Computational Intelligence in Electrical Engineering
Vol. 13, No. 3, 2022
Research Paper

An Intelligent Energy-Efficient Firefighting Strategy in Mobile WSNs

Farzad H. Panahi [1], Fereidoun H. Panahi [2]

[1] Department of Electrical Engineering, Faculty of Engineering, University of Kurdistan, Sanandaj, Iran
[2] Department of Electrical Engineering, Faculty of Engineering, University of Kurdistan, Sanandaj, Iran

Abstract:
With the increased scope of fires and the widespread destruction of the environment and densely populated urban areas in recent years, researchers have investigated the adoption of rapid and effective firefighting solutions, particularly those based on wireless sensor networks (WSNs). In fact, by evaluating various statistical data and developing a new model of sensors, equipment, and intelligent technologies in a fire sensor network, an effective step toward controlling frequent fires on a wide scale and reducing environmental damage can be taken. In the proposed model, the mobile sensors or firefighting robots based on a fuzzy Q-learning (FQL) algorithm and using two learning strategies in the sensor network, namely partial and perfect, could be used to surround fire in firefighting operations. We also formulate a sensor mode selection strategy as an optimization problem to maximize the lifetime of the energy harvesting-enabled WSN. Furthermore, we determine optimal upper and lower bounds for the number of active sensors in the fire detection system, guaranteeing that the target detection and false alarm probabilities are achieved. Computer simulations show that using such a solution in the optimal selection of moving sensors and determining the moving trajectory in rapid firefighting is effective.

Keywords: Wireless Sensor Networks, Intelligent Firefighting Strategy, Mobile Sensors, Fire, Optimization.

http://dx.doi.org/10.22108/isee.2021.124683.1406

مقاله پژوهشی

استراتژی هوشمند و انرژی کارآمد اطفای حریق در شبکه‌های حسگر بی‌سیم متحرک

فرزاد حسین‌پناهی[۱]، فریدون حسین‌پناهی*[۲]

۱- گروه مهندسی برق-الکترونیک و مخابرات، دانشکده مهندسی، دانشگاه کردستان، سنندج، ایران
f.hpanahi@uok.ac.ir

۲- گروه مهندسی برق-الکترونیک و مخابرات، دانشکده مهندسی، دانشگاه کردستان، سنندج، ایران
fereidoun.h.panahi@uok.ac.ir

چکیده: پژوهشگران در سال‌های اخیر با افزایش دامنهٔ آتش‌سوزی‌ها و به‌همراه آن تخریب گستردهٔ محیط زیست و مناطق شهری پرتراکم، به به‌کارگیری راهکارهای سریع و مؤثر در مقابله با حریق، به‌ویژه براساس شبکه‌های حسگر بی‌سیم توجه ویژه‌ای داشته‌اند. درواقع، با تحلیل داده‌های آماری مختلف و طراحی یک مدل نوین از حسگرها، تجهیزات و تکنولوژی‌های هوشمند در یک شبکهٔ حسگر آتشنشان می‌توان گام مؤثری در راستای کنترل آتش‌سوزی‌های مکرر در سطح گسترده و نیز کاهش خسارت‌های زیست‌محیطی آن برداشت. در مدل پیشنهادی، حسگرهای متحرک یا روبوت‌های اطفای حریق بر پایهٔ الگوریتم یادگیری فازی - کیو و به کمک دو سیاست یادگیری کامل و جزئی در شبکهٔ حسگر به محاصرهٔ آتش در عملیات اطفای حریق قادر خواهند بود. در این مدل، محدودیت‌های انرژی در حسگرهای متحرک نیز با طراحی مسئلهٔ بهینهٔ انتخاب مُد عملکرد و با فرض قابلیت برداشت انرژی‌های محیطی قبل از کنترل حرکت به سمت حریق در نظر گرفته شده‌اند که با محاسبهٔ کران‌های بالا و پایین برای تعداد حسگرهای ثابت فعال در تصمیم‌گیری مشارکتی، میزان مطلوب احتمالات آشکارسازی و اعلام اشتباه حریق نیز تضمین‌شدنی است. نتایج شبیه‌سازی‌های کامپیوتری، مؤثربودن اعمال چنین راهکاری در انتخاب بهینهٔ حسگرهای متحرک و همچنین تعیین مسیر حرکت در اطفای سریع حریق را نشان می‌دهند.

واژه‌های کلیدی: شبکه‌های حسگر بی‌سیم، استراتژی هوشمند اطفای حریق، حسگرهای متحرک، آتش، بهینه‌سازی.

۱- مقدمه

آتش‌سوزی، یکی از خطرناک‌ترین پدیده‌هایی است که با زیان‌های شایان توجه جانی و مالی همراه است. همه‌روزه آتش‌سوزی‌های بسیاری در نقاط مختلف جهان رخ می‌دهد که موجب از بین رفتن انسان‌ها و به بار آمدن زیان‌های فراوان زیست‌محیطی می‌شود. با توجه به اینکه اقدامات شناسایی و اطفای حریق در لحظات اولیهٔ شروع حریق دارای اهمیت زیاد و حیاتی است، امروزه طراحی سیستم‌های سریع و کارا بسیار شایان توجه قرار گرفته است. در سال‌های اخیر نیز تحقیقات فراوانی بر این مسئله، به‌ویژه در سطح جنگل‌ها و مناطق شهری پرتراکم متمرکز شده است. کشورهای مختلف با برنامه‌ریزی و سرمایه‌گذاری در بخش فناوری اطلاعات و ارتباطات و نیز گسترش شبکه‌های حسگر بی‌سیم[۱] (WSN)، پروژه‌های مختلفی را برای حفاظت منابع طبیعی در مقابله با آتش‌سوزی به اجرا درآورده‌اند و بدون تردید در آیندهٔ نزدیک، شبکه‌های حسگر بی‌سیم و اینترنت اشیا[۲] به‌عنوان فناوری‌های کلیدی در توسعهٔ شبکه‌های نظارت محیطی مطرح خواهند شد [۱-۴]. در سال‌های گذشته، شبکه‌های حسگر بسیاری با پوشش رادیویی گسترده، به‌ویژه مبتنی بر فناوری زیگبی[۳] و فراپهن باند[۴] (UWB) پیشنهاد داده شده‌اند [۵،۶]. مساحت درخور

---

[۱] تاریخ ارسال مقاله: ۱۳۹۹/۰۶/۱۳
تاریخ پذیرش مقاله: ۱۴۰۰/۰۴/۰۶
نام نویسندهٔ مسئول: فریدون حسین‌پناهی
نشانی نویسندهٔ مسئول: : ایران، سنندج، دانشگاه کردستان، دانشکده مهندسی، گروه مهندسی برق - الکترونیک و مخابرات

توجهی از کرهٔ زمین را کوهستان‌ها، مناطق جنگلی و شهری پرتراکم تشکیل داده است و معمولاً دسترسی به این مناطق به شدت سخت‌تر از مناطق بیابانی خواهد بود. همچنین، این مناطق گسترده اکثراً جمعیت یا پوشش گیاهی پرتراکم دارند که مستعد آتش‌سوزی‌اند. استفاده از شیوه‌های نوین در اطفای حریق و بررسی هوشمند مناطق مذکور می‌تواند راهکار بسیار مناسب، سریع و دقیق برای آتشنشان‌ها باشد. واضح است برای اینکه در مناطق با دسترسی سخت از وقوع آتش اطلاع حاصل شود، مراجعهٔ حضوری بسیار سخت و در بیشتر مواقع کند و پرهزینه است؛ بنابراین، کاهش حضور فیزیکی به‌ویژه در مناطق صعب‌العبور و مدیریت بهینهٔ اقدامات اطفای حریق در لحظات اولیهٔ آتش‌سوزی، محوری‌ترین نکات در بیان ضرورت انجام تحقیقات در این زمینه است. با این توضیحات، امروزه مدل‌سازی یک سامانهٔ کاربردی هوشمند برای شناسایی و کنترل سریع آتش‌سوزی، یک انتخاب نیست، بلکه یک ضرورت است.

	واژگان اختصاری
BS	Base Station
CH	Cluster Head
DC	Decay Coefficient
EEP	Exploration-Exploitation Policy
EH	Energy Harvesting
FQL	Fuzzy Q-Learning
FIS	Fuzzy Inference System
HetNet	Heterogeneous Network
IoT	Internet of Things
MARL	Multi-Agent Reinforcement Learning
MAS	Multi-Agent Systems
PEL	Perfect Learning Policy
PAL	Partial Learning Policy
QL	Q-Learning
RL	Reinforcement Learning
SN	Sensor Node
SE	Square Error
TS	Takagi-Sugeno
UWB	Ultra Wide-Band
WSN	Wireless Sensor Network

۲- پیشینهٔ تحقیق

در راهکارهای ارائه‌شده برای شناسایی حریق، از اطلاعات حسگرهای ثابت و متحرک [۱۰-۷] جمع‌آوری داده در سطح ناحیهٔ تحت پوشش شبکهٔ حسگر استفاده می‌شود که در سریع‌ترین زمان ممکن عوامل و نقاط آتش‌سوزی را تخمین می‌زنند و به پایگاه‌های آتشنشانی اطلاع می‌دهند. به این ترتیب، پیش‌بینی ابعاد آتش‌سوزی و درنتیجه روند عملیات اطفای حریق با برنامه‌ریزی دقیق‌تری امکان‌پذیر خواهد شد. در مواقعی که آتش‌سوزی در مکان‌های صعب‌العبور رخ داده باشد، به‌کارگیری حسگرهای متحرک یا پهپادهای آتشنشان با قابلیت حمل مواد ضد آتش ضروری است. با تحلیل و طراحی یک مدل نوین از حسگرها، تجهیزات و تکنولوژهای هوشمند به‌ویژه برپایهٔ سیستم‌های چندعامل[۵] (MAS) در یک شبکهٔ حسگری آتشنشان، می‌توان گام مؤثری در راستای کنترل آتش‌سوزی‌های مکرر و خسارت‌های زیست‌محیطی فراوان آن برداشت.

سیستم‌های چندعامل شامل چندین عامل[۶] هوشمند در یک محیط‌اند که هرکدام رفتار مستقلی دارند و با دیگر عامل‌ها هماهنگ‌اند [۱۰،۱۱]. این سیستم‌ها می‌توانند به‌عنوان روش جایگزین سیستم‌های متمرکز آتشنشان ظاهر شوند که در آنها چندین عامل هوشمند، وقوع آتش‌سوزی در یک محیط را ازطریق سنسورهای خود درک می‌کنند و رفتارهای مستقل از خود نشان می‌دهند [۱۲]. یکی از زمینه‌هایی که به‌تازگی سیستم‌های چندعاملی را در تحقیقات و مسائل کاربردی مطرح کرده است، مسئلهٔ کنترل و هدایت حسگرها یا روبوت‌های متحرک مبتنی بر سیستم‌های کنترل سطح پایین[۷] است. در مرجع [۱۳] یک الگوریتم مبتنی بر درخت باینری برای طراحی مسیر حرکت روبوت‌ها استفاده شده است. این نوع رویکرد مبتنی بر آگاهی کامل از دینامیک سیستم است و این ساختار با توجه به محدودیت‌های زمانی و مکانی در مسیریابی حریق، قابل پیاده‌سازی نیست. یادگیری تقویتی[۸] (RL) یکی از محبوب‌ترین روش‌های یادگیری در سیستم‌های چندعامل یا MAS است. هدف از یادگیری تقویتی چندعامل[۹] (MARL) به حداکثر رساندن مقادیر تجمعی پاداش[۱۰] است. به این ترتیب، عامل‌ها می‌توانند با محیط ارتباط برقرار کنند و آن را مطابق با الگوی پاداش تغییر دهند. هر عامل در هر مرحلهٔ یادگیری، یک کنش یا اقدام[۱۱] را انتخاب می‌کند و محیط را به سمت حالت[۱۲] جدید سوق می‌دهد [۱۴]. در این

فرایند، تابع پاداش[۱۳] همواره کیفیت گذر حالت را ارزیابی می‌کند [۱۵]. در هر صورت، عملکرد سیستم‌های چندعامل همواره تأثیرگرفته از ابعاد مسئله است و با افزایش تعداد حالت‌ها یا عامل‌ها، به محاسبات و حافظهٔ بیشتری نیاز خواهد داشت. در بیشتر رویکردها، به بازنمایی دقیق مقادیر جفت حالت – کنش به فرم جداول جستجو نیاز است که این مسئله به‌منزلهٔ یک مانع بزرگ، کاربرد این روش‌ها را به مسائل کوچک یا گسسته تقلیل داده است [۱۶]. واضح است در عملیات اطفای حریق، متغیرهای حالت می‌توانند مقادیر با تنوع بیشتر و در یک بازهٔ پیوسته را به خود بگیرند [۹]. این مشکل با تقریب توابع ارزش[۱۴] مدیریت می‌شود [۱۷]. برای مقابله با چنین مشکلی، الگوریتم‌های MARL برپایهٔ شبکه‌های عصبی نیز پیشنهاد شده‌اند که براساس مفهوم تعمیم‌یافتهٔ جدول کیفیت یا Q [۱۸]، تقریب گسسته برای فضای بزرگ حالت – کنش [۱۹]، کوانتیزاسیون برداری برای حالت‌ها یا کنش‌های پیوسته [۲۰]، تکرار تجربه برای MAS [۲۱]، تقریب مبتنی بر یادگیری Q و شبکهٔ گاوسی نرمالیزه [۲۲] و پیش‌بینی پارامترها در عامل‌های ناهمگن [۲۳] نتایج پذیرفتنی حاصل شده است. در مراجع [۲۴،۲۵] نیز از یک مدل دوگانهٔ شبکهٔ عصبی برای نشان‌دادن تابع ارزش و کنترل کننده بهره گرفته شده است. با این حال، موفقیت استراتژی‌های پیشنهادی بستگی زیادی به کاوش کافی دارد و خود این مسئله نیز تابعی از اندازهٔ شبکهٔ عصبی و داده‌های آموزش شبکه است. در تحقیق کنونی ضمن پیشنهاد استراتژی حرکتی مبتنی بر MARL برای حسگرهای متحرک یا روبوت‌های آتشنشان و همچنین ارائهٔ استراتژی بهینهٔ انتخاب این روبوت‌ها براساس محدودیت انرژی، مدل واقعی‌تری از شبکهٔ اطفای حریق در نظر گرفته شده است. درواقع، استراتژی حرکتی یک نسخهٔ اصلاح‌شده برپایهٔ الگوریتم یادگیری Q [۲۶،۲۷] است که در آن تقریب فازی – خطی فضای حالت پیوسته اعمال شده است.

به این ترتیب ساختارها و الگوریتم‌های مطرح‌شده در زمینهٔ شناسایی و اطفای حریق در مراجع مختلف [۴،۵،۱۲،۳۰] در شرایطی ارتقا داده شده‌اند که یا توپولوژی شبکهٔ حسگری در فرایندهای مربوطه ثابت بوده یا روش‌های متمایزی [۱۸–۲۵] با محدودیت‌های مذکور در پیش گرفته شده است؛ اما در عمل، پدیدهٔ حریق نسبتاً تصادفی و آنی است که هم پیاده‌سازی الگوریتم‌های بهینه‌سازی سبُک در عملیات شناسایی و هم تغییر سریع توپولوژی شبکه و چینش گره‌های حسگری در عملیات اطفای حریق ضروری است؛ واقعیتی که به شدت بر دامنه و پراکندگی حریق اثر خواهد گذاشت و در پژوهش‌های مختلف ازجمله مرجع [۳۰] نادیده گرفته شده است. بنابراین، ارائهٔ یک مدل شبکه با ویژگی‌های فوق شامل استراتژی‌های شناسایی مبتنی بر حسگرهای ثابت و اطفای حریق مبتنی بر حسگرها یا روبوت‌های متحرک، اساس نوآوری در تحقیق کنونی است. درواقع، تأثیر الگوریتم‌های مدیریت حسگرها و عامل‌های متحرک با کاربردهای شناسایی یا اطفای حریق مطالعه شده‌اند؛ اما تا کنون پژوهشی مبنی بر مدل‌سازی فرایند هوشمند اطفای حریق بر پایهٔ انتخاب و هدایت بهینهٔ روبوت‌های آتشنشان با قابلیت برداشت انرژی محیطی مبتنی بر فاکتور سطح انرژی و الگوریتم یادگیری فازی – کیو در یک شبکهٔ حسگری با توپولوژی دینامیک و پویا صورت نگرفته است؛ بنابراین در مدل پیشنهادی، تأثیر فرایند مشارکتی شناسایی حریق با حسگرهای ثابت و عملیات بهینهٔ اطفای حریق با حسگرهای متحرک یا روبوت‌های آتشنشان بر پایهٔ الگوریتم یادگیری فازی – کیو و به کمک دو سیاست یادگیری کامل و جزئی در شبکهٔ حسگری با توپولوژی پویا مطالعه خواهند شد. در این تحقیق، محدودیت‌های انرژی در حسگرها نیز با طراحی مکانیزیم بهینهٔ انتخاب روبوت‌های آتشنشان و با فرض قابلیت برداشت انرژی‌های محیطی قبل از کنترل حرکت به سمت حریق، در نظر گرفته خواهند شد. با توضیحات فوق در مجموع، نوآوری اصلی این مقاله در سه بخش تفکیک می‌شود:

- محاسبهٔ کران‌های بالا و پایین برای تعیین تعداد حسگرهای ثابت فعال در تصمیم‌گیری مشارکتی و با هدف دستیابی به احتمالات مشارکتی آشکارسازی و اعلام اشتباه حریق مطلوب.
- محاسبهٔ احتمال بهینهٔ انتخاب مُد عملکرد روبوت‌های آتشنشان در یک شبکهٔ حسگری دینامیک به صورت‌های «شارژ یا برداشت انرژی» یا «حرکت» و

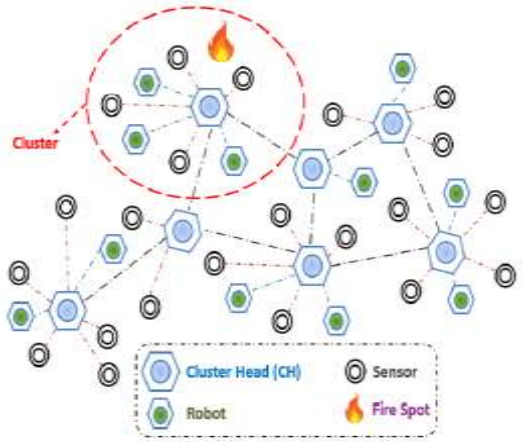
براساس تعریف یک مسئلهٔ سبُک بیشینه‌سازی طول عمر شبکه[۱۵] با محدودیت انرژی روبوت‌ها.

- طراحی استراتژی کنترل حرکت با هدف محاصره و درنهایت اطفای سریع حریق برای سیستم MAS متشکل از حسگرهای متحرک و براساس الگوریتم فازی مبتنی بر یادگیری Q.

۳- مدل شبکه حسگر بی‌سیم

در اینجا یک شبکهٔ حسگر با ساختار سلسله‌مراتبی دولایهٔ ناهمگن[۱۶] (HetNet) متشکل از ایستگاه‌های پایه[۱۷] (BSs)، سرخوشه‌ها[۱۸] (CHs) یا چاهک‌ها[۱۹] جمع‌آوری و پردازش اولیهٔ داده‌ها و نیز گره‌های حسگر[۲۰] (SNs) ثابت و متحرک توزیع‌شده در سراسر شبکه به‌عنوان مدل پایه مطابق با شکل (۱) تعریف می‌شود. براساس این، حسگرهای شبکه به برقراری ارتباط مستقیم با نزدیک‌ترین ایستگاه پایه نیازی ندارند؛ بلکه در این ساختار، حسگرها به خوشه‌ها یا سلول‌هایی تفکیک می‌شوند که در هر خوشه یا کلاستر، یک چاهک یا سرخوشه انتخاب می‌شود. سرگروه‌ها وظیفهٔ جمع‌آوری اطلاعات حسگرهای ثابت هر گروه را بر عهده دارند و در حقیقت نقش رله‌های ارتباطی به‌عنوان واسطه‌های انتقال اطلاعات بین حسگرها و ایستگاه‌های پایه را ایفا می‌کنند. این کار با هدف کاهش ترافیک اطلاعات ارسالی از حسگرها به ایستگاه پایه و درنتیجه، بهبود بازده انرژی شبکه انجام می‌شود. معیارهای مختلف انتخاب سرخوشه و مدیریت پویای توپولوژی شبکه در تحقیقات بسیاری بحث شده‌اند [۲۸]. در مدل ارائه‌شده، هر حسگر دارای یک ناحیهٔ پوشش یا شعاع حسگری است که به نقاط موجود در آن محدوده احاطهٔ کامل دارد. یکی از اهداف شبکه‌های حسگری این است که پوشش حداکثری در یک فضای معین تأمین شود.

در این مدل، گره‌های حسگری براساس کارکردهای متفاوت، شناسایی و اطفای حریق به‌ترتیب به دو دستهٔ حسگرهای ثابت و متحرک (روبوت‌های آتشنشان) تفکیک می‌شوند؛ به طوری که در یک خوشهٔ مفروض به تعداد N_S حسگر و N_r روبوت وجود دارد. در این مقاله، به‌طور کلی برای به حداقل رساندن مصرف انرژی حسگرهای ثابت ویژهٔ شناسایی حریق از دو حالت خواب و بیداری و همچنین با هدف افزایش طول عمر روبوت‌های اطفای حریق دو مد عملکرد «حرکت» و «برداشت انرژی» تعریف شده است. گفتنی است در عمل بر طبق سیاست بهینهٔ اطفای حریق، روبوت‌های اطفای حریق با مد عملکرد «حرکت»، در یک خوشه و به شکل محلی اقدام به محاصرهٔ آتش خواهند کرد و درنهایت پس از عملیات اطفای حریق پیکره‌بندی مجدد شبکهٔ حسگری صورت می‌گیرد.



شکل (۱): مدل پایهٔ شبکهٔ حسگر با ساختار سلسله‌مراتبی شامل سرخوشه‌ها یا چاهک‌ها و حسگرهای ثابت و متحرک درون شبکه

٤- تعیین استراتژی و مسئلهٔ بهینه‌سازی

در مدل پیشنهادی، حسگرهای متحرک یا روبوت‌های آتشنشان بر پایهٔ الگوریتم یادگیری فازی - کیو و به کمک تعریف دو سیاست یادگیری کامل[۲۱] (PEL) یا یادگیری جزئی[۲۲] (PAL) در شبکهٔ حسگری به محاصرهٔ آتش در عملیات اطفای حریق قادر خواهند بود. در سیاست PEL اولویت با یادگیری سریع زاویهٔ حرکت روبوت منتخب (θ_F) نسبت به شعاع حرکتی آن (R_F) در هنگام حرکت به سمت آتش است؛ اما در سیاست PAL فرایند یادگیری دو فاکتور مذکور در طول مسیر و به تدریج صورت می‌گیرد. در مدل پیشنهادی، محدودیت‌های انرژی در روبوت‌ها نیز با طراحی مکانیزم بهینهٔ انتخاب مُد عملکرد و با فرض قابلیت برداشت انرژی[۲۳] (EH) محیطی قبل از کنترل حرکت

بـه سـمت حریـق در نظـر گرفتـه خواهنـد شـد. نتـایج شبیه‌سازی‌های کامپیوتری، مؤثربودن اعمال چنین راهکـاری را در انتخاب بهینهٔ مُـد عملکـرد روبـوت‌هـای آتشنشـان و همچنین طراحی مسیر بهینهٔ حرکت در اطفای سریع حریـق نشان می‌دهند. در اینجا اطلاعات یـا گـزارش‌هـای ارسـالی حسگرها به سرخوشه‌ها برای آشکارسازی مطمئن آتـش در یک منطقهٔ خاص و با فرض احتمـال اعـلام اشـتباه اسـتفاده می‌شوند. در این موارد با توجه به شرایط محیطـی و تـراکم حسـگرها معمـولاً از تسـت نظریـهٔ بـاینری[۲۴] بهـره گرفتـه می‌شود. به این ترتیب هر حسگر تصمیم باینری خـود را از میان حالات وقوع حریق یا عـدم وقـوع حریـق و براسـاس پـردازش سـیگنال‌هـای محیطـی ماننـد دمـا، بـا احتمـال آشکارسازی p^D و احتمال اعـلام اشـتباه p^F ثبـت خواهـد کرد. در اینجا فرض می‌شود تصمیمات باینری مربوط به N_s حسگر از پدیدهٔ وقوع حریق در یک خوشهٔ مدنظر، مسـتقل از هم و به‌ترتیب با احتمالات آشکارسـازی و اعـلام اشـتباه حریق p_i^D و p_i^F $(i \in \{1, \dots, N_s\})$ هستند و اتخاذ تصمیم مشارکتی نهایی براساس ترکیب گزارشات دریافـت‌شـده در سرخوشه مطابق با روابط (۱) و (۲) صورت خواهد گرفت:

$$p_C^D = 1 - \prod_{i=1}^{N_s} \left(1 - A_i . p_i^D\right) \geq p_{TH}^D \qquad (۱)$$

$$p_C^F = 1 - \prod_{i=1}^{N_s} \left(1 - A_i . p_i^F\right) \leq p_{TH}^F \qquad (۲)$$

که در آن $A_i \in \{0,1\}$ شاخص مشـارکت حسـگر i در تصـمیم‌گیـری اسـت و بـه عبـارت دیگـر مـدهای خـواب (غیرفعال) یا بیداری (فعال) به‌ترتیب برابر با مقادیر صـفر و یک هستند. با این توضیحات فرض می‌شود در یـک لحظـهٔ خاص تعداد حسگرهای فعال با شاخص مشـارکت یـک در یک خوشه برابر با M است؛ بنابراین، براساس روابـط (۱) و (۲) و با فرض یکسان‌بودن عملکرد همه حسـگرهای ثابـت تحـت پوشـش یـک خوشـه $(p_i^D = p^D$ و $p_i^F = p^F)$، کران‌های بالا و پایین برای مقادیر مطلوب M برای دسـتیابی به احتمالات مشارکتی آشکارسـازی و اعـلام اشـتباه حریـق معین استخراج می‌شوند:

$$\frac{\log(1 - p_{TH}^F)}{\log(1 - p^F)} \leq M = \sum_{i=1}^{N_s} A_i \leq \frac{\log(1 - p_{TH}^D)}{\log(1 - p^D)} \qquad (۳)$$

واضح است این میزان مشـارکت حسـگرهای ثابـت در یک خوشه براساس مکانیسم‌های متداول خـواب و بیـداری حسگرهای ثابت و غیرمتحرک در یک شبکهٔ حسگری ایستا در تحقیقات انجام‌شده مانند مرجع [۳۰] تحقق‌پذیر است. به ایـن ترتیـب، هـر سرخوشـه قـادر اسـت براسـاس تجمیـع سیگنال‌های دریافتی از M حسگر بیدار از میان N_s حسـگر ثابت نسبت به تعیین مُـد عملکـرد N_r حسـگر متحـرک یـا روبوت آتشنشان اقدام کند.

۴-۱- استراتژی بهینهٔ انتخـاب مُـد عملکـرد روبوت‌ها

در این بخش، یک مسئلهٔ انتخاب مُد عملکـرد روبـوت آتشنشان در شبکهٔ حسگر بی‌سیم مبتنی بر برداشـت انـرژی مطرح می‌شود؛ به طوری که بیشینهٔ طول عمر شبکه تضمین شود. درواقع، هر سرخوشه در تـلاش اسـت روبـوت‌هـای آتشنشـان مربوطـه را در دو مـدل عملکـردی حرکـت یـا برداشت انرژی دسته‌بندی کند. به این ترتیـب، یـک مسـئلهٔ سبُک بیشینه‌سازی طول عمر شـبکه بـا محـدودیت کیفیـت سرویس و طول عمر روبوت به‌عنوان یک مسئلهٔ بهینه‌سازی در نظر گرفته می‌شود. سپس یک احتمال مُد بهینه براسـاس چـارچوب رایـج در مسـائل بهینـه‌سـازی محـدب پیشـنهاد می‌شود. در اینجا فرض شده اسـت روبـوت‌هـای آتشنشـان دارای قابلیت‌های حرکت و برداشت انرژی محیطی به‌ویـژه انرژی خورشیدی [۲۹،۳۰] هستند و در بازهٔ زمانی اتخاذ این استراتژی فقـط یکـی از ایـن مُـدهای عملکـرد اشـاره‌شـده فعال‌اند. به عبـارت دیگـر، در زمـان انتخـاب روبـوت‌هـای آتشنشان، روبوت‌هایی که سطح انرژی قابل قبـول دارنـد بـا احتمال بیشتری برای حرکت به سـمت مکـان آتـش‌سـوزی انتخاب می‌شوند و سایر روبوت‌های خوشهٔ مربوطـه در مُـد برداشت یا شـارژ انـرژی خواهنـد بـود. واضـح اسـت ایـن انتخاب‌ها در حالت ایدئال با مسئلهٔ بهینه‌سازی انجام‌پذیرند. این امکان با در نظر گرفتن یک حد آستانه برای هر روبـوت

آتشنشان در انتهای این بخش فراهم خواهد شد و به این ترتیب محدودیت‌های عملی نیز در اعمال این استراتژی تا حد زیادی برداشته خواهند شد. با این تفاسیر، مسئلهٔ بهینه‌سازی اشاره‌شده به‌صورت رابطهٔ (٤) نوشته می‌شود که مبنای استراتژی بهینهٔ انتخاب مُد عملکرد (حرکت یا برداشت انرژی) هر روبوت آتشنشان است:

$$\begin{aligned} &\max_{p_i,q_i} \left\{ \min_i \{E_i^{t-1} + p_i.H_i^t - q_i.M_i^t\} \right\} \\ &s.t. \quad E_i^{t-1} + p_i.H_i^t - q_i.M_i^t > E_{TH} \\ &p_i, q_i \in \{0,1\}, i \in \{1, \dots, N_r\} \end{aligned} \quad (۴)$$

که تعریف پارامترها و نمادهای اشاره‌شده، در جدول (۱) آمده است. محدودیت‌های مسئلهٔ تعریف‌شده شامل یک حد آستانهٔ کمینه برای انرژی در هر روبوت آتشنشان، مقادیر گسسته برای پارامترهای بهینه‌سازی p_i و q_i از مجموعه اعداد $\{0,1\}$ و درنهایت متریک‌های کنترل کمینهٔ کیفیت آشکارسازی جمعی (p_{TH}^D) و بیشینهٔ احتمال اعلام اشتباه حریق جمعی (p_{TH}^F) در یک خوشه است. پارامترهای گسستهٔ تعریف‌شده، مسئلهٔ بالا را به یک مسئلهٔ برنامه‌ریزی صحیح[۲۵] تبدیل کرده است که در این مسائل معمولاً با نگاشت پارامترهای بهینه‌سازی p_i و q_i به بازهٔ پیوسته $[0,1]$، می‌توان به یک مسئلهٔ بهینه‌سازی ساده‌شده[۲۶] به‌صورت زیر دست یافت:

$$\begin{aligned} &\max_{p_i,q_i} \left\{ \min_i \{E_i^{t-1} + p_i.H_i^t - q_i.M_i^t\} \right\} \\ &s.t. \quad E_i^{t-1} + p_i.H_i^t - q_i.M_i^t > E_{TH} \\ &p_i, q_i \in [0,1], i \in \{1, \dots, N_r\} \end{aligned} \quad (۵)$$

که با تعریف عبارت FN به‌عنوان تابع هدف به شکل رابطهٔ (٦)، داریم:

$$FN = \min_i \{E_i^{t-1} + p_i.H_i^t - q_i.M_i^t\} \quad (۶)$$

جدول (۱): تعریف پارامترهای استراتژی بهینهٔ انتخاب مُد عملکرد

E_i^{t-1}	انرژی اولیهٔ هر روبوت آتشنشان
H_i^t	میزان یک واحد انرژی ذخیره‌شدهٔ متوسط در برداشت انرژی
M_i^t	میزان یک واحد انرژی مصرفی متوسط در صورت حرکت
E_{TH}	آستانهٔ مفروض برای کمینهٔ انرژی در هر روبوت آتشنشان
p_C^D	احتمال آشکارسازی آتش در یک خوشه
p_i^D	احتمال آشکارسازی آتش در یک حسگر ثابت
p_{TH}^D	آستانهٔ مفروض احتمال آشکارسازی آتش در یک خوشه
p_C^F	احتمال اعلام اشتباه حریق در یک خوشه
p_i^F	احتمال اعلام اشتباه حریق در یک حسگر ثابت
p_{TH}^F	آستانهٔ مفروض احتمال اعلام اشتباه حریق در یک خوشه
N_S	تعداد کل حسگرهای ثابت در یک خوشه
N_r	تعداد کل حسگرها یا روبوت‌های متحرک در یک خوشه

اکنون ضمن ساده‌سازی تابع هدف کلی در مسئلهٔ بهینه‌سازی و با در نظر گرفتن نکتهٔ $max_{p_i,q_i,FN}\{FN\} = min_{p_i,q_i,FN}\{-FN\}$ و همچنین لحاظ‌کردن اثر آن به شکل محدودیت $E_i^{t-1} + p_i.H_i^t - q_i.M_i^t > FN$، مسئلهٔ بهینه‌سازی نهایی به شکل رابطهٔ (۷) بازنویسی می‌شود:

$$\begin{aligned} &\min_{p_i,q_i} \{-FN\} \\ &s.t. \quad E_i^{t-1} + p_i.H_i^t - q_i.M_i^t > E_{TH} \\ &p_i, q_i \in [0,1], i \in \{1, \dots, N_r\} \\ &E_i^{t-1} + p_i.H_i^t - q_i.M_i^t > FN \end{aligned} \quad (۷)$$

در ادامه، از روش مبتنی بر تابع لاگرانژ به‌منظور حل این مسئلهٔ بهینه‌سازی محدب بهره خواهیم گرفت که درنتیجهٔ اعمال این روش، مطابق با رابطهٔ (۸) تابع لاگرانژ معادل به دست می‌آید.

$$\begin{aligned} L = -FN &- \sum_{i=1}^{N_r} \alpha_i \big(E_i^{t-1} + p_i.H_i^t - q_i.M_i^t \\ &- E_{TH}\big) \quad - \sum_{i=1}^{N_r} \beta_i\, p_i \\ &+ \sum_{i=1}^{N_r} \gamma_i\, (p_i - 1) - \sum_{i=1}^{N_r} \delta_i\, q_i \\ &+ \sum_{i=1}^{N_r} \varepsilon_i\, (q_i - 1) \\ &- \sum_{i=1}^{N_r} \psi_i \left(E_i^{t-1} + p_i.H_i^t \right. \\ &\left. - q_i.M_i^t - FN\right) \end{aligned} \quad (۸)$$

که در آن $\{\alpha_i\}$، $\{\beta_i\}$، $\{\gamma_i\}$، $\{\delta_i\}$، $\{\varepsilon_i\}$، $\{\psi_i\}$ ضرایب لاگرانژ مرتبط با محدودیت‌های مسئله‌اند. در این تحقیق، فرض می‌شود مُدهای عملکرد تعریف‌شده برای هر روبوت آتشنشان به‌صورت هم‌زمان فعال‌سازی نمی‌شوند (به عبارت دیگر $q_i = 1 - p_i$)؛ بنابراین، نوشته می‌شود:

$$\begin{aligned} &L = -FN - \textstyle\sum_{i=1}^{N_r} \alpha_i (E_i^{t-1} + p_i.H_i^t - \\ &(1 - p_i).M_i^t - E_{TH}) - \textstyle\sum_{i=1}^{N_r} \beta_i\, p_i + \\ &\textstyle\sum_{i=1}^{N_r} \gamma_i\, (p_i - 1) - \textstyle\sum_{i=1}^{N_r} \psi_i\, (E_i^{t-1} + \\ &p_i.H_i^t - (1 - p_i).M_i^t - FN) \end{aligned} \quad (۹)$$

به این ترتیب، پس از اعمال شرایط KKT و ساده‌سازی

ریاضی مقدار بهینهٔ احتمال حرکت روبوت آتشنشان ($q_i^* = 1 - p_i^*$) به‌صورت زیر بیان می‌شود:

$$q_i^* = 1 - \frac{E_{TH} - E_i^{t-1} + M_i^t}{H_i^t + M_i^t} \quad (۱۰)$$

بنابراین، مُد عملکرد روبوت iام در یک خوشهٔ مفروض با احتمال $q_i^* \geq 1/2$ در حالت حرکت به سمت حریق و در غیر این صورت در حالت شارژ یا برداشت انرژی خواهد بود.

٤-٢- استراتژی هوشمند کنترل حرکت ربات‌های آتشنشان

عموماً انواع مختلفی از الگوریتم‌های تقویتی یا RL وجود دارند که برای اتخاذ استراتژی‌های هوشمند در حوزه‌های مختلف ازجمله شبکه‌های حسگری بی‌سیم [۲۲] استفاده می‌شوند. یکی از محبوب‌ترین الگوریتم‌ها، الگوریتم یادگیری کیو[۲۷] است. درواقع، الگوریتم یادگیری کیو با استفاده از تخمین پیوسته[۲۸]، جدولی از تمامی مقادیر $Q(s,a)$ را محاسبه می‌کند که آن را جدول Q می‌نامند. باید توجه داشت $Q(s,a)$ نشان‌دهندهٔ نتایج مورد انتظار است که به‌عنوان فاکتور کیفیت با بردار حالت $S = \{s_1, s_2, \dots, s_N\}$ بعد از انجام کنش a و دریافت پاداش به دست می‌آید. براساس این، جدول Q محاسبه‌شده طبق فرمول بازگشتی (۱۱) به‌روزرسانی می‌شود:

$$Q(s,a) \leftarrow Q(s,a) + \beta.\Delta Q(s,a) \quad (۱۱)$$

که در آن:

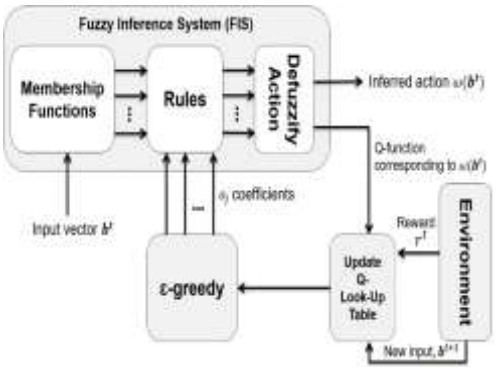
$$\Delta Q(s,a) = r + \lambda \max_{a'} Q(s',a') - Q(s,a) \quad (۱۲)$$

درواقع، بیشینه‌کردن فاکتور کیفیت به بالاترین کیفیت دریافتی توسط هر حسگر متحرک آتشنشان به متغیر کنش a' مرتبط است که ممکن است در حالت بعدی s' رخ دهد. در این رابطه، پارامتر $\beta \in (0,1]$ نیز به‌عنوان نرخ یادگیری تعریف شده است. عملکرد الگوریتم یادگیری کیو در حالت پایه با پیگیری سنجیدهٔ تاریخچهٔ فعال‌سازی حالات، به طرز چشمگیری امکان بهبود دارد. این نکته با پارامتر صلاحیت[۲۹] $\lambda \in [0,1]$ کنترل می‌شود و در حالت کلی این روش را روش یادگیری غنی‌شده $Q(\lambda)$ می‌نامند. پارامتر λ بعد از فعال‌سازی حالت، برای هر حالت بزرگ‌تر می‌شود و بعد از آن، به‌صورت نمایی کاهش می‌یابد تا جایی که حالت مربوطه دیگر اتفاق نیفتد. به این ترتیب، الگوریتم یادگیری کیو در مدل تعریف‌شده، به حسگرهای متحرک یا پهپادهای آتشنشان (شکل (۲)) این اجازه را می‌دهد که از تعامل با محیط، به‌صورت لحظه‌ای آموزش ببینند؛ این نوع از فرایند یادگیری توسط سازوکار تشویق و تنبیه صورت می‌گیرد. با ترکیب راهکارهای کنترل فازی و الگوریتم یادگیری کیو[۳۰] (FQL)، یک روش کارا برای کاربردهای عملی تحقق‌پذیر است (شکل (۳)). در حقیقت، تفاوت عمدهٔ بین الگوریتم یادگیری کیو اصیل و الگوریتم یادگیری FQL را می‌توان در روش ارائهٔ اطلاعات در مدل‌ها پیدا کرد. الگوریتم یادگیری FQL از روش‌های فازی برای ذخیره‌سازی اطلاعات جستجوشده استفاده می‌کند؛ در حالی که الگوریتم یادگیری کیو، آنها را در یک جدول جستجوی ساده (جدول Q) و به‌صورت قواعدی گسسته نگهداری می‌کند. در الگوریتم یادگیری FQL، سیستم استنباطی فازی[۳۱] (FIS) با مجموعه‌ای از قواعد یا ضوابط $\mathcal{R}$ و کنش‌های رقابتی برای هر ضابطه شناخته می‌شود.

شکل (۲): مدل آزمایشگاهی شبکهٔ حسگر آتشنشان شامل ایستگاه پایه، سرخوشه‌ها و حسگرهای توزیع‌شدهٔ ثابت و متحرک

عامل یادگیری (یا همان حسگر متحرک) مجبور است بهترین نتیجه را برای هر ضابطه پیدا کند که این همان کنش

با بهترین مقدار Q بین کنش‌های گسستهٔ ممکن برای هر ضابطه است. مقادیر Q در مقداردهی اولیه، صفر قرار داده می‌شوند و معمولاً در ابتدای فرایند یادگیری قابل توجه نیستند. درواقع، روبوت‌های آتشنشان براساس الگوریتم مبتنی بر یادگیری فازی با گام‌های حرکتی متغیر، به تدریج به سمت آتش حرکت می‌کنند. به این ترتیب، یک استراتژی هوشمند در شناسایی و اطفای حریق مبتنی بر شبکهٔ حسگر بی‌سیم برای هریک از گره‌های متحرک تصور می‌شود. برای تخمین سیاست بهینه[۳۲] نیاز است مقدار تابع کنش – حالت[۳۳] $Q^{\pi}(s,a)$ تقریب زده شود؛ این عبارت، تابعی از مقادیر مورد انتظار در صورت انجام کنش‌های a در حالت مفروض s است که به‌طور کلی یک استراتژی بهینه با سیاست π را دنبال می‌کند.



شکل (۳): بلوک دیاگرام الگوریتم یادگیری فازی – کیو (FQL)

الف) استراتژی کنترل حرکت براساس الگوریتم FQL

در این بخش، یک الگوی کنترل حرکت با هدف محاصره و درنهایت، اطفای سریع آتش برای سیستم MAS متشکل از حسگرهای منتخب یا همان ربات‌های آتشنشان و براساس الگوریتم فازی مبتنی بر روش یادگیری FQL طراحی می‌شود. درواقع، کنترل‌کنندهٔ طراحی‌شده در هر حسگر منتخب براساس الگوریتم FQL، سیگنال اطلاعات آتش (در اینجا زاویهٔ حرکتی) را به‌عنوان متغیر ورودی و پاداشی مثبت یا منفی متناسب با اثر حرکت خود در جهت تصادفی در شبکهٔ حسگر دریافت می‌کند. به این ترتیب در یک بازهٔ زمانی مشخص هر حسگر متحرک سعی دارد در طول حرکت‌های متوالی و تصادفی و با هدف نزدیک‌شدن به مکان آتش، مجموع مقادیر پاداش[۳۴] خود را به بالاترین حد خود برساند. به‌طور کلی در این تحقیق، یک رویکرد کاربردی برای مسیریابی هوشمندانه براساس استراتژی RL و مبتنی بر اطلاعات محیطی برای هر حسگر متحرک اتخاذ شده است. به عبارت دیگر، اعمال قابلیت یادگیری از تجربیات گذشته در هر حسگر متحرک، شبکهٔ حسگر آتشنشان را به یک شبکهٔ هوشمند مبدل می‌کند. همان‌طور که اشاره شد، الگوریتم FQL که بر مبنای روش یادگیری – کیو تعمیم یافته است، الگوی تصمیم‌گیری را با مدل‌سازی مبتنی بر فازی ترکیب می‌کند که در نتیجهٔ آن، حرکت مؤثر حسگر می‌تواند به شکلی کارآمد مدیریت شود. درواقع، در الگوریتم FQL یک سیستم استنباطی فازی (FIS) به کار گرفته شده است که از نظریهٔ مجموعهٔ فازی برای نگاشت ورودی‌ها به خروجی‌ها استفاده می‌کند. سیستم FIS استفاده‌شده در این تحقیق براساس مدل TS[۳۵] مرتبه صفر طراحی شده است؛ زیرا نوع دیگر (مرتبه اول) علاوه بر پیچیدگی بیشتر، هزینه‌های محاسباتی بالایی می‌طلبد؛ بنابراین، برای هر ضابطهٔ تعریف‌شده در سیستم FIS، عامل یادگیری باید بهترین نتیجه را مطابق با مقدار $q(L_j, o_j)$ (که در آن L_j عبارت فازی زبانی[۳۶] و o_j کنش گسسته برای jامین ضابطه[۳۷] تعریف شده است) پیدا کند. به عبارت دیگر، کنشی با بالاترین مقدار q را بین تمامی کنش‌های گسستهٔ احتمالی، برای بردار اطلاعات ورودی $b = [b_1, \dots, b_N]$ بیابد. در هر صورت، عامل یادگیری در راه‌اندازی اولیهٔ الگوریتم با توجه به صفربودن مقادیر q، ممکن است کنش‌های پذیرفتنی نداشته باشد (مقادیر q در جدول جستجو Q ذخیره شده‌اند). فرض می‌شود کنش برای هر ضابطه مطابق با سیاست بهره‌برداری – اکتشاف[۳۸] (EEP) صورت گرفته است. با این سیاست، عامل یادگیری، آن کنشی که باور دارد بهترین است را در بیشتر موارد انتخاب می‌کند؛ اما گاه و بیگاه نیز به‌صورت اتفاقی عمل می‌کند تا شاید پاداش‌های لحظه‌ای بالاتری را دریافت کند. در اینجا از استراتژی $\epsilon - Greedy$ به‌عنوان سیاست EEP برای انتخاب کنش بهره گرفته می‌شود. همان‌طور که مشاهده شد، معماری FQL و اثر متقابل آن با محیط در شکل (۳) نشان

داده شده است. در این شکل، $\omega(b)$ کنش استنباطی (کنش خروجی سیستم تصمیم‌گیری FIS است که جهت حرکت را برای هر حسگر منتخب مشخص می‌کند) برای بردار حالت ورودی b است. همچنین تابع کیفیت Q در این الگوریتم نیز براساس خروجی سیستم FIS تخمین زده می‌شود که از کیفیت (مقدار q) متعلق به کنش گسستهٔ موضعی استنباط شده است و کنش پیوستهٔ سراسری $\omega(b)$ را شکل می‌دهد. درواقع، تابع Q مربوط به $\omega(b)$ (بردار حالت ورودی جدید بعد از انجام کنش $\omega(b)$ برای بردار حالت ورودی b) و پاداش دریافتی r از محیط، با یکدیگر برای به‌روزرسانی مقادیر q در جدول Q استفاده می‌شوند. در این حالت، برای تشخیص مقادیر لحظه‌ای زاویهٔ حرکت حسگر منتخب (θ_F) و شعاع حرکتی آن (R_F)، بردار ورودی $b = [b_1, b_2] = [R_F, \theta_F]$ به سیستم FIS داده می‌شود و مطابق با ضوابط تعریف‌شده برای تعیین کنش یا اقدام $\omega(b)$، تابع کیفیت سیستم FIS به عبارت دیگر تابع $Q(b, \omega(b))$ نیز محاسبه می‌شود.

ب) تنظیمات بلوک تصمیم‌گیری

برای یک حسگر متحرک، هر متغیر ورودی از بردار حالت دوبعدی $\boldsymbol{b} = [R_F, \theta_F]$ در سه زیرمجموعهٔ فازی تقسیم‌بندی می‌شود. درواقع، به تعداد ۹ ضابطه با توجه به توصیفات زبانی سه‌گانهٔ کم (L)، متوسط (M) و زیاد (H) تعریف می‌شود. توصیفات زبانی متعلق به مجموعه‌های فازی مربوط به متغیرهای R_F و θ_F است که به‌صورت توابع عضویت (H, M, L) نشان داده می‌شوند.

الگوریتم (۱): الگوریتم تکرارشوندهٔ حرکت حسگر متحرک بر پایهٔ FQL

۱. مقداردهی اولیه $q(L_j, o_k)$ در جدول Q (که در آن $j \in \mathcal{R}$ و $k \in \mathcal{S}$). در اینجا $\mathcal{S}$ مجموعه تمام کنش‌های گسستهٔ احتمالی برای هر ضابطه است.

۲. دریافت بردار حالت b^t (شامل زاویه و شعاع حرکت حسگر متحرک).

۳. محاسبهٔ درجه صحت[۳۹] بردار حالت b^t یا همان $\alpha_j(b^t)$ برای تمام ضوابط که به‌صورت زیر تعریف می‌شود: $\alpha_j(b^t) = \prod_{n=1}^{N} \mu_{L_j^n}(b_n)$ که $\sum_{j\in\mathcal{R}} \alpha_j(b^t) = 1$. واضح است این مقدار براساس ضابطه jام و مجموعه‌های فازی یا توابع عضویت $\mu_{L_j^N}(b_N)$ تا $\mu_{L_j^1}(b_1)$ برای بردار حالت ورودی b^t محاسبه می‌شود.

۴. انتخاب یک کنش o_j برای هر ضابطه براساس سیاست EEP. به عبارت دیگر:

$$o_j = \begin{cases} argmax_{k\in\mathcal{S}}\, q(L_j, o_k), & P_r = 1-\epsilon \\ random_{k\in\mathcal{S}}\,(o_k), & P_r = \epsilon \end{cases}$$

که در آن مقدار ϵ، مصالحه بین بهره‌برداری و اکتشاف را (به‌ترتیب با احتمال‌های $P_r = 1-\epsilon$ و $P_r = \epsilon$) مشخص می‌کند.

۵. محاسبهٔ کنش استنتاجی $\omega(b^t)$ و کیفیت مربوط به آن $Q(b^t, \omega(b^t))$ که به‌صورت زیر محاسبه می‌شود:

$$Q(b^t, \omega(b^t)) = \sum_{j\in\mathcal{R}} \alpha_j(b^t).\, q(L_j, o_j)$$

۶. اجرای کنش استنتاجی $\omega(b^t)$ و دریافت بردار حالت جدید b^{t+1}.

۷. دریافت سیگنال تقویتی یا پاداش r^t.

۸. محاسبهٔ درجه صحت برای بردار حالت جدید b^{t+1} یا همان $\alpha_j(b^t)$.

۹. محاسبهٔ تابع ارزش براساس بردار حالت جدید:

$$V(b^{t+1}) = \sum_{j\in\mathcal{R}} \alpha_j(b^{t+1}).\, \max_{k\in A} q(L_j, o_k)$$

۱۰. به‌روز رسانی کیفیت اولیه $q(L_j, o_j)$ برای هر ضابطه jام و کنش o_j:

$$q^{t+1}(L_j, o_j) = q^t(L_j, o_j) + \psi\, \alpha_j(b^t)\left(r^t + \lambda V^t(b^{t+1}) - Q^t(b^t, \omega(b^t))\right)$$

که در آن ψ، نرخ یادگیری و $\lambda \in [0,1]$ فاکتور کاهنده است.

۱۱. اتمام تکرار. در صورتی که همگرایی حاصل شده باشد، فرایند یادگیری متوقف می‌شود؛ در غیر این صورت به گام دوم برمی‌گردیم.

در اینجا، توابع عضویت گوسی استاندارد[۴۰] برای بردار حالت b در نظر گرفته می‌شوند. توابع عضویت گوسی، به‌عنوان جایگزینی برای توابع عضویت مثلثی[۴۱] مرسوم، به این منظور ارائه شده‌اند که قابلیت اطمینان و عملکرد سیستم را بهبود ببخشند. در هر دورهٔ تصمیم‌گیری، عامل یا حسگر متحرک، بردار حالت فعلی را در نظر می‌گیرد و اقدام یا کنشی برای ورود به حالت شبکهٔ جدید انجام می‌دهد. به این ترتیب، یک سیگنال پاداش r (مقادیر ثابت مثبت یا منفی به‌ترتیب برای بهبود یا عدم بهبود کنش قبلی حسگر) دریافت می‌شود تا کیفیت این کنش را ارزیابی کند. اطلاعات آموخته‌شده ذخیره خواهد شد و فرایند یادگیری ادامه می‌یابد. خلاصه‌ای از این روش تکرارشونده برای

یادگیری مسیر در الگوریتم (۱) آمده است.

در ادامه، جمع‌بندی فرایند آشکارسازی و عملیات اطفای حریق مبتنی بر شبکهٔ حسگری شامل حسگرهای ثابت و متحرک (روبوت‌های آتشنشان) و همچنین نحوهٔ اعمال استراتژی‌های پیشنهادی تعیین مُد عملکرد و کنترل حرکت روبوت‌ها در الگوریتم (۲) آمده‌اند.

الگوریتم (۲): فرایند آشکارسازی و اطفای حریق

۱. خوشه‌بندی شبکهٔ حسگری و انتخاب سرخوشه‌ها به‌همراه تعیین مُدهای اولیهٔ پیش‌فرض «خواب» برای حسگرهای ثابت و همچنین «برداشت انرژی» برای روبوت‌های متحرک.

۲. فعال‌سازی M حسگر از میان N_S حسگر ثابت براساس کران‌های بالا و پایین محاسبه‌شده در رابطهٔ (۳) با هدف دستیابی به احتمالات مشارکتی قابل قبول آشکارسازی و اعلام اشتباه حریق در هر خوشه.

۳. استراتژی بهینهٔ انتخاب مُد عملکرد روبوت‌ها: تعیین مُد عملکرد N_R روبوت آتشنشان توسط سرخوشه‌ها براساس مقدار بهینهٔ احتمال حرکت روبوت در رابطهٔ (۱۰).

٤. استراتژی هوشمند کنترل حرکت روبوت‌های آتشنشان: هدایت روبوت‌ها بر پایهٔ الگوریتم یادگیری فازی - کیو و به کمک دو سیاست یادگیری کامل و جزئی با هدف محاصرهٔ آتش در عملیات اطفای حریق.

۵. پایان الگوریتم

۵- نتایج شبیه‌سازی کامپیوتری

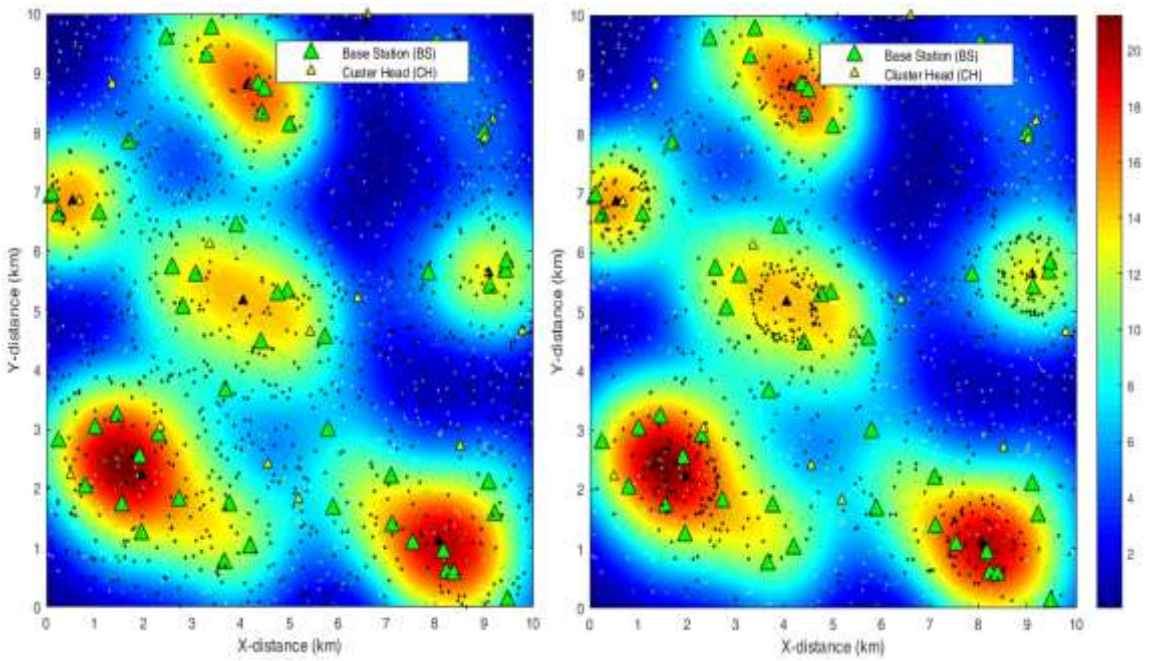
در این بخش، تحلیل نتایج شبیه‌سازی مونت‌کارلو برای ارزیابی عملکرد استراتژی‌های دوگانهٔ پیشنهادی اطفای حریق مبتنی بر انتخاب مُدهای عملکرد و سپس کنترل حرکت به کمک نرم‌افزار MATLAB ارائه خواهد شد. همان‌طور که پیش‌تر نیز مطرح شد، مدل ارائه‌شده در بستر یک شبکهٔ حسگری شامل ایستگاه‌های پایه، سرخوشه‌ها و حسگرهای توزیع‌شدهٔ ثابت و متحرک مطابق شکل (٤) اعمال می‌شود. در اینجا فرض شده است استراتژی اولیه، یعنی انتخاب مُد عملکرد شامل برداشت انرژی یا حرکت به سمت حریق، در دوره‌های زمانی مختلف و براساس مسئلهٔ بهینه‌سازی در رابطهٔ (٤) بر عهدهٔ حسگرهای سرخوشه است. هرکدام از حسگرهای منتخب با مُد عملکرد «حرکت»، به‌عنوان یک عامل یادگیری و مستقل در نظر گرفته می‌شوند. به عبارت دیگر، هرکدام از گره‌های شبکه در هنگام حرکت به سمت حریق همانند یک عنصر هوشمند مطابق با شکل (٤) عمل می‌کنند و براساس الگوریتم یادگیری FQL به تدریج با اصلاح مسیر حرکت خود، فرایند محاصرهٔ آتش را تکمیل خواهند کرد. حسگرهای متحرک یا روبوت‌های اطفای حریق (نقاط سیاه رنگ) بر پایهٔ الگوریتم یادگیری - فازی و به کمک سیاست‌های تعریف‌شدهٔ PEL و PAL در شبکهٔ حسگر، قادر به محاصرهٔ حریق (مثلث‌های سیاه رنگ) در عملیات اطفای حریق خواهند بود. در سیاست PEL اولویت بر یادگیری سریع زاویهٔ حرکت حسگر منتخب (θ_F) نسبت به شعاع حرکتی آن (R_F) در هنگام حرکت به سمت آتش است؛ اما در سیاست PAL فرایند یادگیری دو فاکتور مذکور در طول مسیر و به تدریج صورت می‌گیرد. این امکان با تنظیم مقادیر ضریب افت[٤٢] (DC) یا $\sigma \in [0,1]$ برای احتمال ϵ برای سیاست‌های PEL و PAL به‌ترتیب به‌صورت $\sigma \geq 0.95$ و $\sigma < 0.95$ تعریف می‌شود. در این مدل محدودیت‌های انرژی در روبوت‌ها با طراحی مکانیزم انتخاب گره و در نظر گرفتن قابلیت برداشت انرژی محیطی قبل از مکانیزم حرکت گره اعمال شده‌اند.

مقادیر پارامترهای مربوط به مدل شبکهٔ حسگر و الگوریتم یادگیری FQL در جدول (۲) گردآوری شده‌اند. همان‌طور که اشاره شد، سرعت بالای همگرایی الگوریتم یادگیری FQL و فرایند سبک و غیرپیچیدهٔ آن در مقایسه با سایر سناریوهای متداول دیگر همچون استراتژی مسیر تصادفی[٤٣] و استراتژی انتخاب مسیر مبتنی بر الگوریتم یادگیری کیو[٤٤]، آن را به‌عنوان یکی از بهترین کاندیداها برای طرح یک مدل کاربردی بدل کرده است. به این ترتیب، با توجه به محدودیت‌های شبکهٔ حسگر به‌ویژه طول عمر شبکه، ترکیب فاکتور سرعت در این الگوریتم با فاکتور برداشت انرژی، به طراحی مدلی موفق با نرخ قابل قبول نزدیک‌شدن به آتش منجر شده است (شکل‌های (۵) و (٦)). در اینجا یک تعریف ساده برای ارزیابی میزان اثرگذاری فاکتور سرعت در مدل پیشنهادی در مقایسه با سناریوهای متداول در یک بازهٔ زمانی و توپولوژی یکسان ارائه شده است. درواقع، نرخ کلی نزدیک‌شدن[٤٥] به آتش به‌صورت

$k_m.(d^0_{Fire}/d^t_{Fire})^2$ (که در آن، k_m مقداری ثابت است) فرموله شده است تا هر بهبودی در سرعت محاصرهٔ آتش و درنهایت اطفای حریق برای تمامی حسگرهای متحرک توزیع‌شدهٔ درون شبکه، به‌صورت کمّی نمایش داده شود. در اینجا d^t_{Fire} به‌عنوان متوسط فاصله بین حسگرهای متحرک و مکان آتش‌سوزی در زمان t، فرض شده است. فاکتور سرعت در استراتژی پیشنهادی حرکت به سمت حریق بر پایهٔ الگوریتم FQL با دو نکته استدلال می‌شود که آن را به شکل یک مدل عملی، قابل پیاده‌سازی و با پیچیدگی پایین نشان داده است. نکتهٔ اول، سرعت همگرایی الگوریتم FQL و درنتیجه تکمیل زودهنگام فرایند یادگیری در مقایسه با رویکرد مبتنی بر الگوریتم QL است. نکتهٔ دوم نیز با نحوهٔ طی مسیر حسگر متحرک مرتبط است. درواقع، ویژگی پیوستگی مقادیر کنش - حالت در الگوریتم FQL سبب شده است مسیر حرکت به سمت آتش مستقل از مکان حسگرهای همسایه باشد و با اتخاذ مقادیر پیوسته و دلخواه برای جهت و گام حرکت، انتخاب مسیر مستقیم پس از یادگیری سریع اولیه برخلاف سایر رویکردهای نامبرده ممکن شود.

جدول (۲): مقادیر پارامترها در شبیه‌سازی شبکهٔ حسگر

توان ارسالی ایستگاه پایه	$37dBm$	P_{BS}
توان ارسالی حسگر یا روبوت آتشنشان	$20dBm$	P_S
تراکم ایستگاه پایه	$10\ [km^2]^{-1}$	λ_{BS}
تراکم حسگرها در شبکه	$40\ [km^2]^{-1}$	λ_S
پارامتر فیدینگ رایلی	1	$\mu_{BS}=\mu_S$
حداکثر شعاع پوشش یک حسگر	$200m$	d
حد آستانهٔ برداشت انرژی	$-90dBm$	P^{th}_{EH}

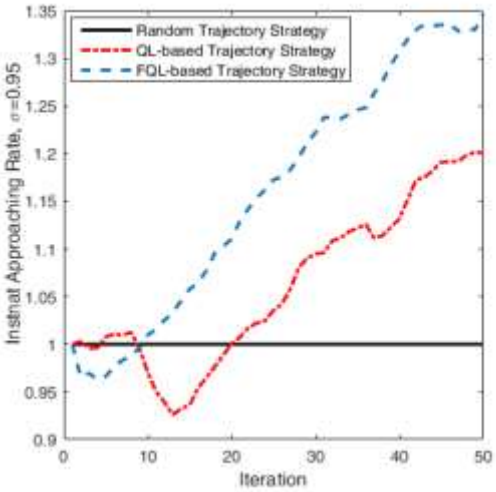


شکل (٤): شبیه‌سازی مدل شبکهٔ حسگر آتشنشان (مثلث‌های سیاه: مکان‌های فرضی دچار حریق) بر پایهٔ استراتژی‌های پیشنهادی انتخاب و حرکت حسگرها (قبل (چپ) و بعد (راست))

همان‌طور که در شکل (۷) نشان داده شده است، مسیر حرکت حسگر یا ربات منتخب (دایرهٔ سبز رنگ) به سمت ناحیهٔ فرضی حریق (ستارهٔ قرمز رنگ) پس از یادگیری سریع اولیه نسبتاً مستقیم بوده است. با این تفاسیر، در مجموع اتلاف زمان هم در فرایند یادگیری و هم در انتخاب مسیر در مقایسه با دو رویکرد دیگر، یعنی تصادفی و مبتنی بر الگوریتم QL، به کمترین میزان خواهد رسید.

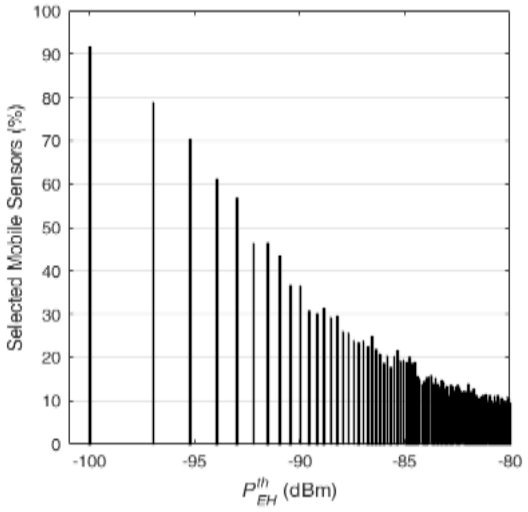
در ادامه، ضمن تمرکز بر رفتار یک حسگر به‌عنوان عامل یادگیری توزیع‌شده در شبکهٔ حسگر، نقش اتخاذ سیاست‌های یادگیری PAL و PEL در استراتژی حرکت

ارزیابی می‌شود و همچنین شاخصه‌های همگرایی و متوسط پاداش بر پایهٔ الگوریتم پیشنهادی FQL بررسی خواهند شد. همان‌طور که اشاره شد، در سیاست PEL اولویت بر یادگیری سریع زاویهٔ حرکت حسگر منتخب (θ_F) نسبت به شعاع حرکتی آن (R_F) در هنگام حرکت به سمت آتش است؛ اما در سیاست PAL، فرایند یادگیری دو فاکتور مذکور در طول مسیر و به تدریج صورت می‌گیرد؛ بنابراین، میزان اثرگذاری مقادیر ضریب افت σ برای احتمال ϵ در سیاست‌های PEL و PAL بر سرعت و دقت استراتژی حرکت مطابق با شکل (۸) مشاهده می‌شود. به‌طور کلی، گفتنی است در سیاست PEL با اُفت دیرهنگام میزان احتمال ϵ، یادگیری دقیق‌تر جهت صحیح حرکت با صرف زمان بیشتر برای دورهٔ آموزش[۴۶] در عامل یادگیری و در عوض، انتخاب زودهنگام مسیر مستقیم به سمت آتش ممکن خواهد شد. این مسئله دربارهٔ سیاست دیگر یعنی PAL با افت زودهنگام مقدار احتمال ϵ، به‌صورت وارون ظاهر خواهد شد. به عبارت دیگر، فرایند یادگیری در طول مسیر به تدریج کامل شده و درنتیجه، مسافت طی‌شده با احتمال بیشتری طولانی‌تر است و همگرایی به مسیر مستقیم تأخیر بیشتری دارد. این تعابیر برای نمودارهای لحظه‌ای و تجمعی تغییرات فاصله از حریق به‌ترتیب بر طبق شکل‌های (۹) و (۱۰) استدلال می‌شوند.

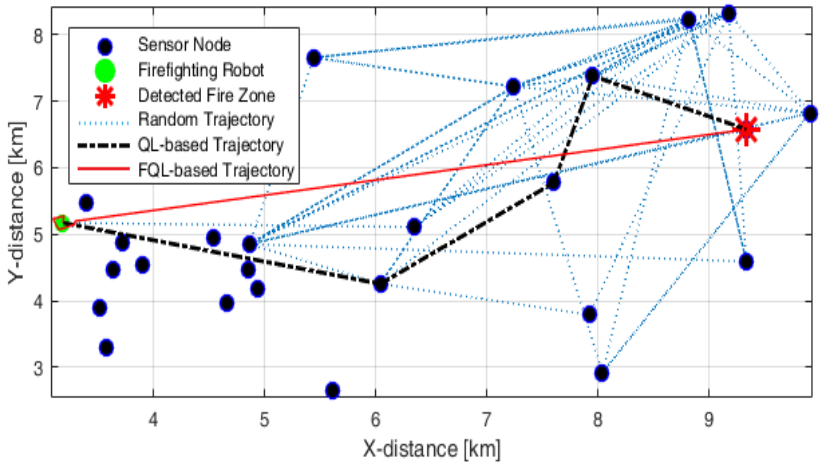


شکل (۵): متوسط نرخ نزدیک‌شدن به آتش ($k_m = 1$) در شبکهٔ حسگر آتشنشان، برای سناریوهای انتخاب مسیر به‌صورت RND، الگوریتم QL و الگوریتم پیشنهادی FQL.

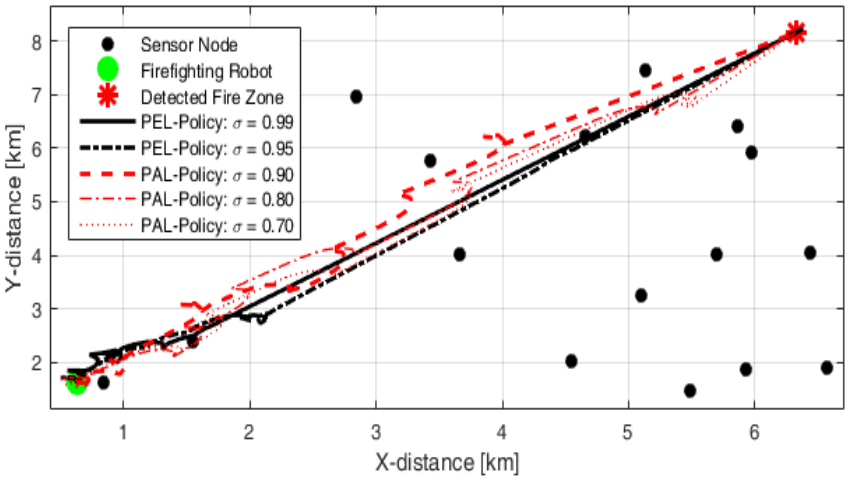
همان‌طور که پیش‌تر نیز اشاره شد، در سیاست PEL اولویت بر یادگیری سریع زاویهٔ حرکت حسگر منتخب نسبت به شعاع حرکتی آن در هنگام حرکت به سمت آتش است؛ اما در سیاست PAL فرایند یادگیری دو فاکتور مذکور در طول مسیر و به تدریج صورت می‌گیرد. استدلال مشابه دربارهٔ نمودارهای لحظه‌ای پاداش حسگر متحرک که به‌صورت عبارت $r = \cos(\theta_F)$ تعریف شده نیز صادق است. شیب صعودی نمودارها (مطابق با شکل (۱۱)) برحسب تکرار در الگوریتم FQL نیز بیان‌کنندهٔ همسوشدن تدریجی جهت حرکت حسگر متحرک در مسیر مستقیم به سمت آتش است که دربارهٔ حالت‌های با سیاست یادگیری PEL این اتفاق در تعداد تکرار کمتری رخ داده است. در حقیقت، نوسانات شیب نمودار در همان تکرارهای آغازین مشاهده می‌شوند و با کامل‌شدن نسبی فرایند یادگیری و شناسایی صحیح جهت حرکت مطابق با سیاست مذکور، شیب مثبت نمودار به‌صورت زودهنگام و با اختلاف چشمگیر در مقایسه با سیاست PAL بر طبق نتایج شبیه‌سازی دیده می‌شود.



شکل (۶): نمودار میله‌ای اثرگذاری میزان سطح آستانهٔ برداشت انرژی، بر تعداد حسگرهای منتخب برای حرکت بر طبق استراتژی بهینهٔ انتخاب مُد عملکرد

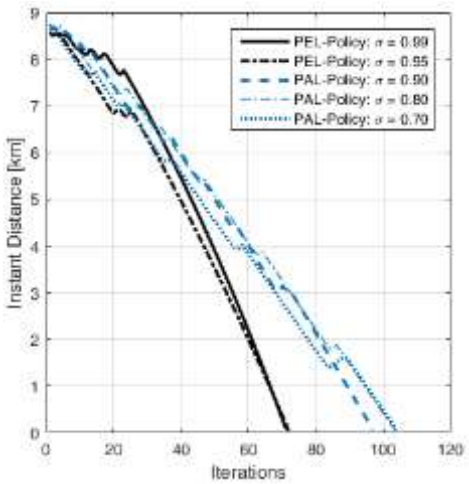


شکل (۷): شبیه‌سازی استراتژی کنترل حرکت به سمت آتش به‌صورت تصادفی، مبتنی بر الگوریتم QL و بر پایهٔ الگوریتم پیشنهادی FQL، در یک حسگر متحرک مطابق با استراتژی بهینهٔ انتخاب مُد عملکرد

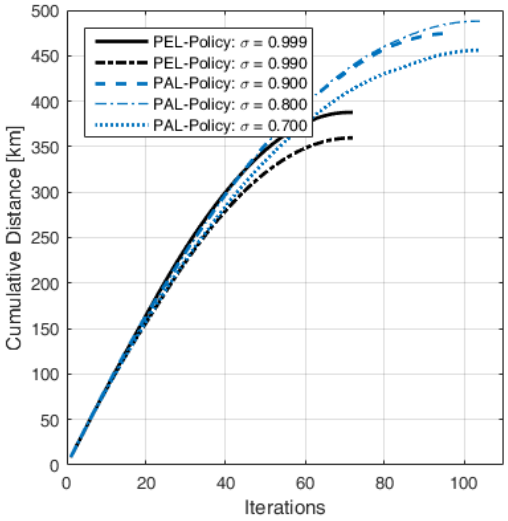


شکل (۸): شبیه‌سازی استراتژی کنترل حرکت به سمت آتش بر پایهٔ الگوریتم پیشنهادی FQL و سیاست‌های یادگیری PAL و PEL

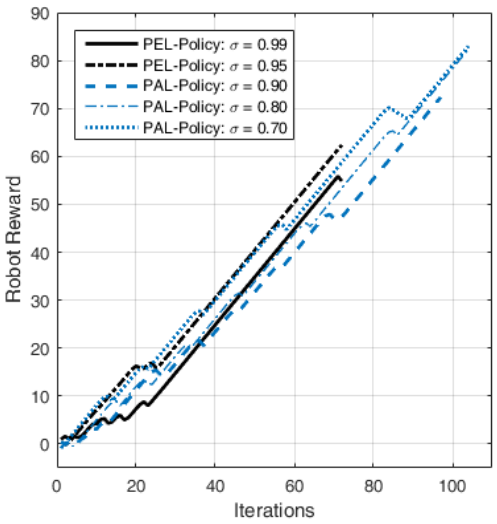
این مقدار برای مکان‌های مفروض وقوع حریق و ربات منتخب، با اختلاف در حدود ۲۵ تکرار مشاهده می‌شود. گفتنی است در اینجا فرایند پاداش‌دهی در یادگیری تقویتی با حضور حسگر متحرک در مکان حریق متوقف شده است. واضح است نقاط با شیب نزولی در نمودار پاداش لحظه‌ای معادل با مقادیر پاداش منفی در حالت‌های اتخاذ اشتباه جهت حرکت است. این موارد در حالت‌های با سیاست حرکتی PEL فقط در شروع حرکت حسگر رخ داده است و به تدریج جهت حرکت عامل یادگیری تثبیت خواهد شد.



شکل (۹): نمودار لحظه‌ای تغییرات فاصله از حریق در استراتژی کنترل حرکت، به سمت آتش بر پایهٔ الگوریتم پیشنهادی FQL

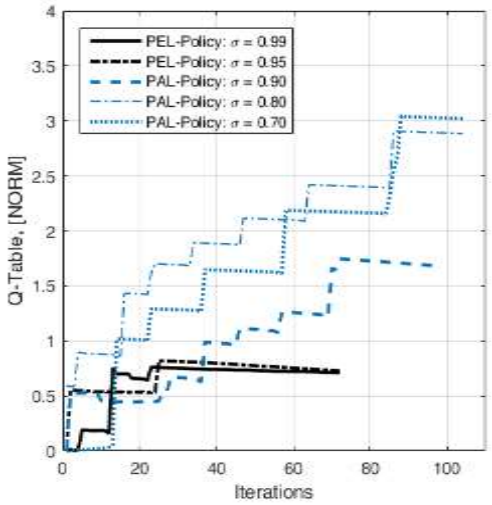


شکل (۱۰): نمودار تجمعی تغییرات فاصله از حریق در استراتژی کنترل حرکت به سمت آتش مبتنی بر الگوریتم پیشنهادی FQL

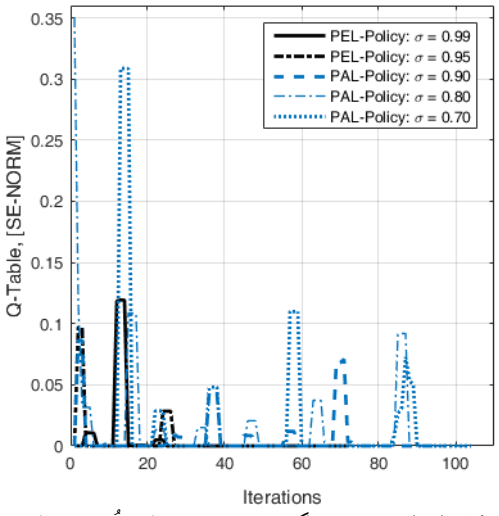


شکل (۱۱): نمودار لحظه‌ای تغییرات پاداش در استراتژی کنترل حرکت به سمت آتش براساس الگوریتم پیشنهادی FQL

در مقابل، این اتفاق مکرراً در بازهٔ بزرگ‌تری برای حالت‌های با سیاست حرکتی PAL درحال وقوع است که این رفتار نیز پیش‌تر انتظار می‌رفت. هر دو نمودار تغییرات نُرم مربوط به بردار بیشینهٔ مقادیر جدول Q در هر کنش (شکل (۱۲)) و نمودار مربع خطای[۴۷] (SE) تغییرات نُرم مربوط به بردار بیشینهٔ مقادیر جدول Q در هر کنش (شکل (۱۳)) نشان‌دهندهٔ همگرایی فرایند یادگیری حدوداً در تکرار شمارهٔ ۳۰ برای سیاست PEL و تقریباً در تکرار شمارهٔ ۷۰ به بعد برای بهترین موارد با سیاست حرکتی PAL هستند.



شکل (۱۲): نمودار تغییرات نُرم مربوط به بردار بیشینهٔ مقادیر جدول Q در هر کنش



شکل (۱۳): نمودار همگرایی تغییرات خطای نُرم مربوط به بردار بیشینهٔ مقادیر جدول Q

۶- نتیجه‌گیری

درواقع، هدف نهایی از مدل پیشنهادی، ارائهٔ یک طرح کاربردی اقدام سریع و هوشمند برای عملیات اطفای حریق مبتنی بر شبکه‌های حسگر بی‌سیم است. در مدل پیشنهادی، فرایند مشارکتی شناسایی حریق با حسگرهای ثابت و عملیات بهینهٔ اطفای حریق با حسگرهای متحرک بر پایهٔ الگوریتم یادگیری فازی-کیو و به کمک دو سیاست یادگیری کامل و جزئی در شبکهٔ حسگری با توپولوژی پویا

مطالعه شد. درنهایت، تحلیل و طراحی مؤثر استراتژی کنترل حرکت بر طبق نتایج بهبودیافتهٔ شبیه‌سازی برای سیستم MAS متشکل از حسگرهای متحرک و براساس الگوریتم پیشنهادی FQL صورت گرفت

مراجع

[1] M. Yang and C. Zhang, "Smoke Alarm System," *Wireless*, Vol. 9, pp. 50-51, 2006.

[2] T. M. Behera, S. K. Mohapatra, U. C. Samal, M. S. Khan, M. Daneshmand and A. H. Gandomi, "I-SEP: An Improved Routing Protocol for Heterogeneous WSN for IoT-Based Environmental Monitoring," in *IEEE Internet of Things Journal*, Vol. 7, No. 1, pp. 710-717, Jan. 2020.

[3] R. Morello, S. C. Mukhopadhyay, Z. Liu, D. Slomovitz and S. R. Samantaray, "Advances on Sensing Technologies for Smart Cities and Power Grids: A Review," in IEEE Sensors Journal, Vol. 17, No. 23, pp. 7596-7610, 1 Dec.1, 2017.

[4] S. Anand and Keetha Manjari.R.K, "FPGA implementation of artificial Neural Network for forest fire detection in wireless Sensor Network," 2017 2nd Int. Conf. on Computing and Comm. Tech. (ICCCT), Chennai, 2017, pp. 265-270.

[5] T. Islam, H. A. Rahman and M. A. Syrus, "Fire detection system with indoor localization using ZigBee based wireless sensor network," 2015 Int. Conf. on Informatics, Electronics & Vision (ICIEV), Fukuoka, 2015, pp. 1-6.

[6] Farzad H. Panahi, Parvin Farhadi & Zhila H. Panahi (2016) Spectral-Efficient Green Wireless Communications via Cognitive UWB Signal Model, Automatika, 57:3, 793-809.

[7] Giglioa, L., Descloitresa, J., Justicec, C.O., Kaufman, Y.J., 2003. An enhanced contextual fire detection algorithm for MODIS. Remote Sensing of Environment 87, 273–282.

[8] V. Sherstjuk, M. Zharikova and I. Sokol, "Forest Fire Monitoring System Based on UAV Team, Remote Sensing, and Image Processing," 2018 IEEE Second Int .Conf .on Data Stream Mining & Processing (DSMP), Lviv, 2018, pp. 590-594.

[9] S. R. Vijayalakshmi and S. Muruganand, "A survey of Internet of Things in fire detection and fire industries," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, 2017, pp. 703-707.

[10] F. A. Hossain, Y. Zhang and C. Yuan, "A Survey on Forest Fire Monitoring Using Unmanned Aerial Vehicles," 2019 3rd International Symposium on Autonomous Systems (ISAS), Shanghai, China, 2019, pp. 484-489.

[11] Stone, P.; Veloso, M. Multiagent systems: A survey from machine learning perspective. Auton. Robots 2000, 8,345–383.

[12] N. K. Ure, S. Omidshafiei, B. T. Lopez, A. [13] Agha-Mohammadi, J. P. How and J. Vian, "Online heterogeneous multiagent learning under limited communication with applications to forest fire management," 2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Sys. (IROS), Hamburg, 2015, pp. 5181-5188.

[14] Rashid, A.T.; Ali, A.A.; Frasca, M.; Fortuna, L. Path planning with obstacle avoidance based on visibility binary tree algorithm. Robot. Auton. Syst. 2013, 61, 1440–1449.

[15] Arel, I.; Liu, C.; Urbanik, T.; Kohls, A.G. Reinforcement learning-based multi-agent system for network traffic signal control. IET Intell. Transp. Syst. 2010, 4, 128–135.

[16] Cherkassky, V.; Mulier, F. Learning from data: Concepts, Theory and Methods; Wiley-IEEE Press: Hoboken, USA, 2007.

[17] Zhang, W.; Ma, L.; Li, X. Multi-agent reinforcement learning based on local communication. Clust. Comput. 2018, 1–10.

[18] Hu, X.; Wang, Y. Consensus of Linear Multi-Agent Sys. Subject to Actuator Saturation. Int. J. Con. Aut. Syst. 2013, 11, 649–656.

[19] Luviano, D.; Yu, W. Path planning in unknown environment with kernel smoothing and reinforcement learning for multi-agent systems. In Proceedings of the 12th Int. Conf. on Electrical Engineering, Computing Science and Automatic Control (CCE), Mexico City, Mexico, 28–30 October 2015.

[20] Abul, O.; Polat, F.; Alhajj, R. Multi-agent reinforcement learning using function approximation. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 2000, 485–497.

[21] Fernandez, F.; Parker, L.E. Learning in large cooperative multi-robots systems. Int. J. Robot. Autom. Spec. Issue Comput. Intell. Tech. Coop. Robots 2001, 16, 217–226.

[22] Foerster, J.; Nardelli, N.; Farquhar, G.; Afouras, T.; Torr, P.H.; Kohli, P.; Whiteson, S. Stabilising experience replay for deep multi-agent reinforcement learning. arXiv 2017.

[23] Tam., H.; Ishi, S. Multi agent reinforcement

[28] F. H. Panahi and T. Ohtsuki, “Optimal channel-sensing scheme for cognitive radio systems based on fuzzy q-learning,” *IEICE Trans. Commun*., Vol. 97, No. 2, pp. 283–294, 2014.
[29] Waleed Ejaz, Muhammad Naeem, Adnan Shahid, Alagan Anpalagan and Minho Jo, "Efficient energy management for the internet of things in smart cities", *IEEE Communications Magazine*, Vol. 55, No. 1, pp. 84-91, 2017.
[30] Zhu, Ch., V. CM L., Lei Shu, and E. C-H. Ngai. "Green internet of things for smart world." *IEEE Access*, Vol. 3, pp. 2151- 2162, 2015.
[31] M. M. Amiri and S. M. H. Andargoli, "Life time maximization in the Wireless Sensor Network with energy harvesting," 2017 IEEE 4th Int. Conf. on Knowledge-Based Engineering and Innovation (KBEI), Tehran, 2017, pp. 0412-0417.
learning applied to a chase problem in a continuous world. Life Robot. 2001, 202–206.
[24] Ishiwaka, Y.; Sato, T.; Kakazu, Y. An approach to pursuit problem on a heterogeneous multiagent system using reinforcement learning. Robot. Auton. Syst. 2003, 43, 245–256.
[25] Radac, M.-B.; Precup, R.-E.; Roman, R.-C. Data-driven model reference control of MIMO vertical tank systems with model-free VRFT and Q-Learning. ISA Trans. 2017.
[26] Pandian, B.J.; Noel, M.M. Control of a bioreactor using a new partially supervised reinforcement learning algorithm. J. Process Control 2018, 69, 16–29.
[27] F. H. Panahi, F. H. Panahi, G. Hattab, T. Ohtsuki and D. Cabric, "Green Heterogeneous Networks via an Intelligent Sleep/Wake-Up Mechanism and D2D Communications," in *IEEE Trans .on Green Comm .and Networking*, Vol. 2, No. 4, pp. 915-931, Dec. 2018.

[1] Wireless Sensor Network
[2] Internet of Things
[3] Zigbee Technology
[4] Ultra Wide-band
[5] Multi-Agent Systems
[6] Agent
[7] Low Level Control Systems
[8] Reinforcement Learning
[9] Multi-Agent Reinforcement Learning
[10] Cumulative Reward
[11] Action
[12] State
[13] Reward Function
[14] Value Functions
[15] Network Lifetime Maximization
[16] Two-Tier Heterogonous Network
[17] Base Stations
[18] Cluster Heads
[19] Sinks
[20] Sensor Nodes
[21] Perfect Learning Policy
[22] Partial Learning Policy
[23] Energy Harvesting
[24] Binary Hypothesis Testing
[25] Integer Programming
[26] Relaxed Problem
[27] Q-Learning Algorithm
[28] Continuous Estimation
[29] Competency Parameter

[30] Fuzzy Q-Learning Algorithm
[31] Fuzzy Inference System
[32] Optimal Policy
[33] State-Action
[34] Reward Values
[35] Takagi-Sugeno
[36] Linguistic Fuzzy Term
[37] IF-THEN Rule
[38] Exploration-Exploitation Policy
[39] Truth
[40] Standard Gaussian Membership
[41] Triangular Membership Functions
[42] Decay Coefficient
[43] Random Trajectory Strategy
[44] QL-based Trajectory Strategy
[45] Total Approaching Rate
[46] Training Period
[47] Square Error