بررسی الگوریتم‌های پردازش کوانتومی در هوش مصنوعی و یادگیری ماشین

حسین کاظمی
کارشناسی‌ارشد فناوری اطلاعات، دانشگاه بین‌المللی امام رضا (علیه‌السلام)

چکیده
الگوریتم‌های پردازش کوانتومی در یادگیری ماشین موضوعی پیشرفته و جذاب است که در سال‌های اخیر تحقیقات زیادی پیرامون این موضوع صورت‌گرفته است. الگوریتم‌های گسترده‌ای نیز در این راستا به جهت دست‌یافتن به راه‌حل‌های سریع در پردازش کوانتومی نیز ارائه گردیده است. این مقاله به بررسی و مقایسهٔ این الگوریتم‌ها و تأثیر آن‌ها بر عملکرد مدل‌های یادگیری ماشین می‌پردازیم. این چکیدهٔ مقاله شامل مروری بر الگوریتم‌های مهم پردازش کوانتومی مانند الگوریتم‌های کوانتومی برای محاسبات خطی، کاهش بعد و یادگیری نمایش داده‌ها، و همچنین بررسی تأثیر این الگوریتم‌ها بر عملکرد مدل‌های یادگیری ماشین است. در انتها نیز دو الگوریتم پرکاربرد این حوزه یعنی آنیل کوانتومی و محاسبات کوانتومی مبتنی بر گیت با یادگیری تقویتی عمیق کلاسیک را باتوجه‌به مقالات ارائه شده در این حوزه بررسی اجمالی خواهیم نمود و هر یک از این دو الگوریتم را بر اساس نتایج ارائه شده در بررسی‌های مختلف با یکدیگر مقایسه می‌نماییم. امید است که این مقاله که تنها به بررسی اجمالی سایر مقالات در حوزه پردازش کوانتومی است مقدمه‌ای برای شروع تحقیقات علاقه‌مندان به این حوزه را فراهم آورد.

واژگان کلیدی: محاسبات کوانتومی، محاسبات کوانتومی مبتنی بر گیت، محاسبات کوانتومی مبتنی بر آنیل کوانتومی، پردازش کوانتومی در یادگیری ماشین

مقدمه

کامپیوترهای کوانتومی[۱]، ماشین‌هایی هستند که مبتنی بر بیت‌های کوانتومی[۲] هستند که از قوانین مکانیک کوانتومی[۳] بهره کامل می‌برند. آنها وعده عملکرد بهتر از قابلیت‌های محاسباتی کلاسیک فعلی را دارند. این انتظار وجود دارد که یادگیری ماشین یکی از تأثیرگذارترین برنامه‌های قاتل برای نسل اول کامپیوترهای کوانتومی تجاری در دسترس باشد و به‌ویژه برای سخت‌افزاری که در طول به‌اصطلاح کوانتومی با مقیاس متوسط پر سروصدا (NISQ) بر چشم‌انداز عصر کامپیوترهای کوانتومی تسلط داشته باشد. پیش‌بینی می‌شود که یادگیری ماشین در کوتاه‌مدت هم توسط آنیل کننده‌های کوانتومی آدیاباتیک (AQA) و هم از کامپیوترهای کوانتومی دروازه‌دار تأثیر می‌پذیرد. AQAهای تجاری با بیش از ۱۰۰ کیوبیت برای بیش از یک دهه در دسترس بوده اند، درحالی‌که در حال حاضر هیچ کامپیوترهای کوانتومی دروازه ای با بیش از ۱۰۰ کیوبیت به‌صورت تجاری در دسترس نیست. در نتیجه، در حال حاضر سابقه ای غنی از به کارگیری AQA در مشکلات یادگیری ماشین وجود دارد. از سوی دیگر، سخت‌افزار کامپیوترهای کوانتومی دروازه‌ای امروزه توجه زیادی را به خود جلب می‌کند، زیرا این پایه‌ای است که یک کامپیوتر کوانتومی باهدف عمومی (و نه یک برنامه خاص) آینده بر آن استوار است. پیشرفت‌های اخیر در سخت‌افزار کامپیوتر کوانتومی ، الگوریتم‌ها، و برنامه‌های کاربردی اثبات مفهوم اولیه نشان می‌دهد که این دستگاه‌ها نیز نفوذ قابل‌توجهی به یادگیری ماشین داشته‌اند و آنها در حال حاضر از سهم بزرگی از انتشارات علمی مرتبط با کوانتومی یادگیری ماشین برخوردار هستند. یکی دیگر از رویکردهایی که در سال‌های اخیر به آن پرداخته شده است مربوط به محاسبات کوانتومی از جمله محاسبات مبتنی بر بر بازپخت[۴] و گیت‌محور[۵] است. پیاده‌سازی این رویکردها برای یافتن خط‌مشی بهینه برای عبور از یک شبکه و مقایسه آنها با رویکرد یادگیری تقویت عمیق کلاسیک ارائه می‌دهد.

روش تحقیق

در این تحقیق حدود ۲۱۰ مقاله که با کلیدواژه‌های ارائه شده مرتبط بود از طریق پایگاه‌های علمی معتبر مورد بررسی قرار گرفت که با بررسی چکیده‌ها و در برخی از موارد متن مقاله بسیاری از آنها از چرخه مطالعه خارج و در نهایت ۵ مقاله به‌عنوان مقاله پایه و از سایر مقالات نیز بخش‌هایی در بعضی از تعاریف مورداستفاده قرار گرفت.

اصول یادگیری ماشینی

یادگیری ماشینی یک علم چندرشته‌ای است که الگوریتم‌های مختلف را برای استخراج اطلاعات معنی‌دار از داده‌های موجود و ارائه راه‌حل‌های خودکار برای مسائل پیچیده محاسباتی مطالعه می‌کند. قدرت الگوریتم‌های یادگیری ماشین در توانایی آن‌ها برای یادگیری از داده‌های موجود نهفته است و به‌این‌ترتیب یادگیری ماشین به‌جای مدل‌سازی، مبتنی بر داده است و دارای سه مدل یادگیری نظارت شده، یادگیری بدون نظارت و یادگیری تقویتی است (Koshka, 2021).

1- Quantum computers (QC)

۲ - کیوبیت

3- Quantum mechanics (QM)

4- annealing-based

5- gate-based

معرفی مدل یادگیری تقویتی

همان‌طور که در بخش قبل بیان شد یکی از مدل‌های یادگیری ماشین، مدل یادگیری تقویتی است. یادگیری تقویتی را می‌توان برای کاربردهای مختلف، از روبات‌های مستقل (Zhu et al., 2016) تا تعیین تعاملات اجتماعی و اقتصادی بهینه (Arel et al., 2010) استفاده کرد. یادگیری تقویتی عوامل هوشمندی را طراحی می‌کند که قادر به تعامل با دنیای بیرون برای انجام موفقیت آمیز وظایف خاص، مانند یافتن یک هدف یا به دست آوردن پاداش های خاص هستند. در طول سال‌ها، یادگیری تقویتی پیشرفت‌های زیادی را به خود دیده است، به‌ویژه استفاده از شبکه‌های عصبی برای رمزگذاری کیفیت ترکیب‌های حالت - عمل. از آن زمان به بعد، با موفقیت در بازی های پیچیده ای مانند Go (Silver et al., 2016) و حل مکعب روبیک (Agostinelli et al., 2019) اعمال شد.

مدل‌های یادگیری تقویتی برای کاوش در یک محیط ناشناخته به‌صورت مقرون‌به‌صرفه مفید هستند. در محیط‌های متخاصم، انتخاب بهترین راه ممکن می‌تواند مرگ و زندگی باشد؛ بنابراین، مدل‌های یادگیری ماشین مصنوعی اغلب به بهترین تصمیم کمک می‌کنند. مدل‌های یادگیری تقویتی می‌توانند یک مسیر مقرون‌به‌صرفه را از طریق محیط‌های ناشناخته پیدا کنند. با ارائه مدل یادگیری تقویتی با یک نمای کلی ساده از محیط، همراه باهدف و مکان‌ها یا مسیرهای احتمالاً خطرناکی که طرف‌های دشمن کنترل می‌کنند، مدل مسیری را جستجو می‌کند که به هدف برسد و درعین‌حال کمترین هزینه را متحمل شود. تعریف هزینه در هر مورداستفاده متفاوت است. اگر بخواهیم مسیری بین دونقطه پیدا کنیم، می‌توانیم هزینه را به‌عنوان طول مسیر یافت شده تعریف کنیم، اما در شرایط خصمانه، باید هزینه را به‌عنوان معیاری برای ایمنی یک مسیر خاص تعریف کنیم. مدل یادگیری تقویتی به طور مؤثر سیاستی را می‌آموزد که دیکته می‌کند کدام عمل باید در یک وضعیت انجام شود. پاداش تجمعی آینده مورد انتظار یک ترکیب حالت - عمل معین، ارائه شده توسط Q(s, a)، کیفیت یک سیاست معین را تعیین می‌کند.

در محیط‌های ساده، حتی بدون محاسبه صریح مقادیر Q، می‌توان خط‌مشی بهینه را به‌راحتی پیدا کرد. بااین‌حال، در محیط‌های پیچیده با متغیرهای زیاد، انسان‌ها در یافتن مسیر بهینه با مشکل مواجه می‌شوند و مدل‌های کامپیوتری کار را به دست می‌گیرند. با پیچیده‌تر شدن محیط‌ها، حتی رایانه‌ها نیز می‌توانند مشکلاتی داشته باشند و قدرت محاسباتی آنها گاهی اوقات کافی نیست (Thompson et al., 2020). این باعث پیشرفت یادگیری می‌شود، مانند یادگیری برنامه درسی که در آن به‌تدریج محیط را پیچیده می‌کند(Bengio et al., 2009). پیشرفت های محاسباتی مبتنی بر سخت‌افزار، مانند یادگیری تقویتی توزیع شده در سیستم های CPU-GPU (Inci et al., 2020) و یادگیری تقویتی کوانتومی می باشند(Dunjko et al., 2017 و Phillipson, 2020).

در سال‌های اخیر در این حوزه‌ها تحقیقات زیادی انجام شده است، هم با استفاده از رایانه‌های کوانتومی مبتنی بر گیت و هم از رایانه‌های کوانتومی مبتنی بر بازپخت به‌عنوان پلت فرم محاسباتی. رویکردهای مبتنی بر گیت می‌توانند از الگوریتم جستجوی گراور برای یافتن بهترین اقدام جدید (Dong et al., 2008a; Paparo et al., 2014) یا مدل‌سازی تعاملات پیچیده بین عامل و محیط در برهم‌نهی استفاده کنند (Dunjko et al., 2017; Jerbi et al., 2019). در اینجا، مرحله یادگیری تا حدی کوانتومی است و خط‌مشی بهینه با استفاده از منابع کلاسیک یا کوانتومی ذخیره می‌شود. رویکرد کوانتومی مبتنی بر بازپخت شامل یک

الگوریتم برای آموزش کارآمد ماشین بولتزمن کوانتومی با استفاده از یک آنیل کوانتومی است (Crawford et al., 2016; Levit et al., 2017). ماشین کوانتومی بولتزمن خطمشی بهینه را ذخیره می‌کند.

سخت‌افزار کوانتومی فعلی هنوز در دست توسعه است و سخت‌افزار معمولاً نویز دارد؛ بنابراین، دستگاه‌های کوانتومی فعلی، دستگاه‌های کوانتومی در مقیاس متوسط نویز[6] نامیده می‌شوند (Preskill, 2018). بااین‌حال، حتی این دستگاه‌های NISQ قبلاً در حل مسائل خاص مفید هستند (Lau et al., 2022). دستگاه‌های NISQ مبتنی بر گیت می‌توانند به شبیه‌سازی سیستم‌های چند بدنه کوانتومی کمک کنند (Ritter, 2019). علاوه بر این، هر دو دستگاه NISQ مبتنی بر گیت و مبتنی بر بازپخت می‌توانند به حل مشکلات بهینه‌سازی کمک کنند. به‌عنوان‌مثال می‌توان به تهدید رمزگذاری AES با فرمول بندی آن به‌عنوان یک مسئله بهینه‌سازی (Wang et al., 2022) و پیاده‌سازی مدل‌های یادگیری ماشین کوانتومی (Hu et al., 2019; N. Neumann et al., 2019) و شبکه های عصبی کوانتومی (Wei et al., 2021) اشاره کرد.

در ادامه قابلیت‌های یادگیری ماشین کوانتومی را برای یادگیری تقویتی تحلیل می‌شود. عملکرد هر دو روش کوانتومی مبتنی بر گیت و مبتنی بر بازپخت را با رویکرد مبتنی بر یادگیری تقویت عمیق کلاسیک مقایسه می‌شود. به طور خاص عواملی را در یک محیط ناشناخته در نظر می‌گیریم که باید به یک هدف برسند. محیط ناشناخته می‌تواند هم حالت‌های مانع و هم حالت‌های مجازات داشته باشد. بازدید از ایالت جریمه هزینه زیادی را به همراه خواهد داشت. همچنین امکان تصادفی بودن را در اقدامات عامل‌ها فراهم می‌کند، باتوجه‌به یک حالت و یک عمل، عامل‌ها فقط با احتمال کمی به حالت موردنظر حرکت می‌کنند و در غیر این صورت به موقعیت شبکه مجاور حرکت می‌کنند. همچنین یک تکنیک یادگیری بهبودیافته به نام یادگیری برنامه درسی را معرفی شده است که در آن محیط به‌تدریج پیچیده‌تر می‌شود. در (Crawford et al., 2016; Levit et al., 2017)، یک رویکرد برای پیمایش شبکه برای عوامل منفرد با استفاده از آنیل کوانتومی ارائه شده است. این کار بعداً به تنظیمات با چندین عامل که به طور جمعی به اهداف خاصی می‌رسند در (N. M. P. Neumann et al., 2020) گسترش یافت.

در بخش‌های بعدی ابتدا کارهای انجام شده در این حوزه، سپس دو رویکرد کوانتومی برای پیاده‌سازی مدل‌های یادگیری تقویتی را مورد بررسی قرار می‌دهیم. سپس، مجموعه آزمایشی که در مقالات این حوزه انجام شده است را توضیح خواهیم داد و نتایج را با یادگیری تقویتی کلاسیک مقایسه و بحث خواهیم کرد.

رویکرد محاسبات کوانتومی

کامپیوترهای کوانتومی از اثرات کوانتومی برای انجام محاسبات استفاده می‌کنند. روشی که رایانه‌های کوانتومی این عملیات را اجرا می‌کنند و اینکه کدام عملیات پشتیبانی می‌شوند، می‌تواند متفاوت باشد. دو رویکرد رایج برای محاسبات کوانتومی، محاسبات کوانتومی مبتنی بر بازپخت و محاسبات کوانتومی مبتنی بر گیت است. این رویکردها به ترتیب مشابه محاسبات آنالوگ کلاسیک و محاسبات دیجیتال کلاسیک هستند.

---

[6] - NISQ

رویکرد محاسبات کوانتومی مبتنی بر آنیل

محاسبات کوانتومی مبتنی بر بازپخت یا آنیل کوانتومی بر اساس کار کادواکی و نیشی مور (Kadowaki & Nishimori, 1998) است. بسیاری از مسائل قبلاً با استفاده از آنیل کوانتومی، ارائه راه‌حل‌های معقول در زمان واقعی (Neukart et al., 2017) یا ارائه راه‌حل‌های بهینه یا بسیار خوب سریع‌تر از جایگزین‌های کلاسیک (Benedetti et al., 2017) حل شده‌اند. کاربردهای آنیل کوانتومی متنوع است و شامل بهینه‌سازی ترافیک (Neukart et al., 2017)، امور مالی (Bhatia & Phillipson, 2021)، مشکلات امنیت سایبری (Neukart et al., n.d.) و یادگیری ماشینی (Benedetti et al., 2017; Crawford et al., 2016; Li et al., 2018) هست. در بازپخت کوانتومی، کیوبیت‌ها در حالت برهم‌نهی اولیه قرار می‌گیرند و پس از آن یک همیلتونی مخصوص مسئله بر روی کیوبیت‌ها اعمال می‌شود. اگر همیلتون به‌اندازه کافی آهسته اعمال شود، کیوبیت‌ها در حالت پایه موردنظر باقی می‌مانند و اندازه‌گیری پاسخ مسئله در نظر گرفته شده را نشان می‌دهد.

رویکرد کوانتومی مبتنی بر بازپخت پیشنهادی به‌صراحت تابع Q را برای تعیین خط‌مشی بهینه محاسبه می‌کند. این تابع Q را می‌توان توسط یک ماشین بولتزمن رمزگذاری کرد که دارای یک شبکه عصبی که در آن همه گره‌ها می‌توانند متصل شوند هست. ماشین‌های محدود بولتزمن نوع خاصی از ماشین‌های بولتزمن هستند که در آنها گره‌ها به گره‌های قابل‌مشاهده v و گره‌های پنهان h تقسیم می‌شوند و اتصالات فقط بین گره‌های گروه‌های مختلف وجود دارد. گره‌های قابل‌مشاهده به حالت‌ها و اقدامات ممکن مربوط می‌شوند. می‌توانیم گره‌های پنهان را در چندین لایه پنهان تقسیم نموده که در آن صورت، اتصالات فقط بین گره‌های لایه‌های بعدی وجود دارد. یال‌ها گره‌های مختلف را به هم متصل می‌کنند و وزن‌هایی را می‌توان به این یال‌ها اختصاص داد. وزن مثبت (منفی) نشان‌دهنده ترجیح دو گره مرتبط برای رسیدن به یک مقدار (مخالف) است. گره‌ها یکی از دو مقدار ممکن ۱± را می گیرند. با استفاده از وزن‌های اختصاص‌داده‌شده به گره‌ها، می‌توانیم اولویت را برای یکی از دو مقدار نشان دهیم.

ماشین‌های محدود بولتزمن مدل‌های Ising تصادفی هستند؛ بنابراین، آنیل کننده‌های کوانتومی می‌توانند به تعیین انرژی مرتبط با یک ماشین محدود بولتزمن کمک کنند. انرژی یک ماشین بولتزمن محدود شده توسط جایی که v_i و h_j ذرات ariable هستند که مقادیر گره‌های مرئی و پنهان را نشان می‌دهند و w_{ij} وزن بین گره‌های i و j را نشان می‌دهد. طبق تعریف، اگر گره‌های i و j در لایه‌های بعدی نباشند، $w_{ij} = 0$ است. همه اوزان دوطرفه هستند.

$$E(v,h) = -\sum_i wii\ vi - \sum_j wjj\ hj - \sum_i \sum_j wijvihj \qquad (۱)$$

برای آموزش یک ماشین محدود شده بولتزمن، ابتدا گره‌های قابل‌مشاهده را تعمیر می‌کنیم که به طور مؤثر ترکیب حالت - عمل را برطرف می‌کند. سپس، از آنیل کوانتومی برای تعیین مؤثر انرژی این جفت استفاده می‌شود و در نهایت، وزن ماشین محدود شده بولتزمن را برای بهبود عملکرد، بر اساس برخی معیارها، به‌روز خواهد شد. متریک استفاده شده می‌تواند بین موارد استفاده متفاوت باشد. جزئیات بیشتر در مورد پیاده‌سازی در (Crawford et al., 2016; Levit et al., 2017; N. M. P. Neumann et al., 2020) آورده شده است.

می‌توان کارایی ماشین محدود شده بولتزمن را با اعمال انباشته کردن ماکت افزایش داد: چندین نسخه از طرح‌بندی یکسان به طور هم‌زمان به سخت‌افزار نگاشت می‌شوند و متغیرهای مربوطه در کپی‌های مختلف جفت می‌شوند. این احتمال یافتن پیکربندی‌های غیربهینه را کاهش می‌دهد. باید توجه داشت که سخت‌افزار موجود، اندازه محیط رمزگذاری شده و تعداد اقدامات طبیعتاً محدودیتی

بر تعداد نسخه‌هایی که می‌توان استفاده کرد، اعمال می‌کند. ماشین محدود بولتزمن و وزن‌های آن خط‌مشی موقت را رمزگذاری می‌کند. با تنظیم وزنه‌ها می‌توانیم خط‌مشی بهتری یاد بگیریم.

رویکرد کوانتومی مبتنی بر دروازه

محاسبات کوانتومی مبتنی بر گیت از بسیاری جهات شبیه کامپیوترهای دیجیتال معمولی است. بیشتر مفاهیم کلاسیک با معادل کوانتومی مستقیم خود جایگزین می‌شوند: بیت‌های کوانتومی (کیوبیت) جایگزین بیت‌ها و عملیات کیوبیت جایگزین عملیات بیت می‌شوند. یک تفاوت کلیدی این است که عملیات کوانتومی باید برگشت‌پذیر باشد. بااین‌حال، تمام عملیات کلاسیک را می‌توان با افزودن بیت‌های اضافی برگشت‌پذیر کرد. کامپیوترهای کوانتومی مبتنی بر گیت عملیات را با دست‌کاری دقیق کیوبیت‌های خاص به ترتیب خاصی انجام می‌دهند. حالت کوانتومی حاصل پاسخ را نگه می‌دارد و اندازه‌گیری تنها یکی از نتایج ممکن را با احتمال متناسب با مجذور دامنه آن نتیجه خاص نشان می‌دهد. برای رویکرد کوانتومی مبتنی بر دروازه، به‌جای مدل‌سازی بر همکنش‌های پیچیده عامل - محیط به روش کوانتومی، همان‌طور که رویکرد گراور (Dong et al., 2008b) چنین است، رویکردی را انتخاب کردیم که از الگوریتم جستجوی گراور (Grover, 1996) برای یافتن بهترین اقدام استفاده می‌کند. برای دستگاه‌های NISQ مناسب تر است. در عوض، پیاده‌سازی که تعاملات پیچیده عامل - محیط را در برهم‌نهی مدل‌سازی می‌کند، نیاز به سربار قابل‌توجهی برای تصحیح خطا دارد.

رویکرد کوانتومی مبتنی بر گیت، یک خط‌مشی موقت را در طول یادگیری ذخیره می‌کند که بهترین عملکرد را برای هر حالت، همراه با پاداش مورد انتظار از آن حالت حفظ می‌کند. این رویکرد پس از پایان آموزش، خط‌مشی ذخیره شده را برمی‌گرداند. در هر تکرار، با استفاده از الگوریتم جستجوی گراور، بهترین اقدام را از یک حالت شروع می‌شود. اجازه داده می‌شود تعداد تکرارهای استفاده شده Grover به پاداش مورد انتظار ترکیب موقت حالت - اقدام سیاست بستگی داشته باشد. به‌این‌ترتیب، اطمینان پیدا می‌شود که پس از یافتن آنها و گنجاندن آنها در خط‌مشی موقت، اغلب اقدامات خوب را پیدا می‌کند.

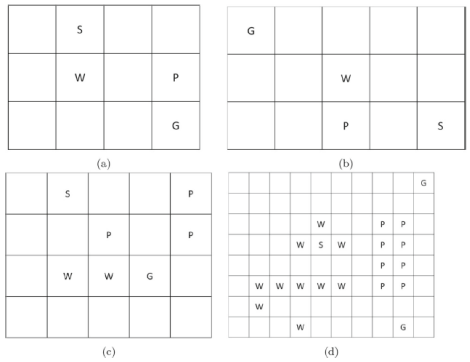
آزمایش‌ها و نتایج بررسی شده

تنظیم آزمایشی

برای پیاده‌سازی مدل یادگیری تقویتی، دو رویکرد کوانتومی وجود دارد، یک رویکرد کوانتومی مبتنی بر بازپخت که یک ماشین بولتزمن محدود را پیاده‌سازی می‌کند و یک رویکرد کوانتومی مبتنی بر دروازه که از الگوریتم جستجوی گراور برای یافتن عمل بهینه استفاده می‌کند. ما هر دو مدل کوانتومی را با یک مدل مقایسه می‌کنیم. رویکرد یادگیری تقویتی عمیق کلاسیک ما عملکرد این رویکردها را در چندین محیط مختلف ارزیابی می‌کنیم. در طول مرحله آموزش هر یک از رویکردها، بهترین خط‌مشی ارائه شده توسط ترکیب دولت - عمل آموخته می‌شود.

محیط‌های مورداستفاده

شکل ۱ محیط‌های مورداستفاده را نشان می‌دهد. محیط ها شامل یک یا چند حالت شروع S و یک یا چند حالت هدف G هستند. یک عامل در حالت شروع، شروع می‌شود و از یک خط‌مشی آموخته شده برای رسیدن به هدف G پیروی می‌کند. این خط‌مشی با استفاده از یکی از سه رویکرد در نظر گرفته شده آموخته می‌شود. علاوه بر این، محیط‌ها شامل حالت‌های مسدود شده W و حالت‌های مجازات P هستند. از هر حالت، عوامل می‌توانند چهار عمل انجام دهند: حرکت به بالا، پایین، چپ و راست. اگر یک عمل باعث شود یک عامل به خارج از محیط حرکت کند یا به حالت مسدود شود، مرحله زمانی بدون تغییر موقعیت عامل پیش می رود. عواملی که از یک حالت پنالتی بازدید می‌کنند، دارای پاداش منفی منهای دویست هستند، درحالی‌که رسیدن به حالت هدف، پاداش مثبت دویست را می‌دهد. بزرگی هر دو مقدار را برابر می‌گیریم، زیرا باید از حالت پنالتی به همان اندازه که باید به هدف رسید نمایندگان اگر قدمی بردارند پاداش منفی کوچکی منهای ده دریافت می‌کنند. به‌صراحت این مقدار را به طور قابل‌توجهی کوچکتر از پاداش در حالت هدف انتخاب می‌شود. این هزینه اندک برای برداشتن یک گام، مسیرهای مستقیم را بر مسیرهای انحرافی ترجیح می‌دهد.



شکل ۱-چهار محیط در نظر گرفته شده. S نشان‌دهنده موقعیت شروع، W نشان‌دهنده حالت مسدود، P نشان‌دهنده وضعیت جریمه، و G نشان‌دهنده وضعیت هدف است.

کنش‌های تصادفی

به عوامل اجازه داده می‌شود که اقدامات قطعی یا تصادفی انجام دهند. با اقدامات تصادفی، یک عامل عملی را با احتمال تنظیم‌پذیر p انجام می‌دهد و با احتمال (1-p)/2 یکی از دو عمل مجاور به‌جای آن انجام می‌شود. به‌عنوان‌مثال، برای p = 0.9، یک عامل با احتمال ۰.۹ به سمت بالا حرکت می‌کند و هر کدام با احتمال ۰.۰۵ به سمت راست یا چپ حرکت می‌کند. اگر p=1 باشد، عمل را قطعی می‌دانیم، در غیر این صورت، آنها را تصادفی می‌نامیم.

اندازه‌گیری عملکرد

می‌توان عملکرد یک خط‌مشی آموخته شده را با استفاده از پاداش مورد انتظار کمی نمود. آثار قبلی نیز وفاداری یک سیاست را در نظر می‌گرفتند. وفاداری برابر با کسری از حالاتی است که عمل صحیحی به آن اختصاص‌داده‌شده است. تعیین اقدام صحیح در هر حالت معمولاً مستلزم ارزیابی صریح محیط است. این فقط برای محیط‌های کوچک و نسبتاً ساده قابل‌اجرا است و بنابراین فقط از وفاداری برای توسعه و آزمایش عملکرد یک رویکرد استفاده می‌شود، نه برای ارزیابی نهایی.

برای محاسبه پاداش مورد انتظار، اقدامات ارائه شده توسط خط‌مشی را دنبال می‌شود، از یک حالت شروع از پیش تعریف شده شروع شده و مجموع پاداش‌ها را در هر مرحله پیگیری می‌شود. ارزیابی این معیار مستقل از اندازه محیط ساده است؛ بنابراین فقط پاداش مورد انتظار را در ارزیابی عملکرد یک سیاست در نظر می‌گیریم.

مرحله آموزش و راهبردهای یادگیری

در مقاله‌های پایه برای تعداد ثابتی از تکرارهای آموزشی تمرین و پاداش مورد انتظار را در طول آموزش محاسبه شده است. در هر دو رویکرد کوانتومی، هر تکرار آموزشی از یک ترکیب حالت - عمل واحد تشکیل شده است. برای رویکرد یادگیری تقویت عمیق کلاسیک، یک تکرار آموزشی یک ارزیابی از یک مسیر از نقطه شروع تا هدف است. برای این تفاوت انتخاب می‌کنیم؛ زیرا در غیر این صورت تعداد تکرارهای آموزشی برای رویکرد تقویت عمیق کلاسیک بسیار زیاد خواهد بود. محاسبه پاداش مورد انتظار، از یک موقعیت شروع S شروع و این خط‌مشی را تا رسیدن به یک حالت هدف یا بازدید حداکثری از وضعیت‌ها دنبال می‌شود. این حداکثر تعداد حالت‌ها به‌عنوان تعداد حالت‌های متمایز در محیط در نظر گرفته می‌شود. در صورت اعمال تصادفی، این عمل را سی بار تکرار و از نتایج میانگین می‌گیریم. در مجموع، به طور مستقل سه رویکرد را ده بار آموزش و میانگین نتایج یافت شده را در این ده اجرا می‌شود. این تغییرات احتمالی را در طول مرحله تمرین جبران می‌کند.

همچنین از دو استراتژی آموزشی متفاوت استفاده می‌شود. در استراتژی آموزشی اول، هر یک از سه رویکرد را با کل محیط ارائه می‌کنیم، درحالی‌که در استراتژی دوم به‌تدریج پیچیدگی محیط را در طول آموزش افزایش می‌دهیم. استراتژی اول را یادگیری مستقیم و استراتژی دوم را یادگیری برنامه درسی (Bengio et al., 2009) نامیده می‌شود. در یادگیری برنامه درسی، محیط در ابتدا فقط شامل حالات انسدادی است. پس از تعداد ثابتی از تکرارهای آموزشی، محیط را با اضافه کردن حالت‌های پنالتی و بعداً اقدامات تصادفی پیچیده می‌کنیم. همچنین اقدامات تصادفی را به‌تدریج معرفی خواهیم نمود، ابتدا با مقادیر p بالا یاد می‌گیریم و به‌تدریج p را کاهش می‌دهیم، جایی که p احتمال انجام عمل صحیح آن است.

انتخاب فراپارامتر

سه رویکرد در نظر گرفته شده دارای برخی فراپارامترهایی هستند که باید تنظیم شود، برای هر دو رویکرد کوانتومی، همچنین باید تعداد نمونه‌های کوانتومی را که در هر تکرار آموزشی گرفته شود، تعیین نماییم. دو ابر پارامتر آخر نرخ یادگیری و ضریب تخفیف هستند. این ضریب تخفیف ارزش فعلی پاداش‌های آینده را می‌سنجد: پاداش آتی کمتر از همان پاداشی است که اکنون به دست می‌آید.

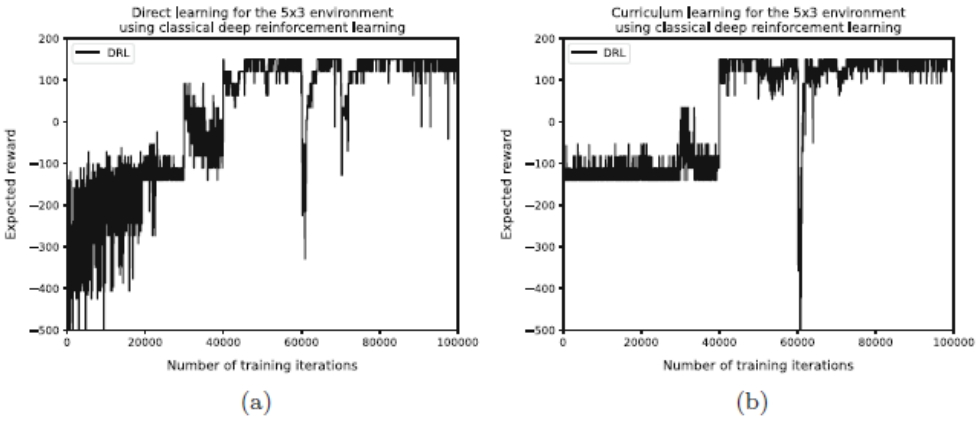
مقادیر کاندید را برای هر یک از فراپارامترها بر اساس (N. M. P. Neumann et al., 2020) انتخاب و یک جستجوی شبکه ای بروی ترکیبات ممکن برای یافتن بهترین تنظیمات انجام می‌شود. کیفیت هر تنظیم را با محاسبه وفاداری خط‌مشی آموخته شده تعیین می‌شود. از محیط ۵×۳- نشان داده شده در شکل b۱ برای یافتن فراپارامترها استفاده شده است، زیرا می‌توان خط‌مشی بهینه را برای این محیط با یک بازرسی سریع بصری پیدا نمود. تنظیمات فراپارامتر را انتخاب شده است که بالاترین میانگین وفاداری را در پنج اجرا مستقل ارائه می‌کند. اگر دو تنظیمات عملکرد متوسط مشابهی را نشان دادند، تنظیمات با پایدارترین عملکرد و سریعترین همگرایی را انتخاب می‌شود.

راه‌اندازی شبیه‌سازی

نتایج هر دو رویکرد کوانتومی را با رویکرد یادگیری تقویتی عمیق کلاسیک مقایسه شده است. در این رویکرد کلاسیک، یک عامل از حالت شروع، شروع به کاوش در محیط می‌کند و امیدوار است که یک حالت هدف پیدا کند. باتوجه‌به زمان تمرین کافی، استراتژی عامل بهبود می‌یابد و او سریع‌تر حالت هدف را پیدا می‌کند. نحوه آموزش با دو رویکرد کوانتومی متفاوت است، به‌ویژه در این که هیچ خط‌مشی صریحی رعایت نمی‌شود و در یک تکرار آموزشی از رویکرد یادگیری تقویتی عمیق کلاسیک، ما مدل را برای یک مسیر کامل به‌جای یک حالت واحد به‌روز می‌شود. رویکرد یادگیری تقویت عمیق کلاسیک را به‌صورت محلی روی یک کامپیوتر شخصی ساده آموزش داده و هر دو رویکرد کوانتومی را شبیه‌سازی شده است: رویکرد مبتنی بر بازپخت کوانتومی با استفاده از مدل‌های بازپخت شبیه‌سازی شده توسط بسته نرم‌افزار Ocean توسط D-Wave و رویکرد کوانتومی مبتنی بر دروازه با استفاده از بسته نرم‌افزاری کوانتومی Qiskit اجرا گردید. انتظار می‌رود که هر دو رویکرد کوانتومی عملکرد مشابهی را روی سخت‌افزار کوانتومی واقعی نشان دهند، حتی در دستگاه‌های NISQ هنوز در معرض نویز هستند.

نتایج یادگیری برنامه درسی در مقابل یادگیری مستقیم

در این بخش، با استفاده از برخی از محیط‌های شکل ۱، عملکرد یادگیری برنامه درسی را با یادگیری مستقیم مقایسه می‌شود. برای هر محیط، پس از نیمی از مراحل آموزشی، حالت‌های جریمه را اضافه و در نتیجه محیط را پیچیده می‌کند. از پاداش مورد انتظار برای تعیین کمیت عملکرد یک سیاست استفاده شده است. در اولین آزمون‌ها، تصادفی بودن بروی صفر قرار گرفته و نتایج یادگیری برنامه درسی را با یادگیری مستقیم مقایسه شده است. نتایج نشان می‌دهد که رویکرد مبتنی بر دروازه نسبتاً سریع یاد می‌گیرد، درحالی‌که رویکرد مبتنی بر آنیل کوانتومی تکرارهای آموزشی بیشتری برای یادگیری نیاز دارد. به طور مشابه، شاهد هستیم که رویکرد یادگیری تقویتی کلاسیک به مراحل آموزشی بیشتری برای یادگیری یک خط‌مشی نیاز دارد. در ابتدا، پاداش یافت شده برای هر سه رویکرد کم است، زیرا عامل به طور مؤثر یک پیاده‌روی تصادفی در محیط انجام می‌دهد. با یادگیری برنامه درسی، هنوز هیچ حالت جریمه‌ای وجود ندارد، بنابراین تنها جریمه از برداشتن گام‌ها ناشی می‌شود.

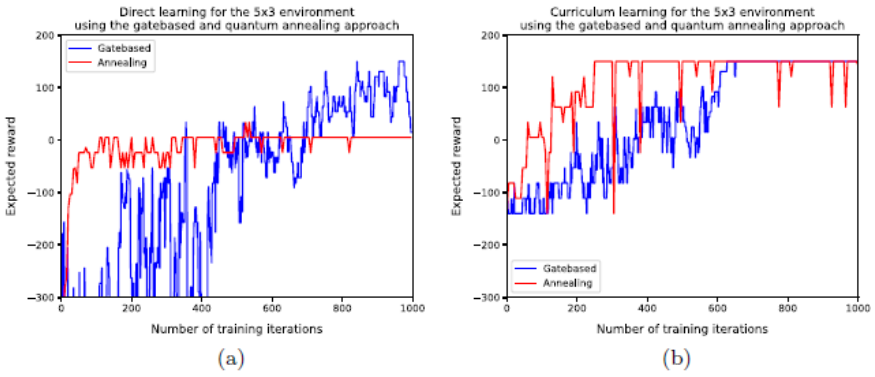


شکل ۲-پاداش مورد انتظار برای یادگیری تقویتی عمیق کلاسیک در محیط ۵×۳ نشان داده شده در شکل b۱

در همه نتایج، همچنین تغییراتی را در پاداش یافت شده مشاهده می‌کنیم که ناشی از تصادفی بودن ذاتی در روش یادگیری است. برای هر دو رویکرد کوانتومی، این ترکیب حالت - عمل انتخابی برای درنظرگرفتن در آن تکرار است و برای رویکرد کلاسیک این مسیری است که برای کاوش انتخاب شده است. علاوه بر این، هنگامی که محیط را پیچیده می‌کنیم، شاهد کاهش پاداش یافت شده هستیم. شکل ۲ و ۳ نتایج را برای محیط ۵×۳ نشان داده شده در شکل b۱ برای یادگیری مستقیم و یادگیری برنامه درسی نشان

می‌دهد. رویکرد یادگیری تقویتی عمیق کلاسیک یک خط‌مشی خوب را می‌آموزد. بااین‌حال، با یادگیری برنامه درسی، عملکرد پایدارتر است. هر دو رویکرد کوانتومی برای یادگیری برنامه درسی یکسان عمل می‌کنند و سیاست بهینه را نسبتاً سریع یاد می گیرند. در یادگیری مستقیم، مشاهده می‌شود که هر دو رویکرد مسیر بهینه را نمی آموزند؛ زیرا این پاداش ۱۵۰: ۲۰۰ از هدف و ۵۰- برای پنج مرحله را تشکیل می‌دهد. رویکرد کوانتومی مبتنی بر دروازه در طول زمان بهبود می‌یابد، درحالی‌که رویکرد مبتنی بر بازپخت کوانتومی به‌سرعت یک خط‌مشی را یاد می‌گیرد، اما پس از آن یادگیری را متوقف می‌کند. یک بازرسی بصری از خط‌مشی متوجه شد که مسیر یافت شده، عامل را قبل از رسیدن به حالت پایانی، از طریق حالت پنالتی هدایت می‌کند که در نتیجه یک پاداش کلی تقریباً صفر است. افت عملکرد در رویکردهای کلاسیک به‌احتمال زیاد ناشی از عدم اطمینان در فرایند یادگیری، یا از دو مسیری است که در ابتدا به یک اندازه خوب به نظر می‌رسند، اما یک مسیر دارای پاداش قابل‌توجهی کمتر است. انتظار می‌رود که دلیل مشابهی باعث افت عملکرد برای رویکردهای کوانتومی مبتنی بر گیت و مبتنی بر بازپخت شود.

حال اگر محیط را پیچیده کنیم، رفتار متفاوتی را می‌بینیم. شکل ۳ و ۴ نتایج را برای محیط ۴ × ۳ نشان داده شده در شکل a۱ نشان می‌دهد. اگرچه این محیط کوچکتر است، اما پیچیده‌تر از محیط ۵ × ۳ است، و ازاین‌رو، قبلاً انتظار عملکرد بدتری را داشتیم. هر دو رویکرد کوانتومی در ابتدا با یادگیری برنامه درسی سریع‌تر از یادگیری مستقیم یاد می گیرند. با یادگیری برنامه درسی، پس از معرفی حالت‌های مجازات در محیط، شاهد کاهش شدید پاداش مورد انتظار هستیم. انتظار داریم که این امر از اولین مرحله یادگیری در یادگیری برنامه درسی که در آن خط مشی‌ای آموخته می‌شود که در محیط پیچیده‌تر کمتر از حد مطلوب است و حذف سریع آن بسیار سخت است، دنبال شود. یک بازرسی بصری از خط‌مشی به دست آمده توسط هر دو رویکرد کوانتومی تحت یادگیری برنامه درسی متوجه شد که تنها یک حالت عملکرد اشتباهی دارد که باعث می‌شود عامل به‌صورت دایره‌ای حرکت کند و پاداش کم پیدا شده را توضیح دهد. در تلاش برای غلبه بر این، عامل گاهی اوقات در حالت پنالتی قرار می‌گیرد، ازاین‌رو پاداش رویکرد مبتنی بر آنیل کوانتومی کاهش می‌یابد.



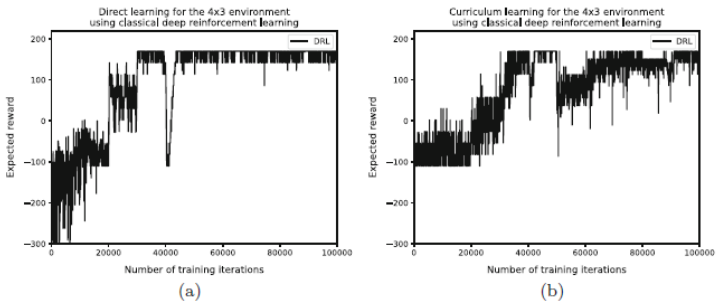
شکل ۳-پاداش مورد انتظار برای یادگیری تقویتی مبتنی بر دروازه و یادگیری تقویتی مبتنی بر بازپخت کوانتومی در محیط ۵ × ۳ نشان داده شده در شکل b۱

کمی‌کردن تأثیر اعمال تصادفی

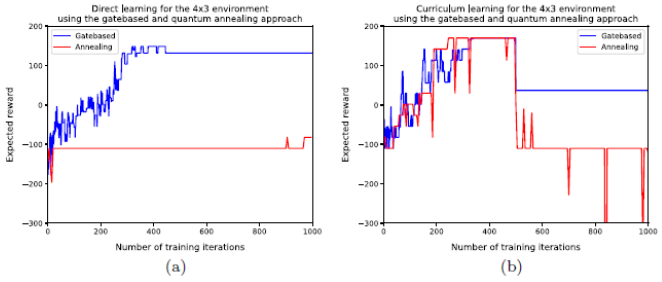
در این بخش، عملکرد رویکردهای خود را در تنظیمات تصادفی تحلیل می‌شود تا ببینیم آیا آنها می‌توانند با آن همکاری کنند یا خیر؛ بنابراین، دو دوره مستقل از یادگیری مستقیم را در نظر گرفتیم، یکی با اقدامات قطعی و دیگری با اقدامات تصادفی. این کار را

برای هر یک از سه رویکرد انجام داده و پاداش مورد انتظار را در طول زمان مقایسه شده است. ازآنجایی‌که عملکرد مطلق تحت کنش‌های تصادفی کمتر است، عمدتاً عملکرد نسبی بین این دو را در نظر گرفته می‌شود. مدل‌ها در ابتدا پاداش کمی پیدا می‌کنند، زیرا عامل به طور مؤثر یک پیاده‌روی تصادفی را انجام می‌دهد. برای برخی از محیط‌ها، مدل در یافتن یک خط‌مشی خوب مشکل دارد و عملکرد در طول مرحله آموزش نسبتاً ثابت می‌ماند، به‌ویژه برای محیط‌های بزرگ‌تر که با هر یک از دو رویکرد کوانتومی ترکیب شده‌اند. یکی دیگر از تأثیرات تصادفی این است که پاداش در طول زمان واریانس بیشتری را نسبت به اقدامات قطعی نشان می‌دهد.

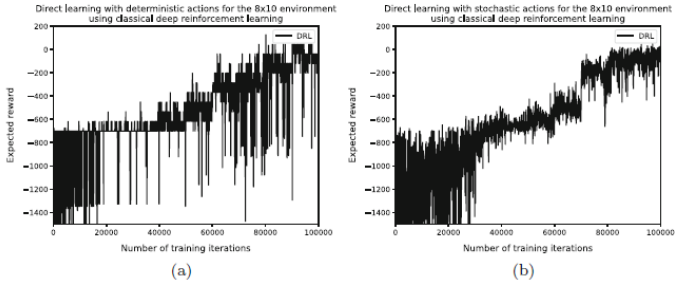
شکل b۶ و b۷ عملکرد رویکرد یادگیری تقویت عمیق کلاسیک و هر دو رویکرد کوانتومی را برای محیط ۸ × ۱۰ نشان می‌دهد. این محیط بزرگترین و پیچیده ترین محیط در نظر گرفته شده است. عملکرد با اقدامات تصادفی مشابه عملکرد با اقدامات قطعی برای هر سه رویکرد است. برای هر دو نسخه کوانتومی، پاداش مورد انتظار پایین‌تر شروع می‌شود، اما شاهد پیشرفت‌هایی به دلیل یادگیری هستیم که نشان می‌دهد پس از مراحل آموزشی کافی، پاداش انتظاری برای اقدامات قطعی و تصادفی هم‌زمان خواهد شد. به‌عنوان آزمون نهایی، اثر ترکیبی دو پسوند در نظر گرفته می‌شود: یادگیری برنامه درسی با اقدامات تصادفی. رویه‌ای که برای یادگیری برنامه درسی اعمال می‌شود به این صورت است که بعد از یک‌چهارم مراحل آموزشی، حالت‌های مجازات را معرفی می‌کنیم. پس از نیمی از مرحله آموزش، اقدامات تصادفی با مقدار p بالا را معرفی می‌شود و پس از سه‌چهارم مراحل آموزش، مقدار p را کاهش می‌دهیم که منجر به تصادفی بیشتر در اقدامات می‌شود. تعداد مراحل آموزشی را برای هر یک از محیط‌های در نظر گرفته شده دوبرابر می‌کنیم.



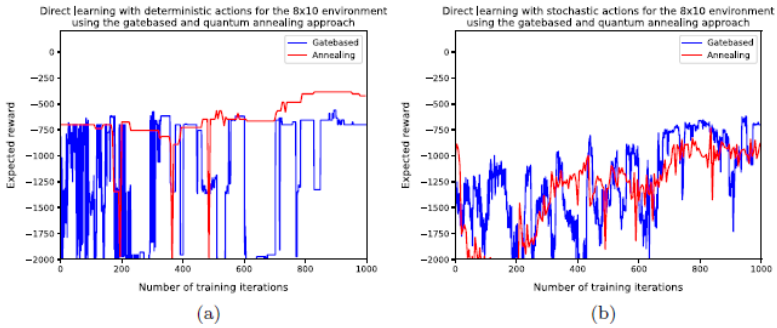
شکل ۴-پاداش مورد انتظار برای یادگیری تقویتی عمیق کلاسیک در محیط ۴×۳ نشان داده شده در شکل a۱



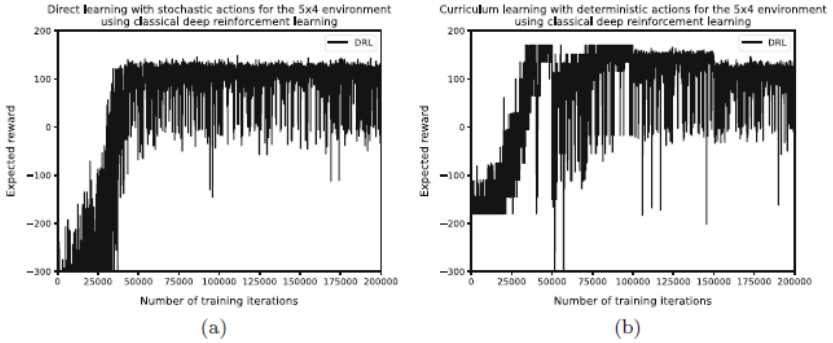
شکل ۵-پاداش مورد انتظار برای یادگیری تقویتی مبتنی بر دروازه و یادگیری تقویتی مبتنی بر بازپخت کوانتومی در محیط ۴ × ۳ نشان داده شده در شکل a۱



شکل ۶-پاداش مورد انتظار برای یادگیری تقویتی عمیق کلاسیک در محیط ۸×۱۰ نشان داده شده در شکل d۱



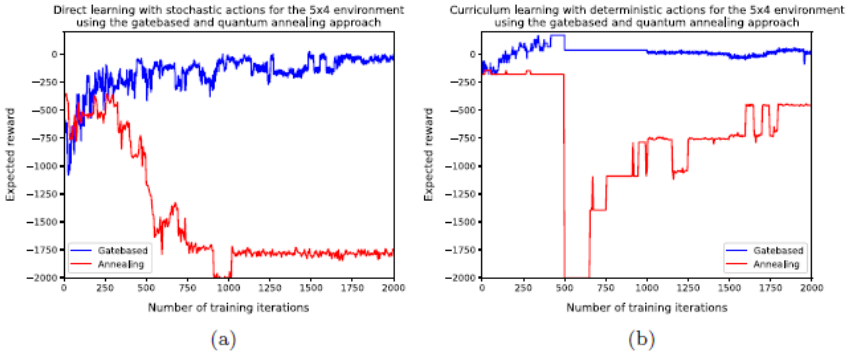
شکل ۷-پاداش مورد انتظار برای یادگیری تقویتی مبتنی بر دروازه و مبتنی بر بازپخت کوانتومی در محیط ۸ × ۱۰ در شکل d۱



شکل ۸-پاداش مورد انتظار برای یادگیری تقویتی عمیق کلاسیک در محیط ۴×۵ نشان داده شده در شکل c۱

شکل‌های a۸ و a۹ عملکرد رویکرد یادگیری تقویتی عمیق کلاسیک و دو رویکرد کوانتومی برای محیط ۴×۵ را نشان می‌دهند که در شکل c۱ نشان داده شده است. مشاهده می‌شود که رویکرد یادگیری تقویت عمیق کلاسیک به پاداش‌های مشابهی می‌رسد، اما این کار را زودتر انجام می‌دهد. . به طور مشابه، رویکرد مبتنی بر دروازه به عملکرد مشابهی برای هر دو استراتژی یادگیری می رسد، اما با یادگیری برنامه درسی، پاداش پایدارتر است. رویکرد مبتنی بر بازپخت کوانتومی هنگام استفاده از یادگیری برنامه درسی نسبت به یادگیری مستقیم، بهبود قابل‌توجهی را نشان می‌دهد، بااین‌حال، عملکرد مطلق کمتر از دو رویکرد دیگر است. یک جنبه جالب این است که با یادگیری برنامه درسی، زمانی که محیط را پیچیده می‌کنیم، شاهد افت قابل‌توجهی در عملکرد هستیم. این افت نشان

می‌دهد که خط‌مشی آموخته شده تا کنون برای محیط‌های پیچیده‌تر نا بهینه بوده است؛ بنابراین، این رویکرد باید بخشی از سیاست این محیط جدید را دوباره یاد بگیرد.



شکل ۹- پاداش مورد انتظار برای یادگیری تقویتی مبتنی بر گیت و مبتنی بر بازپخت کوانتومی در محیط ۵ × ۴ در شکل c۱

بررسی عملکرد نتایج

در بخش‌های قبل، نتایج آزمایش‌های چندگانه برای رویکرد کلاسیک و دو رویکرد کوانتومی برای پیمایش شبکه ارائه شده است. تأثیر یادگیری برنامه درسی و تأثیر اقدامات تصادفی را بر عملکرد روی رویکردها در نظر گرفتیم. دریافتیم که در برخی از محیط‌ها، دو رویکرد کوانتومی به مراحل آموزشی بسیار کمتری نسبت به رویکرد یادگیری تقویت عمیق کلاسیک برای دستیابی به عملکرد مشابه نیاز دارند. در محیط‌های دیگر، عملکرد رویکردهای کوانتومی کمی عقب‌تر بود، بااین‌حال، همچنان از تکرارهای آموزشی کمتری استفاده می‌کرد. یک راه‌حل ممکن، تنظیم بهتر فراپارامترها یا تغییر در تنظیمات یادگیری است. از لحاظ زمانی، هیچ مقایسه‌ای بین رویکردهای مختلف نیست، زیرا رویکردهای کوانتومی شبیه‌سازی شده و از پشتوانه‌های سخت‌افزاری مختلف برای این شبیه‌سازی‌ها و نتایج کلاسیک استفاده شده است. در نتیجه زمان اجرای آزمایش‌ها غیرقابل‌مقایسه و کنار گذاشته شده است.

اولین بسط از مدل‌ها یک تکنیک یادگیری متفاوت است: یادگیری برنامه درسی. با یادگیری برنامه درسی، به‌تدریج محیط را به امید یادگیری سریع‌تر پیچیده می‌شود. به‌استثنای محیط ۴×۳، هنگام استفاده از یادگیری برنامه درسی به‌جای یادگیری مستقیم، همگرایی سریع‌تر به یک خط‌مشی با پاداش مورد انتظار بالا را مشاهده می‌کنیم. مصنوعاتی که در محیط‌های کوانتومی مشاهده می‌شود، زمانی که محیط را پیچیده می‌کنیم، افت شدید عملکرد است. انتظار داریم که دو رویکرد کوانتومی یک خط‌مشی را برای محیط ساده یاد بگیرند و در خارج کردن بخش‌های بهینه این سیاست در محیط پیچیده‌تر مشکل داشته باشند. یک راه‌حل برای این کار انتقال به یک محیط پیچیده‌تر است، به محض اینکه یک سطح عملکرد معین به دست آید، نه پس از تعداد ثابتی از تکرارهای آموزشی. توسعه دوم اجازه اقدامات تصادفی انجام شده توسط عوامل را می‌داد. با برخی احتمالات، اقدامی متفاوت از آنچه در ابتدا در نظر گرفته شده بود انجام می‌شود. دریافتیم که برای هر رویکرد، عملکرد تحت کنش‌های قطعی نزدیک به عملکرد در کنش‌های تصادفی است. هنگام ترکیب هر دو پسوند، می‌بینیم که به‌ویژه رویکرد مبتنی بر بازپخت کوانتومی از استراتژی یادگیری برنامه درسی سود می‌برد. دو رویکرد دیگر نهایی مشابهی را جهت عملکرد برای هر دو استراتژی یادگیری نشان می‌دهند.

نکته قابل‌توجه تفاوت در یادگیری بین رویکرد یادگیری تقویتی عمیق کلاسیک و دو رویکرد کوانتومی است. رویکرد کلاسیک به یک عامل اجازه می‌دهد تا محیط را از حالت شروع بررسی کند و از کل مسیر طی شده برای به‌روزرسانی خط‌مشی برای هر تکرار آموزشی

استفاده کند. دو رویکرد کوانتومی تنها ترکیب‌های تک حالت - عملی را در هر تکرار آموزشی در نظر می‌گیرند و سیاست را فقط بر اساس نتایج آن ترکیب حالت - عمل به‌روز می‌کنند.

نتیجه‌گیری

در این مقاله، دو رویکرد کوانتومی برای پیمایش شبکه با استفاده از یادگیری تقویتی، یک رویکرد مبتنی بر دروازه و یک رویکرد مبتنی بر آنیل کوانتومی در نظر گرفته شده است. مدل‌های قبلی را با گنجاندن اقدامات تصادفی و استفاده از یک تکنیک یادگیری جدید به نام یادگیری برنامه درسی گسترش داده شده است. پاداش مورد انتظار یک خط‌مشی آموخته شده را با رویکرد کوانتومی با خط‌مشی آموخته شده با استفاده از یادگیری تقویت عمیق کلاسیک مقایسه و این کار را برای هر دو پسوند انجام شده است. متوجه شدیم که برای برخی از محیط‌ها، رویکردهای کوانتومی سریع‌تر از رویکرد کلاسیک با ضریب تقریباً صد از نظر تعداد مراحل آموزشی، یاد می‌گیرند. برای محیط‌های دیگر، تفاوت احتمالاً کوچک‌تر است؛ زیرا عملکرد رویکردهای کوانتومی در سطح عملکرد رویکرد کلاسیک نبود. انتظار می‌رود که با بالغ‌شدن سخت‌افزار کوانتومی، بتوانیم این آزمایش‌ها را بر روی سخت‌افزار کوانتومی با عملکرد مشابهی که در شبیه‌سازی‌ها یافت می‌شود، اجرا کنیم. در نتیجه، معتقدیم که شکاف بین رویکرد کلاسیک و رویکردهای کوانتومی می‌تواند برای محیط‌های پیچیده‌تر و سخت‌افزار کوانتومی بهبودیافته افزایش یابد.

همچنین تفاوت عملکرد بین یادگیری برنامه درسی و یادگیری مستقیم تجزیه‌وتحلیل شده است. در اولی، پیچیدگی محیط به‌تدریج افزایش می‌یابد، در دومی، محیط کامل به طور مستقیم ارائه می‌شود. دریافتیم که با یادگیری برنامه درسی، پاداش مورد انتظار بالاتر در همه موارد زودتر به دست می‌آید. رویکرد مبتنی بر بازپخت کوانتومی تفاوت‌هایی را در عملکرد بین محیط‌های مختلف نشان داد. همچنین تأثیر اقدامات تصادفی انجام شده توسط عوامل را در نظر گرفته شده که در آن عوامل فقط با احتمال کمی قدم موردنظر را برمی‌دارند و در غیر این صورت به یکی از حالات مجاور حرکت می‌کنند. مشاهده می‌شود که مدل‌های مورد مقایسه با کنش‌های تصادفی، تحت اعمال قطعی به همان اندازه خوب عمل می‌کنند. یادگیری برنامه درسی در هنگام درنظرگرفتن اقدامات تصادفی پتانسیل خود را نشان داد، زیرا رویکردها می‌توانستند قبل از حرکت به سمت اقدامات تصادفی پیچیده‌تر، یک خط‌مشی معقول را تحت اقدامات قطعی بیاموزند.

توجه داشته باشید که مقایسه عملکرد رویکردهای ما با کارهای قبلی، مانند (Crawford et al., 2016; Levit et al., 2017) دشوار است. در کار قبلی، برداشتن یک گام هزینه‌ای نداشت. بااین‌حال، هزینه اضافی برداشتن یک قدم را اضافه شده است، در نتیجه مسیرهای کوتاه‌تر را نسبت به مسیرهای طولانی‌تر ترجیح دادیم. در نتیجه، پاداش یافت شده در برخی از محیط‌ها کم به نظر می‌رسد، درحالی‌که خط‌مشی نزدیک به بهینه است. در بیشتر موارد، این پاداش کمتر ناشی از یک یا چند حالت با اقدامات نادرست اختصاص‌داده‌شده به آن است که باعث می‌شود عوامل در دایره‌ها سرگردان شوند. تغییرات در استراتژی یادگیری برنامه درسی نیز می‌تواند عملکرد مدل‌ها را با یادگیری سریع‌تر یک خط‌مشی بهینه افزایش دهد. در این مقاله، فقط پس از تعداد ثابتی از مراحل آموزشی، محیط پیچیده شده است. در برخی موارد، این امر منجر به یادگیری سریع سیاست و تقویت این سیاست بهینه شد. سیاستی که بدین ترتیب آموخته می‌شود می‌تواند در محیط‌های پیچیده‌تر نا بهینه یا حتی بد باشد. ثابت شد که به‌روزرسانی خط‌مشی برای برخی از محیط‌ها سخت است. یک استراتژی یادگیری برنامه درسی بهبودیافته، پیچیده کردن محیط پس از دستیابی به عملکرد معین است. بااین‌حال، تعیین آستانه عملکرد متناظر در زمانی که محیط را پیچیده می‌کند، نیازمند دانش بیشتری از محیط در نظر گرفته شده واقعی است. این دانش موردنیاز باهدف اصلی برای استفاده از مدل‌ها در تنظیمات عملیاتی با ورودی دستی کاربر محدود در تضاد است.

فهرست منابع

Agostinelli, F., McAleer, S., Shmakov, A., & Baldi, P. (2019). Solving the Rubik's cube with deep reinforcement learning and search. *Nature Machine Intelligence*, *1*(8), 356–363. https://doi.org/10.1038/s42256-019-0070-z

Arel, I., Liu, C., Urbanik, T., & Kohls, A. G. (2010). Reinforcement learning-based multi-agent system for network traffic signal control. *IET Intelligent Transport Systems*, *4*(2), 128–135. https://doi.org/10.1049/iet-its.2009.0070

Benedetti, M., Realpe-Gómez, J., & Perdomo-Ortiz, A. (2017). *Quantum-assisted Helmholtz machines: A quantum-classical deep learning framework for industrial datasets in near-term devices*. https://doi.org/10.1088/2058-9565/aabd98

Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, 41–48. https://doi.org/10.1145/1553374.1553380

Bhatia, H. S., & Phillipson, F. (2021). Performance Analysis of Support Vector Machine Implementations on the D-Wave Quantum Annealer. In M. Paszynski, D. Kranzlmüller, V. V Krzhizhanovskaya, J. J. Dongarra, & P. M. A. Sloot (Eds.), *Computational Science – ICCS 2021* (pp. 84–97). Springer International Publishing.

Crawford, D., Levit, A., Ghadermarzy, N., Oberoi, J. S., & Ronagh, P. (2016). *Reinforcement Learning Using Quantum Boltzmann Machines*. http://arxiv.org/abs/1612.05695

Dong, D., Chen, C., Li, H., & Tarn, T.-J. (2008a). *Quantum reinforcement learning*. https://doi.org/10.1109/TSMCB.2008.925743

Dong, D., Chen, C., Li, H., & Tarn, T.-J. (2008b). *Quantum reinforcement learning*. https://doi.org/10.1109/TSMCB.2008.925743

Dunjko, V., Taylor, J. M., & Briegel, H. J. (2017). Advances in quantum reinforcement learning. *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, *2017-January*, 282–287. https://doi.org/10.1109/SMC.2017.8122616

Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. *Symposium on the Theory of Computing*. https://api.semanticscholar.org/CorpusID:207198067

Hu, F., Wang, B., Wang, N., & Wang, C. (2019). Quantum Machine Learning with D-Wave Quantum Computer. *Quantum Engineering*, *1*. https://doi.org/10.1002/que2.12

Inci, A., Bolotin, E., Fu, Y., Dalal, G., Mannor, S., Nellans, D., & Marculescu, D. (2020). *The Architectural Implications of Distributed Reinforcement Learning on CPU-GPU Systems*. http://arxiv.org/abs/2012.04210

Jerbi, S., Trenkwalder, L. M., Nautrup, H. P., Briegel, H. J., & Dunjko, V. (2019). *Quantum enhancements for deep reinforcement learning in large spaces*. https://doi.org/10.1103/PRXQuantum.2.010328

Kadowaki, T., & Nishimori, H. (1998). *Quantum Annealing in the Transverse Ising Model Typeset using REVT E X 1*.

Koshka, Y. (2021). Quantum Machine Learning: Strategies Based on Quantum Annealing and Gated Quantum Computing. *2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 1. https://doi.org/10.1109/SPMB52430.2021.9672270

Lau, J. W. Z., Lim, K. H., Shrotriya, H., & Kwek, L. C. (2022). NISQ computing: where are we and where do we go? In *AAPPS Bulletin* (Vol. 32, Issue 1). Springer. https://doi.org/10.1007/s43673-022-00058-z

Levit, A., Crawford, D., Ghadermarzy, N., Oberoi, J. S., Zahedinejad, E., & Ronagh, P. (2017). *Free energy-based reinforcement learning using a quantum processor*. http://arxiv.org/abs/1706.00074

Li, R. Y., Di Felice, R., Rohs, R., & Lidar, D. A. (2018). *Quantum annealing versus classical machine learning applied to a simplified computational biology problem*. https://doi.org/10.1038/s41534-018-0060-8

Neukart, F., Compostella, G., Seidel, C., von Dollen, D., Yarkoni, S., & Parney, B. (2017). *Traffic flow optimization using a quantum annealer*. http://arxiv.org/abs/1708.01625

Neukart, F., Dollen, D. Von, & Seidel, C. (n.d.). *Quantum-assisted cluster analysis*.

Neumann, N. M. P., de Heer, P. B. U. L., Chiscop, I., & Phillipson, F. (2020). Multi-agent reinforcement learning using simulated quantum annealing. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12142 LNCS*, 562–575. https://doi.org/10.1007/978-3-030-50433-5_43

Neumann, N., Phillipson, F., & Versluis, R. (2019). Machine learning in the quantum era. *Digitale Welt*, *3*(2), 24–29. https://doi.org/10.1007/s42354-019-0164-0

Paparo, G. D., Dunjko, V., Makmal, A., Martin-Delgado, M. A., & Briegel, H. J. (2014). *Quantum speedup for active learning agents*. https://doi.org/10.1103/PhysRevX.4.031002

Phillipson, F. (2020). *Quantum Machine Learning: Benefits and Practical Examples*.

Preskill, J. (2018). *Quantum Computing in the NISQ era and beyond*. https://doi.org/10.22331/q-2018-08-06-79

Ritter, M. B. (2019). Near-term Quantum Algorithms for Quantum Many-body Systems. *Journal of Physics: Conference Series*, *1290*(1). https://doi.org/10.1088/1742-6596/1290/1/012003

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489. https://doi.org/10.1038/nature16961

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). *The Computational Limits of Deep Learning*. http://arxiv.org/abs/2007.05558

Wang, Z., Wei, S., Long, G. L., & Hanzo, L. (2022). Variational quantum attacks threaten advanced encryption standard based symmetric cryptography. *Science China Information Sciences*, *65*(10). https://doi.org/10.1007/s11432-022-3511-5

Wei, S., Chen, Y., Zhou, Z., & Long, G. (2021). *A Quantum Convolutional Neural Network on NISQ Devices*. http://arxiv.org/abs/2104.06918

Zhu, Y., Mottaghi, R., Kolve, E., Lim, J. J., Gupta, A., Fei-Fei, L., & Farhadi, A. (2016). *Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning*. http://arxiv.org/abs/1609.05143