دوفصل‌نامه علمی منادی امنیت فضای تولید و تبادل اطلاعات (افتا)
سال ۱۴۰۳، جلد ۲۲، شماره ۲ (۳۴-۴۴)
https://monadi.isc.org.ir

ارائه‌شده در بیست و یکمین کنفرانس بین‌المللی انجمن رمز ایران در دانشگاه تربیت مدرس و پژوهشگاه ارتباطات و فناوری اطلاعات، ۲۵ و ۲۶ مهرماه ۱۴۰۳

هوشمندسازی اجرای سناریوهای تمرین در میدان سایبری با استفاده از روش‌های یادگیری ماشین★

فرنوش کریمی[۱]، بهروز ترک لادانی[۱،*] و بهروز شاه‌قلی قهفرخی[۲]

[۱]گروه مهندسی نرم‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران
[۲]گروه فناوری اطلاعات، دانشکده مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان، ایران

اطلاعات مقاله

کلمات کلیدی:
یادگیری تطبیقی
آموزش امنیت سایبری
میدان سایبری کایپو
یادگیری ماشین
یادگیری تقویتی

نوع مقاله: پژوهشی

چکیده

با افزایش شدت تهدیدات امنیت سایبری در سطح جهانی، نیاز به آموزش متخصصان امنیتی اهمیت بیشتری یافته است. برنامه‌های آموزشی به همراه آزمایشگاه‌ها و انجام تمرین‌های امنیت سایبری، نقش اساسی در بهبود مهارت‌های آفندی و پدافندی ایفا می‌کنند. انجام این تمرین‌ها، به ویژه در شبکه‌های عملیاتی که مناسب آزمایش حملات سایبری نیستند، از اهمیت ویژه‌ای برخوردارند. میدان سایبری، بستر مناسبی برای این تمرین‌ها فراهم می‌کنند. یکی از چالش‌های اساسی در آموزش امنیت سایبری، تطابق برنامه‌های آموزشی با سطوح مختلف مهارت‌آموزان است. یادگیری تطبیقی با استفاده از هوش مصنوعی و سیستم‌های پیشنهاددهنده می‌تواند راه‌حل مناسبی برای ارائه آموزش شخصی‌سازی شده باشد. در این پژوهش، با تمرکز بر میدان سایبری کایپو، به بررسی امکان جایگزینی یا تکمیل نقش مربی با یک عامل پیشنهاد دهنده مبتنی بر هوش مصنوعی پرداخته شده است. هدف از این تحقیق، کاهش نیاز به دخالت انسانی و افزایش کارایی فرآیند آموزش است. بدین منظور، از اطلاعات جمع‌آوری‌شده در میدان سایبری کایپو که توسط دانشگاه ماساریک توسعه یافته، استفاده شده و مدل‌های مختلف یادگیری ماشین به کار گرفته شده است تا فرآیند آموزش به صورت خودکار و بهینه انجام شود. نتایج این پژوهش نشان می‌دهد که استفاده از هوش مصنوعی می‌تواند به بهبود عملکرد سیستم‌های آموزشی و کاهش زمان ارزیابی کمک کند.

© ۱۴۰۳ انجمن رمز ایران

۱ مقدمه

رخدادهای امنیتی[۱] در سراسر جهان نشان می‌دهند که شدت تهدیدات امنیت سایبری در حال افزایش است [۱]. اولین خط دفاعی در برابر این تهدیدات، افزایش آگاهی و مهارت‌های متخصصان امنیتی از جدیدترین تکنیک‌ها و ابزارهای تهدید است. به همین منظور، برنامه‌های آموزشی مانند آزمایشگاه‌ها[۲] و تمرین‌های امنیت سایبری[۳] برای آمادگی متخصصان تعریف می‌شود. این تمرین‌ها با هدف بهبود مهارت‌های آفندی[۴] و پدافندی[۵] در محیط‌های مجازی یا فیزیکی اجرا می‌شوند [۲]. انجام تمرین‌های امنیت سایبری دو هدف کلی دارند [۲]:

برای متخصصان امنیت سایبری: افزایش سطح مهارت در دفاع در برابر رخدادهای امنیتی

★از کمیته علمی بیست و یکمین کنفرانس بین‌المللی انجمن رمز ایران برای داوری این مقاله تشکر می‌شود.
*نویسنده مسئول.
آدرس‌های رایانامه: farnoosh.karimi95@eng.ui.ac.ir (فرنوش کریمی)، ladani@eng.ui.ac.ir (بهروز ترک لادانی)، shahgholi@eng.ui.ac.ir (بهروز شاه‌قلی قهفرخی)
© ۱۴۰۳ تمامی حقوق متعلق به انجمن رمز ایران است.

[1]Security incidents
[2]Labs
[3]Cyber Security Exercises
[4]Attack Scenarios
[5]Defense

برای متخصصان غیرامنیتی و عموم مردم: افزایش آگاهی امنیت سایبری آن‌ها

از آنجایی که شبکه‌های عملیاتی برای آزمایش حملات سایبری مناسب نیستند، به بسترهای آزمایشی[1] مانند میدان سایبری[2] نیاز است تا مهارت‌آموزان بتوانند مهارت‌هایی مانند نظارت بر شبکه[3]، تشخیص و کاهش حملات[4] و تست نفوذ[5] را در یک محیط شبه واقعی تمرین کنند [۲، ۳]. از جمله میدان‌های سایبری موجود، می‌توان به کایپو[6] و آلپاکا[7] اشاره کرد. میدان سایبری کایپو که برای آموزش تیم‌های امنیتی دانشجویان و مهارت‌آموزان طراحی شده است، مبنای محصول میدان سایبری نقش مجاز است که توسط پژوهشکده آپا دانشگاه اصفهان، توسعه داده شده است.

یکی از چالش‌های آموزش امنیت سایبری، انطباق برنامه‌های آموزشی با مهارت‌های مختلف مهارت‌آموزان است. برنامه‌های آموزشی ثابت ممکن است برای برخی بسیار آسان و برای برخی دیگر بسیار دشوار باشد [۴، ۵]. یادگیری تطبیقی می‌تواند راهی برای ارائه مطالب آموزشی به صورت شخصی‌سازی شده باشد که در آن، برنامه‌های آموزشی به صورت خودکار و پویا باتوجه به عملکرد مهارت‌آموز در مراحل قبلی تطبیق می‌یابند [۶، ۷]. بدین منظور مربی می‌تواند در یک محیط شبیه‌سازی شده، پاسخ آزمون‌های مهارت‌آموزان را جمع‌آوری کند و مناسب‌ترین وظیفه را به هر مهارت‌آموز اختصاص دهد اما اگر تعداد مهارت‌آموزان زیاد باشد، این روش بهینه نخواهد بود [۴].

برای حل این مشکل، می‌توان از هوش مصنوعی استفاده کرد تا فرآیند شخصی‌سازی آموزش به طور خودکار انجام شود. سیستم‌های پیشنهاد دهنده مبتنی بر هوش مصنوعی می‌توانند وظایف و محتوای آموزشی مناسب را برای هر فرد تعیین کنند، عملکرد سیستم را بهبود بخشند و زمان ارزیابی و نمره‌دهی را کاهش دهند [۸].

در این پژوهش، قصد داریم نقش مربی را در میدان سایبری نقش مجاز با یک عامل پیشنهاد دهنده مبتنی بر هوش مصنوعی جایگزین کنیم. هدف ما کاهش نیاز به دخالت انسانی و افزایش کارایی و سرعت فرآیند آموزش است. این پژوهش بر روی اطلاعات جمع‌آوری‌شده از میدان سایبری کایپو که توسط دانشگاه ماساریک[8] توسعه یافته است، انجام خواهد شد.

۱.۱ ساختار مقاله

در ادامه مقاله، در بخش ۲ به بیان مفاهیم اولیه پرداخته می‌شود. در بخش ۳، مطالعات پیشین مورد بررسی قرار می‌گیرد. در بخش ۴ طریقه جمع‌آوری داده و ارائه راه‌کار پیشنهادی بیان می‌شود. بخش ۵ کارهای انجام‌شده، ارزیابی و نتایج با یکدیگر مقایسه می‌شوند. در بخش ۶ نیز نتایج کلی بیان می‌شود.

[1] Testbeds [2] Cyber Ranges (CR) [3] Network monitoring [4] Attack detection and mitigation [5] Penetration Testing [6] KYPO [7] Alpaca [8] Masaryk University

۲ مفاهیم اولیه

در این بخش، مفاهیم اولیه مورد نیاز که در این مقاله استفاده می‌شود، به‌اختصار بررسی می‌گردد.

۱.۲ میدان سایبری

میدان سایبری یک محیط آزمایش سایبری مدیریت‌شده و مجازی است که می‌تواند شبکه‌های محلی، سیستم‌ها، ابزارها و برنامه‌ها را شبیه‌سازی کند و به اینترنت شبیه‌سازی شده متصل شود. این محیط برای افزایش مهارت‌های امنیت سایبری و توسعه محصولات ایمن طراحی شده و می‌تواند شامل سخت‌افزار و نرم‌افزار واقعی یا مجازی باشد. تیم‌ها در این محیط می‌توانند با هم همکاری کنند، ایده‌های جدید را آزمایش کنند و مشکلات پیچیده سایبری را حل کنند. انواع مختلف میدان‌های سایبری وجود دارد که بر اساس تنوع و قابلیت، برای آموزش امنیت سایبری استفاده می‌شود. بیشترین استفاده از میدان‌های سایبری در حوزه آموزش و سپس در سازمان‌های تجاری و دفاع نظامی است [۹].

یکی از این میدان‌های سایبری، میدان سایبری کایپو است که توسط تیم امنیتی دانشگاه ماساریک جمهوری چک توسعه یافته است. میدان سایبری کایپو به‌دلیل قابلیت‌های برجسته‌اش، از میان سایر میدان‌های سایبری، برای این پژوهش انتخاب شده است که می‌توان به این موارد اشاره کرد [۳، ۴]:

(۱) منبع باز بودن
(۲) ارائه رابط کاربری گرافیکی برای دسترسی به آموزش‌ها
(۳) پشتیبانی از استفاده چندگانه مانند تحقیق و آموزش
(۴) ارائه محیطی مجازی برای انجام حملات سایبری پیچیده و آموزش تیم‌های امنیتی، دانشجویان و مهارت‌آموزان
(۵) استفاده از فضای ابری OpenStack
(۶) گستردگی سناریوها و امکانات مختلف که می‌توانند برای اهداف مختلف مدل شوند

۲.۲ یادگیری تطبیقی

یادگیری تطبیقی یک فرآیند شخصی‌سازی شده است و تعامل یادگیرنده با محتوای قبلی، ماهیت مطالب در مرحله بعد را تعیین می‌کند و هدف آن این است که یادگیری شخصی را برای مهارت‌آموزان ایجاد کند. هم‌چنین می‌تواند برای مهارت‌آموزان با پیشینه‌های مختلف استفاده شود، عملکرد مهارت‌آموزان را بررسی کرده و جهت پیشرفت آن‌ها، برنامه‌ای متناسب پیشنهاد دهد [۱۰]. برای مثال، مرحله دوم می‌تواند شامل سه وظیفه باشد و اگر عملکرد مهارت‌آموز در مرحله اول خوب باشد، وظیفه یک و درصورتی‌که عملکرد مهارت‌آموز ضعیف باشد، وظیفه سه پیشنهاد می‌شود.

۳.۲ رخدادهای امنیتی

رخدادهای امنیتی شامل حملات و اقداماتی هستند که منجر به از کار انداختن و دسترسی غیرمجاز به سرویس‌های یک سازمان می‌شوند که می‌تواند خرابی‌ها و اختلالات زیادی در سامانه‌ها ایجاد کند. این رخدادها می‌توانند از حملات به سازمان‌های کوچک تا حملات به زیرساخت‌های حیاتی متغیر باشد. باتوجه به ماهرتر شدن مهاجمان و گسترش اینترنت، تعداد و پیچیدگی رخدادهای امنیتی نیز افزایش یافته است. بنابراین امنیت سیستم‌های اطلاعاتی دیگر محدود به استفاده سنتی از دیواره آتش [۱] یا سیستم‌های تشخیص نفوذ [۲] نیست. این امر، نیاز به آموزش متخصصان را برجسته کرده است [۱۱، ۱۲].

۴.۲ یادگیری ماشین

هوش مصنوعی در تلاش است که رفتار انسان‌ها و ماشین‌ها را به صورت مصنوعی در رایانه‌ها شبیه‌سازی کند تا بتواند عملکرد نزدیک به انسان داشته باشد. روش‌های مختلف یادگیری ماشین [۳]، شامل یادگیری عمیق [۴] و یادگیری تقویتی [۵] برای این منظور استفاده می‌شود [۱۳]. در ادامه، تعدادی از روش‌های استفاده شده در مقاله شرح داده می‌شود.

K نزدیک‌ترین همسایه [۶]: یکی از معروف‌ترین الگوریتم‌های یادگیری ماشین است که برای تشخیص الگو استفاده می‌شود. در این الگوریتم، داده‌ها به خوشه‌هایی تقسیم می‌شوند و K نزدیک‌ترین نمونه به نمونه جدید انتخاب می‌شود و برچسب آن‌ها برای پیش‌بینی برچسب نمونه جدید استفاده می‌شود [۱۴].

شبکه عصبی پیچشی [۷]: از جمله رایج‌ترین الگوریتم مورد استفاده در یادگیری عمیق است. مزیت اصلی CNN در مقایسه با پیشینیان خود این است که به طور خودکار ویژگی‌های مربوطه را بدون هیچ نظارت انسانی شناسایی می‌کند. این شبکه، از تعداد پارامترهای کمی استفاده می‌کند که فرآیند آموزش را ساده‌تر می‌کند و سرعت شبکه را افزایش می‌دهد. شبکه عصبی پیچشی، مانند سایر شبکه‌های عصبی، از لایه‌های نورونی با وزن و بایاس با قابلیت یادگیری تشکیل شده است [۱۵].

شبکه عصبی بازگشتی [۸]: یک الگوریتم معروف در زمینه یادگیری عمیق است که برای داده‌های ترتیبی مناسب است و می‌تواند به عنوان یک واحد حافظه کوتاه مدت در نظر گرفته شود. این شبکه، داده‌های متوالی را مدیریت می‌کند و از ساختار تعبیه شده در توالی‌های داده استفاده می‌کند تا اطلاعات زمانی مهم را استخراج کند. این شبکه در هر مرحله از پیمایش، ورودی فعلی و حالت پنهان قبلی را به‌عنوان ورودی می‌گیرد و خروجی فعلی و حالت پنهان بعدی را به‌عنوان خروجی تولید می‌کند [۱۵].

[1] Firewall [2] Intrusion detection system [3] Machine Learning (ML) [4] Deep Learning (DL) [5] Reinforcement Learning (RL) [6] K-Nearest Neighbors (KNN) [7] Convolutional Neural Network (CNN) [8] Recurrent Neural Networks (RNN)

شبکه عصبی حافظه کوتاه‌مدت طولانی [۹]: نوعی خاصی از شبکه عصبی بازگشتی است که مشکل حافظه بلندمدت شبکه RNN را حل می‌کند. این شبکه، سازوکارهایی به‌نام گیت [۱۰] دارد. که جریان اطلاعات را کنترل می‌کند و مشخص می‌کند چه داده‌هایی در توالی مهم هستند و باید حفظ شوند و چه داده‌هایی باید حذف شوند. در نتیجه شبکه اطلاعات مهم را در طول زنجیره توالی عبور می‌دهد تا خروجی مدنظر تولید شود [۱۵].

یادگیری Q [۱۱]: یکی از محبوب‌ترین الگوریتم‌های یادگیری تقویتی است که با تخمین مقدار حالت–اقدام یک تابع Q کار می‌کند. در این الگوریتم، عامل به طور تکراری با محیط تعامل دارد و با انجام اقدامات در محیط، تجربه‌های جدیدی کسب می‌کند و با کمک آن‌ها، تلاش می‌کند بهترین راه‌حل را برای رسیدن به هدف پیدا کند. این الگوریتم بهترین اقدامات برای هر حالت را از محیط یاد می‌گیرد و در جدول Q ثبت می‌کند. هر سلول از این جدول، نشان دهنده حالت فعلی و اقدامات قابل انجام در آن حالت است. همچنین مقدار آن، نشان‌دهنده امتیاز کسب شده از انجام آن اقدام در آن حالت می‌باشد. با کمک رابطه (۱) می‌توان جدول Q را به‌روزرسانی کرد. در این رابطه α نرخ یادگیری و γ فاکتور تخفیف را نشان می‌دهد که هر دو در بازه $[0, 1]$ قرار می‌گیرند [۱۶].

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max Q(s',a') - Q(s,a)] \quad (۱)$$

یادگیری تقویتی عمیق [۱۲]: یکی از الگوریتم‌های یادگیری تقویتی است که یادگیری عمیق را با یادگیری تقویتی ترکیب می‌کند. در DQN، از ساختار مخزن تجربه برای برهم زدن نظم نمونه استفاده می‌شود و با توجه به تجربه به‌دست آمده، مشکل را حل می‌کند. DQN قادر به یادگیری سیاست‌های پیچیده در محیط‌هایی با فضای حالت بزرگ و پیچیده است. این روش از Replay Buffer استفاده می‌کند تا مشکلات یادگیری ناپایدار و همبستگی داده‌ها را بهبود بخشد و به نتایج قابل‌توجهی دست یابد. این الگوریتم بیش‌تر برای حل مسائل کنترل بهره‌وری و تصمیم‌گیری در محیط‌های پویا و پیچیده استفاده می‌شود [۱۷].

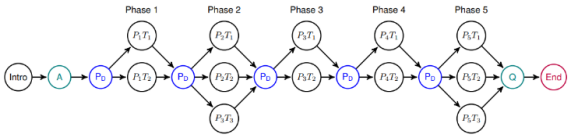
۳ مطالعات پیشین

در این بخش، مطالعات مرتبط با آموزش تطبیقی در سامانه‌های آموزشی مختلف و کار قبلی انجام شده در میدان سایبری کایپو، مرور و مورد بحث قرار می‌گیرند.

۱.۳ آموزش تطبیقی در میدان سایبری کایپو

برای تعریف آموزش تطبیقی در میدان سایبری کایپو، ویکوپال [۱۳] و همکاران [۴]، مدلی به نام مدل مربی را معرفی کرده‌اند. آزمایشگاه هوشمند پیشنهادی در این پژوهش با نام KYPO SLE معرفی شده است که در میدان سایبری کایپو توسعه یافته است. در این مدل، آموزش

[9] Long-Short Term Memory (LSTM) [10] Gate [11] Q-learning [12] Deep Q-Network (DQN) [13] Jan Vykopal



شکل ۱. ساختار نمودار آموزش امنیت سایبری تطبیقی [۴]

می‌تواند از تعدادی مرحله و وظیفه [1] تشکیل شود. در هر یک از مراحل، مهارت‌های یکسانی انجام می‌شود که سختی متفاوتی دارند. همانند شکل ۱، مولفه‌های اصلی این آموزش به‌شرح زیر می‌باشد:

مقدمه: اطلاعات لازم در مورد آموزش را در اختیار مهارت‌آموز قرار می‌دهد.

ارزیابی قبل از آموزش (A): دانش ابتدایی مهارت‌آموزان توسط پرسش‌نامه‌ای سنجیده می‌شود.

مراحل آموزشی (P_x): شامل وظایف مختلف (T_y) است که سختی متفاوتی دارند. برای مثال، T_1 دشوارترین کار و T_n ساده‌ترین کار در همان مرحله می‌باشد.

اجزای تصمیم‌گیری (P_D): عملکرد و دانش مهارت‌آموزان را پردازش می‌کند تا وظیفه مناسب در مراحل بعدی را تعیین کند.

پرسش‌نامه پس از آموزش (Q): این پرسش‌نامه جهت جمع‌آوری بازخورد از مهارت‌آموزان و کمک به مربیان می‌باشد.

این مدل با سه معادله تعریف می‌شود. در این معادلات، متغیر x نشان دهنده هر مرحله، T_y نشان دهنده مناسب‌ترین وظیفه در مرحله x، p نشان دهنده درستی یا نادرستی پاسخ‌ها، m تعداد مراحل آموزشی، k نشان دهنده استفاده مهارت‌آموز از دستورات کلیدی مورد انتظار در خط فرمان، a مشخص کننده پاسخ‌های مورد انتظار به هر وظیفه، t نشان دهنده انجام کار در زمان تعیین‌شده و s نشان دهنده درخواست مهارت‌آموز برای افشای پاسخ تکلیف می‌باشد. همه این متغیرها، به صورت صفر یا یک مقداردهی می‌شوند. برای مثال اگر در زمان تعیین‌شده تکلیف به اتمام برسد، یک و در صورت عدم اتمام، صفر مقداردهی می‌شود.

در مدل پیشنهادی، محتوای آموزشی باید به طور کامل طراحی و آزمایش شود و پارامترهای مدل، تعیین وزن‌ها، محتوای وظایف و روابط بین مراحل توسط مربیان تنظیم شود. انتخاب نادرست وزن‌ها، منجر به اشتباه در مسیر یادگیری می‌شود. شرکت‌کنندگان گزارش داده‌اند که وظایف به خوبی طراحی شده و آموزش را با موفقیت به‌پایان رسانده‌اند.

[1] Tasks

۲.۳ سیستم یادگیری تطبیقی با استفاده از رویکرد توصیه‌ای و یادگیری تقویتی

مدنی [2] و همکاران [۱۸]، سیستم یادگیری الکترونیکی تطبیقی با استفاده از رویکرد توصیه‌ای و یادگیری تقویتی ارائه کرده‌اند. در این سیستم، مهارت‌آموزان وارد حساب کاربری خود می‌شوند. سپس به پیش‌آزمونی پاسخ می‌دهند که نتایج آن برای توصیه مسیر یادگیری استفاده می‌شود. اگر مهارت‌آموزی با شرایط مشابه پیدا شود، می‌توان مسیر یادگیری که به مهارت‌آموز قبلی داده شده را، به مهارت‌آموز فعلی نیز پیشنهاد داد و اگر مهارت‌آموز مشابهی یافت نشود، از رویکرد یادگیری تقویتی برای پیدا کردن مسیر بهینه استفاده می‌شود. این سیستم برای آموزش زبان جاوا به‌کار رفته بود که درنهایت نمرات، عملکرد تحصیلی و رضایت اکثر مهارت‌آموزان افزایش یافته و میانگین زمان پاسخ آن‌ها کم‌تر شده بود.

۳.۳ یادگیری شخصی‌سازی شده مبتنی بر هوش مصنوعی

سید [3] و همکاران [۱۹]، سکوی APPEAL را برای یادگیری الکترونیکی شخصی‌سازی شده ارائه داده‌اند که مسیر یادگیری شخصی را با استفاده از RL بر اساس سبک یادگیری، تلاش و عملکرد مشخص می‌کند. در این سیستم، مهارت‌آموزان به دو گروه تقسیم شدند: گروه اول از روش VARK (تعیین سبک یادگیری که شامل دیداری، شنیداری، متنی و حرکتی است) و گروه دوم از بازی‌وارسازی استفاده کردند. مهارت‌آموزان پس از تعیین سبک یادگیری خود، پیش‌آزمونی را گذراندند. سپس بر اساس سطح خود و با کمک سیستم RL، تمرینات ریاضی با سه سطح دشواری در اختیار مهارت‌آموزان هر دو گروه قرار گرفت. نتایج نشان داد که گروه VARK نسبت به بازی‌وارسازی عملکرد بهتری داشت و نمرات و رضایت مهارت‌آموزان بهبود یافت.

۴.۳ سیستم یادگیری الکترونیکی با استفاده از یادگیری ماشین

لو [4] و همکاران [۲۰]، یک سیستم آموزش الکترونیکی برای شخصی‌سازی محتوای یادگیری انگلیسی (خواندن و شنیدن) بر اساس احساسات و وضعیت عاطفی مهارت‌آموزان ارائه داده‌اند. این سیستم با استفاده از دستگاه EEG بی‌سیم، سیگنال‌های مغزی را جمع‌آوری کرده و حالات عاطفی را تشخیص می‌دهد و با کمک الگوریتم k نزدیک‌ترین همسایه، مهارت‌آموزان به سه دسته منفی، مثبت و خنثی تقسیم می‌شوند. سپس از یادگیری تقویتی برای شخصی‌سازی مواد آموزشی استفاده می‌شود و میزان پاداش، با توجه به وضعیت عاطفی مهارت‌آموزان که توسط اقدامات به‌روز می‌شوند، محاسبه می‌گردد. این پژوهش بر روی ۳۰ مهارت‌آموز انجام شد و نشان داد که رضایت مهارت‌آموزان از سیستم پیشنهادی بالا بود.

[2] Youness Madani [3] Wafaa S. Sayed [4] Xinyang Liu

۵.۳ سیستم یادگیری تطبیقی مبتنی بر یادگیری تقویتی

شوقی[1] و همکاران [۲۱]، روشی برای توصیه مواد آموزشی و پیگیری وضعیت مهارت‌آموزان با استفاده از یادگیری تقویتی ارائه کرده‌اند. این روش شامل سه مرحله است: تعیین حالت فعلی مهارت‌آموز، پیشنهاد مسیر یادگیری و به‌روزرسانی حالت و پاداش‌ها. سیستم ابتدا ماتریس حالت–اقدام را بارگیری می‌کند، سپس حالت فعلی مهارت‌آموز را مشخص می‌کند و اقدام مناسب را پیشنهاد می‌دهد. در نهایت پاداش عمل و حالت جدید را تعیین می‌کند. برای ارزیابی، شبیه‌سازی با ۲۰ حالت و ۲۰ اقدام انجام شد که نتایج نشان داد با افزایش تکرارها، تعداد پاداش‌های مثبت بیش‌تر می‌شود و سیستم بهترین محیط یادگیری را پیشنهاد می‌کند.

۶.۳ سیستم یادگیری الکترونیکی تطبیقی مبتنی بر الگوریتم Q-learning

بوساکسو[2] و همکاران [۲۲]، رویکردی برای دوره‌های یادگیری آنلاین تطبیقی بر اساس فعالیت‌ها و اهداف یادگیری مهارت‌آموزان ارائه کرده‌اند. در این رویکرد، تمرین‌های عملی در اختیار مهارت‌آموزان قرار گرفت که این تمرین‌ها سطح دشواری متفاوتی داشتند. سپس از الگوریتم Q-learning جهت تعیین مسیر یادگیری مهارت‌آموزان استفاده شد که شامل نه حالت و پنج اقدام بود. پاداش هر عمل نیز به صورت متفاوتی محاسبه شد. این الگوریتم، بهترین مسیر را برای یادگیری یک دوره مشخص تعیین می‌کند و مزیت اصلی آن، همگرا شدن سریع بعد از تعدادی تکرار می‌باشد.

۴ روش‌شناسی

در این بخش، طریقه جمع‌آوری داده و روش‌های استفاده شده، مورد بررسی قرار می‌گیرند.

۱.۴ جمع‌آوری داده

همان‌طور که در کار ویکوپال و همکاران [۴] اشاره شد، پارامترهای مدل مربی ابتدا توسط مربیان بر اساس تخصص آن‌ها، محتوای وظایف و روابط بین مراحل، به صورت دستی تنظیم می‌شوند سپس با کمک روابط (۲) تا (۴) مشخص می‌شود که مهارت‌آموز در مرحله بعدی باید چه وظیفه‌ای را انجام دهد. ویکوپال و همکاران، نتایج کار خود را در مجموعه داده‌ای منتشر کرده‌اند [۲۳]. این مجموعه داده شامل نمرات مهارت‌آموزان در پنج مرحله می‌باشد که هر مرحله از سه وظیفه سخت (T_1)، متوسط (T_2) و آسان (T_3) تشکیل شده است. نمرات مهارت‌آموزان به‌شرح زیر ذخیره می‌شوند:

P: این نمره بیانگر پاسخ سوال است که اگر مهارت‌آموز پاسخ صحیح به سوال بدهد نمره ۱ و اگر پاسخ اشتباه به سوال بدهد نمره ۰ ثبت می‌شود.

[1] Doaa Shawky [2] Mohamed Boussakssou

	A	B	C	D	E	F	G	H	I	J	K
1	phase	participants	task_variant	trunc	f(x)	time_in_min	p1	k1	s1	t1	a1
2	1	1	1	1	1	1	1	1	1	0	0
3	1	2	1	1	1	1	1	0	1	1	1
4	1	3	1	1	1	1	1	0	1	1	1
5	1	4	3	4	0	1	0	1	0	0	1
6	1	5	1	1	1	1	1	1	1	0	1
7	1	6	3	4	0	1	0	1	0	0	0
8	1	7	1	1	1	1	1	0	1	1	0
9	1	8	1	1	1	1	1	0	1	1	1
10	1	9	1	1	1	1	1	0	1	1	0
11	1	10	1	1	1	1	1	1	0	0	1
12	1	11	1	1	1	1	1	1	1	1	1
13	1	12	3	4	0	1	0	1	1	0	0
14	1	13	3	4	0	1	0	0	1	1	1
15	1	14	1	1	1	1	1	0	1	0	0
16	1	15	3	4	0	1	0	0	1	1	1
17	1	16	3	4	0	1	0	1	1	1	1
18	1	17	3	4	0	1	0	1	1	0	1
19	1	18	1	1	1	1	1	0	1	1	1
20	1	19	3	4	0	1	0	0	1	1	1
21	1	20	3	4	0	1	0	1	1	1	1
22	1	21	1	1	1	1	1	1	1	1	1
23	1	22	1	1	1	1	1	1	1	1	1
24	1	23	1	1	1	1	1	0	1	1	1
25	1	24	1	1	1	1	1	0	1	1	0
26	1	25	1	1	1	1	1	0	0	1	0
27	1	26	1	1	1	1	1	1	1	0	0
28	1	27	1	1	1	1	1	1	0	1	1
29	1	28	1	1	1	1	1	0	1	0	1

شکل ۲. بخشی از نمای مجموعه‌داده سنتزشده

K: این نمره بیانگر تعداد فرمان‌های صحیح واردشده است که اگر تعداد فرمان‌های صحیح بیش‌تر از ۵ باشد، نمره ۱ و اگر کم‌تر از ۵ باشد، نمره ۰ لحاظ می‌شود.

S: این نمره بیانگر درخواست مهارت‌آموز برای نمایش پاسخ سوال است. اگر مهارت‌آموز درخواست نمایش پاسخ سوال را داده باشد، نمره ۰ و اگر درخواست نمایش پاسخ را نداده باشد نمره ۱ اعمال می‌شود.

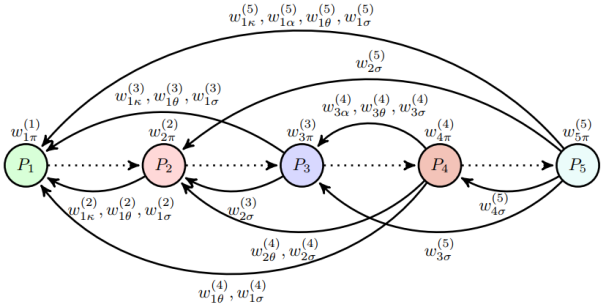
T: این نمره بیانگر این است که مهارت‌آموز در مدت زمان تعیین‌شده پاسخ سوال را ثبت کرده است یا خیر. اگر در مدت زمان تعیین‌شده پاسخ سوال را ثبت کرده باشد، نمره ۱ و در غیر این‌صورت نمره ۰ لحاظ می‌گردد.

A: این نمره بیانگر تعداد کل پاسخ‌های ثبت‌شده است که اگر تعداد پاسخ‌ها بیش‌تر از ۱۰ باشد، نمره ۱ و اگر کم‌تر از ۱۰ باشد نمره ۰ ثبت می‌شود.

Time in min: بیانگر زمانی است که مهارت‌آموز برای انجام تمرین سپری کرده است. این زمان به دقیقه ثبت می‌شود.

باتوجه به این‌که این پژوهش بر روی ۲۴ مهارت‌آموز بررسی شد، برای ادامه کار با مشکل کمبود داده مواجه شدیم. لذا مجموعه داده‌ای شامل ۱۰۰۰۰ سطر با داده‌های سنتزشده ایجاد کردیم. در ابتدا با کمک ابزار اکسل، نمراتی که پیش‌تر مطرح شد، به صورت تصادفی در پنج ستون ثبت شد و با کمک روابط (۲) تا (۴) که از پژوهش آموزش امنیت سایبری تطبیقی در کایپو [۴] اقتباس شده است، وظیفه پیشنهادشده به هر مهارت‌آموز محاسبه گردید (مانند شکل ۲). سپس از این مجموعه داده برای انجام پژوهش استفاده شد.

رابطه (۲) برای ماتریس تصمیم است که وزن مراحل به صورت دستی توسط مربی تنظیم می‌شوند:



شکل ۳. روابط بین تمام مراحل آموزش [۴]

$$W^{(x)} = (\omega_{ij}^{(x)}), \quad i = ۱, \ldots, m, \; j = \pi, k, \alpha, \theta, \sigma \quad (۲)$$

رابطه (۳) عملکرد مهارت‌آموز بر اساس معیارهای تعریف‌شده و وزن آن‌ها را محاسبه می‌کند که مقدار این عملکرد در بازه [۰, ۱] می‌باشد.

$$f(x) = \frac{\sum_{i=۱}^{x}\left[p_i\omega_{i\pi}^{(x)} + s_i\left(k_i\omega_{ik}^{(x)} + a_i\omega_{i\alpha}^{(x)} + t_i\omega_{i\theta}^{(x)} + \omega_{i\sigma}^{(x)}\right)\right]}{\sum_{i=۱}^{x}\left(\omega_{i\pi}^{(x)} + \omega_{ik}^{(x)} + \omega_{i\alpha}^{(x)} + \omega_{i\theta}^{(x)} + \omega_{i\sigma}^{(x)}\right)} \quad (۳)$$

رابطه (۴) نیز، مناسب‌ترین وظیفه y در مرحله x را برای یک مهارت‌آموز خاص به دست می‌آورد $(T_۱, T_۲, T_۳)$.

$$T_y = \begin{cases} n_x & \text{if } f(x) = ۰ \\ trunc(n_x[۱ - f(x)]) + ۱ & \text{otherwise} \end{cases} \quad (۴)$$

شکل ۳، روابط بین مراحل آموزشی و وزن مراحل مختلف را نشان می‌دهد. برای مثال اگر در مرحله سوم قرار داشته باشیم، وزنی که برای نمرات مراحل مختلف اعمال می‌شود، به‌شرح زیر می‌باشد:

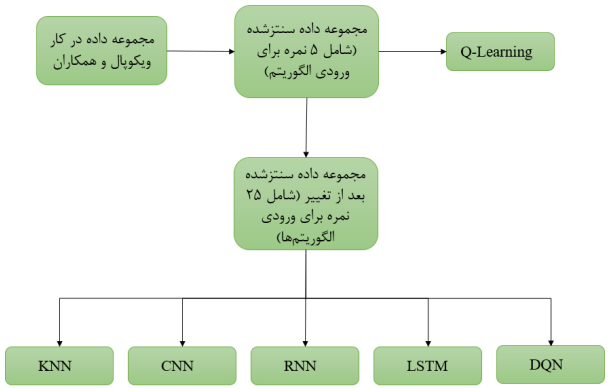
- مرحله اول: نمرات K، T و S با وزن ۱
- مرحله دوم: نمره S با وزن ۱
- مرحله سوم: نمره P با وزن ۱

وزن بقیه نمرات در این مرحله، صفر در نظر گرفته می‌شود.

در قدم بعدی، نمرات مهارت‌آموز به همراه وزن‌های تعیین‌شده، در رابطه (۳) و (۴) قرار می‌گیرد و مشخص می‌شود که در مرحله چهارم، چه وظیفه‌ای باید انجام شود.

۲.۴ روش‌های استفاده شده

برای پیاده‌سازی یادگیری تطبیقی، سیستم باید با توجه به پیشرفت مهارت‌آموز، گزینه‌های مناسبی را پیشنهاد دهد. در این بخش، از شش روش یادگیری ماشین برای پیشنهاد وظیفه مناسب هر مهارت‌آموز متناسب با نمرات کسب‌شده در مراحل قبلی، استفاده شده است که در آن، هوش مصنوعی تصمیم می‌گیرد که فرآیند به چه نحوی پیش برود. خلاصه روش‌های استفاده شده در نمودار شکل ۴، نشان داده شده است.



شکل ۴. نمودار خلاصه روش‌های استفاده شده

۱) Q-Learning برای استفاده از این الگوریتم، ابتدا سطح هر مهارت آموز با کمک رابطه (۵) مشخص می‌شود. این رابطه به صورت تجربی به دست آمده است و در هر مرحله اجرای برنامه، اعداد و ورودی‌های برنامه تغییر کرده است و بهترین نتیجه به‌عنوان رابطه نهایی استفاده شده است. این رابطه مهارت‌آموز را در چهار سطح تقسیم‌بندی می‌کند.

$$G(x) = \frac{(P_x + K_x + T_x)}{\text{time in min}_x} \quad (۵)$$

$$\text{student level} = \begin{cases} (G(x) * ۰/۷۵) + ۱ & \text{if } S_x = ۰ \\ (G(x) * ۱) + ۱ & \text{if } S_x = ۱ \end{cases}$$

در ابتدای استفاده از Q-learning، همه مراحل ارزش یکسان دارند سپس در گذر زمان، سیستم یاد می‌گیرد که کدام وظیفه را اولویت قرار دهد و ارزش بیشتری برای آن قائل شود. همچنین نیاز به پیاده‌سازی تابع پاداش می‌باشد تا اگر سیستم به خوبی عمل کرد، پاداش بگیرید. از آن جایی که آموزش در پنج مرحله انجام می‌شود و چهار گزینه برای سطح هر مهارت‌آموز تعریف شده است، بدین منظور ۲۰ حالت تعریف می‌شود. برای مثال حالت اول، مرحله ۱ و سطح مهارت‌آموز ۱ می‌باشد. حالت دوم، مرحله ۱ و سطح مهارت‌آموز ۲ می‌باشد. حالت سوم، مرحله ۱ و سطح مهارت‌آموز ۳ می‌باشد. حالت چهارم، مرحله ۱ و سطح مهارت‌آموز ۴ می‌باشد.

هر مرحله دارای سه وظیفه در سه دشواری مختلف می‌باشد. بنابراین این الگوریتم شامل سه اقدام می‌باشد. در هر یک از حالت‌ها، امکان انجام سه وظیفه می‌باشد. وظیفه ۱ به‌عنوان سخت‌ترین و وظیفه ۳ به‌عنوان ساده‌ترین وظیفه تعریف شده است.

بنابراین جدول Q تعریف شده همانند جدول ۱ است که مقادیر اولیه این جدول صفر است و در هر بار اجرای برنامه، اطلاعات از مجموعه داده خوانده شده و این جدول را با کمک رابطه (۱)، به‌روز می‌کند. پس از همگرا شدن این الگوریتم، برای مهارت‌آموزان جدید، حالت فعلی مهارت‌آموز سنجیده می‌شود و وظیفه‌ای انتخاب می‌شود که بیشترین مقدار را در جدول داشته باشد.

جدول ۱. بخشی از جدول Q تعریف‌شده در ابتدای کار

اقدامات			حالت‌ها	سطح مهارت‌آموز	مرحله آموزش
وظیفه ۱	وظیفه ۲	وظیفه ۳			
۰	۰	۰	۰	۱	۱
۰	۰	۰	۱	۲	۱
۰	۰	۰	...	...	...
۰	۰	۰	۱۸	۳	۵
۰	۰	۰	۱۹	۴	۵

برای انجام سه روش دیگر، مجموعه داده تغییر داده شد. به نحوی که در هر مرحله، تمام نمرات کسب شده مهارت‌آموز قابل مشاهده است و مراحلی که هنوز نگذرانده است، نمره صفر لحاظ می‌شود. در نتیجه ورودی الگوریتم، شامل ۲۵ نمره می‌باشد. برای مثال اگر مهارت‌آموز در مرحله سوم باشد، سه مرحله اول دارای نمرات کسب شده است و دو مرحله‌ای که هنوز نگذرانده است، نمره صفر ثبت می‌شود.

۲) **K نزدیک‌ترین همسایه** در این بخش از پژوهش، از الگوریتم K نزدیک‌ترین همسایه برای پیش‌بینی مناسب‌ترین وظیفه استفاده شده است. در این روش، تخصیص وظیفه به مهارت‌آموز جدید، بر اساس نزدیک‌ترین نمونه‌های موجود در مجموعه داده موجود تعیین می‌شود. یکی از جنبه‌های مهم در پیاده‌سازی KNN، انتخاب متریک مناسب برای محاسبه فاصله بین نمونه‌هاست. در این مطالعه، از شاخص اقلیدسی برای محاسبه فاصله‌ها استفاده شده است. برای پیاده‌سازی این روش، ابتدا اطلاعات از مجموعه داده، خوانده می‌شود. ٪۸۰ داده‌ها برای مرحله آموزش[1] و ٪۲۰ داده‌ها برای مرحله آزمایش[2] لحاظ می‌شود. سپس نمرات و وظیفه تخصیص داده شده برای هر مهارت‌آموز استخراج می‌گردد.

در این مطالعه، تعداد همسایه‌های نزدیک (K) برابر با ۳ انتخاب شده است. این تعداد همسایه، پس از انجام آزمایش‌های مختلف انتخاب گردیده است.

برای هر مهارت‌آموز جدید، فاصله نمرات مهارت‌آموز با تمامی نمونه‌های موجود در دیتاست با استفاده از فاصله اقلیدسی محاسبه می‌شود. پس از محاسبه فاصله‌ها، سه مهارت‌آموزی که کمترین فاصله را با مهارت‌آموز جدید دارند، به عنوان نزدیک‌ترین همسایه‌ها انتخاب می‌شوند و وظیفه پیشنهادی مهارت‌آموز جدید بر اساس اکثریت وظایف تخصیص داده شده به سه همسایه نزدیک پیش‌بینی می‌شود. به عبارت دیگر، وظیفه‌ای که بیشترین تعداد در بین این سه همسایه دارد، به عنوان وظیفه پیش‌بینی‌شده برای مهارت‌آموز جدید انتخاب می‌شود.

۳) **CNN** در این قسمت از پژوهش، از شبکه عصبی پیچشی برای پیش‌بینی وظایف پیشنهادی استفاده شده است. ابتدا نمرات هر مهارت‌آموز و وظیفه تخصیص داده شده به آن‌ها، از مجموعه داده استخراج می‌شوند. سپس ٪۸۰ این داده‌ها برای مرحله آموزش و ٪۲۰ داده‌ها برای مرحله آزمایش استفاده می‌شوند.

نمرات مهارت‌آموزان به‌عنوان ورودی مدل هستند که شامل ۲۵ ستون است. هدف مدل پیش‌بینی وظیفه مناسب هر مهارت‌آموز است که مقداری بین ۱ تا ۳ دارد.

مدل شامل دو لایه پنهان[3] است که هر یک از این لایه‌ها دارای ۲۴ نورون[4] هستند. نورون‌ها واحدهای پردازشی شبکه عصبی هستند که وظیفه یادگیری و استخراج ویژگی‌های مهم از داده‌ها را بر عهده دارند. استفاده از دو لایه پنهان با ۲۴ نورون، پس از آزمایش و به صورت تجربی به دست آمده است و به مدل این امکان را می‌دهد که ساختارهای پیچیده و الگوهای موجود در داده‌ها را به‌خوبی یاد بگیرد.

لایه خروجی مدل شامل ۳ نورون است. این سه نورون نشان‌دهنده سه کلاس خروجی ممکن هستند، به طوری که هر کلاس نمایانگر یکی از مقادیر ۱، ۲ یا ۳ برای هر وظیفه قابل پیش‌بینی است.

آموزش مدل در طول ۱۰۰ دوره[5] انجام می‌شود که به مدل اجازه می‌دهد تا به تدریج وزن‌های خود را بهینه کرده و به دقت مطلوب در پیش‌بینی‌ها دست یابد.

با استفاده از این ساختار و تنظیمات، مدل توانست به طور موثری داده‌های ورودی را پردازش کرده و خروجی را با دقت قابل قبولی پیش‌بینی کند.

۴) **RNN** در این قسمت از پژوهش، از شبکه عصبی بازگشتی برای پیش‌بینی وظایف مناسب هر مهارت‌آموز استفاده شده است. ابتدا نمرات هر مهارت‌آموز و وظیفه تخصیص داده شده به آن‌ها، از مجموعه استخراج می‌شود. سپس ٪۸۰ این داده‌ها برای مرحله آموزش و ٪۲۰ داده‌ها برای مرحله آزمایش استفاده می‌شود.

نمرات مهارت‌آموزان به‌عنوان ورودی مدل هستند که شامل ۲۵ ستون است. هدف مدل، پیش‌بینی وظیفه مناسب هر مهارت‌آموز است که مقداری بین ۱ تا ۳ دارد.

مدل شامل سه لایه پنهان است که هر یک از این لایه‌ها دارای ۲۰۰ نورون هستند. استفاده از سه لایه پنهان با ۲۰۰ نورون، پس از آزمایش و به صورت تجربی به دست آمده است. لایه خروجی مدل نیز شامل ۳ نورون است. این سه نورون نشان دهنده سه کلاس خروجی ممکن است. هر کلاس نمایانگر یکی از مقادیر ۱، ۲ یا ۳ برای هر وظیفه قابل پیش‌بینی است.

برای جلوگیری از بیش‌برازش[6]، از dropout استفاده شده است. dropout یک روش تنظیم‌سازی است که در هر تکرار آموزشی، تعدادی از نورون‌ها به صورت تصادفی غیرفعال می‌شوند. این کار به مدل کمک می‌کند تا به ویژگی‌های خاص وابسته نشود و توانایی تعمیم‌دهی بهتری داشته باشد.

[1] Train [2] Test

[3] Hidden layers [4] Neuron [5] Epoch [6] Overfitting

مدل به مدت ۱۵۰۰ دوره آموزش داده شد. یک دوره به‌معنای گذر کامل از کل مجموعه داده‌های آموزشی توسط مدل است. این آموزش به مدل اجازه می‌دهد تا به تدریج وزن‌های خود را بهینه کرده و به دقت مطلوب در پیش‌بینی‌ها دست یابد.

با استفاده از این ساختار و تنظیمات، مدل توانست به طور موثری داده‌های ورودی که شامل نمرات مهارت‌آموزان است را پردازش کرده و وظیفه مناسب هر مهارت‌آموز را با دقت قابل قبولی پیش‌بینی کند. استفاده از لایه‌های پنهان متعدد با تعداد کافی نورون و تعداد بالای دوره‌ها، به مدل این امکان را می‌دهد که به تدریج با داده‌ها سازگار شود و به دقت پیش‌بینی بالاتری دست یابد.

۵) **LSTM** در این قسمت از پژوهش، از LSTM برای پیش‌بینی وظایف مناسب هر مهارت‌آموز استفاده شده است. همانند روش‌های قبلی، ٪۸۰ این داده‌های مجموعه داده برای مرحله آموزش و ٪۲۰ داده‌ها برای مرحله آزمایش استفاده می‌شود.

مدل شامل دو لایه پنهان است که هر یک از این لایه‌ها دارای ۱۰۰ نورون هستند. استفاده از دو لایه پنهان با ۱۰۰ نورون، پس از آزمایش و به صورت تجربی به‌دست آمده است. لایه خروجی مدل نیز شامل ۳ نورون است.

برای جلوگیری از بیش‌برازش، از dropout استفاده شده است. هم‌چنین مدل به مدت ۱۵۰۰ دوره آموزش داده شده است که به مدل اجازه می‌دهد تا به تدریج وزن‌های خود را بهینه کرده و به دقت مطلوب در پیش‌بینی‌ها دست یابد.

۶) **DQN** در این قسمت از پژوهش، از DQN برای پیش‌بینی وظایف مناسب هر مهارت‌آموز استفاده شده است. DQN قادر به یادگیری سیاست‌های بهینه برای تصمیم‌گیری در محیط‌های پیچیده می‌باشد.

نمرات مهارت‌آموزان به‌عنوان ورودی شبکه عصبی هستند که شامل ۲۵ ستون است. هدف DQN، پیش‌بینی وظیفه مناسب هر مهارت‌آموز است که مقداری بین ۱ تا ۳ دارد. این شبکه به گونه‌ای تنظیم شده که بتواند از روی نمرات ورودی، بهترین اقدام و وظیفه را برای هر مهارت‌آموز تشخیص دهد.

در طول فرآیند آموزش، مدل DQN با استفاده از داده‌های ورودی به‌روزرسانی می‌شود. هدف مدل، یادگیری سیاستی است که حداکثر پاداش را در محیط مورد نظر فراهم کند. این روش توانسته است به صورت موثری الگوهای پنهان در داده‌ها را شناسایی کرده و وظایف مناسب را برای هر مهارت‌آموز پیش‌بینی کند.

۵ ارزیابی نتایج

در این بخش، عملکرد شش روش یادگیری ماشین استفاده شده در سیستم یادگیری تطبیقی مورد ارزیابی قرار می‌گیرد و نتایج به‌دست آمده با استفاده از معیارهای Accuracy، Precision، Recall و امتیاز F1 مقایسه می‌شوند. برای ارزیابی، از ٪۲۰ داده‌ها برای آزمایش استفاده شده است.

جدول ۲. نتایج ارزیابی روش‌های انجام‌شده در مقاله

	Accuracy	Precision	Recall	F1 Score
Q-Learning	٪۷۵	٪۶۴	٪۷۵	٪۶۶
KNN	٪۸۴	٪۸۵	٪۸۴	٪۸۴
CNN	٪۹۸	٪۹۸	٪۹۸	٪۹۸
RNN	٪۹۶	٪۹۶	٪۹۶	٪۹۶
LSTM	٪۸۲	٪۸۲	٪۸۲	٪۸۲
DQN	٪۹۱	٪۹۲	٪۹۲	٪۹۲

۱.۵ معیارهای ارزیابی

معیار ارزیابی دقت (Accuracy): مطابق رابطه (۶)، این معیار، نشان‌دهنده نسبت تعداد پیش‌بینی‌های صحیح (TP) به تعداد کل پیش‌بینی‌ها می‌باشد.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \qquad (۶)$$

معیار ارزیابی صحت (Precision): مطابق رابطه (۶)، این معیار نشان دهنده نسبت تعداد نمونه‌های مثبت درست پیش‌بینی‌شده به کل نمونه‌های پیش‌بینی‌شده به عنوان مثبت می‌باشد (TP و FP).

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (۷)$$

معیار ارزیابی پوشش (Recall): مطابق رابطه (۸)، این معیار، نشان دهنده نسبت تعداد نمونه‌های مثبت درست پیش‌بینی شده به نمونه‌هایی است که انتظار داشتیم پیش‌بینی شوند ولی برنامه پیش‌بینی نکرده است.

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (۸)$$

معیار ارزیابی F1 Score: این معیار، میانگین وزن‌دار Precision و Recall را با کمک رابطه (۹) محاسبه می‌کند.

$$\text{F1 Score} = \frac{۲ * (\text{Recall} * \text{Precision})}{\text{Recall} + \text{Precision}} \qquad (۹)$$

۲.۵ ارزیابی روش‌ها

نتایج ارزیابی روش‌های استفاده شده در جدول ۲ قابل مشاهده می‌باشد.

با استفاده از Q-Learning، مدل توانست به مرور زمان ارزش هر یک از اقدامات را یاد بگیرد و برای مهارت‌آموز جدید بهترین وظیفه را انتخاب کند. جدول Q به تدریج با هر اجرا به‌روزرسانی می‌شود که این الگوریتم توانست به دقت ٪۷۵ برسد.

الگوریتم KNN با سه همسایه نزدیک و شاخص اقلیدسی برای محاسبه فاصله‌ها، توانست به دقت ٪۸۴ برسد.

شبکه عصبی CNN با دو لایه پنهان و ۲۴ نورون در هر لایه، توانست الگوهای پیچیده داده‌ها را به‌خوبی یاد بگیرد و وظایف را به صورت موثری پیش‌بینی کند و به دقت ۹۸٪ برسد.

شبکه عصبی RNN با سه لایه پنهان و ۲۰۰ نورون در هر لایه به خوبی وظایف مناسب را پیش‌بینی کرد و به دقت پیش‌بینی ۹۵٪ رسید.

شبکه عصبی LSTM با دو لایه پنهان و ۱۰۰ نورون در هر لایه توانست به دقت پیش‌بینی ۸۲٪ برسد.

شبکه عصبی DQN نیز به خوبی وظایف مناسب را پیش‌بینی کرد و به دقت پیش‌بینی ۹۱٪ رسید.

۶ نتیجه‌گیری

با مقایسه شش روش استفاده شده برای سیستم یادگیری تطبیقی، مشاهده می‌شود که هر یک از روش‌ها در پیش‌بینی وظایف، عملکرد قابل قبولی داشته‌اند، اما مدل‌های CNN، RNN و DQN به دلیل توانایی یادگیری الگوهای پیچیده‌تر، دقت و امتیاز F1 بهتری را نشان داده‌اند. الگوریتم Q-Learning نیز با گذر زمان و به‌روزرسانی جدول Q، توانسته است به دقت مناسبی دست یابد. روش‌های LSTM و KNN نیز عملکرد خوبی داشته‌اند اما نسبت به سایر مدل‌های عصبی پایین‌تر بوده‌اند.

۷ کارهای آینده

در ادامه این پژوهش، قصد داریم روش‌های یادگیری ماشین منتخب را در سامانه میدان سایبری نقش مجاز پیاده‌سازی نموده و به صورت میدانی نتایج حاصل از اعمال الگوریتم‌های یادگیری ماشین منتخب در بهبود فرآیند آموزش و تمرین در میدان سایبری را ارزیابی نماییم.

مراجع

[1] Top 20 cybersecurity trends to watch out for in 2023. https://www.simplilearn.com/top-cybersecurity-trends-article. Accessed: Jul. 18, 2024.

[2] Yamin, Muhammad Mudassar, Katt, Basel, and Gkioulos, Vasileios. Cyber ranges and security testbeds: Scenarios, functions, tools and architecture. *Computers & Security*, 88:101636, 2020.

[3] Vykopal, Jan, Ošlejšek, Radek, Čeleda, Pavel, Vizvary, Martin, and Tovarňák, Daniel. Kypo cyber range: Design and use cases. 2017.

[4] Vykopal, Jan, Seda, Pavel, Švábenskỳ, Valdemar, and Čeleda, Pavel. Smart environment for adaptive learning of cybersecurity skills. *IEEE Transactions on Learning Technologies*, 16(3):443–456, 2022.

[5] Seda, Pavel, Vykopal, Jan, Švábenskỳ, Valdemar, and Čeleda, Pavel. Reinforcing cybersecurity hands-on training with adaptive learning. in *2021 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9. IEEE, 2021.

[6] Sungkur, Roopesh Kevin and Maharaj, Manoj S. Design and implementation of a smart learning environment for the upskilling of cybersecurity professionals in mauritius. *Education and Information Technologies*, 26(3):3175–3201, 2021.

[7] Kerr, Philip. Adaptive learning. *Elt Journal*, 70(1):88–93, 2016.

[8] Lu, Jie, Wu, Dianshuang, Mao, Mingsong, Wang, Wei, and Zhang, Guangquan. Recommender system application developments: a survey. *Decision support systems*, 74:12–32, 2015.

[9] Aschmann, Michael Joseph. Towards a capability maturity model for a cyber range. 2020.

[10] Apoki, Ufuoma Chima, Hussein, Aqeel M Ali, Al-Chalabi, Humam K Majeed, Badica, Costin, and Mocanu, Mihai L. The role of pedagogical agents in personalised adaptive learning: A review. *Sustainability*, 14(11):6442, 2022.

[11] Beuran, Razvan, Vykopal, Jan, Belajová, Daniela, Čeleda, Pavel, Tan, Yasuo, and Shinoda, Yoichi. Capability assessment methodology and comparative analysis of cybersecurity training platforms. *Computers & Security*, 128:103120, 2023.

[12] Ukwandu, Elochukwu, Farah, Mohamed Amine Ben, Hindy, Hanan, Brosset, David, Kavallieros, Dimitris, Atkinson, Robert, Tachtatzis, Christos, Bures, Miroslav, Andonovic, Ivan, and Bellekens, Xavier. A review of cyber-ranges and test-beds: Current and future trends. *Sensors*, 20(24):7148, 2020.

[13] Chen, Lijia, Chen, Pingping, and Lin, Zhijian. Artificial intelligence in education: A review. *Ieee Access*, 8:75264–75278, 2020.

[14] Mahesh, Batta. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1):381–386, 2020.

[15] Alzubaidi, Laith, Zhang, Jinglan, Humaidi, Amjad J, Al-Dujaili, Ayad, Duan, Ye, Al-Shamma, Omran, Santa-

maría, José, Fadhel, Mohammed A, Al-Amidie, Muthana, and Farhan, Laith. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74, 2021.

[16] Uc-Cetina, Victor, Navarro-Guerrero, Nicolás, Martin-Gonzalez, Anabel, Weber, Cornelius, and Wermter, Stefan. Survey on reinforcement learning for language processing. *Artificial Intelligence Review*, 56(2):1543–1575, 2023.

[17] Lv, Liangheng, Zhang, Sunjie, Ding, Derui, and Wang, Yongxiong. Path planning via an improved dqn-based learning policy. *IEEE Access*, 7:67319–67330, 2019.

[18] Madani, Youness, Ezzikouri, Hanane, Erritali, Mohammed, and Hssina, Badr. Finding optimal pedagogical content in an adaptive e-learning platform using a new recommendation approach and reinforcement learning. *Journal of Ambient Intelligence and Humanized Computing*, 11:3921–3936, 2020.

[19] Sayed, Wafaa S, Noeman, Ahmed M, Abdellatif, Abdelrahman, Abdelrazek, Moemen, Badawy, Mostafa G, Hamed, Ahmed, and El-Tantawy, Samah. Ai-based adaptive personalized content presentation and exercises navigation for an effective and engaging e-learning platform. *Multimedia Tools and Applications*, 82(3):3303–3333, 2023.

[20] Liu, Xinyang and Ardakani, Saeid Pourroostaei. A machine learning enabled affective e-learning system model. *Education and Information Technologies*, 27(7):9913–9934, 2022.

[21] Shawky, Doaa and Badawi, Ashraf. A reinforcement learning-based adaptive learning system. in *The international conference on advanced machine learning technologies and applications (AMLTA2018)*, pp. 221–231. Springer, 2018.

[22] Boussakssou, Mohamed, Hssina, Bader, and Erittali, Mohammed. Towards an adaptive e-learning system based on q-learning algorithm. *Procedia Computer Science*, 170:1198–1203, 2020.

[23] Seda, Pavel, Vykopal, Jan, Švábenskỳ, Valdemar, and Čeleda, Pavel. Reinforcing cybersecurity hands-on training with adaptive learning. in *2021 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9. IEEE, 2021.

MonadiAfta-Bianual Scientific Journal for Cyberspace Security
2024, Volume 22, Number 2 (pp. 34–44)
https://monadi.isc.org.ir

Presented at the ISCISC 2024 in Tarbiat Modares University & Research Institute for Information and Communication Technology, Tehran, Iran

Intelligent automation of scenario execution in cyber ranges using machine learning techniques★

Farnoosh Karimi [1], Behrouz Tork Ladani [1,*] and Behrouz Shahgholi Ghahfarokhi [2]

[1] *Department of Software Engineering, Faculty of Computer Engineering, Isfahan University, Isfahan, Iran*
[2] *Department of Information Technology, Faculty of Computer Engineering, Isfahan University, Isfahan, Iran*

ARTICLE INFO.

Keywords:
Adaptive learning
Learning cybersecurity
KYPO cyber range
Learning Cybersecurity
Machine learning
Reinforcement learning

Type: Research paper

ABSTRACT

As the intensity of global cybersecurity threats continues to rise, the need for training security professionals has gained greater significance. Educational programs, complemented by laboratories and the execution of cybersecurity exercises, play a fundamental role in enhancing both offensive and defensive capabilities. The execution of such exercises is particularly crucial in operational networks, where testing cyberattacks may not be feasible. Cyber ranges offer an appropriate platform for conducting these exercises. A primary challenge in cybersecurity education is aligning training programs with the diverse skill levels of learners. Adaptive learning, powered by artificial intelligence and recommendation systems, can provide an effective solution for delivering personalized instruction. This study focuses on the KYPO Cyber Range to examine the potential of substituting or augmenting the role of the instructor with an AI-based recommendation agent. The objective of this research is to minimize human intervention and improve the efficiency of the training process. To this end, data collected from the KYPO Cyber Range, developed by Masaryk University, has been utilized, and various machine learning models have been applied to automate and optimize the training process. The results of this research indicate that the integration of artificial intelligence can enhance the performance of educational systems and reduce evaluation time.

© 2024 ISC

★ The ISCISC 2024 Program Committee effort is highly acknowledged for reviewing this paper.
* Corresponding author
Email addresses: farnoosh.karimi95@eng.ui.ac.ir (Farnoosh Karimi), ladani@eng.ui.ac.ir (Behrouz Tork Ladani), shahgholi@eng.ui.ac.ir (Behrouz Shahgholi Ghahfarokhi)
© 2024 ISC. All rights reserved.