مسیریابی ترکیبی وسایل نقلیه و پهپادها جهت کاهش مصرف انرژی با استفاده از الگوریتم های یادگیری تقویتی

تاریخ دریافت:
۱۲ آبان ماه ۱۴۰۳

تاریخ پذیرش:
۱۶ دی ماه ۱۴۰۳

وحید نجف پور*[۱]، صالح یوسفی[۲]، هادی ویسی[۳]

۱. دانشجوی دکتری مهندسی فناوری اطلاعات، گرایش شبکه، دانشگاه ارومیه، ارومیه، ایران.
۲. دانشیار، گروه مهندسی برق و کامپیوتر، دانشگاه ارومیه، ارومیه ایران.
۳. دانشیار، گروه بین رشته ای فناوری، دانشگاه تهران، تهران، ایران.

چکیده

یکی از مباحث مهم که در چند دهه اخیر کاربرد بالایی در هوشمندسازی در شهرهای هوشمند داشته، بحث مسئله مسیریابی سیستم حمل و نقل است. مسئله مسیریابی خودرو و پهپاد درصدد است تا با مدل های ریاضی و بهینه‌سازی به‌گونه‌ای عمل کند که مسافت طی شده، زمان کل سفر و در نهایت تابع هزینه حمل حداقل گردد و در نهایت مصرف انرژی به حداقل برسد. در این مسئله، ابتدا یک مدل ریاضی رسمی از تابع هدف مسیر بهینه پویا بررسی و معیار بهینه‌سازی برای سیستم ارائه می شود. رویکردهای برنامه‌نویسی ریاضی، حداکثر یا به‌حداقل‌رساندن یک عملکرد هدف برای بهبود روند، برای اطمینان از کارایی عملیاتی عمل می‌کند. مسئله حل مسیریابی وسیله نقلیه برای داده‌های بزرگ با محدودیت‌هایی روبرو است. برای افزایش کارایی سیستم‌های مسیریابی وسایل نقلیه، باید مطالعات بیشتری روی همه محدودیت‌های شبکه همچون بازه زمان خدمت رسانی و محدودیت ظرفیت حمل در مسیریابی انجام شود. در این پژوهش مسئله مسیریابی پهپاد و خودرو به منظور امدادرسانی و با هدف کمینه کردن مجموع زمان خدمت رسانی به گره های تقاضا انجام گرفته است و به محدودیت های عملیاتی پهپادها نظیر ظرفیت و مصرف انرژی توجه شده است. ذات مسئله مدیریت حمل و نقل و نیز مسیریابی پهپادها، مسئله VRP است. روش های فراابتکاری جهت مسایل کوچک جوابگو هستند. در مسایل بزرگتر از ۱۰۰ تقاضا، نیاز به الگوریتم های یادگیری خودکار ماشینی پیشنهاد می شود. نتایج با یک مدل آموزش‌دیده روی داده‌های آزمون بزرگ و خارج از توزیع، مقیاس‌پذیری و انعطاف‌پذیری رویکرد پیشنهادی را نسبت به دو الگوریتم فراابتکاری نشان می‌دهد.

واژه‌های کلیدی: مسیریابی پویای وسایل نقلیه، مدیریت حمل و نقل شهری، یادگیری ماشین ، مصرف انرژی، مسیریابی پهپاد.

Combined routing of vehicles and drones to reduce energy consumption using reinforcement learning algorithms

Vahid Najafpoor[1*], Saleh Yousefi[2], Hadi Veysi[3]

1. Ph.D. student of information technology engineering, network orientation, University, Urmia, Iran.
2. Associate Professor, Department of Electrical and Computer Engineering, Urumia University, Urmia, Iran.
3. Associate Professor, Department of Technology, University of Tehran, Tehran, Iran.

Abstract

One of the important topics that has been highly used in smart cities in the last few decades is the issue of transportation system routing. The problem of vehicle and drone routing is trying to work with mathematical models and optimization in such a way that the distance traveled, the total travel time and finally the transportation cost function are minimized and finally the energy consumption is minimized. In this problem, first, a formal mathematical model of the objective function of the dynamic optimal route is investigated and the optimization criterion for the system is presented. Mathematical programming approaches work to maximize or minimize an objective function to improve the process, to ensure operational efficiency. The problem of solving vehicle routing for big data faces some limitations. To increase the efficiency of vehicle routing systems, more studies should be done on all network constraints, problem-related considerations, environment dynamism in routing. In this research, the problem of routing of drones and vehicles for the purpose of providing aid and with the aim of minimizing the total time of service to the demand nodes has been done, and attention has been paid to the operational limitations of drones such as capacity and energy consumption. The essence of the transportation management problem as well as the routing of drones is the VRP problem. Heuristic and meta-heuristic methods are the answer for small problems. In problems larger than 100 requests, the need for automatic machine learning algorithms is suggested. The results with a model trained on large and out-of-distribution test data show the scalability and flexibility of the proposed approach compared to two meta-heuristic algorithms.

Key words: Dynamic routing of vehicle, Urban transportation management, Machine learning, Energy consumption, UAV routing.

ایمیل نویسنده مسئول: v.najafpoor1@gmail.com

۱-مقدمه

سیستم حمل و نقل یکی از مهم ترین حوزه ها در زنجیره تامین است و بهینه سازی آن می تواند موجب کاهش هزینه های عملیاتی و افزایش سودآوری گردد. در این مقاله، در مورد علل و لزوم استفاده از الگوریتم‌های یادگیری ماشین به‌عنوان روش *optimization* برای سیستم حمل و نقل خواهد شد و الگوریتم های یادگیری ماشین برای حل مسایل بزرگ با استفاده از یادگیری عمیق مطرح می گرددکه طول مسیر را برای مسئله *VRP* پویا ابتدا برای یک وسیله و در صورت امکان برای چند ماشین بهینه کنند و تعداد ماشین کمتری نیز مورد نیاز باشد و یا با تعداد کمتری پرواز، پهپاد بتواند به مکان های مد نظر خدمت رسانی کند و در مجموع انرژی مصرفی کاهش یابد. کارهای زیادی در بحث مدیریت حمل و نقل شهری انجام شده است. معماری های جدیدی نیز برای این کار در نظر گرفته شده است که در زیر به برخی از آنها اشاره می شود. حل مسئله مسیریابی ناوگان شهری به دو صورت استاتیک و داینامیک بررسی می شود. برای حل مسیریابی ناوگان از روش های ریاضی، ابتکاری و فراابتکاری استفاده شده است که در مجموع برای حل مسایل بزرگ کارایی لازم را ندارند و نمی توانند در لحظه پاسخگوی ایجاد راه حل برای اینگونه مسائل باشند. به همین دلیل به حل مسایل مسیریابی با استفاده از یادگیری ماشین پرداخته شده است که باید مدل سیستم حمل و نقل بخوبی شناخته شود و در ادامه راه کارهای ممکن جهت حل مسئله مسیریابی پیشنهاد شود.

۲-مرور کارهای گذشته

پهپادها به سرعت در حال تبدیل شدن به ابزاری قدرتمند برای طیف گسترده‌ای از کاربردها، از جمله تحویل کالا، جستجو و نجات و نقشه‌برداری هستند. با این حال، مسیریابی کارآمد و ایمن پهپادها در محیط‌های پیچیده و پویا، چالش بزرگی را به وجود می‌آورد. در این مقاله، به بررسی مسئله مسیریابی خودرو برای پهپادها می‌پردازیم و روش‌های مختلف حل این مسئله را تشریح می‌کنیم. سابقه حل مساله مسیریابی خودرو برای پهپادها به عنوان یک مساله پیچیده و جذاب در زمینه هوافضا و رباتیک مطرح است [۱]. این مساله از اهمیت بالایی برخوردار است زیرا پهپادها به عنوان یکی از تکنولوژی های حیاتی در حوزه نظارت، ارتباطات و حمل و نقل استفاده میشوند. سابقه حل مساله مسیریابی خودرو برای پهپادها شامل تلاش های بسیاری از سوی پژوهشگران و صنعتگران در زمینه طراحی الگوریتم های هوش مصنوعی، شبکه های عصبی، الگوریتم های بهینه سازی و سیستم های کنترل خودکار است. این تلاش ها به منظور بهبود کارایی، دقت و سرعت مسیریابی پهپادها برای انجام وظایف مختلف مانند نظارت، جستجو و نجات، تحویل کالا و غیره انجام شده است. با توجه به پیشرفت های اخیر در حوزه هوافضا و رباتیک، انتظار می رود که روش های پیشرفته تر و دقیقتر برای حل مساله مسیریابی خودرو برای پهپادها توسعه یافته و به کار گرفته شود [۲].

مسیریابی خودرو برای پهپادها یک حوزه مهم در تحقیقات و توسعه فناوری پهپادها است که تاکنون به طور گسترده مورد مطالعه قرار گرفته است. ابتدا، پهپادها به عنوان وسیله ای برای انجام ماموریت های نظامی و امنیتی مورد استفاده قرار گرفتند و نیاز به مسیریابی دقیق برای اجرای ماموریتهای مختلف داشتند. سابقه مسیریابی خودرو برای پهپادها در دنیا شامل تحقیقات و پروژه های مختلف است که از جمله آنها میتوان به موارد زیر اشاره کرد:

تحقیقات دانشگاهی: دانشگاه ها و مراکز تحقیقاتی در سراسر جهان به بررسی الگوریتم های مسیریابی پهپادها و بهبود عملکرد آنها می پردازند. شرکت های فعال در حوزه فناوری و پهپادها نیز به توسعه الگوریتم های مسیریابی پهپادها و بهبود نظام هدایت آنها اختصاص داده اند. نیروهای نظامی و سازمان های امنیتی در سراسر جهان از پهپادها برای اجرای ماموریت های خطرناک و پیچیده استفاده می کنند و نیاز به مسیریابی دقیق برای این پهپادها دارند. با پیشرفت فناوری و تحقیقات در این زمینه، الگوریتم های مسیریابی خودرو برای پهپادها بهبود یافته و قابلیت اجرای ماموریت های مختلف با دقت و کارآمدی بالاتر را دارا شده اند. یکی از کاربردهایی که می توان برای پهپادها متذکر شد، استفاده از آنها در حوزه های بشردوستانه است که میتواند زنجیره تامین دارو، امدادرسانی و موارد مشابه ی را در برداشته باشد. در این پژوهش مسئله مسیریابی پهپاد و وسیله نقلیه به منظور امدادرسانی و یا جمع آوری زباله در محیط، با هدف کمینه کردن مجموع زمان خدمت رسانی به گره های تقاضا انجام گرفته است و به محدودیت های عملیاتی پهپادها نظیر ظرفیت و مصرف انرژی توجه شده است. پهپادها مسافت کمتری را طی می کنند و می توانند خدمت رسانی را در زمان کوتاه تر و با سرعت بیشتری انجام دهند، اگرچه محدودیت هایی نیز دارند.

مدیریت حمل و نقل شهری نیز به دلیل رشد سریع مقدار زباله، هزینه‌های بالای جمع‌آوری زباله[۳]، ظرفیت‌های محدود تصفیه [۴] و مشکلات زیست‌محیطی [۵] به‌عنوان یک مسئله چالش برانگیز برای شهرهای معاصر در نظر گرفته می‌شود [۷-۶]. در چنین شرایط پیچیده‌ای، استفاده از روش‌های برنامه‌نویسی و تحقیق عملیاتی می‌تواند تصمیم‌گیرندگان را بهره‌مند کند [۸]. فعالیت‌های مدیریت حمل و نقل شهری در پنج مرحله از چرخه حیات زباله گروه‌بندی می‌شوند: تولید، جمع‌آوری و حمل‌ونقل، تحول، تصفیه و دفع نهایی [۹]، هزینه جمع‌آوری و حمل‌ونقل زباله ۶۰-۸۰٪ از کل هزینه سیستم مدیریت حمل و نقل را تشکیل می‌دهد، که عامل مهمی در هزینه‌های مالی مدیریت حمل و نقل است، پیشرفت در این زمینه نقش بسزایی در صرفه‌جویی در هزینه‌های شهرداری خواهد داشت[۱۱-۱۰]. به‌عنوان یک قاعده، فعالیت جمع‌آوری و حمل زباله با استفاده از ناوگان وسایل نقلیه با هدف تخلیه سطل‌های زباله بر اساس برنامه‌های از پیش تعیین شده انجام می‌شود [۱۵-۱۲] بااین‌حال، جمع‌آوری زباله معمولی بر اساس بسیاری از گمانه زنی‌ها در مورد سطح پرکردن سطل‌های زباله، پرکردن جزئی تا تخلیه کامل متفاوت است و منجر به مصرف

منابع غیرضروری می‌شود [۱۶]. در [۱۷] مدیریت زباله‌های هوشمند در شهرهای هوشمند بررسی شده است. سطح زباله موجود در سطلهای زباله با کمک سیستم سنسورهای اولتراسونیک شناسایی می‌شود و از طریق سیستم GSM به اتاق کنترل مجاز منتقل می‌شود. این امر باعث کاهش تعداد سفرهای حمل و نقل زباله شهری می‌شود و از این رو کل هزینه‌های مرتبط با حمل و نقل زباله را کاهش می‌دهد. در نهایت به حفظ پاکیزگی در جامعه کمک می‌کند. بنابراین، سیستم مدیریت زباله هوشمند باعث حمل و نقل بهینه تر زباله‌های شهری می‌شود. نویسندگان در [۱۸] از برنامه‌ریزی مسیریابی دینامیکی بهبودیافته استفاده می‌کنند. آنها یک هدایتگر متغیر منطقه‌ای را که با مسئله حمل و نقل زباله سازگار است، تقویت می‌کنند. در [۱۹] برای دستیابی به مسیریابی دینامیکی، مدل مسیریابی ارائه می‌شود که این مدل از الگوریتم کلونی مورچگان(ACS) استفاده می‌کند. آنها محل قرارگیری سطل‌ها را به‌عنوان یک شبکه فاصله‌دار، بهینه‌سازی می‌کنند و برای اینکه بتوانند توزیع سطل‌ها را در مجموعه‌ای از گروه‌ها جمع کنند، از سایر موارد استفاده می‌کنند.

در [۲۰] نویسندگان ترکیب مسیریابی و بهینه‌سازی برنامه‌ریزی را انجام می‌دهند. داده‌های پیشینه کاربردی برای سطل‌های منحصر، مدارهای روزانه نقاط جمع‌آوری شده را بازدید می‌کند. برای مدیریت بهتر سیستم از برنامه‌ریزی استفاده می‌شود. نویسندگان در[۲۱] برنامه‌ریزی مسیریابی پهپادها در نظر گرفتن محدودیت زمانی را بررسی کردند. هدف اصلی این روش به‌حداقل‌رساندن کل هزینه‌های عملیاتی و هزینه‌های ثابت کامیون است. در [۲۲] نویسندگان با فرض مطالبات مشتریان به‌عنوان متغیرهای فازی، مدل مسیریابی دینامیکی را بر اساس خواسته‌های فازی معرفی می‌کنند. مدل رویکرد تجربی مبتنی بر تئوری اعتبار فازی می‌باشد. یک روش فرمول‌بندی ریاضی در [۲۳] برای تدوین برنامه‌ای از مناطق خدمات، تعریف مسیریابی و طراحی برنامه‌ریزی با درنظرگرفتن راه‌حل‌های جایگزین جدید در مدیریت سیستم به‌عنوان یک کل ارائه شده است.

در [۱۱] نویسندگان مسیریابی را با روزنه‌های زمانی ارائه می‌دهند که فعالیت منطقی را در یک شهر تحلیل می‌کنند. مدل مسیرهای بهینه برای یافتن کامیون‌ها برای خالی‌کردن سطل با یک الگوریتم جستجوی منطقه‌ای بزرگ تطبیق پیدا می‌کند. در [۲۴] نویسندگان روش‌های برنامه‌ریزی دینامیکی را که برای حمل و نقل زباله سطل‌های مسیرهای زیرزمینی به کار می‌رود، ارزیابی می‌کنند. با استفاده از مسیریابی دینامیکی، مقادیر دی‌اکسیدکربن موجود در محیط از کامیون‌ها کاهش می‌یابد. در [۲۵] نویسندگان شبیه‌سازی رویداد گسسته برای حمل و نقل زباله از سطل برای مسیر زیرزمینی ارائه می‌دهند. مدل از برنامه‌ریزی دینامیکی برای بهره‌برداری از اطلاعات منتقل شده از طریق سنسورهای حرکتی تعبیه شده در سطل‌های زیرزمینی استفاده می‌کند. نویسندگان در [۲۶] مسیریابی را با یک سیستم اندازه‌گیری سیار روی کامیون‌ها توسعه می‌دهند. آنها مسیریابی دینامیکی احتمالی را انجام می‌دهند که اصلاحات را در طول یا بعد از اجرای مسیرهای موجود ایجاد می‌کند. در شکل ۱، بررسی اجمالی مدل سیستم ارائه شده است.

شکل ۱- بررسی اجمالی مدل سیستم حمل و نقل پسماند شهری

نویسندگان در [۲۷] الگوریتم ممتازی را برای انجام مسیریابی اجباری با فرصت زمان و زمینه تضاد معرفی می‌کنند. مدل ترکیبی از جریان را تشکیل داده و فرمولاسیون جزء بندی را برای دستیابی به بهینه‌سازی چندهدفه تنظیم کرده است. راه‌حل تجربی دیگری در [۲۸] ارائه شده است که در آن نویسندگان جمع‌آوری زباله‌ها را به‌عنوان یک مشکل مسیریابی دوره‌ای کامیون با انبارهای زباله میانی بیان می‌کنند. این مدل شامل جستجوی متغیر منطقه‌ای و برنامه‌نویسی دینامیکی به‌منظور دستیابی به راه‌حل بهینه است. در [۲۹] نویسندگان الگوریتم ژنتیکی چند هدفه را برای حل مسئله مسیریابی دینامیکی پیشنهاد می‌دهند. به طور خاص، مدل فرض می‌کند که مسئله جمع‌آوری زباله می‌تواند به‌عنوان یک مسئله فروشنده سیار (TSP) بهینه شود. سپس الگوریتم ژنتیک TSP را بهینه می‌کند. نویسندگان در[۳۰] یک روش ابتکاری را برای مسیریابی دینامیکی باتوجه‌به پارامترهای مختلف قابل تنظیم پیشنهاد می‌کنند. سنسورها مسیریابی کنترل موجودی معکوس را در شبکه زباله متراکم فعال می‌کنند. این نظریه ابتکاری با عدم قطعیت اثرات روزانه و فصلی سروکار دارد.

مدل ارائه شده در [۳۱] به حمل و نقل زباله‌های پلاستیکی که از سایر پسماندهای جامد شهری متمایز شده است، تخصص دارد. طراحی مجدد ابتکاری از مسیرهای مجموعه با استفاده از یک معیار سازگار با محیط‌زیست و متعادل کردن تبادل بین هزینه‌ها و مسائل زیست‌محیطی حاصل می‌شود. سرانجام، در [۳۲] نویسندگان یک مدل مسیریابی دینامیکی را برای حمل و نقل زباله در یک شهر هوشمند پیشنهاد می‌دهند. به طور خاص، فقط در [۳۳] مسئله حمل و نقل زباله به‌عنوان مشکلی مطرح می‌شود که می‌توان با گنجانیده شدن زیرساخت IoT در شهرهای هوشمند آن را حل کرد.

در این تحقیق تحقیقات ارائه شده در [۱۱] را با معرفی یک مدل سیستم برای جمع‌آوری زباله در یک شهر هوشمند گسترش می‌دهیم. این مدل دارای یک مدل مسیریابی دینامیکی انعطاف‌پذیر است که مدل مسیریابی دینامیکی را اصلاح می‌کند که در [۳۴] گسترش‌یافته است.

تحقیق در مورد روش اکتشافی و روش فرا ابتکاری به طور فزاینده‌ای رو به گسترش است [۳۵]. همراه با بهینه‌سازی ازدحام ذرات گسسته (*PSO*)، راو و همکاران [۳۶] یک روش ابتکاری برای بهبود کیفیت حل ذرات *PSO* برای حل یک مسئله چندهدفه ایجاد کرد. در مقاله [۳۷] یک مسئله مسیریابی وسیله نقلیه را با پنجره زمان (*VRPTW*) جمع‌آوری زباله توسط یک الگوریتم *simulated annealing* حل کرد. ترکولایی و همکاران [۳۸] الگوریتم کلونی مورچگان بهبودیافته برای مدل پیشنهادی مسئله مسیریابی ظرفیت دار (*CVRP*) ارائه کردند. مقاله [۳۹] الگوریتم ترکیبی ژنتیک را برای حل *VRP* برای حمل‌ونقل زباله‌های عفونی طراحی کردند. در[۴۰] یک الگوریتم حریصانه یکپارچه همراه با جستجوی محلی متغیر برای یک مسئله مسیریابی چندهدفه برای جمع‌آوری و حمل‌ونقل زباله پیشنهاد کرد. به طور خلاصه، تحقیقات زیادی در مورد الگوریتم‌ها برای حل مسئله جمع‌آوری زباله انجام شده است. با این وجود، آنها بیشتر الگوریتم‌های منفردی هستند به جای الگوریتم‌های ترکیبی که می‌توانند از هر دو سو بهترین بهره را ببرند. روش حل دقیق برای حل *VRP* متوسط و بزرگ در زندگی واقعی ناکارآمد است [۴۱]. روش‌های بهینه‌سازی زیادی در *SWC* استفاده شده است. بااین‌حال، آنها از نظر همگرایی کند، نظریه پیچیده، قابلیت جستجوی ضعیف و تنظیم پارامترهای خسته‌کننده دارای اشکالاتی هستند. بنابراین، برای دستیابی به نتایج بهتر از نظر هزینه، مسافت سفر و میزان انتشار، توسعه روش‌های بهینه‌سازی کارآمد و قوی در *VRP* موردنیاز است. بخصوص در مورد مسائل واقعی شهرهای بزرگ، با تعداد درخواست بالا، راه‌حل‌های فراابتکاری فوق سرعت مناسبی ندارند.

شرایط کار برداشت به‌صورت زیر ارائه شده است:

۱. زمان در دسترس محدود است (زمان سفر برای هر وسیله نقلیه و پهپاد نباید از زمان کل دردسترس تجاوز کند).

۲. تعداد مخازن هر پست – ناحیه – منطقه (تعداد گره در شبکه راه) موجود و ظرفیت هر یک مشخص است.

۳. میانگین زمان سرویس برای هر گره (مشتری) مشخص است.

۴. ظرفیت هر نوع از وسایل نقلیه و یا پهپاد مشخص و ثابت است و هزینه برای هر نوع مقداری ثابت است.

۵. مسیرهای مجاز (زیر مجموعه‌ای از مسیرهای شبکه راه‌ها) موجود است.

۶. تعداد ماشین‌های در دسترس مشخص است.

۷. فواصل گره‌ها و محل تخلیه زباله (ایستگاه) مشخص است.

۸. هر گره فقط توسط یک ماشین سرویس‌دهی می‌شود.

۹. هر ماشین شروع و پایان مسیرش از یک ایستگاه است.

۱۰. مجموع تقاضا برای هر ماشین نباید از ظرفیتش تجاوز کند.

۱۱. سرعت برای وسایل نقلیه مقداری ثابت است.

۱۲. هزینه هر واحد زمانی کارکرد (اجازه) از هر نوع ماشین مشخص است.

۱۳. میانگین زمان جهت تخلیه در ایستگاه و حداکثر کردن ظرفیت خالی جهت سرویس مجدد، مشخص است.

۱۴. در طول زمان در دسترس، کلیه مخازن باید تخلیه شوند.

مسیریابی چند هدفه برای پهپادها یک حوزه پیچیده و چالش برانگیز است که به دلیل محدودیت های مختلف مانند محدودیت های انرژی، محدودیت های محیطی و پویایی محیط، نیاز به روش ها و الگوریتم های پیشرفته برای مسیریابی دقیق و کارآمد دارد. برخی از چالش ها، روش ها و جهت های اینده در مسیریابی چند هدفه برای پهپادها عبارتند از:

- محدودیت های انرژی: پهپادها معمولاً دارای محدودیت های انرژی هستند که نیاز به الگوریتم های بهینه برای مسیریابی دارند تا از انرژی بهینه استفاده کنند.
- پویایی محیط: محیط پهپادها پویا و تغییرات زود روز دارد که نیاز به الگوریتم های تطبیقی برای مسیریابی دارد.
- تعامل با سایر پهپادها و اجسام: در صورت وجود چندین پهپاد یا اجسام دیگر در محیط، نیاز به هماهنگی و همکاری بین آنها برای مسیریابی دقیق و امن است.
- الگوریتم های هوش مصنوعی: استفاده از الگوریتم های هوش مصنوعی مانند یادگیری تقویتی، شبکه های عصبی و الگوریتم های تکاملی برای بهبود مسیرابی پهپادها.
- الگوریتم های تطبیقی: استفاده از الگوریتم های تطبیقی برای تطبیق با تغییرات محیط و بهبود عملکرد مسیریابی در شرایط پویا.

برای استفاده از یادگیری تقویتی عمیق (*DRL*) در حل *VRP* با پهپاد، ابتدا باید این مسئله را به عنوان یک *MDP* فرمول‌بندی کنیم. در *MDP*، یک عامل هوشمند (پهپاد) در یک محیط پویا عمل می‌کند و در هر گام زمانی، بر اساس اطلاعات گذشته خود، یک عمل را انتخاب می‌کند. انتخاب هر عمل منجر به یک حالت جدید در محیط و دریافت پاداش برای عامل می‌شود. انجام تحقیق با ایجاد شبکه یادگیری تقویتی عمیق با استفاده از *RNN* ها (دو زیر شبکه رمزنگار (*encoder*) و رمزگشا *(decoder)*)با توجه به محدودیت های استخراج شده مسئله حمل و نقل و با ایجاد دیتاست مناسب جهت براورد و تحلیل طول مسیر خروجی مسئله انجام گرفت. در مرحله اول این تحقیق، نمونه گیری و دیتای تست موقعیت مکانی سطل های زباله و مقدار درخواست آنها و نیز بازه زمانی سرویس به هر درخواست بصورت مصنوعی تولید می شود. در ادامه با استفاده از یک الگوریتم شیب سیاست برای بهینه‌سازی پارامترهای آن، مدل و شبکه آموزش‌دیده، راه‌حل را به‌عنوان دنباله‌ای از نقاط توقف مسیر در زمان واقعی تولید می‌کند. بهینه سازی پارامترها و ابرپارامترهای مسئله جهت ارایه خروجی مسیر مناسب جهت مسایل بزرگ در نظر گرفته شد. در انتها مقایسه رویکرد استفاده شده با یادگیری تقویتی با الگوریتم‌های اکتشافی کلاسیک و *OR-Tools* گوگل در مقایسه کیفیت راه‌حل با زمان محاسبات (پس از آموزش)ملاک خوبی برای

ارزیابی بهبود عملکرد با استفاده از یادگیری ماشین به نمایش در آمد.

۳-روش و ابزار گردآوری اطلاعات

تولید داده های آموزشی و تست با استفاده از برنامه‌نویسی پایتون تولید خواهد شد. مسئله مسیریابی برای تعداد نود ۱۰،۲۰،۵۰ و در صورت امکان صد نود حل شد و مقایسه طول مسیر بهینه با خروجی کار های قبلی انجام می گردد.تعداد نمونه جهت یادگیری مساله ، یک میلیون نمونه آموزشی و جهت تست ۱۰۰۰ نمونه می باشد. دیتا بصورت تصادفی تولید می شود. در صورت وجود دیتاست واقعی، جهت مقایسه خروجی، با آن هم امکان تست فراهم می شود. بخشی از موقعیت جغرافیایی و مکانی نیز می تواند از دیتای شهری استفاده شود. جامعه آماری نمونه، مخازن زباله منطقه ۱۱ مدیریت پسماند شهر تهران در نظر گرفته می شود.نمونه گیری بصورت تعداد ۱۲۸ و ۲۵۶ نمونه در هر مرحله یادگیری بود. سیستم جمع‌آوری پسماند خشک (*SWC*) از تعدادی سطل زباله مجهز به سامانه اینترنت اشیا (*smart waste bins*)، ماشین حمل زباله و اطلاعات ترافیک تشکیل شده است. حالت سیستم به عنوان ورودی به شبکه وارد شده و این دیتای حجیم جهت آموزش و ارزیابی نهایی مدل بکار می‌رود. این معماری از ترکیب دو بخش تشکیل شده است. ابتدا دنباله ورودی درخواست‌ها رمزگذاری شده و به برداری از اعداد با طول ثابت تبدیل می‌شود که به آن، بردار زمینه گفته می‌شود. سپس برای رمزگشایی، بردار به دنبالهٔ خروجی یا همان مسیر بهینه تبدیل می‌شود. در ادامه یک الگوریتم یادگیری تقویتی برای مدیریت هوشمند حمل و نقل طراحی می‌کنیم که پارامترهای مدل شبکه عصبی را بدون اطلاع از راه‌حل بهینه برای داده‌های آموزشی با حداقل زمان محاسباتی تعیین کند. مدل رمزنگار – رمزگشا *LSTM* یک شبکه عصبی بازگشتی است که یک دنباله ورودی را به دنباله دیگر نگاشت می‌کند. (یک دنباله که حالت سیستم پسماند شامل درخواست‌ها و موقعیت مخزن و موقعیت فعلی خودرو یا پهپاد است به‌عنوان ورودی دریافت می‌کند و خروجی، دنباله دیگری از عمل‌ها است که نشان دهنده توالی انتخاب نود بعدی برای سرویس دهی است.) به همین دلیل به آن مدل *seq2seq* گفته می‌شود. در این طرح تحقیقاتی نیز به‌واسطه تحلیل دنباله مسیر خودرو یا پهپاد از همین ساختار پیشنهادی استفاده شد. در این پیشنهاد یک متا الگوریتم حریصانه اتخاذ شده، به‌موجب آن یک راه‌حل عملی با افزودن متوالی گره‌ها ساخته می‌شود و برای برآورده ساختن محدودیت‌های مسئله حفظ می‌شود. این معماری یادگیری عمیق جدید ، ویژگی‌های گره‌های شبکه را استخراج می‌کند که به سیاست اجازه می‌دهد تا سودمندی هر گره را از هم تفکیک کند. برای طراحی شبکه عصبی پیشنهادی در شبیه ساز از شبکه پوینتر استفاده می‌کنیم. این شبکه از دو زیر شبکه رمزنگار و رمزگشا تشکیل شده است. شبکه رمزنگار در ابتدا اطلاعات، وضعیت سیستم و مسیر توسعه داده شده را به‌عنوان ورودی می‌گیرد تا سیستم را به ویژگی‌های تعبیه شده برای هر گره مطابق ساختار تبدیل کند.

در این کار ما سیستم مدیریت حمل و نقل را مدل می‌کنیم. شبکه حمل‌ونقل در مسئله WLRP[1] به‌صورت یک گراف $\Phi(v,\varepsilon)$ تعریف می‌شود. به‌طوری‌که هر $i \in v$ در گراف یک نقطه مطلوب را نشان می‌دهد که نشاندهنده مکان خودرو، انبار و مکان دریافت زباله[2] است. همچنین هر جفت $i,j \in \varepsilon$ مختصات جغرافیایی جاده‌هایی است که نقاط $i \in v$ را به هم وصل می‌کند یک مسیر واحد بین دو مکان با هزینه مرتبط[3] که طول آن D_{ij} و مدت‌زمان طی شده تخمینی جاده توسط خودرو یا پهپاد با T_{ij} نشان داده می‌شود[4].

$$V = \left\{ v_t^i = (s^i, d_t^i), i = 0,1,...,n \right\} \quad (۱)$$

که v^0 نود اولیه یا همان مخزن و بقیه موقعیت جغرافیایی پراکنده مشتریان یا همان درخواست‌های حمل و نقل [5] شهری است و n تعداد سطل‌های زباله در سطح شهر است[6]. از نماد $Q \in \bigcup_{k \in K} Q_k$ برای نشان دادن کلیه درخواست‌ها[7] موجود در سیستم که تاکنون بارگیری نشده‌اند استفاده می‌شود. برای هر درخواست $q \in Q$ مختصات دو بعدی مکان درخواست بارگیری زباله با $p_q = (x_q, y_q)$ نشان داده می‌شود. ظرفیت موردنیاز هر درخواست پرشدن زباله برای بارگیری در هر خودرو برابر d_q است و قبل از زمان نهایی T_q باید به انبار جمع‌آوری زباله برسد. اگر بازه‌های زمانی بصورت سخت تنظیم شوند، وسایل نقلیه اجازه ندارند به مشتریان خارج از روزنه‌های زمانی[8] (TW)خدمت دهند. وقتی که زود هنگام رسیدند، $arr.t_i < t_{ei}$ وسیله نقلیه باید تا زمان شروع خدمت صبر کند. در عین حال، این امکان وجود ندارد که دیرتر از زمان انتهای سرویس، یعنی زمانی که در حال حرکت هستیم، به مقصد برسیم t_{li}. هر ورودی v_i به‌صورت دوتایی ترتیبی که شامل پنج ویژگی استاتیک (۲ ویژگی موقعیت مکانی[9] (x_i, y_i)، زمان شروع سرویس t_{ei}، زمان پایان سرویس t_{li}، زمان خدمت‌دهی t_{si})و چهار ویژگی دینامیک (مقدار درخواست d_{it}، ظرفیت وسیله نقلیه q_{it}، زمان ورود $arr.t_i$، زمان خروج $dep.t_i$) درخواست‌ها است، وارد شبکه انکودر که با استفاده از شبکه عصبی بازگشتی (RNN) ساخته شده است، می‌شود که آن را به یک فضای

[1] waste location routing problem
[2] Coordinate of trash bin
[3] Distance weights of the road segments
[4] Road segment between bins
[5] Waste generation nodes
[6] Number of trash bin in town
[7] Waste-holding capacity of each container
[8] Time windows
[9] coordinate

برداری ۱۲۸ بعدی نگاشت می‌کند. برای مخزن $d_0 = 0$ و $t_0 = [a_0, b_0]$ که a_0 و b_0 ، به ترتیب نشان دهنده زودترین خروج ممکن از و آخرین بازگشت به انبار است. خودرو یا پهپاد از انبار خارج می شود، با توجه به محدودیت های ظرفیت و زمان پنجره به مشتریان یا همان سطل زباله هوشمند خدمات ارائه می‌دهد و در نهایت به انبار باز می گردد. دنباله ای از مشتریانی که وسیله نقلیه سرویس داده، یک مسیر حمل و نقل [۱] π است. تابع هدف[۲] ما تلاش می‌کند تا تعداد درخواست‌های لجستیکی وارد شده را ماکزیمم کند و در عین حال میزان تخطی از زمان و نیز طول مسیر طی شده برای اجابت درخواست‌ها کمینه باشد. معادلات ریاضی مربوطه در قبل مورد بحث قرار گرفت. نقاط برداشت[۳] یا همان نقاط توقف به گونه ای انتخاب خواهند شد تا طول مسیر پیمایشی[۴] وسیله نقلیه کمترین و نزدیک به بهینه باشد.

مقدار درخواست[۵] هر مشتری d_{it} یک عدد صحیح است و درخواست مخزن ۰ است. ماکزیمم ظرفیت وسیله[۶] C در هر مسئله با توجه به سایز n متغیر است. این نگاشت اطلاعات ورودی و ویژگی‌های حالت سیستم، از فضای ۹ بعدی به فضای ۱۲۸ بعدی در جزئ نهفته سازی[۷] انجام می‌شود.

$$v_i = \{(x_i, y_i, t_{ei}, t_{li}, t_{si}), (d_{it}, q_{it}, arr\,time_{it}, dep\,time_{it})\} \in R^9 \quad to \quad R^{128} \tag{۲}$$

صرفاً در نظر گرفتن ویژگی‌های یک گره منفرد (موقعیت، مقدار تقاضا و پنجره زمانی) هیچ اطلاعاتی در مورد دنباله مشتری بهینه به دست نمی‌دهد. بنابراین، این ساختار زیربنایی نمی تواند به طور مستقیم توسط شبکه عصبی مورد استفاده قرار گیرد. درعوض، ویژگی‌های گره باید ابتدا در فضایی متفاوت در آنچه ما مرحله تعبیه ورودی می‌نامیم، مجدداً پیش‌بینی شوند. همانند ما یک تعبیه ورودی ساده را پیشنهاد می‌کنیم که در آن هر مجموعه از مشتریان با استفاده از یک لایه کانولوشنی یک بعدی تعبیه شده است. بردار زمینه از یک شبکه کانولوشن گراف[۸] (GCN)تولید می‌شود. با توجه به تغییر ناپذیری تعبیه در اعمال انتخابی، فقط عناصر استاتیک تعبیه شده‌اند. این بخش به درک تأثیر تعبیه ورودی بر همگرایی و عملکرد مدل RL اختصاص یافته است که مقادیر وزن‌های این تبدیل به صورت زیر است:

$$h_i^{(0)} = \begin{cases} w[x_i, y_i, t_{ei}, t_{li}, t_{si}] + b & , i \neq 0 \\ w_0[x_0, y_0, t_{e0}, t_{l0}, t_{s0}] + b_0 & , i = 0 \end{cases} \tag{۳}$$

بردار ترکیبی از جاسازی گره های همه گره های آموخته شده در GCN به عنوان بردار زمینه عمل می کند. اضافه شدن پنجره های زمانی به مسئله VRP باعث افزایش وابستگی گره می شود. امکان سنجی یک مسیر را فقط می توان با در نظر گرفتن توالی گره به عنوان یک کل تعیین کرد و بر اساس π سیاست بهینه خروجی که خروجی توالی مسیر[۹] را نتیجه می‌دهد، احتمال مقادیر ترتیبی خروجی نقاط توقف بعدی[۱۰] به‌صورت زیر محاسبه می‌شود:

$$P(Y \mid S) = \prod_{t=0}^{T'} p(y_{t+1} \mid y(t), S) \tag{۴}$$

$$\Phi_{t+1} = f(y_{t+1}, \Phi_t) \tag{۵}$$

که این احتمال با تابع نشانگر محاسبه می‌گردد و در آن S حالت سیستم است و y_{t+1} نقطه توقف بعدی است. در معادله ۶، حالت بعدی سیستم از حالت فعلی و انتخاب درخواست بعدی با تابع f به صورت یک زنجیره مارکوف حاصل می‌شود. بر اساس توزیع احتمال خروجی از شبکه عملگر[۱۱]، در هر مرحله توالی مشتری‌ها می‌تواند رمزگشایی شود. می‌خواهیم میزان تاثیر هرکدام از گام‌های زمانی (Time step) ورودی بر گام‌های خروجی را مشخص کنیم. برای اینکار از تمامی قسمت‌های شبکه اول (شبکه‌ای که یک تجسم از چینش نودهای مبدا می‌ساخت) به یک شبکه LSTM یک طرفه به صورت وزن‌دار ورودی می‌دهیم. که وزن یاد شده در واقع همان Attention ما است.

لایه توجه: برای توصیف شباهت بین وضعیت فعلی و هر نقطه مشتری استفاده می‌شود. در مرحله t، رمزگشا[۱۲] برای محاسبه حالت پنهان با اطلاعات گراف (s^i, d_t^i) استفاده می‌شود، که در آن $\tilde{v} = (\tilde{s}^i, \tilde{d}_t^i)$ خروجی رمزگذار[۱۳] است. مکانیزم توجه به صورت زیر از روی وزن‌دهی به همسایگان نودها ایجاد می‌شود که حالت نهفته هر نود را از مجموع وزن دار همسایگان آن نود ایجاد می‌کند:

$$u_i^t = v_a^T \tanh(W_a[s_i^r; d_{i,t}^r; h^t]),$$
$$a_t = soft\max(u^t)$$
$$c_t = \sum_{i=0}^{n} a_t \tilde{v}_t \tag{۶}$$

که در آن $[۰;۰]$ عملگر الحاق بین بردارها است[۱۴]و مقادیر v و W وزن‌های قابل یادگیری هستند. در VRP، شبکه برای تعیین اینکه آیا مشتری بعدی محدودیت‌های تقاضا و زمان را برآورده می‌کند یا نه، از ماسک بهره می‌گیرد. نکته قابل توجه اینجاست که در لایه توجه،بر

[1] Waste collection route
[2] objective function
[3] Pickup locations
[4] Waste collection route
[5] Demand
[6] Waste-holding capacity of vehicle
[7] embedding component
[8] graph convolutional network
[9] Waste route
[10] Stop points
[11] Actor network
[12] decoder
[13] encoder
[14] softmax*(·)* is the softmax function also known as normalized exponential function

خلاف مرحله دیکدر، اطلاعات داینامیک حالت سیستم هم استفاده می‌شود. نودهایی که سرویس داده شده‌اند و نیز نودهایی که اندازه درخواست آنها از ظرفیت فعلی وسیله بیشتر است، ماسک صفر می‌گیرند و به شکل زیر نشان داده می‌شود:

$$Mask_i^t = \begin{cases} 1\ , & if\ node\ is\ valid\ at\ time\ t \\ 0\ , & otherwise \end{cases} \quad (۷)$$

با انتخاب نود بعدی، مقادیر ویژگی‌های داینامیک نود انتخاب شده و دیگر نودها به صورت زیر تغییر می کند. مقدار درخواست نود بعدی انتخابی، صفر می‌شود و مقدار درخواست دیگر نودها تغییری نمی‌یابد.

$$d_i^{t+1} = \begin{cases} \max(0, d_i^t - c^t), & if\ \ k = i \\ d_k^t \quad , for\ \ k \neq i \end{cases} \quad (۸)$$

مقدار درخواست نود انتخابی صفر می‌شود و بار فعلی ناوگان هم مقدار درخواست نود از حجم آن طبق فرمول $q_i^{t+1} = \max(0, q_i^t - d_i^t)$ کم می‌شود. زمان ورود $(arr.time_{it})$ نیز با توجه به سرعت ماشین که یک در نظر گرفته شده است، از جمع زمان خروج از نود قبلی و فاصله دو نود متوالی انتخاب شده حاصل می‌شود. زمان خروج $(dep.time_{it})$ از نود نیز مجموع زمان ورود و زمان سرویس برداشت زباله می شود.

طول یک تور[1] را با جایگزینی π به‌صورت زیر تعریف می‌کند:

$$L(\pi \mid s) = \| x_{\pi(n)} - x_{\pi(1)} \|_2 + \sum_{i=1}^{n-1} \| x_{\pi(i)} - x_{\pi(i+1)} \|_2 \quad (۸)$$

که در آن $\|\cdot\|2$ L2 میانگین و $|\pi|$ طول دنباله است. هدف ما کاهش طول مسیر معادله ۹ با استفاده از یک مدل DRL است که بتواند سمت راست معادله ۵ را محاسبه نماید. معماری شبکه عصبی با استفاده از قانون زنجیره‌ای احتمال تور را به‌صورت زیر فاکتور می‌گیرد. مدل DRL_WLRP پیشنهادی، که به وسیله θ پارامتری شده‌است، نیاز به تعریف استراتژی $p(\pi|G)$ و انتخاب یک توالی π است. این یک سیاست تصادفی است و می‌تواند به صورت زیر فاکتور شود:

$$p_\theta(\pi \mid s) = \prod_{i=1}^{n} p_\theta(\pi(t) \mid \pi(<i), s) \quad (۹)$$

شبکه اشاره‌گر شامل دو ماژول شبکه عصبی بازگشتی (RNN)، رمزنگار و رمزگشا است که هر دو از سلول‌های حافظه کوتاه‌مدت (LSTM) تشکیل شده‌اند. در اینجا مساله را به صورت نرم در نظر گرفته و تخطی از زمان شروع۲ ($\delta_{e_t} = \max(t_{ei} - arr.t_i, 0)$) و تخطی برداشت[3] انتهای سرویس دهی۴ ($\delta_{l_t} = \max(arr.t_i - t_{li}, 0)$)را بعنوان المان پنالتی در پاداش نهایی برگشتی از مسیر بصورت وزن‌دار در نظر می‌گیریم. ما سازگاری پرس‌وجو را با همه گره‌ها محاسبه می‌کنیم و گره‌هایی را که نمی‌توان در زمان t بازدید کرد، mask می‌کنیم. برای VRP، این بدان معنی است که ما درخواست‌هایی را که قبلاً بازدید کرده‌ایم و همچنین گره‌هایی مقدار درخواست آنها بیشتر از ظرفیت فعلی وسیله نقلیه است، ماسک می‌کنیم. در هر قدم از انتخاب نود بعدی مسیر ماشین، بایستی مقادیر داینامیک همچون ظرفیت وسیله نقلیه و مقادیر درخواست‌ها و زمان ورود و خروج از نود به روز رسانی شود. عامل ما در مسئله، وسیله نقلیه است که در هر لحظه t حالت را دریافت می کند و بر طبق سیاست تعریف شده $\pi(t)$ را بعنوان عمل[5] بعدی انتخاب می‌نماید.

به روز رسانی پویا: پس از انتخاب مشتری بعدی برای خدمات‌رسانی، حالت سیستم براساس حرکت وسیله نقلیه تغییر می‌کند. وضعیت استاتیکی بدون تغییر باقی می‌ماند، در حالی که وضعیت دینامیکی به هنگام حرکت وسیله نقلیه به جلو به روز رسانی می‌شود. هنگامی که وسیله نقلیه تصمیم می‌گیرد که به مشتری j خدمت کند، متغیر موقعیت برابر با شاخص مشتری انتخابی (idx) است. در همین حال، تقاضای مشتری انتخاب‌شده صفر می‌شود و بار وسیله نقلیه[6] با کم کردن مقدار تقاضا به روز می‌شود $C = C - d_i$. علاوه بر این، زمان پایان این سرویس $dep.t_i = arr.t_i + t_{si}$ است که در آن ، زمان $arr.t_i$ ورود است. تابع پاداش[7]، هدف یک مسئله یادگیری تقویتی را تعریف می کند.

$$J(\theta \mid s) = \alpha * E_\pi[L(\pi \mid s)] + \beta \sum \delta_{ei} + \gamma \sum \delta_{li} \quad (۱۰)$$

برای بهینه‌سازی پارامترهای شبکه نشانگر نشان‌داده‌شده، از تقویت‌کننده سیاست مبتنی بر مدل استفاده شود. هدف آموزشی شبکه، ترکیبی است که شامل طول تور مورد انتظار و نیز مقادیر وزن دار از مجموع زمان تخطی از زمان شروع سرویس و مجموع زمان تخطی از زمان پایان سرویس دهی و برداشت زباله است. در حل مسئله برای طول مسیر و دو تخطی پنجره زمانی با وزن $c1 = 0.30$ و $c2 = 0.70$ در محاسبات پاداش تجمعی فرمول شماره ۱۱ لحاظ گردید که مجموع احتمالات سه وزن برابر یک باشد. باتوجه‌به گراف ورودی، به‌صورت زیر تعریف می‌شود:

$$\text{Re}ward = c1 * \sum_{i,j \in s \cup 0} c_{ij} + (c2/2)(\sum_{i \in \pi} \delta_{ei} + \sum_{i \in \pi} \delta_{li}) \quad (۱۱)$$

[1] bin collection list
[2] Ready_Violation
[3] Pickup delay
[4] Deadlin_Violation
[5] action
[6] Collected waste within the truck
[7] Reward or collection cost

جهت تخمین مقدار طول مسیر در شبکه کمک کننده[1]، مناسب‌ترین مدل یادگیری ماشینی که توانایی تقریبی حل مسئله را دارد، به‌عنوان یک شبکه عصبی با یک‌لایه پنهان و یک تابع فعال‌سازی ReLu انتخاب می‌شود. سپس، ویژگی‌های ورودی که باعث بهبود دقت پیش‌بینی می‌شوند، جستجو می‌شوند. برای بهینه‌سازی پارامترها به روش‌های شیب خط‌مشی و نزولی شیب تصادفی متوسل می‌شویم (REINFORCE algorithm). خروجی شبکه کمک کننده b(S) یک عدد اسکالر است که با دادن S بعنوان ورودی شبکه ، تخمین زده می‌شود. این خروجی شبیه به مقادیر تابع هدف بهینه در یک الگوی یادگیری نظارت شده عمل می کند. با توجه به ماهیت محاسباتی کارآمد انتشار رو به جلو یک شبکه عصبی (به ویژه در CRITNET) چنین روشی زمان محاسبه را تا حد زیادی کاهش می‌دهد.

$$\nabla_{\theta} J(\theta \mid s) = E_{\pi}[(R(\pi \mid s) - b(s))\nabla_{\theta} \log p_{\theta}(\pi \mid s)] \qquad (۱۲)$$

در عمل، مقدار مورد انتظار با میانگین نمونه‌گیری مونت کارلو در یک اندازه دسته B جایگزین می‌شود.

$$\nabla_{\theta} J(\theta \mid s) \approx \frac{1}{B} \sum_{i=1}^{B} [(R(\pi_i \mid s) - b(s))\nabla_{\theta} \log p_{\theta}(\pi_i \mid s)] \qquad (۱۳)$$

میانگین مربعات خطای هدف شبکه کمک کننده توسط ۱۵ محاسبه می شود که به‌صورت شبه کد الگوریتم ۱، آموزش وزن‌های شبکه مدل DRL_VRP انجام می‌شود. در الگوریتم، پارامترهای مدل به طور مکرر به روز می شوند. الگوریتم زمانی خاتمه می یابد که پارامترها همگرا شوند، یا به حداکثر تعداد تکرار از پیش تعریف شده رسیده باشد.

$$\nabla_{\theta_{Bl}} J(\theta_{Bl} \mid s) \approx \frac{1}{B} \sum_{i=1}^{B} \| R(\pi_i \mid s) - b(s) \|_2^2 \qquad (۱۴)$$

مقاله یک مدل نشانگر پویا را با معماری رمزگذار – رمزگشایی[2] ارائه می‌دهد که این مدل را قادر می‌سازد ویژگی‌های گره را به طور دینامیکی کشف کند و از اطلاعات ساختار پنهان به طور مؤثر در مراحل مختلف ساخت بهره‌برداری کند. در هر مرحله ساخت، باتوجه‌به بردار زمینه و تعبیه گره (به جز گره‌هایی که محدودیت‌ها را نقض می‌کند)، رمزگشا توزیع بر روی گره‌ها را پیش‌بینی می‌کند و یکی را برای بازدید و بعنوان نقطه توقف بعدی انتخاب می‌کند.

پارامترهای مسئله، بردار تمام متغیرهای آموزش پذیر مورد استفاده در چارچوب رمزگذار و رمزگشا است. هر نمونه[3] مسئله VRPبه عنوان الحاق مکان ها، ظرفیت تقاضا[4] و پنجره های زمانی برای هر گره تعریف می شود $S = [X, D, T]$. برای آموزش شبکه، از رویکردهای گرادیان خط مشی شناخته شده استفاده می‌کنیم. روش‌های گرادیان خط مشی به طور مکرر از یک گرادیان تخمینی از بازده مورد انتظار برای به روز رسانی پارامترهای خط مشی استفاده می‌کنند. این شبیه‌ساز ابتدا نمونه‌های آزمایشی VRPعظیمی را برای آموزش و اعتبارسنجی تولید می‌کند. علاوه بر تولید داده، شبیه‌ساز باید با وسیله نقلیه در حین حرکت تعامل داشته باشد. وقتی وسیله نقلیه[5] نقطه مشتری بعدی را انتخاب می‌کند، شبیه‌ساز دو جنبه اصلی برای به روز رسانی دارد. یکی به روز رسانی حالت دینامیک و دیگری پنهان کردن نقاط غیرقابل‌دسترس در زمان دیگری است. الگوریتم بعد از آموزش شبکه با یک بار اجرا و باتوجه به موقعیت شروع هر وسیله نقلیه که معمولا از انبار می‌باشد، کلیه نقاط بعدی مسیر وسیله را مشخص می‌کند. شرط پایان اجرا اینست که نقطه درخواست دیگری در سیستم نمانده باشد. ما شبکه را با رویکرد تصادفی[6] آموزش می دهیم و مدل را با استفاده از روش جستجوی حریصانه و پرتو[7] ارزیابی می‌کنیم. شبکه منتقد[8] پاداش را برای هر نمونه از یک حالت مشخص با استفاده از دو لایه متراکم تخمین می‌زند.

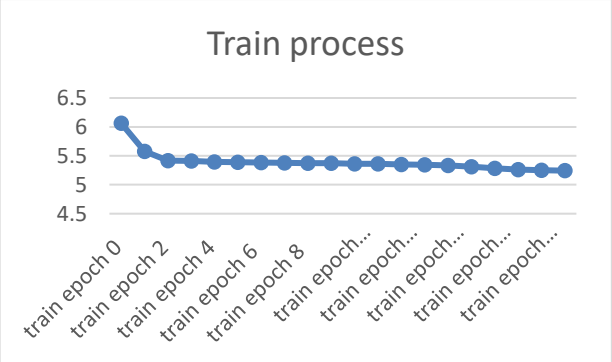
تولید داده‌های آموزشی و تست با استفاده از برنامه‌نویسی پایتون تولید شد. در مرحله اول این تحقیق، نمونه‌گیری و دیتای تست موقعیت مکانی سطل های زباله و مقدار درخواست آنها و نیز بازه زمانی سرویس به هر درخواست بصورت مصنوعی تولید می‌شود. مسئله مسیریابی برای تعداد نود ۱۰،۲۰،۵۰ و صد نود حل شد و مقایسه طول مسیر بهینه با تست ۱۰۰۰ نمونه خروجی کارهای قبلی انجام گردید. Cost ترکیبی نیز محاسبه و با سه الگوریتم مبنا مقایسه شد. تعداد نمونه جهت یادگیری مساله، یک میلیون نمونه آموزشی و جهت تست ۱۰۰۰ نمونه می باشد. دیتا بصورت تصادفی تولید می‌شود. نمونه گیری بصورت تعداد ۱۲۸ و ۲۵۶ نمونه در هر مرحله یادگیری خواهد بود.

دادگان: پارامترهای تولید دیتا به صورت زیر بودند: در یک مثال VRP n_TW_، شبیه‌ساز ابتدا n درخواست و یک انبار ایجاد می‌کند. هر درخواست حمل و نقل دارای ۹ ویژگی است که عمدتا به دو بخش عناصر ایستا و پویا[9] تقسیم می‌شوند. بخش استاتیک از متغیرهای ثابت در طول زمان شامل مختصات دو بعدی مکان درخواست (x_i, y_i) و *مخزن*، پنجره‌های زمانی $[t_{ei}, t_{li}]$ و زمان خدمات، t_{si} تشکیل شده‌است. (x_i, y_i) در بازه [0, 1] × [0, 1] انتخاب شدند. هر دو زمان سرویس t_{si} و طول پنجره زمانی Δt به طور تصادفی از[3, ۰] و [3,10] تولید می‌شوند. زمان سرویس‌دهی عدد ثابت ۰.۱ در نظر گرفته شد. کل بازه خدمت به صورت T =۳۰۰ در نظر گرفته می‌شود. و سرعت خودرو یا پهپاد به صورت v = ۱ در نظر گرفته شده‌است. مقدار عددی درخواست d_i به صورت تصادفی در بازه ۱ تا ۹ انتخاب شد. ظرفیت حمل ماشین پسماند[10] نیز همانند مقاله [22] برای VRP 10_، VRP _50، VRP _20 و VRP _100 مختلف در نظر گرفته شد (LOAD_DICT = { 10: 20, 20: 30, 50: 40, 100: 50}).

[1] critic
2 encoder-decoder architecture
[3] instance
[4] Waste-holding capacity of each container
[5] truck
[6] stochastic approach
[7] greedy and beam search
[8] critic network
[9] *load, demand, arrival, end_service*
[10] Waste-holding capacity of truck

۴-محک زدن

در این تحقیق، یک روش جدید بهینه‌سازی ترکیبی بر مبنای شبکه عصبی با یادگیری تقویتی پیشنهاد داده‌ایم به‌گونه‌ای که برنامه مسیریابی خودرو با حداقل زمان محاسباتی برای سرویس‌های حمل‌ونقل آنلاین در شبکه‌های بزرگ انجام می‌شود. برای حل این مسئله برای تعداد۱۰، ۲۰، ۵۰ و ۱۰۰ درخواست، ما یک شبکه پوینتر طراحی کرده‌ایم تا بتوانیم مسیرهای خودرویی را به‌صورت تکراری بسازیم. ازآنجایی‌که پارامترهای آموزشی آفلاین هستند تولید مسیر با کمترین زمان ممکن صورت می‌پذیرد و می‌توانیم با استفاده از آن مسیرهای خودرویی آنلاین در شبکه‌های بزرگ را بسازیم. بدون استفاده از روش پیشنهادی، ساخت یک مجموعه داده آموزشی با هدف بهینه‌سازی مسیر خودرویی برای یک شبکه عصبی آنلاین با حجم زیاد محاسبات غیرعملی است. همچنین ما از مکانیزم توجه یادگیری تقویتی عمیق در مسئله مسیریابی خودرو یا پهپاد استفاده کردیم تا پارامترهای مدل شبکه عصبی به‌خوبی مقداردهی شوند. برای رسیدن به این هدف، ما از یک شبکه عصبی کریتیک کمک‌کننده[1] استفاده کردیم تا خروجی مورد انتظار داده آموزشی را تخمین بزنیم. در نهایت، با استفاده از احتمال توزیع خروجی شبکه پوینتر یک روش ساختاری چند نمونه‌ای را پیشنهاد می‌دهیم تا مسیرهای خودرویی رمزگشایی شوند. مدل RL زمان بسیار کمتری را برای حل نمونه هایی با تعداد ۵۰ یا ۱۰۰ مشتری خرج می‌کند. در انتها نتایج بر روی یک سرور با مشخصات سخت افزاری زیر تست گردید و نمودار فرایند آموزش هزینه طول مسیرشبکه به صورت شکل ۳ بود. آموزش یک مسئله VRP _TW_10 حدودا ۲۲.۲ ساعت زمان نیاز دارد و با بزرگ شدن مسئله و تعداد نودها، این زمان به ۱۲۰ ساعت برای مسئله با ۱۰۰ درخواست افزایش یافت.



شکل ۳- منحنی طول مسیر در حین آموزش برای هر ۲۰ اپیاک

در شکل ۳ میانگین هزینه راه حل‌های پیش بینی شده در هر دوره را مشاهده می‌کنیم. اگر ما با تنظیم فراپارامتر آزمایش کنیم، در این مورد، می‌توانیم از نرخ یادگیری رو به زوال استفاده کنیم که فرض می‌کنیم بیشتر تابع هزینه را کاهش می‌دهد. برای بهبود بیشتر نتایج، ما از دو رویکرد در طول فرآیند تست، جستجوی حریصانه و جستجوی پرتو استفاده می‌کنیم . برای هر مرحله، استراتژی حریصانه بالاترین احتمال نقطه مشتری را به عنوان اقدام انتخاب می‌کند. با این حال، با توجه به عرض پرتو k، جستجوی پرتو انتخاب‌های بالای k را در یک زمان ذخیره می‌کند و نتایج نهایی را براساس حداکثر احتمال توالی خروجی می‌دهد.

زمان اجرا: مقایسه زمان اجرا سخت اما مهم است. از آنجا که مدل پایان به پایان است، زمان آموزش در هنگام تجزیه و تحلیل عملکرد کلی روش‌های پیشنهادی در نظر گرفته نمی‌شود. زمان اجرای هر روش با اجرای ۱۰۰۰ نمونه در کل محاسبه می‌شود. تمام زمان محاسبه گزارش‌شده در عرض چند ثانیه است. که با توجه به زمان ایجاد مسیر بهینه، این خود انگیزه مناسبی برای استفاده از این مدل ها در مسایل مسیریابی می‌باشد. این روش به طور مؤثری می‌تواند عملکرد سیستم مدیریت حمل و نقل را در تولید مسیر بهینه (feasible solutions)کامیون‌ها بهبود دهد. زمان اجرا برای روش‌های فرااکتشافی[2] به طور تصاعدی با تعداد مشتریان افزایش می‌یابد که دلیل آن ماهیت NP-سخت مسئله است.

مدل DRL_WLRP پیشنهادی با الگوریتم‌های مبنای ذکر شده برای VRP _TW و زمان رسیدن به خروجی به ترتیب در جدول ۱ مقایسه شده‌است. ستون اول مبنای اتخاذ شده برای حل گروه‌های مختلف VRPرا گزارش می‌کند. ستون دوم نوع و نام الگوریتم را بیان می‌کند که در آن DRL مدل DRL_WLRP پیشنهاد شده در بخش۳ می‌باشد. ستون‌های دیگر با عناوین "متوسط طول مسیر"، " تخطی از زمان شروع" و " تخطی از زمان پایان سرویس" ، "درصد رخداد تخطی زمان شروع" و " درصد رخداد تخطی زمان پایان سرویس" را در گروه‌های مختلف گزارش می‌دهند. برای VRP، بهترین نتیجه از حل ریاضی با گروبی حاصل می‌شود البته بعلت افزایش تعداد پارامترها و متغیرهای محدودیت ها حل ریاضی با گروبی نیاز به لایسنس دارد و در حالت عادی جوابگوی حل مسئله نیست. البته در VRP _TW_20 طول مسیر الگوریتم پیشنهادی ما (7.3137) از حل ریاضی نیز بهتر شده است که این موضوع با لحاظ کردن میزان تخطی در هزینه کل قابل توجیه است، چراکه میزان تخطی در حل ریاضی صفر می‌باشد. در جدول3 مشاهده می شود که نتایج بهینه الگوریتم فرابتکاری حتی برای نمونه[3] ۲۰ نودی نیز زمانبر است. برای هر دو VRP s، مدل DRL_WLRP پیشنهادی می‌تواند به نتایج بهتری نسبت به الگوریتم های ساخت مسیر با زمان محاسبه بسیار کوتاه‌تر حتی در اندازه مساله بزرگ دست یابد. روش‌های فرااکتشافی کلاسیک باید هر بار مشکل را از ابتدا حل کنند، درحالیکه مدل آموزش دیده با توجه به اطلاعات قبلی ذخیره شده در مدل، تلاش کمتری برای محاسبات به خرج می‌دهد.

[1] auxiliary critic neural network
[2] Meta heuristic
[3] instance

جدول۱- مقایسه پاداش تجمعی و تخطی مدل DRL_WLRP پیشنهادی با الگوریتم‌های مبنای ذکر شده (Collection Cost: Tour (Length +Ready Time Violations +End Time Violations *

Instance	1000 instances	Tour Length	Cost *
VRP _10	Proposed DRL_WLRP	5.342	7.4451
	GA	5.4463	8.1094
	PSO	5.818	8.83355
	Exact	4.8124	4.8124
VRP _ 20	Proposed DRL_WLRP	7.3137	13.3188
	GA	7.3821	21.0585
	PSO	8.74102	32.2614
	Exact	8.4374	8.4374
VRP _50	Proposed DRL_WLRP	14.6873	84.0876
	GA	16.7918	77.8945
	PSO	22.1877	85.9362
	Exact	-	-
VRP _100	Proposed DRL_WLRP	33.0148	107.076
	GA	61.4745	156.857
	PSO	58.728	222.297
	Exact	-	-

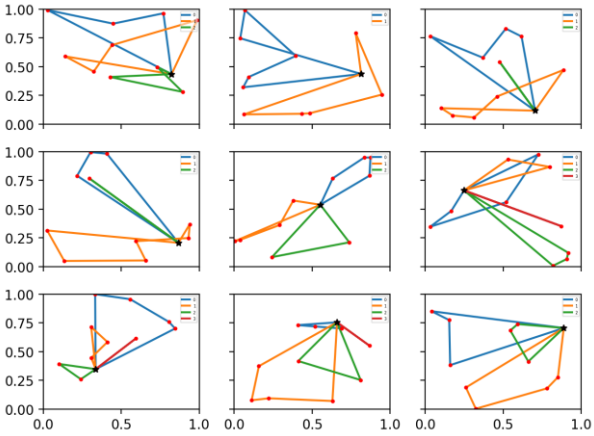
عملکرد مدل DRL_WLRP پیشنهادی خودرا با روش حل ریاضی Grubi و روش‌های فراابتکاری GA_TW و PSO_TW با روش ذکر شده در بالا مقایسه می‌کنیم. جدول ۲ به ترتیب طول متوسط تور، میزان و درصد تخطی های رخداده، و در انتها کل پاداش تجمعی در روش exact با راه‌حل‌های اولیه مختلف را گزارش می‌کنند. ما می‌توانیم مشاهده کنیم که نتایج بهبود یافته با روش‌های DRL_WLRP، از جمله مدل ما ، به طور قابل‌توجهی بهتر از نتایج کسب شده با الگوریتم های فراابتکاری[1] هم در کیفیت راه‌حل طول مسیر و هم در زمان محاسبه و نیز از لحاظ تخطی های صورت گرفته در زمان شروع و پایان سرویس‌دهی است. با در نظر گرفتن VRP _TW_100 به عنوان مثال، نتایج بهبود یافته توسط مدل‌های DRL به میانگین کیفیت جواب طول تور ۳۳.۰۱۴۸ با میانگین زمان حل ۱۰۷.۰۷ ثانیه می‌رسد، در حالی که نتایج بهبود یافته توسط الگوریتم های فراابتکاری به بهترین کیفیت جواب با طول تور ۵۸.۷۲۸ با زمان حل. ۲۲۲.۲۹۷۲ ثانیه می‌رسد. ما بر مقایسه بین PSO با مدل DRL_WLRP خود از نظر کیفیت راه‌حل و زمان محاسبه تمرکز می‌کنیم. با در نظر گرفتن کیفیت راه‌حل با مدل DRL_WLRP برای VRP در تمام ابعاد مساله همچون متوسط طول مسیر یا نیز بصورت پاداش کلی که جمع سه مقدار طول مسیر، تخطی از زمان شروع و پایان در نظر گرفته شود، رقابتی‌تر می‌شود. برای VRP، روش پیشنهادی به نتایج بهتری برای همه گروه‌های تا TW_۱۰۰ دست می‌یابد و همان کیفیت جواب را برای گروه ۲۰_ TW و گروه ۵۰_ TW به دست می‌آورد. برای زمان محاسبه، توجه داشته باشید که این دو مدل DRL دارای ویژگی‌های یکسانی از لحاظ سرعت استنتاج می‌باشند. با افزایش تعداد نمونه های آموزشی به ۲۰۰۰۰۰۰ نمونه، متوسط طول مسیر خروجی شبکه آموزش دیده بهبود می‌یابد و به عدد ۵.۲۹۶۶ رسیدیم. ولی تغییر پارامتر مسئله به ۲۰۰ epoch، تغییر محسوسی در نتیجه خروجی ندارد. در جدول۳ مقایسه متوسط زمان محاسباتی و بازه اطمینان ۹۵ درصد برای طول مسیر تور ارایه شده است. مشاهده می‌شود که با افزایش اندازه مسئله به ۱۰۰ نود، متوسط زمان محاسبه طول مسیر برای الگوریتم‌های کلاسیک بطور معناداری افزایش می‌یابد. زمان تولید تور در مدل DRL_WLRP پیشنهادی زیر یک ثانیه می‌باشد و این خود دلیلی بر کاربرد این راه حل در مسایل شهری بزرگ و زمان واقعی است. زمان محاسبات چارچوب ما تقریباً با افزایش اندازه مسئله از ۱۰ نود به ۲۰، ۵۰ و ۱۰۰ نود یکسان باقی می‌ماند، که بسیار سریع‌تر از معیارها است. در مقابل، زمان اجرا برای روش‌های اکتشافی به‌طور تصاعدی با تعداد مشتریان به دلیل ماهیت NP-hard خود مشکل افزایش می‌یابد. با کاهش اندازه مسئله، روش‌های اکتشافی شروع به نزدیک شدن به نتایج DRL_WLRP می کند. این مورد انتظار بود زیرا در بالا دیدیم که OR-Tools در اندازه های کوچکتر مسایل از چارچوب DRL بهتر عمل می کند.

برای تصور بهتر پیچیدگی و تنوع حل VRP _TW، در شکل ۴ ما یک راه حل تصادفی پیش بینی شده توسط مدل یادگیری عمیق پیشنهادی خود را نشان می دهیم. نمونه خروجی شبکه اموزش دیده تست VRP _TW_10 در شکل ۴ نمونه تست در شکل نمایش داده شده است. نود * سیاه رنگ، مخزن و دیگر نقاط قرمز رنگ روی شکل درخواست‌های سرویس از سوی کاربران و یا مخازن هوشمند می‌باشند. در شکل مسیرهای بهینه ایجاد شده که ماشین توسط یادگیری تقویتی آموزش دیده است را نمایش می‌دهیم. برای تمایز بهتر هر مسیر و برگشت ماشین حمل زباله به محل دپو، هر مسیر با یک رنگ نمایش داده شده است. الگوریتم به نحوی شایسته مسیریابی ناوگان را انجام و مسیری بهینه شامل نقاط درخواست را با کمترین طول ارایه می‌نماید. مشاهده می‌شود که الگوریتم به خوبی مسیر نزدیک به بهینه را در زمان کم خروجی می‌دهد. در یک مسئله یادگیری تقویتی با عاملی روبرو هستیم که از طریق سعی و خطا با محیط تعامل کرده و یاد می‌گیرد تا عملی بهینه را برای رسیدن به هدف انتخاب نماید. یادگیری تقویتی ازاین‌رو موردتوجه است که راهی برای آموزش عامل‌ها برای انجام یک عمل از طریق دادن پاداش و تنبیه است بدون اینکه لازم باشد نحوه انجام عمل را برای عامل مشخص نماییم. متداول‌ترین روش استفاده از یک شبکه عصبی مکرر(RNN) است زیرا آنها به طور گسترده در توابع یادگیری در مورد توالی‌ها استفاده می‌شوند.

یادگیری تقویتی مدل‌هایی مانند فرایندهای تصمیم‌گیری مارکوف را ارائه می‌دهد که به طور طبیعی تکامل سیستم های تصادفی و پویا را به تصویر می‌کشد. فرایندهای تصمیم‌گیری مارکوف چارچوب‌های

[1]Meta heuristic

ریاضی هستند که برای تشریح محیط در یادگیری تقویتی استفاده می‌شوند و تقریباً همه مسائل این حوزه قابل‌ارائه شدن با MDP ها هستند. بعنوان چالش‌های اصلی مدیریت حمل و نقل به‌عنوان مثال وجود چند انبار و نیز محدودیت پنجره زمانی، در نظر گرفتن خرابی وسیله، اعمال مشکل ترافیک- تصادف و است که بایستی در ارایه مدل مسیریابی داینامیک لحاظ گردد. در انتها ضرورت دارد جهت حل مسایل بزرگ و حل مسئله در زمان سریع، به سمت ارایه الگوریتم های یادگیری ماشین و خودکار حرکت کنیم.



شکل ٤- نمایش نمونه راه حل نمونه تصادفی برای اندازه مسئله ۱۰ و مسیر خروجی نمونه تست

٤-خلاصه و نتیجه گیری

فرمول‌بندی VRP با پهپاد به عنوان یک MDP و استفاده از DRL رویکردی نوآورانه برای حل این مسئله پیچیده است. با وجود چالش‌های موجود، این رویکرد پتانسیل بالایی برای بهینه‌سازی مسیرهای پهپاد در سناریوهای مختلف VRP و دستیابی به راه‌حل‌های جدید و کارآمدتر برای مسائل حمل و نقل و لجستیک دارد. مسئله مسیریابی خودرو (VRP) یک مسئله بهینه‌سازی ترکیبی است که به دنبال یافتن بهترین مسیر برای یک یا چند وسیله نقلیه برای بازدید از مجموعه‌ای از نقاط با رعایت محدودیت‌هایی مانند ظرفیت و مسافت است. با ظهور پهپادها، VRP پیچیدگی جدیدی پیدا کرده است، زیرا پهپادها دارای محدودیت‌های خاص مانند برد، ظرفیت حمل و نقل و نیاز به فرود برای شارژ مجدد هستند. در این مقاله پیشنهاد روش جدید بهینه‌سازی با یادگیری تقویتی عمیق جهت مسیریابی ناوگان حمل و نقل با حداقل زمان محاسباتی در شبکه‌های بزرگ با ارایه معماری پویا با توجه به نیاز مسئله حمل و نقل در شهرهای بزرگ و نیز برای پهپادها به همراه ارایه پاداش ترکیبی با توجه به محدودیت های مسئله حمل و نقل در مسئله یادگیری تقویتی عمیق و Feaseability ruleهای بهتر و کامل‌تری برای مسئله مورد بررسی که بعنوان چالش‌های اصلی مدیریت حمل و نقل مطرح گردید. (به‌عنوان مثال وجود چند انبار و نیز محدودیت پنجره زمانی، در نظر گرفتن خرابی وسیله، اعمال مشکل ترافیک- تصادف و ...) . تلاش برای تحقق خودکارسازی بیشتر در مسیریابی پهپادها با استفاده از الگوریتم های هوش مصنوعی، تلاش برای افزایش هماهنگی و همکاری بین پهپادها و سایر اجسام در محیط برای بهبود عملکرد مسیریابی و استفاده از شبکه های سنسور برای جمع آوری داده های محیط و بهبود دقت مسیرابی پهپادها از اهم موارد و جهت گیری های اینده هستند. با توجه به پیشرفت فناوری و تحقیقات در این حوزه، انتظار می رود که روش ها و الگوریتم های پیشرفته تر برای مسیریابی چند هدفه برای پهپادها توسعه یافته و جهت دهنده برای آینده این حوزه باشند. در انتها ارایه مدل شبکه تقویتی عمیق و مکانیزم پاداش و تنبیه با عملکرد بهتر از مسایل ابتکاری و فراابتکاری مقایسه گردیده و حل ریاضی مسئله نیز با استفاده از حل کننده گروبی نیز انجام گرفت.نتایج شبیه‌سازی نشان می‌دهد که مدل DRL_WLRP پیشنهادی نسبت به الگوریتم های ریاضی و دیگر مدل‌های DRL عملکرد بهتری داشته و می‌تواند به طور موثری نرخ هم‌گرایی را سرعت بخشیده و کیفیت راه‌حل را بهبود بخشد. ما آزمایش‌ها جامعی را برای ارزیابی عملکرد چارچوب، با استفاده از الگوریتم GA و PSO انجام می‌دهیم. نتایج تجربی نشان می‌دهد که چارچوب پیشنهادی می‌تواند راهبردهای مسیریابی برجسته‌ای را در مقایسه با جستجوی ابتکاری با راه‌حل‌های اولیه تولید شده توسط سایر الگوریتم ایجاد کند. علاوه بر بهبود کیفیت راه‌حل، تولید مسیر چارچوب پیشنهادی زمان محاسباتی کوتاهی را هزینه می‌کند، که این چارچوب را برای مسایلی با اندازه بزرگ امیدوار کننده می‌سازد. در مورد کارهای تحقیقاتی آینده در مورد آموزش شبکه DRL_WLRP با دیتاهای دنیای واقعی، اعمال یادگیری ماشین در مراحل پیش بینی درخواست و نیز تعداد ناوگان مورد نیاز و افزودن محدودیت‌هایی همچون مصرف انرژی و نیز خودرو یا پهپاد های الکتریکی زمینه های مناسبی خواهد بود. در ادامه، تجزیه و تحلیل عمیق‌تری از ویژگی‌های تعبیه ورودی باید انجام شود و ساختار گرافیکی نمونه‌های مسئله مسیریابی و ارتباط آن با معماری‌های معروف ML مرتبط شود. برخی از راه‌های ممکن برای بهبود عملکرد ترکیب بیشتر الگوریتم‌های کلاسیک با رویکردهای RL باشد.

٥-مراجع

[1] Aggarwal, S., & Kumar, N. (2020). Path planning techniques for unmanned aerial vehicles: A review, solutions, and challenges. Computer communications, 149, 270-299.

[2] Debnath, S. K., Omar, R., & Latip, N. B. A. (2019). A review on energy efficient path planning algorithms for unmanned air vehicles. Computational Science and Technology: 5th ICCST 2018, Kota Kinabalu, Malaysia, 29-30 August 2018, 523-532.

[3] P. Fernandez-Aracil, A. Ortuno-Padilla and J. Melgarejo-Moreno. (2018). "Factors related to municipal costs of waste collection service in Spain," in J. Clean. Prod., 175, 553–560., 2018 .

[4] Khan, A. H., López-Maldonado, E. A., Khan, N. A., Villarreal-Gómez, L. J., Munshi, F. M., Alsabhan, A. H., & Perveen, K. (2022). Current solid waste management strategies and energy recovery in developing countries-State of art review. Chemosphere, 291, 133088 .

[19] بابایی, محسن. (۱۴۰۱). مسیریابی وسایل نقلیه حمل کالا با قابلیت در نظر گرفتن محدودیت ظرفیت و هزینه ثابت بکارگیری ناوگان. فصلنامه مهندسی حمل و نقل.

[20] Chaudhari, M. S., Patil, B., & Raut, V. (2019, March). Iot based waste collection management system for smart cities: an overview. In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 802-805). IEEE.

[21] Kuo, R. J., Lu, S. H., Lai, P. Y., & Mara, S. T. W. (2022). Vehicle routing problem with drones considering time windows. Expert Systems with Applications, 191, 116264.

[22] Phonphoton, N., & Pharino, C. (2019). A system dynamics modeling to evaluate flooding impacts on municipal solid waste management services. Waste Management, 87, 525-536 .

[23] Raucq, J., Sörensen, K., & Cattrysse, D. (2019). Solving a real-life roll-on–roll-off waste collection problem with column generation. Journal on Vehicle Routing Algorithms, 2(1), 41-54 .

[24] Guo, N., Qian, B., Na, J., Hu, R., & Mao, J. L. (2022). A three-dimensional ant colony optimization algorithm for multi-compartment vehicle routing problem considering carbon emissions. Applied Soft Computing, 127, 109326.

[25] Hemidat, S., Oelgemöller, D., Nassour, A., & Nelles, M. (2017). Evaluation of key indicators of waste collection using GIS techniques as a planning and control tool for route optimization. Waste and Biomass Valorization, 8(5), 1533-1554 .

[26] Sulemana, A., Donkor, E. A., Forkuo, E. K., & Oduro-Kwarteng, S. (2018). Optimal routing of solid waste collection trucks: a review of methods. Journal of Engineering, 2018 .

[27] A. Nadizadeha and H. H. Nasaba, . (2014). "Solving the Dynamic Capacitated Location-Routing Problem with Fuzzy Demands by Hybrid Heuristic Algorithm", European Journal of Operational Research, In Press Available Online, Elsevier.

[28] Jorge, D., Antunes, A. P., Ramos, T. R. P., & Barbosa-Póvoa, A. P. (2022). A hybrid metaheuristic for smart waste collection problems with workload concerns. Computers & Operations Research, 137, 105518.

[۲۹] کبگانی, محمدحسین, شاهبندرزاده, حمید, حیدری, ابراهیم, & امیری, سید نورالدین. (۱۴۰۱). ارائه مدل چندهدفه فازی – استوار برای مسأله مکان‌یابی – تخصیص – مسیریابی موجودی سبز در بهینه‌سازی فرایند مدیریت پسماند شهری با استفاده از الگوریتم‌های فرا ابتکاری. فصلنامه مهندسی حمل و نقل, ۱۳(۴), ۱۹۸۳-۲۰۰۸. doi: 10.22119/jte.2021.282693.2525

[30] Vu, H. L., Ng, K. T. W., Fallah, B., Richter, A., & Kabir, G. (2020). Interactions of residential waste composition and collection truck compartment design on GIS route optimization. Waste management, 102, 613-623 .

[31] Abdallah, M., Adghim, M., Maraqa, M., & Aldahab, E. (2019). Simulation and optimization of dynamic waste collection routes. Waste Management & Research, 37(8), 793-802 .

[32] Vasagade, T. S., Tamboli, S. S., & Shinde, A. D. (2017, March). Dynamic solid waste collection and management system based on sensors, elevator and GSM. In 2017 International conference on inventive communication and

[5] Nanda, S., & Berruti, F. (2021). Municipal solid waste management and landfilling technologies: a review. Environmental chemistry letters, 19(2), 1433-1456.

[6] M. Kharat, S. Murthy, S. Kamble, R. Raut, S. Kamble and M. Kharat. (2019)."Fuzzy multi-criteria decision analysis for environmentally conscious solid waste treatment and disposal technology selection.," in Technol. Soc. 57, 20–29 „ „.

[7] P. Andrey, I. Valentina and S. Yuri. (2020). "Strategic approaches to management in the field of solid municipal waste management.," in Talent Dev. Excell. 12, 700–709 „

[8] S. Nesmachnow, D. Rossit and J. Toutouh, . (2018)."Comparison of Multiobjective Evolutionary Algorithms for Prioritized Urban Waste Collection in Montevideo,," in Uruguay. Electron. Notes Discret. Math. , 69, 93–100 .

[9] G. Ghiani, D. Laganà, E. Manni and C. Triki. (2012).". Capacitated location of collection sites in an urban waste management system.," in J. Waste Manag.32, 1291–1296 .

[10] E. Tirkolaee, P. Abbasian, M. Soltani and S. Gha_arian. (2019). " Developing an applied algorithm for multi-trip vehicle routing problem with time windows in urban waste collection: A case study.," in Waste Manag. Res. , 37, 413 .

[11] Hannan, M. A., Akhtar, M., Begum, R. A., Basri, H., Hussain, A., & Scavino, E. (2018). Capacitated vehicle-routing problem model for scheduled solid waste collection and route optimization using PSO algorithm. Waste management, 71, 31-41 .

[12] M. H. M. A. B. R. A. B. H. &. S. E. Akhtar. (2017)."Backtracking search algorithm in CVRP models for efficient solid waste collection and route optimization „in Waste Management, 61, 117-128 ."

[13] A. Expósito-Márquez, C. Expósito-Izquierdo, J. Brito-Santana and J. Moreno-Pérez. (2019). " Greedy randomized adaptive search procedure to design waste collection routes in La Palma.," in Comput.Ind. Eng. , 137 .

[14] Hashemi-Amiri, O., Mohammadi, M., Rahmanifar, G., Hajiaghaei-Keshteli, M., Fusco, G., & Colombaroni, C. (2023). An allocation-routing optimization model for integrated solid waste management. Expert Systems with Applications, 227, 120364.

[15] S. Jaeger, J. Eyckmans, N. Rogge and T. Puyenbroeck. (2011). "Wasteful waste-reducing policies? The impact of waste reduction policy instruments on collection and processing costs of municipal solid waste.," in Waste Manag. , 31, 1429–1440 .

[16] F. Melakessou, P. Kugener, N. Alna_akh, S. Faye and D. Khadraoui. (2020). "Heterogeneous Sensing Data Analysis for Commercial Waste Collection.," in Sensors (Basel, Switzerland) , 20, 978.

[17] Bhor, V., Morajkar, P., Gurav, M., Pandya, D., & Deshpande, A. (2015). Smart garbage management system. International Journal of Engineering Research & Technology (IJERT), 4(03), 2000.

[18] Nižetić, S., Šolić, P., González-de, D. L. D. I., & Patrono, L. (2020). Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. Journal of Cleaner Production, 274, 122877.

[37] İ. Ilhan. (2020). "A population based simulated annealing algorithm for capacitated vehicle routing problem ″„in Turkish Journal of Electrical Engineering & Computer Sciences, 28(3), 1217-1235.

[38] E. Tirkolaee, M. Alinaghian, A. Hosseinabadi, M. Sasi and A. Sangaiah. (2019)." An improved ant colony optimization for the multi-trip Capacitated Arc Routing Problem.," in Comput. Electr. Eng. , 77, 457–470.

[39] N. Wichapa and P. Khokhajaikiat. (2018). "Solving a multi-objective location routing problem for infectious waste disposal using hybrid goal programming and hybrid genetic algorithm.," in Int. J. Ind. Eng. Comput. , 9, 75–98.

[40] L. Delgado-Antequera, R. Caballero, J. Sánchez-Oro, J. Colmenar and R. Martí. (2020). " Iterated greedy with variable neighborhood search for a multiobjective waste collection problem.," in J. Expert Syst. Appl. , 145, 113101.

[41] J Liu, M., Song, Q., Zhao, Q., Li, L., Yang, Z., & Zhang, Y. (2022). A hybrid BSO-ACO for dynamic vehicle routing problem on real-world road networks. IEEE Access, 10, 118302-118312..

computational technologies (ICICCT) (pp. 263-267). IEEE.

[33] Lavigne, C., Inghels, D., Dullaert, W., & Dewil, R. (2023). A memetic algorithm for solving rich waste collection problems. European Journal of Operational Research, 308(2), 581-604.

[34] Bhargava, K., Gupta, R., Singhal, A., & Shrinivas, A. (2019). Genetic Algorithm to Optimize Solid Waste Collection. In Proceedings of the 3rd International Conference of Recent Trends in Environmental Science and Engineering (RTESE'19) .

[35] J. De Armas, B. Melian-Batista, J. Moreno-Perez and J. Brito. (2015). " GVNS for a real-world Rich Vehicle Routing Problem with Time Windows.," in Eng. Appl. Artif. Intell. , 42, 45–56, 2015.

[36] H. Rau, S. Budiman and G. Widyadana. (2018)."Optimization of the multi-objective green cyclical inventory routing problem using discrete multi-swarm PSO method.," in Transp.Res. Part E-Logist. Transp. Rev. , 120, 51–75 .