



Semnan University

Journal of Modeling in Engineering

Journal homepage: <https://modelling.semnan.ac.ir/>



Research Article

Online and Robust Anomaly Detection using Recurrent Neural Network

Morteza Faezinia¹, Maryam Amoozegar^{2,*}, Behrouz Minaei-Bidgoli³

1. Department of Computer engineering, Iran University of Science and Technology, Tehran, Iran

2. Department of Computer and Information Technology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran

3. Department of Computer engineering, Iran University of Science and Technology, Tehran, Iran

*Corresponding Author: Amoozegar_m@yahoo.com

PAPER INFO

Paper history:

Received: 07 December 2022

Revised: 08 April 2023

Accepted: 10 May 2023

Keywords:

Anomaly detection,
Robust model,
Dynamic data,
Recurrent neural network

ABSTRACT

In many real-world applications, data is dynamic and noisy. In such a situation, anomaly detection should be performed with an online and robust model against noise. In recent years, recurrent neural networks have been used on data sequences and have achieved good performance in this field. The existing methods do not have sufficient robustness against noise. This paper presents a method for anomaly detection in dynamic graph data using recurrent neural networks that are robust against noise and have sufficient adaptivity to changes in the data pattern. The proposed robust recurrent neural network extracts and introduces anomalies for the purpose of noise management. At the same time, it learns the original patterns in an online manner and is adapted to the changes. To evaluate the proposed method, some experiments are presented that measure its ability in anomaly detection in addition to the learning and adaptation ability in comparison with the existing methods. The results have confirmed the superiority of the proposed method.

© 2023 Published by Semnan University Press.

DOI: <https://doi.org/10.22075/jme.2023.29201.2370>

How to cite this article:

Amoozegar, M., Faezinia, M., & Minaei_Bidgoli, B. (2023). Online and Robust Anomaly Detection using Recurrent Neural Network. *Journal of Modeling in Engineering*, 21(74), 243-254. doi: 10.22075/jme.2023.29201.2370

تشخیص برخط و استوار ناهنجاری با استفاده از شبکه عصبی بازگشتی

مریم آموزگار^{۱*}، مرتضی فائزی نیا^۲ و بهروز مینایی بیدگلی^۳

چکیده	اطلاعات مقاله
<p>در بسیاری از مسائل دنیای واقعی، داده، پویا و دارای نویز است. در چنین شرایطی تشخیص ناهنجاری باید با یک مدل برخطی که در مقابل نویز استوار است، انجام شود. در سال‌های اخیر، شبکه‌های عصبی بازگشتی بر روی توالی داده‌ها مورد استفاده قرار گرفته و نتایج خوبی بدست آورده‌اند. اما راهکارهای موجود، استواری کافی را در مقابل نویز ندارند. این مقاله، به ارائه راهکاری برای تشخیص ناهنجاری در داده گرافی پویا با استفاده از شبکه‌های عصبی بازگشتی می‌پردازد که در مقابل نویز استوار بوده و تطبیق پذیری کافی را در مقابل تغییرات داده‌ها دارد. نسخه استوار ارائه شده از شبکه عصبی بازگشتی، به هدف مدیریت نویز، همزمان با یادگیری الگوی اصلی و تطبیق با تغییرات، ناهنجاری‌ها را نیز استخراج و معرفی می‌کند. برای بررسی صحت عملکرد روش پیشنهادی، آزمایش‌هایی ارائه شده است که قدرت تشخیص ناهنجاری و توان تطبیق یادگیرنده را در مقایسه با راهکارهای موجود می‌سنجد. نتایج، برتری روش پیشنهادی را تصدیق کرده است.</p>	<p>نوع مقاله: پژوهشی دریافت مقاله: ۱۴۰۱/۰۹/۱۶ بازنگری مقاله: ۱۴۰۲/۰۱/۱۹ پذیرش مقاله: ۱۴۰۲/۰۲/۲۰</p>
	<p>واژگان کلیدی: تشخیص ناهنجاری، مدل استوار، داده پویا، شبکه عصبی بازگشتی.</p>

۱- مقدمه

تشخیص ناهنجاری یکی از مهم‌ترین مسائل مطرح در حوزه داده‌کاوی است. ناهنجاری عبارت است از الگوهایی در داده‌ها که با رفتار طبیعی حاکم بر آن‌ها مطابقت ندارند. به عبارت ساده‌تر داده‌هایی که با اکثریت داده‌ها تفاوت قابل ملاحظه‌ای دارند. ناهنجاری‌ها در دنیای واقعی امری عادی هستند. تشخیص آن‌ها در کاربردهای مختلف از نظر اهمیت و حساسیت، تابع مشخصات فضای مسئله است. ناهنجاری‌ها حاوی اطلاعات بسیار مهمی در مورد داده هستند و در برخی از دامنه‌ها ارزش آن‌ها حتی از داده‌ی اصلی نیز بیشتر است. ناهنجاری ممکن است ناشی از دلایل گوناگون مانند فعالیت‌های مخرب، نفوذ سایبری، وجود توده و تومور در بدن بیمار یا شکست یک سیستم باشد [۱]. برای حل مسئله تشخیص ناهنجاری در داده‌های پویا رویکردهای مبتنی بر تجزیه ماتریس، شبکه‌های عصبی

خودکدگذار و شبکه‌های عصبی بازگشتی ارائه شده‌اند. رویکردهای مبتنی بر تجزیه ماتریس [۲، ۳، ۴ و ۵] و تانسور [۶، ۷، ۸، ۹، ۱۰ و ۱۱]، امکان پردازش برخط را دارند. داده تانسوری در هر برش زمانی در راستای هر بعد، ماتریسی شده، با اعمال تابع بهینه‌سازی، دو مولفه زیرفضای پایه و ماتریس خلوت استخراج می‌شوند. ماتریس خلوت مبنای تشخیص ناهنجاری است. این رویکردها از نظر نوع تجزیه‌ی مورد استفاده، مسئله بهینه‌سازی مطرح شده و الگوریتم ارائه شده با یکدیگر متفاوتند. تشخیص ناهنجاری اخیراً با استفاده از شبکه‌های عصبی، نتایج قابل توجهی را بدست آورده است. شبکه‌های عصبی بخشی از خانواده گسترده‌ای از روش‌ها برای یادگیری بازنمایی هستند که توانایی خودکار شناسایی یا رده‌بندی ویژگی‌ها به صورت خودکار از داده‌های خام را دارند [۱۲]. در این مقاله تمرکز روی شبکه‌های عصبی است که نمایشی بر

* پست الکترونیک نویسنده مسئول: Amoozegar_m@yahoo.com

۱. استادیار، پژوهشکده فناوری اطلاعات و کامپیوتر، پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته

۲. دانش آموخته کارشناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران

۳. دانشیار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران

مبتنی بر شبکه عصبی بازگشتی ارائه شده است. این روش، گرچه به دلیل برخط بودن و لحاظ کردن سابقه داده با تکیه بر حافظه شبکه بازگشتی، امکان تطبیق با پویایی داده را دارد، ولی استواری کافی در برابر نویز و داده پرت را ندارد زیرا مبتنی بر مدل‌های استوار نیست و استراتژی خاصی را در این رابطه در نظر نگرفته است.

بر این اساس، رویکردهای ارائه شده، به صورت برون خط و برخط، ناهنجاری داده‌های ورودی را تشخیص می‌دهند، ولی معمولاً عملکرد پایداری در مواجهه با نویز و داده پرت ندارند. در این میان با این استدلال که داده‌ها به طور مداوم در حال تغییر بوده و ماهیت پویایی دارند، مسئله پیچیده‌تر خواهد شد چرا که مدل ساخته شده باید قابلیت تطبیق با پویایی و تغییرات داده را علاوه بر استواری در برابر نویز و داده پرت داشته باشد. بنابراین مدل‌های یادگیری برون خط، کارایی لازم در این شرایط را نداشته و به مدل‌های یادگیری برخط تطبیق‌پذیر نیاز است. رویکردهای برخط سعی دارند به تدریج و با مشاهده رفتار داده، الگوها را شناسایی و ناهنجاری‌های موجود در داده را تشخیص دهند [۶]. علاوه بر این، مدل‌هایی که در برابر نویز استوار نیستند، دچار ضعف در عملکرد خواهند شد.

با توجه به مطالب فوق، این مقاله رویکرد بی‌ناظری را پیشنهاد می‌دهد که به واسطه لحاظ کردن ماتریس خلوت در مدل، در مقابل داده‌های نویزی استوار است و علاوه بر این به واسطه یادگیری برخط، تطبیق‌پذیری و انعطاف بیشتری در مقابل تغییرات داده و پویایی آنها دارد. همچنین با اتکا به شبکه عصبی بازگشتی و بهره‌گیری از حافظه و ساختار بازگشتی آن، عملکرد مطلوب‌تری، در مقایسه با روش‌های مشابه، بدست آورده است. در نهایت بهبود عملکرد یادگیرنده، منجر به بهبود دقت در تشخیص ناهنجاری شده است.

در ادامه، بخش دوم به مروری بر مبانی پایه می‌پردازد. روش پیشنهادی در بخش سوم و آزمایش‌ها و تحلیل نتایج در بخش چهارم ارائه خواهند شد. بخش پایانی، به نتیجه‌گیری و کارهای آینده می‌پردازد.

۲- مبانی پایه

۲-۱- یادگیری برخط

یادگیری ماشین برخط زیرمجموعه‌ای از یادگیری ماشین است که پارامترهای یادگیری مدل می‌توانند به صورت تدریجی توسط یک یادگیرنده برخط در هنگام ورود داده

اساس ترکیبات غیرخطی از ویژگی‌های ورودی ایجاد می‌کنند [۱۳].

تحقیقات زیادی مبتنی بر شبکه عصبی و یادگیری عمیق ارائه شده است که عمدتاً باناظر هستند [۱۴]. روش‌های باناظر، برای آموزش نیاز به داده برچسب‌دار در حجم بالا دارند که خود یکی از چالش‌های عمده حوزه تشخیص ناهنجاری است. از طرف دیگر این روش‌ها امکان تشخیص ناهنجاری از نوع جدید را ندارند و در مقابل تغییرات داده به اندازه کافی منعطف نیستند [۱۵ و ۱۶]. از این رو این مقاله، بر روی روش‌های بی‌ناظر متمرکز شده است که اخیراً بیشتر مورد توجه قرار گرفته‌اند اما هنوز چالش‌های باز زیادی در آنها مطرح است. در این راستا و در دسته بی‌ناظر، رویکردهایی مبتنی بر شبکه‌های عصبی خودکدگذار ارائه شده‌اند که داده‌های اصلی را به فضای ویژگی پنهان نگاشت کرده و سپس ناهنجاری‌ها را در نقاط با خطای بازسازی بالا، شناسایی می‌کنند [۱۷]. در پژوهش [۱۸] رویکردی برای مواجهه با نویز معرفی شده است که برای استواری در مقابل آن، سنجه بیشینه مقدار کور آنتروپی را با خطای بازسازی خودکدگذار جایگزین می‌کند. این رویکرد با توجه به برون خط بودن و عدم پایداری سنجه پیشنهادی، در مقابل تغییرات، انعطاف کافی را ندارد. در پژوهش [۱۹] رویکرد دیگری مبتنی بر شبکه عصبی خودکدگذار ارائه شده است که با توجه به حذف نویز از ورودی پیش از استفاده در مدل، امکان یادگیری بهتری را در لایه پنهان فراهم می‌کند. بنابراین این رویکرد در مقابل نویز استوار است ولی همانند رویکرد قبلی به صورت برون خط عمل کرده و در فضای پویای داده، عملکرد پایدار و تطبیق‌پذیری ندارد. از آنجا که عملکرد خودرمزگذار به نمایش ویژگی در فضای پنهان متکی است، تکنیک محدودیت منظم‌سازی گراف برای اولین بار در [۲۰] معرفی شد تا اطلاعات ساختار هندسی نیز بررسی شود. بر مبنای این تکنیک، رویکردی تحت عنوان DSGRAE در [۱۷] ارائه شد که با اعمال قید پراکندگی در شبکه به حل مشکل بیش‌برازش و افزایش توانایی نمایش ویژگی‌ها پرداخت. در واقع استفاده از این قیدها ضمن پیچیده کردن شبکه، کارایی را کاهش می‌دهند. این در حالیست که رویکردهای برخط این انعطاف را با هزینه کمتری ایجاد می‌کنند.

نکته مهم دیگر، توجه به سابقه داده و استفاده از شبکه‌های حافظه‌دار است. در این راستا در پژوهش [۲۱] رویکردی

با فرض لایه ورودی، لایه پنهان و لایه خروجی در زمان t به ترتیب x^t ، h^t و o^t می‌توان خروجی شبکه را به صورت زیر محاسبه کرد:

$$\begin{aligned} a^t &= b_1 + Wh^{(t-1)} + Ux^t & (1) \\ h^t &= \sigma(a^t) \\ o^t &= b_2 + Vh^t \end{aligned}$$

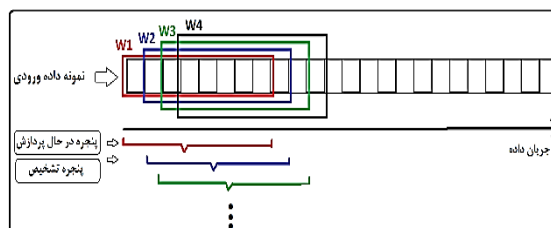
که در آن b_1 و b_2 بردارهای سوگیری هستند. U ، V و W به ترتیب ماتریس‌های وزنی اتصال لایه ورودی به لایه پنهان، اتصال لایه پنهان به لایه خروجی و اتصال لایه پنهان به لایه پنهان بعدی هستند و σ تابع فعال‌سازی است. معمولاً از تابع سیگموئید $\sigma(x) = (1 + e^x)^{-1}$ به عنوان تابع فعال‌سازی استفاده می‌شود. برای مسئله وابستگی‌های طولانی‌مدت در یادگیری شبکه‌های عصبی بازگشتی، روش‌های متعددی پیشنهاد گردیده است. این روش‌ها با تغییر در روش یادگیری و معماری جدید ایجاد شده‌اند. معماری شبکه عصبی حافظه‌ی کوتاه‌مدت طولانی‌مدت (LSTM) و شبکه عصبی واحد بازگشتی گیتی^۳ (GRU) برای مقابله با مسئله ناپدید شدن گرادیان معرفی شدند. این شبکه‌ها به تدریج جای شبکه‌های عصبی بازگشتی استاندارد را گرفته و به محبوب‌ترین نوع شبکه‌های عصبی بازگشتی تبدیل شده‌اند. در رویکرد پیشنهادی این مقاله از شبکه عصبی گیتی (GRU) استفاده شده است. این شبکه توانایی یادگیری توالی‌های طولانی را داراست و همچنین بر مشکل محوشدگی گرادیان غلبه می‌کند. یک واحد GRU بخش‌های ساده‌تری مثل دروازه بروزرسانی^۴ دارد که برای تعیین مقدار اطلاعات قبلی‌ای که باید در امتداد وضعیت بعدی تأثیر داده شوند، استفاده می‌شود. بخش دیگر دروازه فراموشی^۵ است که تصمیم می‌گیرد چه مقدار از اطلاعات گذشته نادیده گرفته شود. به طور کلی مطابق شکل ۳ در GRU ابتدا، گیت فراموشی وارد عمل می‌شود و اطلاعات مربوطه را از مرحله گذشته به حافظه جدید منتقل می‌کند. سپس بردار ورودی و حالت پنهان را با وزن آنها ضرب می‌کند. در مرحله بعد، ضرب بین دروازه تنظیم مجدد و مضرب حالت پنهان قبلی محاسبه می‌شود. پس از جمع‌بندی این مراحل، تابع فعال‌سازی غیرخطی اعمال می‌شود و دنباله بعدی ایجاد می‌شود.

⁴ Forget Gate

جدید، به‌روزرسانی شوند [۲۲].

داده ورودی در حالت برخط می‌تواند با تکنیک‌های مختلفی پردازش شود، پنجره لغزان یکی از روش‌های دریافت ورودی در روش‌های پردازش برخط است. این پنجره روی داده‌ها قرار گرفته و به تدریج روی آن‌ها حرکت می‌کند تا همه داده‌ها مشاهده شوند. پنجره‌ها انواع مختلفی از نظر پویایی اندازه و یا وجود همپوشانی در داده‌های مشترک بین پنجره‌های متوالی، دارند [۲۳].

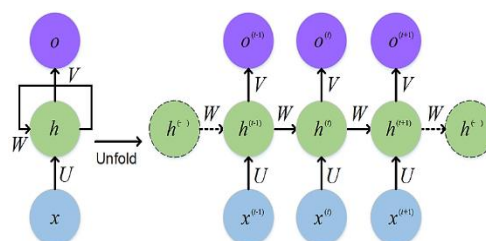
شکل (۱) نوعی از مدل‌های پنجره لغزان را نشان می‌دهد که به صورت همپوشان و با اندازه ثابت است.



شکل ۱- پنجره همپوشان ثابت.

۲-۲- شبکه عصبی بازگشتی

این شبکه‌ها دسته‌ای از شبکه‌های عصبی مصنوعی برای پردازش داده‌های متوالی بوده و برای پردازش دنباله‌های طولانی نسبت به سایر شبکه‌ها، عملکرد قابل قبولی دارند. همچنین این شبکه‌ها برای پردازش توالی‌هایی با طول پویا نیز مناسب هستند. در این شبکه‌ها هر عضو خروجی، به عنوان تابعی از خروجی قبلی است. مطابق شکل (۲) شبکه عصبی بازگشتی دارای یک حلقه بر روی واحد پنهان است و همچنین دارای سه لایه، لایه ورودی x ، لایه پنهان h و لایه خروجی o است. اگر مشابه شکل (۲)، حلقه را بسط دهیم، شبکه عصبی بازگشتی استاندارد را می‌توان به عنوان کپی چندباره از یک ساختار در نظر گرفت و حالت h هر کپی به عنوان ورودی برای جانشین آن در نظر گرفته می‌شود.



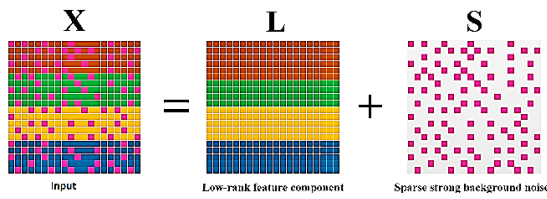
شکل ۲- رفتار یک شبکه عصبی بازگشتی معمولی [24]

¹ Long short-term memory

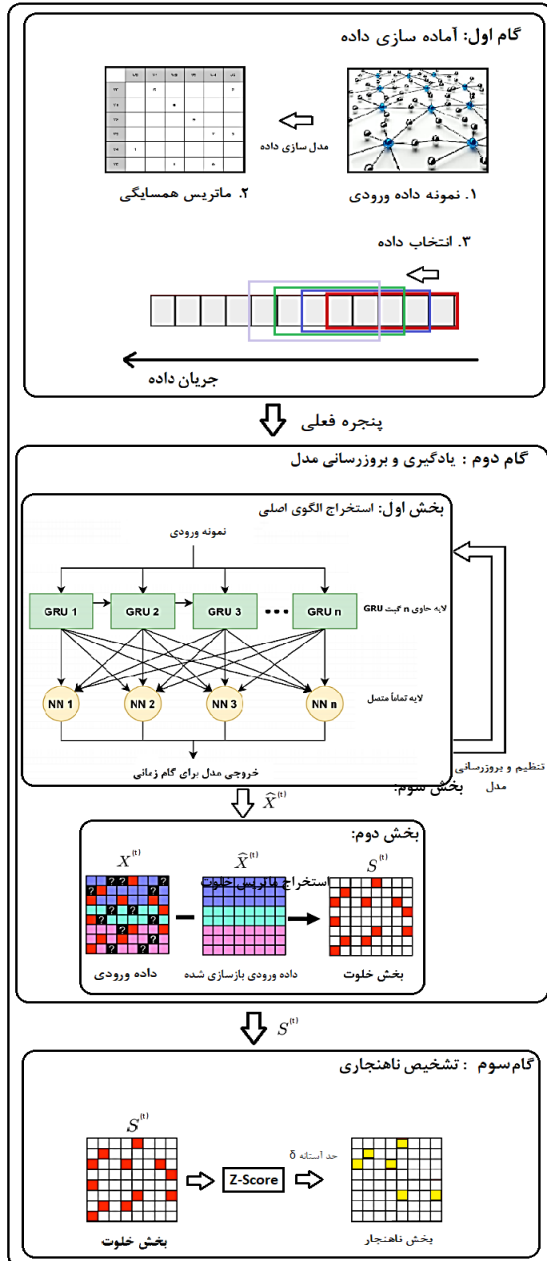
² Gated recurrent unit

³ Update Gate

ماتریس S بازنمایی از داده‌های نویزی ورودی و ماتریس L بازنمایی ورودی بدون نویز است، شکل (۴) تصویری از عملکرد این تکنیک را نشان می‌دهد.



شکل ۴- عملکرد رویکرد استوار [26]



شکل ۵- مراحل رویکرد پیشنهادی

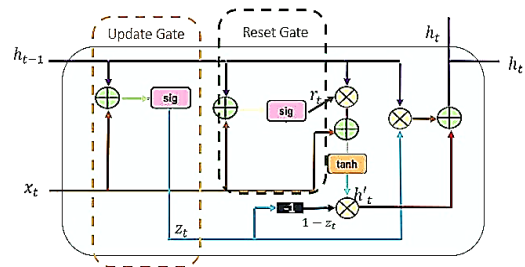
محاسبات مربوط به GRU به صورت زیر است:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

$$\hat{h}_t = \phi(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h)$$

که در آن r_t مربوط به گیت بروزرسانی، z_t مربوط به این مسئله است که چه میزان از اطلاعات گذشته به آینده منتقل شود، h_t خروجی مدل و \hat{h}_t حالت پنهان مدل است.



شکل ۳- اجزای تشکیل دهنده یک بلوک GRU [۲۴]

محاسبات مربوط به GRU به صورت زیر است:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

$$\hat{h}_t = \phi(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h)$$

$$h_t = z_t \odot \hat{h}_t + (1 - z_t) \odot h_{t-1} \quad (2)$$

که در آن r_t مربوط به گیت بروزرسانی، z_t مربوط به این مسئله است که چه میزان از اطلاعات گذشته به آینده منتقل شود، h_t خروجی مدل و \hat{h}_t حالت پنهان مدل است.

۲-۳- مدل‌های استوار

به منظور ردگیری زیرفضای اصلی ورودی، روش‌های طیفی بنا شدند که داده‌های با بعد بالا را به زیرفضایی با مرتبه پایین‌تر نگاشت می‌کنند به طوری که در فضای جدید ناهنجاری به راحتی قابل تشخیص است. دسته دیگر از روش‌ها همانند تحلیل مؤلفه اصلی استوار^۶، توسعه‌ای از رویه آماری به کاررفته در روش تحلیل مؤلفه اصلی^۷ است که هدف آن بازیابی زیرفضای اصلی L و ماتریس خلوت S^A از ورودی X است [۲۵].

$$X = L + S \quad (3)$$

⁷ Sparse

⁵ Robust Principal Component Analysis

⁶ Principal Component Analysis

۳- رویکرد پیشنهادی

از آنجا که داده، همواره با نویز درآمیخته است، یادگیری رفتار و الگوی صحیح داده‌ها، نیازمند یک رویکرد استوار در برابر نویز است. در این میان تشخیص ناهنجاری نیز امکان‌پذیر می‌گردد. علاوه بر این، داده‌های دنیای واقعی به‌صورت مداوم در حال تغییر هستند، لذا الگوهای متعارف دستخوش تغییر شده و این تغییرات بر روی تشخیص ناهنجاری نیز تاثیر می‌گذارند. بنابراین مدلی که توانایی تطبیق بالاتری با تغییرات داشته باشد، هم در امر یادگیری و هم در امر تشخیص ناهنجاری، موفق‌تر عمل خواهد کرد. در راستای تامین این دو هدف یعنی یادگیری استوار و برخط و همچنین تشخیص ناهنجاری، رویکرد پیشنهادی این مقاله تحت عنوان ORURN (مخفف عبارت Online Robust Recurrent Network) ارائه شده است که مطابق شکل (۵)، در این بخش معرفی می‌گردد. رویکرد پیشنهادی از سه گام اساسی آماده‌سازی داده، یادگیری و بروزرسانی مدل و تشخیص ناهنجاری تشکیل شده است.

۳-۱- گام اول: آماده‌سازی داده

در این گام، اقداماتی برای تبدیل داده ورودی به داده مورد نیاز گام‌های بعدی انجام می‌گیرد. هر نمونه داده ورودی، یک گراف است که به یک ماتریس همسایگی تبدیل می‌شود و توالی آن‌ها یک سری زمانی گرافی را می‌سازد. از آنجا که در هر مرحله از یادگیری به تعدادی نمونه آموزشی نیاز است، تکنیک پنجره لغزان استفاده می‌شود. به این شکل که با استفاده از یک پنجره لغزان به طول ثابت، تعدادی از نمونه‌ها برای انجام فرآیند آموزش و تشخیص ناهنجاری برداشته می‌شوند. پنجره لغزان مورد استفاده، از نوع همپوشان است بدین معنی که هر بار فقط به اندازه یک داده، به جلو لغزانده می‌شود و داده‌های قبلی با پنجره قبلی همپوشانی دارند. فرآیند یادگیری بر روی همه داده‌ها و فرآیند تشخیص ناهنجاری بر روی داده جدید انجام می‌شود. بدین ترتیب دقت شبکه و یادگیرنده به واسطه مشاهده چندباره داده افزایش خواهد یافت. از آنجا که روش پیشنهادی به صورت برخط عمل می‌کند، هر داده تنها به تعداد دفعاتی معادل طول پنجره، در اختیار شبکه قرار می‌گیرد. از این رو پیچیدگی آن در مقایسه با روش‌های برون‌خط که داده‌ها را به صورت دسته‌ای و حتی در چندین تکرار پردازش می‌کنند، بسیار کمتر است. این مطلب در گام اول شکل (۵) به نمایش در آمده است.

۳-۲- گام دوم: یادگیری و بروزرسانی مدل

این گام از سه بخش استخراج الگوی اصلی، استخراج ماتریس خلوت و بروزرسانی مدل از روی داده‌ی پاکسازی شده از ناهنجاری، تشکیل شده است. در ادامه به شرح و بررسی این بخش‌ها پرداخته شده است.

• بخش اول: استخراج الگوی اصلی

تشخیص ناهنجاری در این مقاله مبتنی بر پیش‌بینی است. یعنی شبکه بر اساس آنچه از قبل آموخته است انتظار خود را از ورودی تولید می‌کند. آنچه شبکه در اختیار قرار می‌دهد، رفتار هنجار یا همان الگوی اصلی است. در مراحل بعدی، تفاضل آن از داده ورودی می‌تواند مبنای ناهنجاری باشد. از طرف دیگر با توجه اینکه راهکار پیشنهادی به صورت برخط عمل می‌کند، در هر گام زمانی به طور همزمان، هر دو مرحله آزمون (پیش‌بینی و تشخیص ناهنجاری) و یادگیری انجام می‌شوند. اما در بخش اول، داده ورودی دریافت شده از پنجره لغزان به شبکه آموزش دیده شده، داده می‌شود تا پیش‌بینی انجام شود. لذا شبکه، فقط یک خروجی ارائه می‌دهد بدون اینکه در پارامترهای تغییراتی ایجاد شود. خروجی به دست آمده در واقع همان رفتار هنجار قابل انتظار است.

• بخش دوم: استخراج ماتریس خلوت

در بخش دوم، ابتدا ماتریس تفاضل بین ورودی و مقدار قابل انتظار و هنجار بدست آمده از بخش قبل محاسبه می‌شود. در ادامه برای استخراج ماتریس خلوت، بر روی هر درایه آن، عملگر آستانه‌گیری نرم طبق رابطه زیر اعمال می‌گردد:

$$s_i = \text{prox}_{\lambda, l_1}(s_i) = \begin{cases} s_i - \lambda_1 & s_i > \lambda_1 \\ s_i + \lambda_1 & s_i < -\lambda_1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

در این رابطه λ ضریب خلوتی برای مولفه خلوت S و s_i هر درایه ماتریس تفاضل است. این ماتریس دو کاربرد مهم دارد. یکی اینکه مبنای تشخیص ناهنجاری است و در فاز تشخیص ناهنجاری مورد بهره‌برداری قرار می‌گیرد. کاربرد دیگر که در بخش بعد تشریح می‌شود، این است که منجر به آموزش و تعلیم شبکه با داده پاکسازی شده از ناهنجاری می‌شود.

• بخش سوم: تنظیم و بروزرسانی مدل

حال که الگوی مورد انتظار از شبکه بدست آمد و بخش خلوت و ناهنجار آن شناسایی شد، لازم است شبکه وارد مرحله یادگیری شده و بر اساس داده جدید بروزرسانی شود.

می‌شوند. در بخش ارزیابی اثبات می‌شود که روش پیشنهادی عملکرد مناسبی در داده‌های پویا دارای نویز داشته و امکان تشخیص برخط و استوار ناهنجاری را نیز دارد.

۴- ارزیابی روش پیشنهادی

رویکرد پیشنهادی، در واقع یک روش برخط مبتنی بر شبکه عصبی بازگشتی است که در مقابل نویز و ناهنجاری استوار است. برخط بودن سبب می‌شود که فرآیند یادگیری و ساخت مدل به صورت تدریجی-افزایشی بوده و لذا تطبیق‌پذیری بیشتری با تغییرات حاصل شود. از طرف دیگر استخراج ماتریس خلوت و جدا کردن آن از الگوی اصلی، باعث استواری آن در مقابل ناهنجاری‌ها و نویز می‌شود. لذا انتظار می‌رود روش پیشنهادی، دقت بیشتری در تشخیص ناهنجاری داشته باشد. این بخش، با تکیه بر دو ادعا و دستاورد ذکر شده، اثربخشی روش پیشنهادی ORURN را با استفاده از مجموعه‌داده NYC Taxi نشان می‌دهد. در این راستا، لازم است معیارهایی برای ارزیابی در نظر گرفته شوند که دو هدف تطبیق‌پذیری و تشخیص ناهنجاری را بررسی کنند. بدین منظور برای هدف اول، خطای بازسازی یادگیرنده بر مبنای رابطه زیر محاسبه می‌شود تا رفتار یادگیرنده را در هنگام بروز ناهنجاری و خصوصا تغییرات داده نشان دهد.

$$\frac{1}{T} \sum_{t=1}^T \|X_t - Y_t\|_F^2 / \|Y_t\|_F^2 \quad (5)$$

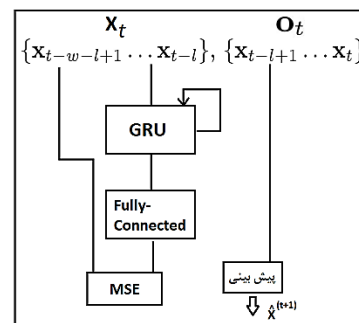
در این رابطه Y_t نمونه داده دریافتی و X_t الگوی ساخته شده توسط شبکه در زمان t است.

هدف دوم، یعنی تشخیص ناهنجاری بر مبنای معیار معروف F1 که میانگین هارمونیک فراخوانی و دقت است، محاسبه می‌شود. فراخوانی، در واقع درصدی از نمونه‌های مرتبط (ناهنجار) است که به درستی ارزیابی و شناسایی شده‌اند و دقت، درصد نمونه‌های مرتبط (ناهنجار) در بین همه موارد شناسایی شده به عنوان مرتبط است.

۴-۱- جزئیات پیاده‌سازی

در پیاده‌سازی رویکرد ORURN از کتابخانه پایتورچ در زبان پایتون استفاده شده است. همچنین برای آزمایش‌ها از بستر گوگل کولب با حافظه‌ی ۱۳ گیگابایت، پردازنده زئون ۲.۳ گیگاهرتزی و پردازنده گرافیکی تسلا T4 استفاده شده

این بار داده ورودی از ماتریس تفاضل بدست‌آمده در بخش قبل کاسته می‌شود تا داده پاکسازی‌شده از ناهنجاری، به شبکه ارسال و تنظیم پارامترها انجام گردد. به این شکل، به طور برخط و در انتهای هر مرحله، شبکه بر اساس دانش جدید بروز شده و تعلیم می‌بیند. از آنجا که یادگیری پس از استخراج ماتریس خلوت و کنار گذاشتن آنها و از روی داده تمیز انجام می‌شود، استواری در مقابل نویز و ناهنجاری، به طور کامل حاصل می‌گردد. شکل (۶) ساختار درونی شبکه را نشان می‌دهد. داده ورودی X_t در ابتدا وارد گیت GRU شده، سپس خروجی آن به یک لایه تماما متصل وارد می‌شود. هر دو مرحله پیش‌بینی و یادگیری، در این شکل نمایش داده شده‌اند.



شکل ۶- ساختار مدل یادگیری ORURN

۳-۳- گام سوم: تشخیص ناهنجاری

در این گام، بخش‌هایی از داده ورودی که در ماتریس خلوت ذخیره شده‌اند، برای تعیین ناهنجاری‌ها مورد تحلیل قرار می‌گیرند. بدین ترتیب که تابع Z-Score بر روی ماتریس خلوت اعمال شده و درایه‌هایی که قدرمطلق مقادیر آنها از یک حدآستانه بیشتر باشند، به‌عنوان ناهنجر گزارش می‌شوند. این حدآستانه در واقع حکم اندازه فیلتر را دارد. مقدار بزرگ برای این حدآستانه، به معنی فیلتر گشادتر، انتخاب داده کمتر به‌عنوان ناهنجر و در نتیجه فراخوانی کمتر و دقت بیشتر در معیار ارزیابی کارایی است. یک راه اصولی برای تعیین مقدار مناسب این پارامتر، مصالحه بین دو معیار مهم ارزیابی کارایی یعنی فراخوانی^۹ و دقت^{۱۰} است. اما با توجه به اینکه Z-score میزان انحراف از میانگین را نشان می‌دهد مقدار ۳ برای قدرمطلق آن مقدار مناسبی معرفی شده است [۲۰].

به طور خلاصه، در هر گام زمانی دو موضوع یادگیری باکیفیت و تشخیص ناهنجاری به‌عنوان هدف در نظر گرفته

¹⁰ Precision

⁹ Recall

است. مطابق بخش ۲-۲ و ۲-۳ براساس معادله مربوط به GRU و یادگیری زیرفضای اصلی پارامترهای زیر با انجام آزمایش‌ها بدست آمده است.

جدول ۱- پارامترهای مدل ORURN

مقدار نرخ یادگیری	پارامتر λ	تعداد گیت‌ها	نوع تابع خطا
۰.۰۳	۰.۰۸	۵	MSE

پردازش مدل در RURN به صورت برخط و با تکیه بر رویکرد پنجره لغزان همپوشان با طول ثابت است. برای رسیدن به بهینه‌ترین طول پنجره، اندازه‌های مختلف بررسی شده و بهترین اندازه‌ها برای این مسئله طول‌های ۱ و ۳ به دست آمده است.

۲-۴- مجموعه داده

مجموعه داده انتخابی در این مقاله مجموعه داده NYC Taxi 11 است که شامل گرافی با ۷۱ یال و ۱۵۶۰ گره است. این مجموعه داده توسط اداره حمل و نقل شهر نیویورک ارائه شده و در حوزه تشخیص ناهنجاری به وفور و از جمله در [۱۳، ۱۴ و ۲۷] مورد استفاده قرار گرفته است. برای اثبات قابلیت تشخیص ناهنجاری روش پیشنهادی، مجموعه داده اصلی ابتدا تمیز شده و سپس ناهنجاری و نویز به آن تزریق می‌شود. روند پاکسازی به شرح زیر است:

(۱) روال موجک ۵-Daubechies بر روی سیگنال حاصل از وزن هر یک از یال‌های گراف در طول زمان اعمال می‌شود تا گراف حاصله هموار گردد.

(۲) برش زمانی از مجموعه داده که روند تقریباً هموار و یکنواختی دارد، انتخاب می‌شود.

(۳) تجزیه CP اعمال می‌شود تا فاکتورهای پایه، استخراج گردند. در ادامه تابع زمان مربوط به هر سیگنال با استفاده از تکنیک‌های برازش منحنی استخراج و بر روی فاکتورهای پایه اعمال می‌شود.

(۴) نویز گوسی به داده‌ها اعمال می‌شود.

سپس بر اساس یک فرآیند مطمئن به شرح زیر به داده تمیز موجود، ناهنجاری تزریق می‌شود. به منظور تغییر زیرفضا، در گام‌های زمانی ۱۰۰، ۲۵۰ و ۳۰۰ روال زیر اعمال می‌شود. دو درصد از عناصر مربوط به ماتریس فاکتورهای

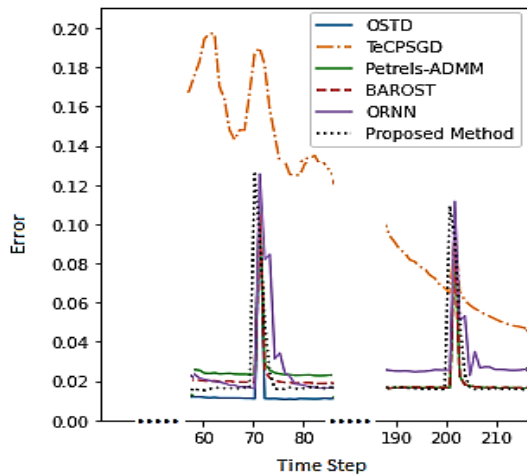
اصلی انتخاب و مقادیر آن‌ها بر اساس فاکتور ضربی، تغییر داده می‌شود. سپس ماتریس‌های فاکتور جدید برای بازسازی و تولید داده جدید استفاده می‌شوند. در انتها، با مقایسه داده‌ها، همه یال‌های تاثیر گرفته از این فرآیند تعیین و به عنوان یال‌های ناهنجاری که به صورت گروهی باعث تغییر زیرفضا شده‌اند، برچسب می‌خورند. در زمان‌های ۵۰، ۱۲۵ و ۱۵۰ تعدادی از درایه‌های ماتریس، معادل یال‌ها در شبکه، انتخاب و وزن آن‌ها به مقدار دیگری تغییر می‌کند. میزان این تغییر توسط یک فاکتور ضربی قابل تنظیم است. با مقادیر مختلف این فاکتور، ناهنجاری‌های با میزان انحراف متفاوت به مجموعه داده تزریق می‌شود. در این پژوهش، فاکتور ضربی بین مقدار ۰.۵ تا ۲ در نوسان بوده است. برای تزریق گره ناهنجار به شبکه، در گام‌های زمانی ۷۰ و ۲۰۰، تعداد محدودی از گره‌های شبکه انتخاب و وزن ده درصد از یال‌های متصل به آن‌ها، متناسب با فاکتور ضربی، تغییر داده می‌شود. بدین ترتیب مجموعه داده مورد نیاز، با برچسب‌های ناهنجار بدست می‌آید.

۳-۴- خطای بازسازی در مواجهه با ناهنجاری

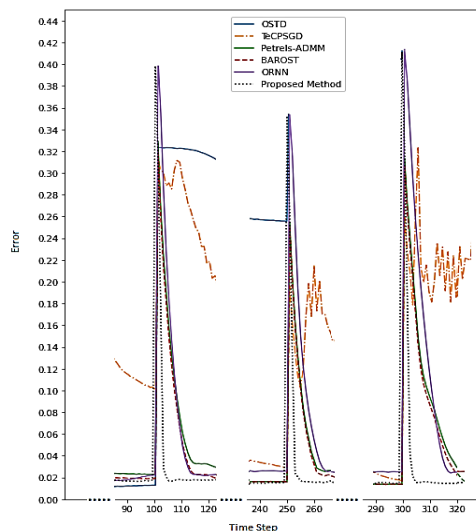
روش‌های قابل مقایسه انتخابی، روش‌های [۶] OSTD، [۲۸] ORNN، [۲۹] TeCPSG، [۳۰] BAROST و مرتبط‌ترین روش اخیراً ارائه شده [۲] PETRELS-ADMM هستند. OSTD یکی از پایه‌ای‌ترین و مهم‌ترین روش‌های ارائه شده استوار و برخط تشخیص ناهنجاری است. این روش در راهکار جامع تشخیص ناهنجاری [۳۱] Lrslibrary، نیز ارائه شده و مورد استفاده محققین قرار می‌گیرد. [۲۱] ORNN استوار نیست اما به دلیل اتکا به حافظه شبکه برگشتی و نزدیکی بسیار ساختاری به روش پیشنهادی، مورد استفاده قرار گرفته است. TeCPSGD به واسطه بررسی عملکرد گردایان کاهشی در بهینه‌سازی و BAROST و PETRELS-ADMM به واسطه استفاده از مشتق مرتبه دوم و بازگشتی انتخاب شده‌اند. بدین طریق راهکارهای انتخابی، وجوه مختلف مساله را پوشش داده و تاثیرگذاری آن‌ها را بررسی می‌کنند. با اعمال مجموعه داده به کلیه روش‌های انتخابی و روش پیشنهادی، نمودار خطای بازسازی مطابق شکل (۷) بدست آمد.

¹¹ <https://github.com/chickenbestlover/RNN-Time-series-Anomaly-Detection>

در شکل (۹) عملکرد روش‌ها را در ناهنجاری از نوع گره نشان می‌دهد.

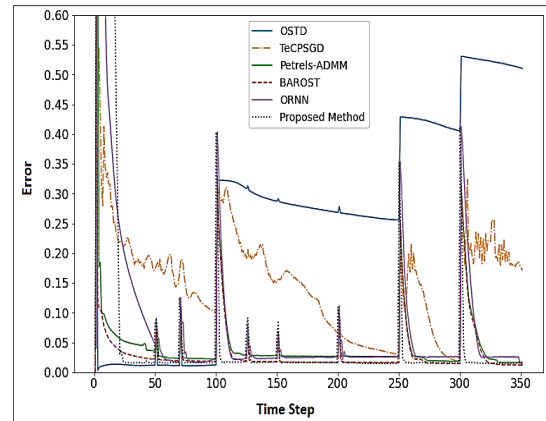


شکل ۹- عملکرد مدل پیشنهادی در گام‌های ۷۰ و ۲۰۰ مطابق شکل (۹) روش پیشنهادی همگرایی سریع‌تری در نقاط با گره‌های ناهنجار داشته است. این همگرایی سریع‌تر، باعث می‌شود تعداد هنجارهای کمتری به اشتباه، به عنوان ناهنجار تشخیص داده شوند. به عبارتی با کاهش نرخ مثبت کاذب، دقت تشخیص ناهنجاری افزایش پیدا خواهد کرد. همانگونه که ادعا شد، انتظار می‌رود رویکرد ارائه شده، زیرفضای اصلی را صحیح‌تر از سایر رویکردها در نقاط تغییر زیر فضا (گام‌های زمانی ۱۰۰، ۲۵۰ و ۳۰۰) پیش‌بینی کند. برای بررسی این موضوع، شکل (۱۰) خطای باقیمانده روش‌های مختلف را در نقاط تغییر زیرفضا نشان می‌دهد.



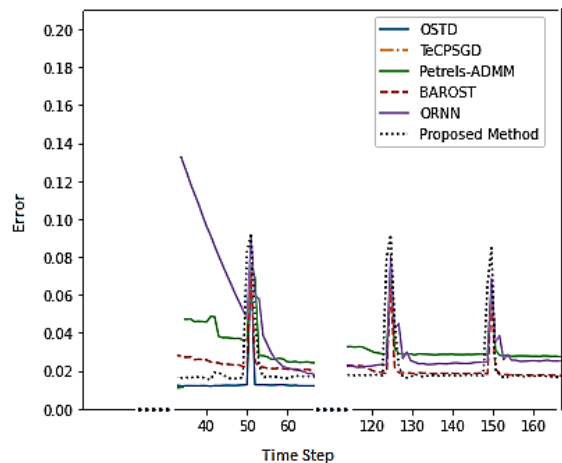
شکل ۱۰- عملکرد مدل پیشنهادی در گام‌های ۱۰۰، ۲۵۰ و ۳۰۰

با توجه به شکل (۱۰)، اگرچه روش OSTD پیش از زمان ۱۰۰ خطای میانگین پایین‌تر را به دست آورده است، اما با



شکل ۷- خطای بازسازی روش‌های مختلف

همانطور که از شکل (۷) مشخص است، روش‌های OSTD و TeCPSGD همگرایی خوبی نداشته‌اند و دچار خطای بالایی شده‌اند. برای روش‌های دیگر که به یکدیگر نزدیکند، لازم است نمودارها در نقاط ناهنجاری بررسی شوند. بنابراین حالت بزرگ‌نمایی شده از شکل (۷) برای گام‌های زمانی با یال ناهنجار در شکل (۸) نمایش داده شده است.



شکل ۸- عملکرد مدل پیشنهادی در گام‌های ۵۰، ۱۲۵ و ۱۵۰

همانطور که از شکل (۸) مشخص است روش پیشنهادی در مجموع همگرایی بهتری نسبت به روش‌های دیگر داشته است. OSTD، گرچه در گام ۵۰ بهتر از بقیه روش‌ها عمل کرده است اما بعد از اولین تغییر زیرفضا در گام ۱۰۰ به حدی دچار انحراف شده که در گام‌های بعدی نیز نتوانسته خود را بازیابد. TeCPSGD نیز بدین دلیل که مبتنی بر مشتق مرتبه اول است، اعوجاج زیادی داشته و توان همگرایی پایینی دارد. نزدیکی روش‌های دیگر، بررسی بیشتر آن‌ها را در بقیه نقاط می‌طلبد. گام‌های ۷۰ و ۲۰۰

تطبيق پذیری یادگیرنده است که نتایج فوق این ادعا را تایید می‌کند.

جدول ۲- مقادیر سنجه F1 برای مقادیر مختلف δ در تشخیص ناهنجاری

حد آستانه	OSTD	TECPSGD	BAROST	PETRELS-ADMM	ORURN
۰.۵	۰.۳۴	۰.۴۵	۰.۶۸	۰.۶۱	۰.۷۹
۱	۰.۴۱	۰.۴۷	۰.۶۹	۰.۶۲	۰.۸۱
۱.۵	۰.۴۸	۰.۴۸	۰.۷	۰.۶۳	۰.۸۴
۲	۰.۴۳	۰.۴۸	۰.۶۷	۰.۶۱	۰.۸۵
۲.۵	۰.۴۶	۰.۴۶	۰.۶۲	۰.۵۷	۰.۸۵
۳	۰.۴۷	۰.۴۵	۰.۶	۰.۵۵	۰.۸۵
۳.۵	۰.۴۷	۰.۴۵	۰.۵۹	۰.۵۳	۰.۸۳
۴	۰.۴۱	۰.۴۴	۰.۵۵	۰.۵۰	۰.۸۲

۵- نتیجه گیری

در این مقاله رویکردی مبتنی بر شبکه عمیق بازگشتی برای شناسایی ناهنجاری در داده گرافی پویا ارائه شد که در مقابل نویز استوار است و به صورت برخط عمل می‌کند. یکی از مهم‌ترین ویژگی‌های روش پیشنهادی، بی‌ناظر بودن آن است که ضرورت استفاده از داده برچسب‌دار را که بسیار پرهزینه و چالش‌برانگیز است، حذف می‌کند. ارزیابی روش پیشنهادی از دو منظر توانایی تطبيق با تغییرات و تشخیص ناهنجاری انجام شد. منظر اول به واسطه بررسی همگرایی یادگیرنده در نقاط تغییر و ناهنجاری و با استفاده از سنجه خطای باقیمانده بررسی شد و توانایی تشخیص ناهنجاری نیز به واسطه سنجه معروف F1، مورد ارزیابی قرار گرفت. باتوجه به ارزیابی‌های انجام شده روش پیشنهادی هم در همگرایی یادگیرنده و هم در تشخیص ناهنجاری به طور مطلوب‌تری عمل کرده است.

برای پژوهش‌های آتی بر روی مدل‌سازی تغییرات زیرفضا و تطبيق پذیری بهتر آن می‌توان کار کرد. همچنین در تشخیص ناهنجاری نیز روش پیشنهادی را تقویت کرد.

اولین تغییر زیر فضا در این زمان، خطای آن افزایش یافته و لذا توان تطبيق با تغییرات را نداشته است. همانطور که مشاهده می‌شود در سایر نقاط تغییر زیرفضا (۲۵۰ و ۳۰۰) نیز این روند تکرار شده است. روش TeCPSGD نیز به واسطه بهره‌گیری از مشتق مرتبه اول، توان تطبيق کمتری داشته است. در این میان روش پیشنهادی در مقایسه با روش‌های دیگر حتی روش‌های بازگشتی و مرتبه دوم سریع‌تر همگرا شده و تطبيق پذیری بالاتری داشته است. در واقع این نتیجه، دستاورد استوارسازی شبکه، علاوه بر بهره‌گیری از حافظه و روابط برگشتی آن است.

همانطور که نتایج این بخش نشان می‌دهد، روش پیشنهادی در مقایسه با بقیه روش‌ها سریع‌تر همگرا شده است. این امر به دلیل تطبيق پذیری بیشتر آن در مواجهه با تغییرات داده است. تطبيق پذیری بیشتر مدل که حاصل یادگیری برخط است و همچنین استواری آن، منجر به انجام بهتر فرآیند یادگیری شده است. انتظار می‌رود این بهبود، دستاورد دیگری هم داشته باشد و آن دقت بیشتر در تشخیص ناهنجاری است. چرا که مدلی که دیرتر با تغییرات تطبيق پیدا کند به اشتباه موارد بیشتری را به عنوان ناهنجار تشخیص داده و دقت کمتری خواهد داشت. در ادامه، در بخش ۴-۴ عملکرد روش‌های مختلف در تشخیص ناهنجاری، مورد بررسی و تحلیل قرار می‌گیرند.

۴-۴ ارزیابی تشخیص ناهنجاری

رویکرد پیشنهادی داده‌های ناهنجار را در ماتریس S ذخیره می‌کند. برای ارزیابی عملکرد مدل در تشخیص ناهنجاری، با اعمال تابع Z-Score با حد آستانه‌های متفاوت بر روی ماتریس S، ناهنجاری‌ها شناسایی می‌شوند. سپس برای سنجش عملکرد، سنجه F1 که میانگین هارمونیک دقت و فراخوانی است استفاده می‌شود. جدول ۲ مقادیر این سنجه را برای روش‌های مختلف نشان می‌دهد.

همانطور که نتایج جدول فوق نشان می‌دهد، روش پیشنهادی در تشخیص ناهنجاری در سطوح آستانه مختلف، از عملکرد بهتری برخوردار می‌باشد. کارایی بیشتر روش پیشنهادی در تشخیص ناهنجاری، حاصل استواری و

مراجع

- [1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys*, Vol. 41, No. 3, 2009, pp. 1–58.
- [2] L. T. Thanh, N. V. Dung, N. L. Trung, and K. Abed-Meraim, "Robust Subspace Tracking with Missing Data and Outliers: Novel Algorithm with Convergence Guarantee," *IEEE Transactions on Signal Processing*, Vol. 69, 2021, pp. 2070–2085.
- [3] Y. Chi, Y. C. Eldar, and R. Calderbank, "PETRELS: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Transactions on Signal Processing*, Vol. 61, No. 23, 2013, pp. 5947–5959.
- [4] J. He, L. Balzano, and A. Szlam, "Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1568–1575.
- [5] P. Narayanamurthy and N. Vaswani, "Provable dynamic robust PCA or robust subspace tracking," *IEEE Transactions on Information Theory*, Vol. 65, No. 3, Mar. 2019, pp. 1547–1577.
- [6] A. Sobral, S. Javed, S. K. Jung, T. Bouwmans, and E. H. Zahzah, "Online Stochastic Tensor Decomposition for Background Subtraction in Multispectral Video Sequences," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 946–953.
- [7] M. Amoozegar, B. Minaei-Bidgoli, H. Fane, and M. Rezaghi, "A Drift-Aware Online Learner for Anomaly Detection from Streaming Data," *Computational Intelligence in Electrical Engineering*, 2021.
- [8] M. Mardani, G. Mateos, and G. B. Giannakis, "Subspace learning and imputation for streaming big data matrices and tensors," *IEEE Transactions on Signal Processing*, Vol. 63, No. 10, 2015, pp. 2663–2677.
- [9] H. Kasai, "Fast online low-rank tensor subspace tracking by CP decomposition using recursive least squares from incomplete observations," *Neurocomputing*, Vol. 347, 2019, pp. 177–190.
- [10] Z. Li, Y. Wang, Q. Zhao, S. Zhang, and D. Meng, "A Tensor-Based Online RPCA Model for Compressive Background Subtraction," *IEEE Transactions on Neural Networks and Learning Systems*, 2022, pp. 1–15.
- [11] M. M. Salut and D. V. Anderson, "Online Tensor Robust Principal Component Analysis," *IEEE Access*, Vol. 10, No. May, 2022, pp. 69354–69363.
- [12] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," 2019. Accessed: Nov. 20, 2019. [Online]. Available: <http://arxiv.org/abs/1901.03407>
- [13] X. Ma, J. Wu, S. Xue, J. Yang, Q. Z. Sheng, and H. Xiong, "A Comprehensive Survey on Graph Anomaly Detection with Deep Learning," Vol. 14, No. 8, Jun. 2021, Accessed: Aug. 15, 2021. [Online]. Available: <https://arxiv.org/abs/2106.07178v1>
- [14] G. Pang, C. Shen, L. Cao, and A. Van Den Hengel, "Deep Learning for Anomaly Detection: A Review," *ACM Computing Surveys*, Vol. 54, No. 2, 2021.
- [15] A. Berroukham, K. Housni, M. Lahraichi, and I. Boulfrifi, "Deep learning-based methods for anomaly detection in video surveillance: a review," *Bulletin of Electrical Engineering and Informatics*, Vol. 12, No. 1, 2023, pp. 314–327.
- [16] Y. Cui, Z. Liu, and S. Lian, "A Survey on Unsupervised Anomaly Detection Algorithms for Industrial Images," *IEEE Access*, Vol. 11, 2023, pp. 55297–55315.
- [17] S. Li, X. Yang, H. Zhang, C. Zheng, and Y. Yi, "DSGRAE: Deep Sparse Graph Regularized Autoencoder for Anomaly Detection BT - Machine Learning for Cyber Security," Y. Xu, H. Yan, H. Teng, J. Cai, and J. Li, Eds., Cham: Springer Nature Switzerland, 2023, pp. 254–265.
- [18] Y. Qi, Y. Wang, X. Zheng, and Z. Wu, "Robust feature learning by stacked autoencoder with maximum correntropy criterion," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2014, pp. 6716–6720.
- [19] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Vol. Part F1296, 2017, pp. 665–674.
- [20] Y. Liao, Y. Wang, and Y. Liu, "Graph regularized auto-encoders for image representation," *IEEE*

Transactions on Image Processing, Vol. 26, No. 6, 2016, pp. 2839–2852.

[21] S. Saurav et al., “Online anomaly detection with concept drift adaptation using recurrent neural networks,” in Proceedings of the ACM India Joint International Conference on Data Science and Management of Data - CoDS-COMAD '18, New York, New York, USA: ACM Press, 2018, pp. 78–87.

[22] P. Zhou and J. Feng, “Outlier-robust tensor PCA,” in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, pp. 3938–3946.

[23] L. I. Kuncheva and I. Žliobaitė, “On the window size for classification in changing environments,” Intelligent Data Analysis, Vol. 13, No. 6, 2009, pp. 861–872.

[24] W. Feng, N. Guan, Y. Li, X. Zhang, and Z. Luo, “Audio visual speech recognition with multimodal recurrent neural networks,” in 2017 International Joint Conference on neural networks (IJCNN), 2017, pp. 681–688.

[25] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, “Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery,” IEEE Signal Processing Magazine, Vol. 35, No. 4, 2018, pp. 32–55.

[26] F. Arrigoni, B. Rossi, P. Fragneto, and A. Fusiello, “Robust synchronization in SO (3) and SE (3) via low-rank and sparse matrix decomposition,” Computer Vision and Image Understanding, Vol. 174, 2018, pp. 95–113.

[27] Y. Cui, C. Surpur, S. Ahmad, and J. Hawkins, “A comparative study of HTM and other neural network models for online sequence learning with streaming data,” in 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 1530–1538.

[28] S. Saurav et al., “Online anomaly detection with concept drift adaptation using recurrent neural networks,” in Proceedings of the ACM India Joint International Conference on Data Science and Management of Data - CoDS-COMAD '18, 2018, pp. 78–87.

[29] M. Mardani, G. Mateos, and G. B. Giannakis, “Dynamic anomalography: Tracking network anomalies via sparsity and low rank,” IEEE Journal on Selected Topics in Signal Processing, Vol. 7, No. 1, 2013, pp. 50–66.

[30] M. Amoozegar, B. Minaei-Bidgoli, M. Rezaghi, and H. Fanaee-T, “Extra-adaptive robust online subspace tracker for anomaly detection from streaming networks,” Engineering Applications of Artificial Intelligence, Vol. 94, Sep 2020 p. 103741.

[31] A. Sobral, T. Bouwmans, and E. Zahzah, “Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos,” Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing, 2016.