



طراحی و پیاده سازی یک چارچوب هوشمند برای تست خودکار نرم افزارهای تحت وب با استفاده از هوش مصنوعی

علیرضا رحیمی پور انارکی^۱

^۱دانشجوی کارشناسی ارشد مهندسی نرم افزار، دانشگاه آزاد اسلامی، واحد تهران مرکزی ، arrahimipourdev@gmail.com

چکیده

با رشد روزافزون پیچیدگی برنامه های تحت وب، نیاز به تست خودکار جامع و مؤثر بیش از پیش احساس می شود. در این مقاله، رویکردهای اخیر تست خودکار وب اپلیکیشن ها با بهره گیری از الگوریتم های هوش مصنوعی (یادگیری ماشین، یادگیری عمیق و شبکه های عصبی) مرور و تحلیل می شوند. پس از بررسی چالش های تست سنتی تحت وب (مانند فرم های پویا و منطق پیچیده)، روش های هوشمند در سه دسته عمده معرفی می شوند: (۱) یادگیری ماشین برای تولید و تحلیل موارد آزمون، (۲) شبکه های عصبی و یادگیری عمیق برای درک عناصر رابط کاربری و آزمون خودکار، و (۳) یادگیری تقویتی عمیق برای کاوش خودگردان فضای حالات وب. نمونه هایی از چارچوب های منتخب، مانند WebRLED و WebExplor، به همراه نتایج تجربی آن ها ارائه می شود. نتیجه گیری نشان می دهد که روش های مبتنی بر هوش مصنوعی می توانند به طور چشمگیری پوشش کد و نرخ کشف خطا را بهبود دهند. چشم اندازهای پژوهشی آینده شامل یکپارچه سازی مدل های زبان بزرگ (LLM) و گسترش دسترسی به دیتاست های واقعی برای آموزش مدرن است. کلید واژه- تست خودکار نرم افزار، برنامه های تحت وب، هوش مصنوعی، یادگیری ماشین، یادگیری عمیق..

می‌شوند. در نهایت، با مقایسه نتایج پژوهش‌های اخیر، کارایی این روش‌ها تحلیل و چشم‌انداز آینده تشریح می‌گردد.

۱- مقدمه

۲- روش‌های سنتی و نقش هوش مصنوعی در تست وب

تست خودکار نرم‌افزار شامل ایجاد، اجرا و تحلیل مجموعه‌ای از آزمون‌ها به صورت خودکار است. در زمینه وب اپلیکیشن‌ها، روش‌هایی مانند تست رندوم (فازینگ)، تست مدل‌محور و تولید آزمون به کمک الگوریتم‌های تکاملی معمولاً استفاده می‌شوند. با این حال، این روش‌ها محدودیت‌هایی دارند: روش‌های رندوم اغلب عملیات نامعتبر تولید کرده و حالات پیچیده را نمی‌یابند؛ روش‌های مدل‌محور به مدلسازی دقیق برنامه وابسته‌اند که در وب پویا دشوار است؛ و روش‌های تکاملی برای محیط‌های ایستا مناسب‌ترند.

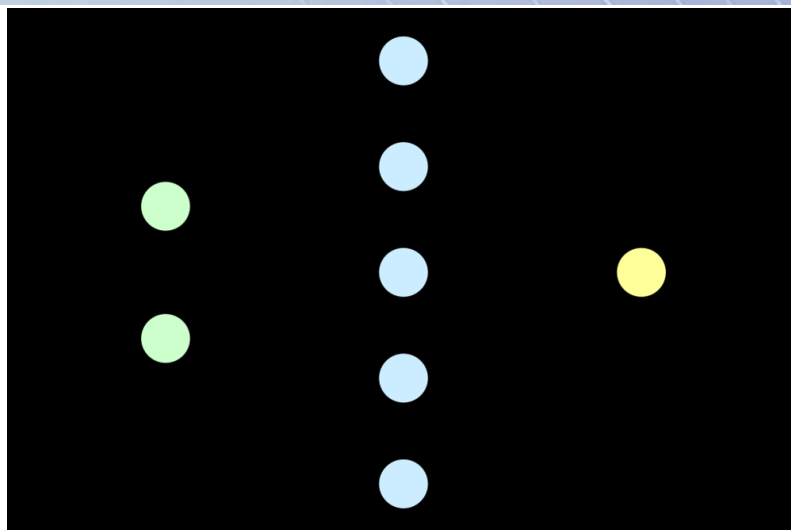
هوش مصنوعی به عنوان راهکاری جدید، امکان بهبود تست خودکار را فراهم می‌سازد. به طور کلی، یادگیری ماشین (ML) و یادگیری عمیق (Deep Learning) می‌توانند با تحلیل داده‌ها و استخراج الگو، در فرایند تولید و اولویت‌بندی آزمون‌ها کمک کنند. به عبارت دیگر، این روش‌ها از حجم بالای داده‌های آزمون و اطلاعات اجرای قبلی می‌آموزند و توانایی خود را بهبود می‌دهند. به عنوان نمونه، الک (۲۰۲۴) بررسی جامعی از یکپارچه‌سازی یادگیری عمیق در اتوماسیون تست ارائه کرده و بر قابلیت یادگیری عمیق در خودکارسازی روندهای پیچیده تست تاکید نموده است.

تعاریف پایه نشان می‌دهد که شبکه‌های عصبی مصنوعی (ANN) با لایه‌های ورودی، پنهان و خروجی می‌توانند نقش کلاس‌سازی عناصر UI یا تولید ویژگی‌های مرتبط با صفحات وب را ایفا کنند. شکل زیر مثالی ساده از یک شبکه عصبی پیش‌خور را نشان می‌دهد که در آن لایه ورودی (سبز)، لایه‌های پنهان (آبی) و لایه خروجی (زرد) به صورت گسسته نمایش داده شده‌اند:

امروزه نرم‌افزارهای تحت وب در حوزه‌هایی مانند تجارت الکترونیک، آموزش و خدمات آنلاین نقش کلیدی دارند. پیچیدگی زیاد منطق کسب‌وکار و واسط کاربری این برنامه‌ها، تست آن‌ها را چالش‌برانگیز کرده است. از یک سو، روش‌های سنتی تست خودکار (رندوم‌محور، مدل‌محور یا مبتنی بر قواعد) در مواقع زیادی دچار کاستی‌هایی مثل پوشش ناکافی و تولید عملیات نامعتبر می‌شوند. از سوی دیگر، تغییرات سریع در وب اپلیکیشن‌ها، نگهداری تست‌های خودکار را سخت می‌کند. گارنتر پیش‌بینی کرده است که تا سال ۲۰۲۱ حدود نیمی از شرکت‌ها از تست خودکار هوشمند مبتنی بر هوش مصنوعی بهره خواهند برد. این پیش‌بینی اهمیت به‌کارگیری AI/ML در تست را نشان می‌دهد.

چالش‌های اصلی در تست وب اپلیکیشن شامل تکمیل خودکار فرم‌های پویا، تعاملات پیچیده کاربر با عناصر صفحه و شناسایی خطاهای نهفته در مسیرهای خاص است. به عنوان مثال، استادزاده و همکارانش بیان کرده‌اند که «مهم‌ترین مشکل در خودکارسازی آزمون برنامه‌های تحت وب، پرکردن فرم است»؛ چرا که بسیاری از قسمت‌های پویای سایت در پشت فرم‌ها مخفی هستند. این مسائل به شکل گسترده‌ای نیاز به روش‌های نوینی را نشان می‌دهند.

در این مقاله به بررسی روش‌های هوشمند تست خودکار وب می‌پردازیم. ابتدا مفاهیم پایه و روش‌های قدیمی‌تر مرور می‌شوند، سپس تکنیک‌های مبتنی بر یادگیری ماشین و شبکه‌های عصبی معرفی خواهند شد. در ادامه، روش‌های یادگیری تقویتی برای کاوش خودکار رابط‌های وب بحث



شکل ۱- یک شبکه عصبی ساده پیش خور. شبکه‌های عصبی همچون این ساختار می‌توانند برای شناسایی الگوهای رابط کاربری و تولید خودکار موارد آزمون مورد استفاده قرار گیرند.

جدول ۱- مقایسه‌ای از رویکردهای مبتنی بر یادگیری تقویتی برای تست خودکار وب

روش	تکنیک AI	مورد آزمایش	نتایج گزارش شده
WebRLED	یادگیری تقویتی عمیق (DQN) + (تجزیه‌ی مربعی)	۱۲ برنامه متن‌باز + ۵۰ وبسایت مشهور	پوشش کد/حالت بالاتر نسبت به روش‌های دیگر، کشف ۶۹۵ شکست منحصر به فرد
WebExplor	RL کنجکاوی محور (-Curiosity) (driven)	۶ پروژه واقعی + یک SaaS تجاری + ۵۰ وبسایت	پوشش و نرخ کشف خطای بالاتر از SOTA، کشف ۱۲ خطای جدید در برنامه تجاری و ۳۴۶۶ استثنا
DRIFT	Q-یادگیری با GNN (شبکه عصبی گراف)	سیستم عامل ویندوز (وظایف GUI)	یادگیری وظایف چندگانه رابط کاربری، برتری دو برابری نسبت به حالت تصادفی در انجام وظایف

می‌توان آزمون‌های جدیدی تولید کرد که به‌طور هوشمند مجموعه حالات بیشتری را پوشش دهند. روش‌های تولید خودکار داده‌های آزمون، مانند استفاده از الگوریتم‌های تکاملی (مثل GA) با به‌کارگیری داده‌های آماری سیستم، می‌توانند بهینه‌سازی پوشش آزمون را امکان‌پذیر سازند. همچنین، تکنیک‌های NLP (مانند مدل‌های زبان بزرگ) اخیراً برای تولید اسکرپت‌های آزمون و جمله‌بندی سناریوهای تست مورد استفاده قرار گرفته‌اند.

در حوزه واسط کاربری وب، یادگیری ماشین می‌تواند برای کلاس‌بندی و شناسایی خودکار عناصر صفحه به‌کار رود. به

جدول ۱ نیز مقایسه‌ای از نتایج برخی روش‌های اخیر مبتنی بر یادگیری تقویتی در تست وب را نشان می‌دهد. در این روش‌ها، عامل هوشمند بر مبنای پاداش، به صورت خودگردان صفحات وب را کاوش می‌کند.

۳- یادگیری ماشین و روش‌های مبتنی بر دیتا

یکی از کاربردهای اولیه ML در تست، تولید خودکار موارد آزمون است. با آموزش مدل‌های یادگیری بر روی داده‌های پیشین (از جمله درخواست‌های کاربر و پاسخ‌های سیستم)،

گرایش‌های نوین است.

۵- یادگیری تقویتی در تست وب

یادگیری تقویتی (RL)، به واسطه توانایی خود در یادگیری از طریق تعامل با محیط و بهبود سیاست بر اساس پاداش، در چند سال اخیر به حوزه تست خودکار وارد شده است. در روش‌های RL، عامل (Agent) به مرور با اجرای عمل‌هایی نظیر کلیک یا وارد کردن داده، صفحات مختلف را کاوش می‌کند و از پاداش (مثلاً افزایش پوشش کد یا مشاهده حالت جدید) هدایت می‌شود.

محققان مختلفی چارچوب‌هایی را پیشنهاد کرده‌اند. «DRIFT» از مایکروسافت (۲۰۱۹) یک چارچوب RL برای تست سیستم‌عامل ویندوز بود که با Q-یادگیری و به‌کارگیری شبکه‌های عصبی گراف (GNN)، نمایانگرهای درختی از رابط کاربری را تحلیل می‌کرد. نتایج نشان داد آزمونگر یادگیرنده توانسته به صورت خودکار وظایف GUI را انجام دهد و بارها نسبت به جستجوی تصادفی موفق‌تر ظاهر شود. به‌طور خاص برای وب، «Zheng et al., 2021» WebExplor رویکردی کنجکاوی‌محور ارائه می‌دهد که با ساخت متناوب یک اتوماتای حالت (برای راهنمایی کاوش) و به‌کارگیری RL، آزمون‌های با کیفیت بالا تولید می‌کند. نویسندگان گزارش کرده‌اند که این روش نسبت به تکنیک‌های پیشین، در اغلب آزمون‌ها نرخ کشف خطا و پوشش بهتری دارد؛ به‌طور مثال، در یک برنامه تجاری SaaS، ۱۲ شکست کشف نشده قبلی را با موفقیت شناسایی کرد. اخیراً، «Wu et al., 2024» WebRLED از یادگیری تقویتی عمیق برای تست رابط وب استفاده کرده است. این روش فضای عملیاتی را بر اساس یک شبکه چهارخانه‌ای (Grid) نمایش می‌دهد تا مسأله همراستایی (misalignment) اقدامات را حل کند، و یک مدل پاداش کنجکاوی‌محور نیز ارائه می‌دهد. ارزیابی‌های انجام شده نشان داد که WebRLED به‌طور قابل ملاحظه‌ای پوشش کد/حالت و نرخ کشف خطا را نسبت به روش‌های مرسوم افزایش داد و ۶۹۵ خطای جدید را در ۵۰ برنامه واقعی کشف کرد. این نتایج موید کارایی بالای استفاده از RL برای اکتشاف موثر در وب اپلیکیشن‌هاست.

عنوان مثال، با آموزش شبکه عصبی بر روی تصاویر عناصر (دکمه، فرم، منو و ...) و ویژگی‌های DOM، می‌توان عامل تست را قادر ساخت تا المان‌های مهم را تشخیص داده و اقدام مناسب (مانند کلیک یا وارد کردن داده) انجام دهد. این امر به ویژه در ابزارهای جدید "خودترمیم" (self-healing) تست، مطرح است؛ به گونه‌ای که هنگام تغییر در طراحی صفحه، مدل ML می‌تواند عنصر متناظر را شناسایی و اسکریپت را بدون دخالت انسان به‌روزرسانی کند.

یکی دیگر از نقش‌های ML، ارزیابی خودکار صحت برنامه (گنجاندن اوراکل) است. در این رویکردها، مدل‌های یادگیری می‌آموزند که بر اساس خروجی‌های مشاهده‌شده، وجود خطا را پیش‌بینی کنند. به عنوان مثال، مدل‌های طبقه‌بندی می‌توانند شناسایی کنند که یک رفتار خاص، رفتار نامطلوب یا خطایی را نشان می‌دهد. مطالعاتی نیز نشان داده‌اند که یادگیری ماشین می‌تواند اوراکل‌های دینامیک تعریف شده توسط انسان را تا حدی بیاموزد و برای تعیین پاس/فیل در آزمون‌های پیچیده مورد استفاده قرار گیرد.

۴- یادگیری عمیق و شبکه‌های عصبی در تست خودکار

یادگیری عمیق با استفاده از شبکه‌های عصبی چندلایه، می‌تواند پردازش اطلاعات پیچیده را به‌خوبی انجام دهد. در زمینه تست وب، کاربردها شامل تولید کد اسکریپت‌های آزمون و پردازش خودکار ویژگی‌های رابط کاربری است. برای نمونه، روش‌های مبتنی بر ترنسفورمرها و RNN برای تبدیل توضیحات متنی UI به اسکریپت‌های آزمون خودکار بررسی شده‌اند. در یک تحقیق، نویسندگان نشان دادند که با استفاده از مدل‌های زبان پایه (مانند GPT-3)، می‌توان آزمون‌های GUI را با کیفیت مناسبی تولید کرد و پوشش کد را بهبود داد.

علاوه بر این، تکنیک‌های بینایی کامپیوتری مانند تجزیه و تحلیل تصویر می‌توانند در شناسایی بصری عناصر UI کاربرد داشته باشند. برخی از ابزارهای صنعتی (مثل Applitools) از یادگیری عمیق برای ارزیابی تغییرات بصری بین نسخه‌های مختلف رابط کاربری استفاده می‌کنند تا اشکالات ظاهری را بیابند. هرچند این مقاله به‌صورت جامع به این موضوع نمی‌پردازد، اما باید اشاره کرد که حوزه تست بصری و مقایسه تصاویر با کمک CNNها، جزء

۶- نتایج و جمع‌بندی

استادزاده، آ.، ضیایی، ن.، ابدی، م. و آخوندی، ب.، «ارائه روشی جهت خودکارسازی تست مبتنی بر مدل در برنامه‌های تحت وب»، اولین همایش بین‌المللی مهندسی برق، علوم کامپیوتر و فناوری اطلاعات، همدان، ۱۳۹۶. کرمی‌سرخه‌چقایی، ف.، «هوش مصنوعی و یادگیری ماشین در توسعه نرم‌افزار»، اولین همایش علوم انسانی با رویکرد نوین، آستارا، ۱۴۰۳.

مرور منابع پژوهشی نشان می‌دهد که هوش مصنوعی در تست خودکار وب توانسته قابلیت‌های سنتی را بهبود بخشد. با به‌کارگیری ML و DL، مراحل تولید اسکریپت و انتخاب موارد آزمون سریع‌تر و دقیق‌تر شده است. استفاده از RL به‌ویژه در ارتقای پوشش آزمون و یافتن خطاهای نهفته مؤثر است. تجارب صنعتی نیز موید این است که روش‌های هوشمند می‌توانند هزینه‌های نگهداری آزمون را کاهش دهند و انعطاف‌پذیری بیشتری ایجاد کنند.

از سوی دیگر، چالش‌هایی باقی است: داده‌ها و مدل‌های آموزشی جامع برای وب اپلیکیشن‌ها محدودند و generalizability مدل‌ها نیاز به بررسی دارد. همچنین، تفسیر نتایج ML برای تضمین کیفیت باید قابل اعتماد باشد. در آینده، مدل‌های زبان بزرگ (LLM) مانند GPT و LLaMA نقش مهمی در خودکارسازی تولید سناریوهای تست خواهند داشت. به‌علاوه، یکپارچه‌سازی هوشمندی در سرتاسر چرخه توسعه (از تحلیل نیازمندی تا ارسال نهایی) و بررسی امنیت و عملکرد با روش‌های AI از موضوعات پژوهشی فعال به‌شمار می‌روند.

۶-۱- سپاسگزاری

نگارنده این مقاله وظیفه خود می‌داند از خانواده محترم خود به‌دلیل حمایت‌ها و همراهی‌های بی‌دریغ‌شان در مسیر علمی و پژوهشی، صمیمانه تشکر و قدردانی نماید. همچنین از داوران گرامی کنفرانس بابت بازخوردهای سازنده و علمی‌شان که موجب ارتقاء کیفیت این مقاله شد، سپاسگزاری می‌نمایم.

مراجع

- [1] N. K. Ale, "Enhancing Test Automation with Deep Learning: Techniques, Challenges and Future Prospects," CS&IT Conf. Proc., vol. 14, no. 15, 2024.
- [2] Z. (Wu) et al., "Deep Reinforcement Learning for Automated Web GUI Testing," arXiv preprint arXiv:2504.19237, 2024.
- [3] Y. Zheng et al., "Automatic Web Testing using Curiosity-Driven Reinforcement Learning," arXiv:2103.06018, 2021.
- [4] L. Harries et al., "DRIFT: Deep Reinforcement Learning for Functional Software Testing," in Proceedings of the Deep RL Workshop, NeurIPS 2019, 2019.
- [5] J. Wang, Y. Huang, C. Chen, Z. Liu, S. Wang, Q. Wang, "Software Testing with Large Language Models: Survey, Landscape, and Vision," arXiv preprint arXiv:2307.07221, 2023.