مجله کنترل
ISSN (print) 2008-8345
ISSN (online) 2538-3752
مقاله پژوهشی
جلد۱۸، شماره ۲، تابستان ۱۴۰۳، صفحه ۱۲-۱

یادگیری تقویتی فازی مبتنی بر تکرار ارزش در ربات تعقیب کننده‌ی هدف

فرزانه نادی[۱]، ولی درهمی[۲] و فریناز اعلمی‌یان هرندی[۳]

[۱] دانشجوی دکتری، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران farzane.nadi@gmail.com
[۲] استاد، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران vderhami@yazd.ac.ir
[۳] پژوهشگر پسا دکتری، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی اصفهان، اصفهان، ایران farinaz.alamiyan@gmail.com

دریافت: ۱۴۰۲/۰۹/۱۰ ویرایش: ۱۴۰۲/۱۲/۰۵ پذیرش: ۱۴۰۳/۰۳/۲۷

چکیده: این مقاله روشی جدید در استفاده از داده‌های جمع آوری شده از حرکت تصادفی عامل در محیط برای تنظیم اولیه‌ی پارامترهای یک کنترلگر با ساختار یادگیری تقویتی فازی ارائه می‌دهد. کندی سرعت آموزش و تعداد شکست بالا در زمان آموزش دو چالش مهم در این قبیل ساختارها هستند. مقداردهی اولیه‌ی پارامترهای سیستم فازی می‌تواند راهکار مناسبی برای رفع این چالش‌ها باشد. در این مقاله با تعمیم روش تکرار ارزش گسسته به پیوسته بدون بهره‌گیری از روش‌های مبتنی بر مشتق، پارامترهای سیستم فازی مقدار دهی اولیه می‌شوند. ابتدا با تعامل تصادفی عامل با محیط داده‌های مرتبط جمع‌آوری می‌شود. با توجه به آنکه فضای حالت پیوسته است، داده‌ها به طور مناسب خوشه بندی شده و هر خوشه به عنوان یک حالت لحاظ می‌گردد. آنگاه با تعمیم روش تکرار ارزش استاندارد به پیوسته ماتریس احتمال انتقال حالت-عمل به حالت بعدی و امید پاداش آنی حالت-عمل به حالت بعدی محاسبه می‌شود. با استفاده از نتایج این مرحله پارامترهای ساختار یادگیری تقویتی فازی مقدار دهی اولیه می‌شوند. پس آز آن پارامترهای این ساختار به صورت برخط با روش یادگیری تقویتی تنظیم نهایی می‌گردند. روش ارایه شده "یادگیری تقویتی فازی مبتنی بر تکرار ارزش" نامیده می‌شود و در مسئله‌ی ربات تعقیب کننده‌ی هدف مورد استفاده قرار می‌گیرد. نتایج آزمایش‌ها حاکی از بهبود قابل توجه عملکرد روش ارائه شده در مسئله‌ی ربات تعقیب کننده‌ی هدف است.

کلمات کلیدی: کنترلگر فازی، یادگیری تقویتی، برنامه‌سازی پویا، خوشه‌بندی، ربات تعقیب کننده‌ی هدف.

Value Iteration based Fuzzy Reinforcement Learning in Target Following Robot

Farzaneh Nadi, Vali Derhami, Farinaz Alamiyan Harandi

Abstract: This paper presents a new method for using data collected from the agent's random movement in the environment for the initial adjustment of parameters of a controller with a fuzzy reinforcement learning structure. Slow learning speed and high failure rates during training are two major challenges in such structures. The initial parameterization of the fuzzy system can be a suitable solution to address these challenges. In this paper, the method of discrete value iteration is extended to continuous without relying on derivative based methods to initialize the parameters of the fuzzy system. First, random interaction with the environment is used to collect relevant data. Since the state space is continuous, the data is appropriately clustered and each cluster is considered as a state. Then, by generalizing the standard value iteration method to the continuous, the transition probability matrix and the immediate reward expectation matrix are calculated. Using the results of this stage, the initial parameterization of the fuzzy reinforcement learning structure is performed. Subsequently, these parameters are fine-tuned using reinforcement learning. The proposed method is called "Value Iteration based Fuzzy Reinforcement Learning" and is used in the problem of target following robots. The experimental results indicate a significant improvement in the performance of the proposed method in the problem of target following robots.

Keywords: Fuzzy Controller, Reinforcement Learning, Dynamic Programming, Clustering, Target Following Robot.

۱- مقدمه

در سال‌های اخیر ربات‌ها به ابزاری قدرتمند برای انسان‌ها تبدیل شده‌اند و در انجام بسیاری از کارها به آن‌ها کمک می‌کنند. ربات "تعقیب کننده‌ی هدف "[1] رباتی است که با استفاده از راه‌حل‌های مختلف سعی در پیگیری هدف و انجام وظایف خود دارد. ربات تعقیب کننده‌ی هدف از جمله ربات‌های پرکاربرد است که این روزها پژوهش‌های متعددی برای بهبود عملکرد آن‌ها انجام شده است. این ربات‌ها بر اساس محیط عملیاتی به سه دسته‌ی زمینی، زیر آب [۱] و هوایی [۲،۳] تقسیم می‌شوند. در این پژوهش، تمرکز بر روی ربات‌های تعقیب کننده‌ی هدف در محیط‌های زمینی است. از جمله ربات‌های این دسته می‌توان به ربات‌های دستیار خانگی، ربات‌های سبد خرید، و ربات‌های کمکی در صنایع و کارخانه‌های هوشمند [۴-۷] اشاره کرد که استفاده از آن‌ها پیوسته در حال گسترش است.

ربات‌ها برای درک محیط پیرامون خود به حسگرها و تجهیزات گوناگونی مجهز می‌شوند. در طراحی ربات‌های تعقیب کننده‌ی هدف نیز می‌توان از حسگرهای متفاوتی استفاده کرد. حسگرهای سونار[2] و پویشگرهای لیزری (از جمله LiDAR) نمونه‌ای از حسگرهای مورد استفاده در این ربات‌هاست [۸،۹] که با بررسی زمان ارسال و دریافت امواج، فاصله‌ی اشیا موجود در محیط را محاسبه می‌کنند. دو چالش مهمِ استفاده از این حسگرها در ربات‌های تعقیب‌کننده، تک‌بُعدی بودن و انحراف موج بازتاب از سطوح خاص و لبه‌ها است که راه‌های رفع این چالش‌ها هزینه‌بر خواهد بود. این روزها، استفاده از دوربین به دلیل ارزان قیمت بودن برغم در دسترس قرار دادن اطلاعات زیاد از محیط بسیار مورد توجه قرار گرفته است. استفاده از تک دوربین به دلیل زاویه‌ی دید کم و عدم دسترسی به اطلاعات فاصله‌ای قابل اطمینان، چالش‌برانگیز است که ترکیب اطلاعات ناشی از این حسگر با حسگر دیگری اعم از پویشگر لیزری و یا استفاده از چند تصویر پشت سر هم، از راه‌های رفع این چالش‌ها می‌باشد [۱۰-۱۲]. پژوهشی در سال ۲۰۱۹، روش جدیدی به نام انتخاب ویژگی بر اساس قطعیت انتقال[3] برای استخراج ویژگی‌های مفید برای کنترل ربات از تصاویر یک دوربین ارائه نموده است [۱۳]. حسگر RGBD گزینه‌ی دیگری است که در این ربات‌ها استفاده می‌شود [۱۴،۱۵]. این حسگر، تصویر رنگی و فاصله‌ی اشیا موجود در محیط تا حسگر (عمق اشیا) را به عنوان خروجی ارائه می‌دهد. با استفاده از این حسگر برای ربات‌های تعقیب کننده‌ی هدف، محل و فاصله‌ی هدفِ مورد تعقیب و اشیا (موانع) موجود در محیط قابل دسترسی است اما با توجه به نحوه‌ی عملکرد این حسگرها (استفاده از اشعه‌ی مادون قرمز برای محاسبه‌ی فاصله اشیا موجود در محیط تا حسگر)، استفاده از آن‌ها برای ربات‌ها در محیط بیرونی امکانپذیر نیست. استفاده از دو دوربین در حالت استریو و تقریب عمق اشیا با روش‌های متفاوت (انجام محاسبات و بدون استفاده از اشعه) راه‌حل دیگری برای دسترسی به اطلاعات اشیا و هدف موجود در محیط است [۱۶-۱۸] که در این پژوهش نیز از این مدل حسگرها استفاده شده است.

یکی از راه‌های کنترل ربات تعقیب کننده هدف استفاده از کنترلگر از نوع PID است. در پژوهشی توسط چوی و همکارانش، از این کنترلگر برای ربات تعقیب کننده‌ی بازیکن گلف بهره برداری شده است [۱۹]. در این ربات از سیگنال‌های رادیویی برای تشخیص هدف و کنترلگر PID برای تعقیب هدف استفاده می‌شود. از چالش‌های استفاده از این کنترلگر، تنظیم بهینه پارامترهای آن است. همچنین، در این پژوهش روشی برای تشخیص و دوری از موانع نیز بیان نشده است.

این روزها برای کنترل ربات تعقیب کننده‌ی هدف و هدایت آن به سمت هدف، استفاده از شبکه‌های عصبی عمیق بسیار مورد توجه قرار گرفته است. از جمله این پژوهش‌ها می‌توان به استفاده از شبکه‌ی عصبی کانولوشن و حسگر استریو [۱۶] اشاره کرد. در پژوهش دیگری که توسط آلگبری و همکارش انجام شده، با استفاده از اطلاعات ناشی از دوربین RGBD و یک شبکه عصبی عمیق (SingleShot MultiBox Detector) تمام افراد محیط شناسایی می‌شوند و سپس با استفاده از هیستوگرام hue-saturation-value فرد هدف تشخیص داده می‌شود [۸]. برای تعقیب هدف سه وضعیت در نظر گرفته شده است: ۱- تعقیب هدف (هدف در ناحیه دید دوربین ربات است)، ۲- گم شدن هدف (خارج شدن هدف از ناحیه دید دوربین ربات) و ۳- جستجوی هدف. در وضعیت ۲ ربات به آخرین موقعیت هدف منتقل می‌شود، اما اگر هدف مشخص نشد وارد وضعیت ۳ می‌شود و به چرخش برای جستجوی هدف می‌پردازد. در وضعیت ۱ و ۲ با استفاده از حسگر LiDAR و نقشه حاصل از محیط، ربات با استفاده از

1 Target Following Robot

2 Sonar sensors

3 Transition Certainty based Feature Selection (TCFS)

کنترلگر PID هدایت می‌شود. علاوه بر چالش تنظیم بهینه‌ی پارامترهای این کنترلگر، به دلیل استفاده از حسگر RGBD، این پژوهش فقط برای تعقیب هدف در محیط‌های داخلی قابل استفاده است.

در پژوهش دیگری، چارچوبی برای ردیابی و شناسایی افراد با استفاده از یک دوربین ارائه شده است [۱۱،۱۲]. ابتدا افراد حاضر در تصویر با استفاده از فیلتر کالمن مشخص می‌شوند. سپس، فرد مورد نظر (هدف) با ترکیبی از ویژگی‌های کانال کانولوشنی و تقویت آنلاین شناسایی می‌گردد. علاوه بر چالش موجود برای آموزش شبکه‌های عصبی عمیق، در این پژوهش از یک کنترلگر ساده برای کنترل ربات استفاده شده است. بدین گونه که ربات به سمت هدف حرکت می‌کند اما در صورت گم کردن مسیر هدف، ربات می‌ایستد و منتظر می‌ماند تا فرد دوباره ظاهر شود. در این پژوهش نحوه‌ی عمل ربات در مواجه به موانع نیز بیان نشده است.

یکی از چالش‌های مهم در مورد استفاده از شبکه‌های عصبی عمیق، نیاز به تعداد زیادی داده‌ی آموزشی برای آموزش شبکه‌های عصبی عمیق است. همچنین، این کنترلگرها در دسته‌ی کنترلگرهای جعبه سیاه قرار می‌گیرند، بنابراین غیر توصیف‌پذیر هستند. بطور کلی غیر توصیف‌پذیری، نیاز به داده‌های آموزشی زیاد، حجم محاسبات بالا، و عدم امکان گنجاندن دانش بشری مهمترین ضعف کنترلگرهای جعبه سیاه است. لذا استفاده از کنترلگرهای جعبه سفید جهت غلبه بر چالش‌های فوق مورد توجه قرار گرفته است. سیستم استنتاج فازی یک ساختار جعبه سفید است که توصیف پذیری و امکان گنجاندن دانش بشری در آن به عنوان دو ویژگی برجسته آن قابل توجه است [۲۰]. بنابراین در پژوهش‌های جدید بسیار مورد توجه قرار گرفته است [۲۱،۲۲].

در پژوهشی که در سال ۲۰۲۲ با بهره‌گیری از کنترلگر فازی بر روی ربات تعقیب کننده‌ی هدف انجام شده، از دوربین و حسگر LiDAR استفاده شده است [۲۳]. این پژوهش یک استراتژی رفتار تطبیقی مقاوم مبتنی بر مکانیزم استنتاج فازی را ارائه نموده تا هدف به خوبی دنبال شود. فاصله تعقیب ایمن و سرعت فعلی ربات ورودی‌های سیستم کنترلی فازی هستند. خروجی سیستم نیز سرعت مناسب ربات است. پژوهش دیگری در راستای همین پژوهش، با هدف ۱- تعریف مقادیر زبانی قابل تفسیر توسط انسان در هر بعد، ۲- نشان دادن رابطه‌ی بین متغیرهای زبانی، و ۳- بدست آوردن خروجی با استفاده از ورودی‌ها با روشی قابل فهم‌تر و ساده‌تر، انجام شده است [۲۴].

یکی از گام‌های مهم بعد از مقداردهی اولیه‌ی کنترلگر فازی، تنظیم نرم تالی قواعد سیستم فازی با یکی از الگوریتم‌های یادگیری است. از پر کاربردترین الگوریتم‌های یادگیری در مسائل رباتیک، الگوریتم یادگیری تقویتی است. آموزش در یادگیری تقویتی تنها با استفاده از یک معیار اسکالر راندمان که سیگنال تقویتی نامیده می‌شود صورت می‌گیرد. در مسئله‌ی ربات تعقیب کننده‌ی هدف هر دو فضای عمل و حالت پیوسته هستند، لذا لازم است از الگوریتم‌های یادگیری تقویتی پیوسته بهره جست. از مهم‌ترین این الگوریتم‌ها می‌توان به یادگیری سارسای فازی [۲۵] اشاره نمود.

دو ضعف عمده الگوریتم‌های یادگیری تقویتی، سرعت پایین آموزش و تعداد شکست بالا در زمان آموزش است. برای غلبه بر این ضعف محققین روش‌های یادگیری تقویتی با ناظر [۱۳،۲۶،۲۷] را ارائه داده‌اند. در دسته‌ای از این کارها [۱۳] ابتدا با استفاده از روش‌های مبتنی بر گرادیانت، پارامترهای سیستم فازی در جهت کاهش مجموع مربعات خطا تنظیم شده و سپس به صورت بر خط با استفاده از روش‌های یادگیری تقویتی فازی پارامترها تنظیم نرم شده‌اند. چند نقطه ضعف برای این روش‌ها لحاظ شده است. پیدا کردن مقدار تالی قواعد با روش‌های مبتنی بر گرادیانت به خصوص زمانی که داده‌های ناسازگار وجود دارد، در برخی حالت‌ها جواب مناسبی نمی‌دهد. به عنوان مثال اگر در یک وضعیت در کنار مانع برای زاویه‌ی حرکت سر ربات در داده‌ی آموزشی یک بار مقدار ۴۵ درجه و یک بار مقدار ۴۵- درجه باشد، مقداری که روش مذکور برمی‌گرداند مقدار صفر است، که به معنی برخورد به مانع است. مشکل دیگر این دسته از روش‌ها گیر افتادن در نقاط مینمم محلی است و در نتیجه مقداردهی تالی قواعد با بهینه‌ترین مقدار ممکن اتفاق نمی‌افتد. برای رفع چالش‌های بیان شده، فتحی‌نژاد و همکارانش [۲۶] یک روش جدید برای تعیین اولیه‌ی مقدار پارامترهای تابع ارزش در یادگیری سارسای فازی را ارائه دادند. روش مذکور راهکاری برای تنظیم توابع عضویت ورودی سیستم فازی ارائه نمی‌دهد. همچنین در هنگام آموزش برخی پیچیدگی‌ها در تعیین ترکیبی از مقدار تالی‌ها که به مقدار خروجی مطلوب برسد وجود دارد. از دیگر چالش‌های موجود در روش‌های با ناظر این است که جمع‌آوری داده‌های آموزشی توسط ناظر در بعضی مسائل با دشواری‌هایی روبه‌رو است. ناسازگاری در داده‌ها و وجود داده‌های نویزی بسیار چالش برانگیز هستند و گاه کیفیت کنترلگر را به شدت تحت تأثیر قرار می‌دهند.

در این پژوهش یک روش جدید ارائه می‌دهیم که با استفاده از داده‌های جمع‌آوری شده از محیط و روش تکرار ارزش فازی که قبلا توسط نویسندگان [۲۸] ارائه شده است، پارامترهای کنترلگر فازی موردنظر را مقداردهی اولیه می‌نماییم. توجه شود که داده‌های جمع‌آوری شده با حرکت تصادفی ربات در محیط بدست می‌آید و در واقع از آنها برای تهیه‌ی مدل گذر از حالت و عمل به حالت بعدی استفاده می‌شود. پس از

مقداردهی اولیه‌ی پارامترهای کنترلگر فازی، با استفاده از روش یادگیری سارسای فازی پارامترها تنظیم نهایی می‌شود. بطور خلاصه سهم علمی مقاله به شرح زیر است: الف) ارایه یک ساختار مبتنی بر تجزیه رفتارها برای ربات تعقیب کننده هدف. ب) بکارگیری ایده روش تکرار ارزش فازی در مقداردهی اولیه پارامترهای کنترلگر فازی. ج) تعریف سیگنال تقویتی مناسب برای مسئله‌ی ربات تعقیب کننده‌ی هدف. د) پیاده‌سازی روش برای مسئله‌ی ربات تعقیب کننده‌ی هدف.

ساختار مقاله بدین گونه است: در بخش دوم مفاهیم پایه مرور می‌شود. چارچوب ارائه شده و روش پیشنهادی به ترتیب در بخش سوم و چهارم شرح داده می‌شود. بخش پنجم بیان کننده‌ی نتیجه‌ی آزمایش‌ها است و در نهایت، نتیجه‌گیری و پیشنهادها در بخش ششم بیان می‌شود.

۲- مفاهیم پایه

در این بخش یادگیری تقویتی فازی و برنامه‌سازی پویا به اختصار بیان می‌شود. در زیربخش یادگیری تقویتی فازی، یادگیری سارسای فازی به عنوان مشهورترین این سیستم‌ها توضیح داده می‌شود.

۲-۱ یادگیری تقویتی فازی

با ترکیب روش‌های یادگیری تقویتی و سیستم‌های فازی به عنوان تقریب زننده‌های تابع، سیستم‌های یادگیری تقویتی فازی ارائه شدند. معماری‌های یادگیری تقویتی فازی در نگاهی گسترده، به سه دسته معماری نقاد-تنها، عملگر-تنها و عملگر-نقاد تقسیم می‌شوند. هر یک از این معماری‌ها مزایا و معایب خاص خود را دارند. اما معماری نقاد-تنها به دلیل درجه‌ی کاوش بالاتر و شفافیت بیشتر در گنجاندن دانش خبره در آن کاربرد وسیع‌تری دارد [۲۹،۳۰].

در معماری نقاد-تنهای یادگیری تقویتی با استفاده از تقریب‌زننده‌های تابع، نگاشت میان فضای حالت و فضای عمل تقریب زده می‌شود. دو روش پایه‌ای در این معماری، یادگیری کیوی فازی[1] و یادگیری سارسای فازی نامیده شده‌اند. این روش‌ها بر اساس مدل فازی سوگنوی مرتبه صفر هستند و راهکاری برای تنظیم برخط تالی قواعد ارائه می‌دهند [۳۱].

یادگیری سارسای فازی بر خلاف یادگیری کیوی فازی که برون سیاست است، یک الگوریتم بر سیاست می‌باشد و ساختار قواعد آن با استفاده از سیستم فازی سوگنوی مرتبه صفر به صورت زیر است:

$$\begin{aligned} &Rule_i: \\ &if\ x_1\ is\ L_{i1}\ and\ \dots\ and\ x_n\ is\ L_{in}, \\ &then\ (o_{i1}\ with\ value\ w^{i1})\ or\ \dots\ or \\ &(o_{im}\ with\ value\ w^{im}), i = 1,2,3,\dots,R \end{aligned} \quad (۱)$$

که در آن R، x_k و n به ترتیب بیان‌کننده تعداد قواعد، ورودی kام، و تعداد متغیرهای ورودی است. $L_i = L_{i1} \times \dots \times L_{in}$ شامل n مجموعه فازی محدب اکیدا نرمال با مرکزهای یکتا برای iامین قاعده است. در این رابطه، m تعداد عمل‌های گسسته ممکن برای هر قاعده، o_{ij} jامین عمل نامزد در قاعده iام و w^{ij} مقدار ارزش تقریب زده شده‌ی آن است. مقدار تالی هر قاعده در هر قدم زمانی با توجه به مقادیر w^{ij} انتخاب می‌شود [۲۹،۳۰]. هدف آموزش پیدا کردن بهترین تالی o_{ij} در هر قاعده است.

۲-۲ برنامه‌سازی پویا

در مبحث یادگیری تقویتی، اصطلاح برنامه‌سازی پویا به مجموعه‌ای از الگوریتم‌ها اتلاق می‌شود که می‌توانند برای محاسبه‌ی مقادیر ارزش و سیاست‌های بهینه در یک مسئله استفاده شوند. الگوریتم‌های برنامه‌سازی پویا به یک مدل کامل از محیط نیاز دارند. مدل محیط شامل مجموعه‌های حالت محیط $(\mathcal{S})$ و عمل $(A(s), s \in \mathcal{S})$ است و پویایی آن‌ها توسط مجموعه‌ای از احتمالات انتقال حالت-عمل به حالت بعدی $(P^a_{ss'})$ و امید پاداش‌های آنی حالت-عمل به حالت بعدی $(R^a_{ss'})$ به صورت زیر تعیین شده است:

$$\begin{aligned} P^a_{ss'} &= \Pr\{s_{t+1} = s' \mid s_t = s, a_t = a\} \\ R^a_{ss'} &= \mathrm{E}\{r_{t+1} \mid s_t = s, a_t = a, s_{t+1} = s'\} \end{aligned} \quad (۲)$$

این رابطه برای تمام $s \in \mathcal{S}$، $a \in A(s)$ و $s' \in \mathcal{S}^+$ ($\mathcal{S}^+$ همان مجموعه $\mathcal{S}$ است که حالت پایانی را نیز شامل می‌شود) برقرار است [۲۹،۳۰].

یک راهکار کلیدی در یادگیری تقویتی، استفاده از توابع ارزش برای پیدا کردن سیاست‌های مناسب است. منظور از سیاست تابع احتمال انتخاب عمل در هر حالت است که در تئوری کنترل آن را کنترلگر می‌نامیم. یکی از راه‌های یافتن توابع ارزش بهینه برای تمام حالت‌های محیط استفاده از روش تکرار ارزش در برنامه‌سازی پویا است:

$$V^*(s) = max_a \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V^*(s')] \quad (۳)$$

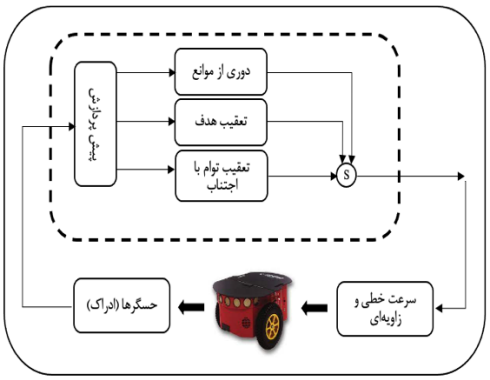
برای تمام $s \in \mathcal{S}$، $a \in A(s)$ و $s' \in \mathcal{S}^+$. در رابطه فوق، $0 \le \gamma \le 1$ فاکتور تخفیف است [۲۹].

با داشتن مقدار تابع ارزش بهینه می‌توان سیاست بهینه را بدست آورد. البته رابطه‌ی فوق برای فضای حالت و عمل گسسته است و در بخش‌های بعد روش ارائه شده برای تعمیم این روش در فضای پیوسته ارائه می‌گردد.

1 Fuzzy Q-Learning (FQL)

۳- چارچوب ارائه شده برای ربات تعقیب کننده‌ی هدف

بحث اصلی در این پژوهش کنترل ربات برای تعقیب هدف متحرک می‌باشد. اگر مسئله‌ی تعقیب هدف را به صورت زنجیره‌ای[1] در مسائل یادگیری تقویتی درنظر بگیریم، تجربه نشان داده که با دشواری‌هایی برای کنترل ربات روبرو خواهیم شد. بنابراین برای کاهش پیچیدگی کنترلگر، از معماری رفتارگرا[2] به عنوان یک معماری کنترلی مبتنی بر رفتار استفاده می‌کنیم. ایده‌ی بروکس در این معماری، ارائه‌ی رفتارهای پیچیده با ترکیب چندین رفتار ساده است [۳۲]. در این پژوهش نیز به دنبال شکستن رفتار کنترلی پیچیده به چندین رفتار ساده هستیم و طراحی کنترلگر رفتارگرا به صورت بلوک نقطه‌چین در شکل ۱ است.



شکل ۱: چارچوب ارائه شده مبتنی بر معماری رفتارگرا برای ربات تعقیب کننده‌ی هدف

ورودی و خروجی این بلوک کنترلگر به ترتیب مقادیر حسگرها و فرمان اعمالی به ربات (سرعت خطی و زاویه‌ای) است. در این کنترلگر سه رفتار «دوری از موانع»، «تعقیب هدف» و «تعقیب توام با اجتناب» در نظر گرفته شده است. با توجه به وضعیت ربات با هدف و موانع، یک از سه رفتار فراخوانی می‌شود، سپس بلوک انتخاب کننده[3] (S) خروجی را انتخاب و به عنوان خروجی بلوک کنترلگر ربات ارائه می‌دهد. پس از اعمال سرعت خطی و سرعت زاویه‌ای خروجی این بلوک به ربات، ادراک وضعیت فعلی توسط حسگرهای ربات انجام می‌شود و به عنوان ورودی به کنترلگر ربات ارسال می‌شود.

اولین رفتار به منظور اطمینان از عدم برخورد ربات به موانع تعبیه شده است و زمانی است که ربات نزدیک موانع است. در این شرایط، ربات باید از موانع اجتناب کند و در عین حال سعی کند به سمت موقعیت هدف حرکت کند. در رفتار «تعقیب هدف» هیچ مانعی در اطراف ربات وجود ندارد و هدف نیز در زاویه دید ربات قرار دارد، بنابراین ربات می‌تواند به سمت هدف بچرخد و سپس ربات مستقیم به سمت هدف حرکت کند. در سایر وضعیت‌های ربات با هدف و موانع نیز رفتار سوم (تعقیب توام با اجتناب) در نظر گرفته شده است. کنترلگر رفتار سوم یک سیستم فازی است که پس از مقداردهی اولیه با بکارگیری یادگیری تقویتی، تنظیم نرم[4] شده است.

۴- روش پیشنهادی برای تنظیم کنترلگر فازی

در این بخش روشی جدید برای تنظیم پارامترهای یک کنترلگر فازی در دو مرحله‌ی تنظیم سخت و تنظیم نرم ارائه می‌گردد. در مرحله‌ی اول هدف آن است که با داشتن داده‌های حسگرها و اقدام تصادفی عامل در محیط پارامترهای کنترلگر فازی مقداردهی اولیه شوند. بدین منظور تعمیمی از روش تکرار ارزش با استفاده از خوشه‌بندی ارائه می‌گردد و سپس با استفاده از نتایج و مشخصات خوشه‌ها و ارزش عمل‌های هر خوشه، یک سیستم فازی با ساختار متناظر با یادگیری سارسای فازی ساخته می‌شود. مشخصات توابع عضویت ورودی با استفاده از مشخصات خوشه‌ها تعیین می‌شود و عمل‌های کاندید و مقدار ارزش عمل‌های کاندید با توجه به مقدار ارزش‌های بدست آمده در روش تکرار ارزش تعمیم یافته حاصل می‌شود. در مرحله‌ی دوم از یادگیری سارسای فازی برای تنظیم نرم پارامترها بصورت بر خط بهره برده می‌شود. روش ارائه شده را یادگیری تقویتی فازی مبتنی بر تکرار ارزش[5] نامگذاری می‌کنیم.

با توجه به اینکه روش ارائه شده برای رفتار تعقیب توام با اجتناب در مسئله‌ی ربات تعقیب کننده‌ی هدف بکار گرفته شده است، توضیحات این بخش بر اساس بکارگیری در این مسئله است. بدین منظور ابتدا ربات در محیط قرار گرفته، ربات به صورت کاملا تصادفی در محیط حرکت داده می‌شود و مقاویر ویژگی‌های درنظر گرفته شده و فرمان‌های کنترلی جمع‌آوری می‌شود. ویژگی‌ها مهم در این مسئله، ۱- فاصله‌ی ربات با مانع سمت راست (RD)، ۲- فاصله‌ی ربات با مانع سمت چپ(LD)، ۳- فاصله‌ی ربات با مانع جلو (FD)، و ۴- زاویه‌ی سر ربات با هدف (Teta)، و فرمان‌های کنترلی اعمالی به ربات نیز، سرعت خطی و سرعت زاویه‌ای

1 Sequential
2 Subsumption architecture
3 Suppressor
4 Fine tune
5 Value Iteration based Fuzzy Reinforcement Learning: VIFRL

است. سپس باید حالات سیستم و سایر اطلاعات تکمیلی از این محیط با استفاده از داده‌های جمع‌آوری شده‌ی پیوسته استخراج شود. بدین منظور از روش تکرار ارزش فازی استفاده می‌شود. به بیانی دیگر، با استفاده از الگوریتم خوشه‌بندی Kmeans، داده‌های جمع‌آوری شده خوشه‌بندی می‌شود و هر خوشه نشان‌دهنده‌ی یک حالت از سیستم می‌باشد. در گام بعد لازم است با بکارگیری برنامه‌سازی پویا، ماتریس احتمال انتقال حالت–عمل به حالت بعدی و امید پاداش آنی حالت–عمل به حالت بعدی با استفاده از داده‌های جمع‌آوری شده استخراج شود. به طور مثال رابطه‌ی زیر برای محاسبه‌ی احتمال انتقال حالت–عمل حالت (خوشه‌ی) S به حالت S' پس از اعمال عمل a_1 است:

$$P_{s,s'}^{a_1} = \frac{\sum \#(s \overset{a_1}{\rightarrow} s')}{\sum \#(s \overset{a_1}{\rightarrow} i) \quad , i = 1,2,\dots,n} \qquad (۴)$$

در رابطه‌ی فوق $\#(s \overset{a_1}{\rightarrow} s')$ بیان کننده‌ی تعداد داده‌های جمع‌آوری شده‌ای است که حالت شروع آن‌ها حالت S است و پس از اعمال عمل a_1 به حالت S' منتقل می‌شوند. تعداد کل حالات مسئله نیز n در نظر گرفته شده است. امید پاداش آنی حالت–عمل حالت S به حالت S' پس از اعمال عمل a_1 نیز طبق رابطه‌ی زیر محاسبه می‌شود:

$$R_{s,s'}^{a_1} = \big(FD(s') - FD(s)\big) + \big((Teta(s') - Teta(s)) * 2\big) \qquad (۵)$$

به ترتیب FD و Teta نشان دهنده‌ی فاصله‌ی ربات تا مانع جلو و زاویه‌ی سر ربات با هدف است. بنابراین هرچه یک انتقال فاصله‌ی ربات تا مانع جلو و زاویه‌ی سر ربات با هدف را بیشتر کاهش دهد، پاداش بیشتری دارد. پس از محاسبه‌ی مقدار توابع ارزش بهینه برای تمام حالت‌های محیط با استفاده از رابطه‌ی ۳ و مقادیر محاسبه شده‌ی فوق، مقدار ارزش حالت–عمل برای تمام جفت حالت–عمل‌های مسئله با رابطه‌ی زیر محاسبه می‌شود:

$$Q^*(s,a) = \sum_{s'} P_{ss'}^{a}[R_{ss'}^{a} + \gamma V^*(s')] \qquad (۶)$$

در رابطه فوق، $0 \le \gamma \le 1$ فاکتور تخفیف است.

سیستم فازی برای کنترلگر ربات از نوع سیستم فازی سوگنوی مرتبه‌ی صفر[1] در نظر گرفته شده است و ساختار کلی قواعد آن طبق رابطه‌ی ۱ است. نواحی قواعد فازی[2] و سپس مقدم‌های قواعد کنترلگر فازی را می‌توان با استفاده از خوشه‌های خروجی روش تکرار ارزش فازی مشخص نمود. در تالی هر قاعده نیز عمل با بیشترین مقدار ارزش حالت–عمل محاسبه شده قرار می‌گیرد. رابطه‌ی زیر شمای کلی از قواعد کنترلگر فازی برای خوشه‌ی iام است:

$$\begin{aligned} &Rule_i: \\ &if LD\ is\ Func(LD_{C_i})\ and\ FD\ is\ Func(FD_{C_i}) \\ &\quad and\ RD\ is\ Func(RD_{C_i})\ and\ Teta\ is\ Func(Teta_{C_i}) \\ &\quad then\ LV = lv_i\ and\ AV = av_i\ with\ W. \end{aligned} \qquad (۷)$$

تابع Func روی هر چهار ویژگی تمام داده‌های هر خوشه اعمال می‌شود و خروجی آن دو مقدار میانگین و انحراف از معیار داده‌ها برای آن ویژگی است. به طور مثال $Func(LD_{C_i})$ بیان کننده‌ی میانگین و انحراف از معیار ویژگی LD برای تمام داده‌های موجود در خوشه‌ی iام است. عمل با بیشترین مقدار ارزش حالت–عمل (مقدار W) در تالی این قاعده، سرعت خطی lv_i و سرعت زاویه‌ای av_i است.

در گام بعدی لازم است تالی قواعد در سیستم فازی اولیه با یکی از الگوریتم‌های یادگیری تقویتی، تنظیم نرم شود. در این پژوهش از روش یادگیری سارسای فازی بدین منظور بهره برده شده است. بدین صورت که چهار عمل دیگر که اطراف عمل با بیشترین ارزش هستند، به همراه ارزش آن‌ها به تالی قواعد اضافه می‌شود. به عنوان مثال، قاعده‌ی موجود در رابطه‌ی ۷ برای شروع پروسه‌ی یادگیری به رابطه‌ی زیر تبدیل می‌شود:

$$\begin{aligned} &Rule_i: \\ &if LD\ is\ Func(LD_{C_i})\ and\ FD\ is\ Func(FD_{C_i}) \\ &\quad and\ RD\ is\ Func(RD_{C_i})\ and\ Teta\ is\ Func(Teta_{C_i}) \\ &\quad then\ \big(LV = lv_{i_1}\ and\ AV = av_{i_1}\ with\ w_{i_1}\big) \\ &\quad\ or\ \big(LV = lv_{i_2}\ and\ AV = av_{i_2}\ with\ w_{i_2}\big) \\ &\quad\ or\ \big(LV = lv_{i_3}\ and\ AV = av_{i_3}\ with\ w_{i_3}\big) \\ &\quad\ or\ \big(LV = lv_{i_4}\ and\ AV = av_{i_4}\ with\ w_{i_4}\big) \\ &\quad\ or\ (LV = lv_{i_5}\ and\ AV = av_{i_5}\ with\ w_{i_5}). \end{aligned} \qquad (۸)$$

ربات در محیط‌های آموزشی قرار می‌گیرد و مراحل زیر برای آموزش این کنترلگر با استفاده از یادگیری سارسای فازی به ترتیب زیر انجام می‌شود:

۱- انتخاب عمل در حالت فعلی: روش انتخاب عمل شبه حریصانه[3] است.

۲- اعمال عمل انتخابی و انتقال به حالت بعدی و بررسی حالت جدید.

۳- انتخاب عمل در حالت جدید.

۴- بروز رسانی مقادیر ارزش تالی قواعد.

1 Zero Order TSK
2 Fuzzy Rule Patch
3 ε-greedy

۵- اعمال عمل جدید، انتقال به حالت بعدی، بررسی حالت جدید، و رفتن به گام ۱.

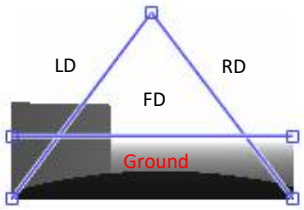
منظور از بررسی حالت جدید در مراحل فوق، حالت رسیدن به هدف یا برخورد با مانع است. در صورتیکه ربات در یک محیط در یکی از این دو حالت قرار گیرد، مقادیر ارزش تالی قواعد به روز رسانی شده و پروسه آموزش متوقف می‌شود. به منظور به روز رسانی مقادیر ارزش، سیگنال تقویتی دریافتی ربات به شرح زیر تعریف شده است:

$$r = \begin{cases} 5 & , \ Goal \\ -5 & , \ Collision \\ -0.01 & , \ Else \end{cases} \quad (۹)$$

پس از اتمام آموزش، کنترلگر نهایی به عنوان رفتار تعقیب توام با اجتناب در چارچوب ارائه شده در بخش ۳ مورد استفاده قرار می‌گیرد.

۵- آزمایش‌ها

با هدف بررسی عملکرد، روش پیشنهادی بر روی ربات تعقیب کننده‌ی هدف و با استفاده از نرم‌افزار شبیه‌ساز ویباتس[1]، پیاده‌سازی شد. ربات استفاده شده، Pioneer 3DX دو چرخ مجهز به دوربین استریوی ZED به عنوان حسگر است. فاصله‌ی ربات تا موانع در هر سه جهت با استفاده از تصاویر عمق دوربین ZED و روش بلوک‌بندی معنایی که قبلا توسط نویسندگان [۱۸] ارائه شده است، محاسبه می‌شود. هدف روش بلوک‌بندی معنایی، تشخیص بهتر موانع در هر سه جهت ربات است. نحوه‌ی بلوک‌بندی در این روش در شکل زیر نشان داده شده است:



شکل ۲: بلوک‌بندی معنایی برای تشخیص بهتر موانع سمت راست، چپ و جلوی ربات [۱۸]

هر تصویر عمق دوربین به چهار ناحیه‌ی LD، FD، RD، و Ground تقسیم می‌شود و فاصله‌ی ربات از موانع در سه ناحیه، برابر با حداقل فاصله‌ی مانع تا ربات در هر بلوک است.

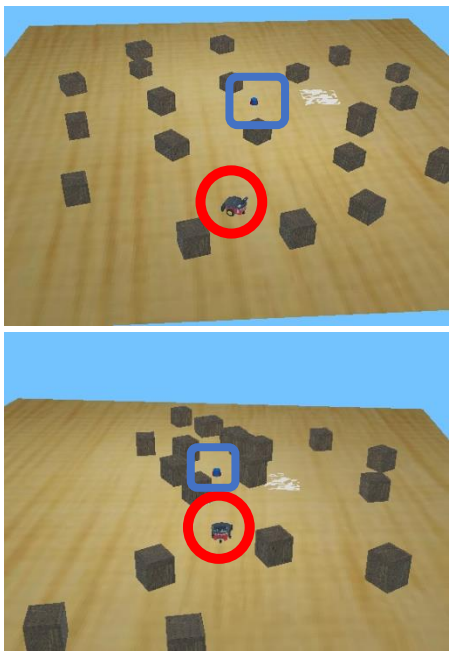
با توجه به چارچوب ارائه شده برای ربات در بخش ۳، سه رفتار دوری از موانع، تعقیب هدف و تعقیب توام با اجتناب وجود دارد که در رابطه‌ی ۱۰ به ترتیب با عنوان O. A.، G. F.، و G. F. with O. A. است. شرایط هر رفتار در رابطه‌ی زیر نشان داده شده است:

$$\begin{cases} min(LD, FD, RD) <= 0.75 & , O.A. \\ (min(LD, FD, RD) \geq Goal\ distance) & \\ \& (-30 <= Goal\ teta <= 30) & , G.F. \\ others & , G.F.\ with\ O.A. \end{cases} \quad (۱۰)$$

همانطور که در رابطه‌ی فوق مشخص است، اگر کمترین فاصله‌ی ربات تا موانع سمت راست (RD)، چپ (LD) و جلو (FD) کمتر از ۷۵ سانتی‌متر شود، رفتار دوری از موانع فعال می‌شود و با توجه به زاویه‌ی سر ربات با هدف، ربات مانع را دور می‌زند. بدین صورت که هر بار، سرعت زاویه‌ای اعمالی به ربات به اندازه زاویه‌ی سر ربات با هدف است و سرعت خطی آن ۲ واحد کاهش داده می‌شود. زمانی رفتار تعقیب هدف فراخوانی می‌شود که حداقل فاصله‌ی ربات تا موانع بیشتر از فاصله‌ی ربات تا هدف باشد و هدف در زاویه‌ی دید ربات قرار داشته باشد (به بیانی دیگر، زاویه‌ی سر ربات با هدف بین ۳۰- و ۳۰ درجه باشد). در این رفتار، سرعت زاویه‌ای به اندازه زاویه‌ی سر ربات با هدف است و سرعت خطی با توجه به فاصله‌ی ربات با هدف تنظیم می‌شود. در غیر اینصورت رفتار تعقیب توام با اجتناب فراخوانی می‌شود و خروجی این رفتار با کنترلگر فازی تولید شده با روش یادگیری تقویتی فازی مبتنی بر تکرار ارزش تولید می‌شود. در این شبیه‌ساز دو محیط برای حرکت ربات و جمع‌آوری داده ایجاد شد که در شکل ۳ نشان داده شده است.

در محیط‌های آموزشی، ربات تعقیب کننده‌ی هدف در دایره‌ی قرمز رنگ و ربات هدف در دایره‌ی آبی رنگ است و جعبه‌ها به عنوان موانع هستند. برای جمع‌آوری داده، ربات کاملا تصادفی در محیط حرکت داده می‌شود. فرمان‌های اعمالی به ربات، سرعت خطی و سرعت زاویه‌ای است. سرعت خطی ربات یکی از مقادیر بازه‌ی $[۰, ۲, ۴, ۶, ۸, ۱۰, ۱۲]$ و سرعت زاویه‌ای آن، یکی از مقادیر بازه‌ی $[۲۰, -۱۰, ۰, ۱۰, ۲۰-]$ است. سرعت خطی با توجه به فاصله‌ی ربات تا هدف و موانع، و سرعت زاویه‌ای ربات با استفاده از دسته کنترل توسط هدایتگر ربات مشخص و سپس به ربات اعمال می‌شود. تعداد کل داده‌های جمع‌آوری شده در این فاز، ۵۶۱۹ است و اطلاعات جمع‌آوری شده، ۱- فاصله‌ی ربات با مانع سمت راست، ۲- فاصله‌ی ربات با مانع سمت چپ، ۳- فاصله‌ی ربات با مانع جلو، ۴- زاویه‌ی سر ربات با هدف، و ۵- سرعت خطی و زاویه‌ای اعمالی به ربات می‌باشد. چهار داده‌ی جمع‌آوری شده‌ی اول در هر حرکت (LD، FD، RD، و Teta)، ویژگی‌های مورد استفاده در حالات سیستم است.

1 Webots



شکل ۳: محیط‌ها برای حرکت ربات و جمع‌آوری داده

برای استخراج حالت‌های سیستم با استفاده از مقادیر این ویژگی‌ها، از الگوریتم خوشه‌بندی Kmeans بهره برده شد. در این روشِ خوشه‌بندی عدد تعداد خوشه‌ها بسیار مهم است بنابراین از معیارهای ارزیابی درونی[1] برای بررسی نتایج خوشه‌بندی با مقادیر متفاوت k استفاده شد. از معیارهای ارزیابی کالینسکی-هارباز[2]، سیهوتی[3]، دیویس-بولدین[4]، و جمع مربعات خطا بدین منظور بهره برده شد و طبق نتایج، مقدار ۶ به عنوان متغیر k در روش خوشه‌بندی Kmeans تعیین شد. بنابراین تمام داده‌های جمع‌آوری شده به ۶ خوشه (۶ حالت سیستم) تقسیم‌بندی شدند. گام بعد محاسبه‌ی مقادیر ماتریس‌های احتمال انتقال حالت-عمل به حالت بعدی و امید پاداش آنی حالت-عمل به حالت بعدی برای تمام حالات با استفاده از برنامه‌سازی پویا و روابط ۲ است. با استفاده از رابطه‌ی ۳ مقدار ارزش هر حالت محاسبه شده، سپس عمل با بیشترین ارزش در هر حالت بدست می‌آید. بدین ترتیب کنترلگر فازی با ۶ قاعده مقداردهی اولیه می‌شود که هر قاعده متناظر با داده‌های یک خوشه است. به عنوان مثال، قاعده‌ی اول کنترلگر فازی اولیه به شکل زیر است:

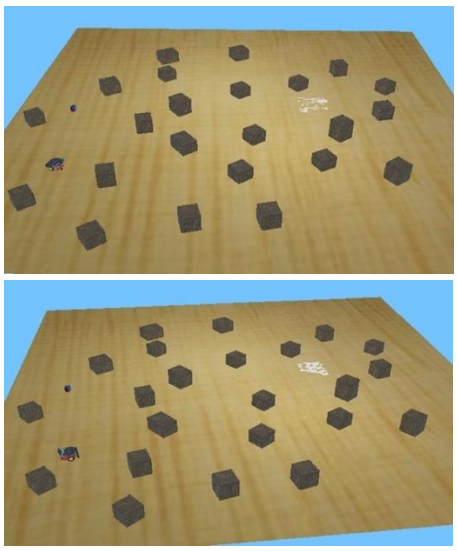
$$\begin{aligned} &if LD\ is\ (0.3,0.1)\ and\ FD\ is\ (0.4,0.2)\\ &and\ RD\ is\ (0.9,0.2)\ and\ Teta\ is\ (-0.5,0.3)\\ &then\ LV=6\ and\ AV=-20\ with\ 3.19. \end{aligned} \quad (۱۱)$$

متغیرهای LD، FD، RD، و Teta در مقدم قاعده‌ی فوق، به ترتیب فاصله‌ی ربات از مانع سمت چپ، جلو، سمت راست و زاویه‌ی سر ربات با هدف را نشان می‌دهد. همانطور که در بخش قبل توضیح داده شد، خروجی تابع Func در قواعد کنترلگر فازی استفاده می‌شود که به ترتیب مولفه‌های میانگین و انحراف از معیار برای تمام داده‌های موجود در یک خوشه است. در مثال فوق، $(0.3, 0.1)$ به ترتیب میانگین و انحراف از معیار ویژگی فاصله‌ی ربات از موانع سمت چپ برای تمام داده‌های موجود در خوشه‌ی شماره‌ی ۱ است. سرعت خطی (LV) و سرعت زاویه‌ای (AV) با بیشترین ارزش نیز در تالی قاعده قرار می‌گیرد.

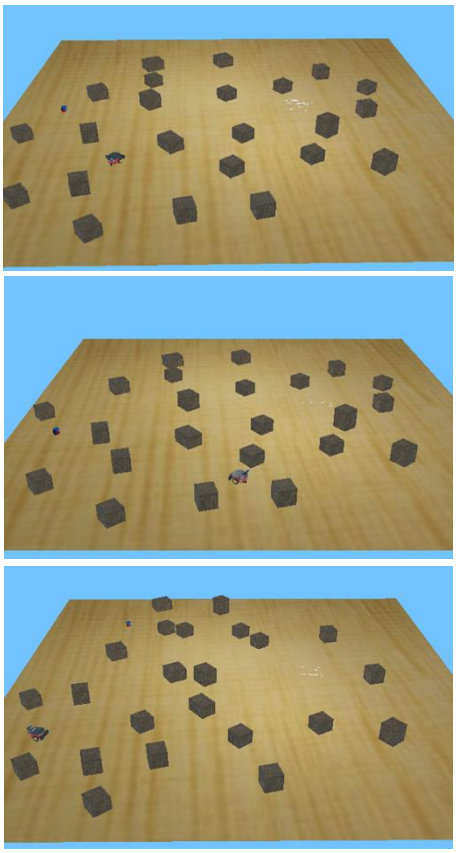
گام بعدی تنظیم نرم تالی قواعد در کنترلگر فازی اولیه است. بدین منظور چهار عمل دیگر که اطراف عمل با بیشترین ارزش هستند، به همراه ارزش آن‌ها به تالی قواعد اضافه می‌شود. به عنوان مثال، قاعده‌ی اول کنترلگر فازی برای شروع پروسه‌ی یادگیری به صورت زیر است:

$$\begin{aligned} &if LD\ is\ (0.3,0.1)\ and\ FD\ is\ (0.4,0.2)\\ &and\ RD\ is\ (0.9,0.2)\ and\ Teta\ is\ (-0.5,0.3)\\ &then\ (LV=6\ and\ AV=-20\ with\ 3.19)\\ &\quad or\ (LV=8\ and\ AV=-20\ with\ 3.08)\\ &\quad or\ (LV=4\ and\ AV=-20\ with\ 2.73)\\ &\quad or\ (LV=6\ and\ AV=-10\ with\ 3.08)\\ &\quad or\ (LV=8\ and\ AV=-10\ with\ 2.87). \end{aligned} \quad (۱۲)$$

پس از آماده سازی تمام قواعد، مراحل آموزش کنترلگر فازی ربات تعقیب کننده‌ی هدف با استفاده از روش یادگیری ارائه شده، انجام می‌شود. بدین منظور ۵ محیط آموزشی زیر در نظر گرفته شده است:



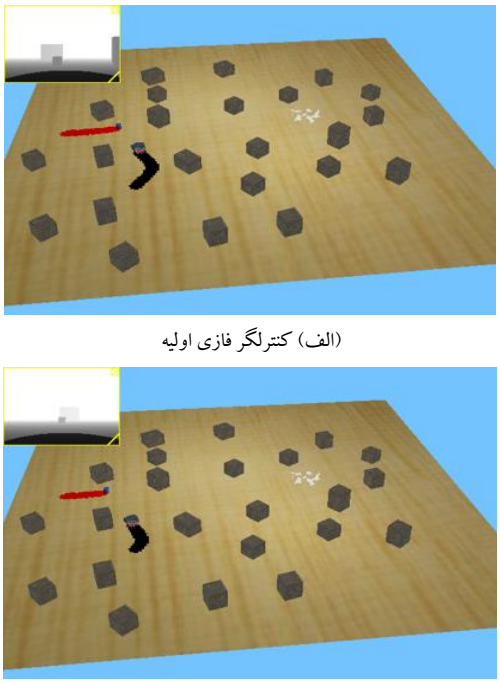
1 Internal Criteria Index
2 Calinski-Harabasz
3 Silhouette
4 Davies-Bouldin



شکل ۴: محیط‌های آماده شده برای آموزش کنترلگر فازی

در صورتیکه ربات بتواند در یک دور از آموزش، در هر ۵ محیط به هدف برسد، فرآیند آموزش کنترلگر و به روز رسانی ارزش تالی قواعد متوقف می‌شود. در این آزمایش‌ها، پس از ۵ دور حرکت ربات در هر ۵ محیط (جمعا ۲۵ اجرا) ربات توانست در تمام محیط‌ها به هدف برسد و آموزش کنترلگر متوقف شد. عمل با بیشترین ارزش برای قاعده‌ی اول از کنترلگر پس از پایان آموزش، مقدار ۸ و ۲۰- به عنوان سرعت خطی و زاویه‌ای است.

به منظور بررسی کارآیی کنترلگر پیشنهادی برای رفتار تعقیب توام با اجتناب در چارچوب ارائه شده، دو محیط متفاوت زیر در شبیه‌ساز آماده و دو روش کنترلگر فازی اولیه و کنترلگر فازی آموزش داده شده برای رسیدن ربات به هدف اجرا شد. چارچوب هر دو کنترلگر یکسان است و تنها کنترلگر استفاده شده در رفتار تعقیب توام با اجتناب برای آن‌ها متفاوت است، بنابراین در این آزمایش‌ها تنها این رفتار مورد بررسی قرار گرفته است. به بیانی دیگر پس از خروج از این رفتار و فراخوانی رفتار بعدی (تعقیب هدف یا دوری از موانع) بررسی متوقف می‌شود. خروجی هر دو روش مورد بررسی در محیط شماره‌ی ۱ در شکل زیر نشان داده است:



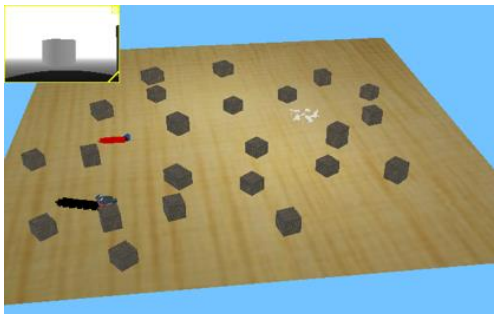
(الف) کنترلگر فازی اولیه

(ب) کنترلگر فازی آموزش داده شده

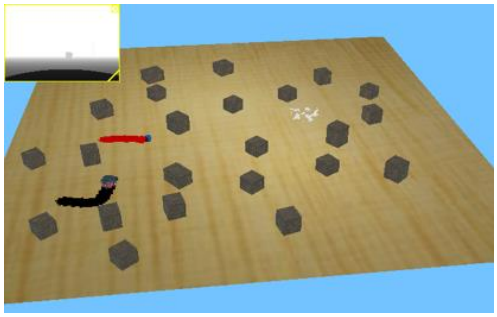
شکل ۵: اجرای دو روش مورد بررسی در محیط شماره‌ی ۱

همانطور که در شکل الف مشخص است، ربات با کنترلگر فازی اولیه نیز توانست خود را به هدف برساند، اما همچنان که انتظار می‌رفت روش کنترلگر فازی آموزش داده شده با الگوریتم VIFRL مسیر کوتاه‌تری برای رسیدن به هدف را طی نمود.

محیط شماره‌ی ۲ نیز برای بررسی عملکرد دو کنترلگر در نظر گرفته شد که خروجی هر روش در شکل ۶ نشان داده شده است. طبق نتایج این آزمایش، در کنترلگر فازی اولیه ربات به مانع برخورد می‌کند، در صورتیکه کنترلگر فازی آموزش داده شده ربات را به خوبی به هدف می‌رساند. این محیط نمونه‌ای از عملکرد بهتر کنترلگر پس از آموزش و تنظیم نرم را نشان می‌دهد.



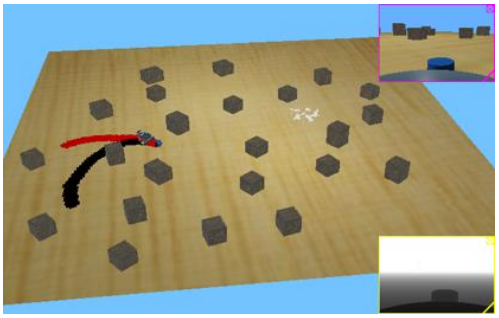
(الف) کنترلگر فازی اولیه



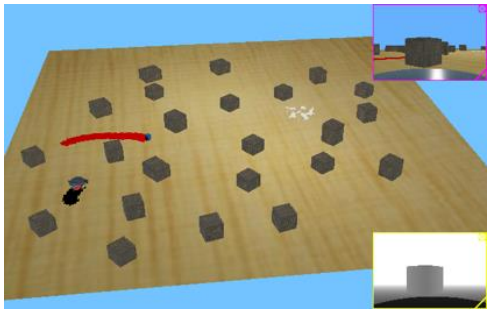
(ب) کنترلگر فازی آموزش داده شده

شکل ۶: اجرای دو روش مورد بررسی در محیط شماره‌ی ۲

حال با آزمایش زیر، کنترلگر فازی ارائه شده با کنترلگر فازی ارائه شده‌ی توآن[1] و همکارانش [۲۴] در ربات تعقیب کننده‌ی هدف بررسی می‌شود. پژوهش ارائه شده توسط توآن و همکارانش جدیدترین پژوهش در زمینه‌ی ربات‌های تعقیب کننده‌ی هدف است که در آن یک کنترلگر فازی (جعبه سفید) برای هدایت ربات پیشنهاد شده است. کنترلگر فازی توآن و همکارانش دو ورودی دارد: ۱- حداقل فاصله‌ی ربات تا نزدیکترین مانع یا هدف (اگر مانع به ربات نزدیکتر باشد مقدار این ورودی، فاصله تا نزدیکترین مانع و درغیراینصورت فاصله‌ی ربات تا هدف است.) و ۲- سرعت خطی فعلی ربات. خروجی این کنترلگر سرعت خطی پیشنهادی برای ربات است. سرعت زاویه‌ای ربات مستقل از کنترلگر فازی و با استفاده از روش پنجره‌گذاری اطراف ربات محاسبه می‌شود. در این روش، پنجره‌هایی با زاویه‌ی $-90 \leq \alpha \leq 90$, α اطراف ربات قرار می‌گیرد و میزان هم‌پوشانی هر پنجره با موانع محاسبه می‌شود. از بین تمام پنجره‌ها، پنجره‌هایی که میزان هم‌پوشانی آن‌ها با موانع از یک حد آستانه کمتر باشد، انتخاب می‌شوند و از بین آن‌ها سرعت زاویه‌ای اعمالی به ربات، زاویه‌ی α پنجره‌ای است که خط میانی آن کمترین فاصله با ربات را دارد. مهم‌ترین چالش در یافتن سرعت زاویه‌ای با این روش، میزان گسسته سازی زاویه‌ی پنجره‌ها (α) است تا هم حجم محاسبات زیادی لازم نباشد و هم پاسخ دقیق محاسبه شود. تعیین میزان بهینه‌ی حد آستانه برای انتخاب پنجره‌های اطراف ربات نیز مهم است. خروجی هر دو روش مورد بررسی در محیط یکسان در شکل ۷ نشان داده شده است. همانطور که در شکل ۷ قسمت الف مشخص است کنترلگر ارائه شده به خوبی هدف را دنبال می‌کند اما کنترلگر توآن و همکارانش [۲۴] به دلیل سرعت خطی پایین از ربات هدف جا می‌ماند.



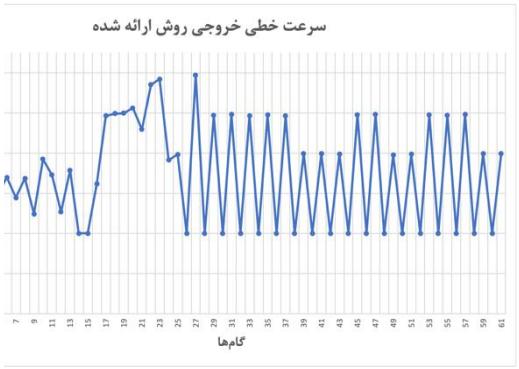
(الف) کنترلگر فازی ارائه شده



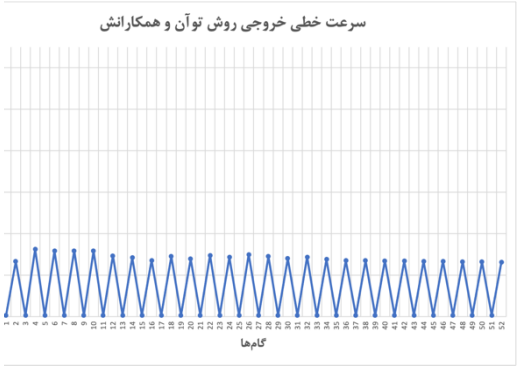
(ب) کنترلگر فازی ارائه شده توسط توآن و همکارانش

شکل ۷: بررسی عملکرد کنترلگر ارائه شده و کنترلگر توآن و همکارانش

در روش توآن و همکارانش فقط سرعت خطی با استفاده از سیستم فازی محاسبه می‌شود، بنابراین سرعت‌های خطی اعمالی به ربات در هر دو روش در آزمایش فوق در نمودار زیر نشان داده شده است:



(الف) کنترلگر فازی ارائه شده



(ب) کنترلگر فازی ارائه شده توسط توآن و همکارانش

1 Toan

شکل ۸: سرعت خطی خروجی روش ارائه شده و روش توآن و همکارانش

در نهایت، طبق بررسی‌ها و آزمایشات اجرا شده، کنترلگر فازی ارائه شده عملکرد بهتری نسبت به دو کنترلگر دیگر دارد. فیلم کاملی از تعقیب هدف با استفاده از چارچوب ارائه شده در این پژوهش و کنترلگر فازی VIFRL در صفحه‌ی آزمایشگاه هوش محاسباتی و رباتیک دانشگاه یزد[1] قابل دسترسی است.

۶- نتیجه‌گیری و پیشنهادها

در این پژوهش، در جهت غلبه بر دو ضعف سرعت پایین آموزش و تعداد بالای شکست‌ها در روش‌های یادگیری تقویتی فازی، یک روش جدید برای مقداردهی اولیه‌ی پارامترهای کنترلگر فازی با استفاده از داده‌های بدست آمده از تعامل عامل با محیط ارائه شد. نشان داده شد که چگونه با بهره‌گیری از مفهوم درجه تعلق در مجموعه‌های فازی می‌توان تعمیمی از روش تکرار ارزش گسسته به فضای پیوسته را بدست آورد و از مقدارهای ارزش بدست آمده و مشخصات خوشه‌های تولید شده در مرحله تعمیم، پارامترهای یک کنترلگر فازی را مقداردهی اولیه کرد. روش ارائه شده که "یادگیری تقویتی فازی مبتنی بر تکرار ارزش (VIFRL)" نام گرفت، برای تنظیم نرم (نهایی) پارامترها از الگوریتم یادگیری سارسای فازی بهره می‌برد. مسئله‌ی مورد مطالعه در این مقاله مسئله‌ی ربات تعقیب کننده‌ی هدف بود. در این مقاله چارچوب مبتنی بر معماری رفتارگرا برای این مسئله ارائه شد. دیده شد که تقسیم رفتارها می‌تواند باعث سادگی در عملکرد شود، چرا که برخی از رفتارهای ساده را می‌توان بدون نیاز به آموزش اجرا نمود و تنها یکی از رفتارها در چارچوب ارائه شده آموزش داده شد. نتایج شبیه‌سازی حاکی از برتری روش ارائه شده نسبت به کنترلگر فازی ارائه شده‌ی توآن و همکارانش [۲۴]، بود. لذا می‌توان نتیجه گرفت که روش ارائه شده تنها با داشتن داده‌های حرکت تصادفی ربات در محیط می‌تواند به خوبی پارمترهای کنترلگر ربات تعقیب کننده‌ی هدف را مقداردهی نماید. همچنین می‌توان نتیجه‌گیری کرد که روش‌هایی که سعی در تعیین ارزش پارامترها دارند، در مقابل روش‌های مبتنی بر مشتق که یک مقدار را برای پارامترها در جهت کمینه کردن مجموع مربعات خطا پیشنهاد می‌دهند، کاوش مناسب‌تری داشته و شانس بالاتری در پیدا کردن جواب بهینه دارند.

از آنجا که روش ارائه شده برای تنظیم پارامترهای کنترلگر می‌تواند بر روی هر ساختار یادگیری تقویتی نقاد-تنها استفاده شود، لذا می‌توان به عنوان کار آینده آن را در این دسته از الگوریتم‌ها بکار برد. فعالیت تحقیقاتی دیگر، توسعه‌ی ایده‌ها برای داشتن عمل‌های پیوسته در مرحله جمع آوری داده است. همچنین می‌توان پیش‌بینی موقعیت بعدی هدف نیز به عنوان یک ورودی به کنترلگر اضافه نمود. در ادامه ایده های مذکور بر روی یک ربات واقعی مورد آزمایش قرار می‌گیرد.

مراجع

[1] M. F. R. Lee, & Y. C. Chen, "Artificial Intelligence Based Object Detection and Tracking for a Small Underwater Robot". Processes, vol. 11, no. 2, pp. 312, 2023.

[2] S. Li, K. Milligan & et al., "Exploring the role of human-following robots in supporting the mobility and wellbeing of older people". Scientific Reports, vol. 13, no. 1, pp. 6512, 2023.

[3] G. Thomas, R. Gade, T. B. Moeslund & et al., "Computer vision for sports: Current applications and research topics". Computer Vision and Image Understanding, vol. 159, pp. 3-18, 2017.

[4] H. Kivrak, F. Cakmak, H. Kose & S. Yavuz, "Social navigation framework for assistive robots in human inhabited unknown environments". The International Journal Engineering Science and Technology, vol. 24, no. 2, pp. 284-298, 2021.

[5] Tempo Walk in Clubcar. Available online: https://www.clubcar.com/en-us/golf-operations/fleet-golf/tempo-walk (accessed on 14 November 2023) .

[6] A. Rudenko, L. Palmieri & et al., "Human motion trajectory prediction: A survey". The International Journal of Robotics Research, vol. 39, no. 8, pp. 895-935, 2020.

[7] M. J. Islam, J. Hong & J. Sattar, "Person-following by autonomous robots: A categorical overview". The International Journal of Robotics Research, vol. 38, no. 14, pp. 1581-1618, 2019.

[8] R. Algabri & M. T. Choi, "Deep-learning-based indoor human following of mobile robot using color feature". Sensors, vol. 20, no. 9, pp. 2699, 2020.

[9] D. Cha & W. Chung, "Human-leg detection in 3D feature space for a person-following mobile robot using 2D LiDARs. International Journal of Precision Engineering and Manufacturing", vol. 21, pp. 1299-1307, 2020.

[10] A. Eirale, M. Martini, & M. Chiaberge, "Human-centered navigation and person-following with omnidirectional robot for indoor assistance and

1 https://aparat.com/v/YnsqD

[22] J. Lin, J. Zhou, M. Lu, H. Wang & A. Yi, "Design of robust adaptive fuzzy controller for a class of single-input single-output (siso) uncertain nonlinear systems". Mathematical Problems in Engineering, pp. 1-11, 2020.

[23] T. V. Nguyen, M. H. Do & J. Jo, "Robust-adaptive-behavior strategy for human-following robots in unknown environments based on fuzzy inference mechanism". Industrial Robot: the international journal of robotics research and application, vol. 49, no. 6, pp. 1089-1100, 2022.

[24] N. Van Toan, M. Do Hoang, P. B. Khoi & S. Y. Yi, "The human-following strategy for mobile robots in mixed environments". Robotics and Autonomous Systems, vol. 160, 2023.

[25] V. Derhami, V. J. Majd & M. N. Ahmadabadi, "Fuzzy Sarsa learning and the proof of existence of its stationary points". Asian Journal of Control, vol. 10, no. 5, pp. 535-549, 2008.

[26] F. Fathinezhad, V. Derhami & M. Rezaeian, "Supervised fuzzy reinforcement learning for robot navigation". Applied Soft Computing, vol. 40, pp. 33-41, 2016.

[27] D. Song, B. Zhu, J. Zhao & et al., (2023). "Personalized Car-Following Control Based on a Hybrid of Reinforcement Learning and Supervised Learning". IEEE Transactions on Intelligent Transportation Systems, 2023.

[28] F. Nadi, V. Derhami & F. Alamiyan-Harandi, "Coarse Tuning of Fuzzy Reinforcement Learning Architecture using Value Iteration Method". Fuzzy Systems and its Applications, vol. 6, no. 1, pp. 109-126, 2023.

[29] V. Derhami, F. Alamiyan-Harandi, & M. Dowlatshahi, "Reinforcement Learning" (In Persian), Yazd University press, 2017.

[30] R. S. Sutton and A. G. Barto. "Reinforcement learning: An introduction". MIT press Cambridge, 1998.

[31] F. Alamiyan-Harandi, V. Derhami, "A reinforcement learning algorithm for adjusting antecedent parameters and weights of fuzzy rules in a fuzzy classifier", Journal of Intelligent & Fuzzy Systems, vol. 30, no. 4, pp. 2339-2347, 2016.

[32] R. A. Brooks, "A robust layered control system for a mobile robot", IEEE Journal of Robotics and Automation 2, pp. 14-23, 1986.

monitoring". Robotics, vol. 11, no. 5, pp. 108, 2022.

[11] J. Liu, X. Chen & et al., "A person-following method based on monocular camera for quadruped robots". Biomimetic Intelligence and Robotics, vol. 2, no. 3, 2022.

[12] K. Koide, J. Miura & E. Menegatti, "Monocular person tracking and identification with on-line deep feature selection for person following robots". Robotics and Autonomous Systems, vol. 124, 2020.

[13] F. Alamiyan-Harandi, V. Derhami, & F. Jamshidi, "A new feature selection method based on task environments for controlling robots". Applied Soft Computing, vol. 85, 2019.

[14] C. A. Yang & K. T. Song, "Control design for robotic human-following and obstacle avoidance using an RGB-D camera". 19th IEEE International Conference on Control, Automation and Systems (ICCAS), pp. 934-939, 2019.

[15] B. J. Lee, J. Choi, C. Baek & B. T. Zhang, "Robust human following by deep Bayesian trajectory prediction for home service robots". IEEE international conference on robotics and automation (ICRA), pp. 7189-7195, 2018.

[16] B. X. Chen, R. Sahdev & J. K. Tsotsos, "Integrating stereo vision with a CNN tracker for a person-following robot". 11th International Conference on Computer Vision Systems, Springer International Publishing, pp. 300-313, 2017.

[17] B. X. Chen, "Real-time Online Human Tracking with a Stereo Camera for Person-Following Robots", 2019.

[18] F. Nadi, F. Alamiyan-Harandi, V. Derhami, F. Taherizade, "Improving Performance of Target Following Robot using Visual Servoing Fuzzy Controller (In Persian), 3rd International Conference on Soft Computing, 2019.

[19] J. H. Choi, K. Samuel, K. Nam & S. Oh, "An autonomous human following caddie robot with high-level driving functions". Electronics, vol. 9, no. 9, pp. 1516, 2020.

[20] X. Gu, J. Han, Q. Shen & P. P. Angelov, "Autonomous learning for fuzzy systems: a review". Artificial Intelligence Review, vol. 56, no. 8, pp. 7549-7595, 2023.

[21] H. Hu, X. Wang & L. Chen, "Impedance with finite-time control scheme for robot-environment interaction". Mathematical Problems in Engineering, 2020.