقعیت جدید و بررسی وضعیت‌های مهم مانند برخورد یا خروج از محدوده.
- کلاس عامل یادگیری تقویتی: پیاده‌سازی الگوریتم‌های یادگیری تقویتی (مانند DDPG/TD3). این کلاس شامل شبکه‌های عصبی (بازیگر و منتقد)، مدیریت حافظه بافر، انتخاب‌کننده عمل و توابع به‌روزرسانی وزن‌ها است. وظیفه اصلی آن انتخاب عمل (تصادفی یا بهینه) بر اساس مشاهدات محیط است.
- توابع پاداش و بررسی اتمام قسمت: وظیفه محاسبه پاداش هر گام زمانی بر اساس وضعیت جدید خودرو و تغییرات نسبت به وضعیت قبل را بر عهده دارند و همچنین شرایط پایان یافتن یک اپیزود را مشخص می‌کنند.
- کلاس شرایط اولیه: وظیفه تعریف و اعمال محدودیت‌ها و شرایط اولیه شروع هر اپیزود یادگیری را دارد، از جمله تعیین موقعیت‌های اولیه متنوع برای افزایش کارایی آموزش.
- کلاس نمایش: مسئول نمایش گرافیکی و بصری فرایند حرکت خودرو در هر اپیزود، شامل نمایش مسیرهای طی شده، موقعیت نهایی و مقایسه آن با مسیر بهینه.



شکل ۷ ساختار کلی شبیه‌سازی عملیات تنظیم موقعیت خودروی مفصلی با استفاده از یادگیری تقویتی

Fig. 7 Overall Structure of Articulated Vehicle Parking Simulation Using Reinforcement Learning in Python

۵-۲- پارامترهای شبیه‌سازی و پیکربندی محیط

برای شبیه‌سازی، پارامترهای مختلفی در نظر گرفته شده‌اند:

- فاصله بین محور جلو کامیون و مفصل (L_t) = ۵ متر
- فاصله بین مفصل و محور عقب تریلر (L_s) = 10 متر
- عرض کامیون و تریلر = ۲/۵ متر
- محدوده زاویه فرمان: ±۵۲°
- محدوده سرعت: رو به جلو تا ۱.۵ متر بر ثانیه، رو به عقب تا ۰.۸ متر بر ثانیه
- مدت گام زمانی: ۱.۰ ثانیه
- آستانه قیچی‌شدن: تعیین به کمک رابطه (۳).
- تشخیص برخورد: حداقل فاصله از هر نقطه وسیله نقلیه به موانع کمتر از ۰.۱ متر

محیط نمایش‌های حالت متفاوت در هر فاز ارائه می‌دهد، متناسب با هدف یادگیری:

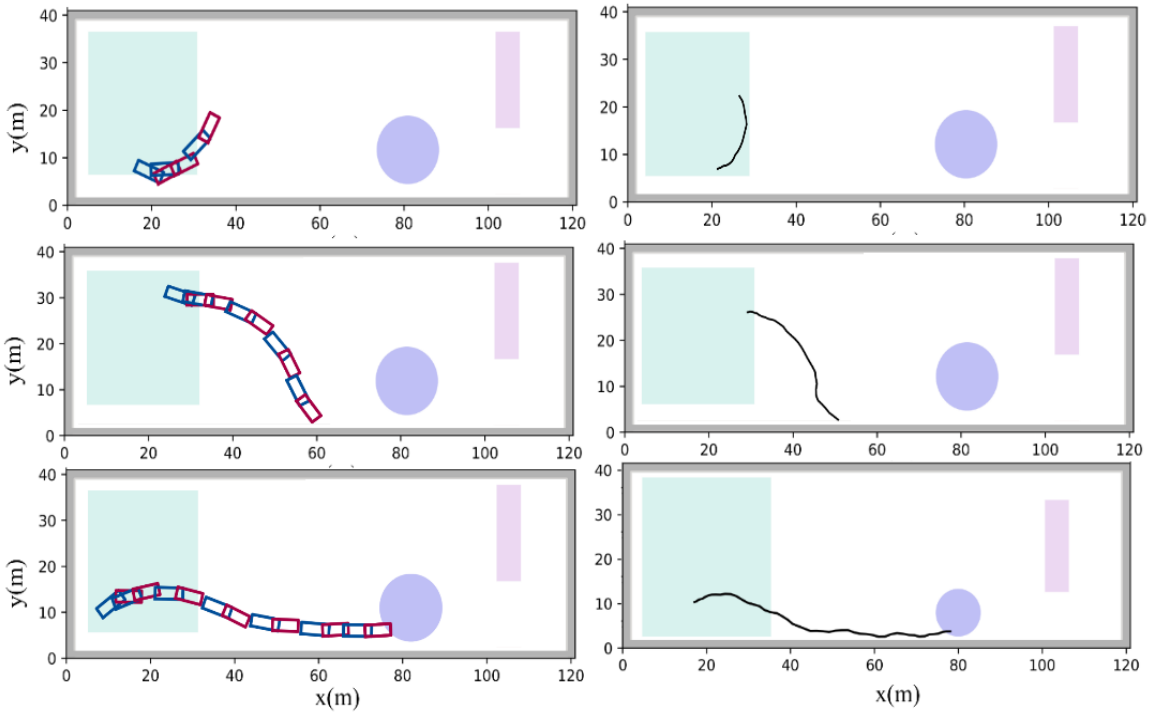
- در فاز ۱، حالت شامل فاصله‌های نسبی در x و y به ناحیه بعدی و زاویه مفصلی است.
- در فاز ۲، حالت شامل فاصله و انحراف زاویه‌ای از یک منحنی بزیه از پیش تعریف‌شده است.

شبیه‌سازی، انتقال خودکار بین فازها را بر اساس شرایط خاتمه تعریف‌شده توسط محدودیت‌های فضایی و هندسی مدیریت می‌کند. این معماری مدولار اجازه می‌دهد هر فاز به طور مستقل آموزش داده شود در حالی که رفتار انتها به انتها هموار هنگام ادغام حفظ شود.

۵-۳- فاز ۱: آماده‌سازی برای تنظیم موقعیت

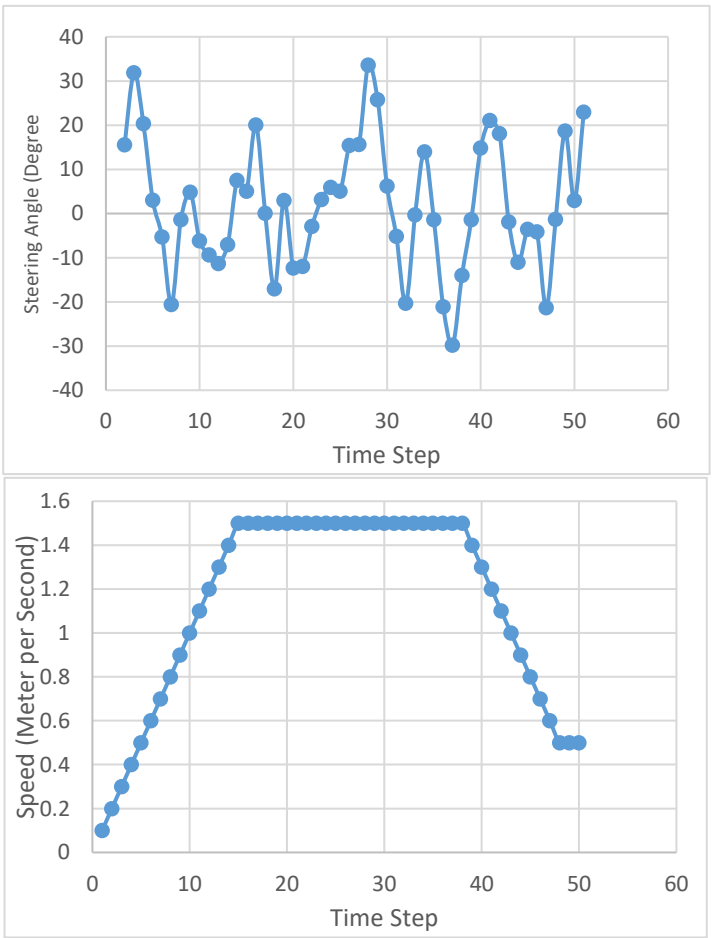
در این فاز وسیله نقلیه با موقعیت و جهت تصادفی از ناحیه ۱ شروع به حرکت می‌کند و هدف آن رسیدن به ناحیه هدف (ناحیه ۲) است. عامل یادگیری تقویتی موظف بود تنظیم زاویه فرمان و سرعت را برای آماده‌سازی وسیله نقلیه جهت هم‌ترازی نهایی بیاموزد. شکست‌های مکرر در فاز ۱ عمدتاً ناشی از برخوردها و رویدادهای قیچی‌شدن بودند. شکل ۸ نمونه‌هایی از قیچی‌شدن، برخورد و رسیدن به هدف در حین فرایند یادگیری را نشان می‌دهد. تحلیل مقایسه‌ای نرخ‌های موفقیت، حوادث مرتبط با ایمنی و پیکربندی‌های شبکه عصبی در جدول ۵ ارائه شده است. بهترین عملکرد با استفاده از الگوریتم TD3 با شبکه پنج‌لایه به دست آمد که به نرخ موفقیت ۹۶.۶٪ دست یافت و پایین‌ترین نرخ‌های برخورد (۶.۵۰٪) و قیچی‌شدن (۲.۲۵٪) را ثبت کرد. شکل ۹ افزایش پاداش‌ها را در طول اپیزودهای آموزش نشان می‌دهد، در حالی که شکل ۱۰ سیگنال‌های کنترل مربوطه (سرعت و زاویه فرمان) را نمایش می‌دهد. لازم به ذکر است که تغییرات حداکثر فرمان به ۲۰ درجه رسید، که این موضوع نشان می‌دهد فضای بیشتری برای دستیابی به کنترل هموارتر از طریق آموزش بیشتر وجود دارد.

به منظور ارزیابی مقاومت الگوریتم TD3 در برابر نویز حسگرها و تغییرپارامتر دو سناریو تست پیاده‌سازی شد.



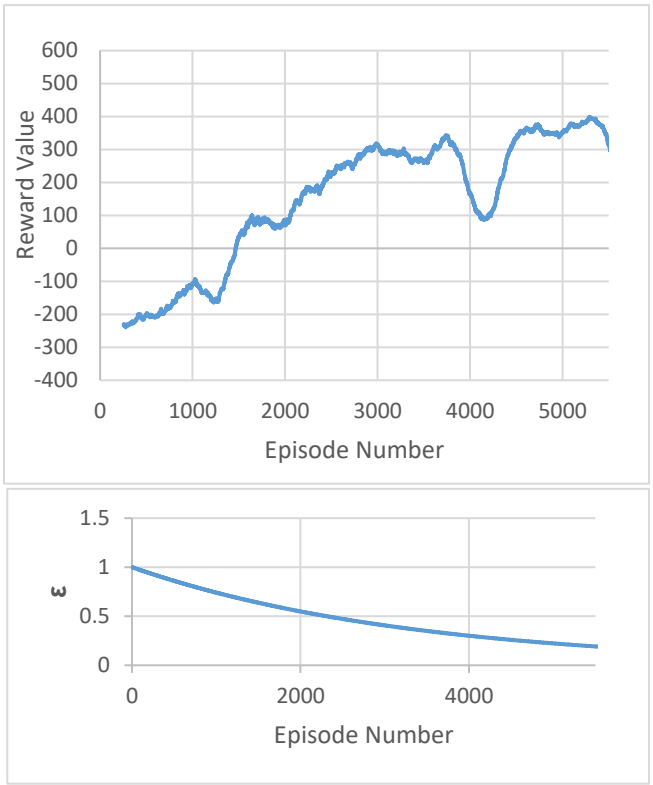
شکل ۸ نمونه‌هایی از رخ دادن پدیده قیچی‌شدن، تصادف و رسیدن به مقصد در حین فرایند یادگیری فاز اول

Fig. 8 Reward, ε, and goal-reaching frequency per episode for Phase 1 motion



شکل ۱۰ نمودار نمونه سرعت و زاویه فرمان بر اساس گام زمانی در پس از فرآیند یادگیری فاز اول

Fig. 10 Sample plots of velocity and steering angle over time steps after the learning process of Phase 1 motion



شکل ۹ نمودارهای پاداش و بر حسب قسمت برای حرکت فاز اول

Fig. 9 Reward, ε per episode for Phase 1 motion

در سناریوی نخست، برای شبیه‌سازی خطای اندازه‌گیری حسگرها، به داده‌های موقعیت و زاویه مفصل نویز گاوسی با میانگین(μ) صفر و انحراف معیارهای به‌ترتیب $\sigma_x = 0.05\ m$ و $\sigma_\gamma = 0.02\ rad$ افزوده شد. عامل TD3 با شبکه عصبی ۵ لایه که بهترین نتایج را کسب کرده بود با اعمال نویز تحت آزمون قرار گرفت. نتایج نشان دادند که نرخ موفقیت در مرحله‌ی آماده‌سازی از ۹۶.۶٪ به ۹۴.۲٪ کاهش یافت. این کاهش محدود بیانگر مقاومت قابل‌قبول الگوریتم در برابر نویز اندازه‌گیری است و نشان می‌دهد ساختار TD3 با سیاست تأخیری توانایی پایدارسازی فرآیند یادگیری را حتی در شرایط غیرایده‌آل حفظ می‌کند.

از میان پارامترهای مدل دینامیکی خودرو، ابعاد هندسی کشنده و تریلر بیشترین تأثیر را بر پویایی سیستم و قابلیت کنترل دارند. از آنجا که عرض وسایل نقلیه معمولاً مقدار ثابتی دارد، در این بخش تغییر در طول کشنده و طول تریلر به‌عنوان عوامل مؤثر در پایداری سیستم مورد بررسی قرار گرفت. افزایش طول، افزایش فاصله بین محورها را به همراه دارد. بنابراین فواصل مفصل تا محور جلوی کامیون (L_t) و محور عقب تریلر (L_s) پارامترهایی هستند که با تغییر آن‌ها مقاومت عامل در برابر تغییر سنجیده می‌شوند. بدین منظور شبیه‌سازی با $Lt=5.5\ m$ و $Ls=11\ m$ (به‌ترتیب حدود ۱۰٪ تغییر نسبت به مقادیر نامی) انجام شد. نتایج نشان داد که نرخ موفقیت در فاز آماده‌سازی از ۹۶٫۶٪ به ۹۵٫۱٪ کاهش یافت. بنابراین می‌توان نتیجه گرفت که الگوریتم *TD3* پیشنهادی در برابر تغییرات جزئی پارامترهای فیزیکی، به‌ویژه ابعاد وسیله نقلیه، از مقاومت و پایداری مناسبی برخوردار است.

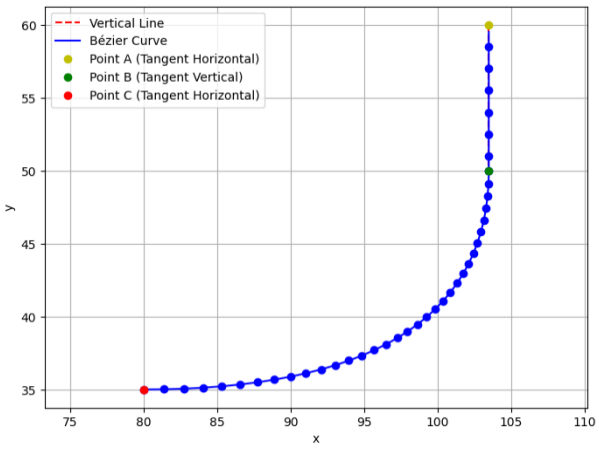
جدول ۵ نتایج مربوط به اجرای عملیات یادگیری فاز اول با ساختارهای شبکه عصبی متفاوت

Table 5 Results of the learning process execution in Phase 1 using different neural network architectures

Algorithm	DDPG	DDPG	DDPG	TD3	TD3	TD3
Number of Neural Network Layers	3	4	5	3	4	5
Number of Neural Network Neurons	32-46	64.128	128-256	32-46	64.128	128-256
Collision (%)	12.22	11.31	8.20	10.19	9.72	8.92
Jackknifing (%)	18.76	22.96	18.44	19.69	21.21	16.53
Deviation (%)	14.32	10.33	9.70	13.22	10.24	9.21
Reaching Destination (%)	54.70	55.40	63.66	56.90	58.83	65.34
Discount Factor	0.997	0.997	0.997	0.997	0.997	0.997
Number of Episodes	5500	5500	5500	5500	5500	5500
Success Rate in Final 500 Episodes (%)	86.4	90.6	94.6	88.4	94.0	94.6

۵-۴- فاز ۲: قرارگرفتن در محل هدف

در فاز دوم، هدف عامل یادگیری تقویتی انجام حرکتی منحنی‌شکل است تا وسیله نقلیه به موقعیت و جهت‌گیری مورد نیاز برسد. مشاهده شد که اگر تنها فاصله هدف به عامل داده شود، صرفاً به سمت هدف حرکت کرده و هرگز با جهت‌گیری مناسب به آن نخواهد رسید. به همین دلیل وظیفه عامل در این فاز ردیابی یک مسیر از پیش تعیین‌شده است، نه برنامه‌ریزی حرکت.

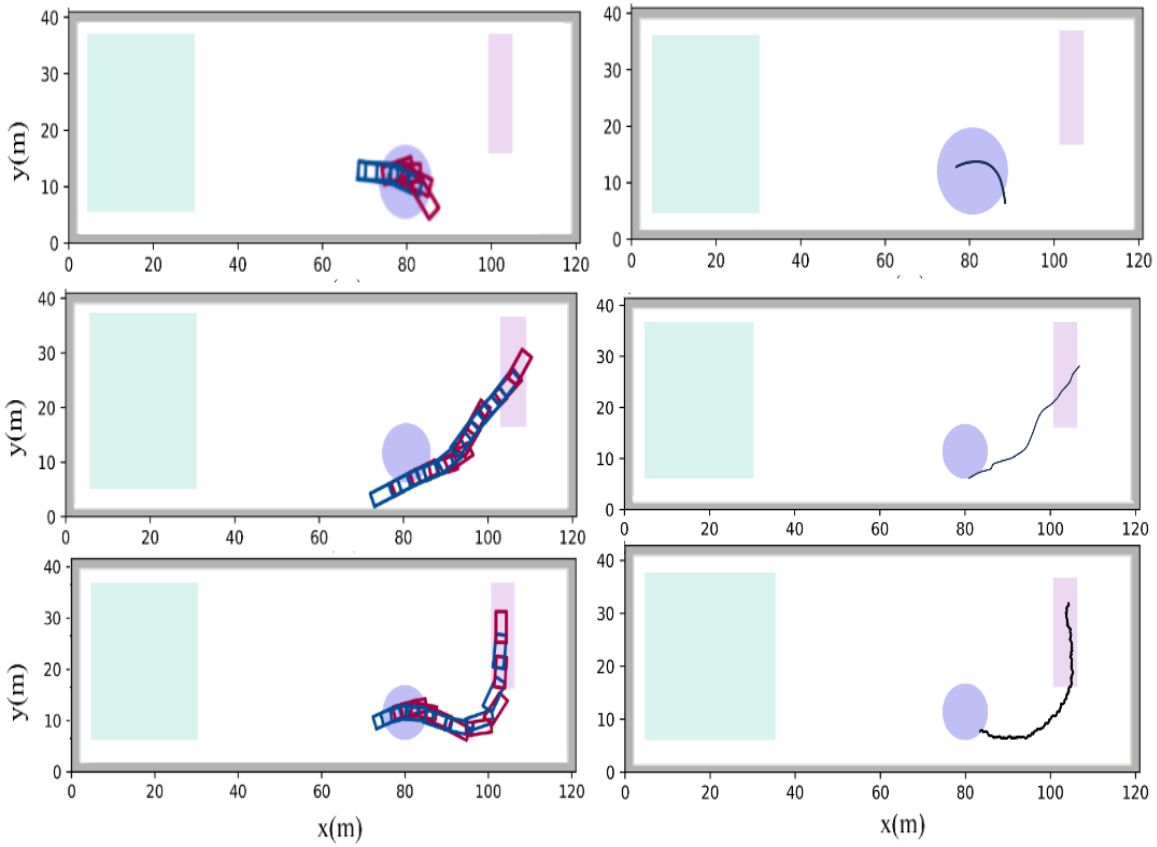


شکل ۱۱ منحنی بزیه برای تعریف مسیر حرکت در فاز دوم

Fig. 11 he Bézier curve used to define the motion path in the second phase.

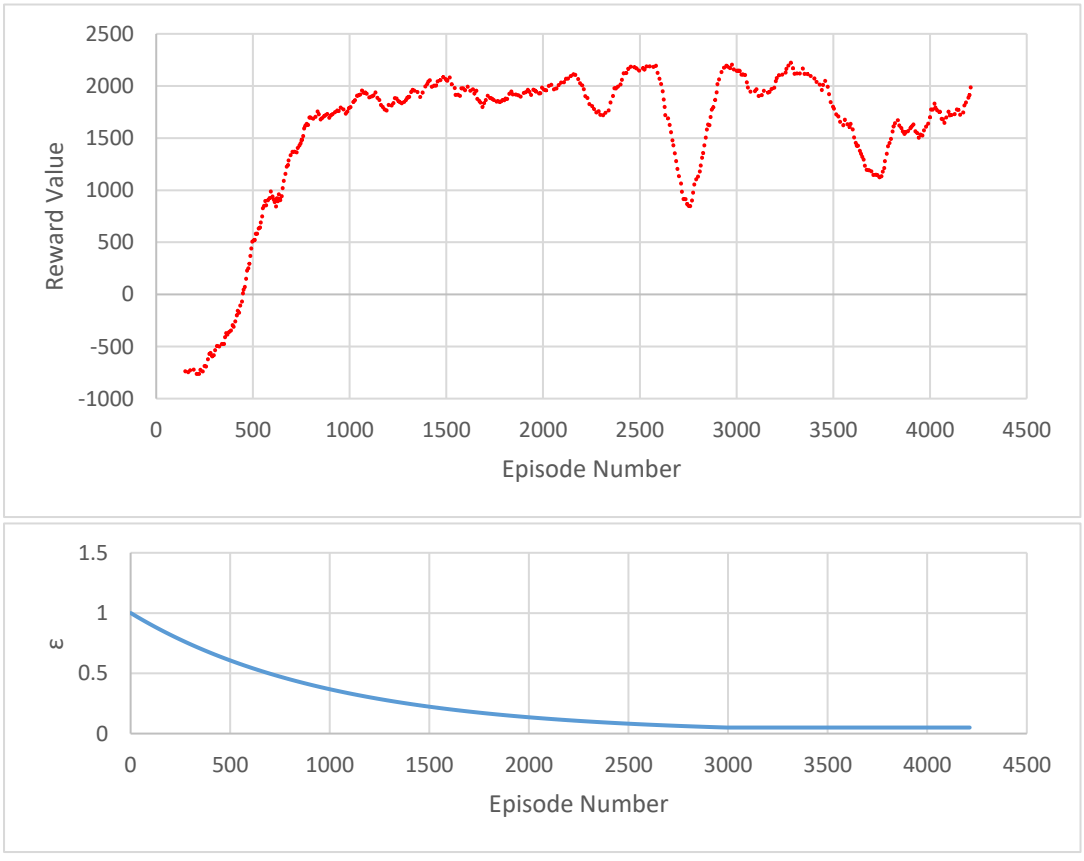
برای طراحی این مسیر، از منحنی بزیه (Bezier Curve) استفاده شد.(شکل ۱۱) منحنی بزیه یک ابزار استاندارد و بسیار مؤثر در مسیریابی و کنترل حرکت است که به دلیل انعطاف‌پذیری هندسی و سادگی در کنترل نقاط مسیر کاربرد گسترده‌ای در رباتیک و خودروسازی دارد. منحنی مورد استفاده با داشتن دو نقطه ثابت A و B و یک نقطه C که بر اساس مبدأ تعریف می‌شود، ایجاد می‌گردد. این منحنی به‌گونه‌ای تعریف می‌شود که مماس بر نقطه B به صورت عمودی و مماس بر نقطه C به صورت افقی باشد.

شکل ۱۲ نمونه‌هایی از قیچی‌شدن، دور شدن از هدف و رسیدن به هدف در حین فرایند یادگیری را نشان می‌دهد. شکل ۱۳ افزایش پاداش‌ها و نرخ موفقیت را در طول اپیزودهای آموزش و شکل ۱۴ سیگنال‌های کنترل در این فاز (سرعت و زاویه فرمان) را نمایش می‌دهد. در اوایل فرآیند آموزش، یکی از چالش‌های برجسته وقوع مکرر رفتارهای ناایمن بود که عمدتاً به شکل برخوردها و قیچی‌شدن ظاهر می‌شد. این شکست‌ها به‌ویژه زمانی مشهود بودند که عامل یادگیری تقویتی تحت سیاست‌های ناقص آموزش‌دیده عمل می‌کرد که فاقد آگاهی محیطی کافی یا دقت کنترل بودند. چنین رفتاری اغلب منجر به تقاطع بدنه وسیله نقلیه یا تریلی با دیوارها یا موانع تعریف‌شده می‌شد. به موازات آن، قیچی‌شدن یک مسئله



شکل ۱۲ رخ دادن پدیده قیچی‌شدن، دور شدن از هدف و رسیدن به مقصد در فاز ۲ در حین فرآیند یادگیری

Fig. 12 Occurrence of jackknifing, deviation from the goal, and successful goal-reaching during the learning process of Phase 2 motion.



شکل ۱۳ نمودارهای پاداش و ε به تارگت بر حسب قسمت برای حرکت فاز دوم

Fig. 13 Reward, ε, and goal-reaching frequency per episode for Phase 2 motion.

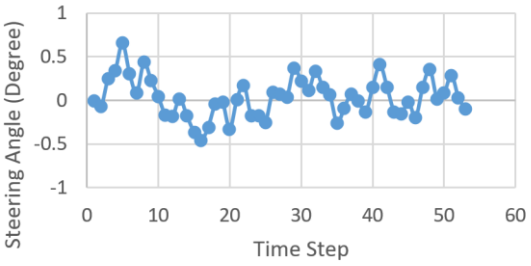
تکراری دیگر بود. این حوادث، پیچیدگی یادگیری سیاست‌های پایدار در مراحل اولیه آموزش را برجسته می‌کنند، به ویژه هنگام مدیریت غیرخطی بودن دینامیک‌های مفصلی.

نتایج اجرای عملیات یادگیری با ساختارهای مختلف شبکه عصبی (جدول ۶) نشان داد که الگوریتم‌های TD3 و DDPG، هر دو با ۵ لایه شبکه عصبی (تعداد نورون‌های ۲۵۶–۱۲۸)، بهترین عملکرد را کسب کردند و به بالاترین درصد موفقیت (۹۴.۶%) رسیدند. همچنین، الگوریتم TD3 با ۵ لایه شبکه، کمترین نرخ وقوع قیچی‌شدن (۱۶.۵۳%) را در بین تمام پیکربندی‌ها نشان داد.

مشابه فاز ۱، به منظور ارزیابی مقاومت الگوریتم TD3 در برابر نویز حسگرها و تغییرات پارامتر، دو سناریو اجرا شد.

برای ارزیابی مقاومت سامانه در برابر نویز، نویز گاوسی با میانگین صفر و انحراف معیارهای به‌ترتیب $\sigma_x = 0.05\ m$ و $\sigma_\gamma = 0.02\ rad$ به داده‌های موقعیت و زاویه مفصل افزوده شد. عامل TD3 با شبکه عصبی ۵ لایه تحت آزمون قرار گرفت و نرخ موفقیت در فاز تنظیم نهایی از ۹۴.۶٪ به ۹۲.۱٪ کاهش یافت. این افت اندک، پایداری TD3را در شرایط غیرایده‌آل تأیید می‌کند.

برای ارزیابی مقاومت سامانه در برابر تغییر پارامترها با تمرکز بر



شکل ۱۴ زاویه فرمان پس از فرایند یادگیری

Fig. 14 Sample plot of steering angle over time steps after the learning process of Phase 2 motion.

ابعاد هندسی، شبیه‌سازی با $Lt=5.5\ m$ و $Ls=11\ m$ (حدود ۱۰٪ تغییر نسبت به مقادیر نامی) انجام شد. نتایج نشان داد نرخ موفقیت در فاز تنظیم نهایی از ۹۴.۶٪ به ۹۳٫۲٪ کاهش یافت، که بیانگر مقاومت مناسب الگوریتم در برابر تغییرات فیزیکی است.

۵-۵- خلاصه نتایج

باتوجه به شبیه‌سازی‌های انجام‌شده، الگوریتم TD3 برای این مسئله مناسب‌تر تشخیص داده شد. این برتری به دلایل زیر قابل توضیح است:

- بهبود درصد موفقیت در ۵۰۰ قسمت نهایی: الگوریتم TD3 در همه حالات آزمایشی توانست درصد موفقیت بالاتر از ۹۵ درصد را در قسمت‌های نهایی به ثبت برساند. این موضوع نشان‌دهنده پایداری بیشتر این الگوریتم در یادگیری و یافتن مسیر بهینه برای تنظیم موقعیت خودرو مفصلی است.

جدول ۶ نتایج مربوط به اجرای عملیات یادگیری فاز دوم با ساختارهای شبکه عصبی متفاوت

Table 6 Results of the learning process execution in Phase 2 using different neural network architectures

Algorithm	DDPG	DDPG	DDPG	TD3	TD3	TD3
Number of Neural Network Layers	3	4	5	3	4	5
Number of Neural Network Neurons	32-46	64.128	128-256	32-46	64.128	128-256
Collision (%)	12..71	10.50	9.80	8.90	7.75	6.50
Jackknifing (%)	4.71	6.45	5.40	2.35	4.30	2.25
Deviation (%)	20.09	19.30	18.50	19.70	16.50	15.30
Reaching Destination (%)	62.49	63.75	66.30	69.05	71.45	75.95
Discount Factor	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997
Number of Episodes	5500	5500	5500	5500	5500	5500
Success Rate in Final 500 Episodes (%)	86.6	92.0	93.6	90.6	94.0	96.6

- کاهش وقوع قیچی‌شدن و برخورد: یکی از چالش‌های مهم این مسئله، جلوگیری از قیچی‌شدن و برخورد وسیله نقلیه در مسیر تنظیم موقعیت بود. نتایج نشان می‌دهد TD3 توانست در تعداد مراحل یادگیری با کاهش تعداد این رخدادها عملکرد بهتری نسبت به DDPG داشته باشد.
- پایداری در یادگیری: استفاده از مکانیزم‌های بهینه‌سازی تأخیری در TD3 باعث کاهش نویز در به‌روزرسانی پارامترهای شبکه عصبی و در نتیجه یادگیری پایدارتر شده است.
- باتوجه به این یافته‌ها، توصیه می‌شود در طراحی سیستم‌های مشابه که نیازمند یادگیری دقیق و امن در شرایط پیچیده هستند، از TD3 به‌عنوان الگوریتم اصلی بهره گرفته شود.
- روش ارائه‌شده در برابر نویز حسگرها و تغییرات پارامترهای فیزیکی مقاوم است، به طوری که با اعمال نویز و تغییر پارامتر، حداکثر ۲.۵ درصد در فاز آماده‌سازی و حداکثر ۳ درصد در فاز تنظیم نهایی دقت کاهش پیدا کرد. این سطح پایین کاهش، پایداری سیستم را در شرایط واقعی تأیید می‌کند و نشان‌دهنده کارایی الگوریتم TD3 در حفظ عملکرد بهینه حتی تحت عدم قطعیت‌های عملی است.

بالاترین درصد موفقیت در فازها به ترتیب ۹۶.۶% و ۹۴.۶% است. بنابراین می‌توان گفت اگر خودرویی در وضعیت اولیه تصادفی قرار گیرد با احتمال ۹۱.۳% می‌تواند به محل تنظیم موقعیت برسد.

روش هندسی مبتنی بر مسیرهای مستقیم و دایره‌ای برای مسیریابی جهت پارک عمودی خودروهای مفصلی [۳۱] نرخ موفقیت ۸۸.۴٪ را گزارش کرده است، در حالی که روش پیشنهادی با نرخ ۹۱٫۳ درصد در مجموع فازهای حرکت عملکرد بهتری داشته است. این برتری عمدتاً ناشی از تقسیم فازهای حرکت، بهره‌گیری از یادگیری تقویتی و بهینه‌سازی تابع پاداش می‌باشد.

۶- جمع‌بندی و پیشنهادات

۶-۱- جمع‌بندی

در این پژوهش، یک چارچوب نوآورانه بر پایه‌ی یادگیری تقویتی عمیق چندمرحله‌ای برای تنظیم موقعیت خودروهای مفصلی طراحی و شبیه‌سازی شد. با توجه به ماهیت پیچیده‌ی دینامیکی و غیرباریک خطی این وسایل نقلیه، روش‌های متداول کنترل و

برنامه‌ریزی مسیر نمی‌توانند به‌طور مؤثر پدیده‌هایی چون قفل‌شدگی مفصل یا انحراف از مسیر را مدیریت کنند. ازاین‌رو، پژوهش حاضر بر طراحی سیستمی هوشمند مبتنی بر عامل تمرکز داشت که بتواند به‌صورت خودکار و تطبیقی، موقعیت خودرو را در فضاهای محدود تنظیم کند.

در مرحله‌ی نخست، مدل سینماتیکی دقیق خودرو مفصلی توسعه یافت و پدیده‌ی قفل‌شدگی مفصل به‌صورت تحلیلی مدل‌سازی شد. سپس مدل به‌وسیله‌ی نرم‌افزار تراک‌سیم اعتبارسنجی گردید تا اطمینان حاصل شود که رفتار حرکتی سیستم با شرایط واقعی منطبق است. برای کاهش پیچیدگی محاسباتی و بهبود همگرایی، فرآیند تنظیم موقعیت به دو فاز مجزا شامل آماده‌سازی اولیه و حرکت رو به جلو برای تنظیم جهت خودرو تقسیم شد تا پیچیدگی یادگیری عملیات کامل کاهش یابد.

دو الگوریتم DDPG و TD3 برای آموزش عامل هوشمند به کار گرفته شدند. نتایج شبیه‌سازی‌ها نشان داد که TD3 به دلیل به‌کارگیری دو شبکه‌ی منتقد و تأخیر در به‌روزرسانی پارامترها، پایداری و دقت بیشتری در فرایند یادگیری دارد. شبکه‌های عصبی با سه تا پنج لایه پنهان طراحی شدند و توابع پاداش اختصاصی برای جلوگیری از برخورد، قفل‌شدگی و انحراف بیش از حد تنظیم گردیدند.

علاوه بر این، تحلیل نتایج نشان داد که روش پیشنهادی منجر به کاهش چشمگیر برخوردها، بهبود پایداری حرکتی و افزایش دقت در رسیدن به موقعیت هدف گردید. سیستم توسعه‌یافته توانست بدون وابستگی به مدل‌های از پیش تعیین‌شده یا قواعد هندسی، استراتژی‌های کنترلی بهینه را از طریق تجربه و تعامل با محیط بیاموزد. این ویژگی، کاربرد چارچوب یادگیری تقویتی عمیق در کنترل و تنظیم موقعیت خودروهای مفصلی واقعی را تقویت می‌کند.

به‌طور کلی، یافته‌های این پژوهش نشان می‌دهد که رویکرد یادگیری تقویتی عمیق، در صورت طراحی دقیق ساختار پاداش و تقسیم فازی یادگیری، می‌تواند جایگزینی مؤثر برای کنترل‌کننده‌های کلاسیک در سیستم‌های خودکار با درجه آزادی بالا باشد.

۶-۲- پیشنهادات

به دلیل محدودیت‌های زمانی، هزینه‌ای و تجهیزاتی، در طراحی و پیاده‌سازی سیستم حاضر چندین ساده‌سازی اعمال شد. با توجه به نتایج به‌دست‌آمده، مسیرهای پژوهشی آینده می‌توانند در جهت گسترش و بهبود عملکرد سیستم به شرح زیر دنبال شوند:

- گسترش مدل دینامیکی: در این پژوهش مدل حرکتی خودرو با فرضیات ساده‌سازی‌شده طراحی شد. در مطالعات آتی، می‌توان مدل را با درجات آزادی بیشتر توسعه داد و اثرات دینامیکی نظیر لغزش چرخ، اصطکاک سطح و تغییرات شرایط جاده را لحاظ کرد. چنین مدلی می‌تواند واقع‌گرایی شبیه‌سازی را افزایش داده و عملکرد عامل یادگیری تقویتی را در شرایط غیرایستا بهبود دهد.
- پیاده‌سازی سخت‌افزاری و آزمون واقعی: توسعه‌ی یک نمونه‌ی فیزیکی از سیستم کامیون–تریلی مجهز به حسگرهای متنوع (دوربین، لایدار و رادار) می‌تواند امکان ارزیابی عملکرد الگوریتم در شرایط واقعی را فراهم کند. این کار به درک دقیق‌تر چالش‌های مرتبط با ادراک محیط، تأخیر حسگر و نویز سیگنال‌ها منجر خواهد شد.
- بهبود و تنظیم الگوریتم TD3: الگوریتم TD3 در این مطالعه عملکرد قابل‌توجهی از خود نشان داد، با این حال، تنظیم دقیق ابرپارامترها، بهبود معماری شبکه‌های بازیگر و منتقد و ترکیب آن با روش‌های بهینه‌سازی تطبیقی می‌تواند سرعت یادگیری و پایداری همگرایی را افزایش دهد.
- مهندسی فضای حالت و طراحی ساختار یادگیری تطبیقی: از آن‌جا که تعریف فضای حالت تأثیر مستقیمی بر کیفیت یادگیری دارد، پیشنهاد می‌شود از روش‌های خودتطبیقی برای فشرده‌سازی و تجرید ویژگی‌ها استفاده شود. این کار می‌تواند موجب کاهش تداخل بین فازهای یادگیری و بهبود تعمیم‌پذیری عامل در سناریوهای متنوع شود.
- تطبیق‌پذیری در محیط‌های پویا: برای انتقال از محیط شبیه‌سازی‌شده به کاربردهای واقعی، نیاز به سازوکار تصمیم‌گیری بلادرنگ وجود دارد. ادغام روش‌های یادگیری تقویتی با کنترل‌کننده‌های کلاسیک مانند PID یا MPC می‌تواند به سیستم‌های هیبریدی منجر شود که ضمن حفظ هوشمندی، از پایداری و واکنش سریع در مواجهه با موانع متحرک یا شرایط اضطراری برخوردار باشند.
- کاربرد در سامانه‌های پارک خودکار: چارچوب دومرحله‌ای ارائه‌شده در این پژوهش می‌تواند پایه‌ای برای توسعه‌ی سیستم‌های پارک خودکار کامل خودروهای مفصلی در محیط‌های واقعی باشد. با افزودن فاز نهایی دنده‌عقب و یکپارچه‌سازی تصمیم‌گیری بین فازها، سیستم می‌تواند فرآیند پارک را به‌صورت انتها به انتها انجام دهد. این رویکرد، گامی مؤثر در جهت استقرار عملی فناوری‌های پارک خودکار در وسایل نقلیه‌ی سنگین هوشمند خواهد بود.

به طور خلاصه، جهت‌گیری‌های زیر برای تحقیقات آینده توصیه می‌شود:

- مدل‌سازی با درجات آزادی اضافی و اثرات لغزش چرخ.
- پیاده‌سازی و آزمایش واقعی جهت ارزیابی عملکرد عامل.
- مطالعات متمرکز بر طراحی فضای حالت برای یادگیری تقویتی.
- ادغام یادگیری تقویتی با کنترل‌کننده‌های کلاسیک برای تصمیم‌گیری هیبریدی.
- استفاده از TD3 برای سناریوهای رانندگی و کنترل پیچیده‌تر.

تأییدیه اخلاقی: محتویات علمی این مقاله حاصل پژوهش نویسندگان است و در هیچ نشریه ایرانی و غیرایرانی منتشر نشده است.

تعارض منافع: در این مقاله از برخی نتایج حاصل از رساله دکتری نویسنده اول استفاده شده است. هیچ تعارض منافع دیگری برای اظهار وجود ندارد.

منابع

[1] H. Rezaei, H. N. Pishkenari, S. Jafari, and E. M. J. Al-Rashed, "Automatic parking of an articulated vehicle using ANFIS," in Proc. IEEE Int. Symp. Ind. Electron. (ISIE), Taipei, Taiwan, 2013, pp. 2519-2524.
[2] R. P. Vieira, E. V. Argento, and T. C. Revoredo, "Trajectory planning for car-like robots through curve parametrization and genetic algorithm optimization with applications to autonomous parking," IEEE Latin Am. Trans., vol. 20, no. 2, pp. 309-316, Feb. 2022, doi: 10.1109/TLA.2022.9661471.
[3] B. Li, T. Acarman, Y. Zhang, Y. Ouyang, C. Yaman, Q. Kong, X. Zhong, and X. Peng, "Optimization-based trajectory planning for autonomous parking with irregularly placed obstacles: A lightweight iterative framework," IEEE Trans. Intell. Transp. Syst., vol. 23, no. 8, pp. 11970-11981, Aug. 2022, doi: 10.1109/TITS.2021.3109011.
[4] H. R. Nedamani, M. Soleymanifard, A. Safaeifar, and P. M. Khiabani, "Soft Computing-Based Driver Modeling for Automatic Parking of Articulated Heavy Vehicles," SAE Int. J. Commer. Veh., vol. 16, no. 03-16-04, pp. 268-281, 2023, doi: 10.4271/02-16-04-0027.
[5] S. Azadi, H. R. Nedamani, and R. Kazemi, "Automatic parking of an articulated vehicle using ANFIS," Global J. Sci. Eng. Technol., vol. 3, pp. 93-104, 2013.
[6] J. Zhang, H. Chen, S. Song, and F. Hu, "Reinforcement learning-based motion planning for automatic parking system," IEEE Access, vol. 8, pp. 154485-154501, Aug. 2020, doi: 10.1109/ACCESS.2020.3017770.
[7] S. Song, H. Chen, H. Sun, and M. Liu, "Data efficient reinforcement learning for integrated lateral planning and control in automated parking system," Sensors, vol. 20, no. 24, p. 7297, Dec. 2020, doi: 10.3390/s20247297.
[8] T. Lei, J. Wang, and Z. Yao, "Modelling and stability analysis of articulated vehicles," Appl. Sci., vol. 11, no. 8, p. 3663, Apr. 2021, doi: 10.3390/app11083663.
[9] B. Saeedi and M. Sadedel, "Implementation of Behavior-Based Navigation Algorithm on Four-Wheel Steering Mobile Robot," Int. J. Adv. Manuf. Technol., vol. 113, no. 3-4, pp. 903-920, Mar. 2021. doi: 10.22059/jcamech.2021.330072.648
[10] A. Molaei and M. Amirkhani, "Distributed Reinforcement Learning Framework for Autonomous Driving on Highways," Iranian Journal of Intelligent Systems, vol. 10, no. 2, pp. 45-59, 2021.
[11] R. Rizehvandi and A. Azadi, "Deep Reinforcement Learning for Highway Driving Decision-Making," Journal of Transportation Systems, vol. 15, no. 3, pp. 112-128, 2024.
[12] A. Adrisi, K. Bagherzadeh Chehreh, and A. Naderi, "Proposing an Innovative Model Based on Reinforcement Learning for Intelligent Vehicle Routing in a Dynamic Network (Case Study: City of Isfahan)," in Proc. Int. Conf. Civil Eng., Archit., Urbanism Contemp. Iran, 2017.
[13] T. T. Li, J. Wang, and Z. Yao, "Nonlinear Dynamic Model for Articulated Vehicles and Stability Analysis," J. Veh. Dyn., vol. 42, no. 1, pp. 56-68, Jan. 2018.
[14] J. Jung, H. Chen, S. Song, and F. Hu, "Side Slip Angle Estimation with Dual Kalman Filters," J. Veh. Control, vol. 25, no. 2, pp. 134-147, Feb. 2019.
[15] J. Doe, H. Chen, S. Song, and F. Hu, "Lyapunov Control for Articulated Systems," IEEE Trans. Control Syst. Technol., vol. 24, no. 5, pp. 1789-1802, Sep. 2016.
[16] K. D. A. van de Vendel, J. Elfring, and A. J. Aertssen, "Path Planning for a Reverse Parking Maneuver using Reinforcement Learning," TU Eindhoven Repository (pure.tue.nl), 2025.
[17] A. Aertssen, R. Huisman, I. Besselink, and M. J. G. R. van de Molengraft, "Real-Time Trajectory Planning for an Autonomous Articulated Commercial Vehicle," in Proc. 43rd Benelux Meeting Syst. Control, 2024, pp. 1-10.
[18] Y. Cheng, X. Hu, K. Chen, X. Yu, and Y. Luo, "Online longitudinal trajectory planning for connected and autonomous vehicles in mixed traffic flow with deep reinforcement learning approach," J. Transp. Eng. A: Syst., vol. 150, no. 7, Jul. 2024, doi: 10.1061/JTEPBS.TEENG-8286.
[19] H. Pei, J. Zhang, Y. Zhang, H. Xu, and L. Li, "Self-organized routing for autonomous vehicles via deep reinforcement learning," IEEE Trans. Veh. Technol., vol. 72, no. 4, pp. 4351-4363, Apr. 2023, doi: 10.1109/TVT.2022.3225808.
[20] F. Ye, S. Zhang, P. Wang, and C.-Y. Chan, "A survey of deep reinforcement learning algorithms for motion planning and control of autonomous vehicles," in Proc. IEEE Intell. Veh. Symp. (IV), Nagoya, Japan, 2021, pp. 1075-1080, doi: 10.1109/IV48863.2021.9575880.
[21] R. Cheng, G. Orosz, R. M. Murray, and J. W. Burdick, "End-to-End Safe Reinforcement Learning through Barrier Functions for Safety-Critical Continuous Control Tasks," arXiv preprint arXiv:1903.08792, 2019.
[22] H. Liu and L. Wang, "TD3 based collision free motion planning for robot navigation," arXiv preprint arXiv:2405.15460, 2024.
[23] T. Xu, Z. Meng, W. Chen, and Z. Li, "End-to-end autonomous driving decision method based on improved TD3 algorithm in complex scenarios," Sensors, vol. 24, no. 15, p. 4962, 2024, doi: 10.3390/s24154962.
[24] Z. Zhang, Y. Luo, Y. Chen, H. Zhao, Z. Ma, and H. Liu, "Automated parking trajectory generation using deep reinforcement learning," arXiv preprint arXiv:2504.21071, 2025.
[25] I. M. Anagnostara, E. Tsardoulias, and A. L. Symeonidis, "Deep reinforcement learning and imitation learning for autonomous parking simulation," Electronics, vol. 14, no. 10, p. 1992, 2025, doi: 10.3390/electronics14101992.
[26] K. Wahdan, N. Ehab, Y. Mansy, and A. El Mougy, "Dynamic path planning for autonomous vehicles using adaptive reinforcement learning TD3 model," in Proc. ICAART, Rome, Italy, 2024, pp. 272-279, doi: 10.5220/0012363300003636.
[27] D. Attard, "Reinforcement learning for autonomous navigation of articulated vehicles," M.Sc. Thesis, University of Malta, 2024.
[28] S. van Dam, L. Wisell, K. Shingade, M. Sadeghi Reineh, and F. Bruzelius, "Safe Control Allocation of Articulated Heavy Vehicles," in Proc. AVEC, LNME, 2024, doi: 10.1007/978-3-031-70392-8_1.
[29] J. Wang, L. Chu, Y. Zhang, Y. Mao, and C. Guo, "Intelligent Vehicle Decision-Making and Trajectory Planning Method Based on Deep Reinforcement Learning in the Frenet Space," Sensors, vol. 23, no. 5, p. 2463, 2023, doi: 10.3390/s23052463.
[30] "Mercedes-Benz Actros 2011 blueprint," Drawing Database, Feb. 2025. [Online]. Available: Drawing Database.
[31] I. Han, "Path Planning for Perpendicular Parking of Large Articulated Vehicles Based on Qualitative Kinematics and Geometric Methods," Autonomous Intelligent Systems, vol. 5, no. 3, p. 48, 2023, doi: 10.3390/vehicles503004.