سیستم های مدیریت اطلاعات و رویدادهای امنیتی مبتنی بر هوش مصنوعی[1]

مرتضی ملکی نژاد شوشتری

دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران

چکیده

وظیفه سیستم های مدیریت اطلاعات و رویداد های امنیتی جمع‌آوری اطلاعات، پردازش و انجام اقدام متناسب با آن است. در رویکرد های سنتی، عموما پردازش و تشخیص اقدام مناسب مبتنی بر قاعده انجام می‌شود و لزوما دقت بالایی به ما نمی‌دهد. این امر می‌تواند منجر به اقدام نامناسب، نیاز به توان پردازشی بالاتر و غیره شود. این مقاله به بررسی چگونگی استفاده از هوش مصنوعی در این سیستم ها و بهبود هایی که منجر می‌شود می‌پردازد.

کلمات کلیدی

سیستم های مدیریت اطلاعات و رویداد های امنیتی، هوش مصنوعی، یادگیری ماشین[2]، سیستم های پاسخ بحران [3]، تحلیل سابقه [4]، تشخیص ناهنجاری [5] سیستم های تشخیص نفوذ [6]

۱ مقدمه

سیستم های مدیریت اطلاعات و رویداد های امنیتی نقش مهمی را برای امنیت سیستم ها دارند. این سیستم ها وظیفه این را دارند تا از بخش های مختلف اطلاعات مختلفی را بگیرند، سپس آن ها را پردازش کنند، بعد اقدام مناسبی را پیدا کنند و در نهایت آن اقدام را انجام دهند. بعد از این هم ممکن است از این داده ها برای کالبدشکافی [7] حادثه در آینده استفاده شود.

ولی این کار بدون چالش نیست. حجم داده ها بسیار زیاد هستند، دقت پایین منجر به نیاز به مداخله انسانی می‌شود، نیاز به توان پردازشی بالا هست و غیره. هوش مصنوعی برای کمک به حل این چالش ها در قسمت های مختلف یک سیستم مدیریت اطلاعات و رویداد امنیتی استفاده می‌شود.

۲ چالش های سیستم های سنتی

یک سیستم یک سابقه تولید می‌کند. سیستم مدیریت اطلاعات این سابقه را جمع می‌کند. جهت کاهش نیاز به توان محاسباتی ممکن است هنگام جمع آوری غربال شوند. سپس همه این سابقه ها در یکجا پردازش می‌شوند و سپس تشخیص و انجام اقدام مناسب رخ می‌دهد.

با توجه به اینکه همه سابقه ها در یکجا جمع می‌شوند بعد نرم‌افزار باید بهینه باشد تا بتواند این حجم داده را پردازش کند. همچنین علیرغم نرم‌افزار بهینه، به سخت افزار مناسبی برای پردازش نیاز هست.

از آنجایی که سیستم های سنتی از روش های مبتنی بر قاعده برای تشخیص ناهنجاری استفاده می‌کنند، تشخیص آن ها به ناچار دچار اشتباه می‌شود. به دلیل حساسیت مسائل

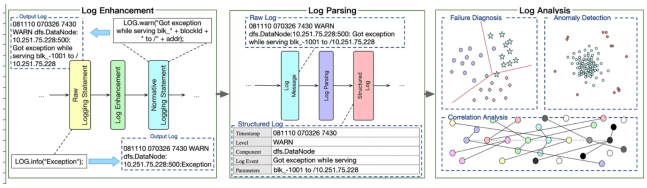
[1] AI based Security information and event management systems
[2] Machine learning
[3] Incident response systems
[4] Log analysis
[5] Anomaly
[6] Intrusion Detection Systems
[7] Postmorterm
[8] False positive

امنیتی، این خطا معمولا به صورت مثبت کاذب [۸] نمود می‌دهد زیرا اولویت بر این است که هیچ ناهنجاری نادیده گرفته نشود یعنی منفی کاذب [۹] نداشته نباشیم. اما این امر منجر به این می‌شود که نیاز به مداخله انسانی بیشتر شود.

پس از تشخیص ناهنجاری اقداماتی که در سیستم های سنتی وجود دارد محدود و بعضا ناموثر هستند زیرا امکان مثبت کاذب هست و ممکن است اقداماتی که تاثیرات سنگینی داشته باشند از گزینه ها خارج شوند. همچنین ممکن است ارتباط بین تشخیص تهدید و تشخیص اقدام دقیق نباشد.

۳ پردازش سابقه با استفاده از هوش مصنوعی

یکی از نکات مهم در پردازش سابقه که این کار را دشوار تر می‌کند عدم ساختار کلی برای سابقه است. اما هوش مصنوعی و مدل های زبانی برای داده های بی‌ساختار عملکرد بسیار خوبی دارند. یک چارچوب [۱۰] پردازش سابقه شامل سه مرحله است؛ بهبود سابقه، پردازش سابقه و تحلیل سابقه.



شکل ۱: قسمت های مختلف یک سیستم پردازش سابقه

هوش مصنوعی در هر سه قسمت می‌تواند به ما کمک کند.

بهبود سابقه

با وجود اینکه کد مربوط به چاپ کردن سابقه درصد کمی از کد را تشکیل می‌دهد ولی نرخ تغییر آن دو برابر سایر قسمت های کد است. حدود ۳۳ درصد تغییرات به صورت واکنشی هستند نه پیشگیرانه. سابقه مناسب بیش از دو برابر تشخیص مشکل را سرعت می‌بخشید. همچنین ۵۷ درصد مشکلات قابل تشخیص با سابقه خطا ثبت نشده‌اند. [۵]

محتوای سابقه

یک سابقه شامل سطح سابقه [۱۱] توضیحات سابقه و متغیر هاست. پژوهش های متعددی روی تعیین سطح سابقه انجام شده که به این می‌پردازد که اگر در بخشی از کد تصمیم به خروجی دادن سابقه گرفته شده است،چه سطحی مناسب آن است.

همچنین متغیر ها نقش مهمی در میزان اطلاعاتی که توسط سابقه به بیرون انتقال داده می‌شوند دارد. حدود ۲۵ درصد تغییرات سابقه مربوط به متغیر هاست. برای این موضوع هم پژوهش هایی برای یافتن یک بهبود دهنده سابقه بر پایه هوش مصنوعی وجود دارد.

توضیحات سابقه منعطف ترین بخش یک سابقه است. وظیفه بخش توضیحاتت، توصیف رویدادی است که اتفاق افتاده. یک توضیح نامناسب می‌تواند منجر به کند شدن و یا گمراه شدن توسعه دهنده هنگام تشخیص خطا باشد. سیستم های هوش مصنوعی با استفاده از مفاهیمی مثل همبستگی بین توضیحات مختلف و غیره به تولید خودکار توضیح سابقه روی آورده‌اند.

یک موضوع نیز مکان سابقه است. یعنی در کجای کد سابقه خروجی داده شود. برای این امر مدل های مبتنی بر درخت تصمیم [۱۲] وجود دارند. یکی از این سیستم ها Log Advisor نام دارد. [۶]

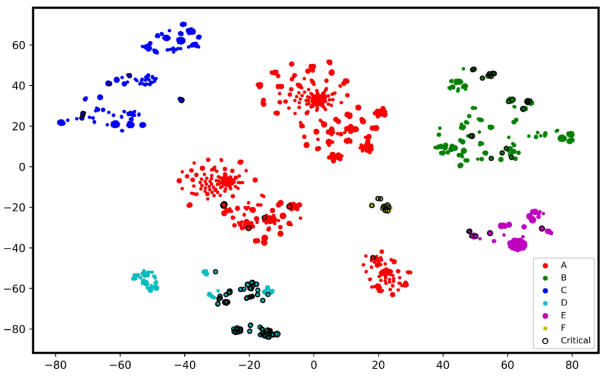
پردازش سابقه

سابقه ها عموما بدون ساختار مشخصی هستند ولی بخش تحلیل سابقه با داده با ساختار بهتر کار می‌کند از این رو در

[9] False negative
[10] framework
[11] Log Level
[12] Decision Tree

بخش پردازش سابقه هدف این است تا سابقه ها به حالت مشخصی در بیایند تا در بخش تحلیل بتوان بهتر عمل کرد.

برای این هدف سیستم های خوشه بندی، استخراج الگوهای پرتکرار و مدل های زبانی وجود دارند تا بتوان حالت داده را به فرم مشخصی در بیاورند. یکی از خروجی های مرسوم به این حالت، درخت است که قسمت های مختلف سابقه گره های آن را تشکیل می‌دهند. [۵]



شکل ۲: نمونه ای از انجام خوشه بندی روی هشدار های یک سامانه [۱]

تحلیل سابقه

سیستم های سنتی بر اساس قاعده تهدید ها را تشخیص می‌دهند. و بر این اساس است که یک فرد حرفه ای قاعده هایی برای تشخیص تهدید تعریف می‌کند. این امر به روز رسانی آن ها را سخت می‌کند و احتمال خطا را بالا می‌برد.

سیستم های مبتنی بر هوش مصنوعی به دنبال این هستند تا الگوی های مشکوک را در سابقه ها تشخیص دهند و از روی آن بتوانند تهدید ها را تشخیص دهند.

تشخیص تهدید می‌تواند به حالت کلی تشخیص ناهنجاری باشد و یا به صورت جزئی تر به نوع تهدید بپردازد. برای این منظور نیز سیستم ها از فناوری های متنوعی مثل درخت تصمیم، SVM[13]، KNN [14] استفاده می‌کنند. [۵]

داده ها و مدل های پردازش سابقه

برای این که یک مدل پردازش سابقه بتواند به خوبی کار کند باید در هر سه مرحله بهبود، پردازش و تحلیل به خوبی عمل کند. ولی عموما هر سه مرحله از مجموعه داده های متفاوتی استفاده نمی‌شود زیرا در عمل یک هدف را دنبال می‌کنند.

برای مثال LogHub [۷] یک مجموعه داده از سابقه های سیستم های مختلف در حالت های مختلف است که جمع آوری شده تا آموزش مدل ها را تسهیل بخشد. به دلیل کمبود داده و یا نبود معیار سنجش [15] در مراحل اولیه استفاده از هوش مصنوعی در این زمینه کمی کند بوده ولی LogHub سعی کرده که مشکل کمبود داده را حل کند.

همچنین یک سیستم تحلیل نیاز دارد تا با بخش های مختلفی ارتباط برقرار کند تا مطمئن شود خروجی غلطی ندهد. برای همین اکثر مدل ها از همه بخش های سابقه مثل فراداده [16]، مکان سابقه در کد و ... استفاده می‌کنند تا بتوانند دقت خود را بالا ببرند [۳] این امر منجر به پیچیدگی فرآیند آموزش می‌شود ولی در صورت پیاده سازی مناسب دقت خروجی را بالاتر می‌برند.

٤ آینده هوش مصنوعی در سیستم ها

سیستم های سنتی یا حتی سیستم های اولیه مبتنی بر هوش مصنوعی روی جمع آوری داده تمرکز داشتند ولی با پیشرفت هوش مصنوعی و افزایش تمرکز عمومی به رویکرد های پیشگیرانه نسبت به واکنشی روی آورد. زیرا نرخ دقت بالاتر رفت و نیاز به نیروی انسانی کمتر شد و از این رو می‌توان با هزینه کمتری کارهای پیشگیرانه بیشتری کرد (که شاید قبلا از لحاظ هزینه منطقی نبود.)

عموما هوش مصنوعی در فاز تحلیل در سه قسمت به کار می‌آید؛ تشخیص الگو، تحلیل پیشگیرانه و کاهش خطر. یعنی در قسمت اول الگو های ناهنجاری را یاد بگیرد و آن ها

[13] Support Vector Machine
[14] K nearest neighbours
[15] Benchmark
[16] metadata

را تشخیص دهد. سپس تشخیص دهد که احتمال وقوع چه مشکلی وجود دارد و سوم با توجه به اینکه تا بروز تهدید چه قدر فرصت هست اقدامات مناسب را انجام بدهد. این اقدام ها می‌توانند شامل قرنطینه کردن، درست کردن مشکل و یا تغییر سیاست سیستم باشند. [٤]

چالش ها

خود سیستم های مبتنی بر هوش مصنوعی نیز چالش هایی دارند. [٢]

- هزینه: هوش مصنوعی از آنجایی که نیاز به توان پردازشی بالایی دارد هزینه های خاص خود را دارد که می‌تواند بسته به نیاز بیشتر و یا کمتر از هزینه نیروی انسانی متناسب با آن باشد. فارغ از هزینه عملیاتی در برخی زمینه ها نیاز است تا تیمی داشت تا بتواند مدل ها را بهبود دهد که خود به هزینه می‌افزاید. همچنین مدل ها هرچه به جلوتر می‌روند سنگین تر و پرهزینه تر می‌شوند.
- پیچیدگی: یکپارچه سازی هوش مصنوعی با سیستم مدیریت اطلاعات و رویدادی که در حال حاظر کار می‌کند می‌تواند امری پیچیده باشد. زیرا نیازمند توانمندی هم در زمینه هوش مصنوعی و هم در زمینه مدیریت اطلاعات و رویداد های امنیتی است. هرچند اگر از سرویس دهنده مسلطی استفاده کرد، این امر می‌تواند حل شود.
- کیفیت داده: مدل های هوش مصنوعی به شدت به کیفیت داده بستگی دارند. هرچقدر داده با کیفیت تر باشد مدل عملکرد بهتری دارد. البته این چالش به مرور با ظهور مجموعه داده های جدید مثل LogHub و غیره کمرنگ تر می‌شود.
- توضیح پذیری: مدل های هوش مصنوعی عموما توضیح پذیر نیستند و نمی‌توانند علت یک تصمیم خود را توضیح دهند. این امر در زمینه امنیت اهمیت دوچندانی دارد.البته اخیرا تلاش های قابل توجهی در زمینه هوش مصنوعی توضیح پذیر [17] وجود داشته که با سرمایه گذاری در این حوزه می‌توان احتمال این چالش را کم کرد.

٥ نتیجه گیری

هوش مصنوعی در ابتدا منجر به بهبود سیستم های فعلی هم در بعد خروجی و هم در بعد مصرف منابع شد. ولی در فاز دوم در نتیجه افزایش دقت، امکاناتی مثل پاسخ خودکار برای حالت های مختلف و یا بهبود کد به وجود آمد.

در آینده احتمالا از هوش مصنوعی برای جواب های پیشگیرانه و به طور خاص برای دفاع در برابر حملات مبتنی بر هوش مصنوعی استفاده شود. یعنی از حالت واکنشی به حالت پیشگیرانه بیاید.

البته خود هوش مصنوعی نیز چالش های متعددی دارد که با سرمایه گذاری در زمینه های مختلف می‌توان تاثیر این چالش ها را تا حد خوبی کم کرد.

٦ مراجع

[1] Tao Ban et al. “Breaking Alert Fatigue: AI-Assisted SIEM Framework for Effective Incident Response”. In: *Applied Sciences* 13.11 (2023). ISSN: 2076-3417. DOI: 10.3390/app13116610. URL: https://www.mdpi.com/2076-3417/13/11/6610.

[17] Explainable AI

[5] J. Zhaoxue, L. Tong, Z. Zhenguo, et al. "A Survey On Log Research Of AIOps: Methods and Trends". In: *Mobile Networks and Applications* 26 (2021), pp. 2353–2364. DOI: 10.1007/s11036-021-01832-3. URL: https://doi.org/10.1007/s11036-021-01832-3.

[6] Jieming Zhu et al. "Learning to Log: Helping Developers Make Informed Logging Decisions". In: *Proceedings of the 37th International Conference on Software Engineering (ICSE)*. Vol. 1. 2015, pp. 415–425.

[7] Jieming Zhu et al. " Loghub: A Large Collection of System Log Datasets for AI-driven Log Analytics ". In: *2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE)*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2023, pp. 355–366. DOI: 10.1109/ISSRE59848.2023.00071. URL: https://doi.ieeecomputersociety.org/10.1109/ISSRE59848.2023.00071.

[2] Vinay Dutt Jangampet. "The Rise of The Machines: AI-Driven SIEM User Experience for Enhanced Decision-Making". In: *International Journal of Computer Engineering and Technology (IJCET)* 12.3 (Sept. 2021). Article ID: IJCET_12_03_009, pp. 74–83. ISSN: 0976-6367. URL: https://iaeme.com/Home/issue/IJCET?Volume=12&Issue=3.

[3] Ruchi Mahindru, Harshit Kumar, and Sahil Bansal. "Log Anomaly to Resolution: AI Based Proactive Incident Remediation". In: *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. 2021, pp. 1353–1357. DOI: 10.1109/ASE51524.2021.9678815.

[4] S. R. Pulyala. "From Detection to Prediction: AI-powered SIEM for Proactive Threat Hunting and Risk Mitigation". In: *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 15.1 (Jan. 2024), pp. 34–43.