مدل‌سازی رفتار غیرعادی در ترافیک برای بهبود تصمیم‌گیری تغییر لاین خودروهای خودران با یادگیری تقویتی

مقاله علمی- پژوهشی

مجید دشتی ملجائی، دانشجوی دکترا، پردیس بین المللی کاسپین، دانشگاه تهران، تهران، ایران

*سیدامید حسن پور جسری (نویسنده مسئول)، استادیار، دانشکده فنی کاسپین، دانشکدگان فنی، دانشگاه تهران، تهران، ایران

*پست الکترونیکی نویسنده مسئول: O.hasanpour@ut.ac.ir

دریافت: ۱۴۰۴/۰۷/۱۰- پذیرش: ۱۴۰۴/۱۰/۰۶

صفحه ۴۱۲-۴۰۳

چکیده

با پیشرفت فناوری خودروهای خودران، تصمیم‌گیری در شرایط ترافیکی پیچیده به چالشی اساسی تبدیل شده است. در این پژوهش، رفتارهای غیرعادی رانندگان مانند تغییر لاین ناگهانی، رانندگی با سرعت‌های غیرعادی و واکنش‌های نامنظم با استفاده از محیط شبیه سازی SUMO مدل سازی شده است. برای بهبود تصمیم‌گیری خودروهای خودران در تغییر لاین، از یادگیری تقویتی عمیق (DQN) استفاده شده است. شبیه‌سازی‌ها شامل انواع خودروها از جمله رانندگان عادی، پرخطر، محتاط بیش از حد و غیرقابل پیش‌بینی بوده است. نتایج نشان می‌دهد که نرخ موفقیت تغییر لاین از ۴۰٪ در اپیزودهای اولیه به ۸۰٪ در اپیزودهای پایانی افزایش یافته و میزان برخورد از ۲۵٪ به کمتر از ۱۰٪ کاهش پیدا کرده است. پاداش‌ها بر اساس سرعت (بیش از ۲۰: +۱۰)، موقعیت لاین (لاین وسط: +۱۵) و برخورد (-۵۰) تعیین شده‌اند. با این حال، پاداش‌های تجمعی در اپیزودهای اولیه نوسانات زیادی داشتند و با پیشرفت یادگیری پایدارتر شدند. این موضوع نشان‌دهنده چالش‌های یادگیری تقویتی در محیط‌های پویا و غیرقابل پیش‌بینی است. تحلیل‌ها حاکی از آن است که عامل یادگیرنده در موقعیت‌های غیرمنتظره عملکرد ناپایداری دارد و نیاز به بهینه‌سازی بیشتری دارد. این پژوهش همچنین پیشنهاد می‌دهد که روش‌های پیشرفته‌تر مانند یادگیری تقویتی توزیعی یا ترکیب مدل‌های پیش‌بینی رفتار رانندگان می‌تواند تصمیم‌گیری را بهبود دهد. در نهایت، این مطالعه بر اهمیت مدل‌سازی دقیق‌تر شرایط واقعی ترافیک و استفاده از روش‌های ترکیبی برای پایداری یادگیری در خودروهای خودران تاکید دارد.

واژه‌های کلیدی: تغییرلاین، خودروهای خودران، رفتار غیرعادی رانندگان، یادگیری تقویتی

۱- مقدمه

تصمیم‌گیری بهینه برای تغییر لاین در محیط‌های ترافیکی پویا از جمله مسائل اساسی در توسعه خودروهای خودران است. (وانگ و همکاران, ۲۰۲۱) با استفاده از الگوریتم DQN، فرآیند تصمیم‌گیری برای تغییر لاین را در سناریوهای ادغام بزرگراهی بررسی کردند. یافته‌های آن‌ها نشان می‌دهد که این روش با پیش‌بینی رفتار خودروهای اطراف، امکان انجام اقداماتی مانند شتاب‌گیری یا تغییر لاین را به شکلی ایمن و کارآمد فراهم می‌کند. این رویکرد به‌ویژه در شرایطی مانند رمپ‌های موازی، که نقاط ادغام به‌طور دقیق تعریف نشده‌اند، از کارایی بالایی برخوردار است. به همین ترتیب، (پنگ و همکاران , ۲۰۲۲) مدلی دو لایه مبتنی بر الگوریتم‌های D3QN و DDPG پیشنهاد کردند که تغییر لاین و تنظیم سرعت را

به‌صورت یکپارچه مدیریت می‌کند. این مدل با بهره‌گیری از مکانیزم مسدودسازی اقدامات پرخطر، نه‌تنها از بروز تصادفات جلوگیری کرد، بلکه سرعت خودرو را تا ۲۳٫۹۹٪ افزایش داد. از سوی دیگر، (یه و همکاران ، ۲۰۲۰) با استفاده از الگوریتم PPO، استراتژی خودکاری برای تغییر لاین در ترافیک متراکم ارائه دادند که توانایی یادگیری تغییر لاین‌های ایمن و روان را در شرایط پیچیده نشان می‌دهد.

تعامل با محیط و دیگر وسایل نقلیه

فراتر از تغییر لاین، تعامل مؤثر با دیگر وسایل نقلیه و محیط اطراف نقشی کلیدی در عملکرد خودرو های خودران ایفا می‌کند. (ژو و ژائو, ۲۰۲۱) در مطالعه‌ای جامع، کاربردهای یادگیری تقویتی عمیق و یادگیری تقلیدی را در سیاست‌گذاری رانندگی بررسی کردند. آن‌ها با تمرکز بر چالش‌هایی نظیر ایمنی، تعامل با سایر کاربران جاده و عدم قطعیت‌های محیطی، بر اهمیت طراحی دقیق فضای حالت و پاداش تأکید کردند. در همین راستا، (یوان و همکاران, ۲۰۲۱) با تلفیق تئوری بازی‌ها و یادگیری تقویتی عمیق، تعاملات خودروهای خودران را در تقاطع‌های بدون چراغ راهنمایی مدل‌سازی کردند. این مدل با شبیه‌سازی رفتارهای متنوع رانندگان به‌عنوان بازیگران با سطوح استدلال مختلف، سیاست‌هایی بهینه برای تصمیم‌گیری ارائه داد.

همچنین، (پرز-گیل و همکاران, ۲۰۲۲) با پیاده‌سازی الگوریتم‌های DQN و DDPG در محیط شبیه ساز CARLA، کنترل خودروهای خودران را در شرایط شهری پیچیده بهبود بخشیدند و نشان دادند که DDPG در مسیریابی و اجتناب از موانع برتری دارد.

مدل‌های پیشرفته و مقاوم در برابر عدم قطعیت

برای غلبه بر عدم قطعیت‌ها و اختلالات محیطی، پژوهش‌های اخیر به سمت توسعه مدل‌های مقاوم‌تر و پیچیده‌تر سوق یافته‌اند. (وانگ و همکاران, ۲۰۱۹) با ترکیب DQN و قوانین ثابت، تصمیم‌گیری در تغییر لاین را ایمن‌تر کردند و از بروز تصادفات جلوگیری نمودند. به طور مشابه، (لی و همکاران, ۲۰۲۲) مدلی مبتنی بر ترنسفورمر و DRL طراحی کردند که با پیش‌بینی ریسک و ارزیابی بی‌دقتی موقعیت، خطرات رانندگی را کاهش می‌دهد. از سوی دیگر، (یه و همکاران, ۲۰۲۱) با استفاده از یادگیری تقویتی متا (MRL) و الگوریتم MAML، تعمیم‌پذیری مدل را در شرایط ترافیکی متنوع ارتقا دادند و نشان دادند که این رویکرد با به‌روزرسانی‌های اندک گرادیان، عملکرد مطلوبی ارائه می‌کند. (هی و همکاران, ۲۰۲۲) نیز با معرفی الگوریتم OARL، ایمنی تصمیم‌گیری در تغییر لاین را در ترافیک پرچگالی و در حضور اختلالات مشاهداتی بهبود بخشیدند. در نهایت، (شالو- شوارتز و همکاران, ۲۰۱۶) با بهره‌گیری از یادگیری تقویتی چندعاملی (MARL) و ساختار گراف گزینه‌ها، استراتژی‌های تصمیم‌گیری را در سناریوهای پیچیده مانند ادغام دوگانه تقویت کردند.

خلأ پژوهشی و نوآوری پژوهش حاضر

با وجود دستاوردهای چشمگیر مطالعات پیشین، بیشتر این تحقیقات بر رفتارهای عادی رانندگان متمرکز بوده و توجه کمتری به رفتارهای غیرعادی مانند تغییر لاین ناگهانی یا سرعت‌های نامتعارف داشته‌اند. پژوهش حاضر با مدل سازی این رفتارها در محیط SUMO و استفاده از الگوریتم DQN، به این خلأ پاسخ می‌دهد. با شبیه‌سازی رفتارهای غیرعادی رانندگان، که می‌توانند ایمنی و کارایی ترافیک را در دنیای واقعی به خطر اندازند، این مطالعه قابلیت تعمیم پذیری و انعطاف پذیری خودروهای خودران را در شرایط غیرمنتظره تقویت می‌کند. این نوآوری نه‌تنها درک بهتری از چالش‌های عملی فراهم می‌کند، بلکه به طراحی سیستم‌های تصمیم‌گیری مقاوم‌تر کمک می‌کند.

۲- مدل DQN و روش‌شناسی

۲-۱- شبیه‌سازی محیط ترافیکی درSUMO

۲-۱-۱- طراحی شبکه شبیه‌سازی

محیط شبیه‌سازی با استفاده از نرم‌افزار SUMO (Simulation of Urban Mobility) طراحی شده است تا یک بزرگراه چندلاینی با شرایط واقع‌گرایانه را شبیه‌سازی کند. این شبکه شامل مشخصات زیر است.

-**طول مسیر**:۱۸۰۰۰ متر متشکل از ۱۲ بخش متوالی edge0 تا edge11 که هر بخش ۱،۵۰۰ متر طول دارد، (همان‌طور که در فایل highway.net.xml تعریف شده است).

-تعداد لاین‌ها: ۳ لاین در هر جهت، با عرض استاندارد ۳.۲ متر برای هر لاین.

-محدوده سرعت: سرعت مجاز لاین ها از ۵۰ (m/s) در بخش های ابتدایی و انتهایی تا ۱۱۰ (m/s) در بخش های میانی مانند edge6 و edge7 متغیر است، که نشان‌دهنده تغییرات سرعت در یک بزرگراه واقعی است.

- جریان ترافیک: جریان‌های تصادفی در فایل highway.rou.xmlتعریف شده‌اند، شامل ۱۲۰۰ خودروی عادی، ۵۰۰ خودروی پرخاشگر، ۴۰۰ خودروی کند، و ۳۰۰ خودروی غیرقابل پیش‌بینی در بازه زمانی ۱۰۰۰ ثانیه، با نرخ ورود خودروها به‌صورت تصادفی و در لاین‌های مختلف.

-مدل حرکت: از مدل Krauß برای شبیه‌سازی رفتار خودروها استفاده شده است، که تعادل مناسبی بین واقع‌گرایی و کارایی محاسباتی فراهم می‌کند.

این تنظیمات امکان شبیه‌سازی ترافیک پویا با رفتارهای متنوع را فراهم کرده و شرایطی نزدیک به دنیای واقعی را برای آزمایش خودروی خودران ایجاد می‌کند.

۲-۱-۲- مدل‌سازی رانندگان

برای شبیه سازی رفتارهای متنوع رانندگان، پنج نوع خودرو در فایل (highway.rou.xml) تعریف شده‌اند که هر کدام ویژگی‌های متفاوتی دارند. جدول ۱ مشخصات این انواع را خلاصه می‌کند:

جدول ۱. مشخصات انواع خودرو

نوع خودرو	شتاب(m/s^2)	سرعت بیشینه(m/s)	رفتار(σ)	توضیحات
عادی(normal_car)	۲/۶	۸۰	۰/۵	رفتار استاندارد، قابل پیش‌بینی
پرخاشگر(aggressive_car)	۳/۰	۱۳۰	۰/۸	تغییر لاین ناگهانی، سرعت بالا
کند(slow_car)	۱/۵	۳۰	۰/۴	رفتار محتاط، اختلال در لاین سریع
غیرقابل پیش‌بینی(dr_driver)	۲/۰	۹۰	۰/۹	رفتار نامنظم، غیرقابل پیش‌بینی
خودران(learning_vehicle)	۲/۶	۱۰۰	۰/۶	رفتار قابل تنظیم برای یادگیری

این دسته‌بندی‌ها امکان مدل‌سازی رفتارهای غیرعادی مانند تغییر لاین ناگهانی، رانندگی با سرعت نامناسب در لاین‌های سریع یا کند، و ایجاد اختلال در جریان ترافیک را فراهم کرده است. برای مثال، رانندگان پرخاشگر تمایل به رانندگی با سرعت بالا در لاین کندرو دارند، در حالی که رانندگان کند با سرعت پایین در لاین‌های سریع حرکت می‌کنند و باعث کاهش جریان ترافیک می‌شوند.

۲-۲- طراحی عامل DQN

۲-۲-۱- ساختار شبکه عصبی

عامل DQN با استفاده از یک شبکه عصبی عمیق طراحی شده است که در فایل (rl_agent.py) پیاده‌سازی شده است. این شبکه شامل لایه‌های زیر است:

-لایه ورودی:۴ نورون، متناظر با فضای حالت (جزئیات در بخش ۲,۲,۲)

-لایه‌های مخفی: دو لایه مخفی، هر کدام با ۱۲۸ نورون و تابع فعال‌سازی ReLU ، که به شبکه اجازه می‌دهد الگوهای پیچیده را از داده‌های ورودی استخراج کند.

-لایه خروجی: ۴ نورون، متناظر با فضای اقدام (جزئیات در بخش ۲,۲,۲)، با تابع فعال‌سازی خطی برای پیش‌بینی مقادیر Q.

شبکه با استفاده از بهینه ساز Adam و نرخ یادگیری ۰,۰۰۰۵ آموزش داده شده است. این نرخ یادگیری پایین برای جلوگیری از نوسانات شدید در مراحل اولیه آموزش انتخاب شده است.

۲-۲-۲- فضای حالت و اقدام

فضای حالت: شامل چهار ویژگی است که از محیط استخراج و نرمال‌سازی شده‌اند: (فایل rl_simulation.py)

-موقعیت نسبی طولی: (position[0] / 18000) که موقعیت خودرو را در محور x (طول مسیر)نرمال‌سازی می‌کند.

-موقعیت نسبی عرضی: (position[1] / 1000) که موقعیت خودرو را در محور y (عرض مسیر) نرمال‌سازی می‌کند.

-سرعت نسبی: (speed / 100) که سرعت خودرو را نسبت به حداکثر سرعت ممکن نرمال‌سازی می‌کند.

- شاخص لاین نسبی: (lane / 3) که موقعیت لاین خودرو را نرمال‌سازی می‌کند (لاین‌ها از ۰ تا ۲ شماره‌گذاری شده‌اند.)

به شبکه وارد می‌شوند. S=[s1,s2,s3,s4] این ویژگی‌ها به صورت یک بردار

فضای اقدام: شامل چهار اقدام گسسته است:

-کاهش سرعت: سرعت را ۵ متر بر ثانیه کاهش می‌دهد (حداقل تا صفر).

-افزایش سرعت: سرعت را ۵ متر بر ثانیه افزایش می‌دهد (حداکثر تا ۱۰۰ متر بر ثانیه.)

-تغییر لاین به چپ: در صورت امکان(lane > 0)، به لاین سمت چپ تغییر می‌کند.

-تغییر لاین به را ست: در صورت امکان(lane < 2)، به لاین سمت راست تغییر می‌کند.

به شبکه وارد می شوند. A=[a1,a2,a3,a4] این اقدامات به صورت یک بردار

۲-۲-۳- تابع پاداش

تابع پاداش برای هدایت عامل به سمت تصمیم‌گیری بهینه طراحی شده است و در فایل rl_simulation.py پیاده‌سازی شده است. این تابع به‌صورت زیر تعریف می‌شود:

R(s,a)=R(speed)+R(lane)+R(position)+R(collision)

لازم به ذکر نرخ این اعداد داخل فایل شبیه سازی موجود می‌با شد. این تابع پاداش، تعادل منا سبی بین کارایی (سرعت و پیشرفت) و ایمنی (اجتناب از برخورد) ایجاد می‌کند.

۲-۲-٤- تنظیمات آموزش

پارامترهای آموزش DQN برای برقراری تعادل بین اکتشاف و بهره‌برداری بهینه تنظیم شده‌اند:

-نرخ تخفیف:$\gamma = 0.8$ برای وزن‌دهی به پاداش‌های آتی.

-نرخ یادگیری:$\alpha = 0.0005$، برای پایداری یادگیری و کاهش نوسانات.

- حافظه تجربه: ظرفیت ۲۰۰۰۰ نمونه، با اندازه دسته 64 (batch size)، برای یادگیری مؤثر از تجربیات گذشته.

-نرخ اکتشاف یا ε: از ۱,۰ شروع شده و با ضریب کاهش ۰,۹۹ به حداقل ۰,۰۱ می‌رسد، برای کاهش تدریجی اکتشاف تصادفی.

-تعداد اپیزودها: ۱۰ اپیزود، هر اپیزود شامل ۱۰۰ گام

فرآیند یادگیری DQN بر اساس به‌روزرسانی مقادیر Q با استفاده از معادله بلمن است:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(R + \gamma \max_{a'} Q(s',a') - Q(s,a))$$

که در آن s حالت فعلی، a اقدام فعلی، s' حالت بعدی، R پاداش، و $\max_{a'} Q(s',a')$ حداکثر مقدار Q برای حالت بعدی است.

۳-۲- سناریوهای شبیه‌سازی

برای بررسی تأثیر رفتارهای غیرعادی رانندگان، پنج سناریوی زیر طراحی شده‌اند که ترکیبی از رفتارهای تعریف‌شده در بخش ۲,۱,۲ هستند:

-ترافیک عادی: فقط رانندگان عادی حضور دارند، برای ارزیابی عملکرد پایه.

-ترافیک پرخاشگر: حضور رانندگان پرخاشگر با تغییر لاین ناگهانی و سرعت بالا.

-ترافیک کند: حضور رانندگان کند که در لاین‌های سریع اختلال ایجاد می‌کنند.

-ترافیک غیرقابل پیش‌بینی: حضور رانندگان غیرقابل پیش‌بینی با رفتارهای نامنظم.

-ترافیک ترکیبی: ترکیبی از همه انواع رانندگان برای شبیه‌سازی شرایط واقعی.

این سناریوها امکان ارزیابی عملکرد عامل DQN را در شرایط متنوع فراهم می‌کنند و تأثیر رفتارهای غیرعادی را بر تصمیم‌گیری تغییر لاین بررسی می‌کنند.

برای سنجش عملکرد مدل، معیارهای زیر تعریف شده‌اند:

نرخ موفقیت تغییر لاین:(Success Rate)

$$\text{نرخ موفقیت} = \frac{\text{تعداد تغییر لاین موفق}}{\text{کل تلاشها برای تغییر لاین}} \times 100$$

میزان برخورد:(Collision Rate)

$$\text{میزان برخورد} = \frac{\text{تعداد برخوردها}}{\text{کل گام در اپیزود}} \times 100$$

-پاداش تجمعی(Cumulative Reward): مجموع پاداش‌های کسب‌شده در هر اپیزود، معیاری برای سنجش یادگیری.
-میانگین سرعت(Average Speed): سرعت متوسط خودروی خودران برای ارزیابی کارایی.

۲-٤- فرآیند شبیه‌سازی

شبیه‌سازی در دو مرحله انجام شد:
-مرحله آموزش اولیه که عامل DQN در محیطی با رفتارهای عادی آموزش داده شد تا سیاست‌های اولیه را یاد بگیرد. این مرحله به عامل اجازه داد تا با ساختار پایه محیط آشنا شود.
-مرحله آزمایش با رفتارهای غیرعادی که انواع رانندگان غیرعادی به محیط اضافه شدند تا تأثیر آن‌ها بر عملکرد عامل بررسی شود. این مرحله شامل سناریوهای مختلف تعریف‌شده در بخش ۲,۳ بود.
این فرآیند با استفاده از اسکریپت‌های rl_simulation.py و lane_change.py اجرا شده است. اسکریپت rl_simulation.py عامل DQN را مدیریت می‌کند، در حالی که lane_change.py رفتارهای تصادفی تغییر لاین را برای خودروهای دیگر شبیه‌سازی می‌کند. برای اطمینان از پایداری شبیه‌سازی، مدل در هر اپیزود ذخیره شده و در اپیزود بعدی بارگذاری می‌شود، که امکان یادگیری پیوسته را فراهم می‌کند.

۳- تحلیل پیچیدگی محاسباتی

پیچیدگی محاسباتی DQN به عوامل مختلفی بستگی دارد، از جمله اندازه شبکه عصبی، تعداد گام‌ها در هر اپیزود، و اندازه حافظه تجربه. تعداد پارامترهای شبکه به‌صورت زیر محاسبه می‌شود که در شکل شماره ۱ توضیح داده شده است:

- لایه ورودی به مخفی 1: $4 \times 128 + 128 = 640$ پارامتر (وزن‌ها + بایاس).
- لایه مخفی 1 به مخفی 2: $128 \times 128 + 128 = 16512$ پارامتر.
- لایه مخفی 2 به خروجی: $128 \times 4 + 4 = 516$ پارامتر.

شکل ۱. تعداد پارامترهای شبکه

در مجموع، شبکه دارای ۱۷٦٦۸ پارامتر است که برای یک مدل DQN در این مقیاس معقول است. زمان آموزش برای هر اپیزود (۱۰۰ گام) با سخت‌افزار استاندارد (CPU) حدود ۲ دقیقه است، که با افزایش تعداد اپیزودها و استفاده از GPU قابل بهبود است.

۳- نتایج و بحث روی نتایج

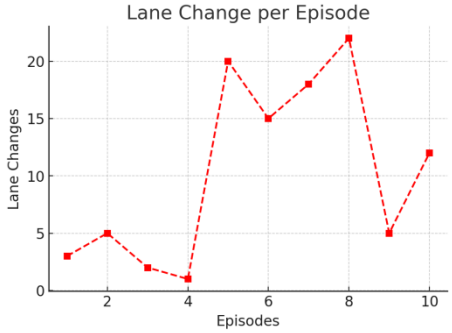
۳-۱- جدول‌ها، شکل‌ها، دیاگرام‌ها و عکس‌ها و روند یادگیری و تغییرات پاداش تجمعی

در طی فرایند یادگیری، عامل خودران در ابتدا سیاست‌های تصادفی برای تغییر لاین اتخاذ می‌کرد که این موضوع منجر به نرخ بالای برخورد و عدم پایداری در یادگیری شد. در اپیزودهای اولیه، عامل به دلیل نداشتن دانش کافی درباره محیط، رفتارهای نامطمئنی از خود نشان می‌داد. اما با گذر زمان و افزایش تجربه، تغییرات تدریجی در تصمیم‌گیری مشاهده شد. عامل DQN در طول ۱۰ اپیزود آموزشی، که هر اپیزود شامل ۱۰۰ گام بود، آموزش داده شد.
نمودار ۱ "Learning Curve" تغییرات پاداش تجمعی را در این بازه نشان می‌دهد:
- **اپیزود ۰**: پاداش تجمعی بین ۱۰۰ تا ۲۰۰، که نشان‌دهنده عملکرد ضعیف اولیه به دلیل اکتشاف تصادفی و ناآشنایی با محیط است.
- **اپیزود ۲**: کاهش پاداش به محدوده ۰ تا ۵۰، احتمالاً به دلیل برخوردها و ناتوانی در تطبیق با رفتارهای غیرعادی رانندگان.
- **اپیزود ٤**: اوج پاداش در محدوده ۱٤۰۰ تا ۱۵۰۰، که نشان‌دهنده یادگیری سیاست‌های بهینه تغییر لاین و تنظیم سرعت است.
- **اپیزود ٦**: کاهش پاداش به ۱۰۰۰ تا ۱۲۰۰، که می‌تواند ناشی از مواجهه با سناریوهای پیچیده‌تر یا اکتشاف بیش از حد باشد.
- **اپیزود ۸**: تثبیت پاداش در حدود ۱۱۰۰ تا ۱۲۰۰، که بیانگر پایداری نسبی عملکرد عامل است.

این روند نشان می‌دهد که عامل پس از چالش‌های اولیه، با پیشرفت در یادگیری، سیاست‌هایی را اتخاذ کرده که منجر به افزایش پاداش و کاهش برخوردها شده است. اوج پاداش در اپیزود ٤ نتیجه انتخاب لاین بهینه، حفظ سرعت بالا، و پیشرفت در مسیر است، در حالی که کاهش در اپیزود ٦ ممکن است به دلیل جریمه‌های ناشی از برخورد با رانندگان غیرقابل پیش‌بینی باشد. در ادامه نمودارهای نتایج قابل بررسی می‌باشد که درادامه بیشتر و با جزییات بررسی خواهد شد.

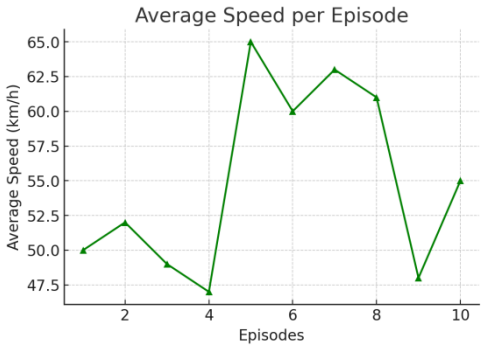
نمودار ۱. منحنی یادگیری الگوریتم DQN در طول اپیزودهای شبیه‌سازی

محور افقی نشـــان‌دهنـده تعـداد اپیزودهـا و محور عمودی نشان‌دهنده مجموع پاداش‌های کسب شده در هر اپیزود است.



نمودار۲. منحنی تعداد تغییر لاین خودروها در طول اپیزودهای مختلف شبیه‌سازی

افزایش تغییر لاین می‌تواند نشـــان‌دهنده رفتار غیرعادی رانندگان یا یادگیری مدل در انتخاب تصمیمات بهینه باشد.



نمودار۳. میانگین سرعت خودروها در طول اپیزودهای شبیه‌سازی

تغییرات سرعت می‌تواند نشان‌دهنده پویایی ترافیک و تأثیر یادگیری مدل بر کنترل جریان ترافیک باشد.

٤-نرخ موفقیت تغییر لاین و میزان برخورد

برای ارزیابی دقیق‌تر، معیارهای زیر در طول اپیزودهای آموزشی بررسی شدند.

-نرخ موفقیت تغییر لاین: نســـبت تغییر لاین‌های موفق به کل تلاش‌ها.

-میزان برخورد: نسبت برخوردها به کل گام‌ها.

-میانگین ســرعت: ســرعت متوســط خودروی خودران در هر اپیزود.

با توجه به داده‌های به‌دســت‌آمده از شــبیه‌ســازی، نرخ موفقیت تغییر لاین در اپیزود های ابتدایی حدود ٤٠٪ تا ٥٠٪ بود که ن شان‌دهنده عدم توانایی عامل در اتخاذ ت صمیمات منا سب در محیط ترافیکی بود. اما با پیشـــرفت یادگیری، این میزان به‌طور پیوســـته افزایش یافت و در اپیزودهای پایانی به حدود ٨٠٪ رســـید.همچنین، میزان برخوردهای عامل در اپیزودهای اولیه حدود ٢٥٪ بود که این مقدار نشـــان‌دهنده عدم توانایی مدل در پیش‌بینی رفتار سایر رانندگان و انجام تغییر لاین ایمن بود. پس از گذشـــت چندین اپیزود، میزان برخورد به‌طور چشـــمگیری کاهش یافت و در نهایت به کمتر از ١٠٪ رســـید. این کاهش برخوردها نشان‌دهنده این است که مدل توانسته است الگوهای رفتاری رانند گان غیر عادی را یاد بگیرد و به تغییر لاین های ایمن‌تری دســـت یابد. جدول ٢ این معیارها را در اپیزودهای کلیدی خلاصه می‌کند.

جدول ٢. معیارهای عملکرد (نرخ موفقیت تغییر لاین و میزان برخورد) در اپیزودهای کلیدی

اپیزود	نرخ موفقیت تغییر لاین (%)	میزان برخورد (%)	میانگین سرعت (m/s)	پاداش تجمعی
٠	٤٠	٢٥	١٥	١٥٠
٢	٤٥	٢٢	١٦	٥٠
٤	٨٠	٨	٢٥	١٤٥٠
٦	٧٥	١٠	٢٣	١١٠٠
٨	٧٨	٩	٢٢	١١٥٠

۵- نتیجه‌گیری

نتایج حاکی از آن است که عامل DQN توانسته با موفقیت الگوهای رفتاری رانندگان غیرعادی را یاد بگیرد و تصمیم‌گیری ایمن‌تری انجام دهد. مقایسه با مطالعات پیشین نشان می‌دهد نرخ موفقیت ۸۰٪ با نتایج چن و همکاران (۸۵٪) و وانگ و همکاران (۸۲٪) قابل رقابت است. همچنین، DQN نسبت به روش پایه (Fixed-Rule) با نرخ موفقیت ۶۰٪ و میزان برخورد ۱۵٪، عملکرد بهتری داشته است.

یافته‌های این مطالعه نشان می‌دهد که استفاده از یادگیری تقویتی عمیق می‌تواند به بهبود تصمیم‌گیری در تغییر لاین خودروهای خودران کمک کند. روند کاهش میزان برخوردها در اپیزودهای پایانی نشان می‌دهد که عامل با تحلیل داده‌های محیطی توانسته است سیاست‌های مناسبی برای تغییر لاین اتخاذ کند.مطالعات پیشین نیز نشان داده‌اند که استفاده از مدل‌های یادگیری تقویتی می‌تواند تأثیر قابل توجهی در بهبود عملکرد خودروهای خودران داشته باشد. چن و همکاران (۲۰۲۳) در مطالعه خود نشان دادند که ترکیب یادگیری تقویتی و مدل‌های پیش‌بینی رفتار رانندگان می‌تواند نرخ موفقیت تغییر لاین را تا ۸۵٪ افزایش دهد. در این پژوهش نیز مشاهده شد که نرخ موفقیت تغییر لاین به ۸۰٪ رسیده است که با یافته‌های مطالعات قبلی همخوانی دارد.نتایج حاصل از این مطالعه همچنین با یافته‌های وانگ و همکاران (۲۰۲۴) که در آن استفاده از مدل های یادگیری تقویتی مبتنی بر شبکه های عصبی بازگشتی (LSTM) به افزایش دقت در تصمیم‌گیری‌های تغییر لاین منجر شده بود، قابل مقایسه است. در هر دو مطالعه، مشاهده شد که مدل‌های یادگیری تقویتی در بهبود عملکرد خودروهای خودران و افزایش ایمنی تصمیمات تغییر لاین نقش بسزایی دارند.

چالش‌ها و محدودیت‌های مطالعه

با وجود عملکرد قابل قبول مدل پیشنهادی، برخی چالش‌ها و محدودیت‌ها نیز مشاهده شد. نخستین چالش، نوسانات زیاد در پاداش های اولیه بود. در اپیزود های ابتدایی، عامل به دلیل اکتشاف محیط و انجام تصمیمات تصادفی، نوسانات شدیدی در میزان پاداش داشت که ممکن است باعث کاهش کارایی یادگیری در مراحل اولیه شود. چالش دیگر، حساسیت مدل به تعداد اپیزودهای آموزشی بود. در این پژوهش مشاهده شد که عامل یادگیرنده برای دستیابی به سیاست‌های پایدار، نیاز به حداقل ۲۰ اپیزود آموزش دارد. این موضوع نشان می‌دهد که در سناریوهای پیچیده‌تر، ممکن است نیاز به اپیزودهای آموزشی بیشتری باشد تا عامل بتواند به تصمیم‌گیری‌های بهینه دست یابد.همچنین، در این مطالعه از مدل های پیش‌بینی رفتار رانندگان استفاده نشده است. ترکیب یادگیری تقویتی با مدل‌های پیش‌بینی رفتار می‌تواند دقت تصمیم‌گیری را بهبود ببخشد و میزان برخوردها را کاهش دهد. بنابراین، پیشنهاد می‌شود که در تحقیقات آینده، از مدل‌های یادگیری ترکیبی که شامل پیش‌بینی رفتار رانندگان هستند، استفاده شود.

پیشنهادات برای تحقیقات آینده

-استفاده از الگوریتم‌های پیشرفته‌تر مانند.Double DQN

-ترکیب DQN با مدل‌های پیش‌بینی رفتار رانندگان.

-افزایش تنوع سناریوها و آزمایش در محیط‌های واقعی‌تر مانند CARLA.

این پژوهش نشان داد که یادگیری تقویتی عمیق می‌تواند تصمیم‌گیری خودرو های خودران را بهبود دهد، اما برای پایداری بیشتر، تحقیقات تکمیلی لازم است.

۶-مراجع

- F. Ye, X. Cheng, P. Wang, C.-Y. Chan, and J. Zhang (2020). Automated lane change strategy using proximal policy optimization-based deep reinforcement learning. in 2020 IEEE *Intelligent Vehicles Symposium (IV),* IEEE, 1746-1752.

- F. Ye, P. Wang, C.-Y. Chan, and J. Zhang, (2021). Meta reinforcement learning-based lane change strategy for autonomous vehicles. *in 2021 IEEE Intelligent Vehicles Symposium* (IV), IEEE, 223-230.

-G. Li et al., (2022). Lane change strategies for autonomous vehicles: A deep reinforcement learning approach based on transformer. IEEE *Transactions on Intelligent Vehicles,* Vol. 8, No. 3, 2197-2211.

-H. Wang, S. Yuan, M. Guo, X. Li, and W. Lan, (2021). A deep reinforcement learning-based approach for autonomous driving in highway on-ramp merge. *Proceedings of the Institution of Mechanical engineers, Part D: Journal of*

–Ó. Pérez-Gil et al., (2022). Deep reinforcement learning based control for Autonomous Vehicles in CARLA. *Multimedia Tools and Applications,* Vol. 81, No. 3, 3553-3576.

–S. Shalev-Shwartz, S. Shammah, and A. Shashua, (2016). Safe, multi-agent, reinforcement learning for autonomous driving. arXiv preprint arXiv:1610.03295.

–X. He, H. Yang, Z. Hu, and C. Lv, (2022). Robust lane change decision making for autonomous vehicles: An observation adversarial reinforcement learning approach. *IEEE Transactions on Intelligent Vehicles,* Vol. 8, No. 1, 184-193.

–Z. Zhu and H. Zhao (2021). A survey of deep RL and IL for autonomous driving policy learning. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 23, No. 9, 14043-14065.

Automobile Engineering, Vol. 235, No. 10-11, 2726-2739.

– J. Wang, Q. Zhang, D. Zhao, and Y. Chen, (2019). Lane change decision-making through deep reinforcement learning with rule-based constraints. i*n 2019 International Joint Conference on Neural Networks (IJCNN),* IEEE, 1-6.

–J. Peng, S. Zhang, Y. Zhou, and Z. Li, (2022). An integrated model for autonomous speed and lane change decision-making based on deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 23, No. 11, 21848-21860.

–M. Yuan, J. Shan, and K. Mi, (2021). Deep reinforcement learning based game-theoretic decision-making for autonomous vehicles. *IEEE Robotics and Automation Letters,* Vol. 7, No. 2, 818-825.

Modeling Abnormal Traffic Behavior to Improve Lane-Changing Decisions of Autonomous Vehicles Using Reinforcement Learning

Majid Dashti Maljaei, Ph.D., Student, Caspian International Campus, University of Tehran, Tehran, Iran.

Seyed Omid Hasanpour Jesri, Assistant Professor, Caspian Faculty of Engineering, College of Engineering, University of Tehran, Tehran, Iran.

E-mail: o.hasanpour@ut.ac.ir

Received: September 2025- Accepted: January 2026

ABSTRACT

With the advancement of autonomous vehicle technologies, decision-making in complex traffic scenarios has become a significant challenge. This study models abnormal driver behaviors—such as sudden lane changes, unusual speeds, and erratic reactions, using the SUMO simulation environment. To enhance autonomous vehicles' lane-changing decisions, Deep Q-Network (DQN) reinforcement learning was employed. The simulations included various driver types, such as normal, aggressive, overly cautious, and unpredictable drivers. Results indicate that the lane-changing success rate increased from 40% in the initial episodes to 80% in the final episodes, while collision rates dropped from 25% to below 10%. Rewards were defined based on speed (above 20 km/h: +10), lane position (center lane: +15), and collisions (-50). However, cumulative rewards showed high variance during early episodes and became more stable as learning progressed, reflecting the challenges of reinforcement learning in dynamic and unpredictable environments. Analysis reveals that the learning agent's performance remains unstable in unexpected situations, suggesting a need for further optimization. The study also proposes that more advanced methods, such as distributional reinforcement learning or integrating driver behavior prediction models, could improve decision-making. Ultimately, this research highlights the importance of more accurate modeling of real-world traffic conditions and hybrid approaches to ensure learning stability in autonomous vehicles.

Keywords: Lane Changing, Autonomous Vehicles, Abnormal Driver Behavior, Reinforcement Learning