



طراحی و پیاده‌سازی کنترل‌کننده هوشمند تطبیقی برای کنترل تقاطع با استفاده از روش یادگیری تقویتی

مجید محمدی^۱، عباس دیدبان^۲، مهناز اثباتی^۳

۱- دانشجوی دکترا، گروه برق و کامپیوتر، دانشگاه سمنان

۲- دانشیار، گروه برق و کامپیوتر، دانشگاه سمنان

۳- دانشجوی کارشناسی ارشد، گروه برق و کامپیوتر، دانشگاه سمنان

چکیده

امروزه با توسعه زندگی شهرنشینی و افزایش نیاز به حمل و نقل، معضل ازدحام ترافیک تبدیل به یکی از بزرگترین مشکلات پیش‌روی جوامع بشری شده است. لذا برای دستیابی به توسعه پایدار در زمینه مدیریت یکپارچه شهری، کنترل شبکه‌های حمل و نقل امری اجتناب‌ناپذیر است. این کنترل شامل دو رویکرد عمده تقاطع‌ها و شریان‌های شهری و شبکه‌های بزرگراهی می‌باشد. وجود دینامیک‌ها و نایقینی‌های زیاد در سیستم کنترل یک تقاطع، مدل‌سازی آن را دشوار می‌سازد. بنابراین رویکرد استفاده از الگوریتم‌های کنترلی مستقل از مدل اهمیت پیدا می‌کند. چارچوب یادگیری تقویتی به عنوان یک روش مستقل از مدل، بستر مناسبی برای کنترل ترافیک فراهم می‌سازد. از ویژگی‌های کنترل یادگیری تقویتی می‌توان به هوشمندی، بهینگی و تطبیقی بودن آن اشاره کرد. پرکاربردترین روشهای پیاده‌سازی یادگیری تقویتی Q-Learning و SARSA هستند که در این مقاله بر مبنای آنها کنترل‌رلهایی برای کنترل ترافیک یک تقاطع ایزوله ارائه و همچنین روش‌های فازی Q-Learning و فازی SARSA مورد مطالعه قرار گرفته‌اند. برای تحلیل کارایی این کنترل‌کننده‌ها، یک سه‌راه در شهر تهران با داده‌های حقیقی در نرم‌افزار Aimsun شبیه‌سازی و آنالیز شده است. نتایج به‌دست آمده نشانگر بهبود عملکرد شبکه نسبت به روش‌های معمول علم مهندسی ترافیک می‌باشد.

کلید واژه: تقاطع، شبکه‌های ترافیک شهری، کنترل هوشمند ترافیک، یادگیری تقویتی.

^۱ دانشجوی دکترا، ۰۹۱۲۵۹۶۲۴۳۴، ma.mohammadi@semnan.ac.ir

^۲ دانشیار گروه کنترل، adideban@semnan.ac.ir



۱- مقدمه

ترافیک و در نتیجه آن گرفتگی و کندی معابر یک واقعیت غیر قابل انکار در اکثر شهرهای بزرگ دنیا است. این مسئله منجر به زیان‌های اقتصادی، آلودگی هوا، آلودگی صوتی و در کل افت کیفیت زندگی می‌شود. راه‌حل اولیه‌ای که برای پاسخگویی به این مشکل به نظر می‌رسد گسترش زیرساخت‌ها مانند تعریض شریان‌ها و ساخت بزرگراه‌های جدید و ایجاد تقاطع‌های غیرهمسطح می‌باشد. اما با توجه به بافت انبوه شهری، هزینه‌های فراوان و ایجاد مزاحمت عبور در طی مراحل ساخت و توسعه راه‌ها و گسترش زیرساخت‌های شهری، این امر روز به روز دشوارتر شده است. از طرفی، گسترش شبکه‌های شهری به صورت مقطعی باعث ایجاد تراکم در شبکه پایین دست خواهد شد. با توجه به این مشکلات، استفاده از روش هوشمند برای مدیریت و هدفمند کردن ترافیک در جهت بازدهی و عملکرد بهتر معابر ضروری به نظر می‌رسد. در این روش با مطالعه ترافیک خودروها و بررسی ظرفیت معابر، محدودیت‌هایی بر حرکت خودروها اعمال می‌شود تا شبکه حمل و نقل در بهترین حالت خود قرار گیرد. استفاده از این راهکار علاوه بر اینکه از نظر صرفه اقتصادی بازدهی بالایی دارد، می‌تواند در کوتاه‌مدت مورد استفاده قرار گیرد و مشکلات ترافیک را به صورت مقطعی حل کند. البته باید توجه داشت که این راه‌حل جایگزین گسترش زیرساخت‌های شهری نمی‌باشد و تنها به‌گونه‌ای عمل می‌کند که از امکانات موجود حداکثر بازدهی استخراج شود.

به طور معمول هدف در این‌گونه مسائل، کاهش زمان سفر و افزایش سرعت جابجایی وسایل نقلیه در شبکه می‌باشد. با توجه به اینکه محیط و اجزای ترافیکی دائماً در حال تغییر هستند، کنترلر بهینه باید به صورت برخط^۱ و تطبیقی^۲ باشد تا بتواند با تغییرات سیستم در زمان موثر باشد. از این رو مدیریت دینامیکی ترافیک^۳ رویکرد اصلی این مقاله است.

از آنجایی که ایجاد یک مدل واقعی و مناسب برای سیستم‌های حمل و نقل به دلیل وسعت جغرافیایی، دینامیک پیچیده، وجود اغتشاشات و نویز امری دشوار و ناممکن به نظر می‌رسد، استفاده از روش‌های مستقل از مدل^۴ اهمیت پیدا می‌کند. از این رو روشی که در این مقاله مدنظر است، حل مسئله کنترل هوشمند ترافیک با کمک یک روش مستقل از مدل به نام یادگیری تقویتی^۵ است. در این مقاله چارچوب یادگیری تقویتی برای کنترل ترافیک یک تقاطع مجزا پیاده می‌گردد. سپس چگونگی پیاده‌سازی کنترلرها در شبیه‌ساز Aimsun توسط ماژول API آن معرفی می‌شود. این آزمایش‌ها بر روی یک سهرای واقعی با داده‌های واقعی انجام و کارایی آن از طریق مقایسه با یک روش متداول مهندسی ترافیک نشان داده می‌شود.

¹ Online

² Adaptive

³ Dynamic Traffic Management

⁴ Model Free

⁵ Reinforcement Learning



۲- پیشینه تحقیق

تاکنون در جهان تحقیقات مختلفی بر روی کنترل ترافیک صورت گرفته است. در [۱] برای کنترل ترافیک در محدوده‌های عملیاتی^۱ در بزرگراه‌ها از یک شبکه عصبی برای پیش‌بینی صف خودروها استفاده شده است که با کمک سیستم VMS این اطلاعات به رانندگان داده می‌شود. در [۲] با استفاده از منطق فازی کنترل‌کننده‌ای برای چندین تقاطع متوالی طراحی شده است که علاوه بر کاهش تاخیر برای خودروها، عابرین پیاده را نیز مدنظر قرار می‌دهد. توابع عضویت^۲ منطق فازی توسط ژنتیک الگوریتم بهینه می‌شوند. در [۳] با در نظر گرفتن مدلی برای یک تقاطع از برنامه‌ریزی پویای تقریبی^۳ برای کنترل تطبیقی سیگنال چراغ راهنمایی استفاده شده است. این روش با تقریب زدن، از حجم محاسبات برای حل برنامه‌ریزی پویا می‌کاهد. در [۴] نیز از همین روش برنامه‌ریزی پویای تقریبی استفاده شده است، با این تفاوت که در این رویکرد، دو شبکه عصبی، تصمیم‌گیری و ارزیابی تصمیم را بر عهده دارند، لذا این روش به مدل احتیاج ندارد. در [۵] یک روش ساده برای تعیین زمان‌بندی چندین چراغ راهنمایی در یک شبکه ارائه شده است، که هدف آن این است که میزان تاخیر برای همه خودروها برابر شود. در [۶] یک ساختار چندعامله سلسله‌مراتبی^۴ معرفی شده است، که در پایین‌ترین سطح هر عامل، کنترل یک تقاطع را به عهده دارد. در سطح میانی، یک عامل، چند کنترلر مربوط به تقاطع‌های درون یک منطقه را هماهنگ می‌کند. در نهایت، در لایه آخر یک عامل مرکزی بر فعالیت همه سیستم نظارت می‌کند. در [۷] مقدماتی از تئوری یادگیری تقویتی ارائه شده و امکان به‌کارگیری آن در کاربردهای حمل و نقلی بررسی شده است. در [۸] یک کنترل‌کننده برای زمان‌بندی یک چراغ راهنمایی با روش Q-Learning برای کنترل یک تقاطع معرفی شده است و با یک چراغ راهنمایی با زمان‌بندی ثابت مقایسه شده است. در [۹] از روش Q-Learning برای کنترل یک کریدور آزادراه-خیابان استفاده شده است. برای این کار روی قسمتی از شهر تورنتو که در نرم‌افزار Paramics مدل و کالیبره شده است، شبیه‌سازی‌ها انجام گرفته شده است. یک روش نوروفازی نیز در [۱۰] ارائه شده که بهینه‌سازی توابع عضویت آن به وسیله یادگیری تقویتی انجام می‌گیرد. در [۱۱] نیز از Q-Learning برای بهینه‌سازی و تصمیم‌سازی بین خروجی‌های قواعد فازی استفاده شده است.

۳- یادگیری تقویتی

یادگیری تقویتی^۵ (RL) روشی برای یافتن عمل بهینه در محیط‌های با عدم قطعیت است [۱۲]. در این روش یک عامل با انجام یک سری عمل و یادگیری از نتایج این اعمال، سعی در یافتن عمل بهینه در یک محیط پویا دارد. در

¹ Work Zone

² Membership Function

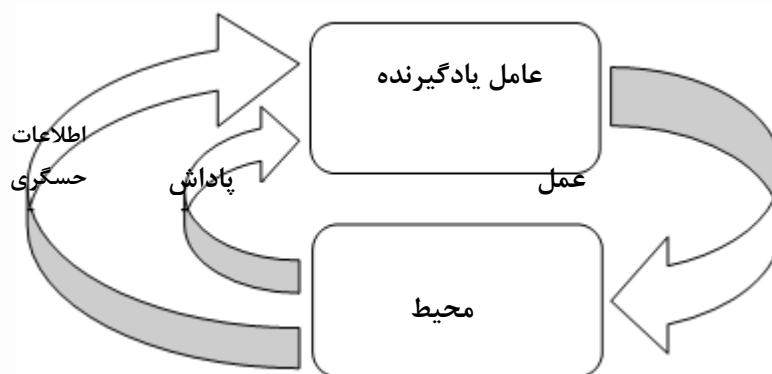
³ Approximate Dynamic Programming

⁴ Hierarchical MultiAgent Structure

⁵ Reinforcement learning



نهایت این عامل یاد می‌گیرد که در هر وضعیت چه عملی را انجام دهد، به بیان دیگر پاسخ بهینه در مسائل یادگیری تقویتی یافتن یک نگاشت از مجموعه وضعیت‌ها به مجموعه اعمال است. در یادگیری تقویتی، برای عامل جواب صحیح موجود نیست؛ بلکه بعد از انجام یک عمل در یک وضعیت، عمل انجام شده بر روی محیط اجرا شده و بر اساس یک معیار معین، میزان خوب یا بد بودن عمل ارزیابی می‌شود و به عامل به‌عنوان یک بازخورد اطلاع داده می‌شود. بدین ترتیب سازگاری پدید می‌آید تا عامل بدون نیاز به دانستن مدل محیط و با گذشت زمان یاد بگیرد که در وضعیت‌های گوناگون، کدام عمل را انتخاب کند.



شکل ۱: پروسه تعاملات عامل با محیط در یادگیری تقویتی

عامل با کمک روش‌های مناسب انتخاب عمل می‌تواند رفتار مناسب را پیاده‌سازی کند. دو روش معروف انتخاب عمل که در این مقاله نیز از آنها استفاده شده است، روش بیشینه نرم^۱ و ϵ -greedy می‌باشند. روش‌های یادگیری تقویتی به سه دسته رویکرد عمده تقسیم‌بندی می‌شوند: [۱۲]
روش برنامه‌ریزی پویا^۲، روش مونت کارلو^۳، روش تفاضل زمانی که از جمله معروف‌ترین روش‌های تفاضل زمانی، الگوریتم‌های Q-Learning و SARSA هستند.

۴- مدل سازی مساله در چارچوب یادگیری تقویتی

در این بخش چارچوب یادگیری تقویتی برای کنترل ترافیک یک تقاطع ایزوله پیاده‌سازی می‌گردد. در ذیل به شرح اجزای مدل مساله در چارچوب یادگیری تقویتی پرداخته می‌شود. در این مدل‌سازی، عامل برای کنترل سیستم، چراغ راهنمایی هوشمند است. تعیین زمان‌بندی چراغ راهنمایی عملی است که عامل می‌تواند بر روی محیط اجرا کند. به طور معمول برای هر فاز یک حداقل زمان سبز و یک حداکثر زمان سبز بر اساس مینیمم و

¹ SoftMax

² Temporal Difference

³ Monte Carlo



ماکزیمم تقاضای خودروها معین می‌گردد که در اینجا فواصل ده ثانیه‌ای در نظر گرفته شده است. این زمان‌بندی به دو روش پیاده می‌شود: ۱- روش چرخه‌ای^۱ که در این روش ترتیب فازها مشخص و ثابت است. ۲- روش غیر چرخه‌ای^۲: در این روش فازها ترتیب مشخص و ثابتی ندارند. در این مقاله روش چرخه‌ای مورد بررسی قرار می‌گیرد. مجموعه یک تقاطع همراه با تمام اجزایی که در آن نقش ایفا می‌کنند، به عنوان محیط برای عامل در نظر گرفته می‌شود. مهمترین مشخصه برای محیط، تعریف وضعیت‌ها است. وضعیت هر تقاطع بر مبنای تعداد خودروهای قرار گرفته در پشت چراغ برای همه مسیرهای ورودی تعریف شده است. با تعریف تعداد خودروها در هر مسیر ورودی به عنوان وضعیت تقاطع، گسسته‌سازی این مقادیر نیز چالش برانگیز است. یک رویکرد که در این مقاله اتخاذ شده است، استفاده از روش‌های یادگیری تقویتی فازی؛ Fuzzy Q-Learning و Fuzzy SARSA است. دو روش گسسته‌سازی دیگر نیز در این مقاله اعمال شده است که در اولین روش، تعداد خودروها به دسته‌های خطی افزایش می‌گردد. در دومین روش، طول دسته‌ها به صورت نمایی افزایش می‌یابد.

پاداشی که برای عامل تعریف شده است، اختلاف تعداد خودروها قبل و بعد از فاز در همه مسیرهای ورودی است. برای مقایسه عادلانه بین الگوریتم‌ها، سه معیار ارزیابی کارایی^۳ (MoE) در نظر گرفته می‌شود. معیار اول میانگین طول صف خودروها برای هر خط در هر مسیر ورودی است. معیار دوم زمان سفر است که برای کل شبکه محاسبه می‌گردد. دو معیار قبلی تعیین‌کننده سطح کیفیت و بهینگی الگوریتم است. اما معیار سوم انحراف از معیار زمان سفر خودروهاست که به نوعی مشخص‌کننده مساوات الگوریتم بین تمام خودروها می‌باشد.

۵- پیاده‌سازی الگوریتم‌ها در شبیه‌ساز AIMSUN

برای مقایسه عملکرد الگوریتم‌ها و بررسی انتقال‌پذیری^۴ چارچوب یادگیری تقویتی، مطالعه موردی^۵ یک سه‌راه حقیقی در شهر تهران مورد آزمایش قرار گرفته است. این سه‌راه در مرکز شهر قرار دارد و تقاطع خیابان آزادی با خیابان رودکی است. این سه‌راه به دلیل موقعیت جغرافیایی دارای حجم ترافیکی بالایی است و خطوط BRT نیز از این مسیر عبور می‌کنند. تمامی موارد ذکر شده و تاکید بر این نکته که می‌توان این سه‌راه را ایزوله فرض کرد، همگی دلایلی هستند که بررسی این تقاطع را موجه می‌کنند. برای مطالعه کنترل‌کننده طراحی شده، در ابتدا هندسه و جغرافیای تقاطع در نرم‌افزار شبیه‌ساز پیاده می‌گردد. شکل ۲ نمای سه‌راه مورد نظر را در نرم‌افزار AIMSUN در حین شبیه‌سازی نشان می‌دهد. در پیاده‌سازی کنترل‌کننده یادگیری تقویتی، هر اپیزود برابر یک روز

¹ Cyclic

² Acyclic

³ Measurement of Effectiveness

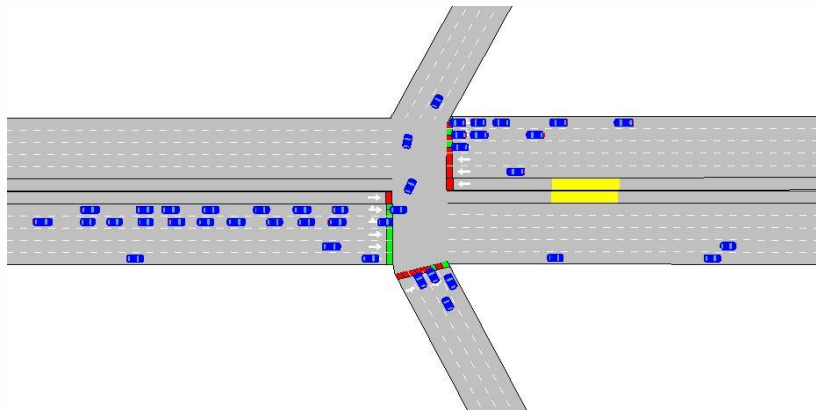
⁴ Transferability

⁵ Case Study

⁶ 3-leg junction



در نظر گرفته شده است و آزمایش برای ۴۵ روز (اپیزود) انجام شده است که ۴۰ روز اول مرحله یادگیری^۱ سیستم است و پنج روز بعدی مرحله تست الگوریتم است.



شکل ۲: نمای تقاطع در نرم افزار AIMSUN

برای پیاده‌سازی الگوریتم‌های یادگیری تقویتی در نرم‌افزار شبیه‌ساز AIMSUN، از ماژول API آن استفاده می‌شود. برای ارتباط با API به منظور نوشتن الگوریتم کنترلی از زبان C++ استفاده شده است.

۶- آزمایش‌ها، نتایج و تحلیل‌ها

در جدول ۱، پارامترهای روش‌های انتخاب عمل در کنترلرهای مختلف و برای تقاطع‌های مختلف آورده شده است:

جدول ۱: پارامترهای روش انتخاب عمل در آزمایش‌های مختلف

		سره‌راه
چرخه‌ای	ϵ -greedy	$\epsilon_0 = 0.7 ; \epsilon_{t+1} = \frac{\epsilon_t}{1.0002}$
	SoftMax	$\tau_0 = 10 ; \tau_{t+1} = \frac{\tau_t}{1.0003}$

در جدول ۲ مقادیر حداقل و حداکثر زمان سبز برای هر فاز و در هر تقاطع آورده شده است.

¹ Training



جدول ۲: مقادیر حداقل و حداکثر زمان سبز برای هر فاز

چرخه‌ای	سه‌راه	
	فاز ۱	{20-150}
	فاز ۲	{20-80}
	فاز ۳	{20-60}

در جدول ۳ نیز نحوه گسسته‌سازی وضعیت محیط نشان داده شده است. در تقاطع سه‌راه، مسیر ورودی ۱ خیابان آزادی غرب به شرق، مسیر ورودی ۲ خیابان آزادی شرق به غرب و مسیر ورودی ۳ خیابان رودکی است. اعداد داخل کروشه، مرزبندی بین وضعیت‌ها را مشخص می‌کند و اعداد ستون سمت راست، تعداد وضعیت‌ها را در هر آزمایش مشخص می‌کند. توابع عضویت فازی نیز که در پیاده‌سازی کنترلر FQL-FSL کاربرد دارند، نمایش داده شده‌اند.

جدول ۳: نحوه گسسته‌سازی وضعیت هر تقاطع

سه‌راه	خطی	مسیر ورودی ۱	{30-40-50-60-70-80-90-100-110-120-130-140-150}	۱۴
		مسیر ورودی ۲	{20-30-40-50-60-70-80-90-100-110-120}	۱۲
		مسیر ورودی ۳	{10-20-30-40}	۵
	نمایی	مسیر ورودی ۱	{25-32-43-60-85-120-167}	۸
		مسیر ورودی ۲	{15-22-33-50-75-110}	۷
		مسیر ورودی ۳	{5-15-30}	۴
	فازی	مسیر ورودی ۱		۵



	مسیر ورودی ۲		۵
	مسیر ورودی ۳		۳

در نهایت در جدول ۴ سایز جدول Q برای هر کنترلر آورده شده است. اولین اعداد نشان‌دهنده تعداد وضعیت‌ها و اعداد آخرین نشان‌دهنده تعداد اعمال هستند.

جدول ۱: سایز جدول Q هر کنترل‌کننده

		چرخه‌ای	
سره‌راه	خطی	$14 \times 12 \times 5 \times (14 + 7 + 5)$	۲۱۸۴۰
	نمایی	$8 \times 7 \times 4 \times (14 + 7 + 5)$	۵۸۲۴
	فازی	$5 \times 5 \times 3 \times (14 + 7 + 5)$	۱۹۵۰

۷- نتایج کنترل‌کننده‌ها

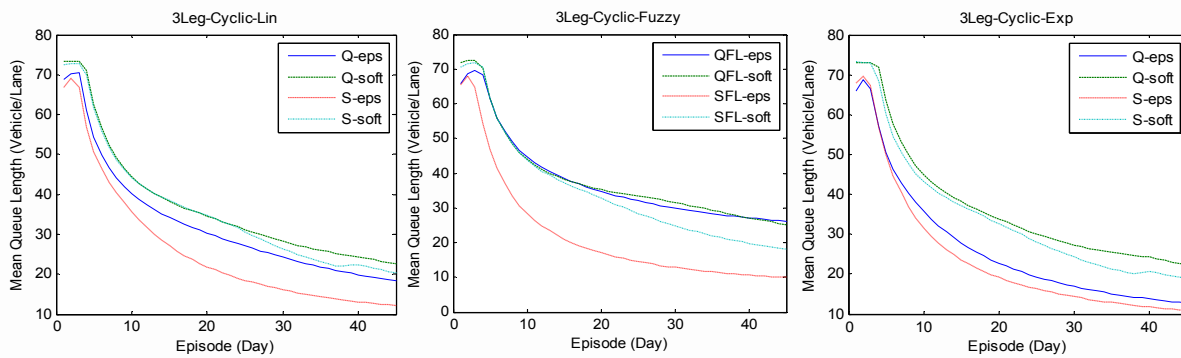
در جدول ۵ نتایج معیار اول ارزیابی عملکرد، که میانگین طول صف است آمده است.

جدول ۲: نتایج معیار اول (میانگین طول صف) برای تمام کنترلرهای مختلف

		وضعیت سرراه		
		خطی	نمایی	فازی
چرخه‌ای	Q-Learning ϵ -greedy	۱۸,۸۷۶۶	۱۳,۱۰۴۵	۲۶,۵۲۶۴
	Q-Learning SoftMax	۲۳,۲۸۶۶	۲۳,۰۴۴۸	۲۵,۹۶۰۲
	Sarsa ϵ -greedy	۱۲,۴۶۹۹	۱۱,۲۳۲۷	۱۰,۲۱۱۹



شکل ۳ نحوه رفتار معیار اول (میانگین طول صف) در طول یادگیری برای تقاطع سه‌راهی نمایش داده شده است. به ترتیب از چپ به راست وضعیت‌های خطی، نمایی و فازی آورده شده است.



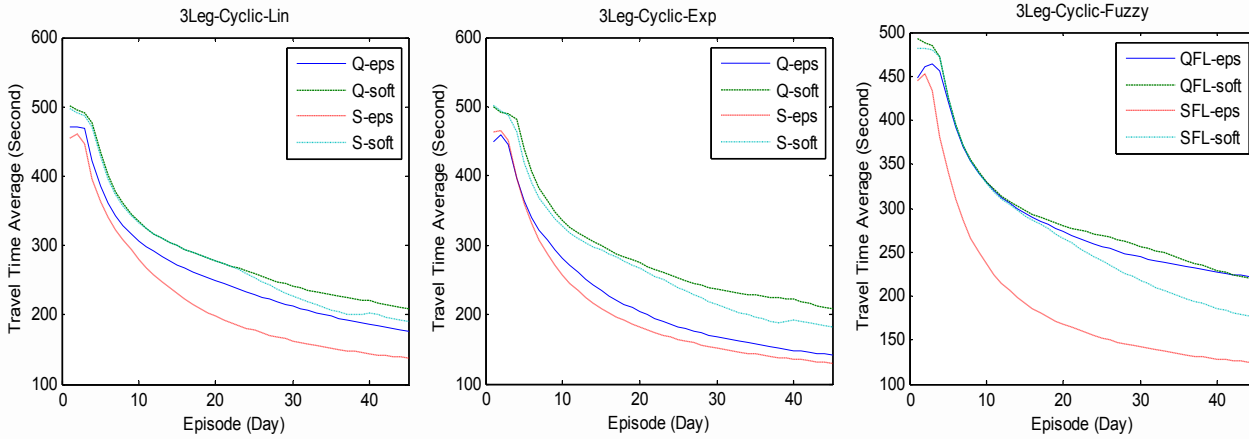
شکل ۳: تغییرات معیار اول ارزیابی عملکرد (میانگین طول صف) در ۴۵ اپیزود یادگیری برای تقاطع سه‌راه

در جدول ۶ نتایج معیار دوم ارزیابی عملکرد، که میانگین زمان سفر است آمده است.

جدول ۳: نتایج معیار دوم (زمان سفر) برای تمام آزمایش‌های مختلف

		وضعیت سه‌راه		
		خطی	نمایی	فازی
چرخه‌ای	Q-Learning ϵ -greedy	۱۷۹,۶۶۴۱	۱۴۴,۲۷۳۷	۲۲۳,۵۱۰۰
	Q-Learning SoftMax	۲۱۲,۹۵۹۴	۲۱۳,۴۷۰۴	۲۲۲,۵۳۶۶
	Sarsa ϵ -greedy	۱۳۹,۹۷۵۲	۱۳۲,۴۱۳۴	۱۲۵,۹۲۶۷
	Sarsa SoftMax	۱۹۵,۰۰۹۰	۱۸۵,۴۸۸۳	۱۷۹,۵۴۲۷

در شکل ۴ نحوه رفتار معیار دوم (میانگین زمان سفر) در طول یادگیری برای تقاطع سه‌راهی نشان داده شده است. به ترتیب از چپ به راست وضعیت‌های خطی، نمایی و فازی آورده شده است.



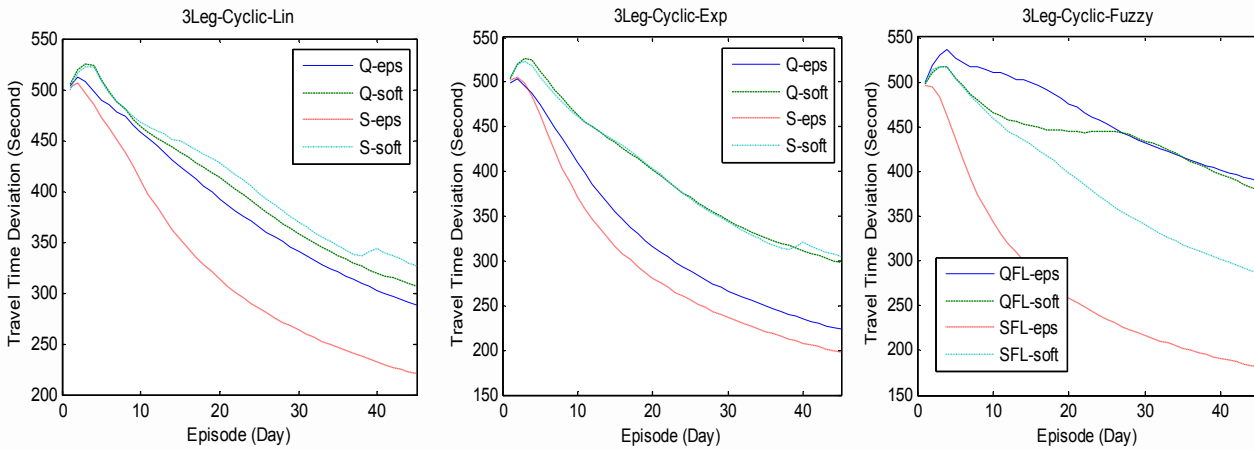
شکل ۳۰: تغییرات معیار دوم ارزیابی عملکرد (میانگین زمان سفر) در ۴۵ اپیزود یادگیری برای تقاطع سه‌راه

در جدول ۷ نتایج معیار سوم ارزیابی عملکرد، که انحراف از معیار زمان سفر است آمده است.

جدول ۴: نتایج معیار سوم (انحراف معیار زمان سفر) برای تمام آزمایش‌های مختلف

		وضعیت سه‌راه		
		خطی	نمایی	فازی
چرخه‌ای	Q-Learning ϵ -greedy	۲۹۳,۶۵۴۳	۲۲۷,۶۰۱۲	۳۹۳,۶۱۶۰
	Q-Learning SoftMax	۳۱۱,۸۶۳۶	۳۰۲,۹۵۱۱	۳۸۶,۰۴۰۰
	Sarsa ϵ -greedy	۲۲۴,۶۶۴۵	۲۰۱,۷۶۳۲	۱۸۵,۱۶۶۸
	Sarsa SoftMax	۳۳۳,۰۵۵۶	۳۱۰,۳۶۷۴	۲۹۱,۸۷۲۷

در شکل ۵ نحوه رفتار معیار سوم (انحراف از معیار زمان سفر) در طول یادگیری برای تقاطع سه‌راهی نمایش داده شده است.



شکل ۵: تغییرات معیار سوم ارزیابی عملکرد (انحراف از معیار زمان سفر) در ۴۵ اپیزود یادگیری برای تقاطع سه‌راه

اکنون برای آنالیز بیشتر، بهترین الگوریتم کنترل ترافیک (SARSA ϵ -greedy) را با روش معمول مهندسی ترافیک که همان چراغ سازگار با ترافیک است مقایسه می‌گردد. نتایج و مقادیر معیارهای ارزیابی برای کنترل تقاطع با چراغ سازگار با ترافیک برای مطالعه موردی در جدول ۸ آورده شده است. در مقایسه با نتایج کنترلر هوشمند، برتری‌های روش پیشنهاد شده به وضوح دیده می‌شود.

جدول ۵: نتایج کنترل تقاطع‌ها با چراغ سازگار با ترافیک

	میانگین طول صف	میانگین زمان سفر	انحراف از معیار زمان سفر
سه‌راه	۲۵,۶۳	۲۳۸,۶۰	۴۱۶,۶۲

۸- نتیجه‌گیری

نتایج بدست آمده در این مقاله را می‌توان به صورت زیر خلاصه نمود:

- در این مقاله مساله کنترل ترافیک در تقاطع‌ها و شریان‌های شهری با استفاده از الگوریتم مستقل از مدل یادگیری تقویتی مطرح و بررسی گردید. سپس چارچوب یادگیری تقویتی برای کنترل ترافیک یک تقاطع ایزوله پیاده‌سازی شد. همچنین با در نظر گرفتن دو الگوریتم Q-Learning و SARSA، دو روش انتخاب عمل ϵ -greedy و بیشینه نرم و سه گونه گسسته‌سازی وضعیت‌ها (خطی، نمایی، فازی) آزمایش‌های متفاوتی مورد بررسی قرار گرفت. این آزمایش‌های (کنترلرهای) متفاوت توسط ماژول API شبیه‌ساز AIMSUN بر روی مطالعه موردی یک سه‌راه در شهر تهران با داده‌های حقیقی پیاده‌سازی شد.



۲. سه معیار ارزیابی میانگین طول صف، میانگین زمان سفر و انحراف از معیار زمان سفر که نمایانگر بهینگی و توانایی برقراری مساوات در بین تمام خودروها بودند معرفی گردید و به مقایسه تفصیلی کنترلرها پرداخته شد. در نهایت الگوریتم SARSA، با روش انتخاب عمل ϵ -greedy، در حالت فازهای چرخه‌ای بهترین روش برای یادگیری کنترل ترافیک شناخته شد.
۳. برای تحلیل بیشتر و نشان دادن کارایی و برتری روش پیشنهادی، بهترین الگوریتم کنترل ترافیک (ϵ -greedy, SARSA) با روش معمول مهندسی ترافیک که همان چراغ سازگار با ترافیک است مقایسه گردید. نتایج حاکی از برتری‌های روش پیشنهادی برای کنترل ترافیک تقاطع بود.
۴. کنترلر پیشنهادی تطبیقی، متغیر با زمان، برخط، بهینه و هوشمند بود و مهمترین ویژگی آن قابلیت عملیاتی شدن برای پیاده‌سازی در تقاطع‌های شهری بود. علاوه بر این کنترلر پیشنهادی قابلیت استفاده در شریان‌های شهری و ویژگی انتقال‌پذیری (استقلال از شرایط جغرافیایی مساله) را نیز دارد.

۹-مراجع

- 1- S .Hooshdar and H .Adeli, 2004,” Toward Intelligent Variable Message Signs in Freeway Work Zones: Neural Network Model,” *Journal of Transportation Engineering* January/February, pp. 83-93,.
- 2- J.D. Schmocker, S. Ahuja, and M.G.H. Bell,2008,“Multi-objective signal control of urban junctions - framework and a London case study,” *Transportation Research Part C*, vol. 16, pp. 454-470.
- 3- C. Cai , C.K. Wong, and B. G. Heydecker, 2009, “Adaptive traffic signal control using approximate dynamic programming,” *Transportation Research Part C*, vol. 17, pp. 456–474.
- 4- T. Li, D. Zhao, and J. Yi, 2007, “Application of ADP to Intersection Signal Control,” *Lecture Notes on Computer Science*, D. Liu et al. (Eds), Part I, vol. 4491, pp. 374–379.
- 5- G. Balan and S. Luke, 2006, “History based Traffic Control,” *AAMAS*, pp. 616-621, Hakodate, Hokkaido, Japan.
- 6- M.C. Choy, D. Srinivasan, and R.L. Cheu , 2003 ,“Cooperative, Hybrid Agent Architecture for Real-Time Traffic Signal Control,” *IEEE Transactions On Systems, Man, And Cybernetics - Part A*, vol. 33, no. 5, pp. 597-607.
- 7- B. Abdulhai, and L. Kattan, 2003,” Reinforcement Learning: Introduction to theory and potential for transport applications,” *Can. J. Civ. Eng.* vol. 30, pp. 981-991.
- 8- B. Abdulhai, R. Pringle, and G.J. Karakoulas, 2003,” Reinforcement Learning for True Adaptive Traffic Signal Control,” *Journal of Transportation Engineering*, vol. 129, no. 3, pp. 278-285.
- 9- C. Jacob, and B. Abdulhai,2009, ” Machine Learning for multi-jurisdictional optimal traffic corridor control,” *Transport. Res. Part A*, doi:10.1016/j.tra.2009.11.001
- 10-E. Bingham,2001,” Reinforcement learning in neurofuzzy traffic signal control,” *European Journal of Operation Research*, vol. 131, pp. 232-241.
- 11-L. Yongquan, and C. Xiangjun, 2009 “Study on Traffic Signal Control Based on Q-Learning,” *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 581-585.
- 12-Richard S. Sutton, and Andrew G. Barto, “Reinforcement Learning: An Introduction,” MIT Press, Cambridge, Massachusetts.



Design and implementation of adaptive intelligent controller for intersection control using Reinforcement Learning

Majid mohammadi, second author name

- 1-PHD. Student ,Control Engineering, Electrical and computer Faculty, Semnan University, Semnan, Iran
2- Associate Professor, Electrical and computer Faculty, Semnan University, Semnan, Iran
3-Mahnaz esbati

Abstract

Nowadays, with development of urbanized life-style and increasing demand to transportation, traffic congestion problem have been become one of the biggest problems facing human societies. Therefore, to achieve sustainable development for integrated urban management, control of transportation networks is inevitable. This control strategy consists of two main approaches such as control of intersections and arterials and control of highway networks. Due to large number of dynamics and uncertainties in the intersection control system, modeling is very complicated task. Thus, approach of using model-free control methods is very important. Reinforcement learning framework as a model-free method seems to be suitable for urban intersections traffic control. Features of reinforcement learning controller may be noted as intelligence, optimality and adaptiveness. Most commonly used methods to implement reinforcement learning are Q-Learning and SARSA that are utilized in this paper to develop variety of controllers for traffic control of isolated junction and the methods of fuzzy Q-Learning and fuzzy SARSA also have been surveyed. To performance evaluation of these controllers, a three-leg junction in Tehran city with the real data has been simulated and analyzed in Aimsun traffic Simulator. The obtained results are all indicate performance improvement comparing traffic engineering common methods.

Keywords: Intersection, Urban traffic networks, Intelligent Traffic control, Reinforcement Learning