دوفصلنامه محاسبات و سامانه های توزیع شده
سال پنجم، شماره۲، شماره پیاپی۱۰، صفحه ۴۹ تا ۶۱ ، سال ۱۴۰۱

ارائه مدلی نوین جهت بهبود تشخیص نفوذ در شبکه با استفاده از روش یادگیری ماشین افزایشی در شبکه‌های عصبی spiking در حال تکامل آنلاین

معصومه شیری[۱]، ناصر مدیری[۲]
۱- دانشجو کارشناسی ارشد – مهندسی کامپیوتر گرایش نرم افزار ، دانشگاه آزاد اسلامی واحد زنجان - زنجان - ایران
۲- استادیار-عضو هیات علمی- دانشگاه آزاد اسلامی واحد زنجان - زنجان - ایران

چکیده

تشخیص نفوذ در تحقیقات سیستمهای کامپیوتری با اهمیت خاصی دنبال می‌شود و برای کمک به مدیران امنیتی سیستم در جهت کشف نفوذ و حمله به کار گرفته می‌شود. اهمیت تشخیص ناهنجاری ناشی از این واقعیت است که ناهنجاری در داده‌ها به اطلاعات مهم قابل استفاده در مجموعه‌ی گسترده‌ای از حوزه‌های کاربردی می‌باشد. روش‌های تشخیص نفوذ در بسیاری از دامنه‌های کاربردی مورد استفاده قرار می‌گیرند و هر دامنه نیازمند روش متفاوتی است. در این پژوهش نیز روشی برای بهبود تشخیص نفود در شبکه‌های رایانه‌ای با استفاده از داده‌های جریانی مبتنی بر شبکه عصبی ارائه می‌شود. برای ارائه روش پیشنهادی از شبکه OeSNN-UAD استفاده شده و دارای لایه‌های ورودی و خروجی است که یک نورون خروجی کاندید را برای هر کدام از داده‌های جدید تولید کرده می‌کند. لایه ورودی این شبکه حاوی GRF و نورون‌های ورودی که GRFها برای فیلتر کردن داده‌های ورودی استفاده شده‌اند. در روش پیشنهادی از الگوریتم ELM برای بهبود روند یادگیری شبکه OeSNN-UAD استفاده شده و این الگوریتم با قرارگیری مابین لایه ورودی و خروجی در شبکه OeSNN-UAD ارتباط بین این دولایه را بهبود داده است. شبیه‌سازی روش پیشنهادی در نرم‌افزار MATLAB انجام شد. در آزمایش اول تأثیر ELM در روش پیشنهادی بر اساس معیارهای دقت، بازخوانی، نمره F، BA، MCC روی دسته‌بندی داده‌های مورد بررسی قرار گرفت و در آزمایش دوم تأثیر اندازه پارامتر W_{size} بر عملکرد نهایی روش پیشنهادی بررسی شد که نتایج بهینه مطلوبی نتیجه داد.

کلمات کلیدی: تشخیص نفوذ، شبکه روش یادگیری ماشین افزایشی، شبکه‌های عصبی spiking در حال تکامل آنلاین و الگوریتم ELM.

A new model for Improvement anomaly detection in network by incremental learning machine in Online Evolving Spiking Neural Networks

Masoumeh shiri[1] Nasser modiri[2]
1. Department Engineering computer,Islamic Azad university of Zanjan Branch, Zanjan, Iran.
2. Faculty member, Assistant Professor, Islamic Azad university, Zanjan, Iran.

Abstract

Intrusion detection is followed with special importance in computer systems research and is used to help system security managers to detect intrusion and attack. The importance of anomaly detection is due to the fact that anomalies in data are important information that can be used in a wide range of application areas. Intrusion detection methods are used in many application domains and each domain requires a different method. In this research, a method for improving intrusion detection in computer networks is presented using stream data based on neural network. OeSNN-UAD network is used to present the proposed method and it has input and output layers that produce a candidate output neuron for each new data. The input layer of this network contains GRF and input neurons, which GRFs are used to filter the input data. In the proposed method, the ELM algorithm is used to improve the learning process of the OeSNN-UAD network, and this algorithm has improved the communication between the two layers by being placed between the input and output layers in the OeSNN-UAD network.The simulation of the proposed method was done in MATLAB software. In the first experiment, the effect of ELM in the proposed method was investigated based on the criteria of accuracy, readability, F score, BA, MCC on data classification, and in the second experiment, the effect of the Wsize parameter on the final performance of the proposed method was investigated, and the optimal results It gave a good result.

تاریخچه مقاله:

تاریخ ارسال: ۱۴۰۱/۱۰/۲۴
تاریخ اصلاحات: ۱۴۰۱/۱۱/۲۵
تاریخ پذیرش: ۱۴۰۱/۱۲/۲۹
تاریخ انتشار: ۱۴۰۱/۱۲/۲۹

Keywords:

Intrusion detection, incremental machine learning network, online evolving spiking neural networks and ELM algorithm

* ایمیل نویسنده مسئول:
masoumehshiri1261@gmail.com

۱ – مقدمه

داده‌های جریانی، جریانی با سرعت بالا و پیوسته از داده‌ها هستند (در واقع جریان داده یک دنباله بالقوه نامحدود و منظم از نمونه‌ها است که در طول زمان می‌رسند.). در سال‌های اخیر، پیشرفت در این فناوری، امکان جمع‌آوری مداوم داده‌ها را تسهیل کرده است. نیازهای اولیه در زندگی روزمره مانند استفاده از تلفن یا وبگردی، بازار سهام منجر به ذخیره سازی خودکار داده‌ها می‌شود[1]. پرداختن به جریان داده‌ها مسئله‌ای چالش برانگیز است زیرا به طور مداوم توسط مجموعه ای از دستگاه ها و منابع در قالب های ناهمگن تولید می شود. دریافت اطلاعات در زمان واقعی پردازش شده یکنواخت از جریان داده‌ها دشوار است. یک نمونه از چالش‌های این نوع داده های تشخیص ناهنجاری است. تشخیص ناهنجاری بدون نظارت، در داده‌های جریان یک موضوع تحقیقاتی مهم است. در این سیستم‌های تشخیص دهنده ابتدا به جمع‌آوری داده‌های آموزشی برچسب‌گذاری شده پرداخته می‌شود. این سیستم‌ها داری فرایند یادگیری با نظارت است. مهم است که در این سیستم، آموزش صحیحی نسبت به تشخیص ناهنجاری‌ها داده شود. این ورودی آموزش می تواند برچسب گذاری شده باشد یا خیر. درصورت نداشتن برچسب فرایند کمی پیچیده تر خواهد شد. در این راستا می توان از یک کلاس-بند شبکه عصبی به نام Spiking با متد تکامل آنلاین برای تشخیص ناهنجاری استفاده نمود. ماسیاگ[1] و همکاران یک شبکه عصبی spiking با متد تکامل آنلاین را برای فرایند تشخیص ناهنجاری بدون نظارت[2] پیشنهاد می‌کنند. زیر مجموعه های مختلفی برای Spiking معرفی شده است. در یک نوع آن روش شبکه عصبی آنلاین OeSNN بدون نظارت ارائه شده است[2].

با رشد زیاد داده‌های تولید شده در اینترنت و زمینه‌های دیگر، طبقه‌بندی جریان داده اخیراً توجه گسترده‌ای را برانگیخته است. امروزه، برخی از چالش‌ها در جریان داده‌ها، مانند مفهوم تشخیص جابه‌جایی و کلاس‌بندی جریان داده با نظارت، پیشرفت زیادی کرده‌است. هرچند، زمان مواجه شدن با جریان داده‌های مختلف(شامل دسته‌بندی و مقادیر عددی) یا نمونه‌های برچسب زده شده در دسترس محدود، بسیاری از روش‌های جریان داده‌ها، توانایی دست‌یابی به یک عملکرد رضایت بخش و یا حتی کار کردن را ندارند[3].

در این مقاله روشی برای بهبود تشخیص نفود در شبکه‌های رایانه‌ای با استفاده از داده‌های جریانی مبتنی بر شبکه عصبی ارائه می‌شود که برای این منظور از دو مقاله به عنوان مقالات پایه استفاده می شود که روش ترکیبی از این مقاله ها است. رویکرد پیشنهادی در پژوهش[2] برای تشخیص ناهنجاری بدون نظارت خود را با معماری شبکه‌های OeSNN پیشین که توسط لوبو و همکاران(۲۰۱۸) معرفی شده‌است. بنابراین، معرفی کردن رویکرد پیشنهادی پیشین با یک پیش‌نمایشی از معماری شبکه OeSNN و اصول کلاس‌بندی خودش است[2]. همچنین مقاله جی[3] و همکاران(۲۰۲۰)، پیشنهاد یک روش ماشین یادگیری افراطی نیمه نظارت افزایشی برای کلاس‌بندی جریان داده‌های مختلط[4] را می‌دهند. در واقع این پژوهش به این سوال پاسخ داده می‌شود که: "ارائه مدلی نوین جهت بهبود تشخیص ناهنجاری در شبکه با استفاده از روش یادگیری ماشین افزایشی در شبکه‌های عصبی spikimg در حال تکامل آنلاین" امکان‌پذیر است؟

۲ – پیشینه تحقیق

روش "تشخیص ناهنجاری در زمان واقعی بدون نظارت برای داده های جریانی" ارائه شده است. نویسندگان این مقاله معتقدند که افزایش چشمگیری در دسترس بودن جریان و داده های سری زمانی وجود دارد. این داده‌ها که عمدتاً ناشی از افزایش منابع داده بی‌درنگ متصل هستند، چالش‌ها و فرصت‌های فنی را ارائه می‌کنند. این داده‌ها که عمدتاً ناشی از افزایش منابع داده بی‌درنگ متصل هستند، چالش‌ها و فرصت‌های فنی را ارائه می‌کنند. یکی از قابلیت‌های اساسی برای تجزیه و تحلیل جریان، مدل‌سازی هر جریان به روشی بدون نظارت و تشخیص رفتارهای غیرعادی و غیرعادی در زمان واقعی است. تشخیص زودهنگام ناهنجاری بسیارارزشمند است، با این حال اجرای قابل اعتماد آن در عمل دشوار است. محدودیت های برنامه نیازمند سیستم ها برای پردازش داده‌ها در زمان واقعی است، نه دسته‌ای. جریان داده ها ذاتاً انحراف مفهومی را نشان می دهد و الگوریتم‌هایی را که به طور مداوم یاد می گیرند، ترجیح می دهد. علاوه بر این، تعداد انبوه جریان‌های مستقل در عمل مستلزم آن است که آشکارسازهای ناهنجاری کاملاً خودکار باشند. در این مقاله یک الگوریتم جدید تشخیص ناهنجاری پیشنهاد می‌شود که این محدودیت‌ها را برآورده می‌کند. این تکنیک بر اساس یک الگوریتم حافظه توالی آنلاین به نام حافظه زمانی سلسله مراتبی (HTM) است. همچنین نتایج این مقاله با استفاده از معیار ناهنجاری (NAB) ارائه می‌شود، معیاری که شامل جریان‌های داده‌های دنیای واقعی با ناهنجاری‌های برچسب‌گذاری شده است. این معیار، اولین در نوع خود، یک محیط منبع باز کنترل شده برای آزمایش الگوریتم های تشخیص ناهنجاری در جریان داده‌ها را فراهم می کند. در این مقاله نتایج و تحلیل‌هایی را برای طیف گسترده‌ای از الگوریتم‌ها در این معیار ارائه شده و در مورد چالش‌های آینده برای حوزه نوظهور تحلیل جریان بحث می‌شود[4].

روش "تشخیص ناهنجاری بدون نظارت در داده‌های حسگرهای جریانی با ادغام مدل‌های آماری و یادگیری عمیق" ارائه شده است.

[1] Maciąg et.al.
[2] Online evolving Spiking Neural Network for Unsupervised Anomaly Detection algorithm (OeSNN-UAD).
[3] Ji et.al.
[4] Incremental Semi-supervised Extreme Learning Machine for Mixed data stream classification (MIS-ELM).

نویسندگان این مقاله معتقدند که در عصر کنونی دستگاه‌های هوشمند، که در آن داده‌های عظیمی از حسگرهای متعدد جمع‌آوری می‌شود، نیاز به تشخیص ناهنجاری بدون نظارت قوی در جریان داده‌ها به سرعت در حال افزایش است. این حسگرها وضعیت داخلی یک ماشین، محیط خارجی و تعامل ماشین ها با ماشین های دیگر و انسان را ثبت می‌کنند. استفاده از این اطلاعات به منظور به حداقل رساندن زمان از کار افتادن ماشین‌ها، یا حتی اجتناب کامل از خرابی با نظارت مداوم، از اهمیت بالایی برخوردار است. از آنجایی که هر دستگاه نوع متفاوتی از جریان داده را تولید می کند، معمولاً نوع خاصی از تکنیک تشخیص ناهنجاری بسته به نوع داده، بهتر از سایرین عمل می کند. برای برخی از انواع داده‌ها و موارد استفاده، تکنیک‌های تشخیص ناهنجاری آماری بهتر عمل می‌کنند، در حالی که برای برخی دیگر، تکنیک‌های مبتنی بر یادگیری عمیق ترجیح داده می‌شوند. در این مقاله، یک تکنیک جدید تشخیص ناهنجاری، FuseAD ارائه می‌شود که از هر دو رویکرد آماری و مبتنی بر یادگیری عمیق با ترکیب آنها با هم به شکلی باقیمانده بهره می‌برد. نتایج به‌دست‌آمده نشان‌دهنده افزایش سطح زیر منحنی (AUC) در مقایسه با روش‌های پیشرفته تشخیص ناهنجاری هنگام آزمایش FuseADبر روی یک مجموعه داده عمومی (معیار Webscope یاهو) است. در واقع نتایج به‌دست‌آمده حاکی از آن است که این تکنیک مبتنی بر همجوشی می‌تواند بهترین‌های هر دو جهان را با ترکیب نقاط قوت و تکمیل نقاط ضعف آنها به دست آورد. همچنین این مقاله یک مطالعه فرسایشی برای تعیین کمیت سهم اجزای فردی درFuseAD، یعنی مدل آماری ARIMA و همچنین مدل شبکه عصبی کانولوشنال مبتنی بر یادگیری عمیق (CNN) انجام شده است[5].

روش "یک رویکرد یادگیری عمیق برای تشخیص ناهنجاری بدون نظارت در سری های زمانی" ارائه شده است. نویسندگان این مقاله معتقدند که تکنیک‌های سنتی تشخیص ناهنجاری مبتنی بر فاصله و چگالی قادر به تشخیص ناهنجاری‌های نقطه‌ای مربوط به فصلی و دوره‌ای نیستند که معمولاً در جریان داده‌ها رخ می‌دهند، و شکاف بزرگی در تشخیص ناهنجاری سری‌های زمانی در عصر فعلی اینترنت اشیا ایجاد می‌کنند. برای پرداختن به این مشکل، در این مقاله یک رویکرد جدید تشخیص ناهنجاری مبتنی بر یادگیری عمیق (DeepAnT) برای داده‌های سری زمانی ارائه می‌شود که به همان اندازه برای موارد غیر جریانی قابل استفاده است. DeepAnT قادر به تشخیص طیف گسترده ای از ناهنجاری ها، به عنوان مثال، ناهنجاری های نقطه، ناهنجاری های متنی، و اختلاف در داده های سری زمانی است .برخلاف روش‌های تشخیص ناهنجاری که در آن ناهنجاری‌ها آموخته می‌شوند، DeepAnT از داده‌های بدون برچسب برای ضبط و یادگیری توزیع داده استفاده می‌کند که برای پیش‌بینی رفتار عادی یک سری زمانی استفاده می‌شود . DeepAnT از دو ماژول پیش بینی سری زمانی و آشکارساز ناهنجاری تشکیل شده است. ماژول پیش‌بینی سری زمانی از شبکه عصبی کانولوشن عمیق (CNN) برای پیش‌بینی مهر زمانی بعدی روی خط افقی تعریف‌شده استفاده می‌کند. این ماژول پنجره ای از سری های زمانی (که به عنوان زمینه استفاده می شود) می گیرد و تلاش می کند تا مهر زمانی بعدی را پیش بینی کند. سپس مقدار پیش‌بینی‌شده به ماژول آشکارساز ناهنجاری ارسال می‌شود، که مسئول برچسب گذاری زمان مربوطه به عنوان عادی یا غیرعادی است. DeepAnT را می‌توان حتی بدون حذف ناهنجاری ها از مجموعه داده های داده شده آموزش داد .به طور کلی، در رویکردهای مبتنی بر یادگیری عمیق، داده های زیادی برای آموزش یک مدل مورد نیاز است. در حالی که درDeepAnT ، یک مدل را می توان در مجموعه داده های نسبتاً کوچک آموزش داد و در عین حال به قابلیت های تعمیم خوب به دلیل اشتراک گذاری موثر پارامتر CNN دست یافت. از آنجایی که تشخیص ناهنجاری در DeepAnTبدون نظارت است، در زمان تولید مدل به برچسب‌های ناهنجاری متکی نیست. بنابراین، این رویکرد می‌تواند مستقیماً در سناریوهای زندگی واقعی اعمال شود که در آن‌ها عملاً غیرممکن است که یک جریان بزرگ از داده‌های حاصل از حسگرهای ناهمگن شامل نقاط عادی و غیرعادی برچسب‌گذاری شود. در این مقاله یک ارزیابی دقیق از ۱۵ الگوریتم بر روی ۱۰ معیار تشخیص ناهنجاری، که در مجموع شامل ۴۳۳ سری زمانی واقعی و مصنوعی است، انجام شده است. آزمایش‌ها نشان می‌دهند که DeepAnT در بیشتر موارد از پیشرفته‌ترین روش‌های تشخیص ناهنجاری بهتر عمل می‌کند، در حالی که عملکردی برابر با دیگران دارد[6].

روش "تشخیص ناهنجاری بدون نظارت با شبکه های عصبی LSTM" ارائه شده است. در این روش تشخیص ناهنجاری را در یک چارچوب بدون نظارت بررسی کرده و الگوریتم‌های مبتنی بر شبکه عصبی حافظه کوتاه‌مدت (LSTM) را معرفی می‌کنند. به طور خاص، با توجه به دنباله های داده با طول متغیر، ابتدا این توالی ها را از ساختار مبتنی بر LSTM خود عبور داده و دنباله هایی با طول ثابت به دست می آید. سپس یک تابع تصمیم برای آشکارسازهای ناهنجاری خود بر اساس ماشین‌های بردار پشتیبان یک‌کلاس (OC-SVMs) و الگوریتم‌های توصیف داده‌های برداری پشتیبانی (SVDD) پیدا می‌شود. به عنوان اولین بار در ادبیات، آنها به طور مشترک پارامترهای معماری LSTM و الگوریتم OC-SVM یاSVDD را با استفاده از روش‌های آموزشی مبتنی بر گرادیان و برنامه‌نویسی درجه دوم آموزش داده و بهینه می‌کردند. برای اعمال روش آموزش مبتنی بر گرادیان، معیارهای هدف اصلی الگوریتم‌های OC-SVM و SVDD اصلاح شده، جایی که هم‌گرایی معیارهای هدف اصلاح‌شده با معیارهای اصلی اثبات می‌کنند. آنها همچنین فرمول‌های بدون نظارت خود را به چارچوب‌های نیمه نظارت شده و کاملاً نظارت شده ارائه می دادند. در انتها آنها الگوریتم‌های تشخیص ناهنجاری را به‌دست می‌آورده که می‌توانند توالی داده‌های با طول متغیر را پردازش کنند و در عین حال عملکرد بالایی را به خصوص

برای داده‌های سری زمانی ارائه دهند. رویکرد این روش عمومی است به طوری که با جایگزینی مستقیم ساختار مبتنی بر LSTM با ساختار مبتنی برGRU، این رویکرد برای معماری واحد بازگشتی دروازه‌ای (GRU)نیز اعمال می‌شود. آنها در آزمایش‌های خود، دستاوردهای عملکرد قابل توجهی را که توسط الگوریتم‌هایشان با توجه به روش‌های مرسوم به دست آمده است، نشان می‌دهند[7].

روش " تشخیص ناهنجاری آنلاین از طریق جریان های داده بزرگ" ارائه شده است. آنها معتقدند که در بسیاری از حوزه‌ها، داده‌های با کیفیت بالا به عنوان پایه ای برای تصمیم گیری استفاده می شوند. یک جزء ضروری برای ارزیابی کیفیت داده ها در تشخیص ناهنجاری نهفته است. در این مقاله طراحی و اجرای یک چارچوب برای آزمایش کیفیت داده بر روی جریان‌های دنیای واقعی در یک شبکه مخابراتی در مقیاس بزرگ توصیف و ارزیابی تجربی شده است. این رویکرد هم عمومی بوده و با استفاده از معیارهای همه منظوره وام گرفته شده از تئوری اطلاعات و آمار و هم مقیاس پذیر از طریق خطوط لوله تشخیص ناهنجاری که در یک محیط توزیع شده بر روی زیرساخت‌های پیشرفته جریان داده های بزرگ و پردازش دسته‌ای اجرا می شوند. آنها به طور تجربی سیستم خود را ارزیابی کرده و با مقایسه آن با تکنیک‌های تشخیص ناهنجاری موجود، با نشان دادن دقت، کارایی بالا و همچنین مقیاس‌پذیری آن در موازی‌سازی عملیات در تعداد زیادی از گره‌ها، مزایا و محدودیت‌های آن را مورد بحث قرار دادند[8].

روش " تشخیص ناهنجاری نظارت شده در جریان های داده شبه دوره ای نامشخص" ارائه شده است. نویسندگان این مقاله معتقدند که جریان های داده نامشخص به طور گسترده در بسیاری از برنامه های کاربردی وب ایجاد شده است. عدم قطعیت در جریان داده‌ها، تشخیص ناهنجاری از جریان داده های حسگر را بسیار چالش برانگیزتر می‌کند. در این مقاله، یک چارچوب جدید ارائه شده که از تشخیص ناهنجاری در جریان‌های داده نامشخص پشتیبانی می‌کند. چارچوب پیشنهادی از روش آستانه نرم موجک برای حذف نویزها یا خطاها در جریان داده استفاده می کند. این روش بر اساس جریان داده‌های تصفیه شده، تکنیک‌های تشخیص الگوی دوره موثر و استخراج ویژگی را برای بهبود کارایی محاسباتی توسعه می دهد. آنها از روش‌های طبقه بندی برای تشخیص ناهنجاری در جریان داده های اصلاح شده استفاده کردند و همچنین به طور تجربی نشان دادند که رویکرد پیشنهادی دقت بالایی در تشخیص ناهنجاری در چندین مجموعه داده واقعی نشان می‌دهد[9].

روش "ارائۀ یک یادگیرندۀ برخط رانش آگاه برای تشخیص ناهنجاری در داده‌های جریانی" ارائه شده است. نویسندگان این مقاله معتقدند داده‌های جریانی در بستر پویا و در حال تغییر، تکامل می‌یابند؛ بنابراین، رانش مفهوم یا تغییر توزیع اساسی داده‌ها با گذشت زمان، یکی از مهم‌ترین چالش‌های این نوع از داده‌ها است. علاوه بر این، رانش مفهوم بر عملکرد فرآیند تشخیص ناهنجاری نیز تأثیر می‌گذارد .تشخیص ناهنجاری در چنین داده‌هایی در حوزه‌های متعددی ازجمله تشخیص نفوذ در شبکه‌های کامپیوتری یا مدیریت ترافیک شبکۀ راهها کاربرد دارد. در سال‌های اخیر، رویکردهایی مبتنی بر تجزیۀ تانسور ارائه شده‌اند که به‌صورت برخط زیرفضا را ردیابی می‌کنند و یادگیرنده را با یک استراتژی ناآگاهانه و به‌طور ضمنی در همۀ گام‌های زمانی، در مقابل تغییرات تطبیق می‌دهند. این مقاله، یک رویکرد برخط را پیشنهاد می‌کند که رانش مفهوم را به‌طور صریح تشخیص می‌دهد و اعلام می‌کند. بدین ترتیب یادگیرنده نیز با یک استراتژی آگاهانه و تنها در گام‌های زمانی لازم با تغییرات و رانش، تطبیق پیدا می‌کند. ارزیابی راهکار پیشنهادی با استفاده از مجموعه‌داده‌های واقعی انجام شد و تحلیل نتایج به‌دست‌آمده، عملکرد روش پیشنهادی را از جنبه‌های یادگیری و تشخیص تأیید می‌کند[10].

روش " ارائه یک مدل بهینه یادگیری عمیق جهت تشخیص ناهنجاری در شبکه های مبتنی بر جریان " ارائه شده است. آنها معتقدند که همگام با توسعه اینترنت و افزایش روز افزون کاربران، محافظت از داده ها در برابر تهدیدات امنیتی مسئله بسیار جدی محسوب میشود و سیستم های تشخیص نفوذ، نقش مهمی در تضمین امنیت اطلاعات دارد . در این مقاله مدلی کارآمد بر مبنای یادگیری عمیق با هدف بهبود دقت و سرعت تشخیص ناهنجاری در شبکه های مبتنی بر جریان ارائه شده است . به دلیل اهمیت ترتیب داده ها در ترافیک شبکه از شبکه‌های عصبی بازگشتی RNN استفاده شده تا ترتیب داده ها حفظ شوند . الگوریتم های RNN RNN DGRUو DLSTM ، SimpleRNNاست. به منظور استفاده از داده های جریانیِ شبکه هایِ متفاوت، از مجموعه داده های NSL-KDD استفاده شده تا به کمک آن عملکرد مدل ارز ی ابی شود . همچن ین عملکرد مدل با روش های سنتی یادگیری ماشین که شامل درخت تصمیم Random Forest و Naïve Bayes ، SVM، J48 مقایسه شده است. در نهایت با مقایسه روشهای موجود و نتایج به دست آمده مدل DGRU-RNN عملکرد بهتری دارد[11].

روش "بهبود تشخیص نفوذ در شبکه با شناسایی ویژگی های موثر بر پایه الگوریتم های تکاملی و دسته بند ماشین بردار پشتیبان " ارائه شده است. نویسندگان این مقاله معتقدند که روند رو به رشد استفاده از اینترنت و وجود نقاط آسیب پذیر در شبکه، استفاده از سیستم های تشخیص نفوذ را به عنوان یکی از مهم ترین عناصر برقراری امنیت درخور توجه قرار داده است. تشخیص نفوذ در اصل مسئله دسته بندی است و شناسایی ویژگی های موثر ازجمله موضوعات با اهمیت در دسته بندی است. در این مقاله یک روش جدید برای انتخاب ویژگی های موثر در تشخیص نفوذ در شبکه، مبتنی بر الگوریتم تخمین توزیع ارائه شده است که از درخت وابستگی احتمالاتی برای شناسایی تعاملات بین ویژگی ها استفاده می کند. به منظور ارزیابی عملکرد این الگوریتم از مجموعه داده NSL-KDD استفاده شده است که در آن، بسته ها به پنج دسته نرمال و نفوذهای نوعDOS ، U2R، R2Lو Prob تقسیم شده اند.

عملکرد الگوریتم ارائه شده به تنهایی و به صورت ترکیبی با سایر الگوریتم های انتخاب ویژگی، مانند انتخاب پیشرو، انتخاب پسرو و الگوریتم ژنتیک، مقایسه و تاثیر پارامترهای الگوریتم، مانند اندازه جمعیت بر میزان دقت تشخیص نفوذ بررسی شده است .براساس نتایج حاصل از این تحلیل و نیز ترکیب نتایج بررسی میزان دقت درون دسته ای حاصل از به کارگیری الگوریتم های انتخاب ویژگی متفاوت، زیرمجموعه ای از ویژگی های موثر در تشخیص نفوذ شناسایی شده است[12].

روش "تشخیص نفوذ در شبکه‌های کامپیوتری به کمک الگوریتم ژنتیک" ارائه شده است. نویسندگان این مقاله معتقدند که سیستم تشخیص نفوذ یک سیستم محافظتی است که خرابکاری های در حال وقوع در شبکه را شناسایی می کند. در این سیستم ها بااستفاده اطلاعاتی مانند پویش پورتها و تشخیص ترافی غیر متعارف، نفوذخرابکاری ها را می توان کشف و به مسئول شبکه گزارش داد. برایآنکه سیستم های تشخیص نفوذ قدرت کشف و تشخیص نفوذهای از قبل تعریف نشده را داشته باشند، به نوعیهوشمندی نیاز دارند. در اینحالت، این سیستم ها قابلیت یادگیری دارند و می توانند بر روی بسته های وارد شده به شبکه تحلیل انجام داده و کاربران عادی و غیر عادی را تشخیص دهند. روش های هوشمند متداول شامل شبکه های عصبی؛ منطق فازی؛ تکنی های داده کاوی و الگوریتم های ژنتیک می باشند. در این مقاله با تشریح الگوریتم ژنتی و ارتباط ان با سیستم های تشخیص نفوذ، پیاده سازی این سیسکتم هکا را به کمک الگوریتم ژنتیک بیان شده و مکانیزم هایی به منظور بهبود عملکرد سیستم های تشخیص نفوذ توسط این الگوریتم بیان می گردد. علاوه بر این مکانیزم ها الگوریتم تطبیقRBNDMبا استفاده از مجموعه دادهKDD99 ارزیابی می شود. این آزمایش نشان می‌دهد که این مدل می تواند نرخ مثبت و واقعیIDSرا افزایش دهد[14].

۳ – راهکار پیشنهادی

کشف ناهنجاری بدون نظارت در داده‌های جریان یک موضوع تحقیقاتی است که کاربردهای عملی مهمی دارد. به عنوان مثال، یک مدیر سیستم اینترنت ممکن است علاقه‌مند به تشخیص فعالیت غیرعادی بالا در یک صفحه وب باشد که بالقوه ناشی از حمله هکر است. استفاده غیرمنتظره از یک واحد CPU در یک سیستم کامپیوتری می-تواند نمونه دیگری از رفتار غیرعادی باشد که ممکن است نیاز به بررسی داشته باشد. تشخیص و طبقه‌بندی صحیح چنین ناهنجاری‌هایی ممکن است بهینه‌سازی عملکرد سیستم کامپیوتری را ممکن سازد. با این حال، در بسیاری از موارد، جمع‌آوری داده‌های آموزشی کافی با ناهنجاری‌های برچسب‌گذاری شده برای یادگیری نظارت‌شده یک تشخیص ناهنجاری به منظور استفاده از آن بعداً برای شناسایی ناهنجاری‌های واقعی در داده‌های جریان آسان نیست. بنابراین، طراحی تشخیص ناهنجاری که می‌توانند به درستی ناهنجاری‌ها را بدون دسترسی به داده‌های آموزشی برچسب‌گذاری‌شده شناسایی کنند، بسیار مهم است. علاوه بر این، از آنجایی که مشخصه یک جریان داده ورودی ممکن است در حال تغییر باشد، آشکارساز ناهنجاری باید در حالت آنلاین یاد بگیرد[2]. در این فصل یک مدل بهبود تشخیص ناهنجاری در شبکه با استفاده از روش یادگیری ماشین افزایشی در شبکه‌های عصبی spiking در حال تکامل آنلاین ارائه می‌شود. در واقع روش پیشنهادی دارای مراحل زیر است و فلوچارت روش پیشنهادی در شکل ۱ نمایش داده شده است.

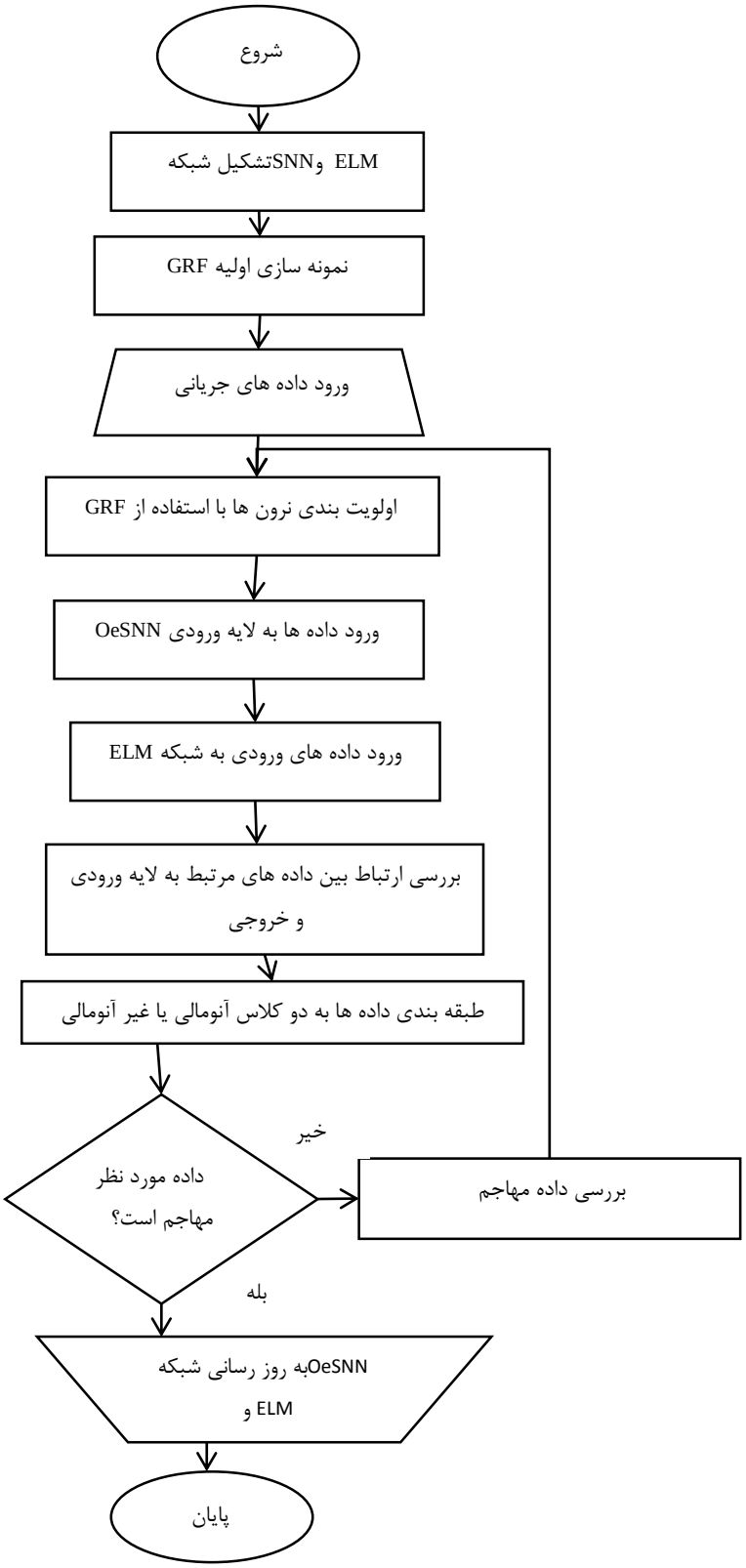
- تشکیل شبکه SNN وELM
- نمونه سازی اولیه GRF
- ورود داده های جریانی
- الویت بندی نرون ها با استفاده از GRF
- ورود داده‌ها به لایه ورودی OeSNN
- ورود داده های ورودی به شبکه ELM
- بررسی ارتباط بین داده های مرتبط به لایه ورودی و خروجی
- طبقه‌بندی دادها به دو کلاس آنومالی و غیر آنومالی

شبکه‌های OeSNN، دارای لایه‌های ورودی و خروجی است و از آن جهت طبقه‌بندی داده‌های جریانی استفاده می‌گردد. با توجه به حجم محاسباتی بالاتر شبکه های عصبی در زمان ورود داده‌های حجیم و مشکلاتی مانند کمبود حافظه سبب شده است تا شبکه‌هایی مانند SNN ایجاد گردند. این شبکه دارای توابع گوسی[1] و نورون‌های ورودی می‌باشد. ازGRF، جهت فیلتر نمودن داده های ورودی برای کاهش حجم محاسبات و انتخاب بهینه نرون ورودی داده استفاده می شود. خروجی این فیلتر پارامتر firing ایجاد می نماید که در واقع برای مقداردهی اولیه وزن سیناپس‌های بین هر نورون ورودی و نورون خروجی کاندید شده استفاده می‌شود. درنهایت نمونه داده‌های ورودی طبقه‌بندی می‌شوند[2].

طبقه‌بندی در OeSNN با محاسبه کردن مقادیر مربوط به پتانسیل پس‌سیناپسی[2] در نورون‌های خروجی به دست می‌آیند. برای هرکدام از نمونه‌های ورودی، ابتدا GRFها ایجاد شده وسپس برای محاسبه کردن ترتیب‌بندی نورون‌های ورودی استفاده می‌شوند. سپس با محاسبه مقادیر مربوط به PSP آن آستانه‌ای که بیش از حد آستانه تعیین شده قرار بگیرد به عنوان کلاس تصمیم‌گیری نمونه ورودی در نظر گرفته می‌شود[2].

[1] Gaussian Receptive Fields(GRFs)

[2] Post- synaptic potential(PSP)



شکل۱: فلوچارت روش پیشنهادی

در ادامه توضیحات مطرح شده، سه ماژول تولید مقادیر خروجی نورون‌های خروجی کاندید، طبقه‌بندی آنومالی و تصحیح مقدار به فرایند اضافه می‌گردد. در ابتدا، فاز اول که برای تشخیص آنومالی است، پنجره w به روز رسانی شده و سپس GRF با بررسی ورودی به اولویت بندی نرون ها می پردازد. از مقدار w برای محاسبه کردن زمان و ترتیب firing در نورون‌های ورودی در لایه ورودی استفاده می-شود[2].

مقادیر ورودی x_t به دو گروه آنومالی و غیر- آنومالی تقسیم-بندی می‌شوند. نورون های لایه خروجی $n_f \in NO$ که در اول fire شده اند محاسبه می شوند. اگر هیچ یک از نورون های خروجی fire نشود، مقدار x_t به عنوان آنومالی کلاس‌بندی می‌شود. در غیراینصورت، مقدار خروجی نورون خروجی که برای اولین بار fire شده‌است به عنوان یک مقدار خروجی جدید y_t گزارش می‌شود. در نهایت، ماژول طبقه بندی آنومالی مقدار ورودی x_t را با استفاده از خطای پیش بینی که تفاوت مطلق بین x_t و y_t است محاسبه نموده با یک مقدار آستانه بررسی می نماید و درصورت مشاهده تفاوت زیاد به صورت آنومالی گزارش می دهد[2].

در این بخش، نورون‌های خروجی کاندید جدید توسط مقدار x_t ایجاد و مقداردهی اولیه می‌شوند. مقداردهی اولیه روند n_c در سه مرحله انجام می‌شود، اول، سیناپس‌ها ارتباطی را بین n_c با تمام نورون-های ورودی در NI برقرار می‌کند. سپس وزن سیناپس‌ها به دست می-آیند. زمان به روزرسانی (τ_{n_c})، نورون‌های خروجی کاندید را برابر با مقدار t در نظر می‌گیرد و در نهایت، واحد تولید مقادیر نورون‌های خروجی کاندید، مقدار خروجی اولیه v_{n_c} در نورون های لایه خروجی را مشخص کرده و مقدار شمارنده M_{n_c} را برابر با ۱ می‌کند. اگر x_t به عنوان غیر آنومالی شناخته شود، خروجی اولیه v_{n_c} تولید شده از نورون خروجی کاندید توسط مقدار ماژول تصحیح، اصلاح خواهد شد[2].

به منظور بهبود روند آموزش OeSNN از الگوریتم ELM استفاده می‌شود. این الگوریتم، دارای یک شبکه عصبی پیشخور به همراه یک لایه پنهان است و وزن‌های بین لایه پنهان و لایه خروجی با استفاده از یک راه‌حل تحلیلی مسأله بهینه‌سازی خطی آموزش می‌دهد. سرعت این الگوریتم به دلیل آنکه همانند شبکه عصبی مصنوعی وزن‌ها را به طور مکرر تنظیم نمی‌کند، بالاتر است و همچنین دقت آن با توجه به آموزش طراحی شده نسبت به SNN بیشتر خواهد بود[2].

در این مقاله به منظور بهبود روش پیشنهادی، از ترکیب روش های الگوریتم ELM و SNN استفاده می‌شود. در واقع الگوریتم ELM مابین لایه ورودی و خروجی در شبکه OeSNN-UAD، به منظور بهبود ارتباط بین لایه ورودی و خروجی استفاده می‌شود.

تشخیص آنومالی در داده‌های جریانی اهمیت زیادی دارد چرا که وجود این داده‌های مخرب در میان حجم انبوهی از داده‌ها می‌تواند مشکلات زیادی را به وجود آورد. به همین منظور استفاده از روش‌هایی که بتوان با کمک آنها این داده‌های مخرب را تشخیص داد مورد نیاز می‌باشد. یکی از این روش‌ها شبکه‌های OeSNN-UAD است. به منظور بهبود تشخیص آنومالی در داده‌های جریانی از روش OeSNN-UAD به

همراه روش ELM استفاده می‌شود. از الگوریتم ELM ما بین لایه ورودی و خروجی در شبکه OeSNN-UAD استفاده می‌شود که موجب شده تا ارتباط بین دو لایه ورودی و خروجی در این شبکه بهبود پیدا کند.

۴- نتایج و تحلیل شبیه‌سازی

شبکه‌های OeSNN، دارای لایه‌های ورودی و خروجی هستند. تعداد لایه‌های خروجی در این شبکه به دلیل الزماتی مانند پردازش حجم بزرگی از داده‌های ورودی و مشکل حافظه که به طبقه‌بندی مقادیر داده‌های جریانی مرتبط است، محدودتر است. این شبکه یک نورون خروجی کاندید برای هر کدام از داده‌های جدید تولید کرده است که مقادیر مربوط به این نورون‌ها یا به مخزن خروجی منتقل شده و یا با یکی از نورون‌های مشابهی که در مخزن وجود دارد ادغام شده‌است. شبکه OeSNN قابلیت کنترل کردن اندازه مخزن را دارد. به این صورت که اگر مخزن پر باشد قدیمی‌ترین نورون موجود در مخزن را حذف می‌کند با اینکار به نورون‌های جدید اجازه می‌دهد تا به مخزن ورود کنند. لایه ورودی در این شبکه، حاوی GRF و نورون‌های ورودی است. از این GRFها برای فیلتر کردن داده‌های ورودی استفاده شده‌است تا با استفاده از آن بتوان حجم محاسبات را کاهش داده و نرون ورودی را به صورت بهینه انتخاب کرد. خروجی این مرحله تولید کردن پارامتر firing می‌باشد و از آنها برای مقداردهی اولیه وزن سیناپس‌ها بین هر نورون ورودی و خروجی کاندید شده استفاده شده‌است. در نهایت، نمونه داده-های ورودی طبقه‌بندی شده‌اند. طبقه‌بندی در این شبکه، از محاسبه کردن مقادیر مربوط به پتانسیل پس سیناپسی[1] در نورون‌های خروجی به دست آمده است. این مهم به اینصورت انجام شده‌است که برای هر کدام از نمونه‌های ورودی، ابتدا GRFها ایجاد می‌شوند سپس از آنها به منظور محاسبه کردن ترتیب‌بندی نورون‌های ورودی استفاده شده‌است. در ادامه، در نورون خروجی با رمزگذاری کردن مقادیر مربوط به PSP مقدار آن به روزرسانی می‌شود. در ادامه اگر PSP به دست آمده از یک کلاس تصمیم‌گیری که به نورون خروجی اختصاص پیدا کرده است بیش از حد آستانه تعیین شده PSP به دست بیاید به عنوان یک کلاس تصمیم‌گیری نمونه ورودی در نظر گرفته شده‌است. به منظور بهبود دادن روند آموزش OeSNN از الگوریتم ELM استفاده شده‌است. الگوریتم ELM یک شبکه عصبی حاوی یک لایه پنهان است که به‌طور تصادفی به لایه‌های ورودی و پنهان این الگوریتم وزن‌هایی را اختصاص داده است. در ادامه، این وزن‌ها ما بین لایه پنهان و لایه خروجی با استفاده کردن از یک راه‌حل تحلیلی مسأله آموزش خطی آموزش داده شده‌است. این الگوریتم با توجه به اینکه وزن‌ها را دائما تنظیم نمی‌کند از سرعت بالاتری برخوردار است. استفاده از این الگوریتم، منجر شده‌است تا ارتباط بین این دو لایه بهبود پیدا کند. در ادامه بلوک دیاگرام روند پژوهش نیز آورده شده‌است.

شکل ۲: روند پژوهش.

مجموعه داده: در این پژوهش از مجموعه داده معیار آنومالی Numeta[2] که در سایت Kaggle موجود است استفاده شده‌است. این مجموعه داده حاوی ۷ گروه از مجموعه داده‌های مصنوعی و واقعی است که در هر کدام از آنها حاوی چند فایل است. هر کدام از این فایل‌های داده دارای پسوند CSV هستند که حاوی دو ستون است یک ستون حاوی اطلاعات مربوط به Timestamp، و ستون بعدی حاوی مقادیری به نام Value می‌باشد. این مجموعه داده حاوی ۵۸ فایل داده است. تمامی سری‌های زمانی در این مجموعه داده نامتوازن هستند. طبقه‌بندی فایل داده‌ها به صورت زیر می‌باشد.

پارامترهای پژوهش: در این پژوهش به منظور اجرای OeSNN پارامترهای مناسب مشخص شده‌اند که همه در جدول ۳ نمایش و توضیح داده شده است. در پارامتر W_{size} و ε دارای بیشترین تأثیر در تشخیص آنومالی هستند. به همین منظور، انتخاب این دو مقدار در این روش بهینه شده‌است.

[1] Post- synaptic potential(PSP)

[2] https://www.kaggle.com/datasets/boltzmannbrain/nab

معیارهای پژوهش: در این پژوهش از ۵ معیار استفاده شده‌است. که این معیارها شامل دقت[1]، یادآوری[2]، F-score، دقت متوازن[3] و ضریب همبستگی متیوز[4] می‌شوند. تعاریف مربوط به این معیارها در جدول به شرح زیر است:

دقت: این معیار نشان دهنده آن است که چند مورد از مقادیر ورودی که به عنوان آنومالی تشخیص داده شده‌اند به صورت آنومالی در فایلهای داده‌ها برچسب خورده‌اند.

$$Precision = \frac{|TP|}{|TP| + |FP|}$$

بازخوانی: این معیار نشان دهنده آن است که چه مقدار از داده‌های موجود که آنومالی هستند توسط مدل به درستی برچسب خورده‌اند.

$$Recall = \frac{|TP|}{|\mathrm{TP}| + |\mathrm{FN}|}$$

F-score: این معیار نشان دهنده میانگین هارمونیک دو معیار دقت و یادآوری است

$$F1 = 2.\left(\frac{Precision.Recall}{Precision + Recall}\right)$$

دقت متوازن: به دلیل نامتوازن بودن مجموعه داده استفاده شده، از معیار BA استفاده شده‌است. این معیار به عنوان میانگین یادآوری شناخته می‌شود و می‌توان آن را معادل مقدار یادآوری در نظر گرفت که بر حسب گروه ورودی‌های غیر آنومالی محاسبه می‌شود. از این معیار بر روی مجموعه داده‌های نامتوازن استفاده می‌شود که از جمله آنها می‌توان به سری‌های زمانی در مجموعه داده معرفی شده اشاره کرد

$$BA = \frac{1}{2}.\left(\frac{|TP|}{|\mathrm{TP}|+|\mathrm{FN}|} + \frac{|TN|}{|\mathrm{TN}|+|\mathrm{FP}|}\right)$$

ضریب همبستگی متیوز: معیار MCC نشان‌دهنده ضریب همبستگی بین برچسب‌های واقعی و برچسب‌های پیش‌بینی شده‌است که بر روی مقادیر ورودی آنومالی و غیر آنومالی تعریف شده‌است

$$MCC = \frac{|\mathrm{TP}| \cdot |\mathrm{TN}| - |\mathrm{FP}| \cdot |\mathrm{FN}|}{\sqrt{(|\mathrm{TP}| + |\mathrm{FP}|)(|\mathrm{TP}| + |\mathrm{FN}|)(|\mathrm{TN}| + |\mathrm{FP}|)(|\mathrm{TN}| + |\mathrm{FN}|)}}$$

در فرمول آورده شده برای دو معیار BA و MCC، $|TP|$ نشان‌دهنده مثبت واقعی است که به معنی آن است که تعداد مقادیر ورودی توسط مدل تشخیص داده شده و برچسب‌گذاری شده‌اند. $|FP|$ نشان دهنده مثبت کاذب است که به معنی تعداد ورودی‌هایی است که به صورت آنومالی تشخیص داده شده‌اند اما برچسب آنومالی نخورده‌اند. مقدار $|\mathrm{FN}|$ که به معنی منفی کاذب است، تعداد مقادیر ورودی را نشان می‌دهد که در آن داده‌ها به عنوان آنومالی برچسب‌گذاری شده، اما به عنوان آنومالی طبقه‌بندی نشده‌اند. مقدار $|\mathrm{TN}|$ به معنی تعداد مقادیر ورودی است که در آن داده‌ها به عنوان آنومالی برچسب‌گذاری و به عنوان آنومالی طبقه‌بندی شده‌اند.

جدول۱: پارامترهای شبیه سازی

ردیف	پارامتر	مقادیر	توضیحات
۱	NI_{size}	۱۰	تعداد نورون های ورودی
۲	NO_{size}	۵۰	حداکثر تعداد نورون های خروجی
۳	mod	./۶	ضریب مدولاسیون[5]
۴	C	./۶	کسری از PSP^{max} مورد نیاز برای شلیک یک نورون خروجی
۵	sim		مقدار آستانه برای شباهت بین وزن بردار های یک نورون خروجی کاندید و یک نورون خروجی، که برای ادغام این نورون ها لازم است
۶	ξ	./۹	فاکتور تصحیح مقدار خروجی
۷	W_{size}	{100,200, …,60	اندازه پنجره
۸	ε	{2,3, …,7}	فاکتور طبقه بندی آنومالی
۹	β و TS		نیازی به تعیین این دو پارامتر نمی‌باشد، چرا که نتیجه طبقه‌بندی به این دو پارامتر بستگی ندارد.

۵- نتایج پژوهش

در این بخش، نتایج به دست آمده از شبیه‌سازی روش پیشنهادی که توسط نرم‌افزار متلب به دست آمده، آورده شده‌است.

آزمایش اول: بررسی تأثیر ELM در روش پیشنهادی در این آزمایش تمامی داده‌ها مورد آزمایش قرار گرفته اند. معیارهایی نیز که در این فصل اشاره شده است برای بررسی عملکرد تاثیر ELM در روش پیشنهادی مورد آزمایش قرار گرفتند. در جدول زیر معیارهای محاسبه شده نشان داده شده‌است. با توجه به اینکه ELM نیاز به آموزش دارد این آزمایش ۱۰ بار تکرار شد و نتایج آن بصورت میانگین اعلام شده است.

الف) بررسی معیار Precision: در این قسمت نتایج مربوط به معیار دقت جهت مقایسه روش پیشنهادی و روش پایه در انواع مختلف دسته‌بندی داده‌ها در جدول ۵ ارائه شده است. همانطور که نتایج آزمایش (الف) نشان می‌دهد مقایسه بین روش پایه و روش پیشنهادی انجام شده است. در دسته‌بندی داده‌های artificialNoAnomaly روش حاضر نسبت به روش پایه به میزان ۴،۸۱ درصد، در داده‌های artificialWithAnomaly به میزان ۲/۲۴ درصد، در داده‌های realAdExchange به میزان ۲/۳۸ درصد، در داده‌های realAWSCloudwatch به میزان ۱/۰۹ درصد، در داده‌های

[1] Precision
[2] Recall
[3] Balanced accuracy(BA)
[4] Matthews correlation coefficient(MCC)
[5] Modulation Factor

realKnownCauses به میزان ۲/۲۹ درصد، در داده‌های realTraffic به میزان ۳/۳۷ درصد و در داده‌های realTweets به میزان ۵/۸۱ درصد عملکرد بهتری از خود نشان داده است. در تمامی این داده‌ها به دلیل خطای کمتر تشخیص که توسط ELM ایجاد شده‌است روش حاضر عملکرد بهتری از خود نشان می‌دهد. در مجموع می‌توان گفت روش حاضر به میزان ۳/۱۴ درصد نسبت به روش پایه دقت بهتری از خود نشان داده است.

جدول ۲: مقایسه نتایج بررسی معیار precision بر روی روش مقاله پایه و روش پیشنهادی.

دسته بندی داده ها	نتایج روش پایه	نتایج استخراجی از روش پیشنهادی
artificialNoAnomaly	۰/۸۳	۰/۸۶
artificialWithAnomaly	۰/۸۹	۰/۹۱
realAdExchange	۰/۸۴	۰/۸۶
realAWSCloudwatch	۰/۹۱	۰/۹۲
realKnownCauses	۰/۸۷	۰/۸۹
realTraffic	۰/۸۹	۰/۹۲
realTweets	۰/۸۶	۰/۹۱

ب) بررسی معیار Recall: در این قسمت نتایج مربوط به معیار بازخوانی جهت مقایسه روش پیشنهادی و روش پایه در انواع مختلف دسته‌بندی داده‌ها در جدول ۳ ارائه شده است. همانطور که نتایج آزمایش (ب) نشان می‌دهد مقایسه بین روش پایه و روش پیشنهادی انجام شده است. در دسته بندی داده‌های artificialNoAnomaly روش حاضر نسبت به روش پایه به میزان ۱۴/۲۸ درصد، در داده‌های artificialWithAnomaly به میزان ۳۷/۰۳ درصد، در داده‌های realAdExchange به میزان ۱۷/۶۴ درصد، در داده‌های realAWSCloudwatch به میزان ۱۲/۵۰ درصد، در داده‌های realKnownCauses به میزان ۱۱/۷۶ درصد، در داده های realTraffic به میزان ۵/۵۵ درصد و در داده‌های realTweets به میزان ۲۱/۷۳ درصد عملکرد بهتری از خود نشان داده است. در تمامی این داده‌ها به دلیل خطای کمتر تشخیص که توسط ELM ایجاد شده‌است روش حاضر عملکرد بهتری از خود نشان می‌دهد. در مجموع می‌توان گفت روش حاضر به میزان ۱۷/۲۱ درصد نسبت به روش پایه یادآوری بهتری از خود نشان داده است.

جدول ۳: مقایسه نتایج بررسی معیار Recall بر روی روش مقاله پایه و روش پیشنهادی.

دسته بندی داده ها	نتایج روش پایه	نتایج استخراجی از روش پیشنهادی
artificialNoAnomaly	۰/۲۱	۰/۱۸
artificialWithAnomaly	۰/۲۷	۰/۱۷
realAdExchange	۰/۱۷	۰/۱۴
realAWSCloudwatch	۰/۱۶	۰/۱۴
realKnownCauses	۰/۱۷	۰/۱۵
realTraffic	۰/۱۸	۰/۱۷
realTweets	۰/۲۳	۰/۱۸

ج) **بررسی معیار F-score**: در این قسمت نتایج مربوط به معیار F-scope جهت مقایسه روش پیشنهادی و روش پایه در انواع مختلف دسته‌بندی داده‌ها در جدول ۴ ارائه شده است. همانطور که نتایج آزمایش (ج) نشان می‌دهد مقایسه بین روش پایه و روش پیشنهادی انجام شده است. در دسته‌بندی داده‌های artificialNoAnomaly روش حاضر نسبت به روش پایه به میزان ۱۴/۲۸ درصد، در داده‌های artificialWithAnomaly به میزان ۳۷/۰۳ درصد، در داده‌های realAdExchange به میزان ۱۷/۶۴ درصد، در داده‌های realAWSCloudwatch به میزان ۱۲/۵۰ درصد، در داده‌های realKnownCauses به میزان ۱۱/۷۶ درصد، در داده‌های realTraffic به میزان ۵/۵۵ درصد و در داده‌های realTweets به میزان ۲۱/۷۳ درصد عملکرد بهتری از خود نشان داده است. در مجموع می‌توان گفت روش حاضر به میزان ۱۷/۲۱ درصد نسبت به روش پایه یادآوری بهتری از خود نشان داده است.

جدول ۴: مقایسه نتایج بررسی معیار F-score بر روی روش مقاله پایه و روش پیشنهادی.

دسته بندی داده ها	نتایج روش پایه	نتایج استخراجی از روش پیشنهادی
artificialNoAnomaly	۰/۰۸	۰/۰۷
artificialWithAnomaly	۰/۰۴	۰/۰۳
realAdExchange	۰/۰۸	۰/۰۶
realAWSCloudwatch	۰/۰۵	۰/۰۳
realKnownCauses	۰/۰۶	۰/۰۴
realTraffic	۰/۰۷	۰/۰۶
realTweets	۰/۰۸	۰/۰۶

د) بررسی معیار *BA*: در این قسمت نتایج مربوط به معیار BA جهت مقایسه روش پیشنهادی و روش پایه در انواع مختلف دسته-بندی داده‌ها در جدول ۵ ارائه شده است. همانطور که نتایج آزمایش (د) نشان می‌دهد در دسته بندی داده‌های artificialNoAnomaly روش حاضر نسبت به روش پایه به میزان ۷/۶۹ درصد، در داده های artificialWithAnomaly به میزان ۱۲/۵۰ درصد، در داده‌های realAdExchange به میزان ۵/۷۶ درصد، در داده های realAWSCloudwatch به میزان ۱۲/۵۰ درصد، در داده‌های realKnownCauses به میزان ۵/۳۵ درصد، در داده‌های realTraffic به میزان ۸/۷۷ درصد و در داده‌های realTweets به میزان ۳/۳۸ درصد عملکرد بهتری از خود نشان داده است. در مجموع می‌توان گفت روش حاضر به میزان ۷/۹۹ درصد نسبت به روش پایه دقت متوازن بهتری از خود نشان داده است.

جدول ۵: مقایسه نتایج بررسی معیار BA بر روی روش مقاله پایه و روش پیشنهادی.

دسته بندی داده ها	نتایج روش پایه	نتایج استخراجی از روش پیشنهادی
artificialNoAnomaly	۰/۵۲	۰/۴۸
artificialWithAnomaly	۰/۵۶	۰/۴۹
realAdExchange	۰/۵۲	۰/۴۹
realAWSCloudwatch	۰/۶۱	۰/۵۷
realKnownCauses	۰/۵۶	۰/۵۳
realTraffic	۰/۵۷	۰/۵۲
realTweets	۰/۵۹	۰/۵۷

ه) بررسی معیار MCC: در این قسمت نتایج مربوط به معیار MCC جهت مقایسه روش پیشنهادی و روش پایه در انواع مختلف دسته-بندی داده‌ها در جدول ۶ ارائه شده است. همانطور که نتایج آزمایش (ه) نشان می‌دهد در دسته‌بندی داده‌های artificialNoAnomaly روش حاضر نسبت به روش پایه به میزان ۱۴/۲۸درصد، در داده‌های artificialWithAnomaly به میزان ۸/۳۳ درصد، در داده‌های realAdExchange به میزان ۶/۶۶ درصد، در داده‌های realAWSCloudwatch به میزان ۲۰/۰۰ درصد، در داده‌های realKnownCauses به میزان ۱۴/۲۸ درصد، در داده‌های realTraffic به میزان ۱۲/۵۰ درصد و در داده‌های realTweets به میزان ۱۶/۶۶ درصد عملکرد بهتری از خود نشان داده است. در مجموع می‌توان گفت روش حاضر به میزان ۱۳/۲۴ درصد نسبت به روش پایه عملکرد بهتری در معیار MCC از خود نشان داده است.

جدول ۶: مقایسه نتایج بررسی معیار MCC بر روی روش مقاله پایه و روش پیشنهادی.

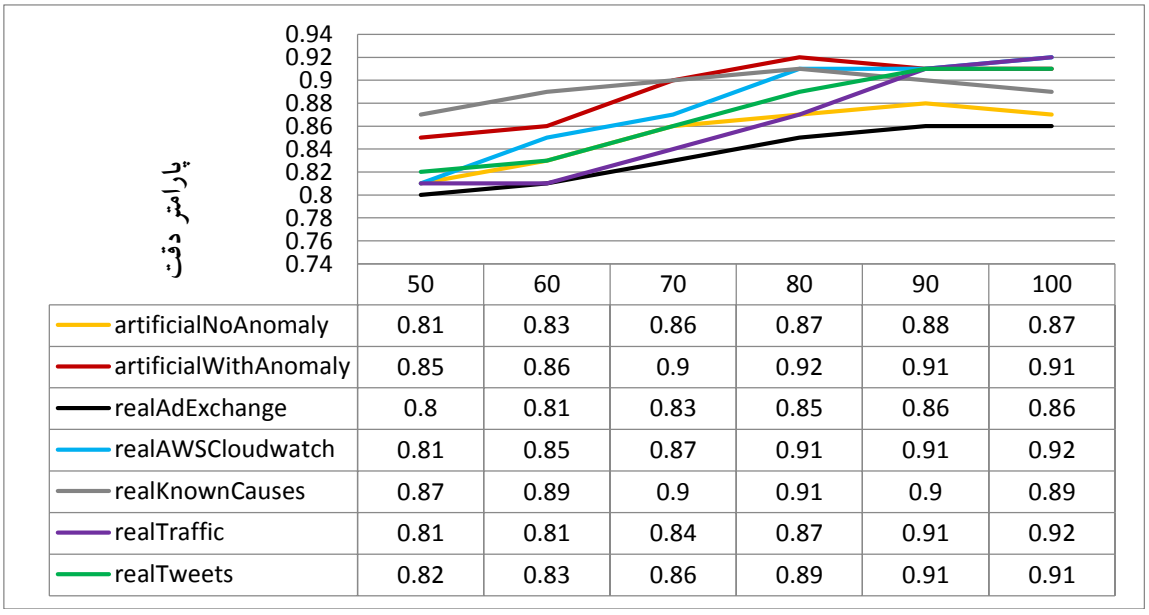
دسته بندی داده ها	نتایج روش پایه	نتایج استخراجی از روش پیشنهادی
artificialNoAnomaly	۰/۱۴	۰/۱۲
artificialWithAnomaly	۰/۱۲	۰/۱۱
realAdExchange	۰/۱۵	۰/۱۴
realAWSCloudwatch	۰/۱۰	۰/۰۸
realKnownCauses	۰/۱۴	۰/۱۲
realTraffic	۰/۱۶	۰/۱۴
realTweets	۰/۱۲	۰/۱۰

آزمایش دوم: بررسی تأثیر پارامتر w_{size} بر عملکرد نهایی: در این آزمایش از ELM جهت بهبود روش OeSNN-UAD برای تشخیص آنومالی استفاده شده‌است. اندازه پنجره از ۵۰ الی ۱۰۰ با گام ۱۰ درنظر گرفته شده است. این آزمایش ۱۰ بار تکرار و نتایج آن بصورت میانگین اعلام شده است. هدف بررسی تأثیر اندازه پارامتر W_{size} بر عملکرد نهایی است.

در این بخش، نتایج به دست آمده از شبیه‌سازی روش پیشنهادی که توسط نرم‌افزار متلب به دست آمده، آورده شده‌است.

آزمایش اول: بررسی تأثیر ELM در روش پیشنهادی در این آزمایش تمامی داده‌ها مورد آزمایش قرار گرفته اند. معیارهایی نیز که در این فصل اشاره شده است برای بررسی عملکرد تاثیر ELM در روش پیشنهادی مورد آزمایش قرار گرفتند. در جدول زیر معیارهای محاسبه شده نشان داده شده‌است. با توجه به اینکه ELM نیاز به آموزش دارد این آزمایش ۱۰ بار تکرار شد و نتایج آن بصورت میانگین اعلام شده است.

الف) بررسی معیار Precision: در این قسمت نتایج مربوط به معیار دقت جهت مقایسه روش پیشنهادی و روش پایه در انواع مختلف دسته‌بندی داده‌ها در جدول ۵ ارائه شده است. همانطور که نتایج آزمایش (الف) نشان می‌دهد مقایسه بین روش پایه و روش پیشنهادی انجام شده است. در دسته‌بندی داده‌های artificialNoAnomaly روش حاضر نسبت به روش پایه به میزان ۴,۸۱ درصد، در داده‌های artificialWithAnomaly به میزان ۲,۲۴ درصد، در داده‌های realAdExchange به میزان ۲,۳۸ درصد، در داده‌های realAWSCloudwatch به میزان ۱,۰۹ درصد، در داده‌های realKnownCauses به میزان ۲/۲۹ درصد، در داده‌های realTraffic به میزان ۳/۳۷ درصد و در داده‌های realTweets به میزان ۵/۸۱ درصد عملکرد بهتری از خود نشان داده است. در تمامی این داده‌ها به دلیل خطای کمتر تشخیص که توسط ELM ایجاد شده‌است روش حاضر عملکرد بهتری از خود نشان می‌دهد. در مجموع می‌توان گفت روش حاضر به میزان ۳/۱۴ درصد نسبت به روش پایه دقت بهتری از خود نشان داده است.



	50	60	70	80	90	100
artificialNoAnomaly	0.81	0.83	0.86	0.87	0.88	0.87
artificialWithAnomaly	0.85	0.86	0.9	0.92	0.91	0.91
realAdExchange	0.8	0.81	0.83	0.85	0.86	0.86
realAWSCloudwatch	0.81	0.85	0.87	0.91	0.91	0.92
realKnownCauses	0.87	0.89	0.9	0.91	0.9	0.89
realTraffic	0.81	0.81	0.84	0.87	0.91	0.92
realTweets	0.82	0.83	0.86	0.89	0.91	0.91

شکل ۲: بررسی تأثیر پارامتر w_{size} بر روی پارامتر دقت.

همانطور که در نمودار شکل ۲ مشخص شده است در پایگاه داده artificialNoAnomaly بالاترین میزان دقت در اندازه پنجره ۹۰ بدست آمده است که نسبت به اندازه پایه درنظر گرفته شده به میزان ۱/۱۴ درصد عملکرد بهتری از خود نشان می‌دهد. در پایگاه داده artificialWithAnomaly بالاترین میزان دقت در اندازه پنجره ۸۰ بدست آمده است که نسبت به اندازه پایه درنظر گرفته شده به میزان ۱/۰۹ درصد عملکرد بهتری از خود نشان می‌دهد. در پایگاه داده realAdExchange بالاترین میزان دقت در اندازه پنجره ۹۰ و ۱۰۰ و به میزان ۹۱/. بدست آمده است. در پایگاه داده realAWSCloudwatch بالاترین میزان دقت در اندازه پنجره ۱۰۰ و به میزان ۰,۹۲ بدست آمده است. در پایگاه داده realKnownCauses بالاترین میزان دقت در اندازه پنجره ۸۰ و به میزان ۲,۲۴ درصد عملکرد بهتری داشته است. در پایگاه داده realTraffic بالاترین میزان دقت در اندازه پنجره ۱۰۰ و به میزان ۹۲/. بدست آمده است. در پایگاه داده realTweets بالاترین میزان دقت در اندازه پنجره ۹۰ و ۱۰۰ و به میزان ۰/۹۱ بدست آمده است.

۶- نتیجه‌گیری و پژوهش‌های آتی

استفاده از داده‌های جریانی در دنیای امروز افزایش پیدا کرده‌است. به همین منظور برقراری امنیت این داده‌ها ضروری است و اگر این داده‌های مخرب تشخیص داده نشوند می‌توانند صدمات جبران ناپذیری را به بار آورند. به همین منظور استفاده از روش‌هایی که بتوانند در تشخیص این داده‌ها کمک کننده باشند مورد نیاز است. شبکه‌هایی مانند OeSNN-UAD به عنوان یکی از این روش‌ها شناخته می‌شود. این شبکه دارای لایه‌های ورودی و خروجی است که تعداد لایه‌های خروجی آن با توجه به دلایلی مانند پردازش حجم زیادی از داده‌ها و مشکل حافظه که به طبقه‌بندی کردن مقادیر داده‌ها مربوط است محدودتر است. در این پژوهش روشی برای بهبود تشخیص نفوذ در شبکه‌های رایانه‌ای با استفاده از داده‌های جریانی مبتنی بر شبکه عصبی ارائه شد. به منظور بهبود تشخیص ناهنجاری روش OeSNN-UAD که با توجه به کلاس‌بندی و رمزگذاری داده‌های ورودی به تشخیص ناهنجاری در شبکه پرداخته است، از روشی به نام MIS-ELM استفاده می‌شود که با این ترکیب به بهبود کلاس‌بندی و رمزگذاری، تشخیص ناهنجاری در شبکه نتیجه شد. شبیه سازی روش پیشنهادی در نرم افزار مطلب انجام شد و در آزمایش اول تأثیر ELM در روش پیشنهادی مورد بررسی قرار گرفت نتایج بررسی معیارهای precision ، Recall ، F-Score ، BA و MCC نشان داد که روش پیشنهادی به نسبت روش پایه بهبود چشمگیری داشت. در آزمایش دوم تأثیر پارامتر w_{size} بر عملکرد نهایی روی چندین مجموعه داده بررسی شد و نتایج نشان داد روش پیشنهادی به نسبت روش پایه نتیجه بهتری داشته است. در مورد کارهای آتی میتوان به موارد زیر اشاره کرد: بهبود تشخیص آنومالی در داده‌های جریانی با استفاده از طبقه‌بند SVM بهبود تشخیص آنومالی در داده‌های جریانی با استفاده از خوشه‌بندی K-means.

۷- مراجع

1. Aditya Ghongade, A.D., Harsh Munot,Parth Lokhande. , Survey Paper on Various Challenges in Data Stream Mining. IJERTV10IS010220 (This work is licensed under a Creative Commons Attribution 4.0 International License.) 2021.

2. Maciąg, P.S., et al., Unsupervised anomaly detection in stream data with online evolving spiking neural networks. Neural Networks, 2021. 139: p. 118-139.

3. Li, Q., et al., Incremental semi-supervised Extreme Learning Machine for Mixed data stream classification. Expert Systems with Applications, 2021. 185: p. 115591.

4. Ahmad, S., et al., Unsupervised real-time anomaly detection for streaming data. Neurocomputing, 2017. 262: p. 134-147.

5. Munir, M., et al., Fusead: unsupervised anomaly detection in streaming sensors data by fusing statistical and deep learning models. Sensors, 2019. 19(11): p. 2451.

6. Munir, M., et al., DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. Ieee Access, 2018. 7: p. 1991-2005.

7. Ergen, T. and S.S. Kozat, Unsupervised anomaly detection with LSTM neural networks. IEEE transactions on neural networks and learning systems, 2019. 31(8): p. 3127-3141.

8. Rettig, L., et al., Online anomaly detection over big data streams, in Applied data science. 2019, Springer. p. 289-312.

9. Ma, J., et al., Supervised anomaly detection in uncertain pseudoperiodic data streams. ACM Transactions on Internet Technology (TOIT), 2016. 16(1): p. 1-20.

10. Amoozegar, M., Minaei Bidgoli, Behrooz, Fanaei, Hadi, Razghi, Mansour. , Provide an informed drift learner online to detect anomalies in flow data. . Computational intelligence in electrical engineering, 2021.

11. Zamanian, S.a.J., Reza and Haddadpajooh, Hamed, , Presenting an Optimal Deep Learning Model for Diagnosis of Anomalies in Stream-Based Networks, . 11th National Conference on Computer Science and Engineering, Information Technology, Babol,, 2020.

12. Sharifian, M., and Expert, H., and Sharifian, S. , Improve network intrusion detection by identifying effective features based on evolutionary algorithms and backup vector machine classification. . Computational Intelligence in Electrical

Engineering (Intelligent Systems in Electrical Engineering), 11 (1), 29-42., 2020.

13. Hosseini, S.M.a.M., Nasser, , Improving network intrusion detection with artificial intelligence algorithms, . the first scientific conference on mechanics, electricity, computer and engineering,, 2020.

14. Pourrajbi, H., Detection of intrusion in computer networks using genetic algorithm, . First National Conference on Computer Engineering.

روش ارجاع به مقاله :
م. شیری، ن. مدیری، ارائه مدلی نوین جهت بهبود تشخیص نفوذ در شبکه با استفاده از روش یادگیری ماشین افزایشی در شبکه‌های عصبی spiking در حال تکامل آنلاین، دوفصلنامه محاسبات و سامانه های توزیع شده، سال پنجم، شماره ۲، شماره پیاپی۱۰، صفحه ۴۹ تا ۶۱ ، سال ۱۴۰۱

How to cite: M. shiri, N. modiri, A new model for Improvement anomaly detection in network by incremental learning machine in Online Evolving Spiking Neural Networks, Journal of Distributed Computing and Systems(JDCS), Vol 5, Issue 2, Page 49-61 , 2023.

معصومه شیری معلم رشته کامپیوتر در استان زنجان، مدرک کارشناسی خود را در رشته کامپیوتر گرایش نرم افزار در تاریخ ۱۳۹۲ از دانشگاه آزاد اسلامی و در سال ۱۴۰۱ مدرک کارشناسی ارشد خود را در رشته مهندسی کامپیوتر گرایش نرم افزار از دانشگاه آزاد اسلامی زنجان اخذ نموده است.

نشانه رایا نامه ایشان عبارتند از :

masoumehshiri1261@gmail.com

ناصر مدیری مدرک کارشناسی ارشد خود را از دانشگاه ساوتهمپتون، انگلستان و مدرک دکترا از دانشگاه ساسکس، انگلستان به ترتیب در سال های ۱۹۸۶ و ۱۹۸۹ اخذ نمود. در سال ۱۹۸۹ به عنوان مهندس اصلی در STC Telecommunications Systems، انگلستان مشغول به کار شد در سال ۱۹۹۰ در شرکت Teknekron Communications Systems آمریکا به عنوان مهندس سیستم در پروژه British Telecom Concert و سپس به عنوان مهندس ارشد در Network Equipment TEchnologies مشغول به کار شد. دکتر مدیری در حال حاضر عضو هیأت علمی دانشگاه آزاد زنجان گرایش توسعه نرم افزارهای امن و امنیت شبکه می باشد. علایق تحقیقاتی او توسعه برنامه های کاربردی برای فناوری هایERP ، RFID، ISO/IEC 27000، ISO/IEC 15408است.

نشانه رایانامه ایشان عبارتند از :

nassermodiri@yahoo.com