



مروری نظام‌مند بر یادگیری تقویتی عمیق در زمان‌بندی و انتقال وظایف در محیط‌های ابری، مه و لبه

پیمان پهلوسای^۱، شیوا رزاق زاده^{۲*}

۱- دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد اردبیل، اردبیل، ایران

Peymanpahloosye@gmail.com

۲- گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد اردبیل، اردبیل، ایران

Shiva.razzaghzadeh@srbiau.ac.ir

خلاصه

با افزایش سریع تولید داده در کاربردهای هوشمند، نیاز به پاسخ‌دهی بلادرنگ و مدیریت تقاضاهای پیچیده در سیستم‌های اینترنت اشیا شدت یافته است. در این میان، محیط‌های رایانش ابری، مه و لبه به‌عنوان زیرساخت‌هایی با قابلیت توسعه‌پذیری و سازگاری، نقش مهمی در اجرای وظایف محاسباتی ایفا می‌کنند. مسئله‌ی زمان‌بندی و انتقال تطبیقی وظایف، به‌ویژه در شرایط ناهمگون و پویا، به یک چالش پژوهشی کلیدی تبدیل شده است. روش‌های سنتی، به دلیل ماهیت ایستا و وابستگی به مدل‌های ثابت، در مواجهه با تغییرات محیطی کارآمد نیستند. در سال‌های اخیر، الگوریتم‌های یادگیری تقویتی عمیق به‌عنوان رویکردی داده‌محور و انطباق‌پذیر برای حل این مسئله، مورد توجه فزاینده‌ای قرار گرفته است. این مقاله مروری با نگاهی نظام‌مند به جدیدترین مطالعات، به بررسی کاربرد یادگیری تقویتی عمیق در زمان‌بندی و انتقال وظایف در محیط‌های ترکیبی ابری-مه-لبه می‌پردازد. مقالات منتخب از نظر نوع الگوریتم یادگیری (مانند Rainbow DQN، DDQN، DQN)، معماری سیستم (متمرکز، توزیع‌شده، سلسله‌مراتبی)، اهداف بهینه‌سازی (مانند کاهش زمان اجرا، مصرف انرژی، نرخ نقض مهلت زمانی) و نوع داده‌های ارزیابی (شبیه‌سازی‌شده یا واقعی مانند Google Workload و داده‌های ترافیکی) طبقه‌بندی شده‌اند. نتایج این مرور نشان می‌دهد که یادگیری تقویتی عمیق، به‌ویژه در شرایط متغیر و با محدودیت‌های بلادرنگ، موجب بهبود عملکرد در معیارهایی نظیر تأخیر، مصرف انرژی و تخصیص منابع می‌گردد. در پایان، چالش‌های باز و مسیرهای آینده برای توسعه راهکارهای تطبیقی و مقیاس‌پذیر ارائه می‌شود.

کلمات کلیدی: یادگیری تقویتی عمیق، زمان‌بندی وظایف، رایانش مه، رایانش لبه، رایانش ابری، انتقال وظایف، بهینه‌سازی منابع، سیستم‌های بلادرنگ

۱. مقدمه

با گسترش سامانه‌های هوشمند توزیع شده و افزایش پیچیدگی وظایف پردازشی، نیاز به تصمیم‌گیری تطبیقی و پاسخ‌دهی سریع به یکی از الزامات اصلی در طراحی زیرساخت‌های محاسباتی مدرن بدل شده است. گسترش پرشتاب فناوری‌هایی



مانند اینترنت اشیا^۱، وسایل نقلیه خودران، واقعیت افزوده و کاربردهای صنعتی بلادرنگ، فشار فزاینده‌ای بر معماری‌های سنتی ایجاد کرده و زیرساخت‌هایی نظیر رایانش ابری، مه و لبه را به گزینه‌هایی پویا و توسعه‌پذیر برای اجرای بارهای محاسباتی تبدیل کرده است [1].

در چنین محیط‌هایی، مسئله‌ی زمان‌بندی و انتقال هوشمند وظایف^۲، به ویژه در شرایط ناهمگون و غیر قطعی، به چالشی اساسی برای تضمین کیفیت خدمات^۳، بهره‌وری منابع و پایداری سیستم‌ها تبدیل شده است. روش‌های کلاسیک مانند اولویت‌دهی، الگوریتم‌های صف‌بندی و حتی متاهیوریستیک‌هایی نظیر بهینه‌سازی گرگ خاکستری^۴ و بهینه‌سازی ازدحام ذرات^۵، به دلیل ساختار ایستا و وابستگی به مدل‌های از پیش تعریف‌شده، توان پاسخ‌گویی مؤثر در محیط‌های پویا را ندارند [2]، [3].

در این راستا، یادگیری تقویتی عمیق^۶ به عنوان رویکردی مبتنی بر تعامل با محیط و بدون نیاز به مدل صریح، امکان استخراج سیاست‌های بهینه در شرایط پیچیده و نامطمئن را فراهم کرده است. مدل‌سازی مسائل زمان‌بندی به صورت فرآیند تصمیم‌گیری مارکوف^۷ و بهره‌گیری از الگوریتم‌هایی نظیر DDQN، Rainbow DQN، DQN و PPO، از جمله رویکردهایی هستند که در مطالعات اخیر مورد استفاده قرار گرفته‌اند. گزارش‌های تجربی نشان می‌دهند که استفاده از یادگیری تقویتی عمیق می‌تواند تا ۳۰٪ مصرف انرژی را کاهش داده و تا ۴۵٪ بهره‌وری منابع را افزایش دهد [4]، [5].

علاوه بر این، نتایج یک مطالعه در سال ۲۰۲۲ نشان داد که به کارگیری یادگیری تقویتی عمیق در محیط‌های ابری تأثیر محسوسی در کاهش زمان اجرا، بهبود توان عملیاتی و صرفه‌جویی داشته است [6]. تحلیل‌های مرور سیستماتیک اخیر نیز بیانگر آن است که بخش عمده‌ای از روش‌های نوین زمان‌بندی در محیط‌های مه و ابر، مبتنی بر یادگیری تقویتی یا ماشینی طراحی شده‌اند [7]. با این حال، بخش زیادی از مطالعات پیشین یا فقط به یک محیط خاص (مانند رایانش لبه) پرداخته‌اند [8]، یا ابعاد زمان‌بندی و انتقال را به صورت منفصل بررسی کرده‌اند.

در این مقاله مروری، با تمرکز بر 20 مقاله منتشر شده بین سال‌های 2021 تا ۲۰۲۵، تلاش شده است رویکردهای یادگیری تقویتی عمیق در زمان‌بندی و انتقال وظایف در محیط‌های ترکیبی ابری-مه-لبه به صورت ساختاریافته و تحلیلی مرور شوند. این مطالعات از نظر نوع الگوریتم یادگیری، ساختار معماری (متمرکز، توزیع‌شده یا سلسله‌مراتبی)، اهداف بهینه‌سازی (مانند کاهش مصرف انرژی، کاهش تأخیر، افزایش نرخ موفقیت در اجرای وظایف) و نوع داده‌های ارزیابی (واقعی یا شبیه‌سازی شده) دسته‌بندی شده‌اند. در پایان نیز چالش‌های باز پژوهشی و مسیرهای توسعه آینده برای طراحی سیستم‌های تطبیقی و مقیاس‌پذیر ارائه شده است.

۲. مروری بر کارهای مرتبط

در سال‌های اخیر، استفاده از الگوریتم‌های یادگیری تقویتی عمیق به عنوان راهکاری نوین در مسائل زمان‌بندی و تخصیص منابع در محیط‌های رایانش ابری، مه و لبه مورد توجه بسیاری از محققان قرار گرفته است. یکی از رویکردهای

¹ IoT

² Task Scheduling and Offloading

³ QoS

⁴ GWO

⁵ PSO

⁶ DRL

⁷ MDP



مهم، بهره‌گیری از مدل‌های یادگیری عمیق در قالب شبکه‌های عصبی برای حل مسائل بهینه‌سازی چندهدفه¹ در زمان‌بندی وظایف بوده است که توانسته‌اند بهبودهایی در کاهش تأخیر و مصرف انرژی ارائه دهند [9].

در این راستا، ترکیب ساختارهایی مانند شبکه عصبی گراف² با الگوریتم‌های تقویتی همچون PPO به منظور درک وابستگی‌های پیچیده بین وظایف و تخصیص بهینه منابع نیز پیشنهاد شده‌اند [10]. برخی دیگر از پژوهش‌ها با تمرکز بر ویژگی‌های کیفیت سرویس و استفاده از رمزگذارهای نویزی در مرحله یادگیری، توانسته‌اند با کاهش نرخ نقض توافق‌نامه سطح خدمات³ و مصرف انرژی، کیفیت سرویس را ارتقا دهند [11]. همچنین استفاده از الگوریتم‌های بازیگر-منتقد⁴ مانند IA3C برای زمان‌بندی تطبیقی در محیط‌های چندابزار⁵، موجب بهبود چشمگیر در معیارهایی نظیر هزینه، انرژی و زمان اجرای وظایف شده است [12].

در حوزه رایانش مه-لبه، طراحی الگوریتم‌هایی مانند DRLMOTS برای تخصیص وظایف بین منابع لبه‌ای و ابری با بهره‌گیری از شبکه‌های Q-learning منجر به بهبود قابل توجهی در تحمل‌پذیری خطا، مصرف انرژی و افزایش کارایی زمان‌بندی شده است [13].

مطالعاتی نیز با هدف ارتقا مقیاس‌پذیری و قابلیت انطباق سریع الگوریتم‌های زمان‌بندی در محیط‌های متغیر، رویکردهای مبتنی بر فرا-یادگیری⁶ را پیشنهاد داده‌اند که نتایج آن‌ها کاهش نیاز به آموزش مجدد را نشان داده‌اند [14]. از سوی دیگر، استفاده از یادگیری خودنظارتی برای زمان‌بندی Zero-Shot در محاسبات لبه موبایل⁷، با ترکیب الگوریتم‌های نظیر SAC و کلاسترینگ وظایف، در بهینه‌سازی مصرف منابع و کاهش تأخیر مؤثر بوده‌اند [15]. در مسائل زمان‌بندی بلادرنگ نیز، رویکردهای مبتنی بر Q-learning دوگانه با هدف دستیابی به تعادل میان زمان‌بندی بهینه و پایداری عملکرد در محیط‌های پویا توسعه یافته‌اند [16].

در برخی مطالعات، با تمرکز بر تنظیم تطبیقی پاداش و نرخ یادگیری در شبکه‌های تقویتی، چارچوب‌هایی مانند EADRL معرفی شده‌اند که در محیط‌های توزیعی و حساس به تأخیر، عملکرد بهتری از خود نشان داده‌اند [17]. در حوزه زمان‌بندی گردش کارهای بلادرنگ، الگوریتم‌هایی نظیر R-DQN ارائه شده‌اند که با در نظر گرفتن وضعیت لحظه‌ای منابع، قادر به تخصیص بهینه وظایف در محیط‌های پویا بوده‌اند [18]. همچنین، در راستای مدیریت بهینه وظایف در محیط‌های دارای منابع محدود، الگوریتم‌هایی مانند DRLIS طراحی شده‌اند که با استفاده از ساختار بازیگر-منتقد و مکانیزم‌های چندسطحی تصمیم‌گیری، توانسته‌اند در محیط‌های لبه‌ای، تعادل بین زمان پاسخ، بار سیستم، و بهره‌وری را بهبود دهند [19].

در نهایت، برای مدیریت وظایف قابل تعویق و زمان‌بندی بهینه آن‌ها در محیط‌های ابری، چارچوب‌های آنلاین مبتنی بر یادگیری تقویتی عمیق نیز توسعه یافته‌اند که تعادل مناسبی میان بهره‌برداری از منابع و رضایت کاربران برقرار می‌سازند [20].

با این حال، علی‌رغم پیشرفت‌های حاصل‌شده، اکثر پژوهش‌های موجود به صورت تک‌بعدی به مسئله زمان‌بندی پرداخته‌اند و یا در محیط‌های کاملاً شبیه‌سازی‌شده و ایستا اجرا شده‌اند. در اغلب موارد، ترکیب هم‌زمان چند معیار واقعی (مانند قابلیت اطمینان، کیفیت سرویس، انرژی، هزینه، مهلت زمان و ...) در محیط‌های پویا و ناهمگن لبه-مه-ابر مورد بررسی

¹ Multi-Objective Optimization

² GNN

³ SLA

⁴ Actor-Critic

⁵ Multi-Cloud

⁶ Meta-RL

⁷ MEC

عمیق قرار نگرفته است. علاوه بر این، روش های پیشنهادی معمولاً فاقد تحلیل حساسیت، ارزیابی نیمه واقعی یا توان تطبیق پذیری در مقیاس بزرگ هستند. از این رو، ضرورت توسعه چارچوب های تطبیقی و چندهدفه مبتنی بر یادگیری تقویتی عمیق که بتواند در شرایط بلادرنگ و در بسترهای لبه محور عمل کنند، همچنان احساس می شود. برای مقایسه ساختارمند پژوهش های بررسی شده در این بخش، جدول شماره ۱ خلاصه ای از ویژگی ها، اهداف، و محیط اجرای مطالعات منتخب را ارائه می دهد.

جدول ۱. مقایسه پژوهش های یادگیری تقویتی عمیق در زمان بندی وظایف

هدف اصلی	نوع الگوریتم یادگیری	محیط اجرا	سال	مرجع
کاهش زمان و انرژی	Multi-DNN	Cloud	2023	[9]
مرور الگوریتم های زمان بندی	PPO + GNN	Cloud	2025	[10]
استفاده از عملکرد لحظه ای سرورها	DQN (Server-Aware)	Cloud	2024	[11]
کاهش هزینه و زمان	IA3C	Multi-Cloud	2024	[12]
کاهش مصرف انرژی و افزایش تحمل پذیری	Q-learning	Fog-Cloud	2025	[13]
تطبیق پذیری در محیط های جدید	Meta-RL	Cloud	2023	[14]
زمان بندی Zero-Shot	SAC + Self-Supervised	Edge	2025	[15]
پایداری و دقت زمان بندی	Double DQN	Cloud	2021	[16]
کارایی در شرایط پویا	Adaptive Actor-Critic	Edge-Cloud	2025	[17]
تخصیص بلادرنگ گردش کارها	R-DQN	Cloud	2024	[18]
تبادل بار و پاسخ در مه	Actor-Critic	Edge-Fog	2024	[19]
مدیریت بارهای قابل تعویق	DQN (Online)	Cloud	2024	[20]

۳. طبقه بندی و تحلیل مقالات منتخب

به منظور تحلیل ساختار یافته مطالعات بررسی شده در حوزه زمان بندی وظایف مبتنی بر یادگیری تقویتی عمیق، مقالات جدول ۱ از سه منظر اصلی شامل محیط اجرا، نوع الگوریتم یادگیری، و اهداف پژوهشی دسته بندی و تحلیل شده اند.

۳.۱. براساس محیط اجرا

بیشتر پژوهش‌های موجود در بستر رایانش ابری پیاده‌سازی شده‌اند، محیطی نسبتاً پایدار و متمرکز که امکان پیاده‌سازی الگوریتم‌ها با تمرکز بر بهینه‌سازی زمان اجرا و مصرف انرژی را فراهم می‌آورد. در مقابل، محیط‌های رایانش مه و لبه که با پویایی، ناهمگنی و محدودیت منابع مواجه‌اند، به‌تازگی مورد توجه قرار گرفته‌اند. برخی مطالعات، از جمله [13] و [19]، تلاش‌هایی اولیه و مؤثر در محیط‌های مه-لبه انجام داده‌اند.

- جمع‌بندی: تمرکز اولیه پژوهش‌ها بر محیط رایانش ابری بوده است، اما تمایل به بهره‌گیری از محیط‌های توزیعی و بلادرنگ در حال افزایش است.
- خلأ تحقیقاتی: نیاز به مطالعات جامع‌تر در محیط‌های ترکیبی و ناهمگن لبه-مه-ابر با سناریوهای نزدیک به واقعیت همچنان احساس می‌شود.

۳.۲. بر اساس نوع الگوریتم یادگیری تقویتی عمیق

مطالعات بررسی‌شده تنوع قابل قبولی از الگوریتم‌های یادگیری تقویتی عمیق را شامل می‌شوند. روش‌های مبتنی بر Q-learning مانند DQN و Double DQN بیشتر در محیط‌های نسبتاً پایدار مورد استفاده قرار گرفته‌اند. در مقابل، الگوریتم‌های بازیگر-منتقد نظیر A3C و نسخه‌های تطبیقی آن برای شرایط پویا و توزیع‌شده مناسب‌تر بوده‌اند. همچنین، پژوهش‌هایی مانند [14] با بهره‌گیری از فرا-یادگیری تلاش کرده‌اند قابلیت انطباق الگوریتم‌ها را در مواجهه با تغییرات محیطی افزایش دهند. مقاله [15] نیز با رویکردی خلاقانه، ترکیب یادگیری خودنظارتی و الگوریتم SAC برای زمان‌بندی Zero-Shot در محاسبات لبه موبایل پیشنهاد داده است.

- جمع‌بندی: انتخاب نوع الگوریتم یادگیری به‌خوبی با ویژگی‌های محیط اجرا هماهنگ شده است.
- خلأ تحقیقاتی: الگوریتم‌های چندعامله^۱ و چارچوب‌های ترکیبی^۲ هنوز به صورت محدود و موردی بررسی شده‌اند.

۳.۳. بر اساس اهداف پژوهشی

اهداف مطالعات مورد بررسی را می‌توان در سه دسته اصلی زیر خلاصه کرد:

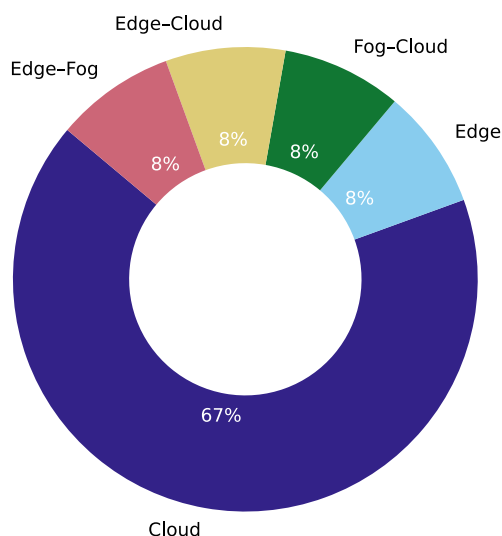
- (۱) کاهش زمان اجرا و انرژی مصرفی،
 - (۲) بهبود پایداری و سازگاری در محیط‌های پویا،
 - (۳) بهینه‌سازی بهره‌وری منابع و زمان پاسخ.
- پژوهش‌هایی مانند [9]، [12] و [20] عمدتاً بر محور اول متمرکز بوده‌اند، درحالی‌که مطالعاتی نظیر [17] و [18]، با استفاده از اطلاعات لحظه‌ای منابع، عملکرد بهتری در شرایط بلادرنگ گزارش کرده‌اند.
- جمع‌بندی: هر یک از پژوهش‌ها بر یکی از اهداف کلیدی تمرکز داشته‌اند؛ اما تنها تعداد معدودی از آن‌ها توانسته‌اند چندین هدف را به طور هم‌زمان بهینه‌سازی کنند.

¹ Multi-Agent DRL

² Hybrid DRL-Heuristic

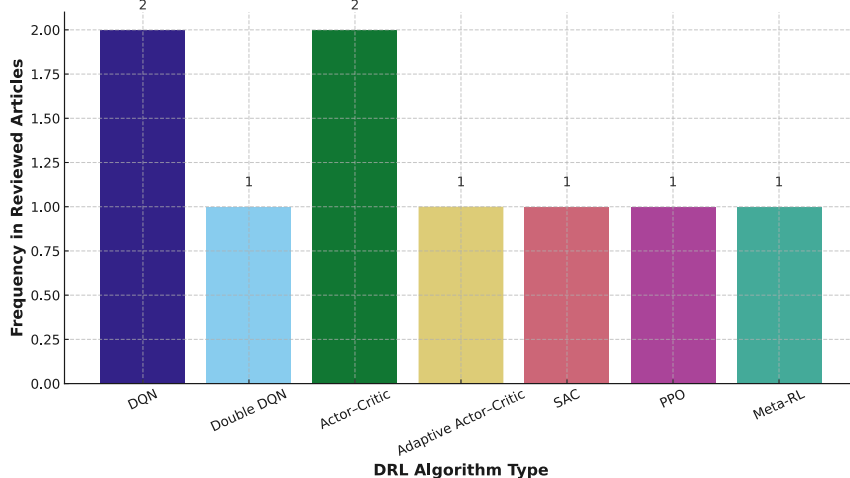
- خلا تحقیقاتی: توسعه رویکردهای چندهدفه که توانایی مدیریت سازش میان معیارهای متضاد را داشته باشند، کمتر مورد توجه قرار گرفته است.
- همانطور که در شکل ۱ دیده می شود، اغلب مطالعات بررسی شده در بستر رایانش ابری انجام شده اند و محیط هایی مانند ابر-مه کمتر مدنظر قرار گرفته اند. همچنین طبق شکل ۲، الگوریتم های DQN و بازیگر-منتقد بیشترین فراوانی را در میان مقالات داشته اند، در حالی که روش هایی مانند PPO، فرا-یادگیری و SAC تنها در یک مطالعه به کار رفته اند.

Distribution of Execution Environments in Reviewed Articles



شکل ۱. توزیع محیط های اجرای مقالات مرور شده در زمینه زمان بندی مبتنی بر یادگیری تقویتی عمیق

Frequency of DRL Algorithms Used in Reviewed Articles



شکل ۲. فراوانی الگوریتم های یادگیری تقویتی عمیق استفاده شده در مقالات مرور شده



۴. جمع بندی تحلیلی و شناسایی خلأهای پژوهشی

بر اساس نتایج تحلیل مقالات مرور شده، چند روند و کاستی مشخص در حوزه زمان بندی مبتنی بر یادگیری تقویتی عمیق در سیستم های لبه-مه-ابر قابل شناسایی است. نخست آن که بسیاری از مطالعات تنها به یک محیط رایانشی مشخص نظیر ابر یا لبه محدود شده اند و از محیط های ترکیبی و سلسله مراتبی مانند لبه-مه-ابر که می توانند عملکرد سیستم را در سناریوهای پویا بهبود دهند، غفلت شده است. این مسئله باعث می شود الگوریتم های توسعه یافته، قابلیت تعمیم و انطباق کافی در سناریوهای عملی نداشته باشند.

دومین خلأ مهم، استفاده محدود از الگوریتم های پیشرفته یادگیری تقویتی عمیق است. بخش عمده ای از مقالات بررسی شده از الگوریتم هایی نظیر DQN، Double DQN و بازیگر-منتقد استفاده کرده اند، در حالی که روش های نوین تری مانند PPO، SAC و فرا-یادگیری که عملکرد بهتری در محیط های پیچیده و غیرقطعی دارند، به ندرت در چارچوب سیستم های لبه-مه-ابر مورد استفاده قرار گرفته اند. این کم تنوعی در انتخاب الگوریتم، میزان نوآوری و انطباق پذیری مدل ها را محدود کرده است.

از منظر اهداف بهینه سازی نیز تمرکز اصلی مقالات بر کاهش انرژی و تأخیر بوده است. در حالی که سایر معیارهای کلیدی مانند کیفیت خدمات، هزینه عملیاتی، نرخ نقض مهلت زمانی و میزان استفاده از منابع کمتر مورد توجه قرار گرفته اند. برای نمونه، در مطالعاتی مانند [12] و [18] تنها دو معیار اصلی در طراحی هدف لحاظ شده اند و سایر ابعاد نادیده گرفته شده اند، که این مسئله می تواند منجر به کاهش تعادل عملکرد سیستم در عمل شود.

در مجموع، سه شکاف عمده پژوهشی شناسایی می شود:

- ۱) فقدان بررسی در محیط های ترکیبی و توزیع شده،
- ۲) استفاده محدود از الگوریتم یادگیری تقویتی عمیق پیشرفته،
- ۳) عدم پوشش جامع معیارهای بهینه سازی.

این خلأها می توانند مسیر روشنی برای تحقیقات آینده در زمینه زمان بندی هوشمند وظایف در سیستم های رایانشی لبه-مه-ابر فراهم آورند.

۵. نتیجه گیری

در این پژوهش مروری، تلاش شد تا با تمرکز بر الگوریتم های یادگیری تقویتی عمیق، روندها، محدودیت ها و فرصت های موجود در حوزه زمان بندی وظایف در محیط های رایانشی مه، لبه و ابر مورد بررسی قرار گیرد. با تحلیل ساختار یافته منابع علمی اخیر، مشخص شد که گرچه استفاده از یادگیری تقویتی عمیق به عنوان رویکردی هوشمند در حال گسترش است، اما هنوز کاستی ها و چالش های قابل توجهی در کاربرد آن در محیط های ترکیبی و ناهمگن وجود دارد. یکی از یافته های کلیدی این پژوهش، نبود پوشش جامع در استفاده از محیط های اجرایی سلسله مراتبی، به ویژه در سناریوهای بلادرنگ، و همچنین فقدان تنوع کافی در الگوریتم های یادگیری تقویتی به کار گرفته شده بود. همچنین مشاهده شد که بسیاری از مطالعات تنها بر مجموعه محدودی از معیارهای بهینه سازی تمرکز داشته اند و ابعاد مهمی نظیر کیفیت خدمات، سازگاری با تغییرات محیطی و بهره برداری بهینه از منابع کمتر مدنظر قرار گرفته اند. بر اساس این تحلیل ها، می توان مسیرهایی برای تحقیقات آینده پیشنهاد کرد. نخست آن که بررسی و توسعه الگوریتم های یادگیری تقویتی عمیق تطبیق پذیر که بتوانند با ساختارهای

ترکیبی لبه-مه-ابر هماهنگ شوند، می‌تواند به بهبود عملکرد زمان‌بندی در شرایط واقعی منجر شود. دوم آن که تلفیق چند معیار بهینه‌سازی در طراحی مدل‌ها، مانند انرژی، هزینه، کیفیت خدمات و میزان استفاده از منابع، می‌تواند موجب افزایش تعادل و پایداری در سیستم شود. در نهایت، به‌کارگیری الگوریتم‌های نوین یادگیری تقویتی عمیق با ساختارهای دوگانه (اکتشاف / بهره‌برداری)، به‌ویژه در محیط‌های پویا، فرصت مناسبی برای ارائه راهکارهای مؤثرتر فراهم می‌سازد.

۶. مراجع

1. N. Khaledian, M. Voelp, S. Azizi, and M. H. Shirvani, "AI-based & heuristic workflow scheduling in cloud and fog computing: a systematic review," *Cluster Comput*, vol. 27, no. 8, pp. 10265–10298, Nov. 2024, doi: 10.1007/s10586-024-04442-2.
2. S. Chandrasiri and D. Meedeniya, "Energy-Efficient Dynamic Workflow Scheduling in Cloud Environments Using Deep Learning," *Sensors*, vol. 25, no. 5, p. 1428, Feb. 2025, doi: 10.3390/s25051428.
3. D. Alsadie and M. Alsulami, "Modified grey wolf optimization for energy-efficient internet of things task scheduling in fog computing," *Sci Rep*, vol. 15, no. 1, Apr. 2025, doi: 10.1038/s41598-025-99837-5.
4. L. Ale, N. Zhang, X. Fang, X. Chen, S. Wu, and L. Li, "Delay-Aware and Energy-Efficient Computation Offloading in Mobile-Edge Computing Using Deep Reinforcement Learning," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 3, pp. 881–892, Sept. 2021, doi: 10.1109/tccn.2021.3066619.
5. B. Wang, F. Liu, and W. Lin, "Energy-efficient VM scheduling based on deep reinforcement learning," *Future Generation Computer Systems*, vol. 125, pp. 616–628, Dec. 2021, doi: 10.1016/j.future.2021.07.023.
6. K. Siddesha, G. V. Jayaramaiah, and C. Singh, "A novel deep reinforcement learning scheme for task scheduling in cloud computing," *Cluster Comput*, vol. 25, no. 6, pp. 4171–4188, Dec. 2022, doi: 10.1007/s10586-022-03630-2.
7. A. A. Ismail, N. E. Khalifa, and R. A. El-Khoribi, "A survey on resource scheduling approaches in multi-access edge computing environment: a deep reinforcement learning study," *Cluster Comput*, vol. 28, no. 3, June 2025, doi: 10.1007/s10586-024-04893-7.
8. D. Hortelano *et al.*, "A comprehensive survey on reinforcement-learning-based computation offloading techniques in Edge Computing Systems," *Journal of Network and Computer Applications*, vol. 216, p. 103669, July 2023, doi: 10.1016/j.jnca.2023.103669.
9. Q. Li, Z. Peng, D. Cui, J. Lin, and H. Zhang, "UDL: a cloud task scheduling framework based on multiple deep neural networks," *J Cloud Comp*, vol. 12, no. 1, July 2023, doi: 10.1186/s13677-023-00490-y.
10. Y. Gu *et al.*, "Deep Reinforcement Learning for Job Scheduling and Resource Management in Cloud Computing: An Algorithm-Level Review," 2025, *arXiv*. doi: 10.48550/ARXIV.2501.01007.
11. J. Wang, S. Li, X. Zhang, F. Wu, and C. Xie, "Deep reinforcement learning task scheduling method based on server real-time performance," *PeerJ Computer Science*, vol. 10, p. e2120, June 2024, doi: 10.7717/peerj-cs.2120.
12. S. Mangalampalli *et al.*, "Efficient deep reinforcement learning based task scheduler in multi cloud environment," *Sci Rep*, vol. 14, no. 1, Sept. 2024, doi: 10.1038/s41598-024-72774-5.

13. P. Choppara and S. Mangalampalli, "An efficient deep reinforcement learning based task scheduler in cloud-fog environment," *Cluster Comput*, vol. 28, no. 1, Feb. 2025, doi: 10.1007/s10586-024-04712-z.
14. X. Xiu, J. Li, Y. Long, and W. Wu, "MRLCC: an adaptive cloud task scheduling method based on meta reinforcement learning," *J Cloud Comp*, vol. 12, no. 1, May 2023, doi: 10.1186/s13677-023-00440-8.
15. P. Khoshvaght *et al.*, "A self-supervised deep reinforcement learning for Zero-Shot Task scheduling in mobile edge computing environments," *Ad Hoc Networks*, vol. 178, p. 103977, Nov. 2025, doi: 10.1016/j.adhoc.2025.103977.
16. S. Swarup, E. M. Shakshuki, and A. Yasar, "Task Scheduling in Cloud Using Deep Reinforcement Learning," *Procedia Computer Science*, vol. 184, pp. 42–51, 2021, doi: 10.1016/j.procs.2021.03.016.
17. J. Anand and B. Karthikeyan, "EADRL: Efficiency-aware adaptive deep reinforcement learning for dynamic task scheduling in edge-cloud environments," *Results in Engineering*, vol. 27, p. 105890, Sept. 2025, doi: 10.1016/j.rineng.2025.105890.
18. J. Pan and Y. Wei, "A deep reinforcement learning-based scheduling framework for real-time workflows in the cloud environment," *Expert Systems with Applications*, vol. 255, p. 124845, Dec. 2024, doi: 10.1016/j.eswa.2024.124845.
19. Z. Wang, M. Goudarzi, M. Gong, and R. Buyya, "Deep Reinforcement Learning-based scheduling for optimizing system load and response time in edge and fog computing environments," *Future Generation Computer Systems*, vol. 152, pp. 55–69, Mar. 2024, doi: 10.1016/j.future.2023.10.012.
20. H. Dong *et al.*, "An Advanced Reinforcement Learning Framework for Online Scheduling of Deferrable Workloads in Cloud Computing," June 03, 2024, *arXiv*: arXiv:2406.01047. doi: 10.48550/arXiv.2406.01047.