Computational Intelligence in Electrical Engineering
Vol. 16, No. 02, 2025
pp. 1- 18
Research Paper

Localization and Tracking of a Moving Signal Source Using a UAV and Reinforcement Learning

Fatemeh Saeidnejad [1], Mahdi Majidi[*2]

[1] Department of Electrical and Computer Engineering, University of Kashan, Kashan, Iran
[2] Department of Electrical and Computer Engineering, University of Kashan, Kashan, Iran

Abstract:
Accurate localization is of significant importance in various applications, such as search and rescue operations. UAVs are regarded as a suitable solution for this purpose due to their agility and greater line-of-sight capabilities. In this research, a UAV is responsible for two-dimensional localizing the ground signal source by collecting received signal strength measurements. As the UAV moves through the environment, it creates virtual reference nodes and estimates the source location in the continuous space using the least squares method. The UAV's trajectory is optimized using the Q-learning algorithm, a reinforcement learning method. This algorithm navigates the UAV within a search space divided into discrete cells. The main innovation of this research is the presentation of a hybrid algorithm that simultaneously integrates the least squares, Kalman filter, and Q-learning algorithms. In this framework, following the estimation of the current position, the Kalman filter predicts the future position of the moving source. Subsequently, the predicted position is discretized, and Q-learning determines the UAV's optimal trajectory. The results indicate that this triple combination provides accurate source localization and the design of an optimal trajectory for the UAV, simultaneously improving localization accuracy and navigation efficiency. Additionally, the integration of the Kalman filter into the proposed framework enhances the tracking accuracy of the moving signal source and leads to a 42% reduction in the mean squared error.

Keywords: UAV, Tracking, Received Signal Strength, Localization, Reinforcement Learning, Q-learning.

https://doi.org/10.22108/isee.2025.145585.1745

مقاله پژوهشی

مکان‌یابی و ردیابی منبع سیگنال متحرک به کمک یک پهپاد و یادگیری تقویتی

فاطمه سعیدنژاد [۱]، مهدی مجیدی [۲*]

۱- کارشناسی ارشد، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه کاشان، کاشان، ایران
f.saeidnejad@grad.kashanu.ac.ir

۲- استادیار، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه کاشان، کاشان، ایران
m.majidi@kashanu.ac.ir

چکیده: در کاربردهایی مختلف مانند عملیات جست‌وجو و نجات، مکان‌یابی دقیق اهمیت زیادی دارد. پهپادها به دلیل چابکی و امکان دید مستقیم بیشتر، گزینه‌ای مناسب برای این هدف هستند. در این پژوهش، یک پهپاد با جمع‌آوری اندازه‌گیری‌های قدرت سیگنال دریافتی، مسئول مکان‌یابی دوبُعدی منبع سیگنال زمینی است. پهپاد با حرکت در محیط، گره‌های مرجع مجازی ایجاد می‌کند و با روش حداقل مربعات، موقعیت منبع را در فضای پیوسته تخمین می‌زند. مسیریابی پهپاد با الگوریتم یادگیری Q بهینه‌سازی شده است که یکی از روش‌های یادگیری تقویتی محسوب می‌شود. این الگوریتم پهپاد را در فضای جست‌وجوی تقسیم‌شده به سلول‌های گسسته هدایت می‌کند. نوآوری اصلی پژوهش ارائهٔ الگوریتمی ترکیبی است که الگوریتم‌های حداقل مربعات، فیلتر کالمن و یادگیری Q را هم‌زمان تلفیق می‌کند. در این چارچوب، پس از تخمین موقعیت فعلی، فیلتر کالمن موقعیت آینده منبع متحرک را پیش‌بینی می‌کند و پس از گسسته‌سازی موقعیت پیش‌بینی‌شده، یادگیری Q مسیر بهینه پهپاد را تعیین می‌کند. نتایج نشان می‌دهد این ترکیب سه‌گانه تخمین دقیق موقعیت منبع را همراه با طراحی مسیر بهینه فراهم می‌کند و موجب ارتقای هم‌زمان دقت مکان‌یابی و کارایی ناوبری می‌شود. همچنین، به‌کارگیری فیلتر کالمن دقت ردیابی منبع سیگنال متحرک را افزایش می‌دهد و کاهش ٤۲ درصدی خطای میانگین مربعات را به همراه دارد.

واژه‌های کلیدی: پهپاد، ردیابی، قدرت سیگنال دریافتی، مکان‌یابی، یادگیری تقویتی، یادگیری Q.

۱- مقدمه

مکان‌یابی و ردیابی منابع سیگنال مسائلی هستند که در کاربردهای تجاری و نظامی مانند کنترل ترافیک هوایی، سنجش از راه دور، نظارت، امداد و نجات استفاده می‌شوند. باتوجه به اینکه دقت و دسترس‌پذیری سیستم‌های موقعیت‌یابی ماهواره‌ای ممکن است در برخی از محیط‌ها، مانند محیط‌های شهری یا داخل ساختمان، به دلیل تضعیف یا انسداد سیگنال به طرزی جالب توجه کاهش یابد، استفاده از فناوری‌ها و روش‌های مکان‌یابی دیگر نیز مورد توجه قرارگرفته است [۱].

بهره‌گیری از ویژگی‌های منابع سیگنال رویکردی مؤثر برای مکان‌یابی در شرایطی است که استفاده از سیستم موقعیت‌یابی جهانی (GPS)[۱] با محدودیت مواجه است. از طرفی، پهپادها به دلیل چابکی، انعطاف‌پذیری در پرواز، دسترسی به نقاط سخت‌گذر و امکان تجهیز به انواع حسگرها و آنتن‌ها، گزینه‌ای ایده‌آل برای موقعیت‌یابی اهداف هستند. بهره‌گیری از پهپادی که مسیر حرکت آن بهینه‌سازی شده و قادر به تخمین بلادرنگ موقعیت منبع متحرک است، راهکاری هوشمند برای افزایش دقت مکان‌یابی و کاهش وابستگی به سامانه‌های ماهواره‌ای محسوب می‌شود. اهمیت این مسئله در کاربردهایی حساس نمایان می‌شود که اطلاعات اولیه محدود است و به واکنش سریع نیاز دارند. در این شرایط، دقت و سرعت در موقعیت‌یابی نقشی تعیین‌کننده در جلوگیری از پیامدهای جبران‌ناپذیر ایفا می‌کند [۲-٤].

۱-۱- مروری بر پیشینهٔ پژوهش

در سیستم‌های مکان‌یابی بی‌سیم، گره‌هایی به عنوان

گره‌های مرجع با موقعیت‌های کاملاً شناخته‌شده، اندازه‌گیری‌های مختلف رادیویی را از سیگنال‌های فرکانس رادیویی (RF)[۲] ساطع‌شده از کاربران شبکه جمع‌آوری و از آنها برای موقعیت‌یابی کاربران استفاده می‌کنند [۵]. در واقع، برای این کار از مشخصات سیگنال ورودی در گره مرجع استفاده می‌شود که به عنوان روش‌های مبتنی بر زمان، زاویه و قدرت سیگنال دریافتی (RSS)[۳] شناخته می‌شوند. روش مکان‌یابی با RSS بر اساس اندازه‌گیری مقدار توان دریافتی سیگنال است. روش‌های مبتنی بر زمان، بر اساس اندازه‌گیری زمان رسیدن (ToA)[۴] و تفاوت زمانی رسیدن (TDoA)[۵] سیگنال در محل گره‌های مرجع عمل می‌کنند. در روش مبتنی بر زاویهٔ ورود (AoA)[۶]، با اندازه‌گیری زاویهٔ رسیدن سیگنال در گره‌های مرجع مختلف، موقعیت منبع سیگنال تخمین زده می‌شود [۱، ۶].

موقعیت‌یابی را می‌توان با ترکیب روش‌های بیان‌شده نیز انجام داد؛ برای مثال، ترکیب RSS و ToA [۵] یا AoA و RSS [۷]. ترکیب این روش‌ها نیاز به استفاده از حسگرهای مختلف و طراحی الگوریتم‌هایی پیچیده‌تر برای محاسبهٔ موقعیت دارد؛ در حالی که مکان‌یابی مبتنی بر RSS به سخت‌افزار پیچیده نیاز ندارد. این روش می‌تواند با فناوری‌های مختلف رادیویی اجرا شود و به همگام‌سازی زمانی بین آنتن‌های فرستنده و گیرنده نیاز ندارد [۸].

در روش‌های رایج مکان‌یابی رادیویی مبتنی بر ویژگی‌های سیگنال، از ایستگاه‌های زمینی برای تعیین موقعیت استفاده می‌شود. برای مثال، در پژوهش‌های [۳] و [۹]، با بهره‌گیری از روش‌های مبتنی بر RSS و در نظر گرفتن گره‌های مرجع ایستا، مکان‌یابی منابع سیگنال انجام شده است. با این‌حال، در شرایط بحرانی، مکان‌یابی با گره‌های مرجع ایستا ممکن است قابل اجرا نباشد. از این رو، استفاده از فناوری‌های هوایی و پهپادها به عنوان گره‌های مرجع هوایی برای مکان‌یابی منبع سیگنال، به عنوان راه‌حلی مقرون‌به‌صرفه، سریع و انعطاف‌پذیر، مورد توجه قرار گرفته است. از مزایای جالب توجه این رویکرد دست‌یابی به دقت بیشتر در مکان‌یابی و توانایی ارائهٔ خدمات موقعیت‌یابی در شرایطی است که دسترسی به GPS محدود شده یا دریافت سیگنال‌های ماهواره‌ای غیرممکن است [۱، ۴، ۱۰].

فرایند مکان‌یابی رادیویی با انتقال مجموعه‌ای از سیگنال‌ها انجام می‌شود. مکان‌یابی می‌تواند یک فرایند دومرحله‌ای باشد که در مرحلهٔ اول اندازه‌گیری‌های مربوط به موقعیت از سیگنال استخراج می‌شود. سپس، با تجزیه‌وتحلیل سیگنال دریافتی و استفاده از روش‌های مختلف موقعیت‌یابی، این اندازه‌گیری‌های مربوط به موقعیت مانند RSS به تخمین موقعیت منجر می‌شوند [۱۱].

روش‌های تخمین موقعیت به چندین دادهٔ اندازه‌گیری‌شده نیاز دارند. در [۸]، مدلی برای مکان‌یابی بر اساس RSS ارائه شده است که در آن با استفاده از چهار پهپاد چندین داده جمع‌آوری می‌شوند و الگوریتم با محاسبهٔ خطوط عبوری از موقعیت منبع، نقطه‌ای را پیدا می‌کند که کمترین فاصله را با تمام خطوط دارد. در [۴]، سامانه‌ای برای مکان‌یابی منبع سیگنال با هدف پشتیبانی از عملیات امداد و نجات طراحی شده است که در آن، هفت پهپاد متحرک اندازه‌گیری‌های TDoA را نسبت به یک پهپاد مرجع انجام می‌دهند. از آنجا که استفاده از تعداد پهپاد کمتر در سیستم‌های موقعیت‌یابی به معنای کاهش هزینه‌های استقرار است، پژوهشگران در [۱۰، ۱۲] مکان‌یابی یک منبع ساطع‌کنندهٔ امواج RF ثابت با استفاده از یک پهپاد و چندین مرحله اندازه‌گیری را انجام داده‌اند.

ممکن است پهپادها در شرایط نامطمئن قادر به تصمیم‌گیری خودکار نباشند؛ از این رو، یادگیری تقویتی (RL)[۷] می‌تواند با افزودن سطحی از هوشمندی، توانایی تصمیم‌گیری خودکار پهپادها را بر اساس شرایط محیطی بهبود بخشد. همچنین، RL می‌تواند دقت مکان‌یابی و ناوبری پهپادها را افزایش دهد [۱۳، ۱۴]. در [۱۳]، یک پهپاد پس از پایش اولیهٔ محیط، مکان تقریبی منبع سیگنال را با استفاده از دست‌کم سه اندازه‌گیری RSS تخمین می‌زند و سپس، با بهره‌گیری از RL، به صورت برخط مسیر خود را با انتخاب نقاط بین‌راهی در طول مسیر بهینه‌سازی می‌کند؛ به‌گونه‌ای که میانگین خطاهای مکان به حداقل برسد. نتایج این مقاله نشان می‌دهد به‌کارگیری RL موجب بهبود چشم‌گیر دقت موقعیت‌یابی می‌شود. روش ارائه‌شده در [۱۳] فقط برای اهداف ثابت طراحی شده است و نیازمند

یک مرحله پایش اولیه از کل محیط است.

مکان‌یابی ایستگاه‌های غیرمجاز رادیویی یکی دیگر از کاربری‌های موقعیت‌یابی رادیویی با استفاده از پهپادها و الگوریتم‌های RL است. در [۱۵]، یک پهپاد مجهز به آنتن جهتی، پس از اندازه‌گیری مکرر RSS، جهت حرکت خود را در بازه‌های زمانی تعیین می‌کند و در نهایت، مکان ایستگاه‌های رادیویی غیرمجاز (IRS)[۸] را با استفاده از الگوریتم یادگیری Q آگاه از جهت تخمین می‌زند. مرجع [۱٦] نیز روشی مبتنی بر یادگیری Q برای مکان‌یابی خودکار منابع تداخل رادیویی توسط پهپاد در کانال رایلی ارائه می‌دهد. با این حال، در این مطالعات، بهینه‌سازی مسیر با استفاده از یک منبع رادیویی ایستا بررسی شده است که مختصات آن مشخص است و پهپاد با اندازه‌گیری RSS به سمت منبع هدایت می‌شود.

در [۱۷–۱۹]، مسئلهٔ مکان‌یابی با اندازه‌گیری RSS و استفاده از الگوریتم یادگیری Q حل شده است. هدف این پژوهش‌ها ناوبری یک پهپاد به سمت منبع سیگنال ثابت با کاربری امداد و نجات در شرایط بحرانی محیط‌های داخلی است. در [۱۷]، هدایت پهپاد به سمت منبع سیگنال با الگوریتم یادگیری Q و در نظر گرفتن کانال محوشدگی رایلی انجام شده است. در [۱۸]، دو الگوریتم یادگیری Q مبتنی بر RSS و الگوریتم یادگیری Q مبتنی بر مکان بررسی شده‌اند. طبق نتایج، یادگیری Q مبتنی بر RSS عملکردی مشابه و قابل رقابت با یادگیری Q مبتنی بر مکان ارائه می‌دهد و در مقایسه با این روش، نوسانات کمتری در طول فرآیند آموزش دارد. مکان‌یابی با الگوریتم‌های RL را با سیگنال‌های اندازه‌گیری‌شده از آنتن‌های جهت‌دار و همه‌جهتی می‌توان اجرا کرد [۱۹].

پهپادها و الگوریتم RL می‌توانند برای پایش زیرساخت‌ها و نظارت هوشمند محیطی مفید باشند. برای مثال، [۱٤] مکان‌یابی منبع تشعشع با یک پهپاد و شبکهٔ عصبی عمیق Q را انجام داده است. مدل قرائت‌های حسگر با استفاده از اصول فیزیک تشعشع برای ایجاد یک مدل تابشی تولید شده است. علاوه بر نظارت زیرساخت‌ها، می‌توان در شرایط بحرانی به کمک پهپادها زیرساخت‌های ارتباطی موقت ایجاد کرد. در [۲۰]، موقعیت‌یابی کاربر زمینی و بهینه‌سازی مکان استقرار و توان ارسال پهپادها به عنوان ایستگاه هوایی به کمک یادگیری Q انجام شده است.

هر منبع سیگنال متحرک علاوه بر مکان‌یابی نیاز به ردیابی نیز دارد. با این حال، پژوهش‌های بیان‌شده مکان‌یابی منابع سیگنال ثابت را انجام داده‌اند و تحرک منبع سیگنال را نادیده گرفته‌اند. فیلتر کالمن یک الگوریتم تخمین پرکاربرد برای ردیابی اهداف است [۲۱].

در [۲۲]، پهپادهای بال چرخان، مجهز به حسگرهای RSS، در یک گروه سازمان‌دهی شده‌اند و فقط با اندازه‌گیری RSS، حرکت یک منبع سیگنال متحرک را دنبال می‌کنند. در [۲۳]، الگوریتم جست‌وجو و مکان‌یابی برای یک منبع سیگنال متحرک زمینی با استفاده از وسیلهٔ هوایی خودکار (AAV)[۹] طراحی و شبیه‌سازی شده است. AAV در حالت جست‌وجوی جهانی به دنبال سیگنال هدف می‌گردد و در حالت مکان‌یابی با حرکت دایره‌ای اطراف هدف، موقعیت آن را با فیلترهای غیرخطی کالمن مانند فیلتر کالمن توسعه‌یافته (EKF)[۱۰] و فیلتر کالمن خنثی (UKF)[۱۱] تخمین می‌زند. برای مدل این مرجع، UKF عملکرد بهتری نسبت به EKF ارائه می‌کند. در [۲٤]، EKF چندمدلی تعاملی برای تخمین دقیق و پیوستهٔ موقعیت و حرکت هدف در سامانهٔ راداری به کار رفته است.

الگوریتم یادگیری Q چندعاملی [۲۵] به سه پهپاد فرصت اکتشاف در محیط را برای یافتن بهترین پیکربندی فضایی می‌دهد. هدف از این پیکربندی مکان‌یابی یک وسیلهٔ نقلیه و ایجاد رلهٔ ارتباطی با ایستگاه‌های پایهٔ زمینی است. همچنین، [۲٦] به کمک الگوریتم RL چندعاملی، پهپادها را برای ردیابی یک منبع سیگنال متحرک هماهنگ کرده است. در [۲۷]، یک رویکرد یادگیری تقویتی عمیق (DRL)[۱۲] برای کنترل پهپادها با هدف نهایی ردیابی امدادگران در محیط‌های چالش‌برانگیز سه‌بُعدی در حضور موانع و انسداد پیشنهاد شده است. نتایج شبیه‌سازی این مقاله نشان می‌دهد کنترل مبتنی بر DRL چندین پهپاد را قادر می‌سازد تا با بهبود عملکرد تخمین‌گر حالت هدف، منبع سیگنال را با دقت زیاد موقعیت‌یابی و ردیابی کنند.

هدف ما در این مقاله مکان‌یابی و ردیابی منبع سیگنال متحرک با استفاده از یک پهپاد به عنوان گره مرجع مجازی

است. در این سیستم، پهپاد مجهز به حسگرهای RSS همه‌جهتی است و می‌تواند فقط با همین اطلاعات از منبع سیگنال، موقعیت آن را مشخص و ردیابی کند. برای این کار، پهپاد در مکان‌های مختلف اندازه‌گیری‌های رادیویی را از هدف RF جمع‌آوری می‌کند. الگوریتم به شکلی طراحی می‌شود که پهپاد با کمک یادگیری Q بتواند به صورت خودکار به سمت منبع سیگنال حرکت کند. در تشخیص موقعیت، الگوریتم موقعیت‌یابی حداقل مربعات (LS)[۱۳] و الگوریتم یادگیری Q ترکیب شده‌اند. همچنین، برای بهبود نتایج تخمین موقعیت و ردیابی منبع سیگنال از فیلتر کالمن در کنار الگوریتم یادگیری Q استفاده شده است. در جدول (۱)، مقایسهٔ روش پیشنهادی با پژوهش‌های مرتبط ارائه شده است.

۲-۱- نوآوری پژوهش

در مقایسه با پژوهش‌های پیشین، نوآوری اصلی این مطالعه تلفیق سه الگوریتم LS، فیلتر کالمن و یادگیری Q در یک چارچوب هماهنگ و منسجم است. این رویکرد امکان به‌روزرسانی مرحله‌ای و پیوستهٔ موقعیت منبع را فراهم می‌کند و هم‌زمان، مسیر پروازی پهپاد را به صورت بلادرنگ و بدون نیاز به شناخت کامل محیط بهینه می‌کند؛ امری که منجر به افزایش دقت و کارایی سامانه در شرایط واقعی می‌شود.

جدول (۱): مقایسهٔ روش پیشنهادی با پژوهش‌های مرتبط

مراجع	تعداد پهپاد	ویژگی سیگنال	روش مکان‌یابی	منبع سیگنال متحرک	استفاده از KF	الگوریتم RL
[۳، ۹]	بدون پهپاد	RSS	LS	×	×	×
[٤]	۷	TDoA	LS	✓	×	×
[۸]	٤	RSS	هندسی با خطوط تقاطع	×	×	×
[۱۰]	۱	RSS	LS	×	×	×
[۱۲]	۱	RSS	فیلتر ذرات	×	×	×
[۱۳]	۱	RSS	LS	×	×	✓
[۱٤–۱۹]	۱	RSS	×	×	×	✓
[۲۰]	۱≤	RSS	LS	×	×	✓
[۲۲]	۳≤	RSS	×	✓	×	×
[۲۳]	۱	×	×	×	✓	×
[۲۵]	۳	RSS	LS	✓	×	✓
[۲٦]	۲≤	RSS	LS	✓	×	✓
[۲۷]	۳≤	RSS	روش بیزین	✓	×	✓
روش پیشنهادی	۱	RSS	LS	✓	✓	✓

نوآوری دیگر این پژوهش آموزش مستقل الگوریتم یادگیری Q برای تمامی ترکیب‌های ممکن از موقعیت‌های اولیهٔ پهپاد و منبع سیگنال است؛ به گونه‌ای که با وجود فضای حالت ۲۵حالته برای هر یک، تمامی حالات ممکن برای شروع و پایان عملیات بررسی شده‌اند. این روش امکان پوشش جامع شرایط مختلف شروع عملیات و ارتقای دقت مکان‌یابی را فراهم می‌کند. در الگوریتم پیشنهادی، خروجی مرحلهٔ پیش‌بینی فیلتر کالمن مکان بعدی منبع سیگنال را تخمین می‌زند. سپس، نزدیک‌ترین نقطهٔ بین‌راهی به این مکان به عنوان مقصد لحظه‌ای پهپاد انتخاب می‌شود. در ادامه، جدول Q متناظر با این ترکیب از موقعیت اولیهٔ پهپاد و مقصد نهایی (که پیش‌تر آموزش داده و ذخیره شده است)، فراخوانی و استفاده می‌شود.

سایر بخش‌های مقاله به صورت زیر سامان‌دهی شده‌اند:

در بخش ۲، مدل سیستم،کانال و توان پهپاد ارائه شده است. در بخش ۳، الگوریتم یادگیری Q معرفی می‌شود. بخش ٤ روش حل مسئله را ارائه داده است. نتایج شبیه‌سازی و مقادیر عددی در بخش ۵ ارائه شده است. در نهایت، نتیجه‌گیری در بخش ٦ بیان می‌شود.

۲- مدل سیستم

یک پهپاد به عنوان ایستگاه هوایی متحرک برای موقعیت‌یابی و ردیابی منبع سیگنال مستقر روی زمین در نظر گرفته شده است. این پهپاد بر فراز یک منطقهٔ نیمه‌شهری در ارتفاع ثابت h پرواز می‌کند. پهپاد در مسیر حرکت خود، در موقعیت‌هایی مشخص که به آنها نقاط بین‌راهی ($W_i = w_1, w_2, \dots, w_n$) گفته می‌شود، RSS منبع سیگنال RF را در محدودهٔ ارتباطی خود اندازه‌گیری و جمع‌آوری می‌کند. در هر نقطهٔ بین‌راهی، پس از ۲۰ اندازه‌گیری، مقدار میانگین RSS محاسبه می‌شود. پهپاد با توجه به مدل کانال استفاده‌شده در این پژوهش، فاصلهٔ خود با جسم را با استفاده از معادلهٔ افت مسیر به دست می‌آورد.

۲-۱- مدل کانال

با توجه به اینکه پهپاد به عنوان گره هوایی برای مکان‌یابی منبع سیگنال در نظر گرفته شده است، به مدل کانال ارتباطی پهپاد و منبع سیگنال هدف نیاز است. کانال‌های ارتباطی بین پهپاد و اهداف عمدتاً شامل پیوندهای دید مستقیم (LoS)[۱٤] و دید غیرمستقیم (NLoS)[۱۵] هستند که هر دو پیوند LoS و NLoS در مدل کانال در نظر گرفته شده‌اند. در این مدل، کانال هوا به زمین که در (۱) فرمول‌بندی شده است، به سایه و افت مسیر با زاویهٔ ارتفاع $\theta = tan^{-1}(\frac{h}{d})$ نیز وابسته است. گفتنی است، این مدل افت مسیر بر اساس دسی‌بل بیان شده است [۲۸، ۲۹].

$$PL = 20\,log_{10}(r) + 20\,log_{10}\left(\frac{4\pi f}{c}\right) + \psi(\theta) \qquad (۱)$$

در این مدل، f و c به‌ترتیب فرکانس سیستم و سرعت نور هستند و $\psi(\theta)$ یک متغیر تصادفی با توزیع گوسی با میانگین μ و واریانس $\sigma^2(\theta)$ است. فرض می‌شود μ صفر باشد و $\sigma^2(\theta)$ طبق (۲) تعریف می‌شود.

$$\sigma^2(\theta) = \mathbb{P}^2_{LoS}(\theta)\sigma^2_{LoS}(\theta) + \mathbb{P}^2_{NLoS}(\theta)\sigma^2_{NLoS}(\theta) \qquad (۲)$$

رابطهٔ میان $\mathbb{P}_{LoS}(\theta)$ و $\mathbb{P}_{NLoS}(\theta)$ برابر (۳) است و $\mathbb{P}_{LOS}(\theta)$ که احتمال داشتن پیوند LOS است با رابطهٔ (٤) بیان می‌شود.

$$\mathbb{P}_{NLoS}(\theta) = 1 - \mathbb{P}_{LoS}(\theta) \qquad (۳)$$

$$\mathbb{P}_{LoS}(\theta) = \frac{1}{1 + a_0.exp(-b_0.\theta)} \qquad (٤)$$

$\sigma^2_{LoS}(\theta)$ و $\sigma^2_{NLoS}(\theta)$ به‌ترتیب با اثر سایهٔ پیوندهای LoS و NLoS بین پهپاد و منبع سیگنال مطابقت دارند و به صورت (۵) و (٦) بیان می‌شوند.

$$\sigma^2_{LoS}(\theta) = a_{LoS}.exp(-b_{LoS}.\theta) \qquad (۵)$$

$$\sigma^2_{NLoS}(\theta) = a_{NLoS}.exp(-b_{NLoS}.\theta) \qquad (٦)$$

در روابط بیان‌شده برای مدل کانال، a_0، b_0، a_{LoS}، b_{LoS}، a_{NLoS} و b_{NLoS} پارامترهای وابسته به محیط هستند [۳۰].

۲-۲- مدل مصرف توان پهپاد

مدل مصرف توان یا به عبارتی مصرف انرژی پهپاد بال چرخان مبتنی بر مدل‌های ارائه‌شده در [۱۰] و [۳۱] استفاده می‌شود. در این مدل، مصرف انرژی مورد نیاز برای تبادل اطلاعات در نظر گرفته نمی‌شود؛ زیرا در عمل، می‌توان از مصرف انرژی مورد نیاز برای ارتباطات و تبادل داده در مقابل مصرف انرژی پیشرانهٔ پهپاد صرف‌نظر کرد [۳۲]. نحوهٔ محاسبهٔ توان مصرفی کل پهپاد در (۷) بیان شده است که مجموع توان‌های سه منبع مصرف توان اصلی در پهپاد بال چرخان است. این سه منبع عبارت‌اند از: توان نمایهٔ پرهٔ پهپاد (P_{blade})، توان پارازیت ($P_{parasite}$) و توان القایی ($P_{induced}$).

$$P_{total} = P_{blade} + P_{parasite} + P_{induced} \qquad (۷)$$

۱-۲-۲- توان نمایهٔ پرهٔ پهپاد

این توان برای چرخاندن پره‌ها مورد نیاز است و به وسیلهٔ (۸) تعریف می‌شود که در آن v سرعت پهپاد، v_b سرعت چرخش پره است و K نیز نشان‌دهندهٔ ثابتی است که به ابعاد پره بستگی دارد.

$$P_{blade} = K(1 + 3\frac{v^2}{v_b^2}) \quad (۸)$$

۲-۲-۲- توان پارازیت

این توان برای مفهوم غلبه بر نیروی پسا[۱۶] ایجادشده هنگام حرکت پهپاد در هوا استفاده می‌شود. نحوهٔ محاسبهٔ توان پارازیت[۱۷] در (۹) تعریف شده است. این توان با مکعب سرعت پهپاد (v) متناسب است و هنگام شناوربودن آن صفر و در سرعت‌های بالا بسیار بزرگ می‌شود. در این رابطه، ρ چگالی هوا و C_{ds} ثابتی است که به ضریب کشش پهپاد و ناحیهٔ مرجع بستگی دارد.

$$P_{parasite} = \frac{1}{2}\rho v^3 C_{ds} \quad (۹)$$

۳-۲-۲- توان القایی

این توان برای بلندکردن پهپاد و غلبه بر نیروی کششی ناشی از گرانش مورد نیاز است. هر زمان که یک پهپاد در حال حرکت است، جریان هوایی که به سمت آن می‌آید، مسیر پهپاد را تغییر می‌دهد و به بلندکردن آن کمک می‌کند. بنابراین، توان القایی لازم نسبتی معکوس با سرعت هوا دارد. هنگام شناورماندن پهپاد، تمام جریان هوای مورد نیاز برای بلندکردن پهپاد باید توسط پره‌های چرخان ایجاد شود که این امر منجر به مصرف انرژی بیشتر می‌شود. توان القایی را می‌توان به صورت (۱۰) نوشت که در آن m و g به‌ترتیب جرم پهپاد و گرانش استاندارد را نشان می‌دهند و v_{ind} نشان‌دهندهٔ میانگین سرعت القایی پروانه‌ها در پرواز رو به جلو است و با (۱۱) بیان می‌شود که در آن متغیر A بیان‌کنندهٔ سطح پهپاد است.

$$\boldsymbol{P}_{induced} = mgv_{ind} \quad (۱۰)$$

$$v_{ind} = \sqrt{\frac{-v^2 + \sqrt{v^4 + \left(\frac{mg}{\rho A}\right)^2}}{2}} \quad (۱۱)$$

در صورت شناورماندن (یعنی زمانی که $v = 0$)، کل توان مصرفی به توان شناور محدود می‌شود و بر اساس (۱۲) محاسبه می‌شود.

$$\boldsymbol{P}_{total} = \boldsymbol{P}_{hover} = K + \sqrt{\frac{(mg)^3}{2\rho A}} \quad (۱۲)$$

۳- یادگیری تقویتی (یادگیری Q)

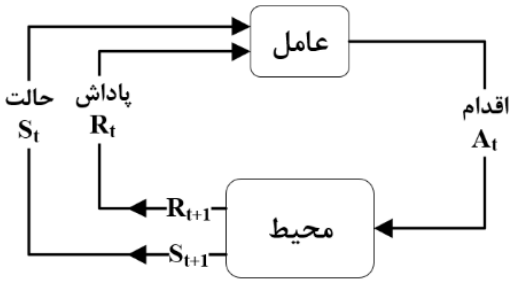
در ادامه، به طور خلاصه، RL را به عنوان یک تکنیک یادگیری ماشینی بررسی می‌کنیم که برای کنترل ماشین‌های مستقلی مانند پهپاد مناسب است.

RL معمولاً توسط فرایندهای تصمیم‌گیری مارکف (MDP)[۱۸] رسمیت می‌یابد. MDP یک چارچوب ریاضی برای مسائل تصمیم‌گیری متوالی در شرایطی است که نتایج تا حدی تصادفی و تا حدودی تحت کنترل عامل شناخته می‌شود. به طور خاص‌تر، عامل و محیط در هر یک از مراحل زمانی t تعامل دارند. در هر مرحلهٔ زمانی t، عامل تعدادی نمایش از حالت محیط $S_t \in \mathcal{S}$ را دریافت می‌کند و بر این اساس، اقدام A_t را از مجموعهٔ تمام اقدامات مجاز $A(s)$ انتخاب می‌کند. در گام زمانی بعدی، عامل به دلیل اقدام انجام‌شده پاداش عددی $R_{t+1} \in \mathcal{R} \subset \mathbb{R}$ را دریافت می‌کند و به حالت جدید می‌رود. MDP و عامل با هم باعث ایجاد دنباله یا مسیری می‌شوند که به صورت (۱۳) است.

$$S_0, A_0, R_1,\ S_1, A_1, R_2,\ S_2, A_2, R_3,\ \ldots \quad (۱۳)$$

در یک MDP محدود، مجموعهٔ حالت‌ها، اقدامات و پاداش‌ها ($\mathcal{S}$، A و R) همگی دارای تعداد عناصر محدود هستند [۳۳]. بر اساس پاداش دریافتی و پس از اجرای مکرر، عامل شروع به بهبود دانش خود از محیط می‌کند و باید بتواند سیاست را به‌نوعی تدوین کند که تعیین کند کدام اقدامات را برای حالت‌های ممکن محیط در نظر بگیرد. چارچوب اصلی RL در شکل (۱) نشان داده شده است

[۳۳، ۳٤].



شکل (۱): تعامل عامل و محیط در RL [۳۳]

الگوریتم یادگیری Q یک الگوریتم یادگیری تقویتی خارج از سیاست است که قانون به‌روزرسانی تابع ارزش آن برای جفت حالت–عمل به طور کلی طبق (۱٤) تعریف می‌شود [۳۵].

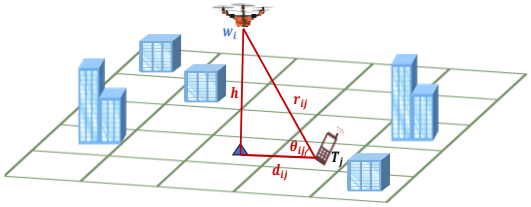
$$Q(S_t,A_t) \leftarrow Q(S_t,A_t) + \alpha\left[R_{t+1} + \gamma \max_a Q(S_{t+1},a) - Q(S_t,A_t)\right] \quad (۱٤)$$

در این رابطه، نرخ یادگیری (α) عددی بین صفر و یک است و تعیین می‌کند تا چه اندازه اطلاعات حالت جدید بر اطلاعات اولیه غلبه می‌کند. همچنین، ضریب تخفیف (γ) عددی بین صفر و یک است و اهمیت پاداش‌های آینده را تعیین می‌کند.

در این الگوریتم، تابع ارزش–عمل آموخته‌شدهٔ Q به طور مستقیم تابع ارزش عمل بهینهٔ q^* را مستقل از سیاستی که دنبال می‌شود، تقریب می‌زند. این موضوع به طرزی چشمگیر تجزیه‌وتحلیل الگوریتم را ساده می‌کند و اثبات همگرایی اولیه را فعال می‌کند. این سیاست همچنان دارای اثری است که تعیین می‌کند کدام جفت‌های حالت–عمل بازدید و به‌روزرسانی می‌شوند. در این الگوریتم، با $\max_a Q(S',a)$ از دانش موجود در محیط بهره‌برداری می‌شود. با این حال، تنها چیزی که برای همگرایی صحیح مورد نیاز است این است که همهٔ این جفت‌ها به‌روز شوند. با این فرض و برقراری شرایط تقریب تصادفی معمول در توالی پارامترهای اندازهٔ گام، Q با احتمال یک به q^* همگرا می‌شود [۳۳، ۳٦].

۴- روش حل مسئله

مکان منبع سیگنال با میانگین RSS جمع‌آوری‌شده در نقاط مختلف با استفاده از روش LS تخمین زده می‌شود. با توجه به شرایط محیطی، پیوند بین پهپاد و منبع سیگنال ممکن است از نوع LoS یا NLoS باشد. در شکل (۲)، فاصلهٔ مستقیم بین پهپاد در نقطهٔ بین‌راهی w_i و منبع سیگنال در نقطهٔ T_j با r_{ij} و فاصلهٔ زمینی با d_{ij} نشان داده شده است. علاوه بر این، زاویهٔ منبع سیگنال با پهپاد با θ_{ij} نشان داده شده است. پوشش‌دهی یکی از پارامترهای عملکردی است که نشان می‌دهد چه میزان از ناحیه در شعاع حسگری قرار گرفته است [۳۷]. محدودهٔ پوشش رادیویی هوایی پهپاد برابر ۷۰۰ متر در نظر گرفته شده است؛ یعنی پهپاد توانایی اندازه‌گیری RSS، از منبعی که در یک کره به شعاع حداکثر ۷۰۰متری در اطراف خود است، را دارد.



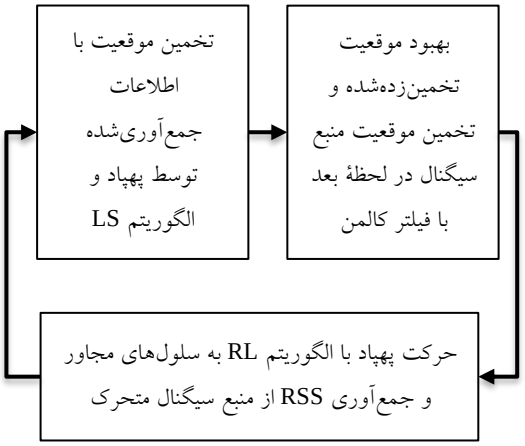
شکل (۲): پارامترهای مدل کانال در جمع‌آوری اندازه‌گیری‌های RSS با استفاده از یک پهپاد برای مکان‌یابی منبع سیگنال

برای شروع عملیات ردیابی، طراحی یک مسیر اولیه برای حرکت پهپاد لازم است؛ زیرا در ابتدا به تشخیص وجود منبع سیگنال و یک موقعیت اولیه از منبع سیگنال نیاز است. از این رو، ابتدا مسیری برای تشخیص کمترین تعداد RSS مناسب برای تخمین مکان منبع سیگنال با LS در نظر گرفته می‌شود. پس از اینکه پهپاد به تعداد کافی RSS در مکان‌های مختلف خود از هدف اندازه گرفت، به پهپاد اجازه داده می‌شود تا با استفاده از روش یادگیری Q، مسیر بهینه تا منبع سیگنال را با عبور از سلول‌های گسستهٔ تعریف‌شده در محیط پیدا کند.

در الگوریتم یادگیری Q که یکی از روش‌های RL است، ناحیهٔ جست‌وجو به سلول‌هایی با ابعاد یکنواخت تقسیم می‌شود و در هر سلول، جهت حرکت پهپاد بر اساس سیاست به‌دست‌آمده از الگوریتم یادگیری Q تعیین می‌شود. برخلاف روش‌های انجام‌شده در پیشینهٔ پژوهش به کمک

یادگیری Q که مکان آغاز حرکت پهپاد و مکان منبع سیگنال ثابت و یک سلول مشخص بود، در روش پیشنهادی امکان وجود منبع سیگنال در کل محیط جست‌وجو و آغاز حرکت پهپاد در هر کدام از سلول‌ها در نظر گرفته شده است. در این راستا، یادگیری با در نظر گرفتن تمام‌حالت‌های ممکن برای مکان پهپاد در آغاز پرواز و موقعیت منبع سیگنال تکرار شده است. خروجی الگوریتم یادگیری Q به عنوان داده‌های اولیه از یادگیری انجام‌شده در مرحلهٔ انجام عملیات موقعیت‌یابی و ردیابی منبع سیگنال استفاده می‌شود.

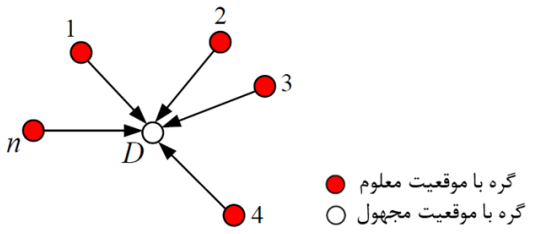
مطابق شکل (۳)، برای در نظر گرفتن تحرک منبع سیگنال، خروجی LS که همان موقعیت هدف اندازه‌گیری‌شده است، به فیلتر کالمن داده می‌شود و خروجی مرحلهٔ به‌روزرسانی اندازه‌گیری در فیلتر کالمن به عنوان تخمین نهایی از موقعیت منبع سیگنال تعیین می‌شود. همچنین، خروجی مرحلهٔ پیش‌بینی فیلتر کالمن برای ردیابی منبع سیگنال متحرک در الگوریتم یادگیری Q به عنوان سلول هدف استفاده می‌شود. در فیلتر کالمن پیشنهادی، موقعیت، سرعت و شتاب حرکت منبع سیگنال در دو بُعد به عنوان حالت‌های فیلتر کالمن تعریف شده‌اند.



شکل (۳): نمودار بلوکی روش پیشنهادی برای موقعیت‌یابی و ردیابی منبع سیگنال متحرک

۴-۱- روش مکان‌یابی LS

وقتی تعداد گره‌های با موقعیت معلوم (n) بیش از سه باشد، می‌توان LS را برای محاسبهٔ مختصات گره ناشناخته $D(x;\ y)$ در شکل (٤) استفاده کرد.



شکل (٤): مثالی از گره‌های معلوم با فواصل مختلف از گره با موقعیت مجهول D [۳۸]

فرض کنید مختصات گره‌های معلوم به‌ترتیب $(x_1;y_1)$، $(x_2;y_2)$، ... و $(x_n;y_n)$ و فاصلهٔ بین D و گره‌های معلوم به‌ترتیب d_1، d_2، ... و d_n است. سپس، می‌توان معادلات را به صورت (۱۵) نوشت.

$$\begin{cases}(x-x_1)^2+(y-y_1)^2=d_1^2\\(x-x_2)^2+(y-y_2)^2=d_2^2\\(x-x_3)^2+(y-y_3)^2=d_3^2\\\vdots\\(x-x_n)^2+(y-y_n)^2=d_n^2\end{cases} \quad (۱۵)$$

با کم‌کردن آخرین معادله از $n-1$ معادلهٔ اول می‌توان معادلات را حل کرد. در نهایت، می‌توان معادلات را به صورت $\boldsymbol{AX}=b$ بازنویسی کرد که در آن $\boldsymbol{A}$، X و b به صورت زیر تعریف می‌شود.

$$X=[x\quad y]^T$$
$$\boldsymbol{A}=2\begin{bmatrix}(x_1-x_n) & (y_1-y_n)\\ \vdots & \vdots\\ (x_{n-1}-x_n) & (y_{n-1}-y_n)\end{bmatrix}$$
$$b=\begin{bmatrix}d_n^2-d_1^2+x_1^2-x_n^2+y_1^2-y_n^2\\ \vdots\\ d_n^2-d_{n-1}^2+x_{n-1}^2-x_n^2+y_{n-1}^2-y_n^2\end{bmatrix} \quad (۱۶)$$

سپس، می‌توان $X=(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^Tb$ را به دست آورد. در واقع، روش LS بسط روش سه‌پهلوبندی است [۳۸]. این روش در مقایسه با الگوریتم مثلث‌سازی، می‌تواند مکان گره‌های مجهول را با دقت بیشتری تخمین بزند [۳۹].

۴-۲- ردیابی هدف با فیلتر کالمن

با توجه به اینکه منبع سیگنال مورد ردیابی هدفی متحرک است، به فرایندی برای تخمین موقعیت هدف در گام‌های بعدی نیاز است. فیلتر کالمن یک الگوریتم پرکاربرد است که برای تخمین حالت‌های پنهان سیستم، حتی زمانی

که اندازه‌گیری‌ها نادقیق و نامطمئن هستند، طراحی شده است. همچنین، فیلتر کالمن وضعیت آیندهٔ سیستم را بر اساس تخمین‌های گذشته پیش‌بینی می‌کند [۲۱]. موقعیت‌های مرحلهٔ پیش‌بینی فیلتر کالمن به عنوان پارامترهای ورودی یادگیری تقویتی برای حرکت پهپاد به سمت هدف استفاده می‌شوند.

فیلتر کالمن با فرض اینکه بتوان مسئله را در قالب (۱۷) مدل کرد، نوشته می‌شود؛ به طوری که $\boldsymbol{x_k}$ بردار حالت در زمان t_k، $\boldsymbol{\Phi_k}$ ماتریس به عنوان ارتباط‌دهندهٔ $\boldsymbol{x_k}$ به $\boldsymbol{x_{k+1}}$ یا به عبارتی حالت فعلی به حالت بعد و $\boldsymbol{w_k}$ بردار نویز سفید با ساختار کوواریانس مشخص $\boldsymbol{Q}$ باشد.

$$\boldsymbol{x_{k+1}} = \boldsymbol{\Phi_k x_k} + \boldsymbol{w_k} \tag{۱۷}$$

مشاهده (اندازه‌گیری) فرایند در زمان‌های گسسته مطابق رابطهٔ خطی (۱۸) فرض می‌شود.

$$\boldsymbol{z_k} = \mathbf{H}_k \boldsymbol{x_k} + \boldsymbol{v_k} \tag{۱۸}$$

$\boldsymbol{z_k}$ بردار اندازه‌گیری در زمان t_k است. $\boldsymbol{H_k}$ ماتریس ارتباط ایده‌آل (بدون نویز) بین اندازه‌گیری و بردار حالت را در زمان t_k ایجاد می‌کند. $\boldsymbol{v_k}$ بردار خطای اندازه‌گیری است که فرض می‌شود دنباله‌ای سفید با ساختار کوواریانس مشخص $\mathbf{R}$ و همبستگی متقابل صفر با دنبالهٔ $\boldsymbol{w_k}$ باشد. گفتنی است که فرض می‌شود $\boldsymbol{\Phi_k}$، $\boldsymbol{H_k}$ و کوواریانس‌های توصیف‌کنندهٔ $\boldsymbol{w_k}$ و $\boldsymbol{v_k}$ را می‌دانیم. ماتریس‌های کوواریانس برای بردارهای $\boldsymbol{w_k}$ و $\boldsymbol{v_k}$ در (۱۹) بیان شده‌اند.

$$\begin{aligned} E[\mathbf{w}_k \mathbf{w}_i^T] &= \begin{cases} \mathbf{Q}_k, & i = k \\ \mathbf{0}, & i \neq k \end{cases} \\ E[\mathbf{v}_k \mathbf{v}_i^T] &= \begin{cases} \mathbf{R}_k, & i = k \\ \mathbf{0}, & i \neq k \end{cases} \\ E[\mathbf{w}_k \mathbf{v}_i^T] &= \mathbf{0}, \quad \text{for all } k \text{ and } i \end{aligned} \tag{۱۹}$$

تخمین در لحظهٔ t_k با استفاده از دانش‌های قبلی انجام می‌شود. با این فرض، تخمین پیشین با $\hat{\boldsymbol{x}}_k^-$ نشان داده می‌شود که در آن علامت «^» نشان‌دهندهٔ تخمین و «−» آن نشان‌دهندهٔ بهترین تخمین و پیش‌بینی قبل از اندازه‌گیری در t_k است. فرض بر این است که ماتریس کوواریانس خطای مرتبط با $\hat{\boldsymbol{x}}_k$ مشخص و برابر (۲۰) است.

$$\boldsymbol{P_k^-} = E[\boldsymbol{e_k^-}\boldsymbol{e_k^-}^T] = E[(\boldsymbol{x_k} - \hat{\boldsymbol{x}}_k^-)(\boldsymbol{x_k} - \hat{\boldsymbol{x}}_k^-)^T] \tag{۲۰}$$

با فرض تخمین پیشین $\hat{\boldsymbol{x}}_k^-$ با اندازه‌گیری $\boldsymbol{z_k}$، تخمین پیشین بهبود داده می‌شود. در این راستا، یک ترکیب خطی از اندازه‌گیری نویز و تخمین پیشین مطابق (۲۱) انتخاب می‌شود.

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^- + \boldsymbol{K_k}(\boldsymbol{z_k} - \boldsymbol{H_k}\hat{\boldsymbol{x}}_k^-) \tag{۲۱}$$

$\hat{\boldsymbol{x}}_k$ تخمین به‌روزرسانی‌شده و $\boldsymbol{K_k}$ بهرهٔ فیلتر کالمن نامیده می‌شود. $\boldsymbol{K_k}$ به صورت زیر محاسبه می‌شود.

$$\boldsymbol{K_k} = \boldsymbol{P_k^-}\boldsymbol{H_k^T}(\boldsymbol{H_k}\boldsymbol{P_k^-}\boldsymbol{H_k^T} + \boldsymbol{R_k})^{-1} \tag{۲۲}$$

به طوری که کوواریانس خطای $\boldsymbol{P_k}$ بعد از مشاهده برابر $\boldsymbol{P_k} = (\boldsymbol{I} - \boldsymbol{K_k}\boldsymbol{H})\boldsymbol{P_k^-}$ است.

در نهایت، تخمین حالت بهینه و ماتریس کوواریانس خطا برای گام زمانی بعد به‌ترتیب با (۲۳) و (۲٤) برابر است.

$$\hat{\boldsymbol{x}}_{k+1}^- = \boldsymbol{\Phi_k}\hat{\boldsymbol{x}}_k \tag{۲۳}$$

$$\boldsymbol{P_{k+1}^-} = \boldsymbol{\Phi_k}\boldsymbol{P_k}\boldsymbol{\Phi_k^T} + \boldsymbol{Q_k} \tag{۲٤}$$

معادلهٔ حالت فیلتر کالمن بر اساس موقعیت، سرعت و شتاب منبع سیگنال در دو بُعد X و Y تدوین شده است. در واقع، معادلات فیلتر کالمن با توجه به معادلات دینامیکی حرکت هدف متحرک بر اساس وابستگی به موقعیت حالت قبل، سرعت و شتاب جسم در حال حرکت نوشته شده‌اند. در (۲۵)، معادلات دینامیکی حالات فیلتر کالمن برای محور X بیان شده‌اند. معادلات حالت برای محور Y نیز به همین ترتیب هستند. در این مدل، فرض می‌شود خطاهای تخمین و نویز در محورهای X و Y همبستگی ندارند [۲۱، ٤۰–٤۲].

$$\begin{aligned} \hat{x}_{k+1} &= \hat{x}_k + \hat{\dot{x}}_k \Delta t + \frac{1}{2}\hat{\ddot{x}}_k \Delta t^2 \\ \hat{\dot{x}}_{k+1} &= \hat{\dot{x}}_k + \hat{\ddot{x}}_k \Delta t \\ \hat{\ddot{x}}_{k+1} &= \hat{\ddot{x}}_k \end{aligned} \tag{۲۵}$$

۳-۴- الگوریتم یادگیری Q پیشنهادی

محیط جست‌وجوی موقعیت و ردیابی منبع سیگنال، یک محیط با مساحت ۱۰۰۰×۱۰۰۰ مترمربع در نظر گرفته

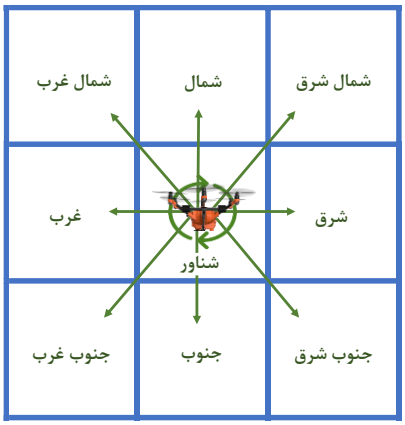
شده است. برای حل الگوریتم یادگیری Q، شبکه‌ای گسسته به ابعاد ۵×۵ در نظر گرفته می‌شود که مقادیر مختصات منبع سیگنال در محیط به مرکز یکی از این سلول‌های گسسته (نقطهٔ بین راهی) نگاشت می‌شود. بنابراین، اگر محیط یادگیری Q به سلول‌هایی گسسته با طول و عرض مشخص تقسیم شود، هر حالت از قرارگیری پهپاد در محیط در حین پرواز با مختصاتی تعریف می‌شود که نشان‌دهندهٔ یک سلول دوبُعدی گسسته در این شبکه است. با این فرض که موقعیت سلول پهپاد در آغاز پرواز و سلول منبع سیگنال مشخص باشد، تعداد حالت‌های محیط یا به عبارتی تعداد سطرهای هر تابع ارزش Q برابر مقدار ۲۵ است.

الگوریتم یادگیری Q با کوچک‌سازی ابعاد مسئله، مسائل را حل می‌کند. برای مثال، [۴۳] روشی مبتنی بر یادگیری Q برای کاهش نویز صوتی به صورت فعال پیشنهاد داده است. در این مقاله، مسئله‌ای که دارای ابعاد بزرگ بوده است، ابتدا به دو یا چند مسئلهٔ کوچک‌تر و مشابه تقسیم شده و در نهایت با تجمیع جواب‌های به‌دست‌آمده، مسئله حل شده است. بنا بر این ایده، در این پژوهش نیز با تقسیم مسئله به چند جزء با جزئیاتی که در ادامه شرح داده می‌شود، مسئله حل شده است.

از آنجا که در مسئلهٔ پیشنهادی، منبع سیگنال و پهپاد در آغاز حرکت می‌توانند در هر کدام از سلول‌های محیط باشند، الگوریتم پیشنهادی به گونه‌ای طراحی شده است که الگوریتم یادگیری Q برای تمام تعداد سلول‌هایی که پهپاد در آغاز پرواز قرار دارد و سلول‌هایی که منبع سیگنال می‌تواند در آن‌ها قرار بگیرد، تکرار شود. تعداد سلول‌های ممکن برای موقعیت شروع پرواز پهپاد ۲۵ و تعداد سلول‌های ممکن برای مکان منبع سیگنال ۲۵ است. در نهایت، خروجی الگوریتم یادگیری Q پیشنهادی که مجموعه‌ای از تابع‌های ارزش اولیه برای مرحلهٔ تست و انجام عملیات است، به عنوان داده‌های خروجی ذخیره می‌شود.

۱-۳-۴- مجموعهٔ اقدامات مسئله

عامل یا همان پهپاد در مسئله می‌تواند در هشت حرکت نشان‌داده‌شده در شکل (۵) به حالت‌های بعدی حرکت کند یا در صورتی که به نزدیکی منبع سیگنال رسیده است و سلول منبع سیگنال با سلول پهپاد برابر است، در همان حالت شناور بماند. عامل پس از انتخاب و انجام هر اقدام، RSS را برای موقعیت‌یابی اندازه‌گیری می‌کند.



شکل (۵): مجموعهٔ اقدام‌های مجاز برای عامل (پهپاد) در هر حالت از الگوریتم یادگیری Q

در مراحل مختلف الگوریتم‌های مبتنی بر RL، گاهی عامل تنها بر اساس دانش خود از محیط به صورت حریصانه عمل می‌کند تا پاداش کل مورد انتظار را بیشینه کند؛ این رویکرد که بهره برداری[۱۹] نام دارد، ممکن است به ویژه در مراحل اولیهٔ یادگیری به انتخاب‌های غیربهینه منجر شود؛ زیرا عامل با دانش محدود خود عمل می‌کند. در نتیجه، برای دست‌یابی به رفتار بهینه، عامل باید اکتشاف[۲۰] انجام دهد. هنگامی که یک عامل اکتشاف می‌کند، لزوماً به بهترین شکل ممکن عمل نمی‌کند، بلکه گزینه‌های مختلف موجود را بررسی می‌کند که توسط یک راهبرد اکتشافی تعیین می‌شوند.

روش‌های انتخاب اقدام مختلفی در RL وجود دارند. با توجه به [۳۳]، سیاست حریصانهٔ اپسیلون[۲۱] رایج‌ترین رویکرد برای ایجاد تعادل بین بهره‌برداری و اکتشاف در RL است. سیاست حریصانهٔ اپسیلون میزان اکتشاف را کنترل می‌کند و تصادفی‌بودن انتخاب اقدام در اکتشاف را تعیین می‌کند. عامل با احتمال ε بین ۰ و ۱ یک عمل تصادفی را انتخاب می‌کند و با احتمال $1-\varepsilon$ به صورت حریصانه یکی از اقدامات بهینهٔ آموخته‌شده در مورد تابع Q را انتخاب می‌کند و از الگوریتم بهره‌برداری می‌کند. اگر این راهبرد

$\pi(s)$ نامیده شود، می‌توان آن را به صورت زیر بیان کرد.

$$\pi(s) = \begin{cases} \varepsilon & \text{Choose a random action} \\ 1-\varepsilon & \text{Be greedy and exploit} \end{cases} \quad (۲٦)$$

در [۱٦] که مسئلهٔ مکان‌یابی توسط پهپاد حل شده است، این روش با روش Softmax مقایسه شده است. روش Softmax به انتخاب هر اقدام یک احتمال می‌دهد. طبق نتایج این مرجع، سیاست حریصانهٔ اپسیلون سریع‌تر همگرا می‌شود. همچنین، در [٤٤]، از سیاست حریصانهٔ اپسیلون برای طراحی الگوریتم یادگیری Q استفاده شده است تا انتخاب اقدام برای مدیریت انرژی و گذردهی در یک شبکهٔ بی‌سیم ناحیهٔ بدنی[۲۲] به طور مؤثر صورت پذیرد.

برای تنظیم پویای اکتشاف و بهره‌برداری از روش پیشنهادی [٤٥]، برای انتخاب ε در هر رویداد[۲۳] از یادگیری استفاده شده است. بنابراین، تعداد نمونه‌های اکتشافی در مراحل اولیهٔ یادگیری بیشتر است و با پیشرفت فرایند به‌تدریج کاهش می‌یابد. در واقع، طبق (۲۷)، ε در هر رویداد با در نظرگرفتن نرخ کاهش[۲٤] و شمارهٔ تکرار رویداد (i) به‌روزرسانی می‌شود؛ بنابراین، برای بهبود سرعت همگرایی با افزایش تعداد رویدادها، بهره‌برداری عامل از محیط هنگام انتخاب بعدی بیشتر می‌شود. در الگوریتم پیشنهادی این پژوهش، با آزمایش در مرحلهٔ یادگیری، مقدار نرخ کاهش ε انتخاب شده است.

$$\varepsilon \longleftarrow \mathrm{e}^{-\varepsilon * decay\ rate * i} \quad (۲۷)$$

تابع پاداش بازخورد ارزشی است که عامل هنگام اکتشاف در محیط دریافت می‌کند. اگر عامل عمل بهینه را انجام دهد، پاداش بزرگ‌تری به دست می‌آید. اگر عامل عملی ضعیف انجام دهد، پاداش کمتری دریافت می‌شود. اقدامات با پاداش زیاد شانس انتخاب‌شدن بیشتری دارند، در حالی که اقدامات با پاداش کم شانس انتخاب‌شدن کمتری خواهند داشت. راهبرد شبکهٔ این پژوهش انتخاب مسیری است که فاصله‌ای کمتر تا منبع سیگنال دارد. الگوریتم انتخاب مقدار پاداش در زمان t در واقع تابع تکه‌ای غیرخطی است که در الگوریتم (۱) تعریف می‌شود. در این الگوریتم، D_c فاصلهٔ هر سلول با سلول بعدی است که با آن فقط در محور x یا y تفاوت دارد. همچنین، موقعیت سلول منبع سیگنال و پهپاد دیگر ورودی‌های این الگوریتم هستند.

Input: (x_{w_i}, y_{w_i}), (x_{T_j}, y_{T_j}), D_c

$\text{distance}_t = \sqrt{(x_{wi} - x_{Tj})^2 + (y_{wi} - y_{Tj})^2}$

if $\text{distance}_t < \text{distance}_{t-1}$:

$\text{reward} = \frac{D_C}{1 + \text{current distance}}$

else:

$\text{reward} = -1$

Output: reward

الگوریتم (۱): انتخاب پاداش با توجه به فاصلهٔ پهپاد تا منبع سیگنال

در الگوریتم پیشنهادی، یادگیری Q منطبق بر سیاست حریصانهٔ اپسیلون پویا به‌ازای تمام حالت‌هایی که ممکن است پهپاد در موقع شروع پرواز آنجا باشد و تمام تعداد حالت‌هایی که منبع سیگنال ممکن است در آنها قرار بگیرد، تکرار می‌شود و سیاست بهینه در ماتریس‌های $Q^*(S,A)$ ذخیره می‌شود. این ماتریس‌ها در مرحلهٔ آزمایش به عنوان ماتریس $Q(S,A)$ اولیه استفاده می‌شوند. شرط توقف الگوریتم یادگیری Q در هر رویداد رسیدن پهپاد به مکان سلول منبع سیگنال است.

۵- نتایج شبیه‌سازی

در این پژوهش، پارامترهای مدل کانال بر اساس مدل نیمه‌شهری در نظر گرفته شده‌اند. مقادیر این پارامترها برای فرکانس کاری ۲۰۰۰ مگاهرتز در جدول (۲) بیان شده‌اند. مقادیر پارامترهای انتخابی برای مصرف توان پهپاد نیز در جدول (۳) ارائه شده‌اند.

جدول (۲): مقادیر پارامترهای وابسته به محیط نیمه‌شهری در مدل کانال [٤٦]

پارامتر	شرح	مقدار
a_{LoS}	ثابت سایه برای LoS	۵
b_{LoS}	ثابت سایه برای LoS	۳/۵
a_{NLoS}	ثابت سایه برای $NLoS$	۱۰
b_{NLoS}	ثابت سایه برای $NLoS$	۲/۵
a_0	پارامتر محیطی برای $\mathbb{P}_{LoS}$	٤۷
b_0	پارامتر محیطی برای $\mathbb{P}_{LoS}$	۲۰

جدول (۳): مقادیر پارامترهای استفاده‌شده برای مدل مصرف توان [۱۰، ۱۳]

پارامتر	شرح	مقدار
K	ثابت توان وابسته به ابعاد پره (kgm^2/s^3)	۵۷۰
m	جرم پهپاد (kg)	۵
ρ	چگالی هوا (kg/m^3)	۱/۲۲۵
v	سرعت پهپاد (km/h)	٤۰
v_b	سرعت چرخش پره (m/s)	۱۰۰
C_{ds}	ضریب کشش پهپاد و ناحیهٔ مرجع (m^2)	۰/٤
g	گرانش استاندارد (m/s^2)	۹.۸
A	نمادی از سطح پهپاد (m^2)	۰/۲۵

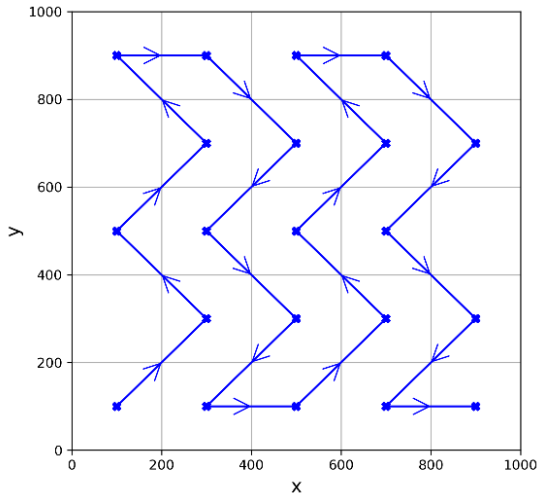
پارامترهای الگوریتم بهینه‌سازی یادگیری Q پیشنهادی با مقادیر بیان‌شده در جدول (۴) تنظیم می‌شوند. برای مثال، در لحظهٔ آغاز، ضریب یادگیری ε برابر یک است و تعداد رویدادهای آموزشی روی ۵۰۰۰ تنظیم شده است.

جدول (٤): مقادیر پارامترهای مسئلهٔ یادگیری Q

پارامتر	مقدار
تعداد کل وضعیت موقعیت شروع پرواز پهپاد	۲۵
تعداد حالات ممکن برای مکان منبع سیگنال	۲۵
نرخ کاهش	۰/۰۰۰۵
بیشینهٔ رویداد	۵۰۰۰
مقدار اولیهٔ $Q(S,A)$	۰
α	۰/۱
γ	۰/۹
مقدار اولیهٔ ε در فاز آموزش	۱

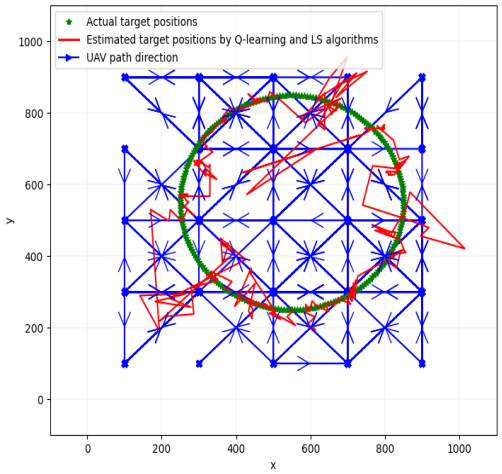
به منظور تشخیص وجود منبع سیگنال در منطقهٔ جست‌وجو، یک پهپاد باید مسیری از پیش تعیین‌شده به عنوان مسیر اولیه در منطقه طی کند تا کمترین تعداد RSS مورد نیاز در محیط را برای مکان‌یابی اندازه‌گیری کند. این مقادیر به عنوان ورودی الگوریتم LS برای یافتن اولین تخمین از موقعیت منبع سیگنال استفاده می‌شوند. طبق معادلات بیان‌شده برای LS، اگر این مسیر به صورت خطی روی محورهای طولی یا عرضی محیط طراحی شود باعث ابهام و افزایش خطا در یکی از مختصات تخمین‌زده‌شدهٔ x و y منبع سیگنال خواهد شد.

از این رو، برای جمع‌آوری مقادیر RSS مورد نیاز برای آغاز، الگوریتم به گونه‌ای طراحی شده است که کمترین تکرار خطی در هر دو محور مختصات در آن رعایت شده باشد. در شکل (۶)، این طراحی به همراه نقاط بین‌راهی و سلول‌های یادگیری Q نمایش داده شده است. پهپاد با توجه به شعاع ارتباطی خود، این مسیر را تا زمانی طی می‌کند که داده‌های RSS به میزان کافی برای راه‌اندازی الگوریتم یادگیری و آغاز فرایند مکان‌یابی و ردیابی جمع‌آوری شوند. پس از این مرحله، ادامهٔ مسیر و تصمیم‌گیری برای حرکت پهپاد بر اساس موقعیت منبع سیگنال و خروجی الگوریتم یادگیری Q صورت خواهد گرفت.

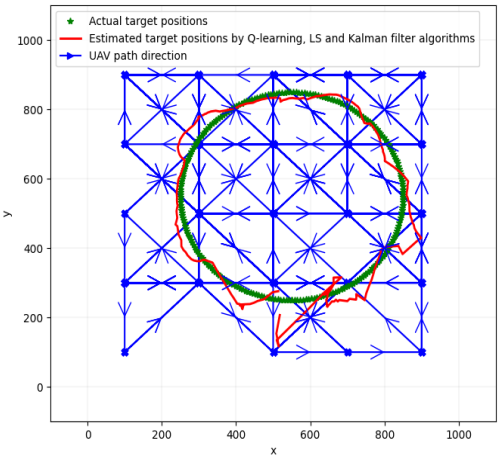


شکل (٦): مسیر اولیهٔ پهپاد برای یافتن کمترین تعداد اندازه‌گیری RSS مورد نیاز برای موقعیت‌یابی و ردیابی منبع سیگنال

یک نمونه از مسیرهای حرکت پهپاد و نتایج مکان‌یابی و ردیابی منبع سیگنال متحرک با استفاده از ترکیب الگوریتم‌های یادگیری Q و LS در شکل (۷) نشان داده شده است. همچنین، شکل (۸) نمونه‌ای از تصویر مسیر حرکت پهپاد و ردیابی منبع سیگنال را در صورت وجود فیلتر کالمن در الگوریتم‌ها نشان می‌دهد. همان‌طور که از این دو تصویر مشخص است، افزودن فیلتر کالمن باعث بهبود ردیابی منبع سیگنال شده است.



شکل (۷): مسیر پهپاد و موقعیت منبع سیگنال در ردیابی منبع سیگنال متحرک با ترکیب یادگیری Q و LS

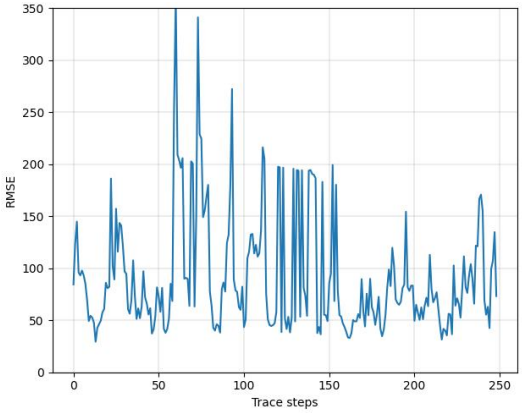


شکل (۸): مسیر پهپاد و موقعیت منبع سیگنال در ردیابی منبع سیگنال متحرک با ترکیب فیلتر کالمن، یادگیری Q و LS

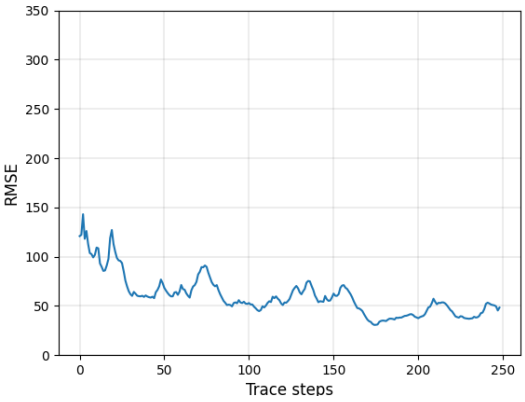
شبیه‌سازی‌ها برای ۵۰ بار تکرار شده‌اند و نتایج RMSE تمام موقعیت‌های منبع سیگنال در محیط بررسی شده است. در تمام مراحل آزمایش، به منظور تعادل میان اکتشاف و بهره‌برداری از محیط، مقدار ε برابر ۰/۵ تنظیم شده است. در نمودارهای نمایش مسیر ردیابی، نقاط سبزرنگ مسیر واقعی حرکت منبع سیگنال و نقاط قرمزرنگ موقعیت‌های تخمین‌زده‌شده در الگوریتم‌ها هستند. نحوهٔ محاسبهٔ RMSE برای منبع سیگنال متحرک در (۲۸) بیان شده است که در آن M تعداد تکرار آزمایش است.

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{j=1}^{M}\left(\hat{x}_{Tj}-x_{Tj}\right)^2+\left(\hat{y}_{Tj}-y_{Tj}\right)^2}{M}} \qquad (۲۸)$$

نمودار RMSE منبع سیگنال در کل عملیات ردیابی در شکل (۹) و شکل (۱۰) نشان داده شده است. این نمودارها اطلاعات RMSE در طول مسیر حرکت منبع سیگنال را به‌ترتیب در صورت عدم وجود و وجود فیلتر کالمن در الگوریتم‌ها نمایش می‌دهند. طبق نتایج جدول (۵)، میانگین RMSE کل مسیر حرکت توسط الگوریتم توسعه‌یافته با فیلتر کالمن ٤۲ درصد کاهش یافته است.



شکل (۹): نمودار RMSE موقعیت منبع سیگنال بدون فیلتر کالمن در ردیابی با ۵۰ آزمایش



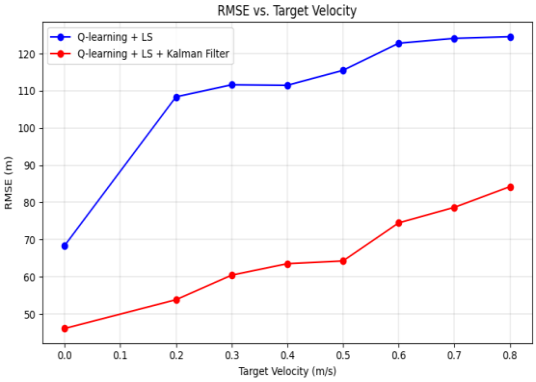
شکل (۱۰): نمودار RMSE موقعیت منبع سیگنال در صورت وجود فیلتر کالمن در الگوریتم‌ها با ۵۰ آزمایش

جدول (۵): بررسی تأثیر فیلتر کالمن بر RMSE کل مسیر منبع سیگنال متحرک با ۵۰ بار آزمایش

RMSE ردیابی منبع سیگنال متحرک بدون وجود فیلتر کالمن در الگوریتم (متر)	۱۰۷
RMSE ردیابی منبع سیگنال متحرک در صورت وجود فیلتر کالمن در الگوریتم (متر)	٦۲

به منظور ارزیابی عملکرد الگوریتم تحت شرایط

دینامیکی متنوع، آزمایش‌هایی با مقادیری مختلف از سرعت هدف انجام شدند. طبق نتایج نشان‌داده‌شده در شکل (۱۱)، با افزایش سرعت حرکت هدف، مقدار RMSE نیز به صورت تقریباً خطی افزایش می‌یابد. این رویداد از منظر نظری نیز قابل توجیه است، زیرا با افزایش سرعت هدف، پیش‌بینی موقعیت آن دشوارتر می‌شود و همین باعث افزایش خطای تخمین می‌شود. همچنین، همان‌طور که مشاهده می‌شود، حضور فیلتر کالمن در الگوریتم باعث کاهش خطا می‌شود.



شکل (۱۱): تغییرات RMSE تخمین مکان برحسب سرعت‌های مختلف هدف با ۵۰ آزمایش

علاوه بر این، تأثیر نسبت سرعت پهپاد به سرعت هدف بر دقت موقعیت‌یابی و مصرف انرژی نیز بررسی شده است. نتایج نشان می‌دهد با افزایش این نسبت، خطای موقعیت‌یابی کاهش می‌یابد؛ به گونه‌ای که با دو برابر شدن سرعت پهپاد نسبت به هدف، RMSE حدود ۱۹ درصد کاهش یافته است. با این حال، این بهبود دقت با افزایش چشمگیر مصرف انرژی همراه بوده است؛ به طوری که میانگین انرژی مصرفی ۵/٤ برابر شده است.

۵-۱- پیچیدگی محاسباتی و حافظهٔ لازم

پیچیدگی زمانی الگوریتم یادگیری Q در این پژوهش به‌ازای هر گام یادگیری برابر $O(|A|)$ است که در آن $|A|$ نشان‌دهندهٔ تعداد اقدامات مجاز برای عامل است. با توجه به اینکه در این مسئله عامل ۹ اقدام ممکن دارد، پیچیدگی زمانی هر گام برابر $O(9)$ است. بنابراین، از منظر زمانی، هر گام یادگیری بسیار سبک و قابل اجرا در زمان واقعی است. در نهایت، با در نظر گرفتن ۵۰۰۰ رویداد یادگیری، حداکثر ۱۱ گام در هر رویداد و ۹ اقدام ممکن در هر وضعیت، پیچیدگی زمانی کل فرایند یادگیری به صورت $O(9\times5000\times11)$ برآورد می‌شود.

فضای حالت الگوریتم یادگیری Q فقط شامل ۲۵ حالت است. همان‌طور که قبلاً بیان شد، به منظور پوشش کامل تمامی ترکیب‌های ممکن از موقعیت اولیهٔ پهپاد و منبع سیگنال، الگوریتم به صورت مستقل برای هر یک از ٦۲۵ حالت ممکن (۲۵ موقعیت اولیه برای پهپاد × ۲۵ موقعیت برای منبع سیگنال) آموزش داده شده است و مقادیر Q حاصل ذخیره شده‌اند. متوسط زمان مورد نیاز برای هر بار یادگیری حدود ۱۷ ثانیه بوده و در مجموع کل فرایند آموزش تقریباً ۱۰۷۰ ثانیه زمان برده است. اجرای الگوریتم یادگیری در محیط ویندوز ۱۰ (نسخهٔ ٦٤بیتی) با استفاده از زبان برنامه‌نویسی Python نسخهٔ ۳.۱۱.۰، بر روی سیستمی مجهز به پردازندهٔ Intel Core i5-4210U و حافظهٔ RAM به ظرفیت ۱۲ گیگابایت انجام شده است.

پیچیدگی فضایی الگوریتم یادگیری Q برابر $O(|S|\times|A|)$ است که در آن $|S|$ نشان‌دهندهٔ تعداد کل حالات محیط و $|A|$ تعداد اقدامات ممکن برای عامل است. این پیچیدگی ناشی از نیاز حافظه‌ای الگوریتم برای ذخیرهٔ جدول Q است که در آن برای هر جفت حالت–اقدام یک مقدار Q منحصربه‌فرد نگهداری می‌شود. بر این اساس، پیچیدگی فضایی الگوریتم پیشنهادی با فرض وجود ۲۵ حالت و ۹ اقدام مجاز برابر $O(25\times9)$ خواهد بود. با در نظر گرفتن تمامی ترکیب‌های ممکن از موقعیت اولیهٔ پهپاد و موقعیت منبع سیگنال، به مجموع ٦۲۵ جدول Q نیاز خواهد بود [۳۳].

۶- نتیجه‌گیری

در این پژوهش، روشی نوآورانه برای مکان‌یابی و ردیابی منابع سیگنال زمینی با استفاده از پهپاد و اندازه‌گیری RSS ارائه شد. ساختار پیشنهادی با ترکیب سه الگوریتم LS، یادگیری Q و فیلتر کالمن توسعه یافته است. این ترکیب امکان بهره‌برداری همزمان از مزایای هر الگوریتم را

فراهم کرده و دقت موقعیت‌یابی و ردیابی را بهبود داده است. الگوریتم یادگیری Q وظیفهٔ ناوبری بهینهٔ پهپاد را به عهده دارد؛ LS به عنوان هستهٔ تخمین موقعیت منبع عمل می‌کند. همچنین، فیلتر کالمن علاوه بر ایفای نقش در هموارسازی تخمین‌ها با پیش‌بینی وضعیت آیندهٔ منبع متحرک، کارایی سیستم را در شرایط پویا ارتقا می‌دهد.

نتایج شبیه‌سازی نشان می‌دهد ترکیب الگوریتم یادگیری Q با روش LS، با بهره‌گیری همزمان از قابلیت یادگیری و دقت تخمین، رویکردی کارآمد برای مکان‌یابی و ناوبری پهپاد فراهم کرده است. همچنین، ترکیب این روش‌ها با فیلتر کالمن در ردیابی منابع متحرک موجب کاهش ٤۲ درصدی RMSE در مقایسه با روش‌های بدون فیلتر کالمن شده است. در مجموع، چارچوب پیشنهادی با تکیه بر تلفیق هدفمند الگوریتم‌های یادگیری و تخمین، توانسته است راهکاری برای مکان‌یابی و ردیابی منابع سیگنال با استفاده از فقط یک پهپاد در شرایط محیطی پیچیده و نیمه‌شهری ارائه دهد. برای مطالعات آتی، پیشنهاد می‌شود از الگوریتم‌های RL عمیق به منظور بهینه‌سازی ناوبری پهپاد و افزایش دقت مکان‌یابی استفاده شود.

مراجع

[1] H. Lee, J. Seo, "Performance evaluation and hybrid application of the greedy and predictive UAV trajectory optimization methods for localizing a target mobile device", In Proceedings of the 2023 International Technical Meeting of The Institute of Navigation, pp. 161-171, 2023. https://doi.org/10.33012/2023.18666.

[2] Y. Chang, Y. Cheng, U. Manzoor, J. Murray, "A review of UAV autonomous navigation in GPS-denied environments", Robotics Autonomous Systems, Vol. 170, p. 104533, 2023. https://doi.org/10.1016/j.robot.2023.104533.

[3] M. I. M. Ismail, R. A. Dzyauddin, S. Samsul, N. A. Azmi, Y. Yamada, M. F. M. Yakub, N. A. B. A. Salleh, "An RSSI-based wireless sensor node localisation using trilateration and multilateration methods for outdoor environment", arXiv preprint arXiv:.07801, 2019. https://doi.org/10.48550/arXiv.1912.07801.

[4] R. A. Khalil, N. Saeed, M. Almutiry, "UAVs-assisted passive source localization using robust TDOA ranging for search and rescue", ICT Express, Vol. 9, No. 4, pp. 677-682, 2022. https://doi.org/10.1016/j.icte.2022.04.011.

[5] O. Esrafilian, R. Gangula, D. Gesbert, "UAV trajectory optimization and tracking for user localization in wireless networks", arXiv preprint arXiv:.14959, 2023. https://doi.org/10.48550/arXiv.2305.14959.

[6] Z. Memarian, M. Majidi, "2D DOA estimation of wideband and FH signals using improved K-means clustering and implementation considerations", Iranian Journal of Electrical and Electronic Engineering, Research Paper, Vol. 21, No. 3, pp. 3523-3523, 2025. https://doi.org/10.22068/IJEEE.21.3.3523.

[7] M. Arabsorkhi, H. Zayyani, M. Korki, "3-D hybrid RSS-AoA passive source localization with unknown path loss exponent", IEEE Sensors Letters, Vol. 7, No. 6, pp. 1-4, 2023. https://doi.org/10.1109/LSENS.2023.3282023.

[8] N. Güzey, "RF source localization using multiple UAVs through a novel geometrical RSSI approach", Drones, Vol. 6, No. 12, p. 417, 2022. https://doi.org/10.3390/drones6120417.

[9] Y. Shi, W. Shi, X. Liu, X. Xiao, "An RSSI classification and tracing algorithm to improve trilateration-based positioning", Sensors, Vol. 20, No. 15, p. 4244, 2020. https://doi.org/10.3390/s20154244.

[10] H. Sallouha, M. M. Azari, S. Pollin, "Energy-constrained UAV trajectory design for ground node localization", In IEEE Global Communications Conference (GLOBECOM), pp. 1-7, 2018. https://doi.org/10.1109/GLOCOM.2018.8647530.

[11] K. Witrisal, C. Anton-Haro, S. Grebien, W. Joseph, E. Leitinger, X. Li, …, T. Wilding, "Localization and Tracking", In Inclusive Radio Communications for 5G and Beyond: Elsevier, pp. 253-293, 2021. https://doi.org/10.1016/B978-0-12-820581-5.00015-8.

[12] M. Hasanzade, Ö. Herekoğlu, R. Yeniçeri, E. Koyuncu, G. İnalhan, "RF source localization using unmanned aerial vehicle with particle filter", In 9th International Conference on Mechanical and Aerospace Engineering (ICMAE), pp. 284-289, 2018. https://doi.org/10.1109/ICMAE.2018.8467555.

[13] D. Ebrahimi, S. Sharafeddine, P.-H. Ho, C. Assi, "Autonomous UAV trajectory for localizing ground objects: A reinforcement

learning approach", IEEE Transactions on Mobile Computing, Vol. 20, No. 4, pp. 1312-1324, 2020. https://doi.org/10.1109/TMC.2020.2966989.

[14]M. Shurrab, R. Mizouni, S. Singh, H. Otrok, "Reinforcement learning framework for UAV-based target localization applications", Internet of Things, Vol. 23, p. 100867, 2023. https://doi.org/10.1016/j.iot.2023.100867.

[15]S. Wu, "Illegal radio station localization with UAV-based Q-learning", China Communications, Vol. 15, No. 12, pp. 122-131, 2018. https://doi.org/10.12676/j.cc.2018.12.010.

[16]W. Li, Y. Xu, "Radio interference source localization with UAV-based Q-Learning", In 2024 IEEE 14th International Conference on Electronics Information and Emergency Communication (ICEIEC), pp. 98-102, 2024. https://doi.org/10.1109/ICEIEC61773.2024.10561748.

[17]B. S. Ciftler, A. Tuncer, I. Guvenc, "Indoor UAV navigation to a Rayleigh fading source using Q-learning", arXiv preprint arXiv:.10375, 2017. https://doi.org/10.48550/arXiv.1705.10375.

[18]M. M. U. Chowdhury, F. Erden, I. Guvenc, "RSS-based Q-learning for indoor UAV navigation", In IEEE Military Communications Conference (MILCOM), pp. 121-126, 2019. https://doi.org/10.1109/MILCOM47813.2019.9020894.

[19]S. Kulkarni, V. Chaphekar, M. M. U. Chowdhury, F. Erden, I. Guvenc, "UAV aided search and rescue operation using reinforcement learning", In SoutheastCon, Vol. 2, pp. 1-8, 2020. https://doi.org/10.1109/SoutheastCon44009.2020.9368285.

[20]D. Mandloi, R. Arya, "Q-learning-based UAV-mounted base station positioning in a disaster scenario for connectivity to the users located at unknown positions", The Journal of Supercomputing, Vol. 79, No. 14, pp. 15643-15674, 2023. https://doi.org/10.1007/s11227-023-05292-2.

[21]A. Becker, "Kalman filter from the Ground Up (1st Ed.)", 2023. [Online]. Available: https://www.kalmanfilter.net/default.aspx

[22]Y. Spyridis, T. Lagkas, P. Sarigiannidis, J. Zhang, "Modelling and simulation of a new cooperative algorithm for UAV swarm coordination in mobile RF target tracking", Simulation Modelling Practice Theory, Vol. 107, p. 102232, 2021. https://doi.org/10.1016/j.simpat.2020.102232.

[23]M. Effati, K. Skonieczny, "EKF and UKF localization of a moving RF ground target using a flying vehicle", In IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1-4, 2017. https://doi.org/10.1109/CCECE.2017.7946749.

[24]D. Soleymani and R. Havangi, "Target tracking in MIMO radar systems using interactive multiple extended Kalman filter and its optimization", Computational Intelligence in Electrical Engineering, Vol. 14, No. 2, pp. 95-110, 2023. https://doi.org/10.22108/isee.2022.132351.1539.

[25]E. Testi, E. Favarelli, A. Giorgetti, "Reinforcement learning for connected autonomous vehicle localization via UAVs", In IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), pp. 13-17, 2020. https://doi.org/10.1109/MetroAgriFor50201.2020.9277630.

[26]Y.-J. Chen, D.-K. Chang, C. Zhang, "Autonomous tracking using a swarm of UAVs: A constrained multi-agent reinforcement learning approach", IEEE Transactions on Vehicular Technology, Vol. 69, No. 11, pp. 13702-13717, 2020. https://doi.org/10.1109/TVT.2020.3023733.

[27]J. Moon, S. Papaioannou, C. Laoudias, P. Kolios, S. Kim, "Deep reinforcement learning multi-UAV trajectory control for target tracking", IEEE Internet of Things Journal, Vol. 8, No. 20, pp. 15441-15455, 2021. https://doi.org/10.1109/JIOT.2021.3073973.

[28]N. Parvaresh, B. Kantarci, "A continuous actor–critic deep Q-Learning-enabled deployment of UAV base stations: toward 6G small cells in the skies of smart cities", IEEE Open Journal of the Communications Society, Vol. 4, pp. 700-712, 2023. https://doi.org/10.1109/OJCOMS.2023.3251297.

[29]A. Al-Hourani, S. Kandeepan, S. Lardner, "Optimal LAP altitude for maximum coverage", IEEE Wireless Communications Letters, Vol. 3, No. 6, pp. 569-572, 2014. https://doi.org/10.1109/LWC.2014.2342736.

[30]A. Al-Hourani, S. Kandeepan, A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments", In IEEE Global Communications Conference, pp. 2898-2904, 2014. https://doi.org/10.1109/GLOCOM.2014.7037

1.
[39]G. Li, E. Geng, Z. Ye, Y. Xu, J. Lin, Y. Pang, "Indoor positioning algorithm based on the improved RSSI distance model", Sensors, Vol. 18, No. 9, p. 2820, 2018. https://doi.org/10.3390/s18092820.
[40]S. Särkkä, L. Svensson, "Bayesian Filtering and Smoothing (2nd Ed.)", Cambridge University Press, 2023.
[41]G. Welch, G. Bishop, "An Introduction to the Kalman Filter", Univ. of North Carolina, 1995.
[42]R. G. Brown, P. Y. Hwang, "Introduction to Random Signals and Applied Kalman Filtering with MATLAB Exercises (4th Ed.)", John Wiley & Sons, 2012.
[43]B. Raeisy, S. Golbahar Haghighi, A. A. Safavi, "Active noise control for narrow-band and broad-band signals using Q-Learning technique", Computational Intelligence in Electrical Engineering, Vol. 4, No. 1, pp. 70-57, 2013.
[44]Z. Rashidi, M. Majidi, "Energy and throughput management in wireless body area network with wireless information and energy transfer using reinforcement learning " International Journal of Engineering, Vol. 37, No. 11, pp. 2314-2324, 2024. https://doi.org/10.5829/ije.2024.37.11b.16.
[45]S. Xu, Y. Gu, X. Li, C. Chen, Y. Hu, Y. Sang, W. Jiang, "Indoor emergency path planning based on the Q-learning optimization algorithm", ISPRS International Journal of Geo-Information, Vol. 11, No. 1, p. 66, 2022. https://doi.org/10.3390/ijgi11010066.
[46]H. Sallouha, M. M. Azari, A. Chiumento, S. Pollin, "Aerial anchors positioning for reliable RSS-based outdoor localization in urban environments", IEEE Wireless Communications Letters, Vol. 7, No. 3, pp. 376-379, 2017. https://doi.org/10.1109/LWC.2017.2778723.
248.
[31]A. Filippone, "Flight Performance of Fixed and Rotary Wing Aircraft". Elsevier, 2006.
[32]Y. Zeng, R. Zhang, "Energy-efficient UAV communication with trajectory optimization", IEEE Transactions on Wireless Communications, Vol. 16, No. 6, pp. 3747-3760, 2017. https://doi.org/10.1109/TWC.2017.2688328.
[33]R. S. Sutton, A. G. Barto, "Reinforcement Learning: An Introduction (2nd Ed.)", MIT Press, 2018.
[34]A. T. Azar, A. Koubaa, N. Ali Mohamed, H. A. Ibrahim, Z. F. Ibrahim, M. Kazim, …, I. A. Hameed, "Drone deep reinforcement learning: A review", Electronics, Vol. 10, No. 9, p. 999, 2021. https://doi.org/10.3390/electronics10090999.
[35]F. H. Panahi, F. H. Panahi, "An intelligent energy-efficient firefighting strategy in mobile WSNs", Computational Intelligence in Electrical Engineering, Vol. 13, No. 3, pp. 37-54, 2022. https://doi.org/10.22108/isee.2021.124683.1406.
[36]Y. Li, "Deep reinforcement learning: An overview", arXiv:1810.06339, 2018. https://doi.org/10.48550/arXiv.1810.06339.
[37]M. Sheikh-Hosseini, S. R. Samareh Hashemi, "Target and areas coverage in wireless sensor networks using analytical and evolutionary algorithms", Computational Intelligence in Electrical Engineering, Vol. 13, No. 1, pp. 39-54, 2022. https://doi.org/10.22108/isee.2020.122866.1372.
[38]G. Han, J. Jiang, C. Zhang, T. Q. Duong, M. Guizani, G. K. Karagiannidis, "A survey on mobile anchor node assisted localization in wireless sensor networks", IEEE Communications Surveys Tutorials, Vol. 18, No. 3, pp. 2220-2243, 2016. https://doi.org/10.1109/COMST.2016.254475

[1] Global Positioning System
[2] Radio Frequency
[3] Received Signal Strength
[4] Time of Arrival
[5] Time Difference of Arrival
[6] Angle of Arrival
[7] Reinforcement Learning
[8] Illegal Radio Station
[9] Autonomous Air Vehicle

[10] Extended Kalman filter
[11] Unscented Kalman Filter
[12] Deep Reinforcement Learning
[13] Least Square
[14] Line of Sight
[15] Non-Line of Sight
[16] Drag
[17] Parasite Power
[18] Markov Decision Process
[19] Exploitation
[20] Exploration
[21] Epsilon Greedy Policy
[22] Wireless Body Area Network
[23] Episode
[24] Decay Rate