چارچوبی جامع مبتنی بر یادگیری تقویتی برای کنترل سیگنال‌های ترافیکی در محیط‌هایV2X

نویسندگان :

Maonan Wang ∗, Xi Xiong †, Yuheng Kan, Chengcheng Xu ‡, Man-On

Wang, Xi Xiong, Yuheng Kan, Chengcheng Xu, Man On Pun

: IEEE Transactions on Vehicular Technology, 2024. DOI / (IEEE Xplore via DOI): https://doi.org/10.1109/TVT.2024.3403879.) IEEE Transactions on Vehicular Technology

6.1-8.6 : impact factor

چکیده

ترافیکِ متراکم یکی از چالش‌های مزمن در نواحی شـهری اسـت که نیازمند توسـعه سـامانه‌های مؤثر کنترل سـیگنال ترافیکی (TSC)[1] می‌باشـد. در حالی که روش‌های مبتنی بر یادگیری تقویتی (RL)[2] نشـان داده‌اند می‌توانند عملکرد TSC را بهبود بخشـند، تعمیم‌پذیری این روش‌ها میان تقاطع‌هایی با سـاختارهای متفاوت همچنان چالش‌برانگیز اسـت. در این کار یک چارچوب عمومی مبتنی بر RL برای محیط‌های خودرو به همه چیز (V2X)[3] ارائه می‌شود. چارچوب پیشـنهادی طراحی نوآورانه‌ای برای عامل (agent) معرفی می‌کند که ماتریس تقاطع (junction matrix) را برای توصـیف وضـعیت تقاطع وارد می‌سـازد، و بدین ترتیب مدل پیشـنهادی برای تقاطع‌های متنوع قابل‌اعمال می‌شـود. به‌منظور توانمندسـازی بیشـتر چارچوب RL پیشـنهادی برای مقابله با سـاختارهای مختلف تقاطع، روش‌های افزایش داده (augmentation) ویژه‌ای برای حالت ترافیکی مناسـب سـیستم‌های کنترل چراغ راهنمایی طراحی شـده‌اند. در نهایت، نتایج تجربی گسـترده حاصـل از پیکربندی‌های مختلف تقاطع، مؤثربودن چارچوب پیشـنهادی را تأیید می‌کنند. کد منبع این کار در دسـترس اسـت : https://github.com/wmn7/Universal_Light

واژگان کلیدی: کنترل سیگنال ترافیکی · مدل‌های عمومی · یادگیری تقویتی · افزایش حالت ترافیکی

[1] Traffic Signal Control
[2] Reinforcement learning
[3] Vehicle to everything

مقدمه

ترافیکِ متراکم به‌عنوان چالشی حیاتی در نواحی شهری مطرح است که منجر به اتلاف زمان شهروندان، مصرف سوخت بیش از حد و افزایش انتشار گازهای گلخانه‌ای می‌شود [۱]. برای کاهش تراکم، روش‌های سنتی کنترل سیگنال ترافیکی مانند کنترل با چرخه ثابت [۲]، روش وبستر [۳]و کنترل خودسازمان‌دهنده چراغ‌ها (SOTL) [۴] توسعه یافته‌اند. با این حال، با رشد شهرها، این رویکردهای سنتی اغلب در مواجهه با افزایش حجم ترافیک و شرایط جاده‌ای پویا ناکافی ظاهر می‌شوند [۵]. ظهور فناوری‌های V2X در سامانه‌های حمل‌ونقل نویدبخش راه‌حل‌هایی است که امکان برقراری ارتباط و تبادل داده میان وسایل نقلیه، زیرساخت‌ها و سایر کاربران جاده‌ای را فراهم می‌آورد [۶]. با استفاده از داده‌های بلادرنگ استخراج‌شده از وسایل نقلیه، زیرساخت‌های مدیریت ترافیک می‌توانند حرکت وسایل نقلیه و عابران را در تقاطع‌ها به‌صورت پویا و بر اساس شرایط جاری ترافیک تنظیم کنند .[۷] برای کنترل چراغ‌ها بر اساس شرایط ترافیکی لحظه‌ای، روش‌های متعددی مبتنی بر RL پیشنهاد شده‌اند. این روش‌ها معمولاً از سه رویکرد اصلی برای تنظیم چراغ‌ها استفاده می‌کنند: «انتخاب فاز بعدی»، «نگه‌داشتن یا تغییر فاز» و «تنظیم مدت‌زمان فاز». به‌طور مشخص، در رویکرد «انتخاب فاز بعدی»، عامل RL فاز بعدی را تعیین می‌کند و اجازه می‌دهد توالی فازها به‌جای ثابت بودن، انعطاف‌پذیر باشد [۸–۲۲]. هرچند این رویکرد انعطاف‌پذیری می‌آورد، ممکن است برای رانندگان نامأنوس یا گیج‌کننده به‌نظر برسد زیرا انتخاب فاز می‌تواند ظاهراً تصادفی شود که منجر به افزایش خطرات حوادث می‌گردد. در رویکرد «نگه‌داشتن یا تغییر فاز»، عامل RL تصمیم می‌گیرد که آیا فاز کنونی را حفظ کند یا تغییر دهد [۲۳–۲۵]. نهایتاً، در رویکرد «تنظیم مدت‌زمان فاز»، عامل طول مدت فاز جاری را از میان مجموعه‌ای از گزینه‌های از پیش تعیین‌شده انتخاب می‌کند [۲۶–۲۸]. از طریق تعامل مستقیم با محیط، عامل RL یاد می‌گیرد که بر مبنای تجربیات واقعی‌زمان، خود را با تغییرات شرایط ترافیک وفق دهد.

با وجود پیشرفت‌های قابل‌توجه روش‌های RL مطرح‌شده، یک محدودیت اصلی این است که بسیاری از آن‌ها برای ساختارهای تقاطع مشخص طراحی شده‌اند. به‌عبارت دیگر، این مدل‌های RL باید هنگام مواجهه با تقاطع‌هایی که از حیث تعداد راه‌های ورودی، خطوط و فازها متفاوت‌اند، از نو طراحی و دوباره آموزش داده شوند که این امر مستلزم صرف منابع قابل‌توجهی برای جمع‌آوری داده، توسعه مدل و آزمایش است [۲۹]. بنابراین توسعه مدل‌های عمومی که بتوان آن‌ها را به آسانی در طیف وسیعی از تقاطع‌ها به‌کار گرفت، اهمیت دارد تا پیاده‌سازی V2X با نیاز کمتر به سفارشی‌سازی و توسعه مجدد در هر تقاطع، به‌طور مؤثر مقیاس‌پذیر شود [۳۰]

در ادبیات موضـــوع، چندین مدل تعمیم‌یافته برای انواع مختلف تقاطع پیشـــنهاد شـــده‌اند [۱۰–۱۲, ۱۸]. با وجود عملکرد خوب آن‌ها، این مدل‌های تعمیم‌یافته تنها برای پیکربندی‌های مشـخصـی که در طراحی آن‌ها مورد نظر قرار گرفته‌اند قابل‌استفاده‌اند و هنگام مواجهه با پیکربندی‌های ناآشنا دچار افت عملکرد می‌شوند. با توجه به این انگیزه‌ها، ما چارچوبی مبتنی بر یادگیری تقویتی با تقویت حالت ترافیکیِ عمومی [1](UniTSA) ارائه می‌کنیم. چارچوب پیشـنهادی امکان آموزش یک عامل عمومی با اسـتفاده از داده‌های افزوده‌شـده را برای کنترل سـیگنال ترافیکی فراهم می‌سـازد. برای رسـیدگی به تقاطع‌هایی با سـاختارهای متنوع، ماتریس تقاطع توسـعه داده شـده است تا وضعیت تقاطع را توصیف کند. با استفاده از این ماتریس نوآورانه، تقاطع‌های دارای ساختارهای مختلف را می‌توان با ماتریس‌هایی از همان اندازه نمایش داد. علاوه بر این، در طراحی عمل (action) در این کار، رویکرد «نگه‌داشـتن یا تغییر فاز فعلی» به‌کار گرفته شـده است تا سـاختار مدل در سـراسـر تقاطع‌های مختلف حفظ شـود. برای مقابله با تقاطع‌های نامشـخص(unseen) ، پنج روش افزایشـــی حالت ترافیکی توسـعه یافته‌اند تا جمع‌آوری داده‌های عامل را در طول فرایند آموزش غنی سازند. این روش‌های افزایش، آموزش جامع‌تری فراهم می‌آورند که در نهایت به عملکرد بهتر در تقاطع‌هایی منجر می‌شـود که پیکربندی آن‌ها در مجموعه آموزشـی وجود نداشـته است. علاوه بر این، سازگارسازی کم‌رتبه (Low-Rank Adaptation — LoRA) [۳۱] برای انجام بهینه سازی بیشتر مدل در تقاطع‌های حیاتی به‌کار گرفته می‌شـــود. در نهایت، آزمایش‌های گســـترده‌ای با اســـتفاده از پلتفرم [2](SUMO)بر روی تقاطع‌هایی با شـمار مختلف راه‌های نزدیک‌شـونده، خطوط و فازها انجام گرفت. نتایج تجربی نشان می‌دهد مدل UniTSA پیشنهادی عملکرد بسیار خوبی حتی در تقاطع‌های ناآشنا ارائه می‌دهد.

مشارکت‌های ما را می‌توان به‌صورت زیر خلاصه کرد:

الف) یک چارچوب تطبیقی کنترل سـیگنال ترافیکی به نام UniTSA برای محیط‌های V2X پیشـنهاد شـده اسـت که بر پایه یک مدل RL عمومی ساخته شده و طراحی عامل نوینی دارد که می‌تواند ساختارهای متنوع تقاطع را مدیریت کند. علاوه بر این، یک مکانیزم فاین‌تیون برای ارتقای عملکرد در تقاطع‌های کلیدی طراحی شده است؛

ب) روش‌های افزایش حالت ترافیکی برای چارچوب TSC پیشـنهادی توسـعه یافته‌اند تا درک عامل از تقاطع‌های مـتـنـوع را افـزایـش داده و عـمـلـکـرد را در هـر دو مـجـمـوعـه آمـوزش و آزمـون بـهـبـود بـخشـــنـــد؛

ج) آزمایش‌های گسـترده روی ۱۲ تقاطع با سـاختارهای گوناگون نشـان می‌دهد که مدل UniTSA پیشـنهادی به‌طور قابل‌توجهی از مدل‌های عمومی مرسوم بهتر عمل می‌کند. افزون بر این، برای یک تقاطع جدید، UniTSA

[1] Universal Traffic State Augmentation
[2] Simulation of Urban MObility

می‌تواند مدل پیش‌آموزش‌یافته خود را با فاین‌تیون کردن به‌سرعت تطبیق دهد و با زمان آموزش به‌مراتب کمتر نسبت به آموزش یک مدل نو از ابتدا، عملکردی قابل‌مقایسه یا بهتر به‌دست آورد.

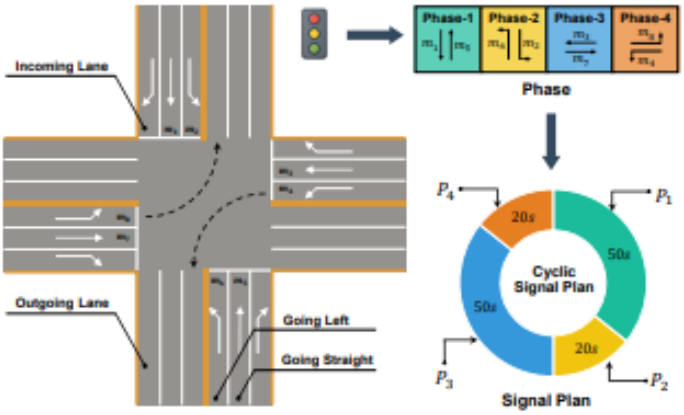
باقی ساختار مقاله به‌صورت زیر است: بخش ۲ مروری بر کارهای مرتبط با TSC ارائه می‌دهد، در حالی که بخش ۳ اصطلاحات مرتبط با راه و چراغ‌های ترافیکی را معرفی می‌کند. سپس بخش ۴ چارچوب UniTSA پیشنهادی و پنج روش افزایش حالت ترافیکی را مطرح می‌سازد و بخش ۵ تنظیمات تجربی و نتایج UniTSA را تشریح می‌کند. در نهایت بخش ۶ مقاله را جمع‌بندی می‌کند.

۲) کارهای مرتبط

پژوهش‌های گسترده‌ای در حوزه حمل‌ونقل برای مطالعه کنترل سیگنال ترافیکی انجام شده است. به‌طور سنتی، کنترل زمان‌ثابت یکی از قدیمی‌ترین و گسترده‌ترین روش‌های TSC است [۳۳]. این روش‌ها بر زمان‌بندی‌های از پیش تعیین‌شده مبتنی هستند که براساس الگوهای تاریخی ترافیک یا دستورالعمل‌های مهندسی تعیین می‌شوند. تکنیک‌های بهینه‌سازی مختلفی برای تعیین برنامه‌های زمان‌ثابت بهینه برای پیکربندی‌های تقاطع خاص پیشنهاد شده‌اند که روش Webster یکی از موفق‌ترین روش‌ها برای مورد یک‌تقاطع است؛ این روش طول چرخه و تقسیم فاز را متناسب با حجم ترافیک در یک بازه معین (مثلاً ۱۵ یا ۳۰ دقیقه گذشته) محاسبه می‌کند. با این حال، چنین روش‌های زمان‌ثابتی اغلب به‌دلیل عدم انطباق با تغییرات پویا شرایط ترافیکی عملکرد زیرحد بهینه دارند. برای مقابله با این مشکل، روش‌های کنترلی عمل‌گرا مانند سیستم هم‌آهنگ‌شده و پویا سیدنی (SCATS) [۳۴]، کنترل فشار حداکثر [۳۵] و SOTL [۴] طراحی شده‌اند تا زمان‌بندی‌ها را براساس تقاضای ترافیکی بلادرنگ تنظیم کنند. با وجود مزایای متعدد، این روش‌های عمل‌گرا نیازمند تنظیمات تخصصی برای هر تقاطع هستند و در سناریوهای پیچیده عملکردشان کاهش می‌یابد.

در سال‌های اخیر، روش‌های مبتنی بر RL توجه زیادی را به‌خود جلب کرده‌اند به‌دلیل توانایی قابل‌توجه آن‌ها در سازگاری با شرایط ترافیک بلادرنگ و قابلیت یادگیری سیاست‌های کنترلی بهینه در سناریوهای پیچیده [۳۶]. به‌صورت کلی، این روش‌ها را می‌توان به سه دسته تقسیم کرد: روش‌های مبتنی بر ارزش [۸و۹و۱۰و۱۱و۱۲و۱۳و۲۳و۲۵]روش‌های مبتنی بر سیاست [۱۵و۱۶و۱۷و۱۸و۱۹] و روش‌های بازیگر-منتقد [۲۰و۲۱و۲۲و۲۸]باوجود عملکرد خوب، اکثر روش‌های RL موجود بر آموزش مدل‌هایی برای پیکربندی‌های تقاطع مشخص متمرکزند. تلاش‌هایی نیز برای ساخت مدل‌های تعمیم‌یافته TSC انجام شده است؛ برای نمونه

[۱۱]مدلی نسبتاً عمومی‌تر با بهره‌گیری از استراتژی متا-یادگیری [۳۷] آموزش می‌دهد. بااین‌حال، مدل نام برده۸ نیازمند بازآموزی پارامترهای مدل برای هر تقاطع جدید است. برای غلبه بر این نقطه‌ضعف، [۱۸, ۱۰, ۱۲] مدل‌های عمومی از طریق اشتراک‌گذاری پارامترها ایجاد کردند؛ اما این روش‌ها ساختار فاز اصلی چراغ راهنما را حفظ نمی‌کنند. در مقابل، روش پیشنهادی ما می‌تواند ساختار چراغ سیگنال موجود را حفظ کند و هم‌زمان با فاین‌تیون تقاطع‌های حیاتی عملکرد را به‌طور چشمگیری بهبود دهد



شکل ۱) نقشه یک چهارراه استاندارد

۳- مفاهیم اولیه

در این بخش، برخی از مفاهیمی که در این کار مورد استفاده قرار گرفته‌اند، با استفاده از یک تقاطع استاندارد چهارراه مطابق شکل ۱ تعریف می‌شوند. این مفاهیم به‌آسانی قابل تعمیم به تقاطع‌هایی با ساختارهای متفاوت هستند.

الف)خط عبور : خط عبور به بخشی از راه گفته می‌شود که مسیر تعریف‌شده‌ای را برای حرکت وسایل نقلیه در یک جهت خاص فراهم می‌کند. در یک تقاطع معمولی دو نوع خط وجود دارد:

- خطوط ورودی l_{in}که وسایل نقلیه از آن وارد می‌شوند
- خطوط خروجی l_{out}که وسایل نقلیه از آن خارج می‌شوند.

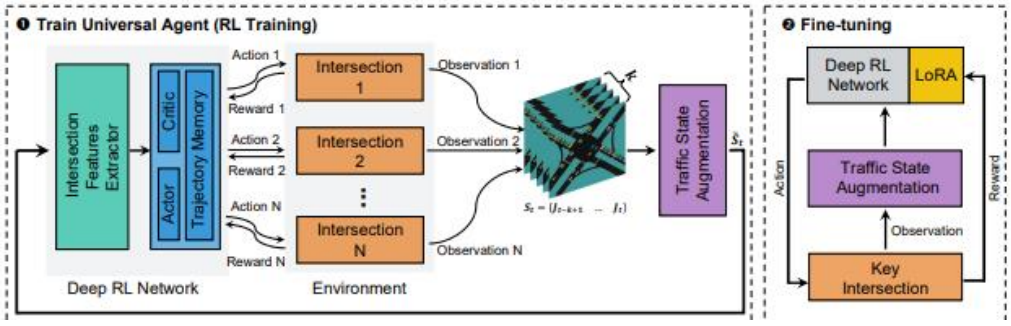
ب)حرکت ترافیکی : حرکت ترافیکی به اتصـــال میان یک خط ورودی l_{in}و یک خط خروجی l_{out}اطلاق می‌شـــود. در یک تقاطع چهارراه متداول در ســـمت چپ شـــکل ۱، مجموعاً ۱۲ حرکت وجود دارد که شـــامل گردش‌های راست، گردش‌های چپ و حرکات مستقیم در چهار جهت می‌باشند.

ج)سیگنال حرکت: سیگنال حرکت بر روی حرکت ترافیکی تعریف می‌شود. چراغ سبز یعنی حرکت مربوطه مجاز است و چراغ قرمز یعنی حرکت ممنوع است. از آنجا که در بسیاری از کشورها گردش راست مستقل از چراغ مجاز اســـت، تنها ۸ ســـیگنال حرکت از ۱۲ حرکت ممکن در یک چهارراه مورد اســـتفاده قرار می‌گیرند. ایـــن ۸ حـــرکـــت کـــه بـــا $M_1, M_2, \dots, M_8$نـــمـــایـــش داده شـــــده‌انـــد عـــبـــارت‌انـــد از: حرکت شمال‌سو(N) ، شمال‌سو با گردش چپ(NL) ، شرق‌سو(E) ، شرق‌سو با گردش چپ(EL) ، غرب‌سو(W) ، غـرب‌ســـو بـا گـردش چـپ(WL) ، جـنـوب‌ســـو(S) ، و جـنـوب‌ســـو بـا گـردش چـپ.(SL) برای مثال، m_8نشان می‌دهد که خودروها می‌توانند از سمت غرب به سمت شمال حرکت کنند.

د) فاز : فاز ترکیبی از چند ســیگنال حرکت اســت. هر فاز مجموعه خاصــی از حرکات ترافیکی را مجاز کرده و سایر حرکات را محدود می‌کند. قسمت بالا-راست شکل ۱ چهار فاز مربوط به یک چهارراه را نشان می‌دهد. مثلاً، فاز ۱ شامل حرکت‌های m_1 و m_5است و اجازه می‌دهد خودروهای شمال–جنوب حرکت کنند، در حالی که سایر حرکات ممنوع می‌باشند.

ه)طرح سیگنال: طرح سیگنال دنباله‌ای از فازها و مدت زمان هر فاز است که برای کنترل چراغ‌های یک تقاطع به‌کار می‌رود.به‌صورت ریاضی به صورت مجموعه‌ای از $\{(p_1, t_1), (p_2, t_2), \dots\}$نمایش داده می‌شود، که p_iفاز و t_iمدت زمان آن است.معمولاً توالی فازها چرخه‌ای است. قسمت پایین-راست شکل ۱ یک طرح چرخه‌ای سیگنال را نشان می‌دهد که در آن مدت زمان هر فاز برابر است با:

t_1
$= 50,$
$= 20,$
$= 20.$



شکل۲) نمای کلی ساختارUniTSA

۴- روش‌شناسی

۴-۱چارچوب

همان‌گونه که در شکل ۲ نشان داده شده است، چارچوب پیشنهادی UniTSA از دو ماژول تشکیل شده است:

الف) ماژول آموزش عامل عمومی

این بخش یک عامل RL عمومی را با اســتفاده از تقاطع‌های متفاوت آموزش می‌دهد. این ماژول از طراحی جدید عامل بهره می‌گیرد که قادر اسـت تقاطع‌هایی با توپولوژی‌ها و طرح‌های سـیگنالی متفاوت را با یک سـاختار ثابت مدل کند.این امر با اســـتفاده از ویژگی‌های حرکت‌ها و اعمال از نوع «فاز بعدی یا نه» و همچنین پنج روش نوآورلنه افزایش حللت ترافیک محقق می‌شـــود. ماژول فاین‌تیون برای تقاطع‌های کلیدی در این مرحله، مدل حاصـــل از ماژول اول برای برخی تقاطع‌های خاص که اهمیت بیشـــتری دارند، به‌صـــورت تطبیقی اصـــلاح می‌شود.جزئیات هر مرحله در بخش‌های بعدی تشریح خواهد شد.

۴-۲ طراحی عامل و ماتریس تقاطع

وضـعیت : تقاطع‌ها ممکن اسـت تعداد خطوط متفاوتی داشـته باشـند که در صـورت ثبت ویژگی‌ها در سـطح خط، فضـای حالت ابعاد متفاوتی خواهد داشـت. همان‌طور که در بخش ۳ اشـاره شـد، مشـاهده می‌شـود که در یک چهارراه تنها هشت سیگنال حرکت معتبر وجود دارد، بدون توجه به تعداد خطوط.

با الهام از این مشـاهده، ما پیشـنهاد می‌کنیم که وضـعیت تقاطع در زمان tبا یک ماتریس تقاطع)به شـکل زیر نمایش داده شود:

$$J_t = \begin{bmatrix} m_1^t \\ m_2^t \\ \vdots \\ m_8^t \end{bmatrix} \qquad (1)$$

که در آن $[\ \cdot\ ]^T$ترانهاده را نشان می‌دهد و بردار m_i^tاز طول ۸، اطلاعات استخراج‌شده از حرکت شماره iدر زمان tرا نشان می‌دهد.

اجزای بردار m_i^t:

بردار هر حرکت شامل سه دسته اطلاعات است:

الف) ویژگی‌های ترافیکی

این ویژگی‌ها سطح تراکم حرکت را با استفاده از پارامترهای زیر توصیف می‌کنند:

جریان متوسط ترافیک $F_{i,t}$

بیشینه اشغال $O_{i,t}^{\max}$ (Occupancy)

میانگین اشغال $O_{i,t}^{\text{mean}}$

ب) ویژگی‌های حرکت

این بخش اطلاعات ذاتی حرکت را ارائه می‌کند:

جهت حرکت I_i^sمثلاً مستقیم بودن یا نبودن

تعداد خطوط اختصاص‌یافته به حرکت L_i

۳- ویژگی‌های سیگنال ترافیکی

این بخش شامل سه پارامتر باینری است:

- $I_{i,t}^{cg}$:آیا حرکت فعلاً چراغ سبز دارد یا نه
- $I_{i,t}^{ng}$:آیا حرکت در فاز بعد چراغ سبز خواهد داشت یا نه
- $I_{i,t}^{mg}$:آیا مدت زمان حداقل چراغ سبز فعلی به پایان رسیده یا نه

این ۸ ویژگی به‌سادگی قابل استخراج هستند، که مدل را برای استفاده عملی قابل اجرا می‌سازد.

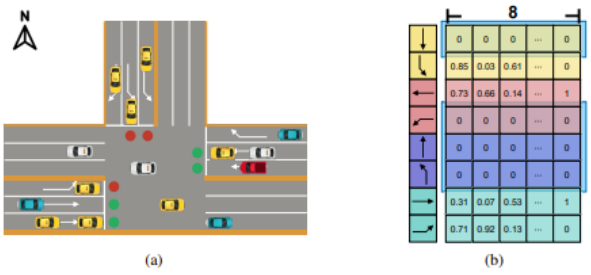
بردار m_i^tبه‌صورت زیر تعریف می‌شود:

$$m_t^i = [Fi, t, Oi, tmax, Oi, tmean, Iis, Li, Ii, tcg, Ii, tng, Ii, tmg]T \qquad (2)$$

۳- افزودن اطلاعات زمانی

مشاهده یک ماتریس J_tدر یک زمان واحد برای درک کامل دینامیک ترافیک کافی نیست. برای رفع این مشکل، ما آخرین K مشاهده را وارد وضعیت می‌کنیم تا عامل بتواند الگوها و روندهای ترافیکی را بهتر تشخیص دهد. در نتیجه، وضعیت عامل در زمان tبه‌صورت زیر تعریف می‌شود:

$$S_t \in \mathbb{R}^{K\times 8\times 8} \qquad (3)$$



شکل ۳) یک تقاطع سه راهی به همراه ماتریس متقاطع آن

در نهایت، شایان ذکر است که هنگامی‌که تعداد حرکت‌ها در یک تقاطع کمتر از ۸ باشد (برای مثال، یک تقاطع سه‌راهی)، از صفرپُرسی (Zero Padding) استفاده می‌شود. برای نمونه، شکل ۳ aیک تقاطع سه‌راهی رایج را نشان می‌دهد که تنها حرکت‌های شرق(E) ، چرخش چپ شرق(EL) ، غرب (W) و چرخش چپ جنوب (SL) فعال هستند. همان‌گونه که در شکل ۳ bدیده می‌شود، سطرهای مربوط به آن چهار حرکتِ بلااستفاده با صفر پر می‌شوند تا اندازهٔ ماتریس، مشابه اندازهٔ ماتریس در تقاطع چهارراهی باقی بماند.

عمل :

یک طراحی اقدام (Action) واقع‌بینانه و قابل‌اجرا باید ایمنی تمامی کاربران ترافیک را در نظر بگیرد. اگرچه طراحی «انتخاب فاز بعدی (choose next phase) «[۱۰و۱۲و۱۸]می‌تواند بازدهی تقاطع را به‌طور چشمگیری افزایش دهد، اما این روش توالی اصـــلی چراغ‌های راهنمایی را برهم می‌زند و در نتیجه، ایمنی رانندگان را تحت‌تأثیر قرار می‌دهد.

در مقابل، این پژوهش از طراحی اقدام »حفظ یا تغییر [۲۳و۲۵و۹] استفاده می‌کند. این طراحی به مفهوم چرخه‌ها وفادار می‌ماند و فازها را به‌صورت ترتیبی اجرا می‌کند (برای مثال: فاز ۱، فاز ۲، فاز ۳، فاز ۴، سپس دوباره فاز ۱، فاز ۲، و به همین ترتیب). عامل (agent) بر اساس وضعیت S_tتصمیم می‌گیرد که:

- فاز فعلی را حفظ کند، یا
- به فاز بعدی منتقل شود.

به دلیل وجود اطلاعات فاز فعلی I^{cg}و اطلاعات فاز بعدی I^{ng}در ماتریس تقاطع، در ادامه نشان داده می‌شود که این طراحی اقدام، مقیاس‌پذیری بسیار خوبی برای تقاطع‌هایی با طرح‌های سیگنال متفاوت ارائه می‌دهد.

پاداش:

منفیِ طول صف متوسط در هر حرکت q_iبه‌عنوان پاداش انتخاب شده است. معیارهایی مانند زمان انتظار، زمان سفر و تأخیر استفاده نشده‌اند، زیرا به‌دست‌آوردن آن‌ها از تجهیزات واقعی تشخیص ترافیک عملی نیست. در نتیجه، تابع پاداش پیشنهادی به‌صورت زیر تعریف می‌شود:

$$r_t = -\left(\sum_{i=1}^{8} q_i\right) - \frac{\mu}{\sigma + \varepsilon} \quad (4)$$

که در آن εعددی کوچک است تا از تقسیم بر صفر جلوگیری شود. علاوه بر این، μو σبه‌ترتیب میانگین و انحراف معیار اولین R مقدار پاداش هستند. این مقادیر به‌صورت زیر محاسبه می‌شوند:

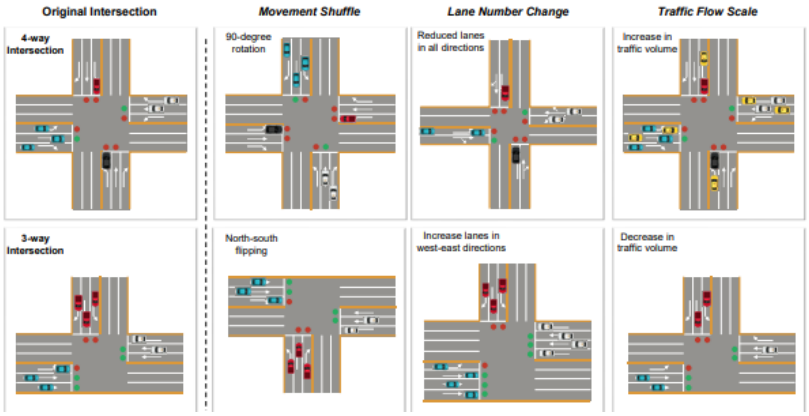
$$\mu = \frac{1}{R-1}\sum_{j=1}^{R-1} r_j \quad (5)$$

$$\sigma = \sqrt{\frac{1}{R-1}\sum_{j=1}^{R-1} (r_j - \mu)^2} \quad (6)$$

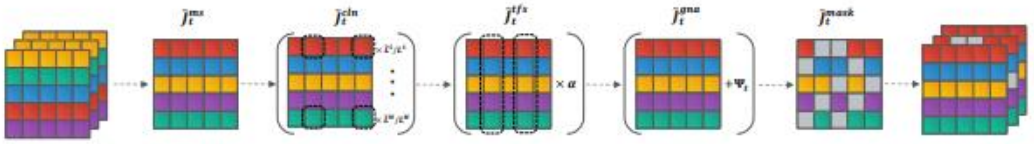
پاداش نرمال‌سازی می‌شود تا روند آموزش سریع‌تر و کارآمدتر گردد.

افزایش حالت ترافیکی:

در پژوهش‌های اخیر، تکنیک‌های افزایش داده اثربخشی خود را در بهبود توانایی تعمیم مدل‌های یادگیری تقویتی نشان داده‌اند [۳۸، ۳۹، ۴۰]. با آموزش عامل‌ها بر مجموعه‌ای متنوع‌تر از داده‌های افزایش‌یافته، مدل‌های یادگیری تقویتی قادر خواهند بود عملکرد بهتری در شرایط گوناگون و دیده‌نشده ارائه دهند.



شکل ۴) تصویرسازی اعمال سه روش اَفزایِش وضعیت ترافیک که بر تقاطع‌های چهارراهی و سه‌راهی



شکل ۵) مراحل دقیق در بلوک افزایش وضعیت ترافیک

قابلیت آن‌ها در مواجهه با وظایف دیده‌نشـــده را افزایش می‌دهد. رایج‌ترین روش‌های افزایش داده که در ادبیات گزارش شده‌اند، شامل افزودن نویز گاوسی و ماسک‌گذاری هستند. در این پژوهش، سه روش نوآورانه دیگر برای افزایش وضــعیت ترافیک، به‌طور خاص برای مســائل کنترل ســیگنال ترافیکی مبتنی بر یادگیری تقویتی (RL-based TSC) ارائه می‌کنیم؛ یعنی جابجایی حرکت‌ها، تغییر تعداد و مقیاس‌بندی جریان ترافیک که در شــکل ۴ نشان داده شده‌اند. شکل ۵ فرایند مرحله‌به‌مرحله اعمال‌شدن پنج روش افزایش داده را بر روی وضعیت ترافیکی S_t نشـان می‌دهد که در نهایت به وضـعیت افزوده‌شـده نهایی $\tilde{S}_t$ منتهی می‌شـود. در طول آموزش، یک مینی‌بچ از داده‌ها به‌طور تصـادفی از حافظهٔ بازپخش یا از مسـیرهای افزوده‌شـده اخیر نمونه‌برداری می‌شـود. اگرچه افزایش داده به‌صــورت تصــادفی روی مینی‌بچ اعمال می‌شــود، اما در میان فریم‌های پشــته‌شــده یکســان باقی می‌ماند. همچنین شــایان ذکر اســت که این پنج روش افزایش وضــعیت ترافیکی می‌توانند مســتقیماً بر ماتریس تقاطع J اعمال شـوند، که باعث می‌شـود عامل بتواند سـناریوها و سـاختارهای مختلف تقاطع را بهتر یاد بگیرد و تطبیق یابد.

جابجایی حرکت ها

در این روش، سطرهای ماتریس تقاطع جابجا می‌شوند تا چرخش‌ها یا تقارن‌های مختلف یک تقاطع شبیه‌سازی شوند. به‌صورت شهودی، جابجایی حرکت‌ها را می‌توان نوعی «چرخش مؤثر» تقاطع در نظر گرفت (طبق شکل ۴). فرض اصلی پشت این روش این است که اقدامی که عامل اتخاذ می‌کند نباید پس از **چرخش تقاطع تغییر کند**.

به‌صورت ریاضی، این عملیات به صورت زیر مدل‌سازی می‌شود:

$$\tilde{J}_t^{ms} = P \cdot J_t, \tag{7}$$

که در آن، J_tماتریس اصلی و $\tilde{J}_t^{ms}$ماتریس افزوده‌شده هستند. ماتریس Pیک ماتریس جایگشتی است که سطرهای J_tرا جابجا می‌کند.

با اعمال «جابجایی حرکت‌ها» بر تمام ماتریس‌های موجود در S_t، حالت جدید به صورت زیر تعریف می‌شود:

$$\tilde{S}_t^{ms} = [\tilde{J}_{t-K+1}^{ms}, \tilde{J}_{t-K+2}^{ms}, \ \dots, \tilde{J}_t^{ms}]. \quad (8)$$

تغییر تعداد خطوط

برای قرار دادن عامل در معرض ترکیب‌های متنوع‌تری از ساختارهای جاده، تعداد خطوط L_iدر هر بردار حرکت $\tilde{m}_i^{ms}$به‌طور تصادفی تغییر داده می‌شود. این روش باعث می‌شود عامل در طول آموزش، با پیکربندی‌های گوناگون خطوط آشنا شود و توانایی تطبیق با تقاطع‌های مختلف را افزایش دهد. همچنین، ویژگی‌های ترافیکی مانند جریان و اشغال، متناسب با ضریب برجای‌ماندن نسبت‌ها مقیاس می‌شوند.

مثالی از این روش در شکل ۴ آمده است: در یک تقاطع ۴راهی، تعداد خطوط هر جهت به ۲ کاهش یافته است؛ و در یک تقاطع ۳راهی تعداد خطوط ورودی شرق و غرب به ۴ افزایش یافته است. در این فرایند، نسبت تعداد وسایل نقلیه در هر جهت حفظ می‌شود؛ بنابراین انتظار می‌رود تصمیم عامل پیش و پس از این تغییر یکسان بماند.

عملیات به صورت زیر تعریف می‌شود:

$$\tilde{m}_i^{cln} = f(\tilde{m}_i^{ms}, \tilde{L}_i), i = 1,2, \dots ,8 \quad (9)$$

که در آن:

- L_iو $\tilde{L}_i$متغیرهای تصادفی یکنواخت هستند که تعداد خطوط اصلی و اصلاح‌شده را نشان می‌دهند،
- تابع fوظیفه تنظیم ویژگی‌های ترافیکی را دارد.

تابع fبه‌صورت زیر تعریف می‌شود:

$$f(\tilde{m}_i^{ms}, \tilde{L}_i) = \begin{cases} [\tilde{m}_i^{ms}]_k \cdot \dfrac{\tilde{L}_i}{L_i}, & k = 1,2,3,5 \\ [\tilde{m}_i^{ms}]_k, & \text{otherwise} \end{cases} \quad (10)$$

- که در آن $[\cdot]_k$ورودیkام بردار را نشان می‌دهد.
- پس از اعمال این روش، حالت جدید برابر است با:

$$\tilde{S}_t^{cln} = [\tilde{J}_{t-K+1}^{cln}, \ \dots, \tilde{J}_t^{cln}] \tag{11}$$

که

$$\tilde{J}_t^{cln} = [\tilde{m}_1^{cln}, \tilde{m}_2^{cln}, \ \dots, \tilde{m}_8^{cln}]^T. \tag{12}$$

مقیاس‌بندی جریان ترافیک

- برای تغییر تمرکز عامل از مقادیر مطلق ترافیک به توزیع نسبی وسایل نقلیه، یک ضریب مقیاس‌بندی جریان معرفی می‌شود. در این روش، مقدار جریان و اشغال هر حرکت در ماتریس تقاطع در یک عدد تصادفی یکنواخت αضرب می‌شود.
- عدد αبرای تمام حرکت‌های یک حالت ثابت باقی می‌ماند تا نسبت‌های بین حرکت‌ها حفظ شوند.
- این روش به عامل کمک می‌کند تا اهمیت نسبی حرکت‌ها را یاد بگیرد و کمتر به مقدارهای مطلق وابسته باشد. لازم به ذکر است که این روش تعداد خطوط را تغییر نمی‌دهد.
- عملیات به‌صورت زیر تعریف می‌شود:

$$\tilde{m}_i^{tfs} = g(\tilde{m}_i^{cln}, \alpha) \tag{13}$$

که تابع gبه‌صورت زیر تعریف می‌شود:

$$g(\tilde{m}_i^{cln}, \alpha) = \begin{cases} [\tilde{m}_i^{cln}]_k \cdot \alpha, & k = 1,2,3 \\ [\tilde{m}_i^{cln}]_k, & \text{otherwise} \end{cases} \tag{14}$$

حالت نهایی این مرحله $\tilde{S}_t^{tfs}$نامیده می‌شود.

افزودن نویز گاوسی:

برای افزودن تصادفی‌بودن به داده‌های آموزشی، نویز گاوسی مستقیماً به ماتریس تقاطع افزوده می‌شود. این نویز می‌تواند روی تمام اجزای ماتریس از جمله ویژگی‌های ترافیکی، ویژگی‌های حرکت و ویژگی‌های سیگنال تأثیر بگذارد. این روش باعث می‌شود عامل در برابر شرایط ناواضح و نامطمئن مقاوم‌تر شود.

عملیات به صورت زیر است:

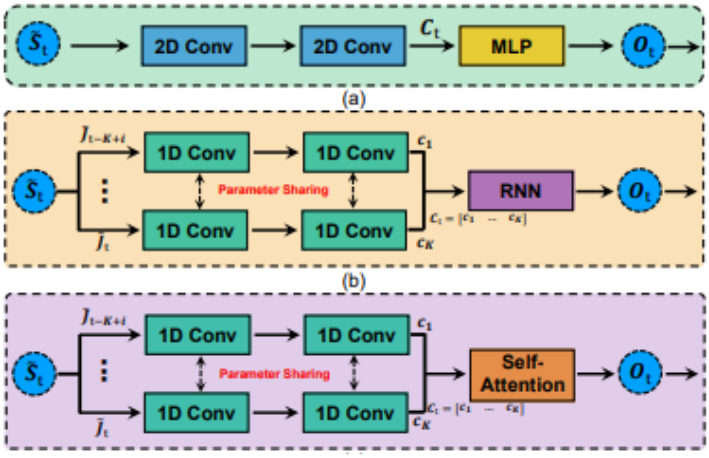
$$\tilde{J}_t^{gna} = \tilde{J}_t^{tfs} + \Psi_t, \qquad (15)$$

که در آن $\Psi_t \sim N(0, I)$ نویز گاوسی است.

پس از این مرحله، حالت جدید $\tilde{S}_t^{gna}$ نامیده می‌شود.

نقاب گذاری :

برای تشویق عامل به یادگیری تغییرات جریان ترافیک، به‌طور تصادفی برخی از مقدارهای موجود در ماتریس تقاطع در یک لحظه زمانی خاص صفر می‌شوند .ماسک‌گذاری عامل را مجبور می‌کند که برای استنتاج دینامیک‌های ترافیک، به اطلاعات قبل و بعد از بازه ماسک‌شده تکیه کند. این روش توانایی عامل را در درک نوسانات ترافیکی افزایش می‌دهد.



شکل ۶) سه نوع بلوک استخراج ویژگی تقاطع (الف) ساختار مبتنی بر CNN؛ (ب) ساختار مبتنی بر RNN؛ (ج) ساختار مبتنی بر ترانسفورماتور

۴-۴استخراج ویژگی‌های تقاطع

در این بخش سه ساختار شبکه عصبی برای استخراج اطلاعات تقاطع از حالات ترافیکی افزوده‌شده به‌کار گرفته می‌شوند، که عبارت‌اند از: شبکه کانولوشنی چندبعدی(Convolutional Neural Network — CNN) ، شبکه بازگشتی (Recurrent Neural Network — RNN) و ترنسفورمرها.(Transformers)

ساختار مبتنی برCNN

همان‌طور که در شکل ۶ (a)نشان داده شده است، از یک ساختار مبتنی بر CNN مجهز به دو لایه کانولوشن دوبعدی استفاده می‌شود تا اطلاعات سری‌زمانی موجود درون یک گره جاده استخراج شود. نمایش پنهان حاصل را C_tمی‌نامیم که به‌صورت زیر تعریف می‌شود:

$$C_t = \mathrm{ReLU}(W_2^{2d}\, W_1^{2d}\, \tilde{S}_t), \qquad (16)$$

که در آن W_1^{2d}و W_2^{2d}پارامترهای قابل‌آموزشِ دو بلوکِ کانولوشن دوبعدی هستند. به‌طور مشخص، بلوک اول W_1^{2d}اطلاعات مربوط به حرکت‌های ترافیکی را استخراج می‌کند و سپس بلوک دوم W_2^{2d}بر اساس آن، اطلاعات خاصِ تقاطع را می‌گیرد. همچنین $\mathrm{ReLU}(\cdot)$نشان‌دهنده تابع فعال‌سازی ReLU است.

سپس C_tاز طریق لایه(های) چندلایه‌ای پرسپترون (MLP) عبور داده شده و بردار ویژگی O_tتولید می‌شود:

$$O_t = W_u C_t + b_u, \qquad (17)$$

که W_uو b_uپارامترهای قابل‌آموزشِ لایه MLP هستند.

ساختار مبتنی برRNN

بر خلاف روش مبتنی برCNN ، ساختار مبتنی بر RNN که در شکل ۶ (b)نشان داده شده از یک لایهٔ کانولوشن ۱ Dبا به‌اشتراک‌گذاری پارامترها برای استخراج اطلاعات از هر ماتریس تقاطع استفاده می‌کند. لایهٔ کانولوشن یک‌بعدی بر هر ماتریس تقاطع $\tilde{J}_{t-K+i}$برای $i = 1,2,\dots,K$اعمال می‌شود که این ماتریس همان حالت

افزوده‌شده در یک گام زمانی خاص در پنجرهٔ تاریخی است. خروجی‌های حاصل که با $c_1, c_2, \dots, c_K$ نشان داده می‌شوند، اطلاعات تقاطع را در K زمان مختلف ضبط می‌کنند و به‌صورت زیر تعریف می‌گردند:

$$c_i = \text{ReLU}(W_2^{1d} W_1^{1d} \tilde{J}_{t-K+i}), i = 1,2, \dots, K, \qquad (18)$$

که در آن W_1^{1d} و W_2^{1d} پارامترهای قابل‌آموزش دو لایهٔ کانولوشن یک‌بعدی هستند. این خروجی‌ها سپس به ماژول RNN خورانده می‌شوند:

$$h_i = \tanh(W_x c_i + h_{i-1} W_h + b_h), i = 1,2, \dots, K, \qquad (19)$$

که W_x، W_h و b_h وزن‌ها و بایاسِ لایه مخفی در ساختار RNN هستند. حالت مخفی نهایی h_K که از خروجی آخرین گام RNN به‌دست می‌آید برای محاسبه ویژگی کل تقاطع در بازهٔ زمانی مورد نظر به‌کار می‌رود:

$$O_t = W_v h_K + b_v, \qquad (20)$$

که W_v و b_v وزن‌ها و بایاسِ لایهٔ خروجی در ساختار RNN می‌باشند.

ساختار مبتنی بر ترنسفورمر (Transformer)

شکل ۶ (c) روش مبتنی بر ترنسفورمر را نمایش می‌دهد. در اینجا، ابتدا از یک شبکه CNN با به‌اشتراک‌گذاری وزن‌ها برای استخراج ویژگی از ماتریس تقاطع در هر گام زمانی استفاده می‌شود (مشابه معادلهٔ (18)). اما به‌جای بلوک RNN، از یک رمزگذار ترنسفورمر برای آشکارسازی وابستگی‌های زمانی بین توالیِ ویژگی‌ها استفاده می‌شود. توالی ویژگی‌ها را $\tilde{C}_t = [c_1, c_2, \dots, c_K]$ تعریف می‌کنیم. برای وارد کردن اطلاعات زمان‌بندی، یک تعبیهٔ قابل‌آموزش با نماد c_{class} به $\tilde{C}_t$ اضافه می‌کنیم به‌طوری‌که:

$$C_t = \langle c_{\text{class}}, \tilde{C}_t \rangle = [\, c_{\text{class}}, c_1, c_2, \dots, c_K \,]. \qquad (21)$$

خروجی رمزگذار ترنسفورمر بر پایهٔ این توالی ورودیِ اصلاح‌شده، به‌عنوان نمایش حالت ترافیکی O_t به‌کار می‌رود. در بلوک رمزگذار ترنسفورمر، مکانیزم Self-attention به‌صورت زیر محاسبه می‌شود:

ابتدا پروجکشن‌های Query ، Key و Value را محاسبه می‌کنیم:

$$Q_C = C_t W_Q, K_C = C_t W_K, V_C = C_t W_V, \qquad (22)$$

و سپس خروجی attention را به‌صورت:

$$Z_t = \varphi(\frac{Q_C K_C^{\top}}{\sqrt{C_d}})V_C, \qquad (23)$$

که در آن W_Q, W_K, W_V پارامترهای پروجکشن مربوطه هستند، C_d بعدی است که برای نرمال‌سازی ضرب در دنباله attention استفاده می‌شود و $\varphi(\cdot)$ عملگر SoftMax است. خروجی نهایی O_t بردار نخست Z_t (معمولاً مرتبط با توکن c_{class}) می‌باشد.

۵-۴ آموزش RL و فاین‌تیون

برای آموزش مدل UniTSA از الگوریتم Proximal Policy Optimization (PPO) استفاده می‌شود. همان‌طور که در شکل ۲ نشان داده شده، عامل در تعامل با سناریوهای ترافیکی مختلف از تقاطع‌های با ساختارهای گوناگون، مسیرهایی متشکل از مشاهدات، اقدامات و پاداش‌ها را جمع‌آوری می‌کند. این مسیرها (trajectories) مبنای محاسبه تابع‌های هزینه سیاست (policy loss) و ارزش (value loss) هستند که برای به‌روزرسانی وزن‌های شبکه‌های Actor و Critic به‌کار می‌روند.

تابع هزینه سیاست (Policy Loss)

تابع هزینه سیاست اختلاف بین سیاست فعلی و سیاست به‌روزرسانی‌شده بر پایهٔ مسیرهای جمع‌آوری‌شده را اندازه‌گیری می‌کند. این تابع عامل را تشویق می‌کند تا احتمال اقداماتی را که منجر به پاداش‌های بالاتر می‌شوند افزایش دهد و احتمال اقدامات منفی را کاهش دهد. به‌صورت ریاضی تابع هزینهٔ سیاست چنین فرموله می‌شود:

$$L_{pf}(\theta) = \widehat{\mathbb{E}}_t[\min(r_t A_t,\ \text{clip}(r_t, 1-\epsilon, 1+\epsilon)\, A_t)], \qquad (24)$$

که $\widehat{\mathbb{E}}_t$ نشان‌دهنده اپراتور امید ریاضی تخمینی است و r_t نسبت میان سیاست جدید $\pi_\theta(a_t \mid s_t)$ و سیاست قدیمی $\pi_{\theta_{\text{old}}}(a_t \mid s_t)$ است:

$$r_t = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)}. \qquad (25)$$

علاوه بر این، $A_t = r_t + \gamma V(s_{t+1}) - V(s_t)$تابع مزیت است و استفاده از تابع $\text{clip}(\cdot)$موجب به‌روزرسانی‌های پایدارتر در سیاست می‌شود.

تابع هزینه ارزش(Value Loss)

تابع هزینه ارزش اختلاف بین مقدار تخمینی تابع ارزش و پاداش‌های واقعی به‌دست‌آمده را اندازه می‌گیرد تا تابع ارزش بهتر بتواند مجموع پاداش‌های آتی مورد انتظار را تقریب بزند. این تابع به‌صورت زیر تعریف می‌شود:

$$L_{vf}(\theta) = \hat{\mathbb{E}}_t[(V_\theta(s_t) - \hat{R}_t)^2], \qquad (26)$$

که در آن $V_\theta(s_t)$مقدار تخمینی تابع ارزش تحت سیاست θاست و $\hat{R}_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$نمایانگر «پاداش تا سرانجام (reward-to-go) «می‌باشد.

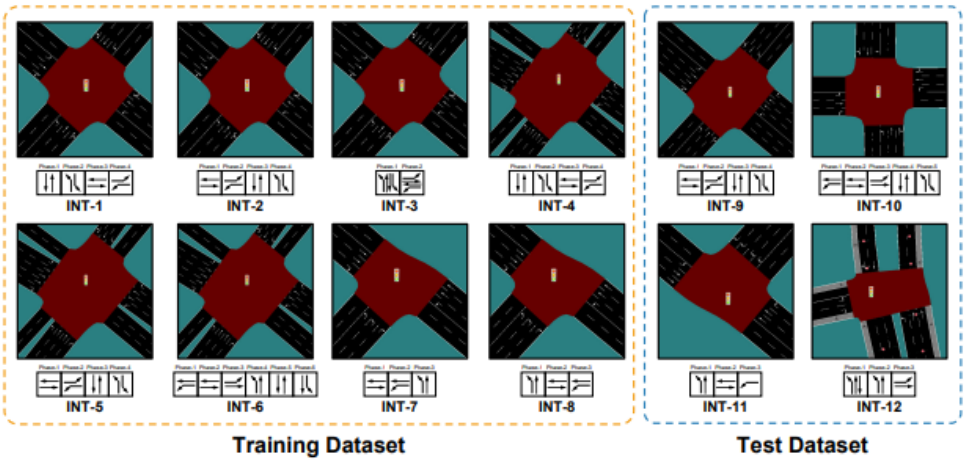
تابع هدف نهایی

پس از محاسبه تابع هزینه سیاست و تابع هزینه ارزش، تابع هدف نهایی به‌صورت زیر تعریف می‌شود:

$$L(\theta) = -L_{pf}(\theta) + \lambda\, L_{vf}(\theta), \qquad (27)$$

که در آن λضریبِ وزن‌دهی به خسارت ارزش است.

این پژوهش یک مدل جهانی مؤثر برای کنترل سیگنال‌های ترافیکی پیشنهاد می‌کند که با کمینه‌سازی تابع هدف فوق از طریق PPO بهینه می‌گردد. علاوه بر این، از آن‌جا که برخی تقاطع‌ها در عمل از اهمیت بیشتری برخوردارند، برای فاین‌تیون عملکرد روی این تقاطع‌های حیاتی از روش LoRAاستفاده می‌شود. به‌طور مشخص، ماژول‌های LoRA به وزن‌های لایه‌های چگال (dense) در شبکه Actor و Critic افزوده می‌شوند (همان‌طور که در شکل ۲ نشان داده شده است). در فرایند فاین‌تیون، وزن‌های پیش‌آموزش‌دیده اصلی ثابت نگه داشته می‌شوند و تنها ماژول‌های LoRA به‌روزرسانی می‌گردند. این طراحی اجازه می‌دهد مدل بدون افزایش قابل‌توجه در تعداد پارامترها به ویژگی‌های ویژه هر تقاطع سازگار شود و در عین حال کارآیی آموزش حفظ شود.



شکل ۷) تمام توپولوژی‌های تقاطع به همراه فازهای موجودشان که برای آموزش و آزمایش استفاده شده‌اند

۵) آزمایش‌ها

۵-۱ تنظیمات آزمایش

برای ارزیابی مدل پیشنهادی، آزمایش‌های گسترده‌ای با استفاده از بسته نرم‌افزاری SUMOانجام شد SUMO . یک شبیه‌ساز ترافیک میکروسکوپی متن‌باز است که برای مدیریت شبکه‌های بزرگ طراحی شده است. این نرم‌افزار رابط Traffic Control Interface (TraCI)را ارائه می‌دهد که امکان کنترل چراغ‌های ترافیکی و دریافت اطلاعات شرایط ترافیکی تقاطع‌ها را فراهم می‌کند. جریان ترافیک و اشغال (occupancy) هر حرکت با تحلیل موقعیت‌ها و مسیرهای خودروها محاسبه می‌شود. برای شبیه‌سازی شرایط واقعی، تنها خودروهایی که در فاصله ۱۵۰ متر از تقاطع قرار دارند، در نظر گرفته شده‌اند. همچنین پس از چراغ سبز، چراغ زرد به مدت ۳ ثانیه نمایش داده شد تا ایمنی راننده حفظ شود .زمان انتظار هر خودرو به‌عنوان معیار عملکرد جهت ارزیابی اثربخشی روش‌های مختلف استفاده شده است؛ مقدار کمتر این معیار نشان‌دهنده عبور سریع‌تر خودروها از تقاطع است.برای آموزش مدل، از پیاده‌سازی الگوریتم Proximal Policy Optimization (PPO)ارائه‌شده توسط کتابخانه Stable Baselines3استفاده شد. برای افزایش سرعت آموزش، از ۳۰ پردازش موازی بهره گرفته شد و تعداد کل گام‌های محیط آموزش برابر با ۱۰ میلیون تعیین گردید. نمایش حالت شامل ۸ نمونه قبلی ماتریس تقاطع بود و فاصله زمانی بین دو اقدام متوالی ۵ ثانیه در نظر گرفته شد.

جدول ۱) تنظیمات ابر پارامترها

Hyper-parameter	Value
Learning rate	0.0001
Trajectory memory size	3000
Clipping range ϵ	0.2
Discount factor γ	0.99
Value function coefficient λ	0.9
Scale hyperparameter α	1
Rank of LoRA module	8

ابر پارامترها نیز مطابق جدول ۱ تنظیم شدند. شبکه‌های Actor و Critic به صورت دو لایه کاملاً متصل طراحی شدند. ابعاد ورودی {۶۴, ۳۲} و ابعاد خروجی {۳۲, ۲} برای Actor و {۳۲, ۱} برای Critic تنظیم شد.

جدول ۲) تمام تنظیمات تقاطع ها

	Training Dataset								Test Dataset			
Intersection ID	INT-1	INT-2	INT-3	INT-4	INT-5	INT-6	INT-7	INT-8	INT-9	INT-10	INT-11	INT-12
roads	4	4	4	4	4	4	3	3	4	4	3	3
lanes per road	(3,3,3,3)	(3,3,3,3)	(3,3,3,3)	(3,4,4,5)	(3,4,4,5)	(3,4,4,5)	(3,3,3)	(3,3,3)	(3,4,3,4)	(3,3,3,3)	(4,3,3)	(2,3,2)
phases	4	4	2	4	4	6	3	3	4	5	3	3

۲-۵ مجموعه داده‌ها

این مطالعه تقاطع‌هایی با ساختارهای متنوع در نظر گرفته است تا یک مدل جهانی واحد برای پیش‌بینی اقدامات در تمامی تقاطع‌ها آموزش داده شود. به‌طور مشخص، ۱۲ تقاطع با تعداد فازها، خطوط عبور و جاده‌های ورودی متفاوت (۳-راه یا ۴-راه) ساخته شده و برای آزمایش استفاده شد. از این ۱۲ تقاطع، هشت تقاطع برای آموزش و چهار تقاطع برای آزمون کنار گذاشته شدند. توپولوژی و فازهای ۱۲ تقاطع در شکل ۷ نشان داده شده و جدول ۲ خلاصهٔ پیکربندی‌های آن‌ها را ارائه می‌دهد. برای مثال، INT-4شامل چهار جاده دوطرفه است با سه خط در جهت شمال-جنوب، پنج خط در جهت غرب-شرق و چهار خط در هر یک از دو جهت دیگر INT-4 .شامل چهار فاز است که هر فاز دو سیگنال حرکت مختلف را ترکیب می‌کند. بنابراین، پیکربندی INT-4 در جدول ۲ شامل چهار جاده با تعداد خطوط مشخص (۳, ۴, ۴, ۵) به ترتیب جهت عقربه‌های ساعت و چهار فاز می‌باشد.

در مجموعه داده آموزش، سه توپولوژی متفاوت وجود دارد:

- INT-1، INT-2و INT-3: تقاطع ۴-راه معمولی با هر جاده شامل سه خط.
- INT-4، INT-5و INT-6: تقاطع ۴-راه بزرگ با بیش از چهار خط در هر جاده.
- INT-7، INT-8و INT-9: تقاطع ۳-راه.

برای تقاطع‌های با توپولوژی مشابه، تقاطع‌های جدید با تغییر ترتیب یا تعداد فازها ساخته شد. به‌طور مثال، INT-1و INT-2 دارای پیکربندی یکسان هستند اما ترتیب فازها متفاوت است. همچنین INT-1 و INT-3 در تعداد فازها متفاوت هستند.

برای ارزیابی عملکرد مدل در تقاطع‌های دیده‌نشده، چهار سناریوی آزمون شامل INT-9 ، INT-10، INT-11و INT-12طراحی شد.

- INT-9و INT-10: تقاطع‌های ۴-راه با پیکربندی خطوط و فازهای متفاوت نسبت به مجموعه آموزش.
- INT-11:تغییر خطوط و فازها بر اساس INT-8.
- INT-12:شبیه‌سازی ترافیک واقعی شهر Ingolstadt، **آلمان** .[43]

علاوه بر تنوع ساختار، برای هر تقاطع ۱۰۰ مسیر منحصربه‌فرد تولید شد که سه‌چهارم آن‌ها برای آموزش و یک‌چهارم برای ارزیابی استفاده شد. هر مسیر دارای مدت ۳۰,۰۰۰ ثانیه (~۸ ساعت) بود.

5.3روش‌های مقایسه‌ای(Compared Methods)

برای ارزیابی عملکرد UniTSA، مدل نهایی با روش‌های کلاسیک و پیشرفته RL برای کنترل ترافیک مقایسه شد:

- **Fix Time [2]**کنترل زمان ثابت با چرخه و فازهای از پیش تعیین‌شده. دو نسخه FixTime-30 :و FixTime-40هر فاز ۳۰ و ۴۰ ثانیه
- **Webster [3]**تعیین طول چرخه و تقسیم فاز بر اساس حجم ترافیک در یک بازه زمانی. می‌تواند زمان سفر را کاهش یا ظرفیت تقاطع را افزایش دهد. برای عدالت، در این مطالعه بر اساس ترافیک لحظه‌ای اعمال شد.

- **SOTL [4]**: کنترل سیگنال خودسازمانده که طول سیگنال‌ها را بر اساس آستانه تعداد خودروهای منتظر تنظیم می‌کند.
- **MPLight [10]**: استفاده از ساختار FRAP و مکانیسم پاداش مبتنی بر فشار، همچنین MLP با اشتراک پارامتر برای افزایش سازگاری با توپولوژی‌های مختلف.
- **AttendLight [18]**: مدل توجه برای آموزش مدل جهانی که قادر به مدیریت تقاطع‌های با ساختار و جریان متفاوت است. شامل دو مدل توجه: اول برای تعداد جاده‌ها و خطوط و دوم برای تصمیم‌گیری بین تقاطع‌ها با تعداد فاز متفاوت.

روش‌های متغیر :**UniTSA**

- **UniTSA (Single)**: آموزش در یک محیط، استفاده از ساختار RNN برای استخراج اطلاعات حالت ترافیک.
- **UniTSA (Multi)**: آموزش همزمان در چند محیط با ساختارهای مختلف شبکه(CNN ، RNN، Transformer).
- **UniTSA (Multi+TSA)**: ترکیب پنج روش تقویت حالت ترافیک (Traffic State Augmentation)با UniTSA (Multi) برای افزایش تنوع داده‌ها و بهبود عملکرد.

جدول ۳) نتایج کمی (میانگین زمان انتظار به ازای هر وسیله نقلیه) تقاطع‌های آموزشی برای مدل‌های جهانی. مقدار کمتر نشان دهنده عملکرد بهتر است و کمترین مقادیر با حروف پررنگ مشخص شده‌اند.

	INT-1	INT-2	INT-3	INT-4	INT-5	INT-6	INT-7	INT-8
Fix-30 [2]	39.458	38.862	8.323	35.123	35.031	51.923	18.920	18.595
Fix-40 [2]	50.696	52.108	10.667	44.888	45.540	61.743	23.411	24.759
Webster [3]	26.466	26.889	5.751	25.413	24.541	40.128	12.755	13.574
SOTL [4]	16.048	16.561	**2.764**	28.169	27.902	27.404	8.639	7.925
MPLight [10]	19.111	14.659	4.469	16.067	19.925	19.115	6.654	7.523
AttendLight [18]	16.483	13.893	3.860	16.903	18.915	20.795	6.532	8.104
UniTSA (Multi+CNN)	13.776	13.580	3.265	14.790	15.437	18.751	6.592	6.894
UniTSA (Multi+RNN)	13.692	13.437	3.495	15.198	14.896	18.612	6.393	6.625
UniTSA (Multi+Trans)	14.976	20.459	2.811	16.489	16.481	23.793	7.552	7.175
UniTSA (Multi+CNN+TSA)	14.071	14.150	3.036	15.990	16.472	24.383	6.329	6.328
UniTSA (Multi+RNN+TSA)	13.450	13.578	3.007	**14.470**	**14.462**	18.689	**6.242**	**6.164**
UniTSA (Multi+Trans+TSA)	**13.335**	**13.314**	3.311	14.648	14.566	**18.579**	6.643	6.571

جدول ۴) نتایج کمی تقاطع‌های آزمون برای مدل‌های جهانی

	INT-9	INT-10	INT-11	INT-12
Fix-30	40.341	38.360	17.136	17.440
Fix-40	60.043	50.580	23.504	20.570
Webster	26.191	27.766	11.981	13.280
SOTL	23.031	28.066	7.452	8.070
MPLight	23.674	21.475	8.447	15.047
AttendLight	18.080	18.501	7.393	12.982
UniTSA (Multi+CNN)	17.877	18.244	7.063	12.820
UniTSA (Multi+RNN)	16.771	15.450	6.807	10.140
UniTSA (Multi+Trans)	17.640	16.269	6.598	9.630
UniTSA (Multi+CNN+TSA)	13.075	14.245	6.794	7.550
UniTSA (Multi+RNN+TSA)	**12.677**	**14.208**	**5.914**	6.730
UniTSA (Multi+Trans+TSA)	13.054	16.186	5.983	**6.190**

۵-۴نتایج تقاطع‌های آموزشی

عملکرد UniTSA با روش‌هایFix Time ، Webster، SOTL، MPLightو AttendLight بر روی تقاطع‌های آموزش مقایسه شد. نتایج متوسط زمان انتظار هر خودرو در جدول ۳ ارائه شده است.

- روش‌های مبتنی بر RL در اکثر تقاطع‌ها عملکرد بهتری نسبت به روش‌های کلاسیک نشان دادند.
- SOTLدر INT-2 کمترین زمان انتظار را داشت اما نیاز به آستانه‌های دستی داشت که تعمیم‌پذیری آن را محدود می‌کند.
- بین روش‌های RL جهانی، UniTSAنسبت به MPLight و AttendLight بهبود قابل‌توجهی داشت؛ به‌طور میانگین، بهبود ۱۵٪ و ۱۲٪ نسبت به MPLight و AttendLight در هشت تقاطع مشاهده شد.

توضیح عملکرد:

- UniTSAعلاوه بر اشتراک پارامتر، با جایگزینی بلوک RNN یا Transformer به جایMLP ، اطلاعات زمانی وضعیت ترافیک را استخراج می‌کند.
- پنج روش تقویت حالت ترافیک به عامل اجازه می‌دهد در طول آموزش با تنوع بیشتری از وضعیت‌های تقاطع مواجه شود.
- بررسی ساختار شبکه‌ها نشان می‌دهد که RNNنسبت به CNN عملکرد بهتری دارد.

- در غیاب تقویت حالت ترافیک، UniTSA (Multi+CNN)و UniTSA (Multi+RNN) عملکرد بهتری نسبت به UniTSA (Multi+Trans) داشتند، زیرا ترنسفورمر به داده‌های بیشتری نیاز دارد.
- با اعمال تقویت حالت ترافیک، مدل می‌تواند با تقاطع‌های متنوع‌تری تعامل داشته باشد و UniTSA (Multi+Trans+TSA)در بسیاری از سناریوها از UniTSA (Multi+RNN) و UniTSA (Multi+RNN+TSA)عملکرد بهتری نشان داد.

فرمول LoRA در UniTSA

اگر ماتریس وزن پیش‌آموزش‌دیده $W \in \mathbb{R}^{n\times m}$با ماژول LoRA $\Delta W = W_A W_B^T$ داشته باشیم، با $W_A \in \mathbb{R}^{n\times d}$، $W_B \in \mathbb{R}^{m\times d}$و $d \ll n$، خروجی لایه به شکل زیر است:

$$z = Wx + \Delta W x = Wx + \alpha r W_A W_B^T x, \qquad (28)$$

که W_Aو W_Bبه ترتیب با ماتریس صفر و ماتریس گوسی با میانگین صفر مقداردهی اولیه شده‌اند، αضریب مقیاس و rرتبه ماژول LoRA است. با ترکیب آموزش RL با PPO و فاین‌تیونLoRA ، مدل جهانی پیشنهادی قادر است چالش‌های تقاطع‌های مهم با ساختارهای متنوع را به‌طور مؤثر مدیریت کند.

۵-۵ نتایج تقاطع‌های آزمون

در مرحله بعد، ویژگی کلیدی UniTSA مورد ارزیابی قرار گرفت تا بررسی شود آیا می‌تولند برای تقاطع‌های دیده‌نشده (unseen intersections) نیز مورد استفاده قرار گیرد. چهار تقاطع شامل INT-9، INT-10، INT-11و INT-12به‌طور ویژه برای اهداف آزمایشی انتخاب شدند. همان‌طور که در بخش ۵.۲ بیان شد، این تقاطع‌ها از نظر تعداد خطوط و تعداد فازها با تقاطع‌های مجموعه آموزش متفاوت هستند.

در زمینه کاهش زمان سفر متوسط خودروها در عبور از تقاطع‌ها، مدل UniTSA (Multi+RNN+TSA)نشان داد که زمان سفر متوسط خودروها به ترتیب تقریباً ۳۲.۹٪ و ۴۱.۳٪ نسبت به MPLight و AttendLight کاهش یافته است. این نتایج نشان‌دهنده اثربخشی روش پیشنهادی در بهینه‌سازی کنترل چراغ‌های ترافیکی (TSC) و بهبود کارایی جریان ترافیک می‌باشد.

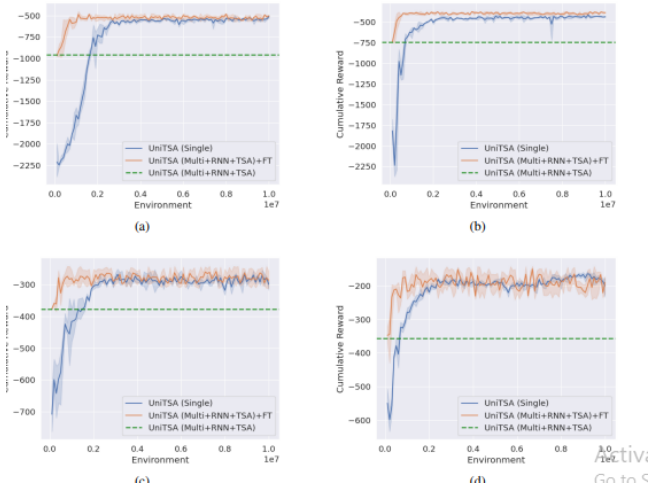
قابل توجه است که استفاده از تکنیک‌های تقویت حالت ترافیک موجب بهبود عملکرد بین واریانت‌های مختلف UniTSAمی‌شود. به‌طور مثال:

- UniTSA (Multi+RNN+TSA)
- UniTSA (Multi+RNN+TSA)
- UniTSA (Multi+Trans+TSA)

به ترتیب بهبودهای ۲۳.۴٪، ۱۹.۸٪ و ۱۷.۹٪ را در مقایسه با(UniTSA (Multi+RNN ، UniTSA (Multi+RNN) و UniTSA (Multi+RNN) نشـان دادند. این بهبود به دلیل افزوده شـدن تنوع بیشـتر در سـناریوهای تقاطع در داده‌های آموزشـی اسـت که از طریق تکنیک‌های تقویت حالت ترافیک، مانند روش "تغییر تعداد خطوط ایجاد شده و امکان تولید ترکیب‌های متنوعی از پیکربندی خطوط را فراهم می‌کند.

جدول ۴ عملکرد روش‌های مختلف، شـامل روش‌های پایه و واریلنت‌های مختلف مدل UniTSA را جمع‌بندی می‌کند. مشابه نتایج به‌دست‌آمده در مجموعه آموزش، الگوریتم‌های مبتنی بر RL عملکرد قابل توجهی نسبت به روش‌های سنتی کنترل ترافیک نشان دادند.

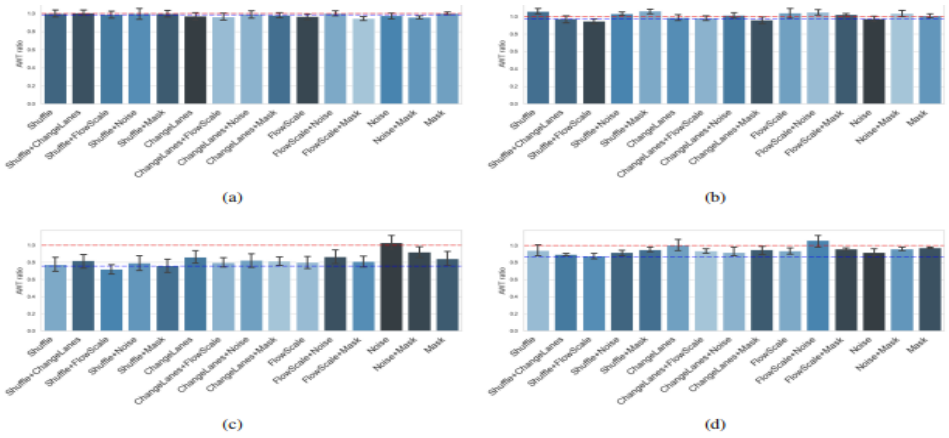
در بین روش‌های جهانی مبتنی برRL ، مدل‌های UniTSA در تمامی تقاطع‌های آزمون عملکرد برتری داشـتند . به ویژه، UniTSA (Multi+RNN+TSA)و UniTSA (Multi+Trans+TSA)در میان تمامی روش‌ها بهترین عملکرد را ارائه کردند.



شکل ۸) مراحل محیطی روش‌های مختلف در چهار تقاطع آزمون. (الف) INT-9. (ب) INT-10. (ج) INT-11. (د) INT-12.

جدول ۵) نتایج کمی تنظیم دقیق در تقاطع‌های آزمایشی

	INT-9	INT-10	INT-11	INT-12
UniTSA (Multi+RNN)	16.771	15.450	6.807	10.140
UniTSA (Multi+RNN+TSA)	12.677	14.208	5.914	6.730
1M Environment Steps				
UniTSA (Single)	23.683	14.863	8.417	5.024
UniTSA (Multi+RNN+TSA) + FT	**10.995**	**12.553**	**4.475**	**3.534**
10M Environment Steps				
UniTSA (Single)	10.353	12.254	4.359	**3.089**
UniTSA (Multi+RNN+TSA) + FT	**10.292**	**11.081**	**4.153**	3.110



شکل ۹) تحلیل مقایسه‌ای روش‌های تقویت وضعیت ترافیک در تقاطع‌های آموزشی و آزمایشی انتخاب‌شده (الف) INT-1. (ب) INT-7. (ج) INT-9. (د) INT-11

۵-۶ نتایج فاین‌تیونینگ در تقاطع‌های آزمون

در سناریوهای عملی، برخی تقاطع‌ها به دلیل اهمیتشان نیاز به توجه ویژه دارند. برای رسیدگی به این موضوع، ما با مدل جهانی آموزش‌دیده توسط UniTSA (Multi+RNN) شروع می‌کنیم. ساختار RNN برای UniTSA انتخاب شد زیرا عملکرد بهتری در اغلب تقاطع‌ها نسبت به ساختارهای CNN و Transformer دارد. در مقایسه با UniTSA (Single)، که روی یک سناریوی واحد آموزش داده شده است، مدل حاصل می‌تواند به سرعت به عملکرد مدل تک‌محیطی برسد یا حتی آن را پس از تنها چند مرحله آموزش پشت سر بگذارد.

شـــکل ۸ تغییرات مجموع پاداش‌ها را طی مراحل آموزش برای مدل‌های مختلف در تقاطع‌های آزمون نشـــان می‌دهد. خط سـبز نقطه‌چین نتایج اعمال مسـتقیم مدل جهانی بر تقاطع‌های جدید بدون هیچ فاین‌تیونینگی را نشـــان می‌دهد. قابل توجه اســـت که مدل حتی بدون هرگونه فاین‌تیونینگ یا یادگیری انتقالی نیز نتایج امیدوارکننده‌ای ارائه می‌دهد.

خط آبی نشـان‌دهنده مدل آموزش‌دیده از صـفر و خط نارنجی نشـان‌دهنده مدل فاین‌تیون‌شـده بر اسـاس مدل جهانی است. مشاهده می‌شود که مدل تک‌محیطی تقریباً در ۳ میلیون مرحله آموزش همگرا می‌شود، در حالی که مدل فاین‌تیون‌شده با تنها حدود ۱ میلیون مرحله آموزش به عملکرد مشابه دست می‌یابد و در نتیجه حدود ٪۶۶ کاهش در زمان محاسبات با حفظ عملکرد مشابه به‌دست می‌آید.

جدول ۵ تحلیل دقیقی از عملکرد مدل پس از فاین‌تیونینگ ارائه می‌دهد. در ۱ میلیون مرحله آموزش، مدل فاین‌تیون‌شده بهبود عملکرد متوسط ٪۳۶ نسبت به UniTSA (Single) در چهار تقاطع آزمون نشان داد. حتی پس از ۱۰ میلیون مرحله آموزش، مدل‌های فاین‌تیون‌شـــده همچنان ٪۳ بهتر از مدل‌های آموزش‌دیده از صـــفر عمل کردند. این ویژگی برای کاربردهای بلادرنگ بسیار جذاب است، زیرا در شبکه‌های جاده‌ای با بیش از ۱۰۰۰ تقاطع می‌توان تعداد تعاملات با محیط را به طور قابل توجهی کاهش داد و در عین حال عملکرد مشـــابهی را حفظ کرد، که به طور چشمگیری کارایی آموزش را افزایش می‌دهد.

۵-۷ تحلیل مقایسه‌ای روش‌های تقویت حالت ترافیک

در این بخش، اثربخشـــی روش‌های تقویت حالت ترافیک در بهبود عملکرد مدل UniTSA (Multi+RNN)تحلیل شده است. آزمایش‌ها با استفاده از ترکیب‌های مختلف دو تکنیک تقویت حالت ترافیک روی تقاطع‌های آموزش و آزمون انجام شـد. برای ارزیابی تأثیر این روش‌ها، نسـبت زمان انتظار متوسـط (AWT ratio) بین مدل‌های با و بدون تقویت حالت ترافیک محاسـبه شـد. مقدار کمتر از ۱ نشـان‌دهنده بهبود عملکرد ناشـی از روش‌های تقویت است.

شـکل ۹ نتایج تحلیل مقایسـه‌ای را برای تقاطع‌های INT-1، INT-7، INT-9و INT-11 ، که نماینده سـاختارهای معمول تقاطع در سـناریوهای واقعی هسـتند، نشـان می‌دهد. هر میله در نمودار مربوط به نسـبت متوسـط AWT برای ترکیب خاصـی از روش‌های تقویت اسـت و نوارهای خطا، فاصـله اطمینان ٪۹۵ را نشـان می‌دهند. خط آبی نقطه‌چین نســـبت AWT مدل UniTSA (Multi+RNN+TSA)را که از تمامی روش‌های تقویت حللت ترافیک استفاده می‌کند، نشان می‌دهد.

شکل ۹ (a)و ۹ (b)نتایج به‌دست‌آمده از تقاطع‌های آموزش INT-1 و INT-7 را نشان می‌دهد. بررسی این نتایج نشان می‌دهد که بهبود متوسط عملکرد از طریق تقویت حالت ترافیک حدود ۲٪ بوده است. همچنین، استفاده از روش‌های Noiseو Mask در INT-7 باعث کاهش عملکرد شد، که می‌تواند ناشی از این باشد که مدل قبلاً الگوها و ویژگی‌های پایه‌ای تقاطع‌های آموزش را تا حد زیادی یاد گرفته است و افزودن تغییرات اضافی از طریق تقویت حالت ترافیک ممکن است مزایای قابل توجهی نداشته باشد.

با این حال، روش‌های تقویت حالت ترافیک هنگام مواجهه با ساختارهای تقاطع دیده‌نشده عملکرد قابل توجهی ارائه دادند. شکل ۹ (c)و ۹ (d)نسبت‌های AWT برای تقاطع‌های آزمون INT-9و INT-11 را نشان می‌دهد. نتایج به وضوح نشان می‌دهد که بیشتر روش‌های تقویت حالت ترافیک عملکرد سیاست پایه را در تقاطع‌های آزمون بهبود دادند .نمونه‌های متنوع آموزشی ایجادشده از طریق تقویت حالت ترافیک باعث بهبود عملکرد مدل شدند. از میان تکنیک‌ها، Movement Shuffleو Traffic Flow Scale به‌طور ویژه‌ای مؤثر بودند، زیرا به مدل اجازه می‌دهند تا از طیف گسترده‌تری از سناریوها یاد بگیرد و خود را تطبیق دهد و در نتیجه عملکرد در مجموعه آزمون افزایش یابد.

۶ نتیجه‌گیری

در این مقاله، یک چارچوب کنترل چراغ ترافیک مبتنی بر RL به‌صورت جهانی (UniTSA) برای ساختارهای مختلف تقاطع در محیط‌های V2X پیشنهاد شد. به طور مشخص، UniTSAامکان آموزش یک عامل RL جهانی را با استفاده از ماتریس تقاطع برای توصیف وضعیت تقاطع فراهم می‌کند. برای مقابله با تقاطع‌های دیده‌نشده، روش‌های جدید تقویت حالت ترافیک پیشنهاد شد تا داده‌های جمع‌آوری‌شده توسط عامل غنی شود و عملکرد و قابلیت تعمیم مدل در تقاطع‌های ناشناخته بهبود یابد. نتیجه این است که UniTSA نیاز به سفارشی‌سازی و توسعه مجدد برای هر تقاطع منفرد را از بین می‌برد و در عین حال پیاده‌سازی ساده، کارآمد و متن‌باز ارائه می‌دهد، که آن را به چارچوبی ارزشمند برای تحقیقات آینده در حوزه کنترل چراغ ترافیک مبتنی بر RL با داده‌کارآمدی و قابلیت تعمیم بالا در محیط‌های V2X تبدیل می‌کند. نتایج تجربی گسترده نشان داد که UniTSAکوتاه‌ترین زمان انتظار متوسط را در میان انواع مختلف تقاطع‌ها به دست می‌آورد، عملکرد بهتری نسبت به روش‌های موجود ارائه می‌دهد و حتی از مدل‌های آموزش‌دیده از صفر با فاین‌تیونینگ پیشی می‌گیرد.

قدردانی(Acknowledgments)

این تحقیق توســط برنامه ملی کلیدی تحقیقات و توســعه چین تحت گرنت شــماره 2020YFB1807700و برنامه Shanghai Pujiangتحت گرنت شماره 21PJD092حمایت شده است.

References

[1] Fehda Malik, Hasan Ali Khattak, and Munam Ali Shah. Evaluation of the impact of traffic congestion based on

sumo. In 2019 25th International Conference on Automation and Computing (ICAC), pages 1–5. IEEE, 2019.

[2] Alan J Miller. Settings for fixed-cycle traffic signals. Journal of the Operational Research Society, 14(4):373–386,

1963.

[3] Thomas Urbanik, Alison Tanaka, Bailey Lozner, Eric Lindstrom, Kevin Lee, Shaun Quayle, Scott Beaird,

Shing Tsoi, Paul Ryus, Doug Gettman, et al. Signal timing manual, volume 1. Transportation Research Board

Washington, DC, 2015.

[4] Carlos Gershenson. Self-organizing traffic lights. arXiv preprint nlin/0411066, 2004.

[5] Ishu Tomar, S Indu, and Neeta Pandey. Traffic signal control methods: Current status, challenges, and emerging

trends. Proceedings of Data Analytics and Management: ICDAM 2021, Volume 1, pages 151–163, 2022.

[6] Wang Tong, Azhar Hussain, Wang Xi Bo, and Sabita Maharjan. Artificial intelligence for vehicle-to-everything:

A survey. IEEE Access, 7:10823–10843, 2019.

[7] Tamás Wágner, Tamás Ormándi, Tamás Tettamanti, and István Varga. Spat/map v2x communication between

traffic light and vehicles and a realization with digital twin. Computers and Electrical Engineering, 106:108560,

2023.

16

[8] Yit Kwong Chin, Lai Kuan Lee, Nurmin Bolong, Soo Siang Yang, and Kenneth Tze Kin Teo. Exploring q-learning

optimization in traffic signal timing plan management. In 2011 third international conference on computational

intelligence, communication systems and networks, pages 269–274. IEEE, 2011.

[9] Guanjie Zheng, Yuanhao Xiong, Xinshi Zang, Jie Feng, Hua Wei, Huichu Zhang, Yong Li, Kai Xu, and Zhenhui

Li. Learning phase competition for traffic signal control. In Proceedings of the 28th ACM international conference

on information and knowledge management, pages 1963–1972, 2019.

[10] Chacha Chen, Hua Wei, Nan Xu, Guanjie Zheng, Ming Yang, Yuanhao Xiong, Kai Xu, and Zhenhui Li. Toward a

thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. In Proceedings of

the AAAI Conference on Artificial Intelligence, volume 34, pages 3414–3421, 2020.

[11] Xinshi Zang, Huaxiu Yao, Guanjie Zheng, Nan Xu, Kai Xu, and Zhenhui Li. Metalight: Value-based metareinforcement learning for traffic signal control. In Proceedings of the AAAI Conference on Artificial Intelligence,

volume 34, pages 1153–1160, 2020.

[12] Enming Liang, Zicheng Su, Chilin Fang, and Renxin Zhong. Oam: An option-action reinforcement learning framework for universal multi-intersection control. In Proceedings of the AAAI Conference on Artificial Intelligence,

volume 36, pages 4550–4558, 2022.

[13] Azzedine Boukerche, Dunhao Zhong, and Peng Sun. A novel reinforcement learning-based cooperative traffic

signal system through max-pressure control. IEEE Transactions on Vehicular Technology, 71(2):1187–1198,

2022.

[14] Liang Zhang, Qiang Wu, Jun Shen, Linyuan Lü, Bo Du, and Jianqing Wu. Expression might be enough:

representing pressure and demand for reinforcement learning based traffic signal control. In International

Conference on Machine Learning, pages 26645–26654. PMLR, 2022.

[15] Yuanhao Xiong, Guanjie Zheng, Kai Xu, and Zhenhui Li. Learning traffic signal control from demonstrations.

In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pages

2289–2292, 2019.

[16] Stefano Giovanni Rizzo, Giovanna Vantini, and Sanjay Chawla. Time critic policy gradient methods for traffic

signal control in complex and congested scenarios. In Proceedings of the 25th ACM SIGKDD International

Conference on Knowledge Discovery and Data Mining, KDD '19, page 1654–1664, New York, NY, USA, 2019.

Association for Computing Machinery.

[17] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. Multi-agent deep reinforcement learning for large-scale

traffic signal control. IEEE Transactions on Intelligent Transportation Systems, 21(3):1086–1095, 2019.

[18] Afshin Oroojlooy, Mohammadreza Nazari, Davood Hajinezhad, and Jorge Silva. Attendlight: Universal attentionbased reinforcement learning model for traffic signal control. Advances in Neural Information Processing Systems,

33:4079–4090, 2020.

[19] Zian Ma, Chengcheng Xu, Yuheng Kan, Maonan Wang, and Wei Wu. Adaptive coordinated traffic control for

arterial intersections based on reinforcement learning. In 2021 IEEE International Intelligent Transportation

Systems Conference (ITSC), pages 2562–2567. IEEE, 2021.

[20] Seyed Sajad Mousavi, Michael Schukat, and Enda Howley. Traffic light control using deep policy-gradient and

value-function-based reinforcement learning. IET Intelligent Transport Systems, 11(7):417–423, 2017.

[21] Mohammad Aslani, Mohammad Saadi Mesgari, Stefan Seipel, and Marco Wiering. Developing adaptive traffic

signal control by actor–critic and direct exploration methods. In Proceedings of the Institution of Civil EngineersTransport, volume 172, pages 289–298. Thomas Telford Ltd, 2019.

[22] Haoran Su, Yaofeng D Zhong, Joseph YJ Chow, Biswadip Dey, and Li Jin. Emvlight: A multi-agent reinforcement learning framework for an emergency vehicle decentralized routing and traffic signal control system.

Transportation Research Part C: Emerging Technologies, 146:103955, 2023.

[23] Elise Van der Pol and Frans A Oliehoek. Coordinated deep reinforcement learners for traffic light control.

Proceedings of learning, inference and control of multi-agent systems (at NIPS 2016), 8:21–38, 2016.

[24] Patrick Mannion, Jim Duggan, and Enda Howley. An experimental review of reinforcement learning algorithms

for adaptive traffic signal control. Autonomic road transport support systems, pages 47–66, 2016.

[25] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. Intellilight: A reinforcement learning approach for

intelligent traffic light control. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge

Discovery & Data Mining, pages 2496–2505, 2018.

[26] Lun-Hui Xu, Xin-Hai Xia, and Qiang Luo. The study of reinforcement learning for traffic self-adaptive control

under multiagent markov game environment. Mathematical Problems in Engineering, 2013, 2013.

17

[27] Mohammad Aslani, Mohammad Saadi Mesgari, and Marco Wiering. Adaptive traffic signal control with actorcritic methods in a real-world traffic network with different traffic disruption events. Transportation Research Part

C: Emerging Technologies, 85:732–752, 2017.

[28] Mohammad Aslani, Stefan Seipel, Mohammad Saadi Mesgari, and Marco Wiering. Traffic signal optimization

through discrete and continuous reinforcement learning with robustness analysis in downtown tehran. Advanced

Engineering Informatics, 38:639–655, 2018.

[29] Halit Bugra Tulay and Can Emre Koksal. Road state inference via channel state information. IEEE Transactions

on Vehicular Technology, pages 1–14, 2023.

[30] Mohamed MG Farag, Hesham A Rakha, Emadeldin A Mazied, and Jayanthi Rao. Integration large-scale modeling

framework of direct cellular vehicle-to-all (c-v2x) applications. Sensors, 21(6):2127, 2021.

[31] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu

Chen. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685, 2021.

[32] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich,

Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. Microscopic traffic simulation using

sumo. In 2018 21st international conference on intelligent transportation systems (ITSC), pages 2575–2582.

IEEE, 2018.

[33] Peter Koonce and Lee Rodegerdts. Traffic signal timing manual. Technical report, United States. Federal Highway

Administration, 2008.

[34] Arthur G Sims and Kenneth W Dobinson. The sydney coordinated adaptive traffic (scat) system philosophy and

benefits. IEEE Transactions on vehicular technology, 29(2):130–137, 1980.

[35] Pravin Varaiya. The max-pressure controller for arbitrary networks of signalized intersections. Advances in

dynamic network modeling in complex transportation systems, pages 27–66, 2013.

[36] Hua Wei, Guanjie Zheng, Vikash Gayah, and Zhenhui Li. A survey on traffic signal control methods. arXiv

preprint arXiv:1904.08117, 2019.

[37] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A

survey. IEEE transactions on pattern analysis and machine intelligence, 44(9):5149–5169, 2021.

[38] Misha Laskin, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel, and Aravind Srinivas. Reinforcement

learning with augmented data. Advances in neural information processing systems, 33:19884–19895, 2020.

[39] Ilya Kostrikov, Denis Yarats, and Rob Fergus. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. arXiv preprint arXiv:2004.13649, 2020.

[40] Nicklas Hansen and Xiaolong Wang. Generalization in reinforcement learning by soft data augmentation. In 2021

IEEE International Conference on Robotics and Automation (ICRA), pages 13611–13617. IEEE, 2021.

[41] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization

algorithms. arXiv preprint arXiv:1707.06347, 2017.

[42] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stablebaselines3: Reliable reinforcement learning implementations. The Journal of Machine Learning Research,

22(1):12348–12355, 2021.

[43] James Ault and Guni Sharon. Reinforcement learning benchmarks for traffic signal control. In Thirty-fifth

Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1), 2021.

[44] Seung-Bae Cools, Carlos Gershenson, and Bart D'Hooghe. Self-organizing traffic lights: A realistic simulation.

Advances in applied self-organizing systems, pages 45–55, 2013.

[45] Hua Wei, Chacha Chen, Guanjie Zheng, Kan Wu, Vikash Gayah, Kai Xu, and Zhenhui Li. Presslight: Learning

max pressure control to coordinate traffic signals in arterial network. In Proceedings of the 25th ACM SIGKDD

International Conference on Knowledge Discovery & Data Mining, pages 1290–1298, 2019